RNA-seq based high throughput sequence and expression analysis of selected root architecture genes in wheat

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

MASTER OF PHILOSOPHY
IN
PLANT GENOMICS & BIOTECHNOLOGY

BY

HIRA HAMEED

DEPARTMENT OF PLANT GENOMICS AND BIOTECHNOLOGY
PARC INSTITUTE OF ADVANCED STUDIES IN AGRICULTURE
NATIONAL AGRICULTURAL RESEARCH CENTRE, ISLAMABAD
QUAID-I-AZAM UNIVERSITY, ISLAMABAD, PAKISTAN

2020

RNA-seq based high throughput sequence and expression analysis of selected root architecture genes in wheat

A Thesis
Submitted to Quaid-i-Azam University, Islamabad in the Partial fulfilment of the requirements for the degree of

MASTER OF PHILOSOPHY
IN
PLANT GENOMICS AND BIOTECHNOLOGY

DEPARTMENT OF PLANT GENOMICS AND BIOTECHNOLOGY
PARC INSTITUTE OF ADVANCED STUDIES IN AGRICULTURE
NATIONAL AGRICULTURAL RESEARCH CENTRE, ISLAMABAD
QUAID-I-AZAM UNIVERSITY, ISLAMABAD, PAKISTAN
2020

CERTIFICATE

This thesis submitted by **Ms. Hira Hameed** from Plant genomics and biotechnology, PARC institute of advance study in Agriculture, Islamabad, Pakistan, is accepted in its current form. This thesis fulfils all the requirement for facilitating her with Degree of Master of Philosophy in **Plant Genomics and Biotechnology**.

Internal Examiner: ____________________

Dr. Muhammad Ramzan Khan
Principal Scientific Officer
NIGAB, NARC, Islamabad

External Examiner: ____________________

Dr. Muhammad Zeeshan Hyder
Associate Professor
Department of Bio Sciences,
Comsats Institute of Information Technology (CIIT),
Islamabad

Chairperson: ____________________

Dr. Shoukat Ali
Principal Scientific Officer
NIGAB, NARC, Islamabad

AUTHOR'S DECLARATION

I **Hira Hameed** hereby declare that I have produced the work presented in thesis, "**RNA-seq based high throughput sequence and expression analysis of selected root architecture genes in wheat**" during the scheduled period of study under the supervision of **Dr. Muhammad Ramzan Khan**. I also declare that I have not taken any material from any source except referred to wherever due. If a violation of HEC rules on research has occurred in this thesis, I shall be liable to punishable action under the plagiarism rules of the HEC.

Date: _______________

Hira Hameed
Reg # 02361713005

DEDICATION

I sincerely dedicate this lifetime achievement to my loving and very caring Husband, and My Parents whose prays and support made my path smooth and comfortable to my goals. They are my mentor and I am proud of them.

TABLE OF CONTENTS

LIST OF FIGURES i

LIST OF TABLES ii

LIST OF ABBREVIATIONS iii

ACKNOWLEDGMENTS v

ABSTRACT vi

INTRODUCTION 1

1.1. Factors affecting quality and yield in wheat 2

1.2. Root architecture and plant organs express genes involved in drought stress 3

1.3. High Throughput Sequencing 4

1.4. Transcriptomic studies 6

1.5. RNA-seq as a resource for characterizing transcriptome 7

1.6. Objectives 7

MATERIALS AND METHODS 8

2.1. Plant material and growth conditions 8

2.2. Tissue sampling 9

2.3. RNA extraction 9

2.4. Gel electrophoresis 10

2.5. RNA sequencing and quality control 10

2.6. Read alignment with HISAT 11

2.7. Transcript assembly and quantification with StringTie 11

2.8. Differential expression analysis with Ballgown 12

2.8.1. Data visualization and analysis of variation in the data 12

2.9. Gene Ontology 12

RESULTS 13

3.1. The sequencing data and quality filtering generated for wheat transcriptome 13

3.2. Assembly and quantification of transcripts with StringTie 16

3.3. Identification of differentially expressed genes using Ballgown 17

3.4. Structure and expression of distinct ARD4-like gene isoforms across all samples 19

3.5. Structure and expression of distinct DRO1-like gene isoforms across all samples 21

3.6. Structure and expression of distinct HKT1-like gene isoforms across all samples 23

3.7. Structure and expression of distinct PIN2-like gene isoforms across all samples 25

3.8. Structure and expression of distinct PSTOL-like gene isoforms across all samples 27

3.9. Structure and expression of distinct RPK1-like gene isoforms across all samples 29

3.10. RNA-seq Alignments 31

3.11. Transcriptome Assembly 31

3.12. Gene Ontology 32

DISCUSSION 35

REFERENCES 38

LIST OF FIGURES

Figure No: **Title:** **Page No**

Fig.3.1: Box and whisker plot of the per base sequence quality generated by FastQC quality control program for six wheat cultivars 15

Fig. 3.2: Distribution of FPKM values across the six wheat cultivars 18

Fig. 3.3: Structure and expression levels of six distinct isoforms of the ARD4 like gene 20

Fig. 3.4: Structure and expression levels of six distinct isoforms of the DRO1 like gene 22

Fig. 3.5: Structure and expression levels of six distinct isoforms of the HKT1 like gene 24

Fig. 3.6: Structure and expression levels of six distinct isoforms of the PIN2 like gene 26

Fig. 3.7: Structure and expression levels of six distinct isoforms of the PSTOL like gene 28

Fig. 3.8: Structure and expression levels of six distinct isoforms of the RPK1 like gene 30

Fig.3.9: Gene Ontology results showing Biological Analysis based on Singular Enrichment Analysis 33

Fig.3.10: Gene Ontology results showing Cellular Analysis based on Singular Enrichment Analysis 34

LIST OF TABLES

Table No: Title: Page No:

Table 2.1: List of wheat cultivars and their characteristics selected for RNA-seq**......8**

Table 3.1: The tab-delimited gene abundances Stringtie output file 16

Table 3.2: Differentially expressed transcripts between varities (q value <5%) 18

Table 3.3: Differentially expressed genes between varities (q value <5%)................19

Table 3.4: Read-mapping statistics on the full data set...31

Table 3.5: Transcriptome assembly statistics 32

LIST OF ABBREVIATIONS

µM	Micromolar
BLAST	Basic Local Alignment Search Tool
Bp	Base pair
cDNA	Complementary Deoxyribonucleic Acid
CDS	Coding DNA Sequence
dNTPS	Deoxy Ribonucleotide Triphosphates
DRO1	Deeper rooting 1
FAOST	Food and Agriculture Organization Statistics
IWGSC	International Wheat Genome Sequencing Consortium
MEGA	Molecular Evolutionary Genetics Analysis
NARC	National Agricultural Research Centre
NIGAB	National Institute for Genomics and Advance Biotechnology
QTL	Quantitative Trait Loci
RNA	Ribonucleic Acid
Rpm	Rotation per minute
Spp	species

Sub-spps	Sub- species
TAE	Tris- Acetate-EDTA
USDA	United States Department of Agriculture
GDP	Gross Domestic Product
FAO	Food and Agriculture Organization
ABA	Abscisic Acid
Pi	Inorganic phosphate
PSTOL1	Phosphorus Starvation Tolerance 1
NGS	Next Generation Sequencing
DNA	Deoxyribonucleic Acid
EST	Expressed Sequence Tags
Yr5 gene	Yellow Rust 5
RNA – seq	RNA sequencing
INRES	Institute of Crop Science and Resource Conservation
RNase	Ribonuclease
PE	Paired end
RPKM	Reads Per Kilo Base Pair of Transcript Per Million of the Mapped Reads

ACKNOWLEDGMENTS

In the name of Allah, Most Gracious, Most Merciful

First and foremost, I am very thankful to Allah Almighty for His countless blessings upon me. I would like to acknowledge my admiration to my university PIASA affiliated with QAU Islamabad and all those who provided me the possibility to complete the research. I would like to express my sincerest gratitude to my Supervisor Dr. Muhammad Ramzan Khan, for his knowledge, excellent supervision, patience, continuous support and motivation during my research work. I consider myself fortunate to work under his kind supervision. Besides my advisor, I am also grateful to Chairman, DG, NARC and all staff members of Plant Genomics and Advanced Biotechnology Department, for facilitating the quality of work in our department.

I wish to express my gratefulness to National Institute for Genomics and Advanced Biotechnology, NARC Islamabad for valued insights, priceless assistance and for mentoring and support in the run up to completion of this work. I take this opportunity to express gratitude to all faculty members of PARC Institute of Advanced Studies in Agriculture, Islamabad for providing me with a high standard education and memorable university experience. I am thankful to my Husband, Parents and Siblings for their support throughout for which my mere expression of thanks is not enough. Lastly, I also record my sense of gratitude to one and all who have lent their hand directly or indirectly to this undertaking.

Hira Hameed

ABSTRACT

Drought is a multigenic trait that has a major global impact on crop productivity. The aim of this research is to reveal the gene variants involved in wheat drought tolerance through transcriptomic analysis based on RNA-seq. In this study 6 wheat cultivars namely, Batis, Blue Silver, Local White, UZ-11-CWA-8, Chakwal 50 and Synthetic S22 were transcriptomically analyzed for identification of root growth genes in the research. The International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v1 by HISAT2 was aligned with the default parameters for the identification of their genomic positions (version 2.0.5). HISAT2 gave overall alignment rate for Batis 92.95%, Blue Silver 92.16%, Local White 92.92%, UZ-11-CWA-8 93.95%, Chakwal-50 94.43% and Syn-22 92.49%, respectively. After alignment, StringTie an RNA-seq abundance estimator assembled and quantified transcripts from alignments generated by HISAT2; isoforms producing primary results in the GTF file comprising assembled and fully covered transcripts that match the reference annotation with the gene abundance files. StringTie produced 6 GTF files for all cultivars with full covered, end-to-end transcripts in the reference annotation, with reads showing total 608 fully covered transcripts for Batis, 617 Blue Silver, 574 Local White, 567 UZ-11-CWA-8, 551 Chakwal-50 and 542 Syn-22, respectively. Similarly, 6 GTF transcript files containing information of the transcripts produced by StringTie from RNA-seq information display a total of 147974 transcripts assembled for Batis, 149383 Blue Silver, 147715 Local White, 148850 UZ-11-CWA-8, 147711 Chakwal-50 and 147802 Syn-22. Ballgown package identified ARD4, DRO1, HKT, PIN2, PSTOL, and RPK1 genes which have major role in root architecture and their differential expressions in wheat homoeologous genomes. RNA-seq studies, showed that expression variations in genes in both drought sensitive and drought tolerant varieties. AgriGO tool used Singular Enrichment Analysis to produce cellular and biological gene ontology analysis which showed gene interaction pathways. Our findings unveiled genes to be tested through functional genomics studies under the impact of drought stress loci. These findings will have future consequences for plant genetic improvement against abiotic stress by editing root architecture genes to improve plant yield.

INTRODUCTION

Wheat is more than one third of the world's population's most significant sustainable food crop, adding more calories and proteins to the world's diet than any other cereal crop (Kumar *et al*.,). Approximately 10,000 years ago, wheat was one of the oldest grown food plants and is presently one of the world's most grown food grains (Shewry, 2009). Over 50% of the daily caloric intake in the world is derived directly from the consumption of cereal grains (Sarwar *et al.,* 2013). Wheat is a major carbohydrate source, with a protein content of about 13%, which is comparatively high compared to other major cereals, but protein quality of essential amino acids is relatively low (Iametti *et al*., 2006).

Wheat is cultivated in more land as compared to other food crop i.e., 220.4 Million hectares, 2014 (Godfray *et al*., 2010), (FAOSTAT, 2014). In 2017, world production of wheat was 730 Million tons, with a production estimate of 766 Million tons for 2019, making it the second most produced cereal after maize (FAOSTAT, 2016). World wheat and other grain crops have tripled since 1960 and expected to continue to rise through the mid-21st century (Godfray *et al*., 2010).

Pakistan is an agricultural country that relies on agriculture for about 70% of its economy (Peter, 2015). Pakistan is the third biggest wheat producing nation in Asia (Brown, 1968); wheat production in 2015-2016 was approximately 25.5 Million tons, although Pakistan is in 9th place in the field of wheat agriculture, based on complete output per hectare. The major crop of Pakistan are wheat, rice, maize, sugarcane and cotton. According to Pakistan's economic survey, the contribution of these crops in the agricultural sector is about 25.6% and the contribution of major crops in GDP is nearly 5.4%. The share of wheat in the agricultural contribution is approximately 10.3% and 2.2% of GDP. In 2010, FAO evaluated that wheat is grown on 4% of the cultivated territory and harvested on 217.0 Million hectares, with a yield of 650.9 Million hectares (Bailey *et al*., 1976; USDA and FAOSTAT, 2014).

1.1. Factors affecting quality and yield in wheat

Wheat, a major cereal crop, is affected by several biotic and abiotic stresses. Globally, these stresses affect the yield of the crop (Suzuki, Rivero, Shulaev, Blumwald, and Mittler, 2014). Biotic stress results from damage done by other living organisms to an organism, like bacteria, viruses, fungi, worms, advantageous and harmful insects, weeds, and cultivated or native plants (Dreher and Callis, 2007); whereas Abiotic stress is the negative impact on living organisms in a particular environment of non-living factors (Bohnert, 2006). The common abiotic factors are High winds, Extreme temperatures, Drought and flood. There are many other factors of abiotic stress that are less recognizable but constantly affect the environment. These include weak edaphic factors such as rock content and pH, excessive radiation, compaction, contamination and many other extremely specific conditions such as rapid seed germination rehydration (United States Patent No. US6559099B1, 2003).

Water shortage circumstances, changes in temperature and elevated salt content are environmental issues that adversely decrease plant growth and yield worldwide (Anjum *et al*., 2011). Plant creates a complicated molecular network to regulate these circumstances by making phenotypic and anatomical modifications (Wang *et al*., 2013).

