Comparative Analysis of Gene Expression in the Developing Roots and Root Hairs Across Vascular Plants

by

Ling Huang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular and Developmental Biology)
in the University of Michigan
2016

Doctoral Committee:

Professor John W. Schiefelbein, Chair
Associate Professor Anuj Kumar
Associate Professor Erik E. Nielsen
Professor Patricia J. Wittkopp

ACKNOWLEDGEMENTS

I have experienced tremendous ups and downs during my PhD study and finally I am able to be here and present my work, which is impossible without the help from my family, friends, and colleagues. I would like to express my sincere gratitude to them.

My advisor, Dr. John Schiefelbein, has always been a great teacher and supportive mentor for my work. He started this interesting and novel research project and guided me through the whole process of the project, from designing experiments to writing manuscripts. John also provided me with unimaginable flexibility in my research. He encouraged me to pursue a master degree in bioinformatics so that I could have a comprehensive understanding of the computational methods needed for my projects. He also granted me an internship opportunity to gain the practical research experience. I learned to think critically and became self-motivated through the training under his supervision. I am glad to have had worked in the Schiefelbein lab.

I would also like to thank my committee members, Dr. Erik Nielsen, Dr. Patricia Wittkopp, and Dr. Anuj Kumar for their intellectual input throughout the project. My project is a combination of different aspects of biology and I would like to thank Erik for his expertise in plant cell biology, Trisha for her valuable suggestions in evolutionary biology, and Anuj for his extensive knowledge in genomics, proteomics, and bioinformatics. I could not have accomplished my projects without their advice.

I would like to thank Dr. Haoxing Xu and Dr. Robert Denver for providing me a rotation position in their lab. During my study in the Xu lab, I learned to keep passion for science and developed my logical thinking skills. During my study in the Denver lab, I learned to pay attention to details and developed my statistical skills. Though I did not end up doing my PhD project in their lab, my rotation experience is an unforgettable gift to me.

I appreciate all the helps I received during my PhD study. The former and current members Dr. Xinhui Shi, Dr. Xiaohua Zheng, Dr. Kook Hui Ryu, and Wenjia wang in the Schiefelbein lab provided enormous amount of suggestions and created a fantastic research environment. Dr. Richard Mc Eachin, as our collaborator in bioinformatics, helped to set up the initial RNA-Seq analysis pipeline and encouraged me to learn more about programming. Dr. Stephen Smith and Dr. Ya Yang in the Smith lab generously shared their phylogenetic analysis pipeline with me. Gregg Sobocinski trained me into a professional in microscopy imaging. I feel fortunate to have worked with them.

I am grateful to the company of my classmates and friends, who have filled the past five years with smiles and laughter. Although I am not able to name all of them, I would remember all the time we have spent together.

Most of all, I would like to give my special thanks to my family. I feel really sorry for the inability to come back and reunion with the family during traditional holidays. But my mom and dad have always been considerate and supportive of me and let me pursue what I am interested in. My two cats, Zura and Naicha, have become an indispensable source of comfort and relief in my spare time. My husband is always right by my side to share my delight and to cheer me on. I dedicate this thesis to my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

The evolution of the plant root is one of the remarkable examples of land plants' adaptation to the terrestrial environment. The root and root hairs are found in almost all vascular plants, making them good models to study the conservation and diversification of adaptations in different vascular lineages. The molecular basis of root and root hair development has been intensively studied in the model organism *Arabidopsis thaliana* and many of the key regulators have been elucidated over the past 20 years. However, little is known about the molecular basis of root and root hair development in other vascular plants, which prevents a detailed understanding of root and root hair adaptation at a mechanistic level. In this thesis project, I compared the gene expression programs in seven vascular plants to find the conserved/diversified regulators in root and root hair development. First, I defined temporal transcript accumulation profiles generated from three root development zones in seven vascular plants and found significant family-wise gene expression similarity between these plants. Next, I found that most of the 133 Arabidopsis key root development genes are used in all vascular plants, including *Selaginella moellendorffii,* which is thought to have evolved roots independently. These results suggest broad conservation in the molecular mechanisms employed during root formation. Next, I defined the root hair transcriptome in Arabidopsis and rice and found significant conservation in the two species. However, a subset of 563 genes regulated by ROOT HAIR DEFECTIVE 6 (RHD6) exhibited less conservation compared to the total root-

hair-expressed genes. This divergence in the family structure and gene expression is found in other vascular plants as well. I further analyzed the Arabidopsis root hair patterning genes in other vascular plants and identified a substantial difference in their expression patterns. These results suggest lineage specific diversification in the structure and expression of the root hair development genes in Arabidopsis. Altogether, this thesis work revealed broad conservation of gene expression programs for root formation in all vascular plants and lineage specific diversification for root-hair cell differentiation in Arabidopsis.

CHAPTER 1

Introduction: Land Plant Root Evolution and Root Hair Patterning

Roots and root hairs are typically underground structures of a plant that help with anchorage, nutrient uptake and storage, and interactions with the rhizosphere microbiome (Petricka et al., 2012; Datta et al., 2011). In addition to these important physiological functions, the root has many features that make it an excellent model to study plant organ formation. First, the root has been found in almost all land plant lineages (except bryophytes), enabling a large-scale comparative analysis across the plant kingdom to understand the evolution of root development. Second, the root has a simple structure despite that it is a multi-cellular organ with different tissue types. The simple development of the root makes it relatively easy to study how cells adopt their fates and further differentiate into different tissues. Third, detailed molecular genetic analyses have been conducted in the model organism *Arabidopsis thaliana*, elucidating the mechanisms of the root and the root hair developmental programs in different developmental stages. This serves as a great resource for comparative studies of other plants' root development. Fourth, the recent advancements in whole-genome sequencing provides important foundation to study the molecular basis of processes in plants other than the model organism Arabidopsis. The combination of the above factors provides a unique opportunity to use the root and the root hairs as models to study organ formation in other land plants.

In this chapter, I will summarize the current knowledge about the root and the root hair

development in Arabidopsis and the recent progress on the study of the molecular mechanisms to understand the root and the root hair development in other plants.

Land Plant Phylogeny

The emergence of land plants (embryophytes) was an important event in the history of evolution, leading to a great increase in species diversity on earth. The land plant kingdom is estimated to contain about 313687 species, 4 times larger than the estimated number of species in Chordata (Hinchliff et al., 2015). It is widely accepted that all land plants share a single common origin from charophycean green algae, as supported by both morphology and molecular data (Bremer et al., 1987; Mishler et al., 1994; Kranz et al., 1995; Duff and Nickrent, 1999). (Figure 1.1)

The land plants that are present on Earth today are divided into two major groups: the basal land plants and the vascular plants. The basal land plants include three bryophyte lineages that probably diverged about 480 million years ago (Kenrick and Crane, 1997): liverworts, hornworts, and mosses. The phylogenetic relationship between these three is still not resolved. The absence of introns in two mitochondrial genes NAD1 and COX2 as well as the inability to convert auxin into amide and ester conjugates indicated that liverworts are less similar to the vascular plants and so they were considered the most basal group (Qiu et al., 1998; Sztein et al., 1995). However, other phylogenetic studies found these three lineages to be monophyletic after correction for a bias in synonymous substitutions (Cox et al., 2014; Karol et al., 2010). The Open Tree of Life project, which combines large-scale genomic data, posits that liverworts and mosses are the most basal monophyletic group and hornworts are more closely related to vascular plants

(Hinchliff et al., 2015) (Open Tree of Life Project: http://www.opentreeoflife.org/), although another recent analysis based on 360 plastid genome sequences and three different complementary genomic matrices suggested that liverworts are the earliest group followed by mosses and hornworts (Ruhfel et al., 2014; Qiu et al., 2006).

The vascular plants arose about 400 million years ago with the distinct features of vascular tissues for water and nutrient transport, a dominant sporophyte phase in its whole life cycle, and branched sporophytes (Pryer et al., 2004b). Early vascular plants, which proliferate by spores, can be classified into two monophyletic lineages: lycophytes (club mosses) and monilophytes (ferns) (Pryer et al., 2001). Molecular evidence suggests lycophytes to be a more basal lineage compared to monilophytes by two facts: an inversion in the chloroplast genome that they share with bryophytes and a large-scale phylogenetic analysis based on two super matrices composed of 4 genomic sequences and 136 morphological characters (Raubeson and Jansen, 1992; Pryer et al., 2001). Vascular plants that bear seeds are called spermatophytes, and they include gymnosperms (seeds without carpels; represented by four living lineages: cycadophyta, gnetophyta, ginkgophytes, and coniferophyta) and angiosperms (seeds enclosed within carpels) (Pryer et al., 2004a). Monilophytes and spermatophytes together are called euphyllophytes.

Angiosperms include by far the largest number of species across the plant kingdom (Pryer et al., 2004a). Since 1998, a large group of plant biologists (Angiosperm Phylogeny Group; APG) have collaborated on the classification and organization of orders and families within angiosperms using mostly molecular phylogenetic analyses (THE ANGIOSPERM PHYLOGENY GROUP, 1998, 2003, 2009, 2016). Their initial paper covered 462 families and 40 orders, which has been recently revised to 416 families and 64 orders (THE ANGIOSPERM PHYLOGENY GROUP, 1998, 2016). The APG tree provides a high-resolution framework of

the angiosperm phylogeny. Although some family positions on the tree remain to be resolved, it is widely agreed that there are eight major lineages within angiosperms: amborellales, nymphaeales, austrobaileyales, magnoliids, chloranthales, monocots, ceratophyllales, and eudicots (THE ANGIOSPERM PHYLOGENY GROUP, 2016).

With the advancement of computational capacity and reduced sequencing cost, there is an increasing amount of phylogenomic analyses published using genome-scale molecular sequences to resolve the fine resolution phylogeny of land plants. However, it is worth pointing out that the sequence data itself might not be able to provide the ultimate answer to land plant evolution. The extinction of many plant lineages makes it almost impossible to acquire sequence data from every sub-clade of species to fully reconstruct the tree of life. Recent studies that combine molecular mechanisms of plant development with phylogenetic evidence to explore the evolutionary history of critical morphological features might be a promising way to better understand the tree of life and find clues to the missing pieces of land plant evolution (Harrison et al., 2005).

The Evolution of Roots

The root is one of the major innovative adaptations that plants made during their transition from water to land. Early land plants like bryophytes do not have well-developed root systems. Instead, they have single (liverworts and hornworts) or multi-cellular (mosses) tubular structures of a uniform tissue type called rhizoids to assist with anchorage (Jones and Dolan, 2012). The evolutionary emergence of roots allowed for efficient transportation of water and nutrients through the whole plant (Pryer et al., 2004b), which is an indispensable advantage to the

prosperity of land plants.

A true root is a multi-cellular organ with layers of different tissue types including stele, cortex, and epidermis, arranged in a radial organization produced by the apical root meristem (Raven and Edwards, 2001) (Figure 1.2). Though it shares these tissue types with the shoot, the root also has a unique structure (the root cap) that is not found in any other plant organs. The root cap carries out many important biological functions. It protects the root apical meristem and lubricates the soil by the production of mucilage during root elongation (Iijima et al., 2003; McCully, 1999). It produces border cells that detach from the root cap and interact with the microbiota of the rhizosphere (Hawes et al., 2000, 2012). The root cap also plays an important role in gravitropism through the redistribution of auxin (Blancaflor et al., 1998; Abas et al., 2006; Massa and Gilroy, 2003). Furthermore, the root cap is the preferential location of the conversion of auxin precursor (indole-3-butyric acid) into active auxin to initiate lateral root development (Roppolo et al., 2011).

The rhizoids and the root hairs are suggested to be functionally homologous structures as both elongate from the plant in a unidirectional manner and function in anchorage and uptake of water and nutrients (though the latter role might be inferior in the rhizoids) (Jones and Dolan, 2012). In addition, there is molecular evidence that the root hairs and the rhizoids may share a similar developmental program. ROOT HAIR DEFECTIVE 6 (AtRHD6) and its paralog RHD6-LIKE 1 (AtRSL1) are critical bHLH transcription factors positively regulating root hair development in Arabidopsis (Masucci and Schiefelbein, 1994; Menand et al., 2007) and the mutants of their orthologs in moss *Physcomitrella patens* (PpRSL1 and PpRSL2) and liverwort *Marchantia polymorpha* (MpRSL1) have similar rhizoid-less phenotypes to the hairless *Atrhd6* and *Atrsl1* double mutant (Menand et al., 2007; Proust et al., 2016). Furthermore, the bryophyte

orthologs are able to rescue the Arabidopsis hairless phenotype when overexpressed in *Atrhd6* and *Atrsl1* double mutant background, indicating conserved molecular function of RHD6 in the root hair/rhizoid development pathway (Menand et al., 2007; Proust et al., 2016).

It is still not clear how the single-tissue-type rhizoids evolved to the multi-tissue-type roots. However, it is accepted that the roots may have evolved more than once in the history of plants evolution based on current limited evidence (Raven and Edwards, 2001; Jones and Dolan, 2012). First, roots have not been found in the fossils of the common ancestor of the lycophytes and the euphyllophytes (Raven and Edwards, 2001; Friedman et al., 2004). Second, lycophyte roots branch dichotomously whereas euphyllophyte roots branch by *de novo* root formation (lateral roots) from the pericycle tissues (Raven and Edwards, 2001; Banks, 2009). These observations suggest an independent evolutionary origin of the lycophyte roots and the euphyllophyte roots. However, due to the fact that it is quite difficult to preserve the underground structures of fossils, the strength of the fossil evidence might be overestimated and a single origin of root evolution is still possible. In order to make more confident conclusion about the root evolution in land plants, it is essential to understand the root cap development process in different plant species, especially in the lycophytes, to see if they use similar molecular mechanisms to euphyllophytes or they evolve *de novo* root cap forming mechanisms.

Root Development in Arabidopsis

Root development has been extensively studied in the model organism *Arabidopsis thaliana* to understand plant organ formation for several reasons. First, the Arabidopsis root is a relatively simple organ with just a few tissue types. Second, most cells are arranged in cell files that

emanate from their corresponding initials/stem cells (Benfey et al., 1993), making it possible to trace the developmental process of one specific cell. Third, the developmental stage of cells can be measured based on their physical distance to the apical meristem quiescent center (QC), leading to three morphologically different development zones as suggested by a recent review (Petricka et al., 2012) (Figure 1.2).

The meristematic zone is the most apical part of a root tip that contains the root cap and the apical meristem. The QC cells are responsible for the maintenance of the stem cell niche in the apical root meristem whose asymmetric cell divisions give rise to four cell type initials: vascular, cortex/endodermal, epidermal/lateral root cap, and columella (Sozzani and Iyer-Pascuzzi, 2014). The further differentiation of these initials (except columella initial) results in a radial organization of different tissue types in cell files along the longitudinal axis in Arabidopsis: vascular tissues (metaxylem, protoxylem, procambium, phloem, and pericycle), endodermis, cortex, epidermis and lateral root cap (from center to outmost layer; Figure 1.2C). The columella initial develops into the root-ward columella root cap, working in collaboration with the lateral root cap to protect the apical meristem during underground root elongation (Iijima et al., 2003; McCully, 1999).

The elongation zone is where the cells start rapid growing (elongation) but remain undifferentiated. The average cell length is 8 μm when exiting the meristematic zone and it is able to grow to over 100 μm at the end of elongation zone (Verbelen et al., 2006). Vacuole expansion has been found to be a feature associated with the rapid cell elongation (Verbelen et al., 2006). Vacuoles can only be found in cells that exit proliferation cycle and their size largely increases in the elongation zone when cell length expands (Verbelen et al., 2006). Other distinct features of the elongation zone include the transverse orientation of cellulose fibrils and cortical

microtubules, which are believed to be the determinants of the mechanical anisotropy in cell wall elongation (Kerstens and Verbelen, 2003; Le et al., 2004). The transverse orientation of cellulose fibrils and cortical microtubules switches to a random distribution in the root-hair cells or a longitudinal orientation in the non-hair cells in the differentiation zone where the cell elongation slows down (Kerstens and Verbelen, 2003; Le et al., 2004). However, the molecular mechanism of the re-orientation of cellulose fibrils and cortical microtubules remains to be uncovered.

The differentiation zone is the region in which cells become fully differentiated. It is in this zone that root hair initiation and extension can be observed in those epidermal cells that acquire root-hair cell fates. Another notable feature of the differentiation zone is the formation of the Casparian strip, which functions as an impermeable barrier between the endodermis and the vascular tissues (Roppolo et al., 2011). This barrier prevents passive diffusion to the central vascular tissues and therefore protects the plant from pathogen invasion and allow active transportation of selected nutrients (Naseer et al., 2012).

Root Hair Patterning of Vascular Plants

The root epidermis is the outermost layer of a root and typically contains two cell types: root-hair cells and non-hair cells. As such, the cell fate determination process of an epidermal cell is reduced to a “to be, or not to be” question. The arrangement of the root-hair cells and non-hair cells in the epidermis provides an easily observed morphological phenotype (root hair patterning) to study the underlying cell fate specification.

There have been extensive studies of root hair patterns in the vascular plants, which show that there are three basic pattern types: Type 1-random pattern, Type 2-alternating pattern, and

Type 3-position-dependent pattern (Datta et al., 2011; Dolan, 1996) (Fig. 1.3). Every epidermal cell of a Type 1 plant may develop into a root-hair cell, in other words, no regular pattern could be observed. The roots of Type 2 plants alternate between root-hair cells and non-hair cells along each epidermal cell file. The root-hair cell is the smaller daughter cell after an asymmetric cell division at the meristematic zone (Kim and Dolan, 2011; Cutter and Hung, 1972). The roots of Type 3 plants alternate between one entire root-hair cell file and one or more non-hair cell files in the epidermis (Cormack, 1947; Dolan et al., 1994; Galway et al., 1994). In this case, the cell fate of the epidermal cells has been reported to be dependent upon positional cues it receives from the environment (Berger et al., 1998b, 1998a; Kwak et al., 2005).

The distribution of species exhibiting three root hair patterns on the tree of life might provide some information about how these patterning mechanisms evolved. According to two recent studies that surveyed a wide range of angiosperms (Clowes, 2000; Pemberton, 2001) (Fig. 1.4), the Type 2 pattern has been found in nymphaeales, magnoliids, and monocots, all of which diverged prior to Eudicots. On the contrary, the Type 3 pattern has been found only in the Eudicots. No lineage specific preference has been found for the Type 1 pattern. Considering these findings, it may be that the Type 1 pattern is the most basal one and evolved into Type 2 and Type 3 patterns at later point(s) during land plants adaptation. Due to the fact that both Type 2 and Type 3 patterns have been found in multiple lineages in which they co-exist with Type 1 pattern, it is equally possible that these lineages lost or gained the patterning ability independently. Therefore, the analysis of the molecular basis of cell fate specification might provide valuable information as how the root hair patterns evolved in different lineages. For example, if the molecular mechanism that regulates the root hair pattern formation is found to be conserved in all Type 3 plants regardless of their position on the species tree, it is more likely for

this trait to be lost in different lineages compared to a convergent/parallel evolution hypothesis.

Molecular Basis of the Position-Dependent Root Hair Patterning in Arabidopsis

Arabidopsis has a Type 3-position-dependent root hair pattern. Epidermal cells localized over the anticlinal cortical cell wall (ACCW; H position) are most likely to differentiate as root-hair cells whereas epidermal cells localized over a single cortical cell (N position) are most likely to differentiate as non-hair cells (Dolan et al., 1994). In Arabidopsis, the morphological differences between the root-hair cells and the non-hair cells can be observed long before the root hair emerges. The immature root-hair cells are found to have greater cytoplasmic density (Dolan et al., 1994; Galway et al., 1994), smaller size and higher rate of cell division (Berger et al., 1998b; Dolan et al., 1994; Masucci et al., 1996), delayed vacuolation (Galway et al., 1994), and distinct chromatin organization (Costa and Shaw, 2006; Xu et al., 2005). Taken together, it is evident that the root epidermal cell fate is determined at an early stage during post-embryonic development based on the positional cues it receives (Berger et al., 1998a, 1998b; Kwak et al., 2005). Furthermore, the positional cues appear to continue to reinforce the cell fate of an epidermal cell until the cell leaves the meristematic zone because an epidermal cell is able to switch its cell fate when forced to occupy a different position over the underlying cortical cells (Berger et al., 1998a).

A dynamic regulatory network of the epidermal cell fate specification has been established over the past 20 years to understand the molecular basis of root hair pattern formation in Arabidopsis (Schiefelbein et al., 2009; Ishida et al., 2008; Schiefelbein et al., 2014; Grebe, 2012) (Fig. 1.5).

The central regulatory core complex consists four transcription factors: the R2R3 Myb protein WEREWOLF (WER) (Lee and Schiefelbein, 1999), two basic helix-loop-helix (bHLH) proteins GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) (Payne et al., 2000; Bernhardt et al., 2003, 2005; Article et al., 2013), and the WD40-repeat protein TRANSPARENT TESTA GLABRA (TTG1) (Galway et al., 1994). Mutants of each of these four transcription factors produce ectopic root hairs in the N positions, suggesting their function in promoting the non-hair cell fate. Studies have shown that the regulatory role of the core complex on the non-hair cells is mediated through an HD-ZIP gene *GLABRA2* (*GL2*) that triggers the non-hair cell fate mostly at the N position (Masucci et al., 1996; Lee and Schiefelbein, 1999). GL2 is preferentially expressed in the non-hair cells to directly repress the transcription activity of five bHLH root hair genes, including the *RHD6* (Masucci et al., 1996; Menand et al., 2007; Lin et al., 2015). Loss of GL2 expression promotes the non-hair cell fate in all positions despite the morphological differences between the root-hair cells and the non-hair cells (Masucci et al., 1996). RHD6 is necessary for the root-hair cell development pathway through an auxin and ethylene-mediated process and the root hair density is reduced to 20% in its absence (Masucci and Schiefelbein, 1994). The lack of accumulation of the WER-bHLH-TTG1 core complex leads to the expression of RHD6 in the epidermal cells at the H positions, which will finally differentiate as root-hair cells. (Cvrčková et al., 2010; Bruex et al., 2012).

Multiple feedback loops at different levels have been found to regulate this position-dependent cell specification process to ensure a robust response and rapid adaptation to the environment cues.

An R3 Myb transcription factor CAPRICE (CPC) is involved in a lateral inhibition mechanism to compete against WER for the binding site of the bHLH-TTG1 complex (Wada et

al., 1997; Song et al., 2011). The mutant of *cpc* exhibits a significant reduction in the root-hair cell formation, implying that CPC inhibits the complex formation in the presumptive root hair cells (Wada et al., 1997). The transcription of CPC is activated by TTG and WER in the non-hair cells and CPC protein then moves to and stays in the root-hair cells through an EGL3-dependent manner to repress *GL2* expression (Wada et al., 2002; Lee and Schiefelbein, 2002; Kang et al., 2013). CPC homologs TRIPTYCHON (TRY) and ENHANCER OF TRY AND CPC1 (ETC1) have been identified to have a partially redundant function as CPC in the lateral inhibition process (Schellmann et al., 2002; Simon et al., 2007; Kirik et al., 2004). TRY is downstream of GL2 and functions in a relatively later stage compared to CPC and ETC1, suggesting its putatively diverged role in the negative feedback loop (Simon et al., 2007; Pesch and Hülskamp, 2011).

A homolog of WER, MYB23 has been found to promote the non-hair cell fate under the regulation of the core WER-bHLH-TTG1 complex (Kang et al., 2009). This positive feedback loop reinforces the epidermal cell fate specification process that leads to a robust Type 3 root hair patterning in Arabidopsis.

SCRAMBLED (SCM) is a receptor-like kinase that putatively transduces the environment cues to the cascading signaling in the unspecified epidermal cells as the position-dependent root hair patterning is destroyed in the *scm* mutant by inhibition of *WER* transcription (Kwak et al., 2005; Kwak and Schiefelbein, 2007). The preferential accumulation of SCM on the plasma membrane at the H position is dependent on the downstream regulatory network including the core WER-bHLH-TTG1 complex and CPC/TRY/ETC1, working as a positive feedback loop to stabilize the positional signaling pathway for epidermal cell fate specification (Kwak and Schiefelbein, 2008).

Arabidopsis Root Hair Genes in Other Angiosperms

Although the Arabidopsis root hair development pathway has been extensively studied, relatively little is known about the molecular mechanism controlling epidermal cell differentiation process in other land plants. An important first step to understand the root hair development pathway in other species is to determine whether the key Arabidopsis regulators are conserved or have diverged. Some recent work in this areas is summarized in this section.

In addition to the conserved RHD6 bHLH family (Group VIII bHLH, RSL class I transcription factor), the LRL family has recently been shown to play a critical role in the rhizoid/root hair elongation in land plants as well. LRL family members encode Group XI bHLH transcription factors that positively regulate the root hair elongation in Arabidopsis, as shown by the fact that the triple mutant of *Atlrl1, Atlrl2, and Atlrl3* has a great defect in the root hair elongation (Karas et al., 2009). Similarly, LRL promotes the rhizoid and caulonema extension in moss because the *Pplrl1* and *Pplrl2* double mutant is completely rhizoidless and does not develop caulonema (Tam et al., 2015). Despite the fact that LRL might be under different regulation in moss compared to Arabidopsis (Karas et al., 2009; Bruex et al., 2012; Tam et al., 2015), the role of LRL transcription factors regulating the root hair/rhizoid elongation is quite conserved in moss. An LRL ortholog in lotus (*Lotus japonicas ROOTHAIRLESS1*, *LjRHL1*) has been identified through genetic screens (Karas et al., 2005). The *Ljrhl1* mutant has a defect in the root hair formation under various growth conditions (Karas et al., 2009). LjLRL and AtLRL have equivalent biochemical properties as over-expression of a single *AtLRL* (either one of the three *AtLRLs*) is able to rescue the hairless phenotype of *Ljrhl1* (Karas et al., 2009). An LRL ortholog in rice (*Oryza sativa ROOTHAIRLESS1*, *OsRHL1*) has been characterized through an

ethyl methanesulfonate-induced mutagenesis screening (Ding et al., 2009). This mutant has a significant reduction in the root hair length with no other root/root hair phenotype observed (Ding et al., 2009).

Another conserved root-hair gene family identified in other plant species is EXPANSIN A7 (EXPA7). AtEXPA7 and its paralog AtEXPA18 have been found to be specifically expressed in the root-hair cells of Arabidopsis (Cho and Cosgrove, 2002). A conserved RHE motif has been characterized in the cis-regulatory region upstream of *AtEXPA7*, *AtEXPA18* and all other orthologs across angiosperms (Kim et al., 2006). The identical expression pattern explicitly in the root-hair cells of *AtEXPA7* promoter-driven GFP and rice native *OsEXPA30* promoter-driven GFP suggests equivalent regulatory function of the promoters (Kim et al., 2006). In addition, the biochemical function of EXPA7 is at least partially conserved in rice. When carrying a non-synonymous point mutation in *OsEXPA17* (a paralog of *OsEXPA30*), the rice seedling has a three-fold reduction in the root hair length that is similar to the phenotype of a knock-down *AtEXPA7* transgenic line in Arabidopsis (Yu et al., 2011; Lin et al., 2011). Over-expression of *AtEXPA7* in the *Osexpa17* mutant background is able to partially restore the short root-hair phenotype (Yu et al., 2011).

In contrast to these relatively conserved families regulating the root hair development (initiation and elongation), the root hair patterning genes that regulate the epidermal cell fate specification are more diverged in other species. While the biochemical function may still remain unchanged, these gene families may not participate in the root hair development regulatory network as in Arabidopsis.

One example is the TTG1 family member in maize (*Zea mays*), petunia (*Petunia hybrida*), and Medicago (*Medicago truncatula*), all of which are able to restore the *ttg* hairy (or

trichomeless) mutant phenotype in Arabidopsis (Carey et al., 2004; Walker et al., 1999; De Vetten et al., 1997; Pang et al., 2009). Another shared feature by all four TTG1 mutants is the regulatory role on anthocyanin biosynthesis pathway as all four mutants have reductions in anthocyanin pigmentation (Carey et al., 2004; Galway et al., 1994; De Vetten et al., 1997; Pang et al., 2009). However, none of these species has a reported mutant phenotype in the root hair development. *Zmttg1* (*Zmpac1*) mutant has the same root hair and trichome abundances as the wild-type siblings (Carey et al., 2004). Similarly, *Mtttg1* (*Mtwd40-1*) has no altered density of trichomes or root hairs (Pang et al., 2009). The study of petunia TTG ortholog focused on its role in the pigmentation biosynthesis and did not report any root or trichome related phenotype (De Vetten et al., 1997).

The maize *R* gene has been found to encode a bHLH transcription factor in the same family as *AtGL3/AtEGL3* with similar biochemical functions (Bernhardt et al., 2003). The expression of maize *R* can induce ectopic non-hair cells in the wild-type Arabidopsis and rescue the *ttg* hairy mutant phenotype (Galway et al., 1994). In addition, maize *R* is able to activate GL2 expression under the regulation of WER (Hung et al., 1998; Lee and Schiefelbein, 1999). However, no root-hair related phenotype has been reported in maize *r* mutant.

