Evolution of bHLH transcription factors that control cell differentiation in plants

Nuno Duque Pires

Thesis submitted to the University of East Anglia for the degree of

Doctor of Philosophy

September 2010

Abstract

Land plants evolved from charophyte algae over 470 million years ago. The body of land plants has changed considerably ever since, but recent genomic analyses have shown that most angiosperm developmental gene families were already present in early land plants. This raises the question of how a conserved set of developmental regulators could direct such a large increase in morphological complexity. I have addressed this problem using two approaches. First, I defined the evolutionary relationships between plant bHLH proteins, a large group of transcription factors that include important development regulators. I identified and analysed bHLH sequences from different species of land plants and algae and showed that these proteins underwent a radiation in the charophyte ancestors of land plants. bHLH subfamilies and their specific interaction domains were conserved throughout land plant evolution, suggesting that the gene regulatory networks in which they participate are very ancient. The second approach was to characterise the evolution of a developmental mechanism that controls the differentiation of rooting cells. In angiosperms, root hair development is controlled by a network of *RSL* class I and class II genes. In mosses, the development of rooting structures is also controlled by *RSL* class I genes. To understand the evolution of the RSL network, I generated *RSL* class II mutants and overexpression lines in mosses and found that they also show defects in rooting cell differentiation. I dissected the transcriptional interactions between the moss *RSL* genes and auxin and found that these are very different in mosses and angiosperms. These results point to the existence of an ancient rooting cell differentiation mechanism, whose transcriptional interactions have changed during land plant evolution. Together, these analyses support the conclusion that the evolution of novel developmental processes in land plants was partly driven by the reutilisation of very ancient developmental networks.

Table of contents

Supplementary files available on CD

Supplementary File 1 – Original names and accession numbers of the bHLH sequences used

Supplementary File 2 – Amino acid sequences of the sequences of the bHLH sequences used

Supplementary File 3 – Amino acid alignment used for the bHLH phylogenetic analyses

Supplementary File 3B – Fasta format alignment of the corresponding full nucleotide sequences

Supplementary File 4 – ML phylogenetic tree and table with bHLH subfamily members

Supplementary File 5 – Table with the non-bHLH motifs

Supplementary File 6 – DNA sequences of the bHLH EST sequences in charophytes

Supplementary File 7 – pHMM of the RSL domain, for use with the program hmmsearch/HMMER

Supplementary File 8 – Amino acid sequences of the sequences used in the RSL alignments and phylogenetic analyses

Supplementary File 9 – Alignment (fasta format) used for the phylogenetic analyses of RSL class I proteins (34 RSL class I proteins + 4 outgroups)

Supplementary File 10 – Alignment (fasta format) used for the phylogenetic analyses of RSL class II proteins (56 RSL class II proteins + 2 outgroups)

Abbreviations

aa	amino acid
aLRT	approximate likelihood-ratio test
amiRNA	artificial micro RNA
bHLH	basic-Helix-Loop-Helix
bp	base pair
CaMV	cauliflower mosaic virus
Col-0	Columbia-0
D-box	destruction box
ER	endoplasmic reticulum
EST	expressed sequence tag
GA	gibberellin
GUS	beta-glucuronidase
IAA	indole-3-acetic acid
kb	kilobase pair
miRNA	microRNA
ML	maximum likelihood
MP	maximum parsimony
n	sample size
NAA	1-naphthaleneacetic acid
NJ	neighbor joining
PCR	polymerase chain reaction
pHMM	profile hidden Markov model
qRT-PCR	quantitative RT-PCR
RSL	RHD SIX-LIKE
RT-PCR	reverse transcriptase PCR
SD	standard deviation
SE	standard error
SH	Shimodaira-Hasegawa
UTR	untranslated region
WT	wild type

Acknowledgements

Chapter 1

Introduction

The colonisation of terrestrial environments by multicellular plants over 470 million years ago was one of the most important events in the history of the planet. Land plants drastically changed atmospheric and geochemical cycles and paved the way for the evolution of terrestrial metazoan groups and complex terrestrial ecosystems. The evolution of plants on land was itself characterised by a series of radical transformations of their body plans, that included the formation of three-dimensional tissues, a *de novo* evolution of a multicellular diploid sporophyte generation, the evolution of multicellular meristems with the capacity for branching, and the development of specialised tissues and organs such as vasculature, roots, leaves, seeds and flowers.

What is the genetic and molecular basis for the complex evolution of plants on land? The recent genome sequencing efforts have revealed that basal plants such as lycophytes and mosses have homologues of most gene families that control specific developmental processes in angiosperms. These findings raise important questions regarding the origin of these developmental mechanisms. In this thesis, this problem is approached in two ways: 1) defining the evolutionary relationships of a large family of plant transcription factors, the bHLH proteins and 2) characterising the evolution of an ancient developmental mechanism that controls the development of rooting cells, using mosses as a model system; these are the subjects of Chapter 2 and 3 of this thesis, respectively. In this chapter, a review of the evolution of developmental mechanisms in land plants is followed by a description of the morphogenesis and development of mosses.

1.1. The history of land plants

Land plants (embryophytes) evolved from freshwater multicellular algae, probably related to the extant charophytes Charales or Coleochaetales (Karol 2001; Lewis and McCourt 2004; Becker and Marin 2009). Together, land plants and charophytes form a monophyletic group, the streptophytes, which is sister to chlorophyte algae (Fig. 1.1A). The most basal and simple charophytes, such as *Mesostigma*, are unicellular, but a progressive transition towards complex multicellularity is clearly traceable in the different groups of streptophytes. Charophytes evolved many features that are plesiomorphic for land plants, such as hexameric cellulose synthases, the phragmoplast, plasmodesmata, apical meristems or a placenta (Graham et al. 2000; Becker and Marin 2009). However, it was the transition of

streptophytes to terrestrial environments that was associated with the evolution of the key features that define land plants, such as a multicellular sporophyte, retention of the zygote and embryo within the female gametophyte and apical cells with three cutting faces that allow the generation of three-dimensional parenchymatous tissues (Graham et al. 2000; Niklas and Kutschera 2009).

The oldest fossil evidence for plants on land comes from spores and tissue fragments extending back through the mid-Ordovician, 470 million years ago (Wellman et al. 2003; Gensel 2008). The morphology of these microfossils suggests an affinity with extant liverworts, although the first macrofossils of liverworts only appear in the middle Devonian, around 390 million years ago (Hernick et al. 2008). The first land plant macrofossils, represented by the sporophytes of *Cooksonia* and similar forms, appear on older mid-late Silurian strata, around 425 million years ago (Fig. 1.1B; Edwards and Feehan 1980; Gensel 2008). Gametophytes have a lower preservation potential than sporophytes, but fossils from the Early Devonian Rhynie Chert indicate that the gametophytes of early land plants were complex (including stomata and conducting elements) and often resembling the gametophytes of extant liverworts (Taylor et al. 2009). The oldest evidence for the existence of vascular plants comes from trilete spores found in upper Ordovician sediments, over 443 million years ago (Steemans et al. 2009), although tracheids can only be identified by the late Silurian, over 415 million years ago (Gensel 2008). By the late Silurian (around 425 million years ago), the now extinct rhyniophytes, zosterophylls and the first lycophytes (which are defined by the presence of microphylls) had evolved (Fig. 1.1B; Kenrick and Crane 1997; Taylor et al. 2009).

The Devonian period (415-360 million years ago) was characterised by an explosion in the diversity of land plants (Kenrick and Crane 1997). This was due to the radiation of vascular plants with a dominant sporophyte generation, which were able to move from damper areas and colonise drier habitats (Bateman et al. 1998). The advantages of an increased dominance of the sporophyte on land were probably due to the potential for the production and air-dispersal of numerous spores after a singular water-dependent fertilisation event. Many important extant groups, including horsetails, ferns and the first seed plants, appeared and diversified during this period (Fig. 1.1B; Kenrick and Crane 1997). The expansion of plants on land during the Devonian period caused drastic changes in the geochemistry and atmospheric composition of the planet. The increased rate of weathering caused by large plants with complex root systems resulted in the formation of deep soils and supplied phosphorus that promoted

A

B
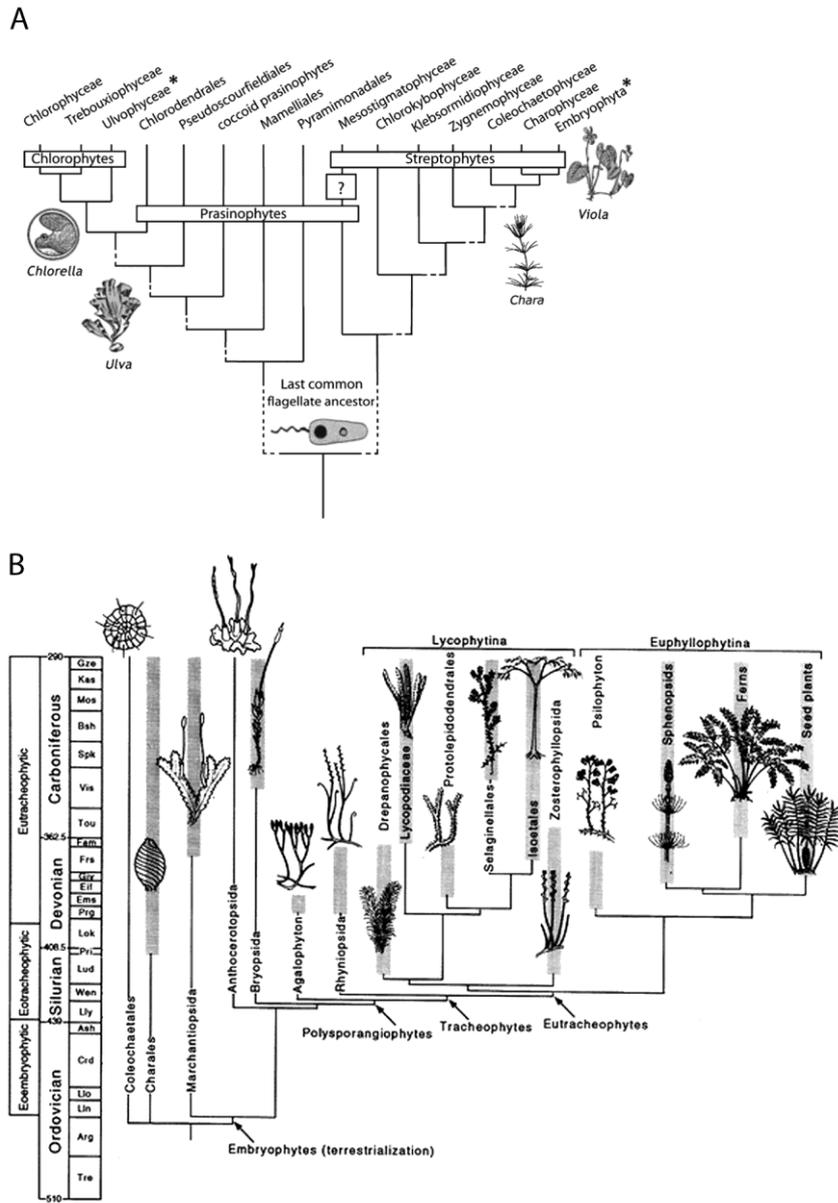
Figure 1.1 – Phylogenetic relationships among the major lineages of plants.

A Phylogenetic relationships between the major extant groups of chlorophytes, prasinophytes and streptophytes (adapted from Niklas and Kutschera 2010). **B** Phylogenetic tree showing the phylogenetic relationships and evolutionary origin of extant and extinct groups of land plants (embryophytes); the thick bars indicate the minimum stratigraphic ranges based on megafossil evidence (adapted from Kenrick and Crane 1997).

terrestrial and marine productivity. This caused a large burial of organic carbon, a decrease in atmospheric CO_2 levels and an increase in the levels of O_2 (Algeo and Scheckler 1998; Lenton 2001).

By the late Carboniferous period, around 300 million years ago, the land surface was covered by large forests of pteridosperms (seed ferns), lycophytes, tree ferns and sphenopsids (Willis and McElwain 2002). Gymnosperms appeared during this period (Taylor et al. 2009) and became dominant in the world flora between the Permian and the late Cretaceous period (between 260 and 70 million years ago) (Willis and McElwain 2002). Basal angiosperms, magnoliids, early monocots and early eudicots appeared almost simultaneously during the early Cretaceous (100-145 million years ago) (Friis et al. 2006; Taylor et al. 2009) and later radiated and became dominant in a majority of habitats from the late Cretaceous (100-65 million years ago) until the present day (Willis and McElwain 2002).