Drought is a significant reason that has lessened crop production globally (Boyer, 1982; Wang *et al*., 2003). The dynamic changes in climate are the primary reason for accelerating drought condition that have profound effects on temperature and moisture that demonstrate reduced crop yields (Ergenand and Budak, 2009; Fleury *et al*., 2010). New wheat cultivars for various abiotic elements are enhanced by using *Triticum* species and higher yielding crops are needed owing to enhanced food requirements (Habash *et al*., 2009; Fleury *et al*., 2010). Diversity is generated in newly grown species by sidestepping drought condition. In research findings linked to drought crops, different species and sub-species reveal genetic diversity plays significant role. Preventing drought conditions adds to crucial role in wheat species research, which also involves a source of gene appropriate to drought and a region of gene that can be used to improve crops (Ashraf and Ozturk, 2008). Genetic improvement serves noteworthy role for stable crop production in those regions where drought is the primary danger.

Advance methods are also the key to identifying and characterizing the gene and gene region appropriate to drought. The drought-related gene of wheat can be identified using transgenic or breeding techniques (Pennisi, 2008).

1.2. Root architecture and plant organs express genes involved in drought stress

The spatial distribution of all root parts is known as root system architecture in development setting and this root architecture is extremely responsive to external environments such as nutrients, water content, temperature and pH (Robbins and Dinneny, 2015). Root supplies nutrients and water for plant growth, and this is the main component of crops. The root of plant comprises of three parts called as primary root, lateral root and the adventitious root. First, the primary root is developed during germination from the basal end of the plant embryo and second, the lateral root developed from the main roots and are highly sensitive to environmental stress and third, the adventitious roots emerge from stem or hypocotyls (Tian *et al*., 2014). Water and nutrients available in soil are the primary source responsible for development and growth of root. It is accountable for discovering root features that can improve plant productivity (Kano *et al*., 2011; Grossman and Rice, 2012). Many researches are underway to enhance the root characteristic so that it can readily absorb more resources from the soil and withstand circumstances of abiotic stress such as drought to enhance plant yield (Kell, 2011; Narayanan *et al*., 2014; Meister *et al*., 2014).

The root is entire complex and adaptable subterranean plant structure (Berntson, 1994). Primary root of plant builds the tap root by expedition, it is the plant's prominent root and accountable for generating many tiny branch roots that make the root system very complicated. The primary root lifetime in monocots like wheat is very brief and it produces numerous adventitious roots, this type of root system is termed as a fibrous root system with a broad mass of roots of the same size. In Plate root system with big horizontal lateral roots continuing to spread under the surface of the soil, from which tiny roots branch vertically downwards (Cannon, 1949). The suitable number and position of lateral roots plays an important role in the absorption of water, nutrients and excellent anchorage (Varney and Canny, 1993; Bailey *et al*., 2002).

Root depth is a significant characteristic of plant growth as it enables greater access to nutrients and water stored in profound soil layers (Wasson *et al*., 2012). Higher yields directly depend on deep and vigorous root system (Lynch and Wojciechowski, 2015). Root characteristics are regulated by polygenes with quantitative impact and are extremely influenced by the environment. The genetic loci which controls these characteristics are the quantitative trait loci (Hunt and Kirkegaard, 2012).

Recent developments have been produced in discovery of genes that control root and shoot growth in relation to drought using both genetics and functional genomics methods. There have been reports of numerous genes and transcription factors engaged in various processes such as closing of stomata, lignification, and root architecture. These genes are induced by concentration of ABA, NO, Pi, CO, Auxin, CK and salts. Genetic methods based on analysis of rice genotypic and phenotypic diversity have enabled identification of significant genes engaged in growth various kinds of root system architectures, such as DRO1 and PSTOL1, with features linked to drought tolerance and low phosphate soil adaptation (Gamuyao *et al*., 2012 ; Uga *et al*., 2013).

1.3. High Throughput Sequencing

During the mid to late 1990s, several new methods for DNA sequencing were developed and implemented by the year 2000 in commercial DNA sequencers. they were collectively called as "next-generation" or "second-generation" sequencing (NGS) methods (Behjati and Tarpey, 2013). Current genetic material sequencing techniques are dedicated to incremental progress in the dideoxy chain termination method of Fred Sanger (Sanger *et al*., 1977). Different genome sequencing strategies have also been developed. Shotgun sequencing approach (Anderson, 1981) is the most skilled technique in genome sequencing projects, which involves breaking large DNA segments into smaller segments, followed by their sequencing. The genome is examined by clone-based physical techniques in clone-by-clone approach and then the gene of interest containing sections are appointed for further analysis (Hattori *et al*., 2000), although the whole genome shotgun strategy carries out sequencing of the entire genome cloned (Fleischmann *et al*., 1995). High-throughput sequencing applies to genome sequencing, genome resequencing, transcriptome profiling (RNA-seq), DNA-

protein interactions (ChIP sequencing) and epigenome characterization (de Magalhães, Finch, and Janssens, 2010).

High throughput sequencing is the technique that is used in many genome projects. The approach is carried out as; extraction and fragmentation of the genome sequence material, generation of cloned libraries by incorporating DNA fragments into vectors, sequencing fragments with NGS tools which generates short sequence reads in large numbers, mapping the reference genome reads to the genome guided assembly and otherwise de-novo assembly of the genome by describing sequence overlaps and finally the annotations of the assembled genome to recognize important characteristics. The growing demand for low-cost sequencing has led to the development of high-performance sequencing technologies that parallel the sequencing process and simultaneously generate thousands or Millions of sequences (Bosch and Grody, 2008).

Genome projects have elegantly dedicated themselves to understanding the genetic and molecular foundation of various life functions. Analysis of global gene expression as in RNA / transcriptome sequencing was effectively studied through a short-read sequencing approach (Mortazavi *et al*., 2008). Gene expression can be evaluated both qualitatively and quantitatively and differs based on tissue and maintenance demands between distinct genes and tissues. The sequencing method has been effectively operational for various agricultural initiatives, including whole-genome sequenced rice mutant resource (Li *et al*., 2017), soybean gene expression transcriptome profiling (Krupp *et al*., 2012).

The growing availability of sequenced genomes for both model and non-model species has accelerated worldwide gene expression research based on NGS and made it possible to characterize gene functions in distinct organisms. Global gene expression studies based on NGS have been promoted by enhancing accessibility of sequenced genomes are ready to represent gene functions in various organisms.

1.4. Transcriptomic studies

The transcriptome is the collection either in single or in population of cells of all RNA molecules. High-quality bread wheat transcriptomics is a powerful technique of enhancing molecular understanding of bread wheat biology. However, the use of NGS techniques in wheat transcriptome analysis is not yet commonly operational (Edwards *et al*., 2013).

Microarray techniques were used to analyze hexaploidy wheat transcriptome. High-density microarrays were generated from a publicly accessible wheat EST resource comprising 26,382 sequences, depending on 35 individual cDNA libraries showing specific developmental phases of distinct grain tissues and plant development (Wilson *et al*., 2004). From this means, a wheat cDNA microarray of 9,155 features was defined and used to monitor changes in the transcriptome of wheat embryo during late grain development and maturation and during the first 48 hours of post-imbibition germination (Wilson *et al*., 2005). In addition, a microarray of 9,000 cDNA wheat gene was generated from cDNA libraries mainly generated from wheat seed development (Gregersen *et al*., 2005).

Affymetrix wheat Gene Chip oligonucleotide array has produced multiple high-quality gene expression datasets with more than 61,127 sample sets representing 55,052 transcripts (Coram *et al*., 2008a). This GeneChip has been used to analyze the grain growth transcriptome (Coram *et al*., 2008b) and to identify the genes engaged in creating low temperature tolerance (Laudencia-Chingcuanco *et al*., 2011). The wheat GeneChip was also genuine in profiling modifications that occurred after inoculation with *Puccinia striiiformis*, the causative agent of orange rust in wheat lines that differed with the Yr5 gene (Coram *et al*., 2008b).

Combining wheat transcriptome information with isolated wheat cell arm assemblies will probably result in higher understanding of composition, expression, and evolution of wheat genome. Such a research will have consequences in polyploid plant genomics, not only in wheat, but also more widely.

1.5. RNA-seq as a resource for characterizing transcriptome

RNA-seq is a transcriptome profiling technique that uses deep-sequencing techniques, and the Illumina sequencing-by-synthesis method is the most popular platform of this technology. It covered the mode of discovering huge amounts of transcriptional information and led to the recognition that even in the simplest eukaryotic genome, gene regulation is highly complex.

Illumina technology is appropriate for transcriptome studies due to extensive reading depth and pair-reading technology. Deep sequence coverage, provided more easily by Illumina's NGS platform than 454 sequencing, is essential for gene detection and gene expression analysis (Varshney *et al*., 2009; Barski *et al*., 2007; Johnson *et al*., 2007).

A population of RNA (complete or fractional, such as poly (A)) is commonly converted into a library of pieces of cDNA with adapters attached to one or both ends. Each RNA molecule is then sequenced from one end (single-end sequencing) or from both ends (pair-end sequencing) in a high-throughput measured way. Depending on the DNA sequencing technology used, the reads are usually 30-400 bp.

RNA – Seq is also expected to distinguish all RNA transcripts generated at a moment as well as variants of transcripts induced by differential gene splicing (Varshney *et al*., 2009). Before using contemporary wheat genetic enhancement instruments, a full transcriptomic profile of wheat with drought relationship is inevitable. This is particularly essential for dryer climatic cultivars.

1.6. Objectives

The main objective of the study was to sequence the transcriptome of various wheat cultivars. Another aim of this thesis was to reveal genetic variations in transcripts to identify genes involved in root traits through RNA-seq based transcriptomic analysis.

MATERIALS AND METHODS

2.1. Plant material and growth conditions

The transcript-level expression analysis was performed using the six varieties of hexaploid wheat (*Triticum aestivum*) including Batis, Local White, Blue Silver, Chakwal-50, UZ-11-CWA-8 and Syn-22. Drought tolerant varieties include UZ-11-CWA-8, Local White, Chakwal 50 and Syn-22, whereas elevated yield but drought-sensitive varieties are Batis and Blue-Silver. The Institute for Bio-resource Conservation, National Agricultural Research Center, Islamabad provided seeds of Local White, Chakwal 50 and Blue Silver. Batis, UZ-11-CWA-8 and Syn22 seeds were obtained from Institute of Crop Science and Resource Conservation (INRES) at University of Bonn, Germany (Table 2.1).

Table 2.1: List of wheat cultivars and their characteristics selected for RNA-seq

Sr. No	Sample	Identifier/Cultivar	Characteristics
1	G1	Batis	Batis, a winter bread wheat variety is released in Czech Republic in 2001.
2	G2	Blue Silver	This bread wheat variety is released in 1971 and suitable for irrigated areas of Pakistan.
3	G3	Chakwal 50	This Drought tolerant, bread wheat variety is released in 2008 and cultivated in rain fed regions of Pakistan.
4	G4	Local White	Local White, is a Drought tolerant land race wheat variety from dryer areas of Pakistan, particularly in Baluchistan.
5	G5	Syn-22	It is Synthetic wheat from INRES, Bonn University.
6	G6	UZ-11-CWA-8	This wheat variety is drought tolerant line from Uzbekistan.

The seeds were grown at the National Institute for Genomics and Advanced Biotechnology, NARC. The growth rooms, chambers were equipped with supplemental lighting necessary to grow the wheat cultivars. 16" x 16" trays were used for the raising the plants seedlings. In order to avoid roots from expanding through the holes in the bottom of the tray, the seed trays were lined with paper towels. A dry, loose organic compost or potting soil layer even 1-1.5 "has been distributed in seed trays. The seeds were planted across the top of the compost in even layers and then slightly pressed into the soil, in a manner that they are not buried completely. To initiate seed germination, soil was held moist. In each tray, plants were distributed into two columns. Each tray constitutes a single collection. The trays were labeled with styrene labels. In total, our experiments contain 6 specimens.

2.2. Tissue sampling

Roots were obtained manually from potting soil and rinsed in the profound plastic ice bucket with ice cold water. Root tips were sliced and put with liquid nitrogen in the properly labeled Styrofoam cup. The plant samples were then stored at -80°C in labeled aluminum foil instead of freeze-drying. The samples were then ground resulting in a homogeneous mixture of tissue from wheat cultivars. The ground material was separated into multiple labeled 2ml screw-top tubes. During grinding and aliquoting, the samples were all kept frozen using liquid nitrogen and placed at -20°C.

2.3. RNA extraction

GeneJET Plant RNA purification Mini reagent kit (Thermo Fisher Scientific, Inc., USA) was used to extract RNA from wheat cultivar root tissues. 100 mg of plant root tissue was placed in liquid nitrogen and thoroughly grinded with pestle and mortar. Tissue powder was transmitted to a 1.5 ml micro-centrifuge tube with 500 µl of Plant RNA Lysis Solution soon after grinding. The sample was then incubated at 56°C for 3 minutes. After incubation, samples were centrifuged at 14,000 rpm for 5 minutes. The supernatant (generally 450-550 µl) was collected in the clean micro-centrifuge tube and added 96% ethanol to 250 µl. The well-mixed prepared solution was then transferred to a purification column that was inserted into a collection tube and centrifuged at 11,000 rpm for 1 minute. The Flow-through solution was removed, whereas column

along with collection tubes were reassembled. After that, in the purification column, 700 µl of Wash Buffer (WB1) was added and centrifuged at 11,000 rpm for 1 minute. Flow-through solution with collection column is then removed. In a new clean 2 ml collection tube the purification column is placed. In this case, Wash Buffer (WB2) 500 µl is added to the purification column and centrifuged at 11,000 rpm for 1 minute. Flow-through solution was discarded yet again, but the column and collection tube were reassembled. According to the manufacturer's instructions, the step with the Wash Buffer (WB2) was repeated and the column was re-spun at 14,000 rpm for 1 minute. The flow-through solution comprising the collection tube was subsequently removed. The purification column was transmitted to a collection tube free of RNase 1.5 ml. To elute the RNA, 50 µL of nuclease-free water was added to center of purification column membrane and centrifuged at 11,000 rpm for 1 minute. The purified RNA obtained was immediately stored at -20°C but for prolonged storage it was kept at -70°C.