There are two members of CPC family in rice, *TRICHOMELESS 1* (*OsTCL1*) and *TRICHOMELESS 2* (*OsTCL2*). The over-expression of *OsTCL1* in the wild-type Arabidopsis leads to higher root-hair density in the epidermis and a reduction in *GL2* expression, both are consistent with the phenotype of *AtCPC* over-expression transgenic line (Zheng et al., 2016). However, no distinguishable morphological difference is observed in rice transgenic line that constitutively expresses *OsTCL1* (Zheng et al., 2016).

All the above evidence suggests that the biochemical function of the orthologs of the root

hair genes (both the patterning genes and the morphogenesis genes) are quite conserved as they are generally able to complement the mutant phenotypes in the root hair development of their corresponding family member in Arabidopsis. Interestingly, mutation of orthologs of the root hair differentiation genes in other plant species produces similar root hair/rhizoid phenotypes compared to their Arabidopsis relatives whereas mutation of orthologs of the root hair patterning genes in other species does not have a mutant phenotype related to the root hair development defect. The diverged biological roles of the root hair patterning genes could be explained by several hypotheses: 1) there have been changes in the regulation of the patterning genes so that they are no longer expressed in the epidermis and therefore lose the control of the epidermal cell fate specification; 2) there have been changes in the regulation or biochemical function of one or more of the other key regulators in the root hair development pathway so that they no longer respond to the regulation of the patterning genes; 3) a combination of both. In order to test these hypotheses or to generate new hypotheses to understand the evolutionary changes in the root hair development pathway in different lineages, it is necessary and important to conduct large-scale comparative transcriptomic analyses on the root hair development process in diverse plant species across the plant kingdom.

Overview of Dissertation Thesis

This work uses RNA-Seq to measure the genome-wide transcription activity of seven plant species across the plant kingdom, including an early vascular plant lycophyte (*Selaginella moellendorffii*), and members of modern angiosperm monocots (*Oryza sative* and *Zea mays*) and eudicots (*Solanum lycopersicum*, *Arabidopsis thaliana*, *Glycine max*, and *Cucumis sativus*), in order to find the conserved/diverged key players regulating the root development process and the

epidermal cell specification regulatory pathway. This is one of the first large-scale studies that focuses on the gene expression similarity between plant species from different orders.

Chapter Two describes the similarity and dissimilarity of gene expression profiles in the root of seven vascular plant species, providing evidence to answer the question of root origin in lycophyte Selaginella. It is widely accepted that Selaginella root has an independent evolutionary origin compared to other plant roots in angiosperms (Raven and Edwards, 2001; Jones and Dolan, 2012). Therefore, it is extremely interesting to test whether roots of different origins still maintain similar expression programs for root development. By using the GreenPhyl-defined gene family composition from seven vascular plants (Conte et al., 2008a, 2008b; Rouard et al., 2011), I compared the gene expression profiles between orthologous genes from different species and found a family-wise conservation in the gene expression profiles across all plant species tested, including Selaginella. I also reconstructed maximum likelihood phylogenetic trees of 133 key root development genes (71 families) identified in Arabidopsis and found that most of them are repeatedly used in Selaginella with a similar expression pattern to their relatives in angiosperms.

Chapter Three focuses on the root-hair specific transcriptional similarity between Arabidopsis and rice and extends the analysis to other vascular plants. Two root-hair specific expressed GFP marker lines were used to isolate the root-hair cells from Arabidopsis and rice (Kim et al., 2006; Brady et al., 2007b), respectively, from which the root hair transcriptomes were measured and compared. In order to study the root hair development pathway in more details, a set of 563 root hair morphogenesis genes that are under the regulation of RHD6 with moderate expression in the root hair cells was identified by a differential expression analysis between the *rhd6* mutant and the wild-type Arabidopsis. I showed that the gene expression

programs of these 563 root hair morphogenesis genes are less conserved in rice, compared to the broad conservation of the total genes that are expressed in the root-hair cells. This divergence in gene family structure and gene expression was not unique to rice, but was also found in other vascular plants tested. More differences in gene family structure and gene expression profiles were found in the Arabidopsis root hair patterning genes, which work at the upstream of RHD6 to regulate the root-hair cell pattern in the epidermis. The expression of these root hair patterning genes in the meristematic zone of root development is positively correlated with the root-hair cell pattern in Arabidopsis, which is not found in any other vascular plants tested. The results suggest the Arabidopsis-specific changes in the gene family structures and gene expression for both the root hair patterning genes and the downstream morphogenesis genes.

The final chapter includes the conclusions and future directions for comparative transcriptomics across plant species to understand the molecular mechanisms and evolutionary origin of the root hair pattern diversification in land plants.

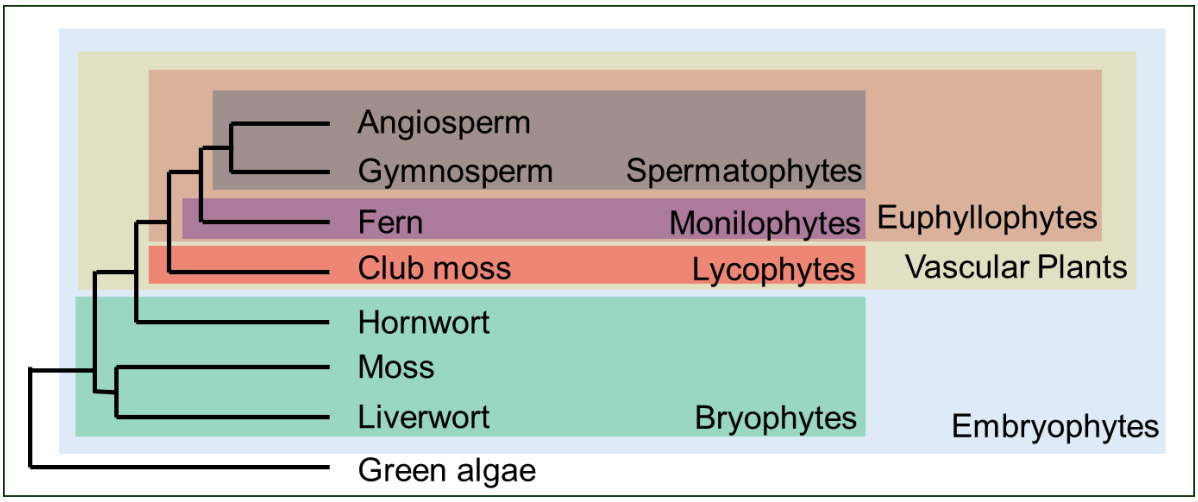

Figure 1.1 Land plants phylogeny

The tree topology is based on Hinchliff et al., 2015 (Open Tree of Life Project: http://www.opentreeoflife.org/)

This is a simplified phylogenetic tree of land plants (embryophytes). The land plants quickly diversified after the water to land transition. The basal land plants include three bryophytes: liverwort, moss, and hornwort, though the intra-clade relationship remains unsolved. The early vascular plants include two lineages proliferated by spores: lycophytes and monilophytes. The seed plants (spermatophytes) contain two large groups: gymnosperms (seeds plants without flower) and angiosperms (seeds plants with flower). Monilophytes and spermatophytes together are called euphyllophytes.

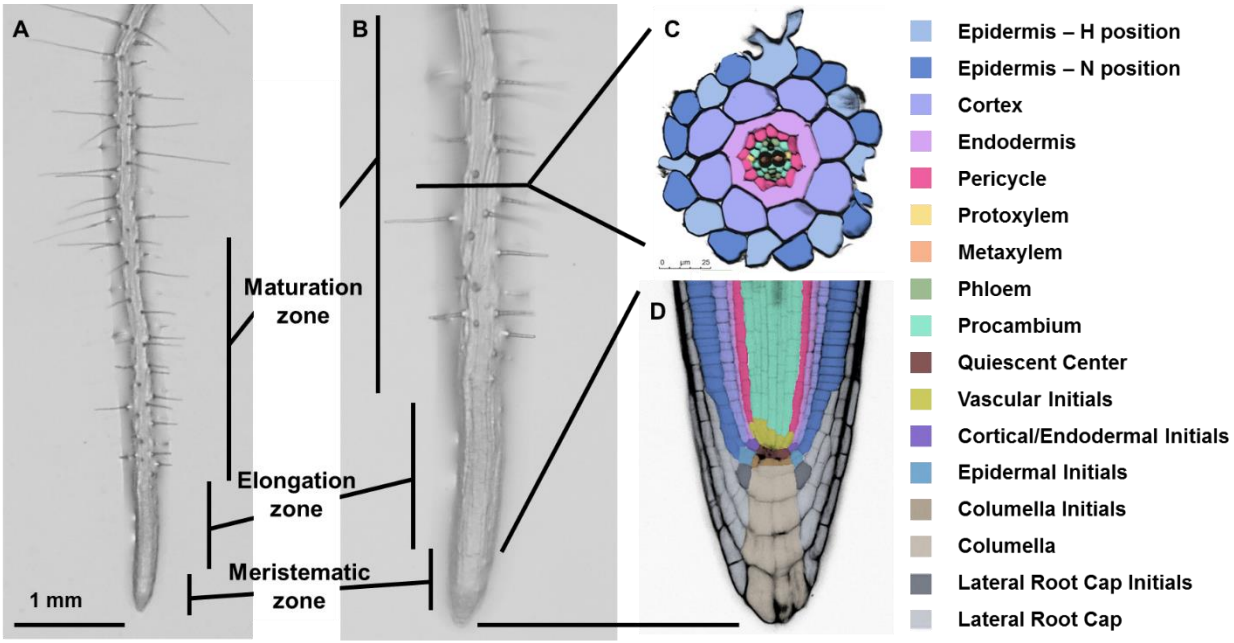

Figure 1.2 Root of *Arabidopsis thaliana*

A. A 5-day-old Arabidopsis seedling. From the most apical part of the root tip to the most distal part is the meristematic zone, elongation zone, and maturation zone.

B. A higher magnification view of the same seedling.

C. A cross section of the Arabidopsis maturation zone. From the outermost layer to the center of the root is the epidermis (H and N positions shown in different colors), cortex, endodermis, pericycle, phloem, procambium, metaxylem, and protoxylem.

D. A longitudinal section of the Arabidopsis root meristematic zone. From the most apical part of the root tip to the most distal part is the columella, columella initials, lateral root cap, lateral root cap initials, epidermal initials, epidermis, cortical/endodermal initials, cortex, endodermis, quiescent center, vascular initials, pericycle, and vascular tissues.

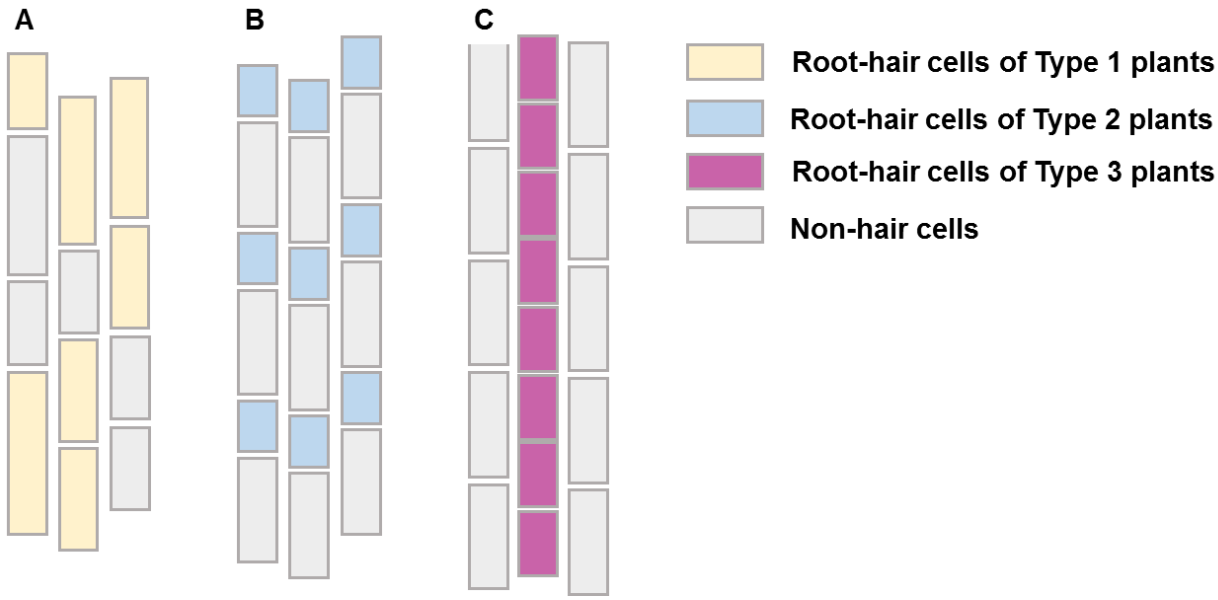

Figure 1.3 Root hair patternings

Diagrams showing the Type 1 (A), Type 2 (B), and Type 3 (C) root hair patternings. Root-hair cells are colored in yellow for Type 1 plants, blue in Type 2 plants, and pink in Type 3 plants whereas non-hair cells are colored in light gray.

Root of the Type 1-random pattern has the root-hair cells randomly distributed in epidermis.

Root of the Type 2-alternate pattern has the root-hair cells and the non-hair cells alternating along each cell file.

Root of the Type 3-position-dependent pattern has the root-hair cell file and the non-hair cell file(s) alternating in epidermis.

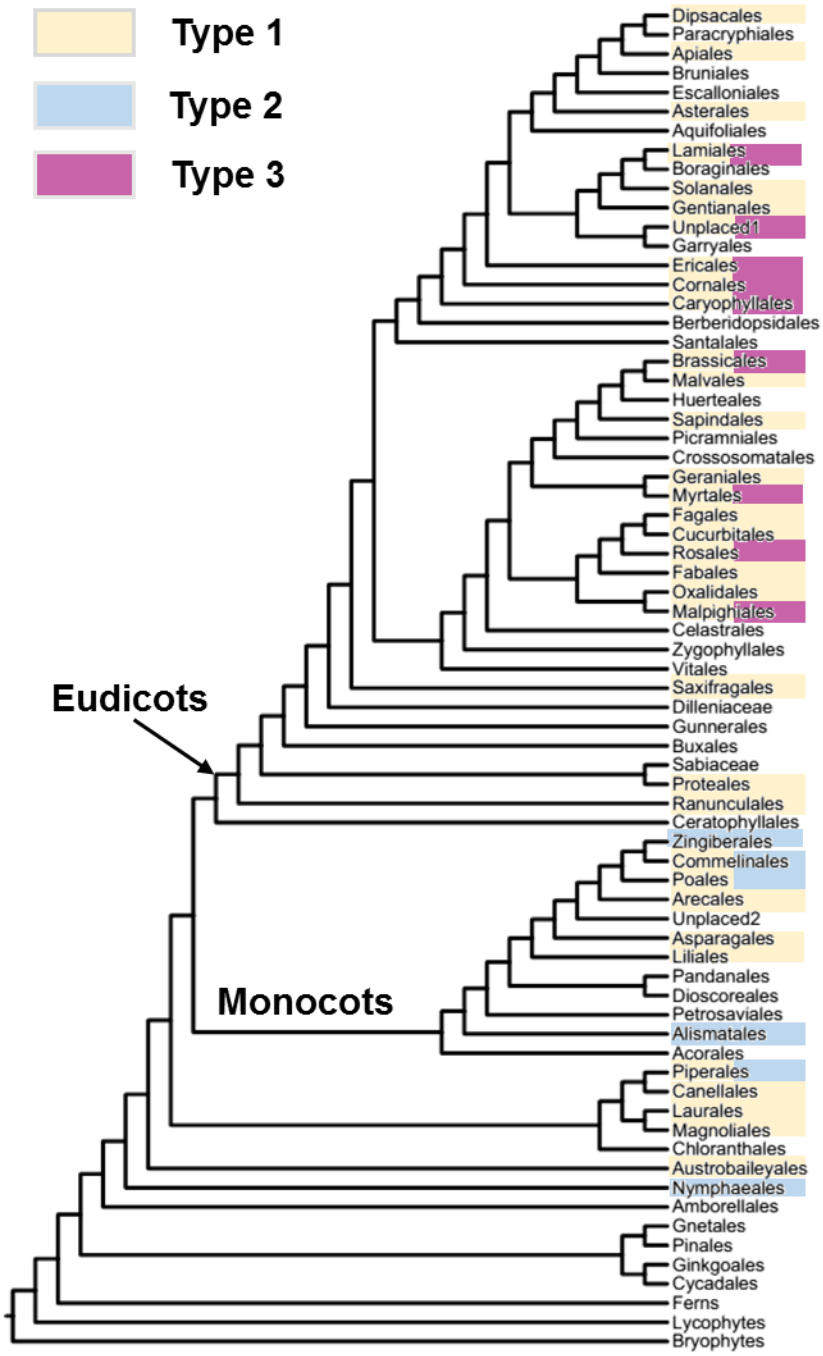

Figure 1.4 Root hair patternings of angiosperms

The species tree topology is based on THE ANGIOSPERM PHYLOGENY GROUP, 2009 with root hair patternings information obtained from Clowes, 2000 and Pemberton, 2001.

Orders examined are shaded in yellow for Type 1, blue for Type 2, and pink for Type 3 plants. The Type 1 plants have been found across all angiosperms. The Type 2 plants have been found in orders that diverged before Eudicots. The Type 3 plants have been found only in Eudicots.

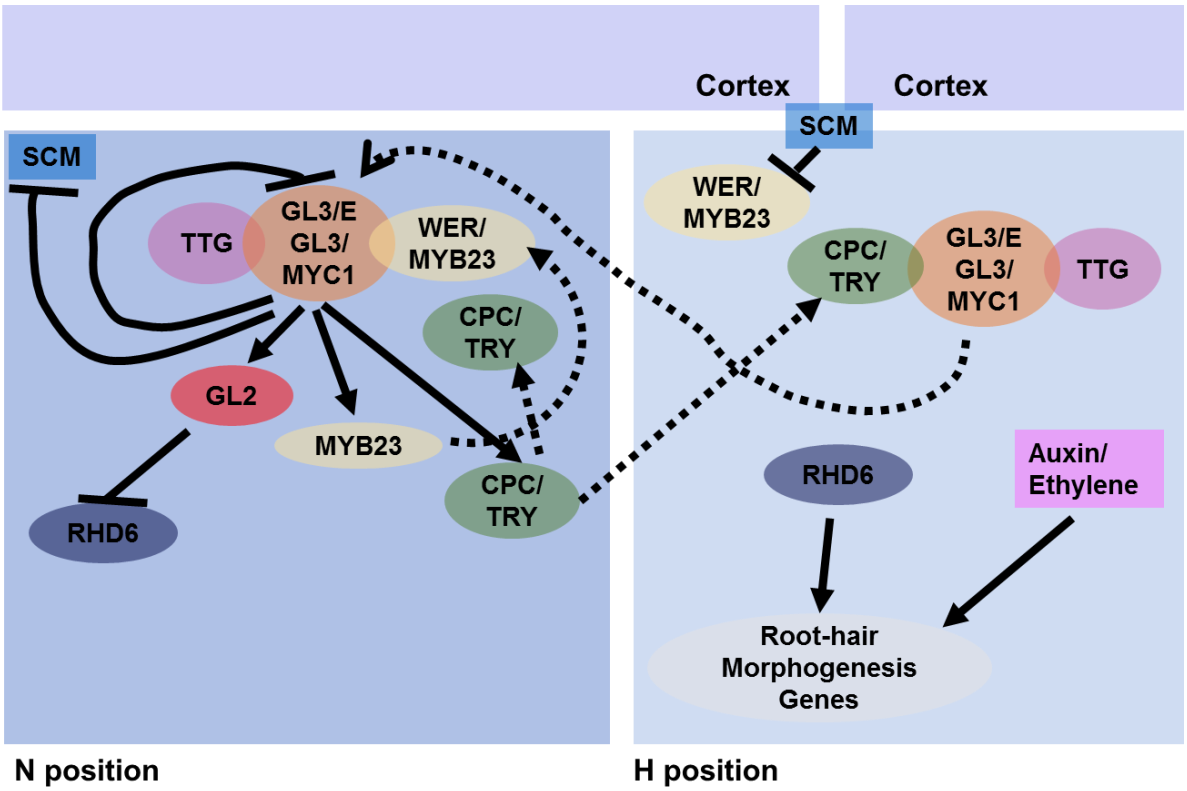

Figure 1.5 Molecular mechanism of the Type 3 - position-dependent root hair patterning of Arabidopsis

A dynamic regulatory network is responsible for the epidermal cell fate specification in Arabidopsis as discussed in this chapter. This figure summarizes the transcriptional effect between the key regulators with activation shown in arrows, inhibition shown in bars, and movements shown in dashed lines.

CHAPTER 2

Conserved Gene Expression Programs in Developing Roots from Diverse Plants

The contents of this chapter were previously published on The Plant Cell (Huang and Schiefelbein, 2015). RNA-Seq samples of rice, maize, soybean, and cucumber was prepared by Xinhui Shi. I prepared all the other samples, performed experiments and analyzed data.

Abstract

The molecular basis for the origin and diversification of morphological adaptations is a central issue in evolutionary developmental biology. Here, we defined temporal transcript accumulation in developing roots from seven vascular plants, permitting a genome-wide comparative analysis of the molecular programs used by a single organ across diverse species. The resulting gene expression maps uncover significant similarity in the genes employed in roots and their developmental expression profiles. The detailed analysis of a subset of 133 genes known to be associated with root development in *Arabidopsis thaliana* indicates that most of these are used in all plant species. Strikingly, this was also true for root development in a lycophyte (*Selaginella moellendorffii*), which forms morphologically different roots and is

thought to have evolved roots independently. Thus, despite vast differences in size and anatomy of roots from diverse plants, the basic molecular mechanisms employed during root formation appear to be conserved. This suggests that roots evolved in the two major vascular plant lineages either by parallel recruitment of largely the same developmental program or by elaboration of an existing root program in the common ancestor of vascular plants.

Introduction

The establishment of plants on land over 400 million years ago represented a critical stage in the history of life on Earth (Kenrick and Crane, 1997; Raven and Edwards, 2001; Gensel, 2008; Doyle, 2013). This transition was associated with numerous physiological and developmental innovations in plants, including in some lineages, the evolution of an exploratory multicellular subterranean organ (the root) suited for effective water and nutrient acquisition and plant anchorage. Considering the fossil record and root morphology in extant plants, it is generally accepted that roots evolved independently on more than one occasion during vascular plant evolution (Kenrick and Crane, 1997; Raven and Edwards, 2001; Kenrick and Strullu-Derrien, 2014).

Roots from extant plant species vary widely in size, cellular anatomy, and physiological properties (Figure 2.1) (Esau, 1965; Rost, 2011; Seago and Fernando, 2013). Furthermore, the overall architecture of a root system can vary (e.g., tap root versus fibrous root system) and typically includes different root types (e.g., primary, lateral, and adventitious roots). Nevertheless, these roots share certain fundamental features, including a set of terminal protective cells (the root cap), a self-sustaining stem cell population (the root apical meristem), a

radial organization of basic tissue types (from outermost epidermis tissue to innermost vascular tissue), the ability to form branch roots, the capacity to acquire and transport water and nutrients, and a tendency to grow in a downward direction (positive gravitropism). Furthermore, a common feature of root development is the spatial separation of the major cellular activities in distinct zones along the longitudinal axis at the root tip, typically including the meristematic zone (MZ; cell division, cellular patterning, and root cap formation), elongation zone (EZ; cell expansion), and differentiation zone (DZ; cell maturation) (Figure 2.1).

The molecular genetic basis for root development has been studied intensively in *Arabidopsis thaliana* and a large collection of genes involved in the patterning, growth, cellular differentiation, and maintenance of roots has been identified in this species (Bennett and Scheres, 2010; Petricka et al., 2012). Furthermore, global gene expression patterns have been defined for specific cells, tissues, and developmental stages of Arabidopsis roots, primarily using microarray-based methods (Brady et al., 2007a; Birnbaum et al., 2005). Molecular studies of root development have also been conducted in other plants (Hochholdinger and Tuberosa, 2009; Jansen et al., 2013; Qiao and Libault, 2013; Karve and Iyer-Pascuzzi, 2015), although the extent to which the molecular mechanisms identified in Arabidopsis roots apply to other plants is not clear.

Here, we describe a broad molecular analysis of gene expression in developing roots of vascular plants, focusing on the tips of early-stage roots prior to branching. This was facilitated by the similar developmental zonation in roots and the availability of genome sequence information from diverse plant species, enabling a detailed comparative analysis of temporal gene expression patterns during root formation in seven plant species. The resulting gene expression maps indicate that, despite considerable variation in the size and cellular anatomy of

roots from different species, these roots share a common developmental program. These data provide a foundation for the use of the plant root as a model for exploring the conservation and diversification of molecular mechanisms during plant organ evolution.

Material and Methods

Biological Material and RNA Isolation

Seeds of *Arabidopsis thaliana* (Columbia wild-type *WER*::*GFP*), tomato (*Solanum lycopersicum* Heinz 1706), soybean (*Glycine max* Williams 82), cucumber (*Cucumis sativus* Gy14), rice (*Oryza sativa spp japonica* cv Nipponbare), maize (*Zea mays* B73), and bulbils of *Selaginella moellendorffii* (Plant Delights) were germinated on agarose-solidified nutrient media under constant light as previously described (Schiefelbein and Somerville, 1990). The growing tips of the angiosperm seedling primary roots and rhizophore-derived *S. moellendorffii* roots were dissected (prior to branching) along the longitudinal axis using landmarks of cell length and root hair production. The MZ segment represented the terminal portion of the root, cut at the position where the length of cells began to exceed their width. The DZ segment included the first initiated root hair until the point where root hairs first reach their full length. The EZ segment represented the root portion between the MZ and DZ. Root sections were frozen immediately after collection, and total RNA was extracted from frozen samples using Qiagen RNeasy Plant Mini Kit. Library construction was performed by the University of Michigan Sequencing Core using the Illumina TruSeq Kit followed by sequencing on Illumina HiSequation 2000 System.

Microscopy Imaging

Plant seedlings were embedded in 3% agarose gel and sections from the DZ were obtained by hand-sectioning. Samples were stained with Fluorescent Brightener 28 (Sigma-Aldrich) for 5 to 20 s prior to examination with an Olympus IX81 microscope.

RNA-Seq Analysis

A total of 1.649 billion 50-bp single-end reads were generated from the 60 RNA samples (average of 27.5 million reads per sample). Reads were assessed by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and the initial 15 bp of the 50-bp reads (containing low quality sequence information) were trimmed before further processing. Raw reads were mapped to the corresponding reference genome by TopHat (version 2.0.3) (Kim et al., 2013), embedded with Bowtie2 (version 2.0.0-beta7) (Langmead and Salzberg, 2012) and SAMtools (version 0.1.19) (Li et al., 2009) with default settings [–segment length 17]. The mapped reads were quantified by Cufflinks2 (version 2.1.1) (Trapnell et al., 2012) with the correction for multiread on [-u -G]. An updated TopHat version (2.0.9 with embedded Bowtie2 2.1.0) was used for combined region analysis using same setting as described above.

Cucumber (v122) and *S. moellendorffii* (v91) reference sequences were downloaded from Phytozome (http://www.phytozome.net/). The other five genome sequences were downloaded from Ensembl (v19, http://plantsensembl.org/index.html).

S. moellendorffii gene expression values were all multiplied by 2 to correct for the duplication of the reference genome sequences.

Transcript sequence data from the Arabidopsis shoot inflorescence meristem were downloaded from http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1946/samples/ (Mantegazza et al., 2014). For consistency in gene comparison, the downloaded raw data was remapped to the reference annotation used for this study using the same parameters listed above. The reprocessed data yielded a Spearman's correlation coefficient of 0.94, relative to the previously published processed data (Mantegazza et al., 2014). A gene was deemed to be inflorescence meristem expressed if the mean FPKM ≥ 0.5 and the FPKM > 0 for each of the two replicates.

Gene Differential Expression Analysis

Raw counts of Arabidopsis were extracted from Cuffdiff2 (v2.1.1 with default setting plus correction for multiread and fragments bias [-b -u -N]) (Trapnell et al., 2012) and analyzed together with the edgeR software package (Robinson et al., 2009) to define gene sets preferentially expressed in zones. Briefly, genes with low expression were filtered out (counts per million should > 1 for at least three out of nine samples). Next, the raw counts were normalized using the upper quartile method, and the sample variation was estimated by tag-wise dispersion. The default trimmed mean of M-values method was not used because the assumption that most genes were not differentially expressed was violated. The resulting P values were corrected by Benjamini-Hochberg method (Benjamini and Hochberg, 1995) for multiple testing. Genes with a FC ≥ 2 as well as a FDR q-value ≤ 0.05 were considered to be zone preferentially expressed.

Comparisons between RNA-Seq and Published Microarray Data

Microarray data were downloaded from the available website (Brady et al., 2007a) (http://www-plb.ucdavis.edu/labs/brady/software/BradySpatiotemporalData/). Expression values from zones 1 to 6, 7 and 8, and 9 to 12 for a given probe were combined to create the mean expression in meristematic, elongation, and differentiation zones, respectively. Next, the data from two independent biological replicates were averaged, and gene expression was measured as the mean across all its corresponding probes. RNA-Seq data were averaged from three independent biological replicates. A value of one was added to all data prior to log2 transformation. A total of 22,262 genes contained data from both data sets and were used for comparisons.

