Discovery and Validation of Single Nucleotide Polymorphisms in Candidate Genes of Drought Tolerance in Rice

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

By

VIJAYA BHASKER REDDY LACHAGARI

Department of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad 500 046
Andhra Pradesh, India

Enrollment No. 03LPPH08

November 2008

Department of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad 500046
INDIA

CERTIFICATE

This is to certify that **Mr.Vijaya Bhasker Reddy Lachagari** has carried out the research work embodied in the present thesis entitled **"Discovery and Validation of Single Nucleotide Polymorphisms in Candidate Genes of Drought Tolerance in Rice"** for the degree of Doctor of Philosophy under my supervision in the Department of Plant Sciences, School of Life Sciences, University of Hyderabad.

This work has not been submitted for the award of any degree or diploma of any other University or Institute.

Prof. Arjula Ramachandra Reddy
Supervisor

28/11/08

Head
Department of Plant Sciences

29/11/08

Dean
School of Life Sciences

29/11/08

DECLARATION

I hereby declare that the work presented in this thesis entitled "**Discovery and Validation of Single Nucleotide Polymorphisms in Candidate Genes of Drought Tolerance in Rice**" has been carried out by me under the supervision of **Prof. Arjula Ramachandra Reddy** in the Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad- 500 046, and that this work has not been submitted for any degree or diploma of any other University or Institute. All the assistance and help received during the course of the investigation have been duly acknowledged.

Vijaya Bhasker Reddy Lachagari

28/11/08

Prof.Arjula Ramachandra Reddy
Supervisor
Department of Plant Sciences

Dedicated to

My thesis Supervisor,

Hon'ble Vice Chancellor of Yogi Vemana University, Kadapa

Prof. Arjula Ramachandra Reddy

Acknowledgements

I express my sincere gratitude to my advisor, ***Professor Arjula Ramachandra Reddy****, for his constant guidance, encouragement and moral support throughout this research and supervising this thesis with such great interest. I am indebted to him for my life time and promise to carry his sense of attitude in my future scientific endeavors. He has been an excellent mentor, introducing me as a scientist, encouraging and getting me funds to work in his lab. The scientific knowledge and training in advanced technological innovations I gained during my stay with Prof. Reddy has helped me in my scientific endeavors at its best.*

I am thankful to the former Deans, School of Life Sciences, ***Prof. T. Suryanarayana*** *and the present Dean,* ***Prof. A. S. Raghavendra****, for extending facilities available in the School for my research work. I would also like to thank the former Head, Department of Plant Sciences,* ***Prof. M.N.V. Prasad, Prof. P.B. Kirti*** *and the present Head* ***Prof. Appa Rao Podile*** *for providing all the required facilities in the department. Special thanks to all the faculty of Dept. of Plant Sciences and School of Life Sciences, who taught me advanced areas of biology during my masters degree, with out which I would not have dared to take up such a challenging task. I would like to thank* ***Dr. G. Padmaja****, Dept. of Plant Sciences for being helpful through out my present work. My special thanks to* ***Prof. C.K Mithra and Prof. Aparna Datta Gupta*** *for their valuable suggestions during my Bioinformatics course. I thank all the* ***non-teaching staff*** *of the School for their help and assistance.*

I am thankful to our collaborators ***Prof. Andrew Paterson****,* ***Dr. Alex Feltus, University of Georgia, USA*** *for extending a collaborative research support for gene expression profiling using cDNA microarrays. I am thankful to our other collaborators* ***Prof. Jeffrey L. Bennetzen, Dr. Ramakrishna, Dr. M.K. Reddy*** *for giving an opportunity to develop and analyze the genetic resources used for the present study on functional genomics of drought tolerance in rice.*

I would like to express my gratitude to ***Dr. Gabor Marth****, Washington University Genome Sequencing Centre for his suggestions in SNP identification and analysis and* ***Dr. Salomini*** *for his valuable suggestions in targeting some of the key target genes for the trait.*

I thank Ocimum Biosolutions Pvt Ltd and Strandls Pvt Ltd for providing the required for software for microarray data mining and analysis.

I wish to thank my colleagues ***Dr. Markandeya Gorantla, Dr. Ravindra Babu Peram, Dr. Chandra Shekar Akila, Dr. Shivarama Prasad Lekkala, Dr. Madana Mohan Reddy Ambavaram, Dr.Linga Reddy Gutha, Dr. Kiran Kumar Velpula, Maheedhar Reddy and Mahinder Thudi*** *for their help and suggestions during my research work. I am thankful to my seniors* ***Dr. Nagabushan Ithal, Dr.Chandrasekhar Boosani*** *for their help and suggestion in trouble shooting of the experiments. I would like to thank* ***B.Venkat Reddy*** *for extending the help whenever needed. I also wish to thank our past research group* ***Dr. B. Karunasree*** *,* ***Dr G. Madhuri*** *for their encouragement. I would like to thank the assistance received from* ***Raju, Ravinder, Poorna, Satish*** *and* ***Narasimha*** *in the lab whenever needed.*

I sincerely acknowledge the ***CSIR, Govt. of India*** *for awarding me with Junior Research Fellowship and Senior Research Fellowship,* ***University of Hyderabad*** *for UPE travel grant award for attending* ***PAG conference at USA.*** *Facilities through* ***DST-FIST*** *funding to the Department of Plant Sciences and the* ***UGC-SAP*** *to School of Life Sciences are gratefully acknowledged.*

I thank all my ***friends*** *for their constant support and great time spent outside my laboratory with fond memories.*

I thank my ***parents, sisters, and brothers*** *for their patience and support. I thank my Uncle and Aunt* ***Mr and Mrs .Raji Reddy Lachagari*** *for their lots of love and affection. I owe a debt of gratitude for their love, supporting spiritually, emotionally and financially, and understanding me along the path of my longest academic pursuits. My best wishes to my cousin* ***Master Aditya Reddy*** *who brings in smiles in me and makes me forget all my worries to begin a new day with great success.*

Date: Friday, November 28, 2008 ***Vijaya Bhasker Reddy Lachagari***

CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

ABA	: Abscisic Acid
ABRE	: ABA Responsive Element
AREB	: ABA Responsive Element Binding
APX	: Ascorbate Peroxidase
ASPCR	: Allele Specific PCR
ASLNA	: Allele Specific LNA
ASMap	: Allele Sharing Map
bHLH	: Basic region Helix-Loop-Helix
bp	: base pair
BAC	: Bacterial Artificial Chromosome
BLAST	: Basic Local Alignment Search Tool
bZIP	: Basic Region Leucine Zipper
CAP3	: Contig Assembly Program version 3
CBF	: CRT/DRE Binding Factor
CRT/DRE	: C-Repeat Dehydration Responsive Element
CTAB	: Cetyltrimethyl Ammonium Bromide
cDNA	: Complementary DNA
CMS	: Cytoplasmic Male Sterility
CAPS	: Cleaved Amplified Polymorphic Sequences
CD-ROM	: Compact Disc Read Only Memory
dbSNP	: database of SNPs at NCBI
DNA	: Deoxyribonucleic Acid
DEPC	: Diethylpyrocarbonate
DTT	: Dithiothreitol
DMSO	: Dimethylsulphoxide
DRE	: Dehydration Responsive Element
DREB	: DRE Binding Protein
DHL	: Double Haploid Lines
EDTA	: Ethylenediamine Tetra Acetic Acid
EST	: Expressed Sequence Tags
FYM	: Farm Yard Manure
GABA	: γ aminobutyric Acid
G X E	: Genetic and Environmental effects
GMOD	: Generic Model Organism Database
H_2O_2	: Hydrogen Peroxide
HSP	: Heat Shock Proteins
IRGSP	: International Rice Genome Sequencing Project
IRRI	: International Rice Research Institute
kDa	: kilo Dalton
LEA	: Late Embryogenesis Abundant proteins
LTRE	: Low Temperature Responsive Element
LNA	: Locked Nucleic Acid
mRNA	: messenger RNA
miRNA	: micro RNA

MAS	: Marker Assisted Selection
MAPK	: Mitogen Activated Protein Kinase
MAPKK	: MAP kinase kinase
MITE	: Miniature Inverted Transposable Elements
NERICA	: New Rice for Africa
N22	: Nagina 22
NCBI	: National Centre for Biotechnological Information
ORYZASNP	: An initiative by IRRI for identifying SNPs in rice
OA	: Osmotic Adjustment
PAC	: P1 Derived Artificial Chromosomes
PCR	: Polymerase Chain Reaction
PHRED	: Phil's Read edit data
PLACE	: Plant cis Acting Elements Database
Pfu	: *Pyrococcus furiosus*
PVP	: Polyvinyl pyrrolidone
QTL	: Quantitative Tract Loci
RNA	: Ribonucleic Acid
rSNPs	: SNPs is in the regulatory region
RAPD	: Randomly Amplified Polymorphic DNA
REMAP	: Retrotransposon Microsatellite Amplified Polymorphism
RWC	: Relative Water Content
ROP	: Rate of Polymorphism
SAM	: S-adenosyn methionine
SNP	: Single Nucleotide Polymorphisms
SRG	: Stress Responsive Genes
STS	: Sequence Tagged Site
SCARS	: Sequence Characterized Amplified Regions
SSCP	: Single Strand Confirmation Polymorphism
SOD	: Superoxide Dismutase
SBE	: Single Base Extension
TMAC	: Tetramethyl Ammonium Chloride
Taq	: *Thermus aquaticus*
UTR	: Untranslated Regions

1. INTRODUCTION

1. INTRODUCTION

Rice, the most important food crop of the world, is a staple food for more than half of the world's population. It has become a model cereal crop for genomics because of its relatively small genome, availability of high density genetic and physical maps and complete draft genome sequence. Various biotic and abiotic factors limit crop production in almost all rice growing areas. Of the abiotic stress factors, the yield losses caused by drought alone accounts to as much as 100% during certain years in some areas. Being a member of a grass family it is expected to cope up with such an adverse environmental condition, but most of the modern rice varieties are not able to withstand water limited conditions. The main reason for this appears to be the process of domestication, may be with the greed for more productivity or to meet the hunger needs of growing population led to biased selection strategies towards high yielding varieties with least focus on other important traits such as stress tolerance. Further, with the limitation in water availability due to inadequate and erratic rainfalls, drought has emerged as major constraint for rice production. Drought tolerance is a complex trait, controlled by a large number of genes dispersed across the genome with a complex regulation. The G X E interactions trigger a set of genes and in turn leading to a complex co-regulation of large number of genes in the network which will lead to adaptation of plants against environmental stresses. A crucial step towards understanding the molecular genetic basis of drought tolerance is identification, functional study of candidate genes and the allelic variants, all of which require large-scale genomic and genetic resources.

Linking a genotype with a phenotype was the most fundamental tenet of Mendelian genetics. The underlying assumption was that a phenotype is appropriated to an effect of a single gene. However, with the remarkable improvements in methods and scale of genetic analysis over the years, many complex traits were shown to be controlled by many non allelic genes throughout the genome. However this 'many-many' relationship remained largely unexplained until recently, largely due to the lack of sophisticated genetic and molecular technologies. Rapid advances in genome technologies have made it possible to take a 'many genes at a time' approach instead of the usual 'one gene at a time' approach and paved the way for unraveling the genetic nature of complex traits. With the availability of the whole genome sequence of Arabidopsis and rice, it became clear that the complex traits and pathways can be dissected using new genomic tools and the underlying genes can be uncovered, thus paving the way to link genes, pathways and phenotypes. With the advent of high-throughput genomic technologies, large amounts of sequence information of rice have become available in the public domain. Large scale EST generation projects helped to uncover the information on gene expression under defined condition during defined developmental stages. This outcome has been utilized for genome annotation, accurate gene prediction gene expression profiling studies and comparative genomics. Significant developments in gene expression profiling technologies in combination with genetic studies have begun to provide leads to the discovery of candidate genes governing complex traits and their regulation. Further, the spectacular progress in plant biochemistry and molecular biology has demonstrated association of these genes with different pathways, all contributing to a complex trait. In parallel, advances in

molecular mapping methods were increasingly used in precisely relating genomic regions to the target complex trait. There are numerous studies involving stress responsive gene regulation in model plants under different environmental stresses. These studies show that plants exhibit various stress responses at gene and protein level. There is increasing evidence for diverse environmental stresses activating similar cell signaling pathways. In general, plants accumulate significantly higher levels of osmoprotectants, antioxidants, LEAs, AREBs, aquaporins and all of these are implicated in stress tolerance. Plants activate signal transduction pathways through activation of relevant enzymes and proteins such as MAP kinases, transcription factors like DREBs and MYBs. Expression analysis of functionally important stress regulated genes has been helpful to elucidate genetic networks in which they participate (Breyne *et al*., 2003). However, genetic analyses of drought tolerance have not been extensively performed; molecular mechanisms of regulating plant genes under drought stress are yet to be uncovered. Recent advances in genomics have provided a new avenue to investigate and utilize allelic variation in target genomic segments associated with complex traits. In rice, gene expression profiling through arrays have been carried out for studies involving salt tolerance (Kawasaki *et al*., 2001; Walia *et al*., 2005), ABA and gibberellins responsive gene expression (Yazaki *et al*., 2003, 2004), cold, drought and high salinity responses (Rabbani *et al*., 2003), organ specific gene expression under high salinity (Jiao *et al*., 2005a). However, all these studies have focused on generating expression profiles on laboratory grown seedlings subjected to stress for a small interval of time or there are studies on tissues during a particular stage in rice life cycle. However there are no studies which show functionally important genes, and how they

are co-regulated and adapt to environmental stresses occurring during the life cycle of rice plant grown in field conditions. As part of the our functional genomic analysis we have utilized the gene resources of rice and pearl millet developed at our laboratory, identified to be involved in stress response through comparative analysis (Gorantla *et al*., 2007) and further studied patterns of their co-regulation across the life cycle of rice plant under field grown conditions and identified sets of drought responsive genes. Further, Identification of allelic variations at such loci will be the basis to develop suitable informative markers for the trait variation and selection.

Though rice has wide diversity with varied field performance, the alleles conferring stress tolerance appears to be masked or lost in domesticated rice during the process of selection. Among the domesticated rice genotypes, some are drought tolerant and some are susceptible. Studies have revealed that upland cultivars possess well developed mechanisms to resist against drought stress (Yadav *et al.,* 1997). This varied performance of genotypes in a given condition is largely attributed to genetic variations, mostly at single nucleotide level. Among such diverse genotypes with varied performance, 99.9% of the DNA is identical and of the 0.1% variation, 90% seemed to be SNPs (Matukumalli *et al*., 2006) which makes the genotype phenotypically unique. A SNP is a substitution of one nucleotide with another, and both versions are observed with the frequency greater than 1% in a population. SNPs may be present within and/or outside genic regions and majority of them do not produce physical changes in the individual with affected DNA. SNPs found within a coding sequence are of particular interest because they are more likely to alter the biological function of a protein. Finding SNPs in genetic region encoding a gene of interest has been prime research

area in parallel to genome sequencing projects. mRNAs containing different bases at SNP sites may vary in their interactions with cellular components involved in mRNA synthesis, maturation, transport, translation, or degradation. Though, only 3 to 5 percent of total DNA codes for the production of proteins, it is observed that the SNPs occur ~1% in case of human and most of the SNPs found outside of coding regions. SNPs that are present outside coding regions, and do not affect RNA consensus and protein sequences have not been analyzed in detail. It is conceivable that such SNPs could also lead to phenotypic effects, most likely through non-consensus-dependent mechanisms. It has been documented that single base-pair substitutions alter or create essential sequence elements for transcription, splicing, processing, translation or steady-state level of mRNA. The importance of SNPs occurring in non coding regions is immense; especially finding SNPs in regulatory regions associated with candidate gene expression is a key research area in post genomic era. The differences in gene expression profiles among different genotypes seemed to be mostly due to SNPs in the regulatory regions called rSNPs. Studies towards identification of such regions and variations are highly significant, particularly in understanding gene regulation of complex traits and further to develop molecular markers for the trait. In particular, rSNPs associated with candidate genes of a trait are of great importance as they are likely to be regulatory switches for the trait as demonstrated by Konishi *et al*. (2006) with BEL-1 type homeobox gene associated with seed shattering. Thus, studies towards identifying such variations are promising to decipher the regulatory switches for the complex traits. Sometimes, the alterations involve a single base pair and are shared by many individuals in the population, and these individuals are likely to share ancestral

history and a common phenotype. Such varying patterns among a group of individuals are the basis for constructing hapmap which is an invaluable tool in associative genetics. Information on the location, frequency, role and distribution of SNPs is limited in plants. In Arabidopsis, Torjek *et al*. (2003) reported 100 SNPs and genotyped in a BC3 population comprising of 48 lines, Schmid *et al*. (2003), reported 8688 SNPs, analyzing ~10,000 EST sequences generated from 6 different cDNA libraries. In Barley, 112 SNPs were reported by Kanazin *et al*. (2002) analyzing 38 loci from 5 genotypes and 2000 SNPs reported by Rostoks *et al*. (2005a) analyzing 877 loci from 6 genotypes. Recently, 5551 SNPs were reported and first transcript map of soybean was generated by Choi *et al*. (2007) re-sequencing 1141 STS regions and intronic regions were well explored in case of pearl millet (Bertin *et al.,* 2005) to convert SSCP-SNP markers. In rice, SNPs were identified for specific loci associated with SOD gene of xa13 (Kottapalli *et al*., 2007), starch synthase gene for gelatinization temperature (Waters *et al*., 2006), waxy gene (Bao *et al*., 2006a), CMS (Duan *et al*., 2007) are being used in MAS. Feltus *et al.* (2004) and Zhao *et al*. (2004) exploited indica and japonica genome sequences for *in silico* comparison to identify large number of SNPs available in the public domain. A total of 2800 SNPs were reported by Nasu *et al.* (2002), analyzing 418 randomly targeted intergenic regions from 3 genotypes, of which 94 SNPs are converted into markers with FPTDI assay and Monna *et al*. (2006) also reported 7805 SNPs randomly targeting 1117 regions in 9 genotypes derived allele sharing maps and integrated with genetic map based BAC clone map of IRGSP. In above two cases, the focus was on identifying large number of SNPs by targeting randomly selected intergenic regions but not relevant to any specific trait. In rice,

reports on dissecting the genomic regions associated with a trait for SNPs and converting them into molecular markers are scanty, and so far, no study has been carried out in case of complex traits such as drought tolerance.

RATIONALE

Identification of allelic variations at single nucleotide level has been well explored in case of model animal systems but there has been only a marginal focus in case of plant systems due to the expensive strategies involved in the process of identification to association of SNPs to the trait. In plants, particularly in grasses, SNP analysis could provide valuable information on genetic control of agronomic traits with unprecedented precision. The abundance, ubiquity and interspersed nature of SNPs make them ideal candidates as molecular markers for marker-assisted plant breeding. Though, the real potential of SNPs in association studies and map saturation has been recognized worldwide, the attempts towards utilization of such a powerful tool seemed to be scanty. Especially, trait based dissection of genomic regions for allelic variations and further utilizing these variations in associating to the trait of interest has been a promising approach as demonstrated in animal systems such as zebra fish (Lamason *et al*., 2005). Such an approach, targeting candidate genes of the desired trait, find SNPs and utilization of these variations as the markers for selecting the trait is more appropriate for complex traits such as drought tolerance in plants. Though there were a few attempts to identify SNPs in rice (Monna *et al*., 2006, Nasu *et al*., 2002) targeting random intergenic regions on the genome, only a limited data has been obtained for association studies for any trait. Realizing the potential of SNPs in molecular genetics, International Rice Research Institute (IRRI) has initiated an International consortium

recently, called ORYZASNP (McNally *et al*., 2006). The main aim of this consortium is to identify all single nucleotide variation in rice genome. But, there has been no attempt so far to dissect the genomic regions associated with drought tolerance and, identify SNPs and associate them with this trait in rice. A very important step towards such an approach is to precisely identify the genomic regions (coding or non-coding) which are involved in response to drought stress. Further, these regions needs to be dissected for the allelic variations to the single nucleotide level and correlated with the phenotype before attempting to utilize them as potential markers.

The present study deals with identification of candidate genes for the drought tolerance in rice, through subtracted cDNA libraries, gene expression profiling and comparative analysis. These genomic regions are dissected to find SNPs in coding and non coding regions including UTRs, introns and, with a major focus on the promoters of the selected drought stress responsive genes. Further, this data has been extensively utilized to convert the candidate SNPs to simplified SNP markers to facilitate marker assisted selection process. This is expected to lead towards the development of a haplotype map with candidate genes of drought tolerance in rice.

Specific Objectives for the present study are

The main objectives set for the present study can be categorized in to two major sections, the first part deals with identification genes and genomic regions involved in drought stress response process and second part deals with dissecting these regions for SNPs and associating with the drought related traits.

1. Identification of Candidate Genes for Drought Tolerance

A. Construction of subtractive cDNA library using highly drought tolerant and susceptible varieties.
B. Sequencing, annotation and identification of drought stress responsive genes.

C. Identification of candidate genes for drought tolerance and their promoters from microarray and comparative analysis.
D. Identification of miRNA genes and their promoters involved in the regulation of stress responsive genes.

2. Identification of allelic variations in candidate genes of drought tolerance

A. *In silico* analysis of identified genes with available unigene dataset for single nucleotide polymorphisms (SNPs)
B. Development of a new software pipeline for identification, analysis and mapping of SNPs in an automated fashion optimized for rice genomic data
C. Targeting putative SRGs and regulatory regions of selected genes, SNP identification through re-sequencing from phenotypically well characterized elite Indian rice cultivars
D. Converting candidate SNPs into markers to associate with the phenotype
E. Genotyping and mapping of identified SNPs in CT9993 X IR62266 DHL population.
F. Effort towards the development of haplotype map of drought responsive genes in rice.

2. LITERATURE REVIEW

2. REVIEW OF LITERATURE

Identification of allelic variations to the single nucleotide level has been well explored in case of model animal systems. Some of the recent interesting findings include a report by Lamason *et al*. (2005) who demonstrated association of an SNP (T111A) in a gene encoding cation exchanger protein, SLC24A5, to skin pigmentation in zebra fish and extended the analysis to humans. Recently, the association of SNPs at various human diseases such as BRCA1 gene related to prostate cancer (Douglas *et al*., 2007), VDR gene to Alzheimer's disease (Gezen-Ak *et al*., 2007). Further, the association of regulatory region SNPs to the various diseases in human beings was reported recently with Cyclophilin A gene to AIDS progression (An *et al*., 2007), apolipoprotein E gene to Alzheimer's disease (Belbin *et al*., 2007), TNF gene to stroke risk in children with sickle cell anemia (Hoppe *et al*., 2007) and resistin gene with obesity and metabolic syndrome (Norata *et al*., 2007). Information on the location, frequency, role and distribution of SNPs is limited in plants compared to that of animal systems, due to the expensive strategies involved in the process of identification to association of SNPs to the trait.

In case of plants, an array of studies though not exhaustive have been carried out in model systems such Arabidopsis, soybean and some cereal crops. In Arabidopsis, Cho *et al*. (1999) exploited genome wide mapping approach and constructed a genetic map with 237 markers. Drenkard *et al*. (2000) reported 67 SNP-AS-PCR markers and Schmid *et al*. (2003) analyzed ~10,000 EST sequences generated from 6 different cDNA libraries and reported 8688 SNPs. Torjek *et al*. (2003) reported 100 SNPs, and genotyped in a BC3 population comprising of 48 lines. Hao *et al*. (2004a, 2004b)

analyzed a drought stress associated gene, CBF4 in 17 *Arabidopsis thaliana* accessions, and reported 1 non synonymous SNP (G/T) resulting in G205V change of the encoded protein and proposed association of a haplotype with 4 SNPs with drought tolerance phenotype. Bakker *et al*. (2006) analyzed leucine-rich repeat (LRR) region from 27 disease responsive (R) genes in 96 *Arabidopsis thaliana* accessions for SNP Identification. Meaux *et al*. (2006) reported 89 SNPs (42 SNPs and 47 indels) studying 5' flanking region of CHS gene in 15 accessions of *Arabidopsis lyrata* and, 53 SNPs from 8 accessions of *Arabidopsis halleri.* Schmid *et al*. (2006) analyzed 115 genome-wide SNPs in 351 *Arabidopsis thaliana* accessions from different geographical regions and found evidence for population structure with in the region and suggested to choose populations from the regional area rather than wide geographical area for LD mapping. The association of fruit-set to 3 haplotype groups based on SNPs in GIGANTIA gene was reported recently (Brock *et al*., 2007). Shultz *et al*. (2007) reported syntenic association of Arabidopsis with that of soybean and found Arabidopsis genome sequence information is accurate enough to identify potential SNPs in soybean genome based on the available EST sequences in soybean.

In soybean, three SNPs with nucleotide diversity (θ) of 0.00085 were reported at Gy_4 glycinin locus (Scallon *et al*., 1987; Xue *et al*., 1992) and two SNPs were reported in A_3B_4 subunit of 789 bp glycinin gene (Zakharova *et al*., 1989). Zhu *et al*. (1995) reported nine SNPs in an RFLP fragment of 400 bp in 3 genotypes with nucleotide diversity of 0.015. Zhu *et al*. (2003) analyzed 28.7 kbp coding, 47.6 kbp non coding regions in 25 genotypes to with an attempt to generate a transcript map of soybean, and reported a total of 233 SNPs and 47 Indels with a less diversity in coding

space (θ = 0.00053) than in non coding regions (θ = 0.00111) with a mean diversity value of 0.00097. Kim *et al*. (2005) demonstrated the power of SNPs as markers to correlate with a phenotype with lipoxygenase and supernodulation gene, GmNARK with AS-PCR based SNAP technique. Choi *et al*. (2007) reported 5551 SNPs and the first transcript map of soybean was generated by re-sequencing 1141 STS regions. Yoon *et al*. (2007) identified a set of 23 informative SNPs to distinguish cultivars located on 19 linkage groups of soybean which are expected to uniquely distinguish 2200 soybean cultivars.

The exploitation of SNPs as tools for MAS in other plants include 37 SNPs in Adh gene (Grivet *et al*., 2003), 58 SNPs in 69 EST contigs (Cordeiro *et al*., 2006) of sugarcane, 3 SNPs in trnK gene of Curcuma (Sasaki *et al.*, 2004), 1 SNP in NCED2 gene involved in carotenoid biosynthetic gene of carrot (Just *et al*., 2007), 1 SNP in psbA gene of onion (Cho *et al*., 2006). Conversion of CAPS marker to an SNP based loop-mediated isothermal amplification (LAMP) marker was reported by Fakuta *et al*. (2006) which is known to have association with shelf life in *Cucumis melo.* Seven SNPs were reported in a gene encoding fatty acid elongase in mustard associated with erucic acid synthesis (Gupta *et al*., 2004). Lai *et al*. (2005) reported 243 DHPLC based SNP markers from sunflower analyzing 535 regions based on the EST generated from *Helianthus annuus, Helianthus paradoxus and Helianthus argophyllus.* Recently, a dense genetic map with 483 SNP based genetic markers was derived in *Vitis vinifera* (Troggio *et al*., 2007). A total of 41 SNPs were reported in *BglA* gene encoding *cyanogenic β glucosidase* and 39 SNPs were reported in *α-hydroxynitrile lyase* gene of cassava and related to the geographic origin of the cultivated Manihot species of

cassava (Olsen *et al*., 2004). Techaprasan *et al*. (2006) analyzed a region of ~2.8 kb spanning 3 genes and reported 45 informative SNP from among the 145 SNPs identified in different species belonging to *Boesenbergia* genus of Zingiberaceae family which can be used for species identification. A total of 58 SNPs from 236 clusters in white clover (Cogan *et al*., 2006) and a set of 53 SNPs including 5 indels were reported (Moreno-Vazquez *et al*., 2003) in a recessive gene conferring resistance to corky root rot in lettuce (*Lactuca sativa*). In tree plants 12,264 SNPs among 6,459 ESTs (Pavy *et al*., 2006) in white spruce (*Picea glauca*), Reale *et al*. (2006) reported a set of 8 SNP markers coupled with 1 CAPS and 2 SCARS used to discriminate origin of cultivars in olive (*Olea europaea*). A set of 24 SNPs were reported in a members of WRKY gene family in *Theobroma cacao* (Borrone *et al*., 2004). A total of 103 SNPs including 58 transitions, 40 transversions, 2 Indels and a trinucleotide deletion was reported in regions of ~3.7 Kb in fruiting mei (*Prunus mume Sieb. et Zucc*) and used to trace origin of cultivars in China (Fang *et al*., 2006).

2.1 Single Nucleotide Polymorphisms as tools in cereal genomics

Relatively, cereal crops were widely studied taking SNPs as tools for genetic and genomic analysis. In maize, Tenaillon *et al*. (2001) surveyed 21 loci distributed on chromosome 1 in 16 land races and 9 inbred lines and reported SNPs with the frequency of 1 polymorphism per every 104 bp with the level of diversity greater than human and Drosophila. The analysis of 14,420 bp aligned sequences in 25 maize lines helped them to identify 522 SNPs, of which 60 were non-synonymous changes and concluded that association studies needs 1 SNP to be identified for every 200 bases in maize. Thornsberry *et al*. (2001) analyzed *Dwarf8* loci in 92 maize inbred lines and identified a total of 38 SNPs, which include 19 synonymous and 19 non-synonymous

SNPs besides, 85 SNPs in 5' upstream region of the gene, of which 48 are point mutations and 37 are insertion/deletions. Further, they attempted to correlate these SNPs to flowering time in maize and succeeded in associating a suite of SNPs to the flowering time which include SNPs at 1964, 3490, 3570 and a 6 bp deletion at 3472 in coding region which spans from 1933 bp to 3680 bp of the 3680 bp analyzed region. The upstream region polymorphisms associated with the analyzed trait include 117 bp deletion and 484 bp MITE at 185, 18 bp deletion at 702, 345 bp MITE at 1044, 181 bp MITE at 1663 on ~2 kb targeted 5' flanking region of the gene. A total of 169 SNPs were reported by Ching *et al*. (2002) analyzing 6935 bp representing 18 loci in 36 inbred maize lines with SNP frequency of 1 per 31 bp in non coding regions and 1 polymorphism per 124 bp in coding space. They further derived 6, 4, 4 haplotype groups based on sequence diversity in Adh1, stearoyl-ACP desaturase, acetyl-CoA C-acyltransferase genes respectively from the analyzed 36 maize accessions. Whitt *et al*. (2002) analyzed the 6 genes involved in starch biosynthetic pathway viz. sucrose synthase (sh1), ADP glucose pyrophosphorylase large sub unit (sh2) & small sub unit (bt2), granule bound starch synthase (wx1), starch branching enzyme IIB (ae1) and debranchin enzyme (su1) from 10 accessions of Z*ea mays* ssp. *Parviglumis* and found relatively less divergence at nucleotide level than other randomly selected set of genes in maize and emphasized the need for the introgression of allelic variation from teosinte at selected genomic regions. Wilson *et al*. (2004) further analyzed these genes in diverse inbred lines of maize and associated SNPs in bt2, C925T causing P22L affecting the oil and protein levels. In sh1, A1210G and T775C associated with amylase content and, in sh2 a 1 bp deletion in intron 8 at 3674, a 11 bp deletion in intron 10, a

67 bp deletion in intron 13 and a 8 bp insertion in the promoter to the kernel composition traits. In, ae1, G1509A causing R58G associated with pasting temperature and T1689C in the intron with amylase content. They further carried out haplotype analysis with sh2 and confirmed the involvement of sh2 in starch viscosity properties and amylose content. Shin *et al*. (2006) further analyzed another panel of maize inbred lines and derived SNAP markers based on the 25 SNPs identified in the above 6 loci associated with starch synthesis. Maize genomic resources also served as basic *in silico* analysis to identify SNP as shown by Batley *et al*. (2003) who explored all the available maize EST sequences in the public domain and identified a set of 14,832 SNPs using autoSNP program and validated a total of 264 SNPs from 27 loci in 4 maize inbred lines.

In wheat, Somers *et al*. (2003) reported SNP frequency of 1 per 540 bp by analyzing ~90,000 EST sequences and designed 45 specific primer pairs for the corresponding alleles of the randomly selected contigs. Zhang *et al*. (2003, 2004) identified and devised 3 ASPCR assays for gamma-gliadin loci to asses the allele at Gli-A1, Gli-B1 and Gli-D1 loci and associated with wheat glutenin content determining the quality of wheat products. Huang and Röder (2005) analyzed puroindoline b (Pinb) gene in 493 European varieties to identify 7 Pinb and associated with hardness of the grain. The SNPs in alpha amylase inhibitors were identified to be associated with active site region (I105V) involved in the amylase inhibition (Wang *et al*., 2005). They have also reported 35 transitions and 10 transversions in this 24 KDa alpha amylase inhibitor causing at least 10 amino acid substitutions. Lagudah *et al*. (2006) analyzed Lr34/Yr18 slow rusting resistance genic region based on the EST sequences and identified a 79 bp

deletion in the intronic region of the gene and attempted to associate it with the rust resistance. Ravel *et al*. (2006a) analyzed a storage protein activating prolamin box binding transcription factor, and reported low level of diversity at this locus (viz. 1,5,1 SNPs in A,B,D genomes of wheat respectively) based on 27 wheat accessions. A total of 64 SNPs were reported by Ravel *et al*. (2006b) by analyzing 21 genes in 26 accessions occurring with the frequency of 1 per 335 bases. The association of Glu-B1-1 to HMW-glutenin content and discrimination of Glu-B-1 as the candidate from Glu-B-1 and spa-B on Glu1Bx was established by Ravel *et al*. (2006c). Schnurbusch *et al.* (2007) reported low frequency of SNPs (1 per 613 bases) in Bo1 QTL associated with tolerance to Boron toxicity based on the 16 gene fragments spanning 19.6 kb region. Zhao *et al*. (2007) analyzed Glu-D3 sub unit influencing the visco-elasticity and extensibility of dough in 8 wheat lines and identified a non sense mutation at 119th position (acc. GluD3-43) resulting in the pseudogene, and a 3 bp insertion (acc. GluD3-41) leading to a glutamine insertion at 249th position of the deduced amino acid sequence.

In barley, a total of 112 SNPs were reported (Kanazin *et al.*, 2002) at 38 loci from 5 genotypes and Bundock *et al*. (2003) reported 16 SNPs analyzing 7 regions encoding Cytochrome P450 genes in 11 barley lines occurring with the frequency of 1 per 131 bases. Kota *et al*. (2003) demonstrated utility of *in silico* analysis to identify SNP markers by extensive analysis of the barley EST collection (2,71,630 ESTs) in the public domain representing 23 barley varieties. They have identified 3069 SNPs by using a novel algorithm called SNiPpER and validated 54 of 63 targeted SNPs through re-sequencing approach. Bundock and Henry (2004) have reported 80 SNPs and 23

indels in 2163 bp region covering promoter, genic and 3' UTR of an alpha amylase, bacterial subtilisin and xylanase inhibitor gene called Isa in 16 barley genotypes. Their results suggest 3' UTRs and promoter regions are highly polymorphic having possibility of high rate of recombination taking place in the promoter regions. A set of 2000 SNPs were reported by Rostoks *et al*. (2005b) at 877 loci from 6 genotypes and utilized Affymetrix array to identify 10,504 expression level polymorphism called single feature polymorphisms (Rostoks *et al*., 2005a). Patokina *et al*. (2006) have analyzed serine carboxypeptidase gene in 12 genotypes to identify 2 SNPs in exonic region (Te931-G/T, TE945-C/T) & 4 SNPs in intronic region (Ti105-A/C, Ui143-G/A, Ui208-G/A, Ui218-G/A). They have designated 4 distinct haplotypes and associated to malting quality. Tacconi *et al*. (2006) analyzed Mlo/mlo gene conferring resistance/susceptibility to powdery mildew in 4 Mlo and 3 mlo cultivars and validated in 46 Mlo and 25 mlo cultivars through a PCR based assay to identify 8 SNPs and 1 indel adjacent to Mlo/mlo gene. Recently, Morrell and Clegg (2007) utilized SNPs as the tools to discriminate geographical distribution and origin of domestication using 25 accession with different geographical origin and assayed at 18 loci to identify 684 SNPs.

The studies carried in other cereal crops include conversion of SSCP-SNP markers in pearl millet (Bertin *et al.*, 2005). Cogan *et al*. (2006) reported 1613 SNPs in 100 genes and attempted to validate a set of 238 SNPs in perennial rye grass (*Lolium perenne*), of which 9 SNPs are present in drought stress responsive *Lp*ASR1a2 gene and 16 SNPs in WRKY family transcription factor. Conversion 1 RAPD marker 1 REMAP (retrotransposon-microsatellite amplified polymorphism) associated with

straw height in Dw6 QTL to SNP-RAPD and SNP-REMAP was reported by Kalendar *et al*., (2006) in Oat (*Avena sativa*). Varshney *et al*. (2007) reported 96 SNPs and 26 indels analyzing 14 targeted regions in rye (*Secale cereale*) occurring with the frequency of 1/58 bp with nucleotide diversity index (π) of 0.0059 to 0.0530 and converted set of 12 SNPs in to CAPS markers.

In rice, a total of 2800 SNPs were reported by Nasu *et al.* (2002), analyzing 418 randomly targeted intergenic regions from 3 genotypes, of which 94 SNPs are converted into markers with FPTDI assay. Monna *et al*. (2006) randomly targeted 1117 regions in 9 genotypes and reported 7805 SNPs, and derived allele sharing maps after integrating with genetic map based BAC clone map of IRGSP. In above two cases, the focus was on identifying a large number of SNPs by targeting randomly selected intergenic regions but not relevant to any specific trait. In rice, the reports on dissecting the genomic regions associated with a complex trait for SNPs and converting them into molecular markers are scanty, and so far, no study has been carried out with drought tolerance. SNPs were identified for specific loci associated with certain traits like *Xanthomonas* resistance with SOD gene for xa13 (Kottapalli *et al*., 2007), starch synthase gene for gelatinization temperature (Waters *et al*., 2006), waxy gene (Bao *et al*., 2006a, 2006b) and CMS (Duan *et al*., 2007), seed shattering (Konishi *et al*., 2006) which can be exploited in MAS. Feltus *et al.* (2004) and Zhao *et al*. (2004) exploited *indica* and *japonica* genome sequences for *in silico* comparison to identify a large number of SNPs available in the public domain.

4. MATERIALS AND METHODS

3. MATERIALS & METHODS

3.1. Materials

3.1.1 Chemicals and reagents used in Molecular Biology

Reagents used in total RNA isolation, cDNA synthesis were obtained from Invitrogen Corporation, USA. Paramagnetic beads for mRNA isolation were from Promega Corporation, USA. All chemicals and reagents used in cDNA library construction were obtained from Stratagene, USA. DNA Sequencing and genotyping reagents were obtained from GE Healthcare, Hong Kong (Formerly Amersham Biosciences), Genetix, UK. *Taq* DNA polymerase, molecular weight markers from Genotex International (I) Pvt. Ltd and RNase A, *Pfu* DNA polymerase were obtained from MBI Fermentas. Oligo nucleotide primers of desalted grade, HPLC purified were synthesized from MWG, Germany, Sigma Genosys and LNA modified primers were obtained from Sigma-Proligo, Singapore. Tris, EDTA, NaCl, CTAB, PVP, β-mercapto-ethanol, ethidium bromide, sodium acetate and Agarose were obtained from Sigma chemicals, USA. Chemicals for bacterial media preparation and solvents chloroform, iso-amyl alcohol, phenol, and ethanol of analytical grade etc. purchased from local sources.

3.1.2. Plant genotypes used for subtracted cDNA library construction

(A) NERICA: NERICA (*Oryza glaberrima* cv CG14 X *Oryza sativa* cv WAB56-104) is the product of interspecific hybridization between the cultivated rice species of Africa and Asia has been obtained from Africa. It is a high yielding anther culture rescued selection with high protein content and highly tolerant to drought. This aroma tasting drought tolerant variety typically mature in 90–100 days, compared to typical improved upland sativa varieties that mature in 120–140 days. It is resistant to African rice gall midge, rice yellow mottle virus and blast disease. Beside it is has got the best

of their parental lines like non-shattering grains, secondary branches on panicles, responsiveness to mineral fertilization. (**B)** IR64 is a drought susceptible genotype belongs to indica subspecies obtained from GKVK, Bangalore.

3.1.3. Plant genotypes used for identification of SNPs in the target regions

A set of 8 genotypes (Table 3.1) with well characterized phenotypic response under drought stress conditions were extensively utilized in the present study for identification of SNPs in the selected target regions involved in drought stress response. These genotypes were obtained from various sources as part of National Rice Biotechnology Network project. These genotypes were earlier utilized for validating EST-PCR based marker based randomly selected drought responsive ESTs (Chandrasekhar, 2005) from normalized cDNA library of Nagina22 (Reddy *et al*, 2002). These genotypes include the parental lines which are widely used for mapping QTLs associated with drought stress response. The DHL population developed with the cross of CT9993 X IR62266 was extensively utilized for mapping traits such as root thickness and penetration index (Zhang *et al*., 2001), a set of root traits (Kamoshita *et al*., 2002), osmotic adjustment under drought (Zhang *et al*., 2001), yield, biomass, osmotic adjustment, roots (Babu *et al*., 2003), yield, yield components, panicle sterility (Lanceras *et al*., 2004), CMS under drought stress (Tripathy *et al*., 2000). Another DHL population developed with a cross of IR64 X Azucena was used for mapping traits associated root morphology, root distribution (Yadav *et al*., 1997), root thickness, root penetration index (Zheng *et al*., 2000), yield and root traits under limited water conditions (Venuprasad *et al*., 2002), leaf rolling, leaf drying, RWC, growth rate (Courtois *et al*., 2000), morphological and physiological traits (Hemamalini *et al*., 2000). Besides Azucena was also utilized in generating RIL population with a

susceptible i.e. Bala to map the traits associated with root, tillers, penetrated root number, dehydration avoidance traits, yield under drought stress (Price and Tomos, 1997; Price *et al*., 2000, 2002; Lafitte *et al*., 2004). Nootripathu was used as another tolerant check and IR20 as susceptible check. Further, we have also included Nagina22 genotype which is an upland drought tolerant indica cultivar for which we have generated ESTs under drought stress condition. Nerica-1 which is known to be highly drought tolerant cultivar was also used as drought tolerance genotype. This set of eight genotypes has been extensively utilized in this study for identification of variations at the loci associated with drought stress response to the single nucleotide level.

S.No.	Genotype	Phenotypic response to drought	Description
1	CT9993	Highly tolerant	Japonica type, upland adapted, long root
2	IR62266	Susceptible	Indica type, shallow root
3	Azucena	Highly tolerant	Japonica type, upland adapted, long root
4	IR64	Susceptible	Indica type, shallow root
5	Nootripathu	Tolerant	Indica type, upland, deep rooted
6	IR20	Susceptible	Indica type, shallow root
7	N22B	Tolerant	Indica type, upland, deep rooted
8	Nerica1	Highly tolerant	African line, interspecific cross of *O.sativa and O.glaberrima*

Table 3.1: Different genotypes used for identification of SNPs

3.1.4. Rice genotypes used in validation of SNP markers

The candidate SNPs were validated in a set of 25 rice genotypes listed below (Table 3.2) with varied response to drought tolerance. The genotypes used in identification of SNPs by sequencing were also utilized in development and validation of PCR based SNP markers. These are represented as first set of 8 genotypes in the panel and another set of 17 genotypes were utilized for validating the AS-PCR markers in the study. This extended set also represents genotypes with diverse phenotype with reference to drought stress.

S.No.	Genotype	Phenotypic response to drought stress
1	CT9993	Highly tolerant
2	IR62266	Susceptible
3	Azucena	Highly tolerant
4	IR64	Susceptible
5	Nootripathu	Moderately tolerant
6	IR20	Susceptible
7	N22B	Moderately tolerant
8	Nerica1	Highly tolerant
9	Purpleputtu	Moderately tolerant
10	Swarna	Susceptible
11	SL44	Highly tolerant
12	WR119	Moderately tolerant
13	N12UP	Moderately tolerant
14	N11BIH	Moderately tolerant
15	TP306	Highly tolerant
16	IR42253	Susceptible
17	Poornima	Moderately tolerant

S.No.	Genotype	Phenotypic response to drought stress
18	Safri	Moderately tolerant
19	MM125	Moderately tolerant
20	DRRH1	Moderately tolerant
21	Mahamaya	Moderately tolerant
22	Co43	Moderately tolerant
23	W1263	Moderately tolerant
24	Nerica2	Moderately tolerant
25	Pusa Basmathi	Susceptible

Table 3.2: Different genotypes used for genotyping based on validated SNPs

3.1.5 Database of target regions associated with drought response

A MySQL database with the target regions was developed based on the comparative expression profiling study we have conducted with the expression profiles of other model plant gene expression profiles (Gorantla *et al*., 2007). A set of genes from this panel of stress responsive genes (SRGs) particularly, class of genes with signal transduction, kinase, hormone responsive and transcription factor activity were selected and considered as target regions. Further, a set of genes showing dramatic gene expression as the drought stress increases were considered from series of field drought stress experiments on Nagina22. This set of genes was complimented with other genes which are reported to be drought responsive besides a set micro RNA genes associated with drought stress response. This database served as the basis for the candidate gene targets and promoters of these regions were extracted from the gene prediction algorithms utilizing rice genome sequencing data.

3.1.6 Primers used to amplify the target regions involved in drought stress response

Oligonucleotide primers used in this study were designed for the target regions using different primer designing algorithms from the above discussed database and synthesized. The set of primers used for the target region amplification and type of the target region considered from comparative analysis, microarray gene expression profiles and other known drought responsive genes were listed in Table 3.3. The regions considered from differentially expressed genes in Nerica-1 are listed in Table 3.4. The set of target regions associated with drought responsive miRNAs and their biogenesis were listed Table 3.5. Type of region is mentioned as 'Gene' if the target is either partial region of the gene or full length gene encoding the corresponding annotated protein. In case of lengthy genes only conserved region or the 3' region of the gene was considered as to capture more variation. The promoters were considered as upstream of the transcription start site and in certain cases 5' UTR were also considered to be the part of the target region and annotated as promoter. This set of target regions considered to be exhaustive to capture the drought responsive variation both in the coding and regulatory regions. Though the focus was more on capturing the regulatory variations, coding, 5' untranslated regions (5' UTRs), 3' untranslated regions (3' UTRs) and introns were also considered in certain cases and used in the study.

S.No	Primer No.	Function associated with the target locus	Forward Primer Sequence	Reverse Primer Sequence	Type of the region
1	1	14-3-3	CGGACAGGATGCCGCCAATCC	TCAAGGAAGCAGCGAAGCCTG	Gene
2	2	Calmodulin E2	ATTTATCGGATCGGGTGCAGC	ACGGCAATGCACGGTTTCCTC	Gene
3	3	Purple acid Phophatase	TCGGATTGTCAAATGCAAGTCG	TTCGGGCACGCCATCCTGGAC	Gene
4	4	RAS	CCATTGCATCATAACTGGACATCC	TGCTTCAGTAGAAACATGCAGTG	Gene
5	5	Calmodulin	TCGACGTTGCACCGTGTGACC	TCGAGCGGTTAGCGAAGAACG	Promoter
6	6	Protein Phosphatase	TGGAGGCGCTTGTTTCGCGTG	AGACGCCGAATCCTCGTCAGG	Gene
7	7	Ca^{+2} binding protein	TCGAGCTGCAAAGGAGCTAGC	ATCCGTGCATACGCTGTGTAG	Gene
8	8	nifU protein	CAATGTGCCTTATCTCGCGAC	TCCTTGGCAGTGGAGTGTGTC	Gene
9	9	Anthocyanidin synthase	ACAGGGGCATTATTGACAGG	ATAGGCCCACTCCTATCCAA	Promoter
10	16	Dihydro flavonase	TGGCGGGTGAGACGTAAA	GATAGCTTGCCTGTTGCTCCT	Promoter
11	25	Dehydration Response Binding protein (DREB)	TACTGATGATCGCGAGTTGG	ACCATAAGCAGAGCTGGCAT	UTR
12	27	Metallothionein	GACTGTGCTGACAAGAGCCA	AGCACGGGTACATCACATCA	Gene
13	37	MAP kinase	GGATCCAGTAAATTTGGCAG	AGCATGTATAATAGCAGATTATTAGTCAG	Promoter
14	38	26S Proteosomal protein	CCCTTCTTTTGTGTTCGAC	CCTTACCCGTAACCTGACAC	Gene
15	39	DNA binding protein	GGAACGGAAAAGGTCTCTC	GGAGAATGGTAGCTTGGC	Promoter
16	40	Ring finger protein	CTTGAACGGATGTTATCTGC	GCTAAGAATGGAGAGATAATTCAG	Gene
17	41	KNOX	TTCAATTTCTAGGCACAACTAAG	TGTTGTTGTAGCAGTTGGG	Promoter
18	42	MYB	CCTTTACTTTAGTTTCACACATCC	GGGCGGAGGTAGTTCATC	Promoter
19	43	Zinc finger protein	TGTGTGTGTTTGTGTCAACC	TGATTTGCTTCATGATTTGTG	Promoter
20	44	Nucleic acid binding protein	CATATCCACTCACCGCTG	CTAAATGTTTATTTTGGGCG	Promoter
21	45	Nucleic acid binding protein	TGAATAGTTTTGAAACTATAGGAATTG	CTAGCGGACAGAAGCATATC	Gene
22	46	Zinc finger protein	TGATGAGATGAACTGCACG	GATCTGAATGAATGCTTTACTTACT	Gene
23	47	RNA binding protein	GGGGCAAAGGTATATGTAAAC	TGTTTGTTTGTCACTTGCC	Promoter
24	48	RNA binding protein	ATCTCTTAATACCCCTTAACTGC	TGGGCATGTGAGAAATTG	Gene
25	49	Zinc finger protein	CCGATGAATCGCCATTAC	TTGGTTGACGGGGAATAC	Promoter

26	50	Defensin	CAGCCGTACCCTGAATGC	GATGAGTGATGAGGCTGCTAC	Promoter
27	51	Defensin	ATGGCTCCGTCTCGTCGCATG	CTAGCAGACCTTCTTGCAGAAGC	Gene
28	52	Zinc finger protein	CCAGAAAATAGAGAATACGTATCG	TTTGGATGGGGTTGTACG	Promoter
29	53	Metallothionein	ATGTCTTGCTGCGGAGGAAGC	TTATTTACTTGTACATATCATCAGCAT	Gene
30	54	Dehydration Response binding protein (DREB1B)	CATCCATGGAGGTGGAGGA	TGCACATTCAAAATCAGATGG	Gene
31	55	Glutathionein reductase	TTCCAATCTTCTCAATCCAC	GCAAGAACTAGCAACCGTG	Promoter
32	56	Glutathionein reductase	TTTGGATGGCATAACTGTG	GCCTTGAAAGTTGCTCCA	Gene
33	57	Cell rescue associated gene	TTGTTGGAGGAGCCCTTA	GTGACACCGTAGCACTGG	Promoter
34	58	Glutathionein S transferase	TGTCTTTTGCTCAACATTAAAGAT	AAACGACAACGAGGGATG	Promoter
35	59	Glutathionein S transferase	CCAGGACCATGCTGTTATT	AACTGCCAGGTCGATTAAC	Gene
36	60	Disease responsive gene	TCGAGGATGGTGTCGATAG	TAGACGGGCAACGATATG	Promoter
37	61	Disease responsive gene	CCACGCTACAAAAGGTCA	TCGGTAGGTGAGGTGAGAA	Gene
38	62	Chitinase	CACAATCGATAGGTAGCTAACTACT	GAGACCAGGCTCATTACCA	Promoter
39	63	Chitinase	AAACACGTTGATTCGATTAAA	TTAGCTACCTATCGATTGTGATG	Gene
40	64	Drought stress responsive Gene	CCTTGCCCTCTGTGTGTC	GAAATGCATTCTCCTACATCAGT	Gene
41	65	Drought stress responsive Gene	CCCATCTCTCTCTCTCAGGAC	GGAGAATGAGATGGAAGTGAC	Promoter
42	66	RNA binding protein	TCTGAACGTCTGAAGCTACAT	TGAATTGCATGCTGAATGT	Gene
43	67	RNA binding hormone responsive gene	CCAGCGTTACAGGTCACTTAT	GAGGAGGGCGTACTCTTG	Promoter
44	68	DIM	CCATCTTCATTGACCTCCTA	TGCTTCTTCTTCCCCAATC	Promoter
45	69	DIM	GGCAGTAGCTGAGACCATAA	AATCAACTTGTCATTGATGTACC	Gene
46	70	KNOX E4	TGAACAAACTTTACAGAGCCTT	AGCTTTTCTCTGTACCCCTG	Gene
47	71	KNOX	CAGAGAAAAGCTTGTGGACAT	GGAAGCTTCTGTCCGAATC	Gene
48	72	Transcription factor-II	GGCCACCTGCACACTTAC	GGTAATGCCAAATGTTGCTA	Gene
49	73	Hypothetical protein	TCTCGGTTCTCGACTTTCTA	AAATTTGCATCAGTTGCG	Gene

50	74	Auxin induced protein	CTGCAGTATTTTGAGAGTTTCTTA	AAAATCAAGAAACCCAAATATGT	Promoter
51	75	Auxin induced protein	AACCTCACTTTGCACAAAGAT	CGGTCATACCGGAAGAAC	Gene
52	76	Hormone responsive protein	CAAGTTAGGATCTGGACGC	CATGACTCTTACCATGAATCCTT	Promoter
53	77	Metallothionein 3UTR	GGCAAGTGAAGTCACGAA	TCTCAGTGCTATATAAACAACGATTA	UTR
54	78	Protein kinase 3 UTR	AATCCAGATGCTGTTCATGT	AAACGCCTCCAAGAGACC	UTR
55	79	Protein kinase	ATGCGCGACCAGTAATTC	AATGATCTGCTACTAAATTTCACA	Promoter
56	80	ABA induced gene	TCAGGACGAATTAATCAATAATAATA	GCTAGTGATGACAATTAGGAGAAT	Promoter
57	81	ABA induced gene 3UTR	TCGTCGATTTGTGTGTGTAAT	CAACTACCAACCACTTTAATTATGAA	UTR
58	82	Inositol phosphate 3UTR	AGTGGTGGGGAGTTTGAT	ATTCTGATTAAGCTACTTCGAATCT	UTR
59	83	Inositol phosphate	CAGTTTTAGTTTCCAGAAATGTTT	ACATTCGAAATATCGGTGATAG	Promoter
60	84	IGPS	TTTCGTAATCTAAGGCTGGA	TCAGTTTATAAAACTTTGGTTAAAGTC	Promoter
61	85	Ethylene responsive gene 3UTR	TTCCCTTTCGAGGGCTAC	GCATAAGAGGTACACGGAATTT	UTR
62	86	Ethylene responsive gene	ACTTCGCAGTCCAGTACCTA	TGGTAAGACAAGTGATATGTGGT	Promoter
63	87	CBF4	ATGGAGAAGAACACCGCC	CTACCTCCTCCATTGAAAAAG	Gene
64	88	Cyclophilin	ATGTCGAACACGAGGGTG	CTAGGAGAGCTGGCCGCAG	Gene
65	89	Cyclophilin	GGTTTCATATGATTCAGGGAGGTGAC	CAAGCTTGCTACATCTCACCACAC	Gene
66	90	Cyclophilin	ATGGAGGGCGGCGGCGAG	GGGATCACTTGCTCTCCTACAGG	Gene
67	91	Hormone responsive gene 3UTR	CTAGCTCATTTGATTGGTGTG	ACACTAAGATAGCAGTGAGCCA	UTR
68	92	Metallothionein	TGGCTTGTTTGGTAGAGTTTC	TGCTTTCTTGTGCTCTGGT	Promoter
69	93	MAP kinase	TCAGTTTGGGGGTGAAATG	GAGTGGCTGGTTATAAAACTTCTTC	Promoter
70	94	betaine aldehyde dehydrogenase E1	ATGGCCACGGCGATCC	CAACCTATGGCCAATTCCTAGTCCAG	Gene
71	95	betaine aldehyde dehydrogenase EL	CGCGAGCGATGCCAGAG	TTACAGCTTGGAAGGGGATTTGTAC	Gene
72	96	betaine aldehyde dehydrogenase	AAGCGGCGGTCAGACTG	ATAGTACTAGTATGTATCGCATCCG	Promoter
73	97	Choline monooxygenase	AGATTGTTCCAAAGGACTTGGAC	AGAGATGGAAATTTGGTTTGGCTC	Promoter
74	98	Choline monooxygenase	GAATTTGTCATATGTCGGGATGC	CAAATTTGGCCAATACAAAAGGCC	Gene

		CD			
75	99	Cyclophilin	AGCTGAAGAACTGATTTCTGATGAG	ATGCCCGTGCTCGAACTC	Promoter
76	100	Ribosomal 6 kinase	CTGGCCGAAAATTCCACTTGC	GCACTACAAAAGGATGGTCAAC	Promoter
77	101	Ribosomal 6 kinase	ATGAAAGTTATGAGGAAGGATAAGATC	CTAGTCTTTCAGCCTAGAGGAAG	Gene
78	102	MAP kinase kinase	ATGCAGTCTCTTTGTAGACACATC	AGAATCAGGTGGAAAGAGAACGAG	Promoter
79	103	MAP kinase kinase	ATGGAGTTCTTTACGGAATACG	TTAGGATTGGAGTTTAGATTGATTATGC	Gene
80	104	M_Ring finger protein	GGATATGAGCCGAGTAAAGCTGCGG	GCTGAAGACGCTAGGGCTTG	Promoter
81	105	M_Ring finger	CTGACTCTGGGCAGGTATCCTAGG	ACAACACCTGTTAAAACTGCTCCAGC	Gene
82	106	M_Gene with MITE	TTCGCGAGACGAATCTATTAAGCC	GCCCACATAATGTGTAAAACCAGTAG	Promoter
83	107	M_Gene with MITE	ATGGACCTCTCCCGCGTC	GGATGGGCCCTTTTGTTGGAACATATG	Gene
84	108	M_Gene with MITE	GAGCTGAATTTTATCCATGACTGCTG	TTAGACAATCAGTGGAACCAAGATAAGTG	Gene
85	109	M_Unknown gene	GCCGTACCTTCCAGAGATTTGC	TCACTGCCTGAAATGCTGCAAAACAG	Gene
86	110	M_Unknown gene	AGTACTAGCTAGTCAGCGTGATC	AGAAGGTGGGTTAATTAGTACTCCG	Promoter
87	111	Mr_NL38C07	ATGTGGAAGGCGCAGCGG	TCACTCATATACACTCATTGCCAATGG	Gene
88	112	Mr_NL38C07_Pro	CGGTCCCAAAATAATCACAAGAGACC	ACATGGCCCATAAGGCTAGGTG	Promoter
89	113	Thioredoxin	ATGGCCGCCGAGGAGGG	TTAGGCAGAAGCAGATGCAGCAG	Gene
90	114	Thioredoxin	TGATCGTTTGATTGTTGGATCACAG	CAGGCGATCACGACTCC	Promoter
91	115	Ferrudoxin	ATGGCGGCGACGGCACT	TTAGATGAGGTCGTCCTCCTTGTGGG	Gene
92	116	Ferrudoxin	ATGTGATGGATGCTTGACATGC	GCGATGTGCAGCCTCTAATAATTTTG	Promoter
93	117	M_Jasmonic acid induced gene	ATGTCGCGGTACGCCGG	TCACTCCTCTGGGAGAGGGCC	Gene
94	118	M_Jasmonic acid induced gene	CCTCTTGGCCTTCCCCAC	GATGAGTGGTCATATTTGCTGGTCG	Promoter
95	119	M_RD22 LE	ATGGCTAGGTCACTCGCTG	TTAATTCAGTTGCGCACCCAGATC	Gene
96	120	M_PSII10KDa LE	ATGGCTGCCTCTGTCATG	CTAAGCCAGTGCACTAGTGTTG	Gene
97	121	M_PSII10KDa	ACTGTCCCGGCAATGTTAGAC	ATGCAGTTCATCATCAGATATCCATGGC	Promoter
98	122	M_Cold Induced gene LE	ATGAGCAGCAGCAGGCTC	CTACTGATGGTAGCCGTTCTTGTTTAC	Gene
99	123	M_Cold Induced gene	CGTTTCGGGAACTTGCAGG	TGGCGCGTGTTATATACGG	Promoter
100	124	CBF4	GCAAACATCGTTGACCGGTG	CGGACCATTTGCTGCGAAATG	Promoter
101	125	bZIP	ATTTTAAGAGAGGAGTGTATTAGCTCTG	AAATCTAGACCTACCACACCC	Promoter

