Evolution of bHLH transcription factors that control epidermal cell development in plants

Bruno Miguel Grou Catarino

Magdalen College
Department of Plant Sciences
University of Oxford
Supervised by Liam Dolan

Thesis submitted for the degree of Doctor of Philosophy
Michaelmas 2017

For Arminda dos Santos Almeida

Evolution of bHLH transcription factors that control epidermal cell development in plants

Bruno Miguel Grou Catarino

Magdalen College
Department of Plant Sciences
University of Oxford

Thesis submitted for the degree of Doctor of Philosophy
Michaelmas 2017

Abstract

The colonization of the arid continental surface by plants was one of the milestones in Earth's history. Morphological innovations, such as the origin of complex 3D tissues, allowed the successful colonization and radiation of plants on land. The epidermis is the outermost plant tissue that constitutes the interface between the plant and the environment. Thus, the evolution of epidermal cells was crucial for the adaptation of plants on the terrestrial arid environment. I undertook a combined approach that aims to understand the evolutionary trends that drove land plant colonization and the genetic mechanisms that underlie the development of the epidermis. This approach includes: 1) analyses of plant transcription factors (TFs) families distribution and diversification, with a particular focus on the basic Helix-Loop-Helix (bHLH) TF family, and 2) functional characterization of a putatively conserved bHLH TF subfamily involved in epidermal cell development in land plants. Here, I showed that there was a stepwise increase in the number of transcription factor (TF) families and bHLH subfamilies that predated the colonization of the terrestrial surface by plants. The subsequent increase in TF number on land was through duplication within pre-existing TF families and subfamilies. Moreover, a similar trend occurred in metazoan bHLH TF, suggesting that the majority of innovation in plant and metazoan TF families occurred in the Precambrian before the Phanerozoic radiation of land plants and metazoans. Furthermore, I demonstrated that the function of IIIf bHLH TFs in controlling the development of the epidermal cell layer is conserved between liverworts and angiosperms. This suggests that IIIf bHLH TFs are ancient and conserved regulators of epidermal cell development since the early colonization of the land by plants. Moreover, these bHLH TFs were recruited during the evolution of land plants to control the development of seemingly unrelated morphological characters in specific lineages of extant land plants. The recruitment of ancient developmental regulators to control distinct and unrelated developmental processes in land plants might underlie the huge morphological and taxonomic radiation of plants on land.

Abbreviations

aLRT	Approximate likelihood ratio test
amiR	Artificial microRNA
bHLH	basic Helix-Loop-Helix
CDS	Coding DNA sequence
CFP	Cyan fluorescent protein
EAR motif	Ethylene-responsive element binding factor-associated amphiphilic repression motif
Ef1α	Elongation factor 1α
gDNA	Genomic DNA
IAA	Indole-3-acetic acid
Kb	Kilo-base pair
L-Kyn	L-kynurenine
miRNA	microRNA
ML	Maximum likelihood
NAA	1-Naphthaleneacetic acid
NLS	Nuclear localisation signal
PCR	Polymerase chain reaction
PI	Propidium iodide
RT-qPCR	Reverse transcription-quantitative PCR
SEM	Scanning electron micrograph
SRDX	EAR-motif repression domain
TAIL-PCR	Thermal asymmetric interlaced PCR
Tak-1	Male wild type *Marchantia polymorpha* accession Takaragaike-1
Tak-2	Female wild type *Marchantia polymorpha* accession Takaragaike-2
T-DNA	Transfer DNA of *Agrobacterium tumefaciens*
TF	Transcription factor
WT	Wild type
YFP	Yellow fluorescent protein

My thanks

I would like to express my sincere gratitude to Liam Dolan, for his support, guidance and advice throughout my DPhil. I have learned so much during these years thanks to his mentorship. This thesis would not be a reality without him. I also thank my second supervisor, Steve Kelly, for advice, suggestions and fruitful collaboration. I thank my examiners Jill Harrison and Hugh Dickinson for fruitful discussion and providing guidelines to improve this thesis.

I am grateful to all the people from the Rhizosphere lab for all the help, support and camaraderie. In particular, Thomas Tam, Sandy Hetherington, Anna Thamm, Victor Jones, Suvi Honkanen, Clement Champion, Holger Breuninger, Réka Tóth and Ana Milhinhos for their insights, technical support and friendship.

I am thankful to Pablo Muñoz Rodríguez and Sabina Sanegre for useful and interesting scientific discussions.

I would also like to thank all the remaining people from the Department of Plant Sciences who are always supportive and encouraging.

Special thanks to Javier Agustí, whose advice, support and friendship helped me tremendously during my DPhil.

Finally, thanks to my family and friends – despite the distance, they are always close to my heart.

Thanks to you all.

October 2017

Oxford

Bruno

Supplementary files available on CD

Supplementary File 1 - Transcription factor family distribution in Archaeplastida

Supplementary FIle 2 - Klebsormidium bHLH accession numbers

Supplementary FIle 3 - Marchantia polymorpha bHLH accession numbers

Supplementary File 4 - bHLH sequences

Supplementary File 5 - bHLH alignment

Supplementary File 6 - ML phylogenetic tree of plant bHLH

Supplementary File 7 - Va Vb XIII XVI Subfamilies sequences

Supplementary File 8 - Va Vb XIII XVI Subfamilies alignment

Supplementary File 9 - IIIf Subfamily sequences

Supplementary File 10 - IIIf Subfamily alignment

Supplementary File 11 – Electronic version of this thesis

Table of Contents

Chapter 3: IIIf bHLH transcription factors are conserved regulators of epidermal cell development in land plants 76

Chapter 1: General Introduction

The colonization of the continental surface by multicellular plants over 470 million years ago was one of the milestones in Earth's history (Gensel, 2008; Kenrick and Crane, 1997; Wellman et al., 2003). The evolution of land plants allowed the subsequent colonization of the terrestrial surface by metazoans and paved the way for the foundation of complex terrestrial ecosystems, transforming drastically the terrestrial surface (Bateman et al., 1998; Berner et al., 2007; Kenrick et al., 2012; Lenton et al., 2012, 2016). Land plants were also responsible for shaping the geochemical and atmospheric cycles by increasing the mineral weathering and gas exchanges between the plant and the environment (Lenton et al., 2012, 2016). One of the key morphological breakthroughs in plant evolution that allowed the successful adaptation and radiation of plants on land was the development of complex three-dimensional tissues produced by meristems that mediate the body plan development (Graham et al., 2000). The origin and evolution of a specialized tissue at the boundary between the plant and the external environment was therefore crucial for the establishment of complex biotic and abiotic interactions, and for the adaptation of plants on the terrestrial realm. The plant epidermis is the outermost cell layer that protects the inner tissues and constitutes the boundary between the plant and the environment. The understanding of the genetic and molecular basis of the development and evolution of epidermal cells provides insights into fundamental mechanisms underlying plant evolution that allowed the colonization and radiation of the first plants on the terrestrial surfaces.

In this thesis I aim to identify the origin and evolution of basic Helix-Loop-Helix (bHLH) plant transcription factors (TFs) which comprise several genetic components required for the development of epidermal cells in land plants and to characterize the ancestral function of one bHLH subfamily that regulates epidermal cell development. These aims will be addressed in Chapters 2 and 3 in order to elucidate the genetic mechanism that underpins the evolution of epidermal cell development in land plants. In this chapter, I provide a brief overview on the history of land plants, epidermal cell development and review some genetic mechanisms underlying bryophyte development and how these mechanisms have evolved during land plant radiation. This review highlights that ancient and conserved developmental regulators underpin the evolution of the body plan in land plants. This is followed by a small section that describes the life-cycle of the liverwort *M. polymorpha*, that I use as a model system to functionally characterize the function of one bHLH subfamily in early divergent land plants and, with analogy to the function in angiosperms, infer the function of these transcription factors in the common ancestor between liverworts and angiosperms – the first land plants.

1.1. Origin and evolution of land plants

One pivotal event in Earth's history was the colonization of land and subsequent adaptation of plants to the terrestrial environment over 470 million years ago (Gensel, 2008; Kenrick and Crane, 1997; Wellman et al., 2003). The colonization of the terrestrial realm by plants had dramatic impacts on the geochemical cycles, atmosphere and climate (Kenrick et al., 2012; Lenton et al., 2012, 2016), and paved the way for the later terrestrial invasion of animals and establishment of the first complex terrestrial ecosystems (Bateman et al., 1998; Berner et al., 2007).

Land plants (embryophytes) are a monophyletic clade that evolved from freshwater multicellular algae (Lewis and McCourt, 2004; Timme et al., 2012; Wickett et al., 2014; Wodniok et al., 2011) (Figure 1.1). The closest living relatives to land plants are the freshwater charophyte algae, which together with land plants constitute a monophyletic clade, the streptophytes (Cox et al., 2014; Finet et al., 2010; Karol et al., 2001) (Figure 1.1). The streptophytes evolved some characters such as hexameric cellulose synthase complexes, presence of phragmoplast, plasmodesmata, and apical cells with less than three cutting edges

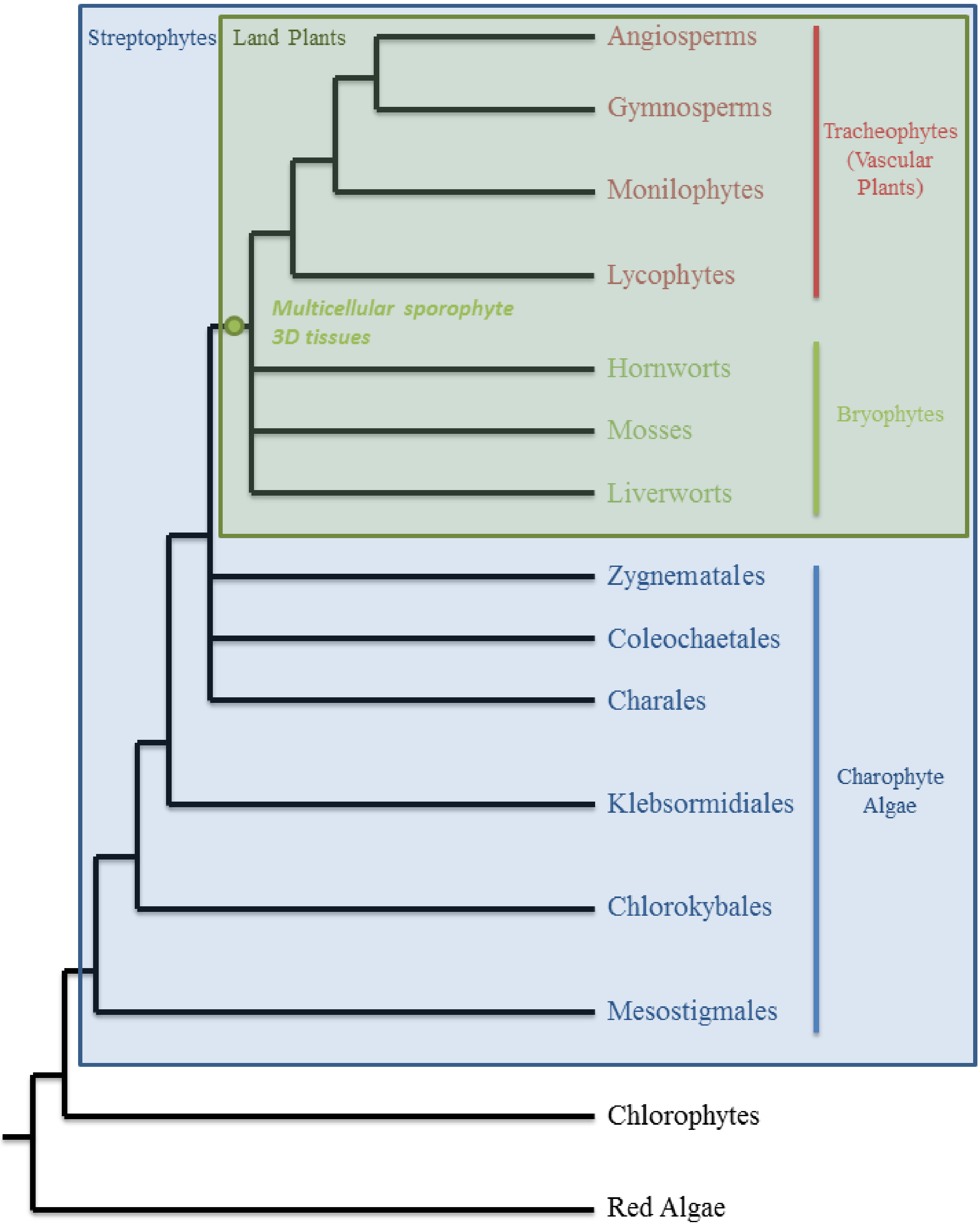


Figure 1.1 - Phylogeny of land plants. Streptophytes (blue box) are a monophyletic clade comprised by charophyte algae and land plants. The charophyte algae Charales, Coleochaetales and Zygnematales constitute a paraphyletic clade that is sister to land plants (Becker and Marin, 2009; Bowman, 2013; Chang and Graham, 2011; Finet et al., 2010; Kenrick and Crane, 1997; Timme et al., 2012; Wickett et al., 2014; Wodniok et al., 2011). Land plants (green box) are a monophyletic clade that evolved over 470 million years ago (Gensel, 2008; Kenrick and Crane, 1997; Wellman et al., 2003). The earliest lineages are the bryophytes that comprise liverworts, mosses and hornworts. The bryophytes are sister to the later divergent clade of land plants – the vascular plants (Kenrick and Crane, 1997; Wickett et al., 2014). The evolution of multicellular sporophyte and 3D tissues were key morphological adaptations that allowed the colonization of the terrestrial surface by plants and it is highlighted at the branch that comprises all land plants.

(Becker and Marin, 2009; Graham et al., 2000; Pires and Dolan, 2012). These characters are not found in the sister lineage of the streptophytes, the chlorophyte algae. Indeed, the colonization and radiation of land plants in the terrestrial realm from a streptophyte aquatic ancestor, were associated with the evolution of key morphological innovations such as: apical cells with three cutting edges and retention of the zygote within the female gametangia (embryo) with intercalated mitotic divisions between fertilization and meiosis (Graham et al., 2000; Haig, 2008; Niklas and Kutschera, 2009). These innovations allowed the origin of multicellular three-dimensional (3D) tissues and multicellular sporophyte, respectively.

It is still unclear which charophyte lineage represents the closest living relatives of the land plants, but phylogenetic and cladistics studies consistently support that the paraphyletic clade formed by Coleochaetales, Charales and Zygnematales is the sister clade to the land plants (Becker and Marin, 2009; Bowman, 2013; Chang and Graham, 2011; Finet et al., 2010; Kenrick and Crane, 1997; Timme et al., 2012; Wickett et al., 2014; Wodniok et al., 2011) (Figure 1.1). Further molecular and phylogenetic analyses are needed to resolve the phylogenetic relationships within the streptophytes and elucidate which one of these three charophyte lineages represents the closest relative lineage of the land plants (embryophytes).

Plants have hugely radiated both taxonomically and morphologically after the colonization of the terrestrial surface by a streptophyte aquatic ancestor. The earliest diverging lineages of extant land plants are the liverworts, mosses and hornworts, collectively known as bryophytes (Kenrick and Crane, 1997; Wickett et al., 2014) (Figure 1.1). The phylogenetic relationship between bryophyte lineages is unresolved, as different studies show disparities in their phylogenetic relationships (Cox et al., 2014). The bryophytes are sister to the clade comprising all the other extant land plant lineages (vascular plants) (Figure 1.1).

The earliest fossil evidences of the presence of plants on land revealed that the first land plants were morphologically similar to extant bryophytes. Microfossils of cryptospores and small tissue fragments from the Middle Ordovician, around 470 million years ago, are the oldest fossil record of land plants (Gensel, 2008; Rubinstein et al., 2010; Wellman et al., 2003). These microfossils show characters, like the wall ultrastructure, that resemble the extant liverworts; however the first macrofossils of liverworts only appear in the Middle Devonian, 420-390 million years ago (Guo et al., 2012; Hernick et al., 2008). The oldest fossil evidence of vascular plants comes from triplete spores from Upper Ordovician, around 443 million years ago (Steemans et al., 2009). Nevertheless, the first macrofossils of vascular plants, represented by the sporophyte of *Cooksonia*, only appear on Mid-Late Silurian sediments (425 million years ago) (Edwards and Feehan, 1980; Gensel, 2008). The first fossils of gametophytes are from the Early Devonian (420-390 million years ago) Rhynie Chert site, and show that complex characters only found in sporophytes, such as stomata, developed in the gametophytes of the first land plants (Kenrick, 1994; Kerp et al., 2004; Taylor et al., 2005, 2009). These fossil evidences demonstrate that, similarly to all extant land plants, the life cycle of the first land plants were comprised of two alternating multicellular generations: the gametophyte and the sporophyte. Moreover, the gametophytes of the first land plants were morphologically complex, resembling extant liverworts.

The origin and early evolution of tracheophytes (vascular plants) with a dominant sporophyte generation on the Devonian period (415-360 million years ago) allowed plants to colonize and adapt to drier habitats (Bateman et al., 1998; Kenrick and Crane, 1997). This period saw the origin and radiation of many extant lineages of land plants: lycophytes, monilophytes and gymnosperms (Kenrick and Crane, 1997; Taylor et al., 2009). The first giant trees, over 50 meters tall, are found in the Carboniferous (359-300 million years ago) coal swamp forests (DiMichele and Phillips, 1994; Thomas and Watson, 1976). These lycophyte trees have highly dichotomous branched root architecture that has been conserved in the small extant lycophyte herbs (Hetherington et al., 2016). Flowering plants (angiosperms) appeared and radiated during the early Cretaceous (146-100 million years ago) colonizing most of the terrestrial habitats and became the dominant flora until the present days (Clarke et al., 2011; Friedman, 2009; Smith et al., 2010).

1.2. Plant epidermis

The epidermis, outer skin, constitutes the outermost cell layer, or layers, that covers and protects the primary plant body. It is the first and primary plant tissue that interacts with the environment, constituting therefore the interface between the plant and the surrounding environment. It develops in all extant land plants and it originated between the divergence of streptophyte algae and land plants (Pires and Dolan, 2012). The evolution of the epidermis was critical for the radiation of plants on the terrestrial realm, by providing a protective layer against desiccation and other exogenous factors like light and temperature. However, some plant tissues do not develop an epidermis, as discussed below. The epidermis is comprised by both unspecialized and specialized (differentiated) cells that carry out a specialized function at the surface of the plant (Datta et al., 2011; Payne, 1979; Yang and Ye, 2013). In this section I will highlight the development, function and evolution of plant epidermis in land plants.

1.2.1. Development of the epidermis in gametophytes

The gametophyte of the earliest divergent lineage, or lineages, of land plants (the bryophytes, the collective name for the liverworts, hornworts and mosses) is an independent free-living organism that develops complex three dimensional tissues, including the epidermis (Figure 1.2). The epidermis in the model liverwort *M. polymorpha*, is comprised of a single layer of cells and it is present in both the dorsal and ventral sides of the thallus (gametophyte). The

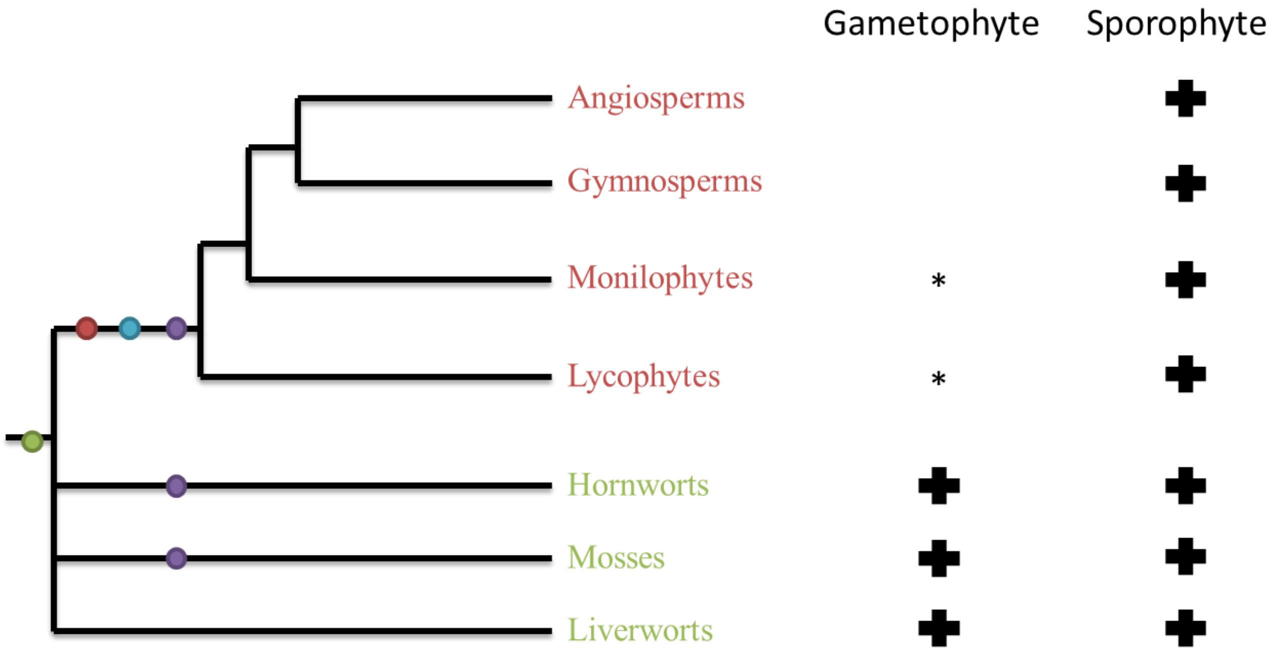


Figure 1.2 – Land plant phylogeny and the evolution of the plant epidermis. The development of complex three-dimensional tissues from meristems was an important morphological innovation that evolved around the time plants colonized the land (Graham et al. 2000). These tissues carried out discrete function such as gas exchange, photosynthesis and water transport. The origin of specialized tissue at the boundary between the plant and the external environment – the epidermis – was one such specialization. The epidermis develops in all extant lineages of land plants but it is absent in some tissues like the protonema in bryophytes and the prothallus in lycophytes and monilophytes. Plus (+) marks the presence of a defined epidermis. Asterisks note land plant lineages that do not develop an epidermis per se, however the development of rhizoids take place in the prothallus. The green, red, blue and purple dots mark the evolution of plant epidermis, root hairs, trichomes and stomata, respectively.

dorsal epidermis derives from anticlinal cell divisions of the dorsal merophyte whereas the ventral epidermis is formed by anticlinal cell divisions of the ventral merophyte (Shimamura, 2016). The epidermis of bryophytes develops specialized epidermal cells that carry out the rooting function in the gametophyte termed rhizoids. Additionally, in some liverwort species, specialized epidermal cells that facilitate gas exchange develop, forming what is termed an air pore complex.

However, some bryophytes do not develop an epidermis in the first developmental stages after spore germination. As an example, the moss *Physcomitrella patens* develops a two-dimensional filamentous structure after spore germination called protonema. These filamentous structures are composed primarily by chloronemata cells with numerous chloroplasts which are photosynthetically very active. Chloronema can also originate branches from subapical cells, giving rise to a network of branched filamentous cells. After several days of chloronemata development, the apical cells start to differentiate into caulonemata cells which are longer and have fewer chloroplasts than chloronemata cells. Caulonema increases the size of the colony and therefore is important to colonize the substrate, carrying therefore a rooting function. Protonema is therefore a tissue that does not comprise an epidermis. However, some of the genes that are active in the epidermis of land plants and are required for the development of filamentous rooting cells (discussed in further sections of this Introduction) control the differentiation of chloronema cells into caulonema cells (Menand et al. 2007, Pires et al. 2013, Tam et al, 2015). This suggests that caulonema development shares similar gene regulatory network and functions like filamentous rooting cells that develop in the epidermis of land plants. The moss *P. patens*, produces buds mainly in caulonema cells, that will give rise to the gametophore that comprises an epidermis which is form after successive spiral cleavages of a tetrahedral apical cell (Harrison et al. 2009).

The fossil record of the pre-vascular plant *Aglaophyton major* from the Rhynie chert, c.a. 400 million years ago, indicates that these gametophytes were complex in structure and anatomy, and developed stomata, unlike the gametophytes of all extant lineages of land plants (Taylor et al., 2005). In lycophytes and monilophytes , unlike the sporophyte, the gametophyte lacks the development of complex tissues, including the epidermis, however these gametophytes still differentiates rhizoids (Banks, 1999, 2009; Schulz et al., 2010) (Figure 1.2). In later divergent lineages of land plants, the seed plants, the gametophyte is reduced in complexity and the three dimensional tissue organization characteristic of the gametophytes of early divergent lineages of land plants is absent (Haig, 2008). There is a gradual loss of tissue organization and complexity in the gametophyte generation that occurs during land plant evolution and, concomitantly, the epidermis becomes absent in the gametophytes of later divergent lineages of land plants (Figure 1.2). The fossil evidence and the morphological characterization of extant land plants lead to the hypothesis that the reduced complexity in gametophyte development, including the epidermis, was a result of loss of gene expression required for epidermis development in the gametophyte, or that this genetic toolkit was recruited to the sporophyte generation during land plant evolution.

1.2.2. Development of the epidermis in sporophytes

The epidermis develops in all extant land plant sporophytes. However, the development of epidermal cells differs in the sporophyte of vascular plants. For example, both upper and lower epidermis of the fern leaves contain chloroplasts; this feature is absent in the epidermis of seed plants (except in the stomata guard cells) (Vasco et al., 2013).

Specialized epidermal cells develop in all sporophytes, except in liverworts. The moss and hornwort sporophytes develop stomata guard cells, and these specialized epidermal cells develop as well in the sporophyte generation of all extant lineages of vascular plants (Payne, 1979). Trichomes, for instance, develop only in the sporophyte of vascular plants (Calvert et al., 1985; Hueber, 1992). Similarly to trichomes, root hairs also develop only in the sporophyte of vascular plants. However, these cells are analogous in function and share a common genetic mechanism with the gametophytic rhizoids (Datta et al., 2011) (Figure 1.2).

The development of the epidermis is ontogenetically different in the two generations and in the different plant organs. In the next section I describe how the epidermis develops in three different organs of the angiosperm *A. thaliana*.

1.2.3. Development of epidermis in *Arabidopsis* embryo, shoot and root

The model plant species, *A. thaliana*, has been extensively studied regarding its development, and provides therefore a suitable model to describe epidermis development in different plant organs. The *A. thaliana* embryo development has been characterized from zygote to full embryo comprising shoot and root apical meristems. The first step of embryo development is characterized by elongation of the zygote, after fertilization, followed by an asymmetric cell division that establishes a small apical cell and a large vacuolated basal cell (De Smet et al., 2010; Wendrich and Weijers, 2013). The basal cell divides symmetrically to form the suspensor, the connective tissue. The apical cell divides symmetrically twice, changing its division plane, to form an octant. Subsequent diagonal divisions form the dermatogen, with the establishment of the radial axis (De Smet et al., 2010). At this stage, the protodermal cells are the outermost cells and will form the epidermal tissue after successive rounds of symmetric cell divisions (ten Hove et al., 2015) (Figure 1.3). The inner cells undergo a series of asymmetric cell divisions to form the vascular and ground tissue precursors. Later in embryo development, at the so-called heart stage, the hypophysis and shoot and root apical meristems are established after a series of cell division and differentiation (ten Hove et al., 2015; De Smet et al., 2010; Wendrich and Weijers, 2013).

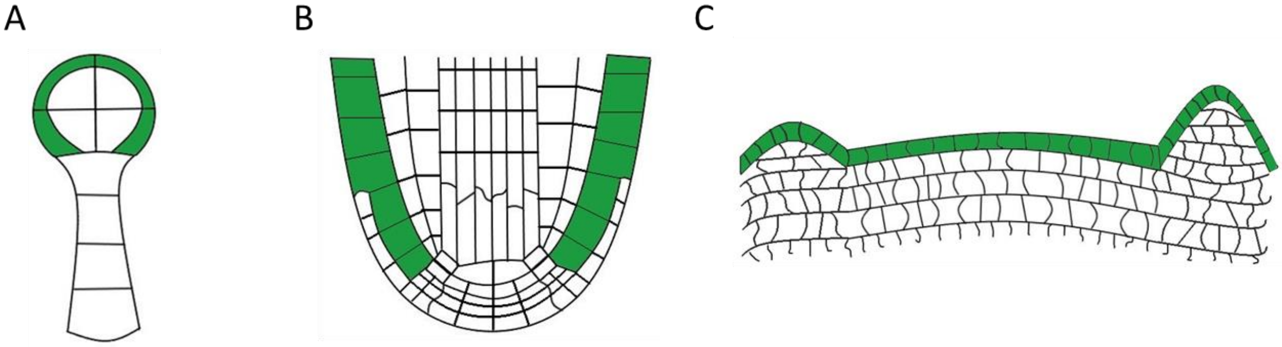


Figure 1.3 – Schematic representation of (A) embryo; (B) root apical meristem and (C) shoot apical meristem showing different ontogenies of epidermis (coloured in green) in *A. thaliana*. In the embryo, at the dermatogen stage, the protodermal cells (in green) are the outermost cells and will form the epidermal tissue after successive rounds of symmetric cell divisions. In the root apex, epidermal cells derive from root cap and epidermal initials which are arranged in a ring and undergo a periclinal division to form the root cap (outer derivative) and an inner derivative that undergoes transverse division to regenerate the initials and form the epidermal cells (in green). In the shoot apex, epidermal cells (in green) are produced by symmetric anticlinal divisions of the outer cell layer (L1).

The shoot meristematic activity gives rise to the above ground plant tissue, including the stem and leaves. Cells of the shoot meristem are arranged in discrete layers in which the outer layer (L1) undergoes predominantly symmetric anticlinal divisions to produce the epidermis. Stem and leaves epidermis are comprised by both unspecialized and differentiated cells (Shapiro et al., 2015) (Figure 1.3). Trichomes and stomata guard cells are two of the differentiated epidermal cells that develop on stems and leaves (Yang and Ye, 2013). Trichomes are unicellular projections that extend from the epidermis of both stem and leaf and provide protection against herbivores and water loss. Despite carrying out similar functions in both the stem and leaves, the structure of trichomes differs between the two organs; trichome cells are tubular on stems but branched on leaves (Yang and Ye, 2013).

The root meristem is arranged into three different tiers of cells that will differentiate the tissues that comprise the root. The lower (outermost) tier of cells, protoderm, comprises the initials of the root cap and epidermis (Dolan et al., 1994). The root cap and epidermal initials are arranged in a ring and undergo a periclinal division to form the root cap (outer derivative) and an inner derivative that undergoes transverse division to regenerate the initials and form the epidermal cells (Dolan and Roberts, 1995) (Figure 1.3). The root epidermal cells are patterned into intercalated files of root hair cells and non-root hair cells. Frequently, the root hair cells lie over two cortical cells whereas the non-root hair cells lie over one cortical cell. The root hair cells undergo further differentiation with the development of a root hair. Root hair development is initiated by a bulging of the cell, followed by a tubular elongation, forming the hair that develops by tip-growth (Berger et al., 1998; Dolan and Roberts, 1995; Dolan et al., 1994). These epidermal projections provide nutrient uptake and anchorage to the substrate.

The genetic mechanism governing the patterning of these specialized cells; trichomes, stomata and root hairs, is described in detail in the section 3.2 of this thesis. In the following sections I highlight the genetic mechanisms underlying the development of epidermal cells and cells with a rooting (rhizoids and root hairs) and gas exchange (air pores and root hairs) functions.

1.2.4. MIXTA transcription factors are conserved regulators of epidermal cell development in angiosperms

Little is known about the genetic regulation of epidermal cell development in land plants with the exception of the model angiosperm *A. thaliana*. However, some studies in different angiosperm species have highlighted that a conserved MYB TF class, called MIXTA, is conserved in angiosperms to control epidermal cell and trichome development (Jakoby et al., 2008; Lashbrooke et al., 2015; Perez-Rodriguez et al., 2005; Di Stilio et al., 2009). It is still unknown whether conservation of genetic mechanisms underlying epidermal development operates across angiosperms, or even, across land plant. The Homerodomain TF AtML1, and its homologs control the development of epidermal cells in the shoot meristem of *A. thaliana* (Abe, 2003; Peterson et al., 2013; Takada et al., 2013). It would be interesting to dissect if this class of TFs also controls the development of epidermal cells in angiosperms and other land plant lineages to understand if different ontogenies of epidermis share a common and conserved regulatory mechanism.

1.2.5. Development of epidermal cells with a rooting function

The evolution of a rooting system was crucially important for the colonization and adaptation of plants in the terrestrial realm. Almost all land plants develop filamentous tip-growing cells that provide nutrient and water uptake from the soil, and anchorage to the substrate (Jones and Dolan, 2012). Unicellular tubular projections that develop in the root epidermis, root hairs, are found in the sporophyte generation of vascular plants. Root hairs have been extensively studied as a model to understand post-mitotic cell growth and two-dimensional tissue patterning (Datta et al., 2011). As a result, the genetic mechanism underlying root hair development is well understood and several molecular components that regulate root hair growth have been identified.

Rhizoids are filamentous tip-growing cells with a rooting function that develop in the gametophyte generation of bryophytes, lycophytes and monilophytes (Duckett et al., 2014; Jones and Dolan, 2012; Pearson, 1969). These cells are analogous to root hairs and some developmental regulators of root hair development are known to be also involved in rhizoid development (Breuninger et al., 2016; Datta et al., 2011; Menand et al., 2007; Pires et al., 2013; Proust et al., 2016; Tam et al., 2015). Three different basic Helix-Loop-Helix (bHLH) subfamilies form a transcriptional regulatory network that has been conserved to control rhizoid and root hair development amongst land plants (Breuninger et al., 2016; Menand et al., 2007; Pires et al., 2013; Proust et al., 2016; Tam et al., 2015). In the angiosperm *A. thaliana*, bHLH TFs Class I and Class II ROOT HAIR DEFECTIVE SIX-LIKE (Class I and Class II AtRSL), *Lotus japonicus* RHL1-LIKE (AtLRL) and the phytohormone auxin form a hierarchal transcriptional network that regulates root hair development (Bruex et al., 2012;

Karas et al., 2009; Menand et al., 2007; Pires et al., 2013; Yi et al., 2010). These molecular components are present in bryophytes and also constitute a transcriptional network that regulates rhizoid development (Breuninger et al., 2016; Jang et al., 2011; Menand et al., 2007; Pires et al., 2013; Proust et al., 2016; Tam et al., 2015). However the ‘wiring’ of this transcriptional regulatory network has changed during land plant evolution (Pires et al., 2013) (Figure 1.4). For example, in the moss *P. patens* auxin positively regulates both Class I Pp*RSL* and Pp*LRL* gene expression but these two bHLH subfamilies are transcriptionally independent of each other (Jang and Dolan, 2011; Jang et al., 2011; Pires et al., 2013; Tam et al., 2015) (Figure 1.4). In contrast, in *A. thaliana*, auxin does not regulate Class I At*RSL* gene expression (Bruex et al., 2012; Yi et al., 2010) (Figure 1.4). At*LRL1* and At*LRL2* gene expression is also independent of auxin, while At*LRL3* is positively regulated by auxin (Bruex et al., 2012) (Figure 1.4). Moreover, At*LRL3* transcription is positively controlled by Class I AtRSL TFs (Karas et al., 2009), while Class I At*RSL*, At*LRL1* and At*LRL2* are transcriptionally independent of each other (Bruex et al., 2012; Karas et al., 2009; Tam et al., 2015) (Figure 1.4). This demonstrates that the bHLH transcriptional regulatory network that controls the development of filamentous tip-growing cells with a rooting function was active in the first land plants and it was recruited and re-wired during the evolution of land plants.

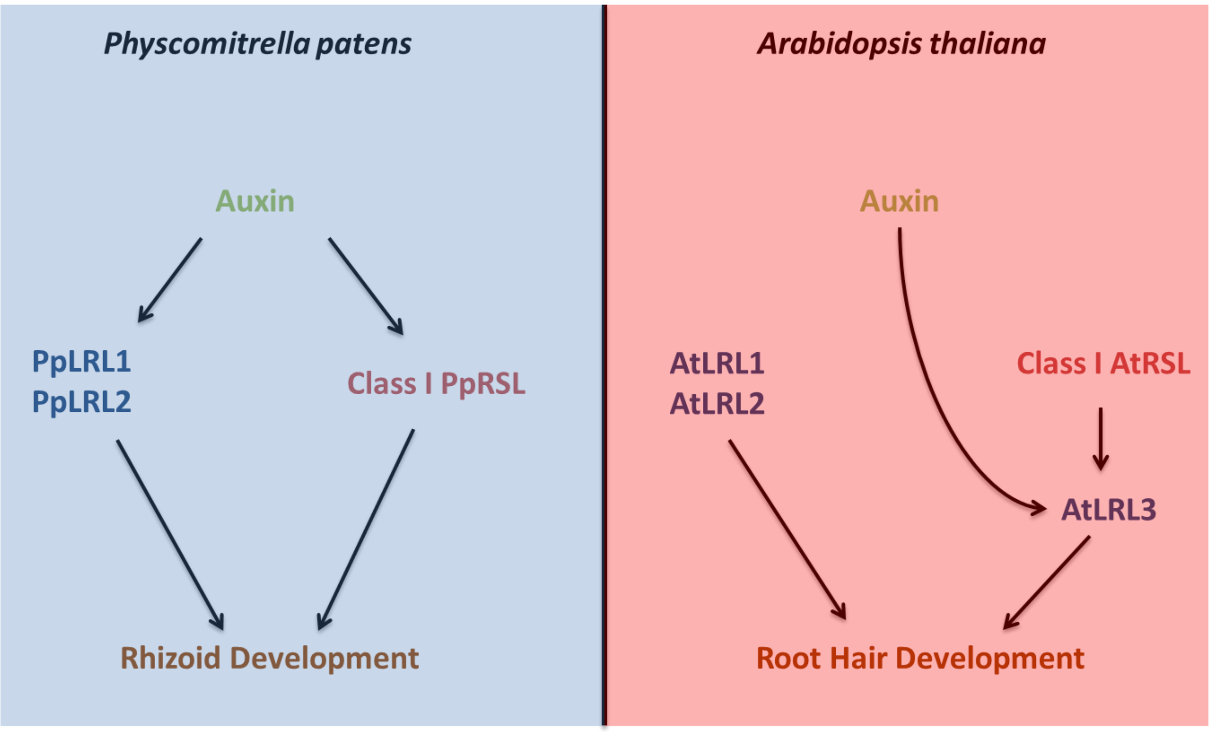


Figure 1.4 – Genetic network controlling the development of cells with a rooting function in land plants. Auxin, RSL and LRL bHLH transcription factors are ancient components of a conserved regulatory network that controls the development of filamentous cells with a rooting function in land plants. The gene regulatory networks of the moss *Physcomitrella patens* and the angiosperm *Arabidopsis thaliana* demonstrate conservatism of this module in land plants.

1.2.6. Development of specialized epidermal structures that facilitate gas exchanges

Land plants develop specialized epidermal cells that facilitate gas diffusion which is crucial for photosynthetic processes. Stomata are pores that are generally surrounded by two guard cells, except in the moss *Funaria* where only one guard cell surrounds the pore (Merced and Renzaglia, 2016). Guard cells are specialized epidermal cells that control the aperture of the pore in response to internal and environmental cues (Hetherington, 2001; Hetherington and Woodward, 2003). These specialized epidermal cells are found in the sporophyte generation of all land plant lineages, except in liverworts where analogous structures to the stomata are found, the air pores, which develop in the gametophyte generation (Ruszala et al., 2011). Air pores are typically constituted of a tier of 4x4 cells that forms a barrel-shape structure that surrounds the pore allowing gas exchange between the parenchymatic tissues in the air chamber, and the environment (Apostolakos et al., 1982). Stomata development has been studied extensively in the past two decades and the genetic mechanism that underpins the stomata morphogenesis is very well understood (Bergmann and Sack, 2007). However, very little is known about the development of air pores in liverworts and therefore it is still unknown whether these two analogous structures, stomata and air pores, share a genetic mechanism at some extent.

The pattern of cell divisions by which stomata develop in bryophytes and lycophytes differs from the one that occurs in angiosperms. While in bryophytes and lycophytes the establishment of guard cells is preceded by only a symmetric or asymmetric cell division, in angiosperms the differentiation of guard cells is preceded by at least one asymmetric division and a subsequent symmetric division (Payne, 1979). Despite the differences in the development of the guard cells in bryophytes and angiosperms, the molecular mechanism has been suggested to be conserved since mosses and angiosperms shared a common ancestor around 410 million years ago.

Stomatal patterning is controlled by a module comprising Epidermal Patterning Factors (EPFs) that interact with membrane-localised receptors in angiosperms (Bergmann and Sack, 2007). This module has been shown to regulate the density and distribution of stomata in the moss *P. patens*, suggesting that stomatal patterning is conserved between mosses and angiosperms (Caine et al., 2016). Moreover, proteins of two bHLH TFs subfamilies, Ia and IIIb, control the stepwise morphogenesis of stomata in angiosperms and act downstream of the module that regulates stomatal patterning (Pillitteri and Torii, 2012). It has been shown that Ia and IIIb bHLH TFs from the moss *P. patens* controls stomata development and Ia bHLH TFs partially complement loss-of-function mutants of Ia bHLH in the angiosperm *A. thaliana*, suggesting that Ia and IIIb bHLH TFs are ancient regulators of stomata development in land plants (Chater et al., 2016; MacAlister and Bergmann, 2011). These evidences demonstrate that the molecular mechanism underpinning stomatal patterning and morphogenesis might be ancient and it was likely to operate in the first plants that develop stomata.

