2018

Transcriptome Analysis of Root Development in Wheat *Triticum Aestivum* Using High Throughtput Sequencing Technologies

Ghana Challa
South Dakota State University

TRANSCRIPTOME ANALYSIS OF ROOT DEVELOPMENT IN WHEAT (*TRITICUM AESTIVUM*) USING HIGH-THROUGHPUT SEQUENCING TECHNOLOGIES

BY

GHANA SHYAM CHALLA

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Biological Sciences

Specialization in Biology

South Dakota State University

2018

TRANSCRIPTOME ANALYSIS OF ROOT DEVELOPMENT IN WHEAT (*TRITICUM AESTIVUM*) USING HIGH-THROUGHPUT SEQUENCING TECHNOLOGIES.

GHANA SHYAM CHALLA

This dissertation is approved as a credible and independent investigation by a candidate for the Doctor of Philosophy degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this dissertation does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Wanlong Li, Ph.D. Date

Dissertation Advisor

Volker S. Brözel, Ph.D. Date

Head, Department of

Biology and Microbiology

Dean, Graduate School Date

Dedicated to my loving brother and my best friend Rakesh

ACKNOWLEDMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor and mentor Dr. Wanlong Li for his invaluable guidance and support throughout my time as a graduate student. Without his continuous motivation and encouragement, this work wouldn't have become a reality. I am most grateful to him for his patience during tough times, be it when I spent two months in NICU for our son or when I was diagnosed with fibromyalgia and depressed. Dr. Li is the best mentor I ever had, and his vast knowledge and experience in wheat genomics helped me to become a better researcher than I am.

I am also thankful to my committee members Dr. Senthil Subramanian, Dr. Jixiang Wu, and Dr. Yang Yen for their helpful advice and assistance during my graduate studies and supporting me in times of need. Without the ongoing support from all these individuals, I could not have finished this work.

I am grateful to all the faculty members, staff and students of the Department of Biology and Microbiology at South Dakota State University for their support and assistance.

My sincere thanks to my lab mates Dr. Zhengzhi Zhang, Dr. Huilan Zhu, Dr. Lei Hua, Dr. Jing Wang, Wei Wang, and Dilkaran Singh for their co-operation and help. I can never forget their collaboration and friendship.

A special thanks to the agencies who provided funding for my work. I thank Department of Biology and Microbiology, South Dakota State University; USDA-Agricultural Experimentation Station and SDSU Scholarly Award. I would also like to extend my sincere thank you to FGCF and its staff at SDSU. I cannot thank enough the staff at University Network Services and Security, especially Brian Moore and Kevin

Brandt who supported me throughout my program and helped me to acquire the computing resources needed for my work.

I express my thanks to all my friends in Brookings, Ramu, Deepthi, Prateek, Suresh, Aravind, Indrani, Shubham, Venkat, Jyothi, Sajag, Bhanu, Dr. Nizampatnam, and everyone who helped me directly or indirectly at different stages with their company or at times with their push to reach the goal. I also thank all my teachers who encouraged me and believed in me. I would like to extend my special thanks to Chandrasekhar Yadav who introduced me to the fun in scientific research and encouraged me at every step of my research career. I thank Dr. Kiran Sharma (ICRISAT, India) and his lab members for teaching me methods and many tips and tricks of plant transformation and tissue culture. I thank Dr. Mukesh Jain (NIPGR and JNU, India) who taught me the basics of plant genomics. I also thank all my friends and seniors at NIPGR for being there in times of need.

Words are not enough to express my gratitude towards my family. Without the support of my parents Prasad and Nagamani, my brother and my closest friend, Rakesh and my wife, Praveena, I wouldn't have reached this stage of my goal. It would be unjust to leave mentioning my uncle Bujji and his wife Latha for their support and encouragement throughout my life. Last but not the least, I wish to thank my son, Lohith, for cheering me up and the fun!

TABLE OF CONTENTS

ABSTRACT

TRANSCRIPTOME ANALYSIS OF ROOT DEVELOPMENT IN WHEAT (*TRITICUM AESTIVUM*) USING HIGH-THROUGHPUT SEQUENCING TECHNOLOGIES

GHANA S CHALLA

2018

Root provides plant water, nutrients and anchorage from soil. Most our knowledge of molecular mechanisms of root development is from the dicot model plant Arabidopsis, but very few studies have done in monocot crop systems like rice, maize, and wheat. We are studying very short root (VSR) phenotype in wheat, and lack of a sequenced reference genome in wheat prompted us to sequence and assemble the root transcriptome of the reference cultivar Chinese Spring (CS). A root transcriptome was assembled from the sequenced reads generated from root tip and the mature root tissues of CS. Approximately 169 million reads were successfully assembled into ~91K transcripts coding for functional proteins. Of these ~91K transcripts, 1,728 were differentially expressed in root tip as compared to the rest of the mature tissues. Generation of the root reference transcriptome and the availability of a reasonable reference genome sequence for wheat enabled us to analyze the gene expression in the long root (LR) and VSR. A total of 4,412 genes were differentially expressed in the VSR compared to the LR root tips. A significant portion of the differentially expressed genes functioning in the hormonal responses, regulation of

transcription, defense response, reactive oxygen species (ROS), abiotic stress response, lignin biosynthesis, calcium signaling, and autophagy pathways were induced. In addition, several negative regulators of cell proliferation, including homologs of the BIGBROTHER E3 ubiquitin ligase, and negative regulators of root cell elongation, such as genes encoding the FERONIA kinases and a RALF peptide hormone, were also up-regulated in VSR. Consistent with this, a large number of genes for chromatin replication and protein syntheses, including those coding for histones and ribosomal proteins, and cell wall remodeling enzymes, were down-regulated in VSR. The ROS and lignin accumulation in the VSR were further validated by histochemical staining. This research revealed several molecular mechanisms of root development, based on which a working model was proposed to explain the VSR development. Although the related pathways identified in Arabidopsis may play a similar role in wheat, the VSR phenotype is probably governed by a unique mechanism that may be cereal- or wheat-specific.

CHAPTER 1

Literature Review

INTRODUCTION

By the year 2050, the world population is projected to reach 9 billion. The increase of 3 billion of the global population will put the food security at risk. In the wake of this, cereals are becoming more important than ever. Wheat is the major cereal food crop consumed worldwide. It is the third major crop in terms of production after Corn and Rice. In United States (US), wheat is produced in almost every state, and the US is the world's fourth largest producer of wheat after China, India and Russian Federation (FAOSTAT, 2016; http://www.fao.org/faostat/). It provides about one-fifth of the calories consumed by the humans worldwide (http://faostat.fao.org). Wheat is cultivated in different environments around the world. It is grown across temperate, Mediterranean and tropical and subtropical parts globally.

Roots are the important part of the plants and form interface between the plants and the soil environment. The most important function of the roots is to uptake water from the surrounding soil and to provide nutrients for the plants to grow. They also act as the first line of sensors that detect the minute changes in the surrounding environment like water deficit, nutrient deficiency, ion toxicity, soil salinity, pH changes, etc. Many genes involved in the regulation root growth and development were identified in the dicot model plant Arabidopsis (de Dorlodot et al. 2007). By contrast, much fewer root regulatory genes have been identified rice, the grass model (Coudert et al. 2010; de Dorlodot et al. 2007; Hochholdinger et al. 2004). Several studies in Arabidopsis and rice revealed the key role plant hormones play in the initiation, regulation, and establishment of the root system. Hormones like auxin, gibberellins, brassinosteroids, jasmonate, and cytokinin regulate the

cell division and elongation in the root meristem and define the root system architecture (Aloni et al. 2006; Benkova and Hejatko 2009; Bishopp et al. 2009; Drisch and Stahl 2015). Reactive oxygen species (ROS) is known to control development by regulating cell elongation and also by interacting with different hormones (Causin et al. 2012; De Tullio et al. 2010; Lv et al. 2018). Unregulated accumulation of ROS is detrimental to the plant cells, and in Arabidopsis it was shown to affect the cell wall integrity and induce lignin deposition in the cell walls (Denness et al. 2011).

This chapter as an introduction will review wheat evolution and genomics, RNA sequencing (RNA-Seq)-based transcriptomics, transcriptome assembly and polyploidy challenges, and root development in cereal crops.

LITERATURE REVIVIEW

Wheat Genomes and Genomics

The wheat genus (*Triticum* L.) contains two diploid (*T. monococcum* and *T. urartu*), two tetraploid (*T. turgidum* and *T. timopheevii*) and two hexaploid species (*T. aestivum* and *T. zhukovskyi*) (van Slageren 1994). One diploid (*T. monococcum*) and both tetraploid species were domesticated, and the two hexaploid species arose under cultivation in Eurasia during the last 10,000 years (Salamini et al. 2002). Hexaploid common wheat, or bread wheat (*T. aestivum*, AABBDD genomes), originated in the Caspian Iran region (Wang et al. 2013) by hybridization between a cultivated form of tetraploid *T. turgidum* (AABB genomes) and diploid goatgrass *Aegilops tauschii* (DD genome) (Kihara 1944; McFadden and Sears 1946). The second hexaploid wheat, T. zhukovskyi arose by hybridization between

tetraploid Timopheevi wheat (*T. timopheevii,* AAGG genomes) and diploid einkorn (*T. monococcum* subsp. *monococcum*,). *T. turgidum* and *T. aestivum* constitute the Emmer lineage, and *T. timopheevii* and *T. zhukovskyi* the Timopheevi lineage (Gill et al. 2002). The diploid wheat *T. urartu* (genome AA) contributed the A genome to the Emmer and Timopheevi lineages (Dvorak et al., 1988; Kerby & Kuspira, 1988) and *A. speltoides* (SS genome), as the female parent, contributed cytoplasm (Wang et al. 1997) and the G genome (Dvorak and Zhang 1992; Kimber 1974) to *T. timopheevii* 0.4 million years ago and to the B genome and the cytoplasm of *T. turgidum* 0.7 Million years ago (Gornicki et al. 2014).

Recent years saw rapid progress in deciphering the wheat genomes. In addition to the reference genome sequences from common wheat cultivar Chinese Spring (CS) (Brenchley et al. 2012; The International Wheat Genome Sequencing Consortium 2014; Zimin et al. 2017a), twenty genomes of common wheat have been sequenced (Chapman et al. 2015; Montenegro et al. 2017). At the diploid level, the A genome (Ling et al. 2013) and the D genome donor species were also sequenced (Jia et al. 2013; Luo et al. 2017; Luo et al. 2013; Zhao et al. 2017; Zimin et al. 2017b) . Last year, IWGSC's efforts to generate a reference genome for wheat resulted in the first ever version of the wheat reference genome (https://www.wheatgenome.org/News/Latest-news/RefSeq-v1.0-URGI). Though the genome sequence and the annotation data are not available for large scale projects, they made it available for the studies involving single or few genes under the Toronto Agreement. To set a gold standard for the wheat genome sequencing efforts, the ~1 Gbp of the 3B chromosome was sequenced, and a pseudomolecule was constructed (Choulet et al. 2014). These sequences resources have greatly broken the bottleneck for map-based cloning of agriculturally important genes from the large, polyploid genome of wheat.

However, the draft, as well as the reference genome, only constitutes ~14 Gbp (~85%) of the hexaploid wheat estimated genome size of 17 Gbp. At the same time, transcriptomes of various wheat tissues and organs have been profiled using the RNA-Seq technology (Choulet et al. 2014; Fox et al. 2015; Liu et al. 2015; Pearce et al. 2015; Pearce et al. 2014; Pfeifer et al. 2014).

RNA sequencing and transcriptomics

The transcriptome is the complete set of transcripts in a cell at a specific developmental stage, cell type or physiological condition. For several decades, analyzing the transcriptome has been an essential tool in exploring the biological function and phenotype variation. Understanding the transcriptome is indispensable for interpreting the functional features of the genome and identifying the molecular make-up of the cells and tissues. The knowledge gained form transcriptome studies often help in understanding development, disease and responses to external stimuli. After the discovery of DNA as the genetic material and RNA as the intermediate in the central dogma of the molecular biology, the methods evolved from analyzing a single gene by Northern blotting to analyzing thousands of genes in a single experiment using microarrays. Though microarrays are high throughput assays, they are limited by the prior information of all the genes (transcriptome) expressed in the sample under consideration and the indirect method of measuring transcript abundance and eventually resulting in noisy data and low reproducibility (Morozova et al. 2009). They do not address the important biological questions like the possibility of novel transcripts. However, they were extensively used in several model systems and non-model systems like crop plants because of their affordability and ease in handling. Alternatively,

DNA sequencing methods, like Sanger's sequencing method, were used to analyze the transcriptome using strategies such as sequencing of the complimentary DNA (cDNA clones) (Kikuchi et al. 2003), expressed sequence tag (EST) sequencing (Manickavelu et al. 2012), which often involves cloning into a sequencing vector, but provided the advantage of novel transcript discovery and the full-length sequence to infer the gene structure and function (Seki et al. 2002). However, these strategies were less accurate, expensive and labor intensive to routinely use in the transcriptome quantification studies. Other methods like Serial Analysis of Gene Expression (SAGE) (Adams 1996; Velculescu et al. 1995), Cap Analysis of Gene Expression (CAGE) (Shiraki et al. 2003) and Massively Parallel Signature Sequencing (MPSS) (Brenner et al. 2000; Reinartz et al. 2002), which are tag-based sequencing approaches, were developed to overcome some of the above-mentioned shortcomings of microarray and low throughput sequencing, but with their own limitations (Harbers and Carninci 2005). After the successful completion of the human genome sequencing project and sequencing of the genomes of other major model system including the Arabidopsis, the model for the plant biology, several new sequencing technologies were developed, and these are collectively referred to as "next-generation sequencing" (NGS) or second-generation sequencing (SGS) methods (Bau et al. 2009; Holt and Jones 2008; Mardis 2008; Shendure and Ji 2008). These platforms include Roche/454 FLX, Illumina Genome Analyzer, ABI SOLiD, and the Heliscope (Mardis 2008). RNA sequencing or simply "RNA-Seq" is the transcriptome analysis approach utilizing high throughput DNA sequencing methods to quantify transcriptomes and also discover novel transcripts, non-coding RNA, detect isoforms, etc. This method has several advantages over the traditional methods of transcriptomics like analysis of several thousand transcripts

in one run, relatively inexpensive and not labor intensive. Many model and non-model organisms were analyzed using the RNA-Seq methods from understanding the transcriptome composition and transcript abundance to discovery of novel coding and non-coding transcripts previously not known to the scientific community (Chao et al. 2017; De Quattro et al. 2018; Hetzel et al. 2016; Marguerat and Bahler 2010; Quattro et al. 2017; Wang et al. 2009; Wilhelm and Landry 2009; Zou et al. 2016). Though these sequencing platforms differ in technology that they use for sequencing of the nucleic acids, they are based on a similar workflow. Although the sequencing technology is developed originally for the genomic DNA sequencing, adaptations were made to use for RNA sequencing. Here, instead of sequencing the RNA directly, double stranded cDNA is generated from RNA and is used as the sequencing template. Typically, the process involves shearing of the nucleic acids into sequencer compatible size and attach the DNA fragments called adapters, which contain unique sequences, at either end of the DNA fragment. These adapters allow the DNA fragment to attach to beads or the slide (called flow cell), depending on the platform, and create a unique locus which allows capturing the sequence data after the completion of the run. This process is called library preparation, and it is the key step in RNA-Seq as it determines the accuracy of the cDNA sequence data (Ansorge 2009). This method is a most straight-forward adaptation of the DNA sequencing for RNA-Seq, and one limitation it has is the loss of strand specificity in the output data. To overcome this hurdle for studies that require strand information several strand-specific library preparation techniques were developed (Lister et al. 2008; Marguerat and Bahler 2010; Parkhomchuk et al. 2009). The RNA-Seq methodology was quickly adapted and extensively utilized for understanding the role of small RNAs in regulation of the gene

expression in the eukaryotic systems. Eventually, it made to the messenger RNA (mRNA) expression studies. RNA-Seq was used in thousands of gene expression studies since its invention. The organisms and systems that were benefited by this technique were from model systems like human, mouse, worm, fission yeast, fruit fly, Arabidopsis, Medicago, etc. to non-model systems like fish, shrimp, dog, soybean, Brachypodium, wheat, barley, bamboo, sunflower, etc.

In addition to the quantification of transcripts in the sample under consideration, RNA-Seq is also being used for other studies. Transcriptome analysis of single-cell or Single-Cell RNA-Seq is one of the most recent applications of the RNA-Seq to understand the gene expression t the single-cell level (Efroni et al. 2015). Targeted RNA-Seq, where a specific set of transcripts were selected for sequencing (Levin et al. 2009; Mercer et al. 2014; Winz et al. 2017). Another application of RNA-Seq in transcriptome studies is for identifying the alternative splice forms and gene fusions (Filichkin et al. 2010). Sequencing of other RNA species like small RNA (Ahmed et al. 2014; Capece et al. 2015; Vidal et al. 2013), long non-coding RNA (Kang and Liu 2015; Lu et al. 2017), circular RNA (Liu et al. 2017; Pan et al. 2018), etc. has been carried out on several organisms, tissues, and physiological conditions. Since its inception a decade ago, RNA-Seq has been widely used and has become an indispensable tool for studying gene expression, understanding transcription, and its regulation and also RNA biogenesis. The third generation sequencing (TGS) methods already made their way into the genomic DNA sequencing studies and technologies like Pacific Bio (PacBio) and Nanopore (Oxford Nanopore Technologies) are already helping the large complex genome sequencing projects like Arabidopsis (Michael et al. 2018) tomato (Schmidt et al. 2017) and wheat (Clavijo et al. 2017; Zimin et al. 2017a;

Zimin et al. 2017b). Recently, full length RNA transcripts were sequenced using PacBio technology in wheat (Giolai et al. 2017; Huo et al. 2018; Zhang et al. 2014b) for transcriptome and gene family studies. The advantage with the TGS methods is the ability to sequence long fragments, several kilo bases, compared to the second-generation sequencers. With these new technologies in place, the transcriptome analysis studies will hugely benefit and can help unravel the complex biological problems in the plant biology and other organismal biology studies.

Transcriptome assembly and challenges with polyploidy

With the advent of the high-throughput sequencing platforms that generate several giga bases (Gbs) of the DNA sequencing DNA in a single run, shot gun genome sequencing and de novo assembly of the genome was made possible with reasonable expense and manpower. However, the cyber infrastructure played a major role in analyzing this kind of data, and several genome assembler tools and software were developed to achieve a draft genome assembly with reasonable quality. In the last decade since the invention of these platforms, several model and non-model genomes were sequenced and annotated to help the ongoing genetic and molecular studies. Wheat, one of the largest (genome size of ~17Gb) genome in plants and a with complex genome composition due to polyploidy was also sequenced by shotgun sequencing of the entire genome and also the shotgun sequencing of the flow sorted individual chromosomes (Brenchley et al. 2012; International Wheat Genome Sequencing 2014; Montenegro et al. 2017).

Our knowledge of the all the genes that are expressed in a cell was resultant of the gene prediction pipelines, EST and cDNA sequencing. This data is very limited and often

incomplete and inaccurate due to limitations in sequencing technologies, and mis-assemblies due to high homology between members of the same gene family, etc. Inspired by the whole genome sequencing, whole transcriptome sequencing or the RNA-Seq approach proved to be useful in understanding the complex landscape and dynamics of the transcriptome. The sensitivity and accuracy of these platforms were unmatched to that of the low-throughput technologies (Martin and Wang 2011). With the deep sequencing of the transcriptome at several hundreds of coverage depth, there is a possibility to capture the snapshot of the entire transcriptome including low and rarely expressly transcripts. Assembly programs to match the size and quantity of the transcriptome data have been developed, such as oases (Schulz et al. 2012), MIRA (http://chevreux.org/thesis/index.html), Edena (Hernandez et al. 2014; Xie et al. 2014), Soaptrans (Xie et al. 2014), IDBA-tran (Peng et al. 2013), Trinity (Grabherr et al. 2011; Haas et al. 2013) etc.

Plant genomes tend to be large and often contain highly repetitive DNA. They also have large gene families with members having high similarity at the nucleotide level. In wheat, addition to polyploidy, it is estimated that 90% of the genome is repetitive (Dvořák 2009). For transcriptome assembly of the polyploid plants is often challenging not only because the repetitiveness requires a large amount of computing resources and but also because there always is a problem of merging the closely related (homoeologous or paralogous) sequences and resulting in mis-assemblies (Krasileva et al. 2013; Schreiber et al. 2012). This limitation is data dependent and also depend on the downstream analysis (Duan et al. 2012; Hornett and Wheat 2012; Mundry et al. 2012; Vijay et al. 2013; Zhao et al. 2011). There were a few studies done in wheat, where the *de novo* assembled

transcriptome was used in identifying differential gene expression (Akhunova et al. 2010; Bouyioukos et al. 2013; Oono et al. 2013).

Root Development

Being underground and often ignored, root is referred to "hidden half" of the plant. A strong and well-established root system plays an important role in plant performance and yield in case of crop plants (Lynch 2011; Paez-Garcia et al. 2015). Molecular mechanisms involved in root development and growth has been extensively studied in Arabidopsis because of the small genome size, easy to grow and handle, short life cycle, ease in genetic transformation and specially, the availability of the genome sequence and as well as tagged mutant lines. Impressive progress has been made in understanding the root development in Arabidopsis. The root system in Arabidopsis consists of one primary root, which produces numerous lateral roots which in turn gives rise to higher-order lateral roots (Hodge et al. 2009). The Arabidopsis primary root has four distinct zones longitudinally. Meristem, where the active cell division occurs, followed by transition zone where the cell grows slowly in length and width, elongation zone with fast cell elongation but no growth in the cell width, and finally the growth terminating zone where the cells slow down the elongation and mature (Takatsuka and Umeda 2014; Ubeda-Tomas et al. 2012; Verbelen et al. 2006) (Figure 1.1). In Arabidopsis, a small number of stem cells give rise to all the different tissues of the root. The continuous growth and development of root is maintained by a pool of undifferentiated cells called the root apical meristem (RAM). The tip of the root meristem has the multipotent stem cells surrounded by a bunch of organizing cells called the quiescent center (QC) (van den Berg et al. 1997) (Figure 1.2). Cereal root system

is a complex structure and consists of several embryonic and post embryonic roots. The postembryonic roots are called nodal or crown roots. In wheat, the primary roots are three to five in number, and later crown roots develop from the node of the coleoptile. The radial and the crown roots branch out up to five degrees of branching. The cereal root system is different from Arabidopsis in many aspects. In additional to the presence of seminal roots, there is no shoot-borne root system, and the root hair pattern in Arabidopsis is regular compared to the cereals. A major difference in RAM anatomy between cereals and Arabidopsis is that there are 800 - 1200 quiescent center (QC) cells in cereal RAM, but Arabidopsis RAM has only four QC cells. At cell type level, there are multiple cortical cell layers in cereal roots, whereas there is only one layer in Arabidopsis. In Arabidopsis, there are eight cortical cells, but number of cortical cells is variable in cereal roots (Coudert et al. 2010; Hochholdinger et al. 2004). Very few studies were carried out to identify the genetic mechanisms involved in cereal root development. Most of the studies were carried out in rice, maize and wheat (Table 1.2). The variation in RSA is very prominent between dicots and monocots. Thus, it is not always possible to extrapolate the knowledge from Arabidopsis studies to the cereal plants (Lynch 1995; Osmont et al. 2007).

Hormonal regulation of root development and growth

Phytohormones play an important role in the regulation of root development and growth. During the globular stage of embryo development, auxin-induced degradation of IAA inducible protein IAA12 releases AUXIN RESPONSE FACTOR 5 (AFR5) from the co-transcriptional repression in provascular cells adjacent to hypophysis (Szemenyei et al. 2008; Weijers et al. 2006), which up-regulates auxin transporter *PINFORMED1* (*PIN1*).

In turn, PIN1-mediated auxin flows into hypophysis and accumulation in the hypophysis regulates the interaction between other ARF/IAA pairs that specify hypophysis (Weijers et al. 2006). At the same developmental stage, cytokinin response is specifically detected in the hypophysis and its apical daughter cell. In the basal hypophysis cell, auxin induces *ARABIDOPSIS RESPONSE REGULATOR7* (*ARR7*) and *ARR15*, repressors of cytokinin response (Muller and Sheen 2008), suggesting that antagonistic interaction between these two phytohormones regulates the establishment of root meristem including the stem cell niche.

During postembryonic root development, two parallel pathways specify the identity of root stem cell niche: PLETHORA (PLT) pathway and SHORT ROOT (SHR)/SCARECROW (SCR) pathway. The SHR/SCR pathway controls the radial patterning by specifying and regulating stem cell function. Both *SCR* and *SHR* encode transcription factors (TFs) of GRAS family regulating asymmetric cell division (Di Laurenzio et al. 1996; Helariutta et al. 2000; Sabatini et al. 2003). SHR and SCR form a heterodimer, which up-regulates SCR transcription in a feed-forward loop (Cui et al. 2007; Nakajima et al. 2001) The SHR also regulates vascular patterning through cytokinin homeostasis (Hao and Cui 2012). PLTs are the double AP2 type TFs acting downstream of the auxin signal pathway. *PLT* expression levels mirror the auxin gradient in the distal root tip and are maximized in the stem cell niche region (Aida et al. 2004). The PLT function is dose-based: high levels promote stem cell identity and maintenance, low levels promote division of stem cell daughters, and very low levels allow cells to differentiate (Galinha et al. 2007; Grieneisen et al. 2007). Therefore, the auxin/PLT pathway regulates the root meristem activity along the longitudinal axis. Transcription factors ARR1- and

SHY2/IAA3-mediated pathways maintain the balance between auxin and cytokinin. ARR1 is cytokinin-responsive, and SHY2/IAA3 is a repressor of auxin signaling. Activation of *SHY2/IAA3* by *ARR1* causes auxin redistribution and promotes cell differentiation (Dello Ioio et al. 2008).

Auxin is also involved in lateral root development through the IAA/ARF modules (De Rybel et al. 2010; De Smet et al. 2010; Fukaki and Tasaka 2009; Okushima et al. 2005), PLT11 and SHR (Lucas et al. 2011). In rice, auxin is a driving force for crown root initiation and development through two pathways. In one pathway, CROWN ROOTLESS4 (CRL4)-mediated auxin transport activates WUSCHEL-related homeobox protein WOX11, which suppresses cytokinin response via an ARR cascade. In another pathway, auxin induced degradation of IAA proteins and release ARFs, which activate *CRL1* expression and crown root initiation (reviewed in (Coudert et al. 2010).

Gibberellins (GAs) are also required for root development and growth (Inada and Shimmen 2000). In Arabidopsis, GAs specifically accumulate in the elongating endodermal cells (Shani et al. 2013), where DELLA protein mediates the GA signal (Ubeda-Tomas et al. 2008). Interacting with auxin and other phytohormones, GAs regulate lateral root formation in poplar (Gou et al. 2010). Opposite to auxin, cytokinin, and GAs, the stress hormone abscisic acid (ABA) and gaseous hormone ethylene inhibit root growth. The ABA effect on root inhibition is achieved through promotion of ethylene biosynthesis (Luo et al. 2014; Ma et al. 2014; Thole et al. 2014) and ethylene signaling pathways (Beaudoin et al. 2000; Ghassemian et al. 2000). Jasmonic acid (JA) or methyl jasmonate (MeJA) also inhibit root growth (Staswick et al. 1992), which block both the G1/S and G2/M transitions in the cell cycle (Swiatek et al. 2002). JAs also mediate root growth

inhibition caused by down-regulation of cellulose biosynthesis (Ellis et al. 2002), which invokes ectopic lignin deposition and defense responses (Cano-Delgado et al. 2003).

In addition, many peptide hormones, mainly belonging to CLV3/ESR-RELATED (CLE), ROOT GROWTH FACTOR (RGF) (Yue et al. 2016) and RAPID ALKALINIZATION FACTOR (RALF) families, are involved in root development and growth (Haruta et al. 2014; Leasure and He 2012; Yamada and Sawa 2013). Peptide RGF1 defines the stem cell niches by functioning upstream of PLTs (Matsuzaki et al. 2010). *CLE40*, expressed in differentiating root cells, restricts and positions the WUSCHEL-related homeobox TF WOX5 for suppressing stem cell fate through RLK and CRINKLY4 (Stahl et al. 2009). CRINKLY4, in turn, interacts with PP2A-3, a catalytic subunit of PP2A phosphatase enzyme and controls formative cell divisions (Yue et al. 2016). A more recent study showed that activation of FER by RALF causes phosphorylation of PM-anchored H+-ATPase, which mediates inhibition of proton transport and suppression of cell elongation in the primary root (Haruta et al. 2014).

ROS regulation of root development and growth

A growing body of evidence indicates that ROS join the phytohormones in the regulation of root development and growth. In the normal Arabidopsis roots, $^{\bullet}O_2^-$ is predominantly located in the apoplast of cell elongation zone and promote root elongation, whereas H_2O_2 accumulates in the differentiation zone and promotes root hair formation (Dunand et al. 2007). Genetic analyses showed that alteration of ROS homeostasis affects root cell proliferation, stem cell niche, meristem maintenance and lateral root formation. *ROOT MERISTEMLESS 1* (*RML1*) encodes γ–glutamylcysteine synthase, which maintains cell

proliferation via regulating redox status by synthesis of glutathione (GSH). Characterization of *rml1* mutant indicated that the redox status regulates the G1-to-S transition (Vernoux et al. 2000). The *rml1* mutant is normal in embryonic root development, but its failure to initiate cell division caused extremely short root (Cheng et al. 1995). UPBEAT1, a TF of bHLH family, directly regulates expression of a set of peroxidase genes that modulate the balance of these ROS in the transition zone between meristematic zone and elongation zone (Tsukagoshi et al. 2010). Cell type-specific transcriptome analysis showed that ROS also plays an important role in lateral root development (Manzano et al. 2014).

Two major sources of ROS have been recognized in root development: PM-localized NADPH oxidases, encoded by *RESPIRATORY BURST OXIDASE HOMOLOGS* (*RBOH*s), and the mitochondrial complex1 or complex3. In Arabidopsis, RbohC is required for root hair tip growth (Foreman et al. 2003; Takeda et al. 2008), and RbohA required for Casparian strip formation (Lee et al. 2013). In common bean, RbohB promotes lateral root elongation (Montiel et al. 2013). In rice, a point mutation in rice *SHORT POSTEMBRONIC ROOT1* (*SPR1*) gene, coding for a mitochondrial protein with an Armadillo-like repeat domain, cause accumulation of ROS in root tips, reduced cell elongation in the postembryonic roots, and lowered iron homeostasis (Jia et al. 2011).

More recently, ROS are found interacting with the phytohormones in regulating root development. In Arabidopsis, ABA overly sensitive mutants *abo6* and *abo8-1*, defective in pentatricopeptide proteins responsible for splicing transcripts of the mitochondrial complex1 genes, showed over accumulation of H2O2 in root tips and reduction of meristem size. The mutant effect on ROS and meristem can be further

enhanced by exogenous ABA treatment but rescued by exogenous GSH (He et al. 2012; Yang et al. 2014). The *abo8-1* mutation also blocks auxin signal and down-regulates *PLT1* and *PLT2* (Yang et al. 2014). In *Medicago truncatula*, a mutation in the *LATERAL ROOT ORGAN DEFECTIVE (LATD)* gene, encoding a nitrate transporter, perturbed the balance of ROS homeostasis by increasing $^{\bullet}O_2^-$ and reducing the H_2O_2 level in both meristem and maturation zones by upregulating the *RBOH* genes (Zhang et al. 2014a). Opposite to the above mutants, exogenous ABA recovered the ROS homeostasis and rescued root elongation in the latd mutant (Zhang et al. 2014a). All these indicated ROS function downstream of ABA in regulating root development and growth. In the RALF signaling pathway, FER either promotes or inhibits ROS depending upon cell types and signal contexts (Li and Zhang 2014). In root hair, FER interacts with guanine nucleotide exchange factors ROPGEF1, which activates the NADPH oxidases for ROS production through RAC/ROP GTPases (Duan et al. 2010). The RAC/ROP GTPases, at the same time, function as mediators for auxin-regulated gene expression (Wu et al. 2011), implicating another cross-talk between ROS and auxin signaling pathways. Under stress conditions, ROS are the universal signals. ROS-auxin cross-talks mediate the stress induced morphogenic responses, which comprises a mixture of growth inhibition and activation (Pasternak et al. 2005; Potters et al. 2007).

Small RNAs in root development

MicroRNAs (miRNAs) are a class of endogenous, small non-coding RNAs (sRNAs), that are ~21 nucleotides in length, and are trans-acting regulatory sequences that regulate gene expression by targeting mRNAs (Bartel 2009). These are present in animals and plants and

also in other eukaryotic organisms (Tarver et al. 2012). They are expressed in a variety of tissues, developmental stages and in different environmental conditions and regulate the developmental and physiological processes (reviewed in (Rubio-Somoza and Weigel 2011; Sunkar et al. 2012; Willmann and Poethig 2007). They are generated from the 3'- or 5'-arm of the single stranded hairpin RNAs. The biogenesis of miRNA in plants starts from the transcription of the primary miRNAs (pri-miRNAs) by RNA Polymerase II. These pri-miRNAs are recognized by the nuclear RNase enzyme called DICER-LIKE 1 (DCL1) and its accessory proteins called SERRATE (SE) and HYPONASTIC LEAVES1 (HYL1) and cleaved into precursor-miRNA (pre-miRNA) (Achkar et al. 2016; Axtell et al. 2011; Rogers and Chen 2013). DCL1 also cleaves the pre-miRNA to generate miRNA/miRNA* duplex. The miRNA/miRNA* is methylated at the 3′ terminus Hua Enhancer 1 (HEN1) and is then exported to the cytoplasm, possibly through HASTY (Achkar et al. 2016; Axtell et al. 2011; Bollman et al. 2003; Park et al. 2005; Rogers and Chen 2013). In cytoplasm, the miRNA/miRNA* duplex gets separated and the guide strand is loaded in to the RNA-induced silencing complex (RISC) by binding to the Argonaute (AGO) proteins (Rogers and Chen 2013). The miRNA production itself is a tightly regulated process and associated with transcription and splicing. The plant miRNA biogenesis and the mode of action in plants is entirely different from the animals and other eukaryotes (Achkar et al. 2016).

The transcriptional regulation by miRNA is considered an ancient evolutionary event, though, there is no concrete evidence to support this claim. A recent study analyzed the deposited miRNAs in miRBase, a miRNA database, concluded lack of evidence to confirm the authenticity of several conserved miRNA from different species (Taylor et al. 2014). However, in the plant kingdom several conserved miRNAs such as miR156 and

miR166 were conserved and they regulate flower development (Luo et al. 2013b). Another example is a dicot-specific miRNA family, miR403, which targets AGO2 and AGO3 (Cuperus et al. 2011) and acts as a feedback loop in its biogenesis. With the advent of high throughput sequencing methods, small RNA (sRNA) was sequenced from several plant species including wheat, and several novel miRNAs were identified (Jin et al. 2008; Kenan-Eichler et al. 2011; Li et al. 2013; Yao and Sun 2012). Comparative genome analysis combined with the sRNA data also revealed several miRNAs that are conserved in the monocots (Dryanova et al. 2008; Wei et al. 2009). In wheat, several studies were done for the discovery of small RNAs that play role in biotic and abiotic stress conditions (Bai et al. 2017; Chen et al. 2017; Gupta et al. 2012; Ragupathy et al. 2016; Song et al. 2017; Tang et al. 2012; Xin et al. 2010).

In Plants, miRNAs are implicated in several aspects of growth and development as regulatory machinery. Many miRNAs identified to be involved in the regulation of plant development are conserved in plant kingdom (Cuperus et al. 2011; Rubio-Somoza et al. 2009). In Arabidopsis, miR165/166 was found to regulate the shoot apical meristem (SAM) by targeting *CLASS III HOMEODOMAIN-LEUCINE ZIPPER* (*HD-ZIP III*) family genes (Barton 2010; Zhou et al. 2015; Zhu et al. 2011). miR394 was shown to target *LEAF CURLING RESPONSIVENESS* (*LCR*) a F-box gene involved in shoot meristem maintenance (Knauer et al. 2013; Litholdo et al. 2016). Another miRNA, named miR156 has been identified as major regulator in controlling phase transition. miR156 accumulates in juvenile plants and the level declines as the plant ages (Luo et al. 2013b; Teotia and Tang 2015). In transgenic Arabidopsis plants constitutively expressing miR156 were in juvenile phase for an extended period and the flowering is delayed. It targets *SQUAMOSA*

PROMOTER BINDING PROTEIN LIKE (*SPL*) family members (Huijser and Schmid 2011).