102	126	bZIP	ATGGGTAGCAGTGGCGCAG	AGAGCTAGAGAGGAAGTTGCAGC	Gene
103	127	bZIP E2	TGGTAAAACCTCTGGGGCATC	CTACAACGGGAAGAAAATATTACAAGTGC	Gene
104	128	M_AP2 family gene LE	ATGTGTGGAGGCGCCATC	TCAGAAAAGGGCGCCGTC	Gene
105	129	M_ AP2 family gene	AGTGGAAGCTCATCTTTGAAGAGTGG	GTGCGGGTGGACAACGTAG	Promoter
106	130	M_dnaJ LE	ATGGAGGGCAACAAGGACG	TTACTGTGCTGTTGCCTCAAATTTCC	Gene
107	131	M_dnaJ	GGTGTTCGCATTTATAATTCTGACAC	CGTAATGATGGCTCCATCGTTTG	Promoter
108	132	M_DREB2 LE	ATGGCTGCAGCTATAGATCTGTCAG	TTAATTGGCGGCGAGGAGG	Gene
109	133	M_DREB2	CCTGCTGACCAAAAGGAAGTTC	TGCAAGCAACCTAACCAAGAAACC	Promoter
110	134	M_PCF3 GCD	AGTTGGCCGGGATGGTG	CCGTTCTACTACGCCGCC	Gene
111	135	M_PCF3	CGTTCACACAAACACTAACAGCAC	CCATATACTTTCAAACCTTGGAGAGC	Promoter
112	136	M_MADS E1	ATGGCGCGGAGGGGGAG	GATACAAGCTGCTGGCTGCTGCTACTG	Gene
113	137	M_MADS	AAATAGTTGCTTTGGGGATTTGATAC	GCAAGCGCAAAAGCCTC	Promoter
114	138	M_RD22	TGCCTTATATTATGGGATGGAGGG	AGGAGTAGGACAGCAGCGAG	Promoter
115	139	M_Catalase	GGCCCTATCCTCCTTGAGGAC	TGGAATTTGACAAGGTGAGGCTTTCC	Gene
116	140	M_Catalase	GCGACGTCAGTTCAGTTTAGTTTGC	GGTTTCGTAGCGTGCTAGTATGCTTC	Promoter
117	141	M_NAM	TTCAGGTTCCACCCGACG	CTTCTTGTGCATCGGCAGGG	Gene
118	142	M_NAM	CCACTGCTTCATTAATTGTTTGTGTAG	TTCTCACGGGAAATTGATTAGTAGG	Promoter
119	143	M_Glycine rich zinc finger protein	ATGGCGGCGGCGGCGAG	TCACTTGTGGCAGTCGCGGGC	Gene
120	144	M_ Glycine rich zinc finger protein	CGTGTATCGGTACCCGGCAG	CCTCGCCTCTCCCTGATCTGATTAG	Promoter
121	145	M_RING-H2 finger ZFP LE	ATGAGCTACCTCCTCTCCTACATC	TCAGAATCCAGGGACGAGCATG	Gene
122	146	M_RING-H2 finger ZFP	GAGAAAGGGTGAAAGGTTTGAAGTG	GTAGTTGATTTGTTCATATCCATTGGTACG	Promoter
123	147	M_HSP90 LE	ATGGGCGGCGGCAAGAATAAG	TCAGATGAGTGGGCAGGAGAAGTC	Gene
124	148	M_HSP90	GATTAAGTCAAAGTGAATTGCACGAG	TGCTATTTATAGGCAGCTCCAGG	Promoter
125	149	M_Unknown gene homologous to CB965622	ATGATGGATAAATATGCCAGATTAATGC	CCATTGATATGCAAAAGCATCCTTG	Gene
126	150	M_Unknown gene homologous to CB965622	GGTTTCAAGGTTATTGTCTGAAGTTG	TGTTCAGTTGTAAGATTCTGGATGC	Promoter
127	151	M_Unknown homologous to CB966119	CAGATGCCGCATATGATCTTCAC	ATGGTTGGCATTTCAAGCACTTG	Gene

128	152	M_Unknown homologous to CB966119	TGTGCTTGCTCAGTTAATTCCTC	TAGGGATACAAAGTTGTTGTCTCG	Promoter
129	153	M_Gigantia like LE	ATGGAAGCTCTCCTCTCCTCG	TTACCTTGATGAGCATGTAAGTTGGC	Gene
130	154	M_Gigantia like gene	ATAATTAATGGCAGTTTCCACATGGC	AATGGTCATCTTCTACCTTTGCTTCTG	Promoter
131	155	M_RS6 kinase GCD	GAAGGATCCATTGGAGATAAGATGG	GGGCGGTCTCTAAAGTCATATTC	Gene
132	156	M_ RS6 kinase 5UTR	CACCGCTTCCCCTCCTTC	CCGCATATACACCACTTAACATAGGCTC	UTR
133	157	M_Water channel GCD	TACATCGTGGCGCAGTGC	TACCGGGACATGGGAGTCG	Gene
134	158	M_Water channel protein	GGTGTCCATCAATAAATATCCAACC	GTTTTGTGAGAAGAACAAGAGGATG	Gene
135	159	DREB1B	CGGGATCCGCGAGGTAAGCCATTAGCGCATG	CGGAATTCGGATGACTCTCTCTGGTTCAC	Promoter
136	160	DREB1B	ATGGCCGACGAAGAAGAAGAC	AGTAGCTCCAGAGCGGCATA	Gene
137	161	DREB_Cons	AAGTTTCGTGAAACTCGTC	TCAATAAGACCAAAG	Gene
138	162	Aquaporin	AGAACTCAGCTTAATTAGTTGCCAAATGCG	GCCGCTCTCCATCACCTC	Promoter
139	163	Aquaporin	ATGGGGAAGGACGAGGTGATG	GCTTCTCTTCTGTTCAATATGCGCTTG	Gene
140	164	Aquaporin LE	GTTTGGCAGCAGCACAGTTTTGTG	TCACGCGTTGCTCCTGAAGG	Gene
141	165	Chalcone Isomerase	GCTATCACTCGTCATTTCCACAAGC	CTAATGTCGACTGTGAGAAATCGGTG	Promoter
142	166	Chalcone Isomerase	ATGGCGGCCGTGTCGG	TCACGCGGACACCGGC	Gene
143	167	Phenylalanine ammonia-lyase	TCTAAGGTGTAGTGACCTTGCACTG	GTTGCCCGCCATTGCTAC	Promoter
144	168	Anthocyanin reductase	TCCTATTCAAGCAGCCCCTAAGTC	ACAGTCAAAGTTGAACACGGAAATCC	Promoter
145	169	Coumaryl CoA Ligase	GCGTGGCTGCATGCATTG	TCGACGGTGGCCATTGG	Promoter
146	170	Pathogen responsive (PR) protein	GAGTGAAAGAGTGTACCTGAGGCTC	CAAGGCAAAGCAAAACCGTGC	Promoter
147	171	Disease responsive gene	GGGAGAAGATAGGCACGTCAAAGAG	ATGATTAAAGTCTCCTTCCGGTGCAG	Promoter
148	172	Wound induced gene	CACAGTACAGTTAGTCCTGATAAGCAGG	GGATCACCAAGAGAGTGCTGC	Promoter
149	173	Thaumatin	GCCAAACGATGGGGACCTTAG	GCTCCATGTTGGAGTGTAGCC	Promoter
150	174	Thaumatin	ATGGCGATGGCGGTGCG	CTAGCGGCGGTGCGGG	Gene
151	175	HSP17	AAGACGCCGGAGAGAAAGTGTG	AGGCGTCGAACGCCAAC	Promoter
152	176	HSP22	CCAAAGTAAATCCTAGGGCCTGTTC	CCCTCCGCGACATGCAG	Promoter
153	177	S-adenosyl-methionine	TCTCTAGCTCTGACTCAAATCGTACTGC	CAACGCTACCAGCAACGAGG	Promoter

		Synthase			
154	178	APX TL29	CTACCATAGTTGATCTATCCAGACCTG	GCGTGAACAGATGTGACAAATATG	Promoter
155	179	APX	ACGGATGGTCAAACGTTGGAC	CGAAAGGCTTGGGCGAATG	Promoter
156	180	HSP82	TCCTGCCAAAATGTTGGCAATATTG	TGAGAGGGCGAGATCGTG	Promoter
157	181	HSP60	GCGTTTGTGTAAAAGTAGCTAGCAGAAC	AGAGGATAAGGCTGCGACCTC	Promoter
158	182	Heat shock Transcription factor	GGCGCGAGACCGAACTAAG	TGGTGGCGCGGAATTTATAGG	Promoter
159	183	CBF1	TGTGACAGTATCGCTTCATAACATGTGG	AGGAAGGCTTTTTCTGTGGGC	Promoter
160	184	CBF1A	ATGTCGAGGCTCGATTTGATGAGAG	TCATCCAAGCACCACGCC	Gene
161	185	Mn superoxide dismutase	CACATTTTCTTTAGGAGCTTAGCTCTTG	TTCCTCGAGGCCAGCG	Promoter
162	186	Dehydroascorbate reductase	TCATTTTTAATCAATGACGTAATTTGTACTG	AAAGCTGCCCAAAATCCATC	Promoter
163	187	Dehydrin	GAAGACAGAGGTAGGTTTTGTTAGAG	AGCTAAAGCTGAAAAAAAACTCTCGGAG	Promoter
164	188	Dehydrin LE	ATGGAGCACCAGGGGCAGC	TCAGTGCTGGCCGGGCAG	Gene
165	189	Helicase	GCGAGAGAGAGTTTTTCCGATTCG	GCCCAGATAGGAAATACAATTTCAACTTGGTG	Promoter
166	190	CBF3	CGCGGAAAACGGAATAAACTATCTCCC	CCGGGTGCCTCGTCTC	Promoter
167	191	CBF3 LE	ATGGAGGTGGAGGAGGC	TTAGTAGCTCCAGAGCGGC	Gene
168	192	CBF3	CGGTTTATGAAAACCGGTCGGTATTAG	GCCGCGTCTCCCTGAAC	Promoter
169	193	CBF3 LE	ATGGACATGGCCGGCCACGAG	TCAGTAGCTCCAGAGCGCGATGTC	Gene
170	194	DREB1A	CAACACAAATCAATCTCGCGATCCAC	CCTCCCGGCATTGCCC	Promoter
171	195	DREB1A LE	ATGTGCGGGATCAAGCAGGAG	CTAGTAGCTCCAGAGTGGGACGT	Gene
172	196	Submergence induced protein	GTTCTCGTTGGTACCATGGGATC	GAGTTGCTTTCCTATTGGGAAACCC	Promoter
173	197	Cinamyl CoA	GGCTGAATTTACCTTGGGTTGAAG	TCACCATTGTTGTCGTTGGAAATGG	Promoter
174	198	CDPK	GCCACCGAATGGTAAGCAGG	TGTGCGGTTGATGAGTAGATAGGCTAG	Promoter
175	199	SAM decarboxylase	GTCGACGCATCCAGCGAAC	CGGTTTCTCTCGACCACGAGAC	Promoter
176	200	WRKY Transcription factor	CCCAGAACCTCACCCCTC	CGGCCATGGTTGGTTCAC	Promoter
177	201	26S Proteosomal protein	GTAGAACTCGACGATACCGACG	CCATCCATCTTCCAAGTTTAGTTAGC	Promoter
178	202	Wound induced protein	CCTTGAACTTTCGTAACCATACGCTC	GCACTACATGAGAACAACATGACACG	Promoter

179	203	Signal recognition particle	GAGTTCAAACTTAGATGCGCCGTC	GAGGAGGAAGGGACAGAAGAGG	Promoter
180	204	RAN	CCTTTTCTCATAGATTTTTATTCGCGC	ATTGGGGGGAGATCGAAGG	Promoter
181	205	Cytochrome P450	GCAAGATAATGTTGGATGAAAAGTGGAGTAG	CACAGGACCACAGGTGATC	Promoter
182	206	Serine threonine kinase	GGCGCTAGTGGTGGTGTG	CTCCCGCAGATCAATCCTCC	Promoter
183	207	Enolase	GGAGGAGGTTGCAAAGAAACAG	GGATCGGATGGATCTTAGATCGGG	Promoter
184	208	R2R3 Myb	GCTTATTTGTAGTCGATTGAGGTTTGCC	CATGTGTGGGTGTTGCATGATAAG	Promoter
185	209	R2R3 Myb LE	ATGTTGGCCTTGGTGGTTTCAC	TTAAGCTAAGAGCAGCAATTGAATTAAGGCG	Gene
186	210	MYB1	GAATCTTAACTACTCTCTCCGTTTCACAATG	CGCGAGCTCGCCGGTATATATG	Promoter
187	211	RAS GTP binding protein	GAGTACTCCCCGTACTCACCC	CCCGACTGGGTCCTAGGC	Promoter
188	212	RACD GTP binding protein	GTGTATTTCTACATGGACTGTAAACTCTAC	CCTATTCGATGAACATGGTGTGATG	Promoter
189	213	RAB2 GTP binding protein	AAGGAACTACTCCGAGTAGAACAAATCAC	GTGCCTCAAGATCTGCACGAG	Promoter
190	214	RIC1 GTP binding protein	GTTAAAGTCTCTTTATGTGGGTCC	CATGAATAGCACTTTATGTGTGAC	Promoter
191	215	Signal recognition particle	GGTGGTGGTGTGTGGTG	GTCTTGAATGTTGTGTAGAAACCATGTCTC	Promoter
192	216	Transcription regulating protein	CTCGTGTGGACAGTTCATATTTCCTG	AACATTTAGCACTTAGCAGTAGTCTCC	Promoter
193	217	Glycine rich protein	TGTGCGAGTCAACTTTTGATCTGATCG	CTGTTATGGCATTTTCCATTTCCTTTTAGCG	Promoter
194	218	Proline rich protein	GGTCTGAACACAGCAGCCG	GCCTCGCCATGGCTGG	Promoter
195	219	Glutathionein reductase	CTGATTAGACGCCGTCCCATTG	GAGGGAGGGAAACCTGTAGAC	Promoter
196	220	Dehydrin	GGGAAGAAGAGTTTAAGGTCAATATAGG	GCGCCATTGTTGAAGGC	Promoter
197	221	Dehydrin	ATGGCGCATAAGCTCCTC	TCAGCTCTCCTCCTCGG	Gene
198	222	Cytokinin oxidase E1	CTTCCATCGTCAGCACACAAACTACAC	TAGATGTGATGGATCTACGTACCGGTGATG	Gene
199	223	Cytokinin oxidase E3	AGGATGGATGTGCTGCGTC	TCATGCGAGTGGTGACGTGAAAATC	Gene
200	224	Cytokinin oxidase	GGCTGAACATACCTAGCGTTCTATGC	GACGAAGCAGTTGAGCATGAGG	Promoter
201	225	M_ Unknown gene homologous to CB964527	CTCATATGCAGTGTTAGTCAATTATGAACTG	CTGATTCCAGTGAAGAAATCACACC	Promoter
202	226	M_ Unknown gene	CCTTTCAGGTTTCCCACCTTGTG	AGGGTTTGGGAATTTTTGGAGTTGATGG	Promoter

		homologous to CB966078			
203	227	Cystein protease	GGAGTGTACAAGAACAATTTAGGCCC	AGGAGCACAAGAGCCAGG	Promoter
204	228	Cystein protease LE	ATGGCGGCTTCAATGAATTCGAAG	TTAAGCGTTCTTGGTTGGGTAGGAG	Gene
205	229	Disease induced protein	CTCTACAGGGCTGGAGTGGC	CTAGCTACAAGGCAGGTGTGAGAAC	Promoter
206	230	Disease induced protein E1	ATGGAGGTTGAGGCTTCCTACAG	GAAGATGCATATACATCATGAGCAACTGG	Gene
207	231	Disease induced protein LE	GCAATCTTCTAGAGACTCCCAAGTCC	TTAAAGAATTGTCGACATGATTATCTCCTTGAC	Gene
208	232	DREB2A LE	ATGGCAGTTTATGATCAGAGTGGAG	TTAGTTCTCCAGATCCAAGTAACTCAAGTC	Gene
209	233	Disease induced protein 19	GACTTCGGATATGGAGGTAGTTCC	GGGGAGTTAGTCGTTGGGAATG	Promoter
210	234	Low temperature induced protein (LT16)	CCGGAGCATACAATGTACCCTG	GCTTGCATTATGGCGAGGTTC	Promoter
211	235	Water stress induced gene 18 LE	ATGGCTTCTCAGCAGGAAC	TTACTGGTAGTCACTCCCAG	Gene
212	236	Water stress induced gene 18	CGACATGTACCAGTACCATGAATCG	CTCACCAACACACGAACTGAAACTAC	Promoter
213	237	AP2	GCTTCTGTGGCTACTATTGTTATCAGC	CGCTAGAGCAACAACAAAATGCTG	Promoter
214	238	AP2 LE	ATGGACGCGGTGGACAG	TTAGTCCTTCCATGCGAAGGTTG	Gene
215	239	ECP40	ATCCGAGTTTGGACGGGTG	CCATAAGTTCGGTCTGCAATAACACATC	Promoter
216	240	Late embryogenesis abundant protein (LEA) LE	ATGGCTTCCCACCAGGAC	CTAGTGATCCCTCGCCGTC	Gene
217	241	LEA protein	GGCCGCTCAATGTGTCATC	CAGTATTTATAAGGGTTGGCATCGAGAC	Promoter
218	242	Cytokinin oxidase	CTGCTTATTTATAGGCCACCTTGTCCC	GGCGTTGGACAGCGTGC	Gene
219	243	DREB1D	CGGGATCCGCGGTCAGAGGATATGGTTCCGTTC	CGGAATTCGACGTGTGCGACGTCT	Promoter
220	244	Aldehyde dehydrogenase	ATGGGGAGCTTCGCGAG	CTAGCCAAAATTTATTCCTTGTGCTAGAGG	Gene
221	245	Aldehyde dehydrogenase	GTGTTTGTAGCGTAATTGCGTAAAGTCC	AAGCTCGGCGAGGAACTG	Promoter
222	246	Drought responsive protein homologous to Os.57533	AGGCACACGTTATATGGTCATCTTCC	GAACGTGAGGGAGTAATTGCAAAGC	Promoter
223	247	Drought induced LEA	GCAACCGAGGCGTACAGAAC	CGTCAATGCAAGCAATAGGGC	Gene

		similar to Os.57520_PG LEA			
224	248	DREB1A similar to Os.57517	CTATCGAAGAATTGACGGTGCC	TGGTGGTCGTTACTCGGTG	Promoter
225	249	DREB1A similar to Os.57517	ATGGAGGTGGAGGAGGCG	TTAGTAGCTCCAGAGCGGCATG	Gene
226	250	EF hand calcium-binding protein similar to Os.75123 and Os.69287	GAACCCTCAAAGCCTCC	GAGATGGAAGAGAAGGGAAGG	Promoter
227	251	EF hand calcium-binding protein similar to Os.75123 and Os.69287 GCD1	CCCCTTCATTTCCGACGGC	TTTGGTATCCCGTGAATCGATTGCG	Gene
228	252	EF hand calcium-binding protein similar to Os.75123 and Os.69287 GCD2	GGCAGGTTATTTCACAGGATTGATC	GGAAGGCGACGAAGCATGTC	Gene
229	253	Early drought induced protein similar to Os.13968	ATCCATACGTTTCTATCGAGTTGGTTGG	GAGAAGACATGGCAGAGTGCC	Promoter
230	254	Early drought induced protein similar to Os.13968 LE	ATGGCGGGAACGGCGAAC	CTAGATGACAAGAATTGGAGCACTCAGG	Gene
231	255	Drought-induced hydrophobic protein similar to Os.46402	TGACTTCCCCAGGGCATC	CTACTTGGTGATGACCCAGACAG	Gene
232	256	Drought-induced hydrophobic protein similar to Os.46402	GTTGCTACAGCTTCATGTAGCC	TTAGCAATGTTCTTGTGGATCTAGAGG	Promoter
233	257	Leucine rich protein similar to Os.46314	CCACAGTCGTCGCTGCC	GCAACCAAATCTGTGGAGCTCTG	Promoter
234	258	Drought-inducible protein1 similar to Os.22276	GTTCATACCACATTCCTCATGAGTGTG	CAACGGCTGTTGGTCAATCG	Promoter
235	259	Drought-induced protein DI-1 similar to Os.12212	CACGAGCAAGACTGGTCACAAAG	AGAGAGGAGTTGGAAAAGCCTTTG	Promoter
236	260	RNase S-like protein similar to Os.9417	GGAGACGATGATGGCAGGC	CCCAACCATTGTGAGGGTATTTATAGG	Promoter
237	261	Early drought induced	CAGATGACCTTTTGTAAAACTTATAAGCAA	CCATTGATCTAATAGTGCTTTTTTGGGTAC	Promoter

		protein R1G1A similar to Os.7840	CC		
238	262	Ubiquitinyl hydrolase similar to BU673177	GCTGACTATAGATTTGTAGCCCGC	GGTTTAGGGATGTTATACCAAGTCGAC	Promoter
239	263	Ubiquitinyl hydrolase similar to BU673177	ATGTCGGTGCCTCTGCTG	CTAAGACGCCTGCGCCTG	Gene
240	264	Leucine rich protein similar to CB967478	GAGAGATTGCGGCGACGG	GCATGACCTCTTCTAAGACAAGGG	Promoter
241	265	CDK activating kinase similar to CB965760	CAGCAAACATATCAGTTTCCCCCATTAC	TTCGAGCCGGAGACTGGC	Promoter
242	266	Tripeptidyl peptidase-II similar to BU672860	GCGTTAGTAGTAGTTGTTTCTTATATTTGGG	GACCGGTTTATGTAGCGCTG	Promoter
243	267	Peptidyl-prolyl cis-trans isomerise similar to BU673613	GACGATCAAGCAAGAGAGATTACATGAGAG	GATTGGACTCTCCTCGGCG	Promoter
244	268	Protein kinase similar to BU673471	GTTGCGGGTGAAATTTGTGTAGTCG	CTGGATGGGACTGTTAGGGG	Promoter
245	269	Proton pump interactor 1 similar to CB965018	GCCTTTGTATCCCTCTCCGTG	CGAGGCACCATACAGCTCC	Promoter
246	270	Tubby like protein similar to CB964631	CTTTCGCCTCCTCTCGCTTC	TCAAATCCTCATCGTGCTCTCGTG	Promoter
247	271	Cyclophilin	GGAATTCCGTTTAGGACAGCGACACGG	CGGGATCCCGATCGCTGCGTGGCCTG	Promoter
248	272	Ramosa-2	CTCGCAGGCAGCTGGAAC	GAGCACATGTGTGCGACTCTC	Promoter
249	273	Ramose-2	ATGGCATCCTCGTCGAGCAC	TCACATGCTGCTGTCTCCTCCTTC	Gene
250	274	Fear like gene	GTCGACGGGCTTGAGGC	GGTGGTGGCTTCCGTGG	Promoter
251	275	Serine threonine kinase	CACCTCAAAGACTGAAGAGCGTG	TCTACCTCCCGAACTGGGC	Promoter
252	276	Lectin like kinase	GTCTTGTATTACCCTCGTCCAGG	TTACCTTCCGAACCATTGTCACC	Promoter
253	277	ICE-1 like transcription factor	CATGGCAGCACAACACTGC	TTGCTCGCCTCCTTGCCTC	Promoter

Table 3.3: Primer details of the target region considered from comparative expression profiling, microarray gene expression profiling under field conditions and known responsive genes for drought stress.

S.No	Primer No.	Function associated with the target locus	Forward Primer Sequence	Reverse Primer Sequence	Type of the region
254	278	NDC1_Receptor like kinase	GATCCCGACCCCTAGTCATAG	CTTCGTGTGAAAATTCCCTCTGTG	Promoter
255	279	NDC3_Phosphoglycerate kinase	AGCAAGCCCCACACGTC	GATCACAATTGATTCATATGCAACGAG	Promoter
256	280	NDC5_Sodium/Hydrogen Exchanger protein	GACTATAGATTTGTACTTCCTATGTCCC	ATCCTCAATGGGTGCACAG	Promoter
257	281	NDC7_Polyubiquitin	GTGTGGTTGGACCGCAAG	GGGGTGGGGTATTTATAGGTGG	Promoter
258	282	NDC10_Squamosa like protein	GTAGTAAAATAGACGGTCGGTCC	GTGGTGTTGGTTCTGTCATC	Promoter
259	283	NDC13_Sac domain-containing inositol hosphatise 3	CCTGCACTATTCAGTTCCCC	GGTTGATGACTCGGTTTTGG	Promoter
260	284	NDC15_γ-aminobutyric acid (GABA)	GACACGGTGCGCTAATTTTG	CCACTGATTCACTGATCTGCC	Promoter
261	285	NDC16_unknown membrane protein	CCAGCACCAGCAAACAATG	GCTCTCTCGAAGTCCAAGTCAAAC	Promoter
262	286	NDC25_DNA binding protein	ACGATGATGCCGCTGGG	CAAGAGGGTAGTAGTCATTTCTATGAG	Promoter
263	287	NDC26_Hypothetical protein	TCTCGCGTCAGACTGGC	GTGATTTTCTGAGGTTTCCTGCG	Promoter
264	288	NDC28_nifU	CGTTTGCAAAACACTGTTGCTGTG	CGAGTCATCATGGGCCACATTAG	Promoter
265	289	NDC29_UDP-glucose pyrophosphorylase 2	AAGCGAGGCGAGGATGC	GATGAGAGGAGAGGAGTGCG	Promoter
266	290	NDC30_Avr9/Cf-9 rapidly elicited protein	GGGGGAAGATGACGTCAAG	GTTACTATCTCCATCCTAAAATGTAAGC	Promoter
267	291	NDC32_Ethylene-upregulated protein ER1	CACGGAAACCTGGGTTTGTC	GTCTTGCAGTTTACAGGCAGAATC	Promoter
268	292	NDC33_GYF like protein	CGTAAGTTTCTTGTGTATGCAGGC	GTGTAACTTTTTATATCGGTCCATCTCC	Promoter
269	293	NDC34_Helicase	GAAGAGGTAAAGCAGGCTACTACT	GCATGAATTATATGTTTGGATGCTTC	Promoter
270	294	NDC39_dnaJ	GACATGAACCCCACCTCACG	CACATGGACCCATATGACACACG	Promoter

271	295	NDC41_Transposan like protein	GCCTGACCAAGACGCATG	CATCGGTTAGCTCTGTTTTCACCATC	Promoter
272	296	NDC42_Glycosyltransferase	GGCGTAGTAAAAGATGATGGCAC	CAACTGCGTGCGTGCTC	Promoter
273	297	NDC46_MAP kinase	GCTGGCTACTTTTTTCCAAGTG	CGCAAGATATGTATGGTTTGTGTTGC	Promoter
274	298	NDC47_Hevein like protein	ATATCCTACCGGGACCCTCC	GTGTGAGAGCCGTGCATTGTG	Promoter

Table 3.4: Primer details of the target regions considered from subtractive cDNA library of Nerica-1 and IR64 developed from field drought stress experiment.

S.No	Primer No.	Function associated with the target locus	Forward Primer Sequence	Reverse Primer Sequence	Type of the region
275	301	MIR159a	GCTAGCACAGCAACTCATCTG	ACAGTAAAGCTGCTCGAGGG	MIR
276	302	MIR159b	GCCTCTGATGTTTGCTTTTGC	GATCTACACATGGGGAACC	MIR
277	303	MIR159c	CACAGTACAACGTAGACACTCAC	CAATTAATCCACTGACACTGCCA	MIR Upstream
278	304	MIR159c	AGAAAAGGCAAGTTCTCTGTCCA	TCAATCTGCCTCCGGTCC	MIR
279	305	MIR159d	ATCAATCTGTCTCCGGTTCC	TAAACCCACCAACCCCTAAG	MIR
280	306	MIR159e	TTGTATACTCACTCCTGTGGTG	CCCCTAAGCAATCGTCGTAC	MIR
281	309	MIR172a	ATAGTATGACAGTGCTCTCTCGTC	ACTCTCTCCGTTTCACCGTTTCAC	MIR
282	311	MIR172d	AACAAAGGGGATTTCTGTAGCTG	TAGCTAAGTCATAACCAGTCCTC	MIR
283	312	MIR172c	GTGTGTACGCGCCGCAAC	TATACTGTACCTGGCCAAAAAGC	MIR
284	313	MIR172d	TTTGACTAGCTCCCACCATG	ATCGTTTAGTCTCTGCTACTGG	MIR
285	314	MIR319a	TACATCAGTCTTGCATACCAG	CTTCGTGATCACTGCAGCTGGG	MIR Upstream
286	316	MIR319b	GATGCTGCATATCATCATGCA	ACAAATCCTCCAAGATCCG	MIR
287	317	MIR398a	ATCGGGGGTCACCGAAATC	AACCTATGTCCACTACAAGCTG	MIR
288	318	MIR398b	CAGAGCGCATTTTTTCCGGG	CATCCTGTATTTCATGCGGACC	MIR Upstream
289	319	MIR398b	GCCATTTAGGAGACCTTTTGGC	AAGTACGACGCCAATCGATGC	MIR

Table 3.5: Primer details of the target regions from microRNA regions associated with abiotic stress response

3.1.7 Primers used for validating the SNPs and in development of AS-PCR markers

The allele specific primer was designed to develop a PCR based assay for validation of SNP and to carryout genotyping in double haploid population. This primer was used as a third primer in the AS-PCR reaction along with the set of primers designed for the target region. In order to increase the specificity at the target allele the 3' base of the oligo was modified to contain and LNA instead of DNA where in a special 2'-O,4' –C methylene bridge is created which locks ribose group into C3' –endo confirmation. This kind of modification is proven to increase the specificity of the allele specific primer and exploited in developing TrueSNP platform by Proligo, Singapore. The designing of these oligos was performed as per the required allele at the target SNP site and the modified oligos were obtained from Proligo, Singapore. The allele specific base was designed either for CT9993 allele or IR62266 allele as to facilitate genotyping in DHL population. Target region primers were used as positive control and allele specific primer was used for producing allele specific product and size was predicted based on the target region and listed in the Table 3.6.

Primer No.	Allele Specific Primer Name	Allele Specific Primer Sequence	Allele in CT9993	Allele in IR62266	Size of the Positive control	AS PCR Product size
37	AS_37_G	CACTCTCCTAGGATACGTG	G	C	524	311
37	AS_37_C	CACTCTCCTAGGATACGTC	G	C	524	311
47	AS_47_c	CCTCCCTCCAAATGATCATC	T	C	670	597
48	AS_48_c	AGGAACTATGGTGCTCCAC	-	C	935	843
49	AS_49_AC	GCTGTAGTTGTAAAATCCTAAAAATGAAC	AC	--	741	317
49	AS_49-2_TC	ATTTTCAGAAGCACTACCAGTTC	TC	--	741	253
51	AS_51_C	GTTGTGTGGTTGTGTGTGCC	C	A	344	201
3	ASLNA_3_A	GAACCACCTAGATCACGACGCCTGA	A	C	499	305
4	ASLNA_4_T	CTTATTGCTGTGCTACAAAAAGTAAACTT	T	A	518	348
5	ASLNA_5_C	GCACATGTATATACATACCTTCAGACGGC	A	C	956	768
7	ASLNA_7_C	GTCATGTCGCCGTCCAGC	A	C	493	212
40	ASLNA_40_C	GAAACCGAACTTTTTTTTTTCCC	C	-	544	368
56	ASLNA_56_T	AGGTTTGTTTTGTTTTGTTAAAAAAAAAT	-T	AT	512	349
61	ASLNA_61_C	TTTTAATAAGATGAACGGTCAAAC	C	T	454	239
65	ASLNA_65-1_C	GATATGGTTAGTTAAAGTAGCATCTC	T	C	510	249
65	ASLNA_65-2_G	CAATCATTTTTAACATTAATTATTAATAG	GA	--A	510	387
67	ASLNA_67_ctC	GCCTACGCTACTATTAACCTTACTC	CTC	---	442	363
70	ASLNA_70_C	GGAAGCTAGTTAAATGCTATGTGATTAC	T	C	771	282
71	ASLNA_71_C	GGTTGATAAACCAATTATTGATCTGC	T	C	581	285
73	ASLNA_73_G	AATGGTAACTGGTGTGTATTG	T	G	579	415
75	ASLNA_74_A	TGAATCTGGATAAGCTGGAAA	-	A	537	140
76	ASLNA_76_C	GATTTTTCAATCCACCTTAATAATAATTAC	C	T	519	371
78	ASLNA_78_c	TCTGCAATATAATACTAAAGAATTGACC	T	C	503	169
79	ASLNA_79_C	AGTTCCTGTAGTCTCTAGC	C	G	468	227
81	ASLNA_81_c	GTACTTTGTACAAGTGAAGCAATATC	T	C	509	294
91	ASLNA_91_c	AACGAATGAATAATATACTATATAATGCGC	T	C	507	309
95	ASLNA_95_G	GACCCCTTTTTTTTGCAAAAGAG	G	A	595	353

98	ASLNA_98_gCg	CTCCTAAAAGCCACGCG	ACA	GCG	806	195
109	ASLNA_109_A	GGTCATGATAGGGTTAATCTATGTGTTA	A	T	349	247
111	ASLNA_111_C	TCAGCGCAGAGAAGCTCAC	C	T	829	272
120	ASLNA_120_g	GAAACAAGAATTTTAACACTGCGCG	A	G	1094	195
147	ASLNA_147_G	CGCCGTCGGGTGGAGG	G	C	427	283
155	ASLNA_155_G	TCATGGGGTTTTGCCTGG	G	C	784	189
114	ASLNA_114_A	ACGACATATTTACGAACACAAAATAATTTGTA	A	G	830	539
116	ASLNA_116_ccgt	CCAGTGTTCTATGGGCCGT	CCGT	----	723	242
118	ASLNA_118_TAC	GTGATTCATGTAAGCTGAGACTTAC	TAC	ATA	1093	701
121	ASLNA_121_c	CATACCTACTACACAATAGTCAGTTCATATAAATC	T	C	676	467
123	ASLNA_123_g	CTGTCTCTGCCACTTCGG	A	G	874	517
129	ASLNA_129_C	TTCAATGTCATAAACTTGATAGCGTATAAAATAC	C	-	667	129
135	ASLNA_135_aggat	TTTCGCCTCCGTGAGGAT	-----	AGGAT	755	363
142	ASLNA_142_G	AAACCGAGTGCGTACGTG	G	T	475	209
144	ASLNA_144_G	AACATGGAGTCAGAAGGACATGGTAATTGG	G	A	929	671
146	ASLNA_146_c	CTCTTAAGGGTCCGCACAC	T	C	846	646
148	ASLNA_148_A	AGATAAATGATTAAAAAGGGAGAGAGTTGA	A	G	655	381
152	ASLNA_152_G	ACCAATGCGTTGGTAGAATTG	G	T	723	156
156	ASLNA_156_T	GTTGGGGGGCGACGTGT	T	G	740	504
158	ASLNA_158_C	GCACATATTTCATAACTCAGCTCAC	C	T	444	281
83	ASLNA_83_A	AGTCCATTAACGCGTAATTAATA	A	T	506	255
86	ASLNA_86_c	TTTTTCACCGATATATGTCTGTATC	A	C	533	184

Table 3.6: List of allele specific primers used for validating and genotyping of the candidate SNPs identified in the target region

3.1.8 Primers used for validating SNPs using Single Base Extension (SBE) technology

The validation of candidate SNPs was performed with the help of SBE primer designed for the SNP region up to the SNP site as such to capture the very first base during the extension and terminated using ddNTPs. The set of primers designed as SBE were synthesized from the commercial oligo synthesizers with high purity. The SBE primers for the corresponding target region and SNP were listed Table 3.7 along with their dbSNP accession numbers deposited from this study. Deletion were denoted with a '-' and multinucleotide polymorphisms were indicated as a single stretch. The complete details of each targeted candidate SNPs can be found in dbSNP division SNP under the handle ARR-VBREDDY with the corresponding SNP number.

S.No	M.Primer No	SBE Primer Name	SBE PRIMER	SNP Type	dbSNP ID
1	147	SB_147	CGCCGTCGGGTGGAG	G/C	69373604
2	142	SB_142	AAACCGAGTGCGTACGT	G/T	69374498
3	156	SB_156	GTTGGGGGGCGACGTG	T/G	69374109
4	169	SB_169	GCTTTGTTCTCCAAGATTCCCAGCA	T/C	69374450
5	171	SB_171_WRKY	GGTTTAATAGGATCAATTGGATCCATG	T/C	69373279
6	180	SB_180-2	GGCTACAATCTGAACATGCC	T/C	69374538
7	183	SB_183_CURE	GCCACTTTGATATGTAATGTAAGCACTGTA	T/C	69374043
8	183	SB_183_SiteII	CGAGAGTTAAGCCATGGCC	C/G	69374041
9	183	SB_183_GATA	GATCTATGGATGGATTCGTGAGTCGAT	T/C	69374039
10	192	SB_192	GGGAAGCCATGGGAGAG	G/-	69374573
11	199	SB_199-2	CCACTATAAATACACGGCGAAGGG	A/G	69373547
12	224	SB_224-2	CCTAGGGTGGCTATACTAACCGT	C/T	69373308
13	78	SB_78	TCTGCAATATAATACTAAAGAATTGAC	T/C	69373327
14	155	SB_155	TCATGGGGTTTTGCCTG	G/C	69374115
15	118	SB_118	GTGATTCATGTAAGCTGAGACT	T/A	69373760
16	123	SB_123	CTGTCTCTGCCACTTCG	A/G	69373966
17	159	SB_159WBOX	GGAAAGTTGCTTTAAAAAATCATATTAGTC	A/C	69374561
18	178	SB_178TBOX	CAATGACATGAGAACCCAGACAAAG	A/G	69373796
19	179	SB_179	TACGAGACGAATCTTTTAAGCCTAATTA	C/T	69373675
20	183	SB_183CGBOX	CTCTCTCCGTCTCTAAAATTTAACGC	A/G	69374044
21	196	SB_196	GGCGAGGTCGCTTTCATTAGAAA	A/C	69373618
22	202	SB_202	GTAAAAGCCCCATCCCTCGATC	T/C	69374425
23	206	SB_206	CGTAAAATCTCGCCCCACATTTGGA	A/T	69373608
24	224	SB_224	TAATTTATTTTTGGATGTGTGGTCTAG	C/T	NS
25	47	SB47_AG	CTGTTCTTCGCTATCAAAGATGG	A/G	NS
26	61	SB61_GT	AATTGCCAAATTTATACTCCCTCC	T/G	69374696
27	65	SB65_AG	GATATGGTTAGTTAAAGTAGCATCT	T/C	NS
28	70	SB70_TC	GGAAGCTAGTTAAATGCTATGTGATTA	T/C	NS
29	73	SB73_TG	GAATGAATGGTAACTGGTGTGTATT	T/G	NS

30	74	SB74_TA	CGTGAATCTGGATAAGCTGGAA	A/-	NS
31	79	SB79_CG	CTGAGTTCCTGTAGTCTCTAG	G/C	69373324
32	98	SB98_AG	GGAGTCCTCCTAAAAGCCAC	A/G	69374167
33	135	SB135_GA	GTTGGTCAAAGATGCACGCAATA	G/A	69374676
34	152	SB152_GT	ATGACCAATGCGTTGGTAGAATT	T/G	NS
35	159	SB159Myb1_CA	AGCCCCAAAAACCCAACAGAAA	GGTTTCTGT/-	69374552
36	180	SB180_CA	CTGAGTGGATTAAAGTGGTTTAGAAT	-/AT	69374536
37	183	SB183CARGCW8_CT	TTGCAACGGGTGAATGCTATTTTAAT	G/A	69374036
38	199	SB199_GA	CCTGGAAGGCAAGAGAAGAAAAGAA	-/AAA	69373542
39	205	SB205_AG	TCTTGTGTAATCTTGGAACGTGTC	T/C	69373351
40	212	SB212_AT	AAATCGTATTCCTTTATGGACTCT	-/AGTCCCCC	69373586
41	192	SB192_GC	ACTATGGTTTCCTTGGGCAT	G/-	69374573
42	206	SB206_TA	TCTCGCCCCACATTTGGA	A/T	69373608
43	202	SB202_AG	TAGCTATGCCACTGTAGGATAC	T/C	69374418
44	155	SB155_CT	ACGCCATGGTTGAATGCCA	G/A	69374116
45	78	SB78_AC	CAGTTGTTGTTCTCCAACGC	T/G	69373328
46	227	SB227_GT	TCTAGATGATACTCCGGAAGAGGA	G/T	NS
47	162	SB162PR1_GC	CAACACGCAACAAATCTCCTCT	C/G	69374243
48	173	SB173Myb1_GT	TACAAACGAATCTAGACAGATG	C/A	69374155
49	81	SB81_AT	GATAACCATCTCTTGGGGACAATC	A/T	69374673
50	116	SB116_CA	GCATAGCCGTACCCGATTCC	G/C	69374400
51	86	SB86_TG	GGAATGAAATTAGGGGAATGATGAATTA	G/T	69373940
52	162	SB162_GC	AACACGCAACAAATCTCCTCT	C/G	69374243
53	173	SB173_GT	TTACAAACGAATCTAGACAGATG	C/A	69374155
54	155	SB155-2_CT	CACGCCATGGTTGAATGCC	G/A	69374116
55	267	SB267_AG	CACCACCAGAAAGTCGTGTAATA	T/C	69373572
56	269	SB269_AC	ATTGGCACGCGAGGCCAT	T/G	69374434

Table 3.7: Set of SBE Primers designed for validating Candidate SNPs

3.1.9 Assembled BAC clone sequences and rice genomic sequence data

Rice genome sequencing data from IRGSP was extensively utilized in anchoring the target regions to BAC/PAC clone map and chromosome maps of the rice genome. This data has been obtained from NCBI, Gramene and IRGSP. The sequence information of BAC clones was stored in MySQL database and utilized in the analysis as required utilizing the custom Perl scripts written for the purpose. The information of the BAC clones of each chromosome and their chromosome mapping, physical and genetic distances were summarized in Table 3.8. This integrated data of physical and genetic maps served as an anchor in SNP analysis and also as a standard reference in identifying and denoting SNPs.

Chromosome	**No. of BACs**	**Size of the chromosome (bp)**	**Size of the chromosome (cM)**
Chromosome 1	394	43261740	181.8
Chromosome 2	383	35949750	157.9
Chromosome 3	323	36192798	166.4
Chromosome 4	263	35498469	129.6
Chromosome 5	270	29663311	122.3
Chromosome 6	293	30727520	124.4
Chromosome 7	297	29636716	118.6
Chromosome 8	276	28424490	121.2
Chromosome 9	223	22710325	93.5
Chromosome 10	191	22685330	83.8
Chromosome 11	230	28389314	117.9
Chromosome 12	242	27533308	109.2

Table 3.8: Details of the rice genome sequence information; Mapping of sequence map with genetic map reveals the correlation of sequence data with that of the genetic data

3.2. Methods

3.2.1. Subtracted cDNA library construction

3.2.1.1. Experimental Setup

For field drought stress induction, plots were prepared by digging the soil up to 2.5 ft, and blocks are built with masonry, to avoid seepage of water into the plots and filled with black soil and FYM. These blocks are raised enough to avoid accidental overflow from the control plots. Rainout shelter was constructed to cover the plots during rainy days to keep the plants out from rainwater during experiment period. The experiment has been conducted during summer (Feb-Mar 2005) in order to impose severe drought stress. Genotypes selected for subtractive cDNA library construction are germinated in laboratory condition up to 15 days and transplanted into field. Irrigation was given to full filed capacity up to 42 days and drought stress has been initiated there after. Control plots were covered with shade net to avoid heat stress in control plants. Rainout shelter was used to avoid rainfall during stress conditions (Fig. 3.1). Relative water content was measured on every alternate day to measure the degree of drought stress. Leaf samples from both stressed and control plots was done when RWC reached to ~60%.

3.2.1.2 Total RNA isolation

Total RNA was isolated from leaves of both stressed and control of all the genotypes using Trizol reagent (Life Technologies, USA). Approximately, 1 gm leaf tissue was disrupted and homogenized using sterile, DEPC treated porcelain mortar and pestle in liquid nitrogen while maintaining RNase free conditions during the entire process. RNaseZap (Ambion, USA) was used to maintain RNAse free conditions during the process and isolated RNA was dissolved in DEPC treated ddH_2O. RNA was quantified by checking absorbance at 260 nm and integrity verified on formaldehyde agarose gel.

Figure 3.1: A view of drought stress induction experiment under rainout shelter. Left side plots are induced stress and right side are controls of the same genotypes. Rainout shelter is in parked position.

3.2.1.3 mRNA capturing and cDNA synthesis from driver

mRNA was isolated from the driver sample was isolated using Promega PolyATract mRNA isolation kit taking ~700 µg total RNA. Captured mRNA was used in synthesis of single strand cDNA using 3 µl Biotin labeled OligodT Primer (50 pm/µl), 6 µl 5X RT buffer, 1.5 µl 0.1 M DTT, 3 µl dNTPs (5 mM), 1 µl RNase (40 U/µl), 2 µl Superscript–III (400 U/µl). Reaction was carried out at 46°C for 2 hrs followed by alkaline hydrolysis incubating at 70°C for 10 min adding 30 µl NaOH (1 N) and neutralized with 30 µl HCl (1 N) and 20 µl 3M Sodium Acetate, pH 5.2. First strand cDNA synthesized in above reaction was purified using S.N.A.P purification columns in order to remove unincorporated dNTPs and hydrolyzed mRNA following manufacturer's protocol and precipitated in 10 µl Sodium acetate (3 M, pH 5.2), 2 µl of glycogen (20 mg/ml), 300 µl ice cold 100% ethanol and placed at -80°C for overnight.

3.2.1.4 mRNA capturing from tester and subtracting with driver cDNA

Single strand cDNA synthesized from driver sample was dissolved in DEPC treated water and used to hybridize the mRNA from tester sample incubating in 16.6 µl KCl (3 M), 1 µl Tris (1 M) for 1 hr at 65°C. ss-cDNA and mRNA hybrids are washed out and mRNA enriched in tester (stress specific) was captured using streptavidin paramagnetic beads. Stress specific mRNA precipitated in 10 µl Sodium Acetate (3 M, pH 5.2), 2 µl of glycogen (20 mg/ml), 300 µl ice cold 100% ethanol and placed at -80°C for overnight.

3.2.1.5 cDNA synthesis from stress specific mRNA

First strand cDNA was synthesized from stress specific mRNA using 2 µl (1.4 µg/µl) OligodT linker primer (Stratagene) with *Xho*I site and incubated at 65°C for 10 min. To this 5 µl of 10X first strand buffer, 3 µl of first strand dNTPs mix (10 mM dATP,

dGTP, and dTTP and 5 mM 5-methyl dCTP) and 1 µl RNase Block ribonuclease inhibitor (40 U/µl) were added and incubated at room temperature for 10 min. To this 1.5 µl Reverse transcriptase (Stratascript 50 U/µl) was added and incubated at 42°C for 60 min. Second strand synthesis was carried out using 20 µl of 10X second strand buffer, 6 µl of second strand dNTP mixture (10 mM dATP, dGTP, and dTTP plus 26 mM dCTP), 116 µl sterile water, 2 µl RNase H (1.5 U/µl) and 11 µl DNA Polymerase-I (9 U/µl). After gentle vortex and spinning reaction was incubated at 16°C for 2 hours 30 minutes in water bath and stored on ice and blunting of the cDNA termini was carried out as described in the next section.

3.2.1.6 cDNA library construction with stress specific cDNA

To the above synthesized ds-cDNA mix 23 µl of blunting dNTP mix (2.5 mM dATP, dGTP, dTTP, and dCTP) and 2 µl *Pfu* DNA polymerase (2.5 U/µl) was added and incubated at 72°C for 30 min. 200 µl phenol-chloroform (1:1 v/v, pH 7-8) was added and vortex and spun for 2 min at room temperature and upper aqueous layer was transferred to new tube and added equal volume of chloroform. This mixture was spun at RT for two minutes and precipitated in 3 M sodium acetate and 400 µl 100% alcohol for overnight at -80°C. Precipitated cDNA was dissolved in 9 µl of *Eco*RI adapter mix (0.4 µg/µl) and 1 µl 10X ligase buffer, 1 µl rATP and 1 µl T4 DNA ligase (4 U/µl) was added and incubated overnight at 8°C. Ligase was heat inactivated after completion of the reaction by heating the tube to 70°C for 30 min. Reaction contents were incubated at room temperature for 5 min and *Eco*RI end s were phosphorylated by adding 1 µl of 10X ligase buffer, 2 µl 10 mM rATP, 5 µl sterile water and 2 µl of T4 Polynucleotide kinase (5 U/µl). Reaction contents were incubated at 37°C for 30 min. Enzyme was heat inactivated by placing the tube at 70°C for 30 min and digested with *Xho*I by

adding 28 µl *Xho*I buffer, 3 µl of *Xho*I (40 U/µl) and incubated at 37°C for 90 min. After 90 min of incubation 5 µl of 10X STE buffer and 125 µl of 100% ethanol was added and precipitated at -80°C for overnight. Following precipitation cDNA was dissolved in 14 µl of 1X STE buffer and size fractionated using sephacryl 400 (G.E.Healthcare, Hong Kong). The collected fraction was purified by adding equal volume of phenol-chloroform (1:1 v/v, pH 7-8) and aqueous phase was collected and precipitated in 100% alcohol for overnight at -80°C. Precipitated pellet was washed with 80% alcohol and dissolved in 5 µl of sterile double distilled water. cDNA concentration was estimated with ethidium bromide plate assay and 100 ng cDNA was used to ligate Uni-ZAP XR vector with pBlueScript (SK-) backbone. 1 µl of (1 µg/µl) Uni-ZAP XR vector digested with *EcoR*I and *Xho*I, 100 ng resuspended cDNA , 0.5 µl of 10X ligase buffer, 0.5 µl of 10 mM rATP (pH 7.5) and 0.9 µl of water, 0.5 µl T4 DNA ligase (4 U/µl) was used to set up ligation reaction and incubated at 12°C for overnight. After ligation 1 µl of the ligated mixture was used for packaging with Gigapack III Gold packaging extract (Stratagene, USA) following manufacturer's protocol. Finally 500 µl of S.M buffer and 20 µl chloroform was added, centrifuged briefly and supernatant was collected.

3.2.1.7 Subtractive cDNA library screening, Purification of PCR positives and size separation

Collected packaged mixture was used to infect XL1 Blue MRF' strain of *E.coli*. Suitable dilution was estimate by calculating pfu (plaque forming units), by plating on XL1 Blue MRF' strain grown in NZY broth with 20% sucrose. Suitable dilution was calculated and used to amplify the library plating on NZY agar media with NZY top agar. Plaques were picked and dissolved in 300 µl of S.M buffer and used in PCR

screening. 3 µl of dissolved plaque, 2 µl of 10X PCR buffer (100 mM Tris-HCl pH 9.0, 500 mM KCl, 25 mM $MgCl_2$, 1% Triton X-100), 1 µl 5 mM dNTPs, 0.6 µl of M13 forward and reverse primers (5 pm/µl), 12.6 µl sterile water and 1 U of Taq DNA polymerase (Genotex International Pvt. Ltd, India) used to amplify the c-DNA from λ phage. The samples were denatured initially at 95°C for 3 min followed by 32 cycles of 1 min denaturation at 94°C, 1 min of primer annealing at 50°C and 1 min of synthesis at 72°C, with a final extension step of 72°C for 10 min. The programs ended at 4°C for cooling. The PCR products were analyzed on a 1% agarose gel and PCR positives were size separated using 200 bp DNA ladder (Genotex International (I) Pvt. Ltd).

3.2.1.8 cDNA Sequencing and identification of stress responsive genes

PCR products were purified to remove unincorporated dNTPs and excess primers remaining in the reaction by passing through the Wizard SV96 (Promega Corp, USA) columns using 100 µl of binding buffer, and washed with 300 µl of washing buffer and eluted into 40 µl MQ water. Agarose gel electrophoresis was performed on 1% agarose and 25-60 ng purified PCR was used for sequencing purpose from each genotype. The purified PCR products were used for DNA sequencing using one of the vector primers in the sequencing reaction using standard sequencing protocol for capillary electrophoresis. The differentially expressed genes in Nerica-1 were sequence characterized using dye terminator chemistry, on MegaBACE 500 sequencer using DYEnamic ET terminator kit. Treated PCR products were sequenced with Dye terminator reagent (GE Healthcare, Hong Kong), combining 2-3 µl of the purified PCR product diluted to (~20 ng/µl) with 0.5 µl of either of the PCR primer (5 pm/µl), 2 µl of Dye terminator reagent and 2 µl of halfTS-II (Genetix, UK) and 2.5 µl ddH2O. The reactions were cycled for 28 cycles at 95°C for 20 sec., 50°C for 15 sec., 60°C for 1

min. Reactions were ethanol precipitated and dissolved in MegaBACE loading solution. Prior to loading, the samples were re-suspended in 10 µl of loading solution (70% Formamide, 1 mM EDTA) by gentle vortexing. Reaction products were loaded and analyzed on the MegaBACE 500 using an injection of 2 kv for 45 seconds. Approximately 300-600 bases of data were collected for each sample. Each template was sequenced from the 3' end of the clone using the vector specific primer and contigs were generated in order to generate the unigenes. The sequence characterized clones were searched for the homology at NCBI and annotated using BLAST algorithm (Altschul *et al*., 1997).

3.2.2 SNP Genotyping

3.2.2.1. Selection of elite Genotypes

Rice cultivars that were used in different breeding programs across India were obtained from various sources (Table 3.1). The list represents genotypes which are phenotypically well characterized with reference to drought stress response, representing both indica and japonica cultivars. A total of 8 genotypes representing 3 highly drought tolerant, 2 moderately tolerant and 3 drought susceptible genotypes were used in SNP genotyping. A doubled-haploid line (DHL) mapping population of rice (CT9993-5-10-1-M/IR62266-42-6-2, designated as IR68586) developed at IRRI and adopted for mapping of QTLs associated with drought resistance of rice along with parents was obtained from IRRI was used in mapping.

3.2.2.2. Genomic DNA isolation using modified CTAB method

Genomic DNA was isolated from leaf tissue by the procedure described by Murray and Thompson, 1980. Freshly harvested leaf tissue (10 gm) was quick frozen in liquid nitrogen and 30 ml of pre-heated (to 90°C) 2X CTAB buffer (2% CTAB, 100 mM Tris-HCl pH 8.0, 20 mM EDTA pH 8.0, 1.4 M NaCl, 40 mM 2-mercaptoethanol), was

added. This mixture was incubated in a water bath at 65°C for 90 min with mild shaking and 30 ml of chloroform: isoamylalcohol (24:1) was added and further incubated for 20 min at room temperature (25°C -28°C) with mild shaking. The sample was centrifuged at 6000 rpm for 10 min at room temperature. The aqueous phase was separated and DNA was precipitated with 0.7 volumes of isopropanol at -20°C for 20 min. The DNA pellet was spooled out with a glass hook, washed with 5 ml of 76% aqueous ethanol, 0.2 M sodium acetate, followed by 5 ml of 70% aqueous ethanol. The DNA pellet was air dried for 20 min and dissolved in 1 ml of sterile TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). To the above, 1/100 volume of RNase A stock solution (10 mg/ml), and incubated at 37°C for 1 hr. DNA solution was purified with phenol: chloroform method. DNA solution was mixed with phenol: chloroform (1:1), and mixed gently. The sample was centrifuged at 5000 rpm for 10 min at room temperature. The aqueous phase was separated and mixed with equal volume of chloroform, mixed gently and centrifuge at 5000 rpm for 10 min at room temperature. The aqueous phase was separated and mixed with twice volumes of absolute ethanol and incubate at -20°C for 20 min. The DNA pellet was spooled out with a glass hook, wash with 70% aqueous ethanol. The DNA pellet was air dried for 20 min and dissolved in an appropriate volume of sterile TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) and stored at 4°C. The concentration of DNA was determined spectrophotometrically and the quality of DNA was checked through agarose gel electrophoresis.

3.2.2.3 Identification and selection of candidate genes and primer designing

Candidate genes for drought tolerance were identified from the gene resource developed earlier in our laboratory from drought stress induced seedlings of *Oryza*

sativa cv Nagina22 (Reddy *et al*., 2002 and Babu *et al*., 2002) and microarray gene expression profiles of published literature and unpublished microarray studies (Markandeya and Lachagari *et al*., Unpublished). Selected candidate genes were analyzed with available unigene data set from NCBI and genotype specific contigs were generated with a custom script developed called 'gtCluster'. As a result of high throughput EST sequencing projects of rice, huge amount of EST data being generated and deposited in public databases like NCBI, DDBJ resulting in redundancy of sequence information. This will mislead direct readout of sequence data for SNP identification. In order to overcome this problem gtCluster and a pipeline for automated analysis and identification of SNP has been developed as a part of my research work during this period. gtCluster is a Perl script, developed in order to have genotype specific clusters and their by to develop genotype specific contigs which can be used to identify SNPs *in silico*. gtCluster has been embedded in pipeline developed for SNP identification, analysis, and physical mapping further to develop HapMaps with selected candidate genes. This program is developed to run in UNIX platform using Redhat Linux 9.0 and supported by MySQL database at the backend. The identification of likely paralogous (or repetitive) sequences within clusters and SNP detection was performed by an adaptation of the PolyBayes SNP analysis approach. FastPCR and Consed were used to design the primers for the targeted regions and synthesized at MWG, Germany or Sigma Genosys, Bangalore with desalted grade.

3.2.2.4 Targeting putative promoter regions of the candidate genes

Rice genome sequence data was used to identify putative promoter region of targeted candidate genes. BAC/PAC clones sequence data of spanning targeted candidate gene region has been obtained by BLASTN program (Altschul *et al*., 1997) at NCBI and

gene structure was predicted using FGENESH program. Gene structure including transcription start site, introns, exons, poly adenylation site was predicted and stored in MySQL database and used in the SNP identification pipeline. This data was used to tag gene structure information and visualized in Consed during SNP identification analysis. Candidate gene promoters were targeted based on the tags generated in SNP pipeline approximately 500-900 bp region at 5' upstream to transcription start site and primers were designed with Consed and used in SNP genotyping. A new database has been developed called RiceSNPs to store the information generated from automated SNP identification pipeline. This database contains information of SNPs, BAC/PAC localization, type of SNPs, Flanking region of SNP site, probability of the SNP identified based on Bayesian statistics. Now this database is being updated to store the information generated from re-sequencing the targeted region from different genotypes which will be used in developing HapMap with candidate genes.