It is still unknown whether the molecular mechanisms that control stomata development are also involved in air pore morphogenesis in liverworts. However, recent studies have identified components of the molecular mechanism that controls air pore development in the liverwort *Marchantia polymorpha* that are not known to regulate stomata development in other land plants. Loss-of-function mutants of the E3 ubiquitin ligase NOPPERABO1 (NOP1) exhibit a dorsal epidermal layer without any air chamber comprising air pores, suggesting that NOP1 and its target proteins constitute a molecular pathway that regulates the formation of air chambers and air pores in *M. polymorpha* (Ishizaki et al., 2013a). The MpWIP TF is a putative downstream target of the NOP1-regulatory mechanism that controls air chambers and air pores development (Jones and Dolan, 2017). Plants with reduced Mp*WIP* expression develop abnormal air pores in *M. polymorpha*, despite the presence of air chambers, which is consistent with the hypothesis that MpWIP operates downstream of NOP1 (Jones and Dolan, 2017). These lines of evidence suggest that, despite the analogous function in facilitating gas exchange between the plant and the environment, air pores and stomata have different molecular mechanisms underlying their development.

1.3. Genetic mechanisms underlying bryophyte development

Morphological innovations, such as the origin of complex 3D tissues, allowed the successful colonization and radiation of plants on land (Graham et al., 2000; Haig, 2008; Niklas and

Kutschera, 2009). The evolution of the genetic mechanisms underlying these morphological novelties was therefore crucial for the development of the first land plants and subsequent land plant lineages. Comparative analyses between the development of the earliest and later divergent lineages of land plants allow the inference of developmental mechanisms that operated in the common ancestor of extant land plants – the first land plants. Bryophytes are the earliest divergent lineage of land plants and constitute thus far an excellent model system to uncover the evolution of developmental mechanisms that orchestrated the development of the body plan of early land plants.

Most of the current knowledge regarding the molecular mechanisms underpinning morphogenesis and development in plants relies on studies in angiosperms. Nevertheless, during the past decade, studies dissecting the genetic and molecular framework of bryophyte development provided useful insights into the evolution of ancient developmental mechanisms in plants. The following sections highlight some genetic mechanisms underlying land plant development and evolution, with a special focus on bryophyte development.

1.3.1. Alternation of generations

The life cycle of land plants is comprised by two alternating multicellular generations, the haploid (gametophyte) and diploid (sporophyte) generations and constitutes an autapomorphic feature of land plants (Bower, 1890; Graham et al., 2000; Haig, 2008; Hofmeister, 1862). Since charophyte algae do not have a multicellular sporophyte, and the sporophyte of bryophytes is small and dependent on the dominant gametophyte generation, it is accepted that the sporophyte increased in size and dominance during land plant evolution to become the dominant generation in later divergent lineages of land plants, the vascular plants (Bower, 1890). In bryophytes, the sporophyte is a small multicellular body that is dependent on the dominant haploid generation where the majority of the photosynthesis occurs. The diploid sporophyte develops after fertilization and undergoes successive mitotic divisions and differentiation until it produces haploid spores through meiosis. The haploid spore will then undergo multiple mitotic divisions in order to develop the dominant multicellular gametophyte that produces gametes. During land plant evolution the sporophyte increased in size and complexity and became the dominant generation of the life cycle of later divergent lineages of land plants, where the gametophyte generation is confined to just a few cells.

The transition from one generation to another can also occur, but less frequently, without meiosis or fertilization taking place. The irregular sporophyte-to-gametophyte transition results from the development of gametophyte tissue from a sporophyte without meiosis and subsequent production of spores (apospory) (Bell, 1992). The developing aposporous gametophyte is diploid (2n), in contrast with the normal gametophyte that is haploid (1n). The transition from gametophyte-to-sporophyte can rarely take place without fertilization of gametes (apogamy) (Bell, 1992). In this scenario, somatic gametophyte cells generate a sporophyte-like body which is haploid (1n). Apospory and apogamy can be induced in bryophytes grown on severe conditions and can occasionally take place in natural environments.

The genetic mechanisms regulating the sporophyte and gametophyte program are different. Moreover, the transition between these two distinct developmental programs is genetically regulated by repressing one developmental program and activating the other. Class 2 KNOX (KNOX2) Homeodomain TFs are important regulators of sporophyte development in angiosperms (Furumizu et al., 2015; Li et al., 2012). In the moss *Physcomitrella patens*, KNOX2 TFs repress the gametophytic developmental program during sporophyte development (Sakakibara et al., 2013). The expression of these genes in the egg cells and sporophyte of the moss is necessary for sporophyte body plan development (Sakakibara et al., 2013). *P. patens* plants that lack functional KNOX2 are unable to develop a sporophyte capable of producing spores. Instead, after fertilization, the embryo develops a gametophyte-like body (filamentous cells), which is a result of apospory (Sakakibara et al., 2013).

KNOX2 TFs heterodimerise with other Homeodomain TF, BELL, and together they control the development of the diploid generation of the life cycle in plants (Bhatt et al., 2004; Hay and Tsiantis, 2010; Lee et al., 2008). This heterodimerization also occurs in the moss *P. patens* and the BELL TF, when ectopically expressed, is sufficient to trigger the development of a sporophyte-like structure in the haploid generation without fertilization taking place (apogamy) (Horst et al., 2016). The apogamous sporophyte is able to undergo meiosis and produce spores which are less viable than WT spores (Horst et al., 2016).

The expression of BELL is regulated by Polycomb Repressive Complex 2 (PRC2) proteins, namely CURLY LEAF (CLF) and FERTILIZATION-INDEPENDENT ENDOSPERM (FIE) (Pereman et al., 2016). PRC2 proteins epigenetically repress (by mediating tri-methylation of H3K27) the expression of BELL Homeodomain in *P. patens* (Pereman et al., 2016). Moreover, plants that lack functional PRC2 proteins develop sporophyte-like structures in the haploid phase of the life cycle of *P. patens* (Mosquna et al., 2009; Okano et al., 2009).

The genetic mechanism comprising the TFs KNOX2 and BELL and the PRC2 proteins CLF and FIE controls the transition of generations in the life cycle of the moss *P. patens*. This regulatory machinery has been shown to be mainly involved in sporophyte development and the transition of the two morphological distinct generations of the life cycle in angiosperms (He et al., 2012). It is therefore possible to infer that this genetic mechanism has been employed early in the evolution of land plants to regulate the developmental switch between one multicellular generation and another. The temporal regulation of this genetic mechanism during land plant evolution contributed therefore to the increase of size and dominance of the sporophyte over the gametophyte generation.

1.3.2. Sporophyte development

The *de novo* origin of a multicellular diploid sporophyte generation was one of the key developmental innovations that characterised the early colonization of the land by plants (Graham et al., 2000; Haig, 2008). While the sporophyte in vascular plants is the dominant generation of the life cycle, the sporophyte in bryophytes is a small gametophyte-dependent multicellular body (Kenrick and Crane, 1997). The evolution of the sporophyte in bryophytes, and other land plants, is postulated to derive from the retention of the zygote in the female gametophyte (Becker and Marin, 2009; Bowman, 2013). Successive mitotic divisions that are intercalated between the fertilization and meiosis allowed the increase in size and complexity of the sporophyte diploid body plan that is able to produce spores. The mature sporophyte in bryophytes is comprised by three major parts: sporangium, seta and foot. The production of spores via meiosis occurs in the sporangium which is attached to a seta that serves as a stalk to the sporangium and a foot which holds the sporophyte to the gametophyte.

Several genetic regulators of sporophyte development in flowering plants have been characterised in moss which regulate the zygote outgrowth, first zygotic cell division, subsequent mitotic divisions and spore production. After fertilization, the zygote expands until it fills the archegonial cavity before undergoing the first asymmetric cell division. The homeodomain TFs belonging to the WOX subfamily are required for zygote expansion prior to the first asymmetric cell division in *P. patens* (Sakakibara et al., 2014). Pp*wox* loss-of-function mutant zygotes resemble unfertilized egg cells, even with fertilization taking place, and are therefore unable to develop sporophytes (Sakakibara et al., 2014). Consistently, WOX TFs in *P. patens* upregulate the expression of cell wall loosening genes, which are required for cell outgrowth (Sakakibara et al., 2014). WOX TFs have been shown to control meristem and early stages of embryo development in the angiosperm *A. thaliana* (Breuninger et al., 2008; Haecker et al., 2004). Since WOX TFs control embryo development in bryophytes and angiosperms, it is plausible that WOX controlled the first embryonic cell division in the first land plants. This suggests that WOX TFs might have contributed for the origin of the first embryo and establishment of multicellular sporophytes during the early evolution of land plants.

The ability of the diploid zygote undergoing a first cell division was a key innovation in the evolution of the land plants, which preceded the subsequent evolution of the complex sporophyte body plan that undergoes cell division and differentiation (Graham et al., 2000). After the first zygotic cell division a basal and apical cell are formed and the latter undergoes multiple mitotic divisions (Tanahashi et al., 2005). The TFs belonging to the LEAFY (LFY) family of TFs in the moss *P. patens* are necessary for the first zygotic cell division (Tanahashi et al., 2005). Plants that lack functional *LFY* genes develop a zygote that is arrested in the single cell stage; therefore the sporophyte body plan development does not occur (Tanahashi et al., 2005). LFY TFs also control the sporophyte development in angiosperms (Busch et al., 1999), suggesting that LFY TFs are conserved regulators of sporophyte development in land plants and controlled the development of the multicellular sporophytes of the fist land plants.

The Homeodomain KNOX1 TFs are important regulators of shoot apical meristem initiation and maintenance in angiosperms (Hay and Tsiantis, 2006, 2009; Scofield and Murray, 2006; Tsuda and Hake, 2015). The discovery that heterologous expression of a moss KNOX1 gene in the angiosperms *Arabidopsis thaliana* is sufficient to phenocopy ectopic expression of *A. thaliana* KNOX1 genes gave rise to the hypothesis that a sporophytic function of KNOX1 in land plants was conserved (Sakakibara et al., 2008). Indeed, *P. patens* plants that lack KNOX1 gene expression develop an abnormal sporophytic body (Sakakibara et al., 2008). The number of cells in the developing sporophyte is reduced and the pattern of cell division is also impaired when compared to wild type plants resulting in smaller sporangium and seta (Sakakibara et al., 2008). Nevertheless, meiosis still occurs and spores are produced, despite being in a few numbers and lacking a normal surface ornamentation (Sakakibara et al., 2008). These observations indicate that in the moss *P. patens*, *KNOX1* are not necessary for sporophyte stem cell initiation or maintenance. Instead, these TFs are required for cell proliferation in the sporophyte and are conserved regulators of cell proliferation in the sporophyte of land plants (Tsuda and Hake, 2015).

1.3.3. Meristem/stem cell development

Stem cells have the potential to self-renew and differentiate to give rise to the body plan in both distinct morphological generations in plants. During land plant evolution, plants have acquired different stem cells with the ability to give rise to different parts of the body plan which accounted for its increasing complexity (Graham et al., 2000). In bryophytes, the meristem is comprised by a single stem cell; and several meristems are distributed in different growing regions of the developing plant, in both gametophyte and sporophyte generations; e.g. sporophyte apical cell, antheridium apical cell, archegonium apical cell, etc (Kofuji and Hasebe, 2014).

Differentiated cells have the potential to become stem cells, which again can give rise to different tissues. In the moss *P. patens*, induction of stem cells with the ability to develop tip-growing filaments occurs when the leaf tissue is detached from the gametophore (Sakakibara et al., 2014). This dedifferentiation process is mediated by WOX TFs (Sakakibara et al., 2014). Moreover, the development of the first apical cell from a protoplast and protonema apical cells is defective in the Pp*wox* loss-of-function mutants (Sakakibara et al., 2014). These defective phenotypes are developmentally similar to the sporophytic defective phenotype of Pp*wox* loss-of-function mutants described above and suggest that WOX TFs are positive regulators of stem cell development and proliferation in the *P. patens*. As mentioned above, WOX TFs regulate meristem identity in angiosperms (Etchells et al., 2013; Gallois et al., 2004; Sarkar et al., 2007). These TFs are involved in stem cell development and proliferation in bryophytes and angiosperms; hence, they have been conserved since the first land plants to regulate one of the major innovations that accompanied land plant evolution – the origin of multicellular meristems.

The transition from two-dimensional to three-dimensional growth is mediated by stem cells with three cutting edges (Graham et al., 2000; Harrison et al., 2009). In the moss *P. patens*, this transition is mediated by the establishment of a gametophore stem cell in the protonema. The gametophore stem cell undergoes an asymmetric cell division and after successive cell divisions it develops into the gametophore bud that will give rise to the three-dimensional structure that will produce gametes (Harrison et al., 2009). The establishment of the gametophore stem cell is mediated by hormonal signaling (Ashton and Cove, 1977; Prigge et al., 2010; von Schwartzenberg et al., 2007). Auxin and cytokinin are positive regulators of gametophore bud development in *P. patens* (Aoyama et al., 2012; Prigge et al., 2010). The APETALA2 (AP2)-type TFs in *P. patens* are required and sufficient for the two-dimensional to three-dimensional growth transition, i.e. gametophore stem cell establishment (Aoyama et al., 2012). Moreover, these TFs are required for the auxin and cytokinin-mediated induction of gametophore stem cells (Aoyama et al., 2012). The transcription of *AP2-type* genes is positively regulated by auxin and the cytokinin-dependent regulation of gametophore bud development requires AP2-type TFs function (Aoyama et al., 2012). Interestingly, AP2-type of TFs act downstream of auxin and cytokinin in the angiosperm *A. thaliana* to control stem cell programming (Dello Ioio et al., 2008; Mähönen et al., 2014). Therefore, since the auxin-cytokinin-AP2-type TFs module is conserved between bryophytes and angiosperms, it is likely that this module functioned early in land plant evolution to regulate stem cell development and, subsequently, the formation of 3D tissues.

1.3.4. Branching

The origin of multicellular meristems during land plant evolution increased radically plant morphological variety, allowing diversification of the body plan and plant architecture, such as the development of branched aerial axes (Kenrick and Crane, 1997; Sussex and Kerk, 2001). The ability to branch in the aerial plant axes (shoots) contributed largely for the increase of photosynthetic tissue and production of reproductive structures (Kenrick and Crane, 1997). Shoot branching evolved independently in the sporophyte of flowering plants and in the gametophyte of bryophytes. The evolution of branched sporophyte with multiple sporangia occurred after the divergence of bryophytes and tracheophytes (Kenrick and Crane, 1997; Wickett et al., 2014). Three hormonal cues and TFs are involved in the regulation of branching in the sporophyte of flowering plants and they have been shown to also control the branching in bryophytes.

Auxin, cytokinin and strigolactone are three hormones that regulate branching in the shoot of flowering plants (Brewer et al., 2009; Crawford et al., 2010; Ferguson and Beveridge, 2009; Ongaro and Leyser, 2008). It has been demonstrated that these three hormones also regulate gametophore branching in the moss *P. patens* (Coudert et al., 2015). The lateral branches are formed from branch stem cells that develop from reprogrammed gametophore epidermal cells (Harrison et al., 2009). While cytokinin positively regulates the development of gametophore branching, auxin and strigolactone repress this developmental program (Coudert et al., 2015). Strigolactone suppresses branching at the base of the gametophore while auxin acts as a global suppressor of gametophore side branching (Coudert et al., 2015). Basipetal and acropetal transport of auxin is required for auxin-regulation of gametophore branching (Coudert et al., 2015). This observation contrasts with the mechanism operating in the sporophytic shoot of flowering plants, where only the basipetal transport of auxin is necessary. This highlights that despite branching had evolved independently in the sporophyte of flowering plants and the gametophyte of bryophytes, the hormonal cues that regulates branching were co-opted during land plant evolution.

One observation that corroborates the co-option of hormonal cues to control both gametophore and sporophyte branching during land plant evolution is the fact that auxin also suppresses sporophyte branching in *P. patens*. Branched sporophytes develop upon treatment with the auxin transport inhibitor 1-N-naphthylphthalamic acid (Bennett et al., 2014). Moreover, loss-of-function mutants of auxin transporters, Pp*pin* mutants, exhibit branched sporophytes (Bennett et al., 2014). This suggests that auxin movement in the sporophyte of *P. patens* is required for the development of the unbranched sporophyte. However, it is still unclear if the genetic mechanism that acts downstream of auxin signaling is similar in both gametophyte and sporophyte branching.

There are several TFs that regulate shoot branching in flowering plants (Aguilar-Martínez et al., 2007). TEOSINTE-BRANCHED1/CYCLOIDEA/PCF (TCP) TFs control axillary bud development and branching in several flowering plants (Aguilar-Martínez et al., 2007; Studer et al., 2011). Two members of this family of TFs are preferentially expressed during sporophyte development in *P. patens* and Pp*tcp5* loss-of-function mutants develop sporophytes bearing two or more capsules that were anchored to the gametophore by the same foot (Ortiz-Ramírez et al., 2016). This means that auxin and TCP TFs are two components of an ancient mechanism that represses sporophyte branching and the evolution of this mechanism might have accounted for the evolution of a branching sporophyte, and plant architecture diversity, after the split between bryophytes and vascular plants. Since TCP TFs do not regulate gametophore branching in the moss (Ortiz-Ramírez et al., 2016), it is possible to infer that indeed the origin and evolution of gametophyte and sporophyte branching involved the recruitment of distinct genetic mechanisms that act downstream of auxin signaling, which is conserved to regulate branching in both generations.

1.3.5. Development of water-conducting and structural support cells

One of the major challenges that plants faced when they colonized the terrestrial surface was the limited water availability, important for their physiological processes and mechanical support. The evolution of water-conducting and structural supporting cells was a key innovation that allowed land plants to colonize and adapt in the arid terrestrial realm (Kenrick and Crane, 1997). Bryophytes develop water-conducting and structural support tissues that differ from the ones found in tracheophytes (vascular plants) (Bateman et al., 1998). Vascular plants develop xylem vessels that confer mechanical support and allow efficient water transport across the plant. Xylem cells undergo secondary cell wall thickening with the deposition of lignin and programmed cell death. Specialized cells with similar function develop in mosses; hydroids which conduct water and stereids which provide structural support. Hydroid cells are elongated cells devoid of cellular content but their secondary cell wall lacks lignin and stereid cells develop thickened cell walls.

NAM, ATAF, and CUC (NAC) TFs have been shown to control the development of xylem vessels in flowering plants, and hydroids and stereid cells in the moss *P. patens* (Xu et al., 2014). In moss, *NAC* genes are expressed in tissues where hydroids and stereids differentiate and loss-of-function mutants develop abnormal thickened walls and cells comprise cellular content, suggesting that NAC TFs are necessary for secondary cell wall thickening and programmed cell death in the moss (Xu et al., 2014). Moreover, NAC TFs positively regulate the expression of genes involved in secondary cell wall synthesis which are shared between the moss *P. patens* and the flowering plant *A. thaliana* (Xu et al., 2014). This suggests that programmed cell death and secondary cell wall thickening mediated by

NAC TFs were highly important developmental processes that allowed the differentiation of cells that carry water-conducting and structural support functions in the early land plants.

1.3.6. Leaf development

The origin of organs, like leaves, that increase the area of photosynthetic tissue, and therefore allowing the plant to produce higher amounts of carbon source, was essential for the plant colonization and adaptation to the terrestrial environment. In the bryophyte lineage, some species produce gametophores bearing phyllids that are analogous in function to the microphylls and megaphylls that develop in lycophytes and euphyllophytes, respectively. The oldest vascular plants found in the fossil record do not develop leaves in their branched stems, suggesting that microphylls and megaphylls evolved independently during land plant evolution (Edwards and Feehan, 1980; Gensel, 2008). Strikingly, the formation of microphylls and megaphylls shares a common regulatory mechanism. KNOX TFs are necessary for the maintenance of the indeterminacy of the shoot apical meristem (Tsuda and Hake, 2015). The initiation of the leaf primordia requires the transition from indeterminate growth to a determinate growth program in the shoot apex (Harrison et al., 2005). This developmental transition requires a negative regulation of KNOX genes, which is mediated by ASYMMETRIC LEAVES1/ROUGHSHEATH2/PHANTASTICA (ARP) proteins (Harrison et al., 2005). The KNOX-ARP mechanism has been shown to be involved in both microphyll and megaphyll development, and it is an example of a co-option of a regulatory mechanism during land plant evolution to control two analogous structures (Harrison et al., 2005).

The phyllids in bryophytes are comprised by a single cell layer of cells (lamina), which increases substantially the area of photosynthetic tissue, providing energy source to the plant. *miR166* and its target *HD-ZIP III*, which encodes a homeodomain TF, constitute an ancient regulatory network in land plants and operate in the phyllids of the moss *P. patens* in order to regulate cell proliferation (Bao et al., 2004; Carlsbecker et al., 2010; Floyd and Bowman, 2004; Yip et al., 2016). Plants that constitutively express the *miR166* develop shorter phyllids with altered shape when compared to wild type plants (Yip et al., 2016). The *miR166 - HD-ZIP III* pathway also acts in megaphylls development in angiosperms but it is involved in organ polarity instead of cell proliferation, as observed in moss phyllids (Emery et al., 2003). This might suggest that the *miR166 -HD-ZIP III* pathway was co-opted to regulate the development of laminar structures in land plants despite the different role in the development of these structures.

The examples of genetic mechanisms underlying the evolution of morphological characters highlighted in the current and previous sections show that most of these mechanisms have been conserved since the early radiation of land plants. One of the open questions relates to the conservation of the genetic mechanism regulating the development of analogous specialized epidermal structures, like epidermal pores (stomata and air pores) that allow gas exchange between the plant and the environment. Moreover, another open question lies on the genetic basis underlying the development of plant epidermis, despite its different ontogenies throughout land plant evolution. In order to address these questions, I aim to use *M. polymorpha* as a model organism to study the function of a genetic component that has been identified to control the development of the epidermis and specialized epidermal cells in angiosperms. *M. polymorpha* is a thalloid liverwort with prominent dorsiventrality, which develops specialized epidermal cells in the upper and lower epidermis of the thallus. IIIf bHLH TFs regulate the development of root hairs, stomata and trichomes in *A. thaliana*. I hypothesize that IIIf bHLH TFs regulate the development of specialized epidermal cells in *M. polymorpha* and, by analogy with their function in angiosperms, I infer that these TFs are ancestral regulators of epidermal cell development since the radiation of the early land plants. The function of IIIf bHLH TFs in angiosperms is discussed in Chapter 3, section 3.2.

1.4. *M. polymorpha* as a model system

Bryophytes (liverworts, mosses and hornworts) are the earliest diverging lineage of land plants (Kenrick and Crane, 1997; Wickett et al., 2014). Therefore, the analogy between the development of bryophytes and other lineages of plants constitute an excellent system to study the evolution of plant development in order to infer the genetic mechanisms that control the development of the first land plants. The liverwort *M. polymorpha* is an emerging model bryophyte species that has a relatively short life cycle and it is easy to grow under laboratory conditions. The analogy between gene function in *M. polymorpha* and in later divergent lineages of land plants allows the inference of the ancestral role of IIIf bHLH TFs in land plants. Moreover, *M. polymorpha* is a thalloid liverwort that develops specialized epidermal cells on the dorsal (air pore complexes) and on the ventral (rhizoids) sides of the thallus, constituting therefore a good model system to study the development and evolution of epidermal cell differentiation. Furthermore, gene function analyses are possible in *M. polymorpha* due to the establishment of transformation protocols and molecular techniques that allow the gene expression manipulation in this liverwort. Research in *M. polymorpha* has been improved substantially in the last years due to the establishment of resources like: efficient *Agrobacterium*-mediated transformation protocols (Ishizaki et al., 2008) and genetic/molecular manipulation tools (Althoff et al., 2014; Ishizaki et al., 2013b; Sugano et al., 2014), which allow reverse and forward genetic studies in *M. polymorpha*. The further sections describe the life cycle of *M. polymorpha* and how auxin regulates various developmental processes, such as the developmental of epidermal cells, in this liverwort.

1.3.1. The life cycle of *M. polymorpha*

M. polymorpha is a complex thalloid liverwort that belongs to the order Marchantiales (Shimamura, 2016). The life cycle of *M. polymorpha* is predominantly haploid (gametophyte) and the diploid generation (sporophyte) is ephemeral and dependent on the haploid generation (Figure 1.5). Both sexual and asexual reproduction occurs in *M. polymorpha*, producing separate male and female plants (dioecious). The beginning of the life cycle occurs with the germination of the haploid spore that carries either an X chromosome in the female, or Y chromosome in the male individual. During spore germination, the spore undergoes a first asymmetric cell division that gives rise to a rhizoid cell (basal cell) and a cell that will further divide (apical cell) to give rise to the photosynthetic tissue (sporeling) (O'Hanlon, 1926). After sporeling development, a thallus is developed with prominent dorsiventrality, with two apical notches where the active sites of growth and bifurcation occur, leading to the branching of the thallus (Binns and Maravolo, 1972). Rhizoids and membranous scales develop on the ventral (lower) epidermis of the thallus, while gemma cups and air chambers comprising air pores are produced on the dorsal (upper) surface of the thallus. The rhizoids in *M. polymorpha* are unicellular tubular projections that develop from rhizoid precursor cells and provide nutrient uptake and anchorage to the substrate (Jones and Dolan, 2012; Proust et al., 2016). Air chamber is a schizogenous intracellular space and it is the gas-exchange tissue between the environment and the plant (Apostolakos et al., 1982). Each air chamber comprises an aperture on the epidermis with four tiers of four cells which form a barrel-shape (4x4 cells) structure that forms the air pore (Apostolakos et al., 1982).

The asexual life cycle of *M. polymorpha* occurs via the development of small propagules, called gemma, that are isogenic individuals and develop from a single cell into a lenticular disc of cells in the cup-like structure on the dorsal epidermis of the thallus, the gemma cup (Voth, 1941) (Figure 1.6). Gemma remains dormant while it is in the gemma cup (Eklund et al., 2015). When the dormancy is disrupted, rhizoids develop on both, lower and upper, surfaces of gemma. After a few days of growth, dorsiventrality is established and rhizoids only develop from the lower epidermis (Binns and Maravolo, 1972). Similarly to the sporeling, the gemma develops into a mature thallus with two apical notches where active growth occurs. Gametophores are modified thallus branches where gametangia are formed at the apex (Durand, 1908). In the male individuals, antheridiophores are gametophores that develop antheridia where motile biflagellate sperm is produced. Archegoniophores are gametophores that develop in the female individuals and contain archegonia where the egg-cell develops. Fertilization is mediated by water and the motile biflagellate sperm swims to the archegonia and fertilizes the egg-cell. After fertilization, the zygote (diploid) undergoes several mitotic divisions and develops into a sporophyte that is dependent on the gametophyte tissue. Meiotic divisions take place in the sporophyte and spores carrying either the X or Y chromosome are produced. These spores are subsequently dispersed and germinate to produce the next generation.

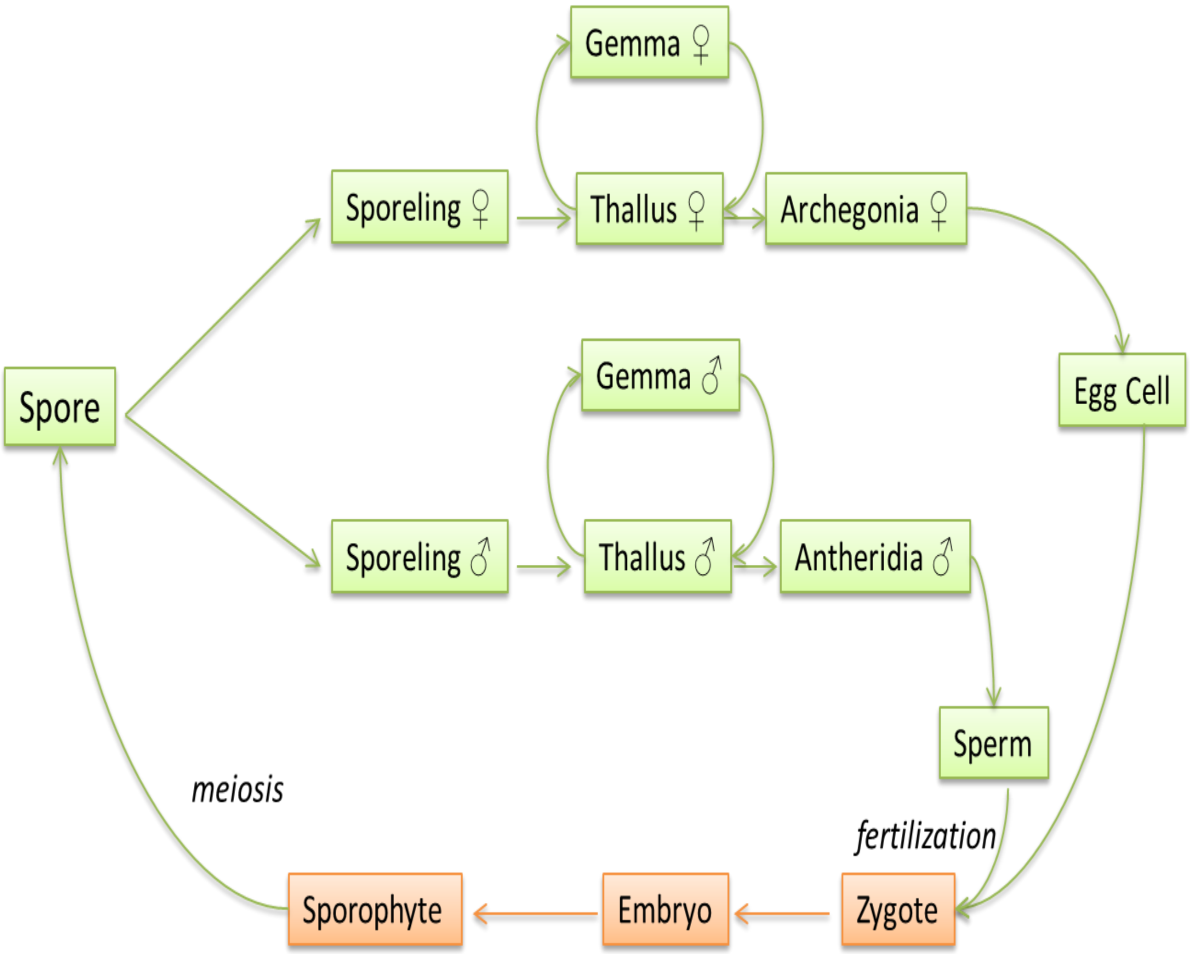


Figure 1.5 – Life cycle of *M. polymorpha*. The gametophyte (haploid) generation is highlighted in green while the sporophyte (diploid) is in orange. For details see Section 1.3.1.

.

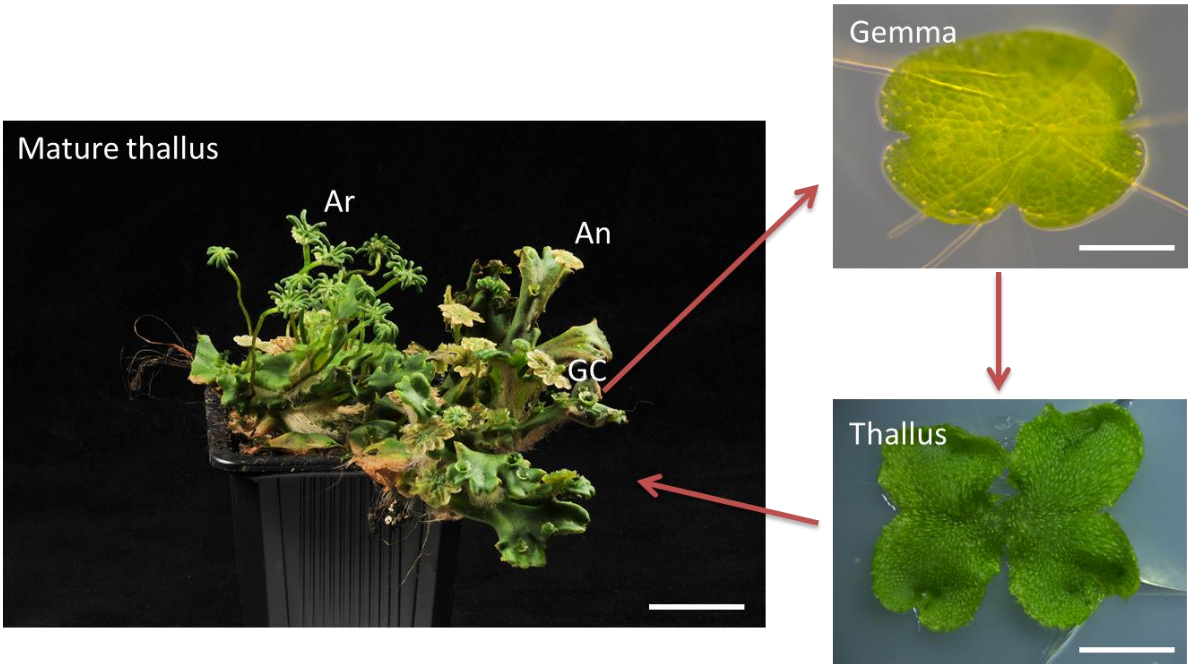


Figure 1.6 – Asexual life cycle of *M. polymorpha*. The mature thallus develops aerial gametophores that produce at the apex female and male gametangia, archegonia (Ar) and antheridia (An), respectively. Gemma cups (GC) produce lenticular discs – gemmae that develop into a thallus. Scale bars are 5 cm (mature thallus), 5mm (thallus) and 500 µm (gemma).

1.3.2. Auxin effect in *M. polymorpha* development

Plant hormones regulate a plethora of morphological and developmental characters in land plants (Wolters and Jürgens, 2009). Auxin (indole-3-acetic acid [IAA]) is one of the most important plant hormones and has been extensively studied throughout land plant phylogeny (Benjamins and Scheres, 2008). With the advent of *M. polymorpha* has a model species to study developmental biology, recent studies have shed light on the auxin biosynthesis and signalling pathways in the liverwort lineage. The effects of exogenous auxin treatments on *M. polymorpha* development vary on the concentration applied. Low concentration of auxin increases thallus area, disrupts gemma dormancy and produces ectopic meristems in dormant gemma, disruption of dorsiventral polarity, tissue patterning (disruption of air pores and air chambers shape) and branching (Davidonis and Munroe, 1972; Halbsguth and Kohlenbach, 1953; Kaul et al., 1962; Tarén, 1958). Application of higher concentration of auxin is sufficient to promote the formation of rhizoids on the dorsal surface of the thallus that resembles a callus-like tissue with no discernible dorsiventrality (Ishizaki et al., 2012). The genetic basis of one of the auxin effects on *M. polymorpha* development has been demonstrated by studying the role of auxin (IAA) biosynthesis and distribution. The biosynthesis of auxin in *M. polymorpha* is mainly active in the meristematic notches of the thallus and in developing gemma cups (Eklund et al., 2015). Auxin biosynthetic pathway components in *M. polymorpha* are necessary to maintain gemma dormancy in the gemma cups by maintaining appropriate levels of IAA in the thallus (Eklund et al., 2015). This observation is consistent with the auxin distribution in *M. polymorpha*, that can be visualized by using the *GRETCHEN HAGEN 3* (*GH3*) promoter, $_{pro}$*GmGH3*, which is active in sites of auxin-mediated transcriptional activation (Bierfreund et al., 2003). In *M. polymorpha*, $_{pro}$*GmGH3* is active in gemma cup, gametophyte-sporophyte junction and sporophyte tissue (Ishizaki et al., 2012). Altogether, these evidences suggest that IAA accumulates preferentially in developing gemma cups, promoting auxin-mediated transcriptional activity that is necessary to maintain gemma dormancy in the gemma cups.

The auxin-mediated transcriptional mechanism is conserved amongst land plants. *M. polymorpha* contains a minimal set of components necessary for the auxin-mediated transcriptional machinery. There is a single TIR1/AFB (MpTIR1), a single AUX/IAA (MpIAA), and three ARFs (MpARF1, MpARF2, and MpARF3) proteins encoded in the genome of *M. polymorpha* (Flores-Sandoval et al., 2015; Kato et al., 2015). MpIAA represses the auxin-mediated transcriptional pathway by inhibiting the action of MpARFs (Flores-Sandoval et al., 2015; Kato et al., 2015). Also, MpTPL co-repress the auxin-mediated transcriptional pathway by repressing MpARFs dependent-transcription regulation (Flores-Sandoval et al., 2015). At high auxin concentration, MpTIR1 targets MpIAA for ubiquitin-mediated degradation, resulting in the de-repression of MpARFs and therefore activation of the downstream auxin-mediated transcription that control many developmental processes in *M. polymorpha* (Flores-Sandoval et al., 2015; Kato et al., 2015). *M. polymorpha* mutants of the components of the auxin signalling pathway are impaired in dorsiventrality establishment, gemma cups, rhizoids and dorsal epidermal cell development, branching and thallus growth (Flores-Sandoval et al., 2015; Kato et al., 2015)

1.5. Approach and aims of the thesis

In this thesis I aim to identify evolutionary trends that occurred during land plant evolution and the genetic mechanisms that underpin the origin and evolution of the tissue that constitute the boundary between the plant and the environment – the epidermis. Respectively, I investigate the evolution of key regulators of plant development – TFs; and I infer the ancestral function of one TF that is known to regulate epidermal cell development in angiosperms. To this extent, I undertake a combined approach that includes: 1) phylogenetic analyses of plant TFs families, with a particular focus on the bHLH TF family, and; 2) functional characterization of a putative conserved bHLH TF subfamily involved in epidermal cell development in land plants, using *M. polymorpha* as a model bryophyte.

Chapter 2 includes the phylogenetic analyses of plant TFs families, where the origin and evolution of one of the largest families of TFs, the bHLH, is considered. In Chapter 3, I carry out a reverse genetic approach, using the model bryophyte *M. polymorpha*, aiming to identify the ancestral function of a bHLH subfamily – IIIf bHLH. This subfamily of bHLH TFs has only been functionally characterized in angiosperms, where it regulates the development of epidermal cells. Thus, the comparison between the previously reported function of IIIf bHLH in angiosperms with the function of IIIf bHLH in the liverwort *M. polymorpha*, enables me to infer whether this bHLH subfamily has been involved in regulating the development of the epidermis of the first land plants, or if this regulation evolved later in the evolution of land plants, i.e. after liverworts and angiosperms shared a common ancestor. Therefore, this analysis is crucial to assess if the genetic mechanism that regulates epidermal cell development in angiosperms is conserved in plants since the early colonization of the land or if it was recruited to control distinct developmental mechanism during the radiation of land plants.

Chapter 2: Evolution of Plant bHLH TFs

2.1. Abstract

The colonization of the land by streptophytes and their subsequent radiation is a major event in Earth history. I report a stepwise increase in the number of transcription factor (TF) families and basic Helix-Loop-Helix (bHLH) subfamily in Archaeplastida before the colonization of the land. The subsequent increase in TF number on land was through duplication within existing TF families and subfamilies. Almost all bHLH subfamilies had evolved before the radiation of extant land plant lineages from a common ancestor. I demonstrate that the evolution of bHLH TFs independently followed similar trends in both plants and metazoans; almost all extant bHLH subfamilies were present in the first land plants and in the last common ancestor of bilaterians. These findings reveal that the majority of innovation in plant and metazoan TF families occurred in the Precambrian before the Phanerozoic radiation of land plants and metazoans.

2.2. Introduction

TFs play a key role in the regulation of developmental mechanisms in eukaryotes. The basic Helix-Loop-Helix (bHLH) TFs are one of the major families of TFs in both plant and metazoan lineages (Carretero-Paulet et al., 2010; Gyoja, 2014; Pires and Dolan, 2010; Simionato et al., 2007). In metazoans, bHLH TFs control a plethora of developmental processes, including myogenic initiation, terminal neuron differentiation, primitive hematopoiesis and cholesterol metabolism (Simionato et al., 2007). Plant bHLH TFs play major roles in a plethora of developmental processes such as vascular tissue establishment, development of cells with a rooting function, response to environmental cues and modulation of secondary metabolism pathways (Bai et al., 2012; Broun, 2005; MacAlister et al., 2007; Menand et al., 2007; Pires et al., 2013; De Rybel et al., 2013; Yi et al., 2010). Moreover, some of these processes, such as the development of cells with a rooting function, have been conserved during land plant evolution and bHLH proteins have been shown to be ancient regulators of these developmental processes (Breuninger et al., 2016; Menand et al., 2007; Pires et al., 2013; Proust et al., 2016; Tam et al., 2015).

All plant bHLH proteins are monophyletic and can be divided into distinct subfamilies that are defined not only by their phylogenetic relationship but also by the presence of subfamily-specific motifs (Carretero-Paulet et al., 2010; Pires and Dolan, 2010). The bHLH proteins comprise a 50-60 highly conserved amino acid domain that defines this family of TFs (Long et al., 2010). The bHLH domain can be divided in two distinct motifs: the basic region and two amphipathic α-helices separated by a loop. The basic region is enriched in basic amino acids and binds to the hexanucleotide E-box nucleotide motif CANNTG, where N corresponds to any nucleotide (Buck and Atchley, 2003; Toledo-Ortiz et al., 2003). The two amphipathic α-helices are constituted by hydrophobic amino acids and promote the formation of homo- and heterodimers with other bHLH proteins (Buck and Atchley, 2003).

