Identification, characterization and validation of the genes involved in salt tolerance of the rice (*Oryza sativa* L. indica) landrace *Horkuch* and its segregating population using RAD-genotyping and RNA-sequencing under salinity stress

Ph.D. Thesis

Sabrina Moriom Elias

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY
FACULTY OF BIOLOGICAL SCIENCES
UNIVERSITY OF DHAKA
DHAKA, BANGLADESH

OCTOBER 2017

Identification, characterization and validation of the genes involved in salt tolerance of the rice (*Oryza sativa* L. indica) landrace *Horkuch* and its segregating population using RAD-genotyping and RNA-sequencing under salinity stress

BY

Sabrina Moriom Elias

A DISSERTATION SUBMITTED TO THE UNIVERSITY OF DHAKA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN BIOCHEMISTRY AND MOLECULAR BIOLOGY

FACULTY OF BIOLOGICAL SCIENCES
DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY
UNIVERSITY OF DHAKA
DHAKA, BANGLADESH

REGISTRATION NUMBER: 83
SESSION 2012-13

CERTIFICATE

This is to certify that Sabrina Moriom Elias has conducted her thesis work entitled "Identification, characterization and validation of the genes involved in salt tolerance of the rice (*Oryza sativa* L. indica) landrace *Horkuch* and its segregating population using RAD-genotyping and RNA-sequencing under salinity stress" under my supervision for fulfillment of the degree 'Doctor of Philosophy in Biochemistry and Molecular Biology' from the University of Dhaka. Furthermore, her works are also co-supervised by Dr. Harkamal Walia, Associate Professor, University of Nebraska Lincoln, NE, USA. I would like to acknowledge with gratitude Dr. Thomas Juenger, Professor, University of Texas, Austin and Dr. Abdelbagi Ismail, Head, Genetics and Biotechnology Division, International Rice Research Institute (IRRI) for also hosting Sabrina for parts of her research work and providing their valuable inputs and criticisms. The work or any part of the thesis has not been submitted anywhere for any other degree.

Supervisor

.....10/10/2017.....

Date:

Zeba Islam Seraj, Ph.D.
Professor,
Department of Biochemistry and
Molecular Biology,
University of Dhaka
Ph: +880-2-9661920x7647
zebai@du.ac.bd

Harkamal Walia

Co-supervisor

Oct 9, 2017

.....

Date:

Harkamal Walia, Ph.D.
Associate Professor,
Department of Agronomy and Horticulture
University of Nebraska, Lincoln, USA
Ph: +1402-472-1162
hwalia2@unl.edu

Dedications

To

My Parents

Dr. S. M. Elias and Mrs. Anwara Elias

&

My Country, Bangladesh

Acknowledgements

The aim of a PhD study should be how much you are contributing to the body of knowledge. I believe learning is a continuous process with no end, but in the completion of a study, we focus on how much contribution is made to make the existing knowledge more resourceful. I am afraid the acknowledgment section is very long since it would not be possible for me to complete this dissertation without support from my excellently supportive mentors, peers, friends and family. My great admiration and thanks to Almighty Allah for making me capable to complete the study and submit this dissertation.

Prof. Zeba Islam Seraj, Ph.D, Department of Biochemistry and Molecular Biology, University of Dhaka, encouraged by her dedication to science and implementation of the basic findings, in reality, was the moto behind my decision to pursue the PhD with the salt tolerance of rice. These particulars sound interesting to me as it has the scope of both applied as well as basic investigations. My heartiest gratitude to my Supervisor for her guidance, direction and support all through my PhD session.

I am greatly indebted to my mentors Dr. Thomas Juenger, Professor, Department of Integrative Biology, University of Texas, Austin, USA, Thanks for the great support from designing the experiment to conducting the study as well as your help during adjustment in an unknown continent, without which it would be never possible. My appreciation and thanks to my co-supervisor Dr. Harkamal Walia, Associate Professor, Department of Agronomy and Horticulture, University of Nebraska Lincoln, Lincoln, USA for his kind suggestions in designing experiments and supports for using the Nebraska innovation campus facility for image phenotyping. My heartfelt thanks to Dr. Abdelbagi Ismail for being supportive and his wise and directional advice during my part of the study in International Rice Research Institute,(IRRI), Los Banos, Laguna, for providing supports to complete the comparative analysis and allele mining study.

I would like to thank Monsanto's Beachell Borlaug International Scholars' program for providing me the funds for my PhD study and making an excellent platform to build the network by meeting experts in this field. This fellowship program was a great opportunity for me as it was designed to get professional leadership training and also to participate in prestigious conferences to build the network of peers. Specifically I would like to thank Dr. Edward Runge, Program Director & Judging Panel Chair, Monsanto's Beachell-Borlaug International Scholars' Program and Professor Emeritus, Department of Soil & Crop Sciences, Texas A & M University, College Station, TX, USA, for being extremely supportive, appreciative and inspiring through his words and deeds.

I would like to extend my heartfelt thanks to NSF-PEER (Partnerships for Enhanced Engagement in Research) grant awarded to Dr. Seraj, Dr. Juenger and Dr. Ismail for

funding, of which this PhD study was also a part of. I am much thankful to my colleagues Taslima Haque and Samsad Razzaque for sharing their part of works in the PEER project with me to complete this study. Thanks to Dr. Sazzadur Rahman from BRRI and the BRRI physiology division for providing help in carrying out the phenotyping works.

The Dhaka University team, Sudip Biswas, fellow PhD student Sumaiya Farah Khan, Nazrul Islam, Shamim Hossain, Alamin Hossain, Robin Sorkar, Raju Ahmed, please accept my heartfelt appreciation and thanks. I am profoundly thankful to the UT Austin team and UNL team, specifically Tierney Logan, Dr. John Lovell, Dante, Malachy Campbell, Aaron Schmitz, Alexandar, Sarah, and Dr. Vincent Stroetger (manager, Nebraska Innovation Campus green house). Special thanks to Martha Row for making the complicated things easier in UNL regarding formalities of foreign transportation of seeds as well as transports of samples between institutes. Thanks a lot to the IRRI team specifically Dr. John Damien Platten for sharing the rice donor sequence data, Rexie, Joshua Dizon and Mr. Froilan.

My thanks to our chairpersons, Department of Biochemistry and Molecular Biology, University of Dhaka, Dr. Haseena Khan and Dr. Mohammad Sayedul Islam for being extremely supportive during my PhD study. My heartfelt gratitude to Dr. Imdadul Haque, Dean, Faculty of Biological Sciences, University of Dhaka and his office staffs for their support in administrative paper-works.

I would like to acknowledge Texas Advanced Computing Center (TACC), Holland Computing Center in USA and Computing and Archiving Research Environment (CoARE) project, ASTI server, The Philippines the three servers where I needed to complete the big data analyses during different period of the study

Pursuing a doctoral study for long period, proper compilation, and presenting the gist in the correct order was not as easy as I expected before entering this journey. Frustration and depressions were there but there also existed the word reborn, which helped get out of frustrations and completing the works. Since the study was conducted involving four different institutes, adjustment in new places and interaction with new people was a major part of the period apart from the study. My time was made enjoyable in large part due to the many friends and groups that became a part of my life. I am grateful for the time spent with my friends (junior and senior) during my long stay in Lincoln, Nebraska. Without their mental support, it would not be possible to continue the study. Also the supports of unknown angels who became friends later, to ease my first steps in the unknown country in Texas, USA and Los Banos, The Philippines will always be remembered with love and gratefulness based on which I finished up my degree, and for many other people and memories. My time in this period was also enriched by the graduate BSA (Bangladesh Student Association) in UNL and AFSTRI (The Association of Fellows, Scholars, Trainees and Resident of IRRI) group in IRRI. Heartfelt thanks to all PBTLAB members especially Jewel and Sarah for their kind help. My gratitude to Dr. Richard Malo for lending his laptop (Snowy) without which completing my dissertation would be difficult and thanks to both him and Farida for supporting my mental journey through the long process. Due to space limitation, I might

have unintentionally missed anyone, rest assured that my heartfelt personal thanks will reach you.

Finally, I have no words to thank my family members who have always wished to see me successful and prosperous by providing continuous moral support. My dedication of this thesis is to my parents and my country. My father, Dr. S.M. Elias, was my first encouragement to pursue higher studies so that I can obtain his exact name. He had guided me and encouraged me always to explore this excellent field of interdisciplinary knowledge, I hope you are watching me from the other world and will be happy seeing this accomplishment. My mother, Mrs. Anwara Elias, I really appreciate your patience and support throughout the study, without your support, it would never be possible you know. I am thankful to my brothers Shaquil Elias and Shoeb Elias and their family for being supportive all through my life and career.

I entered the journey to achieve another academic progress, but along with academic knowledge, this period fulfilled me with way more lessons and accomplishments. I always wanted to work with the resources of Bangladesh. MBBISP with their excellent scheme of PhD study encouraging collaboration between developing and developed world made this possible. It helped me accessing the cutting edge technologies that we are still building to study the diverse rice germplasm resources in Bangladesh. My beloved country, this study is dedicated to you. I hope the proper implementation of the findings of the study will prove useful for the growing population of the world.

Sincerely
Sabrina Moriom Elias

Abstract

Plants being sessile organisms, possess excellent means of sensing and regulation to adjust with the environment and its change. Molecular exploration of indigenous landraces adapted to various stresses in such environments can therefore be useful in understanding their mechanism of adaptation and tolerance. Due to climate change the world is losing more and more cultivable lands affected by drought, flooding salinity, etc., leading to food insufficiencies. The rice landrace *Horkuch (Oryza sativa*) is adapted to the saline conditions in the southern coast of Bangladesh and maintains efficient photosynthesis and detoxification under salt stress, but does not have enough yield to meet the farmers' needs. Therefore, this study was aimed at finding the genetic loci and genes responsible for this tolerance and their heritability in populations derived from it. To this end, reciprocal crosses were made from salt tolerant *Horkuch* and high yielding but sensitive IR29. A set of the reciprocal $F_{2:3}$ population containing 300 individuals was genotyped using DArTSeq™ for discovering SNP markers and phenotyped under salt stress, both at seedling and reproductive stages when rice is most sensitive. DArTSeq™ is a restriction site associated, species-based intelligently designed next generation sequencing method which allowed reliable discovery of a large number of SNPs. A robust combined Linkage map was generated from the polymorphic SNPs identified from both populations followed by identification of multiple novel QTLs. QTLs for root length, potassium content and sodium content at seedling stage as well as spikelet fertility, days to maturity, seed breadth, plant height, grain weight, harvest index, etc., have been identified at the reproductive stage. Selected plants from the population, containing several phenotypic QTLs (pQTLs) from both seedling and reproductive stage were identified. These can therefore be used for introgression and pyramiding of multiple tolerance loci into commercial genotypes.

A subset of the same combined population were studied for the gene expression patterns under 150mM stress at seedling and reproductive stage using 3′ tag-based RNAseq method. Certain correlations were observed between the superior physiological performance of the tolerant versus sensitive progenies and the differential pattern of gene expression in leaf and root tissues under salt stressed and control. Potential list of candidate genes were generated based on the differential expression related to G protein coupled receptor signaling pathway, carbohydrate metabolic process, potassium transport, calcium dependent signaling, water deprivation, better ROS scavenging, growth and photosynthesis.

The DNA and expressional polymorphism were then associated to discover expression QTLs (eQTLs), where the relative importance of specific genes could be deduced. Multiple cis and trans-eQTLs were identified for seedling and reproductive stages in leaf and roots with and without salt. eQTL hotspots were identified mostly in Chr 1, 2, 3, 5, 9 and 10 which were then

used to construct networks of co-expressed genes. Hotspots harboring genes involved in photosynthesis, response to salt stress and water deprivation were identified. Moreover overlapping eQTLs in the pQTL regions were detected and significant genes underlying the relevant phenotypes can be proposed. For example, 2 cis-eQTLs for calcium transporting ATPase and calmodulin related calcium sensor protein were overlapped with the potassium content QTL region in seedling stage suggesting importance of calcium sensing in sodium-potassium homeostasis.

An image based non-destructive automated and continuous phenotyping over 3 weeks of salt stress was carried out on a selected F_3 and F_5 sub-populations followed by QTL identification for the digital traits and relative growth rates. Since the digital phenotyping gives a longitudinal result rather than an end point one, it could separate the early and late responses to salt stress as well as correlate with the manually identified pQTLs from F_3. Genes for cation transport, chloride transport, stomatal movement, auxin stimulus and calcium signaling was identified in the QTL regions.

In-depth analysis of allelic diversity at key loci in salt tolerant donor genomes was carried out for some traits thought to be linked to salt tolerance that helped to determine whether unique accessions like *Horkuch* have better and novel alleles. Specific missense variants were identified for some genes involved in ROS scavenging, water deprivation, etc. For example, a deletion was observed in the 3′ downstream region of OsRCI2-6 mostly in salt tolerant varieties. This comparative and integrative study provided a comprehensive landscape of the genetic determinants of gene expression under salinity stress in rice responsible for tolerance. The positive effect of the *IR29* nuclear allele (cytoplasmic effect) for yield trait QTLs and the strongly positive *Horkuch* nuclear allele effect for QTL models of desired traits e. g., K, will help define the combinations of cytoplasm and nuclear-donor materials for breeding plants with desired traits. The identified network of genes in the progenies also support the segregation of better photosynthetic and detoxifying ability of *Horkuch* in the breeding lines. Implementation of this knowledge in modern breeding technology by selection, breeding and genetic modifications can facilitate production of high yielding salt tolerant crop varieties, which are adaptable and cultivable in the changing environment.

Table of Contents

	List of Tables
Table 3.1	Formulas used for calculation of physiological traits
Table 3.2	Seedling stage phenotyping data statistics
Table 3.3	Reproductive stage phenotyping data statistics
Table 3.4	QTLs identified in seedling stage
Table 3.5	QTLs identified in reproductive stage
Table 3.6	Plants containing highest number of QTLs
Table 4.1	List of primers used in the library preparation
Table 4.2	Primer sequences for real time qPCR validation
Table 4.3	List of potential candidate genes for further investigation
Table 4.4	Number of eQTLs identified in the genome with respect to spatial and temporal factors
Table 4.5	Hotspot regions harboring major networks for trans-eQTLs
Table 4.6	Cis-eQTL hotspots genes and their biological functions
Table 4.7	Significant cis-eQTLs as a list of candidate genes
Table 4.8	eQTLs associated with pQTL regions
Table 5.1	Specification of the cameras used for taking images
Table 5.2	Formula and description of the digital traits used in the study
Table 5.3	Formula and description for the growth rate related traits
Table 5.4	Adjusted mean of the traits and percent decrease r increase under salinity in parents, F_3 , progenies and F5 progenies
Table 5.5	List of selected candidate genes identified from association between image QTL regions and expression QTL regions
Table 6.1	List of the genomes used in the study
Table 6.2	Mapping statistics of the BAM files
Table 6.3a	Number of variants found in the combined haplotype calling of 14 genomes
Table 6.3b	Number of variants found in the combined haplotype calling of 14 genomes after filtering
Table 6.4	List of Alleles in specific SNP markers
Table 6.5	List of INDELS in the differentially expressed gene regions associated with cis-eQTLs in chr 10
Table 6.6a	List of Homozygous SNPs in *Horkuch* genome compared to *IR29* in the differentially expressed gene regions associated with cis-eQTLs in chr 10
Table 6.6b	List of significant variant effect containing Heterozygous SNPs in *Horkuch* genome compared to *IR29* in the differentially expressed gene regions associated with cis-eQTLS in chr 10

	List of Figures
Fig 1.1	Rice Production and consumption scenario of top 5 rice producing countries (according to World Rice Statistics, IRRI, 2017)(Values are in million metric ton)
Fig 1.2	Soil salinity map of Bangladesh (BARC, 2009)
Fig 1.3(A)	Schematic experiment Design with thesis organization
Fig 1.3(B)	The experiment design in details
Fig 2.1	Diversity in the salt tolerance of various species, shown as increases in shoot dry matter
Fig 2.2	Rice salt tolerance in different developmental stages (IRRI rice knowledge bank)
Fig 2.3	Rice growing seasons in Bangladesh
Fig 2.4	Understanding of symplastic and apoplastic route
Fig 2.5	Transport of ion emphasizing the transporters (adapted from Reddy et al, 2017)
Fig 2.6	Approaches for development of salt tolerant rice (adopted and redrawn from Reddy et al, 2017)
Fig 2.7	Main mapping populations for self-pollinating species (Adopted from Collard et al , 2005)
Fig 2.8	Expression analysis method timeline
Fig 2.9	Schematic representation of cis and trans eQTL
Fig 2.10	Classification of lights according to spectrum (Fahlgren et al, 2015)
Fig 2.11	Four major type of images taken for plant phenotyping
Fig 3.1	Seedling stage phenotyping
Fig 3.2	Reproductive stage phenotyping
Fig 3.3	Evaluation of qualitative traits in F_2 without any treatment.
Fig 3.4	Correlation among the traits in F_2 population based on cytoplasm (A) and tolerance (B). Only the tolerance trait in 3.4 (B) was measured from salt stressed F_3 population.
Fig 3.5	Distribution of seedling stage phenotyping data in reciprocal populations (here Hor = Horkuch♀ and IR = IR29♀)
Fig 3.6	Distribution of reproductive stage phenotyping data in reciprocal populations
Fig 3.7	Correlation between traits in seedling stage
Fig 3.8	Principal component analysis among the physiological traits. (A) Seedling stage, (B) Reproductive stage)
Fig 3.9	Correlation between traits in reproductive stage
Fig 3.10	Genetic map constructed from the combined Horkuch♀ and IR29♀ cross population
Fig 3.11(A)	Map of identified QTLs of seedling stage in chromosomes
Fig 3.11(B)	Map of identified QTLs of reproductive stage in chromosomes
Fig 3.12 (A)	Selected F_5 plants have superior performance under stress in seedling stage
Fig 3.12 (B)	Selected F_5 plants have superior performance under stress in reproductive stage

Chapter 1

Introduction

Life requires continuous internal homeostasis against fluctuations of the environment in order to survive. Disruption in such homeostasis due to external challenges often causes living organisms to experience stress. Since plants are sessile organisms, they either develop adaptive responses to re-establish homeostasis or embrace death. Excess salinity in soil or water is an environmental stress (abiotic) for plants which results in water deprivation as well as salt accumulation at toxic levels in sensitive organs, disrupting homeostasis. This thesis focuses on the elucidation of the salt tolerance determinants, their characterization and to some extent, validation of these determinants in a salinity stress tolerant Bangladeshi rice landrace *Horkuch* and its breeding population with sensitive but high yielding *IR29*. The study was undertaken at the two most saline sensitive developmental stages of rice, seedling and reproductive. The breeding population was studied in order to understand the inheritance of the tolerance determinants of the donor *Horkuch* in its crossed progenies as well as to observe combination of alleles from both parents which would give high yields under saline stress.

1.1 Climate change and food scarcity of the growing population

The population in our green planet earth is increasing so fast that scientists are now exploring options for colonization in another planet initiating the "Mars One" project. Between 2010 and 2014 , the world's population grew at a rate of 1.2% per annum, and on its current trajectory, the population is expected to reach 8.1 billion in 2025 and 9.6 billion in 2050 (world population datasheet 2014 by UN). These additional people will need food, fiber and fuel for their livelihood. Moreover there are extra demands for resource intensive agricultural products like livestock and dairy products, exerting further pressure on land, water and biodiversity resources (UN World Economic and Social Survey 2013). Agricultural production has to increase by 60% over the next 40 years to meet the rising demand for food (OECD and FAO 2012). Unfortunately, our world has come into a situation that 40% of the arable land has been degraded to some degree and will be further affected adversely by climate change (IFAD).

Excess salinity is one of the major abiotic stresses, along with drought, temperature rise, etc., negatively affecting food production. Together the abiotic and biotic stress factors result in 65-87% crop failures worldwide as estimated by (Buchanan et al. 2002). Land clearing, excessive irrigation as well as attributes related to climate change, like salt intrusion into coastal region resulting from sea-level rise and increased number of storms are making a considerable proportion of cultivable land more saline (Smajgl et al. 2015). About 45 million ha (20%) of the current irrigated land and more than 6% of the total lands are salt affected and 1.5 million ha are being added each year. (Munns and Tester 2008). Lack of available agricultural land due to gradually increased salinity as a consequence of continued unsustainable cultivation practices and climate changes has raised concern about securing enough food for the world

population (Wassmann et al. 2009; Qadir et al. 2014). Not only soil salinity, but increasing levels of salinity in ground water has made the situation worse. This has necessitated the use of saline water in irrigation which also imposes the necessity for producing more salt tolerant crops.

1.2 Excess salinity: dissecting the tolerance determinants

Salts are compound substances e.g. NaCl, $CaCl_2$, KCl, $MgCl_2$, $MgCO_3$, etc., but plants uptake them as soluble ions. Concentration of various salts can be elevated in soil, e.g. irrigation can lead to high concentration of Calcium and Magnesium carbonates. On the other hand, areas with geologically recent marine history with low enough rainfall to prevent leaching out of salts, soil can also be high in NaCl. Besides, saline soils often have toxic concentrations of weak acid boric acid. The major focus of the research for understanding salt tolerance mechanism is however based on elevated levels of NaCl, because it is the major salt in saline water. Moreover, excessive Na^+ prevents uptake of K^+ ions, the latter being essential for a plants survival. A soil is considered saline, when the electrical conductivity of its saturation extract (ECe) is above $4dSm^{-1}$ (USDA-ARS, 2008), which is equivalent to approximately 40mM NaCl and generates an osmotic pressure of approximately 0.2 MPa (Munns and Tester 2008).

Excess salinity stresses plants in 2 ways; (a) high concentrations of salts in the soil make it harder for roots to extract water causing internal dehydration or osmotic stress similar to drought and (b) the elevated ion concentrations in cells leading to toxicity particularly in photosynthetic cells perturbing homeostasis. These cause reduction of growth as well as transpiration rates and ultimately cause premature leaf senescence. Both osmotic and ionic components are likely to impact the yield which is the major concern for humankind. Three essential Na^+ transport systems have been extensively studied: (a) Na^+ efflux from roots to the rhizosphere, (b) Na^+ sequestration in vacuoles, and (c) Na^+ loading and unloading at the xylem, all of which contribute to lowering of shoot Na^+. Moreover knowledge on associated ROS production, signaling mechanisms, transcription factors, osmoprotectants, transporters and many more functionally related genes and their associated pathways have opened the door to pinpoint tolerance mechanisms. Usually tolerance is due to multiple mechanisms, involving multiple genes. Different combinations, levels of expression of these genes and their timing ultimately determine the tolerance response of a plant. It may be mentioned here that sources of such genes which play a role in conferring tolerance are mostly found in landraces, or traditional genotypes which have evolved to adapt to local environmental conditions over many years (Gamuyao et al. 2012). Therefore tolerance mechanisms need to be understood for implementation in conventional or new breeding technologies. Such implementation by selection, breeding and genetic modifications can facilitate production of high yielding salt

tolerant crop varieties, which are adaptable and cultivable to the changing environment. Currently available commercial genotypes usually have the high yielding phenotype, very few reported with tolerance to stress.

1.3 Strategies for developing salt tolerant plants

Most cereal crops are glycophytes that cannot tolerate salinity stress, though their tolerance level varies. By contrast, halophytes benefit from high salt conditions and grow in soil or a growth medium that has substantial amount of salt as reviewed by (Ismail and Horie 2017). Halophytes share many comparable aspects with their closely related glycophytes, though part of the tolerance mechanisms in halophytes depends on distinct structure of tissues and organs. For example, *Porteresia coarctata* (a halophyte) and *Oryza sativa* (a glycophyte) are very closely related and belong to the same grass family Poaceae. Therefore learning from the salt loving halophytes can be an important strategy (Shabala 2013). Another popular halophyte with nutritious seed is Quinoa.

Similarly there are tolerant glycophytic crops, which are endemic to salt prone region and in course of evolution have developed the mechanism to withstand salt, but they do not produce enough yield, may be due to their modified morphological and genome structure. Normally they require less leaching to remove salt from topsoil and can be used to reclaim saline and sodic soils (Lisa et al. 2004; Munns et al. 2006; Ismail and Horie 2017). These naturally salt tolerant cultivars can be used as salt tolerance donor in breeding programs. So another emerging approach is to dissect the underlying mechanisms of these naturally tolerant crops or landraces as they are referred to.

Several strategies have been adopted to develop salt tolerant cereals and economically important crops. Two main strategies now being utilized are breeding, to be more specific, Marker Assisted Selection (MAS) followed by backcrossing breeding (MABC) and transgenic approaches by overexpressing a desired gene, which is usually regulatory in nature (Wilson et al. 2005). Conventional breeding often takes about 10-15 years to release a new stable variety and sometimes with less than the expected results. In contrast, MAS-based precision breeding along with development of sequencing technologies are now reducing the breeding cycles with precise outputs. Moreover MABC ensures that the background parental high yielding quality is preserved in addition to the desired trait from the donor. Adaptation of statistical models during Recombinant Inbred Line (RIL) production and selection, e.g the $F_{2:3}$ concept has reduced the waiting period till complete homozygosity. Advancement in sequencing technologies and discovery of genome wide SNP markers with precision breeding has further facilitated QTL identification. To this toolbox, the addition of genotype-phenotype association through GWAS has further facilitated useful allele discovery for tolerance traits.

There has been tremendous development in last decade to understand the genetic mechanism of salinity tolerance and making transgenic plants with overexpression of responsible genes. The consumer acceptance issue with transgenic i.e., genetically modified Organisms (GMOs) and the polygenic nature of the salinity tolerance trait have shifted the focus towards the marker assisted breeding, QTL mapping and QTL cloning. In most crops, tolerance is developmentally regulated, several traits and many genes are involved with each trait. Knowledge of the processes that are hindered in specific developmental stages can facilitate targeting particular genes for tolerance. Differential expression analysis strategies have been developed and facilitated by technologies like microarray, RNAseq, etc. Advancement in sequencing technologies have paved the way for using whole genome information in pinpointing specific loci responsible for salinity tolerance. Integrating the SNP genotyping information with the expression variation by mapping Expression QTLs (eQTL) as well as Genome Wide Association Studies (GWAS) has emerged as a novel strategy to narrow down specific regions in the genome instead of the use of laborious and expensive fine mapping strategies. The goals of some of the emerging technologies is to combine the MAB and transgenic approaches to incorporate significant regions of the genome responsible for desired traits and to pyramid genes to get high yielding salt tolerant varieties.

Reliable high throughput phenotyping is necessary to validate the genetic basis of selection for overall tolerance, for dissecting specific traits associated with tolerance and subsequent breeding. Non-destructive imaging techniques are being developed for discovery of genes underlying plant development and stress response. In-depth analysis of allelic diversity at key loci for traits underpinning salt tolerance will help determine if unique accessions have better and novel alleles. Along with the adopted strategies researchers have also now started to focus on quality of the seeds under stress, which was often overlooked in earlier studies. Particular emphasis has also being made to evaluate breeding lines in farmer's fields (participatory breeding) which ensures selection and formal release by ensuring local adaptation and farmer preference.

1.4 Rice production and salinity

Since its domestication, about 8000 years ago, rice has played a part in building civilization, shaping societies and most of all, feeding the growing world (Elert 2014). On global average, 42% of the daily energy comes from cereal crops where rice alone contribute 19% followed by wheat i.e., 18% (Elert 2014). About 3 billion people depend on rice as a staple food which is about half of the world's population. The major rice producers of the world grow more than enough rice to feed their own people. The excess ends up as exports, livestock feed, seed stock or waste (spoiled during transport or storage). However, 5 countries in Asia, such as China, India, Indonesia, Bangladesh and Vietnam are the major producers, consumers and exporters according to FAO report 2017. (Fig 1.1). In 2010, approximately 154 million ha of rice were harvested worldwide, of which 137 million ha (88%) were in Asia - and 48 million ha (31%) were harvested in Southeast Asia alone (FAOSTAT, 2012). Nevertheless with the rise of the global population, a 26% increase in rice yield is predicted to be required to meet global demands in the next 25 years (IRRI, 2013). This signifies the importance of increase in rice productivity on salinized water.

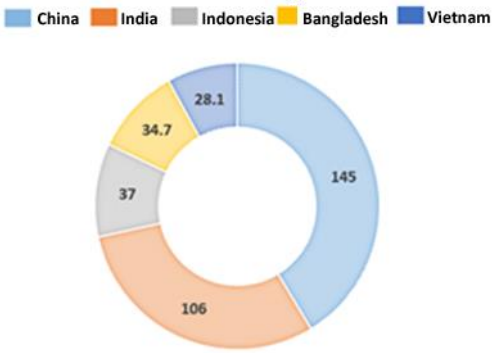


Fig 1.1 Rice Production and consumption scenario of top 5 rice producing countries (according to World Rice Statistics, IRRI, 2017) *(Values are in million metric ton)*

Rice *(Oryza)* belongs to the Poaceae family which has species spread all over the world like *Oryza sativa, Oryza glaberima, Oryza austreliensis etc.* The rice plant is generally salt sensitive, with low thresholds of 1.9–3.0 decisiemens per meter (dS/m) (19–30 mM), above which growth and yield are affected (Grattan et al. 2002). Despite this high sensitivity rice can adapt to saline ecosystems which results from rice's ability to grow well even in flooded soils, where the free salts are diluted and leached from the mostly shallow rooting zone (reviewed in (Ismail and Horie 2017). But the main concern of the producers and consumers is that yield of rice in the saline zone is low. Efforts like, leaching the field with fresh water, utilizing different effective land preparation techniques and applying organic materials to reclaim soil properties have been used but which proved expensive and non- economical. So, focus is now being given to genetic approaches like breeding, transgenic and marker assisted breeding in order to develop salt tolerant rice.

Bangladesh is not only the fourth largest rice producing country (WRS 2017), it is also the home of many genetically diverged rice germplasm. But farmers cannot afford to cultivate local salt tolerant rice due to their low yield. About 1 million of the available 9 million hectares of total cultivable land is in the saline prone region. Bangladesh is a land-strapped country, and any further increase in rice production will have to come either from hybrid rice or growth of rice in suboptimal soils such as the salinity affected areas of the coastal belt. Recent studies have shown that production of high yielding rice varieties in the country will decrease by 15.6% in coastal districts where soil salinity is predicted to exceed 4 (dSm^{-1}) by 2050 (Dasgupta et al. 2014). Figure 1.2 shows the increase of soil salinity barrier towards the mainland of the country from 1973 to 2009. Even modest increase in rice production for the moderate saline zones (0.3 Mha) would go a long way in ensuring food safety for the local small-holding farmers in the high salinity areas. Hence salt tolerant crops is a dire need for stabilizing the country's rice production.

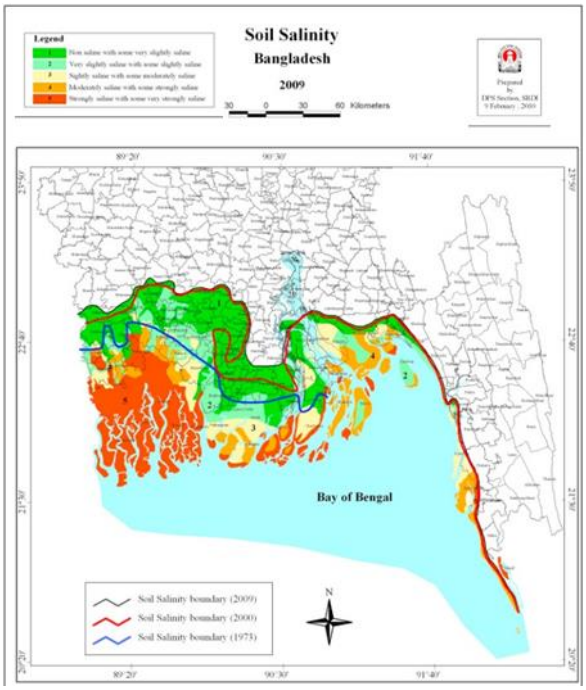


Fig 1.2 Soil salinity map of Bangladesh (BARC, 2009)

1.5 Development of salt tolerant rice: modern varieties meet landraces

A list of salt tolerant varieties released in Bangladesh, India and Philippines has been provided by (Ismail and Horie 2017). To date, about 32 salt tolerant high-yielding rice varieties with good grain quality have been developed and released for farmers (Mishra et al. 2012). These varieties are mostly developed using conventional breeding which took 10 -15 years for varietal release where the salt tolerant donors are mainly Pokkali or Nonabokra. Numerous studies have identified salinity tolerance QTLs. Among which *Saltol* (derived from Pokkali) and *SKC1* (derived from Nona bokra) came into limelight and later, *SKC1* was found to be associated with shoot potassium concentration, controlled by *OsHKT1;5* which resides in the Saltol region. Study of the genes in the Saltol region have identified clusters of genes involved in salinity tolerance (Walia et al. 2005; Nutan et al. 2017). More diverged salt tolerance donors are however needed to be utilized in modernized marker assisted selection-based breeding and for identification of novel tolerance loci. Due to lack of precision breeding the released varieties show low to moderate tolerance and some of them possess undesired traits like shattering, long

awn, poor grain quality, etc. In view of the predicted increase in salinity levels in Bangladesh, more tolerant varieties are needed, preferentially with a wider genetic base.

Whole genome re-sequencing and BAC clone sequencing of some rice landraces with low yield has shown that they possess novel genes or alleles that are different or absent in the modern germplasm. They have already been used as the source of tolerant QTLs for development of submergence, salt, drought or phosphorous deficiency tolerant high yielding varieties (Bonilla et al. 2002; Septiningsih et al. 2009; Gamuyao et al. 2012). Recent searches for salt tolerant donors from the rice diversity panels have found moderately to highly salt tolerant donors from all over the world, most of which fall in the indica subspecies (Platten et al. 2013; Rahman et al. 2016). These landraces are distinct with distinct alleles for tolerance. They could be narrowed down to some potential donors from Bangladesh, India and the Philippines region which are going under trial. Lisa et al. (2004) reported potential tolerant donors from the southern coast of Bangladesh among which the rice landrace *Horkuch* was foremost. *Horkuch*, is also popular with some farmers in the coastal areas in Satkhira, but it has low yields. *Horkuch* was initially identified to be salt tolerant at the seedling stage (Lisa et al. 2004). Its yield-related traits under stress were also found to be superior (Lisa et al. 2011). The gene expression pattern of *Horkuch* seedling in 150mM was analyzed and attempts were made to explain its superior physiological performance under salt stress. However further validation of the genes differentially regulated under salt stress in *Horkuch* is essential to understand the any unique mechanism of salinity tolerance in order to choose specific genes for future breeding strategies using *Horkuch* as a donor.

1.6 Aim of the study

In order to identify the tolerance loci in *Horkuch* and their heritability, a reciprocal population derived from *Horkuch* and high-yielding *IR29* was characterized in depth. Another aim of characterizing the population was to find progenies with transgressive variation where they performed better than either parent and to see the effect of cytoplasm (maternal effects) on the values of important traits. The latter was especially important because in order to breed better varieties we needed to characterize progenies with good yields under stress. For this reason, all analysis were performed at the two developmental stages, when rice is most sensitive, seedling and reproductive. Therefore, in this study, an in-depth analysis has been carried out on a reciprocally crossed population derived from the Bangladeshi salt landrace rice *Horkuch* and High Yielding rice *IR29*. According to the $F_{2:3}$ design both the *Horkuch*♀ and *IR*♀ population was genotyped in the F_2 population using DArT based SNP genotyping and phenotyped in the F_3 population. Moreover, a subset of the tolerant, sensitive and some intermediate varieties with

contrasting tissue, time points, developmental stage, treatment were subjected to RNA sequencing for differential expression analysis. Integration of the genotyping with morphological phenotyping values and expression values were used to map phenotypic QTL and expression QTLs respectively. A more detailed schematic diagram of the experiment design is provided in Fig 1.3.

The specific objectives of the study are:

1. SNP genotyping of the segregating F_2 population of both cross directions for generating large number of molecular markers across the whole genome using DArT-based markers to produce a combined linkage map.
2. Phenotyping of a randomly selected set (300 individuals) of F_3 population of both cross direction in both seedling and reproductive stage for quantitative estimation of specific traits known to correlate with salt tolerance and yield under stress.
3. Mapping phenotypic QTLs for both cross populations combining genotyping and phenotyping information.
4. Differential expression analysis of the selected subset (104 individuals) of the F_3 population using RNAseq with contrasting tissues (leaf, root), developmental stages (seedling, reproductive), treatments (salt stress, without stress) and time (24hr stress and 72 hr stress in case of seedling).
5. Mapping expression QTLs for both cross population and developmental stages combining genotyping and expression information.
6. Image based non-destructive phenotyping of a subset of the F_3, F_4 and F_5 population in 2 different years for generations of digital traits that focuses on growth under salt stress and can separate the osmotic and ionic stage of tolerance focusing on identification as well as validation of the determinants.
7. Allele based diversity mining of a portion of the identified salt tolerance determinants in the rice diversity panel for variant analysis in other tolerant and sensitive genomes.

The thesis includes results from all the objectives above. Analysis data for the linkage map construction, QTL mapping, and a part of the model fitting data for differential expression analysis was joint work, majority of which has been kindly shared by peer researchers of the author, who worked on the same project.

1.7 Experiment design and organization of the thesis

A schematic representation of the experiment design is provided in Fig 1.3 (A and B). The thesis is divided in the following chapters:

Chapter 1: Introduction, which is the current chapter and provides the background to the study and explains the aims and objectives.

Chapter 2: Review of Literature, which includes a concise literature review of rice salt tolerance mechanism, current scenario of identified candidate genes and strategies for the development of salt tolerant high yielding rice.

Chapter 3: A combined chapter for the manual phenotyping study followed by the SNP-based genotyping and reporting of the linkage map and mapped QTLs in the combined cross population data.

Chapter 4: Illustrates the differential expressed candidate genes identified from RNAseq expression analysis in contrasting time, tissue, developmental stage and treatment followed by mapping of expression QTLs combining genotyping and expression data.

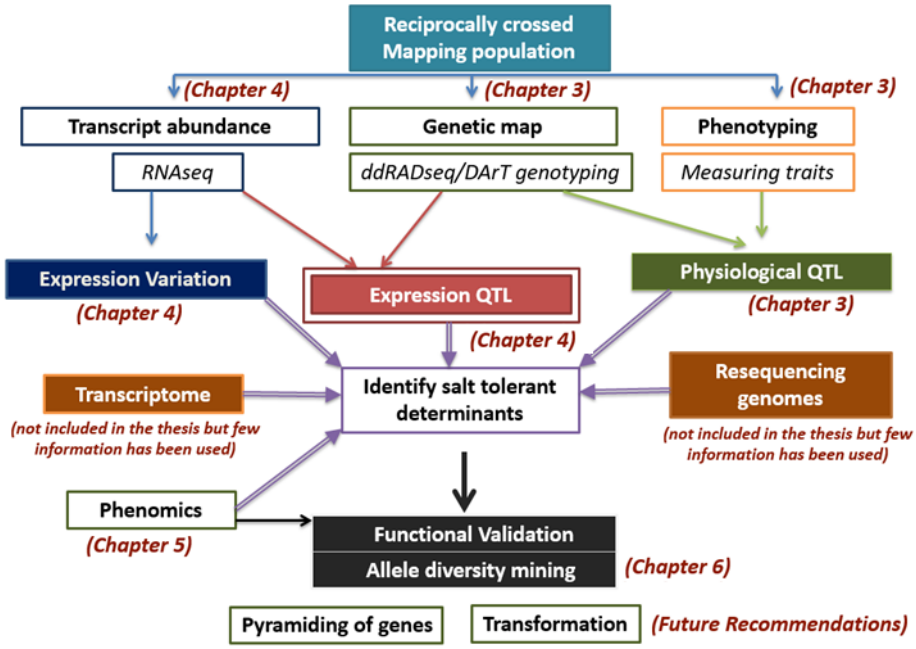


Fig: 1.3(A): Schematic experiment Design with thesis organization

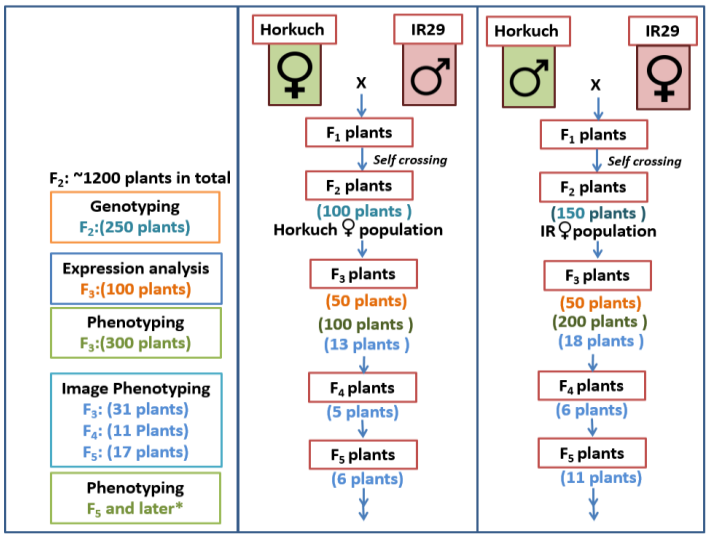


Fig: 1.3(B): The experiment design in details

**Not included in the thesis*

Chapter 5: Image based phenotyping for validation of tolerant plants and identification of digital traits correlated with salt stress for a closer look on the growth rate

Chapter 6: In silico validation of a small set of the identified candidates using allele diversity mining.

Chapter 7: A large number of information has been generated from the study which needs validation as well as linking to metabolic pathways as follow-up work. The last chapter includes conclusion and future prospective of the study.

1.8 Beauty of this experiment

This study on reciprocal cross population of salt tolerant *Horkuch* and High yielding IR29 is of immense significance from many aspects, as pointed below

1. A comparative approach has been used in the study. Since tolerance and sensitive progenies have been derived from the same segregating population the system here is biologically very similar but differ in the outcome of their stress response, the side by side comparisons made in differential expression and other studies have the potential to pinpoint correct targets.
2. This study is also integrative in the sense that different developmental stages of the stress response were studied in the same system. Moreover spatial and temporal differences in tissue and timing of stress were also considered. Most importantly it includes the data from the non-stressed condition which can exclude the false positive tolerance determinants.

3. Most studies in rice have not provided detail screening and expression information on reproductive stage which is important with respect to yield and long term survival under salt stress. The study deals with yield under salinity stress.
4. Reciprocally crossed population was considered in the study, i.e., two populations [*Horkuch*(♀) and *IR29*(♀)] data were combined which provided important information regarding maternal cytoplasm on specific traits.
5. $F_{2:3}$ design has been followed for QTL mapping which accelerated plant selection in earlier generation rather than waiting till homozygous generation which takes 6 to 8 generations.
6. Integration of the expression information with the DNA polymorphism data by mapping eQTLs could narrow down significant candidate genes.
7. The nondestructive image data under stress has the potential to identify important targets for osmotic components of the tolerance.
8. In silico validation via allele mining and comparison with other salt tolerant donors in specific alleles provided clues to the causative mechanism underlying the variation.
9. The wealth of knowledge of genes and metabolic networks associated with the critical processes involved in salt tolerance can help fast-track breeding and pyramiding of genes in future.

Chapter 2

Review of Literature

2.1 Overview

Development of salt tolerant rice with sufficient yield is an urgent necessity in the context of ensuring food security for the overgrowing population of the world. This is especially pertinent as global warming and climate change is reducing the number of cultivable land. In this chapter the current understanding on salt tolerant rice development has been reviewed concentrating on the following areas:

1. The diversity of salt tolerance in crops, with emphasis on *Horkuch*, the salt tolerant rice, which is the subject of this study. A brief review on other tolerant donors and their bred and released varieties *(Section 2.2).*
2. Review on existing knowledge on how a rice plant responds to salinity stress and the mechanism adopted by salt tolerant glycophytes, especially rice *(Section 2.3).*
3. The modern approaches to develop salt tolerant rice with respect to the current study has been reviewed *(Section 2.4).*
4. The physiological, genomic, transcriptomic and phenomics strategies for identification of tolerance determinants and the strategies thereof used in the study are discussed *(Section 2.5).*
5. Finally the previously known determinants responsible for salt tolerance identified using the above strategies has been reviewed in case studies on rice *(Section 2.6).*

2.2 Diversity of salt tolerance in crops

Cultivated species and their wild relatives, including halophytes significantly vary in their salt tolerance. Cereal crops (like barley, triticale and wheat) and other crops (e.g cotton, canola and sugar beet) are relatively more salt tolerant, on the other hand corn, rice and beans are considered salt sensitive (Fig 2.1). In most crops, tolerance is developmentally regulated ultimately regulated by multiple traits and numerous genes. However the complex mechanism of salt tolerance and its interaction with other agronomic traits e.g. yield and grain quality and poor understanding on

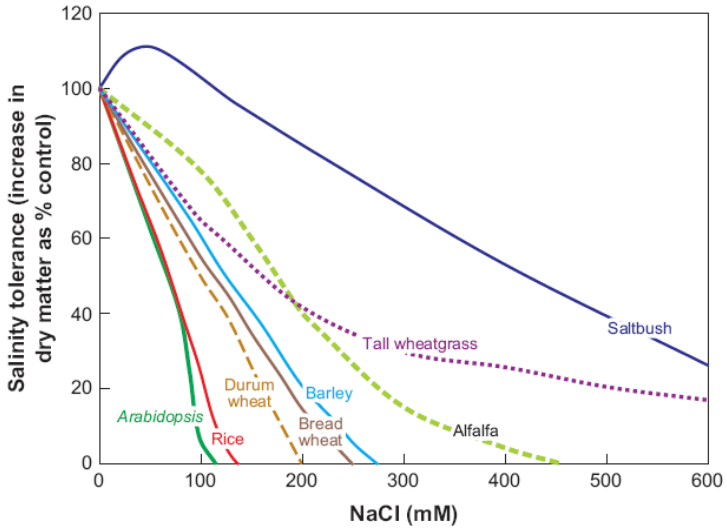


Fig 2.1:* Diversity *in the salt tolerance of various species, shown as increases in shoot dry matter (Munns and Tester, 2008)

the genetics and physiology is still hindering to get the complete scenario for exploitation of this natural variation.

2.2.1 Rice salt tolerance diversity

In spite of its high sensitivity (low threshold of 1.9-3.0 dS/m to affect growth and yield), rice is grown in salt affected and alkaline/sodic soils in coastal areas and inland, respectively (Ismail and Tuong 2009). This adaptation to the saline ecosystem results from the ability of rice to grow well on flooded soils as the free salts become diluted and leached from the mostly shallow rooting zone (Ismail and Horie 2017). Moreover rice has relatives from other species like *Oryza glaberrima*, *Oryza australiensis*, etc, which also possesses diversified tolerance. Recently, (Platten et al. 2013) evaluated a set of 550 accessions from the salt affected areas worldwide from the genetic resource center at the International Rice Research Institute, which identified ~103 moderate to highly salt tolerant genotypes. Most of them originated from the indica subspecies with a significant number from aromatic, a few from aus and tropical japonica subgroups and 12 from *O. glaberrima* (Ismail and Horie 2017). Rahman et al. (2016) evaluated a set of 111 rice accessions assembled from Bangladesh, India, Sri Lanka, and the Philippines, most of which were collected from salt-affected areas, and observed a similar abundance of tolerance within the *indica* subspecies. The two most common salt tolerant donors are Pokkali and Nonabokra. Some other tolerant cultivars are Cheriveruupu, SR26B, Damodar, Getu but with poor agronomic characteristics. Ketumbar (Indonesia), Khao Shetha (Thailand) and Soc Nau (Vietnam) have better grain quality (fine white grain, intermediate amylose) than Pokkali (Coarse red grains and high amylose) but have not been characterized adequately (Gregorio et al. 2002). Lisa et al. (2004) has studied 31 landraces from the Bangladesh coastal region and characterized them genetically and physiologically. They identified *Horkuch* along with some other moderate to highly tolerant genotypes as potential donor for breeding salt tolerant rice. In chapter 6 of this thesis some salt tolerant donors mentioned above have been used for mining allele diversity among the donors. The genetic diversity in salt tolerance in different donors with novel genes and alleles are important for development of resilient varieties using modern breeding and genomics tools.

2.2.2 Response of rice at different growth stages

Rice is relatively tolerant during germination, tillering, grain filling and maturity but sensitive during the seedling stage, panicle initiation, flowering and pollination (Fig 2.2). It was found that under varying salinity stress, rice showed delayed flowering, reduced number of productive tillers, less number of fertile florets per panicle, reduced weight per grain and subsequently poor yield (Khatun et al. 1995). Also for rice and some other crops like tomato the seedling stage does not correlate with tolerance during reproduction or other stages (Moradi et al. 2003).

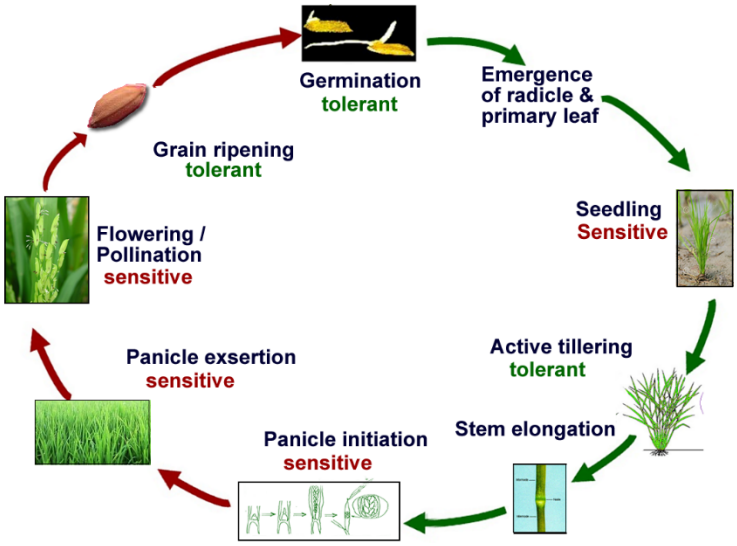


Fig 2.2 Rice salt tolerance in different developmental stages (IRRI rice knowledge bank)

2.2.3 *Horkuch* –a salt tolerant landrace

The rice landrace *Horkuch* (*Oryza sativa* L.) from the Bangladesh coast was previously identified as a potential but novel donor (Lisa et al. 2004; Lisa et al. 2011). The rice landrace *Horkuch* has evolved over many years in the saline-affected areas of Khulna and Satkhira and has been popular with farmers, despite its low yield. *Horkuch* was confirmed to be salt tolerant at the developmental stages where rice is known to be extremely sensitive to salt stress, i.e., seedling and reproductive stages under laboratory conditions. The gene expression pattern of *Horkuch* seedling in 150mM was analyzed (Lisa et al. 2011) and attempts were made to explain its superior physiological performance under salt stress. *Horkuch* was grouped with aromatic rice in a study by (Rahman et al. 2016) and hence it is diverse to the common *indica* donors of salt tolerance traits, such as *Pokkali* and *Nona Bokra*. In all breeding lines so far developed at IRRI and Bangladesh, the latter two are the only common donor of salt tolerance traits. *Horkuch* was previously shown to have less reduction in shoot biomass, low Na^+: K^+ ratio in flag leaf, less reduction in yield and good partitioning of Na^+ in the older leaves as well as high levels of Ca^{2+} and Mg^{2+} in flag leaves under stress at the reproductive stage. Lisa et al. (2011) also showed better photosynthetic capacity of *Horkuch* under salt stress.

2.2.4 Progress in development of salt tolerant rice

Despite the complex mechanism of salt tolerance and diverse developmental regulation, reasonable progress has been made in developing salt tolerant varieties. Several salt tolerant rice varieties were recently released and commercialized. (Ismail and Horie 2017) provides a list of released varieties, which includes BRRI dhan47 BRRI dhan53, 54, 55, BRRI dhan61, BINA Dhan 8, BINA Dhan 10, from Bangladesh, Salinas 6, 8 , 9 10 -20 from The Philippines, CRdhan - 402, 403 , 405, 406, CSR 43, NDRK 5088 from India spanning the years 2007 to 2014. They showed moderate to high salinity tolerance. All of these salt tolerant varieties originated from advanced salt tolerant IRRI lines (www.brri.gov.bd and www.bina.gov.bd). They are rice season-specific and cannot be grown in all seasons as shown in Fig 2.3. In salt affected areas these varieties have shown increase in grain yield (by 0.5 to more than 2 tons per hectares in farmers' fields (Islam et al. 2016; Singh et al. 2016), stable production and the ability to bring abandoned lands into production. However, these also took 10 to 15 years of rigorous evaluation for development incurring high cost. Moreover, the adoption of some of these varieties are not encouraging due to some undesirable traits like shattering (BRRI dhan47), long awn (BRRI dhan40 and BRRI dhan41), poor grain qualities (bold grains of BRRI dhan47, BINA Dhan8 and BINA Dhan10) (Chowdhury et al. 2012), less resistance to diseases and pests and most importantly low yield in high stress (> 6dS/m) condition. BRRI dhan61 and BRRI dhan67 are relatively new releases so their level of acceptability is still unknown.

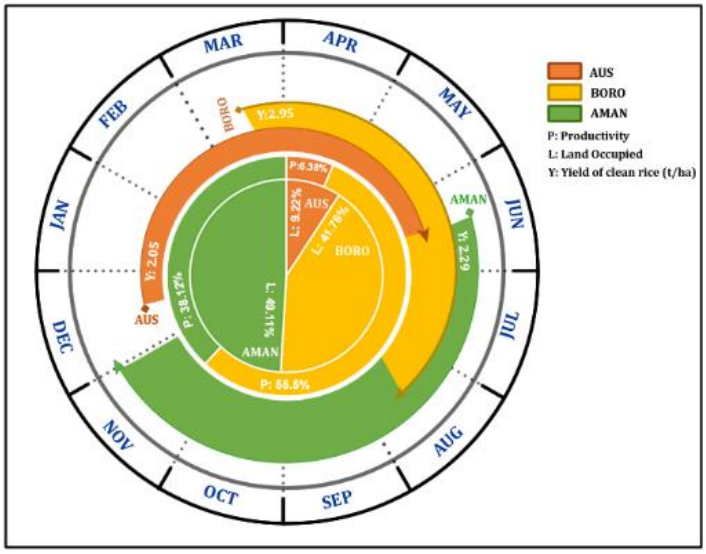


Fig 2.3 Rice growing seasons in Bangladesh

2.3 Physiological responses under stress

What happens when a rice plant encounters salt stress? Negrão et al. (2011) has discussed this response in detail dividing the response into main three stages; short term, long term and at whole plant level, which were further classified into osmotic and ionic phases.

2.3.1 Short term response

2.3.1.1 Osmotic phase: After exposure to high salinity, a temporary drop in stomatal conductance and growth rate occurs (Moradi and Ismail 2007). Sensitive varieties encounter a slower reduction in stomatal conductance. An increase in abscisic acid (ABA) concentration also occurs which is the likely cause of stomatal closure. ABA acts as a long distance signal, even though ABA levels were low in tolerant rice (Zhang et al. 2006). Na^+ might also act as an early signal triggering responses to salinity resulting in successful adaptation (Ismail et al. 2014). In longer exposures, Na^+ causes toxicity which in turn builds up depending on timing of the calcium signal (discussed more later). The signal for osmotic stress is perceived by stretch-activated channels, cytoskeleton-related mechano-sensors, stretch dependent ion (calcium) channels, redox-mediated systems and by transmembrane protein kinases, such as two component histidine-kinases (Kacperska 2004) and wall-associated kinases (Anderson et al. 2001). After a few hours of acclimation, a partial recovery occurs gradually, mostly in tolerant genotypes. The new stabilized growth rate depends on the concentration of the salt solution and on genotypic sensitivity (Moradi and Ismail 2007). Root growth recovers better than leaf growth.

2.3.1.2 Ionic phase: The ionic stress follows the osmotic one. Growth vigor is a strategy used by plants to avoid the toxic effect of salinity (Kumar et al, 2013). The concentration of transported Na^+ will be lower in faster growing genotypes, which is an avoidance mechanism. Over the timescale of hours after salt stress imposition, Na^+ ions make their way into roots through apoplastic and symplastic pathways. Water enters root hair, then travels from cell to cell through intracellular regions across the cortex to deliver salt in xylem in the apoplastic pathway (Fig 2.4). In the symplastic pathway, Na^+ enters via ion channels or carrier proteins and passes through the cell wall and epidermal plasma membrane to the cytoplasm. It then travels from cell to cell through plasmodesmata to pass the casparian bands ultimately discharging to the xylem (Das et al. 2015). In rice, a major part of sodium transport however occurs via the apoplastic pathway (Krishnamurthy et al. 2009). The influx of salt into roots activates perception and signaling mechanisms that tend to (i) inhibit the entry of further Na^+ into the roots, (ii) reduce long-distance Na^+ transport from root to shoot and (iii) restore leaf ion homeostasis.

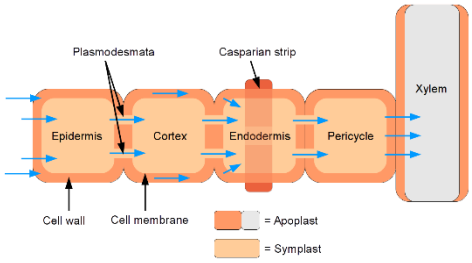


Fig 2.4: Understanding of symplastic and apoplastic route

Na^+ influx is passively mediated by voltage-independent (or weakly voltage dependent) non selective cation channels (NSCCs) (Demidchik and Maathuis 2007). The influx into roots and movement towards leaves strongly competes with K^+ uptake and activates ion transport systems that have high affinities for K^+ and low affinities for Na^+. These includes inward rectifying K^+ channels (KIRCs) like AKT1, outward rectifying K^+ channels (KORCs) and the KUP HAK family of K^+ H^+ symporters (Blumwald et al. 2000). *OsHKT1* was found to control the entry of Na^+ into roots and maintain a low Na^+ /K^+ ratio. The gene was later found to be within a major QTL region for salt tolerance called SKC1 (Ren et al. 2005). It was also found that the K^+ selective intracellular uptake is also Ca^{2+} dependent which enhances K^+ selective accumulation (Maathuis and Sanders 1996). Rahman et al. (2001) suggested that the expansion of root cap cells is an adaptive response to salinity in rice plants. Length of the root cap is shorter in the sensitive lines and is more prominent in earlier stages suggesting that proliferation of root cap cells might be related to exclusion of Na^+ or its avoidance to entry into roots. The rapid expansion also help stopping the building up of salt in high concentrations (Munns 2002).

The concentration of Na^+ in root cells is regulated by SOS1 (Salt overly sensitive), a Na^+/H^+ antiporter on the cell membrane which is coordinated by the SOS pathway. In this pathway the Ca^{2+} binding protein SOS3 interacts with and activates SOS2, a serine threonine protein kinase. SOS3 recruits SOS2 on the cell membrane, where SOS2-SOS3 complex phosphorylates SOS1 which extrudes Na^+ out of the cell (Qiu et al. 2004; Martínez-Atienza et al. 2007).

2.3.2 Long term response

2.3.2.1 Osmotic adjustment: High Na^+ concentration leads to cell dehydration as well as protein denaturation and destabilization of cellular structure. To alleviate these effects nontoxic metabolites called “compatible solutes” accumulate in the cytoplasm. These includes sugars like fructose, glucose, sucrose, complex sugars like trehalose, raffinose, fructans, sugar alcohols like mannitol, glycerol and amino acids and their derivatives like proline, glycine betaine, proline, betaine etc.(Flowers and Colmer 2015). In rice, proline accumulation rate increases in tolerant genotypes confirming its role (Demiral and Türkan 2006). Mohanty et al. (2002) suggested that transgenic rice harboring the *codA* (choline oxidase) gene is highly tolerant to salinity, though this gene is not present endogenously in rice. *codA* introduced into transgenic rice plants promotes the synthesis of glycine betaine and results in the enhancement of salt tolerance in plants (Su et al. 2006) . Zang et al. (2011) isolated 11 *OsTPS* (trehalose phosphate synthase) genes from rice. Li et al. (2011) demonstrated that overexpression of *OsTPS1* in rice

plants enhances salt tolerance by increasing the amount of trehalose and proline. But in later studies it was found that osmoprotectants might have roles in osmotic adjustment but has minimum effect on yield enhancement (Serraj and Sinclair 2002). This mechanism is therefore of low practical use since the main concern is salinity stress tolerance which results in significant yield gains. Sakurai et al. (2008) showed that rice aquaporins have a role in maintaining the water potential in different tissues and cells.

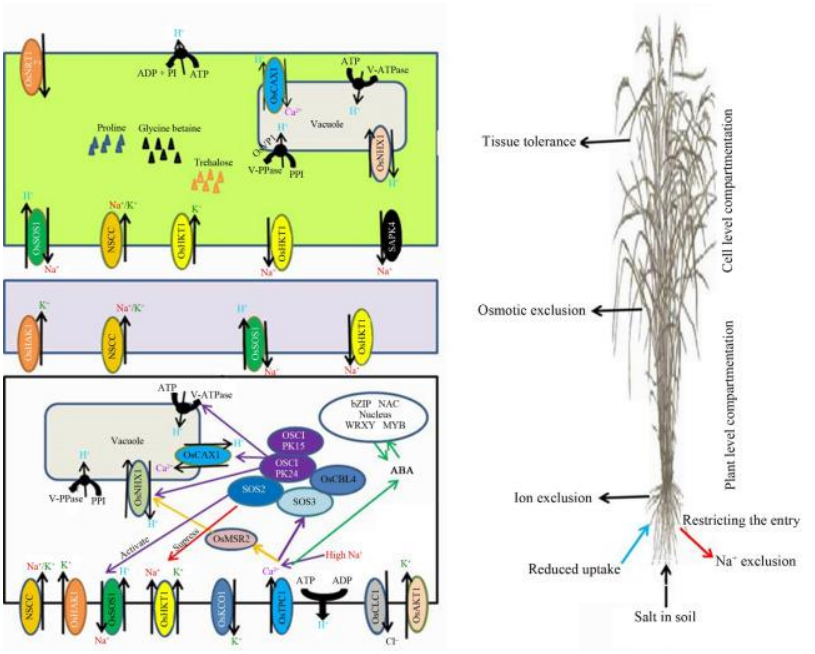


Fig 2.5 Transport of ion emphasizing the transporters (adapted from Reddy et al, 2017)

2.3.2.2 Ionic homeostasis*:* Plants use 3 main strategies to achieve a high cytosolic K^+/Na^+. These are exclusion of Na^+ ions entry, extrusion of Na^+ ions out of the cells and vacuolar compartmentalization of Na^+ ions. Rice can exclude at least 94% of the soil Na^+ from the transpiration stream (Munns 2005). At the whole plant level, exclusion of Na^+ ions takes place in different ways, such as, 1) selective uptake by root cells, 2) preferential loading of xylem with K^+ rather than Na^+, and 3) retention of Na^+ in the upper part of the root system and in the lower part of the shoot through exchange of K^+ for Na^+ in the cells lining the transpiration stream(Negrão et al. 2011).

In rice, *OsHKT1* mediates leaf Na^+ exclusion by removing Na^+ from the xylem sap(Davenport et al. 2007). Transcript expression of *OsHKT* genes is much higher in shoots than roots except for *OsHKT4* and *OsHKT1*. *OsHKT2* mediates Na^+-K^+ co transport. For intracellular compartmentalization, rice sequesters extra Na^+ into vacuoles. The transport of Na^+ into the vacuoles is mediated by a Na^+/H^+ antiporter driven by a proton electrochemical gradient generated by the vacuolar H^+-translocating enzymes, H^+-ATPase and H^+-pyrophosphatase(Blumwald 1987). *OsNHX1* overexpression in rice (Fukuda et al. 2004; Amin et al. 2016) showed enhanced tolerance to salt.

2.3.2.3 Signaling and antioxidant system

In all these responses signaling molecules like Brassinosteroids, Abscisic acid (ABA) plays important roles in regulating plant growth and development under stress. Depending on ABA two pathways of signaling exists. In the **ABA dependent pathway** mitogen activated protein kinase (MAPK) cascades, calcium- dependent protein kinases (CDPK), receptor-like kinases (RLK), sucrose non-fermenting-1 (SNF1) related protein kinases, transcription factors and microRNAs (miRNAs) play important roles. Transcription factor *OsNAC22* plays a positive role in salt tolerance through modulating an ABA-mediated pathway. Rice transgenic plants overexpressing miR393 showed an increment in tillering and early flowering stage, together with decreased tolerance to salt (Xia et al. 2012). **ABA-independent pathway** includes transcription factors like *OsDREB1*, ROS-producing/scavenging enzymes and stress-related genes. Reports suggested that rice transgenic plants overexpressing *OsDREB1A*, *OsDREB1F* and *OsDREB2A* show improved salt tolerance (Wang et al. 2008; Mallikarjuna et al. 2011). ABA-independent ROS scavenging system is also involved in salinity tolerance. Plants have devised different system for scavenging of ROS by using the detoxifying enzymes like super-oxide dismutase (SOD), peroxidases, catalases and antioxidants like ascorbate and reduced glutathione. (Ismail et al. 2014) proposed that depending on timing, Ca^{2+} plays a major role in signaling. A delay in generation and dissipation of salinity triggered Ca^{2+} relative to a signal by ROS will lead to activation of Jasmonate (JA) signaling initiating cell death, in contrast if properly timed Ca^{2+} initiate adaptive response like sequestration and extrusion of sodium and activate the ABA signaling pathway.

2.3.3 Response in whole plant level

Over the course of about a week, or so, change in the cellular ionic and osmotic status affects photosynthesis and growth. Increased concentration of organic osmolytes and growth regulators such as ABA, reduced membrane permeability, lower intercellular CO_2 partial pressure, lower guard cell turgor and stomatal conductance, there is a big decrease in efficiency of the photosynthetic apparatus and feedback inhibition due to reduced sink activity (Zhang et al. 2006; Chaves et al. 2009) (Moradi et al, 2007, Chaves et al, 2009). Under salt stress shoots, are affected more than root growth. Root cytokinin concentration is reduced and transport of auxin from shoot to root is induced. The overall downregulation of photosynthesis and growth leads to decreased yields (Zheng et al. 2011) (Zheng et al., 2001). For instance, even at a low salt concentration (30 mM NaCl), a long exposure affects the number of tillers per plant, number of spikelets per panicle, spikelet fertility, grain weight and the final grain production (Grattan et al. 2002).

When the tolerance mechanism cannot effectively exclude salt from the transpiration stream, old leaves accumulate the salt in toxic level causing death. New leaf growth is supported through the export of carbon from mature leaves, they maintain source (carbon of mature leaves) – sink (growth of new leaves) relationship. Plant death may occur when this relationship is disrupted. The ultimate aim of any crop plant is to complete life cycle by reaching successful reproduction. Under <50mM stress salt stress induces delayed panicle initiation and flowering (Grattan et al. 2002) and also leads to poor seed set through reduced pollen viability and other yield parameters like tillering, spikelet number, sterility and grain weight (Khatun et al. 1995). Tolerant genotypes tend to exclude salt from flag leaves (crucial organs for carbohydrate synthesis at the reproductive stage) and panicles (Moradi et al. 2003). Tolerant varieties keep lower salt concentration in panicle with the lowest concentration in grains compared to husks and rachis (www.knowledgebank.irri) so the weight of unfilled grain as well as rachis can indicate genotypic tolerance. Tolerant varieties tend to maintain lower concentration of salt in the flag leaves (Ahmed et al. 2013). Selection for healthier flag leaf at flowering could be used to screen for salt tolerance at the reproductive stage (www.knowledgebank.irri).