Gene Family Information

The composition of gene families in the seven plant species was obtained from GreenPhyl (v4) (Rouard et al., 2011) and is presented in Appendix B Supplemental Data Set 2.

Statistical Analyses

The distribution of root-expressed and non-root-expressed genes among families in Arabidopsis was analyzed by χ^2 tests, controlling for gene family size (2, 3, 4, 5, or 6 Arabidopsis genes/family) and assuming random distribution. In addition, the distribution of genes preferentially expressed in the MZ, EZ, or DZ among Arabidopsis families and the distribution of root-expressed and non-root-expressed genes from the six angiosperm species

among families were analyzed by χ^2 tests. All statistical analysis and graph plotting was performed in the R environment (http://www.R-project.org) unless mentioned specifically. Figures were plotted using "fmsb" (Nakazawa, 2014), "ggplot2" (Wickham, 2009), and "gplots" (http://crantastic.org/packages/gplots/versions/36507) packages in R.

PCA

Gene expression FPKM values of samples from the same species were all elevated by1 and then transformed to log2 scale, mean centered for PCA by "prcomp" function in R. The first and second principal components were found to account for over 78% of the total variation in the data set. PC1 and PC2 from all species were plotted on the same figure.

Gene Expression Pattern Type Assignment

Root-expressed genes (defined as FPKM $\geq$ 0.5 in at least one zone and FPKM > 0 for at least two out of three biological replicates) with FC < 2 between all zones were assigned expression pattern type 0. The remaining root-expressed genes' expression profiles were standardized such that mean = 0 and SD = 1, followed by fuzzy C-means clustering using R package "Mfuzz" (Kumar and E Futschik, 2007). After monitoring the minimum distances between cluster centroids, the number of groups was optimized as 9. Genes were assigned the pattern type (types 1 to 9) with which they exhibited the highest affinity value.

For the two-zone expression profile comparisons, five pattern types were assigned to the root-expressed genes. FC was calculated as EDZ FPKM/MZ FPKM (FC was set at 10 for genes

with MZ FPKM = 0): type 1, FC ≥ 3; type 2, 1.5 < FC < 3; type 3, 0.67 ≤ FC ≤ 1.5; type 4, 0.33 < FC < 0.67; type 5, 0 ≤ FC ≤ 0.33.

Fisher's Exact Test Analyses

To test for association between genes in different species in a given gene family by expression pattern type, a Fisher's exact test was performed. Given two species, A and B, and two expression profile types, i_a and i_b (i belongs to 1 to 9 for angiosperm or 1 to 5 for *S.moellendorffii* comparison), the number of GreenPhyl gene families that possess exactly one A gene and one B gene (with an expression profile type from 1-9), one A gene and two B genes, or two A genes and one B gene were counted as the background total. For these three circumstances, overlapping gene families were those containing at least one gene from A with pattern i_a and at least one gene from B with pattern i_b. Nonoverlapping gene families were those containing at least one gene from A with pattern type i_a, but no gene from B with pattern type i_b and vice versa. When comparing types from the same species, only families possessing exactly two genes were considered. Overlapping gene families were those containing each of the types of interest. Nonoverlapping gene families were those containing only one type of interest. The Fisher's test was performed in R by "fisher.test". The resulted P values were corrected by Benjamini-Hochberg method (Benjamini and Hochberg, 1995) to avoid multiple testing errors.

Generation and Analyses of Supergenes

Supergene expression was obtained by summing the transcript expression values from all of genes from the same species within a given family. Expression profile types were then assigned

by clustering using the "Mfuzz" program (Kumar and E Futschik, 2007) as above or by FC (described below). Comparisons within species for the same expression profile types were assigned a P value of 0, whereas comparisons within species for different expression profile types were assigned a P value of 1 because each species was only able to have one expression profile type within a given family in this analysis. Overlapping gene families were counted as ones possessing supergenes from two species with the same profile types. Nonoverlapping gene families were counted as the ones possessing supergenes with different profile types in the two species being considered. The total number of gene families possessing supergenes from two species was used as background.

For supergene expression comparisons across all seven species, the 5027 GreenPhyl families that possessed at least one root-expressed supergene from each of the seven species were analyzed. The EDZ/MZ gene expression FC was calculated for each species' supergene within these 5027 families, and these were used to construct a matrix containing the pairwise Pearson's correlation coefficients (Figure 2.3C). To create angiosperm expression profiles, the FC values from dicot/monocot supergenes were averaged, respectively. Next, the mean of the dicot and monocot FC values were used to represent the overall angiosperm expression profiles and compared with the corresponding *S. moellendorffii* supergene expression profiles from the same families.

Connectivity

CNV was calculated as the average pairwise Pearson's correlation coefficient of a given gene's expression profile to all the genes' profiles within a given family, similar to the reported

methods (Koenig et al., 2013).

Hierarchical Clustering

For each assigned gene expression pattern type, the overlapping ratio of expression pattern types between two species was calculated as the number of families that possessed at least one gene from each species with a given expression pattern type divided by the number of families that possessed at least one gene from either species with a given expression pattern type. The dissimilarity was measured as 1 - overlapping ratio. The average dissimilarity across nine expression profile types was used for hierarchical clustering by the "hclust" function in R with default "complete" method.

GO Term Enrichment Test

GO term enrichment analysis was performed by DAVID (http://david.abcc.ncifcrf.gov/) with P value corrected by Benjamini-Hochberg FDR method (Benjamini and Hochberg, 1995). The analysis was performed on the Arabidopsis genes present in the gene families with CNVs > 5.9 (10.5% of total) that had been clustered into nine groups by the "Mfuzz" program (Kumar and E Futschik, 2007). The "gene family expression profile" in the three developmental zones was generated by summing the standardized expression values of each super gene within a given family. Significantly enriched terms with Benjamini-Hochberg (Benjamini and Hochberg, 1995) corrected P value ≤ 0.01 were found for five of these nine groups (3, 4, 5, 7, and 8).

Phylogenetic Analysis

BLAST (v2.2.26+) (Camacho et al., 2009) was used to search for candidate homologous genes (e-value ≤ 1) in a given database composed of protein sequences from five or six species. Then, pairwise comparisons between all candidate genes were performed by a fast Smith-Waterman search (SWIPE version 2.0.7) (Rognes, 2011) to cluster sequences into homolog groups by a connected component clustering method essentially as previously described (Bernardes et al., 2015) (Appendix B Supplemental Methods 1). The assumption in the clustering approach was that orthologous sequences should be more closely related compared with nonorthologous sequences. After clustering, one longest protein sequence for each gene model was retained. Evolution models were tested by Modelgenerator (version 0.85) (Keane et al., 2006). All related sequences were first aligned by MAFFT (version 6.864b) (Katoh and Standley, 2013) with setting [--genafpair --ep 0 --maxiterate 1000], and the maximum likelihood tree was computed by RAxML (version 7.7.8) (Stamatakis, 2006) with JTT model using empirical base frequencies, gamma distribution of rate heterogeneity, and 100 rapid bootstrap test [-m PROTGAMMAJTTF -f a -N 100] or, if the tree size exceeded 100, first alignment was conducted with FastTree (Price et al., 2009) default setting with [-gamma] option on. Next, sequences from a well-supported clade (defined as a monophyletic group including genes from all species, unless the closely related genes were included in another well-supported clade) with bootstrap ≥ 70 or FastTree local support ≥ 0.85 together with sequence from an outgroup (neighboring) clade or gene (a closely related Arabidopsis gene with the smallest BLAST e-value) were realigned by MAFFT with the same settings. Alignments were trimmed using trimAI (version 1.2rev59) (Capella-Gutiérrez et al., 2009) with setting [-automated1] or [-gappyout] (for WOX5 family tree only). Alignments are provided in Appendix B, using the SeaView (v.4.3.3)

platform (Gouy et al., 2010). The final trees were reconstructed by RAxML with 1000 rapid bootstrap [-m PROTGAMMAJTTF -f a -N 1000]. Most trees were rooted between two well-supported clades. If no related well-supported neighbor clade could be identified, the tree was rooted by an Arabidopsis outgroup gene. If neither was available, the tree was rooted between monocots and eudicots (for the angiosperm species trees) or between lycophytes and euphyllophytes (for the vascular plant trees). Norway spruce protein sequences were downloaded from http://congenie.org/start/ (Nystedt et al., 2013).

Accession Numbers

Sequence data from this chapter can be found in the Gene Expression Omnibus under accession number GSE64665 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64665).

Results

Gene Expression in Root Development Zones of Arabidopsis

We first analyzed gene expression in developing Arabidopsis primary roots by sequencing mRNA from longitudinal sections of the MZ, EZ, and DZ (Figure 2.1A; see Methods). This yielded a total of 21,037 root-expressed genes (mean fragments per kilobase per million mapped reads [FPKM] ≥ 0.5 in MZ, EZ, or DZ; expression detected in at least two out of three biological replicates), exhibiting diverse transcript accumulation patterns in the three zones (see sequence submission information; Appendix B Supplemental Data Set 1). As validation, we surveyed the literature and found our expression data matches each of 19 genes' transcript accumulation profiles previously determined by in situ RNA hybridization (Appendix B Supplemental Table 1)

(Birnbaum et al., 2005). More broadly, our RNA-Seq-based gene expression values positively correlate with previous microarray-based expression values (Brady et al., 2007a) obtained from equivalent Arabidopsis developmental zones (r values: 0.72 [MZ], 0.65 [EZ], and 0.7 [DZ]; see Methods for details), although as expected, our RNA-Seq data exhibit substantially greater dynamic range, permitting greater accuracy for extreme expression values (Figure 2.2).

Using previously defined Arabidopsis gene family assignments (GreenPhyl v4) (Rouard et al., 2011), we discovered that the root-expressed Arabidopsis genes are not randomly distributed, but tend to cluster among families ($P < 0.001$, χ^2 test; Appendix B Supplemental Data Sets 2 and 2.3). Similarly, we observed a nonrandom (clustering) distribution among families for those genes exhibiting preferential root zone expression (≥2.0 fold-change [FC]; false discovery rate [FDR] ≤ 0.05) in the MZ, the EZ, or the DZ ($P < 0.001$ for each gene set, χ^2 test; Appendix B Supplemental Data Sets 3 and 4). The tendency for related Arabidopsis genes to possess similar root expression characteristics implies that root expression patterns tend to be conserved in gene lineages.

Root Gene Expression Is Conserved across Angiosperms

To analyze root developmental gene expression across angiosperms, we obtained MZ, EZ, and DZ transcriptomes (three biological replicates from each zone) from the primary roots of five additional angiosperm species: three eudicots (tomato [*Solanum lycopersicum*], soybean [*Glycine max*], and cucumber [*Cucumis sativus*]) and two monocots (rice [*Oryza sativa*] and maize [*Zea mays*]) (Figure 2.1 and 2.3A). Principal component analysis (PCA) shows that the major variation in transcript accumulation across these samples is explained by differential expression in different development zones (PC1 + PC2 accounted for >78% total variation). This

pattern of developmental zone-driven gene expression variation is highly correlated across all six of the angiosperms tested (Figure 2.3B).

To further investigate whether patterns of primary root gene expression are conserved among these species, we examined the distribution of root-expressed genes across angiosperm gene families. Among the 6613 gene families that possess at least one gene from each of these six angiosperms (Rouard et al., 2011), we discovered a nonrandom distribution of root-expressed genes ($P < 0.001$; χ^2 test), with root-expressed genes from different species tending to associate in families (Appendix B Supplemental Data Sets 2 and 3). This suggests that gene families devoted to root functions have been conserved during angiosperm evolution.

Next, we evaluated the possibility that related genes from different species possess similar temporal root expression profiles across the three developmental zones. Using a fuzzy C-means clustering algorithm, root-expressed genes exhibiting variation in transcript accumulation across the MZ, EZ, and DZ (≥ 2.0 FC between any two zones) from each of the six species were assigned to one of nine dominant gene expression profile types (Figure 2.4A; Appendix B Supplemental Data Set 5). Pairs of species were then compared to determine whether matching profile types were preferentially observed for genes in the same families (using gene families possessing exactly one or two genes from each of the six species; see Methods). Indeed, the most statistically significant familial associations were found for genes with the same expression profile type (Figure 2.4B). Specifically, among the 135 pairwise interspecies comparisons for the same profile type, 132 of them exhibited significant familial association (corrected $P < 0.01$, Fisher's exact test; Figure 2.4B). Furthermore, we found that the average ratio of overlapping expression types in a given family between species mirrored their known phylogenetic relationships (Figure 2.4C; see Methods). These results suggest conservation of the regulation of

gene expression during primary root development in angiosperms.

A difficulty in accurately comparing gene expression between species is the variation in gene number per species within families. To address this and enable an aggregate comparison of gene expression profiles in multigene families, we generated "supergenes" for each species, by summing root transcript accumulation values (separately for the MZ, EZ, and DZ) for all genes from the same species within a given family (Appendix B Supplemental Data Set 6). These supergenes were clustered into nine expression profile types, similar to above, and using pairwise species comparisons, we again observed statistically significant familial association of supergenes bearing the same expression type (Figure 2.4D).

Next, we generated connectivity values (CNVs) for the supergenes by analyzing the correlation in expression profile between each supergene and the supergenes in its family (CNV ranges from -1 to 1; see Methods). The CNVs for all six supergenes in a given family were summed (for the 6161 families that contain root-expressed supergenes from each of the six angiosperms), as an estimate for expression profile similarity across species' genes in the family. Strikingly, a large fraction of these 6161 families exhibit very high summed CNVs, relative to families constructed by random assignment of genes (Figure 2.4E; Appendix B Supplemental Data Set 7). This provides additional evidence for the conservation of root gene expression profiles at the family level across all angiosperm species.

To further study those families exhibiting the greatest conservation in root gene expression, we selected families with summed supergene CNV ≥ 5.9 (10.5% of total families). After summing the standardized supergene expression data in each of these families, the resulting "family expression profiles" were clustered into nine expression types, similar to above, and Gene Ontology (GO) term enrichment analysis was conducted for each group. Interestingly,

distinct sets of significantly enriched terms were obtained from the different expression profile types (Appendix B Supplemental Table 2). For example, the largest proportion of highly significant GO terms from cluster type 7 (high expression in DZ, relative to MZ and EZ) was related to transcriptional regulation (Appendix B Supplemental Table 2). Considering their high level of expression profile conservation across angiosperms, these families are likely to include previously unidentified genes important for angiosperm root function.

To focus on specific genes likely involved in root development, we used available root gene information from Arabidopsis. We identified 133 Arabidopsis genes previously reported to have a function in primary root development, comprising 71 families, encoding transcription factors and other putative regulatory proteins (Appendix B Supplemental Table 3) (Petricka et al., 2012). Phylogenetic trees were constructed that contain these Arabidopsis genes and their relatives from all six angiosperm species (Figure 2.4F; Appendix B Supplemental Figure 1). In addition, our root gene expression data was mapped onto these trees, and CNVs were calculated for the genes in each clade, enabling an assessment of sequence and expression relationships. Among these 71 families, 67 of them yielded well-supported clades containing the known Arabidopsis root development gene(s) together with a root-expressed gene(s) from each of the other five angiosperm species (Appendix B Supplemental Figure 1). Among the remaining families, one was eudicot specific and three lacked a related gene in one angiosperm species. Furthermore, the genes in these families generally exhibited conservation in root expression profiles, as demonstrated by their high family CNV values, compared with randomly constructed families (Figure 2.4E). Together, the results suggest that regulators of primary root development genetically defined in Arabidopsis are employed similarly by other angiosperms.

Analysis of Root Gene Expression in the Lycophyte *Selaginella moellendorffii*

Roots are found in two major clades of extant vascular plants: euphyllophytes (including angiosperms, gymnosperms, and monilophytes [ferns]) and lycophytes (a non-seed plant clade that diverged from the euphyllophyte lineage; ~400 million years ago) (Banks, 2009). Therefore, to compare root development more broadly, we analyzed root gene expression in a sequenced lycophyte, *S. moellendorffii* (Banks et al., 2011). We defined the transcriptomes of *S. moellendorffii* roots (produced from rhizophores via bulbils) from the MZ and the combined EZ + DZ (EDZ) (necessary due to the superimposition of EZ and DZ characters in the *S. moellendorffii* root; Figure 2.1A) for three biological replicates each (Figures 2.5A and 2.5B; Appendix B Supplemental Data Set 8). Among the 5465 gene families containing at least one gene from *S. moellendorffii* and each of the six angiosperms (defined by GreenPhyl) (Rouard et al., 2011), we discovered a significant association by family for root-expressed genes in *S. moellendorffii* and root-expressed genes in angiosperms ($P<0.001$; Fisher's exact test; Appendix B Supplemental Data Sets 2 and 3). Specifically, 81.6% of families that contained a root-expressed gene from each of the six angiosperms also contained a root-expressed *S. moellendorffii* gene.

We next compared root gene expression profiles from the six angiosperms and *S. moellendorffii* by converting the three-zone angiosperm transcript data to two-zone (by combining the EZ and DZ expression values; ≥ 0.5 FPKM in at least one zone, expression detected in at least two out of three biological replicates) (Appendix B Supplemental Data Set 8). We then assigned each gene to one of five expression profile types, based on expression fold change between the MZ and EDZ (FPKM EDZ/FPKM MZ; type 1, $FC \geq 3$; type 2, $1.5 < FC < 3$; type 3, $0.67 \leq FC \leq 1.5$; type 4, $0.33 < FC < 0.67$; type 5, $FC \leq 0.33$; Appendix B Supplemental

Data Set 9). Using pairwise species comparisons to assess the frequency of matching expression patterns within families, we discovered significant familial association when the same expression types were compared between species (Fisher's exact test; Figure 2.5C), indicating that gene expression profiles are generally conserved within families containing *S. moellendorffii* and angiosperm genes.

To compare the overall degree of similarity in expression profiles in these species, we generated supergenes for each species' genes in a given family (by summing expression values by zone) and compared supergene expression FC between the MZ and EDZ in *S. moellendorffii* and the six angiosperms by family (Appendix B Supplemental Data Set 10). We assigned each of the supergenes to one of the five expression profile types based on their expression FC and, using pairwise species comparisons, we again observed a significant familial association of genes from different species exhibiting the same profile type (Fisher's exact test; Figure 2.5D; Appendix B Supplemental Data Set 10). We also discovered an overall positive correlation between gene expression profiles in *S. moellendorffii* and each angiosperm (r values: 0.49 to 0.57), albeit lower than for intra-angiosperm comparisons, consistent with the greater evolutionary divergence between *S.moellendorffii* and the angiosperms (Figure 2.5E). Next, we calculated the average EDZ/MZ expression for supergenes from each of the six angiosperms (in a weighted manner; see Methods) within a given family to generate a combined angiosperm EDZ/MZ expression value that was compared with the corresponding *S. moellendorffii* supergene's value from the same family. We found a strong correlation in these EDZ/MZ values (F-statistic $P < 0.001$; Figure 2.5F), providing further evidence for family-dependent similarity in root gene expression profiles between *S. moellendorffii* and angiosperm genes. Interestingly, mapping angiosperm family CNV onto these results shows that supergenes with extreme

expression FC values tend to exist in high-CNV families (Figure 5F), implying that the most conserved expression patterns are the ones that exhibit the greatest difference between these developmental zones.

To assess relationships among specific genes likely to encode root regulators, we analyzed *S. moellendorffii* genes related to the 133 known Arabidopsis root development genes (Appendix B Supplemental Table 3). Phylogenetic trees were constructed that contain these 71 families of Arabidopsis genes and related genes from rice (as a representative monocot), *S. moellendorffii*, Norway spruce (*Picea abies*, a gymnosperm), and the moss *Physcomitrella patens* (a nonvascular, rootless plant), and we mapped our root expression data and calculated gene CNVs for the resulting clades (Appendix B Supplemental Figure 2). Strikingly, we discovered that 67 of the 71 well-supported clades containing the Arabidopsis root genes also possessed a root-expressed *S. moellendorffii* gene(s) (Figures 2.6 and 2.7; Appendix B Supplemental Figure 2). Furthermore, in 53 of these 67 clades, at least one *S.moellendorffii* gene matched the expression profile type of the Arabidopsis gene(s). These results indicate that *S.moellendorffii* largely possesses and expresses the same genes known to be critical for root development in angiosperms. In addition, we found that 70 of these 71 clades possessed a related gene from Norway spruce (Appendix B Supplemental Figure 2), suggesting that conservation of the developmental gene program extends to gymnosperms, although we do not have root expression data to fully support this suggestion.

Among these 71 clades, two of them possess Arabidopsis genes regulating root cap formation. The root cap is believed to be a root-specific innovation with no shoot counterpart (Barlow, 2002; Bennett and Scheres, 2010). The *FEZ* gene of Arabidopsis promotes root cap stem cell activity (Willemsen et al., 2008) and, interestingly, it is part of a vascular plant specific

clade with representatives sharing preferential MZ expression (Figure 7). The *SOMBRERO*, *BEARSKIN1*, and *BEARSKIN2* genes participate in root cap maturation in Arabidopsis (Bennett et al., 2010) and are included in a large clade with a preferential MZ-expressed gene(s) from each vascular plant species (Appendix B Supplemental Figures 1 and 2). These results are consistent with the possibility that root cap-associated gene functions are shared in vascular plants.

Lastly, we considered the possibility that the similarity in root gene expression among these species is due to a general molecular program acting in all developing organs. To examine this, we compared our root transcriptome data sets to an available transcriptome data set from the shoot inflorescence meristem of Arabidopsis (Mantegazza et al., 2014). We identified Arabidopsis gene families that lack any shoot meristem-expressed genes and then analyzed the distribution of root-expressed genes from the seven species within these families (Appendix Supplemental Data Set 3). We discovered a statistically significant association of root-expressed genes by family for each pairwise comparison of Arabidopsis and each of the other species (corrected $P < 0.001$ for each comparison, Fisher's exact test), implying conserved root expression characteristics for these families of genes that are not expressed in all developing organs. For example, 96% of the families (102/106) possessing a *S. moellendorffii* root-expressed gene (and lacking an Arabidopsis shoot meristem-expressed gene) also possessed a root-expressed gene from at least one of the angiosperm species. Although limited by its use of a single non-root developing organ, this analysis indicates that the similarity in root gene expression we observed among these species is not likely to be due solely to a general molecular program shared by all developing plant organs.

Discussion

In this study, we defined gene expression maps from the developing roots of seven different plant species, enabling a comprehensive comparative analysis of the molecular genetic control of a developing organ type in plants. The most general finding is that, despite the vastly different sizes and cellular structures of roots from these seven species (Figure 2.1), there is substantial conservation in the usage and expression of their genes in root development. Regarding conservation in gene usage, we discovered a statistically significant degree of overlap in the gene families containing root-expressed members from the various species, indicating that related genes are used for root development across all species. Regarding conservation in gene expression, we found significant similarities in the expression profiles from the root developmental zones for related genes in the same family across different species. These findings were observed both in a genome-wide analysis of all root-expressed gene families as well as in the specific analysis of 71 families containing 133 Arabidopsis genes encoding known root developmental regulators. These results suggest a common molecular program, employing similar genes and gene regulation, is used in developing roots across vascular plants.

These findings provide insight into the history of gene innovation/recruitment during root evolution in vascular plants (Figure 2.8). A large number of gene families (6004), including a large fraction of the known Arabidopsis root developmental gene families (67/71), contain root-expressed genes from each of the seven vascular plant species, implying that representatives of these families were recruited to participate in root development at an early stage (Figure 2.8). Smaller numbers of root gene families possess root-expressed genes in specific plant subgroups (e.g., angiosperms, eudicots, and monocots) or from a single species, likely reflecting later gene gains and losses in distinct lineages. These lineage-specific gene families may be responsible, in

part, for the differing root characteristics that exist in particular plant clades. Extending this study to include additional species with varying root architecture (e.g., fibrous versus tap root) and to include different root types (e.g., primary versus lateral versus adventitious) will likely link specific genes/families to particular root characters and provide a more complete picture of root evolution. We also note that some of these families possess a relatively large number of angiosperm genes, indicating substantial gene expansion in certain angiosperm lineages (as previously reported for other developmental gene families) (Feller et al., 2011), which may also explain some of the variation in root phenotypes.

It is remarkable that the roots of *S. moellendorffii* and angiosperms appear to share a similar molecular developmental program because the lycophyte and euphyllophyte lineages of plants are generally thought to have evolved roots independently (Kenrick and Crane, 1997; Raven and Edwards, 2001). This view is supported by the available fossil evidence, which indicates that early euphyllophytes lacked roots at a time when lycophytes possessed them (Raven and Edwards, 2001; Friedman et al., 2004). Furthermore, lycophyte roots exhibit some unusual developmental features, including branching by bifurcation rather than the endogenous lateral root formation typical of euphyllophytes (Raven and Edwards, 2001; Banks, 2009).

We consider two general explanations for the similar root gene expression patterns in lycophytes and angiosperms: parallel recruitment of largely the same developmental program independently in the lycophyte and euphyllophyte lineages or the existence of a primitive root developmental program in their common ancestor. Regarding the possibility of parallel recruitment, strong selective pressures and a limited genetic "toolkit" may have restricted the evolutionary path for root formation in both lineages. In this vein, it is notable that roots and shoots of extant plants deploy many of the same (or closely related) developmental genes

(Benfey, 1999; Stahl and Simon, 2010), likely due to their common origin from a primitive telomic axis (Kenrick and Crane, 1997; Gensel and Berry, 2001; Friedman et al., 2004; Ligrone et al., 2012; Tomescu et al., 2014). Similarly, roots that evolved independently in separate lineages might still be expected to share a substantial fraction of their developmental program due to recruitment from a largely common pool of organ development genes. To examine this issue rigorously, it will be necessary to define developmental transcriptomes, equivalent to the root developmental transcriptomes analyzed here, from multiple organs of lycophytes and angiosperms. Related to this, we note that the observed similarity in gene usage and expression at the family level reported here may overestimate the degree of functional similarity in these genes because individual genes within a given family may have undergone substantial functional diversification.

On the other hand, the possibility that a primitive root program existed prior to the divergence of lycophytes and euphyllophytes is also consistent with the substantial similarity in root gene expression profiles and, in particular, the gene families associated with root cap formation. The root cap was a unique evolutionary innovation, not present in the shoot (or in the presumed telomic axis precursor), so genes for its specification and formation would be expected to be distinct if roots evolved independently. Indeed, a detailed molecular dissection of the Arabidopsis root meristem has led to the proposal that the root cap, and its associated meristematic cells, is a structure that evolved separately from the major portion of the root, representing a later innovation that enabled the root to more effectively penetrate soil and generate modern-day “true roots” (Bennett and Scheres, 2010). In addition, the strength of the fossil evidence supporting independent root evolution has been called into question, due to its incomplete nature and the poor preservation of fossilized roots, leading some to consider the

origin of roots an unsettled issue (Gensel, 2008). Thus, it is conceivable that the common ancestor of lycophytes and euphyllophytes had already possessed a rudimentary root developmental program, perhaps generating a transitional "rooty structure" (Doyle, 2013) that was subsequently modified. It will be necessary to conduct detailed studies of individual root genes identified here (e.g., the root cap genes) to determine whether their similarity in sequence and expression across species is mirrored by similarity in developmental function.

The gene expression data sets described here represent a useful resource for future studies of root molecular biology. In addition to eliciting and testing evolutionary hypotheses, these data should assist in the identification of new root-expressed genes and functionally related homologs of previously defined root genes. In particular, the genes with strongly conserved root expression profiles across all the species described here are likely to include novel regulators of root development and function.