Several miRNAs were reported to be involved in the regulation of the root development in plants. Stem cell identity in root meristem is maintained by *PLT*s and were negatively regulated by GROWTH REGULATORY FACTOR (GRF) TFs. *GRF*s were regulated by miR396 which is known to control cell proliferation. In root meristem, *PLT* activates miR396 to repress the *GRF* expression thereby removing the suppression of *PLT*s thus forming a regulatory loop (Galinha et al. 2007; Rodriguez et al. 2015; Rodriguez et al. 2010; Rodriguez et al. 2016). It was shown that many of the auxin-related miRNAs control root development by modulating the auxin induced expression of root development genes in the root meristem. For example, miR160 family was reported to be regulating the root cap formation, gravity sensing, and root tip growth. miR160 regulates auxin responsive TFs, ARF10, 16, and 17 (Mallory et al. 2005; Nizampatnam et al. 2015; Wang et al. 2005). Another miRNA also targeting *ARF*s, miR167 was reported to control adventitious rooting in Arabidopsis and a crosslink between auxin and JA signaling pathways (Gutierrez et al. 2009; Gutierrez et al. 2012). In rice, miR167 also regulates *ARF8* and *ARF12*, which act on *GH3* to control auxin signaling and Root architecture (Meng et al. 2010). In Arabidopsis, miRNA165/166 were mobile and directs the gradient of cell differentiation along the xylem axis. This phenomenon works in a dose-dependent manner and were activated by *SHR* (Carlsbecker et al. 2010; Miyashima et al. 2011).

REFERENCES

Achkar NP, Cambiagno DA, Manavella PA (2016) miRNA Biogenesis: A Dynamic Pathway. Trends Plant Sci 21:1034-1044

Adams MD (1996) Serial analysis of gene expression: ESTs get smaller. Bioessays 18:261-262

Ahmed F, Senthil-Kumar M, Lee S, Dai X, Mysore KS, Zhao PX (2014) Comprehensive analysis of small RNA-seq data reveals that combination of miRNA with its isomiRs increase the accuracy of target prediction in Arabidopsis thaliana. RNA Biol 11:1414-1429

Aida M, Beis D, Heidstra R, Willemsen V, Blilou I, Galinha C, Nussaume L, Noh YS, Amasino R, Scheres B (2004) The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. Cell 119:109-120

Akhunova AR, Matniyazov RT, Liang H, Akhunov ED (2010) Homoeolog-specific transcriptional bias in allopolyploid wheat. BMC Genomics 11:505

Aloni R, Aloni E, Langhans M, Ullrich CI (2006) Role of cytokinin and auxin in shaping root architecture: regulating vascular differentiation, lateral root initiation, root apical dominance and root gravitropism. Ann Bot 97:883-893

Ansorge WJ (2009) Next-generation DNA sequencing techniques. N Biotechnol 25:195-203

Axtell MJ, Westholm JO, Lai EC (2011) Vive la difference: biogenesis and evolution of microRNAs in plants and animals. Genome Biol 12:221

Bai JF, Wang YK, Wang P, Duan WJ, Yuan SH, Sun H, Yuan GL, Ma JX, Wang N, Zhang FT, Zhang LP, Zhao CP (2017) Uncovering Male Fertility Transition Responsive miRNA in a Wheat Photo-Thermosensitive Genic Male Sterile Line by Deep Sequencing and Degradome Analysis. Front Plant Sci 8:1370

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215-233

Barton MK (2010) Twenty years on: the inner workings of the shoot apical meristem, a developmental dynamo. Dev Biol 341:95-113

Bau S, Schracke N, Kranzle M, Wu H, Stahler PF, Hoheisel JD, Beier M, Summerer D (2009) Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays. Anal Bioanal Chem 393:171-175

Beaudoin N, Serizet C, Gosti F, Giraudat J (2000) Interactions between abscisic acid and ethylene signaling cascades. Plant Cell 12:1103-1115

Benkova E, Hejatko J (2009) Hormone interactions at the root apical meristem. Plant Mol Biol 69:383-396

Bishopp A, Help H, Helariutta Y (2009) Cytokinin signaling during root development. Int Rev Cell Mol Biol 276:1-48

Bollman KM, Aukerman MJ, Park MY, Hunter C, Berardini TZ, Poethig RS (2003) HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. Development 130:1493-1504

Bouyioukos C, Moscou MJ, Champouret N, Hernandez-Pinzon I, Ward ER, Wulff BB (2013) Characterisation and analysis of the Aegilops sharonensis transcriptome, a wild relative of wheat in the Sitopsis section. PLoS One 8:e72782

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705-710

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18:630-634

Cano-Delgado A, Penfield S, Smith C, Catley M, Bevan M (2003) Reduced cellulose synthesis invokes lignification and defense responses in Arabidopsis thaliana. Plant J 34:351-362

Capece V, Garcia Vizcaino JC, Vidal R, Rahman RU, Pena Centeno T, Shomroni O, Suberviola I, Fischer A, Bonn S (2015) Oasis: online analysis of small RNA deep sequencing data. Bioinformatics 31:2205-2207

Carlsbecker A, Lee JY, Roberts CJ, Dettmer J, Lehesranta S, Zhou J, Lindgren O, Moreno-Risueno MA, Vaten A, Thitamadee S, Campilho A, Sebastian J, Bowman JL,

Helariutta Y, Benfey PN (2010) Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. Nature 465:316-321

Causin HF, Roqueiro G, Petrillo E, Lainez V, Pena LB, Marchetti CF, Gallego SM, Maldonado SI (2012) The control of root growth by reactive oxygen species in Salix nigra Marsh. seedlings. Plant Sci 183:197-205

Chao YT, Yen SH, Yeh JH, Chen WC, Shih MC (2017) Orchidstra 2.0-A Transcriptomics Resource for the Orchid Family. Plant Cell Physiol 58:e9

Chapman JA, Mascher M, Buluç A, Barry K, Georganas E, Session A, Strnadova V, Jenkins J, Sehgal S, Oliker L, Schmutz J, Yelick KA, Scholz U, Waugh R, Poland JA, Muehlbauer GJ, Stein N, Rokhsar DS (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biol 16:26

Chen XY, Yang Y, Ran LP, Dong ZD, Zhang EJ, Yu XR, Xiong F (2017) Novel Insights into miRNA Regulation of Storage Protein Biosynthesis during Wheat Caryopsis Development under Drought Stress. Front Plant Sci 8:1707

Cheng JC, Seeley KA, Sung ZR (1995) RML1 and RML2, Arabidopsis genes required for cell proliferation at the root tip. Plant Physiol 107:365-376

Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Le Gouis J, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Safar J, Dolezel J, Rogers J, Vandepoele K, Aury JM, Mayer K, Berges H, Quesneville H, Wincker P, Feuillet C (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science 345:1249721

Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H, Lipscombe J, Barker T, Lu FH, McKenzie N, Raats D, Ramirez-Gonzalez RH, Coince A, Peel N, Percival-Alwyn L, Duncan O, Trosch J, Yu G, Bolser DM, Namaati G, Kerhornou A, Spannagl M, Gundlach H, Haberer G, Davey RP, Fosker C, Palma FD, Phillips AL, Millar AH, Kersey PJ, Uauy C, Krasileva KV, Swarbreck D, Bevan MW, Clark MD (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. Genome Res 27:885-896

Coudert Y, Perin C, Courtois B, Khong NG, Gantet P (2010) Genetic control of root development in rice, the model cereal. Trends Plant Sci 15:219-226

Cui H, Levesque MP, Vernoux T, Jung JW, Paquette AJ, Gallagher KL, Wang JY, Blilou I, Scheres B, Benfey PN (2007) An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. Science 316:421-425

Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and functional diversification of MIRNA genes. Plant Cell 23:431-442

de Dorlodot S, Forster B, Pages L, Price A, Tuberosa R, Draye X (2007) Root system architecture: opportunities and constraints for genetic improvement of crops. Trends Plant Sci 12:474-481

De Quattro C, Mica E, Pe ME, Bertolini E (2018) Brachypodium distachyon Long Noncoding RNAs: Genome-Wide Identification and Expression Analysis. Methods Mol Biol 1667:31-42

De Rybel B, Vassileva V, Parizot B, Demeulenaere M, Grunewald W, Audenaert D, Van Campenhout J, Overvoorde P, Jansen L, Vanneste S, Moller B, Wilson M, Holman T, Van Isterdael G, Brunoud G, Vuylsteke M, Vernoux T, De Veylder L, Inze D, Weijers D, Bennett MJ, Beeckman T (2010) A novel aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity. Curr Biol 20:1697-1706

De Smet I, Lau S, Voss U, Vanneste S, Benjamins R, Rademacher EH, Schlereth A, De Rybel B, Vassileva V, Grunewald W, Naudts M, Levesque MP, Ehrismann JS, Inze D, Luschnig C, Benfey PN, Weijers D, Van Montagu MC, Bennett MJ, Jurgens G, Beeckman T (2010) Bimodular auxin response controls organogenesis in Arabidopsis. Proc Natl Acad Sci U S A 107:2705-2710

De Tullio MC, Jiang K, Feldman LJ (2010) Redox regulation of root apical meristem organization: connecting root development to its environment. Plant Physiol Biochem 48:328-336

Dello Ioio R, Nakamura K, Moubayidin L, Perilli S, Taniguchi M, Morita MT, Aoyama T, Costantino P, Sabatini S (2008) A genetic framework for the control of cell division and differentiation in the root meristem. Science 322:1380-1384

Denness L, McKenna JF, Segonzac C, Wormit A, Madhou P, Bennett M, Mansfield J, Zipfel C, Hamann T (2011) Cell wall damage-induced lignin biosynthesis is regulated by a reactive oxygen species- and jasmonic acid-dependent process in Arabidopsis. Plant Physiol 156:1364-1374

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN (1996) The SCARECROW gene regulates an

asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. Cell 86:423-433

Drisch RC, Stahl Y (2015) Function and regulation of transcription factors involved in root apical meristem and stem cell maintenance. Front Plant Sci 6:505

Dryanova A, Zakharov A, Gulick PJ (2008) Data mining for miRNAs and their targets in the Triticeae. Genome 51:433-443

Duan J, Xia C, Zhao G, Jia J, Kong X (2012) Optimizing *de novo* common wheat transcriptome assembly using short-read RNA-Seq data. BMC Genomics 13:392

Duan Q, Kita D, Li C, Cheung AY, Wu HM (2010) FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development. Proc Natl Acad Sci U S A 107:17821-17826

Dunand C, Crevecoeur M, Penel C (2007) Distribution of superoxide and hydrogen peroxide in Arabidopsis root and their influence on root development: possible interaction with peroxidases. New Phytol 174:332-341

Dvořák J (2009) Triticeae Genome Structure and Evolution. In: Muehlbauer GJ, Feuillet C (eds) Genetics and Genomics of the Triticeae. Springer US, New York, NY, pp 685-711

Dvorak J, Anderson OD, Bennetzen JL, Dai X, Dawson MW, Deal KR, Devos KM, Dolezel J, Gu YQ, Huo N, Kugler KG, Leroy P, Liang Y, Liu Z, Luo M-C, Lyons E, Ma Z, Mao L, Mayer KFX, McCombie WR, McGuire P, Mueller HG, Ouyang S, Pertea G, Puiu D, Salzberg SL, Soderlund C, Sun Q, Twardziok SO, Wang H, Wang Y, Wang Z, Wicker T, Xiao L, Xiao M, You FM, Zhu T, Zimin A (2016)

Reference Sequence of the Genome of Aegilops tauschii, the Progenitor of the Wheat D Genome. Plant and Animal Genome XXIV, San Diego, p W972

Dvorak J, Zhang HB (1992) Reconstruction of the phylogeny of the genus Triticum from variation in repeated nucleotide sequences. Theor Appl Genet 84:419-429

Efroni I, Ip PL, Nawy T, Mello A, Birnbaum KD (2015) Quantification of cell identity from single-cell gene expression profiles. Genome Biol 16:9

Ellis C, Karafyllidis I, Wasternack C, Turner JG (2002) The Arabidopsis mutant cev1 links cell wall signaling to jasmonate and ethylene responses. Plant Cell 14:1557-1566

Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong WK, Mockler TC (2010) Genome-wide mapping of alternative splicing in Arabidopsis thaliana. Genome Res 20:45-58

Foreman J, Demidchik V, Bothwell JH, Mylona P, Miedema H, Torres MA, Linstead P, Costa S, Brownlee C, Jones JD, Davies JM, Dolan L (2003) Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. Nature 422:442-446

Fox SE, Geniza M, Hanumappa M, Naithani S, Sullivan C, Preece J, Tiwari VK, Elser J, Leonard JM, Sage A, Gresham C, Kerhornou A, Bolser D, McCarthy F, Kersey P, Lazo GR, Jaiswal P (2015) De novo transcriptome assembly and analyses of gene expression during photomorphogenesis in diploid wheat *Triticum monococcum*. PLoS One 9:e96855

Fukaki H, Tasaka M (2009) Hormone interactions during lateral root formation. Plant Mol Biol 69:437-449

Galinha C, Hofhuis H, Luijten M, Willemsen V, Blilou I, Heidstra R, Scheres B (2007) PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. Nature 449:1053-1057

Ghassemian M, Nambara E, Cutler S, Kawaide H, Kamiya Y, McCourt P (2000) Regulation of abscisic acid signaling by the ethylene response pathway in Arabidopsis. Plant Cell 12:1117-1126

Gill BS, FAO RIPPaPD, eng, Friebe B, Curtis BC, (ed.), Rajaram S, (ed.), Gomez Macpherson H, (ed.) (2002) Cytogenetics, phylogeny and evolution of cultivated wheats

Giolai M, Paajanen P, Verweij W, Witek K, Jones JDG, Clark MD (2017) Comparative analysis of targeted long read sequencing approaches for characterization of a plant's immune receptor repertoire. BMC Genomics 18:564

Gornicki P, Zhu H, Wang J, Challa GS, Zhang Z, Gill BS, Li W (2014) The chloroplast view of the evolution of polyploid wheat. New Phytol 204:704-714

Gou J, Strauss SH, Tsai CJ, Fang K, Chen Y, Jiang X, Busov VB (2010) Gibberellins regulate lateral root formation in Populus through interactions with auxin and other hormones. Plant Cell 22:623-639

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol 29:644-652

Grieneisen VA, Xu J, Maree AF, Hogeweg P, Scheres B (2007) Auxin transport is sufficient to generate a maximum and gradient guiding root growth. Nature 449:1008-1013

Gupta OP, Permar V, Koundal V, Singh UD, Praveen S (2012) MicroRNA regulated defense responses in Triticum aestivum L. during Puccinia graminis f.sp. tritici infection. Mol Biol Rep 39:817-824

Gutierrez L, Bussell JD, Pacurar DI, Schwambach J, Pacurar M, Bellini C (2009) Phenotypic plasticity of adventitious rooting in Arabidopsis is controlled by complex regulation of AUXIN RESPONSE FACTOR transcripts and microRNA abundance. Plant Cell 21:3119-3132

Gutierrez L, Mongelard G, Flokova K, Pacurar DI, Novak O, Staswick P, Kowalczyk M, Pacurar M, Demailly H, Geiss G, Bellini C (2012) Auxin controls Arabidopsis adventitious root initiation by regulating jasmonic acid homeostasis. Plant Cell 24:2515-2527

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494-1512

Hao Y, Cui H (2012) SHORT-ROOT regulates vascular patterning, but not apical meristematic activity in the Arabidopsis root through cytokinin homeostasis. Plant Signal Behav 7:314-317

Harbers M, Carninci P (2005) Tag-based approaches for transcriptome research and genome annotation. Nat Methods 2:495-502

Haruta M, Sabat G, Stecker K, Minkoff BB, Sussman MR (2014) A peptide hormone and its receptor protein kinase regulate plant cell expansion. Science 343:408-411

He J, Duan Y, Hua D, Fan G, Wang L, Liu Y, Chen Z, Han L, Qu LJ, Gong Z (2012) DEXH box RNA helicase-mediated mitochondrial reactive oxygen species production in Arabidopsis mediates crosstalk between abscisic acid and auxin signaling. Plant Cell 24:1815-1833

Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, Jung J, Sena G, Hauser MT, Benfey PN (2000) The *SHORT-ROOT* gene controls radial patterning of the Arabidopsis root through radial signaling. Cell 101:555-567

Hernandez D, Tewhey R, Veyrieras JB, Farinelli L, Osteras M, Francois P, Schrenzel J (2014) De novo finished 2.8 Mbp Staphylococcus aureus genome assembly from 100 bp short and long range paired-end reads. Bioinformatics 30:40-49

Hetzel J, Duttke SH, Benner C, Chory J (2016) Nascent RNA sequencing reveals distinct features in plant transcription. Proc Natl Acad Sci U S A 113:12316-12321

Hochholdinger F, Park WJ, Sauer M, Woll K (2004) From weeds to crops: genetic analysis of root development in cereals. Trends Plant Sci 9:42-48

Hodge A, Berta G, Doussan C, Merchan F, Crespi M (2009) Plant root growth, architecture and function. Plant and Soil 321:153-187

Holt RA, Jones SJ (2008) The new paradigm of flow cell sequencing. Genome Res 18:839-846

Hornett EA, Wheat CW (2012) Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. BMC Genomics 13:361

Huijser P, Schmid M (2011) The control of developmental phase transitions in plants. Development 138:4117-4129

Huo N, Zhu T, Altenbach S, Dong L, Wang Y, Mohr T, Liu Z, Dvorak J, Luo MC, Gu YQ (2018) Dynamic Evolution of alpha-Gliadin Prolamin Gene Family in Homeologous Genomes of Hexaploid Wheat. Sci Rep 8:5181

Inada S, Shimmen T (2000) Regulation of elongation growth by gibberellin in root segments of Lemna minor. Plant Cell Physiol 41:932-939

International Wheat Genome Sequencing C (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345:1251788

Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R, Zhang C, Ma Y, Gao L, Gao C, Spannagl M, Mayer KF, Li D, Pan S, Zheng F, Hu Q, Xia X, Li J, Liang Q, Chen J, Wicker T, Gou C, Kuang H, He G, Luo Y, Keller B, Xia Q, Lu P, Wang J, Zou H, Zhang R, Xu J, Gao J, Middleton C, Quan Z, Liu G, Wang J, Consortium. IWGS, Yang H, Liu X, He Z, Mao L, Wang J (2013) Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91-95

Jia L, Wu Z, Hao X, Carrie C, Zheng L, Whelan J, Wu Y, Wang S, Wu P, Mao C (2011) Identification of a novel mitochondrial protein, short postembryonic roots 1 (SPR1), involved in root development and iron homeostasis in Oryza sativa. New Phytol 189:843-855

Jin W, Li N, Zhang B, Wu F, Li W, Guo A, Deng Z (2008) Identification and verification of microRNA in wheat (Triticum aestivum). J Plant Res 121:351-355

Kang C, Liu Z (2015) Global identification and analysis of long non-coding RNAs in diploid strawberry Fragaria vesca during flower and fruit development. BMC Genomics 16:815

Kenan-Eichler M, Leshkowitz D, Tal L, Noor E, Melamed-Bessudo C, Feldman M, Levy AA (2011) Wheat hybridization and polyploidization results in deregulation of small RNAs. Genetics 188:263-272

Kihara H (1944) Discovery of the DD-analyser, one of the ancestors of Triticum vulgare (Japanese). Agric and Hort (Tokyo) 19:13 -14

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T, Foundation of Advancement of International Science Genome S, Analysis G, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K, Riken, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y, Yasunishi A (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376-379

Kimber G (1974) A reassessment of the origin of the polyploid wheats. Genetics 78:487-492

Knauer S, Holt AL, Rubio-Somoza I, Tucker EJ, Hinze A, Pisch M, Javelle M, Timmermans MC, Tucker MR, Laux T (2013) A protodermal miR394 signal defines a region of stem cell competence in the Arabidopsis shoot meristem. Dev Cell 24:125-132

Krasileva KV, Buffalo V, Bailey P, Pearce S, Ayling S, Tabbita F, Soria M, Wang S, Consortium I, Akhunov E, Uauy C, Dubcovsky J (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. Genome Biol 14:R66

Leasure CD, He ZH (2012) CLE and RGF family peptide hormone signaling in plant development. Mol Plant 5:1173-1175

Lee Y, Rubio MC, Alassimone J, Geldner N (2013) A mechanism for localized lignin deposition in the endodermis. Cell 153:402-412

Levin JZ, Berger MF, Adiconis X, Rogov P, Melnikov A, Fennell T, Nusbaum C, Garraway LA, Gnirke A (2009) Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts. Genome Biol 10:R115

Li S, Zhang Y (2014) To grow or not to grow: FERONIA has her say. Mol Plant 7:1261-1263

Li YF, Zheng Y, Jagadeeswaran G, Sunkar R (2013) Characterization of small RNAs and their target genes in wheat seedlings using sequencing-based approaches. Plant Sci 203-204:17-24

Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo M, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87-90

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133:523-536

Litholdo CG, Jr., Parker BL, Eamens AL, Larsen MR, Cordwell SJ, Waterhouse PM (2016) Proteomic Identification of Putative MicroRNA394 Target Genes in Arabidopsis thaliana Identifies Major Latex Protein Family Members Critical for Normal Development. Mol Cell Proteomics 15:2033-2047

Liu T, Zhang L, Chen G, Shi T (2017) Identifying and Characterizing the Circular RNAs during the Lifespan of Arabidopsis Leaves. Front Plant Sci 8:1278

Liu Z, Xin M, Qin J, Peng H, Ni Z, Yao Y, Sun Q (2015) Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). BMC Plant Biol 15:152

Lu Z, Xia X, Jiang B, Ma K, Zhu L, Wang L, Jin B (2017) Identification and characterization of novel lncRNAs in Arabidopsis thaliana. Biochem Biophys Res Commun 488:348-354

Lucas M, Swarup R, Paponov IA, Swarup K, Casimiro I, Lake D, Peret B, Zappala S, Mairhofer S, Whitworth M, Wang J, Ljung K, Marchant A, Sandberg G, Holdsworth MJ, Palme K, Pridmore T, Mooney S, Bennett MJ (2011) Short-Root regulates primary, lateral, and adventitious root development in Arabidopsis. Plant Physiol 155:384-398

Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, McGuire PE, Liu S, Long H, Ramasamy RK, Rodriguez JC, Van SL, Yuan L, Wang Z, Xia Z, Xiao L, Anderson OD, Ouyang S, Liang Y, Zimin AV, Pertea G, Qi P, Bennetzen JL, Dai X, Dawson MW, Muller HG, Kugler K, Rivarola-Duarte L, Spannagl M, Mayer KFX, Lu FH, Bevan MW, Leroy P, Li P, You FM, Sun Q, Liu Z, Lyons E, Wicker T, Salzberg SL, Devos KM, Dvorak J (2017) Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. Nature 551:498-502

Luo MC, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, Huo N, Wang Y, Wang J, Chen S, Jorgensen CM, Zhang Y, McGuire PE, Pasternak S, Stein JC, Ware D, Kramer M, McCombie WR, Kianian SF, Martis MM, Mayer KF, Sehgal SK, Li W, Gill BS, Bevan MW, Simková H, Dolezel J, Weining S, Lazo GR, Anderson OD, Dvorak J (2013a) A 4-gigabase physical map unlocks the structure and evolution of the complex genome of Aegilops tauschii, the wheat D-genome progenitor. Proc Natl Acad Sci U S A 110:7940-7945

Luo X, Chen Z, Gao J, Gong Z (2014) Abscisic acid inhibits root growth in Arabidopsis through ethylene biosynthesis. Plant J 79:44-55

Luo Y, Guo Z, Li L (2013b) Evolutionary conservation of microRNA regulatory programs in plant flower development. Dev Biol 380:133-144

Lv B, Tian H, Zhang F, Liu J, Lu S, Bai M, Li C, Ding Z (2018) Brassinosteroids regulate root growth by controlling reactive oxygen species homeostasis and dual effect on ethylene synthesis in Arabidopsis. PLoS Genet 14:e1007144

Lynch J (1995) Root Architecture and Plant Productivity. Plant Physiol 109:7-13

Lynch JP (2011) Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. Plant Physiol 156:1041-1049

Ma B, Yin CC, He SJ, Lu X, Zhang WK, Lu TG, Chen SY, Zhang JS (2014) Ethylene-induced inhibition of root growth requires abscisic acid function in rice (Oryza sativa L.) seedlings. PLoS Genet 10:e1004701

Mallory AC, Bartel DP, Bartel B (2005) MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. Plant Cell 17:1360-1375

Manickavelu A, Kawaura K, Oishi K, Shin IT, Kohara Y, Yahiaoui N, Keller B, Abe R, Suzuki A, Nagayama T, Yano K, Ogihara Y (2012) Comprehensive functional analyses of expressed sequence tags in common wheat (Triticum aestivum). DNA Res 19:165-177

Manzano C, Pallero-Baena M, Casimiro I, De Rybel B, Orman-Ligeza B, Van Isterdael G, Beeckman T, Draye X, Casero P, Del Pozo JC (2014) The Emerging Role of Reactive Oxygen Species Signaling during Lateral Root Development. Plant Physiol 165:1105-1119

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387-402

Marguerat S, Bahler J (2010) RNA-seq: from technology to biology. Cell Mol Life Sci 67:569-579

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12:671-682

Matsuzaki Y, Ogawa-Ohnishi M, Mori A, Matsubayashi Y (2010) Secreted peptide signals required for maintenance of root stem cell niche in Arabidopsis. Science 329:1065-1067

McFadden ES, Sears ER (1946) The origin of Triticum spelta and its free-threshing hexaploid relatives. J Hered 37:81 107

Meng Y, Chen D, Ma X, Mao C, Cao J, Wu P, Chen M (2010) Mechanisms of microRNA-mediated auxin signaling inferred from the rice mutant osaxr. Plant Signal Behav 5:252-254

Mercer TR, Clark MB, Crawford J, Brunck ME, Gerhardt DJ, Taft RJ, Nielsen LK, Dinger ME, Mattick JS (2014) Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat Protoc 9:989-1009

Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP, Lanz C, Loudet O, Weigel D, Ecker JR (2018) High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. Nat Commun 9:541

Miyashima S, Koi S, Hashimoto T, Nakajima K (2011) Non-cell-autonomous microRNA165 acts in a dose-dependent manner to regulate multiple differentiation status in the Arabidopsis root. Development 138:2303-2313

Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan CK, Visendi P, Lai K, Doležel J, Batley J, Edwards D (2017) The pangenome of hexaploid bread wheat. Plant J 90:1007-1013

Montiel J, Arthikala MK, Quinto C (2013) Phaseolus vulgaris RbohB functions in lateral root development. Plant Signal Behav 8:e22694

Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet 10:135-151

Muller B, Sheen J (2008) Cytokinin and auxin interaction in root stem-cell specification during early embryogenesis. Nature 453:1094-1097

Mundry M, Bornberg-Bauer E, Sammeth M, Feulner PG (2012) Evaluating characteristics of de novo assembly software on 454 transcriptome data: a simulation approach. PLoS One 7:e31410

Nakajima K, Sena G, Nawy T, Benfey PN (2001) Intercellular movement of the putative transcription factor SHR in root patterning. Nature 413:307-311

Nizampatnam NR, Schreier SJ, Damodaran S, Adhikari S, Subramanian S (2015) microRNA160 dictates stage-specific auxin and cytokinin sensitivities and directs soybean nodule development. Plant J 84:140-153

Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Lui A, Nguyen D, Onodera C, Quach H, Smith A, Yu G, Theologis A (2005) Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: unique and overlapping functions of ARF7 and ARF19. Plant Cell 17:444-463

Oono Y, Kobayashi F, Kawahara Y, Yazawa T, Handa H, Itoh T, Matsumoto T (2013) Characterisation of the wheat (Triticum aestivum L.) transcriptome by de novo assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. BMC Genomics 14:77

Osmont KS, Sibout R, Hardtke CS (2007) Hidden branches: developments in root system architecture. Annu Rev Plant Biol 58:93-113

Paez-Garcia A, Motes CM, Scheible WR, Chen R, Blancaflor EB, Monteros MJ (2015) Root Traits and Phenotyping Strategies for Plant Improvement. Plants (Basel) 4:334-355

Pan T, Sun X, Liu Y, Li H, Deng G, Lin H, Wang S (2018) Heat stress alters genome-wide profiles of circular RNAs in Arabidopsis. Plant Mol Biol 96:217-229

Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS (2005) Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A 102:3691-3696

Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res 37:e123

Pasternak T, Potters G, Caubergs R, Jansen MA (2005) Complementary interactions between oxidative stress and auxins control plant growth responses at plant, organ, and cellular level. J Exp Bot 56:1991-2001

Pearce S, Huttly AK, Prosser IM, Li YD, Vaughan SP, Gallova B, Patil A, Coghill JA, Dubcovsky J, Hedden P, Phillips AL (2015) Heterologous expression and transcript

analysis of gibberellin biosynthetic genes of grasses reveals novel functionality in the GA3ox family. BMC Plant Biol 15:130

Pearce S, Tabbita F, Cantu D, Buffalo V, Avni R, Vazquez-Gross H, Zhao R, Conley CJ, Distelfeld A, Dubcovsky J (2014) Regulation of Zn and Fe transporters by the GPC1 gene during early wheat monocarpic senescence. BMC Plant Biol 14:368

Peng Y, Leung HC, Yiu SM, Lv MJ, Zhu XG, Chin FY (2013) IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics 29:i326-334

Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, Consortium. IWGS, F. MK, Olsen OA (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345:1250091

Potters G, Pasternak TP, Guisez Y, Palme KJ, Jansen MA (2007) Stress-induced morphogenic responses: growing out of trouble? Trends Plant Sci 12:98-105

Quattro C, Enrico Pe M, Bertolini E (2017) Long noncoding RNAs in the model species Brachypodium distachyon. Sci Rep 7:11252

Ragupathy R, Ravichandran S, Mahdi MS, Huang D, Reimer E, Domaratzki M, Cloutier S (2016) Deep sequencing of wheat sRNA transcriptome reveals distinct temporal expression pattern of miRNAs in response to heat, light and UV. Sci Rep 6:39373

Reinartz J, Bruyns E, Lin JZ, Burcham T, Brenner S, Bowen B, Kramer M, Woychik R (2002) Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Brief Funct Genomic Proteomic 1:95-104

Rodriguez RE, Ercoli MF, Debernardi JM, Breakfield NW, Mecchia MA, Sabatini M, Cools T, De Veylder L, Benfey PN, Palatnik JF (2015) MicroRNA miR396 Regulates the Switch between Stem Cells and Transit-Amplifying Cells in Arabidopsis Roots. Plant Cell 27:3354-3366

Rodriguez RE, Mecchia MA, Debernardi JM, Schommer C, Weigel D, Palatnik JF (2010) Control of cell proliferation in Arabidopsis thaliana by microRNA miR396. Development 137:103-112

Rodriguez RE, Schommer C, Palatnik JF (2016) Control of cell proliferation by microRNAs in plants. Curr Opin Plant Biol 34:68-76

Rogers K, Chen X (2013) Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell 25:2383-2399

Rubio-Somoza I, Cuperus JT, Weigel D, Carrington JC (2009) Regulation and functional specialization of small RNA-target nodes during plant development. Curr Opin Plant Biol 12:622-627

Rubio-Somoza I, Weigel D (2011) MicroRNA networks and developmental plasticity in plants. Trends Plant Sci 16:258-264

Sabatini S, Heidstra R, Wildwater M, Scheres B (2003) SCARECROW is involved in positioning the stem cell niche in the Arabidopsis root meristem. Genes Dev 17:354-358

Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the near east. Nat Rev Genet 3:429-441

Schmidt MH, Vogel A, Denton AK, Istace B, Wormit A, van de Geest H, Bolger ME, Alseekh S, Mass J, Pfaff C, Schurr U, Chetelat R, Maumus F, Aury JM, Koren S,

Fernie AR, Zamir D, Bolger AM, Usadel B (2017) De Novo Assembly of a New Solanum pennellii Accession Using Nanopore Sequencing. Plant Cell 29:2336-2348

Schreiber AW, Hayden MJ, Forrest KL, Kong SL, Langridge P, Baumann U (2012) Transcriptome-scale homoeolog-specific transcript assemblies of bread wheat. BMC Genomics 13:492

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086-1092

Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K, Shinagawa A, Shinozaki K (2002) Functional annotation of a full-length Arabidopsis cDNA collection. Science 296:141-145

Shani E, Weinstain R, Zhang Y, Castillejo C, Kaiserli E, Chory J, Tsien RY, Estelle M (2013) Gibberellins accumulate in the elongating endodermal cells of Arabidopsis root. Proc Natl Acad Sci U S A 110:4834-4839

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135-1145

Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajska A, Harbers M, Kawai J, Carninci P, Hayashizaki Y (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A 100:15776-15781

Song G, Zhang R, Zhang S, Li Y, Gao J, Han X, Chen M, Wang J, Li W, Li G (2017) Response of microRNAs to cold treatment in the young spikes of common wheat. BMC Genomics 18:212

Stahl Y, Wink RH, Ingram GC, Simon R (2009) A signaling module controlling the stem cell niche in Arabidopsis root meristems. Curr Biol 19:909-914

Staswick PE, Su W, Howell SH (1992) Methyl jasmonate inhibition of root growth and induction of a leaf protein are decreased in an Arabidopsis thaliana mutant. Proc Natl Acad Sci U S A 89:6837-6840

Sunkar R, Li YF, Jagadeeswaran G (2012) Functions of microRNAs in plant stress responses. Trends Plant Sci 17:196-203

Swiatek A, Lenjou M, Van Bockstaele D, Inze D, Van Onckelen H (2002) Differential effect of jasmonic acid and abscisic acid on cell cycle progression in tobacco BY-2 cells. Plant Physiol 128:201-211

Szemenyei H, Hannon M, Long JA (2008) TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis. Science 319:1384-1386

Takatsuka H, Umeda M (2014) Hormonal control of cell division and elongation along differentiation trajectories in roots. J Exp Bot 65:2633-2643

Takeda S, Gapper C, Kaya H, Bell E, Kuchitsu K, Dolan L (2008) Local positive feedback regulation determines cell shape in root hair cells. Science 319:1241-1244

Tang Z, Zhang L, Xu C, Yuan S, Zhang F, Zheng Y, Zhao C (2012) Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing. Plant Physiol 159:721-738

Tarver JE, Donoghue PC, Peterson KJ (2012) Do miRNAs have a deep evolutionary history? Bioessays 34:857-866

Taylor RS, Tarver JE, Hiscock SJ, Donoghue PC (2014) Evolutionary history of plant microRNAs. Trends Plant Sci 19:175-182

Teotia S, Tang G (2015) To bloom or not to bloom: role of microRNAs in plant flowering. Mol Plant 8:359-377

The International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345

Thole JM, Beisner ER, Liu J, Venkova SV, Strader LC (2014) Abscisic acid regulates root elongation through the activities of auxin and ethylene in *Arabidopsis thaliana*. G3 (Bethesda) 4:1259-1274

Tsukagoshi H, Busch W, Benfey PN (2010) Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. Cell 143:606-616

Ubeda-Tomas S, Beemster GT, Bennett MJ (2012) Hormonal regulation of root growth: integrating local activities into global behaviour. Trends Plant Sci 17:326-331

Ubeda-Tomas S, Swarup R, Coates J, Swarup K, Laplaze L, Beemster GT, Hedden P, Bhalerao R, Bennett MJ (2008) Root growth in Arabidopsis requires gibberellin/DELLA signalling in the endodermis. Nat Cell Biol 10:625-628

van den Berg C, Willemsen V, Hendriks G, Weisbeek P, Scheres B (1997) Short-range control of cell differentiation in the Arabidopsis root meristem. Nature 390:287-289

Van Norman JM, Breakfield NW, Benfey PN (2011) Intercellular communication during plant development. Plant Cell 23:855-864

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:484-487

Verbelen JP, De Cnodder T, Le J, Vissenberg K, Baluska F (2006) The Root Apex of Arabidopsis thaliana Consists of Four Distinct Zones of Growth Activities: Meristematic Zone, Transition Zone, Fast Elongation Zone and Growth Terminating Zone. Plant Signal Behav 1:296-304

Vernoux T, Wilson RC, Seeley KA, Reichheld JP, Muroy S, Brown S, Maughan SC, Cobbett CS, Van Montagu M, Inze D, May MJ, Sung ZR (2000) The ROOT MERISTEMLESS1/CADMIUM SENSITIVE2 gene defines a glutathione-dependent pathway involved in initiation and maintenance of cell division during postembryonic root development. Plant Cell 12:97-110

Vidal EA, Moyano TC, Krouk G, Katari MS, Tanurdzic M, McCombie WR, Coruzzi GM, Gutierrez RA (2013) Integrated RNA-seq and sRNA-seq analysis identifies novel nitrate-responsive genes in Arabidopsis thaliana roots. BMC Genomics 14:701

Vijay N, Poelstra JW, Kunstner A, Wolf JB (2013) Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. Mol Ecol 22:620-634

Wang GZ, Miyashita NT, Tsunewaki K (1997) Plasmon analyses of Triticum (wheat) and Aegilops: PCR-single-strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs. Proc Natl Acad Sci U S A 94:14570-14577

Wang J, Luo MC, Chen Z, You FM, Wei Y, Zheng Y, Dvorak J (2013) *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. New Phytol 198:925-937

Wang JW, Wang LJ, Mao YB, Cai WJ, Xue HW, Chen XY (2005) Control of root cap formation by MicroRNA-targeted auxin response factors in Arabidopsis. Plant Cell 17:2204-2216

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57-63

Wei B, Cai T, Zhang R, Li A, Huo N, Li S, Gu YQ, Vogel J, Jia J, Qi Y, Mao L (2009) Novel microRNAs uncovered by deep sequencing of small RNA transcriptomes in bread wheat (Triticum aestivum L.) and Brachypodium distachyon (L.) Beauv. Funct Integr Genomics 9:499-511

Weijers D, Schlereth A, Ehrismann JS, Schwank G, Kientz M, Jurgens G (2006) Auxin triggers transient local signaling for cell specification in Arabidopsis embryogenesis. Dev Cell 10:265-270

Wilhelm BT, Landry JR (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48:249-257

Willmann MR, Poethig RS (2007) Conservation and evolution of miRNA regulatory programs in plant development. Curr Opin Plant Biol 10:503-511

Winz ML, Cahova H, Nubel G, Frindert J, Hofer K, Jaschke A (2017) Capture and sequencing of NAD-capped RNA sequences with NAD captureSeq. Nat Protoc 12:122-149

Wu HM, Hazak O, Cheung AY, Yalovsky S (2011) RAC/ROP GTPases and auxin signaling. Plant Cell 23:1208-1218

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam TW, Li Y, Xu X, Wong GK, Wang J (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660-1666

Xin M, Wang Y, Yao Y, Xie C, Peng H, Ni Z, Sun Q (2010) Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (Triticum aestivum L.). BMC Plant Biol 10:123

Yamada M, Sawa S (2013) The roles of peptide hormones during plant root development. Curr Opin Plant Biol 16:56-61

Yang L, Zhang J, He J, Qin Y, Hua D, Duan Y, Chen Z, Gong Z (2014) ABA-mediated ROS in mitochondria regulate root meristem activity by controlling PLETHORA expression in Arabidopsis. PLoS Genet 10:e1004791

Yao Y, Sun Q (2012) Exploration of small non coding RNAs in wheat (Triticum aestivum L.). Plant Mol Biol 80:67-73

Yue K, Sandal P, Williams EL, Murphy E, Stes E, Nikonorova N, Ramakrishna P, Czyzewicz N, Montero-Morales L, Kumpf R, Lin Z, van de Cotte B, Iqbal M, Van Bel M, Van De Slijke E, Meyer MR, Gadeyne A, Zipfel C, De Jaeger G, Van Montagu M, Van Damme D, Gevaert K, Rao AG, Beeckman T, De Smet I (2016) PP2A-3 interacts with ACR4 and regulates formative cell division in the Arabidopsis root. Proc Natl Acad Sci U S A 113:1447-1452

Zhang C, Bousquet A, Harris JM (2014a) Abscisic acid and lateral root organ defective/NUMEROUS INFECTIONS AND POLYPHENOLICS modulate root

elongation via reactive oxygen species in Medicago truncatula. Plant Physiol 166:644-658

Zhang W, Ciclitira P, Messing J (2014b) PacBio sequencing of gene families - a case study with wheat gluten genes. Gene 533:541-546

Zhao G, Zou C, Li K, Wang K, Li T, Gao L, Zhang X, Wang H, Yang Z, Liu X, Jiang W, Mao L, Kong X, Jiao Y, Jia J (2017) The Aegilops tauschii genome reveals multiple impacts of transposons. Nat Plants 3:946-955

Zhao QY, Wang Y, Kong YM, Luo D, Li X, Hao P (2011) Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics 12 Suppl 14:S2

Zhou Y, Honda M, Zhu H, Zhang Z, Guo X, Li T, Li Z, Peng X, Nakajima K, Duan L, Zhang X (2015) Spatiotemporal sequestration of miR165/166 by Arabidopsis Argonaute10 promotes shoot apical meristem maintenance. Cell Rep 10:1819-1827

Zhu H, Hu F, Wang R, Zhou X, Sze SH, Liou LW, Barefoot A, Dickman M, Zhang X (2011) Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. Cell 145:242-256

Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL (2017a) The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. Gigascience 6:1-7

Zimin AV, Puiu D, Luo MC, Zhu T, Koren S, Marcais G, Yorke JA, Dvorak J, Salzberg SL (2017b) Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res 27:787-792

Zou C, Wang Q, Lu C, Yang W, Zhang Y, Cheng H, Feng X, Prosper MA, Song G (2016) Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (Gossypium arboreum). Sci China Life Sci 59:164-171

FIGURE LEGENDS

Figure 1.1 A schematic representation of the root of Arabidopsis thaliana. The apex of the Arabidopsis root consists of four distinct zones of growth activities: the division zone or meristem, the transition zone, the elongation zone, the differentiation or mature zone. The figure has been reproduced from Ubeda-Tomás et al. (2012).