2.3.4 Mechanism of salt tolerance

In rice, considering the above physiological responses mechanism of salt tolerance can be portrayed as follows

(1) Salt exclusion: By selective absorption avoid taking excess Na^+

(2) Salt reabsorption: Translocation of Na^+ to shoot is avoided by reabsorption

(3) Root-shoot translocation: Maintain high electrolyte content in root and low content in shoot

(4) Salt translocation: Tolerant plants are able to translocate lower amount of Na^+ to the shoot

(5) Salt compartmentalization: Excess salt is transported from younger to older leaves

(6) Tissue tolerance: Compartmentalize the excess Na^+ in vacuole to lessen the deleterious effect on growth. It might include ABC transporters and detoxification mechanisms through removal of oxygen radical as well.

(7) Salt dilution: By fast growth rate and high water content in the shoot dilute the salt.

2.4 Strategies for engineering salt tolerant rice

Recently (Reddy et al. 2017) has illustrated the common approaches being undertaken to develop salt tolerant rice (Fig 2.6). All analysis and studies normally are leading to two main approaches described below.

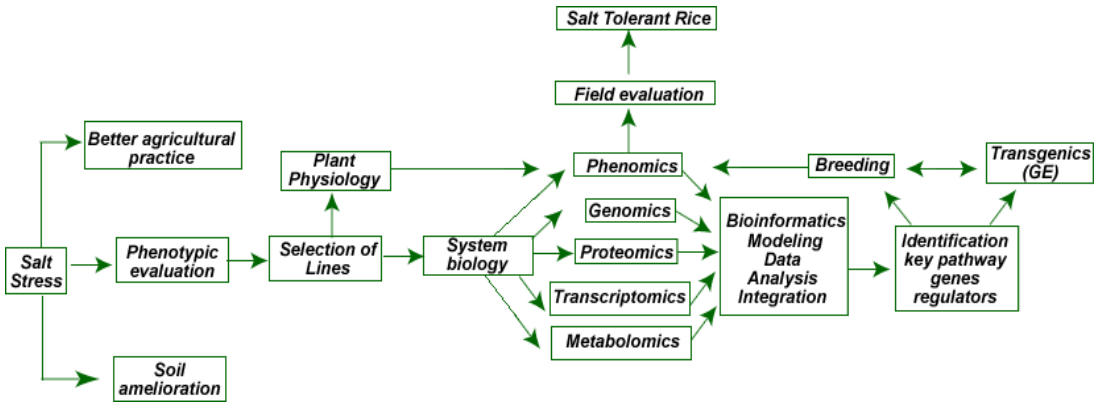


Fig 2.6 Approaches for development of salt tolerant rice *(adopted and redrawn from (Reddy et al. 2017)*

2.4.1 Marker Assisted Selection (MAS) and Molecular breeding

In contrast to conventional breeding, marker assisted breeding involves the selection of plants by polymorphism at the level of DNA and its relationship with phenotypic polymorphism. Selecting plants from a segregating population that contain appropriate combinations of genes is a critical component of plant breeding. Marker-assisted selection (also 'marker-assisted breeding' or 'marker-aided selection') greatly increases the efficiency and effectiveness of plant breeding compared to conventional breeding methods. Since the polymorphic markers are linked to the trait of interest, these will co-segregate with the trait, during segregation of genes after crossing. Hence by tracking a linked marker in the progenies one can track the desired trait and ultimately select the plant possessing the trait of interest. For quantitative traits multiple markers may be linked to responsible loci and can be identified as QTLs (Quantitative Trait Loci). Markers that are tightly linked to genes or QTLs of interest need to be identified prior to field evaluation of large numbers of plants. Thereafter, breeders use specific DNA marker alleles as a diagnostic tool to identify plants carrying the desired genes or QTLs. However, due to the high cost of utilizing several QTLs, only markers that are tightly linked to no more than three QTLs are typically used, although there have been reports of up to 5 QTLs being introgressed into tomato via MAS (Lecomte et al. 2004). After selection using these markers, backcrossing is normally adopted to recover the recurrent parent genome. Using conventional breeding it takes 6-8 backcross to recover the recurrent genome compared to only 2-3 in MABC. Pyramiding of desired genes and QTLs is now proving to be a promising and effective strategy, particularly with advancement of technologies (Pang et al. 2017).

2.4.2 Transgenic approach

The introgression of genomic portions (QTLs) involved in stress tolerance often brings along undesirable agronomic characteristics from the donor parents due to lack of precise knowledge

of the key genes underlying the QTLs. Therefore, the development of genetically engineered plants by the introduction and/or overexpression of selected genes seems to be a viable option to hasten the breeding of ‘‘improved’’ plants. Also it is a faster way to insert beneficial genes than through conventional or molecular breeding and would be the only option when genes of interest originate from cross barrier species, distant relatives, or from non-plant sources. Since the abiotic stress tolerance are controlled by polygenic traits and as a consequence, multiple genes, transformation of master regulators such as stress inducible transcription factors are of great interest (Kasuga et al. 1999).

Transgenic approaches uses recombinant DNA techniques to create plants with new and desirable characteristics. This approach has been successfully applied in producing salt tolerant plants by introducing new genes and testing them against salinity stress. Transgenic plants overexpressing *OsPP1a* show enhanced tolerance to high salt treatment, and *SnRK1A*, *OsNAC5* and *OsNAC6* are up-regulated in these transgenic plants (Liao et al. 2016). Also enhanced tolerance was observed in overexpression of *OsNHX1* (Fukuda et al. 2004; Amin et al. 2016) and *OsTPS1* associated with trehalose-6 phosphate synthase (Li et al, 2011). Sahoo et al (2014) reported the use of OsSUV3 which functions as DNA and RNA helicase and provides salinity stress tolerance by maintaining photosynthesis and antioxidant machinery. Also overexpression of *PDH45* which has both DNA and RNA helicase activity and probably stabilizes gene expression under salt stress, showed enhanced tolerance (Amin et al. 2012). Development of stress tolerance in rice varieties is however not enough, the plant also will need to have high yields. So care must therefore be taken in the choice of promoter and vector used to drive the expression of salt tolerance transgenes, so as not to disrupt or negatively affect the plant phenotype. In recent days Crispr-Cas9 technology is facilitating the transgenic approach. The latter technology uses the host editing system to incorporate the transgene into the genome leaving little trace of the donor. Therefore this approach may remove many consumer acceptance issues associated with genetically engineered transgenic crops.

2.5 Strategies for identification of salt tolerant determinants

2.5.1 Screening for tolerance: identification of phenotypic polymorphism

2.5.1.1 Morphological traits important for measuring salinity tolerance

Plants’ response under salt stress helps us identify target traits for improvement of salinity tolerance. Pires et al. (2015) concluded that none of the three salinity tolerance mechanisms i.e., shoot-ion independent tolerance (or osmotic tolerance), ion exclusion and tissue tolerance is predominant among rice genotypes and that individual genotypes may have adapted to use a

combination of different strategies. In a recent review by (Negrão et al. 2017), the impact of salinity on physiological trait measurement techniques by both destructive and non-destructive ways under stress has been discussed. Destructive harvest involves separation of plants into parts, measurement of shoot and root length, fresh and dry biomass of shoot and root, leaf area, length, chlorophyll, water relations, ion composition, yield components, senescence, etc. The simple visual salt injury scoring (Gregorio et al. 1997) is widely used for characterization as it reflects the overall plant's response to salt stress. (De Leon et al. 2015) suggested that instead of considering only salt injury score, other parameters, like ion leakage, chlorophyll concentration, shoot length, shoot K^+ concentration and shoot Na^+/K^+ ratio could be unbiased parameters for assessing salinity tolerance. Rather than individual Na^+ and K^+ content of the plant, the K^+/Na^+ ratio is of greater relevance to study salinity tolerance in rice. These ideas are also supported by studies from (Asch and Wopereis 2001) and (Garcia et al. 1997). Plants with different biomass can have similar growth reduction, even though plant with more biomass get allow more dilution of toxicity and therefore of tolerating the stress better. As a precaution Negrão et al (2017) suggested to consider the mass of salt while measuring the biomass, as this can become a significant fraction of the total mass and proposed use of ash-free or ethanol soluble dry mass instead. Third leaf sodium content is normally higher in plants with high SES score (Platten et al. 2013) though other physiological responses can differ. The inherent subjectivity and the quantitative nature of salinity tolerance complicate its evaluation. Physiological components such as the maintenance of plant water status, transpiration (T) and transpiration use efficiency (TUE) (Harris et al. 2010; Barbieri et al. 2012), leaf area (Maggio et al. 2007), seed germination, production of antioxidants (Ashraf and Akram 2009), early seedling growth, and harvest index (HI) are also likely to contribute to salinity tolerance. The hypothesized plausible reasons for changes in HI may include a lower shoot biomass reduction, maintenance of tiller number (Zeng and Shannon 2000) or earlier flowering (Saade et al. 2016). Besides HI, other parameters shown to be affected by salinity stress includes yield and yield components including seed/fruit mass, spikelet per panicle (for cereals), spikelet fertility rates in the spikes, 1000-grain mass, etc. (Gholizadeh et al. 2014).

In the non-destructive approach, images of plants are taken at defined time intervals over several days and the biomass is deduced from pixel counts (Berger et al. 2012). Recent advancements in imaging technology and high throughput phenotyping adds digital traits to assess the biomass, growth, chlorophyll concentration, growth rate and growth rate derived tolerance scores from the projected shoot area, convex hull area, height, weight and image moments (Campbell et al. 2015). Temperature and water status of the plants captured by IR, NIR and hyperspectral cameras open up a new dimension for linking and correlation with salt tolerance estimates. The use of Principal Component Analysis (PCA) and correlation based

clustering can provide an indication of the most important traits contributing to salinity tolerance in the materials and conditions under study. Though measurements of yield of tuber crops as well as other roots are still little challenging, novel methods are also developing for the latter. In a nutshell, shortcomings of morphological markers which makes the screening for a quantitative trait like salt tolerance difficult, can be now overcome by considering a combination of the strategies discussed above.

2.5.1.2 The system for conducting screening

Creation and maintenance of salt stress conditions for appropriate screening are major issues which can dramatically change results. Though field screening is the most relevant and appropriate one, in the initial stages, a controlled environment is needed to avoid the climatic effects. Plant's developmental stage at which the stress is imposed is also very important as the sensitivity varies in different stages. For seedling stage, the hydroponic system and for reproductive stage the soil-based system is well adapted in breeding programs. However, flaws and benefits remain, as in hydroponic systems, the transpiration is too low to be representative of field conditions, and the imposed stress may not be gradual enough or too severe (Flowers et al. 2000). Detailed screening procedures have been proposed by (Gregorio et al. 1997) which normally targets salinity tolerance as a whole or assess salt effect on yield components. Application of gradual stress is also proposed to avoid salt shock (Shavrukov 2013). Addition of calcium chloride as a supplement along with NaCl has been also suggested (Walia et al. 2005).

2.5.2 Identification of genotypic polymorphism

2.5.2.1 Genetic markers and different genotyping methods

Genetic markers represent genetic differences between individuals or species. They may not affect the phenotype of the trait themselves as they remain linked mostly in the non-coding regions. There are 3 major type of markers, i.e morphological or visible, biochemical and DNA markers. Unlike morphological or biochemical markers, DNA markers are practically unlimited in number and are not affected by environmental factors or developmental stage of the plant (Winter and Kahl, 1995). These markers are called polymorphic when they can discriminate between genotypes and monomorphic if they cannot. Polymorphic markers may also be described as codominant or dominant depending on whether they can distinguish the homozygotes and heterozygotes. DNA markers may be broadly divided into three classes based on the method of their detection: (1) hybridization-based; (2) polymerase chain reaction (PCR)-based and (3) DNA sequence-based (Gupta et al., 1999; Jones et al., 1997; Joshi et al., 1999;

Winter & Kahl, 1995). Thus polymorphic markers can become the identity of a specific individual organism. This provides the opportunity of genotyping as a set of polymorphic markers can discriminate one individual from another. Based on the DNA markers, genotyping can be done using RFLP (Restriction Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), SSR (Simple Sequence Repeat) or SNP (Single Nucleotide Polymorphism) markers.

2.5.2.2 Genotyping by next generation sequencing

With the current advancement in technology it is not impossible but tough to sequence the whole genome of all individual living things in the world. Hence a logical and easier solution comes with genotyping by sequencing which uses restriction enzymes to reduce genome complexity and genotype multiple DNA samples together (multiplexity). After digestion, fragments are increased in number by PCR (Polymerase Chain Reaction), multiple samples are pooled using different adaptors from different samples and lastly sequences generated using next generation technology. Depending on the variation in methods, Genotyping by Sequencing (GBS), Restriction Associated DNAseq (RADseq), and double digest restriction site associated DNA seq (ddRADseq), etc. the naming can vary. For example, ddRAD uses one rare cutter and one frequent cutter restriction enzyme, while RADseq uses only one cutter. The variation also depends on the exact size selection for sequencing. These methods are relatively inexpensive and increasingly being used in plant molecular breeding. These approaches are one-step procedure of SNP discovery and genotyping and constitutes a rapid, high-throughput and cost-effective tool for a genome-wide analysis of genetic diversity, especially for non-model species and germplasm sets. Many processes have been advanced to cut down genome complexity; nonetheless, the DArTseq method has brought a significant improvement via intelligent selection of the genome fraction to correspond mainly with active and genes and polymorphic loci. DArTseq™ refers to Diversity Arrays Technology based sequencing. Selection is achieved by using a combination of Restriction enzyme which separate low copy sequences (most informative for marker discovery and typing) from the repetitive fraction of the genome. DArT was implemented initially on the microarray platform which involved fluorescent labeling of representations and hybridization to dedicated DArT arrays. Diversity Arrays Technology (DArT) offers an inexpensive and high throughput whole-genome genotyping technique as initially shown for rice (Jaccoud et al. 2001). Classic DArT markers have been substituted by DArTseq markers based on genotyping by sequencing. DArTseq for a new organism starts with optimization of complexity reduction using combination of restriction enzymes. This is advantageous over the array version when high marker densities (tens of thousands of markers)

are required. DArTseq and SNP markers based on GBS technology have been successfully applied for linkage mapping, QTL identification in biparental mapping population, genome wide association studies (GWAS), genetic diversity studies in wheat (Akbari et al. 2006; Baloch et al. 2017) and many other crops (Grzebelus 2015) and as well marker-assisted and genomic selection.

2.5.3 Detection of QTL: Correlating phenotypic and genotypic polymorphism

The various components of salinity tolerance appear to be polygenically controlled. QTL analysis is based on the principle of detecting an association between phenotype and the genotype of markers (Collard et al. 2005) . QTL mapping requires a mapping population where the genes are segregated from the parental ones. Genetic polymorphism of the mapping population in form of a genetic map or linkage map and subsequent QTL mapping can then define both genotyping and phenotypic polymorphism among the parents by correlation.

2.5.3.1 Mapping population

Selection of two genetically divergent parents showing clear difference for one or more trait of interest is the first step in creating a mapping population. But precaution should be taken, avoiding too much divergence, as that may cause sterility and show very high level of segregation distortion during linkage analysis. Progenies from second filial generation (F_2), backcross (BC), recombinant inbred lines (RIL), double haploids (DHs) and near isogenic lines (NIL) can be used for genetic mapping in self pollinations species (Burr et al. 1988; Doerge 2002). F_2s are derived by selfing F_1 hybrids originated by crossing 2 parents while BC population is produced by crossing F_1 back into one of the recipient parent. RILs are developed by single-seed selections from individual plants of an F_2 population; such selections continue for six to eight generations. Repeating backcross for at least 6 generations derives more than 99% of the recurrent parent genome in BC_6 and above (R. et al. 2004). Selfing of selected BC_7F_1 individuals produce BC_7F_2, homozygous for target genes, and nearly isogenic with the recipient parent (NILs). A DH population is produced by doubling the gametes of F_1 or F_2 population followed by regeneration using tissue culture. Selection of mapping population affect linkage analysis and size of the genetic map. RILs, NILs

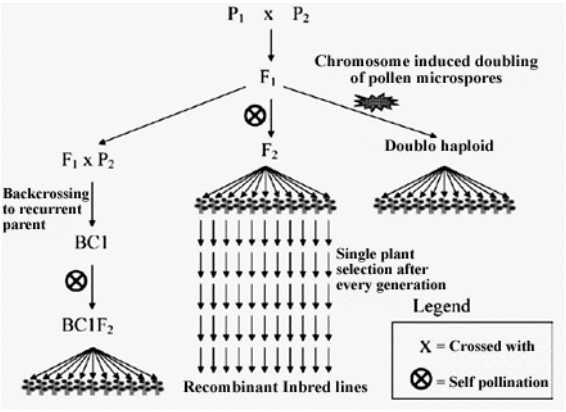


Fig 2.7 Main mapping populations for self-pollinating species (*Adopted from Collard et al , 2005*)

and DHs are permanent populations and homozygous while F_2 and BC are temporary-type and size of mapping populations can exert an influence on the accuracy of genetic maps (Ferreira et al. 2006).

Sometimes in QTL studies reciprocally crossed mapping populations are used. These are crosses where the sexes of the parents are reversed e.g. in the case of strains *A* and *B*, (*A*×*B*) and (*B*×*A*) are reciprocal crosses. Reciprocal crosses can show difference in inheritance of maternally supplied growth factors, nutrients, episomes, e.g. mitochondrial DNA, uniparental epigenetic marks and in inheritance of sex chromosomes. Hence these may be useful in mapping loci that underlie parent-of-origin effects. Evaluation of the cytoplasmic component or direction of the cross, may significantly affect the agronomic performance related to the nuclear genotypic value of the hybrid. Genetic information carried by the cytoplasmic organelles (mitochondria, chloroplast) and their interaction with the nuclear genetic information can contribute to a specific phenotype and can favor the evolutionary co-adaptation of high-fitness combinations of nuclear and cytoplasmic alleles (Wolf 2009; Wolf 2009; Wolf and Brandvain 2014). Considering quantitative traits, Edwards et al. (1996) reported cytoplasmic effect in maize. The cytosolic genetic background could also affect the ability to identify genetic loci in mapping populations (Moison et al. 2010; Tang et al. 2013).

2.5.3.2. Genetic polymorphism and linkage mapping

Linkage map construction requires generation of mapping population, identification of genetic polymorphism and linkage analysis. A segregating population contain mixture of parental and recombinant genotypes. The frequency of these recombinant genotypes can be used to calculate recombination fractions and ultimately which is used to predict the genetic distance between the markers. Mapping functions are used to convert recombination fractions into map units called centi morgans (cM). So, the linkage map basically indicate the position of polymorphic markers in the genome available from the genotyping assay. Linkage between markers is usually calculated using odds ratios i.e the ratio of linkage versus no linkage and expressed as a logarithm value (LOD). A LOD (logarithm of odd) value of 3 between 2 markers indicates that linkage is 1000 times more likely than no linkage (Collard et al. 2005). Mapping functions are require to convert recombination fractions into centiMorgans (cM) as the recombination frequency and the frequency of crossing over are not linearly related (Hartl and Jones 2001). Most commonly used 2 mapping functions are the kosambi (assumes recombination events influence the occurrence of adjacent recombination events) and the Haldane (assumes no interference between crossover events)(Hartl and Jones 2001). The distance in a linkage map is also not directly related to a physical distance between genetic markers. It depends on the plant genome size, recombination's hotspots and cold spots. Commonly used software are

Mapmaker, MapManager, Joinmap, R/QTL etc. Linked markers clusters together to make linkage groups representing chromosomal segments.

2.5.3.3 Principle of QTL detection

The closer a marker is from a QTL, the lower the chance of recombination occurring between marker and QTL. Three widely-used methods for detecting QTLs are single-marker analysis, simple interval mapping and composite interval mapping (Tanksley 1993; Lin et al. 1998; Liu 1998). Single-marker analysis (also 'single-point analysis') is the simplest method for detecting QTLs associated with single markers. For single marker analysis, normally *t*- tests, analysis of variance (ANOVA) and linear regression are use. The coefficient of determination (R^2) from the marker explains the phenotypic variation arising from the QTL in linear regression, so it is commonly used. However, the major disadvantage with this method is that the further a QTL is from a marker, the less likely it will be detected. This is because recombination may occur between the marker and the QTL. This causes the magnitude of the effect of a QTL to be underestimated (Tanksley 1993). The use of a large number of segregating DNA markers covering the entire genome (usually at intervals less than 15 cM) may minimize both problems (Tanksley 1993). In this context composite interval mapping (CIM) is statistically much more powerful as it combines interval mapping with linear regression and include additional genetic markers in the model (Jansen 1993).

2.5.4 Gene expression polymorphism

Gene expression analysis refers to the study of the way genes are transcribed to synthesize functional gene products- functional RNA species or protein products. Up or down regulation of expression of the same gene under stress and non-stress condition is a very important measurement for identification of candidate genes. The variation observed in level of expression values of a gene in different samples is called expression polymorphism, commonly known as differential expression.

2.5.4.1 Various expression analysis methods

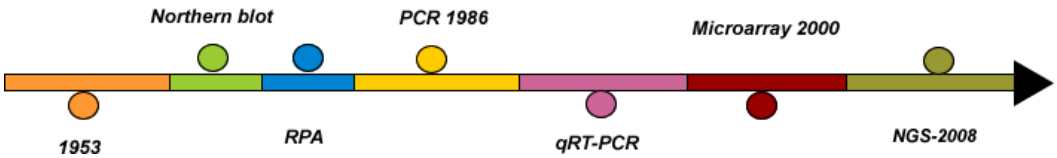


Fig 2.8: Expression analysis method timeline

In the past, technical approaches have been developed in two major ways to characterize gene expression regulation. These detection technology ranges from traditional analysis of one gene at a time using northern blotting, reverse transcription PCR, real time quantitative PCR(qPCR),

to large scale analysis of many genes using microarray (Minning et al. 2003) (Ta et al, 2003), and more recently RNA sequencing (Wang et al. 2009)(Fig2.8).

2.5.4.2 RNA sequencing

RNA-seq is replacing other methods of quantifying transcript expression, including microarray platforms (Martin et al. 2013) as it overcomes some of their limitations, such as detection of only those transcripts that are represented on microarrays, low dynamic range (limited upper and lower limits of detection), and thus provides more accurate quantification of differential transcript expression. Multiplexity is a measure of how many different genes can be practically assessed by a given experiment and for quantification of expression of a gene various analysis methods are recognized e.g. tag based, shotgun method, targeted method etc. In brief, after preparation sequencing library from RNA or cDNA, it is sequenced to a specified depth. These reads are aligned to the genome or transcriptome and can be counted to determine differential gene expression or, alternative splicing and isoform expression. Usually paired-end 50-100 bp are sequenced. The exception is microRNA sequencing as this only requires single-end 36bp sequencing.

2.5.4.3 3′ tag based RNAseq

The high sequencing cost of whole mRNAseq leading to low powered sample design makes the whole mRNAseq unsuitable for studies with large number of sample. 3′ TagSeq method used by (Meyer et al. 2011) is an efficient low cost method that can resolve the problem of screening large number of samples. It focuses on sequencing the 3′ end of mRNA fragments which reduces the sequencing efforts required to characterize a population of mRNAs in a biological sample. Lohman et al. (2016) later modified the method to avoid the skewness in distribution of RN fragments in the library, with respect to both fragment size and GC content (Aird et al, 20011). They also compared the TagSeq method with industry standard Truseq/NEBNext which reliably measures moderate and high abundance mRNA in a sample. And found it trustworthy.

2.5.4.3.1 Advantages of the TagSeq method

1. **Increased sample number:** Increased number of samples can be used in gene expression analyses.
2. **Increased statistical power:** It also increases the statistical power to detect subtle difference in transcript abundance than traditional whole mRNAseq methods.

3. **Improved accuracy:** It promises improved accuracy when measuring medium and low abundance RNAs. More differentially expressed transcripts with 3′ RNA-seq has been identified suggesting a greater power to detect expression variation
4. **Uniform distribution of fragments:** In total RNAseq even if two transcripts are expressed at identical levels, random fragmentation and priming leads to greater representation of longer fragments in sequencing libraries (Trapnell et al. 2010; Roberts et al. 2011) resulting in a biased estimation of expression levels. TagSeq only primes the 3 ′ poly-A tail, generating an essentially uniform distribution of fragments with respect to original RNA length.
5. **Cost Effective:** The improved Tagseq method is more cost effective than whole mRNAseq. At maximal efficiency (32 individuals per sequencing lane) it could produce highly accurate, transcriptome-wide gene counts for only ~$33 per sample, including sequencing cost (HiSeq 2500 1 9 100 V3 chemistry with ~5.6 M raw reads per sample.

2.5.4.3.2 Limitations of the TagSeq method

The drawbacks are negligible when compared to the cost effectiveness of the procedure.

1. It does not generate optimal numbers of clusters on the HiSeq platforms
2. Slightly lower number of raw reads.
3. No alternative splicing: TagSeq cannot distinguish between alternatively spliced transcripts from a single locus,
4. No allele specific expression: It does not identify polymorphism or allele-specific expression in much of a gene's coding sequence
5. Genotyping still ambiguous.

2.5.5 Expression QTL and polymorphism

Genetic variation often lead to alternative gene expression which may influence phenotypic variation. Expression of genes is a dynamic process which depends on different tissue, cells, developmental stages, genetic backgrounds and growth environments (Wang et al. 2010). The loci controlling transcript abundance have been termed expression QTL (eQTL). Transcript abundance can be termed as an eTrait treating it as a quantitative trait, just like other phenotypes (pTraits), e.g plant height, yield etc. eQTLs can be of two types, those mapped to the same genetic location as the gene whose transcript is measured generally are termed cis-eQTL and generally indicate the presence of a cis acting regulatory polymorphism in the gene. eQTL that

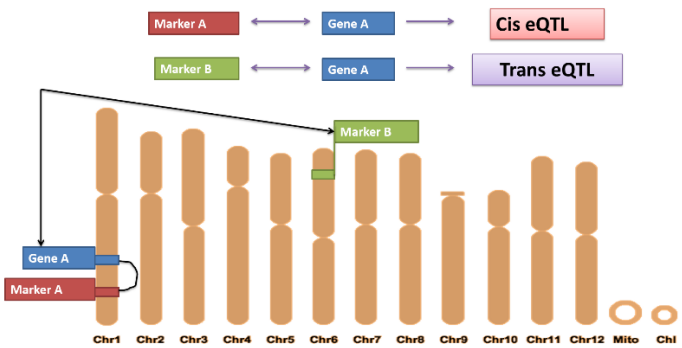


Fig 2.9: Schematic representation of cis and trans eQTL

map distant to the location of the gene being assayed most likely identify the location of trans-acting regulators and termed trans-eQTL (Fig 2.9). These may control the expression of multiple genes elsewhere in the genome. In principle, a cis-eQTL is detected due to polymorphism within a neighboring range of the gene, including polymorphism within the gene itself and alter the cis-acting elements for mRNA stability or polymorphisms between parents in the promoter region (Smith et al. 2011). But it can also act in trans as in the case of the AMN1 gene *S. cerevisiae* that was found to be a local but trans-acting through feedback regulation due to coding polymorphism in its transcript (Ronald et al. 2005).

2.5.5.1 Significance of expression QTLs

Several recent studies have shown that variation in transcript abundance is the cause of variation in phenotypes that include disease resistance, insect resistance, glucosinolate biosynthesis and activation, phosphate sensing, flowering time, circadian rhythm and plant development (Piffanelli et al. 2004; Wentzell et al. 2007). Analysis of eQTLs which treats the expression values as quantitative trait to perform QTL analysis is a useful strategy for suggesting regulatory relationship between genes (Kliebenstein 2009). Combining this approach with co-expression analysis facilitates the identification of regulator candidates underlying eQTLs, which may gently refine the resolution of the analysis (Flassig et al. 2013). eQTL studies are gaining popularity as they have the potential to bypass the tedious positional cloning process (Hansen et al. 2008). They can be used to generate a sequence-based genetic framework map (Potokina et al. 2008) as well as can be directly coupled to candidate gene identification (Shi et al. 2007). eQTL analysis also opens the door for revealing network of genes under common regulatory control. It also provides the possibility of correlating observed variation in the abundance of mRNA transcripts with variation observed in simple or complex phenotypes revealing the molecular basis of the phenotype (Kliebenstein et al. 2006). Furthermore exploring eQTLs using system analyses of a biological phenomena is of much significance genome wide. eQTL study in a segregating population can also detect significant genes that are/are not differentially expressed between the parental lines under stress, thus indicating the relative importance of a genetic locus in conferring tolerance.

2.5.5.2 Challenges in studying eQTL

The major challenge in expression QTL analysis lies in the high dimensionality of genome-wide gene expression data and their complex correlation, otherwise it shares the same basics as the QTL study with physiological phenotypes (termed pQTL in the context of eQTL). The RNAseq data carry both genetic marker and transcript abundance information. But proper statistical models are needed to extract the genetic polymorphism information from the RNAseq or microarray data, though there are some reports on SFP (Single feature polymorphism)

marker generation from hybridization which led to false prediction of cis regulators. eQTL study provides useful data for identification of candidate genes for pQTLs. eQTL and pQTL generated from the same population is a very useful resource as these genes are co-located within the pQTL and these are differentially expressed among the parent and there segregating progenies. Validation methods of the role of these candidate genes in polygenic processes need to be improved. The large amount of expression data is also a big issue compared to pTraits which requires different algorithms efficient analysis. Several tools are available for eQTL analysis e.g R/eQTL, eMAP, Merlin, FastMap, compared to which Matrix eQTL showed the fastest performance (Shabalin 2012).

2.5.6 Association genetics GWAS

Another important tool apart from the eQTLs, that can offer a valuable first insight into trait architecture as well as candidate loci for subsequent validation is Genome wide association study i.e. Genome Wide Association Study (GWAS) (Korte and Farlow 2013) . QTL mapping suffers from two fundamental limitations; only allelic diversity that segregates between the parents of the particular F_2 cross or within the RIL population can be assayed, and second, the amount of recombination that occurs during the creation of the RIL population places a limit on the mapping resolution. If an advanced intercross RIL is used this resolution can be dramatically improved by several generation of intercrossing. Also there are population e.g. Multi-parent Advanced Generation Inter-Cross (MAGIC) and *Arabidopsis* multi-parent RIL (AMPRIL) where the allelic diversity is increased within a mapping population by intercrossing multiple genetically diverse accession before establishing RIL(Kover et al. 2009).

GWAS overcomes the two main limitations of QTL analysis mentioned above, but introduces several other drawbacks as a trade-off. However GWAS is gaining popularity among plant scientists as sequencing and resequencing efforts now provide SNPs and InDels at the right density for GWAS study (Delseny et al. 2010). Generally, after identifying a phenotype of interest, GWAS can serve as a foundation experiment to explore the genetic architecture of the trait thus allowing an informed choice of parents for QTL analysis and suggesting candidates for transgenic and mutagenesis. Thus, GWAS are often complementary to QTL mapping and, when conducted together, they mitigate each other's limitations.

2.5.7 High Throughput Phenotyping

Phenotyping or measurement of morphological traits is very important as plant breeding and genetic studies require hundreds or even thousands of measurements to select superior individuals. This selection eventually leads to identification of genomic regions controlling the trait. Phenome is the expression of the genome as traits in a given environment, as defined in (Furbank and Tester 2011). High throughput phenotyping involves the screening of phenotypes

of a large number of plants, as the name implies. The high throughput nature comes from automation rather than laborious manual phenotyping. Typically, it uses an image- capture system in which the plants are photographed regularly or at specific time intervals on an automated conveyor belt system. Here they are automatically watered and moved around in the green house area to avoid spatial biasness. The system is under continuous computer monitoring and regulation. In this way, millions of images of the side and top view of the plants from different angles are captured. This system is also applicable to open fields with a modified arrangement. Downstream analysis of the image pixels and intensity and their differences among plants can produce phenome information like growth, thermal or water dynamics of the plant depending on the cameras used to take images.

2.5.7.1 Advantages and limitations of high throughput phenotyping

Advantages of high throughput phenotyping are as below

1. Manual phenotyping is laborious and requires human involvement, time and more instruments which can be avoided using automated phenotyping. Sometimes a measurement needs to be taken within short time for large number of plants which requires involvement of multiple individuals and more than one instruments. Automated phenotyping can solve this problem.
2. It can takes measurement accurately and frequently, nullifying the human error.
3. Automated phenotyping is non-invasive. For example, to measure chlorophyll content it requires the sacrificing of leaves and the same plant might not be used for taking measurements in subsequent experiments. Which is not the case in automated phenotyping. Measurement of chlorophyll content can be availed from the image analysis.
4. In diverse environments many of the commercially valuable traits which are quantitative, exhibit high variability. Also throughout the plant's life cycles changes are observed which is not possible to capture by manual phenotyping in a certain period but is possible by continuous monitoring using image-based high throughput phenotyping.
5. Furthermore many phenotypic traits are not amenable for visual scoring and observations on a large scale. This phenotype-genotype gap can be bridged by using the high throughput imaging system.
6. Integration of mechanical automation along with this optical-based system enhances the ability to capture quantitative traits on a temporal and spatial scale with lower technical error.

Among the limitations, cost is still a major factor. Moreover the storage and handling of huge amount of data from everyday imaging makes downstream analysis difficult. But this is being taken care of with advent of new analysis tools and use of statistical models to explain the digital traits. Widespread adaptation in field is still underway (Hairmansis et al. 2014).

2.5.7.2 Classification of images and the hidden traits

Different types of cameras are available to capture signals from the visible and infrared spectrum of light. VIS cameras detect light in the visible range from ~400 to 700 nm and are used to measure the morphological, geometric and color properties of plants. VIS camera sensor arrays have three color sensors (blue, green and red) that are used to estimate the true color of each pixel. Infrared cameras capture wavelength above ~700nm (Fig 2.10, Fig 2.11). Wavelengths in the near infrared (NIR) sub-ranges are strongly reflected by plant tissue, relative to visible light (Seelig et al. 2008). The NIR and short-wave infrared (SWIR) covers ~900-1700 nm have many major and minor water absorption troughs that can be used to measure leaf relative water content index though it can also be influenced by leaf thickness also (Seelig et al. 2008; Fahlgren et al. 2015). Hyper-spectral imaging covers a long range of ~350nm to 2500 nm and is very promising in detecting abiotic and biotic stress symptoms. The latter however still has computational challenges and is relatively slower in capturing images compared to the others (Fahlgren et al. 2015).Chen et al. (2015) used steady-state chlorophyll fluorescence excited by blue light as a relative measure of photosynthetic health in a drought stress experiment in barley.

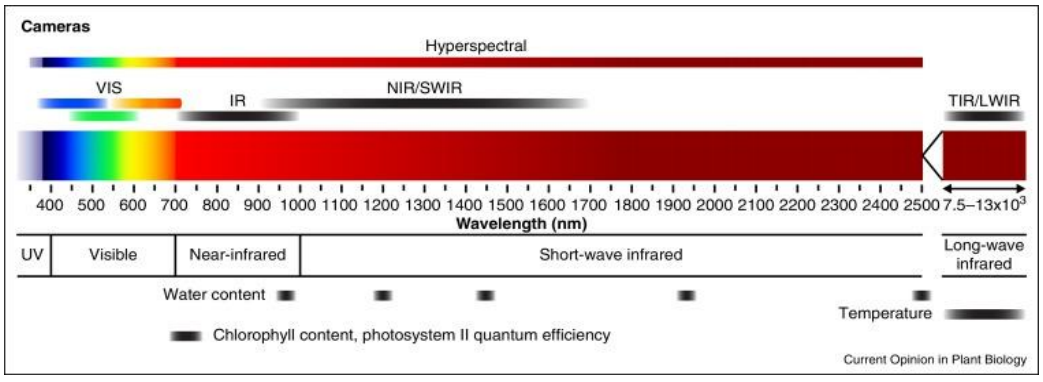


Fig: 2.10 Classification of lights according to spectrum (Fahlgren et al. 2015)

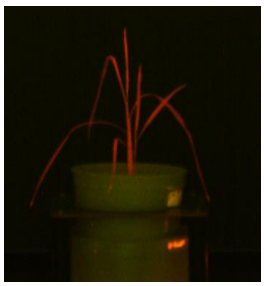
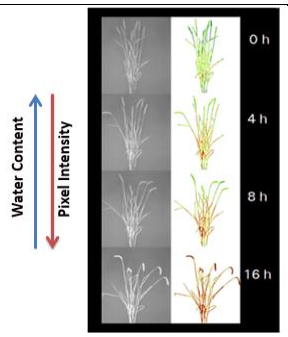
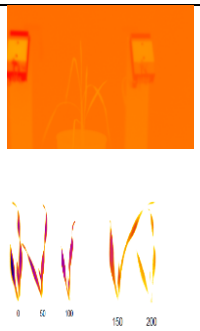
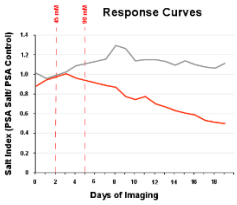
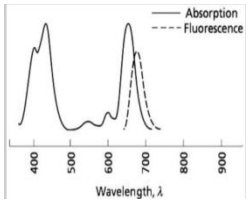
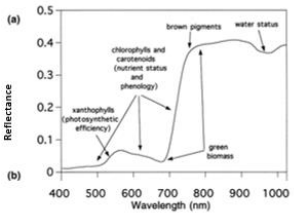
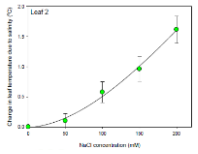
Visible camera (VIS): measuring growth dynamics, colors and geometry	*Fluorescence (Flo): Chlorophyll content (exclusion, PSII efficiency and tissue tolerance)*	*Near Infrared (Plant water content: for evaluating osmotic stress tolerance (more water in leaves)*	*Infrared (IR) leaf temperature. During salinity stress it increases*
			
			

Fig 2.11: Four major type of images taken for plant phenotyping

Many traits related to plant growth, architecture and development have been quantified from digital color imaging, while leaf color is a simple, under-utilized trait that indicates plant health and leaf senescence (Berger et al. 2012) and can be used in assessing the levels of both biotic and abiotic stress.

2.5.7.3 Analysis of images: pixels to trait

Development and application of high throughput phenomics technologies are now the focus of many research centers (Yang et al. 2013). Automated phenotyping facilities are available in the Australian Plant Phenomics Facility, European Plant phenotyping network, PHENOME: The French Plant Phenomics Network (FPPN), High Throughput Rice Phenotyping Facility (HRPF, Huazhign University, China), Donald Danforth Plant Science Center, University of Nebraska Lincoln Innvoation Campus, Leasy Scan at ICRISAT, etc. This normally involves availability

of a series of cameras that are used to capture images which help estimate plant growth, temperature, water content or chlorophyll characteristics on a large scale. The dynamic physiological changes occurring during the course of plant development (over several days) can be revealed with lower technical error.

After capturing the image, the underlying digital traits need to be extracted. Many image analysis tools are available. (Klukas et al. 2014) and Fahlgren et al. (2015) have both developed software for extracting biological information from images captured with LemnaTec platforms (http://www.lemnatec.com). Some of the available tools are HTPpheno, PlantCV, IAP, and others (Tanabata et al. 2012; Pound et al. 2013; Brown et al. 2014; Fahlgren et al. 2015; Müller-Linow et al. 2015). The number of files generated in a triplicated experiment with 3 weeks of multi camera imaging. About 300 genotypes can easily exceed 3 million which makes the file management tasks very difficult. Image harvest is an open source Python library and uses algorithm from OpenCV and SciPy, has been used in the current analysis which makes the hurdles easier (Knecht et al. 2016). IH has also been integrated in the open science grid providing computing resource to users at no cost. Another time consuming challenge in plant phenomics is correlation of image-extracted traits to empirically-measured and biologically relevant traits. More data can be generated from an image dataset with new algorithms analogous to reanalyzing sequencing data (Fahlgren et al. 2015). Accordingly, there are only relatively few well-curated high-throughput image phenotyping datasets available to the public e.g., plantCV danforthcenter page as phenome data repository.

2.5.7.4 Models to describe growth under salinity stress

Shortly after application of salt, just as stomata shut, inhibition of plant growth also occurs rapidly, which is independent of the accumulation of the salt. This provides an opportunity to separate salt-dependent and salt-independent components of plant responses to salinity. Manual phenotyping to detect incremental changes in growth rate during the osmotic stress and slight shifts in leaf color due to ionic stress are difficult to quantify for a large number of genotypes, but is feasible in everyday imaging technology. Al-Tamimi et al. (2016) have proposed that plant growth was neither exponential nor logistic throughout the imaging period, in particular, in the salt treated plants. So they fitted cubic splines (functions used in applications requiring data interpolation and/or smoothing) to data to characterize growth as suggested by (Li and Sillanpää) in order to describe complex growth trends. Splines can make no *a priori* assumptions about the shape of the curve thus giving growth and transpiration rates in an unbiased way while removing transient influences on the growth as described by(Al-Tamimi et al. 2016). In the current study spline fitting strategy has been applied to the data (detail in chapter 5). PSA positively correlates strongly with shoot biomass (Campbell et al. 2017) . Al-Tamimi et al. (2016) and Campbell et al. (2017) have taken most attention in functional analysis

from digital traits. Campbell et al. (2017) used a functional modeling of temporal trends in projected shoot area, modified from (Paine et al. 2012). Adoption of equations to describe growth to develop statistical modelling strategies have the ability to describe the growth dynamics more precisely.

2.5.8 Sequencing based allele diversity mining

For modern crop science, an important tool to improve crops is knowledge about nucleotide sequence variation so that it can be related to differences in trait. This variation can either be natural, from divergent populations, or induced through treatment with mutagens. Addition of a new allele to a population may cause that to become more competent to survive, or sometimes less capable. The variation can be insertions or deletions (together called InDels) or single nucleotide polymorphism (SNP). Allele mining is a promising way to dissect naturally occurring allelic variants of candidate genes and link it with essential agronomic qualities in different varieties. This allows rapid identification and characterization of allelic variants (Latha et al. 2004).

Sonia et al. (2012) genotyped 392 rice accessions by EcoTILLING targeting five key salt-related genes involved in mechanisms such as Na^+/K^+ ratio equilibrium, signaling cascade and stress protection, and they found 40 new allelic variants in coding sequences. They identified 11 significant SNPs related to salinity signifying allelic variants affecting salinity tolerance. Recent advances in sequencing technology facilities the allele diversity mining using whole genome sequence. Allele diversity mining in the stress associated QTL regions are being carried out using the salt tolerant donor and recipient parents (Platten et al, unpublished work) The OryzaSNP project and 3K genome sequence project(2014)facilitates this effort (McNally et al. 2009). Combination of screening of the diverse rice germplasm, phenotype characterization for identifying important alleles will result most out of it.

2.6 Current understandings on salt tolerant determinants

2.6.1 QTLs mapped for salt tolerance

Several salt tolerance-related quantitative trait locus (QTLs) associated with parameters such as Na^+ and K^+ ion uptake, ionic concentration and Na^+/K^+ ratio has been reported in rice (Koyama et al. 2001; Lin et al. 2004; Singh et al. 2007; Sabouri et al. 2009; Ul Haq et al. 2010). These QTLs are detected repeatedly on chromosomes 1, 4, 6, and 7. None have been found on chromosomes 8 and 11 and very few on chromosomes 2, 3, 5, 9, 10 and 12 (Negrão et al. 2011). However, it should be highlighted that most of these studies have been conducted under

hydroponic conditions and only one used field conditions (Takehisa et al. 2004). Salinity stress was shown to have different impact in plants grown under hydroponic and soil system. A segment of the short arm of chromosome 1 concentrates a large number of QTLs. A major QTL *Saltol* has been identified from a cross of IR29/Pokkali on chromosome 1 in rice, associated with seedling-stage salt tolerance and Na^+/K^+ ratio explaining 43-70% of phenotypic variation.(Gregorio et al. 1997; Bonilla et al. 2002). There are reports of introgression of this *Saltol* QTL into rice varieties like BR11, BRRIdhan28, Q5DB, IR64, AS996 and PB1121(Gregorio 2013; Huyen et al. 2013; Hasan et al. 2015; Babu et al. 2017). However since the advantage gained was mainly at the seedling stage, considerable loss in grain yield under salt stress was observed. A <1 Mb chromosomal fragment from Pokkali was revealed at 10.6-11.5 Mb (Kim et al, 2009) containing shoot K^+ content 1 (SKC1) and other Pokkali derived loci (Thomson et al. 2010).SKC1 was originally mapped as a QTL, qSKC1(Lin et al. 2004) later fine mapped containing *OsHKT1;5* gene encoding for a sodium transporter regulating K+ homeostasis (Ren et al. 2005)

Since the focus is now shifting towards pyramiding of genes or QTLs very recently (Pang et al. 2017) carried out a designed QTL pyramiding experiment to develop high yielding Green Super rice, with significantly improved tolerance to salt stress and grain yield.

2.6.2 Candidate genes co-segregating with QTL regions

Genes co-segregated with known salt stress associated QTL are of most important and Negrão et al. (2011) has summarized them in 3 main groups of QTL traits (ion uptake, plant development and salt tolerance). Interestingly QTLs co-segregating with candidate genes encoding transcription factors concentrate in a small region in the long arm of chromosome 1 in which 2 QTLs have been identified (NQR 2092 and DSD2085). *OsHKT8* co segregates with 13 QTLs and is involved in Na^+ removal from the xylem sap, guaranteeing salt exclusion from leaves. salT, is a mannose binding lectin overlapping with 10 QTLs. Os01g25280 codes for a protein with unknown function and co segregates with 9 QTLS (Negrão et al. 2011). All these three genes overlap with the *Saltol* QTL. Nutan et al. (2017) has narrowed down 14 transcription factors from the *Saltol* QTL region and classified their roles in salinity response and in plant growth and development.

2.6.3 Candidate genes identified from transcriptomic studies conferring salt tolerance

Transcriptomic studies on tolerant and sensitive rice cultivars have revealed differential expression of genes from many functional classes in spatial and temporal manner (Walia et al. 2005; Walia et al. 2007; Kumari et al. 2008; Senadheera et al. 2009). This emphasize existence

of diverse mechanisms in tissues in response to salt. Negrão et al. (2011) published a precise table of genes shown to be responsible in tolerance mechanism approved by semiRT, RT and northern blot. Here very briefly the classes of such genes will be reviewed.

2.6.3.1 Signaling:

As the first step in response to salt, transient increase in cytosolic Ca^{2+} has been recorded. Major groups of Calcium binding proteins includes Calmodulin (CaM), CAM like proteins (CML), Ca^{2+} dependent protein kinase (CDPKs) and calcineurin B like proteins (CBLs)(DeFalco et al. 2009). In rice five putative CaMs and 32 CMLs proteins were identified in a genome wide sequence homology study(Boonburapong and Buaboocha 2007) and (Phean et al. 2005) reported induction of Calmodulin-1 transcript levels by salt stress. Within the CDPK family *OsCPK13 (OsCDPK7)* was found to be overexpressed in both cold and salt stress (Saijo et al. 2000). CBL proteins are mainly regulatory subunits for plant specific serine/ threonine protein kinases, known as CBL-interacting protein kinase (CIPKs) (Weinl and Kudla 2009). The classical example is SOS pathway. AtCBL4 (SOS3) binds AtCIPK24 (SOS2) in a Ca^{2+} dependent manner. Rice homologues for SOS2 and SOS3, *OsCIPK24* and *OsCBL4* were identified (Martínez-Atienza et al. 2007). *OsCIPk15* overexpression in rice enhanced tolerance. And in FL478 it was shown as downregulated and no change in IR29 (Senadheera et al. 2009). Mitogen activated protein kinase (MAPK) cascades are composed by 3 types of kinases. MAPKKK are activated by signal plasma membrane receptor and activate downstream MAPKK in turn activating MAPK. The downstream targets of MAPK can be TFs, phospholipases or cytoskeletal proteins. In rice evidence of MAPK is not as much as Arabidopsis but the overexpression of *OsMAP5a* leads to salt drought and cold stress increased tolerance (Xiong and Yang 2003). A Na^+ dependent receptor like kinase (RLK), *OsSIK1* was found to be involved in scavenging and detoxification of reactive oxygen species and over expression of which induces salinity and drought tolerance by increasing antioxidant activity (Ouyang et al. 2010).

2.6.3.2 Ion Homeostasis

Membrane transport proteins are of immense importance for their role in regulation of fluxes of ions, nutrients and other molecules across the membrane. An important mechanism of salt tolerance is the selective uptake of K^+ over Na^+ by root system, (Maathuis and Amtmann 1999). The HKT (High Affinity K^+ Transporter) family resides at the plasma membrane and permeable to either K^+ or Na^+ or to Na^+ only (Rodríguez-Navarro and Rubio 2006). Rice has 7 to 9 genes of HKT family depending on the cultivar (Garciadeblas et al. 2003). *OsHKT8* is involved in regulating K^+Na^+ homeostasis in the shoots (Ren et al. 2005) found allelic difference at *OsHKT8* between Nona Bokra and Koshihikari in 6 nucleotide substitutions in the coding region leading to 4 amino acid changes enhanced overall Na^+ transport activity. Miyamoto et al (2015) found that *OsHKT2:1* overexpression promotes shoot Na^+ accumulation under low

K^+ supply but under sufficient K^+ supply, *OsHKT2;1*-overexpressing rice plants accumulates Na^+ in roots but not in shoots. This led them to suggest that Na^+ transferred from roots to shoots may be regulated by another Na^+ transporter. On the other hand, rapid inactivation of *OsHKT1* was found in under high Na^+ concentration. Works by Horie et al. (2007) suggested that Na^+ transport activity of this gene is post transcriptionally regulated.

Shaker family proteins are also involved in ion homeostasis. Both *OsAKT1* and *OsKAT1* are K^+ channels that maintain a low Na^+/K^+ ratio due to increase of K^+ uptake and decreasing Na^+ (Fuchs et al. 2005; Obata et al. 2007). In tolerant varieties, *OsAKT1* is down-regulated, while in the sensitive it is not repressed (Golldack et al. 2003). It's expression does not respond to external K^+ concentrations and the Na^+ accumulation or exclusion in the whole plant depends on the AKT1 expression in specific cells of salt-stressed plants(Fuchs et al. 2005). In arabidopsis (Lee et al. 2007) described upon phosphorylation *CIPK23* kinase interacts with the ankyrin domain of AKT1 activating it.

Increased expression of antiporter genes by salt stress has been reported for both glycophytes and halophytes (Blumwald and Poole, 1985, Gaxiola et al, 1999, Hamada et al, 2001). The acidic environment inside membrane bound vacuoles in plant cells allows the efficient compartmentalization of Na^+ in the vacuole through vacuolar Na^+/H^+ antiporter (Apse et al, 1999). They use proton motive force to catalyze the exchange of Na^+ for H^+ across vacuolar membranes (Zhang and Blumwald, 2001). This vacuolar sequestration lowers Na^+ concentration in the cytoplasm as well as contributes to osmotic adjustment in the cytoplasm to maintain water uptake from saline solutions (Zhu, 2003). Fukuda et al, (2004), Amin et al, (2016) clearly demonstrated that overexpression of *OsNHX1* improved the salt tolerance of transgenic rice plants. They found Na^+ content in seedlings and young leaves were similar in wild type and transgenics but different in mature leaves which suggested that the Na^+ compartmentation in mature tissues could prevent young tissues from being more severely injured. Both tonoplast and plasma membrane antiporters exclude Na^+ from the cytosol by proton motive force generated by the plasma membrane H^+ATPase and by vacuolar membrane H^+ ATPase and H^+-pyrophosphatase and it has been shown that the activity of these proteins responds to salinity In work of (Senadheera et al. 2009) *OsCHX11* and *OsCAX4* involved in cation transport have shown differential expression after salt treatment.

2.6.3.3 LEA and other stress-induced proteins

Salt and drought share the water deficit condition, and also some of the genes are also shares. Late Embryogenesis proteins (LEA) accumulates to high level both during the last stage of seed maturation and during water deficit in vegetative organ. In rice 25 new LEA genes have been identified (Wang et al. 2007) and was grouped in 7 distinct families according to pfam classification. Salinity induced overexpression of *OsLEA* was shown by (Wang et al. 2007; Hu

2008). LEA group as protector molecules under water limitation and also in providing a water like rich environment. SalT gene is one of the most reported co-localizing with the QTL Saltol on chromosome 1 and is regulated by both ABA dependent and independent pathway (Garcia et al. 1998). Role of SalT protein is not restricted to environmental stress, but in older plants higher expression level is found in oldest leaves (Garcia et al. 1998).

2.6.3.4 Transcription Regulators

Transcription factors are proteins with sequence specific DNA binding and capable of activate or repress gene expression. In plants, AP2/EREBP (apetala2/ethylene Responsive Element Binding protein), NAC(NAM, ATAF and CUC), ZF-HD(Zinc finger homeodomain), AREB/ABF(ABA responsive element binding protein, AREB/ABF(ABA – responsive element binding protein /ABA binding factor) and MYC (Myeolocytomatosis oncogene)/MYB (myeoloblastosis oncogene) have been the most responsive to abiotic stresses. *OsDRB1A* and *OsDREB1F* are induced by high salinity and cold (Dubouzl et al, 2003, Liu et al, 2007) but *OsDREB2A* is induced by salt drought and temperature but not cold (Matsukura et al, 2010). So DREB1CBFs are ABA independent and related to low temperature and DREB2 is ABA dependent and more related to osmotic stress. Overexpression of *OsBHD1* reduces salt tolerance (Oh et al, 2009).

Recently Ambavaram et al (2014) have narrowed down a master regulator directly affecting the yield and named it as Higher Yield Rice (HYR). Under water deprivation, in the HYR lines higher RWC was observed and they related it with accumulation of soluble carbohydrates, leading to osmotic adjustment. Functional analysis on the HYR protein activities revealed its involvement in direct transcriptional activation of multiple photosynthesis related process and activation of auxin responsive TFs.

Three genes encoding zinc finger proteins have been reported as responsive to high salinity. (Huang et al, 2008; Huang et al 2007; Sun et al, 2010). *ZFP179* and *ZFO182* induced while SRZ1 repressed by salt stress working as positive and negative regulator of salt tolerance (Negrão et al. 2011).

Four genes encoding bZIPs (*OsABF1*,*OsABI5*, *OsbZIP23* and *OsBZ8*) are induced by salt stress response and positively responsive to ABA. (Hossain et al, 2010, Zou et al, 2008). Overexpression of OsbZIP23 showed improved salt tolerance whereas OsABI5 overexpression leads to high salt sensitivity stating it as a negative regulator (Zou et al, 2008). Other TFs like OsMYB3R-2, OsGTγ-1 and OsHsfA2e genes are induced by high salinity and improved tolerance in Arabidopsis (Fang et al. 2010). Overexpression of OsMYB3R-2 also increased tolerance to low temperature and drought proving them as a master switch in stress tolerance (Dai et al. 2007).

Out of 42 TF responsive to salt stress in rice, 25 belong to the NAC (NAM, ATAF1/2, CUC2) family (Mohanty et al. 2002). Negrão et al. (2011) suggested *OsNAC* genes are normally

expressed at low levels and may be induced when plants are subjected to adverse environmental conditions. Transgenic plants overexpressing SNAC1 are more sensitive to ABA and close more stomatal pores to avoid loss of water (Hu et al. 2006). The SNAC2 gene was found to be associated with blast disease resistance apart from salt (Nakashima et al. 2007). A subset of the NAC members *OsNTL2* to *OsNTL6* is induced by high salinity and mannitol and are responsive at transcriptional level.

Small RNAs have an important role controlling the transcript level of many TFs involved in abiotic stress responses (Sunkar et al. 2012).

2.6.3.5 Post translational protein modifications (PTM) in salt stress response

PTMs in plant response to salt stress is regulation of the Salt Overly Sensitive (SOS) pathway. Rice SOS1 was isolated and shown to be phosphorylated and activated by the SOS2/SOS3 protein kinase complex (Martínez-Atienza et al. 2007). This pathway also require N-myristoylation of SOS3 to function, which is, a co-translational protein modification occurring in N terminal glycine residue of some proteins. Analysis of rice phosphorylated proteome resulted in the identification of over 30 phospho proteins differentially regulated in response to salt stress (Chitteti and Peng 2007). Other important PTMs in abiotic stress response are ubiquitination and binding to ubiquitin like proteins like SUMOylation and are important in regulation of cellular processes for their role in protein stability, cellular localization and in protein-protein interactions. Ubiquitin-activating enzyme (E1), Ubiquitin-conjugating enzyme (E2) and Ubiquitin ligase are involved in this process. Soybean E2 enzyme GmUBc2 is found to be induced by both salt and drought stress and its overexpression in Arabidopsis resulted in tolerance improvement to both stresses (Zhou et al. 2010). It also increase transcripts of AtNHX1 and AtCLCa known to be involved in ion homeostasis.

SUMOlyation is the conjugation of Small Ubiquitin related modifier (SUMO) to proteins via an enzymatic pathway. In rice, direct evidence of SUMOylation is less compared to other species, e.g. OsSZ2. Nevertheless rice SUMO-conjugates gradually accumulate during the first 12h of 250 mM NaCl exposure and decreases after 24 h (Chaikam and Karlson 2010). Other PTMs like histone acetylation is expected to play important roles in the control of gene expression during salt stress conditions. In rice, histone deacetylase (HDC) family did not show a central roles in salt stress response (Fu et al. 2007), but is present in Arabidopsis.

2.6.3.6 Osmo-protectants

Proline is an essential amino acid and needed to maintain pH of cytosolic redox of cell and as antioxidant or singlet oxygen quencher (Matysik et al. 2002). And it increases in salt tolerant transgenic rice plants (Karthikeyan et al. 2011). Mohanty et al. (2002) suggested that transgenic rice harboring the *codA* (choline oxidase) gene is highly tolerant to salinity. *codA* introduced into transgenic rice plants promotes the synthesis of glycine betaine and enhancement of salt

tolerance in plants (Su et al. 2006; Li et al. 2011) demonstrated that overexpression of *OsTPS1* in rice plants enhances salt tolerance by increasing the amount of trehalose and proline. Trehalose has now been shown to be a regulatory molecule and over-expression of trehalase causes closure of stomata and protects plants from high salt-water intrusion by lowering transpiration. In rice, the expression of aquaporin *OsTIP1;1* is up-regulated during response to salinity stress (Sakurai et al. 2005). Sakurai et al. (2008) showed that rice aquaporins have a role in maintaining the water potential in different tissues and cells. Despite their role in salinity stress involvement of osmoregulators are shaky when it comes to enough yield under stress (Serraj and Sinclair 2002).

2.6.3.7 Other candidate genes

The up-regulation of *OsNRT1;2*, a gene coding nitrate transporter, advances the accumulation of NO_3^- in the old leaves of rice plants under salinity (Wang et al. 2012). Xu et al. (2016) indicated that nitrogen metabolism under salt stress is enhanced and rearranged to synthetize more amino acids as the compatible solute to cope with adverse conditions. *PDH45* is involved in the regulation of Na^+ level, ROS production, Ca^{2+} homeostasis, cell viability and cation transporters in roots of *PDH45* transgenic rice and consequently provides salt tolerance (Amin et al. 2012)

Expression level of *OsBURP* in various tissues and organs revealed different temporal spatial pattern and was found to be responsive to ABA and salt treatments. Response of these genes to salt stress is extremely evident suggesting then as excellent candidates for breeding purpose (Negrão et al. 2011).

2.6.4 Expression QTL mapping: case studies in rice

Most published studies on plant eQTL have been con-ducted in Arabidopsis (Keurentjes et al. 2007; West et al. 2007), maize (Schadt et al. 2003; Shi et al. 2007), wheat (Jordan et al. 2007). Wang et al. (2010) have conducted an eQTL study on rice shoots at 72h after germination from 110 recombinant inbred lines (RIL) derive from a cross between Zhenshan97 and Minghui 63 and identified 26051 eQTLs including both cis and trans eQTLs and 171 eQTL hotspots. They reported enrichment of DNA metabolic process related e-traits in hotspots residing in chromosome 3, 5 and 10. Correlation between shoot dry weight and eQTLs revealed possible candidate genes for the trait. Another eQTL search in rice was carried out by Wang et al. (2014) using 210 RILs derived from same parent as above on the RNA samples of flag leaves at the heading stage with Affiymetrix whole-genome arrays They identified 13647 eQTLs and 138 trans eQTL hotspots and putative regulators in flowering and yield, suggesting candidate genes. Liu et al. (2010) identified 565 genes combining gene expression variations and InDel result togethers demonstrating both InDel variations and differential expression between rice *indica* and *japonica* subspecies. Horiuchi et al. (2015) observed that expression differences of most

global DE genes were regulated by *cis*-eQTLs and expression evolution of changed tissue DE genes were regulated not by *cis*-eQTLs, but by complicated *trans*-eQTLs.

2.6.5 Removing the genotype-phenotype gap: case studies in rice

Non-invasive image phenotyping facilities have created the opportunities for researchers to study the plant early growth under stressed conditions. Recent studies by Moore et al. (2013) , Yang et al. (2013) and Würschum et al. (2014) utilized image-based phenomics to identify QTLs involved with temporal developmental dynamics in Arabidopsis, rice and wheat, respectively. Recently Al-Tamimi et al. (2016) have conducted a phenome study on indica and aus panel of rice followed by GWAS which identified a promising region at chromosome 11 at ~ 3.62-3.76 Mb residing a receptor kinase (Os11g07230) and a serine threonine protein kinase BRI1 like 2 precursor along with other important genes. They have identified candidate genes involved in TUE (Transpiration Use Efficiency) in early and late stage of growth revealing that early intervals of 2-6 and 2-9 days after treatment mainly relates to encoding of signaling proteins, supporting the hypothesis that the early response to salt stress are related to signaling mechanism. Several significant peaks were identified in chromosome 11 of the indica panel and chromosome 5, 9, 11 of the aus panel at the early time interval (2–6 days after treatment) associated with TUE under saline conditions that disappeared from later time intervals. This explains the beauty of high throughput phenotyping study as TUE is a derived complex trait that has often been phenotyped using surrogate traits. The high throughput method could differentiate the Aus and indica panel, showing that a greater decrease of transpiration rate occurred in the aus panel(Al-Tamimi et al. 2016).

While revealing the genetic architecture of shoot growth dynamics, Campbell et al. (2017) identified Gibberellic acid as a main component for natural variation for shoot growth dynamics in rice in their study on ~360 diverse rice accessions belonging to each of the five major rice subpopulations (*indica, aus, aromatic, temperate japonica*, and *tropical japonica*). Accessions belonging to the major allele group displayed significantly higher expression of OsGA2ox7, slower growth rate during the early tillering stage and lower GA4 levels consistent with the report by Lo et al. (2008). Interestingly they have found a significant positive relationship between the early and active tillering stage using Projected Shoot Area (PSA) suggesting that growth characteristics observed at early tillering stage partly persists during the latter tillering stages. In ~3Mb on chromosome 8, the largest effect on projected shoot area during the active tillering stage was observed.

For digital traits $Den1_{TV}$ and GH1 (which can be considered as proxies for canopy density and erect growth habit) significant peaks were identified in chromosome 6 and 4, notably a rice gene OsRCN4 (LOC_Os04g33570) similar to *TERMINAL FLOWER 1 (TFL1)* and several

other genes related to flowering time and morphology in a study on 373 accessions using GWAS in with 44k SNP data (Knecht et al. 2016)

In another study Campbell et al. (2015)could identify the fluorescent signal classes as digital traits between the 7th day and 11th day after 90mM NaCl stress in a study on rice. Genetic analysis at discrete time points could identify time-specific QTLs that partially influence the final phenotype. Since this was only detected at a specific time of the growth stage, this suggested that longitudinal phenotypes have advantage over cross sectional (i.e., endpoint) phenotyping data.

Chapter 3

Phenotyping and QTL mapping

3.1 Overview

The primary goal of the study was to map salt tolerance determinants in *Horkuch*, the rice landrace which is the subject of this study. To track down these determinants one promising strategy is mapping of the quantitative trait loci (QTL) responsible for its salt tolerance. Salt tolerance is however governed by multiple genes and termed as a quantitative (polygenic/multifactorial/complex) traits. We therefore need to look at multiple phenotypic traits associated with tolerance in order to explore the trait-to-gene relationship. The regions in the genome containing genes associated with quantitative traits is termed QTL. Tracking down these loci requires the following steps:

(1) Construction of a segregating population of progenies crossed from two contrasting parents,
(2) Analyzing polymorphic markers between the parents,
(3) Scoring the progenies as homo or heterozygotes with respect to parental alleles,
(4) Mapping relative orders and distance of the polymorphic markers,
(5) Analyzing the linkage between markers by calculating the recombination frequency and
(6) Associating the trait of interest with the linked markers by tracking polymorphism in phenotype as well as genotype.

In the study *IR29* was used as the 2nd parent. This is salt sensitive as well as high yielding i.e. it possess contrasting characters compared to *Horkuch* as the ultimate focus was to derive information not only on vegetative salt tolerance, but also on obtaining high yields under stress. This chapter focuses on

1. Information on mapping population construction and phenotypic screening of F_3 population generated from a reciprocal (2-way cross) of *Horkuch* and *IR29*, i.e., *Horkuch*♀ and *IR29*♀ population. Extensive screening was performed at both seedling and reproductive stages. Specific traits affected by and related to salt stress tolerance as well as traits related to yield under stress were measured. Their distribution, correlation with each other and heritability was also observed.
2. Genotyping of the segregating F_2 population from SNP (Single Nucleotide Polymorphism) markers derived from DArTseq technology and construction of a linkage map.
3. QTL mapping of traits relevant to salt tolerance in both seedling and reproductive stages, with emphasis on yield traits under salt stress for the latter.

The constructed linkage map and the genotype information are important resources that have been used in the subsequent chapters as a basis for identification of eQTLs (chapter 4) and QTLs for digital traits derived from automated phenotyping(Chapter 5).The mapped QTLs also identify significant regions important for salt tolerance.

3.2 Materials and methods

3.2.1 Generation of the mapping Population

The rice cultivar *Horkuch* (IRGC 31804) and *IR29* (IRGC 30412) were collected from International Rice Research Institute (IRRI) Genebank (collection site/origin of *Horkuch* was mentioned as Bangladesh). The collected seeds were sown at IRRI crossing block during wet season of 2011 (June-July) after breaking dormancy at 70˚C for 5 days in an oven. Two reciprocal crosses were made between *Horkuch* and *IR29* during October to November, 2011 (registered as IR102583 with GID: 3539832 [F_1(Horkuch / IR29)], and IR102584 with GID:3539833 [F_1(IR29 / Horkuch)],). F_1 plants were confirmed by SSR marker RM493 at the University of Dhaka, Bangladesh and advanced to F_2 in Plant Physiology Division of BRRI (Bangladesh Rice Research Institute). Using single seed descendent method F_2 plants were selfed to F_3 and so on.

3.2.2 Evaluation of F_2 plants

About 1200 F_2 progenies from both crosses (600 from each cross) were planted in the field at BRRI for advancing to the generation F_3. *Horkuch* has some distinct characters in contrast to *IR29* e.g., red and bold grain, tall height, red culm, compact tiller etc. Therefore the F_2 plants were evaluated for these characteristics and the segregation was observed without any applying any stress. Simple observation of some characters were noted, such as grain color (Red/Green/Red-like), grain size (Bold/Slender), visual height (not measured, but qualitative, e.g. – Tall/Intermediate/Short), culm color (Red/Green), tiller pattern(Compact/Spread), grain weight (High/Low).200 plants from *IR29*♀ and 100 plants from *Horkuch* ♀ were then randomly selected for DNA isolation and genotyping.