Previous phylogenetic analyses characterized the distribution of TF families, and concretely the diversification of bHLHs, in land plants (Carretero-Paulet et al., 2010; Pires and Dolan, 2010). However, due to the lack of genome sequencing in key streptophyte taxa, these studies were unable to infer the origin and evolution of bHLH TFs in the streptophytic aquatic ancestor of land plants. The recent availability of the genome sequence of the aquatic streptophyte alga (belonging to the clade that is sister to the land plants), *Klebsormidium flaccidum,* and the liverwort, one of the earliest divergent lineages of land plants, *Marchantia polymorpha* allowed me to track the change in the number and size of TF families and bHLH subfamilies during streptophyte evolution. Specifically, the addition of these two species in the phylogenetic analyses is essential to trace the origin and evolution of TF families and bHLH subfamilies in the streptophytic aquatic ancestor of land plants and in the first land plants.

In this Chapter, I demonstrate that the origin of the majority of TF families and bHLH subfamilies predate the radiation of extant land plant lineages. The stepwise origin of the majority of plant TFs and bHLH subfamilies occurred during the Precambrian before the radiation of the extant lineages of land plants. Moreover, metazoan bHLH evolution follows similar trends of those observed in plants, suggesting that bHLH diversification in the Precambrian preceded two major radiations in organismal diversity in distant branches of the tree of life.

2.3. Materials and methods

2.3.1. Plant TF family analysis

Protein sequences from *Klebsormidium flaccidum* (Hori et al. 2014) (http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/), *Marchantia polymorpha* (NCBI GenBank accession LVLJ00000000.1), *Cyanidioschyzon merolae* (Matsuzaki et al., 2004) (http://merolae.biol.s.u-tokyo.ac.jp/), *Porphyridium purpureum* (Bhattacharya et al., 2013), *Chondrus crispus* (Collén et al., 2013) and *Chlorella variabilis* (Blanc et al., 2010) were obtained. Protein sequences from *Ostreococcus tauri* (Derelle et al., 2006; Palenik et al., 2007) were obtained from the JGI Genome Portal (Grigoriev et al., 2012). Protein sequences from the remaining species were obtained from Phytozome (Goodstein et al., 2011). TFs (TFs) were identified using the same rules for required and forbidden PFAM domains (Finn et al., 2016) as has been previously described (Jin et al., 2014) (Supplementary File 1). The origin of each TF family was mapped to the node in the tree that provides the most parsimonious explanation of their distribution in extant species. The SAP family of TFs was excluded due to the existence within it of promiscuous domains (WD40-like and F-box) which prevented its origin being established with certainty.

2.3.2. Basic Helix-Loop-Helix (bHLH) phylogenetic analysis

In order to search the *K. flaccidum* (Hori et al., 2014) and *M. polymorpha* (NCBI GenBank accession LVLJ00000000.1) genomes and proteomes a blast query was assembled from the previously characterized bHLH (Pires and Dolan, 2010) protein sequences in *At, Arabidopsis thaliana*; *Os, Oryza sativa*; *Sm, Selaginella moellendorffii*; *Pp, Physcomitrella patens*; *Cr, Chlamydomonas reinhardtii*; *Ot, Ostreococcus tauri*; *Vc, Volvox carteri; Cm, Cyanodioschyzon merolae*. In total this search resulted in 540 bHLH protein sequences, which were aligned using MAFFT (Katoh and Frith, 2012; Katoh and Standley, 2013) and manually trimmed in Bioedit (Hall, 1999) in order to use only the conserved bHLH domain for the Blast search. The *K. flaccidum* and *M. polymorpha* proteomes were interrogated using a BlastP local blast search using an e-value cut-off of 0.1. This low stringency search ensured that all possible bHLH sequences in the proteome dataset were investigated and evaluated. The results of the blast search were analyzed manually to ensure the presence of the bHLH conserved domain using SMART (http://smart.embl-heidelberg.de/) and Pfam (http://pfam.sanger.ac.uk/search) (Supplementary File 2; Supplementary File 3).

All protein sequences (Supplementary File 4) were aligned using MAFFT (Katoh and Frith, 2012; Katoh and Standley, 2013) and further manually aligned in Bioedit (Hall, 1999). The resulting alignment was used for subsequent phylogenetic analysis (Supplementary File 5). The Maximum likelihood analysis was carried using PhyML 3.0 (Guindon et al., 2010), using Jones, Taylor and Thorton (JTT) amino acid substitution model and a predicted gamma distribution. Branch support was tested using a Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-like aLRT). The generated unrooted trees were visualized using MEGA6 (Tamura et al., 2013).

2.3.3. Detection of conserved motifs

To identify amino acid motifs that are specific for each bHLH subfamily, the amino acid sequence of all proteins from each subfamily were individually analyzed using MEME (Multiple Em for Motif Elicitation) (Bailey and Elkan, 1994). The presence of the subfamily motifs were further verified on the alignment of each individual subfamily and a consensus sequence for each motif was obtained.

2.4. Results

2.4.1. 47 of the 48 land plant TF families evolved before the colonization of the land

Forty eight (Supplementary File 1) TF families were identified in the genomes of 15 species of Archaeplastida (the lineage that includes red algae, chlorophytes and streptophytes) (Adl et al., 2005; Archibald, 2009). These taxa comprise the red algae *Cyanidioschyzon merolae*; *Porphyridium purpureum*, *Chondrus crispus* the chlorophytes *Ostreococcus tauri*, *Micromonas pusilla*, *Chlorella variabilis*, *Coccomyxa subellipsoidea*, *Volvox carteri* and *Chlamydomonas reinhardtii*; the streptophyte alga *Klebsormidium flaccidum* and the streptophyte land plants *Marchantia polymorpha*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Oryza sativa* and *Arabidopsis thaliana* (Supplementary File 1). TF genes were defined by the presence of DNA-binding domains and categorisation of TFs was on the basis of their domain architecture and sequence similarity (Finn et al., 2016; Jin et al., 2014). The number of TF families that are present at each internal node in the phylogeny was inferred using parsimony. This revealed that the number of TF families increased progressively from node 1 to node 6 (Figure 2.1). There were 17 TF families in the common ancestor of all species in this analysis (node 1 in Figure 2.1) and 28 in the common ancestor of chlorophytes and streptophytes (node 2 in Figure 2.1). 39 TF families were present in the aquatic common ancestor of *K. flaccidum* and land plants (node 3 in Figure 2.1) and 47 TF

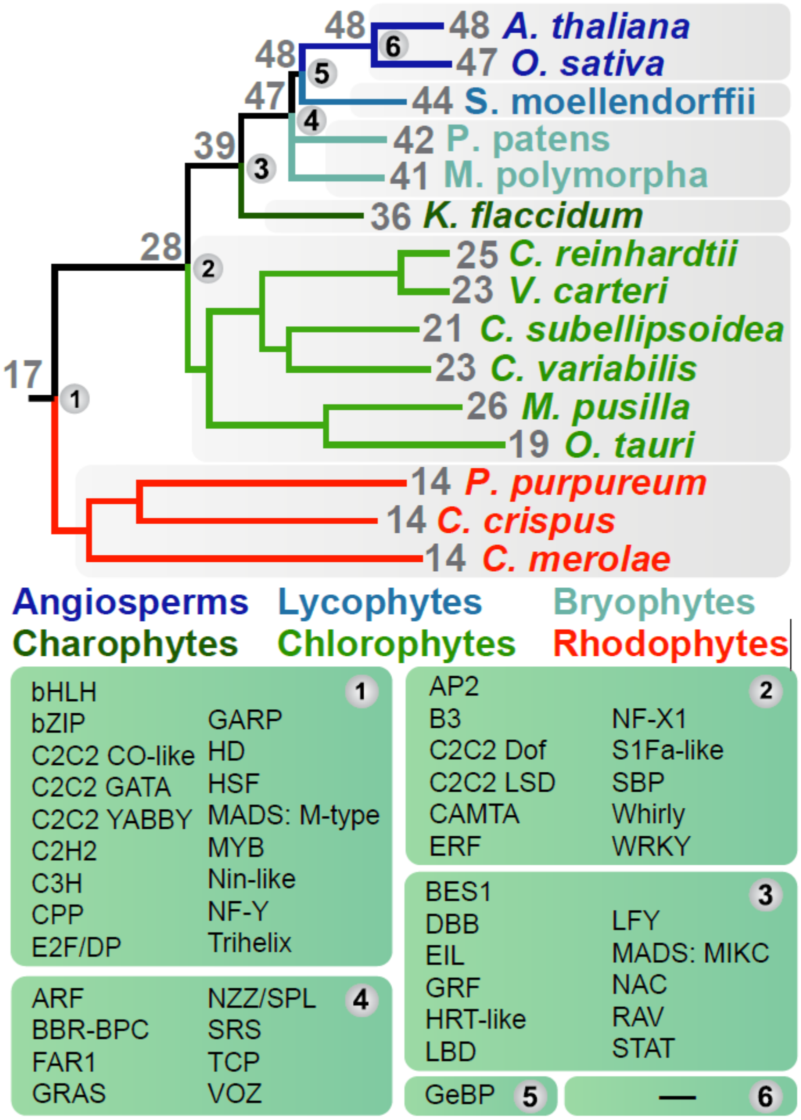


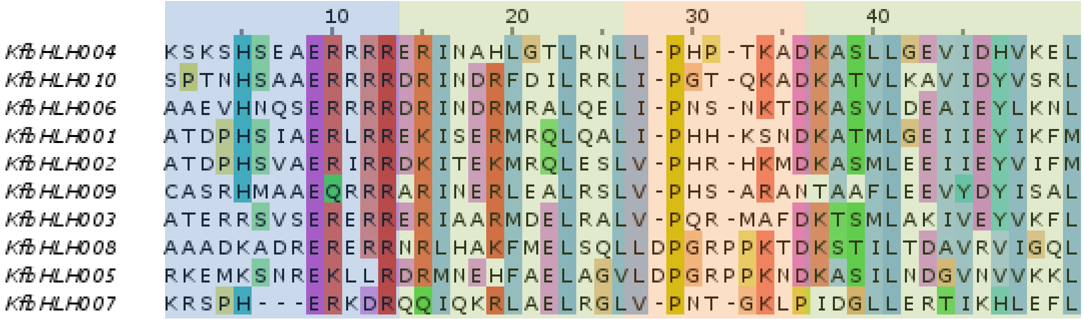
Figure 2.1 Distribution of TF families in plants. Cladogram of Archaeplastida phylogeny (based on Finet et al. 2010, Cox et al. 2014, Wickett et al. 2014) with the origin of the TF families (within green boxes) and total number of TF families (grey) present at each ancestral node (circled numbers) and in extant species (coloured according to clade). The number of TF families increases in later divergent Archaoplastida lineages.

families had evolved in the first land plants (node 4 in Figure 2.1). The stepwise origin of TF families in the aquatic ancestors of land plants – from 17 TFs in the last common ancestor of all species at node 1 to 47 in the first land plants – contrasts strikingly with the subsequent evolution of TF families on land. During the radiation of plants in the terrestrial realm TF number increased from 47 to 48; only one TF family, GeBP (node 5 in Figure 2.1), evolved after the divergence of bryophytes and vascular plants (Figure 2.1).This indicates that there was a stepwise increase in the number of TF families in Archaeplastida before the radiation of land plants and since that time there has been relatively little change in TF family number (Figure 2.1).

2.4.2. *Klebsormidium flaccidum* encodes 10 bHLH proteins

K. flaccidum is a key taxa for phylogenetic analysis of gene families, being one representative of the lineage that is sister to all land plants, the charophyte algae. The genome of *K. flaccidum* is the only genome of a streptophyte alga that has been sequenced so far. It encodes many proteins, including signaling cascade proteins and hormone biosynthesis enzymes, which are found in land plant species (Figure 2.1). In order to retrieve all bHLH sequences present in the proteome of *K. flaccidum,* a BlastP local blast search using all previously identified bHLH sequences from *Arabidopsis thaliana, Oryza sativa, Selaginella moellendorffii*, *Physcomitrella patens, Chlamydomonas reinhardtii*, *Ostreococcus tauri, Volvox carteri* and *Cyanodioschyzon merolae* was performed in the proteome of *K.flaccidum*. This search resulted in 10 sequences (Supplementary File 2) that were further validated for the presence of the bHLH domain using PFAM and alignment analysis (Figure 2.2).

The number of bHLH proteins encoded in the genome of *K. flaccidum* is higher than the number of bHLH proteins encoded in the genome of all examined chlorophytes species so far. However, it contains fewer bHLH proteins than land plants (Pires and Dolan, 2010). This suggests that the number of plant bHLH proteins has increased after the divergence of



KfbHLH004	KSKSHSEAERRRRERINAHLGTLRNLL-PHP-TKADKASLLGEVIDHVKEL
KfbHLH010	SPTNHSAAERRRRDRINDRFDILRRLI-PGT-QKADKATVLKAVIDYVSRL
KfbHLH006	AAEVHNQSERRRRDRINDRMRALQELI-PNS-NKTDKASVLDEAIEYLKNL
KfbHLH001	ATDPHSIAERLRREKISERMRQLQALI-PHH-KSNDKATMLGEIIEYIKFM
KfbHLH002	ATDPHSVAERIRRDKITEKMRQLESLV-PHR-HKMDKASMLEEIIEYVIFM
KfbHLH009	CASRHMAAEQRRRARINERLEALRSLV-PHS-ARANTAAFLEEVYDYISAL
KfbHLH003	ATERRSVSERERRERIAARMDELRALV-PQR-MAFDKTSMLAKIVEYVKFL
KfbHLH008	AAADKADREREERRNRLHAKFMELSQLLDPGRPPKTDKSTILTDAVRVIGQL
KfbHLH005	RKEMKSNREKLLRDRMNEHFAELAGVLDPGRPPKNDKASILNDGVNVVKKL
KfbHLH007	KRSPH---ERKDRQQIQKRLAELRGLV-PNT-GKLPIDGLLERTIKHLEFL

Figure 2.2 Alignment of the bHLH domain of the 10 *K. flaccidum* bHLH proteins. The basic, helix and loop regions are highlighted with blue, green and orange shaded boxes, respectively.

chlorophyte and streptophyte algae, and subsequently after the divergence of streptophyte algae and land plants.

2.4.3. *Marchantia polymorpha* encodes 49 bHLH proteins

The livewort *M. polymorpha* is a representative of the earliest divergent lineage of extant land plants, the bryophytes. The only bryophyte that has been studied regarding the presence and distribution of bHLH TFs in its genome is the moss *P. patens* (Pires and Dolan, 2010). Since the phylogeny of bryophytes is unresolved, the inclusion of a liverwort in the phylogenetic analysis of the evolution of Archaeplastida bHLH proteins increases the robustness of this study, allowing to infer with more certainty what was the bHLH content of the first land plants. Recent studies in *M. polymorpha* have revealed that, commonly, the number of genes in each gene family is low when compared to other land plant species (Flores-Sandoval et al., 2015; Inoue et al., 2016). To determine the total number of bHLH sequences present in the proteome of *M. polymorpha*, a BlastP local blast search using all previously identified bHLH sequences from *Arabidopsis thaliana, Oryza sativa, Selaginella moellendorffii*, *Physcomitrella patens, K. flaccidum, Chlamydomonas reinhardtii*, *Ostreococcus tauri*, *Volvox carteri* and *Cyanodioschyzon merolae* was performed in the proteome of *M. polymorpha*. This search resulted in 49 sequences (Supplementary File 3) that were further validated for the presence of the bHLH domain using PFAM and alignment analysis (Figure 2.3).

The genome of *M. polymorpha* encodes fewer bHLH proteins than other land plants that have been sequenced. There are 158 and 173 bHLH proteins in the genomes of the angiosperms *A. thaliana* and *O. sativa,* respectively, while there are 103 and 98 in the lycophyte *S. moellendorffii* and the moss *P. patens,* respectively. This suggests that the number of bHLH proteins increased during streptophyte evolution and that this increase accompanied the radiation of plants on land.

```
MpbHLH001  TVSKNLVSER-KRRKKLNEGLYTLRALV-PRISKMDKASIIGDAIDYVREL
MpbHLH048  TVSKNLVSER-KRRKKLNDGLYTLRALV-PKISKMDKASIVGDAIDYVREL
MpbHLH011  HLSKNLVAER-KRRKKLNERLYALRALV-PKITKMDRASILGDAIEYVREL
MpbHLH029  CSSRNLFSER-KRRKKMNDGLYSLRSIV-PNISKMDKASIIGDAIKYVMEL
MpbHLH016  LPAKNLMAER-RRRKKLNDRLYMLRSVV-PKITKMDRASILGDAIEYLKEL
MpbHLH017  LPAKNLMAER-RRRKKLNDRLYMLRSVV-PKITKMDRASILGDAIDYLRDL
MpbHLH026  EPLNHVQAER-QRREKLNQRFYALRSVV-PNVSKG------------FQDL
MpbHLH046  EPLNHVQAER-QRREKLNQRFYALRSVV-PNVSKMDKASLLGDAIAYIQEL
MpbHLH032  HTQDHIMAER-KRREKLSQRFIALSAIV-PGLKKMDKASVLGDAIKYVKQL
MpbHLH041  HTQDHIMAER-KRREKLSQRFIALSAIV-PGLKKMDKASVLGDAIKYVKSL
MpbHLH002  AAVSHMLAER-RRRIKQNENFSQLKALI-PSATKMDKATVLAETIVYINSL
MpbHLH004  AAINHMLAER-KRREKQKENFSALRALI-PFVSKTDRASILGDAIQYVKQL
MpbHLH003  QARTHMLAER-RRRVRENEHYDALRKLI-PPTSRMDKATVLAETIKLVKKL
MpbHLH012  LNVSHVLAER-RRREKLNDRFMSLRALV-PFVTKMDKASILGDAIEYVKQL
MpbHLH034  QRENHILAER-QRREEMNEKFSALRAMI-PKATKKDKASIVGDTIAYVLEL
MpbHLH005  GHLEHVQRER-QRRDDMSAKIATLETLL-PPASKRDRVSIVHDGIHYVRAL
MpbHLH023  DQVEHVIRER-QRRDDMSYKIATLERLL-PQGPKRDRASIVHDSVQYVKSL
MpbHLH047  EQVEHVLRER-QRRDDMSCKIAILEGLL-PPCPKRDRASIVHDSVQYVKSL
MpbHLH006  AAEVHNLSER-RRRDRINEKMKALQELI-PNSNKTDKASMLDEAIEYLKML
MpbHLH020  SIEPQSVAAR-HRRKKISERVRVLEKLI-PGGNKMDTASMLDEAIEYVKFL
MpbHLH039  PSRSLKVAESVEQDQDLESRFVALQRLV-PGGSNMDTSSLLDETADFVMFL
MpbHLH040  SKDPQSVAAR-HRRERISDRIRVLQRLV-PGGTKMDTASMLDEAISYVKYL
MpbHLH014  ANDPQSIAAR-HRRERISERLKTLQDLV-PNGSKVDLVTMLEKAINYVKFL
MpbHLH028  ATDPQSVYAR-HRRERITERLKILQHLV-PNGAKVDIVTMLEEAISYVKFL
MpbHLH022  YTNPRSIIEK-TRRQRITEKLRELQELV-HSVKPDNTVYMLDQVINYVEGL
MpbHLH025  ATNPRSIAER-LRRIRITERLRKLQALL-PHTHNKDTAYMLDEIIDYIEYL
MpbHLH027  ATHPRSIAER-VRRTRISERMRKLQELV-PNMDKQNTADMLDEAVEYVKQL
MpbHLH043  ATDPHSIAER-LRRERIAERMKALQELV-PNSNKTDKASMLDEIIEYVKFL
MpbHLH015  ATDSHSLAER-VRREKISERMKFLQDLV-PGCSKVGKAVMLDEIINYVQSL
MpbHLH021  ATDSHSLAER-VRREKISERMKFLQDLV-PGCSKVGKAVMLDEIINYVQSI
MpbHLH033  ATDPQSIAAR-TRRERFSDRIRILQSLV-PNGERLDTVSMLGHTLEYVRFL
MpbHLH018  AVHPRSIAER-VRRSRIAERVKILEAKL-PYPTKKHTDEILERTVNYIKWL
MpbHLH010  FLDHPHMDER-TRQE--NEKFQVLRAMLQPHTCKGDKATVLGDAIEHMRDL
MpbHLH030  EKAEHVQRER-ERRIQLRAMYKTLDTLV-PATAKRDRTTIVNDALNYVKEM
MpbHLH031  QKSDHMRRER-DRRIALKLMYKTLDSMI-PKKSKRDRTAIVDDALKYVKGM
MpbHLH024  QRESHIWSER-QRRKGMNHLFSTLRSLL-PQTSKADKSTVVSEIIKYIQGL
MpbHLH038  DHEIHIWTER-ERRKKMNGMFSTLHSLL-PHNSKADKSTIVDEAINYIQTL
MpbHLH008  PVSSTFAREK-LRREKLTENYNTLKSML-PKHTKVDRASILQATIARVKLL
MpbHLH009  YEGMTHAQEILCRPKPVLS----------PSFAKADRASILEETIAHVKNL
MpbHLH036  KPLNHFLTWK-KKLEDLNEKYQTLRSLV-ANPTKADRVSILSDVINMIHDL
MpbHLH037  KPLHHFATER-QRREHLNEKYQTLRSLV-PNPTKADRASIVSDAINRIHDL
MpbHLH035  HRNAHIVAER-LRRKQMNEHFSALKDII-PSQTKLDQATILAAAVNLVRDL
MpbHLH042  DSISSGSARK-SRRDKIKKTVKLLRNII-PGGDCMDTAIVLDEAIHYVKLL
MpbHLH045  QTPNVPFKMQ-VRKEKLGERITALQQLV-SPFGKVDTASVLLEAIGYIKFL
MpbHLH007  ASKSHSEAER-RRRERINTHLATLRNNL-PSSTKLDKASLLGEVIDHLKFL
MpbHLH044  PRSKHSATEQ-RRRSKINDRFQMLRELV-PHDQKRDKASFLLEVIEYIQVL
MpbHLH019  PKLGDTLRPRPKDRQQIQDRVRELREIV-PNACKCSIDSLLERTIRHMHFL
MpbHLH013  GAGSKACREK-MRRDRLNDRFVELSSALEPGPPKSDKATILCDAVRILNQL
MpbHLH049  RKVHKADREK-LRRDRLNEQFAELAGVLDPDRPKNDKATILGDSVQVVKEL
```

Figure 2.3 Alignment of the bHLH domain of the 49 *M. polymorpha* bHLH proteins. The basic, helix and loop regions are highlighted with blue, green and orange shaded boxes, respectively.

2.4.4. 26 of the 30 land plant basic Helix-Loop-Helix subfamilies were present in the first land plants

The number of bHLH TFs increased during Archaeplastida evolution. In order to test if this increase is paralleled with the origin of novel TF subfamilies that accompanied the radiation of plants on land, a gene tree analysis was conducted on one of the most representative TF families in plants, the bHLH family (Degnan et al., 2009; Gyoja, 2014; Gyoja and Satoh, 2013; Pires and Dolan, 2010; Sakakibara, 2016; Simionato et al., 2007). The phylogenetic analysis of the gene tree of Archaeplastida bHLH genes resolved 30 streptophyte bHLH subfamilies (Figure 2.4). The classification of plant bHLH subfamilies consisted on the phylogenetic relationships between bHLH proteins and the presence of subfamily specific non-bHLH motifs (Table 2.1). This analysis revealed that there is a single bHLH protein in the red alga, *C. merolae* and 3 bHLH proteins encoded in the genomes of the chlorophytes *C. reinhardtii and V. carteri* (Table 2.2). However, none of the 30 streptophyte bHLH subfamilies were found in either red algae or chlorophytes (Figure 2.4; Supplementary File 6). There are 10 *K. flaccidum* bHLH TFs in 6 subfamilies all of which were previously described in land plants – IVb, IVc,Vb, VII(a+b), XI and XIII. The genome of the liverwort *M. polymorpha* encodes 49 bHLH proteins which comprise 25 of the 30 plant bHLH subfamilies (Figure 2.4, Table 2.2). The *M. polymorpha* sequences, together with the sequences of *P. patens*, *S. moelendorffii*, *O. sativa* and *A. thaliana*, defined 4 new monophyletic bHLH subfamilies that were not identified in previous studies: IVd(2), VIIIc(3), XVI and XVII. 23 of the 30 bHLH subfamilies are monophyletic clades supported by SH-like aLRT values greater than 0.85 (Supplementary File 6). The analysis of the distribution of bHLH subfamilies in the earliest divergent land plants (*M. polymorpha* and *P. patens*) allows to infer that 26 bHLH subfamilies were present in the first land plants. Some of these subfamilies contain bHLHs

Table 2.1 Subfamily specific non-bHLH motifs.

Subfamily	Motif	Consensus sequence
Ia	1	EQLLQSLEAQKRRKAYAD
	2	TVADHTVLYSFSVK
	3	VDEIAAAVVHQIFSII
Ib(1)	None	-
Ib(2)	None	-
II	None	-
III(a+c)	None	-
IIIb	4	IHMFCARRPGL
	5	CFNGFALDVFRAEQC
	6	GPEVRGPEEIKAVLLHTAGCDYHG
III(d+e)	7	EEEQEMRKRVLRELNSLISG
IIIf	8	ELSADELGLQRSEQLRELYESLSAAEDGS
IVa	None	-
IVb	9	RSEVKRLKKEQTSLLDESRDLTQEKNELREEKAALKSEIEELQN
IVc	10	RAEAQKLKESNEQLREKIKDLKAEKNELRDEKLRLKAEKERLEQ
IVd(1)	None	-
IVd(2)	None	-
Va	11	QEKVQKYETSEQGRHQERLKSMPW
	12	IQGGVISLSSVYSQGLLDTLT
Vb	13	PVPTEADELTVD
	14	EDGRLVIKASLCCEDRPDLLPDLIRALKSLRLRTLKAEIATLGGRVRNVLLLTRDE
VII(a+b)	15	DDDLVELAWFGVADPRRKGGNGQVV
	16	LHHHHHDGVLELGSPQNLFIDDDLLTLARAANGGGAAGDRGEMASWLYPIVGCHRGAAPGPHHPLDQQIDDL

VIIIa	None	-
VIIIb	17	DPAAVEDPSSRPPPPNSRPKRRNVRI
VIIIc(1)	18	KVLATDEFWPAQGGKAAPCTALPDISQELLELAGPEFAGIDGKNTEEPPAKPVKDAL
VIIIc(2)	19	ADPIAYNGMDIG
VIIIc(3)	20	SASEVKCEKWKEFIDAQT
IX	None	-
X	21	DLRSRGLCLVPVSCTLQVAND
XI	22	PSQDNNGGGGLASALTERQVAKLMEEDMGSA
	23	MQYLQSKGLCLMPI
XII	24	DYIHVRARRGQ
	25	PGPSSVRLDFGNPDIDGLLCDVSKDM
XIII	26	VTKHADKLKASG
	27	EVGSKSMGCPIIVEDLNQPR
	28	EIADFIRGLGLTILKGVMETSRKDKIWADDRF
XIV	29	MHEDTEEIDALLYSDDD
XV	None	-
XVI	30	ESLGLDVSRCSITKLADRLVECVITAK
XVII	31	AALEECKVEVIQSN

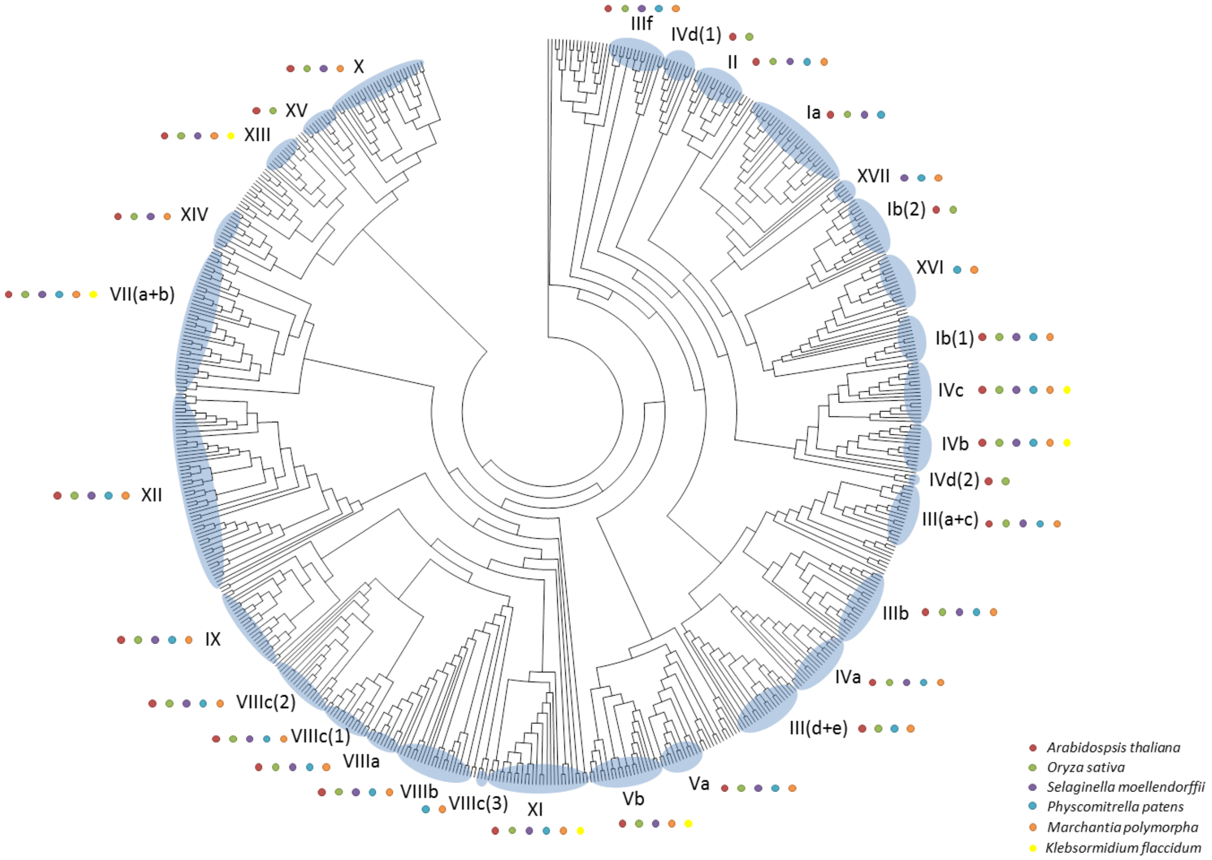


Figure 2.4 Phylogenetic analysis of bHLH Archaeplastida TFs shows that the majority of bHLH subfamilies were present in the common ancestor of extant land plants. Circular cladogram representing the Maximum Likelyhood (ML) analysis of Archaeplastida basic Helix-Loop-Helix TF proteins. The tree is an unrooted ML generated with the software PhyML 3.0, (Guindon *et al.*, 2010) using the JTT model of amino acid substitution, an estimated gamma distribution parameter and a SH-like aLRT test. Subfamiles are grouped with blue ellipses. Coloured dots represent protein presence in each subfamily. *Arabidopsis thaliana* (orange); *Oryza sativa* (green); *Selaginella moellendorffii* (purple); *Physcomitrella patens* (blue)*; Marchantia polymorpha* (orange) and *K. flaccidum* (yellow). See Supplementary File 6 for fully annotated ML tree.

Table 2.2 bHLH TFs distribution in plants.

Subfamily	*At*	*Os*	*Sm*	*Pp*	*Mp*	*Kf*	*Vc*	*Cr*	*Ot*	*Cm*
Ia	10	12	7	2						
Ib(1)	1	2	5	3	2					
Ib(2)	12	5								
II	4	2	3	2	3					
III(a+c)	4	3	2	5	3					
IIIb	4	5	7	3	2					
III(d+e)	8	5		4	2					
IIIf	4	7	2	2	1					
IVa	4	6	2	4	2					
IVb	3	4	1	3	1	1				
IVc	4	4	4	3	1	1				
IVd(1)	1	7								
IVd(2)	1	2								
Va	3	3	2	3	1					
Vb	5	9	4		1	1				
VII(a+b)	14	14	7	4	1	1				
VIIIa	3	1	2	2	1					
VIIIb	5	7	2	6	1					
VIIIc(1)	2	3	6	2	1					
VIIIc(2)	4	7	2	4	1					
VIIIc(3)				2	1					
IX	6	5	4	7	2					
X	9	14	4		1					
XI	5	7	6	2	1	3				
XII	17	17	6	13	2					
XIII	3	3	3		1	1				
XIV	4	3	4		1					
XV	4	6								
XVI				12	3					
XVII			4	1	1					
Orphans	14	10	14	9	12	2	3	3	1	1
Total	158	173	103	98	49	10	3	3	1	1

Note - Numbers of TFs in each bHLH subfamily. *At*, *Arabidopsis thaliana*; *Os, Oryza sativa*; *Sm, Selaginella moellendorffii*; *Pp, Physcomitrella patens*; *Mp, Marchantia polymorpha; Kf, Klebsormidium flaccidum Cr, Chlamydomonas reinhardtii*; *Ot, Ostreococcus tauri*; *Vc, Volvox carteri; Cm, Cyanodioschyzon merolae*.

involved in developmental mechanisms and morphological characters that are characteristic of land plants. The stoma is one morphological character that did not evolve in the lineage that gave rise to the extant liverworts. Interestingly, subfamily Ia is not present in the liverwort *M. polymorpha*. This subfamily contains *A. thaliana* FAMA (At3g24140), MUTE (At3g06120), and SPEECHLESS (At5g53210), which are known regulators of the stepwise development of stomatal guard cells, including asymmetric cell division and cell-type differentiation (MacAlister et al., 2007; Vatén and Bergmann, 2012). The absence of Ia subfamily in liverworts could account for the lack of stomata development in this early divergent lineage of land plants.

The phylogenetic analysis of bHLH gene family indicates that some new bHLH subfamilies evolved while other subfamilies were lost during the course of land plant evolution. Subfamilies VIIIc(3) and XVI are present only in bryophytes (liverworts and mosses) suggesting that subfamilies VIIIc(3) and XVI were either lost before the origin of vascular plants or that they evolved independently in the case that bryophyte constitute a monophyletic clade (Cox et al., 2014; Wickett et al., 2014). There are only non-seed plant proteins in subfamily XVII. The most parsimonious interpretation of this result is that subfamily XVII was lost in the lineage giving rise to the seed plants after the divergence of the lycophytes from the seed plants (Figure 2.4, Table 2.2). Moreover, subfamilies Ib(2), IVd(1), IVd(2) and XV are present in angiosperms only (Figure 2.4, Table 2.2), suggesting that these 4 subfamilies evolved in the lineage leading to the angiosperms after the divergence of lycophytes and seed plants. These data suggest that there were losses and origins of new bHLH subfamilies during the evolution of plants on land. The presence of 6 bHLH subfamilies in *K. flaccidum* compared to 26 bHLH subfamilies in bryophytes (*M. polymorpha* and *P. patens*) suggests that the majority of bHLH subfamilies originated around the time that plants colonised the land, or at very least after the time when *K. flaccidum* and early diverging land plants last shared a common ancestor. In sum, the morphological diversification of land plants was predated by the evolution of the majority of bHLH subfamilies.

2.4.5. Three bHLH subfamilies that regulate periclinal cell division were lost in mosses

The phylogenetic analysis of Archaeplastida bHLH proteins revealed that there were losses and origins of bHLH subfamilies in the course of land plant evolution. Some of these losses appear to be lineage specific (Figure 2.5 and Table 2.1). Interestingly, four subfamilies (Vb, X, XIII and XIV) are absent in the moss *P. patens* but are present in all the other land plants examined (including the liverwort *M. polymorpha*). Moreover, two of these subfamilies (Vb and XIII) are present in the streptophyte alga *K. flaccidum*, indicating that the origin of these subfamilies predated the colonization of the land by plants (Figure 2.5). Three of the subfamilies absent in *P. patens* (Vb, XIII and XIV) form a transcriptional network that controls periclinal cell divisions in the angiosperm *A. thaliana* (Vera-Sirera et al., 2015). These subfamilies include TARGET OF MONOPTEROS 5 (TMO5), LONESOME HIGHWAY (LHW) and SAC51-LIKE (SACL) TFs. These bHLH form heterodimers amongst them and these heterodimers regulate periclinal cell divisions in the hypocotyl and the root of *A. thaliana* (Vera-Sirera et al., 2015). This suggests that an entire transcriptional network was lost in *P. patens*.

I examined the genome of *Sphagnum fallax* (which was not available at the time that the high level phylogenetic analysis of bHLH TF distribution and diversification was performed) to know whether the losses of these subfamilies occurred at a bigger taxonomic level. The genome of *S. fallax* has been recently release (*Sphagnum fallax* v0.5, DOE-JGI, http://phytozome.jgi.doe.gov/) and I have searched the *S. fallax* genome for bHLH proteins belonging Vb, XIII and XIV subfamilies. This search revealed that, similarly to *P. patens*, none of these subfamilies are present in the genome of *S. fallax*. *S. fallax* and *P. patens*

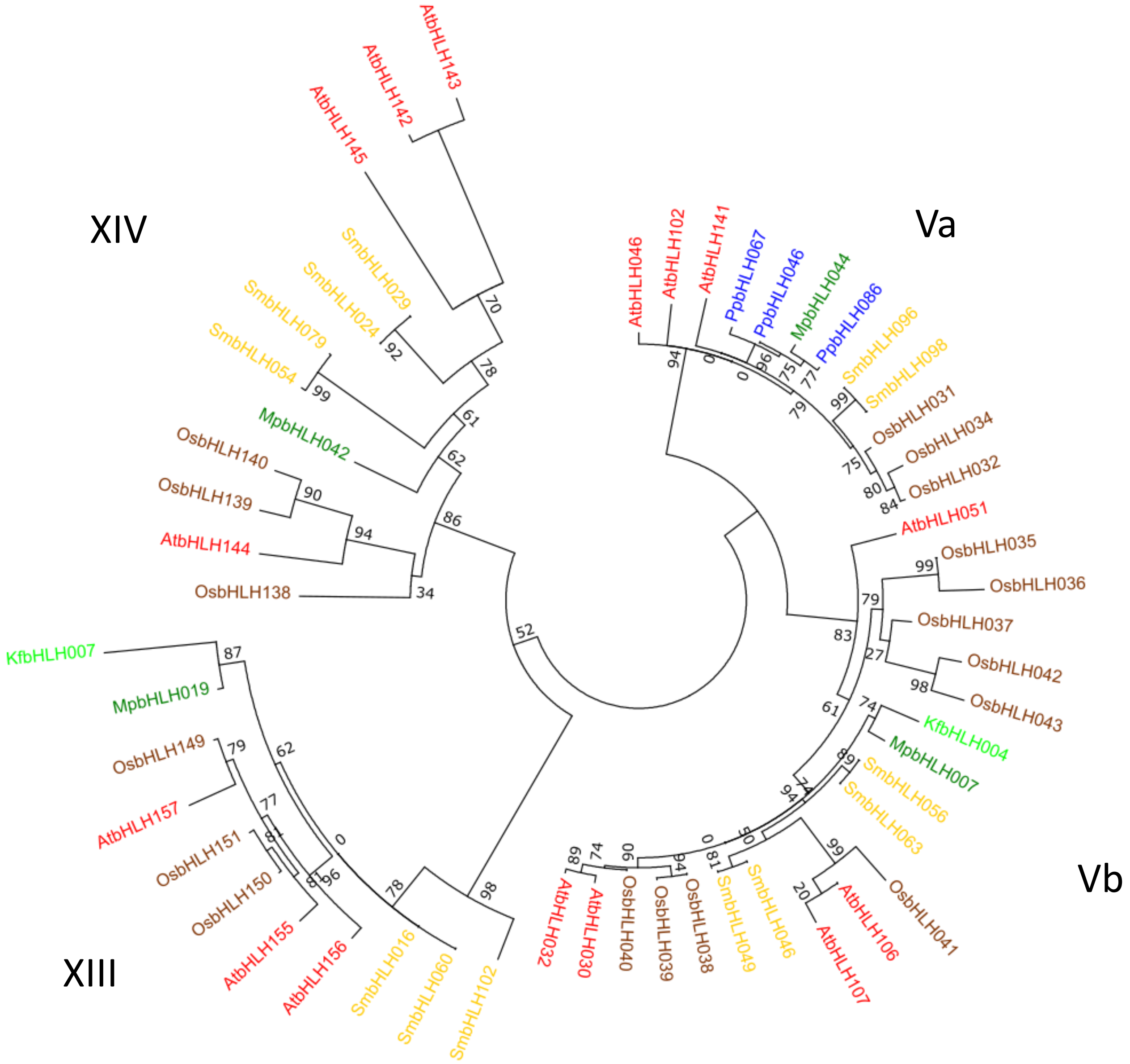


Figure 2.5 Phylogenetic analysis of bHLH subfamilies Va, Vb, XIII and XVI indicates that subfamilies Vb, XIII and XIV were lost in the moss lineage. Va, Vb, XIII and XVI sequences (Supplementary File 7) were aligned using MAFFT (Katoh and Frith, 2012; Katoh and Standley, 2013). The resulting alignment (Supplementary File 8) was subsequently used for phylogenetic analysis. The tree is an unrooted ML generated with the software PhyML 3.0, (Guindon *et al.*, 2010) using the JTT model of amino acid substitution, an estimated gamma distribution parameter and a SH-like aLRT test. Coloured taxa represent proteins from *Arabidopsis thaliana* (red); *Oryza sativa* (brown); *Selaginella moellendorffii* (gold); *Physcomitrella patens* (blue)*; Marchantia polymorpha* (dark green) and *K. flaccidum* (light green). Numbers at the branch are SH-like aLRT support values.

belong to different classes of mosses, Sphagnopsida and Bryopsida (Shaw et al., 2010), respectively, and it is possible that these subfamilies were lost in the early radiation of the moss clade . This suggests that several components of a gene regulatory network were lost in the moss lineage, or at least during the early radiation of the lineage that comprises the two distinct moss classes; Sphagnopsida and Bryopsida. Indeed, the loss of such regulatory components might suggest that mosses and other lineages of land plants, such as the angiosperms, have a different transcriptional mechanism that regulates periclinal cell divisions.