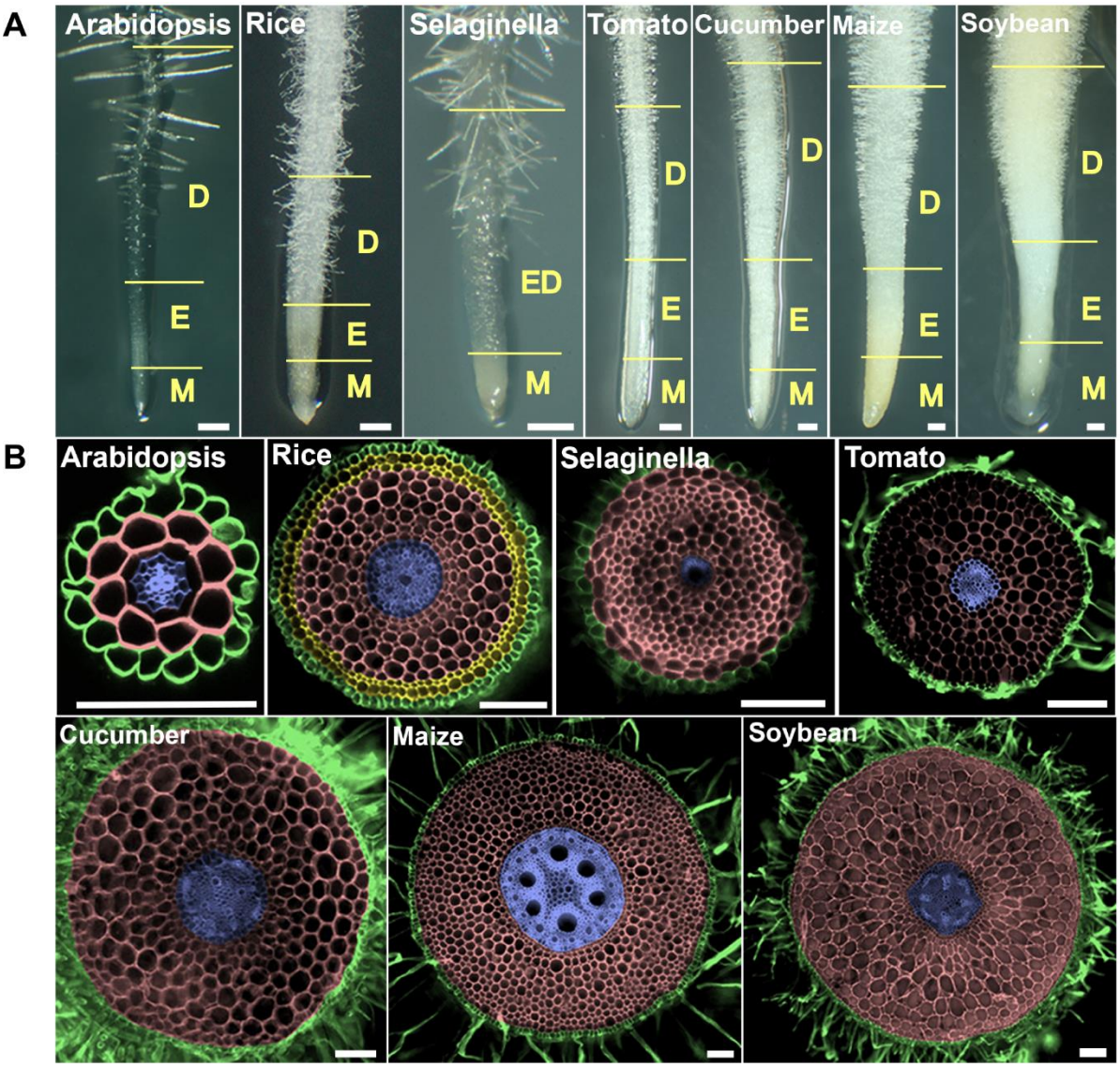

Figure 2.1 Root development in land plant species

A. Primary seedling roots of Arabidopsis (4 d), rice (4 d), Selaginella (grown from bulbils, 21 d), tomato (6 d), cucumber (4 d), and soybean (4 d). M = meristematic, E = elongation, D = differentiation, ED = overlapping elongation and differentiation. Scale bar = 0.1 mm.

B. Transverse section of primary seedling roots of each species. Pseudo-color green = epidermis, red = cortex, blue = endodermis + pericycle + vascular tissue, and yellow = exodermis + sclerenchymatous layer. Scale bar = 0.1 mm.

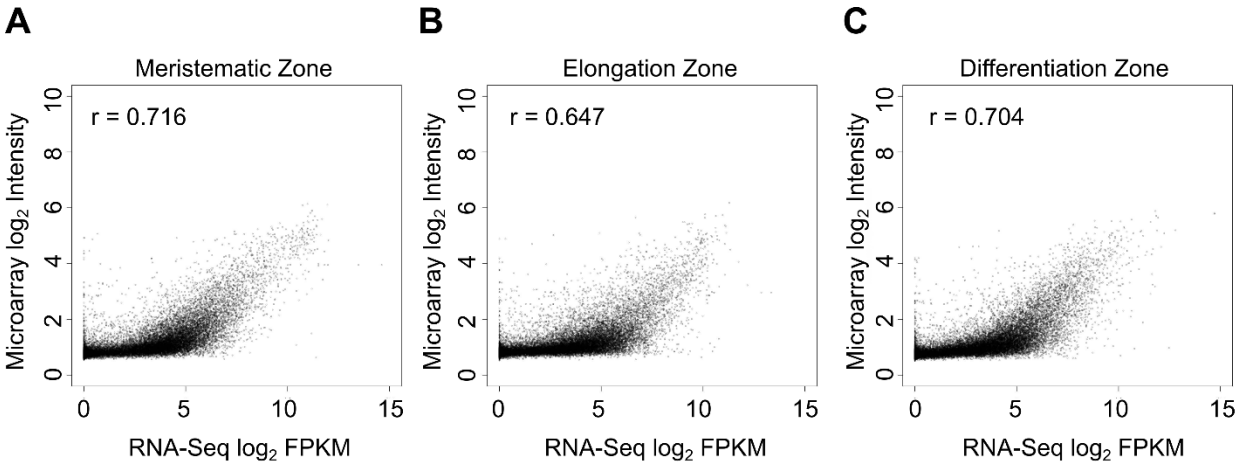

Figure 2.2 Comparison of RNA-Seq results to published microarray data

A = meristematic zone, B = elongation zone, and C = differentiation zone. r = Pearson's Correlation Coefficient.

RNA-Seq FPKM values were averaged from three independent biological replicates. Microarray data was averaged from two independent biological replicates. One was added to all values prior to log2 transformation.

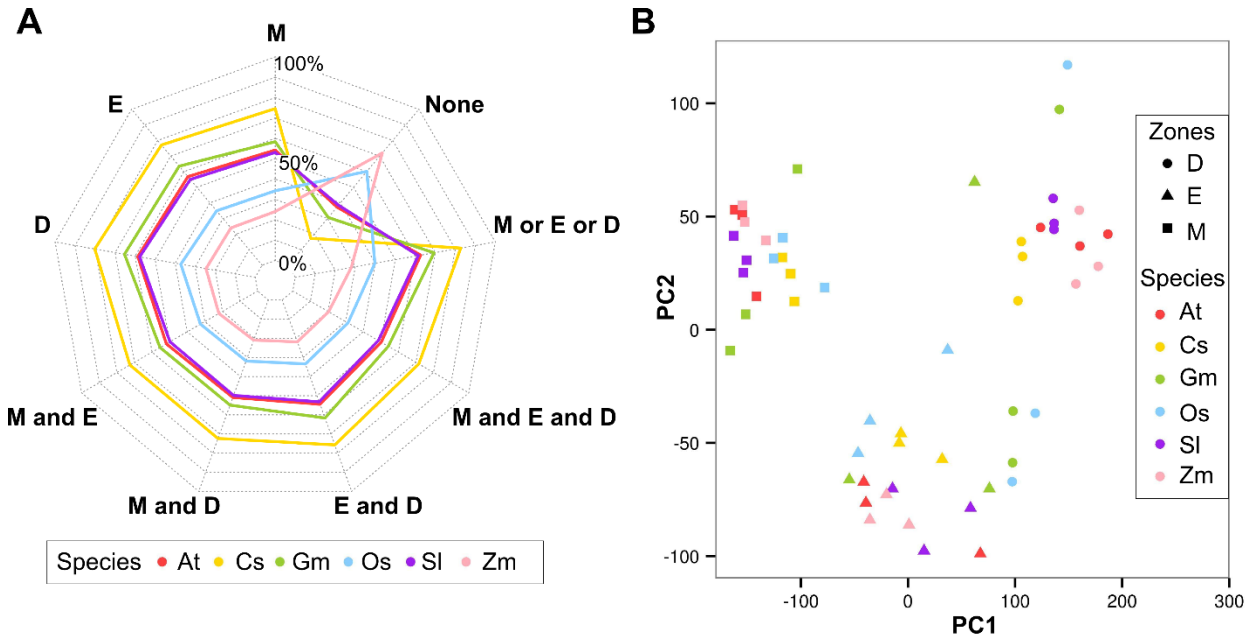

Figure 2.3 Gene expression preferences in three development zones across six angiosperm species

A. Ratio of genes expressed (average FPKM >= 0.5, at least 2 replicates have expression) in the root development zones across species. At = Arabidopsis, Cs = cucumber, Gm = soybean, Os = rice, Sl = tomato, Zm = maize. M = meristematic, E = elongation, and D = differentiation.

B. Merged Principal Component Analysis (PCA) of gene expression in the root development zones from six angiosperm species samples. PCA was performed on all individual biological replicates from the same species and then plotted to the same figure.

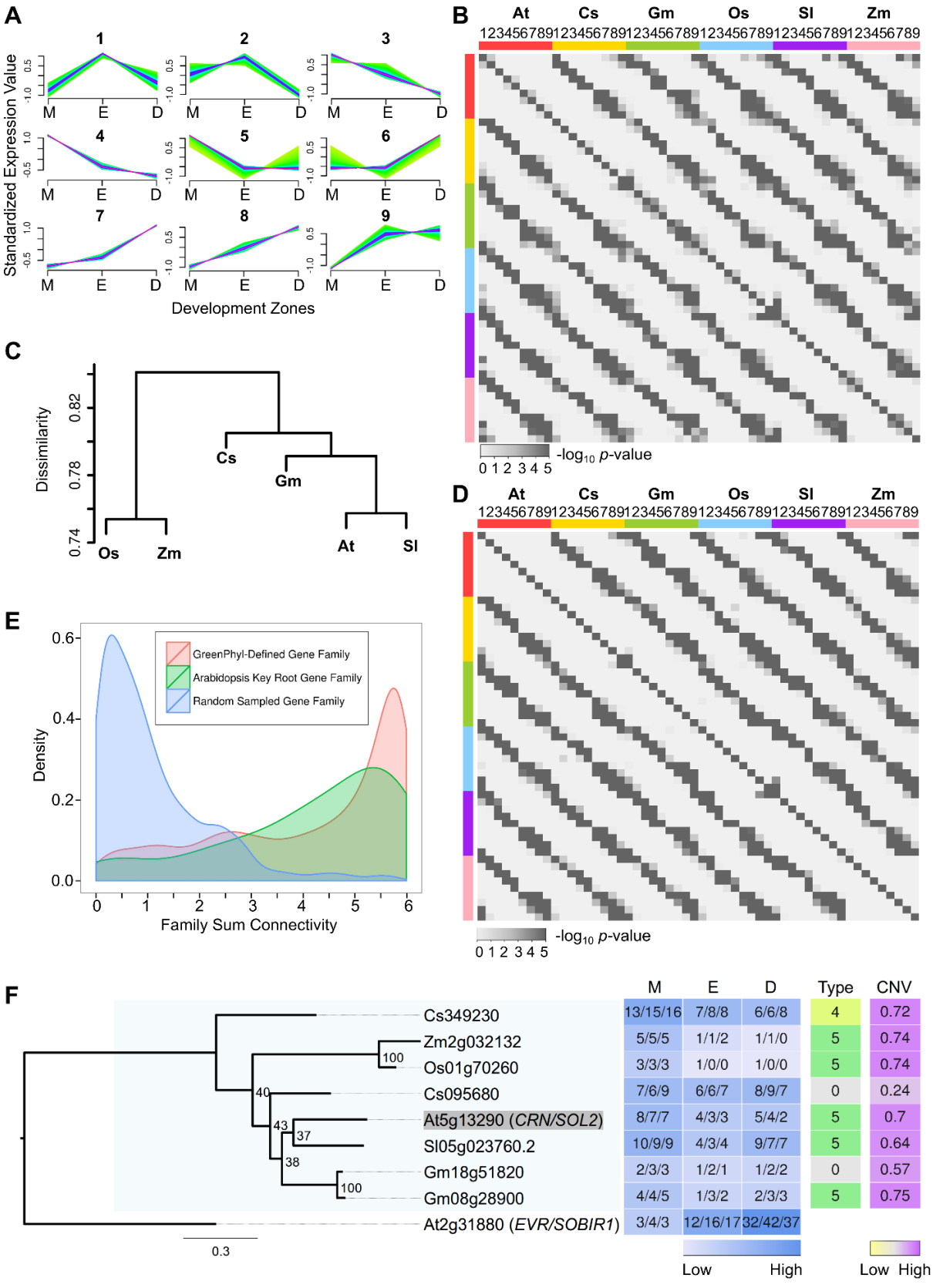
A
Standardized Expression Value
Development Zones
M
E
D
B
At
Cs
Gm
Os
Sl
Zm
-log10 p-value
C
Dissimilarity
Cs
Gm
Os
Zm
At
Sl
D
E
GreenPhyl-Defined Gene Family
Arabidopsis Key Root Gene Family
Random Sampled Gene Family
Density
Family Sum Connectivity
F
Cs349230
Zm2g032132
Os01g70260
Cs095680
At5g13290 (CRN/SOL2)
Sl05g023760.2
Gm18g51820
Gm08g28900
At2g31880 (EVR/SOBIR1)
M
E
D
Type
CNV
13/15/16 7/8/8 6/6/8 4 0.72
5/5/5 1/1/2 1/1/0 5 0.74
3/3/3 1/0/0 1/0/0 5 0.74
7/6/9 6/6/7 8/9/7 0 0.24
8/7/7 4/3/3 5/4/2 5 0.7
10/9/9 4/3/4 9/7/7 5 0.64
2/3/3 1/2/1 1/2/2 0 0.57
4/4/5 1/3/2 2/3/3 5 0.75
3/4/3 12/16/17 32/42/37
Low
High
Low
High

Figure 2.4 Comparison of root gene expression across angiosperms

A. Nine expression profile types in the MZ, EZ, and DZ assigned by fuzzy C-Means clustering. Gene expression with at least two fold-change between zones was standardized to have mean of 0 and standard deviation of 1. High affinity to the cluster centroid is shown in purple and low in green.

B. Heatmap of corrected one tail p-values of pair-wise Fisher's Exact Test for association of expression profile types within gene families.

C. Hierarchical clustering dendrogram based on the average dissimilarity ratio of gene families not sharing the same expression profile types.

D. Heatmap of corrected one tail p-values of pair-wise Fisher's Exact Test for association of expression profile types within supergene families.

E. Density plot of family total connectivity, which summed the connectivity values (CNV) of the 'supergenes' of each species within given families. Red = 6161 GreenPhyl-defined families with root-expressed supergenes from all six angiosperms; Green = 73 GreenPhyl-defined families containing Arabidopsis known key root development genes; Blue = 1000 gene families with randomly assigned member genes.

F. Example of one of the 71 maximum likelihood phylogenetic trees reconstructed for relatives of Arabidopsis key root development genes. Tree reconstructed for Arabidopsis (At) CRN (shaded) and its relatives from rice (Os), maize (Zm), cucumber (Cs), tomato (Sl), and soybean (Gm). Also included are a heatmap indicating gene expression in the three development zones (M = meristematic zone; E = elongation zone; D = differentiation zone) (values indicate FPKM), the gene expression profile types (Type), and connectivity values (CNVs). The well-defined CRN clade with family members from all of the angiosperm species is shaded in light blue.

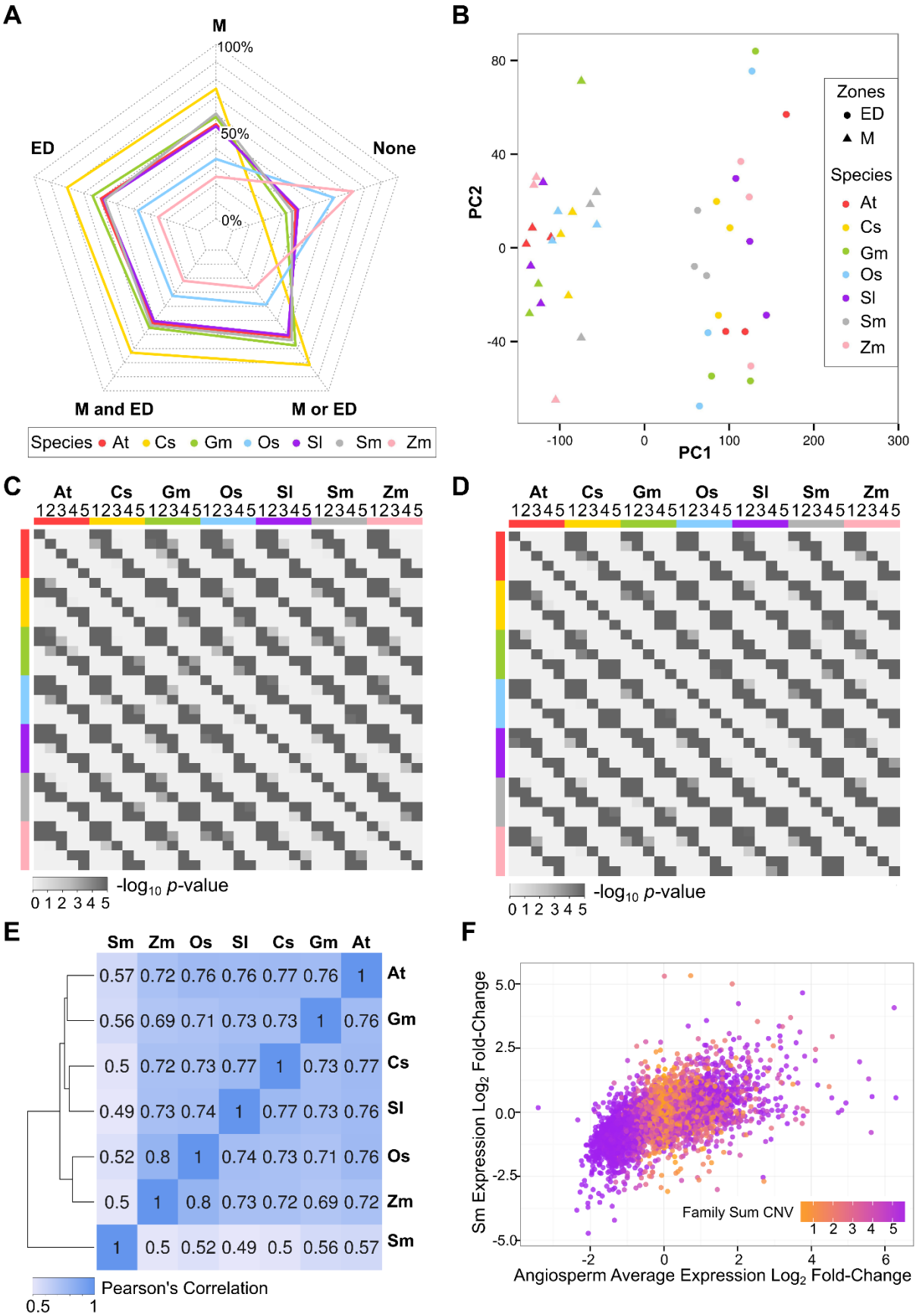
A
M
100%
50%
0%
None
M or ED
M and ED
ED
Species At Cs Gm Os Sl Sm Zm
B
Zones
ED
M
Species
At
Cs
Gm
Os
Sl
Sm
Zm
PC1
PC2
C
At Cs Gm Os Sl Sm Zm
-log10 p-value
0 1 2 3 4 5
D
At Cs Gm Os Sl Sm Zm
-log10 p-value
0 1 2 3 4 5
E
Sm Zm Os Sl Cs Gm At
0.57 0.72 0.76 0.76 0.77 0.76 1 At
0.56 0.69 0.71 0.73 0.73 1 0.76 Gm
0.5 0.72 0.73 0.77 1 0.73 0.77 Cs
0.49 0.73 0.74 1 0.77 0.73 0.76 Sl
0.52 0.8 1 0.74 0.73 0.71 0.76 Os
0.5 1 0.8 0.73 0.72 0.69 0.72 Zm
1 0.5 0.52 0.49 0.5 0.56 0.57 Sm
Pearson's Correlation
0.5 1
F
Sm Expression Log2 Fold-Change
Angiosperm Average Expression Log2 Fold-Change
Family Sum CNV
1 2 3 4 5

Figure 2.5 Comparison of root gene expression across seven vascular plants

A. Ratio of genes expressed (average FPKM ≥ 0.5, at least 2 replicates have expression) in the root development zones across species. Sm = Selaginella; ED = overlapping elongation and differentiation zone.

B. Merged PCA of gene expression in the root development zones from seven angiosperm species samples. PCA was performed on all individual biological replicates from the same species and then plotted to the same figure.

C. Heatmap of corrected one tail p-values of pair-wise Fisher's Exact Test for association of expression profile types assigned by EDZ/MZ fold-change within gene families.

D. Heatmap of corrected one tail p-values of pair-wise Fisher's Exact Test for association of expression profile types assigned by EDZ/MZ fold-change within supergene families.

E. Correlation matrix of supergene expression log2 EDZ/MZ fold-change in 5027 GreenPhyl-defined families. Heatmap was re-ordered according to the hierarchical clustering result.

F. Scatter plot comparing supergene expression log2 EDZ/MZ fold-change between the average of all angiosperms versus Selaginella, for 5027 GreenPhyl-defined families. Colors indicate family total connectivity values (orange, low; purple, high) (see also Figure 2.2E). Angiosperm average was calculated as the mean of the eudicot average and the monocot average, giving equal weight to the two clades. Regression F-statistic $p < 0.001$.

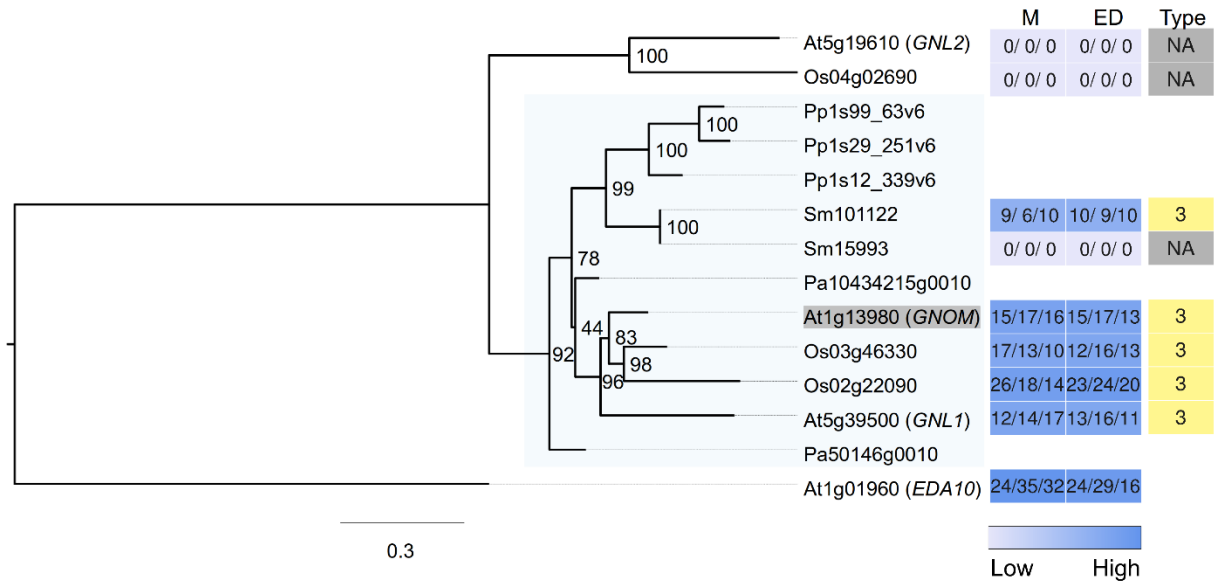

Figure 2.6 Maximum likelihood tree constructed for relatives of Arabidopsis *GNOM*

The gene expression level from the two development zones (M=meristematic zone, ED=combined elongation + differentiation zone) are shown as heatmaps (values indicate FPKM), together with the expression fold-change type. The closely related Arabidopsis gene *EDA10* is used as outgroup to root the tree. Expression fold-change types (1-5) for the genes are shown in different colors, whereas light grey indicates that no root expression was detected. The target gene (Arabidopsis *GNOM*) is shaded in grey and the target clade with family members from all species is shaded in light blue. Related genes are also included from two species (Pa = Norway spruce, Pp = Physcomitrella) which did not have their root expression assessed, indicated by blank spaces in the heatmaps.

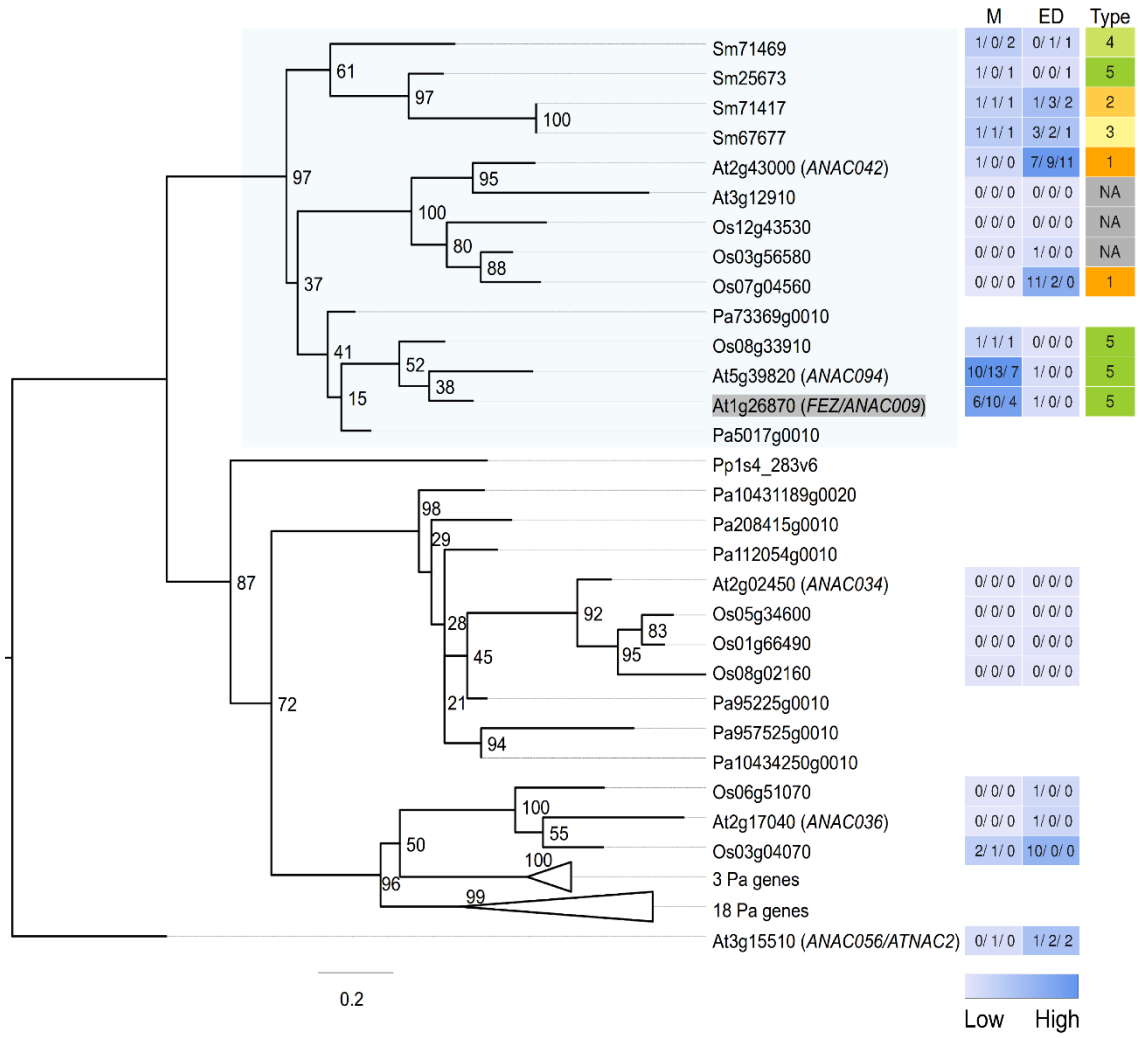

Figure 2.7 Maximum likelihood tree constructed for relatives of Arabidopsis *FEZ*

The gene expression level from two development zones (M and ED) are shown as heatmaps, together with the expression fold-change type. The closely related Arabidopsis gene *ANAC056* is used as outgroup to root the tree. Expression fold-change types (1-5) for the genes are shown in different colors, whereas light grey indicates that no root expression was detected. The target gene (Arabidopsis *FEZ*) is shaded in gray and the target clade with family members from all vascular plant species is shaded in light blue. Related genes are also included from two species (Pa = Norway spruce, Pp = Physcomitrella) which did not have their root expression assessed, indicated by blank spaces in the heatmaps.