2.4. Gel electrophoresis

The total RNA was run on 1.5% agarose gel prepared in 1% TAE buffer to verify the quality. For gel preparing, 1.5 gm agarose was dissolved in 100ml of TAE buffer. Ethidium bromide has been introduced in gel for RNA staining. The RNA samples were mixed with 6X loading dye (bromophenol blue) and loaded into gel wells. For further analysis, images were taken using the gel documentation scheme. NanoDrop 2000 spectrophotometer (ND-2000, Thermo Fisher Scientific, Inc., USA) was also used to measure RNA concentration. An Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., CA, USA) evaluated the integrity of RNA. Finally, for RNA sequencing, the equal quantity of RNA of each variety was used for the same treatment.

2.5. RNA sequencing and quality control

Commercial company (GenXPRo Frankfurt, Germany) sequenced RNA samples of wheat cultivars. Subsequently, paired end (PE) RNA-seq libraries were prepared and High-throughput sequencing was performed on an Illumina NGS HiSeq2000 platform in accordance with standard protocol at commercial company (GenXPRo Frankfurt, Germany). FastQC (https:/www.bioinformatics.babraham.ac.uk/projects/fastqc/) evaluated the quality of the raw data acquired for sequencing and clean reads were

obtained. Trimmomatic (http:/www.usadellab.org/cms/index.php?page=trimmomatic) was used afterwards to remove the contamination of adaptor, reads of low quality (quality scores < 20), reads more than 10% bases with ambiguous "N" bases and reads less than 15 bases. The number of reads per lane was substantially high, ranging from 9.5 Million to 12.3 Million per sample. Average reads were about 10 Million per sample, which is a good number. Reads were 100 nt in length. Individual sample read qualities were high as FastQC software had verified. All quality average scores were 20 or greater than 30.

2.6. Read alignment with HISAT

Pre-processed reads fetched by FastQC (http:/www.bioinformatics.babraham. ac.uk/projects/fastqc) can assist guarantee the initial spliced alignment phase is supplied with high-quality data. Six varieties of clean reads have been aligned with the wheat reference genome The International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v1.0(International Wheat Genome Sequencing Consortium (IWGSC) *et al*., 2018) by HISAT2 (http:/cccb.jhu.edu/software/hisat2/index.shtml) with the default parameters for identifying their genomic positions (version 2.0.5)(Kim, Langmead, and Salzberg, 2015). HISAT2 program requires raw reads (fastQ files) as an input and eventually sorted BAM files are generated using SAM tools (http:/samtools.sourceforge.net/) this mapping data enables the collection of subsets of reads corresponding to each gene, assembly and quantification of transcripts represented by those reads.

2.7. Transcript assembly and quantification with StringTie

StringTie (http://ccb.jhu.edu/software/stringtie) assembles RNA-seq transcripts reads aligned with the genome, first grouping the reads into distinct gene loci and afterwards assembling every locus in as much isoforms as necessary to explain the data. The annotation file used by StringTie includes a specification of the exon – intron structure for ' known ' genes, including names for genes as a guide for assembly, and thus genes have been tagged and assembled with names from that file. Reconstruction of low-abundance genes was feasible with annotation. Subsequently, the merge function of

StringTie was conducted to merge all assemblies together. Gffcompare utility (https:/github.com/gpertea/gffcompare) was used after assembling transcripts with StringTie to examine transcript statistics in a manner that matches assembled transcripts with annotated genes, in whole or in part, and calculate how many are completely novel.

2.8. Differential expression analysis with Ballgown

Differential expression genes were identified using Ballgown (R Package) by offering the abundance data produced by StringTie and the phenotype information describing the samples and analysis was performed using a linear model. Typically, the FPKM values attached to transcripts are extremely skewed; hence, the built-in features of Ballgown apply a log transformation to stabilize the variance and then match conventional linear models that can be used to test for differential expression. The adjusted p-value < 0.05 and fold-changes > 2 (Log2 (treatment / control) = 1) have been used as thresholds for differential expressed analysis.

2.8.1. Data visualization and analysis of variation in the data

The R programming language (https:/www.r-project.org) and the Bioconductor software suite are hosting an extensive collection of tools for handling functions ranging from raw information plotting to standardization and downstream statistical modeling. Some other R packages such as RSkittleBrewer (for setting up colors), genefilter (for quick calculation of means and variances), dplyr (for sorting and arranging outcomes) and devtools (for reproducibility and package installation) helped to display the outcomes. As part of the R / Bio-conductor package, Ballgown involves plotting tools that make plots easier to view and compare expression information.

2.9. Gene Ontology

AgriGO offers an intuitive and fair interface for analyzing genomic data from more th an 30 plant species in Gene Ontology (GO). The AgriGO (Tian *et al*., 2017) has determined GO conditions that are considerably over-represented in each cluster. AgriGO used Singular Enrichment Analysis (SEA) to identify over-represented GO categories in each cluster relative to the entire genes.

RESULTS

3.1. The sequencing data and quality filtering generated for wheat transcriptome

A commercial company (Frankfurt, Germany) has sequenced an RNA sample of wheat cultivars. RNA sequencing was subsequently carried out using the platform Illumina NGS HiSeq2000. Input Read Pairs were 48628107, respectively, obtained from Batis samples. Trimmomatic tool was used to trim and clean the reads. After trimming, there were 47172834 numbers of cleaned reads, which is 97.01% of the total reads. We used FastQC to estimate quality of the sequence, which aims to provide a simple way to raw sequence data quality control. FASTQC report stated that eligibility of Batis for both forward and reverse pair per baseline sequence quality test shown in Fig. 3.1(A). HISAT2 sensitive alignment program was used to map next-generation sequencing reads against reference genome. The overall alignment rate produced by HISAT2 for Batis is 92.95%.

Input Read Pairs obtained from Blue-Silver samples respectively were 48509084. Trimmomatic tool was used to clean and trim the reads. After trimming, there were 46664681 numbers of cleaned reads, which is 96.20% of the total. FastQC was used to predict sequence quality, which seeks to provide an easy way to raw sequence data quality control. FASTQC report mentioned that per base sequence quality test was qualified for both forward and reverse pair for Blue Silver shown in Fig. 3.1(B). HISAT2, a fast alignment program for mapping next-generation sequencing reads against reference genome. The overall alignment rate produced by HISAT2 for Blue Silver is 92.16%.

Input Read Pairs obtained from samples from Chakwal-50 were respectively 48373716. 46617180 clean reads were obtained after trimming with the Trimmomatic tool, which is 96.37% of total reads. FASTQC report showed that the quality test per base sequence was qualified for the Chakwal-50 forward and reverse pair shown in Fig. 3.1(C). HISAT2 generated 92.92% overall alignment rate for Chakwal 50.

Input Read Pairs obtained from Local-White samples were 52373682, respectively. Cleaned read numbers were 50646812, which is 96.70% of the total. FastQC was used

throughout all sequences to predict sequence quality. FASTQC report mentioned that the quality test for the Local-White forward and reverse pair is qualified per base sequence shown in Fig. 3.1(D). HISAT2, aligner produced 93.95% overall alignment rate for Local White.

Input Read Pairs acquired from Syn-22 samples were 50556166, respectively. The number of cleaned reads was 48779601 after trimming, which is 96.49 % of the total. FastQC is used to predict sequence quality. FASTQC report mentioned that the quality test per base sequence is qualified for the Syn-22 forward and reverse pair shown in Fig. 3.1(E). HISAT2 produces 94.43% overall alignment rate for Syn-22.

Input Read Pairs acquired from UZ-11-CWA-8 specimens were 51537979, respectively. The number of reads cleaned after trimming was 49806318, which is 96.64% of the total. FastQC provides an easy way to regulate raw sequence data quality. The FASTQC report mentioned that the quality test per base sequence is qualified for forward and reverse pair shown in Fig. 3.1(F). HISAT2, generated overall alignment rate of 92.49% for UZ-11-CWA-8.

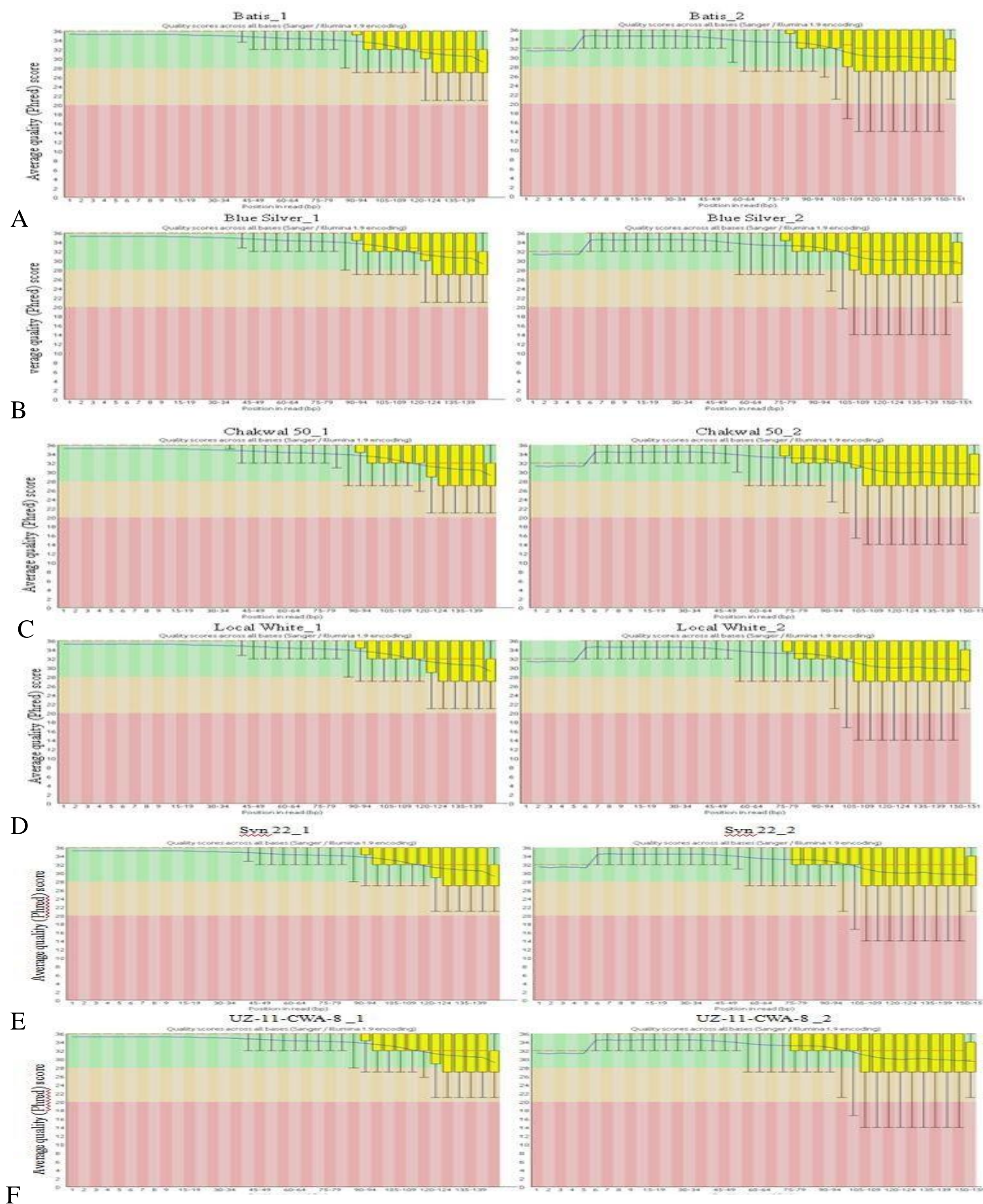

Fig.3.1: Box and whisker plot of the per base sequence quality generated by FastQC quality control program for six wheat cultivars (1) forward paired ends (2) reverse pair ends x-axis shows base position in reads all samples contain 135-139 reads for forward pair ends and 150-151 for reverse pair ends. Whereas, y-axis shows quality score for all samples, the quality values do not drop lower than 30 which is good quality range. The background green, orange and red colors represents base calls of good, reasonable and poor quality, respectively.

3.2. Assembly and quantification of transcripts with StringTie

StringTie takes input binary SAM /BAM file hence assembles transcripts from RNA-seq reads that were aligned with the genome, this instrument employs a network flow algorithm that starts with the most widely expressed transcript that has been simultaneously assembled and quantified. The annotation file used by StringTie incorporates a specification of the exon – intron structure for ' recognized ' genes, along with names for those genes as an assembly guide, and therefore genes have been tagged and assembled with names from that file. Reconstruction of low-abundance genes was feasible with annotation. StringTie produces 6 GTF files for all cultivars in the reference annotation with full covered, end-to- end transcripts, with a total of 608 fully covered transcripts for Batis, 617 Blue-Silver, 574 Chakwal-50, 567 Local-White, 551 Syn-22 and 542 UZ-11-CWA-8, respectively. Likewise, 6 GTF transcript files containing RNA-seq transcript information provided by StringTie displaying a total of 147974 transcripts assembled for Batis, 149383 Blue-Silver, 147715 Chakwal-50, 148850 Local-White, 147711 Syn-22 and 147802 UZ-11-CWA-8. Subsequently, the merge feature of StringTie was carried out to merge all assemblies together. The gffcompare utility is used to compare one or more GTF files generated by StringTie with a reference annotation file that eventually makes it possible to verify how the transcripts anticipated relate to an annotation file. The tool produces gffcmp.stats file also includes several other measures, such as number of novel exons, introns contained in StringTie output.