2.4.6. The majority of plant and metazoan bHLH subfamilies evolved in the Precambrian

To independently assess if morphological diversification was predated by the evolution of the majority of bHLH subfamilies in another eukaryotic radiation I contrasted the evolution of the bHLH family in metazoans (Degnan et al., 2009; Larroux et al., 2008). Strikingly, in metazoans, like plants, there was an early origin followed by a stepwise increase of the majority of bHLH subfamilies before the radiation of bilaterians (Figure 2.6). This observation indicates that the evolution of the majority of bHLH subfamilies in the Precambrian predated the major morphological diversification in both plants (radiation of land plants) and metazoans (radiation of bilaterians) in the Phanerozoic (Clarke et al., 2011; Erwin et al., 2011; Kenrick and Crane, 1997; Parfrey et al., 2011).

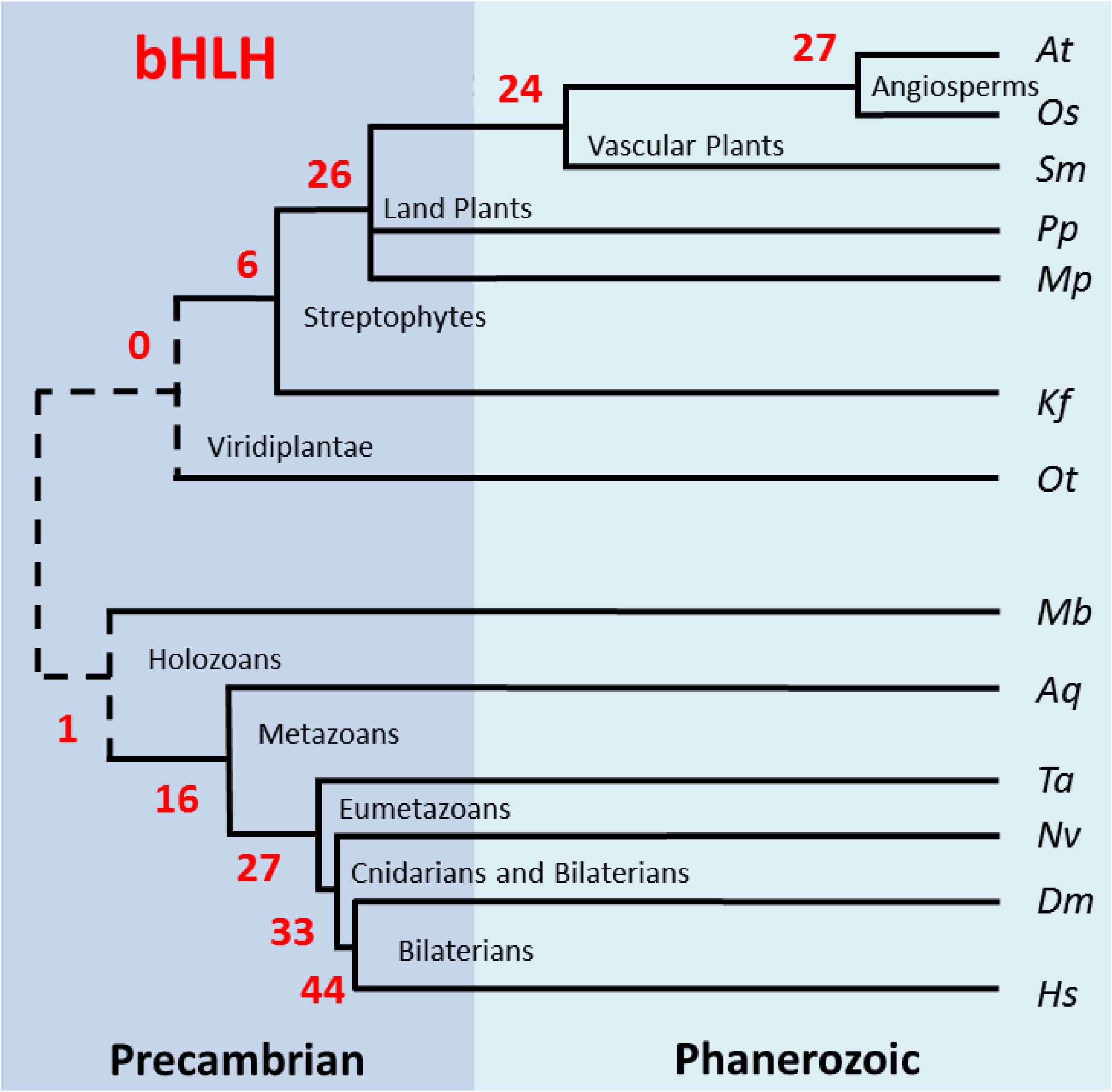


Figure 2.6 A Precambrian origin the majority of HD and bHLH TF subfamilies. The number of bHLH (red) subfamilies present at each node of a simplified time calibrated phylogeny of animals and Archaeplastida. Subfamily numbers for Archaeplastida are based on the phylogenetic analysis of this Chapter and for metazoans based on (Degnan et al., 2009; Larroux et al., 2008). Phylogenetic tree for Archaeplastida based on (Cox et al., 2014; Wickett et al., 2014), time calibrations for land plants based on (Clarke et al., 2011), time calibration for the origin of streptophytes based on (Parfrey et al., 2011). Phylogeny and time calibration of animals based on (Erwin et al., 2011). Dashed lines lack time calibrations. Abbreviated taxa names are *At, Arabidopsis thaliana*; *Os, Oryza sativa*; *Sm, Selaginella moellendorffii*; *Pp, Physcomitrella patens*; *Mp, Marchantia polymorpha; Kf ,Klebsormidium flaccidum; Ot, Ostreococcus tauri*; *Aq, Amphimedon queenslandica; Nv, Nematostella vectensis; Sp, Strongylocentrotus purpuratus; Hs, Homo sapiens; Dm, Drosophila melanogaster; Lg, Lottia gigantea.*

There is a clear trend of diversification within ancient subfamilies rather than expansion through the origin of novel subfamilies during the Phanerozoic in both eukaryotic lineages. Although plants differ from metazoans in many ways, one of their notable differences is the ability of plants to continue to evolve higher taxonomic units throughout the Phanerozoic, whereas animals had established the great majority of their phyla by the end of the Cambrian (Figure 2.6) (Erwin et al., 2011; Niklas et al., 1983; Valentine et al., 1991). In plants, the diversification of higher taxonomic units, such as the origins of lycophytes, gymnosperms and angiosperms within the Phanerozoic, did not involve a significant increase in bHLH subfamily number (Figure 2.6). The finding of a Precambrian origin of the majority of bHLH subfamilies relies on molecular clock timing estimates which are debated (Clarke et al., 2011; Heckman et al., 2001; Hedges et al., 2004; Kenrick et al., 2012; Parfrey et al., 2011; Sanderson, 2003; Yoon et al., 2004; Zimmer et al., 2007). The study by Clarke et al., (2011) was chosen for the comparative analysis because of their use of 17 minimum and maximum fossil age calibration points, which are essential for reliably dating plant phylogeny (Kenrick, 2011). In addition, as almost the full complement of bHLH subfamilies is already present it *M. polymorpha*, even with later estimates of the origin of land plants (Parfrey et al., 2011; Sanderson, 2003; Yoon et al., 2004), it is possible to hypothesise that the origin of these subfamilies occurred in the Precambrian streptophytic ancestor. Altogether, these observations suggest that the major transformations of the body plan in both multicellular eukaryotic lineages in the Phanerozoic were preceded by a stepwise increase of the bHLH regulatory toolkit during the Precambrian.

2.5. Discussion

The analysis of TF family distribution across Archaeplastida revealed that there was a stepwise increase in TF family before the colonization of the land by multicellular, photosynthetic, eukaryotes. The in-depth analysis of one of the largest families of TF in Archaeplastida, the bHLH family, indicated that the majority of bHLH subfamilies had evolved before the radiation of plants on the land in a stepwise manner. Moreover, this evolutionary trend is shared in both plants and metazoans; almost all extant bHLH subfamilies were present in the first land plants and in the last common ancestor of bilaterians.

Land plants radiated both taxonomically and morphologically after the colonization of the land (Bateman et al., 1998; Finet et al., 2010; Kenrick and Crane, 1997; Nickrent et al., 2000; Qiu et al., 2006; Wickett et al., 2014). However, this analysis revealed that this radiation was not accompanied by an increase in the number of novel TF families. Instead, the number of TFs within each family increased as land plants evolved after the colonization of the land (de Mendoza et al., 2013; Mukherjee et al., 2009; Pires and Dolan, 2010). The majority of TF families had evolved in the streptophyte aquatic ancestor of all land plants and only one TF family, GLABROUS1 enhancer-binding protein (GeBP), had its origin during the evolution of plants in the terrestrial realm just before the radiation of vascular plants. The GeBP family of TFs are known regulators of cytokinin response in angiosperms (Chevalier et al., 2008). Cytokinin signalling is likely to have evolved in the streptophyte aquatic ancestor of land plants (Hori et al., 2014) and this observation suggests that plants acquired new regulatory components during the evolution on land to regulate pre-existing signalling cascades. Nevertheless, the massive origin of TF family diversity occurred before the radiation of extant lineages of land plants, suggesting that the gene toolkit involved mainly in transcriptional regulation was established since all land plants shared a common ancestor with streptophyte algae and that diversification of TFs occurred within ancient families. These ancient families were therefore reused, recruited and co-opted to regulate developmental mechanisms since the early colonization of the land by plants.

The majority of plant bHLH subfamilies had evolved before the radiation of extant lineages of land plants. This means that the bHLH gene toolkit that controls developmental processes in later divergent lineages of land plants was recruited, co-opted and deployed from pre-existing subfamilies in the first land plants. Some bHLH subfamilies have been functionally characterised in both bryophytes and angiosperms (Table 2.3) (Bergmann and Sack, 2007; Breuninger et al., 2016; Castillon et al., 2007; Inoue et al., 2016; MacAlister and Bergmann, 2011; Menand et al., 2007; Pires et al., 2013; Proust et al., 2016; Tam et al., 2015; Yi et al., 2010). The function of these bHLH subfamilies appears to be conserved since all land plants shared a common ancestor. Amongst these there are three bHLH subfamilies that regulate the development of filamentous tip growing cells with a rooting function across land plants. bHLH subfamilies VIIIc(1), VIIIc(2) and XI are components of a transcriptional regulatory network that was likely to regulate the development of the first rooting systems in land plants (Table 2.3) (Bruex et al., 2012; Kim and Dolan, 2016; Pires et al., 2013; Tam et al., 2015). It is tempting to hypothesize that the function of other bHLH subfamilies has also been conserved throughout land plant evolution and that they regulated critical developmental processes in the early plant colonizers. Therefore, I postulate that the evolution of bHLH TFs was crucially important to promote body plan transitions that allowed the successful colonization and subsequent radiation of plants on land. The functional characterization of more bHLH subfamilies across land plants will be critical to validate this hypothesis.

One of the bHLH subfamilies encoded in the genome of the streptophyte alga *K. flaccidum* is the subfamily VII(a+b). This subfamily has been shown to regulate plant

Table 2.3 bHLH subfamilies have conserved functions in land plants.

Subfamily	Function	References
Ia	Regulate stomata morphogenesis in mosses and angiosperms. Moss Ia proteins can rescue Ia loss-of-function phenotypes in the angiosperm *A. thaliana*	MacAlister and Bergmann, 2011 Bergmann and Sack, 2007 Chater et al, 2016
VIIIc(1)	Required for the development of filamentous tip growing cells with a rooting function (root hairs, rhizoids and caulonema) across land plants	Jang et al., 2011; Menand et al., 2007; Proust et al., 2015
VIIIc(2)	Involved in the development of filamentous tip growing cells with a rooting function (root hairs, rhizoids and caulonema) in land plants. They constitute a transcriptional network with VIIIc(1) bHLH that has been conserved since mosses and angiosperms shared a common ancestor	Pires et al., 2013; Yi et al., 2010
XI	Conserved regulators of tip growing cells with a rooting function in land plants	Breuninger et al., 2016;Bruex et al., 2012; Karas et al., 2009; Tam et al., 2015
VII(a+b)	Involved in light-mediated developmental regulation in liveworts and angiosperms	Castillon et al., 2007; Inoue et al., 2016

development downstream of light signaling (Castillon et al., 2007; Inoue et al., 2016; Zhiponova et al., 2014). Light is one the most important environmental (exogenous) cues for plant growth and development (Castillon et al., 2007). It is a source of energy for photosynthetic processes and metabolism, and it regulates plant development via phytochromes that perceive light and regulate gene expression that it is further translated into developmental processes. For example, in the livewort *M. polymorpha*, light quality affects, amongst other developmental processes, thallus growth, gemma germination and tissue regeneration (Nakazato et al., 1999; Nishihama et al., 2015). There is a single phytochrome protein in *M. polymorpha*, MpPHY, that inhibits the transcriptional network regulated by the single PHYTOCHROME INTERACTING FACTOR, MpPIF, bHLH TF (Inoue et al., 2016). Under red light, MpPHY interacts with MpPIF in the nucleus and triggers MpPIF degradation via the 26S proteasome, resulting in the repression of the MpPIF transcriptional-dependent pathway that negatively regulates thallus growth, gemma germination and tissue regeneration. This mechanism is also conserved in angiosperms, suggesting that the first land plants used the PHY-PIF module to regulate light-dependent developmental programs (Castillon et al., 2007; Inoue et al., 2016). Since VII(a+b) bHLH are encoded in the genome of streptophyte alga, it is plausible that this mechanism predated the colonization of the land by plants and that VII(a+b) bHLH regulated light signalling in the streptophyte aquatic ancestor of land plants.

During land plant evolution, there were origins and losses of bHLH subfamilies. Moreover, some of these losses appear to be lineage specific. The moss lineage has lost three bHLH subfamilies that are known to control periclinal cell division in angiosperms. The losses of bHLH subfamilies have also been reported in metazoan evolution. Three metazoan bHLH subfamilies (pearl, amber and peridot) were lost during bilaterian evolution (Gyoja and Satoh, 2013). This contrasts with the evolution of other TF families, like the

Homeodomain TF family, in both eukaryotic lineages, where losses of Homeodomain subfamilies did not occurred during the evolution of metazoans and plants (Mukherjee et al., 2009; Nam and Nei, 2005).

Metazoans and plants independently evolved multicellularity and have subsequently radiated morphologically and taxonomically to their huge diversity today (Knoll, 2011). These lineages seem to be marked by the expansion of the number of TFs within ancient subfamilies rather than by the origin of novel subfamilies. From examining molecular clock analyses it is also apparent that the increase in subfamily number occurred prior to the morphological radiation in both of these lineages (Clarke et al., 2011; Erwin et al., 2011; Kenrick and Crane, 1997; Niklas et al., 1983). Based on the molecular clocks, the full subset of bHLH subfamilies were in place at roughly 670 million years ago (Clarke et al., 2011). The comparison of bHLH family evolution in plants and metazoans indicates that the evolution of complex multicellular body plans was driven by the co-option of existing subfamilies of TFs. I hypothesize that new developmental mechanisms were a product of TFs diversification within ancient subfamilies present over 600 million years ago in two multicellular eukaryotic lineages.

Whole genome sequencing of more species spanning Archaeplastida will allow the performance of more robust phylogenetic analyses. These will enable to track with more precision the evolution of gene families and disclose evolutionary trends that encompass the morphological radiation of plants. It seems that the TF gene toolkit is ancient and its origin predated the morphological radiation of land plants. However, increasing the sampling of phylogenetic analyses with the addition of more taxa is indispensable in order to infer the scale (if at a species or at higher taxonomic levels) of evolutionary events in gene families, such as losses of entire subfamilies during the course of plant evolution.

Chapter 3: IIIf bHLH transcription factors are conserved regulators of epidermal cell development in land plants

3.1. Abstract

The evolution of complex 3D tissues during the early colonization of the land was a critical morphological breakthrough in plant body plan that enabled the successful colonization and radiation of plants on land. The origin of the epidermis occurred concomitantly with the evolution of complex 3D tissues. The epidermal cell layer constitutes the outermost cell layer and it functions as the interface between the plant and the environment. Therefore, the evolution of the epidermis was crucial for the adaptation and radiation of plants on the terrestrial realm. IIIf bHLH TFs are important regulators of epidermal cell development in angiosperms. Here, I demonstrate that the function of IIIf bHLH TFs in controlling the development of the epidermal cell layer is ancient and conserved in land plants, predating the radiation of extant lineages of land plants. Furthermore, I show that there is a negative feedback loop between auxin and IIIf bHLH TF that regulate thallus growth and dorsal epidermis and tissue differentiation in *M. polymorpha*. Moreover, these bHLH TFs were recruited during the evolution of land plants to control the development of seemingly unrelated morphological characters in specific lineages of extant land plants. The recruitment of ancient developmental regulators to control distinct and unrelated developmental processes in land plants might underlie the huge morphological and taxonomic radiation of land plants.

3.2. Introduction

The epidermis of angiosperms is comprised by non-specialized pavement or non-hair cells and specialized cells: trichomes, stomata and root hairs. The development of these epidermal cells is regulated by a spatial patterning mechanism that defines the epidermal cells that undergo differentiation (Baudry et al., 2006; Bernhardt et al., 2005; Gonzalez et al., 2008; Ramsay and Glover, 2005). Interestingly, the genetic mechanism that controls the pattern of these specialized cells in the epidermis of different plant organs is conserved in angiosperms (Zhang et al., 2003). A conserved set of MYB, IIIf bHLH and WD-40 TFs function as a transcriptional complex that regulates epidermal cell development in different plant organs in angiosperms. The WD40 protein, TRANSPARENT TESTA GLABRA 1 (TTG1), is a component of this transcriptional complex providing a scaffold for the interaction between IIIf bHLH and MYB TFs (Figure 3.1) (Walker et al., 1999; Zhang et al., 2003). In the *A. thaliana* mutant At*ttg1* the development of trichomes, root hairs and stomata is defective, evidencing the importance and general role of TTG1 for the assembly of the transcription complex that controls epidermal cell differentiation (Walker et al., 1999).

The above-mentioned transcriptional complex (formed by MYB, IIIf bHLH homodimers or heterodimers and WD-40) is able to either positively or negatively regulate the transcription of downstream targets, thus triggering the differentiation of epidermal cells (Qing and Aoyama, 2012). The MYB TF that is present in the complex determines whether this transcriptional complex positively or negatively regulates the expression of downstream genes, and is thus a key regulator of epidermal cell fate (Figure 3.1) (Kirik et al., 2004; Schiefelbein, 2003; Zhang et al., 2003). In the case of the differentiation of stomata in the hypocotyl and root hairs in the root, the presence of a R2R3 MYB in the complex is able to positively regulate the expression of a Homeodomain TF, GLABRA 2 (GL2) (Figure 3.1) (Hung et al., 1998; Lee and Schiefelbein, 1999, 2001). When GL2 is expressed, the genetic mechanism that regulates the development of stomata and root hairs is repressed, which results in the differentiation of non-stomata cells in the hypocotyl and non-root hair cells in the root (Masucci et al., 1996; Qing and Aoyama, 2012). R3 MYB TFs compete with the R2R3 MYB to interact with the IIIf bHLH and WD-40 TFs. When the R3 MYB forms a complex with the IIIf bHLH and WD-40 TFs, the complex is able to negatively regulate the expression of GL2 (Figure 3.1) (Ryu et al., 2013; Schellmann et al., 2002; Wada, 2002; Wada et al., 1997). In this scenario, the genetic mechanism that drives the differentiation of stomata and root hairs is positively regulated, resulting in the development of such specialized epidermal cells. The competition between R2R3 MYB and R3 MYB proteins is a result of lateral inhibition mediated by a cell to cell movement of the R3 MYB (Kurata et al., 2005; Schellmann et al., 2002; Simon et al., 2007) (Figure 3.1). The R2R3MYB-IIIf bHLH-WD-40 complex positively regulates the expression of R3 MYB, which moves into and acts in the neighbouring cell (Figure 3.1). This means that the cell where R3 MYB is expressed remains as a non-differentiated cell whereas the neighbouring cell will differentiate into either a stoma in the hypocotyl or root hair in the root (Ryu et al., 2005). Conversely, the differentiation of trichomes occurs when the GL2 TF is expressed, i.e. when the R2R3 MYB is present in the transcriptional complex (Figure 3.1) (Kirik et al., 2004; Rerie et al., 1994; Schellmann et al., 2002; Walker et al., 1999; Yang and Ye, 2013). The complex comprised by R3 MYB-IIIf bHLH-WD40 represses the expression of GL2, resulting in the non-differentiation of trichomes (Figure 3.1).

The IIIf bHLH subfamily is one of the genetic components that provide the developmental machinery necessary to regulate the development of epidermal cells. As described above, these IIIf TFs interact with the WD-40 TF TTG1 and with the R2R3 or R3 MYB TFs to form a transcriptional complex that regulates epidermal patterning in angiosperms. In *A. thaliana* there are four IIIf bHLH proteins: GLABRA 3 (GL3), ENHANCER OF GLABRA 3 (EGL3), MYC1 and TRANSPARENT TESTA 8 (TT8) which can form homodimers or heterodimers amongst themselves with exception of MYC1 (Baudry et al., 2006; Bernhardt et al., 2003, 2005; Carretero-Paulet et al., 2010; Payne et al., 2000; Pires and Dolan, 2010; Zhang et al., 2003). Recent studies have shown that a domain in the N-terminus of the protein sequence of these TFs are required for the dimerization with the R2R3 MYB or R3 MYB, indicating that the presence of this domain located at the N-terminus of the protein is necessary for the biochemical function of this bHLH subfamily (Zhao et al., 2012).

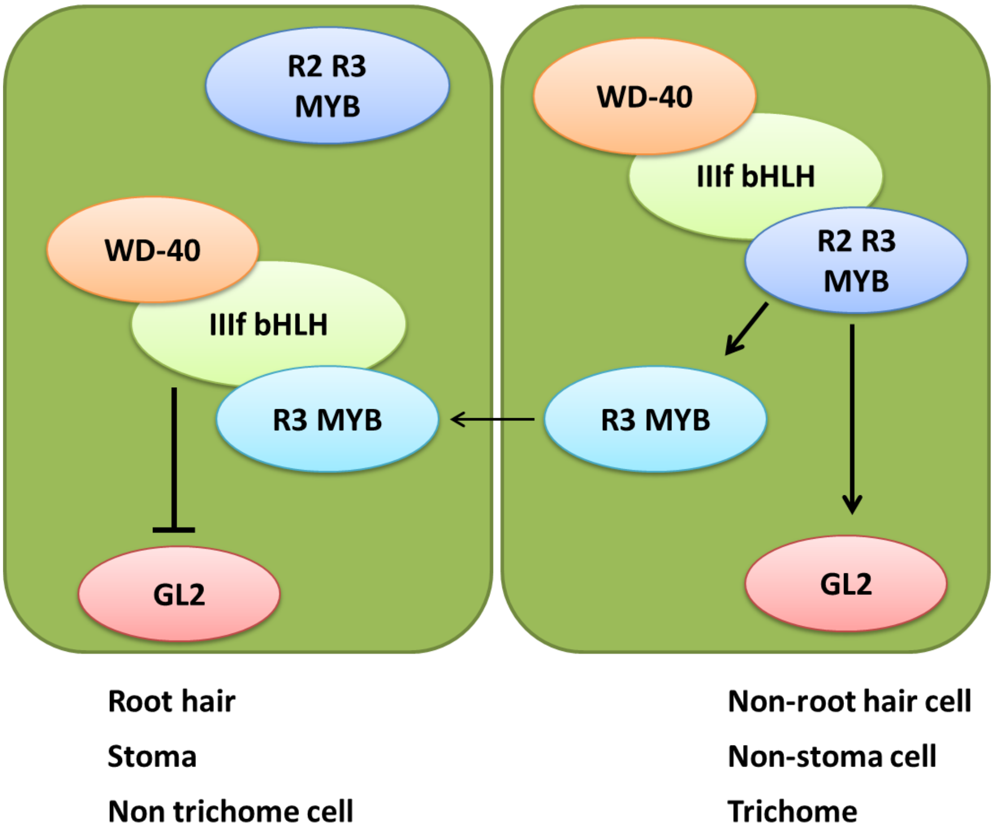


Figure 3.1 Genetic mechanism of epidermal cell differentiation in *A. thaliana*. A transcriptional complex formed by WD-40, IIIf bHLHs and R2R3 MYB, or R3 MYB, regulate epidermal cell differentiation, namely root hair, stoma and trichome, in different organs of *A. thaliana*.

The four *A. thaliana* IIIf bHLH TFs are differentially expressed in the plant (Baudry et al., 2006; Gonzalez et al., 2008). These different expression patterns correlate with the function of the gene inferred by each corresponding loss-of-function mutant phenotype. A clear example of this is found, for instance, in leaves, the TT8 gene is only expressed at the leaf margin and its function is restricted to the control of the development of marginal leaf trichomes in this specific leaf domain (Baudry et al., 2006).

To date, there is no functional characterization of IIIf bHLH TFs in other land plant lineages, apart from angiosperms. In Chapter 2, I have shown that the origin of IIIf bHLH subfamily predated the radiation of plants on land. Moreover, it has been shown that the function of some bHLH subfamilies is conserved to regulate specific developmental processes across land plants, suggesting that these functions might be ancient and were active in the first land plants to regulate crucial developmental processes. Since IIIf bHLH TFs regulate epidermal cell development in angiosperms, I hypothesize that this function might be conserved across land plants and that IIIf bHLH TFs regulated the development of the outermost cell layer – the epidermis; critical for the successful colonization and radiation of plants on land. To test this hypothesis, I assessed the function of IIIf bHLH subfamily in one of the earliest divergent lineage of land plants: the liverwort *M. polymorpha*. The analogy between the function of IIIf bHLHs in *M. polymorpha* and the previously characterized function in angiosperms allows me to infer the ancestral role of IIIf bHLH TFs in land plants. Moreover, *M. polymorpha* is a thalloid liverwort that develops specialized epidermal cells on the dorsal (air pore complexes) and on the ventral (rhizoids) sides of the thallus, constituting therefore a good model system to study the development and evolution of epidermal cell differentiation. Furthermore, gene function analyses are possible in *M. polymorpha* due to the establishment of transformation protocols and molecular techniques that allow the gene expression manipulation in this liverwort (Ishizaki et al., 2013b, 2015; Kubota et al., 2013; Sugano et al., 2014).

In summary, I aim to assess whether the genetic mechanism that regulates epidermal cell development in angiosperms is conserved in plants since the early colonization of the land or whether it was recruited to control distinct developmental mechanism during the radiation of land plants. To this extent I carried out a reverse genetic approach to characterize the function of IIIf bHLH in *M. polymorpha*. As explained in great detail in the next sections of this chapter, the functional characterization of this bHLH subfamily, together with the analogy with the previously described function in angiosperms, allows me to infer whether IIIf bHLH TFs are ancient regulators of epidermal cell development in land plants.

3.3. Materials and methods

3.3.1. Sequence retrieval and phylogenetic analysis

In order to search and retrieve IIIf bHLH sequences in land plant species for subsequent phylogenetic analysis of IIIf bHLH subfamily, a BlastP search was performed using the IIIf bHLH AtbHLH001 as a querry Phytozome v9.1 (Goodstein et al., 2011), with exception of *M. polymorpha* (NCBI GenBank accession LVLJ00000000.1). The genomes and proteomes of *At, Arabidopsis thaliana*; *Fv, Fragaria vesca*; *Cs Capsella rubella*; *Cs, Citrus sinensis*; *Aq, Aquilegia coerulea*; *Mt, Medicago truncatula*; *St, Solanum tuberosum*; *Si Setaria italica*; *Os, Oryza sativa*; *Bd, Brachypodium distachyon*; *At, Amborella trichopoda*; *Sm, Selaginella moellendorffii*; *Pp, Physcomitrella patens; Sf, Sphagnum fallax* and *M. polymorpha* were assessed. The results of the blast search were analyzed manually to ensure the presence of the bHLH and IIIf conserved domains. The AtbHLH003 (Subfamily III(d+e)) and AtbHLH033 (Subfamily IIIb) were included as an outgroup.

All protein sequences (Supplementary File 9) were aligned using MAFFT (Katoh and Frith, 2012; Katoh and Standley, 2013) and further manually aligned in Bioedit (Hall 1999). The resulting alignment (Supplementary File 10) was used for subsequent phylogenetic analysis. The Maximum likelihood analysis was carried using PhyML 3.0 (Guindon et al., 2010), using Jones, Taylor and Thorton (JTT) amino acid substitution model and a predicted gamma distribution. Branch support was tested using a Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-like aLRT). The generated unrooted trees were visualized using MEGA6 (Tamura et al., 2013).

3.3.2. Plant and bacterial materials

The wild-type *Marchantia polymorpha* used were male accession Takaragaike-1 (Tak-1) and female accession Takaragaike-2 (Tak-2), described in Okada et al. (2000) and Ishizaki et al. (2008). These were a gift of Kimitsune Ishizaki (Kobe University, Japan). *Escherichia coli* XL-1 Blue and *Agrobacterium tumefaciens* GV3101 were used in all cloning and transformations.

3.3.3. Plant growth conditions

M. polymorpha was grown in axenic culture in Petri dishes containing modified Johnson's medium (Johnson et al. 1957) (6 mM KNO3 (Sigma); 0.5mM MgSO4 (Sigma); 4 mM Ca(NO3)2 (Sigma); 25 µM KCl (Sigma); 10 µM H3BO3 (Fisher Scientific); 1 µM MnSO4 (Fisher Scientific); 1 µM ZnSO4 (Sigma); 0.25 µM CuSO4 (Sigma); 0.25 µM (NH4)6Mo7O24 (Sigma); 25 µM FeSO4 (BDH Laboratory Supplies, UK); 25 µM Na2-EDTA (Sigma); 0.6 mM NH4H2PO4 (Sigma); 0.4 mM (NH4)2SO4 (Sigma); 2.7 mM 2-(N-morpholino) ethanesulfonic acid (Sigma); 29 mM sucrose (Sigma); pH adjusted to 5.8 with KOH; solidified with 1.4% Agar (Sigma, Cat. No. A9799)). The Petri dishes were placed horizontal in a Sanyo growth cabinet at 23°C under continuous white light (light intensity of 38 µE.m-2.s-1).

Plants to be used for sexual reproduction, and subsequent sporangia collection, were grown on a 1:3 mix of vermiculite and John Innes No. 2 compost (Arthur Bower's) at 18°C under a 16h light: 8h dark photoperiod with a white light intensity of 50 µmol photons m^{-2} s^{-1}, supplemented with far red light (Phillips GreenPower LED HF module, 740 nm peak emission, 6 µmol photons m^{-2} s^{-1}).

3.3.4. DNA extraction

Four entire gemmalings were frozen and ground to powder with liquid nitrogen. Subsequently, 500 µL DNA extraction buffer (100 mM TRIS pH8; 1.4 M $NaCl_2$; 20 mM EDTA; 2% w/v CTAB; 1% w/v polyvinylpyrrolidone, 0.3% v/v β-mercaptoethanol) was added to fine-powder and incubated at 60°C for 30 minutes. 250 µL chloroform:isoamyl alcohol 24:1 was added, the sample was mixed and centrifuged at 13,000 rpm for 1 minute. The upper phase was transferred to a new microcentrifuge tube. This was repeated one more time and 500 µL isopropanol was added. The sample was centrifuged at 13,000 rpm for 15 minutes. The supernatant was removed and the pellet washed with 200 µL 70% ethanol. The sample was then centrifuged at 13,000 rpm for 2 minutes. The supernatant was removed and the pellet air-dried for 15 minutes, then dissolved in 90 µL water.

3.3.5. RNA extraction and cDNA synthesis

Four entire gemmalings were flash frozen and ground to powder in liquid nitrogen and 700 µl Trizol ® was added. The samples were mixed thoroughly and incubated at 56°C for 10 minutes with gentle shaking. Posteriorly, samples were centrifuged at 13,000 rpm for 1 minute. The supernatant was transferred to a new microcentrifuge tube and 700 µL ethanol was added and mixed. RNA isolation proceeded using the Direct-zol RNA MINIprep kit (Zymo Research) according to the manufacturer's instructions. RNA was eluted in 35 µL nuclease-free water.

RNA concentration was measured using a NanoDrop ND1000 spectrophotometer (Thermo Fisher Scientific). The concentration of RNA in each sample was equalized. Turbo DNA-free kit (Ambion, Life Technologies) was used according to the manufacturer's instructions to digest contaminating genomic DNA. cDNA was synthesized using ProtoScript II reverse transcriptase (New England Biolabs) according to the manufacturer's instructions, with oligo(dT) primer.

3.3.6. RT-qPCR

RT-qPCR analysis was carried out using the SensiFAST SYBR Hi-ROX Kit (Bioline) and the Applied Biosystems 7300 Real Time PCR system. Three technical replicates were used for each sample. MpACT, MpEF1α were used as reference genes, on the basis of an analysis of *M. polymorpha* reference genes suitable for RT-qPCR (Saint-Marcoux et al., 2015). cDNA of genes of interest was detected using the primers in Appendix 1 (includes primer names and sequences). The cycling conditions were 95°C for 2 min, then 45 cycles of 95°C for 15s, 60°C for 15s, 72°C for 1 min, with detection during the 72°C step. At the end of the cycles, a dissociation curve analysis was performed to confirm that presence of a single amplicon. The bundled 7000 System SDS Software (Applied Biosystems) was used to export the ΔRn for analysis using LinRegPCR v2012.0 (Ruijter et al. 2009). In LinRegPCR, the window of linearity was determined for each amplicon and the mean reaction efficiency calculated for each reaction. From this an N0 was computed, which reflects the initial amount of template included in the reaction.

3.3.7. Cloning

Generation of $_{pro}$MpGL3 reporter construct

3.9 kb of sequence upstream of the start codon of the CDS of MpGL3 was amplified from gDNA using Phusion High-Fidelity DNA Polymerase (New England Biolabs) with the primer pairs pMpGL3-PmeI_F and pMpGL3-SmaI-PmeI_R (See Appendix 1) and subcloned into pGEM-T. The fragment containing the 3.9 Kb sequence upstream of MpGL3 CDS was then digested with PmeI (New England Biolabs) and subsequently cloned into a binary vector (pCAMBIA1300). The resulting plasmid was digested with SacI (New England Biolabs), dephosphorylated with Antarctic Phosphatase (New England Biolabs), and the Gateway Vector Conversion System (Thermo Fisher) was used to ligate GW Cassette C.1 in between the promoter and terminator, to generate a destination vector containing a 3.9 Kb sequence upstream of MpGL3 CDS. An LR reaction was carried out between the destination vector and an entry vector containing 3xCFP-NLS. The resulting expression vector was designated $_{pro}$MpGL3:3xCFP-NLS:Term that confers gentamycin resistant in plants.

Generation of MpGL3 artificial microRNAs

The Mp*miR160* pre-miR backbone was used as the basis of amiR design (Flores-Sandoval et al., 2016), with the endogenous *miR160* sequence replaced with 21 nt targeting the MpGL3 transcript. miRs were designed using the WMD3 software (http://wmd3.weigelworld.org/), with the minimum number of included targets set to 1, the accepted off-site targets set to 3, the full-length MpGL3 transcript as the target, and the transcript library set to Marchantia_polymorpha_v183.mRNA.PUT.fasta. The two highest-ranked amiR, targeting MpGL3 were chosen (*amiR*-Mp*GL3-1*^{Mpmir160} *amiR*-Mp*GL3-2*^{Mpmir160}) (See Appendix 2). The amiR* was designed to have mismatches with the amiR sequence at positions 7, 13 and 18. These were then used to replace the native miR160 miR and miR* sequences in the backbone (See Appendix 2). This backbone was flanked by attB1 and attB2 sites and synthesized by Life Technologies and it was recombined with pDONR221 using BP Clonase II (Invitrogen) to create an entry vector, which was recombined with a destination vector containing the constitutively active $_{pro}$Mp*Ef1α* using LR Clonase II (Thermo Fisher) to generate the expression clones, the amiR constructs $_{pro}$Mp*Ef1α:amiR*-Mp*GL3-1*^{Mpmir160} and $_{pro}$Mp*Ef1α:amiR*-Mp*GL3-2*^{Mpmir160} that confer hygromycin resistant in plants.

Generation of Constitutive expression of MpGL3

Mp*GL3* CDS was amplified from wild type cDNA using Phusion High-Fidelity DNA Polymerase (New England Biolabs) with primers MpGL3_F and MpGL3_R (See Appendix 1) and recombined into the pCR8/GW/TOPO Gateway entry vector (Invitrogen) to create an entry clone with the Mp*GL3* CDS. To create a binary plasmid for plant transformation, an LR reaction was carried out between this plasmid and a destination vector containing the constitutively active $_{pro}$Mp*Ef1α* using Gateway LR Clonase II (Invitrogen), creating the expression vector that contains $_{pro}$Mp*Ef1α:*Mp*GL3:Term* and confers plant hygromycin resistance.

Generation of MpGL3-SRDX fusion construct

To generate a fusion between the C-terminus of Mp*GL3* and the EAR-motif repression domain (SRDX), the Mp*GL3* CDS was amplified from cDNA with Phusion High-Fidelity DNA Polymerase with the primers MpGL3_F and MpGL3-SRDX (See Appendix 1). MpGL3-SRDX replaces the STOP codon of Mp*GL3* with a sequence encoding the SRDX domain (LDLDLELRLGFA*). This product was recombined into the pCR8/GW/TOPO Gateway entry vector to create an entry clone containing MpGL3-SRDX fusion. An LR reaction was carried out between this entry clone and a destination vector containing the constitutively active $_{pro}$Mp*Ef1α* to generate an expression clone that contains $_{pro}$Mp*Ef1α:*Mp*GL3-SRDX :Term* and confers hygromycin resistance in plants.

3.3.8. Transformation of *M. polymorpha* sporelings

The method of transformation was based on that presented by Ishizaki and coworkers (2008). Wild-type sporangia obtained from crosses between Tak-1 and Tak-2 plants were surfaced sterilized with 1% sodium dichloroisocyanurate (Sigma) for 5 minutes, then rinsed 4 times with sterile water. Three sterilized sporangia were broken open in 500 µL sterile water, added to a 125ml vented tissue culture flask containing 25 mL M51C medium (19.8 µM KNO_3 (Sigma); 5 µM NH_4NO_3 (Sigma); 1.5 µM $MgSO_4$ (Fisher Scientific); 2 µM $CaCl_2$ (Sigma); 2 µM KH_2PO_4 (Sigma); 110 nM EDTA-Fe(III) (Sigma); 1 µM Na_2MoO_4 (Sigma) ; 100 nM $CuSO_4$ (Sigma); 100 nM $CoCl_2$ (Sigma); 7 µM $ZnSO_4$ (Sigma); 40 µM $MnSO_4$ (Fisher Scientific); 49 µM H_3BO_3 (Fisher Scientific); 555 µM *myo*-inositol (Sigma); 8 µM nicotinic acid (Sigma); 5 µM pyridoxine hydrochloride (Sigma); 30 µM thiamine hydrochloride (Sigma); 58 mM sucrose (Sigma); 2 mM L-glutamine (Sigma); 0.1% w/v casein hydrolysate (Fluka Analytical); pH 5.8), and cultured without shaking at 23°C under continuous white light (light intensity of 38 µE.m-2.s-1) for 6 days.

A single colony of *Agrobacterium tumefaciens* strain GV3101 carrying the vector of interest was inoculated into 5 ml LB medium containing 25 µg/mL gentamycin, 100 µg/mL rifampicin, and 50 µg/mL kanamycin (or 100 µg/mL spectinomycin for $_{pro}$Mp*GL3:3xCFP-NLS:Term*), and cultured with shaking at 27 °C for 2 days. The bacteria were sedimented by centrifugation at 2000 rpm for 15 minutes, the supernatant discarded, the bacteria resuspended in 10 ml M51C liquid medium containing 100 µM acetosyringone (Sigma Aldrich, 97%), and cultured for a further 6 h with shaking at 27 °C. The 7 day old cultures of *M. polymorpha* sporelings were then supplemented with 100 µM acetosyringone, and 1 mL of the bacterial culture was added. The sporelings and bacteria were co-cultivated with shaking at 120 rpm under continuous white light (56 µmol photons m^{-2} s^{-1}) at 18°C for 2 days. Sporelings were then collected on a sterile 40 µm cell strainer (Fisher Scientific) and rinsed thoroughly with 100 ml sterile distilled water. The sporelings were then spread on antibiotic selection plates of Johnson's agar with 100 µg/mL cefotaxime to kill *A. tumefaciens* and 10 µg/mL hygromycin (or 100 µg/mL gentamycin for $_{pro}$Mp*GL3:3xCFP-*

NLS:Term) to select for transformed plants. After 14 days transformants could be distinguished from antibiotic-sensitive non-transformed plants, and were transferred, at regular spacing, onto 12 cm square petri dishes of Johnson's agar with appropriate antibiotics for further growth and selection. Establishment of transformant lines from a single gemma of a primary transformant was done to avoid chimaerism.

