The origin and evolution of vascular tissue regulators TMO5 and LHW

Nicole van 't Wout Hofland

Propositions

1.   Vascular plants evolved through the redeployment of ancient developmental regulators TMO5 and LHW (this thesis).

2.   Homo- to heterodimerization transitions may be a key determinant for increased organismal complexity (this thesis).

3.   Given that the RNA world hypothesis involves atoms that are abundant throughout the universe, it is likely that life originated in the same manner on more than one planet.

4.   The focus on water during the search for extra-terrestrial life is understandable, but limited, since water may not be required for life in different systems.

5.   With abundant modern, internet-based information sources, communication by scientists to the general public is more important than ever.

6.   Gender equality can only be reached when society stops pushing gender specific roles.

Propositions belonging to the thesis, entitled
"The origin and evolution of vascular tissue regulators TMO5 and LHW"

Nicole van 't Wout Hofland

The origin and evolution of vascular tissue regulators TMO5 and LHW

Nicole van 't Wout Hofland

The origin and evolution of vascular tissue regulators TMO5 and LHW

Nicole van 't Wout Hofland

Table of Contents

Chapter 1

Introduction

1

1. The use of model organisms in comparative biology

One of the fascinating goals in biology is understanding how the morphological novelties leading to big evolutionary leaps and thriving new species can come into existence. Model organisms such as the yeast *Saccharomyces cerevisiae*, the fruit fly *Drosophila melanogaster*, the nematode *Caenorhabditis elegans*, the zebrafish *Danio rerio* and the plant *Arabidopsis thaliana*, provide useful platforms for unravelling cellular and molecular mechanisms and have contributed greatly to the knowledge of the science community. However, investigating the emergence of novel body plans or organs such as the coelom in animals and seeds in plants, cannot be studied using a single model species. The onset of large-scale genomic and transcriptomic sequencing projects now makes it possible and attractive to move away from model species and study how certain traits or molecular mechanisms arose and evolved over time. While other species are being studied, *Arabidopsis thaliana* is considered the primary model species within the plant lineage. However, many evolutionary innovation have taken place in the 1 billion years that the green lineage has been in existence (Moczydlowska et al., 2011). Understanding of phylogenetic relationships between major plant lineages allows comparative biological research to such innovations, and shed light on how certain major evolutionary leaps could have occurred.

2. Evolution and early innovations in the green lineage

The green lineage (Viridiplantae) was established by an endosymbiotic event between 1.5 and 1 billion years ago (Alberts et al., 2008). During this event, a eukaryotic cell captured and integrated an ancestral cyanobacterium which ultimately developed into a chloroplast. Diversification of an unicellular ancestor approximately 1000 to 540 million years ago, created a lineage, which can be divided in two main clades: Chlorophytes and Streptophytes (**Figure 1**) (Floyd and Bowman, 2007). The former consists of a wide range of marine, freshwater and terrestrial green algae (**Figure 2B**), whereas the latter comprises the Charophyte green algae and all land plants (Embryophytes) (**Figure 1**). Charophytes emerged after Chlorophytes split from their ancestor (**Figure 2A**) (Yoon et al., 2004), and are exclusively freshwater and terrestrial green algae. Both Chlorophytes and Charophytes are morphologically diverse and contain unicellular and multicellular algal species. As such, the transition from a unicellular ancestor to simple multicellular organisms occurred within green algae. This innovation likely evolved as a result of predation, since algae above a certain threshold size are able to escape microscopic predators such as the unicellular *Paramecium tetraurelia* (Herron et al., 2019). Multicellularity also allowed diverse morphological assortments, such as filamentous and branched forms, to evolve. Approximately 460 million years ago, many additional adaptation occurred within an ancestral lineage of Charophytes, allowing them to cope with terrestrial conditions (Reviewed by Delwiche and Cooper, 2015). This supported them to survive and reproduce while being exposed to the atmosphere. These Charophytes are considered
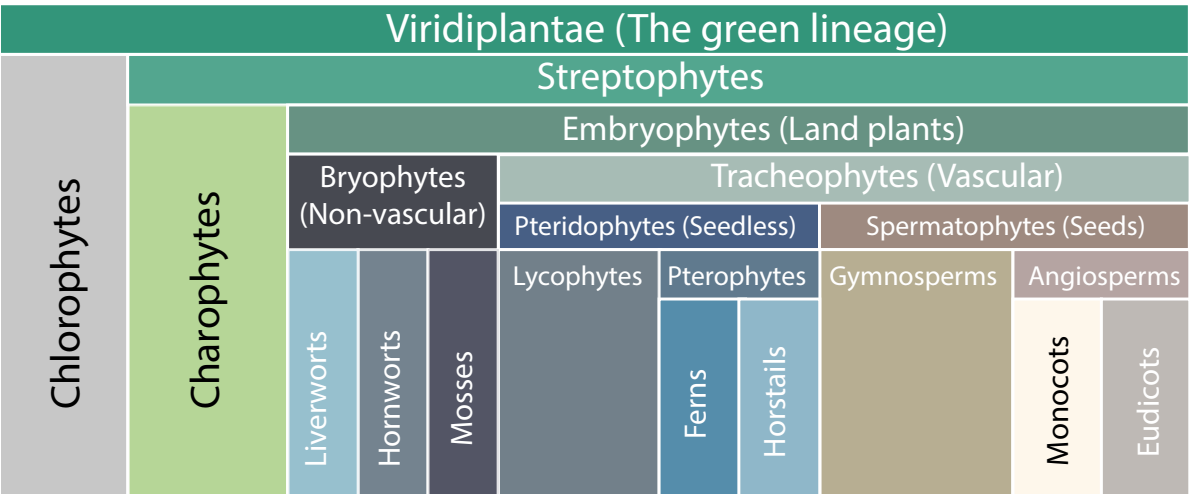
Figure 1. Overview of the plant taxonomy within the plant kingdom.

as the closest living relatives to Embryophytes (**Figure 2A**).

Charophytic algae can be divided in six main lineages: Mesostigmatophyceae, Chlorokybophyceae, Klebsormidiophyceae, Charophyceae, Zygnematophyceae and Coleochaetophyceae (**Figure 2A**) (Mattox and Stewart, 1984; McCourt et al., 2004). Phylogenetic studies have indicated that Mesostigmatophyceae and Chlorokybophyceae are the earliest diverging lineages within the Charophytes (**Figure 2A**) (Bhattacharya and Medlin, 1998; Lemieux et al., 2007; Simon et al., 2006). Indeed, morphologically, these are the simplest two Streptophyte algae: Mesostigma (the only genus in the Mesostigmatophyceae) is unicellular containing two flagella and Chlorokybus (the only member of the Chlorokybophyceae) forms packets of a few cells. Filamentous algae are found in the Klebsormidiophyceae lineage (**Figure 2C**), which diverged after Mesostigmatophyceae and Chlorokybophyceae (**Figure 2A**). These filamentous algae generally do not contain specialized or differentiated cells and branching is very rare (Mikhailyuk et al., 2014). Instead, branching as well as increased morphological complexity in general, is mostly observed in the three later diverging Charophytes. Charales (the only order in the Charophyceae) for example, are large morphologically complex macroscopic algae with differentiated cells and an apical meristem (**Figure 2D**). They generally contain a central stalk with many branches (**Figure 2D**). Furthermore, other innovations are present in the later diverging Charophytes, contributing to complexity. For example, Charales and Coleochaetophyceae develop plasmodesmata, allowing cell-cell communication. While the three early diverging lineages use furrowing during the cytokinesis step of cell division, Zygnematophyceae, Charophyceae and Coleochaetophyceae, produce a phragmoplast, similar to Embryophytes. Zygnematophyceae and Coleochaete are thought to be the closest related to Embryophytes (**Figure 2A**). However, the Zygnematophyceae lineage appears to have lost many traits such as plasmodesmata, flagella and extensive branching.
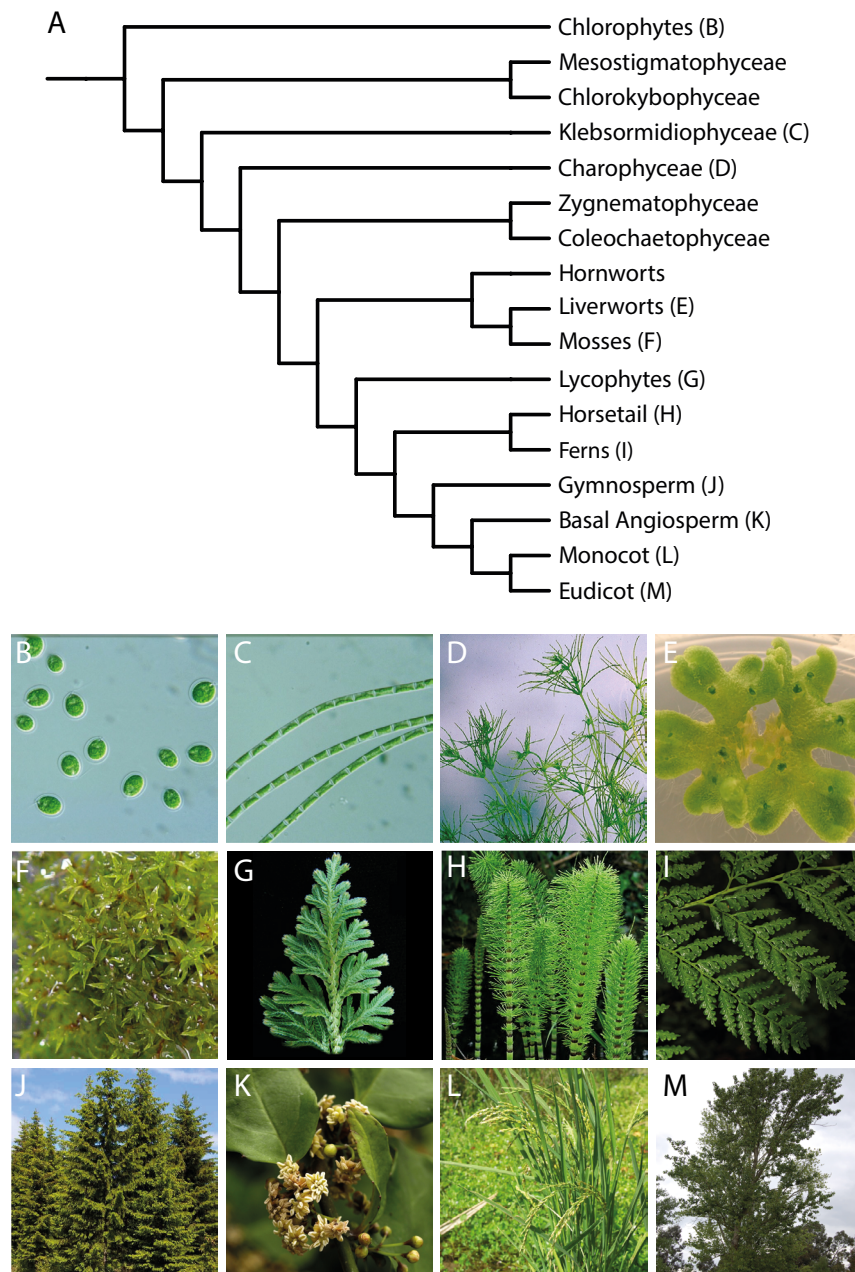
Figure 2. Species within the green lineage and their phylogeny. (A) Phylogenetic relationships within the green lineage. Letters refer to Figure 2B-M. **(B)** The unicellular Chlorphytes *Chlamydomonas reinhardtii* (Protist Information Server). (C) The filamentous Charophyte *Klebsormidium nitens* (Protist Information Server). **(D)** The complex Charophyte *Chara (Charophyceae*. Courtesy of Prof. K. Fitzsimmons, University of Arizona **(E)** The liverwort *Marchantia Polymorpha*. **(F)** The moss *physcomitrella patens* (http://greenc.sciencedesigners.com/wiki/Species:Physcomitrella_patens) **(G)** The Lycophyte *Selaginella moelendorffii.* Courtesy of Jing-Ke Weng, Purdue University. (H) the horsetail *Equisetum telmateia* (https://en.wikipedia.org/wiki/Equisetum). **(I)** The fern *Culcita macrocarpa*. Coutesy of Jose Luis Perez Calo. **(J)** Gymnosperms *Picea abies* (https://www.scienceview.gr/news/mapping-a-giant-s-genome-picea-abies/). **(K)** The basal Angiosperm *Amborella trichopoda*. Courtesy of dr. Scott Zona (Florida International University). **(L)** The monocot *Oryza Sativa* (http://knowledge.taibif.tw/node/15500). **(M)** The eudicot *Populus trichocarpa*. Courtesy of Davis Landscape Architecture, London.

The colonization of land by an ancestor related to Charophyte algae, was followed by the emergence of the first land plants, approximately 460 million years ago (Lemieux et al., 2007; Lewis and McCourt, 2004; Wellman et al., 2003). The earliest diverging Embryophytes are mosses, liverworts and hornworts, collectively known as Bryophytes (**Figure 1**). This lineage holds a key position in evolution, as they diverged after the transition from water to land, but prior to the emergence of vascular tissues. Indeed, Bryophytes are plants without lignified vascular systems or true leaves and are generally a few millimetres to a few centimetres tall (**Figure 2E, F**). While Bryophytes are considered to be simple plants, terrestrialization did require a number of physiological and structural adaptations to the plant body, such as the development of a water-repellent cuticle and stomata to regulate water evaporation (Kenrick, 2008). Furthermore, Bryophytes develop unicellular (in liverworts and hornworts) or multicellular (in mosses) rhizoids to anchor to the soil. Another developed feature in the ancestor of Bryophytes is a life cycle in which a haploid sexual generation alternates with a diploid asexual generation. Unlike many other plants, the haploid generation is the dominant generation in Bryophytes.

Bryophytes are non-vascular plants since they lack an elaborate lignified water transportation system. Instead, early diverging land plants take up water by diffusion through the cell wall, which is subsequently conducted over short distances by water-conducting cells (WCCs). WCCs are dead cells, devoid of cytoplasmic contents, containing smooth walls and small plasmodesmata-derived perforations. Somewhat more advanced water-conducting conduits, called hydroids, can be found in the central strand of certain mosses (peristomate mosses) (Kenrick and Crane, 1997; Ligrone et al., 2000; Richardson, 1981). Hydroids are elongated cells that die at maturity and subsequently lose their cytoplasmic contents (Ligrone et al., 2000; Richardson, 1981). In contrast to other WCCs, hydroids lack perforations (Scheirer, 1975). Instead, oblique cell walls between two adjacent hydroids are highly permeable, allowing efficient water conduction. Although some hydroid walls become thickened (Scheirer, 1975), they lack secondary cell wall patterns containing the polyphenolic polymer lignin, responsible for mechanical strength. Interestingly, lignin-like polymers have been identified in Bryophytes (Espiñeira et al., 2011; Ligrone et al., 2008). However, they are not cell-type specific and do not contain mechanical strength properties. It has been suggested that these polymers function in microbe protection instead (Gunnison and Alexander, 1975). Although the lack of structural support impedes vertical growth, it does allow Bryophytes to collapse during desiccation and recover quickly upon water availability (Ligrone et al., 2012; Lucas et al., 2013).

Conduction of carbohydrates and amino acids in Bryophytes is realized by food-conducting cells (FCCs), and the more elaborate leptoids, which develop in close proximity to hydroids and WCCs (Thomas et al., 1988). FCCs form files of elongated, polarly organized cells with trumpet-shaped ends containing high amounts of plasmodesmata, facilitating cell-cell transport (Ligrone et al.,

2000; Lucas et al., 2013).

3. The emergence of vascular plants

With the emergence of vascular plants (Tracheophytes), highly advanced lignified WCCs and FCCs, respectively termed xylem and phloem, evolved in shoots and roots. Xylem is responsible for the transport and storage of water and nutrients. At the same time its lignin-containing secondary cell walls provide rigidity, supporting the plant body and allowing it to grow to greater heights. The specialized FCC phloem is responsible for circulation of sucrose, proteins and mRNA involved in plant growth and development. The establishment of vascular tissues and roots allowed plants to reach water deep in the soil and transport it throughout the plant, reducing the need for moist habitats.

Extant Tracheophytes consist of three main lineages (**Figure 1**): Lycophytes, Pterophytes (containing ferns and horsetails) and seed plants (Spermatophytes). While root systems are present in nearly all Tracheophytes, the common ancestor of Lycophytes and Euphyllophytes (Pterophytes and Spermatophytes) was a rootless plant, indicating that roots evolved at least twice (Kenrick and Crane, 1997; Raven and Edwards, 2001). Thus, Lycophytes, which have the longest evolutionary history of all Tracheophytes (**Figure 2A**), are distinct from all other vascular plants. Overall, Lycophytes have a simple morphology containing a stem, a lignified vascular system with simple arrangements of xylem and phloem and microphylls ('leaves') containing a single, unbranched mid-vein (**Figure 2G**). Each microphyll contains a kidney-shaped sporangium near the point where the microphyll is attached to the stem. With the genome sequencing of the Lycophyte *Selaginella moellendorffii* in 2011 (Banks et al., 2011), Selaginella has become a well-studied representative of the Lycophyte lineage.

Other basal vascular plants belong to the Pterophyte lineage (**Figure 1**). Similar to Lycophytes, Pterophytes do not contain seeds or true leaves, but rather microphylls (horsetails) or fronds (ferns) (**Figure 2H, I**). Reproduction occurs by spores, which are dispersed via wind. While Pterophytes and Lycophytes were thriving plant clades on Earth during the Paleozoic period (up to 251 million years ago), competition with later diverging Spermatophytes (Gymnosperms and Angiosperms) resulted in reduced amounts of species within these lineages.

One of the reasons Gymnosperms and Angiosperms are thriving and can easily compete with earlier diverging land plants, is the innovation of reproduction by seeds, which eliminates their dependence on water for reproduction. In addition, seeds provide protection and nourishment to the embryo and allows dormancy for long periods of time. As such, germination can occur when growth conditions are favourable. The emergence of Spermatophytes is correlated with other major trends. One example is the separation of male and female spores. Many seedless plants produce a

1

single spore type (homospory), whereas Spermatophytes produce a large female macrospore and a small male microspore (heterospory), which mature into egg-producing female gametophytes and sperm-producing male gametophytes respectively.

It is believed that the common ancestor of all Spermatophytes were vegetatively similar to seed plants, while they reproduced in a manner similar to Pterophytes, i.e. by releasing spores. This group of plants is termed Progymnosperms (Beck, 1966; Rothwell and Erwin, 1987). The oldest known seed plant, identified by fossils, is the extinct *Elkinsia polymorpha,* which emerged between 385 and 365 million years ago (Rothwell et al., 1989; Serbet and Rothwell, 1992). This plant protected its ovules by enclosing it in cup-like structures called cupules (Rothwell et al., 1989). Elkinsia is classified as a seed fern, which is the extinct ancestral group of which all Spermatophytes evolved. The earliest diverging extant Spermatophytes are Gymnosperms such as pine and spruce (**Figure 2A, J**). In contrast to Angiosperms, their seeds are exposed to the environment and are not enclosed within an ovary.

While the first Angiosperm fossils date back to approximately 136 million years ago (Hickey and Doyle, 1977), the initial split of Angiosperm and Gymnosperm lineages may have occurred more than 310 million years ago (Doyle, 2012; Magallón et al., 2013). Today, Angiosperms represent one of the dominant terrestrial radiations due to their accumulated advanced tissues and organs such as flowers, the endorsperm and enclosed seeds (**Figure 2K, L, M**). Although flowers are the defining feature of Angiosperms, their origin and subsequent diversification is poorly understood. It has however been proposed that a whole genome duplication event in an early ancestor of flowering plants, followed by extensive gene diversification, provided novel gene functions that led to flower development (Amborella Genome Project, 2013; Callaway, 2013; Jiao et al., 2011).

Some angiosperms develop a single cotyledon during embryogenesis, whereas others develop two. Based on this morphological characteristic, Angiosperms are divided in monocots and eudicots (**Figure 1, 2A**). Other differences between monocots and eudicots include difference in pollen structures, leaf vein morphology and vascular patterning.

4. Vascular tissue development in the Tracheophyte Arabidopsis

Recent studies estimated that plants contribute to approximately 80% of all biomass on Earth (Bar-On et al., 2018). Due to the successful terrestrial radiation of Tracheophytes, most of this plant biomass is comprised of vascular plants (Bar-On et al., 2018). As discussed previously, vascular systems make up elaborate networks of conducting tissues that stretches throughout the entire plant body. In Arabidopsis, the development of vascular tissues can be subdivided in several phases: specification of cell identity, proliferation, patterning, maintenance and differentiation.

4.1. Vascular tissue specification and proliferation

In Arabidopsis, cell divisions early during embryogenesis create four cells with a vascular identity in the centre of the embryo (Scheres et al., 1994). These will eventually give rise to all cells in the root vascular bundle of the mature plant. However, in order to generate a specialized vascular cylinder, the vascular initial cells require radial expansion both during and after embryogenesis. In plants, where cells are immobile, such increase of tissue dimensions requires organized cell divisions. Cell divisions that occur parallel to the surface of the plant body and result in increased girth are called periclinal (perpendicular to the plant body axis and parallel to the periphery) and radial (parallel to the plant body axis and perpendicular to the periphery) cell divisions, whereas anticlinal divisions result in longitudinal growth (**Figure 3**, See glossary). Periclinal and radial cell divisions are thus crucial for the radial expansion during the establishment of vascular tissues and require tight regulation.

At the end of embryogenesis, periclinal/radial cell divisions created a vascular bundle of approximately 30 cells. At this time, all vascular cell type identities, including xylem- and phloem initials, are established (Bauby et al., 2007; Bonke et al., 2003). These cells will later differentiate to form the main conduits for long-distance transport of water, nutrients, and signalling molecules within the plant.
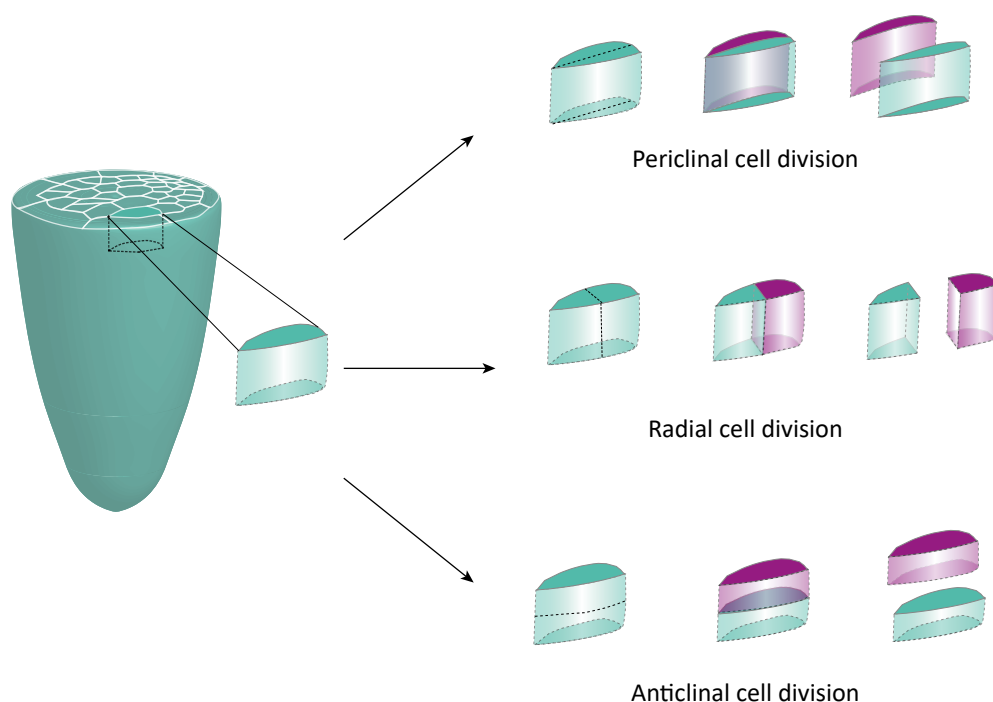
Figure 3. A schematic overview of three possible division planes of a cell. The vascular bundle of the root is shown with one highlighted cell indicating the different divisions planes. Periclinal and radial cell divisions give rise to additional cell files and facilitate radial growth. During periclinal cell divisions (top), the division plane runs parallel to the surface of the plant body, whereas during radial cell divisions (middle) the division plane is perpendicular to the plant body. Anticlinal cell divisions (bottom) result in longitudinal growth.

4.2. Vascular patterning

Auxin and cytokinin play crucial roles in regulating patterning and it is believed that their signalling early in embryogenesis is the first step towards a bilateral symmetry pattern (De Rybel et al., 2014). Several studies suggest that asymmetry in auxin signalling, which occurs upon cotyledon initiation, controls vascular patterning, in which cells that receive high auxin levels ultimately differentiate into xylem cells (Bishopp et al., 2011; De Rybel et al., 2014; Mähönen et al., 2006a; Mellor et al., 2016). With the establishment of xylem and phloem initials, the vascular pattern is modified. While morphologically indistinguishable, the specification of two opposite phloem poles at the periphery and a centrally localized xylem axis induces a shift from a radial to a bilateral symmetry pattern.

The vascular organization with two phloem and two xylem poles as found in Arabidopsis is referred to as a diarch pattern. In addition, plant species containing one (monarch), three (triarch), four (tetrarch) or many (polyarch) poles exist. The relatively simple diarch patterning which exists in Arabidopsis, is a reflection of the narrow vascular bundle in this species. In contrast, species with wider vascular cylinders, such as *Medicago sativa*, are triarch (Winter, 1932). Another example comes from pea roots. When grown in cultures, these normally triarch patterned roots showed a reduced root diameter and a shift to diarch or even monarch patterns (Torrey, 1955).

4.3. Vascular cell differentiation

During vascular cell differentiation, (pro)cambium daughter cells cease dividing and mature into specialized xylem or phloem cells. Three mature xylem tissues are: xylary fibres, xylem parenchyma cells and xylem tracheary elements (TEs) (**Figure 4**). Xylary fibres are narrow, highly elongated cells that undergo extensive secondary cell wall thickening and lignification, making them suitable to provide structural support. Xylem parenchyma cells are involved in multiple processes such as storing metabolites and promoting lignification in neighbouring vessel elements and fibres. While these cells assist in lignification processes, they do not contain extensive secondary cell walls themselves. Finally, TEs are hollow dead cells, devoid of cytoplasmic contents, with reinforcing secondary cell walls and perforated cell terminal ends. This makes them specialized for water and solute transport between organs. TEs are formed both during early and later stages of vascular development, which results in two distinct types of TEs: proto- and metaxylem respectively. In Arabidopsis roots, these cells can be distinguished morphologically, by the pattern of their secondary cell wall deposition, as well as on their spatial organization. Protoxylem develops in the periphery of the vascular bundle during primary plant growth and is characterized by the deposition of ring-shaped or spiral cell wall thickenings, allowing continued cell elongation (**Figure 4**). During later stages of development, the shorter and broader metaxylem TEs are formed in the centre with pitted or reticulate patterns of secondary cell wall thickenings (**Figure 4**).

Besides xylem tissues, procambium cells also give rise to mature phloem tissues, most of which are alive after maturation. Phloem tissue types include sieve tube elements, phloem companion cells, and phloem fibres (**Figure 4**). The long distance transportation function is executed by sieve tubes, which are comprised of several elongated cells, devoid of most of their cellular components. These so-called sieve elements are connected to each other via large pores at the end walls. Given that sieve elements lack many organelles, their survival depends on neighbouring companion cells, which are
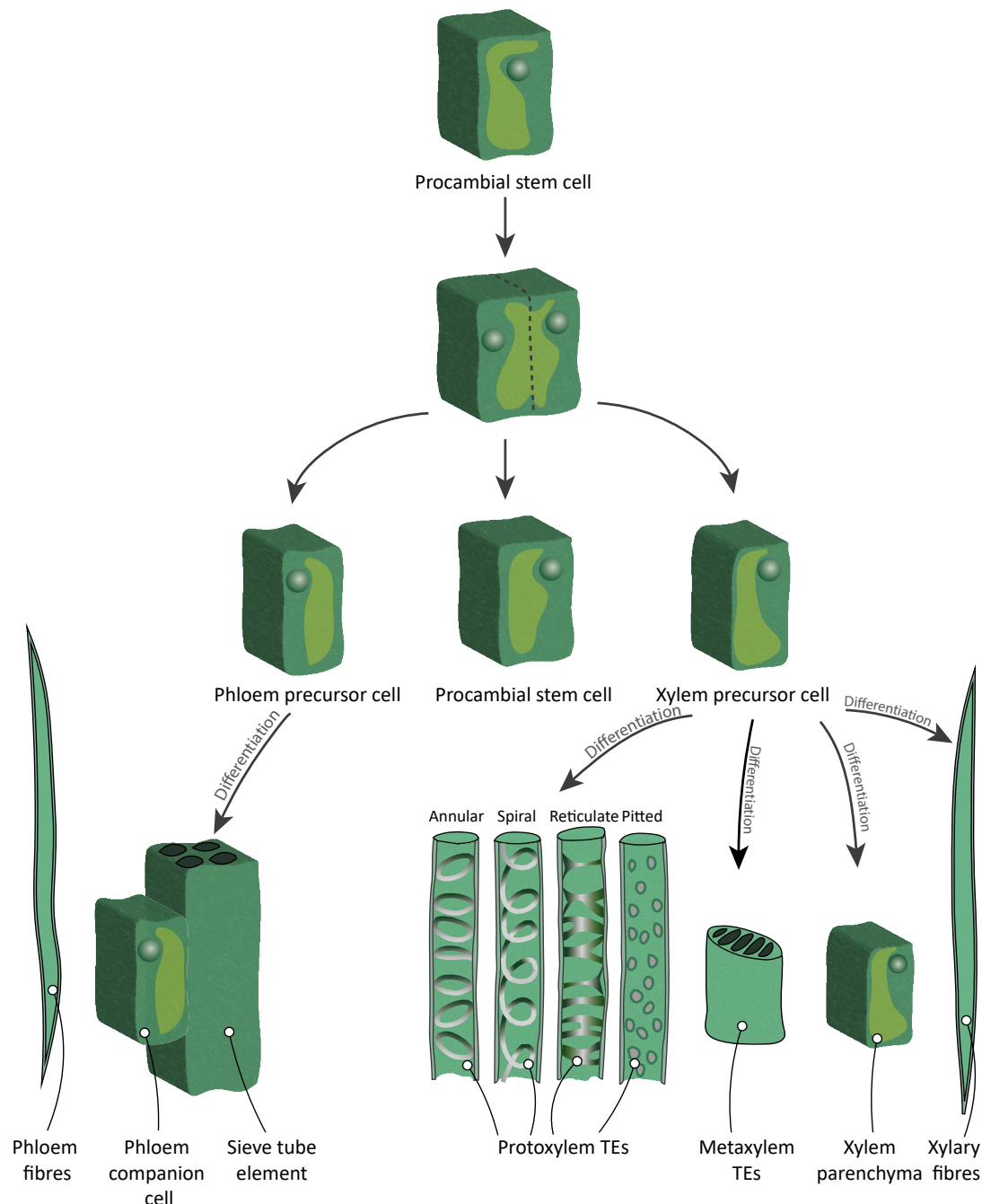
Figure 4. Schematic overview of xylem and phloem differentiation. Stem cells in the procambium divide and give rise to phloem or xylem precursor cells. Differentiation of these precursor cells creates specialized cells in the phloem (phloem companion cells, sieve tube elements, phloem fibres) and xylem (Tracheary elements or TEs, paranchyma and xylary fibres). Reconstructed from Schuetz et al., 2013.

connected through plasmodesmata. Similar to fibres in the xylem, phloem fibres provides strength to the conductive cells.

5. Molecular mechanisms of vascular development in Arabidopsis

5.1. Vascular proliferation

In Arabidopsis, the first vascular initial cells are specified early during embryogenesis (Scheres et al., 1994). However, in order to generate a vascular pattern with differentiated vascular cells, these initial vascular cells require radial expansion both during and after embryogenesis. Once specified, xylem and phloem precursor cells remain quiescent, while cell proliferation continues in procambium cells.

In 2008, Matsumoto-Kitano and colleagues showed that cambium formation was impaired in a quadruple mutant that disrupted ATP/ADP isopentenyl transferases (IPTs), which catalyse the first step in cytokinin biosynthesis. Moreover, a single IPT mutant, in which cytokinin levels were decreased compared to wild type, displayed reduced root girth (Matsumoto-Kitano et al., 2008). Similarly, the *wooden leg* (*wol*) mutant, defective in the cytokinin receptor ARABIDOPSIS HISTIDINE KINASE 4 (AHK4), shows significant reduction in vascular cell file numbers (Mähönen et al., 2000). Furthermore, mutations in three cytokinin receptors result in small vascular bundles, solely comprised of protoxylem (Mähönen et al., 2006). Cytokinin thus plays a crucial role in cambium proliferation. It was recently shown that DOF-type transcription factors act downstream of cytokinin in the vasculature and control specific procambium cell divisions (Miyashima et al., 2019; Smet et al., 2019).

Local cytokinin signalling in vascular cells is thus key in organized periclinal cell divisions and vascular proliferation. A major pathway responsible for local cytokinin production involves two basic HELIX-LOOP-HELIX (bHLH) transcription factors: TARGET OF MONOPTEROS 5 (TMO5) and LONESOME HIGHWAY (LHW) and is relatively well-understood (**Figure 5**). This pathway is already active in the first vascular initial cells during embryogenesis. Like many processes in plants, it depends on auxin signalling. Following biosynthesis, auxin is actively transported downwards via polar PIN-FORMED (PIN) proteins, during which it moves through vascular (initial) cells. Upon auxin perception in these vascular cells, transcriptional inhibitors Auxin/Indole-3-Acetic Acids (AUX/IAAs), which bind AUXIN RESPONSE FACTORS (ARFs), are ubiquitinated and subsequently degraded. Consequentially, ARF proteins are free to perform their function. One of these ARF proteins, ARF5 – also known as MONOPTEROS (MP), binds the promoter of *TMO5* and activates its transcription. TMO5 and its homologs form heterodimer complexes with LHW and its homologs. This heterodimer binds the promoter region of *LONELY GUY 4* (*LOG4*)

as well as the homolog, *LOG3*, which encode 5'-monophosphate phosphoribohydrolases, responsible for the catalysis of the final step in cytokinin biosynthesis. Cytokinin subsequently triggers periclinal and radial divisions in the neighbouring procambium cells (**Figure 5**).
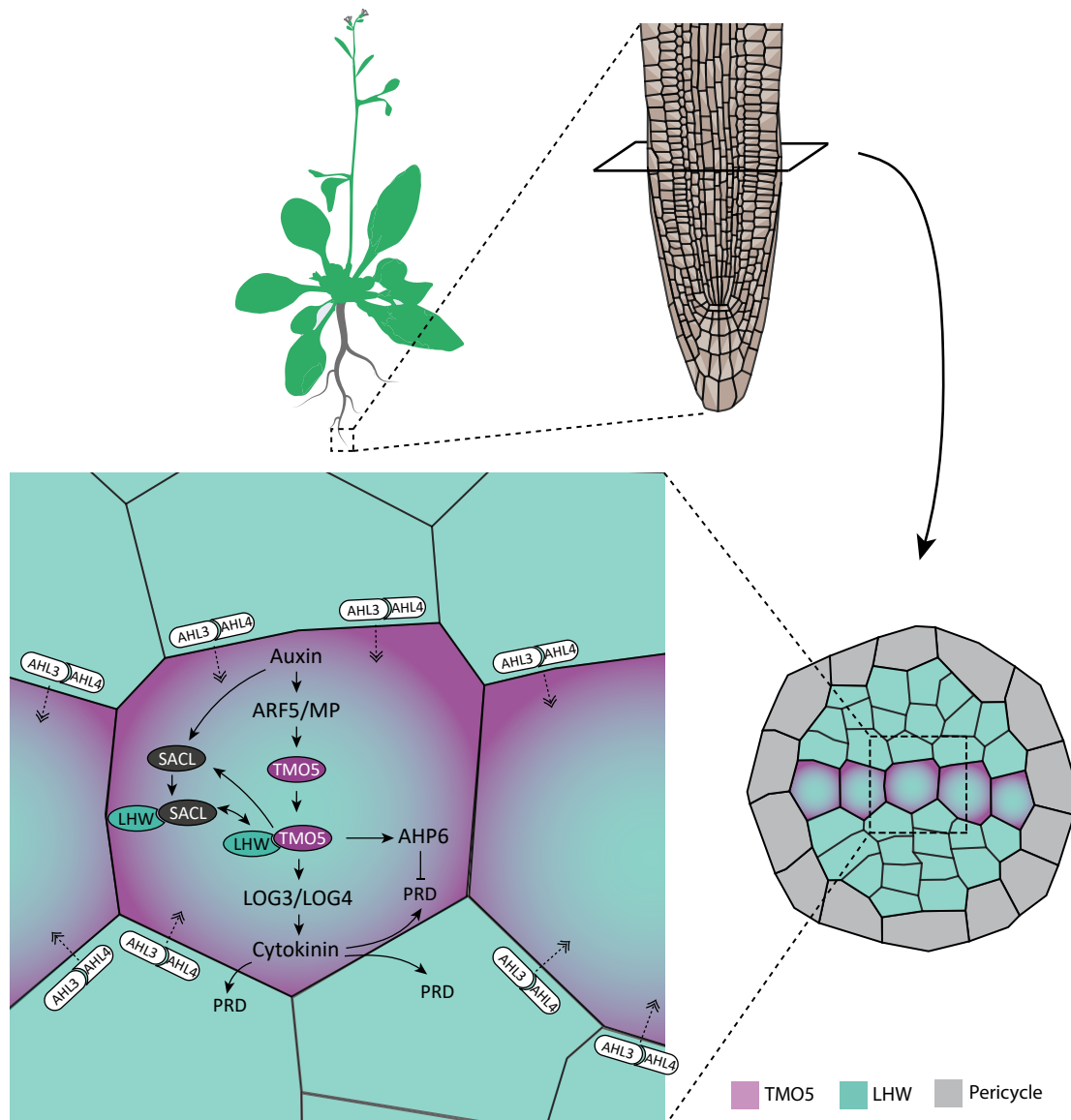
Figure 5. Schematic representation of the genetic pathways controlling vascular proliferation and vascular boundaries that take place in the Arabidopsis root meristem. An *Arabidopsis thaliana* plant is shown with an enlargement the root (top panels). A cross section of the Arabidopsis root is presented in the bottom right panel, showing the endodermis (grey) and the vascular bundle (green/purple). Expression patterns of the vascular regulators TARGET OF MONOPTEROS 5 (TMO5) and LONESOME HIGHWAY (LHW) are indicated in purple and green respectively. The genetic pathways controlling vascular proliferation and vascular boundaries is indicated in the bottom left panel. Within the xylem axis (purple/green), high auxin levels activate AUXIN RESPONSE FACTOR 5 (ARF5), also known as MONOPTEROS (MP). MP induces TMO5 levels, allowing the formation of TMO5/LHW dimers. I turn, TMO5/LHW heterodimers induce LONELY GUY 3 (LOG3) and LOG4, leading to cytokinin biosynthesis. Cytokinin subsequently triggers periclinal/radial cell divisions (PRD). ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN 6 (AHP6), induced by TMO5/LHW prevents PRDs in the xylem axis. A negative feedback loop, involving SAC51-LIKE (SACL) proteins, is promoted by auxin and TMO5/LHW. The boundary between xylem and procambium cells is regulated by heterodimers of AT-HOOK MOTIF NUCLEAR LOCALIZED 3 (AHL3) and AHL4 that move from the procambium to the xylem.

The spatial restriction of cytokinin biosynthesis is coordinated by the bHLH heterodimer TMO5/LHW. *LHW* and its homologs are broadly expressed in the embryo and root meristem, whereas *TMO5* and homologs are expressed in young xylem cells (**Figure 5**). As a consequence, the TMO5/LHW heterodimer and subsequent cytokinin production are restricted to xylem cells. In this fashion, TMO5 and LHW are responsible for the bilateral symmetry (diarch) pattern in Arabidopsis. Indeed, in *lhw* single or *tmo5 t5l1* double mutants, reduction of periclinal cell divisions result in smaller vascular bundles with less complex (monarch) patterning (De Rybel et al., 2013; Ohashi-Ito and Bergmann, 2007).

In order to control the amount of vascular cell proliferation, a negative feedback loop limits TMO5/LHW activity. A key protein in this feedback loop is the bHLH protein SAC51-LIKE (SACL), which antagonizes TMO5/LHW activity (**Figure 5**). Indeed, misexpression of SACL genes phenocopies *TMO5* and *LHW* loss-of-function mutants with reduced vascular cell files (Katayama et al., 2015; Vera-Sirera et al., 2015). In contrast, a mutation in the *SACL* gene results in increased vascular proliferation and xylem differentiation. Interestingly, both the TMO5/LHW dimer and its repressor are controlled by auxin (Hanzawa et al., 2000). The auxin inducible ACAULIS 5 (ACL5) encodes a spermine synthase, involved in thermospermine biosynthesis. In the presence of thermospermine, the translational repression of *SAC51* and other *SACLs* is released, allowing transcription and translation of these bHLH proteins. Subsequently, SACLs form heterodimer complexes with LHW, thereby competing with TMO5/LHW heterodimerization (**Figure 5**). The efficiency of the negative feedback loop is, in part, achieved by the high interlinkage of the forward and feedback loop. Besides being controlled by auxin, both ACL5 and SACLs are upregulated by TMO5/LHW dimers, most likely by direct binding and activation of their promoters (Katayama et al., 2015; Vera-Sirera et al., 2015).

Recently, a mechanism was identified regulating vascular proliferation as well as the sharp boundary between dividing and more-quiescent cells. PHLOEM EARLY DOF 1 (PEAR1) and its homologs PEAR2, DOF6, TMO6, OBP2 and HCA2, collectively termed PEAR proteins, play a central role in vascular proliferation. This was demonstrated by the reduction of procambium and phloem cell files in *pear1 pear2 dof6 tmo6* quadruple mutants (Miyashima et al., 2019). The PEAR protein expression pattern is regulated by an elaborate pathway that is initiated by SHORTROOT (SHR) production in the stele. SHR is a mobile protein that moves to the neighbouring endodermis, where it interacts with the transcription factor SCARECROW (SCR) to activate the expression of microRNA (miRNA) 165 and miRNA166 (Carlsbecke et al., 2010; Cui et al., 2007). These mobile miRNA diffuse towards the vascular cylinder, creating a gradient from the vascular periphery (high) towards the centre of the vascular bundle (low). Here they target HD-ZIP III transcription factors for degradation, creating an inverse gradient with high levels in the central cells and low level at the periphery (Carlsbecke et al., 2010). Subsequently, HD-ZIP III proteins inhibit PEAR proteins,

resulting in high PEAR protein levels in procambial cells (Miyashima et al., 2019). Here, PEAR proteins induce periclinal/radial cell division through activation of downstream targets in a non-cell-autonomous manner.

5.2. Establishing boundaries

The diarch pattern in Arabidopsis roots is established by zones of high auxin signalling in the xylem axis and high cytokinin signalling in procambium cells. Given that TMO5 and LHW in the xylem precursor cells induce cytokinin biosynthesis, additional signals are necessary to maintain the established boundaries and prevent cytokinin responses, such as induced cell divisions, in the meristematic xylem axis. One such signal that establishes boundaries of hormonal output is an inhibitory feedback loop between auxin and cytokinin. In xylem precursor cells, T5L1/LHW, and most likely their homologs, bind the promoter regions of ARABIDOPSIS HISTIDINE PHOSPHOTRANSFER PROTEIN 6 (AHP6) and promote its expression (Ohashi-Ito et al., 2014). AHP6 functions as a cytokinin signalling inhibitor and thus maintains xylem precursor cells in a nondividing state (**Figure 5**). On the other side of the feedback loop, cytokinin promotes PIN depletion in certain parts of the membrane, rearranging the PIN polarity and directing the auxin flow towards the central xylem cells.

A second pathway that contributes to sharp boundaries between the procambium and xylem cells involve the mobile transcription factors AT-HOOK MOTIF NUCLEAR LOCALIZED 3 (AHL3) and its homolog AHL4. AHL3 and AHL4 form heterodimers and loss of function of either of these AT-HOOK factors results in additional proto- and metaxylem (Zhou et al., 2013). Both genes are expressed in the procambium, after which they move towards the xylem to regulate the boundary between these cell types (**Figure 5**) (Zhou et al., 2013). Although the exact mechanism remains unclear, AT-HOOK factors are thought to function via modifying chromatin accessibility.

Together, these pathways create a vascular bundle consisting of approximately 30 cell files with well-defined boundaries between cell types. Subsequent secondary growth further increases the vascular cell file numbers to around 150 in Arabidopsis. Overall, highly organized periclinal and radial cell divisions are crucial for proper plant development and growth.

1

Scope of the thesis

Tracheophytes have contributed extensively to the environmental conditions and ecosystems on our planet. Their evolution from the common ancestor of Bryophytes and Tracheophytes raises the key question of what genetic innovations were responsible for the establishment of elaborate vascular tissues in Tracheophytes. One fundamental difference in the conducting tissues between Bryophytes and Tracheophytes is its dimensions. While conducting tissues comprise only few cell files in Bryophytes, vascular bundles in Tracheophytes contain many cell files and increase even further in width during secondary growth (Kenrick and Crane, 1997; Ligrone et al., 2000, 2012; Ohtani et al., 2017; Ragni and Greb, 2018). In this thesis, we investigate if proliferation of vascular cells could be the key innovation that led to elaborate vascular tissues and thus also the rise of Tracheophytes.

In order to study the role of cell proliferation in vascular tissue emergence and evolution, we investigate the origin and evolution of known vascular tissue proliferation regulators: TMO5 and LHW. In **Chapter 2**, we describe an *in silico* study used to identify conservation of TMO5 and LHW orthologues in the green lineage. We report that both TMO5 and LHW have deep evolutionary origins that can be traced back to Charophytes.

In **Chapter 3**, we discuss how the capacity to induce periclinal/radial cell divisions in the vasculature (termed the vascular function - see glossary) emerged using the TMO5 and LHW orthologues identified in Chapter 2. We demonstrate a two-step evolutionary process which gave rise to two dimer components capable of controlling vascular proliferation in Tracheophytes. First, the vascular function of TMO5 emerged in the common ancestor of Bryophytes and Tracheophytes. Simultaneously, heterodimerization capacities were established. Secondly, LHW obtained its vascular function in the ancestor of Tracheophytes. We report that the TMO5/LHW dimer capable of performing the vascular function emerged simultaneously with the rise of elaborate vascular tissues.

Functional protein innovations can be the result of protein domain rearrangements or innovations. Hence, the domain architectures of TMO5 and LHW are investigated in **Chapter 4.** We explore the conservation and innovations of the protein domains. We further demonstrate that bHLH and ACT-like domains, present in both TMO5 and LHW are crucial for the vascular function. We further identify the basic region of the TMO5 bHLH domain as the region responsible for the gain of vascular function.

In **Chapter 5**, we dig deeper into the domain structure of TMO5 (the bHLH domain and the ACT-like domain) and LHW (the GAF-like domain, bHLH domain and the ACT-like domain) and further speculate on their role in dimerization and hormonal cross-talk.

Finally, in **Chapter 6**, we place our findings in a broader context and discuss how our results contribute to answering the main question of this thesis of how vascular tissues emerged in plants.

Chapter 2

Identification of TMO5 and LHW orthologues in the plant kingdom

Nicole van 't Wout Hofland[1], Klaas Vandepoele[2,3], Dolf Weijers[1] and Bert De Rybel[1,2,3]

1. Wageningen University, Laboratory of Biochemistry, Stippeneng 4, 6708 WE Wageningen, the
Netherlands
2. Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71,
9052 Ghent, Belgium
3. VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium

Abstract

The emergence of vascular tissues is one of the key innovations facilitating the wide radiation of plants. These conducting tissues ensure survival in habitats with limited water access and their mechanical support allows increased vertical growth. Vascular tissues can thus be regarded as one of the major evolutionary innovations in the plant kingdom. Yet, the molecular innovations that led to the evolution of these conductive tissues are unknown. In Arabidopsis, TARGET OF MONOPTEROS 5 (TMO5) and LONESOME HIGHWAY (LHW) play an important role in vascular tissue proliferation by inducing periclinal and radial cell divisions. Here, we investigate if TMO5 and LHW could be one of the key innovations leading to the emergence of vascular tissues. We use bioinformatics tools to study whether the emergence of *TMO5* and *LHW* genes is correlated with the emergence of complex vascular tissues. We reveal deep evolutionary origins of *TMO5* and *LHW* genes, pre-dating Tracheophytes. Consequently, we conclude that the presence of *TMO5* and *LHW* genes is not the determinant for complex vascular tissue formation.

Introduction

The colonization of land through the emergence of land plants (Embryophytes), approximately 460 million years ago, had major impacts on our planet. It significantly decreased the atmospheric CO_2 concentration, modified geological cycles of deposition and erosion, supplied nutrients and allowed the rise of other major clades of eukaryotes, creating complex ecosystems (Lentona et al., 2016; Piombino, 2016). Successful colonization however, requires the evolution of specialized cells, tissues and organs to cope with challenges such as drought, UV radiation and gravity (Floyd and Bowman, 2007; Waters, 2003). Some adaptive mechanisms, such as multicellularity, plasmodesmata and cellulosic cell walls, can also be found in algal species, indicating that their development most likely predates the origin of Embryophytes (Franceschi et al., 1994; Northcote et al., 1958). Adaptation to cope with limited water access however, were not established in algae or early diverging land plants. This may have been one of the major hurdles limiting plants from vast colonization of land. Specialized cells able to transport water, nutrients, hormones and other signalling molecules posed the solution to this problem and are unique to vascular plants (Tracheophytes) (Lucas et al., 2013; Raven, 1977). This conducting ability, together with the mechanical support it provides allowed plants to colonize a larger range of environmental habitats, thereby extensively impacting our planet. The acquisition of conducting tissues can therefore be regarded as one of the major evolutionary innovations in the plant kingdom.

Initial land colonization is speculated to have occurred by simple plants that diverged from a Charophytic (freshwater green alga) ancestor, resembling extant Bryophytes (Finet et al., 2010; Glime, 2015; Qiu et al., 2006). Bryophytes are the earliest diverging land plant lineages: liverworts, mosses, and hornworts (**Figure 1**) (Kenrick and Crane, 1997; Ligrone et al., 2012; Qiu et al., 2006). While attempts have been made to describe the body plan of the Bryophytic ancestor, this remains challenging since exact phylogeny within the Bryophytes remains controversial (Ligrone et al., 2012; Mishler and Churchill, 1985). Many studies suggest that liverworts represent the most basal lineage, and mosses and hornworts form the living sister group to Tracheophytes (Ligrone et al., 2012; Qiu et al., 2006), whereas some molecular and morphological analyses place hornworts as the earliest diverging land plants (Duff and Nickrent, 1999; Hedderson et al., 1996; Nickrent et al., 2000). Recently, Bryophytes were proposed to be monophyletic and sister to all Embryophytes (Puttick et al., 2018). Within the vascular plant lineage, evolutionary relationships are better understood. Lycophytes, such as Selaginella, form the most basal group, while Pterophytes (horsetails and ferns) form a monophyletic clade, sister to seed plants (**Figure 1**) (Kenrick and Crane, 1997; Pryer et al., 2001). Seed plants in turn, are comprised of two sister clades: Gymnosperms and Angiosperms (**Figure 1**). Within the Angiosperm lineage, monocots branched off from eudicot ancestors, creating two separate clades (**Figure 1**) (Chaw et al., 2004).

Despite the differences in classification, appearance and growth, all Tracheophytes, ranging from

Lycophytes to Angiosperms, have the same distinct cell types within their vascular bundles: xylem, phloem and (pro)cambium cells. Xylem transports water, nutrients and phytohormones, while phloem cells function as the distribution network of photosynthetic products, like sucrose, and of growth- and development signals such as proteins and mRNAs. Both cell types are produced via stem cells in the cambium region. In contrast to Tracheophytes, Bryophytes do not possess an elaborate vascular system containing xylem or phloem. Some early diverging land plants do however develop simple water-conducting cells (WCCs) and food conducting cells (FCCs) (see **Chapter 1**).
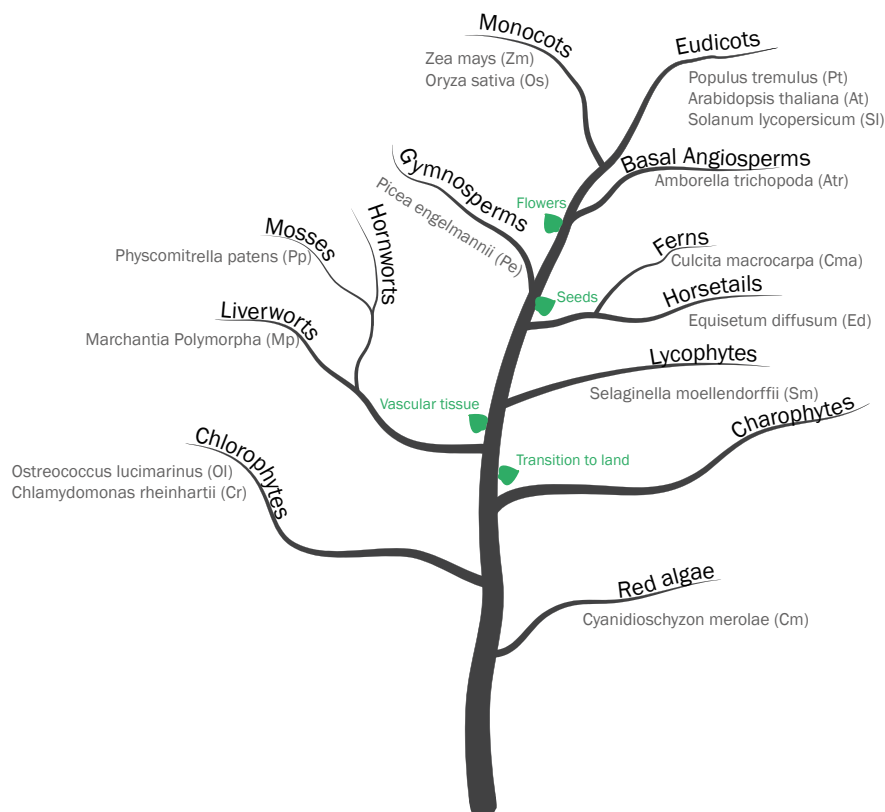
Figure 1. Schematic representation of the phylogenetic relationships in the plant lineage. Representative species used in step one and step two of this study are indicated. Key evolutionary events are marked in green.

Both functions of the vasculature, i.e. conducting water and solutes and providing mechanical support, are dependent on the width of the vascular bundle. Indeed, one fundamental difference in the conducting tissues between Bryophytes and Tracheophytes is its dimensions. Dimensions of plant tissues can be controlled by coordinated orientated cell divisions. Such proliferation of vascular cells may thus be a key link that led to elaborate vascular tissues. In Arabidopsis, this proliferation is controlled by an auxin flow that induces expression of the transcription factor AUXIN RESPONSE FACTOR 5 (ARF5), also known as MONOPTEROS (MP). One of its downstream targets, the basic-Helix-Loop-Helix (bHLH) protein TARGET OF MONOPTEROS 5 (TMO5) will form a transcriptional heterodimer complex with its bHLH partner LONESOME HIGHWAY (LHW) (De Rybel et al., 2013). Heterodimers of TMO5 and LHW, as well as their respective homologues, activate transcription of *LONELY GUY 4* (*LOG4*) and its most similar

homologue *LOG3*, encoding enzymes in the cytokinin biosynthesis pathway. Cytokinin is proposed to act as a mobile signal, triggering neighbouring procambium cells to initiate periclinal cell divisions (De Rybel et al., 2014; Ohashi-Ito et al., 2014). In addition, TMO5/LHW activity is controlled by a negative feedback loop involving SUPPRESSOR OF ACAULIS 51 (SAC51) and its three homologues (SAC51-LIKEs, or SACLs), which are also direct target genes of the TMO5/ LHW dimer. SACL proteins compete for LHW binding, thereby reducing activity of the TMO5/ LHW heterodimer and balancing radial growth.

Besides a report in Poplar (Immanen et al., 2016), our current knowledge of this TMO5/LHW dependent regulatory pathways underlying vascular development is mostly derived from the flowering plant *Arabidopsis thaliana.* It is unknown whether this regulatory pathway is universal in Tracheophytes or confined to a subset of plant lineages. In this chapter, we investigate the origin of the *TMO5* and *LHW* genes as a first step towards understanding the evolutionary trajectory of the dimer. We use bioinformatics tools to study whether the emergence of these genes is correlated with the emergence of complex vascular tissues.

Results

Data set selection

To test the hypothesis that emergence of TMO5 and LHW proteins have contributed to the establishment of complex vascular tissues during land plant evolution, we applied a phylogenomic strategy to identify orthologues of both genes in the green lineage. It should be noted that while the definition of functional orthologues is not trivial, we here define orthologous genes as homologous genes between different species, with a common evolutionary origin, not necessarily sharing the same function (Fitch, 1970).

We used full-length TMO5 and LHW Arabidopsis protein sequences as well as their paralogs, TMO5-LIKE1 to 4 (T5L1to T5L4) and LHW-LIKE1 to 3 (LL1 to LL3), as queries to search against transcriptome assemblies and annotated genome sequences of a representative group of plants ranging from Chlorophytes to Angiosperms (see details in Methods): the red unicellular alga *Cyanidioschyzon merolae (Cm)*, the green unicellular algae *Ostreococcus lucimarinus (Ol)* and *Chlamydomonas reinhardtii (Cr),* the liverwort *Marchantia polymorpha (Mp)*, the moss *Physcomitrella patens (Pp)*, the Lycophyte *Selaginella moellendorffii (Sm)*, the Pterophytes *Culcita macrocarpa (Cma)*, and *Equisetum diffusum (Ed)*, the Gymnosperm *Picea engelmannii (Pe)*, the basal Angiosperm *Amborella trichopoda (Atr)*, the Angiosperm monocots *Oryza sativa (Os)* and *Zea mays (Zm)* and the Angiosperm eudicots *Arabidopsis thaliana (At), Solanum lycopersicum (Sl) and Populus trichocarpa (Pt)* (**Figure 1**). The complete set of closest orthologues obtained comprised 602 and 525 protein sequences of TMO5 and LHW respectively. Because of the large-scale data set, we developed a

bioinformatical pipeline in which phylogenetic analyses were performed in two phases (**Figure 2**). First, a phylogenetic tree was computed from all selected amino acid sequences. Next, the clades containing our proteins of interest were isolated and a second full analysis was run using these amino acid sequences.
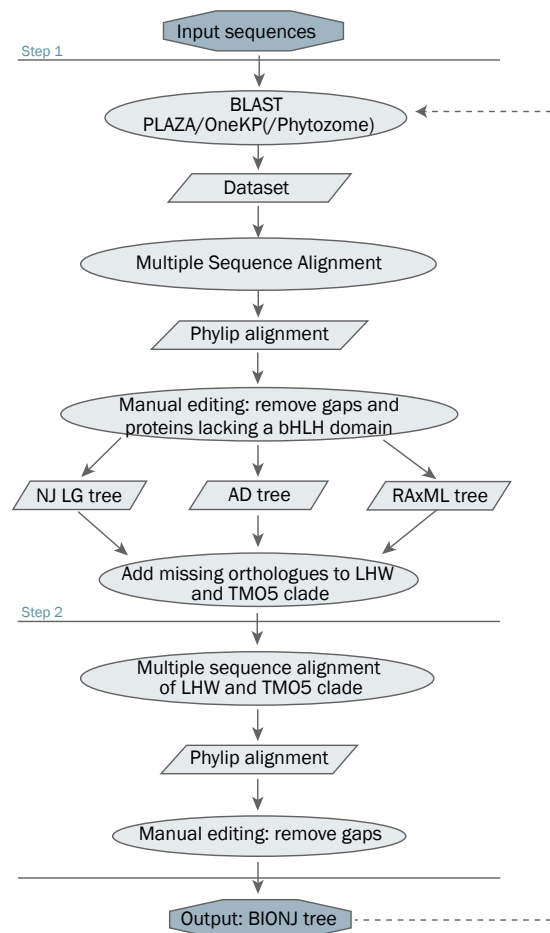
Figure 2. Schematic pipeline of the bioinformatical analyses performed in this chapter to identify TMO5 and LHW orthologues in the plant lineage. The input sequences (protein sequences of TMO5, LHW and their homolougs) were used as queries to BLAST against transcriptome assemblies and annotated genome sequences of plant species. Subsequently, Multiple Sequence Alignments (MSAs) were constructed and the phylip file was edited to remove gaps. Three phylogenetic trees were constructed and from the consensus TMO5 or LHW clade, phylogenetic relationships were estimated. The dotted line represents step 3: increase of taxon sampling.