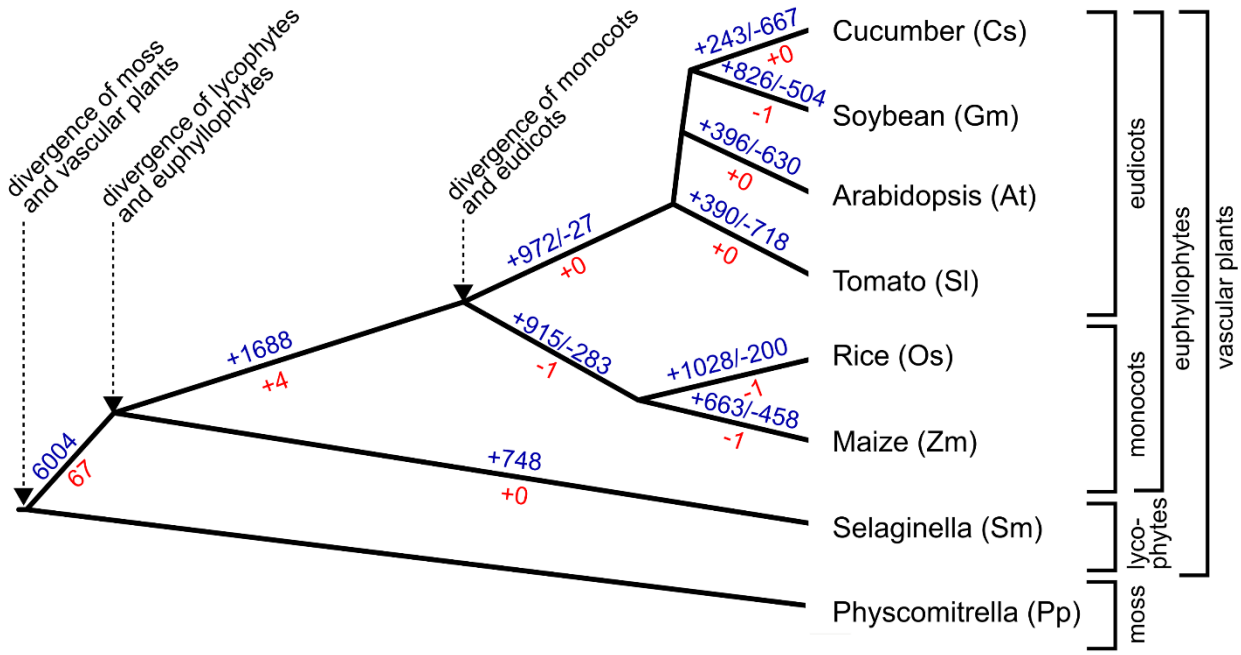

Figure 2.8 Evolutionary history of root-expressed gene families

Depiction of the phylogenetic relationships of the species examined in this study and putative origin of their root-expressed gene families. Blue = GreenPhyl-defined gene families; Red = Gene families containing an Arabidopsis gene known to be associated with root development (reconstructed by maximum likelihood). Positive numbers refer to putative lineage-specific gain of families containing root-expressed genes; negative numbers refer to putative lineage-specific loss of families containing root-expressed genes. See text for discussion.

CHAPTER 3

Diversification of Root Hair Development Genes in Vascular Plants

The contents of this chapter has been submitted for consideration to be published as a research manuscript in the journal, Plant Physiology. RNA-Seq samples were prepared by Xinhui Shi. I performed all the other experiments and analyzed the data.

Abstract

The molecular genetic program for root hair development has been intensively studied in Arabidopsis. To understand the extent to which this program might operate in other plants, we conducted a large-scale comparative analysis of root hair development genes from diverse vascular plants, including other eudicots, monocots, and a lycophyte. Combining phylogenetics and transcriptomics, we discovered conservation of a core set of Arabidopsis root hair genes across all vascular plants, which may represent an ancient program for unidirectional cell growth coopted for root hair development during vascular plant evolution. Interestingly, we also discovered preferential diversification in the structure and expression of root hair development genes, relative to other root hair expressed genes, among these species. These differences enabled the definition of sets of genes and gene functions that were acquired or lost in specific

lineages during vascular plant evolution. In particular, we found substantial divergence in the structure and expression of genes used for Arabidopsis root hair patterning, suggesting that the transcriptional regulatory mechanism used in Arabidopsis to specify root-hair cells is not used in these other species.

Introduction

A fundamental feature of organismal evolution is the creation and diversification of cell-type specific differentiation programs. These programs are responsible for generating the cellular diversity, and associated division of labor, that is the hallmark of multicellular organisms (Arendt, 2008). However, we still know relatively little about the evolution of the genetic and molecular mechanisms that establish cell differentiation programs and how they differ between different species.

The root hair cell is a useful single cell type for experimental studies in plant biology, and its development, physiology, and cell biology have been intensively studied in many plant species (Cormack, 1935; Datta et al., 2011; Qiao and Libault, 2013; Grierson et al., 2014). Root hairs are long tubular extensions of root epidermal cells that greatly increase the root surface area and thereby assist in water and nutrient absorption. The development of root hairs occurs in three basic stages; specification of the root hair cell fate, initiation of a root hair outgrowth, and elongation of the hair via tip growth. Root hairs are found in nearly all vascular plants, including angiosperms, gymnosperms, and lycophytes, suggesting a common evolutionary origin. However, different plant species are known to differ in their root hair distribution pattern and in their root hair (Clowes, 2000; Pemberton, 2001; Datta et al., 2011), suggesting genetic

differences exist in their root hair developmental program.

Root hairs have ben intensively studied in Arabidopsis. In particular, molecular genetic analyses have led to identification of numerous root hair genes which provide insight into the mechanisms of Arabidopsis root hair development (Balcerowicz et al., 2015; Bruex et al., 2012; Grierson et al., 2014; Salazar-Henao et al., 2016; Gu and Nielsen, 2013). Root-hair-bearing cells in Arabidopsis are specified by a set of early-acting patterning genes that generate a cell-position-dependent distribution of root-hair cells and non-hair cells via a complex transcriptional regulatory network (Grierson et al., 2014; Salazar-Henao et al., 2016). Once specified, the presumptive root-hair cells initiate the outgrowth of the root hair through the action of the *ROOT HAIR DEFECTIVE6* (*RHD6*) gene, which encodes a bHLH transcription factor that induces an extensive root hair gene expression program through activation of additional regulatory genes (Bruex et al., 2012; Masucci and Schiefelbein, 1994; Menand et al., 2007; Yi et al., 2010). This suite of downstream root hair morphogenesis genes generates the unidirectional expansion (tip growth) of the root hair (Balcerowicz et al., 2015; Datta et al., 2011). These genes encode proteins involved in secretory activities, cell wall synthesis, ion transport, reactive oxygen species regulation, and many other processes (Balcerowicz et al., 2015; Salazar-Henao et al., 2016). The expression profiles of the patterning genes, initiation genes, and morphogenesis genes differ along the longitudinal length of the root tip, which reflects their temporal importance in root hair development (Datta et al., 2011; Grierson et al., 2014).

The wealth of knowledge concerning the genetic control of root hair development in Arabidopsis provides an opportunity to evaluate the similarity in root hair development programs in other plants and thereby address fundamental issues regarding the evolution of cell differentiation mechanisms. Several focused studies have begun to investigate this issue, by

analyzing individual root hair genes/families in Arabidopsis and selected species to examine their molecular relationships (Brady et al., 2007b; Ding et al., 2009; Karas et al., 2009; Kim et al., 2007, 2006). In general, the results from these studies suggest that root hair developmental genes tend to share similar function in different species, implying conservation in their root hair development programs.

In the present study, we sought to comprehensively analyze root hair differentiation programs across vascular plants. We first defined the root hair transcriptome and root hair development genes in Arabidopsis and then analyzed the distribution and expression of these genes in six other vascular plant species. Although we found that many root hair genes are conserved across these species and therefore likely share similar roles, we also discovered significant differences in the structure and/or expression of some root hair development genes. In particular, we found poor conservation of Arabidopsis patterning genes, implying that root hair cell specification in these other plants does not employ the same regulatory proteins. These findings provide new insight into the conservation and diversification of plant cell differentiation programs in vascular plants.

Material and Methods

Biological Material and Growth Condition

Transgenic lines used for this study included *Atrhd6 WER::GFP* (Masucci and Schiefelbein, 1994), Arabidopsis WT *WER::GFP* (Lee and Schiefelbein, 1999), Arabidopsis WT *COBL9::GFP* (Brady et al., 2007b), and rice WT *EXPA30::GFP* (Kim et al., 2006). Seeds were surface sterilized and germinated on agarose-solidified MS media under constant light at 22 ^{o}C

as previously described (Schiefelbein and Somerville, 1990).

Microscopy

Young seedlings of Arabidopsis and rice (4-5 days after plating) were stained with propidium iodide for 1 min and the roots were examined with a Leica SP5 laser scanning confocal microscope. The excitation wavelength was 488 nm for the detection of GFP signals and 561 nm for the propidium iodide.

Young seedlings of all vascular plants were stained with toluidine blue for 5-10 sec and the roots and the root hairs were examined with a Leica Laborlux S microscope or a Wild M420 Makroskop.

For analysis of root hair distribution, the root epidermis of each species was examined with an Olympus IX81 after the root was stained with Fluorescent Brightener 28 for 30-60 sec or propidium iodide for 1 min. The root hair cells were pseudo-colored in purple and the non-hair cells were pseudo-colored in yellow.

RNA Isolation

RNA was isolated from protoplasts after fluorescence-activated cell sorting (FACS) as described previously (Birnbaum et al., 2005; Bruex et al., 2012). In brief, root tips of 4- or 5-day-old seedlings were pooled and digested using cell wall degrading enzymes. GFP positive cells were collected using FACS machine at the University of Michigan Flow Cytometry Core. Total RNA was extracted using Qiagen RNAeasy Plant Micro Kit. RNA quality was assessed by the

RNA Integrity Number provided by an Agilent Bioanalyzer and RNA samples with a score ≥ 8 were used for cDNA library construction by Illumina TruSeq Kit. Library samples were sequenced on Illumina HiSequation 2000 System. The library construction and the sequencing were performed at the University of Michigan Sequencing Core.

RNA-Seq Processing and Differential Expression Analysis

Sequencing reads were processed and analyzed as previously described (Huang and Schiefelbein, 2015). In short, the first 15 bp of each 50 bp-long read was trimmed before mapping to a reference genome using TopHat (version 2.0.3) (Kim et al., 2013) with default settings (--segment length 17). Gene expression was calculated using Cufflinks2 (version 2.1.1) (Trapnell et al., 2013) with multiread correction (-u -G). Reads generated from rice samples were processed using an updated version of TopHat (version 2.0.9) with the other steps unchanged.

Reference genome and annotation of Arabidopsis and rice were both downloaded from Ensembl Plant database (v19, http://plants.ensembl.org/index.html).

The number of raw counts mapped to each gene was quantified by HTSeq (version 0.6.1) (Anders et al., 2015) with setting (-m intersection-strict -s no -f bam) and analyzed using edgeR (Robinson et al., 2010) for differential expression analysis. First, genes with expression lower than the cutoff (counts per million > 1 for at least three out total six samples) were filtered out. Second, raw counts were normalized using the default trimmed mean of M-values method and the variation was modeled using a tag-wise dispersion. Next, the calculated P values were corrected for multiple testing by Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Significant differentially expressed genes were identified using a cutoff of fold-change (FC) >= 2

and a False Discovery Rate (FDR) q-value <= 0.01.

The $\log_2$ scaled gene expression value was added 1 before the $\log_2$ transformation.

Statistical Analyses

All statistical analyses and graph plotting were performed in the R statistical computing environment (R core team, https://www.R-project.org) unless mentioned otherwise.

The built-in R function "fisher.test" was used to calculate the P value for the Fisher's Exact Test. Background total was Green-Phyl defined families with a specific family size. A total of 9 combinations of different family sizes were tested (permutations drawn from 1-3 At genes and 1-3 Os genes). All families met size requirement were divided into four groups for the test: GFP expression in both species; GFP expression in Arabidopsis only; GFP expression in rice only; no GFP expression in either. The alternative hypothesis was that the observed data had greater association than expected from the null. The resulted P values were corrected by Bonferroni method (Dunn, 1961).

The Fisher's Exact Test for the association of the temporal expression profiles between *AtRH* and *OsRH* genes followed the previous analysis (Huang and Schiefelbein, 2015) with the background total to be the families with exact one *AtRH* gene and one *OsRH* gene, exact one *AtRH* gene and two *OsRH* genes, and exact two *AtRH* genes and one *OsRH* gene with the expression profile types 1-9. *AtRHM* families were used to test the association between *AtRHM* and *OsRH* genes.

Family Size Analysis

The composition of gene families in the seven plant species was obtained from GreenPhyl (v4) (Rouard et al., 2011) and is presented in Supplemental Data Set 2.

Each time, a set of 543 genes was randomly drawn from the total 12449 *AtRH* genes (excluding genes that are not included in the GreenPhyl database) and the number of families of these 543 genes was recorded. The process was repeated 1000 times and the distribution of the number of families was plotted as a histogram.

Construction and Analysis of Subfamilies Containing *AtRHM* and *OsRH* Genes

To analyze Arabidopsis-rice subfamilies of the *AtRHM* GreenPhyl-defined families, protein sequences from all seven vascular plants were obtained from each of the 304 GreenPhyl families that possess at least one *AtRHM* gene and one *OsRH* gene. Multiple sequence alignment was generated by MAFFT (-6.86b, --genafpair --op 0 --maxiterate 1000 for tree size < 200 or -auto for tree size ≥ 200) (Katoh and Standley, 2013). Phylogenetic trees were reconstructed using FastTree (v2.1, -gamma) (Price et al., 2009). The trees were rooted between two vascular clades, if applicable, or at the mid-point of the total tree and were plotted by the "ete2" package in Python (Huerta-Cepas et al., 2010). Well-supported (>0.85) subfamilies containing at least one Arabidopsis and one rice gene were identified, and the distribution of *AtRHM*, *AtRH(-RHM)*, *non-AtRH*, *OsRH*, and *non-OsRH* genes were analyzed within these subfamilies.

Supergene Expression Analysis

Supergene expression was calculated as previously described (Huang and Schiefelbein, 2015). In brief, the FPKM expression values were summed for genes from the same family in a given species. For this analysis, only the expression values in the root-hair cells were processed.

Expression Dissimilarity Analysis

The dissimilarity between the temporal expression profiles was measured as the absolute difference in the expression profile types between the *AtRHM* genes and their relatives in other vascular plants.

For each *AtRHM* family, the expression difference between each of its *AtRHM* gene and every gene from the other species was calculated and the minimal value was reported. The heatmap was generated by "gplots" package (https://cran.r-project.org/web/packages/gplots/index.html). The angiosperm data is based on comparison of gene expression in three developmental zones (i.e. 10 profile types), and the Selaginella data is based on two zone comparisons (i.e. 5 profile types).

The classification of *AtRHM* families according to their expression similarity were defined as follows: "vascular plant conserved", score = 0-4 in all vascular plants; "angiosperm conserved", score = 0-4 in all angiosperms and score=5 or 9 in Selaginella; "eudicot conserved", score = 0-4 in all eudicots and scores = 5 or 9 in maize, rice, and Selaginella; "Arabidopsis specific", score = 5 or 9 in all vascular plants.

Gene Ontology (GO) Term Enrichment Test

GO term enrichment analysis was performed by DAVID (Huang et al., 2009) (http://david.abcc.ncifcrf.gov/) on the 563 *AtRHM* genes versus the background total 33550 At genes on the genome.

Phylogenetic Analysis

Phylogenetic trees were reconstructed using maximum likelihood (ML) or an approximate maximum likelihood similar to previous methods published (Huang and Schiefelbein, 2015). Briefly, homologous sequences were identified using BLAST (v2.2.26+) (Camacho et al., 2009) and then clustered into groups. Groups with more than 200 members were aligned using MAFFT (version 6.864b) (Katoh and Standley, 2013) (-auto option) whereas smaller groups were aligned using MAFFT (--genafpair --ep 0 --maxiterate 1000 option). Next, large family alignment (>= 100) was sent to FastTree (Price et al., 2009) (v2.1.9, -gamma) for the approximate maximum likelihood tree reconstruction. A well-supported clade (a monophyletic clade with at least one member from all species, unless the members were included in another well-supported clade) with local support value >= 0.85 and its neighboring well-supported clade (or the most similar Arabidopsis gene as outgroup) were re-aligned using MAFFT (--genafpair --ep 0 --maxiterate 1000) option. Alignment was then trimmed using trimAl (version 1.2rev59) (Capella-Gutiérrez et al., 2009) with (--automated 1) option. Finally, trees were reconstructed using RAxML (version 7.7.8) (Stamatakis, 2006) (-m PROTGAMMAJTTF -f a -N 1000). Trees were rooted between two well-supported clades or after the Arabidopsis outgroup. The heatmap aligned with the tree was generated using the "ggtree" package (https://www.bioconductor.org/packages/release/bioc/html/ggtree.html).

Accession Numbers

The raw sequencing data, the processed FPKM values of gene expression and the result of the differential expression analysis were all deposited at the Gene Expression Omnibus under the accession number GSE85516 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85516).

Results

Arabidopsis Root Hair Genes

To compare root hair gene activity across plant species, we first defined the genes expressed in differentiating root-hair cells of Arabidopsis. A transgenic line containing the green fluorescent protein (GFP) and β-glucuronidase (GUS) reporters under the control of the *COBRA-LIKE9* promoter (*AtCOBL9::GFP* (Brady et al., 2007b)) was used for this purpose, because it specifically accumulates GFP in root-hair cells beginning in the elongation zone (EZ; prior to hair emergence) through the maturation zone (MZ; hair maturation) (Figure 3.1A). The GFP-expressing cells were isolated by protoplasting and fluorescence-activated cell sorting (FACS) of the *AtCOBL9::GFP* root tips, and their transcripts were purified and subjected to RNA-Seq analysis (using 3 biological replicates; see Materials and Methods for details). Transcripts from 12,691 genes were identified from among the total of 33,550 Arabidopsis genes (TAIR10) surveyed (designated 12691 *AtRH* (*Arabidopsis thaliana* root hair) genes; mean fragments per kilobase per million mapped reads (FPKM) ≥ 3; had expression for at least two out of three biological replicates). As validation, we found that the *AtRH* genes covered all 17 of the individual genes previously reported to be root-hair-specific using non-transcriptome methods,

and it possesses 90-97% overlap with four previously reported root hair gene datasets (505-1814 genes/dataset) defined by transcriptome-based methods (Becker et al., 2014; Brady et al., 2007a; Lan et al., 2013; Li et al., 2015).

It is likely that many *AtRH* genes are associated with functions common to most/all cells (i.e. "housekeeping genes"). To identify the subset of *AtRH* genes closely associated with root hair cell differentiation, we assessed their transcript levels in the hairless *rhd6* mutant, relative to wild-type (WT). We also included in these lines a WER::GFP marker, which accumulates GFP in the entire developing root epidermis (Lee and Schiefelbein, 1999), to focus on transcript changes in the epidermal cells. Following protoplasting/FACS and RNA-Seq analysis, we compared transcript accumulation in *rhd6 WER::GFP* versus WT *WER::GFP* roots (3 biological replicates per line) and identified 563 *AtRH* genes that are significantly downregulated in *rhd6* mutant (fold-change (FC) ≥2, false discovery rate (FDR) ≤ 0.01, and at least 3 out of 6 samples counts per million ≥1; Figure 3.2A). Because their expression is *RHD6*-dependent and associated with root hair formation, these 563 *AtRH* genes are designated as *AtRHM* (*Arabidopsis thaliana* root hair morphogenesis) genes. As validation, we found that all six of the root hair genes previously shown to be positively regulated by RHD6 based on non-transcriptome methods are included here, and 122 of the 126 genes (97%) previously reported to be downregulated in *rhd6* in a microarray study (Bruex et al., 2012) are also present in the *AtRHM* gene set. Further, as expected, gene ontology (GO) analysis of the *AtRHM* genes showed significant overrepresentation (FDR<0.01) of root hair-associated categories, including "root hair cell differentiation", "unidimensional cell growth", and "trichoblast differentiation" (data not shown).

Given that they are positively regulated by RHD6, the *AtRHM* genes might exhibit preferential expression in root-hair cells, as compared to the remainder of the *AtRH* genes

(designated *AtRH(-RHM)* genes). To test this, we calculated the ratio of transcript accumulation in the root-hair cells (FPKM from *AtCOBL9::GFP*) to transcript accumulation in the entire root elongation zone and differentiation zone (FPKM from previously published root segments (Huang and Schiefelbein, 2015)) for each of these genes. The distribution of these values differ significantly between the *AtRHM* and the *AtRH(-RHM)* gene groups ($p<10^{-15}$, Wilcoxon rank-sum test), indicating that, as a whole, the AtRHM genes possess a relatively greater degree of preferential root hair expression than the other *AtRH* genes (Figure 3.2B).

We also analyzed the temporal expression profiles of the *AtRHM* genes, relative to the *AtRH(-RHM)* genes. Because they are associated with root hair formation, *AtRHM* genes might be expected to exhibit relatively high transcriptional activity in the differentiation zone of the root, where root hairs emerge and grow (Grierson et al., 2014). To examine this, we compared each gene's transcript accumulation in the three major longitudinal root zones (meristematic zone (MZ), elongation zone (EZ), and differentiation zone (DZ)), using transcriptome data previously reported from these Arabidopsis root segments (Huang and Schiefelbein, 2015). We found that a majority of the *AtRHM* genes (79%), but not the *AtRH(-RHM)* genes (30%), exhibit temporal expression profiles associated with relatively high transcript accumulation in the differentiation zone (expression profile types of 6-9), a statistically significant enrichment ($p< 0.01$, χ^2 test, Bonferroni corrected; Figure 3.2C)

Next, we sought to determine whether the *AtRHM* genes tend to be related to one another in sequence. To assess this, we analyzed the distribution of the *AtRHM* genes among Arabidopsis gene families. Using an established plant gene family database (GreenPhyl v4 (Rouard et al., 2011)), we found that 543 of the 563 *AtRHM* genes have been assigned to a total of 397 GreenPhyl-defined families. This number of families is substantially less than the numbers

obtained from 1000 random draws of 543 genes from the 12449 *AtRH* gene set that are included in the GreenPhyl database (Figure 3.2D), indicating that *AtRHM* genes tend to be related to one another and therefore cluster in families. Consistent with this, we observed several families that contain high proportions of *AtRHM* genes, including 6 two-gene families in which both members are *AtRHM* genes and, in the most extreme case, an eleven-member family composed entirely of *AtRHM* genes (data not shown). This suggests conservation in *RHD6*-regulated root hair gene expression in certain gene lineages.

Rice Root Hair Genes

To determine whether the root hair development genes identified in Arabidopsis is similar in other plants, we defined the root hair transcriptome of rice. RNA was extracted from protoplasting/FACS isolated cells from a rice transgenic line, *OsEXPA30::GFP* (Kim et al., 2006) that specifically accumulates GFP in root-hair cells, in a manner similar to the GFP accumulation in the *AtCOBL9::GFP* line (Figure 3.1B). Following RNA-Seq analysis (3 biological replicates), we identified a total number of 13,342 genes that were expressed in the rice *OsEXPA30::GFP* sorted cells (designated 13342 *OsRH* genes (*Oryza sativa* root hair genes); same expression cutoff used for the *AtCOBL9::GFP* analysis). As validation, we found all six of the previously reported root-hair-specific genes defined by non-transcriptome methods in rice to be included in the *OsRH* genes (data not shown).

We tested whether the root hair genes in Arabidopsis are related to the root hair genes in rice by analyzing the distribution of the *AtRH* and non-*AtRH* genes relative to the *OsRH* and non-*OsRH* genes within GreenPhyl-defined gene families. Among families that possess at least one Arabidopsis gene and at least one rice gene, we discovered a statistically significant non-random

distribution (controlling for family size), indicating preferential association of *AtRH* genes with *OsRH* genes and *non-AtRH* genes with *non-OsRH* genes in these families ($p < 0.05$, Bonferroni corrected; Table 3.1). This familial association indicates that the root hair expressed genes tend to be conserved in these two plant species.

Next, we compared the relative level of root hair expression from Arabidopsis and rice genes present in the same family. To avoid complications associated with differences in gene number per family between these species, we calculated the total root hair transcript accumulation for all Arabidopsis genes (by summing FPKM values from the *AtCOBL9::GFP* dataset) and for all rice genes (by summing FPKM values from the *OsEXPA30::GFP* dataset) from each individual family (data not shown). These aggregate genes are referred to as "supergenes". A comparison of the transcript level for Arabidopsis and rice supergenes from the same families reveals a strong positive correlation (Pearson's Correlation Coefficient $r = 0.72$; Figure 3.3), indicating similar total root hair expression for Arabidopsis and rice genes in the same family.

We also analyzed the possibility that gene expression profile of gene expression is conserved for those *AtRH* and *OsRH* genes present in the same families. Using Fisher's Exact Test, we discovered a significant preferential familial association of expression profile types between *AtRH* and *OsRH* genes ($p<0.01$, Bonferroni corrected). Thus, in addition to possessing sequence similarity, Arabidopsis and rice root hair genes from the same families also tend to exhibit similar transcript levels and development pattern of gene expression.

In our next serious of experiments, we compared the rice root hair (*OsRH*) genes to the subset of Arabidopsis root hair genes associated with root hair morphogenesis (i.e. *AtRHM* genes). As above, we first assessed the association between the *AtRHM* genes and *OsRH* genes within gene families. Surprisingly, unlike the strong familial association of *AtRH* genes with

OsRH genes (and *non-AtRH* genes with *non-OsRH* genes), the *AtRHM* genes do not strongly associate with *OsRH* genes within families ($p<0.05$, Bonferroni corrected; Table 3.1). This suggests that, as a group, the *AtRHM* genes exhibited less similarity to rice genes than do the other *AtRH* genes.

To extend this analysis, we constructed phylogenetic trees for each of the *AtRHM* GreenPhyl families and identified well-supported subfamilies within these that possess at least one Arabidopsis and one rice gene (See Materials and Methods). Consistent with our overall family-level results, we discovered that subfamilies containing *AtRH(-RHM)* genes preferentially included *OsRH* genes rather than non-*OsRH* genes ($p<0.01$; Fisher's Exact Test), but subfamilies containing *AtRHM* genes di not exhibit a statistically significant preference for *OsRH* genes over non-*OsRH* genes ($p=0.45$, Fisher's Exact Test).

These results suggest greater diversification of the genes in the *AtRHM* families, relative to the *AtRH(-RHM)* families, between Arabidopsis and rice. If so, we might expect that a greater fraction of the *AtRHM* families would lack rice gene members entirely, relative to the *AtRH* families. Indeed, controlling for family size (1-3 Arabidopsis genes/family), 25.4% of the *AtRHM*-containing GreenPhyl gene families lack a rice gene, whereas only 12.0% of the *AtRH(-RHM)*-containing gene families lack a rice gene.

We also compared the expression level of the *AtRHM* and *AtRH(-RHM)* supergenes versus *OsRH* supergenes from the same family, to determine whether there is diversification in gene expression level in these families (Figure 3.3). We discovered a significant difference ($p<0.01$, t-test, Bonferroni corrected) between mean transcript levels between *AtRHM* and *OsRH* supergenes from common families, but not between *AtRH(-RHM)* and their related *OsRH* supergenes (data not shown). Further, as shown in the Figure 3.3, the adjusted R^2 for the *AtRHM*

vs. *OsRH* is smaller than for *AtRH* vs. *OsRH* (0.39 vs. 0.52), showing that *AtRHM* vs. *OsRH* exhibits more variation (greater "scatter" in the plot) that cannot be explained by the regression model. These results indicate less conservation of gene expression between Arabidopsis and rice genes in the *AtRHM* families, as compared to *AtRH(-RHM)* families.

We also analyzed the degree of similarity in gene expression profiles for families containing *AtRHM* and *OsRH* genes. In contrast to the results from this test using the entire set of *AtRH* genes, we did not find a significant association of Arabidopsis and rice genes possessing the same expression profile within these *AtRHM* families (data not shown).

Together, these findings indicate that *AtRHM* genes are less conserved in rice, as compared to *AtRH(-RHM)* genes, suggesting substantial divergence in the root hair developmental program used by Arabidopsis and rice.

Root Hair Gene Relatives in Other Plants Species

To determine whether rice is unique among vascular plants in its dissimilarity to the *AtRHM* genes, we analyzed related root hair genes in four additional angiosperm species (cucumber, soybean, tomato, and maize) and in a lycophyte species (Selaginella) (Figure 3.4A). First, we analyzed the composition of *AtRHM* and *AtRH(-RHM)* gene families to determine whether these additional species possess related genes. Consistent with our results with rice, we found a greater fraction of the *AtRHM* families lack genes from these species (approximately 2-fold difference for each species), as compared to the *AtRH(-RHM)* families (data not shown). Thus, preferential divergence of *AtRHM*-related genes does not appear to be unique to rice.

Next, we analyzed the overall degree of conservation of *AtRHM*-related genes and gene

expression in these seven species. For each of the 397 GreenPhyl-defined *AtRHM* gene family, we assigned each species a similarity score based on whether the species possesses a related gene in that family and the degree to which its family member matches the expression profile of the *AtRHM* gene (See Material and Methods). The comparative analysis of these similarity scores yielded a species-wise hierarchical clustering with a tree topology that mirrored the evolutionary relationships between the species (Figure 3.5), indicating that changes in the gene family structure and expression are positively correlated with the divergence time from common ancestors. The family-wise groupings, generated by hard cutoffs of the similarity scores, produced distinct clusters of gene families with common across-species *AtRHM* relationships (Figure 3.5; see Materials and Methods).