3.3.9. Histological sections and *in situ* hybridization on longitudinal sections

10 day old thallus of TAK-1 and $_{pro}$Mp*Ef1α:amiR*-Mp*GL3-1*^{Mpmir160} plants were fixed using a 4% paraformaldehyde (PFA) solution in PBS. After 72h the samples were dehydrated in an ethanol series of 30%, 50%, 70%, 95% and 2x 100% ethanol. Fixed plants were subsequently embedded in paraffin, sectioned, and stained with toluidine blue. Slides were imaged with an Olympus BX50 compound microscope using DIC optics. Plants embedded in paraffin and sectioned were also used for *in situ* hybridization as described in Kim and Dolan, 2016.

3.3.10. Whole-mount *in situ* hybridization

Plants were grown from gemmae on modified Johnson's medium in 50 mm deep petri dishes (Fisher Scientific) for 10 days. Plants were fixed on 4% Paraformaldehyde for 50 minutes, followed by 2x of 100% methanol and 3x 100% ethanol for five minutes each. *In situ* hybridization was used to determine the cells in which Mp*GL3* mRNA accumulated. Antisense probes for the Mp*GL3* gene were made by amplification from the start codon of Mp*GL3* until the bHLH domain coding sequence by PCR. PCR fragments were cloned into pGEM-T Easy Vector (Promega). To generate specific antisense probes, full length cDNA was transcribed from the terminal SP6 or T7 RNA polymerase priming sites of these vectors, then treated DNase and precipitated RNA by ethanol. Full length RNA was hydrolysed by carbonate buffer (200 mM, pH 10) to create 150 bp fragments for hybridization. In vitro transcription, digoxigenin labelling of RNA probes and *in situ* hybridization were carried out as previously described (Derbyshire et al., 2008).

3.3.11. Microscopy

Stereomicroscopy

Images were obtained of plants growing in axenic culture using a Leica M165FC stereomicroscope with a Leica M series Plan APO 1.0x objective (Leica, Gemany). Images were captured using a Leica DFC310 FX camera.

Confocal laser scanning microscopy

Plants were grown from gemmae on modified Johnson's medium in 50 mm deep petri dishes (Fisher Scientific). For ventral side visualization, the petri dishes were positioned upside down so that the ventral side could be imaged. Cell walls were stained by immersing plants in 15 µM propidium iodide solution for 20 minutes, which was then removed by submergence of the plants in water. Images were acquired with a Leica SP5 confocal microscope and the Leica Application Suite (LAS) software, using either a Leica HCX APO L20x/0.50 W U-V-I dipping lens. YFP and CFP fluorescence was excited at 514 nm using an argon laser and emission was measured between 524 and 561 nm (YFP) and between 463 and 561 nm (CFP) using an Acousto Optical Tunable Filter. PI was excited at 543 nm using a helium-neon laser and emission was measured between 561 and 640 nm. The images were processed using FIJI software to create brightest-point 3d projections (Schindelin et al. 2012).

Scanning electron microscopy

Tissue samples were fixed using a 4% paraformaldehyde (PFA) solution in PBS. After 72h the samples were dehydrated in an ethanol series of 30%, 50%, 70%, 95% and 2x 100% ethanol before critical point drying. The samples were dried at the critical point using a Tousimis Autosamdri-815, then mounted on aluminium stubs (Agar Scientific) using double-sided carbon adhesive discs (Agar Scientific), and coated with a mixture of gold and palladium using an SC7640 sputter coater (Quorum Technologies). The samples were imaged immediately with a JSM-5510 scanning electron microscope (JEOL).

Image analysis

Image analyses and plant/cell measurements were performed in ImageJ.

3.3.12. Exogenous auxin and auxin biosynthesis inhibitor treatment

Plants were grown from gemmae on modified Johnson's medium in petri dishes (Fisher Scientific) for 21 days. They were further transferred to modified Johnson's media supplemented with 10µM of 1-Naphthaleneacetic acid (NAA) (Sigma-Aldrich) and 3-Indoleacetic acid (IAA) (Sigma-Aldrich), and 100 µM of L-Kynurenine (L-Kyn) (Sigma-Aldrich) for 2 days. As a mock control, plants were transferred to DMSO. Plant tissue harvesting, RNA extraction, cDNA synthesis and RT-qPCR were performed as described in sections 3.2.5 and 3.2.6 of this Chapter.

3.4. Results

3.4.1. *Marchantia polymorpha* encodes one IIIf bHLH

In the previous chapter I showed that IIIf bHLH TFs evolved after the divergence between the charophyte algae *K. flaccidum* and land plants. The phylogenetic analysis of bHLH TFs performed in Chapter 2 does not allow me to determine with precision the distribution and diversification of IIIf bHLH TFs in land plants due to the limited taxon sampling. In order to investigate the diversification of IIIf bHLH TFs in land plants, I conducted a phylogenetic analysis that includes more taxa, using the following species: the angiosperms *Arabidopsis thaliana*, *Fragaria vesca, Capsella rubella, Citrus sinensis, Aquilegia coerulea, Medicago truncatula, Solanum tuberosum, Setaria italica, Oryza sativa, Brachypodium distachyon, Amborella trichopoda*; the lycophyte *Selaginella moellendorffii,* the mosses *Physcomitrella patens* and *Sphagnum fallax*; and the liverwort *Marchantia polymorpha*. The Maximum likelihood analysis of IIIf bHLH proteins shows that there are two monophyletic clades of IIIf bHLH TFs in land plants, designated Class I and Class II IIIf bHLHs (Figure 3.2). Class I IIIf is a well-supported clade, with 94 aLRT branch support value, that comprises only angiosperm proteins (Figure 3.2). Class II IIIf is a monophyletic clade with a 90 aLRT branch support value that contains angiosperm, lycophyte and bryophyte proteins (Figure 3.2). The most parsimonious explanation for these results is that Class I IIIf bHLHs evolved after the divergence between lycophytes and angiosperms. However, it is also plausible, but less likely, that Class I IIIf bHLHs evolved in the first land plants but were lost in the bryophyte and lycophyte lineages.

The phylogenetic analysis of IIIf bHLH subfamily revealed that there is only one IIIf bHLH encoded in the genome of *M. polymorpha* (Figure 3.2). The full length amino acid sequence of MpGL3 is 30.3%, 31.2%, 28.3% and 39.8% identical to AtbHLH001, AtbHLH002, AtbHLH012 and AtbHLH042, respectively. IIIf bHLH TFs are characterized by the presence of non-bHLH conserved domains at the N-terminal and C-terminal end of the proteins in angiosperms (Feller et al., 2006; Zhao et al., 2012). These conserved domains have been demonstrated to be required for the function of IIIf bHLH, being crucial for homo- and heterodimerization with other TFs (Feller et al., 2006). I analysed the full length sequence of MpGL3 to investigate whether the *M. polymorpha* IIIf bHLH contains these domains. To this extent, I created a full length sequence alignment of MpGL3 and *A. thaliana* IIIf bHLH TFs (AtGL3, AtEGL3, AtMYC1 and AtTT8). The alignment of *M. polymorpha* and *A. thaliana* IIIf bHLH proteins shows that MpGL3 shares non-bHLH conserved domains with all *A. thaliana* IIIf bHLH proteins (Figure 3.3). There are two N-terminal and one C-terminus domains that are conserved between liverworts and angiosperms IIIf bHLH proteins. This observation suggests that the biochemical properties of

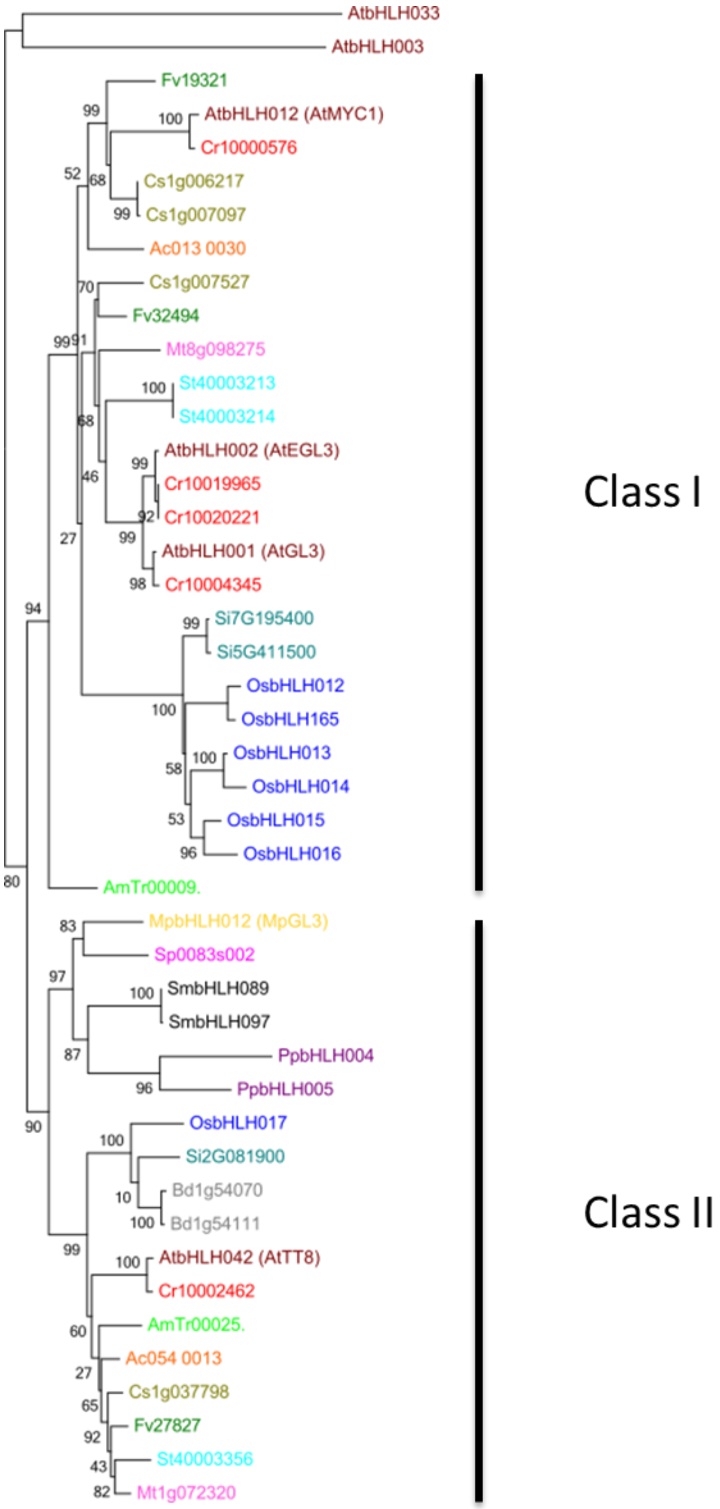


Figure 3.2 Phylogenetic analysis of bHLH subfamily IIIf reveals two monophyletic classes. MpGL3 belongs to IIIf Class II. The tree is a ML generated with the software PhyML 3.0, (Guindon *et al.*, 2010) using the JTT model of amino acid substitution, an estimated gamma distribution parameter and a SH-like aLRT test. Coloured taxa represent bHLH proteins from *Arabidopsis thaliana* (maroon); *Fragaria vesca* (green); *Capsella rubella* (red); *Citrus sinensis* (olive); *Aquilegia coerulea* (orange); *Medicago truncatula* (fuchsia); *Solanum tuberosum* (aqua); *Setaria italica* (teal); *Oryza sativa* (blue); *Brachypodium distachyon* (grey); *Amborella trichopoda* (lime); *Selaginella moellendorffii* (black); *Physcomitrella patens* (purple)*; Sphagnum fallax* (pink) and *Marchantia polymorpha* (gold). AtbHLH003 (Subfamily III(d+e)) and AtbHLH033 (Subfamily IIIb) were used as an outgroup. Numbers at the branch are SH-like aLRT support values. The *M.polymorpha* IIIf bHLH was named MpGL3.

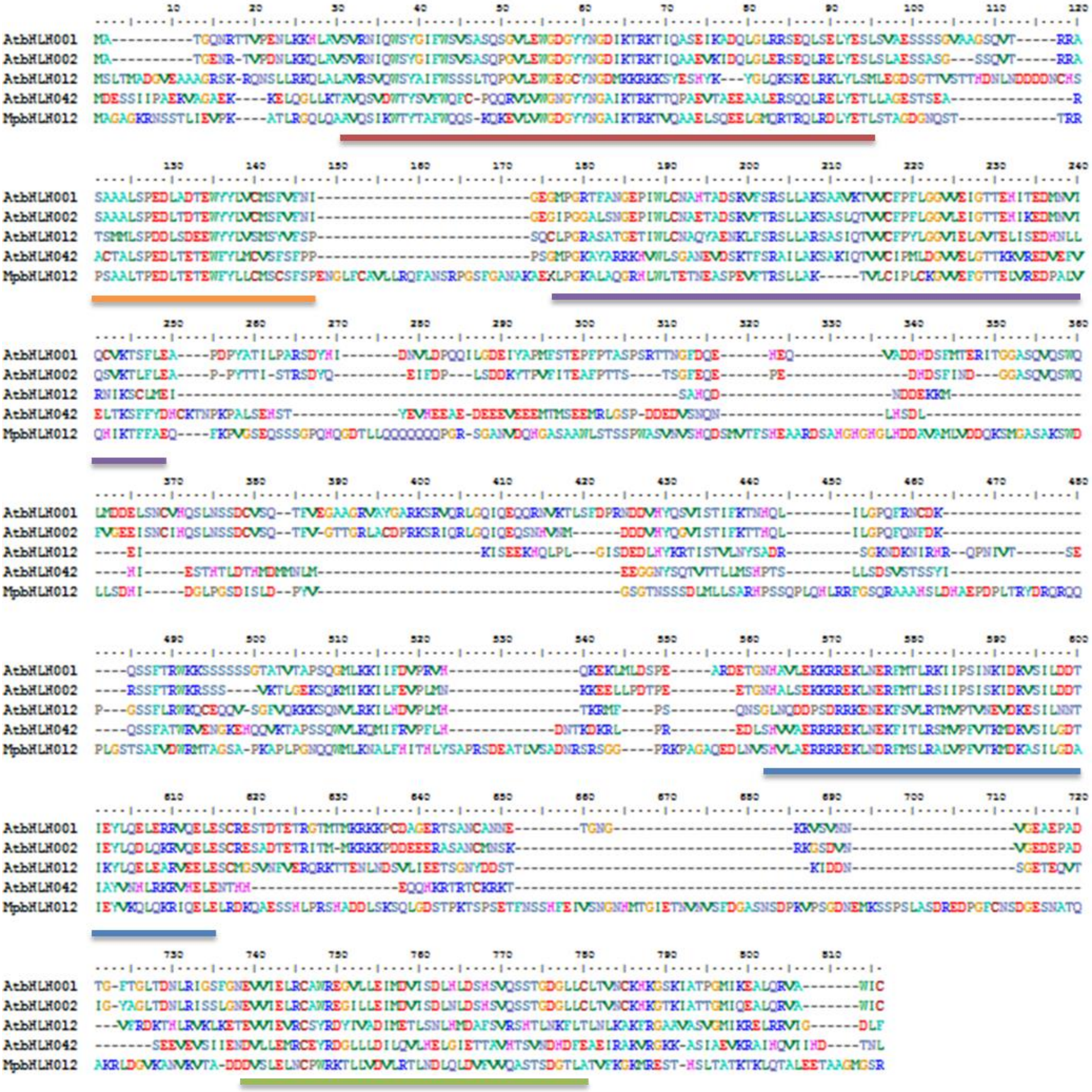


Figure 3.3 MpGL3 shares conserved domains with IIIf bHLH from *A. thaliana*. Full protein alignment of MpbHLH012 (MpGL3) with IIIf bHLH subfamily homologs from Arabidopsis: AtbHLH001 (AtGL3), AtbHLH002 (AtEGL3), AtbHLH012 (AtMYC1) and AtbHLH042 (AtTT8) shows that MpGL3 contains conserved amino acid domains characteristic of IIIf bHLH transcription factors. Conserved motifs are underlined in red (N-terminal-1), orange (N-terminal-2), purple (N-terminal-3) blue (bHLH) and green (C-terminal).

IIIf bHLH proteins might be conserved across land plants. Together, these results indicate that there is one IIIf bHLH TF in *M. polymorpha* (referred hereafter as MpGL3) and that IIIf bHLH subfamily is a conserved group of TFs that evolved highly conserved domains before the radiation of extant lineages of land plants. In the next sections I describe the functional characterization of Mp*GL3* in the development of *M. polymorpha* thallus in order to infer the ancestral function of IIIf bHLH TFs in land plants.

3.4.2. *MpGL3* mRNA accumulates preferentially in the dorsal thallus

The largest (dominant) phase of the *M. polymorpha* life cycle is the gametophyte which comprises a flattened structure (thallus) and aerial structures (antheridiophore and archaegoniophores). The thallus develops several days after spore or gemmae germination (Tarén, 1958). Within six days of thallus development no dorsiventrality is established and rhizoids are growing from both dorsal and ventral epidermis cells. Later in development, the upper (dorsal) and lower (ventral) sides of the thallus become morphologically different. Rhizoids develop from the ventral epidermis cells while air pore complexes develop from dorsal epidermal cells.

Since IIIf bHLH genes control the development of epidermal cells in angiosperms, I hypothesized that Mp*GL3* is active in the epidermis of *M. polymorpha.* To assess if Mp*GL3* is expressed in the dorsal and ventral epidermis, I performed whole-mount *in situ* hybridization of Mp*GL3* in 10-day-old thallus, at which point dorsiventrality is already established. Whole-mount *in situ* hybridization revealed specific expression patterns in the thallus of *M. polymorpha* (Figure 3.4). Mp*GL3* transcript was detected preferentially in the dorsal epidermis of the thallus (Figure 3.4A). In the ventral epidermis, Mp*GL3* transcript was only detected in the meristematic notch (Figure 3.4C). As a control, non-hybridizing sense

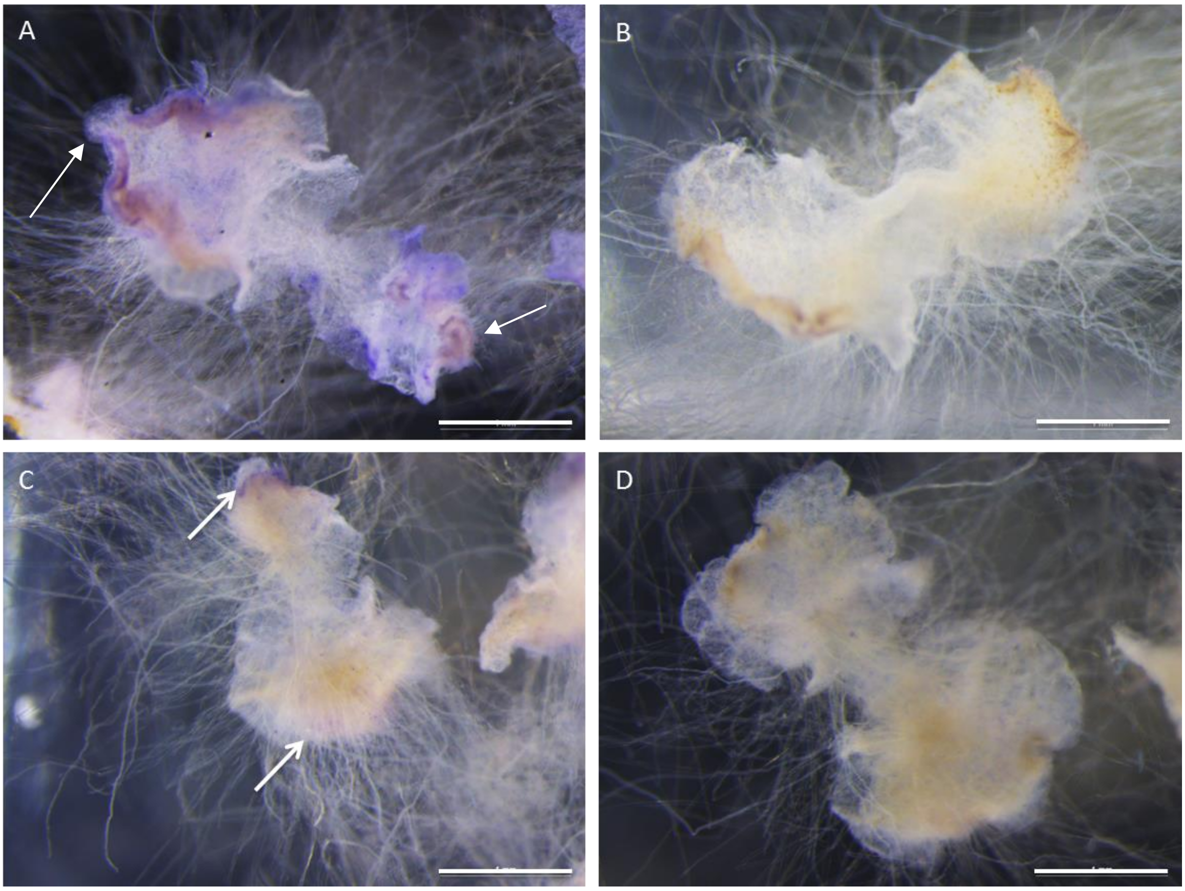


Figure 3.4 Mp*GL3* transcript accumulates preferentially in the dorsal epidermis and meristematic notch. Whole-mount i*n situ* mRNA hybridization of Mp*GL3* transcript in 10-day-old thallus shows preferentially accumulation of Mp*GL3* transcript in the dorsal side of the thallus and meristematic notches in both dorsal and ventral side of the thallus. (A) dorsal view of plants hybridized with antisense probe. (B) dorsal view of plants hybridized with sense probe (negative control). (C) ventral view of plants hybridized with antisense probe. (D) ventral view of plants hybridized with sense probe (negative control). White arrows point to the meristematic notch. Scale bar is 500 µm

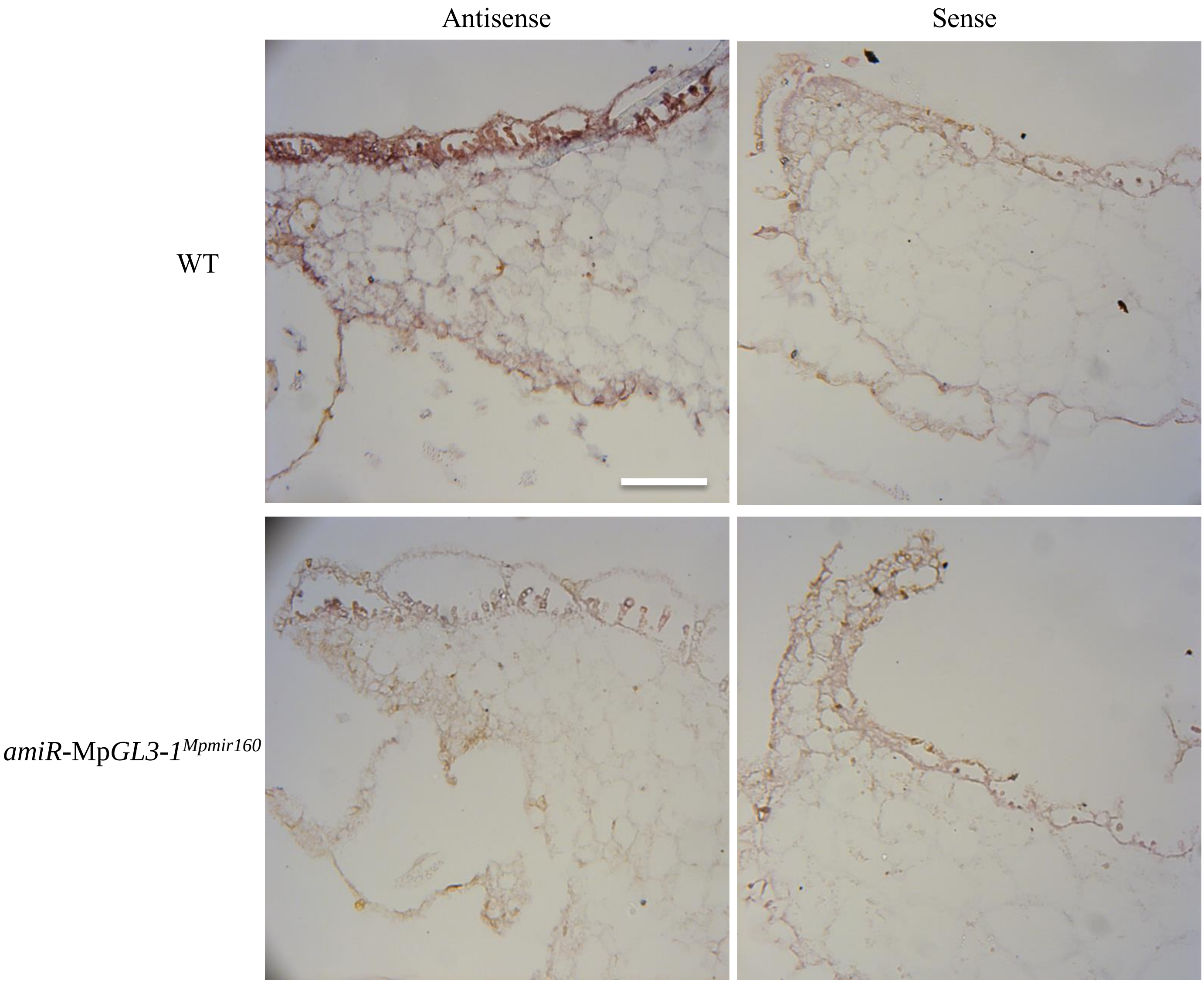


Figure 3.5 Mp*GL3* transcript accumulates preferentially in the dorsal epidermis and in air chamber. *In situ* mRNA hybridization of Mp*GL3* transcript in median longitudinal sections of 10-day-old thallus shows that Mp*GL3* transcript accumulates prefrentially in the dorsal side of the thallus, including the epidermis and the air chambers. No detectable signal is present in the sense control or in the antisense and sense hybridized *amiR*-Mp*GL3-1*^{Mpmir160} Median longitudinal sections. Arrow points to air chamber. Scale bar is 100 µm.

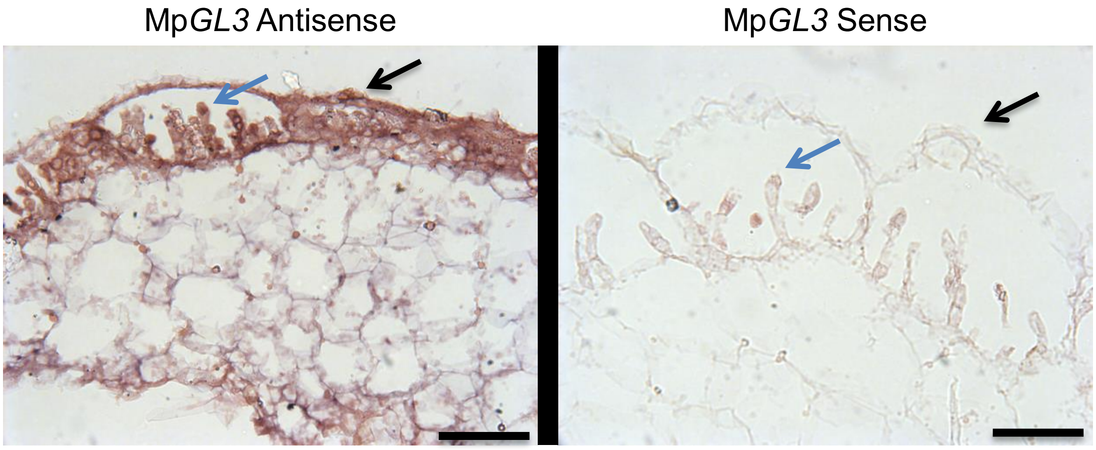


Figure 3.6 Mp*GL3* transcript accumulates preferentially in photosynthetic filaments, bottom and lateral sides of the air chamber. *In situ* mRNA hybridization in median longitudinal section of 10-day-old thallus of Mp*GL3* transcript hybridized with antisense and sense (control) probes. On the dorsal side of the thallus, Mp*GL3* transcript was detected in cells that comprise the air chamber; epidermis, air pore complex, photosynthetic filaments and the cells that line the bottom of the air chamber from which the filaments develop and cells that line the sides of the air chamber. Black arrows points to air pore complex and blue arrows point to photosynthetic filaments. Scale bar is 50 µm.

probe shows that there is no detectable signal in both dorsal and ventral epidermis (Figures 3.4B and 3.4D). This indicates that Mp*GL3* transcript accumulates preferentially in the dorsal side of the thallus, and the expression domain in the ventral side of the thallus is restricted to the meristematic notch. To further validate these observations, I performed *in situ* hybridization on median longitudinal sections of the thallus in WT and loss-of-function mutant (see section 3.4.4 of this section). Mp*GL3* transcript accumulation in the WT is enriched in the dorsal side of the thallus being particularly expressed in the dorsal epidermis, air pore complex cells, base and sides of the air chambers and in the assimilatory filaments of the air chambers (Figures 3.5 and 3.6). In the loss-of-function mutant Mp*GL3* transcript accumulation is practically undetectable, confirming the nature of this mutant (reduced gene expression) (Figure 3.5). Taken together, these data suggest that Mp*GL3* acts in the dorsal side of the thallus and that it might control the development of dorsal epidermal cells and air chambers.

3.4.3. The Mp*GL3* promoter activity is enriched in the dorsal thallus during gemmaling development

The accumulation of the Mp*GL3* transcript preferentially in the dorsal side of the thallus suggests that the promoter of Mp*GL3* is preferentially active in the dorsal epidermis. To investigate where the Mp*GL3* promoter is active in both the upper and lower sides of the thallus during thallus development from gemma, I transformed *M. polymorpha* with a construct that expresses 3xCFP-NLS under the control of a 3.9 kb fragment of gDNA upstream of the CDS of Mp*GL3*. Twelve transformants showed a strong expression of CFP throughout the dorsal thallus.

To investigate the spatiotemporal pattern of $_{pro}$Mp*GL3* activity in detail, I imaged $_{pro}$Mp*GL3:3xCFP-NLS* plants during the establishment of dorsiventrality using a confocal scanning laser microscope. To this end, I analysed the expression of 3xCFP-NLS in 0, 2, 4

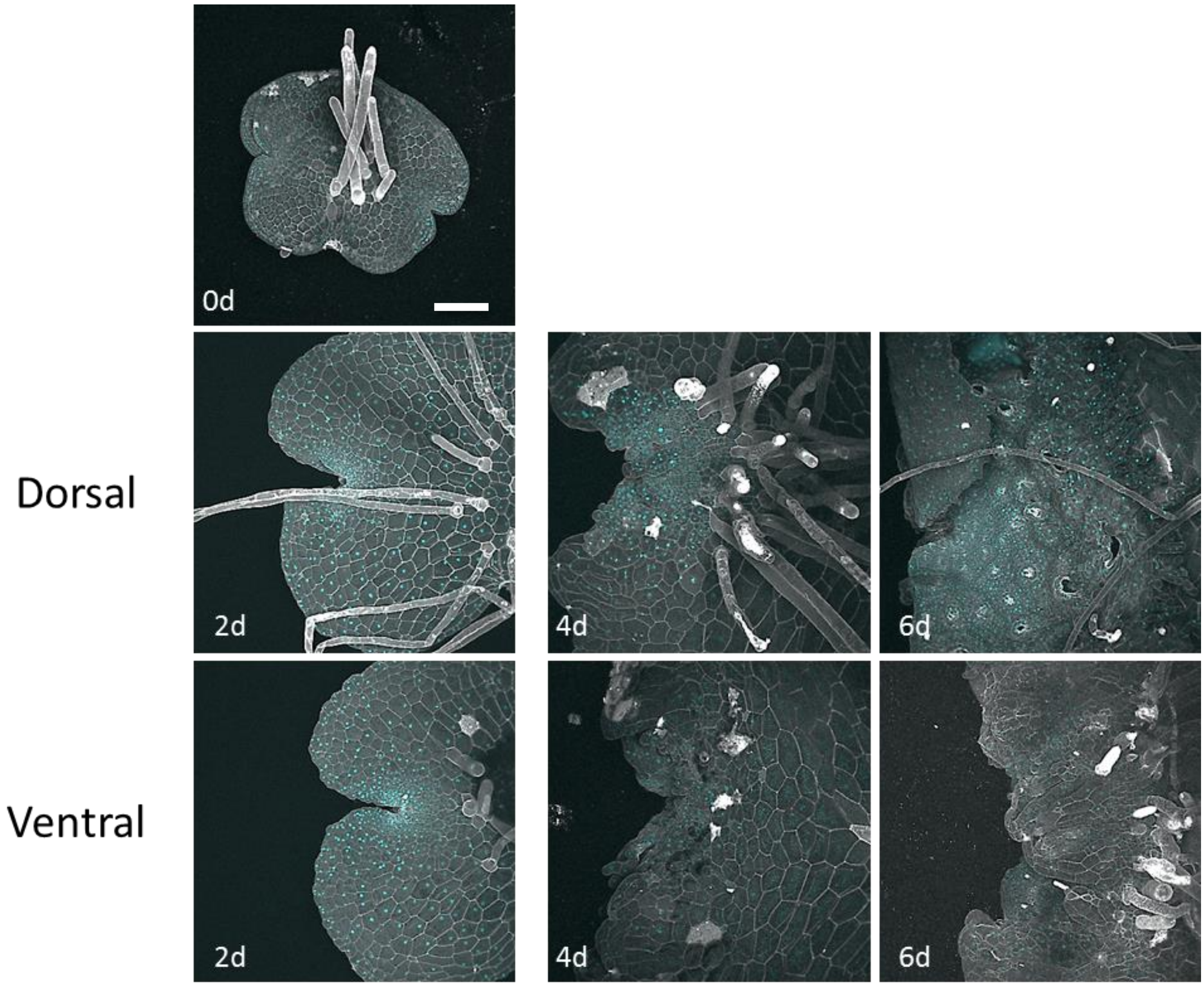


Figure 3.7 Mp*GL3* promoter is preferentially active in the dorsal side of the thallus in developing gemma. The Mp*GL3* promoter is expressed in all cells at the dorsal and ventral side of the thallus at 0d and 2d after gemmae plating, whereas it becomes preferentially expressed in the dorsal side of the thallus at 4d and 6d after gemmae plating. Confocal laser imaging of $_{pro}$Mp*GL3:3xCFP-NLS* gemmalings. Cells were stained with PI (propidium iodide). Nuclear localised CFP signal is shown in blue. Scale bar is 100 μm.

and 6 day-old gemmalings. At 0-days after gemma germination there is no dorsiventral polarity established. By 6 days after germination, dorsiventrality has developed with characteristic air pore complexes developing on the dorsal surface and rhizoids on the ventral surface (Figure 3.7). CFP fluorescence was detected throughout the gemma in 0-day-old gemma with no distinguishable dorsal and ventral thallus (Figure 3.7). 2 days after gemma germination, the CFP is expressed throughout the dorsal and ventral sides of the gemmaling, but at 4 days after gemma germination, the CFP fluorescence is higher in the dorsal side of the gemmaling compared to the ventral side, being actively repressed in the ventral side of the old tissue (Figure 3.7). At 6 days, dorsiventraliy has been established and air pores develop in the dorsal side of the thallus (Figure 3.7). At this stage, CFP fluorescence is mainly enriched in the dorsal side of the thallus while there is just a weak CFP fluorescence detected in the ventral side of the thallus (Figure 3.7). These results indicate that $_{pro}$Mp*GL3* is active in both dorsal and ventral sides of young gemma, but as gemma develops and dorsiventrality is established in the new tissue, the $_{pro}$Mp*GL3* becomes preferentially active in the dorsal surface of the thallus.

To test whether fluorescent proteins can be detected in both dorsal and ventral sides of the thallus during gemma development, I imaged $_{pro}$ *Ef1α:3xYFP-NLS* (ubiquitous promoter driving the expression of *3xYFP-NLS*) plants during the establishment of dorsiventrality using a confocal scanning laser microscope. In these plants, fluorescence could be detected in both dorsal and ventral sides of the thallus in the developing gemmaling at 0, 2, 4 and 6 days after gemma germination (Figure 3.8). This observation validates the specific pattern of $_{pro}$Mp*GL3* activity during early stages of thallus development.

Together, these data show that $_{pro}$Mp*GL3* activity is mainly restricted to the dorsal side of the thallus during and after the establishment of dorsiventrality. This is consistent with the hypothesis that Mp*GL3* regulates the development of dorsal epidermal cells.

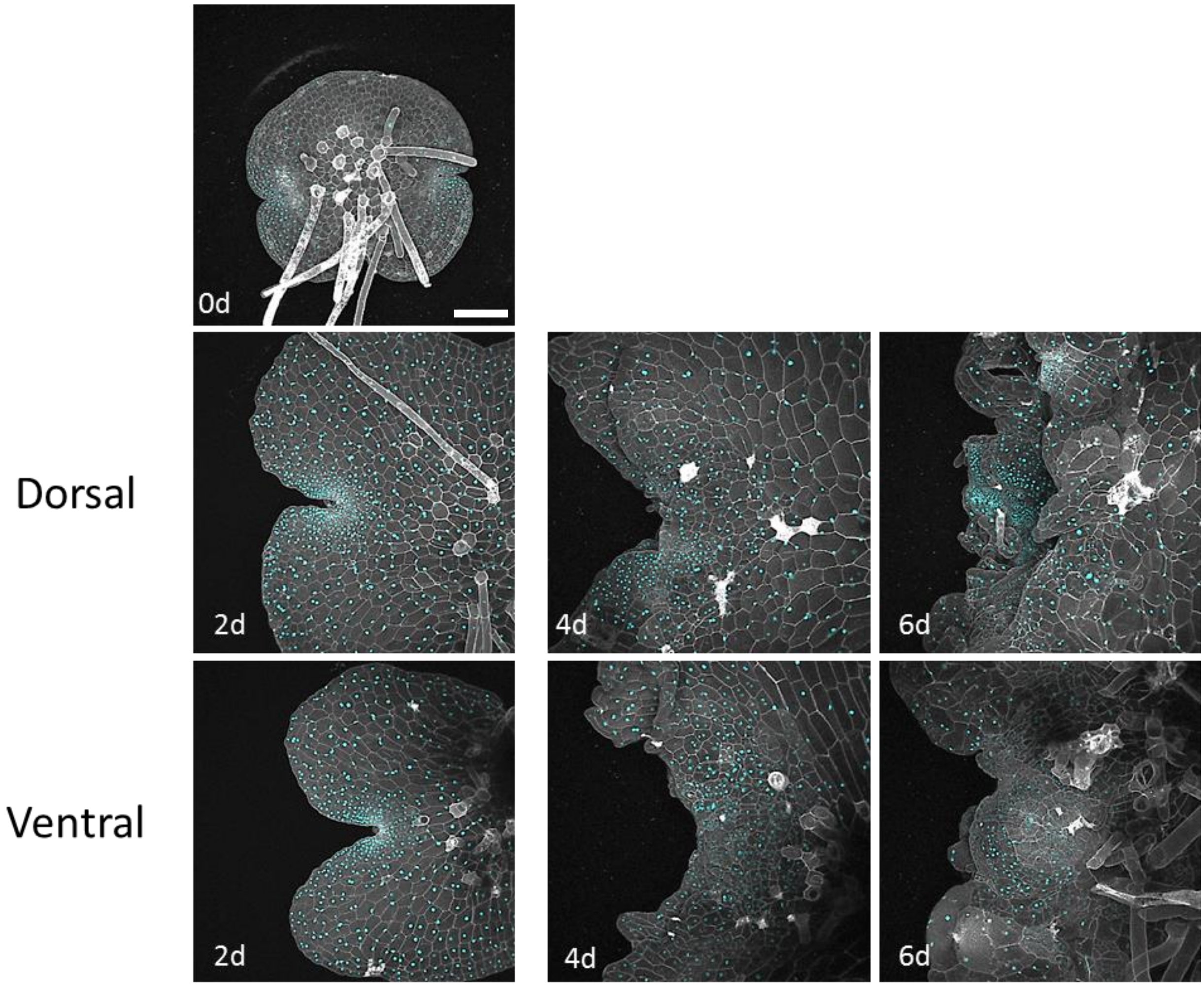


Figure 3.8 Mp*Ef1α* promoter is active in the dorsal and ventral sides of developing gemma. Mp*Ef1α* promoter is active in every cell of the ventral and dorsal thallus at 0d, 2d, 4d and 6d old gemmalings. Confocal laser imaging of ${}_{pro}$Mp*Ef1α :3xYFP-NLS* gemmalings. Cells were stained with PI (propidium iodide). Nuclear localised YFP is shown in blue. Scale bar is 100 μm.