Step 1: Phylogenetic reconstruction and validation

Following extensive BLAST searches, Multiple Sequence Alignments (MSAs) and elimination of candidates lacking any homology to the bHLH domain, the resulting 600 and 498 putative orthologues of respectively TMO5 and LHW were used for phylogenetic reconstruction. A Maximum likelihood (ML) phylogenetic tree of TMO5 classified the 600 potential orthologues in fourteen clades, representing different bHLH subfamilies (**Figure 3**). Using the Roman nomenclature as proposed by Pires and Dolan (Pires and Dolan, 2010), bHLH subfamilies were found to cluster

in a similar manner as described previously (Pires and Dolan, 2010; Toledo-Ortiz et al., 2003; Zhang et al., 2018). AtTMO5 and its homologues were clustered within the Vb subfamily together with 52 protein sequences of 10 different species (**Figure 3**). In contrast to TMO5 phylogeny, only four bHLH subfamilies could be distinguished within the LHW ML tree, belonging to bHLH subfamilies III(d-f), VIIIc, XI, and XIII (**Figure 4**). Despite our efforts to eliminate pentatricopeptide repeat and non-bHLH proteins (see Methods for details), two non-bHLH clades were formed (WNK and PPR/TPR). Remaining sequences appeared as orphan clades. As expected, all Arabidopsis LHW homologues clustered with the XIII subfamily. Moreover, no additional Arabidopsis proteins clustered within this family. In total, the XIII (LHW) clade contained 49 proteins of 11 different species. Overall, the high support clustering of bHLH subfamilies, similar
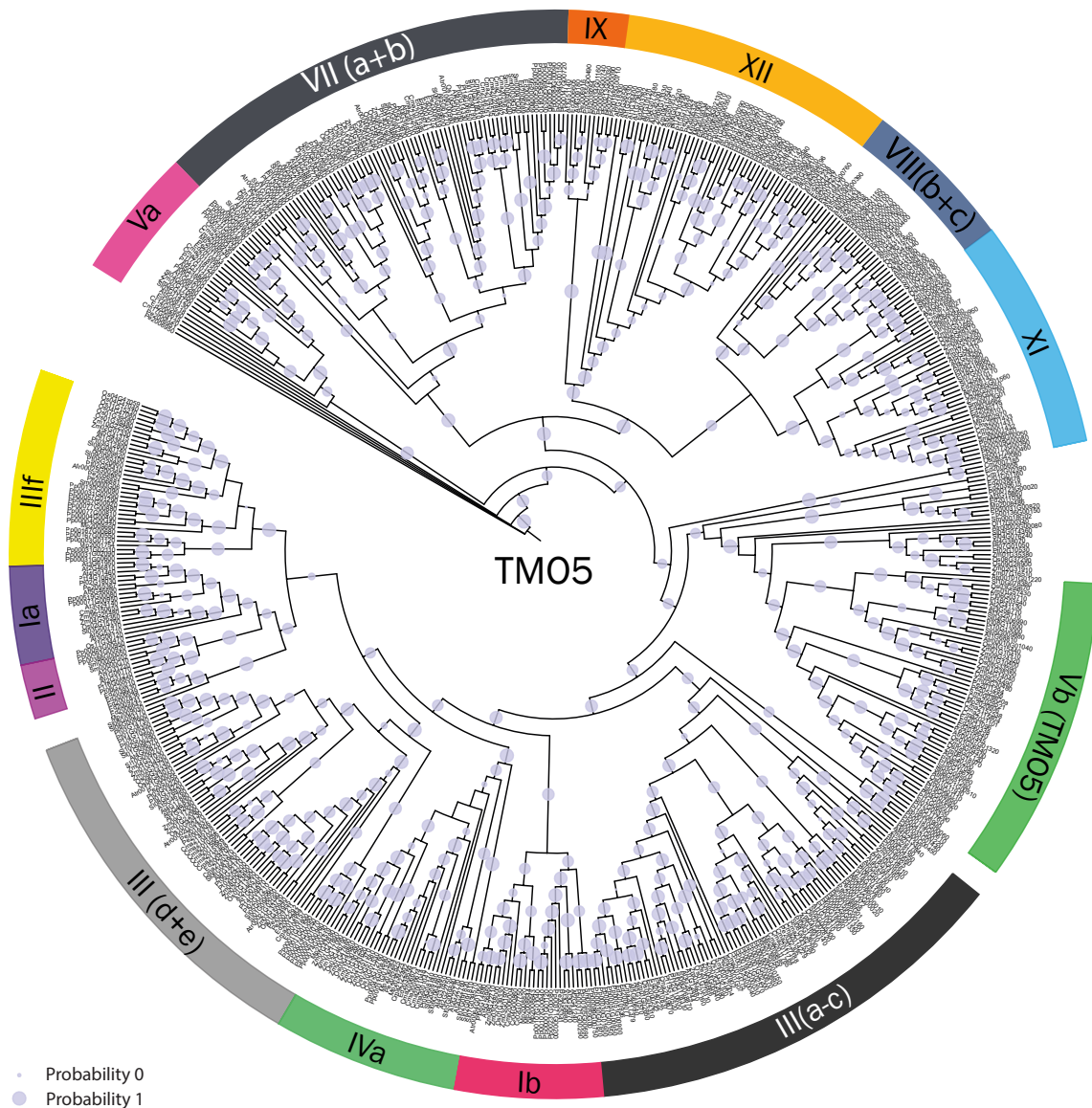
Figure 3. ML phylogenetic analysis of 600 proteins related to TMO5 and its homologues. The Roman numbers indicates the name of the bHLH subfamily, each marked with a seperate colour. TMO5 and its homologues cluster within subfamily Vb (green). Likelihood values are represented as blue dots, where the size of the dot is proportional to the likelihood value.

to literature, and the fact that no other Arabidopsis proteins clustered within the LHW or TMO5 clade besides the known direct homologues, support overall successful phylogenetic assembly.

In an attempt to further validate our phylogeny and assess robustness, Neighbour-joining (NJ), Average Distance (AD) analyses and Randomized Accelerated Maximum Likelihood (RAxML) analyses were performed for both LHW and TMO5 data sets. The TMO5 tree estimates produced from AD and RAxML analyses were generally concordant and largely consistent with the relationships reflected in the initial ML tree (**Supplemental Figure 1, Supplemental Table 1**). In contrast, LHW clades found by AD and RAxML analyses were more confined compared to the initial analysis (**Supplemental Figure 2, Supplemental Table 2**). Putative orthologues identified
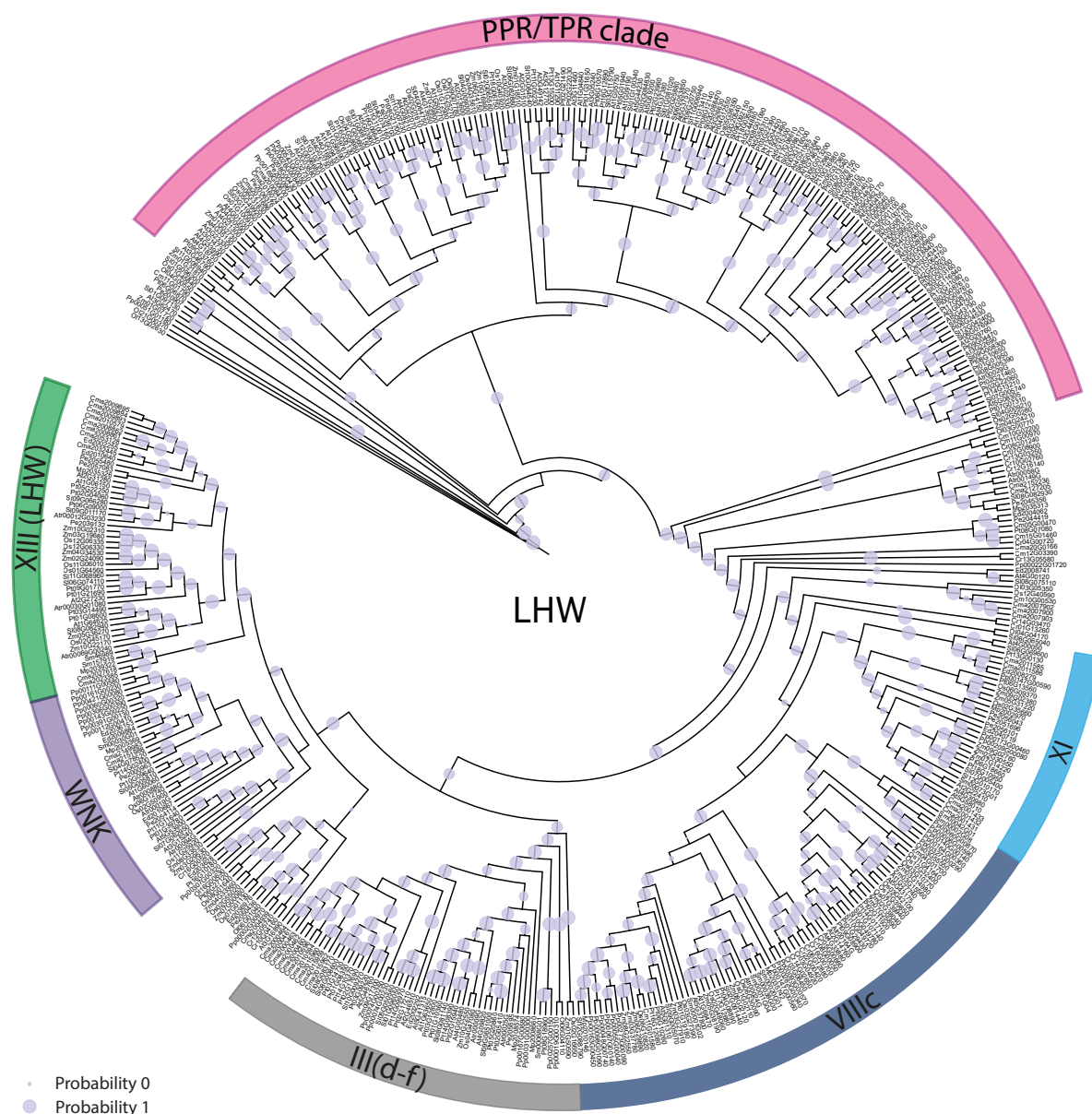
Figure 4. ML phylogenetic analysis of 498 proteins related to LHW and its homologues. The Roman numbers indicate the name of the bHLH subfamily, each marked with a seperate colour. Clades containing non-bHLH proteins are marked in purple (WNK) and pink (PPR/TPR). LHW and its homologues cluster within subfamily XII (green). Likelihood values are represented as blue dots, where the size of the dot is proportional to the likelihood value.

by both AD and RAxML, but not our initial analysis, were included in the sequence selection. This was the case for three TMO5 orthologue sequences, belonging to Solanum, Selaginella and Amborella (**Supplemental Figure 1, Supplemental Table 1**). No additional LHW orthologues were identified with this supplementary analysis (**Supplemental Figure 2, Supplemental Table 2**). As we are trying to answer evolutionary questions using a high amount of taxonomic sampling, variations between different analyses to some extent is to be expected. Indeed, four LHW orthologues identified in our study were not found back in the LHW clade for both AD and RAxML analyses. However, since experimental verification of potential orthologues will be performed during the next phase of this study, false positives are preferred over dismissal of a potential orthologue. As such, no sequences were omitted from our initial selection.
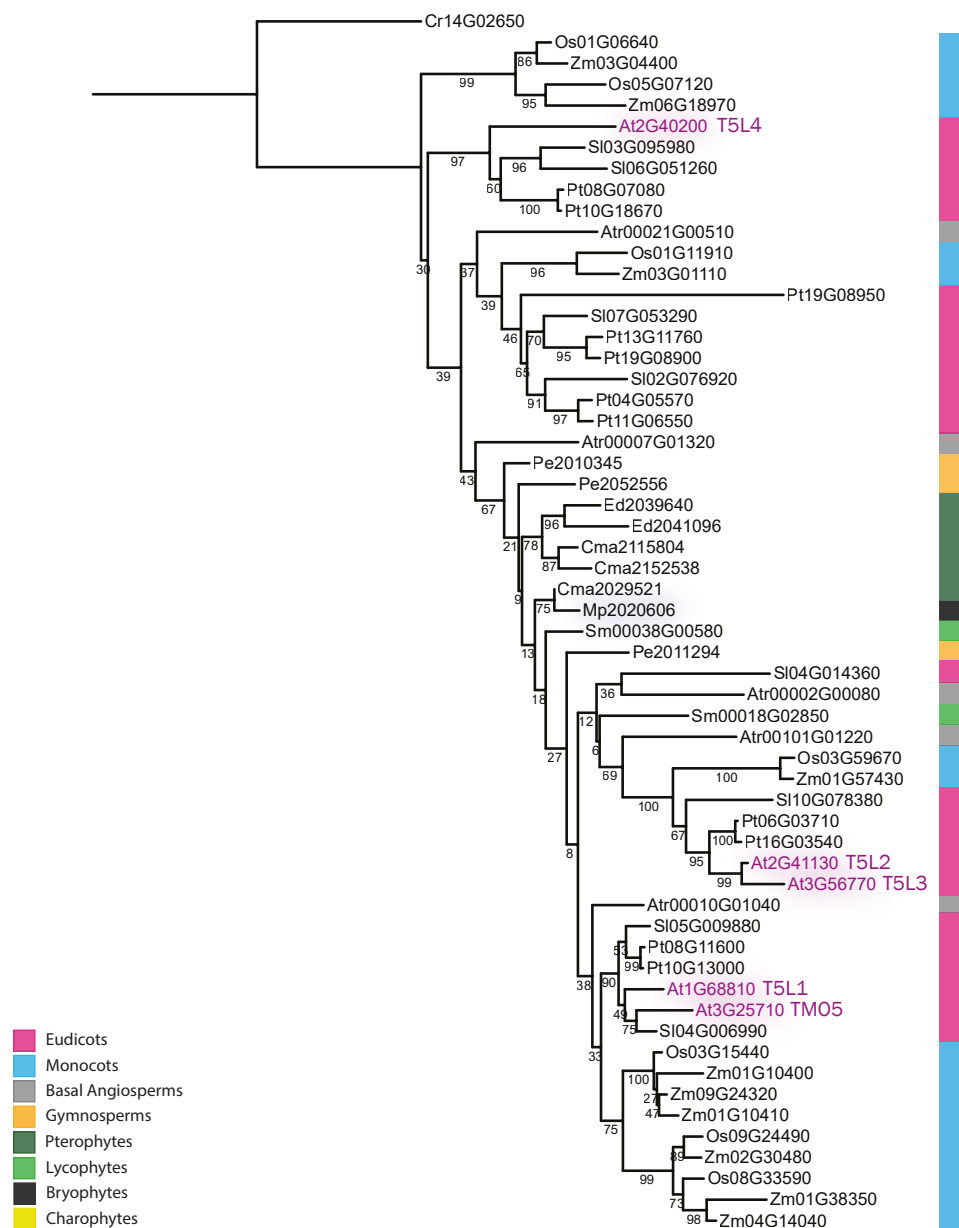
Figure 5. ML Phylogentic tree of direct TMO5 orthologues. Colours in the bar on the right represent different taxonomic groups.

Step 2: Phylogeny of TMO5 and LHW subfamilies

During the second phase of our two-step approach, all unedited protein hits within subfamilies Vb (TMO5) and XIII (LHW) were realigned. A single green algal representative, clustering outside of the LHW or TMO5 clade, was included as an outgroup to root the trees. After minimal manual editing, TMO5 (**Figure 5**) and LHW (**Figure 6**) specific ML phylogenetic trees were generated using 100 bootstrap replicas. Phylogenetic analyses show that TMO5 and LHW homologues form clades with dicots such as Solanum and Populus, whereas monocots (Oryza and Zea) formed aseparate sister clade. In agreement with literature (Moore et al., 2007; Soltis et al., 1999; Stefanović et al., 2004), genes belonging to the basal Angiosperm Amborella were generally placed as the sister group to the Angiosperms (**Figure 5, 6**). Together, this supports the reliability of the tree topologies. Strikingly, the phylogenetic analyses demonstrated that TMO5 and LHW orthologues are not solely present in Tracheophytes, but also in the Liverwort Marchantia, which does not possess any elaborate vascular tissues. Given that the other Bryophyte species included in this study (the moss Physcomitrella) does not have a TMO5 or LHW orthologue (**Figure 5, 6**), this posed the questioned
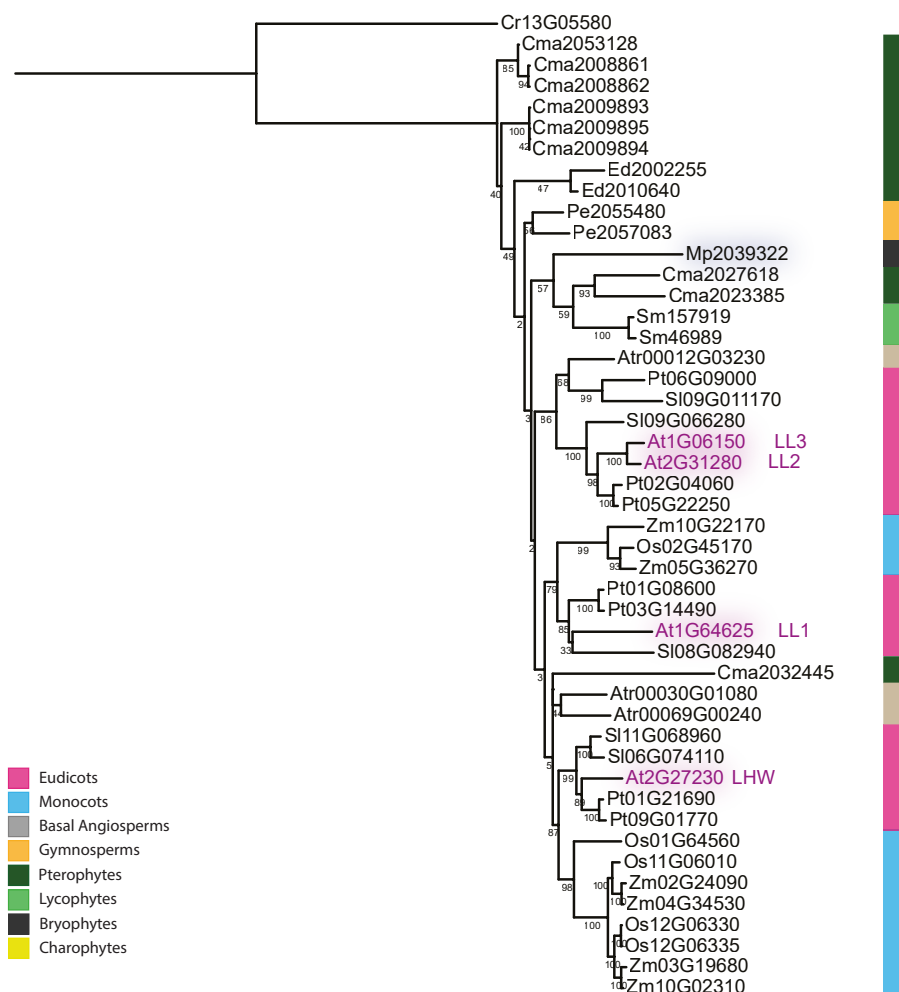
Figure 6. ML Phylogentic tree of direct LHW orthologues. Colours in the bar on the right represent different taxonomic groups.

whether the identified orthologues of Marchantia are false positives or if Physcomitrella, or mosses in general, are the exception within the Bryophytes. Physcomitrella is known for its high evolution rate and it could have lost its orthologues over time. From these results we conclude that the current taxon sampling is not sufficient to accurately pinpoint the emergence of *TMO5* and *LHW*. It does however give more insight into where sampling has to be increased.

Step 3: increase of taxon sampling

Ambiguity regarding presence of TMO5 and LHW orthologues occurred bordering the Bryophytes. Accordingly, a higher taxon sampling of hornworts, mosses and liverworts was performed, using all Bryophyte transcriptomes in the OneKP database (Methods, **Supplemental Table 5**). In order to achieve a clear boundary, all Lycophytes, the earliest diverging vascular land plants, as well as all green algae of the same database were included (Methods, **Supplemental Table 5**). Since Charophytes are thought to be the closest relatives of the Embryophytes, we included the annotated genome of the Charophyte *Klebsormidium nitens*. Given the large database size, the BLAST searches for potential orthologues were performed less extensively, i.e. only AtTMO5 and AtLHW, not their homologues, were used as query in the search and solely the top hits were selected. This resulted in 91 and 40 additional TMO5 and LHW hits respectively. To identify which of these belong to subfamilies Vb (TMO5) and XIII (LHW), the complete MSA and phylogenetic analyses were repeated with 693 and 565 sequences of TMO5 (**Figure 7**) and LHW (**Figure 8**) respectively. The same bHLH subfamilies were distinguished for TMO5 (**Figure 5, 7**) and LHW (**Figure 6, 8**) compared to the initial phylogenetic estimation. Likewise, alternative methods of phylogenetic tree construction (AD and RAxML), as means of verification, gave highly similar outputs for TMO5 and a somewhat more confined output for LHW (**Supplemental Figure 3, 4, Supplemental Table 3, 4**). Using the same majority ruling as before, we determined that no proteins were lacking from the TMO5 or LHW clade.

We next realigned the isolated full-length protein sequences of the TMO5 and LHW clades (**Supplemental Figure 5, 6**) and constructed the final TMO5 and LHW phylogenetic trees, containing direct orthologues (**Figure 9, 10**). The LHW ML tree was largely in concordance with the generally accepted evolutionary history. For example, within the LHW subfamily, a clade of Charophytes branches off first, followed by a monophylectic clade of Bryophytes (**Figure 10**). A large Lycophyte family formed a sister clade to all other Tracheophytes. In the latter, Pterophytes branch off creating a separate Gymnosperm/Angiosperm family in which mono- and eudicots form monophylectic clades. While phylogenetic relationships within the TMO5 tree were less accurate, monocots and eudicots did cluster as monophylectic clades, sister to basal Angiosperms (**Figure 9**). Furthermore, all Lycophytes were grouped together in a single clade. Together, this supports the reliability of the phylogenetic estimations.

2
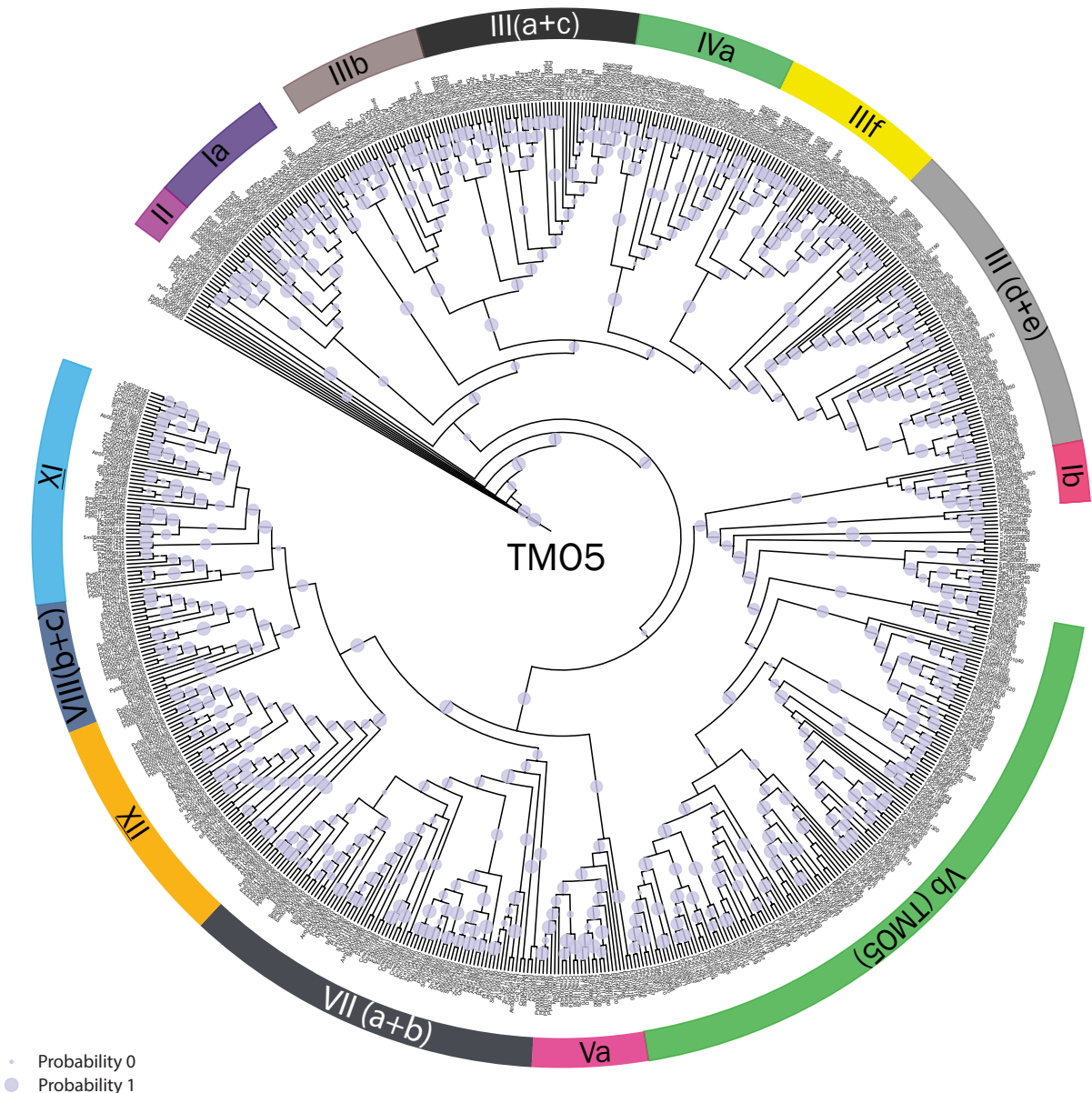
Figure 7. ML Phylogenetic tree of TMO5 and related proteins with increased taxon sampling. The Roman numbers indicates the name of the bHLH subfamily, each marked with a seperate colour. TMO5 and its homologues cluster within subfamily Vb (green). Likelihood values are represented as blue dots, where the size of the dot is proportional to the likelihood value.

Step 4: Identification of TMO5 and LHW orthologues

The in-depth analyses contained a wide range of plant species including a high taxon sampling of Lycophytes, Bryophytes, Charophytes and Chlorophytes, derived from a large transcriptome data set and a set of genome-based sequences. Strikingly, the phylogenetic analyses demonstrated that TMO5 and LHW orthologues are present in all major plant lineages including Charophytes, but not Chlorophytes (**Figure 9, 10, Supplemental Table 5**). Given the wide range of Chlorophyte species included in our analysis, we can state with high certainty that *TMO5* and *LHW* genes did not emerge prior to the split of Streptophytes from their ancestor with extant Chlorophytes.
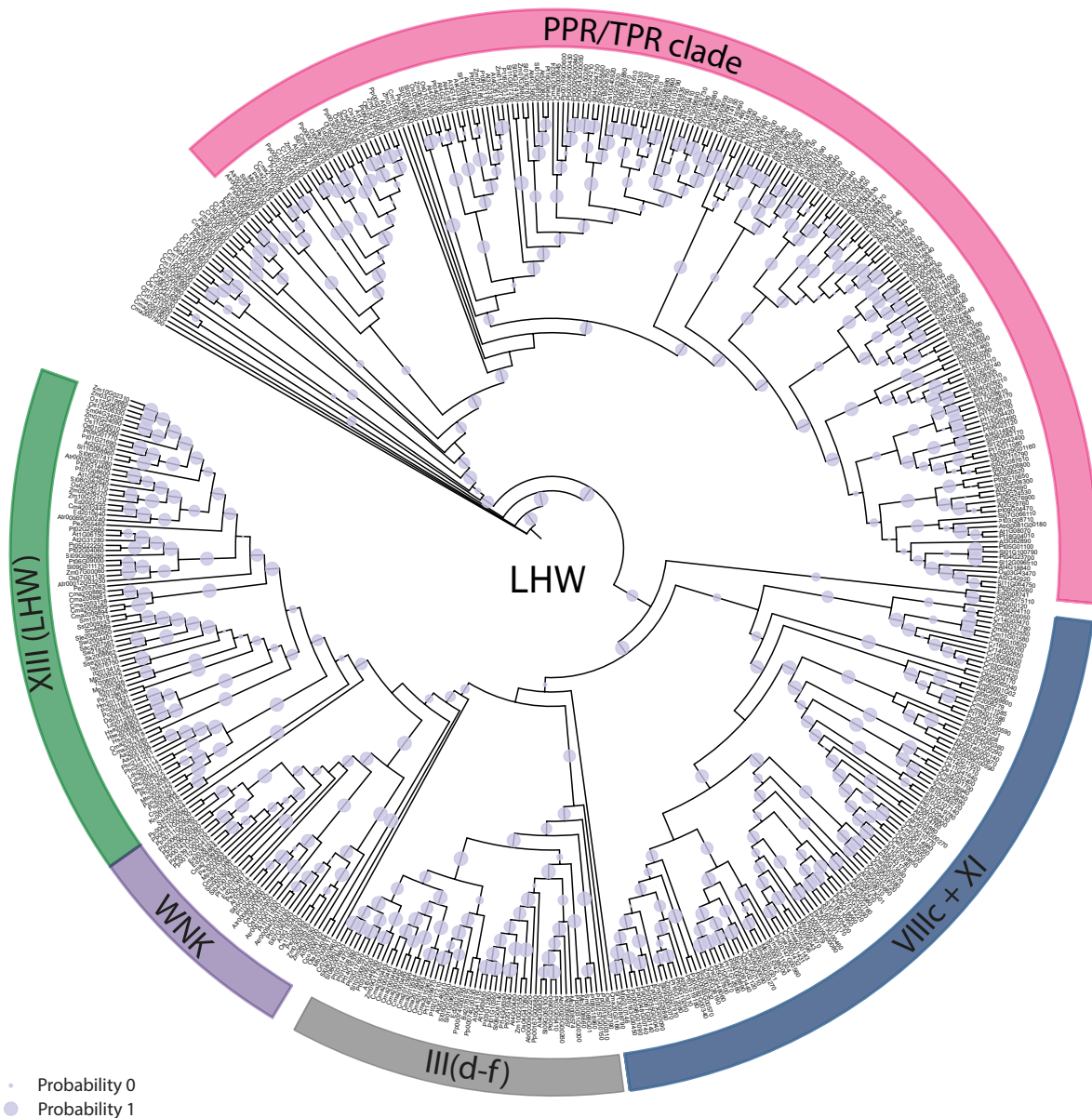
Figure 8. ML Phylogenetic tree of LHW and related proteins with increased taxon sampling. The Roman numbers indicates the name of the bHLH subfamily, each marked with a seperate colour. LHW and its homologues cluster within subfamily XIII (green). Likelihood values are represented as blue dots, where the size of the dot is proportional to the likelihood value.

We further observed that most Charophytes and Bryophytes contain only a single copy of each gene, whereas the number of *TMO5* and *LHW* orthologues increased in the Tracheophyte lineage (**Figure 11, Supplemental Table 5**). For example, Klebsormidium, and early diverging land plants such as Marchantia, Takakia and some Selaginella species contain a single copy of the *TMO5* and *LHW* genes, whereas the monocot Oryza contains seven and six respectively. In summary, we show that the TMO5 and LHW genes did not emerge with the establishment of vascular tissues during land plant evolution, but have a more ancient origin.
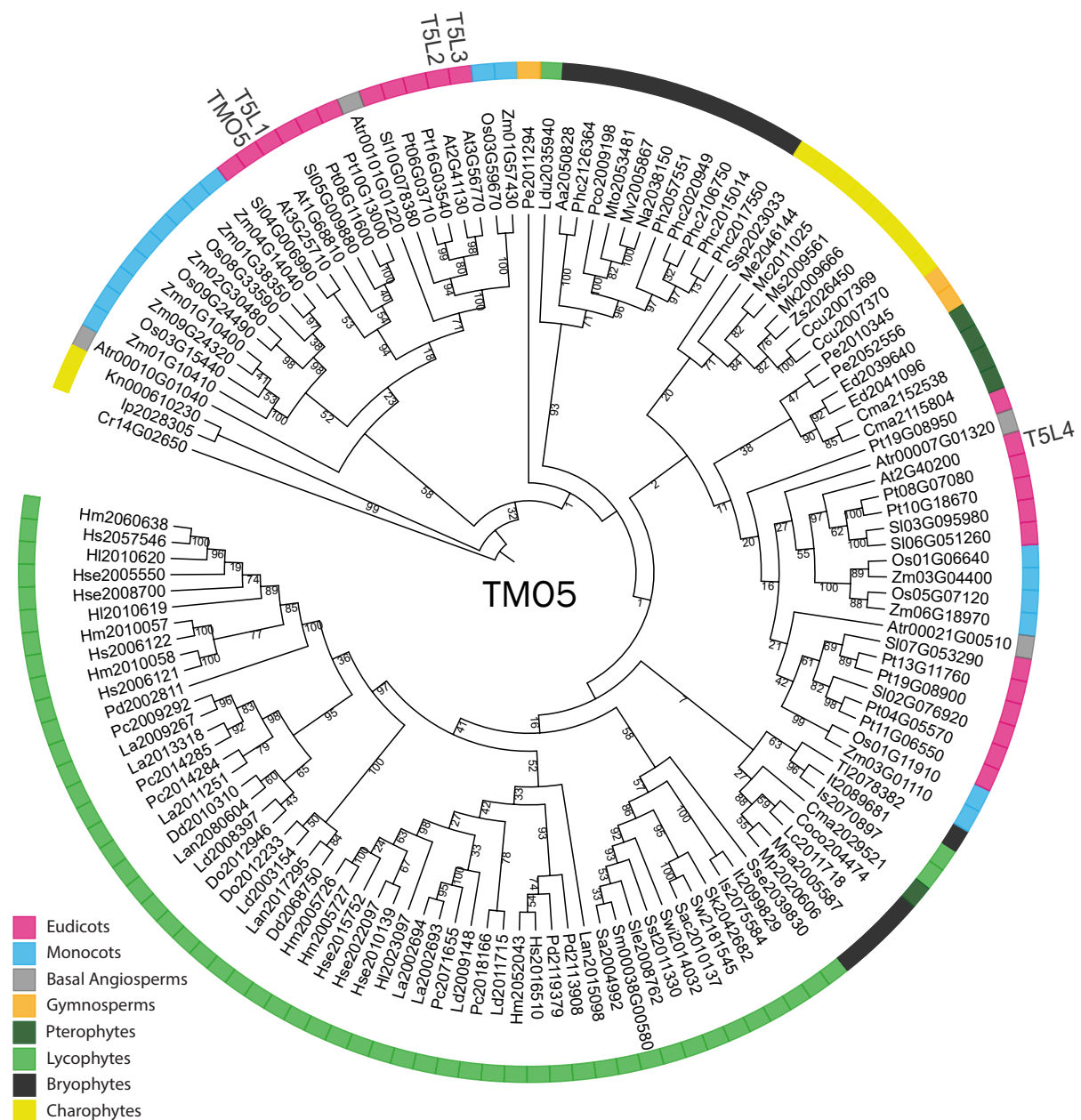
Figure 9. ML Phylogenetic tree of the TMO5 clade, as determined by the analysis presented in Figure 5. Colours in the ring represent different taxonomic groups. TMO5 orthologues are encoded in the genome of Tracheophytes as well as Bryophytes (black) and Charophytes (yellow). Bootstap values (n=100) are represented by numbers. The Multiple Sequence Alignment (MSA) used to calculate the phylogeny is shown in Supplemental Figure 5. Note that Mp2020606 is also known as Mapoly0039s0068.

Identification of LOG4 orthologues in Marchantia and Klebsormidium

The finding that potential TMO5 and LHW orthologues are present in Bryophytes and Charophytes, raises the question of whether the whole cytokinin pathway is conserved in these early diverging organisms. Here we focussed on the liverwort Marchantia and the Charophyte Klebsormidium, since they both contain a copy of the *TMO5* and *LHW* gene and are evolutionary positioned around two important events: the transition to land and the gain of vascular tissues (**Figure 1**). In addition, both species have well-annotated genomes. If the pathway is indeed conserved, MpTMO5/

Figure 10. ML Phylogenetic tree of the LHW clade, as determined by the analysis presented in Figure 5. Colours in the ring represent different taxonomic groups. LHW orthologues are encoded in the genome of Tracheophytes as well as Bryophytes (black) and Charophytes (yellow). Bootstap values (n=100) are represented by numbers. The Multiple Sequence Alignment (MSA) used to calculate the phylogeny is shown in Supplemental Figure 5. Note that Mp2039322 is also known as Mapoly0088s0049.

MpLHW and KnTMO5/KnLHW dimers will activate transcription of a downstream target involved in cytokinin biosynthesis. We tried to determine whether orthologues of the AtTMO5/AtLHW dimer targets are present in Marchantia and Klebsormidium. We used the full-length protein sequence of the AtLHW/AtTMO5 downstream targets LOG3 and LOG4 to search against the genome sequences of Marchantia and Klebsormidium. A single LOG4 orthologue (Mp0103s0059) was identified in Marchantia, whereas two orthologues (Kn006470030 and Kn000870020) were found in the Klebsormidium genome. No additional hits were found using

AtLOG3 as query. Sequence identity of Mp0103s0059, Kn006470030 and Kn000870020 to AtLOG4 were 65%, 54% and 23% respectively. When blasting LOG4 orthologues found against the Arabidopsis genome, either AtLOG4 or AtLOG3 was the first (Kn006470030; Mp0103s0059) or second (Kn000870020) hit in Phytozome. Moreover, when searching the Marchantia and Klebsormidium genomes using Mp0103s0059 or Kn006470030/Kn000870020 as query respectively, no additional hits were obtained.

Next, AtLOG3, AtLOG4, Mp0103s0059, Kn006470030, Kn000870020 and At1G50575 were aligned and a phylogenetic tree was estimated (**Figure 12A**). The strong bootstrap supported phylogenetic tree revealed a clear LOG clade, containing AtLOG3 and AtLOG4. Mp0103s0059 and Kn006470030 clustered within this clade, confirming their homology to AtLOGs. As such, these will henceforth be referred to as MpLOG and KnLOG.

Taken together, our results indicate that TMO5 and LHW as well as their downstream target LOG4, may predate the rise of Tracheophytes and even Emybrophytes in general. As such, part of the potential of land plant development may have been present as early as in Charophytes. Whether these individual components were already acting in the same pathway remains to be studied.

		TMO5	LHW
	Arabidopsis thalia	5	4
	Populus tremulus	11	8
	Solanum lycopersicum	7	5
	Oryza sativa	7	6
	Zea Mays	10	7
	Amborella trichopoda	4	3
	Picea (engelmannii / abies)	3	2
	Culcita macrocarpa	3	8
	Equisetum diffusum	2	2
	Huperzia selago	5	2
	Selaginella moellendorffii	1	2
	Megaceros tosanus	1	1
	Leiosporoceros dussii	1	2
	Takakia lepidozioides	1	1
	Marchantia polymorpha	1	1
	Klebsormidium nitens	1	1
	Cylindrocystis cushleckae	1	1
	Chlamydomonas reinhardtii	-	-
	Ostreococcus lucimarinus	-	-
	Cyanidioschyzon merolae	-	-

Figure 11. Phylogenetic relationships of several representative species used in this study and the number of TMO5 and LHW orthologues identified in these species. See Supplemental table 5 for an overview of all species and the amount of TMO5 and LHW orthologues.

Identification of SACL orthologues in Marchantia and Klebsormidium

A fourth important regulator within the TMO5/LHW pathway is the SACL family, composed of SAC51 and three SACL bHLH proteins. *SACL* genes, activated by the TMO5/LHW dimer, are

part of the negative feedback loop, restricting TMO5/LHW activity by forming heterodimers with the LHW protein. Given that TMO5, LHW and LOG3/LOG4 all seem to have orthologues in the early diverging land plant Marchantia and the Charophyte Klebsormidium, we next wondered if SACL proteins, and thus the negative feedback loop, is conserved as well. We searched the Marchantia genome sequences using all SACL proteins (SACL51, SACL1, SACL2 and SACL3) as query and the top three hits were selected. This resulted in four unique sequences: Mp0048s0099, Mp0039s0059, Mp0024s0106 an Mp0073s0051. In the same manner, we obtained four sequences for Klebsormidium: Kn000120300, Kn003550070, Kn000120330 and Kn000860300. A bidirectional BLAST revealed ROOT HAIR DEFECTIVE 6-LIKE 2 (RSL2)/ RSL4, PHYTOCHROME INTERACTING FACTOR 3 (PIF3) and HECATE 1 (HEC1) as the most similar Arabidopsis homologues for the respective Marchantia proteins (HEC1 being the closest hit for both Mp0024s0106 and Mp0073s0051). Lotus japonicus ROOTHAIRLESS1-LIKE 3 (LRL3), BES1-INTERACTING MYC-LIKE PROTEIN 2 (BIM2), LRL1 and PIF3 were found for the respective Klebsormidium proteins. To identify if any of the identified proteins clusters with AtSACL proteins, a ML phylogenetic tree was calculated using a single green alga bHLH protein as outgroup (**Figure 12B**). This analysis identified a single Marchantia gene orthologous to AtSACL proteins: Mp0048s0099 and will henceforth be called MpSACL. MpSACL clearly clustered within the SACL clade rather than with its top blast hits: RSL4 and RSL2. None of the four Klebsormidium proteins clustered within the SACL clade. Given the fact that the total number of Klebsormidium

Figure 12. ML Phylogenetic tree of the LOG and SACL orthologues. (A) Phylogenetic analysis of *LOG* related genes in Arabidopsis, Marchantia and Klebsormidium. A Charophytic gene was used as an outgroup. The clade containing Arabisopsis LOG3 and LOG4 is marked in blue. The numbers represent bootstrap values (n=100). **(B)** Phylogenetic relationship of SACL related genes in Arabidopsis, Marchantia and Klebsormidium. A Charophytic gene was used as an outgroup. The clade containing Arabisopsis SACLs is marked in blue. The numbers represent bootstrap values (n=100). The Roman numbers indicates the name of the bHLH subfamily.

proteins identified by BLAST was four, it rules out the possibility of under sampling. We conclude that the regulatory negative feedback loop is not conserved in Charophytes, since SACL proteins emerged after Charophytes diverged from the common ancestor.

Discussion

The heterodimeric TMO5/LHW transcriptional complex is necessary and sufficient for radial root growth in Arabidopsis (De Rybel et al., 2014). Given that this pathway is restricted to the vascular bundle, we questioned if, and to what extent, the emergence of the dimer contributed to evolution of complex vascular tissues. The novel OneKP data set containing transcriptomes of over 1,000 plants (Matasci et al., 2014), now allows in-depth evolutionary studies with early diverging plant species. Here, we used a bioinformatical approach to identify orthologues of TMO5 and LHW throughout the plant kingdom as a first step towards uncovering the role of the TMO5/LHW dimer in facilitating the establishment of complex vascular tissues.

The identification of TMO5 and LHW orthologues in all studied vascular plants, supports the notion of a conserved genetic network for vascular tissue proliferation by cytokinin biosynthesis through TMO5 and LHW, in all vascular plant lineages. Intriguingly, our data shows that TMO5 and LHW orthologues are not restricted to Tracheophytes, but have earlier evolutionary origins. Indeed, orthologues of both genes were identified in Bryophytes and even in the Charophytes Klebsormidium and Cylindrocystis. We observed that most Charophytes and Bryophytes contain only a single copy of each gene, whereas the number of TMO5 and LHW orthologues increased in the tracheophyte lineage (**Supplemental Table 5**). Our data suggest that a single *TMO5* and *LHW* gene were present in the Charophytic ancestor and was retained in Bryophytes, Lycophytes, Pterophytes and Gymnosperms. Phylogenetic clustering indicated that extra copies found in some of these species are most likely the result of more recent duplications. Consistent with the whole-genome duplication (WGD) event in the ancestor of *Amborella* (Amborella Genome Project, 2013), duplication in the ancestor of the basal Angiosperms resulted in four and three copies of *TMO5* and *LHW* in Amborella respectively. A final division, resulting in TMO5, T5L1, T5L2, T5L3 and LHW, LL1, LL2 and LL3 (following Arabidopsis nomenclature), most likely occurred within the Angiosperms.

It is striking that TMO5 and LHW orthologues are found in Charophytes and early diverging land plants, especially because significant bHLH protein expansion did not occur prior to the emergence of Embryophytes (Carretero-paulet et al., 2010; Pires and Dolan, 2010). Indeed, while 98 – 173 bHLH proteins are encoded within Embryophyte species (ranging from Physcomitrella to Oryza), Chlorophytes contain no more than four bHLH proteins (Carretero-paulet et al., 2010; Pires and Dolan, 2010). The Charophyte Klebsormidium encodes ten bHLH proteins (Jin

et al., 2017; Pires and Dolan, 2010), suggesting that TMO5 and LHW were among the first bHLH proteins established in the plant lineage. However, identification of two Arabidopsis bHLH orthologues (TMO5 and LHW) amongst ten Klebsormidium bHLHs should be interpreted with caution. Perhaps, we selected the Klebsormidium precursor bHLHs of which TMO5 and LHW evolved after gene duplications. Indeed, unlike many other genes, duplications of transcription factor genes are often retained within the genome (Edger and Pires, 2009), leading to division of existing functions (subfunctionalization) or the gain of a new function (neofunctionalization). Nevertheless, these Klebsormidium genes did cluster closer to TMO5 or LHW than any other Arabidopsis bHLH protein included in our phylogeny study, making them promising genes for future studies.

Given that both dimer components have evolutionary origins in ancestors of Bryophytes and Charophytes, this poses the question whether the genetic network, in which cytokinin is produced through TMO5 and LHW, is conserved in these early diverging species. If true, ancestral TMO5 and LHW should contain the ability to act as a transcriptional complex as heterodimer, activating cytokinin biosynthesis. Several Bryophytes as well as Klebsormidium were shown to produce low levels of cytokinin and other plant hormones (Drábková et al., 2015; Hori et al., 2014; Stirk et al., 2013). Moreover, cytokinin hormone receptors are present in Klebsormidium and Marchantia (Gruhn et al., 2014; Hori et al., 2014). Given that our data revealed a LOG orthologue, responsible for cytokinin biosynthesis in Arabidopsis, in Klebsormidium and Marchantia genomes, one could thus postulate that TMO5 and LHW are ancient cytokinin biosynthesis regulators that may have evolved in the common ancestor of Charophytes and Embryophytes. In contrast, the TMO5/LHW repressor SACL, is encoded by Marchantia, but not Klebsormidium. As such, one could speculate that the ancestral pathway recruited *SACL* after Embryophytes split from their ancestor. An alternative hypothesis is that ancestral TMO5 and LHW target, either collectively or independently, one or more different targets and regulate different pathways than in Arabidopsis.

Surprisingly, within the Bryophytes, TMO5 or LHW orthologues are not found in mosses other than Takakia, despite the fact that mosses are the most species-rich within the Bryophytes and are well represented in the OneKp database. Since Takakia is a separate moss family that forms a sister clade to all other mosses (Chang and Graham, 2011; Liu et al., 2019; Newton et al., 2000; Volkmar and Knoop, 2010), this suggests that the ancestral *TMO5* and *LHW* genes were lost in mosses, after they split from Takakia. Therefore, the function of TMO5 and LHW in Takakia, and potentially other Bryophytes, is not required in other mosses, or is replaced by other genes. In addition, eighteen species, belonging to Lycophytes, Bryophytes and Charophytes, were shown to contain a TMO5, but no LHW orthologue (**Supplemental Table 5**). It should, however, be noted that transcriptome data derived from the OneKP database was obtained by only a single round

of sequencing, using one specific tissue type per sample. As a consequence, transcripts may have limited quality. In addition, it is possible that the gene is simply not transcribed in the sampled tissue, resulting in false negatives. Alternatively, the lack of LHW could be the result of gene loss or the single TMO5 hits could be false positives. Indeed, the parameters set in our phylogenetic analyses were not very stringent to ensure identification of all potential orthologues. We only set one restriction after our initial analysis, namely the presence of a bHLH domain. As the objective of this study is to identify orthologues rather than resolve the exact phylogenetic relationship, false positives were favoured over false negatives. However, additional conditions, for examples restrictions based on domain conservations, could filter out false positives.

In summary, we aimed to get more insight in the evolutionary conservation of the TMO5/LHW cytokinin biosynthesis pathway and whether this is correlated to the emergence of vascular plants. Here we showed that three of the cytokinin biosynthesis/periclinal and radial cell division pathway components, namely TMO5, LHW and LOG may have been present in non-vascular land plants and Charophytes. This suggests that the presence of TMO5 and LHW bHLH proteins is not directly correlated with increase in vascular tissue complexity. It will be exciting to study the co-operation of ancestral *TMO5* and *LHW* within a common pathway as well as the functional conservation of these genes.

Methods

Sequence data

Genomic sequences of plant and algae species *Cyanidioschyzon merolae*, *Ostreococcus lucimarinus*, *Chlamydomonas reinhardtii*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Amborella trichopoda*, *Oryza sativa*, *Zea mays*, *Arabidopsis thaliana*, *Solanum lycopersicum and Populus tremulus* were retrieved from the comparative genomics database PLAZA 3.0, PLAZA 2.5 and Pico-PLAZA (Proost et al., 2015; Vandepoele et al., 2013; http://bioinformatics.psb.ugent.be/plaza). Genomic data of *Marchantia polymorpha* was accessed via Phytozome v12 (https://phytozome.jgi.doe.gov/pz/portal.html). *Klebsormidium nitens* genomic sequences were collected from the *Klebsormidium nitens* NIES-2285 genome project (Hori et al., 2014; http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/). Transcriptomic data of *Culcita macrocarpa*, *Equisetum diffusum, Picea engelmanni, Marchantia polymorpha* and a wide range of Lycophytes, Bryophytes and algae were derived from the transcriptome database OneKP, to which access was kindly provided by the developers of OneKP (Matasci et al., 2014). A list of all species included can be found at: http://www.onekp.com/samples/list.php. For extensive LHW and TMO5 analyses full length protein sequences of TMO5, TMO5-LIKE1 (T5L1), TMO5-LIKE2 (T5L2), TMO5-

LIKE3 (T5L3), TMO5-LIKE4 (T5L4), LHW, LHW-LIKE1 (LL1) and LHW-LIKE2 (LL2) were used. PPR regions of LHW-LIKE3 (LL3) were excluded from the search query in order to reduce PPR domain-derived superfamily hits. Transcriptome assembly outputs resulting from tBLASTn searches were translated into protein sequences prior to alignments.

Sequence alignments

Full-length proteins were aligned using MAFFT version 7 (Katoh et al., 2002; http://mafft.cbrc.jp/alignment/server/), with default parameters. Gaps at positions containing less than 20% (for large MSA) or 12% (for TMO5 and LHW clades only) overall conservation were manually removed using BioEdit v7.2.5 (Hall, 1999). Sequences lacking any homology to a bHLH domain were removed from the MSA. MSA in the figures were visualized by Jalview v2.10.2 using clustal based colour coding.

Phylogenetic tree construction

ML phylogenetic trees were estimated using the program PhyML version 3.0 (Guindon et al., 2010; http://www.atgc-montpellier.fr/phyml/), applying the LG amino-acid replacement matrix using 20 substitution rate categories and an estimated gamma distribution parameter. A Nearest Neighbour Interchange (NNI) topology improvement was applied. For phylogenetic reconstructions, a ML tree was calculated with an Approximate Likelihood-Ratio Test (aLRT). For confined TMO5 and LHW trees 100 bootstrap replicas were calculated. RAxML verification analyses using an LG substitution matrix were performed on the RAxML BlackBox web server using default settings (Stamatakis et al., 2008; http://www.genome.jp/tools/raxml/). AD verification trees, applying BLOSUM62 weighing matrices, were estimated by Jalview v2.10.2. Phylogenetic trees were visualized using the program iTol v3.6 (Letunic and Bork, 2016; https://itol.embl.de/).

Acknowledgments

Supplemental information

A

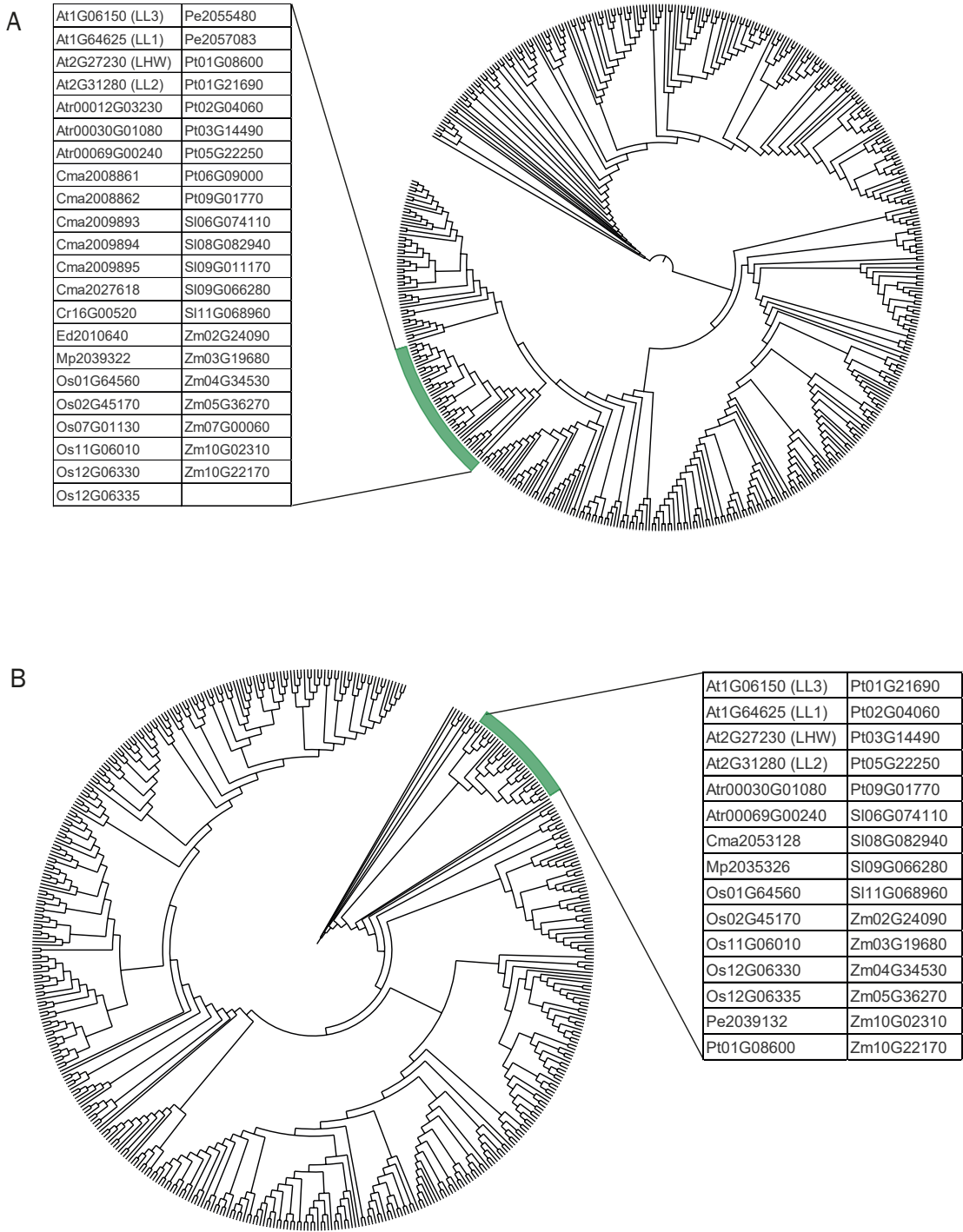|	
At1G68810 (T5L1)	Pt06G03710
At2G40200 (T5L4)	Pt08G07080
At2G41130 (T5L2)	Pt08G11600
At3G25710 (TMO5)	Pt10G13000
At3G56770 (T5L3)	Pt10G18670
Atr00007G01320	Pt11G06550
Atr00010G01040	Pt13G11760
Atr00101G01220	Pt16G03540
Atr00002G00080	Pt19G08900
Atr00021G00510	SL02G076920
Cma2152538	SL03G095980
Ed2039640	SL04G006990
Ed2040338	SL04G014360
Ed2041096	SL05G009880
Mp2020606	SL06G051260
Os01G11910	SL07G053290
Os03G15440	SL10G078380
Os03G59670	Sm00018G02850
Os08G33590	Sm00038G00580
Os08G37290	Zm01G35380
Os09G24490	Zm01G57430
Os09G28900	Zm02G30480
Pe2010345	Zm03G01110
Pe2011294	Zm04G14040
Pe2052556	Zm07G16510
Pt04G05570	Zm09G24320

B

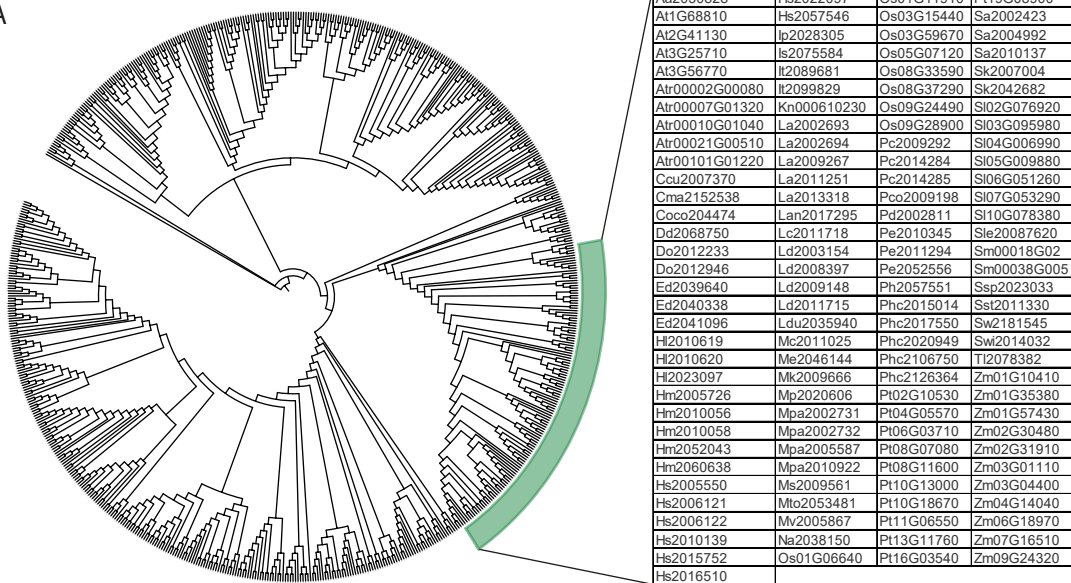|	
At1G68810 (T5L1)	Pt04G05570
At2G40200 (T5L4)	Pt06G03710
At2G41130 (T5L2)	Pt07G01050
At3G25710 (TMO5)	Pt08G07080
At3G56770 (T5L3)	Pt08G11600
AT4G38070	Pt10G13000
Atr00007G01320	Pt10G18670
Atr00010G01040	Pt11G06550
Atr00101G01220	Pt13G11760
Atr00002G00080	Pt16G03540
Atr00021G00510	Pt19G08900
Cma2021101	Pt19G08950
Cma2029521	SL02G076920
Cma2115804	SL03G095980
Cma2152538	SL04G006990
Ed2039640	SL04G014360
Ed2040338	SL04G076240
Ed2041096	SL05G009880
Mp2005672	SL06G051260
Mp2020606	SL07G053290
Ol17G00340	SL10G078380
Os01G06640	Sm00018G02850
Os01G11910	Sm00038G00580
Os03G15440	Zm01G10400
Os03G59670	Zm01G10410
Os05G07120	Zm01G38350
Os08G33590	Zm01G57430
Os09G24490	Zm02G30480
Pe2010345	Zm03G01110
Pe2011294	Zm03G04400
Pe2024173	Zm04G14040
Pe2052556	Zm06G18970
Pt02G10530	Zm09G24320

Supplemental Figure 1. Verification of TMO5 phylogeny using AD (A) and RAxML (B) phylogenies. The TMO5 clade is indicated in green. Genes identified in both AD and RAxML analyses, but not in the initial ML calculation are marked in orange.
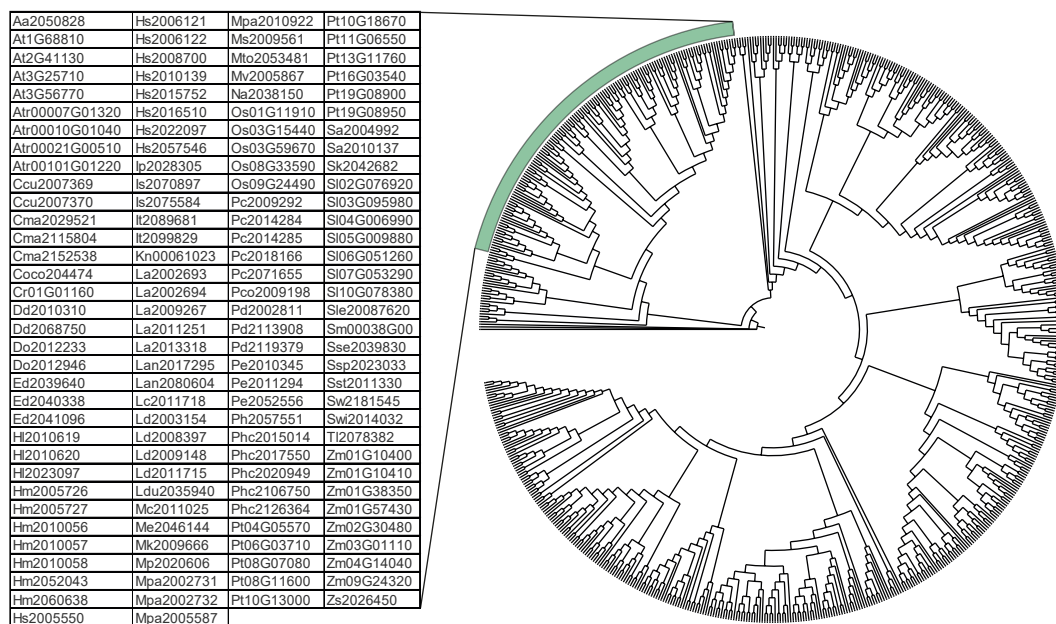
A