1.2. Evolution of developmental mechanisms in plants

1.2.1. Sporophyte and gametophyte

A key characteristic of land plants is that their life cycle is composed of two distinct multicellular generations: a haploid gametophyte and a diploid sporophyte. By contrast, in charophyte algae only the zygote cell is diploid. An alternation of two multicellular generations has evolved several times in different groups of algae (John 1994) but it appears to have evolved only once in the streptophytes. Historically, two major theories have addressed the origin of the alternation of generations in land plants: the homologous (or transformation) and the antithetic (or interpolation) theories (reviewed in Blackwell 2003). The homologous theory states that land plant ancestors had an alternation of isomorphic generations; this theory has currently little support, except for the discovery of early Devonian fossils with almost isomorphic generations (Kenrick and Crane 1997). The antithetic theory (Bower 1908), by contrast, suggests that the sporophyte generation originated through the intercalation of mitotic divisions in the zygote before meiosis, resulting on a diploid embryo being retained on a gametophytic thallus. The sporophyte would then gradually evolve from a parasitic dependence on the gametophyte into a dominant, physiologically independent

organism. The antithetic theory is well supported by the phylogenetic relationships of streptophytes, and is now widely accepted.

Recent insights from evolutionary developmental studies also support the antithetic theory of interpolation of a multicellular diploid phase in an ancestral haplontic life cycle. In the unicellular chlorophyte *Chlamydomonas*, the formation of a heterodimeric complex with the proteins Gsm1 and Gsp1 is sufficient to initiate the diploid phase of the life cycle (Lee et al. 2008); Gsm1 and Gsp1 are members of the TALE superclass of homeobox proteins, which in plants include the KNOX and BEL classes (Mukherjee et al. 2009); KNOX and BEL proteins are important regulators of sporophyte development in land plants (Hake et al. 2004; Sakakibara et al. 2008), suggesting that the function of TALE proteins is restricted to the diploid phase of the life cycle. Similarly, the floral meristem regulator LEAFY also functions specifically in the sporophytes of mosses and angiosperms (Tanahashi et al. 2005). In contrast to the TALE and LEAFY proteins specific role in sporophyte development, other genes and regulatory mechanisms appear to have been recruited from the gametophyte to the sporophyte generation. A good example is the family of type II MADS-box transcription factors. In different charophyte algae, a single type II MADS-box gene functions during haploid reproductive cell differentiation (Tanabe et al. 2005). Type II MADS-box genes radiated in land plants and formed two groups: the MIKC^c and MIKC* (Henschel et al. 2002). MIKC* genes have retained a gametophyte function in bryophytes (Zobell et al. 2010) and angiosperms (Kofuji et al. 2003); however, MIKC^c genes are expressed in both the gametophyte and sporophyte tissues in mosses and ferns (Münster et al. 1997; Quodt et al. 2007; Singer et al. 2007), but are mostly restricted to the sporophyte in *Arabidopsis* (Kofuji et al. 2003), where they have become specialised to the point of being the most important floral homeotic genes. Another example was the discovery that *RSL* class I genes control the differentiation of root hairs in the angiosperm *Arabidopsis thaliana*, and rhizoid and caulonema cells in the moss *Physcomitrella patens* (Menand et al. 2007b). Root hairs, rhizoids and caulonemata are structures that fulfil similar rooting functions, but root hairs are tubular projections from epidermal cells of the root in the sporophytic life cycle stage, whereas rhizoids and caulonemata are filamentous structures that grow in the gametophytic life cycle stage. This suggests that *RSL* class I genes that controlled the development of rooting filaments in the gametophyte of early land plants were later recruited to control the development of root hairs in the sporophyte generation (Menand et al. 2007b). These examples suggest that the

elaboration of the sporophyte generation, and particularly the large radiation of morphology forms during the Devonian period, was partially achieved through the recruitment of genes and genetic mechanisms that had previously evolved and functioned in the gametophyte generation of charophytes and early land plants.

The cues that mediate the transition of the different stages of the life cycle are likely to be predominantly epigenetic. Okano et al. (2009) and Mosquna et al. (2009) have shown that *PpCLF* and *PpFIE*, moss genes encoding putative subunits of a Polycomb group complex that regulates epigenetic states through chromatin modification, are required for the correct establishment of sporophyte and gametophyte identity: loss of - function *Ppclf* and *Ppfie* mutants develop sporophyte-like bodies in the place of gametophores. The phenomenon of apogamy (development of sporophytes from gametophytes without fertilization) is long known to occur in fern and bryophyte species (Bell 1992), but the discoveries of Okano et al. and Mosquna et al. provide a glimpse into the molecular mechanisms that provide the epigenetic context of life cycle transitions.

1.2.2. Leaf evolution

Leaves have evolved multiple times during land plants evolution. Mosses and liverworts have leaf-like structures, but these are unrelated to the complex leaves of vascular plants; these have evolved independently as microphylls in lycophytes and several times as megaphylls in euphyllophytes. Microphylls, small and simple leaves with a single unbranched vein, are hypothesised to have evolved during the Silurian / early Devonian (Taylor et al. 2009) through either the vascularisation of stem enations, the reduction of flattened lateral branches or the sterilisation of sporangia (Crane and Kenrick 1997). The first leaves of euphyllophytes (megaphylls), which have a complex venation pattern, had evolved multiple times by the Late Devonian / Carboniferous, probably through the planation (flattening) and webbing of branch systems (Taylor et al. 2009). The megaphylls of seed plants and ferns are clearly not homologous, but determining the total number of independent origins of leaves in euphyllophytes depends on the resolution of a complex phylogeny of Devonian and Carboniferous fossils (Friedman et al. 2004; Boyce 2009).

The formation of leaves involves a transition from indeterminate growth in shoot apical meristems into a determinate growth programme. The indeterminacy of the shoot apical meristem is maintained by KNOX transcription factors; in different

angiosperms, the initiation of leaf determinate growth requires that KNOX genes are negatively regulated by ARP proteins in leaf primordia. Surprisingly, a similar KNOX-ARP mechanism operates during microphyll development in lycophytes, despite the independent origin of microphylls and megaphylls (Harrison et al. 2005). This suggests that the mechanisms for regulating determinacy and indeterminacy were present in the common ancestor of vascular plants and were recruited independently to control leaf initiation. By contrast, a mechanism involving Class III HD-Zip transcription factors that patterns stem vasculature was co-opted for the adaxial/abaxial patterning of leaves in seed plants but not in lycophytes (Floyd and Bowman 2006), reflecting the independent origin of microphylls and seed plant megaphylls.

Unlike in plants with simple leaves, where KNOX expression is restricted to the shoot apical meristem, KNOX genes are expressed in the leaf primordia of fern fronds (Bharathan et al. 2002; Harrison et al. 2005) or reactivated during the development of compound leaves in seed plants (reviewed in Hay and Tsiantis 2009). This possibly reflects a delayed determinacy and meristematic properties of compound leaves, which in turn allow the generation of complex leaf morphologies. Supporting this hypothesis, a high degree of natural variation in the morphology of tomato compound leaves has been found to be caused by a dosage effect of a KNOX gene (Kimura et al. 2008). The reactivation of KNOX expression during compound leaf development appears to have had multiple evolutionary origins (Hay and Tsiantis 2009) and provide a remarkable example of how small changes in the spatial expression of a transcription factor can cause the evolution of a multitude of different morphologies.

1.2.3. Root evolution

The successful colonisation of terrestrial environments involved the evolution of multicellular organs that actively penetrate the substrate, anchor the plant, weather the soil and retrieve the mineral nutrients necessary for plant growth. The rooting function in free-living gametophytes (and to some extent in a few aquatic charophytes and chlorophyte algae) is performed by a system of rhizoid filaments. However, true roots comprising a specialised axis, a root cap, an endodermis, and an endogenous origin of lateral branches are only found in the sporophytes of vascular plants (Raven and Edwards 2001). The earliest vascular plants did not have specialised root axes, but the zosterophyllophyte-lycophyte clade had evolved roots by the early Devonian; on the

other hand, there is no evidence of roots in other vascular plants until the middle Devonian (Gensel et al. 2001; Raven and Edwards 2001). This suggests that roots evolved at least twice in land plants and that the occurrence of an endodermis, endogenous branching and an endodermis in the roots of both lycophytes and euphyllophytes can be interpreted as the result of convergent evolution. Within the euphyllophytes, a fundamental difference in the anatomy of embryonic roots has suggested the existence of a further independent origin of roots in seed plants and in free-sporing plants (Raven and Edwards 2001). As discussed above, the development of root hairs in the root of angiosperms and of rhizoids in the gametophyte of mosses is controlled by RSL class I proteins (Menand et al. 2007b). Given the independent origin of roots in lycophytes and (possibly) in ferns, it will be interesting to determine if the development of root hairs/rhizoids in these groups has also involved the recruitment of *RSL* class I genes. Many questions remain regarding the exact homology of shoots and roots: in angiosperms, several regulatory factors that control shoot apical meristem (such as WUS and CLV3) have homologues that regulate the function of the root apical meristem (reviewed in Stahl and Simon 2010). It will be interesting to determine if these regulators are also required for the development of the root meristem of monilophytes and lycophytes.

1.2.4. Flower evolution

The evolution of seeds and flowers were major hallmarks in land plant evolution and the most important factors responsible for the dominance of gymnosperms and angiosperms on land floras for the past 250 million years. Little is known regarding the genetic mechanisms that guided the evolution of seeds in the Middle Devonian, but the evolution of flowers has received considerable attention by developmental and evolutionary biologists alike.

The earliest fossils of flowers are from the early Cretaceous (around 125 million years ago), and indicate that a rapid diversification of floral forms occurred very early in angiosperm evolution (Friis et al. 2006). The evolution of flowers in angiosperms involved the transformation of unisexual gymnosperm reproductive structures into a hermaphrodite structure. Different theories providing an explanation for this transformation have been proposed (reviewed in Specht and Bartlett 2009). One of the most sounding is based on the discovery that homologues of floral homeotic genes are

present in gymnosperms: classes B and C, in particular, are expressed in gymnosperm reproductive structures, with class B genes being specifically expressed in the male structures (reviewed in Melzer et al. 2010). According to the out of male (or out of female) hypothesis (Theißen et al. 2002), changes in the spatial expression pattern of class B genes in male (or female) cones, could have given rise to the hermaphroditic precursors of flowers.

Floral homeotic genes are central to the specification of flower organ identities and were probably a major driver of flower evolution: the 'sliding boundary' (Bowman 1997) and the 'fading borders' (Buzgo et al. 2004) models propose that changes in the spatial expression domains of homeotic genes result in gradual transitions in organ morphology. Different floral organs could also potentially arise through changes in the protein interactions of floral homeotic genes or in the promoters of their target genes (Irish 2009). Another factor that has promoted flower evolution was the multiple evolution of a floral axis of asymmetry (zygomorphy) in different plant lineages. Interestingly, a mechanism involving the TCP transcription factor CYCLOIDEA was independently recruited multiple times to establish bilateral symmetry in eudicots families (reviewed in Preston and Hileman 2009).

1.2.5. Evolution of transcription factors and signalling pathways

The sequencing of different plant genomes opened the door for powerful comparative genomic analyses. These have shown that most developmental genes are highly conserved in land plants. Nearly all the 50-60 transcription factor gene families found in angiosperms are present in basal land plants, but only 15-30 are present in chlorophyte algae (Richardt et al. 2007; Riaño-Pachon et al. 2008). This indicates that there was a large increase in the number of transcription factor families in the streptophyte lineage, and that the core set of land plant transcription factors is highly conserved. Nevertheless, the average size of each transcription factor gene family is substantially smaller in mosses (less than 10 genes per transcription factor family) than in angiosperms (20-25 genes per family) (Richardt et al. 2007). This suggests that there was a large expansion and diversification of transcription factor families on land, possibly associated with the elaboration of the multicellular body. Interestingly, despite having a smaller set of transcription factors, mosses appear to have more elaborated two-component signalling systems (involving histidine kinases and response regulators)

than angiosperms (Rensing et al. 2008). The increase in the complexity of plant transcription factor families is reminiscent of the evolution of transcription factors in metazoans: a wide range of transcription factor families and classes are present in demosponges (the most basal metazoan group), but not in choanoflagelates (unicellular organisms that are the sister group to metazoans)(Rokas 2008; Degnan et al. 2009). The ancestors of bilaterians and cnidarians later underwent an expansion and diversification of transcription factor families, which correlates with an increase in morphological and cell type complexity (Rokas 2008; Degnan et al. 2009).