3.2.2.5. Amplification of targeted regions from genomic DNA of selected genotypes

Specific primer sequences designed targeting region of interest using either Consed or FastPCR software and oligos were synthesized. Targeted regions were amplified from 8 elite genotypes listed above (Table 3.1) with PCR. PCR was performed using a DNA engine (PTC 200, M J Research, Inc.) or GeneAmp (PCR system 2700, ABI) with 60 ng of plant genomic DNA in a 50 µl reaction containing 5 µl 10X PCR buffer, 2.5 mM of each dNTP, 5 µM of primers, 2.5 U of *Pfu* DNA polymerase. The samples were denatured initially at 95°C for 5 min followed by 32 cycles of 1min denaturation at 94°C, 1 min of primer annealing at 55°C -65°C and 1 min of synthesis at 72°C, with a final extension step of 72°C for 10 min. The programs ended at 4°C for cooling. The PCR products were analyzed on a 1% agarose gel.

3.2.2.6 Purification of PCR amplicons

PCR products were purified to remove unincorporated dNTPs and excess primers remaining in the reaction either by 2 µl of ExoSAP (G.E. Healthcare, Hong Kong) incubated at 37°C for 15 min and enzyme was heat inactivated at 70°C for 30 min or passing through the Wizard SV96 (Promega corp., USA) columns using 100 µl of binding buffer, and washed with 300 µl of washing buffer and eluted into 40 µl MQ water. Agarose gel electrophoresis was performed on 1% agarose and 25-60 ng purified PCR was used for sequencing purpose from each genotype.

3.2.2.7 Sequencing of purified PCR products

Samples were sequenced using big-dye terminator chemistry, on MegaBACE 500 sequencer using DYEnamic ET terminator kit. Treated PCR products were sequenced with Dye terminator reagent (GE Healthcare, Hong Kong), combining 2-3 µl of the purified PCR product diluted to (~20 ng/µl) with 0.5 µl of either of the PCR primer (5 pm/µl), 2 µl of Dye terminator reagent and 2 µl of halfTS-II (Genetix, UK) and 2.5 µl ddH2O. The reactions were cycled for 28 cycles at 95°C for 20 sec., 50°C for 15 sec., 60°C for 1 min. Reactions were ethanol precipitated and dissolved in MegaBACE loading solution. Prior to loading, the samples were re-suspended in 10 µl of loading solution (70% Formamide, 1 mM EDTA) by gentle vortexing. Reaction products were loaded and analyzed on the MegaBACE 500 using an injection of 2 kV for 45 seconds. Approximately 300-600 bases of data were collected for each sample. Each template was sequenced in both strands using one of the specific primer each time in order to discriminate sequencing errors and to achieve high degree of accuracy.

3.2.2.8. Sequence Processing & SNP Identification

High quality sequences were generated from sequence chromatograms obtained on MegaBACE 500 sequencer using Phred (Ewing and Green, 1998). The low quality

regions present at the beginning and end of each sequence were trimmed using Phred 20 cutoff value. Obtained sequences are processed with custom automated SNP pipeline developed for the purpose. Full pair-wise comparison between all reads was performed with 'xsact' and clusters were created following standard parameters. Cluster members were multiply aligned with Phrap to genomic anchor sequence of *O.sativa cv* Nipponbare (IRGSP). A Bayesian statistical approach is used to determine if cluster members represent a unique genomic location with the adaptation of the Polybayes SNP analysis approach. A Bayesian statistical approach is used to determine if a slice of a multiple alignment is likely to be polymorphic given the base calls and base quality values of each sequence contributing to the site under examination. Rate of polymorphism was calculated as per the standard procedure with cumulative number of SNPs and the number of bases examined.

3.2.2.9. Generation of an integrated allele sharing map with abiotic stress QTL map

The SNPs identified were utilized in generating allele sharing map with in-house developed custom scripts utilizing Perl, MySQL and HTML and visualized with Firefox browser under Redhat Linux operating system. The IRGSP sequence data, BAC maps, genetic maps and QTL data were extensively utilized in generating an integrated map. The QTL data integration was restricted to abiotic stress responsive QTLs with a trait category filter in the program.

3.2.2.10. SNP validation, genotyping and conversion to PCR based markers with LNA

Interesting pattern of haplotypes observed in the targeted regions was analyzed and polymorphisms observed were validated with allele specific polymerase chain reaction (AS-PCR) with TrueSNP technology (Sigma-Proligo, Singapore). For analyzing allelic

variants, allele specific primers were designed (Table 3.6), with the 3′ base of each primer matching only one of the biallelic SNP bases to be evaluated and this base was modified with locked nucleic acid (LNA) to increase the stringency and specificity (Latorra *et al*., 2003). Allele specific primers with 3' LNA modification were used along with regular primers for the amplicon and PCR reaction performed using a DNA engine (PTC 200, M J Research, Inc.) with 60 ng of plant genomic DNA in a 20 µl reaction containing 100 mM Tris-HCl (pH 9.0), 500 mM KCl, 25 mM MgCl2, 1% Triton X-100, 2.5 mM of each dNTP, 5 µM of primers specific primers for the region and an allele specific primer (3 primers) , 0.8 U of *Taq* DNA polymerase. The samples were denatured initially at 95°C for 5 min followed by 32 cycles of 1 min denaturation at 94°C, 1 min of primer annealing at 55°C -65°C (varies with primers set) and 1 min of synthesis at 72°C, with a final extension step of 72°C for 10 min. The programs ended at 4°C for cooling. The PCR products were analyzed on a 1% agarose gel and scored for the expected band with allele specific primer in 8 genotypes which are sequenced and also another panel of 17 genotypes listed in Table 3.2.

3.2.2.11. SNP genotyping and evaluation in mapping population

Polymorphisms observed between CT9993 and IR62266 are further used in genotyping 154 doubled-haploid line (DHL) mapping population of rice (CT9993-5-10-1-M/IR62266-42-6-2, designated as IR68586) developed at IRRI with Allele specific PCR (AS-PCR) described above with 15 ng genomic DNA from 154 DHL lines and analyzed on a 1% agarose gel and scored for polymorphisms to map these next generation molecular markers. The segregation pattern of the alleles in mapping population was scored on 154 Doubled Haploid Lines (DHLs) and documented. The genetic map of IR68586 population was updated using mapmaker/exe software.

3.2.2.12 Marker integration and linkage map construction

A genetic linkage map previously constructed by Zhang *et al.* (2001) and revised by Nguyen (2002) with 280 markers (134 RFLPs, 131 AFLPs and 15 SSRs) was used for our study, as anchor markers (genotypic data was kindly provided by Dr. Henry T. Nguyen, TTU, USA.). All the informative markers segregation data generated in present study along with the anchor marker data was used for map saturation and construction of linkage maps. Map construction was done using the MAPMAKER/EXE V.3.0 (Lander *et al.,* 1987 and Lincoln *et al.,* 1992) program following Kosambi function (Kosambi, 1944). Linkage groups were determined using "group" and "error detection on" commands with an LOD score of 3.0 and a recombination fraction of 0.5. The "compare" and "order" commands in Mapmaker were used to identify the most probable marker order within a linkage group. The "ripple" command was used to verify and confirm marker order as determined by multipoint analysis.

3.2.2.13. Validation of SNPs with SBE assay

Single base extension (SBE) assays were designed and performed with fluorescently labeled ddNTPs on GeneAmp Thermocycler 2700 and capillary electrophoresis was performed on MegaBACE DNA sequencer with the help of SNP genotyping module. Target regions were amplified using the set of primers designed for the target region identified to contain candidate SNPs through sequencing reaction. In most cases it is ensured that the target region is with size ranging from 350 bp to 1 Kb in order to avoid interference in the multiple injections. The amplified products were purified either with 2 µl of ExoSAP (G.E. Healthcare, Hong Kong) incubated at 37°C for 15 min and enzyme was heat inactivated at 70°C for 30 min or passing through the Wizard SV96

(Promega corp., USA) columns and diluted to 5-10 ng/µl. An additional primer for the SNP target regions were designed either on the sense or antisense strand up to the SNP site and ensured that the very first base to extend is the site of target SNP. The reaction mixture was prepared by combining 4 µl SNuPe mix, 1 µl 2.5 pm/ µl l SNP specific SBE primer (Table 3.7), 4 µl of ddH2O and 1 µl of template at the concentration of 5-20 ng/ µl concentration. SBE mixture was cycled to generate single base extension product in thermal cycler following 96°C, 10 seconds, 50°C, 5 seconds, 60°C, 10 seconds for 25 cycles. A final cleanup was performed to remove excess terminators and salts before injection of the SBE reaction product using ethanol precipitation using 1/10 volume of 7.5 M ammonium acetate and 2.5 times of absolute ethanol and washed with 70% ethanol, dried and pellet was dissolved in loading dye along with the multiple injection marker. An interval of 100 seconds at 9 kV was used for each injection along with the standard control in order to calibrate the SNP analyzer software. Alleles were detected with laser and analyzed with SNP analyzer using standard calibration matrix for allele detection.

4. RESULTS

4. RESULTS

4.1. Identification of candidate genes for drought tolerance

Drought tolerance is a multigenic trait in rice and regulated in a complex manner by an array of metabolic pathways. A crucial step in identifying alleles associated with the trait is to identify the genomic regions involved in drought stress response from ~430 MB of rice genome. The strategy adapted in the present study to identify the target regions, include construction of subtractive cDNA libraries, analysis of microarray gene expression profiles at various water stress regimes, and comparative analysis of ESTs generated from drought stressed normalized cDNA libraries with reported microarray data besides extensive literature survey. This set of data from diverse array of experiments expected to provide a snapshot of abundantly expressed transcripts under drought stressed condition. Further, mining of alleles at these target loci and their regulatory regions has helped us to identify candidate SNPs which are expected to serve as reliable and precise SNP markers associated with drought stress response.

4.1.1. Identification of differentially expressed genes in drought tolerant genotype

Two rice genotypes having a contrasting phenotype with reference to drought tolerance were selected to identify differentially expressed genes. A highly drought tolerant genotype, Nerica-1 and a drought susceptible genotype, IR64 were used for subtractive cDNA library construction. The drought stress induction was carried out under a rainout shelter under field condition. The water stress condition was mimicked by withholding water in one of the experimental plot at 45 days after germination for 10 days and relative water content was measured with two days interval after stress induction. The sampling has been done once the RWC reached 60% and considered for subtractive cDNA library

construction. The total RNA was isolated from field drought stress induced Nerica-1 and IR64 genotypes and integrity was verified in denatured agarose gel (Fig. 4.1). First strand cDNA pool synthesized from mRNA isolated from IR64 genotype was used as the driver grown under control condition and the mRNA isolated from drought stressed Nerica-1 genotype, grown synchronously with IR64 was used as tester in subtractive cDNA library construction. Differential mRNA was captured and tester specific cDNA was cloned into Uni-ZAP XR vector and phage library was screened with PCR. ESTs generated from this library are to be deposited in the dbEST division of NCBI. A total of 1200 PCR positives were amplified and 580 cDNAs were sequence characterized to identify stress responsive genes differentially expressed in Nerica-1. A representative image of agarose gel of PCR screening is shown in Fig. 4.2. Vector sequences were trimmed and sequences were assembled with the help of cap3 (Huang and Madan, 1999) program in order to generate the contigs. The differentially expressed cDNA clones were annotated based on the homology analysis using BLAST program (Altschul *et al*., 1997) at NCBI. Forty two unigenes involved in various stress response processes such as signal perception, signal transduction, transcription activation and downstream genes involved in stress response were identified besides some novel genes (Table 4.1). The list of target genes and their regulatory considered from the differentially expressed genes for SNP identification are listed in table 3.4 and set of targets successfully amplified and sequenced in the panel of genotypes are listed in table 4.4.

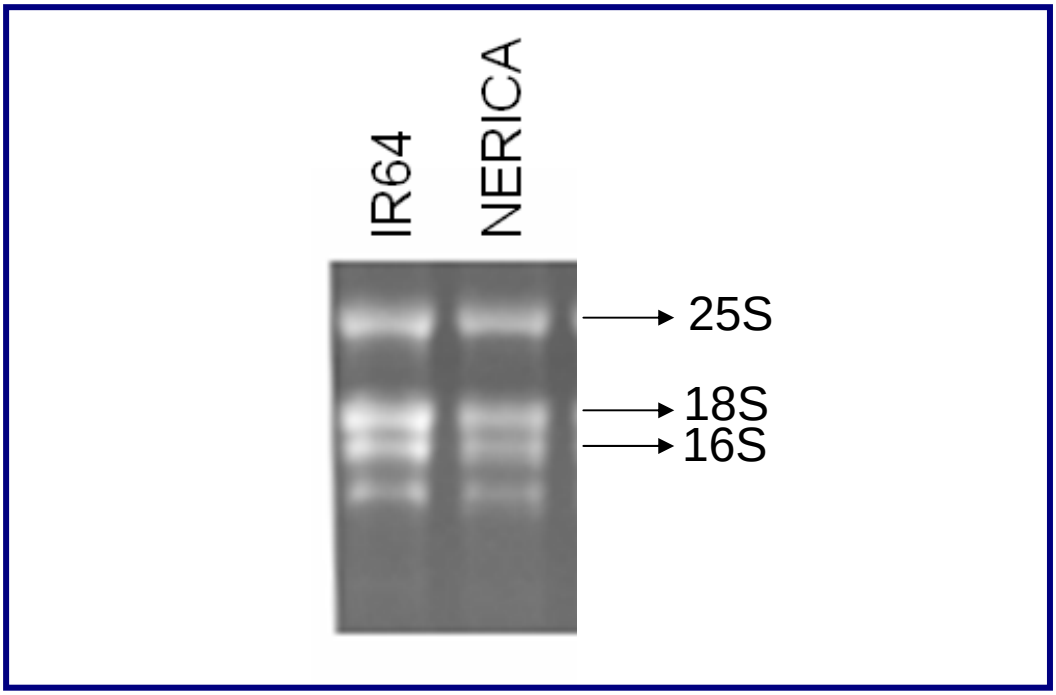


Figure 4.1: Denatured agarose gel image of total RNA isolated from Nerica-1 and IR64 genotypes.

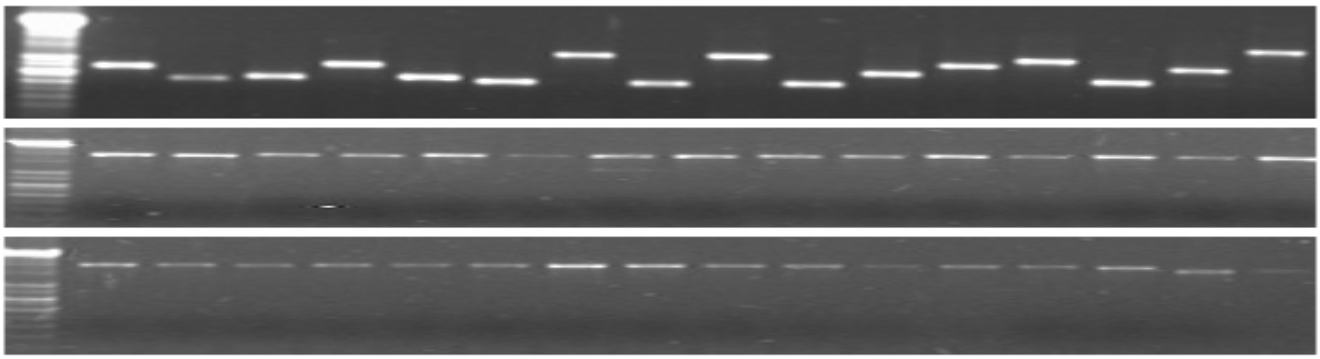


Figure 4.2: Agarose gel image showing positive cDNA clones in PCR screening of differentially expressed genes in Nerica-1

Contig No.	No. of ESTs	Length	Putative Function	Homologous sequence Acc. No.
Contig1	1	374	Receptor like STK	NM_001066888
Contg2	214	1785	23S ribosomal RNA	
Contig3	1	257	phosphoglycerate kinase 3' UTR	AAU44053
Contig5	1	518	Na+/H+ antiporter	NP_001067400
Contig6	23	696	25s rRNA	AAR10852
Contig7	1	527	polyubiquitin (RUBQ1)	AF184279
Contig8	1	378	Hypothetical protein	AK108021
Contig9	15	436	ribosomal protein S15	
Contig10	1	611	squamosa promoter binding protein-like 1	NM_001049328
Contig11	5	223	ribosomal protein S15	
Contig12	9	480	23S ribosomal RNA	
Contig13	1	286	Sac domain-containing inositol phosphatase 3	NM_001063583
Contig14	1	571	Possible bacterial contamination in library	--
Contig15	1	696	Putative GABA-specific permease	BAD88019
Contig16	1	645	unknown membrane protein	NP_001051551
Contig17	1	205	Unknown	Os01g0507000
Contig18	1	351	Possible bacterial contamination in library	--
Contig19	1	603	Possible bacterial contamination in library	--
Contig20	1	424	rubp small unit	
Contig21	5	611	CytP450	
Contig23	1	640	Possible bacterial contamination in library	--
Contig24	4	455	rubp small unit	
Contig25	1	116	carbohydrate transporter	NP_195397
Contig26	2	371	senescence-associated protein	AAO72638
Contig28	2	499	nifU	BAD53892
Conti29	1	481	UDP-glucose pyrophosphorylase	BAB69069
Contig30	1	487	Avr9/Cf-9 rapidly elicited protein	BAD27662
Contig31	1	603	Possible bacterial contamination in library	--
Contig32	1	346	ER1 ethylene responsive	
Contig33	1	571	GYF domain	
Contig34	1	361	(dreb2 like) Helicase	
Contig35	1	464	Possible bacterial contamination in library	--
Contig36	1	313	ribosomal protein S15	
Contig39	1	391	dnaJ protein	NM_001057314
Contig40	5	633	ribosomal protein S15	
Contig41	1	516	transposon protein	AAX96696

Contig42	1	542	transferase/avr9	
Contig43	1	56	Putative protein	
Contig44	1	411	sigma rna polymerase	
Contig45	1	414	PSII 32 KDa	
Contig46	3	500	MAPK2 (MEK1)	NP_001060061
Contig47	1	141	Defence related PR protein BARWIN Hevein	NP_001068190

Table 4.1: Differentially expressed genes under drought stress condition in Nerica-1 genotype compared to IR64.

4.1.2. Microarray gene expression profiling and comparative analysis

The gene expression profiles of ESTs characterized from normalized cDNA library constructed from drought stressed seedlings of Nagina22 genotype were also considered for SNP identification. Some of the interesting genes from the gene clusters showing dramatic changes at various time course experiments in three different drought stress regimes at various developmental stages were considered as the drought stress response genes (Gorantla *et al*., 2005). In addition, some members of interesting gene families identified from the comparative analysis of Nagina22 ESTs with the microarray gene expression profiles of orthologous genes from our recent report (Gorantla *et al*., 2005, 2007) were also considered for the study. The set of genes and their associated regulatory regions targeted for SNP identification are listed in table 3.3 and the successfully amplified and sequence characterized regions are listed in table 4.3. Besides, some of the reported drought responsive genes were also considered with a major focus on identifying SNPs in the regulatory regions of the targeted regions.

4.1.3. miRNA regions associated stress response process

Gene regulation is a complex process under defined stress conditions. Particularly, negative regulation of many genes and subsequent triggering/suppression of other genes is one of the critical areas of study to understand the mechanism behind such regulation.

This kind of regulation is mainly attributed to the involvement of small non protein coding mRNAs called miRNAs. In order to characterize the regulatory switches triggering biogenesis of such miRNAs, the regions involved in encoding miRNAs and their biogenesis were considered. Genomic regions encoding members of MIR159, MIR319, MIR172 and MIR398 were targeted along with their 5' upstream regions (Table 4.2). The set of genomic regions associated with miRNA and their biogenesis considered for allele mining are listed in table 3.5. The set of successful amplified regions are sequence characterized and their details are summarized in table 4.5 along with their genomic localization.

miRNA	**miRNA Acc.**	**Size**	**Genome localization**			
			Chr.	**Chr. Acc.**	**From**	**Up to (bp)**
osa-MIR159a	MI0001092	272	1	AP008207	1,76,78,032	1,76,78,303
osa-MIR159b	MI0001093	188	1	AP008207	12,14,030	12,14,217
osa-MIR159c	MI0001094	197	1	AP008207	65,55,244	65,55,048
osa-MIR159d	MI0001095	189	1	AP008207	65,62,944	65,62,756
osa-MIR159e	MI0001096	191	1	AP008207	97,59,874	97,59,684
osa-MIR159f	MI0001097	188	1	AP008207	66,92,112	66,92,299
osa-MIR319a	MI0001098	191	1	AP008207	2,68,16,730	2,68,16,540
osa-MIR319b	MI0001099	197	1	AP008207	66,76,766	66,76,570
osa-MIR172a	MI0001139	109	9	AP008215	2,25,80,439	2,25,80,547
osa-MIR172b	MI0001140	238	1	AP008207	4,29,16,331	4,29,16,094
osa-MIR172c	MI0001141	111	7	AP008213	1,30,58,016	1,30,57,906
osa-MIR172d	MI0001154	130	2	AP008208	3,51,62,159	3,51,62,030
osa-MIR398a	MI0001051	115	10	AP008216	87,39,976	87,39,862
osa-MIR398b	MI0001052	88	7	AP008213	1,45,45,125	1,45,45,038

Table 4.2: microRNAs reported to be associated with abiotic stress response considered as candidate regions for stress in response.

4.2 *In silico* analysis of Nagina22 SRGs with available unigenes set

We have extensively utilized set of unigenes sequences of stress responsive genes we have reported recently (Gorantla *et al*., 2007) to identify SNPs *in silico*. Custom Perl scripts were developed to automatically process the available unigenes set to identify SNPs. This pipeline script called gtCluster was developed to overcome the problem of redundancy with in the unigenes belonging to the same cultivar. The program was designed to generate contigs considering the source genotype in view and compares the genotype specific contigs to identify SNPs. The flow of the data in the analysis pipeline and the procedure for integration of various tools is shown in Fig. 4.3. The SNPs with significant threshold value are fed into a newly developed MySQL database called ric*e*SNP after manual inspection of SNP site. We attempted to analyze 586 SRGs but we could retrieve unigenes set for only 87 regions. Analysis of these 87 regions enabled us to identify 432 SNPs with the probability values greater than 0.9 of which 162 are transitions, 250 are transversions and 20 are insertion/deletion (indels). The *in silico* methodology is a free of cost approach. However, factors such as limited availability of ESTs at targeted loci, incomplete genotype data and difficulty in genotype to phenotype correlation makes researchers to compromise with the available data. In view of the above we have sequenced the targeted regions from the selected genotypes for reliability and accuracy.

4.3 Amplification and sequence characterization of target regions

The regions targeted were successfully amplified using sets of primers designed for each of the target region after rigorous standardization of PCR conditions optimal for amplification. Particularly, annealing temperature for each of the primer was fixed based

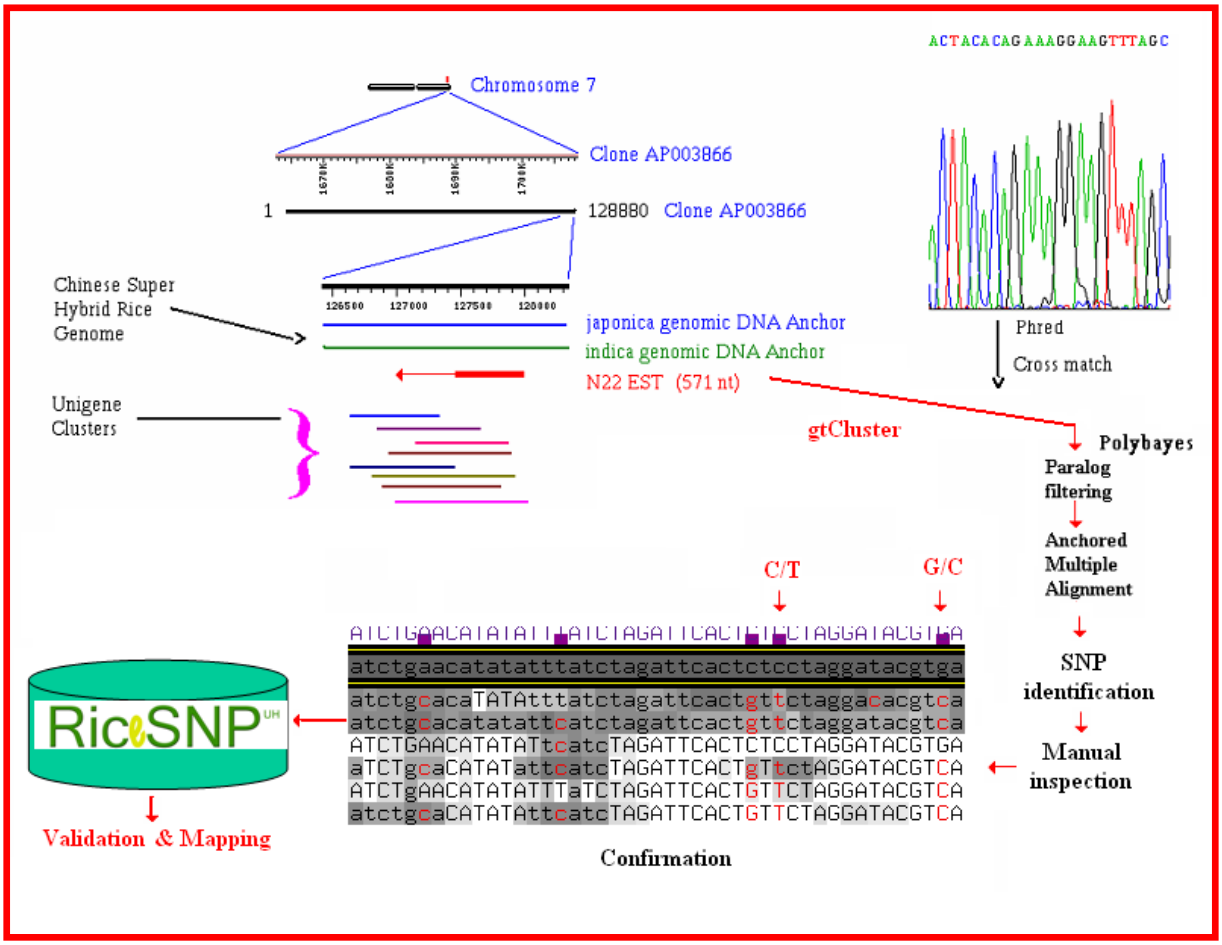


Figure 4.3: Data flow and processing diagram of gtCluster pipeline for *in silico* SNP analysis.

on Tm of the primer set, and after a series of gradient thermal cycling conditions. Non specifically amplified primer sets were not considered for subsequent processing and eliminated if single amplicon was not obtained even on addition of adjuvant after standardizing the annealing temperature. Most of the primer sets were standardized with the annealing temperature ranging between 50 to 55^{o}C with 20 mM TMAC as an adjuvant in the PCR reaction. The list of Oligonucleotide primer sets and optimal conditions used for each of the primer set were listed in table 4.3, 4.4 and 4.5 along with their localization in IRGSP sequence map to the base pair level. The PCR products were verified on agarose gel electrophoresis (Fig. 4.4) after amplification in 8 genotypes and used for sequencing after purification of PCR reaction as described in materials and methods. The sequence characterization of over 1 million bases at the target regions from a panel of 8 genotypes was carried out and chromatograms were utilized identification allelic variation. A snapshot of electropherogram generated on automated DNA sequencer is shown in Fig. 4.5.

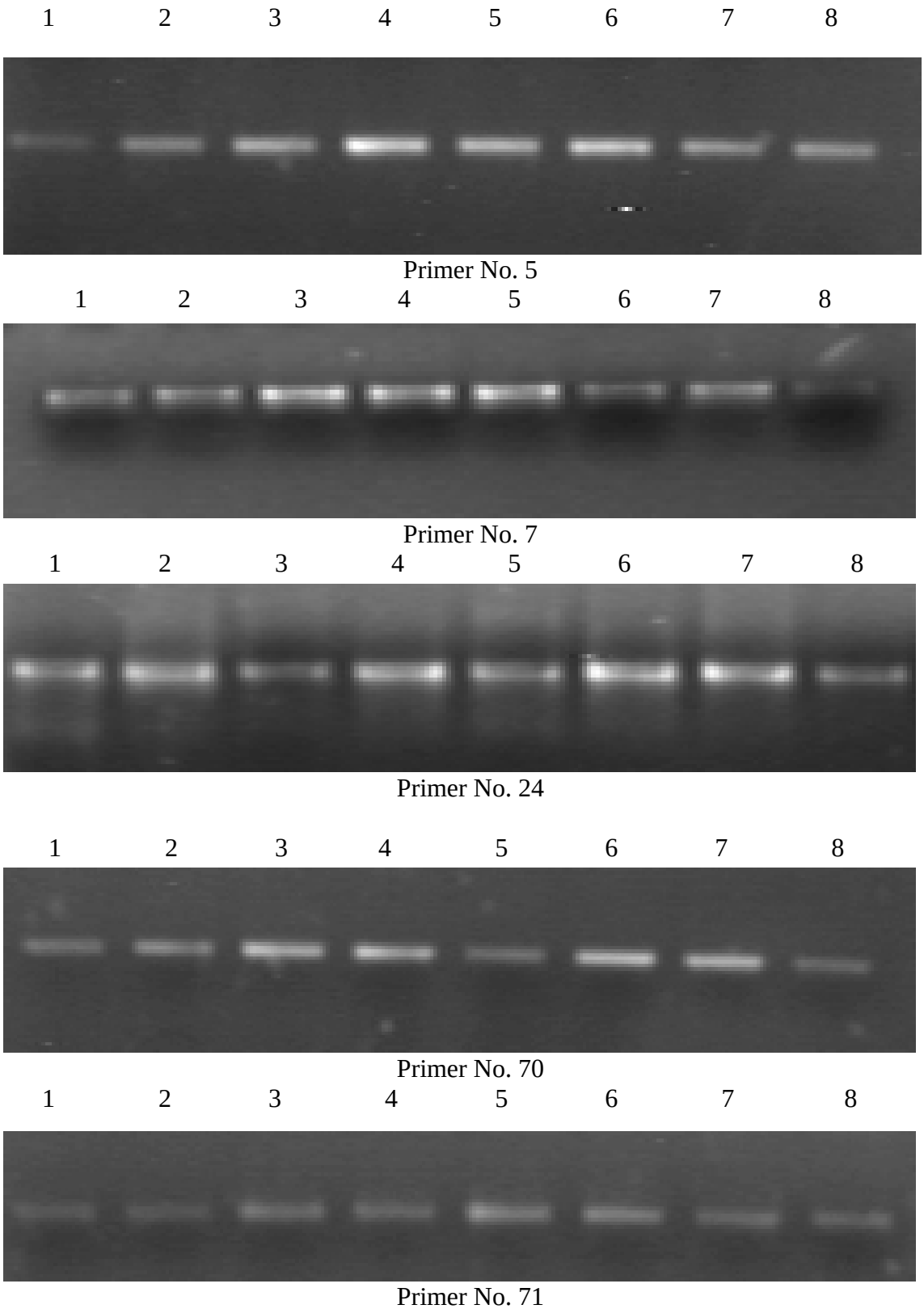


Figure 4.4: Agarose gel image of successfully amplified target regions from a panel of 8 genotypes; 1- CT9993, 2- IR62266, 3- Azucena, 4- IR64, 5- Nootripathu, 6- IR20, 7- N22B, 8- Nerica-1.

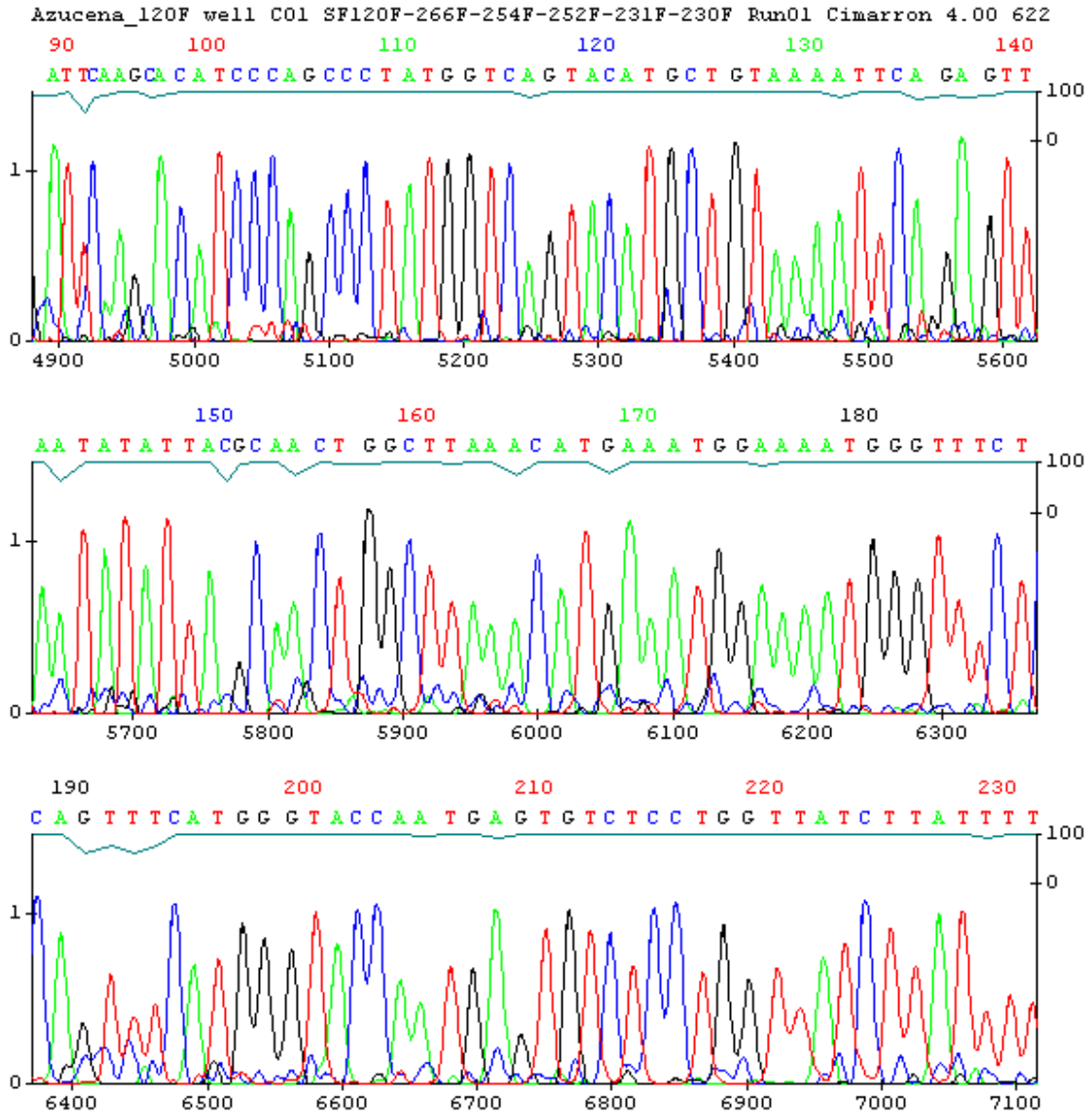


Figure 4.5: Snapshot of chromatogram generated from automated DNA sequencer.

Primer No.	Description of the target region	Forward Primer	Reverse Primer	Ta (°C)	*PCR Adj.	Chromosome No.	#Chromosome Acc.	Spanning	
								From	To
4	RAS	CCATTGCATCATAACTGGACATCC	TGCTTCAGTAGAAACATGCAGTG	63	Nil	1	AP008207	4161693	4162209
5	Calmodulin	TCGACGTTGCACCGTGTGACC	TCGAGCGGTTAGCGAAGAACG	68	Nil	3	AP008209	11491926	11492880
25	DREB	TACTGATGATCGCGAGTTGG	ACCATAAGCAGAGCTGGCAT	52	Nil	6	AP008212	1434437	1434737
37	MAP kinase	GGATCCAGTAAATTTGGCAG	AGCATGTATAATAGCAGATTATTAGTCAG	55	Nil	3	AP008209	8548441	8547947
39	DNA binding protein	GGAACGGAAAAGGTCTCTC	GGAGAATGGTAGCTTGGC	55	Nil	4	AP008210	22388311	22389193
41	KNOX	TTCAATTTCTAGGCACAACTAAG	TGTTGTTGTAGCAGTTGGG	52	Nil	6	AP008212	26031787	26031155
43	Zinc finger protein	TGTGTGTGTTTGTGTCAACC	TGATTTGCTTCATGATTTGTG	52	Nil	12	AP008218	23549540	23549997
46	Zinc finger protein	TGATGAGATGAACTGCACG	GATCTGAATGAATGCTTTACTTACT	55	Nil	5	AP008211	753322	752642
50	Defensin	CAGCCGTACCCTGAATGC	GATGAGTGATGAGGCTGCTAC	55	Nil	2	AP008208	25208506	25209037
51	Defensin	ATGGCTCCGTCTCGTCGCATG	CTAGCAGACCTTCTTGCAGAAGC	55	Nil	2	AP008208	25209067	25209409
59	Glutathionein S transferase	CCAGGACCATGCTGTTATT	AACTGCCAGGTCGATTAAC	55	Nil	3	AP008209	1933274	1932753
60	Disease responsive gene	TCGAGGATGGTGTCGATAG	TAGACGGGCAACGATATG	50	Nil	11	AP008217	3899228	3898815
61	Disease responsive gene	CCACGCTACAAAAGGTCA	TCGGTAGGTGAGGTGAGAA	50	Nil	11	AP008217	3898627	3898175
63	Chitinase	AAACACGTTGATTCGATTAAA	TTAGCTACCTATCGATTGTGATG	50	Nil	5	AP008211	8803867	8803401
78	Protein kinase 3 UTR	AATCCAGATGCTGTTCATGT	AAACGCCTCCAAGAGACC	45	Nil	1	AP008207	5512582	5512081
79	Protein kinase	ATGCGCGACCAGTAATTC	AATGATCTGCTACTAAATTTCACA	45	Nil	1	AP008207	5509989	5509523
81	ABA induced gene 3UTR	TCGTCGATTTGTGTGTGTAAT	CAACTACCAACCACTTTAATTATGAA	45	Nil	11	AP008217	3259128	3258620
83	Inositol phosphate	CAGTTTTAGTTTCCAGAAATGTTT	ACATTCGAAATATCGGTGATAG	45	Nil	3	AP008209	21481595	21482102
86	Ethylene responsive gene	ACTTCGCAGTCCAGTACCTA	TGGTAAGACAAGTGATATGTGGT	45	Nil	5	AP008211	24301650	24301119
91	Hormone responsive	CTAGCTCATTTGATTGGTGTG	ACACTAAGATAGCAGTGAGCCA	56	Nil	8	AP008214	28371312	28371817
92	Metallothionein	TGGCTTGTTTGGTAGAGTTTC	TGCTTTCTTGTGCTCTGGT	55	Nil	1	AP008207	5477077	5477566
95	Betaine aldehyde dehydrogenase EL	CGCGAGCGATGCCAGAG	TTACAGCTTGGAAGGGGATTTGTAC	55	Nil	8	AP008214	20377994	20378587
98	Choline monooxygenase	GAATTTGTCATATGTCGGGATGC	CAAATTTGGCCAATACAAAAGGCC	56	Nil	6	AP005931	140557	141361
99	Cyclophilin	AGCTGAAGAACTGATTTCTGATGAG	ATGCCCGTGCTCGAACTC	48	Nil	3	AP008209	33785903	33785163
100	Ribosomal 6 kinase	CTGGCCGAAAATTCCACTTGC	GCACTACAAAAGGATGGTCAAC	55	Nil	7	AP008213	28803714	28804514
102	MAP kinase kinase	ATGCAGTCTCTTTGTAGACACATC	AGAATCAGGTGGAAAGAGAACGAG	55	Nil	2	AP008208	1855756	1855043

104	M_Ring finger protein	GGATATGAGCCGAGTAAAGCTGCGG	GCTGAAGACGCTAGGGCTTG	55	Nil	1	AP008207	40109667	40110266
105	M_Ring finger	CTGACTCTGGGCAGGTATCCTAGG	ACAACACCTGTTAAAACTGCTCCAGC	55	Nil	1	AP008207	40113643	40114178
108	M_Gene with MITE	GAGCTGAATTTTATCCATGACTGCTG	TTAGACAATCAGTGGAACCAAGATAAGTG	55	Nil	10	AP008216	20832943	20833956
109	M_Unknown gene	GCCGTACCTTCCAGAGATTTGC	TCACTGCCTGAAATGCTGCAAAACAG	50	Nil	4	AP008210	25361560	25361907
110	M_Unknown gene	AGTACTAGCTAGTCAGCGTGATC	AGAAGGTGGGTTAATTAGTACTCCG	50	Nil	4	AP008210	25358300	25358833
111	Clone similar to NL38C07	ATGTGGAAGGCGCAGCGG	TCACTCATATACACTCATTGCCAATGG	55	Nil	10	AP008216	8691446	8690663
112	Clone similar to NL38C07	CGGTCCCAAAATAATCACAAGAGACC	ACATGGCCCATAAGGCTAGGTG	50	Nil	10	AP008216	8692464	8691459
114	Thioredoxin	TGATCGTTTGATTGTTGGATCACAG	CAGGCGATCACGACTCC	55	Nil	7	AP008213	4609084	4608256
115	Ferrudoxin	ATGGCGGCGACGGCACT	TTAGATGAGGTCGTCCTCCTTGTGGG	57	Nil	8	AP008214	236198	235779
116	Ferrudoxin	ATGTGATGGATGCTTGACATGC	GCGATGTGCAGCCTCTAATAATTTTG	55	Nil	8	AP008214	237462	236741
118	M_Jasmonic acid induced	CCTCTTGGCCTTCCCCAC	GATGAGTGGTCATATTTGCTGGTCG	50	Nil	4	AP008210	14092304	14093397
119	M_RD22 LE	ATGGCTAGGTCACTCGCTG	TTAATTCAGTTGCGCACCCAGATC	55	Nil	5	AP008211	7050557	7049557
120	M_PSII10KDa LE	ATGGCTGCCTCTGTCATG	CTAAGCCAGTGCACTAGTGTTG	55	Nil	7	AP008213	2486556	2487649
121	M_PSII10KDa	ACTGTCCCGGCAATGTTAGAC	ATGCAGTTCATCATCAGATATCCATGGC	55	Nil	7	AP008213	2485817	2486491
123	M_Cold Induced gene	CGTTTCGGGAACTTGCAGG	TGGCGCGTGTTATATACGG	50	Nil	5	AP008211	25886082	25886956
124	CBF4	GCAAACATCGTTGACCGGTG	CGGACCATTTGCTGCGAAATG	50	Nil	6	AP008212	3308467	3309432
129	M_ AP2 family gene	AGTGGAAGCTCATCTTTGAAGAGTGG	GTGCGGGTGGACAACGTAG	50	Nil	3	AP008209	4333592	4332927
131	M_dnaJ	GGTGTTCGCATTTATAATTCTGACAC	CGTAATGATGGCTCCATCGTTTG	55	Nil	1	AP008207	20982956	20983753
133	M_DREB2	CCTGCTGACCAAAAGGAAGTTC	TGCAAGCAACCTAACCAAGAAACC	55	Nil	9	AP008215	12162031	12162933
135	M_PCF3	CGTTCACACAAACACTAACAGCAC	CCATATACTTTCAAACCTTGGAGAGC	50	Nil	11	AP008217	3742631	3741878
137	M_MADS	AAATAGTTGCTTTGGGGATTTGATAC	GCAAGCGCAAAAGCCTC	50	Nil	1	AP008207	40336055	40336576
139	M_Catalase	GGCCCTATCCTCCTTGAGGAC	TGGAATTTGACAAGGTGAGGCTTTCC	50	Nil	6	AP008212	30429626	30430539
141	M_NAM	TTCAGGTTCCACCCGACG	CTTCTTGTGCATCGGCAGGG	50	Nil	8	AP008214	26760135	26760635
142	M_NAM	CCACTGCTTCATTAATTGTTTGTGTAG	TTCTCACGGGAAATTGATTAGTAGG	50	Nil	8	AP008214	26759035	26759508

144	M_ Glycine rich zinc finger protein	CGTGTATCGGTACCCGGCAG	CCTCGCCTCTCCCTGATCTGATTAG	50	Nil	2	AP008208	1105145	1104218
145	M_RING-H2 finger ZFP LE	ATGAGCTACCTCCTCTCCTACATC	TCAGAATCCAGGGACGAGCATG	50	Nil	4	AP008210	9348289	9347936
146	M_RING-H2 finger ZFP	GAGAAAGGGTGAAAGGTTTGAAGTG	GTAGTTGATTTGTTCATATCCATTGGTACG	57	Nil	4	AP008210	9349460	9348614
147	M_HSP90 LE	ATGGGCGGCGGCAAGAATAAG	TCAGATGAGTGGGCAGGAGAAGTC	57	Nil	3	AP008209	189525	189950
148	M_HSP90	GATTAAGTCAAAGTGAATTGCACGAG	TGCTATTTATAGGCAGCTCCAGG	50	Nil	3	AP008209	188751	189404
149	M_Unknown gene homologous to CB965622	ATGATGGATAAATATGCCAGATTAATGC	CCATTGATATGCAAAAGCATCCTTG	50	Nil	10	AP008216	20836576	20837530
150	M_Unknown gene homologous to CB965622	GGTTTCAAGGTTATTGTCTGAAGTTG	TGTTCAGTTGTAAGATTCTGGATGC	48	Nil	10	AP008216	20834095	20834783
151	M_Unknown homologous to CB966119	CAGATGCCGCATATGATCTTCAC	ATGGTTGGCATTTCAAGCACTTG	50	Nil	2	AP008208	12335170	12334868
152	M_Unknown homologous to CB966119	TGTGCTTGCTCAGTTAATTCCTC	TAGGGATACAAAGTTGTTGTCTCG	50	Nil	2	AP008208	12339349	12338628
154	M_Gigantia like gene	ATAATTAATGGCAGTTTCCACATGGC	AATGGTCATCTTCTACCTTTGCTTCTG	50	Nil	1	AP008207	42303672	42304087
155	M_RS6 kinase GCD	GAAGGATCCATTGGAGATAAGATGG	GGGCGGTCTCTAAAGTCATATTC	48	Nil	6	AP008212	25620974	25621759
156	M_ RS6 kinase 5UTR	CACCGCTTCCCCTCCTTC	CCGCATATACACCACTTAACATAGGCTC	46	1 µl	6	AP008212	25619492	25620231
157	M_Water channel GCD	TACATCGTGGCGCAGTGC	TACCGGGACATGGGAGTCG	48	Nil	7	AP008213	15323801	15324025
159	DREB1B	CGGGATCCGCGAGGTAAGCCATTAGCGCATG	CGGAATTCGGATGACTCTCTCTGGTTCAC	55	1 µl	NL	NL	NL	NL
162	Aquaporin	AGAACTCAGCTTAATTAGTTGCCAAATGCG	GCCGCTCTCCATCACCTC	55	1 µl	7	AP008213	15352101	15353001
163	Aquaporin	ATGGGGAAGGACGAGGTGATG	GCTTCTCTTCTGTTCAATATGCGCTTG	55	1 µl	7	AP008213	15352972	15353847
164	Aquaporin LE	GTTTGGCAGCAGCACAGTTTTGTG	TCACGCGTTGCTCCTGAAGG	55	1 µl	7	AP008213	15355301	15355854
167	Phenylalanine ammonia-lyase	TCTAAGGTGTAGTGACCTTGCACTG	GTTGCCCGCCATTGCTAC	55	1 µl	2	AP008208	24996557	24995684
169	Coumaryl CoA Ligase	GCGTGGCTGCATGCATTG	TCGACGGTGGCCATTGG	55	1 µl	8	AP008214	24846355	24847148
170	Pathogen responsive (PR) protein	GAGTGAAAGAGTGTACCTGAGGCTC	CAAGGCAAAGCAAAACCGTGC	55	1 µl	6	AP008212	833352	832482
171	Disease responsive gene	GGGAGAAGATAGGCACGTCAAAGAG	ATGATTAAAGTCTCCTTCCGGTGCAG	55	1 µl	7	AP008213	381762	382417
172	Wound induced gene	CACAGTACAGTTAGTCCTGATAAGCAGG	GGATCACCAAGAGAGTGCTGC	55	1 µl	1	AP008207	1760848	1759955

173	Thaumatin	GCCAAACGATGGGGACCTTAG	GCTCCATGTTGGAGTGTAGCC	55	1 µl	6	AP008212	28440969	28441656
177	S-adenosyl-methionine Synthase	TCTCTAGCTCTGACTCAAATCGTACTGC	CAACGCTACCAGCAACGAGG	56	1 µl	5	AP008211	2027923	2028785
178	APX TL29	CTACCATAGTTGATCTATCCAGACCTG	GCGTGAACAGATGTGACAAATATG	55	1 µl	4	AP008210	30418455	30418965
179	APX	ACGGATGGTCAAACGTTGGAC	CGAAAGGCTTGGGCGAATG	56	1 µl	3	AP008209	9812639	9813451
180	HSP82	TCCTGCCAAAATGTTGGCAATATTG	TGAGAGGGCGAGATCGTG	55	1 µl	9	AP008215	18464263	18463660
181	HSP60	GCGTTTGTGTAAAAGTAGCTAGCAGAAC	AGAGGATAAGGCTGCGACCTC	55	1 µl	6	AP008212	816364	815471
182	Heat shock Transcription	GGCGCGAGACCGAACTAAG	TGGTGGCGCGGAATTTATAGG	40	2 µl	2	AP008208	19333141	19332191
183	CBF1_Pro	TGTGACAGTATCGCTTCATAACATGTGG	AGGAAGGCTTTTTCTGTGGGC	55	1 µl	6	AP008212	1427414	1426503
187	Dehydrin	GAAGACAGAGGTAGGTTTTGTTAGAG	AGCTAAAGCTGAAAAAAAACTCTCGGAG	55	1 µl	11	AP008217|	14768060	14768961
189	Helicase	GCGAGAGAGAGTTTTTCCGATTCG	GCCCAGATAGGAAATACAATTTCAACTTGGTG	55	1 µl	4	AP008210	32050327	32049488
190	CBF3	CGCGGAAAACGGAATAAACTATCTCCC	CCGGGTGCCTCGTCTC	55	1 µl	9	AP008215	20096274	20095574
192	CBF3	CGGTTTATGAAAACCGGTCGGTATTAG	GCCGCGTCTCCCTGAAC	55	1 µl	9	AP008215	20100271	20099576
196	Submergence induced protein	GTTCTCGTTGGTACCATGGGATC	GAGTTGCTTTCCTATTGGGAAACCC	55	1 µl	3	AP008209	3319327	3320163
198	CDPK	GCCACCGAATGGTAAGCAGG	TGTGCGGTTGATGAGTAGATAGGCTAG	55	1 µl	4	AP008210	29503811	29504439
199	SAM decarboxylase	GTCGACGCATCCAGCGAAC	CGGTTTCTCTCGACCACGAGAC	55	1 µl	2	AP008208	24064262	24064720
200	WRKY Transcription factor	CCCAGAACCTCACCCCTC	CGGCCATGGTTGGTTCAC	55	1 µl	4	AP008210	30578526	30579269
201	26S Proteosomal protein	GTAGAACTCGACGATACCGACG	CCATCCATCTTCCAAGTTTAGTTAGC	55	1 µl	7	AP008213	29387424	29386664
202	Wound induced protein	CCTTGAACTTTCGTAACCATACGCTC	GCACTACATGAGAACAACATGACACG	55	1 µl	8	AP008214	16336971	16336134
203	Signal recognition particle	GAGTTCAAACTTAGATGCGCCGTC	GAGGAGGAAGGGACAGAAGAGG	56	1 µl	10	AP008216	18521489	18522191
205	Cytochrome P450	GCAAGATAATGTTGGATGAAAAGTGGAGTAG	CACAGGACCACAGGTGATC	55	1 µl	1	AP008207	25046140	25045477
206	Serine threonine kinase	GGCGCTAGTGGTGGTGTG	CTCCCGCAGATCAATCCTCC	55	1 µl	3	AP008209	1499614	1498636
207	Enolase	GGAGGAGGTTGCAAAGAAACAG	GGATCGGATGGATCTTAGATCGGG	55	1 µl	3	AP008209	7822164	7821328
208	R2R3 Myb	GCTTATTTGTAGTCGATTGAGGTTTGCC	CATGTGTGGGTGTTGCATGATAAG	55	1 µl	5	AP008211	769785	770475
209	R2R3 Myb LE	ATGTTGGCCTTGGTGGTTTCAC	TTAAGCTAAGAGCAGCAATTGAATTAAGGCG	55	1 µl	5	AP008211	770978	771661

210	MYB1	GAATCTTAACTACTCTCTCCGTTTCACAATG	CGCGAGCTCGCCGGTATATATG	40	2 µl	5	AP008211	20957119	20956352
212	RACD GTP binding protein	GTGTATTTCTACATGGACTGTAAACTCTAC	CCTATTCGATGAACATGGTGTGATG	55	1 µl	2	AP008208	35898401	35899199
214	RIC1 GTP binding protein	GTTAAAGTCTCTTTATGTGGGTCC	CATGAATAGCACTTTATGTGTGAC	55	1 µl	1	AP008207	21144727	21143946
216	Transcription regulating protein	CTCGTGTGGACAGTTCATATTTCCTG	AACATTTAGCACTTAGCAGTAGTCTCC	55	1 µl	5	AP008211	19535791	19536650
218	Proline rich protein	GGTCTGAACACAGCAGCCG	GCCTCGCCATGGCTGG	55	1 µl	5	AP008211	7648605	7647766
221	Dehydrin	ATGGCGCATAAGCTCCTC	TCAGCTCTCCTCCTCGG	55	1 µl	4	AP008210	22795734	22796370
224	Cytokinin oxidase	GGCTGAACATACCTAGCGTTCTATGC	GACGAAGCAGTTGAGCATGAGG	56	1 µl	1	AP008207	5275433	5274469
225	M_ Unknown gene homologous to CB964527	CTCATATGCAGTGTTAGTCAATTATGAACTG	CTGATTCCAGTGAAGAAATCACACC	55	1 µl	10	AP008216	20835802	NL
226	M_ Unknown gene homologous to CB966078	CCTTTCAGGTTTCCCACCTTGTG	AGGGTTTGGGAATTTTTGGAGTTGATGG	56	1.5 µl	2	AP008208	NL	12337761
227	Cystein protease	GGAGTGTACAAGAACAATTTAGGCCC	AGGAGCACAAGAGCCAGG	54	Nil	1	AP008207	42847076	42848111
229	Disease induced protein	CTCTACAGGGCTGGAGTGGC	CTAGCTACAAGGCAGGTGTGAGAAC	55	1 µl	1	AP008207	42837274	42838216
230	Disease induced protein E1	ATGGAGGTTGAGGCTTCCTACAG	GAAGATGCATATACATCATGAGCAACTGG	55	1 µl	1	AP008207	42838260	42839223
231	Disease induced protein LE	GCAATCTTCTAGAGACTCCCAAGTCC	TTAAAGAATTGTCGACATGATTATCTCCTTGAC	55	1 µl	1	AP008207	42839314	42840264
233	Disease induced protein 19	GACTTCGGATATGGAGGTAGTTCC	GGGGAGTTAGTCGTTGGGAATG	54	Nil	2	AP008208	11885912	11885042
234	Low temperature induced protein (LT16)	CCGGAGCATACAATGTACCCTG	GCTTGCATTATGGCGAGGTTC	55	1 µl	7	AP008213	26348060	26349057
235	Water stress induced gene 18 LE	ATGGCTTCTCAGCAGGAAC	TTACTGGTAGTCACTCCCAG	55	Nil	1	AP008207	29227384	29226526
236	Water stress induced gene 18	CGACATGTACCAGTACCATGAATCG	CTCACCAACACACGAACTGAAACTAC	55	Nil	1	AP008207	29228315	29227399
237	AP2	GCTTCTGTGGCTACTATTGTTATCAGC	CGCTAGAGCAACAACAAAATGCTG	48	Nil	2	AP008208	25616869	25615966
239	ECP40	ATCCGAGTTTGGACGGGTG	CCATAAGTTCGGTCTGCAATAACACATC	55	1 µl	11	AP008217	14648172	14647353
241	Late embryogenesis abundant protein (LEA)	GGCCGCTCAATGTGTCATC	CAGTATTTATAAGGGTTGGCATCGAGAC	55	1 µl	5	AP008211	26719821	26720794
245	Aldehyde dehydrogenase	GTGTTTGTAGCGTAATTGCGTAAAGTCC	AAGCTCGGCGAGGAACTG	55	1 µl	9	AP008215	16284650	16283654
246	Drought responsive protein homologous to Os.57533	AGGCACACGTTATATGGTCATCTTCC	GAACGTGAGGGAGTAATTGCAAAGC	57	Nil	9	AP008215	953570	952771

248	DREB1A similar to Os.57517	CTATCGAAGAATTGACGGTGCC	TGGTGGTCGTTACTCGGTG	52	0.5 µl	9	AP008215	20096649	20095760
252	EF hand calcium-binding protein similar to Os.75123 and Os.69287 GCD2	GGCAGGTTATTTCACAGGATTGATC	GGAAGGCGACGAAGCATGTC	55	1 µl	2	AP008208	8360400	8361451
254	Early drought induced protein similar to Os.13968 LE	ATGGCGGGAACGGCGAAC	CTAGATGACAAGAATTGGAGCACTCAGG	55	1 µl	5	AP008211	2160064	2159503
255	Drought-induced hydrophobic protein similar to Os.46402	TGACTTCCCCAGGGCATC	CTACTTGGTGATGACCCAGACAG	52	0.5 µl	7	AP008213	26349490	26350507
256	Drought-induced hydrophobic protein similar to Os.46402	GTTGCTACAGCTTCATGTAGCC	TTAGCAATGTTCTTGTGGATCTAGAGG	51	0.5 µl	7	AP008213	26347994	26348965
258	Drought-inducible protein1 similar to Os.22276	GTTCATACCACATTCCTCATGAGTGTG	CAACGGCTGTTGGTCAATCG	55	1 µl	1	AP008207	33310217	33311185
260	RNase S-like protein similar to Os.9417	GGAGACGATGATGGCAGGC	CCCAACCATTGTGAGGGTATTTATAGG	52	0.5 µl	9	AP008215	20857497	20856558
262	Ubiquitinyl hydrolase similar to BU673177	GCTGACTATAGATTTGTAGCCCGC	GGTTTAGGGATGTTATACCAAGTCGAC	52	0.5 µl	4	AP008210	27338793	27337778
266	Tripeptidyl peptidase-II similar to BU672860	GCGTTAGTAGTAGTTGTTTCTTATATTTGGG	GACCGGTTTATGTAGCGCTG	55	1	2	AP008208	26970811	26971724
267	Peptidyl-prolyl *cis*-trans isomerise similar to BU673613	GACGATCAAGCAAGAGAGATTACATGAGAG	GATTGGACTCTCCTCGGCG	51	0.5 µl	2	AP008208	32039270	32038283
268	Protein kinase similar to BU673471	GTTGCGGGTGAAATTTGTGTAGTCG	CTGGATGGGACTGTTAGGGG	47	1	10	AP008216	20277758	20276726
269	Proton pump interactor 1 similar to CB965018	GCCTTTGTATCCCTCTCCGTG	CGAGGCACCATACAGCTCC	55	1	8	AP008214	18467321	18468152
270	Tubby like protein similar to CB964631	CTTTCGCCTCCTCTCGCTTC	TCAAATCCTCATCGTGCTCTCGTG	47	1	5	AP008211	25296268	25297070
276	Lectin like kinase	GTCTTGTATTACCCTCGTCCAGG	TTACCTTCCGAACCATTGTCACC	55	1 µl	1	AP008207	32988572	32989235

Table 4.3: Set of target regions successfully amplified from gene expression profiling studies and sequenced to identify SNPs in a panel of 8 genotypes.