At1G06150 (LL3)	Pe2055480
At1G64625 (LL1)	Pe2057083
At2G27230 (LHW)	Pt01G08600
At2G31280 (LL2)	Pt01G21690
Atr00012G03230	Pt02G04060
Atr00030G01080	Pt03G14490
Atr00069G00240	Pt05G22250
Cma2008861	Pt06G09000
Cma2008862	Pt09G01770
Cma2009893	Sl06G074110
Cma2009894	Sl08G082940
Cma2009895	Sl09G011170
Cma2027618	Sl09G066280
Cr16G00520	Sl11G068960
Ed2010640	Zm02G24090
Mp2039322	Zm03G19680
Os01G64560	Zm04G34530
Os02G45170	Zm05G36270
Os07G01130	Zm07G00060
Os11G06010	Zm10G02310
Os12G06330	Zm10G22170
Os12G06335	

B

At1G06150 (LL3)	Pt01G21690
At1G64625 (LL1)	Pt02G04060
At2G27230 (LHW)	Pt03G14490
At2G31280 (LL2)	Pt05G22250
Atr00030G01080	Pt09G01770
Atr00069G00240	Sl06G074110
Cma2053128	Sl08G082940
Mp2035326	Sl09G066280
Os01G64560	Sl11G068960
Os02G45170	Zm02G24090
Os11G06010	Zm03G19680
Os12G06330	Zm04G34530
Os12G06335	Zm05G36270
Pe2039132	Zm10G02310
Pt01G08600	Zm10G22170

Supplemental Figure 2. Verification of LHW phylogeny using AD (A) and RAxML (B) phylogenies. The LHW clade is indicated in green. No additional genes were identified in both AD and RAxML analyses, but not in the initial ML calculation.
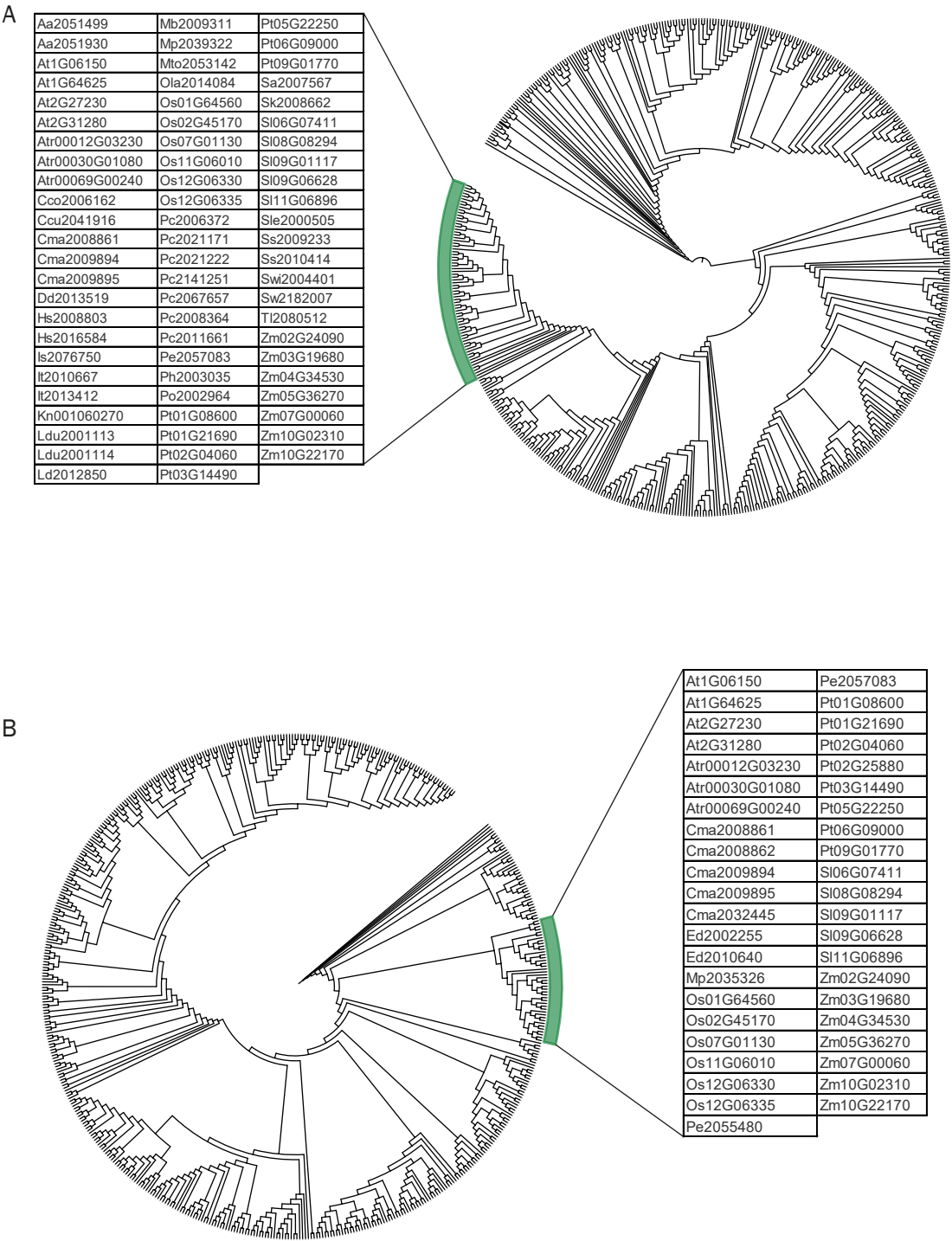
A

Aa2050828	Hs2022097	Os01G11910	Pt19G08900
At1G68810	Hs2057546	Os03G15440	Sa2002423
At2G41130	Ip2028305	Os03G59670	Sa2004992
At3G25710	Is2075584	Os05G07120	Sa2010137
At3G56770	It2089681	Os08G33590	Sk2007004
Atr00002G00080	It2099829	Os08G37290	Sk2042682
Atr00007G01320	Kn000610230	Os09G24490	Sl02G076920
Atr00010G01040	La2002693	Os09G28900	Sl03G095980
Atr00021G00510	La2002694	Pc2009292	Sl04G006990
Atr00101G01220	La2009267	Pc2014284	Sl05G009880
Ccu2007370	La2011251	Pc2014285	Sl06G051260
Cma2152538	La2013318	Pco2009198	Sl07G053290
Coco204474	Lan2017295	Pd2002811	Sl10G078380
Dd2068750	Lc2011718	Pe2010345	Sle20087620
Do2012233	Ld2003154	Pe2011294	Sm00018G02
Do2012946	Ld2008397	Pe2052556	Sm00038G02
Ed2039640	Ld2009148	Ph2057551	Ssp2023033
Ed2040338	Ld2011715	Phc2015014	Sst2011330
Ed2041096	Ldu2035940	Phc2017550	Sw2181545
Hl2010619	Mc2011025	Phc2020949	Swi2014032
Hl2010620	Me2046144	Phc2106750	Tl2078382
Hl2023097	Mk2009666	Phc2126364	Zm01G10410
Hm2005726	Mp2020606	Pt02G10530	Zm01G35380
Hm2010056	Mpa2002731	Pt04G05570	Zm01G57430
Hm2010058	Mpa2002732	Pt06G03710	Zm02G30480
Hm2052043	Mpa2005587	Pt08G07080	Zm02G31910
Hm2060638	Mpa2010922	Pt08G11600	Zm03G01110
Hs2005550	Ms2009561	Pt10G13000	Zm03G04400
Hs2006121	Mto2053481	Pt10G18670	Zm04G14040
Hs2006122	Mv2005867	Pt11G06550	Zm06G18970
Hs2010139	Na2038150	Pt13G11760	Zm07G16510
Hs2015752	Os01G06640	Pt16G03540	Zm09G24320
Hs2016510			

B

Aa2050828	Hs2006121	Mpa2010922	Pt10G18670
At1G68810	Hs2006122	Ms2009561	Pt11G06550
At2G41130	Hs2008700	Mto2053481	Pt13G11760
At3G25710	Hs2010139	Mv2005867	Pt16G03540
At3G56770	Hs2015752	Na2038150	Pt19G08900
Atr00007G01320	Hs2016510	Os01G11910	Pt19G08950
Atr00010G01040	Hs2022097	Os03G15440	Sa2004992
Atr00021G00510	Hs2057546	Os03G59670	Sa2010137
Atr00101G01220	Ip2028305	Os08G33590	Sk2042682
Ccu2007369	Is2070897	Os09G24490	Sl02G076920
Ccu2007370	Is2075584	Pc2009292	Sl03G095980
Cma2029521	It2089681	Pc2014284	Sl04G006990
Cma2115804	It2099829	Pc2014285	Sl05G009880
Cma2152538	Kn00061023	Pc2018166	Sl06G051260
Coco204474	La2002693	Pc2071655	Sl07G053290
Cr01G01160	La2002694	Pco2009198	Sl10G078380
Dd2010310	La2009267	Pd2002811	Sle20087620
Dd2068750	La2011251	Pd2113908	Sm00038G00
Do2012233	La2013318	Pd2119379	Sse2039830
Do2012946	Lan2017295	Pe2010345	Ssp2023033
Ed2039640	Lan2080604	Pe2011294	Sst2011330
Ed2040338	Lc2011718	Pe2052556	Sw2181545
Ed2041096	Ld2003154	Ph2057551	Sw2014032
Hl2010619	Ld2008397	Phc2015014	Tl2078382
Hl2010620	Ld2009148	Phc2017550	Zm01G10400
Hl2023097	Ld2011715	Phc2020949	Zm01G10410
Hm2005726	Ldu2035940	Phc2106750	Zm01G38350
Hm2005727	Mc2011025	Phc2126364	Zm01G57430
Hm2010056	Me2046144	Pt04G05570	Zm02G30480
Hm2010057	Mk2009666	Pt06G03710	Zm03G01110
Hm2010058	Mp2020606	Pt08G07080	Zm04G14040
Hm2052043	Mpa2002731	Pt08G11600	Zm09G24320
Hm2060638	Mpa2002732	Pt10G13000	Zs2026450
Hs2005550	Mpa2005587		

Supplemental Figure 3. Verification of TMO5 phylogeny using AD (A) and RAxML (B) phylogenies. The TMO5 clade is marked in green. No additional genes were identified in both AD and RAxML analyses, but not in the initial ML calculation.

A

Aa2051499	Mb2009311	Pt05G22250
Aa2051930	Mp2039322	Pt06G09000
At1G06150	Mto2053142	Pt09G01770
At1G64625	Ola2014084	Sa2007567
At2G27230	Os01G64560	Sk2008662
At2G31280	Os02G45170	Sl06G07411
Atr00012G03230	Os07G01130	Sl08G08294
Atr00030G01080	Os11G06010	Sl09G01117
Atr00069G00240	Os12G06330	Sl09G06628
Cco2006162	Os12G06335	Sl11G06896
Ccu2041916	Pc2006372	Sle2000505
Cma2008861	Pc2021171	Ss2009233
Cma2009894	Pc2021222	Ss2010414
Cma2009895	Pc2141251	Swi2004401
Dd2013519	Pc2067657	Sw2182007
Hs2008803	Pc2008364	Tl2080512
Hs2016584	Pc2011661	Zm02G24090
Is2076750	Pe2057083	Zm03G19680
It2010667	Ph2003035	Zm04G34530
It2013412	Po2002964	Zm05G36270
Kn001060270	Pt01G08600	Zm07G00060
Ldu2001113	Pt01G21690	Zm10G02310
Ldu2001114	Pt02G04060	Zm10G22170
Ld2012850	Pt03G14490	

B

At1G06150	Pe2057083
At1G64625	Pt01G08600
At2G27230	Pt01G21690
At2G31280	Pt02G04060
Atr00012G03230	Pt02G25880
Atr00030G01080	Pt03G14490
Atr00069G00240	Pt05G22250
Cma2008861	Pt06G09000
Cma2008862	Pt09G01770
Cma2009894	Sl06G07411
Cma2009895	Sl08G08294
Cma2032445	Sl09G01117
Ed2002255	Sl09G06628
Ed2010640	Sl11G06896
Mp2035326	Zm02G24090
Os01G64560	Zm03G19680
Os02G45170	Zm04G34530
Os07G01130	Zm05G36270
Os11G06010	Zm07G00060
Os12G06330	Zm10G02310
Os12G06335	Zm10G22170
Pe2055480	

Supplemental Figure 4. Verification of LHW phylogeny using AD (A) and RAxML (B) phylogenies. The LHW clade is marked in green. No additional genes were identified in both AD and RAxML analyses, but not in the initial ML calculation.

Supplemental Figure 5. Multiple Sequence Alignment of all TMO5 orthologues.
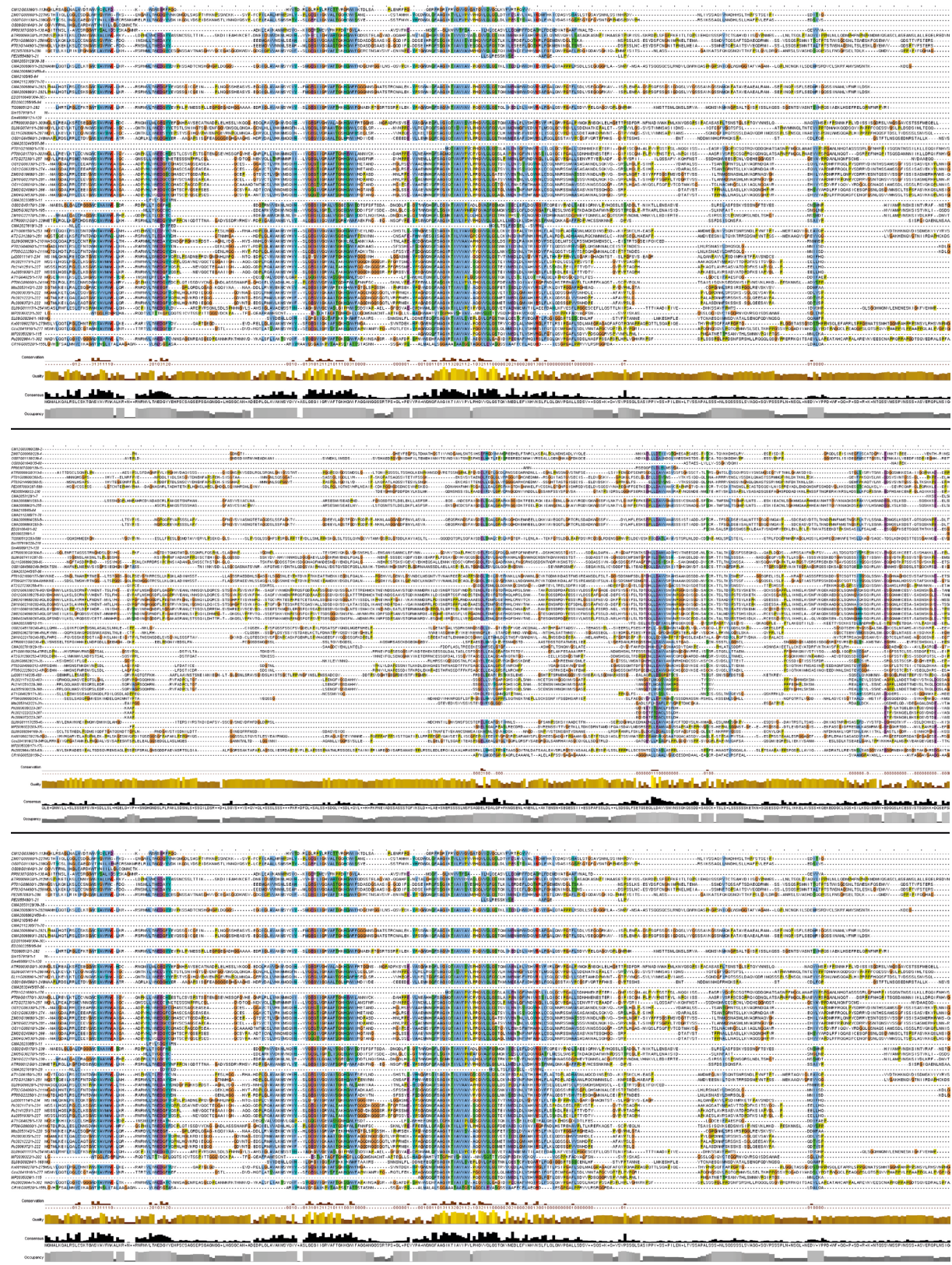
Supplemental Figure 6. Multiple Sequence Alignment of all LHW orthologues.

Supplemental Table 1. Genes within the TMO5 clade determined by ML, AD and RAxML. Purple circles indicate genes clustering within the TMO5 subfamily.

Gene	ML	AD	RAxML	Gene	ML	AD	RAxML	Gene	ML	AD	RAxML
At1G68810	●	●	●	Os03G15440	●	●	●	Sl02G07692	●	●	●
At2G40200	●	●	●	Os03G59670	●	●	●	Sl03G09598	●	●	●
At2G41130	●	●	●	Os05G07120	●		●	Sl04G00699	●	●	●
At3G25710	●	●	●	Os08G33590	●	●	●	Sl04G01436		●	●
At3G56770	●	●	●	Os08G37290			●	Sl04G07624			●
At4G38070			●	Os09G24490	●	●	●	Sl05G00988	●	●	●
Atr00007G01320	●	●	●	Os09G28900			●	Sl06G05126	●	●	●
Atr00010G01040	●	●	●	Pe2010345	●	●	●	Sl07G05329	●	●	●
Atr00021G00510	●	●		Pe2011294	●	●	●	Sl10G07838	●	●	●
Atr00101G01220	●	●	●	Pe2024173			●	Sm00018G02850		●	●
Atr00002G00080		●		Pe2052556	●	●	●	Sm00038G00580	●	●	●
Atr00021G00510			●	Pt02G10530			●	Zm01G35380		●	
Cma2021101			●	Pt04G05570	●	●	●	Zm01G10400	●		●
Cma2029521	●		●	Pt06G03710	●	●	●	Zm01G10410	●		●
Cma2115804	●			Pt07G01050			●	Zm01G38350	●		●
Cma2152538	●	●	●	Pt08G07080	●	●	●	Zm01G57430	●	●	●
Ed2039640	●	●	●	Pt08G11600	●	●	●	Zm02G30480	●	●	●
Ed2040338	●	●	●	Pt10G13000	●	●	●	Zm03G01110	●	●	●
Ed2041096	●	●	●	Pt10G18670	●	●	●	Zm03G04400	●		●
Mp2005672			●	Pt11G06550	●	●	●	Zm04G14040	●	●	●
Mp2020606	●	●	●	Pt13G11760	●	●	●	Zm07G16510		●	
Ol17G00340			●	Pt16G03540	●	●	●	Zm06G18970	●		●
Os01G06640	●			Pt19G08900	●	●	●	Zm09G24320	●	●	●
Os01G11910	●	●	●	Pt19G08950	●		●				

Supplemental Table 2. Genes within the LHW clade determined by ML, AD and RAxML. Purple circles indicate genes clustering within the LHW subfamily.

Gene	ML	AD	RAxML	Gene	ML	AD	RAxML
At1G06150	●	●	●	Os12G06330	●	●	●
At1G64625	●	●	●	Os12G06335	●	●	●
At2G27230	●	●	●	Pe2039132	●		●
At2G31280	●	●	●	Pe2055480	●	●	
Atr00012G03230	●	●		Pe2057083	●	●	
Atr00030G01080	●	●	●	Pt01G08600	●	●	●
Atr00069G00240	●	●	●	Pt01G21690	●	●	●
Cma2008861	●	●		Pt02G04060	●	●	●
Cma2008862	●	●		Pt03G14490	●	●	●
Cma2009893	●	●		Pt05G22250	●	●	●
Cma2009894	●	●		Pt06G09000	●	●	
Cma2009895	●	●		Pt09G01770	●	●	●
Cma2010579	●			Sl06G07411	●	●	●
Cma2023385	●			Sl08G08294	●	●	●
Cma2027618	●	●		Sl09G01117	●	●	
Cma2032445	●			Sl09G06628	●	●	●
Cma2053128	●		●	Sl11G06896	●	●	●
Cr16G00520		●		Sm157919	●		
Ed2002255	●			Sm46989	●		
Ed2010640	●	●		Zm02G24090	●	●	●
Mp2035326	●		●	Zm03G19680	●	●	●
Mp2039322	●	●		Zm04G34530	●	●	●
Os01G64560	●	●	●	Zm05G36270	●	●	●
Os02G45170	●	●	●	Zm07G00060		●	
Os07G01130		●		Zm10G02310	●	●	●
Os11G06010	●	●	●	Zm10G22170	●	●	●

Supplemental Table 3. Genes within the TMO5 clade after increased taxon sampling, determined by ML, AD and RAxML. Purple circles indicate genes clustering within the TMO5 subfamily.

Gene	ML	AD	RAxML
Aa2050828	●	●	●
At1G68810	●	●	●
At2G40200	●		
At2G41130	●	●	●
At3G25710	●	●	●
At3G56770	●	●	●
Atr00002G00080		●	
Atr00007G01320	●	●	●
Atr00010G01040	●	●	●
Atr00021G00510	●	●	●
Atr00101G01220	●	●	●
Ccu2007369	●		●
Ccu2007370	●	●	●
Cma2029521	●		●
Cma2115804	●		●
Cma2152538	●	●	●
Coco204474	●	●	●
Cr01G01160			●
Dd2010310	●		●
Dd2068750	●	●	●
Do2012233	●	●	●
Do2012946	●	●	●
Ed2039640	●	●	●
Ed2041096	●	●	●
Hl2010619	●	●	●
Hl2010620	●	●	●
Hl2023097	●	●	●
Hm2005726	●	●	●
Hm2005727	●		●
Hm2010056	●	●	●
Hm2010057	●		●
Hm2010058	●	●	●
Hm2052043	●	●	●
Hm2060638	●	●	●
Hs2006121	●	●	●
Hs2006122	●	●	●
Hs2016510	●	●	●
Hs2057546	●	●	●
Hse2005550	●	●	●
Hse2008700	●		●
Hse2010139	●	●	●
Hse2015752	●	●	●
Hse2022097	●	●	●
Ip2028305	●	●	●
Is2070897	●		●
Is2075584	●	●	●
It2089681	●	●	●
It2099829	●	●	●
Kn000610230	●	●	●
La2002693	●	●	●

Gene	ML	AD	RAxML
La2002694	●	●	●
La2009267	●	●	●
La2011251	●	●	●
La2013318	●	●	●
Lan2015098	●		
Lan2017295	●	●	●
Lan2080604	●		●
Lc2011718	●	●	●
Ld2003154	●	●	●
Ld2008397	●	●	●
Ld2009148	●	●	●
Ld2011715	●	●	●
Ldu2035940	●	●	●
Mc2011025	●	●	●
Me2046144	●	●	●
Mk2009666	●	●	●
Mp2020606	●	●	●
Mpa2002731	●	●	●
Mpa2002732	●	●	●
Mpa2005587	●	●	●
Mpa2010922	●	●	●
Ms2009561	●	●	●
Mto2053481	●	●	●
Mv2005867	●	●	●
Na2038150	●	●	●
Os01G06640	●	●	
Os01G11910	●	●	●
Os03G15440	●	●	●
Os03G59670	●	●	●
Os05G07120	●	●	
Os08G33590	●	●	●
Os08G37290		●	
Os09G24490	●	●	●
Os09G28900		●	
Pco2009198	●	●	●
Pc2009292	●	●	●
Pc2014284	●	●	●
Pc2014285	●	●	●
Pc2018166	●		●
Pc2071655	●		
Phc2015014	●	●	●
Phc2017550	●	●	●
Phc2020949	●	●	●
Phc2106750	●	●	●
Phc2126364	●	●	●
Pd2002811	●	●	●
Pd2113908	●		●
Pd2119379	●		●
Pe2010345	●	●	●
Pe2011294	●	●	●

Gene	ML	AD	RAxML
Pe2052556	●	●	●
Ph2057551	●	●	●
Pt02G10530		●	
Pt04G05570	●	●	●
Pt06G03710	●	●	●
Pt08G07080	●	●	●
Pt08G11600	●	●	●
Pt10G13000	●	●	●
Pt10G18670	●	●	●
Pt11G06550	●	●	●
Pt13G11760	●	●	●
Pt16G03540	●	●	●
Pt19G08900	●	●	●
Pt19G08950	●		
Sa2002423		●	
Sa2004992	●	●	●
Sa2010137	●	●	●
Sk2007004		●	
Sk2042682	●	●	●
Sl02G076920	●	●	●
Sl03G095980	●	●	●
Sl04G006990	●	●	●
Sl05G009880	●	●	●
Sl06G051260	●	●	●
Sl07G053290	●	●	●
Sl10G078380	●	●	●
Sle20087620	●	●	●
Sm00018G02850		●	
Sm00038G00580	●	●	●
Sse2039830	●		●
Ssp2023033	●	●	●
Sst2011330	●	●	●
Sw2181545	●	●	●
Swi2014032	●		●
Tl2078382	●	●	●
Zm01G10400	●		●
Zm01G10410	●	●	●
Zm01G38350	●		●
Zm01G35380		●	
Zm01G57430	●	●	●
Zm02G30480	●	●	●
Zm02G31910		●	
Zm03G01110	●	●	●
Zm03G04400	●	●	
Zm04G14040	●	●	●
Zm06G18970	●	●	
Zm07G16510		●	
Zm09G24320	●	●	●
Zs2026450	●		●

Supplemental Table 4. Genes within the LHW clade after increased taxon sampling, determined by ML, AD and RAxML. Purple circles indicate genes clustering within the TMO5 subfamily.

	ML	AD	RAxML
Aa2051499	●	●	
Aa2051930	●	●	
At1G06150	●	●	●
At1G64625	●	●	●
At2G27230	●	●	●
At2G31280	●	●	●
Atr00012G03230	●	●	●
Atr00030G01080	●	●	●
Atr00069G00240	●	●	●
Cco2006162		●	
Ccu2041916	●	●	
Cma2008861	●	●	●
Cma2008862	●		●
Cma2009894	●	●	●
Cma2009895	●	●	●
Cma2023385	●		●
Cma2027618	●		
Cma2032445	●		
Cma2053128	●		
Dd2013519	●	●	
Ed2002255	●		●
Ed2010640	●		●
Hm2013431	●		
Hse2008802	●		
Hse2008803	●	●	
Hs2016584	●	●	
Is2076750	●	●	
It2010667	●	●	
It2013412	●	●	

	ML	AD	RAxML
Kn001060270	●	●	
Ldu2001113	●	●	
Ldu2001114	●	●	
Ld2012850	●		
Mb2009311		●	
Mp2035326	●		●
Mp2039322	●	●	
Mto2053142	●	●	
Ola2014084		●	
Os01G64560	●	●	●
Os02G45170	●	●	●
Os07G01130	●	●	●
Os11G06010	●	●	●
Os12G06330	●	●	●
Os12G06335	●	●	●
Phc2006372	●	●	
Phc2021171	●	●	
Phc2021222	●	●	
Phc2141251	●	●	
Pco2067657	●	●	
Pc2008364	●	●	
Pc2011661	●	●	
Pd2128658	●		
PE2055480	●		●
PE2057083	●	●	●
Ph2003035	●	●	
Po2002964	●	●	
Pt01G08600	●	●	●
Pt01G21690	●	●	●

	ML	AD	RAxML
Pt02G04060	●	●	●
Pt02G25880	●		●
Pt03G14490	●	●	●
Pt05G22250	●	●	●
Pt06G09000	●	●	●
Pt09G01770	●	●	●
Sac2007567	●	●	
Sk2008662	●	●	
Sl06G07411	●	●	●
Sl08G08294	●	●	●
Sl09G01117	●	●	●
Sl09G06628	●	●	●
Sl11G06896	●	●	●
Sle2000505	●	●	
Sm157919	●		
Sm46989	●		
Sst2003544	●		
Sst2009233	●	●	
Sse2010414	●	●	
Swi2004401	●	●	
Sw2182007	●	●	
Tl2080512	●	●	
Zm02G24090	●	●	●
Zm03G19680	●	●	●
Zm04G34530	●	●	●
Zm05G36270	●	●	●
Zm07G00060	●	●	●
Zm10G02310	●	●	●
Zm10G22170	●	●	●

Supplemental Table 5. Species used in this study and the amount of TMO5 and LHW orthologues identified.
Note that transcriptomes derived from the OneKP are only indicated when at least one orthologue was identified.

Species	Abbreviation	Clade	Order	Database	LHW orthologues	TMO5 orthologues
Arabidopsis thaliana	At	Angiosperm	Brassicales	PLAZA 3.0	4	5
Populus trichocarpa	Pt	Angiosperm	Malpighiales	PLAZA 3.0	8	11
Solanum lycopersicum	Sl	Angiosperm	Solanales	PLAZA 3.0	5	8
Oryza sativa	Os	Angiosperm	Poales	PLAZA 3.0	6	7
Zea Mays	Zm	Angiosperm	Poales	PLAZA 3.0	7	10
Amborella trichopoda	Atr	Basal angiosperm	Amborellales	PLAZA 3.0	3	4
Picea engelmannii	Pe	Gymnosperm	Pinales	PLAZA 3.0	2	3
Culcita macrocarpa	Cma	Leptosporangiate Monilophytes (ferns)	Cyatheales	PLAZA 3.0	8	3
Equisetum diffusum	Ed	Eusporangiate Monilophytes (horsetails)	Equisetales	PLAZA 3.0	2	2
Isoetes sp	Is	Lycophytes	Isoetales	OneKP	1	2
Isoetes tegetiformans	It	Lycophytes	Isoetales	OneKP	2	2
Dendrolycopodium obscurum	Do	Lycophytes	Lycopodiales	OneKP	-	2
Diphasiastrum digitatum	Dd	Lycophytes	Lycopodiales	OneKP	1	2
Huperzia lucidula	Hl	Lycophytes	Lycopodiales	OneKP	-	3
Huperzia myrisinites	Hm	Lycophytes	Lycopodiales	OneKP	1	6
Huperzia selago	Hse	Lycophytes	Lycopodiales	OneKP	2	5
Huperzia squarrosa	Hs	Lycophytes	Lycopodiales	OneKP	1	4
Lycopodiella apressa	La	Lycophytes	Lycopodiales	OneKP	-	5
Lycopodium annotinum	Lan	Lycophytes	Lycopodiales	OneKP	-	3
Lycopodium deuterodensum	Ld	Lycophytes	Lycopodiales	OneKP	1	4
Phylloglossum drummondii	Pd	Lycophytes	Lycopodiales	OneKP	1	3
Pseudolycopodiella caroliniana	Pc	Lycophytes	Lycopodiales	OneKP	3	5
Selaginella acanthonota	Sac	Lycophytes	Selaginellales	OneKP	1	1
Selaginella apoda	Sa	Lycophytes	Selaginellales	OneKP	-	1
Selaginella kraussiana	Sk	Lycophytes	Selaginellales	OneKP	1	1
Selaginella lepidophylla	Sle	Lycophytes	Selaginellales	OneKP	1	1
Selaginella moellendorffii	Sm	Lycophytes	Selaginellales	PLAZA 2.5	2	1
Selaginella selaginoides	Sse	Lycophytes	Selaginellales	OneKP	1	1
Selaginella stauntoniana	Sst	Lycophytes	Selaginellales	OneKP	2	1
Selaginella wallacei	Sw	Lycophytes	Selaginellales	OneKP	1	1
Selaginella willdenowii	Swi	Lycophytes	Selaginellales	OneKP	1	1
Anthoceros agrestis	Aa	Hornworts	Anthocerotales	OneKP	2	1
Paraphymatoceros hallii	Ph	Hornworts	Anthocerotales	OneKP	1	1
Megaceros tosanus	Mto	Hornworts	Dendrocerotales	OneKP	1	1
Megaceros vincentianus	Mv	Hornworts	Dendrocerotales	OneKP	-	1
Nothoceros aenigmaticus	Na	Hornworts	Dendrocerotales	OneKP	-	1
Phaeomegaceros coriaceus	Pco	Hornworts	Dendrocerotales	OneKP		1
Leiosporoceros dussii	Ldu	Hornworts	Leiosporocerotal	OneKP	2	1
Phaeoceros carolinianus	Phc	Hornworts	Notothyladales	OneKP	4	5
Physcomitrella patens	Pp	Mosses	Funariales	PLAZA3.0	-	-
Takakia lepidozioides	Tl	Mosses (Takakiopsida)	Takakiales	OneKP	1	1
Conocephalum conicum	Coco	Liverworts	Marchantiales	OneKP	-	1
Lunularia cruciata	Lc	Liverworts	Marchantiales	OneKP	-	1
Marchantia paleacea	Mpa	Liverworts	Marchantiales	OneKP	-	1
Marchantia polymorpha	Mp	Liverworts	Marchantiales	OneKP	1	1
Klebsormidium nitens	Kn	Charophyte	Klebsormidiales	NIES-2285	1	1
Interfilum paradoxum	Ip	Charophyte	Klebsormidiales	OneKP	-	1
Cylindrocystis cushleckae	Ccu	Charophyte	Zygnematales	OneKP	1	1
Mesotaenium caldariorum	Mc	Charophyte	Zygnematales	OneKP	-	1
Mesotaenium endlicherianum	Me	Charophyte	Zygnematales	OneKP	-	1
Mesotaenium kramstae	Mk	Charophyte	Zygnematales	OneKP	-	1
Mougeotia sp	Ms	Charophyte	Zygnematales	OneKP	-	1
Spirotaenia sp	Ssp	Charophyte	Zygnematales	OneKP	-	1
Zygnemopsis sp	Zs	Charophyte	Zygnematales	OneKP	-	1
Planotaenium ohtanii	Po	Charophyte	Zygnematales	OneKP	-	-
Chlamydomonas reinhardtii	Cr	Chlorophyte	Chlamydomonad	Pico-PLAZA	-	-
Ostreococcus lucimarinus	Ol	Chlorophyte	Mamiellales	Pico-PLAZA	-	-
Cyanidioschyzon merolae	Cm	Rhodophytes	Cyanidiales	Pico-PLAZA	-	-

Chapter 3

Functional conservation of the TMO5/LHW heterodimer complex

Nicole van 't Wout Hofland[1], Dolf Weijers[1] and Bert De Rybel[1,2,3]

1. Wageningen University, Laboratory of Biochemistry, Stippeneng 4, 6708 WE Wageningen, the Netherlands
2. Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, 9052 Ghent, Belgium
3. VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium

Abstract

In the vascular plant Arabidopsis, the transcription factors TARGET OF MONOPTEROS 5 (TMO5) and LONESOME HIGHWAY (LHW) function as a heterodimer to induce vascular proliferation. Orthologues of these vascular tissue regulators are not only present in Tracheophytes, but also in Bryophytes and Charophytes (Chapter 2). One can thus question if the vascular-specific function of these transcription factors is conserved in Bryophytes and Charophytes. Here, we study when the TMO5 and LHW proteins obtained the capacity to induce periclinal/radial cell divisions (the 'vascular function'). We performed complementation assays in which we attempted to rescue Arabidopsis *tmo5 t5l1* double and *lhw* single mutants with *TMO5* and *LHW* orthologues respectively. Interestingly, TMO5 proteins in the Bryophyte *Marchantia polymorpha* are able to perform the vascular function, suggesting that the TMO5 function is highly conserved. In contrast, the vascular function of LHW is restricted to Tracheophytes. We further show that heterodimerization properties of TMO5 and LHW were established in the ancestors of Bryophytes, but that this Bryophytic dimer (MpTMO5/MpLHW) is inadequate to induce vascular proliferation. In summary, our results show that the vascular function of the TMO5/LHW heterodimer is correlated with the emergence of Tracheophytes. This leads to the hypothesis that TMO5 and LHW may have played a role in the emergence of vascular tissues.

58

Introduction

Transcription factors are DNA-binding proteins that enhance or suppress transcription of target genes. They can coordinate a wide range of biological processes, such as cell differentiation and development. Their functionality, and in some cases their DNA-binding capacities, are often dependent on interaction with a protein partner. Depending on the interaction partner, a dimer can activate an array of downstream targets leading to specific cellular outputs, ensuring proper development and functioning of the organism. Each monomer contains a specific DNA-binding motif. Combining two monomers with two distinct DNA-binding domains creates a heterodimer and allows recognition of novel targets and can increase DNA-binding specificity.

Transcription factors TARGET OF MONOPTEROS 5 (TMO5) and LONESOME HIGHWAY (LHW) form a heterodimer that is involved in radial growth of the vascular tissue in Arabidopsis. This growth is induced by a controlled auxin flow activating TMO5 through AUXIN RESPONSE FACTOR 5 (ARF5), also known as MONOPTEROS (MP). TMO5, as well as the three TMO5-LIKE (T5L) homologues, form a transcriptional heterodimer complex with LHW, or one of its two LHW-LIKE (LL) homologues. Together, they activate transcription of LONELY GUY 4 (LOG4) and its homologue LOG3, which encode an enzyme in the biosynthesis pathway of cytokinin (De Rybel et al., 2013; Ohashi-Ito et al., 2014). Subsequently, a signal is sent to neighbouring procambium cells to initiate periclinal and radial cell divisions (De Rybel et al., 2014; Ohashi-Ito et al., 2014). Loss of function of TMO5, LHW or their homologues, results in reduced periclinal and radial cell divisions and, by extension, a reduction in vascular cell files (De Rybel et al., 2013; Ohashi-Ito et al., 2013). In contrast, introducing extra copies of *TMO5* and *LHW* induces overproliferation, resulting in additional vascular cell files (De Rybel et al., 2013; Ohashi-Ito et al., 2014).

As key transcription factors of vascular tissue proliferation, TMO5 and LHW may have been involved in the evolution of complex vascular tissues. In the previous chapter, we investigated proteins of the plant kingdom with high amino acid sequence homology to TMO5 and LHW, referred to as TMO5- and LHW orthologues. Orthologues of both proteins were identified in the plant lineage ranging from Angiosperms to Charophytes. While homologous sequences often share similar structures and function, functionality cannot be inferred solely based on homology, especially between evolutionary distant species. In this chapter, we therefore investigate the evolutionary conservation of the biological role of TMO5 and LHW orthologues (functional conservation). We found that functional conservation of the dimer as a whole is restricted to Tracheophytes, whereas the vascular specific function of TMO5 was already established in the ancestor of Bryophytes.

Results

TMO5 function is highly conserved in evolution, while LHW function is restricted to Tracheophytes

The in depth-phylogenetic analysis of *TMO5* and *LHW*, discussed in Chapter 2, identified *TMO5-* and *LHW* orthologues in the plant lineage from Charophytes to Angiosperms. Although these genes appear similar in sequence to respectively At*TMO5* and At*LHW*, they may not have the capacity to regulate periclinal cell division in the vasculature, henceforth referred to as the 'vascular function'. In order to study when the vascular function of TMO5 and LHW developed, we used an inter-species complementation assay in which we evaluate the degree of complementation by introducing TMO5 and LHW orthologues in the Arabidopsis *tmo5 t5l1* double or *lhw* single mutant respectively. These mutants are compromised in vascular cell proliferation resulting in a monarch pattern, containing one xylem and one phloem pole (**Figure 1B, J**), compared to the diarch pattern found in wild type roots (**Figure 1A**), with two xylem and two phloem poles. Thus, introducing an orthologue able to perform the vascular function should restore the diarch pattern in the mutant.

We selected *TMO5* and *LHW* orthologous genes from a range of representative species, based on the calculated phylogenetic trees described in Chapter 2 (**Chapter 2: Figure 9, 10**). *TMO5* orthologous genes were selected from the species *Populus trichocarpa, Oryza sativa, Picea engelmannii, Equisetum diffusum, Selaginella moellendorffii, Marchantia polymorpha* and *Klebsormidium nitens*. Per species, one or two genes were selected, which were predicted to be closely related to AtTMO5, AtT5L1, AtT5L2 or AtT5L3. Of the eleven Populus genes, we selected two non-paralogous genes, one which clustered within the TMO5 and T5L1 clade (Pt08G11600), and one within the T5L2 and T5L3 clade (Pt16G03540). Furthermore, we selected an Oryza gene from the monocotyledon clade, sister to TMO5, T5L1, T5L2 and T5L3 (Os03G15440). The Gymnosperm Picea only has three TMO5 orthologues, two of which are likely paralogues. As such, we selected one of the paralogues (Pe2010345) and the third gene (Pe2011294). However, because of technical reasons we used the related *Picea abies* (Pa132680g0010 and Pa65818g0010 respectively) for experimental work. Since the two TMO5 orthologues of Equisetum appear to be paralogs, we selected a single gene (Ed2039640). Given that Selaginella, Marchantia, and Klebsormidium only contain one *TMO5* orthologue (Sm00038G00580, Mp2020606 – also known as Mapoly0039s0068 - and Kn000610230 respectively), single genes were selected for these species.

We attempted to complement the Arabidopsis *tmo5 t5l1* double mutant by driving the TMO5 orthologues from the native At*TMO5* promoter (see Material and Methods section for details). Vascular patterns were studied in T1 and T2 lines (**Figure 1A-I, Supplemental Table 1**). Contradictory to our expectations, not only the TMO5 orthologous genes of Tracheophytic species
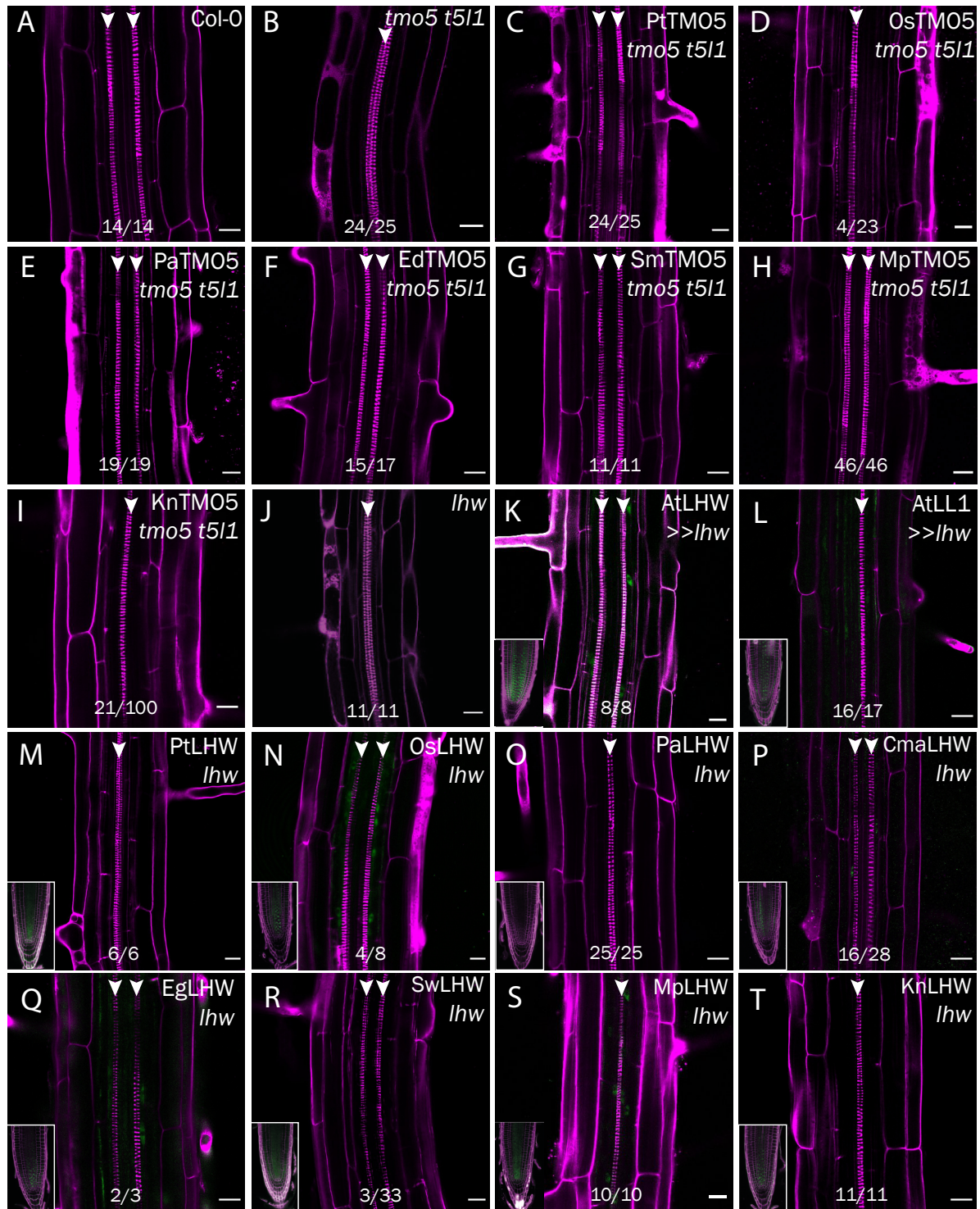
Figure 1. Evolutionary analysis of TMO5 and LHW function in vascular tissue proliferation in the Arabidopsis root. (A) Wild type Arabidopsis root with two xylem poles. **(B)** Arabidopsis *tmo5 t5l1* double mutant root showing a monarch pattern. **(C - I)** Attempted rescues of the *tmo5 t5l1* double mutant phenotype by TMO5 orthologues of *Populus trichocarpa* (C), *Oryza sativa* (D), *Picea abies* (E), *Equisetum diffusum* (F), *Selaginella moellendorffii* (G), *Marchantia polymorpha* (H) and *Klebsormidium nitens* (I) using the Arabidopsis p*TMO5* promoter. Note that heterozygous lines were used for this assay (see Material and Methods). **(J)** Arabidopsis *lhw* mutant root showing a monarch pattern. **(K)** Positive control, complementation of the *lhw* mutant by AtLHW. **(L)** *lhw* complementation by AtLL1 **(M - T)** Attempted rescues of the *lhw* mutant phenotype by LHW orthologues of *Populus trichocarpa* (M), *Oryza sativa* (N), *Picea abies* (O), *Culcita macrocarpa* (P), *Equisetum giganteum* (Q), *Selaginella wallacei* (R), *Marchantia polymorpha* (S) and *Klebsormidium nitens* (T), fused to a YFP and expressed under the Arabidopsis p*LHW* promoter. Numbers at the bottom indicate the number of times the represented xylem pattern was observed per total lines screened. Arrowheads mark protoxylem poles. Scale bars represent 20 µm.

Populus (Pt08G11600: PtTMO5, **Figure 1C**), Picea (Pa65818G0010, henceforth referred to as PaTMO5, **Figure 1E**), Equisetum (EdTMO5, **Figure 1F**) and Selaginella (SmTMO5, **Figure 1G**), but also the TMO5 orthologue of the Bryophyte Marchantia (MpTMO5, **Figure 1H**), was able to restore the diarch architecture of the *tmo5 t5l1* double mutant. In contrast, the evolutionary more distant *TMO5* gene of Klebsormidium (KnTMO5) was unable to restore the diarch pattern in the *tmo5 t5l1* phenotype (**Figure 1I**). In addition, no full rescue of the phenotype was observed by the TMO5 orthologue of Oryza (**Figure 1D**). Considering that the TMO5 vascular function appears to be conserved in all Tracheophyte species tested, it is likely that one of the two other Oryza genes (Os03G59670 or Os08G33590) is the true orthologue providing the vascular function. Interestingly, of the two selected TMO5 orthologues of Populus and of Picea only one was able to rescue the mutant (Pt08G11600 and Pa65818g0010), while the other orthologue (Pt16G03540 and Pa132680g0010) did not show any complementation (**Supplemental Figure 1**). In both cases only the gene closest related to AtTMO5 had the capacity to perform the vascular function. Possibly, functional divergence occurred during evolution, explaining the increased evolutionary distance to AtTMO5 as well as their lack of vascular function. In summary, we found evidence that the vascular function of TMO5 was established in the ancestor of Bryophytes, approximately 530 million years ago, prior to vascular tissue formation.

In parallel, we determined conservation of the LHW vascular function in a similar complementation study using the *lhw* mutant. *LHW* orthologous genes were selected from *Arabidopsis thaliana, Populus trichocarpa, Oryza sativa, Picea engelmannii, Equisetum diffusum, Culcita macrocarpa, Selaginella wallacei, Marchantia polymorpha* and *Klebsormidium nitens*. Given that orthologues of one species generally clustered within the same clade, with the same evolutionary distance to LHW, only one gene was selected per species. We chose Pt01G21690, one of the two closest orthologues of At*LHW*. Furthermore, Os01G64560, one of the four genes closest to At*LHW*, was selected to represent the monocotyledon sister clade of LHW. Since the Picea genome contains only two LHW genes, which are likely the result of a recent duplication, a single gene was selected. Like for the TMO5 orthologues, we used *Picea abies* for experimental work (Pa45581G0020). The sister clade of LHW, LL1, LL2 and LL3 was formed by Equisetum and Culcita. From this clade one gene of Equisetum (Ed2002255) and one of Culcita (Cma2009894) was selected. However, the *Equisetum diffusum LHW* orthologue was replaced by the related *Equisetum giganteum* (Eg1939), as the sequence of the *Equisetum diffusum LHW* orthologous genes were incomplete. Given that all Lycophytes clustered together in one clade with the same evolutionary distance to AtLHW, we selected a Lycophytic gene with the highest sequence identity to AtLHW (37%): Sw2182007. The single Marchantia (Mp2039322, also known as Mapoly0088s0049) and Klebsormidium (Kn001060270) *LHW* orthologues were included to represent the non-vascular plant lineage. Lastly we included the Arabidopsis LHW and LL1 as positive and negative controls respectively.

We performed interspecies genetic complementation experiments using the selected *LHW* orthologues fused to a YFP. Intriguingly, of the LHW orthologues only the positive control, AtLHW, showed a full rescue of the monarch pattern in the root (**Figure 1J, K, Supplemental Table 1**). Introducing the closest Arabidopsis LHW homologue, LL1, did not restore the diarch pattern of the Arabidopsis root (**Figure 1L**). This suggests that putative orthologues which are evolutionary closer to AtLL1 than to AtLHW, probably will not be able to complement the *lhw* phenotype. Furthermore, these results indicate that complementation analyses of species where multiple paralogs exist have to be taken with caution as false negative findings are possible. Os01G64560 (OsLHW), Cma2009894 (CmaLHW), Eg1939 (EgLHW) and Sw2182007 (SwLHW) all expressed YFP in the nuclei of the root meristem (**Supplemental Figure 2**) and restored the diarch architecture in 50% (**Figure 1N**), 59% (**Figure 1P**), 66% (**Figure 1Q**), and 9% (**Figure 1R**) of the lines respectively. Strikingly, even within independent lines, complementation was variable in the next generation (**Supplemental Table 1**). For example, while less than 60% of the CmaLHW lines showed complementation in the T1 generation, all independent T2 lines were able to induce cell proliferation in the vasculature in at least 10% of the roots (**Supplemental Table 1**). Even though the level of complementation is less strong in comparison to the TMO5 rescue experiments, these results show that LHW orthologues of Tracheophytes, ranging from Angiosperms to Lycophytes, are able to induce periclinal cell divisions in the vasculature. This demonstrates that the vascular function is conserved throughout Tracheophyte species. Two of the selected Tracheophytes (Pt01G21690 and Pa45581g0020) were unable to restore the diarch pattern in *lhw* mutants. The latter protein was not expressed in the root meristem (**Figure 1O**, inset, **Supplemental Figure 2**) and could thus not induce cell divisions in the vasculature. Most likely incorrect paralogues, with diverged functions, were selected from the Populus and Picea genome (**Figure 1M, O**). Notably, the expression of the *LHW* orthologous genes of the Bryophyte and Charophyte Marchantia and Klebsormidium was strong in the meristem, but failed to rescue the *lhw* monarch pattern (**Figure 1S, T, Supplemental Figure 2**). Since only one LHW orthologue exists in Marchantia and Klebsormidium, it can be stated with high certainty that the vascular function of LHW developed after Bryophytes and Tracheophytes split. In conclusion, our analysis showed that the vascular function of LHW emerged in the Tracheophytes, simultaneously with the emergence of vascular tissue.

To get more insight into the partial rescue of LHW orthologues as well as to verify the binary outcome of the complementation assays of TMO5 and LHW, cell file numbers in the vascular tissue were quantified. Radial sections, taken through the middle of the meristem (see Material and Methods section for details) were used for vascular cell file quantification (**Figure 2A-J**). While wild type roots were comprised of, on average, 45 vascular cell files, the *tmo5* single and *tmo5 t5l1* double mutant were made up of 39 and 23 cell files respectively (**Figure 2K**). Complementation of
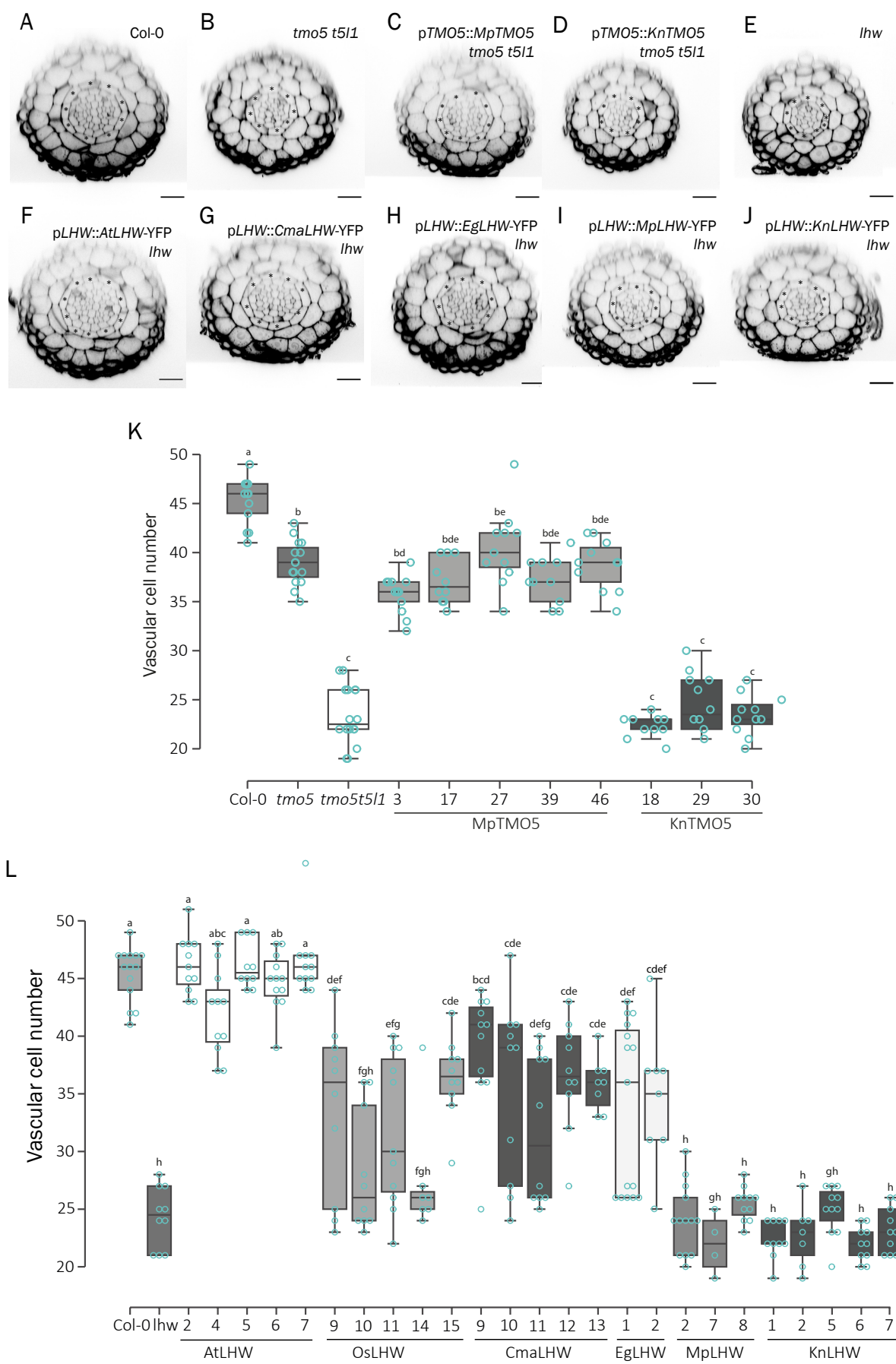
◄ **Figure 2. Quantification of vascular cell proliferation in *tmo5 t5l1* and *lhw* complementation lines. (A-J)** Representative radial sections of col-0 (A), *tmo5 t5l1* double mutant (B), complementations of the *tmo5 t5l1* double mutant with MpTMO5 (C) and KnTMO5 (D), *lhw* single mutant (E), complementation of the *lhw* mutant with AtLHW (F), and orthologues CmaLHW (G), EgLHW (H), MpLHW (I) and KnLHW (J). Asterisks indicate the endodermis. Scale bars represent 20 μm **(K)** Vascular cell file quantification in radial sections of 5 day-old roots of the aforementioned TMO5 orthologues. **(L)** Vascular cell file quantification in radial sections of 5 day-old roots of the aforementioned LHW orthologues. Turquoise circles indicate individual roots. Lines indicated with different letters can be distinguished from each other (one-way ANOVA with multiple comparisons corrected using a Tukey test, $P < 0.05$)

MpTMO5, which resulted in a diarch pattern, restored the vascular cell file number to an average of 38 cell files, similar to the *tmo5* single mutant (**Figure 2K**). Consistent with the initial TMO5 analysis, introduction of KnTMO5 into the *tmo5 t5l1* double mutant did not significantly increase the vascular cell files (**Figure 2K**). These outcomes fit with the results of the complementation assay and thus support the theory that the vascular function of TMO5 developed prior to the emergence of Tracheophytes.

Similar to the *tmo5 t5l1* double mutant, the *lhw* mutant was comprised of 23 cell files in the vascular tissue. In accordance with our initial complementation study, vascular cell file numbers did not increase when MpLHW or KnLHW were introduced, while AtLHW was able to fully restore the mutant phenotype to 45 vascular cell file numbers (**Figure 2L**). OsLHW, CmaLHW and EgLHW, which were shown to partially complement the *lhw* phenotype, also showed partial rescue of the number of vascular cell files (**Figure 2L**). While not restoring the cell files to 45, a significant

	TMO5		LHW	
Arabidopsis thalia	●	5	●	4
Populus tremulus	●	11	⊗	8
Solanum lycopersicum	○	7	○	5
Oryza sativa	⊗	7	●	6
Zea Mays	○	10	○	7
Amborella trichopoda	○	4	○	3
Picea (engelmannii / abies)	●	3	⊗	2
Culcita macrocarpa	○	3	●	8
Equisetum (diffusum / gigateum)	●	2	●	2
Huperzia selago	○	5	○	2
Selaginella (moellendorffii / wallacei)	●	1	●	2
Megaceros tosanus	○	1	○	1
Leiosporoceros dussii	○	1	○	2
Takakia lepidozioides	○	1	○	1
Marchantia polymorpha	●	1	⊗	1
Klebsormidium nitens	⊗	1	⊗	1
Cylindrocystis cushleckae	○	1	○	1
Chlamydomonas reinhardtii				
Ostreococcus lucimarinus				
Cyanidioschyzon merolae				

Tree labels (left to right on phylogeny): Eudicots, Monocots, Basal angiosperms, Gymnosperms, Horsetails, Ferns, Lycophytes, Hornworts, Mosses, Liverworts, Charophytes, Chlorophytes

Figure 3. Reconstruction of the TMO5 and LHW evolution. Phylogeny of taxonomic classes are shown on the left. The purple and green circle in the tree indicate the emergence of the vascular function of TMO5 and LHW respectively. On the right, circles indicate the presence of a putative orthologue. Numbers indicate the amount of orthologues identified by phylogenic studies (Chapter 2). Filled circles represent the capacity of performing the vascular function, while crosses indicate the absence. Open circles were not tested.

increase was observed for most lines. Vascular cell files ranged from 23 up to 42, indicating these orthologues are indeed capable of mediating vascular proliferation. Overall, this quantification confirmed the initial analysis of TMO5 and LHW orthologues and demonstrates that the binary monarch/diarch system is a representative output for the number of vascular cell files.

In conclusion, *TMO5* and *LHW* orthologues genes emerged in Charophytes, coding for one of the earliest bHLH plant proteins (**Figure 3**). Interestingly, TMO5 already acquired the vascular function in the ancestor of Bryophytes, prior to vascular emergence, as was illustrated by the complementation of the *tmo5 t5l1* phenotype by MpTMO5 (**Figure 3**). In contrast, the vascular function of LHW did not emerge until Tracheophytes and Bryophytes split (**Figure 3**). As a consequence, the vascular function of the TMO5/LHW dimer emerged in the ancestor of vascular plants, and may therefore have played an important role in creating complex vascular tissues.