Figure 1.2 Schematic of Cell Types in the Arabidopsis Root Tip. (A) Transverse section. (B) Median longitudinal section. The figure has been reproduced from Van Norman et al. (2011).

FIGURES

Figure 1.1 A schematic representation of the root of Arabidopsis thaliana. The apex of the Arabidopsis root consists of four distinct zones of growth activities: the division zone or meristem, the transition zone, the elongation zone, the differentiation or mature zone. The figure has been reproduced from Ubeda-Tomás et al. (2012).

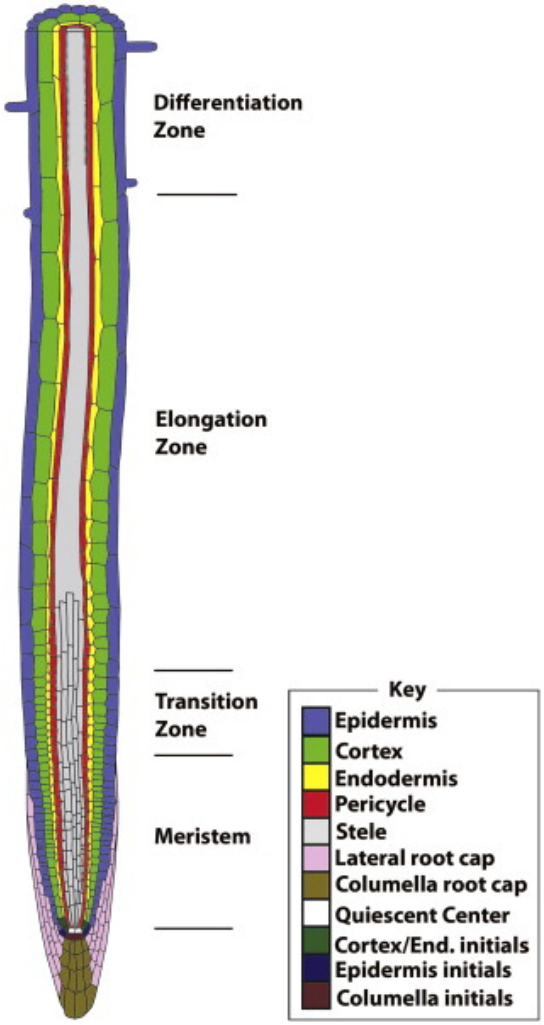
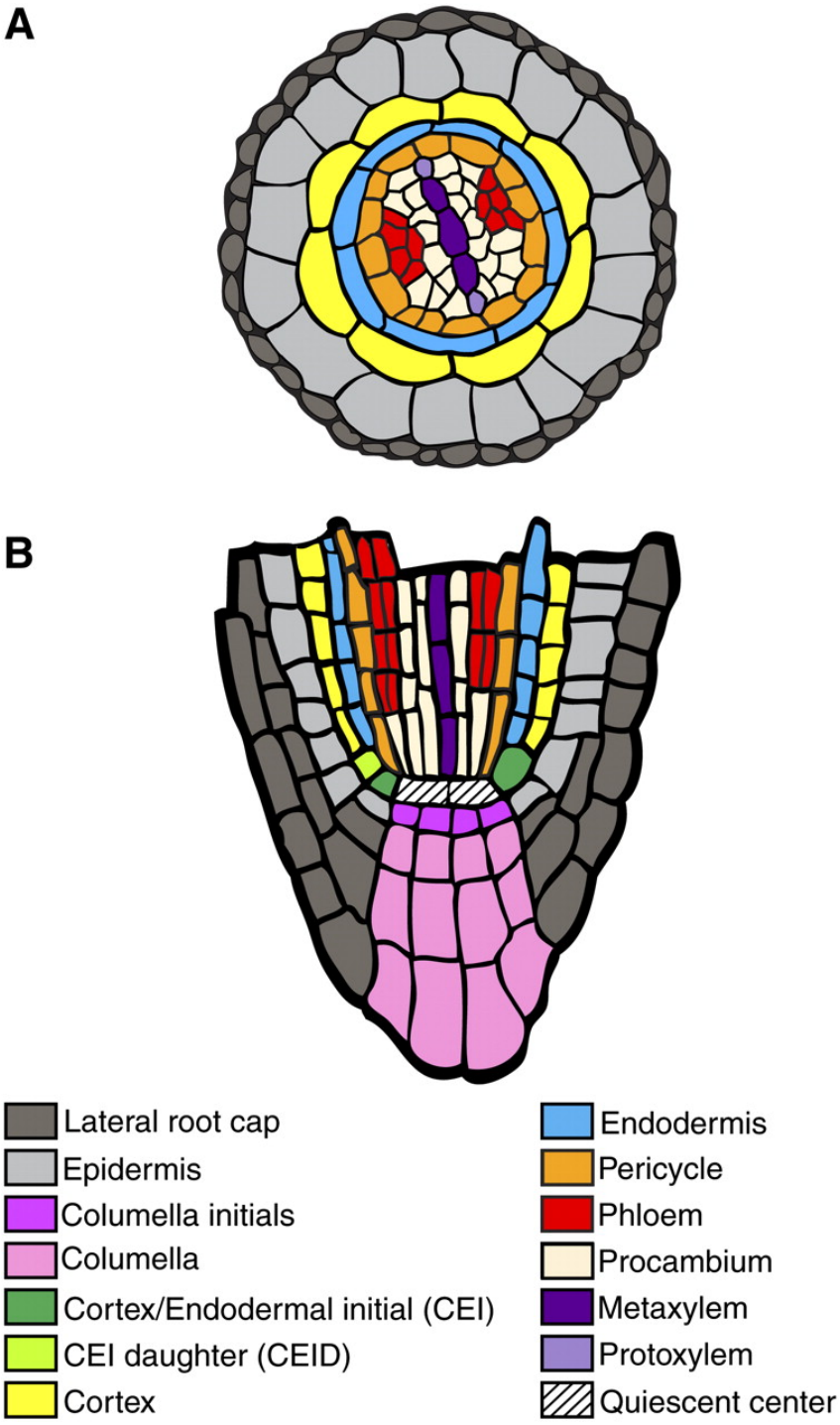


Figure 1.2 Schematic of Cell Types in the Arabidopsis Root Tip. (A) Transverse section. (B) Median longitudinal section. The figure has been reproduced from Van Norman et al. (2011).



CHAPTER 2

De Novo Assembly of Wheat Root Transcriptomes and Transcriptional Signature of Longitudinal Differentiation

ABSTRACT

Hidden underground, root systems constitute an important part of the plant for its development, nourishment and sensing the soil environment around it, but we know very little about its genetic regulation in crop plants like wheat. In the present study, we *de novo* assembled the root transcriptomes in reference cultivar Chinese Spring from RNA-Seq reads generated by the 454-GS-FLX and HiSeq platforms. The FLX reads were assembled into 24,986 transcripts with the completeness of 54.84%, and the HiSeq reads were assembled into 91,543 high-confidence protein-coding transcripts, 2,404 low-confidence protein-coding transcripts， and 13,181 non-coding transcripts with the completeness of >90%. Approximately 7% of the coding transcripts and ~2% non-coding transcripts are not present in the current wheat genome assembly. Functional annotation of both assemblies showed similar gene ontology patterns and that ~7% coding and >5% non-coding transcripts are root-specific. Transcription quantification identified 1,728 differentially expressed transcripts between root tips and maturation zone, and functional annotation of these transcripts captured a transcriptional signature of longitudinal development of wheat root. With the transcriptomic resources developed, this study provided the first view of wheat root transcriptome under different developmental zones and laid a foundation for molecular studies of wheat root development and growth using a reverse genetic approach.

Keywords: Differential expression, RNA-Seq, Root, Transcriptome assembly, Wheat

INTRODUCTION

As the "hidden half" of a plant, root systems provide plant water, nutrients, and an anchorage from the soil, produce growth regulators and sense soil environmental changes such as pH, moisture, and mineral content. A well-developed root system is critical for sustainable crop production. Despite the important roles in plant development and growth, our understanding of root development and growth is still very limited as compared to the aboveground half. Nevertheless, most knowledge of root biology comes from the model plant Arabidopsis. Rich genomic resources, non-soil cultivation and anatomical simplicity make the Arabidopsis root state of the art in plant biology at both molecular and cellular levels, including identification of many genes involving various aspects of root development, characterization of hormone interaction, cell type definition, and environmental responses (Petricka et al. 2012). Dicots and monocots differ significantly in root system architecture and cellular organization. Compared to the tap root system in dicots, monocot roots are fibrous with large quiescent centers, the separate origin of endodermis and cortex in ground tissue, multiple layers of cortical cells with variable cell numbers, and multiple-tissue occurrence of lateral roots (Hochholdinger and Zimmermann 2008). With a finished genome and relative ease of genetic transformation, rice has emerged a model for grass root biology study and provided a good amount of information (Coudert et al. 2010; Rebouillat et al. 2008). In contrast, very little information is available in the small-grain crops, including barley, oats, rye, and wheat, which grow in relatively dry conditions and have large genomes.

Common wheat or bread wheat (*Triticum aestivum* L., genomes AABBDD) is a hexaploid species of relatively recent origin and one of the most important food sources, providing ~20% daily caloric consumption. As the most widely adapted crop, wheat plays an important role in the global food security. Mainly due to climate change, however, wheat production is facing numerous challenges from biotic stress and abiotic stress. Understanding the molecular mechanisms underlying root development, growth and the environmental response is a prerequisite for improving tolerance to the soil-borne stress, such as drought and waterlogging, using biotech approaches. Functional genomics has long been expected to play an important role in wheat root studies. Of the ~1.3 million wheat expressed sequence tags (ESTs) from 147 complementary DNA (cDNA) libraries, 26,849 ESTs of 25 cDNA libraries were made from the root tissues of reference genotype Chinese Spring (CS; http://ncbi.nlm.nih.gov). But they are far apart from covering the root transcriptome, particularly for those transcripts that are low in abundance but important in function, such as transcription factors (TFs). Compared to the traditional EST development and microarray hybridization, RNA-Seq offers unprecedented capacity and resolution in revealing the landscape and dynamics of complex transcriptomes. As the sequencing cost continues to drop, RNA-Seq has been the favorite choice for transcriptome analysis of the non-model plant species (Strickler et al. 2012). Without finished genome sequences, the transcriptomes of the non-model species are assembled *de novo*. Although draft genome sequences of common wheat cultivar Chinese Spring (CS) (Brenchley et al. 2012; The International Wheat Genome Sequencing Consortium 2014) and the A-genome (Ling et al. 2013) and D-genome progenitors of wheat (Jia et al. 2013) were reported recently, their

utility in transcriptome analysis remains to be tested. In another aspect, a *de novo* transcriptome assembly will also benefit annotation of the wheat genome.

To gain a global view of the allelic interaction and its effect on the root transcriptome at large and to lay a foundation for root functional genomics, we initiated a wheat transcriptome project. As the first stage, we sequenced the three root RNA samples of CS using 454 GS-FLX (Roche, Branford, CT, USA) and HiSeq (Illumina, San Diego, CA, USA) platforms. Assembly and quantification of the wheat root transcriptomes provide the first view of the transcriptional landscape of wheat root development. Here we report the *de novo* assembly of the root transcriptomes, characterization of the assembled transcripts, expression profiling of the genes in the root tip and the mature part of the roots, and their implication to wheat root development and growth.

MATERIAL AND METHODS

Plant material and RNA extraction

RNA was extracted from the roots of CS seedlings in two germination experiments. In experiment 1, CS seeds were germinated in germination box on the tap water-wetted paper towels, and 3-mm root tips were harvested from the 3 days-old seedlings and frozen in liquid nitrogen. In experiment 2, the CS seeds were germinated in deep pots containing sands, and root tips of ~3mm (meristematic zone) and rest of the roots, mainly the maturation zone, were collected from seven-day-old seedlings separately, snap frozen in liquid nitrogen. Three biological replicates were included for each developmental zone. RNA was extracted using Trizol (Thermo Fisher Scientific, Waltham, MA) according to

the manufacturer's instruction. The RNA samples were purified using the RNeasy mini kit (Qiagen, Valencia, CA). Concentration and integrity of the purified RNA samples were quantified was confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA), and samples with an RNA integrity number (RIN) greater than eight were used in the subsequent analyses.

454 GS Titanium FLX sequencing

Messenger RNA (mRNA) was extracted from total RNA derived from the experiment 1 using a mRNA-only kit (Epicentre, Madison, WI, USA). The purified RNA was submitted to the Integrated Genomics Facility at Kansas State University, Manhattan, KS, for cDNA synthesis using random primers, for construction of a DNA sequencing library using a standard cDNA rapid library construction kit from 454/Roche and a sequencing run on a 454/Roche Titanium platform.

Illumina Sequencing

RNA samples extracted from root tips and rest of the root tissue (mainly maturation zone) from plants grown in experiment 2 were submitted the DNA Core Facility at University of Missouri, Columbia, MO, for cDNA synthesis, sequencing library construction and sequencing. Six barcoded sequencing libraries for three biological replicates for the meristematic zone and three biological replicates for maturation zone were prepared using the TruSeq RNA Library Prep Kit (Illumina). These six libraries were pooled and sequenced in one lane on the HiSeq 2000 platform (Illumina) to generate 100 bp single-end reads.

Quality control and preprocessing

Adapter sequences used during the library preparation were trimmed from the 454-GS-FLX reads using a Perl script from NGSQC toolkit (Patel and Jain 2012). Trimming of the HiSeq reads were performed using a Java-based program Trimmomatic (Lohse et al. 2012). The adapter-free reads were further filtered based on the quality using the prinseq (Schmieder and Edwards 2011). The parameters for quality trimming were set for a minimum mean quality of Q20 across the read and to trim low-quality bases at 3' end. The minimum read length of 100 bp for the FLX reads, and 50 bp for HiSeq reads was used as cutoffs for length filtering. For the FLX reads with homopolymer sequences were trimmed using a Perl script from the NGSQC toolkit (Patel and Jain 2012). The reads corresponding to rRNA sequences were filtered using Ribopicker Perl script (Schmieder et al. 2012) using a plant rRNA sequence dataset generated from the rDNA sequences retrieved from NCBI (http://www.ncbi.nlm.nih.gov), TAIR (https://www.arabidopsis.org) and the rice genome annotation database (http://rice.plantbiology.msu.edu).

***De novo* assembly of the transcriptome**

All the assembling work was done on a server with 24 cores and 128GB RAM or 64 cores and 512GB RAM. The clean reads obtained from the 454 sequencing were assembled using Newbler program v2.6 from Roche, TGICL v2.1 (http://sourceforge.net/projects/tgicl/) (Pertea et al. 2003) and MIRA v3.9.17 (http://mira-assembler.sourceforge.net) (Chevreux et al. 2004). The assembly with Newbler was carried out with six different overlap percentages of identity, i.e., 95 -100% keeping the number of bases in overlap constant as

80bp, and a read was only assigned to one contig. TGICL and MIRA assemblies were done using the 98 % identity over a stretch of 80 bp and keeping the rest of the parameters default.

The contigs and singletons from the Newbler 98% identity assembly were used to assemble with the 35,042 ESTs from 26 CS root-only libraries deposited in DFCI gene index, NCBI EST database (http://www.ncbi.nlm.nih.gov/genbank/dbest/dbest_access/) and Komugi wheat EST database (http://www.shigen.nig.ac.jp/wheat/komugi/ests/tissueBrowse.jsp). The hybrid assembly was carried out using CAP3 assembly program (Huang and Madan 1999) with a 98% identity across a minimum of 80-bp overlap.

The purged HiSeq reads were assembled using Velvet/Oases program version 1.0.14 (http://www.ebi.ac.uk/~zerbino/velvet/) (Schulz et al. 2012) with *k*-mer values 31, 41, 51, 61, 71, 81 to get a better assembly. The contigs from all the *k*-mer assemblies were clustered using CD-HIT-EST (Li and Godzik 2006) at 99% identity (–c 0.95 –n 8 –T 0 –M 0 –gap -2) to remove redundant contigs generated by different *k*-mers. To further extend the contigs, the non-redundant contigs from these multiple *k*-mer assemblies were assembled using TGICL program (Pertea et al. 2003) with 99% identity and 100-bp overlap.

Evaluation of assemblies

Both the FLX and the HiSeq reads used for the assemblies were mapped onto the corresponding assembled sequences using mapping tool in CLC Bio Genomic Workbench v6.0.1 (Qiagen, Carlsbad, CA) with parameters global alignment at 95% identity. The

quality of the assembly was evaluated by aligning the assembled contigs to the full-length (FL) cDNA sequences of wheat from TriFLDB database (Riken, Japan). The FL-cDNA sequences were downloaded from TriFLDB, and redundant sequences with an identity of 99% were removed using CD-HIT program (Li and Godzik 2006). Eventually, 17,094 non-redundant cDNA sequences were used for evaluating the completeness of our assembly. For evaluating the completeness of both the Newbler and Velvet assemblies, the program CEGMA was run on both the assemblies to determine the percent of the conserved core eukaryotic genes were assembled (Parra et al. 2007).

Prediction of open reading frames and coding potential

The root assemblies were aligned with eight proteomes from finished genomes, wheat protein sequences from TriFLDB, and barley protein sequences from TriFLDB and MIPS (http://pgsb.helmholtz-muenchen.de/plant/barley) databases using BLASTX algorithm. The sequences of the finished plant genomes, including those of Arabidopsis, rice, Brachypodium, sorghum, foxtail millet, maize, and switchgrass were retrieved from the Phytozome database (v11.0; https://phytozome.jgi.doe.gov/pz/portal.html). The BLASTX results were used to predict open reading frames (ORFs) by the findorf program (Krasileva et al. 2013). A second prediction was performed on the already predicted sequences by masking the first ORF to identify the misassembled transcripts that may arise during the *de novo* assembly. TransDecoder (https://github.com/TransDecoder/TransDecoder) was used to predict ORF from the leftover transcripts. The coding potential of the transcripts without predicted ORFs were analyzed using a potential coding calculator (CPC) with default setting using a webserver (http://cpc.cbi.pku.edu.cn/).

Functional annotation and GO assignment

The assembled transcripts were annotated by performing a BLASTX search against the NCBI non-redundant (nr) protein database with an E-value of 10E-6 and minimum coverage of 100bp or 33aa. Gene Ontology (GO) assignment was performed using Blast2go software (www.blast2go.com). The assembled transcripts were further aligned against the Wheat Unigene dataset build 60 (ftp://ftp.ncbi.nih.gov/repository/UniGene/Triticum_aestivum/) and against Arabidopsis, Rice and *Brachypodium* proteomes using command line BLASTX from NCBI v2.2.26 with an e-value of 10E-6. The transcripts were also searched against the Triticeae Repeat sequence database (TREP) to identify the transposable elements (TEs) in the wheat root transcriptome.

Separating homoeologous transcripts from the *de novo* assembled transcriptome

To separate the homoeologous transcripts, we used the pipeline reported by (Krasileva et al. 2013) using Freebayes (https://github.com/ekg/freebayes) and Hapcut programs (Bansal and Bafna 2008) to phase the reads based on the SNPs found in the homoeologous genes of wheat. The phased reads were assembled into contigs using a Perl script, which employs the MIRA assembler v3.4.1.1 (Chevreux et al. 2004) and CAP3 (Huang and Madan 1999).

Differential expression analysis in root tip and the mature root tissues

The HiSeq reads from both the root tip and maturation zone samples were mapped to the *de novo* assembled transcriptome using the read mapping tool in CLC Bio Genomic

Workbench v6.5.1 (Qiagen, Carlsbad, CA, USA) with parameters set as 95% identity along the length of the read. Multiple mapping of the reads is limited to ten. The transcript abundance was calculated in terms of reads per kilobase of the transcript per million (RPKM) and transformed by adding a constant "1" to avoid zero values. The transformed expression values were normalized by median scaling method across all the biological replicates of both the samples. The transcripts differentially expressed in both the tissues were identified with a fold change of at least two and a false discovery rate (FDR) *p*-value of 0.05. The normalization, statistical tests, and the p-value correction were done using the inbuilt tools in the CLC Bio Genomics Workbench. The differentially expressed genes were mapped to the MapMan bins using the Mercator tool (http://mapman.gabipd.org/web/guest/app/mercator) (Lohse et al. 2014) and were represented on the metabolic pathways (http://mapman.gabipd.org/web/guest/mapman; v3.6.0RC1) (Usadel et al. 2009).

RESULTS

Wheat root transcriptome datasets

We sequenced the transcriptome of the CS root tip using the 454 GS-FLX platform (Roche), which generated 1,086,240 raw reads from a single pyrosequencing run. As the evolution of sequencing technologies, we subsequently sequenced six libraries, three for the root tips and three from the rest of root tissues using HiSeq 2000 platform (Illumina), which generated 192,767,620 single-end sequence reads of 100 bp length. All these sequence-reads went through the processing pipeline for trimming adapters/primer sequences at the ends of the reads and low-quality bases at the 3' end of the reads and filtering all the reads with low quality (average Phred quality score of <20) and rRNA contamination. The quality filtering and removing rRNA contamination resulted in 808,117 (74.4%) FLX reads and 169,286,239 (87.82%) HiSeq reads of high quality (Figure 2.1 and Table 2.1).

De novo assembly of FLX reads and annotation of wheat root transcriptome

High quality reads from 454 sequencing were *de novo* assembled using Newbler software with different identity thresholds, from 95% through 100% of identity across 80 bp overlap to place two reads into a contig. The assemblies were analyzed for various parameters, including the number of reads used, the total number of contigs generated, number of contigs longer than 200bp, N50 length, longest contig length and average contig length, and mapped the reads back onto the assembled contigs to estimate the number of unmapped reads. A total of six assemblies were generated (Table 2.2). As expected, an increase of

sequence identity reduces N50, longest contig and average contig size, number of reads used and size of the assemblies, but increases the number of contigs and singletons. One exception to this is the largest contig length for the assembly with 97% identity, which is smaller than that of the assembly with the identity of 98%. The FLX reads were also assembled with other programs including MIRA and TGICL separately, and the quality of these assemblies was analyzed using the same output parameters used for the Newbler assemblies. The TGICL assembly generated more contigs (78,413) than any of the assemblies from Newbler or the MIRA. But the largest contig assembled and N50 was the smallest compared to the other assemblies. The assemblies generated by TGICL and MIRA are larger (50.3 and 52.43 Mbp) than the six assemblies generated by the Newbler. Although the Newbler assembly with 95 % identity has the largest contig size and N50, use of a lower identity would increase the probability of merging the homoeologous transcripts. With all the parameters considered, the Newbler assembly with the 98% identity is overall desirable (Table 2.2) and used for further analysis. The distribution of the size of transcripts assembled in this assembly was shown in figure 2.2.

To improve our assembly of the root transcriptome generated from the FLX reads, we performed a hybrid assembly using the 24,986 contigs from the Newbler assembly with 98% identity (Table 2.2) and 35,042 ESTs from the CS root. This merged 5,863 Newbler contigs with 11,940 ESTs into 4,812 CAP3 contigs. As a result, hybrid assembly reduced contig number from 24,986 to 23,935 and increased N50 from 815 to 887 and the longest contig size improved from 6699 bp to 6,747 bp. At the same time, 19,123 Newbler contigs and also 23,102 ESTs found no match, indicating that 454 sequencing expanded CS root transcriptome significantly, but its coverage is still low. This low coverage is confirmed by

the CEGMA assay, which showed the root transcriptome assembled from the FLX reads has a completeness of 54.84 % for full length conserved eukaryotic genes (CEGs) and 85.08 % for the partial CEGs.

Approximately, 87% of the transcripts had BLASTX hits in NCBI nr protein database, of which 78% of the total transcripts were assigned with GO terms and 18% were assigned with enzyme commission (EC) annotation. For biological processes, >70% of GO items fall in top five categories, i.e., organic substance metabolic process (7,227), primary metabolic process (7,224), cellular metabolic process (5,388); biosynthetic process (3,458) and nitrogen compound metabolic process (2,852). For molecular functions, the top five categories account for >80% of the total GO items, i.e., heterocyclic compound binding (5,726), organic cyclic compound binding (5,726), small molecule binding (3,727), transferase activity (3,269) and hydrolase activity (3177). For cellular localization, >90% of the GO items were from the top three groups: intracellular (11,533), membrane-bounded organelle (6,923) and membrane (3,686) localization (Figure S1.1; Table S1.1).

De novo assembly of HiSeq sequence reads

We *de novo* assembled the clean reads that were obtained from the Illumina sequencing using the velvet program, which assembles short reads using the De Bruijn graph, with six different *k*-mers. Multiple *k*-mer assemblies generated a total of 1,372,996 sequence contigs. Contig files from all the assemblies with *k*-mer lengths of 31, 41, 51, 61, 71 and 81 were concatenated, and the redundant contigs generated by different *k*-mer assemblies were clustered into the corresponding longest contigs using CD-HIT-EST. The concatenation resulted in 504,839 non-redundant sequences. These sequences were further

assembled again using TGICL program with an identity of 99% across a minimum overlap of 100 bp to extend the contigs and generated a final assembly of 148,984 transcripts, including 68,589 extended and merged contigs and 80,395 unextended sequences. After filtering the transcripts with a length of less than 200 bp, a total of 146,165 transcripts were assembled from the 169,282,312 quality HiSeq reads. We evaluated this assembly for various features. It has an N50 of 1,865 bp with the largest transcript of 21,400 bp and an assembly size of 210,848,484 bp. A run of the CEGMA program indicated that the root transcriptome assembled from the HiSeq reads had a completeness of 90.32% for full-length CEGs and 92.4 % for the partial CEGs.

Anatomy of wheat root transcriptome

Alignment against the Triticeae Repeat database found that 6,692 assembled transcripts originated from or containing repetitive DNA sequences were expressed in the root. These repeated sequences include 3,421 miniature inverted transposable elements (MITEs), 2,401 retrotransposons, 659 DNA transposons, 35 Helitron and 176 transposable elements of unknown classes (Figure 1.3). Also, 495 transcripts were found to contain repetitive sequences with transcript coverage of 90% or more, including LTRs, LINES, CACTA, Helitron and unknown classes of transposable elements. Compared to other TE species, MITEs are much smaller in size and mainly located in 3' untranslated regions (UTRs).

To predict the ORFs in the 142,894 non-TE transcripts, we performed a BLASTX run against various protein databases, and the BLASTX outputs were used with the findorf program to predict the coding sequences (cds) and protein sequences encoded by the transcripts. The program predicted ORFs in 116,833 transcript sequences, and the

remaining 26,061 sequences had no coding capacity. Of the 116,833 ORF-containing transcripts, 4,727 sequences had premature stop codons, and 18,000 sequences had frame-shifts in their ORFs, suggesting that these 22,727 sequences were transcribed from pseudogenes. For the 94,106 transcripts that contain normal ORFs, running an iterative step of findorf with the first ORF masked found that 6,158 sequences contained a second ORF, suggesting that they were derived from misassemblies during the *de novo* assembly process. Therefore, a total of 87,948 transcripts contain unique ORFs. Further annotation of the 26,061 transcripts, from which no ORFs were predicted by findorf, using outputs of BLASTX against NCBI nr database predicted ORFs in 9,987 transcripts. Of these 9,987 transcripts, 4,244 transcripts were found to be pseudogenes with a frameshift or a premature stop codon in the ORF. And an iterative run with the masked ORF sequence found 2,148 transcripts were containing a second ORF. Thus, 3,595 transcripts were identified with a functional ORF, increasing the total transcripts with a predicted functional ORF to 91,543. These transcripts were considered high-confidence (HC) protein-coding transcripts.

The findorf program didn't predict any ORF in the remaining 16,074 transcripts. Using TransDecoder (https://github.com/TransDecoder/TransDecoder), we identified only a single putatively functional ORFs in 2,404 of these 16,074 transcript sequences based on the pfam domain and the BLASTP hit against SWISSPROT database. These 2,404 transcripts are therefore considered low-confidence (LC) protein-coding transcripts.

The remaining 13,181 transcripts were left over without any predicted ORF present and further analyzed using the potential coding calculator (CPC). Of the 13,181 transcripts, 189 showed coding potential with the score ranging from 3.999 to 0.008, 12,705 showed

no coding potential with a potential coding score ranging from -0.008 to -1.572, and 287 transcripts had no results returned by CPC. Considering that LC proteins are not confirmed in other plant genomes and due to the very low CPC for the noncoding transcripts, they were pooled and referred as non-ORF transcripts hereafter.

We aligned the 91,543 ORF-containing transcripts and 16,074 non-ORF transcripts with the current version of the wheat genome assembly. The results showed that 58,341 (63.7%) ORF-transcripts found matches in whole genome sequence with >97% identity and >50% length coverage. A majority (51,610) of these ORF-transcripts had hits in the predicted gene models (Figure S1.2), 6,252 ORF-transcripts did not show any sequence similarity to the predicted cDNA sequences, and 536 showed no homology to the wheat genome assembly. Of the 16,074 non-ORF transcripts, 10,931 hit the whole genome sequences with the above criteria, and 360 did not show any match in the wheat genome assembly. Of the 10,931 matched non-ORF transcripts, 2,343 hit the predicted cDNA sequences with same parameters and remaining 8,588 only found matches in the wheat genome assembly but not in the predicted cDNA, suggesting that they are located either in the intergenic regions or introns. To further validate the 536 ORF-transcripts and 360 non ORF-transcripts that are not found in the IWGSC draft genome and gene build, we did a BLASTn search of these sequences against the 5x wheat genome sequences assembled using 454 sequencing platform (Brenchley et al. 2012). Only 20 HC protein-coding transcripts and 43 non-ORF transcripts were not found. All these indicate that the all the transcripts are present in the wheat genome, but the current wheat genome assembly and annotation is incomplete.

To gain insights into the organ specificity of the transcripts, we aligned the wheat root transcriptome assembly with the RNA-Seq reads from the aboveground tissues, i.e., leaf, stem, spike, and grain, of wheat plants, which are deposited in NCBI SRA database. Results showed that 6,222 (6.8%) of the 91,543 protein-coding transcripts and 834 (5.2%) of the 16,074 non-coding transcripts did not show significant similarity, indicating that they are root specific.

Common wheat is a hexaploid species containing the A, B, and D genomes. During the *de novo* assembly of the reads into transcripts, the reads corresponding to the homoeologous genes can be merged into a single transcript rather than into separate transcripts due to high sequence similarity between the homoeologous genes (Brenchley et al. 2012; The International Wheat Genome Sequencing Consortium 2014). In our assembly pipeline, we merged multiple *k*-mer assemblies, which reduced the redundancy in the assembled contigs. This strategy also merged homoeologs with high sequence similarity into one contig. With available assembly algorithms and *de novo* assembly programs, however, it is difficult to assemble highly similar sequences into separate contigs. Using the homoeolog separation pipeline (Krasileva et al. 2013), we identified a total of 13,664,029 polymorphic reads corresponding to the 34,506 of the 91,543 assembled transcripts with a predicted functional ORF. These reads were assembled into 115,692 homoeologous blocks using the phasing information provided by the hapcut program.

To gain an understanding of the sub-genome specific expression of the assembled root transcriptome, we pooled the chromosomes and the gene models in the draft genome into subsets of the A, B, and D genomes and aligned the ORF-transcripts and the non-ORF transcrits with them using the same parameters as above (Figure 2.4 and 2.5). The results

showed that 52,486 ORF-transcripts and 8,737 non-ORF transcripts had a hit in the genome and that 55,704 ORF-transcripts and 2,415 non-ORF transcripts had a hit and the cDNA. All these corroborate that the current assembly and annotation is incomplete for each sub-genome.

Functional annotation, classification, and comparative genomics

The assembled transcripts were annotated by aligning against the NCBI nr protein database. Out of the 91,543 *de novo* assembled transcripts predicted with a functional ORF, 86,477 (94.47%) transcripts have at least one hit in the nr database, and 5,066 (5.53%) transcripts with a predicted ORF don't have a hit in the database. GO terms were assigned based on the annotation of the nr database, and 71,031 (77.59%) transcripts were assigned to at least one GO term. For 15,446 (16.87%) transcripts, there is a hit in the nr database, but no GO term is assigned. For biological processes, the top five GO groups account for >75% of the GO-assigned transcripts. These include macromolecule metabolic (13,687), organic cyclic compound metabolic (10,005), cellular aromatic compound metabolic (9,995), heterocycle metabolic (9,979) and cellular nitrogen compound metabolic process (9,970) (Figure S2.3A and Table S2.2). For molecular functions, top three GO groups, nucleoside phosphate binding (14213), nucleic acid binding (9898) and transferase activity, transferring phosphorus-containing groups (6356) account >42% of total GO-assigned transcripts (Figure S2.3B and Table S2.2). The assembled root transcriptome has 6,594 transcripts coding for transcription factors (TFs) of 55 families. The C2H2 TF family is the largest with 1,409 members followed by Myb-HB-like (600), bHLH (517), HAP3/NF-YB (410) and AP2/EREBP (378) in the top five families (Figure 2.6 and Table S2.3).