Table 3.1: The tab-delimited gene abundances Stringtie output file

Gene ID	Gene Name	Ref	Strand	Start	End	Coverage	FPKM	TPM
ENSRNA 050013913	5_8S_rRNA	1A	+	7060	7215	2.756413	0.236575	0.313334
ENSRNA 050013874	LSU_rRNA _eukarya	1A	+	12317	13827	2.95571	0.25368	0.335989
ENSRNA 050013873	LSU_rRNA _eukarya	1A	+	15322	16832	6.776106	0.581573	0.770271
ENSRNA 050013836	SSU_rRNA _eukarya	1A	+	20009	21818	38.70854	3.322241	4.400175
ENSRNA 050013914	5_8S _rRNA	1A	+	22045	22200	3.647436	0.313049	0.414621
ENSRNA 050013853	SSU_rRNA _eukarya	1A	+	28456	30266	29.76136	2.55433	3.383109
ENSRNA 050013918	5_8S _rRNA	1A	+	30493	30648	0.083333	0.007152	0.009473

3.3. Identification of differentially expressed genes using Ballgown

StringTie is intended to create a connected collection of tables that can be read straight into R using the Ballgown package features. Ballgown package is meant to promote the assessment of RNA-seq information with flexible differential expression. It also offers features for the organization, visualization and analysis of transcriptome assembly expression measurements. It requires data loading in the form of csv (comma-separated values) file, which includes RNA-seq sample information. The phenotypic data must be loaded to describe each sample in one line of the file, and one variable should be included in each column. Expression information were read afterwards, calculated by StringTie, and genes with low abundance were filtered for removal. A prevalent problem with RNA-seq information is that genes often have very few or zero numbers, so all transcripts with less than one sample variance have been deleted. Transcripts showing statistically significant differences between groups were recognized using the Ballgown stat-test feature. The statistical test of Ballgown is a conventional linear model-based comparison using a cumulative upper quartile standardization (Paulson, Stine, Bravo, and Pop, 2013). Similarly, genes showing statistically significant differences between groups were recognized using Ballgown's stattest feature. The most important step is to add gene names and Gene I d's to Transcripts, and the result obtained is sorted from the lowest to the largest P value, producing a table of Differentially expressed transcripts with q value <5%. It was shown that gene abundances (measured as FPKM values) are distributed across samples, colored by variation shown in Fig. 3.2. Ballgown holds a range of measurements that can be compared and viewed, here FPKM measurements were compared for the transcripts. Better visualization is provided by log 2 conversion has been performed which adds one to all FPKM value as log 2 (0) is undefined and has been generated as plot.

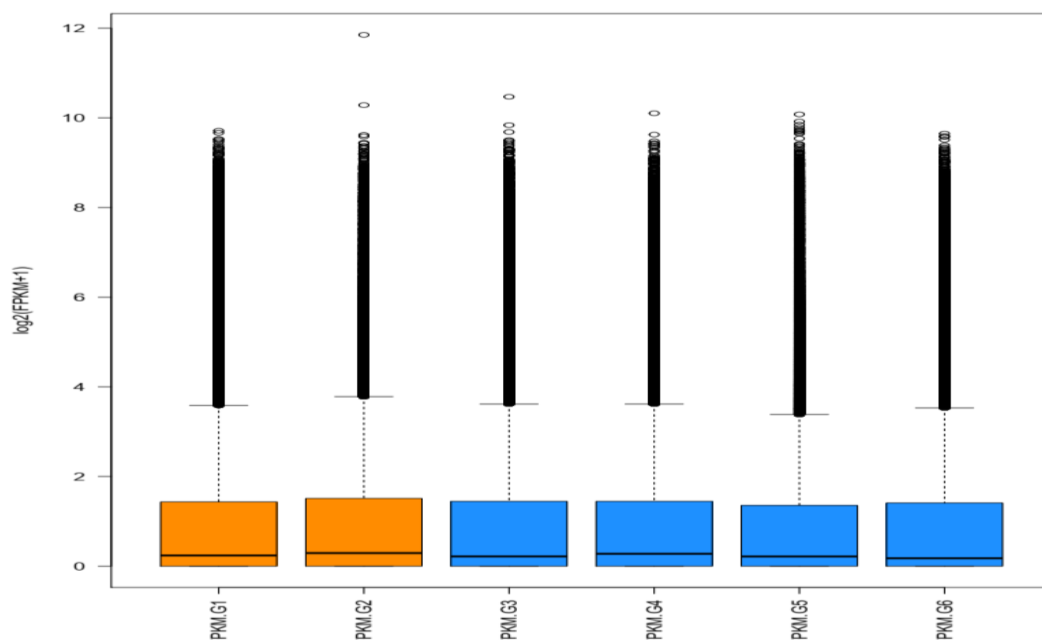

Fig. 3.2: Distribution of FPKM values across the six wheat cultivars This figure of six wheat cultivars i.e., (left to right) Batis, Blue-Silver, Chakwal-50, Local-White, Syn-22, and UZ-11-CWA-8 shows the distribution of gene abundances measured as FPKM values across samples, colored by category i.e., drought tolerant in represented in blue, and drought sensitive is represented in orange color.

Table 3.2: Differentially expressed transcripts between varities (q value <5%)

Gene names	Gene id	Feature	Id	Fold change	P-value	Q-value
.	MSTRG.54886	transcript	137261	0.304847	1.84E-11	2.90E-06
.	TraesCS6D02G078900	transcript	152889	0.574415	2.79E-10	2.20E-05
LSU_rRNA_eukarya	ENSRNA050018652	transcript	193469	0.816101	5.11E-06	0.013436
U1	ENSRNA050023660	transcript	114861	0.941366	1.46E-05	0.027521
U2	ENSRNA050022760	transcript	82382	0.979597	2.14E-05	0.035818
Plant_SRP	ENSRNA050023735	transcript	92278	0.961883	9.13E-05	0.08844
SSU_rRNA_eukarya	MSTRG.167	transcript	42	0.588263	0.000564	0.257878
5S_rRNA	MSTRG.12741	transcript	31203	0.990905	0.000657	0.275193
SNORD25	MSTRG.60613	transcript	151292	0.392912	0.00089	0.309083
Intron_gpII	MSTRG.29933	transcript	74189	0.703406	0.000954	0.319723
5_8S_rRNA	MSTRG.75824	transcript	192835	0.76753	0.000983	0.323986
rps15-A	MSTRG.38030	transcript	94842	0.998844	0.001348	0.367046

Table 3.3: Differentially expressed genes between varities (q value <5%)

Feature	Id	Fold change	P-value	Q-value
Gene	MSTRG.54886	0.304864	2.14E-11	7.36E-07
Gene	MSTRG.20291	0.317939	3.50E-07	0.005305
Gene	MSTRG.28511	0.188768	5.69E-07	0.005305
Gene	MSTRG.54860	0.207911	7.11E-07	0.005305
Gene	MSTRG.69745	0.249286	7.71E-07	0.005305
Gene	MSTRG.43109	5.900321	2.18E-06	0.010867
Gene	MSTRG.54854	0.386005	2.65E-06	0.010867
Gene	MSTRG.24411	0.282035	3.24E-06	0.010867
Gene	MSTRG.15070	0.230089	3.47E-06	0.010867
Gene	MSTRG.49295	1.811291	3.75E-06	0.010867

3.4. Structure and expression of distinct ARD4-like gene isoforms across all samples

A putative candidate gene encoding for "acireductone dioxygenase (OsARD4)," is engaged in ethylene and polyamine biosynthesis in a shallow-rooted rice genotype, ASD16, has shown its role in stimulating crown/lateral root development and, thereby, modulating rooting patterns (Chen *et al.*, 2013). The properties of gene make it important constituent of root architecture as extensive rooting patterns provides better tendency to withstand stress. According to the Id column, 6,720th transcript in the Ballgown object reveals that TraesCS1A02G375800.1 is an isoform of ARD4-like gene known to be much more expressed in Chakwal-50, Local-White, UZ-11-CWA-8 than in Blue-Silver and less expressed in Syn-22 than in Batis in genome A shown in Fig 3.3 (A). Similarly,15492th transcript showed that TraesCS1B02G396200.1 ARD4 like gene's isoform is more expressed in Chakwal-50, Local-White,UZ-11-CWA-8 than in Blue-Silver and less expressed in Syn-22 than in Batis in genome B shown in Fig 3.3 (B). Whereas, 23704th transcript, i.e.,TraesCS1D02G383100.1 isoform showed that ARD4 like gene is highly expressed in Chakwal-50,Local-White,UZ-11-CWA-8 than in Blue-Silver and less expressed in Batis than Syn-22 at genome D shown in Fig 3.3 (C).

Average expression for all ARD4 transcripts for genome A, B and D indicates that a drought-tolerant organization has highest transcript expression than a drought-sensitive variety organization shown in Fig 3.3 (D).

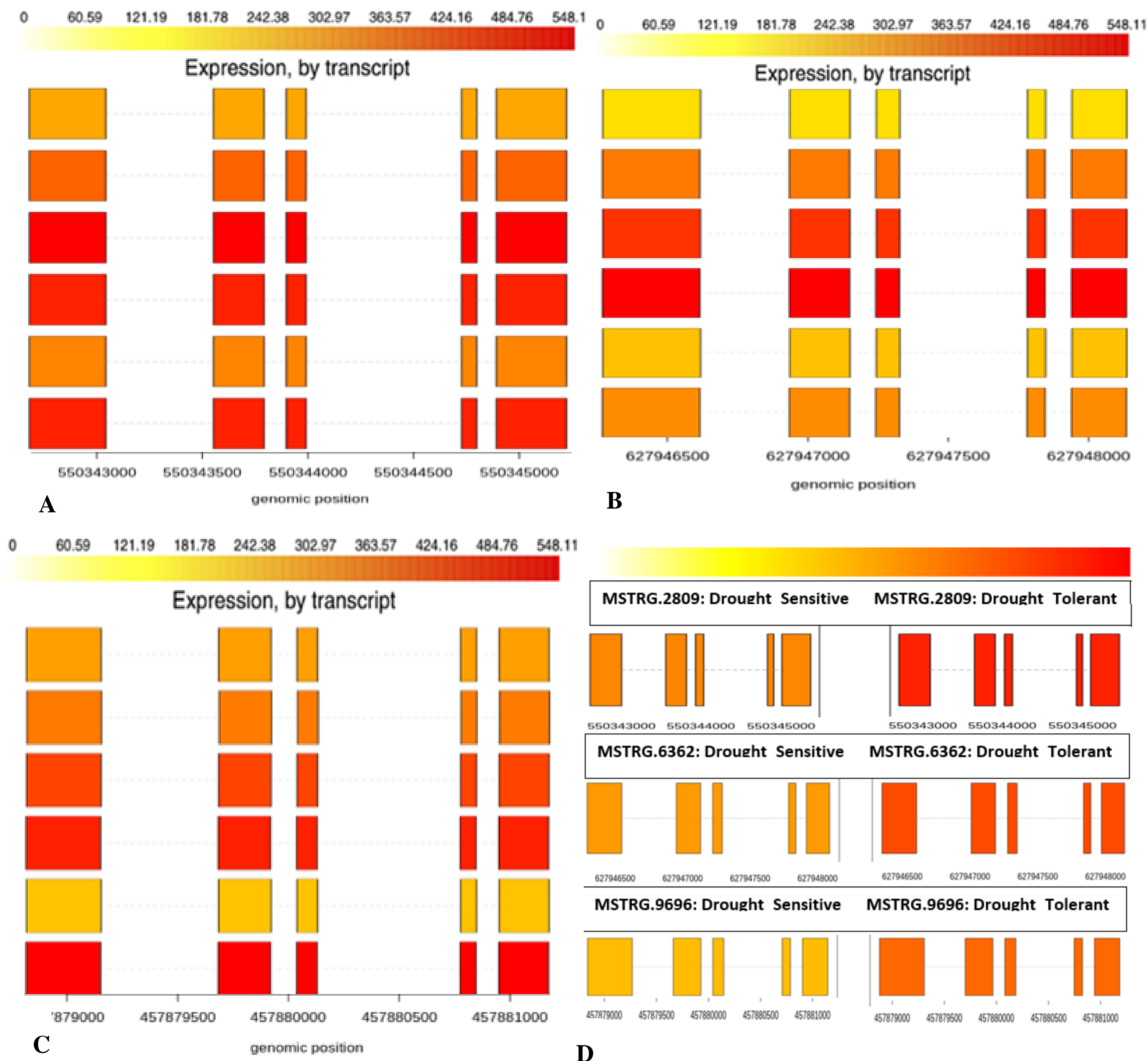

Fig. 3.3: Structure and expression levels of six distinct isoforms of the ARD4 like gene (A) genome A (B) genome B(C) genome C (D) Average expressions of all transcripts for A, B and D genome. Each plot of six wheat cultivars i.e., (top to bottom) Batis, Blue-Silver, Chakwal-50, Local-White, Syn-22, and UZ-11-CWA-8 show single transcript on each row, colored by its FPKM level. Darker the colour of transcript on map higher will be the expression.

3.5. Structure and expression of distinct DRO1-like gene isoforms across all samples

DRO1 is negatively controlled by auxin and is engaged in root tip cell elongation causing asymmetric root development and root downward bending in reaction to gravity. Higher DRO1 expression raises the angle of root development, with roots growing more downward (Uga *et al*., 2013) hence, make DRO1 a significant gene in root architecture. Recent studies in *Oryza sativa* (rice) recognized a function for DEEPER ROOTING 1 (DRO1) in affecting the root system orientation, leading to improvements in grain yields under water-limited circumstances (Guseman, Webb, Srinivasan, and Dardick, 2017). According to ballgown object, 111036th transcript in showed that TraesCS5A02G213300 is isoform of the Dro1-like gene, which is more pronounced in Blue-Silver,Chakwal-50th than in Local-White,UZ-11-CWA-8 and less expressed in Batis than in Syn-22 in genome A shown in Fig 3.4 (A). Similarly, 120789th transcript i.e., TraesCS5B02G210500 Dro1-like gene's isoform known to be more expressed in Batis than in Chakwal-50, Local-White, Blue-Silver and UZ-11-CWA-8 and not expressed in Syn-22 in genome B shown in Fig 3.4 (B).Where as, 130823th transcript i.e., TraesCS5D02G218700 is more expressed in Batis, Blue-Silver, Local-White and UZ-11-CWA-8 than in Chakwal-50 and Syn-22 in D shown in Fig 3.4 (C).

FPKM distributions in DRO1 like gene for transcript TraesCS5A02G213300.1, TraesCS5B02G210500.1 and TraesCS5D02G218700.1 is highly expressed in drought sensitive. For all transcripts in Dro1 such as genome A, average expression indicates that a drought-tolerant organization has minimal transcript expression. Whereas, genome B and D shows that varieties that are susceptible to drought have elevated concentrations of transcript expression shown in Fig 3.4 (D).