NL- Not accurately localized; *Nil- No Adjuvant was used; Adjuvant if mentioned 400 mM TMAC (Tetramethyl ammonium chloride) at the final concentration of 10-40 mM in PCR reaction; #Accession number of the chromosome refers to version of 1 corresponding rice chromosome sequence.

278	NDC1_Receptor like kinase	GATCCCGACCCCTAGTCATAG	CTTCGTGTGAAAATTCCCTCTGTG	57	1 µl	7	AP008213	26010350	26009512
284	NDC15_γ-aminobutyric acid (GABA)	GACACGGTGCGCTAATTTTG	CCACTGATTCACTGATCTGCC	55	1 µl	1	AP008207	41544790	41543893
285	NDC16_unknown membrane protein	CCAGCACCAGCAAACAATG	GCTCTCTCGAAGTCCAAGTCAAAC	55	1 µl	3	AP008209	32892981	32892013
286	NDC25_DNA binding protein	ACGATGATGCCGCTGGG	CAAGAGGGTAGTAGTCATTTCTATGAG	55	1 µl	3	AP008209	33629517	33628685
288	NDC28_nifU	CGTTTGCAAAACACTGTTGCTGTG	CGAGTCATCATGGGCCACATTAG	57	1 µl	6	AP008212	28608862	28609840
290	NDC30_Avr9/Cf-9 rapidly elicited protein	GGGGGAAGATGACGTCAAG	GTTACTATCTCCATCCTAAAATGTAAGC	55	1 µl	2	AP008208	7609755	7610565
292	NDC33_GYF like protein	CGTAAGTTTCTTGTGTATGCAGGC	GTGTAACTTTTTATATCGGTCCATCTCC	55	1 µl	NL	NL	NL	NL
293	NDC34_Helicase	GAAGAGGTAAAGCAGGCTACTACT	GCATGAATTATATGTTTGGATGCTTC	55	1 µl	11	AP008217	270213	271194
295	NDC41_Transposan like protein	GCCTGACCAAGACGCATG	CATCGGTTAGCTCTGTTTTCACCATC	55	1 µl	11	AP008217	8462830	8463722
297	NDC46_MAP kinase	GCTGGCTACTTTTTTCCAAGTG	CGCAAGATATGTATGGTTTGTGTTGC	55	1 µl	7	AP008213	23091108	23092059

Table 4.4: Set of target regions successfully amplified based on the genes differentially expressed in Nerica-1 genotype when compared to IR64 and sequenced to identify SNPs in a panel of 8 genotypes.

NL- Not accurately localized; *Nil- No Adjuvant was used; Adjuvant if mentioned 400 mM TMAC (Tetramethyl ammonium chloride) at the final concentration of 10-40 mM in PCR reaction; #Accession number of the chromosome refers to version of 1 corresponding rice chromosome sequence.

301	MIR159a	GCTAGCACAGCAACTCATCTG	ACAGTAAAGCTGCTCGAGGG	48	Nil	1	AP008207	17677550	17678325
309	MIR172a	ATAGTATGACAGTGCTCTCTCGTC	ACTCTCTCCGTTTCACCGTTTCAC	48	Nil	9	AP008215	22580244	22580733
311	MIR172d	AACAAAGGGGATTTCTGTAGCTG	TAGCTAAGTCATAACCAGTCCTC	50	Nil	1	AP008207	42915860	42916460
313	MIR172d	TTTGACTAGCTCCCACCATG	ATCGTTTAGTCTCTGCTACTGG	48	Nil	2	AP008208	35161749	35162544
314	MIR319a	TACATCAGTCTTGCATACCAG	CTTCGTGATCACTGCAGCTGGG	56	Nil	1	AP008207	26817938	26818518
317	MIR398a	ATCGGGGGTCACCGAAATC	AACCTATGTCCACTACAAGCTG	57	Nil	10	AP008216	8739587	8740273
318	MIR398b	CAGAGCGCATTTTTTCCGGG	CATCCTGTATTTCATGCGGACC	56	Nil	7	AP008213	14548086	14548716
319	MIR398b	GCCATTTAGGAGACCTTTTGGC	AAGTACGACGCCAATCGATGC	56	Nil	7	AP008213	14544981	14545364

Table 4.5: Set of target regions successfully amplified from the set of genomic regions targeted based on the miRNA associated with drought stress and sequenced to identify SNPs in a panel of 8 genotypes.

NL- Not accurately localized; *Nil- No Adjuvant was used; Adjuvant if mentioned 400 mM TMAC (Tetramethyl ammonium chloride) at the final concentration of 10-40 mM in PCR reaction; #Accession number of the chromosome refers to version of 1 corresponding rice chromosome sequence.

4.4. Identification of SNPs in the targeted region associated with the trait

Our current approach enabled us to target a set of 300 genomic regions (Promoters, CDS, Introns, and UTRs) to dissect the allelic variation to the single nucleotide level. A total of 150 target regions were successfully amplified from a panel of eight rice genotypes with well characterized phenotype with reference to drought tolerance (Table 4.3, 4.4, 4.5). This set of genotypes used for identification of SNPs can be broadly categorized into two panels, drought tolerant and drought susceptible. The members falling in to the drought tolerant panel includes CT9993, Azucena, Nootripathu, N22B, Nerica-1 and drought susceptible panel is represented by IR62266, IR64 and IR20. These includes parental lines two populations viz. CT9993 X IR62266 and Azucena X IR64, which are widely used for mapping drought related QTLs. A total of 150 regions were amplified and sequence characterized from eight genotypes (Genbank Acc. EF556551- EF558346 and EU868928-EU869180, Table 4.6), and analyzed for identification of SNPs using standard sequence processing programs coupled with in-house developed Perl scripts in polybayes approach (Marth *et al*., 1999; Marth, 2003).

S.No	Genotype Name	No. of bases analyzed	Accession numbers of Genbank submission from the present study
1	CT9993	1,28,959	EF557001 to EF557223, EU868929 to EU868941, EU869038 to EU869055
2	IR62266	1,46,771	EF557460 to EF557708, EU868960 to EU868975, EU869079 to EU869100
3	Azucena	1,05,580	EF556797 to EF557000 , EU868928, EU869034 to EU869037
4	IR64	1,32,967	EF557709 to EF557926 , EU868976 to EU868990, EU869101 to EU869123
5	Nootripathu	1,32,433	EF558115 to EF558340, EU869006 to EU869017, EU869139 to EU869158
6	IR20	1,38,067	EF557225 to EF557456, EU868942 to EU868959, EU869056 to EU869078

7	N22B	1,04,876	EF557930 to EF558114, EF556551 to EF556552, EU868991 to EU869005, EU869124 to EU869138
8	Nerica-1	1,41,798	EF556657 to EF556796, EU869018 to EU869033, EU869159 to EU869180
9	DRRH1	497	EF557224
10	IR42253	1,611	EF557459
11	Mahamaya	712	EF557927
12	MM125	1,406	EF557929
14	Poornima	989	EF558342
15	Safri	1,160	EF558344
16	Swarna	1,574	EF558346

Table 4.6: Summary of the sequence analysis and genbank accession numbers of each genotype.

Screening of more than one million bases at the targeted sites enabled us to capture the informative sites and categorized based on the type of change observed (Fig. 4.6). All the SNPs were screened with a quality cut off parameter using Bayesian probabilistic algorithm and SNPs having probability greater than 0.99 were considered as SNP and rest of them were eliminated. A total of 601 transitions, 461 transversions and 548 Indels were identified (Fig. 4.6) of which 318 are identified in coding regions, 8 are in intronic regions and 69 are in UTRs. Majority of the SNPs (1203) are found in promoter regions as the major focus was on 5' upstream region of genes, and this includes SNPs found in the regulatory regions of the promoters and surrogate SNPs as well. Most of the variations were observed to be a single nucleotide substitution or indels besides a few (50) multi-nucleotide indels (MNPs). Amongst single nucleotide changes majority of them are A/G (308) and T/C (289) changes. A total of 143 (T/A), 109 (A/C), 102 (C/G) and 110 (T/G) variations were observed (Fig. 4.7). Further, analysis of SNPs between specific parental lines such as CT9993 and IR62266, Azucena and IR64 helped us to identify specific variation between the parental lines used in mapping QTLs associated with drought tolerance. Table 4.7a shows the number of bases considered for comparison between the genotypes to identify the SNPs. A total of 587 SNPs were identified between CT9993 and IR62266 and 495 SNPs were identified between Azucena and IR64. These SNPs can be utilized to convert them into molecular markers to genotype the available population. Interestingly, comparison of drought tolerant Nerica-1 with drought susceptible IR64 revealed 681 SNPs and with that of IR62266 revealed 706 SNPs. The comparison of Nerica-1 with drought tolerant Azucena revealed only 336 SNPs and with that of CT9993 revealed only 496. The rate of polymorphism in these

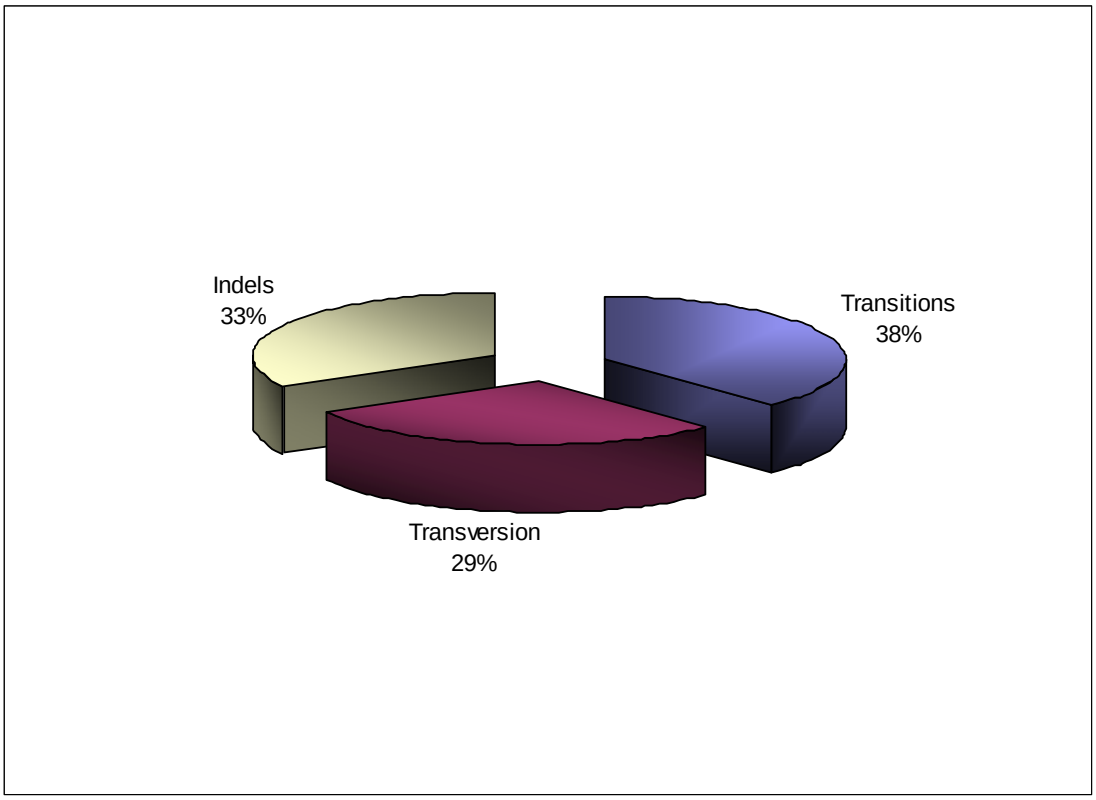


Figure 4.6: Categorization of SNPs based on the type of change.

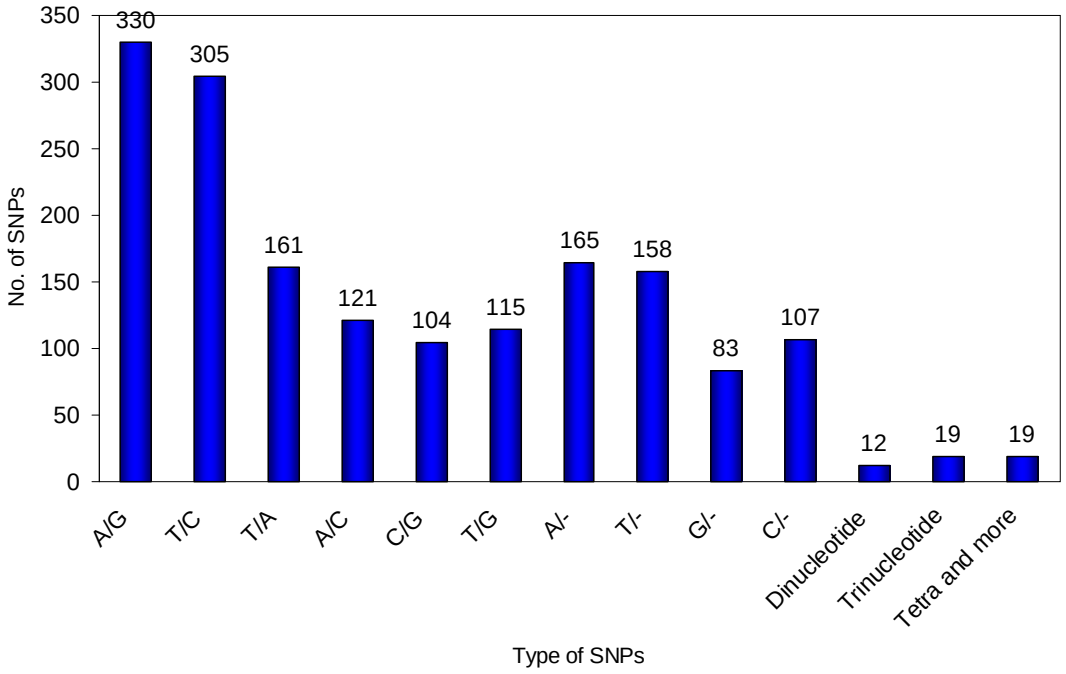


Figure 4.7: Classification of SNPs identified at the target loci based on the type of observed change.

cases is relatively low when compared to any of the other susceptible genotypes. Similarly, comparison among the drought susceptible varieties revealed relatively less number of SNPs. The number SNPs and rate of polymorphism between the genotypes is shown in table 4.7b. Rate of polymorphism (ROP) between genotypes ranged from 0.313 to 0.928. Highest rate of polymorphism was observed between Nipponbare and Nootripathu which is basically a land race rich in flavonoids. Significantly, a low rate of polymorphism was observed between Nerica and Azucena which are known to be drought tolerant cultivars.

	Nippon-bare	CT9993	IR62266	Azucena	IR64	Nootri.	IR20	N22B
CT9993	93902							
IR62266	98130	110607						
Azucena	72089	104696	108952					
IR64	93208	110907	109313	106081				
Nootripathu	93245	108479	110243	103224	107892			
IR20	94396	108579	109548	105432	108123	107589		
N22B	78837	105313	106308	95672	102851	104553	104950	
Nerica-1	99675	112487	113814	108055	111837	110090	110526	108413

Table 4.7a: Number of bases compared among the genotypes with in the panel

	Nippon Bare	CT9993	IR62266	Azucena	IR64	Nootri.	IR20	N22B
CT9993	0.493 (463)							
IR62266	0.833 (818)	0.541 (599)						
Azucena	0.375 (270)	0.367 (385)	0.404 (441)					
IR64	0.830 (774)	0.469 (521)	0.351 (384)	0.466 (495)				
Nootripathu	0.928 (865)	0.532 (578)	0.386 (426)	0.527 (544)	0.393 (425)			
IR20	0.904 (854)	0.594 (645)	0.354 (388)	0.480 (507)	0.362 (392)	0.367 (395)		
N22B	0.800 (631)	0.552 (582)	0.437 (465)	0.422 (404)	0.552 (568)	0.427 (447)	0.432 (454)	
Nerica-1	0.467 (466)	0.451 (508)	0.637 (725)	0.313 (339)	0.631 (706)	0.689 (759)	0.675 (747)	0.606 (657)

Table 4.7b: Number of SNPs and rate of polymorphism (ROP) between the genotypes calculated based on the target regions analyzed. ROP was calculated as ROP=100*number of SNPs identified/number of bases analyzed between the genotypes. The value in the bracket indicates the number of SNPs between the respective genotypes.

The high quality SNPs were confirmed after manual inspection of the alignment and submitted to dbSNP division of NCBI under the handle ARR-VBREDDY (dbSNP ss#69373227 to 69374767; 105107054 to 105107119; 105111345 to 105111433) and released to the public domain as part of the SNP build 128 and 130 of dbSNP. Though, the analysis revealed more than 2700 SNPs in the targeted regions, the application Bayesian probabilistic algorithm with appropriate cut-off (0.99 probability) limited us to identify 1699 SNPs (dbSNP ss#69373227 to 69374767; 105107054 to 105107119; 105111345 to 105111433). The dbSNP submission file was prepared by using custom developed Perl script using SNP database in the backend. Figure 4.8 shows the format of dbSNP submission file generated automatically using the custom Perl scripts using extracted SNPs from the analysis. The details of the each of the SNP and their flanking regions along with the submission files are provided as soft copy (Table S1) in the attached CD-ROM and representative image of reference clusters generated from the submitted SNPs under the handle ARR-VBREDDY is shown in Figure 4.9. A full list of SNPs along with the reference cluster details are presented in the attached CD-ROM (Table S2), which can be also accessible from dbSNP division of NCBI. The sequences clustered and the total number of SNPs identified in each of the target region is summarized in table 4.8 along with the genbank sequence numbers and dbSNP accessions numbers of the corresponding candidate loci deposited at Genbank division of NCBI.

```
TYPE:          CONT
HANDLE:        ARR-VBREDDY
NAME:          Arjula.R.Reddy and V. B. Reddy Lachagari
FAX:           +9140-23010120
TEL:           +9140-23134553
EMAIL:         arjulsl@uohyd.ernet.in,vijaylg2002@yahoo.com
LAB:           Molecular Genetics and Functional Genomics Laboratory
INST:          University of Hyderabad
ADDR:          Prof.Arjula.R.Reddy, Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad, A.P., India, 500046
||
TYPE:          PUB
HANDLE:        ARR-VBREDDY
TITLE:
Identification and Validation of Single Nucleotide Polymorphisms in Candidate Genes of Drought Tolerance in Rice
AUTHORS:
V.B.Reddy,Lachagari; Markandeya,Gorantla; Prasanna Kumar,J.;Venkat Reddy,B.;Reddy, A.R.
YEAR:          2008
STATUS:        1
||
TYPE:          METHOD
HANDLE:        ARR-VBREDDY
ID:            VBREDDY-RESEQ8-24
METHOD_CLASS:          Sequence
SEQ_BOTH_STRANDS:      YES
TEMPLATE_TYPE:         DIPLOID
MULT_PCR_AMPLIFICATION:        YES
MULT_CLONES_TESTED:            NO
METHOD:
PCR reactions were performed with genomic DNA and products were analysed by DNA sequencing.
PARAMETER:
Template:  50 ng genomic DNA
Primer:    each 0.25 uM
dNTPs:     each 200 uM
TMAC:      20-40 mM
PCR Buffer: 5 ul (10X), Mg 2+ 1.5 mM, Taq DNA Polymerase: 2.0 U
||
TYPE:          POPULATION
HANDLE:        ARR-VBREDDY
ID:            8PLDrSNP
POPULATION:
This file contains SNPs identified between 6 (CT9993, IR62266, Azucena, IR64, Nootripathu, IR20) Parental lines, which are widely used for identification drought responsive QTLs in rice, and 1 upland indica cultivar N22B and Nerica1 (O. glaberrima X O. sativa)
PHENOTYPE:             Drought Tolerant and susceptible genotypes are used in the study
||
TYPE:          SNPASSAY
HANDLE:        ARR-VBREDDY
BATCH:         VBREDDY-ReSeq8-24-BATCH2
MOLTYPE:       Genomic
METHOD:        VBREDDY-RESEQ8-24
SAMPLESIZE:            18
ORGANISM: Oryza sativa
||
SNP:           ARR-VBREDDY|OsU-AP003834-90708
SYNONYM:       P319-1794,AP003834-90708
SAMPLESIZE:            18
COMMENT:
               Forward PCR primer: GCCATTTAGGAGACCTTTTGGC
               Reverse PCR primer: AAGTACGACGCCAATCGATGC
               Region: MIR
               note: MIR398b
LENGTH:        ?
5'_FLANK:
CCGCGCCATTTAGGAGACCTTTTGGCAAGAGAAAATAATCAACATATACATAATCAACCAGAGAGTCCCGGCAGGGGCGACCTGAGAACACACGAAAGAC
OBSERVED: C/-
3'_FLANK:
GACCGCCTCATCACCGTGTGTTCCCAGCTCGCCCCTGTAGGAACTCCCTCGCGTTCCATCTCTCTGTCTCCATCGCCGCTGCCGATGCCGATAACGATGC
||
SNP:           ARR-VBREDDY|OsU-AP003834-90731
SYNONYM:       P319-1795,AP003834-90731
SAMPLESIZE:            18
COMMENT:
               Forward PCR primer: GCCATTTAGGAGACCTTTTGGC
               Reverse PCR primer: AAGTACGACGCCAATCGATGC
               Region: MIR
               note: MIR398b
LENGTH:        ?
5'_FLANK:
GGCAAGAGAAAATAATCAACATATACATAATCAACCAGAGAGTCCCGGCAGGGGCGACCTGAGAACACACGAAAGACCAGACCGCCTCATCACCGTGTGT
OBSERVED: T/-
3'_FLANK:
CCAGCTCGCCCCTGTAGGAACTCCCTCGCGTTCCATCTCTCTGTCTCCATCGCCGCTGCCGATGCCGATAACGATGCCGATTCCGATTCGTTCCCTCCCC
||
```

Figure 4.8: Format of dbSNP submission generated automatically using custom perl scripts

refSNP ID: rs54411397	
Organism:	rice (*Oryza sativa*)
Molecule Type:	Genomic
Created/Updated in build:	128/128
Map to Genome Build:	4.1

Allele	
Variation Class:	SNP: single nucleotide polymorphism
RefSNP Alleles:	C/T
Ancestral Allele:	Not available
Clinical Association:	unknown

SNP Details are organized in the following sections:

Submission | Fasta | Resource | GeneView | Map | Diversity | Validation

Submitter records for this RefSNP Cluster

The submission **ss69373824** has the longest flanking sequence of all cluster members and was used to instantiate sequence for **rs54411397** during BLAST analysis for the current build.

NCBI Assay ID	Handle|Submitter ID	Validation Status	ss to rs Orientation /Strand	Alleles	5' Near Seq 30 bp	3' Near Seq 30 bp	Entry Date	Update Date	Build Added	Molecule Type
ss69373824	ARR-VBREDDY|Os04-32049989		fwd/B	C/T	ccaaagccaccgccttctcctcctcaaaaa	atccgccgcaatcaaaatctcccgcaaatc	04/04/07	04/04/07	128	Genomic

Fasta sequence (Legend)

>gnl|dbSNP|rs54411397|allelePos=101|totalLen=201|taxid=4530|snpclass=1|alleles='C/T'|mol=Genomic|build=128

GCCCACGATC GCTTAAAACT TCCCCCCAAA ACCGCCCAAT CCACCAAAAC CCTACCTAAG
CGAAATCCGC CCAAAGCCAC CGCCTTCTCC TCCTCAAAAA
Y
ATCCGCCGCA ATCAAAATCT CCCGCAAATC GCAACAACCC TCTCACACCT CCCCGAAACA
AAACATCTCA CAAAAGCATT CCATCCATTT GAACCAAAAC

GeneView

GeneView via analysis of contig annotation: Os04g0629300 Os04g0629300
View variations for gene: Include clinically associated: in gene region | cSNP | has frequency | double hit Go

GeneView via analysis of contig annotation: Os04g0629400 Os04g0629400
View variations for gene: Include clinically associated: in gene region | cSNP | has frequency | double hit Go

Group Label	Contig->mRNA	Gene Model (contig mRNA transcript) Color Legend
reference	NC_008397->NM_001060501 sv function	
reference	NC_008397->NM_001060502 sv function	

Group label	Contig-->mRNA-->Protein	Contig position	mRNA orientation	mRNA pos	Function	dbSNP allele	Protein residue	Codon pos	Amino acid pos
reference	NC_008397->NM_001060501->NP_001053966	32049989:32049990	reverse		intron				
reference	NC_008397->NM_001060502->NP_001053967	32049989:32049990	reverse		3' near gene				

GeneView: no link established by BLAST analysis of mRNA sequences

Integrated Maps:

NCBI MapViewer: rs54411397 maps exactly once on NCBI rice chromosome 4

Chromosome	Contig accession	Contig position	Chromosome position	Hit orientation	Contig Allele	Assembly Type	Group label	Contig label	Neighbor SNP	SNP_flank position
4	NC_008397.1	32049989..32049990	32049989..32049990	plus	TG	ref_assembly	reference	reference	view	100

Figure 4.9: Details of reference cluster generated from the SNPs submitted under ARR-VBREDDY handle at NCBI.

P. No.	Description of the sequenced region	Type	Genbank Accession numbers of the sequenced regions	No. of SNPs	NCBI ss Numbers of identified SNPs
100	Ribosomal 6 kinase	P	EF558176, EF557985, EF557986, EF557293, EF557292, EF557291, EF557290, EF556618, EF557984, EF557056, EF556619, EF558177, EF558178, EF558179, EF556862, EF556861, EF556860, EF556859, EF556620, EF557530, EF557768, EF556621, EF557769, EF557529, EF557528, EF557770, EF557531	2	69374377 to 69374378
102	MAP kinase kinase	P	EF557771, EF557057, EF557058, EF556622, EF556863, EF557532, EF558180, EF557294, EF556623	1	69373482
104	M_Ring finger protein	P	EF556624, EF557059, EF556864	3	69373395 to 69373397
105	M_Ring finger GCD	G	EF556553, EF556554, EF557709, EF557001, EF557002, EF556797, EF557930, EF557460, EF557461, EF558116, EF557226, EF557225, EF558115	4	69373398 to 69373401
108	M_Gene with MITE	G	EF557931, EF557710, EF558117, EF557003, EF557228, EF557463, EF557462, EF557004, EF557005, EF558118, EF556799, EF556798, EF557227, EF556555, EF556556	3	69374653 to 69374655
109	M_Unknown gene	G	EF556800, EF557006, EF557007, EF558120, EF558119, EF556557, EF556801, EF557230, EF557465, EF557711, EF557229, EF556558, EF557712, EF557932, EF557933, EF557464	2	69373785 to 69373786
110	M_Unknown gene	P	EF556866, EF558182, EF558181, EF557534, EF557533, EF556626, EF557987, EF557295, EF556865, EF557060, EF557061, EF556625, EF557772	4	69373781 to 69373784
111	Mr_NL38C07	G	EF558122, EF557466, EF557935, EF557934, EF556802, EF557231, EF557232, EF556559, EF558121, EF557467, EF557714, EF557008,, EF557009, EF557713, EF556803, EF556560	14	69374593 to 69374606
112	Mr_NL38C07_Pro	P	EF557988, EF557297, EF557773, EF556627, EF557535, EF557296, EF556867, EF557774,	24	69374607 to 69374630
114	Thioredoxin	P	EF556628, EF556629, EF558184, EF558183, EF557775, EF557063, EF557062, EF557990, EF557536, EF557537, EF557989, EF556868, EF557776, EF557298, EF557299, EF556869	23	69374195 to 69374217
115	Ferrudoxin	G	EF557468, EF557715, EF556561, EF557233, EF558123, EF557469, EF556804, EF557716, EF557936, EF557010	1	69374392
116	Ferrudoxin	P	EF556871, EF556870, EF558185, EF557778, EF557992, EF557991, EF557777, EF558186, EF557539, EF557538, EF557064, EF557065, EF557300, EF556630, EF557301	17	69374393 to 69374409
118	M_Jasmonic acid induced gene	P	EF557540, EF557303, EF557302, EF558187, EF557541, EF557543, EF557542, EF557304,	30	69373748 to

			EF557066, EF556631		69373777
119	M_RD22 LE	G	EF557472, EF557471, EF557470, EF557011, EF557235, EF557717, EF556564, EF557234, EF558125, EF556562, EF557237, EF557236, EF558124, EF557718, EF556563, EF557938, EF557937,	19	69373894 to 69373912
120	M_PSII10KDa LE	G	EF557476, EF556806, EF556805, EF558126, EF558127, EF558128, EF556565, EF556566, EF556567, EF557941, EF557940, EF557939, EF557240, EF557474, EF557475, EF557473, EF557241, EF557239, EF557012, EF557238, EF557720, EF557013, EF557719	5	69373264 to 69373268
121	M_PSII10KDa	P	EF557779, EF556632, EF557782, EF557993, EF557545, EF557780, EF557068, EF557544, EF557547, EF557067, EF557546, EF557070, EF557781, EF556633, EF557994, EF557069, EF557306, EF557305, EF556634, EF556635, EF557308, EF558188, EF558189, EF558190, EF557307	7	69373254, 69373257 to 69373259, 69373261 to 69373263
123	M_Cold Induced gene	P	EF557071, EF557995, EF556636, EF558191, EF556872, EF557548, EF557309, EF557783	6	69373962 to 69373967
124	CBF4	P	EF557549, EF557311, EF556874, EF557996, EF556873, EF557073, EF557072, EF557310, EF557997, EF557784, EF556637, EF558192, EF557550, EF556638, EF558193, EF557785	35	69374018 to 69374032, 69374050 to 69374069
127	bZIP E2	G	EF556808, EF557014, EF557478, EF557477, EF556569, EF557721, EF557016, EF557015, EF556807, EF556570, EF558129, EF557243, EF556568, EF558130, EF557942, EF557242	0	Nil
129	M_ AP2 family gene	P	EF556875, EF556876, EF557551, EF556640, EF557074, EF557552, EF558194, EF557313, EF557786, EF557998, EF557999, EF557787, EF557075, EF556639, EF558195, EF557312	5	69373621 to 69373625
131	M_dnaJ	P	EF557790, EF556641, EF557553, EF558196, EF557314, EF557788, EF557789, EF557076, EF557315, EF558000, EF556877	20	69373227 to 69373246
133	M_DREB2	P	EF558197, EF558198, EF557555, EF556642, EF557791, EF557554, EF558001, EF557316, EF556879, EF556878	3	69374517 to 69374519
135	M_PCF3	P	EF557556, EF557317, EF557318, EF557557, EF558199, EF558200, EF557792, EF557793, EF557078, EF556644, EF557077, EF556643	3	69374675 to 69374677
137	M_MADS	P	EF558003, EF558002, EF557321, EF557082, EF557080, EF557796, EF558004, EF558202, EF557081, EF558201, EF557560, EF557561, EF557319, EF557795, EF557794, EF557320, EF556647, EF558005, EF557079, EF557322, EF557559, EF557558, EF556646, EF556645, EF556880, EF556881, EF556882, EF556883	4	69374107

139	M_Catalase	G	EF557479, EF557244, EF557245, EF557018, EF557723, EF556809, EF557017, EF557943, EF557944, EF557480, EF557722, EF556810	2	69374176 to 69374177
141	M_NAM	G	EF557247, EF557019, EF556811, EF556812, EF557246, EF557724, EF556571, EF556573, EF557020, EF557945, EF558131, EF557481, EF557482, EF556572	3	69374506 to 69374508
142	M_NAM	P	EF557563, EF557562, EF558204, EF556884, EF556648, EF556649, EF558203, EF557798, EF557797, EF557083, EF558006, EF558007, EF557084, EF557323, EF557324	11	69374495 to 69374505
144	M_ Glycine rich zinc finger protein	P	EF557085, EF558009, EF556650, EF558205, EF557086, EF556885, EF556886, EF558206, EF556651, EF558008, EF557564, EF557326, EF557800, EF557799, EF557325, EF557565	17	69373465 to 69373481
145	M_RING-H2 finger ZFP LE	G	EF557248, EF557725, EF557726, EF557483, EF556814, EF557249, EF557021, EF557947, EF557946, EF558133, EF558132, EF557484, EF556813	1	69373724
146	M_RING-H2 finger ZFP	P	EF557087, EF558209, EF557088, EF557804, EF557803, EF557802, EF557801, EF556889, EF556888, EF558207, EF557089, EF558208, EF557328, EF556654, EF556652, EF556653, EF557329, EF557566, EF557327, EF558010, EF558011, EF556887	23	69373725 to 69373747
147	M_HSP90 LE	G	EF556816, EF558135, EF557485, EF558134, EF557022, EF556815, EF557727, EF557948, EF557949, EF557251, EF557023, EF557250, EF557486, EF557728, EF556574	4	69373602 to 69373605
148	M_HSP90	P	EF557568, EF557330, EF556655, EF556656, EF557331, EF558012, EF558013, EF556890, EF557567, EF557090, EF557806, EF557805, EF556891, EF557091, EF558210	9	69373593 to 69373601
149	M_Unknown gene homologous to CB965622	G	EF557730, EF558345, EF557729, EF557950, EF557253, EF557252, EF556575, EF557025, EF557024, EF558136, EF558137, EF556576, EF557951, EF557487, EF556817, EF556818, EF557457, EF557488	6	69374664 to 69374669
150	M_Unknown gene homologous to CB965622	P	EF556657, EF556658, EF557092, EF556893, EF557807, EF558211, EF558213, EF558212, EF557093, EF556892, EF557808, EF557570, EF557332, EF557333, EF558015, EF557569, EF558014	3	69374656 to 69374658
151	M_Unknown homologous to CB966119	G	EF557952, EF557026, EF557731, EF556578, EF557953, EF556577, EF558139, EF556819, EF557254, EF557490, EF558138, EF557489	1	69373515
152	M_Unknown homologous to CB966119	P	EF557810, EF557334, EF558016, EF557335, EF556659, EF557572, EF557571, EF557809, EF556660, EF558214, EF556894, EF556895, EF557094	14	69373528 to 69373541
154	M_Gigantia like gene	P	EF558017	7	69373426 to 69373432

155	M_RS6 kinase GCD	G	EF556821, EF557491, EF556820, EF557028, EF556580, EF557954, EF556579, EF557255, EF557733, EF557492, EF557027, EF557256, EF557955, EF558140, EF558141, EF557732	6	69374112 to 69374117
156	M_RS6 kinase GCD	U	EF556784, EF557448, EF557701, EF556785, EF557447, EF557217, EF556991, EF557700, EF556992, EF557218, EF558335, EF558334, EF557920, EF557919	4	69374108 to 69374111
157	M_Water channel GCD	G	EF558142, EF556823, EF557734, EF557956, EF557030, EF556822, EF557029, EF557258, EF557257, EF557493, EF556582, EF556581	55	69373704 to 69373723, 69374218 to 69374239, 69374266 to 69374278
158	M_Water channel	G	EF557494, EF557957, EF557495, EF557736, EF557735, EF556824, EF558143, EF557958, EF557031, EF556825, EF558144, EF557260, EF557259, EF556583, EF556584, EF557032	0	Nil
159	DREB1B	P	EF558020, EF558218, EF556899, EF558019, EF557337, EF558217, EF557099, EF557336, EF556551, EF557100, EF556662, EF557811, EF557574, EF556663, EF557575	11	69374541 to 69374543, 69374552, 69374556 to 69374562
162	Aquaporin	P	EF556901, EF558219, EF558220, EF557102, EF556900, EF557339, EF556665, EF557338, EF557101, EF557812, EF556664, EF557576, EF558022, EF558021, EF557577	23	69374240 to 69374262
163	Aquaporin	G	EF556585, EF557737, EF558145, EF557033, EF557034, EF558146, EF557959, EF557261, EF556586, EF557960, EF556827, EF557262, EF556826, EF557496, EF557497	6	69374263 to 69374265, 69374279 to 69374281
164	Aquaporin LE	G	EF556829, EF558147, EF556828, EF557738, EF557961, EF557498, EF556587, EF557263, EF557035	19	69374282 to 69374300
167	Phenylalanine ammonia-lyase	P	EF557813, EF556903, EF556902, EF558023, EF558221, EF556666, EF557578, EF557340, EF557104, EF557103, EF556667	12	69373551 to 69373562
168	Anthocyanin reductase	P	EF557105, EF556904	0	Nil
169	Coumaryl CoA Ligase	P	EF558223, EF558024, EF558025, EF557341, EF556905, EF556906, EF557107, EF557342, EF557579, EF558222, EF557814, EF557815, EF557106, EF556669, EF556668	65	69374440 to 69374494
170	Pathogen responsive (PR) protein	P	EF557580, EF558224, EF557343, EF556670, EF557108, EF557816, EF558026, EF556907	9	69374009 to 69374017

171	Disease responsive gene	P	EF557581, EF556671, EF556672, EF557817, EF558027, EF557109, EF557345, EF558225, EF556909, EF556908, EF557344	7	69373276 to 69373282
172	Wound induced gene	P	EF557346, EF557347, EF557111, EF558227, EF558226, EF558028, EF557110, EF558029, EF557583, EF556910, EF557582, EF556674, EF557818, EF556673, EF557819	15	69373287 to 69373301
173	Thaumatin	P	EF557820, EF557821, EF556912, EF556911, EF557112, EF557585, EF556675, EF556676, EF557348, EF557349, EF558228, EF558229, EF557113, EF557584, EF558031, EF556913, EF558030	22	69374140 to 69374161
177	S-adenosyl-methionine Synthase	P	EF557822, EF557350, EF558230	2	69373882 to 69373883
178	APX TL29	P	EF557587, EF557352, EF557351, EF556914, EF556678, EF558231, EF558232, EF557586, EF556677, EF558033, EF558032, EF557115, EF557823, EF557824, EF557114	2	69373795 to 69373796
179	APX	P	EF557353, EF557825, EF556679, EF557116, EF557588, EF558034, EF558233	21	69373658 to 69373678
180	HSP82	P	EF557827, EF557826, EF557354, EF557355, EF556915, EF558036, EF556680, EF557118, EF556681, EF558234, EF556916, EF557117, EF557589, EF558235, EF558035	11	69374530 to 69374540
181	HSP60	P	EF558236, EF558237, EF557120, EF557119, EF557356, EF556683, EF558037, EF556682, EF557828, EF557357, EF557590, EF557591, EF556917, EF558038	32	69373977 to 69374008
182	Heat shock Transcription factor	P	EF557592, EF558239, EF557829, EF557830, EF556684, EF557593, EF558238, EF556918, EF557121	7	69374169 to 69374175
183	CBF1	P	EF556685, EF556686, EF556919, EF558240, EF558039, EF557831, EF557122, EF557594, EF557358	13	69374033 to 69374045
185	Mn superoxide dismutase	P	EF557123, EF557832, EF556920, EF556687	0	Nil
187	Dehydrin	P	EF556922, EF558241, EF557595, EF557359, EF556688, EF556689, EF556923, EF556921, EF557597, EF558242, EF558243, EF557360, EF557124, EF557361, EF557125, EF557126, EF557596	10	69373247 to 69373253, 69373255 to 69373256, 69373260
189	Helicase	P	EF556926, EF558043, EF558042, EF558041, EF558040, EF557363, EF557362, EF556925, EF557129, EF556927, EF556691, EF556690, EF556692, EF557835, EF557834, EF557127, EF557128, EF557833, EF556924, EF558245, EF558246, EF557601, EF557600, EF557599, EF557598, EF558244	13	69373819 to 69373831

190	CBF3	P	EF557132, EF558247, EF557364, EF556693, EF557837, EF557836, EF558248, EF557131, EF557365, EF557602, EF557603, EF557604, EF558044, EF557130, EF556929, EF556930, EF556928	10	69374544 to 69374551, 69374553, 69374555
192	CBF3	P	EF558045, EF558250, EF558249, EF557605, EF556932, EF557133, EF558046, EF557366, EF556931, EF556695, EF556694	7	69374571 to 69374577
196	Submergence induced protein	P	EF557606, EF558048, EF557838, EF556697, EF556696, EF558251, EF558252, EF557607, EF557839, EF556933, EF557367, EF556934, EF557135, EF557134, EF557368, EF558047	6	69373615 to 69373620
198	CDPK	P	EF557608, EF558254, EF556935, EF556936, EF558253, EF557609, EF557840, EF557136, EF558049	4	69373791 to 69373794
199	SAM decarboxylase	P	EF556699, EF557370, EF557842, EF557841, EF557611, EF558256, EF557369, EF558050, EF556937, EF556938, EF557137, EF557138, EF558255, EF556698, EF557610,	9	69373542 to 69373550
200	WRKY Transcription factor	P	EF558052, EF557371, EF557843, EF556939, EF556940, EF557612, EF557139, EF557613, EF556700, EF558257, EF558258, EF557140, EF556701, EF558051	22	69373797 to 69373818
201	26S Proteosomal protein	P	EF558260, EF558259, EF557845, EF556941, EF556942, EF556703, EF556702, EF557615, EF557844, EF557614, EF558054, EF557373, EF557141, EF557142, EF557372, EF558053	13	69374379 to 69374391
202	Wound induced protein	P	EF557846, EF557374, EF557143, EF557144, EF558262, EF556943, EF556705, EF556704, EF558261, EF558056, EF558055, EF557616	22	69374410 to 69374431
203	Signal recognition particle	P	EF557375, EF556706, EF558263	7	69374634 to 69374640
205	Cytochrome P450	P	EF557145, EF557376, EF558057, EF556944	37	69373348 to 69373384
206	Serine threonine kinase	P	EF557147, EF557617, EF556707, EF556946, EF556945, EF558059, EF557146, EF557618, EF558264, EF556708, EF557378, EF557377, EF558058	4	69373606 to 69373609
207	Enolase	P	EF556709, EF557380, EF557379, EF556947, EF556948, EF557619, EF558265, EF557148, EF558266, EF556710, EF557149, EF558061, EF558060, EF557848, EF557847	10	69373626 to 69373635
208	R2R3 Myb	P	EF556711, EF557849, EF557620, EF558267, EF557151, EF556949, EF557150, EF557381	15	69373852 to 69373866
209	R2R3 Myb LE	G	EF558148, EF557499, EF557962, EF557265, EF557963, EF557264, EF557037, EF557739, EF557036	15	69373867 to 69373881
210	MYB1	P	EF558063, EF557851, EF558062, EF556950, EF556712, EF557850	3	69374752 to 69374754

212	RACD GTP binding protein	P	EF557853, EF558269, EF557383, EF557152, EF557382, EF557852, EF557621, EF558268, EF557622, EF558064, EF556714, EF556952, EF556713, EF556951	18	69373575 to 69373592
213	RAB2 GTP binding protein	P	EF556715	0	Nil
214	RIC1 GTP binding protein	P	EF557624, EF558065, EF558271, EF557153, EF558270, EF556953, EF557623, EF556716, EF557854, EF556717, EF557384, EF557385, EF557855	18	69373330 to 69373347
216	Transcription regulating protein	P	EF557627, EF556955, EF556954, EF557154, EF557155, EF558067, EF558066, EF557626, EF557625	7	69373932 to 69373938
218	Proline rich protein	P	EF556719, EF558068, EF557857, EF557628, EF557386, EF558272, EF557856, EF557157, EF557156, EF557629, EF556718, EF557387	6	69373913 to 69373918
221	Dehydrin	G	EF557741, EF556589, EF556588, EF558149, EF557038, EF557267, EF558150, EF557500, EF557266, EF556830, EF556831, EF557740, EF557039	3	69373778 to 69373780
224	Cytokinin oxidase	P	EF557388, EF557160, EF557630, EF557631, EF557389, EF557158, EF557159, EF557859, EF557858, EF558069, EF556720, EF558070, EF558273, EF556721	7	69373302 to 69373308
225	M_ Unknown gene homologous to CB964527	P	EF558072, EF556723, EF557161, EF556722, EF556956, EF558071, EF558275, EF558274, EF556957, EF557632, EF557860, EF557633, EF557391, EF557390, EF557861	5	69374659 to 69374663
226	M_ Unknown gene homologous to CB966078	P	EF557392, EF556724, EF557634, EF557862, EF558073, EF558276, EF557162	12	69373516 to 69373527
227	Cystein protease	P	EF557864, EF558277, EF558278, EF557635, EF557636, EF557163, EF556725, EF557863	21	69373444 to 69373464
229	Disease induced protein	P	EF557865, EF557164, EF557637, EF557394, EF557393, EF556726, EF557638, EF556727	3	69373433 to 69373435
230	Disease induced protein	G	EF558152, EF558151, EF557742, EF557743, EF557744, EF557040, EF557041, EF556832, EF557965, EF557964, EF556592, EF557269, EF557270, EF556591, EF557502, EF557501, EF557268, EF556590, EF558153, EF557503	3	69373436 to 69373438
231	Disease induced protein LE	G	EF557272, EF557271, EF556833, EF556834, EF557504, EF556593, EF556594, EF557042, EF557505, EF557043, EF557506, EF558154, EF557745, EF557746	5	69373439 to 69373443
233	Disease induced protein 19	P	EF557639, EF557395, EF556728, EF557166, EF557866, EF557640, EF558279, EF558280, EF557165	7	69373508 to 69373514
234	Low temperature induced protein (LT16)	P	EF557641, EF556730, EF558281, EF557867, EF558074, EF556729, EF557396	2	69374355 to 69374356
235	Water stress induced gene 18 LE	G	EF556596, EF556595, EF557967, EF557966, EF558155, EF557747, EF556836, EF557507,	3	69373385 to

			EF557273, EF556835		69373387
236	Water stress induced gene 18	P	EF556731, EF556732, EF557397, EF557642, EF556958, EF557167	1	69373388
237	AP2	P	EF556734, EF556733, EF557168, EF557643, EF558282	3	69373563 to 69373565
239	ECP40	P	EF558075, EF557398, EF556735, EF557169, EF557644	6	69374735 to 69374740
241	LEA protein	P	EF557171, EF558076, EF558284, EF557869, EF558283, EF557170, EF557399, EF557868, EF556737, EF556736, EF557645, EF557646	9	69373968 to 69373976
243	DREB1D	P	EF557871, EF557648, EF556552, EF557647, EF557870	0	Nil
245	Aldehyde dehydrogenase	P	EF557400, EF557872, EF557172, EF557649, EF558077, EF556959, EF558285	10	69374520 to 69374529
246	Drought responsive protein homologous to Os.57533	P	EF557173, EF557873, EF558078, EF558286, EF557401, EF556738, EF557650	4	69374513 to 69374516
248	DREB1A similar to Os.57517	P	EF557874, EF557402, EF558079, EF557651, EF557174, EF557652	9	69374554, 69374563 to 69374570
25	DREB	U	EF556994, EF557219, EF557449, EF558110, EF557702, EF556993, EF556786, EF557450	4	69374046 to 69374049
252	EF hand calcium-binding protein similar to Os.75123 and Os.69287	G	EF556597, EF557748, EF557044, EF557508, EF557275, EF557045, EF557046, EF557276, EF557750, EF558158, EF558157, EF558156, EF557749, EF557968, EF557509, EF557274, EF556599, EF556598, EF557510	2	69373506 to 69373507
253	Early drought induced protein similar to Os.13968	P	EF558080, EF557875, EF557654, EF557403, EF557653, EF558287, EF557175	0	Nil
254	Early drought induced protein similar to Os.13968 LE	G	EF556601, EF557047, EF558159, EF557278, EF558160, EF556602, EF557511, EF557048, EF556600, EF557512, EF556603, EF557969, EF557970, EF557971, EF557277, EF557752, EF557753, EF556837, EF556838, EF557751	10	69373884 to 69373893
255	Drought-induced hydrophobic protein similar to Os.46402	G	EF557516, EF556604, EF556605, EF557050, EF557755, EF557049, EF558161, EF556606, EF557754, EF558163, EF557515, EF557514, EF557513, EF557279, EF557973, EF557972, EF558162, EF556840, EF556841, EF557280, EF556839, EF556842	20	69374357 to 69374376
256	Drought-induced hydrophobic protein similar to Os.46402	P	EF557404, EF558290, EF558289, EF558288, EF557656, EF557405, EF557876, EF557658, EF557655, EF556739, EF558081, EF557406, EF557657	20	69374335 to 69374354
258	Drought-inducible protein1 similar to	P	EF556740, EF558082, EF557407, EF558291, EF557659, EF557877	4	69373391 to

	Os.22276				69373394
260	RNase S-like protein similar to Os.9417	P	EF558293, EF556960, EF557177, EF557408, EF557409, EF557176, EF556741, EF558084, EF557660, EF557878, EF557879, EF558083, EF558292	15	69374578 to 69374592
262	Ubiquitinyl hydrolase similar to BU673177	P	EF558295, EF557179, EF557410, EF556743, EF556742, EF558294, EF557881, EF557178, EF557880	4	69373787 to 69373790
266	Tripeptidyl peptidase-II similar to BU672860	P	EF556744, EF558296, EF556961, EF558297, EF557663, EF556747, EF556745, EF557883, EF557882, EF556962, EF557661, EF557662, EF556746, EF557182, EF558085, EF557414, EF557413, EF557412, EF557411, EF557181, EF557180	6	69373566 to 69373571
267	Peptidyl-prolyl *cis*-trans isomerise similar to BU673613	P	EF558086, EF557664, EF556748, EF557183, EF557884	3	69373572 to 69373574
268	Protein kinase similar to BU673471	P	EF557185, EF556964, EF558087, EF557416, EF557415, EF556963, EF557184, EF558298, EF557666, EF557665, EF557885, EF557886, EF556749, EF558299, EF556750	12	69374641 to 69374652
269	Proton pump interactor 1 similar to CB965018	P	EF557668, EF557667, EF557418, EF557417, EF558088, EF557186, EF558089, EF557888, EF558300, EF557887	5	69374432 to 69374436
270	Tubby like protein similar to CB964631	P	EF557890, EF557889, EF558090, EF557670, EF557420, EF557419, EF556965, EF557669, EF558301, EF556752, EF558302, EF556751	19	69373943 to 69373961
276	Lectin like kinase	P	EF557421, EF557671	2	69373389 to 69373390
278	NDC1_Receptor like kinase	P	EF557422, EF557672, EF557423, EF556753, EF557892, EF557891	24	69374311 to 69374334
284	NDC15_γ-aminobutyric acid (GABA)	P	EF557424, EF558304, EF558303, EF556754, EF557673, EF557674, EF556755, EF556966, EF557188, EF557187	24	69373402 to 69373425
285	NDC16_unknown membrane protein	P	EF557675, EF556756, EF557425, EF557189, EF556967, EF557190, EF558305	3	69373687 to 69373689
286	NDC25_DNA binding protein	P	EF558307, EF557676, EF558306, EF557191, EF556757, EF556758	9	69373690 to 69373698
288	NDC28_nifU	P	EF558309, EF557192, EF557193, EF558092, EF557426, EF558308, EF558091, EF557427, EF557677, EF557678, EF557893, EF557894, EF556968	5	69374162 to 69374166
290	NDC30_Avr9/Cf-9 rapidly elicited protein	P	EF556759, EF556969, EF557195, EF557895, EF556970, EF557194, EF557896, EF557429, EF558311, EF557428, EF557679, EF558310	23	69373483 to 69373505
292	NDC33_GYF like protein	P	EF557680, EF556971, EF558312, EF556760, EF558313, EF557897, EF557196, EF557681, EF558093, EF557197, EF557430	17	69374178 to 69374194

293	NDC34_Helicase	P	EF556972, EF557198, EF558314, EF556761	11	69374741 to 69374751
295	NDC41_Transposan like protein	P	EF557200, EF557898, EF557683, EF557682, EF558315, EF557431, EF557432, EF557199, EF558094, EF558095, EF556762, EF556763, EF556974, EF556973	15	69374720 to 69374734
297	NDC46_MAP kinase	P	EF558316, EF557899, EF556764, EF557201, EF557202, EF558317, EF557684, EF556975, EF557433, EF557900	10	69374301 to 69374310
3	Purple acid Phophatase	G	EF557517, EF556843	0	Nil
37	MAP kinase	P	EF557927, EF556796, EF557203, EF557685, EF556976, EF557901, EF557929, EF556765, EF558342, EF558344, EF558096, EF557224, EF558346, EF558318	22	69373636 to 69373657
39	DNA binding protein	P	EF556767, EF557686, EF557687, EF557205, EF557204, EF556766, EF557435, EF557902, EF557903, EF558098, EF558319, EF558320, EF558097, EF557434, EF556977	4	69373283 to 69373286
4	RAS	G	EF558341, EF557928, EF558164, EF556844, EF557458, EF558343, EF556607, EF556795	37	69374070 to 69374106
41	KNOX	P	EF556978, EF557436, EF557904, EF556768, EF558322, EF557688, EF558321, EF557206	22	69374118 to 69374139
43	Zinc finger protein	P	EF557905, EF558100, EF558323, EF557906, EF556769, EF558324, EF558099, EF557207, EF557437	13	69374755 to 69374767
46	Zinc finger protein	G	EF556846, EF556845, EF557051, EF558165, EF558166, EF557518, EF556608, EF557756, EF557281, EF557283, EF557974, EF557282, EF557757	20	69373832 to 69373851
5	Calmodulin	P	EF558101, EF557689, EF557907, EF556770, EF557459	1	69373273
50	Defensin	P	EF557209, EF557909, EF558325, EF557691, EF556980, EF556979, EF558326, EF558103, EF556772, EF557208, EF556771, EF557438, EF557690, EF557908, EF558102	4	69373269 to 69373272
51	Defensin	G	EF558168, EF558167, EF557519, EF556609, EF556848, EF557975, EF557976, EF556847, EF557052	2	69373274 to 69373275
59	Glutathionein S transferase	G	EF557758, EF557759, EF557520, EF557521, EF556610, EF556611, EF557053, EF557284, EF557978, EF557977, EF556850, EF558169, EF556849	5	69373610 to 69373614
60	Disease responsive gene	P	EF556981, EF556982, EF558327, EF558104, EF556773, EF556774, EF557692, EF557210, EF558328, EF557211, EF557910, EF557439, EF557440, EF557911	18	69374702 to 69374719
61	Disease responsive gene	G	EF557761, EF556613, EF556612, EF557760, EF557286, EF558171, EF558170, EF556851, EF556852, EF557979, EF557522, EF557523, EF557285	24	69374678 to 69374701

63	Chitinase	G	EF557525, EF556853, EF556854, EF557287, EF557288, EF557981, EF557762, EF557980, EF558173, EF558172, EF556614, EF557763, EF557524	13	69373919 to 69373931
78	Protein kinase 3 UTR	U	EF558112, EF558337, EF557921, EF557922, EF558111, EF557703, EF557704, EF557452, EF558336, EF556995, EF556787, EF556788, EF556996, EF557451	3	69373327 to 69373329
79	Protein kinase	P	EF557912, EF558105, EF558106, EF557442, EF556984, EF557441, EF557693, EF557694, EF556775, EF557913, EF556983, EF556776, EF558329, EF558330	4	69373323 to 69373326
81	ABA induced gene 3UTR	U	EF557706, EF556789, EF556790, EF556998, EF556791, EF556792, EF558113, EF557923, EF557705, EF557453, EF557221, EF556997, EF557454, EF557924, EF558338, EF557220	5	69374670 to 69374674
83	Inositol phosphate	P	EF558107, EF556985, EF557695, EF556777, EF557914, EF557212, EF558331, EF557443	6	69373681 to 69373686
86	Ethylene responsive gene	P	EF557445, EF557444, EF558333, EF558108, EF558332, EF558109, EF557213, EF557916, EF556987, EF556779, EF556778, EF557915, EF557697, EF556986, EF557696	4	69373939 to 69373942
87	CBF4	G	EF556855, EF556856, EF557765, EF557764	0	Nil
91	Hormone responsive gene 3UTR	U	EF557925, EF556793, EF556999, EF557000, EF557926, EF557708, EF558339, EF558114, EF557222, EF557455, EF557456, EF558340, EF556794, EF557707, EF557223	4	69374509 to 69374512
92	Metallothionein	P	EF556780, EF556989, EF557215, EF557214, EF556988, EF556781	14	69373309 to 69373322
95	betaine aldehyde dehydrogenase EL	G	EF557766, EF556857, EF556615, EF557982, EF556858, EF557054, EF557526, EF558174, EF557289	3	69374437 to 69374439
98	Choline monooxygenase CD	G	EF557767, EF556616, EF556617, EF557983, EF557055, EF558175, EF557527	2	69374167 to 69374168
99	Cyclophilin	P	EF557699, EF557698, EF557918, EF557216, EF557446, EF556990, EF556782, EF556783, EF557917	7	69373679 to 69373680, 69373699 to 69373703
24	DREB1A	P	EU869018, EU868960, EU868961, EU868976, EU868928, EU868942, EU869019, EU869006, EU868992, EU868991, EU868929	5	105111356 to 105111360
42	MYB	P	EU869029, EU868955, EU868971, EU869030, EU868954, EU868985, EU868938, EU868937, EU869014, EU869013, EU868986	0	NA

72	Transcription factor-II	G	EU868963, EU869022, EU868978, EU868993, EU868945, EU869021, EU868977, EU869007, EU868964, EU868931, EU868930, EU868946, EU869008, EU868994	7	105111427 to 105111433
82	Inositol phosphate 3' UTR	U	EU868975, EU868959, EU868958, EU868974, EU868990, EU868941, EU869005, EU869017	4	105111394 to 105111397
93	MAP kinase	P	EU869002, EU868956, EU868957, EU869003, EU868939, EU868988, EU869015, EU868940, EU869016, EU868987, EU868972, EU868973, EU869031, EU869032	0	NA
97	Choline monooxygenase	P	EU869033, EU868989, EU869004	23	105111398 to 105111420
101	Ribosomal 6 kinase	G	EU869020, EU868962, EU868944, EU868943	0	NA
104	M_Ring finger protein	P	EU869009, EU868995, EU869024, EU868996, EU869023, EU868947, EU869010, EU868965, EU868966, EU868980, EU868979, EU868932	9	105111395 to 105111426
140	M_Catalase	P	EU869012, EU869026, EU868981, EU868967, EU868968, EU869025, EU869011, EU868949, EU868950, EU868934, EU868982, EU868997, EU868948, EU868998, EU868933	11	105111345 to 105111355
165	Chalcone Isomerase	P	EU868969, EU868984, EU868983, EU868999, EU868951, EU868970, EU868935	13	105111381 to 105111393
219	Glutathionein reductase	P	EU869027, EU868936, EU868952, EU868953, EU869000, EU869028, EU869001	20	105111361 to 105111380
301	MIR159a	M	EU869079, EU869139, EU869038, EU869057, EU869056, EU869140, EU869039, EU869160, EU869159, EU869080, EU869102, EU869101	7	105107073 to 105107080
302	MIR159b	M	EU869125, EU869161, EU869104, EU869103, EU869124, EU869041, EU869081, EU869040, EU869141, EU869082, EU869058, EU869162, EU869059, EU869035, EU869034	0	NA
304	MIR159c	M	EU869042, EU869143, EU869142, EU869163, EU869083, EU869106, EU869061, EU869105, EU869084, EU869060,EU869043	0	NA
305	MIR159d	M	EU869108, EU869107, EU869126, EU869164, EU869085, EU869145, EU869127, EU869063, EU869044, EU869165, EU869045, EU869062, EU869144	0	NA
306	MIR159e	M	EU869109, EU869166, EU869064, EU869128, EU869086, EU869146	0	NA

309	MIR172a	M	EU869167, EU869168	4	105107116 to 105107119
311	MIR172d	M	EU869087, EU869147, EU869066, EU869065, EU869169, EU869148, EU869170, EU869110, EU869111, EU869088	10	105107083 to 105107092
312	MIR172c	M	EU869172, EU869171, EU869112, EU869089, EU869046, EU869090, EU869068, EU869067, EU869130, EU869129, EU869113, EU869047, EU869036, EU869149, EU869150	0	NA
313	MIR172d	M	EU869114, EU869174, EU869091, EU869115, EU869092, EU869173, EU869069, EU869152, EU869131, EU869132, EU869049, EU869048, EU869151, EU869070	6	105107092 to 105107098
314	MIR319a	MU	EU869055, EU869097, EU869138, EU869137, EU869121, EU869120, EU869157, EU869158, EU869098, EU869076, EU869075, EU869178, EU869179, EU869054	2	105107081, 105107082
317	MIR398a	M	EU869116, EU869050, EU869051, EU869154, EU869134, EU869133, EU869175, EU869071, EU869176, EU869153, EU869093, EU869072, EU869117, EU869094	19	105107099 to 105107115
318	MIR398b	MU	EU869077, EU869078, EU869180, EU869122, EU869100, EU869099, EU869123	17	105107058 to 105107073
319	MIR398b	M	EU869155, EU869156, EU869177, EU869136, EU869135, EU869053, EU869118, EU869037, EU869096, EU869095, EU869074, EU869073, EU869052, EU869119	4	105107054 to 105107057

P- Promoter; G- Genic region; U- UTR; M- miRNA; MU- UTR region of miRNA

Table 4.8: Details of the target successfully analyzed regions along with the accession numbers deposited at genbank division and ss numbers of the SNPs deposited at dbSNP division of NCBI.