A major factor driving transcription factor evolution in plants is the frequent occurrence of gene duplications, particularly whole genome duplications through autopolyploidy and allopolyploidy events. Many land plant species are polyploid (Cui et al. 2006), and possibly almost all angiosperm species (including *Arabidopsis thaliana*) are paleopolyploids, i.e. diploids with polyploid ancestors. The detection of collinearity between triplicate regions in rosid and asterid species suggests that there was a hexaploidy event in the common ancestral of the main eudicot lineages, around 150 million years ago (Tang 2008). Additional duplications occurred later independently in several lineages, including two duplications in the Brassicales ancestors of *Arabidopsis thaliana* (Van de Peer et al. 2009a). It is estimated that the three whole genome duplications are directly responsible for the generation of 60% of the *Arabidopsis* genes during the last 150 million years (Maere et al. 2005; Van de Peer et al. 2009a). Regulatory genes (including genes involved in transcription and signal transduction) are preferentially retained after large scale duplication events than after small scale gene duplications, probably because of dosage effects and the importance of maintaining a correct stoichiometric balance in protein complexes (Blanc et al. 2004; Maere et al. 2005). The three whole genome duplication events are calculated to be responsible for 90% of the *Arabidopsis* transcription factors created over the last 150 million years (Maere et al. 2005). Interestingly, many whole genome duplications occurred independently in several plant groups during the Cretaceous-Tertiary boundary 65 million years ago, a period of mass extinctions followed by extensive radiations (Fawcett et al. 2009). This suggests that whole genome duplications may confer a competitive advantage under changing environments and enhance the diversification potential of a lineage (Van de Peer et al. 2009b). Gene duplication can fuel evolution because a duplicate copy is free to evolve a novel function (neofunctionalisation) without compromising the function of the original gene. However, most retained duplicates

probably undergo a subfunctionalisation process instead, in which complementary loss-of-function mutations occur such that both genes are required to produce the full complement of functions of the single ancestral gene (Prince and Pickett 2002).

Mechanisms of gene regulation by RNA silencing are also conserved in land plants. The microRNA (miRNA) machinery appears to have evolved independently in animals and in plants (Axtell and Bowman 2008). In plants, miRNAs have been identified in the unicellular chlorophyte *Chlamydomonas*, but these are not homologous to any land plant miRNA (Molnár et al. 2007; Zhao et al. 2007). Dozens of miRNA families have been identified in land plants (Axtell and Bowman 2008). At least 16 of these miRNA families were present in the common ancestor of mosses and vascular plants and are highly conserved in other land plants (Tanzer et al. 2010). However, the majority of miRNAs is lineage specific and non-conserved (Axtell et al. 2007). In contrast to miRNAs, short interfering RNAs (siRNAs) are widely present in eukaryotic organisms; accordingly, they have been identified in chlorophytes and mosses (Zhao et al. 2007; Cho et al. 2008). The plant specific class of *trans*-acting siRNA (ta-siRNA) is present in mosses (Talmor-Neiman et al. 2006) and, less clearly, in *Chlamydomonas* (Zhao et al. 2007).

The major components of auxin signalling in land plants (*AUX-IAA*, *ARFs* and *TIR1-AFBs*) are absent from chlorophyte algae but present in land plants (Lau et al. 2009), suggesting that it evolved in the streptophyte lineage. The auxin signalling response is functional in mosses (Bierfreund et al. 2003; Hayashi et al. 2008) and auxin polar transport has been found to occur in different moss structures (discussed later in this Chapter). All the genetic components required for cytokinin signalling are also present in mosses but not in chlorophytes (Pils and Heyl 2009). Nevertheless, there was an important expansion of most of the gene families involved in cytokinin signalling in vascular plants (Pils and Heyl 2009). *P. patens* has candidate gibberellin (GA) biosynthetic genes and GA–DELLA signalling components (Hirano et al. 2007; Vandenbussche et al. 2007; Yasumura et al. 2007; Anterola and Shanle 2008), but it does not have GA-dependent GID1–DELLA signalling. This mechanism appears to have evolved in vascular plants (Hirano et al. 2007; Yasumura et al. 2007), while the characteristic DELLA-GA mediated growth restraint probably evolved only after the divergence of lycophytes from other vascular plants (Yasumura et al. 2007). This suggests that the GA signalling pathways evolved gradually in land plants. An abscisic acid (ABA) signalling response is present in mosses (Knight et al. 1995; Komatsu et al. 2009) and, accordingly, and the genome of *P. patens* encodes homologues of receptors

and transcription factors involved in ABA signalling (Rensing et al. 2008). The analysis of the *P. patens* genome suggests that signalling through jasmonic acid, ethylene or brassinosteroids evolved after the divergence of mosses from other land plants (Rensing et al. 2008).

Most of the evidence used to infer the evolutionary origin of different signalling pathways is based on the genomic identification of homologues to known biosynthetic enzymes receptors or signal transducers; however, there are several reports of developmental effects caused by some of these hormones in basal plants that do not fit with the genomic predictions. It is possible that separate plant lineages have evolved slightly different signalling pathways, and it will take more that comparative genomics to identify these mechanisms.

1.3. Moss development

Liverworts, mosses and hornworts are the earliest diverging groups of land plants. Their phylogenetic position and the simplicity of their morphology and development make them excellent systems to study the evolution of developmental processes in plants. The moss *Physcomitrella patens*, in particular, has been intensively studied over the past decade: its full genomic sequence has been determined (Rensing et al. 2008) and a high rate of homologous recombination allows the precise targeting of genomic sequences upon transformation; this makes the generation of knockout mutants possible for virtually every gene (Schaefer 2001). The following sections are a brief description of moss development.

1.3.1. Protonema

Mosses begin their haploid development from a single-celled spore (Schofield 1985). Upon germination, a uniseriate filament grows by successive divisions of a tip growing apical cell. The filament, composed of chloronema cells with numerous chloroplasts, can also originate branches from subapical cells. After several days of chloronemata development, some apical cells begin to differentiate a second cell type, the caulonema. Unlike chloronema, caulonema cells have few and elongated chloroplasts, and their growth rate is much higher. Together, chloronema and caulonema filaments constitute the protonema, the first stage of gametophytic

development. There are several differences between the two cell types that underlie two distinct functions: chloronemata carry out high rates of photosynthesis, while caulonemata increase the size of the colony.

Chloronema cells have numerous and well-developed chloroplasts disposed around a large central vacuole (Duckett et al. 1998). Apical cells extend at a rate of 2-6 μm.h^{-1} and divide every 10-12 hours (Schumaker and Dietrich 1997; Duckett et al. 1998; Cove 2005). Unlike most plant cells, chloronema cells arrest in the G2/M transition of the cell cycle (Schween et al. 2003). Caulonema cells have characteristic oblique cross cell walls, a strong cytoplasmic polarity and their nuclei typically undergo endoreplication (Duckett et al. 1998). Plasmodesmata in cross walls have expanded central lamina, like those associated with intense trafficking in vascular plants (Duckett et al. 1998). Like chloronema, caulonema filaments grow by divisions of a tip growing apical cell (Menand et al. 2007a). They extend at a rate greater than 20 μm.h^{-1} and divide every 6-8 hours (Schumaker and Dietrich 1997; Duckett et al. 1998; Cove 2005). Caulonema and chloronema cells show distinct phototropic responses (Cove and Knight 1987) which are mediated by phytochromes (Mittmann et al 2004).

Subapical cells can develop side-branch initials that give rise to secondary chloronema, caulonema or buds (Cove 2005). Side-branch initials always arise from the distal side of subapical cells, just behind the cross wall and in the flank opposite the nucleus; in a caulonema filament, this generally occurs in the third cell from the tip (Schmiedel and Schnepf 1979). In *P. patens* and in *Physcomitrium turbinatum*, caulonema side-branch initials develop near the wide end of the oblique cross walls (Jensen 1981; Schumaker and Dietrich 1997). In *P. turbinatum*, particularly, the orientation of the oblique cross wall is determined by the direction of light; this means that the place where a side-branch initial emerges is determined hours before, when the subapical cell was originated (Jensen 1981). However, in F. *hygrometrica*, the cell wall flank where the side-branch initial develops is independent of the orientation of the oblique cross cell wall (Schmiedel and Schnepf 1979). The first visible sign is a bulging on the cell wall, caused by apposition of new wall material. The new cell wall later pierces the older one; the nucleus migrates and divides in the base of the outgrowth, and a new cross wall forms (Schmiedel and Schnepf 1979).

Protonemal filaments exhibit a strong polarity, more pronounced in caulonema cells, manifested in the cell division asymmetries, side-branch initials positioning and in cytoplasmic organization. The cytoskeleton has been shown to be crucial in the

establishment of this polarity. Protonema apical cells contain microtubule arrays that extend to the apical dome; treatment with the microtubule-depolymeriser drug cremart causes the cell to bend, develop new 'tips' or swell at the tip (Doonan et al. 1988), indicating that growth can still occur in the absence of microtubules, but not in a polarized manner. It was proposed that microtubules can respond to environmental signals like light and gravity and accordingly reorientate the filament growth (Doonan 1991). Actin microfilaments are also axially oriented towards the apical dome. Experiments with the drug cytochalasin first showed that actin microfilaments are required for tip growth in protonema filaments (Schmiedel and Schnepf 1980; Doonan et al. 1988). More recently, several knockouts and knockdowns of actin associated proteins such as profilin (Vidali et al. 2007), myosin XI (Vidali et al. 2009), actin depolymerising factor (ADF) (Augustine et al. 2008), Arp2/3 complex subunits (Finka et al. 2008; Harries et al. 2005; Perroud and Quatrano 2006) and Wave/SCAR complex subunits (Perroud and Quatrano 2008) have been shown to cause a disruption of the subapical actin network and to inhibit tip growth. Interestingly, the development of gametophores is unaffected in many of these mutants, despite the strong defects that they have in the protonema and rhizoid tip growing filaments.

Cell differentiation in the protonema

Protonema development is extremely plastic: caulonema can give rise to chloronema, chloronema rise to caulonema, and both structures rise to buds; in addition, isolated cells can regenerate an entire colony (Cove et al. 2006). The chloronema-caulonema transition, in particular, is modulated by an array of endogenous and exogenous factors.

A rich substrate or the addition of ammonium tartrate as the nitrogen source causes the preferential development of chloronema filaments, while poor media generally results in an increase in the number of caulonema filaments. However, Thelander et al. (2005) observed that high energy conditions such as external glucose or high light induce caulonemata formation and proposed an homeostatic model for the regulation of chloronemata and caulonemata formation: under low energy conditions, chloronemata growth is stimulated, leading to higher rates of photosynthesis; when energy is readily available, the colony can afford to produce caulonemata and that way increase the size of the colony. Interestingly, a knockout of a sulfite reductase causes an

inhibition of caulonema differentiation (Wiedemann et al. 2010), indicating that not only carbon and nitrogen, but also sulphate metabolism, play an important role in moss morphogenesis.

Light is a major factor in determining protonema patterning. The formation of side-branches is mediated by at least 4 distinct photoreceptor systems: nuclear-localised cryptochrome (blue light receptors) and red light receptors induce branch formation, while plasma membrane-localised phototropins (blue light receptors) and phytochromes (red and far-red light receptors) determine the position of branches (Imaizumi et al. 2002; Uenaka et al. 2005). The blue-light response is mediated by SBP-box transcription factors, which are negatively regulated by cryptochromes (Riese et al. 2008). Cryptochrome blue light signals inhibit auxin-induced caulonema differentiation and the expression of auxin inducible genes (Imaizumi et al. 2002), suggesting the existence of close links between light and auxin signalling during protonema development.

Auxin treatment increases the number of caulonema filaments and rhizoids (Johri and Desai 1973; Ashton et al. 1979). Immunoenzymatic and immunofluorescence assays initially showed that auxin levels were higher in caulonema than in chloronema filaments (Atzorn et al. 1989; Bopp and Atzorn 1992). More recently, a fusion of the promoter of the *Glycine max GH3* gene (which is responsive to auxin) with GUS has been extensively used to detect the sites of auxin response in *P. patens*: these are predominantly caulonema filaments, gametophore buds, the base of the gametophore stem, basal and mid-stem rhizoids, zygote and embryos, and the seta and foot of mature sporophytes (Bieurfreund et al. 2003; Fujita et al. 2008; Eklund et al. 2010). The genome of *P. patens* encodes the major gene families that are implicated in auxin homeostasis and signalling in angiosperms (Rensing et al. 2008) and some of these have been shown to be functionally conserved: GH3 proteins, which synthesise auxin conjugates, regulate auxin homeostasis in *P. patens* (Ludwig-Müller et al. 2009), while the responses of *P. patens* to an auxin antagonist suggest that the TIR1/AUX-IAA/ARF auxin signalling pathway is functionally conserved in mosses (Hayashi et al. 2008). Polar transport of auxin has been demonstrated to occur in caulonema filaments, rhizoids and sporophytes (Rose and Bopp 1983; Bopp and Atzorn 1992; Poli et al. 2003, Fujita et al. 2008), but different pieces of evidence suggest that it is local auxin biosynthesis that plays a major role in auxin peak formation in mosses: 1) there is no polar transport of auxin in gametophore shoots (Fujita et al. 2008); 2) the expression of *SHI* genes, which regulate auxin biosynthesis, coincides with the sites of auxin response (Eklund et al. 2010) and 3)

the 3 *Physcomitrella* PIN proteins are functionally related to the PIN5-type proteins that regulate subcellular homeostasis of auxin, and not to the PIN1-type proteins that are responsible for auxin efflux from cell to cell (Mravec et al. 2009).