TMO5/LHW acts strictly as an obligate heterodimer in Arabidopsis

Given the asymmetrical conservation between the vascular function of TMO5 and LHW, we set out to investigate if TMO5/LHW acts strictly as an obligate heterodimer in Arabidopsis. Promoter-reporter analyses demonstrated that *TMO5* and its homologues are strongly enriched in the (proto) xylem cells within the vascular bundle, while *LHW* and homologues are more broadly expressed (**Figure 4A**, De Rybel et al., 2013). Interactions between one broad or ubiquitous expressed transcription factor, such as LHW, with several tissue-restricted transcription factors creates an elegant mechanism to tissue specify functions with limited amount of transcription factors. Hence, the possibility may exist that LHW has other interaction partners besides TMO5. Consistent with the overlapping expression of *TMO5* and *LHW*, both Arabidopsis *tmo5 t5l1* double and *lhw* single mutants show strong vasculature phenotypes, indicating that the *TMO5* and *LHW* family genes have a major cooperative function in vascular tissues (De Rybel et al., 2013). In addition, it has been shown that only simultaneous over-expression of TMO5 and LHW generates dramatic phenotypic changes, suggesting they function as heterodimers (De Rybel et al., 2013). While a strong reduction in the number of vascular cell files and a monarch root anatomy is observed in *lhw* single mutants, vascular cell proliferation is almost fully absent in *lhw lhw-like1 (ll1)* double mutants, resulting in an extremely dwarfed plant (De Rybel et al., 2013). Despite these strong defects in vascular proliferation, the neighboring cell files appear normal (De Rybel et al., 2013), suggesting LHW family proteins have no additional function outside of the vascular tissues.

To further investigate whether LHW indeed only functions inside the vasculature, we designed a complementation experiment to rescue the *lhw ll1* double mutant by expressing *LHW* only in the xylem cells (p*TMO5*::LHW, **Figure 4B-D**). The p*TMO5*::LHW in *lhw ll1* showed almost full

rescue of the phenotype not only in the root vasculature (**Figure 4C, D**), but also in the shoot phenotype and the total growth (**Figure 4B**), suggesting that xylem cell specific expression of *LHW* is sufficient for its functions. One alternative hypothesis is that LHW might be a mobile protein which can migrate to other cells and in this way rescue the *lhw ll1* mutant. To verify the mobility of LHW protein, we studied the p*TMO5*::LHW-GFP expression pattern and only observed the fluorescence within the xylem cells (**Figure 4D**). Although it is still possible that the addition of the GFP tag interferes with mobility, these results suggest that LHW is a cell-autonomous protein.
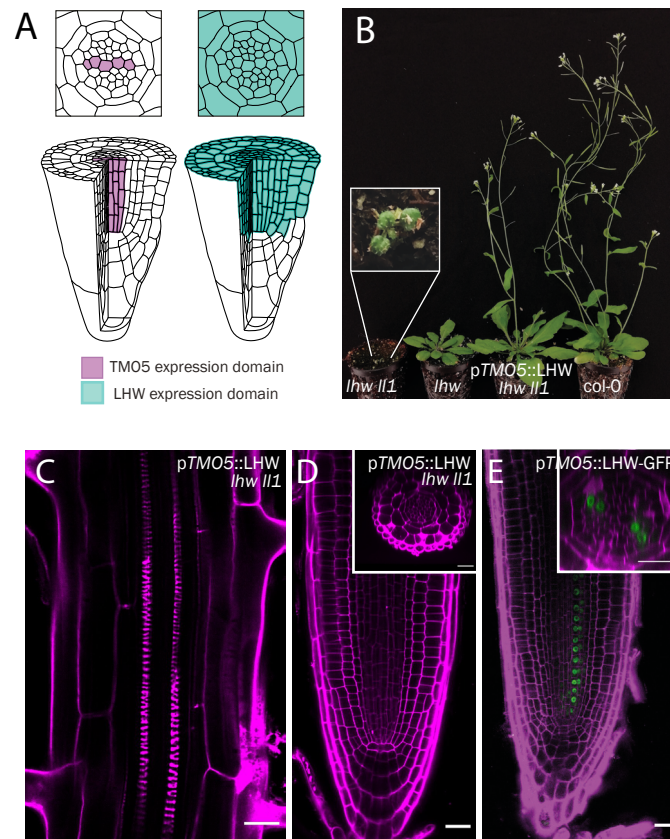
Figure 4. TMO5 and LHW strictly function as an obligate heterodimer in Arabidopsis. (A) The expression pattern and co-localization of TMO5 (in xylem precursor cells) and LHW (more broadly localized) in the primary root meristems. Image designed by Ykä Helariutta (Bonke et al., 2003), modified with permission. **(B-D)** Rescue of the *lhw ll1* double mutant phenotype in the shoot (B), the root meristem (C) and roottip (D). **(E)** Protein localization of LHW fused to GFP under the TMO5 promoter. Scale bars represent 20 µm.

In summary, our data suggests that despite its broad expression pattern in the root meristem, the absolute requirement of LHW is restricted to the xylem precursor cells. In concert with the fact that *TMO5* and *LHW* only overlap in xylem cells and that the double over-expression phenotype is much stronger than the single over-expression lines, we conclude that TMO5/LHW most likely acts as an obligate heterodimer complex in Arabidopsis.

Heterodimerization properties of TMO5/LHW were established in Bryophyte ancestors

Given that TMO5 and LHW are obligated to heterodimerize in order to initiate periclinal divisions in the vascular tissues, vascular functions of MpLHW and KnTMO5 may have been compromised due to the inability to form heterodimers with AtTMO5 or AtLHW. In order to study the heterodimerization abilities, we used Bimolecular Fluorescence Complementation (BiFC) in *N. benthamiana* (**Figure 5**). YFP halves (nYFP and cYFP) were fused to the C-terminus of TMO5 and LHW proteins. The positive control, AtTMO5-nYFP with AtLHW-cYFP, showed clear YFP signals (**Figure 5A**), validating the assay. Self-assembly, i.e. folding of the two YFP halves in the absence of a protein-protein interaction, of all used constructs was ruled out by co-expression with empty vectors (**Supplemental Figure 3**). Yet, Spontaneous unspecific and irreversible interactions of the fragmented halves of YFP proteins may occasionally occur in the absence of interaction. As such, sporadic or infrequent weak fluorescence was not considered as a positive interaction in this study. Fluorescence complementation indicated that not only the vascular functional MpTMO5, but also KnTMO5 was able to interact with AtLHW (**Figure 5B, C**). These results indicate that the dimerization ability of TMO5 is highly conserved and even present in Charophytes. Therefore, dimerization of KnTMO5 was not the limiting factor to perform the vascular function. Instead, other factors, such as DNA binding or RNA polymerase recruitment may be prevented in KnTMO5 in the Arabidopsis system. In contrast to TMO5, dimerization ability of LHW was found to be less conserved. Both MpLHW and KnLHW were unable to form dimers with AtTMO5 (**Figure 5D, E**). It is thus likely that MpLHW and KnLHW are unable to restore the vascular cell file number of the *Arabisopsis lhw* mutant because they cannot interact with AtTMO5.
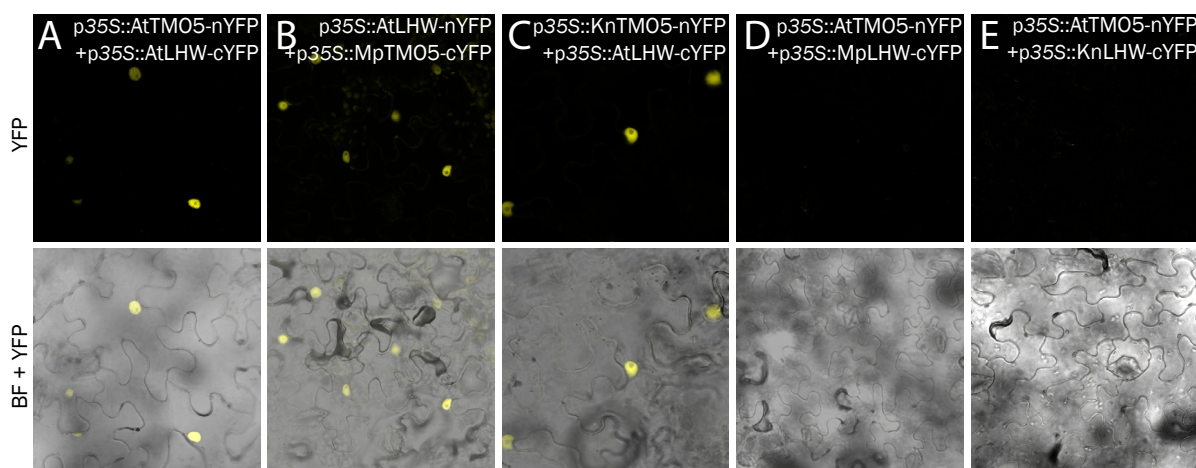
Figure 5. BiFC in *N. benthamiana* leaves reveals that TMO5, but not LHW, has a conserved dimerization capacity. Top panel: YFP signal, bottom pannel: YFP merged with Bright Field (BF). **(A)** Positive control of AtTMO5 and AtLHW. **(B and C)** BiFC of AtLHW with MpTMO5 (B) and KnTMO5 (C). **(D and E)** BiFC of AtTMO5 with MpLHW (D) and KnLHW (E).

Even though we did not detect interaction between AtTMO5 and LHW from Marchantia (MpLHW) or Klebsormidium (KnLHW) in our system, TMO5 and LHW may still have co-evolved as a heterodimer. An alternative hypothesis is that the two components have evolved separately to form a heterodimer in Tracheophytes. Given that our phylogenetic and complementation data demonstrated that TMO5 is more conserved in the plant lineage than LHW, we hypothesized that early diverging TMO5 proteins, might be functional as homodimers. In order to validate this hypothesis, we investigated the homo- and heterodimerization properties of TMO5 and LHW proteins in the plant kingdom. We selected Arabidopsis, Marchantia and Klebsormidium as representative species of Charophytes, Bryophytes and Angiosperms respectively. Not only are these model species with fully sequenced genomes, but these are species of interest since they are located on the border of the vascular function of either TMO5 or LHW. BiFC data revealed the homodimerization potential of TMO5 proteins in Klebsormidium and Marchantia (**Figure 6A**). In contrast, we were unable to detect homodimerization of AtTMO5, AtLHW and MpLHW proteins (**Figure 6A**). Intriguingly, BiFC data identified an interaction between TMO5 and LHW in Marchantia, but not in Klebsormidium, showing that heterodimerization properties of TMO5 and LHW were already established in Bryophytes, but most likely not in Charophytes (**Figure 6A**). Interestingly, while TMO5 and LHW orthologues in Klebsormidium are unable to heterodimerize,
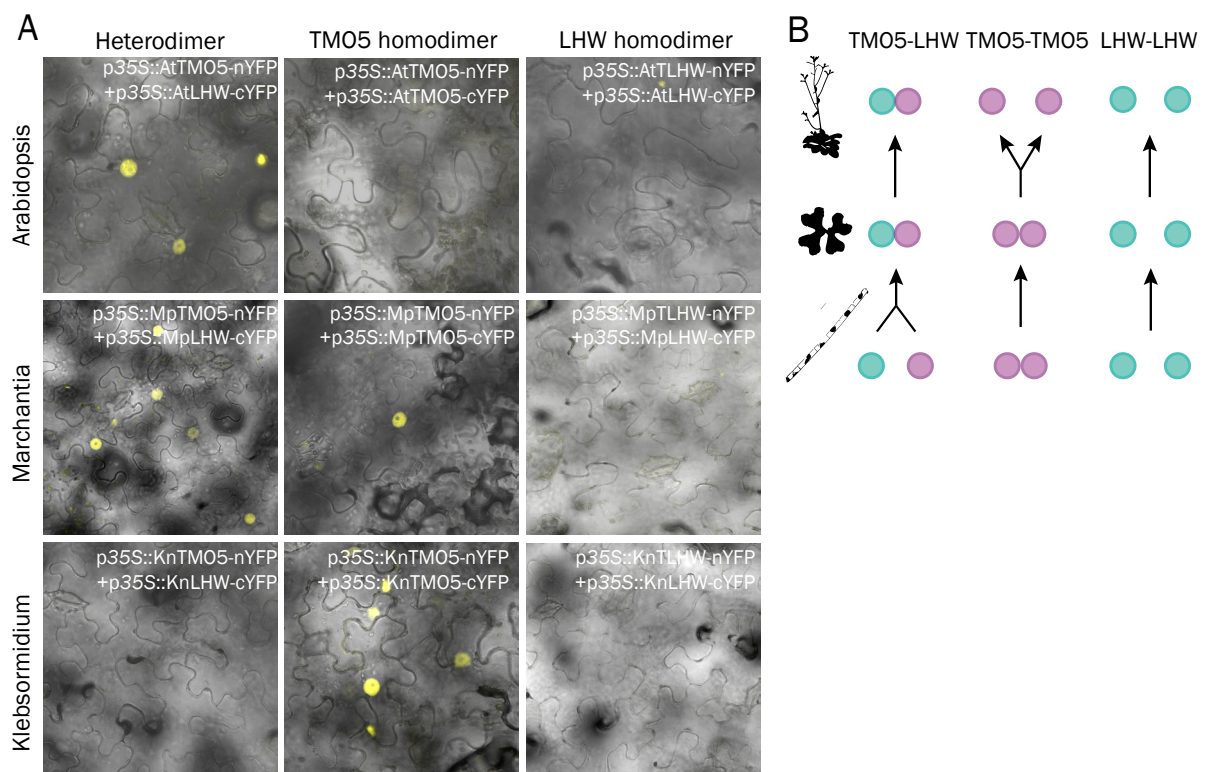
Figure 6. The evolution of dimerization capacities of TMO5 and LHW. (A) BiFC in *N. benthamiana* leaves of TMO5 and LHW hetero- and homodimerization capacity in Arabidopsis (top panel), Marchantia (middle panel) and Klebsormidium (bottom panel). **(B)** Model of TMO5 and LHW dimerization capacity over the course of evolution. Purple circles represent TMO5, green circles represent LHW.

KnTMO5 is able to form dimers with LHW of Marchantia (**Supplemental Figure 4**) and Arabidopsis (**Figure 5C**). This suggests that KnTMO5 already contains the ability to heterodimerize and that changes specifically to the *LHW* gene were responsible for the emergence of heterodimerization between TMO5 and LHW.

It should be noted that the strong *35S* promoter, used for BiFC analyses, results in artificial conditions in which high protein concentration may promote protein interaction that may not occur in their natural condition. However, this experiment suggests that MpTMO5 and MpLHW are capable of forming a heterodimer. Its biological relevance however, is yet to be determined.

In conclusion, our data suggests that the TMO5/LHW dimer was established in Bryophytes, prior to the emergence of the vascular function of LHW. We observe a trend in which TMO5 proteins evolve from a homodimer in Charophytes, towards a heterodimer with LHW in Bryophytes and eventually losing its homodimerization capacity in Tracheophytes (**Figure 6B**).

The Marchantia TMO5/LHW dimer cannot execute the vascular function

Previously we found that MpLHW cannot interact with AtTMO5 and therefore cannot induce cell proliferation in the vascular bundle of Arabidopsis. Knowing that TMO5 of Marchantia can interact with itself as well as with MpLHW, we hypothesized that MpLHW, together with MpTMO5, might be able to promote cell proliferation in Arabidopsis. In addition, we questioned whether MpTMO5 alone, as homodimer, could activate the downstream target At*LOG4*. To this end, we used a nuclear triple GFP (n3GFP) promoter fusion of the downstream TMO5/LHW target *LOG4* (p*LOG4*::n3GFP) in a wild type background. In the Arabidopsis root, *LOG4* expression is confined to the xylem, xylem pole pericycle and xylem pole endodermis cells (**Figure 7A**)(De Rybel et al., 2014). However, upon overexpression of the AtTMO5/AtLHW heterodimer, *LOG4* expression is triggered in the entire root meristem (**Figure 7B**) (De Rybel et al., 2014). A similar expression pattern is expected when overexpressing a homo- or heterodimer capable of performing the vascular function. When overexpressing MpTMO5 under the strong *35S* promoter, the *LOG4* expression pattern was extended, to some extent, within the vascular tissue of the root meristem, but not to the epidermis or cortex (**Figure 7C**). Given that no *LOG4* activation occurred outside the vasculature and MpTMO5 can interact with AtLHW, this extended vascular expression could be caused by a MpTMO5/AtLHW dimer activating *LOG4* in the overlapping region (the vascular tissue). From these results we can conclude that MpTMO5 can solely perform the vascular function as heterodimer.

In order to verify that MpLHW alone cannot form functional homodimers, we performed the same experiment with p*35S*::MpLHW. As expected, no extended *LOG4* expression was observed

(**Figure 7D**). We next crossed p*35S*::MpTMO5 with p*35S*::MpLHW, both in a p*LOG4*::n3GFP background, to obtain the double overexpression line p*35S*::MpTMO5 x p*35S*::MpLHW in p*LOG4*::n3GFP. In contrast to our expectations, co-expression of Mp*TMO5* and Mp*LHW* did not extent the expression pattern of p*LOG4* in the Arabidopsis root (**Figure 7E**). Thus, the Marchantia TMO5/LHW dimer is unable to promote cell proliferation in Arabidopsis. Strikingly, *LOG4* expression was more restricted in the double overexpression line compared to the single MpTMO5 overexpression. Possibly, MpTMO5 has a higher affinity to MpLHW than to AtLHW, creating a preference for MpTMO5/MpLHW heterodimers, which are incapable of activating the Arabidopsis *LOG4* promoter.
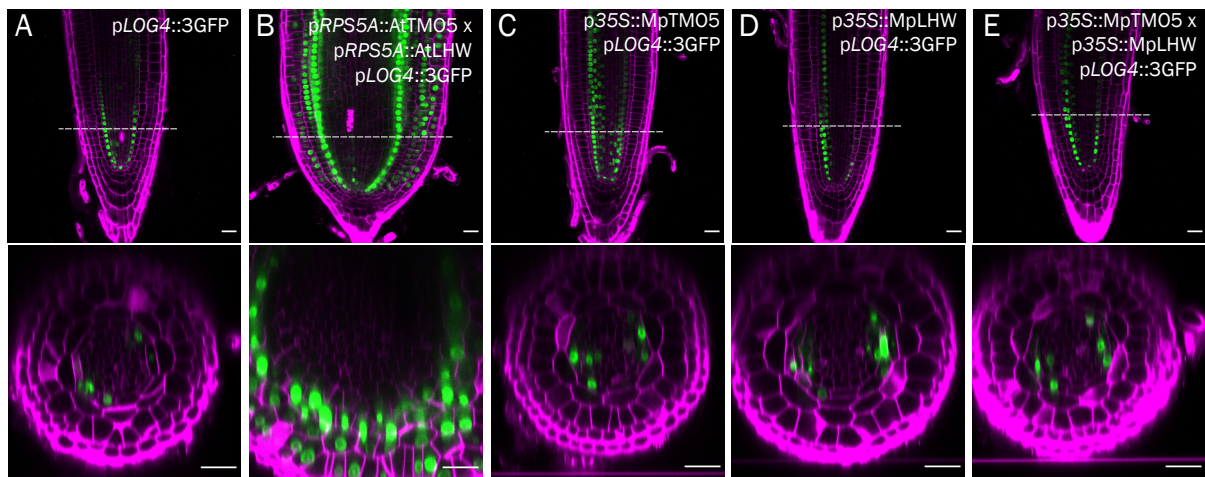
Figure 7. The Marchantia TMO5/LHW dimer cannot activate *LOG4* in Arabidopsis (A-E) Expression of p*LOG4*::n3GFP reporter in: wild type (A), AtTMO5/AtLHW overexpression (B), MpTMO5 overexpression (C), MpLHW overexpression (D) and in a MpTMO5 x MpLHW overexpression (E) root meristem. Dotted lines indicate the position of the optical radial section as shown in the bottom panels. Scale bars represent 20 μm.

Taken together, these results demonstrate that while TMO5 and LHW of Marchantia can heterodimerize, they cannot perform the same function as the Arabidopsis TMO5 and LHW. Given that the vascular function of TMO5 was already established in the ancestor of Marchantia and dimerization occurs between MpTMO5 and MpLHW, we can state that LHW did not obtain the vascular function prior to the emergence of Tracheophytes. The evolution of the LHW protein towards the vascular regulating protein thus occurred in two steps: first, LHW gained the ability to interact with TMO5. Subsequently, it obtained the activity to bind and regulate the *LOG4* promoter to induce periclinal and radial cell divisions.

Conservation of LOG protein sequence and structure, but not function, in plant ancestors

Within the vascular function of Arabidopsis, essential local cytokinin production is regulated by LOG3 and LOG4. TMO5 and LHW in Marchantia and Klebsormidium are unable to execute this vascular function. Consequently, we focused on the proteins further downstream: LOG3

and LOG4. Indeed, the *LOG* activation by TMO5/LHW may not be conserved, but the vascular function might still exist in Marchantia and Klebsormidium. Cytokinin as well as signaling pathway components have been detected in plants ranging from Angiosperms to Charophytes, including Klebsormidium (Hori et al., 2014; Stirk et al., 2013). Furthermore, LOG orthologues were identified in Tracheophytes, Bryophytes and Charophytes (Chapter 2). We thus set out to investigate whether the LOG protein family members have a conserved function as cytokinin-specific phosphoribohydrolases.

We first explored the conservation of LOG proteins in the plant lineage. Based on the sequence alignments, AtLOG4 shares a sequence identity of 57% with MpLOG and 54% with KnLOG (**Figure 8A**). In order to examine if structural features are also conserved between the LOG proteins, structural models of all three LOGs were constructed using Phyre2. Overall structures of AtLOG4, MpLOG and KnLOG were predicted to be nearly identical, with the exception of helix 3 of KnLOG (**Figure 8B-E**). In order to study whether the function of LOG proteins is conserved we used a misexpression assay in Arabidopsis. Upon misexpression of AtLOG4 in Arabidopsis,
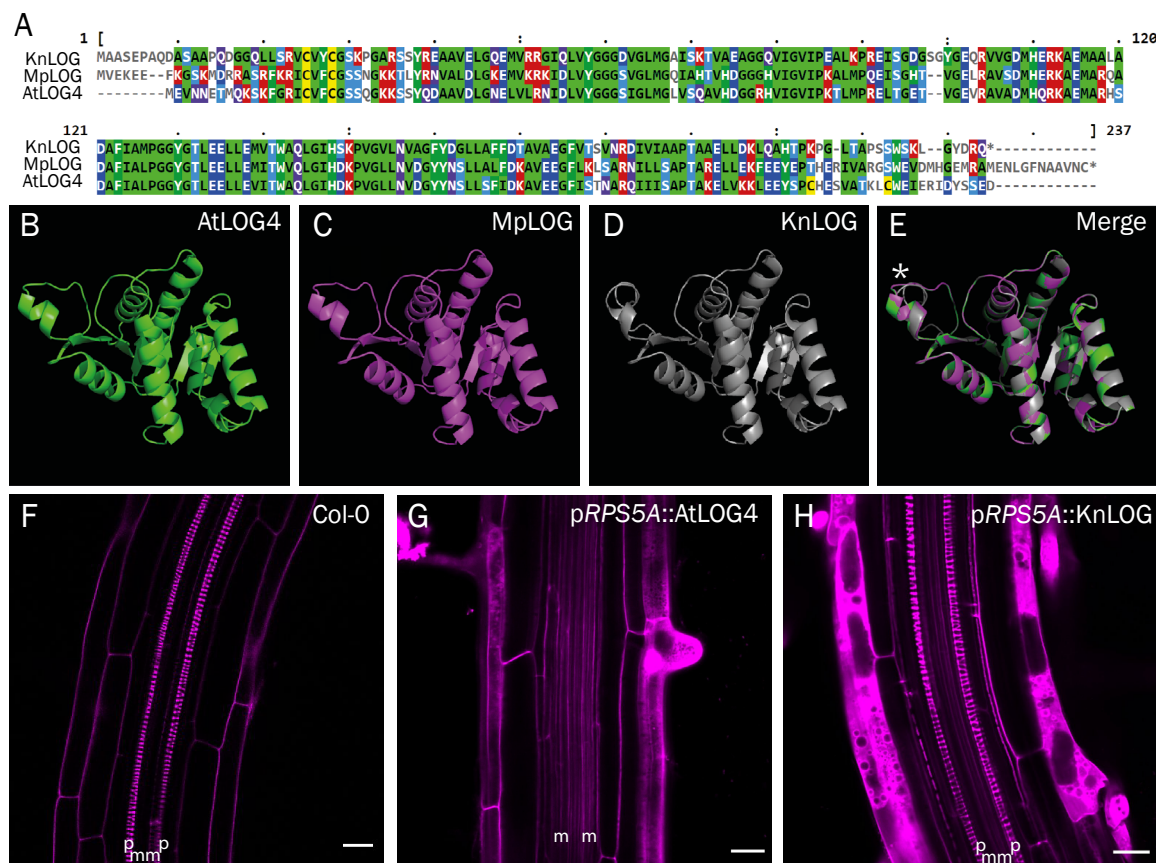
Figure 8. LOG proteins functionally diverged from Klebsormidium to Arabidopsis. (A) Sequence alignment of LOG proteins of Klebsormidium (KnLOG), Marchantia (MpLOG) and Arabidopsis LOG4 (AtLOG4) as displayed by the MView alignment. **(B-E)** Structural homology predictions of the aforementioned AtLOG4 (B), MpLOG (C) and KnLOG (D) and merged (E). The asterisk indicates the deviating helix of KnLOG. **(F-H)** Arabidopsis roots of wild type (F), AtLOG misexpression (G) and KnLOG misexpression (H). m: metaxylen, p: protoxylem. Scale bars represent 20 μm.

increased cytokinin activity inhibits protoxylem differentiation (**Figure 8G**) (De Rybel et al., 2014). We attempted to construct overexpression lines of the LOG protein of Marchantia and Klebsormidium, fused to a YFP (p*RPS5A*::MpLOG-YFP and p*RPS5A*::KnLOG-YFP respectively). However, Marchantia LOG overexpression lines were not obtained for technical reasons. When ectopically expressing KnLOG in Arabidopsis, proto- and meta xylem developed normally (**Figure 8H**). This suggests that the Arabidopsis LOG4 protein function, is not conserved in the early diverging charophytic LOG protein. It would be worthwhile to measure cytokinin levels in this system to confirm the inability of KnLOG to catalyze cytokinin production.

Discussion

One of the major elements controlling vascular cell proliferation in the Arabidopsis root is the TMO5/LHW pathway (De Rybel et al., 2013; Ohashi-Ito et al., 2013). While, well studied in Arabidopsis, little is known about the TMO5/LHW pathway in other plant species. In Chapter 2, we identified the heterodimer components TMO5 and LHW in a range of plant species. Surprisingly, *TMO5* and *LHW* orthologous genes are present in non-vascular species, such as the Bryophyte Marchantia and the Charophyte Klebsormidium. Here, we studied when the vascular function, i.e. the capacity to induce periclinal cell divisions in the vasculature, of TMO5 and LHW evolved. Our data demonstrated a two-step innovation process resulting in two dimer components capable of controlling vascular proliferation in Tracheophytes. The first step is the innovation of the TMO5 vascular function, which originated in the common ancestor of Bryophytes and Tracheophytes. Simultaneously, a shift from homo- to heterodimerization was initiated, allowing TMO5 and LHW to interact, while TMO5 homodimers were still able to form. The second functional innovation was acquired in the ancestor of Tracheophytes: the gain of LHW vascular function. This transitioned the TMO5/LHW dimer to a vascular tissue regulating complex in Tracheophytes.

In this chapter we showed successful inter-species complementation experiments, confirming the functionality of our experimental set-up. It is highly compelling that a TMO5 protein from the basal land plant Marchantia, lacking any complex vasculature, was able to perform a vascular specific function in Angiosperms. Thus, TMO5 function is highly conserved in Embryophytes. Furthermore, we can state that the TMO5/LHW pathway is not restricted to Angiosperms but is conserved in many, if not all, Tracheophytes. Functional divergence of paralogues, which can already be seen in the Arabidopsis paralogues LHW and LL1 (**Figure 3C**), makes identifying functionally conserved orthologues more challenging. This may explain why no vascular functional LHW proteins were identified in Populus and Picea. Alternatively, certain species within the Tracheophytes, evolved a novel vascular regulating pathway, independent of LHW. Another consideration to be taken into account is that Angiosperm platforms were used for all complementation assays. Certain

suppressor proteins may have evolved in Arabidopsis, or in Angiosperms in general, inhibiting the vascular function of earlier diverging TMO5 or LHW proteins, resulting in false negative outputs. Furthermore, Angiosperms may possess innovations or facilitating factors such as cofactors and enhancers that were not present in earlier diverging lineages and vice versa. Differentiated or newly formed cofactors or enhancers important for TMO5 and LHW function may not be compatible with early diverging orthologues and can thus result in false negative results.

When rescue of the *lhw* mutant was observed, quantification analysis of *lhw* complementation often revealed partial complementation. In general two populations, rather than a single cluster, was observed within independent lines. These two populations correlate to the monarch and diarch architectures observed in T2 complementation lines, where the cut-off value for a diarch structure is likely located around 32 vascular cell files. Given that vascular proliferation is a tightly regulated process, one could speculate that a certain cytokinin threshold needs to be reached in order to initiate the cytokinin response. Putative LHW orthologues, when bound to AtTMO5, may not have the same binding affinity to the *LOG* promoters as the endogenous AtTMO5/AtLHW dimer. Therefore, transcriptional initiation of *LOG3* and *LOG4* may be impaired resulting in limited LOGs and, consequently, limited cytokinin in the vasculature. When cytokinin levels oscillate around the cytokinin threshold, this could results in one monarch and one diarch population. This is one explanation of why, within independent lines, sometimes two populations exist.

Overall, our experimental data showed that the TMO5/LHW dimer did not co-evolve from the land plant ancestor. Indeed, Klebsormidium TMO5 and LHW orthologues do not possess the vascular function, nor the capacity to heterodimerize. *TMO5-* and *LHW* orthologous genes in Klebsormidium are most likely derived from ancestral genes from which *TMO5* and *LHW* later evolved. The ancestral TMO5 and LHW protein belong to the ten first emerging bHLH proteins (Jin et al., 2017). As such, these ancestral proteins may have played major regulatory roles in Charophyte algae. Moreover, the inability to heterodimerize strongly supports the notion that ancestral TMO5 and LHW performed independent function, either as homodimer or as heterodimer with another interaction partner. Further research is needed to reveal this ancestral function in the basal plant lineage.

Since *TMO5-* and *LHW-* like genes of Klebsormidium most likely performed independent functions, it is evident that these basal bHLH proteins underwent mutations and functional innovations during evolution. One of the main drivers of such functional innovation is gene duplication followed by functional divergence, a process called neofunctionalization. Neofunctionalization is assumed to be relatively common, since 90% of all eukaryotic genes are the result of gene duplication (Lynch and Conery, 2000; Teichmann and Babu, 2004). However, given that only single *TMO5* and *LHW* genes are present in Bryophytes and even in certain Lycophytes, neofunctionalization may not have

been the driver for TMO5 and LHW evolution. We speculate that TMO5 performed an essential, algae-specific function in the ancestor of land plants. During the evolution towards land plants, TMO5 and LHW may have escaped purifying selection as its alga-specific function was no longer essential, allowing gradual mutations to create functional divergence.

Simultaneously, dimerization innovation occurred through alterations in the *LHW* gene, allowing TMO5 and LHW to interact (**Figure 6B**). Such innovation combines different DNA-binding domains and may allow regulation of novel targets. Promoter activation assays in Arabidopsis however, demonstrated that Arabidopsis *LOG4* cannot be activated by the Marchantia dimer. This experiment was performed in an Angiosperm system using the Arabidopsis *LOG* promoter. Potential co-evolution of TMO5 and LHW with the *LOG* gene cannot be dismissed. Recently, RNA seq data revealed that upon overexpression of MpTMO5 and MpLHW in Marchantia, the endogenous LOG protein is not upregulated (Kuan-Ju Lu, unpublished results), concluding that dimerization capacity of TMO5 and LHW is probably not sufficient to bind and regulate *LOG* expression.

Nine different LOG proteins exist in Arabidopsis and all, except AtLOG6 and AtLOG9, play a central role in cytokinin activation (Tokunaga et al., 2012). Their functions are redundant and all seven At*LOG* genes have overlapping expression patterns (Kuroha et al., 2009). The phosphoribohydrolase activity of LOGs has previously been shown in *Medicago truncatula* and Oryza (Kurakawa et al., 2007; Mortier et al., 2014). However, we were unable to identify this conserved cytokinin-specific phosphoribohydrolase function of the LOG protein in the aquatic ancestor of Embryophytes, even though cytokinin signaling had evolved in Charophytes (Hori et al., 2014). Surprisingly, LOG orthologues were recently identified outside the plant kingdom, in almost all major lineages of prokaryotes (Naseem et al., 2015; Samanovic et al., 2015) as well as in the archaea *M. oralis* (Poehlein and Seedorf, 2016). Like plant and Charophyte LOG proteins, LOG orthologues in prokaryotes were shown to be structurally highly conserved. In contrast to KnLOGs, phosphoribohydrolase activity was detected in prokaryotic LOG proteins. One could speculate that the phosphoribohydrolase activity of LOG proteins evolved independently in prokaryotes and eukaryotes. A second, more appealing, hypothesis is that LOG proteins and their phosphoribohydrolase activity are ancient and highly conserved during evolution, but was lost in the Klebsormidium lineage, after it split from its ancestor. Indeed, recent trends in cytokinin studies allude to cytokinin as a cross-kingdom communication molecule (reviewed by Naseem et al., 2015). In order to elucidate the conservation of the LOG proteins in the cytokinin pathway, further analysis of LOG activity in Bryophytes or other early diverging lineages is needed.

One important notation is that not only protein sequence variation, but also orthologue expression patterns may be of importance for functional innovation of transcription factors such as TMO5 and

LHW. Recently it was shown that both *TMO5* and *LHW* are expressed ubiquitously in Marchantia and not restricted to, for example, water-coducting cells (WCCs) (Kuan-Ju Lu, unpublished results). Thus, the characteristic tissue-specific expression pattern of *TMO5* was not obtained prior to the emergence of Tracheophytes. As such, MpTMO5 may be involved in controlling general processes, rather than tissue-specific processes such as rhizoid- or gammae formation. Given the vascular-specific role of TMO5 and LHW in Tracheophytes, we expect that *TMO5* expression was restricted to vascular cells once vascular identity was established in Tracheophytes. It remains to be investigated when and how the expression patterns of the TMO5 transcription factor shifted.

Overall, our data support the notion that Tracheophyte ancestors acquired new regulatory components, i.e. TMO5 and LHW, to regulate the pre-existing cytokinin signaling cascade, potentially including LOG proteins. This process of re-purposing existing genetic units, also known as co-option, provides a relatively easy and elegant mechanism to create new transcriptional networks and contribute to morphological complexity (Pires et al., 2013). In conclusion, a step-wise functional innovation of TMO5 and LHW, achieving heterodimerization capacity, as well as the recruitment of the dimer to a novel pathway may have played pivotal roles in creating complex vascular tissues in Tracheophytes. Further work is however needed to investigate the ancestral role of TMO5 and LHW in Bryophytes and Charophytes.

Material and methods

Plant material and growth conditions

All seeds were surface sterilized and grown on ½ MS plates containing 1% sucrose and 0.8% Daishin agar (Duchefa) after a two day vernalization at 4ºC. 5-7 Day old seedlings were transferred to soil and grown under long day conditions (16 hours light, 8 hours dark) at 22ºC. Arabidopsis ecotype Columbia-0 (Col-0) was used as wild type. *lhw, lhw ll1, tmo5 and tmo5 t5l1* mutants as well as p*LOG4*::n3GFP lines were generated by and obtained from De Rybel et al. (2013, 2014)

Cloning and plant transformations

Complementation vectors expressing p*TMO5* or p*LHW* were constructed through conventional cloning of pPLV28. The *RPS5A* promoter in pPLV28 (Wendrich et al., 2015) was removed by flanking KpnI restriction sites and replaced by a fragment, 3 kb upstream of either *TMO5* or *LHW*. Promoter regions were amplified from existing plasmids. All further cloning procedures were performed using Seamless Ligation Cloning Extract (SLiCE), with 15 homologous bases (Zhang et al., 2015). Using primers with flanking LIC sites, a YFP was inserted in the LIC site of p*LHW*::LIC, creating p*LHW*::YFP.

cDNA of *Populus trichocarpa, Oryza sativa, Picea abies* and *Selaginella moellendorffii* was used to amplify orthologues, while MpTMO5, MpLHW, KnLHW, EgLHW, EdTMO5, SwLHW, CmaLHW and Cma2008862 were gene synthesized by GeneArt or GenScript. KnTMO5 CDS was amplified from genomic DNA (see below) by stitching PCR. The cDNAs of TMO5 orthologues were amplified and introduced in pGIIB-p*TMO5*::LIC-NOSt. The cDNAs of LHW orthologues, excluding the stop codon, were amplified and cloned into pGIIB-p*LHW*::YFP-NOSt. BiFC plasmids were constructed by cloning the orthologues into a modified pPLV22 or pPLV27 vector containing a p*35S*::LIC-nYFP or p*35S*::LIC-cYFP respectively. All constructs were verified by sequencing.

Complementation assays

For complementation assays, plasmids carrying the LHW orthologues were transformed into *lhw* plants by simplified floral dipping (De Rybel et al., 2011). Rescue was analyzed in T1 as well as T2 plants. Approximately 10 individual seedlings were screened per T2 line.

Plasmids carrying the TMO5 orthologues were transformed to *tmo5 t5l1* plants, homozygous for *tmo5*, heterozygous for *t5l1* (obtained by crossing *tmo5* mutant plants with *tmo5 t5l1* double mutant plants). In T1 lines, rescue was considered when 80% or more of the independent lines showed complementation. Homozygous *tmo5 t5l1* double mutant backgrounds were selected in MpTMO5 and KnTMO5 T2 lines by screening vascular pattern in seedlings that lost the transgene by segregation.

DNA extraction Klebsormidium

Genomic DNA was extracted from freshly grown Klebsormidium on BCD agar medium. Klebsormidium was collected with a 10 μl inoculation loop and inoculated in 200ul undiluted Edwards solution (200 mM Tris-HCl (pH 7.5), 250 mM NaCl, 25 mM EDTA, and 0.5% SDS) for 1 hour at 90°C. After centrifugation, the supernatant was diluted 20 times to obtain a working stock Klebsormidium DNA.

Bimolecular Fluorescence Complementation (BiFC)

Agrobacterium tumefaciens containing BiFC plasmids were cultured overnight at 28°C/250 rpm in 5ml LB medium containing 50 μg/ml kanamycin, 25 μg/ml rifampicin, 2 μg/ml tetracyclin and 200 μM acetosyringone. The bacteria were collected by centrifugation (4000rpm, 10 minutes) and resuspended in MMA infiltration medium (20 g/l sucrose, 5 g/l MS-salts, 2 g/l MES, pH 5.6) containing 200μM acetosyringone to an optical density (OD_{600}) of 0.3. BiFC samples were mixed in a 1:1 ratio to a total OD_{600} of 0.6. Samples were incubated for 1-2 hours at room temperature (RT) under continuous shaking. The abaxial side of the two youngest, fully expanded leaves of 5 –

6 week old *Nicotiana benthamiana* plants were infiltrated with a 1 ml needless syringe. Infiltrated leaves were harvested after 2 – 3 days and imaged with a confocal microscope.

Plant imaging

For confocal imaging of monarch or diarch root architecture, 5 – 7 day-old-seedlings were stained with 10 μg/ml Propidium Iodide. A Leica SP5 confocal microscope, containing an Argon and a DSS561 diode laser, was used for imaging. GFP, YFP and Propidium Iodide were excited at 488 nm, 514 nm and 561 nm respectively. Emission was visualized with Hybrid detectors at 500-550 for GFP, 520-550 for YFP and 600-700 for Propidium Iodide. When multiple fluorescent markers were imaged, sequential scanning was used.

Quantification of vascular cell numbers

5 day-old roots were fixed and stained using a modified Pseudo Schiff – Propidium Iodine (mPS-PI) staining (Truernit et al., 2008). In short, whole seedlings were fixed overnight in fixative (50% methanol and 10% acetic acid) at 4ºC, rinsed with MQ and incubated in 1% periodic acid at RT for 40 min. After a second washing step, seedlings were placed in Schiff reagent (100 mM sodium metabisulphite and 0.15 N HCl) containing 15 mg/ml Propidium Iodide for 1 – 2 hours. The samples were transferred onto a microscope slides and covered with a chloral hydrate solution (4 g chloral hydrate, 1 mL glycerol, and 2 mL water). Two small coverslips (20x20) we placed on the sides of the slide, with a larger coverslip (20x40) on top covering the samples, to avoid crushing the roots. Slides were kept overnight in the dark at RT prior to imaging. Radial cross section were obtained by a SP5 confocal microscope. All radial sections were taken in the middle of the root meristem: half way between the QC and the first elongating cortex cell. The data was visualized using BoxPlotR.

Homology modeling

Sequence alignments were constructed using ClustalO and visualized using MView. Structural homology predictions through homology modeling was done using Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2). Resulting protein models were studied and imaged using Pymol.

Acknowledgments

Supplemental information
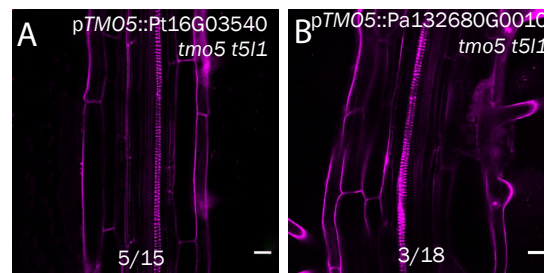
Supplemental Figure 1. Complementation assay of the Arabidopsis *tmo5 t5l1* double mutant with the Populus gene Pt16G03540 (A) and the Picea gene Pa132680G0010 (B). Numbers at the bottom indicate the number of times the represented xylem pattern was observed per total lines screened. Note that heterozygous lines were used in this study (see Material and Methods for details). Scale bars represent 20 μm.
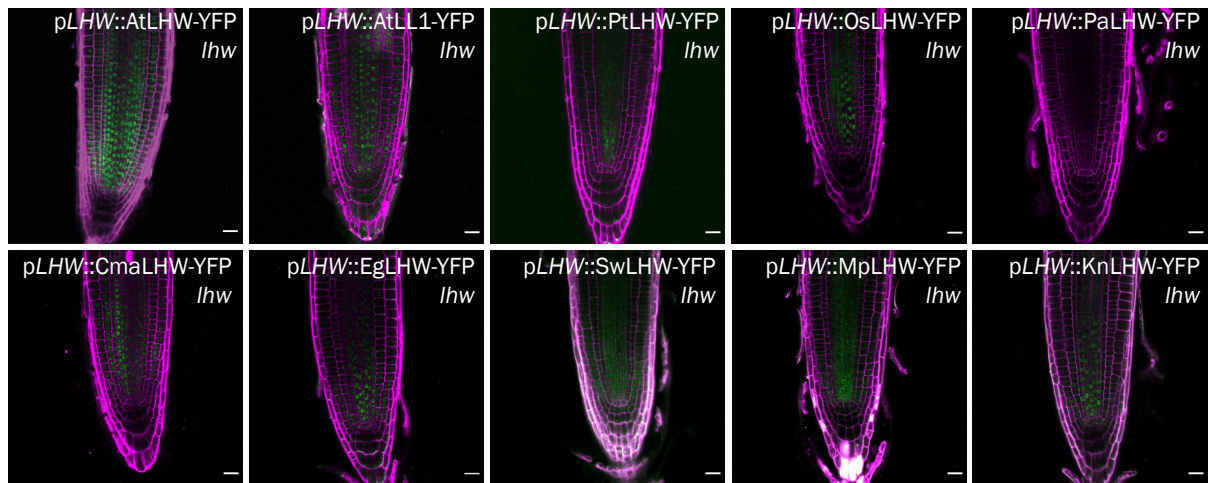
Supplemental Figure 2. Expression of LHW, LL1 and their orthologues driven from the native p*LHW* promoter in *lhw* mutants.
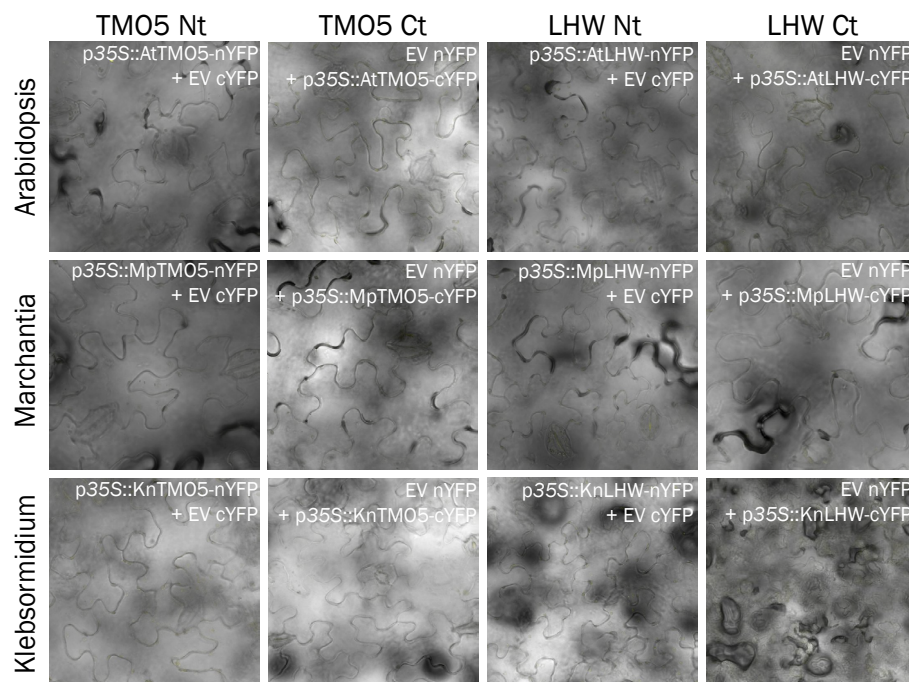
Supplemental Figure 3. BiFC negative controls of Arabidopsis, Marchantia and Klebsormidium TMO5 and LHW. Fusions to nYFP or cYFP were tested with Empty Vector (EV) cYFP or nYFP respectively.

p35S::KnTMO5-nYFP + p35S::MpLHW-cYFP

YFP	BF + YFP
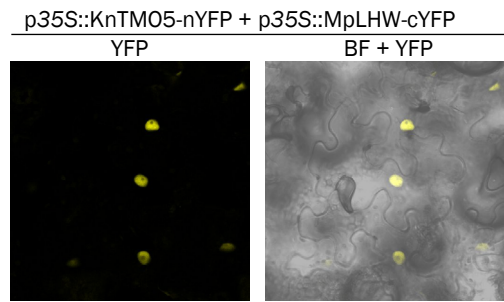
Supplemental Figure 4. BiFC of Klebsormidium TMO5 (KnTMO5) with Marchantia LHW (MpLHW). YFP signals are sown in the left panel. A YFP and Bright Field (BF) merge is shown in the right panel.

Supplemental tabel 1. Percentage rescue in T2 lines of *tmo5 t5l1* and *lhw* Arabidopsis mutants by TMO5 and LHW orthologues respectively.

	Line	% complementation	n
LHW	2	100%	10
	4	100%	10
	5	100%	10
	6	100%	10
LL1	1	0%	10
	2	0%	10
	3	0%	10
	4	0%	10
	5	10%	10
	6	0%	10
	7	0%	10
	8	0%	10
	9	0%	10
	11	0%	10
	13	0%	10
	14	0%	10
	16	0%	10
	17	0%	10
	18	0%	10
OsLHW	1	50%	10
	2	50%	10
	3	40%	10
	4	30%	10
	5	60%	10
	6	0%	10
	7	40%	10
	9	80%	10
	10	20%	10
	11	60%	10
	14	60%	10
	15	90%	10
	16	30%	10
	17	40%	10
	18	0%	10
	19	80%	10

	Line	% complementation	n
PaLHW	19	0%	10
	20	0%	10
	22	0%	10
	23	0%	10
	24	0%	10
	26	0%	10
	27	0%	10
	28	0%	10
	29	0%	10
	30	0%	10
	31	0%	10
	32	0%	10
EgLHW	1	70%	10
	2	92%	12
	3	90%	10
CmaLHW	1	10%	10
	2	90%	10
	3	100%	10
	4	100%	10
	5	50%	10
	6	50%	10
	7	50%	10
	9	100%	10
	10	50%	10
	11	50%	10
	12	90%	10
	13	100%	10
	14	90%	10
	15	20%	10
	16	10%	10
	18	70%	10
	21	30%	10
	22	10%	10
	28	40%	10
	29	40%	10
	31	10%	10

	Line	% complementation	n
SwLHW	1-33 (T1)	9%	33
MpLHW	2	0%	10
	7	0%	10
	8	0%	10
KnLHW	7	0%	9
SmTMO5	8	100%	10
	9	100%	10
	11	100%	10
MpTMO5	7	100%	10
	8	100%	10
	9	100%	10
	10	95%	10
	11	75%	10
	13	100%	10
	23	100%	10
	24	95%	20
	26	100%	10
KnTMO5	15	0%	2
	18	0%	10
	30	0%	10
	60	0%	10
	68	0%	10

Supplemental tabel 2. Primers used in this study.

Primer pairs used for complementation assays

Gene number	Gene name	Organism	Primer orientation	Sequence
AT3G25710	AtTMO5	*A. thaliana*	Sense	5'-TAG TTG GAA TAG GTT C ATG TAC GCA ATG AAA GA
			Antisene	5'-AGT ATG GAG TTG GGT TC CTA ATT ATA ACA TCG ATT CAC
Pt08G11600	PtTMO5	*P. tremulus*	Sense	5'-TAG TTG GAA TAG GTT C ATG TGT GGA CTT AAA GAA
			Antisene	5'-AGT ATG GAG TTG GGT TC GGT ACT CAT GGT CTC
Os03G15440	OsTMO5	*O. sativa*	Sense	5'-TAG TTG GAA TAG GTT CAT GGG TGC TCA TGG AGA C
			Antisene	5'-AGT ATG GAG TTG GGT TCT TAG ATA GAC CTG TTC TC
Pa65818g0010	PaTMO5	*P. abies*	Sense	5'-TAG TTG GAA TAG GTT ATG GAA CGC TGT TTC GCT GCT G
			Antisene	5'-AGT ATG GAG TTG GGT TTC AAC CTA CCT GAC GTT GTC TTT T
Sm00038G00580	SmTMO5	*S. moellendorffii*	Sense	5'-TAG TTG GAA TAG GTT CAT GAC CCC GCA AGA GAT A
			Antisene	5'-AGT ATG GAG TTG GGT TCT TAG TAG GAC GAG GAG GA
Mp2020606/ Mapoly0039s0068	MpTMO5	*M. polymorpha*	Sense	5'-TAG TTG GAA TAG GTT ATG GGT GGA GAA GCA ACG GGG CGC
			Antisene	5'-AGT ATG GAG TTG GGT T TTA CAT CGA AGG GCT CGA CGA ATC
Kn000610230	KnTMO5	*K. nitens*	Sense	5'-TAG TTG GAA TAG GTT C ATG TGT GGA CTT AAA GAA
			Antisene	5'-AGT ATG GAG TTG GGT TC GGT ACT CAT GGT CTC
Pt16G03540	-	*P. tremulus*	Sense	5'-TAG TTG GAA TAG GTT C AAC ATG CAG CCT GAA
			Antisene	5'-AGT ATG GAG TTG GGT TC ACC AGA CTG TAA CAC C
Pa132680g0010	-	*P. abies*	Sense	5'-TAG TTG GAA TAG GTT ATG AGA GGG TTC AGA GAT CAA
			Antisene	5'-AGT ATG GAG TTG GGT TTT AAA TGG CAG AGC TTG TTG CAT C
AT2G27230	AtLHW	*A. thaliana*	Sense	5'-TAG TTG GAA TAG GTT ATG GGA GTT TTA CTA AGA GAA GCT TTA GGG
			Antisene	5'-CTT GCT CAC CAT GTT CAT TGA ACA GCC ACC AGT AAC CGG T
AT1G64625	AtLL1	*A. thaliana*	Sense	5'-TAG TTG GAA TAG GTT ATG GGT TCA GAG TAT AGC
			Antisene	5'-CTT GCT CAC CAT GTT TGA TAA TAA ATC ATC
Os01G64560	OsLHW	*O. sativa*	Sense	5'-TAG TTG GAA TAG GTT GTC AAA TCC CAT GGC GGG CGC CGC C
			Antisene	5'-CTT GCT CAC CAT GTT CAA GCA GTT GGA GAA ACC AGA GGC T
Eg1939	EgLHW	*E. giganteum*	Sense	5'-TAG TTG GAA TAG GTT ATG GCT CTG CAG
			Antisene	5'-CCC TTG CTC ACC ATG TTT TGC AAC A
Cma2009894	CmaLHW	*C. macrocarpa*	Sense	5'-TAG TTG GAA TAG GTT ATG GCT CTT CAT CAG A
			Antisene	5'-CTT GCT CAC CAT GTT AGC ATG AAG TGA TGG
Sw2182007	SwLHW	*S. wallacei*	Sense	5'-TAG TTG GAA TAG GTT ATG GCT GAG GTG CTC AAG CA
			Antisene	5'-CTT GCT CAC CAT GTT CCA GTT AGC CGA GGA CAC CA
Mp2039322/ Mapoly0088s0049	MpLHW	*M. polymorpha*	Sense	5'-TAG TTG GAA TAG GTT ATG GCA ATG GTG CTG CAG CAG GCG T
			Antisene	5'-CTT GCT CAC CAT GTT TCG GGC TGT CAT CTG GAG AGG TAC A
Kn001060270	KnLHW	*K. nitens*	Sense	5'-TAG TTG GAA TAG GTT ATG TCC TTA GTG CTG CAG CA
			Antisene	5'-CTT GCT CAC CAT GTT GTA CGG ATC GTC CAG GTA CA
Pt01G21690	-	*P. tremulus*	Sense	5'-GCT AGT TGG AAT AGG TTA TGC AGA CAA TTG CTG TTA TTC CTG
			Antisene	5'-CTT GCT CAC CAT GTT CTG CAA ACT ACT AGG CAT ACC AGT T
Pa45581g0020	-	*P. abies*	Sense	5'-TAG TTG GAA TAG GTT ATG CAG ATC ATG GAG AAT TTG GAG
			Antisene	5'-CTT GCT CAC CAT GTT TTC ATA GGC AGA TCC CAG CAG ATC
Cma2008862	-	*C. macrocarpa*	Sense	5'-TAG TTG GAA TAG GTT ATG GCA GTT GCT TCT TG
			Antisene	5'-CTT GCT CAC CAT GTT GAT AGA AAA ACA T

Primer pairs used for BiFC cloning

Gene number	Gene name	Organism	Primer orientation	Sequence
AT3G25710	AtTMO5	*A. thaliana*	Sense	5'-TAG TTG GAA TAG GTT C ATG TAC GCA ATG AAA GA
			Antisene	5'-GTA TGG AGT TGG GTTC AT TAT AAC ATC GAT TCA CCA TC
AT2G27230	AtLHW	*A. thaliana*	Sense	5'-TAG TTG GAA TAG GTT C ATG GGA GTT TTA CTA AGA GA
			Antisene	5'-GTA TGG AGT TGG GTT CCA TTG AAC AGC CAC CAG TAA CC
Mp2020606/ Mapoly0039s0068	MpTMO5	*M. polymorpha*	Sense	5'-TAG TTG GAA TAG GTT ATG GGT GGA GAA GCA ACG GGG CGC
			Antisene	5'-GTA TGG AGT TGG GTTC CAT CGA AGG GCT CGA CGA ATC
Mp2039322/ Mapoly0088s0049	MpLHW	*M. polymorpha*	Sense	5'-TAG TTG GAA TAG GTT ATG GCA ATG GTG CTG CAG CAG GCG T
			Antisene	5'-GTA TGG AGT TGG GTT CTC GGG CTG TCA TCT GGA GAG
Kn000610230	KnTMO5	*K. nitens*	Sense	5'-TAG TTG GAA TAG GTT ATG GAG AGC TCA AAA TCT GG
			Antisene	5'-GTA TGG AGT TGG GTTC CGA GGC GGT CAC TGT TTG TCC
Kn001060270	KnLHW	*K. nitens*	Sense	5'-TAG TTG GAA TAG GTT ATG TCC TTA GTG CTG CAG CA
			Antisene	5'-GTAT GGA GTT GGG TTC GTA CGG ATC GTC CAG GTA CAT G

Other primer pairs

Gene number	Gene name	Organism	Primer orientation	Sequence
kn006470030	KnLOG	*K. nitens*	Sense	5'-TAG TTG GAA TAG GTT ATG GCT GCC TCG GAG CCT GCT
			Antisene	5'-GTA TGG AGT TGG GTT CTG CCT GTC GTA GCC CAG CTT
AT2G27230	AtLHW(-GFP)	*A. thaliana*	Sense	5'-TAG TTG GAA TAG GTT ATG GGA GTT TTA CTA AGA GAA GCT TTA GGG
			Antisene	5'-TCC TCG CCC TTG CTC ACC ATC ATT GAA CAG CCA CCA G
-	(AtLHW-)GFP	-	Sense	5'-CTG GTG GCT GTT CAA TGA TGG TGA GCA AGG GCG AGG A
			Antisene	5'-GTA TGG AGT TGG GTT TTA CTT GTA CAG CTC GTC CAT GCC G

3

Chapter 4

Evolution of domain architecture in TMO5 and LHW proteins

Nicole van 't Wout Hofland[1], Paul Abrahams[1] Dolf Weijers[1] and Bert De Rybel[1,2,3]

1. Wageningen University, Laboratory of Biochemistry, Stippeneng 4, 6708 WE Wageningen, the Netherlands

2. Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, 9052 Ghent, Belgium

3. VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium

Abstract

Insight into protein evolution can be achieved by studying structural and functional units of proteins: the protein domains. Most proteins are comprised of one or more distinct domains in a specific arrangement (domain architecture). Changes in domain architecture by recombination of existing domains is an efficient method by which novel protein functions develop. As such, analyses of protein domain architecture dynamics can provide insight in the protein function evolution. Previously, we identified functional innovations of TARGET OF MONOPTEROS 5 (TMO5) and LONESOME HIGHWAY (LHW) in the ancestors of Charophytes and Bryophytes respectively (Chapter 3). Here, we investigate how these novel functions emerged, by studying the domain architectures of TMO5 and LHW. The ancestral LHW proteins (containing a GAF-like and a bHLH domain) gained an ACT-like domain in the ancestor of Tracheophytes. We show that this ACT-like domain, together with the bHLH domain, is necessary and sufficient for the LHW function. However, gain of the ACT-like domain in the ancestor of Tracheophytes was not the sole determinant for the development of the vascular function. In contrast to LHW, the domain architecture of TMO5 (consisting of a bHLH and an ACT-like domain) is conserved from Charophytes to Tracheophytes. We report that in algal ancestors of Embryophytes, mutations in the DNA-binding domain (DBD) of the ancestral TMO5 resulted in the vascular function of this protein.

Introduction

In the previous chapters, we studied the evolutionary course of the vascular tissue regulators *TARGET OF MONOPTEROS 5* (*TMO5*) and *LONESOME HIGHWAY* (*LHW*). The origin of these genes was found in the common ancestor of green algae (Charophytes) and land plants (Embryophytes). However, the vascular function that these proteins perform in Arabidopsis is not encoded in the Charophycean proteins, and was therefore likely not yet established in the ancestral gene. Functional innovation of TMO5 occurred around 530 million years ago, in the common ancestor of all land plants, resulting in a TMO5 protein able to induce periclinal and radial cell divisions in the vascular tissues of Arabidopsis. *LHW* orthologous genes did not obtain the vascular function until approximately 460 million years ago, in the ancestor of Tracheophytes. The mechanism behind the evolutionary innovations are, however, still elusive.

Proteins, such as TMO5 and LHW, consist of one or more discrete domains, connected by linker regions. Domains were first described in the 1960's, when distinct regions were identified in the three-dimensional structure of a ribonuclease (Kartha et al., 1967). Domains are now defined as spatially distinct conserved modules of 35 to 250 amino acids that can function and fold independently in isolation (Ponting and Russell, 2002). Most eukaryotic proteins are built up of multiple domains (Tordai et al., 2005). Their arrangement, in a N- to C-terminal direction, is termed "domain architecture". The emergence of new domain architectures is thought to be a major mechanism to generate new protein functions or molecular connections (Fong et al., 2007; Hegyi and Gerstein, 2001). Such novel organizations may arise from mutations in the amino acid sequence or through domain rearrangements. Domain rearrangements result from events such as recombination (fusion and fission), insertions, deletions and domain-shuffling. Proteins with identical domain architectures are likely to have similar structures and function. Therefore, studying the conservation of these architectures provides a helpful tool to investigate the evolution of protein functions.