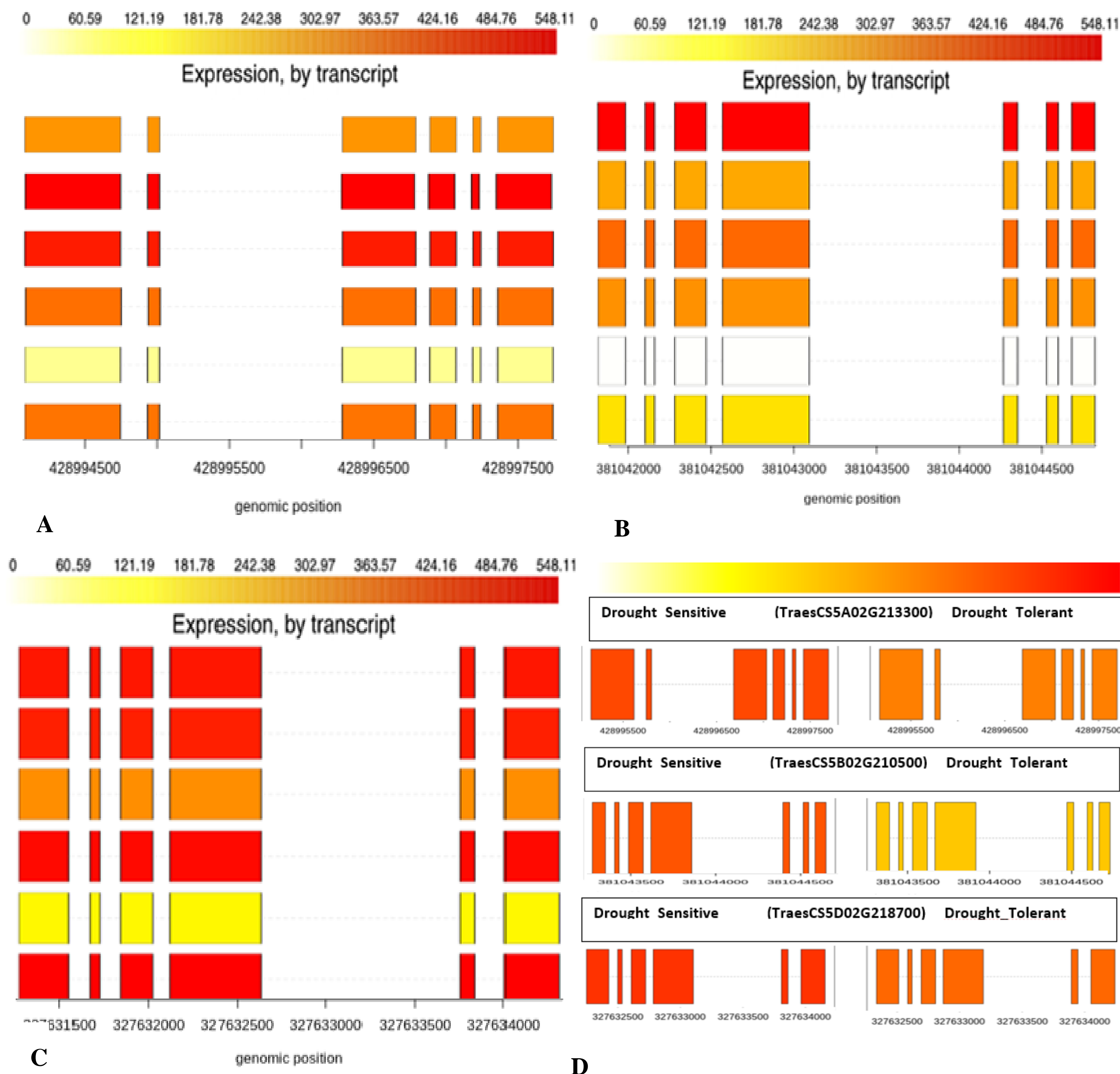

Fig. 3.4: Structure and expression levels of six distinct isoforms of the DRO1 like gene (A) genome A (B) genome B(C) genome C (D) Average expressions of all transcripts for A, B and D genome. Each plot of six wheat cultivars i.e., (top to bottom) Batis, Blue-Silver, Chakwal-50, Local-White, Syn-22, and UZ-11-CWA-8 show single transcript on each row, colored by its FPKM level. Darker the colour of transcript on map higher will be the expression.

3.6. Structure and expression of distinct HKT1-like gene isoforms across all samples

HKT1 was initially categorized as the K+-H+ symporter that intermediated the high-affinity K+ acceptance in wheat roots (Schachtman and Schroeder, 1994), but it was later found to co-transport Na+-K+ when expressed in yeast (*Saccharomyces cerevisiae*) or Xenopus oocytes (Rubio and Rodensj 1995). Even though this co-transport was least shown to activate in the roots of any cereal (Hayes, Smith, and Walker, 2001) , the concept that HKT1 is a high-affinity root K+ transporter still persists (Horie and Schroeder, 2004). According to 146418th transcript i.e., TraesCS6B02G182600.1 is an isoform of the HTK1-like gene is more expressed in Batis, and Local-White than in Blue-Silver, Chakwal-50,Syn-22,UZ-11-CWA-8 in Genome B shown in Fig 3.5 (B). Similarly, 154034th transcript i.e., TraesCS6D02G144500.1 is more pronounced in Local-White than in Chakwal-50, Batis,UZ-11-CWA-8,Blue-Silver and not expressed in Syn-22 Genome D shown in Fig 3.5 (C). Correspondingly, 165681th i.e., TraesCS7A02G418200.1 is not expressed in any of the genome A varieties shown in Fig 3.5 (A).

FPKM distributions for HKT1 like gene that is known to be highly expressed in drought sensitive. Average expression for all HKT1 transcripts, suggests no transcript expression for genome A whereas genome B transcripts suggests that variety groups sensitive to drought have minimal transcript expression, furthermore genome D shows that varieties that are susceptible to drought have elevated concentrations of transcript expression shown in Fig 3.5 (D).

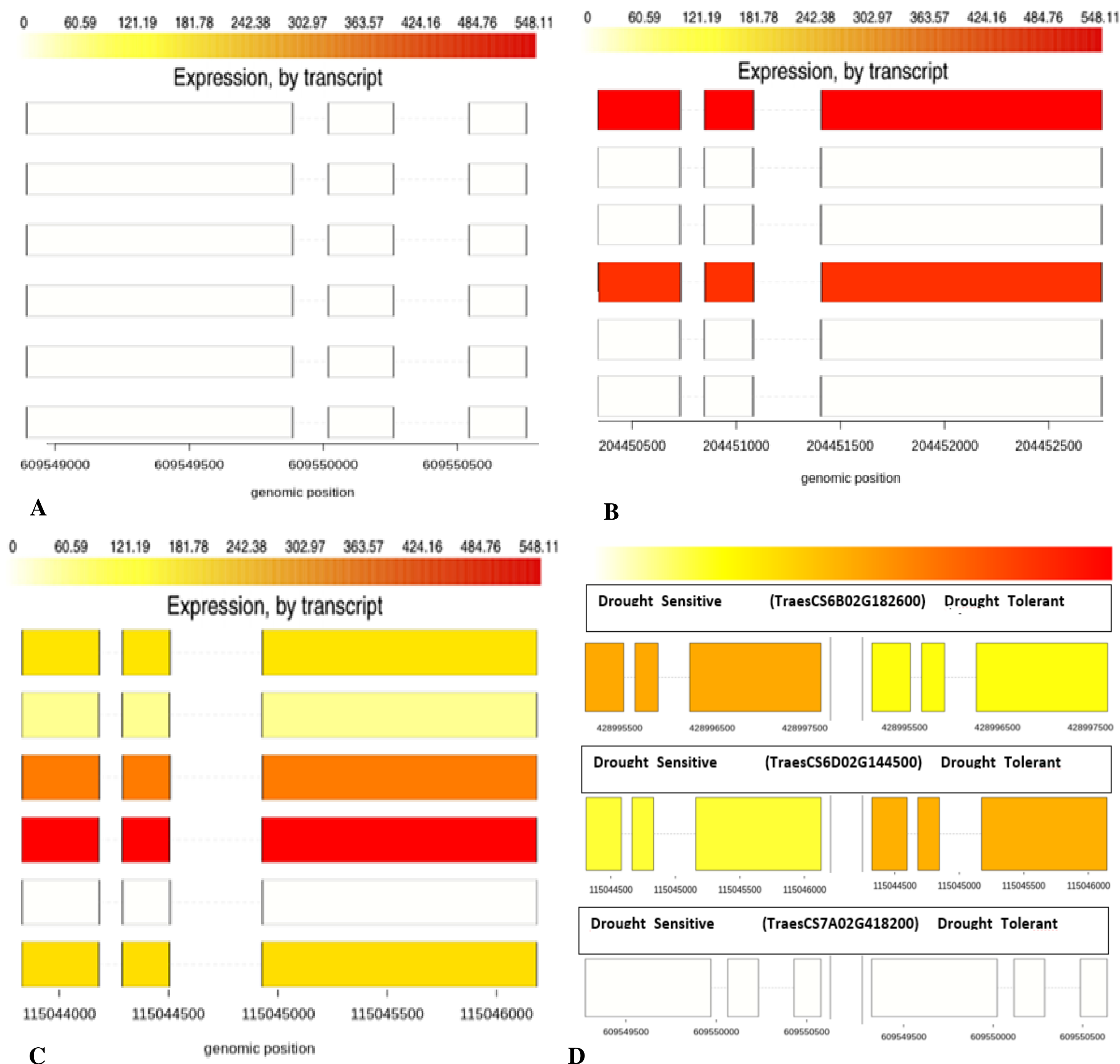

Fig. 3.5: Structure and expression levels of six distinct isoforms of the HKT1 like gene (A) genome A (B) genome B(C) genome C (D) Average expressions of all transcripts for A, B and D genome. Each plot of six wheat cultivars i.e., (top to bottom) Batis, Blue-Silver, Chakwal-50, Local-White, Syn-22, and UZ-11-CWA-8 show single transcript on each row, colored by its FPKM level. Darker the colour of transcript on map higher will be the expression.

3.7. Structure and expression of distinct PIN2-like gene isoforms across all samples

PIN2 is a major component forbasipetal auxin transport, being involved in root gravitropism. PIN genes actively control the distribution of auxins in order to regulate cell division and primary root (Blilou *et al*., 2005). Hence the gene has direct participation in root development and growth, therefore, it can be considered as vital gene involved in root architecture along with other gene in network. According to the I d column in the table, 98679th transcript in the Ballgown object reveals that TraesCS4B02G318200 has 3 transcripts and is an isoform of the PIN2 gene, the first transcript does not appear in Genome B for any variety, but the second and third transcripts show variable expressions. In contrast, the second transcript is highly expressed in Blue-Silver, and transcript displays the same amount of expressions in the remaining variants. Similarly, in contrast, the third transcript is highly expressed in Batis, Blue-Silver and Chakwal-50, and in the remainder of the variants, transcript displays equal expressions shown in Fig 3.6 (B). Similarly, 105554th transcript i.e., TraesCS4D02G314500 has two transcripts isoform of the gene PIN2. For all varieties, the first transcript has minimal appearance in Genome D, second transcript is highly expressed in Batis, Blue-Silver, Chakwal-50 and Syn-22 and is less expressed in UZ-11-CWA-8 than in Local-White shown in Fig 3.6 (C). Likewise, 115339th transcript i.e., TraesCS5A02G494100 has 3 transcripts and is a PIN2-like gene isoform, the first transcript does not appear in Genome A for any variety where the second transcript has a maximum level of expression and the third transcript has a minimum level of expression for all varieties shown in Fig 3.6 (A).

FPKM distributions for transcript TraesCS5A02G494100, TraesCS4B02G318200 and TraesCS4D02G314500 from PIN2 like gene that is known to be more highly expressed in drought sensitive. The average expression for all three PIN2 transcripts such as genome A, B and genome D indicates that a drought-tolerant organisation has the same transcript expression as a drought-sensitive variety organisation shown in Fig 3.5 (D).