3.4.4. Mp*GL3* expression is required for dorsal epidermal cell development

The expression pattern of Mp*GL3* suggests that this TF encoding gene is active in the meristem and the dorsal side of the developing thallus. To determine the function of Mp*GL3* in the development of *M. polymorpha* thallus, I generated two different artificial microRNA (amiR) constructs under the ubiquitous Mp*Ef1α* promoter in order to generate loss-of-function mutants by repressing the expression of MpGL3. These two amiR - *amiR*-Mp*GL3-1*^{Mpmir160} and *amiR*-Mp*GL3-2*^{Mpmir160} - target two different regions of the Mp*GL3* transcript (Appendix 2). These constructs were transformed and plants were assessed for phenotypical defects and compared to the WT plants. To test whether the expression of Mp*GL3* was indeed reduced in the amiR plants, I performed qRT-PCR to quantify the gene expression. This shows that steady state level of Mp*GL3* transcript in both independent amiR mutant lines is three times lower than in WT (Figure 3.9).

Plants transformed with the amiR constructs (*amiR*-Mp*GL3-1*^{Mpmir160} and *amiR*-Mp*GL3-2*^{Mpmir160}) developed smaller thalli than WT (Figure 3.9). Rhizoids are formed in the lower side of the thallus in both WT and amiR plants (Figure 3.10). The dorsal epidermis of WT plants develops air chambers comprising air pores in the middle (Figure 3.9 and 3.10). The development of these specialized dorsal epidermal cells is defective in amiR plants (Figure 3.9 and 3.10). Scanning electron micrographs and longitudinal median cross sections of the thallus show that indeed the dorsal epidermal cell layer of the amiR plants is defective when compared to the WT dorsal epidermis and the development of air chamber, comprising assimilatory filaments, is also defective in the amiR plants compared to WT plants (Figure 3.9 and 3.10). Dorsal epidermal cell shape and organization is impaired in the loss-of-function mutants, assimilatory filaments within air chambers are reduced in size and number, and air pore complex differentiation is defective (Figures 3.9, 3.10 and 3.11). Indeed, the cells that surround the

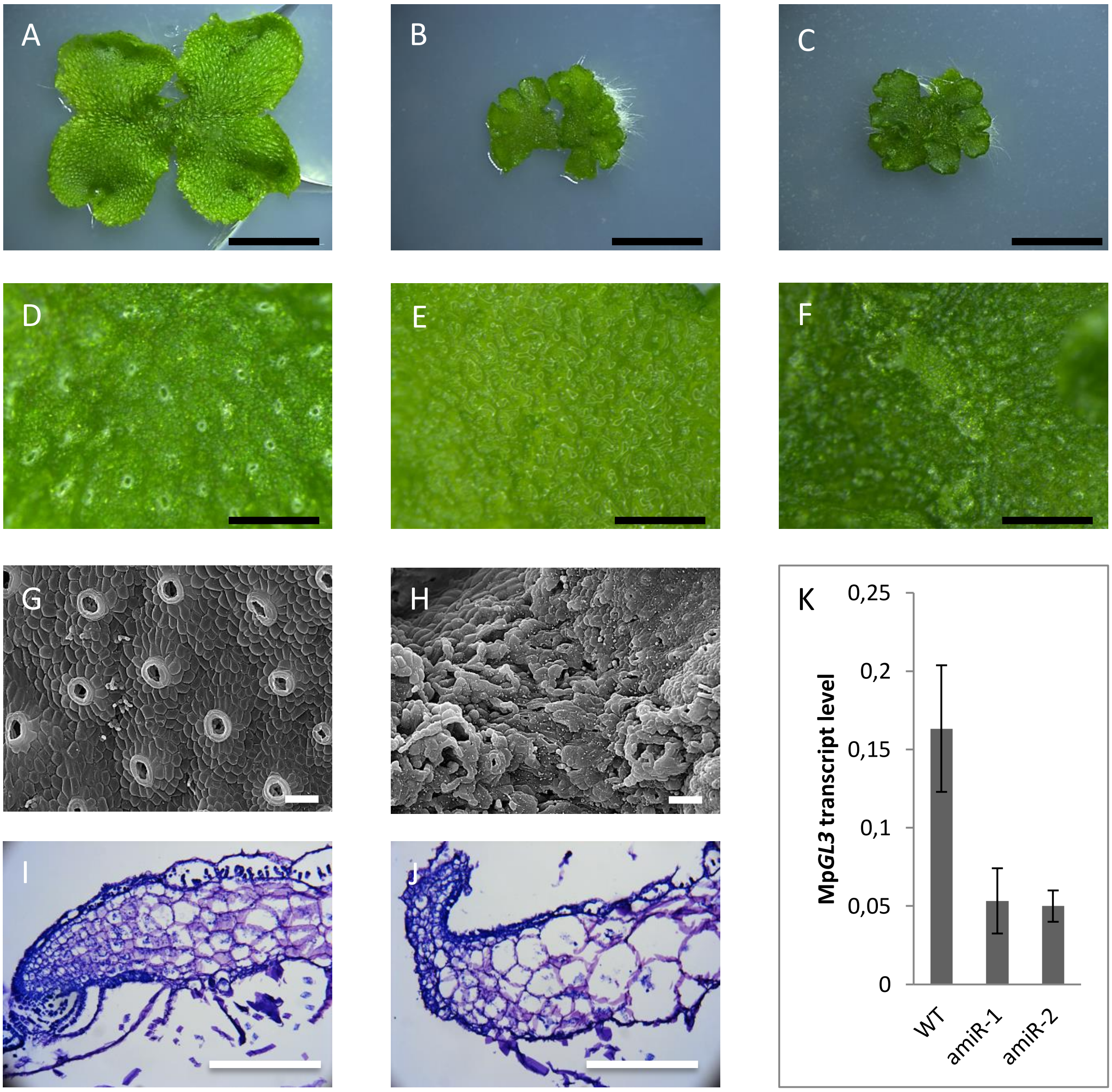


Figure 3.9 Reduced Mp*GL3* expression causes defective dorsal epidermal cell development. (A-C) Phenotype of 21-day-old thallus imaged on a stereomicroscope shows that *amiR*-Mp*GL3-1*^{Mpmir160} and *amiR*-Mp*GL3-2*^{Mpmir160} plants are smaller and more branched than WT plants. (A) WT; (B) *amiR*-Mp*GL3-1*^{Mpmir160} and (C) *amiR*-Mp*GL3-2*^{Mpmir160}. (D-F) Phenotype of 21-day-old thallus dorsal epidermis imaged on a stereomicroscope shows that air pore complex development is defective in *amiR*-Mp*GL3-1*^{Mpmir160} and *amiR*-Mp*GL3-2*^{Mpmir160} plants compared to WT plants. (D) WT; (E) *amiR*-Mp*GL3-1*^{Mpmir160} and (F) *amiR*-Mp*GL3-2*^{Mpmir160}. (G and H) Scanning electron microscope images of 21-day-old thallus shows that *amiR*-Mp*GL3-1*^{Mpmir160} plants have defective tissue differentiation (epidermis and air pore complex) at the dorsal side of the thallus compared to WT plants (G) WT and (H) *amiR*-Mp*GL3-1*^{Mpmir160} dorsal epidermis. (I and J) Median longitudinal sections of 10-day-old *amiR*-Mp*GL3-1*^{Mpmir160} plants show that air chamber and assimilatory filaments development is defective compared to WT plants (I) WT and (J) *amiR*-Mp*GL3-1*^{Mpmir160}. (K) Mp*GL3* expression level is reduced in *amiR*-Mp*GL3-1*^{Mpmir160} and *amiR*-Mp*GL3-2*^{Mpmir160} plants compared to WT plants. Mp*Actin* and Mp*Ef1α* were used as reference genes. Scale bars are 5 mm (A-C), 1 mm (D-F) and 100 µm (G-J). Scale bars are 5 mm (A-C), 1 mm (D-F) and 100 µm (G-J).

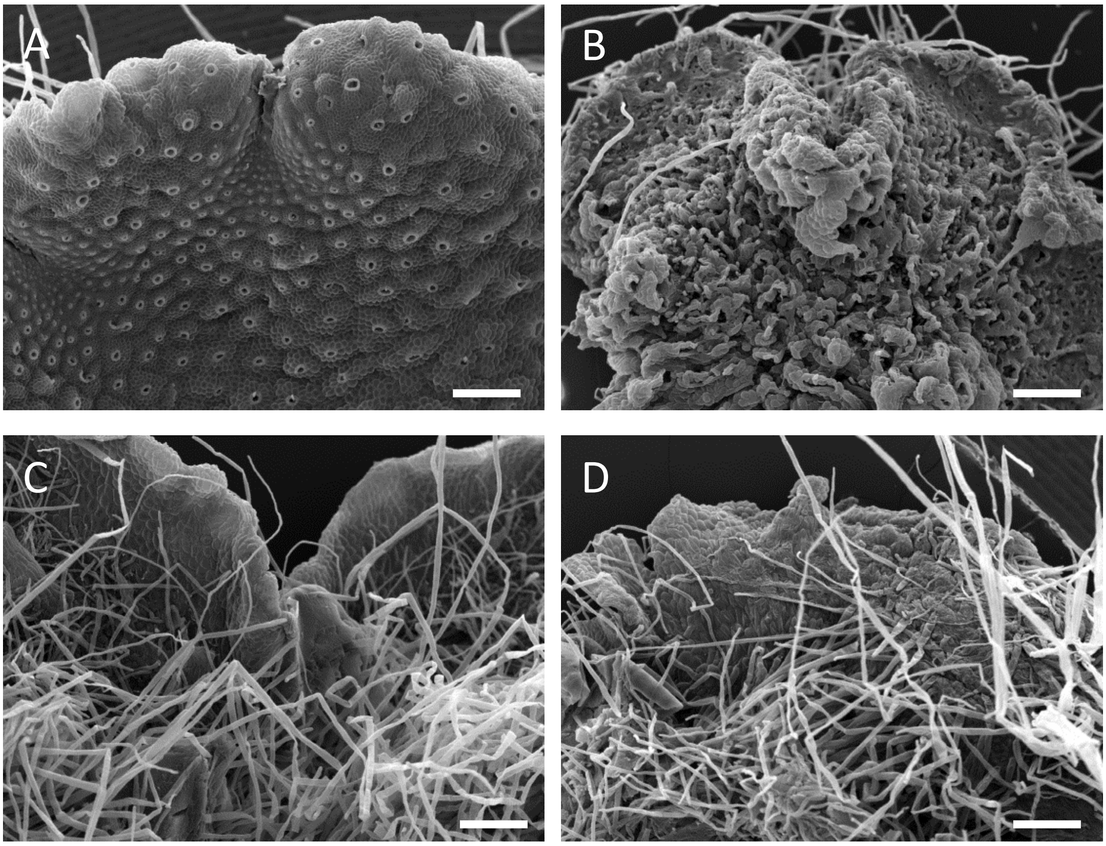


Figure 3.10 Reduced Mp*GL3* expression impairs dorsal epidermis differentiation but does not impair rhizoid development in the ventral side. Scanning electron microscope imaging of 21-day-old (A and C) WT and (B and D) *amiR*-Mp*GL3-1*^{Mpmir160} shows that the differentiation of air pore complexes and epidermal cells from the notch to the differentiation region is defective in *amiR*-Mp*GL3-1*^{Mpmir160} plants compared to WT, while there is no apparent defective phenotype at the ventral side of *amiR*-Mp*GL3-1*^{Mpmir160} plants compared to WT plants. (A-B) Dorsal view of the thallus and (C-D) ventral view of the thallus. Rhizoids develop in the ventral thallus in the WT and in the *amiR*-Mp*GL3-1*^{Mpmir160} mutant. Asterisks mark the meristematic notch. Scale bar is 500 µm.

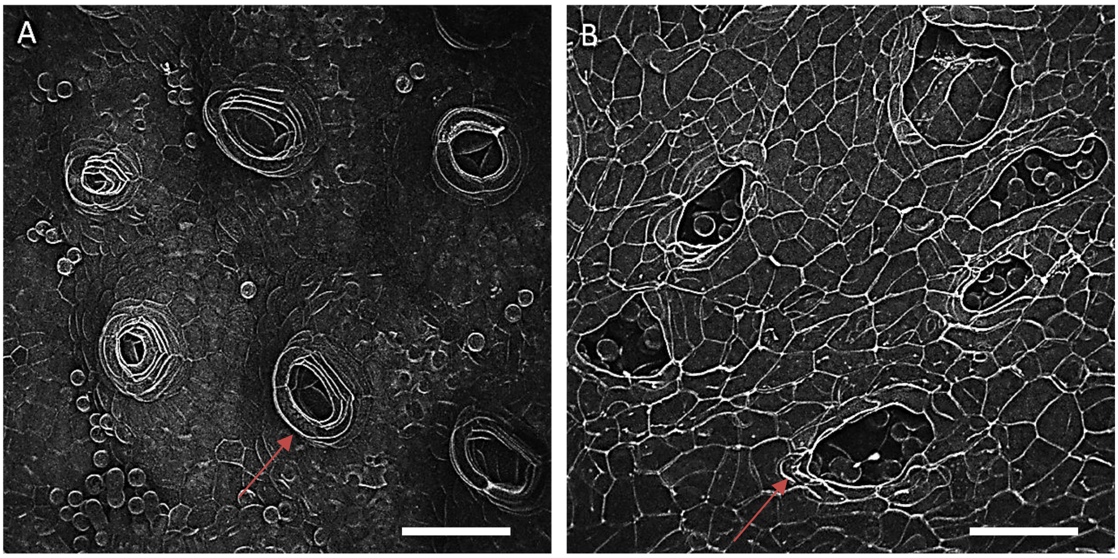


Figure 3.11 Reduced Mp*GL3* expression impairs air pore development. Confocal laser imaging of 14 day-old (A) WT and (B) *amiR*-Mp*GL3-1*^{Mpmir160} thallus shows that epidermal cell differentiation is defective and that the stereotypical arrangement of 4x4 barrel of cells that comprise the air pore complex is impaired in *amiR*-Mp*GL3-1*^{Mpmir160} plants compared to WT plants. Cells were stained with PI (propidium iodide). Red arrow points to air pore complex. Scale bars are 100 µm.

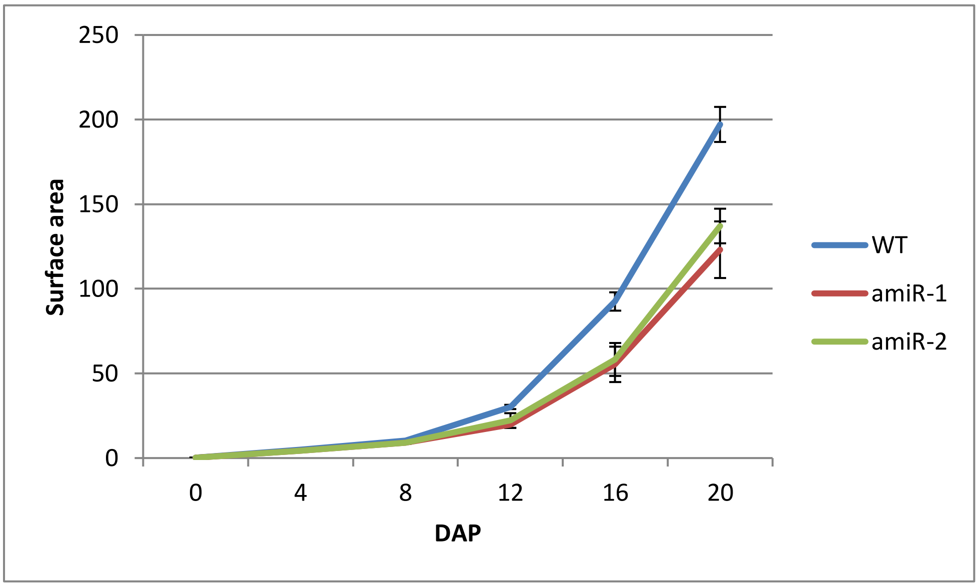
pore are defective in shape and often there are more than four cells comprising the ring of cells that surrounds the pore in the amiR mutant when compared to the WT (Figure 3.11). Moreover, the stereotypical cell divisions that give rise to the tier of four cells that forms the barrel-shape that surrounds the pore in the WT is absent in the amiR mutant (Figure 3.11). These observations are consistent with the hypothesis that MpGL3 controls the development of epidermal cells and dorsal structures in *M. polymorpha*. Together, these data suggest that Mp*GL3* gene expression is required for epidermal cell and air chamber differentiation in the dorsal thallus of *M. polymorpha*.

3.4.5. Mp*GL3* regulates thallus growth and branching

To understand why plants with reduced MpGL3 function are smaller than wild type I characterised the development of the thallus. I measured the increase in thallus surface area for 20 days after gemma germination. The increase in surface area of young plants developing from gemmae is identical in WT and amiR plants for the first 8 days after germination (surface area of 9 mm^2) (Figure 3.12). After 8 days of gemma development, the amiR plants surface area is reduced compared to WT. At 12, 16 and 20 days the WT shows 30, 93 and 187 mm^2 of surface area, while the amiR plants show approximately 20, 57 and 130 mm^2 of surface area (Figure 3.12). These results suggest that while dorsiventrality is not established, there is no difference between WT and amiR plants regarding thallus surface area, but after the establishment of dorsiventral polarity in WT, the growth of the thallus in the amiR plants is impaired.

amiR plants also develop meristematic notches than wild type (Figure 3.9). Meristematic notches are surrounded by two growing lobes. Lobes constitute the growing regions of the thallus and two neighbouring lobes develop and grow from the apical meristematic notch and the lobes further branch in the apex by the production of a new apical

A



B

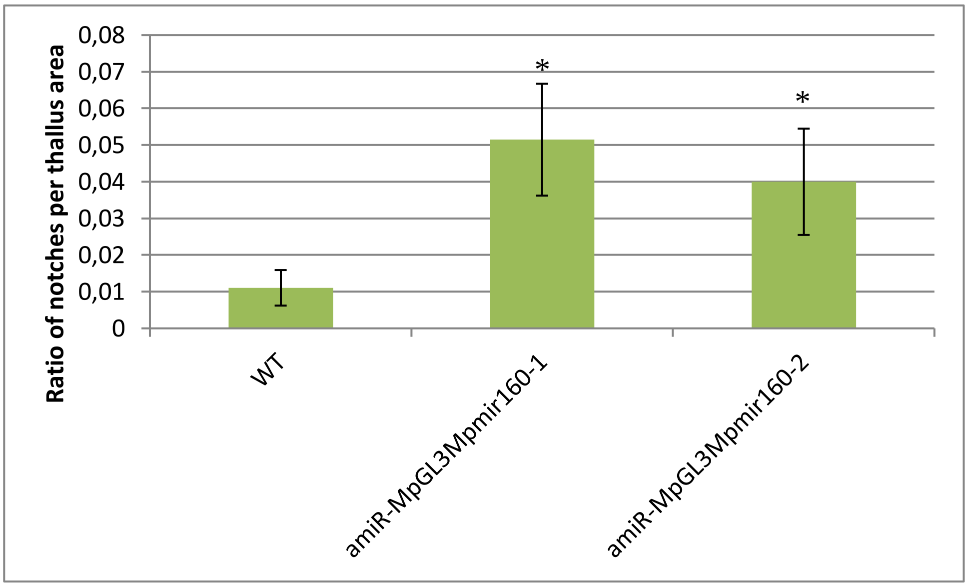


Figure 3.12 Reduced Mp*GL3* expression causes defects in thallus growth and branching. *amiR*-Mp*GL3-1*^{Mpmir160} and *amiR*-Mp*GL3-2*^{Mpmir160} plants are smaller and have more notches per thallus area compared to WT. (A) Thallus surface area of WT, *amiR*-Mp*GL3-1*^{Mpmir160} and *amiR*-Mp*GL3-2*^{Mpmir160} plants. The area is indicated in mm^2, time in days after plating (DAP). Error bars indicate SD of 10 plants. (B) Ratio of meristematic notches per thallus area in 21-day-old WT, *amiR*-Mp*GL3-1*^{Mpmir160} and *amiR*-Mp*GL3-2*^{Mpmir160} plants. Error bars indicate SD of 10 plants. Asterisks represent significant difference compared to WT (t-test $P < 0.005$).

meristematic notch that develops new lobes (Davidonis and Munroe, 1972). The ratio between the number of lobes and the thallus area is four to five times higher in the amiR mutants compared to WT (Figure 3.12), suggesting that Mp*GL3* activity is necessary to represses thallus branching and meristematic activity. Taken together, these results suggest that Mp*GL3* expression is required for the development of dorsal epidermal cells and is necessary to regulate thallus growth and branching.

3.4.6. Mp*GL3* is not sufficient to regulate thallus development

Given that Mp*GL3* expression is necessary to regulate thallus growth and the differentiation of dorsal epidermal cells, I postulated that plants whose expression of Mp*GL3* is more abundant can develop bigger thallus, with fewer lobes and abnormal epidermal cell development when compared to the WT. To test whether ectopic constitutive expression of Mp*GL3* is sufficient to promote the development of thallus and epidermal cell differentiation, I generated two transgenic plants that contain the ubiquitous *Ef1α* promoter driving the CDS of Mp*GL3*; $_{pro}$ *Ef1α:*Mp*GL3-1* and $_{pro}$ *Ef1α:*Mp*GL3-2.* To test whether Mp*GL3* transcript is more abundant in these two transgenic plants, I quantified gene expression using qRT-PCR. This shows that Mp*GL3* transcript is more abundant (up to seven fold increase) in both transgenic mutant plants when compared to the WT (Figure 3.13).

There is no evident difference between the phenotype of WT and the two overexpressing lines (Figure 3.13). Thallus size and shape of the overexpressing lines are similar to those of WT (Figure 3.13). Furthermore, the differentiation of dorsal epidermal cells is not affected in both overexpressing lines when compared to WT (Figure 3.13). Moreover, I measured the thallus surface area since gemma germination in order to validate that gemma growth is not affected in the transgenic Mp*GL3*; $_{pro}$ *Ef1α:*Mp*GL3-1* and $_{pro}$ *Ef1α:*Mp*GL3-2* plants. Indeed, the thallus surface area is not affected in Mp*GL3*; $_{pro}$

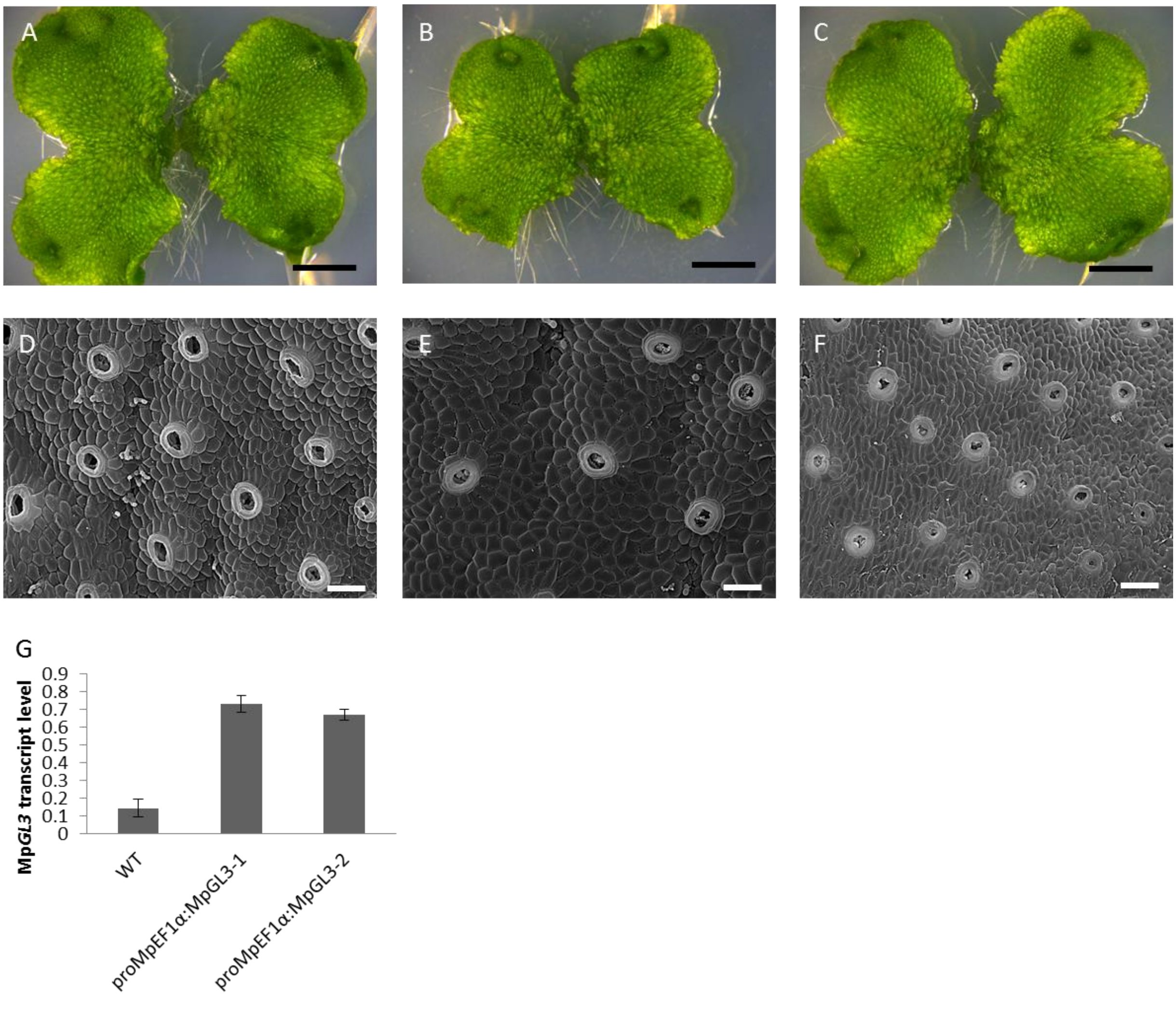


Figure 3.13 Ectopic expression of *MpGL3* does not affect thallus development. Thallus growth and epidermal cell development is similar in $_{pro}$ *Ef1α:*Mp*GL3-1*, $_{pro}$ *Ef1α:*Mp*GL3-2* and WT plants. (A-C) Phenotype of 21-day-old thallus imaged on a stereomicroscope. (A) WT; (B) $_{pro}$ *Ef1α:*Mp*GL3-1* and (C) $_{pro}$ *Ef1α:*Mp*GL3-2*. (D-F) Scanning electron microscope images of 21-day-old (D) WT; (E) $_{pro}$ *Ef1α:*Mp*GL3-1* and (F) $_{pro}$ *Ef1α:*Mp*GL3-2* dorsal epidermis. (G) Mp*GL3* expression level is elevated in $_{pro}$ *Ef1α:*Mp*GL3-1* and $_{pro}$ *Ef1α:*Mp*GL3-2* plants. Mp*Actin* and Mp*Ef1α* were used as reference genes. Scale bars are 5 mm (A-C), 100 µm (D-F).

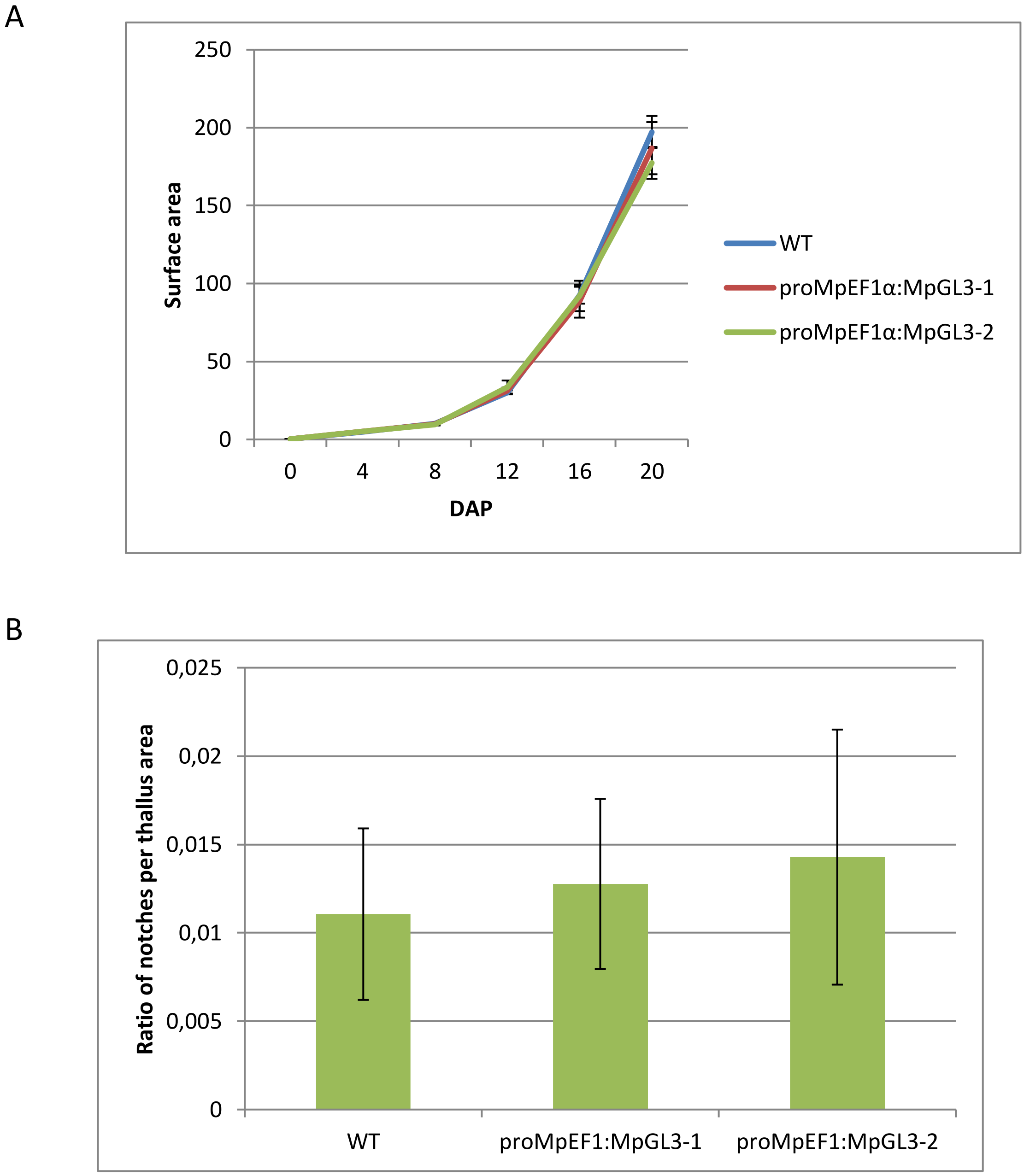


Figure 3.14 Ectopic expression of *MpGL3* does not affect thallus growth and branching. (A) Thallus surface area of WT, $_{pro}$ *Ef1α:*Mp*GL3-1* and $_{pro}$ *Ef1α:*Mp*GL3-2* plants is similar. The area is indicated in mm^2, time in days after plating (DAP). Error bars indicate SD of 10 plants. (B) Ratio of meristematic notches per thallus area in 21-day-old WT, $_{pro}$ *Ef1α:*Mp*GL3-1* and $_{pro}$ *Ef1α:*Mp*GL3-2* plants is identical. Error bars indicate SD of 10 plants.

*Ef1α:*Mp*GL3-1* and $_{pro}$ *Ef1α:*Mp*GL3-2* mutants, suggesting that the growth of gemma since germination is not altered in these transgenic plants when compared to the WT (Figure 3.14). To further validate that branching and meristematic activity are not affected, I quantified the ratio of meristematic notches per thallus area and both transgenic lines and the WT show similar ratios of meristematic notches per thallus area (Figure 3.14). Taken together, these data suggest that ectopic expression of Mp*GL3* does not affect the development of the thallus. Altogether, these evidences suggest that Mp*GL3* is not sufficient to regulate thallus growth and branching in *M. polymorpha*.

3.4.7. Expression of the dominant repressor *MpGL3-SRDX* impairs the development of gemma cups

I have demonstrated that Mp*GL3* expression is required but not sufficient to regulate thallus growth and the development of dorsal epidermal cells. To further investigate the role of Mp*GL3* in the development of thallus in *M. polymorpha*, I generated dominant negative mutants that dominantly repress the transcription of the target genes of MpGL3. To this end, I fused the EAR (ethylene-responsive element binding factor-associated amphiphilic repression) repressor domain (Hiratsu et al., 2003), SRDX, to the C-terminal end of Mp*GL3* CDS under the control of the *Ef1α* promoter. Two independent transformant lines ($_{pro}$ *Ef1α:*Mp*GL3-SRDX-1* and $_{pro}$ *Ef1α:*Mp*GL3-SRDX-2*) were isolated and further phenotypically analysed. In order to confirm that these two transformant lines expressed the Mp*GL3-SRDX* transgene, the transcript abundance was quantified using qRT-PCR. Both transgenic lines showed expression of the Mp*GL3-SRDX* transgene (Figure 3.15),

These plants develop normal thallus with no apparent defective epidermal cell development (Figure 3.15). However, these transformant lines develop abnormal gemma

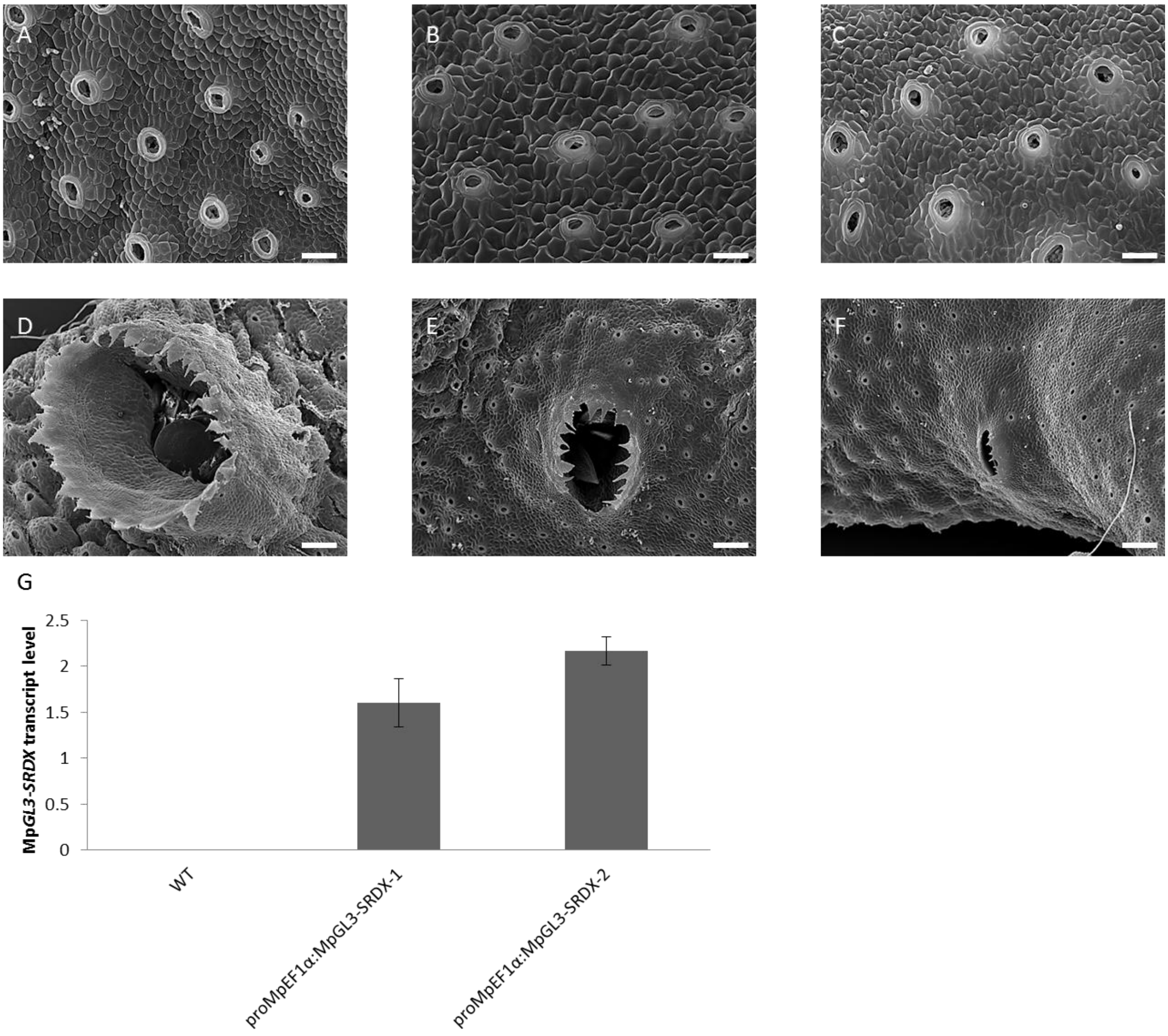


Figure 3.15 Expression of the dominant repressor *MpGL3-SRDX* causes the development of defective gemmae cups. (A-C) Scanning electron microscope (SEM) images of 28-day-old (A) WT; (B $_{pro}$ *Ef1α:*Mp*GL3-SRDX-1* and (C) $_{pro}$ *Ef1α:*Mp*GL3-2* dorsal epidermis shows that air pore complex and epidermal cell development is unaffected in transgenic plants compared to WT. (D-F) Scanning electron microscope images of 21-day-old shows that gemmae cup development is impaired in transgenic plants compared to WT. The "cup-shape" structure fails to develop in transgenic plants (D) WT; (E) $_{pro}$ *Ef1α:*Mp*GL3-SRDX-1* and (F) $_{pro}$ *Ef1α:*Mp*GL3-SRDX-2* gemma cups. Scale bars are 100 µm (A-C) and 30 µm (D-F). (G) Mp*GL3-SRDX* is expressed in $_{pro}$ *Ef1α:*Mp*GL3-SRDX-1* and $_{pro}$ *Ef1α:*Mp*GL3-SRDX-2* plants (not detected in WT). Mp*Actin* and Mp*Ef1α* were used as reference genes.

cups with the absence of the cup-like structure that develops in the WT plants (Figure 3.15). These results show that the expression of the dominant repressor Mp*GL3-SRDX* causes the development of impaired gemma cups. Taken together, these results suggest that Mp*GL3* is a positive regulator of thallus growth, and morphological characters that develop in the dorsal side of the thallus.

3.4.8. **MpGL3 positively regulates the expression of the air pore regulator Mp*WIP***

To assess if MpGL3 regulates the expression of TFs that control thallus growth and air pore development, I performed RT-qPCR to quantify the transcript abundance of Mp*RKD* (positive regulator of thallus growth), Mp*WIP* (positive regulator of air pore complex development) and Mp*RSL1* (positive regulator of rhizoid development) in the WT and *amiR*-Mp*GL3-1*^{Mpmir160} (amiR mutant) (Jones and Dolan, 2017; Proust et al., 2016; Rövekamp et al., 2016). The Mp*WIP* transcript abundance is two times higher in the WT comparing with the amiR mutant (Figure 3.16), suggesting that Mp*WIP* transcription is positively regulated by MpGL3. There are no differences in relative expression of Mp*RKD* and Mp*RSL1* in the WT and *amiR*-Mp*GL3-1*^{Mpmir160}, suggesting that Mp*RKD* and Mp*RSL1* act transcriptionally independent of MpGL3. Moreover, this result supports the hypothesis that MpGL3 does not control rhizoid development, since the transcript abundance of Mp*RSL1* is similar in the WT and amiR mutant. Altogether, these observations suggest that MpGL3 positively regulate the transcription of Mp*WIP* to control the differentiation of air pore complexes in *M. polymorpha*.

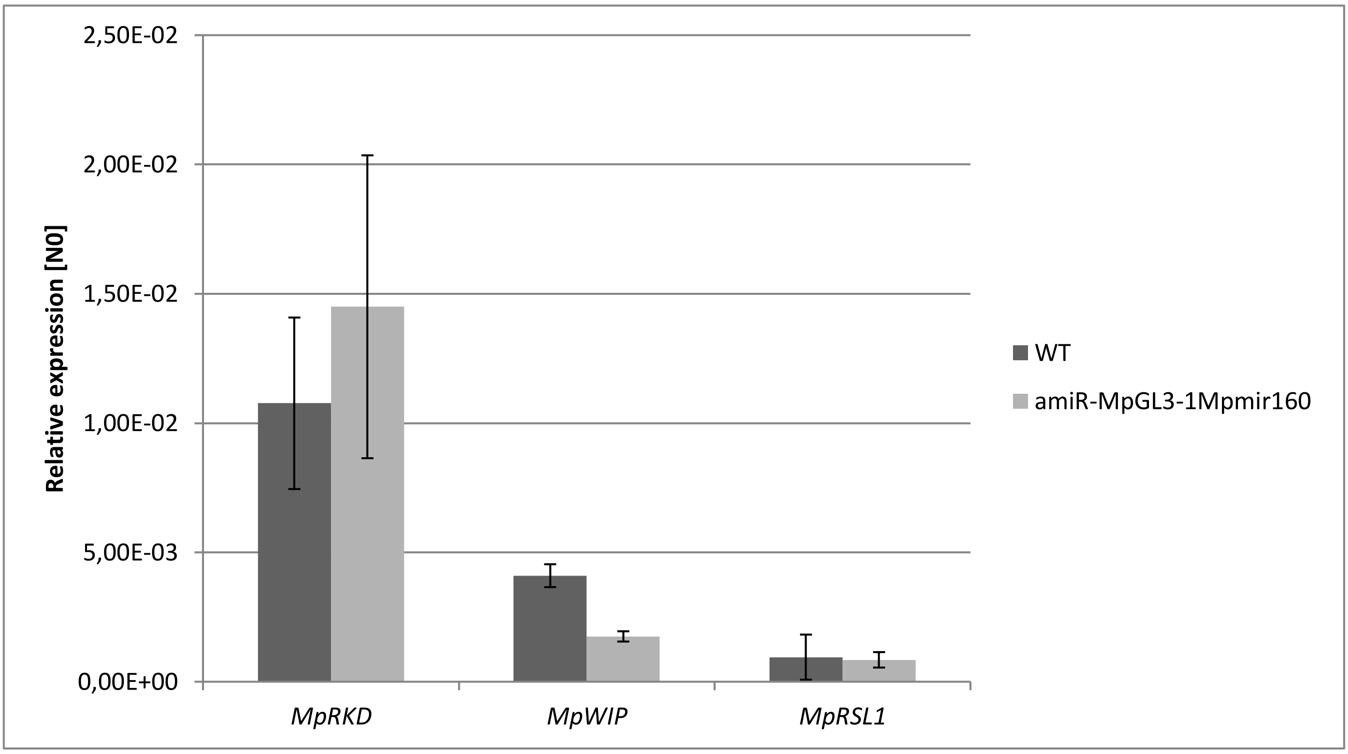


Figure 3.16 MpGL3 positively regulates the expression of Mp*WIP*. Mp*RKD*, Mp*WIP* and Mp*RSL1* transcript levels measured by RT-qPCR in WT and *amiR*-Mp*GL3-1*^{Mpmir160} plants shows that Mp*WIP* transcript is reduced in *amiR*-Mp*GL3-1*^{Mpmir160} plants compared to WT. Mp*Actin* was used as reference gene. Error bars show standard deviation of 3 biological replicates.