As mentioned earlier, the genome of *P. patens* encodes candidate GA biosynthetic genes signalling components. However, apart from a possible effect in spore germination (Anterola et al. 2009), *P. patens* lacks a detectable growth response to GA and there is no evidence for a functional conservation of the GA-DELLA signalling (Hirano et al. 2007; Yasumura et al. 2007). Cytokinins and cyclic adenosine monophosphate (cAMP) are thought to be involved in the differentiation of chloronemata cells. Although its role in plants is not very clear (Taiz and Zeiger 2002), cAMP was shown to be present (Handa and Johri 1976) and rapidly metabolized (Sharma and Johri 1982) in *F. hygrometrica*. The addition of exogenous cAMP inhibits the effect of exogenous auxin (Handa and Johri 1979), suggesting that the two substances have an antagonistic role. Addition of exogenous cytokinins also results in an inhibition of caulonema formation and increase of chloronema branching (Thelander et al. 2005), but their main effect is bud induction, which is discussed below.

1.3.2. Gametophores and rhizoids

A few weeks after spore germination, some protonemal side-branch initials develop into gametophore buds, initiating the second stage of gametophytic development. Buds develop preferentially from caulonema filaments, although they can also develop, more rarely, from chloronemata. Bud formation is induced by red light (Imaizumi et al. 2002) and by exogenous cytokinin (Ashton et al. 1979). Calcium plays a major role in bud formation. There is an increase in membrane-associated calcium in the presumptive initial cell site after addition of exogenous cytokinin (Saunders and Hepler 1981), while treatment of protonemata with calcium ionophores or calcium channel agonists results in initial cell formation on virtually every cell (Saunders and Hepler 1982; Conrad and Hepler 1988).

Gametophore formation changes the filamentous, two dimensional organization of protonema into a three dimensional structure. A side-branch initial first undergoes four divisions that result in the formation of a tetrahedral-shaped shoot initial cell. This shoot initial cell then undergoes asymmetric, self-replacing divisions in 3 planes that generate leaf initials in a spiral pattern (Harrison et al. 2009). Application of exogenous

auxin causes an increase in the stem length (Hayashi et al. 2008; Fujita et al. 2008); conversely, *Ppshi* mutants (which have reduced auxin biosynthesis) and cryptochrome blue light signals (which inhibits auxin responses) induce smaller internodal lengths (Imaizumi et al. 2002; Eklund et al. 2010), indicating that auxin also plays a role in patterning gametophore development. Gametophore leaves consist of a single layer of cells, except in the conducting tissue (midrib) that develops in adult leaves. The planar growth of leaves is generated by asymmetric, self replacing divisions of leaf initials in 2 planes. Further unevenly distributed divisions in the leaf contribute to the final leaf shape (Harrison et al. 2009).

During the transition from filamentous (protonema) to intercalary (gametophores) growth, the cytoskeleton changes dramatically to adopt a typical angiosperm organization: interphase microtubules become predominantly cortical and the pre-prophase band, a land plant specific structure, appears (Doonan et al. 1987). A knockout of PpTON1, a protein required for the assembly of the pre-prophase band, causes strong morphological defects in gametophores, but not in the protonema (Spinner et al. 2010).

Rhizoids are pigmented filaments that develop from epidermal cells of gametophores. The first rhizoids form at the base of the gametophores (basal rhizoids). Later, rhizoids (mid-stem rhizoids) develop also from epidermal cells below adult leaves (Sakakibara et al. 2003). Like caulonema filaments, rhizoids have distinctive oblique cross walls, few or no chloroplasts and elongate by fast tip-growth (Duckett et al. 1998). Many mutations or treatments (such as auxin) that affect caulonema differentiation also affect rhizoid formation (Ashton et al. 1979; Sakakibara et al. 2003; Menand et al. 2007b). In fact, there are many similarities between caulonema filaments and rhizoids, and the two cell types are readily interchangeable. However, they can usually be easily distinguished based on their anatomical origin and branching frequency (most rhizoids do not form branches).

Later in development, sex organs (antheridia and archegonia) are produced in the apex of the shoot. The antheridium produces numerous biflagellate male gametes, while the archegonium encloses a single egg (Schofield 1985). If water is available, antherozoids swim to the archegonium and fertilise the egg, originating the diploid zygote that initiates the sporophytic generation.

1.3.3. Sporophyte

The initiation of sporophytic versus gametophytic development is controlled by an epigenetic mechanism involving a Polycomb group complex, which has been proposed to repress the initial stages of sporophyte development in the haploid generation (Mosquna et al. 2009; Okano et al. 2009). After fertilisation, the zygote undergoes a transverse division, which is controlled by homologues of the floral regulator LEAFY (Tanahashi et al. 2005). The basal cell resulting from this first division further divides to become the foot, while the apical cell is the precursor of the remainder of the sporophyte (Schofield 1985). The division of this apical cell is controlled by KNOX class 1 genes (Sakakibara et al. 2008), which also regulate shoot apical meristem development in the sporophyte of angiosperms. The cells of the lower part of the archegonium divide and form a gametophytic layer (the calyptra) that protects the embryo. As the embryo grows, the calyptra is teared apart from the gametophore, forming a protective cap around the sporophyte tip (Schofield 1985). The seta elongates through the activity of a subapical intercalary meristem (French and Paolillo 1975); seta elongation can be increased through the application of exogenous auxin (Poli et al. 2003). Unlike in the gametophore, a polar transport (basipetal) of auxin has been detected in the sporophytes of different moss species (Poli et al. 2003; Fujita et al. 2008).

The mature sporophyte consists of a foot that penetrates the gametophore, a seta and a sporangium (capsule) where meiotic divisions produce haploid spores. The sporangium has a basal apophysis region, aerenchymal tissue, stomata (with single guard cells) in the base, spore sacs surrounding a central columella and peristome teeth covered by an operculum (Sack and Paolillo 1983; Miller and Miller 2006). Unlike *F. hygrometrica* and most other mosses, the sporophyte of *P. patens* lacks an operculum and a peristome, and it has a very short seta. A mature *P. patens* sporangium produces about 4000 spores (Cove 2005).

Chapter 2

Evolution of bHLH proteins in plants

2.1. Introduction[1]

The basic helix-loop-helix (bHLH) domain is a highly conserved amino acid motif that defines a group of transcription factors. It was originally described in animals (Murre et al. 1989) and soon discovered in all of the major eukaryotic lineages. Proteins that contain a bHLH domain (referred to as bHLH proteins) are involved in a myriad of regulatory processes. Their functions include regulating neurogenesis, myogenesis and heart development in animals (Massari and Murre 2000; Jones 2004), controlling phosphate uptake and glycolysis in yeast (Robinson and Lopes 2000), or modulating secondary metabolism pathways, epidermal differentiation and responses to environmental factors in plants (Ramsay and Glover 2005; Castillon et al. 2007).

The bHLH domain consists of 50-60 amino acids that form two distinct segments: a stretch of 10-15 predominantly basic amino acids (the basic region) and a section of roughly 40 amino acids predicted to form two amphipathic α-helices separated by a loop of variable length (the helix-loop-helix region). Structural analyses of mammalian and yeast bHLH proteins showed that the basic region forms the main interface where contact with DNA occurs, while the two helices promote the formation of homo- or hetero-dimers between bHLH proteins, a prerequisite for DNA binding to occur (Jones 2004).

Phylogenetic analyses have classified the diversity of bHLH proteins into a number of distinct groups. Over 50 bHLH proteins are encoded in the genomes of most animals (metazoans) and are typically classified into 6 major groups (A-F), based on their ability to bind DNA (Atchley and Fitch 1997; Ledent and Vervoort 2001; Jones 2004). Detailed analyses using whole genome sequences showed that animal bHLH could be further classified in several smaller subfamilies that are highly conserved across major metazoan lineages (Ledent and Vervoort 2001; Simionato et al. 2007). Phylogenetic analyses indicate that 44 of these subfamilies were present in the common ancestor of all bilaterians, which is thought to have existed sometime before 600 million years ago (Simionato et al. 2007). The genomes of *Arabidopsis thaliana* and *Oryza sativa* (rice) encode even more bHLH sequences than animals. Different phylogenetic studies proposed the classification of plant bHLH into 15-25 subgroups (Buck and Atchley 2003;

[1] The results in this chapter (except section 2.3.6.) were published in:
Pires N, Dolan L (2010) Origin and Diversification of Basic-Helix-Loop-Helix Proteins in Plants. Mol Biol Evol 27:862-74; Pires N, Dolan L. (2010) Early evolution of bHLH proteins in plants. Plant Signal Behav 5 (7).

Heim et al. 2003; Toledo-Ortiz et al. 2003; Li et al. 2006b). However, the origin and evolutionary history of these groups cannot be understood using *A. thaliana* and *O. sativa* sequences alone. The characterization of the evolution of plant bHLH diversity requires the phylogenetic analysis of bHLH proteins from a more diverse selection of plants, including algae, bryophytes and different lineages of vascular plants.

In this chapter it is shown that the plant bHLH family is monophyletic and underwent a major radiation before the evolution of the mosses[2]. The bHLH groups established in the early land plants over 400 million years ago were conserved during subsequent plant evolution, although there were many gene duplications and losses within these groups. This analysis defines 26 subfamilies that represent deep evolutionary relationships between plant bHLH proteins.

---

[2] Plants are defined as the organisms that are likely to have been derived from the primary endosymbiotic event that gave rise to the red algae, chlorophytes and land plants (Rodríguez-Ezpeleta et al. 2005)

2.2. Materials and methods

2.2.1. Sequence retrieval

The *Arabidopsis thaliana* bHLH reported by Bailey et al. (2003), Heim et al. (2003) and Toledo-Ortiz et al. (2003) were retrieved from TAIR (http://www.arabidopsis.org/). A clear bHLH domain was not found in At1g31050 (AtbHLH111) and At1g22380 (AtbHLH152), so they were not further used in this study; At2g20095 (AtbHLH133) and At4g38071 (AtbHLH131) could not be found in any database. A dataset of predicted *Oryza sativa* L. ssp. *japonica* bHLH proteins was retrieved from the Plant TFDB (Guo et al. 2008) and combined with the bHLH protein sequences reported by Li et al. (2006b), retrieved from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/). Eleven new proteins were numbered following the nomenclature style of Li et al. (2006b), while a clear bHLH was not found in Os01g65080 (OsbHLH033), Os04g35000 (OsbHLH145), Os11g02054 (OsbHLH160) and Os12g02020 (OsbHLH161). A dataset of predicted *Physcomitrella patens* bHLH was retrieved from the plant TFDB (Guo et al. 2008). A direct search of genes annotated as bHLH was performed on the genome assembly of *Selaginella moellendorffii* v1.0 (http://www.jgi.doe.gov/). HMMsearch (Eddy 1998) was used to screen the genome assemblies of *Cyanidioschyzon merolae* (Matsuzaki et al. 2004), *Chlamydomonas reinhardtii* v3.0 (Merchant et al. 2007), *Ostreoscoccus tauri* v2.0 (Palenik et al. 2007), *Thalassiosira pseudonana* v3.0 (Armbrust et al. 2004), and the draft assemblies of *Chlorella vulgaris* C-169 and *Volvox carteri* (http://www.jgi.doe.gov/) with the PFAM profile Hidden Markov Model (pHMM) HLH_ls.hmm (http://pfam.sanger.ac.uk/).

Five *Homo sapiens* and four *Amphimedon queenslandica* (demosponge) representative sequences of the major metazoan groups of bHLH proteins (based on Jones 2004; Simionato et al. 2007) were retrieved from Genbank; group F proteins are not clearly alignable to other bHLH (Ledent et al. 2002) and so they were not used in this study. The *Saccharomyces cerevisiae* bHLH proteins reported by Robinson and Lopes (2000) were retrieved from http://www.yeastgenome.org/.

For simplicity, all sequences were renamed according to the Supplementary File 1[3]. The complete amino acid sequence of all proteins can be found in Supplementary File 2.