The largest cluster, designated "vascular plant conserved", includes 266 *AtRHM* families that possess a root-expressed gene from each of the plant species tested, indicating that these are the most ancient families and likely contain genes with common root hair functions shared by all vascular plants (Figure 3.5). This cluster includes many of the well-characterized Arabidopsis root hair genes (e.g. *EXPA7*, *IRT2*, *AHA7*, *RHD2*, *LRX1*, *COW1*, *MRH1*, *MRH6*, *IRE*, *PIP5K3*), and includes a disproportionate share (93%) of the *AtRHM* genes encoding secretory pathway activities. It is noteworthy that the degree of conservation of the *AtRHM* root developmental expression profile varies among these families (Figure 3.5), suggesting that the regulation or developmental role of these genes has diverged in some of the families.

A second cluster of gene families, "angiosperm conserved", possesses root-expressed *AtRHM*-related genes from all six angiosperms, but Selaginella either lacks a related gene or lacks root expression of its gene (Figure 3.5), suggesting that these root hair gene functions arose after the lycophyte-euphyllophyte split or they have been lost during Selaginella evolution.

These *AtRHM* genes encode a relatively high proportion (40%) of putative regulatory proteins (e.g. AP2-, GATA-, and WRKY-related transcription factors and various protein kinases), which may have evolved to provide angiosperms new mechanisms to control root hair growth.

A cluster designated "eudicot conserved" includes 13 families of *AtRHM*-related genes that possess root-expressed members exclusively from the four eudicot species tested. Another cluster, "Arabidopsis specific", includes 34 families that do not possess a root-expressed *AtRHM*-related gene from any of the other six species tested. These two clusters are dominated (10/13 and 22/34) by genes encoding unknown/uncharacterized proteins, which contribute to novel species- or lineage-specific root hair features. The Arabidopsis specific cluster also contains six families encoding cell wall-related proteins, including an arabinogalactan protein (AGP3) and several proline-rich family proteins.

A final cluster of gene families, designated "Other", contains unusual distributions of *AtRHM*-related genes among the species, consistent with relatively rare lineage-specific gene loss/gain (Figure 3.5). For instance, the family containing the Arabidopsis *FERRIC REDUCTION OXIDASE4* (*FRO4*) and *FRO5* genes include root-expressed genes from all vascular plant species tested except rice and maize. This implies loss of this root-hair-related gene activity during monocot evolution, perhaps associated with distinct strategies used by grasses for iron acquisition (Jain et al., 2014).

We also analyzed these GreenPhyl-defined *AtRHM* families for the presence of related genes from the moss Physcomitrella. Interestingly, although moss lacks roots and root hairs, we found that most of the *AtRHM* gene families (277/397) contain a Physcomitrella gene (Figure 3.5), implying that these root hair developmental genes evolved from genes possessing a related function in a root-hairless ancestor of vascular plants. Conversely, we found that 20 of the

AtRHM families lack a Physcomitrella gene member, yet possess a root-expressed gene from Selaginella and at least one angiosperm, which defines families likely to have arisen during vascular plant evolution coincident with the evolution of root hairs (data not shown).

To more rigorously analyze gene families containing *AtRHM* genes with a demonstrated role in root hair development, we performed maximum likelihood (ML) phylogenetic analysis based on the protein sequences from relatives of the 19 *AtRHM* genes that, when mutated, exhibit an abnormal root hair phenotype (data not shown). For this analysis, we included related genes from the seven vascular plant species and their associated root developmental zone transcript data, as well as related genes from Physcomitrella and Norway spruce (*Picia abies*). A total of 15 trees were generated from these 19 *AtRHM* genes (Figure 3.6), and overall, the gene relationships largely mirrored the results obtained from the similarity score clustering analysis described above. For example, in the *COBL9* family (one of the "vascular plant conserved" families from Figure 3.5), each species possesses a gene with strong similarity in sequence and transcript accumulation to the Arabidopsis *COBL9* (Figure 3.6A). In a few trees, we observed variation in structure or expression of *AtRHM*-related genes in certain species. The *EXPA7* family contains a well-supported clade including root-expressed *EXPA7/EXPA18*-related genes from all vascular plant species, but not from Physcomitrella (Figure 3.6B), which suggests that this expansin subgroup may have evolved (in part) for use in root hair development. Consistent with this, the two rice genes from this clade (*Os10g39110* and *Os06g01920*) have previously been shown to participate in root hair differentiation (Kim et al., 2006; Yu et al., 2011). In another case, the *ROOT HAIR SPECIFIC8* (*RHS8*) gene is part of a well-supported clade that lacks rice and maize members but contains root-expressed genes from all other species tested (Figure 3.6C), which suggests a monocot-lineage-specific loss of this *AtRHM* gene/function. Altogether, these results

indicate that the *AtRHM*-related gene divergence uncovered in the similarity score matrix analysis (Figure 3.5) likely represents an underestimate of the actual variation in *AtRHM*-related gene function across vascular plants. More generally, these results demonstrate the utility of a combined phylogenetic and transcriptomic approach, enabling a high resolution view of the likely evolutionary and functional relationships between genes in large families.

We also generated a ML tree for the *RHD6*-related genes from these species. We find that *RHD6* is included in a well-supported clade that contains the partially functionally redundant Arabidopsis *RSL1* gene (Menand et al., 2007), as well as root-expressed genes from each of the other species examined (Figure 3.6D), consistent with a previous study showing broad conservation of *RHD/RSL* gene sequence (Pires et al., 2013). It is notable that each of these species possesses an *RHD6*-related gene with transcript accumulation in the meristematic region of the root, similar to the Arabidopsis *RHD6* (Figure 3.6D), implying that each of these species might use an *RHD6* homolog to regulate early root-hair cell differentiation.

Root Hair Patterning Gene Relatives

In addition to root hair morphogenesis genes, we also sought to determine whether root hair patterning genes are conserved across vascular plant species. Arabidopsis is unique among the plants analyzed in this study because its root hair pattern is position dependent, with root-hair cells limited to longitudinal cell files in particular locations (Type 3), whereas the other six species produce root-hair and non-hair epidermal cells in a random distribution (Type 1; Figure 3.4B) (Balcerowicz et al., 2015; Clowes, 2000; Pemberton, 2001; Salazar-Henao et al., 2016). We generated ML trees and analyzed root gene expression for 12 Arabidopsis patterning genes (present in seven gene families). In Arabidopsis, each of these genes acts early in root epidermis

development (beginning in the meristematic zone) and ultimately regulates *RHD6* transcription to specify the root-hair cell pattern (Balcerowicz et al., 2015; Grierson et al., 2014; Salazar-Henao et al., 2016).

The *CPC/TRY/ETC1* patterning genes encode small one-repeat MYB proteins (Kirik et al., 2004; Wada et al., 1997), and we found they are all present in a clade that includes genes from all euphyllophytes, but only cucumber and soybean genes share similar consistent meristem zone transcript accumulation (Figure 3.7A). The *GL2* gene encodes a HD-Zip transcription factor that promotes the non-hair fate (Masucci et al., 1996), and it occupies a well-supported clade containing root-expressed genes from cucumber and soybean only (Figure 3.7B). The *GL3/EGL3/MYC1* genes encode partially redundant bHLH proteins (Bernhardt et al., 2003; Bruex et al., 2012), and our ML tree shows they reside in a subgroup that contains meristem-zone-expressed genes from eudicots only (Figure 3.7C). Similarly, we found conservation of gene structure and root expression in eudicots only for our ML tree containing the *TTG2* gene (data not shown), which encodes a WRKY transcription factor(Johnson et al., 2002). These four trees are similar in showing conservation among (some) eudicots only, suggesting functional divergence for these patterning genes during eudicot evolution or possibly loss of the gene/function in the monocot lineage.

Two other patterning genes exhibit potential conservation. The *TTG1* gene, encoding a WD protein required to repress root hair specification (Galway et al., 1994), is in a clade with similar root-expressed genes from all vascular plants tested (data not shown). The *SCM* (aka *SUB*) gene encodes a receptor-like kinase that influences the positional expression of the other patterning genes (Kwak and Schiefelbein, 2007; Kwak et al., 2005), and we found root-expressed *SCM*-related genes in each of the vascular plant species tested (data not shown). However, *SCM*'s

preferential meristematic transcript accumulation is only shared by *SCM*-related genes from eudicots.

The R2R3 MYB transcription factors *WEREWOLF* (*WER*) and *MYB23* are partially redundant early acting patterning genes that negative transcriptionally regulate root hair genes (Lee and Schiefelbein, 1999; Kang et al., 2009). Our ML analysis places these two MYBs in a clade (previously defined as MYB subgroup 15 (Stracke et al., 2001)) that also includes GL1 (Oppenheimer et al., 1991), but does not include related genes from any of the other plant species tested.

These results indicate substantial divergence in the structure and expression of the Arabidopsis patterning genes in these vascular plant species, suggesting they are not generally used for specifying root-hair cells in all vascular plants. Altogether, our analysis of the *AtRHM* genes and the patterning genes provides a broad outline of the evolution of genes controlling root hair development in vascular plants (Figure 3.8).

Discussion

This large-scale study combined phylogenetic and transcriptome analyses to define and compare root hair genes from seven diverse vascular plant species, including eudicots, monocots, and a lycophyte. A major finding was that most root hair development genes are similar in structure and expression in all species tested, suggesting that a core program for root hair development is conserved across the vascular plants. Further, we found that nearly all of these vascular plant conserved genes (251/266) possess close relatives in the rootless moss Physcomitrella, implying that the core root hair program did not evolve *de novo* in the vascular

plant lineage but likely was coopted from a preexisting program in a land plant ancestor. An attractive possibility is that this ancient program was responsible for unidirectional cellular growth (tip growth) of exploratory or invasive cell types in the ancestral species and was recruited for tip-growing root-hair cells during vascular plant evolution. Related to this, we found that a disproportionate share (93%) of the *AtRHM* genes encoding predicted secretory pathway proteins (likely involved in tip growth) are among this vascular plant conserved group. Further, cellular and physiological mechanisms employed by tip-growing cells are similar across different groups of organisms, including fungi, bryophytes, and vascular plants, consistent with the possibility of an evolutionarily ancient underlying program (Geitmann and Emons, 2000; Jones and Dolan, 2012; Rounds and Bezanilla, 2013; Nezhad and Geitmann, 2013).

In addition to identifying conserved root hair genes, we also discovered significant diversification in the genetic program associated with root hair development among the vascular plants. We initially discovered this by comparing root-hair-expressed genes from Arabidopsis and rice. Specifically, we found that the Arabidopsis root hair morphogenesis (*AtRHM*) genes exhibit significantly greater divergence in their structure and expression from their rice relatives (within the same gene families), as compared to *non-AtRHM* root-hair-expressed genes. This was unexpected because we had previously found that root-expressed genes are generally conserved between Arabidopsis and rice (Huang and Schiefelbein, 2015). The underlying reason for preferential divergence of the root hair development genes is unclear. It may be that, as a single cell type, the root hair may be relatively less constrained in its developmental options, due to minimal coordination with neighboring cells. Another possibility is that, as a cell that extends from the plant body into the rhizosphere, the root hair may evolve and utilize multiple developmental strategies to effectively interact with and adapt to a varying environment. In

support of this, root hair growth in many species is known to be strongly influenced by nutrient availability (Nestler et al., 2016; Salazar-Henao et al., 2016; Perry et al., 2007).

Overall, approximately one-third of the Arabidopsis root hair development genes differ substantially in structure or expression in one or more of the other six vascular plant species tested. Considering that the conserved root hair genes may define a core root hair growth program (as discussed above), then these diverged genes may be responsible for regulating and/or modifying this core program in ways appropriate for particular species or lineages. The proportion of diverged genes within species largely followed phylogenetic lines, with Selaginella exhibiting the greatest differences in gene structure and expression (Figure 3.8). It is notable that, among the genes not shared with Selaginella, those encoding putative regulatory proteins were highly represented, suggesting that new mechanisms of root hair developmental control evolved following the divergence of lycophytes and euphyllophytes. Among the genes not shared with any other tested species (the Arabidopsis-specific root hair genes), those encoding proteins with unknown or uncharacterized functions were overrepresented, which may prove fruitful for further study to understand evolution of novel cell-type developmental activities or characteristics.

The analysis of Arabidopsis genes controlling root hair patterning was of particular interest in this study, because Arabidopsis differs from the other analyzed species by producing a particular pattern of root-hair cells (dependent on cell position; Type 3), rather than a random distribution of root-hair cells (Type 1) in the root epidermis (Clowes, 2000; Pemberton, 2001). Consistent with this, we detected greater divergence in gene structure and expression within the seven families of Arabidopsis genes involved in patterning, as compared to families containing *AtRHM* genes. In particular, five of these seven families possess a clade that includes the

Arabidopsis patterning gene(s) but lacks a related root-expressed gene from one or more of the other angiosperm species. These results strongly suggest a linkage between the structure/expression of these patterning genes and evolution of the Type 3 root hair pattern in Arabidopsis. Further, this implies that the Type 1 root hair distribution mechanisms relies on other, as yet unknown, cell fate regulators. In this respect, it is notable that all of these species possess and express an *RHD6*-related bHLH gene similar to the Arabidopsis *RHD6* (Figure 3.6D and Figure 3.8). Indeed, it has been previously shown that *RHD6* homologs are widespread and function similarly in divergent species (Menand et al., 2007; Pires et al., 2013), suggesting that *RHD6* acts as the critical regulator of root hair initiation in all vascular plants. Given that the Arabidopsis root hair patterning genes specify cell fate via transcriptional regulation of *RHD6* (Balcerowicz et al., 2015; Grierson et al., 2014), Type 1 species may similarly achieve their roo-hair cell distribution by regulating their *RHD6* homologs, but employing a different mechanism(s) to do so.

Among the seven families containing Arabidopsis patterning genes, the WER/MYB23 family was unique in possessing its patterning genes in an Arabidopsis-specific subgroup (previously defined as MYB subgroup 15 (Stracke et al., 2001)). Thus, it is tempting to speculate that evolution of the WER/MYB23 genes was the critical factor in the origin of the Arabidopsis Type 3 pattern. However, a recent extensive analysis of MYB genes in multiple species showed that subgroup 15 includes members from several Type 1 eudicots (Du et al., 2015), complicating the potential linkage between this subgroup and the Type 3 pattern. Interestingly, the patterning of epidermal hairs (trichomes) on the leaf surface of Arabidopsis also relies on a member of this MYB subgroup 15, the *GLABROUS1* (*GL1*) gene (Larkin et al., 1993), implying shared evolution of these patterning mechanisms. This study provides a foundation for further analyses

of evolutionary events responsible for the origin of these cell type patterns.

Table 3.1 Results of Fisher's Exact Test analyzing the familial association between Arabidopsis and rice root hair gene sets. Test was performed in 9 groups when the family size was fixed (a combination of 1-3 At genes and 1-3 Os genes per family). The design table for the test was listed below. *AtRH* families = families with at least one *AtRH* gene, *Non-AtRH* families = families without any *AtRH* gene, *OsRH* families = families with at least one *OsRH* gene, *Non-OsRH* families = families without any *OsRH* gene, *AtRHM* families = families with at least one *AtRHM* gene, *AtRH(-RHM)* families = families do not contain any *AtRHM* gene and contain at least one *AtRH* gene. The P values were corrected by Bonferroni method to control for multiple testing error rate.

Fisher's Exact Test on familial association between the *AtRH* genes and the *OsRH* genes.

Test design table	AtRH families	Non-AtRH families
OsRH families		
Non-OsRH families		

	Os gene(s) per family		
At gene(s) per family	1	2	3
1	1.6699E-152	5.66047E-34	1.125E-09
2	1.85754E-25	6.73908E-17	6.9186E-10
3	7.94086E-06	1.15231E-05	0.006749827

Fisher's Exact Test on familiar association between the *AtRH(-RHM)* genes and the *OsRH* genes.

Test design table	AtRH(-RHM) families	Non-AtRH families
OsRH families		
Non-OsRH families		

	Os gene(s) per family		
At gene(s) per family	1	2	3
1	3.2434E-154	1.68647E-34	1.1001E-09
2	3.9281E-25	6.8512E-17	1.37369E-09
3	3.58292E-06	1.45424E-05	0.006405968

Fisher's Exact Test on familiar association between the *AtRHM* genes and the *OsRH* genes.

Test design table	AtRHM families	Non-AtRH families
OsRH families		
Non-OsRH families		

	Os gene(s) per family		
At gene(s) per family	1	2	3
1	1	1	1
2	0.000578923	0.239459721	0.016056802
3	1	0.072914349	0.905572755

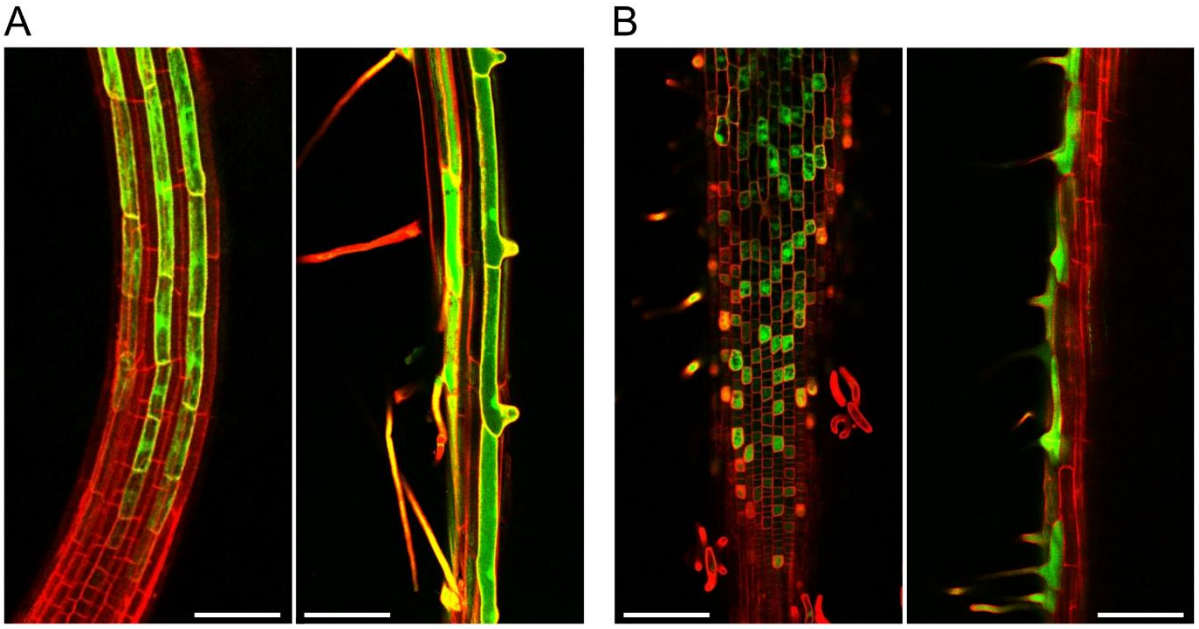

Figure 3.1 Root-hair cell specific expression of GFP marker lines in Arabidopsis and rice roots.

A. GFP accumulation in the *AtCOBL9::GFP* line in immature root-hair cells in the elongation zone (left) and in the differentiation zone (right) of Arabidopsis roots.

B. GFP accumulation in the *OsEXPA30::GFP* line in immature root-hair cells in the elongation zone (left) and in the differentiation zone (right) of rice roots. Roots were stained with propidium iodide (red fluorescence).

Scale bars = 100 μm.

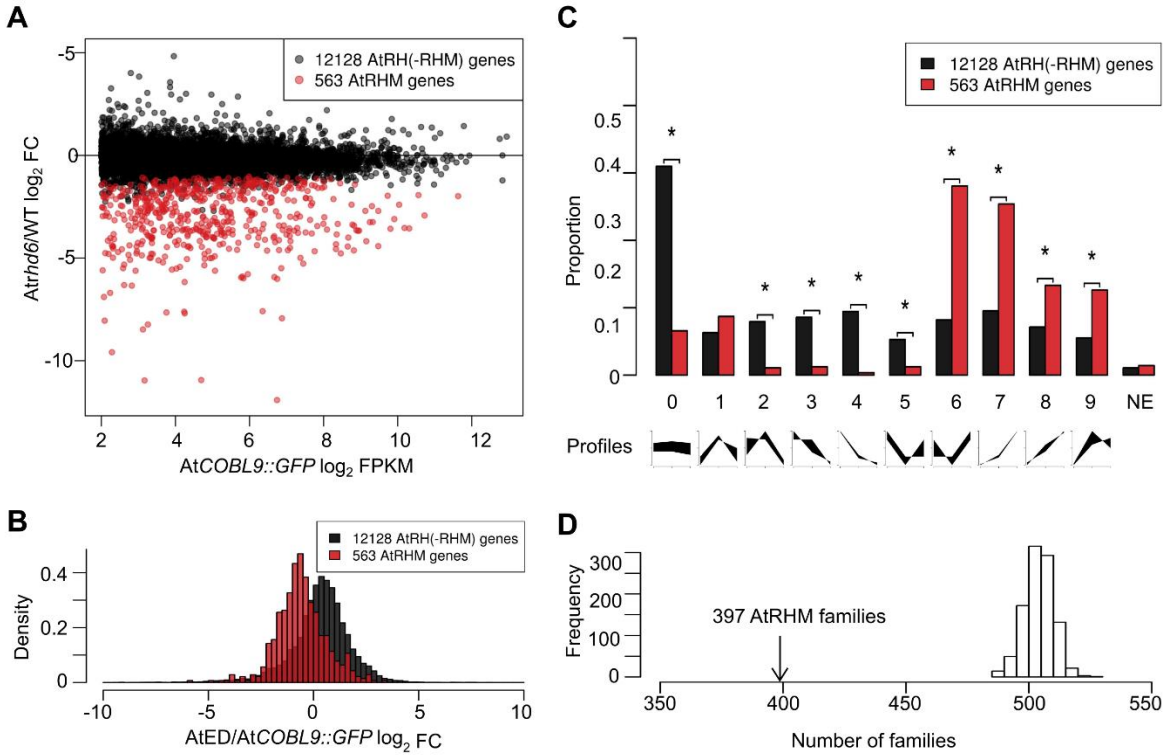

Figure 3.2 Analysis of *AtRH* and *AtRHM* genes.

A. Distribution of *AtRH(-RHM)* genes and *AtRHM* genes, based on transcript level in the FACS-purified root-hair cells of *AtCOBL9::GFP* and $\log_2$ fold-change in transcript level from FACS-purified cells of *rhd6 WER::GFP* versus *WT WER::GFP*. Data is the mean from three biological replicates.

B. Distribution of *AtRH(-RHM)* genes and *AtRHM* genes, based on $\log_2$ fold-change in transcript level from FACS-purified root-hair cells of *AtCOBL9::GFP* versus wild-type root elongation zone and differentiation zone segments. Data is the mean from three biological replicates.

C. Distribution of *AtRH(-RHM)* genes and *AtRHM* genes, based on relative transcript level in the meristematic, elongation, and differentiation zones of wild-type roots (i.e. expression profiles). The nine expression profile types (defined in (Huang and Schiefelbein 2015)) are indicated in the graphs below the figure (left to right; meristematic, elongation, differentiation zones). NE=no expression detected. Asterisks indicate profile types with significantly different proportions between the two groups ($p<0.01$, chi-square test, Bonferroni corrected).

D. Distribution of the number of gene families resulting from 1000 random draws of 543 genes from the 12449 *AtRH* genes in the GreenPhyl family database. The observed number of gene families (397) that contain the 543 *AtRHM* genes are indicated.

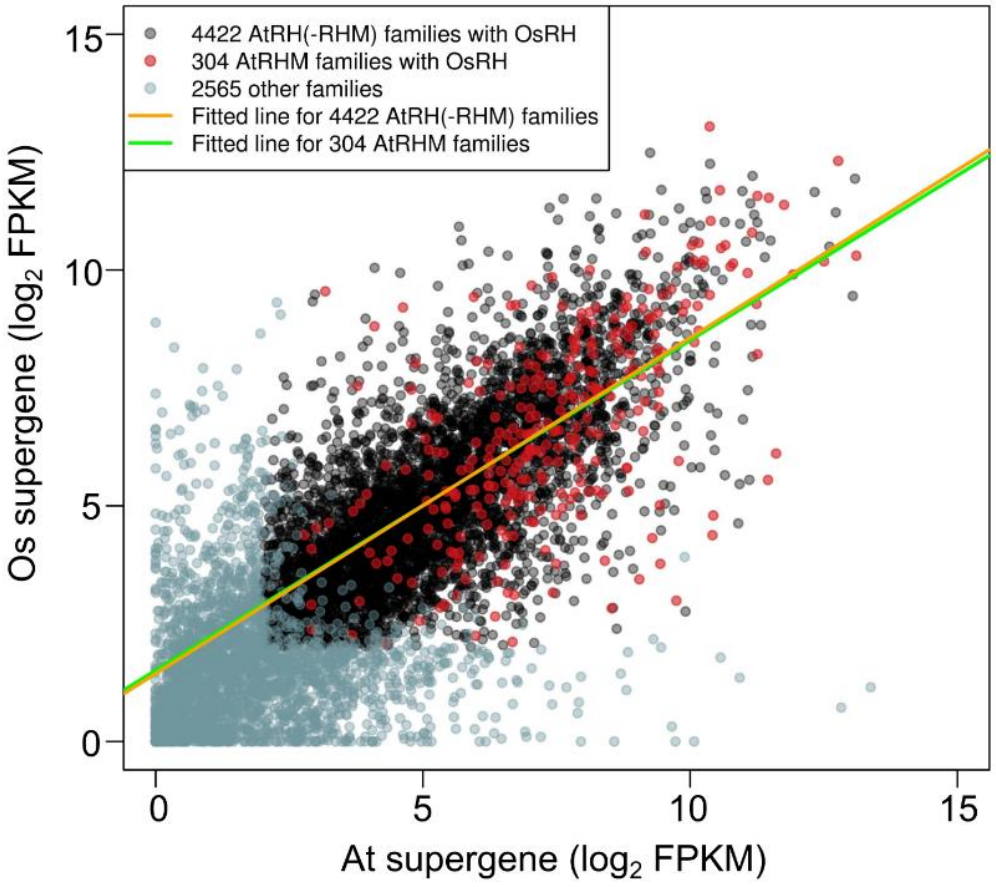

Figure 3.3 Comparison of Arabidopsis and rice supergene expression from common families.

Distribution of GreenPhyl-defined gene families, based on combined transcript level (FPKM; log_2 scaled) for all Arabidopsis genes (from FACS purified *AtCOBL9::GFP*) and for all rice genes (from FACS purified *OsEXPA30::GFP*) from each of the 7291 families that possess at least one Arabidopsis and one rice gene. Pearson's correlation coefficient r = 0.77 for the total 7291 families. Least-square fitted lines were generated for the 304 *AtRHM* families containing ≥1 *OsRH* gene (red dots; gene line; adjusted R^2=0.39) and the 4422 *AtRH(-RHM)* families containing ≥1 *OsRH* gene (black dots; orange line; adjusted R^2=0.52).

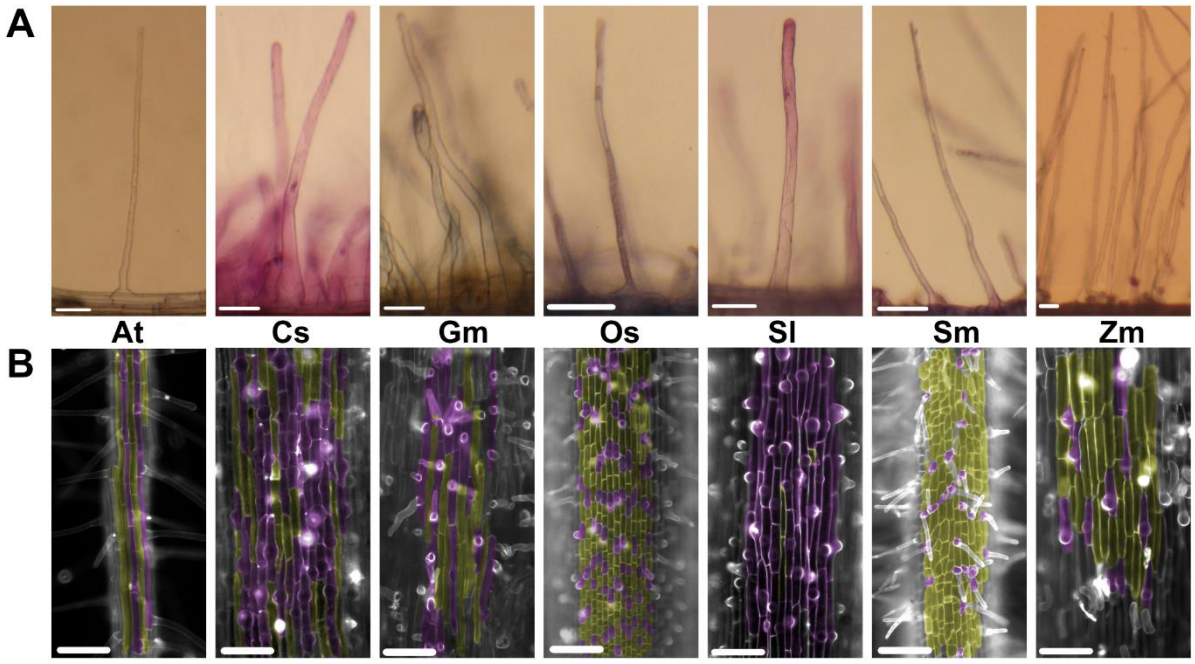

Figure 3.4 Root hairs in diverse vascular plants.