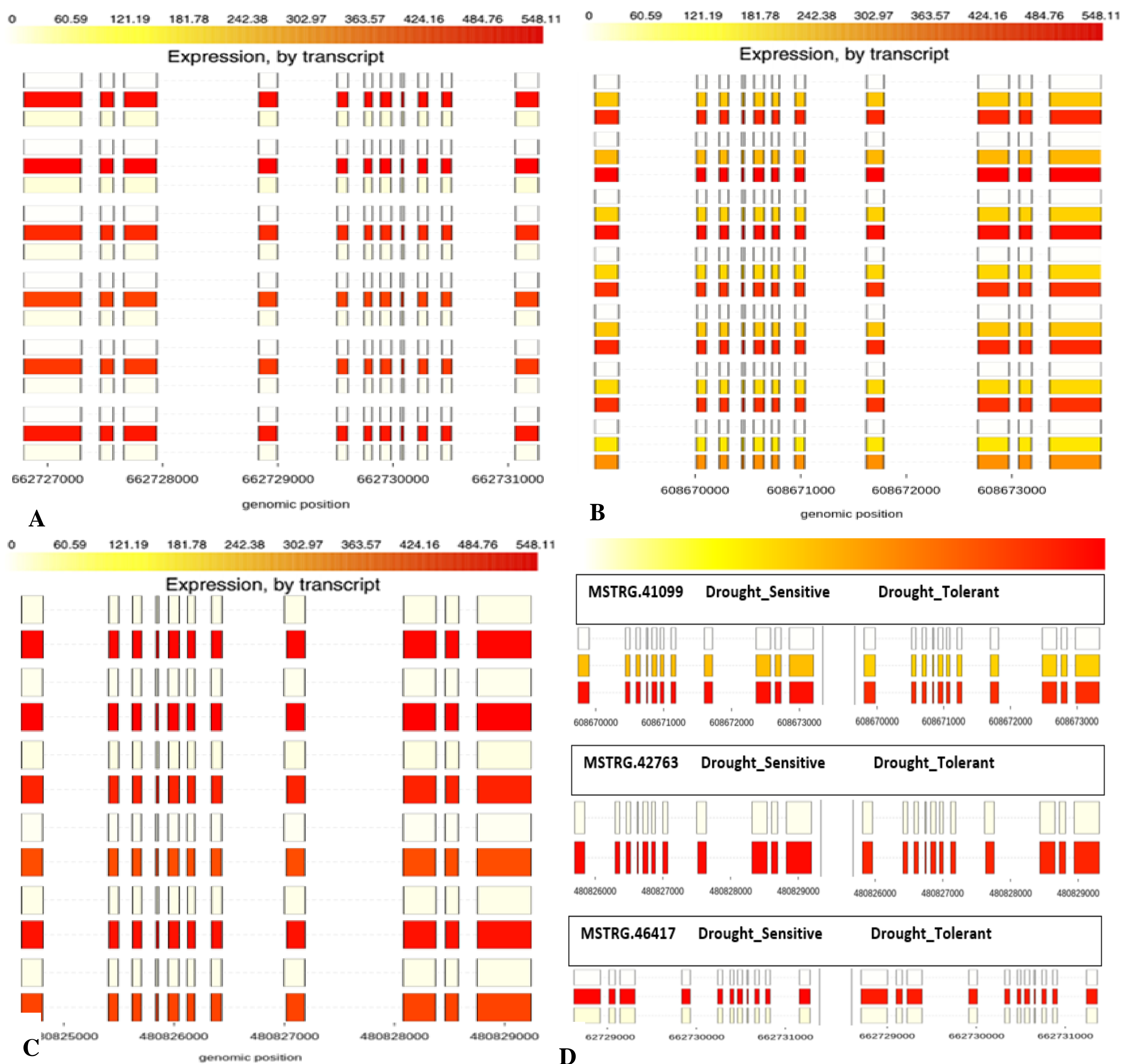

Fig. 3.6: Structure and expression levels of six distinct isoforms of the PIN2 like gene (A) genome A (B) genome B(C) genome C (D) Average expressions of all transcripts for A, B and D genome. Each plot of six wheat cultivars i.e., (top to bottom) Batis, Blue-Silver, Chakwal-50, Local-White, Syn-22, and UZ-11-CWA-8 show single transcript on each row, colored by its FPKM level. Darker the colour of transcript on map higher will be the expression.

3.8. Structure and expression of distinct PSTOL-like gene isoforms across all samples

Rice strains with this genomic introgression carrying the PSTOL gene display enhanced biomass, improved root growth, improved tiller number and yield increases of up to 30% when grown under low P limits, whereas no adverse effects were seen when grown under normal circumstances of soil fertility (Milner *et al*., 2018). The 520th transcript in the Ballgown object i.e., TraesCS1A02G018000.1 is a PSTOL-like gene that is extremely expressed in Chakwal-50 and less expressed in Batis. The transcript indicates no expression at Genome A for the remaining variants shown in Fig 3.7 (A). Likewise, 8188th transcript i.e., TraesCS1B02G022400.1 is extremely expressed in Local-White and indicates lowest variable expression in Blue-Silver, Local-White, UZ-11-CWA-8 and no expression in Batis and Syn-22 in B as shown in Fig 3.7 (B). In addition, 17221th transcript i.e., TraesCS1D02G017800.1 is highly expressed in Local-White and demonstrates the lowest variable expression in Chakwal-50, Batis, Blue-Silver and no expression in Syn-22 and UZ-11-CWA-8 in D shown in Fig 3.7 (C).

FPKM distributions in both varieties for transcript TraesCS1A02G018000.1, from PSTOL like gene that is known to be more highly expressed in drought sensitive, but TraesCS1B02G022400.1 and TraesCS1D02G017800.1 are more expressed in drought tolerant. The average expression for all PSTOL transcripts, such as genes A, B and D, shows both organisations, has the same transcript expression shown in Fig 3.7 (D).

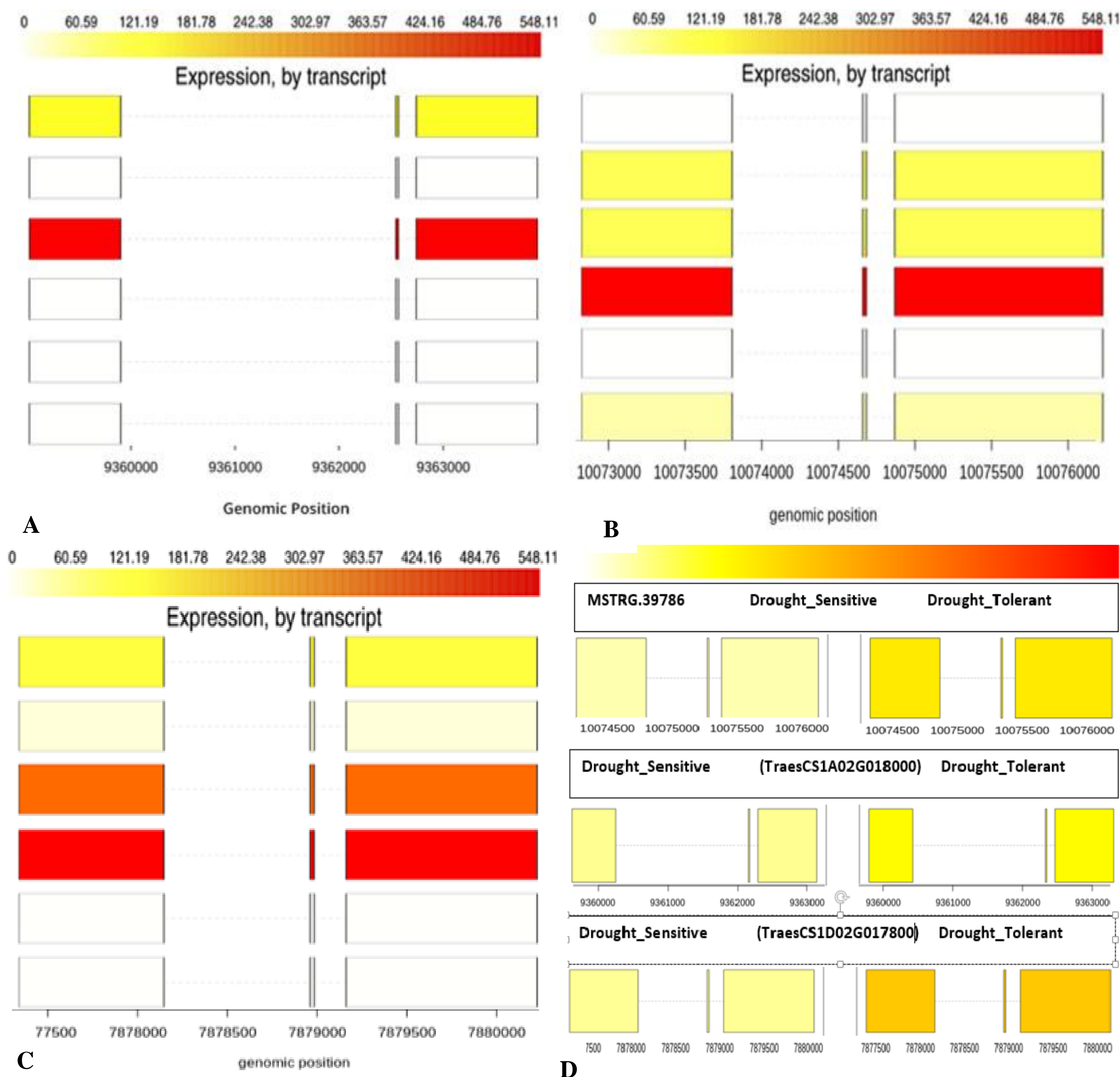

Fig. 3.7: Structure and expression levels of six distinct isoforms of the PSTOL like gene (A) genome A (B) genome B(C) genome C (D) Average expressions of all transcripts for A, B and D genome. Each plot of six wheat cultivars i.e., (top to bottom) Batis, Blue-Silver, Chakwal-50, Local-White, Syn-22, and UZ-11-CWA-8 show single transcript on each row, colored by its FPKM level. Darker the colour of transcript on map higher will be the expression.

3.9. Structure and expression of distinct RPK1-like gene isoforms across all samples

The gene RPK1 is expressed flowers, stems, leaves and roots. Some environmental stresses such as dehydration high salt, and low temperature also instantly induce the gene, implying that the gene is involved in a general response to stress (Hong, Jon, Kwak, and Nam, 1997). The 27566 th transcript i.e., TraesCS2A02G176500.1 has two transcripts and is an RPK1 gene isoform. The first transcript has minimal expression in Blue-Silver and Chakwal-50 and no expression in any of Genome A's other variants, whereas the second transcript is extremely expressed in Blue-Silver, Chakwal-50 and UZ-11-CWA-8 and less expressed in Batis and Syn-22 than in Local-White shown in Fig 3.8 (A). Similarly, 39614th transcript i.e., TraesCS2B02G281400.1 has 2 transcripts and is an isoform of the RPK1 gene, the first transcript does not appear in Genome B for any variety where the second transcript has a maximum level of expression for UZ-11-CWA-8 than in Batis, Blue-Silver, Chakwal-50, Local-White and the least expressed in Syn-22 shown in Fig 3.8 (B) Whereas, 48176th transcript i.e., TraesCS2D02G183900.1 has 2 transcripts and is an RPK1-like gene isoform, the first transcript does not appear in Genome D for any variation whereas the second transcript has the highest amount of expression for Blue-Silver, Chakwal-50, Local-White and UZ-11-CWA-8 and the lowest level of expression for Batis and Syn-22 shown in Fig 3.8 (C).

FPKM distributions in both varieties for transcript TraesCS2A02G176500.1from RPK1 like gene that is known to be more highly expressed in drought sensitive where TraesCS2B02G281400.1 and TraesCS2D02G183900.1 has no expression. Average expression for two RPK1 transcripts such as genome A indicates the same transcript expression for both groups, whereas average expression Genome B and D for RPK1 as the for genome B and D of two transcripts indicates the highest amount of expression for organization tolerant of drought shown in Fig 3.8 (D).

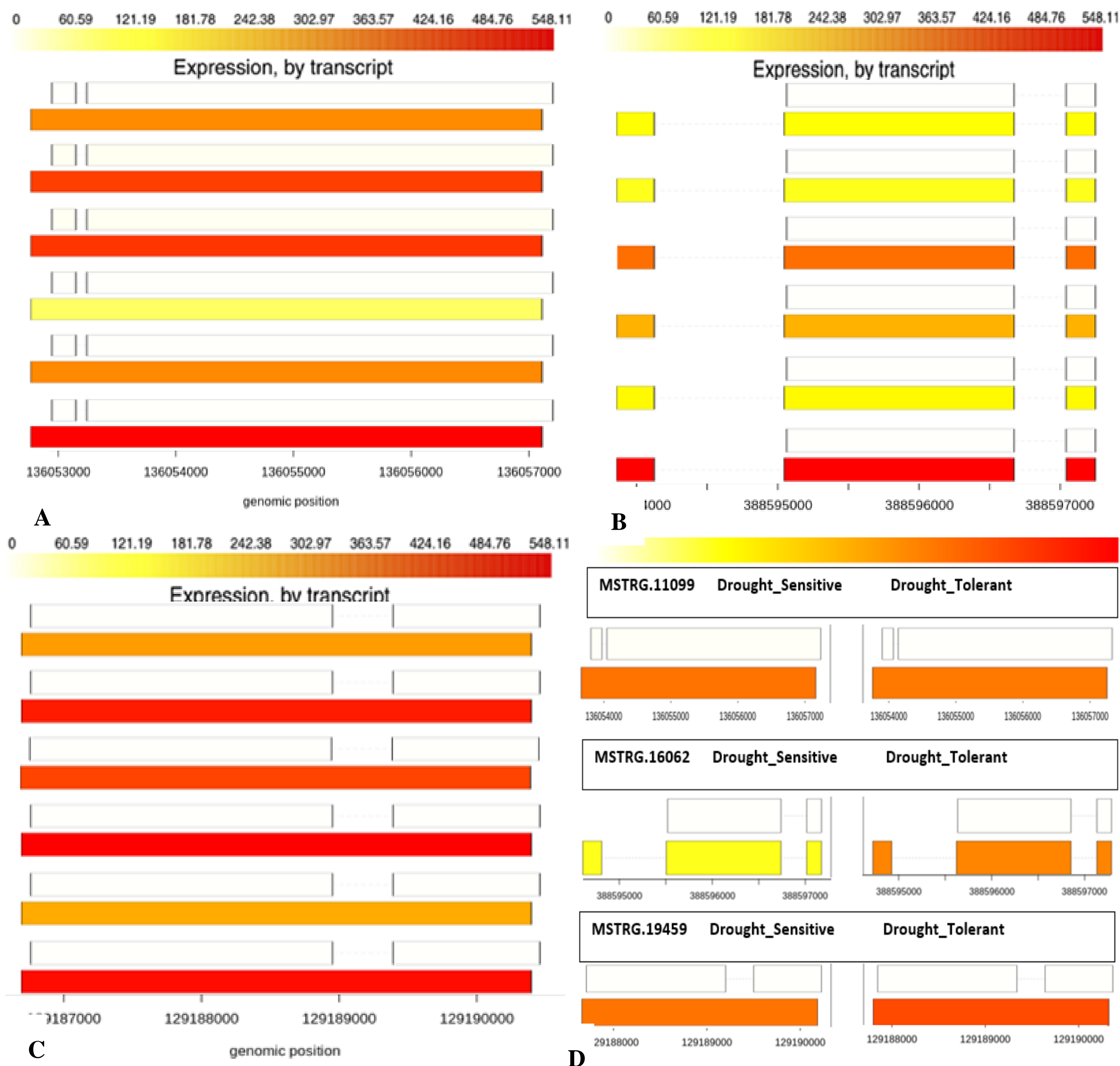

Fig. 3.8: Structure and expression levels of six distinct isoforms of the RPK1 like gene (A) genome A (B) genome B(C) genome C (D) Average expressions of all transcripts for A, B and D genome. Each plot of six wheat cultivars i.e., (top to bottom) Batis, Blue-Silver, Chakwal-50, Local-White, Syn-22, and UZ-11-CWA-8 show single transcript on each row, colored by its FPKM level. Darker the colour of transcript on map higher will be the expression.