To further investigate the similarity of the wheat root transcriptome with the finished and draft genomes of model plants and other crops, we aligned the root transcripts with proteins sequences from Arabidopsis, Brachypodium, rice, sorghum, maize, *Ae. tauschii*, and *T. urartu* from NCBI and protein sequences for wheat and barley from RIKEN and MIPS using BLASTX. The hits from each database are compared. In the finished genomes, 74,302 (81.16%) transcripts had a match in all the genomes while 30, 96, 156, 286, 1,182 transcripts were unique to Arabidopsis, sorghum, maize, rice, and Brachypodium, respectively. Whereas in the draft genomes and ESTs, only 50,210 (54.84%) had a match owing to the incompleteness of the genomes (Figure 2.7).

Differential expression analysis of root tip and the mature root tissues

The reads from the libraries corresponding to the root tip and the mature part of the root were mapped to assembled transcripts, and their abundance was quantified in these two tissues. Of the 107,617 transcripts assembled, a total of 1,728 transcripts were found differentially expressed between the root tip and mature root tissues according to a comparison of expression levels with fold change (FC) ≥ 2 and a false discovery rate (FDR) of ≤ 0.05. Out of these 1,728, 1083 transcripts were more abundant in root tips, and 645 transcripts were more abundant in the matured part of the root. A search of the NCBI nr database and the Arabidopsis TAIR database annotated 1,647 of the 1,728 differentially expressed transcripts (DETs). Remaining 81 transcripts had no annotation in both the databases, of which 27 transcripts have functional ORFs but do not have a match in the two databases used, whereas 54 were non-ORF transcripts, representing putative noncoding transcripts. Of the 27 transcripts containing ORFs but no annotation, 18 were

enriched in root tips and 9 in the mature root; and out of the 54 non-ORF transcripts, 25 were enriched in root tips and 29 in mature root tissue.

Of the 1,728 DETs in the root tips, 82 transcripts were without any predicted ORF and considered noncoding. Interestingly, 41 transcripts were up-regulated and 41 down-regulated. For 15 transcripts upregulated in root tips transdecoder predicted single putative functional ORF and for another six transcripts were predicted with more than one ORF. In the down-regulated transcripts, 11 transcripts were predicted with a single ORF, and three transcripts were predicted with more than one ORF.

We annotated the DETs by BLASTX against the protein databases and mapped them onto the metabolic pathways using MapMan. Full annotation of the DETs is listed in Table S2.4 and an overview of the metabolic pathways in which the differentially expressed genes in root tip and mature root were illustrated in Figure 2.8. Genes in several metabolic pathways showed consistent differential expression, including fatty acid (FA) metabolism, secondary metabolism, glycolysis and tricarboxylic acid (TCA) cycle, cell wall biosynthesis and degradation (Figure 2.8). A total of 248 DETs were represented on the overview pathway map (Figure 2.8). Of the 248 mapped transcripts, 51 were involved in the secondary metabolism, 43 in lipid metabolism, 38 in cell wall metabolism, 23 in amino acid metabolism, 20 in starch and sucrose metabolism, 20 in minor carbohydrate metabolism, 13 in glycolysis and TCA cycle and 15 in the mitochondrial electron transport pathway.

Root tips include apical meristem, which maintains the high activity of cell division. In agreement with this, a significant number of up-regulated transcripts in the root tips were involved in the protein synthesis. These transcripts encode the ribosomal subunit

proteins (93 transcripts), translation (52 transcripts), chromatin structural proteins like histone proteins (39 transcripts), RNA binding and splicing components (27 transcripts), transcription factors (25 transcripts), and transport (21 transcripts) (Table S2.4). Several metabolic pathways were up-regulated in root tips: TCA cycle and mitochondrial electron transport pathways, FA synthesis, terpene synthesis, and biosynthesis of aromatic amino acids Phe, Tyr and Trp. In contrast, mature root mainly functions in cell elongation, differentiation, the formation of root hairs and lateral roots and transportation of water and minerals. Accordingly, nine genes in the phenylpropanoid pathway for lignin biosynthesis were enriched in the mature root tissue in agreement with its function in water conduction. These include those encoding a phenyl ammonia lyase (PAL), a 4-hydroxycinnamoyl CoA ligase (4CL), a hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase (HCT), a cinnamoyl-CoA reductase (CCR), a Caffeate O-methyltransferase (COMT) and four 4CL-like proteins. Except for the *COMT*, expression of these genes was induced in mature root tissues. Closely related to phenylpropanoid pathway, expression of the flavonoid pathway genes was also increased in the mature root. Pathways for FA degradation and biosynthesis of polar uncharged amino acids, Ser, Gly, and Cys, were also up-regulated in root tips (Table S2.4).

Of the metabolites, carbohydrate metabolism regulates root development in numerous ways apart from providing energy and structural components, including gravitropism, osmotic adjustment, and sugars that often act as regulatory signals and are required for lateral root initiation. We found that 22 transcripts corresponding to the different enzymes in the starch and sucrose metabolism were differentially expressed in root tips and mature root tissues (Figure 2.8). Transcripts (TC039764, TC088166, and

TC088167) encoding for the AGPases, starch synthases and starch branching enzymes in the starch biosynthesis pathway were induced or up-regulated in the root tips. At the same time, transcripts encoding enzymes of starch degradation, such as starch D enzyme, starch phosphorylase, and heteroglycan glucosidase were induced in the root tips, indicating active starch metabolism in the root tip tissue. In another aspect, three transcripts (TC001776, TC071737, and TC110592) encoding sucrose synthase were induced in the mature root.

Phytohormones, particularly auxin, brassinosteroids (BRs), jasmonic acid (JA) and abscisic acid (ABA), regulate almost every aspect of root development and growth. Numerous transcripts encoding hormone biosynthetic enzymes and transporters were also differentially transcribed between root tips and the mature root portion. Five auxin-promoting transcripts, one encoding the auxin efflux carrier PINFORMED 2 (PIN2), similar to OsPIN2 of rice and AtPIN7 of Arabidopsis, and four coding for an auxin-inducible 5NG4/Nodulin21-like protein (TC144456) and O-fucosyltransferases, were up-regulated in the root tip compared to the mature root. In contrast, six auxin-suppressing transcripts, three encoding Aux/IAA proteins homologous to OsIAA2, OsIAA6, and OsIAA21 of rice, and three coding for indole-3-acetic acid (IAA)-amido synthase-like proteins, which prevent free IAA accumulation, were up-regulated in the mature root. Downstream in the auxin pathway, two transcripts, TC056398 and TC018213, encoding SMALL AUXIN UPREGULATED (SAUR) proteins were differentially expressed with the former induced in the root tip and the latter induced in the mature root. Three transcripts encoding ATP binding cassette subfamily B/multi-drug-resistance/P-glycoprotein (ABCB/MDR/PGP) were up-regulated in the root tips. These proteins were identified to

create auxin gradient together with other auxin influx carriers (Benkova and Hejatko 2009). In the BR biosynthesis pathway, four transcripts, one coding for a secologanin synthase, two for cycloartenol synthases, and one for the DWF1 protein, which is involved in the conversion of early brassinosteroid precursor 24-methylenecholesterol to campesterol (Choe et al. 1999), were up-regulated in the root tips. These results suggest that a higher auxin and BR level is maintained in root tips compared to the matured zone. In the JA signaling pathway, three transcripts encoding the sulfotransferases similar to AtST2A, a protein involved in the reduction of the endogenous levels of 12-OH-JA (a by-product of switching off JA signaling) (Wasternack 2007), were up-regulated in the mature root, suggesting an opposite pattern for JA as compared to auxin and BRs. A complicated scenario was observed for ABA biosynthetic pathway. Three transcripts homologous to Arabidopsis ABA DEFICIENT 2 (ABA2)/SHORT-CHAIN DEHYDROGENASE/REDUCTASE 1 (SDR1) and one homologous to aldehyde oxidase 2 (AAO2), a putative ABA aldehyde oxidase that may be functional in the last step of ABA biosynthesis (Kataoka et al. 2004), were induced in the mature root. Two transcripts coding for TETRATRICOPEPTIDE-REPEAT THIOREDOXIN-LIKE 1 (TTL1) were up-regulated in root tips. TTL1 in Arabidopsis is required for elongation and organization of the root meristem and is involved in ABA signaling (Rosado et al. 2006). Two transcripts encoding for cytokinin receptor HISTIDINE KINASE 3 were induced in the mature root.

Transcription factors (TFs) are important regulators of gene expression. Expression of 112 transcripts encoding TFs of 21 families was altered in wheat root along the longitudinal axis. The major classes include AP2, bHLH, bZIP, MYB and MYB-related, homeodomain (HD), NAC families, and numbers and expression patterns of these TF

transcripts are shown in Figure 2.9. Notably, all 38 members of nine TF families, including three members of the GRAS family and 28 members of the NF-YB family, were induced in the root tips. By contrast, all 21 members of five TF families, including 12 members of the NAC family, four members of the HD family, were only induced in the mature root tissue. For the remaining seven TF families, such as the MYB family, 28 members were up-regulated, and 25 members were down-regulated in root tips (Figure 2.10). Several differentially expressed TFs are homologous to the known genes functioning in root development in the model plants, including two members of the STY-LRP1 family upregulated in the mature root tissue, suggesting their involvement in lateral root development. Of the four members of the AP2 family that up-regulated in root tips, three are homologous to AINTEGUMENTA-like 5 of *Ae. taushcii* (AIL5; EMT02119) and another homologous to BABY BOOM 2 (BBM2; EMS64473) of *T. urartu*. Two transcripts encoding for the ARFs homologous to AUXIN RESPONSE FACTOR 6 Arabidopsis thaliana (AtARF6) were up-regulated in root tip, and another transcript encoding for ARF homologous to AtARF11 was induced in mature root part. One transcript (TC084552) encoding the Argonaute family member homologous to AtAGO4 that is associated with 24-nt small RNA and involved in RNA dependent DNA methylation (Zilberman et al. 2003) was induced in root tips.

DISCUSSION

Growing and functioning underground complicates root studies by using traditional approaches, leaving a gap in our understanding of wheat development and growth.

Transcriptome analysis by RNA-Seq technology promises new opportunities for studying root development. RNA-Seq technology has been used to characterize the response of wheat root transcriptome to phosphate starvation (Oono et al. 2013) and infection of *Gaeumannomyces graminis* var. *tritici*, a pathogen of take-all root rot disease (Yang et al. 2015), but a reference transcriptome of wheat root and developmental expression pattern are not available. The present study developed and characterized a *de novo* assembly of wheat root transcriptome containing 94,106 transcripts that contain unique ORFs and identified 1,728 differentially expressed transcripts between the root tip and mature root tissues. All this will provide a global view of wheat root transcriptome and start point for a molecular understanding of root development and improving soil-related stress tolerance in a reverse genetics approach.

Root transcriptome assemblies

We assembled the FLX reads into a transcriptome of 19,123 Newbler contigs with >50% completeness and the HiSeq reads into a transcriptome of 146,165 transcripts with >90% completeness. For the FLX reads, the Newbler assemblies performed better overall on the statistics metrics than TGICL and Mira. Compared to the recently reported transcriptome assemblies of wheat (Cantu et al. 2011), barley (Bedada et al. 2014), *Persea Americana* (Reeksting et al. 2014) and smooth cordgrass (Bedre et al. 2016), our Newbler assembly showed comparable or even better statistic metrics including N50 value and percentage of assembled reads. Compared to the Newbler assembly of the pyrosequencing reads, the assembly of the HiSeq reads had a much greater N50 value, assembly size, and completeness mainly due to the large read number. A total of 1,749 transcripts from the

HiSeq assembly found matches in the wheat genome sequences but did not get hits in the publicly available RNA-Seq reads from the wheat roots. This discrepancy is mainly due to the enrichment of them in root tips by separation of root tips from the rest of root in the present study. All these corroborate sound quality and high content of information of the HiSeq assembly of the wheat root transcriptome.

Common wheat is a hexaploid species with A, B, and D genomes and a total of 94,000 to 120,000 protein-coding genes (Brenchley et al. 2012; The International Wheat Genome Sequencing Consortium 2014). Of the 91,543 transcripts, 34,506 were separated into 115,692 homoeologous blocks. If each of these 34,506 transcripts was derived from merging of at least two homoeologous transcripts, the total number of transcripts in the root assembly would be >126,049, excessing the total gene number, implying the existence of isoforms of transcripts due to alternative splicing, which is enhanced in polyploid wheat (Akhunov et al. 2013). In another aspect, 6.8% protein-coding transcripts did not find a match in the current assembly of the wheat genome, indicating the incompleteness of wheat genome assembly. In these respects, the wheat root transcriptome assemblies from this research can be used for improving wheat genome assembly and annotation.

Of the 146,165 transcripts in the final assembly of the HiSeq reads, 91,543 transcripts contain predicted functional ORFs, and 13,181 transcripts have no coding capacity and do not show homology to degenerated TEs, suggesting that they were transcribed as polyadenylated long non-coding RNAs (lncRNAs). LncRNAs condition gene expression in plants by regulating histone modification, transcription machinery, RNA processing machinery and posttranscriptional (Liu et al. 2015a). The 13,181 lncRNA transcripts, particularly the 55 lncRNA transcripts differentially expressed between root tip

and mature root, are an important resource for studying lncRNA regulation of root development.

Gene expression and root development

Although root has a much simpler anatomical structure as compared to the shoot and flower, it grows in a very different environment, underground, implying the existence of root-specific expression patterns including a set of root-specific genes. We found that 6.8% of the protein-coding genes are specifically expressed in root, not in the aboveground portion of wheat plants. Further characterization of these root-specific genes using reverse genetics approaches will shed new light on root development.

Current assembly of wheat root transcriptome contains 91,543 HC protein-coding transcripts and 16,074 non-ORF transcripts, but only a small fraction of the transcriptome, 1.17%, was differentially expressed in the root tip and mature root tissues, similar to the result obtained in rice (Kyndt et al. 2012). In rice, 1,761 of the 2,067 DETs showed higher transcription level in the mature root tissue (Kyndt et al. 2012). Opposite to the finding in rice, 1,083 of 1,728 wheat DETs were up-regulated or induced in the root tips.

Root tip and mature root tissues differ in several functional aspects, and these differences are reflected at the transcriptome level. First of all, root tips contain apical meristem for maintaining cell division capacity. Consistent with this, several TFs for maintaining meristem indeterminacy, such as GRAS TFs homologous to AtHAM2 and AtHAM3 of Arabidopsis (Engstrom et al. 2011) and AP2 TFs homologous to AIL5 (Nole-Wilson et al. 2005) and BABY BOOM (Galinha et al. 2007), were up-regulated in root tips. Besides, numerous genes related auxin transport and response are up-regulated in root tips and auxin catabolic, and auxin signal suppressor genes were down-regulated in root

tips. BR is critical in the regulation of cell expansion (Jaillais and Vert 2016), and increased expression of three BR biosynthetic genes in root tips was probably due to the partial inclusion of elongation zone in the root tip samples. Another important function of root tips is to percept gravitropism, which is achieved through starch statoliths (Fitzelle and Kiss 2001). In agreement with this function, transcription of 19 starch metabolic genes was up-regulated in root tips (Figure 2.8). In another aspect, the matured root part mainly functions in transporting water and minerals, which is achieved by development of lateral roots, root hairs, and vascular system. For lateral root development, four lateral root-promoting TF genes including two LRP1 (Smith and Fedoroff 1995), a KUODA1 (Lu et al. 2014), and an AtNAC1 homolog, were up-regulated in the mature zone, and an AtMBY93 of Arabidopsis, a negative regulator of lateral root (Gibbs and Coates 2014), was down-regulated in the mature zone. Increased expression of sucrose synthase in the mature zone may also be related to lateral root development as seen in soybean (Liu et al. 2015b). Another difference of mature zone from root tips lies in the differentiation of vascular bundles. In this respect, nine lignin biosynthetic genes and a homolog of *SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN 2*, encoding a NAC TF activating the lignin biosynthetic genes (Hussey et al. 2011), were up-regulated in the mature root portion.

Development of the root transcriptome assembly and identification of the DETs lay a foundation for molecular studies of wheat root biology and for improving soil-borne stress tolerance. In this respect, the recent development of sequence-cataloged TILLING libraries (Krasileva et al. 2017) will be very helpful in validating the function of DETs and homologs of root regulators identified in the model plant Arabidopsis and rice. Genome

editing technologies also can be used for targeting the candidate genes in wheat for functional validation (Wang et al. 2014).

In summary, we assembled a wheat root transcriptome containing 91,543 protein-coding and 16,074 non-ORF transcripts, 6.8% and 5.2% of which, respectively, are root specific. Approximate 6.8% of coding transcripts and ~2.2% of non-ORF transcripts were not found in the current wheat genome assembly. We also identified 1,728 transcripts differentially transcribed in root tip and mature root tissues. Annotation of these DETs provides a blueprint of molecular regulation of wheat root development. Thus, they are important candidates for in-depth analysis of wheat root development by TILLING, genome editing or other reverse genetics approaches.

Data availability. The raw FLX reads, and HiSeq reads are deposited in the sequence read archive at the National Center for Biotechnology Information under the bioproject id PRJNA419079. The Newbler assembly of the FLX reads, and the final assembly of the HiSeq reads are deposited in GrainGenes database.

REFERENCES

Akhunov ED, Sehgal S, Liang H, Wang S, Akhunova AR, Kaur G, Li W, Forrest KL, See D, Simkova H, Ma Y, Hayden MJ, Luo M, Faris JD, Dolezel J, Gill BS (2013) Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. Plant Physiol 161:252-265

Bansal V, Bafna V (2008) HapCUT: an efficient and accurate algorithm for the haplotype assembly problem. Bioinformatics 24:i153-159

Bedada G, Westerbergh A, Müller T, Galkin E, Bdolach E, Moshelion M, Fridman E, Schmid KJ (2014) Transcriptome sequencing of two wild barley (*Hordeum spontaneum* L.) ecotypes differentially adapted to drought stress reveals ecotype-specific transcripts. BMC Genomics 15:995

Bedre R, Mangu VR, Srivastava S, Sanchez LE, Baisakh N (2016) Transcriptome analysis of smooth cordgrass (*Spartina alterniflor*a Loisel), a monocot halophyte, reveals candidate genes involved in its adaptation to salinity. BMC Genomics 17:657

Benkova E, Hejatko J (2009) Hormone interactions at the root apical meristem. Plant Mol Biol 69:383-396

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705-710

Cantu D, Pearce SP, Distelfeld A, Christiansen MW, Uauy C, Akhunov E, Fahima T, Dubcovsky J (2011) Effect of the down-regulation of the high Grain Protein Content (GPC) genes on the wheat transcriptome during monocarpic senescence. BMC Genomics 12:492

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14:1147-1159

Choe S, Dilkes BP, Gregory BD, Ross AS, Yuan H, Noguchi T, Fujioka S, Takatsuto S, Tanaka A, Yoshida S, Tax FE, Feldmann KA (1999) The Arabidopsis dwarf1 mutant is defective in the conversion of 24-methylenecholesterol to campesterol in brassinosteroid biosynthesis. Plant Physiol 119:897-907

Coudert Y, Perin C, Courtois B, Khong NG, Gantet P (2010) Genetic control of root development in rice, the model cereal. Trends Plant Sci 15:219-226

Engstrom EM, Andersen CM, Gumulak-Smith J, Hu JO, E., Sozzani R, Bowman JL (2011) Arabidopsis homologs of the petunia hairy meristem gene are required for maintenance of shoot and root indeterminacy. Plant Physiol 155:735-750

Fitzelle K, J., Kiss JZ (2001) Restoration of gravitropic sensitivity in starch-deficient mutants of Arabidopsis by hypergravity. J Exp Bot 52:265-275

Galinha C, Hofhuis H, Luijten M, Willemsen V, Blilou I, Heidstra R, Scheres B (2007) PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. Nature 449:1053-1057

Gibbs DJ, Coates JC (2014) AtMYB93 is an endodermis-specific transcriptional regulator of lateral root development in arabidopsis. Plant Signal Behav 9:e970406

Hochholdinger F, Zimmermann R (2008) Conserved and diverse mechanisms in root development. Curr Opin Plant Biol 11:70-74

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

Hussey SG, Mizrachi E, Spokevicius AV, Bossinger G, Berger DK, Myburg AA (2011) SND2, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in Arabidopsis fibres and increases fibre cell area in Eucalyptus. BMC Plant Biol 11:173

Jaillais Y, Vert G (2016) Brassinosteroid signaling and BRI1 dynamics went underground. Curr Opin Plant Biol 33:92-100

Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R, Zhang C, Ma Y, Gao L, Gao C, Spannagl M, Mayer KF, Li D, Pan S, Zheng F, Hu Q, Xia X, Li J, Liang Q, Chen J, Wicker T, Gou C, Kuang H, He G, Luo Y, Keller B, Xia Q, Lu P, Wang J, Zou H, Zhang R, Xu J, Gao J, Middleton C, Quan Z, Liu G, Wang J, Consortium. IWGS, Yang H, Liu X, He Z, Mao L, Wang J (2013) Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91-95

Kataoka T, Hayashi N, Yamaya T, Takahashi H (2004) Root-to-shoot transport of sulfate in Arabidopsis: evidence for the role of SULTR3;5 as a component of low-affinity sulfate transport system in the root vasculature. Plant Physiol 136:4198–4204

Krasileva KV, Buffalo V, Bailey P, Pearce S, Ayling S, Tabbita F, Soria M, Wang S, Consortium I, Akhunov E, Uauy C, Dubcovsky J (2013) Separating homeologs by phasing in the tetraploid wheat transcriptome. Genome Biol 14:R66

Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, Simmonds J, Ramirez-Gonzalez RH, Wang X, Borrill P, Fosker C, Ayling S, Phillips AL, Uauy C, Dubcovsky J (2017) Uncovering hidden variation in polyploid wheat. Proc Natl Acad Sci U S A 114:E913-E921

Kyndt T, Denil S, Haegeman A, Trooskens G, De Meyer T, Van Criekinge W, Gheysen G (2012) Transcriptome analysis of rice mature root tissue and root tips in early development by massive parallel sequencing. J Exp Bot 63:2141-2157

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659

Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo M, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87-90

Liu J, Wang H, Chua NH (2015a) Long noncoding RNA transcriptome of plants. Plant Biotechnol J 13:319-328

Liu W, Han X, Zhan G, Zhao Z, Feng Y, Wu C (2015b) A novel sucrose-regulatory mads-box transcription factor gmnmhc5 promotes root development and nodulation in soybean (*Glycine max* [L.] Merr.). Int J Mol Sci 16:20657-20673

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res 40(Web Server issue):W622-627

Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B (2014) Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ 37:1250-1258

Lu D, T. W, Persson S, Mueller-Roeber B, Schippers JH (2014) Transcriptional control of ROS homeostasis by KUODA1 regulates cell expansion during leaf development. Nat Commun 5:3767

Nole-Wilson S, Tranby TL, Krizek BA (2005) AINTEGUMENTA-like (AIL) genes are expressed in young tissues and may specify meristematic or division-competent states. Plant Mol Biol 57:613-628

Oono Y, Kobayashi F, Kawahara Y, Yazawa T, Handa H, Itoh T, Matsumoto T (2013) Characterisation of the wheat (*Triticum aestivum* L.) transcriptome by de novo assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. BMC Genomics 14:77

Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061-1067

Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7:e30619

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2003) TIGR Gene Indices clustering

tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651-652

Petricka JJ, Winter CM, Benfey PN (2012) Control of Arabidopsis root development. Annu Rev Plant Biol 63:563-590

Rebouillat J, Dievart A, Verdeil JL, Escoute J, Giese G, Breitler JC, Gantet P, Espeout S, Guiderdoni E, Périn C (2008) Molecular Genetics of Rice Root Development. Rice 2:15-34

Reeksting BJC, N., Mahomed W, Engelbrecht J, van den Berg N (2014) De novo sequencing, assembly, and analysis of the root transcriptome of Persea americana (Mill.) in response to Phytophthora cinnamomi and flooding. PLoS One 9:e86399

Rosado A, Schapire AL, Bressan RA, Harfouche AL, Hasegawa PM, Valpuesta V, Botella MA (2006) The Arabidopsis tetratricopeptide repeat-containing protein TTL1 is required for osmotic stress responses and abscisic acid sensitivity. Plant Physiol 142:1113–1126

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863-864

Schmieder R, Lim YW, Edwards R (2012) Identification and removal of ribosomal RNA sequences from metatranscriptomes. Bioinformatics 28:433-435

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086-1092

Smith DL, Fedoroff NV (1995) LRP1, a gene expressed in lateral and adventitious root primordia of arabidopsis. Plant Cell 7:735-745

Strickler SR, Bombarely A, Mueller LA (2012) Designing a transcriptome next-generation sequencing project for a nonmodel plant species. Am J Bot 99:257-266

The International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345

Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg AS, M. (2009) A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. Plant Cell Environ 32:1211-1229

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nature Biotechnology 32:947-951

Wasternack C (2007) Jasmonates: An update on biosynthesis, signal transduction and action in plant stress response, growth and developmen. Ann Bot 100:681-697

Yang L, Xie L, Xue B, Goodwin PH, Quan X, Zheng C, Liu T, Lei Z, Yang X, Chao Y, Wu C (2015) Comparative transcriptome profiling of the early infection of wheat roots by *Gaeumannomyces graminis* var. *tritici*. PLoS One 10:e0120691

Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science 299:716-719

Table 2.1 Quality control and filtering of reads from 454 and Hiseq sequencing

Sequencing runs	Total Raw Reads	Quality Filter	Contaminants	High-quality reads
454 Reads (root tips)	1,086,240	196,306	81,817	808,117 (74.4%)
Root tip Replicate 1	31,803,479	4,040,700	382,197	27,380,582 (86.09%)
Root tip Replicate 2	34,133,444	3,794,965	419,609	29,918,870 (87.65%)
Root tip Replicate 3	34,873,477	3,632,386	454,067	30,787,024 (88.28%)
Mature root replicate 1	25,348,308	2,801,923	233,118	22,313,267 (88.03%)
Mature root replicate 2	37,575,294	3,886,507	379,931	33,308,856 (88.65%)
Mature root replicate 3	29,033,618	3,166,220	293,685	25,573,713 (88.08%)

Table 2.2 *De novo* assemblies of the FLX reads using Newbler, TGICL and Mira

Assembly parameters	95% identity	96% identity	97% identity	98% identity	99% identity	100% identity	TGICL	Mira
Total contigs (>200bp)	23,418	23,497	24,140	24,986	25,548	22,544	78,413	73,084
Avg contig size	749.87	742.22	722.04	696.92	675.54	660.86	641.53	717.46
N50 (bp)	905	893	857	815	787	737	672	762
Large Contigs >500bp	14,720	14,640	14,866	15,122	15,400	14,314	52,763	53,271
% Large Contigs	62.86	62.31	61.58	60.52	60.28	63.49	67.29	72.89
Large contigs N50	1,046	1,033	994	951	912	835	747	828
Largest Contig size (bp)	7,528	6,892	5,977	6,699	5,598	3,787	3,451	5,199
Assembly size (bp)	17,560,564	17,439,938	17,429,939	17,413,219	17,258,803	14,898,493	50,304,196	52,434,582
Reads used in assembly	697,311	691,259	681,502	662,618	624,961	518,627	686,622	506,290
% Assembled reads	86.31	85.56	84.35	82.02	77.35	64.19	84.99	62.67
Singletons	95,262	100,789	109,525	125,932	159,474	234,805	121,295	110,157
% Singletons	11.79	12.48	13.56	15.59	19.74	29.06	15.01	13.63

Table 2.3 Assembly statistics for the Newbler, TGICL and Mira assemblies with 98% identity

Parameters	Newbler	TGICL	Mira
Total Sequences (contigs+singletons) (>200bp)	122,086	181,533	155,628
Avg contig size	443.45	487.93	530.41
N50 (bp)	450	493	546
Largest Contig size (bp)	6,699	3,451	5,199
Large Contigs >500bp	22,047	58,467	57,927
% Large Contigs	18.06	32.21	37.22
Large contigs N50	810	725	808
Assembly size (Mbp)	54.14	88.58	82.55

Table 2.4 Hybrid assembly details

Input:	
Newbler contigs	30,047
454 singletons	125,932
Sanger ESTs	35,042
Output:	
Assembly size (>200bp)	49.45 Mbp
Total CAP3 contigs	43,109
Extended Newbler or new contigs	24,149
Newbler only contigs	18,960
454 singletons	58,020
N50	489 bp
Average contig size	490 bp
Largest contig size	6,747 bp

Figure legends

Figure 2.1 A flowchart of the assembly and annotation strategy for root transcriptome. Parameters for assemblies by Velvet/Oases, CD-HIT and TGICL-CAP3 are indicated in the parentheses.

Figure 2.2 Distribution of transcript length of the Newbler assembly of the root transcriptome. Numbers in the X-axis indicate the length of the transcript in bp, and numbers in the Y-axis indicate the quantity of the transcripts.

Figure 2.3 Transposable elements expressed in root transcriptome. The pie chart presents the different classes of the transposable elements expressed in the root tissues. The numbers after the transposon class are the number of transcripts in each class expressed in the root transcriptome. DNA_unknown, unknown DNA transposons; MITE, miniature inverted-repeat transposable elements; LINE, long interspersed elements; and SINE, short interspersed elements.

Figure 2.4 Venn diagram showing the distribution of protein-coding and non-ORF transcript nucleotide sequence alignment with the IWGSC draft genome sequences separated into sub-genomes. A) Distribution of the alignment of protein-coding transcripts with the sub-genome separated chromosome sequences. B) Distribution of the alignment of non-ORF transcripts with the sub-genome separated chromosome sequences. The sub-

genomes A, B, and D are represented by color, and the numbers within the circles indicate the number of transcripts aligned in each of the sub-genomes.

Figure 2.5 Venn diagram showing the distribution of protein-coding and non-ORF transcript nucleotide sequence alignment with the IWGSC cDNA sequences separated into sub-genomes. A) Distribution of the alignment of protein-coding transcripts with the sub-genome separated cDNA sequences. B) Distribution of the alignment of non-ORF transcripts with the sub-genome separated cDNA sequences. The sub-genomes A, B, and D are represented by color, and the numbers within the circles indicate the number of transcripts aligned in each of the sub-genomes.

Figure 2.6 Transcription factor (TF) families expressed in both the root tissues used in the study. Numbers of transcripts in each TF family are indicated on the X-axis, and names of the TF families are indicated on the Y-axis.

Figure 2.7 Venn diagrams showing the similarity of wheat root transcriptome with finished and draf genomes of model and crop plants. A) Comparision of the root transcripts against the protein sequences of finished genomes. B) Comparision of root transcripts against the protein sequences of draft genomes and assembled ESTs. Arabidopsis (TAIR v10); Rice (RGAP v 7); Brachypodium (Pyhtozome, Bd192); Sorghum (Phytozome, Sb79); Maize (Phytozome, Zm181); *Ae. Taushii* (Jia et al., 2012); Urartu (*Triticum urartu*) (Ling et al., 2013); Wheat_RIKEN_ESTs and Barley_RIKEN – translated protein sequences from assembled ESTs at RIKEN (http://trifldb.psc.riken.jp/v3/index.pl); Barley_MIPS – protein

sequences from barley genome from MIPS (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/).

Figure 2.8 An overview of the differentially expressed transcripts mapped onto the metabolic pathways. The differentially expressed genes were mapped onto the metabolic pathways using MAPMAN software. Each box represents a transcript, and the red colored ones are the up-regulated in the mature root tissues, and the blue colored ones are induced in the root tips. A fold change scale is indicated in the lower right corner.

Figure 2.9 Differentially expressed genes in root tips and mature root involved in the starch biosynthesis. The transcripts encoding for the enzymes involved in the starch and sucrose metabolism were represented each by a box. The blue colored are induced in the root tips and the mature root tissue. A fold change scale is indicated in the upper right corner.

Figure 2.10 Transcription factor (TF) families differentially expressed in root tip and the mature root tissues. Numbers of transcripts in each TF family are indicated on the X-axis, and names of the TF families are indicated on the Y-axis. The striped bars are the transcripts induced in mature root and the solid black bars in the root tips.

Supplementary figure legends

Figure S2.1 Gene Ontology (GO) classification of the de novo assembled 454 contigs. (A) Biological processes, (B) Molecular functions, and (C) Subcellular localization. The GO categories are indicated on the X axis, and the number of transcripts in each category is indicated on the Y axis.

Figure S2.2 Alignment statistics of the de novo assembled root transcriptome against the genomic and the predicted cDNA sequences from the hexaploid wheat draft genome. (A) A clustered and stacked bar chart showing the coverage percent and the identity percent of the root transcripts with predicted ORFs against the genomic and the DNA sequences of the draft genome. (B) A clustered and stacked bar chart showing the coverage percent and the identity percent of the root transcripts with without predicted ORFs against the genomic and the DNA sequences of the draft genome. The bars are stacked by the percent identity (represented by color) of the alignment and are clustered by the percent of the query covered in the alignment (shown by the labels on the top of each cluster). The databases against which the transcripts were compared are indicated on the X-axis. The number of transcripts in each bin are indicated on Y-axis.

Figure S2.3 Gene Ontology (GO) classification of the Root transcripts with predicted ORFs. (A) Biological processes (B) Molecular functions (C) Subcellular localization. The GO categories are indicated on the X-axis, and the number of transcripts in each category is indicated on Y-axis.

FIGURES

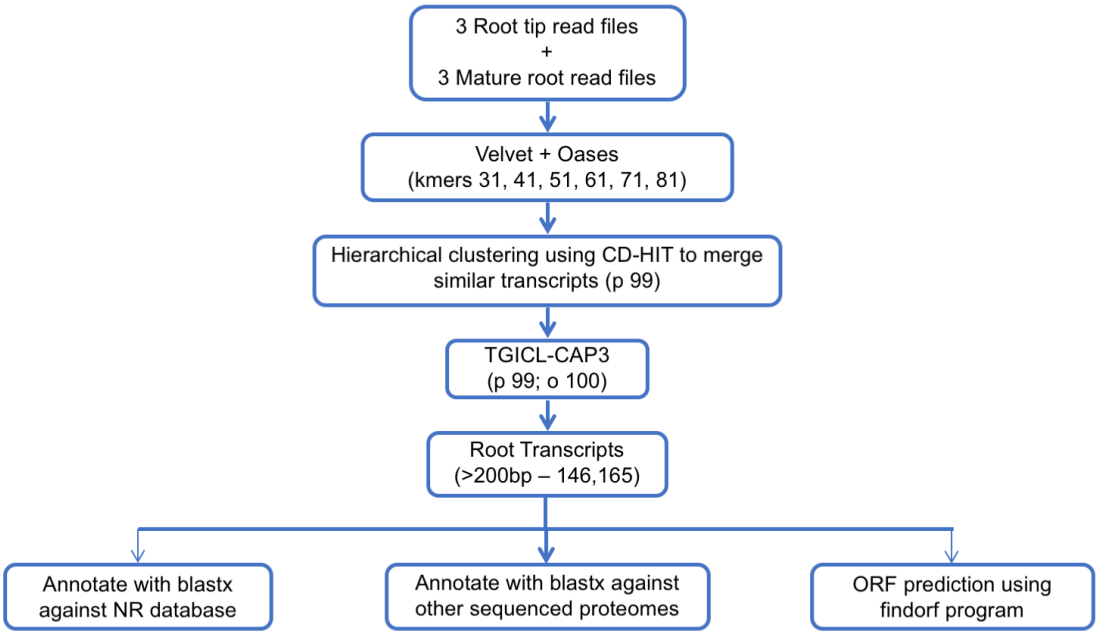


Figure 2.1 A flowchart of the assembly and annotation strategy for root transcriptome. Parameters for assemblies by Velvet/Oases, CD-HIT, and TGICL-CAP3 are indicated in the parentheses.