2.2.2. Alignment and phylogenetic analysis

Protein sequences were pre-aligned using HMMalign (Eddy 1998) and the pHMM HLH_ls.hmm from PFAM (http://pfam.sanger.ac.uk/). The bHLH region was then extensively manually aligned in BioEdit (http://www.mbio.ncsu.edu/BioEdit/ BioEdit.html). Unambiguous aligned positions were used for the subsequent phylogenetic analyses (Supplementary File 3; Supplementary File 3B). The JTT model was selected as the best-fitting amino acid substitution model with the Akaike Information Criterion implemented in ProtTest (Abascal et al. 2005). The maximum likelihood (ML) analyses were done with the program PhyML version 3.0.1 (Guindon and Gascuel 2003) using the JTT model of amino acid substitution, an estimated gamma distribution parameter and a Shimodaira-Hasegawa-like aLRT. The PHYLIP package version 3.67 (Felsenstein 1989) was used to perform 100 bootstrap replicas of a neighbour joining tree based on a JTT distance matrix. PAUP* version 4.0b10 (Swofford 2003) was used to perform 100 bootstrap replicas of a maximum parsimony tree. The Bayesian analysis was performed with MrBayes version 3.1.2 (http://mrbayes.csit.fsu.edu/): two independent runs were computed for 10 million generations, at which point the standard deviation of split frequencies was less than 0.01; one tree was saved every 100 generations and 75,000 trees from each run were summarized to give rise to the final cladogram. The reconciliation tree was calculated using Notung 2.6 (Vernot et al. 2008). All trees were visualized using the program Figtree (http://tree.bio.ed.ac.uk/software/figtree/).

Alignments of the bHLH domain of related sequences were used to build pHMMs with HMMbuild from the HMMER2.0 suite of programs (Eddy 1998). The pHMMs were used to identify and classify protein sequences not used in the phylogenetic analyses in the plant bHLH subfamilies.

---

[3] Supplementary Files are available from the CD that accompanies this thesis.

2.2.3. Detection of conserved motifs

The MEME and FIMO software (Bailey and Elkan 1994) were used to discover patterns in the complete amino acid sequences of plant bHLH proteins. Each motif was individually checked so that incorrect or insignificant matches were discarded. The complete plant amino acid sequences were also screened against the PFAM 23.0 database (http://pfam.sanger.ac.uk/).

2.3. Results

2.3.1. All the major groups of land plants have large numbers of bHLH proteins

Previous phylogenetic analyses of plant bHLH proteins were based on the genome sequences of *A. thaliana* and *O. sativa* (Buck and Atchley 2003; Heim et al. 2003; Toledo-Ortiz et al. 2003; Li et al. 2006b). This provided a useful, but limited, phylogenetic framework for the classification of bHLH proteins in flowering plants (angiosperms). Nevertheless, it provided no insight into the diversity of this family in the earlier diverging groups of land plants. To determine if these subfamilies were angiosperm specific or if they arose earlier in plant evolution and to understand the deeper evolutionary history of this family in plants, bHLH protein coding sequences were searched in the complete genome of the lycophyte *Selaginella moellendorffii*, the moss *Physcomitrella patens*, the chlorophytes *Volvox carteri*, *Chlamydomonas reinhardtii*, *Chlorella vulgaris* and *Ostreoscoccus tauri* and the red alga *Cyanidioschyzon merolae*. These sequences were combined with the previously reported *A. thaliana* and *O. sativa* sequences to generate a primary dataset consisting of 544 bHLH sequences representing the major evolutionary lineages of plants (Fig. 2.1). This dataset was then extended to include proteins from selected eukaryotic groups: the full set of bHLH proteins encoded in the genomes of the diatom *Thalassiosira pseudonana* and the fungi *Saccharomyces cerevisiae*, plus representative bHLH sequences from the sponge *Amphimedon queenslandica* and *Homo sapiens* (Fig. 2.1).

There are large numbers of bHLH proteins in all species of land plants (embryophytes) sequenced to date. *A. thaliana* and *O. sativa* have over 150 bHLH sequences in their genomes, making it the second largest family of transcription factors in angiosperms (Xiong et al. 2005). Approximately 100 bHLH proteins are encoded in the genomes of the lycophyte *S. moellendorffii* and the moss *P. patens* (Fig. 2.1). In contrast, there are less than five bHLH-encoding sequences in the genome of each chlorophyte and red alga examined. Other unicellular eukaryotic organisms such as the diatom *T. pseudonana* and *S. cerevisiae* also have small numbers (less than 10) of bHLH proteins (Fig. 2.1; Robinson and Lopes 2000). In animals, the sponge *A. queenslandica* has 16 bHLH-encoding genes while most bilaterians have over 50 genes (Simionato et al. 2007).

Species	Number of bHLH
Oryza sativa	173
Arabidopsis thaliana	158
Selaginella moellendorffii	103
Physcomitrella patens	98
Volvox carteri	3
Chlamydomonas reinhardtii	3
Chlorella vulgaris	4
Ostreococcus tauri	1
Cyanidioschyzon merolae	1
Thalassiosira pseudonana	5
Saccharomyces cerevisiae	8[a]
Amphimedon queenslandica	16[b]
Homo sapiens	118[b]
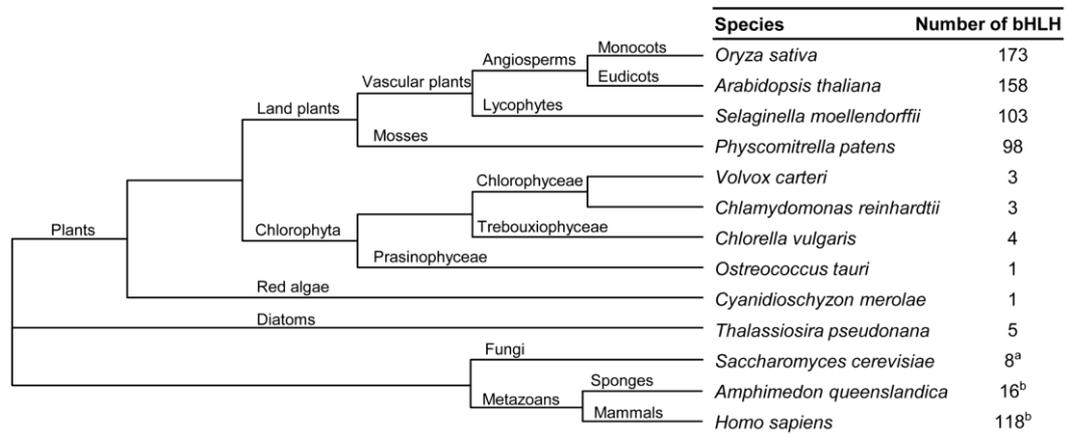
Figure 2.1 – Number of bHLH sequences in present in the different species.

The total number of bHLH proteins found in the genome of each species is indicated. The cladogram is based on the current view of plant and eukaryotic phylogeny (Baldauf 2003; Lewis and McCourt 2004; Rodríguez-Ezpeleta et al. 2005); [a] (Robinson and Lopes 2000); [b] (Simionato et al. 2007).

Animals and land plants have considerably more bHLH sequences than other eukaryotic organisms. This suggests that the increase in the number of bHLH proteins occurred independently during the evolution of plants and animals.

2.3.2. Key amino acid residues are highly conserved between plant and metazoan bHLH proteins

To characterize the molecular evolution of plant bHLH proteins, the retrieved amino acid sequences were aligned in the conserved bHLH region (Fig. 2.2, Supplementary File 3). The first 10-15 amino acids correspond to the basic region, where most interactions with the DNA are made (Ferré-D'Amaré et al. 1993). Most animal bHLH proteins bind to hexanucleotide sequences (5'-CANNTG-3') known as E-boxes. All E-box binding bHLH proteins have a glutamic acid (E) residue at position 9 that directly contacts the DNA at the CA nucleotides of the hexanucleotide sequence (Ferré-D'Amaré et al. 1993; Atchley et al. 1999). In plants, the critical E_9 residue is present in 74% of the proteins analysed (Supplementary File 3). Other positions of the basic region allow a better discrimination of the target DNA sequences, and are easily distinguishable in the major animal bHLH groups (Atchley and Fitch 1997; Ledent and Vervoort 2001; Jones 2004; Atchley and Zhao 2007). Animal group A proteins bind the CAGCTG (or CACCTG) E-box configuration and have a diagnostic arginine (R) at position 8. Animal group B proteins have a lysine (K) or histidine (H) residue at position 5 and an R at position 13 and bind the CACGTG (or CATGTTG) E-box configuration. In plants, 53% of the bHLH proteins have the characteristic animal group B configuration H_5-E_9-R_{13} and only 8% have the typical R_8-E_9 found in animal group A. This suggests that most plant bHLH proteins also bind to E-boxes. Indeed, a number of plant bHLH proteins have been shown to bind the CACGTG sequence (e.g. Martínez-García et al. 2000; Toledo-Ortiz et al. 2003; Qian et al. 2007), which is classically known in plants as a G-box motif (Giuliano et al. 1988). Group E animal proteins, that bind N-boxes (CACGCG or CACGAG), have the same H_5-E_9-R_{13} configuration as group B and a proline (P) at position 6. This configuration is absent in all the 544 plant bHLH proteins analysed. The remaining animal bHLH groups C and F proteins contain extra PAS and COE domains, not found in plant bHLH proteins, while group D proteins are atypical bHLH without a basic domain. 11% of the plant proteins have a conserved Q_5-A_9-R_{13} motif (Supplementary File 3), not present in animals.

Figure 2.2 – Alignment of the bHLH domain of representative plant proteins.

A representative of each of the 26 subfamilies of plant bHLH is shown, together with the human protein Max, a well characterized bHLH protein. The shaded boxes indicate the position of the DNA-binding basic region, the two α-helixes and the variable loop region (Ferré-D'Amaré et al. 1993). The numbering of the amino acids follows (Atchley and Fitch 1997). This is a subset of the full alignment with all the proteins used in this study (Supplementary File 3).

This raises the interesting possibility that these proteins bind to a novel target DNA sequence. Other frequent basic amino acids in animal bHLH, such as R in positions 10 and 12, are also highly conserved in plants (73% and 90%, respectively).

The α-helices promote the formation of homo- or hetero-dimeric complexes between bHLH proteins. The structure of a dimer is stabilized by the hydrophobic amino acids isoleucine (I), leucine (L) and valine (V) in conserved positions in the bHLH domain (Ferré-D'Amaré et al. 1993). These positions are highly conserved in animals (Atchley et al. 1999) and in plants (Fig. 2.2). An L residue is present in sites 23 and 64 in 99% and 96% of the plant proteins, and in 98% and 80% of the animal proteins, respectively. Sites 54 and 61 have an I, L or V in 99% and 93% of the plant proteins, and in 98% and 93% of the animal proteins, respectively. A conserved P breaks the first helix and starts a loop of variable length (usually 6-9 residues in plants). Some loop residues are also conserved: site 47 is K or R in 88% of the plant proteins (Supplementary File 3) and 82% of the animal proteins (Atchley et al. 1999).

The high degree of sequence similarity between the bHLH domain of plant and animal proteins, particularly in key DNA-interacting basic amino acids and in helix-stabilising hydrophobic amino acids, indicates that the molecular structure and transcription factor activity of bHLH proteins are conserved between animals and plants.