An essential domain present in all transcription factors (TFs) is the DNA-binding domain (DBD), since transcription factors interact with *cis*-regulatory elements in the promoter sequence of a target gene to control gene expression. Thus, TFs must recognize specific nucleotide sequences in the double helical DNA structure. Within the DNA, each nucleotide exposes part of its characteristic groups to the outside of the double helix, in the major and minor grooves (Alberts et al., 2008). As such, proteins can recognize sequence information in the DNA without opening the helical structure (Alberts et al., 2008). Given that most of the DNA only allows moderate degree of bending, proteins often conform to the DNA. This mostly occurs by entering the DNA groove via an α-helical part of the protein. In addition, α-helices are the most common structural element for nucleotide recognition. Proportions of such helices are optimal for the presentation of amino acid

side chains (Church et al., 1977). Maximum contact and thus specificity can be achieved when the α-helix is inserted into the major groove with a 90 degree angle between the α-helical and DNA axis (Suzuki and Gerstein, 1995). The length of DNA-contacting DBDs depends on multiple factors, such as its conformation and monomeric, dimeric or oligomeric state. Typically, the DBD interacts with four to ten base pairs (reviewed by Garvie and Wolberger, 2001). These interactions are mediated by hydrogen bonds, electrostatic, van der Waals and hydrophobic interactions. In addition, contact is made with the negative phosphates of the DNA backbone. However, the majority of sequence specificity is thought to be decided by hydrogen bond interactions.

To date, the only domain that has been identified in TMO5 and LHW proteins is the domain responsible for DNA interaction: the basic Helix-Loop-Helix (bHLH) domain. bHLH domains contain two distinct regions: a ~15 amino acid basic stretch at the N-terminal region of the bHLH, which functions as the DBD, and two amphipathic helices of the HLH, which induce dimerization (Atchley et al., 1999). bHLH proteins are thought to bind to a consensus core element called the E-box (CANNTG). TMO5 and LHW for example, bind a DNA region containing the most common E-box form: the G-box (CACGTG), in the promoter of their target gene *LOG4* (De Rybel et al., 2014).

In this chapter we scan the TMO5 and LHW protein sequences for other protein domains and we study the domain architecture dynamics of these proteins and their orthologues. We show that while a transition of the LHW domain architecture occurred in the course of plant evolution, this alone cannot explain the functional innovation of LHW observed in Tracheophytes. Furthermore we demonstrate that, substitution in the DBD of TMO5 led to the acquisition of the TMO5 vascular function in the ancestor of Embryophytes.

Results

LHW, but not TMO5, domain architecture evolved during plant evolution

In Chapter 3, the functional origin of the TMO5 and LHW vascular function was found to date back to Bryophytes and Lycophytes, respectively. Given that domain architecture modification is one of the mechanisms leading to altered protein function, we investigated whether domain rearrangements were responsible for the gain of vascular function of TMO5 and LHW. No domains other than the bHLH domain have been reported in TMO5 or LHW (De Rybel et al., 2013; Ohashi-Ito and Bergmann, 2007). In order to assess the domain architecture, as well as the level of domain architecture conservation among orthologues, we used Phyre2 algorithms (Kelley and Sternberg, 2009). We identified a bHLH domain and a downstream ACT-like domain in the

C-terminal region of AtTMO5 (**Figure 1**). The same bHLH and ACT-like domains were found to be present in the C-terminal region of AtLHW. In addition, a GAF-like domain is localized at the N-terminus of AtLHW (**Figure 1**). We next compared the domain architecture of TMO5 orthologues to assess conservation of these newly identified domains. Interestingly, domain architecture of all TMO5 orthologues were identical, i.e. all orthologues, contained a bHLH domain followed by an ACT-like domain (**Figure 1, Supplemental Figure 1**). These results indicate that domain recruitment was not responsible for the gain of vascular function of TMO5 (**Figure 1**). In contrast, dynamics in the domain architecture was identified within LHW evolution: the LHW orthologue in Charophytes, liverworts and mosses lack an ACT-like domain (**Figure 1, Supplemental Figure 2**). Intriguingly, gain of the ACT-like domain coincides with the emergence of vascular function, raising the question if recruitment of the ACT-like domain might have been the driver for this functional innovation.
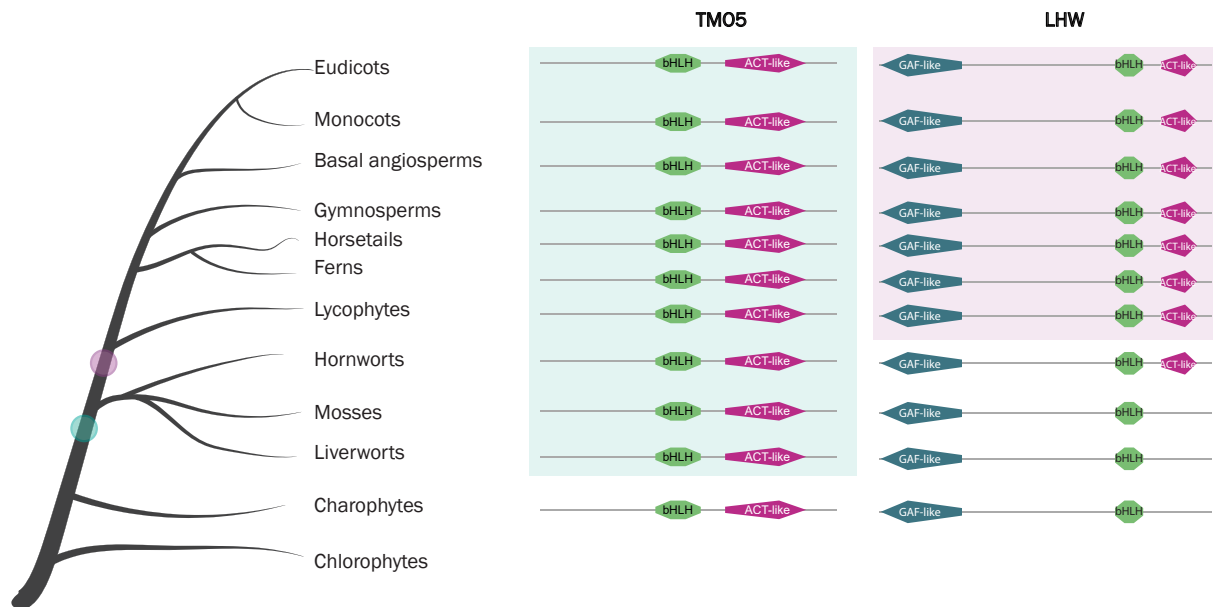
Figure 1. Domain architecture evolution of TMO5 and LHW proteins from Charophytes to Angiosperms. A phylogenetic tree illustrating the phylogenetic relationships is depicted on the left. The emergence of the vascular function of TMO5 and LHW are marked in green and purple respectively. bHLH domains are depicted in green, ACT-like domains in magenta and GAF-like domains in blue.

The bHLH and ACT-like domain of LHW are necessary and sufficient for *in vivo* function

In order to study if any of the identified domains are key to the *in vivo* function of LHW, we constructed domain deletions and truncated versions of AtLHW, fused to a YFP, driven by its native promoter (p*LHW-LHW*truncation-YFP; **Figure 2A**). Functionality was determined by genetic complementation of the Arabidopsis *lhw* single mutant and quantification of the number of vascular cell files in the root meristem (**Figure 2, Supplemental Table 1**). In all lines, protein expression in the root meristem was confirmed by YFP fluorescence (**Figure 2D – K**, inset, **Supplemental Figure 3**). As shown previously, the full-length AtLHW protein was able to induce cell proliferation

in the vasculature and restore the diarch pattern (**Figure 2B – D, M, Supplemental Table 1**). In contrast, LHW proteins with a deletion of the bHLH domain (LHW$^{\Delta bHLH}$) or the ACT-like domain (LHW$^{\Delta ACT}$) failed to rescue the *Arabidopsis* mutant phenotype (**Figure 2E – F, Supplemental Table 1**). Moreover, vascular cell file numbers of LHW$^{\Delta bHLH}$ and LHW$^{\Delta ACT}$ were similar to those of the *lhw* mutant (**Figure 2M**). While several roots of LHW$^{\Delta bHLH}$ line 4 showed some proliferation in the vascular tissue, this did not result in an overall significant difference with the *lhw* mutant. Similar to the aforementioned truncations, the bHLH domain of LHW alone (LHW^{bHLH}) was unable to induce periclinal cell divisions and rescue the mutant phenotype (**Figure 2G, L, M**). It should be noted that this partial protein was not detected in the nucleus (**Figure 2G**, inset, **Supplemental Figure 3**). Therefore, its functionality as a transcription factor could not be assessed. Deletion of the N-terminal GAF-like domain (LHW$^{\Delta GAF}$) could partially restore the reduced vascular bundle size (**Figure 2M**). The majority of roots of line 5 and 28, displayed a significant increase in vascular cell file numbers compared to the mutant. In accordance with this data, we observed that nine of the 31 independent lines were able to restore the diarch phenotype (**Figure 2H**). Deletion of the approximately 300 amino acid linker between the GAF-like and bHLH domain (LHW$^{\Delta linker}$) could restore the diarch phenotype in four out of ten independent lines (**Figure 2I**). Quantification analysis revealed that wild-type vascular cell numbers were reached in certain lines, while in others no vascular proliferation was observed (**Figure 2M**). Indeed, vascular cell file numbers of three lines (2, 6 and 7) were indistinguishable from wild type roots, while two lines (8 and 9) were identical to the *lhw* mutant. We observed that, for this construct, protein expression levels were correlated with vascular proliferation, i.e. lines with low LHW$^{\Delta linker}$-YFP protein levels were unable to induce periclinal cell divisions. Intriguingly, solely the bHLH domain and the ACT-like domain (LHW$^{\Delta GAF\Delta linker}$) were sufficient for LHW function in three out of fifteen independent lines and showed increased vascular cell file levels (**Figure 2J, M**). Finally, removal of the 28 amino acid C-terminal tail of LHW (LHW$^{\Delta C\text{-terminus}}$) did not affect vascular function and fully rescued the mutant monarch phenotype as well as the vascular cell file number (**Figure 2K, M**). From these analyses, we infer that the bHLH and ACT-like domains are necessary and sufficient for LHW protein function.

Figure 2. Domain analysis of the Arabidopsis LHW protein. ►
(A) Schematic representation of LHW protein domain deletion constructs expressed as YFP fusions and driven by the *LHW* promoter. bHLH domains are depicted in green, ACT-like domains in magenta and GAF-like domains in blue. The scale bar represents 50 amino acids. **(B)** Arabidopsis wild type root illustrating a diarch phenotype. **(C)** Arabidopsis *lhw* mutant with a monarch phenotype. **(D-K)** Complementation attempts of the *lhw* mutant using full length LHW (D), LHW$^{\Delta ACT}$ (E), LHW$^{\Delta bHLH}$ (F), LHW^{bHLH} (G), LHW$^{\Delta GAF}$ (H), LHW$^{\Delta linker}$ (I), LHW$^{\Delta GAFlinker}$ (J), LHW$^{\Delta C\text{-terminus}}$ (K) arrowheads mark protoxylem strands. **(L)** Representative radial sections of 5 day-old roots of Arabidopsis *lhw* complementation lines with LHW domain deletions. Asterisks indicate the endodermis. Scale bars represent 20 μm. **(M)** Vascular cell file quantification in radial sections of 5 day-old roots of the aforementioned domain deletions. Line numbers are indicated. Turquoise circles indicate individual roots. Lines indicated with different letters are significantly different (one-way ANOVA, P<0.05).
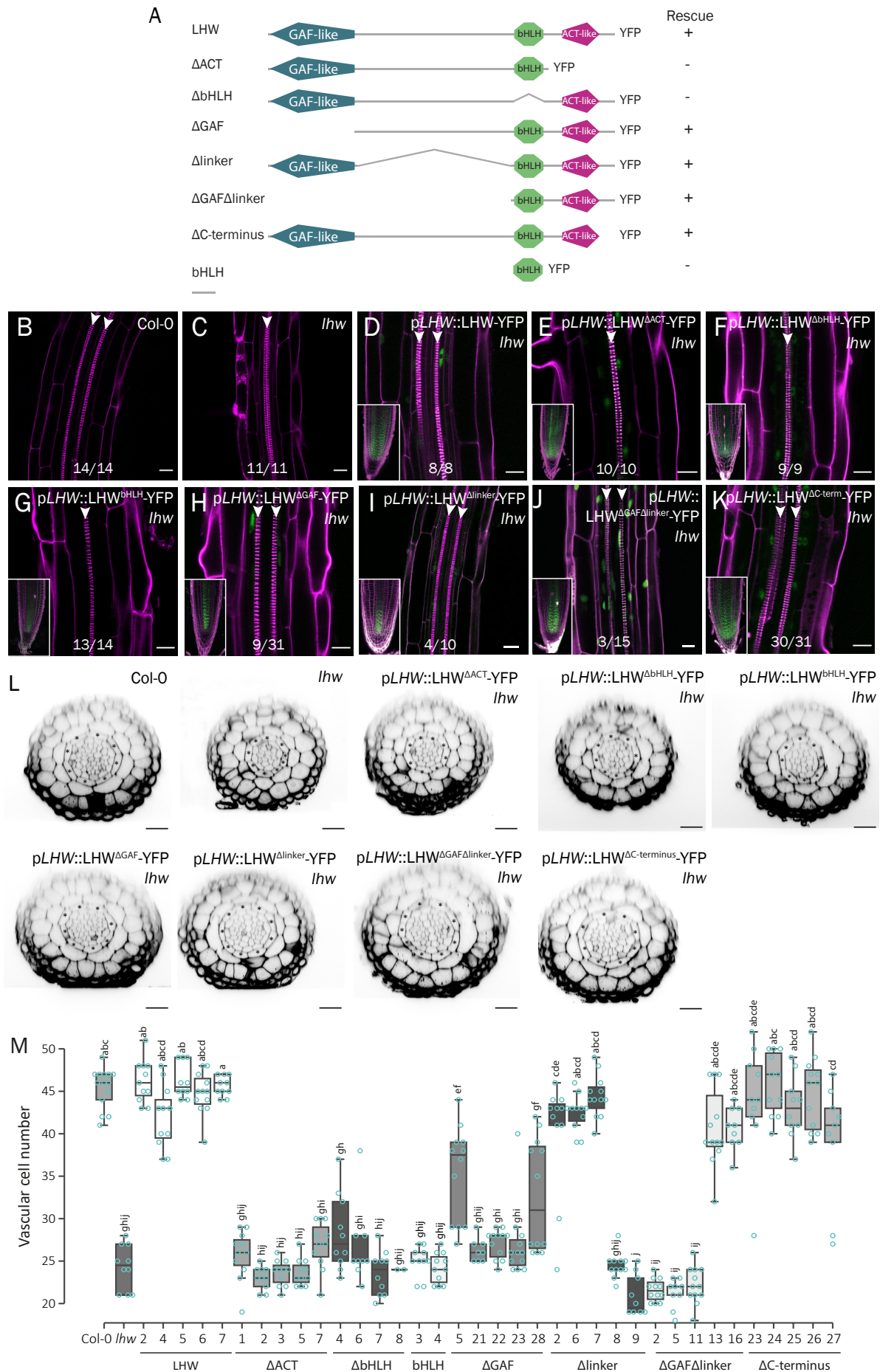
The ACT-like domain emerged through sequence divergence

It has been reported that both plant and animal bHLH protein diversification occurred through extensive sequence divergence and domain shuffling, a process involving domain insertion and rearrangement. We identified that the essential ACT-like domain of LHW is present in all Tracheophyte LHW orthologues. Within early diverging land plants and Charophytes, this ACT-like domain is only found in the LHW orthologues of hornworts. We set out to investigate if the ACT-like domain was introduced by insertion or through mutations of an existing sequence. To this end, we performed a sequence alignment of ACT-like domains of Arabidopsis LHW with all Bryophytic as well as the two Charophytic LHW orthologues, belonging to Klebsormidium and Cylindrocystis (**Figure 3A**). All hornwort species showed conserved amino acid sequences of approximately 50% sequence identity. Although no ACT-like domain was identified in mosses or liverworts, both Takakia (Tl2080512) and Marchantia (Mp2039322) showed high amino acid sequence conservation to the ACT-like domain of Arabidopsis (55% and 67% identity respectively). Discrepancies were distributed equally over the predicted protein structure (**Figure 3B**). Moreover, 30% and 35% of the ACT-like domain amino acids were already present C-terminal of the bHLH in Klebsormidium and Cylindrocystis, respectively (**Figure 3A**). These ACT-like resembling sequences will be referred to as "protoACT domains". From these results we can state that the ACT-like domain of LHW did not emerge by domain shuffling, but rather through a series of point mutations.
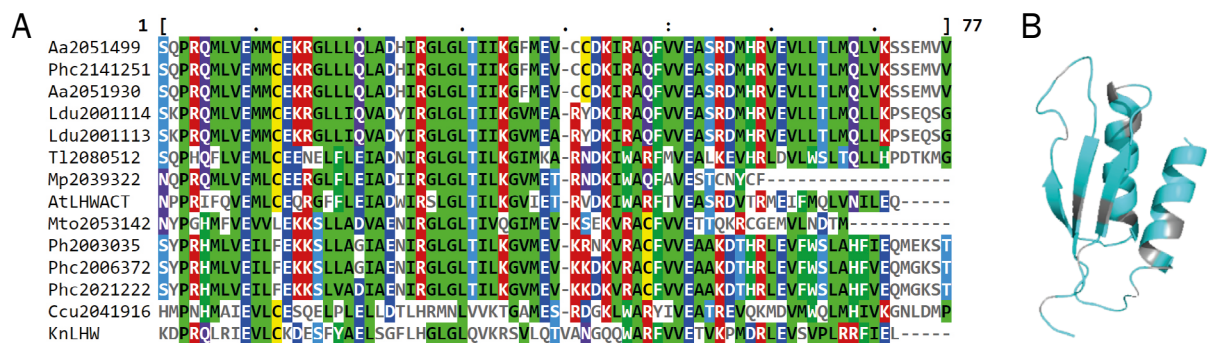
Figure 3. The evolution of the ACT-like domain of LHW proteins. (A) A multiple sequence alignment of ACT-like domains of LHW orthologues of Arabidopsis (At), various Bryophytes (See Chapter 2, Supplemental Table 5 for abbreviations), Klebsormidium (Kn) and Cylindrocystis (Ccu). **(B)** Three-dimensional fold of the Arabidopsis LHW ACT-like domain. Grey areas mark the variation with Bryophytic ACT-like domains.

We next verified the three-dimensional structure using a second prediction algorithm: I-TASSER. These results confirmed the divergent protein folding of the ACT-like domain of Arabidopsis and protoACT domains of Marchantia and Klebsormidium. While the LHW ACT-like domain contains the characteristic α-helices and β-sheets on opposite sides, the protoACT domains were predicted to be made up of α-helices only (**Figure 4B - E**). Next, we investigated *in silico* which amino acid positions in the protoACT domain may have contributed to the gain of the ACT-like

fold. Accordingly, we applied selective point mutations in the amino acid sequence of the Marchantia LHW protoACT, based on the alignment with AtLHW ACT-like (**Figure 4A**). Of the 72 amino acids, 31 are different between the ACT-like domain and the protoACT. We first created 'protoACT swap1' by substituting fourteen of these amino acids to match the ACT-like domain of Arabidopsis (**Figure 4F**). Modeling of protoACT swap1 revealed an ACT-like fold, similar to the LHW ACT-like domain, confirming our expectations (**Figure 4G**). We next restricted the point mutations to the β-sheets, based on the protoACT 3D model, creating protoACT swap2. ProtoACT swap2, in which a minimal seven amino acids were altered, was predicted to fold like an ACT-like domain (**Figure 4H, I**). Together, these results reveal that the ACT-like domain of LHW may have evolved through sequence divergence by point mutations in the protoACT, transforming short α-helical structures into β-sheets.
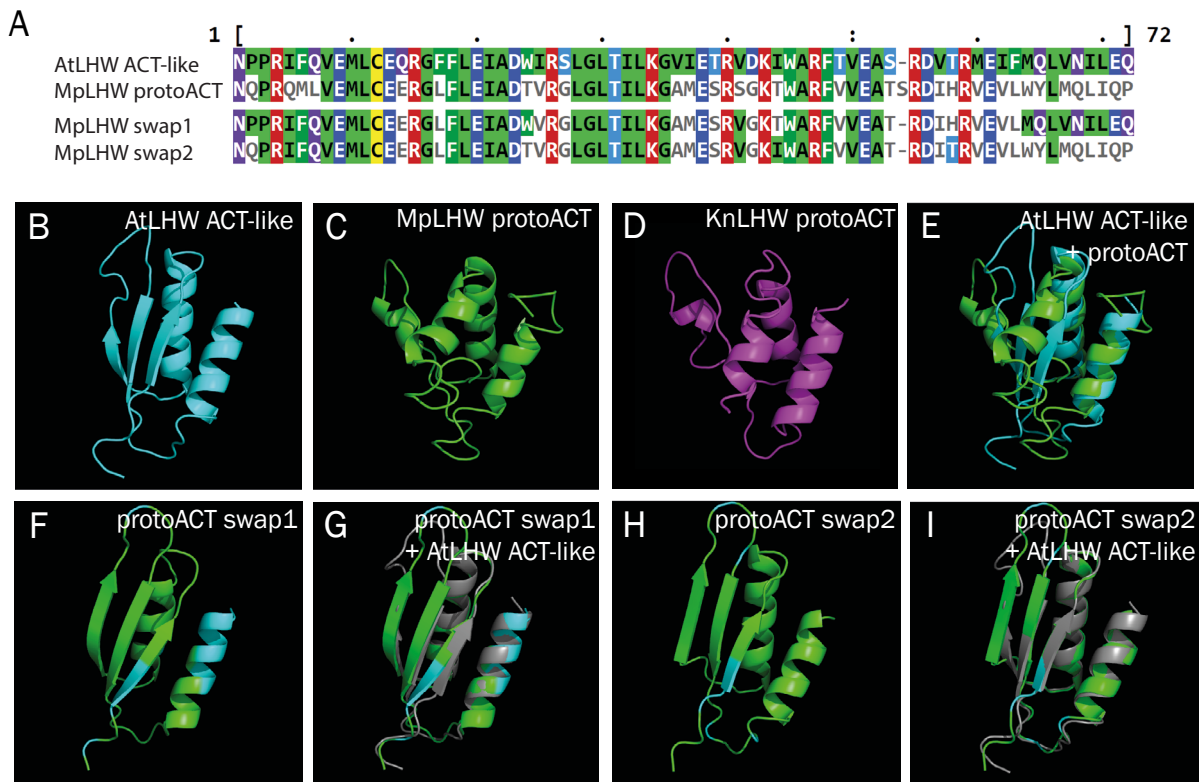
Figure 4. The ACT-like domain evolved through point mutations. (A) A sequence alignment of the ACT-like domain of AtLHW and the protoACT of MpLHW. Amino acid substitutions in MpLHW swap1 and MpLHW swap2 are indicated below. **(B-D)** Predicted domain folds of AtLHW ACT-like (B), MpLHW protoACT and KnLHW protoACT (D). **(E)** The alignment of the LHW ACT-like domain and a protoACT domain. **(F-G)** Structural prediction of protoACT swap1 (F) and the alignment with AtLHW ACT-like (G). The protoACT is depicted in green, the altered amino acids in blue and the AtLHW ACT-like fold in grey. **(H-I)** Structural prediction of protoACT swap2 (H) and the alignment with AtLHW ACT-like (I). The protoACT is depicted in green, the altered amino acids in blue and the AtLHW ACT-like fold in grey.

Gain of the ACT-like domain is not sufficient for development of the vascular function

Our data indicates that the protoACT evolved into a correctly folded ACT-like domain approximately at the time in evolution where the vascular function of LHW was established. However, whether

the emergence of the ACT-like domain was causal to the formation of complex vascular tissue remains unclear. We therefore investigated whether the gain of the ACT-like domain may have contributed to the modern vascular function of LHW. To that end, we constructed the chimeric LHW protein LHW-C1 (**Figure 5A**). In LHW-C1, the ACT-like domain of AtLHW was introduced in MpLHW, replacing the protoACT. When introducing the chimeric *LHW* gene into the *lhw* background of *Arabidopsis*, none of the nineteen T1 roots were restored to a diarch phenotype (**Figure 5 B-D**). This experiment illustrates that, while domain architecture of LHW evolved, the emergence of the LHW vascular function was not solely determined by acquisition of the ACT-like domain.
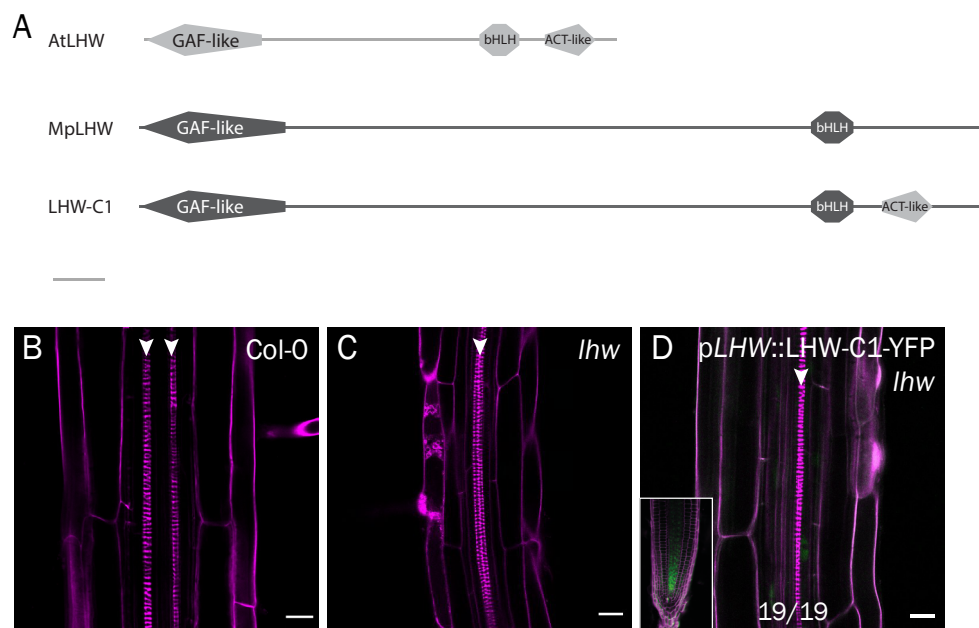
Figure 5. Functional analysis of a Arabidopsis/Marchantia chimeric LHW protein. (A) Schematic representation of AtLHW, MpLHW and the chimeric LHW: LHW-C1. Protein domains of AtLHW are shown in light grey and MpLHW domains in dark grey. The scale bar represents 50 amino acids. **(B)** A wild type Arabidopsis root with two xylem poles. **(C)** A *lhw* mutant with a single xylem pole. **(D)** Complementation analysis of the *lhw* mutant by LHW-C1 fused to a YFP. YFP expression is shown in green. White arrowheads mark protoxylem strands. Scale bars represent 20 μm.

Evolution of the DNA-binding region in TMO5 gave rise to the vascular function

In contrast to LHW, TMO5 domain architecture and protein sequence appears extremely conserved from Charophytes to Angiosperms. Yet, the vascular function is not shared by a Charophycean orthologue. To pinpoint which part of the protein was responsible for the novel biological function of TMO5, we compared the Charophycean KnTMO5 with the land plant MpTMO5 (**Figure 6A**). While domains are conserved, KnTMO5 lacks the N-terminal region upstream of the bHLH domain. Domain swap experiments were designed in which domains or regions of Kn*TMO5* were replaced by those of Mp*TMO5*, creating chimeric proteins. Chimeric TMO5 proteins were constructed in which the KnTMO5 N-terminal tail, the bHLH domain, the ACT-like domain, or

combinations thereof were exchanged with those of MpTMO5 (**Figure 6A**). The chimeras were introduced into the *tmo5 t5l1* double mutant and driven by the Arabidopsis *TMO5* promoter (see Material and Methods for details), to assess their functionality by mutant complementation (**Figure 6B - H, Supplemental Table 2**). Addition of the N-terminal tail (TMO5-C1) was not sufficient to complement the mutant phenotype (**Figure 6D, Supplemental Table 2**). When swapping both the bHLH and ACT-like domain (TMO5-C2), we observed a gain of vascular function, meaning one or both of these domains is responsible for the vascular function of TMO5 (**Figure 6E, Supplemental Table 2**). Single domain swaps of the ACT-like domain (TMO5-C3) or the bHLH domain together with the N-terminus (TMO5-C4) identified the bHLH domain as the causal domain, as it was able to rescue the mutant phenotype, whereas TMO5-C3 was not (**Figure 6F, G, Supplemental Table 2**). We therefore examined the bHLH domain of KnTMO5 more closely. Overall, no substantial differences exist within the TMO5 HLH region, when comparing Klebsormidium, Marchantia
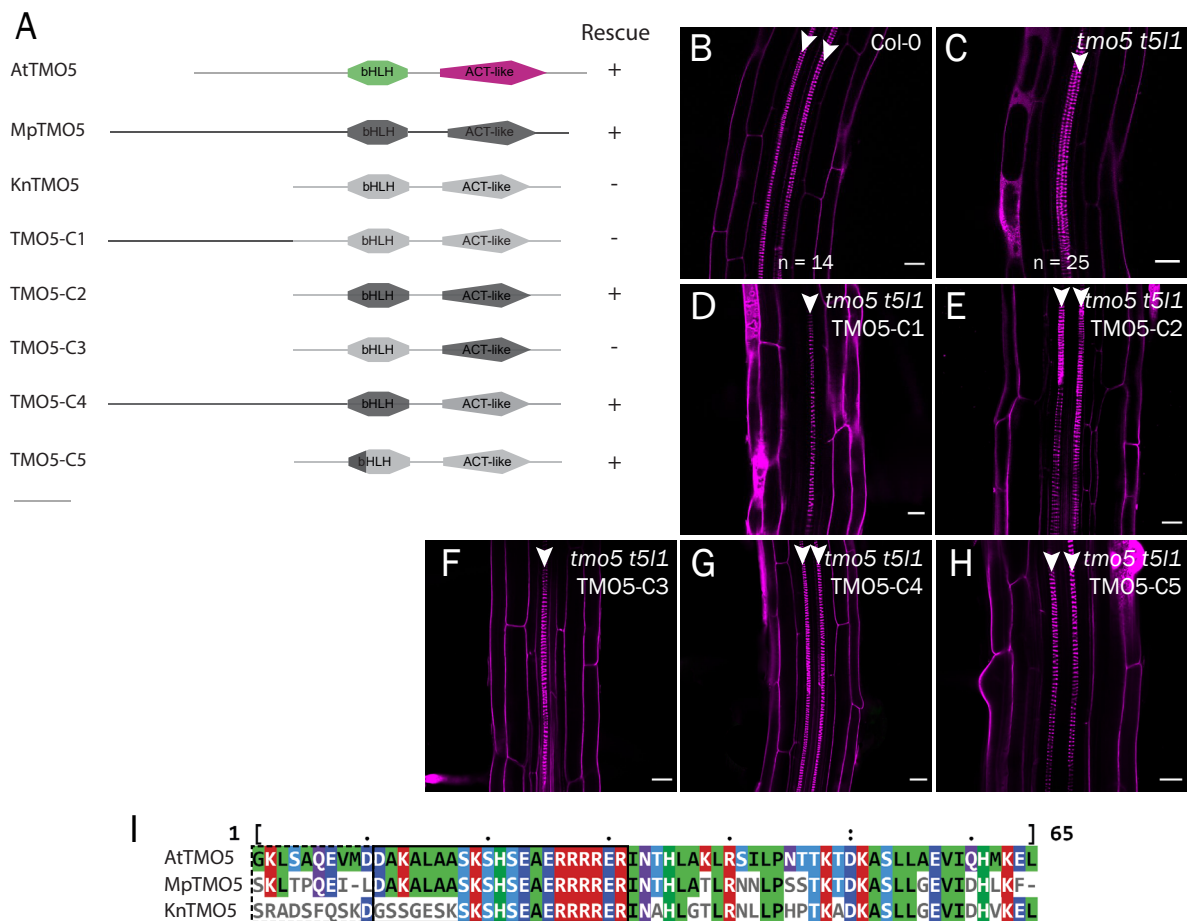
Figure 6. Functional analysis of Klebsormidium/Marchantia TMO5 chimeric proteins. (A) Schematic representation of AtTMO5, MpTMO5, KnTMO5 and the chimeric TMO5 proteins. Protein domains of MpTMO5 are shown in dark grey and KnTMO5 domains in light grey. The scale bar represents 50 amino acids. **(B)** A Wild type Arabidopsis root with two xylem poles. **(C)** A *tmo5 t5l1* double mutant with a single xylem pole. **(D-H)** Complementation analysis of the *tmo5 t5l1* double mutant by TMO5-C1 (D), TMO5-C2 (E), TMO5-C3 (F), TMO5-C4 (G) and TMO5-C5 (H) using the Arabidposis p*TMO5* promoter. White arrowheads mark protoxylem strands. Scale bars represent 20 μm. **(I)** Sequence alignment of Arabidopsis, Marchantia and Klebsormidium TMO5 bHLH domains. The black box indicate the basic region of the bHLH. The dashed box marks the leader sequence.

and Arabidopsis (**Figure 6I**). Furthermore, a large part of the basic region of the Klebsormidium, Marchantia and Arabidopsis TMO5 proteins are identical. However, a seven amino acid stretch at the N-terminus of the basic amino acid region is highly conserved in Marchantia and Arabidopsis, but not in Klebsormidium (**Figure 6I**). The leader sequence, i.e. the ten amino acids upstream of the bHLH also show high sequence variation. We hypothesized that evolution of the basic region, rather than of the HLH, was responsible for the achievement of the vascular function. Another, more specific, chimeric protein (TMO5-C5) was designed in which the seven amino acids of the basic domain as well as nine amino acids of the leader sequence of MpTMO5 were introduced into KnTMO5 (**Figure 6A, I**). Intriguingly, interchanging only these sixteen amino acids of the basic domain and its upstream region was sufficient to create a KnTMO5 protein that could complement the Arabidopsis mutant (**Figure 6H, Supplemental Table 2**). It is fascinating that such small alterations were able to transform a green alga gene to perform a vascular-specific function. Especially, since Klebsormidium diverged from the land plant ancestor 850 million years ago, 400 million years prior to the emergence of vascular tissue. We thus conclude that evolution of the basic region of the TMO5 bHLH in the common ancestor of Bryophytes and Tracheophytes gave rise to the vascular function.

Overall, we show that evolution of the LHW protein domain structure was dynamic and an ACT-like domain was acquired in the ancestor of Tracheophytes or in the common ancestor of Tracheophytes and Bryophytes. While gain of the vascular function coincided with the formation of the ACT-like domain, emergence of this domain was not the absolute determinant for the evolution of a vascular function. In contrast, both domain architecture and the vascular function of the TMO5 protein are extremely conserved over the course of plant evolution. We were able to pinpoint that mutations specifically in the basic region of the bHLH domain gave rise to the vascular function in the common ancestor of Bryophytes and Tracheophytes.

Discussion

Because tertiary structures of proteins tend to be more conserved than the primary sequence, research into protein domains and domain architecture have received increasing attention in recent years. In this chapter, we studied the domain architecture of TMO5 and LHW as well as its conservation throughout evolution. We identified bHLH and ACT-like domains in both TMO5 and LHW proteins and an additional GAF-like domain in LHW. The bHLH and ACT-like domains were found to be conserved in TMO5 orthologues of Embryophytes and Charophytes. In contrast, ACT-like domains of LHW orthologues were subjected to change. We further demonstrated that the ACT-like domain as well as the bHLH domain are necessary and sufficient for vascular function of the LHW protein. While a correlation exists between the timing of ACT-like domain emergence

and the acquisition of vascular function of LHW, this was not a causal relationship, Indeed, introducing an ACT-like domain in MpLHW was not sufficient to convey a vascular function. We further managed to capture the evolutionary trajectory of the TMO5 vascular function by pinpointing a sixteen amino acid stretch in the basic and upstream region of the bHLH domain, responsible for the gain of vascular function.

Given that the basic region of the bHLH is involved in DNA binding and the HLH region in dimerization (Murre et al., 1989), it is not surprising that this domain is crucial for conveying the vascular function. bHLH transcription factors require dimerization and subsequent DNA interaction to initiate transcription of the target gene.

ACT-like domains have been identified in bHLH proteins before and were found to have significant roles in dimerization and regulatory specificity (Feller et al., 2006; Kong et al., 2012). One can thus speculate that such essential functions are conserved in all ACT-like domains, including the ACT-like domain of LHW. While the exact role of the ACT-like domain in LHW is yet to be determined, our work did conclude that gain of the ACT-like domain alone was not sufficient to evolve the vascular function.

One important consideration is that amino acid sequences of the ACT-like domain of Arabidopsis and the protoACT of Marchantia were rather well conserved. While a few point mutations can, in theory, alter domain folding, the possibility of a conserved domain architecture between Bryophytes and Tracheophytes remains possible. Since the domain architecture analyses were predictions performed using Hidden Markov Models (HMMs) based on fold recognition, the conclusions regarding the presence or absence of the ACT-like domain are strongly dependent on thresholds set within these algorithms. Nevertheless, even a MpLHW with a correctly folded, functional ACT-like domain was unable to induce cell divisions in the vascular tissue, suggesting that other or additional regions were involved in vascular function emergence.

An alternative hypothesis that could explain the discrepancy between Tracheophytic and Bryophytic LHW involves the length of the linker situated between the GAF-like domain and the bHLH. This linker is twice as long in MpLHW compared to AtLHW. Even though the linker of AtLHW can be removed without compromising the vascular function, an extended linker may affect protein folding or interaction and thereby its function.

Similar to LHW, the bHLH domain of TMO5 is crucial for its functionality. During TMO5 evolution, mutations in the basic DBD ultimately resulted in the acquisition of the vascular function. While sixteen amino acids were mutated (seven within the basic region and nine in the leader) we suspect that the seven amino acids of the DBD, rather than the leader, generated functional divergence. Indeed, DBDs contain base-contacting residues which determine DNA-binding

preferences and consequentially define transcription factor targets. It has been shown previously that substitutions in specific residues of the DBD can alter the DNA target of transcription family members (Badis et al., 2009; Berger et al., 2008; Noyes et al., 2008). Given the negative surface charge of DNA, basic amino acids are highly important for protein-DNA interactions. However, implemented mutations in the DBD of KnTMO5, did not increase the overall amount of basic amino acids, but rather modified the distribution of basic residues. While highly speculative, this redistribution may have altered the exposure of one or more basic amino acids to the DNA. Furthermore, the AtTMO5 DBD contains many hydrophobic residues. While DNA recognition is often mediated by hydrogen bonds between amino acids and nucleotides, hydrophobic interactions also contribute to sequence recognition (Rohs et al., 2010; Suzuki, 1994). These hydrophobic interactions are mostly formed between hydrophobic amino acids and thymine or cytosine. The six base pair G-box upstream of *LOG4*, to which TMO5 and LHW bind, contains one thymine and two cytosines. As such, it is tempting to speculate that increased amount of hydrophobic amino acids in the DBD of TMO5, may have increased binding preference to the *LOG4* G-box or created a more stable protein-DNA complex. Besides being hydrophobic, the aforementioned amino acids have high helix-forming propensities. Helix-promoting residues are not present in the studied DBD region of KnTMO5 and may therefore have evolved after Charophytes split from their ancestor. While purely speculative, one could contemplate the effects of structural rearrangement that occurred in the DBD of TMO5 during the course of evolution. While β-sheets and disordered sequences can interact with DNA, interaction through α-helices is much more prevailing (Alberts et al., 2008; Garrett and Grisham, 1997; Rohs et al., 2010; Yamasaki et al., 2005). Extension of the α-helical DBD of TMO5 may have contributed to DNA-binding affinity or specificity. In addition, a helical structure may influence the orientation of base-contacting amino acids and expose a different set of residues to the DNA, compared to a disordered or β-sheet organized DBD. One example of structural rearrangement and its influence on transcription factors, has been shown for the *Arabidopsis* transcription factors DEHYDRATION-RESPONSIVE ELEMENT-BINDING (DREB) and ETHYLENE RESPONSE FACTOR (ERF). Residues involved in protein structure, rather than in DNA interaction, are responsible for the sequence-recognition diversification between these two transcription factors (Hao et al., 2002; Liu et al., 2006; Yang et al., 2009).

Overall, substitutions in the DBD of transcription factors can create functional divergence in diverse ways. In 2005, Maizel and colleagues reported that the plant specific transcription factor LEAFY is present in the moss *Physcomitrella patens*. While the specific floral fate induction activity of Angiosperms was not conserved in this basal land plant, this activity was partially gained upon substitutions in the DBD domain of *P. patens* LEAFY. A similar mechanism was discovered by Higo et al. (2018) for the DUO1 protein, involved in sperm differentiation. Point mutations in the DBD of the algal DUO1 orthologue enabled recognition of the target gene in the sperm differentiation

program. It can thus be expected that mutations in the DBD domains of transcription factors are common and have played major roles in transcription factor network evolution.

Nine residues of the mutated sixteen amino acid stretch were positioned upstream of the bHLH domain. While we expect the mutations in the basic DBD to be the main driver for TMO5 evolution, the involvement of the leader sequence cannot be excluded. Many DNA-interacting proteins contain regions outside the DBD to assist DNA docking. For example, N-terminal arms of the bacteriophage λ repressor and homeodomain proteins contact bases in the major and minor groove respectively (Jordan and Pabo, 1988; Joshi et al., 2007). The zinc finger of the Tramtrack protein of *Drosophila melanogaster* also contains an N-terminal region upstream of the zinc finger. It is believed that this region is responsible for protein orientation along the DNA, aiding the search for specific binding sites (Kamashev et al., 2000). Given that in addition to the basic region, the leader sequence of TMO5 is conserved in all land plants, this region might have a similar role in protein positioning or docking.

Taken together, we report that modifications in the DBD, and potentially the upstream region, of TMO5 was the determinant of protein function divergence and led to the vascular function in the ancestor of Bryophytes. We speculate that base-contacting residues, exposure of specific amino acids, secondary protein structure, protein positioning or combinations thereof led to the recognition of novel target genes. This ultimately resulted in recognition and regulation of the *LOG3* and *LOG4* genes and by extent, complex vascular tissues in Tracheophytes.

Material and methods

Plant material and growth conditions

Seeds were sterilized in 25% bleach and 75% ethanol solutions for 8 minutes, rinsed twice with 70% ethanol and once with 96% ethanol. Died seeds were grown on half strength Murashige and Skoog (MS) medium containing 1% sucrose and 0.8% Daishin agar (Duchefa) after a two day stratification at 4°C. 15 mg/l Phosphinothricin (PPT) was added in order to select transformed seeds. Seedlings were transferred to soil and grown under long day conditions (16 hours light, 8 hours dark) at 22°C. *Arabidopsis* ecotype Columbia-0 (Col-0) was used as wild-type. *lhw, tmo5 and tmo5 55l1* mutants, used for complementation studies, were generated by and obtained from De Rybel et al. (2013, 2014)

Cloning

All cloning procedures were performed using Seamless Ligation Cloning Extract (SLiCE), with 15 homologous bases (Zhang et al., 2015). Primers used for cloning are listed in Supplemental Table

2. Truncated and chimeric versions of LHW proteins were constructed using existing plasmids as template to perform (overlap extension) PCR with Phusion Flash PCR Master Mix (Thermo Scientific). Coding regions, without a stop codon, were cloned into pGIIB-p*LHW*::YFP-NOSt (constructed as described in Chapter 3). TMO5 chimeric proteins were constructed in a similar manner, with a stop codon, and cloned into pGIIB-p*TMO5*::LIC-NOSt (constructed as described in Chapter 3).

Complementation assays

For complementation assays, plasmids carrying the *LHW* truncations or *LHW* chimeras were transformed into *lhw* plants by simplified floral dipping (De Rybel et al., 2011). Rescue was analyzed in T1 as well as T2 plants. Approximately 10 individual seedlings were screened per T2 line.

Plasmids carrying the TMO5 orthologues were transformed to *tmo5 t5l1* plants, homozygous for *tmo5*, heterozygous for *t5l1* (obtained by crossing *tmo5* mutant plants with *tmo5 t5l1* double mutant plants). In T1 lines, rescue was considered when 80% or more of the independent lines showed complementation. Homozygous *tmo5 t5l1* double mutant backgrounds were selected in p*TMO5*::*TMO5*-C5 T2 lines by screening vascular phenotypes in seedlings that lost the transgene by segregation.

Quantification of vascular cell numbers

5 Day-old-roots were fixed and stained using a modified Pseudo Schiff – Propidium Iodine (mPS-PI) staining as explained in Chapter 3 (Truernit et al., 2008). Radial cross section were obtained by a SP5 confocal microscope in the middle of the root meristem: half way between the QC and the first elongating cortex cell. The data was visualized using BoxPlotR.

Plant imaging

To assess monarch and diarch phenotypes, 5 – 7 day-old-seedlings were stained with 10 µg/ml Propidium Iodide. A Leica SP5 confocal microscope, containing an Argon and a DSS561 diode laser, was used for imaging. YFP and Propidium Iodide were excited at 514 nm and 561 nm respectively. Emission was visualized with Hybrid detectors at 520-550 for YFP and 600-680 for Propidium Iodide. When multiple fluorescent markers were imaged, sequential scanning was used.

Domain identification and protein modeling

Domain identification by structural homology predictions was done using Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2). Protein modeling was performed using I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). Protein models were studied and imaged using Pymol.

Sequence alignments

Sequence alignments were constructed using ClustalO or MAFFT version 7 (Katoh and Standley, 2013; http://mafft.cbrc.jp/alignment/server/) using default parameters. Alignments were visualized using MView.

Acknowledgments

4

Supplemental information
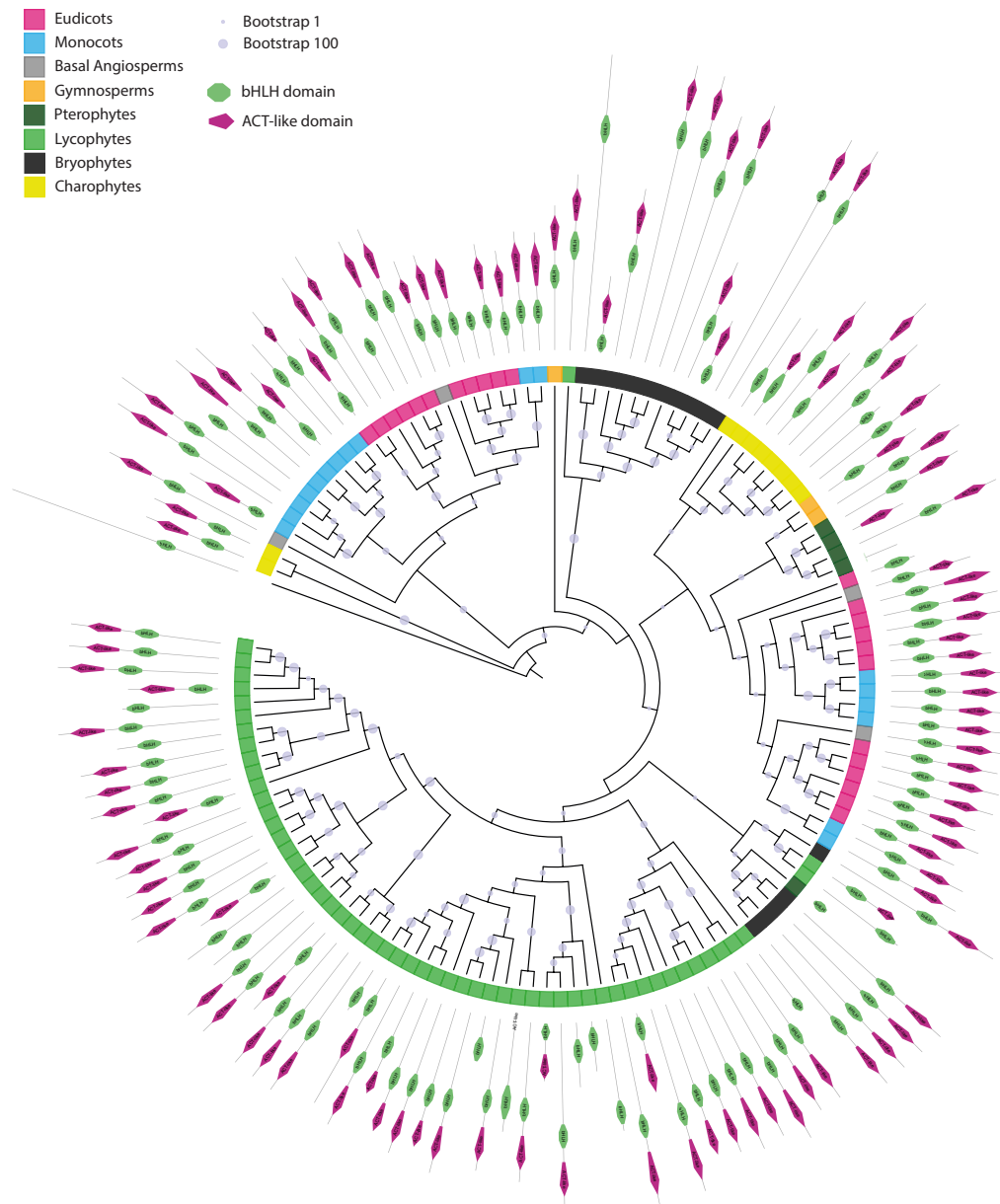
Supplemental Figure 1. Domain architecture of all TMO5 orthologues, as identified in Chapter 2. The bHLH domain is indicated in green and the ACT-like domain in magenta. Colours in the ring represent different taxonomic groups. Bootstap values are represented by blue dots.
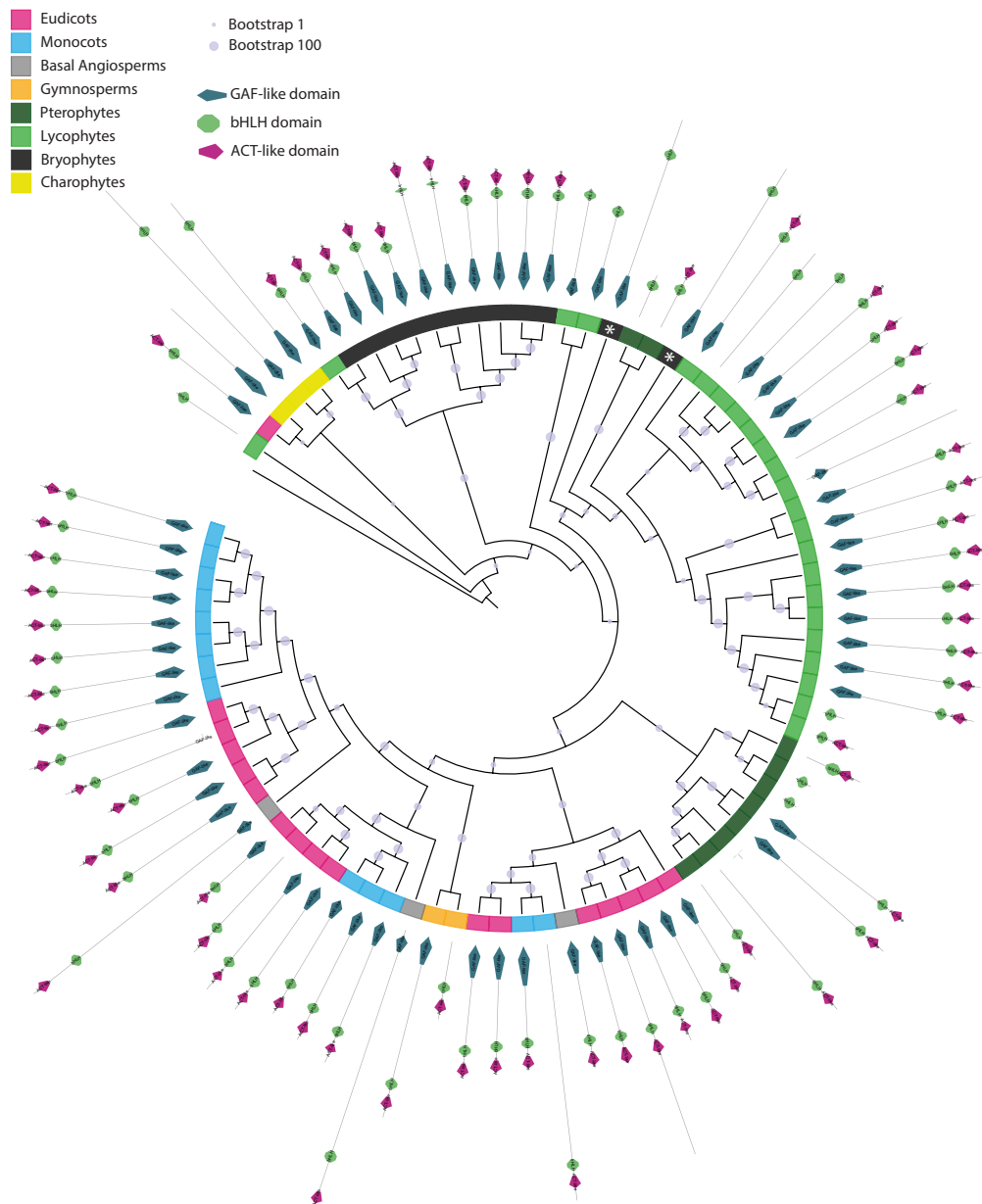
Supplemental Figure 2. Domain architecture of all LHW orthologues, as identified in Chapter 2. The GAF-like domain is indicated in blue, the bHLH domain in green and the ACT-like domain in magenta. Colours in the ring represent different taxonomic groups. White asterisks indicate the moss and liverwort lacking the ACT-like domain. Bootstap values are represented by blue dots.
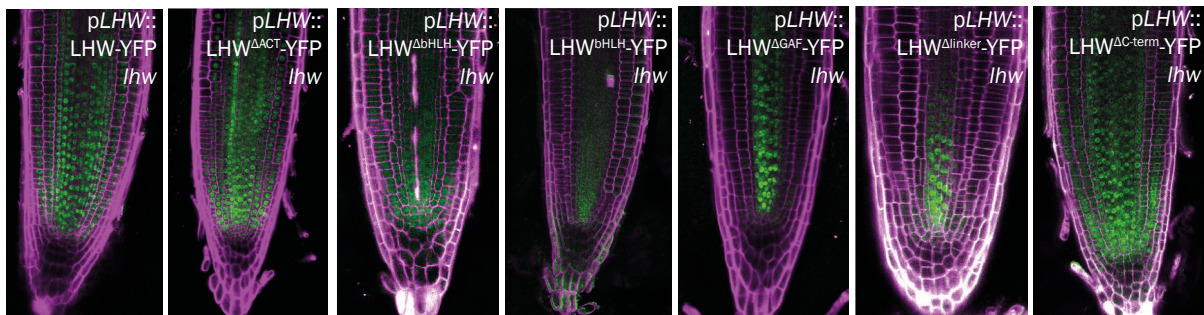
Supplemental Figure 3. Expression of truncated LHW variants driven from the native p*LHW* promoter in *lhw* mutants.

Supplemental tabel 1. Percentage rescue in T2 lines of Arabidopsis *lhw* mutants by LHW domain deletions and truncations.

	Line	% complementation	n
LHW	2	100%	10
	4	100%	10
	5	100%	10
	6	100%	10
LHW^ΔACT	2	0%	10
	5	0%	10
	6	0%	10
LHW^ΔbHLH	4	0%	8
	6	0%	10
	7	0%	10
	8	0%	5
LHW^ΔGAF	5	60%	10
	10	10%	10
	27	0%	10
	28	60%	10
	30	0%	10
	34	30%	10
	36	40%	10
	39	20%	10
LHW^Δlinker	2	80%	10
	4	78%	0
	6	100%	10
	8	0%	10
	9	0%	10

	Line	% complementation	n
LHW^ΔGAFΔlinker	1	0%	10
	2	0%	10
	4	0%	10
	8	0%	10
	11	0%	10
	13	90%	10
	14	0%	10
	15	0%	10
	16	100%	10
	17	0%	10
LHW^ΔC-terminus	23	100%	5
	24	100%	9
	25	88%	8
	26	100%	6
LHW^bHLH	3	0%	10
	4	0%	10
	5	0%	10
	6	0%	10
	7	0%	10

Supplemental tabel 2. Percentage rescue in T1 and T2 lines of Arabidopsis *tmo5 t5l1* double mutants by chimeric *TMO5* genes.

	Line	% complementation	n	Background (%diarch)	Genotype	n
TMO5-C1	1-22 (T1)	82%	22	75%	segregating	
TMO5-C2	16	100%	10			10
	17	100%	10			10
	18	100%	10			10
	19	80%	10			10
	20	100%	10			10
	21	100%	10			10
TMO5-C3	1-9 (T1)	78%	9	75%	segregating	
TMO5-C4	2	100%	10	88%	tmo5-/- t5l1+/-	25
	4	70%	20	79%	tmo5-/- t5l1+/-	19
	5	100%	10	87%	tmo5-/- t5l1+/-	15
	8	96%	10	65%	tmo5-/- t5l1+/-	25
TMO5-C5	1	100%	10			
	3	100%	20			
	4	100%	10			
	5	100%	10	100%	tmo5-/- t5l1+/+	13
	6	33%	15	0%	tmo5-/- t5l1-/-	5
	7	100%	10			
	8	100%	7	100%	tmo5-/- t5l1+/+	5
	9	100%	8			
	10	100%	10	90%	tmo5-/- t5l1+/-	10
	11	75%	20	0%	tmo5-/- t5l1-/-	5
	12	100%	17	75%	tmo5-/- t5l1+/-	8

Supplemental tabel 3. Primers used in this study.