4.5. Physical mapping of SNPs to IRGSP sequence map and analysis of informative sites

The regions which are successfully analyzed were mapped on to the rice sequence map (IRGSP) with the help of BLAST algorithm (Fig. 4.10). It has been observed that most of the target regions were mapped on to the chromosome 1, 6 and 7. Least number of regions was targeted from chromosome 12. A set of 81 SNPs remained unmapped as the corresponding BAC sequences were not mapped to the finished rice chromosome map. However, after generating the reference clusters at dbSNP database only 53 SNPs out of 1540 reference SNPs generated in build 128 remained unmapped. More number of transitions was observed on chromosome 1, 2, 8, 10 and 11 than other variations (Fig. 4.12). Chromosome 3, 4 and 5 were observed to have more transversions than transitions and indels. In all the chromosomes a significant number of indels were identified, in fact chromosome 6 was observed to contain more number of indels than any of the other variations. However, this pattern and distribution can not be generalized as the target regions are selectively chosen and not randomly picked up. All the informative sites are also mapped on to rice genome sequence map (Fig. 4.11). Further, the SNPs mapped onto the each chromosome were classified based on the type of variation. A physical map of analyzed regions along with the SNPs identified was generated with the help of CMap (GMOD, sourceforge) program which further helped in understanding the analyzed regions and SNPs positions at the genome level. The physical maps of the chromosome 1, 2, 3 are displayed in Fig. 4.13; chromosome 4, 5, 6 are in Fig. 4.14; 7, 8, 9 are in Fig. 4.15 and chromosomes 10, 11, 12 are displayed in Fig.4.16. Due to the pixel limitation annotation of all SNPs at the target site are not displayed. Each SNP is indicated by a

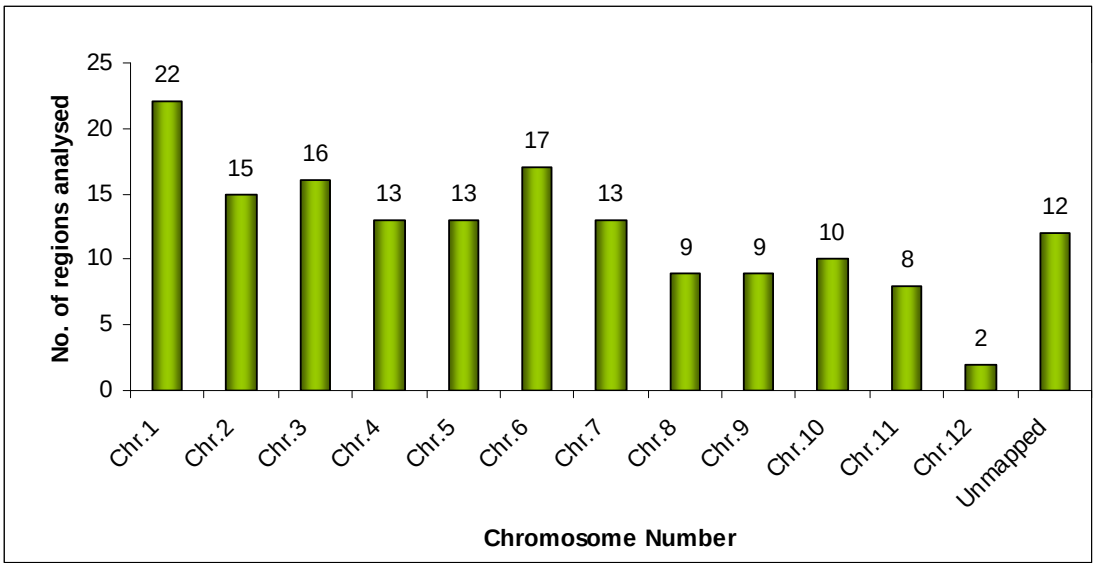


Figure 4.10: Chromosomal distribution of analysed regions on rice chromosomes; Unmapped- Regions which are not mapped precisely onto any of the rice chromosomes.

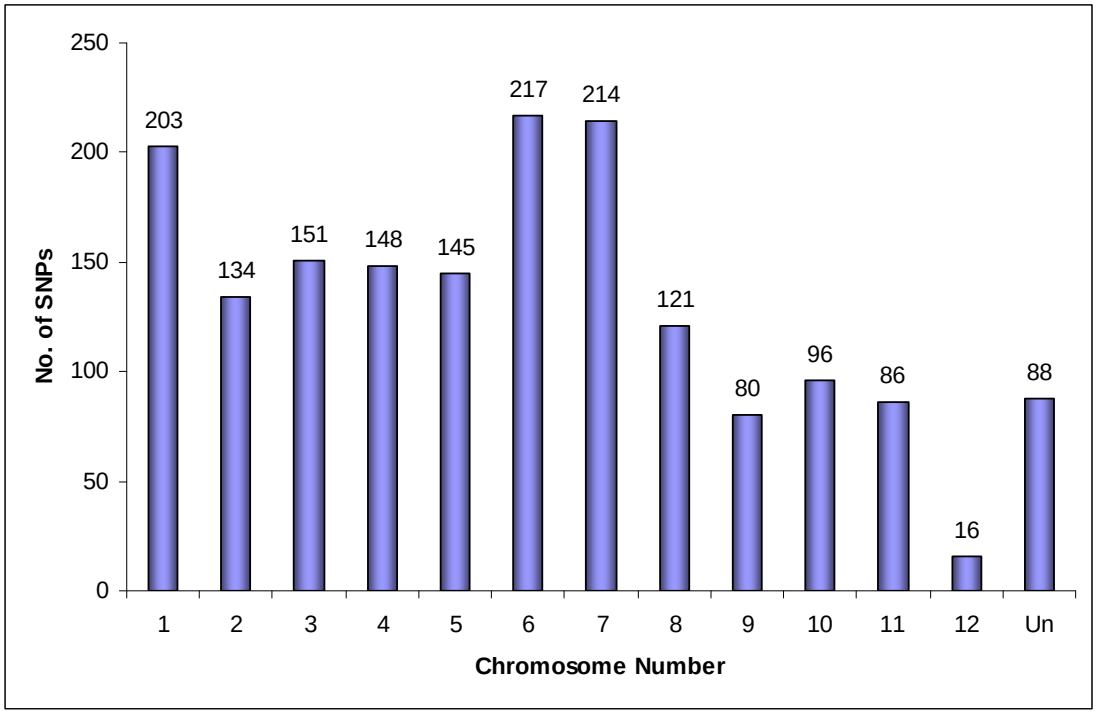


Figure 4.11: Chromosomal distribution of SNP on rice chromosomes; Un: SNPs which are not mapped precisely onto any of the rice chromosomes.

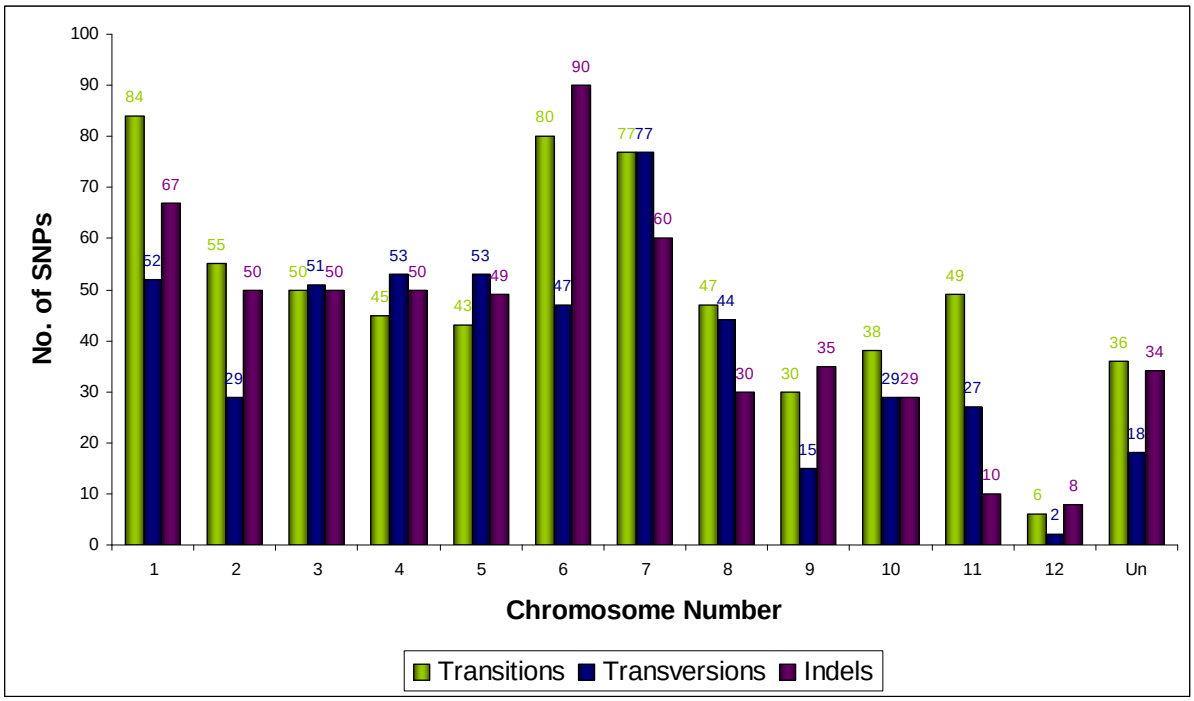


Figure 4.12: Chromosomal distribution of SNPs at the candidate loci categorised based on the type of variation.

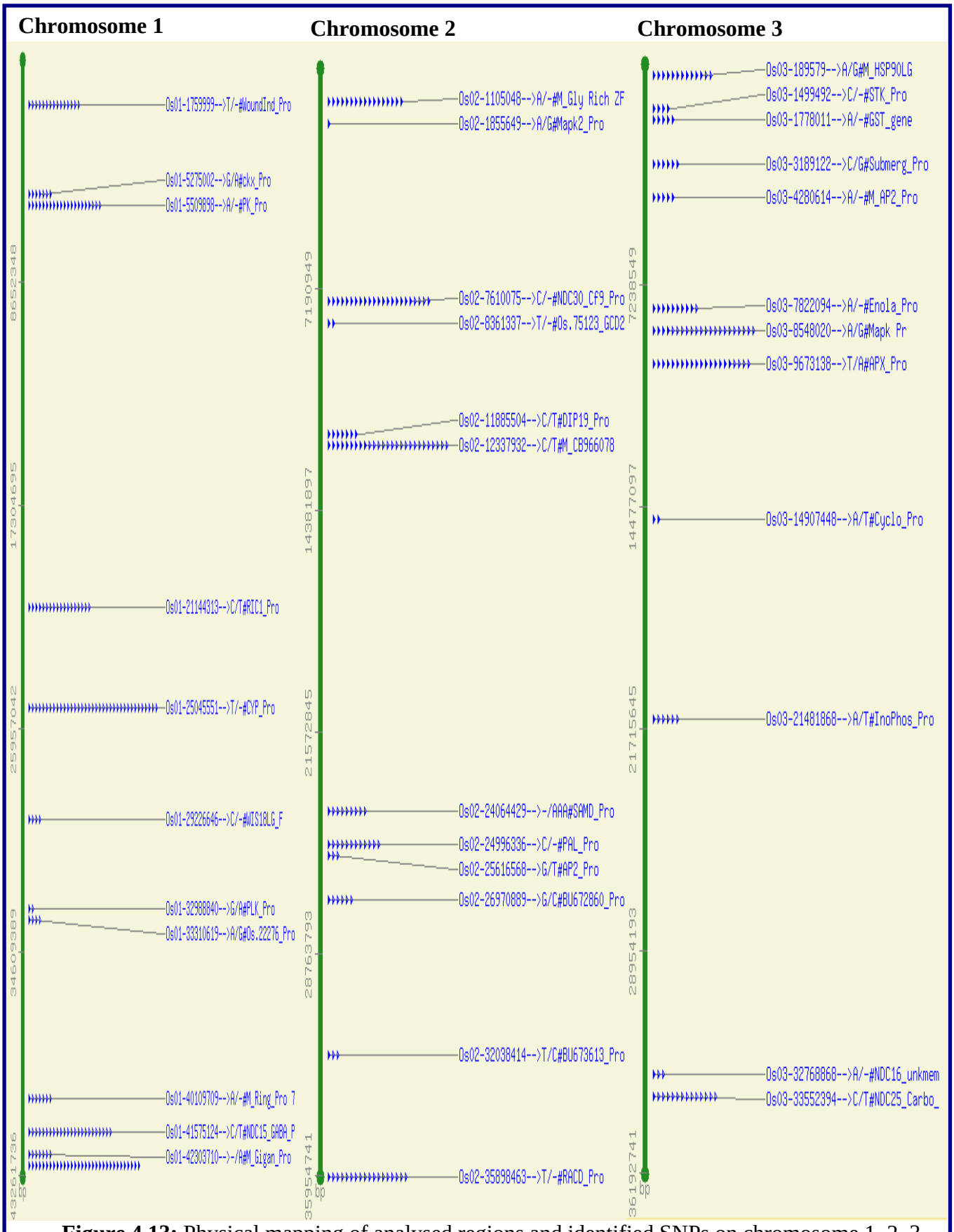


Figure 4.13: Physical mapping of analysed regions and identified SNPs on chromosome 1, 2, 3.

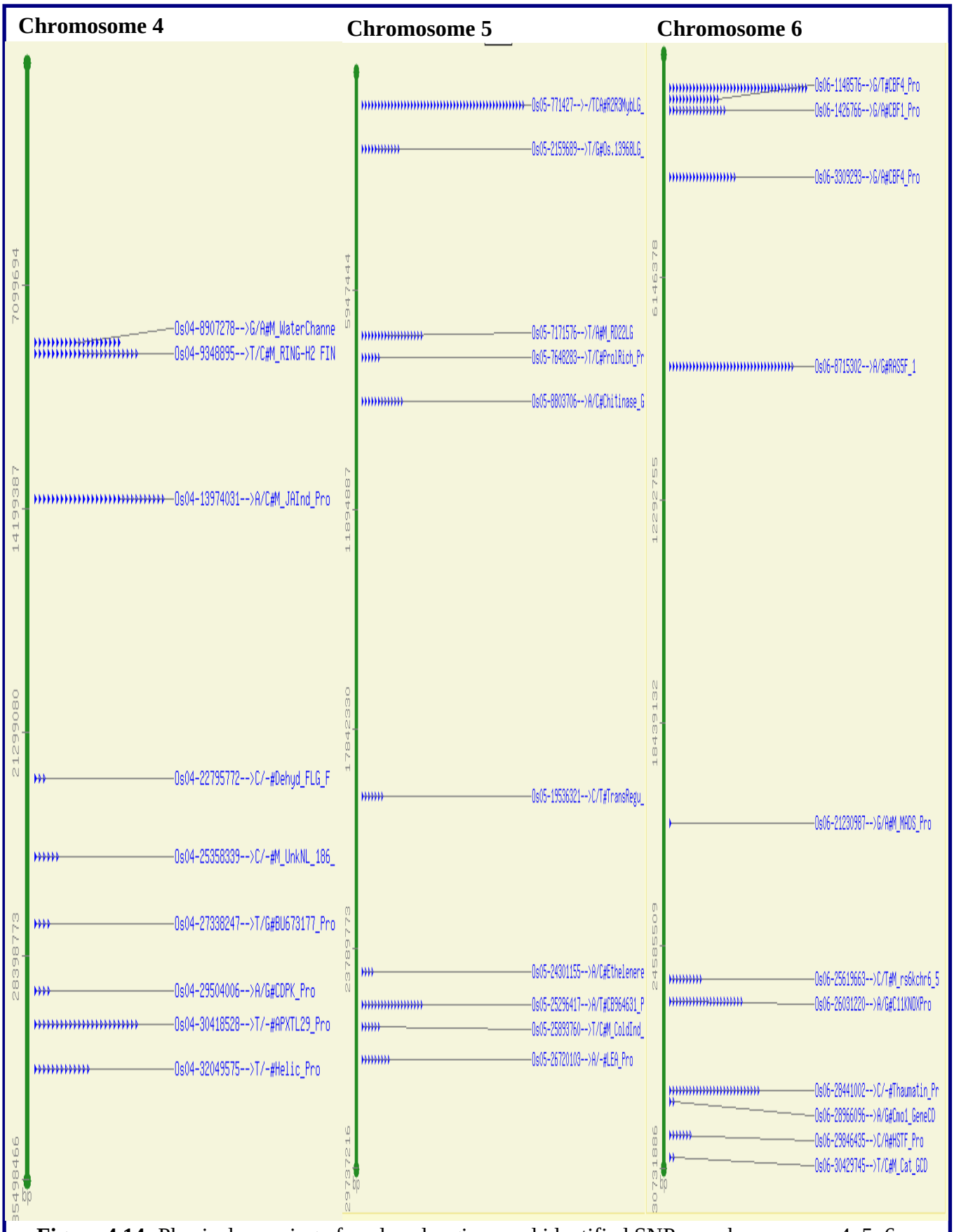


Figure 4.14: Physical mapping of analysed regions and identified SNPs on chromosome 4, 5, 6.

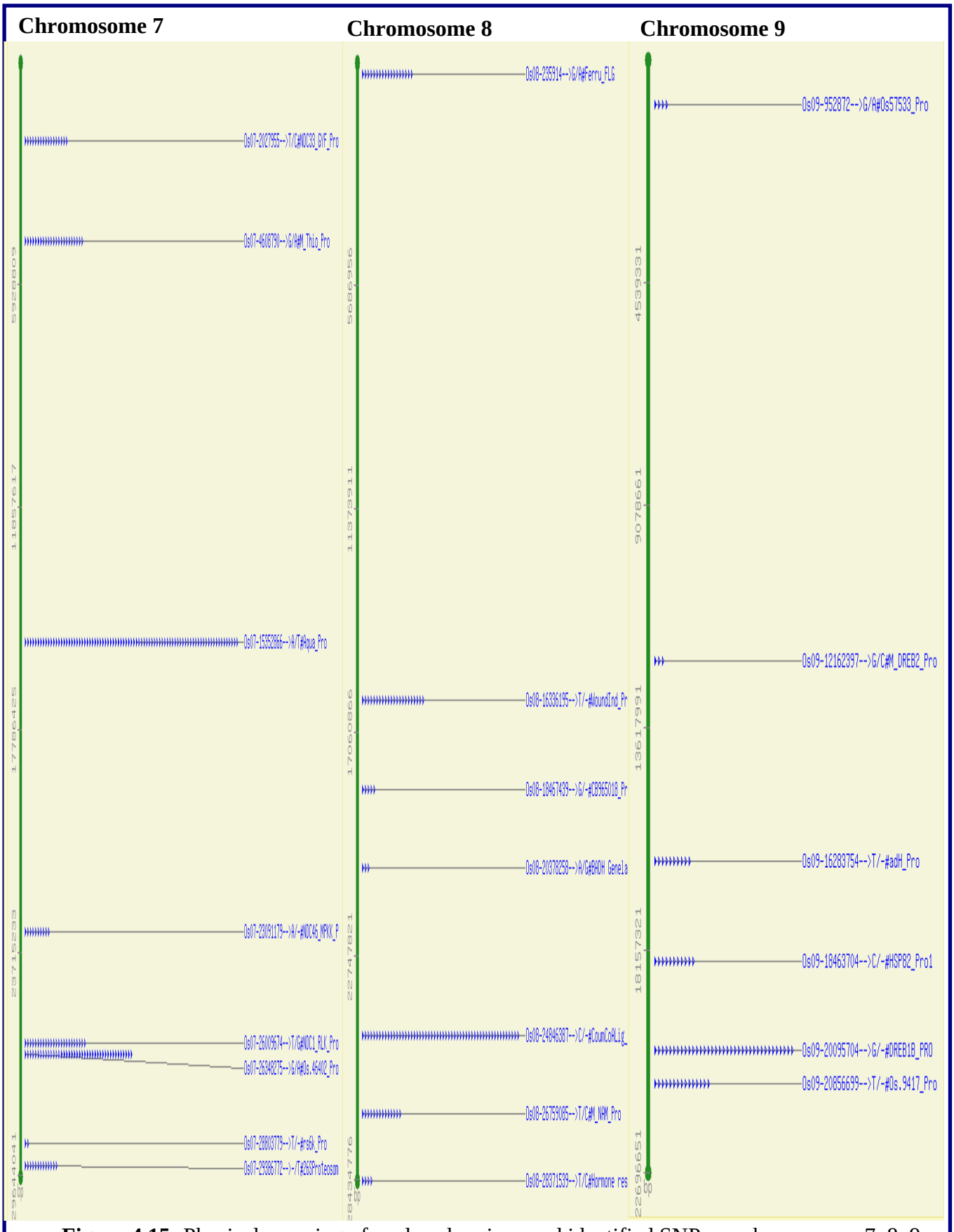


Figure 4.15: Physical mapping of analysed regions and identified SNPs on chromosome 7, 8, 9.

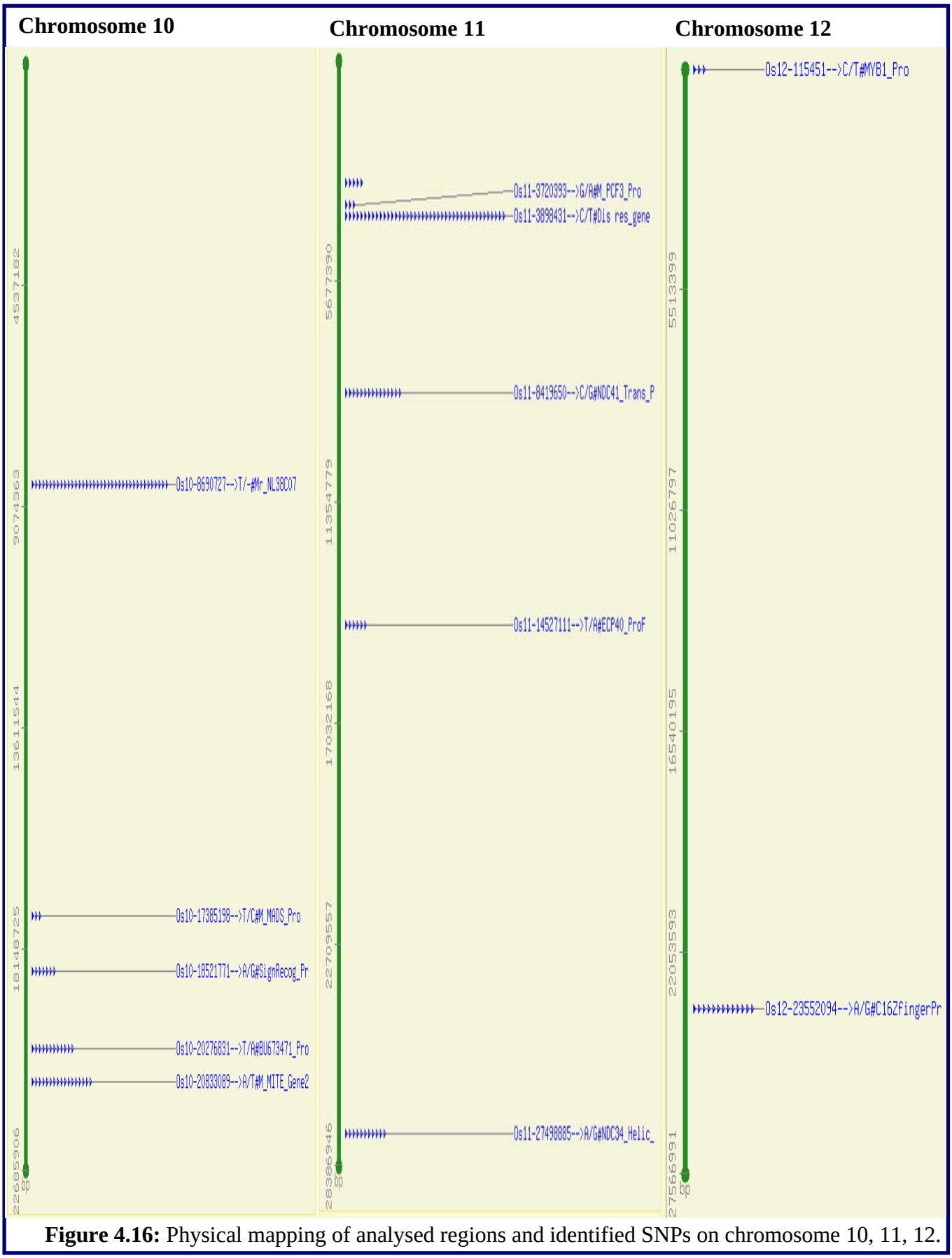


Figure 4.16: Physical mapping of analysed regions and identified SNPs on chromosome 10, 11, 12.

triangle and base positions were indicated along with chromosome vertically. SNP positions were identified and denoted following a unique nomenclature referring chromosome number followed by base position on the rice genome for all the regions which could be mapped on to the rice sequence map. The unmapped SNPs were denoted with the 'Un' prefix followed by corresponding BAC accession number and SNP position in the BAC sequence. This physical mapping and unique nomenclature system has greatly helped in downstream processing of the informative sites in order to generate an integrated map with SNPs.

4.6. Development of allele sharing map

The alleles at the SNP site were extracted from each of the SNP site and formulated orderly for each of the genotype and stored in the SNP database. The consensus was derived from the sequencing data of both the strands with the help of custom Perl scripts at each of the SNP site. Further, with the help of physical mapping data and alleles identified at the SNP site, an attempt was made to develop a genome wide haplotype map with the selective regions and allele sharing maps were generated with the help of an in-house developed software tool called AS-map V1.0. This newly developed program enabled us to develop allele sharing maps of candidate genes and integrated with rice sequence map as well as genetic map utilizing the data available in the public domain. The data from the physical and genetic map integration (IRGSP, 2005) helped in the process of integrating allelesharing maps with that of the genetic maps. This level integration helped further in correlating QTL maps from Gramene database (http://www.gramene.org) and finally to integrate the QTL data from the Gramene database. Further, a filter was developed and used in the program to restrict the

integration of only abiotic stress associated QTLs from all the QTL data available at Gramene. This entire process helped greatly in generating allelesharing maps and integrating with sequence, genetic and abiotic stress QTL map. The integrated allelesharing maps helped us not only in assessing allelesharing patterns but also in genome wide view of the analyzed regions along with the abiotic stress responsive QTLs. The allelesharing maps were color coded for each of the allele as such to easily distinguish allelesharing patterns among the drought tolerant and susceptible genotypes. Green color background was used to denote an allele having a pattern similar to drought tolerant line and red color background was used to denote a pattern similar to a susceptible parent. Grey color was used as the background where allele was ambiguous and denoted with 'N' and not considered in color coding based on the phenotype. White background was used for the alleles which do not follow a unique pattern similar to either of the parental lines. The allele from the IRGSP sequence data of Nipponbare was used as the reference point for each of the allelesharing pattern and displayed in the first column with yellow background. Further, the color used for the alleles were also coded with specific colors for each of the nucleotide. The standard colors of the nucleotides such as black for 'G', red for 'T', blue for 'C' and green for 'A' were used in the map. The SNP type was displayed immediately after the allelesharing pattern and dbSNP number given for the SNP follows the type of SNP. The genetic map distance was shown with arrow showing the range and genetic distance was shown in centimorgans (cM). This is followed by the QTLs mapped to the region displayed using appropriate filter along with the accession numbers of the QTL in the Gramene database. The integrated allelesharing maps are shown in Fig. 4.17*A* to Fig. 4.17*L* with high compression as to display all the

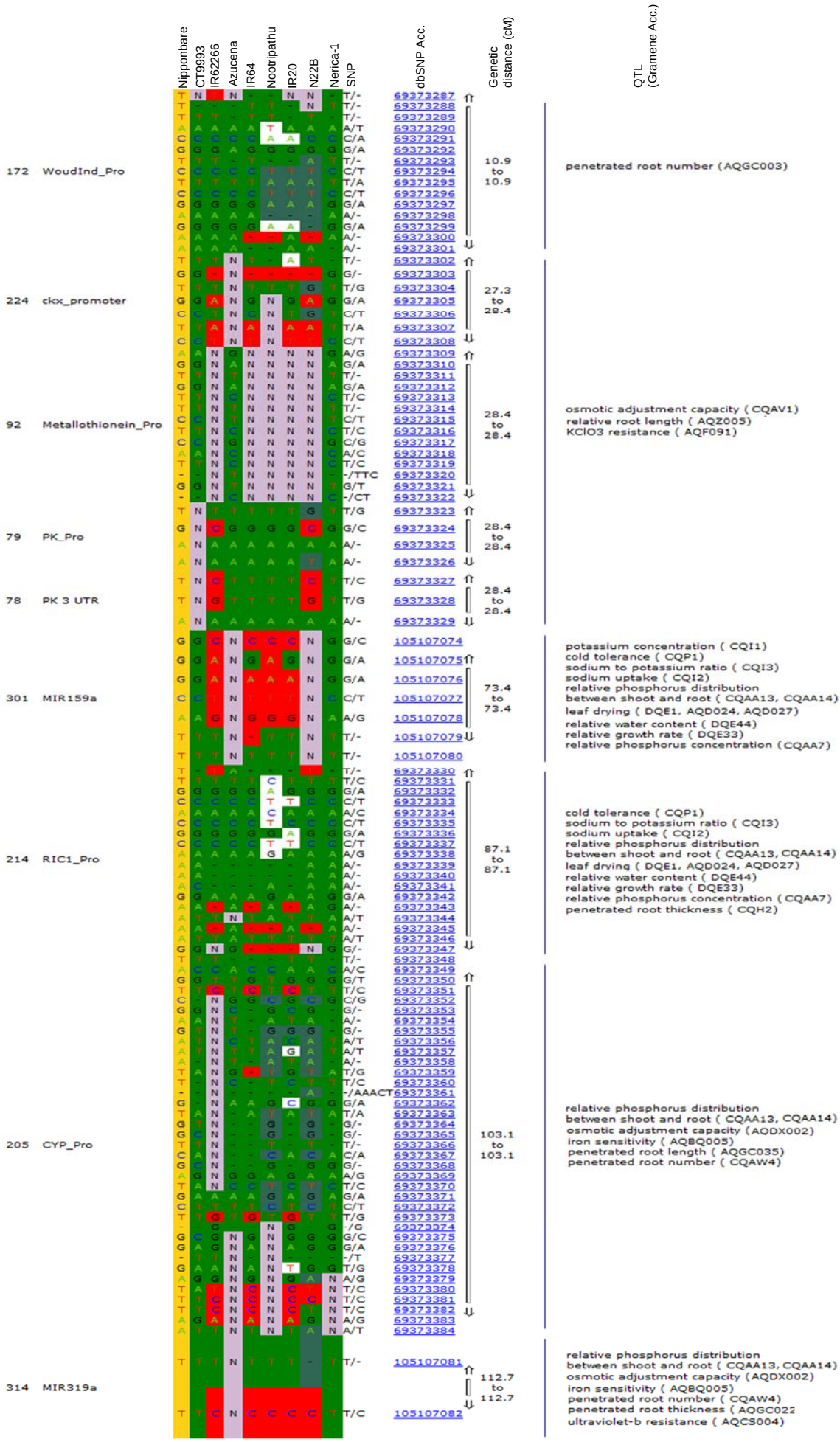


Fig. 4.17A: Integrated allele sharing map of chromosome 1

Conti..

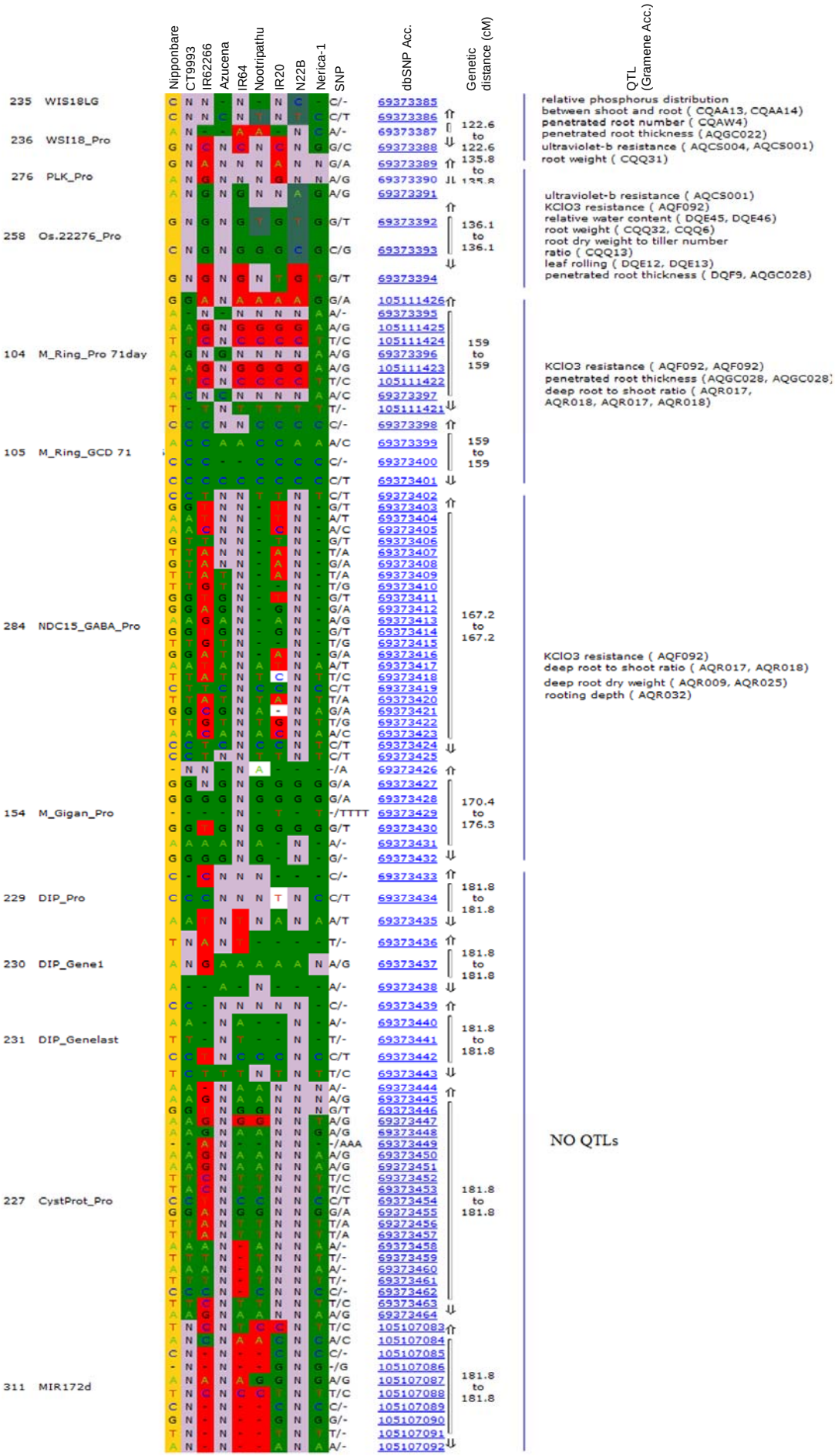


Fig.4.17A: Integrated allele sharing map of chromosome 1

Continued from previous page

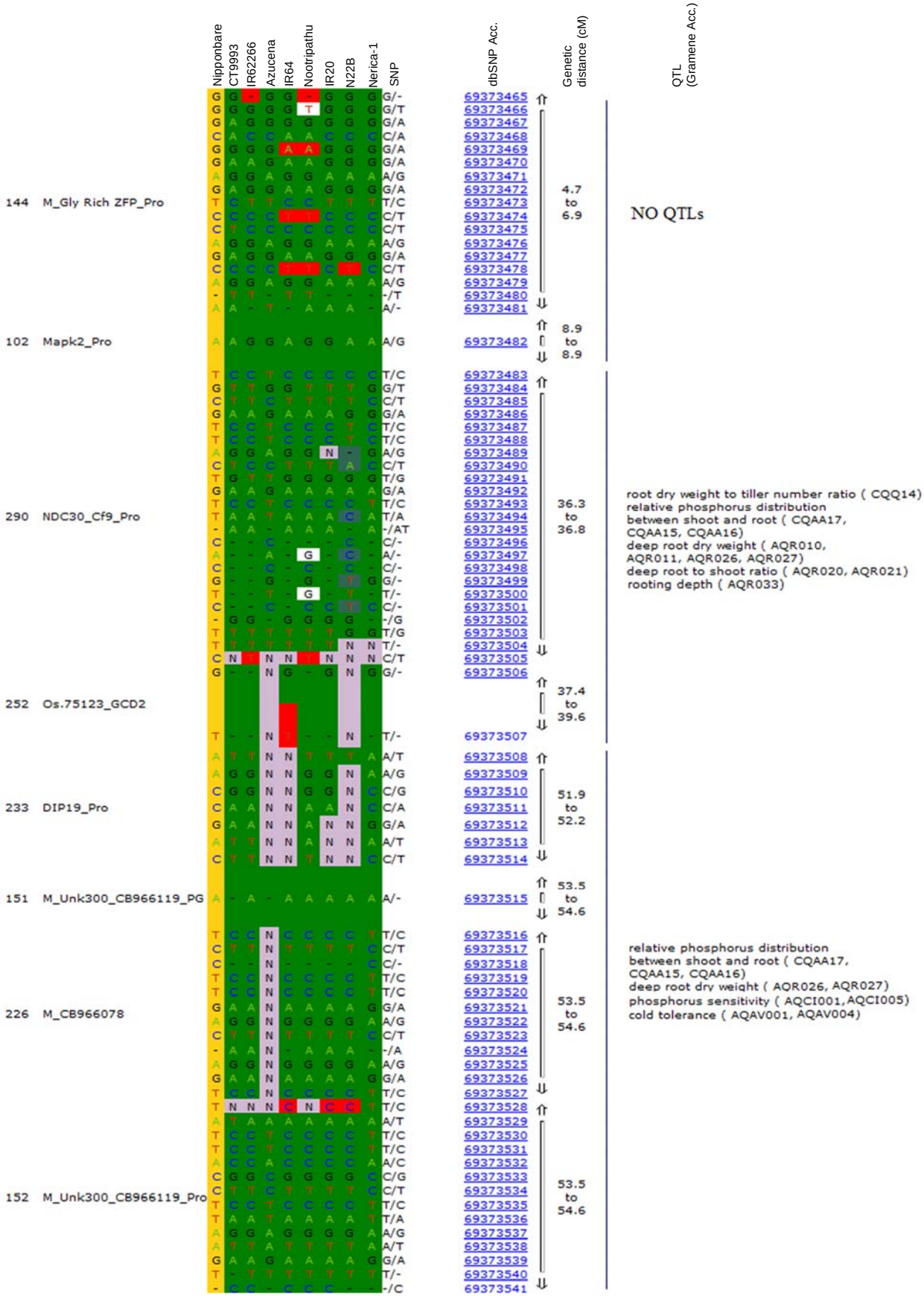
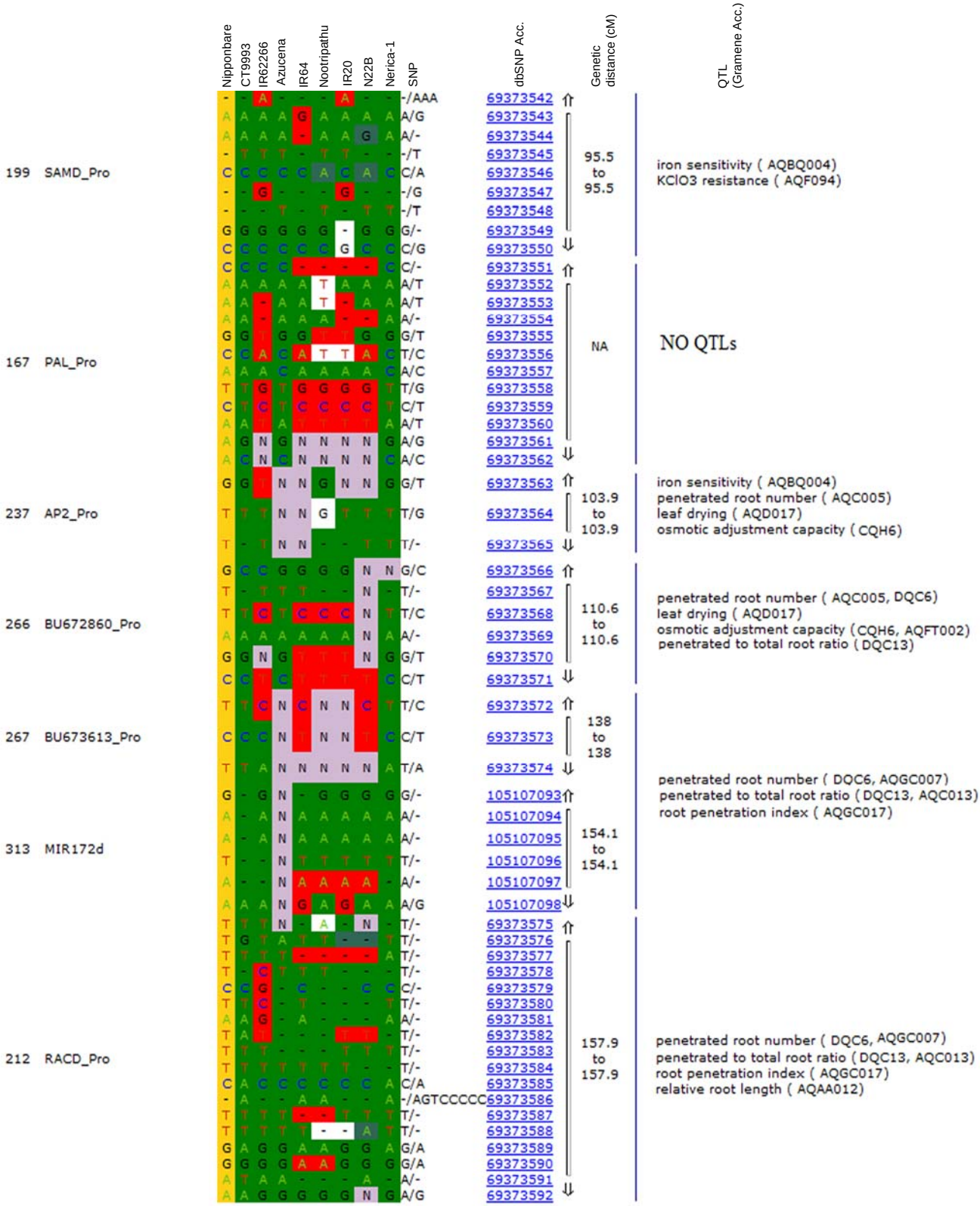


Fig. 4.17B: Integrated allele sharing map of chromosome 2

Conti..



Integrated allele sharing map of chromosome 2

Continued from previous page

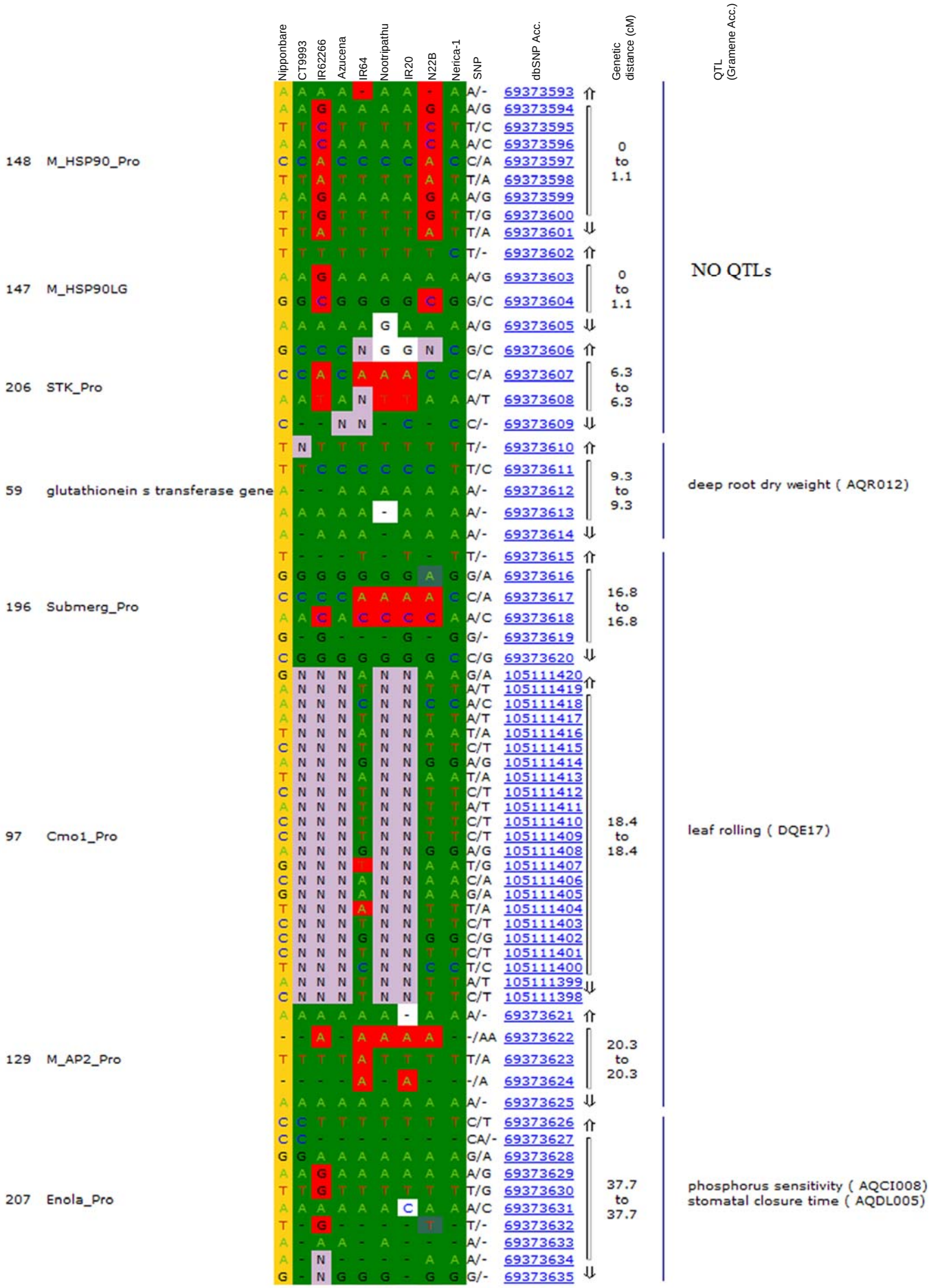


Fig. 4.17C: Integrated allele sharing map of chromosome 3

Conti..

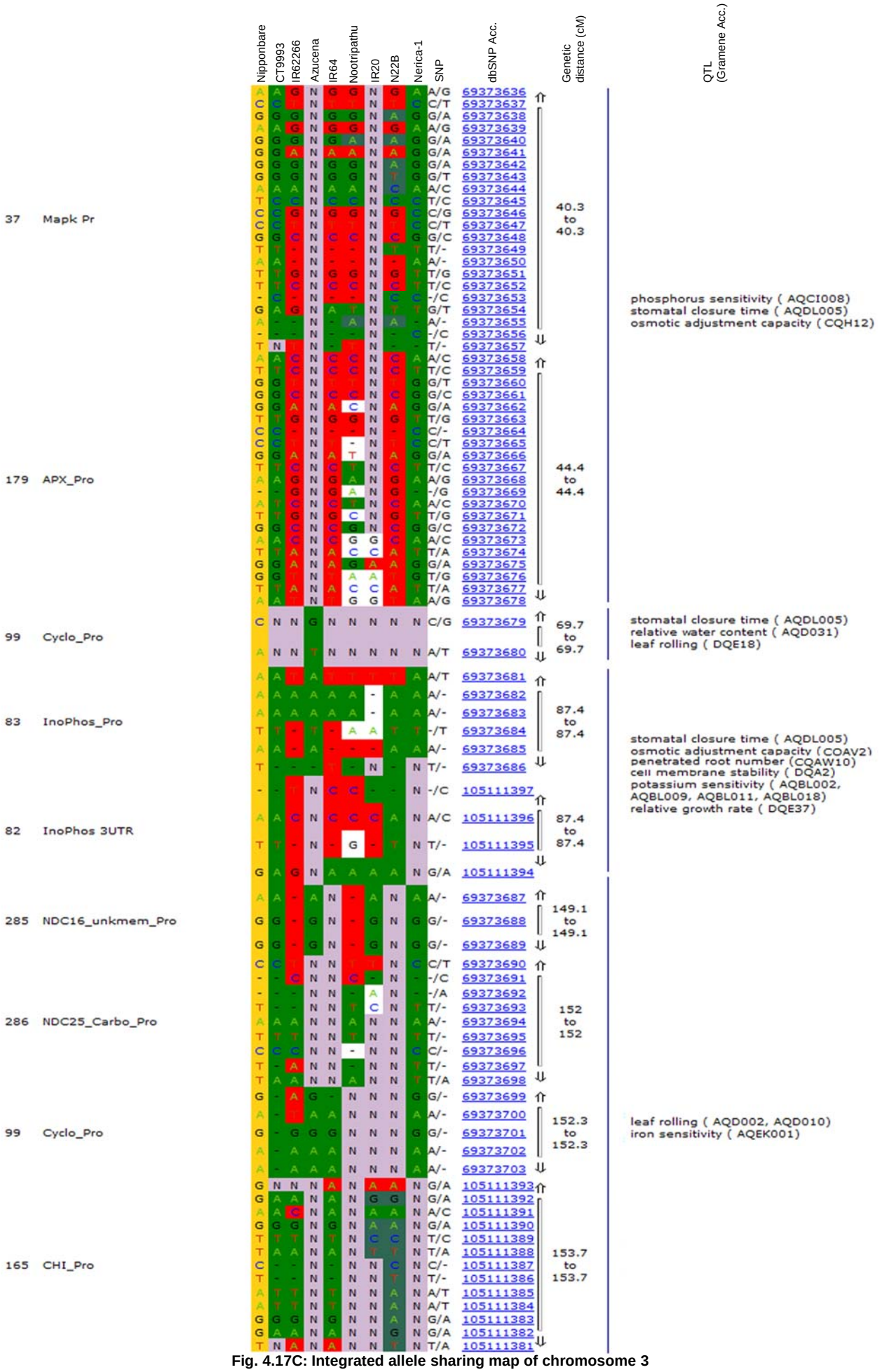


Fig. 4.17C: Integrated allele sharing map of chromosome 3

Continued from previous page

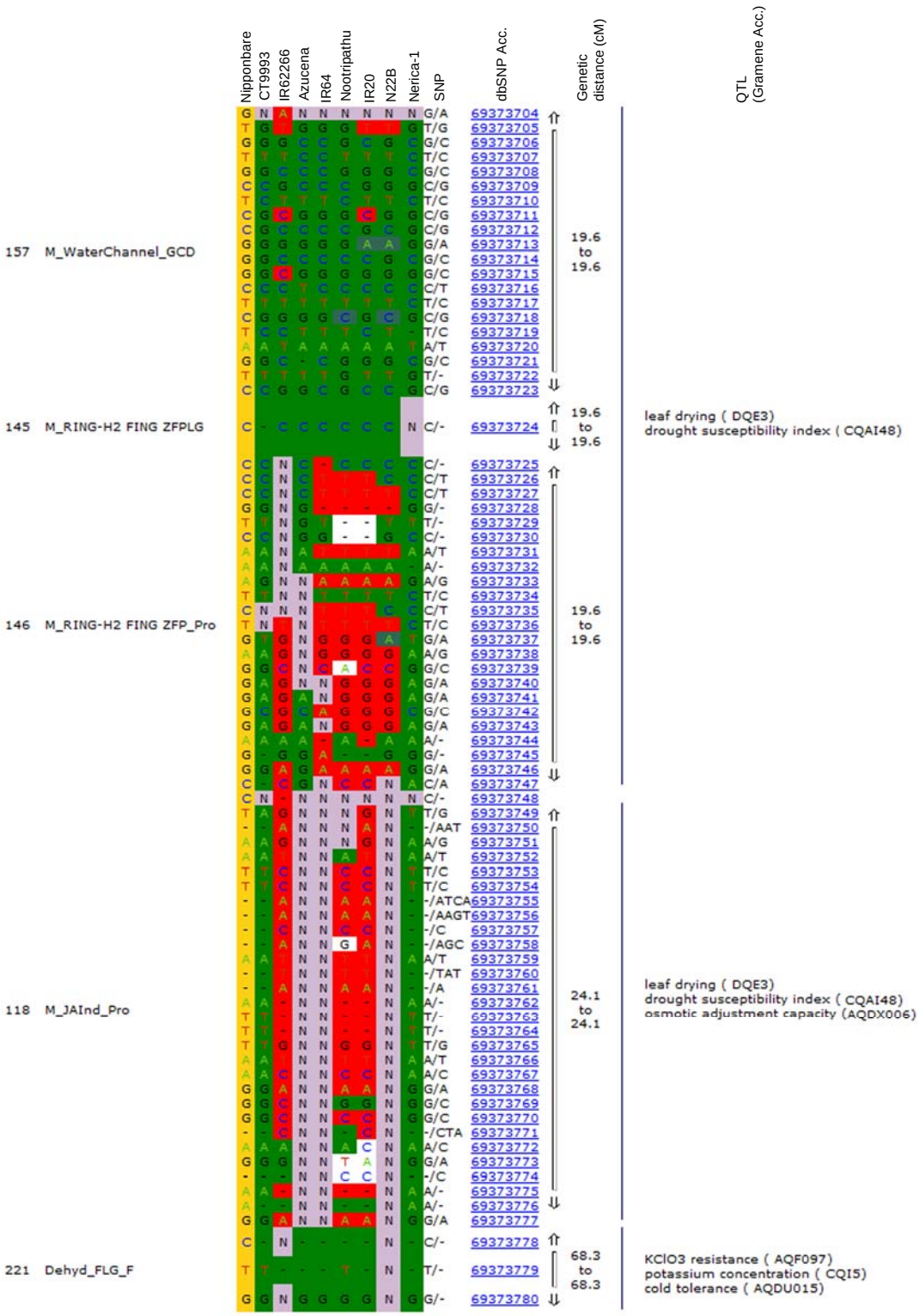


Fig. 4.17D: Integrated allele sharing map of chromosome 4

Conti..

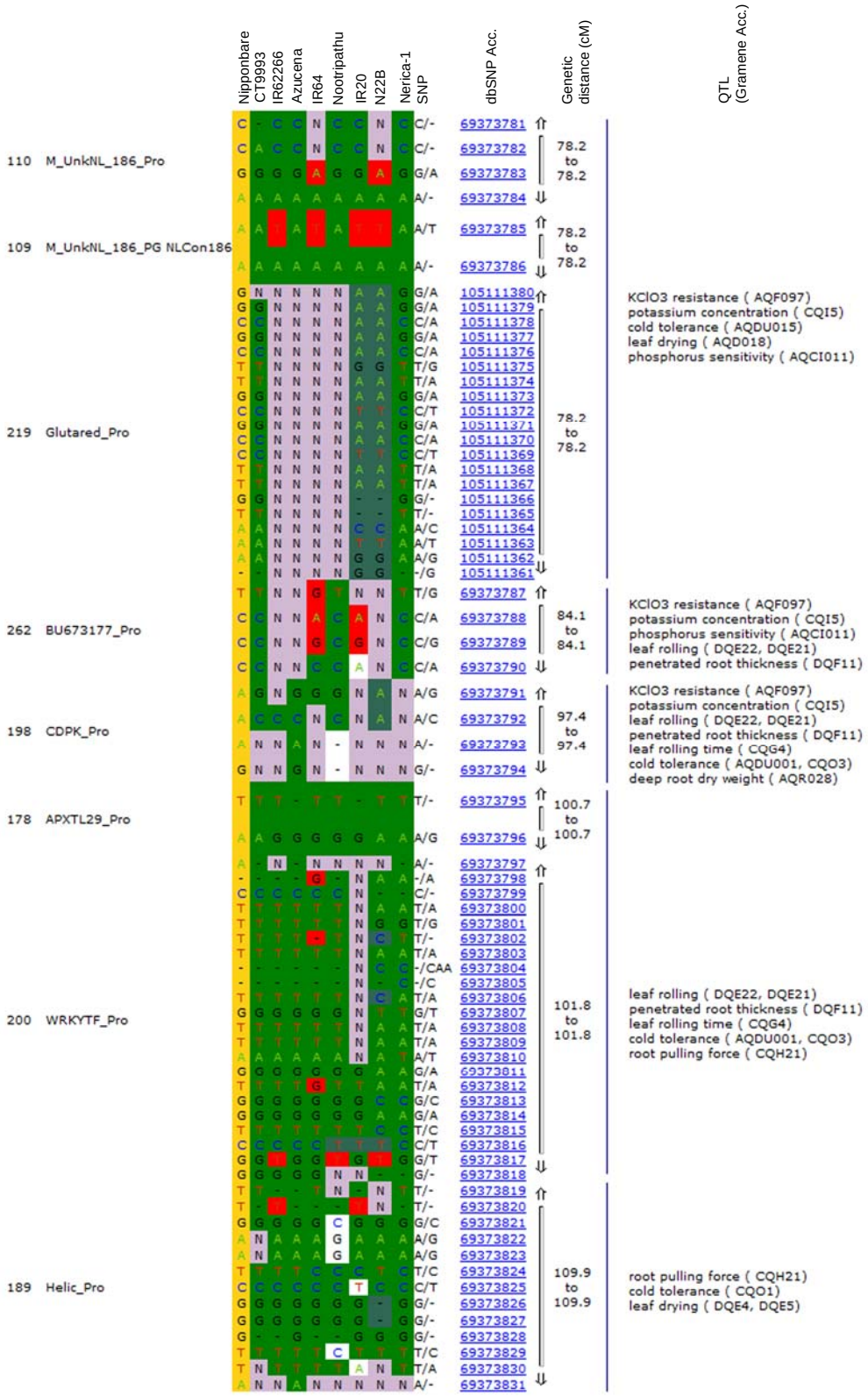


Fig. 4.17D: Integrated allele sharing map of chromosome 4

Continued from previous page

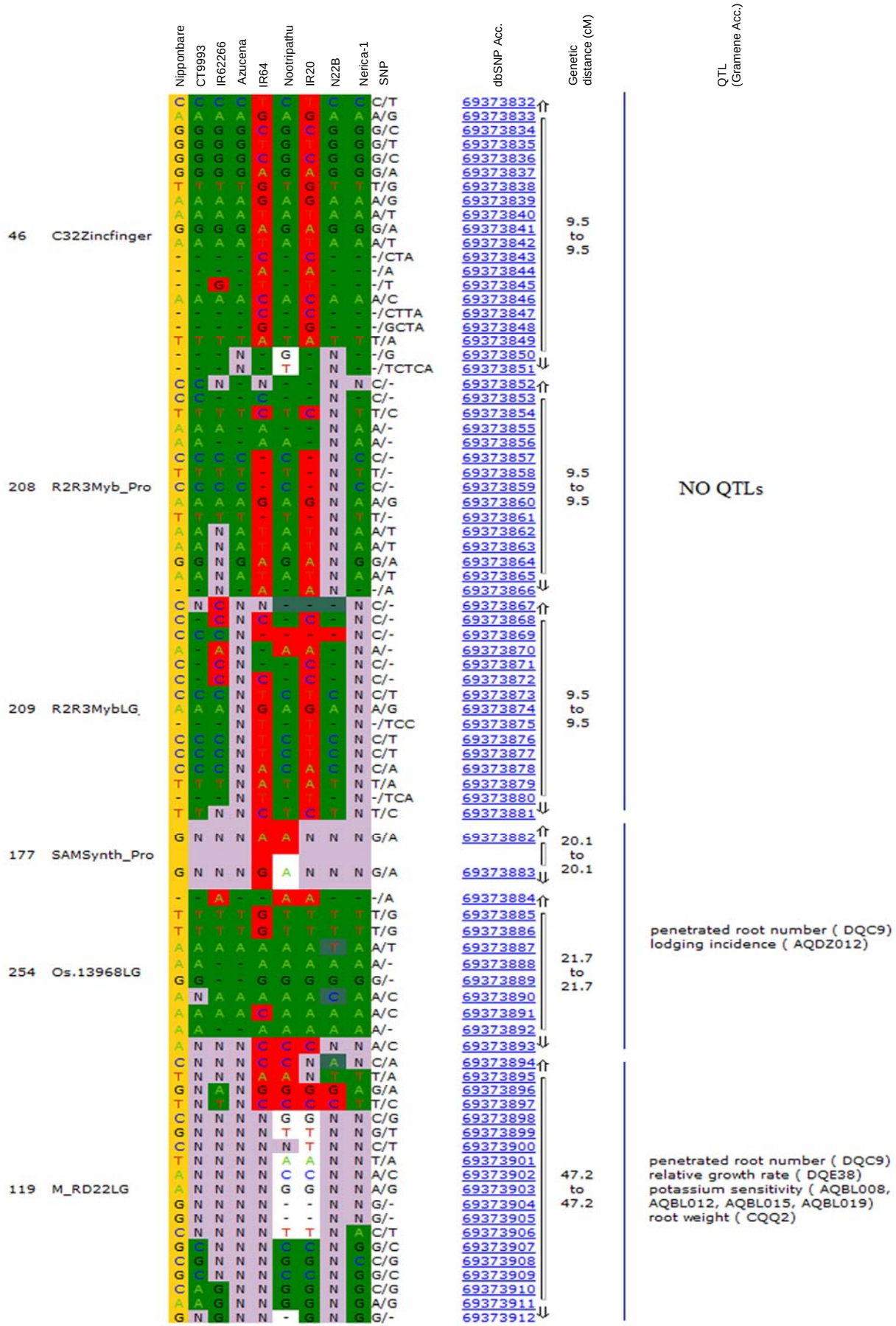


Fig. 4.17E: Integrated allele sharing map of chromosome 5

Conti..