A. Photos of individual root hairs from Arabidopsis (At), cucumber (Cs), soybean (Gm), rice (Os), tomato (Sl), Selaginella (Sm) and maize (Zm). Scale bar: 50 μm.

B. Root epidermis from Arabidopsis, rice, cucumber, soybean, tomato, maize, and Selaginella roots, stained with fluorescent dye (Fluorescent Brightener 28 or propidium iodide) and false colored to indicate the root-hair cells (purple) and the non-hair cells (yellow). Only the Arabidopsis root possesses the longitudinal file-specific (Type 3) pattern of root-hair cells. the pattern of root-hair cells vs. non-hair cells in the epidermis is different in these species (Arabidopsis=file-specific pattern (Type 3); rice=random distribution (Type 1)). Scale bar: 100 μm.

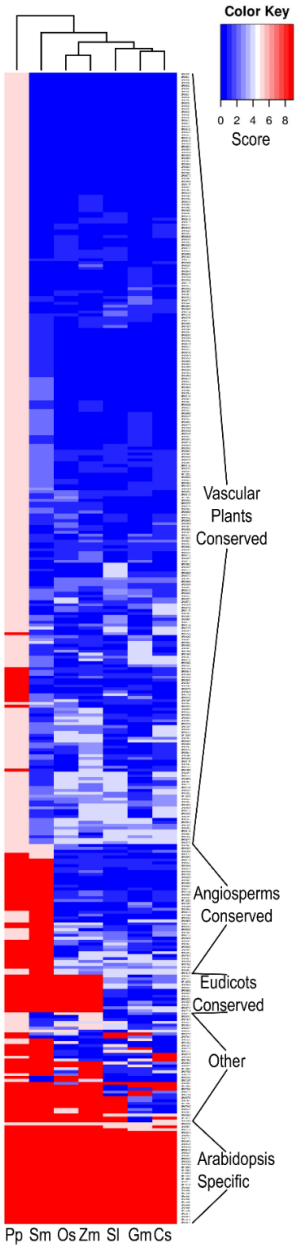

Figure 3.5 Conservation of Arabidopsis root hair morphogenesis genes in other plants.

A differential matrix heat map was generated for the *AtRHM*-containing GreenPhyl gene families. Each species (from left to right: Physcomitrella, Selaginella, rice, maize, tomato, soybean, cucumber) was scored for its degree of conservation (blue=highest; red=lowest) based on presence/absence of a gene and its expression profile, relative to the *AtRHM* gene in each family. Major categories of *AtRHM* genes are indicated.

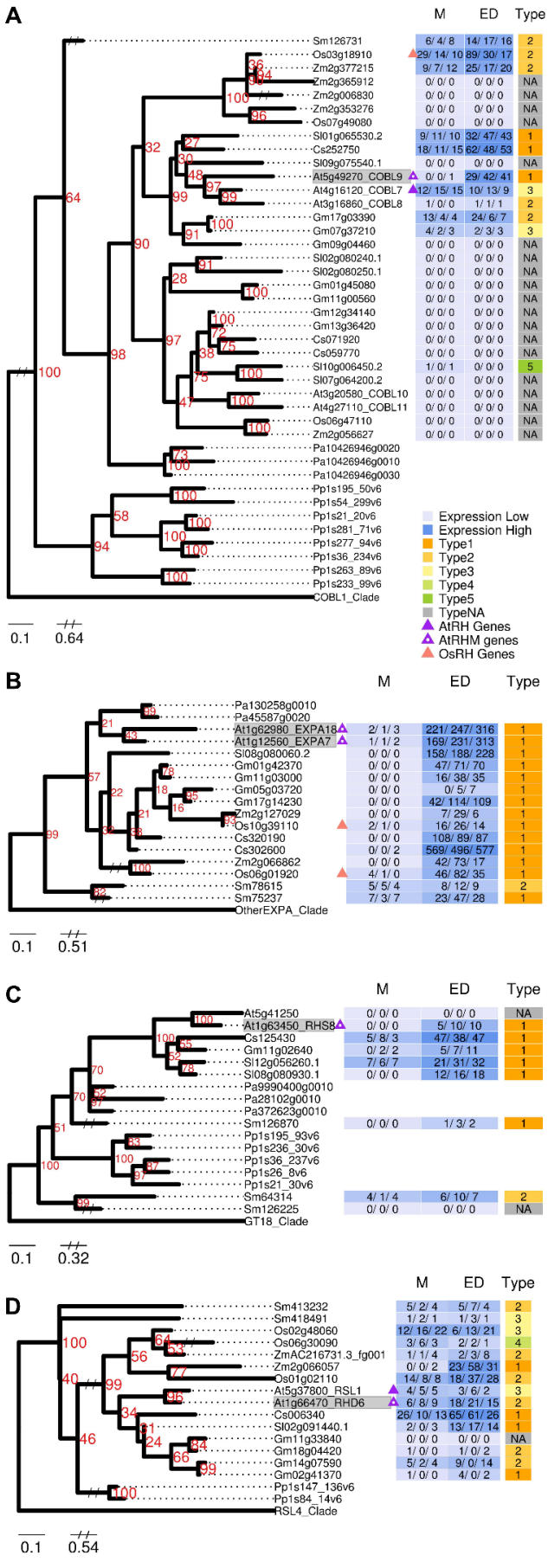
A
M
ED
Type
Sm126731
Os03g18910
Zm2g377215
Zm2g365912
Zm2g006830
Zm2g353276
Os07g49080
Sl01g065530.2
Cs252750
Sl09g075540.1
At5g49270_COBL9
At4g16120_COBL7
At3g16860_COBL8
Gm17g03390
Gm07g37210
Gm09g04460
Sl02g080240.1
Sl02g080250.1
Gm01g45080
Gm11g00560
Gm12g34140
Gm13g36420
Cs071920
Cs059770
Sl10g006450.2
Sl07g064200.2
At3g20580_COBL10
At4g27110_COBL11
Os06g47110
Zm2g056627
Pa10426946g0020
Pa10426946g0010
Pa10426946g0030
Pp1s195_50v6
Pp1s54_299v6
Pp1s21_20v6
Pp1s281_71v6
Pp1s277_94v6
Pp1s36_234v6
Pp1s263_89v6
Pp1s233_99v6
COBL1_Clade
Expression Low
Expression High
Type1
Type2
Type3
Type4
Type5
TypeNA
AtRH Genes
AtRHM genes
OsRH Genes
0.1
0.64
B
M
ED
Type
Pa130258g0010
Pa45587g0020
At1g62980_EXPA18
At1g12560_EXPA7
Sl08g080060.2
Gm01g42370
Gm11g03000
Gm05g03720
Gm17g14230
Zm2g127029
Os10g39110
Cs320190
Cs302600
Zm2g066862
Os06g01920
Sm78615
Sm75237
OtherEXPA_Clade
0.1
0.51
C
M
ED
Type
At5g41250
At1g63450_RHS8
Cs125430
Gm11g02640
Sl12g056260.1
Sl08g080930.1
Pa9990400g0010
Pa28102g0010
Pa372623g0010
Sm126870
Pp1s195_93v6
Pp1s236_30v6
Pp1s36_237v6
Pp1s26_8v6
Pp1s21_30v6
Sm64314
Sm126225
GT18_Clade
0.1
0.32
D
M
ED
Type
Sm413232
Sm418491
Os02g48060
Os06g30090
ZmAC216731.3_fg001
Zm2g066057
Os01g02110
At5g37800_RSL1
At1g66470_RHD6
Cs006340
Sl02g091440.1
Gm11g33840
Gm18g04420
Gm14g07590
Gm02g41370
Pp1s147_136v6
Pp1s84_14v6
RSL4_Clade
0.1
0.54

Figure 3.6 Representative maximum-likelihood phylogenetic trees of *AtRHM* gene families.

A. *COBL9*.

B. *EXPA7*.

C. *RHS8*.

D. *RHD6*.

For each tree, the defining Arabidopsis *AtRHM* genes associated with a mutant phenotype are shaded in gray. Gene expression FPKM values are shown for each replicate and converted to a heatmap with high expression in darker blue and low expression in light blue. Expression profile types generated from the fold-change between two developmental zones are shown in different colors as indicated in legend. Triangles indicate *AtRHM* genes (purple with white dot), *AtRH* genes (solid purple), and *OsRH* genes (pink). Numbers in red indicate support for 1000 bootstrap. Gene IDs are abbreviated (Arabidopsis: At; cucumber: Cs; soybean: Gm; rice: Os; tomato: Sl; Selaginella: Sm; maize: Zm; Physcomitrella: Pp; Norway spruce: Pa).

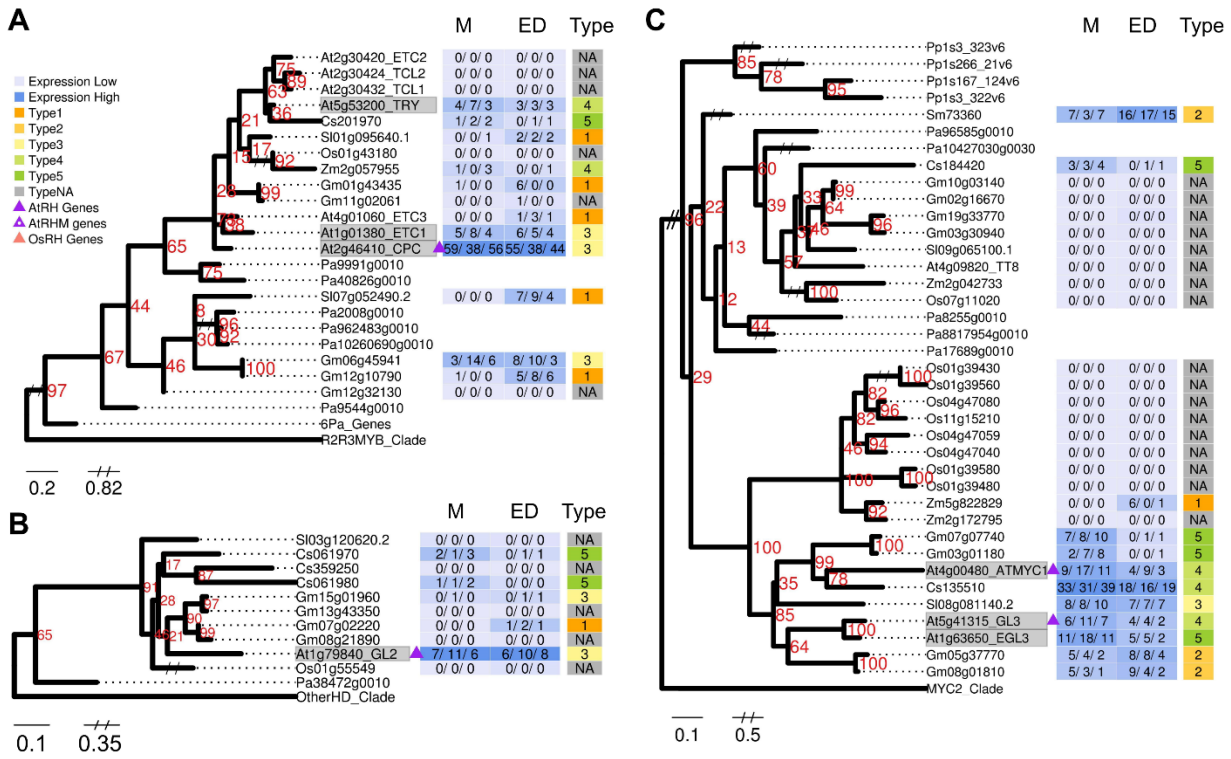

Figure 3.7 Representative maximum-likelihood phylogenetic trees of Arabidopsis root-hair patterning gene families.

A. *CPC*/TRY/ETC1.

B. *GL2*.

C. *GL3/EGL3/MYC1*.

For each tree, the defining Arabidopsis root-hair patterning genes *AtRHM* genes associated with a mutant phenotype are shaded in gray. Gene expression FPKM values are shown for each replicate and converted to a heatmap with high expression in darker blue and low expression in light blue. Expression profile types generated from the fold-change between two developmental zones are shown in different colors as indicated in legend. Triangles indicate *AtRHM* genes (purple with white dot), *AtRH* genes (solid purple), and *OsRH* genes (pink). Numbers in red support for 1000 bootstrap. Gene IDs are abbreviated (Arabidopsis: At; cucumber: Cs; soybean: Gm; rice: Os; tomato: Sl; Selaginella: Sm; maize: Zm; Physcomitrella: Pp; Norway spruce: Pa).

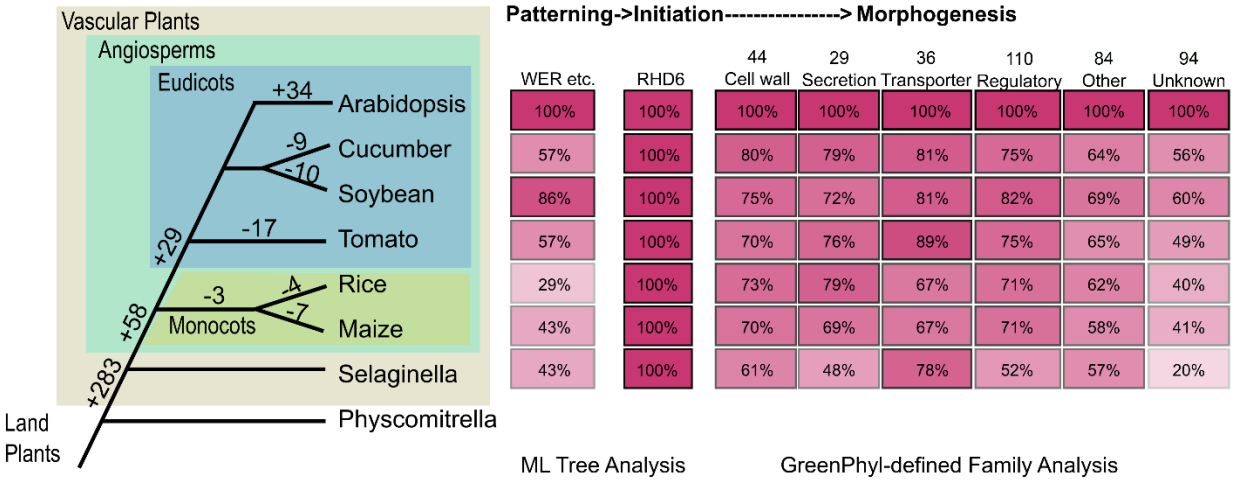

Figure 3.8 Summary of the evolution and relationships between Arabidopsis root hair gene families.

The tree indicates the possible origin of 404 gene families (397 *AtRHM* + 7 patterning gene families), by displaying the number of GreenPhyl-defined gene families with at least one root-expressed member in each species at each point in the phylogeny. Within the tree, vascular plants are shaded in gray, angiosperms in cyan, monocots in yellow, and eudicots in blue. Arabidopsis morphogenesis gene families (397 *AtRHM* gene families) were subdivided into cell wall (44 families), secretion (29 families), regulatory (110 families), transporter (36 families), other (84) and unknown (94 families) categories, based on their predicted protein functions. The values in the boxes represent the fraction of related genes with similar expression profiles (score = 0 or 1) in each of the non-Arabidopsis species using ML tree groupings (for the patterning genes and *RHD6* families) or GreenPhyl families (for the *AtRHM* families).

CHAPTER 4

Conclusions and Future Directions

Conclusions

In this thesis project, I generated a high-resolution, large-scale transcriptional landscape of genes that are expressed in various root development stages (meristematic, elongation, and differentiation) and different cell types (root-hair cells vs. total cells from root) across vascular plants (Arabidopsis, cucumber, soybean, rice, tomato, Selaginella, and maize). The information obtained from these studies was utilized to compare gene expression profiles during root and root hair development to understand the degrees to which these developmental programs have either remained conserved or have diversified in different plant lineages.

The goal of the research presented in the first chapter was to study root evolution by comparative analysis of gene expression programs in Selaginella and other vascular plants. The root is one of the greatest adaptations shared by almost all land plants, except the early land plants (bryophytes). In contrast, bryophytes possess single-celled structures, called rhizoids, what are believed to help with anchorage and nutrient uptake. The root, as a multi-cellular organ, evolved in the vascular plant lineage that diverged after bryophytes. Due to this universal existence throughout the plant kingdom and its relatively simple structures, the root is a good model to understand the molecular basis for multi-cellular organ formation and adaptation in evolutionary developmental biology. Previous studies concerning root evolution have posited

that the root may have evolved at least twice; once in the lycophytes (Selaginella) and once in the euphyllophytes, due to the lack of fossil support and the differences in root morphologies (Kenrick and Crane, 1997; Raven and Edwards, 2001; Friedman et al., 2004; Kenrick and Strullu-Derrien, 2014). In this study, I compared the root transcriptome of Selaginella to other vascular plants and found significant conservation in the gene family structures and gene expression profiles during root development in the two lineages. In addition, I analyzed 133 key genes known to be critical for root development in Arabidopsis and found that they are used in all vascular plants, including Selaginella, and they show conserved expression profiles during root development. Particularly surprising, was the observation that root cap genes display conserved gene expression in Selaginella. The root cap is a structure unique to the root not found in any other plant organs and so it is less likely to be shared by roots of different evolutionary origin. Overall, the results show significant conservation of gene expression program in the lycophyte Selaginella, despite its possible independent root evolutionary origin. Thus, the basic molecular mechanisms for root development appear to be conserved in all vascular plants tested. There are two possible interpretations for the evolution of plant roots based on these findings: 1) roots from different origins may have recruited similar developmental programs from a limited number of functional gene pools; 2) a single root developmental program may have existed in the last common ancestor of Selaginella and other vascular plants.

A second research project, presented in Chapter 2, examines the conservation/diversification of the root hair development programs in vascular plants. Root hairs are tubular structures protruding from the root epidermis that help with various root functions. Similar to roots, root hairs are also found in almost all plant lineages, thereby allowing for large-scale comparative studies across plant species. In addition, root hairs are single-celled structures that can serve as a

good model to study cellular differentiation during organ formation. Previous studies suggest that root hairs are homologous structures to the rhizoids in bryophytes and they use the same regulators (e.g. *ROOT HAIR DEFECTIVE 6*, *RHD6*) for cell elongation (Menand et al., 2007; Jones and Dolan, 2012; Proust et al., 2016). However, plants differ in their root hair distributions in the epidermis and may also have variable root hair sizes (Clowes, 2000; Pemberton, 2001; Dittmer, 1949; Gahoonia et al., 1997), indicating lineage specific differences exist in root hair development programs. In this study, I examined the root hair transcriptome in Arabidopsis and rice and find broad conservation in the genes that are expressed in the root-hair cells of these two species. By identifying a subset of 563 root-hair expressed genes that are regulated by RHD6 (Arabidopsis root hair morphogenesis genes), I further analyzed the gene expression in root hair development in Arabidopsis and rice. I found that the Arabidopsis root hair morphogenesis genes show less conservation in the gene family structure and gene expression to their relatives in rice, compared to the total root-hair expressed genes. Other vascular plants tested exhibit a similar trend in the divergence of family structure and gene expression as observed in rice. The root hair patterning genes, which work upstream of RHD6 to regulate root hair pattern formation, also show substantial differences in the gene family structure and gene expression profiles compared to other vascular plants tested. Taken together, the results suggest Arabidopsis lineage specific diversification of root hair patterning genes and RHD6-regulated root hair morphogenesis genes.

Future Directions

One unexpected result of our experiment is the lack of detectable expression of an *RHD6*-related gene in the rice *OsEXPA30::GFP*-expressing cells. If RHD6 homologs regulate root hair

elongation in all vascular plants, then we would have expected to observe its transcript accumulation in this rice line. One explanation for this is that the *RHD6*-related genes in rice may be inhibited prior to the expression of the *COBL9*-related gene(s). Another possibility is that the *RHD6*-related genes in rice are not related to, or involved in, root hair development. Though both two explanations support the divergence of the *RHD6* expression in rice, they have distinct biological significances about the function of RHD6 in root hair development when considering the conservation of RHD6 in early land plants and Arabidopsis (Menand et al., 2007; Proust et al., 2016). In order to distinguish between these two possibilities, it may be necessary to utilize *in situ* hybridization to examine the mRNA accumulation patterns of the *RHD6*-related genes in rice during root epidermal development. In addition, we could analyze the phenotype in root hair development when the RHD6-related genes are mutated. If the *RHD6*-related genes are expressed prior to root hair initiation in rice and the mutant possesses defect in root hair initiation, it would be consistent with the conclusion that RHD6 has conserved function in rice. In this case, additional studies can focus on the mechanism that inhibits the expression of *RHD6* in the root-hair cells. If the *RHD6*-related genes are not expressed in the root epidermis and there is no root hair development phenotype associated with these genes, then it would suggest that *RHD6*-related genes are not regulating the root hair development in rice as they do in Arabidopsis and the early land plants. In this case, many interesting questions can be asked. For example, which gene(s) is responsible for regulating root hair morphogenesis in rice? Is this divergence rice-specific or common to all monocots? If the *RHD6*-regulated genes are not expressed in the root epidermis yet exhibit a mutant phenotype in root hair development, it is possible that the *RHD6*-regulated genes may function in an underlying tissue, and in this case, new studies can focus on the mechanisms how the signal cascades from inner cell layers to the

epidermis.

The molecular basis of root hair pattern formation is much less studied compared to the root hair morphogenesis in plant species other than Arabidopsis. This work presents the first large-scale study about the Arabidopsis root hair patterning genes in other plants to understand the evolution of root hair pattern formation. I have shown that the presence and the expression of the Arabidopsis root hair patterning genes is unique to Arabidopsis and it is positively correlated with the Type 3 root hair pattern in Arabidopsis, in contrast to the Type 1 root hair pattern in other vascular plants. In order to further verify the relationship between the Arabidopsis root hair patterning genes and the Type 3 root hair pattern, we might need to examine the presence and expression of these patterning genes in the close relatives of Arabidopsis (in the Brassicales order) that have been reported to have, or to not have, a Type 3 root hair pattern. *Carica papaya* is the closest plant species to Arabidopsis in Brassicales with a non-Type 3 root hair pattern, so it can be used as a negative control for our analysis. If the presence and expression of the Arabidopsis root hair patterning genes is found to be conserved in all Type 3 plants but not in papaya, it is likely that all other Type 3 plants in Brassicales use largely the same molecular program for its root hair pattern formation as Arabidopsis. In this case, the split of papaya from the common ancestor of the Brassicales is a starting point to track the divergence of root hair pattern formation in Brassicales. For example, we could analyze more plant species that evolved between papaya and the Type 3 Brassicales to estimate the time of this divergence in the Brassicales evolutionary history.

The Type 3 root hair pattern has been found in multiple lineages across the eudicots (Figure 1.4) (Clowes, 2000; Pemberton, 2001). However, it is not clear whether the Type 3 pattern evolved multiple times in the evolution of the eudicots or the Type 3 pattern existed in their

common ancestors. To address this, we might need to check the presence and expression profiles of the Arabidopsis root hair patterning genes in the Type 3 plants in other lineages to determine whether conservation could be found. Consequently, we may generate a new hypothesis about the evolutionary origin of the Type 3 root hair pattern formation in eudicots. For example, if the Arabidopsis root hair patternings genes are repeatedly used by all Type 3 plants with conserved expression profiles, it is less likely for this pattern to have independent evolutionary origins, as it requires the collaboration of multiple transcription factors to be expressed in the same time and location (and their downstream targets conserved as well). Instead, it might be better explained by parallel recruitment of the same set of genes from the existing "toolkit" (e.g. the WD40-bHLH-MYB complex that is used in pigmentation biosynthesis) or by strong positive selection from the common ancestors of the eudicots.

Appendices

Appendix A Abbreviations

ACCW	anticlinal cortical cell wall
bHLH	basic helix-loop-helix
bp	base pair
CPC	CAPRICE
DNA	deoxyribonucleic acid
DE	differential expression / differentially expressed
EGL3	ENHANCER OF GLABRA3
EM	Expectation Maximization
ETC1	ENHANCER OF TRY AND CPC1
EXPA7	EXPANSIN A7
FACS	Fluorescence-activated cell sorting
FPKM	fragments per kilo-base per million mapped reads
d	day
GL2	GLABRA2
GL3	GLABRA3
GFP	green fluorescent protein
H-cell	hair cell
H-file	hair cell file
kb	kilobase

LRL	RHL1-LIKE
mb	mega base
min	minute
ML	maximum likelihood
nm	nanometer
N-cell	non-hair cell
N-file	non-hair cell file
NGS	next generation sequencing
RHD6	ROOT HAIR DEFECTIVE6
RHL1	ROOT HAIR LESS1
RNA	ribonucleic acid
RNA-Seq	ribonucleic acid sequencing
RSL1	RHD6-LIKE1
SCM	SCRAMBLED
TCL1	TRICHOMELESS1
TCL2	TRICHOMELESS2
TF	transcription factor
TRY	TRIPTYCHON
TTG1	TRANSPARENT TESTA GLABRA1
TTG2	TRANSPARENT TESTA GLABRA2
WER	WEREWOLF
WT	wild-type

Appendix B Supplemental Materials of Chapter 2

Supplemental Data

The following Supplemental Data is available from the Plant Cell website by the link:

http://www.plantcell.org/content/suppl/2015/08/10/tpc.15.00328.DC1/TPC2015-00328-LSBR3_tpc00238_Supplemental.pdf

Supplemental Figure 1 Phylogenetic trees for six angiosperm species

Supplemental Figure 2 Phylogenetic trees for five plant species

Supplemental Table 1 Comparison of mRNA accumulation of published *in situ* hybridization and RNA-Seq result.

Supplemental Table 2 Gene Ontology enrichment terms for supergene families with a specific expression profile

Supplemental Table 3 List of Arabidopsis gene families with genes exhibiting a root developmental mutant phenotype

Supplemental Methods 1 Custom script for clustering related sequences

Supplemental Data Sets

The following Supplemental Data Sets are available through the link: http://dx.doi.org/10.5061/dryad.68686.

Supplemental Data Set 1 Processed RNA-Seq gene expression values

Supplemental Data Set 2 GreenPhyl-defined gene families

Supplemental Data Set 3 Distribution of root-expressed genes by family

Supplemental Data Set 4 Arabidopsis genes preferentially expressed in root development zones

Supplemental Data Set 5 Clustering of root-expressed genes

Supplemental Data Set 6 Supergene expression in three developmental zones

Supplemental Data Set 7 Supergene connectivity values

Supplemental Data Set 8 Processed RNA-Seq gene expression values

Supplemental Data Set 9 Classification of genes by expression profile

Supplemental Data Set 10 Supergene expression and profile types

Supplemental Data Set 11 Alignment of gene sequences from six angiosperm species

Supplemental Data Set 12 Alignment of gene sequences from five plant species

References

Abas, L., Benjamins, R., Malenica, N., Paciorek, T., Wiśniewska, J., Moulinier-Anzola, J.C., Sieberer, T., Friml, J., and Luschnig, C. (2006). Intracellular trafficking and proteolysis of the Arabidopsis auxin-efflux facilitator PIN2 are involved in root gravitropism. Nat. Cell Biol. **8**: 249–256.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics **31**: 166–169.

Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. Nat. Rev. Genet. **9**: 868–882.

Article, L.B., Simon, M., Bruex, A., Kainkaryam, R.M., Zheng, X., Huang, L., Woolf, P.J., and Schiefelbein, J. (2013). Tissue-Speci fi c Pro fi ling Reveals Transcriptome Alterations in Arabidopsis Mutants Lacking Morphological Phenotypes. **25**: 1–12.

Balcerowicz, D., Schoenaers, S., and Vissenberg, K. (2015). Cell Fate Determination and the Switch from Diffuse Growth to Planar Polarity in Arabidopsis Root Epidermal Cells. Front. Plant Sci. **6**: 1–13.

Banks, J.A. (2009). Selaginella and 400 million years of separation. Annu. Rev. Plant Biol. **60**: 223–238.

Banks, J.A. et al. (2011). The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science **332**: 960–963.

Barlow, P.W. (2002). The root cap: Cell dynamics, cell differentiation and cap function. J. Plant Growth Regul. **21**: 261–286.

Becker, J.D., Takeda, S., Borges, F., Dolan, L., and Feijó, J. a (2014). Transcriptional Profiling of Arabidopsis root hairs and pollen defines an apical cell growth signature. BMC Plant Biol. **14**: 197.

Benfey, P.N. (1999). Is the shoot a root with a view? Curr. Opin. Plant Biol. **2**: 39–43.

Benfey, P.N., Linstead, P.J., Roberts, K., Schiefelbein, J.W., Hauser, M.-T., and Aeschbacher, R.A. (1993). Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. Development **119**: 57–70.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B **57**: 289 – 300.

Bennett, T. and Scheres, B. (2010). Root development-two meristems for the price of one?

Curr. Top. Dev. Biol. **91**: 67–102.

Bennett, T., van den Toorn, A., Sanchez-Perez, G.F., Campilho, A., Willemsen, V., Snel, B., and Scheres, B. (2010). SOMBRERO, BEARSKIN1, and BEARSKIN2 regulate root cap maturation in Arabidopsis. Plant Cell **22**: 640–654.