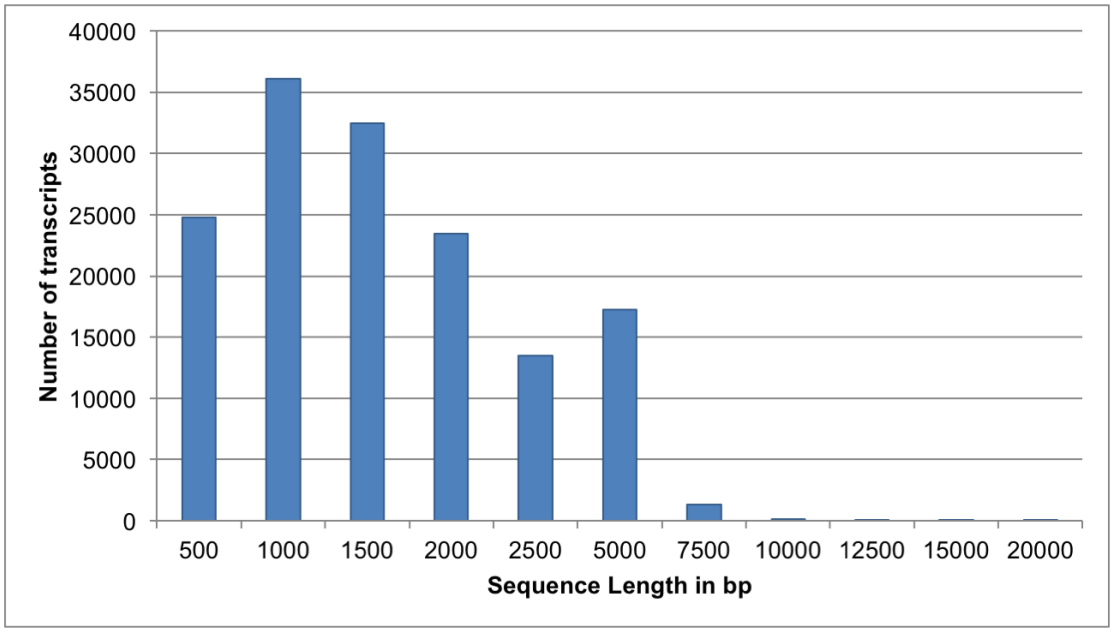


Figure 2.2 Distribution of transcript length of the Newbler assembly of the root transcriptome. Numbers in the X-axis indicate the length of the transcript in bp, and numbers in the Y-axis indicate the quantity of the transcripts.

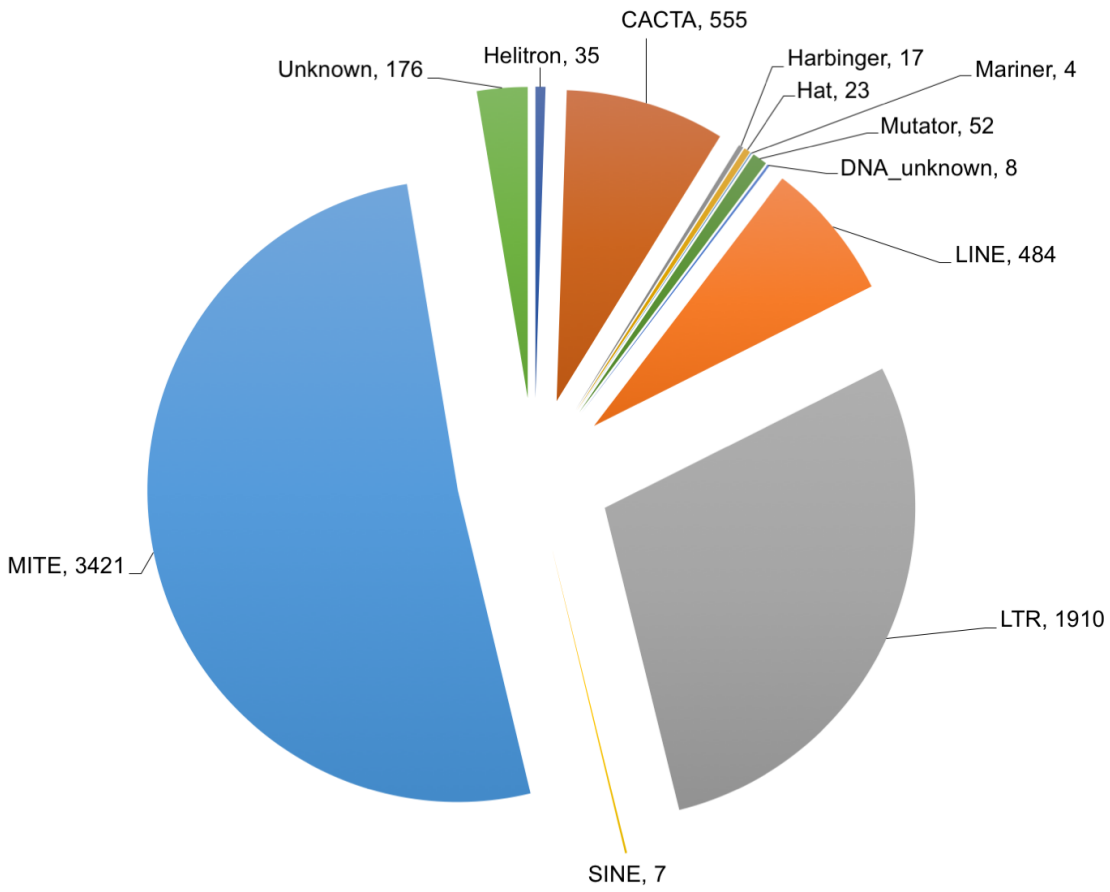


Figure 2.3 Transposable elements expressed in root transcriptome. The pie chart presents the different classes of the transposable elements expressed in the root tissues. The numbers after the transposon class are the number of transcripts in each class expressed in the root transcriptome. DNA_unknown, unknown DNA transposons; MITE, miniature inverted-repeat transposable elements; LINE, long interspersed elements; and SINE, short interspersed elements.

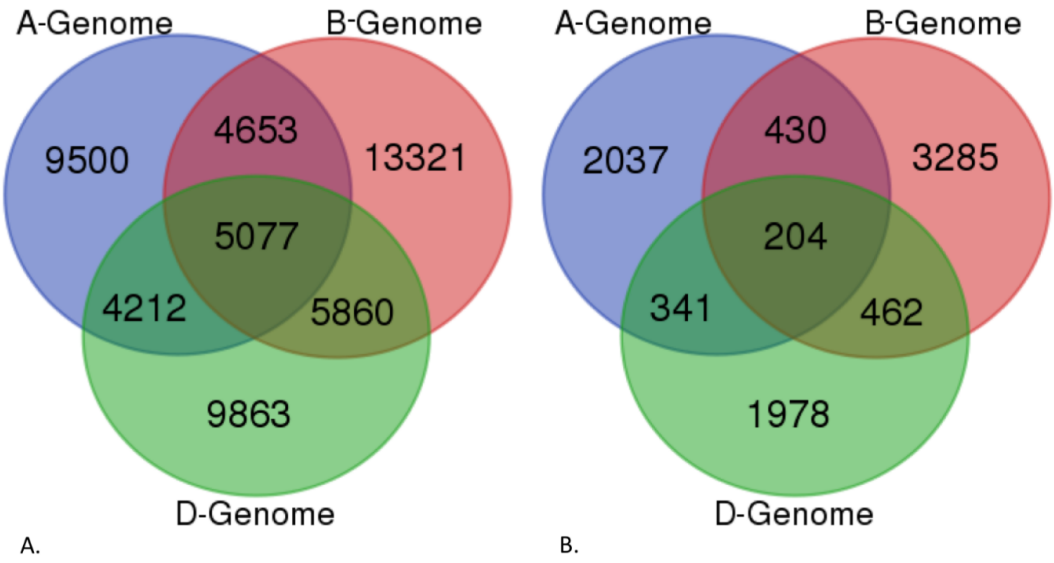


Figure 2.4 Venn diagram showing the distribution of ORF and non-ORF transcript nucleotide sequence alignment with the IWGSC draft genome sequences separated into sub-genomes. A) Distribution of the alignment of protein-coding transcripts with the sub-genome separated chromosome sequences. B) Distribution of the alignment of non-ORF transcripts with the sub-genome separated chromosome sequences. The sub-genomes A, B, and D are represented by color, and the numbers within the circles indicate the number of transcripts aligned in each of the sub-genomes.

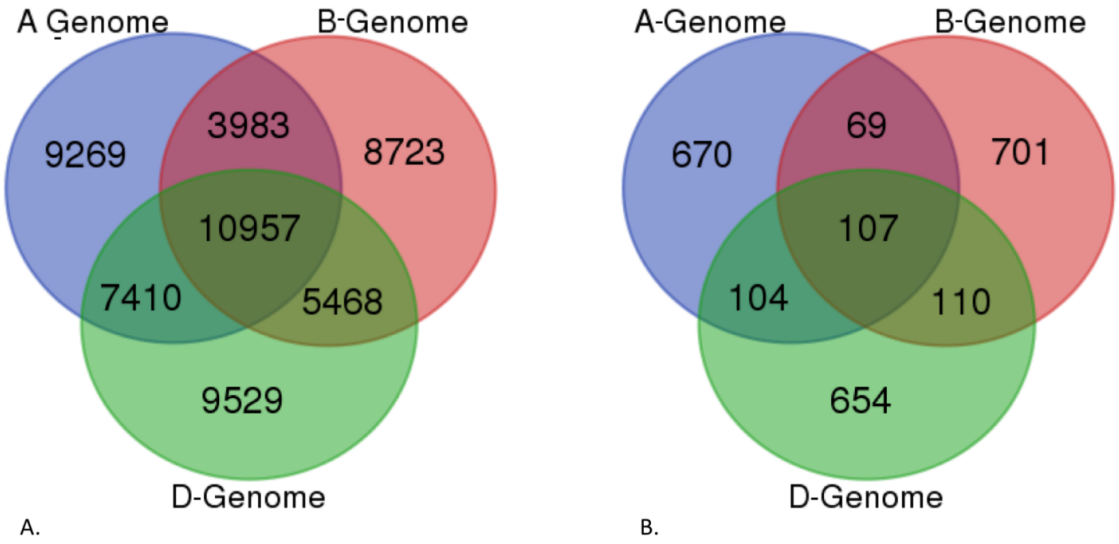


Figure 2.5 Venn diagram showing the distribution of ORF and non-ORF transcript nucleotide sequence alignment with the IWGSC cDNA sequences separated into sub-genomes. A) Distribution of the alignment of protein-coding transcripts with the sub-genome separated cDNA sequences. B) Distribution of the alignment of non-ORF transcripts with the sub-genome separated cDNA sequences. The sub-genomes A, B, and D are represented by color, and the numbers within the circles indicate the number of transcripts aligned in each of the sub-genomes.

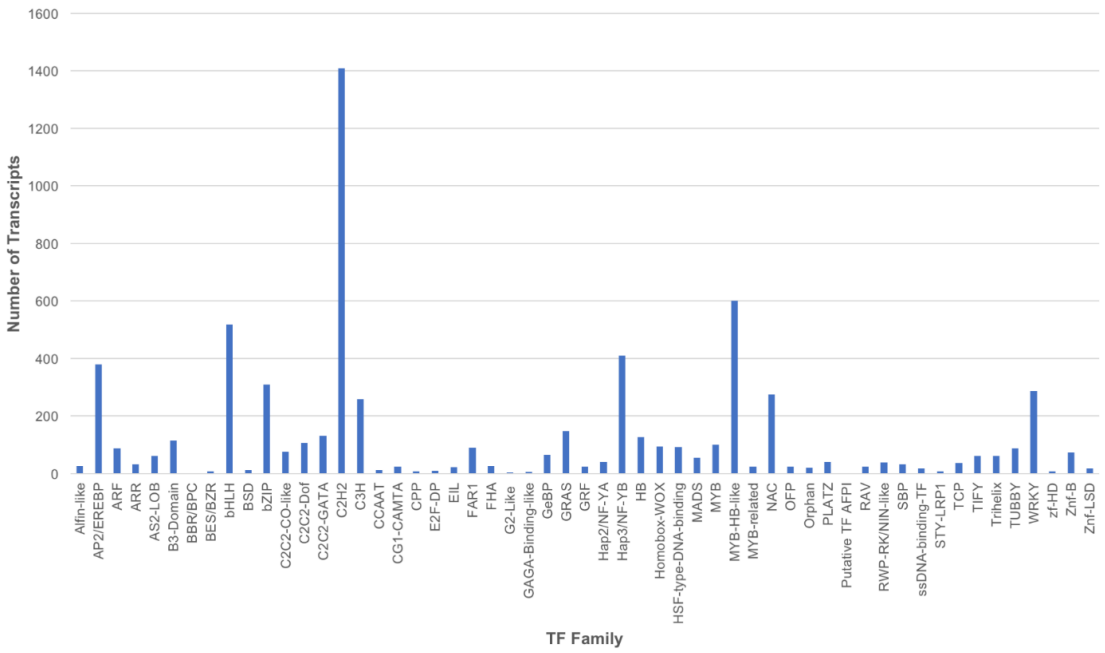


Figure 2.6 Transcription factor (TF) families expressed in both the root tissues used in the study. Numbers of transcripts in each TF family are indicated on the X-axis, and names of the TF families are indicated on the Y-axis.

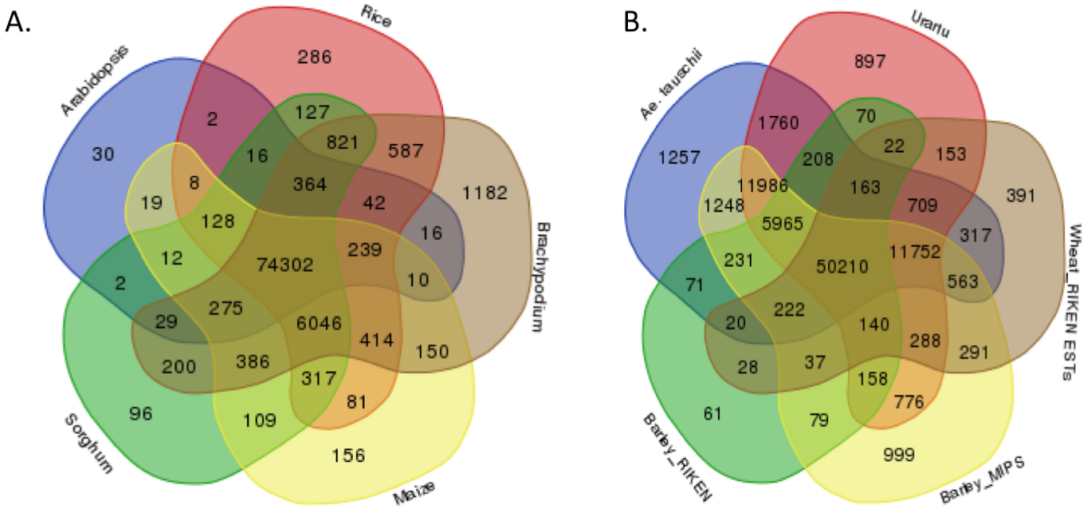


Figure 2.7 Venn diagrams showing the similarity of wheat root transcriptome with finished and draft genomes of model and crop plants. A) Comparision of the root transcripts against the protein sequences of finished genomes. B) Comparision of root transcripts against the protein sequences of draft genomes and assembled ESTs. Arabidopsis (TAIR v10); Rice (RGAP v 7); Brachypodium (Pyhtozome, Bd192); Sorghum (Phytozome, Sb79); Maize (Phytozome, Zm181); *Ae. Taushii* (Jia et al., 2012); Urartu (*Triticum urartu*) (Ling et al., 2013); Wheat_RIKEN_ESTs and Barley_RIKEN – translated protein sequences from assembled ESTs at RIKEN (http://trifldb.psc.riken.jp/v3/index.pl); Barley_MIPS – protein sequences from barley genome from MIPS (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/).

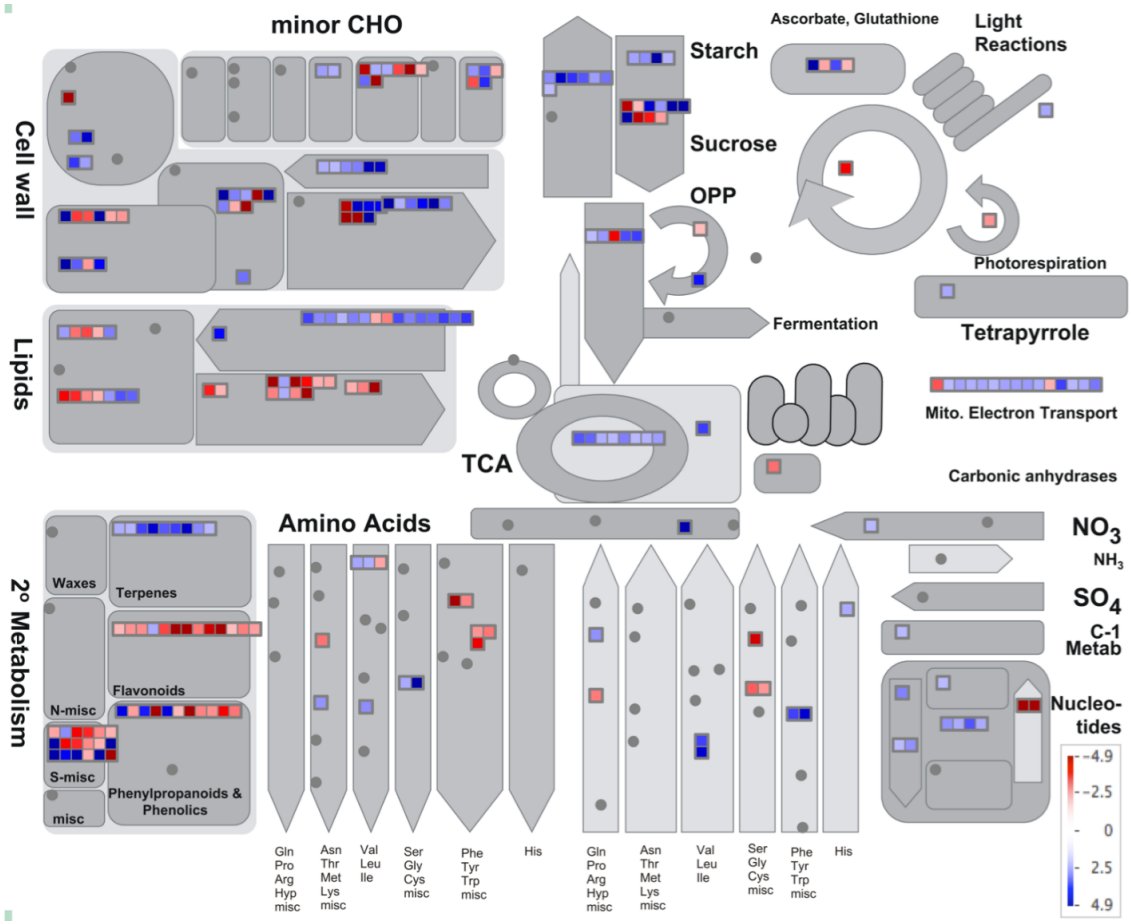


Figure 2.8 An overview of the differentially expressed transcripts mapped onto the metabolic pathways. The differentially expressed genes were mapped onto the metabolic pathways using MAPMAN software. Each box represents a transcript, and the red colored are the up-regulated in the mature root tissues and the blue colored are induced in the root tips. A fold change scale is indicated in the lower right corner.

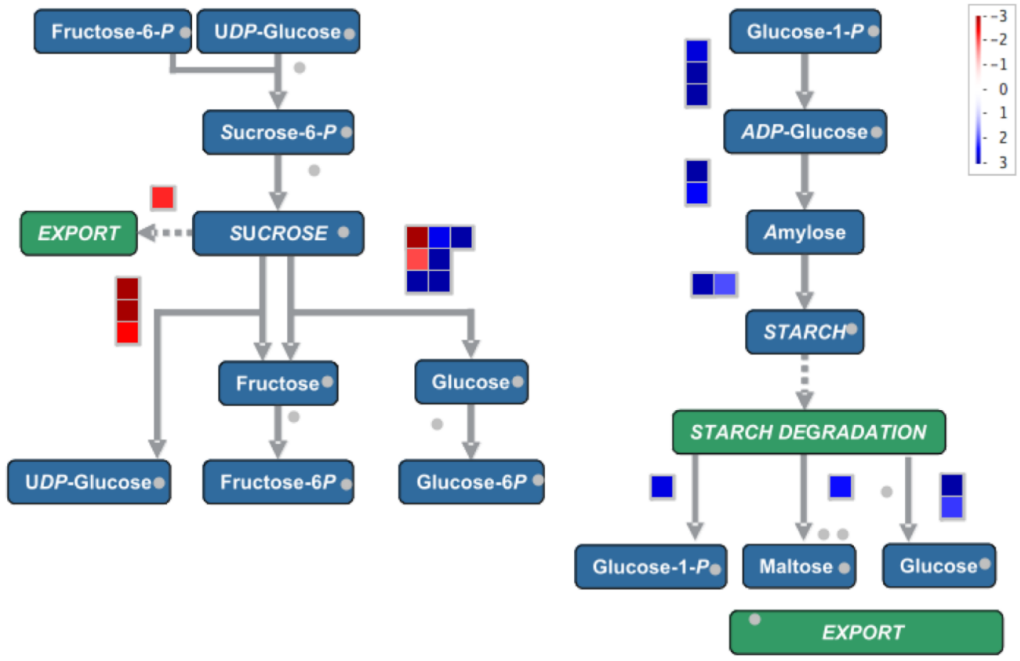


Figure 2.9 Differentially expressed genes in root tips and mature root involved in the starch biosynthesis. The transcripts encoding for the enzymes involved in the starch and sucrose metabolism were represented each by a box. The blue colored ones are induced in the root tips and the mature root tissue. A fold change scale is indicated in the upper right corner.

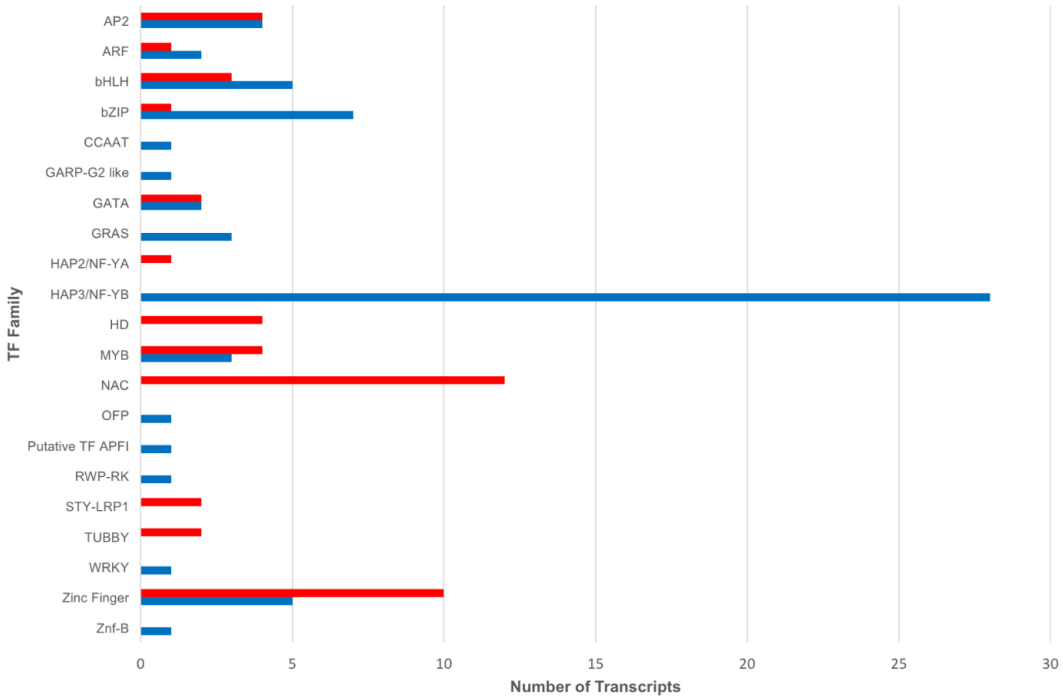


Figure 2.10 Transcription factor (TF) families differentially expressed in root tip and the mature root tissues. Numbers of transcripts in each TF family are indicated on the X-axis, and names of the TF families are indicated on the Y-axis. The striped bars are the transcripts induced in mature root and the solid black bars in the root tips.

SUPPLEMENTARY FIGURES

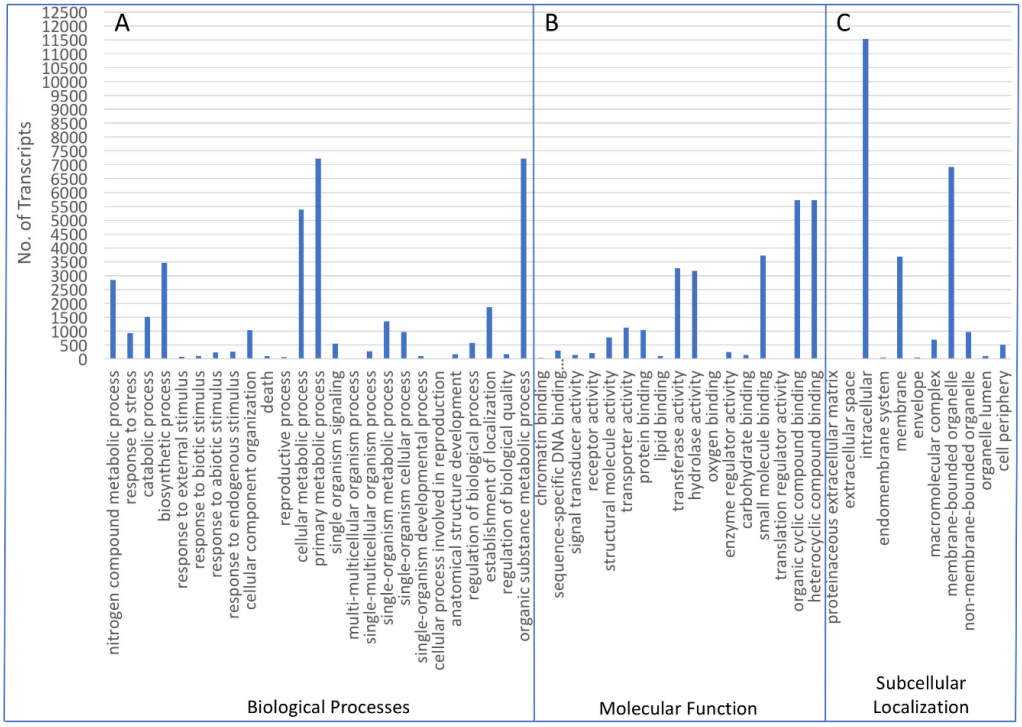


Figure S2.1 Gene Ontology (GO) classification of the de novo assembled 454 contigs. (A) Biological processes (B) Molecular functions (C) Subcellular localization. The GO categories are indicated on the X-axis, and the number of transcripts in each category is indicated on Y-axis.

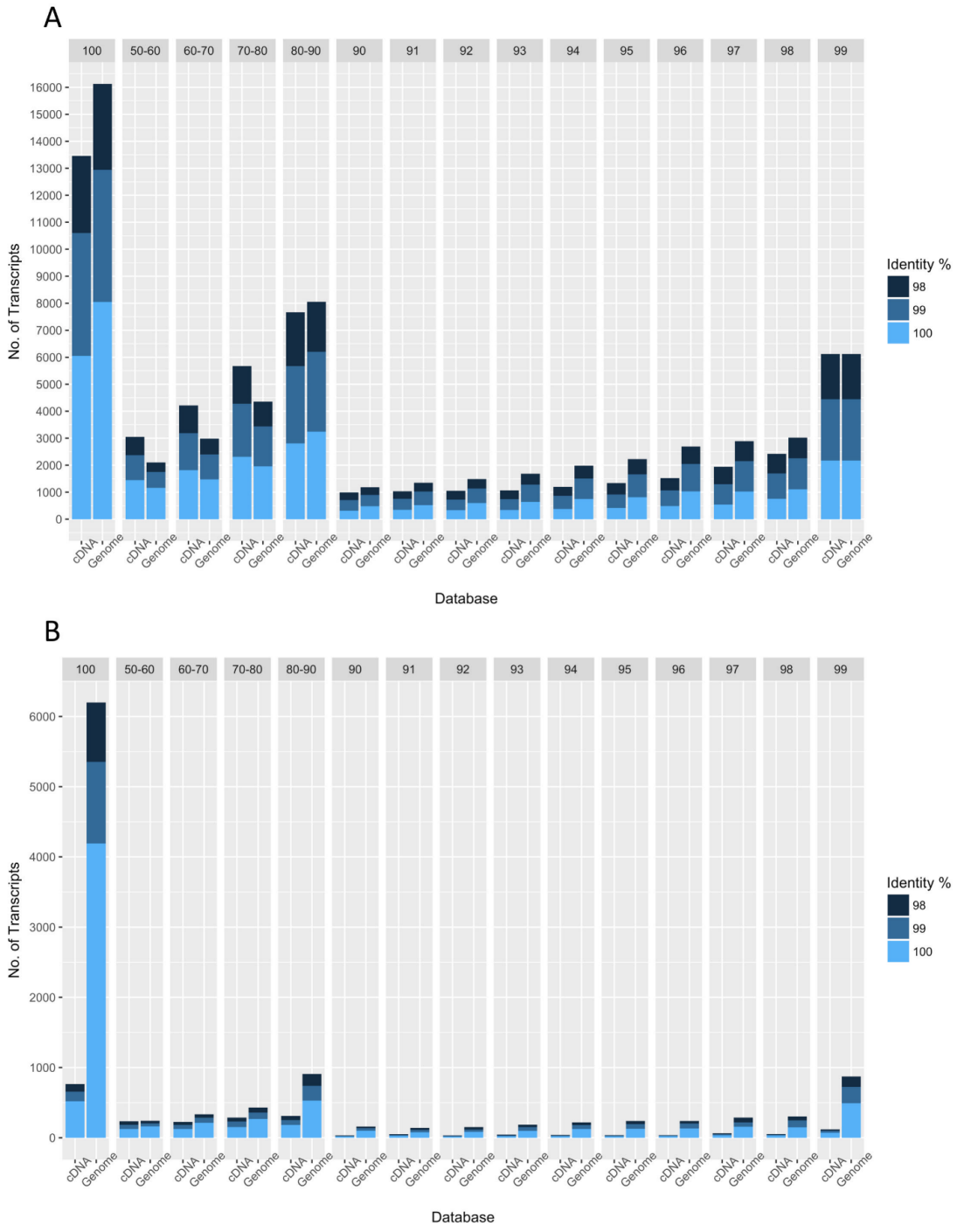
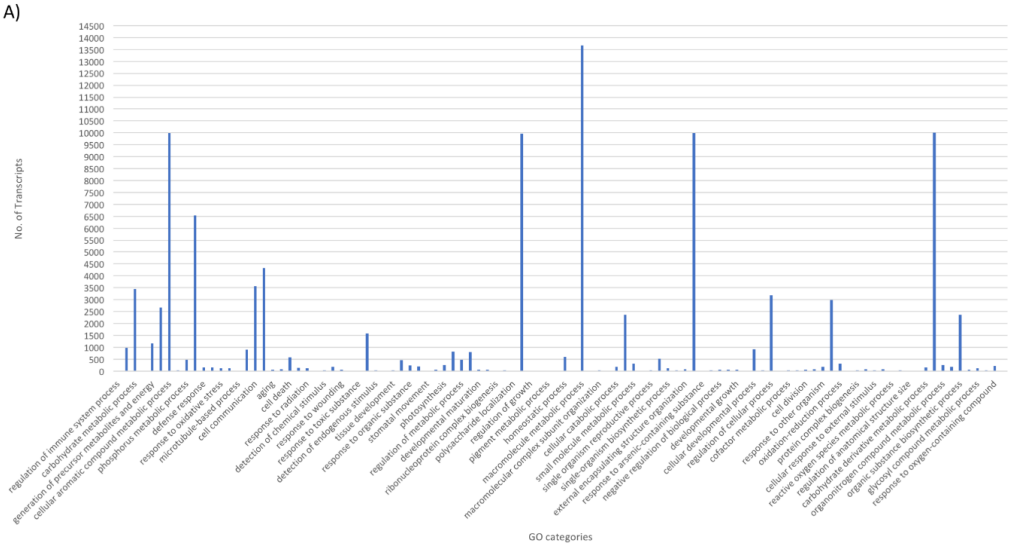
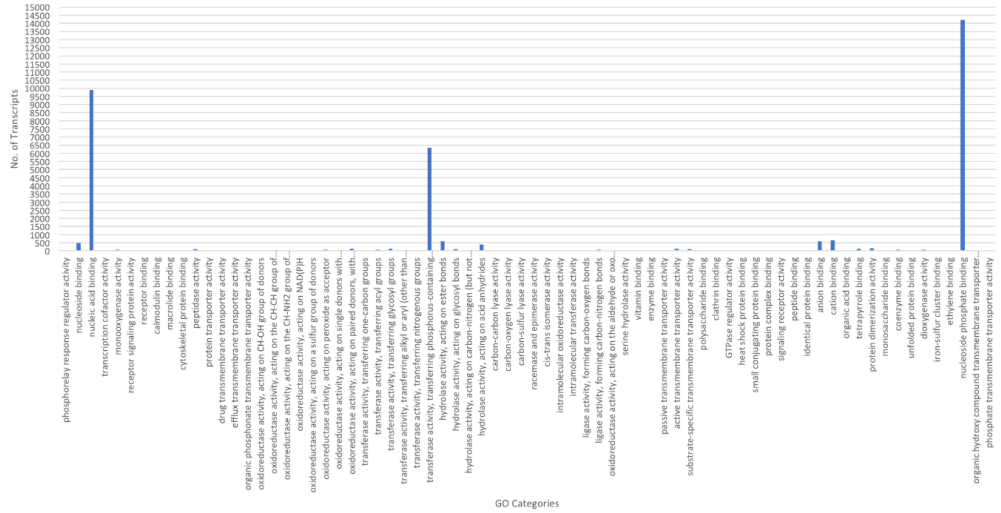

A
100
50-60
60-70
70-80
80-90
90
91
92
93
94
95
96
97
98
99
No. of Transcripts
Database
cDNA
Genome
Identity %
98
99
100
B
No. of Transcripts
Database
Identity %
98
99
100

Figure S2.2 Alignment statistics of the de novo assembled root transcriptome against the genomic and the predicted cDNA sequences from the hexaploid wheat draft genome. (A) Aclustered and stacked bar chart showing the coverage percent and the identity percent of the root transcripts with predicted ORFs against the genomic and the DNA sequences of the draft genome. (B) A clustered and a stacked bar chart showing the coverage percent and the identity percent of the root transcripts with without predicted ORFs against the genomic and the DNA sequences of the draft genome. The bars are stacked by the percent identity (represented by color) of the alignment and are clustered by the percent of the query covered in the alignment (shown by the labels on the top of each cluster). The database against which the transcripts were compared is indicated on the X-axis. The number of transcripts in each bin is indicated on Y-axis.

A)


No. of Transcripts
GO categories

B)


No. of Transcripts
GO Categories

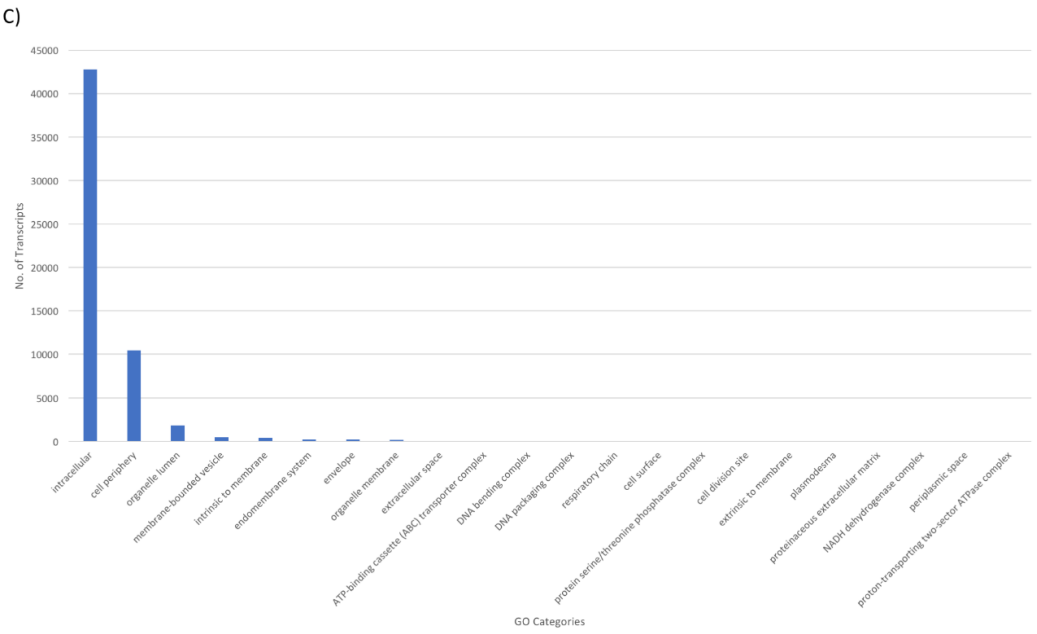


Figure S2.3 Gene Ontology (GO) classification of the Root transcripts with predicted ORFs. (A) Biological processes (B) Molecular functions (C) Subcellular localization. The GO categories are indicated on the X-axis, and the number of transcripts in each category is indicated on Y-axis.