2.3.3. Twenty bHLH subfamilies found in flowering plants were also present in early land plants

To understand the evolutionary relationships between plant bHLH proteins, conserved regions of the alignment shown in Supplementary File 3 were used to compute phylogenetic trees. A maximum likelihood (ML) analysis shows that proteins from different species cluster together in compact clades with high support values (Fig. 2.3; Supplementary File 4). Maximum parsimony (MP) and neighbour joining (NJ) analyses support the existence of most of these clades (Supplementary File 4). Based on the topology of the trees, clade support values, branch lengths and visual inspection of the bHLH amino acid sequences, 26 subfamilies of bHLH proteins were defined (Fig. 2.3; Supplementary File 4). These subfamilies are mostly consistent with the groups proposed by previous phylogenetic analyses of plant bHLH using *A. thaliana* and *O. sativa* sequences alone (Buck and Atchley 2003; Heim et al. 2003; Toledo-Ortiz et al. 2003; Li et al. 2006b).
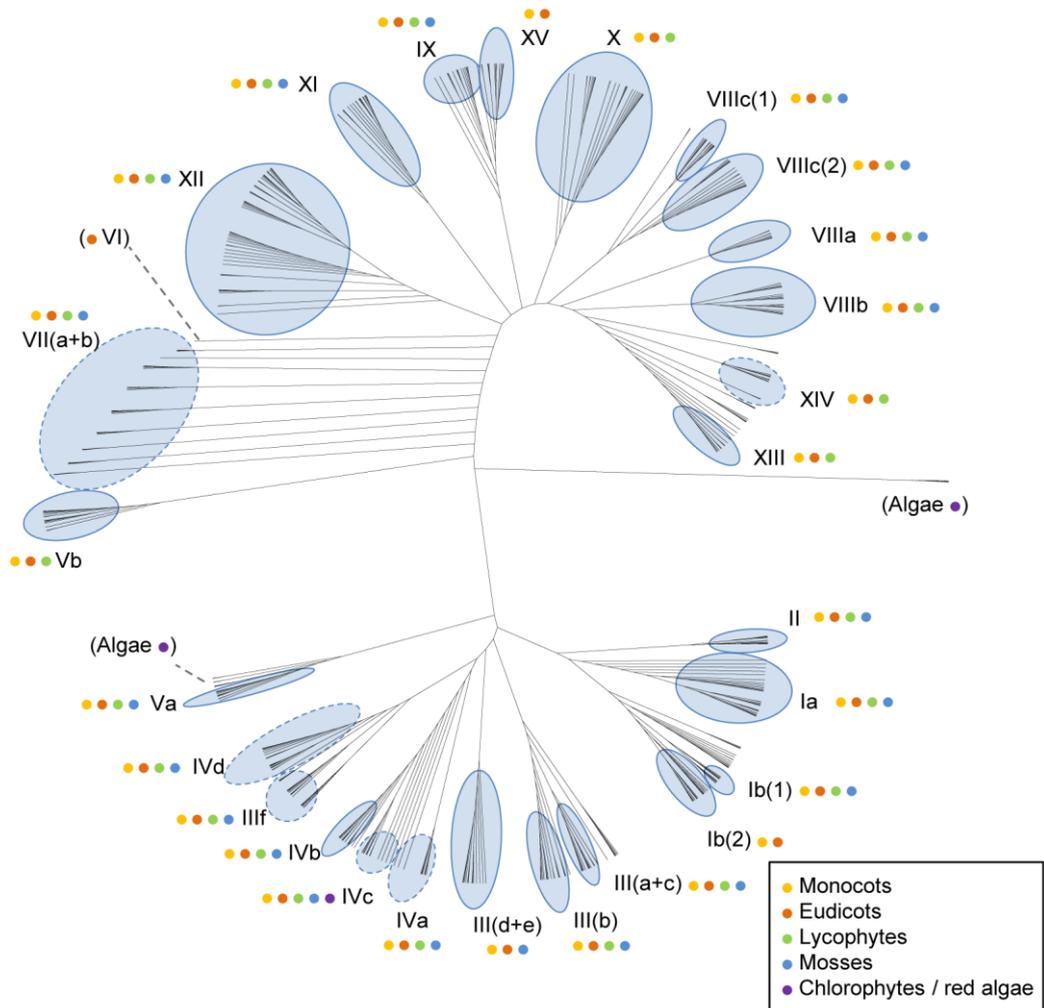
Figure 2.3 – 20 subfamilies of bHLH were already established in the common ancestral of vascular plants and mosses.

Maximum likelihood analysis of 544 plant bHLH, shown as an unrooted cladogram. The blue balloons delineate the 26 subfamilies of plant bHLH proteins; ballons encircled by a dashed line indicate that there is little evidence for monophyly. Coloured dots symbolize the species to which the proteins in each group belong (yellow: *Oryza sativa* (monocot); red: *Arabidopsis thaliana* (eudicot); green: *Selaginella moellendorffii* (lycophyte); blue: *Physcomitrella patens* (moss); purple: *Volvox carteri*, *Chlamydomonas reinhardtii*, *Chlorella vulgaris*, *Ostreococcus tauri*, *Cyanidioschyzon merolae* (chlorophytes and red algae). A full tree with protein names, proportional branch lengths and clade support values is given in Supplementary File 4.
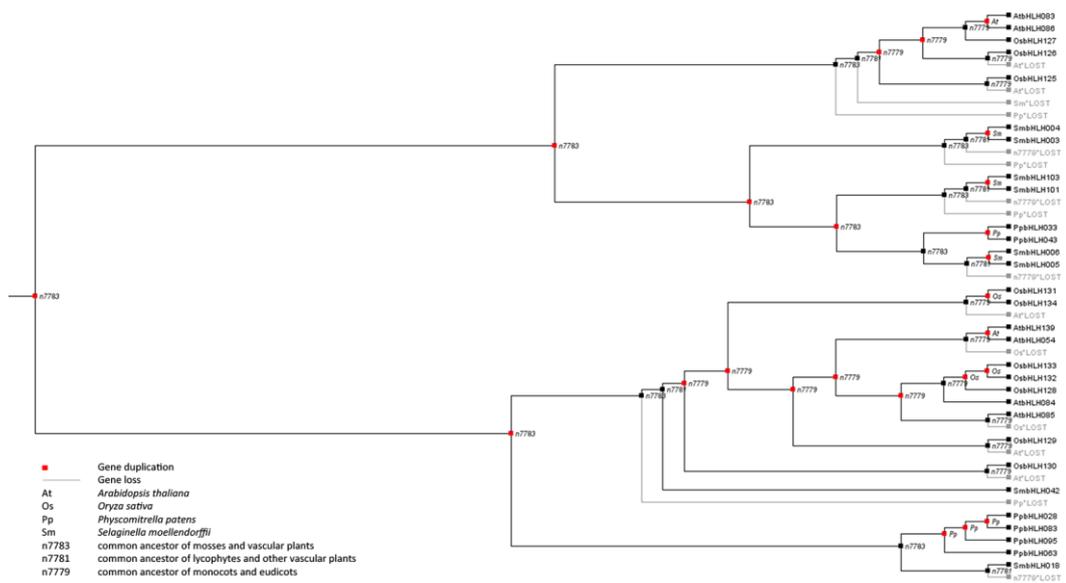
Figure 2.4 – Reconciliation tree of subfamilies VIIIc(1) and VIIIc(2) showing multiple gene duplication and gene loss events.

The reconciliation tree was calculated using the ML gene tree shown in Supplementary File 4 and the species tree shown in Fig. 2.1, although only the branches containing subfamilies VIIIc(1) and VIIIc(2) are shown here. Nodes where a gene duplication has occurred are indicated in red, whereas lost branches are indicated in grey.

The *A. thaliana* bHLH group nomenclature proposed by Heim et al. (2003) was adopted to label these subfamilies, with some modifications, e.g. Ib was divided in Ib(1) and Ib(2), and IIIa and IIIc were combined into III(a+c). Three new groups (XIII, XIV and XV) that include 28 *A. thaliana* sequences not present in the Heim et al. analysis were also defined. Of the 544 proteins analyzed, 10% do not clearly fall in any of the 26 subfamilies and were classified as 'orphans' (Supplementary File 4). These proteins often have a high degree of sequence divergence from other bHLH: this may be due to lineage specific specializations or, alternatively, they may correspond to pseudogene sequences. One of the *A. thaliana* groups proposed by Heim and colleagues (group VI, consisting of only two proteins) falls in this 'orphan' category.

Of the 26 plant bHLH subfamilies, 3 include only angiosperm proteins and 23 include angiosperm and lycophyte proteins (Fig. 2.3). Since the last common ancestor of angiosperms and lycophytes lived sometime in the late Silurian period before 415 million years ago (Kenrick and Crane 1997), this implies that these 23 bHLH subfamilies are at least 415 million years old. Interestingly, 20 of these subfamilies include not only vascular plants, but also moss proteins. Given that the oldest evidence for the existence of vascular plants is trilete spores in Upper Ordovician sediments (Steemans et al. 2009), this suggests that these subfamilies are more than 443 million years old. A bHLH protein from the chlorophyte algae *O. tauri* is a member of subfamily IVc (Fig. 2.3). This suggests that this subfamily may be over one billion years old (Heckman et al. 2001).

A clade composed of *V. carteri*, *C. reinhardtii* and *C. vulgaris* bHLH proteins is sister to the proteins in subfamily Va. However, these chlorophyte proteins were not included into the Va subfamily as this relationship is not strongly supported. Nevertheless, this relationship suggests that subfamily Va is phylogenetically closer to chlorophyte proteins than to any other land plant proteins. Another group of *V. carteri*, *C. reinhardtii* and *C. vulgaris* proteins forms a clade that is clearly distinct from other plant proteins (Fig. 2.3); this probably represents a group that evolved amongst the chlorophytes or, alternatively, was present in the common ancestors of the chlorophytes and land plants but maintained among the chlorophytes and lost in the ancestors of land plants. The only bHLH-encoding gene found in the genome of the red algae *C. merolae* could not be allocated to any chlorophyte or land plant bHLH clade.

A closer analysis of the outer branches of the different subfamilies in the ML tree shows the existence of many paralog genes, caused by multiple events of gene duplication and gene loss. To further understand the patterns of gene duplication, a

reconciliation tree was calculated for subfamilies VIIIc(1) and VIIIc(2) (Fig. 2.4). Reconciliation trees extract information from the topological incongruence between gene and species trees to infer duplications and losses in the history of a gene family (Vernot et al. 2008). Reconciliation is dependent on a strict determination of the correct phylogenetic relationships and should be interpreted with caution in this example. Nevertheless, the reconciliation tree indicates the existence of at least 10-13 independent gene duplications and 7-9 gene losses throughout the history of these two subfamilies.

In summary, the phylogenetic analysis shows that plant bHLH proteins form 26 distinct subfamilies, or evolutionary lineages; 20 of these subfamilies were already present in early land plants 443 million years ago, by which time the mosses had diverged from the vascular plants. Despite several rounds of gene duplications and losses in different plant lineages, these subfamilies have been highly conserved throughout plant evolution.

2.3.4. Plant bHLH proteins are monophyletic

The phylogenetic information contained in the 50-60 amino acids of the bHLH allows delimitation of major evolutionary lineages of proteins in plants, but does not allow good resolution of deeper nodes that represent the phylogenetic relationships between different bHLH subfamilies; these basal nodes often have low support values (Supplementary File 4) and vary when using NJ or MP analyses (data not shown). Similar poor resolution was observed in previous classifications of bHLH proteins in other groups of organisms (Atchley and Fitch 1997; Ledent and Vervoort 2001; Buck and Atchley 2003; Toledo-Ortiz et al. 2003; Li et al. 2006b). Thus, the inter-subfamily relationships shown in Fig. 2.3 should be interpreted cautiously. Non-plant bHLH sequences were initially incorporated in the ML analysis. However, the large number of proteins and the great evolutionary distances (and consequent high degree of sequence divergence) caused the non-plant proteins to form very long branches, nested within plant clades with no obvious sequence similarity (data not shown). To circumvent this problem, a phylogenetic analysis was performed on a simplified alignment (Supplementary File 3) that includes chlorophytes, red algae, diatom and yeast proteins plus representatives of the 26 plant subfamilies and of the higher-order metazoan groups (Atchley and Fitch 1997; Simionato et al. 2007).