3.2.3 Phenotypic evaluation at F_3 population

3.2.3.1 Seedling stage phenotyping:

3.2.3.1.1 Growth condition

One and two hundred F_3 progenies from *Horkuch*♀ and *IR29*♀ crosses were randomly chosen for phenotypic characterization against salinity stress at seedling stage. Seeds were germinated on a wet filter paper in Petri dishes and were kept inside a seed germinator at 30˚C and 75% relative humidity. Well- germinated seedlings of the same size were planted on netted Styro-foam sheet float in a plastic tray (Fig 3.1) containing Yoshida culture solution (Yoshida, et al. 1976) in 3 replicates following Incomplete Block Design (balanced). Each block accommodated 19 lines and two parents and one tolerant check (FL378), Fig 3.1. Only 4

progenies from each line/parent/check were placed in rows per block and all 22 genotypes in each block were subjected to complete randomization. The Yoshida solution was changed once a week and pH of the solution was always kept 5.0±0.5 throughout the experimental period. Salt stress was applied 12 days after seeding by changing the normal Yoshida solution by saline solution, but the salt was gradually applied at 2dS increment per day, starting from $6dSm^{-1}$ on 13 day-old seedlings till $12dSm^{-1}$ on day 4. The progenies from the two reciprocal crosses were screened in two experimental setups, in overlapping time periods; IR29♀ (May 18th to June 16th) and Horkuch♀ (June 1st to July 1st, 2013). Temperature and humidity recorded at that period were more or less similar (averaging 27°C at night and 34.9°C, at day-time) so that any experimental differences between the 2 populations was not caused due to separate experimental setups.

3.2.3.1.2 Screening of the traits

Parameters like SES (Standard Evaluation System) score, Total chlorophyll content (TChlr), stomatal conductance (SC), shoot relative water content (SRWC), root relative water content (RRWC), shoot length (SL), root length (RL), shoot dry weight (SDW), root dry weight (RDW), total sodium (TNa), total potassium (TK) were measured in the salt-stressed plants.

SES (Standard Evaluation System) scoring: The overall salinity tolerance at seedling stage was evaluated mainly based on the value of leaf damage score named as SES score (IRRI, 1996) where a scale of 1-9 corresponds from highly tolerant to extremely sensitive.

SC (Stomatal Conductance): Stomatal conductance of fully opened young leaves were measured after 7 days of salt stress by a Decagon Leaf Porometer (sensor serial LPS1283) (Decagon inc., USA) (Fig3.1) during a bright sunny day from 11 am to 2 pm.

RWC (Relative Water Content): Relative water content of shoot and root was measured from the percent ratio of the difference between fresh (W) and dry weight (DW) and the difference between turgid and dry weight. For measuring turgid weight (TW), samples were hydrated by floating on de-ionized water in a closed petri dish for 24h and after taking out were well dried of any surface moisture lightly with tissue paper followed by weighing immediately. For the dry weight(DW) measurement, the samples were oven-dried at 80°C for 24h and weighed (after being cooled down in a desiccator (Barrs and Weatherley 1962).The formula for calculation is provided in Table 3.1.

TChlr (Total Chlorophyll): For measuring the chlorophyll content, a specifically weighed leaf was cut into ~$1cm^2$ pieces and soaked in 20mL 80% acetone in dark. After 48 h, the absorbance was taken at three different wavelengths; 645 nm, 663 nm for Chlorophyll a and b and 652 nm for total chlorophyll (Yoshida, et al. 1976). Calculation was done using rice physiological study manual (Yoshida et al.) following Hiscox and Israelstam(1979) (Table 3.1).

Na^+ and K^+ ion measurement: Plants were washed in flowing tap water for 30 sec and oven dried for the measurement of sodium and potassium concentrations in seedling shoot and root. Dried leaves from each replicate were pooled, ground and analyzed by a flame photometer Sherwood 410 (Sherwood, UK) after 48 h of extraction with 1N HCl following the procedure described by Yoshida, et al. 1976. Concentrations of Na^+ and K^+ were expressed as percent of dry weight and mmole/g dry weight.

Length and Weight: Shoot and root lengths were measured using meter scale after harvesting the plants. Shoot and root dry weights were measured after drying in a hot air circulating oven for 72h at 70°C (ALP, Japan).

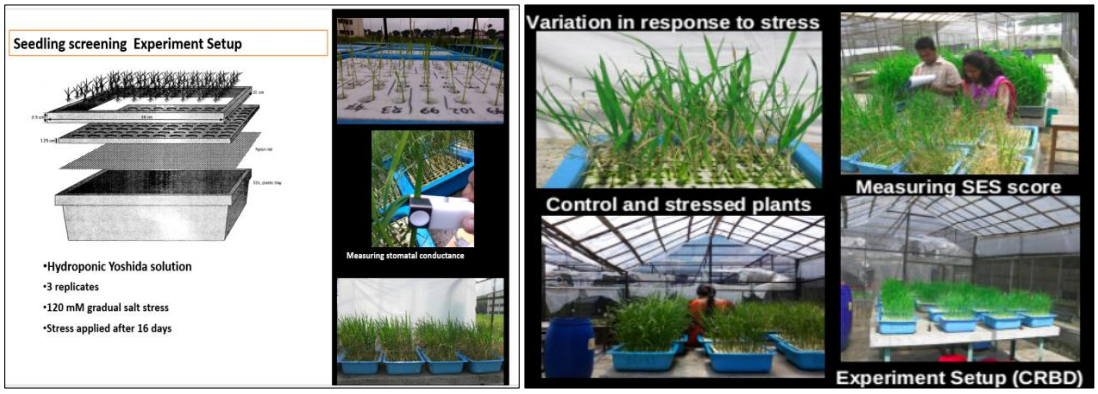


Fig 3.1: Seedling stage phenotyping

Table 3.1: Formulas used for calculation of physiological traits

Trait	**Formula**
Relative water content	**RWC = [(W-DW)/(TW-DW)] × 100,** *where, W = sample fresh weight, TW = Sample turgid weight, DW = Sample dry weight*
Chlorophyll content	Total Chl(A+B) =0.00802 × $Abs_{.\,663nm}$ + 0.0202 x $Abs_{\,645nm}$ Chl_A= 0.0127× $Abs_{.\,663nm}$ – 0.00269 × $Abs_{\,645nm}$ Chl_B = 0.0229 × $Abs_{\,645nm}$ - 0.00468× $Abs_{.\,663nm}$ Total Chl = $\frac{Abs652nm}{34.5}$ mg chl per gram fresh weight (Totalchl) = TotalChl/Leaf fresh weight
Thousand grain weight	$\frac{\text{Filled grain weight} \times 1000}{\text{Filled grain number}}$
Total grain weight	Filled grain weight + unfilled grain weight

Spikelet fertility	$\frac{\text{Filled spikelet Number} \times 1000}{Total\ Spikelet\ Number}$
Total Straw weight	Summation of leaf, straw and stem weights
Harvest index	$\frac{\text{Total grain weight}}{\text{Total grain weight} + \text{Total weight of above ground tissue}}$

3.2.3.2 Reproductive stage Phenotyping

3.2.3.2.1 Growth condition

One hundred lines from both cross *Horkuch* ♀and *IR29*♀ totaling 200 progenies were selected based on SES scores during seedling stage phenotyping (25% of the extreme tails) for reproductive stage characterization. Phenotyping was carried out in a net house condition with a controlled saline environment by the method described in Gregorio, et al. 1997. Soil was collected from a rice field, sun-dried, ground by mortar and pestle and placed in small perforated plastic buckets kept within a large plastic bowl which is filled with water or salt water as described later. Each bowl accommodated 6 pots. At least four replicates to represent each F_2 plants were used. NaCl salt was applied at 35 days after seeding of sprouted seeds to the soil according to Gregorio, et. al. 1997. Salt was applied by replacing tap water with saline water (8 dS/m) in all bowls and the top level was marked in order to keep the water constant in each bowl by topping-up the evaporated water. Two populations were phenotyped in the *T. Aman* season (July to December) of two different years, where the IR29♀population and Horkuch♀ population were characterized in 2013 and 2014 respectively. All cultural managements i.e. fertilizer, weed and disease-insect managements were done according to the recommendation of Gregorio et al. 1997 and BRRI, 2013.

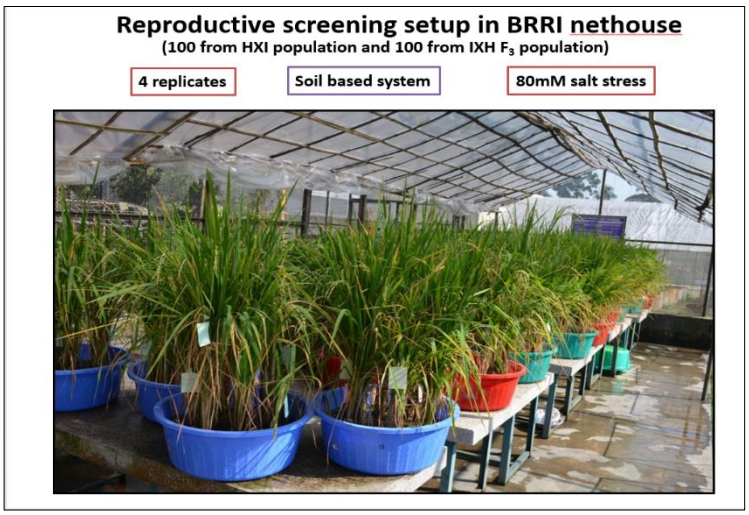


Fig 3.2 Reproductive stage phenotyping

3.2.3.2.2 Screening of the traits

The following phenotypes were measured at maturity: Days to flowering, days to maturity, plant height, total number of tillers, percent of effective tiller, panicle exsertion, panicle length, panicle damage, primary and secondary branching, filled and unfilled grain number, filled and unfilled grain weight, thousand grain weight, spikelet fertility, rachis weight, seed length and breadth, flag leaf length and weight, second leaf length and weight, third leaf length and weight, leaf sheath weight, lower leaf weight, stem weight, total straw weight, harvest index, total sodium and potassium content in flag leaf. Measurement of sodium and potassium was carried out using the dried flag leaf tissue, following the same procedure as described above for the seedling stage.

3.2.4 DNA extraction and genotyping by DArTseq

Genomic DNA of the F_2 population and parents was extracted by using CTAB method(Doyle and Doyle, 1987) from 1g of fresh leaf tissue after freezing in liquid nitrogen and grinding. Extra precaution was taken during the purification steps. Genotyping was done by the DArTseq technique as described by(Akbari et al. 2006). The DNA samples were sent to Diversity array technology Pty. Limited, Australia for DArT genotyping

3.2.5 Linkage map construction

All the analysis for linkage map construction were done with qtlTools (Lovell) and R/qtl (Broman et al. 2003) packages. We have filtered all the DArTseq SNPs and retained only loci that, a) are homozygous for both the parents b) are polymorphic between parents and c) represent minimum 50% of the individuals in the population. Among the 2235 loci found, 818 markers which showed significant segregation distortion by chi-square test [P value < $1e^{-2}$] were removed. Similar markers (649 markers) were further dropped using dropSimilarMarkers function using qtlTools package (Lovell et al, 2016) in R with a minimum recombination fraction threshold of 0.03.

Alleles were masked by calc.errorlod function using r/qtl package that have error.prob threshold less than 0.005 using “kosambi” map function. To get the correct order of markers we used tspOrder function of qtlTools package which applies a traveling salesperson problem solver using Hamiltonian circuit to fix marker order. To validate maker order with the annotation of SNP positions in a physical map we used matchMarkerOrder function of qtlTools package and removed those markers that have different orders in genetic map vs. physical map. Before estimating the linkage map, a few alleles were masked manually that showed erroneous call

and finally this linkage map was estimated by est.map function with “kosambi” map function using an error probability threshold of 0.001.

3.2.6 QTL analysis

To test the assumption of normality, Shapirio-Wilk test was performed for each trait. For QTL analysis, rank transformation was done for traits that did not have a normal distribution [P value < 0.05]. Recombination probabilities were calculated for each locus and inferred for a gap of more than 1 cM by placing a pseudo-molecule. QTL mapping was executed using the Haley-Knott regression algorithm in the R/qtl package (Broman et al, 2003). Multiple-QTL models were generated using a penalized stepwise model selection approach where significance was determined for each trait by 1000 permutations (significance threshold, alpha=0.05). Cytoplasm/cross-direction was found to be significantly co-related with many traits. However, it was not initially included as a covariate during the computation of permutation threshold as it may reduce the power to detect a local or small-effect QTL. The cytoplasm was used later as a covariate to explain residual variance while fitting the best model. Confidence intervals (1.5 LOD drop) for each QTL were calculated using lodind function of the R/qtl package (Broman et al, 2003) expanding it to a true marker on both sides of the QTL.

3.3 Results

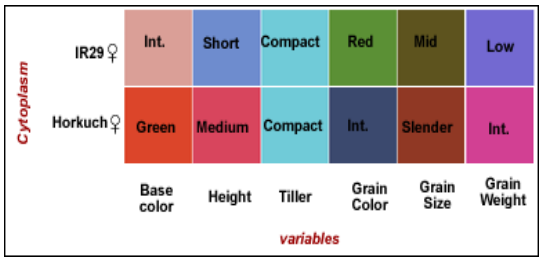
3.3.1 F_2 qualitative trait evaluation

In F_2 population the segregation was observed in many traits unique to *Horkuch* or *IR29* and an initial assessment of *Horkuch*-like and *IR29*-like parents were noted. For example *Horkuch* showed red culm, red and bold seed and *IR29* has greenish culm as shown in figure 3.3.

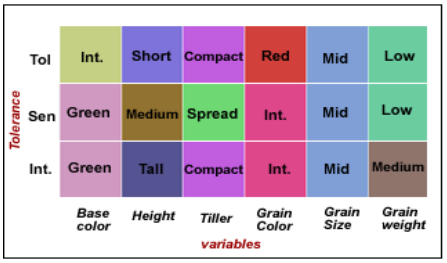
Fig 3.3 Evaluation of qualitative traits in F_2 without any treatment.

Observation of the qualitative character showed segregation among the progenies. The qualitative traits were analyzed for categorical variable correlation analysis using melt function of R. It was interesting to note that, though the characters were spread among all the progenies

in both populations, in the *Horkuch*♀ population, green-base color, medium height, intermediate grain color, medium grain size, low weight was prominent and in *IR29*♀ population intermediate base color, short height, red grain color, slender size, intermediate weight was prominent among progenies. In both populations compact tiller and tolerant phenotypes were dominant (Fig 3.4 A). Later when the F_3 plants derived from these F_2 progenies were screened under stress they were assigned a tolerance score (SES). Combining the F_2 qualitative trait observation and F_3 salt tolerance score it was noted that, in tolerant phenotypes mostly intermediate base color, short height, compact tiller, red grain color, low thousand grain weight (as only this yield component was measured) was observed, in sensitive varieties green base, medium height, spread tiller, intermediate grain color, low grain weight was observed and in intermediate tolerant varieties, green base, tall height, compact tiller, intermediate grain color, intermediate grain weight was observed. Midsize grain was found in most progenies irrespective of their tolerance character (Fig 3.4 B). These observations can help in plant selection with respect to plant architecture.



(A)



(B)

Fig 3.4 Correlation among the traits in F_2 population based on cytoplasm (A) and tolerance (B). Only the tolerance trait in 3.4 (B) was measured from salt stressed F_3 population.

3.3.2 F_3 population screening under stress

3.3.2.1 Distribution of phenotypic data

For the seedling stage, 14 traits were measured under salinity treatment which were related to survivability, photosynthesis ability and mineral elements, e.g., Standard Evaluation System (SES) score, Total Chlorophyll Content (Tchlr), Total Sodium (TNa), Total Potassium (TK), etc. (Table 3.2). Root Relative Water Content (RRWC), SES, Root Dry Weight (RDW), Tchlr, Total Content of Chlorophyll, Chlorophyll A by B (ChlrAbyB), Stomatal Conductance (SC), TNa, TK and Potassium by Sodium (KbyNa) were found to differ significantly in the two crosses for seedling stage treatment (Fig 3.5).

For the reproductive stage, 37 traits were measured. Focus was given to yield-related parameters, e.g., Effective Tiller Number (ET), Filled Grain Weight (FGW), Total Grain Number (TGN), Flag Leaf Weight (FLW), etc. under salt stress (Table 3.3). For the reproductive stage treatment, the traits, e.g., Panicle Exsertion (PE), Total Tiller Number (TT), ET, Primary Branching (PB), Secondary Branching (SB), TGN, Total Grain Weight (TGW), Thousand Grain Weight (THGW), Total Sodium of Flag leaf (TNaF), Potassium by Sodium of Flag Leaf (TNaFByTKF), etc., were found to differ significantly among the two crosses (Fig 3.6).

Generation of extreme phenotypes in a crossing population or transgressive segregation, is a rule rather than the exception in many plants which may be due to the effect of complementary genes, over-dominance or epistasis (Rieseberg et al. 1999). In this experiment we found many traits such as: SES, SRWC, RRWC,

TChlr, TK, PD, PB, SB, TGN, TGW, etc., which clearly segregated transgressively (Table 1). Interestingly, some traits e.g. TK, TChlr, TGN, TGW, TNaF also showed significant differences in the two reciprocal cross directions with respect to the segregation pattern which is indicative of an important role of the cytoplasm on transgressive segregation (Fig 3.5, 3.6).

Table 3.2 Seedling stage phenotyping data statistics

Phenotypes	**Abbrv.**	**H^2 (%)**	**F_2 population**		**Parents**	
			Mean	**SD**	***Horkuch***	***IR29***
Shoot Relative Water Content (%)	SRWC	17.36	70.08	5.11	67.72	62.43
Root Relative Water Content ***(%)	RRWC ♦	31.97	80.88	7.98	75.42	89.39
Standard Evaluation System **(number score)	SES ♦	59.35	6.35	1.13	5.6	7.8
Shoot Length (cm)	SL ♦	72.22	38.07	4.58	41.78	21.35
Root Length (cm)	RL ♦	72.14	10.68	2.1	11.51	7.6
Shoot Dry Weight (gm)	SDW♦	60.00	0.69	0.15	0.72	0.19
Root Dry Weight (gm) ***	RDW	4.76	0.12	0.09	0.07	0.03
Total Chlorophyll ***(mg chl per gram fresh weight)	Tchlr ♦	11.50	4.97	0.91	5.31	3.38
Total content of chlorophyl A plus B***(mg chl per gram fresh weight)	A+B ♦	35.90	4.87	1.01	5.49	3.25
Chlorophyll A by B ***(ratio)	ChlrAbyB	13.8	2.87	0.35	3.03	2.84
Stomatal Conductance ***(mmol/(m^2·s))	SC	48.98	390.44	123.96	366.43	317.28
Total Sodium *** (mmol/g dry wt)	TNa ♦	47.09	3.12	0.76	3.08	6.23

Total Potassium *** (mmol/g dry wt)	TK	67.00	0.38	0.08	0.38	0.21
Sodium by Potassium (ratio) **	NabyK ♦	54.13	9.41	3.1	8.1	29.6
Potassium by Sodium (ratio) **	KbyNa ♦	54.34	0.14	0.04	0.13	0.04

*Here H^2 denotes broad sense heritability of the traits, the asterisks denotes significant correlation with the trait to cross direction (*** for P < 0.001, ** for P <0.01 and * for P <0.05). ♦ denotes significantly different between parents in p<0.05*

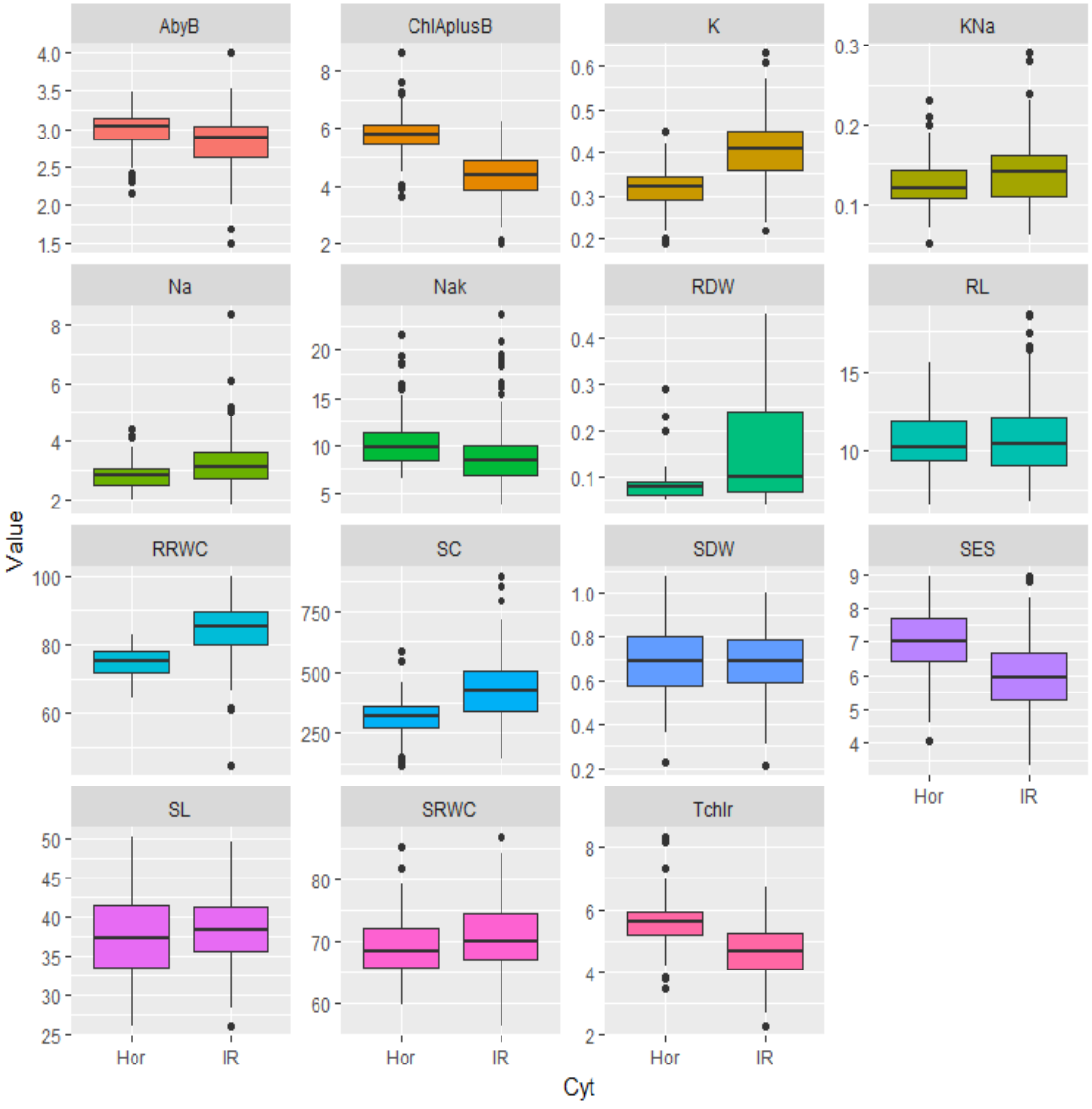


Fig 3.5 Distribution of seedling stage phenotyping data in reciprocal populations (here Hor = Horkuch♀ and IR = IR29♀)

Table 3.3 Reproductive stage phenotyping data statistics

Phenotypes	**Abbrv.**	**H^2 (%)**	**F_2 population**		**Parents**	
			Mean	**SD**	***Horkuch***	***IR29***
Plant Height (cm)	PH ♦	80.43	105.02	17.95	129.81	75.5
Panicle Exsertion (%) ***	PE	61.09	96.22	4.25	100	98.81
Total Tiller*** (number)	TT	59.84	4.78	1.23	5.07	4.36
Effective Tiller (number)	ET ♦	50.74	3.48	0.82	3.87	3.23
Percent of effective Tiller (%) ***	PT	57.98	75.43	13.81	75.25	79.76
Panicle Length (cm) ***	PL ♦	72.75	19.36	2.29	20.09	17.43
Panicle Damage (cm) ***(%)	PD	70.41	8.20	8.19	6.74	7.38
Primary Branching (number) ***	PB ♦	63.50	7.24	1.10	5.89	7.33
Secondary Branching (number) ***	SB ♦	53.05	5.96	2.06	4.09	5.39
Filled Grain Number (number) ***	FGN	77.22	155.48	105.63	156.29	149.85
Unfilled Grain Number (number) ***	UGN	53.05	60.73	26.04	32.87	29.54
Total grain number (number) ***	TGN	69.33	229.39	93.90	183.46	195.79
Filled Grain Weight (gm) ***	FGW ♦	76.42	3.13	2.17	3.91	2
Unfilled Grain Weight (gm) ***	UGW	38.65	0.25	0.10	0.15	0.13
Total Grain Weight (gm) ***	TGW ♦	78.02	3.50	2.22	4.21	2.33
Thousand Grain Weight (gm) ***	THGW ♦	52.83	20.37	3.30	25.31	15.2
Spikelet Fertility (%) ***	SF	72.07	48.86	17.08	49.4	45.08
Rachis Weight (gm) ***	RW♦	15.25	0.23	0.10	0.24	0.18
SeedLength (cm)	SeL	28.57	0.84	0.04	0.82	0.84
Seed Breadth (cm)	SeB ♦	19.35	0.29	0.01	0.3	0.27
Flag Leaf Length (cm) ***	FLL	60.94	25.19	3.56	25.52	24.99
Flag Leaf Weight (gm) ***	FLW	71.67	0.27	0.10	0.3	0.27
Days to Flower (day) ***	DF ♦	80.28	71.70	9.75	101.93	66.6
Days to Maturity (day) **	DM ♦	66.89	113.36	13.87	149.21	101.47
Total Straw Weight (gm)	TstW ♦	59.35	9.15	2.95	14.02	5.03
Harvest Index***(ratio)	HI ♦	64.91	0.28	0.15	0.23	0.32
Second Leaf Length (cm) **	SLL ♦	59.35	43.38	5.55	44.15	38.44
Second Leaf Weight (gm) ***	SLW ♦	51.22	0.42	0.11	0.45	0.34
Third Leaf Length (cm) ***	TLL ♦	67.00	57.54	7.63	58.5	44.86
Third Leaf Weight (gm) *	TLW ♦	53.70	0.48	0.14	0.59	0.34
Lower Leaf Weight (gm) *	LLW ♦	18.70	0.29	0.11	0.36	0.18

Leaf sheath Weight (gm) ***	LSW ♦	55.95	2.07	0.74	4.42	0.93
Stem Weight (gm) *	SW ♦	65.40	2.37	1.06	3.51	1.05
Straw Weight (gm)	StW ♦	51.92	3.31	1.40	4.39	1.91
Total Sodium of Flag leaf **(mmol/g dry wt)	TNaF	50.74	0.35	0.15	0.4	0.47
Total Potassium of Flag Leaf * (mmol/g dry wt)	TKF	45.36	0.29	0.24	0.16	0.23
Sodium by Potassium of Flag Leaf **(ratio)	TNaFByTKF	55.36	2.08	5.46	2.6	2.02
Potassium by Sodium of Flag Leaf **(ratio)	TKFByTNaF	39.02	1.37	1.29	0.38	0.49

*Here H^2 denotes broad sense heritability of the traits, the asterisks denotes significant correlation with the trait to cross direction (*** for P < 0.001, ** for P <0.01 and * for P <0.05). ♦ denotes significantly different between parents in p<0.05*

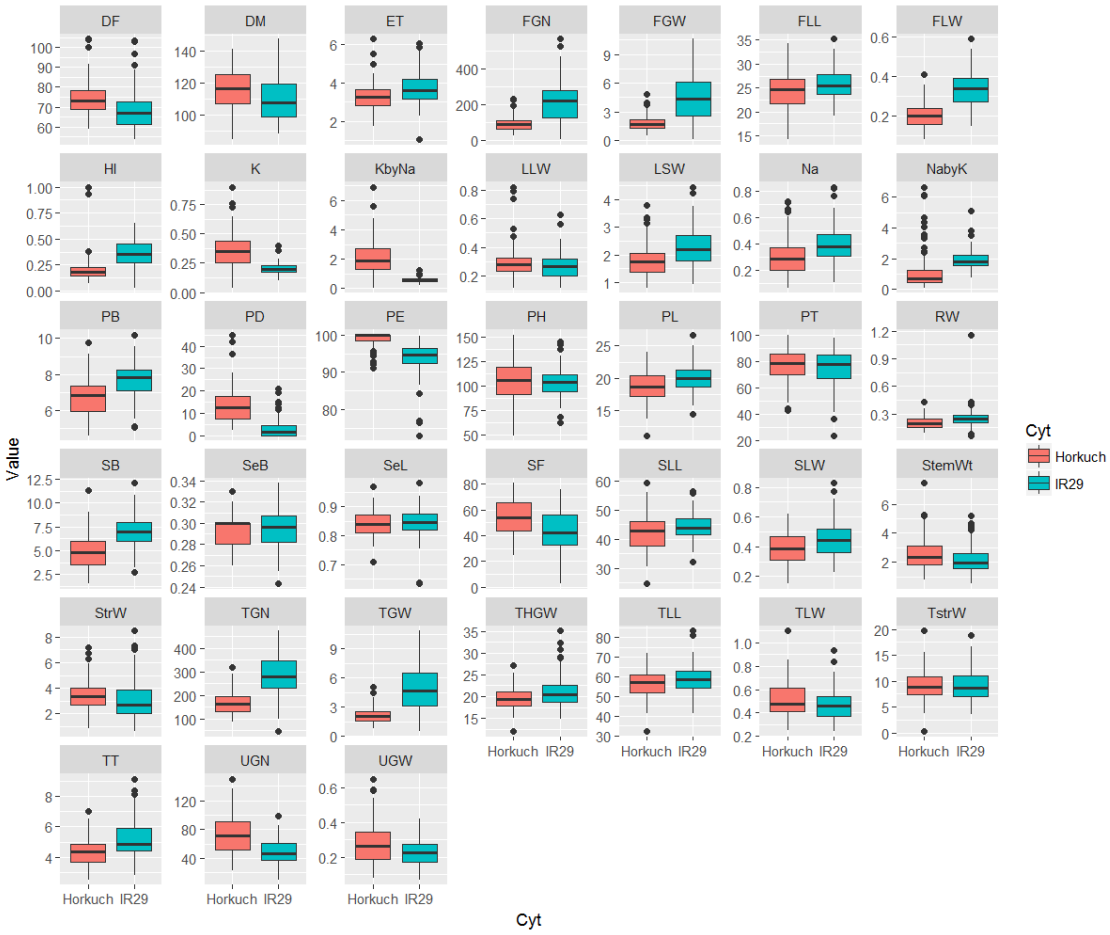


Fig 3.6 Distribution of reproductive stage phenotyping data in reciprocal populations

3.3.2.2 Correlation, heritability and principal components: exploring the trends in the traits

For the seedling stage phenotypes, significant negative correlation of SES scores with Shoot Length (SL)(-0.41), Root Length (RL) (-0.36), Shoot Dry Weight (SDW) (-0.55), Root dry weight (RDW)(-0.36), Stomatal conductance (SC) (-0.32), Total potassium (K) (-0.54) and potassium by sodium (KNa) (-0.58)were observed, whereas a significant positive correlation was found between SES scores and Na (0.63)(Fig 3.7). As expected, Tchlr and Total content of chlorophyll A plus B(A+B) were highly correlated as these two traits were the measure of Chlorophyll A and Chlorophyll B in two different ways.

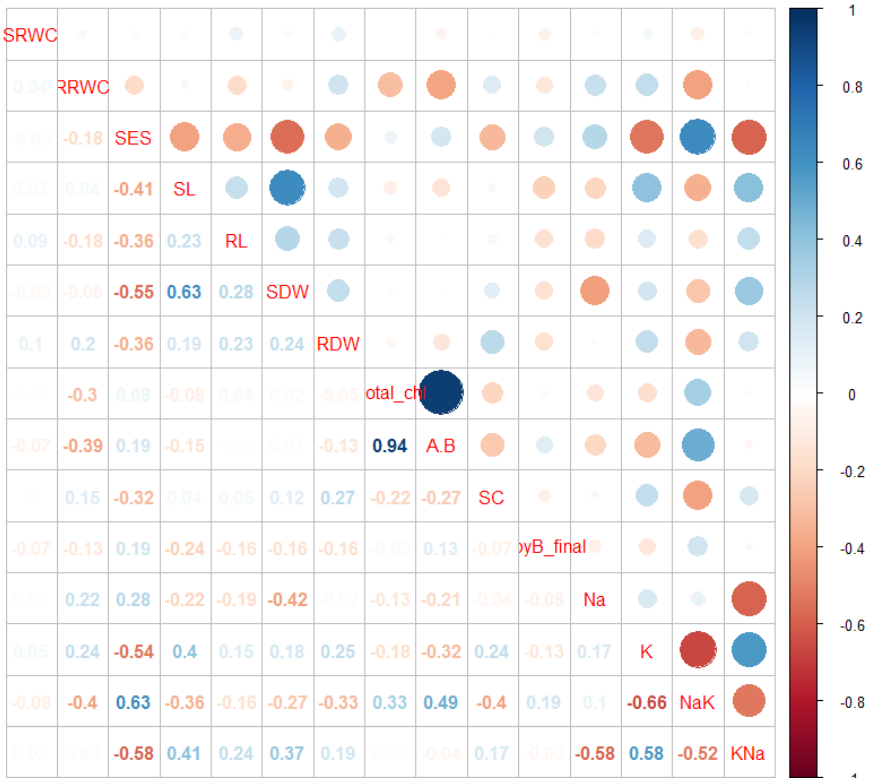


Fig 3.7 Correlation between traits in seedling stage

High broad sense heritability was observed for SL (72.2%) and RL (72.1%) followed by total potassium (67%) and SES (59.3%) in the seedling stage traits (Table 3.2). Principal component analysis clearly separated chlorophyll-related traits in one direction, opposite to RRWC. Moreover, TK, SL, RL, SDW, RDW, TK/TNa, SC clustered to one direction opposite to SES (Fig 3.8 A) as also observed in the correlation graph.

For the reproductive stage treatment, FGW was significantly and positively correlated with ET-effective tiller(0.5), PT-percentage of effective tiller(0.32), PL-panicle length (0.41),PB-primary branching (0.55) , SB –secondary branching (0.56), FGN – filled grain number(0.97),

TGN- total grain number(0.82), TGW – total grain weight (0.99), THGW – thousand grain weight (0.25), SF - spikelet fertility (0.35), FLW- flag leaf weight(0.74) etc. whereas it has significant negative correlation with PD-Panicle Damage (-0.45), DF -Days to Flowering (-0.44) and DM - Days to Maturity (-0.48) (Fig 3.9).

High heritability was observed for PH (80.4%), DF (80.2%), TGW (78%), FGN(77.2%), FGW(76.4%), PL(72.7%), SF(72%), PD(70.4%). The abbreviations are explained in table 3.2. Principal component analysis clustered the flowering and maturity related traits opposite to the harvest index and other grain weight related traits, indicating importance of early flowering and maturity on HI which is a component of yield. Leaf and other biomass related traits clustered together. (Fig 3.8B)

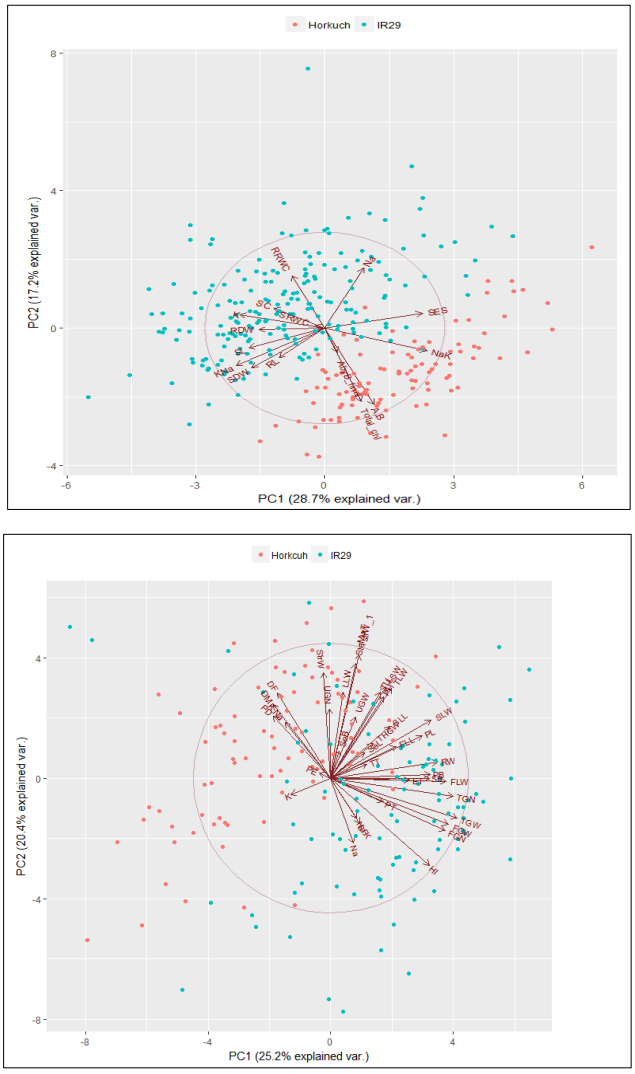


Fig 3.8 Principal component analysis among the physiological traits. (A) Seedling stage, (B) Reproductive stage)

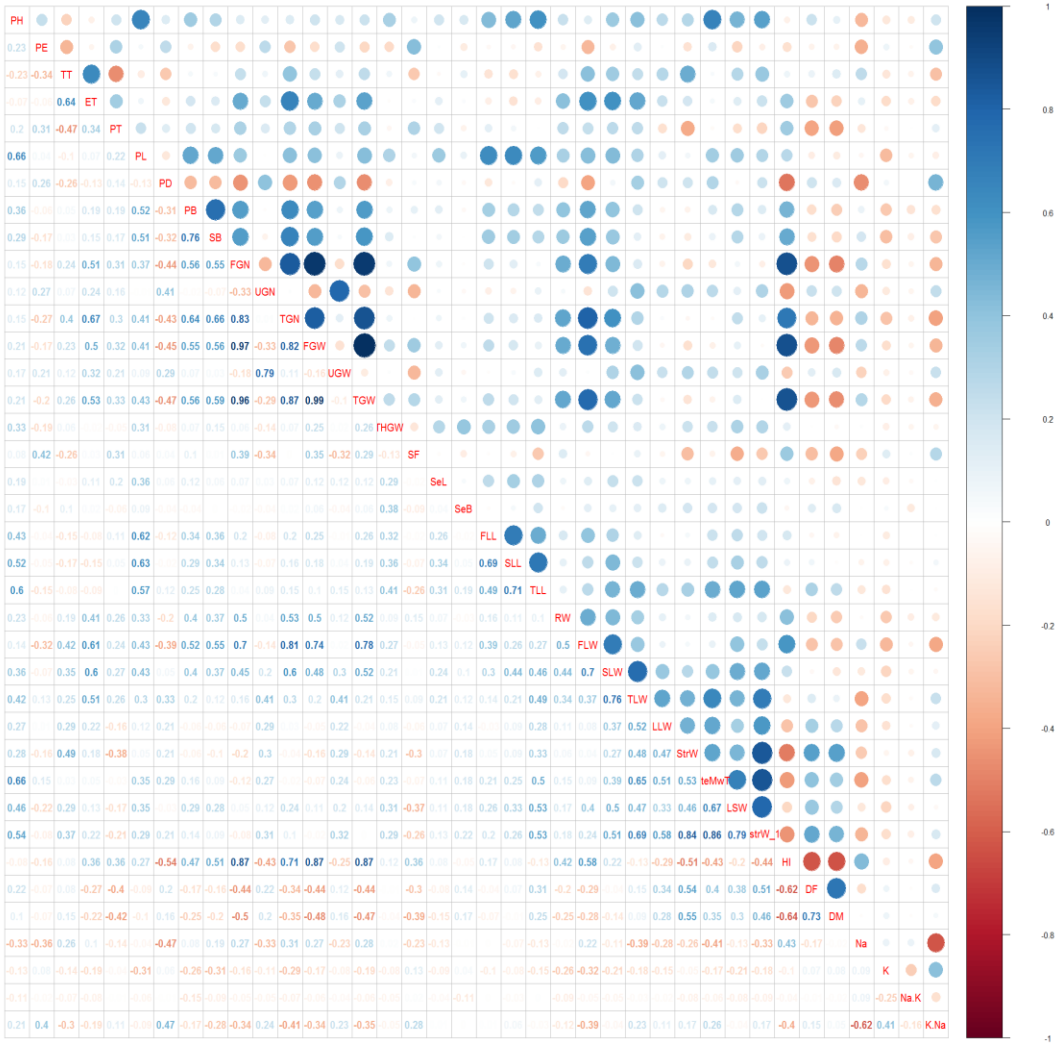


Fig 3.9 Correlation between traits in reproductive stage

3.3.3 Genetic Map

After filtration and omitting individuals with the same genotype as parents, a total of 139 indi viduals from IR29♀ and 65 from *Horkuch*♀ totaling 204 individuals were used for the genetic map construction. The map size was estimated as 2423.6 cM with 469 DArT markers distribut ed over the 12 chromosomes (Fig 3.10). Highest number of markers was found in Chromosom e 1 (60) followed by chromosome 3 (54) and chromosome 6 (51). Genotype percentage was A A:23.7,AB:50.8 and BB:25.5 (A = Horkuch, B = IR29). Maximum spacing between the mark ers were observed in chromosome 6 (30.8) and the range of average spacing in all chromosom e was 3.8 – 8.4. The average number of crossovers for the *Horkuch*♀ cross was 54.3, and rang ed from 32 to 152 with median of 49 whereas the average number of crossovers for the *IR29*♀ cross was 42.3, and ranged from 14 to 149 with median of 33.

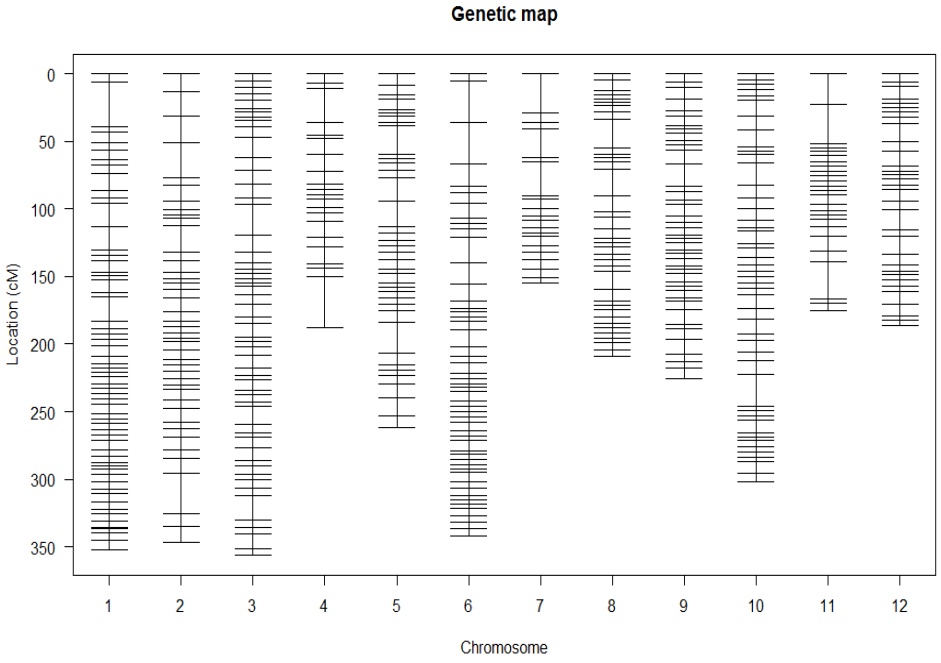


Fig 3.10 Genetic map constructed from the combined Horkuch♀ and IR29♀ cross population

3.3.4 QTL identification

As many of the traits showed cytoplasmic bias, it was used as a covariate during QTL mapping, but at a later stage of the analysis, it was used to explain the residual variance while fitting the best model. For the seedling stage treatment, QTLs were found for 5 traits. Root relative water content QTL was observed in chromosome 11 t around ~82cM, where the allele came from IR29. RWC is a good indicator of the plant status. It has been reported that the ionic balance of K^+ to Na^+ as well as total content of each ion plays a major role for seedling stage salinity tolerance in rice (Wang et al. 2012; Zheng et al. 2015). Two QTL regions have been identified

for potassium ion at chromosome 3 and 8 from Horkuch as well as for sodium ion, both in chromosome 1 and in close positions, one from Horkuch parent and the other from IR29 (Table 3.4). QTLs for shoot and root length were observed in chromosome 1, 3 and chromosome 2 respectively (Fig 3.11). Vigor under stress in early stage is important for the survivability of the plant to cope up salt stress.

Table 3.4 QTLs identified in seedling stage

Trait	QTL name: Chrom@position (cM)	pLOD	formula	TotLOD	TotVar	Donor
RRWC	Q1:11@82	0.5	y ~ covar + Q1	20.35	37.12	I
SL	Q1:1@212 Q2:3@268	9.22	y ~ covar + Q1 + Q2	16.93	32.03	H
RL	Q1:2@165.7	6.35	y ~ covar + Q1	10.33	20.99	H
K	Q1:3@215.47, Q2:8@158.03	2.9	y ~ covar + Q1 + Q2	63.91	76.7	H
Na	Q1:1@120.85, Q2:1@122	0.16	y ~ covar + Q1 + Q2	14.21	27.67	H/I

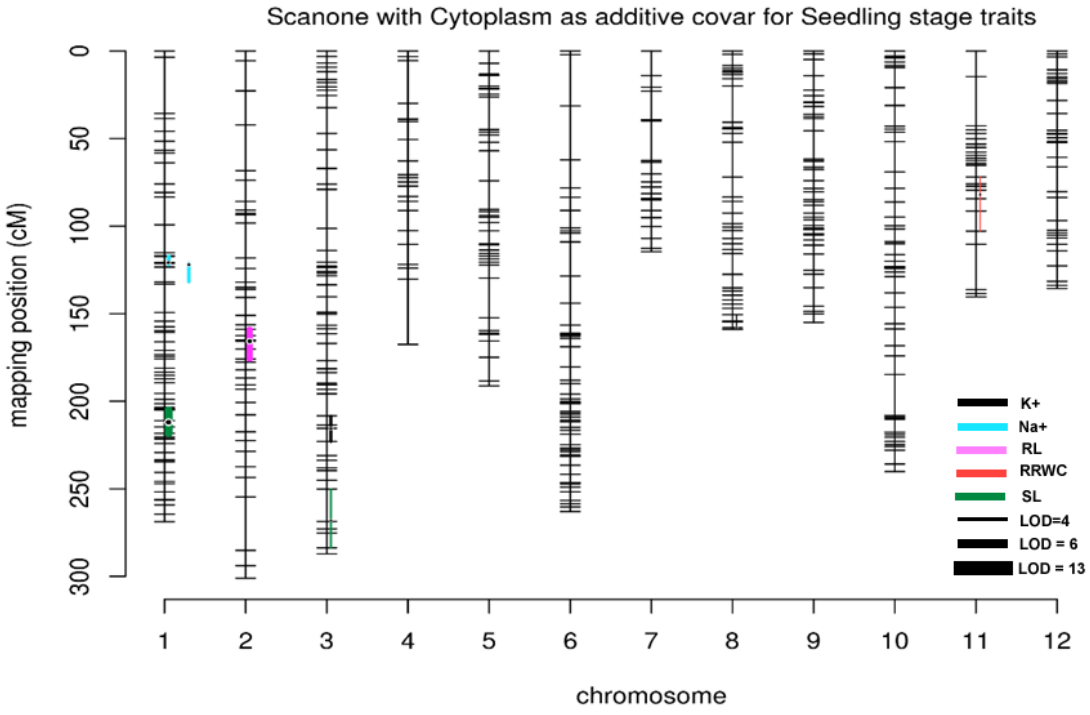


Fig 3.11(A) Map of identified QTLs of seedling stage in chromosomes

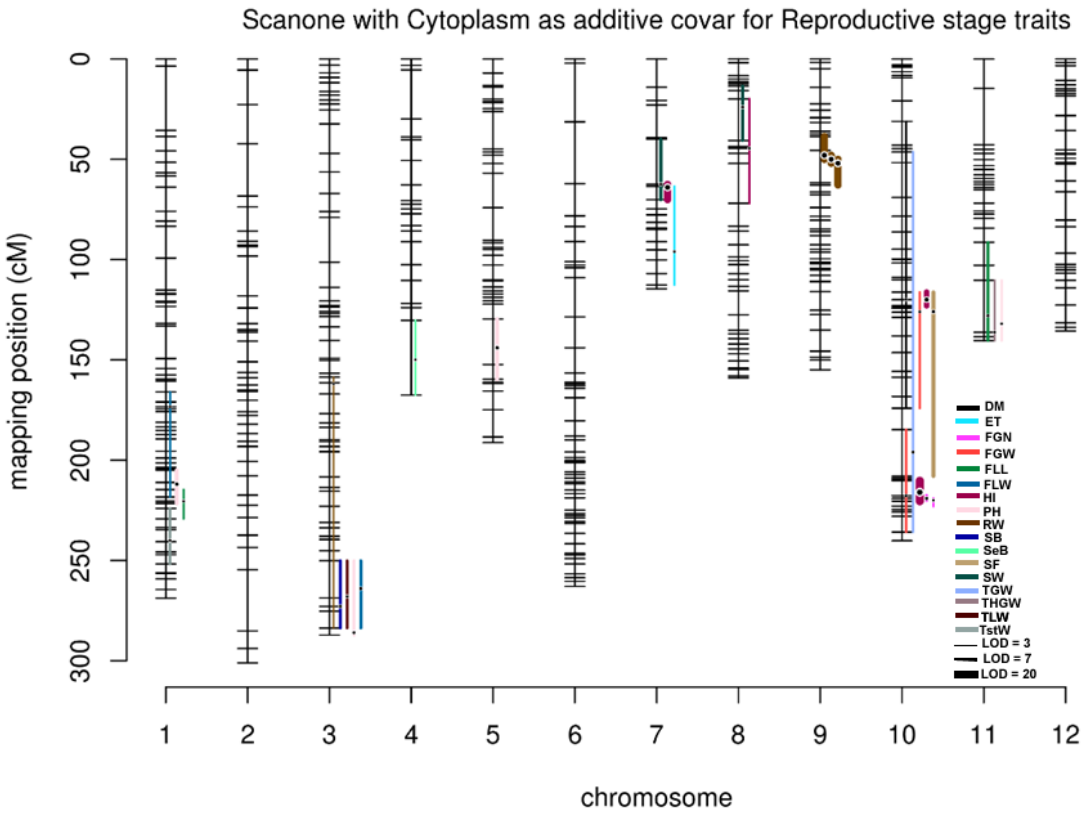


Fig 3.11(B) Map of identified QTLs of reproductive stage in chromosomes

For reproductive stage, the primary focus was yield under salt stress so the traits measured were related to yield and plant architecture under salt stress mostly. QTLs for 17 traits have been identified. In chromosome 10, multiple QTLs for multiple traits has been observed which have high correlation among themselves as observed in Fig 3.9 e.g TGW, FGN, FGW, SF, HI and DM. Except SF allele donors for these mentioned traits are IR29 (Table 3.5) . Spikelet fertility QTL under stress, where Horkuch is the donor allele, is a very significant yield component in this context. Similarly for plant height the QTL where IR29 was donor, has been considered as the short height is a desired good plant architecture.

Table 3.3 QTLs identified in reproductive stage

Trait	QTL name: Chrom@position (cM)	pLOD	formula	TotLOD	TotVar	Donor
PH	Q1:1@212, Q2:3@286, Q3:5@144, Q4:11@132	4.01	y ~ covar + Q1 + Q2 + Q3 + Q4	19.1	49.17	I
ET	Q1:7@96	0.16	y ~ covar + Q1	5.51	17.73	H
SB	Q1:3@272.93	1.31	y ~ covar + Q1	12.92	36.73	H
FGN	Q1:10@219.05, Q2:10@220	0.9	y ~ covar + Q1 + Q2	20.81	52.15	I
FGW	Q1:10@126, Q2:10@217.72	1.46	y ~ covar + Q1 + Q2	21.91	53.98	I
TGW	Q1:10@196	0.7	y ~ covar + Q1	19.14	49.23	I
THGW	Q1:11@136.26	0.07	y ~ covar + Q1	5.62	18.06	H
SF	Q1:10@126	2.33	y ~ covar + Q1	15.07	41.36	H
RW	Q1:3@162, Q2:9248, Q3:9@50, Q4:9@52	14.41	y ~ covar + Q1 + Q2 + Q3 + Q4	29.14	64.39	H/I
SeB	Q1:4@150	0.1	y ~ covar + Q1	4.82	15.69	H
FLL	Q1:1@220.42, Q2:11@128	0.07	y ~ covar + Q1 + Q2	9.49	28.55	H
FLW	Q1:1@173.37, Q2:3@264	0.72	y ~ covar + Q1 + Q2	24.78	58.43	H/HI
DM	Q1:10@120.01	0.09	y ~ covar + Q1	4.85	15.79	I
TstW	Q1:1@240	0.89	y ~ covar + Q1	15.61	42.48	H
HI	Q1:7@64, Q2:8@44.54, Q3:10@120. Q4:10@216	10.64	y ~ covar + Q1 + Q2 + Q3 + Q4 + Q1:Q4	45.65	80.15	I
TLW	Q1:3@268	1.03	y ~ covar + Q1	5.71	18.43	H
StW	Q1:7@62.47, Q2:8@24	2.34	y ~ covar + Q1 + Q2	10.36	30.91	H

From the mapped QTLs, plants have been identified which have maximum number of QTLs from both developmental stages (Table 3.6). Based on the donor allele and the desired ideal plant architecture like short height, reduced days to maturity like *IR29*, increased spikelet

fertility under stress, high sodium potassium partitioning mechanism like *Horkuch* plants are being selected for QTL validation and ultimate pyramiding. Superior performance of the selected plants are represented in Fig 3.12 (A) for seedling stage and Fig 3.12 (B) for reproductive stage.

Table 3.6 Plants containing highest number of QTLs

PlantID	QTLs	No.
I_095	RRWC, SL, RL, Na(2), PH(2), FGN(2), FGW, TGW, RW(4), FLL, FLW, TstW, HI(3)	21
I_107	RRWC, K, Na(2), PH, FGN(2), FGW(2), TGW, RW(4), FLW, DM, HI(4)	20
I_139	RRWC, SL, RL, K(2), Na(2), ET, SB, RW(4), SeB, FLL, FLW(2), TstW, TLW, StW	20
I_040	RRWC, SL(2), K(2), PH, ET, SB, THGW, RW, SeB, FLL(2), FLW(2), TstW, TLW, StW(2)	19
I_150	SL, ET, SB, FGN(2), FGW(2), TGW, THGW, RW(4), FLL, FLW, DM, HI(2), TLW	19
I_141	RRWC, RL, PH(2), FGN(2), FGW(2), TGW, THGW, RW, SeB, FLL, FLW, DM, HI(3)	18
H_027	SL, PH(2), ET, FGN(2), FGW, THGW, RW(4), FLL, FLW, HI(2), TLW	17
I_023	RRWC, SL, K, ET, FGW, THGW, RW(4), FLL(2), FLW, DM, TstW, HI, StW	17
I_176	RRWC, K, FGN(2), FGW(2), THGW, RW(3), FLL, FLW, DM, HI(3), StW,	17
I_033	SL(2), PH, SB, FGN(2), FGW, TGW, RW, FLL, FLW, TstW, HI(3), TLW	16
I_083	RRWC, SL, PH(3), FGN(2), FGW, TGW, RW(4), FLL, HI(2)	16
I_005	SL, K, PH, SB, FGN(2), FGW(2), TGW, RW, FLW, DM, HI(2), TLW	15
H_012	K(2), ET, FGN(2), FGW, THGW, RW(3), FLL, FLW, HI, StW	14
H_031	RRWC, RL, Na(2), PH, SB, FGN(2), FGN, FGW, RW(3), HI(2)	14
H_047	RRWC, K, Na(2), PH, FGN(2), FGW, THGW, RW, SeB, FLL, FLW, HI	14
I_075	RL, Na(2), FGN(2), FGW(2), TGW, RW, SeB, FLW, DM, HI(2)	14

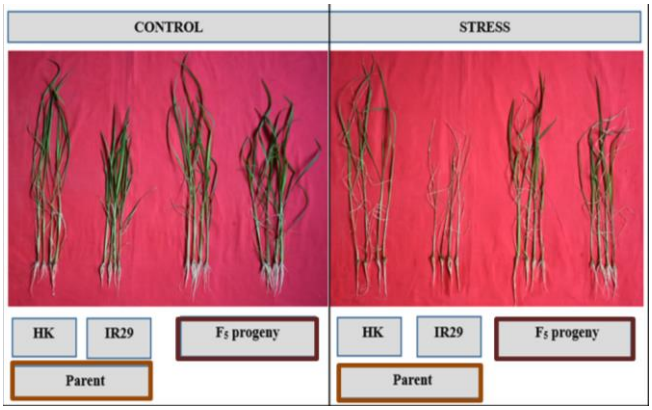


Fig 3.12 (A) Selected F_5 plants have superior performance under stress in seedling stage

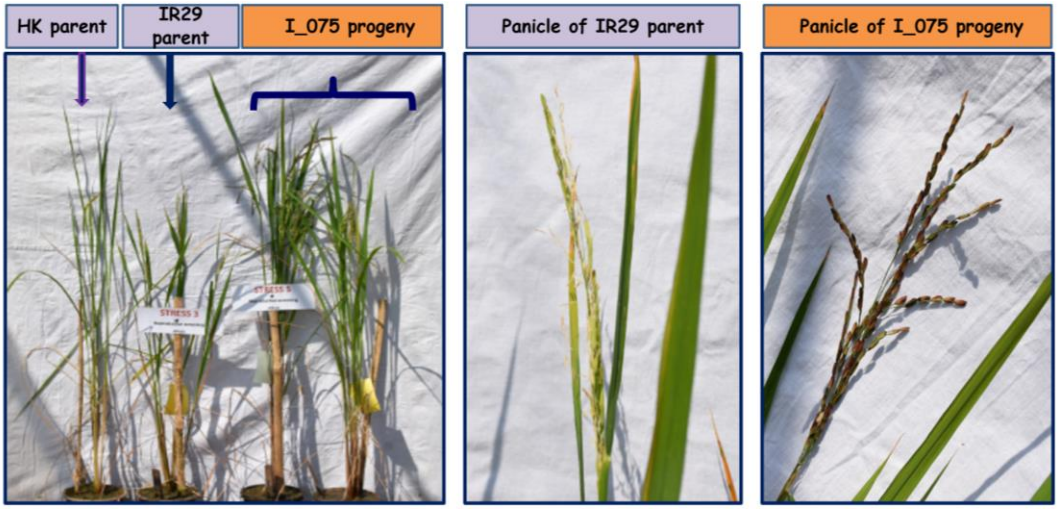


Fig 3.12 (B) Selected F_5 plants have superior performance under stress in reproductive stage

3.3.5 Discussion

A widespread measurement of rice phenotypic traits (total 51 traits) under salt stress was documented in both seedling and reproductive developmental stages followed by genotyping, linkage map construction and QTL identification. The plants were genotyped in F_2 generation and phenotyped in F_3 generation with the concept that all F_3 progeny derived from the same F_2 plant belong to the same $F_{2:3}$ family. This model reduces the time for breeding tremendously by cutting the waiting period till full homozygosity in a recombinant inbred lines but with high accuracy to map QTL (Zhang et al, 2004). Observation of segregation of different traits unique to the parents can help associate the traits under stress to the mother F_2. Both selection for F_2 and F_3 plants were totally random to avoid biasness.

We have incorporated a reciprocally crossed population into a combined analysis of genetic determinant identification. This is an uncommon protocol and novel factors might arise when constructing linkage map of progenies with parents from the two genetically distant groups. The cytosolic genetic background could also affect the ability to identify genetic loci in mapping populations (Tang et al, 2013). Cytoplasmic biasness was very obvious for some of the traits. For identification of QTLs, therefore, cytoplasm was used as a covariate. For rice, Tao et al (2004) has reported the cytoplasmic effects on yield, width of flag leaf, and low temperature tolerance on nuclear substitution lines in an alien cytoplasm. QTL pleiotropy was also observed for yield related traits in chromosome 10. We found the positive effect of *IR29* nuclear allele (cytoplasmic effect) for yield trait QTLs, on the other hand strong positive *Horkuch* nuclear allele effect for QTL models of desired traits e. g. TK .This will help define the combinations of cytoplasm and nuclear-donor materials that may favor breeding for desired traits. But the cytoplasmic effect obviously makes plant selection misleading to some extent, as often introgression of the QTL region only might not give the desired result.

Transgressive segregation was observed in the study when the combined reciprocal population was studied. Transgressive segregation refers to the formation of extreme phenotypes observed in segregated population compared to the phenotypes observed in the parents. Transgressive segregation are common in interspecific crossing programs (Marri et al. 2005, Septiningsih et al. 2003) where genetic distances between the parents are large. Due to transgressive segregation plants with much higher yield as well as some plants with higher salt tolerance at seedling stage were observed suggesting inheritance of tolerance trait from Horkuch as well as the desired yield from IR29.

One aim of the study was to identify QTLs for yield under salt stress. Though screening for individual physiological traits is more accurate than screening of complex agronomic traits, but salinity stress tolerance is a complex process involving multiple steps starting from salt sensing by root to grain formation in the panicle. What we need is the appropriate combination of these components to achieve increased yield even under salt stress and selection for the high yield under stress should be the appropriate strategy in breeding salt tolerant rice. Progenies with multiple QTLs will act as a better donor in salt tolerance breeding programs compared to their original salt tolerant donor, as they contain the superior traits from both parents.

Moreover in this study both seedling and reproductive stage QTLs have been identified and plants are being selected based on combination of the desired traits form both developmental stages. In this way the plants with multiple QTL shows better survivalist status during seedling stage when exposed to salt stress as well as produce enough yield even after being salt stressed

during panicle initiation stage. In a word, the diverged tolerance in seedling and reproductive stage can be overcome by choosing plants with multiple QTLs from both developmental stages signifying the advantage of this strategy.

In our unpublished work with another restriction site associated DNA based marker (ddRAD) construction of genetic map was much difficult due to segregation distortion of the markers as well as lack of coverage in the whole genome. DArT markers are intelligently designed, where the locations are supervised from previous diversity array based information. These RADseq based genotyping (both DArTSeq and ddRAD) are very useful in reducing the complexity of the genome, by sequencing the desired regions, ultimately generating a large number of robust SNP markers.

The phenotyping in F_3 population was performed only under salt stress without having data for non-stress condition which may raise the speculation about actual effect of stress on the QTL. For example, the question may be made whether the phenotypic effect is actually due to the salt stress or just due to the genetic differences between the progenies. The identified phenotypic effects were later supported by 1) the superior performance of selected plants (with multiple qtls) in control and stress condition in F_5 generation (details not mentioned in the thesis) 2) evidence from the eQTL study (Chapter 4) and 3) QTL identification for digital traits via automated phenotyping in F_5 (Chapter 5) where control samples were also taken for phenotyping. For example, at the F_5 seedling stage, QTLs for growth habit was found in chromosome 10 only under stress and not in absence of salt. Interestingly, this region overlaps with the yield related QTL regions identified from manual phenotyping. So these QTLs are real and helping the plants in combatting salt stress.

In a nutshell the F_3 population of the reciprocally crossed parents *Horkuch* and *IR29* was used to document a resource of phenotypes under stress at two important stages of rice under stress. The F_2 population has been genotyped and a combined linkage map has been constructed using DArTseq markers. Finally the association between the genotypes and the phenotypes aka QTLs have been identified in both stages, in seedling for traits linked to survivability under osmotic and ionic stress and in the reproductive stage for traits related to producing yield under stress.

Chapter 4

Expression analysis and eQTL mapping

4.1 Overview

The central dogma of life relies on transmission of genetic information from DNA to RNA to proteins. In- depth analysis of gene expression levels reveal important information about candidates responsible for the variation in tolerance. Hence, a selected subset of the F_3 population used for phenotyping and QTL mapping (Chapter 3) were examined for transcript abundance under salt stress. A five way model was adopted for studying developmental (seedling/reproductive), spatial (leaf/root), temporal (24hr vs 72hr stress), tolerance (tolerant vs sensitive) and cytoplasmic (Horkuch♀ vs IR29♀ population) differential expression pattern. The quantitative expression values of genes were then used as traits and correlated with the genetic QTL map obtained earlier (Chapter 3). The information from these correlation analysis is termed as eQTL mapping. eQTLs link the genetic polymorphism in DNA with expression polymorphism in the RNA. This narrows down candidate genes within a QTL locus, whose differential expression is actually responsible for tolerance or sensitivity. Therefore correlating eQTL information with phenotypic QTLs (pQTL) mapped in chapter 3 also helps determine the functional network of genes underlying the tolerance mechanism. A cost effective RNAseq method was used for expression studies with the above outcomes in mind. Specific objectives of the work described in this chapter are as follows:

1. Differential expression analysis of the five way model (described above) to narrow down list of genes which were up- or down- regulated in the population under control and salt stress.
2. Mapping expression QTLs (both cis and trans-eQTLs) by combining the transcript abundance with the polymorphic SNP marker information from the genetic map constructed in chapter 3. Cis and trans-eQTLs are two types of the eQTL which gives information on local (*SNP marker and affected genes are lying closely to each other in the genome*) and distant (*SNP marker and affected genes are located beyond a threshold distance in the genome*) eQTLs respectively.
3. Investigate the region around the major eQTL hotspots to identify important genes within a location-specific (or in other words QTLs linked to important phenotypic traits). This allows identification of crucial regions in the genome responsible for tolerance which could be narrowed down from this multidimensional study.
4. Exploring the eQTLs overlapping with the pQTLs to propose a functional network of genes responsible for the tolerance mechanism

4.2 Materials and methods

4.2.1 Plant Materials

For the seedling stage, plants were assigned as tolerant, sensitive and intermediate based on their SES (Standard Evaluation System) score obtained from the F_3 phenotypic screening under stress (Chapter 3) which is basically a measure of overall plant damage scoring by human eye. For the reproductive stage, the thousand grain weight (THGW) was used as the basis for assigning tolerant and sensitive as THGW represents one of the components of yield performance of a plant. For the reproductive stage, yield under stress was selected as the major parameter to assess the stress tolerance for future breeding targets. Detail of the phenotypic screening can be found in chapter 3. Plants were grouped as tolerant, sensitive and intermediate, taking the representatives from the extreme two tails of the distribution curve (Fig 4.1A and Fig 4.1B) and 46 plants from each cross population along with the parents were subjected to RNA sequencing.

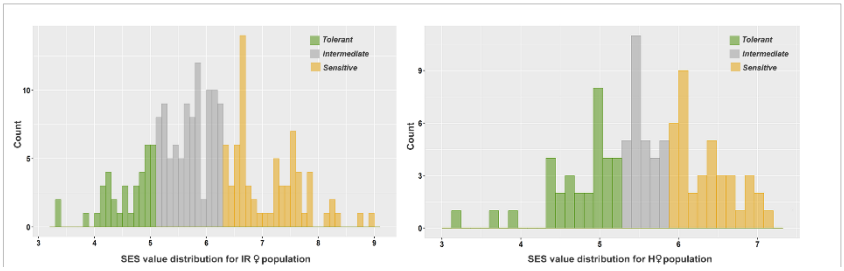


Fig 4.1 (A) Distribution of the reciprocal cross populations based on SES score,

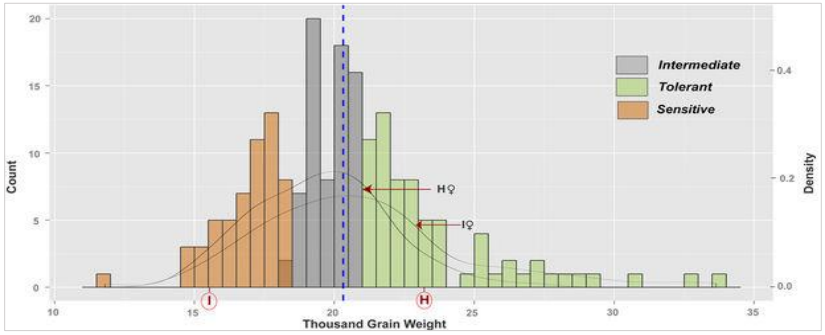


Fig 4.1 (B) Distribution of the whole population from both reciprocal cross based on thousand grain weight(THGW).