Primer pairs used for LHW truncations

Gene number	Gene name	Organism	Truncation	Primer orientation	Sequence
AT2G27230	AtLHW	*A. thaliana*	LHWΔACT	Sense	5'- TAG TTG GAA TAG GTT ATG GGA GTT TTA CTA AGA GAA GCT TTA AGG
				Antisene	5'-CTT GCT CAC CAT GTT GGT TTG CTT CAG CTT ATC AGA ATG C
AT2G27230	AtLHW	*A. thaliana*	LHWΔbHLH	Sense	5'- TAG TTG GAA TAG GTT ATG GGA GTT TTA CTA AGA GAA GCT TTA AGG
				Antisene	5'-ATT CCC CGG TCT CTC CTG GTT TAA GCC TTT TCC GG
				Sense	5'-ACC AGG AGA GAC CGG GGA ATC CAA GAT TAT GAA AG
				Antisene	5'-CTT GCT CAC CAT GTT CAT TGA ACA GCC ACC AGT AAC CGG T
AT2G27230	AtLHW	*A. thaliana*	LHWΔGAF	Sense	5'-TAG TTG GAA TAG GTT ATG CTC GGA TGT GTT CCG GGG GCT
				Antisene	5'-CTT GCT CAC CAT GTT CAT TGA ACA GCC ACC AGT AAC CGG T
AT2G27230	AtLHW	*A. thaliana*	LHWΔspacer	Sense	5'-TAG TTG GAA TAG GTT ATG GGA GTT TTA CTA AGA GAA GCT TTA AGG
				Antisene	5'-TTG GGT TCT CGA GCT GCA GGA TAA GAC CCT TCA C
				Sense	5'-CCT GCA GCT CGA GAA CCC AAG ACC AAG GCC TAA A
				Antisene	5'-CTT GCT CAC CAT GTT CAT TGA ACA GCC ACC AGT AAC CGG T
AT2G27230	AtLHW	*A. thaliana*	LHWΔGAFΔspacer	Sense	5'-TAG TTG GAA TAG GTT ATG GAG AAC CCA AGA CCA AGG CCT A
				Antisene	5'-CTT GCT CAC CAT GTT CAT TGA ACA GCC ACC AGT AAC CGG T
AT2G27230	AtLHW	*A. thaliana*	LHWΔC-terminus	Sense	5'-TAG TTG GAA TAG GTT ATG GGA GTT TTA CTA AGA GAA GCT TTA AGG
				Antisene	5'-CTT GCT CAC CAT GTT TGT CTG CTC CAA AAT ATT CAC TAG
AT2G27230	AtLHW	*A. thaliana*	LHW bHLH	Sense	5'-TAG TTG GAA TAG GTT ATG GAG AAC CCA AGA CCA AGG CC
				Antisene	5'-CTT GCT CAC CAT GTT GGT TTG CTT CAG CTT ATC AGA ATG C

Primer pairs for chimeras

Gene number	Chimera	Organism	Primer orientation	Sequence
Mp2039322/ Mapoly0088s0049	LHW-C1	*M.polymorpha*	Sense	5'-TAG TTG GAA TAG GTT ATG GCA ATG GTG CTG CAG CAG GCG T
			Antisene	5'-CTT CTA CGA CAA TTG GAC AAC CAT TAT TTG GTT TGT CA
AT2G27230		*A. thaliana*	Sense	5'-TGA CAA ACC AAA TAA TGG TTG TCC AAT TGT CGT AGA AG
			Antisene	5'-CTT GCT CAC CAT GTT CAT TGA ACA GCC ACC AGT AAC CGG T
Mp2020606/ Mapoly0039s0068	TMO5-C1	*M.polymorpha*	Sense	5'-TAG TTG GAA TAG GTT ATG GGT GGA GAA GCA ACG GGG CGC
			Antisene	5'-CCA GAT TTT GAG CTC TCC ATA ACG AGC CCG TCA TCG TAA C
Kn000610230		*K. nitens*	Sense	5'-ATG GAG AGC TCA AAA TCT GGA G
			Antisene	5'-AGT ATG GAG TTG GGT TTC ACG AGG CGG TCA CTG TT
Kn000610230	TMO5-C2	*K. nitens*	Sense	5'-TAG TTG GAA TAG GTT ATG GAG AGC TCA AAA TCT GG
			Antisene	5'-AGC CAG TGC CTT GGC TCC ATC TTT GCT CTG GAA GGA
Mp2020606/ Mapoly0039s0068		*M.polymorpha*	Sense	5'-GCC AAG GCA CTG GCT GCC TC
			Antisene	5'-GTC CGC CTG CCG CTG GAG AAA CTT AAA TG
Kn000610230		*K. nitens*	Sense	5'-CAT TTA AAG TTT CTC CAG CGG CAG GCG GAC
			Antisene	5'-GCT CGA TAC ACC TGC GCC CCT TCT TGC
Mp2020606/ Mapoly0039s0068		*M.polymorpha*	Sense	5'-GCA GGT GTA TCG AGC GTC TCT TTG CTG
			Antisene	5'-GCT TCT GGT CGC TAC CCT CAT CGC TT
Kn000610230		*K. nitens*	Sense	5'-GTA GCG ACC AGA AGC TGC CGA AGG ACG
			Antisene	5'-AGT ATG GAG TTG GGT TTC ACG AGG CGG TCA CTG TT
Kn000610230	TMO5-C3	*K. nitens*	Sense	5'-TAG TTG GAA TAG GTT ATG GAG AGC TCA AAA TCT GG
			Antisene	5'-GCT CGA TAC ACC TGC GCC CCT TCT TGC
Mp2020606/ Mapoly0039s0068		*M.polymorpha*	Sense	5'-GCA GGT GTA TCG AGC GTC TCT TTG CTG
			Antisene	5'-GCT TCT GGT CGC TAC CCT CAT CGC TT
Kn000610230		*K. nitens*	Sense	5'-GTA GCG ACC AGA AGC TGC CGA AGG ACG
			Antisene	5'-AGT ATG GAG TTG GGT TTC ACG AGG CGG TCA CTG TT
Mp2020606/ Mapoly0039s0068	TMO5-C4	*M.polymorpha*	Sense	5'-TAG TTG GAA TAG GTT ATG GGT GGA GAA GCA ACG GGG CGC
			Antisene	5'-GTC CGC CTG CCG CTG GAG AAA CTT AAA TG
Kn000610230		*K. nitens*	Sense	5'-CAT TTA AAG TTT CTC CAG CGG CAG GCG GAC
			Antisene	5'-AGT ATG GAG TTG GGT TTC ACG AGG CGG TCA CTG TT
Kn000610230	TMO5-C5	*K. nitens*	Sense	5'-TAG TTG GAA TAG GTT ATG GAG AGC TCA AAA TCT GG
			Antisene	5'-GGC ATC GAG GAT CTC CTG AGG AGT CAA CTT CGA AGC CCG GCT GAG GCT GTT ATC C
			Sense	5'-GGA GAT CCT CGA TGC CAA GGC ACT GGC TGC CTC GAA AAG CCA CAG CGA AGC AG
			Antisene	5'-AGT ATG GAG TTG GGT TTC ACG AGG CGG TCA CTG TT

Chapter 5

Functional characterization of LHW and TMO5 protein domains

Nicole van 't Wout Hofland[1], Robin Vanden Bossche[2,3], Nathalie Clark[4], Rosangela Sozzani[4], Alain Goossens[2,3], Dolf Weijers[1] and Bert De Rybel[1,2,3]

1. Wageningen University, Laboratory of Biochemistry, Stippeneng 4, 6708 WE Wageningen, the Netherlands

2. Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, 9052 Ghent, Belgium

3. VIB Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium

4. Department of Plant and Microbial Biology, North Carolina State University, 2731 Pillsbury Cir, Raleigh, NC 27607, United States

Abstract

Protein domains are the main units of a protein and form the primary level at which protein function and interactions can be understood. Studying individual protein domains thus can give a great deal of insight into the protein function. The vascular tissue regulators TARGET OF MONOPTEROS 5 (TMO5) and LONESOME HIGHWAY (LHW) belong to one of the largest plant transcription factor families: the basic Helix-Loop-Helix (bHLH) family. Besides the defining bHLH motif, additional domains were identified in TMO5 (an ACT-like domain) and LHW (a GAF-like domain and an ACT-like domain) (Chapter 4). Here, we dig deeper into the domains present within the TMO5 and LHW proteins in order to increase the understanding of these proteins. We report that the LHW GAF-like domain structurally resembles a JAZ-interaction domain (JID). These domains can mediate interactions with JAZ proteins, which are known transcriptional inhibitors that are degraded upon Jasmonic acid (JA) perception. Although we show that certain JAZ proteins may affect TMO5/LHW transcriptional activities and that JAZ1 and LHW may be present in a complex, no conclusive evidence for JA and cytokinin (CK) crosstalk in the TMO5/LHW pathway is found. We further report that the bHLH and ACT-like domains of TMO5 and LHW are redundant, homotypic dimerization domains. In addition, we provide evidence that the LHW ACT-like domain may be a potential interaction specificity determinant, restricting the broad interaction capacity of LHW and directing it towards heterodimerization with TMO5.

Introduction

TARGET OF MONOPTEROS 5 (TMO5) and LONESOME HIGHWAY (LHW) belong to one of the largest plant transcription factor families: the basic Helix-Loop-Helix (bHLH) family (Toledo-Ortiz et al., 2003). These proteins often contain other conserved domains in addition to the bHLH motif. In Chapter 4, domains within TMO5 and LHW were identified and used as a tool to explain functional innovations. We found that both TMO5 and LHW harbour an ACT-like domain downstream of the bHLH motif. In addition, a GAF-like domain is present within the N-terminal region of the LHW protein. However, specific functions of these domains remain unclear.

The ACT domain motif was first visualized in 1995, with the crystal structure of D-3-phosphoglycerate dehydrogenase from *Escherichia coli*. However, it was not recognized as a conserved domain until it was detected in diverse, mostly metabolic, proteins of archaea, bacteria and plants in 1999 (Aravind and Koonin, 1999; Chipman and Shaanan, 2001). It was named after three proteins containing this domain: aspartokinase (A), chorismate mutase (C) and TyrA (T) (Aravind and Koonin, 1999). The signature fold of the 60-80 amino acid ACT domains is a αβ sandwich (βαββαβ) which folds to form an α-helix side and an antiparallel β-sheet side (Al-Rabiee et al., 1996; Chipman and Shaanan, 2001). However, variations in this arrangement have been reported (Curien et al., 2008; Gai et al., 2016; Saxton et al., 2016). As such, structurally related, but not identical, folds are termed ACT-like domains (Chipman and Shaanan, 2001). ACT-like domains are found in plant bHLH proteins, often localized downstream of the bHLH domain (Anantharaman et al., 2001; Curien et al., 2008; Feller et al., 2006). While ACT domains were characterized in proteins that function in the control of metabolism, solute transport, and signal transduction (Anantharaman et al., 2001; Aravind and Koonin, 1999; Ettema et al., 2002), ACT-like domains may have different or additional functionalities. For example, the ACT-like domains of the Maize (*Zea mays*) bHLH protein RED1 (R) and the Arabidopsis Aspartate Kinase 1 (AK1), act as homodimerization domains (Feller et al., 2006; Mas-Droux et al., 2006). In addition, the ACT-like domain of R can suppress DNA-binding activity (Feller et al., 2006; Kong et al., 2012).

GAF-like motifs are complex folds that resemble a PAS domain, in which centrally localized β-sheets are surrounded by α-helices and loops (Ho et al., 2002). Similar to the ACT domain, the GAF domain obtained its name from three proteins in which this motif was identified: cGMP-specific cyclic nucleotide phosphodiesterase (G), Adenylyl cyclase (A), and FhlA (F). These GAF-like domains are not restricted to plants, but are widespread and can be found in many bacteria and eukaryotes (Anantharaman et al., 2001). In some of these species, GAF-like domains are known for their small-molecule binding capacities and were found to be involved in sensory systems, protein regulation and signal transduction (Anantharaman et al., 2001; Aravind and Ponting, 1997).

In mouse phosphodiesterase PDE2A, one of the two GAF domains is involved in dimerization (Martinez et al., 2002). In plants, GAF-like domains are found in ethylene receptors, phytochromes, redox-responsive transcription factors and in NIN-like proteins (NLPs) (Cao et al., 2017; Kim et al., 2011). However, the role of this domain within these proteins has not been well elucidated.

In this chapter, we explore the role of the GAF-like domain of LHW and the bHLH and ACT-like domains of TMO5 and LHW. We identify the bHLH and ACT-like domains as redundant homotypic heterodimerization domains. Furthermore, we provide data suggesting that the ACT-like domain may restrict the dimerization specificity of LHW.

Results

The LHW GAF-like domain is not involved in heterodimerization with TMO5

Protein domains are modular units of a protein, responsible for a particular function or interaction and thereby contributing to the overall protein function. Besides the family-defining bHLH domain, LHW contains a GAF-like and an ACT-like domain (**Figure 1A**). Since dimerization is an essential event in bHLH transcription factor function (Murre et al., 1989) and some GAF-like domains play a role in dimerization (Martinez et al., 2002), we first investigated the dimerization potential of the LHW GAF-like domain. However, given that TMO5 does not contain a GAF-like domain to engage in homotypic interactions (**Figure 1B**), heterodimerization was not expected to be mediated by the GAF-like domain. We constructed a version of LHW without a GAF-like domain (LHW$^{\Delta GAF}$) and investigated its ability to interact with full length TMO5 using Bimolecular Fluorescence Complementation (BiFC). In accordance with our expectations, LHW was able to interact with TMO5 even without its GAF-like domain, while negative controls did not indicate any interaction (**Figure 1C-E**). We therefore conclude that the GAF-like domain is not involved in heterodimerization of LHW and TMO5.

The LHW GAF-like domain resembles a JID domain

To gain insight in the GAF-like domain of LHW, we compared the predicted domain structure to known folds. Interestingly, the GAF-like domain highly resembles a JAZ-interacting domain (JID) (**Figure 2A**). JID domains are present in a range of bHLH transcription factors such as MYC2, MYC3, MYB21, GLABRA 3 (GL3), ENHANCER OF GL3 (EGL3) and TRANSPARENT TESTA8 (TT8). These proteins can interact with a range of JASMONATE ZIM DOMAIN (JAZ) proteins through their JID domain (Fernández-Calvo et al., 2011; Qi et al., 2011; Song et al., 2011). Through this physical interaction, and recruitment of co-repressors, JAZ proteins inhibit
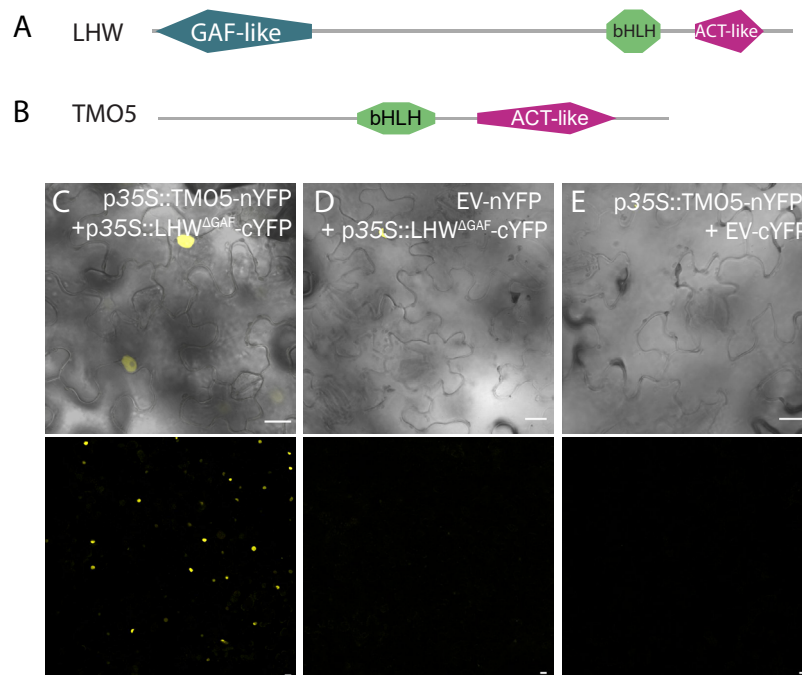
Figure 1. BiFC studies indicate that the GAF-like domain is not involved in TMO5/LHW dimerization. (A) A schematic representation of the LHW domain architecture with the GAF-like (blue), bHLH (green) and ACT-like (magenta) domain. **(B)** A schematic representation of the TMO5 domain architecture containing the bHLH (green) and ACT-like (pink) domain. **(C)** BiFC of TMO5 and LHW without the GAF-like domain of LHW (LHW^ΔGAF). Close-up merge of YFP and brightfield is shown in the top panel, low magnification YFP signal is shown in the bottom panel. **(D-E)** The negative controls of LHW^ΔGAF-cYFP with an empty vector (EV) (D) and TMO5-nYFP with an EV (E). In the former, sporadic fluorescence complementation was obserbed. Close-up merge of YFP and brightfield is shown in the top panel, low magnification YFP signal is shown in the bottom panel. Scale bars represent 20 µm.

transcription of the target genes (Chini et al., 2007; Pauwels et al., 2010). However, upon Jasmonic Acid (JA) perception, JAZ proteins are degraded via the ubiquitin/26S proteasome-dependent proteolytic pathway, resulting in transcription of downstream targets (Chini et al., 2007; Thines et al., 2007). Given the high resemblance to the JID domain, we asked whether the GAF-like domain of LHW is a functional JID domain, able to interact with JAZ proteins. This would imply a cross-talk between the auxin, cytokinin (CK) and JA pathways in which LHW is regulated through JA and JAZ, and TMO5 through auxin and MONOPTEROS (MP), ultimately leading to CK production. Interestingly, JA was found to play a role in secondary vascular regulation in Arabidopsis stems (Sehr et al., 2010). JAZ7 and JAZ10 act as repressors of secondary growth, and addition of JA increases cambium activity (Sehr et al., 2010). Given that JA signaling triggers periclinal cell divisions in the cambium of the stem, it is possible that this mechanism is also employed in the procambium. To investigate if the LHW GAF-like domain can function as a JID domain, we performed interaction studies between LHW and JAZ proteins using the cross Number and Brightness (N&B) method (**Figure 2B, C, Supplemental Figure 1B**). Cross N&B is a method used to determine protein-protein interaction and binding ratios by cross-correlating pixels over space and time (see Material and Methods for more details) (Clark et al., 2016; Clark

and Sozzani, 2017). JAZ1, fused to mScarlet, and LHW, fused to mTurquoise, were overexpressed in *Nicotiana benthamiana* leaves from the *35S* promoter (**Figure 2B**). In the negative control using LHW, fused to mTurquoise, and an empty vector containing mScarlet, no cross-correlation was measured (**Figure 2C**, left panel, **Supplemental Figure 1A**). In contrast, the positive control,
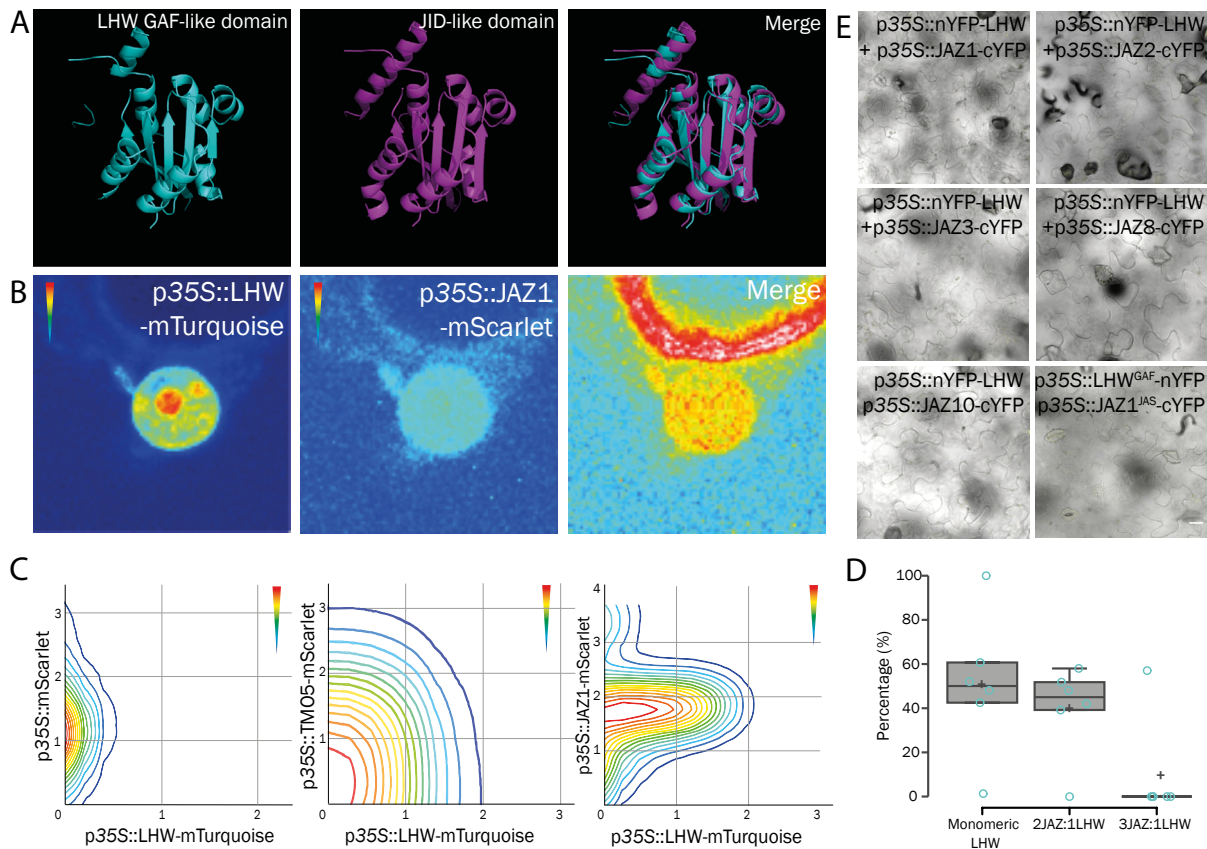
Figure 2. The LHW GAF-like domain is similar to a JID domain and might interact with JAZ proteins. (A) Predicted domain fold of the LHW N-terminal GAF-like domain (cyan), the portein fold of the MYC3 JID domain (magenta) and a merge of both domains. **(B-D)** Cross N&B analysis of LHW and JAZ1 and controls. **(B)** Representative expression of p*35S*::LHW-mTurquoise and p*35S*::JAZ1-mScrlet in a *N. benthamiana* nucleus. **(C)** Stoichiometry histogram of the cross N&B analyses. The colors of the curves represent the proportions of each complex. For LHW/JAZ1, the orange line at (1,2) represents a high proportion of monomeric LHW bound to dimeric JAZ1 (58%), while blue lines at intersections represent extremely low to no interactions. **(D)** Boxplot displaying detected protein complexes of LHW with and without JAZ1 (n=6). Turquoise circles indicate individual samples. Crosses (+) indicate average values **(E)** BiFC analysis of LHW or the GAF-like domain of LHW (LHW^GAF) with different JAZ proteins or the JAS domain of JAZ1 (JAZ1^JAS).

TMO5-mScarlet and LHW-mTurquoise indicated heterodimerization in a 1:1 ratio and minimal dimerization in a 2:1 ratio (**Figure 2C**, middle panel). When p*35S*::JAZ1-mScarlet and p*35S*::LHW-mTurquoise were co-expressed, five out of six samples indicated a JAZ1:LHW interaction ratio of 2:1 (**Figure 2C**, right panel, **Supplemental Figure 1B**). One sample did not indicate any cross correlation and in one sample a 3:1 as well as a 2:1 ratio was identified. On average, 39.8% of the LHW proteins was found in a JAZ:LHW 2:1 complex, 9.5% in a 3:1 complex and 50.8% was present as monomeric LHW (**Figure 2D**). This data hints that LHW and JAZ1 might be

present within the same complex. One drawback of the cross N&B method is that correlation of pixel movement of two fluorescent proteins is determined, indicating protein clustering within a complex, but not physical interaction between two proteins. Furthermore, the assay relies, for a large part, on computational analysis and does not have a direct read-out. We therefore also performed the more direct Bimolecular fluorescence complementation (BiFC) assay to probe interaction between LHW and JAZ1, as well as JAZ2, JAZ3, JAZ8, JAZ9 and JAZ10 (**Figure 2E, Supplemental Figure 1C**). In none of the LHW-JAZ samples, fluorescence complementation of YFP was detected. Moreover, no interaction was observed between the LHW GAF-like domain and the JAZ1 JAS domain (**Figure 2E**). These results suggest that a direct interaction between LHW and JAZ proteins may not exist.

TMO5/LHW target genes are upregulated upon JA treatment

Since direct interaction assays between LHW and JAZ proteins were not conclusive, we investigated the effect of exogenous JA on genes regulated by TMO5 and LHW. If JAZ proteins regulate LHW function, application of JA, and subsequent degradation of JAZ proteins, should result in altered transcriptional activity of the TMO5/LHW dimer. Hickman and colleagues (2017) performed a high-resolution RNAseq analysis on Arabidopsis leaves treated with the volatile derivative of JA, Methyl jasmonate (MeJA) (Hickman et al., 2017). We used this whole-genome expression data to investigate the effect of MeJA on transcript levels of the TMO5/LHW targets. We studied the expression levels of the few downstream targets of the dimer that have been identified to date: *LONELY GUY 3* (*LOG3)* (Ohashi-Ito et al., 2014), *LOG4* (De Rybel et al., 2014; Vera-Sirera et al., 2015), *GLYCOSYL HYDROLASE 10* (*GH10)* (AT4G38650) (De Rybel et al., 2014) and *SACL3* (Katayama et al., 2015). Furthermore, the expression levels of the defense gene *PLANT DEFENSIN 1.2* (*PDF1.2*) were studied as a control. *PDF1.2* is a well-studied marker for JA signaling, which is upregulated upon JA treatment. Given that transcription levels of *LOG4* increase within 30 minutes upon induction of *TMO5* and *LHW* (Smet et al., 2019), gene expression levels up to 1 hour following MeJA application were studied (**Figure 3**). As expected, *PDF1.2* expression was upregulated shortly after MeJA treatment. Expression levels started increasing after 0.25 hours and a fourfold average change was observed at 0.5 hours post treatment (**Figure 3**). Interestingly, *LOG3* as well as *LOG4* expression showed a similar trend to *PDF1.2* (**Figure 3**). 0.5 Hours after MeJA treatment, expression levels were increased three and four times respectively. In contrast, MeJA application did not affect the TMO5/LHW target genes *SACL3* and *GH10*. This RNA sequencing data thus suggests that JA, and potentially JAZ proteins, might play a role in the TMO5/LHW pathway.
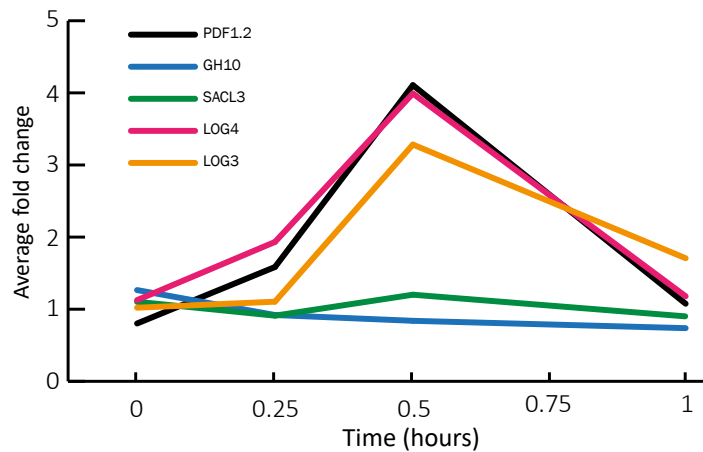
Figure 3. Temporal expression levels of MYC2 target gene PDF1.2 and TMO5/LHW target genes GH10, SACL3, LOG3 and LOG4 following exogenous application of MeJA. Reconstructed with permission from Hickman et al., 2017.

We next examined the effect of JAZ proteins on TMO5/LHW activity *in vivo*. To this end, we overexpressed tandem Tomato (tdT) fusions of *JAZ1* and *JAZ10* from the *RPS5A* promoter (**Figure 4C,G**). If JAZ proteins inhibit LHW activity, the overexpression phenotype of JAZ proteins should resemble a *LHW* loss-of-function phenotype with reduced vascular cell files containing a single xylem pole. However, the diarch architecture of Arabidopsis roots was not affected upon ectopic expression of *JAZ1* or *JAZ10* (**Figure 4A, D, H**). In contrast, ectopic expression of *JAZ1* resulted in a slight, but significant, upregulation of the number of cell files within the vasculature in one line (from 45 to 51 cells on average) (**Figure 4B, E, K**), suggesting increased rather than inhibited activity of TMO5 and LHW. This was however not observed for *JAZ10* misexpression (**Figure 4I, K**). Furthermore, upon overexpression of a single *JAZ* gene, shoot phenotypes were indistinguishable from wild (**Figure 4F, J**). To further assess the effect of JAZ proteins on the activity of the TMO5/LHW dimer, we compared the transcriptional activity of the TMO5/LHW dimer in the absence and presence of a series of JAZ proteins by transient expression assays (TEAs) in *Nicotiana tabacum* protoplasts (**Figure 5**). Given that *LOG3* and *LOG4* genes are upregulated upon MeJA treatment, p*LOG4*, fused to a *firefly luciferase* (*fLUC*) gene, was used as a read-out. Luciferase levels were not altered when TMO5 or LHW alone were added to p*LOG4*::*fLUC*. However, for one experiment (**Figure 5A**, right panel), an increase of p*LOG4*::*fLUC* was measured when LHW alone was co-expressed, suggesting the read-out of this experiment might be skewed. As expected, co-expression of both TMO5 and LHW with p*LOG4*::*fLUC*, significantly increased luciferase levels. While overexpression of *JAZ1* in Arabidopsis increased vascular cell files, co-expression of JAZ1 with TMO5 and LHW did not affect luciferase levels in the TEA (**Figure 5A**). However, p*LOG4*::*fLUC* levels were significantly increased upon co-expression with JAZ5, JAZ6 or JAZ7 (**Figure 5A**). Significant decrease of *LOG4* was only detected in three cases: JAZ10, PPD2 and TIFY8 (**Figure 5A**). Given that TEAs are generally variable, co-expression with JAZ5, JAZ6, JAZ10 and TIFY8
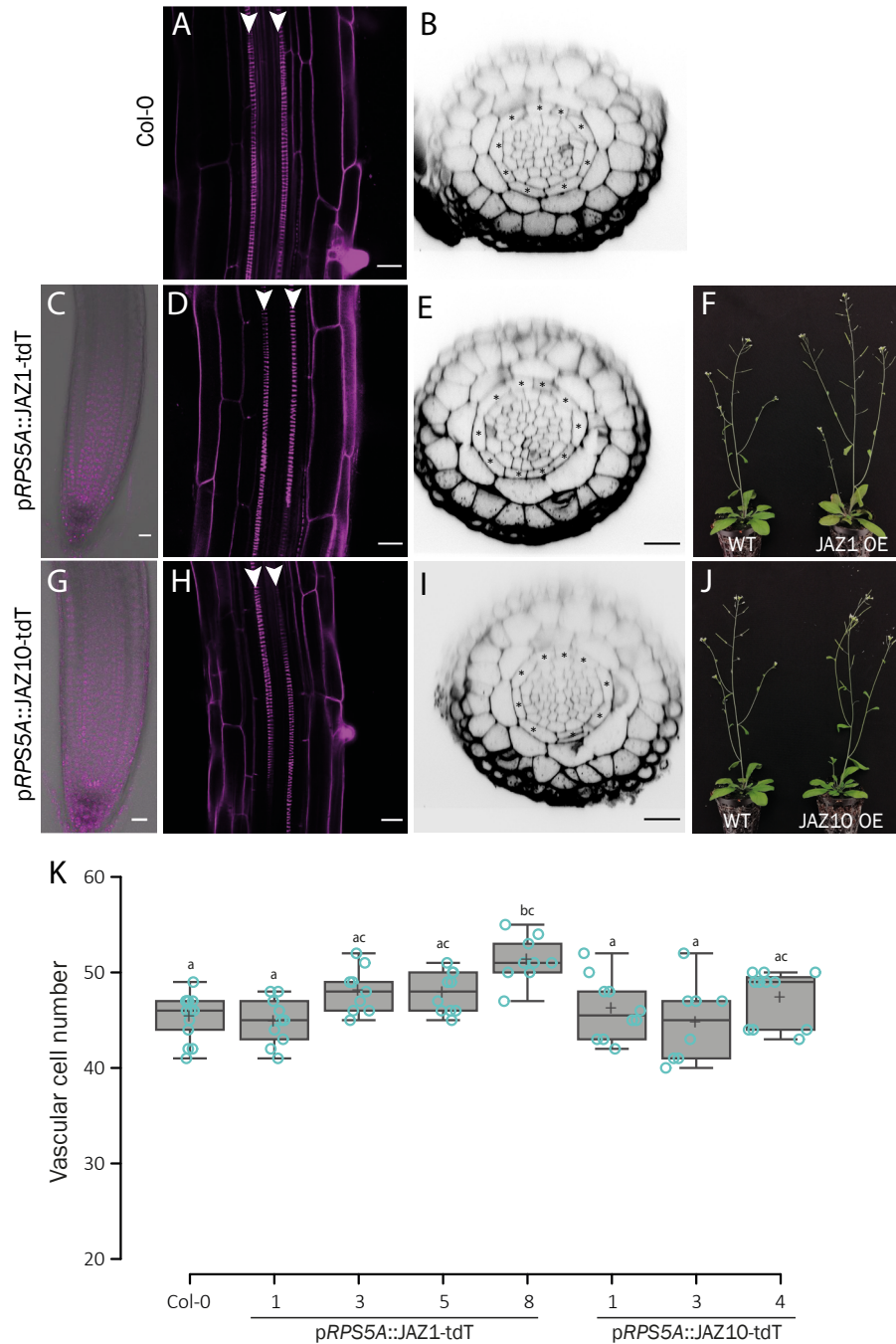
Figure 4. Single misexpression of JAZ1 (C-F) and JAZ10 (G-J) in Arabidopsis does not inhibit vascular proliferation. **(A-B)** Arabidopsis wild type roots with a diarch pattern. Arrowheads indicate xylem poles. Asterisks indicate endodermis. **(C)** p*RPS5A*::JAZ1-tdT expression in the Arabidopsis root tip. **(D)** The vascular phenotype of the JAZ1 overexpression line. Arrowheads indicate xylem strands. **(E)** A radial section of the JAZ1 overexpression line in the middle of the meristem. Asterisks indicate endodermis. **(F)** The shoot phenotype of the JAZ1 misexpression line (right) compared to a wild type plant (left). **(G)** p*RPS5A*::JAZ10-tdT expression in the Arabidopsis root tip. **(H)** The vascular phenotype of the JAZ10 overexpression line. Arrowheads indicate xylem poles. **(I)** A radial section of the JAZ10 overexpression line in the middle of the meristem. Asterisks indicate endodermis. **(J)** The shoot phenotype of the JAZ1 misexpression line (right) compared to a wild type plant (left). **(K)** Vascular cell file quantification in radial sections of 5 day-old p*RPS5A*::JAZ1-tdT and p*RPS5A*::JAZ10-tdT roots. The numbers indicate the line and turquoise circles represent individual roots. Crosses (+) indicate average values. Lines indicated with different letters are significantly different (one-way ANOVA with multiple comparisons corrected using a Tukey test, P<0.01). Scale bars represent 20 μm.

were repeated in an independent experiment (**Figure 5B**). Consistent with the initial analysis, JAZ5, JAZ6 significant increased *LOG4* expression. Furthermore, no decrease of p*LOG4*::f*LUC* was detected for any of the co-expressions. As such, these TEAs do not provide solid evidence that JAZ proteins inhibit activity of the TMO5/LHW dimer.

Taken together, the LHW GAF-like domain may structurally be similar to a JID domain, which can mediate interactions with JAZ proteins. However, no conclusive evidence was found to support JA and JAZ involvement in the TMO5/LHW pathway. N&B analyses indicate JAZ1 and LHW may be present within one complex, but direct interaction between JAZ proteins and LHW could not be demonstrated using other experimental approaches.
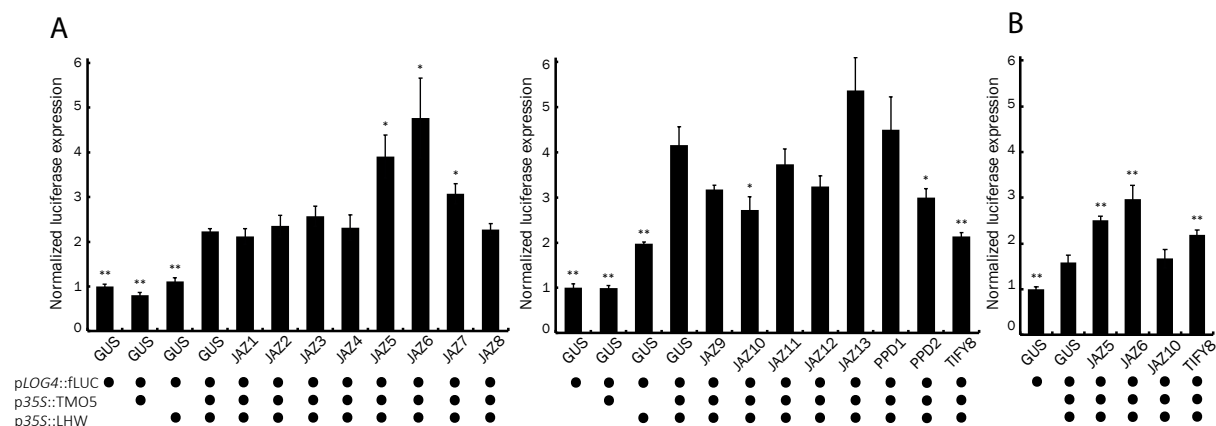
Figure 5. Dual luciferase assay in *Nicotiana tabacum* protoplasts to investigate the effect of JAZ proteins on the activity of the TMO5/LHW dimer (A) and a duplo of samples which showed a significant difference (B). Dots at the bottom indicate which of the constructs were co-transfected. Error bars represent the standard error of 4 technical replicas. Statistically significant differences, compared to p*LOG4* in combination with both TMO5 and LHW, are indicated by a single asterisk (*) at a level of $P < 0.05$ and (**) at $P < 0.01$.

The bHLH and ACT-like domains are redundant, homotypic dimerization domains

In the previous chapter, we identified the ACT-like and the bHLH domains as being crucial for LHW function. Given that dimerization is essential for DNA binding by bHLH transcription factors (Murre et al., 1989), and because the GAF-like domain does not seem to be involved in this (**Figure 1C**), we hypothesized that either the bHLH and/or the ACT-like domain is responsible for the formation of heterodimers. While it is well known that the HLH signature motif is involved in homo- or heterodimer formation, additional domains may also participate in protein-protein interactions. To investigate if the HLH or the ACT-like domains are responsible for the heterodimerization capacity of TMO5 and LHW, we performed BiFC analyses in which either the bHLH or the ACT-like domain was removed (**Figure 6A-D**). YFP halves (nYFP or cYFP) were fused to TMO5 and LHW variants and were expressed in *N. benthamiana* leaves under the strong *35S* promoter. The positive control, the full length TMO5 and LHW, showed clear interactions

(**Figure 6A**). To determine whether the heterodimerization is mediated by the HLH motif, bHLH domains of both TMO5 and LHW were deleted (TMO5[ΔbHLH] and LHW[ΔbHLH] respectively). Interestingly, removal of the bHLH domains did not impede the heterodimerization capacity of TMO5/LHW, indicating that the HLH motif is not essential in dimer formation (**Figure 6B**). We therefore expected the C-terminal ACT-like domain to be the main dimerization domain. Surprisingly, BiFC analysis revealed that heterodimerization capacities were maintained upon deletion of the LHW ACT-like domain (LHW[ΔACT]) (**Figure 6C**). These results demonstrated that removal of either the bHLH or the ACT-like domain does not affect the heterodimerization state of TMO5 and LHW. Dimerization of TMO5 and LHW was only abolished when one bHLH (TMO5[ΔbHLH]) and one ACT-like domain (LHW[ΔACT]) were deleted simultaneously (**Figure 6D**). This strongly supports the notion that both the HLH and ACT-like domain can mediate heterodimerization in a homotypic fashion.
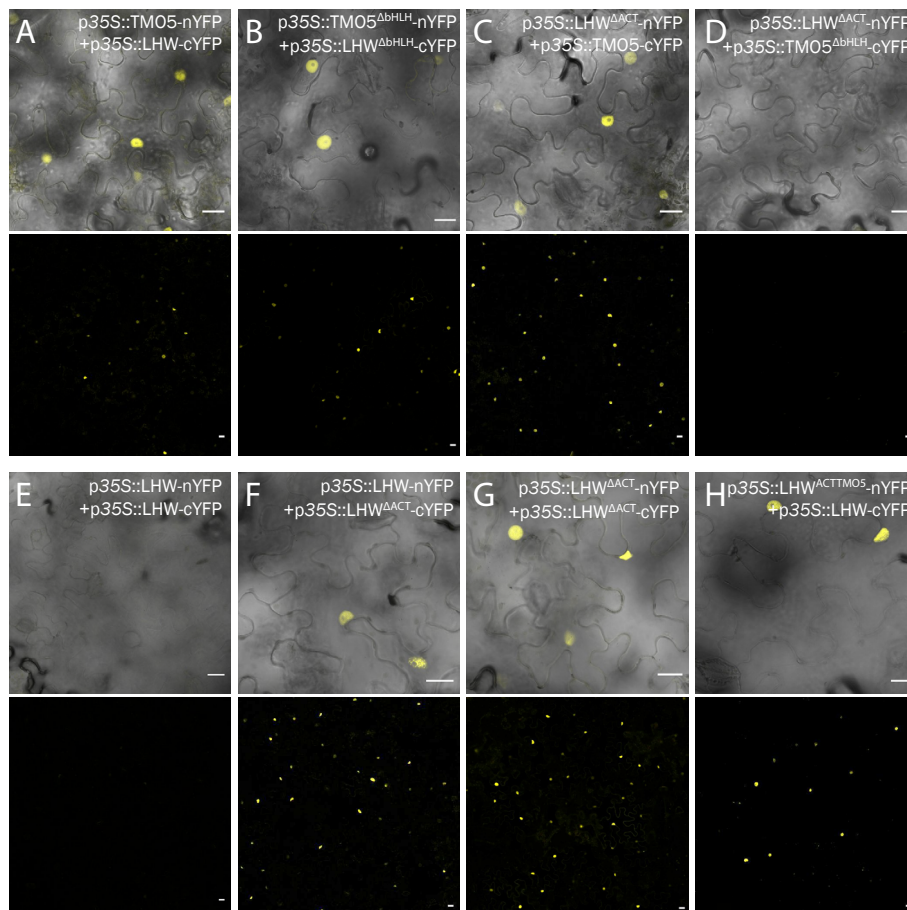
Figure 6. The role of the ACT-like domain in dimerization demonstrated by BiFC in *Nicotiana benthamiana*. **(A-D)** Heterodimerization potential of wild type TMO5 and LHW (A), TMO5 and LHW without bHLH domains (B), without the LHW ACT-like domain (C) or without the LHW ACT-like domain and the TMO5 bHLH domain (D). **(E-G)** Homodimerization capacities of LHW with (E) or without (F, G) the ACT-like domain. **(H)** Homodimerization capacity of LHW when the ACT-like domain of one LHW is replaced by the ACT-like domain of TMO5. Close-up merge of YFP and brightfield are shown in the upper panel, low magnification YFP signals are shown in the bottom panel. Scale bars represent 20 μm.

The ACT-like domain is a potential interaction specificity determinant

Having established that the ACT-like domain plays a role in heterodimerization, we further investigated its function in homodimerization using BiFC. Surprisingly, while two wild-type LHW proteins cannot interact, removal of one or both ACT-like domains resulted in homodimer complexes (**Figure 6E-G**). This could mean that LHW in its natural state can form dimers in an orientation that is incompatible with fluorescence complementation. Alternatively, the ACT-like domain may actively prevent homodimerization of LHW proteins, whilst deletion of one of these domains eliminates this function. We investigated this hypothesis by replacing the ACT-like domain of LHW with the ACT-like domain of TMO5 (LHW^ACTTMO5), creating a LHW with an overall similar structure to the native LHW, but lacking its own ACT-like domain. Fluorescence complementation was observed when LHW^ACTTMO5, fused to nYFP, was co-expressed with the wild-type LHW, fused to cYFP (**Figure 6H**). We therefore hypothesize that the ACT-like domain of LHW is an interaction specificity determinant, promoting certain interactions, while actively preventing others.

Following the results above, we reasoned that two LHW, and potentially two TMO5, ACT-like domains, i.e. two identical ACT-like domains, actively prevent interactions. To investigate this hypothesis, we first studied the interaction potential of the TMO5/LHW dimer when two identical or two distinct ACT-like domains were present. To this end, BiFC analyses were performed using TMO5 and LHW with interchanged ACT-like domains (TMO5^ACTLHW and LHW^ACTTMO5), as well as wild-type TMO5 and LHW. Interaction between wild type TMO5 and LHW was used as positive control (**Figure 7A**). In contrast to our expectations, TMO5^ACTLHW and LHW^ACTTMO5 (with two distinct ACT-like domains), were unable to interact (**Figure 7D**). Furthermore, when two identical LHW ACT-like domains were present (TMO5^ACTLHW and LHW), fluorescence complementation was detected, indicating successful heterodimerization (**Figure 7B**). In contrast, no fluorescence complementation was detected in the presence of two TMO5 ACT-like domains (TMO5/LHW^ACTTMO5) (**Figure 7C**). It is thus apparent that two identical ACT-like domains do not prevent dimerization per se.

We further verified these results in Arabidopsis plants using the biological role of the TMO5/LHW dimer as output, i.e. inducing periclinal cell divisions in the vascular tissues of the Arabidopsis root. Overexpression of TMO5 and LHW results in high levels of the active dimer and in significantly more vascular cell files (De Rybel et al., 2013). We simultaneously overexpressed the dimer components in Arabidopsis wild type plants in which the ACT-like domain was swapped in either *TMO5, LHW,* or both (**Figure 7F-I**). Optical radial sections were taken in the middle of the meristem and the number of vascular cell files was determined (**Figure 7J**). On average, wild type roots contain 45 vascular cell files (**Figure 7E, J**). In accordance with previously published data,
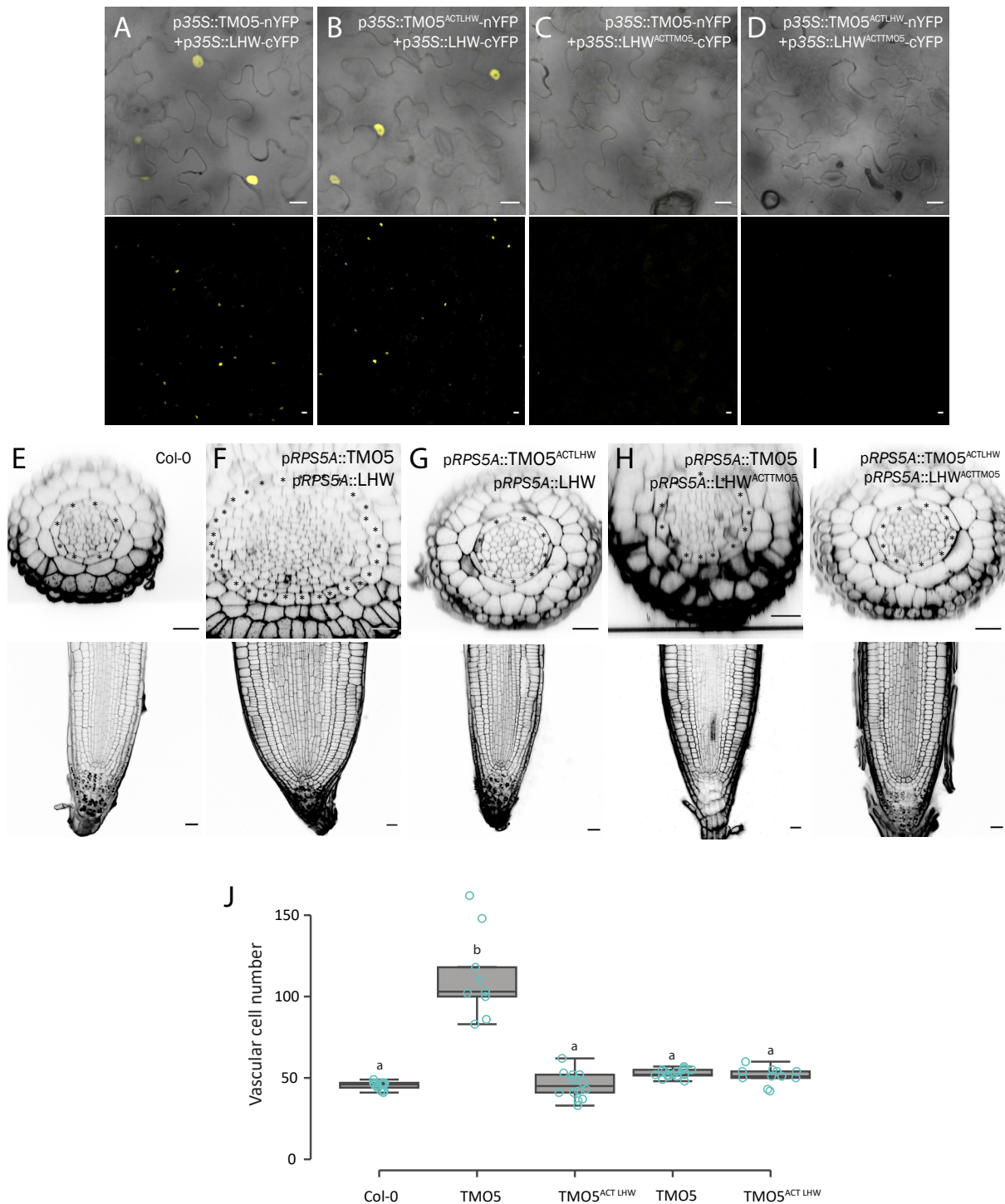
Figure 7. The ACT-like domains of TMO5 and LHW are highly specific and cannot be interchanged. (A-D) BiFC analyses of TMO5 and LHW where none (A), the TMO5 ACT-like domain alone (B), the LHW ACT-like domain alone (C) or both ACT-like domains (D) are replaced with the ACT-like domain of their interaction partner. Close-up merge of YFP and brightfield are shown in the top panel, low magnification YFP signals are shown in the bottom panel. **(E-I)** Wild type roots and roots containing overexpression lines of the aforementioned constructs. Radial sections of the root, in the middle of the meristem are shown in the upper panel. Root tips are shown in the lower panel. Asterisks indicate the endodermis. **(J)** Box plot showing vascular cell file quantification in radial sections of 5 day-old roots of the aforementioned ACT-like domain swap constructs. Turquoise circles indicate individual roots. Lines indicated with different letters can be distinguished from each other (one-way ANOVA with multiple comparisons corrected using a Tukey test, P < 0.01). Scale bars represent 20 µm.

overexpression of the native *TMO5* and *LHW* genes significantly increased the cell file number (112 vascular cell files on average) (**Figure 7F, J**). In agreement with our BiFC data, interchanging the LHW ACT-like domain (TMO5/LHW[ACTTMO5] and TMO5[ACTLHW]/LHW[ACTTMO5]) resulted in a loss of function (46 And 51 cell files respectively), possibly due to the loss of heterodimerization capacity (**Figure 7H, I, J**). While replacement of the TMO5 ACT-like domain alone did not affect dimerization potential, co-overexpression of this TMO5[ACTLHW] with LHW did not increase cell proliferation to the extent of the wild-type TMO5 and LHW proteins (**Figure 7G, J**). These results reveal that the role of the ACT-like domain regarding partner selection is complex and dependent on the protein context.

A hypothetical model for ACT-like domain interactions

Protein-protein interactions are mediated by a complex array of intermolecular and inter-surface forces, such as hydrogen bonds, hydrophobic and electrostatic interaction (Lavigne et al., 1995). To understand the determinants of the ACT-like mediated dimerization, we studied the electrostatics and hydrophobicity of the ACT-like domains of TMO5 and LHW. While hydrophobic and hydrophilic residues were divided equally in both ACT-like domains (**Supplemental Figure 2**), a generally positively charged and a generally negatively charged side exists in both ACT-like domains (**Figure 8A, B**). One observed difference is the highly negative patch in the LHW ACT-like domain, while the negative charge in more evenly distributed in the TMO5 ACT-like domain (**Figure 8A, B**). Nevertheless, the high similarity in electrostatic potential in the same N- to C- orientation, urges the question how these two ACT-like domains can mediate heterodimerization. One possible explanation is a different domain orientation, in which the positively charged side of one ACT-like domain faces the negatively charged side of the second ACT-like domain (**Figure 8C**). In opposite orientation, with two identical charges facing each other, dimerization might be prevented due to repulsion (**Figure 8C**). This hypothesis may explain why the LHW ACT-like domain engages in homodimerization in some cases, while hindering it in others.

Domain-domain orientations are, in part, determined by the sequence and conformational preferences of Inter-Domain Linkers (IDLs). α-Helical linkers often act as rigid spacers maintaining the domain orientation, but prevent domain-domain interactions (George and Heringa, 2003). In contrast, softness in linkers allows flexible hinge regions and can affect domain-domain orientation (George and Heringa, 2003). We investigated if structural differences in IDLs of TMO5 and LHW exists using secondary structure predictions via Jpred4 (Drozdetskiy et al., 2015). Interestingly, two or three β-sheets were predicted to be present in the IDL of LHW in different plant species, while TMO5 IDLs does not contain significant secondary structures and are mostly disorganized (**Figure 8D**). As such, linker structure and accordingly, ACT-like domain orientation, might be different

between TMO5 and LHW proteins. One could speculate that interchanging the LHW IDL with the TMO5 IDL, is sufficient to induce homodimerization, as this could expose a positively charged ACT side to a negatively charged side (**Figure 8C**). To investigate if the IDL of LHW indeed has a dominant role in dimerization specification of LHW, a *LHW* gene was created in which the IDL of LHW was replaced by that of TMO5 (LHW^{linkerTMO5}). This LHW^{linkerTMO5}, under the *35S* promoter, was fused to a cYFP and its interaction was determined by BiFC. While LHW^{linkerTMO5} was able to dimerize with wild-type TMO5 (**Figure 8E**), no reconstitution of YFP was observed when LHW^{linkerTMO5} was co-overexpressed with wild type LHW (**Figure 8F**). As such, the IDL between the bHLH and ACT-like domain of LHW is likely not the sole determinant for ACT-like specificity.

In summary, our data illustrates that the LHW bHLH domain is apt to homodimerize, but that the C-terminal localized ACT-like domain may restrict this capacity. Therefore, this ACT-like domain may be a driver for interaction specificity. The ACT-like domain orientation relative to the protein, determined by IDLs, might be involved in this, but additional studies are required to pinpoint its contribution and exact function in the heterodimerization process between TMO5 and LHW proteins.
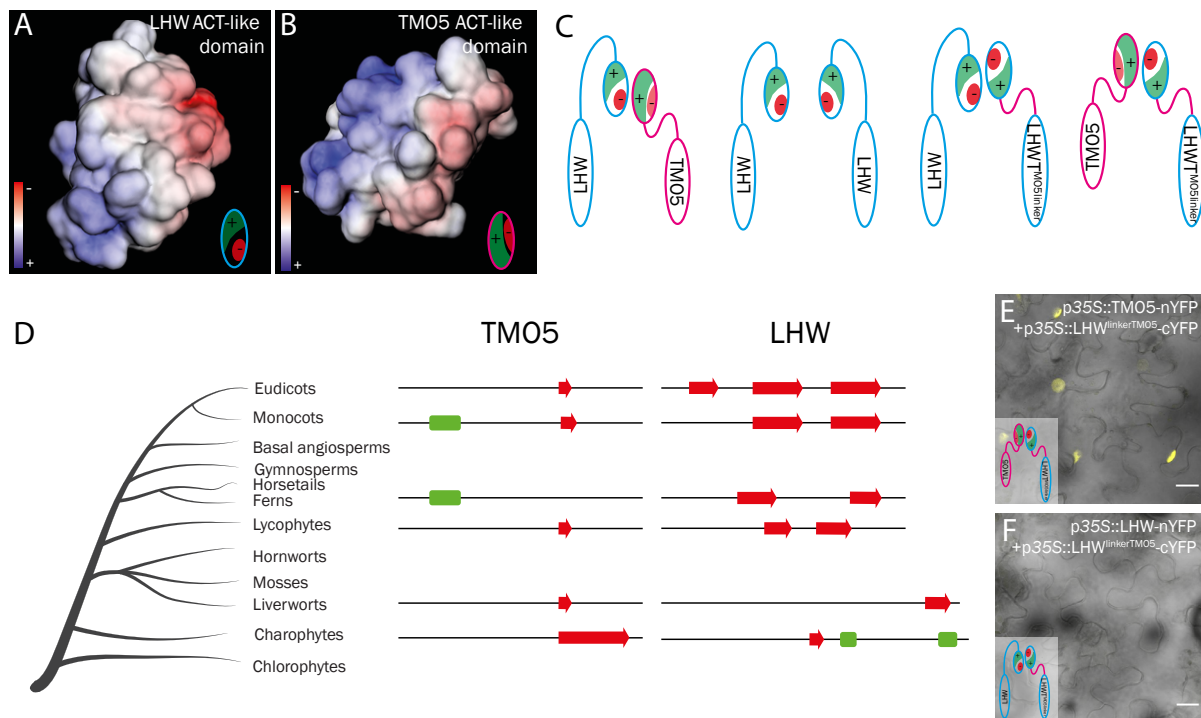
Figure 8. The study of the role of the IDLs between the bHLH and ACT-like domain in LHW and TMO5 in dimerization. (A-B) Electrostatic surfaces showing that LHW (A) and TMO5 (B) ACT-like domains possess a positively charged (in blue) and a negatively charged (in red) face. **(C)** Hypothesis of the IDL role in partner selection through the ACT-like domain. **(D)** Secondary structure prediction of the IDLs of TMO5 and LHW in different plant and algae species as determined by Jpred4. Red arraow indicate β-sheets and green tubes represent α-helices. **(E-F)** BiFC analyses of TMO5 and LHW heterodimerization (E) and LHW homodimerization (F) upon interchanging the LHW IDL with the TMO5 IDL. Scale bars represent 20 μm.

Discussion

A putative role for the GAF-like domain in LHW

Plant hormones, also referred to as phytohormones, play pivotal roles in plant growth and development. These phytohormones compose signaling networks and regulate signaling and metabolic pathways. Extensive research has revealed many individual pathways. However, during the past decades, more and more evidence has been found for complex, interconnected hormonal networks. Such cross talk allows fine-tuning of cellular processes. For example, the *JAZ1* gene is not only inducible by JA, but also by auxin (Grunewald et al., 2009). *AUXIN RESPONSE FACTOR 6 (ARF6)* and *ARF8* also form a bridge between JA and auxin, as these auxin inducible genes promote JA production in Arabidopsis flowers (Nagpal et al., 2005).

In this chapter, we investigated the domains present within the TMO5 and LHW proteins. Given that the N-terminal domain of LHW resembles a domain capable of interacting with JAZ proteins, we investigated the possibility of an interconnected network of JA and auxin that regulates vascular proliferation. We hypothesized that JAZ proteins inhibit the LHW function by interacting with the N-terminal domain. Upon JA perception, JAZ proteins would be degraded allowing LHW, together with TMO5, to transcribe its target genes. Our data indicated that JAZ1 and LHW may interact, possibly in a 2:1 ratio. While alternative methods are needed to verify this interaction, it is worth highlighting that the twelve Arabidopsis JAZ proteins are able to homo- or heterodimerize through their ZIM domain (Chini et al., 2009; Chung and Howe, 2009) and it has been suggested that JAZ proteins bind their target as a dimer (Dadalto et al., 2014; Reviewed in Memelink, 2009). This is in agreement with our findings, suggesting that LHW might interact with a homodimeric form of JAZ1. If JAZ proteins indeed need to homo- or heterodimerize in order to bind their target, removal of the ZIM domain most likely interferes with its interaction. This may explain why no fluorescence complementation was observed when only single domain JAS-nYFP and LHW[GAF]-cYFP were expressed. Furthermore, we found that TMO5/LHW target genes *LOG3* and *LOG4* are upregulated upon JA treatment. However, LOG4 levels were not suppressed when single *JAZ* genes were overexpressed. If some JAZ proteins indeed require heterodimerization with another JAZ protein, rather than homodimerization, overexpression of a single *JAZ* may not block TMO5/LHW activity. In order to investigate whether JAZ heterodimers can affect TMO5/LHW functionality, the combination of multiple JAZ proteins in TEA experiments could be tested in the future.

While we were unable to confirm that vascular proliferation in the root meristem is regulated through crosstalk between auxin and JA, some results do hint towards involvement of JA. From a biological perspective, the involvement of JA in root girth might be an elegant mechanism to fine-tune radial growth. JA is produced upon tissue damage by pathogens or insects, but also as a consequence of mechanical perturbations (Ellis et al., 2002; Hamann et al., 2009). Therefore,

the pressure the root encounters from growing into the soil, may provide a JA-induced signal to promote radial growth, creating a sufficiently strong root able to penetrate the soil.

Interestingly, JAZ1, JAZ5 and JAZ6 appeared to slightly stimulate, rather than inhibit, vascular proliferation or the expression levels of the TMO5/LHW downstream target *LOG4*. The question remains however, whether they do so in the biological system when not overexpressed. Depending on their expression pattern, these protein may not interact with LHW inside the vascular tissue and may therefore not affect vascular proliferation in their natural state. However, little is known about the expression pattern of JAZ family members such as JAZ5 and JAZ6. The JAZ1 expression pattern has been studied and was shown not to be present in Arabidopsis root tips (Pérez Cuéllar et al., 2014). Studying the expression patterns of JAZ proteins may thus be a next step in uncovering the role of these proteins in LHW signaling.

While the JAZ/LHW interaction and thus JA influence on vascular proliferation is an attractive hypothesis, we cannot exclude a different, JA-independent function of the GAF-like domain. GAF-like domains are often associated with small molecule binding and (cyclic-nucleotide-dependent) signaling (Anantharaman et al., 2001; Auldridge and Forest, 2011; Rockwell and Lagarias, 2006; Unden et al., 2013). As such, the LHW GAF-like domain could potentially bind a small molecule or co-factor. However, it has been shown that the GAF, and the highly related PAS, fold have given rise to domains with a similar fold, but very different function, such as the actin-binding protein profilin (Ho et al., 2002; Schlüter et al., 1997). The profilin GAF-like domain contributes to protein-protein interaction, rather than to small-molecule interactions (Schlüter et al., 1997). This strengthens the hypothesis that the LHW GAF-like domain is involved in protein-protein interaction. Potentially, it can interact with a yet undetermined protein containing a GAF-like domain. Immunoprecipitation coupled to mass spectrometry (IP-MS), of full length LHW as well as LHW lacking the GAF-like domain, is currently being performed to provide more detailed information on potential interaction partners of LHW through the GAF-like domain.