3.10. RNA-seq Alignments

Explicit assembly and quantification of transcripts is explicitly linked with the peculiarity of the spliced alignments. If a large percentage of the reads are not aligned, it will be difficult for the transcriptome assembler to reconstruct genes, particularly those expressed at low concentrations The alignment mode promoted in HISAT2 makes a small load of false positives which helps transcript assemblers drastically enhance its computational time and memory usage requirements. The findings show the cost of aligning the entire information set by aligning all reads with the original data sets to the entire genome. Although > 95% of the reads are generally aligned, a large percentage of them are mapped in more than one area. Multi-mapped reads can have an acute impact on the calculated gene expression levels and will be vigilant in deducting differentially expressed genes that integrate an excessive percentage of multimedia reads (Robert and Watson, 2015).

Table 3.4. Read-mapping statistics on the full data set

Cultivars	Total reads	Uniquely aligned reads(%)	Multimapped reads (%)	Overall alignment Rate (%)	Q20(%)	Q30(%)	GC Content (%)
Batis	47172834	75.58	7.14	92.95	99.2	94.7	48.5
Blue Silver	46664681	74.62	7.10	92.16	98.6	95.6	47.5
Chakwal 50	46617180	75.95	6.54	92.92	89.2	95	48
Local White	50646812	77.16	6.65	93.95	98.8	92	49.5
Syn-22	48779601	76.67	7.51	94.43	89.3	93.7	49.5
UZ-11-CWA-8	49806318	75.50	6.01	92.49	98.1	94.6	49

3.11. Transcriptome Assembly

Gffcompare tool helps to collect summary statistics that describe how assembled transcripts assess the annotation as shown in Table3.5. The number of annotated genes that could be expressed differs across the samples, showing the gene expression's tissue-unique nature. The gene and transcript models are merged after the assembly of all the samples. When the annotation record is used to merge, the output will include all transcripts in the annotation report, including those with 0 levels of expression, as well as all novel transcripts that ultimately help StringTie accurately collect transcripts, mainly those expressed at low expressions. When annotation is given, the range of transcripts matching the annotation rises dramatically. The results also demonstrate that there are almost as many novel transcripts (isoforms) in each sample during these documents as recognized transcripts.

Table 3.5. Transcriptome assembly statistics

Cultivars	Number of assembled genes	Novel genes	Transcripts matching annotation	Novel Transcripts
Batis	234418	113674	57077	48599
Blue Silver	244662	123918	56033	48599
Chakwal 50	234863	114119	545010	48599
Local White	235179	114434	61040	69274
Syn 22	229223	108479	58190	48599
UZ-11-CWA-8	234362	113618	54311	50491

3.12. Gene Ontology

The gene ontology (GO) terms of DEGs have been retrieved using the agriGO platfor m, which has served the scientific community for more than 10 years, focusing specifi cally on plant and agricultural species enrichment analysis of gene ontology (GO) (Tian *et al*., 2017).

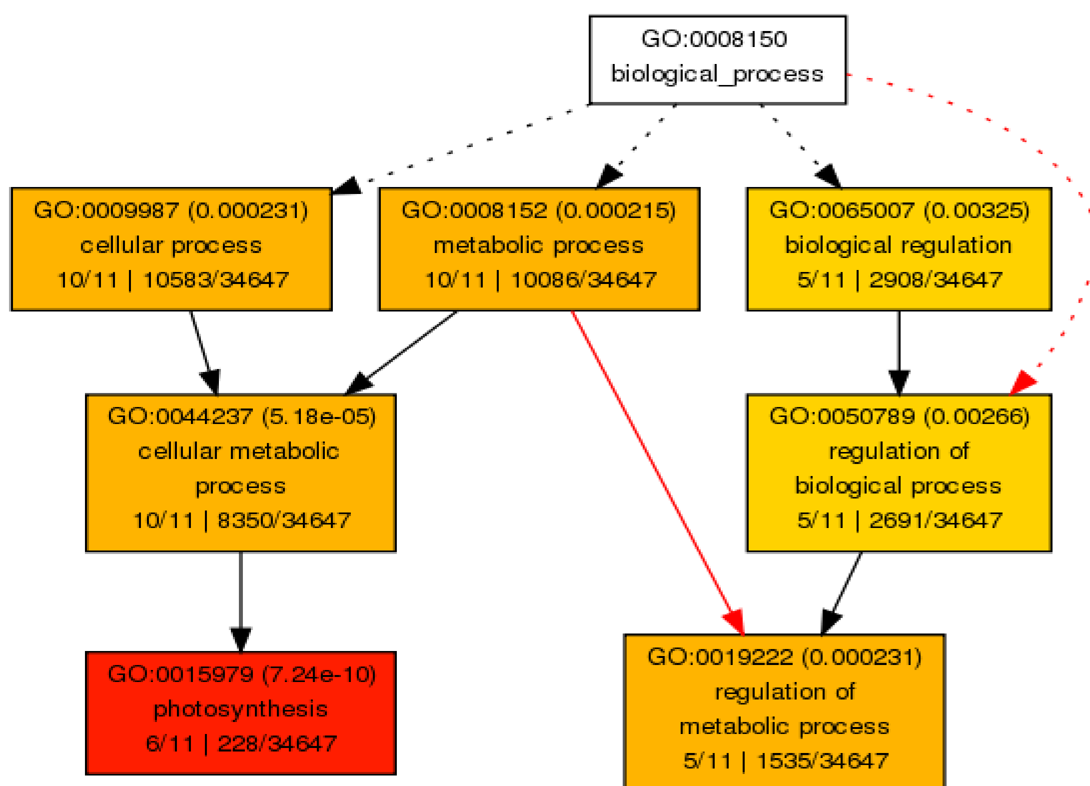

Fig.3.9: Gene Ontology Results showing Biological Analysis based on Singular Enrichment Analysis

Singular enrichment analysis was performed in AgriGO to identify gene ontologies. Each box shows the GO term number, the p-value in parenthesis, and GO term. First pair of numerals represents number of genes in input list associated with that GO term and number of genes in input list. Box colors indicates levels of statistical significance: yellow=0.05; orange=e-05; and red=e-09. Level 1 p = 0.05, Level 2 p = 5 × 10−3, Level 3 p = 5 × 10−4, Level 4 p = 5 × 10−5, Level 5 p = 5 × 10−6, Level 6 p = 5 × 10−7, Level 7 p = 5 × 10−8, Level 8 p = 5 × 10−9, Level 9 p = 5 × 10−10. Solid, dashed, and dotted lines represent two, one and zero enriched terms at both ends connected by the line, respectively.

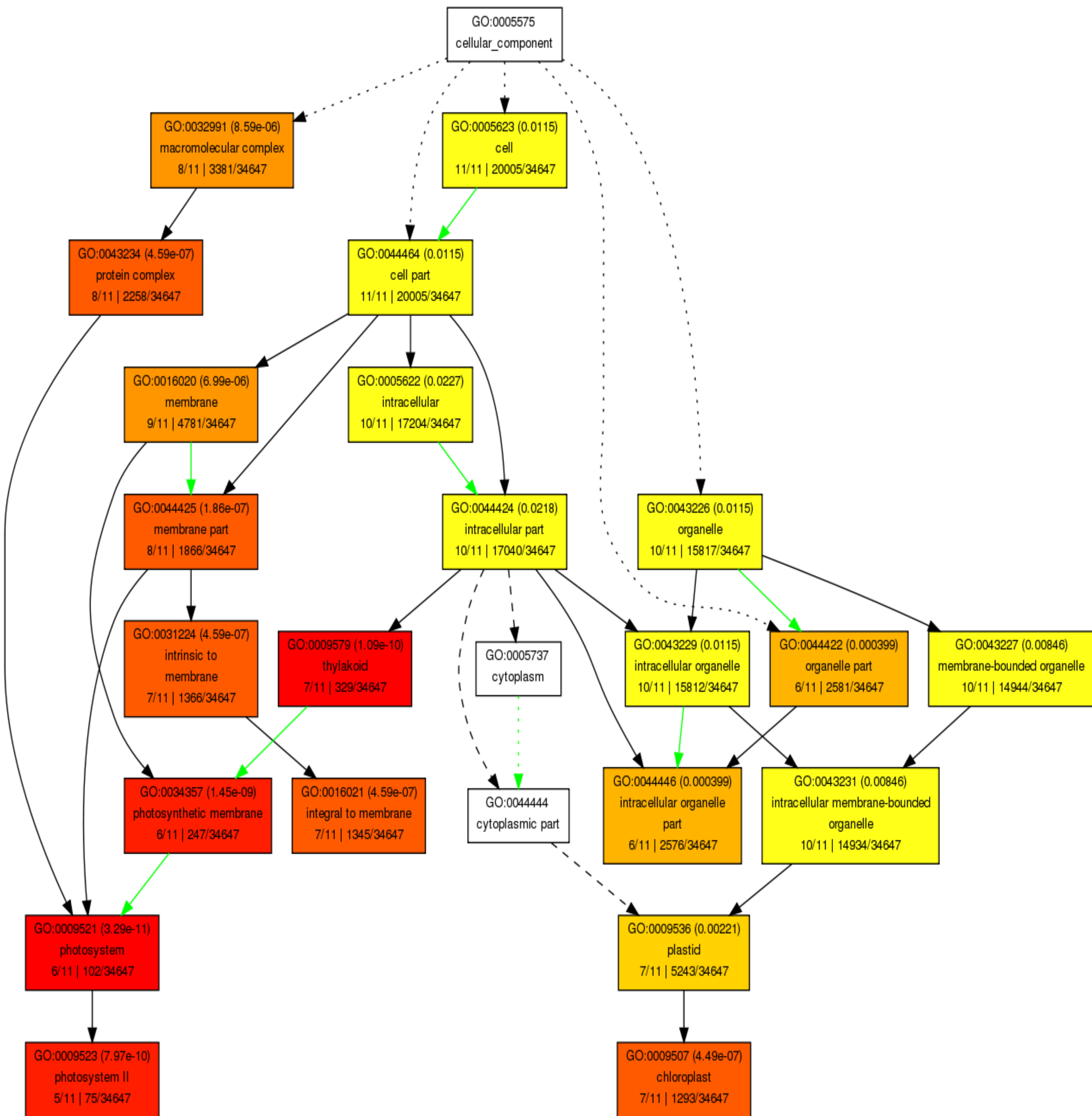

Fig.3.10: Gene Ontology Results showing Cellular Analysis based on Singular Enrichment Analysis

DISCUSSION

Drought and metrological pressure are major disasters that affect the productivity of plants in both rain-fed and irrigated areas. These factors restrict the yield potential mainly because of the lack of root architecture that can withstand these stresses. Roots provide necessary soil water and nutrient uptake and anchorage and stability for the entire plant. Root architecture greatly affects the development of agriculture and crop production (Kong ., 2014). This research was therefore performed to distinguish the wheat transcriptome of drought tolerant and other wheat cultivars. RNA-seq was conducted and analyzed using various tools of bioinformatics.

Our HISAT2 NGS alignment system analysis reads against genome population produced maximum alignment frequency for Batis 92.95%, Blue Silver 92.16%, Chakwal 50 92.92%, Local White 93.95%, and Syn-22 94.43% and UZ-11-CWA-8 92.49%, respectively. To encapsulate, Ballgown computing program for differentially expressed genes in RNA-seq studies, using the output of StringTie, shows that expression for differential genes in drought is more sensitive than in varieties tolerant to drought. This study focused primarily on the quest for genes involved in root structure, for their genomic location and expression in wheat homologous genomes (A, B, and D). Studies showed that root architecture and anatomical traits appear instead to be quantitative and controlled by many genes (Price and Tomos 1997, Mano 2007, Topp 2013, Burton 2015, Zurek 2015). Regulation of auxin distribution by PIN transporters is key in the dynamic modulation of root growth and branching (Kochian ., 2005). PSTOL1, which is associated with enhanced root growth and increased grain yield on phosphorous deficient soils. The PSTOL1 act as an enhancer of early root growth and promotes more phosphorus uptake (Gamuyao ., 2012). In Arabidopsis, genetic and biochemical studies have revealed the essential roles of the members of PIN, AUX1/LAX (AUXIN-INSENSITIVE1/LIKE AUX1), and ABCB (B subfamily of ABC transporters) families in mediating polar auxin transport and root gravitropism (Geisler *et al.*, 2014). The average expression for all three PIN2 and PSTOL transcripts for genome A, B and D along with RPK1 transcripts for genome A shows that a

drought-tolerant organization, and drought-sensitive variety organization, has the same transcript expression whereas average expression Genome B and D for RPK1 indicates the highest amount of expression for organization tolerant of drought. According to a research, RPK1 overproduction in these transgenic plants had increased their tolerance to drought stress receptor-like kinases TOAD2/RPK2 and RPK1 regulate root growth by controlling cell proliferation and affecting meristem size (Sentenac ., 1992; Hirsch ., 1998). Similarly, average expression for all ARD4 transcripts for genome A, B and D indicates that a drought-tolerant organization has highest transcript expression. Currently, the only genes reported to control a root architecture in crop species were found in rice: deeper rooting 1 ( ) (Uga . 2013) and phosphorus starvation tolerance 1 ( ) (Gamuyao . 2012). Overexpression of DRO1 in led to deeper-rooting phenotypes. Collectively, gene networks bear potential application for DRO1-related genes to alter root architecture for drought avoidance (Guseman 2017). For all transcripts in Dro1, genome A, average expression indicates that a drought-tolerant organization has minimal transcript expression but genome B and D shows that varieties that are susceptible to drought have elevated concentrations of transcript expression. HKT1 is expressed in root cortical cells and in cells adjacent to the vascular tissue in leaves, indicative of a role in K+ transport (Schachtman and Schroeder, 1994). Average expression for all HKT1 transcripts, suggests no transcript expression for genome A, whereas genome B transcripts suggests that drought sensitive group have minimal transcript expression, furthermore genome D showed that varieties susceptible to drought have elevated concentrations of transcript expression.