4.2.2 Plant sample collection and RNA isolation

All selected F_3 progenies were grown in hydroponics in Yoshida culture solution (Yoshida et al. 1976.), in 3 replicates per F_2 individual. Temperature and humidity recorded at that period did not vary much averaging 27°C at night and 34.9°C at daytime. For the seedling stage, selected F_3 progenies were planted randomly following incomplete block design (balanced) to avoid possible

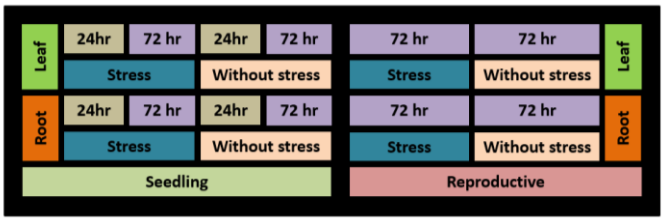


Fig 4.2: Design of RNAseq samples

environmental and positional bias, in a net house at the University of Dhaka, Bangladesh. After 15 days of germination, 15 dS/m of salt stress (NaCl) was applied gradually, in 5 dS/m increments per day. Samples (replicated pools of 3) were collected at two time points, at 24 h and 72 h after applying the maximum salt (NaCl) stress along with the samples without stress at the same time points. Leaf and root samples were collected separately. Therefore the total number of tissue samples for the seedling stage was 752 [(2 directions × 2 treatments × 2 time points ×2 tissues ×46 progenies) + 2 parents × (2 treatments × 2 time points ×2 tissues)] (Fig 4.2)

For the reproductive stage, the plants were individually grown in Yoshida culture solution in pails covered by flat bowls with sieves though which the rice was grown. As the rice plants grew taller, the culms were supported by surrounding them with pebbles placed on top of the sieves. Bamboo columns with wide wire fencing were also placed at regular intervals among the pails to prevent plants from toppling over. The pH of the Yoshida culture solution was adjusted daily and maintained at 5.5. The design was similar except only one time point i.e., 72 h was selected for the experiment due to space limits in the nethouse. Hence the number of samples were 368 [2 directions × 2 treatments ×2 tissues ×45 progenies) + 2 parents × (2 treatments ×2 tissues)]. For each plant, flag leaf and roots were collected separately.

RNA sample collection **Seedling stage plants** **Reproductive stage plants**

Fig 4.3 Seedling and reproductive stage plants and sample collection for RNAseq study

The hydroponic solutions were changed at three day intervals. During the reproductive stage, when flag leaves emerged, salt stress (15 dSm^{-1} NaCl) was applied gradually in 5 dSm^{-1} increments per day. Flag leaf and root samples were collected at 72 h after completion of the salt (NaCl) stress application. Samples (both root and flag leaf) from no stress treatment were also collected at the same time point. After tissue collection for RNAseq, plants were kept in the respective culture solution until grains were set from all panicles. Seed weights were evaluated to cross-check with the results obtained from the soil experiments.

RNA was isolated using TRIZOL following manufacturer's protocol (Invitrogen, USA) and quantified using Nanodrop® spectrophotometer ND-1000 (Thermo Fisher Scientific inc.).

RNA samples were shipped to University of Texas at Austin using RNAstable® following manufacturer's protocol (Biomatrica, USA). RNAs were treated with DNAse I (Promega, USA) at a concentration of 1 unit/µg of total RNA. Total RNA purity and degradation were checked on 1% agarose gels before proceeding further.

4.2.3 Library preparation and sequencing

To estimate transcript abundance, each RNA sample was processed for short-read sequencing using an mRNA tag-seq approach developed by Meyer et al. (2011) and extended to the Illumina platform by Des Marais et al. (2015).This protocol focuses sequencing on a smaller portion of transcripts compared to traditional whole-RNA illumina preparations by enriching the cDNA pool for 300-500bp fragments at the 3′ end. 1 µg of total RNA for each sample were used in the mRNA-Seq library construction following the protocol described by Meyer et al. (2011) at UT, Austin. This protocol requires the following steps

1. Heat fragmentation of RNA to remove biases resulting from variations in transcript length
2. Synthesis of first strand cDNA using a modified oligo-dT primer aiming at the 3′ end
3. Preparation of cDNAs and amplification with sample–specific oligonucleotide barcodes. Over amplification of the cDNA was avoided as this is a crucial step, which can distort expression ratios. Fewer cycles are preferable and optimization of the cycles were done
4. The PCR products were purified and then adaptor extension and size selection step was carried out using sample specific barcodes. From agarose gel the specific size range was cut as seen in an example figure in Fig 4.4

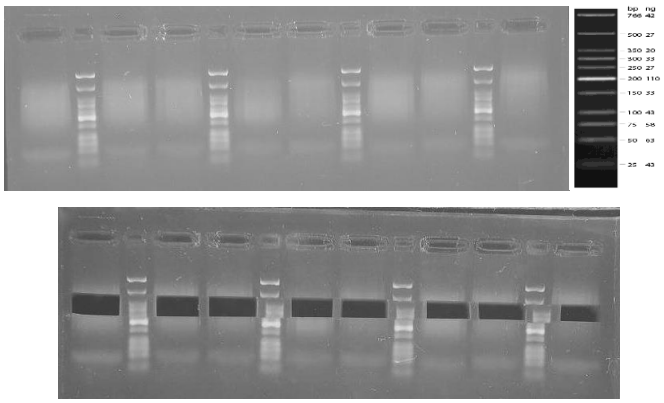


Fig 4.4 Size selection using agarose gel

The size selected samples were then quantified and pooled prior sequencing. Table 4.1 lists some of the primers used.

Table 4.1 1 List of primers used in the library preparation

Primer	Use	Sequence	Note
3ILL-30TV	cDNA synthesis and amplification	ACGTGTGCTCTTCCGATCTAATTTTTTTTTT TTTTTTTTTTTTTTTTTTTV	V = [ACG]
5-Ill-swMW	cDNA synthesis	ACCCCAUGGGGCUACACGACGCUCUUCCG AUCUNNMWGGG	RNA oligo M = [AC]W = [AU]
5ILL	cDNA amplification	CTACACGACGCTCTTCCGATCT	
TruSeq-Mpx-2n	Barcoding	AATGATACGGCGACCACCGAAAAATACAC TCTTTCCCTACACGACGCTCTTCCGAT	Extends the linker at the 5' of the cDNA
ILL -BC (BC=NNNNNN)	Barcoding	CAAGCAGAAGACGGCATACGAGATNNNNNN NGTGACTGGAGTTCAGACGTGTGCTCTTCC GAT	Barcode underlined (62bp length)
anti-T30	qPCR	AAATTAGATCGGAAGAGCACAC	
IC-P7	Final check	CAAGCAGAAGACGGCATACGA	
IC-P5	Final check	AATGATACGGCGACCACCGA	

For single lane, 32 samples were pooled for sequencing by illumina next generation sequencing platform (Hiseq) as 1 x 100 bp reads at the University of Texas, Austin's Genome Sequencing and Analysis Facility (GSAF). Each lane generated an average of ~140-190 million raw reads from both leaf and root. The average sequencing depth of leaf and root samples was ~6.6 and ~6.5 million respectively. A small number of poor libraries resulted in low read counts which were subsequently omitted for further studies.

4.2.4 Data Processing: Quality check, filtering, mapping and transcript counts

Quality of the raw reads were initially evaluated using FastQC (Andrews) and a variety of quality metrics were utilized for filtering or retaining reads. Filtering or trimming occurred when reads exhibited homopolymer stretches or polyA counts > 20 or 10% of the read length, respectively. Reads with average sequence quality >20, fewer than 5 unknown bases, and with a minimum of 30 bp in final length were retained. K-mer

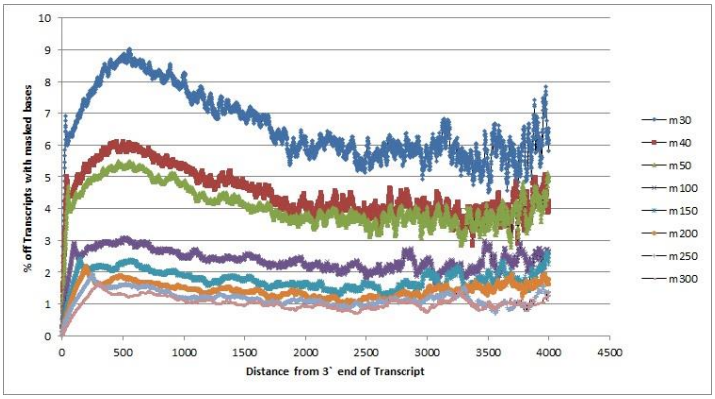


Fig 4.5: k-mer profiling with available rice transcriptome data. It shows that 90% of rice transcripts can be uniquely captured using 30bp transcript sequence (Published at Razzaque et al, 2017)

profiling with available rice transcriptomes was completed to determine the minimum cut-off length for implementation in our filter. These analyses revealed that about 90% of rice transcripts are unique at 30 bp length.

Filtered libraries were mapped against the genome of Nipponbare, a cultivar of the japonica subspecies of *Oryza sativa*, using the bowtie/1.0.0 short read aligner (Langmead 2010) fixing the maximum mismatches at 1 (value ranges from 1–3).The mapping efficiency varied from ~89–95% among the samples. Unmapped reads were likely poorly sequenced, containing many missing values (n's) and unannotated gene sequences. Reads with poor quality score were removed by using "bq" cut-off value of 10 which eliminated another 8–9% of the aligned reads that were non-uniquely mapped. Rice annotation file (*Oryza sativa)* from Phytozome V.9 (Goodstein et al. 2012) was used to generate count files from the sequenced RNA data using NGSTools (Kyle).

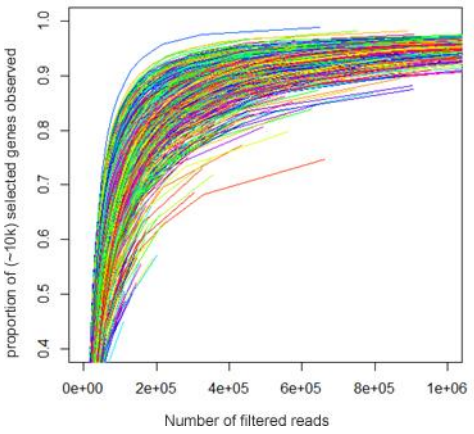


Fig 4.6: Rarefaction curve showing capturing of ~70% of the transcripts with minimum 100k reads (Published at Razzaque et al, 2017)

4.2.5 Normalization and filtering of the count data

The RNAseq count file was primarily filtered to remove with low counts (sum of counts less than 100 across samples). These filtered samples were then KDMM (Kernel Density Mean of M-component) normalized using JMP genomics 7.0 (SAS, Cary NC) and further filtered by removing rows with 40% zero counts. However, to avoid bias due to outliers present in the experiment, a rarefaction curve was plotted. This plot showed that samples with a minimum of 100k filtered reads can capture ~70% of the gene occurring in the transcript pool as seen in Fig 4.6. So samples with fewer than 100k read counts were removed. The filtering resulted in ~13500 transcripts for leaf samples and ~14900 for root samples at reproductive stage and ~11000 transcripts for leaf samples and ~7700 transcripts for root samples

4.2.6 Differential gene counts and modelling

For differential expression analysis, number of samples were reduced to tolerant and sensitive varieties, hence keeping 15 tolerant and 15 sensitive progenies from each reciprocally crossed population. Therefore, the total number of samples in seedling stage became (2 phenotypes × 2

directions × 4 treatments (2 salt + 2 control) × 15 biological replicates = 240 samples × 2 tissues = 480 (+15 extra) =495 + 2 parents (× 4 (2 salt + 2 control) × 2 tissues) = 511 and in reproductive stage the number of samples were (2 phenotypes × 2 directions × 2 treatments (salt/control) × 15 biological replicates = 120 samples × 2 tissues = 240 + 2 parents (×2 control/stress, ×2 tissues) = 248 − 6 missing samples from control tissues = 242.

A simple generalized linear mixed model was fitted on the normalized count data for each transcript using a negative binomial distribution and a log link function. The model included factors for cytoplasm (*IR29*♀ or *Horkuch*♀), treatment (salt stress or control), and phenotype (tolerant or sensitive). Pairwise interactions included treatment × cross direction and treatment × phenotype, along with a random effect for sequencing lane. Given their complexity, higher-order interactions were not considered but rather pooled in the residual of the model. For seedling stage the model was like Expression = constant + Treatment + Cross-Direction + Phenotype + Time + Treatment*Cross-Direction + Treatment*Phenotype + Treatment*Time + Cross-Direction*Phenotype + Cross-Direction*Time + Phenotype*Time + Lane + Error. The false-discovery rate (FDR) was set as 0.05 in the study using the Benjamin-Hochberg method.

4.2.7 Validation by q/RT-PCR

Validation studies have shown that read counts from this library preparation show very high correspondence with estimates of transcript abundance from q/RT-PCR. Hence, in the current study expression data was validated using qPCR with only two selected genes, instead of validating all. For the qPCR experiment and analysis, Biorad CFX96^{TM} realtime thermo cycler and CFX96 manager software were utilized using LightCycler®480 SYBR green master mix. Two technical replicates and four biological replicates were used for the Tolerant and Sensitive category across control and stressed samples. A putative proteasome subunit gene (LOC_Os03g63430) in rice was used as internal control.

Sequence of the primers designed are stated in Table 4.2

Table 4.2 Primer sequences for real time qPCR validation

Primer name	Sequence
LOC_Os05g04700_F	5′ GAACTGCATCGACATCCTCA 3′
LOC_Os05g04700_R	5′ ATGGCGTAGATGGCGTAGAT 3′
LOC_Os03g05290_F	5′ GGCCTCCTCTACTGGATCG 3′
LOC_Os03g05290_R	5′ ATGACGATCTCCAGCACCA 3′

4.2.8 Gene Ontology annotation and enrichment analysis

Differentially expressed genes as well as genes lying under eQTL region were evaluated for gene ontology enrichment analysis using AgriGO (DU et al. 2010) using hypergeometric test after Hochberg FDR correction with a significance level of $p < 0.05$. GO enrichment sets were further summarized using ReviGO (Supek et al. 2011) by removing redundant GO names. We also used the Rice Oligonucleotide Array Database for further GO association studies (Cao et al. 2012).

4.2.9 Expression QTL mapping

To map expression QTLs (eQTL) both expression polymorphism and DNA polymorphism data were needed. The normalized value of transcript abundance data was used as the expression values. Transcripts with zero were removed. However, to get a standard set of expression traits with an acceptable pValue distribution, a threshold value was adjusted which caused a reduction in the number of genes. However, different sets of analyses were conducted with the whole data sets as well as the reduced ones. A better distribution was generated using a stringent threshold. For the DNA polymorphism data, the genetic map generated using DArTseq SNP assay (mentioned in chapter 3) was used. To adjust with the number of progenies which were subjected to RNAseq, genotype information of selected progenies were considered. SNPs with minor allele frequency were filtered out. Outliers in expression values were further quantile normalized which is a technique for making two distributions identical in statistical properties. In total 469 SNP markers were used in the study.

The current analyses has been done using matrixeQTL, an R package which is significant for its fast performance compared to other available tools (Shabalin 2012). The modelLINEAR function of the package was used, which considers the effect of genotype as additive linear (the genotype is assumed to have only additive effect on expression) and the significance is tested using t-statistics. Cytoplasm information was used as a covariate during the mapping process due to observed cytoplasmic effect on the differential expression study. False discovery rate was estimated with Benjamini–Hochberg procedure. P value was considered as the main determinant to categorize and determine the significance of the eQTL. The maps were generated using the eMAP package of R (Wei 2009).

4.2.10 Candidate gene identification in phenotypic QTLs

For deciphering the biological relevance of the identified eQTLs in the hotspots as well as to assess whether the eQTLs fall in the same region as the phenotypic QTLs (pQTLs) identified in chapter 3, functional annotation of the selected eQTLs was carried out. The network-assisted gene prioritization was carried out using RiceNet v2, which is an improved network

prioritization server for rice genes (Lee et al. 2015). The database provides a network of 25,765 genes (70.1% of the coding genome) and 1,775,000 co-functional links. The precision of co-functional links were measured by odds ratio (OR). $OR = \frac{(\#positive\ gene\ pairs\ in\ network\ )/(\#negative\ gene\ pairs\ in\ nework)}{\#total\ positive\ gene\ pairs/\#totalnegatvegenepairs}$

4.3 Results

4.3.1 Differential expression analysis

*(Authors note: The differential expression analysis part was a joint work, part of which is published in doi: 10.1038/srep46138 (2017) titled : "**Reproductive stage physiological and transcriptional responses to salinity stress in reciprocal populations derived from tolerant (Horkuch) and susceptible (IR29) rice",** so a detailed analysis on expression profile is omitted and only salient points are presented)*

4.3.1.1 Categorization of plants for tolerance

To study the variation in gene expression from the physiological point of view the combined populations containing both Horkuch♀ and IR29♀, were categorized into tolerant and sensitive classes. For seedling stage the parameter for categorization was the SES (Standard Evaluation System) score and for reproductive stage samples it was THGW (THousand Grain Weight). Since SES is a reverse value i.e. the higher the SES score the more sensitive the plant is, the correlations appeared negative. The correlation coefficients of SES with other seedling traits and correlation of THGW with other yield related trait supports the categorization based on this traits as shown in Fig 4.7.

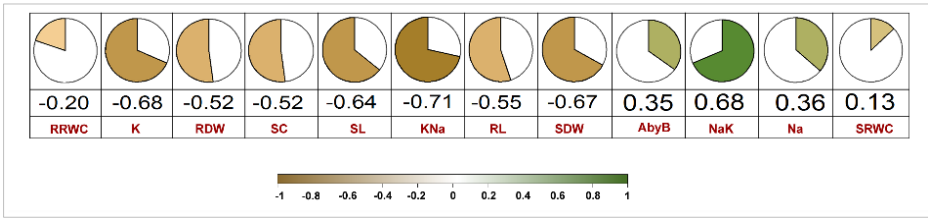


Fig 4.7(A) Categorization basis for tolerance and sensitive groups. Correlation of SES score with other seedling traits

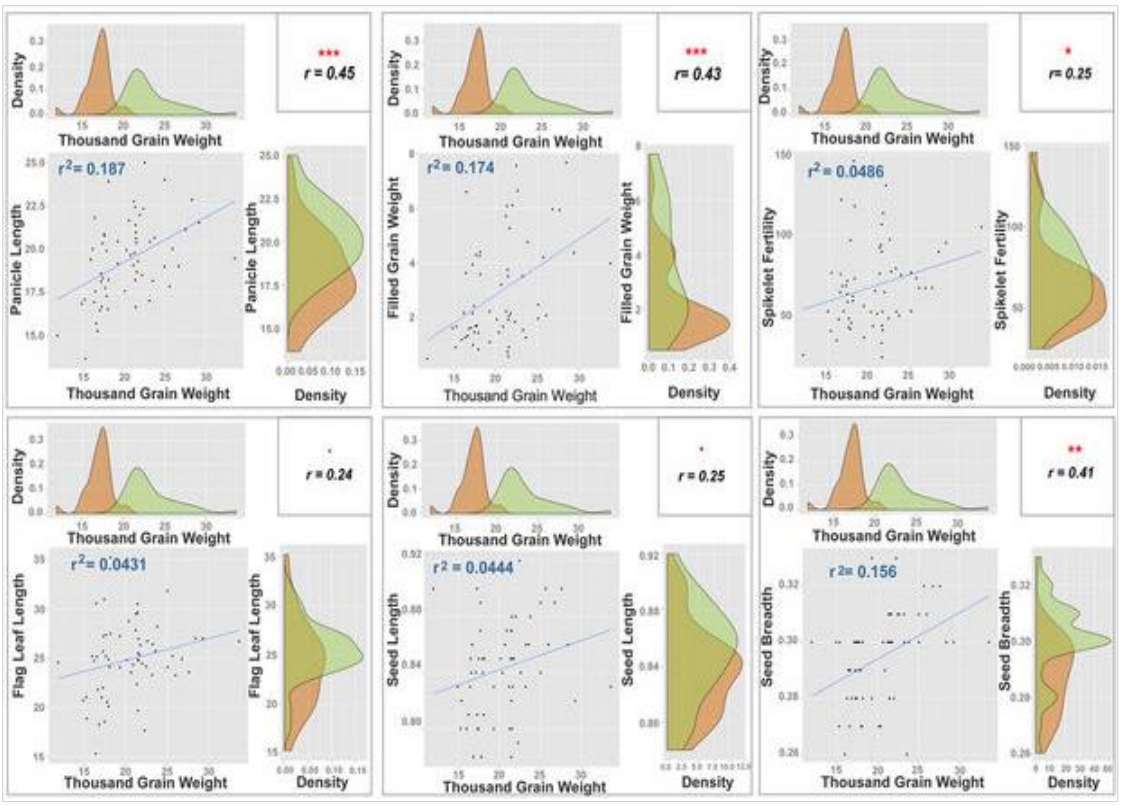


Fig 4.7(B) Categorization basis for tolerance and sensitive groups. Correlation of THGW with other yield related traits at reproductive stage

4.3.1.2 General pattern of transcript counts

Simple correlation among the experimental factors revealed that samples from leaf and root were poorly correlated in both seedling (0.45-0.61) and reproductive stage (0.54 - 0.6), compared to those within same tissue types (0.9 -0.94). So leaf and root tissue were analyzed separately. Also seedling and reproductive stage showed different pattern of differential expression distribution. Fig 4.8 shows the number of differentially expressed (DE) genes in leaf and root tissues in seedling and reproductive stage. Here number of gene up or down regulated in comparison models for cross direction, treatment, phenotype, (Cross direction ×Treatment) and (Phenotype × Treatment) are noted. It was interesting to observe that in general the reproductive stage showed different pattern of up and down regulation than the seedling stage with respect to the number of DE genes. For example, under stress vs control model, higher number of down-regulated genes were observed compared to up-regulated genes in seedling stage, which was opposite in the reproductive stage. The tolerance criteria was different for the two developmental stage as well as additional 24h stressed samples were included in seedling stage samples which might affect the number of DE genes. Whether it's an artifact or actual pattern need further investigation. However, the profile of DE gene numbers in both stages show higher number of differentially expressed genes in leaf compared to root.

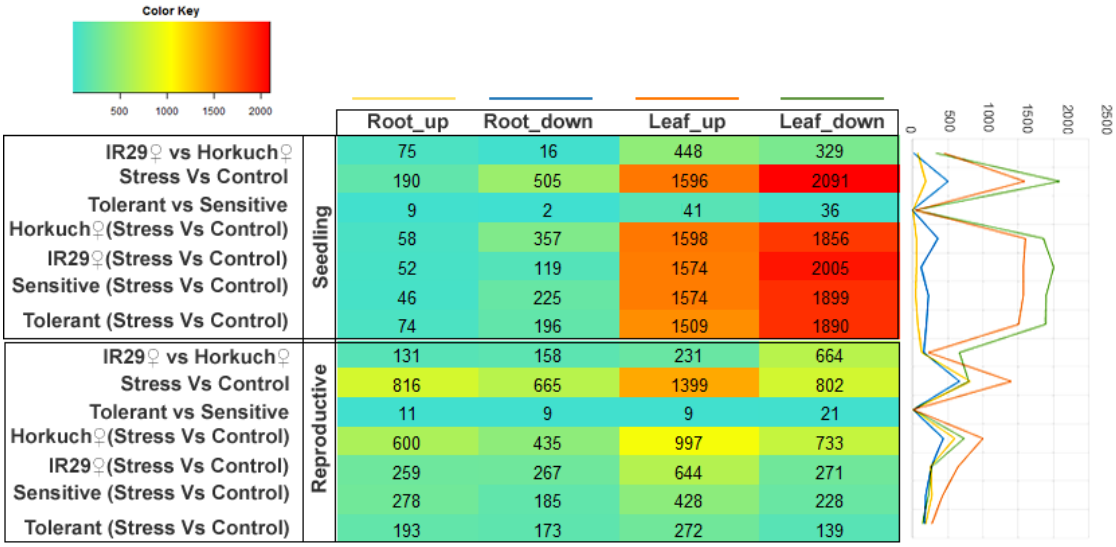


Fig 4.8 Number of differentially expressed genes in comparative models

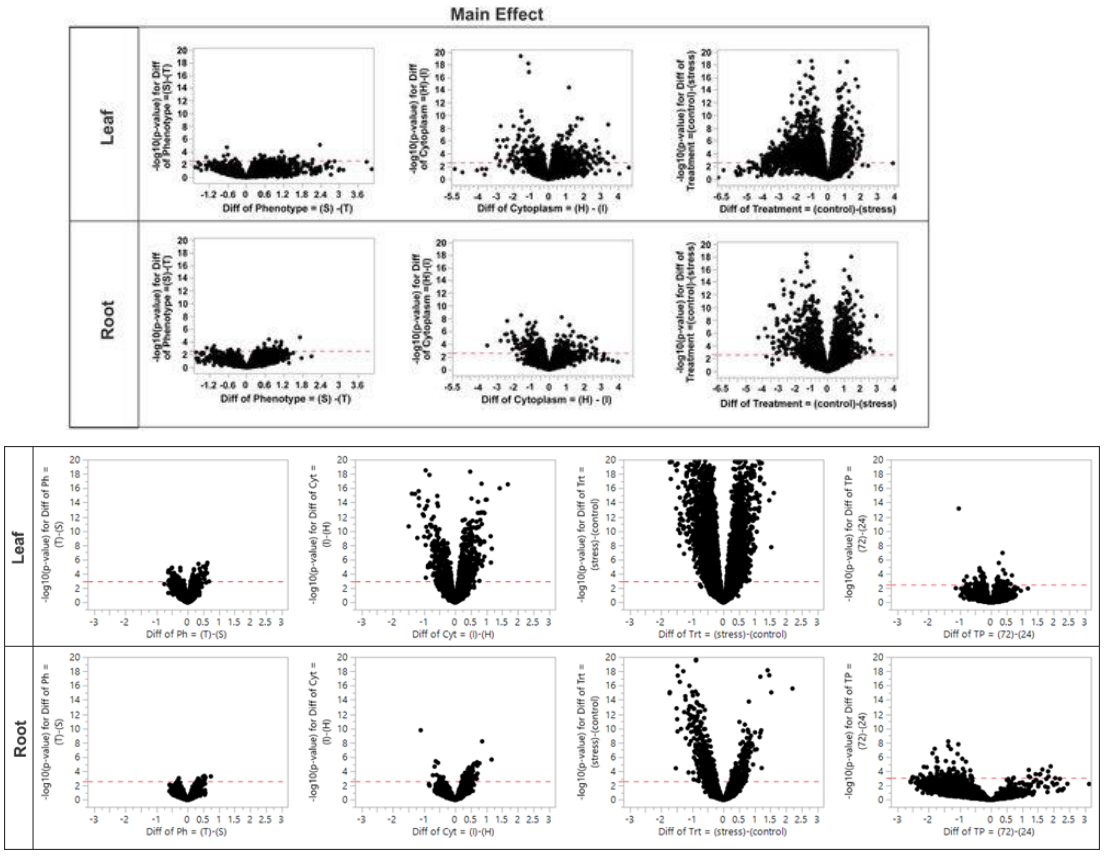


Fig 4.9 Volcano plots of significant genes in leaf and root tissue from main experimental factors. The x-axis represents the natural logarithm of fold change (Fc) and the y-axis represents log10 of the P-value of each gene. Upper panel shows seedling stage and bottom panel shows reproductive stage expression (H = Horkuch ♀, I = IR29♀ S = Sensitive, T= Tolerant, Trt = Traeatment, Tip = time point

4.3.1.3 Effect of 24h an 72h stress on gene expression variation

The goal for taking stressed samples in two different hour i.e. 24h and 72 h was to track down the different set of genes responsible in relatively early and late response to stress in seedling stage. Expression variation between 24 h and 72 h was strikingly lower (~10×) in leaf tissues (27) compared to root (258). In seedling stage also an interesting pattern of transcript abundance was noted between 24 h and 72 h stress. In leaf tissue of tolerant plants an increase in abundance was observed after 72h stress whereas in sensitive plants a decrease was observed (Fig 4.10). For root tissue no such pattern was detected. When the comparison was done between tolerant and sensitive plants, it was found that the tolerant plants show increased expression level of genes associated with ion transport, water deprivation, salt stress, sugar transport after 24 hour of stress whereas the sensitive plants showed increased expression of protein amino acid phosphorylation related genes under stress. Supporting the observation of differential expression of 10× higher genes in root after 72 h compared to leaf, more genes were found to be differentially regulated under 72 h stress compared to 24 h. Tolerant plants showed increased expression of GTPase mediated signal transduction, cell wall biogenesis and cellulose synthase related genes. This indicates during initial sensing of salt a tolerant and a sensitive plant does not vary much, but after sensing salt for 3 days they differ in signaling system from root. After 24 h the osmotic adjustment and ionic homeostasis-related mechanism are already enhanced in tolerant varieties and after 72h of stress protein amino acid phosphorylation, metabolic processes and translation related genes show higher differential expression than the sensitive plants which are mainly enriched in nucleotide-sugar transport, ribosome biogenesis related pathways, rRNA processing, nonsense mediated decay etc. after 72h stress. Even when the stress application was not considered, tolerant and sensitive plants showed expression variation in different sets of genes indicating their built-in differences in genomic regulation.

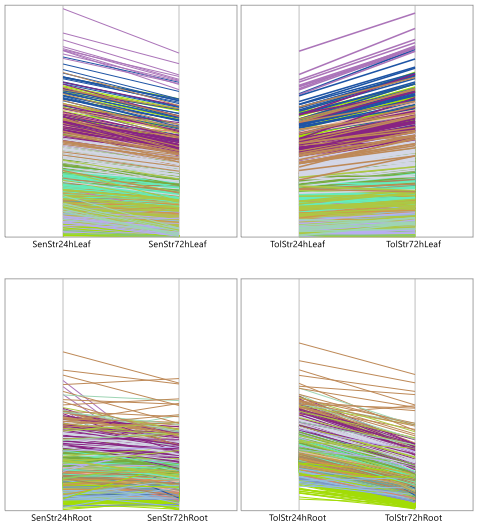


Fig 4.10 Differential pattern of expression in two different time points in seedling

4.3.1.4 Comparison of DEGs under stress between *Horkuch*♀ and *IR29*♀ population

Both populations contributed to tolerant and sensitive progenies, so studying the cross direction × treatment effect would allow us to obtain a global picture of gene expression variations. Hence, under salt stress the two cross direction populations were compared and more

DEgenes from the *Horkuch*♀ population than the *IR29*♀ population were detected, in both leaf and root at reproductive stage. The scenario was not quite same in the seedling stage which showed approximately similar number of DEGs. Both *Horkuch*♀ and *IR29*♀ backgrounds had different functional enrichments under salinity stress even though they showed some similar gene expression patterns under stress at the seedling stage, the number of photosynthesis related genes from leaf tissues were higher in the *Horkuch*♀ population than the *IR29*♀ population as also observed when cytoplasm and stress interaction was considered. On the other hand, *IR29*♀ progenies showed higher number of differential genes enriched for localization, signaling, methylation, ion transport, etc. compared to the *Horkuch*♀ population. In root tissues *IR29*♀ progenies also had some enriched functional enrichments compared to the *Horkuch*♀ ones, such as, response to water and ion transport, etc. In reproductive stage it was found that *Horkuch*♀ progenies responded to stress by adjustments via signal transduction, DNA-dependent regulation and consequential increase in transport activities. At the same time, they adjusted to stress by slowing nucleosome organization and chromatin assembly as inferred from their differential expression.

4.3.1.5 Comparison of DEGs under stress between tolerant and sensitive phenotype

Salt stress had more marked effects on the sensitive population compared to the tolerant ones at the reproductive stage. The number of up- or down-regulated DEGs in sensitive progenies was higher in both leaves and roots in comparison with the tolerant group. Interestingly, only in tolerant progenies, a putative endomembrane Na^+/H^+ co-transporter (LOC_Os05g02240) is associated with cation transport which is enhanced (MSU Release 7). Also a stress responsive protein Osr40g2 is associated with the higher activity of G-protein coupled receptor protein signaling pathway only in tolerant progenies (MSU Release 7). On the other hand, the potassium channel protein AKT1 was observed to be associated with downregulation of ion transport in tolerant roots. On the other hand, GO names for downregulated genes like signal transduction, auxin-mediated signaling, electron transport chain and DNA repair only in sensitive roots reflects that these are unable to successfully respond to the applied stress. A greater number of glutathione-S-transferases were associated with downregulated genes in the sensitive population, again indicating lack of response to stress (Fig 4.11)

At the seedling stage, cellular protein metabolic processes were down-regulated in sensitive plants but up-regulated in tolerant plants under stress. Moreover, in tolerant plants photosynthesis related GO association were found in the up-regulated genes list which is completely absent in sensitive plants. Down-regulated genes unique to sensitive plants were associated to various types of transport function including ion transport, ATP biosynthetic process, kinase activity, phosphorylation, phosphate- phosphorus metabolic process, gene

expression, biological regulation, NADH dehydrogenase activity and photosynthesis. On the other hand, in tolerant plants down-regulated genes were enriched with sodium ion transport, potassium ion transport, intra-cellular protein transport, amino acid transport as well as cellular protein metabolic processes. Upregulated genes unique to sensitive plants were enriched in cellular protein metabolic processes such as methionine metabolic process, oxidase activity, malate metabolic process, translational initiation-elongation, protein polymerization and oxidation reduction, etc. On the other hand, in tolerant plants, upregulated genes were associated with different types of transport, including ATP synthesis- coupled transport, photosynthetic electron transport, ATP biosynthetic- metabolic process, trehalose biosynthetic process, calcium ion binding, etc. Apoptosis and cell death related genes were upregulated more in sensitive plants (4 genes) compared to tolerant ones (Fig 4.11).

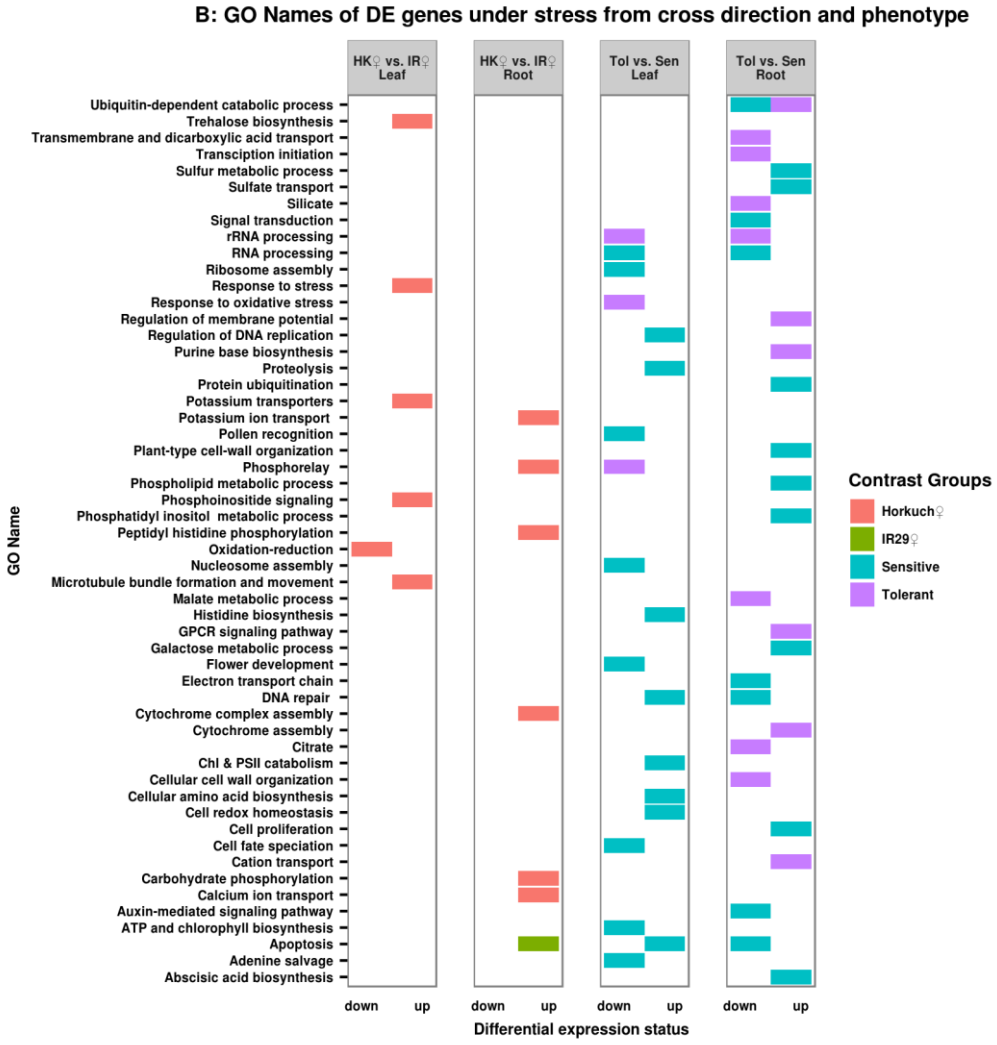


Fig 4.11 GO names of the DEGs of contrasting genotypes were compared and their pattern of expression (up or down-regulation, root or shoot) is shown

4.3.1.6 Candidate gene list narrowed down for further investigation

Studying the expression difference in the progenies from multivariable factors generated a vast resource of candidate genes. Their individual and cumulative function as network of genes helps to understand the crucial points of the salt tolerance mechanism. With this global picture of the transcriptional variation a list of potential candidates is also needed for future functional validation by overexpression or knockout studies. In table 4.3 a list of narrowed down genes have been proposed considering the major and some interacting effects as well as the significant fold changes. Some of these genes have unknown function but show significantly high level of fold change and were included as they need further functional characterization.

Table 4.3 List of potential candidate genes for further investigation

Accession	Description	Significance
LOC_Os01g11250.1	potassium channel KAT1	TOL↑24h-stress-L-Seedling
LOC_Os07g44180.1	OsRCI2-10 - Hydrophobic protein LTI6A	TOL↑24h-stress-L-Seedling TOL↑72h – R- reproductive
LOC_Os07g39280.1	transporter-related	TOL↑24h-stress-L-Seedling
LOC_Os07g43370.1	amine oxidase family protein	TOL↑24h-stress-L-Seedling
LOC_Os01g52240.1	chlorophyll A-B binding protein	TOL↑24h-stress-L-Seedling
LOC_Os04g24430.1	sucrose synthase	TOL↑24h-stress-L-Seedling
LOC_Os02g49340.1	nitrate-induced NOI protein	TOL↑24h-stress-L-Seedling
LOC_Os10g36710.1	CAMK_CAMK_like.40 - CAMK includes calcium/calmodulin dependent protein kinases	TOL↑72h-stress-R-Seedling
LOC_Os10g32980.1	CESA7 - cellulose synthase	TOL↑72h-stress-R-Seedling
LOC_Os01g45990.1	Potassium channel protein AKT1	TOL↓-R-Repro
LOC_Os07g43240.1	SKP1-like protein 1B	TOL↑-R-Repro
LOC_Os07g02250.1	Expressed protein	TOL↑-R-Seedling (fold change high)
LOC_Os12g12115.1	Expressed protein	TOL↑-L-Seedling (fold change high)

4.3.1.7 Validation of sample gene expression variation by qRT-PCR

The expression data was validated by quantitative real time PCR with two selected genes. In RNAseq experiment they showed significant variation between sensitive and tolerant progenies under stress compared to no-stress condition. Aquaporin (LOC_Os03g05290) showed higher up-regulated expression under stress in tolerant plants compared to the sensitive progenies. On the other hand OsRCI2-6 (LOC_Os05g04700) showed higher up regulated expression under stress in sensitive plants compared to the tolerant plants. In the qPCR experiment both genes showed similar trend of expression values (Fig 4.12).

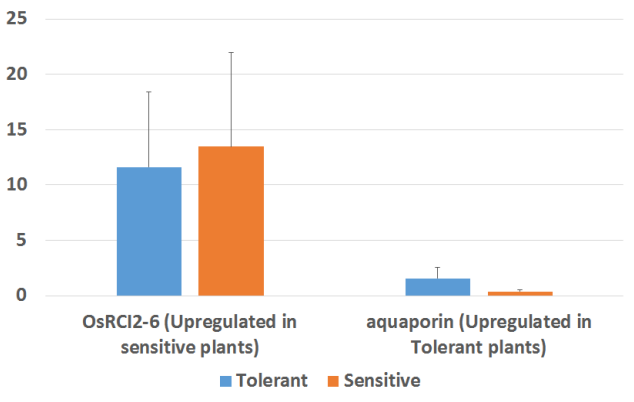


Fig 4.12 qPCR validation of genes upregulated in tolerant and sensitive plants

4.3.2 Expression QTL (eQTL) mapping

An eQTL is a chromosomal region that drives variation in gene expression patterns between individuals of a genetic mapping population Depending upon the proximity to the gene being regulated, eQTL can be classified into two groups, cis-eQTL when the physical location of an eQTL coincides with the location of the regulated gene, and trans-eQTL when an eQTL is located at a different position from the gene being regulated. This section describes eQTLs that were mapped for both seedling and reproductive stage transcripts of the subset of F_3 population by combining this with the genetic polymorphism data (polymorphic DNA used to produce the linkage map) from chapter 3.

4.3.2.1 Mapping statistics of the eQTLs

Expression QTLs in control condition refers to the regions in the genome which indicates DNA polymorphism affecting expression polymorphism irrespective of any stress or added environmental factors. Hence this is important in identifying regions polymorphic among the progenies, but not necessarily impacting any stress response. eQTLs under stress condition is the desired one in context of our experiment as it indicates the polymorphic regions that can crucially affect expression of genes under stress. In addition the eQTLs studied with the differential expression values of stress and control condition is most significant as it narrows down the candidates truly affected by stress irrespective of the genetic variation. As illustrated in Table 4.4 in reproductive stage 362 cis and 6086 trans- eQTLs were identified in leaf tissue as well as 482 cis and 7524 trans eQTLs were observed in root tissue. In the seedling stage 353

cis and 6758 trans-eQTLs were identified for 24 hour stressed leaf tissue and 290 cis and 4914 trans-eQTLs were observed in 24 hour stressed root tissue.

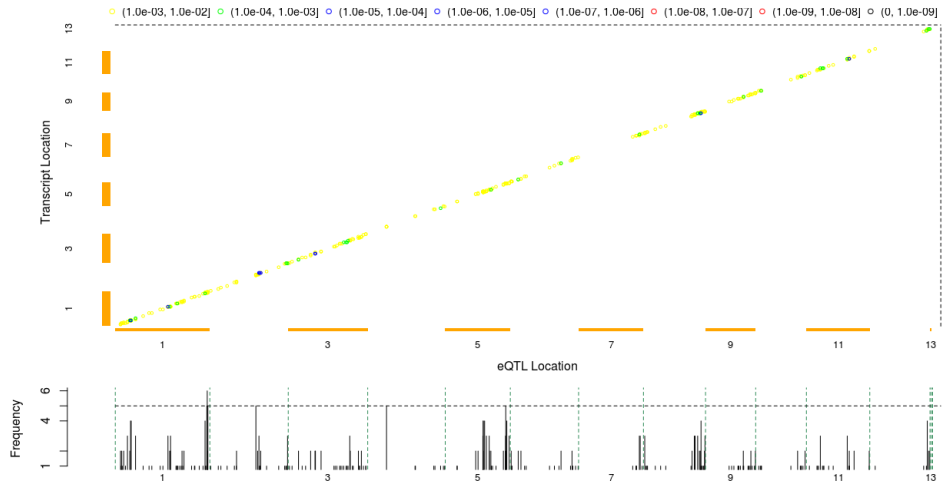
Table 4.4: Number of eQTLs identified in the genome with respect to spatial and temporal factors

		Seedling (24 hour)			**Seedling (72 hour)**			**Reproductive (72 hour)**		
		Ctrl.	**Stress**	**Diff.**	**Ctrl.**	**Stress**	**Diff.**	**Ctrl.**	**Stress**	**Diff.**
Leaf	**Cis**	465 SeLC24 Ci	338 SeLS24 Ci	353 SeLD24 Ci	242 SeLC72 Ci	333 SeLS72 Ci	265 SeLD72 Ci	317 ReLC72 Ci	349 ReLS72 Ci	362 ReLD72 Ci
	Trans	7389 SeLC24 Tr	4488 SeLS24 Tr	6758 SeLD24 Tr	4078 SeLC72 Tr	4251 SeLS72 Tr	4403 SeLD72 Tr	5775 ReLC72 Tr	5279 ReLS72 Tr	6086 ReLD72 Tr
Root	**Cis**	281 SeRC24 Ci	272 SeRS24 Ci	290 SeRD24 Ci	212 SeRC72 Ci	186 SeRS72 Ci	250 SeRD72 Ci	407 ReRC72 Ci	479 ReRS72 Ci	482 ReRD72 Ci
	Trans	4747 SeRC24 Tr	4724 SeRS24 Tr	4914 SeRD24 Tr	2879 SeRC72 Tr	2799 SeRS72 Tr	3347 SeRD72 Tr	4835 ReRC72 Tr	5538 ReRS72 Tr	7524 ReRD72 Tr

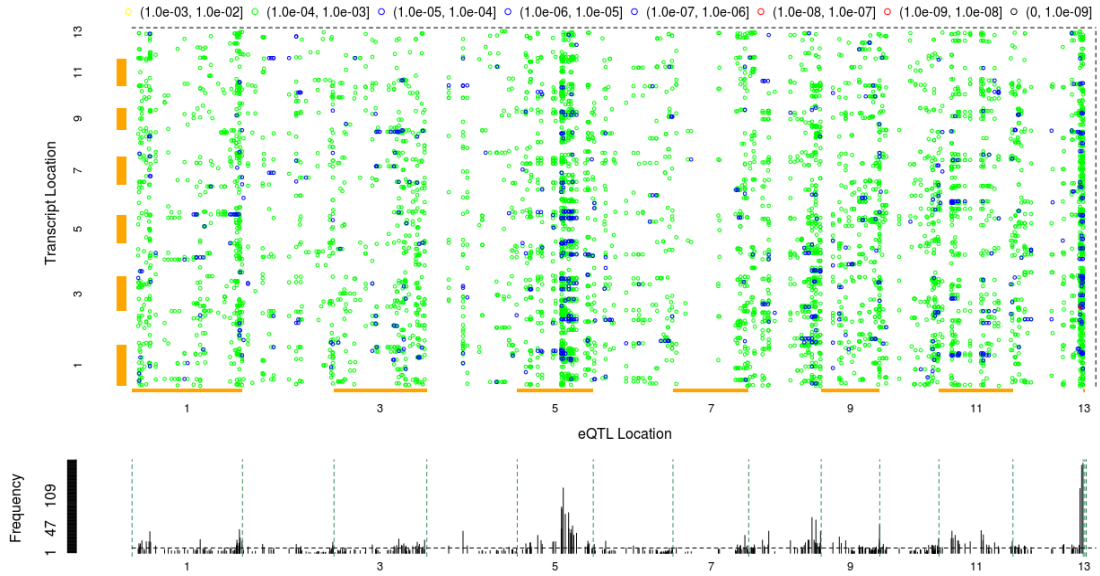
(here ctrl = Control, diff = differential i.e difference between stress and control condition. Also a code name has been assigned to refer individual eQTLs below each number where SeLC24Ci means Seedling-leaf-control-24h-cis, and so on. It follows the pattern **Se/Re**(seedling/reproductive)**L/R**(leaf/root)**C/S/D**(control/stress/Diff)**24/72**(stress hour)**Ci/Tr**(cis/trans)

Fig 4.13 shows the full eQTL map for reproductive stage cis and trans-eQTLs for leaf tissue after 72 hour (ReLD72Ci and ReLD72Tr). The difference between stress and controlled condition was captured.

The upper panel shows association of transcript location and eQTL location and the lower panel shows the hotspot regions in the chromosome with multiple eQTLs. For cis-eQTL a straight line is found as the SNP and the transcripts are collocated within 1Mb. Trans-eQTLs have been found in distant locations from the SNP. The significance values are indicated in colors.



(A)



(B)

Fig 4.13: Full eQTL map for ReLD72Ci (cis) (A) and ReLD72Tr (trans) (B) (high resolution figures in Appendix)

4.3.2.2 eQTL hotspots

eQTL hotspots can be defined as the region in the genome where frequency of cis and eQTLs are the highest. Precisely, one SNP marker in that region, can affect more than one cis eQTLs and multiple number of trans eQTLs. For cis eQTLs 5 and for trans eQTLs 10 was set as threshold to identify hotspots in the maps. Normally it has been seen that the trans eQTLs affected by one SNP marker are connected to each other. In the analysis multiple hotspots were identified in the maps (Fig 4.14). The maps show chromosomes in the x axis and frequency of eQTLs in the y axis. Hotspots are mostly found in chr 1, 3, 5, 10, 11 (Fig 4.14 (A-C)). Only the most significant ones are highlighted below and some master regulators are listed in Table 4.5.

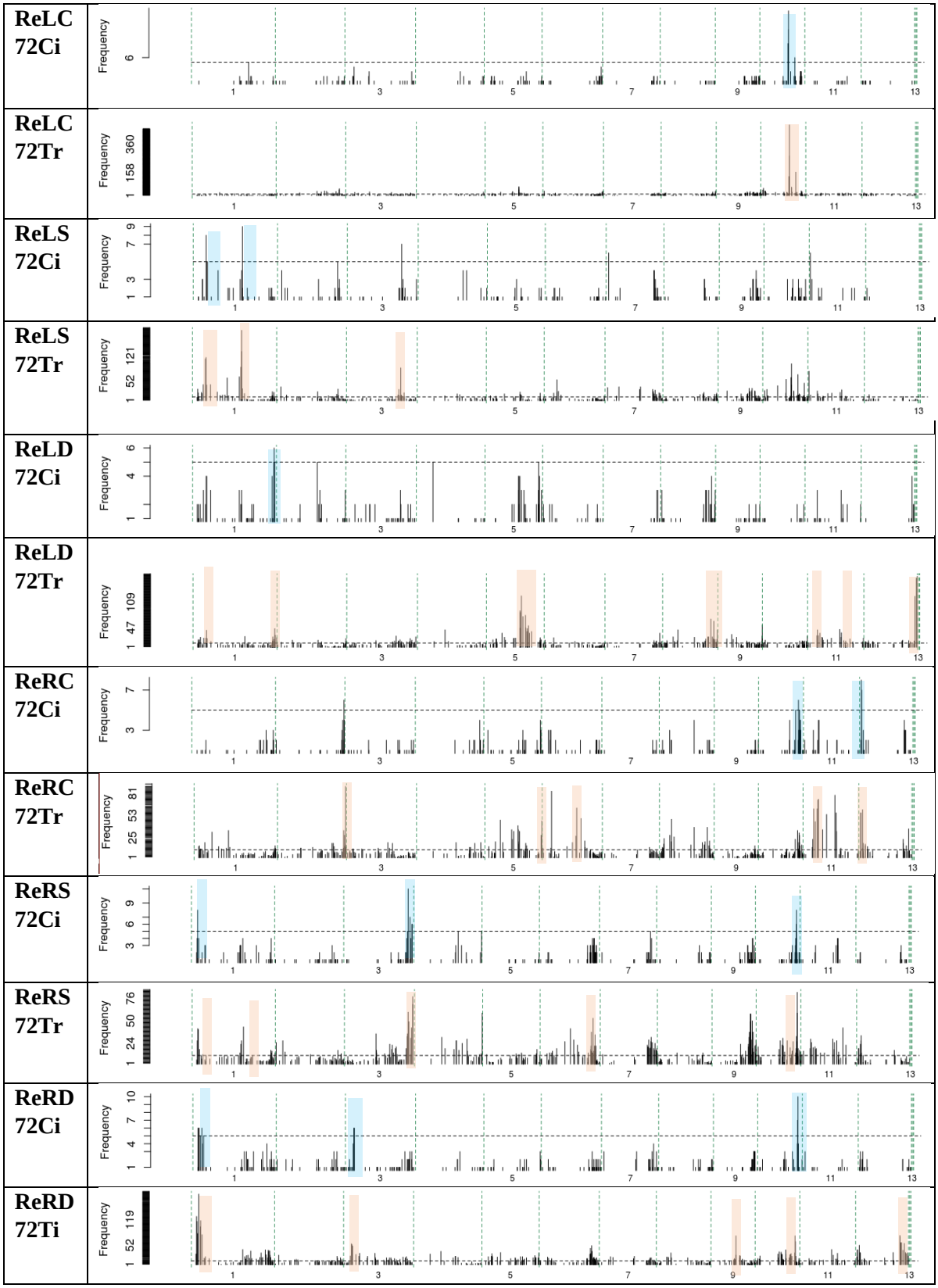


Fig 4.14 (A) Reproductive stage leaf and root eQTL hotspots

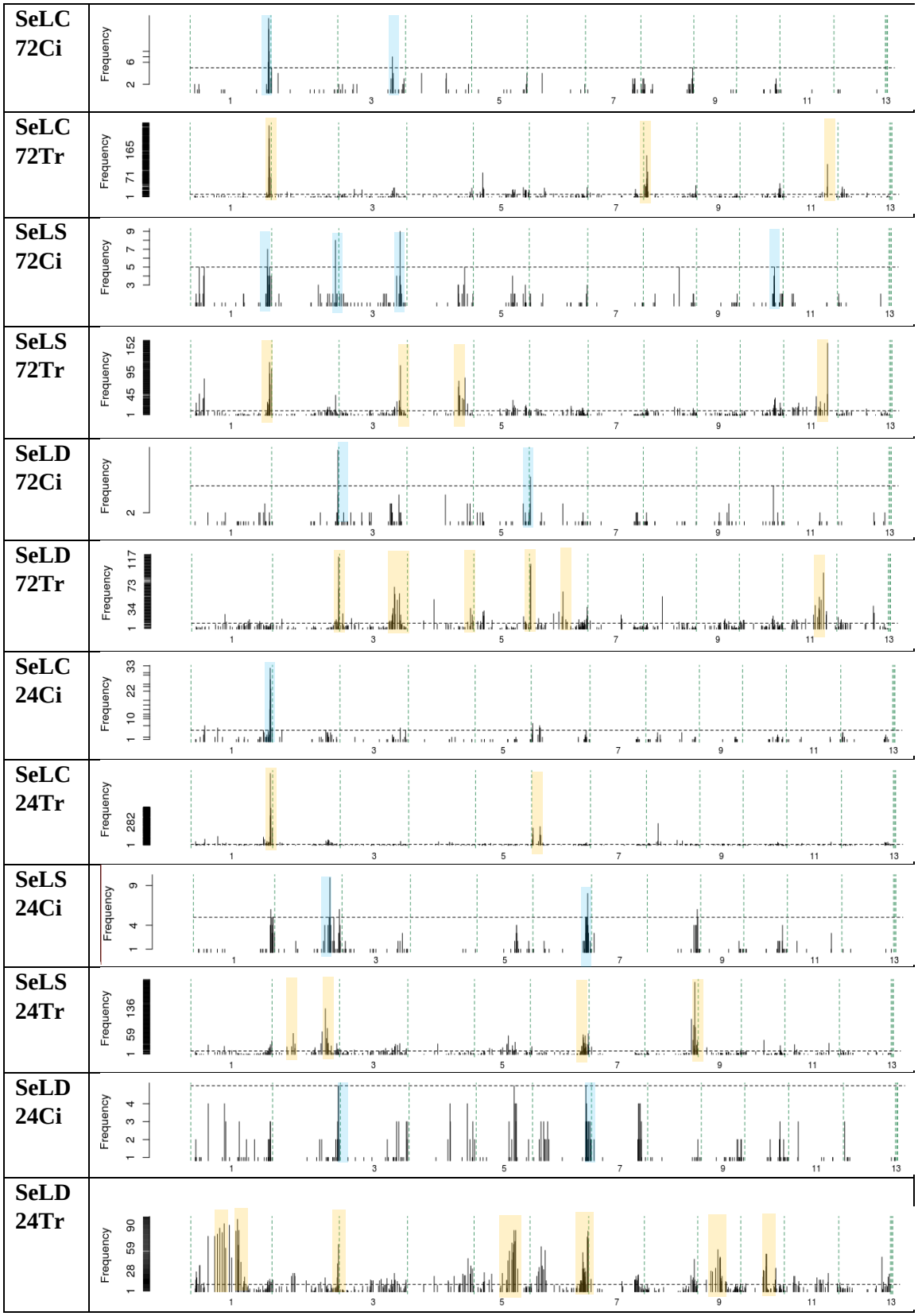


Fig 4.14(B) Seedling stage leaf eQTL hotspots in different time points

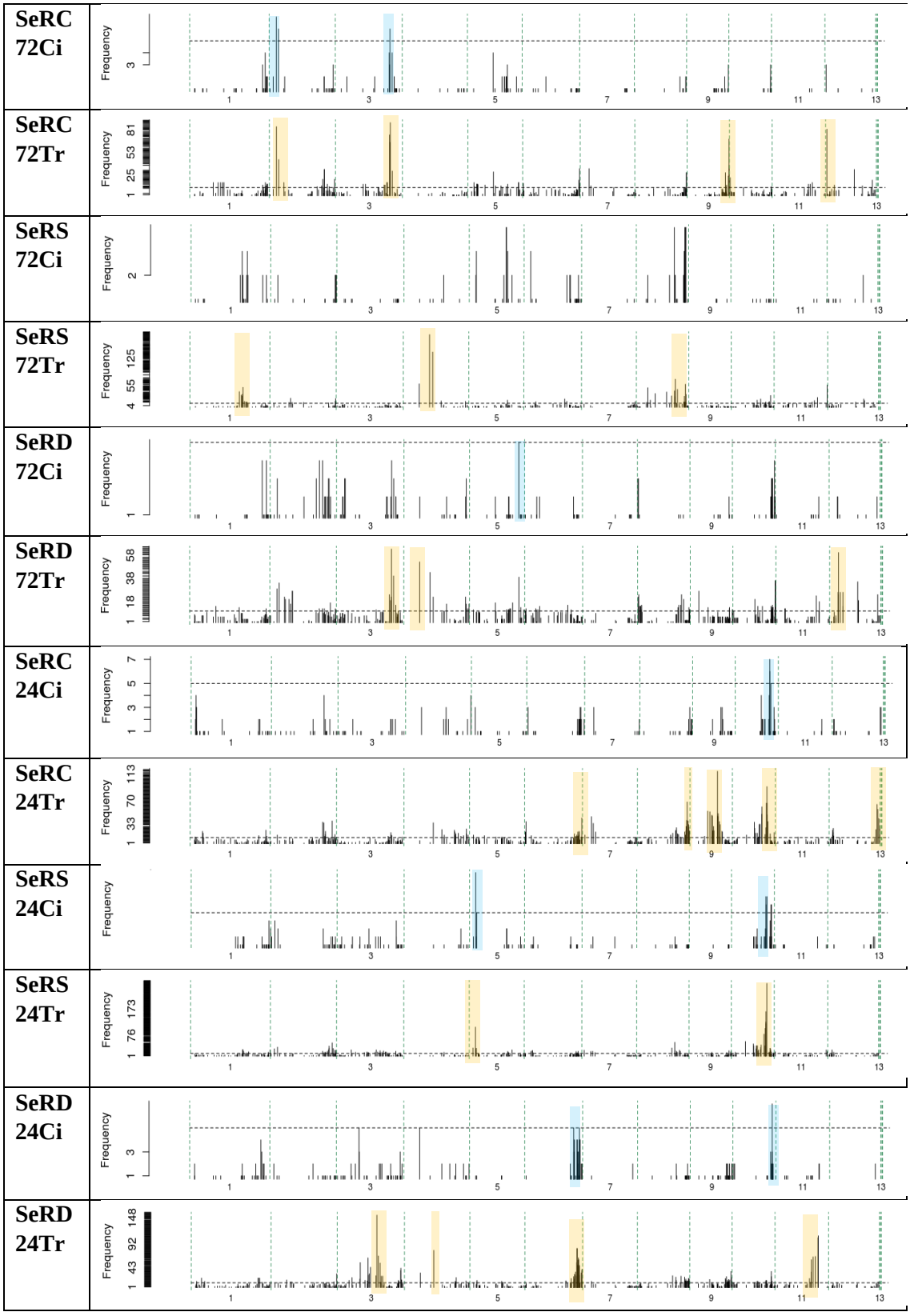


Fig 4.14 (C) Seedling stage root eQTL hotspots in different time points

4.3.2.2.1 Hotspot 1: Effect on photosynthetic machineries

For reproductive stage leaf tissue in control condition (ReLC72Tr) *(refer to Table 4.4 for the naming convention)* a significant hotspot of trans-eQTLs was found in chromosome 10 near ~14.76 Mb (Fig 4.14 A). Functional network analysis of the genes affected by that region was analyzed. About 300 guided functional network of genes could be formed consisting of 496 trans-eQTL of that hotspot region connected to a specific SNP. Also 16 cis-eQTLs (ReLC72Ci) were found in that region which increased the significance of the hotspot much more, as cis eQTLs coincide with the physical location of the SNP. Since *Horkuch's* photosynthesis capacity is already known to be better than *IR29*, this region might be of broader interest irrespective of the effect of stress. Out of the 300 networks, the highest ranking networks showed enrichment of translation and RNA biosynthetic process. Besides several networks formed related to flower development, morphogenesis, protein folding, mega gametogenesis, etc. A group of 51 genes were found connected with each other via a master regulator gene of NAD dependent epimerase/dehydratase family protein with a score of 150.73 (Table 4.5). This particular network of genes functions in pentose phosphate shunt, response to blue light, plastid organization, photosynthesis, chlorophyll biosynthetic process, light reaction, plastid translation, etc. Apart from this, several other master regulators were found that can form network of genes together and function in photosynthesis enhancement. These include ferredoxin-NADP reductase, chloroplast precursor, photosystem II reaction center PSB28 protein, chloroplast precursor, bifunctional thioredoxin reductase, thylakoid lumenal 20 kDa protein chlorophyll A- B binding protein etc. (Table 4.5). Besides several small networks of eQTLs formed related to calcium and other cation transport relating to stress and carbohydrate metabolism pathway. One such network is shown in Fig 4.15 where Calcineurin B acts as a master regulator and together the genes function in calcium sensing, potassium homeostasis leading to salinity response.

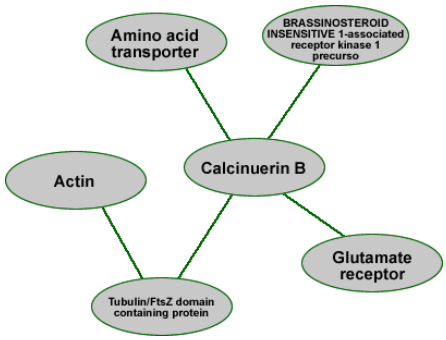


Fig 4.15 Network of calcium sensing, potassium homeostasis, salinity response

4.3.2.2.2 Hotspot 2: Network of salt stress tolerance regulator

In chromosome 1 a cluster of trans-eQTLs is found under stressed condition (ReLS72Tr). A network of 24 genes is shown in Fig 4.16 which are involved in protein folding, response to stress, defense response, response to heat, response to water deprivation, response to salt stress, flower development, heat acclimation, leaf development, protein stabilization and cellular response to calcium ion. A heat shock protein (LOC_Os09g30412) functions as the master regulator in the network. Besides this hotspot harbors functional networks involved in

proteolysis, ubiquitin-dependent protein catabolic process, oxidation reduction, tricarboxylic acid cycle, electron transport chain, glutamine biosynthetic process, nitrogen compound metabolic process etc. An interesting network of glycolysis and carbohydrate metabolism formed where hexokinase (LOC_Os05g44760) functioned as master regulator for the affected genes of this hotspot.

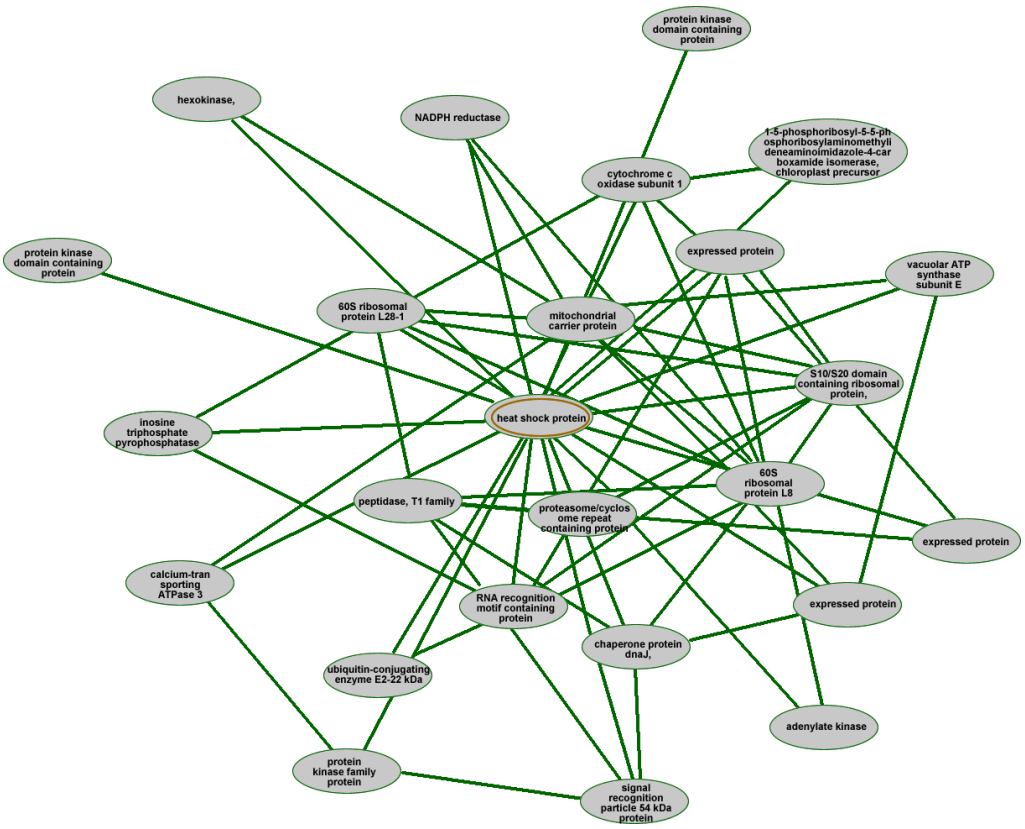


Fig 4.16: Functional network of salt tolerance regulators by trans-eQTLs found in chromosome 1 in reproductive leaf under stressed condition (Table 4.5) (high resolution figure in Appendix)

4.3.2.2.3 Hotspot 3: Network for vacuole organization and salt response

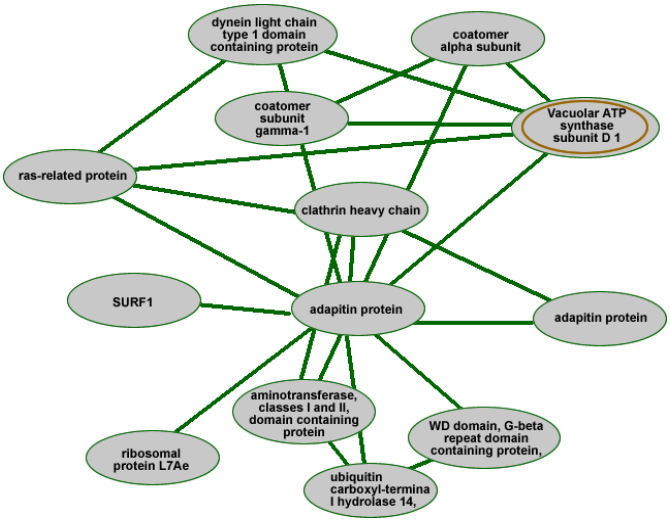


Fig 4.17 Network relating vacuole organization and salt stress response

Another hotspot was found near ~35.3Mb of chromosome 2 for the seedling stage under differential stress after 72 hour. The trans-eQTLs (SeLD72Tr) affected by this region form multiple functional networks of genes. One of them is for glucose catabolic process, protein targeting to vacuole, calcium ion transport, Golgi

organization vacuole organization, response to salt stress, ATP hydrolysis coupled proton transport under the master regulator vacuolar ATP synthase subunit D 1, LOC_Os04g55040 with 14 other genes as shown in Fig 4.17. It seems these genes might act together to cope with the salt stress at the seedling stage. Beside this network of genes, this hotspot harbors many other networks of gene that are involved in translation, salt stress response, protein folding, protein metabolic process, pollen germination, etc.