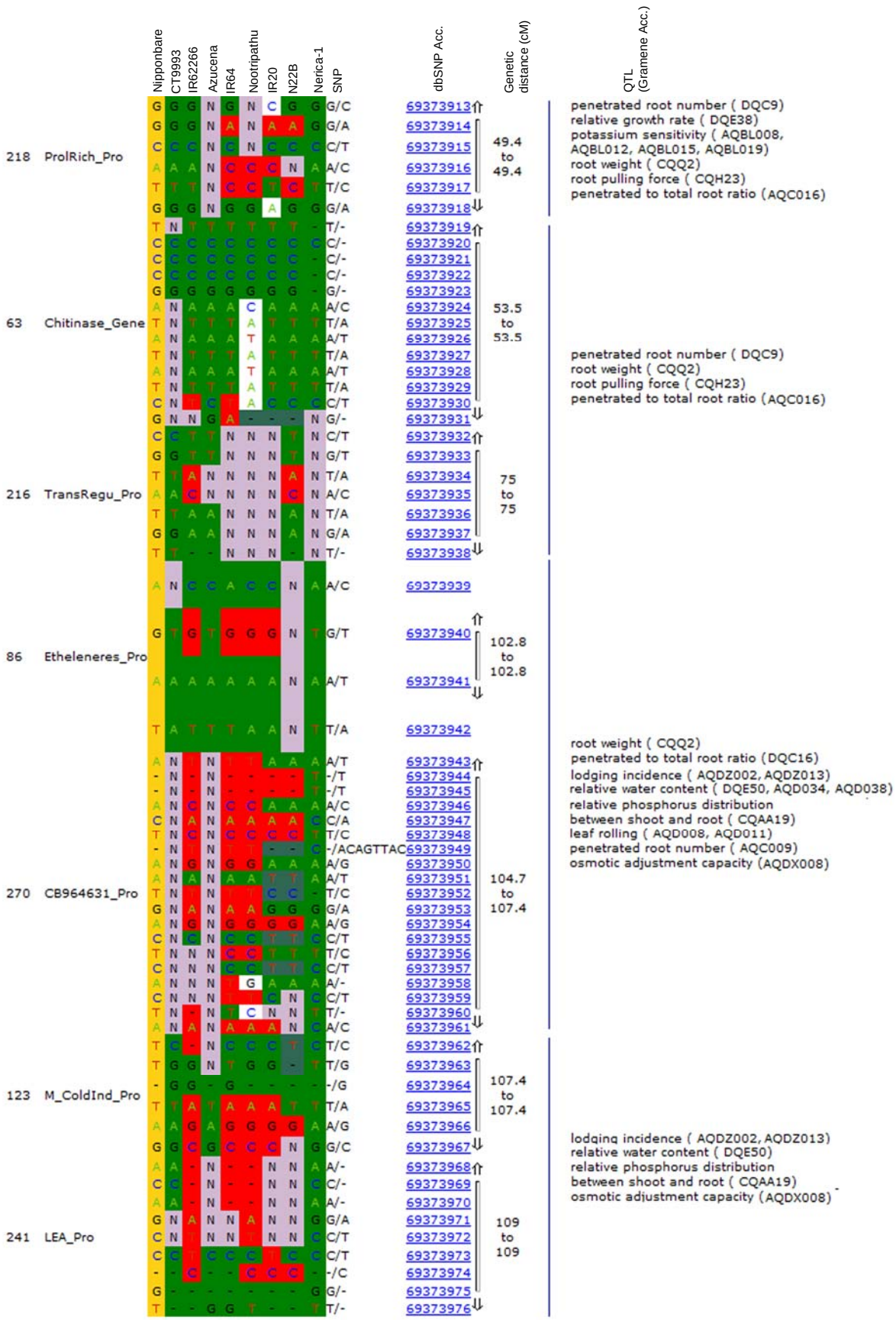


Fig. 4.17E: Integrated allele sharing map of chromosome 5

Continued from previous page

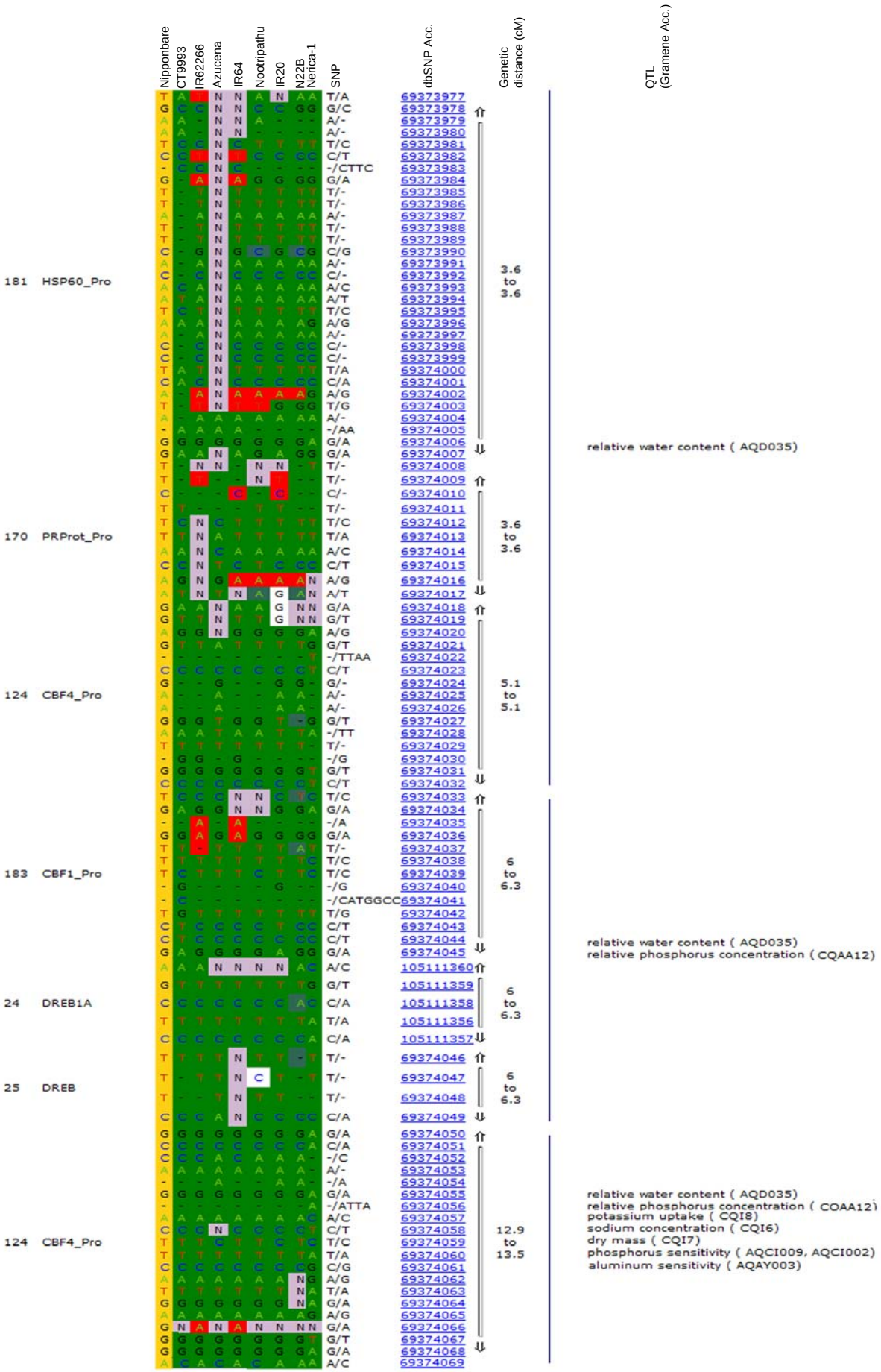


Fig. 4.17F: Integrated allele sharing map of chromosome 6

Conti..

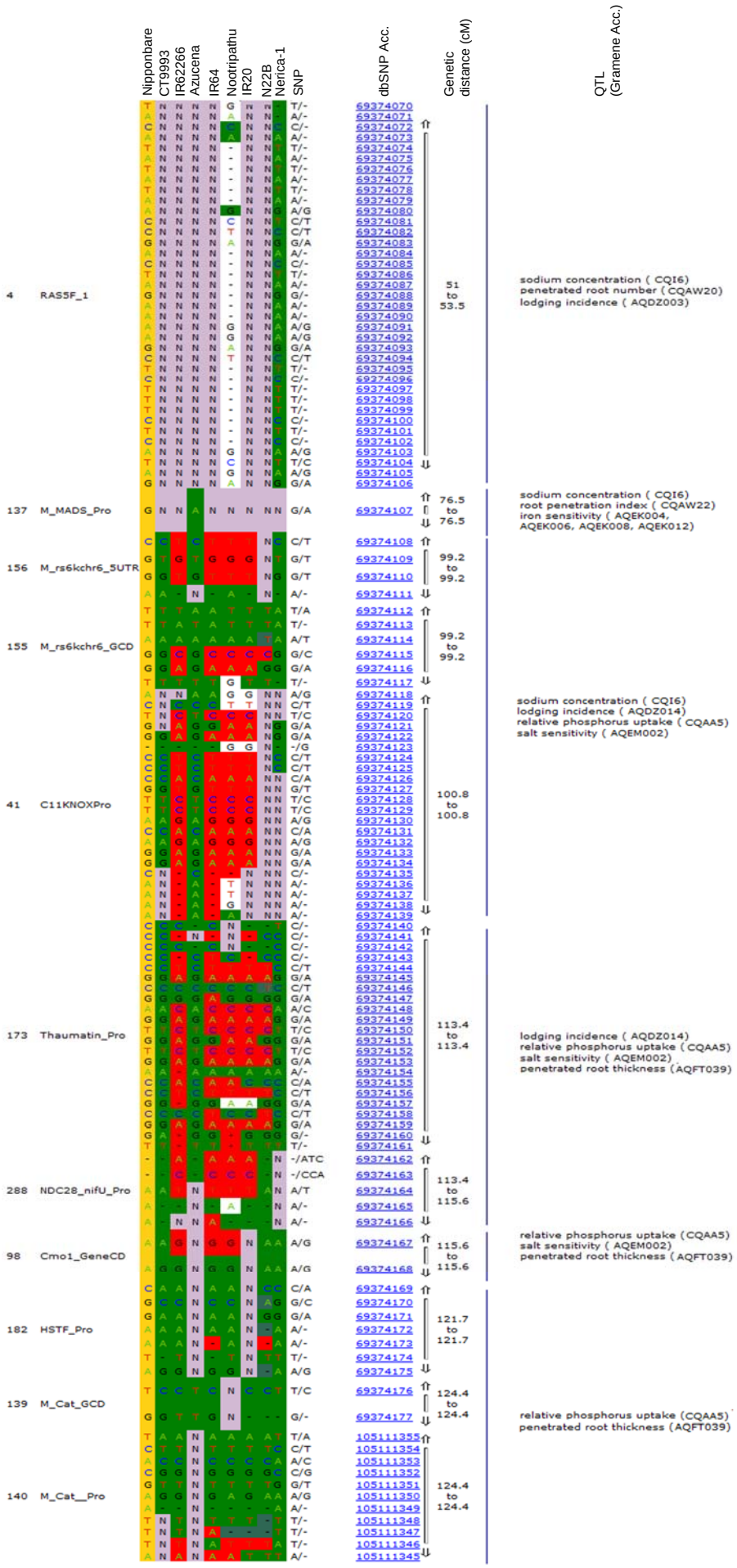


Fig. 4.17F: Integrated allele sharing map of chromosome 6

Continued from previous page

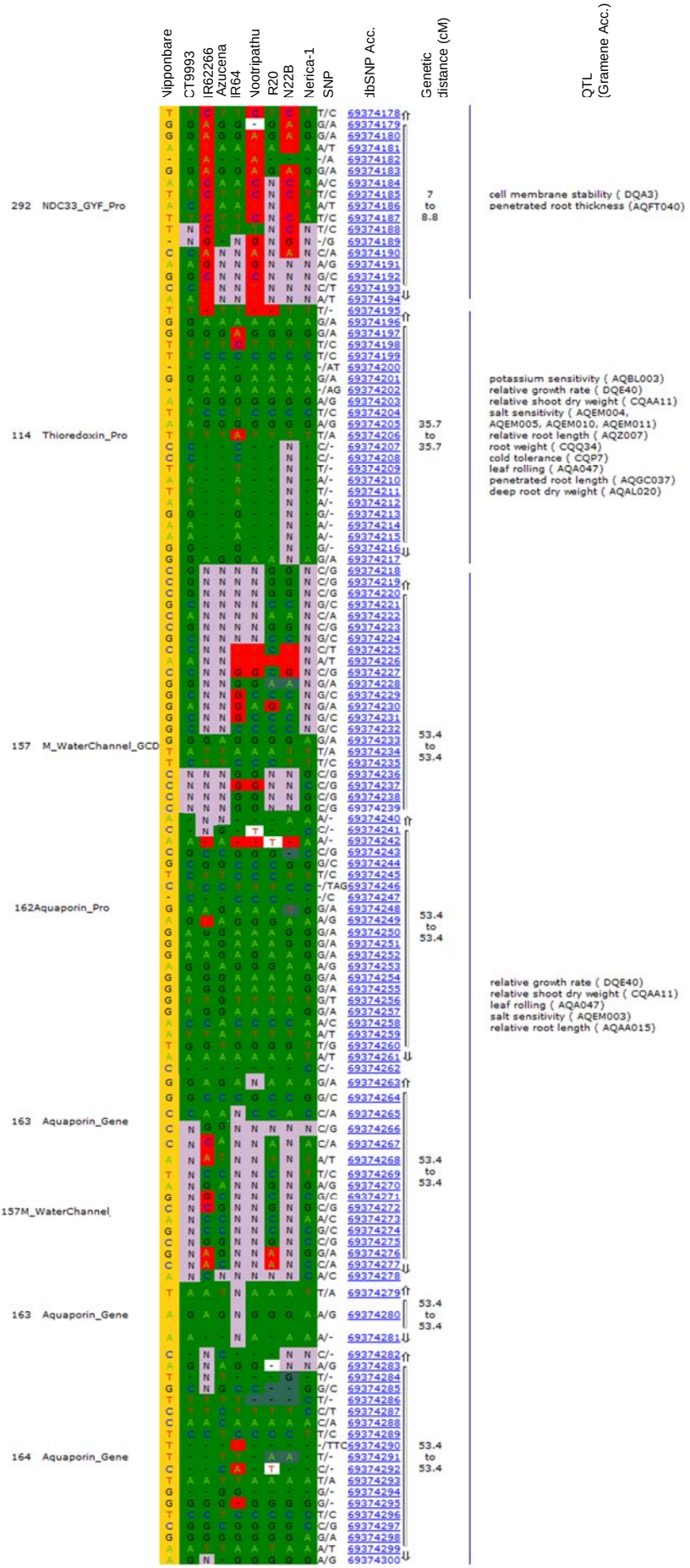


Fig. 4.17G: Integrated allele sharing map of chromosome 7

Conti..

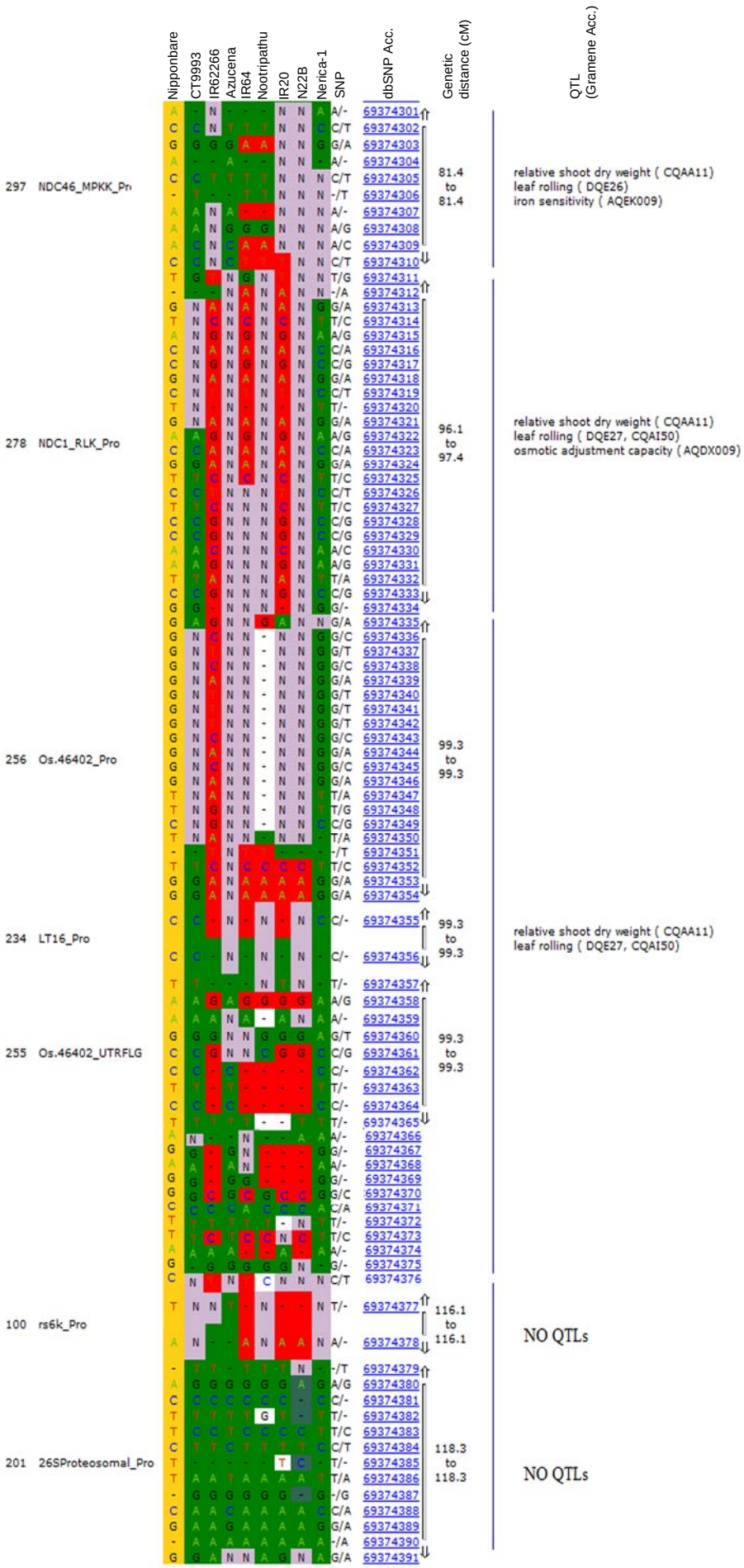


Fig. 4.17G: Integrated allele sharing map of chromosome 7

Continued from previous page

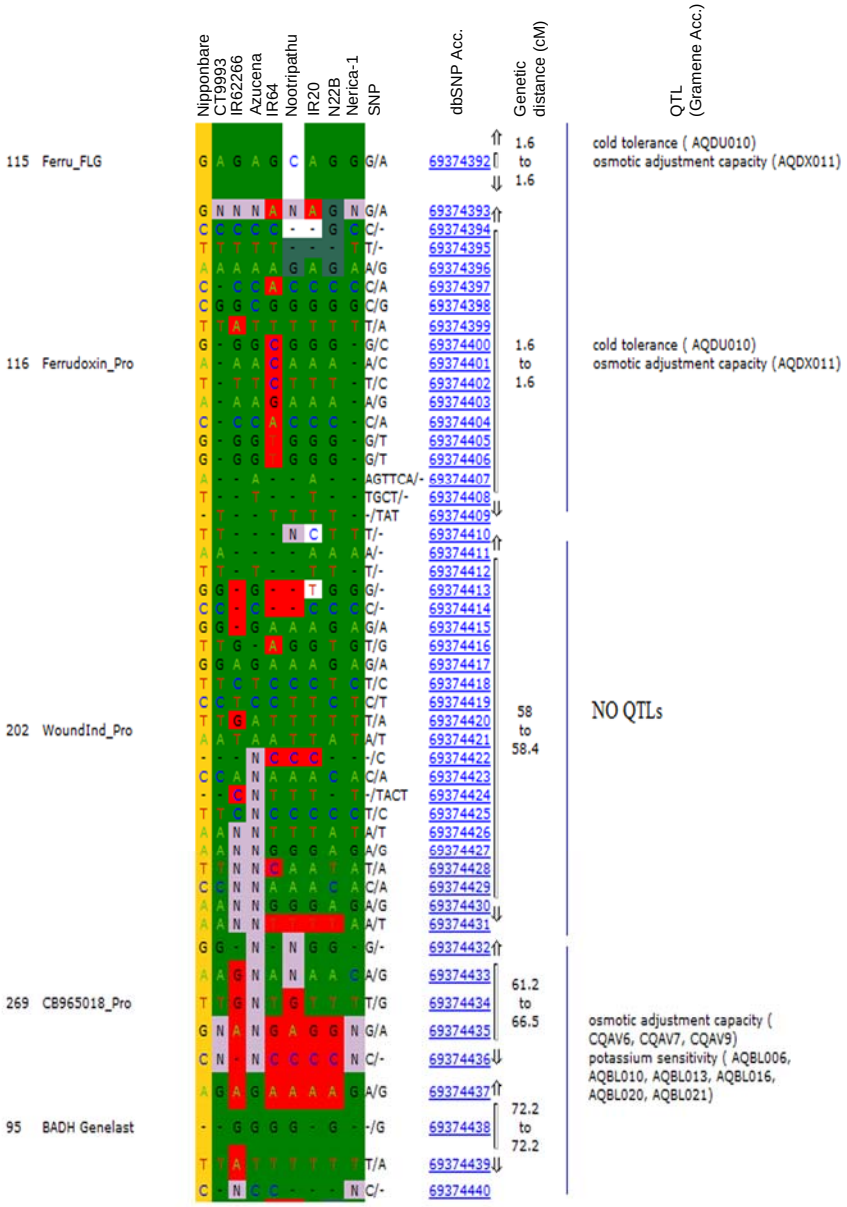


Fig. 4.17H: Integrated allele sharing map of chromosome 8

Conti..

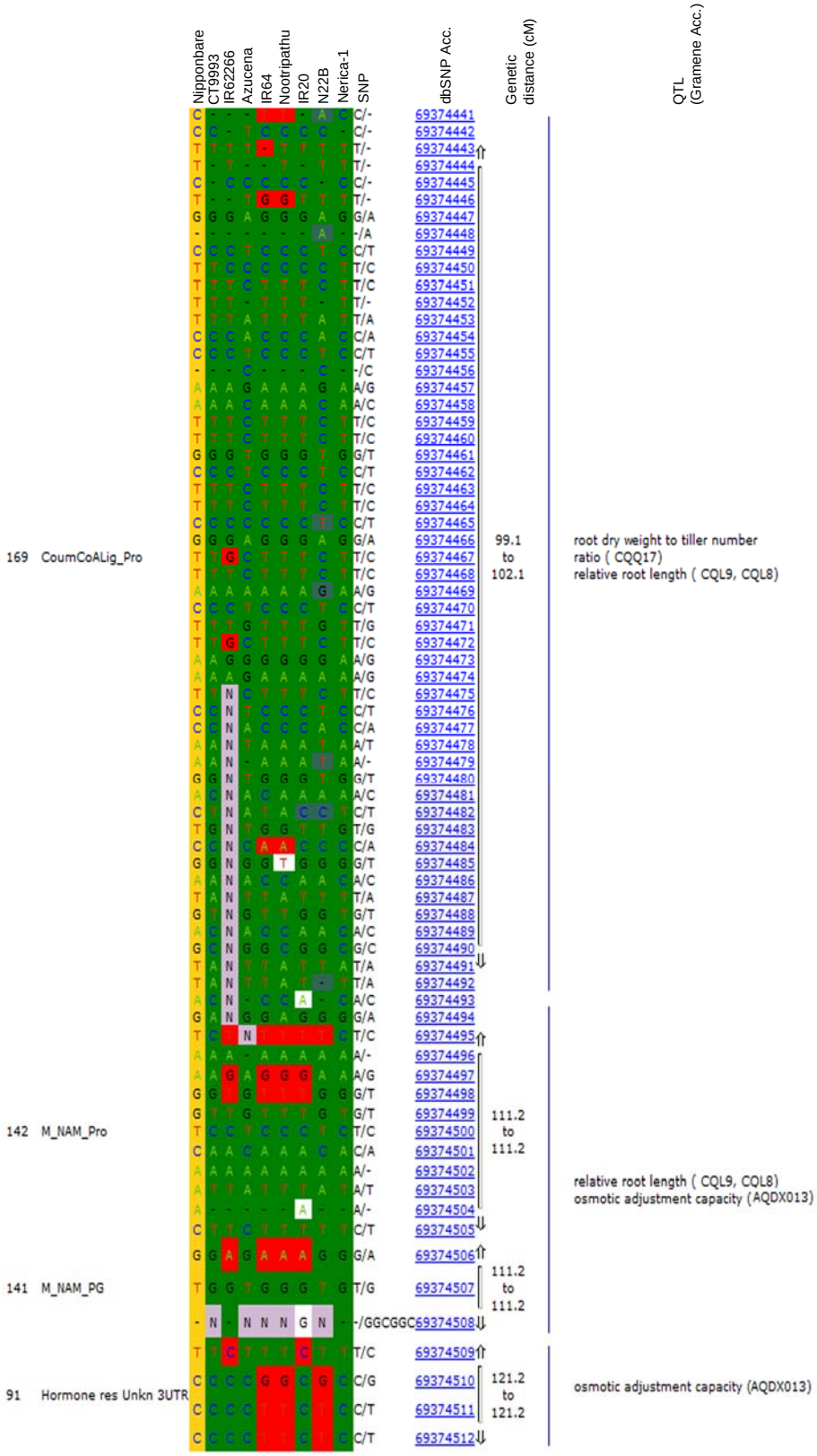


Fig. 4.17H: Integrated allele sharing map of chromosome 8

Continued from previous page

Fig. 4.17I: Integrated allele sharing map of chromosome 9

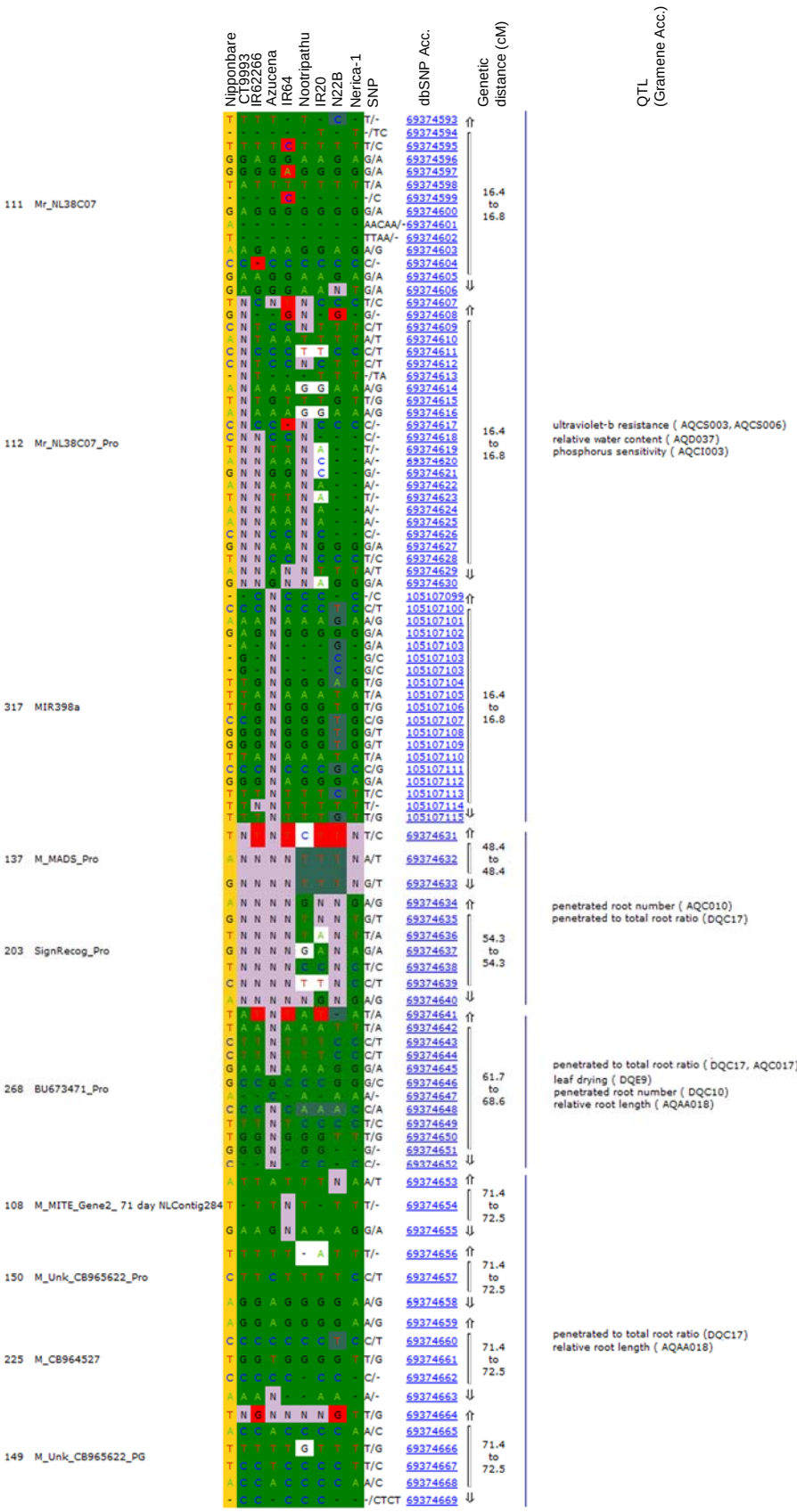


Fig. 4.17J: Integrated allele sharing map of chromosome 10

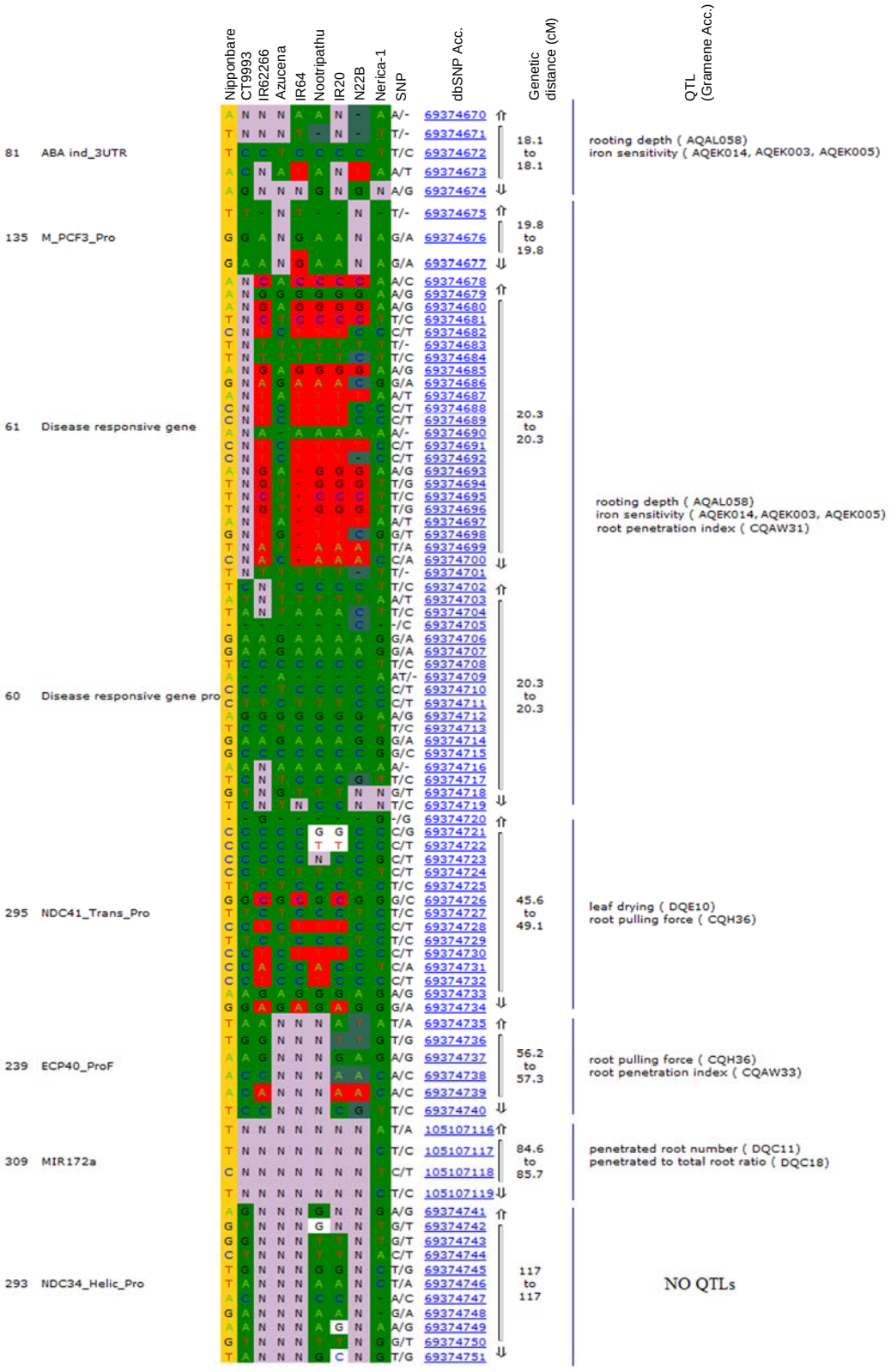


Fig. 4.17K: Integrated allele sharing map of chromosome 11

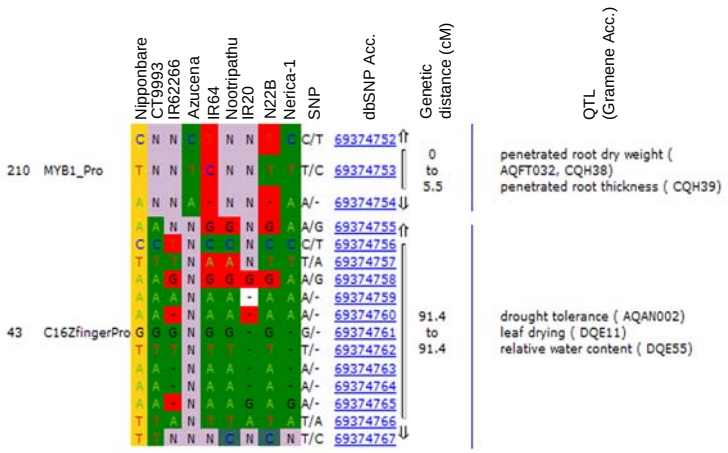


Fig. 4.17L: Integrated allele sharing map of chromosome 12

SNPs identified in the allelesharing map. High quality SNP maps are provided in the attached CD-ROM and also available on an interactive web based intranet server.

4.7 Integration of allelesharing maps with rice QTL data

In order to dissect and associate these SNPs we have integrated abiotic stress responsive QTLs to our allele sharing map. This strategy enabled us to identify candidate SNPs in the genomic regions associated with drought stress response, spanning abiotic stress responsive QTLs. SNPs in the QTL regions associated with relative water content (RWC), osmotic adjustment capacity (OA), leaf rolling, leaf drying, stomatal closure time, drought susceptibility index and dehydration tolerance beside large number of root related QTLs were identified. Majority of the target regions were found to be in the QTL regions associated with leaf rolling, osmotic adjustment capacity, relative root length, penetrated root number, relative water content, deep root dry weight, penetrated root thickness, root weight, leaf drying, penetrated to total root ratio. Few regions were also found spanning the QTLs associated with stomatal closure time, drought susceptibility index, leaf rolling time. The QTL regions spanning the target regions targeted for SNP identification and number of regions associated with each of the abiotic stress QTLs are shown in table 4.9. It has been observed that most of the target regions spans more than one abiotic stress responsive QTL. Further, the SNPs were analyzed in each of the spanning abiotic stress responsive QTLs (Fig. 4.18). This strategy helped in visualizing the SNPs from QTL point of view, dissecting the candidate genes to the single nucleotide level. As the most of the target regions are associated with regulatory regions of the candidate genes, our strategy helped in dissecting the QTL to the candidate genes and in

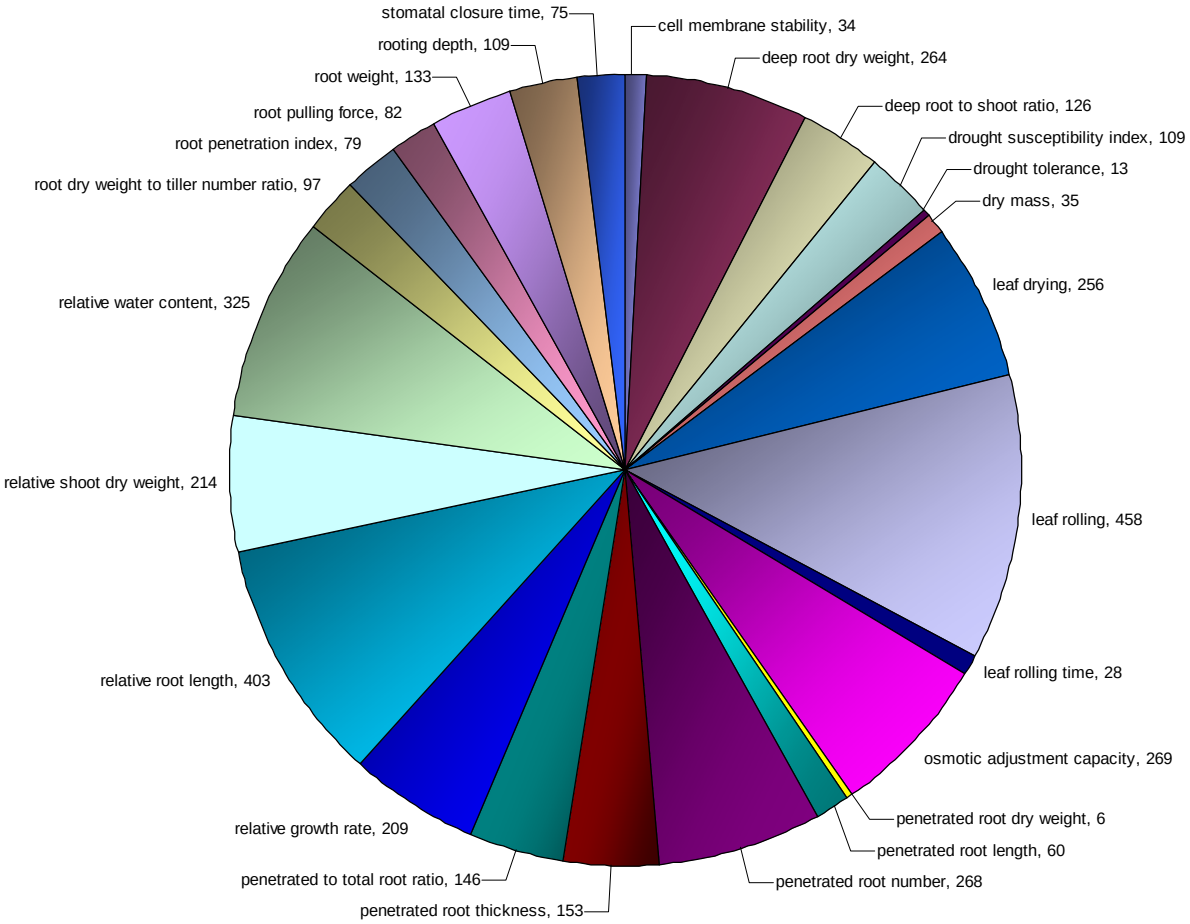


Figure 4.18: The number of SNPs identified in the abiotic stress responsive QTLs spanning at the target regions

identifying SNPs in the regulatory region likely to be associated with variation in gene expression among genotypes resulting in the variation in the phenotype.

S.No.	Name of the QTL	Number of regions spanning the QTL
1	Leaf rolling	40
2	Osmotic adjustment capacity	28
3	Relative root length	27
4	Penetrated root number	25
5	Relative water content	25
6	Deep root dry weight	24
7	Penetrated root thickness	22
8	Root weight	17
9	Leaf drying	17
10	Penetrated to total root ratio	17
11	Relative growth rate	13
12	Deep root to shoot ratio	12
13	Rooting depth	9
14	Root penetration index	8
15	Root pulling force	7
16	Potassium uptake	6
17	Root dry weight to tiller number ratio	6
18	Stomatal closure time	6
19	Drought susceptibility index	4
20	Cell membrane stability	3
21	Leaf rolling time	3
22	Penetrated root length	2
23	Penetrated root dry weight	2
24	Drought tolerance	1
25	Dry mass	1

Table 4.9: The QTLs associated with the abiotic stress response and the number of regions targeted in the corresponding QTL region.

4.8. Identification of SNPs in regulatory regions

It is well established that a set of genes show altered gene expression in a defined stress condition. However, this altered gene expression may not be exactly same in different genotypes having contrast phenotypes under that particular condition. This differential gene expression of a set of genes greatly influences phenotype of the plant which is known to be due to the variation in regulatory regions of one more genes in the gene cluster. The main focus of this study was to identify such variations which are likely to play a major role in phenotypic expression of trait which in turn may be due to a set of SNPs found in promoter regions of the targeted genes. These were identified by extensive analysis for their association with *cis* regulatory regions of which can be effectively converted and functionally associated with the trait. A total of 1203 SNPs identified in the promoter region were analyzed for their association with any of the *cis* regulatory element present in the promoter of the target genes. The analysis of targeted promoter regions with the SNP site at PLACE (Higo *et al*., 1999) database revealed at least 40 SNPs associated with various *cis* elements. Interestingly, twenty three are found to be associated with stress response related *cis* acting elements such as Myb, Myc, CRT/DRE, WRKY, ABRE etc. The list of the SNPs, associated *cis* regulatory element, spanning QTLs and corresponding references were listed in table 4.10. This helped us in functionally establishing the connection between the QTL and regulatory regions finally dissecting the variable regions with contrasting phenotype under study. The consensus of the *cis* element is shown highlighting the site of variation along with the type of variation in each of the candidate SNPs. This set of SNPs are designated as regulatory SNPs

(rSNP) which are likely to be the master switches involved in the target gene expression which can be converted in to potential markers to associate with the trait.

P. No.[a]	QTL[b]	Gene / Promoter	*cis* element	*cis* element sequence (Ref)	SNP[c]
37	SCT, OA	MAPK	ABRE	TACGTGTC (Hobo *et al.*, 1999)	A/G (69373648)
114	RRL, RW, LR, PRL, DRDW	Thioredoxin	COREOS	AAKAATWYRTAWATAAAAMTTTTATWTA (Tsukamoto *et al.*, 2005)	-/AT (69374200)
116	OA,CT	Ferrudoxin	SITEII	TGGCY (Welchen *et al.*, 2006)	CC/AA (69374406)
118	LD, DSI, OA	JA induced protein	EEC LCR1	GANTTNC (Yoshioka *et al.*, 2004)	---/ATA (69373760)
123	RWC, OA	Cold induced protein	ARR1	NGATT (Ross *et al.*, 2004)	G/A (69373966)
142	RRL,OA	NAM	ABRELATERD1	ACGTG (Nakashima *et al.*, 2006)	G/T (69374498)
144	LS,Pan Len	Glycine rich ZFP	WRKY,Myb2, LTRE	TTGAC (Xu *et al*, 2006),YAACKG (Abe *et al*, 2003), CCGAC (Kim *et al.*, 2002)	A/G (69373473), T/C (69373477), G/A (69373475)
159	RRL	DREB1B (CBF1)	WBOX,MYB2,MYB1,ERE	TTGAC (Xu *et al.*, 2006),YAACKG (Abe *et al.*, 2003), WAACCA,AWTTCAAA (Rawat *et al.*, 2005)	T/G (69374561), C/A (NS), a--/CCA (69374553), C/T (69374562)
162	LR, RRL, RGR	Aquaporin	PR10a,ARFAT,Dof	YTGTCWC (Boyle *et al.*, 2001),TGTCTC (Hagen *et al.*, 2002), AAAG (Yanagisawa, 2000)	G/C_G/C_T/C (69374243-69374245), G/C_T/C (69374244, 69374245),T/G_A/G (69374256, 69374257)
167	---	Phenyl ammonia lyase	MART	TTWTWTTWTT (Gasser *et al.*, 1989)	--/TT (69373554)
171	---	Disease responsive gene	WRKY	TTGAC (Xie *et al.*, 2005)	G/A (69373279)
173	PRT	Thaumatin	WRKY, Myb1	TTGAC (Xie *et al.*, 2005), WAACCA (Abe *et al.*,2003)	G/A (69374149), C/A (69374155)
178	LR, PRT, DRDW	APX TL29	TBOX	ACTTTG (Chan *et al.*, 2001)	A/G (69373796)

183	RWC	CBF1	CGBOX,CURE,GATA, CARGCW8	VCGCGB (Yang *et al*., 2002), GTAC (Kropat *et al*., 2005), GATA (Rubio-Somoza, 2006), CWWWWWWWWG (Tang *et al*., 2003)	A/G (69374044), A/G (69374043), A/G (69374039), A/T (69374036)
288	PRT	nifU	Dof	AAAG (Yanagisawa, 2000)	T/A (69374164)
292	Cell mem. stab., PRT	GYF domain containing protein	Myb, MycRD22	CNGTTR (Solano, 1995) ,CACATG (Abe *et al*., 1997)	G/A_A/C (69374183,69374184), C/T (69374185)
227	Pl. hei, leaf wid	Cystein protease	Dof,MybRD22	AAAG (Yanagisawa, 2000),CTAACCA (Abe *et al*., 1997)	-/GAAAGAAAGAAA (NS), A/G (69373464)
234	RSDW, LR	LT16	CRTDRE	RYCGAC (Svensson, 2006)	G/A (NS)
236	PRN,PRT, R.we	WSI18 LEA	CE3OSOSEM	AACGCGTGTC (Hobo *et al*., 1999)	G/C (69373388)
245	R.We,RGR,	aldehyde dehydrogenase	WRKY,SORLPIT	TGAC (Xie *et al*., 2005),GCCAC (Jiao, 2005)	AC/-- (69374522), ----/GCCAC (69374521)
246	---	drought inducible protein (Os.57533)	Dof	AAAG (Yanagisawa, 2000)	G/A (69374515)
248	RRL	DREB1A (CBF3)	GATA,ACGTATERD1	GATA (Rubio-Somoza, 2006), GRWAAW (Buchel *et al*., 1999), ACGT (Simpson *et al*, 2003)	T/G (69374566), T/C (69374564)
146	LD, DSI	Ring H2 finger protein	MYBCOREATCYCB1, CellCycle	AACGG (Planchais *et al*., 2002),CACGAAAA (Nasmyth, 1990)	C/T (69373740), C/T (69373741)
256	RSDW, LR	RCI2B (Os.46402)	CRTDRE, Myb	RYCGAC (Svensson *et al*., 2006), YAACKG (Abe *et al*., 2003)	G/A (69374354), G/A (69374353)
266	PRN, LD, OA, PRN, OA	Tripeptidyl-peptidase (BU672860)	PREATPRODH	ACTCAT (Weltmeier *et al*., 2006)	C/T (69373571)

Table 4.10: Set of SNP found in the *cis* regulatory regions of the target genes. The abiotic stress responsive QTL spanning the target region, name of the *cis* regulatory element were listed along with the associated gene description. SNP identified in the consensus of the *cis* regulatory element is highlighted with red color. QTL names are abbreviated as per the Gramene abbreviations.

[a]- Laboratory accession number of the analyzed region
[b]- QTL name abbreviated as per the Gramene nomenclature
[c]-SNP identified from the present study spanning the corresponding *cis* element.

4.9. Validation of SNPs, and development of SNP-ASPCR markers

The candidate SNPs which were found to be polymorphic between CT9993 and IR62266 were utilized for developing PCR based markers for genotyping candidate SNPs to validate and also to utilize in genotyping. The complexity in typing SNPs was minimized by converting the SNPs into PCR based markers. A set of another 17 genotypes including some of the wild relatives of rice like *O.nivara and O.rufipogan* were also utilized in genotyping along with the 8 genotypes used for the re-sequencing purpose (Table 3.1 and 3.2). The problem of non specific annealing of the allele specific primer at the 3' end was minimized by modification of allele specific base with locked nucleic acid (LNA) wherein a methylene bridge is created between 2' and 4' carbons bringing ribose group into C3' endoconformation (Latorra *et al*., 2003). Further specificity of the reaction was achieved with the addition of varied standardized concentrations of TMAC ranging 10 mM to 60 mM and following optimum annealing temperature conditions. Allele specific bands were designed as scorable markers with ease in genotyping using third allele specific primer along with the primers of the target regions (Fig. 4.19). The presence of allele specific marker was scored in a set of 25 parental lines (Fig. 4.20) and the polymorphic markers between CT9993 and IR62266 were used in genotyping a subset of 152 DHL lines (Fig. 4.21) of the cross. A total of 82 SNPs were considered based on various parameters such as haplotype pattern, allele sharing pattern, association with *cis* elements and spanning the QTL regions for converting them into SNP-PCR markers. Of these, a set of 38 markers were converted into PCR based marker using AS-PCR technique with LNA modification. The list of the markers with their corresponding allele specific primer and allele specific PCR product size along with the positive control used

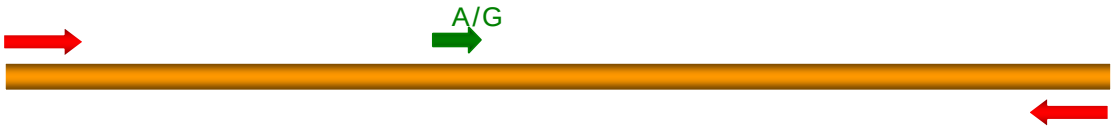


Figure 4.19: Schematic diagram showing the allele specific reaction for converting SNPs to AS-PCR markers

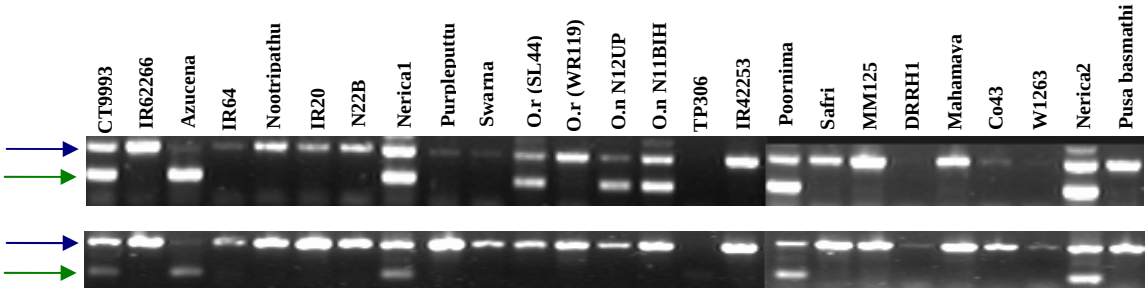


Figure 4.20: Agarose gel image showing allele specific PCR results of SNP in 25 parental lines used for validating SNPs.

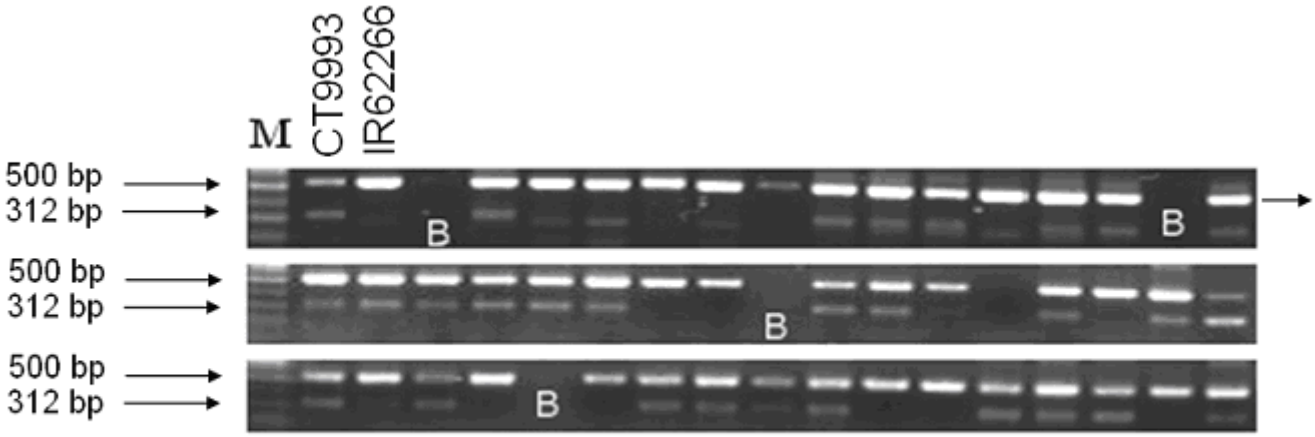


Figure 4.21: Genotyping of DHL population of CT9993 X IR62266 with SNP-ASPCR markers. 'B' indicates blank wells. Upper 500 bp band is the positive control amplified with corresponding forward and reverse primer and lower 312 bp band is the allele specific band.

are listed in table 4.11. The established SNP markers with AS-PCR were used for screening a set of 25 lines (Table 3.2) and scored usability of the marker (Table 4.12). A distance matrix has been prepared using Dice co-efficient values for the scored markers using SNP-ASPCR technique and a dendrogram (Fig. 4.22) was plot with Dice Co-efficient values (Table 4.13). As the criteria for selecting these markers was the polymorphism between CT9993 and IR62266 at the sequence level, the correlation co-efficient value between these two lines was expected and observed to be zero. The dendrogram revealed two distinct groups of genotypes one drought tolerant group having CT9993, Azucena, Nerica-1, Nerica-2, TP309 and a second group consisting of all other genotypes. Besides, Pusa Basmathi, the second group genotypes were grouped into two distinct sub clusters wherein *O. nivara* and *O. rufipogan* were grouped together as single cluster along with DRRH1 which is a member of *O.sativa.* All other genotypes were grouped into a single cluster and N22B was distinctly grouped with in this sub cluster. IR62266 was not considered in evaluating the groups and calculating the distances in grouping as all the markers under study are polymorphic between these two genotypes. The SNPs, which could not be converted in to PCR based markers due to low complexity at the targeted site are converted in to single base extension (SBE) markers using SNuPE technique.

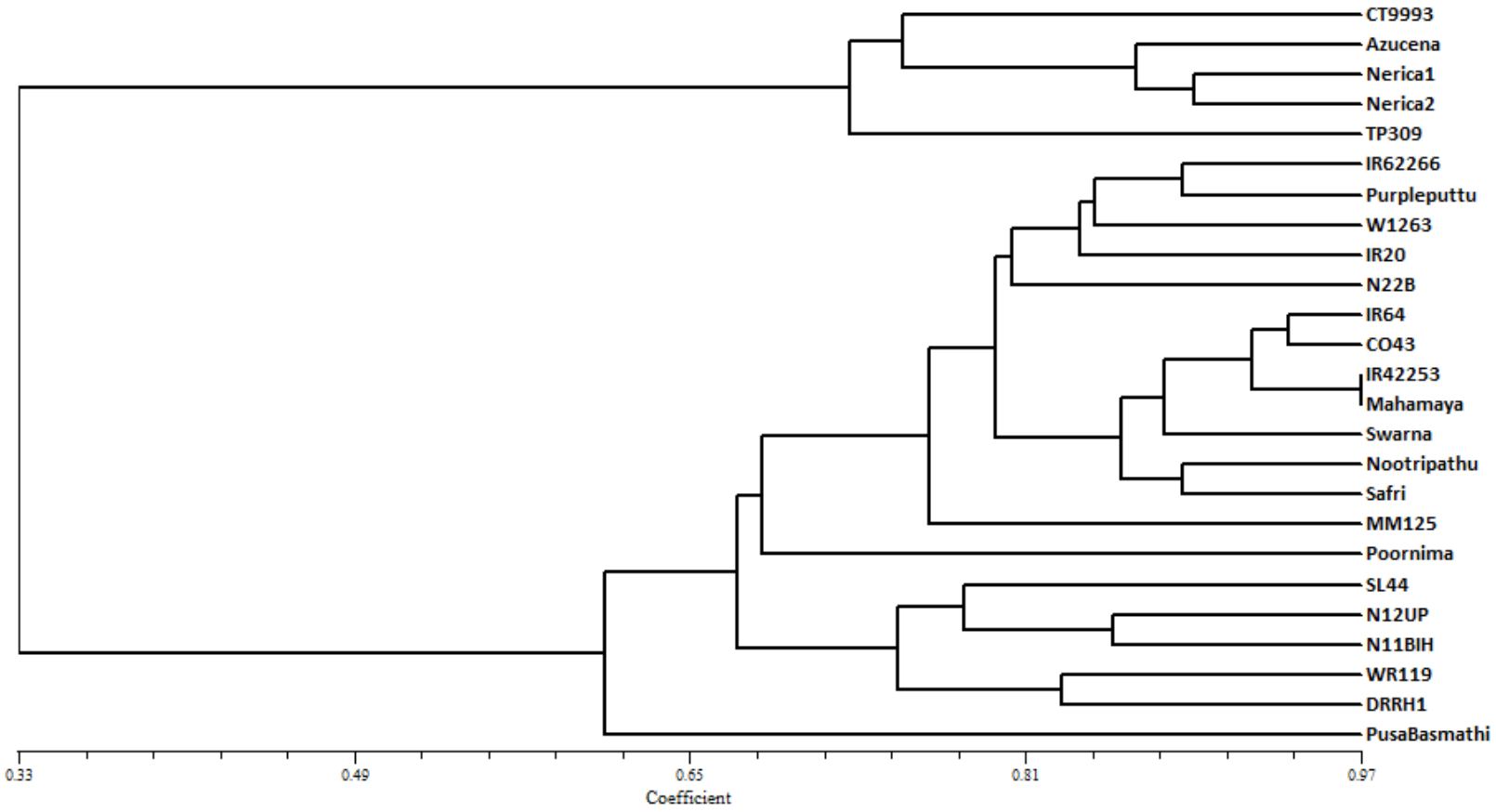


Figure 4.22: Dendrogram of parental lines developed using the SNP-ASPCR markers.