SUPPLEMENTARY TABLES

Table S2.1: GO classification of the 454 de novo assembled root transcripts

A) Biological processes		
GO Term	**GO Name**	**No. of Transcripts**
GO:0071704	organic substance metabolic process	7227
GO:0044238	primary metabolic process	7224
GO:0044237	cellular metabolic process	5388
GO:0009058	biosynthetic process	3458
GO:0006807	nitrogen compound metabolic process	2852
GO:0051234	establishment of localization	1867
GO:0009056	catabolic process	1526
GO:0044710	single-organism metabolic process	1350
GO:0016043	cellular component organization	1034
GO:0044763	single-organism cellular process	963
GO:0006950	response to stress	928
GO:0050789	regulation of biological process	580
GO:0044700	single organism signaling	547
GO:0044707	single-multicellular organism process	266
GO:0009719	response to endogenous stimulus	265
GO:0009628	response to abiotic stimulus	236
GO:0048856	anatomical structure development	171

GO:0065008	regulation of biological quality	162
GO:0016265	death	107
GO:0009607	response to biotic stimulus	106
GO:0044767	single-organism developmental process	105
GO:0009605	response to external stimulus	77
GO:0022414	reproductive process	56
GO:0044706	multi-multicellular organism process	16
GO:0048610	cellular process involved in reproduction	10
B) Molecular function		
GO Term	**GO Name**	**No. of Transcripts**
GO:1901363	heterocyclic compound binding	5726
GO:0097159	organic cyclic compound binding	5726
GO:0036094	small molecule binding	3727
GO:0016740	transferase activity	3269
GO:0016787	hydrolase activity	3177
GO:0005215	transporter activity	1130
GO:0005515	protein binding	1038
GO:0005198	structural molecule activity	767
GO:0003700	sequence-specific DNA binding transcription factor activity	295
GO:0030234	enzyme regulator activity	243

GO:0004872	receptor activity	208
GO:0004871	signal transducer activity	136
GO:0030246	carbohydrate binding	135
GO:0008289	lipid binding	101
GO:0003682	chromatin binding	38
GO:0019825	oxygen binding	6
GO:0045182	translation regulator activity	1
C) Subcellular localization		
GO Term	**GO Name**	**No. of Transcripts**
GO:0005622	intracellular	11533
GO:0043227	membrane-bounded organelle	6923
GO:0016020	membrane	3686
GO:0043228	non-membrane-bounded organelle	964
GO:0032991	macromolecular complex	689
GO:0071944	cell periphery	506
GO:0043233	organelle lumen	107
GO:0012505	endomembrane system	47
GO:0031975	envelope	47
GO:0005578	proteinaceous extracellular matrix	4
GO:0005615	extracellular space	1

Table S2.2: GO classification of the root transcripts with predicted ORFs.

A) Biological processes		
GO Term	**GO Name**	**No. of Transcripts**
GO:0043170	macromolecule metabolic process	13687
GO:1901360	organic cyclic compound metabolic process	10005
GO:0006725	cellular aromatic compound metabolic process	9995
GO:0046483	heterocycle metabolic process	9979
GO:0034641	cellular nitrogen compound metabolic process	9970
GO:0006810	transport	6533
GO:0007275	multicellular organismal development	4335
GO:0007154	cell communication	3574
GO:0005975	carbohydrate metabolic process	3446
GO:0050794	regulation of cellular process	3180
GO:0051716	cellular response to stimulus	2992
GO:0006629	lipid metabolic process	2656
GO:0044249	cellular biosynthetic process	2369
GO:1901576	organic substance biosynthetic process	2362
GO:0009653	anatomical structure morphogenesis	1586
GO:0006091	generation of precursor metabolites and energy	1156
GO:0003006	developmental process involved in reproduction	971
GO:0048869	cellular developmental process	915
GO:0007049	cell cycle	894

GO:0016049	cell growth	820
GO:0019748	secondary metabolic process	803
GO:0042592	homeostatic process	602
GO:0008219	cell death	572
GO:0044703	multi-organism reproductive process	516
GO:0006793	phosphorus metabolic process	478
GO:0019222	regulation of metabolic process	471
GO:0009991	response to extracellular stimulus	459
GO:0044281	small molecule metabolic process	308
GO:0055114	oxidation-reduction process	305
GO:0015979	photosynthesis	258
GO:1901564	organonitrogen compound metabolic process	257
GO:0010033	response to organic substance	225
GO:1901700	response to oxygen-containing compound	206
GO:0010035	response to inorganic substance	186
GO:0009606	tropism	179
GO:0051707	response to other organism	174
GO:1901575	organic substance catabolic process	171
GO:0044248	cellular catabolic process	167
GO:0006970	response to osmotic stress	161
GO:0006952	defense response	153
GO:1901135	carbohydrate derivative metabolic process	147
GO:0009266	response to temperature stimulus	125

GO:0044711	single-organism biosynthetic process	110
GO:0009314	response to radiation	109
GO:0006979	response to oxidative stress	108
GO:1901657	glycosyl compound metabolic process	105
GO:0006996	organelle organization	104
GO:0071554	cell wall organization or biogenesis	82
GO:0045229	external encapsulating structure organization	73
GO:0051641	cellular localization	72
GO:0008104	protein localization	68
GO:0071495	cellular response to endogenous stimulus	67
GO:0048519	negative regulation of biological process	54
GO:0022607	cellular component assembly	52
GO:0010817	regulation of hormone levels	50
GO:1901615	organic hydroxy compound metabolic process	49
GO:0009611	response to wounding	47
GO:0048583	regulation of response to stimulus	46
GO:0051301	cell division	45
GO:0048589	developmental growth	44
GO:0007568	aging	43
GO:0021700	developmental maturation	43
GO:0080167	response to karrikin	42
GO:0051186	cofactor metabolic process	41
GO:0043933	macromolecular complex subunit organization	40

GO:0009888	tissue development	39
GO:0048518	positive regulation of biological process	38
GO:0050793	regulation of developmental process	36
GO:1901698	response to nitrogen compound	32
GO:0070271	protein complex biogenesis	31
GO:0051239	regulation of multicellular organismal process	30
GO:0006790	sulfur compound metabolic process	28
GO:0071496	cellular response to external stimulus	28
GO:0044712	single-organism catabolic process	27
GO:0009593	detection of chemical stimulus	26
GO:0044702	single organism reproductive process	25
GO:0023051	regulation of signaling	25
GO:0009726	detection of endogenous stimulus	23
GO:0072593	reactive oxygen species metabolic process	21
GO:0040008	regulation of growth	19
GO:0042493	response to drug	17
GO:0010118	stomatal movement	16
GO:2000241	regulation of reproductive process	16
GO:1900673	olefin metabolic process	15
GO:0042440	pigment metabolic process	14
GO:0061024	membrane organization	14
GO:0022613	ribonucleoprotein complex biogenesis	13
GO:0002682	regulation of immune system process	12

GO:0044093	positive regulation of molecular function	12
GO:0051128	regulation of cellular component organization	10
GO:0042330	taxis	10
GO:0048609	multicellular organismal reproductive process	9
GO:0009636	response to toxic substance	8
GO:0009629	response to gravity	8
GO:0009410	response to xenobiotic stimulus	7
GO:0009812	flavonoid metabolic process	7
GO:0044403	symbiosis, encompassing mutualism through parasitism	7
GO:0033037	polysaccharide localization	7
GO:0043062	extracellular structure organization	6
GO:0090066	regulation of anatomical structure size	6
GO:0007017	microtubule-based process	6
GO:0006081	cellular aldehyde metabolic process	5
GO:0046685	response to arsenic-containing substance	5
GO:0043476	pigment accumulation	5
GO:0007585	respiratory gaseous exchange	4
GO:0043900	regulation of multi-organism process	4
GO:0032196	transposition	4
GO:0032879	regulation of localization	4
GO:0030029	actin filament-based process	4
GO:0010876	lipid localization	3

GO:0009292	genetic transfer	3
GO:0044092	negative regulation of molecular function	2
GO:0071941	nitrogen cycle metabolic process	2
GO:2001057	reactive nitrogen species metabolic process	2
GO:0035821	modification of morphology or physiology of other organism	2
GO:0006403	RNA localization	2
GO:0006928	cellular component movement	1
GO:0046209	nitric oxide metabolic process	1
GO:0097438	exit from dormancy	1
GO:0007155	cell adhesion	1
GO:0009581	detection of external stimulus	1
GO:0009582	detection of abiotic stimulus	1
B) Molecular function		
GO Term	**GO Name**	**No. of Transcripts**
GO:1901265	nucleoside phosphate binding	14213
GO:0003676	nucleic acid binding	9898
GO:0016772	transferase activity, transferring phosphorus-containing groups	6356
GO:0043169	cation binding	641
GO:0043168	anion binding	597

GO:0016788	hydrolase activity, acting on ester bonds	572
GO:0001882	nucleoside binding	461
GO:0016817	hydrolase activity, acting on acid anhydrides	390
GO:0046983	protein dimerization activity	146
GO:0016757	transferase activity, transferring glycosyl groups	129
GO:0022804	active transmembrane transporter activity	123
GO:0046906	tetrapyrrole binding	114
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	113
GO:0008233	peptidase activity	107
GO:0022891	substrate-specific transmembrane transporter activity	100
GO:0016798	hydrolase activity, acting on glycosyl bonds	89
GO:0050662	coenzyme binding	75
GO:0004497	monooxygenase activity	69
GO:0016879	ligase activity, forming carbon-nitrogen bonds	69
GO:0016684	oxidoreductase activity, acting on peroxide as acceptor	67
GO:0016746	transferase activity, transferring acyl groups	58
GO:0051213	dioxygenase activity	52
GO:0005102	receptor binding	46
GO:0016741	transferase activity, transferring one-carbon groups	39

GO:0042802	identical protein binding	35
GO:0051082	unfolded protein binding	34
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	33
GO:0031072	heat shock protein binding	32
GO:0019899	enzyme binding	27
GO:0016830	carbon-carbon lyase activity	27
GO:0016667	oxidoreductase activity, acting on a sulfur group of donors	25
GO:0043177	organic acid binding	25
GO:0015238	drug transmembrane transporter activity	21
GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	21
GO:0016810	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	19
GO:0017171	serine hydrolase activity	18
GO:0016903	oxidoreductase activity, acting on the aldehyde or oxo group of donors	18
GO:0016860	intramolecular oxidoreductase activity	18
GO:0003712	transcription cofactor activity	17
GO:0030695	GTPase regulator activity	17
GO:0030247	polysaccharide binding	16

GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	16
GO:0016835	carbon-oxygen lyase activity	16
GO:0005516	calmodulin binding	15
GO:0022803	passive transmembrane transporter activity	15
GO:0016769	transferase activity, transferring nitrogenous groups	14
GO:1901618	organic hydroxy compound transmembrane transporter activity	14
GO:1901677	phosphate transmembrane transporter activity	14
GO:0016875	ligase activity, forming carbon-oxygen bonds	14
GO:0051536	iron-sulfur cluster binding	13
GO:0016651	oxidoreductase activity, acting on NAD(P)H	13
GO:0019842	vitamin binding	12
GO:0048029	monosaccharide binding	12
GO:0016866	intramolecular transferase activity	12
GO:0016638	oxidoreductase activity, acting on the CH-NH2 group of donors	11
GO:0032403	protein complex binding	10
GO:0038023	signaling receptor activity	9
GO:0008092	cytoskeletal protein binding	9
GO:0000156	phosphorelay response regulator activity	8
GO:0016859	cis-trans isomerase activity	8
GO:0016846	carbon-sulfur lyase activity	8

GO:0005057	receptor signaling protein activity	8
GO:0030276	clathrin binding	7
GO:0008565	protein transporter activity	7
GO:0015604	organic phosphonate transmembrane transporter activity	6
GO:0005527	macrolide binding	6
GO:0032182	small conjugating protein binding	6
GO:0051740	ethylene binding	5
GO:0015562	efflux transmembrane transporter activity	5
GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	5
GO:0016854	racemase and epimerase activity	5
GO:0042277	peptide binding	5
GO:0016679	oxidoreductase activity, acting on diphenols and related substances as donors	4
GO:0016645	oxidoreductase activity, acting on the CH-NH group of donors	3
GO:0016725	oxidoreductase activity, acting on CH or CH2 groups	3
GO:0016877	ligase activity, forming carbon-sulfur bonds	3
GO:0061135	endopeptidase regulator activity	3
GO:0030414	peptidase inhibitor activity	3
GO:0016885	ligase activity, forming carbon-carbon bonds	3

GO:0005484	SNAP receptor activity	2
GO:0019779	APG8 activating enzyme activity	2
GO:0016872	intramolecular lyase activity	2
GO:0015232	heme transporter activity	2
GO:0016782	transferase activity, transferring sulfur-containing groups	2
GO:0016661	oxidoreductase activity, acting on other nitrogenous compounds as donors	2
GO:0030742	GTP-dependent protein binding	1
GO:0080132	fatty acid alpha-hydroxylase activity	1
GO:0015197	peptide transporter activity	1
GO:0019887	protein kinase regulator activity	1
GO:0016840	carbon-nitrogen lyase activity	1
GO:0016722	oxidoreductase activity, oxidizing metal ions	1
GO:0016824	hydrolase activity, acting on acid halide bonds	1
C) Subcellular localization		
GO Term	**GO Name**	**No. of Transcripts**
GO:0005622	intracellular	42785
GO:0071944	cell periphery	10434
GO:0043233	organelle lumen	1811
GO:0031988	membrane-bounded vesicle	483

GO:0031224	intrinsic to membrane	371
GO:0012505	endomembrane system	223
GO:0031975	envelope	215
GO:0031090	organelle membrane	140
GO:0005615	extracellular space	24
GO:0043190	ATP-binding cassette (ABC) transporter complex	12
GO:1990104	DNA bending complex	20
GO:0044815	DNA packaging complex	20
GO:0070469	respiratory chain	6
GO:0009986	cell surface	7
GO:0008287	protein serine/threonine phosphatase complex	4
GO:0032153	cell division site	4
GO:0019898	extrinsic to membrane	5
GO:0009506	plasmodesma	3
GO:0005578	proteinaceous extracellular matrix	2
GO:0030964	NADH dehydrogenase complex	2
GO:0042597	periplasmic space	1
GO:0016469	proton-transporting two-sector ATPase complex	1

Tabe S2.3 Transcription factor (TF) families expressed in both the root tissues used in this study.

TF Family	Number of Transcripts
Alfin-like	26
AP2/EREBP	378
ARF	88
ARR	32
AS2-LOB	61
B3-Domain	113
BBR/BPC	1
BES/BZR	7
bHLH	517
BSD	11
bZIP	308
C2C2-CO-like	74
C2C2-Dof	105
C2C2-GATA	131
C2H2	1409
C3H	257
CCAAT	11
CG1-CAMTA	23
CPP	8
E2F-DP	10

EIL	22
FAR1	89
FHA	26
G2-Like	4
GAGA-Binding-like	6
GeBP	65
GRAS	146
GRF	23
Hap2/NF-YA	40
Hap3/NF-YB	410
HB	126
Homobox-WOX	94
HSF-type-DNA-binding	92
MADS	54
MYB	100
MYB-HB-like	600
MYB-related	23
NAC	273
OFP	23
Orphan	20
PLATZ	39
Putative TF AFPI	1

RAV	24
RWP-RK/NIN-like	37
SBP	31
ssDNA-binding-TF	18
STY-LRP1	8
TCP	36
TIFY	60
Trihelix	61
TUBBY	88
WRKY	286
zf-HD	8
Znf-B	73
Znf-LSD	18

Table S1.4: Annotation and the expression profiles of the DETs

The table was uploaded to the proquest as supplementary material.

CHAPTER 3

Transcriptome Analysis of a Very Short Root Phenotype

ABSTRACT

Roots play a major role in the plant growth and development. The knowledge of root biology is limited in cereals like wheat because the invasive methods are usually required for these studies. To gain insights into molecular regulation of a very short root (VSR) phenotype in wheat, we sequenced the six mRNA libraries from three long root segregant pools and three VSR segregant pools, by the HiSeq2000 platform. Mapping of the cleaned reads to the reference genome of Chinese Spring wheat and transcript quantification identified 4,412 differentially expressed transcripts between the VSR and LR, of which 3,635 were up-regulated and 777 down-regulated in VSR. Of the up-regulated genes, a significant portion belongs to the hormonal responses, regulation of transcription, defense response, reactive oxygen species (ROS), abiotic stress response, lignin biosynthesis, calcium signaling, and autophagy pathways. In addition, several negative regulators of cell proliferation, including homologs of the BIGBROTHER E3 ubiquitin ligase, were also up-regulated. Consistent with this, a large number of genes for chromatin replication and protein syntheses, such as those coding for histones and ribosomal proteins, were down-regulated in VSR. Transcription of genes encoding the FERONIA kinases and a RALF peptide hormone, negative regulators of root cell elongation, was elevated in VSR. In accordance, several cell wall remodeling genes, including those encoding xyloglucan endotransglucosylase/hydrolase, pectin lyase, expansin and cellulose synthase, were down-regulated in VSR. Regarding root development and hormone signaling, expression of several genes involved in auxin efflux, ethylene biosynthesis, JA signaling, and lateral root initiation was upregulated, but transcription of a bHLH transcription factor involved in GA

signaling was down-regulated in VSR. Although several homologs of PLETHORA, a major regulator of root growth in Arabidopsis, exist in the wheat genome, their transcription was not affected by the *Vsr1* state, suggesting that VSR is controlled by a different mechanism from what have found in the model plant Arabidopsis. We also validate the ROS and lignin by histochemical staining. Based on our results, we proposed a working model to explain the mechanisms underlying the VSR development.

INTRODUCTION

Roots are the underground part of a plant and play a significant role in plant growth and development. They are the major organ in the acquisition of water and nutrients, sensing the changes in the surrounding environment, even the micro level modulations, interacting with microbial populations in the rhizosphere, and providing anchoring. Plant roots "keep an eye" on the continuously fluctuating surrounding environment in soil throughout their growth and development and relay that information to the rest part of the plant to respond and adapt accordingly. Compared to the above-ground parts of the plant, the knowledge about how roots perform all their functions is much less known. The most available knowledge of root development and growth is derived from the model plant Arabidopsis, a dicot. The research on the root growth, development, and functions in crop plants like soybean, rice, wheat, etc. is nowhere near to the work done in Arabidopsis, and most of the available information mainly is translated from the Arabidopsis research.

The root growth in higher plants is mainly achieved by tangible elongation of the cells that descend from the root apical meristem (RAM). RAM is a specialized tissue present in the tip of the growing root and has a reservoir of undifferentiated cells at the tip. These stem cells are surrounded by a group of organizing cells that maintain the undifferentiated state and are called quiescent center (QC) (van den Berg et al. 1995; van den Berg et al. 1997). RAM produces all the cell types essential for the postembryonic root development along both longitudinal and radial axes (Dolan et al. 1993; van den Berg et al. 1998). The meristematic cells in RAM can stay pluripotent forever and produce different root cell types and simultaneously self-renewing. Thus, these cells together with the QC

are regarded as the stem-cell niche, which is formed during embryogenesis (Dinneny and Benfey 2008; Luo et al. 2017; Perilli et al. 2012; van den Berg et al. 1997). In Arabidopsis, there are 15 cell types that make the root, and all of them are differentiated from the asymmetric cell division of the post-embryonic stem cells (Jiang and Feldman 2005).

Common wheat or bread wheat (*Triticum aestivum* L., genomes BBAADD) is one of the most important cereal crops in the world. It provides about one-fifth of the calories consumed by the humans worldwide (http://faostat.fao.org). Bread wheat is a hexaploid originated from the spontaneous hybridization between a domesticated form of tetraploid wheat (*T.turgidum* L., genomes AABB) and diploid goatgrass (*Aegilops tauschii* Coss., genomes DD) (Kihara 1944; McFadden and Sears 1946), which took place in the Caspian Iran region (Wang et al. 2013) approximately 8,000 years ago (Nesbitt and Samuel, 1996). From the Caspian Iran region, wheat spread throughout the world. Wheat is known to be grown in diverse environments though the majority of its production comes from temperate regions with multiple seasons. Wheat is expected to play an important role in global food security. The ongoing population growth is projected to increase the demand for wheat by 40% in the year 2030 (Dixon et al. 2009). This increased demand has to be met with the production achieved with less acreage and less water than today. The traditional plant breeding methods have been proven slow and are only based on the selection for yield improvement (Richards et al. 2002). For plants to perform better under water-deficit conditions, a deep root system is required for increased water uptake. A faster and more extensive root growth will be more important for the plants to achieve better growth and yield under the most adverse soil conditions. A deeper and more extensive root system will

help uptake available water more efficiently from the soil in dry conditions, increase nutrient uptake, compete with weeds, and result in improved yield (Richards 2008).

The root system architecture is determined by the internal genetic component of the plant and shaped by the external environmental stimuli as well. The root development is of very adaptive plasticity, making it more complicated to understand the genetic pathways involved in root growth and development. Availability of thorough knowledge of the genes involved in the root development can enable breeding programs to focus more on the varieties with the improved root system. Cereal root system also vastly differs from that of the model plant Arabidopsis. It is a dense fibrous root system comprised of primarily of postembryonic adventitious roots. Several genes were identified to be involved in the cereal root development in rice. These include *OsGNOM1/CROWN-ROOTLESS4* encoding ADP-ribosylated factor G protein (ARF-GEF) involved in polar auxin transport and affects adventitious root formation (Kitomi et al. 2008; Liu et al. 2009) and *ADVENTITIOUS ROOTLESS1 (ARL1)/ CROWN ROOTLESS1 (CRL1)*, which codes for a Lateral Organ Boundary (LOB) domain protein and induced by auxin in the stem base where crown roots are formed and is involved in the lateral root formation and gravitropism (Inukai et al. 2005; Liu et al. 2005). A few quantitative trait loci (QTL) were identified in rice, maize, and wheat that control root traits (Coudert et al. 2010; Sanguineti et al. 2007). More insights are needed to understand the molecular mechanisms underlying cereal root development and growth, and a genome-wide transcriptome approach is expected to play an important role in this respect.

We discovered a very short root (VSR) phenotype in F_1 hybrid derived from a cross between common wheat landrace Chinese Spring (CS) and synthetic wheat accession

TA4152-71 (TA) (Figure 3.1). This root phenotype in the hybrid has blunt, swollen and yellow or brown colored root tips compared to white and sharp root tips of the parental lines. The root growth rate was significantly reduced, and lateral roots started to appear very early compared to either of the parental lines. In our previous paper that reports the discovery of the VSR, we showed that a non-additive interaction of the parental genomes in the F_1 hybrid resulted in VSR phenotype. Genetic analysis of the VSR placed the gene controlling this trait, i.e., *Vsr1*, in the distal region of the long arm of 5D chromosome of wheat (Li et al. 2013). Transcriptome study of the VSR would provide an opportunity for us to understand wheat root development and growth by identifying the genes and pathways underlying. In the present research, we focused on identifying the genes that are differentially expressed in the VSR compared to the long root (LR) by RNA-Seq analysis of their root tips and genetic pathways contributing to the VSR development. Thus, this research slated the process to reveal the molecular mechanisms underlying the VSR phenotype.

MATERIALS AND METHODS

Plant material and RNA extraction

The seeds from the BC_2F_2 generation were germinated in germination boxes with paper towels wetted with tap water. One root tip (~3mm, meristematic zone) from the primary root of each of the germinated seeds was collected individually after 2 d of germination, frozen in liquid nitrogen, and stored at -80^{o}C until further use. The seedlings and the vials with root tips were numbered correspondingly. The seedlings with remaining two intact

roots were allowed to grow for a few more days, and the phenotype was scored from the remaining roots. The root tips were pooled for the VSR seedlings and the long root (LR) seedlings. Each of the VSR and the LR samples included three biological replicates, and each biological replicate contained ~40 root tips. Due to a large number of root tips were included in each segregant pool (SP), the LR and VSR SPs had similar genetic background except for the *Vsr1* region, mimicking near isogenic lines. Total RNA was extracted from the root tips of all the six SPs using Trizol (Thermo Fisher Scientific, Waltham, MA) following the manufacturer's instruction. The RNA samples were purified using the RNeasy mini kit (Qiagen, Valencia, CA). Concentration and integrity of the purified RNA samples were quantified was confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA), and samples with an RNA integrity number (RIN) greater than eight were used in the subsequent analyses.

Illumina Sequencing

RNA samples extracted from root tips from both the LR and the VSR SPs were submitted for sequencing library construction and sequencing to the DNA Sequencing & Genotyping Center, Delaware Biotechnology Institute, Newark, DE. Six barcoded sequencing libraries for three biological replicates for the LR sample and three biological replicates for the VSR sample were prepared using the TruSeq RNA Library Prep Kit (Illumina, San Diego, CA). These six libraries were pooled and sequenced in one lane on the HiSeq 2000 platform (Illumina, San Diego, CA) to generate 100 bp single-end reads. The same libraries were also sequenced to generate 100 bp paired-end reads at a lower depth.

Quality control and preprocessing

The remnants of the adapters used during the library preparation were trimmed from both the single- and paired-end reads using Trimmomatic (Bolger et al. 2014) with default settings. The adapter-free reads were filtered for low quality and shorter reads and trimmed at the 3' end for low quality bases using the Prinseq script (Schmieder and Edwards 2011b). The parameters used for quality filtering and trimming were set for a minimum mean quality of Q20 across the read and to trim low-quality bases at 3' end. The minimum read length 50 bp was used as cutoffs for length filtering. The reads generated from rRNA sequences were filtered using Ribopicker Perl script (Schmieder et al. 2012) with a plant rRNA sequence dataset generated from the rDNA sequences retrieved from NCBI (http://www.ncbi.nlm.nih.gov), TAIR (https://www.arabidopsis.org) and the rice genome annotation database (http://rice.plantbiology.msu.edu) as a reference. The reads were also filtered for any contaminations like reads generated from bacterial DNA were filtered using Deconseq with *E. coli* strain K12 genome as a reference (Schmieder and Edwards 2011a).

Differential gene expression analysis of the VSR and LR

The high quality reads from both the LR and VSR libraries were aligned to Wheat RefSeq v1.0 (https://www.wheatgenome.org/News/Latest-news/RefSeq-v1.0-URGI) using HISAT2 v2.0.5 (Kim et al. 2015; Pertea et al. 2016) (with parameters --dta -k 20 --no-mixed and rest of the options as default). The aligned RNA-Seq reads were assembled into transcripts using Stringtie v1.3.3b (Pertea et al. 2015) with default parameters. The read count for the transcripts annotated in the reference genome and those assembled by the Stringtie were retrieved using the HTseq v0.9.1 (Anders et al. 2015). The gene expression

comparison of the transcripts in the two samples and the statistical analysis was done using the edgeR (v3.20.9) package in Bioconductor (https://bioconductor.org/packages/release/bioc/html/edgeR.html) (Robinson et al. 2010). The differentially expressed genes (DEGs) with an adjusted *p*-value of less than or equal to 0.05 and fold change of at least 2-fold were considered to be significant.

Functional annotation and GO assignment of differentially expressed genes in VSR

The DEGs between the LR and VSR were functionally annotated by performing a BLASTp search against the NCBI non-redundant (nr) protein database (http://www.ncbi.nlm.nih.gov), Rice proteome from Rice Genome Annotation Project (RGAP v7) (http://rice.plantbiology.msu.edu) and Arabidopsis proteome from The Arabidopsis Information Resource (TAIR v10) (https://www.arabidopsis.org). Gene Ontology (GO) terms were assigned using Blast2GO software (www.blast2go.com). The transcript sequences were further annotated using the Mercator 4 (v0.3) (http://www.plabipd.de/portal/web/guest/mercator-ii-alpha-version-) and assigned the MapMan bins (https://mapman.gabipd.org/). The metabolic pathways were visualized in the MapMan (v3.5).

Small RNA Sequencing

RNA samples extracted from root tips from both the LR and the VSR root tips were submitted to the DNA Sequencing & Genotyping Center, Delaware Biotechnology Institute, Newark, DE, for sequencing library construction and sequencing of the small RNA (sRNA) transcriptomes. Six barcoded sequencing libraries for three biological

replicates for the LR and three biological replicates for very short root (VSR) were prepared using the TruSeq Small RNA Library Preparation Kit (Illumina). These six libraries were pooled and sequenced in one lane on the HiSeq 2000 platform (Illumina) to generate 50 bp single-end reads.

Preprocessing of the small RNA libraries

The sequences from the six libraries were passed through our preprocessing pipeline to clean and filter the low quality and the contaminant bases and/or reads. First, the adapters were trimmed from the 3' end of the reads using Skewer (Jiang et al. 2014). The low-quality reads with an average Q score less than 20 were removed using the Prinseq tool (Schmieder and Edwards 2011b) and the reads shorter that 15 bp or longer than 25 bp were removed. The reads generated from the RNA groups such as rRNA, tRNA, snRNA, snoRNA and scRNA were removed from the analysis using the Deconseq tool (Schmieder and Edwards 2011a).

Differential expression analysis of the sRNA from VSR and LR

The remaining 15-25 nt reads were counted and merged using the small RNA analysis pipeline in CLC Genomics Workbench v11 (www.clcbio.com). The merged counts were annotated using the miRBase v22 (Kozomara and Griffiths-Jones 2011) and the microRNAs annotated in the Wheat Refseq v1.0. The counts were normalized and the differential expression of the sRNAs was determined by the empirical analysis of DGE tool available in the CLC genomics Workbench. The differentially expressed sRNA were filtered by at least two-fold change in expression with adjusted *p*-value of less than 0.05.

The targets for the differentially expressed sRNA were determined by searching against the wheat gene models using psRNATarget webserver (http://plantgrn.noble.org/psRNATarget/) (Dai et al. 2018).

Staining for hydrogen peroxide and lignin

The root tips of the 3-day old LR and VSR seedlings were stained with 3,3'-Diaminobenzidine (DAB) to detect hydrogen peroxide. DAB staining solution was freshly prepared by dissolving 1mg/ml DAB in distilled water adjusted to pH 3.6 using 0.1N HCl. The root tip samples were incubated in DAB stain for 30 min at room temperature and observed under a microscope.

The root tip samples of 1-, 2-, and 3-day old LR and VSR seedlings were stained for lignin using phloroglucinol stain. The stain solution is prepared by dissolving about 2g of phloroglucinol in 80 ml of 20% ethanol solution and then add 20 ml of concentrated HCl (12 N) to it. The root tips were incubated in the stain for 2-5 min and observed under light microscope and photographed.

RESULTS

Sequencing of RNA from the VSR and LR root tips

We sequenced the transcriptomes of the root tips from the LR and VSR SPs using illumine Hiseq 2000 platform. Sequencing six libraries, three for the long root (LR) and three for the very short root (VSR), generated 151,069,941 raw single-end reads, and 39,202,663 raw paired-end reads. The sequencing depth for the single-end reads range from ~23 to ~27

million reads per sample whereas for paired-end reads was ~6 million to ~7 million. Both the single- and the paired-end reads were preprocessed using our preprocessing pipeline for trimming adapters or the primer sequences, low-quality bases at the 3' end of the reads, filtering of the low-quality reads with an average PHRED score of <20 and contaminants like rRNA or bacterial DNA. The quality filtering resulted in 140,700,938 (93.14%) single-end reads and 34,624,093 (88.32%) paired-end reads and accounts for 92.14% of the total raw reads (Table 3.1).

Differential gene expression analysis of the LR and VSR root tips

The reads from the LR and VSR libraries were mapped to the wheat reference genome that was made available recently and the root transcriptome assembly developed in the previous chapter. The mapping details were listed in Table 3.2. The RefSeq 1.0 has 110,790 loci of high confidence gene models, and 158,793 loci of low confidence gene models coding for proteins were predicted in the sequenced reference genome (only ~14 Gb of the ~17 Gb of the wheat genome was captured in the current reference genome). For this analysis, we merged the high and low confidence gene models into a single annotation with 269,583 loci and used as the reference annotation for gene expression analysis. A total of 4,415 (3 loci are non-coding according to the Refseq1.0 annotation and not included in the further functional annotation analysis) loci were found differentially expressed between the VSR and LR with a cutoff of fold change (FC) ≥ 2 (also expressed as $|\log_2 FC| \geq 1$) and the adjusted *p*-value of ≤ 0.05. Of these 4,412 loci, 3,635 loci were up-regulated and 777 down-regulated in the VSR root tips as compared to the LR. Our pipeline for differential gene expression analysis uses the read mapping to the reference genome and assembly of the

root transcriptome, which allows the discovery of the new transcript loci that are not annotated or predicted in the reference gene models. As a result, a total of 245 DEGs were identified as novel transcripts, of which 72 and 173 transcripts up-regulated and down-regulated in VSR, respectively. These 245 transcripts were assembled based on the reads mapped to the reference genome sequence rather than the annotated gene models (Table 3.2).

Functional and Gene Ontology Annotation of the DEGs

Of the 3,563 up-regulated transcripts, the BLASTp search against NCBI-nr, Rice (RGAP v7) and Arabidopsis (TAIR10) databases provided the functional annotation for the 3561, 3,560 and 3,560 transcripts, respectively. For the 604 down-regulated genes, 602, 601 and 602 transcripts had a match in nr, rice, and Arabidopsis databases, respectively (Table 3.2). Three loci in up-regulated genes were annotated as noncoding RNA in the RefSeq 1.0 annotation. Gene ontology terms were assigned to each of the DEGs. For the up-regulated genes, 2,566 were assigned with at least one GO term, and for down-regulated genes, 399 were assigned with GO terms. In biological process categories, cellular metabolic processes, biosynthetic processes, localization, biological regulation and response to stress were the top five categories in the up-regulated genes. Ion binding, transferase activity, heterocyclic compound binding, transporter activity, and antacid antioxidant activity were the top five categories in the molecular function section. Membrane and part of the membrane were the top two categories in the cellular component section. In down-regulated genes, metabolic processes, cellular processes, cellular component biogenesis, response to stimulus and localization were top five categories in biological processes;

catalytic processes, binding processes, antioxidant activity, transporter activity, and molecular function regulator were top five categories in molecular function. The DEGs were mapped to MapMan classification bin using the Mercator tool. Of the 4,412 DEGs in VSR, 1,953 were assigned with at least one annotation or pathway bin. The overview of metabolic processes was visualized with MapMan (Figure 3. 2). The majority of the up-regulated genes in VSR root tips belong to transporters (271), transcription factors (262), reactive oxygen species (ROS) generation and scavenging (200), defense related (194), E3 ligases and other components of 26S proteasome pathway (124), signaling pathway related receptor kinases (113), protein kinases (95), and calcium signaling (83), cell wall modifying and degrading enzymes (67), lignin biosynthesis pathway (70), autophagy and cell death (43) (Table S3.1 and Figure 3.3).

Hormone-related gene expression

Plant growth and development has been inextricably linked to phytohormone signaling and regulation. In root development hormones such as auxin, ethylene, cytokinin, abscisic acid, gibberellin, brassinosteroids, jasmonic acid are shown to regulate root development either acting individually or interacting with other hormones synergistically and/or antagonistically depending on the zone/tissue and cell type context (Benkova and Hejatko 2009; Iyer-Pascuzzi and Benfey 2009). Phytohormone auxin plays an important role in many of the plant's development processes including root development. Auxin and its gradient across the root longitudinal axis and the radical axis is implicated in many aspects of root growth and development like embryonic root formation, maintenance of RAM, cell division and expansion, tropisms, vascular differentiation, lateral root initiation, etc.

(reviewed in (Overvoorde et al. 2010; Petricka et al. 2012b; Woodward and Bartel 2005). In VSR, a total of 29 genes involved in auxin biosynthesis and signaling were induced or up-regulated. Seven genes that belong to the auxin-responsive SAUR gene family and another seven genes coding for the AUX/IAA proteins IAA8, IAA17, IAA18, and IAA26 were induced in VSR. An auxin receptor, auxin binding protein 1 (ABP1), which controls cell elongation and cell division, and an F-Box protein, auxin signaling F-BOX 3 (AFB3), which is a component of SCF (ASK-cullin-F-box) E3 ubiquitin ligase complexes and mediate the proteasomal degradation of Aux/IAA proteins, were up-regulated in the VSR. Two Auxin Responsive Factor (ARF) genes homologous to AtARF19 were also up-regulated in the VSR. An auxin efflux carrier protein (TraesCS5D01G361600) involved in polar auxin transport was induced in the VSR. Apart from these, a flavin-binding monooxygenase/ indole-3-pyruvate monooxygenase similar to AtYUCCA11 that catalyzes the indole-3-pyruvic acid (IPyA) to indole-3-acetaldehyde step in auxin biosynthesis pathway, two genes encoding for enzyme that catalyzes the synthesis of indole-3-acetic acid (IAA)-amino acid conjugates, GH3.2/YDK1 and two genes coding for IAA beta-glucosyl transferase (IAGLU) were up-regulated in VSR root tips. Two genes that encode IAA-amido hydrolase homologs of Arabidopsis IAA-leucine-resistant (ILR1) and IAA-leucine-resistant (ILR1)-like 3 (ILL3), which hydrolyzes amino acid conjugates of IAA were induced in the VSR root tips. In addition to this, two genes belonging to a detoxification enzyme that protects the cells against auxin-induced oxidative stress, flavodoxin-like quinone reductase 1 (FQR1) was also induced in the VSR root. All this indicated that the auxin pathway is overall up-regulated in the VSR.