3.4.9. Mp*GL3* transcription is dependent on auxin concentration

Auxin mediates various developmental processes in the thallus of *M. polymorpha* (Eklund et al., 2015; Flores-Sandoval et al., 2015; Kato et al., 2015; Kaul et al., 1962; Maravolo, 1976). These include thallus growth and shape, establishment of dorsiventral polarity, epidermal cell differentiation and gemma cup development, which I have demonstrated that are regulated by Mp*GL3*. I hypothesize that Mp*GL3* acts downstream of the auxin-mediated transcriptional pathway that regulates these morphological and developmental characters. Therefore, I postulate that the Mp*GL3* transcript abundance in the thallus of *M. polymorpha* is dependent on the auxin concentration. To test this hypothesis, I have performed RT-qPCR on plants grown on media containing 10 µM of the auxins Indole-3-acetic acid (IAA) and 1-Naphthaleneacetic acid (NAA), and 100 µM of the auxin biosynthesis inhibitor L-Kynurenine (L-Kyn). The results of the RT-qPCR show that the transcript of Mp*GL3* is less abundant in plants grown on 10 µM of the auxins IAA and NAA, and 100 µM of the auxin biosynthesis inhibitor L-Kyn (Figure 3.17). These results suggest that the abundance of Mp*GL3* transcript decreases when auxin homeostasis is disturbed compared to untreated plants. Taken together, these data indicate that auxin modulates the expression of Mp*GL3* in a concentration-dependent manner (Figure 3.17). This suggests that Mp*GL3* might act downstream of auxin signalling in order to control thallus growth and the development of dorsal characters in *M. polymorpha*.

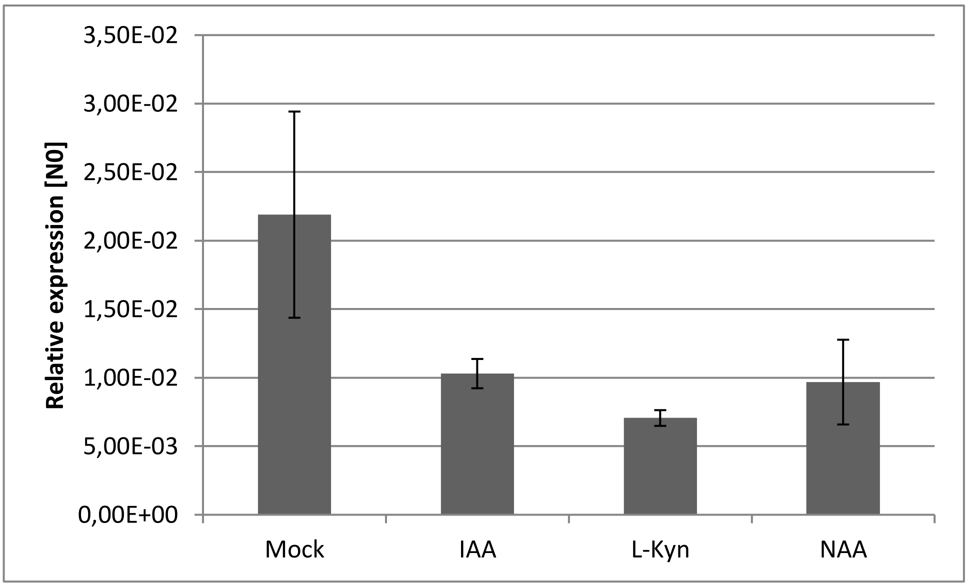


Figure 3.17 Mp*GL3* transcript abundance is dependent on auxin concentration. *MpGL3* transcript levels are reduced in exogenous auxin treated plants and in plants treated with auxin biosynthesis inhibitor comparing with mock control plants. *MpGL3* transcript was detected using RT-qPCR on 21-day-old WT plants grown on 10 µM of NAA, IAA and 100 µM of L-Kyn for 2 days. The constitutive and ubiquitous Mp*ACT* gene was used as reference. Error bars show standard deviation of 3 biological replicates.

3.4.10. An auxin - Mp*GL3* transcriptional negative feedback loop regulate thallus growth and dorsal tissue differentiation

To further test the auxin-mediated transcriptional regulation of Mp*GL3* transcript abundance, I quantified the Mp*GL3* transcript abundance in a variety of mutants with defective auxin biosynthesis or signalling. These include Mp*arf1*, *pro*EF1α:amiRMpAUX/IAA, *pro*EF1α:MpYUC2 and *pro*EF1α:MpTPL-D34$^{\mathrm{MpIAA}}$. The loss-of-function Mp*arf1* is insensitive to auxin, and the auxin signalling is therefore downregulated (Flores-Sandoval et al., 2015; Sugano et al., 2014). The *pro*EF1α:amiRMpAUX/IAA mutant, is hypersensitive to auxin, and the auxin signalling is constitutively upregulated (Flores-Sandoval et al., 2015). The *pro*EF1α:MpYUC2 mutant produces more auxin than the WT, increasing the concentration of endogenous auxin in the plant (Eklund et al., 2015). The *pro*EF1α:MpTPL-D34$^{\mathrm{MpIAA}}$ is insensitive to auxin, therefore the auxin signalling cascade is downregulated (Flores-Sandoval et al., 2015). The Mp*GL3* transcript abundance, assessed by RT-qPCR, is higher (more than 3-fold increase) in *pro*EF1α:amiRMpAUX/IAA and *pro*EF1α:MpYUC2 plants, whereas there is no difference in transcript abundance between the WT and Mp*arf1* and *pro*EF1α:MpTPL-D34$^{\mathrm{MpIAA}}$ (Figure 3.18). This result suggests that Mp*GL3* expression is positively regulated by auxin and acts downstream of auxin signalling.

In order to test whether there is a transcriptional feedback mechanism between auxin and MpGL3, I performed RT-qPCR in WT and *amiR*-Mp*GL3-1*^{Mpmir160} mutant to detect the transcript abundances of Mp*IAA*, Mp*YUC2,* Mp*ARF3* and Mp*TPL* in these two genetic backgrounds. The transcript abundances of Mp*TPL* are two-fold lower in *amiR*-Mp*GL3-1*^{Mpmir160} compared to WT, whereas Mp*IAA*, Mp*YUC2* and Mp*ARF3* are not different between the WT and *amiR*-Mp*GL3-1*^{Mpmir160} (Figure 3.18). This suggests that MpGL3 positively regulates the transcription of the auxin co-repressor Mp*TPL* in the thallus of *M. polymorpha*. Altogether, these results suggest that auxin signalling positively regulates the expression of

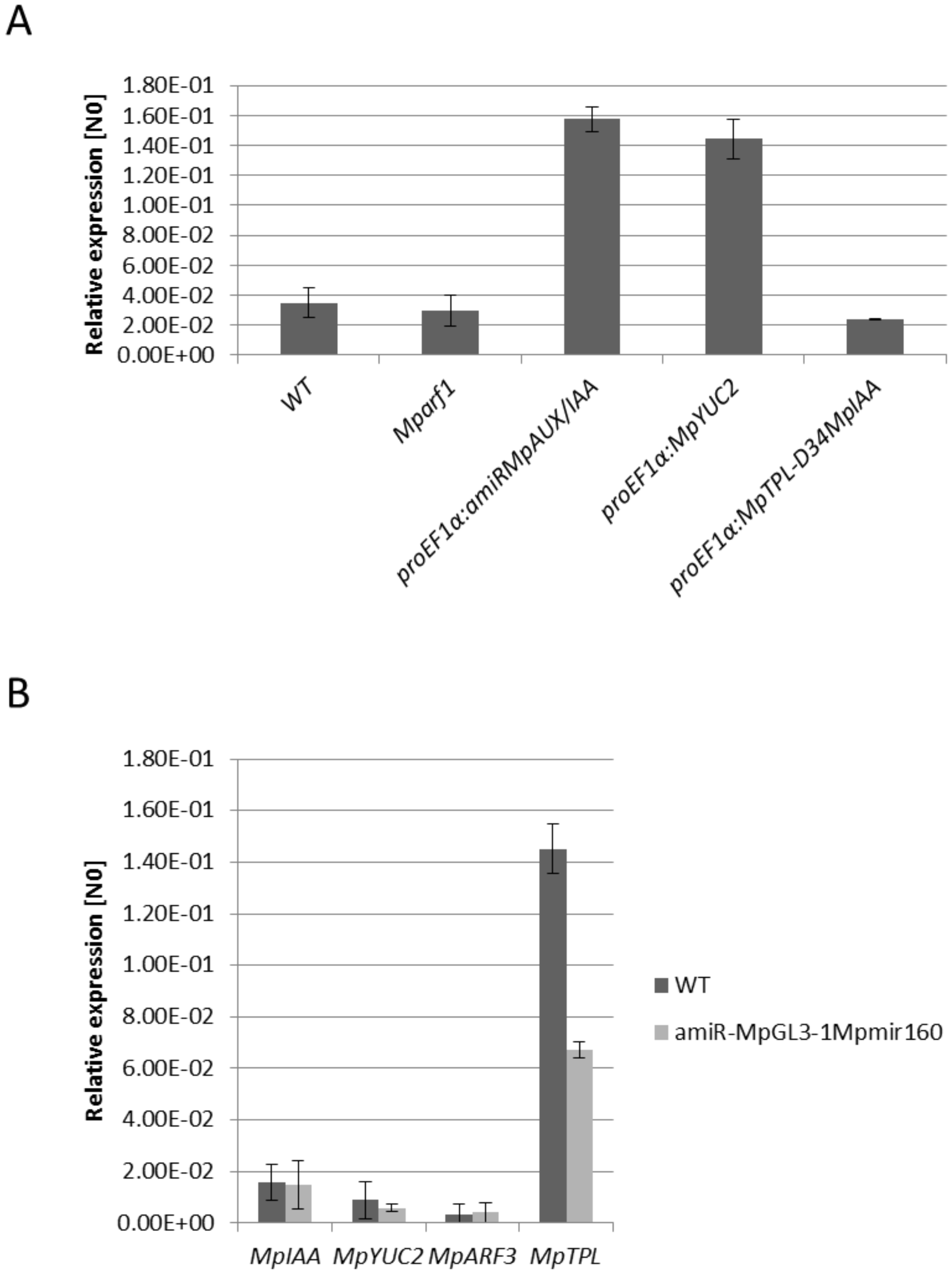


Figure 3.18 Auxin and MpGL3 transcriptional regulation. Mp*GL3* transcript levels are more abundant in *pro*EF1α:amiRMpAUX/IAA and *pro*EF1α:MpYUC2 plants compared to WT plants. Mp*TPL* transcript is less abundant in *amiR*-Mp*GL3-1*^{Mpmir160} compared to WT plants. (A) Mp*GL3* transcript was detected using RT-qPCR on 21-day-old WT, Mp*arf1*, *pro*EF1α:amiRMpAUX/IAA, *pro*EF1α:MpYUC2 and *pro*EF1α:MpTPL-D34$^{\text{MpIAA}}$ mutants. The constitutive and ubiquitous Mp*ACT* gene was used as reference. (B) Mp*IAA*, Mp*YUC2*, Mp*ARF3* and Mp*TPL* transcript were detected using RT-qPCR on 21 day-old WT and *amiR*-Mp*GL3-1*^{Mpmir160} mutant. The constitutive and ubiquitous Mp*ACT* gene was used as reference. Error bars show standard deviation of 3 biological replicates.

Mp*GL3*, which in turn, positively regulate the transcription of the auxin signalling co-repressor Mp*TPL,* which results in a negative regulation of auxin signalling in the thallus of *M. polymorpha*. It is possible to hypothesise that this negative feedback loop, between auxin and MpGL3, provides developmental robustness in governing thallus growth and dorsal epidermis and tissue differentiation.

3.5. Discussion

The origin of the epidermal cell layer occurred concomitantly with the evolution of complex multicellular 3D tissues during the early colonization of the land by plants (Graham et al., 2000; Kenrick and Crane, 1997). IIIf bHLH TFs form a transcriptional complex together with WD-40 and MYB TFs that regulates epidermal cell development in angiosperms (Zhang et al., 2003). IIIf bHLH TFs originated before or during the early colonization of the land by plants (Chapter 2), suggesting that IIIf bHLH TFs might have controlled epidermal cell development in the first land plants. To test this hypothesis I carried out functional analysis of IIIf bHLH TF in the liverwort *M. polymorpha* in order to infer the ancestral function of IIIf bHLH TFs. This can be achieved by comparing the function in *M. polymorpha* with the previously described function of this bHLH subfamily in angiosperms. The analyses presented in this Chapter support the hypothesis that IIIf bHLH TFs are ancient regulator of epidermal cell development, predating the diversification of extant lineages of land plants, and that they have been recruited during land plant evolution to control other developmental processes in specific land plant lineages.

The IIIf bHLH TF subfamily originated sometime between the divergence of streptophyte algae and land plants, and has been conserved throughout land plant evolution (Chapter 2). The phylogenetic analysis of IIIf bHLH TF subfamily in land plants demonstrated that this subfamily has diverged into two classes (I and II) during the course of land plant evolution; Class I originated after the divergence of lycophytes and angiosperms while Class II originated in a common ancestor of extant land plant lineages. These two classes are functionally equivalent, since members of both classes complex with WD-40 and MYB TFs to control epidermal cell differentiation in angiosperms (Baudry et al., 2006; Bernhardt et al., 2005; Zhang et al., 2003). The liverwort *M. polymorpha* encodes one IIIf bHLH TF that belongs to the IIIf Class II and contains highly conserved domains present in other IIIf bHLH TFs, suggesting that the biochemical properties, and therefore the function, of these TFs might be conserved across land plants (between bryophytes and angiosperms).

If epidermal cell development was regulated by IIIf bHLH TFs in early land plants, I would expect that IIIf bHLH TFs control epidermal cell development in the earliest divergent lineages of land plants (the bryophytes) as they do in the later divergent lineages of land plants, the angiosperms. The IIIf bHLH TF in the liverwort *M. polymorpha,* here designated Mp*GL3,* is preferentially expressed in the meristematic notch and in the dorsal side of the thallus, giving support to the hypothesis that IIIf bHLH TF control the development of dorsal epidermal cells in early divergent lineages of land plants. The analysis of the gene function, using mutants with decreased Mp*GL3* expression, shows that indeed this TF encoding gene is required for the development of epidermal cells that develop in the dorsal surface of the thallus (both air pore and non-air pore cells). These amiR mutants also developed smaller thallus with more lobes than the WT, which is consistent with the expression of Mp*GL3* observed in the meristematic notch. Altogether these results suggest that MpGL3 controls the development of epidermal cells and thallus growth (Figure 3.19). Since IIIf bHLH TFs regulate epidermal cell development across land plants, I postulate that they were active in the outermost cell layer of the first land plants to regulate its development.

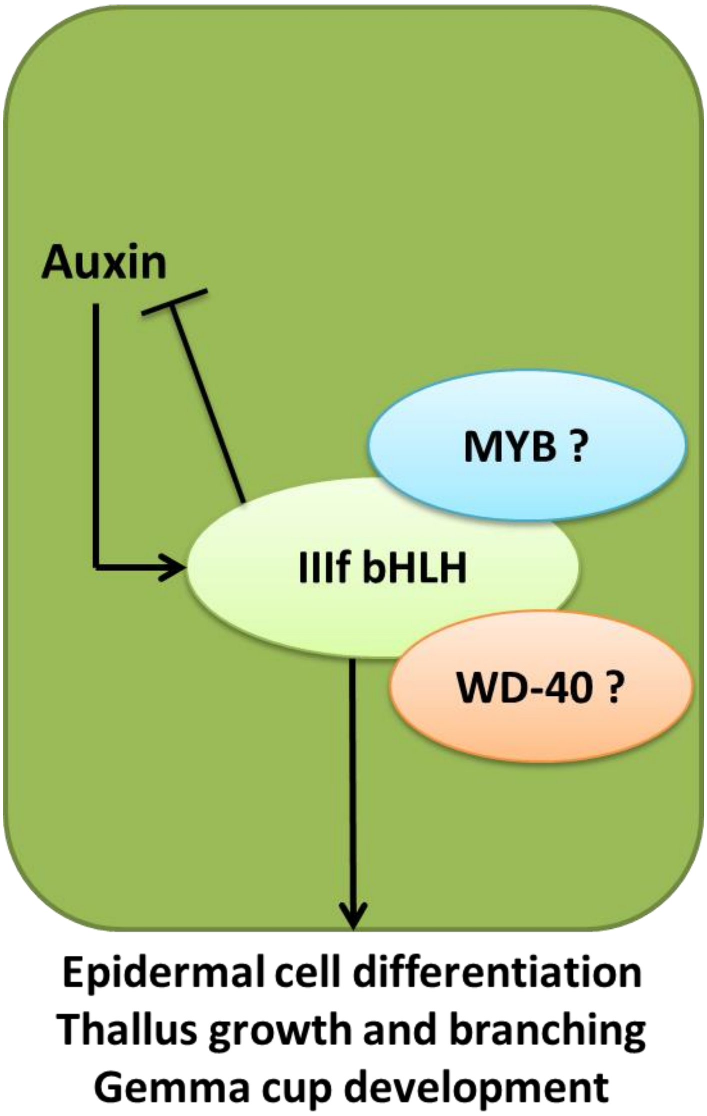


Figure 3.19 Proposed model of IIIf bHLH function in *M. polymorpha*. The expression of IIIf bHLH is modulated by auxin, which in turn transcriptionally repress auxin signalling. IIIf bHLH TF regulate epidermal cell differentiation, thallus growth and branching, and gemma cup development. The *M. polymorpha* IIIf bHLH TF contains domains that are required for heterodimerization with MYB and WD-40 TFs in angiosperms. However, the presence and role of the IIIf bHLH-MYB-WD-40 transcriptional complex in *M. polymorpha* is not known. For detailed description please read section 3.5.

Analysis of the constitutively expressed dominant repressor mutant Mp*GL3*-SRDX shows that gemma cups are defective when compared to the WT. This suggests that Mp*GL3* is required for the development of dorsal characters. However, contrarily to the phenotype of amiR mutant plants, the dorsal epidermal cell layer is not defective in the Mp*GL3*-SRDX. Moreover, the amiR mutants develop gemma cups that resemble WT gemma cups (data not shown). The difference between the amiR and Mp*GL3*-SRDX dominant repressor mutant phenotypes could be explained by genetic redundancy in the control of gemma cup development. Another possibility is that reduced expression of Mp*GL3* alone is not required for normal development of gemma cups. Furthermore, constitutive expression of the Mp*GL3*-SRDX dominant repressor could repress the expression of genes involved in gemma cup development by unspecific binding to their regulatory sequences. However, these two alternative scenarios are just hypothetical explanations for the difference between the amiR and Mp*GL3*-SRDX dominant repressor mutant phenotypes. In sum, IIIf bHLH TFs appear to regulate dorsal morphological characters development in *M. polymorpha* (Figure 3.19).

IIIf bHLH TFs have been recruited during land plant evolution to control other developmental pathways in specific lineages of land plants. I demonstrated that Mp*GL3* regulates thallus branching in *M. polymorpha*. Since Mp*GL3* is expressed in the meristematic notch, it is possible that MpGL3 is active in the meristem to regulate lobe growth and establishment of a new meristem in the growing lobes. This means that Mp*GL3* positively regulates the growth of the lobe and inhibits the formation of new meristems in the growing lobes. This is consistent with the observation that reduced expression Mp*GL3* impairs thallus growth, suggesting that indeed Mp*GL3* is required for thallus growth.

Moreover, within liverworts, compound air pores develop only in members of the class Marchantiopsida , in the subclass Marchantiidae (Crandall-Stotler et al., 2009). This suggests that air pore complex is a morphological innovation that occurred specifically in the

Marchantiidae lineage. This means that IIIf bHLH TFs were likely to be active in the epidermal cell of the first land plants and they were recruited during the diversification of the liverwort lineage to control the development of air pore complexes. Moreover, MpGL3 positively regulate the transcription of Mp*WIP*, a positive regulator of air pore complex development. Altogether, these observations suggest that MpGL3 was recruited to control air pore complex development by positively regulate Mp*WIP* expression.

Further research would provide useful insights into the function of MpGL3 that could explain the phenotypical differences observed in the amiR and Mp*GL3*-SRDX dominant repressor plants. It would be particularly useful to generate and analyse a complete loss-of-function mutant of *MpGL3*. CRISPR/Cas9 and homologous recombination can be used to create complete loss-of-function in *M. polymorpha* and these molecular techniques can be employed to generate Mp*gl3* mutant plants that are unable to express functional MpGL3 (Ishizaki et al., 2013b; Sugano et al., 2014). Although it is possible that a complete loss-of-function of Mp*GL3* is lethal due to the fact that this TF encoding gene is expressed in the meristematic notches and it is involved in the regulation of thallus growth.

Thallus development in the constitutively expressed Mp*GL3* mutant is identical to the WT thallus. This suggests that while IIIf bHLH TF is required for the development of dorsal characters and thallus branching and growth, this bHLH TF is not sufficient to promote these developmental processes. This could be explained by the fact that IIIf bHLH TF heterodimerizes and forms a transcriptional complex with other TFs and expression of all components of this transcriptional complex is required to promote thallus development in *M. polymorpha* (Figure 3.19). This is a plausible explanation as IIIf bHLH TFs form heterodimers with WD-40 and MYB TFs that control epidermal cell differentiation in angiosperms (Zhang et al., 2003). Moreover, MpGL3 contains domains that are required for heterodimerization of IIIf bHLH TFs angiosperms. However, *M. polymorpha* does not encode homologs of the MYB TF class known to heterodimerize with IIIf bHLH TFs in *A. thaliana* (John Bowman et al. unpublished). Nevertheless, it is possible that other MYB TF classes form a transcriptional complex, together with MpGL3, that controls thallus branching, growth and the development of dorsal characters. The potential presence and activity of such transcriptional complex can be further examined by performing a forward mutagenesis screen in order to identify and characterize mutants that are defective in thallus growth and dorsal epidermal cell development. Furthermore, protein-protein interaction assays can reveal putative developmental regulators that heterodimerize with MpGL3 in order to control thallus development in *M. polymorpha*. These experimental approaches can demonstrate that the heterodimerization of IIIf bHLH TFs with WD-40 and MYB TFs constitutes an ancient transcriptional complex that regulates epidermal cell differentiation across land plants.

Auxin regulates the development of *M. polymorpha* thallus in a dosage-dependent manner (Eklund et al., 2015; Flores-Sandoval et al., 2015; Kato et al., 2015; Kaul et al., 1962; Maravolo and Voth, 1966). This developmental regulation includes dorsal epidermal cell differentiation and thallus growth and branching, which I have shown to be controlled by MpGL3. The analysis of Mp*GL3* transcript abundance on plants grown on media supplemented with different exogenous auxin and auxin biosynthesis inhibitor suggests that auxin mediates Mp*GL3* transcript abundance in a dosage-dependent manner. This result implies that auxin-mediated transcriptional regulation acts upstream of IIIf bHLH TF in order to regulate the development of *M. polymorpha* thallus (Figure 3.19). Moreover, MpGL3 positively regulates the expression of the auxin signalling co-repressor Mp*TPL*, suggesting that a negative feedback loop between auxin and MpGL3 operates in the thallus of *M. polymorpha* to regulate growth, and dorsal epidermis and tissue differentiation (Figure 3.19). Interestingly, no transcriptional regulation between auxin and IIIf bHLH TFs has been reported in angiosperms, suggesting that the transcriptional pathway that operates upstream of IIIf bHLH TFs has changed since *M. polymorpha* and *A. thaliana* shared a common ancestor, i.e. the first land plants.

The degree of conservation of the transcriptional pathway that acts downstream of IIIf bHLH TFs in land plants is still unknown. This can be tested by transcriptome analysis of WT and IIIf bHLH TF loss-of-function mutants in both *M. polymorpha* and *A. thaliana* to determine the genes whose expression is misregulated in the loss-of-function mutants in comparison with the WT. This analysis would reveal how the IIIf bHLH dependent-transcriptional network has changed during land plant evolution and if conserved downstream regulators could account for the conserved function in the control of epidermal cell development. Differences in the transcriptional pathway that acts downstream of IIIf bHLH TFs could explain the recruitment of IIIf bHLH TF function to control unrelated developmental processes during land plant evolution.

Taken together, the results presented in this chapter suggest that the differentiation of epidermal cells of the first land plants is controlled by IIIf bHLH TFs. These bHLH TFs were recruited during the evolution of land plants to control other functions in specific lineages of extant land plants (such as air pore complexes development, branching and thallus growth in the liverwort *M. polymorpha*). This recruitment can be caused by the rewiring of IIIf bHLH TFs transcriptional pathways that act upstream of IIIf bHLH TFs in different land plant lineages. The recruitment of ancient developmental regulators to control distinct developmental processes in land plants might underlie the morphological and taxonomic radiation of land plants.

Chapter 4: General Discussion

Plants radiated both morphologically and taxonomically since the colonization of the terrestrial realm by the first land plants (Bateman et al., 1998; Kenrick and Crane, 1997; Steemans et al., 2009). Evidence suggests that the evolution of plant body plan was underpinned by a set of highly conserved regulatory proteins. How can such conservation of regulatory proteins account for the generation of gene regulatory mechanisms diversity that is ultimately translated into the morphological and taxonomic land plant diversity? It is currently postulated and experimentally validated that components of gene regulatory networks are conserved at a higher taxonomic level and these components regulate homologous, and functionally and morphologically analogous, characters during the evolution of land plants. The rewiring of these conserved networks might contribute for the morphological and taxonomical diversity of land plants. One example that supports this evolutionary trend is the bHLH transcriptional pathway, comprised by components of three different bHLH subfamilies, that regulates the development of filamentous tip-growing cells with a rooting function (Breuninger et al., 2016; Bruex et al., 2012; Menand et al., 2007; Pires et al., 2013; Proust et al., 2016; Tam et al., 2015). These bHLH TFs are ancient regulators of cells with a rooting function but the transcriptional pathway has been rewired during the evolution of land plants (Pires et al., 2013; Tam et al., 2015). The rewiring of this gene regulatory network can account for the developmental plasticity and be a driver of morphological disparity of land plant rooting systems. Another example is the independent recruitment of the meristem-to-leaf developmental transition mediated by KNOX-ARP interaction in land plants. This mechanism was recruited from a leafless ancestral to regulate microphyll and megaphyll formation in lycophytes and euphylophytes, respectively (Harrison et al., 2005). However, the recruitment and rewiring of a gene regulatory network, during land plant radiation, controlling the development of functionally and morphologically analogous characters, is only one evolutionary mechanism that accounts for the morphological and taxonomic diversity of land plants.

In this thesis I demonstrate that other evolutionary mechanisms also contribute for land plant evolution and diversity. In Chapter 2, I presented a phylogenetic analysis of TF families that demonstrates that the origin of TF families occurred in a stepwise manner before the colonization of the land by plants. I also demonstrate that the radiation of plants on land was not accompanied by a significant origin of new TF families. Moreover, this evolutionary trend also occurs within the evolution of bHLH TF family, where the majority of subfamilies originated before the radiation of extant lineages of land plants. This is consistent with the hypothesis that the origin of bHLH subfamilies predated the morphological and taxonomic diversification of plants on land and suggests that ancient bHLH subfamilies underpinned developmental mechanisms critical for the successful colonization and adaptation of plants on land.

I also showed that there were some origins and losses of bHLH subfamilies during land plant evolution. These evolutionary events, such as lineage-specific losses that occur at a high taxonomic level, can change or abolish the development of specific morphological characters. This can account for the morphological and taxonomic disparity between different lineages of land plants. Moreover, this evolutionary trend in plant bHLH TFs is shared with metazoan bHLH TFs evolution, where losses of metazoans subfamilies also occurred in specific lineages of metazoans (Gyoja and Satoh, 2013). This suggests that common evolutionary trends occurred in distantly related multicellular eukaryotic lineages. Indeed, the origin of the majority of metazoan bHLH subfamilies occurred in a stepwise manner and predated the morphological radiation of bilaterians. I have shown that this is similar to the evolution of plant bHLH subfamilies where there was a stepwise origin of the majority of plant bHLH subfamilies that predated the radiation of extant lineages of land plants. This similar trend in the evolution of bHLH TFs in distant branches of the tree of life suggests that morphological disparity, diversity and complexity within land plants and bilaterians were a product of TFs reuse and diversification within ancient subfamilies.

The epidermis constitutes the interface between the plant and the environment. Thus, the evolution of plant epidermis was essential for the successful colonization and adaptation of plants on the terrestrial realm. Different specialized epidermal cells mediate plant-environment interactions in the liverwort *M. polymorpha* and in the angiosperm *A. thaliana*. *M. polymorpha* develops rhizoids and air pores while *A. thaliana* develops epidermal cells that are not found in liverworts, such as root hairs, stomata and trichomes. Air pore complex is a morphological innovation that occurred specifically in the Marchantiidae lineage (Crandall-Stotler and Stotler, 2008). Therefore, the control of its development is likely to have been recruited from a pre-existing gene regulatory mechanism. Indeed, it appears that ancient regulatory mechanisms were recruited independently in different lineages of land plants to control the development of different and novel morphological characters. In chapter 3, I showed that IIIf bHLH TFs are conserved regulators of epidermal cell development and were likely to be active in the epidermis of the first land plants to promote the development of cells that constitute the interface between the plant and the environment. Furthermore, not only they control morphologically distinct epidermal cell types in different lineages of land plants, but they also regulate processes seemingly unrelated to epidermal cell development, such as thallus branching in *M. polymorpha*. The ability of IIIf bHLH TFs to control distinct processes in different lineages of land plants could be explained by a remodelling of the transcriptional network that acts upstream and downstream of IIIf bHLH TFs. Consistent with this hypothesis is the fact that auxin seems to mediate the expression of IIIf bHLH TFs in *M. polymorpha*, and such negative feedback loop between auxin and IIIf bHLH TFs has not been reported in *A. thaliana*. This means that IIIf bHLH TFs act in different transcriptional pathways in distantly related lineages of land plants, which can account for the ability of IIIf bHLH TFs to be recruited and co-opted to regulate unrelated development processes during land plant evolution. In sum, this illustrates that a functionally conserved genetic regulator can account for the morphological and taxonomic diversification of land plants through its recruitment to control unrelated lineage-specific developmental processes.

In conclusion, this thesis contributes to the understanding of evolutionary trends that underpinned land plant evolution and the radiation of extant lineages of land plants at both morphological and taxonomic level. I showed that the origin of ancient regulators of plant development predated the radiation of land plants. I identified one of these regulators has a conserved regulator of epidermal cell development that acquired the ability to regulate other developmental processes in specific lineages of land plants. Further analyses that comprise more diverse taxa will provide a solid framework for phylogenetic and developmental studies that will lead to the consolidation of proposed evolutionary theories and shed light on evolutionary mechanisms that underpin the radiation of land plants.

Appendices

Appendix 1 – List of primers used

Amplification of Mp*GL3* CDS

MpGL3_F	ATGGCTGGAGCGGGCAAAAG
MpGL3_R	TTATCTGGAGCCCATCCCAG

Amplification of Mp*GL3*-SRDX

MpGL3_F	ATGGCTGGAGCGGGCAAAAG
MpGL3-SRDX	CTACGCAAAGCCCAGGCGCAGTTCCAGATCCAGATCCAGTCTGGAGCCCATCCCAGCAGC

Amplification of *pro*Mp*GL3*

pMpGL3-PmeI-F	GTTTAAACTCGAATCCATGAAACGTAC
pMpGL3-SmaI-PmeI-R	GTTTAAACCCGGGCGTGTCGTAGCACACATTTCC

RT-qPCR and RT-sqPCR

MpEF1α__RT_F	AGGTTGTCACCATGGGAAAGGAGA
MpEF1α__RT_R	TCACACGCTTGTCAATACCTCCCA
MpACT_F	GAGCGCGGTTACTCTTTCAC
MpACT_R	GACCGTCAGGAAGCTCGTAG
MpGL3_RT_F	CAAGCTTCAACGAGTGATGG
MpGL3_RT_R	CGGTTTCTTCCAAAGCTGTC
SRDX_R	CGCAAAGCCCAGGCGCAGTT
MpRSL1_RT_F	AGATGAGTCTGGGGCAACC
MpRSL1_RT_R	GGATGAGCGCTTTAGAGTG
MpRKD__RT_F	TCGAGCTTTGGCAATGCATA
MpRKD__RT_R	TTGCGGATTCCTCGTGACAC
MpWIP_RT_F	CGCTACAACAACATGCAGATGC
MpWIP_RT_R	GGGTGGTCGATGTTGTTCC
MpYUC2_F	AGATTTTTGGCACGTCAACC
MpYUC2_R	CCGACCTTGTCTTTCAGCTC
MpIAA_F	TCAGCGAGAAGAAGCCATTT
MpIAA_R	CCTGCTGCTGTTGTTGTTGT
MpARF3_F	ATGAATATGCAAGGCCAAGG
MpARF3_R	AACCCACCTGGGAGTTTACC
MpTPL_F	TATTGGGTACCGCGAGATTC
MpTPL_R	GCAGTAGCTCCGCTAACACC

Appendix 2 –*amiR*-Mp*GL3-1*^{Mpmir160} and *amiR*-Mp*GL3-2*^{Mpmir160} sequences

MpmiR160

TGCCTGGCTCCCTGTATGCCA
ACGTACCGGGGGACTTACGGT

Miss match in bp 7/13/18

>MpmiR160

GCACCTCCTCTCTCCGACTGCAGCCCGTTTCGAGATCCGAGGACTTGCTCGACGCGA

CTAATTGGGGAGGCCAGACTGCACTTGCCTGGCTCCCTGTATGCCAACTGAGGAGCT

CCTCAGAGACCTTGACAGGCTCCGTAGCTGGCATTCAGGGGGCCATGCAGGAGGAAG

TCGCTACCTCCCGCAAGGTGCGACTAGCTTTCTGTCTTGGGTGCACACCTCACTGAT

GTTTGATAGATTTACTTA

amiR-MpGL3-1Mpmir160 (5-3) TAAGTTTCGTATAAATCCCGC
amiR-MpGL3-1Mpmir160* (3-5) ATTtAAAGtATATTcAGGGCG

amiR-MpGL3-1Mpmir160* (5-3) GCGGGACTTATATGAAATTTA

> *amiR*-Mp*GL3-1*^{Mpmir160}

GCACCTCCTCTCTCCGACTGCAGCCCGTTTCGAGATCCGAGGACTTGCTCGACGCGACTAATTGGGGA
GGCCAGACTGCACTTAAGTTTCGTATAAATCCCGCACTGAGGAGCTCCTCAGAGACCTTGACAGGCTC
CGTAGCGCGGGACTTATATGAAATTTAGGAGGAAGTCGCTACCTCCCGCAAGGTGCGACTAGCTTTCT
GTCTTGGGTGCACACCTCACTGATGTTTGATAGATTTACTTA

amiR-Mp*GL3-2*^{Mpmir160} (5-3) TAGATCGTTCAAAGTTCGCAA
amiR-Mp*GL3-2*^{Mpmir160} * (3-5) ATCcAGCAgGTTTCgAGCGTT

amiR-Mp*GL3-2*^{Mpmir160} * (5-3) TTGCGAGCTTTGGACGACCTA

> *amiR*-Mp*GL3-2*^{Mpmir160}

GCACCTCCTCTCTCCGACTGCAGCCCGTTTCGAGATCCGAGGACTTGCTCGACGCGACTAATTGGGGA
GGCCAGACTGCACTTAGATCGTTCAAAGTTCGCAAACTGAGGAGCTCCTCAGAGACCTTGACAGGCTC
CGTAGCTTGCGAGCTTTGGACGACCTAGGAGGAAGTCGCTACCTCCCGCAAGGTGCGACTAGCTTTCT
GTCTTGGGTGCACACCTCACTGATGTTTGATAGATTTACTTA

References

Abe, M. (2003). Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development *130*, 635–643.

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G., Fensome, R.A., Fredericq, S., et al. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. *52*, 399–451.

Aguilar-Martínez, J.A., Poza-Carrión, C., and Cubas, P. (2007). Arabidopsis BRANCHED1 acts as an integrator of branching signals within axillary buds. Plant Cell *19*, 458–472.

Althoff, F., Kopischke, S., Zobell, O., Ide, K., Ishizaki, K., Kohchi, T., and Zachgo, S. (2014). Comparison of the MpEF1α and CaMV35 promoters for application in Marchantia polymorpha overexpression studies. Transgenic Res. *23*, 235–244.

Aoyama, T., Hiwatashi, Y., Shigyo, M., Kofuji, R., Kubo, M., Ito, M., and Hasebe, M. (2012). AP2-type transcription factors determine stem cell identity in the moss Physcomitrella patens. Development *139*, 3120–3129.

Apostolakos, P., Galatis, B., and Mitrakos, K. (1982). Studies on the Development of the Air Pores and Air Chambers of Marchantia paleacea 1 . Light Microscopy. Ann. Bot. *49*, 377–396.

Archibald, J.M. (2009). The Puzzle of Plastid Evolution. Curr. Biol. *19*, 81–88.

Ashton, N.W., and Cove, D.J. (1977). The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, Physcomitrella patens. Mol. Gen. Genet. *154*, 87–95.

Bai, M.-Y., Fan, M., Oh, E., and Wang, Z.-Y. (2012). A triple helix-loop-helix/basic helix-loop-helix cascade controls cell elongation downstream of multiple hormonal and environmental signaling pathways in Arabidopsis. Plant Cell *24*, 4917–4929.

Bailey, T.L., and Elkan, C. (1994). Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers. Proc. Second Int. Conf. Intell. Syst. Mol. Biol. 28–36.

Banks, J.A. (1999). GAMETOPHYTE DEVELOPMENT IN FERNS. Annu. Rev. Plant Physiol. Plant Mol. Biol. *50*, 163–186.

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

Bao, N., Lye, K.W., and Barton, M.K. (2004). MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. Dev. Cell *7*, 653–662.

Bateman, R.M., Crane, P.R., DiMichele, W.A., Kenrick, P.R., Rowe, N.P., Speck, T., and Stein, W.E. (1998). Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation. Annu. Rev. Ecol. Syst. *29*, 263–292.

Baudry, A., Caboche, M., and Lepiniec, L. (2006). TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in Arabidopsis thaliana. Plant J. *46*, 768–779.

Becker, B., and Marin, B. (2009). Streptophyte algae and the origin of embryophytes. Ann. Bot. *103*, 999–1004.

Bell, P.R.. (1992). Apospory and Apogamy : Implications for Understanding the Plant Life

Cycle. Int. J. Plant Sci. *153*, 123–136.

Benjamins, R., and Scheres, B. (2008). Auxin: The Looping Star in Plant Development. Annu. Rev. Plant Biol. *59*, 443–465.

Bennett, T.A., Liu, M.M., Aoyama, T., Bierfreund, N.M., Braun, M., Coudert, Y., Dennis, R.J., O'Connor, D., Wang, X.Y., White, C.D., et al. (2014). Plasma membrane-targeted PIN proteins drive shoot development in a moss. Curr. Biol. *24*, 2776–2785.

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

Bergmann, D.C., and Sack, F.D. (2007). Stomatal development. Annu. Rev. Plant Biol. *58*, 163–181.

Berner, R.A., Vandenbrooks, J.M., and Ward, P.D. (2007). Oxygen and evolution. Science (80-.). *316*, 557–558.

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

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

Bhatt, A.M., Etchells, J.P., Canales, C., Lagodienko, A., and Dickinson, H. (2004). VAAMANA--a BEL1-like homeodomain protein, interacts with KNOX proteins BP and STM and regulates inflorescence stem growth in Arabidopsis. Gene *328*, 103–111.