Marker	Description of the region	Region	Chrom.	Forward Primer	Reverse Primer	Allele Specific Primer	CT 9993 Allele	IR62266 Allele	AS-PCR-product
*P37	MAP kinase	P	3	GGATCCAGTAAATTTGGCAG	AGCATGTATAATAGCAGATTATTAGTCAG	CACTCTCCTAGGATACGTG	G	C	
*P49-1	Zinc finger protein	P	6	CCGATGAATCGCCATTAC	TTGGTTGACGGGGAATAC	GCTGTAGTTGTAAAATCCTAAAAATGAAC	AC	--	317
*P40	Ring finger protein	G	1	CTTGAACGGATGTTATCTGC	GCTAAGAATGGAGAGATAATTCAG	GAAACCGAACTTTTTTTTTTCCC	C	-	368
*P67	RNA binding hormone responsive gene	P	10	CCAGCGTTACAGGTCACTTAT	GAGGAGGGCGTACTCTTG	GCCTACGCTACTATTAACCTTACTC	CTC	---	363
*P71	KNOX	G	6	CAGAGAAAAGCTTGTGGACAT	GGAAGCTTCTGTCCGAATC	GGTTGATAAACCAATTATTGATCTGC	T	C	285
*P3	Purple acid Phophatase	G	1	TCGGATTGTCAAATGCAAGTCG	TTCGGGCACGCCATCCTGGAC	GAACCACCTAGATCACGACGCCTGA	A	C	305
*P5	Calmodulin	P	3	TCGACGTTGCACCGTGTGACC	TCGAGCGGTTAGCGAAGAACG	GCACATGTATATACATACCTTCAGACGGC	A	C	768
*P4	RAS	G	6	CCATTGCATCATAACTGGACATCC	TGCTTCAGTAGAAACATGCAGTG	CTTATTGCTGTGCTACAAAAAGTAAACTT	T	A	348
*P49-2	Zinc finger protein	P	6	CCGATGAATCGCCATTAC	TTGGTTGACGGGGAATAC	ATTTTCAGAAGCACTACCAGTTC	TC	--	253
*P7	Ca^{+2} binding protein	G	9	TCGAGCTGCAAAGGAGCTAGC	ATCCGTGCATACGCTGTGTAG	GTCATGTCGCCGTCCAGC	A	C	212
*P65-2	Drought stress responsive Gene	P	11	CCCATCTCTCTCTCTCAGGAC	GGAGAATGAGATGGAAGTGAC	CAATCATTTTTAACATTAATTATTAATAG	GA	--A	387
*P76	Hormone responsive protein	P	8	CAAGTTAGGATCTGGACGC	CATGACTCTTACCATGAATCCTT	GATTTTTCAATCCACCTTAATAATAATTAC	C	T	371
*P83	Inositol phosphate	P	3	CAGTTTTAGTTTCCAGAAATGTTT	ACATTCGAAATATCGGTGATAG	AGTCCATTAACGCGTAATTAATA	A	T	255
*P91	Hormone responsive gene 3UTR	U	8	CTAGCTCATTTGATTGGTGTG	ACACTAAGATAGCAGTGAGCCA	AACGAATGAATAATATACTATATAATGCGC	T	C	309
*95	betaine aldehyde dehydrogenase EL	G	8	CGCGAGCGATGCCAGAG	TTACAGCTTGGAAGGGGATTTGTAC	GACCCCTTTTTTTTGCAAAAGAG	G	A	353
*P111	Mr_NL38C07	G	10	ATGTGGAAGGCGCAGCGG	TCACTCATATACACTCATTGCCAATGG	TCAGCGCAGAGAAGCTCAC	C	T	272
*P114	Thioredoxin	P	7	TGATCGTTTGATTGTTGGATCACAG	CAGGCGATCACGACTCC	ACGACATATTTACGAACACAAAATAATTTGTA	A	G	539
*P121	M_PSII10KDa	P	7	ACTGTCCCGGCAATGTTAGAC	ATGCAGTTCATCATCAGATATCCATGGC	CATACCTACTACACAATAGTCAGTTCATATAAATC	T	C	467
*P120	M_PSII10KDa LE	G	7	ATGGCTGCCTCTGTCATG	CTAAGCCAGTGCACTAGTGTTG	GAAACAAGAATTTTAACACTGCGCG	A	G	195

*P109	M_Unknown gene	G	4	GCCGTACCTTCCAGAGATTTGC	TCACTGCCTGAAATGCTGCAAAACAG	GGTCATGATAGGGTTAATCTATGTGTTA	A	T	247
*P129	M_ AP2 family gene	P	3	AGTGGAAGCTCATCTTTGAAGAGTGG	GTGCGGGTGGACAACGTAG	TTCAATGTCATAAACTTGATAGCGTATAAAATAC	C	-	129
*P144	M_ Glycine rich zinc finger protein	P	2	CGTGTATCGGTACCCGGCAG	CCTCGCCTCTCCCTGATCTGATTAG	AACATGGAGTCAGAAGGACATGGTAATTGG	G	A	671
*P148	M_HSP90	P	3	GATTAAGTCAAAGTGAATTGCACGAG	TGCTATTTATAGGCAGCTCCAGG	AGATAAATGATTAAAAAGGGAGAGAGTTGA	A	G	381
*P158	M_Water channel protein	G	7	GGTGTCCATCAATAAATATCCAACC	GTTTTGTGAGAAGAACAAGAGGATG	GCACATATTTCATAACTCAGCTCAC	C	T	281
*P159-2	DREB1B	P	9	CGGGATCCGCGAGGTAAGCCATTAGCGCATG	CGGAATTCGGATGACTCTCTCTGGTTCAC	GGAAAACGGAATAAACTATCTCCCAATCTC	A	C	612
*P162-2	Aquaporin	P	7	AGAACTCAGCTTAATTAGTTGCCAAATGCG	GCCGCTCTCCATCACCTC	GAGAAATGAATGGTTAGGATTGAAAGG	A	G	635
*P173-1	Thaumatin	P	6	GCCAAACGATGGGGACCTTAG	GCTCCATGTTGGAGTGTAGCC	GTCCGTCACTAGGTGTCTTCCTTATG	G	A	296
*P216	Transcription regulating protein	P	5	CTCGTGTGGACAGTTCATATTTCCTG	AACATTTAGCACTTAGCAGTAGTCTCC	GGCATATTGGAGGAAAAAGTGAAACGG	G	T	354
*P207-1	Enolase	P	3	GGAGGAGGTTGCAAAGAAACAG	GGATCGGATGGATCTTAGATCGGG	GGGATGAAGATTTGAAGAGTGAAGATTTCTG	C	-	390
*P146	M_RING-H2 finger ZFP	P	4	GAGAAAGGGTGAAAGGTTTGAAGTG	GTAGTTGATTTGTTCATATCCATTGGTACG	CTCTTAAGGGTCCGCACAC	T	C	646
*P167	Phenylalanine ammonia-lyase	P	2	TCTAAGGTGTAGTGACCTTGCACTG	GTTGCCCGCCATTGCTAC	AAAGTAAAGTAGGATGGAAGTTATATATC	A	-	272
*P292	NDC33_GYF like protein	P	7	CGTAAGTTTCTTGTGTATGCAGGC	GTGTAACTTTTTATATCGGTCCATCTCC	CAAGCAACAAGTGAACTAGGAAGG	T	G	449
*P288`-1	NDC28_nifU	P	6	CGTTTGCAAAACACTGTTGCTGTG	CGAGTCATCATGGGCCACATTAG	TCATTCCCAGTGGCCCCA	T	A	713
*P288-2	NDC28_nifU	P	6	CGTTTGCAAAACACTGTTGCTGTG	CGAGTCATCATGGGCCACATTAG	GCTGCAGTTCCTCCCTCT	-	C	181
*P236	Water stress induced gene 18	P	1	CGACATGTACCAGTACCATGAATCG	CTCACCAACACACGAACTGAAACTAC	CGGCCACAACGCGTGTCG	G	C	235
*P181	HSP60	P	6	GCGTTTGTGTAAAAGTAGCTAGCAGAAC	AGAGGATAAGGCTGCGACCTC	GGGTGTCAATAAATCTAGACGAGAG			393
*P245-2	Aldehyde dehydrogenase	P	9	GTGTTTGTAGCGTAATTGCGTAAAGTCC	AAGCTCGGCGAGGAACTG	GCGGTGGCTGTGGCTGG	-	T	848
*P168	Anthocyanin reductase	P	4	TCCTATTCAAGCAGCCCCTAAGTC	ACAGTCAAAGTTGAACACGGAAATCC	- (STS)	-	-	840

Table 4.11: AS-PCR markers validated, established between CT9993 and IR62266, and used for segregation analysis in DHL population.

S.No.	Primer No.	NCBI ss#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	37	69373648	+	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-
2	49	NS	+	-	+	-	-	-	+	+	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-
3	49	NS	+	-	-	-	+	-	+	+	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	+	-
4	40	NS	+	-	+	-	-	-	-	+	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	+	-
5	67	NS	+	-	+	-	-	-	M	+	-	-	+	+	+	-	+	-	-	-	-	-	-	-	-	+	-
6	71	NS	-	+	-	+	+	+	+	-	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	-	-
7	3	NS	+	-	+	-	-	-	-	+	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	+	+
8	5	69373273	-	+	-	+	-	-	M	-	-	M	-	-	M	M	-	-	-	+	+	+	-	+	M	-	+
9	4	NS	+	-	+	-	-	-	M	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-
10	7	NS	-	+	-	+	+	+	+	-	+	+	-	+	-	+	-	+	+	+	+	+	+	+	-	-	-
11	65	NS	+	-	+	-	+	-	-	+	-	-	+	-	-	-	+	-	+	-	-	-	-	-	-	+	-
12	76	NS	+	-	+	+	+	-	-	+	-	-	-	-	-	+	-	+	+	+	+	-	+	+	-	+	-
13	83	69373681	+	-	+	-	-	+	-	+	-	-	M	-	-	-	+	-	+	-	-	-	-	-	-	+	-
14	91	69374509	-	+	-	M	-	+	M	-	-	+	-	-	-	-	M	-	-	-	-	M	-	-	M	-	+
15	95	69374437	+	-	+	M	-	-	-	+	-	-	-	+	-	-	M	-	+	+	-	+	-	-	-	+	+
16	111	69374596	+	-	+	-	-	-	+	-	+	-	+	+	+	+	+	-	+	-	+	-	-	-	-	-	+
17	114	69374199	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
18	121	69373257	-	+	-	+	+	+	+	-	+	+	+	+	+	+	-	+	+	+	+	-	+	+	+	-	+
19	120	69373267	-	+	-	+	+	+	+	-	+	+	M	M	-	-	-	+	+	+	+	-	+	+	+	-	+
20	109	69373785	+	-	+	-	-	-	-	+	-	-	-	-	+	-	+	-	+	+	+	-	-	+	+	+	-
21	129	69373622	+	-	+	-	-	-	-	+	-	-	-	-	-	-	M	-	+	-	-	-	-	-	-	+	-
22	144	69373473	+	-	-	M	+	-	-	-	-	+	-	+	M	-	-	+	+	+	+	M	+	+	-	-	-
23	148	69373599	+	-	+	+	+	+	M	+	-	-	-	-	-	-	+	+	+	+	+	-	+	+	-	+	+
24	158	69374235	+	-	-	+	+	+	+	-	-	+	-	-	-	-	-	+	+	+	-	M	+	+	+	-	+
25	159	69374553	-	+	-	+	+	+	+	+	+	+	+	+	M	M	M	+	-	+	+	+	+	+	+	-	-
26	162	69374257	-	+	+	-	-	-	-	+	+	-	M	+	+	+	M	-	+	-	+	-	-	-	-	+	-
27	173	69374149	+	-	+	-	-	-	-	+	-	-	+	-	+	+	M	-	+	-	-	M	-	-	M	+	-
28	216	69373933	+	-	-	M	M	M	-	M	M	M	M	M	+	+	M	-	-	+	-	M	M	1	M	M	M
29	207	69373627	+	-	-	-	+	-	M	+	M	M	M	-	+	+	M	-	-	+	-	M	-	-	M	-	-

30	146	69373741	-	+	-	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
31	167	69373554	+	-	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+
32	292	69374185	-	+	M	-	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
33	288	69374164	-	+	M	+	+	+	-	-	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	-	+
34	288	69374163	-	+	-	+	+	+	+	-	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	-	+
35	236	69373388	-	+	-	+	+	+	+	-	+	+	-	-	-	-	-	+	+	+	+	-	+	+	+	-	-
36	181	NS	-	+	M	+	M	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
37	245	69374521	-	+	M	+	+	+	+	+	+	M	M	M	M	+	-	M	+	+	+	M	+	M	+	-	-
38	164	NS	+	-	+	M	M	-	-	-	M	M	-	M	-	-	-	-	-	-	M	-	-	M	M	-	-

Table 4.12: Validation and scoring of AS-PCR markers in a set of 25 parental lines

	CT9993	IR62266	Azucena	IR64	Nootripathu	IR20	N22B	Nerica1	Purpleputtu	Swarna	SL44	WR119	N12UP	N11BIH	TP309	IR42253	Poornima	Safri	MM125	DRRH1	Mahamaya	CO43	W1263	Nerica2	Pusa Basmathi
CT9993	1.00																								
IR62266	0.00	1.00																							
Azucena	0.79	0.21	1.00																						
IR64	0.18	0.82	0.21	1.00																					
Nootripathu	0.31	0.69	0.22	0.84	1.00																				
IR20	0.14	0.86	0.21	0.85	0.77	1.00																			
N22B	0.22	0.78	0.21	0.79	0.79	0.81	1.00																		
Nerica1	0.73	0.27	0.85	0.24	0.37	0.30	0.35	1.00																	
Purpleputtu	0.11	0.89	0.29	0.78	0.76	0.80	0.83	0.31	1.00																
Swarna	0.15	0.85	0.10	0.93	0.81	0.85	0.79	0.15	0.82	1.00															
SL44	0.34	0.66	0.52	0.54	0.60	0.63	0.70	0.53	0.77	0.60	1.00														
WR119	0.24	0.76	0.39	0.71	0.67	0.68	0.71	0.35	0.88	0.75	0.77	1.00													
N12UP	0.35	0.65	0.45	0.53	0.55	0.55	0.59	0.48	0.77	0.61	0.83	0.81	1.00												
N11BIH	0.31	0.69	0.44	0.65	0.64	0.60	0.65	0.49	0.82	0.63	0.73	0.75	0.85	1.00											
TP306	0.67	0.33	0.77	0.25	0.29	0.40	0.42	0.73	0.41	0.26	0.70	0.48	0.63	0.45	1.00										
IR42253	0.22	0.78	0.26	0.94	0.91	0.86	0.81	0.31	0.82	0.88	0.63	0.76	0.59	0.66	0.34	1.00									
Poornima	0.39	0.61	0.53	0.70	0.71	0.68	0.59	0.46	0.69	0.61	0.59	0.68	0.59	0.61	0.47	0.76	1.00								
Safri	0.32	0.68	0.21	0.88	0.89	0.73	0.75	0.41	0.71	0.79	0.50	0.68	0.62	0.64	0.30	0.84	0.71	1.00							
MM125	0.24	0.76	0.36	0.79	0.77	0.67	0.74	0.39	0.80	0.70	0.55	0.74	0.64	0.66	0.31	0.81	0.73	0.81	1.00						
DRRH1	0.19	0.81	0.25	0.79	0.69	0.74	0.70	0.35	0.77	0.83	0.68	0.83	0.72	0.76	0.44	0.77	0.58	0.77	0.63	1.00					
Mahamaya	0.19	0.81	0.21	0.91	0.89	0.89	0.77	0.30	0.80	0.85	0.59	0.74	0.58	0.66	0.30	0.97	0.73	0.84	0.78	0.74	1.00				
CO43	0.25	0.75	0.24	0.94	0.85	0.80	0.73	0.29	0.76	0.85	0.55	0.71	0.64	0.65	0.32	0.92	0.72	0.92	0.83	0.77	0.91	1.00			
W1263	0.16	0.84	0.21	0.83	0.77	0.84	0.79	0.31	0.84	0.87	0.71	0.70	0.76	0.68	0.39	0.84	0.66	0.81	0.72	0.79	0.81	0.87	1.00		
Nerica2	0.73	0.27	0.88	0.24	0.31	0.30	0.26	0.89	0.29	0.15	0.47	0.38	0.48	0.46	0.73	0.33	0.51	0.35	0.39	0.35	0.30	0.31	0.28	1.00	
Pusa Basmathi	0.38	0.62	0.42	0.73	0.51	0.65	0.61	0.32	0.60	0.70	0.59	0.62	0.55	0.51	0.50	0.64	0.54	0.59	0.53	0.68	0.59	0.63	0.66	0.43	1.00

Table 4.13: Correlation matrix of SNP-ASPCR markers in different parental lines using Dice correlation co-efficiency.

4.10 Conversion of SNPs in to Single Base Extension (SBE) markers

SNPs found in the low complexity regions and low GC content regions were observed to be relatively difficult to establish as markers through AS-PCR technique. Such SNPs were validated using Single Base Extension (SBE) technique. A set of 38 SBE primers were analyzed and using SNuPe assay and 27 found to be polymorphic between CT9993 and IR62266. The representative images of SBE product peaks were analyzed in MegaBACE SNP Analyzer are shown in Fig 4.23a to Fig.4.23h. Fig. 4.23a shows a C/A SNP in Myb1 element of DREB1B promoter deposited with the accession number 69374552 in dbSNP database of NCBI wherein 'C' is observed in CT9993 and 'A' is observed in IR62266 in sequencing and also validated with SNuPe assay. Similarly, dbSNP ss No.69374561 (Fig. 4.23b), 69374041 (Fig. 4.23c), 69374036 (Fig. 4.23d) were validated with this assay, which are associated with various *cis*-regulatory elements of DREB promoter region. Some of the other important observations include 69374109 in RS6K UTR region (Fig. 4.23e), 69373586 in GTP binding RACD (Fig. 4.23f), 69374536 in HSP82 (Fig. 4.23g), 69373604 in HSP 90 (Fig. 4.23h). The set of all the validated SNPs along with the corresponding accession numbers and SNP specific primer sequences were listed in Table 4.14.

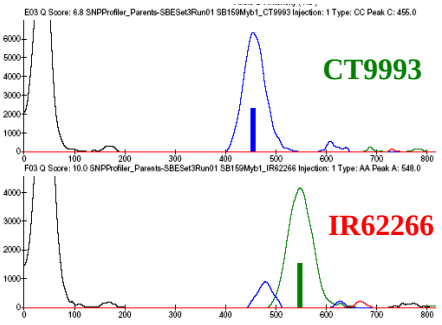


Fig. 4.23a : C/A SNP (69374552) in Myb1 binding site of DREB1B promoter

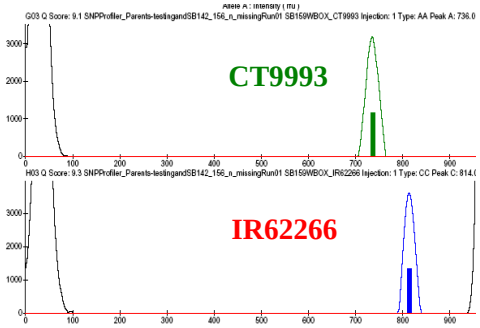


Fig. 4.23b :A/C SNP (69374561) in WRKY box of DREB1B promoter

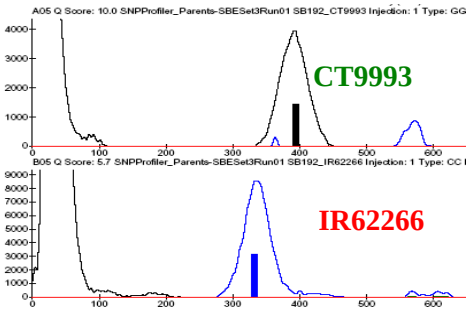


Fig. 4.23c :G/C SNP (69374041) in DREB1A promoter

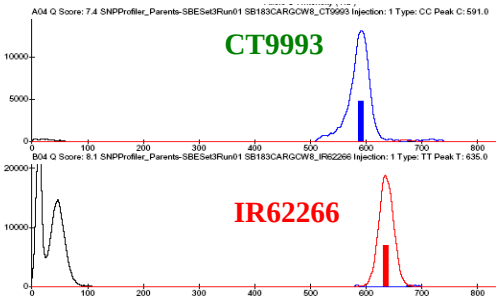


Fig. 4.23d: C/T SNP (69374036) in CARGCW8 element of DREB1B promoter

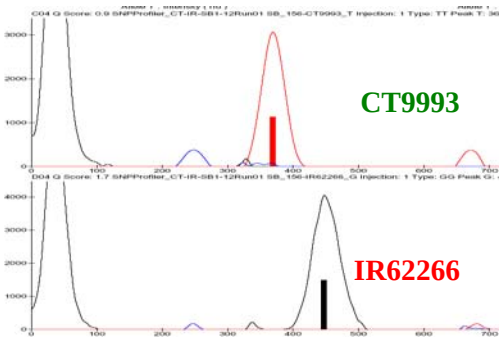


Fig. 4.23e :T/G SNP (69374109) in RS6K 5'UTR

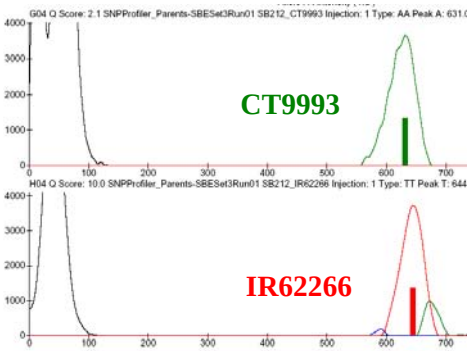


Fig. 4.23f : A/T SNP (69373586) in RACD

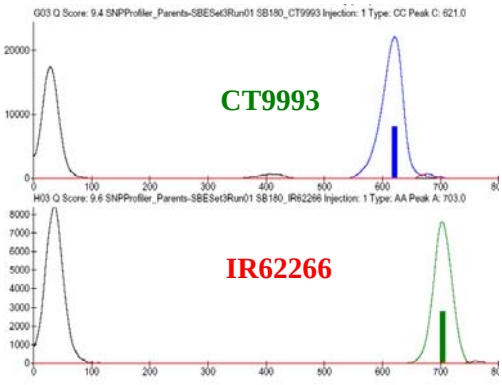


Fig. 4.23g:C/A SNP (69374536) in HSP82 Promoter

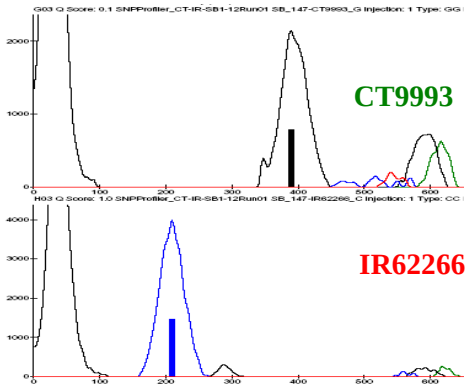


Fig. 4.23h :G/C SNP (69373604) in HSP90 CDS

Figure 4.23: Validation of SNPs with Single base extension in CT9993 and IR62266 lines.

S.No	M.Primer No	SBE Primer Name	SBE PRIMER	SNP Type	Validation Remarks	dbSNP ID
1	147	SB_147	CGCCGTCGGGTGGAG	G/C	Polymorphic	69373604
2	142	SB_142	AAACCGAGTGCGTACGT	G/T	Polymorphic	69374498
3	156	SB_156	GTTGGGGGGCGACGTG	T/G	Polymorphic	69374109
4	169	SB_169	GCTTTGTTCTCCAAGATTCCCAGCA	T/C	Polymorphic	69374450
5	171	SB_171_WRKY	GGTTTAATAGGATCAATTGGATCCATG	T/C	Polymorphic	69373279
6	180	SB_180-2	GGCTACAATCTGAACATGCC	T/C	Polymorphic	69374538
7	183	SB_183_CURE	GCCACTTTGATATGTAATGTAAGCACTGTA	T/C	Polymorphic	69374043
8	183	SB_183_SiteII	CGAGAGTTAAGCCATGGCC	C/G	Polymorphic	69374041
9	192	SB_192	GGGAAGCCATGGGAGAG	G/-	Monomorphic	69374573
10	199	SB_199-2	CCACTATAAATACACGGCGAAGGG	A/G	Polymorphic	69373547
11	224	SB_224-2	CCTAGGGTGGCTATACTAACCGT	C/T	Monomorphic	69373308
12	78	SB_78	TCTGCAATATAATACTAAAGAATTGAC	T/C	Monomorphic	69373327
13	118	SB_118	GTGATTCATGTAAGCTGAGACT	T/A	Polymorphic	69373760
14	123	SB_123	CTGTCTCTGCCACTTCG	A/G	Polymorphic	69373966
15	159	SB_159WBOX	GGAAAGTTGCTTTAAAAAATCATATTAGTC	A/C	Polymorphic	69374561
16	178	SB_178TBOX	CAATGACATGAGAACCCAGACAAAG	A/G	Heterozygous	69373796
17	179	SB_179	TACGAGACGAATCTTTTAAGCCTAATTA	C/T	Polymorphic	69373675
18	183	SB_183CGBOX	CTCTCTCCGTCTCTAAAATTTAACGC	A/G	Polymorphic	69374044
19	196	SB_196	GGCGAGGTCGCTTTCATTAGAAA	A/C	Monomorphic	69373618
20	202	SB_202	GTAAAAGCCCCATCCCTCGATC	T/C	Monomorphic	69374425
21	206	SB_206	CGTAAAATCTCGCCCCACATTTGGA	A/T	Monomorphic	69373608
22	224	SB_224	TAATTTATTTTTGGATGTGTGGTCTAG	C/T	Polymorphic	NS
23	47	SB47_AG	CTGTTCTTCGCTATCAAAGATGG	A/G	Polymorphic	NS
24	61	SB61_GT	AATTGCCAAATTTATACTCCCTCC	T/G	Monomorphic	69374696
25	73	SB73_TG	GAATGAATGGTAACTGGTGTGTATT	T/G	Monomorphic	NS
26	74	SB74_TA	CGTGAATCTGGATAAGCTGGAA	A/-	Heterozygous	NS
27	135	SB135_GA	GTTGGTCAAAGATGCACGCAATA	G/A	Polymorphic	69374676
28	152	SB152_GT	ATGACCAATGCGTTGGTAGAATT	T/G	Monomorphic	NS
29	159	SB159Myb1_CA	AGCCCCAAAAACCCAACAGAAA	GGTTTCTGT/-	Polymorphic	69374552
30	180	SB180_CA	CTGAGTGGATTAAAGTGGTTTAGAAT	-/AT	Polymorphic	69374536

31	183	SB183CARGCW8_CT	TTGCAACGGGTGAATGCTATTTTAAT	G/A	Polymorphic	69374036
32	205	SB205_AG	TCTTGTGTAATCTTGGAACGTGTC	T/C	Polymorphic	69373351
33	212	SB212_AT	AAATCGTATTCCTTTATGGACTCT	-/AGTCCCCC	Polymorphic	69373586
34	192	SB192_GC	ACTATGGTTTCCTTGGGCAT	G/-	Polymorphic	69374573
35	202	SB202_AG	TAGCTATGCCACTGTAGGATAC	T/C	Polymorphic	69374418
36	155	SB155_CT	ACGCCATGGTTGAATGCCA	G/A	Polymorphic	69374116
37	162	SB162PR1_GC	CAACACGCAACAAATCTCCTCT	C/G	Polymorphic	69374243
38	173	SB173Myb1_GT	TACAAACGAATCTAGACAGATG	C/A	Polymorphic	69374155

Table 4.14: Set of validated SBE markers along with their dbSNP accession numbers and polymorphism data (‘-‘ indicates the deletion; M.Primer No. refers to the Primer number of the amplified target region)

4.11. Segregation analysis and updating genetic map of IR68586 DHL population

A double haploid population (DHL) of rice with 154 lines was extensively utilized in conversion of these markers and segregation analysis. The allele specific PCR products were run on 1-2% agarose and gel allele specific PCR bands were scored. A positive control was always used for each of the PCR reaction either by using a third primer or a standardized set of primers at the specific PCR conditions. This data was utilized in updating genetic map of IR68586 population using set backbone markers for the population (Chandrasekhar *et al*, 2005). A set of 13 markers were mapped onto the existing map with reported backbone markers Fig. 4.24

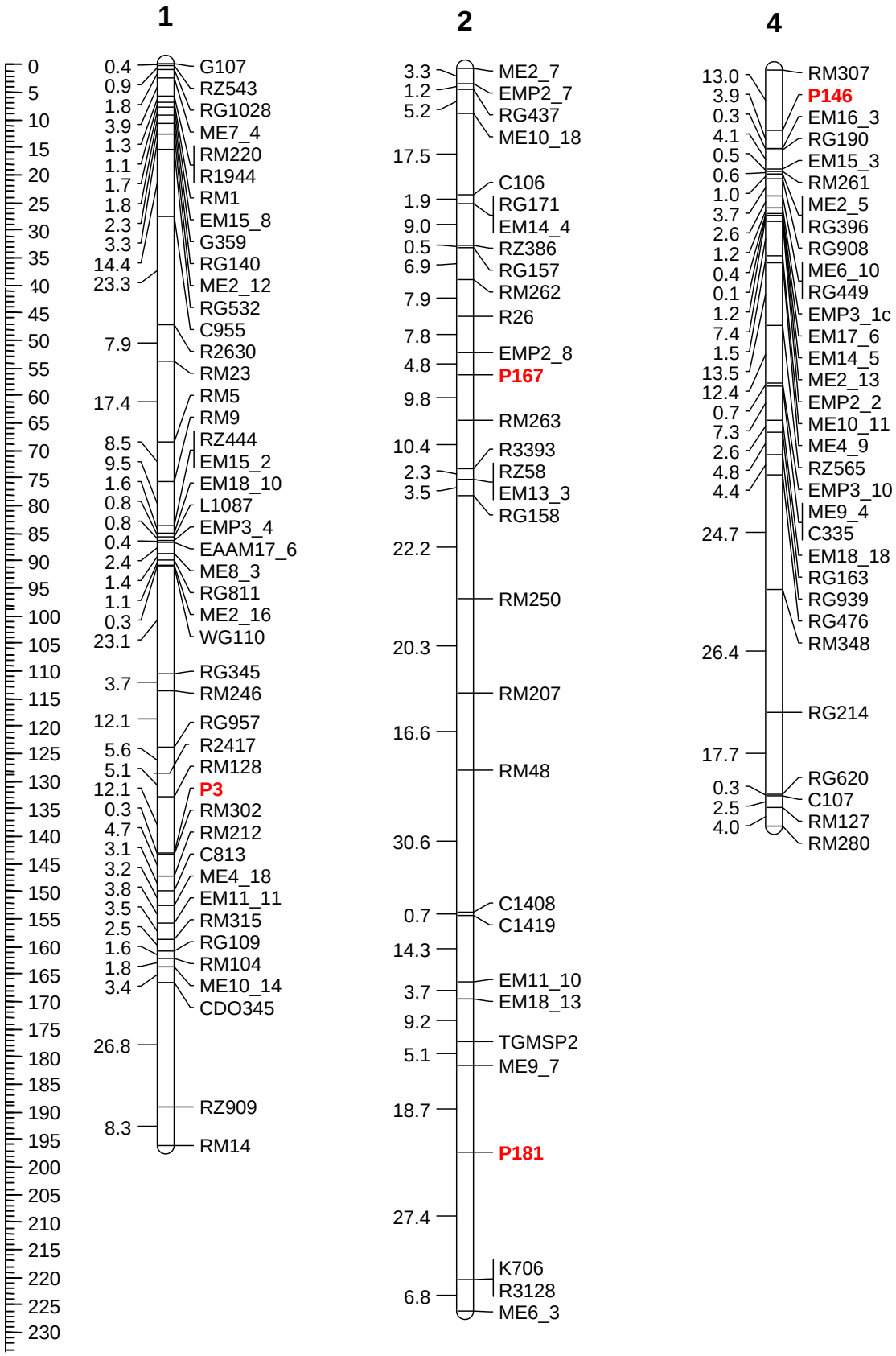


Figure 4.24: Updated genetic map of chromosome 1, 2 and 4 of IR68586 population with SNP markers.

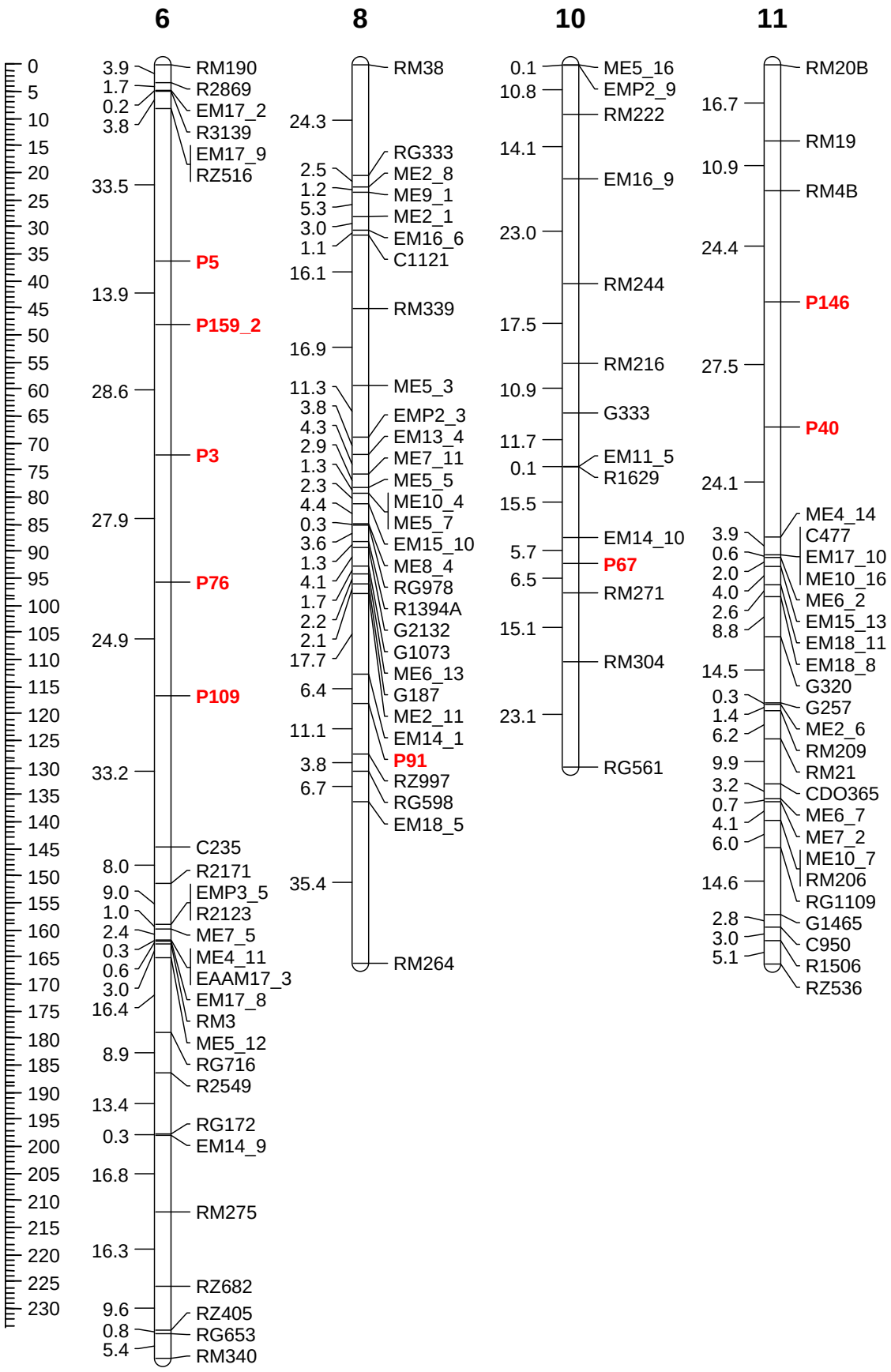


Figure 4.25: Updated genetic map of chromosome 6, 8, 10 and 11 of IR68586 population with SNP markers.

5. DISCUSSION

5. DISCUSSION

5.1 Candidate genes for stress response were identified using a multi-disciplinary approach

A multi-disciplinary approach was taken up in the present study to identify the regions associated with the drought stress response and attempted to dissect these regions to identify single nucleotide polymorphisms. The approach includes most appropriate methodologies such as microarray gene expression profiles, subtractive cDNA libraries and comparative analysis of genes identified from the normalized cDNA library besides the regions associated with post transcriptional regulation. As the drought tolerance known to be a complex, multigenic trait it is difficult to associate a set of SNPs in single locus to the trait. Hence, it is highly essential to scan multiple set of genes in diverse pathways associated with the trait in order to derive a set candidate SNP which may be correlated to the trait under study using molecular genetic approaches. The genes identified from subtracted cDNA library of Nerica-1 with IR64 revealed some novel stress responsive genes including some known stress responsive genes. Though we could efficiently capture differentially expressed genes in Nerica-1 genotype under drought stress condition, the sequencing of target cDNAs revealed the library is enriched with genes encoding rRNA genes besides abundantly present RUBP carboxylase which is commonly seen in most of the differential libraries not subjected to normalization. The set of genes identified to be differentially expressed include kinases and phasphatases such as serine threonine kinase homologous to NM_001066888, phosphoglycerate kinase homologous to AAU44053, mitogen activated protein kinase kinase (MAPK2) homologous to NP_001060061, sac domain-containing inositol phosphatase 3

homologous to NM_001063583, transcription factors like Squamosa promoter binding like gene homologous to NM_001049328, dnaJ like protein homologous to NM_001057314, GYF domain containing protein. The genes belonging to transport associated proteins include Na+/H+ antiporter, GABA specific permease, carbohydrate transporter etc. besides a set of stress responsive defence protein such as senescence associated genes homologous to AAO72638, Avr9/Cf-9 rapidly elicited protein homologous to BAD27662, ER1 ethylene responsive protein, hevein like defence related protein homologous to NP_001068190. This represents genes associated at various levels of signal transduction pathway to activation of downstream cascade of genes mediated through a set of transcription factors. Further, the differential gene expression observed between these genotypes may be attributed mostly to the variations in the promoter regions of some genes, particularly the regulatory regions of the promoter which regulate the gene expression. Hence, the promoters regions of the selected genes were also screened for SNP identification in the present study and resulted in very informative and interesting pattern of haplotype. Further, the extensive analysis of microarray data of Nagina22 genotype enabled us to populate a list of candidates showing dramatic changes. Though a large number of co-regulated genes were identified, the application fold change parameter with a cut off value of 4 fold limited us to select small group of genes. Since it is difficult to target all the members from a cluster, a few representative genes were selected. This resulted in identification of a small set of candidate genes representing the cluster with possibly having a common or similar regulatory mechanism. Our microarray gene expression profiles suggest at certain stress regimes during specific developmental stages show dramatic down regulation of genes. The set of genomic regions analyzed

based on microarray expression profiles include protein kinases such as MAP kinase, MAP kinase kinase, ribosomal 6 kinase, serine threonine kinase, lectin like kinase, CDP kinase and other receptor like protein kinases. The regions associated with the signal transduction mechanism include RAS GTP binding protein, RACD GTP binding protein, RIC1 GTP binding protein, Ca^{+2} calmodulin binding protein, inositol phosphate transferase. Signal transduction genes also induced by hormones include ABA induced gene, Ethylene responsive genes, jasmonic acid induced genes. The regions associated with transcription factor activity include DREB1A, DREB1B, DREB2, CBF1, CBF3, CBF4, KNOX, zinc finger binding protein, dnaJ, NAM, R2R3 Myb binding protein, PCF3, WRKY transcription factor and other member of AP2 genes family. The downstream set of genes activated by the transcription factors include set of genes associated with detoxification process such as glutathionein s transferase, thioredoxin, catalase, proline rich proteins, glycine rich proteins, thaumatin, SAM synthase, SAM decarboxylase, APX, metallothioneins and cytochrome P450. The known downstream genes associated with the drought stress response include dehydrin, rd22, late embryogenesis abundant proteins (LEA) and aldehyde dehydrogenases (Adh). The chaperonins and heat shock proteins associated with stress response process include cyclophilins, peptidyl cis-trans isomerases, heat shock protein 90 (HSP90), HSP60 and HSP82 etc. The structural genes playing a key role in ion and other solute transport include genes such as aquaporins, Na^{+}/H^{+} antiporter, water channel proteins, proton pump interacting genes etc. Glycine betain, and its precursors, choline and glycine betaine aldehyde are known to be involved in the drought stress response through osmotic protection. Hence, the set of genes associated with glycine betain biosynthesis

such as choline monoxygenase, betain aldehyde dehydrogenease were targeted. The stress response process include activation secondary metabolic pathway initially in order to produce secondary metabolites which play a key role in defense mechanisms. In order to identify SNPs associated with the stress induced secondary metabolic pathway genes such as phenyl ammonia lyase (PAL), coumaryl co-A ligase, anthocyanin synthase were analyzed. Few regions associated with photosynthetic activity were also targeted such as PSII 10 KDa protein, ferrudoxin etc, besides other stress responsive genes such as cold induced genes, low temperature induced genes, gigantia like genes, pathogen responsive (PR) genes. The recent advances in transcriptome analysis revealed the role of miRNAs in post transcriptional regulation particularly in negative regulation of many target genes. This kind of negative regulation could be mediated through the involvement of miRNAs for specific target genes and intern mediate alteration of a cascade of genes. The regions encoding miRNAs, particularly, the coding space involved in stem loop and mature miRNA formation is believed to be evolutionarily highly conserved. Hence, we attempted to identify members from specific miRNA families and targeted upstream regions of these genes. These factors enabled us to focus on dissecting the promoter regions rather than the coding space which is more likely to play a major role in biogenesis of miRNA. The set of miRNA genes analyzed which are associated with stress response include MIR159a, MIR172a, MIR172d, MIR319a, MIR398a and MIR398b.

Unlike other recent reports (Monna *et al*., 2006; Nasu *et al*., 2002) wherein large number of intergenic regions was randomly targeted to find thousands of SNPs, our intention was to target few regions but associated with the trait under study to associate them with the phenotype (Fig. 5.1). In the present study we also attempted to analyze the

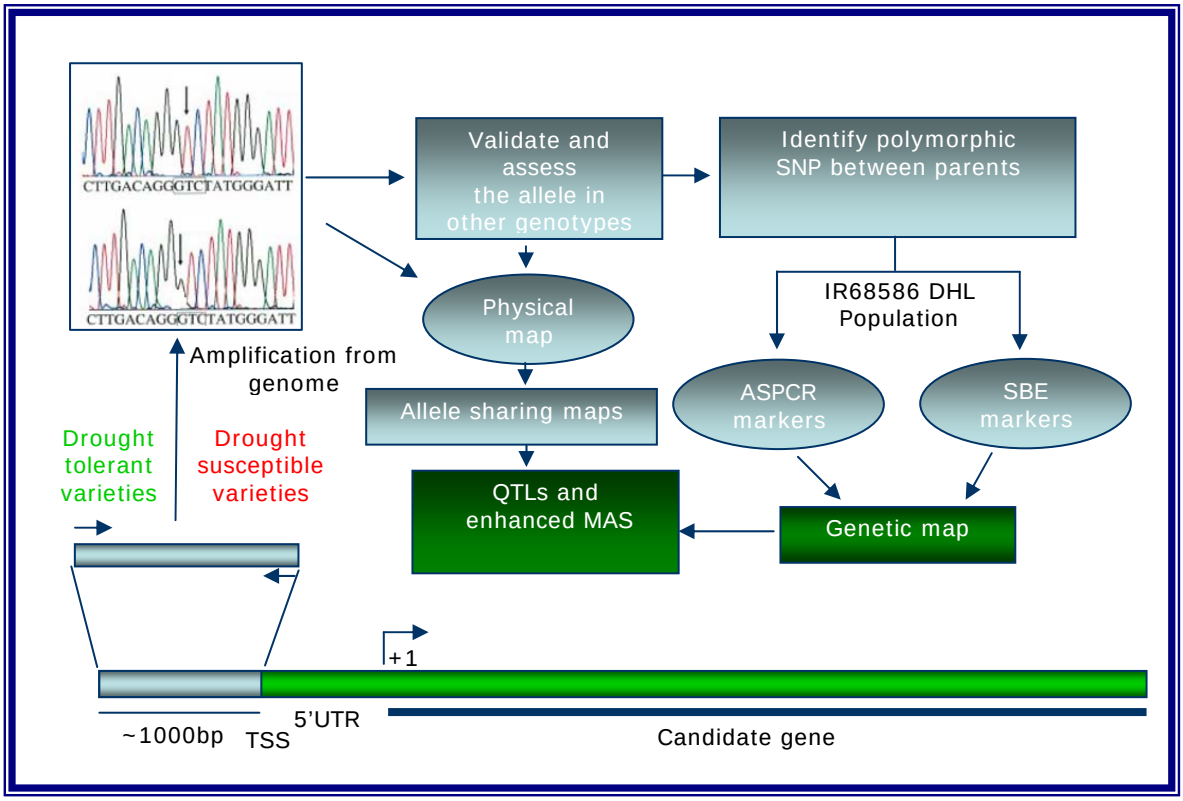


Figure 5.1: The experimental plan for identification of regulatory SNPs.

set of target regions with available sequence data in the public domain using a genotype specific clustering approach with the help of custom Perl scripts. This approach and scripts helped in handling the redundancy of the sequence information at the target locus at some of the target loci though comprehensive information of any of the target loci could not be obtained for all the target genotypes under the study. MySQL database, Perl programming language and bioinformatics approach helped us to handle the huge amount of data to utilize it in our programming pipeline. The resources from public databases such as dbEST, unigene and nucleotide divisions of NCBI helped us to curate the database with all the required information.

5.2 *In silico* analysis of unigenes is inadequate for trait based dissection of genomic regions

Our initial analysis with available unigene data from the public domain at the target regions revealed a good number of SNPs. But the inherent problems with the approach such as limited availability of the data at the target loci, incomplete details of source genotype and redundant sequence information of certain ESTs restricted our analysis to compromise with available data initially. Though most of such issues were addressed by gtCluster program developed as part of the current study we preferred sequencing of target regions from the selected genotypes mainly due to unavailability of sequence information at the majority of our target regions (i.e. 5' upstream regions of the target genes) from the selected genotypes. gtCluster program helps in analyzing the unigene clusters with polybayes program if the chromatograms of the corresponding regions are available for at least one sequence under study. The problem with highly redundant sequences was well addressed with the program using CAP3 (Huang and Madan, 1999)

in order to develop the consensus based on the redundancy score. Further, the program was developed to present the results in a web interface dynamically which facilitates ease in scanning the identified SNPs at each locus. Though the program was developed for Linux operating system (Redhat Linux 9.0), it is portable to other operating systems as the code was written using Perl language with minor modifications to the script for compatibility. Though we could get the sequence information of 14 percent of the targeted sequences again it found to be inadequate due to unavailability of data for selected genotypes. Within the available sequences most of the sequences were found to have sequence data from Nipponbare only hence not much information could be obtained from the *in silico* analysis for the other genotypes. From the 432 SNPs identified from such analysis most of them were transversions and a very few were indels. Though, probability filter of 0.99 was used to obtain these SNPs, these were not submitted into the public domain as the sequence information could not be verified at the trace level using chromatograms.

5.3 Dissecting targeted regions for identification of SNPs through re-sequencing approach

The SNPs are well recognized as markers of choice due to their high frequency, availability different methodologies with amenability for automation. SNPs are known to occur at one for every 154 bases in rice (Liu and Zhang, 2006) between indica and japonica subspecies and they can be found in closely related cultivars as well. The genotypes we have selected for SNP identification are phenotypically well characterized and widely used as parental lines in breeding programs in India and worldwide for drought tolerance. Based on the response to drought stress we have categorized these

genotypes into three panels, the first panel represent highly drought tolerant CT9993, Azucena and Nerica-1, the second panel is represented by drought susceptible genotypes, IR62266, IR64 and IR20, and the third panel is represented by moderately tolerant N22B and a landrace Nootripathu. Though a set of 300 genomic regions were targeted for the analysis in the present study we are successful in amplifying only 50% of the targeted regions. Various parameters were standardized in order to obtain a single amplicon in the PCR reaction. Particularly, annealing temperature was standardized optimal for primer annealing for each of the primer set after a gradient PCR reaction. Though the amplification was successfully obtained in more than 85% of the targeted regions approximately 35% of the amplicons were resulted in multiple bands even on rigorous standardization changing various parameters and hence eliminated as they can not be used for direct sequencing. Various adjuvants were used in order to obtain a single amplicon by increasing the specificity of the reaction. Of the various adjuvant tried tetra methyl ammonium chloride (TMAC) was found to be most appropriate to increase the specificity of the reaction. Though a concentration range of 10 mM to 40 mM was tried, for most of the PCR reactions 20 mM of TMAC was found optimal to increase the specificity of the reaction. Dissection of successfully amplified target regions from three panels of genotype revealed many SNP having a common pattern sharing with in the panel. Though a set of 8 genotypes were used for the sequence characterization, a set of 16 genotypes were used initially in order to asses the appropriateness of the target regions for the trait under study and possibility of correlation with the phenotype (Table 4.6). The association of targeted regions to the panel of genotype was evidenced by the analysis one of the randomly selected region. A dendrogram was drawn with the sequence

information obtained from mitogen activated protein kinase UTR sequence and SNPs identified on comparison with 16 different genotypes falling into 3 panels (Fig. 5.2). The dendrogram reveals distinct group of genotypes falling into 3 major clusters. Cluster 1 represents highly drought tolerant genotypes such as CT9993, Azucena and Nerica1 which are known to be highly drought tolerant cultivars. CT9993 and Azucena are tolerant parents widely used in mapping the traits associated with drought stress response. Nerica-1 is an interspecific hybrid of *Oryza sativa X Oryza glaberrima* for which a subtractive cDNA library has been generated in the present study. However, another line of Nerica i.e. Nerica-2 was found to be in the second cluster. Cluster 2 represents genotypes such as N22B, Nerica-2, Nootripathu and DRRH1 which are identified as moderately tolerant when compared to the panel of highly drought tolerant lines. The third cluster represents almost all drought susceptible lines such as IR62266, IR42253, IR20, IR64, Swarna, Mahamaya and Safri. This preliminary result has provided leads to restrict the number of genotype analyzed to eight based on the haplotype patterns also provided fundamental evidences of appropriateness of the target regions. The sequences were sorted based on the genotype (Table 4.6) and deposited in the nucleotide division of NCBI instead of GSS (Genome Scan Sequences) division as most of the characterized sequences are upstream (5' UTR or Promoter) or downstream (3' UTR) sequences associated with a gene. Application Bayesian algorithm for screening of SNPs in aligned DNA sequences helped in eliminating the most unlikely SNPs based on the probability values. As most of the targeted regions belong to 5'UTR or promoter regions majority of SNPs were identified belonging to promoter regions. A significant proportion of SNPs were found to be insertion deletions as multinucleotide variations were identified in most

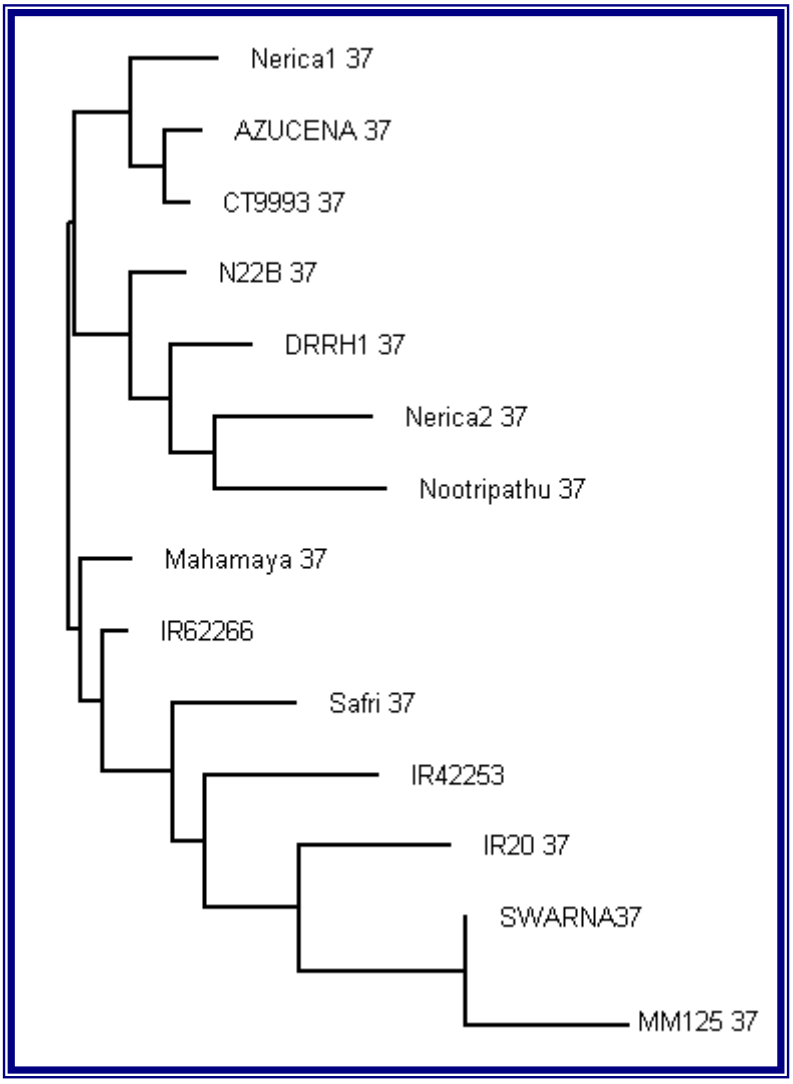


Figure 5.2: Assessing the appropriateness of the target regions – candidate gene haplotyping approach

of the cases. Though a set of 587 SNPs were identified between CT9993 and IR62266 a set of 65 SNPs were converted to into markers either by AS-PCR or SBE technology which are having high probability score. SNPs were identified in members of different gene families like receptor kinases, protein kinases, signal transduction pathway genes, transcription factors, osmoprotectant associated genes, structural genes and drought stress responsive cascade of downstream genes. Some of the interesting observations include SNPs identified in members of protein kinase and signal recognition protein family such as signal recognition particle (dbSNP ss# 69374634-69374640), receptor like kinase (69374311-69374334), lectin like kinase (69373389, 69373390), CDPK (69373791-69373794), ribosomal 6 kinase (69374377 and 69374377), MAP kinase kinase (69373482, 69374301-69374310), MAP kinase (69373636-69373657) and serine threonine kinase (69373606-69373609). SNPs identified from signal transduction gene family include variations spanning in the genomic regions associated with RACD GTP binding protein (69373575-69373592), RIC1 GTP binding protein (69373330-69373347), RAS GTP binding protein (69374070-69374106), EF hand Ca^{+} binding protein (69373506, 69373507), ABA induced gene (69374670-69374674), Inositol phosphate transferase (69373681-69373686), ethylene responsive gene (69373939-69373942). SNPs were also identified in some of the well known transcription factors associated with stress response process such as ring finger protein promoter (69373395-69373397), CBF4 (69374018-69374032, 69374050-69374069), AP2 family transcription factor (69373621-69373625), DREB2 (69374517-69374519), CBF1 (69374033-69374045), Heat shock induced transcription factor (69374169-69374175), CBF3 (69374544-69374551, 69374553-69374555, 69374571-69374577), WRKY transcription

factor (69373797-69373818), Myb1 (69374752-69374754), Transcription regulating protein (69373932-69373938), AP2 family gene (69373563-69373565), DREB1A (69374554, 69374563-69374570, 105111356-105111360, 69374046-69374049) , R2R3Myb binding transcription factor (69373852-69373881) and KNOX (69374118-69374139). The SNPs identified from osmoprotectants and detoxification gene family include thioredoxin (69374195-69374217), catalase (69374176, 69374177, 105111345-105111355), thaumatin (69374140-69374161), APXTL29 (69373795, 69373796), APX (69373658-69373678), cytochrome P450 (69373348-69373384), proline rich protein (69373913-69373918), glutathionein s transferase (69373610-69373614), betaine aldehyde dehydrogenase (69374437-69374439), choline monoxygenase (69374167, 69374168, 105111398-105111420), metallothionein (69373309-69373322). SNPs identified from chaperonins and other stress induced genes include HSP90 (69373593-69373605, 69373593-69373601), HSP60 (69373977-69374008), HSP82 (69374530-69374540), Peptidyl-propyl cis-trans isomerase (69373572-69373574), cyclophilin (69373679, 69373680, 69373699-69373703), pathogen responsive protein (69374009-69374017), disease responsive gene (69373276-69373282), wound induced gene promoter (69373287-69373301), dehydrin (69373247-69373253, 69373255, 69373256, 69373260, 69373778-69373780), SAM synthase (69373882, 69373883), SAM decarboxylase (69373542-69373550), wound induced protein (69374410-69374431), defensin (69373269-69373272, 69373274, 69373275), chitinase (69373919-69373931), cold induced gene (69373962-69373967), disease induced protein (69373433-69373443, 69373508-69373514), low temperature induced protein (69374355, 69374356), water stress induced protein 18 (69373385-69373388), late embryogenesis abundant protein

(69373968-69373976), aldehyde dehydrogenase (69374520-69374529), early drought induced protein homologous to Os.13968 (69373884-69373893), drought induced hydrophobic protein homologous to Os.46402 (69374357-69374376, 69374335-69374354), drought induced protein homologous to Os.22276 (69373391-69373394), γ aminobutyric acid (69373402-69373425), GYF domain containing protein (69374178-69374194) and RD22 (69373894-69373912). The SNPs associated with structural genes and secondary metabolic pathway include water channel protein (69373704-69373723, 69374218-69374239, 69374266-69374278), Aquaporin (69374240-69374265, 69374279-69374300), proton pump interacting protein (69374432-69374436), phenylalanine ammonia lyase (69373551-69373562), coumaryl coA ligase (69374440-693744494), chalcone isomerase (105111381-105111393). Some of the regions were though not directly associated with stress response but associated with other important agronomic traits such as yield locus (Ashikari *et al*., 2005) were identified such as SNPs in cytokinin oxidase (69373302-69373308).