Ethylene (ET) is produced in the plant cells from the amino acid methionine and an intermediate S-adenosylmethionine (SAM) which is part of the Yang cycle. The first committed and the rate-limiting step of the ethylene biosynthesis is the conversion of SAM to 1-aminocyclopropane-1-carboxylic acid (ACC) by the enzyme ACC synthase. ACC is then converted to ethylene by the enzyme ACC oxidase. We found two genes encoding for the enzyme ACC synthase and seven genes encoding for the enzyme ACC oxidase, including two homologs of AtACO1 and one homologous to OsACO2, were up-regulated in the VSR. Both ACC synthase genes were orthologous to the Arabidopsis ACC synthase 6 (ACS6). Two of the seven ACC oxidases up-regulated in VSR are orthologous to rice and Arabidopsis ACC oxidase 1, (ACO1) and the other five are the rice ACC oxidase like homolog genes. Three genes that encode for ethylene receptor or sensor, Ethylene Insensitive 1 (EIN1), EIN2 and EIN4 in Arabidopsis, are induced in VSR root tips. Also, 13 genes that encode ethylene response factors (ERFs), which are members of the AP2 family of transcription factors, were induced in the VSR root tips. However, two of the genes homologus to rice ACO homolog 4 were down-regulated in VSR. These results indicated that the gaseous phytohormone ET biosynthesis and its overall signaling pathways were induced in the VSR root tips suggesting a major role of ethylene in VSR development.

Gibberellin or gibberellic acid (GA) is an essential phytohormone and involved in the regulation of plant growth and development. It is also known to regulate root growth and development. Gibberellin Insensitive Dwarf 1-like 2 (GID1L2) which are alpha/beta hydrolases and function as GA receptors. Five genes that encode homologs of rice GID1L2 were induced in the VSR. In VSR root tips, we also observed drastic up-regulation (8-

fold) of the GA catabolism enzyme gibberellin 2-oxidase similar to ATGA2OX1 and GA20OX1. A gene coding for the enzyme Ent-kaurenoic acid oxidase 1 (KAO1), which catalyzes three successive oxidations of ent-kaurenoic acid producing gibberellin 12 (GA12), a key step in GAs biosynthesis, was induced in the VSR. These results indicate a finely tuned role of GA in the manifestation of VSR.

Cytokinins (CK) act negatively on the root growth (Werner et al. 2003; Yang et al. 2003). In this study, we observed that 11 genes coding for the enzyme cytokinin-N-glucosyltransferase, which catalyze the formation of CK-N-glucoconjugates and inactivate the CK, were induced in VSR, of which TraesCS6A01G018000 and TraesCS5B01G500700 were up-regulated by 15- and 16-fold, respectively. Interestingly, three genes coding for proteins that are homologous to Arabidopsis Histidine Kinase (AHK3) and one gene for AHK5, the CK receptors, and a gene for the negative regulator of CK signaling called the A-type response regulator (RR) and similar to Arabidopsis ARR12, were induced in the VSR root tips.

Brassinosteroids (BRs) are the plant hormones derived from campesterol and are involved in various plant growth and development processes. The BR signals are perceived by membrane-anchored leucine-rich repeat receptor-like kinases (LRR-RLKs) BRASSINOSTEROID-INSENSITIVE 1 (BRI1) (He et al. 2000) and SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 (SERK1) (Karlova et al. 2006). BRs are involved many aspects of root development (Wei and Li 2016). In the VSR, 12 BR signaling genes were induced. These include two genes coding for the homologs of BRI1-LIKE 1(*BRL1*), and six genes homologous to SERK1, two genes homologous to the *BR*-

SIGNALING KINASE 1 (*BSK1*), a member of membrane-bound receptor-like cytoplasmic kinases (RLCKs), and a BR-responsive RING-H2 or *BRH1*.

Jasmonic acid (JA) belongs to the family of oxylipins and are involved in defense responses and the restriction of plant organ growth. Lipoxygenases (LOX) are the enzymes that facilitate the conversion of the fatty acid α-linolenic acid to the hydroperoxides and are part of the JA biosynthetic pathway. Root growth is inhibited when exogenous JA is applied. In the VSR root tips, five genes coding for LOX1, three genes for LOX3 and three genes for LOX4 were induced. Another enzyme in JA biosynthesis pathway, 4-coumarate-CoA ligase-like 9, converts 12-oxo-phytodienoic acid (OPDA) into OPDA-CoA was also induced in the VSR. The enzyme allene oxide synthase (AOS) is a cytochrome P450 belonging to the class CYP74A that catalyzes the conversion of 13-hydroperoxides to allene oxides, a key step in JA biosynthesis pathway. We found that three of the Jasmonic acid-amido synthetase (*JAR1*) genes were induced in the root tips of the VSR The JAR1 genes belong to the GH3 family and code for a protein which catalyzes the synthesis of jasmonate-amino acid conjugates by adenylation and is induced by auxin. In VSR root tips, six genes coding for the cytochrome P450 proteins were induced. Apart from this, two protein kinases, which are induced in response to JA, salicylic acid (SA), pathogen infection and wounding, were also up-regulated in VSR root tips. The Jasmonate ZIM-domain (JAZ) proteins are the TIFY transcriptions factors and negative regulators of the JA signaling and act by repressing the JA responsive genes in the absence of the JA. In VSR root tips, three genes that are homologs of *JAZ1*, one gene for *JAZ10* and three of *JAZ12* were induced. Along with these, a gene in each of the family transcriptional repressors and negative regulators of JA responses, i.e., NINJA and TOPLESS-related 2,

were also induced in the VSR. Two genes that encode for UDP-glucosyltransferases that catalyze the formation of both SA 2-O-beta-D-glucoside (SAG) and SA glucose ester (SGE) were induced in VSR root tips. Three genes that encode for the 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase/AtDMR6 which converts SA to 2,3-dihydroxybenzoic acid (2,3-DHBA) were induced in the VSR. And also, there were five genes homologous to AtNPR3, a homolog of AtPAD4, three homologs of DMR6 and two AP4.3A protein kinases with elevated expression in the VSR root tips. These results indicated the elevated JA synthesis and responsive genes along with SA responses in the VSR.

Abscisic acid (ABA) is an important phytohormone for abiotic stress response and negatively regulate the root growth (Yang et al. 2014). In the VSR root tips, several ABA-related genes were found to be induced. Among the ABA-responsive genes were five genes belonging to the Abscisic acid-responsive (TB2/DP1, HVA22) protein family, several of the GRAM domain-containing proteins, two genes for ABA responsive element-binding factor 1 (ABF1). Apart from these responsive genes, two genes that encode regulatory components of ABA receptor 3 (RCAR3/PYL8), the receptors for ABA along with a protein phosphatase 2C (PP2C) gene a homolog of highly ABA-induced PP2C gene 2 (HAI2) were also induced in the VSR root tips.

We observed the elevated expression of the receptors for several plant hormones and their responsive genes. Also, several enzymes that take part in their biosynthetic pathways in the VSR root tips. These results suggest that the VSR phenotype is a result of concerted interaction between the seven phytohormones in these root tips and their effect on the root meristem and the transition zone.

Reactive Oxygen Species-related genes

ROS plays an indispensable role in the regulation of plant development. In plants, ROS are generated from molecular oxygen by reduction. Singlet oxygen ($^1O_2$), hydroxyl radical ($HO^{\cdot}$), hydrogen peroxide (H_2O_2), and superoxide (O_2^-) are the ROS found in plants and are generated by reduction reactions in peroxisomes, chloroplast, and mitochondria. Many plant metabolic processes like photosynthesis, respiration, etc. were known to generate ROS as a part of their process. In VSR root tips, 184 genes encoding for several enzymes involved in either ROS generation like respiratory burst oxidase homolog B (RBOHB), RBOHD, and RBOHF or the antioxidant systems like peroxidase (POD), catalase (CAT), glutathione S-transferase (GST), ascorbate peroxidase (APX), etc. Of these, 75 genes belonging to the GST gene family were up-regulated in the VSR root tips. The GST genes belonging to Tau (46), Phi (16), Zeta (6), lambda (3) and microsomal (3) were found to be induced in the VSR root tips. Interestingly, one GST gene belonging to the Tau class was the only GST gene downregulated in the VSR. Another ROS scavenging gene family that was induced in the VSR is the POD gene family. We found 64 POD genes were induced in the VSR. Of these 64 POD genes up-regulated, 20 genes code for the homolog of Arabidopsis RARE COLD INDUCIBLE 3 (RCI3) gene. We also found 20 POD genes were down-regulated in the VSR. Apart from these, two genes that code for glutathione synthetase, an enzyme that catalyzes the ligation of glycine to the γ-glutamylcysteine in the biosynthesis of glutathione were also induced in the VSR. One gene coding for the catalase (CAT) and three monodehydroascorbate reductases (MDAR) genes were induced. Three genes that encode a peptide met sulfoxide reductase 4 (PMSR4) were also induced in the

VSR. Of the genes that are involved in the ROS generation, four genes belonging to the respiratory burst oxidase homolog D (RBOHD) and two RBOHF genes were induced in the VSR root tips. Together we found both genes for ROS generation and genes for ROS scavenging are induced or upregulated in VSR, indicating ROS played an important role in the development of the VSR phenotype.

Lignin biosynthesis gene families

Lignin is one of the major components of the plant secondary cell wall, and its biosynthesis in plants is a multi-step pathway involving multiple enzymes. In VSR root tip transcriptome, we surprisingly found 56 genes that encode several enzymes involved in the biosynthesis of lignin were up-regulated (Figure 3.4). Of these, 14 genes code for the enzyme phenylalanine ammonia lyase 1 (PAL1) that catalyzes the first step of the phenylpropanoid pathway. Thirty-one genes coding for the other enzymes involved in the lignin biosynthesis pathway that were up-regulated in the VSR. They are cinnamic acid 4-hydroxylase or C4H (4), 4-coumarate-coenzyme A ligase or 4CL (1), caffeoyl coenzyme A 3-O-methyltransferase 1 or CCoAoMT (3), cinnamoyl-CoA reductase or CCR (8), cinnamyl alcohol dehydrase or CAD (3), hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase or HCT (10). In addition to these biosynthetic enzymes, 12 laccase (*LAC*) genes were also up-regulated. The *LAC* genes are involved in the polymerization of the lignin, and the 12 *LAC* genes belong to LAC5, 7, 12, and 17 classes.

Transcription factors

Transcription factors (TFs) play an important role in the plant growth and development by regulating several processes. In line with this role, we found 260 genes encoding for several TF families were differentially expressed in the VSR root tip compared to the root tips of the long root. Of these 260 TFs differentially expressed, 245 genes were induced, and 17 were down-regulated in the VSR suggesting there is a major transcriptomic reprogramming occurred to manifest this phenotype. Of all the TF families differentially expressed in VSR, the WRKY TF family was highly enriched. There were 40 genes that code for WRKY TFs were induced in the VSR root tips and followed by NAC (29), bZIP (21), AP2 (21), MYB (18), bHLH (20), Zinc Finger (ZF) (17), GRAS (17), Homeobox (17) and TIFY (10). The genes belonging to TF families bHLH (4), MYB (3), B3 (2), ARF (1), bZIP (1), E2F/DP (1) and ZF (1) were the only genes down-regulated in the VSR (Table 3.2 and Figure 3.5). Interestingly, all the members of the TF familes belonging to ASL/LOB (6), CAMTA (5), GATA (4), GRAS (17), Homeobox (17), NF-YA (2), Ovate (2), PLATZ (3), RWP-RK (1), SBP (3), TIFY (10), TUBBY (3), VOZ (2), and WRKY (40) were only up-regulated in the VSR root tips.

Cell proliferation and cell death related genes

Cell death is part of the growth and development of plants. In the VSR root tips, genes that encode for proteins belonging to the cell death and autophagy, as well as regulators of cell division, cell elongation, and cell proliferation, were also up regulated. We found several genes belonging autophagy like Autophagy-related proteins (ATG) including ATG2, ATG6, ATG9, ATG13, ATH18H, vacuolar processing enzymes (VPE), cysteine rich

receptor like kinases (CRKs), etc. were induced in the VSR root tips. In addition to these, several genes involved in proteasome degradation pathway, disease resistance, and defense response were also up-regulated in VSR. Surprisingly, three genes encoding each of the BAX Inhibitor 1 and BAX Inhibitor like 4 (BIL4) which are inhibitors of the programmed cell death in plants were induced in the VSR root tips.

In the VSR root tips, we found genes that are involved in cell proliferation and elongation up-regulated. Cysteine rich small peptides called RAPID ALKALINIZATION FACTORS (RALFs), specifically RALF33, and their receptor FERONIA, a receptor kinase were induced in the VSR. We also found two genes coding for the E3 ligase BIG BROTHER induced in the VSR root tips.

Histochemical straining of root tips for ROS and lignin

Expression of 184 ROS-related genes was increased in VSR root tips (Table S3.1), suggesting that ROS are involved in the development of the VSR phenotype. We tested this hypothesis by staining root tips with DAB, which detects H_2O_2, and found that VSR of the F_1 hybrid accumulated much more H_2O_2 in the tips of primary and lateral roots compared to the long roots of its parents CS and CS-*Vsr1b* (Figure 3.6). This result indicated that *Vsr1* enhanced ROS production, and accumulated ROS pose an oxidative stress to root growth.

Our transcriptome quantification showed that transcription of 56 lignin biosynthetic genes was increased in VSR (Table S3.1). Thus, we conducted a lignin assay of the root tips by pholoroglucinol staining. The result showed that no lignin is detected in LR tips at all the time points (Figure 3.7a and b). No lignin was detected in VSR tips 1 d after

germination (Figure 3.7c), but lignin deposition was found in VSR tips 2 d after germination (Figure 3.7d). Root tips bended and "hooked up" 3 d after germination (Figure 3.7e). The root "hook" was most probably due to asymmetrical deposition of lignin within root tip, more lignin was found on the inner side of the "hook" (Figure 3.7e and f). The cells in the outer side of the "hook" are significantly longer than those in the inner side, and cells above the bent are much longer than those below the curve in the same profiles (Figure 3.7f). This indicates that root tips bended in the elongation zone due to the ectopic lignification of the cell wall, which limited root elongation more severely in the inner side than in the outer side.

Small RNA sequencing and differential expression analysis

We sequenced the six libraries on the Hiseq 2000. We generated ~28 to ~38 million reads per replicate of LR and VSR small RNA libraries. After cleaning the raw reads using our pipeline and removing the reads corresponding to rRNA, tRNA, snRNA, snoRNA and scRNA were removed. Additionally, reads of length smaller than 15 bp and longer than 25bp were filtered. This resulted in ~7 to ~12 million reads per replicate. The reads were counted and merged by sequence similarity and were annotated using the miRBase v 22 and the miRNAs predicted in the wheat reference sequence. We identified 153 miRNAs were differentially expressed in the VSR root tips. Of these, 67 were up-regulated and 86 were down-regulated. The top five up-regulated miRNAs were miR1135-3p-271 (~535-fold), miR159b-1 (~230-fold), miR1439-3p-6 (~149-fold), miR5721 (~141-fold), and mir1121-3p-53 (~123-fold) and 33 miRNAs were induced at least 10-fold in VSR root tips (Fig. 3.8). The top five down-regulated (~2-fold) miRNAs were miR159b isoform,

miR319b, miR5048a/miR5048b, 319b, miR159b-2/miR159g. We found members of the miR159 family both induced and repressed in the VSR root tips. The miR159 targets MYB TFs and is known to repress these TFs during vegetative and reproductive development in Arabidopsis. The miR159 regulates the GAMYBs and inhibit the growth and promotes programmed cell death in Arabidopsis. Members of the miR159 family were functionally redundant (Allen et al. 2007; Alonso-Peral et al. 2010; Alonso-Peral et al. 2012). The up-regulated members of the miR159 family in the VSR root tips were induced at least 8-fold to 230-fold whereas the down-regulated were at 2-fold level (Table S3.2). In addition to miR159, another miRNA known to regulate root meristem size and thus root development is miR396 (Rodriguez et al. 2015), which also showed different expression patterns among its members. There were four members of the miR396 family were differentially expressed and two were induced and two were repressed in VSR. The miRNA family miR167, which regulates the *ARF*s and control lateral root formation (Gutierrez et al. 2012), was induced in the VSR root tips. One member each of miR394, miR827, miR5062a, miR5083, miR9652, miR9778, and miR9654b families and all the members of miR1436, miR1128, miR1122, miR166, miR319b, miR5048, miR9658, miR9664, miR9674, and miR9772 families were down regulated in the VSR root tips. Whereas, all members of miRNA1121, miR1125, miR1127b, miR1432, miR1439, miR393, miR408d, miR531, and miR9776 and one member each of miR528, miR5721, miR397, miR1137-3p-153, and miR1127b were induced in the VSR root tips.

DISCUSSION

Roots are a plant organ that grows underground and is indispensable in plant's growth and development. They play a crucial role in the acquisition of water and nutrients and support the growth of aerial parts and provide mechanical support by forming an anchor. They form a determinant of the plant's potential to grow and successfully establish in the highly variable soil environment. A well established and healthy root system is identified as one of the key contributors to the maximizing yield in many plant models (Gechev et al. 2006; Glover et al. 2007). The exploration of root biology is far behind as compared to above-ground structures. Despite the growing knowledge of genetic and molecular mechanisms of root development in dicot model Arabidopsis, very little information is known in cereal crops. Owing to lack of a complete and fully annotated genome sequence and the complexity of the polyploid nature, the knowledge of genes and gene networks involved in root development in common wheat is limited. In recent years, studies at transcriptome level using the latest high throughput sequencing methods were carried out to understand molecular mechanisms involved in wheat root response to environment stress (Camilios-Neto et al. 2014; Oono et al. 2013). But, no research has been performed to investigate the transcriptome effect of a gene locus on root development. In this study, we showed the transcriptome level changes in VSR and showed that VSR development is different from other short root phenotypes in the model plants, providing new insight into wheat root development and growth.

The VSR phenotype is thought to be a result of non-additive interaction between the parental genomes, and the phenotype is only observed when there is a heterozygous

composition at this locus (Li et al. 2013). We selected the pooled segregants for the analysis of the transcriptome using the high throughput sequencing of RNA. The pooled segregants were derived from the F_2 population of the backcross (BC_2) generation. The samples used in this study represent closely to the near-isogenic lines as the background genome is normalized by pooling ~40 root tips per replicate. Thus, the samples for the LR and VSR share a nearly identical background genome composition except for variation in the *Vsr1* locus. We observed 3,566 genes induced and 604 genes down regulated in the VSR root tips (encompassing the meristem and the elongation zone). We were able to annotate 3,561 up-regulated and 602 down-regulated genes with at least one hit from Arabidopsis or rice or NCBI-nr protein databases. The remaining 300 the up- and 68 down-regulated genes were annotated as uncharacterized or domain of unknown function suggesting that there is a need for functional characterization of such genes in model plant systems.

In understanding the transcriptome level reprogramming that occurred in the root tips of the hybrid plants and manifestation of the VSR phenotype, we identified only a minute portion (1.55%) of the transcriptome expressed differentially. The DEGs in VSR root tips can be categorized into response to hormones, response to pathogen and defense, ROS generation and scavenging, regulation of transcription, transport, signaling, lignin biosynthesis, flavonoid metabolism, cell wall modification, cell wall degradation, primary and secondary metabolism, regulation of cell cycle and cell division, cell death and autophagy, protein trafficking, protein degradation etc.

Many transcription factors were induced in the VSR root tips (Figure 3.5) suggesting a major reprogramming in the gene expression profiles. However, TFs like PLETHORA1 (PLT1), PLT2, PLT3, BABY BOOM (BBM) SHORT ROOT (SHR),

SCARECROW (SCR), MAGPIE (MGP), JACKDAW (JKD), PHABULOSA (PHB) which play major role in the stem cell regulation and proper formation of the formation of root (Aida et al. 2004; Cruz-Ramirez et al. 2012; Di Laurenzio et al. 1996; Galinha et al. 2007; Helariutta et al. 2000; Moreno-Risueno et al. 2015; Prigge et al. 2005; Sabatini et al. 2003; Welch et al. 2007; Zhong et al. 1999), were unperturbed in the VSR. TFs of the families NAC and MYB were identified as the regulators of the lignification and the secondary cell wall formation in Arabidopsis (Wang and Dixon 2012; Zhong et al. 2010). In VSR, 31 members of the NAC TF family were differentially expressed along with 21 members of the MYB TF family. Along with these TFs, several enzymes in the lignin biosynthetic pathway were induced in the VSR root tips (Figure 3.4; Table S3.1). And in our study of the transcriptome analysis of the root tips verses mature root transcriptome in the previous chapter, we noticed that activation of the lignin biosynthetic enzymes was found in the mature root rather than the root tips. Typically, the lignification of the cell walls and the secondary cell wall formation is not observed in the root tips and on the contrary, cell division and the cell elongation related genetic pathways are expected to play a major role in this zone (Somssich et al. 2016). In together with the transcriptome profiles of TFs and lignin biosynthesis pathway genes and the ectopic lignin deposition observed in the VSR root tips suggests the VSR phenotype is due to the premature stabilization of the cell walls in the meristematic zone.

In the VSR root tips, a number of genes belonging to the all the plant hormones, including ABA, auxin, BR, CK, ET, GA, and JA, were differentially expressed and mostly up-regulated (Table S3.1). This situation suggests the *Vsr1* influences the transcription of the regulators that induce or repress hormone responsive gene expression. Auxin is one of

the crucial plant hormone implicated in the root development. Though we didn't find any auxin biosynthesis genes differentially expressed in the VSR root tips, several genes that belong to gene families involved in the auxin sensing, signaling and homeostasis were differentially expressed (Table S3.1). Along with auxin, several hormones related gene expression and their interactions with other hormone responses orchestrate the regular maintenance of the root meristem and the stem cell niche and proper transition into the elongation and differentiation (Benkova and Hejatko 2009; Luijten and Heidstra 2009; Moreno-Risueno et al. 2015; Petricka et al. 2012a; Petricka et al. 2012b).

ROS such as super-oxide anion ($^{\bullet}O2^{-}$), hydrogen peroxide (H2O2), etc., are the by-products continuously generated during normal metabolic processes in plants, such as photosynthesis, photorespiration, and cellular respiration. ROS at an optimal level is required for the plant development and acts a secondary signaling molecule and help the plants in their response to external stimuli. At levels higher than the optimal, ROS is very detrimental to the plants (Foreman et al. 2003). In VSR root tips, we observed high accumulation of the ROS evident from DAB staining and also correlated with the induction of a large number of genes that code for the ROS scavenging mechanisms (Foyer and Noctor 2013). ROS also acts as a signal for the programmed cell death in plant cells (Gechev et al. 2006). Related to this, we also detected DEGs in the autophagy pathway. Recent studies in Arabidopsis found crosstalk between hormones, such as ABA and BRs, and ROS signaling (Yang et al. 2014; Lv et al. 2018). Our results on VSR found signification perturbation both in phytohormones and ROS, suggesting a potential interaction between them in VSR development

In addition to dynamics of over 200 TF genes, the hormonal signaling, lignin, ROS and autophagy pathways, (Fig 3.5), expression in genes coding protein kinases such as FERONIA (FER), ubiquitin E3 ligases such as BIGBROTHER (BB), and transporter proteins was up-regulated in VSR (Table S3.1). Perturbation of these pathways eventually down regulate cell proliferation and cell elongation as evidenced by down regulation of cell division machinery protein genes such as histones, In the model plant Arabidopsis, BB represses cell proliferation (Li and Baven 2004), FER interact with peptide hormone RALF to inhibit cell elongation (Haruta et al. 2014). We also observed up-regulation of RALF homolog of wheat in VSR root tips. In Arabidopsis, ROS also can activate lignin deposition (Denness et al. 2011). We also observed over accumulation of ROS in VSR root tips (Figure 3.6.) and deposition of lignin in the elongation zone (Figure 3.7). Based on our results and publications from the model plant Arabidopsis, we propose a model to explain the VSR development (Figure 3.9). In this model, we hypothesize that the *Vsr1* gene is expressed only in *Vsr1aVsr1b* heterozygote, the *Vsr1* gene product interacts with other proteins, i.e., the Vsr1-interacting proteins (VIP s), and the VIPs participate in over-activation of the hormonal signaling, ROS accumulation, lignification, and protein kinase Feronia (FER) and BigBrother (BB) E3 ligase. The coordinated expression of these genetic pathways eventually suppresses cell proliferation and elongation, leading to the VSR phenotype (Figure 3.9).

Recently, several studies provided the evidence for miRNA's role in root development. In VSR root tips, several isomiRs of miR167 were upregulated. miR167 targets TFs *ARF6* and *ARF8* along with *IAA-ALA-RESISTANT 3* (*IAR3*) in Arabidopsis (Gutierrez et al. 2012; Kinoshita et al. 2012). Interestingly, miR167 represses the *IAR3*

which represses the primary root (Kinoshita et al. 2012). miR319 was reported to repress cell proliferation in leaf development by targeting TCP4 (Schommer et al. 2014). In VSR four members of miR319 were down-regulated. miR159, one of the most abundant miRNA in plants, that targets GA related MYB TFs and inhibits growth and promotes cell death was recently identified as the repressor of root growth (Alonso-Peral et al. 2010; Alonso-Peral et al. 2012; Xue et al. 2017). The meristem size in the *miR159ab* mutants in Arabidopsis was increased and was thought to be acting on the targets related to the cell cycle progression (Xue et al. 2017). Though several isomiRs of miR159 were repressed, the induced isomiRs in the VSR root tips were up-regulated at much higher level (~230-fold induction). Also, miR159 was shown to act independent of the abundance levels in Arabidopsis seeds (Alonso-Peral et al. 2012). miR396 was reported to regulate the root development by reducing the cell cycle time in the root tissue by interacting with the *GRF*s which in turn repress *PLT*s (Bazin et al. 2013; Rodriguez et al. 2015). Interestingly, two isoforms of miR396 were induced and two were repressed in the VSR. The bidirectional expression patterns among the miR396 members may suggest fine tuning function of these members in different cell types during the development of VSR phenotype. Lack of knowledge on the direct role of other differentially expressed miRNAs in the regulation of the root development makes it more complicated to understand VSR phenotype.

We have shown that VSR expression is not beneficial to root growth and plant development. But why the *Vsr1* gene is still present in the wheat population? We hypothesize that the *Vsr1* gene is important for some developmental processes by triggering cell death, but it is mis-activated in root tips of the F_1 hybrid between CS and TA. The Vsr1-dependenr expression profiles could be due to a combination of the

regulatory elements from the two parental genomes or modification of epigenetic status in the regulatory regions.

VSR represents an important and interesting mechanism of gene expression regulation. Although we can imagine many fascinating and possible scenarios but still have no idea how the regulation of *Vsr1* expression and *Vsr1* activation of those genetic pathways that suppress cell proliferation and elongation. To reveal the complete VSR mechanism, we will need to clone the *Vsr1* gene. Despite the availability of the wheat reference genome from CS, we still need the sequence from the corresponding region of the TA genome. To this end, the 5D chromosome may be sorted from TA and sequenced using the long-read technology, such as Nanopore (Oxford Nanopore Technologies, Oxford Science Park, UK) and aligned with the CS reference genome to identify the variation. At the same time, the root tip transcriptome from the VSR also needs to be sequenced using the long-read technologies to a much deeper level or using sequence capture strategy to increase the sequence depth as the expression of the *Vsr1* locus may be very low or coding for non-coding RNA. The *Vsr1* candidate gene identified in this way can be validated by traditional genetic complementation or genome editing such CRSIPR/Cas9, which can cause DNA deletions from 1-bp to several hundred kb (Weeks et al. 2015).

REFERENCES

Aida M, Beis D, Heidstra R, Willemsen V, Blilou I, Galinha C, Nussaume L, Noh YS, Amasino R, Scheres B (2004) The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. Cell 119:109-120

Allen RS, Li J, Stahle MI, Dubroue A, Gubler F, Millar AA (2007) Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. Proc Natl Acad Sci U S A 104:16371-16376

Alonso-Peral MM, Li J, Li Y, Allen RS, Schnippenkoetter W, Ohms S, White RG, Millar AA (2010) The microRNA159-regulated GAMYB-like genes inhibit growth and promote programmed cell death in Arabidopsis. Plant Physiol 154:757-771

Alonso-Peral MM, Sun C, Millar AA (2012) MicroRNA159 can act as a switch or tuning microRNA independently of its abundance in Arabidopsis. PLoS One 7:e34751

Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166-169

Bazin J, Khan GA, Combier JP, Bustos-Sanmamed P, Debernardi JM, Rodriguez R, Sorin C, Palatnik J, Hartmann C, Crespi M, Lelandais-Briere C (2013) miR396 affects mycorrhization and root meristem activity in the legume Medicago truncatula. Plant J 74:920-934

Benkova E, Hejatko J (2009) Hormone interactions at the root apical meristem. Plant Mol Biol 69:383-396

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120

Camilios-Neto D, Bonato P, Wassem R, Tadra-Sfeir MZ, Brusamarello-Santos LC, Valdameri G, Donatti L, Faoro H, Weiss VA, Chubatsu LS, Pedrosa FO, Souza EM (2014) Dual RNA-seq transcriptional analysis of wheat roots colonized by Azospirillum brasilense reveals up-regulation of nutrient acquisition and cell cycle genes. BMC Genomics 15:378

Coudert Y, Perin C, Courtois B, Khong NG, Gantet P (2010) Genetic control of root development in rice, the model cereal. Trends Plant Sci 15:219-226

Cruz-Ramirez A, Diaz-Trivino S, Blilou I, Grieneisen VA, Sozzani R, Zamioudis C, Miskolczi P, Nieuwland J, Benjamins R, Dhonukshe P, Caballero-Perez J, Horvath B, Long Y, Mahonen AP, Zhang H, Xu J, Murray JA, Benfey PN, Bako L, Maree AF, Scheres B (2012) A bistable circuit involving SCARECROW-RETINOBLASTOMA integrates cues to inform asymmetric stem cell division. Cell 150:1002-1015

Dai X, Zhuang Z, Zhao PX (2018) psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res

Denness L, McKenna JF, Segonzac C, Wormit A, Madhou P, Bennett M, Mansfield J, Zipfel C, Hamann T (2011) Cell wall damage-induced lignin biosynthesis is regulated by a reactive oxygen species- and jasmonic acid-dependent process in Arabidopsis. Plant Physiol 156:1364-1374

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN (1996) The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. Cell 86:423-433

Dinneny JR, Benfey PN (2008) Plant stem cell niches: standing the test of time. Cell 132:553-557

Dixon J, Braun H-J, Kosina P, Crouch J (2009) Wheat Facts and Futures 2009. CIMMYT, Mexico, p 105

Dolan L, Janmaat K, Willemsen V, Linstead P, Poethig S, Roberts K, Scheres B (1993) Cellular organisation of the Arabidopsis thaliana root. Development 119:71-84

Foreman J, Demidchik V, Bothwell JH, Mylona P, Miedema H, Torres MA, Linstead P, Costa S, Brownlee C, Jones JD, Davies JM, Dolan L (2003) Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. Nature 422:442-446

Foyer CH, Noctor G (2013) Redox signaling in plants. Antioxid Redox Signal 18:2087-2090

Galinha C, Hofhuis H, Luijten M, Willemsen V, Blilou I, Heidstra R, Scheres B (2007) PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. Nature 449:1053-1057

Gechev TS, Van Breusegem F, Stone JM, Denev I, Laloi C (2006) Reactive oxygen species as signals that modulate plant stress responses and programmed cell death. Bioessays 28:1091-1101

Glover JD, Cox CM, Reganold JP (2007) Future farming: a return to roots. Sci Am 297:82-89

Gutierrez L, Mongelard G, Flokova K, Pacurar DI, Novak O, Staswick P, Kowalczyk M, Pacurar M, Demailly H, Geiss G, Bellini C (2012) Auxin controls Arabidopsis

adventitious root initiation by regulating jasmonic acid homeostasis. Plant Cell 24:2515-2527

Haruta M, Sabat G, Stecker K, Minkoff BB, Sussman MR (2014) A peptide hormone and its receptor protein kinase regulate plant cell expansion. Science 343:408-411

He Z, Wang ZY, Li J, Zhu Q, Lamb C, Ronald P, Chory J (2000) Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. Science 288:2360-2363

Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, Jung J, Sena G, Hauser MT, Benfey PN (2000) The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling. Cell 101:555-567

Inukai Y, Sakamoto T, Ueguchi-Tanaka M, Shibata Y, Gomi K, Umemura I, Hasegawa Y, Ashikari M, Kitano H, Matsuoka M (2005) Crown rootless1, which is essential for crown root formation in rice, is a target of an AUXIN RESPONSE FACTOR in auxin signaling. Plant Cell 17:1387-1396

Iyer-Pascuzzi AS, Benfey PN (2009) Transcriptional networks in root cell fate specification. Biochim Biophys Acta 1789:315-325

Jiang H, Lei R, Ding SW, Zhu S (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15:182

Jiang K, Feldman LJ (2005) Regulation of root apical meristem development. Annu Rev Cell Dev Biol 21:485-509

Karlova R, Boeren S, Russinova E, Aker J, Vervoort J, de Vries S (2006) The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 protein complex includes BRASSINOSTEROID-INSENSITIVE1. Plant Cell 18:626-638

Kihara H (1944) Discovery of the DD-analyser, one of the ancestors of Triticum vulgare (Japanese). Agric and Hort (Tokyo) 19:13 -14

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357-360

Kinoshita N, Wang H, Kasahara H, Liu J, Macpherson C, Machida Y, Kamiya Y, Hannah MA, Chua NH (2012) IAA-Ala Resistant3, an evolutionarily conserved target of miR167, mediates Arabidopsis root architecture changes during high osmotic stress. Plant Cell 24:3590-3602

Kitomi Y, Ogawa A, Kitano H, Inukai Y (2008) CRL4 regulates crown root formation through auxin transport in rice. Plant Root 2:19-28

Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39:D152-157

Li W, Zhu H, Challa GS, Zhang Z (2013) A non-additive interaction in a single locus causes a very short root phenotype in wheat. Theor Appl Genet 126:1189-1200

Liu H, Wang S, Yu X, Yu J, He X, Zhang S, Shou H, Wu P (2005) ARL1, a LOB-domain protein required for adventitious root formation in rice. Plant J 43:47-56

Liu S, Wang J, Wang L, Wang X, Xue Y, Wu P, Shou H (2009) Adventitious root formation in rice requires OsGNOM1 and is mediated by the OsPINs family. Cell Res 19:1110-1119

Luijten M, Heidstra R (2009) Arabidopsis Root Development. Annual Plant Reviews Volume 37: Root Development

Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, McGuire PE, Liu S, Long H, Ramasamy RK, Rodriguez JC, Van SL,

Yuan L, Wang Z, Xia Z, Xiao L, Anderson OD, Ouyang S, Liang Y, Zimin AV, Pertea G, Qi P, Bennetzen JL, Dai X, Dawson MW, Muller HG, Kugler K, Rivarola-Duarte L, Spannagl M, Mayer KFX, Lu FH, Bevan MW, Leroy P, Li P, You FM, Sun Q, Liu Z, Lyons E, Wicker T, Salzberg SL, Devos KM, Dvorak J (2017) Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. Nature 551:498-502

McFadden ES, Sears ER (1946) The origin of Triticum spelta and its free-threshing hexaploid relatives. J Hered 37:81 107

Moreno-Risueno MA, Sozzani R, Yardimci GG, Petricka JJ, Vernoux T, Blilou I, Alonso J, Winter CM, Ohler U, Scheres B, Benfey PN (2015) Transcriptional control of tissue formation throughout root development. Science 350:426-430

Oono Y, Kobayashi F, Kawahara Y, Yazawa T, Handa H, Itoh T, Matsumoto T (2013) Characterisation of the wheat (Triticum aestivum L.) transcriptome by de novo assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. BMC Genomics 14:77

Overvoorde P, Fukaki H, Beeckman T (2010) Auxin control of root development. Cold Spring Harb Perspect Biol 2:a001537

Perilli S, Di Mambro R, Sabatini S (2012) Growth and development of the root apical meristem. Curr Opin Plant Biol 15:17-23

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650-1667

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290-295

Petricka JJ, Schauer MA, Megraw M, Breakfield NW, Thompson JW, Georgiev S, Soderblom EJ, Ohler U, Moseley MA, Grossniklaus U, Benfey PN (2012a) The protein expression landscape of the Arabidopsis root. Proc Natl Acad Sci U S A 109:6811-6818

Petricka JJ, Winter CM, Benfey PN (2012b) Control of Arabidopsis root development. Annu Rev Plant Biol 63:563-590

Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE (2005) Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. Plant Cell 17:61-76

Richards RA (2008) Genetic Opportunities to Improve Cereal Root Systems for Dryland Agriculture. Plant Production Science 11:12-16

Richards RA, Rebetzke GJ, Condon AG, van Herwaarden AF (2002) Breeding Opportunities for Increasing the Efficiency of Water Use and Crop Yield in Temperate Cereals. Crop Sci 42:111-121

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140

Rodriguez RE, Ercoli MF, Debernardi JM, Breakfield NW, Mecchia MA, Sabatini M, Cools T, De Veylder L, Benfey PN, Palatnik JF (2015) MicroRNA miR396

Regulates the Switch between Stem Cells and Transit-Amplifying Cells in Arabidopsis Roots. Plant Cell 27:3354-3366

Sabatini S, Heidstra R, Wildwater M, Scheres B (2003) SCARECROW is involved in positioning the stem cell niche in the Arabidopsis root meristem. Genes Dev 17:354-358

Sanguineti MC, Li S, Maccaferri M, Corneti S, Rotondo F, Chiari T, Tuberosa R (2007) Genetic dissection of seminal root architecture in elite durum wheat germplasm. Annals of Applied Biology 151:291-305

Schmieder R, Edwards R (2011a) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6:e17288

Schmieder R, Edwards R (2011b) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863-864

Schmieder R, Lim YW, Edwards R (2012) Identification and removal of ribosomal RNA sequences from metatranscriptomes. Bioinformatics 28:433-435

Schommer C, Debernardi JM, Bresso EG, Rodriguez RE, Palatnik JF (2014) Repression of cell proliferation by miR319-regulated TCP4. Mol Plant 7:1533-1544

Somssich M, Khan GA, Persson S (2016) Cell Wall Heterogeneity in Root Development of Arabidopsis. Front Plant Sci 7:1242

van den Berg C, Weisbeek P, Scheres B (1998) Cell fate and cell differentiation status in the Arabidopsis root. Planta 205:483-491

van den Berg C, Willemsen V, Hage W, Weisbeek P, Scheres B (1995) Cell fate in the Arabidopsis root meristem determined by directional signalling. Nature 378:62-65

van den Berg C, Willemsen V, Hendriks G, Weisbeek P, Scheres B (1997) Short-range control of cell differentiation in the Arabidopsis root meristem. Nature 390:287-289

Wang HZ, Dixon RA (2012) On-off switches for secondary cell wall biosynthesis. Mol Plant 5:297-303

Wang J, Luo MC, Chen Z, You FM, Wei Y, Zheng Y, Dvorak J (2013) *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. New Phytol 198:925-937

Weeks Donald P, Spalding Martin H, Yang B (2015) Use of designer nucleases for targeted gene and genome editing in plants. Plant Biotechnology Journal 14:483-495

Welch D, Hassan H, Blilou I, Immink R, Heidstra R, Scheres B (2007) Arabidopsis JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. Genes Dev 21:2196-2204

Werner T, Motyka V, Laucou V, Smets R, Van Onckelen H, Schmulling T (2003) Cytokinin-deficient transgenic Arabidopsis plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. Plant Cell 15:2532-2550

Woodward AW, Bartel B (2005) Auxin: regulation, action, and interaction. Ann Bot 95:707-735

Xue T, Liu Z, Dai X, Xiang F (2017) Primary root growth in Arabidopsis thaliana is inhibited by the miR159 mediated repression of MYB33, MYB65 and MYB101. Plant Sci 262:182-189

Yang L, Zhang J, He J, Qin Y, Hua D, Duan Y, Chen Z, Gong Z (2014) ABA-mediated ROS in mitochondria regulate root meristem activity by controlling PLETHORA expression in Arabidopsis. PLoS Genet 10:e1004791

Yang S, Yu H, Xu Y, Goh CJ (2003) Investigation of cytokinin-deficient phenotypes in Arabidopsis by ectopic expression of orchid DSCKX1. FEBS Lett 555:291-296

Zhong R, Lee C, Ye ZH (2010) Global analysis of direct targets of secondary wall NAC master switches in Arabidopsis. Mol Plant 3:1087-1103

Zhong R, Taylor JJ, Ye ZH (1999) Transformation of the collateral vascular bundles into amphivasal vascular bundles in an Arabidopsis mutant. Plant Physiol 120:53-64

TABLES

Table 3.1 Details of the sequenced reads for the Long Root and VSR datasets.