4.3.2.2.4 Miscellaneous other hotspots: Salt stress, cold stress, response to water

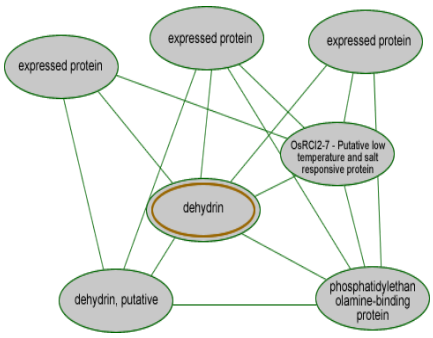


Fig 4.18 Network of gene for response to water and salt stress

In SeLS24Tr i.e in seedling stage leaf tissues under 24 hour stress among the trans eQTLs in chromosome 8 at 26.5 Mb a hotspot affects 213 trans- eQTLs. Multiple salt stress related networks formed along with translational elongation and glycolysis from this region. Interestingly this hotspot is co-located with the region for potassium content pQTL (Q2). Curiously, this hotspot was not present after 72 hour stress. Some chaperones remain in the system until the system balances or re-sets. Persistence of these proteins rather indicates that the plant will not recover from stress because these chaperones interfere with normal metabolic process. These are good until the plant stabilizes. In chromosome 1 ~3.85 Mb, 187 trans-eQTLs were found in a hotspot at reproductive stage of root tissue under stress. This hotspot also harbors multiple network of genes related to translational machinery reformation. There exists one network of 25 genes solely related to salt stress response with a master regulator nascent polypeptide-associated complex subunit alpha (LOC_Os05g31000). In chromosome 5 at ~18 Mb, a network for response to salt and water can probably form in reproductive leaf after 72 hour stress (ReLD72Tr) (Fig 4.18). In SeLD24Tr chromosome 1 harbors multiple hotspots. Since the *saltol* QTL (~9Mb -16Mb) is from chromosome 1, the relevant hotspots in that region was considered. At ~12.7 Mb multiple network of small number of genes was found related to chloroplast relocation, amine metabolic process, inflammatory response, etc. At ~9.26 Mb glycolysis, flower development, vegetative to reproductive phase transition of meristem, phyllome development, response to salt stress, etc. with a master regulator fructose-bisphosphate aldolase isozyme LOC_Os05g33380 was observed. Table 4.5 lists information on some of the trans eQTL networks.

Table 4.5 Hotspot regions harboring major networks for trans-eQTLs *(here for location X:X indicates chromosome: position, M indicate million base pair, Bold accessions can also act as master regulators)*

Hotspot region	Significant Master regulators	No	Functional annotation
10:~14.7M ReLC72Tr	NAD dependent epimerase/dehydratase family protein **(LOC_Os07g11110)** LOC_Os01g26039, LOC_Os01g32830, LOC_Os01g59090, LOC_Os01g64270, LOC_Os01g71190, LOC_Os02g02550, LOC_Os02g03010, LOC_Os02g07350, LOC_Os02g09940, LOC_Os02g35090, LOC_Os02g42700, LOC_Os02g42960, LOC_Os02g46980, LOC_Os02g51030, LOC_Os03g08570, LOC_Os03g10060, LOC_Os03g17470, LOC_Os03g20100, LOC_Os03g21560, LOC_Os03g29730, LOC_Os03g61090, LOC_Os04g33630, LOC_Os04g33970, LOC_Os04g36760, LOC_Os04g44830, LOC_Os04g45490, LOC_Os04g53230, LOC_Os04g54790, LOC_Os06g01850, LOC_Os06g06770, LOC_Os06g49160, LOC_Os07g05000, LOC_Os07g07540, LOC_Os07g28400, LOC_Os07g32560, LOC_Os07g37550, LOC_Os07g46410, LOC_Os07g46460, LOC_Os08g21700, LOC_Os08g31750, LOC_Os08g37940, LOC_Os09g04790, LOC_Os09g39390, LOC_Os09g39680, LOC_Os10g35810, LOC_Os11g24450, LOC_Os11g32500, LOC_Os12g33080, LOC_Os12g34890, LOC_Os12g43810, LOC_Os12g44370	51	pentose phosphate shunt, response to blue light, plastid organization, photosynthesis, chlorophyll biosynthetic process, light reaction, plastid translation etc
10:~14.7M ReLC72Tr	ferredoxin--NADP reductase, chloroplast precursor **(LOC_Os06g01850)** LOC_Os01g17150, LOC_Os01g32830, LOC_Os01g59090, LOC_Os01g64270, LOC_Os02g03010, LOC_Os02g35090, LOC_Os02g42700, LOC_Os02g43600, LOC_Os03g10060, LOC_Os03g13540,LOC_Os03g29730, LOC_Os03g31490, LOC_Os03g61090, LOC_Os04g33630, LOC_Os04g36760, LOC_Os04g44830, LOC_Os04g53230, LOC_Os06g01460, LOC_Os06g49160, LOC_Os07g07540, LOC_Os07g10490, LOC_Os07g11110, LOC_Os07g32560, LOC_Os07g37550, LOC_Os07g46410, LOC_Os07g46460, LOC_Os08g27010, LOC_Os08g37940, LOC_Os09g39180, LOC_Os10g30870, LOC_Os10g35810, LOC_Os11g32500, LOC_Os12g33080, LOC_Os12g34890	34	stomatal complex morphogenesis, photosynthesis, photosystem, photorespiration, plastid organization, MAPK regulated signaling, jasmonic acid signaling pathway
10:~14.7M ReLC72Tr	uncharacterized protein ycf53 (**LOC_Os11g16550)** LOC_Os01g64270, LOC_Os01g71190, LOC_Os01g72205, LOC_Os02g09940, LOC_Os02g36570, LOC_Os02g43600, LOC_Os02g49326, LOC_Os03g10060, LOC_Os03g18070, LOC_Os03g20100, LOC_Os03g21560, LOC_Os03g31300, LOC_Os03g61090, LOC_Os04g45490, LOC_Os04g55230, LOC_Os07g07540, LOC_Os07g09800, LOC_Os07g32560, LOC_Os07g37550, LOC_Os07g46460, LOC_Os12g34890	18	chloroplast-nucleus signaling pathway, photosystem II assembly, chlorophyll biosynthetic process
10:~14.7M ReLC72Tr	Calcineurin B **(LOC_Os05g45810)** LOC_Os02g54640, LOC_Os03g16010, LOC_Os03g56810, LOC_Os06g42720, LOC_Os10g36650	5	detection of calcium ion, , response to water deprivation, response to salt stress, cellular potassium ion homeostasis, hypotonic salinity response
1:~25.2M ReLS72Tr	heat shock protein **(LOC_Os09g30412)** LOC_Os01g46980, LOC_Os01g51020, LOC_Os01g56600, LOC_Os01g70140, LOC_Os02g27110, LOC_Os02g30900, LOC_Os03g03820, LOC_Os03g44620, LOC_Os03g52090, LOC_Os04g51910, LOC_Os04g59000, LOC_Os05g23720, LOC_Os05g33260, LOC_Os05g44760, LOC_Os06g04290, LOC_Os06g21470, LOC_Os07g13280, LOC_Os08g43540,	24	protein folding, response to stress, defense response, response to heat, response to water deprivation, response to salt stress, flower development, heat acclimation, leaf

	LOC_Os09g19700, LOC_Os09g33780, LOC_Os09g38620, LOC_Os10g31940, LOC_Os12g33946, LOC_Os12g38000		development, protein stabilization, cellular response to calcium ion
1:~25.2M ReLS72Tr	Hexokinase **(LOC_Os05g44760)** LOC_Os01g52790, LOC_Os03g13380, LOC_Os05g23720, LOC_Os06g36990, LOC_Os07g49320, LOC_Os09g30412	6	carbohydrate metabolic process, glycolysis, hexokinase-dependent signaling, sugar mediated signaling pathway, glucose mediated signaling pathway,
2:~35.5M SeLD72Tr	vacuolar ATP synthase subunit D1 (**LOC_Os04g55040)** LOC_Os01g15260, LOC_Os01g17190, LOC_Os02g26850, LOC_Os02g32120, LOC_Os03g12590, LOC_Os03g23950, LOC_Os03g50340, LOC_Os03g63090, LOC_Os06g33530, LOC_Os06g44374, LOC_Os07g09600, LOC_Os07g31370, LOC_Os08g33710, LOC_Os11g26910	14	glucose catabolic process, protein targeting to vacuole, calcium ion transport, Golgi organization, vacuole organization, response to salt stress, ATP hydrolysis coupled proton transport
5:~18M ReLD72Tr	Dehydrin (**LOC_Os11g26790)** LOC_Os05g03130, LOC_Os05g39250, LOC_Os10g36180, LOC_Os11g07911, LOC_Os11g26780, LOC_Os11g32890	6	response to stress, response to water
8:~26.5M SeLS24Tr	fructose-bisphosphate aldolase isozyme (**LOC_Os01g67860)** LOC_Os01g02880, LOC_Os01g25610, LOC_Os01g49290, LOC_Os02g01560, LOC_Os03g05780, LOC_Os03g22950, LOC_Os03g58050, LOC_Os05g43360, LOC_Os06g20390, LOC_Os06g36160, LOC_Os07g10660, LOC_Os08g03640	12	Glycolysis, response to salt stress, response to cadmium ion
1:~3.85M ReRS72Tr	nascent polypeptide-associated complex subunit alpha(**LOC_Os05g31000**) LOC_Os01g18050, LOC_Os01g19400, LOC_Os01g32790, LOC_Os01g38970, LOC_Os02g08370, LOC_Os02g37862, LOC_Os02g45820, LOC_Os02g48390, LOC_Os02g57540, LOC_Os03g11410, LOC_Os04g39700, LOC_Os05g40780, LOC_Os06g16290, LOC_Os06g21480, LOC_Os07g03960, LOC_Os07g33997, LOC_Os07g42450, LOC_Os07g46750, LOC_Os08g06040, LOC_Os08g41300, LOC_Os08g45010, LOC_Os11g05562, LOC_Os11g06750, LOC_Os11g37080, LOC_Os12g21798	25	response to salt stress
1:~9.26M SeLD24Tr	fructose-bisphosphate aldolase isozyme(**LOC_Os05g33380**) LOC_Os01g07870, LOC_Os02g57260, LOC_Os11g43970	3	Glycolysis, response to salt stress, response to cadmium ion
1:~12.7M SeLD24Tr	DCL, chloroplast precursor **(LOC_Os08g21700)** LOC_Os06g48600, LOC_Os09g04440	2	Amine metabolic process

4.3.2.3 Cis eQTLs

Cis eQTLs refers to the expression QTLs located within a threshold distance (here 1Mb) of the DNA polymorphism or SNP. Number of cis-eQTLs are listed in Table 4.4. Cis eQTLs are very significant as they co-locate with the SNP region, meaning that there might be polymorphism within the gene itself that is causing the expression variation. Cis eQTLs which are clustered in same region of the genome can be called cis-eQTL hotspots. Table 4.6 lists some of the cis-eQTL hotspots identified, along with their gene description and functions.

Table 4.6 Cis-eQTL hotspots genes and their biological functions *(here for location X:X indicates chromosome: position, M indicate million base pair, Bold accessions can also act as master regulators)*

Location	Biological function	Accession	Gene
10:~14.7M ReLC72Ci 16 eQTLs	amino acid transmembrane transport	**LOC_Os10g30090**	amino acid permease
		LOC_Os10g29470	dehydrogenase
		LOC_Os10g28360	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase protein
	cellular amino acid biosynthetic process	**LOC_Os10g28360**	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase protein
		LOC_Os10g30090	amino acid permease
	fructose 6-phosphate metabolic process, glycolysis	LOC_Os10g26570	phosphofructokinase
	chlorophyll catabolic process	LOC_Os10g28370	chlorophyllase-2, chloroplast precursor
	oxidation reduction	LOC_Os10g27340	oxidoreductase
1:~25.2M ReLS72Ci 9 eQTLs	carbohydrate metabolic process	LOC_Os01g43160	polygalacturonase
	protein amino acid dephosphorylation	LOC_Os01g43100	protein phosphatase 2C
	negative regulation of cell proliferation, lateral root morphogenesis	LOC_Os01g43040	amino acid transporter
1:~42M ReLD72Ci 6 eQTLs	two-component signal transduction system, circadian rhythm, response to cytokinin stimulus	LOC_Os01g72330	OsRR4 type-A response regulator
	double-strand break repair via homologous recombination,	LOC_Os01g71960	endonuclease
	proline biosynthetic process, response to salt stress, oxidation reduction, L-proline biosynthetic process	LOC_Os01g71990	pyrroline-5-carboxylate reductase
		LOC_Os01g72490	LRP1
		LOC_Os01g72170	Glutathione S-transferase
12:~0.8M ReRC72Ci 8 eQTLs	TOR signaling pathway, embryonic development ending in seed dormancy, cell growth	LOC_Os12g01922	WD domain and HEAT domain containing protein
	protein amino acid phosphorylation, signal transduction, negative regulation of abscisic acid mediated signaling pathway	LOC_Os12g02200	CAMK KIN1/SNF1/Nim1 like.6 - CAMK includes calcium/calmodulin dependent protein kinases
	carbohydrate metabolic process, metabolic process	LOC_Os12g02520	OsMan09-Endo-Beta-Mannanase,_expressed
	xylan catabolic process, L-arabinose metabolic process	LOC_Os12g03480	alpha-N-arabino furanosidase
3:~33.4M ReRS72Ci 11 eQTLs		LOC_Os03g60560	ZOS3-21-C2H2 zinc finger protein
	metabolic process	LOC_Os03g57640	gibberellin_receptor_GID1L2
		LOC_Os03g60340	Leaf senescence related protein
	DNA replication initiation, regulation of DNA	LOC_Os03g59470	Stage II sporulation protein E

	replication, methylation-dependent chromatin silencing, RNA interference, regulation of cell cycle, metabolic process		
		LOC_Os03g57320	expp1 protein precursor
		LOC_Os03g58980	Cupin domain containing protein
	nucleus organization,poly(A)+ mRNA export from nucleus	LOC_Os03g57430	expressed_protein
		LOC_Os03g59225	expressed_protein
	actin filament depolymerization, response to stress	LOC_Os03g60580	actin-depolymerizing factor, putative
	response to abscisic acid stimulus	LOC_Os03g58790	ATPase
		LOC_Os03g60520	Expressed protein
10:~20.9M ReRD72Ci 10 eQTLs		LOC_Os10g40530	LTPL146 Protease_inhibitor /seed_storage / LTP family protein precursor
	protein import into mitochondrial outer membrane	LOC_Os10g37430	Expressed protein
		LOC_Os10g39740	Glutathione S-transferase,_putative,_expressed
	sexual reproduction	LOC_Os10g40730	Expansin precursor
	biosynthetic process	LOC_Os10g39710	strictosidine_synthase
	response to oxidative stress, oxidation reduction	LOC_Os10g39170	peroxidase_precursor
	activation of protein kinase C activity by G-protein coupled receptor protein signaling pathway	LOC_Os10g37280	diacylglycerol_kinase
	leucine catabolic process, response to absence of light	LOC_Os10g37210	FAD dependent oxidoreductase domain containing protein
	L-phenylalanine biosynthetic process	LOC_Os10g37980	Prephenate dehydratase domain containing protein

Besides the hotspots there are several significant cis-eQTLs which can be associated with relevant biological functions associated with salt tolerance mechanism. Some of the over represented cis-eQTLs, worth investigating are listed in Table 4.7

Table 4.7 Significant cis-eQTLs as a list of candidate genes

Accessions	**Gene Description**	**Cis-eQTL**
LOC_Os01g42090	Nodulin like mtn3	ReLD72Ci
LOC_Os02g58350	OsRR3 – response regulators	ReLD72Ci
LOC_Os02g50060	OsCML20	ReLD72Ci
LOC_Os02g41860	Aquaporin	ReLD72Ci
LOC_Os05g09480	OsIAA16 Auxin	ReLD72Ci
LOC_Os11g12810	Sucrose phosphate synthase	ReLD72Ci
LOC_Os01g72170	Glutathione-s transferase	ReLD72Ci
LOC_Os10g39870	ATG1 involved in autophagy	SeLD72Ci
LOC_Os08g02700	Fructose bisphosphate aldolase	SeLD72Ci
LOC_Os01g49720	Glutathione s transferase	SeLD72Ci
LOC_Os03g52090	Calcium transporting ATPase	SeLD72Ci
LOC_Os03g22550	Cation efflux family protein	SeLD72Ci
LOC_Os09g20990	Trehalose – 6 –phosphate synthase	SeLD72Ci
LOC_Os03g52090	calcium-transporting ATPase 3	SeLD24Ci
LOC_Os06g47840	ZOS6-08 - C2H2 zinc finger protein	SeLD24Ci

LOC_Os01g43410	CAMK like.9 - CAMK includes calcium/calmodulin depedent protein kinases	SeLD24Ci
LOC_Os01g57260	vacuolar protein sorting-associated protein 28 homolog 2	SeLD24Ci
LOC_Os04g52950	nitrate-induced NOI protein	SeLD24Ci
LOC_Os01g09790	IQ calmodulin-binding motif domain containing protein	ReRD72Ci
LOC_Os01g32120	OsCML11 - Calmodulin-related calcium sensor protein	ReRD72Ci
LOC_Os01g61780	vacuolar ATP synthase 98 kDa subunit	ReRD72Ci
LOC_Os01g74450	aquaporin protein	ReRD72Ci
LOC_Os05g34150	glutathione S-transferase	ReRD72Ci
LOC_Os10g30156	starch synthase	ReRD72Ci
LOC_Os10g32980	CESA7 - cellulose synthase	ReRD72Ci
LOC_Os12g05180	avr9/Cf-9 rapidly elicited protein	ReRD72Ci
LOC_Os09g24980	vesicle transport v-SNARE protein	SeRD24Ci
LOC_Os01g68589	LTPL39 - Protease inhibitor/seed storage/LTP family protein precursor	SeRD24Ci
LOC_Os10g38214	chloroplast 30S ribosomal protein S4	SeRD24Ci
LOC_Os01g63230	growth regulator related protein	SeRD24Ci
LOC_Os06g43130	tic21- works in protein translocation	SeRD24Ci
LOC_Os03g61760	OsSPL6 - SBP-box gene family member	SeRD24Ci
LOC_Os03g08230	sodium/calcium exchanger protein	SeRD24Ci
LOC_Os06g46410	auxin response factor	SeRD24Ci
LOC_Os09g34900	bile acid sodium symporter family protein	SeRD24Ci
LOC_Os10g38850	amine oxidase	SeRD72Ci
LOC_Os01g67860	fructose-bisphospate aldolase isozyme	SeRD72Ci
LOC_Os01g08660	aquaporin protein	SeRD72Ci
LOC_Os02g47320	vacuolar ATPase G subunit	SeRD72Ci
LOC_Os05g37690	OsFBL23 - F-box domain and LRR containing protein	SeRD72Ci
LOC_Os01g72230	stromal membrane-associated protein	SeRD72Ci
LOC_Os03g53200	OsCML4 - Calmodulin-related calcium sensor protein	SeRD72Ci
LOC_Os02g51110	aquaporin protein	SeRD72Ci

Several glutathione-s-transferases, aquaporin, calmodulin related protein, nodulin proteins were identified as cis eQTLs. Some cis-eQTLs were found to function in protein import into chloroplast stroma e.g. tic21. A cis eQTL for trehalose-6-phosphate (T6P) synthase was found, trehalose is a nonreducing disaccharide and T6P can act as a signal of sucrose status and as a regulatory molecule especially in sugar influx and metabolism. Several cis eQTLs for calcium/calmodulin dependent protein kinases were found which acts in calcium dependent signaling. In Arabidopsis Zhou et al. (2016) have shown that calmodulins can serve as signals in plant salt resistance by promoting NO accumulation through the binding and inhibition of GSNOR (*S*-nitrosoglutathione reductase). Zinc finger proteins and glutathione-s-transferases are also over represented. One cis eQTL for a sodium/calcium exchanger was found after 24 our stress in and one cis eQTL was found for cation efflux family protein after 72 hour stress in seedling stage (Table 4.7). Nodulin-like genes in non-nodulating plants suggests a possible role for nodulin-like proteins in regulating plant growth and development, although the functions of most nodulin-like proteins remain unclear. Some aquaporin can act as non-selective cation transporter when expressed in root (Byrt et al. 2017).

4.3.2.4 Functional network of eQTLs harbored in pQTL region

Identification of eQTLs located in the pQTL region identified in chapter 3 is an intelligent choice to explain the genes involved in regulating a phenotype. Cis and trans eQTLs affected

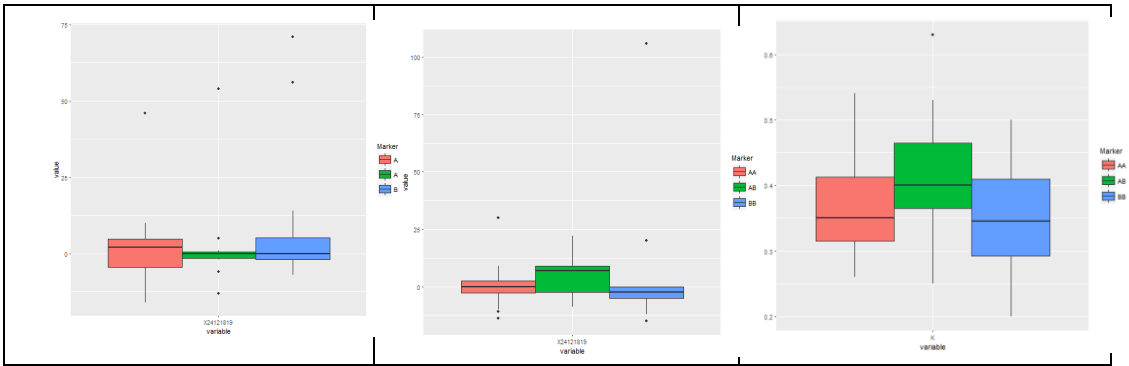
by the specific polymorphic region in the DNA can be more informative than the list of underlying genes in the pQTL region. Moreover they can narrow down relevant genes. pQTL regions for important traits like root length, potassium content and sodium content from seedling stage as well as spikelet fertility, day to maturity and thousand grain weight from reproductive stage were investigated for the eQTLs.

4.3.2.4.1 Seedling stage

Potassium Content

For the first QTL for K content, one calcium-transporting ATPase 3 was found at the SeLD72Ci condition, which participates in cation transport, calcium ion transport and root development. A transporter family protein (LOC_Os03g39710), a cation-transporting ATPase (LOC_Os05g33390), two heat shock proteins (LOC_Os06g50300 andLOC_Os09g30412) were also found to be involved. A potassium transporter (LOC_Os09g21000) was also identified in seedling root control tissue.

The expression levels and phenotypic screening values of few sample genes were visualized to track the reliability of the eQTLs identified. For example, a Calcium transporting ATPase3 was found to be associated with the regions which is responsible for the Potassium content pQTL in the seedling stage. Distribution of expression values (left) and potassium content (right) for *Horkuch, IR29* and Hetero alleles showed similar pattern after 72 hour stress. It is interesting to note that after 24 hr stress the pattern was slightly different (Fig 4.19)



24h	72 h	After harvest
calcium-transporting ATPase 3 expression		**K^+ content**

Fig 4.19 Association of distribution patterns for expression and phenotypic measurements for potassium content. (*Here AA = Horkuch, AB = Heterozygote, BB = IR29*)

The second K^+ content QTL was found to affect more genes at 24h rather than 72h which involves nuclear mRNA splicing, via spliceosome network with the master regulator (a splicing factor) (LOC_Os05g07050) and fructose-bisphosphate aldolase isozyme (LOC_Os01g678600) to regulate glycolysis, response to cadmium ion and salt stress. This co-locates with a hotspot of trans-eQTLs.

Root Length

Differential expression of an OsIAA31-Auxin-responsive Aux/IAA gene family member was observed in leaf after 24 hour of stress, which acts in response to water deprivation, gravitropism, auxin-mediated signaling pathway, jasmonic acid and lateral root morphogenesis with another IAA gene (LOC_Os12g40900). A cullin family domain-containing protein (LOC_Os04g40830) was identified in root after 72 hour salt stress which can act with a tetratricopeptide repeat domain-containing protein (LOC_Os06g41750). This was discovered from network-guided genes in cell division, root cap development, auxin stimulus, root meristem specification, etc. A OsFBX339-F-box domain containing protein (LOC_Os09g35680) along with IAA (LOC_Os12g40900) and another F box domain protein (LOC_Os04g32460) can function in primary root development, stamen development, root development (Table 4.8).

Sodium Content

A network of sodium transport was observed involving cytochrome P450, dehydration-response related protein, calcium-binding protein, FGGY family of carbohydrate kinases and some other genes listed in Table 4.8 where the master regulator is a citrate transporter.

Table 4.8 eQTLs associated with pQTL regions

Root Length	• OsIAA31-Auxin-responsive Aux/IAA gene • IAA (LOC_Os12g40900)	after 24 hour of stress water deprivation, gravitropism, auxin mediated signaling pathway, jasmonic acid, lateral root morphogenesis
	• Cullin family domain containing protein LOC_Os04g40830 • tetratricopeptide repeat domain containing protein LOC_Os06g41750	Cell division, root cap dev, auxin stimulus
	• OsFBX339-F-box domain containing protein (LOC_Os09g35680) • IAA (LOC_Os12g40900) • F box domain protein (LOC_Os04g32460)	Primary root, stamen development
Potassium Content	• calcium-transporting ATPase 3 cis eQTL	cation transport, calcium ion transport, root development
	• A transporter family protein (LOC_Os03g39710) • a cation-transporting ATPase (LOC_Os05g33390) • two heat shock proteins (LOC_Os06g50300 and LOC_Os09g30412)	
	• Potassium transporter (LOC_Os09g21000)	
Sodium Content	• **citrate transporter (LOC_Os09g31130)** • cytochrome P450 (LOC_Os01g41820) • dehydration response related protein (LOC_Os02g45310) • UDP-glucuronosyl and UDP-glucosyl transferase domain containing protein(LOC_Os05g42020) • acyl-coenzyme A oxidase 1.2 (LOC_Os06g01390) • glucosyltransferase (LOC_Os07g10190) • calcium-binding protein (LOC_Os07g39630) • FGGY family of carbohydrate kinases (LOC_Os07g44660)	Sodium transport

	• MBTB23- Bric-a-Brac (LOC_Os08g13070), GIL1(LOC_Os10g36400) • cytokinin-N-glucosyltransferase(LOC_Os11g25454)	
Spikelet fertility	• ZOS10-06-C2H2 zinc finger protein LOC_Os10g35190 cis eQTL	
	• chlorophyll A-B binding protein **LOC_Os04g38410** • LOC_Os01g59080, LOC_Os02g34810, LOC_Os07g04840, LOC_Os08g45190, LOC_Os09g30340	calcium ion transport, Golgi organization, response to salt stress, photosynthesis, light harvesting , nonphotochemical quenching, photosynthesis, cysteine biosynthetic process light reaction, regulation of protein amino acid dephosphorylation
FGW	• **ATG1 (**LOC_Os10g39870)	Cytochrome c biogenesis related protein
DM	• **Leucine carboxyl methyl transferase(LOC_Os06g04820)** • suppressor of Mek LOC_Os01g23590 • ubiquitin fusion degradation protein LOC_Os02g08480 • AMK_CAMK_like.20 - CAMK includes calcium/calmodulin dependent protein kinases LOC_Os03g41460 • AGAP007115-PA, LOC_Os05g43560	Methylation Photoperiodism flowering
Thousand Grain Weight	• gibberellin 20 oxidase 2 (LOC_Os01g66100)	oxidation reduction, short-day photoperiodism, flowering , response to gibberellin stimulus, unidimensional cell growth, leaf development
	• mitochondrial prohibitin complex protein 1 (LOC_Os02g37000	response to stress, mitochondrion organization, response to salt stress, response to auxin stimulus, cell growth, lateral root development, cell division, response to ethylene stimulus
	• protein spotted leaf 11(LOC_Os12g38210)	protein ubiquitination, defense response, flower development, negative regulation of defense response, negative regulation of programmed cell death, regulation of long-day photoperiodism, flowering
	• cation efflux family protein (LOC_Os02g53490) cis eQTL	Cation transport

4.3.2.4.2 Reproductive stage

Spikelet Fertility and Filled grain weight (SF, FGW)

The spikelet fertility and filled grain weight Qtl1 or pQTLs overlaps with each other and resides at chr 10 ~18.6 Mb. A cis eQTL was found for C2H2 zinc finger protein for this SNP and a trans-eQTL network was found with master regulator chlorophyll A-B binding protein (LOC_Os04g38410) with other genes to form a network for calcium ion transport, Golgi organization, response to salt stress, photosynthesis, light harvesting, etc. A maltose metabolic

pathway as well as multiple photosynthesis and water deprivation- related networks were formed. This might be related to better photosynthesis under salt stress to have better spikelet fertility. For QTL2 of filled grain weight at chromosome 10 at ~21.6 Mb a cis eQTL was found for Cytochrome C biogenesis-related protein, ATG1.

Day to Maturity (DM)

For days to maturity the pQTL was found at chromosome 10: ~18 Mb. Together with two methyl transferases, a network for methyl transferase and photoperiodism was observed (Table 4.8). Also a network of positive regulation of flowering formed with LOC_Os08g08830, a pre mRNA splicing factor (LOC_Os05g07050.1) and a transcription factor jumonji (LOC_Os05g10770.1).

Thousand grain weight (THGW)

THGW QTL was found in chromosome 11 at ~28.1 Mb. Interestingly, some very important eQTLs are found to be affected by this region also at the seedling stage after 72 hour differential stress. This might or might not have relation with the reproductive stage thousand grain weight which is a yield component, but seemed worth investigating. The interesting genes affected included a gibberellin 20 oxidase 2 (LOC_Os01g66100.1) associated with oxidation reduction, short-day photoperiodism, flowering response to gibberellin stimulus, unidimensional cell growth, leaf development, a mitochondrial prohibitin complex protein 1 (LOC_Os02g37000.1) associated with response to stress, mitochondrion organization response to salt stress, response to auxin stimulus, cell growth, lateral root development, cell division and response to ethylene stimulus. A protein spotted leaf 11 (LOC_Os12g38210) involved in protein ubiquitination, protein ubiquitination, defense response, flower development, negative regulation of defense response, negative regulation of programmed cell death, regulation of long-day photoperiodism, flowering, defense response to bacterium and a cation efflux family protein (LOC_Os02g53490) involved in cation transport was also observed.

4.4 Discussion

Discovery of the heritable combination of genes and their expression pattern differences under stress is very important in terms of understanding mechanism of tolerance to stress. When this kind of comparative study is performed on donor parents and their progenies from a breeding population, a large amount of information becomes available for comparison and contrasting purpose, to understand the mechanisms and in being able to choose the right donors in breeding programs.

An interesting aspect of this study is that we tried to divide the population into tolerant and sensitive groups which made the differential expression study much more informative. For seedling stage this categorization was done using SES score representing the survival ability of the plants during stress and for reproductive stage the categorization was done using thousand grain weight under stress which is a component of yield. Since high yield under stress is the desired tolerance parameter, this justifies the categorization. A rice plant is highly sensitive in seedling and reproductive stage of its life rather than any other stages and the tolerance ability lies from salt sensing in soil, stable growth using better photosynthesis machinery despite stress, finally leading to enough yield. So the mechanisms revealed for tolerating stress in these two stages should help outlining the context of the salt tolerance response of a plant. Interestingly this categorization in both stages also correlated with tolerant and sensitive parents' expression pattern as expected.

In this study a 3′ tag based RNAseq protocol was adopted which is cost effective when dealing with large number of samples. Because this does not require sequencing of the whole mRNA. By using only the 3′ end signal and sequencing a specific length from that end picks up most transcripts and allows counting of the transcript abundance. This is satisfactorily used for differential expression comparison or determining the actual number of transcripts of specific genes. However, it makes the detection of alternative splicing as well as RNAseq based genotyping difficult.

A high number of biological samples for differential expression analysis was a plus point of this study because a subset of the population containing a minimum of 15 samples for tolerant, sensitive, *Horkuch*♀ and *IR*29♀ category were present. If the tissue, developmental stage or treatment categories are considered, minimum number of biological samples were even higher. To our knowledge this is an uncommon but very promising experimental design for deciphering the candidate genes with high confidence and has not been implemented before. Moreover differential expression of genes due to cytoplasmic background of parents were also observed, which suggested the use of cytoplasm as covariate and the eQTL mapping was carried out following both.

At the whole expression level study many members of the OsRCI2 (Rare Cold Inducible) also known as LT (Low Temperature) family was found to be differentially regulated. Rice homologues of this protein from Arabidopsis is reported in the MSU 7.0 database which reports existence of 12 members in rice spanning chromosome 1, 3, 5, 6, 7 and 9. Significant up-regulated expression was observed for OsRCI2-10 /LT16A and OsRCI2-6/LT16B under stress in leaves of the sensitive progenies at reproductive stage, whereas the former was also upregulated without treatment in tolerant roots. Interestingly in seedling stage LT16A was

found to be upregulated in the leaf tissue of tolerant progenies but in reproductive stage it was in tolerant roots. Over expression of LT16B under stress in leaves of sensitive progenies at reproductive stage compared to the tolerant progenies was also confirmed by qPCR experiment. Moreover in the allele diversity mining study (Chapter 6) a 13bp deletion was found in most tolerant donors in the 3′ downstream region of this particular gene, which might give a potential explanation for the downregulation of the gene in tolerant plants under stress, involving a possible role of noncoding RNAs. These are hydrophobic membrane potential proteins with a possible function in water deprivation. Further investigation of this gene in context of relative water content QTL found from IR29 the sensitive variety (Chapter3) can give insight on its function maintaining water potential.

In contrast to sensitive progenies, tolerant plants showed upregulation of an aquaporin gene under stress in the reproductive stage. Aquaporins have been known to act as pores by transporting water across membranes in plants and animals playing a critical role in controlling the water content of cells. Byrt et al (2016) recently reported role of arabidopsis aquaporin AtPIP2,1 in non selective cation transport in certain conditions, regulated by Ca^{2+} and pH showing possible role in transport of sodium ions.

The differential expression analysis from sensitive and tolerant plants revealed that while enhancement in signaling is a hallmark of tolerant roots, downregulation of photosynthesis related and ATP biosynthesis genes are unique to levels of sensitive plants. Moreover constitutive expression of some genes related to regulation of membrane potential was observed in tolerant roots speculating crucial mechanisms for protection against high salt and low water conditions.

We observed higher number of signals in leaf compared to root tissues. It has been found in some other studies also that more differentially expressed genes are observed in sensitive plants compared to the tolerant plants which is also the case observed in our study. This indicates a tolerant plant might be already occupied with necessary machineries to combat the stress whereas a sensitive plant needs to up or down regulate expression of some more genes (irrespective of success) upon stress exposure generating high number of DEGs in sensitive plants.

Having a 5-way contrasting model for tissue, time point, treatment, developmental stage and cytoplasm helped in pinpointing specific eQTLs belonging to specific metabolic categories. For example some hotspots were found after 24 hour of salt stress in seedling stage which were absent after 72 hour stress. The eQTLs are significant sets of genes, supported by both DNA

and expression polymorphisms' association and can narrow down a promising set of candidate genes.

The eQTL hotspots identified specific regions of the genome which affects a set of genes that might function together as network of co-expressed genes to manipulate a specific phenotype. For example networks of genes could be identified with possible role in photosynthesis regulation, carbohydrate metabolism, salt stress response, etc. Interestingly the photosynthesis related hotspot region (under non stress condition) overlaps with the pQTL for yield components under stress (chapter 3). Identification of an eQTL hotspot for photosynthesis also support the better photosynthetic ability of *Horkuch* under stress as previously reported (Lisa et al , 2011). In chromosome 1, some specific regions harbor signals for salt stress responsive genes as both cis and trans eQTLS which might be correlated to previous identification of vegetative stage salt tolerant QTLs like *Saltol*, SKC1, etc. in chromosome 1.

The beauty of this eQTL study is that it reveals both cis and trans-eQTLs. Since trans-eQTLs are distantly located genes whose expression is also affected by the specific DNA level polymorphism i.e. the specific SNP region, they are more powerful in explaining a functional association. It expands the scope of identification of potential candidate genes from all over the genome rather than in the limited specific DNA polymorphic region where a QTL is identified. On the other hand cis eQTLs are more accurate as usually they possess structural variation also leading to the expression difference as well the DNA polymorphism.

Most interesting were the insights on pQTL (physiological QTLs from chapter 3) and eQTL association, which revealed potential role of some genes with specific functions relevant to the physiological traits. Both cis and trans-eQTL provide more informative results than mining the genes underlying a pQTL region. For example the region associated with days to maturity showed several genes associated with flowering, photoperiodism, morphogenesis, seedling to reproductive stage maturity, etc. A region associated with potassium content showed cis eQTL for calcium-transporting ATPase 3, which regulates cation transport and a trans-eQTLs for a potassium transporter. Further investigation of specific pathways should reveal a model for the underlying mechanisms.

In a nutshell, the expression study has provided the information on differentially expressed genes in 5 contrasting models for developmental stage, tissue, treatment, cytoplasm and timepoint. The eQTL study could dig out more by identifying hotspots as well as relating the pQTLs and expression difference of the genes. The functional annotation of the candidate genes as well as the functional network analysis could associate the genome to transcriptome explaining the phenotypes also.

Chapter 5

Automated Image Phenotyping

5.1 Overview

The selected plants from the bi-parental F_3 and F_5 generations were phenotyped using automated high-throughput image phenotyping. The goal of this phenome analysis was as described below:

1. Validation of the selected tolerant and sensitive lines with respect to the F_3 manual phenotyping (Chapter 3) by comparing their digital growth traits with tolerance scores.
2. Using the non-invasive method of imaging, development of digital traits using continuous monitoring to capture the gap between osmotic and ionic phase under salt stress.
3. Identification of potential Quantitative trait loci (QTL) for the digital traits using the linkage map data (Chapter 3) and functional analysis by the eQTL data (Chapter 5)

High throughput phenotyping of plants have opened the door for investigating the growth dynamics of a plant in temporal basis. Instead of cross sectional endpoint data, this longitudinal data gives the opportunity to investigate the osmotic components involved in early response to salt stress which cannot be identified in later stages in endpoint phenotyping. In this chapter, focus has been given to the growth dynamics extracted from the RGB (Red Green Blue) images, followed by measurement of growth rate related traits. Furthermore, using the genotyping data and expression data from the previous chapters few potential regions in the genome has been identified that are associated with the digital traits, followed by downstream functional analysis.

5.2 Material and Methods

5.2.1 Plant Materials

Two individual experiments were carried out during 2015 (May, 2015 – September 2015) and 2016 (June 2016 – October 2016) in the High Throughput Phenotyping Greenhouse of Nebraska Innovation Campus, UNL, NE, USA. For the first year experiment, selected F_3 plants and some F_4 plants of IR × Horkuch reciprocal cross population were imported from DU, Bangladesh to UNL, Nebraska, USA with USDA permit no PCIP-14-00687, received 5th Dec, 2014). For the second year experiment (2016), F_5 plants were selected based upon QTL mapping results and a subset was sent to UNL, NE, USA permit no PCIP-16-00133, received 4th may, 2016. A list of selected samples used in both study is mentioned in Table XX.

5.2.2 Growth condition

5.2.2.1 Seed treatment and germination: Seeds were sent to USDA first according to the rule of the seed import permit as after treatment those were considered as Quarantined material. Before import, the seeds were fumigated with Phostoxin for 72 hours @ one pellet per 0.5 cubic meter at Bangladesh Rice Research Institute. In USDA PGIS (Plant Germplasm Inspection Station), the seeds were inspected and subjected to hot water treatment at 56°C for 15 minutes

followed by drying. Upon receiving, the seeds were de-husked and surface sterilized with a 0.525 percent sodium hypochlorite solution insuring that all seed surfaces are thoroughly wetted. The rice seeds were germinated on blotters moistened with water in enclosed petri dishes, allowing transmission of light. According to the regulation by the permit the seeds were equidistant from each other and no more than 20 seeds per 9 cm. culture plate were allowed. (Fig 5.1) The seeds were germinated in an incubator with 28 degrees C.

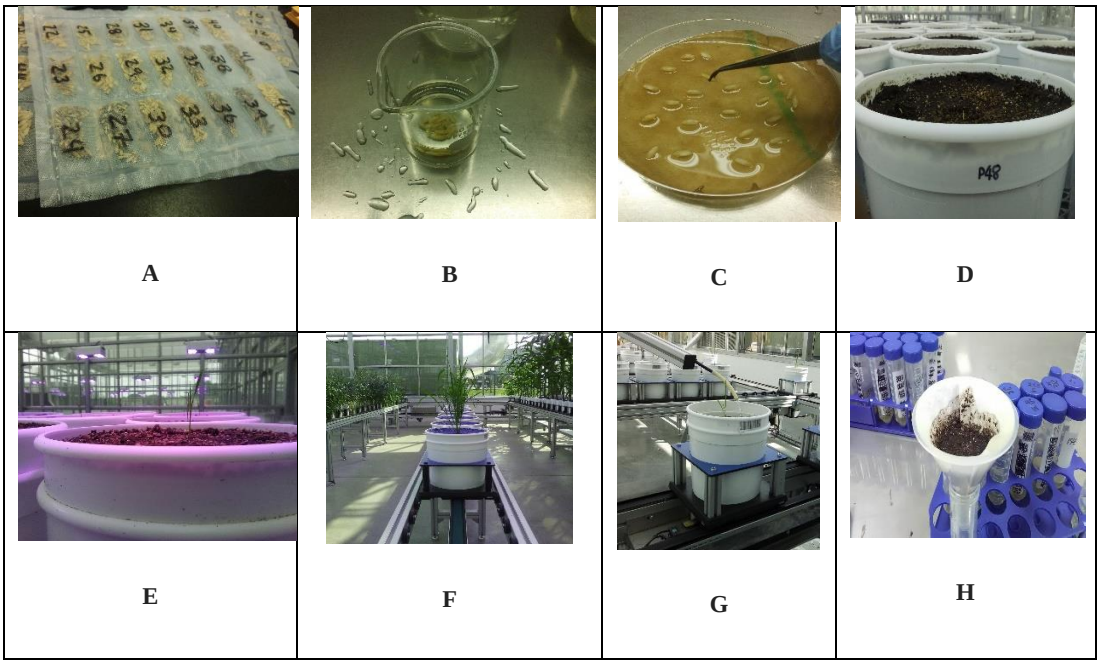


Fig 5.1: Experiment setup for plants in the NIC greenhouse From seed genrminationto potting of plants and downstream analysis.

5.2.2.2 Soil preparation: Soil used in the experiment was Fafard germination mix, (60-70% peat, fine perlite, vermiculite, starter nutrients, limestone). 4 pounds each of Osmocote 15-9-12 and 17-5-11 was added to this with micronutrients at pH 6.3. This mixture made sure the consistency and did not vary experimentally.

5.2.2.3 Pot arrangement: In experiment 1 two pots were used per plant. One pots was perforated and was set into another pot to allow draining of water. In experiment 2 the two pot system was not used and only one intact pot was used per plant. In addition to avoid the algae contamination in the images a blue mesh pad was used on each pot, to ease the top view of the fluorescent images. The pots were able to contain about 5 kg of soil. For experiment 2 soil were filled upto 4200 kg. The soil was saturated with enough water to make it prepared for the plant transplantation. Arrangement of the pots are shown in Fig 5.1 F.

5.2.2.4 Plant growth condition: After two days the plants were transferred to a growth chamber with a light/dark cycle of 12 hours in 20 /-2 degrees C. Four Healthy seedlings were transplanted to soil. After 7 to 8 days Plants of unequal size in control and stress designated plant were discarded keeping only one plant per pot of equal size. In total 35 F_3 and about 30 F_5 progenies

were analyzed through the system along with the parents in three replicates and two treatment condition that is control and salt stress imaging 6 plants per accession. The green house condition was set at 25-27°C during day and 20-22°C at night.

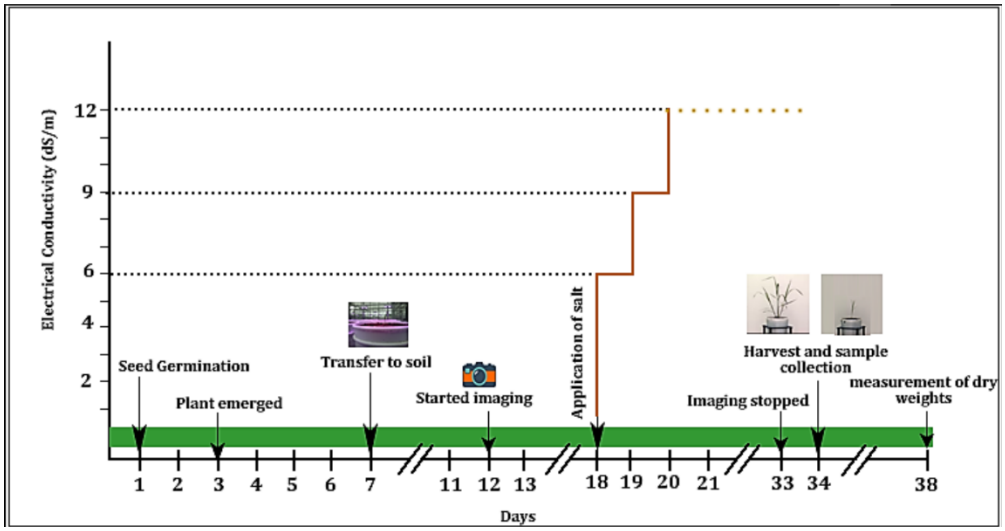


Fig 5.2 Schematic diagram of the days for the imaging and salt stress application

5.2.3 Using the Lemnatec scanalyzer system for imaging

The pots were loaded in image belts after 7 days of growth in the bench, when the plants were 11 days old. For loading, the pots were put on each carrier of the belt, the barcodes were scanned to set the entry into the database. The individual pots then can move through the conveyor belt to their respective spot in the greenhouse. Each lane of the belt can accommodate 20 carriers. Images were taken daily in the RGB, Fluorescent and infrared chambers. They were weighted automatically each day and according to the weight balance amount of water needed was calculated automatically and according to that extra water needed was poured by the system. A target weight for the pots were fixed from the very first which was calculated from the dry and saturated weight of the soil, that how much water the soil can hold.

Images were taken in four chambers. The specifications for the cameras are in Table 5.1

Table 5.1: Specification of the cameras used for taking images

Camera	Resolution	Pixels	Spectral Range	Measures	View
Visible light	2454× 2056	24bit	VIS 400-700nm	Plant height, width, growth rate, stem diameter, Biomass, internode length, leaf rolling, leaf tracking over time, color classification, Tip burn, leaf area	Side, Top

Infrared	640x480	24 bit	IR 8 to 14uM	Plant temperature, stomatal aperture, transpiration	Side Top
Fluorescent	1390x1038	24bit	Excitation range :- 400 – 500 nm Spectral range : Vis 620-900 nm	Chlorophyll fluorescence under continuous blue light	Side Top
Hyperspectral	Spectral resolution 320px line width	Spectral resolution 242 bands	Spectral range VNIR 545-1700nm	Substance specific reflectance learned finger printing or pattern, pure pattern finding	Side

5.2.4 Salt application

Plants were salt stressed after 7 days of imaging and were imaged for total 21 days. A schematic diagram of plant growth and salt application date are shown in Fig 5.2. To measure the water capacity of the soil, an experiment was carried out, measuring the weight of the over-dried soil and the saturated soil. Same amount of soil was oven dried at 72° C for 4 days and then weighed. Another pot was water saturated and was weighed. Deducting the over-dried soil from the water saturated soil gave the water content capacity of the soil. This water volume was used to calculate the final salt concentration in the soil. NaCl solution was prepared calculating the final concentration to be 120mM. At first 60 mM and then gradually increased to 90mM and in the third day, final concentration of 120mM. NaCl and $CaCl_2$ was used in the ration of 100:3.3mM.Salt was applied as a solution on top of the soil.

5.2.5 Manual measurement of biomass

After completion of imaging the plants were thoroughly cleaned in distilled water and subjected to dry. Length and fresh weight of shoot, root were taken followed by incubation for 5 days in a 72° C oven and measurement of dry weights. For the F_5 plants Electrical conductivity of the saturated soil was also measured by drying the soil, making saturated paste followed by filtration and measurement by EC meter.

5.2.6 Image Data Analysis

For extracting the raw pixel information and generating digital traits from the images, open source tool image harvest (knecht et al, 2016) was used which is an open source python library and uses algorithms from OpenCV. Due to large size of the dataset HPC server tusker was used where the images were uploaded and the workflow for image processing and analysis were

developed. The JavaScript Object Notation file config.json fles has been configured to fit the image information and crawl.json file has been adjusted with the individual directory structure. The ih-crawl script crawled through the directories and made a database of the directory structure. Four distinct modules were used for the processing steps i.e (1) background removal, (2) noise reduction, (3) cropping and (4) data gathering using functions like, crop(), colorFilter(), contourCut(), contourChop() functions to extract the region of interest (ROI). After modification of the imageproc.json file the ih-run command generated files ready for Pegasus submission. It was then sql aggregated and the final pixel information was retrieved from the sql database.

5.2.6.1 Extraction of digital traits:

Digital traits were extracted from the pixel and centroid information. Here 22 digital traits have been considered. The definition of the traits are given in the Table 5.2. They were used to describe plant morphological qualities.

Table 5.2 Formula and description of the digital traits used in the study *(Adapted from Campbel et al, 2015)(SV: side view; TV: top view; n: number of side view images)*

Digital Trait	**Class**	**Formula**	**Description**
$Area_{TV}$	Biomass	$Area_{TV} = Pixels_{TV}$	Plant pixels from extracted from top view (TV) image
PSA	Biomass	$PSA = Pixels_{TV} + \sum_i Pixels_{SV_i}$	Summation of the plant pixels from all side view images and top view
$Ht{:}Width_{SV}$	Plant Shape	$Ht{:}Width_{SV} = \frac{\sum_i Ht_{SV_i}}{n} \div \frac{\sum_i Width_{SV_i}}{n}$	Ratio of the average height of the cropped SV image to the average width of the cropped SV image with dimFromROI()
CA_{TV}	Plant Size	$CA_{TV} = Ht_{TV} \times Width_{TV}$	Product of the height and width of the cropped TV image with dimFromROI()
$Den1_{SV}$	Density/ Compactness	$Den1_{SV} = \frac{\sum_i Pixels_{SV_i}}{\frac{\sum_i Convex\ Hull\ Area_{SV_i}}{n}}$	Ratio of the summation of the plant pixels from all SV images and the average convex hull area of all SV images
$Den2_{SV}$	Density/ Compactness	$Den2_{SV} = \frac{\sum_i Pixels_{SV_i}}{\frac{\sum_i Min.Enclosing\ Circle\ Radi}{n}}$	Summation of the plant pixels from all SV images to the average radius of the minimum enclosing circle of all SV images

$Den3_{SV}$	Density/ Compactness	$Den3_{SV} = \frac{\sum_i Pixels_{SV_i}}{\frac{\sum_i Ht_{SV_i}}{n} \times \frac{\sum_i Width_{SV_i}}{n}}$	Summation of the plant pixels from all SV images to the average side view cropped area (CA_{SV}) with dimFromROI()
$Den1_{TV}$	Density/ Compactness	$Den1_{TV} = \frac{Pixels_{TV}}{Convex\ Hull\ Area_{TV}}$	Ratio of the plant pixels extracted from TV image to the convex hull area of the TV image
$Den2_{TV}$	Density/ Compactness	$Den2_{TV} = \frac{Pixels_{TV}}{Min.\ Enclosing\ Circle\ Radius_{TV}}$	Ratio of the plant pixels extracted from TV image to the radius of the minimum enclosing circle of the TV image
$Den3_{TV}$	Density/ Compactness	$Den3_{TV} = \frac{Pixels_{TV}}{Ht_{TV} \times Width_{TV}}$	Ratio of the plant pixels extracted from TV image to the top view cropped area (CA_{TV})
Ht2	Height	$Ht2 = \frac{\sum_i Ht_{SV_i}}{n}$	Average height of the final cropped image from all side views determined with dimFromROI()
Ht1	Height	$Ht1 = \frac{M_{01}}{M_{00}}$	Center of mass about the *y*-axis determined from the raw image moments (*M*) of side view images
GH1	Plant Shape	$GH1 = \frac{(Convex\ Hull\ Area_{TV})(M_{00\,SV})}{M_{01\,SV}}$	The ratio of the convex hull area of the TV image to the center of mass about the *y*-axis of all side view images
GH2	Plant Shape	$GH2 = \frac{(Convex\ Hull\ Area_{TV})}{Ht_{SV}}$	The ratio of the convex hull area of the TV image to the average height of all final cropped SV images determined with dimFromROI()
GH3	Plant Shape	$GH3 = \frac{(Min.\ Enclosing\ Circle\ Radius_{TV})(M_{00\,SV})}{M_{01\,SV}}$	The ratio of the minimum enclosing circle radius of the TV image to the average center of mass about the y-axis of all SV images
GH4	Plant Shape	$GH4 = \frac{(Min.\ Enclosing\ Circle\ Radius_{TV})}{Ht_{SV}}$	The ratio of the minimum enclosing circle radius of the TV image to the average height all final cropped SV images determined with dimFromROI()
Width SV	Plant Size	$Width_{SV} = \frac{\sum_i Width_{SV_i}}{n}$	The average width of all final cropped SV images

CASV	Plant Size	$CA_{SV} = \frac{\sum_i Ht_{SV_i}}{n} \times \frac{\sum_i Width_{SV_i}}{n}$	Product of the average height and width of all cropped SV images determined with dimFromROI()
CHTV	Plant Size	$CH_{TV} = Convex\ Hull\ Area_{TV}$	Convex hull area of the TV image
CHSV	Plant Size	$CH_{SV} = \frac{\sum_i Convex\ Hull\ Area_{SV_i}}{n}$	Average convex hull area of all SV images
CRTV	Plant Size	$CR_{TV} = Min.Enclosing\ Circle\ Radius_{TV}$	The radius of the minimum enclosing circle radius of the TV image
CRSV	Plant Size	$CR_{SV} = \frac{\sum_i Min.Enclosing\ Circle\ Radius_{SV_i}}{n}$	The average radius of the minimum enclosing circle radius of all SV images

The traits were calculated from the four side view at SV72, 144, 216, 288 and top view TV90.

The four side views were combined to get the average SV values.

5.2.6.2 Measurement of Growth rate related traits

In addition to the 22 digital traits regarding plant shape and size, the growth related traits have been calculated from the projected shot area value using the formulas in Table 5.3

Table 5.3: Formula and description for the growth rate related traits

Trait	Unit	Calculation	Description
TOL	Kilo pixel/day	$RGR_c - RGR_s$	Stress Tolerance
EGRI		RGR_s/RGR_c	Early Growth Response index, related to Osmotic tolerance
SWP	$\sqrt{Kpixel/day}$	$RGR_s/\sqrt{RGR_c}$	Stress Weighted Performance
SSI		$1 - (\frac{RGRs}{RGRc})/(RGRav_s/RGRav_c)$	Stress Susceptibility Index
STI		$RGR_c * RGR_s/(RGRav_c)^2$	Stress tolerance index
AGR, AGRc, AGRs	day^{-1}	$AGR_{(t_j,t_k)} = \frac{PSA_{t_k} - PSA_{t_j}}{t_k - t_j}$	Absolute Growth Rate (c and s refers to control and stress respectively)
RGR, RGRc, RGRs	day^{-1}	$RGR_{(t_j,t_k)} = \frac{\ln(PSA_{t_k}) - \ln(PSA_{t_j})}{t_k - t_j}$	Relative Growth Rate (c and s refers to control and stress respectively)

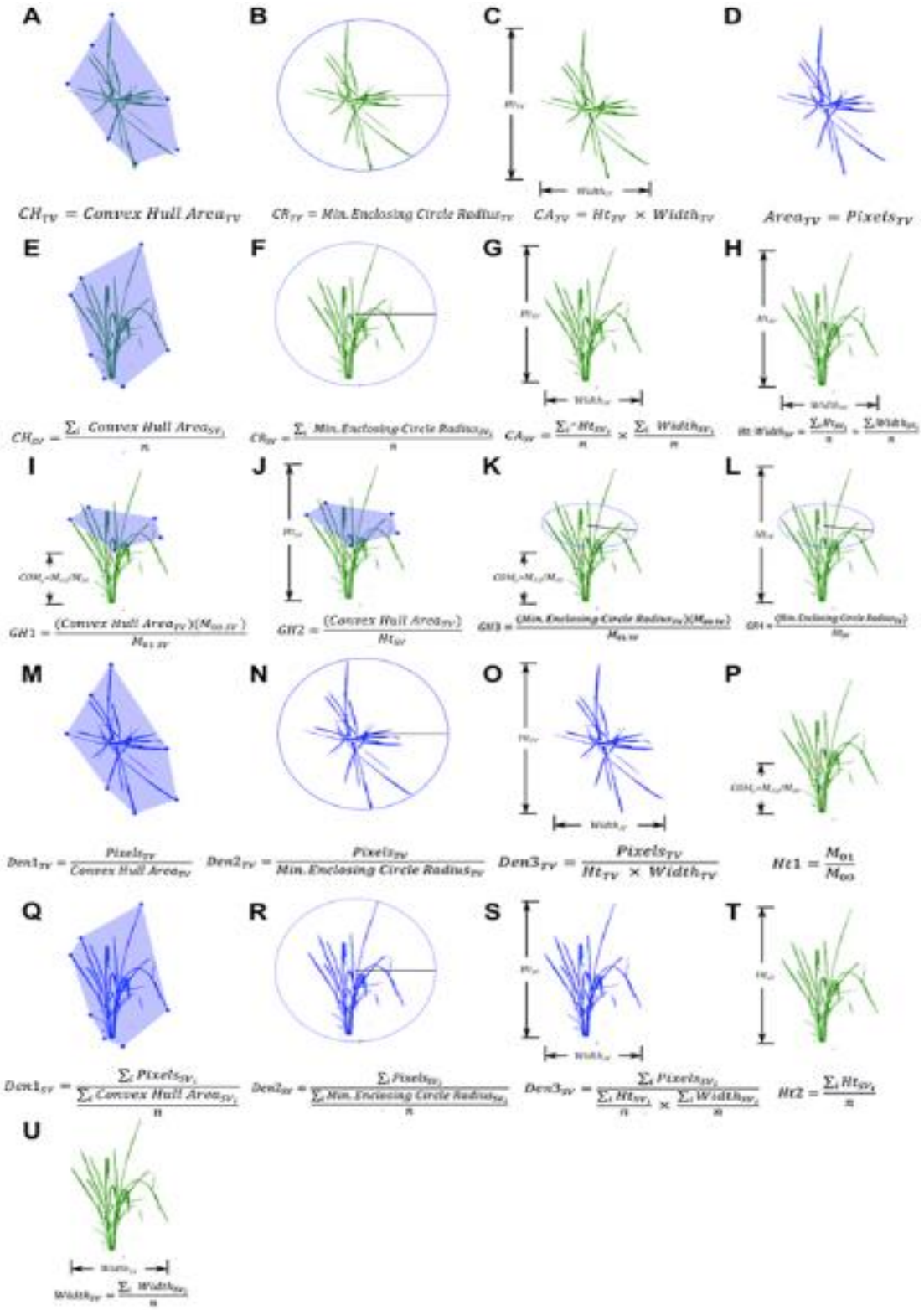


Fig 5.3 Visualization of 21 of the 22 digital traits derived from IH metrics (A-U). Convex hull area is indicated with the light blue shaded bounding polygon superimposed over portions of the plant. COM_y: Center of mass about the *y*-axis; *Ht*: Height; *CH*: Convex Hull Area; *CR*: Radius of minimum enclosing circle; *Den*: Density; *GH*: Growth Habit; *M*: raw moment; *n*: number of side view imaging perspectives. ***(Adapted from Campbel et al, 2015)(Appendix)***

5.2.7 Smoothing of spline curves for RGR

PSA or Projected Shoot Area data for the F_5 population was smoothed with cubic spline curve using the Image data package of R using the splitspline() function. And the smoothed AGR and RGR values were calculated. Profile plots were created using the longiplot() function using aesthetic options of ggplot2 package of R.

5.2.8 Correlation and Distribution analysis

Adjusted mean of the parents and the progenies were calculated using the lsmean package of R. The distribution plots and Pearson correlation calculations as well as plots were generated using the boxplot function of ggplot2 package, corr() and corrplot() function.

5.2.9 Mapping of QTL and Identification of associated genes

From the genotype map generated previously (Chapter 3), QTLs were mapped for the digital traits using the R/qtl package. A few significant SNP markers were identified and associated eQTLs were extracted from the eQTLs identified in Chapter 5.

5.3 Results:

5.3.1 Generation of digital traits and their distribution

Table 5.4 Adjusted mean of the traits and percent decrease under salinity in parents, F_3 , progenies and F_5 progenies.

Digital Traits	Unit	Horkuch C	Horkuch S	IR29 C	IR29 S	F_3-C	F_3-S	F_5C	F_5-S	HK (% decrease)	IR (% decrease)	F_3(% decrease)	F_5(%decrease)
PSA	*kp*	120.0	33.0	105.8	21.3	506.6	122.1	262.3	124.2	72.49	79.86	75.90	52.65
ht1	*p*	269.8	209.1	255.0	159.2	367.4	226.7	365.3	281.0	22.50	37.59	38.31	23.09
ht2	*p*	480.2	384.8	405.3	243.3	606.7	355.7	611.7	462.1	19.87	39.96	41.38	24.46
Width$_{av}$	*p*	421.5	221.3	246.6	230.3	782.8	397.7	572.6	417.5	47.51	6.61	49.19	27.08
Htbywd	*p*	1.3	1.7	1.8	1.1	0.8	1.0	1.2	1.2	-35.51	40.31	-18.19	1.39

CASV	*kp*	205.1	84.9	129.0	64.6	500.5	171.8	370.4	213.3	58.60	49.92	65.68	42.42
CATV	*kp*	970.6	348.6	521.8	19.7	3197.1	659.7	1717.0	808.2	64.09	92.41	79.36	52.93
CHSV	*kp*	125.0	51.8	84.8	42.7	337.3	97.3	259.8	145.5	58.57	49.64	71.16	43.98
CHTV	*kp*	461.6	126.1	306.4	3.5	2210.6	388.9	1065.0	459.0	72.69	93.84	82.41	56.90
CRSV	*p*	284.5	210.9	215.4	154.3	436.8	220.5	374.8	283.0	25.88	28.36	49.52	24.49
CRTV	*p*	402.6	427.5	449.2	133.2	1108.0	463.3	802.7	521.5	-6.18	70.34	58.18	35.04
DEN1sv	*p*	0.5	0.6	0.9	0.6	0.6	0.7	0.6	0.6	-9.08	25.82	-15.11	1.16
DEN2sv	*p*	0.1	0.0	0.2	0.4	0.1	0.2	0.1	0.1	58.82	-105.62	-9.71	-11.59
DEN3_SV	*p*	218.5	130.9	235.6	117.9	467.1	253.1	363.4	250.9	40.12	49.96	45.82	30.94
DEN1_TV	*p*	94.3	14.8	71.1	9.6	249.0	86.7	135.4	64.2	84.25	86.49	65.17	52.58
DEN2_TV	*p*	0.3	0.3	0.5	0.3	0.4	0.5	0.4	0.4	-3.98	31.67	-11.43	5.55
DEN3_TV	*p*	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1	71.94	-8.15	3.66	-4.19
GH1	*p*	2146.9	645.5	993.4	25.8	6053.0	1324.7	2834.5	1437.5	69.93	81.34	78.11	49.29
GH2	*p*	1246.4	357.2	625.0	16.4	3550.0	835.4	1657.3	841.1	71.34	81.65	76.47	49.25
GH3	*p*	2.1	2.1	2.1	1.0	3.1	2.1	2.2	1.8	0.53	51.38	33.68	19.82
GH4	*p*	1.2	1.2	1.3	0.6	1.9	1.3	1.3	1.1	2.07	50.14	32.07	19.46
TVpxl	*kp*	54.0	5.7	38.6	1.3	292.7	54.6	118.6	45.5	89.53	96.60	81.35	61.64

From the RGB images 22 digital traits have been extracted mostly from the height, width, top view radius and moment of the images. In the Horkuch parent which is salt tolerant, damage in growth due to salinity was less as depicted in 69.93% decrease of GH1 compared to IR29 parent which decreased 87.4% under salt stress (Table 5.4). GH1 is a plant shape related trait which is the ratio of the convex hull area of the TV image to the center of mass about the *y*-axis of all side view images. The data was an adjusted mean of replicates on the 11th day after application of salt. The percent decrease in GH1 due to salt stress was 78.11% and 49.29% in F_3 and F_5 population respectively. The projected shoot area (PSA) also showed similar trend (Table 5.4). Htbywd, DEN1sv, $DEN2_{SV}$, $DEN2_{TV}$ and $DEN3_{TV}$ are measurements of density and compactness in ratios and in the sensitive parents increased under salt stress

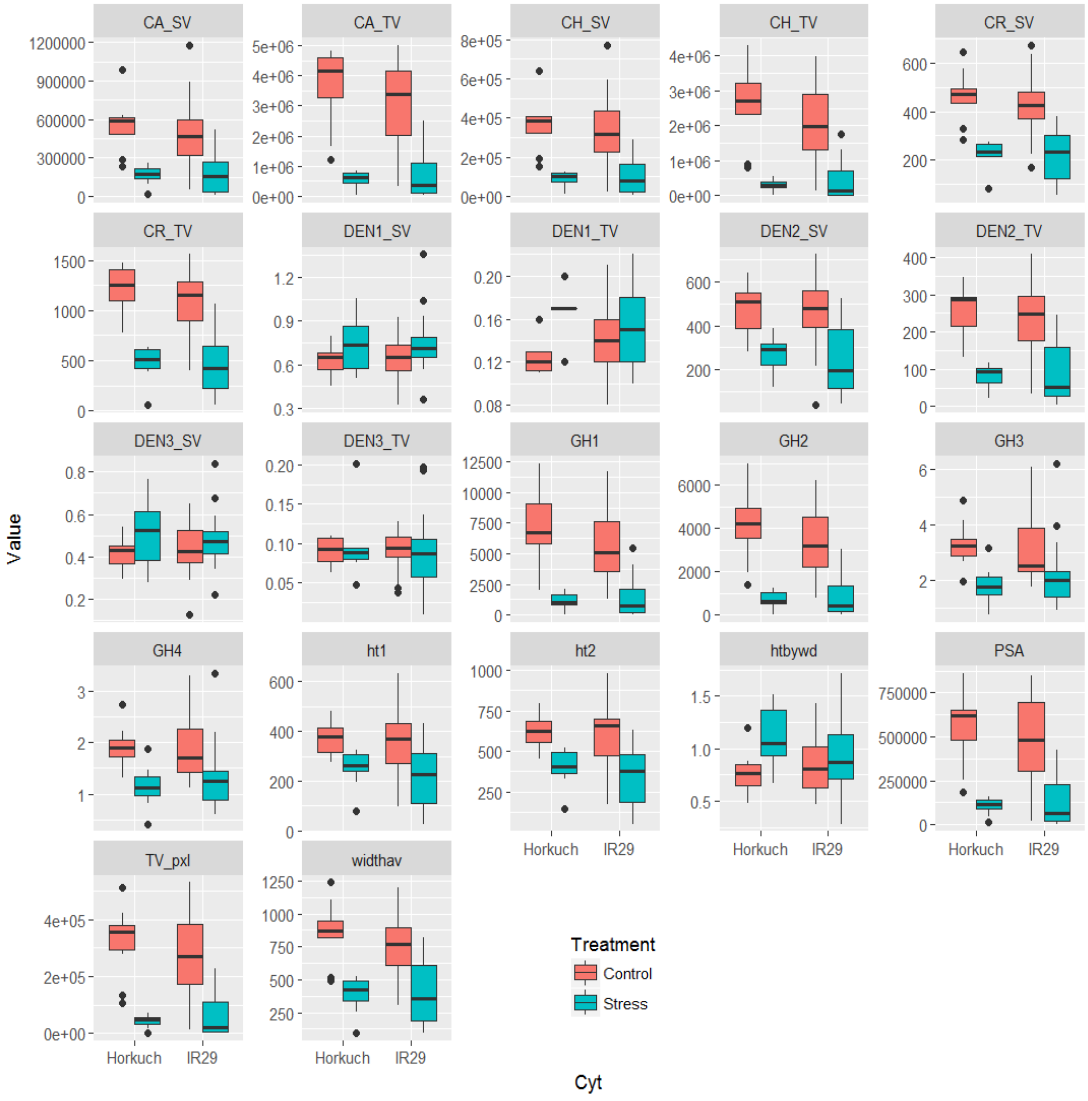


Fig 5.4: Distribution of Digital traits in F_3 progeny plants

The distribution of the digital traits for F_3 and F_5 progenies are given in Fig 5.4 and Fig 5.5 respectively which shows less damage of the traits in F_5 population as also shown in the Table 5.4. For PSA, the IR29♀ F_3 population showed greater range of minimum and maximum values compared to the Horkuch ♀ population.