5.4 Rate of Polymorphisms reveals the association of target regions with the trait

Rate of polymorphism (ROP) was observed to be high between drought tolerant and susceptible varieties in most of the cases, when compared with in the panel suggesting tight association of the targeted regions with the trait. Rate of polymorphism between CT9993 and IR62266 was observed to be 0.544 and between Azucena and IR64 was observed to be 0.466. Relatively low ROP was observed between either of the tolerant lines such as CT9993 and Azucena (0.367) or susceptible lines such IR62266 and IR64 (0.351). Similarly, low ROP was observed between IR20 and IR62266 (0.354) and also between IR20 and IR64 (0.362) with in the panel of susceptible genotypes. High ROP

was observed between Nerica-1 and any of the other line either within or across the panel other than Azucena indicating more closeness of Azucena with Nerica-1 as the ROP was observed to be very low. The panel of moderately tolerant lines has ROP neither low nor high when compared to either of the tolerant or susceptible panel of genotypes other than Nipponbare. Further, it has been observed that a common pattern was observed with in the panel of genotypes in most of the targeted regions. This suggests the tight association of the set of candidate SNPs with the trait. However, they further needs to be studied through association genetic approaches.

5.5 Physical mapping of SNPs onto IRGSP map helped in finding allelesharing pattern

Sequence information from IRGSP and extensive utilization BAC/PAC contig mapping data helped us in physically mapping identified SNPs directly on to IRGSP sequence map. A unique nomenclature was used for the first time with the help of an automated tool developed in the present study. In fact this can be utilized by the other research groups for automated screening and reporting of SNPs. Further, direct mapping of SNPs to the base level on the integrated map enabled us to generate a physical map of SNPs with CMAP software. Haplotype extraction from the alignment data helped in generating an allele sharing pattern between the 8 genotypes studied for SNP identification.

5.6 Allele sharing maps and QTL integration helped in connecting functional genomics approaches to molecular genetics approach for the trait dissection

The concept of allele sharing maps was proposed by Monna *et al*. (2006) by integrating allele sharing map with the genetic map. We have developed here a further improved version of allele sharing maps by integrating with rice sequence map, genetic map and

with that of QTL maps. This will help geneticists and rice breeders to precisely select the QTL region of interest and further look in to the candidate gene to select readily available marker displaying allele sharing and haplotype pattern. The allele sharing maps we have reported here can be utilized efficiently to convert gene specific markers spanning the target abiotic stress related QTL regions. The allele sharing maps helped us in identifying the pattern of variation and sharing between the genotypes. As shown in the allele sharing map of chromosome 1-12 (Fig. 4.17A-4.17L) many SNPs were found to have a common pattern with in the panel of genotypes. Some of the important observations with a strict sharing pattern with in the panel of susceptible or tolerant genotypes include SNPs associated with RIC1 GTP binding promoter dbSNP ss#69373343 wherein all the drought tolerant varieties have 'A' and susceptible varieties show a deletion. Similarly, C/T (69373351) shows tolerant lines show 'T' and susceptible lines show 'C' in the same region. The map at this position also reveals the region is spanning many drought responsive QTLs such as leaf drying (Gramene Acc. No. DQE1, AQD024, AQD027), relative water content (DQE44), relative growth rate (DQE33), penetrated root thickness (CQH2), sodium potassium ratio (CQ13) etc at 87.1 cM region of chromosome 1 of the genetic map. Similarly, two SNPs associated with promoter region of the cytochrome P450 gene viz. T/C (69373351) and T/G (69373373) show a common pattern with in the panel of genotypes spanning the QTLs like osmotic adjustment capacity (AQDX002), penetrated root length (AQGC035), penetrated root number (CQAW4), relative phosphorous distribution between shoot and root (CQAA13, CQAA14) at 103.1 cM of chromosome 1. Some important observations on chromosome 6 include C/- (69374141, 69374141) in the promoter region of thaumatin gene spanning QTLs like lodging

incidence (AQDZ014), relative phosphorous uptake (CQAA5), salt sensitivity (AQEM002), penetrated root thickness (AQFT039) at 113.4 cM of genetic map. Chromosome 10 shows A/T SNP (69374641) having similar pattern with an exception in Nipponbare, associated with promoter region of protein kinase similar to Os. 10505 spanning QTLs like penetrated root ratio (DQC17, AQC017), leaf drying (DQE9), penetrated root number (DQC10), relative root length (AQAA018) spanning 61.7 to 68.6 cM on genetic map. Two SNPs are identified on chromosome 11, viz. G/C (69374726), G/A (69374734) having an association with the phenotype based on the panel of genotypes analyzed at 45.6 to 49.1 cM region on genetic map in the region associated with the promoter of a transcription factor differentially expressed in Nerica library developed in the present study. These SNPs could be regulatory switches for differential expression of this transcription factor and downstream activation of several other genes involved in the process of conferring tolerance.

The allele sharing maps also reveal many SNPs having a common pattern in Nipponbare, CT9993, Azucena and Nerica1 but not in either or both the moderately tolerant lines such as Nootripathu and N22. These include A/- (69373345) in RIC1 promoter region associated with QTLs such as leaf drying (DQE1, AQD024, AQD027), relative water content (DQE44), relative growth rate (DQE33), penetrated root thickness (CQH2), sodium potassium ratio (CQ13), G/T (69373394) in the promoter of drought induced protein similar to Os. 22276 spanning QTLs like relative water content (DQE45, DQE46), root weight (CQQ32, CQQ6), root dry weight to tiller number ratio (CQQ13), leaf rolling (DQE12, DQE13), penetrated root thickness (DQF9, AQGC028), UV-B resistance (AQCS001) and $KClO_3$ resistance (AQF092) etc. at 136.1 cM on chromosome

1. Chromosome 3 also shows some SNPs with above pattern such as C/A (69373607), A/T (69373608) in promoter region of serine threonine kinase but not spanning any of the stress responsive QTLs, one indel (-/AA; 69373622) in a region associated with AP2 gene promoter spanning leaf rolling (DQE17) associated QTL. A complete haplotype pattern was also observed in MAP kinase region at 40.3 cM (69373636, 69373637, 69373639, 69373641, 69373646-69373652) also in APX promoter region at 44.4 cM region with a set of 32 SNPs (69373658 - 69373678) spanning QTLs such as stomatal closure time (AQDL005), osmotic adjustment capacity (CQH12), phosphorous sensitivity (AQCI008). Such haplotype also observed at regions associated with other signal transduction associated genes such as inositol phosphate transferase at 87.4 cM on genetic map spanning other drought associated QTLs such as penetrated root number (CQAW10), cell membrane stability (DQA2), potassium sensitivity (AQBL002, AQBL009, AQBL011, AQBL018), relative growth rate (DQE37) besides stomatal closure time (AQDL005), osmotic adjustment capacity (CQAV2). Interestingly, a similar pattern was also observed in an unknown membrane associated protein differentially expressed in Nerica cDNA library localized at 149.1 cM associated leaf rolling (AQD002, AQD010), iron sensitivity (AQEK001) associated QTLs. Chromosome 4 also shows similar haplotype pattern at genomic regions associated promoter regions of ring H2 finger associated protein (69373726-69373728, 69373735-69373744), jasmonic acid induced protein (69373747-69373771, 69373775, 69373777) at 19.6 to 24.1 cM region associated with drought responsive QTLs such as leaf drying (DQE3), drought susceptibility index (CQAI48), osmotic adjustment capacity (AQDX006) which are found to be highly expressed in the microarray gene expression profiling experiments

with different degrees of drought stress induction (Gorantla, 2005). Similarly, A/T variation (69373785) at 78.2 cM region associated with an unknown protein identified to be over expressed during the increased drought stress condition spanning QTLs associated with potassium concentration (CQ15), leaf drying (AQD018), phosphorous sensitivity (AQCI011), $KClO_3$ resistance (AQF097) etc. An insertion deletion (Indel) at 9.5 cM was identified with C/- variation (69373868) with such pattern but not associated with any of the stress responsive QTLs in the promoter regions of R2R3 MYB binding protein. A deletion (A/- ; 69373884) was observed in the SAM synthase promoter with this pattern spanning the QTLs associated with penetrated root number (DQC9) and lodging incidence (AQDZ012). A common haplotype pattern was observed at the promoter region of a tubby like transcription factor homologous Os.10160 with a set of 11 SNPs (69373943- 69373954) at 104.7 cM region spanning several drought stress associated QTLs such as root weight (CQQ2), penetrated to total root ratio (DQC16), relative water content (DQE50, AQD034, AQD038), leaf rolling (AQD008), penetrated root number (AQC009), osmotic adjustment capacity (AQDX008). A similar pattern was also observed at 107.4 cM region in the promoter regions of cold induced gene (69373965-69373967) associated with drought stress associated QTL regions such as relative water content (DQE50), osmotic adjustment (AQDX008), relative phosphorous distribution between shoot and root (CQAA19). An Indel was observed in the promoter region of late embryogenesis abundant protein (LEA) at 109 cM region (69373974) spanning in the QTLs such as relative water content (DQE50), osmotic adjustment capacity (AQDX008). Chromosome 6 observed to have some of the interesting haplotype patterns (69374108-69374110, 69374115, 69374116) though not shared by Nootripathu

in the region encoding rs6 kinase and its promoter region at 99.2 cM spanning QTLs associated with salt tolerance like salt sensitivity (AQEM002), sodium concentration (CQ16). This haplotype also found to be shared by promoter of KNOX gene (69374120-69374134), thaumatin (69374144, 69374145, 69374148- 69374153, 69374155, 69374156, 69374159 spanning the same QTLs at 100.8 to 113.4 cM. Chromosome 7 has set of 16 SNPs (60374178-69374194) having a common haplotype pattern spanning drought tolerance associated QTLs such as cell membrane stability (DQA3), penetrated root thickness (AQFT040) at 7-8.8 cM in the promoter region GYF domain containing gene, differentially expressed in the Nerica library. One indel (A/-; 69374242) was observed in Aquaporin promoter region having this pattern spanning many drought responsive QTLs such as relative growth rate (DQE40), relative shoot dry weight (CQAA11), leaf rolling (AQAO47), relative root length (AQAA015) at 53.4 cM on the genetic map. This pattern and haplotype was also observed in the promoter region of receptor like kinase (69374312-69374325) which is also differentially expressed in Nerica library found spanning the drought stress response QTLs such as relative shoot dry weight (CQAA11), leaf rolling (DQE26, CQAI50), Osmotic adjustment capacity (AQDX009) at 96.1 - 97.4 cM on the genetic map. This haplotype pattern (69374351-69374355, 69374361-69374364, 69374367-69374370, 69374373, 69374374) was also found extended at 99.3 cM region associated with the promoter region of low temperature induced protein (LT16) similar to Os.46402 spanning drought responsive QTLs such as relative shoot dry weight (CQAA11), leaf rolling (DQE27, CQAI50). Chromosome 8 shows two indels (69374413, 69374414) in the promoter region of wound induced protein at 58 cM region but not associated with any of the stress responsive

QTLs. A set of 3 SNPs (69374434-69374436) were mapped at 61.2 to 72.2 cM region associated with a proton pump interacting gene homologous to Os.10450 and one SNP (69374436) associated with betaine aldehyde dehydrogenase genic region spanning two drought responsive QTLs viz. Osmotic adjustment (CQAV6, CQAV7, CQAV9), potassium sensitivity (AQBL006, AQBL010, AQBL013, AQBL016, AQBL020, AQBL021). A set of 3 SNPs (69374497, 69374498, 69374506) were found spanning QTLs such as relative root length (CQL9, CQL8), osmotic adjustment capacity (AQDX013) in the region associated with NAM transcription factor. Many drought induced transcription factors were mapped on chromosome 9 spanning several QTLs in the region spanning 40.1 to 78.5 cM on the genetic map. A common haplotype was observed in DREB2 promoter region (69374517-69374519) at 40.1 cM and also in aldehyde dehydrogenase promoter (69374520-69374525) at 58.3 to 60.8 cM region spanning drought associated QTLs like root weight (CQQ39, CQQ8), relative growth rate (DQE42), rooting depth (AQR040). A similar haplotype pattern was also observed in the promoter region of HSP82 with 5 SNPs (69374533-69374536, 69374538, 69374540) at 68.2 cM region spanning QTLs like cell membrane stability (DQA7), penetrated root thickness (CQH33, DQF12), root dry weight to tiller number ratio (CQQ16). A set of 6 SNPs were found with a common haplotype pattern in the promoter region of CBF3 (69374552-69374554, 69374558, 69374561 and 69374562) which is a well known drought tolerance associated transcription factor spanning drought associated QTLs such as relative root length (AQAA017), potassium uptake (CQ19) at 78.5 cM on the genetic map. A 9 bp deletion was also observed in this region (69374552) observed in all the drought susceptible lines in the Myb1 binding site. A similar haplotype was also observed

in another CBF-3 gene tandemly duplicated at this region with 11 SNPs (69374563-69374570, 69374573, 69374575). Chromosome 10 found to have two SNPs with a common sharing pattern across the genotypes. A 'T/C' variation (69374631) observed in the promoter region of promoter of MADS transcription factor at 48.4 cM spanning two drought associated QTLs viz. penetrated root number (AQC010), penetrated to total root ratio (DQC17). Another SNP (69374641) was observed in the promoter region of a protein kinase gene homologous to Os.10505 at 61.7 to 68.6 cM region spanning four drought associated QTLs viz. penetrated to total root ratio (DQC17, AQC017), leaf drying (DQE9), penetrated root number (DQC10), relative root length (AQAA018). Chromosome 11 has common haplotype (69374678, 69374680-69374682, 69374685-69374689, 69374691- 69374700) in a region encoding a disease responsive gene spanning QTLs such as rooting depth (AQAL058), root penetration index (CQAW31) at 20.3 cM. Six SNPs (69374726-69374728, 69374730, 69374731, 69374732, 69374734) were observed in the region associated with the promoter of a differentially expressed transcription factor in Nerica-1 library spanning QTLs such as leaf drying (DQE10), root pulling force (CQH36) at 45.6 - 49.1 cM on the genetic map. Of this, a G/C variation (69374726) showed 'C' in all susceptible lines and 'G' in all tolerant lines. Another SNP with G/A variation (69374734) has 'G' in all tolerant lines and 'A' in all susceptible lines. Chromosome 12 found to have 3 SNPs (69374755, 69374757, 69374758) with a common pattern in the promoter of a zinc finger protein spanning three drought tolerance associated QTLs viz. drought tolerance (AQAN002), leaf drying (DQE11), relative water content (DQE55) at 91.4 cM on the genetic map. All such informative SNPs showing polymorphism between CT9993 and IR62266 were attempted to convert into PCR based

markers using ASPCR technology using LNA as ASPCR primer or using SBE technology. The allelesharing mapping approach connecting the abiotic stress responsive QTLs was effective in extracting haplotype patterns and mapping SNPs to the spanning QTLs.

5.7 Promoter scanning at the SNP sites revealed regulatory variations

The promoter regions of candidate genes dissected to identify variation have revealed many SNPs associated with many *cis* regulatory elements associated with stress response. This strategy coupled with allelesharing maps allowed us to establish a direct link from the QTL level to gene and regulatory region variations at the targeted loci. Analysis of CBF3 promoter spanning the QTL associated with relative root length (RRL) revealed a 'T/C' variation (dbSNP ss#69374564) in the 4th nucleotide of cis element with the core motif sequence of 'ACGT' associated with ABRE binding transcription factors. It is reported that ACGT element coupled with Myb1 element is essential for induction of genes associated with dehydration (Simpson *et al*., 2003). Though, CBF3 is reported for its role in ABA independent pathway (Oh *et al*., 2005), variations in such element may play a key role in determining the expression CBF3 gene in ABA dependent manner and which in turn activate many downstream genes involved in dehydration response process through both the pathways. Analysis of SNPs observed in the promoter region of NAM gene spanning the QTLs associated with RRL, Osmotic adjustment capacity (OA) revealed one SNP (G/T; 69374498) in ABRE *cis* element with *cis* regulatory motif of ACGTG where in SNP was observed in the 5th nucleotide of the motif. All tolerant lines found to have a motif 'ACGTG' where as susceptible lines have 'ACGTT' due to the SNP. Though, ACGT is essential for ABRE to bind, there could be altered binding

affinity with this change. Similar haplotype exists in this region but none of them seems to be associated with any *cis* regulatory element. In Arabidopsis, Nakashima *et al*, 2006, reported ABRELATERD1 element with *cis* regulatory sequence of 'ACGTG' is involved in binding bZIP transcription factor in response to ABA. Though, ACGT motif is reported to be the core motif required for binding ABF transcription factors in ABA dependent pathway, the flanking regions also likely to play a key role in determining the binding affinity. Such variations are likely to contribute in determining the level of gene expression. The analysis of low temperature and drought stress induced promoter of LT16a revealed variation in CRT/DRE element which plays a key role in activating the gene expression in ABA independent manner. The CRT/DRE element with the core sequence motif of 'RYCGAC' is known to be involved in the binding of DREB transcription factors (Yamaguchi Shinozaki and Shinozaki, 1994; Svensson *et al*., 2006). We observed a 'G/A' variation (69374354) in the 4th nucleotide of core motif of CRT/DRE element wherein G is an essential element required for binding of DREB transcription factors. This region was also found spanning the QTLs associated with root dry weight, leaf rolling. Another 'G/A' variation (69374353) in flanking the Myb binding region (YAACKG; Abe *et al*., 2003) was also found in this promoter region which is known to be involved in ABA dependent pathway (Oh *et al*., 2005). Many variations were observed in the Myb binding regions of some transcription factors and downstream genes as well. Promoter region of DREB1B (CBF1) which is known to be differentially regulated in during osmotic stress, oxidative stress, salicylic acid, ABA, and cold (Gutha and Reddy, 2008) revealed a variation (C/A) at the 2nd nucleotide of Myb2 core motif (YAACKG; Abe *et al*., 2003) and a two base pair deletion (69374553) in Myb1 binding

site (WAACCA; Abe *et al*., 2003) affecting 4th to 6th nucleotides in the core motif. Variations were also observed in the other *cis* regulatory elements of this promoter including WBOX (T/G; 69374561) and ethylene responsive element (ERE; 69374562). Interestingly, the integrated allelesharing map shows this region is spanning in the QTL associated with 'relative root length (RRL). Variations were also observed in the Myb2 binding region (YAACKG) of promoter other transcription factors such as Glycine rich zinc finger protein (T/C; 69373477) affecting 1st nucleotide of the core motif. Variations in the flanking region of Myb binding region were also observed in the GYF domain containing protein (G/A; 69374183, A/C; 69374184). SNPs (C/T; 69374185) in the MycRD22 binding region (CACATG; Abe *et al*., 1997) were also observed in this promoter region affecting the 5th nucleotide of the core motif. The allelesharing maps revealed that the region is spanning the QTLs associated with the penetrated root thickness (PRT) and cell membrane stability. Variations were also observed in the MybRD22 (CTAACCA; Abe *et al*., 1997) binding region of cystein protease promoter affecting the 4th nucleotide in the core motif spanning the QTLs associated with plant height (PH) and leaf width. The promoter analysis of Ring H2 finger gene revealed variation (C/T; 69373740) in the MYBCOREATCYCB1 (AACGG; Planchais *et al*., 2002) affecting 3rd nucleotide in the core motif. Variations were observed in the flanking region of Dof binding site (AAAG; Yanagisawa, 2000) in the promoter region of Aquaporin (T/G; 69374256, A/G; 69374257). Variation (T/A; 69374164) was also observed in Dof binding site of nifU promoter region affecting the 3rd nucleotide of core motif spanning the QTL associated with penetrated root thickness. A complete deletion of Dof element was observed in the promoter region of cystein protease promoter (-

/GAAAGAAAGAAA). The Dof binding site of drought inducible promoter region homologous to Os.57533 was also found have a G/A variation (69374515) affecting the 4th base of the core motif. Several other *cis* elements were identified to have SNPs in the promoter regions of various candidate genes analyzed in the study such as CORE, EEC LCR1, SITEII, ARR1, WRKY, WBOX, MART, TBOX, GATA and SORLPIT etc. which may alter the expression of the target genes in different genotypes. SNPs observed in the promoter region of thioredoxin, an antioxidant gene involved in oxidative stress shows 'A/G' variation (69373648) in CORE element involved MAPKK mediated gene regulation (Tsukamoto *et al*., 2005). Their report based on different deletion constructs suggests this element is necessary for expression antioxidant genes such as thioredoxin, SOD, glutaredoxin etc. Promoter of the cold inducible gene revealed a 'G/A' variation (69373966) in the ARR1 element, affecting second nucleotide of core sequence (NGATT; Ross *et al*., 2004). It is observed that essential 'GATT' element is conserved in drought tolerant lines and susceptible lines have 'GGTT' due the SNP. This motif is reported to be essential for cytokinin induced gene expression mediated through Myb (Ross *et al*., 2004).

5.8 SNP-ASPCR technology facilitated development of rapid, simple yet cost effective markers for SNP genotyping.

The simple LNA-ASPCR-SNP markers we report here spanning the QTLs associated with candidate genes and their *cis* acting elements having interesting allele sharing pattern are expected to serve as precise markers in breeding for drought tolerance in rice and other crops as well. The markers developed with ASLNA-PCR technique (Latorra *et al*., 2003) is a step towards bridging the gap between functional genomics and molecular

genetics. The preferential choice for selecting allele specific PCR was to facilitate utilization of SNP markers in Marker Assisted Selection (MAS) with minimal biotechnological facilities. These markers were used in genotyping of the DHL population but only a very few markers were effectively mapped onto the genetic map as per the physical map. Most of the markers were found to be mapped at more than one chromosome or remained lonesome. Further, these need to be analyzed in a set of larger population in order to map them more precisely. Though we attempted to convert all targeted SNP sites into AS-PCR markers, inherent problems with the assay such as low complexity at the SNP site and restriction at 3' end for primer design have become bottleneck for the success. Thus, we have adapted single base extension strategies with where we could validate most of the candidate SNPs. Further, these SBE markers need to be converted into simplified markers in order to use them in MAS program cost effectively.

6. SUMMARY

6. Summary

The main objective of the present study was to identify the genes involved in drought stress response process, and to dissect these regions to find single nucleotide polymorphisms (SNPs) having correlation with drought tolerance, and converting them in to molecular markers to facilitate marker assisted selection (MAS). As the trait is multigenic showing a complex regulation, our main objective is to identify the genomic regions involved in drought stress response from ~430 MB of the rice genome. Various strategies have been employed in the present study to identify the target genomic regions which include construction of subtractive cDNA libraries, analysis of microarray gene expression profiles at various water stress regimes, and comparative analysis of ESTs generated from drought stressed normalized cDNA libraries with reported microarray data besides extensive literature survey. Two rice genotypes having a contrasting phenotype with reference to drought tolerance were selected to identify differentially expressed genes. A highly drought tolerant genotype, Nerica-1 and a drought susceptible genotype (IR64) were used for subtractive cDNA library construction. ESTs generated from this library were deposited in the dbEST division of NCBI. Forty two unigenes involved in various stress response processes such as signal perception, signal transduction, transcription activation and downstream genes involved in stress response were identified besides some novel genes. The variation observed in respect to gene expression between these genotypes is attributed to the variations in the promoter regions, particularly in the regulatory elements which drive the gene expression. In addition, gene expression profiles of ESTs characterized from normalized cDNA library constructed from drought stressed seedlings of Nagina22 genotype were considered. Some of the interesting genes from the gene clusters showing dramatic changes at various time

course experiments in three different drought stress regimes at various developmental stages were considered as the drought stress response genes. In addition, some members of interesting gene families identified from the comparative analysis of Nagina22 ESTs with the microarray gene expression profiles of orthologous genes were also considered for the study. The gene regulation observed to be complex under defined stresses condition, particularly, negative regulation of many genes and subsequent triggering/suppressing of other genes is one of the critical areas to understand the mechanism behind such regulation. This kind of regulation is mainly attributed to the involvement of small non-protein coding mRNAs called miRNAs. In order to characterize the regulatory switches triggering biogenesis of such miRNAs, the regions involved in encoding miRNAs and their biogenesis were considered. Besides, some of the reported drought responsive genes were also considered with a major focus on identifying SNPs in the regulatory regions of the targeted regions.

A set of 300 genomic regions (Promters, CDS, Intron, UTRs) were targeted to dissect the allelic variation to the single nucleotide level. Eight rice genotypes with well characterized phenotype with reference to drought tolerance have been used for identification of SNPs that can be broadly categorized into two panels, drought tolerant and drought susceptible. The members falling in to the drought tolerant panel includes CT9993, Azucena, Nootripathu, N22B, Nerica-1 and, IR62266, IR64, IR20 represent the drought susceptible panel. These includes parental lines two populations viz. CT9993 X IR62266 and Azucena X IR64, which are widely used for mapping drought related QTLs. A total of 150 regions were amplified and sequence characterized from all the eight genotypes (Genbank Acc. EF556551- EF558346; EU868928-EU869180), and analyzed for SNPs using standard sequence processing programs coupled with in-house developed perl scripts. Screening of more than one

million bases at the targeted sites enabled us to capture the informative sites. All the informative sites are mapped on to rice genome sequence map. SNP positions were identified and denoted following a unique nomenclature referring chromosome number followed by base position on the rice genome for the first time. Though, the analysis revealed 2700 SNPs in the targeted regions, the application Bayesian probabilistic algorithm with cut-off probability value of 0.99 limited us to identify 1699 SNPs (dbSNP, NCBI, Handle ARR-VBREDDY, NCBI_ss#69373227 to 69374767; 105107054 to 105107119; 105111345 to 105111433). A total of 601 transitions, 461 transversions and 548 Indels were identified of which 318 are identified coding regions, 8 are in intronic regions and 69 are in UTRs. A majority of the SNPs i.e 1203 are found in promoter regions as our major focus was on 5' upstream region of the gene, and this includes SNPs found in the regulatory regions of the promoters and surrogate SNPs as well. Rate of polymorphism (ROP) between genotypes ranged from 0.313 to 0.928. Interestingly, ROP was observed to be high between drought tolerant and susceptible varieties, when compared with in the panel suggesting tight association of the targeted regions with the trait. Further, common pattern was observed with in the panel of genotypes in most of the targeted regions. Based on the successful results, we developed a genome wide haplotype and allele sharing map with the help of an in-house developed software tool called AS-map V1.0. This newly developed program enabled us to construct allele sharing maps of candidate genes and integrate with rice sequence map as well as genetic map. Further, in order to dissect and associate these SNPs, we have integrated abiotic stress responsive QTLs to our allele sharing map. This enabled us to identify haplotype tagged SNPs in the genomic regions associated with drought stress response, spanning abiotic stress responsive QTLs. SNPs in the QTL regions associated with relative

water content (RWC), osmotic adjustment capacity (OA), leaf rolling, leaf drying, stomatal closure time, drought susceptibility index and dehydration tolerance beside large number of root related QTLs were identified. All such SNPs found in promoter regions of the targeted genes were analyzed for their association with cis acting regulatory regions. A total of 40 SNPs are found to be associated with various cis elements, interestingly, twenty three are found to be associated with drought stress response related cis acting elements such as Myb, Myc, CRT/DRE, WRKY, ABRE, CORE etc. These findings were utilized for developing SNP markers for the drought tolerance trait. The complexity in typing SNPs was minimized by attempting to convert these SNPs into PCR based markers. A set of another 17 genotypes including some of the wild relatives of rice (*O.nivara and O.rufipogan*) and a double haploid population (DHL) of rice with 154 lines were extensively utilized in conversion of these markers and segregation analysis. A total of 82 SNPs were considered based on various parameters such as haplotype pattern, allele sharing pattern, association with cis elements and spanning the QTL regions for converting in to SNP-PCR markers. Of these, we could convert 38 SNP markers into PCR based marker using AS-PCR technique with LNA modification to enhance the specificity. A total of 25 parental lines and IR68586 DHL population with 154 lines were genotyped using these SNP-PCR markers. The SNPs, which could not be converted in to PCR based markers due to low complexity at the targeted site are converted in to single base extension (SBE) markers using SNuPE technique. In conclusion, we report 38 SNP-ASPCR, 26 SBE markers identified using a multidisciplinary approach dissecting the genomic regions associated with drought stress response. To our knowledge, this is the first report of its kind in rice, bridging the gap between the functional genomics and the molecular genetics for the trait under study.

7. REFERENCES

7. References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell **15:** 63–78

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research **25:** 3389-3402

An P, Wang LH, Hutcheson-Dilks H, Nelson G, Donfield S, Goedert JJ, Rinaldo CR, Buchbinder S, Kirk GD, O'Brien SJ, Winkler CA (2007) Regulatory Polymorphisms in the Cyclophilin A Gene, PPIA, Accelerate Progression to AIDS. PLoS Pathog. **3:** e88

Babu PR, Chandra Sekhar A, Ithal N, Markandeya G, Reddy A (2002) Annotation and BAC/PAC localization of nonredundant ESTs from drought-stressed seedlings of an indica rice. Journal of Genetics **81:** 25-44

Babu RC, Nguyen BD, Chamarerk V, Shanmugasundaram P, Chezhian P, Jeyaprakash P, Ganesh SK, Palchamy A, Sadasivam S, Sarkarung S, Wade LJ, Nguyen HT (2003) Genetic Analysis of Drought Resistance in Rice by Molecular Markers: Association between Secondary Traits and Field Performance. Crop Science **43:** 1457-1469

Bakker EG, Toomajian C, Kreitman M, Bergelson J (2006) A genome-wide survey of R gene polymorphisms in Arabidopsis. Plant Cell **18:** 1803-1818

Bao JS, Corke H, Sun M (2006) Microsatellites,single nucleotide polymorphisms and a sequence tagged site in starch-synthesizing genes in relation to starch physicochemical properties in nonwaxy rice (*Oryza sativa* L.). Theor Appl Genet **113:** 1185-1196

Bao JS, Corke H, Sun M (2006) Nucleotide diversity in starch synthase IIa and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physicochemical properties in rice (*Oryza sativa* L.). Theor Appl Genet **113:** 1171-1183

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D (2003) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiology **132:** 84-91

Belbin O, Dunn JL, Ling Y, Morgan L, Chappell S, Beaumont H, Warden D, Smith DA, Kalsheker N, Morgan K (2007) Regulatory region single nucleotide polymorphisms of the apolipoprotein E gene and the rate of cognitive decline in Alzheimer's disease. Hum Mol Genet **16:** 2199-2208

Bertin I, Zhu JH, Gale MD (2005) SSCP-SNP in pearl millet;a new marker system for comparative genetics. Theor Appl Genet **110:** 1467-1472

Borrone JW, Kuhn DN, Schnell RJ (2004) Isolation,characterization,and development of WRKY genes as useful genetic markers in *Theobroma cacao*. Theor Appl Genet **109:** 495-507

Boyle B, Brisson N (2001) Repression of the defense gene PR-10a by the single-stranded DNA binding protein SEBF. Plant Cell **13:** 2525–2537

Breyne P, Dreesen R, Cannoot B, Rombaut D, Vandepoele K, Rombauts S, Vanderhaeghen R, Inzé D, Zabeau M (2003) Quantitative cDNA-AFLP analysis for genome-wide expression studies. Mol Genet Genomics **269:** 173-179

Brock MT, Tiffin P, Weinig C (2007) Sequence diversity and haplotype associations with phenotypic responses to crowding: GIGANTEA affects fruit set in *Arabidopsis thaliana*. Mol. Ecology **16:** 3050-3062

Buchel AS, Brederode FT, Bol JF, Linthorst HJ (1999) Mutation of GT-1 binding sites in the Pr-1A promoter influences the level of inducible gene expression in vivo. Plant Molecular Biology **40:** 387–396

Bundock PC, Christopher JT, Eggler P, Ablett G, Henry RJ, Holton TA (2003) Single

nucleotide polymorphisms in cytochrome P450 genes from barley. Theor Appl Genet **106:** 676-682

Bundock PC, Henry RJ (2004) Single nucleotide polymorphism,haplotype diversity and recombination in the Isa gene of barley. Theor Appl Genet **109:** 543-551

Chan CS, Guo L, Shih MC (2001) Promoter analysis of the nuclear gene encoding the chloroplast glyceraldehyde-3-phosphate dehydrogenase B subunit of *Arabidopsis thaliana*. Plant Molecular Biology **46:** 131–141

Chandrasekhar A (2005) Molecular mapping using SSRs and EST-PCR based markers, and Identification of quantitative trait loci (QTLs) for drought tolerance in a doubled-haploid population of rice (*Oryza sativa* L.). Ph.D thesis, School of Life Sciences, University of Hyderabad

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency,haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics **3:** 19

Cho K-S, Yang T-J, Hong S-Y, Kwon Y-S, Woo J-G, Park H-G (2006) Determination of cytoplasmic male sterile factors in onion plants (*Allium cepa* L.) using PCR-RFLP and SNP markers. Mol Cells **21:** 411-417

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang WH, Hubbell E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat Genet **23:** 203-207

Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon M-S, Hwang E-Y, Yi S-I, Young ND, Shoemaker RC, van Tassell CP, Specht JE, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics **176:** 685-696

Cogan NOI, Ponting RC, Vecchies AC, Drayton MC, George J, Dracatos PM, Dobrowolski MP, Sawbridge TI, Smith KF, Spangenberg GC, Forster JW (2006) Gene-associated single nucleotide polymorphism discovery in perennial ryegrass (*Lolium perenne* L.). Mol Genet Genomics **276:** 101-112

Cordeiro G, Eliott F, McIntyre C, Casu R, Henry R (2006) Characterisation of single nucleotide polymorphisms in sugarcane ESTs. Theor Appl Genet **113:** 331-343

Courtois B, McLaren G, Sinha PK, Prasad K, Yadav R, Shen L (2000) Mapping QTLs associated with drought avoidance in upland rice. Molecular Breeding **6:** 55-66

de Meaux J, Pop A, Mitchell-Olds T (2006) Cis-regulatory evolution of chalcone-synthase expression in the genus Arabidopsis. Genetics **174:** 2181-2202

Douglas JA, Levin AM, Zuhlke KA, Ray AM, Johnson GR, Lange EM, Wood DP, Cooney KA (2007) Common variation in the BRCA1 gene and prostate cancer risk. Cancer Epidemiol Biomarkers Prev **16:** 1510-1516

Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. Plant Physiology **124:** 1483-1492

Duan S-H, Li S-Q, Li Y-S, Xiong Y, Zhu Y-G (2007) [Distribution and SNPs of the rice CMS-related gene in AA-genome of *Oryza species*]. Yi Chuan **29:** 455-461

Ewing B, Green P (1998) Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. Genome Research **8:** 186--194

Fang J, Twito T, Zhang Z, Chao CT (2006) Genetic relationships among fruiting-mei (*Prunus mume Sieb*. et Zucc.) cultivars evaluated with AFLP and SNP markers. Genome **49:** 1256-1264

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for

rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Research **14:** 1812-1819

Fukuta S, Mizukami Y, Ishida A, Kanbe M (2006) Development of loop-mediated isothermal amplification (LAMP)-based SNP markers for shelf-life in melon (*Cucumis melo* L.). J Appl Genet **47:** 303-308

Garcés-Claver A, Fellman S, Gil-Ortega R, Jahn M, Arnedo-AndrÃs Ma (2007) Identification, validation and survey of a single nucleotide polymorphism (SNP) associated with pungency in Capsicum. Theor Appl Genet **115:** 907-916

Gasser SM, Amati BB, Cardenas ME, Hofmann JF (1989) Studies on scaffold attachment sites and their relation to genome function. Int Rev Cytol **119:** 57–96

Gaudet M, Fara A-G, Sabatti M, Kuzminsky E, Mugnozza G (2007) Single-reaction for SNP Genotyping on Agarose Gel by Allele-specific PCR in Black Poplar (*Populus nigra*). Plant Molecular Biology Reporter **25:** 1-9

Gezen-Ak D, Dursun E, Ertan T, Hanaasi Hm, Gürvit H, Emre M, Eker E, Oztürk M, Engin F, Yilmazer S (2007) Association between vitamin D receptor gene polymorphism and Alzheimer's disease. Tohoku J Exp Med **212:** 275-282

Gorantla M (2005) Functional genomics of drought tolerance: large scale EST generation, annotation, physical mapping and gene expression profiling in rice (*Oryza sativa* sub sp., *indica* cv. Nagina 22). Ph.D thesis, School of Life Sciences, University of Hyderabad

Gorantla M, Babu PR, Lachagari VBR, Reddy AMM, Wusirika R, Bennetzen JL, Reddy AR (2007) Identification of stress-responsive genes in an indica rice (*Oryza sativa* L.) using ESTs generated from drought-stressed seedlings. J Exp Bot **58:** 253-265

Grivet, Grivet L, Glaszmann, Glaszmann JC, Vincentz, Vincentz M, da S, da Silva F, Arruda, Arruda P (2003) ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. Theor Appl Genet **106:** 190-197

Gupta V, Mukhopadhyay A, Arumugam N, Sodhi YS, Pental D, Pradhan AK (2004) Molecular tagging of erucic acid trait in oilseed mustard (*Brassica juncea*) by QTL mapping and single nucleotide polymorphisms in FAE1 gene. Theor Appl Genet **108:** 743-749

Gutha LR and Reddy AR (2008) Rice DREB1B promoter shows distinct stress-specific responses, and the overexpression of cDNA in tobacco confers improved abiotic and biotic stress tolerance. Plant Mol Biol. doi 18754079

Hagen G, Guilfoyle T (2002) Auxin-responsive gene expression: genes, promoters and regulatory factors. Plant Molecular Biology **49:** 373–385

Hao G-P, Wu Z-Y, Cao M-Q, Pelletier G, Brunel D, Huang C-L, Yang Q (2004) [Nucleotide polymorphism in the drought induced transcription factor CBF4 region of *Arabidopsis thaliana* and its molecular evolution analyses]. Yi Chuan **31:** 1415-1425

Hao G-P, Zhong-Yi W, Chen M-S, Ming-Qing C, Brunel D, Pelletier G, Cong-Lin H, Qing Y (2004) Single nucleotide polymorphisms of CBF4 locus region of *Arabidopsis thaliana* correspond to drought tolerance. Chinese Journal of Agricultural Biotechnology **1:** 181-190

Hemamalini GSSHE, Hittalmani S (2000) Molecular marker assisted tagging of morphological and physiological traits under two constrasting moisture regimes at peak vegetative stage in rice (*Oryza sativa* L.). Euphytica **112:** 69-78

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Research **27:** 297-300

Hobo T, Asada M, Kowyama Y, Hattori T (1999) ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. Plant J **19:** 679–689

Hoppe C, Klitz W, D'Harlingue K, Cheng S, Grow M, Steiner L, Noble J, Adams R, Styles L, in Sickle Cell Anemia Investigators SPT (2007) Confirmation of an association

between the TNF(-308) promoter polymorphism and stroke risk in children with sickle cell anemia. Stroke **38:** 2241-2246

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Research **9:** 868-877

Huang X-Q, Röder MS (2005) Development of SNP assays for genotyping the puroindoline b gene for grain hardness in wheat using pyrosequencing. J Agric Food Chem **53:** 2070-2075

Jiao Y, Jia P, Wang X, Su N, Yu S, Zhang D, Ma L, Feng Q, Jin Z, Li L, Xue Y, Cheng Z, Zhao H, Han B, Deng XW (2005) A tiling microarray expression analysis of rice chromosome 4 suggests a chromosome-level regulation of transcription. Plant Cell **17:** 1641-1657

Jiao Y, Ma L, Strickland E, Deng XW (2005) Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and Arabidopsis. Plant Cell **17:** 3239–3256

Just BJ, Santos CAF, Fonseca MEN, Boiteux LS, Oloizia BB, Simon PW (2007) Carotenoid biosynthesis structural genes in carrot (*Daucus carota*): isolation,sequence-characterization,single nucleotide polymorphism (SNP) markers and genome mapping. Theor Appl Genet **114:** 693-704

Kalendar PTR, Laurila J, Schulman AH, Manninen O, Kiviharju E (2006) Generation of SNP markers for short straw in oat (*Avena sativa* L.). Genome **49:** 282-287

Kamoshita, Kamoshita A, Wade, Wade L, Ali, Ali M, Pathan, Pathan M, Zhang, Zhang J, Sarkarung, Sarkarung S, Nguyen, Nguyen H (2002) Mapping QTLs for root morphology of a rice population adapted to rainfed lowland conditions. Theor Appl Genet **104:** 880--893

Kanazin V, Talbert H, See D, DeCamp P, Nevo E, Blake T (2002) Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Molecular Biology **48:** 529-537

Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ (2001) Gene expression profiles during the initial phase of salt stress in rice. Plant Cell **13:** 889-905

Kim H-J, Kim Y-K, Park J-Y, Kim J (2002) Light signalling mediated by phytochrome plays an important role in cold-induced gene expression through the C-repeat/dehydration responsive element (C/DRE) in *Arabidopsis thaliana*. Plant J **29:** 693–704

Kim MY, Van K, Lestari P, Moon JK, Lee SH (2005) SNP identification and SNAP marker development for a GmNARK gene controlling supernodulation in soybean. Theor Appl Genet **110:** 1003-1010

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. Science **312:** 1392-1396

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen **12:** 172-175

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). Mol Genet Genomics **270:** 24-33

Kottapalli KR, Kottapalli P, Agrawal GK, Kikuchi S, Rakwal R (2007) Recessive bacterial leaf blight resistance in rice: complexity,challenges and strategy. Biochem Biophys Res Commun **355:** 295-301

Kropat J, Tottey S, Birkenbihl RP, Depège N, Huijser P, Merchant S (2005) A regulator of nutritional copper signaling in Chlamydomonas is an SBP domain protein that recognizes the GTAC core of copper response element. Proc Natl Acad Sci. **102:** 18730–18735

Lafitte Hr AHPBC (2004) Yield response to water deficit in an upland rice mapping population: associations among traits and genetic markers. Theor Appl Genet **109:** 1237-1246

Lagudah ES, McFadden H, Singh RP, Huerta-Espino J, Bariana HS, Spielmeyer W (2006) Molecular genetic characterization of the Lr34/Yr18 slow rusting resistance gene region in wheat. Theor Appl Genet **114:** 21-30

Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet **111:** 1532-1544

Lamason RL, Mohideen M-APK, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G, Makalowska I, McKeigue PM, O'Donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ, Shriver MD, Canfield VA, Cheng KC (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science **310:** 1782-1786

Lanceras JC, Pantuwan G, Jongdee B, Toojinda T (2004) Quantitative Trait Loci Associated with Drought Tolerance at Reproductive Stage in Rice. Plant Physiology **135:** 384-399

Lander E.S GPAJBADMJLSE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary linkage maps of experimental and natural populations. Genomics **1:** 174-181

Latorra D, Campbell K, Wolter A, Hurley JM (2003) Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers. Hum Mutat **22:** 79-85

Lincoln SDM, Lander E (1992) Mapping Genes Controlling Quantitative Traits with MAPMAKER/QTL 1.1. *In*. Whitehead Institute Technical Report.,2nd edition.

Liu CG, Zhang GQ (2006) Single nucleotide polymorphism (SNP) and its application in rice. Yi Chuan **28(6):** 737-744

Marth GT (2003) Computational SNP discovery in DNA sequence data. Methods Mol Biol **212:** 85-110

Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitziel NO, Hillier L, Kwok PY, Gish WR (1999) A general approach to single-nucleotide polymorphism discovery. Nat Genet **23:** 452-456

Matukumalli LK, Grefenstette JJ, Hyten DL, Choi I-Y, Cregan PB, Tassell CPV (2006) Application of machine learning in SNP discovery. BMC Bioinformatics **7:** 4

McNally KL, Bruskiewich R, Mackill D, Buell CR, Leach JE, Leung H (2006) Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. Plant Physiology **141:** 26-31

Monna L, Ohta R, Masuda H, Koike A, Minobe Y (2006) Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars (*Oryza sativa* L.) and a wild accession (*Oryza rufipogon* Griff.). DNA Research **13:** 43-51

Moreno-Vázquez S, Ochoa OE, Faber N, Chao S, Jacobs JME, Maisonneuve B, Kesseli RV, Michelmore RW (2003) SNP-based codominant markers for a recessive gene conferring resistance to corky root rot (*Rhizomonas suberifaciens*) in lettuce (*Lactuca sativa*). Genome **46:** 1059-1069

Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proc Natl Acad Sci. **104:** 3289-3294

Nakashima K, Fujita Y, Katsura K, Maruyama K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2006) Transcriptional regulation of ABI3-and ABA-responsive genes including RD29B and RD29A in seeds,germinating embryos,and seedlings of Arabidopsis. Plant Molecular Biology **60:** 51–68

Nasmyth K, Adolf G, Lydall D, Seddon A (1990) The identification of a second cell cycle control on the HO promoter in yeast: cell cycle regulation of SW15 nuclear entry. Cell **62:** 631–647

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002)

Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Research **9:** 163-171

Nguyen, Nguyen V, Nguyen B, Sarkarung, Sarkarung S, Martinez, Martinez C, Paterson, Paterson A, Nguyen H (2002) Mapping of genes controlling aluminum tolerance in rice: comparison of different genetic backgrounds. Molecular Genetics and Genomics **267:** 772-780

Norata GD, Ongari M, Garlaschelli K, Tibolla G, Grigore L, Raselli S, Vettoretti S, Baragetti I, Noto D, Cefalù AB, Buccianti G, Averna M, Catapano AL (2007) Effect of the -420C/G variant of the resistin gene promoter on metabolic syndrome, obesity, myocardial infarction and kidney dysfunction. Journal of Internal Medicine **262:** 104-112

Oh Se-Jun SISYSKH-JJSYKMKY-KKBHN, Kim J-K (2005) Arabidopsis CBF3/DREB1A and ABF3 in Transgenic Rice Increased Tolerance to Abiotic Stress without Stunting Growth. Plant Physiology **138:** 341–351

Olsen KM (2004) SNPs,SSRs and inferences on cassava's origin. Plant Molecular Biology **56:** 517-526

Pavy N, Parsons LS, Paule C, MacKay J, Bousquet J (2006) Automated SNP detection from a large collection of white spruce expressed sequences: contributing factors and approaches for the categorization of SNPs. BMC Genomics **7:** 174

Planchais S, Perennes C, Glab N, Mironov V, Bergounioux DIC (2002) Characterization of cis-acting element involved in cell cycle phase-independent activation of Arath;CycB1;1 transcription and identification of putative regulatory proteins. Plant Molecular Biology **50:** 111–127

Potokina E, Prasad M, Malysheva L, Röder MS, Graner A (2006) Expression genetics and haplotype analysis reveal cis regulation of serine carboxypeptidase I (Cxp1),a candidate gene for malting quality in barley (*Hordeum vulgare* L.). Funct Integr Genomics **6:** 25-35

Price A.H KASBJMRGWJ (2002) Upland rice grown in soil-filled chambers and exposed to contrasting water-deficit regimes. II. Mapping quantitative trait loci for root morphology and distribution. Field Crops Research **76:** 25-43

Price A.H SKAMBJBPBCLJ (2000) A combined RFLP and AFLP linkage map of upland rice (*Oryza sativa* L.) used to identify QTLs for root-penetration ability. Theor Appl Genet **100:** 49-56

Price AH, Tomos AD (1997) Genetic dissection of root growth in rice (*Oryza sativa* L.). II: mapping quantitative trait loci using molecular markers. Theor Appl Genet **95:** 143-152

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiology **133:** 1755-1767

Ravel C, Nagy IJ, Martre P, Sourdille P, Dardevet M, Balfourier F, Pont C, Giancola S, Praud S, Charmet G (2006) Single nucleotide polymorphism,genetic mapping,and expression of genes coding for the DOF wheat prolamin-box binding factor. Funct Integr Genomics **6:** 310-321

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhoub B, Brunel D, Beckert M, Charmet G (2006) Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). Genome **49:** 1131-1139

Ravel C, Praud S, Murigneux A, Linossier L, Dardevet M, Balfourier F, Dufour P, Brunel D, Charmet G (2006) Identification of Glu-B1-1 as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study. Theor Appl Genet **112:** 738-743

Rawat R, Xu Z-F, Yao K-M, Chye M-L (2005) Identification of cis-elements for ethylene and circadian regulation of the *Solanum melongena* gene encoding cysteine proteinase. Plant

Molecular Biology **57:** 629–643

Reale S, Doveri S, Díaz A, Angiolillo A, Lucentini L, Pilla F, Martín A, Donini P, Lee D (2006) SNP-based markers for discriminating olive (*Olea europaea* L.) cultivars. Genome **49:** 1193-1205

Reddy AR, Ramakrishna W, Sekhar AC, Ithal N, Babu PR, Bonaldo MF, Soares MB, Bennetzen JL (2002) Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice (*Oryza sativa* L. subsp. indica cv. Nagina 22). Genome **45:** 204–211

Ross EJH, Stone JM, Elowsky CG, Arredondo-Peter R, Klucas RV, Sarath G (2004) Activation of the *Oryza sativa* non-symbiotic haemoglobin-2 promoter by the cytokinin-regulated transcription factor,ARR1. J Exp Bot **55:** 1721–1731

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R (2005) Single-feature polymorphism discovery in the barley transcriptome. Genome Biology **6:** R54

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Genet Genomics **274:** 515-527

Rubio-Somoza I, Martinez M, Abraham Z, Diaz I, Carbonero P (2006) Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. Plant J **47:** 269–281

Sasaki Y, Fushimi H, Komatsu K (2004) Application of single-nucleotide polymorphism analysis of the trnK gene to the identification of Curcuma plants. Biol Pharm Bull **27:** 144-146

Scallon BJ, Dickinson CD, Nielsen NC (1987) Characterization of a null-allele for the Gy4 glycinin gene from soybean. Molecular and General Genetics MGG **208:** 107-113

Schmid KJ, OttóTörjék, Meyer R, Schmuths H, Hoffmann MH, Altmann T (2006) Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. Theor Appl Genet **112:** 1104-1114

Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. Genome Research **13:** 1250-1257

Schnurbusch T, Collins N, Eastwood R, Sutton T, Jefferies S, Langridge P (2007) Fine mapping and targeted SNP survey using rice-wheat gene colinearity in the region of the Bo1 boron toxicity tolerance locus of bread wheat. Theor Appl Genet

Shin J-H, Kwon S-J, Lee JK, Min H-K, Kim N-S (2006) Genetic diversity of maize kernel starch-synthesis genes with SNAPs. Genome **49:** 1287-1296

Shultz JL, Ray JD, Lightfoot DA (2007) A sequence based synteny map between soybean and *Arabidopsis thaliana*. BMC Genomics **8:** 8

Simpson SD, Nakashima K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Two different novel cis-acting elements of erd1,a clpA homologous Arabidopsis gene function in induction by dehydration stress and dark-induced senescence. Plant J **33:** 259–270

Solano R, Nieto C, Avila J, Cañas L, Diaz I, Paz-Ares J (1995) Dual DNA binding specificity of a petal epidermis-specific MYB transcription factor (MYB.Ph3) from Petunia hybrida. EMBO J **14:** 1773–1784

Somers DJ, Kirkpatrick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome **46:** 431-437

Svensson JT, Crosatti C, Campoli C, Bassi R, Stanca AM, Close TJ, Cattivelli L (2006) Transcriptome analysis of cold acclimation in barley albina and xantha mutants. Plant Physiology **141:** 257–270

Tacconi G, Baldassarre V, Collins NC, Bulgarelli D, Stanca AM, Valè G (2006) Haplotype characterization and markers at the barley Mlo powdery mildew resistance locus as tools for marker-assisted selection. Genome **49:** 864-872

Tang W, Perry SE (2003) Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. J Biol Chem **278:** 28154–28159

Techaprasan J, Ngamriabsakul C, Klinbunga S, Chusacultanachai S, Jenjittikul T (2006) Genetic variation and species identification of *Thai Boesenbergia* (Zingiberaceae) analyzed by chloroplast DNA polymorphism. J Biochem Mol Biol **39:** 361-370

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays* L.). Proc Natl Acad Sci. **98:** 9161-9166

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet **28:** 286-289

Törjék O, Berger D, Meyer RC, Müssig C, Schmid KJ, Sörensen TR, Weisshaar B, Mitchell-Olds T, Altmann T (2003) Establishment of a high-efficiency SNP-based framework marker set for Arabidopsis. Plant J **36:** 122-140

Tripathy JNZJRSNTT, Nguyen HT (2000) QTLs for cell-membrane stability mapped in rice (*Oryza sativa* L.) under drought stress. Theor Appl Genet **100:** 1197-1202

Troggio M, Malacarne G, Coppola G, Segala C, Cartwright DA, Pindo M, Stefanini M, Mank R, Moroldo M, Morgante M, Grando MS, Velasco R (2007) A dense SNP-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring Pinot Noir BAC contigs. Genetics

Tsukamoto S, Morita S, Hirano E, Yokoi H, Masumura T, Tanaka K (2005) A novel cis-element that is responsive to oxidative stress regulates three antioxidant defense genes in rice. Plant Physiology **137:** 317–327

Varshney RK, Beier U, Khlestkina EK, Kota R, Korzun V, Graner A, Börner A (2007) Single nucleotide polymorphisms in rye (*Secale cereale* L.): discovery,frequency,and applications for genome mapping and diversity studies. Theor Appl Genet **114:** 1105-1116

Venuprasad RSHEHS, Hemamalini GS (2002) Tagging quantitative trait loci associated with grain yield and root morphological traits in rice (*Oryza sativa* L.) under contrasting moisture regimes. Euphytica **128:** 293-300

Walia H, Wilson C, Condamine P, Liu X, Ismail AM, Zeng L, Wanamaker SI, Mandal J, Xu J, Cui X, Close TJ (2005) Comparative transcriptional profiling of two contrasting rice genotypes under salinity stress during the vegetative growth stage. Plant Physiology **139:** 822-835

Wang J-R, Wei Y-M, Yan Z-H, Zheng Y-L (2005) Detection of single nucleotide polymorphisms in 24 kDa dimeric alpha-amylase inhibitors from cultivated wheat and its diploid putative progenitors. Biochim Biophys Acta **1723:** 309-320

Waters DLE, Henry RJ, Reinke RF, Fitzgerald MA (2006) Gelatinization temperature of rice explained by polymorphisms in starch synthase. Plant Biotechnol J **4:** 115-122

Welchen E, Gonzalez DH (2005) Differential expression of the Arabidopsis cytochrome c genes Cytc-1 and Cytc-2. Evidence for the involvement of TCP-domain protein-binding elements in anther-and meristem-specific expression of the Cytc-1 gene. Plant Physiology **139:** 88–100

Weltmeier F, Ehlert A, Mayer CS, Dietrich K, Wang X, Schütze K, Alonso R, Harter K, Vicente-Carbajosa J, Dröge-Laser W (2006) Combinatorial control of Arabidopsis proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors. EMBO J **25:** 3133–3143

Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES (2002) Genetic diversity and

selection in the maize starch pathway. Proc Natl Acad Sci. **99:** 12959-12962

Wilson LM, Whitt SR, Ibáñez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell **16:** 2719-2733

Xie Z, Zhang Z-L, Zou X, Huang J, Ruas P, Thompson D, Shen QJ (2005) Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. Plant Physiology **137:** 176–189

Xu X, Chen C, Fan B, Chen Z (2006) Physical and functional interactions between pathogen-induced Arabidopsis WRKY18,WRKY40,and WRKY60 transcription factors. Plant Cell **18:** 1310–1326

Xue Z-T, Xu M-L, Shen W, Zhuang N-L, Hu W-M, Shen SC (1992) Characterization of a Gy4 glycinin gene from soybean *Glycine max* cv. Forrest. Plant Molecular Biology **18:** 897-908

Yadav R, Courtois B, Huang N, McLaren G (1997) Mapping genes controlling root morphology and root distribution in a doubled-haploid population of rice. Theor Appl Genet **94:** 619-632

Yamaguchi-Shinozaki K, Shinozaki K (1994) A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, lowtemperature, or high-salt stress. Plant Cell **6:** 251–264

Yanagisawa S (2000) Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. Plant J **21:** 281–288

Yang T, Poovaiah BW (2002) A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. J Biol Chem **277:** 45049–45058

Yazaki J, Kishimoto N, Nagata Y, Ishikawa M, Fujii F, Hashimoto A, Shimbo K, Shimatani Z, Kojima K, Suzuki K, Yamamoto M, Honda S, Endo A, Yoshida Y, Sato Y, Takeuchi K, Toyoshima K, Miyamoto C, Wu J, Sasaki T, Sakata K, Yamamoto K, Iba K, Oda T, Otomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S (2003) Genomics approach to abscisic acid- and gibberellin-responsive genes in rice. DNA Research **10:** 249-261

Yazaki J, Shimatani Z, Hashimoto A, Nagata Y, Fujii F, Kojima K, Suzuki K, Taya T, Tonouchi M, Nelson C, Nakagawa A, Otomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S (2004) Transcriptional profiling of genes responsive to abscisic acid and gibberellin in rice: phenotyping and comparative analysis between rice and Arabidopsis. Physiol Genomics **17:** 87-100

Yoon MS, Song QJ, Choi IY, Specht JE, Hyten DL, Cregan PB (2007) BARCSoySNP23: a panel of 23 selected SNPs for soybean cultivar identification. Theor Appl Genet **114:** 885-899

Yoshioka S, Taniguchi F, Miura K, Inoue T, Yamano T, Fukuzawa H (2004) The novel Myb transcription factor LCR1 regulates the CO2-responsive gene Cah1,encoding a periplasmic carbonic anhydrase in *Chlamydomonas reinhardtii*. Plant Cell **16:** 1466–1477

Zakharova ES, Epishin SM, Vinetski YP (1989) An attempt to elucidate the origin of cultivated soybean via comparison of nucleotide sequences encoding glycinin B4 polypeptide of cultivated soybean, *Glycine max*, and its presumed wild progenitor, *Glycine soja*. Theor Appl Genet **78:** 852-856

Zhang JZHGAAPGNTTTTNSAKRSBRCN (2001) Locating genomic regions associated with components of drought resistance in rice: comparative mapping within and across species. Theor Appl Genet **103:** 19–29

Zhang W, Gianibelli MC, Ma W, Rampling L, Gale KR (2003) Identification of SNPs and development of allele-specific PCR markers for gamma-gliadin alleles in *Triticum aestivum*. Theor Appl Genet **107:** 130-138

Zhang W, Gianibelli MC, Rampling LR, Gale KR (2004) Characterisation and marker development for low molecular weight glutenin genes from Glu-A3 alleles of bread wheat (*Triticum aestivum*. L). Theor Appl Genet **108:** 1409-1419

Zhao H, Li QZ, Zeng CQ, Yang HM, Yu J (2005) Neighboring-nucleotide effects on the mutation patterns of the rice genome. Genomics Proteomics Bioinformatics **3:** 158-168

Zhao W, Wang J, He X, Huang X, Jiao Y, Dai M, Wei S, Fu J, Chen Y, Ren X, Zhang Y, Ni P, Zhang J, Li S, Wang J, Wong GK-S, Zhao H, Yu J, Yang H, Wang J (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. Nucleic Acids Research **32:** 377-382

Zhao XL, Xia XC, He ZH, Lei ZS, Appels R, Yang Y, Sun QX, Ma W (2007) Novel DNA variations to characterize low molecular weight glutenin Glu-D3 genes and develop STS markers in common wheat. Theor Appl Genet **114:** 451-460

Zheng HBRCPMSALHNCB, H.T N (2000) Quantitative trait loci for root penetration ability and root thickness in rice: Comparison of genetic backgrounds. Genome **43:** 53-61

Zhu T, Shi L, Doyle JJ, Keim P (1995) A single nuclear locus phylogeny of soybean based on DNA sequence. Theor Appl Genet **90:** 991-999

Zhu YL, Song QJ, Hyten DL, Tassell CPV, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics **163:** 1123-1134