Contribution of the ACT-like domain in TMO5/LHW interactions

Gene regulation is a complex mechanism in which transcription factors with similar dimerization domains and even similar DNA binding motifs, can result in different outputs. While DNA binding specificity is a primary factor of transcription factor function, dimerization specificity also plays a crucial role. The fundament of such dimerization specificities are domain-domain interactions. In this chapter we found that in TMO5 and LHW, both the bHLH and the ACT-like domains can independently promote dimerization. ACT-like domains in bHLH proteins and their role in protein-protein interaction have been described before (Feller et al., 2006; Kong et al., 2012; Mas-

Droux et al., 2006). While most reports concern ACT-like domains mediating homodimerization, the Arabidopsis TMO5 and LHW ACT-like domains appear to mediate heterodimer formation. Given that multiple variants of specific domains, such as the ACT-like domain, exist in a range of proteins, a certain level of specificity is required for protein domains mediating protein-protein interaction (Feller et al., 2006). Lack of specificity can lead to promiscuous interactions and activation of improper targets (Zarrinpar et al., 2003). Therefore, a type of negative selection is essential to coordinate regulatory pathways without the problem of cross-reaction. We demonstrated that the LHW bHLH domain contains the intrinsic ability to homodimerize, but that the C-terminal located ACT-like domain restricts its dimerization potential and could potentially determine specificity to TMO5. We propose that the LHW ACT-like domain is a specificity determinant that recognizes its partner, while discriminating against related but incorrect domains. A similar mechanism was demonstrated for the bacterial bHLH protein ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR (ARNT). While the bHLH domain of ARNT on its own is able to homodimerize, in the presence of the C-terminal localized PAS domain, strictly heterodimers with the aryl hydrocarbon (dioxin) receptor are formed (Pongratz et al., 1998). In animals, the bHLH protein E12 contains an inhibitory domain, N-terminal to the bHLH, that prevents homodimerization and stimulates the formation of a heterodimer complex with MyoD (Shirakata and Paterson, 2018). The mammalian apoptosis regulating B-cell lymphoma 2 (Bcl-2) protein is another example of a protein that actively prevents homodimerization (Conus et al., 2000). In its natural state, it buries the homodimerization domain, thereby avoiding homodimer formation (Conus et al., 2000). Other examples of additional, C-terminal localized, dimerization domains able to modulate dimerization specificities include the Zip motif of bHLH-Zip factors such as Transcription factor E3 (TFE3), activating enhancer binding protein 4 (AP-4), C-myc, and myc-associated factor X (max) (Beckmann and Kadesch, 1991; Deppmann et al., 2004; Hu et al., 1990; Rishi et al., 2004). Often, changes at the amino acid level of different domains contribute to binding specificity, as was shown for the Arabidopsis ATP SULFURYLASE 1 (APS1) (Hädrich et al., 2012). However, in the case of the ACT-like domain, the same domain appears to behave differently depending on the protein it is found in (**Figure 6**).

Another method to obtain interaction specificity is by charged and polar residues (Moll et al., 2002, 2000; Newman and Keating, 2003; O'Shea et al., 1993, 1992). We identified a positive and negative interaction face in the LHW ACT-like domain. Such polar electrostatic surface potential could function as a 'molecular magnet' in which one side can interact with the oppositely charged and positioned face. It is conceivable that when similarly charged interfaces are facing each other, electrostatic repulsion occurs, while interaction takes place when oppositely charged sides are exposed to one another. A similar directional protein interaction exists within ARF5 and ARF7, where the polar electrostatic Phox and Bem1 (PB1) domain drives protein-protein interaction

through electrostatic contacts (Jez et al., 2014).

Although we were unable to pinpoint the exact mechanism behind the specification character of the ACT-like domain, given the polar electrostatic surface potential, we expected domain-domain orientation to play a major role. In the on-going IP-MS experiments, full length LHW will be compared with LHW lacking the ACT-like domain. This may give insight into the specification character of the ACT-like domain.

In summary, our results and previous reported data show that ACT-like domains appear to be complex domains due to their multitude of functions. ACT-like domains can either promote homo- or heterodimerization, prevent dimerization or simply assist, but not initiate, dimerization (Yang et al., 2012). Within the LHW protein, the ACT-like domain restricts the broad interaction capacity and possibly direct LHW towards heterodimerization with TMO5.

Material and methods

Cloning

All cloning procedures were performed using Seamless Ligation Cloning Extract (SLiCE), with 15 homologous base pairs (Zhang et al., 2015). JAZ misexpression lines were constructed by amplification of the coding sequence from Arabidopsis cDNA using Phusion Flash polymerase (Thermo Scientific). Coding sequences, without a stop codon, were cloned into the pGIIB p*RPS5A*::LIC-tdT-tNOS plasmid (pLC007). For double overexpression of TMO5 and LHW with ACT-like domain swaps, pGIIB p*RPS5A*::LIC-YFP-tNOS-p*RPS5A*::LIC-tdT-tNOS (pLC043) was used. TMO5[ACTLHW] and LHW[ACTTMO5] were generated by overlap extension PCR using existing plasmids as template. For BiFC plasmids, JAZ coding sequences were amplified from Arabidopsis cDNA, and TMO5 and LHW variants were amplified from existing plasmids by (overlap extension) PCR, without a stop codon. Coding sequences were cloned into a modified pPLV22 or pPLV27 vector containing a p*35S*::LIC-nYFP or p*35S*::LIC-cYFP respectively. For N&B plasmids, the same JAZ1 and TMO5 PCR fragment was used to create pGIIB p*35S*::*JAZ1*-mScarlet-tNOS and pGIIB p*35S*::*TMO5*-mScarlet-tNOS from pPLV25 (De Rybel et al., 2011). The N&B plasmid pGIIB p*35S*::*LHW*-mTurquoise-tNOS was constructed from pPLV25 by overlap extension PCR (De Rybel et al., 2011). All constructs were sequenced. Primers used for cloning are listed in Supplemental Table 1. The pGreenII 0800 LUC plasmid with 2 kb of *LOG4* promoter was kindly provided by Paco Vera Sirera (Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV), Valencia, Spain.

Plant material and growth conditions

Arabidopsis seeds were surface sterilized and sown on half-strength Murashige and Skoog (MS) medium with antibiotics when needed (15 µg/ml Phosphinothricin or 0.1 µg/ml Methotrexate) and placed at 4°C for 24-48 hours. Subsequentially, seeds were placed in a growth room at 22°C under long-day conditions (16 hours light, 8 hours dark). Seedlings were transferred to soil approximately 7 days after germination. Arabidopsis ecotype Columbia-0 (Col-0) was used as wild type.

Structural homology prediction and protein modeling

The primary amino acid sequence of ACT-like domains of TMO5 and LHW were submitted to the PHYRE2 protein fold recognition server (http://www.sbg.bio.ic.ac.uk/ phyre/) and to I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/), an internet based structure prediction program. The output of these analyses, a predicted structure was viewed using Pymol. Hydrophobic regions were analyzed by CAVER analyist 2.0 (Jurcik et al., 2018). Electrostatic calculations were performed using PDB2PQR with default settings (Dolinsky et al., 2004).

Bimolecular Fluorescence Complementation (BiFC)

Agrobacterium strain GV3101 (pSoup) containing BiFC plasmids were cultured overnight at 28°C/250 rpm in 5ml LB medium containing 50 µg/ml kanamycin, 25 µg/ml rifampicin, 2 µg/ml tetracyclin and 200 µM acetosyringone. The bacteria were collected by centrifugation (4000rpm, 10 minutes) and resuspended in MMA infiltration medium (20 g/l sucrose, 5 g/l MS-salts, 2 g/l MES, pH 5.6) containing 200µM acetosyringone to an optical density (OD_{600}) of 0.3. BiFC samples were mixed in a 1:1 ratio to a total OD_{600} of 0.6 and incubated for 1-2 hours at room temperature (RT) under continuous shaking. The abaxial side of the two youngest, fully expanded leaves of 5 – 6 week old *N. benthamiana* plants were infiltrated with a 1 ml needless syringe. Infiltrated leaves were harvested after 2 – 3 days and imaged using a confocal microscope.

Cross Number and Brightness (N&B) Analysis

Agrobacterium containing N&B plasmids were cultured and infiltrated in *N. benthamiana* leaves as described for BiFC. Leaves were harvested 2-3 days after infiltration and subjected to imaging and N&B analysis. A detailed step-by-step protocol is described in Clark and Sozzani (2017).

Two particles, marked with different fluorescent proteins (mScarlet and mTurquoise) were followed over time by a Leica SP5 confocal microscope. Microscope settings used were: 12 bit, xyt, 256x256 pixels, 200 Hz, line average 1, 1AU pinhole, no averaging, monodirectional scanning and 100nm pixel size using a 63x objective. We first performed single-channel N&B on monomeric variants of mScarlet (p*35S*::mScarlet) and mTurquoise (p*35S*::*LHW*-mTurquoise) as controls. Monomeric p*35S*::*LHW*-mTurquoise was excited at 458 nm (10% laser power) and detected by a HyD3 detector

between 465 and 560 nm. Control samples of p*35S*::mScarlet were excited at 561 nm (3% laser power) and detected by a PMT4 detector at 580 to 650 nm. Time series of 100 frames (minimized time per frame) were collected. For cross N&B analysis, images of nuclei containing both proteins were acquired with the same microscope settings as described above.

The SimFCS Software, developed at the Laboratory for Fluorescence Dynamics (www.lfd.uci.edu), was used to perform cross N&B analysis on scans obtained by confocal microscopy. The B1-B2 channel, within the raster image correlation spectroscopy (RICS) menu, was used to determine the correlation of pixels in the cyan channel with pixels in the red channel. The cursor was placed in the monomeric area of mTurquoise and mScarlet, as determined by monomeric controls. Brightness of both channels was determined at each correlated pixel using the Brightness cross correlation (Bcc) option. Upon activating the "Scan B1-B2" function, a stoichiometry plot was computed, displaying the most likely stoichiometry of the protein-protein complex. The percentage of monomer, 2:1 and 3:1 complexes was calculated by dividing the number of pixels inside the monomer, 2:1 or 3:1 complex cursor by the total number of fluorescent pixels.

Quantification of vascular cell numbers

5 day-old Arabidopsis roots were fixed and stained using a modified Pseudo Schiff – Propidium Iodine (mPS-PI) staining (Truernit et al., 2008). In brief, whole seedlings were fixed overnight in fixative (50% methanol and 10% acetic acid) at 4ºC. Next, seedlings were washed with MQ and incubated in 1% periodic acid at RT. After 40 min, seedlings were rinsed again with MQ and placed in Schiff reagent (100 mM sodium metabisulphite and 0.15 N HCl) containing 15 mg/ml Propidium Iodide. Seedlings were incubated for 1 – 2 hours. The samples were transferred onto a microscope slides and covered with a chloral hydrate solution (4 g chloral hydrate, 1 mL glycerol, and 2 mL water). Two small coverslips (20x20) we placed on the sides of the slide, with a larger coverslip (20x40) on top covering the samples, to avoid crushing the roots. Slides were kept overnight in the dark at RT prior to imaging. Radial cross section were obtained by a SP5 confocal microscope. All radial sections were taken in the middle of the root meristem: half way between the QC and the first elongating cortex cell. The data was visualized using BoxPlotR.

Plant imaging

5 – 7 Day-old seedlings were stained with 10 μg/ml Propidium Iodide to assess monarch and diarch phenotypes. A Leica SP5 confocal microscope, containing an Argon and a DSS561 diode laser, was used for imaging. YFP, Propidium Iodide and tandem Tomato were excited at 514 nm, 561 nm and 561 respectively. Emission was visualized with Hybrid detectors at 520-550 for YFP, 600-680 for Propidium Iodide and 570-650 for tandem Tomato.

Protoplast preparation

Tobacco BY-2 cell suspension cultures were grown in BY-2 cell suspension culture medium (43% (w/v) Murashige and Skoog (MS) medium, 30 mM KH_2PO_4, 3% (w/v) sucrose, pH 5.8) containing 1ml/L BY-2 vitamin mix (18 mM 2,4-D, 3 mM thiamine, 0.6 M myoinositol) in the dark at 25ºC while gently shaking (150rpm). BY-2 protoplasts were prepared from 100 ml, 3-day-old suspension cultures. Cells were collected by centrifugation (145 x g, 5 min, RT). The cell pellets were resuspended in freshly prepared cell wall-degrading enzyme mixture (1% cellulase Y-C, 0.1% pectolyase (Kyowa Chemical Products, Osaka, Japan), 0.4 M mannitol, 5 mM MES, pH 5.7, filter sterilized) and incubated at 25°C under gentle shaking in the dark for 1-2 hours. Protoplasts were centrifuged (145 x g, 5 min, RT) and washed three times in washing buffer (0.4 M mannitol, 2.5 mM $CaCl_2$, 1 mM MES, pH5.7). Finally, 30 ml MaMg solution (0.4 M mannitol, 15 mM MgCl2, 5 mM MES, pH 5.7) was gently added and the pellet was carefully dissolved.

Dual luciferase assay

2µg of each plasmid was distributed into a 48-well plate. The effector plasmid p2GW7-GUS was used as a control sample in order to limit technical variations and equalize the total amount of DNA. 100 µl Protoplast suspension was added per well and incubated for 10 min at RT. 120 ml PEG solution (40 % (w/v) polyethylene glycol (PEG) 3350 (Sigma-Aldrich), 0.4 M mannitol, 0.1 M $Ca(NO_3) 2·4H_2O$) was added per well and mixed gently. Following a 10 min incubation, 600 µl W5 medium (0.4 M mannitol, 154 mM NaCl, 125 mM $CaCl_2·2H_2O$, 5 mM KCl, 5 mM glucose, 1.5 mM MES, pH 5.8, filter sterilized) was added. After 10 minutes, the medium was replaced by 800 µl BY-2 medium without the vitamin mix and incubated in the dark for 20 h at RT with gentle shaking. BY-2 medium was removed the next day and protoplasts were lysed in cell culture lysis reagent (CCLR) buffer (Promega) by vigorous pipetting. 70 µl of the lysate was transferred to a new well in a 96-well readout plate. Subsequently, 100 µl Luciferase assay reagent II (LARII) (Promega) was added and immediately luminescence was measured for each well using a LUMIstar Galaxy microplate luminometer (BMG LABTECH GmbH, Ortenberg, Germany). Next, Stop&Glo reagent (Promega) was added and luminescence was measured once more. The Firefly luciferase (fLUC) activity data was normalized using Renilla luciferase (rLUC), expressed under the *35S* pomoter (p*35S*::rLUC).

Acknowledgments

Supplemental information
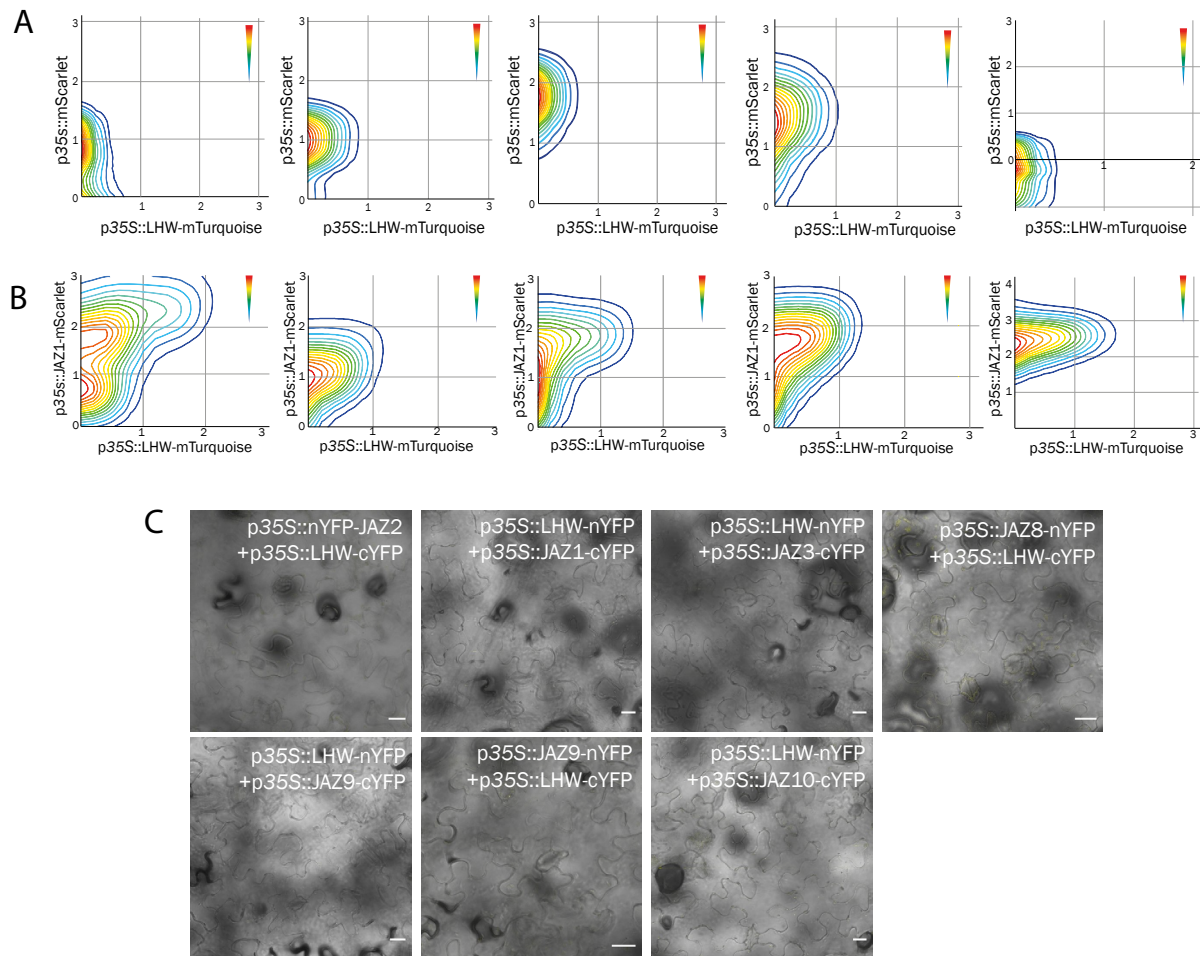
Supplemental Figure 1. Interaction studies of LHW with JAZ proteins. (A) Stoichiometry histograms of the cross N&B analysis of LHW and an empty vector (EV) containing mScarlet as negative control. **(B)** Stoichiometry histograms of the cross N&B analysis of LHW and JAZ1. The colors of the level curves represent the proportions of each complex. Red to orange lines represent a high proportion of interactions, while blue lines at intersections represent extremely low to no interactions. **(C)** BiFC analysis of LHW with various JAZ proteins. Scale bars represent 20 μm.
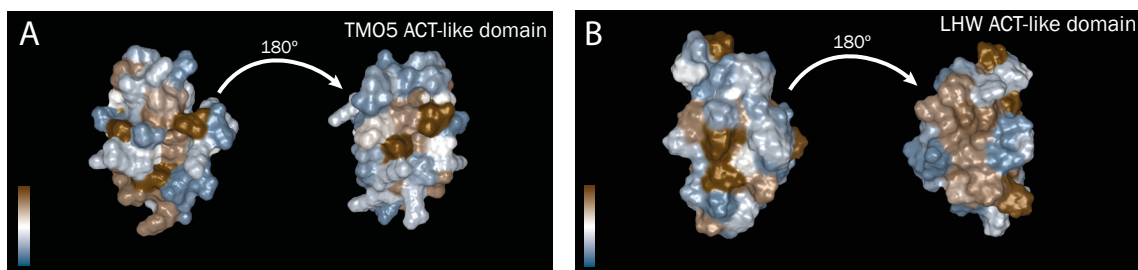
Supplemental Figure 2. Distribution of hydrophobic (brown) and hydrophylic (grey) residues in the ACT-like domain of TMO5 (A) and LHW (B).

Supplemental Tabel 1. Primers used in this study.

Primer pairs used for overexpression lines

Gene number	Gene name	Primer orientation	Sequence
AT1G19180	AtJAZ1	Sense	5'-TAG TTG GAA TAG GTT ATG TCG AGT TCT ATG GAA TGT TCT G
		Antisene	5'-GTA TGG AGT TGG GTT TAT TTC AGC TGC TAA ACC GAG CC
AT5G13220	AtJAZ10	Sense	5'-TAG TTG GAA TAG GTT ATG TCG AAA GCT ACC ATA GAA C
		Antisene	5'-GTA TGG AGT TGG GTT GGC CGA TGT CGG ATA GTA AG
AT3G25710/	TMO5^{ACTLHW}	Sense	5'-CTC TCA CCA CAG CCG ATG TAC GCA ATG AAA GAA GA
AT2G27230		Antisene	5'-AAA ATG CGA GGC GGG TTC ACC AAG TTT CCT TCC T
		Sense	5'-AGG AAG GAA ACT TGG TGA ACC CGC CTC GCA TTT T
		Antisene	5'-TTC ATC ATT ATT ATG AAC CTG CTC CAA AAT ATT CAC
		Sense	5'-GTG AAT ATT TTG GAG CAG GTT CAT AAT AAT GAT GAA
		Antisene	5'-TAT GGA GTT GGA TTT ATT ATA ACA TCG ATT CAC CAT CTT
AT2G27230/	LHW^{ACTTMO5}	Sense	5'-TTC TCT CAC CAC AGC CGA TGG GAG TTT TAC TAA G
AT3G25710		Antisene	5'-GCA AAA GGA TGC TCT TAT TAT ATC TTC TAC GAC AAT
		Sense	5'-ATT GTC GTA GAA GAT ATA ATA AGA GCA TCC TTT TGC
		Antisene	5'-TCC TCC ACA TTT CAT TGT ACA CTT CTC TAT AAC CGC
		Sense	5'-GCG TTA TAG AGA AGT GTA CAA TGA AAT GTG GAG GA
		Antisene	5'-GTA TGG AGT TGG GTT CCA TTG AAC AGC CAC CAG TAA CCG G

Primer pairs used for cross N&B and BiFC

| Gene number	Gene name	Primer orientation	Sequence
AT2G27230	LHW$^{\Delta GAF}$	Sense	5'-TAG TTG GAA TAG GTT ATG CTC GGA TGT GTT CCG GGG GCT
		Antisene	5'-GTA TGG AGT TGG GTT CCA TTG AAC AGC CAC CAG TAA CCG G
AT2G27230	(nYFP-)LHW	Sense	5'-TAG TTG GAA TAG GTT C ATG GGA GTT TTA CTA AGA G
		Antisene	5'-AGT ATG GAG TTG GGT TCT TAC ATT GAA CAG CCA CC
AT2G27230	LHW	Sense	5'-TAG TTG GAA TAG GTT C ATG GGA GTT TTA CTA AGA GA
		Antisene	5'-GTA TGG AGT TGG GTT CCA TTG AAC AGC CAC CAG TAA CC
AT1G19180	JAZ1	Sense	5'-TAG TTG GAA TAG GTT ATG TCG AGT TCT ATG GAA TGT TCT G
		Antisene	5'-GTA TGG AGT TGG GTT CTA TTT CAG CTG CTA AAC CGA GCC AC
AT1G74950	(nYFP)-JAZ2	Sense	5'-TAG TTG GAA TAG GTT ATG TCG AGT TTT TCT GCC GA
		Antisene	5'-GTA TGG AGT TGG GTT GCC GTG AAC TGA GCC AAG CTG
AT3G17860	JAZ3	Sense	5'-TAG TTG GAA TAG GTT ATG GAG AGA GAT TTT CTC GGG TT
		Antisene	5'-GTA TGG AGT TGG GTT CGG TTG CAG AGC TGA GAG AAG A
AT1G30135	JAZ8	Sense	5'-TAG TTG GAA TAG GTT ATG AAG CTA CAG CAA AAT TGT GAC TTG
		Antisene	5'-GTA TGG AGT TGG GTT CTC GTC GTG AAT GGT ACG GTG AA
AT1G70700	JAZ9	Sense	5'-TAG TTG GAA TAG GTT ATG GAA AGA GAT TTT CTG GGT TTG AG
		Antisene	5'-GTA TGG AGT TGG GTT CTG TAG GAG AAG TAG AAG AGT AAT TC
AT5G13220	JAZ10	Sense	5'-TAG TTG GAA TAG GTT ATG TCG AAA GCT ACC ATA GAA C
		Antisene	5'-GTA TGG AGT TGG GTT CGG CCG ATG TCG GAT AGT AAG G
AT2G27230	LHW^{GAF}	Sense	5'-TAG TTG GAA TAG GTT C ATG GGA GTT TTA CTA AGA GA
		Antisene	5'-GTA TGG AGT TGG GTT CAC ATC CGA GCT GCA GGA TAA GAC CC
AT1G19180	JAZ1^{JAS}	Sense	5'-TAG TTG GAA TAG GTT ATG CCA TTG ACA GAA CTT CCT ATT G
		Antisene	5'-GTA TGG AGT TGG GTT CTA TTT CAG CTG CTA AAC CGA GCC AC
AT3G25710	AtTMO5	Sense	5'-TAG TTG GAA TAG GTT C ATG TAC GCA ATG AAA GA
		Antisene	5'-GTA TGG AGT TGG GTTC AT TAT AAC ATC GAT TCA CCA TC
AT3G25710	TMO5$^{\Delta bHLH}$	Sense	5'-TAG TTG GAA TAG GTT C ATG TAC GCA ATG AAA GA
		Antisene	5'-GTA TGG AGT TGG GTT CAT TAT AAC ATC GAT TCA CCA TCT TAC
AT2G27230	LHW$^{\Delta bHLH}$	Sense	5'-TAG TTG GAA TAG GTT C ATG GGA GTT TTA CTA AGA GA
		Antisene	5'-GTA TGG AGT TGG GTT CCA TTG AAC AGC CAC CAG TAA CCG G
AT3G25710	TMO5$^{\Delta ACT}$	Sense	5'-TAG TTG GAA TAG GTT C ATG TAC GCA ATG AAA GA
		Antisene	5'-GTA TGG AGT TGG GTT CTT TTA GCT CCT TCA TGT GTT GGA TCA
AT2G27230	LHW$^{\Delta ACT}$	Sense	5'-TAG TTG GAA TAG GTT C ATG GGA GTT TTA CTA AGA GA
		Antisene	5'-GTA TGG AGT TGG GTT CGG TTT GCT TCA GCT TAT CAG AAT GC
AT2G27230/	LHW^{linkerTMO5}	Sense	5'-TAG TTG GAA TAG GTT ATG GGA GTT TTA CTA AGA GAA GCT TTA AGG
AT3G25710		Antisene	5'-ATC TGT GAT GTT TGT CGT TTT TGC TTC AGC TTA TCA GAA T
		Sense	5'-ATT CTG ATA AGC TGA AGC AAA AAC GAC AAA CAT CAC AGA T
		Antisene	5'-AAA ATG CGA GGC GGG TTC ACC AAG TTT CCT TCC T
		Sense	5'-AGG AAG GAA ACT TGG TGA ACC CGC CTC GCA TTT T
		Antisene	5'-GTA TGG AGT TGG GTTC CAT TGA ACA GCC ACC AGT AAC C

Chapter 6

General discussion

The transition from non-vascular to vascular land plants

The first land plants (Embryophytes) were Bryophyte-like plants, unable to grow beyond several centimetres. The evolution of a lignin-containing vascular system throughout the plant body, provided increased conductivity and strength and likely helped overcome size constraints. The establishment of complex vascular systems depends on acquiring vascular identity, the subsequent proliferation of specified vascular cells and the ultimate differentiation into specialized cells for conducting and supporting purposes. While lignified vascular tissues are restricted to Tracheophytes, several pathways involved in different aspects of vascular development predate the emergence of vascular plants. For example, functional auxin and cytokinin signalling pathways, which are involved in vascular proliferation, can be found in Bryophytes (Mutte et al., 2017; Pils and Heyl, 2009). Furthermore, secondary cell wall formation capacity through VASCULAR-RELATED NAC-DOMAIN (VND) transcriptional regulators is conserved in the moss *Physcomitrella patens* (Xu et al., 2014), suggesting that this pathway also predates the emergence of Tracheophytes. Finally, lignin-like compounds, which are important components of secondary cell walls, are present in Bryophytes (Espiñeira et al., 2011; Ligrone et al., 2008). Evidently, complex vascular tissue evolution was not limited by these hormonal and differentiation pathways. We thus questioned which factors may have facilitated the emergence of elaborate vascular tissues.

While water-conducting cells (WCCs) found in Bryophytes have a similar conducting function to vascular tissues, they consist of limited cell files, lacking differentiation into vessel elements. Interestingly, when TMO5 and/or LHW levels in Arabidopsis are below a critical threshold, vascular bundles with limited cell files, lacking differentiated xylem and phloem are formed (De Rybel et al., 2014; Ohashi-Ito et al., 2013). This phenotype resembles Bryophytic WCCs to some degree (**Figure 1**). Normal levels of TMO5 and LHW proteins induce periclinal and radial cell divisions in the vascular bundle and allow the formation of vascular patterns, such as the diarch pattern in Arabidopsis with two xylem and two phloem poles (**Figure 1**) (Dolan et al., 1993). Maximized TMO5 and LHW levels in Arabidopsis by double overexpression increases cell file numbers, resembling vascular bundles of larger plants such as *Populus tremula x Populus alba* (**Figure 1**). Thus, the trend observed with increasing TMO5/LHW levels in Arabidopsis resembles the evolutionary trend of vascular tissues (**Figure 1**). This leads to the hypothesis that the acquisition of vascular cell division capacity, leading to a larger bundle of cells, may have been a crucial innovation permitting vasculature tissue emergence during land plant evolution.

In this thesis we provide evidence that the evolution of the TMO5/LHW dimer was important for the development of increased vascular bundle sizes through induction of periclinal and radial cell divisions. In contrast to the previously discussed factors (hormone responses, xylem differentiation and lignin), the emergence of this regulatory dimer is clearly correlated with the emergence of
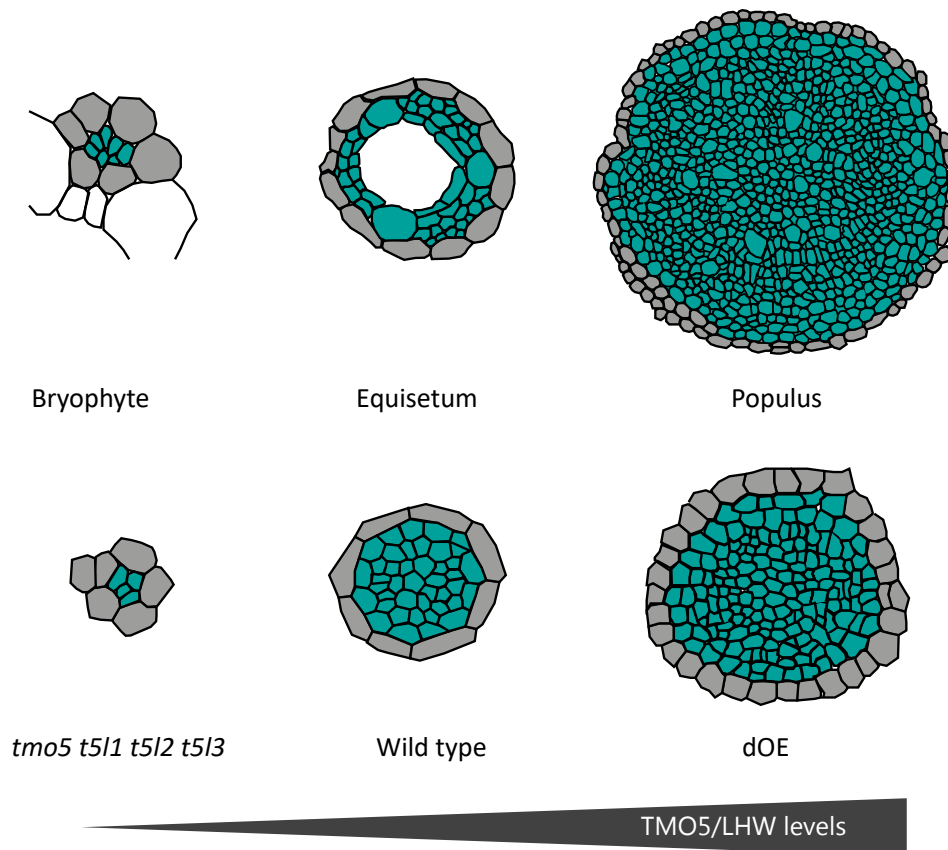
Figure 1. The evolution of vascular tissues (top) can be mimicked by TMO5/LHW expression level gradients in Arabidopsis (bottom). Top panel: schematic representation of cross sections through the vascular bundle of of the Bryophyte *Physcomitrella patens*, the early diverging Tracheophyte *Equisetum intermedium* and the tree hybrid *Populus tremula* x *Populus alba*. Bottom panel: Schematic representation of cross sections through the Arabidopsis root containing different levels of TMO5 and LHW. Water conducting and vascular cells are marked in green.

Tracheophytes (**Chapter 3**). We present a two-step model for the evolution of the TMO5/LHW dimer components in the plant lineage (**Figure 2**). In the first step, before the emergence of Embryophytes, point mutations in the basic region of the bHLH domain conferred the modern function to TMO5 (**Chapter 4**), which has been maintained ever since (**Chapter 3**). At the same time, changes to the *LHW* gene enabled heterodimerization with TMO5 (**Chapter 3**). In the second step, gain of an ACT-like domain and other, yet unknown, modifications to the *LHW* gene (**Chapter 4**), resulted in the gain of the vascular function and a preference for heterodimerization with TMO5 (**Chapter 3, 5**).

Intriguingly, the extreme conservation of TMO5 protein sequences as well as its function (**Chapter 2, 3**), suggests that modifications specifically to LHW were required to develop the vascular function of the TMO5/LHW complex. One prominent innovation within the LHW protein is the gain of the ACT-like domain with the emergence of Tracheophytes (**Chapter 4**). However, we have shown that additional evolution events must have occurred to create the vascular LHW function. Since the ACT-like and bHLH domain of LHW are sufficient to induce vascular proliferation (**Chapter 4**),

this presents the possibility that modifications in the basic region of the LHW bHLH contributed to the modern TMO5/LHW function, in a similar manner as seen for TMO5.

While evolution of vascular proliferation appears to be correlated with the rise of Tracheophytes, one can question how the emergence of vascular periclinal cell divisions can give rise to elaborate vascular systems. It should be noted that vasculatures are not solely shaped by the number of vascular cells, but also by their pattern. Experimental as well as modelling data have shown a correlation
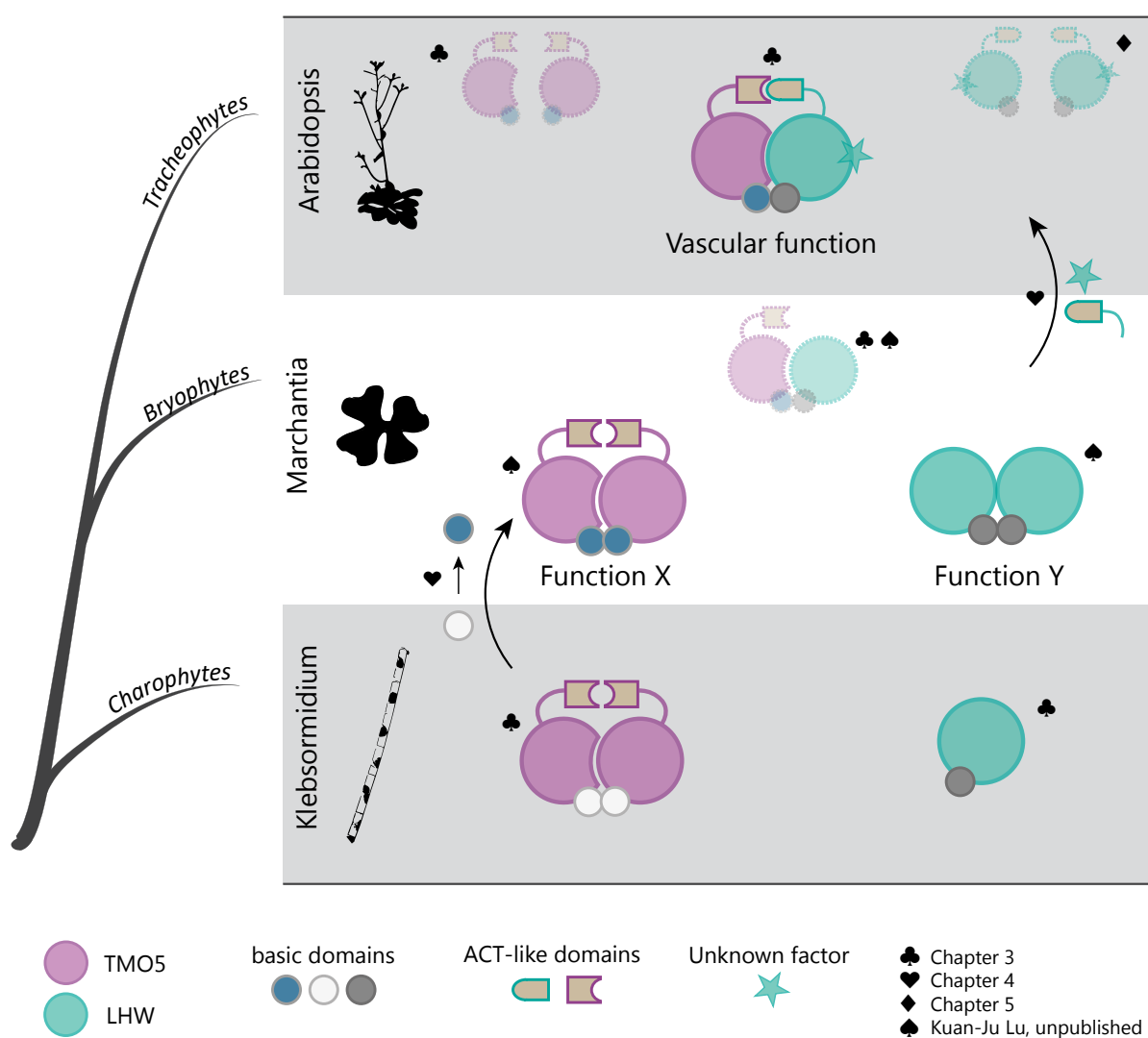
Figure 2. A two-step model of TMO5/LHW evolution, depicting the innovations in TMO5 (purple) and LHW (green) proteins towards the modern vascular function. The presumed ancestral TMO5 and LHW proteins (represented by orthologues in the extant Charophyte Klebsormidium) can form homodimers, but are unable to heterodimerize. With the emergence of Embryophytes, TMO5 (represented by the orthologue in the extant Bryophyte Marchantia) gained its modern vascular function potential through mutations in its basic domain (white and blue circles). At the same time, TMO5 and LHW acquired the capacity to form heterodimers, yet control development independent from one another. A critical innovation to LHW in the ancestor of Tracheophytes (represented by the extant Angiosperm Arabidopsis) favoured obligate heterodimerization with TMO5 and completed the modern vascular function of this dimer. These changes include mutations in the ACT-like domain (yellow/brown shapes) together with another, yet unknown modification (star). Along with this transition to obligate heterodimerization, functions of each homodimer were lost. Hearts, diamonds and clubs indicate references to chapters of this thesis. Spades refer to unpublished work by Kuan-Ju Lu (Wageningen University, the Netherlands).

between the size of the vascular cylinder during primary growth and the number of xylem and phloem poles (Mellor et al., 2019; Torrey, 1955). Thus, the patterning and complexity of vascular bundles is in part determined by spatial constraints. Reducing these constraints, by TMO5/LHW induced cell divisions, allows the formation of (additional) xylem and phloem poles. With these additional poles, vascular bundles acquire more mechanical strength. Whether this also increases axial water flow remains controversial. Some reports indicate that higher pole numbers have the potential to improve axial water flow (McMichael et al., 1985), while other studies do not observe significant changes in water flux between roots containing four and five xylem poles (Oosterhuis and Wullschleger, 1987). In contrast, a clear and direct correlation exists between the number of xylem poles and the number of lateral roots (McMichael et al., 1987, 1985). Indeed, xylem-adjoining pericycle cell files give rise to lateral roots, and an increase in xylem poles thus results in a greater potential for lateral root initiation (McMichael et al., 1987, 1985). In turn, higher numbers of lateral roots enables the plant to establish a complex root system allowing more secure anchoring into the soil and increased water and nutrient uptake under varying soil conditions. In summary, the establishment of additional vascular cell files can lead to more complex patterns and elaborate root systems.

In this thesis we used Arabidopsis mutants as a model to study the vascular function of TMO5 and LHW orthologues and variants, by quantifying vascular proliferation as well as the number of xylem poles of complementation lines (**Chapter 3, 4**). With this method we recorded transitions of a small vascular bundle, to a slightly larger bundle, resulting in little additional complexity. Even upon strong overexpression of TMO5 and LHW in Arabidopsis, anatomical complexity and the number of vascular cell files do not reach the levels of those of large, woody trees (**Figure 1**). In these trees, root girth and the high transportation capacity are mostly determined by secondary growth, rather than by primary proliferation. Unlike primary vascular growth, secondary growth is not present in all Tracheophytes, but is restricted to seed plants (Spermatophytes) (Ragni and Greb, 2018). Recently, it was shown that secondary width growth is abolished in Arabidopsis *TMO5* loss-of-function mutants, suggesting that the TMO5/LHW dimer controls secondary growth in the vascular cambium of Spermatophytes (Eliana Mor, unpublished results). This data proposes dynamics of TMO5/LHW dimer evolution within Tracheophytes and opens up the exciting possibility to explore this evolutionary track of TMO5 and LHW within Tracheophytes.

Homo- to heterodimerization transitions may be a key determinant for increased organismal complexity

While organismal complexity is challenging to define or measure accurately, comparative genome analyses revealed that 'apparent' complexity of an organism's body plan (morphology) is not

determined by the number of genes within a genome. Rather, this complexity arises from progressively more elaborate transcriptional networks, which allow the emergence of novel signalling pathways (de Mendoza et al., 2013). While such elaboration could be accomplished by duplication and expansion of transcription factors, the transcription factor number is not always proportional to organismal complexity. For example, the nematode *Caenorhabditis elegans*, comprised of approximately 1000 somatic cells and only 28 different cell types (Corsi et al., 2006; Liu et al., 2013), encodes just over 900 transcription factors (Reece-Hoyes et al., 2005). In contrast, the morphologically more complex fruit fly *Drosophila melanogaster* (with 64 different cell types) only encodes ~700 (Hammonds et al., 2013; Liu et al., 2013). Moreover, the significantly more complex humans, were reported to encode only about double the number of transcription factors compared to Drosophila (~1600 transcription factors) (Wingender et al., 2013). Similar phenomena are observed within the plant lineage: the basal land plants *Physcomitrella patens* and *Marchantia polymorpha* encode close to respectively 1200 and 400 transcription factors in their genome (Jin et al., 2017). In contrast, the more morphologically complex vascular plant *Selaginella moelendorffii* and the coniferous tree *Picea abies* encode ~700 and ~1100 transcription factors respectively (Jin et al., 2017). The expansion of transcription factor families thus appears to correlate with increasing complexity to some extent, but it is evident that other components also contribute to refined transcriptional networks and increased complexity of life.

In this thesis we showed that increased morphological complexity in plants, by the acquisition of elaborate vascular tissues, coincided with the transition from homo- to heterodimerization of transcription factors TMO5 and LHW (**Chapter 3**). Often, heterodimeric transcription factors originate from a duplicated shared ancestral protein, followed by diversification of both paralogues, while retaining dimerization capacities. (Moraitis, 1999; Tan et al., 2017; Winter et al., 2002). The obligate heterodimer formed by AtTMO5 and AtLHW (**Chapter 3**), presents an intriguing, contrasting case. These proteins emerged roughly simultaneously, during the rise of Charophytes (**Chapter 2**), but share little resemblance and belong to different bHLH subfamilies (**Chapter 2**). Moreover, dimerization capacities of TMO5 and LHW could not be demonstrated in Charophytes (**Chapter 3**). It is thus unlikely that the path towards the TMO5/LHW obligate heterodimer followed the canonical pattern described above.

In our model (**Figure 2**), we indicate a transition from a homodimerization to heterodimerization ability with the emergence of Bryophytes as a result of changes within the LHW protein (**Chapter 3**). However, in Bryophytes, this heterodimer appears to have limited biological functions (Kuan-Ju Lu, unpublished results). Obliged heterodimerization between TMO5 and LHW, as observed in Arabidopsis, thus emerged in Tracheophytes (**Chapter 3**). Currently, little is known about how obligate heterodimers are established. We propose that one of the drivers for the obligate heterodimerization between TMO5 and LHW is the ACT-like domain, found in both proteins

(**Chapter 5**). As a specificity determinant, this domain could prevent inappropriate interactions such as LHW homodimers, and promote heterodimerization with TMO5 (**Figure 2**). The correlation between the emergence of the ACT-like domain, the transition from homo- to heterodimers and the development of vascular tissues presents an interesting case. We therefore suspect that the transition from homodimeric to obligate TMO5/LHW heterodimeric complexes might be one of the key changes in transcriptional regulation of plants that resulted in enhanced morphological complexity.

Given the general trend of morphological complexity increase during evolution and the universal role of transcription factors in this process, it is tempting to extrapolate the findings of this specific case to the entire plant lineage or even beyond the plant kingdom. In other words, the evolution of obligate heterodimer transcriptional regulators might be one of the drivers for increasing elaborate transcriptional networks and, consequentially, increased organismal complexity. This is supported by the fact that transcription factors in morphologically simple organisms, such as bacteria and bacteriophages, often interact with another copy of themselves, creating homodimers (Balleza et al., 2009), whereas obligate heterodimers are more frequent in morphologically more complex organisms (Dioum et al., 2002; Schuster et al., 2015). Examples of such heterodimers include the mammalian transcription factor Neuronal PAS domain protein 2 (NPAS2), which interacts with its obligate dimeric partner Brain and Muscle ARNT-Like 1 (BMAL1) to regulate circadian rhythm (Dioum et al., 2002), the Arabidopsis flower development B-class genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) (Riechmann, 1996) and HECATEs-SPATULA, required for carpel fusion control (Schuster et al., 2015). Biologically, heterodimers have major advantages over homodimers. One such advantage is the significant increase in the amount of interactions theoretically possible in a combinatorial manner. This creates new combinations of DNA-binding regions, allowing interaction with novel targets. For example, overexpression of MpTMO5 and MpLHW homodimers in Marchantia were shown to result in respectively 29 and 172 differentially expressed genes (DEGs) compared to wild type (Kuan-Ju Lu, unpublished results). Significantly more DEGs were identified when AtTMO5/AtLHW heterodimers were overexpressed in Arabidopsis, namely 822 DEGs compared to wild type levels (Kuan-Ju Lu, unpublished results). Another advantage of heterodimeric complexes is the fact that the two protein subunits can be regulated independently at transcriptional and (post-) translational level. This could create, for instance, spatially appropriate patterns in which broad or ubiquitous expressed transcription factor interact with tissue-restricted partners. Finally, monomers with unique properties can generate a heterodimer with characteristics very distinct from the homodimeric forms.

Conclusion

In this thesis, we revealed that vascular tissue regulators TMO5 and LHW have deep evolutionary origins and that a two-step innovation process resulted in the vascular function of the obligate heterodimer. The results in this research provided evidence that this specific vascular function of the dimer, i.e. the periclinal and radial cell division inducing competency, could be one of the key evolutionary innovation that enabled the emergence of Tracheophytes. We further provide an indication that increase of organismal complexity may be reached by creating novel heterodimer complexes of existing transcription factors.

6

6

Appendices

English summary

Glossary of terms

References

English summary

The emergence of vascular plants (Tracheophytes) has had great impacts on our planet and now make up the majority of the biomass on Earth. The innovation of vascular tissues in plants allowed the colonization of a larger habitats and subsequently created the opportunity for new ecosystems to arise. Vascular tissues can thus be regarded as one of the key evolutionary innovations in the plant lineage. Yet, the molecular innovations that led to the evolution of these conductive tissues are unknown.

Chapter 1 introduced the evolutionary trajectory of plants starting at the green algal ancestor. We discussed the closest relatives to land plants (Embryophytes): the fresh water green algae Charophytes, as well as the earliest diverging land plants lacking elaborate vascular tissues: Bryophytes. In addition, we introduced water-conducting cells (WCCs) in Bryophytes and the vascular tissues in Tracheophytes. We described the molecular mechanisms underlying vascular cell proliferation in Arabidopsis, in which the TARGET OF MONOPTEROS 5 (TMO5) / LONESOME HIGHWAY (LHW) heterodimer induces periclinal and radial cell divisions in the procambium through cytokinin (CK) production.

In **Chapter 2**, we investigated the origin of the *TMO5* and *LHW* genes using bioinformatical analyses. Phylogenetic trees revealed that both genes can be found in Tracheophytes, Bryophytes and Charophytes, suggesting that both vascular regulatory genes have ancient origins and are not restricted to Tracheophytes. Since TMO5 and LHW operate in a pathway that also includes other proteins such as LONELY GUY (LOG) proteins and SAC51-LIKE (SACL) proteins, we next investigated the conservation of these proteins using phylogenetic studies. We revealed that LOG proteins have similar deep evolutionary origins as TMO5 and LHW. In contrast, SACL proteins are likely restricted to Embryophytes. We concluded that the presence of *TMO5* and *LHW* genes is not the determinant for complex vascular tissue formation and that the different genes involved in the TMO5/LHW pathway in Arabidopsis emerged at different timepoints in plant evolution.

Having established when TMO5 and LHW emerged approximately, we questioned if the establishment of the vascular-specific function of TMO5 and LHW was the determinant for complex vascular tissue formation. In **Chapter 3**, we performed complementation assays, using *tmo5 t5l1* double and *lhw* single mutants, in order to assess the competence of TMO5 and LHW orthologues to induce periclinal and radial cell divisions in Arabidopsis. We showed that the vascular function of TMO5 was established in the ancestor of Bryophytes, whereas LHW obtained its modern function in the ancestor of Tracheophytes. Bimolecular Fluorescence Complementation (BiFC) assays showed that heterodimerization capacities of TMO5 and LHW predate Tracheophytes. However, using single and double misexpressions of Marchantia TMO5 and LHW in Arabidopsis, combined with a p*LOG4*::3GFP, we showed that the Bryophytic heterodimer does not have the

same biological function as the Arabidopsis dimer. Together, these results suggest that the TMO5/LHW dimer may have been an important driver for the establishment of complex vascular tissues.

The functional innovations of TMO5 and LHW, uncovered in Chapter 3, raised the question which changes to the protein were at the basis for these functional innovations. In **Chapter 4**, we investigated this based on domain architectures. We demonstrated that TMO5 contains two domains (a bHLH and an ACT-like domain), which are highly conserved. LHW was shown to contain three domains in Tracheophytes and hornworts (a GAF-like, bHLH and an ACT-like domain) and two domains in Charophytes , liverworts and mosses. The ACT-like domain was likely acquired in the ancestor of Tracheophytes or Bryophytes. While the ACT-like domain, together with the bHLH domain, is necessary and sufficient for the vascular function of LHW, we reported that the acquisition of the ACT-like domain was not the sole determinant for the development of the vascular function. We further revealed that mutations to the DNA-binding domain of the TMO5 bHLH conveyed the vascular function.

In **Chapter 5**, we dug deeper into the domains of TMO5 and LHW. While the LHW GAF-like domain structurally resembles a JAZ-interaction domain (JID), no conclusive evidence was found to proof a hormonal cross-talk between Jasmonic acid (JA) and CK in the TMO5/LHW pathway. We did reveal that the bHLH and ACT-like domains of TMO5 and LHW are redundant, homotypic dimerization domains. In addition, we showed that the bHLH domains of LHW proteins have the propensity to homodimerize, but that the C-terminal ACT-like domain restricts this capacity. We thus hypothesize that the ACT-like domain of LHW could be a specification determinant, promoting heterodimerization with TMO5.

Finally, in **Chapter 6**, we combined the information of this thesis and reflected on how a transition from non-vascular to vascular plants could have occurred. We proposed that TMO5 and LHW may have been one of the main determinants for the emergence of elaborate vascular tissues, since the TMO5/LHW dimer able to execute the vascular function, emerged in the ancestor of Tracheophytes. We highlighted that increased vascular girth by periclinal and radial cell divisions contributes to vascular complexity. We further speculated that the homo- to heterodimerization transition during the evolutionary trajectory, as seen for TMO5 and LHW, may be extrapolated to to the entire plant lineage. The evolution of obligate heterodimer transcriptional regulators might be one of the drivers for increasing elaborate transcriptional networks and, consequentially, increased morphological complexity.

Glossary of terms

Anticlinal divisions
: A cell division process where the division plane is perpendicular to the plant surface, adding new cells to an existing cell file and thus results in longitudinal growth.

Axial water flow
: Water movement through the plant, parallel to the plant axis.

Bryophytes
: Small green non-vascular land plants (mosses, liverworts and hornworts) that form the most basal group of extant land plants.

Cambium
: A layer of actively dividing cells (meristem) that gives rise to secondary xylem and secondary phloem during secondary growth in stems and roots.

Cellular differentiation
: The processes by which a cell transforms from one cell type into a different, often more specialized, cell type.

Charophyte
: A group of freshwater green algae that are the closest living relatives to Embryophytes.

Chlorophytes
: A diverse group of uni- and multicellular green algae in marine, freshwater and terrestrial habbitats.

Clade
: See monophyletic group.

Cotyledons
: The embryonic leaf/leaves of a plant.

Cuticle
: A protective waxy layer covering the epidermis of leaves, young shoots and other aerial plant organs to limit water loss.

Cyanobacteria
: Aquatic bacteria that obtain their energy through photosynthesis (also referred to as blue-green algae).

Differentiation
: See Cellular differentiation.

Embryophytes
: Land plants.

Endosymbiotic event
: The evolutionary event where a unicellular protist engulfed a free-living prokaryotes and retaining it, eventually leading to mitochondria and plastids.

Flagella
: A microscopic hair-like rotary apparatus used for movement by microorganisms.

Fronds
: Large, heavily divided leaves commonly found on ferns and palms.

Furrowing
: Generation of the cleavage furrow: a shallow groove in the cell surface that marks the site of cytokinesis.

Gametophyte
: The haploid, sexual phase in the two alternating phases in the life cycle of plants and certain algae.

Green algae — A paraphyletic group consisting of all Chlorophytes and Charophytes.

Gymnosperms — A group of woody flowerless plants that produce 'naked' seeds.

Homologue — Genes or proteins that have a common ancestor and can either be orthologues or paralogues.

Hornworts — Small, green, non-flowering and non-vascular plants with pipe-like structures and unicellular rhizoids that belong to the Bryophytes.

Liverworts — Small, green, non-flowering and non-vascular plants with lobed thallus (leaf-like structure) and unicellular rhizoids that belong to the Bryophytes.

Meristem — Tissue in plants containing undifferentiated cells that actively divide and allow indeterminate growth.

Metaxylem — Primary xylem that differentiates after protoxylem, characterized by broad cells with pitted secondary cell wall depositions.

Monophyletic (group) — A group of organisms, sometimes called a clade, that includes the most recent common ancestor and all of its descendents.

Mosses — Small, green, non-flowering and non-vascular plants with a central midrib and multicellular rhizoids that belong to the Bryophytes.

Orthologue — Homologous genes between two species (that evolved from a common ancestral gene), but not necessarily able to perform the same function.

Paralogue — Homologous genes within a species, as a result of a recent duplication.

Patterning — The establishment of organized cell fates within a tissue in the plant through organized cell divisions and cell differentiation.

Periclinal divisions — Cell divisions that occur parallel to the surface of the plant body and perpendicular to the plant body axis, resulting in radial growth.

Phragmoplast — A structure of cytoskeletal arrays that positions the new cell wall during cytokinesis in plants and some algae.

Plasmodesmata — Small pores in cell walls that function as microscopic channels, facilitating cell-cell communication and transport in plants and some algae.

Plastid — A small, self-reproducing organelle of plants and algae such as chloroplasts.

Procambium — Part of the primary meristem of a plant that forms cambium and primary xylem and phloem.

Protoxylem — Primary xylem, developed from the procambium and is characterized by spiral or ring-shaped cell wall thickenings.

Pterophytes — Seedless, vascular plants containing horsetails and ferns.

Radial divisions   Cell divisions that occur perpendicular to the surface of the plant body and parallel to the plant body axis, resulting in radial growth.

Rhizoids   Uni- or multicellular roothair-like structures in Bryophytes, allowing anchoring to the soil.

Secondary growth   Radial growth of stems and roots that occurs after primary growth by divisions in cambial cells in seed plants.

Sporangium   An enclosed structure in which reproductive spores are formed.

Sporophyte   The diploid, multicellular phase in the two alternating phases in the life cycle of plants and certain algae.

Stomata   Small pores in the epidermis of plants that can be opened or closed to regulate gas exchange.

Streptophytes   All land plants and Charophytes together.

Tracheophytes   All vascular plants.

Viridiplantae   A clade of eukaryotic organisms made up of all green algae (Chlorophytes and Charophytes) and all land plants.

References

Al-Rabiee, R., Lee, E.J., Grant, G.A., 1996. The mechanism of velocity modulated allosteric regulation in D-3-phosphoglycerate dehydrogenase: Cross-linking adjacent regulatory domains with engineered disulfides mimics effector binding. J. Biol. Chem. 271(22), 13013–13017.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2008. Molecular biology of the cell, fifth ed.

Amborella Genome Project, T., 2013. The Amborella Genome and the Evolution of Flowering Plants. Science. 342, 1467–1477.

Anantharaman, V., Koonin, E. V, Aravind, L., 2001. Regulatory Potential , Phyletic Distribution and Evolution of Ancient , Intracellular Small-molecule- binding Domains. J. Mol. Biol. 307, 1271–1292.

Aravind, L., Koonin, E. V., 1999. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J. Mol. Biol. 287(5), 1023–1040.

Aravind, L., Ponting, C.P., 1997. The GAF domain: an evolutionary link between diverse phototransducing proteins. Plant Cell 04(97), 17–18.

Atchley, W.R., Terhalle, W., Dress, A., 1999. Positional dependence, cliques, and predictive motifs in the bHLH protein domain. J. Mol. Evol. 48(5), 501–516.

Auldridge, M.E., Forest, K.T., 2011. Bacterial phytochromes: More than meets the light. Crit. Rev. Biochem. Mol. Biol. 46(1), 67–88.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C.F., Coburn, D., Newburger, D.E., Morris, Q., Hughes, T.R., Bulyk, M.L., 2009. Diversity and complexity in DNA recognition by transcription factors. Science. 324(5935), 1720–1723.

Balleza, E., López-Bojorquez, L.N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y.I., Encarnación, S., Collado-Vides, J., 2009. Regulation by transcription factors in bacteria: Beyond description. FEMS Microbiol. Rev. 33(1), 133–151.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., DePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., Ashton, N.W., Axtell, M.J., Barker, E., Barker, M.S., Bennetzen, J.L., Bonawitz, N.D., Chapple, C., Cheng, C., Correa, L.G.G., Dacre, M., DeBarry, J., Dreyer, I., Elias, M., Engstrom, E.M., Estelle, M., Feng, L., Finet, C., Floyd, S.K., Frommer, W.B., Fujita, T., Gramzow, L., Gutensohn, M., Harholt, J., Hattori, M., Heyl, A., Hirai, T., Hiwatashi, Y., Ishikawa, M., Iwata, M., Karol, K.G., Koehler, B., Kolukisaoglu, U., Kubo, M., Kurata, T., Lalonde, S., Li, K., Li, Y., Litt, A., Lyons, E., Manning, G., Maruyama, T., Michael, T.P., Mikami, K., Miyazaki, S., Morinaga, S.I., TakashiMurata, Mueller-Roeber, B., Nelson, D.R., Obara, M., Oguri, Y., Olmstead, R.G., Onodera, N., Petersen, B.L., Pils, B., Prigge, M., Rensing, S.A., Riaño-Pachón, D.M., Roberts, A.W., Sato, Y., Scheller, H.V., Schulz, B., Schulz, C., Shakirov, E. V, Shibagaki, N., Shinohara, N., Shippen, D.E., Sørensen, I., Sotooka, R., Sugimoto, N., Sugita, M., Sumikawa, N., Tanurdzic, M., Theißen, G., Ulvskov, P., Wakazuki, S., Weng, J.K., Willats, W.W.G.T., Wipf, D., Wolf, P.G., Yang, L., Zimmer, A.D., Zhu, Q., Mitros, T., Hellsten, U., Loqué, D., Otillar, R., Salamov, A., Schmutz, J., Shapiro, H., Lindquist, E., Lucas, S., Rokhsar, D., Grigoriev, I. V, 2011. The selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science. 332(6032), 960–963.

Bar-On, Y.M., Phillips, R., Milo, R., 2018. The biomass distribution on Earth. Proc. Natl. Acad. Sci. 25, 6506–6511.

Bauby, H., Divol, F., Truernit, E., Grandjean, O., Palauqui, J.C., 2007. Protophloem differentiation in early Arabidopsis thaliana development. Plant Cell Physiol. 48, 97–109.

Beck, C.B., 1966. On the Origin of Gymnosperms. Taxon 15(9), 337–339.

Beckmann, H., Kadesch, T., 1991. Search : The leucine zipper of TFE3 dictates helix-loop-helix dimerization. Genes Dev 5, 1057–1066.

**Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S.A., Morris, Q.D., Bulyk,

M.L., Hughes, T.R., 2008. Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. Cell 133(7), 1266–1276.

Bhattacharya, D., Medlin, and L., 1998. Algal Phylogeny and the Origin of Land Plants. Plant Physiol. 116, 9–15.

Bonke, M., Thitamadee, S., Mähönen, A.P., Hauser, M.T., Helariutta, Y., 2003. APL regulates vascular tissue identity in Arabidopsis. Nature 426, 181–186.

Callaway, E., 2013. Shrub genome reveals secrets of flower power. Nat. News.

Cao, H., Qi, S., Sun, M., Li, Z., Yang, Y., Crawford, N.M., Wang, Y., 2017. Overexpression of the maize ZmNLP6 and ZmNLP8 can complement the Arabidopsis nitrate regulatory mutant nlp7 by restoring nitrate signaling and assimilation. Front. Plant Sci. 8, 1703.

Carlsbecke, A., Lee, J.-Y., Roberts, C.J., Dettmer, J., Lehesranta, S., Zhou, J., Lindgren, O., Moreno-Risueno, M.A., Vatén, A., Siripong, T., Campilho, A., Sebastian, J., Bowman, J.L., Helariutta, Y., Benfey, P.N., 2010. Cell signalling by microRNA165/6 directs gene dose-dependent root. Nature 465(7296), 316–321.