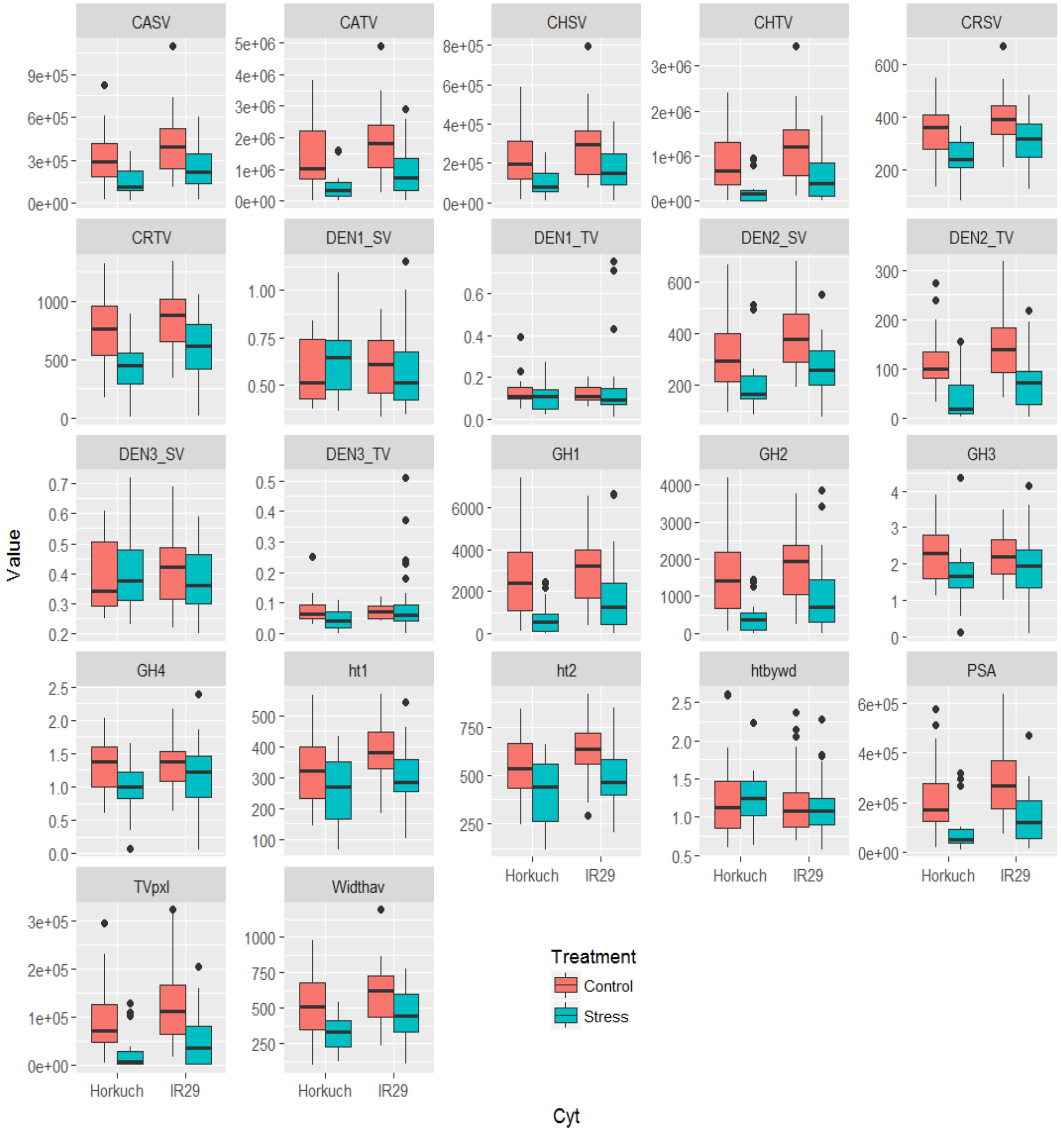


Fig 5.5: Distribution of Digital traits in F_5 progeny plants

5.3.2 Correlation of the digital traits with manual phenotyping traits

The digital traits showed good negative correlation with the salinity injury score of the plants (SES) ranging from (0.44 to 0.62). Since SES is a scoring of damage, high score means a sensitive plant. Only Height by width, $DEN1_{TV}$ and $DEN3_{TV}$ showed positive correlation with SES which indicates that these increase in tolerant plants under stress.

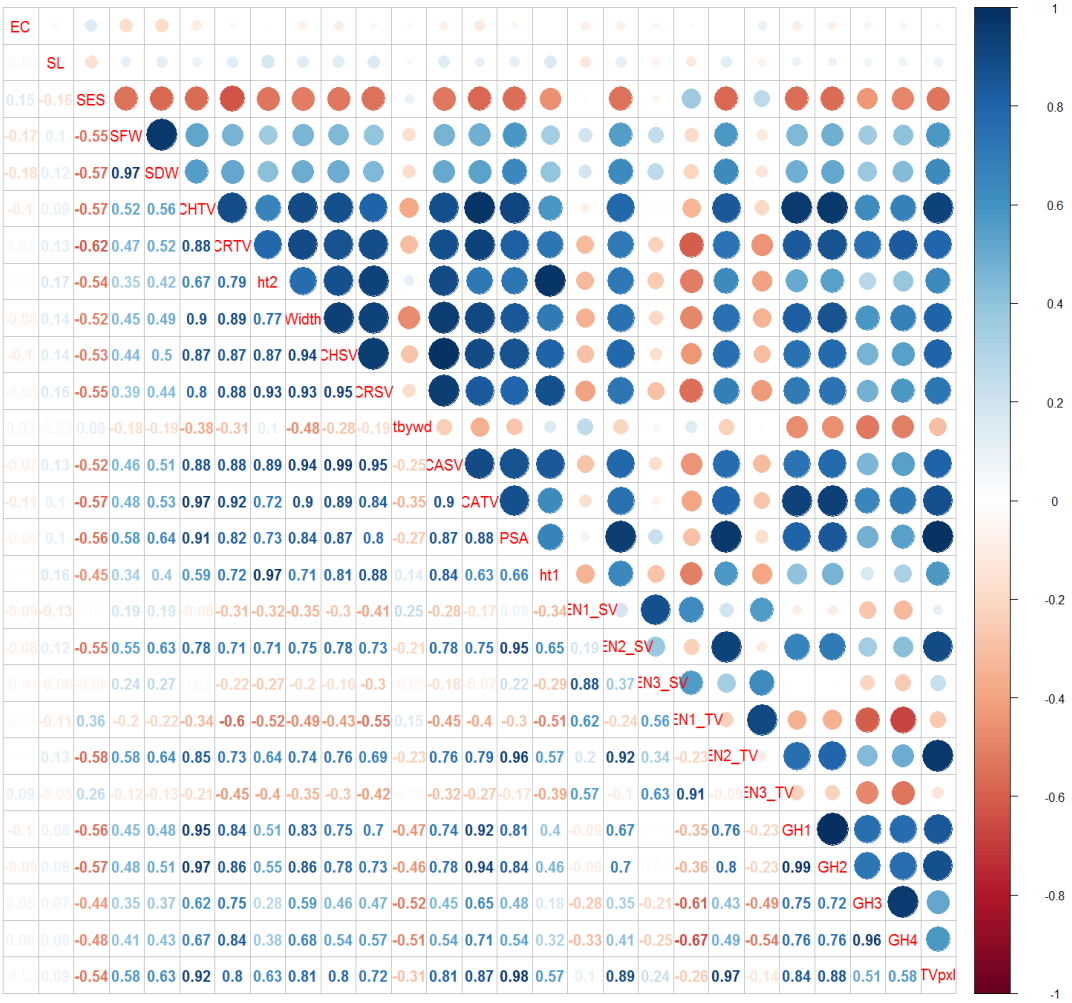


Fig 5.6: Distribution of Digital traits in F_3 progeny plants

Under control condition, Shoot Fresh weight (SFW), shoot dry weight (SDW) showed very good correlation with PSA (0.92, 0.95), $DEN2_{SV}$(0.78, 086)m $DEN2_{TV}$(0.81, 0.9) and TVpxl(0.9, 0.94). The biomass also showed good correlation with GH1 (0.57, 0.55) and GH2 (0.62, 0.59) respectively

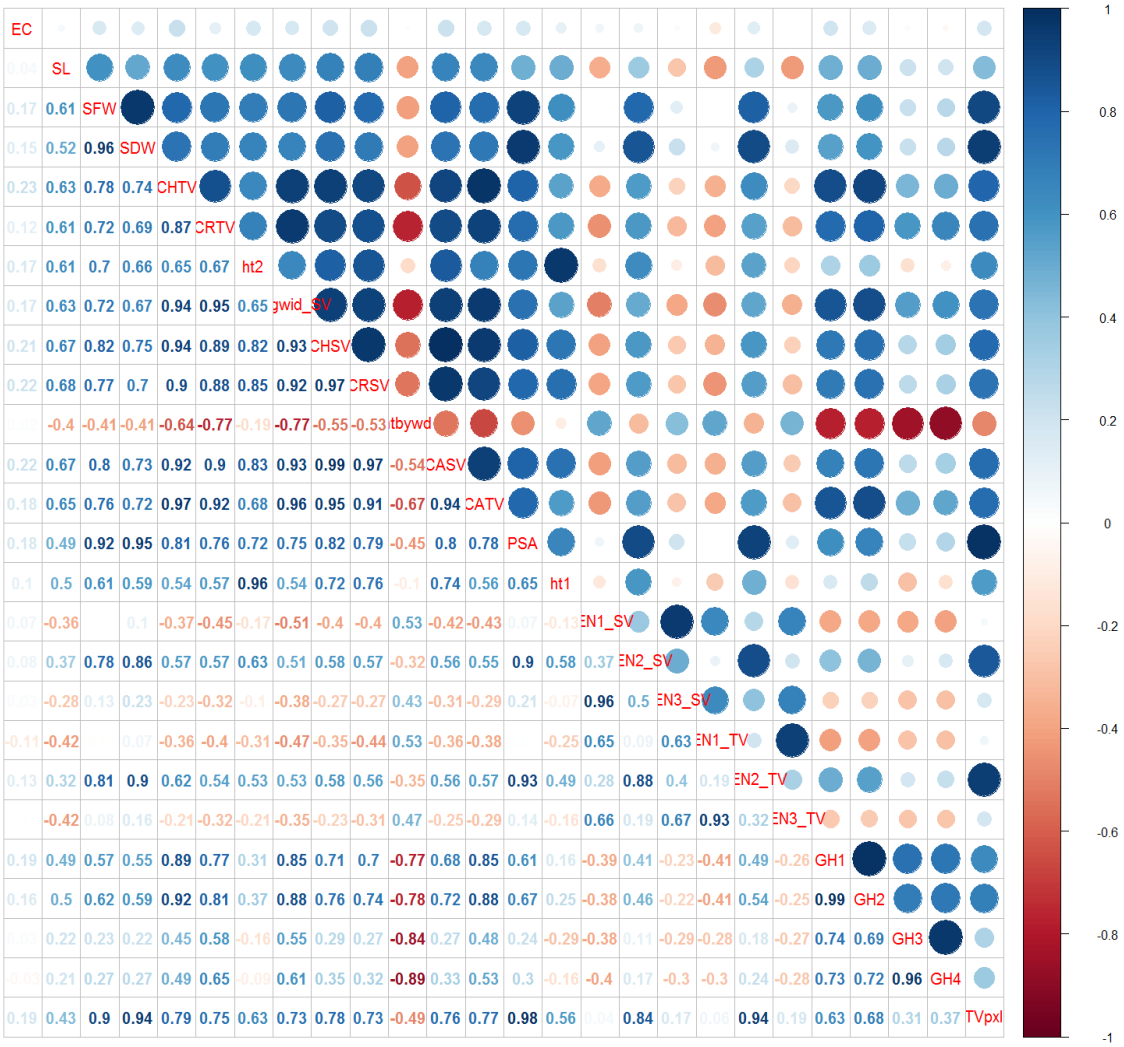


Fig 5.7: Distribution of Digital traits in F_5 progeny plants

A good negative correlation was also observed between htbywidth and GH1(-0.77), GH2(-0/78), GH3(-/0.84) and GH4(-0.89). So PSA can be a proxy of digital biomass of the plants and can give important insights about the growth under stress condition. The distribution of one digital trait PSA is shown in Fig 5.5 to visualize the growth trend over number of days. From the 9th day of imaging which is the 1st day after addition of salt the trend for salt treated plants showed reduction in growth rate over time.

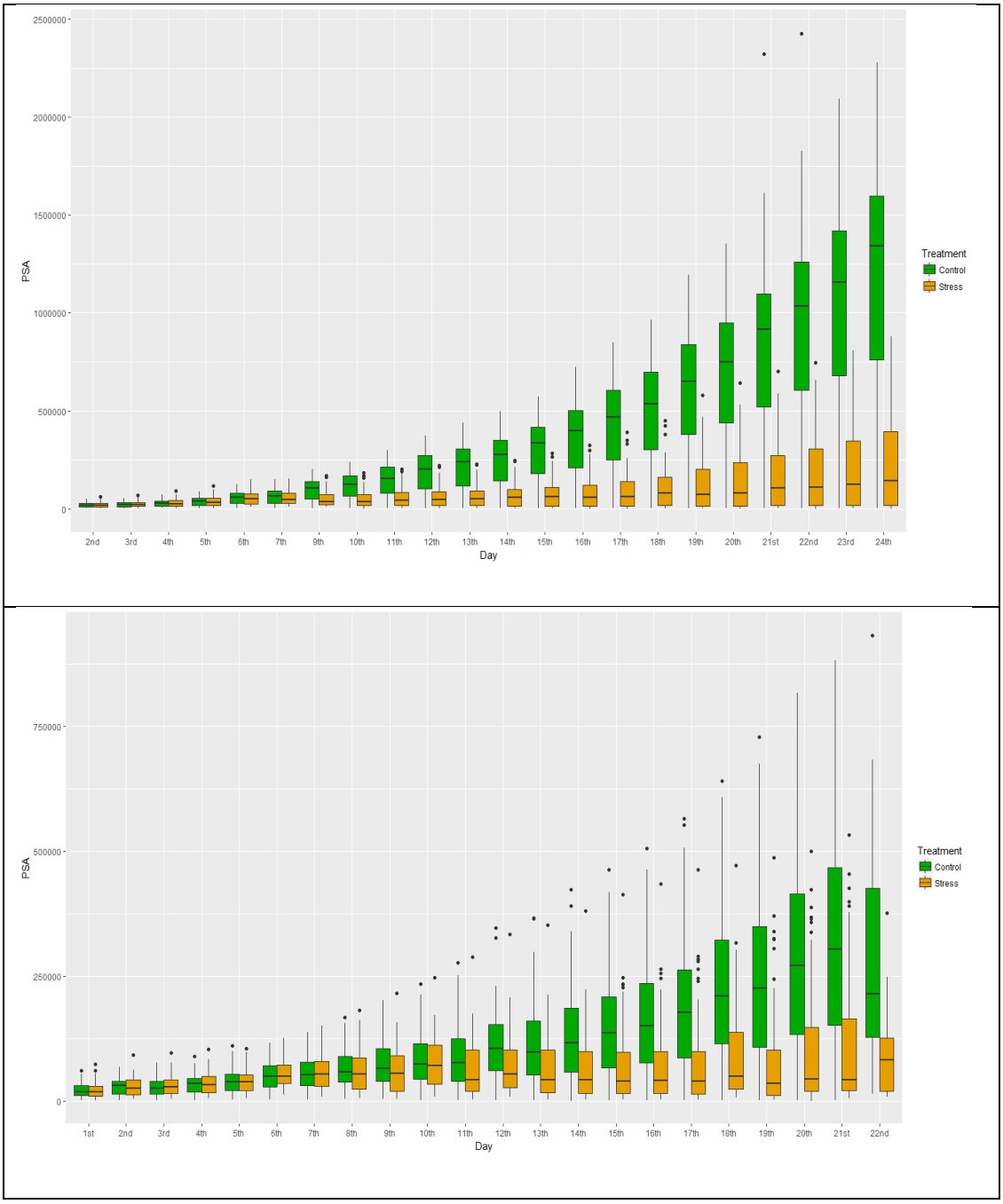


Fig 5.8 Longitudinal data for PSA over time, (green = control and brown = stress)

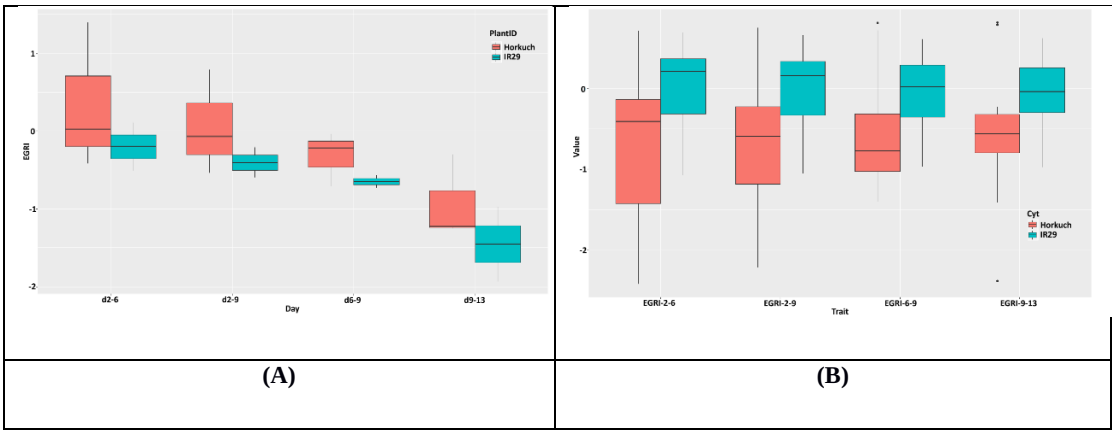
5.3.3 Cubic spline smoothing of Relative Growth Rate gives a better dataset

To visualize the growth rate more properly the relative growth rate was measured for each individual, except for the 1st day. Relative growth rate is expected to be decreased in salt treated plant. From the projected shoot area the absolute growth per day (AGR), relative growth rate (RGR) was measured under control and stressed condition. Relative growth rate in plants show exponential trend so RGR was measured using the difference of natural logarithm function ln(x)

of PSA between two days. As ln(x) is the inverse of exponential function, it explain the growth properly. But as explained by Al-Tamimi et al (2016), the growth rate in the population does not always show a proper exponential growth, so according to their suggestion cubic spline was fitted to the data to smoothen it. .The PSA, AGR and RGR curves, before and after the spline smoothing are shown in fig 5.6 showing the advantage of the strategy to explain plant growth rate more accurately. It shows that the relative growth rate shows greater reduction over time under salt stress compared to the control.

5.3.4 Correlation between growth rate indices

Four intervals were chosen following the strategy used in Al-Tamimi et al (2016) as 2-6, 6-9, 1-9 and 9-13 days interval. RGR declines less rapidly under control and saline conditions during 6-9 days after treatment. During 9-13 days interval AGR increases less in control compared with salt treated plants. Six more indices were derived from the RGR data to cumulatively represent the control and stress ratio data. These are EGRI (Early growth related index), STI (Stress tolerance index), SWP (Stress weighted performance), TOL(Stress Tolerance), SSI (Stress susceptibility index). The early growth response index (EGRI) gives an estimate of early responses of plants to salinity. It has been reported that the tolerant variety Pokkali also maintain a higher EGRI compared to sensitive IR28 (Al-Tamimi et al, 2016). For the earlier intervals, Horkuch parent showed better EGRI than IR29 and with the accumulation of salt in the plant tissues, in the later intervals EGRI decreased in lower rate compared to IR29 (Fig 5.10A). It indicates that Hokuch might possess a greater tolerance to stress in earlier osmotic phase also. EGRI study in the F_5 population revealed that IR♀ population maintain a better EGRI than the Horkuch mother population in all four intervals.



(A) **(B)**

Fig 5.10 EGRI profile of (A) Horkuch and IR29 parents and (B) F_5 population Horkuch♀ and IR♀ population in four intervals

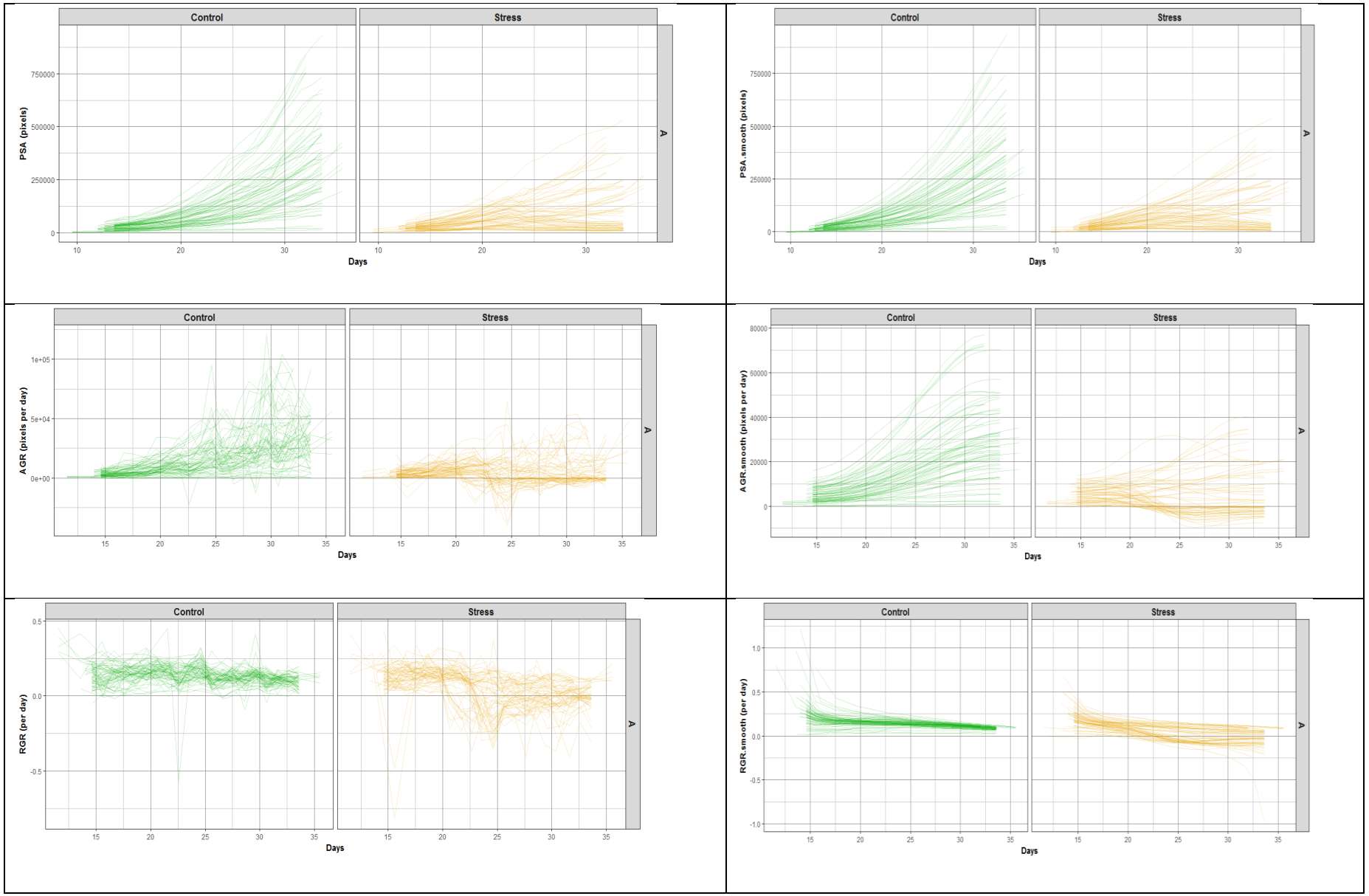


Fig 5.9 : Projected shoot Area, AGR and RGR data showing before smoothing in the left panel and after smoothing in the right panel.

According to the SES (Standard Evaluation System) scores the tolerant and sensitive plants of a subset was taken and their RGR in 4 interval days showed less reduction in tolerant plants under stress as shown in Fig 5.11.

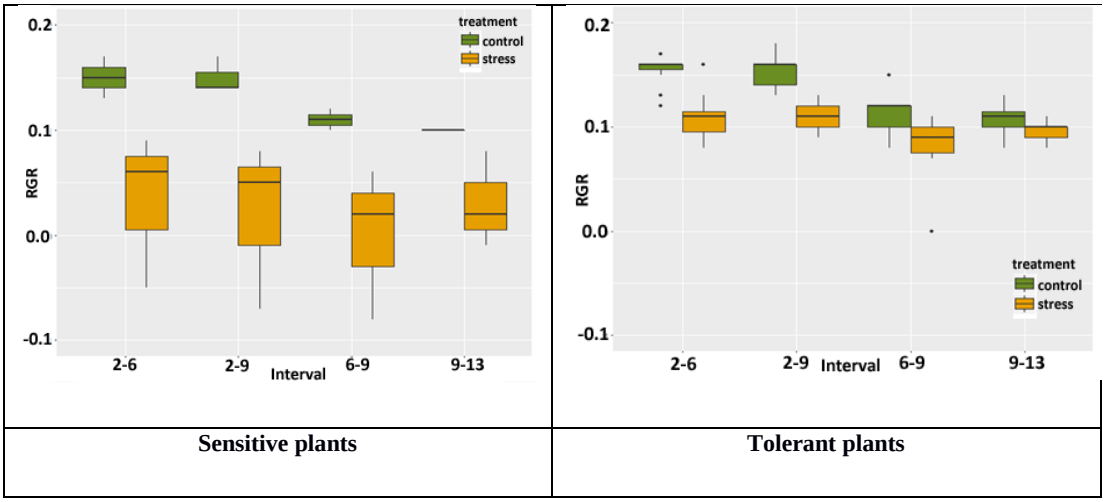


Fig 5.11 RGR profile of selected tolerant and sensitive plants

Pearson correlation analysis of the RGR derived traits showed that TOL is negatively associated with EGRI and other indices. Stress susceptibility index during 2 to 6 day interval is negatively correlated with EGRI, meaning in the earliest interval increase in EGRI decreases the susceptibility index, but in later stages SSI increases with increase of EGRI.

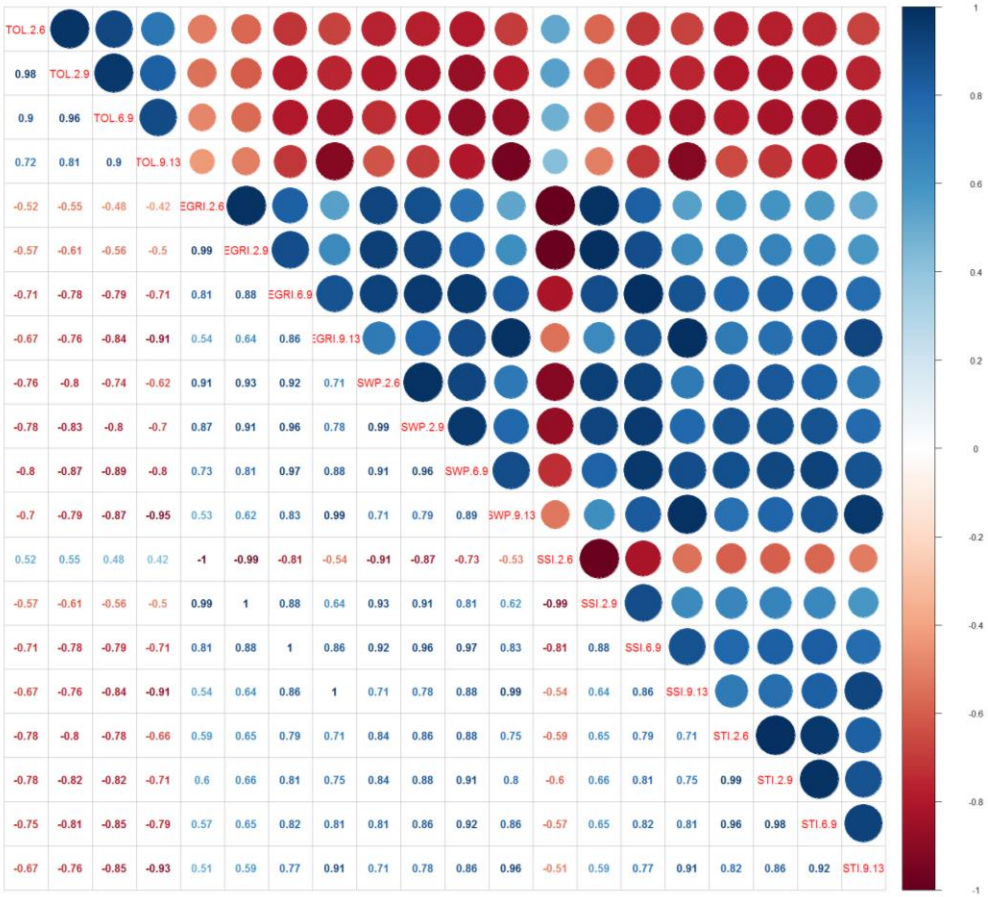


Fig 5.12 Correlation plot between the RGR derived growth rate indices

5.3.5 Mapping QTL and identification of candidate gene

The size and shape related digital traits and the RGR growth related traits were combined with the genotyping data obtained (Chapter 3) to map QTLs. Although, lower number of plants were used for image analysis compared to the QTL map described previously from the manual phenotyping, it gave signals for same region in the genome, some of which with above 3.5 LOD threshold are noted in Table 5.5. For the plant size related traits, on the 18th day of salinity 3 significant markers in chromosome 1, comprising ~43.1 Mb were found to be associated with CHTV, CATV, PSA, GH1, GH2, TVpxl in control condition plants, to be specific GH1 (*Growth Habit 1*) was found to be associated with the SNP marker DART_1_43100556. In Gramene.org rice QTL database qtls for vigor, spikelet fertility and yield was found near the location 42.95 Mb.

Functional analysis of the associated trans-eQTLs (identified from chapter 4) could associate multiple networks related with cation transport, including two members of *OsHKT2,3* and *OsHKT2,4*. Interestingly *OsHKT2,3* gave signal in 24 hour tissue and *OsHKT2,4* in 72 hour

tissue in control condition though they have about 93% sequence similarity, indicating variation of the expression level of these genes in the population. Several networks of genes were formed involving morphogenesis, like pollen development, formation of patterns for xylem, phloem, carbohydrate metabolism, cell wall organization, vegetative to reproductive phase transition of meristem and shoot development, growth hormone stimulus, stomatal movement etc., when plugged into the Rice net functional network database.

Under stress condition for vigor related traits GH1, CATV and CHTV in chromosome 10, ~21.9Mb, a region was identified which is nearby the regions identified for spikelet fertility, filled grain weight in the physiological QTL mapping (Chapter 3).

Analysis of the growth related traits under stress conditions identified a signal for marker DART_3_561226 with multiple traits in the early stage of salinity, to be specific , for RGRs, TOL, EGRI, SSI in the 2-6, 6-9, and 2-9 interval but not in the 9-13 interval indicating role of this region in early growth stages. Mainly cation transport, chloride transport, response to ABA stimulus, response to salt stress, response to water deprivation etc. gene networks were formed. For STI in early stage genes enriched with response to stress, ROS, auxin stimulus response to osmotic stress, desiccation, potassium ion transport, stomatal movement, salt stress, abscisic acid stimulus, response to calcium ion were found. For 9-13 interval two good signals were found for EGRI and SSI involved in carbohydrate metabolism, cell division, response to salt stress, oxidative stress. Based on the functional analysis few important candidate genes are proposed in table 5.5

Table 5.5 List of selected candidate genes identified from association between image QTL regions and expression QTL regions

Trait Category	Marker Position	Functional network	Selected genes
Vigor in control stage, GH1, CATV, CHTV, GH2, PSA	DART_1_43100556 Chr 1, ~43.1Mb	morphogenesis, like pollen development, formation of patterns for xylem, phloem, carbohydrate metabolism, cell wall organization, vegetative to reproductive phase transition of meristem and shoot development, growth hormone stimulus, stomatal movement etc.	LOC_Os01g34850 LOC_Os06g48800 LOC_Os08g44590, LOC_Os09g28650
GH1, CATV, CHTV under salt stress	Chr 10 ~ 21.9 Mb	Not investigated	

Early growth related 2-9 days RGRs, TOL, EGRI, SSI	DART_3_561226	cation transporter, chloride transport, response to ABA stimulus, response to salt stress, response to water deprivation etc. gene network was formed	LOC_Os03g03020 LOC_Os03g14520 LOC_Os03g14690
Early growth related STI 2-9	DART_5_17378471	Response to stress, ROS, auxin stimulus.	LOC_Os04g47270 LOC_Os09g07350
Early growth related RGRs, TOL, EGRI, SSI	DART_5_17365097	Response to osmotic stress, dessication, potassium ion transport, stomatal movement, salt stress , abscisic acid stimulus, response to calcium ion	LOC_Os01g69830 LOC_Os01g72530 LOC_Os03g08230 LOC_Os05g03130 LOC_Os05g28190 LOC_Os05g35410 LOC_Os07g01810 LOC_Os07g26630 LOC_Os11g02240
EGRI and SSI 9-13 interval	DART_1_33731460	Carbohydrate metabolism, cell division, response to salt stress, oxidative stress auxin stimulus, post embryonic organ development, translation	LOC_Os01g49190 LOC_Os02g49160 LOC_Os01g59840 LOC_Os01g56800 LOC_Os03g38000

5.4 Discussion

This chapter describes a comprehensive study on the same breeding line using an emerging technique called High throughput phenotyping. Both the F_3 and F_5 populations were studied under this system in two subsequent years. The most important advantage of this phenotyping is, it gives longitudinal data rather than only the endpoint. In this way the day to day growth rate could be observed in the plants and important regions could be identified in downstream analysis with the data generated from previous chapters. Moreover the digital traits which were generated mainly using convex hull, projected shoot area, width height and moment of the images, have proved to be a good means to study the vigor, growth shape and architecture of the plants.

As reported in earlier studies, digital trait PSA showed very good correlation with manually measured shoot biomass. Distribution of the digital traits in F_3 and F_5 population showed a greater variation in F_3 and lesser in F_5 as most of the F_5 plants were selected based on physiological QTL maps, though other random plants were included in the F_5 also. In the distribution boxplots *Horkuch* and *IR29* mother population did not differentiate much except for PSA. So for studying the PSA later cytoplasm can be used as covariate.

For growth rate measurement the data were smoothed using cubic spline smoothing method, which improved the data quality and then 5 relative growth related traits were derived associated with early growth and stress indices. *Horkuch* as a tolerant plant showed better EGRI than *IR29*, but in the populations the *IR29* mother population showed a greater variation, indicating cytoplasmic effect.

QTL identification using the genotype data from chapter 3 and downstream functional analysis using the eQTL data from chapter 5 identified six important regions in chr 1, 3, 5 and 10. Interestingly the region in chromosome 10 resides very close to the physiological QTLs identified for grain weight. Also the region responsible for vigor related traits like GH1, PSA under control condition, resides near the end of chromosome 1 is very close to QTLs identified for early vigor reported in literature. Under stress, QTL for GH1 was identified with strong effect on a trans- eQTL for gibberellin 20 oxidase important in growth as supported by Campbel et al (2017). The relative growth rate related indices were analyzed in 4 overlapping intervals and could pinpoint a difference in the early and late stage after salt application. 3 QTL regions were identified for early growth indices in the 2-9 days interval after stress, which mainly composed of transport, osmotic response etc. and the later stage QTL was enriched with genes associated with growth, carbohydrate metabolism, organ development etc.

Image based phenotyping is a more accurate and informative phenotyping system. Due to temporal and spatial limitation thorough analysis of the both populations was beyond the scope of this thesis. To the author's knowledge, this is the first study where imaged-based analysis was used to identify candidate genes instead of GWAS support, but by using an eQTL approach. The main limitation of this study was the number of plants which were not enough for a QTL study, but still the QTLs were mapped to scan important signals. It's possible that some important regions might be missed due to this limitation, but the signals appeared to be promising in the downstream analyses. The fluorescence images generated should give more information regarding the photosynthetic mechanism in *Horkuch* in future studies.

In a nutshell this study could pinpoint some important candidate genes separating early and late response as well as genes representing the growth dynamics. Also it could validate QTL region identified from manual phenotyping.

Chapter 6

Variant analysis and allele mining

6.1 Overview

For modern crop science, an important tool to improve crops is knowledge about nucleotide sequence variation so that it can be related to differences in traits of interest. This variation can either be natural, from divergent populations, or induced through treatment with mutagens. Addition of a new allele to a population may cause that to become more competent to survive, or sometimes less capable. The variation can be insertions or deletions (together called InDels) or single nucleotide polymorphism (SNP). Allele mining is a promising way to dissect naturally occurring allelic variants of candidate genes and link it with essential agronomic qualities in different varieties. This allows rapid identification and characterization of allelic variants (Latha et al. 2004).

In this chapter, the cis expression QTLs identified from the DArTseq based genotyping data of the IR29♀ population and reproductive stage RNAseq expression data (both samples without stress and 72hr salt stress at 150mM NaCl) was used. These narrowed down specific regions of the genome which were then used in the study to analyze and validate target regions. The analysis was supported by whole genome resequencing data of *Horkuch* and *IR29* which allowed variant analysis and allele mining. Genomes of representative salt tolerant donor and recipient lines were also included in the study for a comparative analysis. Moreover few of the candidate genes listed as the output of candidate gene search in chapter 4 have been analyzed for significant variants if any.

Output from this study will give insights on the following:

1. Validate the specific SNP markers associated with the eQTLs and their positions in the genome.
2. Variant analysis among the specific genes and their regulatory regions identified in the eQTL analysis can provide a basis for their differential expression pattern under stress.
3. Apart from specific regions, the data generated has immense potential to draw hypothesis about common mechanism of tolerance among the salt tolerant donor and recipient varieties and can group them accordingly.

6.2 Method and Materials

6.2.1 Expression QTL analysis

Expression variation data of reproductive stage for F_3 progenies of IR♀ population were analyzed for control stage (38 replicates) and 72 hour stress stage (39 replicates) with the DArTseq genotyping data. It should be noted here that the reciprocal cross population data is

not used here, but the linkage map (shared from another PhD fellow's work) constructed from only *IR*♀ has been used. eQTLs for control and stressed samples were identified in separate runs using matrix eQTL package of R (Shabalin 2012) using the linear model (*detail of method can also be found in chapter 4*). Apart from these the differentially expressed cis and trans eQTLs were also identified. The specific regions of the genome harboring these eQTLs have been used in the analysis as subsets.

6.2.2 Whole Genome Sequences

The following whole genome sequences were used in the study for comparison purpose. The origin of the sequences are listed in Table 6.1. The genome cultivars were chosen based on their tolerance to salt i.e. salt tolerant, sensitive and moderately tolerant rice cultivars.

Table 6.1: List of the genomes used in the study

Cultivar Name	Accession	Stress tolerance	Origin of sequence
Capsule	N/A	Salt tolerant	IRRI
Damodar	IRGC 17038-1	Salt tolerant	3k genome database
FL478	N/A	Salt tolerant	IRRI
FL478CX219	CX219	Salt tolerant	IRRI
Horkuch	IRGC 31804	Salt tolerant	PBTLAB, DU
Hasawi	N/A	Salt tolerant	IRRI
IR29	IRGC 30412	Salt sensitive	PBTLAB, DU
Ir29	N/A	Salt sensitive	IRRI
IR64	N/A	Moderately salt tolerant	IRRI
Kasalath	Cx227	Tolerant to drought and phosphate deficiency	IRRI
Nonabokra	IRGC 22710-C1	Salt tolerant	3K genome database
Pokkali8948	IRGC 8948-1	Salt tolerant	IRRI
Pokkali26869	N/A	Salt tolerant	IRRI
SR26B	IRGC 10803-2	Salt tolerant	IRRI

6.2.3 Filtering the raw reads

Raw reads for Nonabokra and Damodar were downloaded from 3Kgenome ftp site (Li et al. 2014). *Horkuch* and *IR29* genome were re-sequenced (Unpublished data). All were sequenced in illumina hiseq 2500. *Horkuch* and *IR29* genomes were sequenced using 4 libraries of 180 bp, 500 bp, 3000 bp and 10000 bp. They were filtered for base quality of more than Q20 and true paired sequences and singletons were separated before mapping. Read lengths were 150 bp.

The genome sequences provided by IRRI were already in the Binary Alignment Map (BAM) format. These were mapped to the MSU 7 rice reference genome. The sequences were filtered for base quality of more than Q20. The paired reads were analyzed to filter the true paired reads from single reads (without proper pair). Read length was of variable size and these were provided as sorted bam files.

6.2.4 Mapping to the reference and post-processing

Reads of *Horkuch, IR29, Damodar and Nonabokra* genome were mapped to the same customized version of the reference Nipponbare genome MSU7.0 as IRRI provided BAMs. As IRRI provided sequences were already in BAM format they did not need the upstream analysis before mapping. Burrows wheeler Aligner (BWA)(Li and Durbin 2009) was used to map the reads. Readgroups were added to the filtered raw reads so that the duplicate reads can be identified and the sample names can be easily distinguished in the downstream files. Sequence Alignment Map (SAM) files were generated for each library and lane separately to track the readgroups and mark the duplicates. Then the SAM files were sorted according to the coordinate using the picard tool (Li et al. 2009). Duplicate reads were marked using the picard mark duplicate tool.

The individual SAM files were converted to BAM using the samtools (Li et al. 2009) followed by merging to get one binary alignment map (BAM) per genotype. The BAM files were filtered again for low quality mapping reads. Some of the reads with different quality encoding were fixed with Genome Analysis Toolkits.

For the IRRI-provided BAM files, the readgroups were added later using AddorReplaceReadgroup tool of GATK (McKenna et al. 2010), and duplicate reads were marked in the BAM files. Some of the BAM files had Un and Sy coordinates which were not present in the customized reference file. Those maps were removed using GATK countloci tool, samtools and bash scripts to keep consistency among the genomes. Statistics from all the BAM

files were generated to get idea about the coverage, mapping percentage, etc. and was visualized using qualimap_v2.2.1 (Okonechnikov et al. 2016).

6.2.5 Variant Calling

For variant calling, a modified version of the GATK best practice (McKenna et al. 2010) was applied (Fig 1). The duplicate marked, sorted and map-filtered bam files were used for variant calling using HaplotypeCaller of GATK. To avoid false positive and false negative variants, three methods were adopted to generate variant call files to assess the one with the best call score (Fig 6.1). Since GATK best practice suggested the use of known site variant-call files to recalibrate the base quality scores before calling haplotypes, the following three different methods were used.

Method 1: the original sorted deduped and map-filtered bam files were passed through BQSR where *Oryza sativa* available variant file was used as known sites. The pre- and post-recalibration tables were compared for their performance. Once this was done with satisfaction, the recalibration table was applied on the bam files and a new set of bam files were generated with the recalibrated base quality score. If the process was un-satisfactory then another recalibration was performed. These were used to generate individual gvcfs followed by a combined vcf (variant calling format) file.

Method 2: The original sorted deduped and map-filtered bam files were directly used to generate gvcf file using haplotype caller of gatk without base quality recalibration. The separated gvcf files were then used to generate individual vcfs and a combined vcf file. The combined vcf files were used as the output of method 2.

Method 3: The individual vcfs generated from method 2 were passed through a hard filtering value for SNPs and INDELs and a second set of individual vcf files were generated containing variants which only could pass the hard filter. The hard filter used for SNP was QD < 5.0 || FS > 10.0 || SOR>3.0 || MQ < 50.0 || MQRankSum < -5.0 || ReadPosRankSum < -4.0 . For INDELs, the filter criteria was "QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0"

The original sorted deduped and map-filtered bam files were then passed through BQSR recalibration tool of GATK, where the hard filtered vcf files from method 2 were used as the known sites. When satisfied, this recalibration table of base quality was applied to the bam files and a recalibrated set of bam files were generated using the Printread tool of GATK. Finally from these bam files individual gvcf files followed by separate vcf files and a combined vcf file was generated (Fig 6.1)

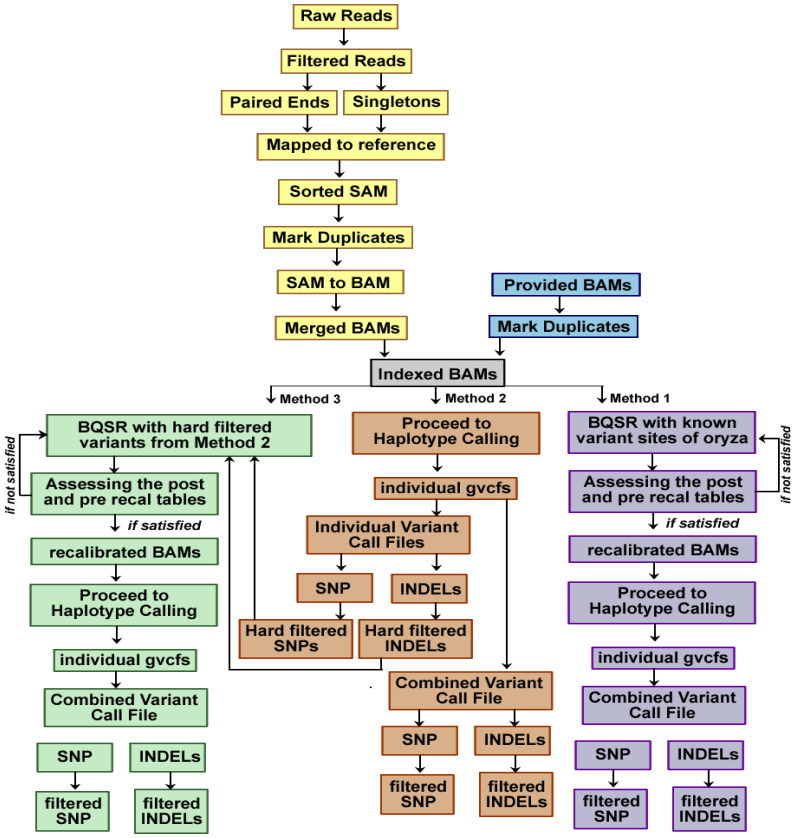


Fig 6.1: Schematic diagram of variant discovery method applied in the study

6.2.6 Variant Filtration

The final combined vcf files from Method 1, 2 and 3 were later split into INDEL and SNP files using the select variant option of GATK, to apply a low filtration criteria. For low filtering the filter applied was

For SNP -- QD < 2.0 || FS > 60.0 || SOR > 4.0 ||MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"

For INDEL -- "QD < 2.0 || FS > 200.0 || SOR >10.0 || ReadPosRankSum < -20.0"

6.2.7 Variant annotation

Variant annotation was done using the snpeff tool (Cingolani 2012), where the *Oryza sativa* public database was used to generate variant statistics, chromosome-wise distribution and annotation of genic and regulatory regions.

6.2.8 Generation of subsets and cluster analysis

Subsets of the variant files were generated for individual chromosomes, identified cis_eQTL regions (referred to as *cis_dart_SNP* and *cis_dart_genes* in subsequent analysis). Using SNPphylo (Lee et al. 2014) clusters were generated from the variant files. For a comparison, variants across the whole genome, variants in the *cis_dart_SNP, cis_dart_genes* and variants in individual chromosomes were clustered.

6.2.9 Allele mining and validation of SNP marker positions

For allele mining, the specific regions of the genomes were extracted from the variant analysis output file using snpSift (Cingolani 2012) and vcftools (Danecek et al. 2011) followed by close visualization using Integrative genomics viewer (IGV) (Robinson et al. 2011). Firstly, the SNP marker positions provided from DArTseq study were validated and their polymorphic nature between *Horkuch* and *IR29* were confirmed. Secondly, the regions of the genes affected by these markers as hypothesized from the eQTL analysis was also extracted and variants were listed for SNP marker within a coding or regulatory region of any gene. For the genic regions a subset i.e. only identified genes from chromosome 10 were chosen and Indels, homozygous SNPs and heterozygous SNPs with significant variant effect were precisely analyzed. Chromosome 10 was chosen because maximum number of QTLs for yield were observed in Chromosome 10.

6.3 Results

6.3.1 Expression QTL (eQTL) regions for allele mining

Analysis of the F_2 SNP genotyping and reproductive stage RNAseq expression data could be narrowed down and associated with some expression QTL regions in the genome. Positions of the SNP markers and their associated differentially expressed genes were selected for validation. The genomic regions of cis-eQTLs observed from control and environmental effect were chosen (Fig 6.2). Moreover if the SNP marker itself was located in the genic region, the corresponding gene was also considered for analysis. Few genes from the candidate list generated in chapter 4 was included for variant analysis.

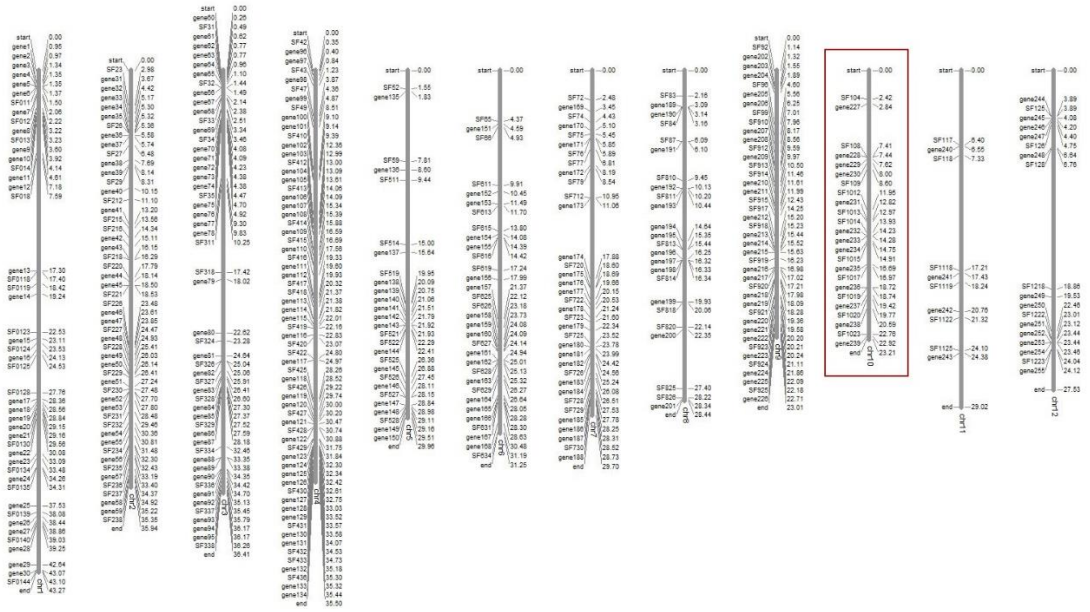


Fig 6.2: Positions of the cis-eQTLs regions chosen for further validation (high resolution figure in Appendix)

6.3.2 Mapping statistics and sequence depth

The mapping statistics for the final BAM files used are stated in Table 6.2. In the analysis of provided BAM files, the duplication rate was higher and therefore was marked in subsequent processing

Table 6.2: Mapping statistics of the BAM files

Sample	Coverage	SD	No. of Reads	Mapped Reads (%)	Mapped reads both in pair (%)	Mapped reads in singletons (%)	Min read length	Max read length	mean read length	Duplicate rate (%)	Mean MQ
Capsule	12.26	37.65	48042868	97.55	95.95	0.59	30	101	99.6	0.72	48.17
Damodar	11.95	11.7	54499671	100	98.94	0.86	30	83	82.9	1.96	53.05
FL478	29.63	80.77	122145934	70.18	66.26	0.48	30	151	122.7	2.02	48.17
FL478CX219	15.49	22.25	69988315	100	98.48	1.52	83	83	83	4.79	50.55
Horkuch	201.3	331.04	649941847	100	92.49	0.56	20	151	118.7	1.45	54.27
Hasawi	62.72	150.55	244512820	75.44	70.89	0.52	30	151	123	4.45	48.05

IR29	202.1	389.4	657503338	100	91.66	0.70	20	151	118.2	14.9	53.49
Ir29 (2)	13.39	37.76	53641207	95.91	95.22	0	30	101	98.6	9.0	41.5
IR64	38.98	100.99	205343426	94.84	94.10	0.57	30	76	75.7	2.34	45.09
Kasalath	94.44	391.94	371894186	97.44	88.69	0.30	30	102	98.8	7.94	45.96
Nonabokra	9.09	16.95	41408506	100	98.95	0.85	30	82	82.9	2.39	52.84
Pokkali 29869	26.89	81.74	81568777	98.06	92.38	0.5	30	151	28.9	2.73	47.51
Pokkali 8948	9.06	12.08	40953538	100	98.69	1.30	83	83	83	8.37	50.54
SR26B	7.57	12.22	34198782	100	98.67	1.32	83	83	83	7.49	50.15

6.3.3 Comparison among the three methods

A comparative analysis was performed to choose the best output from the three methods (Table 6.3a and 6.3b). Number of variants along with the view in IGV showed that the Base quality recalibration did not greatly affect the output. However, using only the known sites may result in missing of rare variants, so the output of Method 3 was adopted in further analysis.

Table 6.3a: Number of variants found in the combined haplotype calling of 14 genomes

Parameters	Method1	Method2	Method3
No. of variants before filter	7,456,962	8,303,203	8,183,627
No. of variants after filter	7,361,462	8,188,159	8,072,781
No. of effect	27,218,111	29,942,189	29,571,877
SNP	5,895,947	6.636,704	6539193
INS	715,981	759,267	753262
DEL	749,534	792,188	780326
Ts/Tv	2.53	2.54	2.54

Table 6.3b: Number of variants found in the combined haplotype calling of 14 genomes after filtering

Sample	SNP			INDEL		
	Method 1	Method 2	Method3	Method 1	Method 2	Method3
Capsule	2,663,561	2,633,289	2,659,364	514,116	510,934	512,692
Damodar	2,396,075	2,361,251	2,393,705	471,825	465,062	471,219
FL478	2,692,650	2,678,978	2,692,032	522,500	525,369	521,722
FL478CX219	1,716,789	1,696,365	1,714,511	427,140	421,648	425,973
Horkuch	3,520,854	3,625,873	3,498,812	677,693	690,469	669,899
Hasawi	3,054,668	3,093,385	3,053,272	601,079	610,177	599,878
IR29	3,142,016	3,162,461	3,119,187	654,048	657,495	650,309
Ir29	1,832,315	1,774,013	1,829,891	334,507	328,626	333,482
IR64	2,354,259	2,458,424	2,458,838	427,140	528,523	528,413
Kasalath	2,747,437	2,757,921	2,745,511	593,919	595,670	592,303
Nonabokra	2,400,099	2,360,251	2,398,127	449,443	442,758	448,154
Pokkali8948	1,682,330	1,649,812	1,678,549	392,975	384,931	392,079
Pokkali26869	2,755,318	2,734,165	2,755,371	536,638	537,585	535,921
SR26B	1,563,298	1,533,956	1,560,536	340,530	331,286	337,168

6.3.4 Distribution of variants across chromosomes

After analyzing the variant distribution of the eQTL regions compared to the whole genome variants, it was observed that variants were well distributed in chromosome 1, 2, 3, 4, 6 followed by chromosome 8, 9 and 12. Chromosome 5, 7, 10 and 11 did not show many variants in the eQTL regions (Fig 6.3)

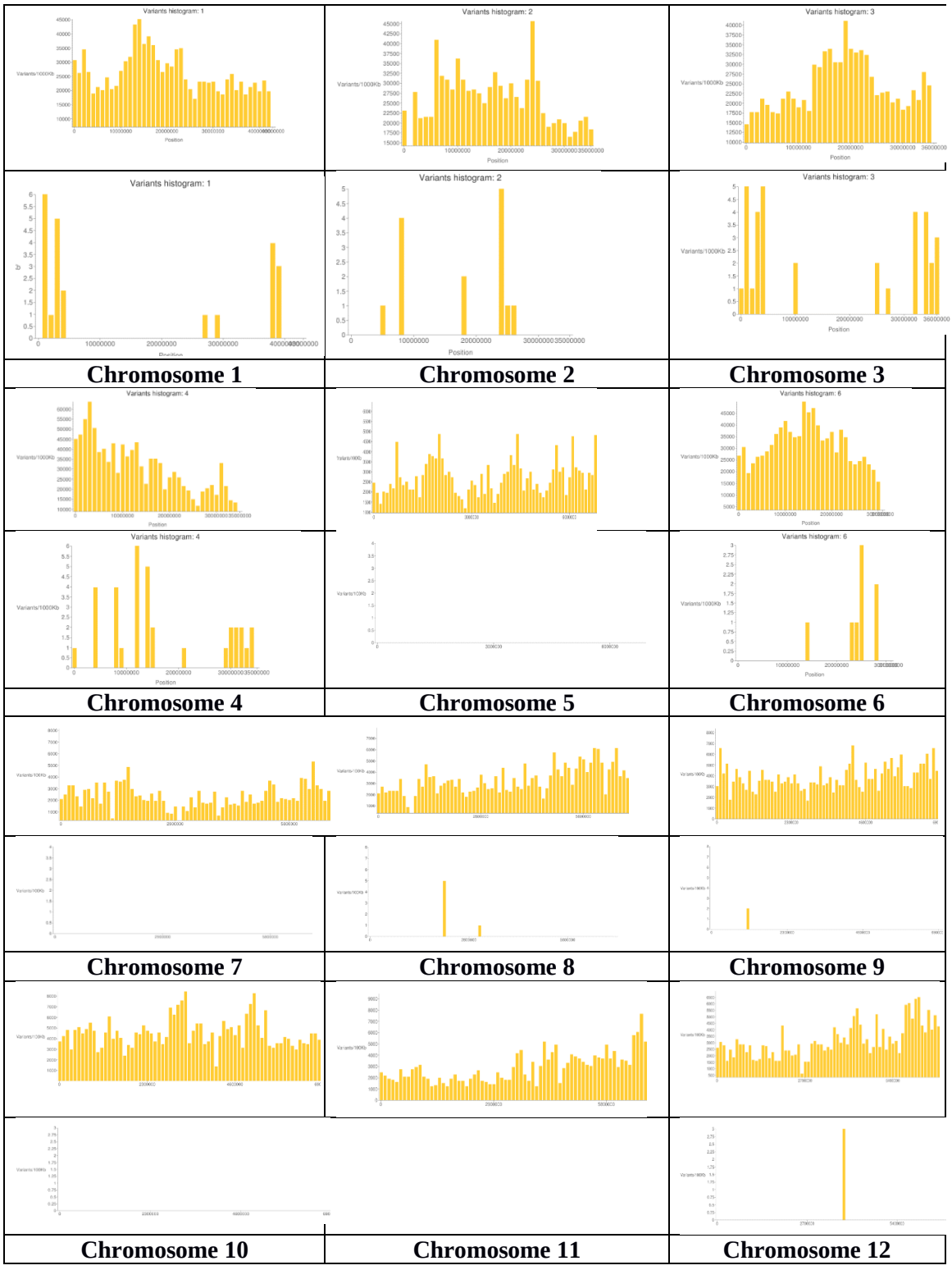
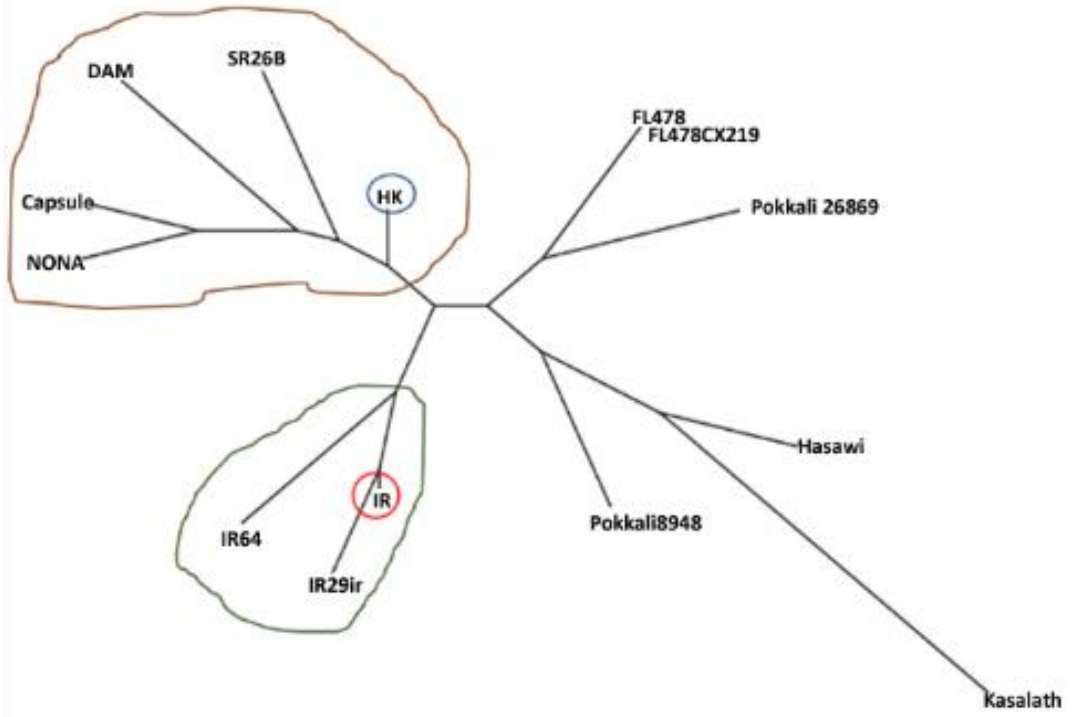


Fig 6.3: Distribution of variants across chromosomes. For each chromosome block, the upper row depicts all variants in the genome and the bottom row depicts only the variants found in eQTL regions. (high resolution figures available in Appendix)

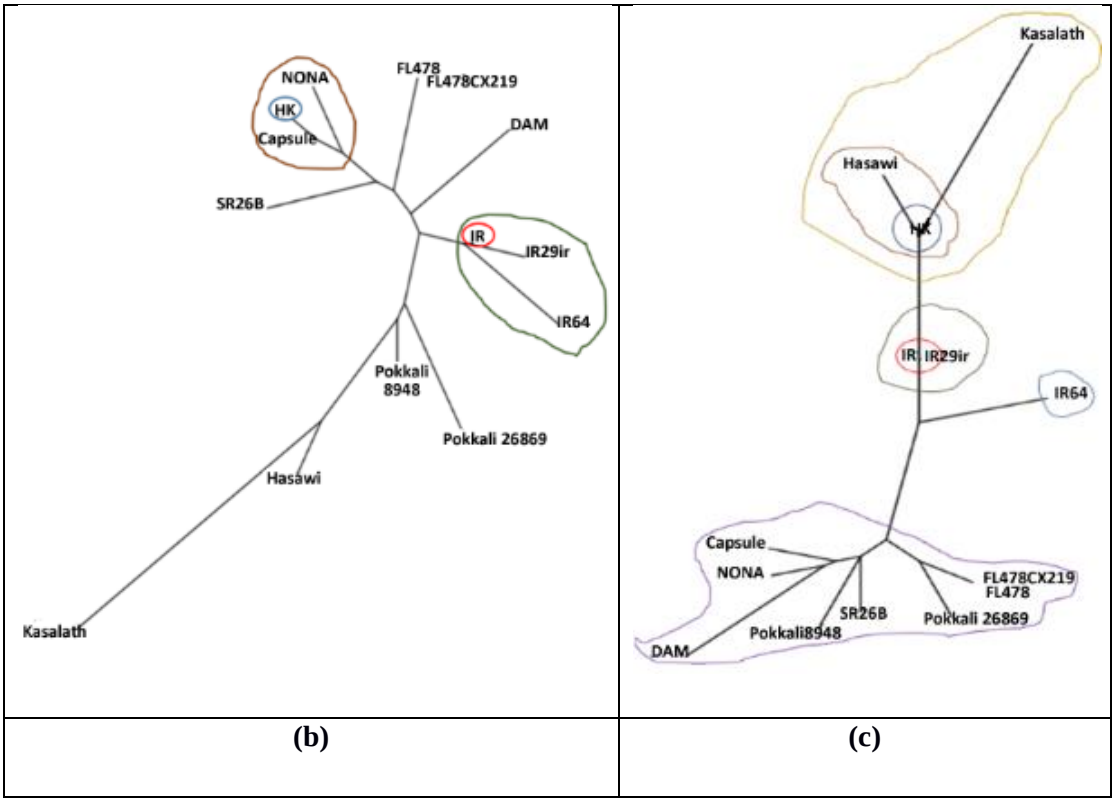
6.3.5 Phylogeny among the genomes

Horkuch grouped with the other salt tolerant landraces *Nonabokra, Damodar, Capsule , SR26B* and *IR29* grouped with *IR64*. There were differences observed in the sequences of IRRI-provided IR29 (ir29) and the newly sequenced IR29 which might be due to the very low sequencing depth of the ir29 file compared to the latter one.