The deep evolutionary relationships between many proteins were still not resolved: most branches in the Bayesian phylogenetic tree had low support values (Fig. 2.5). However, some close relationships between different plant bHLH subfamilies (Fig. 2.3) were supported by this analysis. For example, subfamilies IVc and Va were probably established in the common ancestors of chlorophyte algae and land plants; subfamily IVb possibly evolved later among land plants from subfamily IVc proteins. Pairs of subfamilies such as VIIIc(1)/VIIIc(2), XI/XII and III(a+c)/IIIb seem to form monophyletic lineages. Interestingly, the five diatom sequences and a sponge group A protein form a well supported clade. Closer examination of the amino acid sequence of the five diatom bHLH proteins reveals that each of these proteins have an arginine in position 8 of the bHLH domain, a defining characteristic of group A proteins (Atchley and Fitch 1997).
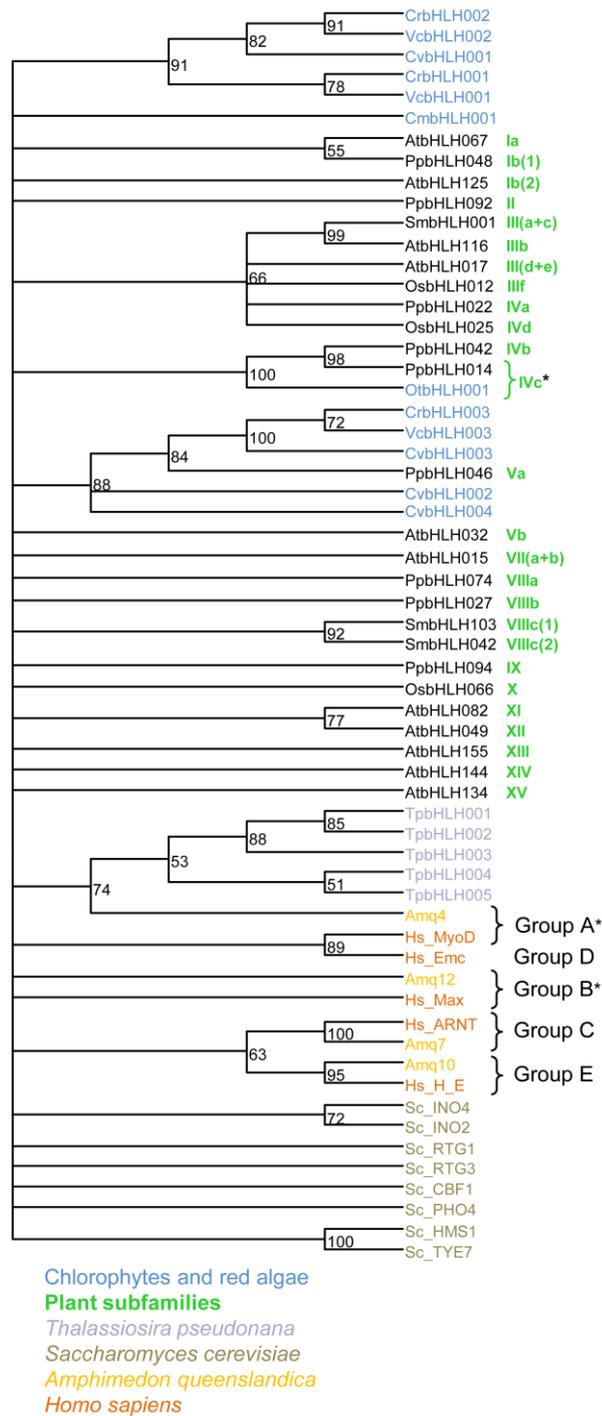
Figure 2.5 – Plant bHLH do not group with other eukaryote bHLH.

A Bayesian analysis was performed on an alignment of the bHLH sequence of one representative of each of the 26 subfamilies of plant bHLH, all the chlorophyte and red algae proteins, 5 proteins found in diatom *Thalassiosira pseudonana*, 8 *Saccharomyces cerevisiae* proteins and representatives of 5 major groups of metazoan bHLH in the sponge *Amphimedon queenslandica* and *Homo sapiens*. The tree is unrooted. The asterisks indicate possible monophyly conflicts. The numbers in the clades are posterior probability values; clades with less than 50% support were collapsed.

Although beyond the scope of this study, this suggests that group A might pre-date the origin of opisthokontes, the eukaryotic lineage that includes fungi and animals.

No clustering of plant proteins with proteins from other eukaryotic organisms is found on the Bayesian tree (Fig. 2.5). The small number of bHLH proteins found in the genomes of different chlorophytes and red algae (Fig. 2.1) suggests that the first plants had one or a few bHLH proteins, from which all modern plant bHLH descended and radiated. This view is consistent with previous analyses that highlighted the distant relationship of angiosperm and animal bHLH proteins (Ledent and Vervoort 2001; Buck and Atchley 2003; Toledo-Ortiz et al. 2003). The lack of discernible phylogenetic relationships between bHLH subfamilies in plants and other eukaryotic organisms supports the hypothesis that plant bHLH proteins are monophyletic.

2.3.5. Conserved non-bHLH motifs are present in most plant bHLH subfamilies

The amino acids sequences outside the bHLH region are generally divergent, even in closely related proteins from the same species. Nevertheless, it has been reported that short conserved amino acid motifs are often present in related plant bHLH proteins (Heim et al. 2003; Li et al. 2006b). If this plant bHLH classification were correct then it would be expected that such motifs would be conserved within subfamilies. To determine if non-bHLH motifs were conserved throughout plant evolution, a search for for amino acid patterns was conducted in the dataset of plant bHLH proteins. Twenty-eight motifs that are represented in both angiosperm and non-angiosperm proteins were found (Supplementary File 5). The relative position of each of these motifs is conserved (Fig. 2.6): most are located C-terminal to the bHLH domain, which itself is generally located towards the C-terminal half of plant proteins. Each of these motifs is only found in members of the same subfamily, apart from motif 9, which is found in both IVb and IVc proteins (Fig. 2.6). None of the 28 conserved motifs corresponds to known domains in the PFAM database. Motifs 14 and 15, present in several proteins of subfamily VII(a+b), overlap with the APB (active phytochrome binding) motif, shown to mediate the binding of several *A. thaliana* bHLH proteins to phytochrome B (Khanna et al. 2004).
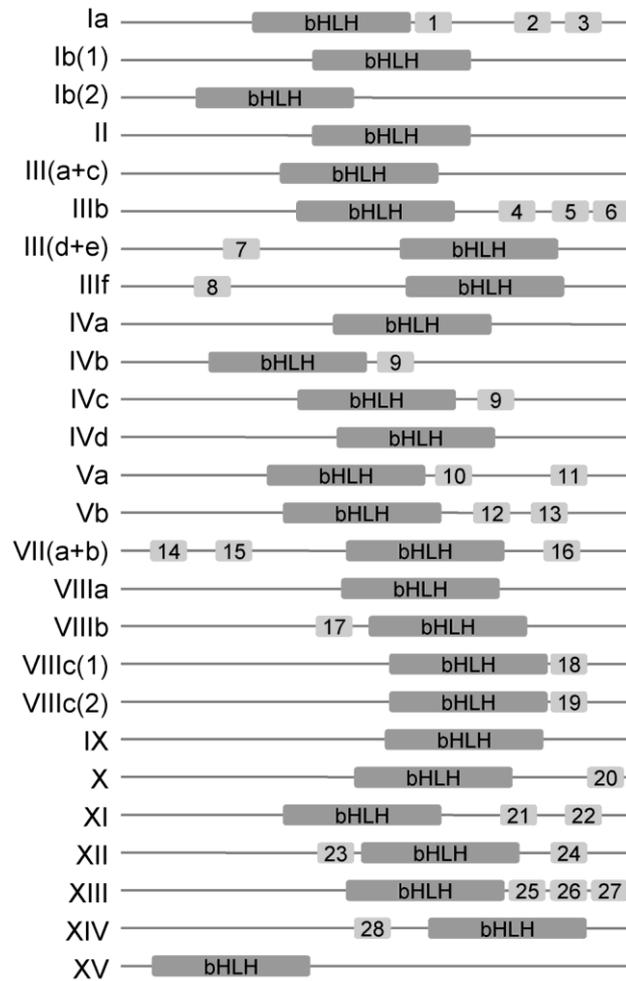
Figure 2.6 – Non-bHLH amino acid motifs are highly conserved in each bHLH subfamily.

An idealized representation of a typical member of each bHLH subfamily is shown, with the bHLH domain and other conserved motifs drawn as shaded boxes. The diagrams are not drawn to scale. The sequences of each motif in individual proteins are given in Supplementary File 5.

Motif 9 (present in IVb and IVc proteins) has a typical LZ (leucine zipper) conformation. The LZ is a dimerization domain that occurs in several regulatory proteins and consists of a periodic repetition of leucine followed by six other residues (Bornberg-Bauer et al. 1998).Several animal bHLH proteins also have a LZ immediately C-terminal to the second helix (Atchley and Fitch 1997). However, its presence in unrelated bHLH proteins suggested a multiple origin of the LZ domain in animal bHLH proteins (Atchley and Fitch 1997; Morgenstern and Atchley 1999). No similarities could be found between the bHLH sequences of IVb/IVc proteins and animal bHLH-LZ proteins. Therefore, it is likely that the acquisition of a LZ motif in bHLH proteins occurred independently in plant and animals. The occurrence of conserved domains outside the bHLH domain strongly supports the classification made on the basis of alignments of the bHLH sequence.

The PFAM database of protein domains was queried with the 544 plant bHLH proteins: significant matches to an ACT domain were found in several unrelated proteins (OsbHLH036, VcbHLH001, CrbHLH002, OsbHLH170, VcbHLH002 and PpbHLH097). The ACT is a regulatory ligand-binding domain found in a diverse group of proteins, mostly metabolic enzymes (Chipman and Shaanan 2001). The occurrence of the ACT domain in plant bHLH proteins was previously reported (Anantharaman et al. 2001) and an ACT-like domain was found to mediate homo-dimerization of the maize R protein (Feller et al. 2006). Feller et al. also found ACT-like domains in over 30 *A. thaliana* proteins using low stringency structure-based searches, but this could not be confirmed using stringent motif-based search methods. The occurrence of the ACT domain in a few proteins from different bHLH subfamilies suggests that the ACT-bHLH association occurred multiple times, possibly through domain-shuffling processes. Such mechanisms have been proposed to play an important role in the evolution of several metazoan bHLH proteins (Morgenstern and Atchley 1999; Ledent and Vervoort 2001).

The presence of highly conserved motifs among proteins of the same subfamily supports the phylogenetic relationships inferred from the bHLH domain sequence alone. The conservation of these extra domains during plant evolution suggests that they are essential for the function of the bHLH proteins in the respective subfamilies. Nevertheless, the presence of the ACT domain in a few unrelated proteins also indicates that domain-shuffling processes may have played a small role in plant bHLH evolution.

2.3.6 At least six bHLH subfamilies evolved in the ancestors of land plants

The phylogenetic analysis of plant bHLH proteins allowed the definition of the 26 major subfamilies of plant bHLH, 20 of which were already present in early land plants. It would be particularly interesting to determine if these twenty bHLH subfamilies evolved in land plants during the colonisation of land or if they were inherited from their algal ancestors. Land plants evolved from charophyte-like algae over 470 million years ago (Lewis and McCourt 2004; Gensel 2008); therefore, identifying bHLH protein sequences in charophyte species would allow the testing of these hypotheses.

Unfortunately, no genomic sequences are currently available from charophyte algae species, but a large collection of EST sequences is now being generated in the lab of Charles Delwiche, University of Maryland (personal communication). A blast search of one member of each bHLH subfamily on this EST collection identified 14 bHLH sequences from 6 charophyte species (Supplementary File 6). The incorporation of these sequences in the previously generated bHLH alignments and the generation of phylogenetic trees allowed an unequivocal classification of these bHLH proteins in different bHLH subfamilies (Fig. 2.7). This analysis reveals that proteins from the subfamilies IVb, IVc, VII(a+b), IX, XII and XIII are present in either Chlorokybales, Klebsormidiales or Zygnematales species. This indicates that these 6 bHLH subfamilies were already present in the common ancestral of Zygnematales and Choleochatales, long before the colonisation of land by plants.

	bHLH subfamily						
	IVb	IVc	VII(a+b)	IX	XII	XIII	others
Land plants	+	+	+	+	+	+	+
Nitella hyalina	-	-	-	-	-	-	-
Coleochaete orbicularis	-	+	-	-	-	-	-
Chaetospheridium globosum	-	+	-	-	-	-	-
Spirogyra pratensis	-	+	+	+	+	-	-
Penium margaritaceum	-	+	-	+	-	-	-
Klebsormidium flaccidum	-	+	-	-	-	+	-
Chlorokybus atmophyticus	+	+	-	-	-	-	-
Mesostigma viride	-	-	-	-	-	-	-
Chlorophyta	-	+	-	-	-	-	-
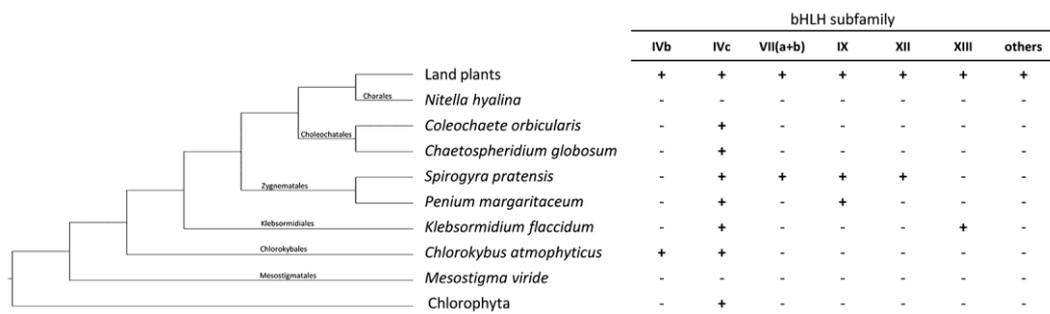
Figure 2.7 – bHLH sequences identified in different charophyte species.

The cladogram indicating the phylogenetic relationships between the difference charophyte groups is based in Karol et al. (2001). The table indicates the presence in each species of EST sequences from different bHLH subfamilies (The '+' symbol indicates the presence of one or more sequences).