Carretero-paulet, L., Galstyan, A., Roig-villanova, I., Martínez-García, J.F., Bilbao-castro, J.R., Robertson, D.L., 2010. Genome-Wide Classification and Evolutionary Analysis of the bHLH Family of Transcription Factors in. Plant Physiol. 153(July), 1398–1412.

Chang, Y., Graham, S.W., 2011. Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. Am. J. Bot. 98(5), 839–849.

Chaw, S.M., Chang, C.C., Chen, H.L., Li, W.H., 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. J. Mol. Evol. 58(4), 424–441.

Chini, A., Fonseca, S., Chico, J.M., Fernández-Calvo, P., Solano, R., 2009. The ZIM domain mediates homo- and heteromeric interactions between Arabidopsis JAZ proteins. Plant J. 59(1), 77–87.

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J.M., Lorenzo, O., García-Casado, G., López-Vidriero, I., Lozano, F.M., Ponce, M.R., Micol, J.L., Solano, R., 2007. The JAZ family of repressors is the missing link in jasmonate signalling. Nature 448(7154), 666.

Chipman, D., Shaanan, B., 2001. The ACT domain family David M Chipman* and Boaz Shaanan. Curr. Opin. Struct. Biol. 11, 694–700.

Chung, H.S., Howe, G.A., 2009. A Critical Role for the TIFY Motif in Repression of Jasmonate Signaling by a Stabilized Splice Variant of the JASMONATE ZIM-Domain Protein JAZ10 in Arabidopsis. Plant Cell Online 21(1), 131–145.

Church, G.M., Sussman, J.L., Kim, S., 1977. Secondary structural complementarity between DNA and proteins. Proc. Natl. Acad. Sci. U. S. A. 74(4), 1458–1462.

Clark, N.M., Hinde, E., Winter, C.M., Fisher, A.P., Crosti, G., Blilou, I., Gratton, E., Benfey, P.N., Sozzani, R., 2016. Tracking transcription factor mobility and interaction in arabidopsis roots with fluorescence correlation spectroscopy. Elife 5(JUN2016), 1–25.

Clark, N.M., Sozzani, R., 2017. Measuring protein movement, oligomerization state, and protein–protein interaction in Arabidopsis roots using scanning fluorescence correlation spectroscopy (Scanning FCS). In: Methods in Molecular Biology. pp. 251–266.

Conus, S., Kaufmann, T., Fellay, I., Otter, I., Rosse, T., Borner, C., 2000. Bcl-2 is a monomeric protein: prevention of homodimerization by structural constraints. EMBO J. 19(7), 1534–1544.

Corsi, A.K., Markaki, M., Tavernarakis, N., Sawa, H., Korswagen, H.C., 2006. A biochemist's guide to C. elegans. Anal. Biochem. 359(1), 1–17.

Cui, H., Levesque, M.P., Vernoux, T., Jung, J.W., Paquette, A.J., Gallagher, K.L., Wang, J.Y., Blilou, I., Scheres, B., Benfey, P.N., 2007. An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. Science. 316, 421–425.

Curien, G., Biou, V., Mas-Droux, C., Robert-Genthon, M., Ferrer, J.L., Dumas, R., 2008. Amino acid biosynthesis: New architectures in allosteric enzymes. Plant Physiol. Biochem. 46(3), 325–339.

References

Dadalto, S., Alves, M., Fietto, L., Barros, V., de Souza, G., Gonçalves, A., 2014. Transcription Factor Functional Protein-Protein Interactions in Plant Defense Responses. Proteomes 2(1), 85–106.

de Mendoza, A., Sebe-Pedros, A., Sestak, M.S., Matejcic, M., Torruella, G., Domazet-Loso, T., Ruiz-Trillo, I., 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. 110(50), E4858–E4866.

De Rybel, B., Adibi, M., Breda, A.S., Wendrich, J.R., Smit, M.E., Novak, O., Yamaguchi, N., Yoshida, S., Van Isterdael, G., Palovaara, J., Nijsse, B., Boekschoten, M. V, Hooiveld, G., Beeckman, T., Wagner, D., Ljung, K., Fleck, C., Weijers, D., 2014. Integration of growth and patterning during vascular tissue formation in Arabidopsis. Science. 345(6197), 1255215–1255215.

De Rybel, B., Möller, B., Yoshida, S., Grabowicz, I., Barbier de Reuille, P., Boeren, S., Smith, R.S., Borst, J.W., Weijers, D., 2013. A bHLH Complex Controls Embryonic Vascular Tissue Establishment and Indeterminate Growth in Arabidopsis. Dev. Cell 24(4), 426–437.

De Rybel, B., van den Berg, W., Lokerse, A.S., Liao, C.-Y., van Mourik, H., Moller, B., Llavata-Peris, C.I., Weijers, D., 2011. A Versatile Set of Ligation-Independent Cloning Vectors for Functional Studies in Plants. Plant Physiol. 156(3), 1292–1299.

Delwiche, C.F., Cooper, E.D., 2015. The evolutionary origin of a terrestrial flora. Curr. Biol. 25(19), R899–R910.

Deppmann, C.D., Acharya, A., Rishi, V., Wobbes, B., Smeekens, S., Taparowsky, E.J., Vinson, C., 2004. Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: A comparison to Homo sapiens B-ZIP motifs. Nucleic Acids Res. 32(11), 3435–3445.

Dioum, E.M., Rutter, J., Tuckerman, J.R., Gonzalez, G., Gilles-Gonzalez, M.A., McKnight, S.L., 2002. NPAS2: A gas-responsive transcription factor. Science. 298(5602), 2385–2387.

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

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., Baker, N.A., 2004. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32, W665–W667.

Doyle, J.A., 2012. Molecular and Fossil Evidence on the Origin of Angiosperms. Annu. Rev. Earth Planet. Sci. 40, 301–326.

Drábková, L.Z., Dobrev, P.I., Motyka, V., 2015. Phytohormone profiling across the bryophytes. PLoS One 10(5), 1–19.

Drozdetskiy, A., Cole, C., Procter, J., Barton, G.J., 2015. JPred4: A protein secondary structure prediction server. Nucleic Acids Res. 43, W389–W394.

Duff, R.J., Nickrent, D.L., 1999. Phylogenetic Relationships of Land Plants using Mitochondrial Small-Subunit rDNA Sequences. Am. J. Bot. 86(3), 372–386.

Edger, P.P., Pires, J.C., 2009. Gene and genome duplications : the impact of dosage-sensitivity on the fate of nuclear genes. Chromosom. Res. 17(5), 699–717.

Ellis, C., Karafyllidis, I., Wasternack, C., Turnera, J.G., 2002. The Arabidopsis Mutant cev1 Links Cell Wall Signaling to Jasmonate and Ethylene Responses. Plant Cell Online 14, 1557–1566.

Espíñeira, J.M., Novo Uzal, E., Gómez Ros, L. V., Carrión, J.S., Merino, F., Ros Barceló, A., Pomar, F., 2011. Distribution of lignin monomers and the evolution of lignification among lower plants. Plant Biol. 13, 59–68.

Ettema, T.J.G., Brinkman, A.B., Tani, T.H., Rafferty, J.B., Van Oost, J. Der, 2002. A novel ligand-binding domain involved in regulation of amino acid metabolism in prokaryotes. J. Biol. Chem. 277(40), 37464–37468.

Feller, A., Hernandez, J.M., Grotewold, E., 2006. An ACT-like domain participates in the dimerization of several plant basic-helix-loop-helix transcription factors. J. Biol. Chem. 281(39), 28964–28974.

Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J.-M., Gimenez-Ibanez, S., Geerinck, J., Eeckhout, D., Schweizer, F., Godoy, M., Franco-zorrilla, J.M., Pauwels, L., Witters, E., Puga, M.I., Paz-ares, J., Goossens, A., Reymond, P., De Jaeger, G., Solano, R., 2011. The Arabidopsis bHLH Transcription Factors

MYC3 and MYC4 Are Targets of JAZ Repressors and Act Additively with MYC2 in the Activation of Jasmonate Responses. Plant Cell 23(February), 701–715.

Finet, C., Timme, R.E., Delwiche, C.F., Marlétaz, F., 2010. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. Curr. Biol. 20(24), 2217–2222.

Fitch, W.M., 1970. Distinguishing Homologous from Analogous Proteins. Syst. Zool. 19(2), 99–113.

Floyd, S.K., Bowman, J.L., 2007. The Ancestral Developmental Tool Kit of Land Plants. Int. J. Plant Sci. 168(1), 1–35.

Fong, J.H., Geer, L.Y., Panchenko, A.R., Bryant, S.H., 2007. Modeling the Evolution of Protein Domain Architectures Using Maximum Parsimony. J. Mol. Biol. 366(1), 307–315.

Franceschi, V.R., Ding, B., Lucas, W.J., 1994. Mechanism of plasmodesmata formation in characean algae in relation to evolution of intercellular communication in higher plants. Planta 192(3), 347–358.

Gai, Z., Wang, Q., Yang, C., Wang, L., Deng, W., Wu, G., 2016. Structural mechanism for the arginine sensing and regulation of CASTOR1 in the mTORC1 signaling pathway. Cell Discov. 2, 1–13.

Garrett, R.H., Grisham, C.M., 1997. Biochemistry, Fifth. ed.

Garvie, C.W., Wolberger, C., 2001. Recognition of specific DNA sequences. Mol. Cell 8(5), 937–946.

George, R.A., Heringa, J., 2003. An analysis of protein domain linkers: their classification and role in protein folding. Protein Eng. 15(11), 871–879.

Glime, J.M., 2015. Chapter 7-1 Water Relations: Conducting Structures. In: Bryophyte Ecology. pp. 1-24.

Gruhn, N., Halawa, M., Snel, B., Seidl, M.F., Heyl, A., 2014. A Subfamily of Putative Cytokinin Receptors Is Revealed by an Analysis of the Evolution of the Two-Component Signaling System of Plants. Plant Physiol. 165, 227–237.

Grunewald, W., Vanholme, B., Pauwels, L., Plovie, E., Inzé, D., Gheysen, G., Goossens, A., 2009. Expression of the Arabidopsis jasmonate signalling repressor JAZ1/TIFY10A is stimulated by auxin. EMBO Rep. 10(8), 923–928.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0 1–37.

Gunnison, D., Alexander, M., 1975. Basis for the susceptibility of several algae to microbial decomposition. Appl. Environ. Microbiol. 29(6), 729–738.

Hädrich, N., Hendriks, J.H.M., Kötting, O., Arrivault, S., Feil, R., Zeeman, S.C., Gibon, Y., Schulze, W.X., Stitt, M., Lunn, J.E., 2012. Mutagenesis of cysteine 81 prevents dimerization of the APS1 subunit of ADP-glucose pyrophosphorylase and alters diurnal starch turnover in Arabidopsis thaliana leaves. Plant J. 70(2), 231–242.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41(41), 95–98.

Hamann, T., Bennett, M., Mansfield, J., Somerville, C., 2009. Identification of cell-wall stress as a hexose-dependent and osmosensitive regulator of plant responses. Plant J. 57, 1015–1026.

Hammonds, A.S., Bristow, C.A., Fisher, W.W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E., Celniker, S.E., 2013. Spatial expression of transcription factors in Drosophila embryonic organ development. Genome Biol. 14(12), R140.

Hanzawa, Y., Takahashi, T., Michael, A.J., Burti, D., Pineiro, D.L.M., Coupland, G., Komeda, Y., 2000. ACAULIS5, an Arabidopsis gene required for stem elongation, encodes a spermine synthase. EMBO J. 19(16), 4248–4256.

Hao, D., Yamasaki, K., Sarai, A., Ohme-Takagi, M., 2002. Determinants in the sequence specific binding of two plant transcription factors, CBF1 and NtERF2, to the DRE and GCC motifs. Biochemistry 41(12), 4202–4208.

Hedderson, T. a., Chapman, R.L., Rootes, W.L., 1996. Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. Plant Syst. Evol. 200(3–4), 213–224.

References

Hegyi, H., Gerstein, M., 2001. Annotation transfer for genomics: Measuring functional divergence in multi-domain proteins. Genome Res. 11(10), 1632–1640.

Herron, M.D., Borin, J.M., Boswell, J.C., Walker, J., Knox, C.A., Boyd, M., Rosenzweig, F., Ratcliff, W.C., 2019. De novo origin of multicellularity in response to predation. Sci. Rep. 9(1), 2328.

Hickey, L.J., Doyle, J.A., 1977. Early cretaceous fossil evidence for angiosperm evolution. Bot. Rev. 43, 3–104.

Hickman, R., van Verk, M.C., Van Dijken, A.J.H., Pereira Mendes, M., Vroegop-Vos, I.A., Caarls, L., Steenbergen, M., Van Der Nagel, I., Wesselink, G.J., Jironkin, A., Talbot, A., Rhodes, J., de Vries, M., Schuurink, R.C., Denby, K., Pieterse, C.M.J., Van Wees, S.C.M., 2017. Architecture and Dynamics of the Jasmonic Acid Gene Regulatory Network. Plant Cell 29(9), 2086–2105.

Higo, A., Kawashima, T., Borg, M., Zhao, M., López-Vidriero, I., Sakayama, H., Montgomery, S.A., Sekimoto, H., Hackenberg, D., Shimamura, M., Nishiyama, T., Sakakibara, K., Tomita, Y., Togawa, T., Kunimoto, K., Osakabe, A., Suzuki, Y., Yamato, K.T., Ishizaki, K., Nishihama, R., Kohchi, T., Franco-Zorrilla, J.M., Twell, D., Berger, F., Araki, T., 2018. Transcription factor DUO1 generated by neo-functionalization is associated with evolution of sperm differentiation in plants. Nat. Commun. 9(1), 5283.

Ho, Y.-S.J., M.Burden, L., Hurley, J.H., 2002. Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. EMBO J. 19(20), 5288–5299.

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., Moriyama, T., Ikeuchi, M., Watanabe, M., Wada, H., Kobayashi, K., Saito, M., Masuda, T., Sasaki-Sekimoto, Y., Mashiguchi, K., Awai, K., Shimojima, M., Masuda, S., Iwai, M., 2014. Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. Nat. Commun. 5, 3978.

Hu, Y.F., Luscher, B., Admon, A., Mermod, N., Tjian, R., 1990. Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes Dev. 4(10), 1741–1752.

Immanen, J., Nieminen, K., Smolander, O.P., Kojima, M., Alonso Serra, J., Koskinen, P., Zhang, J., Elo, A., Mähönen, A.P., Street, N., Bhalerao, R.P., Paulin, L., Auvinen, P., Sakakibara, H., Helariutta, Y., 2016. Cytokinin and Auxin Display Distinct but Interconnected Distribution and Signaling Profiles to Stimulate Cambial Activity. Curr. Biol. 26(15), 1990–1997.

Jez, J.M., Lee, S.G., Nanao, M.H., Dumas, R., Hagen, G., Strader, L.C., Westfall, C.S., Korasick, D.A., Guilfoyle, T.J., 2014. Molecular basis for AUXIN RESPONSE FACTOR protein interaction and the control of auxin response repression. Proc. Natl. Acad. Sci. 11(14), 5427–5432.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., Depamphilis, C.W., 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473, 97–100.

Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., Gao, G., 2017. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 45, D1040–D1045.

Jordan, S.R., Pabo, C.O., 1988. Structure of the lambda complex at 2.5 Å resolution: Details of the repressor-operator interactions. Science. 242(4880), 893–899.

Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B., Mann, R.S., 2007. Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure. Cell 131(3), 530–543.

Jurcik, A., Bednar, D., Byska, J., Marques, S.M., Furmanova, K., Daniel, L., Kokkonen, P., Brezovsky, J., Strnad, O., Stourac, J., Pavelka, A., Manak, M., Damborsky, J., Kozlikova, B., 2018. CAVER Analyst 2.0: Analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. Bioinformatics 34(20), 3586–3588.

Kamashev, D.E., Balandina, A. V., Karpov, V.L., 2000. Tramtrack protein-DNA interactions: A cross-linking study. J. Biol. Chem. 275(46), 36056–36061.

Kartha, G., Bello, J., Harker, D., 1967. Tertiary structure of ribonuclease. Nature 213(5079), 862.

Katayama, H., Iwamoto, K., Kariya, Y., Asakawa, T., Kan, T., Fukuda, H., Ohashi-Ito, K., 2015. A negative feedback loop controlling bHLH complexes is involved in vascular cell division and differentiation in the root

apical meristem. Curr. Biol. 25(23), 3144–3150.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30(4), 772–780.

Kelley, L.A., Sternberg, M.J.E., 2009. Protein structure prediction on the web: A case study using the phyre server. Nat. Protoc. 4(3), 363.

Kenrick, P., 2008. Alternation of generations in land plants: new phylogenetic and palaeobotanical evidence. Biol. Rev. 69, 293–330.

Kenrick, P., Crane, P.R., 1997. The origin and early evolution of plants on land. Nature 389(4), 33–39.

Kim, H., Helmbrecht, E.E., Stalans, M.B., Schmitt, C., Wen, C.-K., Wang, W., Binder, B.M., Physiology, S.P., May, N., Patel, N., 2011. Ethylene receptor ETHYLENE RECEPTOR1 domain requirements for ethylene responses in Arabidopsis seedlings. Plant Physiol. 156(1), 417–429.

Kong, Q., Pattanaik, S., Feller, A., Werkman, J.R., Chai, C., Wang, Y., Grotewold, E., Yuan, L., 2012. Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. Proc. Natl. Acad. Sci. 109(30), E2091–E2097.

Kurakawa, T., Ueda, N., Maekawa, M., Kobayashi, K., Kojima, M., Nagato, Y., Sakakibara, H., Kyozuka, J., 2007. Direct control of shoot meristem activity by a cytokinin-activating enzyme. Nature 445(7128), 652–655.

Kuroha, T., Tokunaga, H., Kojima, M., Ueda, N., Ishida, T., Nagawa, S., Fukuda, H., Sugimoto, K., Sakakibara, H., 2009. Functional Analyses of LONELY GUY Cytokinin-Activating Enzymes Reveal the Importance of the Direct Activation Pathway in Arabidopsis. Plant Cell 21(10), 3152–3169.

Lavigne, P., Kondejewski, L.H., Houston, M.E., Sönnichsen, F.D., Lix, B., Sykes, B.D., Hodges, R.S., Kay, C.M., 1995. Preferential heterodimeric parallel coiled-coil formation by synthetic Max and c-Myc leucine zippers: A description of putative electrostatic interactions responsible for the specificity of heterodimerization. J. Mol. Biol. 254(3), 505–520.

Lemieux, C., Otis, C., Turmel, M., 2007. A clade uniting the green algae Mesostigma viride and Chlorokybus atmophyticus represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. BMC Biol. 5(1), 2.

Lentona, T.M., Dahl, T.W., Daines, S.J., Mills, B.J.W., Ozaki, K., Saltzman, M.R., Porada, P., 2016. Earliest land plants created modern levels of atmospheric oxygen. Proc. Natl. Acad. Sci. 113(35), 9704–9709.

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44(1), W242–W245.

Lewis, L.A., McCourt, R.M., 2004. Green algae and the origin of land plants. Am. J. Bot. 91(10), 1535–1556.

Ligrone, A.R., Duckett, J.G., Renzaglia, K.S., Philosophical, S., Biological, T., Ligrone, R., Duckett, J.G., Renzaglia, K.S., 2000. Conducting Tissues and Phyletic Relationships of Bryophytes 355(1398), 795–813.

Ligrone, R., Carafa, A., Duckett, J.G., Renzaglia, K.S., Ruel, K., 2008. Immunocytochemical detection of lignin-related epitopes in cell walls in bryophytes and the charalean alga Nitella. Plant Syst. Evol. 270, 257–272.

Ligrone, R., Duckett, J.G., Renzaglia, K.S., 2012. Major transitions in the evolution of early land plants: A bryological perspective. Ann. Bot. 109(5), 851–871.

Liu, G., Mattick, J.S., Taft, R.J., 2013. A meta-analysis of the genomic and transcriptomic composition of complex life. Cell Cycle 12(13), 2061–2072.

Liu, Y., Johnson, M.G., Cox, C.J., Medina, R., Devos, N., Vanderpoorten, A., Hedenäs, L., Bell, N.E., Shevock, J.R., Aguero, B., Quandt, D., Wickett, N.J., Shaw, J.A., Goffinet, B., 2019. Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. Nat. Commun. 10(1), 1485.

Liu, Y., Zhao, T.J., Liu, J.M., Liu, W.Q., Liu, Q., Yan, Y. Bin, Zhou, H.M., 2006. The conserved Ala37 in the ERF/AP2 domain is essential for binding with the DRE element and the GCC box. FEBS Lett. 580(5), 1303–1308.

Lucas, W.J., Groover, A., Lichtenberger, R., Furuta, K., Yadav, S.R., Helariutta, Y., He, X.Q., Fukuda, H., Kang,

J., Brady, S.M., Patrick, J.W., Sperry, J., Yoshida, A., López-Millán, A.F., Grusak, M.A., Kachroo, P., 2013. The Plant Vascular System: Evolution, Development and Functions. J. Integr. Plant Biol. 55(4), 294–388.

Lynch, M., Conery, J.S., 2000. The Evolutionary Fate and Consequences of Duplicate Genes 290(November), 1151–1156.

Magallón, S., Hilu, K.W., Quandt, D., 2013. Land plant evolutionary timeline: Gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. Am. J. Bot. 100, 556–573.

Mähönen, A.P., Bonke, M., Kauppinen, L., Riikonen, M., Benfey, P.N., Helariutta, Y., 2000. A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root. Genes Dev. 14(23), 2938–2943.

Mähönen, A.P., Higuchi, M., Törmäkangas, K., Miyawaki, K., Pischke, M.S., Sussman, M.R., Helariutta, Y., Kakimoto, T., 2006. Cytokinins Regulate a Bidirectional Phosphorelay Network in Arabidopsis. Curr. Biol. 16(11), 1116–1122.

Maizel, A., Busch, M.A., Tanahashi, T., Perkovic, J., Hasebe, M., Weigel, D., 2005. The floral regulator LEAFY evolves by substitutions in the DNA binding domain 308(5719), 260–263.

Martinez, S.E., Wu, A.Y., Glavas, N.A., Tang, X.-B., Turley, S., Hol, W.G.J., Beavo, J.A., 2002. The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. Proc. Natl. Acad. Sci. 99, 13260–13265.

Mas-Droux, C., Curien, G., Robert-Genthon, M., Laurencin, M., Ferrer, J.-L., Dumas, R., 2006. A Novel Organization of ACT Domains in Allosteric Enzymes Revealed by the Crystal Structure of Arabidopsis Aspartate Kinase. Plant Cell Online 18(7), 1681–1692.

Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Wafula, E.K., Der, J.P., DePamphilis, C., Roure, B., Philippe, H., Ruhfel, B.R., Miles, N.W., Graham, S.W., Mathews, S., Surek, B., Melkonian, M., Soltis, D., Soltis, P., Rothfels, C.J., Pokorny, L., Shaw, J.A., DeGironimo, L., Stevenson, D.W., Villarreal, J.C., Chen, T., Kutchan, T.M., Rolf, M., Baucom, R.S., Deyholos, M.K., Samudrala, R., Tian, Z., Wu, X., Sun, X., Zhang, Y., Wang, J., Leebens-Mack, J., Wong, G.K.-S., 2014. Data access for the 1,000 Plants (1KP) project. Gigascience 3(1), 17.

Matsumoto-Kitano, M., Kusumoto, T., Tarkowski, P., Kinoshita-Tsujimura, K., Vaclavikova, K., Miyawaki, K., Kakimoto, T., 2008. Cytokinins are central regulators of cambial activity. Proc. Natl. Acad. Sci. 105(50), 20027–20031.

Mattox, K.R., Stewart, K.D., 1984. Classification of the green algae: a concept based on comparative cytology. In: Systematics of the Green Algae. pp. 29–72.

McCourt, R.M., Delwiche, C.F., Karol, K.G., 2004. Charophyte algae and land plant origins. Trends Ecol. Evol. 19(12), 661–666.

McMichael, B.L., Burke, J.J., Berlin, J.D., Hatfield, J.L., Quisenberry, J.E., 1985. Root vascular bundle arrangements among cotton strains and cultivars. Environ. Exp. Bot. 25(1), 23–30.

McMichael, B.L., Quisenberry, J.E., Upchurch, D.R., 1987. Lateral root development in exotic cottons. Environ. Exp. Bot. 27(4), 499–502.

Mellor, N., Vaughan-Hirsch, J., Kümpers, B.M.C., Help-Rinta-Rahko, H., Miyashima, S., Mähönen, A.P., Campilho, A., King, J.R., Bishopp, A., 2019. A core mechanism for specifying root vascular patterning can replicate the anatomical variation seen in diverse plant species. Development 146(6), dev172411.

Memelink, J., 2009. Regulation of gene expression by jasmonate hormones. Phytochemistry 70(13–14), 1560–1570.

Mikhailyuk, T., Holzinger, A., Massalski, A., Karsten, U., 2014. Morphology and ultrastructure of Interfilum and Klebsormidium (Klebsormidiales , Streptophyta) with special reference to cell division and thallus formation 49(May 2015), 395–412.

Mishler, B.D., Churchill, S.P., 1985. Transition to a land flora: phylogenetic relationships of the green algae and Bryophytes. Cladistics 1(4), 305–328.

Miyashima, S., Roszak, P., Sevilem, I., Toyokura, K., Blob, B., Heo, J. ok, Mellor, N., Help-Rinta-Rahko, H., Otero, S., Smet, W., Boekschoten, M., Hooiveld, G., Hashimoto, K., Smetana, O., Siligato, R., Wallner, E.S., Mähönen, A.P., Kondo, Y., Melnyk, C.W., Greb, T., Nakajima, K., Sozzani, R., Bishopp, A., De Rybel, B., Helariutta, Y., 2019. Mobile PEAR transcription factors integrate positional cues to prime cambial growth. Nature 565(7740), 490–494.

Moczydlowska, M., Landing, E., Zang, W., Palacios, T., 2011. Proterozoic phytoplankton and timing of Chlorophyte algae origins. Palaeontology 54, 721–733.

Moll, J.R., Bonovich, M., Acharya, A., Mir, A.A., Myakishev, M., Vinson, C., 2002. Classification of Human B-ZIP Proteins Based on Dimerization Properties. Mol. Cell. Biol. 22(18), 6321–6335.

Moll, J.R., Olives, M., Vinson, C., 2000. Attractive interhelical electrostatic interactions in the proline- and acidic-rich region (PAR) leucine zipper subfamily preclude heterodimerization with other basic leucine zipper subfamilies. J. Biol. Chem. 275(44), 34826–34832.

Moore, M.J., Bell, C.D., Soltis, P.S., Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl. Acad. Sci. U. S. A. 104(49), 19363–8.

Moraitis, A.N., 1999. Transition from Monomeric to Homodimeric DNA Binding by Nuclear Receptors: Identification of RevErbA Determinants Required for ROR Homodimer Complex Formation. Mol. Endocrinol. 13, 431–439.

Mortier, V., Wasson, A., Jaworek, P., De Keyser, A., Decroos, M., Holsters, M., Tarkowski, P., Mathesius, U., Goormachtig, S., 2014. Role of LONELY GUY genes in indeterminate nodulation on *Medicago truncatula*. New Phytol. 202(2), 582–593.

Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C. V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., Weintraub, H., Baltimore, D., 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell 58(3), 537–544.

Mutte, S., Kato, H., Rothfels, C., Melkonian, M., Wong, G.K.-S., Weijers, D., 2018. Origin and evolution of the nuclear auxin response system. Elife 7, e33399.

Nagpal, P., Ellis, C.M., Weber, H., Ploense, S.E., Barkawi, L.S., Guilfoyle, T.J., Hagen, G., Alonso, J.M., Cohen, J.D., Farmer, E.E., Ecker, J.R., Reed, J.W., 2005. Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. Development 132(18), 4107–4118.

Naseem, M., Sarukhanyan, E., Dandekar, T., 2015. LONELY-GUY Knocks Every Door: Crosskingdom Microbial Pathogenesis. Trends Plant Sci. 20(12), 781–783.

Newman, J.R.S., Keating, A.E., 2003. Comprehensive identification of human bZIP interactions with coiled-coil arrays. Science. 300(5628), 2097–2101.

Newton, A.E., Cox, C.J., Duckett, J.G., Wheeler, J.A., Goffinet, B., Hedderson, T.A.J., Mishler, B.D., 2000. Evolution of the major moss lineages: phylogenetica analyses based on multiple gene sequences and morphology. Bryologist 103(2), 187–211.

Nickrent, D.L., Parkinson, C.L., Palmer, J.D., Duff, R.J., 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol. Biol. Evol. 17(12), 1885–1895.

Northcote, D.H., Goulding, K.J., Horne, R.W., 1958. The Chemical Composition and Structure of the Cell Wall of Chlorella pyrenoidosa. Biochem. J. 70(3), 391–7.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., Wolfe, S.A., 2008. Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. Cell 133(7), 1277–1289.

O'Shea, E.K., Lumb, K.J., Kim, P.S., 1993. Peptide 'Velcro': Design of a heterodimeric coiled coil. Curr. Biol. 3(10), 658–667.

O'Shea, E.K., Rutkowski, R., Kim, P.S., 1992. Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. Cell 68(4), 699–708.

Ohashi-Ito, K., Bergmann, D.C., 2007. Regulation of the Arabidopsis root vascular initial population by LONESOME HIGHWAY. Development 134(16), 2959–2968.

References

Ohashi-Ito, K., Matsukawa, M., Fukuda, H., 2013. An atypical bHLH transcription factor regulates early xylem development downstream of auxin. Plant Cell Physiol. 54(3), 398–405.

Ohashi-Ito, K., Saegusa, M., Iwamoto, K., Oda, Y., Katayama, H., Kojima, M., Sakakibara, H., Fukuda, H., 2014. A bHLH complex activates vascular cell division via cytokinin action in root apical meristem. Curr. Biol. 24(17), 2053–2058.

Ohtani, M., Akiyoshi, N., Takenaka, Y., Sano, R., Demura, T., 2017. Evolution of plant conducting cells: perspectives from key regulators of vascular cell differentiation 68(1), 17–26.

Oosterhuis, D.M., Wullschleger, S.D., 1987. Water Flow Through Cotton Roots in Relation to Xylem Anatomy. J. Exp. Bot. 38(11), 1866–1874.

Pauwels, L., Barbero, G.F., Geerinck, J., Tilleman, S., Grunewald, W., Pérez, A.C., Chico, J.M., Bossche, R. Vanden, Sewell, J., Gil, E., García-Casado, G., Witters, E., Inzé, D., Long, J.A., De Jaeger, G., Solano, R., Goossens, A., 2010. NINJA connects the co-repressor TOPLESS to jasmonate signalling. Nature 464, 788–91.

Pérez Cuéllar, A., Nagels Durand, A., Pauwels, L., Persiau, G., Goossens, A., De Clercq, R., Vanden Bossche, R., Van Wees, S.C.M., De Jaeger, G., Pieterse, C.M.J., Gevaert, K., Cuéllar Pérez, A., 2014. The Non-JAZ TIFY Protein TIFY8 from Arabidopsis thaliana Is a Transcriptional Repressor. PLoS One 9(1), e84891.

Pils, B., Heyl, A., 2009. Unraveling the Evolution of Cytokinin Signaling. Plant Physiol. 151(2), 782–791.

Piombino, A., 2016. The Heavy Links between Geological Events and Vascular Plants Evolution: A Brief Outline. Int. J. Evol. Biol. 2016, 9264357.

Pires, N., Dolan, L., 2010. Origin and Diversification of Basic-Helix-Loop-Helix Proteins in Plants. Mol. Biol. Evol 27, 862–874.

Poehlein, A., Seedorf, H., 2016. Draft Genome Sequences of Methanobrevibacter curvatus DSM11111, Methanobrevibacter cuticularis DSM11139, Methanobrevibacter filiformis DSM11501, and Methanobrevibacter oralis DSM7256: TABLE 1. Genome Announc. 4(3), e00617-16.

Pongratz, I., Antonsson, C., Whitelaw, M.L., Poellinger, L., 1998. Role of the PAS domain in regulation of dimerization and DNA binding specificity of the dioxin receptor. Mol. Cell. Biol. 18(7), 4079–88.

Ponting, C.P., Russell, R.R., 2002. The Natural History of Protein Domains. Annu. Rev. Biophys. Biomol. Struct. 31(1), 45–71.

Proost, S., Bel, M. Van, Vaneechoutte, D., Van De Peer, Y., Inzé, D., Mueller-Roeber, B., Vandepoele, K., 2015. PLAZA 3.0: An access point for plant comparative genomics. Nucleic Acids Res. 43(D1), D974–D981.

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

Puttick, M.N., Morris, J.L., Williams, T.A., Schneider, H., Pisani, D., Donoghue, P.C.J., 2018. The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte. Curr. Biol. 28, 733–745.

Qi, T., Song, S., Ren, Q., Wu, D., Huang, H., Chen, Y., Fan, M., Peng, W., Ren, C., Xie, D., 2011. The Jasmonate-ZIM-Domain Proteins Interact with the WD-Repeat / bHLH / MYB Complexes to Regulate Jasmonate-Mediated Anthocyanin Accumulation and Trichome Initiation in Arabidopsis thaliana 23(May), 1795–1814.

Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., Estabrook, G.F., Hendry, T.A., Taylor, D.W., Testa, C.M., Ambros, M., Crandall-Stotler, B., Duff, R.J., Stech, M., Frey, W., Quandt, D., Davis, C.C., 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. 103(42), 15511–15516.

Ragni, L., Greb, T., 2018. Secondary growth as a determinant of plant shape and form. Semin. Cell Dev. Biol. 79, 58–67.

Raven, J.A., 1977. The Evolution of Vascular Land Plants in Relation to Supracellular Transport Processes. Adv. Bot. Res. 5, 153–219.

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

Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A., Walhout, A.J.M., 2005. A compendium of Caenorhabditis elegans regulatory transcription factors: A resource for mapping transcription regulatory networks. Genome Biol. 6(13), R110.

Richardson, D.H.S., 1981. The Biology of Mosses. Halsted Press.

Riechmann, J., 1996. DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Res. 24(16), 3134–3141.

Rishi, V., Gal, J., Krylov, D., Fridriksson, J., Boysen, M.S., Mandrup, S., Vinson, C., 2004. SREBP-1 dimerization specificity maps to both the helix-loop-helix and leucine zipper domains: Use of a dominant negative. J. Biol. Chem. 279(12), 11863–11874.

Rockwell, N.C., Lagarias, J.C., 2006. The Structure of Phytochrome: A Picture Is Worth a Thousand Spectra. Plant Cell Online 18(1), 4–14.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., Mann, R.S., 2010. Origins of Specificity in Protein-DNA Recognition. Annu. Rev. Biochem. 79(1), 233–269.

Rothwell, G.W., Erwin, D.M., 1987. Origin of Seed Plants : An Aneurophyte / Seed-Fern Link Elaborated. Am. J. Bot. 74(6), 970–973.

Rothwell, G.W., Scheckler, S.E., Gillespie, W.H., 1989. Elkinsia gen. nov., a Late Devonian Gymnosperm with Cupulate Ovules. Bot. Gaz. 150(2), 170–189.

Samanovic, M.I., Tu, S., Novák, O., Iyer, L.M., McAllister, F.E., Aravind, L., Gygi, S.P., Hubbard, S.R., Strnad, M., Darwin, K.H., 2015. Proteasomal Control of Cytokinin Synthesis Protects Mycobacterium tuberculosis against Nitric Oxide. Mol. Cell 57(6), 984–994.

Saxton, R.A., Chantranupong, L., Knockenhauer, K.E., Schwartz, T.U., Sabatini, D.M., 2016. Mechanism of arginine sensing by CASTOR1 upstream of mTORC1. Nature 536(7615), 229.

Scheirer, D.C., 1975. Anatomical Studies in the Polytrichaceae II. Histochemical Observations on Thickened Lateral Walls of Hydroids of Dendroligotrichum. Bryologist 120(9), 113–123.

Scheres, B., Wolkenfelt, H., Willemsen, V., Terlouw, M., Lawson, E., Dean, C., Weisbeek, P., 1994. Embryonic origin of the Arabidopsis primary root and root meristem initials. Development 2487, 2475–2487.

Schlüter, K., Jockusch, B.M., Rothkegel, M., 1997. Profilins as regulators of actin dynamics. Biochim. Biophys. Acta - Mol. Cell Res. 1359(2), 97–109.

Schuetz, M., Smith, R., Ellis, B., 2013. Xylem tissue specification, patterning, and differentiation mechanisms. J. Exp. Bot. 64(1), 11–31.

Schuster, C., Gaillochet, C., Lohmann, J.U., 2015. *Arabidopsis HECATE* genes function in phytohormone control during gynoecium development. Development 142(19), 3343–3350.

Sehr, E.M., Agusti, J., Lehner, R., Farmer, E.E., Schwarz, M., Greb, T., 2010. Analysis of secondary growth in the Arabidopsis shoot reveals a positive role of jasmonate signalling in cambium formation. Plant J. 63(5), 811–822.

Serbet, R., Rothwell, G.W., 1992. Characterizing the Most Primitive Seed Ferns. I. A Reconstruction of Elkinsia polymorpha. Int. J. Plant Sci. 153(4), 602–621.

Shirakata, M., Paterson, B.M., 2018. The E12 inhibitory domain prevents homodimer formation and facilitates selective heterodimerization with the MyoD family of gene regulatory factors. EMBO J. 14(8), 1766–1772.

Simon, A., Glöckner, G., Felder, M., Melkonian, M., Becker, B., 2006. EST analysis of the scaly green flagellate Mesostigma viride (Streptophyta): Implications for the evolution of green plants (Viridiplantae). BMC Plant Biol. 6(1), 2.

Smet, W., Sevilem, I., Balaguer, M.A. de L., Wybouw, B., Mor, E., Miyashima, S., Blob, B., Roszak, P., Jacobs, T.B., Boekschoten, M., Hooiveld, G., Sozzani, R., Helariutta, Y., Rybel, B. De, 2019. DOF2.1 Controls Cytokinin-Dependent Vascular Cell Proliferation Downstream of TMO5/LHW. Curr. Biol. 29(3), 520–529.

Soltis, P.S., Soltis, D.E., Chase, M.W., 1999. Angiosperm phylogeny inferred from multiple genes as a tool for

comparative biology. Nature 402(6760), 402–404.

Song, S., Qi, T., Huang, H., Ren, Q., Wu, D., Chang, C., Peng, W., Liu, Y., Peng, J., Xie, D., 2011. The Jasmonate-ZIM Domain Proteins Interact with the R2R3-MYB Transcription Factors MYB21 and MYB24 to Affect Jasmonate-Regulated Stamen Development in Arabidopsis. Plant Cell 23(3), 1000–1013.

Stefanović, S., Rice, D.W., Palmer, J.D., 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? BMC Evol. Biol. 4, 35.

Stirk, W.A., Ördög, V., Novák, O., Rolčík, J., Strnad, M., Bálint, P., van Staden, J., 2013. Auxin and cytokinin relationships in 24 microalgal strains. J. Phycol. 49(3), 459–467.

Suzuki, M., 1994. A framework for the DNA-protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. Structure 2(4), 317–326.

Suzuki, M., Gerstein, M., 1995. Binding geometry of α-helices that recognize DNA. Proteins Struct. Funct. Bioinforma. 23(4), 525–535.

Tan, B.X., Liew, H.P., Chua, J.S., Ghadessy, F.J., Tan, Y.S., Lane, D.P., Coffill, C.R., 2017. Anatomy of Mdm2 and Mdm4 in evolution. J. Mol. Cell Biol. 9(1), 3–15.

Teichmann, S.A., Babu, M.M., 2004. Gene regulatory network growth by duplication 36(5), 492–496.

Thines, B., Katsir, L., Melotto, M., Niu, Y., Mandaokar, A., Liu, G., Nomura, K., He, S.Y., Howe, G.A., Browse, J., 2007. JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. Nature 448(7154), 661.

Thomas, R.J., Schiele, E.M., Scheirer, D.C., 1988. Translocation in Polytrichum commune (Bryophyta). I. Conduction and Allocation of Photoassimilates. Am. J. Bot. 75(2), 275–281.

Tokunaga, H., Kojima, M., Kuroha, T., Ishida, T., Sugimoto, K., Kiba, T., Sakakibara, H., 2012. Arabidopsis lonely guy (LOG) multiple mutants reveal a central role of the LOG-dependent pathway in cytokinin activation. Plant J. 69(2), 355–365.

Toledo-Ortiz, G., Huq, E., Quail, P.H., 2003. The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell 15(8), 1749–70.

Tordai, H., Nagy, A., Farkas, K., Bányai, L., Patthy, L., 2005. Modules, multidomain proteins and organismic complexity. FEBS J. 272(19), 5064–5078.

Torrey, J.G., 1955. On the Determination of Vascular Patterns During Tissue Differentiation in Excised Pea Roots. Am. J. Bot. 42(2), 183–198.

Truernit, E., Bauby, H., Dubreucq, B., Grandjean, O., Runions, J., Barthelemy, J., Palauqui, J.-C., 2008. High-Resolution Whole-Mount Imaging of Three-Dimensional Tissue Organization and Gene Expression Enables the Study of Phloem Development and Structure in Arabidopsis. Plant Cell Online 20(6), 1494–1503.

Unden, G., Nilkens, S., Singenstreu, M., 2013. Bacterial sensor kinases using Fe-S cluster binding PAS or GAF domains for O2sensing. Dalt. Trans. 42(9), 3082–3087.

Vera-Sirera, F., De Rybel, B., Úrbez, C., Kouklas, E., Pesquera, M., Álvarez-Mahecha, J.C., Minguet, E.G., Tuominen, H., Carbonell, J., Borst, J.W., Weijers, D., Blázquez, M.A., 2015. A bHLH-Based Feedback Loop Restricts Vascular Cell Proliferation in Plants. Dev. Cell 35(4), 432–443.

Volkmar, U., Knoop, V., 2010. Introducing Intron Locus cox1i624 for Phylogenetic Analyses in Bryophytes : On the Issue of Takakia as Sister Genus to All Other Extant Mosses. J. Mol. Evol. 70(5), 506–518.

Waters, E.R., 2003. Molecular adaptation and the origin of land plants. Mol. Phylogenet. Evol. 29(3), 456–463.

Wellman, C.H., Osterloff, P.L., Mohiuddin, U., 2003. Fragments of the earliest land plants. Nature 425(6955), 282–285.

Wendrich, J.R., Liao, C.-Y., van den Berg, W.A.M., De Rybel, B., Weijers, D., 2015. Ligation-Independent Cloning for Plant Research. In: Alonso, J.M., Stepanova, A.N. (Eds.), Plant Functional Genomics: Methods and Protocols. Springer New York, New York, NY, pp. 421–431.

Wingender, E., Schoeps, T., Dönitz, J., 2013. TFClass: An expandable hierarchical classification of human transcription factors. Nucleic Acids Res. 41(1), D165–D170.

Winter, C.W., 1932. Vascular System of Young Plants of Medicago sativa. Bot. Gaz. 94(1), 152–167.

Winter, K.U., Weiser, C., Kaufmann, K., Bohne, A., Kirchner, C., Kanno, A., Saedler, H., Theißen, G., 2002. Evolution of class B floral homeotic proteins: Obligate heterodimerization originated from homodimerization. Mol. Biol. Evol. 19(5), 587–596.

Xu, B., Ohtani, M., Yamaguchi, M., Toyooka, K., Wakazaki, M., Sato, M., Kubo, M., Nakano, Y., Sano, R., Hiwatashi, Y., Murata, T., Kurata, T., Yoneda, A., Kato, K., Hasebe, M., Demura, T., 2014. Contribution of NAC Transcription Factors to Plant Adaptation to Land Contribution of NAC Transcription Factors to Plant Adaptation to Land. Sci. reports 343(6178), 1505–1508.

Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Matsuda, T., Tomo, Y., Hayami, N., Terada, T., Shirouzu, M., Tanaka, A., Seki, M., Shinozaki, K., Yokoyama, S., 2005. Solution Structure of an Arabidopsis WRKY DNA Binding Domain. Plant Cell Online 17(3), 944–956.

Yang, S., Wang, S., Liu, X., Yu, Y., Yue, L., Wang, X., Hao, D., 2009. Four divergent Arabidopsis ethylene-responsive element-binding factor domains bind to a target DNA motif with a universal CG step core recognition and different flanking bases preference. FEBS J. 276(23), 7177–7186.

Yang, Y., Labesse, G., Carrère-Kremer, S., Esteves, K., Kremer, L., Cohen-Gonsaud, M., Blanc-Potard, A.B., 2012. The C-Terminal Domain of the Virulence Factor mgtc Is a Divergent ACT Domain. J. Bacteriol. 194(22), 6255–6263.

Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., Bhattacharya, D., 2004. A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. Mol. Biol. Evol. 21(5), 809–818.

Zarrinpar, A., Park, S.H., Lim, W.A., 2003. Optimization of specificity in a cellular protein interaction network by negative selection. Nature 426(6967), 676–680.

Zhang, C., Feng, R., Ma, R., Shen, Z., Cai, Z., Song, Z., Peng, B., Yu, M., 2018. Genome-wide analysis of basic helix-loop-helix superfamily members in peach. PLoS One 13(4), 1–20.

Zhang, Y., Werling, U., Edelmann, W., 2015. Seamless Ligation Cloning Extract (SLiCE) Cloning Method Yongwei. In: Methods Mol Biol. pp. 235–244.

Zhou, J., Wang, X., Lee, J.-Y., Lee, J.-Y., 2013. Cell-to-Cell Movement of Two Interacting AT-Hook Factors in Arabidopsis Root Vascular Tissue Patterning . Plant Cell 25(1), 187–201.

About the author

Acknowledgements

Curriculum vitae

Publications

Education statement

Acknowledgements

Some of you might know that rock climbing, and more specifically bouldering, is one of my hobbies. During the four years of my thesis I have come to realise that a PhD project is similar to climbing an (extremely steep and high) wall. This is one of the reasons I included a boulder-like structure on the cover of my thesis. A climbing route over a wall or rock is often referred to as a 'problem', and climbing the rock to the top thus equals solving the problem.

Four years ago, I started a giant bouldering problem that was my PhD research. Before starting a boulder problem, it is important to 'read' the route: what steps do you need to take to solve the problem? Once you start the project, it can be hard to keep the overview of the steps you need to take and sometimes the planned steps simply do not work. You might make a mistake half way to the top, fall down and have to start all over again. This is where a coach can come in handy. In the case of my PhD project I was lucky enough to have two brilliant coaches, who already came up with some steps to solve the PhD boulder problem before I started it. They have always been there to coach me during the project. I would like to thank Prof. Dr. Dolf Weijers and Prof. Dr. Bert De Rybel for their coaching and supervision during my project.

Dolf, already in the second year of my Bachelor studies, you inspired me and convinced me how much fun biochemistry is. Therefore, I was thrilled to do my Master thesis at the biochemistry lab. Later, you were always there to give me advice when I was stuck during career choices. Thank you for all your help and guidance during this time and suggesting the PhD position with Bert. I cannot thank you enough for your time, your scientific input and guidance during my PhD. You have created an amazing team and an inspiring workplace. Working in your lab has been a great pleasure and privilege.

Bert, I enjoyed working with you a lot. I am grateful for the opportunity you gave me to join your team and work on the evolution of vascular tissues. Even though you were just starting your group at the beginning of my thesis, your scientific insights, professionality, involvement and enthusiasm already made you an accomplished group leader. Somehow you managed to be an amazing and involved daily supervisor even from 200 km away. I enjoyed our regular Skype (and face-to-face) discussions very much. Thank you for all your input and coaching. I might not be ready to 'free solo' just yet, but you brought me one step closer.

I also want to thank Sacco, who was the department head during the first part of my PhD. Thank you for welcoming me in the Biochemistry lab. I had a great time planning the PhD trip to Barcelona together. I also want to thank you for being the superman

that comes to the rescue when the growth rooms are failing. On top of that you allowed me to cross off an item on my bucket list: flying a plane. Thank you for everything!

A great help during the climb of a boulder is a climbing partner. He or she can try to help solve the problem by working on certain difficult steps together or give advice on a specific move. During my PhD, this partner has been Dr. Kuan-Ju Lu. I would like to express my gratitude to Kuan-Ju, my paranymph, who contributed greatly to my PhD project. I have enjoyed working with you a lot. It has been great to have you as an office mate for all these years. First in the ladies office and, later on, in the best biochemistry office.

Officially, climbing is an induvial sport, but it never feels like it is. Even though everyone works on their own problem, climbers in the gym are there to encourage each other and give advice when needed. While I suppose this might not be the case for all laboratories, I can say with certainty that the same team spirit can be felt in the groups of Dolf and Bert. I want to thank all my colleagues in Wageningen and Ghent for their assistance, advice and encouragement. In the last four years we have become good friends and I enjoyed the time we spent together in- and outside the lab. I have had great fun with all activities such as the lab weekends, board game evenings, movie nights, the PhD trip to Barcelona, cocktail nights and many more. I could not have asked for better colleagues. I will miss you all!

However, I might sell some people short with this acknowledgement. Therefore, I particularly want to mention a few people. First of all, my paranymph Sumanth. Thank you for all your help with the bioinformatic analyses. You have made my PhD climb a lot lighter! I have come to know you as a great scientist and I am sure you have a bright scientific future ahead of you. Margo, my fellow climber, you have become a good friend in the last years. You were there to talk and complain to, about work-related and not so work-related topics. Thank you for all the conversations and the accompanying beers. I wish you all the best in California! Tanya and Maritza, my former office buddies in the 'ladies office': it has been a pleasure to have you as my colleagues. Thanks for the very warm welcome in the lab and your support during the last four years. Next, what is a climber without her gear? Another special thanks goes out to Willy. Without you, the 'gear' in the lab would certainly fall apart. You are the man that saves the day when lab equipment fails or when we are suddenly out of stock. Thank you for all the help! Cathy, merci beaucoup for all your help in the lab and during enzymology practicals. Mark and Sjef, thank you for your assistance with IP-MS. A special thanks goes to Laura, who is always there to help with bureaucracy complications and never gets tired from hearing the same stupid question over and over. Laura, thank you for everything!

It is customary for climbers to train in one, maybe two, indoor climbing halls. Mostly because of

the close proximity. However, exploring other gyms and outside areas can be a very useful training to experience different styles or get used to the feeling of actual rocks. During my project I explored such different styles in Ghent, where I worked with Prof. Dr. Klaas Vandepoele on bioinformatics. I would like to offer my special thanks to Klaas. Thank you for all your help, time and patience.

During my PhD I had the pleasure of working with three bachelor students. Paul Abrahams, Mitchel Keereweer and Eline de Ridder: thank you for your help on my project. I learned a lot during the time I was your supervisor and I hope that I, in return, managed to pass on some of my knowledge to you. I wish all of you the best of luck in your future endeavours.

I am also grateful to my friends who have supported me along the way. I specifically want to mention Anne, Ko, Laurens, Marnix, Vivian and Koen.

Furthermore, I would like to show my appreciation to my parents, my sister Janine and brother-in-law Koen for always supporting me. Pap en mam, ik vind het moeilijk om mijn dankbaarheid te verwoorden. Jullie hebben me altijd gestimuleerd om het beste uit mezelf te halen en hebben school en opleiding altijd bovenaan de prioriteitenlijst gehad. Ik wil jullie bedanken voor al jullie hulp, steun en betrokkenheid bij mijn hele opleiding. Zonder dit alles zou ik nu zeker niet op het punt staan om mijn PhD te verdedigen. Ik ben ook ontzettend dankbaar voor alle steun die ik ontvangen heb van de rest van mijn familie. Bedankt!

Last but not least, I want to thank my (actual climbing) partner Ruud. In climbing and all other sports, training is just as important as recovery and relaxation. Thank you for distracting me from my work from time to time and helping me to relax. I also want to thank you for your patience and encouragement during the writing process. Your believe in me has made the whole process a lot easier. I am incredibly happy to have you in my life and I cannot wait to start the next chapter of our lives together.

Curriculum vitae

Nicole van 't Wout Hofland was born on the 2nd of March 1990 in Heerlen, the Netherlands. In 2008, she graduated from her secondary school St. Janscollege in Hoensbroek. Shortly after she moved to Wageningen to study Biotechnology at Wageningen University (Wageningen, the Netherlands). After completing her Bachelor programme in 2011, she continued at Wageningen University to obtain her Master degree in biotechnology. During this time she specialized herself in medical biotechnology as well as in cellular- and molecular biotechnology. She performed her master theses in the laboratory of Biochemistry (Wageningen University), where she worked on cell-type-specific transcriptome analysis in the Arabidopsis plant embryo, and in laboratory of Nematology (Wageningen University), where she attempted to optimize human cytokine IL-10 production in tobacco plants. In 2013, she spent 7 months at Généthon (Paris, France), a company that develops gene therapy for rare diseases, working on the baculovirus system for rAAV vector production. After obtaining her MSc degree in 2014 she returned to France to work as a researcher at Centre Energie Atomique (CEA), focussing on designing Interferon(IFN)-β variants to treat multiple sclerosis (MS) in humans. In 2015, she returned to the Netherlands to join the Biochemistry group of Prof. Dr. Dolf Weijers at Wageningen University. She started a PhD project under the supervision of Prof. Dr. Dolf Weijers and Prof. Dr. Bert De Rybel to perform research on the origin and evolution of two known vascular tissue regulators and their role in the evolution of vascular tissues. On the 8th of October she will present and defend the results of this thesis. Since June 2019 she works in science communication and science journalism at Wageningen University & Research.

Publications

Nicole van 't Wout Hofland, Kuan-Ju Lu, Eliana Mor, Sumanth Mutte, Paul Abrahams, Hirotaka Kato, Klaas Vandepoele, Dolf Weijers and Bert De Rybel. "Evolution of vascular plants through redeployment of ancient developmental regulators." PNAS (Under review).

Joakim Palovaara, Shunsuke Saiga, Jos R. Wendrich, **Nicole van 't Wout Hofland**, J. Paul van Schayck, Friederike Hater, Sumanth Mutte, Jouke Sjollema, Mark Boekschoten, Guido J. Hooiveld and Dolf Weijers. "Transcriptome dynamics revealed by a gene expression atlas of the early Arabidopsis embryo." Nature plants 3.11 (2017): 894.

Education Statement of the Graduate School

Experimental Plant Sciences

The Graduate School
EXPERIMENTAL PLANT SCIENCES

Issued to:	**Nicole van 't Wout Hofland**
Date:	**08 October 2019**
Group:	**Biochemistry**
University:	**Wageningen University & Research**

1) Start-Up Phase	*date*	*cp*
► **First presentation of your project**		
Project proposal: The contribution of LHW/TMO5 recrutement to complexity increase in vascular tissue of plants during evolution.	12 dec 2015	1,5
► **Writing or rewriting a project proposal**		
► **Writing a review or book chapter**		
► **MSc courses**		
Subtotal Start-Up Phase		1,5

2) Scientific Exposure	*date*	*cp*
► **EPS PhD student days**		
EPS Get2Gether 2016	28 - 29 Jan 2016	0,6
EPS Get2Gether 2018	15 - 16 Feb 2018	0,6
► **EPS theme symposia**		
EPS theme 1 'Developmental Biology of Plants'	21 Jan 2016	0,3
EPS theme 1 'Developmental Biology of Plants'	28 Feb 2017	0,3
EPS theme 1 'Developmental Biology of Plants'	30 Jan 2018	0,3
► **Lunteren Days and other national platforms**		
ALW meeting 'Experimental Plant Sciences', Lunteren	11 - 12 Apr 2016	0,6
ALW meeting 'Experimental Plant Sciences', Lunteren	10 - 11 Apr 2017	0,6
ALW meeting 'Experimental Plant Sciences', Lunteren	9 - 10 Apr 2018	0,6
► **Seminars (series), workshops and symposia**		
Seminar: PP2A phosphatase as a regulator of stress restistance in plants	28 May 2015	0,1
EPS Flying Seminar: Regulation of root morphogenesis in tomato species in the face of a changing environment (Dr. Siobhan Brady, UC Davis)	09 Sept 2015	0,1
EPS seminar: An integrated structural biology approach to flower development (Dr. Francois Parcy, IRTSV, Grenoble, France)	05 Oct 2015	0,1
Joined meeting groups Scheres and Weijers	22 Jan 2016	0,2
EPS Flying seminar: How do plants read their own shape (Dr. Olivier Hamant, ESN Lyon, France)	16 Mar 2016	0,1
Seminar: 'Auxin Signaling: Inputs and Outputs' (Prof. Mark Estelle, UC San Diego, CA, USA)	07 Apr 2016	0,1
Joined meeting groups Scheres and Weijers	16 Dec 2016	0,2
Seminar: 'Embryo development' at Ghent univeristy (Prof. Dolf Weijers, Wageningen UR, the Netherlands)	06 Jan 2017	0,1
Symposium Molecules @ WURk (Forum, Wageningen UR)	08 Feb 2017	0,2
Seminar: A new family of Microtubule-binding proteins (dr. Katharina Bürstenbinder, Halle, Germany)	19 Jun 2017	0,1
Seminar Towards understanding higher order structures in the cytoplasm using genetically encoded probes (Dr. Arnold Boersma, Groningen)	07 Oct 2017	0,1
Seminar: Underground metabolism and the predictability of evolution (dr. Richard Notebaart, Food Microbiology group, WUR)	02 Nov 2017	0,1
Seminar: Condensation of the ARF19 transcription factor regulates its activity (Lucia Strader; University of St. Louis, MO, USA)	10 Sept 2018	0,1
Seminar: Adding Action and Color to Electron Microscopy: New Tools for Correlated Light and Electron Microscopy (Stephen Adams, University of California San Diego, USA)	11 Sept 2018	0,1
► **Seminar plus**		
► **International symposia and congresses**		
International PhD School Plant development, Retzbach - Wurzburg, Germany	07 - 09 Oct 2015	0,9
Vascular get-to-gether, Oud-Turnhout, Belgium	24 - 25 Nov 2015	0,6
EMBO workshop: New model systems for early land plant evolution, Vienna, Austria	22 - 24 Jun 2016	0,9
Vascular get-to-gether, Oud-Turnhout, Belgium	15 - 16 Nov 2016	0,6
International PhD School Plant development, Retzbach - Wurzburg, Germany	04 - 06 Oct 2017	0,8
Vascular get-to-gether, Oud-Turnhout, Belgium	13 - 15 Nov 2017	0,5
Plant Embryogenesis congress, Haigerloch, Germany	07 - 09 May 2018	0,8
EMBO workshops: New shores in land plant evolution, Lisbon, Portugal	20 - 23 Jun 2018	1,2
Vascular get-to-gether, Oud-Turnhout, Belgium	12-14 Nov 2018	0,5
► **Presentations**		
Poster presentation at PhD school Plant development Retzbach (Germany)	07 - 09 Oct 2015	1,0
Oral presentations at Vascular get-to-gether (Belgium)	24 Nov 2015	1,0
Oral presentation at International PhD school Retzbach (Germany)	04 Oct 2017	1,0
Oral presentation at Vascular get-to-gether (Belgium)	14 Nov 2017	1,0
Oral presentation at EPS theme 1 day	30 Jan 2018	1,0
Oral presentation at Nematology, WUR (the Netherlands)	12 Apr 2018	1,0
Oral presentation at embryogenesis congress Haigerloch (Germany)	08 May 2018	1,0
Poster presentation at EMBO workshop (Lisbon, Portugal)	22 Jun 2018	1,0
► **IAB interview**		
► **Excursions**		
Scientific PhD excursion to Agrasys, ICFO, CRAG and Synchrotron, Barcelona, Spain	15 - 22 May 2017	1,0
Company visit Koppert	26 Oct 2018	0,2
Subtotal Scientific Exposure		21,6

3) In-Depth Studies	*date*	*cp*
► **Advanced scientific courses & workshops**		
Postgraduate course 'Transciption Factors and Transcriptional Regulation'	12-14 Dec 2016	1,0
Summer School in Advanced Light Microscopy, Ghent, Belgium	12 - 16 Jun 2017	1,5
► **Journal club**		
Weekly journal club	2015 - 2019	3,0
► **Individual research training**		
Bioinformatics training, Gent University, Belgium	04 - 12 Jun 2015	2,1
Subtotal In-Depth Studies		7,6

4) Personal Development	*date*	*cp*
► **General skill training courses**		
EPS introduction course	11 Feb 2016	0,3
Scientific Publishing (SP)	13 Apr 2017	0,3
Brain Training	07 Feb 2018	0,3
Adobe Indesign Essential training	04-05 Jun 2018	0,6
Career Orientation course	02 Oct - 01 Nov 2018	1,5
Infographics and Iconography	04 Dec 2018	0,3
► **Organisation of meetings, PhD courses or outreach activities**		
Organisation Scientific PhD excursion biochemistry	15 - 22 May 2017	1,5
Oral presentation 'Lunchlezing' Alchimica, molecular life science (WUR, the Netherlands)	22 Nov 2017	0,5
► **Membership of EPS PhD Council**		
Subtotal Personal Development		5,3

TOTAL NUMBER OF CREDIT POINTS*	**36,0**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.

* A credit represents a normative study load of 28 hours of study.