The transcriptomics studies have advanced considerably because of the explosive growth in the underlying sequencing technology (Abdel-Ghany ., 2016; Wang ., 2016). Transcriptomics offers important insights on gene structure, expression, and regulation and has been widely studied in many organisms (Jain, 2012; Casamassimi ., 2017; Lowe ., 2017). Such a study will have implications not just in wheat, but also more broadly in polyploid plant genomics.

In this review, we found that the characteristics of transcriptomic sequence variations in different cultivars and genes showed mutations in drought-related genes that can contribute

directly to drought tolerance. Additional analysis of this data includes alternate splicing, variant identification and allele-specific expression, pathway analysis, co-expression network analysis, in order to have a complete picture of differences in root growth and used in genetic improvement programs through CRISPR / Cas9 system.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome biology*, 11(10), R106.

Anderson, S. (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic acids research*, 9(13), 3015-3027.

Anjum, S. A., Xie, X. Y., Wang, L. C., Saleem, M. F., Man, C., and Lei, W. (2011). Morphological, physiological and biochemical responses of plants to drought stress. *African Journal of Agricultural Research*, 6(9), 2026-2032.

Athar, H. R., and Ashraf, M. (2009). Strategies for crop improvement against salinity and drought stress: An overview. In Salinity and water stress. *Springer*, 1-16.

Bailey, L. H. (1976). *Hortus Third: A concise dictionary of plant cultivated in US and Canada* (No. 582.03/B155). Mac'milan publishing Co. Inc.

Bailey, P. H., Currey, J. D., and Fitter, A. H. (2002). The role of root system architecture and root hairs in promoting anchorage against uprooting forces in Allium cepa and root mutants of Arabidopsis thaliana. *Journal of Experimental Botany*, *53*(367), 333-340.

Ball, P. (2013). DNA: Celebrate the unknowns. *Nature*, *496*(7446), 419.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823-837.

Belderok, B., Mesdag, J., and Donner, D. A. (2000). *Bread-making quality of wheat: a century of breeding in Europe*. Springer Science and Business Media.

Berntson, G. M. (1994). Modelling root architecture: are there tradeoffs between efficiency and potential of resource acquisition? *New Phytologist*, *127*(3), 483-493.

Boyer, J. S. (1982). Plant productivity and environment. *Science*, *218*(4571), 443-448.

Blilou, I., Xu, J., Wildwater, M., Willemsen, V., Paponov, I., Friml, J., Scheres, B. (2005). The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. Nature, 433(7021), 39–44.

Breiman, A. G. D. (1995). Wheat evaluation. *Israel J Plant Sci*, *43*, 58-95.

Cannon, W. A. (1949). A tentative classification of root systems. *Ecology*, *30*(4), 542-548.

Coram, T. E., Brown-Guedira, G., and Chen, X. (2008a). Using transcriptomics to understand the wheat genome. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, *3*(83), 1-9.

Coram, T. E., Wang, M., and Chen, X. (2008b). Transcriptome analysis of the wheat–Puccinia striiformis f. sp. tritici interaction. *Molecular Plant Pathology*, *9*(2), 157-169.

Chen, X., Shi, J., Hao, X., Liu, H., Shi, J., Wu, Y., Mao, C. (2013). OsORC3 is required for lateral root development in rice. *The Plant Journal*, *74*(2), 339–350

Curtis, B., Rajaraman, S., and Macpherson, H. (2002). Bread wheat. Food and Agriculture Organization of the United Nations. *FAO Plant production and Protection series*, 30.

Edwards, D., Batley, J., and Snowdon, R. J. (2013). Accessing complex crop genomes with next-generation sequencing. *Theoretical and Applied Genetics*, 126(1), 1-11.

Ergen, N. Z., and Budak, H. (2009). Sequencing over 13 000 expressed sequence tags from six subtractive cDNA libraries of wild and modern wheats following slow drought stress. *Plant, cell and environment*, *32*(3), 220-236.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., ... and Merrick, J. M. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*, *269*(5223), 496-512.

Fleury, D., Jefferies, S., Kuchel, H., and Langridge, P. (2010). Genetic and genomic tools to improve drought tolerance in wheat. *Journal of experimental botany*, *61*(12), 3211-3222.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., ... and Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *Science*, *327*(5967), 812-818.

Gregersen, P. L., Brinch-Pedersen, H., and Holm, P. B. (2005). A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. *Transgenic research*, *14*(6), 887-905.

Grossman, J. D., and Rice, K. J. (2012). Evolution of root plasticity responses to variation in soil nutrient distribution and concentration. *Evolutionary Applications*, *5*(8), 850-857.

Guseman, J. M., Webb, K., Srinivasan, C., and Dardick, C. (2017). DRO1 influences root system architecture in Arabidopsis and Prunus species. *The Plant Journal*, *89*(6), 1093–1105.

Habash, D. Z., Kehel, Z., and Nachit, M. (2009). Genomic approaches for designing durum wheat ready for climate change with a focus on drought. *Journal of Experimental Botany*, *60*(10), 2805-2815.

Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T., Park, H. S., and Soeda, E. (2000). The DNA sequence of human chromosome 21. *Nature*, *405*(6784), 311.

Hayes, D. E., Smith, F. A., and Walker, N. A. (2001). High-affinity potassium transport into wheat roots involves sodium—A role for HKT1? *Functional Plant Biology*, *28*(7), 645–654.

Hong, S. W., Jon, J. H., Kwak, J. M., and Nam, H. G. (1997). Identification of a receptor-like protein kinase gene rapidly induced by abscisic acid, dehydration, high salt, and cold treatments in Arabidopsis thaliana. *Plant Physiology*, *113*(4), 1203–1212.

Horie, T., and Schroeder, J. I. (2004). Sodium Transporters in Plants. Diverse Genes and Physiological Functions. *Plant Physiology*, *136*(1), 2457–2462.

Hunt, J. R., and Kirkegaard, J. A. (2012). Re-evaluating the contribution of summer fallow rain to wheat yield in southern Australia. *Crop and Pasture Science*, *62*(11), 915-929.

International Wheat Genome Sequencing Consortium (IWGSC), IWGSC RefSeq principal investigators: Appels, R., Eversole, K., Feuillet, C., Keller, B., Uauy, C. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science (New York, N.Y.)*, *361*(6403).

Jacob, F., and Monod, J. (1961). On the regulation of gene activity. In Cold Spring Harbor Symposia on Quantitative Biology. *Cold Spring Harbor Laboratory Press,* 26, 193-211.

Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316(5830), 1497-1502.

Kano, M., Inukai, Y., Kitano, H., and Yamauchi, A. (2011). Root plasticity as the key root trait for adaptation to various intensities of drought stress in rice. *Plant and Soil*, *342*(1-2), 117-128.

Kell, D. B. (2011). Breeding crop plants with deep roots: their role in sustainable carbon, nutrient and water sequestration. *Annals of Botany*, *108*(3), 407-418.

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360.

Krupp, M., Marquardt, J. U., Sahin, U., Galle, P. R., Castle, J., and Teufel, A. (2012). RNA-seq Atlas—a reference database for gene expression profiling in normal tissue by next-generation sequencing. *Bioinformatics*, *28*(8), 1184-1185.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, *10*(3), R25.

Laudencia-Chingcuanco, D., Ganeshan, S., You, F., Fowler, B., Chibbar, R., and Anderson, O. (2011). Genome-wide gene expression analysis supports a developmental model of low temperature tolerance gene regulation in wheat (Triticum aestivum L.). *BMC genomics*, *12*(1), 299.

Li, G., Jain, R., Chern, M., Pham, N. T., Martin, J. A., Wei, T., and Jiang, L. (2017). The sequences of 1,504 mutants in the model rice variety Kitaake facilitate rapid functional genomic studies. *The Plant Cell*, tpc-00154.

Lynch, J. P., and Wojciechowski, T. (2015). Opportunities and challenges in the subsoil: pathways to deeper rooted crops. *Journal of Experimental Botany*, *66*(8), 2199-2210.

Meister, R., Rajani, M. S., Ruzicka, D., and Schachtman, D. P. (2014). Challenges of modifying root traits in crops for agriculture. *Trends in Plant Science*, *19*(12), 779-788.

Michael, T. P., and Jackson, S. (2013). The first 50 plant genomes. *The Plant Genome*, *6*(2).

Milner, M. J., Howells, R. M., Craze, M., Bowden, S., Graham, N., and Wallington, E. J. (2018). A PSTOL-like gene, TaPSTOL, controls a number of agronomically important traits in wheat. *BMC Plant Biology*, *18*.

Morozova, O., and Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, *92*(5), 255-264.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature methods*, *5*(7), 621.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature methods*, *5*(7), 621.

Narayanan, S., Mohan, A., Gill, K. S., and Prasad, P. V. (2014). Variability of root traits in spring wheat germplasm. *PLoS One*, *9*(6), e100317.

Nevo, E., Korol, A. B., Beiles, A., and Fahima, T. (2002). Genetic Variation in Agronomic Traits. In *Evolution of Wild Emmer and Wheat Improvement* (pp. 179-237). Springer, Berlin, Heidelberg.

Paulson, J. N., Stine, O. C., Bravo, H. C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, *10*(12), 1200–1202

Pennisi, E. (2008). The blue revolution, drop by drop, gene by gene. *Science*, *320*(5873), 171-173.

Ramsköld, D., Wang, E. T., Burge, C. B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS computational biology*, *5*(12), e1000598.

Raychaudhuri, S., Stuart, J. M., and Altman, R. B. (1999). Principal components analysis to summarize microarray experiments: application to sporulation time series. In *Biocomputing 2000* (pp. 455-466).

Robbins, N. E., and Dinneny, J. R. (2015). The divining root: moisture-driven responses of roots at the micro-and macro-scale. *Journal of experimental botany*, *66*(8), 2145-2154.

Robert, C., and Watson, M. (2015). Errors in RNA-seq quantification affect genes of relevance to human disease. *Genome Biology*, *16*(1), 177.

Rubio, C. A., and Rodensj, M. (1995). P53 overexpression in flat serrated adenomas and flat tubular adenomas of the colorectal mucosa. *Journal of Cancer Research and Clinical Oncology*, *121*(9–10), 571–576.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, *74*(12), 5463-5467.

Schachtman, D. P., and Schroeder, J. I. (1994). Structure and transport mechanism of a high-affinity potassium uptake transporter from higher plants. *Nature*, *370*(6491), 655–658.

Shewry, P. R. (2009). Wheat. *Journal of experimental botany*, *60*(6), 1537-1553.

Smyth, G. K., Ritchie, M., Thorne, N., and Wettenhall, J. (2005). LIMMA: linear models for microarray data. In Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Statistics for Biology and Health.

Tanno, K. I., and Willcox, G. (2006). How fast was wild wheat domesticated? *Science*, 311(5769), 1886- 1886.

Tian, H., De Smet, I., and Ding, Z. (2014). Shaping a root system: regulating lateral versus primary root growth. *Trends in plant science*, *19*(7), 426-431.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Su, Z. (2017). agriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*, *45*(Web Server issue), W122–W129

United Nations, Food and Agriculture Organization, Statistics Division (FAOSTAT). 2016. Retrieved 8 February 2018.

Usman, M. (2016). Contribution of agriculture sector in the GDP growth rate of Pakistan. *Journal of Global Economics*.

Uga, Y., Sugimoto, K., Ogawa, S., Rane, J., Ishitani, M., Hara, N., Yano, M. (2013). Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nature Genetics*, *45*(9), 1097–1102.

Varney, G. T., and Canny, M. J. (1993). Rates of water uptake into the mature root system of maize plants. *New Phytologist*, *123*(4), 775-786.

Varshney, R. K., Kayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in biotechnology*, *27*(9), 522-530.

Wang, C., Deng, P., Chen, L., Wang, X., Ma, H., Hu, W., and He, G. (2013). A wheat WRKY transcription factor TaWRKY10 confers tolerance to multiple abiotic stresses in transgenic tobacco. *Plops one*, *8*(6), e65120.

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

Wasson, A. P., Richards, R. A., Chatrath, R., Misra, S. C., Prasad, S. S., Rebetzke, G. J., ... and Watt, M. (2012). Traits and selection strategies to improve root systems and water uptake in water-limited wheat crops. *Journal of experimental botany*, *63*(9), 3485-3498.

Weinstein, J. N., Myers, T. G., O'connor, P. M., Friend, S. H., Fornace, A. J., Kohn, K. W., ... and Buolamwini, J. K. (1997). An information-intensive approach to the molecular pharmacology of cancer. *Science*, *275*(5298), 343-349.

Wilson, I. D., Barker, G. L., Beswick, R. W., Shepherd, S. K., Lu, C., Coghill, J. A., ... and Lenton, J. R. (2004). A transcriptomics resource for wheat functional genomics. *Plant Biotechnology Journal*, *2*(6), 495-506.

Wilson, I. D., Barker, G. L., Lu, C., Coghill, J. A., Beswick, R. W., Lenton, J. R., and Edwards, K. J. (2005). Alteration of the embryo transcriptome of hexaploid winter wheat (Triticum aestivum cv. Mercia) during maturation and germination. *Functional and integrative genomics*, *5*(3), 144-154.

APPENDICES

Thesis MPhil Hira Hameed

ORIGINALITY REPORT

10%	3%	5%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

#	Source	%
1	Submitted to Higher Education Commission Pakistan Student Paper	2%
2	Mihaela Pertea, Daehwan Kim, Geo M Pertea, Jeffrey T Leek, Steven L Salzberg. "Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown", Nature Protocols, 2016 Publication	2%
3	Submitted to Queen Mary and Westfield College Student Paper	1%
4	www.plantphysiol.org Internet Source	1%
5	Submitted to North West University Student Paper	1%
6	Berkman, P. J., K. Lai, M. T. Lorenc, and D. Edwards. "Next-generation sequencing applications for wheat crop improvement", American Journal of Botany, 2012. Publication	1%