Bhattacharya, D., Price, D.C., Chan, C.X., Qiu, H., Rose, N., Ball, S., Weber, A.P., Arias, M.C., Henrissat, B., Coutinho, P.M., et al. (2013). Genome of the red alga Porphyridium purpureum. Nat Commun *4*, 1941.

Bierfreund, N.M., Reski, R., and Decker, E.L. (2003). Use of an inducible reporter gene system for the analysis of auxin distribution in the moss Physcomitrella patens. Plant Cell Rep. *21*, 1143–1152.

Binns, A., and Maravolo, N. (1972). Apical dominance, polarity, and adventitious growth in Marchantia polymorpha. Am. J. Bot. *59*, 691–696.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., et al. (2010). The Chlorella variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell *22*, 2943–2955.

Bower, F.O. (1890). On antithetic as distinct from homologous alternation of generations in plants. Ann. Bot. *3*, 347–370.

Bowman, J.L. (2013). Walkabout on the long branches of plant evolution. Curr. Opin. Plant Biol. *16*, 70–77.

Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M., and Laux, T. (2008). Differential expression of WOX genes mediates apical-basal axis formation in the Arabidopsis embryo. Dev. Cell *14*, 867–876.

Breuninger, H., Thamm, A., Streubel, S., Sakayama, H., Nishiyama, T., and Dolan, L. (2016). Diversification of a bHLH transcription factor family in streptophytes led to the evolution of antagonistically acting genes controlling root hair growth. Curr. Biol. *26*, 920–934.

Brewer, P.B., Dun, E.A., Ferguson, B.J., Rameau, C., and Beveridge, C.A. (2009). Strigolactone Acts Downstream of Auxin to Regulate Bud Outgrowth in Pea and Arabidopsis. PLANT Physiol. *150*, 482–493.

Broun, P. (2005). Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. Curr. Opin. Plant Biol. *8*, 272–279.

Bruex, A., Kainkaryam, R.M., Wieckowski, Y., Kang, Y.H., Bernhardt, C., Xia, Y., Zheng, X., Wang, J.Y., Lee, M.M., Benfey, P., et al. (2012). A Gene Regulatory Network for Root Epidermis Cell Differentiation in Arabidopsis. PLoS Genet. *8*, e1002446.

Buck, M.J., and Atchley, W.R. (2003). Phylogenetic analysis of plant basic helix-loop-helix proteins. J. Mol. Evol. *56*, 742–750.

Busch, M. a, Bomblies, K., and Weigel, D. (1999). Activation of a floral homeotic gene in Arabidopsis. Science *285*, 585–587.

Caine, R., Chater, C.C., Kamisugi, Y., Cuming, A.C., Beerling, D.J., Gray, J.E., and Fleming, A.J. (2016). An ancestral stomatal patterning module revealed in the non-vascular land plant *Physcomitrella patens*. Development dev.135038.

Calvert, H.E., Pence, M.K., and Peters, G.A. (1985). Ultrastructural ontogeny of leaf cavity trichomes in Azolla implies a functional role in metabolite exchange. Protoplasma *129*, 10–27.

Carlsbecker, A., Lee, J.-Y., Roberts, C.J., Dettmer, J., Lehesranta, S., Zhou, J., Lindgren, O., Moreno-Risueno, M. a, Vatén, A., Thitamadee, S., et al. (2010). Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. Nature *465*, 316–321.

Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martinez-Garcia, J.F., Bilbao-Castro, J.R., Robertson, D.L., Martínez-García, J.F., Bilbao-Castro, J.R., and Robertson, D.L. (2010). Genome-Wide Classification and Evolutionary Analysis of the bHLH Family of Transcription Factors in Arabidopsis, Poplar, Rice, Moss, and Algae. PLANT Physiol. *153*, 1398–1412.

Castillon, A., Shen, H., and Huq, E. (2007). Phytochrome Interacting Factors: central players in phytochrome-mediated light signaling networks. Trends Plant Sci. *12*, 514–521.

Chang, Y., and 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*, 839–849.

Chater, C.C., Caine, R.S., Tomek, M., Wallace, S., Kamisugi, Y., Cuming, A.C., Lang, D., MacAlister, C.A., Casson, S., Bergmann, D.C., et al. (2016). Origin and function of stomata in the moss Physcomitrella patens. Nat. Plants *2*, 16179.

Chevalier, F., Perazza, D., Laporte, F., Le Henanff, G., Hornitschek, P., Bonneville, J.-M., Herzog, M., and Vachon, G. (2008). GeBP and GeBP-Like Proteins Are Noncanonical Leucine-Zipper Transcription Factors That Regulate Cytokinin Response in Arabidopsis. Plant Physiol. *146*, 1142–1154.

Clarke, J.T., Warnock, R.C.M., and Donoghue, P.C.J. (2011). Establishing a time-scale for plant evolution. New Phytol. *192*, 266–301.

Collén, J., Porcel, B., Carré, W., Ball, S.G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K., et al. (2013). Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. Proc. Natl. Acad. Sci. U. S. A. *110*, 5247–5252.

Coudert, Y., Palubicki, W., Ljung, K., Novak, O., Leyser, O., and Jill Harrison, C. (2015). Three ancient hormonal cues co-ordinate shoot branching in a moss. Elife *2015*, 1–26.

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

Crandall-Stotler, B., and Stotler, R.E. (2008). Morphology and classification of the Marchantiophyta. In Bryophyte Biology, A. Shaw, and B. Goffinet, eds. (New York: Cambridge University Press), pp. 1–54.

Crandall-Stotler, B., Stotler, R.E., and Long, D.G. (2009). Phylogeny and classification of the Marchantiophyta. Edinburgh J. Bot. *66*, 155.

Crawford, S., Shinohara, N., Sieberer, T., Williamson, L., George, G., Hepworth, J., Mueller, D., Domagalska, M.A., and Leyser, O. (2010). Strigolactones enhance competition between shoot branches by dampening auxin transport. Development *137*, 2905–2913.

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

Davidonis, G., and Munroe, M.H. (1972). Apical Dominance in Marchantia : Correlative Inhibition of Neighbor Lobe Growth. Bot. Gaz. *133*, 177–184.

Degnan, B.M., Vervoort, M., Larroux, C., and Richards, G.S. (2009). Early evolution of metazoan transcription factors. Curr. Opin. Genet. Dev. *19*, 591–599.

Derbyshire, P., Drea, S., Shaw, P.J., Doonan, J.H., and Dolan, L. (2008). Proximal-distal patterns of transcription factor gene expression during Arabidopsis root development. J. Exp.

Bot. *59*, 235–245.

Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A.Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., et al. (2006). Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. Proc. Natl. Acad. Sci. U. S. A. *103*, 11647–11652.

DiMichele, W.A., and Phillips, T.L. (1994). Paleobotanical and paleoecological constraints on models of peat formation in the Late Carboniferous of Euramerica. Palaeogeogr. Palaeoclimatol. Palaeoecol. *106*, 39–90.

Dolan, L., and Roberts, K. (1995). The Development of Cell Pattern in the Root Epidermis. Philos. Trans. R. Soc. B Biol. Sci. *350*, 95–99.

Dolan, L., Duckett, C.M., Grierson, C., Linstead, P., Schneider, K., Lawson, E., Dean, C., Poethig, S., and Roberts, K. (1994). Clonal relationships and cell patterning in the root epidermis of Arabidopsis. Development *2474*, 2465–2474.

Duckett, J.G., Ligrone, R., Renzaglia, K.S., and Pressel, S. (2014). Pegged and smooth rhizoids in complex thalloid liverworts (Marchantiopside): structure, function and evolution. Bot. J. Linn. Soc. *174*, 68–92.

Durand, E.J. (1908). The Development of the Sexual Organs and Sporogonium of Marchantia polymorpha. Bull. Torrey Bot. Club *35*, 321–335.

Edwards, D., and Feehan, J. (1980). Records of Cooksonia-type sporangia from late Wenlock strata in Ireland. Nature *287*, 41–42.

Eklund, D.M., Ishizaki, K., Flores-Sandoval, E., Kikuchi, S., Takebayashi, Y., Tsukamoto, S., Hirakawa, Y., Nonomura, M., Kato, H., Kouno, M., et al. (2015). Auxin Produced by the

Indole-3-Pyruvic Acid Pathway Regulates Development and Gemmae Dormancy in the Liverwort *Marchantia polymorpha*. Plant Cell *27*, tpc.15.00065.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial Patterning of Arabidopsis Shoots by Class III HD-ZIP and KANADI Genes. Curr. Biol. *13*, 1768–1774.

Erwin, D.H., Laflamme, M., Tweedt, S.M., Sperling, E.A.E. a, Pisani, D., and Peterson, K.J. (2011). The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science *334*, 1091–1097.

Etchells, J.P., Provost, C.M., Mishra, L., and Turner, S.R. (2013). WOX4 and WOX14 act downstream of the PXY receptor kinase to regulate plant vascular proliferation independently of any role in vascular organisation. Development *140*, 2224–2234.

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

Ferguson, B.J., and Beveridge, C.A. (2009). Roles for Auxin, Cytokinin, and Strigolactone in Regulating Shoot Branching. PLANT Physiol. *149*, 1929–1944.

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

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44*, D279–D285.

Flores-Sandoval, E., Eklund, D.M., and Bowman, J.L. (2015). A simple auxin transcriptional response system regulates multiple morphogenetic processes in the liverwort Marchantia polymorpha. PLOS Genet. *11*, e1005207.

Flores-Sandoval, E., Dierschke, T., Fisher, T.J., and Bowman, J.L. (2016). Efficient and inducible use of artificial microRNAs in Marchantia polymorpha. Plant Cell Physiol. *57*, 281–290.

Floyd, S.K., and Bowman, J.L. (2004). Ancient microRNA target sequences in plants. Nature *428*, 485–486.

Friedman, W.E. (2009). The meaning of Darwin's 'abominable mystery'. Am. J. Bot. *96*, 5–21.

Furumizu, C., Alvarez, J.P., Sakakibara, K., and Bowman, J.L. (2015). Antagonistic Roles for KNOX1 and KNOX2 Genes in Patterning the Land Plant Body Plan Following an Ancient Gene Duplication. PLoS Genet. *11*, 1–24.

Gallois, J., Nora, F.R., Mizukami, Y., and Sablowski, R. (2004). WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem. Genes Dev. 375–380.

Gensel, P.G. (2008). The earliest land plants. Annu. Rev. Ecol. Evol. Syst. *39*, 459–477.

Gonzalez, A., Zhao, M., Leavitt, J.M., and Lloyd, A.M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. Plant J. *53*, 814–827.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., et al. (2011). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. *40*, D1178–D1186.

Graham, L.E., Cook, M.E., and Busse, J.S. (2000). The origin of plants: body plan changes contributing to a major evolutionary radiation. Proc. Natl. Acad. Sci. U. S. A. *97*, 4535–4540.

Grigoriev, I. V, Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R. a, et al. (2012). The Genome Portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res. *40*, D26–D32.

Guindon, S.S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Guo, C.-Q., Edwards, D., Wu, P.-C., Duckett, J.G., Hueber, F.M., and Li, C.-S. (2012). Riccardiothallus devonicus gen. et sp. nov., the earliest simple thalloid liverwort from the Lower Devonian of Yunnan, China. Rev. Palaeobot. Palynol. *176–177*, 35–40.

Gyoja, F. (2014). A genome-wide survey of bHLH transcription factors in the Placozoan Trichoplax adhaerens reveals the ancient repertoire of this gene family in metazoan. Gene *542*, 29–37.

Gyoja, F., and Satoh, N. (2013). Evolutionary Aspects of Variability in bHLH Orthologous Families: Insights from the Pearl Oyster, Pinctada fucata. Zoolog. Sci. *30,* 868–876.

Haecker, A., Gross-Hardt, R., Geiges, B., Sarkar, A., Breuninger, H., Herrmann, M., and Laux, T. (2004). Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in Arabidopsis thaliana. Development *131*, 657–668.

Haig, D. (2008). Homologous versus antithetic alternation of generations and the origin of sporophytes. Bot. Rev. *74*, 395–418.

Halbsguth, W., and Kohlenbach, H.-W. (1953). Einige versuche über die wirkung von heteroauxin auf die sym- metrieentwicklung der brutkörperkeimlinge von Marchanta polymorpha L. Planta *42*, 349–366.

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

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

Harrison, C.J., Roeder, A.H.K., Meyerowitz, E.M., and Langdale, J.A. (2009). Local cues and asymmetric cell divisions underpin body plan transitions in the moss Physcomitrella patens. Curr. Biol. *19*, 461–471.

Hay, A., and Tsiantis, M. (2006). The genetic basis for differences in leaf form between Arabidopsis thaliana and its wild relative Cardamine hirsuta. Nat. Genet. *38*, 942–947.

Hay, A., and Tsiantis, M. (2009). A KNOX family TALE. Curr. Opin. Plant Biol. *12*, 593–598.

Hay, A., and Tsiantis, M. (2010). KNOX genes: versatile regulators of plant development and diversity. Development *137*, 3153–3165.

He, C., Chen, X., Huang, H., and Xu, L. (2012). Reprogramming of H3K27me3 Is Critical for Acquisition of Pluripotency from Cultured Arabidopsis Tissues. PLoS Genet. *8*, 1–13.

Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L., and Hedges, S.B. (2001). Molecular evidence for the early colonization of land by fungi and plants. Science *293*, 1129–1133.

Hedges, S.B., Blair, J.E., Venturi, M.L., and Shoe, J.L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol. Biol. *4*, 2.

Hernick, L.V., Landing, E., and Bartowski, K.E. (2008). Earth’s oldest liverworts—Metzgeriothallus sharonae sp. nov. from the Middle Devonian (Givetian) of eastern New York, USA. Rev. Palaeobot. Palynol. *148*, 154–162.

Hetherington, A.M. (2001). Guard cell signaling. Cell *107*, 711–714.

Hetherington, A.M., and Woodward, F.I. (2003). The role of stomata in sensing and driving environmental change. Nature *424*, 901–908.

Hetherington, A.J., Berry, C.M., and Dolan, L. (2016). Networks of highly branched stigmarian rootlets developed on the first giant trees. Proc. Natl. Acad. Sci. 201514427.

Hiratsu, K., Matsui, K., Koyama, T., and Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. Plant J. *34*, 733–739.

Hofmeister, W.F.B. (1862). On the germination, development, and fructification of the higher Cryptogamia: And on the fructification of the Coniferae (London: Ray Society).

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., et al. (2014). Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. Nat. Commun. *5*, 3978.

Horst, N.A., Katz, A., Pereman, I., Decker, E.L., Ohad, N., and Reski, R. (2016). A single homeobox gene triggers phase transition, embryogenesis and asexual reproduction. Nat. Plants *2*, 15209.

ten Hove, C.A., Lu, K.-J., and Weijers, D. (2015). Building a plant: cell fate specification in the early Arabidopsis embryo. Development *142*, 420–430.

Hueber, F.M. (1992). Thoughts on the Early Lycopsids and Zosterophylls Author. Ann. Missouri Bot. Gard. *79*, 474–499.

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

Inoue, K., Nishihama, R., Kataoka, H., Hosaka, M., Manabe, R., Nomoto, M., Tada, Y., Ishizaki, K., and Kohchi, T. (2016). Phytochrome signaling is mediated by PHYTOCHROME INTERACTING FACTOR in the liverwort Marchantia polymorpha. Plant Cell *28*, tpc.01063.2015.

Dello Ioio, R., Nakamura, K., Moubayindin, L., Perilli, S., Taniguchi, M., T. Morita, M., Aoyama, T., Costantino, P., and Sabatini, S. (2008). A genetic framework for genetic control of cell division and differentiation in the root meristem. Science (80-.). 1543–1547.

Ishizaki, K., Chiyoda, S., Yamato, K.T., and Kohchi, T. (2008). Agrobacterium-mediated transformation of the haploid liverwort Marchantia polymorpha L., an emerging model for plant biology. Plant Cell Physiol. *49*, 1084–1091.

Ishizaki, K., Nonomura, M., Kato, H., Yamato, K.T., and Kohchi, T. (2012). Visualization of auxin-mediated transcriptional activation using a common auxin-responsive reporter system in the liverwort Marchantia polymorpha. J. Plant Res. *125*, 643–651.

Ishizaki, K., Mizutani, M., Shimamura, M., Masuda, A., Nishihama, R., and Kohchi, T. (2013a). Essential role of the E3 ubiquitin ligase nopperabo1 in Schizogenous intercellular space formation in the liverwort Marchantia polymorpha. Plant Cell *25*, 4075–4084.

Ishizaki, K., Johzuka-Hisatomi, Y., Ishida, S., Iida, S., and Kohchi, T. (2013b). Homologous recombination-mediated gene targeting in the liverwort Marchantia polymorpha L. Sci. Rep.

3, 1532.

Ishizaki, K., Nishihama, R., Ueda, M., Inoue, K., Ishida, S., Nishimura, Y., Shikanai, T., and Kohchi, T. (2015). Development of gateway binary vector series with four different selection markers for the liverwort marchantia polymorpha. PLoS One *10*, 1–13.

Jakoby, M.J., Falkenhan, D., Mader, M.T., Brininstool, G., Wischnitzki, E., Platz, N., Hudson, A., Hülskamp, M., Larkin, J., and Schnittger, A. (2008). Transcriptional profiling of mature Arabidopsis trichomes reveals that NOECK encodes the MIXTA-like transcriptional regulator MYB106. Plant Physiol. *148*, 1583–1602.

Jang, G., and Dolan, L. (2011). Auxin promotes the transition from chloronema to caulonema in moss protonema by positively regulating PpRSL1 and PpRSL2 in Physcomitrella patens. New Phytol. *192*, 319–327.

Jang, G., Yi, K., Pires, N.D., Menand, B.B., and Dolan, L. (2011). RSL genes are sufficient for rhizoid system development in early diverging land plants. Development *138*, 2273–2281.

Jin, J., Zhang, H., Kong, L., Gao, G., and Luo, J. (2014). PlantTFDB 3.0: A portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. *42*, 1182–1187.

Jones, V.A.S., and Dolan, L. (2012). The evolution of root hairs and rhizoids. Ann. Bot. *110*, 205–212.

Jones, V.A.S., and Dolan, L. (2017). Mp *WIP* regulates air pore complex development in the liverwort *Marchantia polymorpha*. Development *144*, 1472–1476.

Karas, B., Amyot, L., Johansen, C., Sato, S., Tabata, S., Kawaguchi, M., and Szczyglowski, K. (2009). Conservation of Lotus and Arabidopsis Basic Helix-Loop-Helix Proteins Reveals

New Players in Root Hair Development. PLANT Physiol. *151*, 1175–1185.

Karol, K.G., McCourt, R.M., Cimino, M.T., and Delwiche, C.F. (2001). The closest living relatives of land plants. Science (80-.). *294*, 2351–2353.

Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J.L., Nishihama, R., and Kohchi, T. (2015). Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in Marchantia polymorpha. PLoS Genet. *11*, 1–26.

Katoh, K., and Frith, M.C. (2012). Adding unaligned sequences into an existing alignment using MAFFT and LAST. Bioinformatics *28*, 3144–3146.

Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. *30*, 772–780.

Kaul, K.N.N., Mitra, G.C., Tripathi, B.K., Trioathi, B.K., and Tripathi, B.K. (1962). Responses of Marchantia in aseptic culture to well-known auxins and antiauxins. Ann. Bot. *26*, 447.

Kenrick, P. (2011). Timescales and timetrees. New Phytol. *192*, 3–6.

Kenrick, P.R. (1994). Alternation of generations in land plants: new phylogenetic and palaeobotanical evidence. Biol. Rev. *69*, 293–330.

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

Kenrick, P., Wellman, C.H., Schneider, H., and Edgecombe, G.D. (2012). A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. Philos. Trans. R. Soc. B Biol. Sci. *367*, 519–536.

Kerp, H., Trewin, N.H., and Hass, H. (2004). New gametophytes from the Early Devonian Rhynie chert. Trans. R. Soc. Edinb. Earth Sci. *94*, 411–428.

Kim, C.M., and Dolan, L. (2016). ROOT HAIR DEFECTIVE SIX-LIKE Class I Genes Promote Root Hair Development in the Grass Brachypodium distachyon. PLoS Genet. *12*, 1–18.

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

Knoll, A.H. (2011). The Multiple Origins of Complex Multicellularity. Annu. Rev. Earth Planet. Sci. *39*, 217–239.

Kofuji, R., and Hasebe, M. (2014). Eight types of stem cells in the life cycle of the moss Physcomitrella patens. Curr. Opin. Plant Biol. *17*, 13–21.

Kubota, A., Ishizaki, K., Hosaka, M., and Kohchi, T. (2013). Efficient Agrobacterium-Mediated Transformation of the Liverwort Marchantia polymorpha Using Regenerating Thalli. Biosci. Biotechnol. Biochem. *77*, 167–172.

Kurata, T., Ishida, T., Kawabata-Awai, C., Noguchi, M., Hattori, S., Sano, R., Nagasaka, R., Tominaga, R., Koshino-Kimura, Y., Kato, T., et al. (2005). Cell-to-cell movement of the CAPRICE protein in Arabidopsis root epidermal cell differentiation. Development *132*, 5387–5398.

Larroux, C., Luke, G.N., Koopman, P., Rokhsar, D.S., Shimeld, S.M., and Degnan, B.M. (2008). Genesis and expansion of metazoan transcription factor gene classes. Mol. Biol. Evol. *25*, 980–996.

Lashbrooke, J.G., Adato, A., Lotan, O., Alkan, N., Tsimbalist, T., Rechav, K., Fernandez Moreno, J.-P., Widemann, E., Grausem, B., Pinot, F., et al. (2015). The Tomato MIXTA-like Transcription Factor Coordinates Fruit Epidermis Conical Cell Development and Cuticular Lipid Biosynthesis and Assembly. Plant Physiol. *169*, 2553–2571.

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

Lee, M.M., and Schiefelbein, J. (2001). Developmentally distinct MYB genes encode functionally equivalent proteins in Arabidopsis. Development *128*, 1539–1546.

Lee, J.-H., Lin, H., Joo, S., and Goodenough, U. (2008). Early sexual origins of homeoprotein heterodimerization and evolution of the plant KNOX/BELL family. Cell *133*, 829–840.

Lenton, T.M., Crouch, M., Johnson, M., Pires, N., and Dolan, L. (2012). First plants cooled the Ordovician. Nat. Geosci. *5*, 86–89.

Lenton, T.M., Dahl, T.W., Daines, S.J., Mills, B.J.W., Ozaki, K., Saltzman, M.R., and Porada, P. (2016). Earliest land plants created modern levels of atmospheric oxygen. Proc. Natl. Acad. Sci. *113*, 201604787.

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

Li, E., Bhargava, A., Qiang, W., Friedmann, M.C., Forneris, N., Savidge, R.A., Johnson, L.A., Mansfield, S.D., Ellis, B.E., and Douglas, C.J. (2012). The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in Arabidopsis and is functionally conserved in Populus. New Phytol. *194*, 102–115.

Long, T.A., Tsukagoshi, H., Busch, W., Lahner, B., Salt, D.E., and Benfey, P.N. (2010). The bHLH Transcription Factor POPEYE Regulates Response to Iron Deficiency in Arabidopsis Roots[W][OA]. Plant Cell *22*, 2219–2236.

MacAlister, C.A., and Bergmann, D.C. (2011). Sequence and function of basic helix-loop-helix proteins required for stomatal development in Arabidopsis are deeply conserved in land plants. Evol. Dev. *13*, 182–192.

MacAlister, C. a, Ohashi-Ito, K., and Bergmann, D.C. (2007). Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. Nature *445*, 537–540.

Mähönen, A.P., Tusscher, K. ten, Siligato, R., Smetana, O., Díaz-Triviño, S., Salojärvi, J., Wachsman, G., Prasad, K., Heidstra, R., and Scheres, B. (2014). PLETHORA gradient formation mechanism separates auxin responses. Nature *515*, 125–129.

Maravolo, N.N.C. (1976). Polarity and localization of auxin movement in the hepatic, Marchantia polymorpha. Am. J. Bot. *63*, 526–531.

Maravolo, N.C., and Voth, P.D. (1966). Morphogenic effects of three growth substances on Marchantia gemmalings. Bot. Gaz. *127*, 79–86.

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

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S.-Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., et al. (2004). Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. Nature *428*, 653–657.

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

de Mendoza, A., Sebé-Pedrós, A., Šestak, M.S., Matejcic, M., Torruella, G., Domazet-Loso, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. U. S. A. *110*, 1–9.

Merced, A., and Renzaglia, K.S. (2016). Patterning of stomata in the moss *Funaria* : a simple way to space guard cells. Ann. Bot. *117*, 985–994.

Mosquna, A., Katz, A., Decker, E.L., Rensing, S.A., Reski, R., and Ohad, N. (2009). Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. Development *136*, 2433–2444.

Mukherjee, K., Brocchieri, L., Burglin, T.R., and Bürglin, T.R. (2009). A comprehensive classification and evolutionary analysis of plant homeobox genes. Mol. Biol. Evol. *26*, 2775–2794.

Nakazato, T., Kadota, A., and Wada, M. (1999). Photoinduction of spore germination in Marchantia polymorpha L . is mediated by photosynthesis. Plant {&} Cell Physiol. *40*, 1014–1020.

Nam, J., and Nei, M. (2005). Evolutionary change of the numbers of homeobox genes in bilateral animals. Mol. Biol. Evol. *22*, 2386–2394.

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

Niklas, K.J., and Kutschera, U. (2009). The evolution of the land plant life cycle. New Phytol. *185*, 27–41.

Niklas, K., Tiffney, B., and Knoll, A. (1983). Patterns in vascular land plant diversification. Nature *303*, 614–616.

Nishihama, R., Ishizaki, K., Hosaka, M., Matsuda, Y., Kubota, A., and Kohchi, T. (2015). Phytochrome-mediated regulation of cell division and growth during regeneration and sporeling development in the liverwort Marchantia polymorpha. J. Plant Res. *128*, 407–421.

O'Hanlon, M.E. (1926). Germination of Spores and Early Stages in Development of Gametophyte of Marchantia polymorpha. Bot. Gaz. *82*, 215–222.

Okano, Y., Aono, N., Hiwatashi, Y., Murata, T., Nishiyama, T., Ishikawa, T., Kubo, M., and Hasebe, M. (2009). A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. Proc. Natl. Acad. Sci. U. S. A. *106*, 16321–16326.

Ongaro, V., and Leyser, O. (2008). Hormonal control of shoot branching. J. Exp. Bot. *59*, 67–74.

Ortiz-Ramírez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., Feijó, J.A., and Becker, J.D. (2016). A Physcomitrella patens comprehensive transcriptome atlas provides insights into the evolution and development of land plants. Plant Cell, Submitt. 205–220.

Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., et al. (2007). The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. U. S. A. *104*, 7705–7710.

Parfrey, L.W., Lahr, D.J.G., Knoll, A.H., and Katz, L. a (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. U. S. A. *108*, 13624–13629.

Payne, W.W. (1979). Stomatal patterns in embryophytes: their evolution, ontogeny and interpretation. Taxon *28*, 117–132.

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

Pearson, H.M. (1969). Rhizoids and root hairs of ferns. Am. Fern J. *59*, 107–122.

Pereman, I., Mosquna, A., Katz, A., Wiedemann, G., Lang, D., Decker, E.L., Tamada, Y., Ishikawa, T., Nishiyama, T., Hasebe, M., et al. (2016). The Polycomb group protein CLF emerges as a specific tri-methylase of H3K27 regulating gene expression and development in Physcomitrella patens. Biochim. Biophys. Acta - Gene Regul. Mech. *1859*, 860–870.

Perez-Rodriguez, M., Jaffe, F.W., Butelli, E., Glover, B.J., and Martin, C. (2005). Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of Antirrhinum majus flowers. Development *132*, 359–370.

Peterson, K.M., Shyu, C., Burr, C. a, Horst, R.J., Kanaoka, M.M., Omae, M., Sato, Y., and Torii, K.U. (2013). Arabidopsis homeodomain-leucine zipper IV proteins promote stomatal development and ectopically induce stomata beyond the epidermis. Development *140*, 1924–1935.

Pillitteri, L.J., and Torii, K.U. (2012). Mechanism of stomatal development. Annu. Rev. Plant Biol. *63*, 12.1-12.4.

Pires, N., and Dolan, L. (2010). Origin and diversification of basic-helix-loop-helix proteins in plants. Mol. Biol. Evol. *27*, 862–874.

Pires, N.D., and Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *367*, 508–518.

Pires, N.D., Yi, K., Breuninger, H., Catarino, B., Menand, B.B., and Dolan, L. (2013). Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. Proc. Natl. Acad. Sci. *110*, 9571–9576.

Prigge, M.J., Lavy, M., Ashton, N.W., and Estelle, M. (2010). Physcomitrella patens auxin-resistant mutants affect conserved elements of an auxin-signaling pathway. Curr. Biol. *20*, 1907–1912.

Proust, H., Honkanen, S., Jones, V.A.S., Morieri, G., Prescott, H., Kelly, S., Ishizaki, K., Kohchi, T., and Dolan, L. (2016). RSL class I genes controlled the development of epidermal structures in the common ancestor of land plants. Curr. Biol. *26*, 93–99.

Qing, L., and Aoyama, T. (2012). Pathways for epidermal cell differentiation via the homeobox gene GLABRA2: update on the roles of the classic regulator. J. Integr. Plant Biol. *54*, 729–737.

Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. U. S. A. *103*, 15511–15516.

Ramsay, N. a, and Glover, B.J. (2005). MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. *10*, 63–70.

Rerie, W.G., Feldmann, K. a, and Marks, M.D. (1994). The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in Arabidopsis. Genes Dev. *8*, 1388–1399.

Rövekamp, M., Bowman, J.L., and Grossniklaus, U. (2016). Marchantia MpRKD Regulates the Gametophyte-Sporophyte Transition by Keeping Egg Cells Quiescent in the Absence of Fertilization. Curr. Biol. *26*, 1782–1789.

Rubinstein, C. V., Gerrienne, P., de la Puente, G.S., Astini, R.A., and Steemans, P. (2010). Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). New Phytol. *188*, 365–369.

Ruszala, E.M., Beerling, D.J., Franks, P.J., Chater, C., Casson, S.A., Gray, J.E., and Hetherington, A.M. (2011). Land plants acquired active stomatal control early in their evolutionary history. Curr. Biol. *21*, 1030–1035.

De Rybel, B., Möller, B., Yoshida, S., Grabowicz, I., Barbier de Reuille, P., Boeren, S., Smith, R.S., Borst, J.W., and Weijers, D. (2013). A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in Arabidopsis. Dev. Cell *24*, 426–437.

Ryu, K.H., Kang, Y.H., Park, Y., Hwang, I., Schiefelbein, J., and Lee, M.M. (2005). The WEREWOLF MYB protein directly regulates CAPRICE transcription during cell fate specification in the Arabidopsis root epidermis. Development *132*, 4765–4775.

Ryu, K.H., Zheng, X., Huang, L., and Schiefelbein, J. (2013). Computational modeling of epidermal cell fate determination systems. Curr. Opin. Plant Biol. *16*, 5–10.

Saint-Marcoux, D., Proust, H., Dolan, L., and Langdale, J. a. (2015). Identification of reference genes for real-time quantitative PCR experiments in the liverwort Marchantia polymorpha. PLoS One *10*, e0118678.

Sakakibara, K. (2016). Chapter One - Technological Innovations Give Rise to a New Era of Plant Evolutionary Developmental Biology. In Genomes and Evolution of Charophytes, Bryophytes, Lycophytes and Ferns, (Elsevier Ltd), pp. 3–35.

Sakakibara, K., Nishiyama, T., Deguchi, H., and Hasebe, M. (2008). Class 1 KNOX genes are not involved in shoot development in the moss Physcomitrella patens but do function in sporophyte development. Evol. &amp; Dev. *10*, 555–566.

Sakakibara, K., Ando, S., Yip, H.K., Tamada, Y., Hiwatashi, Y., Murata, T., Deguchi, H., Hasebe, M., and Bowman, J.L. (2013). KNOX2 Genes Regulate the Haploid-to-Diploid Morphological Transition in Land Plants. Science. *339*, 1067–1070.

Sakakibara, K., Reisewitz, P., Aoyama, T., Friedrich, T., Ando, S., Sato, Y., Tamada, Y., Nishiyama, T., Hiwatashi, Y., Kurata, T., et al. (2014). WOX13-like genes are required for reprogramming of leaf and protoplast cells into stem cells in the moss Physcomitrella patens. Development *141*, 1660–1670.

Sanderson, M. (2003). Molecular data from 27 proteins do not support a Precambrian origin of land plants. Am. J. Bot. *90*, 954–956.

Sarkar, A.K., Luijten, M., Miyashima, S., Lenhard, M., Hashimoto, T., Nakajima, K., Scheres, B., Heidstra, R., and Laux, T. (2007). Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers. Nature *446*, 811–814.

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

Schiefelbein, J. (2003). Cell-fate specification in the epidermis: a common patterning mechanism in the root and shoot. Curr. Opin. Plant Biol. *6*, 74–78.

Schulz, C., Little, D.P., Stevenson, D.W., Bauer, D., Moloney, C., and Stützel, T. (2010). An Overview of the Morphology, Anatomy, and Life Cycle of a New Model Species: The Lycophyte Selaginella apoda (L.) Spring. Int. J. Plant Sci. *171*, 693–712.

von Schwartzenberg, K., Núñez, M.F., Blaschke, H., Dobrev, P.I., Novák, O., Motyka, V., Strnad, M., Nunez, M.F., Blaschke, H., Dobrev, P.I., et al. (2007). Cytokinins in the Bryophyte Physcomitrella patens: Analyses of Activity, Distribution, and Cytokinin Oxidase/Dehydrogenase Overexpression Reveal the Role of Extracellular Cytokinins. PLANT Physiol. *145*, 786–800.

Scofield, S., and Murray, J.A.H. (2006). KNOX Gene Function in Plant Stem Cell Niches. Plant Mol. Biol. *60*, 929–946.

Shapiro, B.E., Tobin, C., Mjolsness, E., and Meyerowitz, E.M. (2015). Analysis of cell division patterns in the Arabidopsis shoot apical meristem. Proc. Natl. Acad. Sci. U. S. A. *112*, 4815–4820.

Shaw, A.J., Cox, C.J., Buck, W.R., Devos, N., Buchanan, A.M., Cave, L., Seppelt, R., Shaw, B., Larraín, J., Andrus, R., et al. (2010). Newly resolved relationships in an early land plant lineage: Bryophyta class Sphagnopsida (peat mosses). Am. J. Bot. *97*, 1511–1531.

Shimamura, M. (2016). Marchantia polymorpha : Taxonomy , Phylogeny and Morphology of a Model System Special Focus Issue – Review. Plant Cell Physiol. *57*, 230–256.

Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B.M., and Vervoort, M. (2007). Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol. Biol. *7*, 33.

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

De Smet, I., Lau, S., Mayer, U., and Jürgens, G. (2010). Embryogenesis - The humble beginnings of plant life. Plant J. *61*, 959–970.

Smith, S. a, Beaulieu, J.M., and Donoghue, M.J. (2010). An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. Proc. Natl. Acad. Sci. U. S. A. *107*, 5897–5902.

Steemans, P., Hérissé, A. Le, Melvin, J., Miller, M. a, Paris, F., Verniers, J., and Wellman, C.H. (2009). Origin and radiation of the earliest vascular land plants. Science *324*, 353.

Di Stilio, V.S., Martin, C., Schulfer, A.F., and Connelly, C.F. (2009). An ortholog of MIXTA-like2 controls epidermal cell shape in flowers of Thalictrum. New Phytol. *183*, 718–728.

Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. Nat. Genet. *43*, 1160–1163.

Sugano, S.S., Shirakawa, M., Takagi, J., Matsuda, Y., Shimada, T., Hara-Nishimura, I., and Kohchi, T. (2014). CRISPR/Cas9-mediated targeted mutagenesis in the liverwort Marchantia polymorpha L. Plant Cell Physiol. *55*, 475–481.

Sussex, I., and Kerk, N. (2001). The evolution of plant architecture. Curr. Opin. Plant Biol. *4*, 33–37.

Takada, S., Takada, N., and Yoshida, A. (2013). ATML1 promotes epidermal cell differentiation in Arabidopsis shoots. Development *140*, 1919–1923.

Tam, T.H.Y., Catarino, B., and Dolan, L. (2015). Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. Proc. Natl. Acad. Sci. *112*, E3959–E3968.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. *30*, 2725–2729.

Tanahashi, T., Sumikawa, N., Kato, M., and Hasebe, M. (2005). Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss Physcomitrella patens. Development *132*, 1727–1736.

Tarén, N. (1958). Factors regulating the initial development of gemmae in Marchantia polymorpha. Bryologist *61*, 191–204.

Taylor, T.N., Kerp, H., and Hass, H. (2005). Life history biology of early land plants: deciphering the gametophyte phase. Proc. Natl. Acad. Sci. U. S. A. *102*, 5892–5897.

Taylor, T.N., Taylor, E.L., and Krings, M. (2009). Paleobotany: The Biology and Evolution of Fossil Plants (London: Academic Press).

Thomas, B.A., and Watson, J. (1976). A rediscovered 114-foot Lepidodendron from Bolton, Lancashire. Geol. J. *11*, 15–20.

Timme, R.E., Bachvaroff, T.R., and Delwiche, C.F. (2012). Broad phylogenomic sampling and the sister lineage of land plants. PLoS One *7*, e29696.

Toledo-Ortiz, G., Huq, E., and Quail, P.H. (2003). The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell *15*, 1749–1770.

Tsuda, K., and Hake, S. (2015). Diverse functions of KNOX transcription factors in the diploid body plan of plants. Curr. Opin. Plant Biol. *27*, 91–96.

Valentine, J.W., Tiffney, B.H., and Sepkoski, J.J. (1991). Evolutionary dynamics of plants and animals: a comparative approach. Palaios *6*, 81–88.

Vasco, A., Moran, R.C., and Ambrose, B. a (2013). The evolution, morphology, and development of fern leaves. Front. Plant Sci. *4*, 345.

Vatén, A., and Bergmann, D.C. (2012). Mechanisms of stomatal development: an evolutionary view. Evodevo *3*, 11.

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., et al. (2015). A bHLH-Based Feedback Loop Restricts Vascular Cell Proliferation in Plants. Dev. Cell *35*, 432–443.

Voth, P.D.. (1941). Gemmae-Cup Production in Marchantia polymorpha and Its Response to Calcium Deficiency and Supply of Other Nutrients. Bot. Gaz. *103*, 310–325.

Wada, T. (2002). Role of a positive regulator of root hair development, CAPRICE, in Arabidopsis root epidermal cell differentiation. Development *129*, 5409–5419.

Wada, T., Tachibana, T., Shimura, Y., and Okada, K. (1997). Epidermal cell differentiation in Arabidopsis determined by a Myb homolog, CPC. Science (80-.). *277*, 1113–1116.

Walker, a R., Davison, P. a, Bolognesi-Winfield, a C., James, C.M., Srinivasan, N., Blundell, T.L., Esch, J.J., Marks, M.D., and Gray, J.C. (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. Plant Cell *11*, 1337–1350.

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

Wendrich, J.R., and Weijers, D. (2013). The Arabidopsis embryo as a miniature morphogenesis model. New Phytol. *199*, 14–25.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci U S A *111*, E4859–E4868.

Wodniok, S., Brinkmann, H., Glöckner, G., Heidel, A.J., Philippe, H., Melkonian, M., and Becker, B. (2011). Origin of land plants: do conjugating green algae hold the key? BMC Evol. Biol. *11*, 104.

Wolters, H., and Jürgens, G. (2009). Survival of the flexible: hormonal growth control and adaptation in plant development. Nat. Rev. Genet. *10*, 305–317.

Xu, B., Ohtani, M., Yamaguchi, M., Toyooka, K., Wakazaki, M., Sato, M., Kubo, M., Nakano, Y., Sano, R., Hiwatashi, Y., et al. (2014). Contribution of NAC transcription factors to plant adaptation to land. Science *343*, 1505–1508.

Yang, C., and Ye, Z. (2013). Trichomes as models for studying plant cell differentiation. Cell. Mol. Life Sci. *70*, 1937–1948.

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

Yip, H.K., Floyd, S.K., Sakakibara, K., and Bowman, J.L. (2016). Class III HD-Zip activity coordinates leaf development in Physcomitrella patens. Dev. Biol. 1–14.

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

Zhang, F., Gonzalez, A., Zhao, M., Payne, C.T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development *130*, 4859–4869.

Zhao, H., Wang, X., Zhu, D., Cui, S., Li, X., Cao, Y., and Ma, L. (2012). A single amino acid substitution in IIIf subfamily of basic helix-loop-helix transcription factor AtMYC1 leads to trichome and root hair patterning defects by abolishing its interaction with partner proteins in Arabidopsis. J. Biol. Chem. *287*, 14109–14121.

Zhiponova, M.K., Morohashi, K., Vanhoutte, I., Machemer-Noonan, K., Revalska, M., Van Montagu, M., Grotewold, E., and Russinova, E. (2014). Helix-loop-helix/basic helix-loop-helix transcription factor network represses cell elongation in Arabidopsis through an apparent incoherent feed-forward loop. Proc. Natl. Acad. Sci. U. S. A. *111*, 2–7.

Zimmer, A., Lang, D., Richardt, S., Frank, W., Reski, R., and Rensing, S. a (2007). Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Mol. Genet. Genomics *278*, 393–402.