Berger, F., Haseloff, J., Schiefelbein, J., and Dolan, L. (1998a). Positional information in root epidermis is defined during embryogenesis and acts in domains with strict boundaries. Curr. Biol. **8**: 421–430.

Berger, F., Hung, C.Y., Dolan, L., and Schiefelbein, J. (1998b). Control of cell division in the root epidermis of Arabidopsis thaliana. Dev Biol **194**: 235–245.

Bernardes, J.S., Vieira, F.R., Costa, L.M., and Zaverucha, G. (2015). Evaluation and improvements of clustering algorithms for detecting remote homologous protein families. BMC Bioinformatics **16**: 34.

Bernhardt, C., Lee, M.M., Gonzalez, A., Zhang, F., Lloyd, A., and Schiefelbein, J. (2003). The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. Development **130**: 6431–6439.

Bernhardt, C., Zhao, M., Gonzalez, A., Lloyd, A., and Schiefelbein, J. (2005). The bHLH genes GL3 and EGL3 participate in an intercellular regulatory circuit that controls cell patterning in the Arabidopsis root epidermis. Development **132**: 291–298.

Birnbaum, K., Jung, J.W., Wang, J.Y., Lambert, G.M., Hirst, J. a, Galbraith, D.W., and Benfey, P.N. (2005). Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. Nat. Methods **2**: 615–619.

Blancaflor, E.B., Fasano, J.M., and Gilroy, S. (1998). Mapping the functional roles of cap cells in the response of Arabidopsis primary roots to gravity. Plant Physiol **116**: 213–222.

Brady, S.M., Orlando, D. a, Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007a). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science **318**: 801–806.

Brady, S.M., Song, S., Dhugga, K.S., Rafalski, J.A., and Benfey, P.N. (2007b). Combining Expression and Comparative Evolutionary Analysis. The *COBRA* Gene Family. Plant Physiol. **143**: 172–187.

Bremer, K., Humphries, C.J., Mishler, B.D., and Churchill, S.P. (1987). On Cladistic Relationships in Green Plants. Taxon **36**: 339–349.

Bruex, A., Kainkaryam, R.M., Wieckowski, Y., Kang, Y.H., Bernhardt, C., Xia, Y., Zheng, X., Wang, J.Y., Lee, M.M., Benfey, P., Woolf, P.J., and Schiefelbein, J. (2012). A gene regulatory network for root epidermis cell differentiation in Arabidopsis. PLoS Genet. **8**: 1–59.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics **10**: 421.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**: 1972–1973.

Carey, C.C., Strahle, J.T., Selinger, D. a, and Chandler, V.L. (2004). Mutations in the pale aleurone color1 regulatory gene of the Zea mays anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in Arabidopsis thaliana. Plant Cell **16**: 450–464.

Cho, H.-T. and Cosgrove, D.J. (2002). Regulation of root hair initiation and expansin gene expression in Arabidopsis. Plant Cell **14**: 3237–3253.

Clowes, F. a L. (2000). Pattern in root meristem development in angiosperms. New Phytol. **146**: 83–94.

Conte, M.G., Gaillard, S., Droc, G., and Perin, C. (2008a). Phylogenomics of plant genomes: a methodology for genome-wide searches for orthologs in plants. BMC Genomics **9**: 183.

Conte, M.G., Gaillard, S., Lanau, N., Rouard, M., and Périn, C. (2008b). GreenPhylDB: A database for plant comparative genomics. Nucleic Acids Res. **36**: 991–998.

Cormack, R.G.H. (1947). A comparative study of developing epidermal cells in white mustard and tomato roots. Am. J. Bot. **34**: 310–314.

Cormack, R.G.H. (1935). Investigations on the development of root hairs. New Phytol **34**: 30–54.

Costa, S. and Shaw, P. (2006). Chromatin organization and cell fate switch respond to positional information in Arabidopsis. Nature **439**: 493–6.

Cox, C.J., Li, B., Foster, P.G., Embley, T.M., and Civan, P. (2014). Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. **63**: 272–279.

Cutter, E.G. and Hung, C.Y. (1972). Symmetric and asymmetric mitosis and cytokinesis in the root tip of Hydrocharis morsus-ranae L. J. Cell Sci. **11**: 723–737.

Cvrčková, F., Bezvoda, R., and Zárský, V. (2010). Computational identification of root hair-specific genes in Arabidopsis. Plant Signal. Behav. **5**: 1407–1418.

Datta, S., Kim, C.M., Pernas, M., Pires, N.D., Proust, H., Tam, T., Vijayakumar, P., and Dolan, L. (2011). Root hairs: Development, growth and evolution at the plant-soil interface. Plant Soil **346**: 1–14.

Ding, W., Yu, Z., Tong, Y., Huang, W., Chen, H., and Wu, P. (2009). A transcription factor with a bHLH domain regulates root hair development in rice. Cell Res. **19**: 1309–1311.

Dittmer, H. (1949). Root hair variations in plant species. Am. J. Bot. **36**: 152–155.

Dolan, L. (1996). Pattern in the root epidermis: an interplay of diffusible signals and cellular geometry. Ann. Bot. **77**: 547–553.

Dolan, L., Duckett, C.M., Grierson, C., Linstead, P., Schneider, K., Lawson, E., Dean, C.,

Poethig, S., and Roberts, K. (1994). Clonal relationships and cell patterning in the root epidermis of Arabidopsis. Development **2474**: 2465–2474.

Doyle, J.A. (2013). Phylogenetic Analyses and Morphological Innovations in Land Plants.

Du, H., Liang, Z., Zhao, S., Nan, M.-G., Phan Tran, L.-S., Lu, K., Huang, Y.-B., and Li, J.-N. (2015). The Evolutionary History of R2R3-MYB Proteins Across 50 Eukaryotes: New Insights Into Subfamily Classification and Expansion. Sci. Rep. **5**: 11037.

Duff, R.J. and Nickrent, D.L. (1999). Phylogenetic Relationships of Land Plants Using Mitochondrial Small-Subunit RDNA Sequences. Am. J. Bot. **86**: 372–386.

Dunn, O.J. (1961). Multiple Comparisons Among Means. J. Am. Stat. Assoc. **56**: 52–64.

Esau, K. (1965). Plant Anatomy (New York, John Wiley & Sons).

Feller, A., MacHemer, K., Braun, E.L., and Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J. **66**: 94–116.

Friedman, W.E., Moore, R.C., and Purugganan, M.D. (2004). The evolution of plant development. Am. J. Bot. **91**: 1726–1741.

Gahoonia, T.S., Care, D., and Nielsen, N.E. (1997). Root hairs and phosphorus acquisition of wheat and barley cultivars. Plant Soil **191**: 181–188.

Galway, M.E., Masucci, J.D., Lloyd, a M., Walbot, V., Davis, R.W., and Schiefelbein, J.W. (1994). The TTG gene is required to specify epidermal cell fate and cell patterning in the Arabidopsis root. Dev. Biol. **166**: 740–754.

Geitmann, A. and Emons, A.M.C. (2000). The cytoskeleton in plant and fungal cell tip growth. J. Microsc. **198**: 218–45.

Gensel, P.G. (2008). The Earliest Land Plants. Annu. Rev. Ecol. Evol. Syst. **39**: 459–477.

Gensel, P.G. and Berry, C.M. (2001). Early lycophyte evolution. **91**: 74–98.

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. **27**: 221–4.

Grebe, M. (2012). The patterning of epidermal hairs in Arabidopsis-updated. Curr. Opin. Plant Biol. **15**: 31–37.

Grierson, C., Nielsen, E., Ketelaarc, T., and Schiefelbein, J. (2014). Root Hairs. Arab. B. **12**: e0172.

Gu, F. and Nielsen, E. (2013). Targeting and regulation of cell wall synthesis during tip growth in plants. J. Integr. Plant Biol. **55**: 835–846.

Harrison, C.J., Corley, S.B., Moylan, E.C., Alexander, D.L., Scotland, R.W., and Langdale, J. a (2005). Independent recruitment of a conserved developmental mechanism during leaf evolution. Nature **434**: 509–514.

Hawes, M.C., Curlango-Rivera, G., Xiong, Z., and Kessler, J.O. (2012). Roles of root border

cells in plant defense and regulation of rhizosphere microbial populations by extracellular DNA “trapping.” Plant Soil **355**: 1–16.

Hawes, M.C., Gunawardena, U., Miyasaka, S., and Zhao, X. (2000). The role of root border cells in plant defense. Trends Plant Sci. **5**: 128–133.

Hinchliff, C.E. et al. (2015). Synthesis of phylogeny and taxonomy into a comprehensive tree of life. Proc. Natl. Acad. Sci. **112**: 201423041.

Hochholdinger, F. and Tuberosa, R. (2009). Genetic and genomic dissection of maize root development and architecture. Curr. Opin. Plant Biol. **12**: 172–177.

Huang, D.W., Sherman, B.T., and Lempicki, R. a (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. **4**: 44–57.

Huang, L. and Schiefelbein, J. (2015). Conserved Gene Expression Programs in Developing Roots from Diverse Plants. Plant Cell **27**: tpc.15.00328.

Huerta-Cepas, J., Dopazo, J., and Gabaldón, T. (2010). ETE: a python Environment for Tree Exploration. BMC Bioinformatics **11**: 24.

Hung, C.Y., Lin, Y., Zhang, M., Pollock, S., Marks, M.D., and Schiefelbein, J. (1998). A common position-dependent mechanism controls cell-type patterning and GLABRA2 regulation in the root and hypocotyl epidermis of Arabidopsis. Plant Physiol. **117**: 73–84.

Iijima, M., Higuchi, T., Barlow, P.W., and Bengough, A.G. (2003). Root cap removal increases root penetration resistance in maize (Zea mays L.). J. Exp. Bot. **54**: 2105–2109.

Ishida, T., Kurata, T., Okada, K., and Wada, T. (2008). A genetic regulatory network in the development of trichomes and root hairs. Annu. Rev. Plant Biol. **59**: 365–386.

Jain, A., Wilson, G.T., and Connolly, E.L. (2014). The diverse roles of FRO family metalloreductases in iron and copper homeostasis. Front. Plant Sci. **5**: 100.

Jansen, L., Hollunder, J., Roberts, I., Forestan, C., Fonteyne, P., Van Quickenborne, C., Zhen, R.G., Mckersie, B., Parizot, B., and Beeckman, T. (2013). Comparative transcriptomics as a tool for the identification of root branching genes in maize. Plant Biotechnol. J. **11**: 1092–1102.

Johnson, C.S., Kolevski, B., and Smyth, D.R. (2002). TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. Plant Cell **14**: 1359–1375.

Jones, V. a S. and Dolan, L. (2012). The evolution of root hairs and rhizoids. Ann. Bot. **110**: 205–12.

Kang, Y.H., Kirik, V., Hulskamp, M., Nam, K.H., Hagely, K., Lee, M.M., and Schiefelbein, J. (2009). The MYB23 gene provides a positive feedback loop for cell fate specification in the Arabidopsis root epidermis. Plant Cell **21**: 1080–1094.

Kang, Y.H., Song, S.-K., Schiefelbein, J.W., and Lee, M.M. (2013). Nuclear Trapping Controls the Position-Dependent Localization of CAPRICE in the Root Epidermis of Arabidopsis. Plant Physiol. **163**: 193–204.

Karas, B., Amyot, L., Johansen, C., Sato, S., Tabata, S., Kawaguchi, M., and Szczyglowski, K. (2009). Conservation of lotus and Arabidopsis basic helix-loop-helix proteins reveals new players in root hair development. Plant Physiol. **151**: 1175–1185.

Karas, B., Murray, J., Gorzelak, M., Smith, A., Sato, S., Tabata, S., and Szczyglowski, K. (2005). Invasion of Lotus japonicus root hairless 1 by Mesorhizobium loti involves the nodulation factor-dependent induction of root hairs. Plant Physiol. **137**: 1331–1344.

Karol, K.G. et al. (2010). Complete plastome sequences of Equisetum arvense and Isoetes flaccida: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol. Biol. **10**: 321.

Karve, R. and Iyer-Pascuzzi, A.S. (2015). Digging deeper: High-resolution genome-scale data yields new insights into root biology. Curr. Opin. Plant Biol. **24**: 24–30.

Katoh, K. and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. **30**: 772–780.

Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., and Mclnerney, J.O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6**: 29.

Kenrick, P. and Crane, P.R. (1997). The origin and early evolution of plants on land. Nature **389**: 33–39.

Kenrick, P. and Strullu-Derrien, C. (2014). The Origin and Early Evolution of Roots. Plant Physiol. **166**: 570–580.

Kerstens, S. and Verbelen, J.P. (2003). Cellulose orientation at the surface of the Arabidopsis seedling. Implications for the biomechanics in plant development. J. Struct. Biol. **144**: 262–270.

Kim, C.M. and Dolan, L. (2011). Root hair development involves asymmetric cell division in Brachypodium distachyon and symmetric division in Oryza sativa. New Phytol. **192**: 601–610.

Kim, C.M., Park, S.H., Je, B. Il, Park, S.H., Park, S.J., Piao, H.L., Eun, M.Y., Dolan, L., and Han, C. (2007). OsCSLD1, a cellulose synthase-like D1 gene, is required for root hair morphogenesis in rice. Plant Physiol. **143**: 1220–1230.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. **14**: R36.

Kim, D.W., Lee, S.H., Choi, S.-B., Won, S.-K., Heo, Y.-K., Cho, M., Park, Y.-I., and Cho, H.-T. (2006). Functional conservation of a root hair cell-specific cis-element in angiosperms with different root hair distribution patterns. Plant Cell **18**: 2958–2970.

Kirik, V., Simon, M., Huelskamp, M., and Schiefelbein, J. (2004). The ENHANCER of TRY and CPC1 gene acts redundantly with TRIPTYCHON and CAPRICE in trichome and root hair cell patterning in Arabidopsis. Dev. Biol. **268**: 506–513.

Koenig, D. et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato.

Kranz, H.D., Mikš, D., Siegler, M.L., Capesius, I., Sensen, C.W., and Huss, V.A.R. (1995). The origin of land plants: Phylogenetic relationships among charophytes, bryophytes, and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. J. Mol. Evol. **41**: 74–84.

Kumar, L. and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation **2**: 5–7.

Kwak, S.H. and Schiefelbein, J. (2008). A Feedback Mechanism Controlling SCRAMBLED Receptor Accumulation and Cell-Type Pattern in Arabidopsis. Curr. Biol. **18**: 1949–1954.

Kwak, S.H. and Schiefelbein, J. (2007). The role of the SCRAMBLED receptor-like kinase in patterning the Arabidopsis root epidermis. Dev. Biol. **302**: 118–131.

Kwak, S.-H., Shen, R., and Schiefelbein, J. (2005). Positional signaling mediated by a receptor-like kinase in Arabidopsis. Science **307**: 1111–1113.

Lan, P., Li, W., Lin, W.D., Santi, S., and Schmidt, W. (2013). Mapping gene activity of Arabidopsis root hairs. Genome Biol **14**: R67.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**: 357–359.

Larkin, J., Oppenheimer, D., Pollock, S., and Marks, M. (1993). Arabidopsis GLABROUS1 Gene Requires Downstream Sequences for Function. Plant Cell **5**: 1739–1748.

Le, J., Vandenbussche, F., Van Der Straeten, D., and Verbelen, J.P. (2004). Position and cell type-dependent microtubule reorientation characterizes the early response of the Arabidopsis root epidermis to ethylene. Physiol. Plant. **121**: 513–519.

Lee, M.M. and Schiefelbein, J. (2002). Cell pattern in the Arabidopsis root epidermis determined by lateral inhibition with feedback. Plant Cell **14**: 611–618.

Lee, M.M. and Schiefelbein, J. (1999). WEREWOLF, a MYB-related protein in Arabidopsis, is a position-dependent regulator of epidermal cell patterning. Cell **99**: 473–483.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**: 2078–2079.

Li, W., Lan, P., and 3.948 (2015). Re-analysis of RNA-seq transcriptome data reveals new aspects of gene activity in Arabidopsis root hairs. Front Plant Sci **6**: 421.

Ligrone, R., Duckett, J.G., and Renzaglia, K.S. (2012). Major transitions in the evolution of early land plants: A bryological perspective. Ann. Bot. **109**: 851–871.

Lin, C., Choi, H.S., and Cho, H.T. (2011). Root hair-specific expansin A7 is required for root hair elongation in arabidopsis. Mol. Cells **31**: 393–397.

Lin, Q., Ohashi, Y., Kato, M., Tsuge, T., Gu, H., Qu, L.J., and Aoyama, T. (2015).

GLABRA2 Directly Suppresses Basic Helix-Loop-Helix Transcription Factor Genes with Diverse Functions in Root Hair Development. Plant Cell **27**: 2894–2906.

Mantegazza, O., Gregis, V., Chiara, M., Selva, C., Leo, G., Horner, D.S., and Kater, M.M. (2014). Gene coexpression patterns during early development of the native Arabidopsis reproductive meristem: Novel candidate developmental regulators and patterns of functional redundancy. Plant J. **79**: 861–877.

Massa, G.D. and Gilroy, S. (2003). Touch modulates gravity sensing to regulate the growth of primary roots of Arabidopsis thaliana. Plant J. **33**: 435–445.

Masucci, J.D., Rerie, W.G., Foreman, D.R., Zhang, M., Galway, M.E., Marks, M.D., and Schiefelbein, J.W. (1996). The homeobox gene GLABRA2 is required for position-dependent cell differentiation in the root epidermis of Arabidopsis thaliana. Development **122**: 1253–1260.

Masucci, J.D. and Schiefelbein, J.W. (1994). The rhd6 Mutation of Arabidopsis thaliana Alters Root-Hair Initiation through an Auxin- and Ethylene-Associated Process. Plant Physiol. **106**: 1335–1346.

McCully, M.E. (1999). ROOTS IN SOIL: Unearthing the Complexities of Roots and Their Rhizospheres. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50**: 695–718.

Menand, B., Yi, K., Jouannic, S., Hoffmann, L., Ryan, E., Linstead, P., Schaefer, D.G., and Dolan, L. (2007). An ancient mechanism controls the development of cells with a rooting function in land plants. Science **316**: 1477–1480.

Mishler, B.D., Lewis, L.A., Buchheim, M.A., Renzaglia, K.S., Garbary, D.J., Delwiche, C.F., Zechman, F.W., Kantz, T.S., and Chapman, R.L. (1994). Phylogenetic Relationships of the "Green Algae" and "Bryophytes." Ann. Missouri Bot. Gard. **81**: 451.

Naseer, S., Lee, Y., Lapierre, C., Franke, R., Nawrath, C., and Geldner, N. (2012). Casparian strip diffusion barrier in Arabidopsis is made of a lignin polymer without suberin. Proc. Natl. Acad. Sci. **109**: 10101–10106.

Nestler, J., Keyes, S.D., and Wissuwa, M. (2016). Root hair formation in rice (Oryza sativa L.) differs between root types and is altered in artificial growth conditions. J. Exp. Bot. **67**: erw115.

Nezhad, A.S. and Geitmann, A. (2013). The cellular mechanics of an invasive lifestyle. J. Exp. Bot. **64**: 4709–4728.

Nystedt, B. et al. (2013). The Norway spruce genome sequence and conifer genome evolution. Nature **497**: 579–84.

Oppenheimer, D.G., Herman, P.L., Sivakumaran, S., Esch, J., and Marks, M.D. (1991). A myb gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. Cell **67**: 483–493.

Pang, Y.Z. et al. (2009). A WD40 Repeat Protein from Medicago truncatula Is Necessary for Tissue-Specific Anthocyanin and Proanthocyanidin Biosynthesis But Not for Trichome Development. Plant Physiol. **151**: 1114–1129.

Payne, C.T., Zhang, F., and Lloyd, A.M. (2000). GL3 encodes a bHLH protein that regulates trichome development in arabidopsis through interaction with GL1 and TTG1. Genetics **156**: 1349–1362.

Pemberton, L. (2001). Epidermal Patterning in Seedling Roots of Eudicotyledons. Ann. Bot. **87**: 649–654.

Perry, P., Linke, B., and Schmidt, W. (2007). Reprogramming of root epidermal cells in response to nutrient deficiency. Biochem. Soc. Trans. **35**: 161–3.

Pesch, M. and Hülskamp, M. (2011). Role of TRIPTYCHON in trichome patterning in Arabidopsis. BMC Plant Biol. **11**: 130.

Petricka, J.J., Winter, C.M., and Benfey, P.N. (2012). Control of Plant Root Development. Annu. Rev. Plant Biol.: 563–590.

Pires, N.D., Yi, K., Breuninger, H., Catarino, B., Menand, B., and Dolan, L. (2013). Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. Proc. Natl. Acad. Sci. U. S. A. **110**: 9571–6.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. **26**: 1641–1650.

Proust, H., Honkanen, S., Jones, V.A.S., Morieri, G., Prescott, H., Kelly, S., Ishizaki, K., Kohchi, T., and Dolan, L. (2016). RSL Class i Genes Controlled the Development of Epidermal Structures in the Common Ancestor of Land Plants. Curr. Biol. **26**: 93–99.

Pryer, K.M., Schneider, H., and Magall, S. (2004a). chapter 10. The Radiation of Vascular Plants. Assem. tree life: 138–153.

Pryer, K.M., Schneider, H., Smith, A.R., Cranfill, R., Wolf, P.G., Hunt, J.S., and Sipes, S.D. (2001). Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature **409**: 618–622.

Pryer, K.M., Schuettpelz, E., Wolf, P.G., Schneider, H., Smith, A.R., and Cranfill, R. (2004b). Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Am. J. Bot. **91**: 1582–1598.

Qiao, Z. and Libault, M. (2013). Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology. Front. Plant Sci. **4**: 484.

Qiu, Y.-L. et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. **103**: 15511–15516.

Qiu, Y.L., Cho, Y., Cox, J.C., and Palmer, J.D. (1998). The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature **394**: 671–674.

Raubeson, L. a and Jansen, R.K. (1992). Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science **255**: 1697–1699.

Raven, J. a and Edwards, D. (2001). Roots: evolutionary origins and biogeochemical significance. J. Exp. Bot. **52**: 381–401.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139–140.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139–140.

Rognes, T. (2011). Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. BMC Bioinformatics **12**: 221.

Roppolo, D., De Rybel, B., Dénervaud Tendon, V., Pfister, A., Alassimone, J., Vermeer, J.E.M., Yamazaki, M., Stierhof, Y.-D., Beeckman, T., and Geldner, N. (2011). A novel protein family mediates Casparian strip formation in the endodermis. Nature **473**: 380–3.

Rost, T.L. (2011). The organization of roots of dicotyledonous plants and the positions of control points. Ann. Bot. **107**: 1213–1222.

Rouard, M., Guignon, V., Aluome, C., Laporte, M.A., Droc, G., Walde, C., Zmasek, C.M., Périn, C., and Conte, M.G. (2011). GreenPhylDB v2.0: Comparative and functional genomics in plants. Nucleic Acids Res. **39**: 1095–1102.

Rounds, C.M. and Bezanilla, M. (2013). Growth mechanisms in tip-growing plant cells. Annu. Rev. Plant Biol. **64**: 243–65.

Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., and Burleigh, J.G. (2014). From algae to angiosperms - inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. **14**: 23.

Salazar-Henao, J.E., Vélez-Bermúdez, I.C., and Schmidt, W. (2016). The regulation and plasticity of root hair patterning and morphogenesis. Development **143**: 1848–1858.

Schellmann, S., Schnittger, A., Kirik, V., Wada, T., Okada, K., Beermann, a., Thumfahrt, J., Jürgens, G., and Hülskamp, M. (2002). *TRIPTYCHON* and *CAPRICE* mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. EMBO J. **21**: 5036–5046.

Schiefelbein, J., Huang, L., and Zheng, X. (2014). Regulation of epidermal cell fate in Arabidopsis roots: the importance of multiple feedback loops. Front. Plant Sci. **5**: 47.

Schiefelbein, J., Kwak, S.H., Wieckowski, Y., Barron, C., and Bruex, A. (2009). The gene regulatory network for root epidermal cell-type pattern formation in arabidopsis. J. Exp. Bot. **60**: 1515–1521.

Schiefelbein, J. and Somerville, C. (1990). Genetic Control of Root Hair Development in Arabidopsis thaliana. Plant Cell **2**: 235–243.

Seago, J.L. and Fernando, D.D. (2013). Anatomical aspects of angiosperm root evolution. Ann. Bot. **112**: 223–238.

Simon, M., Lee, M.M., Lin, Y., Gish, L., and Schiefelbein, J. (2007). Distinct and overlapping roles of single-repeat MYB genes in root epidermal patterning. Dev. Biol. **311**: 566–578.

Song, S.-K., Ryu, K.H., Kang, Y.H., Song, J.H., Cho, Y.-H., Yoo, S.-D., Schiefelbein, J., and Lee, M.M. (2011). Cell Fate in the Arabidopsis Root Epidermis Is Determined by Competition between WEREWOLF and CAPRICE. Plant Physiol. **157**: 1196–1208.

Sozzani, R. and Iyer-Pascuzzi, A. (2014). Postembryonic control of root meristem growth and development. Curr. Opin. Plant Biol. **17**: 7–12.

Stahl, Y. and Simon, R. (2010). Plant primary meristems: shared functions and regulatory mechanisms. Curr. Opin. Plant Biol. **13**: 53–58.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**: 2688–2690.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3 - MYB gene family in Arabidopsis thaliana. Curr. Opin. Plant Biol. **4**: 447–456.

Sztein, A.E., Cohen, J.D., Slovin, J.P., and Cooke, T.J. (1995). Auxin metabolism in representative land plants. Am. J. Bot. **82**: 1514–1521.

Tam, T.H.Y., Catarino, B., and Dolan, L. (2015). Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. Proc. Natl. Acad. Sci.: 201416324.

THE ANGIOSPERM PHYLOGENY GROUP (1998). AN ORDINAL The Angiosperm Phylogeny Group' CLASSIFICATION FOR THE FAMILIES OF FLOWERING PLANTS. Ann. Missouri Bot. Gard. **85**: 531–553.

THE ANGIOSPERM PHYLOGENY GROUP (2003). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Bot. J. Linn. Soc. **141**: 399–436.

THE ANGIOSPERM PHYLOGENY GROUP (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc. **161**: 105–121.

THE ANGIOSPERM PHYLOGENY GROUP (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc. **181**: 1–20.

Tomescu, A.M.F., Wyatt, S.E., Hasebe, M., and Rothwell, G.W. (2014). Early evolution of the vascular plant body plan - the missing mechanisms. Curr. Opin. Plant Biol. **17**: 126–136.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. **31**: 46–53.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. **7**: 562–78.

Verbelen, J.-P., De Cnodder, T., Le, J., Vissenberg, K., and 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.

De Vetten, N., Quattrocchio, F., Mol, J., and Koes, R. (1997). The an11 locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. Genes Dev. **11**: 1422–1434.

Wada, T., Kurata, T., Tominaga, R., Koshino-Kimura, Y., Tachibana, T., Goto, K., Marks, M.D., Shimura, Y., and Okada, K. (2002). Role of a positive regulator of root hair development, CAPRICE, in Arabidopsis root epidermal cell differentiation. Development **129**: 5409–5419.

Wada, T., Tachibana, T., Shimura, Y., and Okada, K. (1997). Epidermal cell differentiation in Arabidopsis determined by a Myb homolog, CPC. Science **277**: 1113–1116.

Walker, A.R., Davison, P.A., Bolognesi-Winfield, A.C., James, C.M., Srinivasan, N., Blundell, T.L., Esch, J.J., Marks, M.D., and Gray, J.C. (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. Plant Cell **11**: 1337–1349.

Willemsen, V., Bauch, M., Bennett, T., Campilho, A., Wolkenfelt, H., Xu, J., Haseloff, J., and Scheres, B. (2008). The NAC Domain Transcription Factors FEZ and SOMBRERO Control the Orientation of Cell Division Plane in Arabidopsis Root Stem Cells. Dev. Cell **15**: 913–922.

Xu, C.-R., Liu, C., Wang, Y.-L., Li, L.-C., Chen, W.-Q., Xu, Z.-H., and Bai, S.-N. (2005). Histone acetylation affects expression of cellular patterning genes in the Arabidopsis root epidermis. Proc. Natl. Acad. Sci. U. S. A. **102**: 14469–14474.

Yi, K., Menand, B., Bell, E., and Dolan, L. (2010). A basic helix-loop-helix transcription factor controls cell growth and size in root hairs. Nat. Genet. **42**: 264–267.

Yu, Z., Kang, B., He, X., Lv, S., Bai, Y., Ding, W., Chen, M., Cho, H.T., and Wu, P. (2011). Root hair-specific expansins modulate root hair elongation in rice. Plant J. **66**: 725–734.

Zheng, K., Tian, H., Hu, Q., Guo, H., Yang, L., Cai, L., Wang, X., Liu, B., and Wang, S. (2016). Ectopic expression of R3 MYB transcription factor gene OsTCL1 in Arabidopsis, but not rice, affects trichome and root hair formation. Sci. Rep. **6**: 19254.

Nakazawa, M. (2014). Practices of Medical and Health Data Analysis using R. Boston, Pearson Education.

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. New York, Springer.