Interestingly when all DArT markers were considered, *Horkuch* was grouped with *Nonabokra* and capsule but when only cis eQTL were considered it grouped with *Hasawi* and *Kasalath*. Here it can be mentioned that cis eQTLs are regions where the SNP markers affect a nearby gene causing differential expression. Later, detailed analysis of the regions showed similarity between *Horkuch* and *Hasawi* genome in those particular regions where heterozygous haplotype calls were dominant. This can be a point for further investigation.



(a)



(b) (c)

Fig 6.4: (a) Whole genome variant based phylogeny (b) All eQTL variant based phylogeny. (c) cis eQTL variant based phylogeny

For having an idea of their grouping, a chromosome wise phylogeny was also carried out, which reveals close relationship between *Horkuch* and *Hasawi* genome in Chromosome 1 and chromosome 12 (Fig 6.5). In chromosome 4 and 9, *Horkuch* and *IR29* clustered closely explaining less variation. In chromosome 5 *Horkuch* and other salt tolerant varieties clustered closely except *Hasawi*.

6.3.6 Allele mining and validation of the specific SNP alleles

The position of the SNP markers identified from the F_2 genotyping study which were associated with the cis-eQTLs were validated. The exact position varied from 6 to 30 bp (plus or minus). The cis-DArT positions as well as the variant positions found from the genome analysis are compiled in Table 6.4. This will be helpful in designing allele specific markers more confidently.

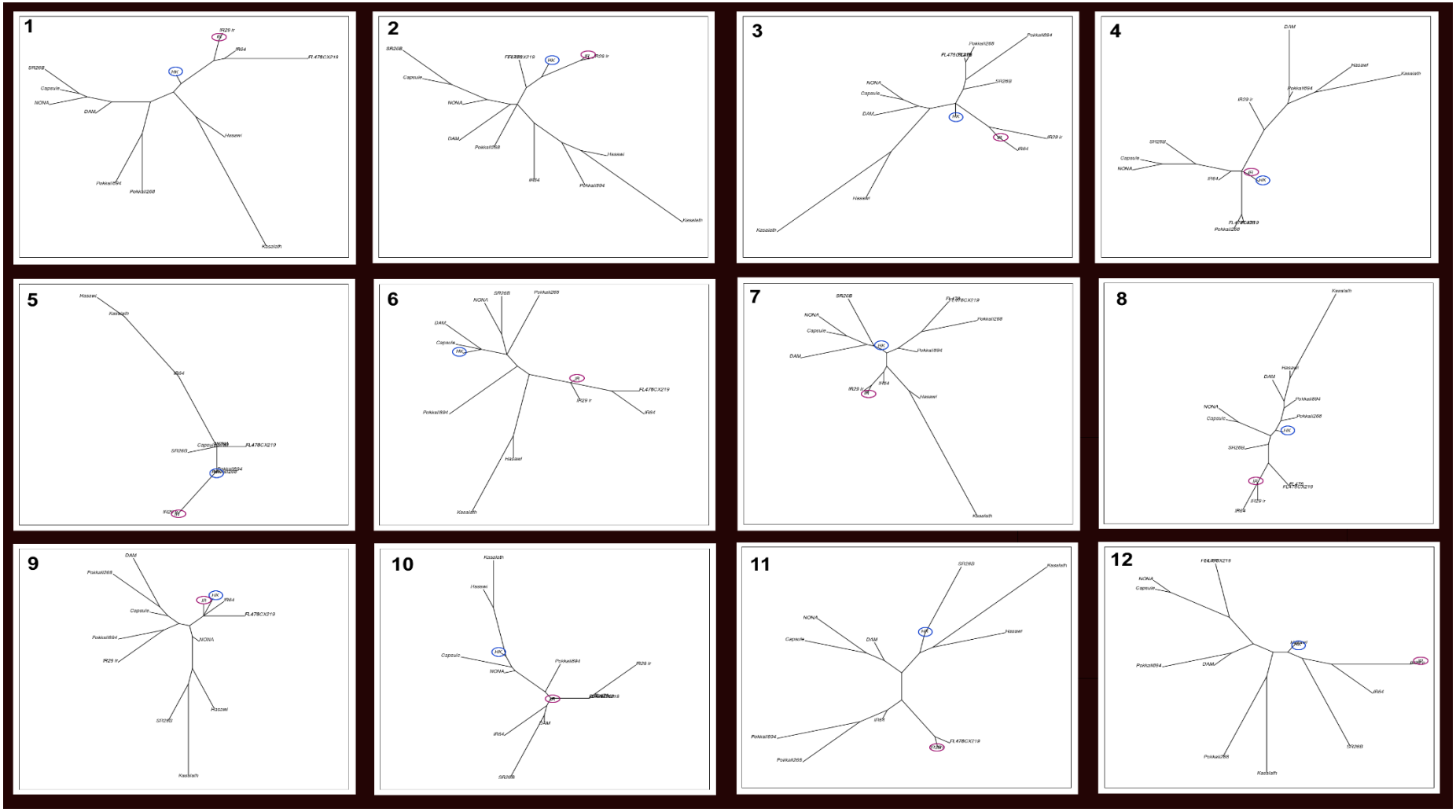


Fig 6.5: Phylogenetic relationship based on identified variants among salt tolerant rice genomes in individual chromosome. *Horkuch* marked with red circle and *IR29* with blue circle (High resolution figure in Appendix).

The allele change noticed in *Horkuch* and *IR29* compared to the reference was also analyzed in the other genomes included in this study and the specific allele in that position was extracted. Finally the MSU ID of the gene that possess any of the SNP markers were listed in the 'Gene' column of Table 6.4. Some specific markers were identified where *Horkuch* shared the specific allele with other salt tolerant varieties like SF013, SF125, SF430, SF720. Interestingly for SF47 IR varieties i.e. IR64 and IR29 shared the same allele while the others have the opposite.

Table 6.4: List of Alleles in specific SNP markers

SNP	DART position	Position	Changed Allele	St	HK	IR	Other Genomes*											Gene
							C	D	N	S	P2	P8	F	FC	I64	Hs	K	
SF013	1:32289876	3229006	G>A	+	G	A	A	A	G	G	A	A	A	A	A	A	G	LOC_Os01g06800 promoter
SF0125	1:24533617	1:24533677	A>T	+	A	T	T	T	T	T	.	T	T	T	T	T	T	
		1:24533702	G>C	?	G	C	G	G	G	G	G	C	G	G	G	G	G	
SF0128	1:27759059	1:27759066	A>C T>G	-	T	G	G	G	G	G	G	G	G	G	G	T	T	LOC_Os01g48440
SF0130	1:29556664	1:29556686	A>T (T>C)		T	C	C	C	C	C	C	C	C	C	C	T	T	
SF0134	1:33484711	Not found																
SF0139	1:38079382	NOT RELIABLE	A>G															
SF0140	1:39031989	1:39031993	(C>T)	-	T	T	T	T	T	T	T	T	T	T	T	T	C	LOC_Os01g67250
		1:39032037	T>C	+	**T/C**	T	T	T	T	T	T	T	T	T	T	T	T	
SF26	2:5359375	2:5359418	A>G (T>C)	-	T	C	T	T	T	T	T	T	T	T	T	T	T	LOC_Os02g10230
SF29	2:8308254	2:8308263	T>A	+	T	A	T	T	T	T	T	T	T	T	T	T	T	LOC_Os02g14900
		2:8308262	T>A CT/C?															
SF220	2:17785171	Not found																
Sf32	3:1435501	3:1435506	C>T (G>A)	-	G	A	A	A	A	A	A	A	A	A	A	A	A	Not reliable
SF33	3:2513308	3:2513321	C>T	+	T	T	T	T	T	T	T	T	T	T	T	T	T	LOC_Os03g05180
SF34	3:3460935	3:3460914	T>G	+	T	G												Nearby INDEL
SF35	3:4467064	3:4467097	A>C (G>T)	-	T	G	G	G	G	G	G	G	G	G	G	G	G	PROMOTER LOC_Os03g08670

SF334	3:32456249	3:32456285	A>G		**A**	G	G	G	G	G	G	**A**	G	G	G	G	**A**	
		3:32456264	G>A	?	**G**	A	A	A	A	A	A	**G**	A	A	A	A	A	
SF336	3:34415755	3:34415778	T>C	+	T	T	C	C	T	C	C	C	C	C	C	T	T	LOC_Os03g60550
		3:34415767	C>T	?	C	T	C	C	T	C	C	C	C	C	C	C	T	
SF337	3:35445145	3:35445162	T>C A>G		T	C	C	C	C	C	T	C	T	T	T	T	C	LOC_Os03g62630
SF47	4:4361849	4:4361855	G>T (T>C)		T	**C**	T	T	T	T	T	T	T	T	**C**	T	T	UTR of LOC_04g08200
SF414	4:15878250	4:15878283	A>G (T>C)	-	C	C												
		4:15878260	(C>T)	?	T	C	C	C	C	C	C	C	C	C	C	C	C	LOC_Os04g26850
SF415	4:16687477	Not found	C>T															
SF418	4:21366297	4:21366345	A>G															
SF430	4:32605170	4:32605195	C>G	+	C	G	C	C	C	C	C	C	C	C	G	G	C	LOC_Os04g54830
SF431	4:33568865	4:33568894	C>A (G>T)	-	G/T	T												3' utr of LOC_Os04g56300
SF432	4:34525966	4:34525987	T>C	+	T	C	C	C	C	C	C	C	C	C	C	T	T	INTERGENIC
SF521	5:219330870	5:21930884	G>A	+	G	A	A	A	A	A	A	A	A	A	A	A	A	LOC_Os05g37470
SF627	6:24142189	Not found	T>A															
SF631	6:28299218	6:28299253	T>G (A>C)	-	A	C	C	C	C	C	C	C	C	C	C	C	A	
SF720	7:18599556	7:18599577	A>G (T>C)	-	T	**C**	T	T	T	T	T	T	T	**C**	**C**	**C**	**C**	LOC_Os07g31390
SF722	7:20525007	7:20525028	C>T (G>A)	-	A	G		A					A	A				LOC_Os07g34280
SF723	7:21596128	7:21596156	C>G	+	C/G	C	C	C	C	C	C	C	C	C	C	C	C	LOC_Os07g36150
SF726	7:24559292	7:24559302	T>A	+	A	T	T	T	T	T	T	T	T	T	T	T	T	Promoter of LOC_Os07g41060
SF83	8:2158243	8:2158242	G>A	?	C	A	A	A	A	A	A	A	A	A	A	A	C	LOC_Os08g04400
		8:2158250		?	A	G	G	G	G	G	G	A	G	G	G	G	A	
		8:2158287		?	T	C	C	C	C	G	G	T	G	G	G	G	T	
SF918	9:15228188	9:15228192	A>G (T>C)	-	T	C	C	C	C	C	C	C	C	C	C	C	T	
SF923	9:20210872	9:20210877	A>G (T>C)	-	T	T	T	T	T	T	T	T	T	T	T	T/C	T	

SF924	9:21112097	9:21112119	G>T)C>A)	-	C/A	C	C	C	C	C	C	C	C	C	C	C	C	LOC_Os09g36608 CDS
SF1020	10:19770527	Not found																

*(**HK** = Horkuch, **IR**= IR29 (high coverage), **C** = Capsule, **D**= Damodar, **N** = Nonabokra, **S** = SR26B, **P2** = Pokkali26869, **P8** = Pokkali8948, **F** = FL478, **FC**= FL478CX219, **I64**= IR64, **Hs** = Hasawi, **K** = Kasalath, **St** = Strand)*

6.3.7 Allele mining in the differentially expressed gene regions associated with cis-eQTL

For this part, only cis-eQTL associated genes from Chromosome 10 were chosen for precise analysis. Variants were analyzed in three categories, i.e. (i) INDELS (Table 6.5), (ii) Homozygous SNPs (Table 6.6a) and (iii) significant Heterozygous SNPs (Table 6.6b). A short-list of SNP calls were generated by choosing homozygous alleles that are polymorphic between *Horkuch* and IR29 and heterozygous alleles in *Horkuch* that have significant variant effect.

Two deletions were observed in LOC_Os10g27090 where *Horkuch* has the same allele as reference and all others have the alternate allele (Table 6.5) Functional annotation of this gene was not found in the existing databases. In downstream region of one glutathione transferase gene (LOC_Os10g38580) alternate insertions were observed in *Horkuch* and *IR*. In OsFBK23 - F-box domain and kelch repeat containing protein alternate allele was observed in *IR29* with 33 bp insertion. Insertion alleles were also observed in a putative Guanylate Cyclase gene (LOC_Os10g13694) in exon region (Table 6.5).

In Table 6.6a SNPs observed between *Horkuch* and *IR29* were extracted where *Horkuch* has homozygous alleles. It was interesting to note that the homozygous variants were found in regulatory regions like 5′ UTR, 3′ UTR, upstream and downstream regions. Very few were found as missense and splice site variant effect. For example, one missense variant was identified in Rf1 mitochondrial precursor (LOC_Os10g35090 where *Horkuch* shares the reference allele (A) with Capsule, whereas all other genomes contain the alternate allele (G) (Table 6.6a)

Table 6.6b lists the allelic variant haplotypes of heterozygous SNPs, which are polymorphic between *Horkuch* and *IR29*. These are listed based on their significant effect as a variant, i.e. whether they fall in intron, intergenic, splice site regions or whether they are synonymous or missense SNPs in exonic regions. Missense variants are specifically highlighted here as they are important in disrupting function of a gene. Missense variants were observed in 5 genes. Among them LOC_Os10g14011 (contains transmembrane region, but no specific annotation), LOC_Os10g35090 (Rf1 mitochondrial precursor) and LOC_Os10g31820 (fluG, putative Glutamine synthetase) were prominent. The other 2 genes 1, 2, dihydroxy 3 – keto 5 methyl theopentene dioxygenase protein (LOC_Os10g28350) and OsFBK23 (LOC_Os10g26990) also

contain 2 missense variants. Other variants were found as synonymous or in the splice site or intronic variant sites.

Table 6.5: List of INDELS in the differentially expressed gene regions associated with cis-eQTLs in chr 10

Position	Reference	Alternate	Gene in the eQTL region	Gene description and Putative function	C	D	F	FC	HK	Hs	IR	i29	I64	K	N	P2	P8	S	Annotation
Insertion																			
2845833	CATATAT	C,CAT,CATATATATATATATATATATAT,CATATATATATATATATATATAT	LOC_Os10g05680	Expressed Protein	B	.	E	R	C	B	H	R	.	A	R	D	.	.	DOWN
7436563	A	AAAGC,AAAGCTTGTCAAATCAATCAAGG	LOC_Os10g13694	Guanylate cyclase putative	R	.	A	.	H2	.	A	.	.	.	.	.	.	.	EXN
7436567	T	TC,TCAATCAAGGTATAAAAGTAGA	LOC_Os10g13694	Guanylate cyclase putative	R	.	B	.	R	.	A	.	.	.	.	.	.	.	EXN
7617322	T	TCGAATCCGCATCCGGCAAGATCAAAGACA	LOC_Os10g14011	Expressed Protein	R	R	R	R	H	R	R	R	R	R	R	R	.	R	DOWN
14228072	T	TAGAGAGCAGAAGGGAAGAAACTTGGACTCTTA	LOC_Os10g26990	OsFBK23	A	A	R	R	H/R	R	A	.	A	R	A	R	A	A	EXN
14754537	A	ATTATGTAGTTTAACAACGTAT			A	A	.	A	H	A	A	A	A	A	A	.	H	A	
7616589	C	CCGGTGGCGGTAGCGG	LOC_Os10g14011	Expressed Protein	R	A	.	A	H	.	A	.	.	A	A	.	A	R	EXN
14283211	A	AGCACTTCGATTAG	LOC_Os10g27090	Expressed Protein	A	A	A	A	R	A	.	.	A	A	A	A	A	A	EXN
20591335	G	GTCAGCTAGTGGCGAACCCA,GCCAGCTAGTGGCGAACCCA	LOC_Os10g38580	Glutathione s transferase	A	A	A	A	C	H	A	.	A	R	A	H	A	A	DOWN
Deletion																			
14282018	CTACTCCCTCAGTCCCAAAT	C	LOC_Os10g27090	Expressed Protein	A	A	.	.	R	.	A	.	A	A	.	A	.	.	EXN
14283400	TAGG	T	LOC_Os10g27090	Expressed Protein	A	A	A	A	R	A	A	.	A	A	A	A	A	A	DOWN

UP: *Upstream gene variant* ***DOWN:*** *Downstream gene variant,* ***5-UTR:*** *5' UTR variant* ***3-UTR:*** *3' UTR variant,* ***INT:*** *Intronic variant* ***SYN :*** *Synonymous Variant,* ***MIS:*** *Missense variant,* ***SPLICE****: Splice site variant,* ***STOP :*** *Stop gained,* ***EXN****: Exonic region* ***R:*** *Reference Allele,* ***A:*** *Alternate Allele,* ***H:*** *Heterozygote Allele "."* ***:*** *No haplotype called* ***HK*** *= Horkuch,* ***IR****= IR29 (high coverage),* ***C*** *= Capsule,* ***D****= Damodar,* ***N*** *= Nonabokra,* ***S*** *= SR26B,* ***P2*** *= Pokkali26869,* ***P8*** *= Pokkali8948,* ***F*** *= FL478,* ***FC****= FL478CX219,* ***I64****= IR64,* ***Hs*** *= Hasawi,* ***K*** *= Kasalath, I 29 = IR29 previous*

Table 6.6A: List of Homozygous SNPs in *Horkuch* genome compared to *IR29* in the differentially expressed gene regions associated with cis-eQTLs in Chr 10

Position	Reference	Alternate	Gene in the eQTL region	Gene description and Putative function	Capsule	Damodar	FL478	FL478CX21	Horkuch	Hasawi	IR29	Ir29	IR64	Kasalath	Nonabokra	Pokkali2686	Pokkali8948	SR26B	Variant Annotation
7435339	C	T	LOC_Os10g13694	Expressed protein	R	R	A	A	R	A	H	A	R	A	R	A	A	R	5-UTR
7436971	A	G	LOC_Os10g13694	Expressed protein	R	.	.	A	R	A	A	.	A	A	.	A	.	.	3-UTR
14227097	T	C	LOC_Os10g26990	OsFBK23	A	A	A	.	R	A	A	.	A	A	A	A	A	A	UP
14227107	G	T	LOC_Os10g26990	OsFBK23	A	A	A	A	R	A	A	.	A	A	A	A	A	A	UP
14227121	A	G	LOC_Os10g26990	OsFBK23	A	A	A	A	R	A	A	.	A	A	A	A	A	A	UP
14227136	A	G	LOC_Os10g26990	OsFBK23	A	A	A	A	R	A	A	.	A	A	A	A	A	A	UP
14227160	T	C	LOC_Os10g26990	OsFBK23	A	A	A	A	R	A	.	.	A	A	A	A	A	A	UP
14227688	T	C	LOC_Os10g26990	OsFBK23	A	A	A	A	R	A	A	A	A	A	A	A	A	A	UP
14230067	G	A	LOC_Os10g26990	OsFBK23	.	.	A	A	R	A	.	.	.	A	A	A	.	A	SPLICE
14231504	A	C, *	LOC_Os10g26990	OsFBK23	.	.	.	.	R	A	*	.	.	A	.	.	.	.	INT
14231516	C	A	LOC_Os10g26990	OsFBK23	R	.	.	.	R	R	A	.	.	.	.	.	.	.	INT
14231525	A	G	LOC_Os10g26990	OsFBK23	R	R	R	.	R	R	A	.	.	R	R	.	.	R	INT
14231583	T	C	LOC_Os10g26990	OsFBK23	A	A	.	A	R	A	A	.	A	A	A	.	.	.	INT
14231585	A	C	LOC_Os10g26990	OsFBK23	A	A	.	A	R	A	A	.	A	A	A	.	.	.	INT

1428 0083	C	T	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	.	.	A	.	.	.	.	.	INT
1428 0089	T	C	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	.	.	A	.	.	.	.	.	INT
1428 0195	T	G	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	A	.	A	.	.	.	.	.	INT
1428 0202	A	T	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	A	.	A	.	.	.	.	.	INT
1428 0221	C	T	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	A	.	A	.	.	.	.	.	INT
1428 0233	T	C	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	A	.	A	.	.	.	.	.	INT
1428 0243	T	C	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	A	.	A	.	.	.	.	.	INT
1428 0304	C	T	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	A	A	A	.	.	.	.	.	SYN
1428 0318	T	C	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	A	A	A	.	.	.	.	.	MIS
1428 0325	A	G	LOC_ Os10g 27090	Expressed protein	.	.	.	.	R	.	A	A	A	.	.	.	.	.	SYN
1428 1817	G	T	LOC_ Os10g 27090	Expressed protein	A	A	A	.	R	A	A	A	A	A	A	A	.	.	3-UTR
1428 1843	T	C	LOC_ Os10g 27090	Expressed protein	A	A	A	.	R	A	A	A	A	A	A	A	.	.	3-UTR
1428 1855	A	G	LOC_ Os10g 27090	Expressed protein	A	A	A	.	R	A	A	A	A	A	A	A	.	.	3-UTR
1428 1876	T	A	LOC_ Os10g 27090	Expressed protein	A	A	A	.	R	A	A	A	A	A	A	A	.	.	3-UTR
1428 1878	A	C	LOC_ Os10g 27090	Expressed protein	A	A	A	.	R	A	A	A	A	A	A	A	.	.	3-UTR
1428 3155	C	T	LOC_ Os10g 27090	Expressed protein	A	A	A	A	R	A	A	A	A	A	A	A	A	A	3-UTR
1428 3163	T	C	LOC_ Os10g 27090	Expressed protein	A	A	A	A	R	A	A	A	A	A	A	A	A	A	3-UTR
1428 3175	G	C	LOC_ Os10g 27090	Expressed protein	A	A	A	A	R	A	A	A	A	A	A	A	A	A	3-UTR
1428 3687	T	A	LOC_ Os10g 27090	Expressed protein	A	A	A	A	R	A	A	A	A	A	A	A	.	.	DOWN
1668 7844	C	A	LOC_ Os10g 31820	fluG, Glutamine synthetase	A	A	A	A	R	A	A	A	A	R	R	A	A	R	INT
1871 7340	G	C	LOC_ Os10g 35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	A	A	A	A	A	.	A	UP
1871 7412	T	C	LOC_ Os10g 35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	A	A	A	A	A	A	A	UP

18717525	G	A	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	.	A	A	A	A	A	A	UP
18717535	G	A	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	.	A	A	A	A	A	A	UP
18717615	A	G	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	A	A	A	A	A	A	A	UP
18717637	T	G	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	A	A	A	A	A	A	A	UP
18717906	C	T	LOC_Os10g35090	Rf1,_mitochondrial_precursor	A	R	R	R	A	R	R	R	R	R	R	R	R	R	UP
18717907	T	A	LOC_Os10g35090	Rf1,_mitochondrial_precursor	A	R	R	R	A	R	R	R	R	R	R	R	R	R	UP
18718667	A	C	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	H	H	A	R	H	A	A	H	A	A	A	A	H	5-UTR
18718805	A	T	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	H	R	A	A	.	A	A	H	A	H	A	5-UTR
18718979	C	T	LOC_Os10g35090	Rf1,_mitochondrial_precursor	A	R	R	R	A	R	R	.	R	R	R	R	R	R	5-UTR
18719308	G	T	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	A	A	A	A	A	A	A	5-UTR
18719749	A	G	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	A	A	A	A	A	A	A	MIS
18720002	C	A	LOC_Os10g35090	Rf1,_mitochondrial_precursor	A	R	R	R	A	R	R	.	R	R	R	R	R	R	SPLICE
18720470	C	T	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	A	A	A	A	A	A	A	3-UTR
18720474	C	T	LOC_Os10g35090	Rf1,_mitochondrial_precursor	A	R	R	R	A	R	R	R	R	R	R	R	R	R	3-UTR
18720524	G	A	LOC_Os10g35090	Rf1,_mitochondrial_precursor	A	R	R	R	A	R	R	R	R	R	R	R	R	R	3-UTR
18720704	G	A	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	A	A	A	A	A	A	A	3-UTR
18720827	C	G	LOC_Os10g35090	Rf1,_mitochondrial_precursor	A	R	R	R	A	R	R	R	R	R	R	R	R	R	3-UTR
18720850	T	C	LOC_Os10g35090	Rf1,_mitochondrial_precursor	A	R	R	R	A	R	R	R	R	R	R	R	R	R	3-UTR
18721783	G	T	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	.	A	A	A	A	A	A	SYN
18721863	T	A	LOC_Os10g35090	Rf1,_mitochondrial_precursor	R	A	A	A	R	A	A	A	A	A	A	A	A	A	3-UTR
19414561	T	G	LOC_Os10g36350	MSP_domain_containing_protein	A	A	A	A	R	A	A	.	A	A	A	A	A	A	UP

1941 4601	C	T	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	A	A	A	A	R	A	A	A	A	A	A	A	.	.	UP
1941 4602	A	G	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	A	A	A	A	R	A	A	A	A	A	A	A	A	.	UP
1941 4616	A	T	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	A	A	A	A	R	A	A	A	A	A	A	A	A	.	UP
1941 4617	A	G	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	A	A	A	A	R	A	A	A	A	A	A	A	A	.	UP
1941 4631	A	G	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	A	A	A	A	R	A	A	A	A	A	A	A	A	.	UP
1941 4668	G	A	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	A	A	A	A	R	A	A	A	A	A	A	A	A	.	UP
1941 4691	G	T	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	A	A	A	A	R	A	A	A	A	A	A	A	A	.	UP
1941 6047	G	T	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	A	.	A	.	R	.	A	.	.	A	.	A	.	.	INT
1941 6074	T	A	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	.	.	A	.	R	.	A	.	.	A	.	A	.	.	INT
1941 6099	A	T	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein,	.	.	A	.	R	.	A	.	.	A	.	A	.	.	INT
1941 6126	T	C	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	.	.	A	.	R	.	A	.	.	A	.	A	.	.	INT
1941 6158	A	C	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein	.	.	A	.	R	.	A	.	.	A	.	A	.	.	INT
1941 6167	A	G	LOC_ Os10g 36350	MSP_dom ain_contai ning_prot ein,	.	.	A	.	R	.	A	A	.	A	.	A	.	.	INT

UP: *Upstream gene variant* ***DOWN:*** *Downstream gene variant,* ***5-UTR:*** *5' UTR variant* ***3-UTR:*** *3' UTR variant,* ***INT:*** *Intronic variant*

SYN: *Synonymous Variant,* ***MIS:*** *Missense variant,* ***SPLICE****: Splice site variant,* ***STOP :*** *Stop gained*

R: *Reference Allele,* ***A:*** *Alternate Allele,* ***H:*** *Heterozygote Allele "."****:*** *No haplotype called*

Table 6.6B: List of significant variant effect containing Heterozygous SNPs in *Horkuch* genome compared to *IR29* in the differentially expressed gene regions associated with cis-eQTLs in Chr 10

Position	R	A	Gene in the eQTL region	Gene description and Putative function	C	D	F	FC	HK	Hs	IR	Ir	I64	K	N	P2	P8	S	Annotation
7616268	C	T	LOC_Os10g14011	Transmembrane region	A	R	R	R	H	R	R	R	R	R	R	R	R	R	SYN
7616305	C	T	LOC_Os10g14011	Transmembrane region	A	R	R	R	H	R	R	R	R	R	R	R	R	R	SYN
7616327	C	T	LOC_Os10g14011	Transmembrane region	A	R	R	R	H	R	R	R	R	R	R	R	R	R	MIS
7616350	G	A	LOC_Os10g14011	Transmembrane region	R	A	A	A	H	A	A	A	A	A	A	A	A	A	MIS
7616373	C	T	LOC_Os10g14011	Transmembrane region	A	R	R	R	H	R	R	R	R	R	R	.	R	R	SYN
7616408	C	T	LOC_Os10g14011	Transmembrane region	A	R	R	R	H	R	R	R	R	R	R	R	R	R	MIS
7616432	T	C	LOC_Os10g14011	Transmembrane region	R	A	A	A	H	.	A	A	A	A	A	.	A	A	MIS
7616448	G	A	LOC_Os10g14011	Transmembrane region	R	A	A	A	H	.	A	A	A	A	A	.	A	A	SYN
7616459	G	A	LOC_Os10g14011	Transmembrane region	R	A	A	A	H	.	A	A	.	A	A	.	A	A	MIS
7616461	C	T	LOC_Os10g14011	Transmembrane region	A	R	R	R	H	R	R	R	.	R	R	R	R	R	STOP
8000897	G	A	LOC_Os10g15300	Transmembrane region	A	R	R	R	H	R	R	R	R	R	R	R	R	R	MIS
14228859	T	G	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14228893	T	C	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14228953	G	A	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14228968	G	A	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14228980	C	A	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229007	G	A	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	H	SYN
14229031	A	C	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	H	SYN
14229037	C	T	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	H	SYN
14229058	C	T	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	H	SYN
14229124	C	T	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229169	G	A	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229211	A	G	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229229	T	C	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229259	C	T	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229291	G	A	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229355	G	A	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN

14229499	A	G	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229541	A	G	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229556	C	T	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229586	C	G	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229637	T	C	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	H	SYN
14229640	T	C	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	H	SYN
14229745	C	T	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229856	G	A	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14229945	A	G	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	MIS
14229973	A	G	LOC_Os10g26990	OsFBK23	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
14231824	A	C	LOC_Os10g26990	OsFBK23	.	A	.	.	H	.	A	.	.	A	A	.	.	.	MIS
14231825	A	G	LOC_Os10g26990	OsFBK23	.	A	.	.	H	.	A	.	.	A	A	.	.	.	SYN
14283312	C	T	LOC_Os10g26990	OsFBK23	A	A	R	R	H	R	A	A	A	R	A	R	A	A	SPLICE
14753471	T	G	LOC_Os10g28350	1,2-dihydroxy-3-keto-5-methylthiopentene_dioxygenase_protein,_putative,_expressed	A	A	A	A	H	A	A	A	A	A	A	A	A	A	MIS
16688428	G	A	LOC_Os10g31820	fluG,Glutamine synthetase	A	A	A	A	H	A	A	A	A	R	R	A	A	R	SYN
16688603	G	T	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	R	R	H	A	R	R	R	R	R	R	MIS
16688604	C	T	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	R	R	H	A	R	R	R	R	R	R	MIS
16688872	G	A	LOC_Os10g31820	fluG,Glutamine synthetase	A	A	A	A	H	A	A	A	A	R	.	A	A	.	SPLICE
16689064	C	T	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	R	R	H	A	R	R	R	R	R	R	SYN
16690112	T	G	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	H	R	R	R	R	A	A	R	R	A	SPLICE & SYN
16690256	C	T	LOC_Os10g31820	fluG,Glutamine synthetase	A	A	A	A	H	A	A	A	A	R	R	A	A	R	SYN
16690267	G	A	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	R	R	H	A	R	R	R	R	R	R	MIS
16690292	C	T	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	R	R	H	A	R	R	R	R	R	R	SYN

16690497	G	A	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	R	R	H	A	R	R	R	R	R	R	SYN
16690504	G	A	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	R	R	H	A	R	R	R	R	R	R	MIS
16690638	T	C	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	R	R	H	A	R	R	R	R	R	R	SYN
16690898	G	A	LOC_Os10g31820	fluG,Glutamine synthetase	R	R	R	R	R	R	H	A	R	R	R	R	R	R	MIS
18719787	A	T	LOC_Os10g35090	Rf1,_mitochondrial_precursor,_putative,_expressed	R	A	A	A	R	A	H	.	A	H	A	A	A	A	MIS
18719866	T	C	LOC_Os10g35090	Rf1,_mitochondrial_precursor,_putative,_expressed	R	H	H	H	R	A	H	.	H	H	A	H	H	A	MIS
18721415	C	T	LOC_Os10g35090	Rf1,_mitochondrial_precursor,_putative,_expressed	R	A	A	A	R	A	H	.	A	H	A	.	A	A	MIS
19417608	C	T	LOC_Os10g36350	MSP_domain_containing_protein,_putative,_expressed	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SPLICE
19417856	A	C	LOC_Os10g36350	MSP_domain_containing_protein,_putative,_expressed	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
19417895	G	A	LOC_Os10g36350	MSP_domain_containing_protein,_putative,_expressed	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SPLICE & SYN
19418246	A	G	LOC_Os10g36350	MSP_domain_containing_protein,_putative,_expressed	A	A	A	A	H	A	A	.	A	A	A	A	A	A	SYN
19418481	G	A	LOC_Os10g36350	MSP_domain_containing_protein,_putative,_expressed	A	A	A	A	H	A	A	.	A	A	A	A	A	A	SYN
19418495	G	T	LOC_Os10g36350	MSP_domain_containing_protein,_putative,_expressed	A	A	A	A	H	A	A	.	A	A	A	A	A	A	SPLICE
19418652	A	G	LOC_Os10g36350	MSP_domain_containing_protein,_putative,_expressed	A	A	A	A	H	A	A	A	A	A	A	.	A	.	SYN
22924138	C	G	LOC_Os10g42510	expressed_protein	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SPLICE

22932214	T	C	LOC_Os10g42510	expressed_protein	A	A	A	A	H	A	A	A	A	A	A	A	A	A	SYN
22932345	C	T	LOC_Os10g42510	expressed_protein	A	A	A	A	H	A	A	A	A	A	A	A	A	A	MIS

SYN : *Synonymous Variant,* ***MIS:*** *Missense variant,* ***SPLICE****: Splice site variant,* ***STOP :*** *Stop gained (****HK*** *= Horkuch,* ***IR****= IR29 (high coverage),* ***C*** *= Capsule,* ***D****= Damodar,* ***N*** *= Nonabokra,* ***S*** *= SR26B,* ***P2*** *= Pokkali26869,* ***P8*** *= Pokkali8948,* ***F*** *= FL478,* ***FC****= FL478CX219,* ***I64****= IR64,* ***Hs*** *= Hasawi,* ***K*** *= Kasalath,,* ***R:*** *Reference Allele,* ***A:*** *Alternate Allele,* ***H:*** *Heterozygote Allele "." :* *No haplotype called*

6.3.8 Allele mining in the candidate gene regions

The identified candidate genes from chapter 4 were also taken into consideration for variant analysis. Some of them indicated promising InDels. OsRCI2-6 showed a 12 bp deletion in the 3′ downstream region (Fig 6.6). This deletion is shared by mostly (but not all) salt tolerant varieties. Another gene LOC_Os07g02250 (Expressed protein) which showed high expression value in tolerant roots in seedling stage compared to the sensitive plants, has shown a G>GGGA insertion in the last intron shared by most of the salt tolerant varieties except *FL478* and *Pokkali*. Fructose-bisphosphate aldolase isozyme (LOC_Os01g67860) has shown a downstream deletion of ~72bp which is shared by most salt tolerant varieties except *FL478* and *Hasawi*. This gene was identified as an eQTL multiple times in the experiments. Trehalose-6-phosphate synthase (LOC_Os09g20990) showed a variation in sequence TGA in the 5′ UTR region in most salt tolerant varieties except *Pokkali* and *Hasawi*.

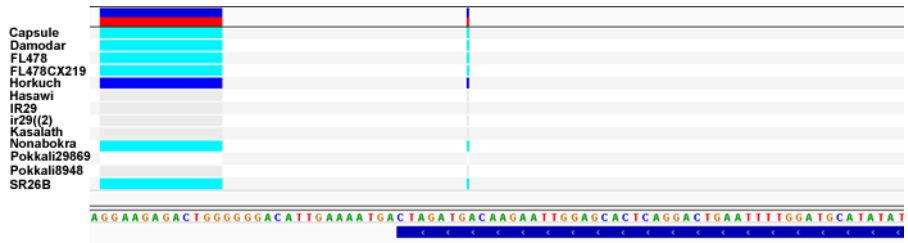


Fig 6.6 ~12 bp deletion in 3′ downstream region of OsRCI2-6 gene

6.4 Discussion

Variant discovery is an interesting way of finding genetic variation among the genomes to infer the causative variants for expression variations. It has the potential to identify subtle regions in regulatory region of the genes that can disrupt its function. Moreover chunks of insertions or deletions in the genome can be identified e.g. the discovery of the role of protein kinase (PSTOL1) in tolerance to phosphate deficiency (Gaumyao et al, 2012), which was absent from the reference genome, Nipponbare. Moreover the traits for tolerance to abiotic stress are more

abundant in rice varieties native to the certain stressed environment. So variant discovery and allele diversity mining is an excellent mean of comparative analysis among the genomes.

Analysis of the eQTL regions and differentially expressed genes in the breeding population of *Horkuch* and *IR29* identified numerous number of putative regions in the genome associated with salt tolerance and yield. Goal of the allele mining and variant analysis was to narrow down the list by comparing with other donor and recipient varieties to identify the most significant ones. Moreover the variant analysis study could validate the important SNP markers identified in the QTL regions by confirming their position in the genome. This will facilitate the design of precise allele-specific markers. Variations identified in differentially expressed gene regions provided some rational behind their differential expression. Some of the significant variants are mentioned in the results section including the deletion in downstream region of OsRCI2-6 and Fructose-bisphosphate aldolase isozyme gene. Phylogeny based on the variants could give insight on the genomic architecture of the *Horkuch* and *IR29* genome with other salt tolerant and sensitive varieties. Distribution of variants in chromosome and comparative analysis in individual chromosomes showed the importance of certain chromosomes for the traits of interest.

A short subset of the genome was analyzed in the study. However it contains immense potential to identify specifically important regions associated with salt tolerance if specific genes and QTL regions associated with salt tolerance are chosen and analyzed precisely. In depth analysis of the combined population eQTLs in all regions will give valuable information on allele diversity for salt tolerance.

Chapter 7

Conclusion

This study took a multidimensional approach to find the genetic loci and genes responsible for salt tolerance of *Horkuch*, a salt tolerant rice landrace with low yields endemic to the saline coast in the southwest of Bangladesh, north of the Sundarbans. This was done by analyzing a set of salt tolerant, sensitive and intermediate category plant siblings derived from *Horkuch* and high-yielding but sensitive IR29 with the aim of finding and the heritability of salt tolerant traits in the population. Since this set of progenies share a similar genetic background due to their inheritance from same parents differing only in few characteristics e.g. salinity tolerance, they were an appropriate and informative set for deciphering the mechanism and salt tolerance determinants.

The parents were reciprocally crossed to make two populations so that the mother's cytoplasmic effect could also be tracked. The set of reciprocal $F_{2:3}$ population containing 300 individuals was genotyped using DArTSeq™ for discovering SNP markers and phenotyped under salt stress, both at seedling and reproductive stages when rice is most sensitive to salt stress. A linkage map was generated from the genotyping data and quantitative trait loci (pQTL) were mapped associating the phenotype data. Moreover, using 3′ tag based RNAseq, differentially expressed genes among the tolerant and sensitive plants were identified in both developmental stages, from leaf and roots under salt stress and no stress condition. Expression QTLs (eQTL) were mapped in order to associate genetic and expression polymorphism. This generated a set of candidate genes and informative regions in the genome which are proposed to affect the salt tolerance traits of *Horkuch*. The functional annotation of the candidate genes, as well as their functional network analysis could also relate to the tolerant phenotypes. An image-based non-destructive automated and continuous phenotyping over 3 weeks of salt stress was carried out on a selected F_3 and F_5 sub-populations followed by QTL identification for the digital traits and relative growth rates. Since the digital phenotyping gives a longitudinal result rather than an endpoint one, it could separate the early and late responses to salt stress as well as correlate with the manually identified pQTLs from F_3. Finally the comparative variant analysis using allele mining with other salt tolerant donors and recipient genomes helped in deciphering regulatory signals in the identified regions responsible for salt tolerance.

The study aimed at identifying loci affecting yield under stress as high yield under stress is the desired phenotype most important for breeding salt tolerant crops. The inclusion of two sensitive developmental stages, i.e. seedling and reproductive as well as inclusion of without stress condition measurements in expression analysis and automated phenotyping, facilitated the distinction between the general effects of salt on plant response, from the genetic variation causing tolerance. The effect of mother's cytoplasm on specific traits helped in identifying the combination of best traits conferring tolerance. The next step will be to incorporate the specific

regions responsible for better tolerance and yield by pyramiding, selection followed by breeding to introgress the specific regions into commercial varieties.

Plant selection for use as superior tolerance donors will be a crucial process which will based on the proper combination of the identified regions for pyramiding. Here, the QTL effect and parent of origin of the allele will also play a decisive role. High yield is already associated with better relative water content and we also obtained a QTL for relative water content from the high yielding parent. Selection of plants containing this QTL in seedling stage along with potassium content QTL of seedling and yield related QTLs of reproductive stage can be a good strategy for superior plant selection. It was observed that the trait controlling potassium content is affected by the cytoplasm of the mother plant. This needs to be kept in consideration while selecting this QTL, even though cytoplasm was used as a covariate while mapping the QTLs. The positive effect of the *IR29* nuclear allele (cytoplasmic effect) for the yield trait QTLs and the strongly positive *Horkuch* nuclear allele effect for QTL models of desired traits e. g., K^+, will help define the combinations of cytoplasm and nuclear-donor materials for breeding plants with desired traits. The identified network of genes in the progenies also support the segregation of better photosynthetic and detoxifying ability of *Horkuch* in the breeding lines. The resolution of the multidisciplinary approach was at the level of specific genes. Many genes could be associated with the tolerant response, including certain master regulator proteins. These and other relevant genes can be validated using knockout or overexpression strategies and can be genetically engineered using transformation. Implementation of the knowledge obtained from this study for modern breeding by selection, pyramiding and genetic modifications will facilitate production of high yielding salt-tolerant crop varieties, which are adaptable and cultivable in the changing environment.

Appendix

Appendix

- **Accession codes:** RNA-seq data analyzed here has been deposited in NCBI database under BioProject ID: PRJNA317262. SRA submission id: SRP090879 and can be accessed at https://trace.ncbi.nlm.nih.gov/Traces/study/?acc=SRP090879.

- Since the supplementary files are large in size they are dumped in a website. The supplementary files are temporarily available in the following URL http://www.pbtlabdu.net/SMETHESIS/thesis_sme.zip

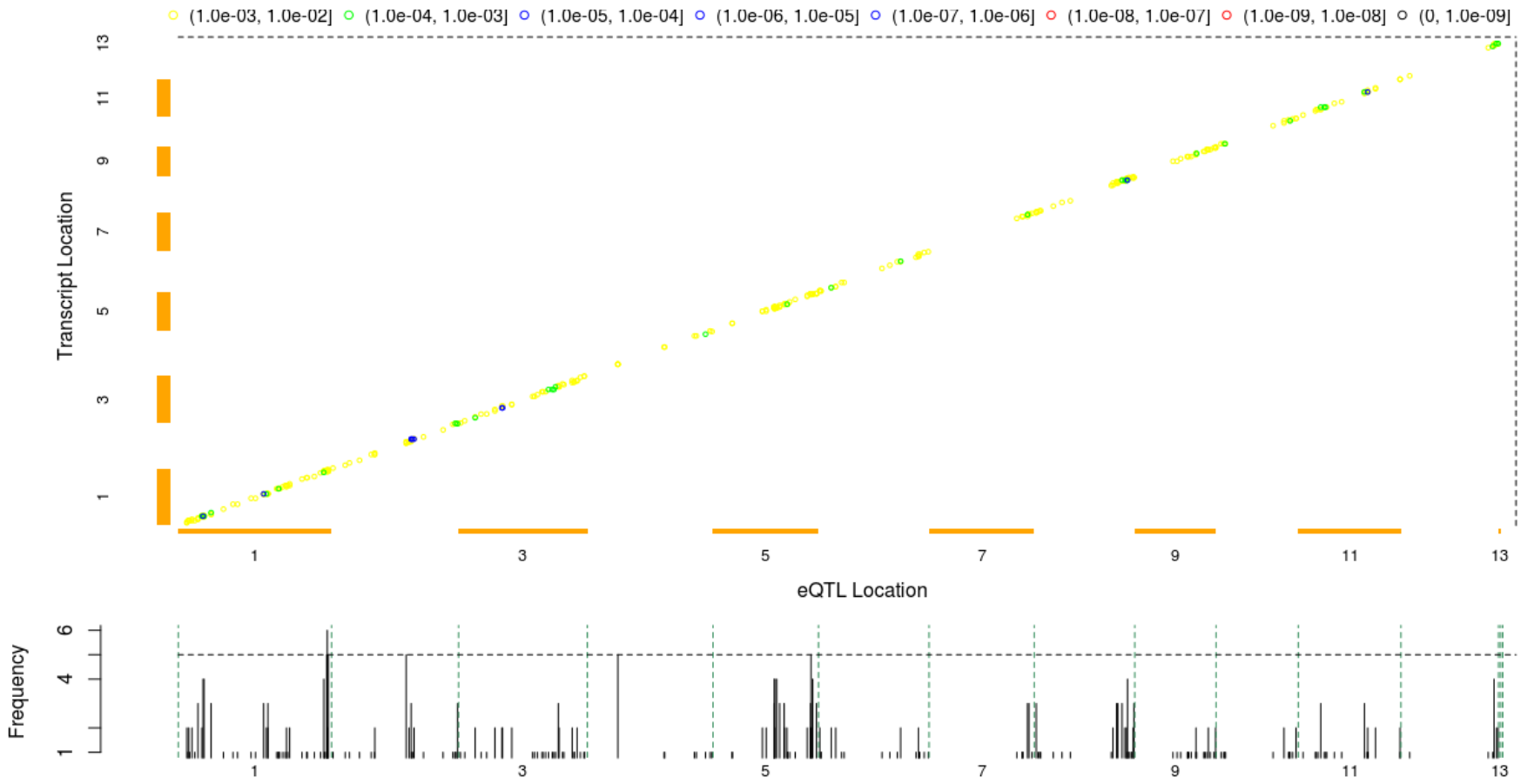


Fig 4.13: Full eQTL map for ReLD72Ci (cis) (A)

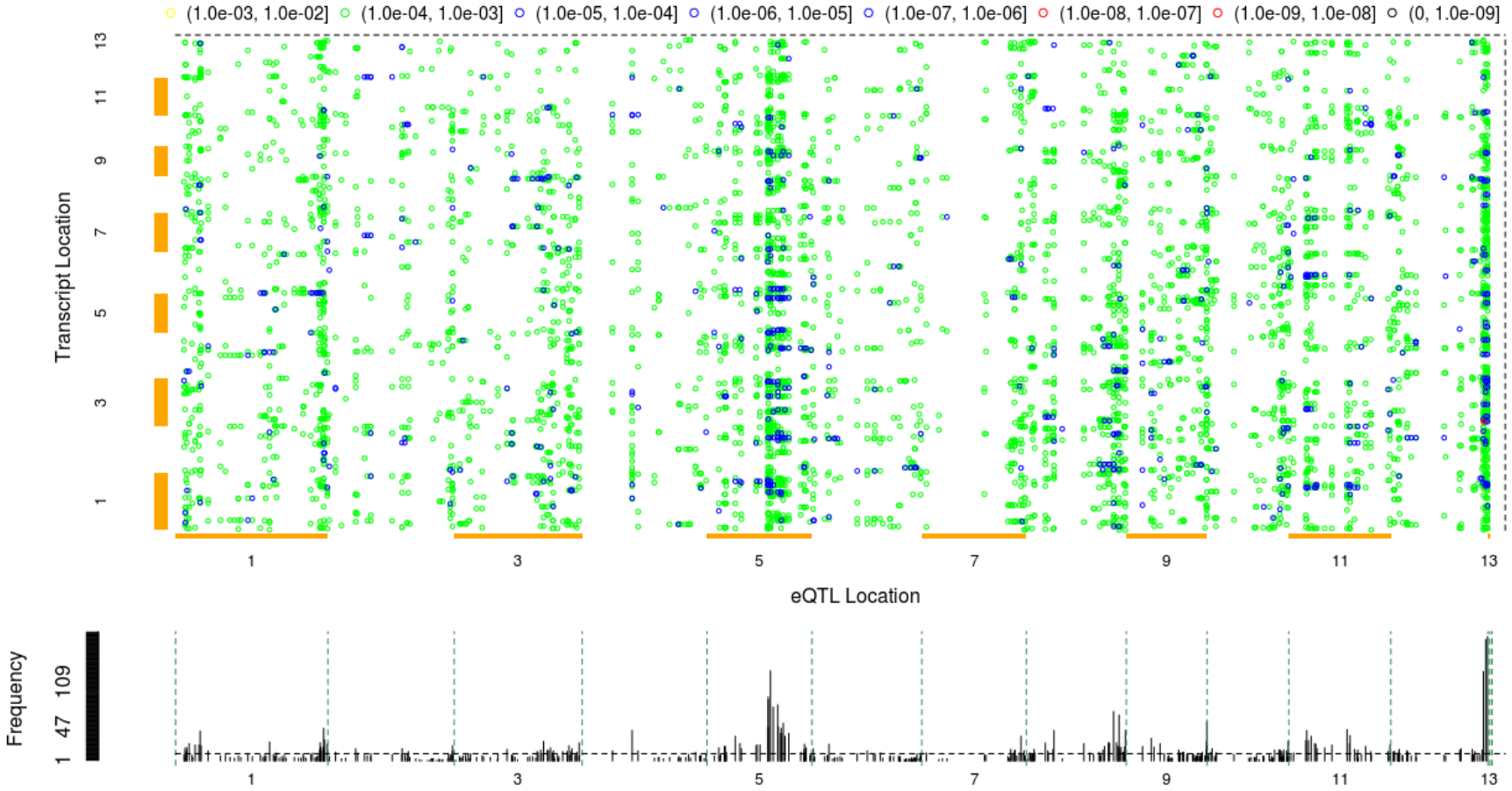


Fig 4.13: Full eQTL map for ReLD72Tr (trans) (B)

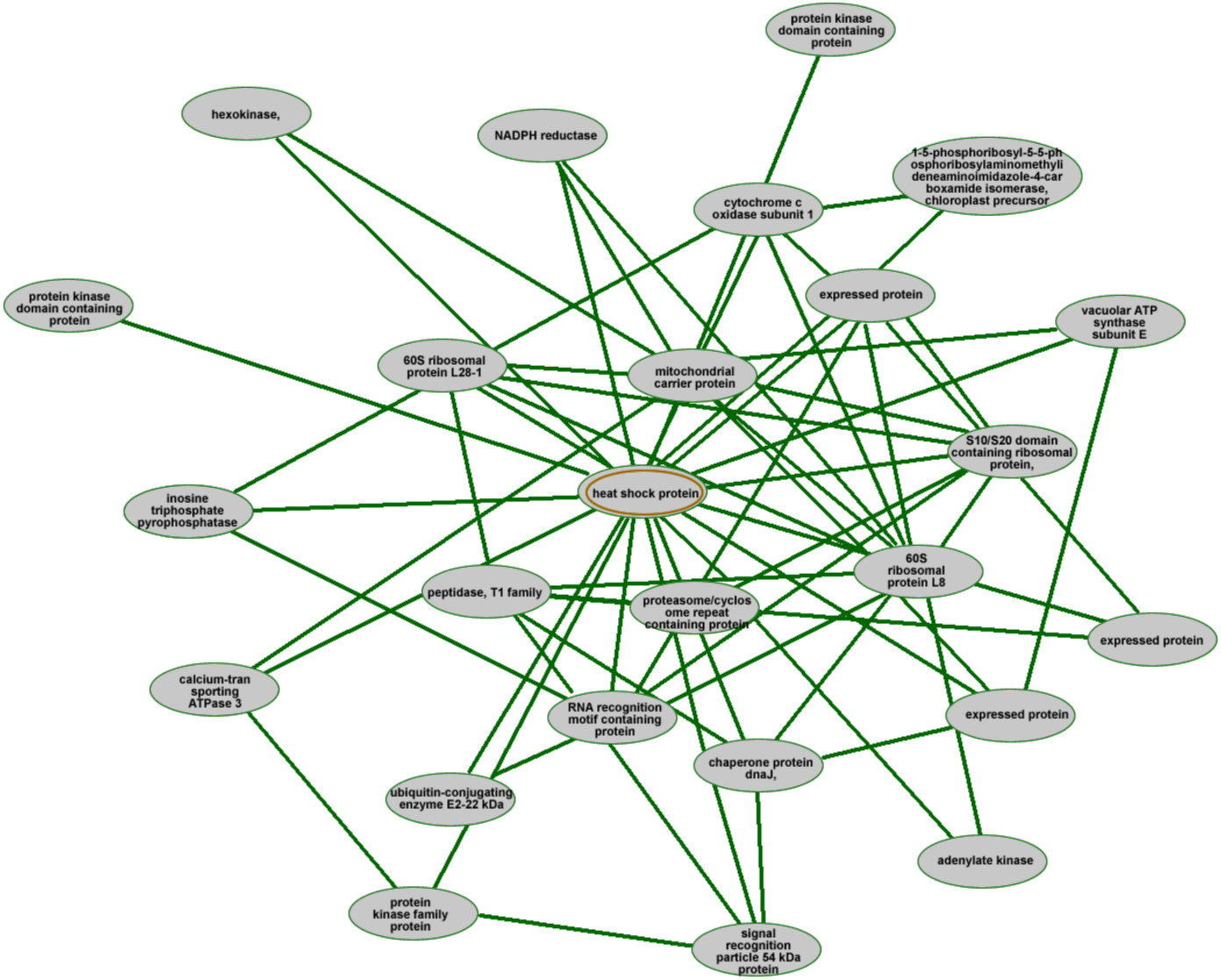


Fig 4.16: Functional network of salt tolerance regulators by trans-eQTLs found in chromosome 1 in reproductive leaf under stressed condition (Table 4.5)

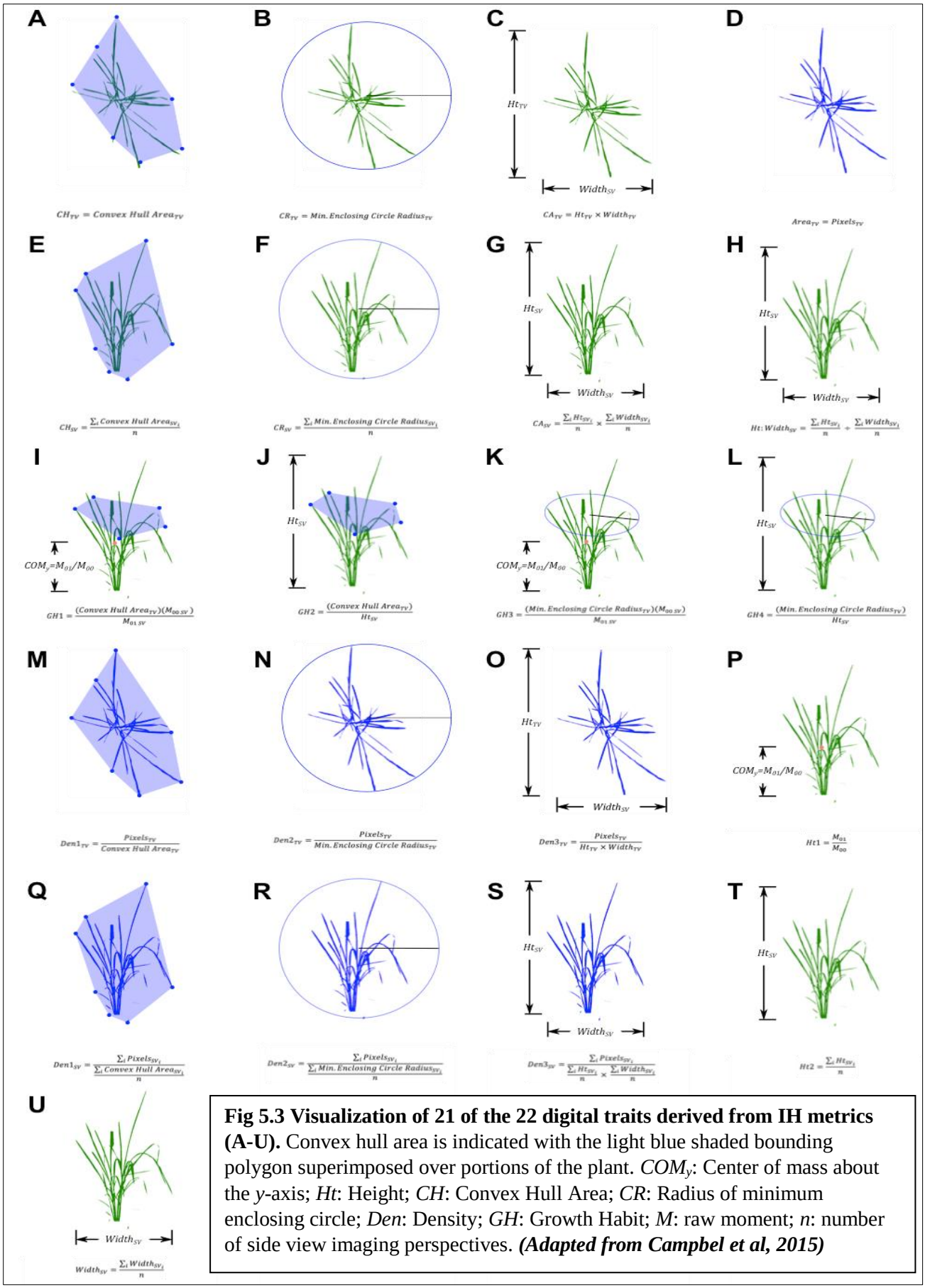


Fig 5.3 Visualization of 21 of the 22 digital traits derived from IH metrics (A-U). Convex hull area is indicated with the light blue shaded bounding polygon superimposed over portions of the plant. COM_y: Center of mass about the *y*-axis; *Ht*: Height; *CH*: Convex Hull Area; *CR*: Radius of minimum enclosing circle; *Den*: Density; *GH*: Growth Habit; *M*: raw moment; *n*: number of side view imaging perspectives. ***(Adapted from Campbel et al, 2015)***

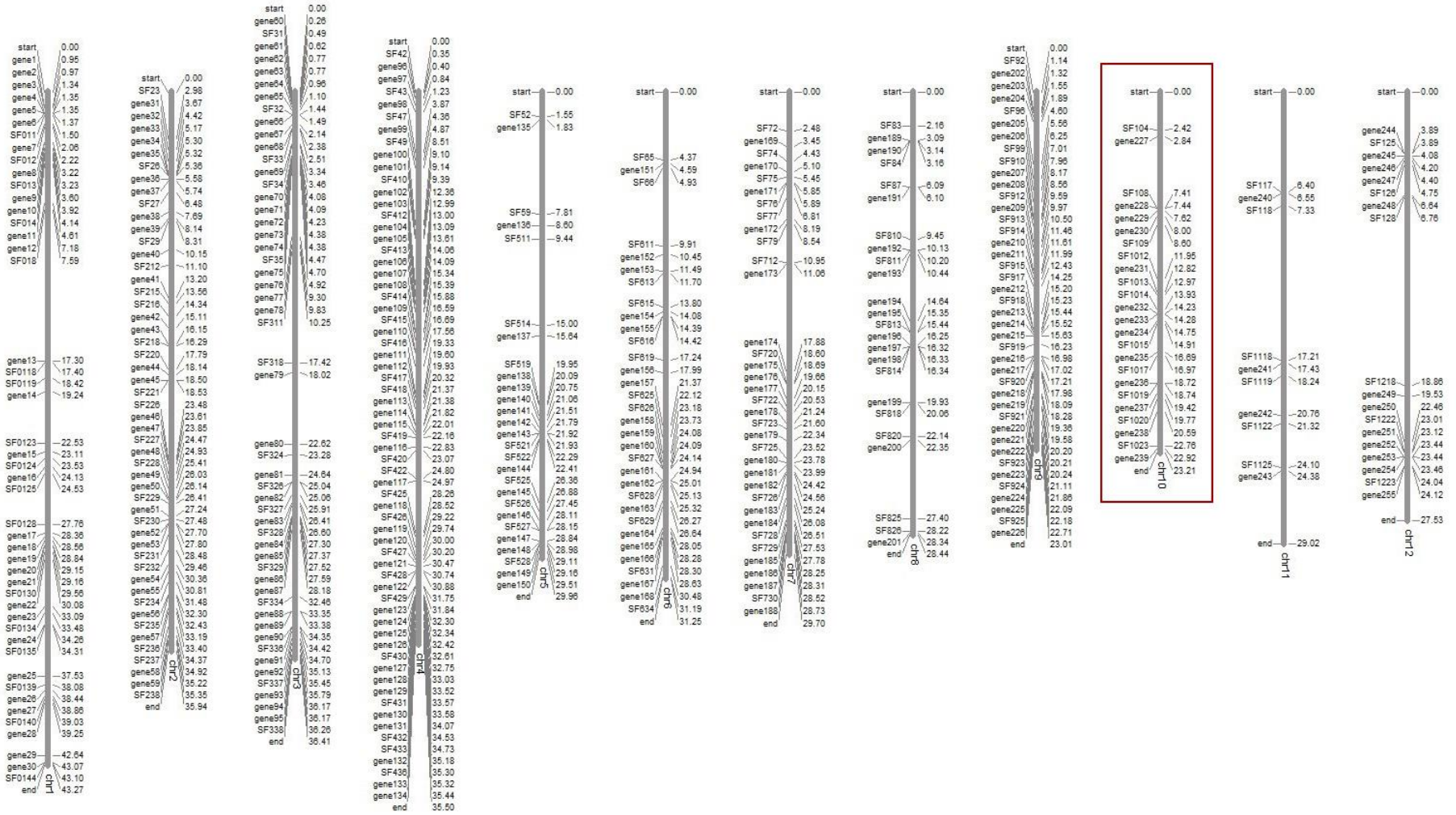


Fig 6.2: Positions of the cis-eQTLs regions chosen for further

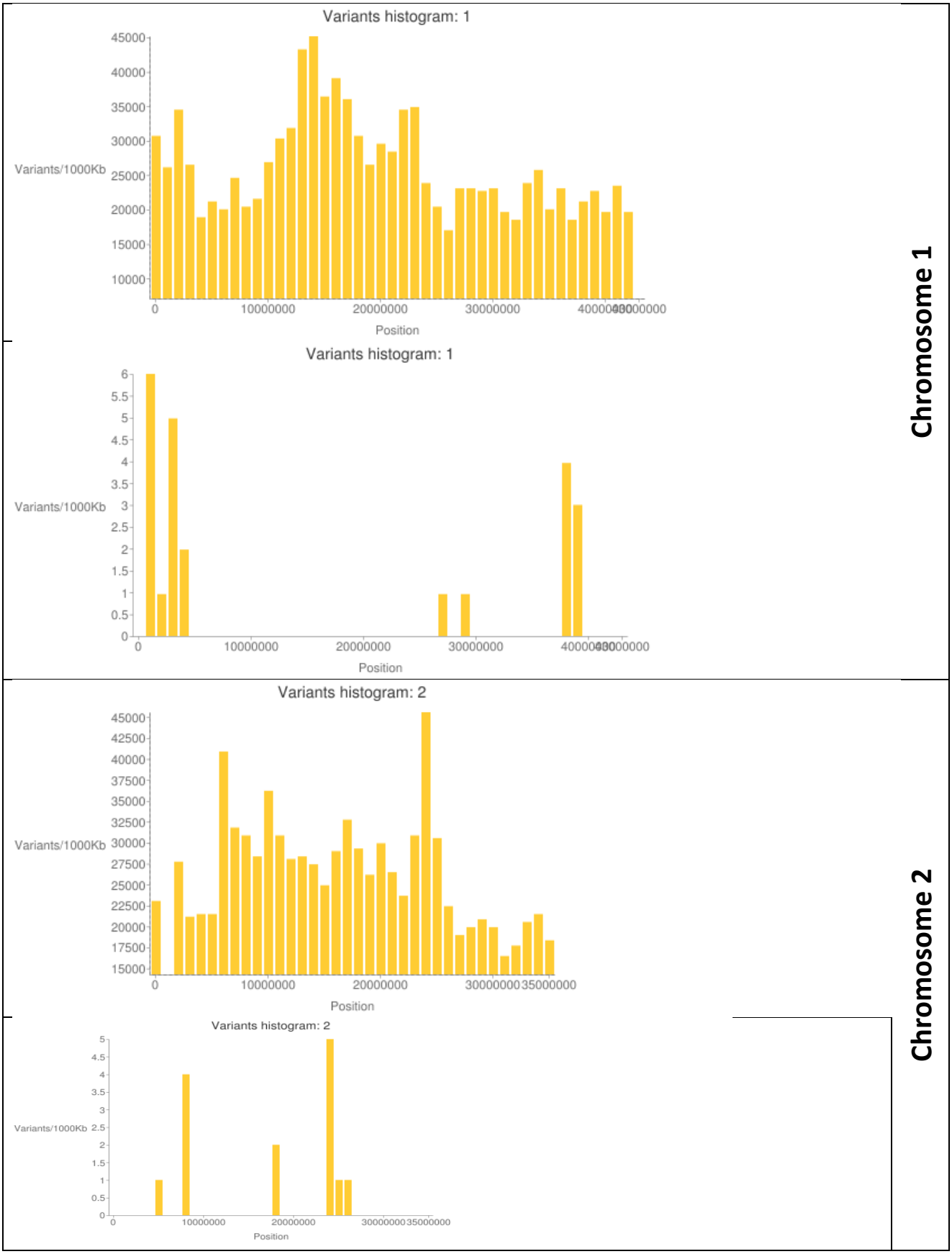
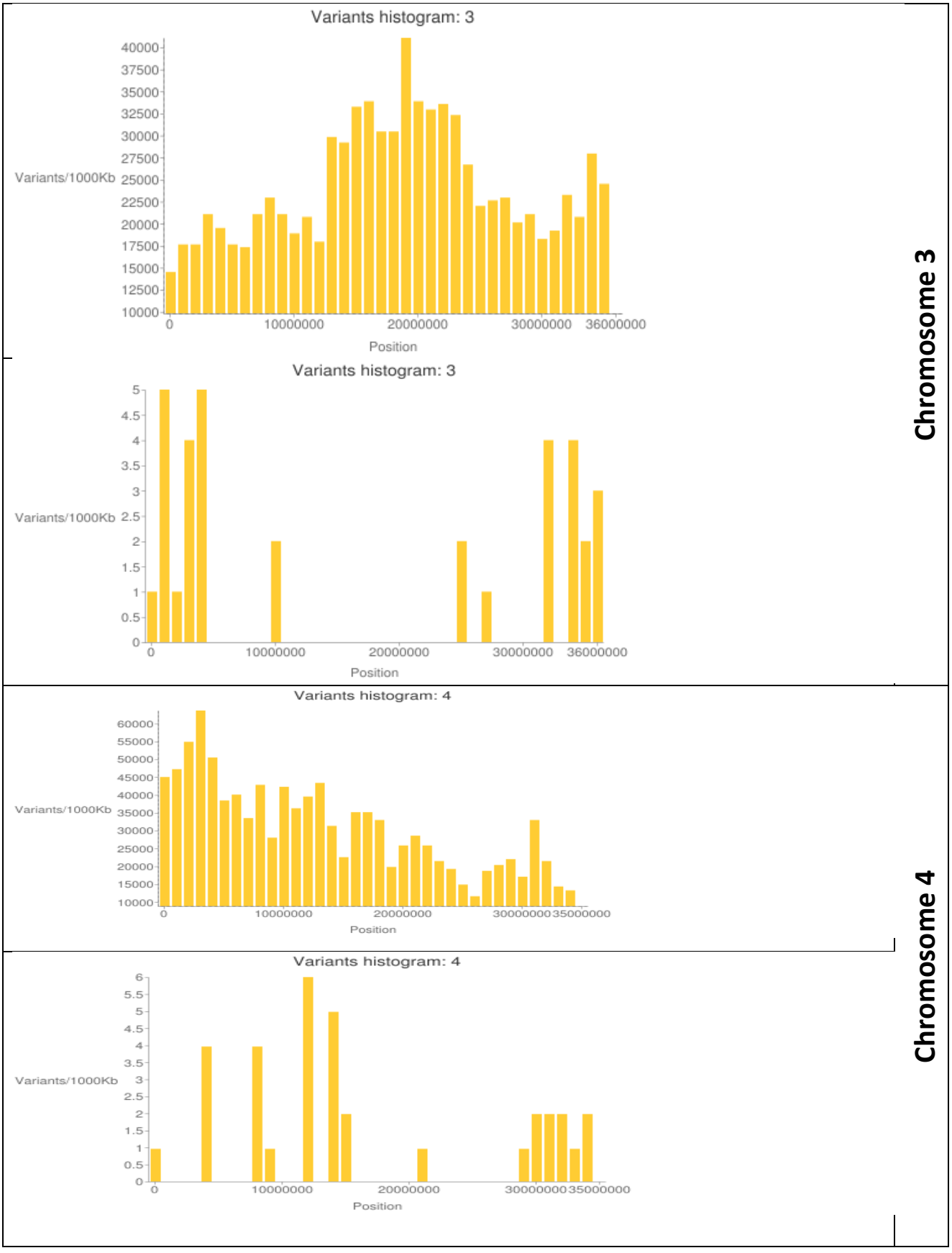
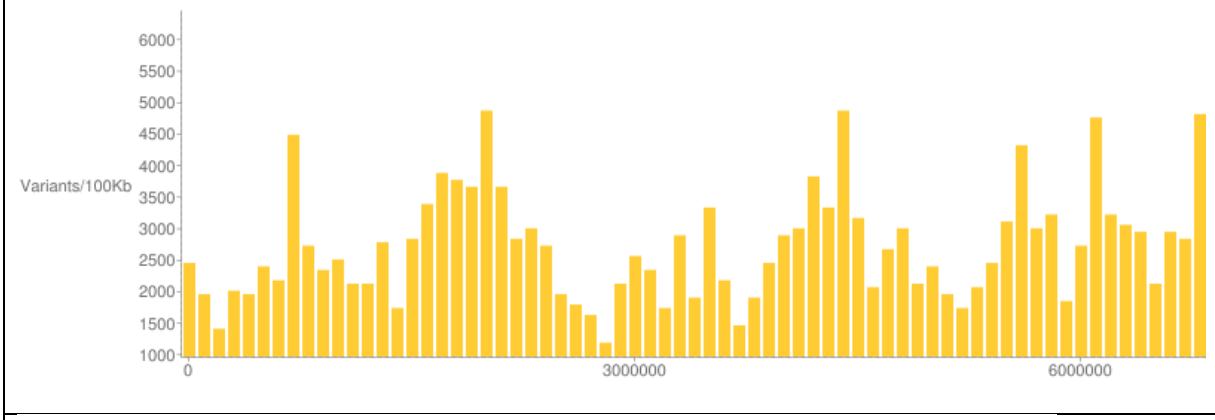
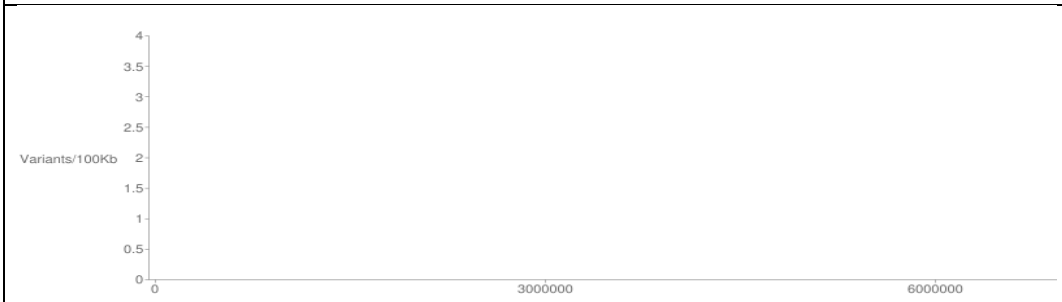


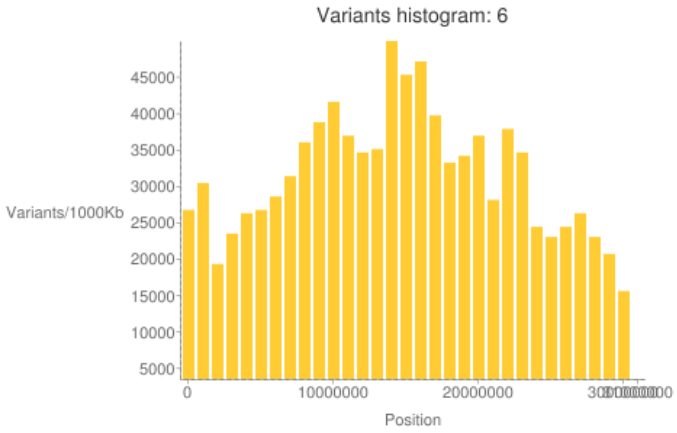
Fig 6.3: Distribution of variants across chromosomes. For each chromosome block, the upper row depicts all variants in the genome and the bottom row depicts only the variants found in eQTL regions.