Sample	Raw reads		Cleaned high quality reads	
	Single-end	Paired-end	Single-end (%)	Paired-end (%)
LR Replicate 1	25,352,326	6,657,382	22,638,023 (89.29)	5,621,234 (84.44)
LR Replicate 2	27,085,540	7,042,388	26,335,159 (97.23)	6,526,716 (92.68)
LR Replicate 3	27,114,351	6,928,016	25,904,592 (95.54)	6,324,452 (91.29)
VSR replicate 1	24,387,317	6,381,164	22,452,573 (92.07)	5,556,556 (87.08)
VSR replicate 2	23,032,193	6,039,935	21,323,360 (92.58)	5,274,329 (87.32)
VSR replicate 3	24,098,214	6,153,778	22,047,231 (91.49)	5,320,806 (86.46)
Total	151,069,941	39,202,663	140,700,938 (93.14)	34,624,093 (88.32)

Table 3.2 Details of the Functional annotation of the DEGs

	# gene models	
Refseq 1.0 High confidence (HC) models	110,790	
Refseq 1.0 Low confidence (LC) models	269,583	
Total gene models	380,373	
	Up	**Down**
DEGs	3,635	777
• **HC**	3,418	567
• **LC**	145	37
• **Stringtie Transcripts**	72	173

FIGURE LEGENDS

Figure 3.1 VSR phenotypes in seedlings and adult plants. (a) CS BC2F1 seedlings 20 days after planting carrying genotypes *Vsr1aVsr1a* (left) and *Vsr1aVsr1b* (right). (b) A 10 days-old Bobwhite (BW)-BC5F1 seedling (Vsr1aVsr1b) showing very short primary roots, very short crown roots and very short lateral roots. (c) A root tip of a 10 days-old BW seedling (*Vsr1aVsr1a*). No lateral roots were observed. (d) BW BC5F1 plants carrying genotypes *Vsr1aVsr1a* (left) and *Vsr1aVsr1b* (right). The scale bars indicate 1 cm.

Figure 3.2 Overview of metabolic processes to which differentially expressed genes in VSR root tips belong. Each box represents a gene. The up-regulated genes were represented in blue color and down-regulated genes in red color.

Figure 3.3 Overview of various cellular responses to which differentially expressed genes in VSR root tips belong. Each box represents a gene. The up-regulated genes were represented in blue color and down-regulated genes in red color.

Figure 3.4 Overview secondary metabolism pathways to which differentially expressed genes in VSR root tips belong. Each box represents a gene. The up-regulated genes were represented in blue color and down-regulated genes in red color.

Figure 3.5 Transcription factors differentially expressed in the VSR root tips. The X-axis represents the transcription factor families, and the Y-axis represents the number of

transcripts belonging to each family. The blue colored bars are up-regulated genes, and the orange color bars are down-regulated genes in the VSR root tips.

Figure 3.6. DAB staining of root tips. (a) Primary roots of CS-*Vsr1a*, (b) primary root of CS-*Vsr1b*, (c) primary root of F_1 hybrid between CS-*Vsr1a* and CS-*Vsr1b*, and (d) lateral roots of the F_1 hybrid.

Figure 3.7. Lignin deposition in VSR. Root tips of CS (a), TA (b) and their F_1 hybrids of 1 d (c), 2 d (d) and 3 d after germination (e and f). Lignin was stained in red (arrows) by 1% phloroglucinol (d, e and f). The scale bars in a through e indicates 1 mm; the scale bar indicates 100 μm.

Figure 3.8 miRNAs upregulated by at least 10-fold in VSR root tips. The X-axis represents the miRNA and the Y-axis represents the fold change.

Figure 3.9. A working model for Vsr1-mediated root inhibition. The arrows indicate promotion and the inverted 'T's indicate suppression. VIPs: Vsr1-interacting proteins.

FIGURES

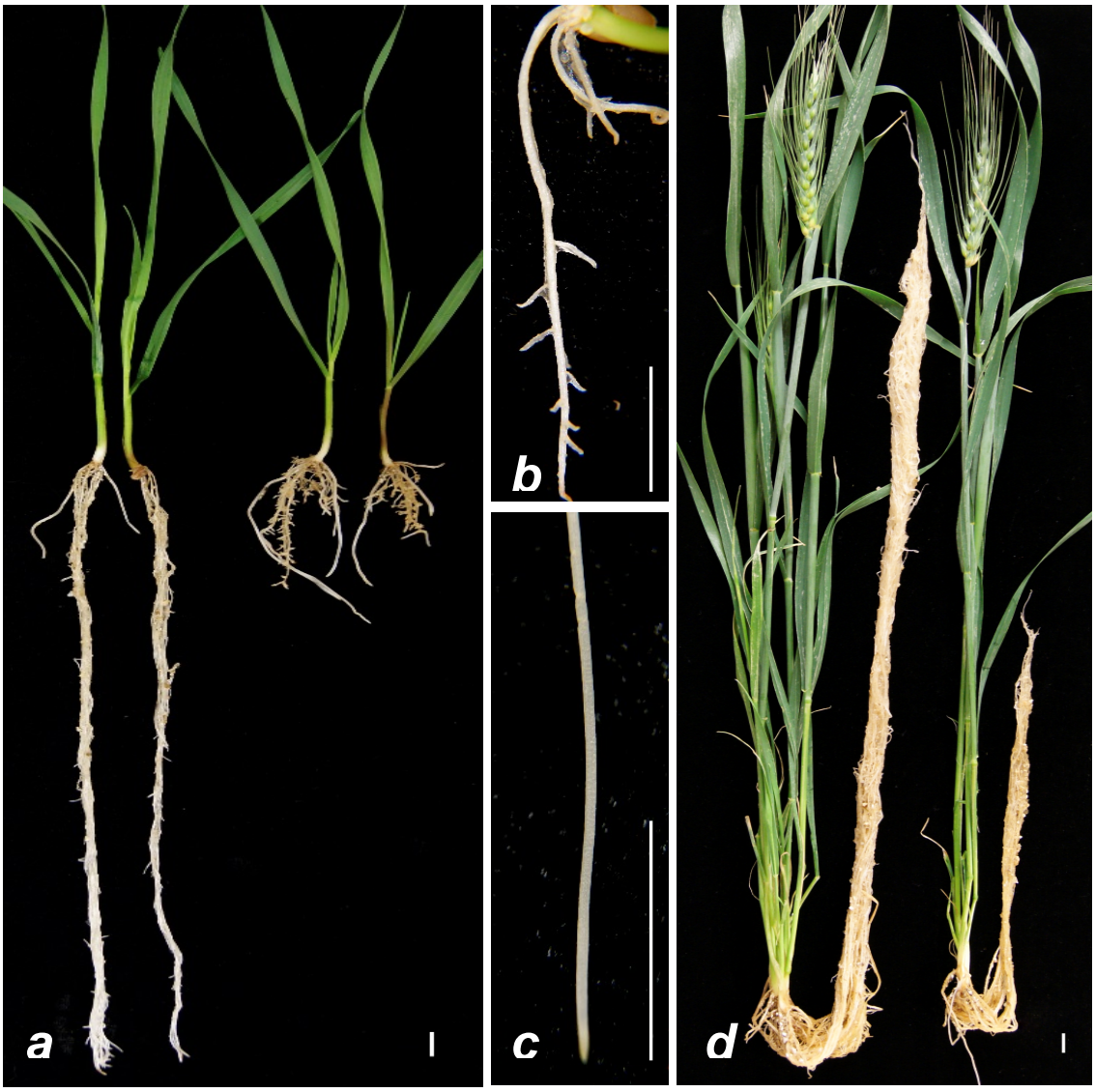


Figure 3.1 VSR phenotypes in seedlings and adult plants. (a) CS BC2F1 seedlings 20 days after planting carrying genotypes *Vsr1aVsr1a* (left) and *Vsr1aVsr1b* (right). (b) A 10 days-old Bobwhite (BW)-BC5F1 seedling (Vsr1aVsr1b) showing very short primary roots, very short crown roots and very short lateral roots. (c) A root tip of a 10 days-old BW seedling (*Vsr1aVsr1a*). No lateral roots were observed. (d) BW BC5F1 plants carrying genotypes *Vsr1aVsr1a* (left) and *Vsr1aVsr1b* (right). The scale bars indicate 1 cm.

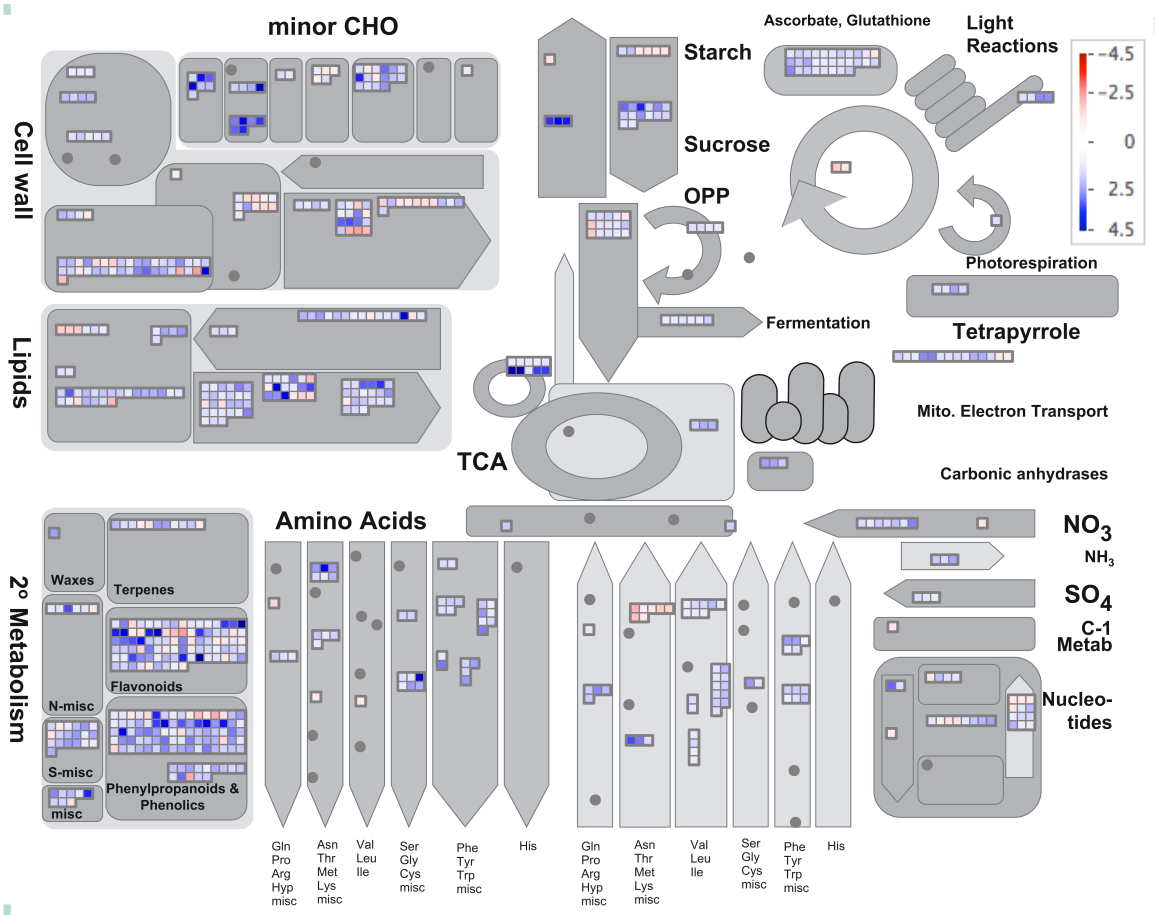


Figure 3.2 Overview of metabolic processes to which differentially expressed genes in VSR root tips belong. Each box represents a gene. The up-regulated genes were represented in blue color and down-regulated genes in red color.

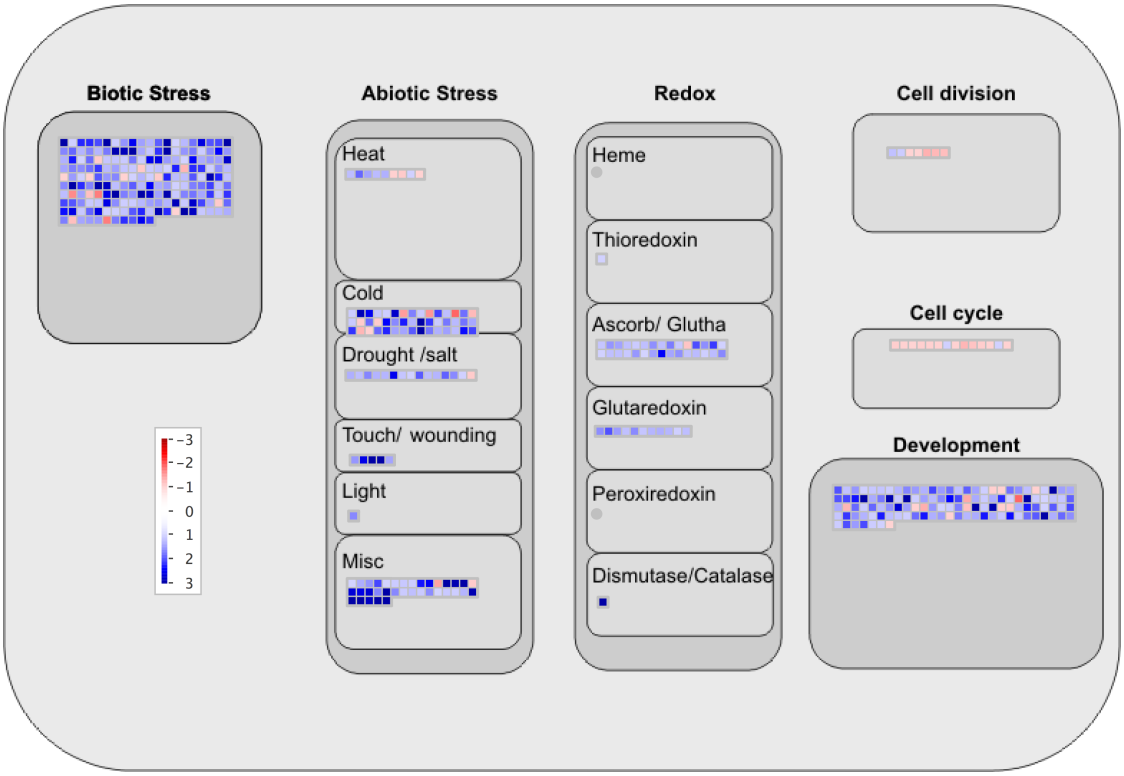


Figure 3.3 Overview of various cellular responses to which differentially expressed genes in VSR root tips belong. Each box represents a gene. The up-regulated genes were represented in blue color and down-regulated genes in red color.

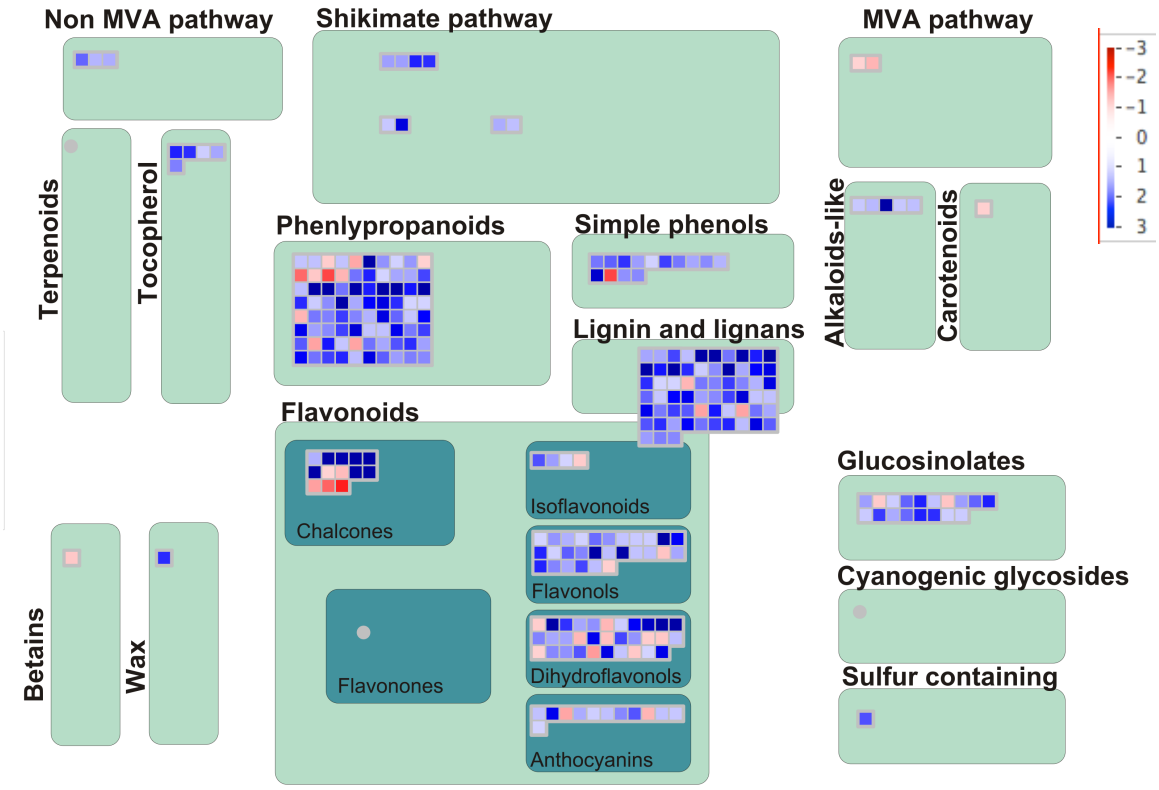


Figure 3.4 Overview secondary metabolism pathways to which differentially expressed genes in VSR root tips belong. Each box represents a gene. The up-regulated genes were represented in blue color and down-regulated genes in red color.

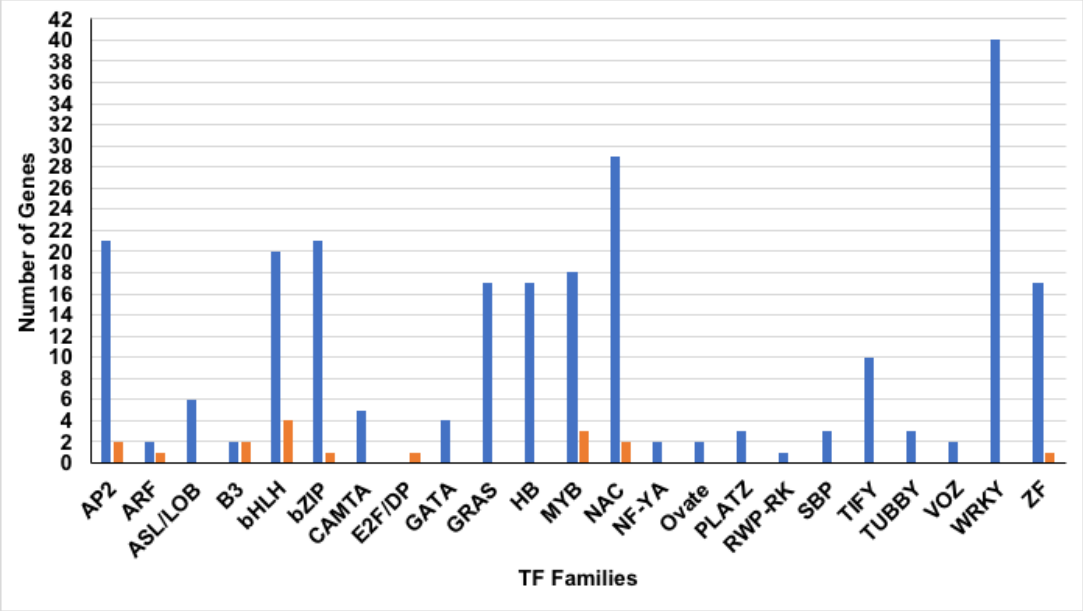


Figure 3.5 Transcription factors differentially expressed in the VSR root tips. The X-axis represents the transcription factor families, and the Y-axis represents the number of transcripts belonging to each family. The blue colored bars are up-regulated genes, and the orange color bars are down-regulated genes in the VSR root tips.

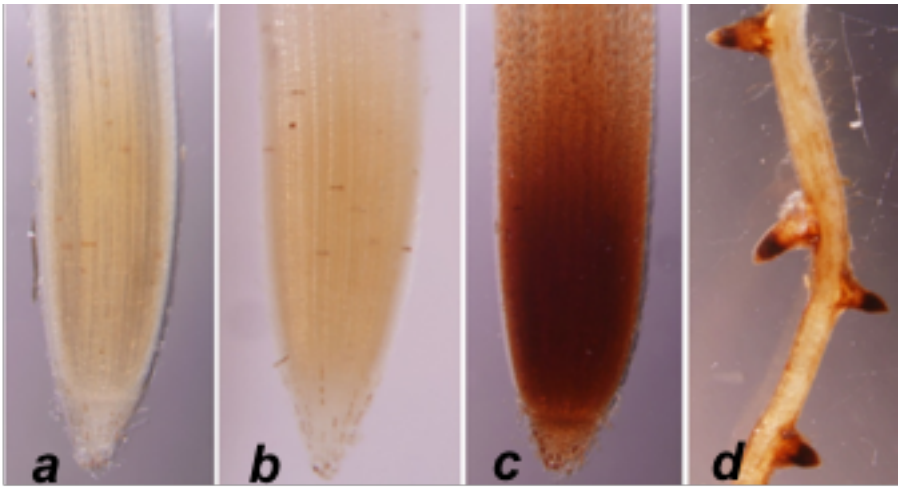


Figure 3.6. DAB staining of root tips. (a) Primary roots of CS-*Vsr1a*, (b) primary root of CS-*Vsr1b*, (c) primary root of F_1 hybrid between CS-*Vsr1a* and CS-*Vsr1b*, and (d) lateral roots of the F_1 hybrid.

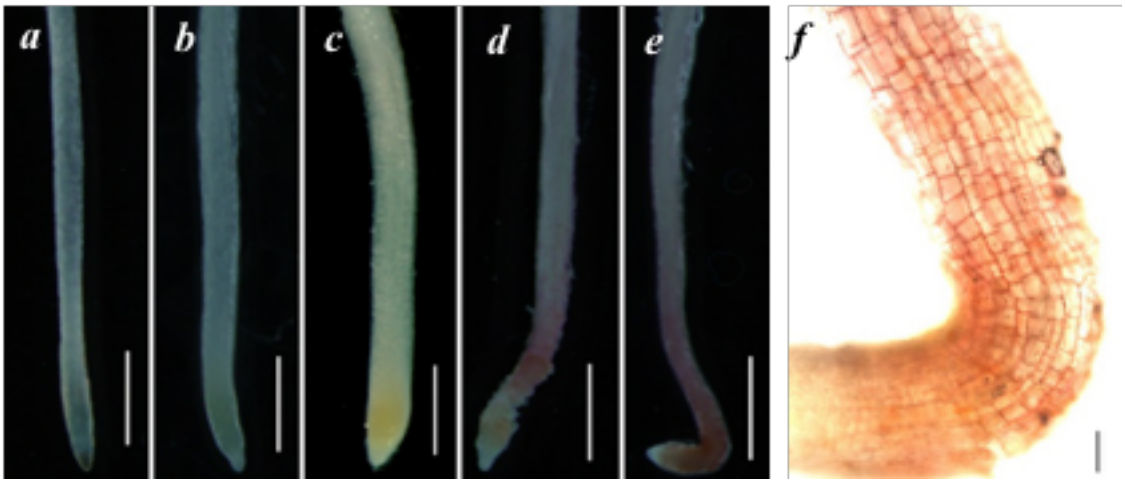


Figure 3.7. Lignin deposition in VSR. Root tips of CS (a), TA (b) and their F_1 hybrids of 1 d (c), 2 d (d) and 3 d after germination (e and f). Lignin was stained in red (arrows) by 1% phloroglucinol (d, e and f). The scale bars in a through e indicates 1 mm; the scale bar indicates 100 µm.

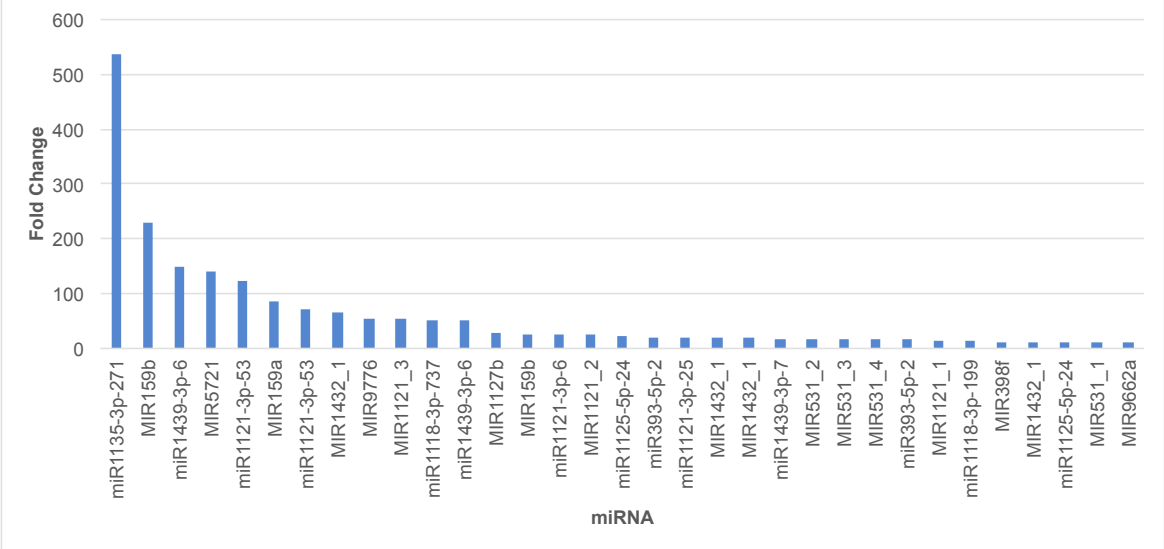


Figure 3.8 miRNAs upregulated by at least 10-fold in VSR root tips. The X-axis represents the miRNA and the Y-axis represents the fold change.

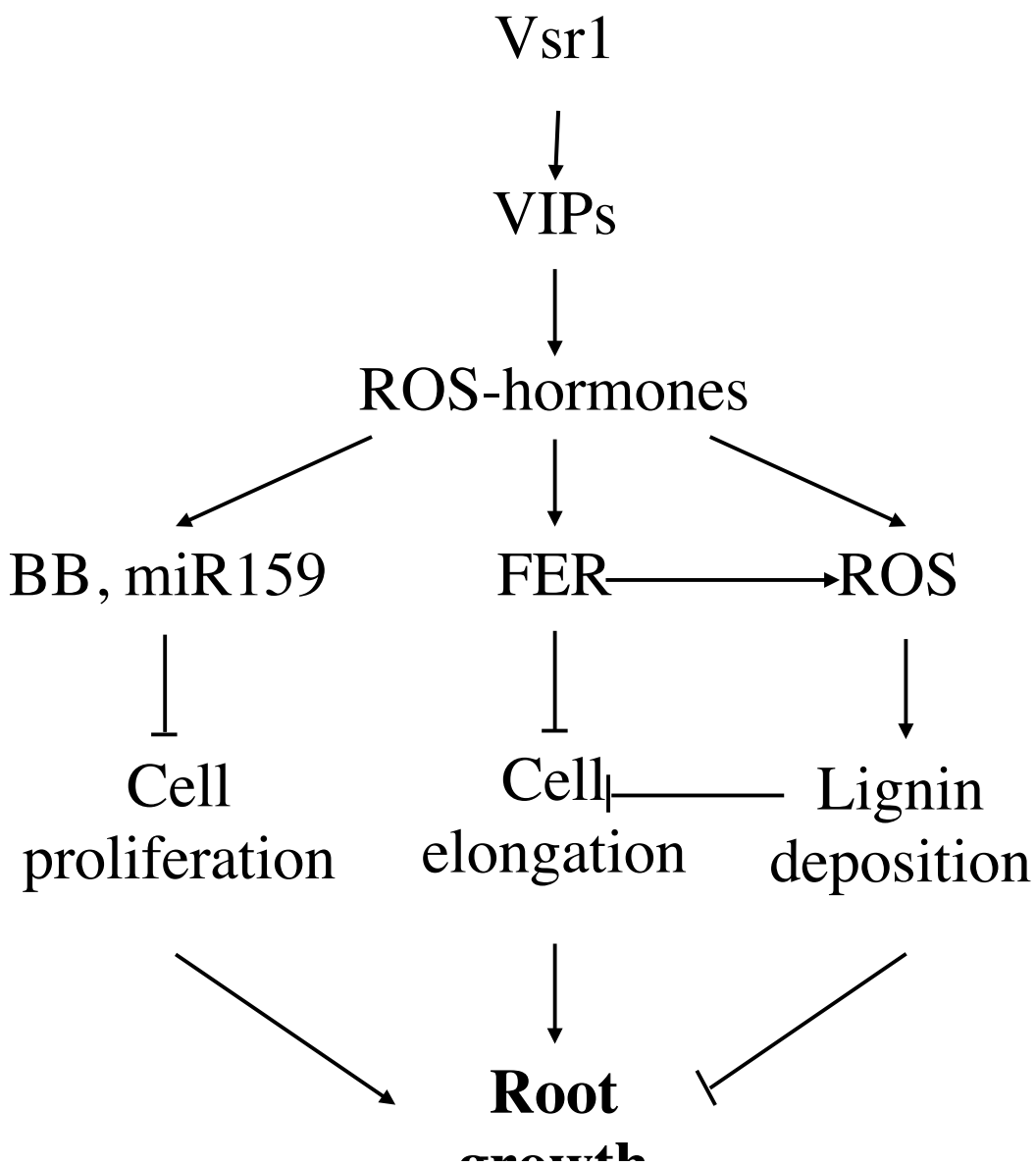


Figure 3.9. A working model for Vsr1-mediated root inhibition. The arrows indicate promotion and the inverted 'T's indicate suppression. VIPs: Vsr1-interacting proteins.

SUPPLEMENTARY TABLES

Table S3.1. Annotation and the expression profiles of the DEGs

Table S3.2 Annotation and the expression profiles of the differentially expressed miRNA

Both the tables were uploaded to proquest as supplementary material.

CHAPTER 4

Conclusions and Future Directions

CONCLUSIONS AND FUTURE DIRECTIONS

This study was initiated due to the lack of information on root, development of cereals, more specifically of wheat. It was aimed at throwing light on the anatomy of the root tip transcriptome and helping further studies to elucidate the molecular mechanism involved in the manifestation of the VSR phenotype. The transcriptome of the wheat root was generated using the high-throughput sequence data and *de novo* assembly strategy. *De novo* assembly of the root transcriptome was compared to the draft genome sequences that were made available by IWGSC and it not only showed that the *de novo* transcriptome is of reasonable good quality, but also proved that the gene models predicted in the draft genome were incomplete. In addition, we were able to identify several transcripts that were not coding for a functional protein and that could be putative non-coding transcripts and were differentially expressed in the root tip. Our study identified differentially expressed transcripts in the root tips compared to the mature root and also several homologs known to participate in root development in the model systems were also identified in the wheat root transcriptome. We also identified several proteins that were previously not known to regulate root development or that were not previously assigned any function were identified. We strongly believe *de novo* assembled root transcriptome paved a path for improving the genome annotation of wheat where it can identify novel transcripts that otherwise not predicted by the *ab initio* gene prediction tools and provides a new evidence for the low confidence gene models predicted by the genome annotation pipelines. In addition, it also provides the transcripts that were not present in the genome assembly and thus providing information of the chromosome fragments that were missed in the genome

sequence. For further understanding the molecular mechanism of the cereal root development and improving the wheat genome annotation, I suggest following research be considered:

a. Identifying the homologs of the putative non-coding transcripts expressed in the root, in other sequenced plant genomes
b. Heterologous expression of the differentially expressed non-coding transcripts in Arabidopsis to identify their role in root development
c. Several differentially expressed transcripts in root tip has no function assigned to them or coding for proteins of unknown function. Functional characterization of these genes will provide more information to resolve the complexity of the pathways involved in the root development and also can identify the missing links in the present knowledge
d. Further computational studies to identify the transcription start sites can be initiated with the current data to improve the genome annotation

The root transcriptome assembly initiated the study of root biology and is an initial milestone for understanding the molecular mechanisms involved in the cereal root development. This paved a path to understand the very interesting root phenotype identified in our lab, i.e., VSR. A transcriptome topology of the VSR root tips was elucidated using the recently available reference genome for wheat and high throughput sequencing technology. This study identified the interplay between hormones, ROS and developmental pathways in the occurrence of the VSR phenotype. VSR represents an important and interesting mechanism of gene expression regulation. Although we can imagine but still

have no idea how the *Vsr1* expression is regulated in the root tip and how *Vsr1* can mis-activate those genetic pathways that suppress cell proliferation and elongation. The interesting entanglement of several hormones with ROS and other stresses makes the delineation of *Vsr1* mediated pathway an important next step. To achieve this, I suggest following research to be considered.

a. We will need to clone the *Vsr1* gene.
b. Despite the availability of the wheat reference genome from CS, we still need the sequence from the corresponding region of the TA genome. To this end, the 5D chromosome may be sorted from TA and sequenced using the long-read technology, such as the Nanopore sequencing and aligned with the CS reference genome to identify the variation.
c. At the same time, the root tip transcriptome from the VSR also needs to be sequenced using the long-read technologies to a much deeper level or using sequence capture strategy to increase the sequence depth as the expression of Vsr1 may be very low or coding for non-coding RNA.
d. The *Vsr1* candidate gene identified in this way can be validated by traditional genetic complementation or genome editing such CRSIPR/Cas9, which can cause DNA deletions from 1-bp to several hundred kb.