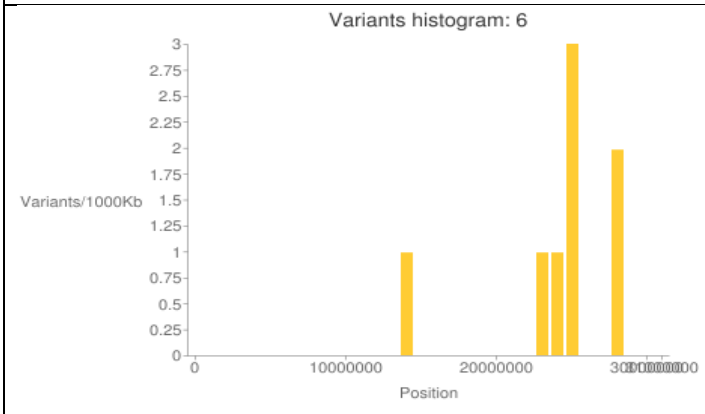

Variants histogram: 3
Variants/1000Kb
Position
Variants histogram: 3
Variants/1000Kb
Position
Chromosome 3
Variants histogram: 4
Variants/1000Kb
Position
Variants histogram: 4
Variants/1000Kb
Position
Chromosome 4





Chromosome 5





Chromosome 6

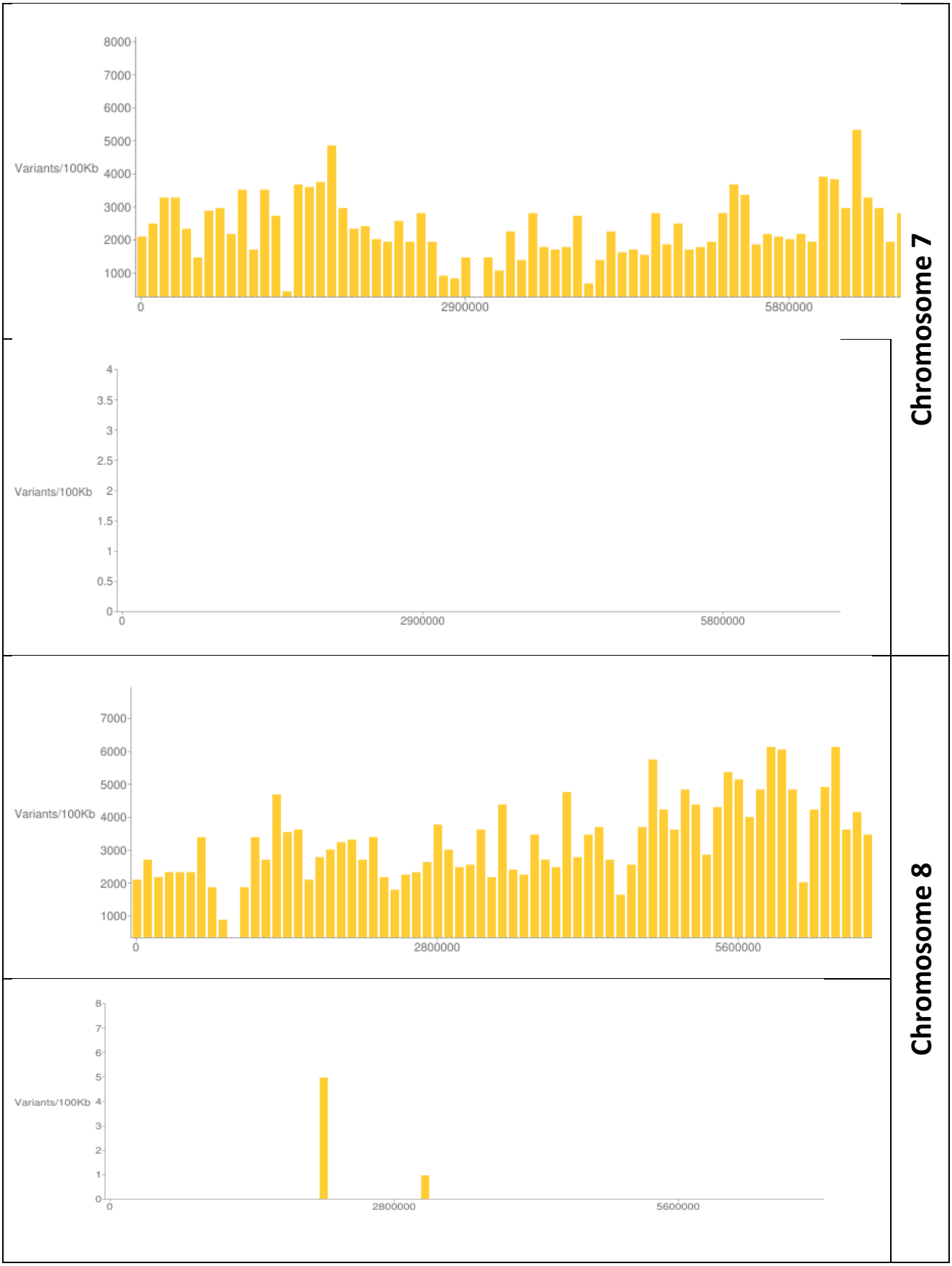


Fig 6.3: Distribution of variants across chromosomes. For each chromosome block, the upper row depicts all variants in the genome and the bottom row depicts only the variants found in eQTL regions.

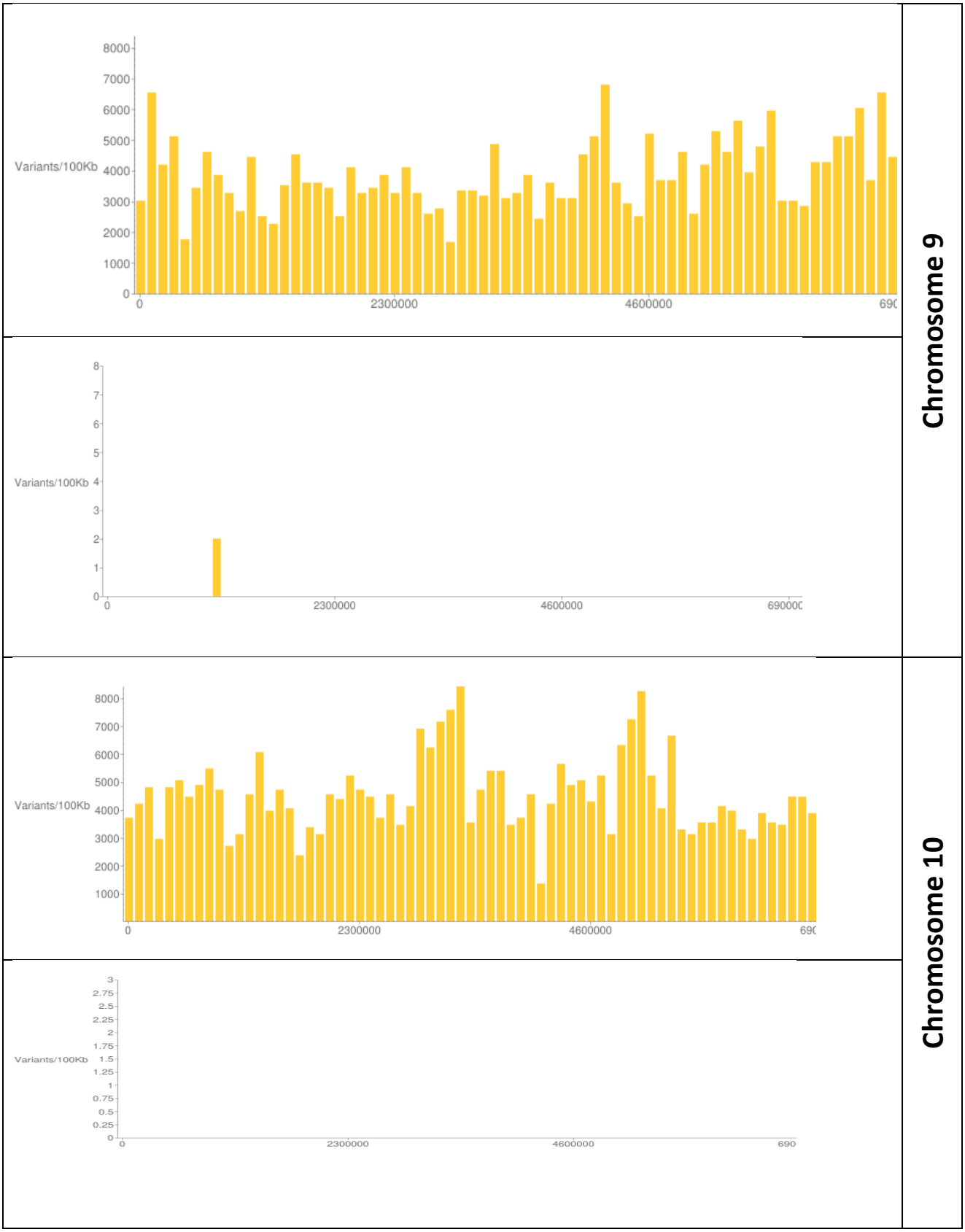


Fig 6.3: Distribution of variants across chromosomes. For each chromosome block, the upper row depicts all variants in the genome and the bottom row depicts only the variants found in eQTL regions.

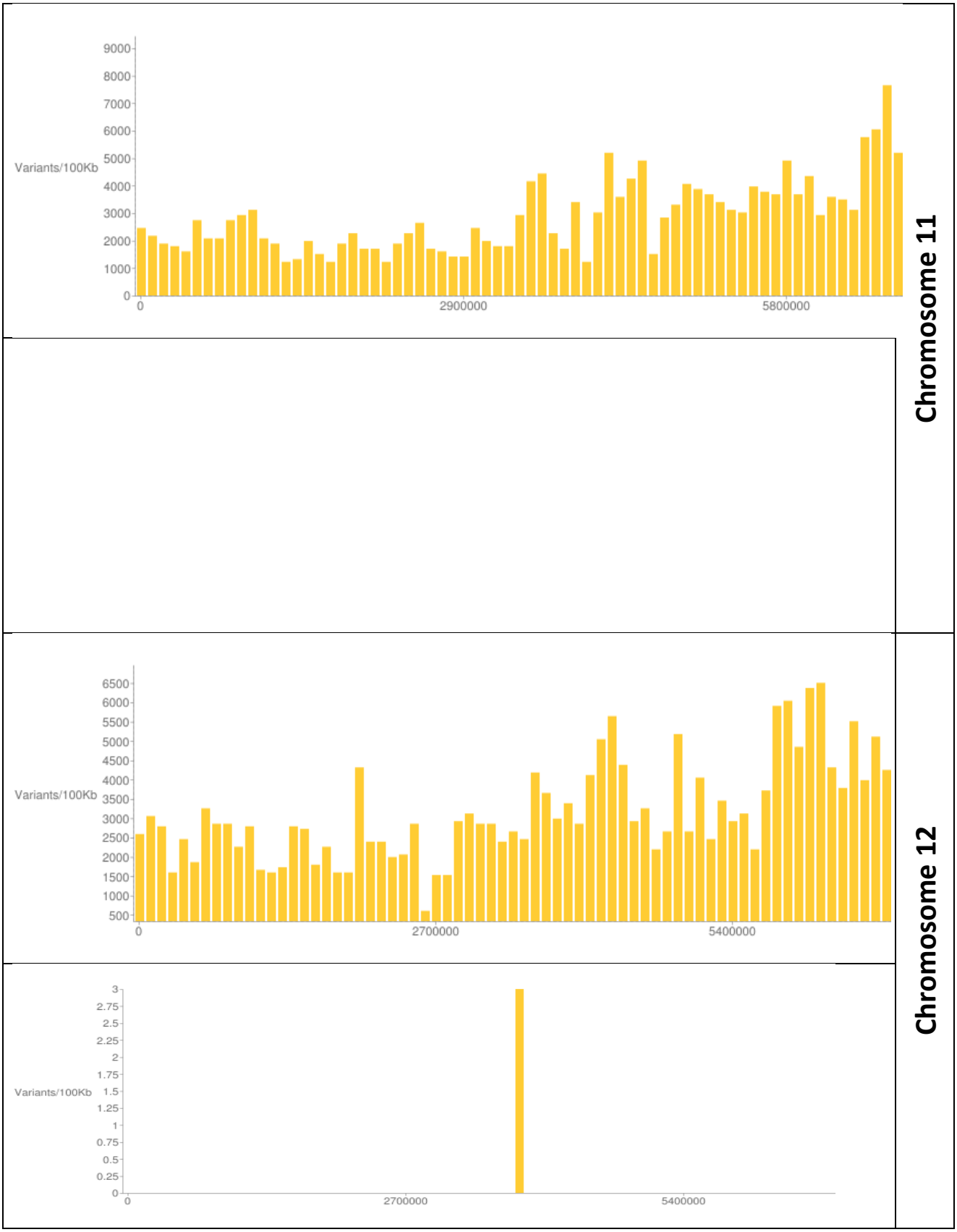
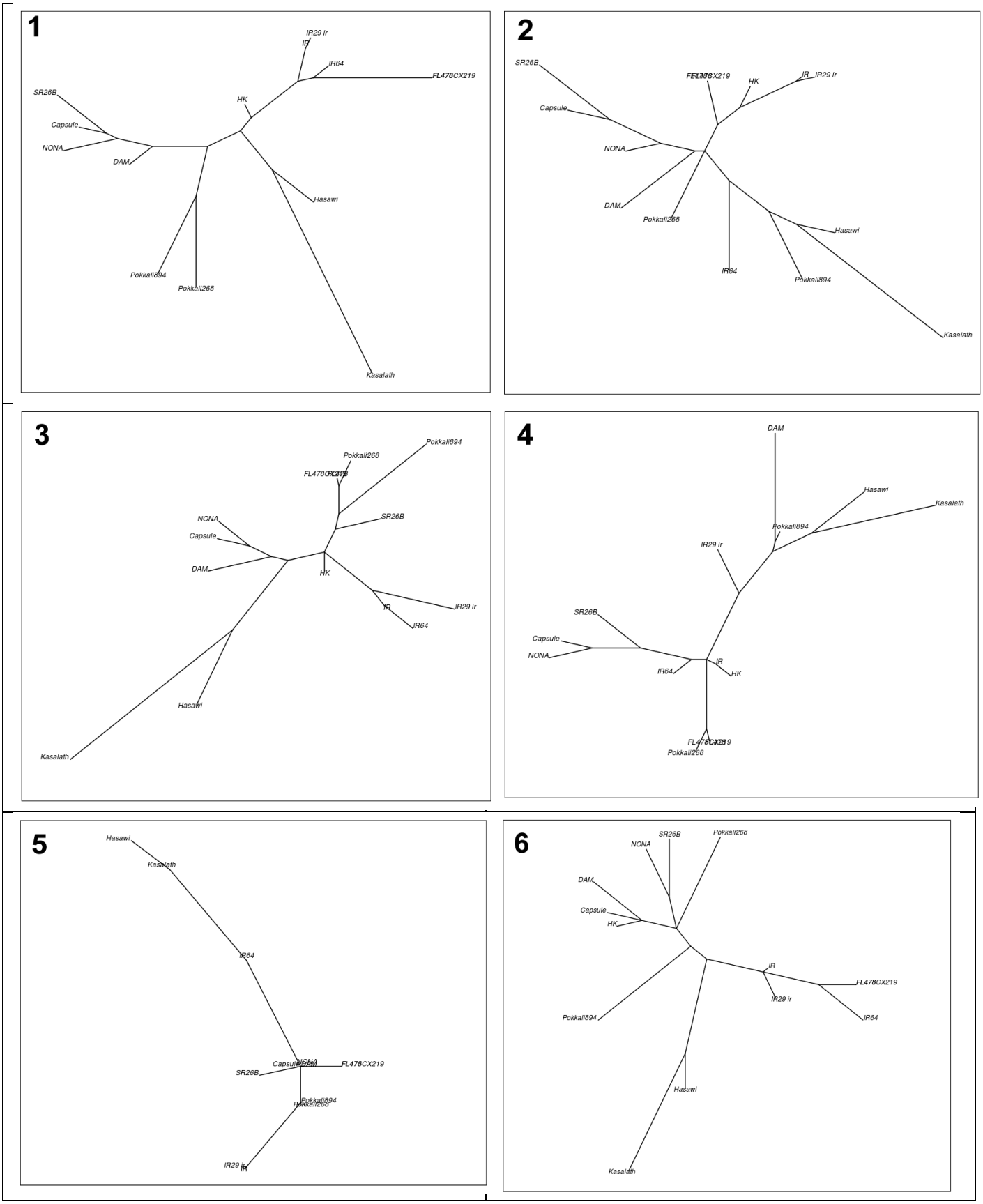


Fig 6.3: Distribution of variants across chromosomes. For each chromosome block, the upper row depicts all variants in the genome and the bottom row depicts only the variants found in eQTL regions.


1
IR29 ir
IR
IR64
FL478CX219
SR26B
HK
Capsule
NONA
DAM
Hasawi
Pokkali894
Pokkali268
Kasalath
2
SR26B
FL478CX219
HK
IR
IR29 ir
Capsule
NONA
DAM
Pokkali268
Hasawi
IR64
Pokkali894
Kasalath
3
Pokkali894
Pokkali268
FL478CX219
SR26B
NONA
Capsule
DAM
HK
IR
IR29 ir
IR64
Hasawi
Kasalath
4
DAM
Hasawi
Kasalath
Pokkali894
IR29 ir
SR26B
Capsule
NONA
IR
IR64
HK
FL478CX219
Pokkali268
5
Hasawi
Kasalath
IR64
SR26B
FL478CX219
Pokkali894
IR29 ir
IR
6
SR26B
Pokkali268
NONA
DAM
Capsule
HK
IR
FL478CX219
IR29 ir
Pokkali894
IR64
Hasawi
Kasaldh

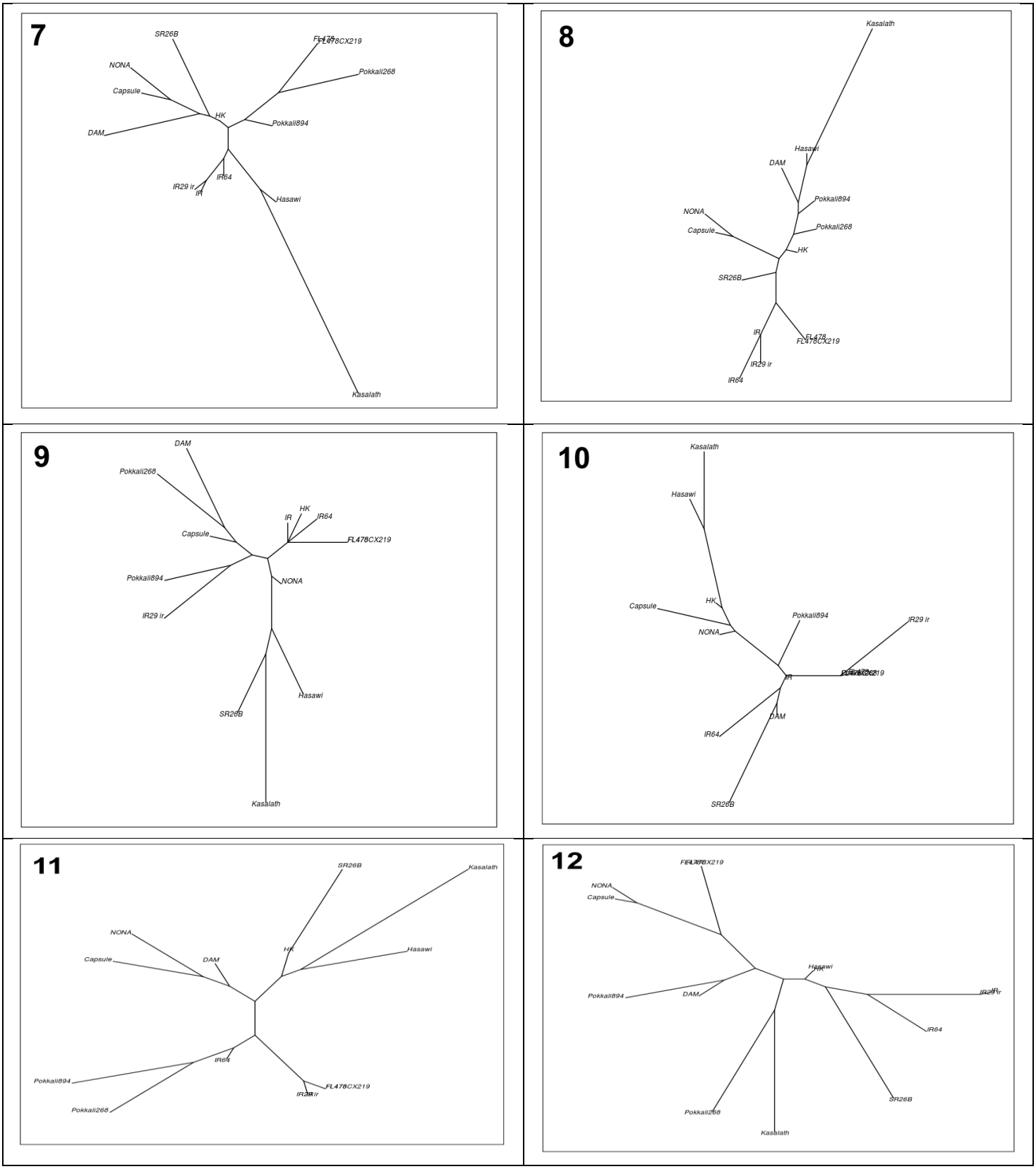


Fig 6.5: Phylogenetic relationship based on identified variants among salt tolerant rice genomes in individual chromosome. Horkuch marked with red circle and IR29 with blue circle

References

(2014). "The 3,000 rice genomes project." GigaScience **3**(1): 7.
Ahmed, I. M., F. Cao, et al. (2013). "Difference in Yield and Physiological Features in Response to Drought and Salinity Combined Stress during Anthesis in Tibetan Wild and Cultivated Barleys." PLOS ONE **8**(10): e77869.
Akbari, M., P. Wenzl, et al. (2006). "Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome." Theoretical and Applied Genetics **113**(8): 1409-1420.
Al-Tamimi, N., C. Brien, et al. (2016). "Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping." **7**: 13342.
Amin, M., S. M. Elias, et al. (2012). "Over-expression of a DEAD-box helicase, PDH45, confers both seedling and reproductive stage salinity tolerance to rice (Oryza sativa L.)." Molecular Breeding **30**(1): 345-354.
Amin, U. S. M., S. Biswas, et al. (2016). "Enhanced Salt Tolerance Conferred by the Complete 2.3 kb cDNA of the Rice Vacuolar Na(+)/H(+) Antiporter Gene Compared to 1.9 kb Coding Region with 5′ UTR in Transgenic Lines of Rice." Frontiers in Plant Science **7**: 14.
Anderson, C. M., T. A. Wagner, et al. (2001). "WAKs: cell wall-associated kinases linking the cytoplasm to the extracellular matrix." Plant Molecular Biology **47**(1): 197-206.
Andrews, S. " FastQC: A quality control tool for high throughput sequence data."

Asch, F. and M. C. S. Wopereis (2001). "Responses of field-grown irrigated rice cultivars to varying levels of floodwater salinity in a semi-arid environment." Field Crops Research **70**(2): 127-137.
Ashraf, M. and N. A. Akram (2009). "Improving salinity tolerance of plants through conventional breeding and genetic engineering: An analytical comparison." Biotechnology Advances **27**(6): 744-752.
Babu, N. N., S. G. Krishnan, et al. (2017). "Marker Aided Incorporation of Saltol, a Major QTL Associated with Seedling Stage Salt Tolerance, into Oryza sativa 'Pusa Basmati 1121'." Frontiers in Plant Science **8**: 41.
Baloch, F. S., A. Alsaleh, et al. (2017). "A Whole Genome DArTseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat from Central Fertile Crescent." PLOS ONE **12**(1): e0167821.
Barbieri, G., S. Vallone, et al. (2012). "Stomatal density and metabolic determinants mediate salt stress adaptation and water use efficiency in basil (Ocimum basilicum L.)." Journal of Plant Physiology **169**(17): 1737-1746.
Barrs, H. and P. Weatherley (1962). "A Re-Examination of the Relative Turgidity Technique for Estimating Water Deficits in Leaves." Australian Journal of Biological Sciences **15**(3): 413-428.
Berger, B., B. de Regt, et al. (2012). High-Throughput Phenotyping of Plant Shoots. High-Throughput Phenotyping in Plants: Methods and Protocols. J. Normanly. Totowa, NJ, Humana Press**:** 9-20.
Blumwald, E. (1987). "Tonoplast vesicles as a tool in the study of ion transport at the plant vacuole." Physiologia Plantarum **69**(4): 731-734.
Blumwald, E., G. S. Aharon, et al. (2000). "Sodium transport in plant cells." Biochimica et Biophysica Acta (BBA) - Biomembranes **1465**(1): 140-151.
Bonilla, P. P. R. R. I., Munoz, Nueva Ecija (Philippines)) Dvorak, J., D. Mackell, et al. (2002). "RFLP and SSLP mapping of salinity tolerance genes in chromosome 1 of rice (Oryza sativa L.) using recombinant inbred lines." Philippine Agricultural Scientist (Philippines) **v. 85**.
Boonburapong, B. and T. Buaboocha (2007). "Genome-wide identification and analyses of the rice calmodulin and related potential calcium sensor proteins." BMC Plant Biology **7**(1): 4.
Broman, K. W., H. Wu, et al. (2003). "R/qtl: QTL mapping in experimental crosses." Bioinformatics **19**(7): 889-890.

Brown, T. B., R. Cheng, et al. (2014). "TraitCapture: genomic and environment modelling of plant phenomic data." Current Opinion in Plant Biology **18**(Supplement C): 73-79.

Buchanan, G. W, et al. (2002). Biochemistry & Molecular Biology of Plants. , American Society of Plant Physiology.

Burr, B., F. A. Burr, et al. (1988). "Gene mapping with recombinant inbreds in maize." Genetics **118**(3): 519.

Byrt, C. S., M. Zhao, et al. (2017). "Non-selective cation channel activity of aquaporin AtPIP2;1 regulated by Ca2+ and pH." Plant Cell Environ **40**(6): 802-815.

Campbell, M. T., Q. Du, et al. (2017). "A Comprehensive Image-based Phenomic Analysis Reveals the Complex Genetic Architecture of Shoot Growth Dynamics in Rice (Oryza sativa)." The Plant Genome **10**(2).

Campbell, M. T., A. C. Knecht, et al. (2015). "Integrating Image-Based Phenomics and Association Analysis to Dissect the Genetic Architecture of Temporal Salinity Responses in Rice." Plant Physiology **168**(4): 1476.

Cao, P., K.-H. Jung, et al. (2012). "The Rice Oligonucleotide Array Database: an atlas of rice gene expression." Rice **5**: 17.

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

Chaves, M. M., J. Flexas, et al. (2009). "Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell." Annals of Botany **103**(4): 551-560.

Chen, D., K. Neumann, et al. (2015). "Dissecting the Phenotypic Components of Crop Plant Growth and Drought Responses Based on High-Throughput Image Analysis." The Plant Cell **26**(12): 4636.

Chitteti, B. R. and Z. Peng (2007). "Proteome and Phosphoproteome Differential Expression under Salinity Stress in Rice (Oryza sativa) Roots." J Proteome Res **6**(5): 1718-1727.

Chowdhury, A. K. M. H. U., M. E. Haque, et al. (2012). "Adoption of BRRI Dhan47 in the Coastal Saline areas of Bangladesh." Agricultural Journal **7**: 286-291.

Cingolani, P. (2012). snpEff: Variant effect prediction.

Collard, B. C. Y., M. Z. Z. Jahufer, et al. (2005). "An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts." Euphytica **142**(1): 169-196.

Dai, X., Y. Xu, et al. (2007). "Overexpression of an R1R2R3 MYB Gene, OsMYB3R-2, Increases Tolerance to Freezing, Drought, and Salt Stress in Transgenic Arabidopsis." Plant Physiology **143**(4): 1739.

Danecek, P., A. Auton, et al. (2011). "The variant call format and VCFtools." Bioinformatics **27**(15): 2156-2158.

Das, P., K. K. Nutan, et al. (2015). "Understanding salinity responses and adopting 'omics-based' approaches to generate salinity tolerant cultivars of rice." Frontiers in Plant Science **6**: 712.

Dasgupta, S., M. M. Hossain, et al. (2014). "Climate Change, Soil Salinity, and the Economics of High-Yield Rice Production in Coastal Bangladesh " World Bank Policy Research Working Paper No. 7140.

Davenport, R. J., A. MuÑOz-Mayor, et al. (2007). "The Na+ transporter AtHKT1;1 controls retrieval of Na+ from the xylem in Arabidopsis." Plant, Cell & Environment **30**(4): 497-507.

De Leon, T. B., S. Linscombe, et al. (2015). "Genetic variation in Southern USA rice genotypes for seedling salinity tolerance." Frontiers in Plant Science **6**: 374.

DeFalco, Thomas A., Kyle W. Bender, et al. (2009). "Breaking the code: Ca²⁺ sensors in plant signalling." Biochemical Journal **425**(1): 27.

Delseny, M., B. Han, et al. (2010). "High throughput DNA sequencing: The new sequencing revolution." Plant Science **179**(5): 407-422.

Demidchik, V. and F. J. M. Maathuis (2007). "Physiological roles of nonselective cation channels in plants: from salt stress to signalling and development." New Phytologist **175**(3): 387-404.

Demiral, T. and I. Türkan (2006). "Exogenous glycinebetaine affects growth and proline accumulation and retards senescence in two rice cultivars under NaCl stress." Environmental and Experimental Botany **56**(1): 72-79.

Des Marais, D. L., W. D. Skillern, et al. (2015). "Deeply Diverged Alleles in the Arabidopsis AREB1 Transcription Factor Drive Genome-Wide Differences in Transcriptional Response to the Environment." Molecular Biology and Evolution **32**(4): 956-969.

Doerge, R. W. (2002). "Mapping and analysis of quantitative trait loci in experimental populations." Nat Rev Genet **3**(1): 43-52.

Doyle, J. J. and J. L. Doyle (1990). "Isolation of plant DNA from fresh tissue." Focus **12**(1): 13-15.

Edwards, J. W. I. S. U., Ames, IA.), J. O. Allen, et al. (1996). "Teosinte cytoplasmic genomes. I. Performance of maize inbreds with teosinte cytoplasms." Crop science (USA) **v. 36**.

Elert, E. (2014). "Rice by the numbers: A good grain." Nature **514**(7524): S50-S51.

Fahlgren, N., M. Feldman, et al. (2015). "A Versatile Phenotyping System and Analytics Platform Reveals Diverse Temporal Responses to Water Availability in Setaria." Molecular Plant **8**(10): 1520-1535.

Fang, Y., K. Xie, et al. (2010). "Systematic analysis of GT factor family of rice reveals a novel subfamily involved in stress responses." Molecular Genetics and Genomics **283**(2): 157-169.

Ferreira, A., M. F. d. Silva, et al. (2006). "Estimating the effects of population size and type on the accuracy of genetic maps." Genetics and Molecular Biology **29**: 187-192.

Flassig, R. J., S. Heise, et al. (2013). "An effective framework for reconstructing gene regulatory networks from genetical genomics data." Bioinformatics **29**(2): 246-254.

Flowers, T. J. and T. D. Colmer (2015). "Plant salt tolerance: adaptations in halophytes." Annals of Botany **115**(3): 327-331.

Flowers, T. J., M. L. Koyama, et al. (2000). "QTL: their place in engineering tolerance of rice to salinity." Journal of Experimental Botany **51**(342): 99-106.

Fu, W., K. Wu, et al. (2007). "Sequence and expression analysis of histone deacetylases in rice." Biochemical and Biophysical Research Communications **356**(4): 843-850.

Fuchs, I., S. Stölzle, et al. (2005). "Rice K+ uptake channel OsAKT1 is sensitive to salt stress." Planta **221**(2): 212-221.

Fukuda, A., A. Nakamura, et al. (2004). "Function, Intracellular Localization and the Importance in Salt Tolerance of a Vacuolar Na+/H+ Antiporter from Rice." Plant and Cell Physiology **45**(2): 146-159.

Furbank, R. T. and M. Tester (2011). "Phenomics – technologies to relieve the phenotyping bottleneck." Trends in Plant Science **16**(12): 635-644.

Gamuyao, R., J. H. Chin, et al. (2012). "The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency." Nature **488**(7412): 535-539.

Garcia, A. B., J. Engler, et al. (1997). "Effects of Osmoprotectants upon NaCl Stress in Rice." Plant Physiology **115**(1): 159.

Garcia, A. B., J. A. Engler, et al. (1998). "The expression of the salt-responsive gene salT from rice is regulated by hormonal and developmental cues." Planta **207**(2): 172-180.

Garciadeblas, B., M. E. Senn, et al. (2003). "Sodium transport and HKT transporters: the rice model." Plant J **34**(6): 788-801.

Gholizadeh, A., H. Dehghani, et al. (2014). "Determination of the Most Effective Traits on Wheat Yield under Saline Stress. ." Agricultural Advances, North America.

Golldack, D., F. Quigley, et al. (2003). "Salinity stress-tolerant and -sensitive rice (Oryza sativa L.) regulate AKT1-type potassium channel transcripts differently." Plant Molecular Biology **51**(1): 71-81.

Grattan, S. R., L. Zeng, et al. (2002). "Rice is more sensitive to salinity than previously thought." California Agriculture **56**(6): 189-198.

Gregorio, G. B., Islam, M.R., Vergara, G.V. and Thirumeni, S (2013). "Recent advances in rice science to design salinity and other abiotic stress tolerant rice varieties. ." SABRAO Journal of Breeding and Genetics **45**(1): 31-41.

Gregorio, G. B., D. Senadhira, et al. (1997). "Screening rice for salinity tolerance." IRRI Discussion Paper Series No. 22. International Rice Research Institute, Los Baños,Philippines.

Gregorio, G. B., D. Senadhira, et al. (2002). "Progress in breeding for salinity tolerance and associated abiotic stresses in rice." Field Crops Research **76**(2): 91-101.

Grzebelus (2015). Diversity Arrays Technology (DArT) Markers for Genetic Diversity, Plants. Sustainable Development and Biodiversity, vol 7. Springer, Cham.

Hairmansis, A., B. Berger, et al. (2014). "Image-based phenotyping for non-destructive screening of different salinity tolerance traits in rice." Rice **7**(1): 16.

Hansen, B. G., B. A. Halkier, et al. (2008). "Identifying the molecular basis of QTLs: eQTLs add a new dimension." Trends in Plant Science **13**(2): 72-77.

Harris, B. N., V. O. Sadras, et al. (2010). "A water-centred framework to assess the effects of salinity on the growth and yield of wheat and barley." Plant and Soil **336**(1): 377-389.

Hartl, D. and E. Jones (2001). Genetics: Analysis of Genes and Genomes, Jones and Bartlett Publishers, Sudbury, MA.

Hasan, M. M., M. Y. Rafii, et al. (2015). "Marker-assisted backcrossing: a useful method for rice improvement." Biotechnology, Biotechnological Equipment **29**(2): 237-254.

Hiscox JD, Israelstam, GF. (1979). a method for the extraction of chlorophyll from leaf tissue without maceration. canadian journal of botany 57: 1332-133

Horie, T., A. Costa, et al. (2007). "Rice OsHKT2;1 transporter mediates large Na(+) influx component into K(+)-starved roots for growth." The EMBO Journal **26**(12): 3003-3014.

Horiuchi, Y., Y. Harushima, et al. (2015). "Global expression differences and tissue specific expression differences in rice evolution result in two contrasting types of differentially expressed genes." BMC Genomics **16**: 1099.

Hu, H., M. Dai, et al. (2006). "Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice." Proceedings of the National Academy of Sciences **103**(35): 12987-12992.

Hu, T. Z. (2008). "OsLEA3, a late embryogenesis abundant protein gene from rice, confers tolerance to water deficit and salt stress to transgenic rice." Russian Journal of Plant Physiology **55**(4): 530-537.

Huyen, L. T. N., L. M. Cuc, et al. (2013). "Introgression the SALTOL QTL into Q5DB, the Elite Variety of Vietnam Using Marker- Assisted - Selection (MAS)." American Journal of BioScience. **1**(4): 80-84.

Islam, M. R., M. R. A. Sarker, et al. (2016). "Assessment of adaptability of recently released salt tolerant rice varieties in coastal regions of South Bangladesh." Field Crops Research **190**: 34-43.

Ismail, A., M. Seo, et al. (2014). "Salt adaptation requires efficient fine-tuning of jasmonate signalling." Protoplasma **251**(4): 881-898.

Ismail, A. and T. Tuong (2009). Brackish water coastal zones of the monsoon tropics: challenges and opportunities. Natural resource management for poverty reduction and environmental sustainability in rice-based systems. S. Haefele and A. Ismail. Los Banos, Philippines, International Rice Research Institute.

Ismail, A. M. and T. Horie (2017). "Genomics, Physiology, and Molecular Breeding Approaches for Improving Salt Tolerance." Annual Review of Plant Biology **68**(1): 405-434.

Jaccoud, D., K. Peng, et al. (2001). "Diversity Arrays: a solid state technology for sequence information independent genotyping." Nucleic Acids Research **29**(4): e25-e25.

Jansen, R. C. (1993). "Interval mapping of multiple quantitative trait loci." Genetics **135**(1): 205.

Jordan, M. C., D. J. Somers, et al. (2007). "Identifying regions of the wheat genome controlling seed development by mapping expression quantitative trait loci†." Plant Biotechnol J **5**(3): 442-453.

Kacperska, A. (2004). "Sensor types in signal transduction pathways in plant cells responding to abiotic stressors: do they depend on stress intensity?" Physiologia Plantarum **122**(2): 159-168.

Karthikeyan, A., S. K. Pandian, et al. (2011). "Transgenic indica rice cv. ADT 43 expressing a Δ1-pyrroline-5-carboxylate synthetase (P5CS) gene from Vigna aconitifolia demonstrates salt tolerance." Plant Cell, Tissue and Organ Culture (PCTOC) **107**(3): 383-395.

Kasuga, M., Q. Liu, et al. (1999). "Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor." Nat Biotech **17**(3): 287-291.

Keurentjes, J. J. B., J. Fu, et al. (2007). "Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci." Proceedings of the National Academy of Sciences **104**(5): 1708-1713.

Khatun, S., C. A. Rizzo, et al. (1995). "Genotypic variation in the effect of salinity on fertility in rice." Plant and Soil(173): 239-250.

Kliebenstein, D. (2009). "Quantitative Genomics: Analyzing Intraspecific Variation Using Global Gene Expression Polymorphisms or eQTLs." Annual Review of Plant Biology **60**(1): 93-114.

Kliebenstein, D. J., M. A. L. West, et al. (2006). "Genomic Survey of Gene Expression Diversity in <em>Arabidopsis thaliana</em&gt." Genetics **172**(2): 1179.

Klukas, C., D. Chen, et al. (2014). "Integrated Analysis Platform: An Open-Source Information System for High-Throughput Plant Phenotyping." Plant Physiology **165**(2): 506.

Knecht, A. C., M. T. Campbell, et al. (2016). "Image Harvest: an open-source platform for high-throughput plant image processing and analysis." Journal of Experimental Botany **67**(11): 3587-3599.

Korte, A. and A. Farlow (2013). "The advantages and limitations of trait analysis with GWAS: a review." Plant Methods **9**(1): 29.

Kover, P. X., W. Valdar, et al. (2009). "A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in Arabidopsis thaliana." PLOS Genetics **5**(7): e1000551.

Koyama, M. L., A. Levesley, et al. (2001). "Quantitative Trait Loci for Component Physiological Traits Determining Salt Tolerance in Rice." Plant Physiology **125**(1): 406.

Krishnamurthy, P., K. Ranathunge, et al. (2009). "The role of root apoplastic transport barriers in salt tolerance of rice (Oryza sativa L.)." Planta **230**(1): 119-134.

Kumari, S., V. Panjabi nee Sabharwal, et al. (2008). "Transcriptome map for seedling stage specific salinity stress response indicates a specific set of genes as candidate for saline tolerance in Oryza sativa L." Functional & Integrative Genomics **9**(1): 109.

Latha, R., L. Rubia, et al. (2004). "Allele mining for stress tolerance genes in Oryza species and related germplasm." Molecular Biotechnology **27**(2): 101-108.

Lecomte, L., P. Duffe, et al. (2004). "Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds." Theor Appl Genet **109**(3): 658-668.

Lee, S. C., W.-Z. Lan, et al. (2007). "A protein phosphorylation/dephosphorylation network regulates a plant potassium channel." Proceedings of the National Academy of Sciences **104**(40): 15959-15964.

Lee, T.-H., H. Guo, et al. (2014). "SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data." BMC genomics **15**(1): 162.

Lee, T., T. Oh, et al. (2015). "RiceNet v2: an improved network prioritization server for rice genes." Nucleic Acids Res **43**(W1): W122-127.

o construct a phylogenetic tree from huge SNP data." BMC genomics 15(1): 162.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows–Wheeler transform." Bioinformatics 25(14): 1754-1760.

Li, H.-W., B.-S. Zang, et al. (2011). "Overexpression of the trehalose-6-phosphate synthase gene OsTPS1 enhances abiotic stress tolerance in rice." Planta **234**(5): 1007-1018.
Li, J.-Y., J. Wang, et al. (2014). "The 3,000 rice genomes project: new opportunities and challenges for future rice research." GigaScience **3**(1): 8.
Li, Z. and M. J. Sillanpää "Dynamic Quantitative Trait Locus Analysis of Plant Phenomic Data." Trends in Plant Science **20**(12): 822-833.
Li, H., B. Handsaker, et al. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics 25(16): 2078-2079.
Liao, Y.-D., K.-H. Lin, et al. (2016). "Oryza sativa protein phosphatase 1a (OsPP1a) involved in salt stress tolerance in transgenic rice." Molecular Breeding **36**(3): 22.
Lin, H. X., M. Z. Zhu, et al. (2004). "QTLs for Na+ and K+ uptake of the shoots and roots controlling rice salt tolerance." Theoretical and Applied Genetics **108**(2): 253-260.
Lin, S., T. Sasaki, et al. (1998). "Mapping quantitative trait loci controlling seed dormancy and heading date in rice, Oryza sativa L., using backcross inbred lines." Theor Appl Genet **96**.
Lisa, L. A., S. M. Elias, et al. (2011). "Physiology and gene expression of the rice landrace Horkuch under salt stress." Functional Plant Biology **38**(4): 282-292.
Lisa, L. A., Z. I. Seraj, et al. (2004). "Genetic variation in microsatellite DNA, physiology and morphology of coastal saline rice (Oryza sativa L.) landraces of Bangladesh." Plant and Soil **263**(1): 213-228.
Liu, B. H. (1998). Statistical Genomics: Linkage, Mapping, and QTL Analysis, CRC press.
Liu, F., W. Xu, et al. (2010). "Gene Expression Profiles Deciphering Rice Phenotypic Variation between Nipponbare (Japonica) and 93-11 (Indica) during Oxidative Stress." PLOS ONE **5**(1): e8632.
Lo, S.-F., S.-Y. Yang, et al. (2008). "A Novel Class of Gibberellin 2-Oxidases Control Semidwarfism, Tillering, and Root Development in Rice." The Plant Cell **20**(10): 2603-2618.
Lohman, B. K., J. N. Weber, et al. (2016). "Evaluation of TagSeq, a reliable low-cost alternative for RNAseq." bioRxiv.
Lovell, J. T. "qtlTools." from https://github.com/jtlovell/qtlTools.
Maathuis, F. J. M. and A. Amtmann (1999). "K + Nutrition and Na + Toxicity: The Basis of Cellular K + /Na + Ratios." Annals of Botany **84**(2): 123-133.
Maathuis, F. J. M. and D. Sanders (1996). "Mechanisms of potassium absorption by higher plant roots." Physiologia Plantarum **96**(1): 158-168.
Maggio, A., G. Raimondi, et al. (2007). "Salt stress response in tomato beyond the salinity tolerance threshold." Environmental and Experimental Botany **59**(3): 276-282.
Mallikarjuna, G., K. Mallikarjuna, et al. (2011). "Expression of OsDREB2A transcription factor confers enhanced dehydration and salt stress tolerance in rice (Oryza sativa L.)." Biotechnology Letters **33**(8): 1689-1697.
Martin, L. B. B., Z. Fei, et al. (2013). "Catalyzing plant science research with RNA-seq." Frontiers in Plant Science **4**: 66.
Martínez-Atienza, J., X. Jiang, et al. (2007). "Conservation of the Salt Overly Sensitive Pathway in Rice." Plant Physiology **143**(2): 1001.
Matysik, J., Alia, et al. (2002). "Molecular mechanisms of quenching of reactive oxygen species by proline under stress in plants." Current Science **82**(5): 525-532.
McKenna, A., M. Hanna, et al. (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome research **20**(9): 1297-1303.
McNally, K. L., K. L. Childs, et al. (2009). "Genomewide SNP variation reveals relationships among landraces and modern varieties of rice." Proceedings of the National Academy of Sciences **106**(30): 12273-12278.
Meyer, E., G. V. Aglyamova, et al. (2011). "Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure." Mol Ecol **20**(17): 3599-3616.

Minning, T. A., J. Bua, et al. (2003). "Microarray profiling of gene expression during trypomastigote to amastigote transition in Trypanosoma cruzi." Molecular and Biochemical Parasitology **131**(1): 55-64.

Mishra, B., R. K. Singh, et al. (2012). "Advances in rice genetics." World Scientific Publishing Co.

Mohanty, A., H. Kathuria, et al. (2002). "Transgenics of an elite indica rice variety Pusa Basmati 1 harbouring the codA gene are highly tolerant to salt stress." Theoretical and Applied Genetics **106**(1): 51-57.

Moison, M., F. Roux, et al. (2010). "Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in Arabidopsis thaliana." The Plant Journal **63**(5): 728-738.

Moore, C. R., L. S. Johnson, et al. (2013). "High-Throughput Computer Vision Introduces the Time Axis to a Quantitative Trait Map of a Plant Growth Response." Genetics **195**(3): 1077.

Moradi, F. and A. M. Ismail (2007). "Responses of Photosynthesis, Chlorophyll Fluorescence and ROS-Scavenging Systems to Salt Stress During Seedling and Reproductive Stages in Rice." Annals of Botany **99**(6): 1161-1173.

Moradi, F., A. M. Ismail, et al. (2003). "Salinity tolerance of rice during reproductive development and association with tolerance at seedling stage." Indian Journal of Plant Physiology **8**.

Müller-Linow, M., F. Pinto-Espinosa, et al. (2015). "The leaf angle distribution of natural plant populations: assessing the canopy with a novel software tool." Plant Methods **11**(1): 11.

Munns, R. (2002). "Comparative physiology of salt and water stress." Plant, Cell & Environment **25**(2): 239-250.

Munns, R. (2005). "Genes and salt tolerance: bringing them together." New Phytologist **167**(3): 645-663.

Munns, R., R. A. James, et al. (2006). "Approaches to increasing the salt tolerance of wheat and other cereals." Journal of Experimental Botany **57**(5): 1025-1043.

Munns, R. and M. Tester (2008). "Mechanisms of Salinity Tolerance." Annual Review of Plant Biology **59**(1): 651-681.

Nakashima, K., L. S. Tran, et al. (2007). "Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice." Plant J **51**(4): 617-630.

Negrão, S., B. Courtois, et al. (2011). "Recent updates on salinity stress in rice: from physiological to molecular responses." Critical Reviews in Plant Sciences **30**.

Negrão, S., S. M. Schmöckel, et al. (2017). "Evaluating physiological responses of plants to salinity stress." Annals of Botany **119**(1): 1-11.

Nutan, K. K., H. R. Kushwaha, et al. (2017). "Transcription dynamics of Saltol QTL localized genes encoding transcription factors, reveals their differential regulation in contrasting genotypes of rice." Functional & Integrative Genomics **17**(1): 69-83.

Obata, T., H. K. Kitamoto, et al. (2007). "Rice Shaker Potassium Channel OsKAT1 Confers Tolerance to Salinity Stress on Yeast and Rice Cells." Plant Physiology **144**(4): 1978.

OECD and FAO (2012).

Okonechnikov, K., A. Conesa, et al. (2016). "Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data." Bioinformatics 32(2): 292-294.

Ouyang, S.-Q., Y.-F. Liu, et al. (2010). "Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (Oryza sativa) plants." The Plant Journal **62**(2): 316-329.

Paine, C. E. T., T. R. Marthews, et al. (2012). "How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists." Methods in Ecology and Evolution **3**(2): 245-256.

Pang, Y., K. Chen, et al. (2017). "Simultaneous Improvement and Genetic Dissection of Salt Tolerance of Rice (Oryza sativa L.) by Designed QTL Pyramiding." Frontiers in Plant Science **8**: 1275.

Phean, O. P. S., P. Punteeranurak, et al. (2005). "Calcium signaling-mediated and differential induction of calmodulin gene expression by stress in Oryza sativa L." J Biochem Mol Biol **38**(4): 432-439.

Piffanelli, P., L. Ramsay, et al. (2004). "A barley cultivation-associated polymorphism conveys resistance to powdery mildew." Nature **430**(7002): 887-891.

Pires, I. S., S. Negrao, et al. (2015). "Comprehensive phenotypic analysis of rice (Oryza sativa) response to salinity stress." Physiol Plant **155**(1): 43-54.

Platten, J. D., J. A. Egdane, et al. (2013). "Salinity tolerance, Na+ exclusion and allele mining of HKT1;5 in Oryza sativa and O. glaberrima: many sources, many genes, one mechanism?" BMC Plant Biology **13**(1): 32.

Potokina, E., A. Druka, et al. (2008). "Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation." Plant J **53**(1): 90-101.

Pound, M. P., A. P. French, et al. (2013). "RootNav: Navigating Images of Complex Root Architectures." Plant Physiology **162**(4): 1802.

Qadir, M., E. Quillérou, et al. (2014). "Economics of salt-induced land degradation and restoration." Natural Resources Forum **38**(4): 282-295.

Qiu, Q.-S., Y. Guo, et al. (2004). "Regulation of Vacuolar Na+/H+ Exchange in Arabidopsis thaliana by the Salt-Overly-Sensitive (SOS) Pathway." Journal of Biological Chemistry **279**(1): 207-215.

R., B., N. S. K., et al. (2004). Integrating marker-assisted selection in crop breeding: Prospects and challenges. Bangalore, INDE, Current Science Association.

Rahman, M. A., M. J. Thomson, et al. (2016). "Exploring novel genetic sources of salinity tolerance in rice through molecular and physiological characterization." Annals of Botany **117**(6): 1083-1097.

Rahman, S. M., T. Matsumuro, et al. (2001). "Effects of Salinity Stress on the Seminal Root Tip Ultrastructures of Rice Seedlings (Oryza sativa L.)." Plant Production Science **4**(2): 103-111.

Reddy, I. N. B. L., B.-K. Kim, et al. (2017). "Salt Tolerance in Rice: Focus on Mechanisms and Approaches." Rice Science **24**(3): 123-144.

Ren, Z.-H., J.-P. Gao, et al. (2005). "A rice quantitative trait locus for salt tolerance encodes a sodium transporter." Nat Genet **37**(10): 1141-1146.

Rieseberg, L. H., M. A. Archer, et al. (1999). "Transgressive segregation, adaptation and speciation." Heredity **83**(4): 363-372.

Roberts, A., C. Trapnell, et al. (2011). "Improving RNA-Seq expression estimates by correcting for fragment bias." Genome Biology **12**(3): R22.

Robinson, J. T., H. Thorvaldsdóttir, et al. (2011). "Integrative genomics viewer." Nature biotechnology **29**(1): 24-26.

Rodríguez-Navarro, A. and F. Rubio (2006). "High-affinity potassium and sodium transport systems in plants." Journal of Experimental Botany **57**(5): 1149-1160.

Ronald, J., R. B. Brem, et al. (2005). "Local Regulatory Variation in Saccharomyces cerevisiae." PLOS Genetics **1**(2): e25.

Saade, S., A. Maurer, et al. (2016). "Yield-related salinity tolerance traits identified in a nested association mapping (NAM) population of wild barley." Scientific Reports **6**: 32586.

Sabouri, H., A. M. Rezai, et al. (2009). "QTLs mapping of physiological traits related to salt tolerance in young rice seedlings." Biologia Plantarum **53**(4): 657-662.

Saijo, Y., S. Hata, et al. (2000). "Over-expression of a single Ca2+-dependent protein kinase confers both cold and salt/drought tolerance on rice plants." The Plant Journal **23**(3): 319-327.

Sakurai, J., A. Ahamed, et al. (2008). "Tissue and Cell-Specific Localization of Rice Aquaporins and Their Water Transport Activities." Plant and Cell Physiology **49**(1): 30-39.

Sakurai, J., F. Ishikawa, et al. (2005). "Identification of 33 Rice Aquaporin Genes and Analysis of Their Expression and Function." Plant and Cell Physiology **46**(9): 1568-1577.

Schadt, E. E., S. A. Monks, et al. (2003). "Genetics of gene expression surveyed in maize, mouse and man." Nature **422**(6929): 297-302.

Seelig, H. D., A. Hoehn, et al. (2008). "The assessment of leaf water content using leaf reflectance ratios in the visible, near-, and short-wave-infrared." International Journal of Remote Sensing **29**(13): 3701-3713.

Senadheera, P., R. K. Singh, et al. (2009). "Differentially expressed membrane transporters in rice roots may contribute to cultivar dependent salt tolerance." Journal of Experimental Botany **60**(9): 2553-2563.

Septiningsih, E. M., A. M. Pamplona, et al. (2009). "Development of submergence-tolerant rice cultivars: the Sub1 locus and beyond." Annals of Botany **103**(2): 151-160.

Serraj, R. and T. R. Sinclair (2002). "Osmolyte accumulation: can it really help increase crop yield under drought conditions?" Plant, Cell & Environment **25**(2): 333-341.

Shabala, S. (2013). "Learning from halophytes: physiological basis and strategies to improve abiotic stress tolerance in crops." Annals of Botany **112**(7): 1209-1221.

Shabalin, A. A. (2012). "Matrix eQTL: ultra fast eQTL analysis via large matrix operations." Bioinformatics **28**(10): 1353-1358.

Shavrukov, Y. (2013). "Salt stress or salt shock: which genes are we studying?" Journal of Experimental Botany **64**(1): 119-127.

Shi, C., A. Uzarowska, et al. (2007). "Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint × Flint maize recombinant inbred line population." BMC Genomics **8**: 22-22.

Shi, C., A. Uzarowska, et al. (2007). "Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint × Flint maize recombinant inbred line population." BMC Genomics **8**(1): 22.

Singh, R. K., G. B. Gregorio, et al. (2007). "QTL mapping for salinity tolerance in rice." Physiol. Mol. Biol. Plants **13**(2): 87-99.

Singh, Y. P., V. K. Mishra, et al. (2016). "Productivity of sodic soils can be enhanced through the use of salt tolerant rice varieties and proper agronomic practices." Field Crops Research **190**: 82-90.

Smajgl, A., T. Q. Toan, et al. (2015). "Responding to rising sea levels in the Mekong Delta." Nature Clim. Change **5**(2): 167-174.

Smith, R. M., H. Alachkar, et al. (2011). "Nicotinic [alpha]5 receptor subunit mRNA expression is associated with distant 5[prime] upstream polymorphisms." Eur J Hum Genet **19**(1): 76-83.

Sonia, N., A. M. Cecilia, et al. (2012). "Use of EcoTILLING to Find Genes Related to Salt Tolerance in Rice. W622 " Plant and Animal Genome conference. 20, San Diego, États-Unis, 14 January 2012/18 January 2012.

Su, J., R. Hirji, et al. (2006). "Evaluation of the stress-inducible production of choline oxidase in transgenic rice as a strategy for producing the stress-protectant glycine betaine." Journal of Experimental Botany **57**(5): 1129-1135.

Sunkar, R., Y.-F. Li, et al. (2012). "Functions of microRNAs in plant stress responses." Trends in Plant Science **17**(4): 196-203.

Supek, F., M. Bošnjak, et al. (2011). "REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms." PLOS ONE **6**(7): e21800.

Takehisa, H., T. Shimodate, et al. (2004). "Identification of quantitative trait loci for plant growth of rice in paddy field flooded with salt water." Field Crops Research **89**(1): 85-95.

Tanabata, T., T. Shibaya, et al. (2012). "SmartGrain: High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis." Plant Physiology **160**(4): 1871.

Tang, Z., Z. Yang, et al. (2013). "Cytonuclear epistatic quantitative trait locus mapping for plant height and ear height in maize." Molecular Breeding **31**(1): 1-14.

Tanksley, S. D. (1993). "Mapping Polygenes." Annual Review of Genetics **27**(1): 205-233.

Thomson, M. J., M. de Ocampo, et al. (2010). "Characterizing the Saltol Quantitative Trait Locus for Salinity Tolerance in Rice." Rice **3**(2): 148-160.

Trapnell, C., B. A. Williams, et al. (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nat Biotech **28**(5): 511-515.

Ul Haq, T., J. Gorham, et al. (2010). "Dynamic quantitative trait loci for salt stress components on chromosome 1 of rice." Functional Plant Biology **37**(7): 634-645.

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

Walia, H., C. Wilson, et al. (2007). "Genome-wide transcriptional analysis of salinity stressed japonica and indica rice genotypes during panicle initiation stage." Plant Molecular Biology **63**(5): 609-623.

Wang, J., H. Yu, et al. (2014). "An expression quantitative trait loci-guided co-expression analysis for constructing regulatory network using a rice recombinant inbred line population." Journal of Experimental Botany **65**(4): 1069-1079.

Wang, J., H. Yu, et al. (2010). "A global analysis of QTLs for expression variations in rice shoots at the early seedling stage." Plant J **63**(6): 1063-1074.

Wang, Q., Y. Guan, et al. (2008). "Overexpression of a rice OsDREB1F gene increases salt, drought, and low temperature tolerance in both Arabidopsis and rice." Plant Molecular Biology **67**(6): 589-602.

Wang, X.-S., H.-B. Zhu, et al. (2007). "Genome-scale identification and analysis of LEA genes in rice (Oryza sativa L.)." Plant Science **172**(2): 414-420.

Wang, Y.-Y., P.-K. Hsu, et al. (2012). "Uptake, allocation and signaling of nitrate." Trends in Plant Science **17**(8): 458-467.

Wang, Z., Z. Chen, et al. (2012). "QTL Analysis of Na⁺ and K⁺ Concentrations in Roots and Shoots under Different Levels of NaCl Stress in Rice (<italic>Oryza sativa</italic> L.)." PLoS ONE **7**(12): e51202.

Wang, Z., M. Gerstein, et al. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nature reviews. Genetics **10**(1): 57-63.

Wassmann, R., S. V. K. Jagadish, et al. (2009). "Climate change affecting rice production: the physiological and agronomic basis for possible adaptation strategies." Adv Agron **101**.

Wentzell, A. M., H. C. Rowe, et al. (2007). "Linking Metabolic QTLs with Network and cis-eQTLs Controlling Biosynthetic Pathways." PLOS Genetics **3**(9): e162.

West, M. A. L., K. Kim, et al. (2007). "Global eQTL Mapping Reveals the Complex Genetic Architecture of Transcript-Level Variation in Arabidopsis." Genetics **175**(3): 1441.

Wei, s. (2009). "eMap."

Wilson, A., J. Latham, et al. (2005). "Regulatory regimes for transgenic crops." Nat Biotech **23**(7): 785-785.

Wolf, J. (2009). "Cytonuclear interactions can favor the evolution of genomic imprinting." Evolution **63**(5): 1364-1371.

Wolf, J. B. (2009). "CYTONUCLEAR INTERACTIONS CAN FAVOR THE EVOLUTION OF GENOMIC IMPRINTING." Evolution **63**(5): 1364-1371.

Wolf, J. B. and Y. Brandvain (2014). "Gene interactions in the evolution of genomic imprinting." Heredity **113**(2): 129-137.

WRS, I. (2017). "World Rice Statistics."

Würschum, T., W. Liu, et al. (2014). "Adult Plant Development in Triticale (× Triticosecale Wittmack) Is Controlled by Dynamic Genetic Patterns of Regulation." G3: Genes|Genomes|Genetics **4**(9): 1585-1591.

Xia, K., R. Wang, et al. (2012). "OsTIR1 and OsAFB2 Downregulation via OsmiR393 Overexpression Leads to More Tillers, Early Flowering and Less Tolerance to Salt and Drought in Rice." PLOS ONE **7**(1): e30039.

Xiong, L. and Y. Yang (2003). "Disease Resistance and Abiotic Stress Tolerance in Rice Are Inversely Modulated by an Abscisic Acid–Inducible Mitogen-Activated Protein Kinase." The Plant Cell **15**(3): 745.

Xu, X.-J., C. Chen, et al. (2016). "Characterization of a newly isolated strain Pseudomonas sp. C27 for sulfide oxidation: Reaction kinetics and stoichiometry." **6**: 21032.

Yang, W., L. Duan, et al. (2013). "Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies." Current Opinion in Plant Biology **16**(2): 180-187.

Yoshida, S., D. A. Forno, et al. (1976.). "Laboratory manual for physiological studies of rice." IRRI, Los Banos, Laguna, Philippines: 61-66.

Zang, B., H. Li, et al. (2011). "Analysis of trehalose-6-phosphate synthase (TPS) gene family suggests the formation of TPS complexes in rice." Plant Molecular Biology **76**(6): 507-522.

Zeng, L. and M. C. Shannon (2000). "Salinity Effects on Seedling Growth and Yield Components of Rice." **40**(4): 996-1003.

Zhang, A., M. Jiang, et al. (2006). "Mitogen-Activated Protein Kinase Is Involved in Abscisic Acid-Induced Antioxidant Defense and Acts Downstream of Reactive Oxygen Species Production in Leaves of Maize Plants." Plant Physiology **141**(2): 475-487.

Zheng, H., H. Zhao, et al. (2015). "QTL analysis of Na+ and K+ concentrations in shoots and roots under NaCl stress based on linkage and association analysis in japonica rice." Euphytica **201**(1): 109-121.

Zheng, T. C., X. K. Zhang, et al. (2011). "Genetic gains in grain yield, net photosynthesis and stomatal conductance achieved in Henan Province of China between 1981 and 2008." Field Crops Research **122**(3): 225-233.

Zhou, G.-A., R.-Z. Chang, et al. (2010). "Overexpression of soybean ubiquitin-conjugating enzyme gene GmUBC2 confers enhanced drought and salt tolerance through modulating abiotic stress-responsive gene expression in Arabidopsis." Plant Molecular Biology **72**(4): 357-367.

Zhou, S., L. Jia, et al. (2016). "Arabidopsis CaM1 and CaM4 Promote Nitric Oxide Production and Salt Resistance by Inhibiting S-Nitrosoglutathione Reductase via Direct Binding." PLOS Genetics **12**(9): e1006255.