2.4. Discussion

This analysis shows that most of the major subfamilies of plant bHLH transcription factors were already present in early land plants, before the separation of the moss and vascular plant lineages that occurred over 440 million years ago. In contrast, only a handful of bHLH sequences are encoded by the genomes of chlorophytes or red algae. This indicates that the large radiation of plant bHLH proteins occurred sometime between the separation of the chlorophyte / streptophyte lineages and the establishment of plants on land (Fig. 2.8).

The recent advent of large scale sequencing projects has shown that many of the gene families that control angiosperm development were present in early land plants: a recent phylogenetic study of plant homeobox proteins has come to similar conclusions: 14 classes of plant homeodomain proteins are present in both mosses and vascular plants, but fewer are present in unicellular chlorophytes or red algae (Mukherjee et al. 2009). Previously, Floyd and Bowman (2007) had shown that many of the gene families that control angiosperm development were also present in mosses and lycophytes. Similarly, Richardt et al. (2007) found that there are more transcription factor families in land plants than in unicellular algae. Nevertheless, many of these families (such as the MIKC^C MADS-box and TCP transcription factors) diversified after the divergence of lycophytes from the other vascular plants (Floyd and Bowman 2007). Two major hypotheses could explain the early radiation of the bHLH proteins in plants. The first is that the radiation occurred in parallel with the evolution of multicellularity in charophyte algae, long before the transition of plants to terrestrial environments. Charopyhtes are freshwater algae and the closest relatives to land plants (Fig. 2.8). The first charophytes were probably unicellular (Qiu 2008), but a gradual transition towards a complex multicellular body took place during charophyte evolution. The increase in the number of cell types and morphological complexity brought about by multicellularity would have been programmed by increasingly elaborate gene regulatory networks. bHLH proteins, with their ability to hetero-dimerize and differentially control gene expression, might have become an ideal tool to assemble such complex regulatory pathways. Consistent with this view is the observation that the first large radiation of the bHLH family in metazoans may have accompanied the evolution of multicellularity (Simionato et al. 2007).
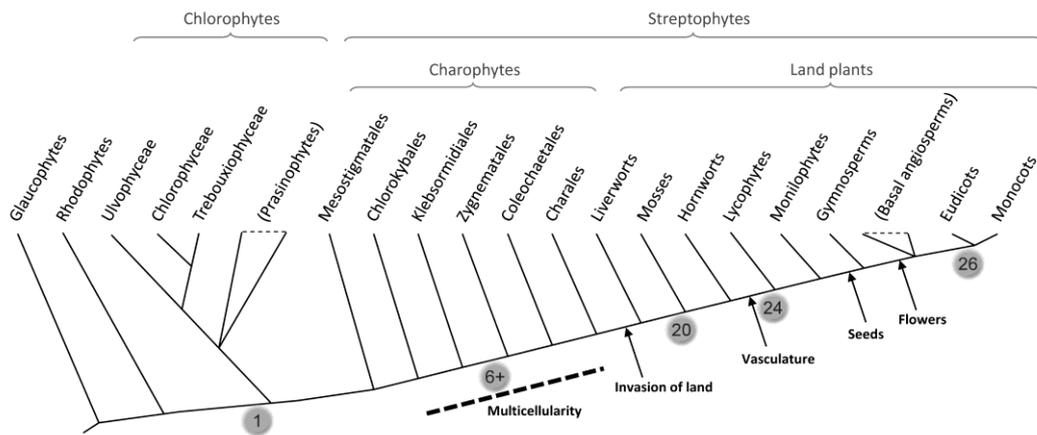
Figure 2.8 – Evolution of bHLH proteins in plants.

The simplified cladogram showing the phylogenetic relationships between the major groups of plants is based on Karol et al. (2001), Lewis and McCourt (2004) and Qiu et al. (2006). The grey balloons indicate the number of modern bHLH subfamilies predicted to be present at different nodes of plant evolution. Some of the most important events in land plant evolution are indicated.

A second hypothesis is that the diversification of plant bHLH proteins accompanied the colonization of the land. The challenges faced by plants in a dry terrestrial environment led to the evolution of many novel structures and physiological mechanisms, orchestrated by versatile gene regulatory networks. Distinguishing between these alternatives will require knowledge of the number of bHLH proteins encoded in the genomes of multicellular algae. The sequence of a charophycean (multicellular aquatic algae, sister group to land plants) genome would allow the testing of these hypotheses but unfortunately only EST sequences are currently available. However, the available ESTs are enough to show that at least six bHLH subfamilies were already present in the common ancestral of the Zygnematales and the Coleochaetales (Fig. 2.7). This supports the idea that the bHLH radiation started long time before the colonisation of land by plants.

All sequenced genomes of chlorophytes and red algae encode few bHLH proteins (Fig. 2.1). Three distinct evolutionary lineages were detected in chlorophytes (Fig. 2.3). One lineage includes both chlorophytes and land plants (subfamilies IVc and IVb), implying that it pre-dates the divergence of chlorophytes from the ancestors of land plants, over one billion years ago (Heckman et al. 2001). Interestingly, a characteristic of these two subfamilies is the presence of a LZ motif associated with the bHLH domain. This association has also occurred, independently, in animals. A second lineage of chlorophyte proteins is more similar to subfamily Va than to any other bHLH subfamily, although support for monophyly is poor. A third lineage is distinct from all other plant bHLH proteins and possibly evolved only in chlorophytes. The only bHLH protein found in red algae could not be clearly allocated to any clade. This suggests that none of the 26 subfamilies of plant bHLH proteins was established at the time of divergence of red algae from other plants, 1.5 billion years ago (Yoon et al. 2004). Alternatively, these protein lineages were lost in a *C. merolae* ancestor but are still present in other red algae; the availability of additional whole genome sequences from red algae will help to clarify this. However, the small number of bHLH found in all the chlorophytes and red algae examined (Fig. 2.1) and the lack of clear phylogenetic relationships with other eukaryotic bHLH proteins (Fig. 2.5) imply that all bHLH proteins found in plants evolved after the primary endosymbiotic event that led to the evolution of plastids and are not represented in other eukaryotic groups.

Table 2.1 – Functionally characterised bHLH proteins from different plant species.

Name	bHLH number	Function	Reference
Subfamily Ia			
AtMUTE AtFAMA AtSPCH	AtbHLH045 AtbHLH097 AtbHLH098	control sequential cell fate specification during stomatal differentiation	Nadeau, 2009; Serna, 2009
OsMUTE OsFAMA OsSPCH2	OsbHLH055 OsbHLH051 OsbHLH053	control stomata development	Liu et al. 2009
Subfamily Ib(1)			
RGE1/ZHOUPI	AtbHLH095	regulates embryonic development and endosperm breakdown	Kondou et al. 2008; Yang et al. 2008
Subfamily Ib(2)			
OsIRO2	OsbHLH056	regulates genes involved in Fe uptake under Fe-deficiency conditions	Yuko et al. 2007
Subfamily III(a+c)			
FIT	AtbHLH029	required for the up-regulation of responses to iron deficiency in *Arabidopsis* roots	Bauer et al. 2007
RERJ1	OsbHLH006	involved in the rice shoot growth inhibition caused by jasmonic acid	Kiribuchi et al. 2004
Subfamily IIIb			
ICE/SCRM ICE2/SCRM2	AtbHLH116 AtbHLH033	control stomatal development; implicated in the cold acclimation response and freezing tolerance	Chinnusamy et al. 2003; Kanaoka et al. 2008; Fursova et al. 2009
TaICE41 TaICE87	Wheat[a]	potential activators of the cold-responsive genes	Badawi et al. 2008
Subfamily III(d+e)			
MYC2/ JAI1/JIN1	AtbHLH006	involved in abscisic acid, jasmonic acid and light signalling pathways	Abe et al. 2003; Lorenzo et al. 2004; Yadav et al. 2005
AIB	AtbHLH017	involved in abscisic acid signalling	Li et al. 2007
PsGBF	Pea[a]	regulates phenylpropanoid biosynthetic pathway	Qian et al. 2007
Subfamily IIIf			
TT8 GL3 EGL3	AtbHLH042 AtbHLH001 AtbHLH002	partially redundantly regulate anthocyanin biosynthesis, trichome and root hair development	Nesi et al. 2000; Payne et al. 2000; Bernhardt et al. 2003; Zhang et al. 2003
Ra/OSB1 Rb Rc OSB2 Lc IN1 An1	OsbHLH013 OsbHLH165 OsbHLH017 OsbHLH016 Maize[a] Maize[a] Petunia[a]	regulate the anthocyanin biosynthetic pathway	Ludwig et al. 1989; Burr et al. 1996; Hu et al. 2000; Spelt et al. 2000; Sakamoto et al. 2001; Sweeney et al. 2006
Subfamily IVa			
NAI1	AtbHLH020	required for the formation of an endoplasmic reticulum-derived structure, the ER body	Matsushima et al. 2004
Subfamily IVc			
ILR3	AtbHLH105	modulate metal homeostasis and auxin-conjugate metabolism	Rampey et al. 2006
Subfamily Va			

BIM1	AtbHLH046		
BIM2	AtbHLH102	implicated in brassinosteroid signalling	Yin et al. 2005
BIM3	AtbHLH141		
Subfamily VII(a+b)			
PIF1/PIL5	AtbHLH015	bind to activated phytochromes and mediate light and gibberellin signalling responses; PIF4 was recently shown to also mediate plant architecture responses to high temperatures	Castillon et al. 2007; de Lucas et al. 2008; Leivar et al. 2008; Koini et al. 2009
PIF3	AtbHLH008		
PIF4	AtbHLH009		
PIF5/PIL6	AtbHLH065		
PIF7	AtbHLH072		
HFR1	AtbHLH026	mediate both phytochrome and cryptochrome signalling	Duek and Fankhauser, 2003
SPATULA	AtbHLH024	regulator of carpel margin development; mediator of germination responses to light and temperature	Heisler et al. 2001; Penfield et al. 2005
ALCATRAZ	AtbHLH073	required for the formation of a cell layer necessary for fruit dehiscence	Rajani and Sundaresan, 2001
UNE10	AtbHLH016	involved in the fertilization process	Pagnussat et al. 2005
BP-5	OsbHLH102	involved in the regulation of amylose synthesis in the rice endosperm	Zhu et al. 2003
Subfamily VIIIb			
HEC1	AtbHLH088	redundantly control the development of the transmitting tract and stigma; each of these proteins can form hetero-dimers with SPATULA	Gremski et al. 2007
HEC2	AtbHLH037		
HEC3	AtbHLH043		
LAX	OsbHLH123	regulator of axillary meristem generation in rice	Komatsu et al. 2003
INDEHISCENT	AtbHLH040	required for the differentiation, in the *Arabidopsis* fruit, of three cell types involved in seed dispersal	Liljegren et al. 2004
Subfamily VIIIc(1)			
AtRHD6	AtbHLH083	required for the formation of root hairs	Menand et al. 2007
AtRSL1	AtbHLH086		
PpRSL1	PpbHLH043	redundantly required for the development of rhizoids and caulonemata	Menand et al. 2007
PpRSL2	PpbHLH033		
Subfamily VIIIc(2)			
AtRSL2	AtbHLH085	partially redundant and involved in root hair development	Yi, 2008
AtRSL3	AtbHLH084		
AtRSL4	AtbHLH054		
AtRSL5	AtbHLH139		
Subfamily XI			
UNE12	AtbHLH059	involved in the fertilization process	Pagnussat et al. 2005
PTF1	OsbHLH096	involved in the responses to phosphate deficiency stress	Yi et al. 2005
Subfamily XII			
ZCW32/BPE	AtbHLH031	controls petal size	Szecsi et al. 2006
BEE1	AtbHLH044	redundant positive regulators of brassinosteroid signalling	Friedrichsen et al. 2002
BEE2	AtbHLH058		
BEE3	AtbHLH050		
CIB1	AtbHLH063	shown to interact with the blue-light receptor CRY2 and promote floral initiation	Liu et al. 2008
CIB5	AtbHLH076		
Subfamily XIII			
LHW	AtbHLH156	regulates the size of the vascular initial population in the root meristem	Ohashi-Ito and Bergmann, 2007
Subfamily XIV			

SAC51	AtbHLH142	involved in a spermidine synthase mediated stem elongation process	Imai et al. 2006

Subfamily XV

PRE1	AtbHLH136		
PRE2	AtbHLH134		
PRE3	AtbHLH135	proposed to act as positive regulators of gibberellin signalling	Lee et al. 2006
PRE4	AtbHLH161		
PRE5	At3g28857[a]		
PRE6	At1g26945[a]		
KIDARI	At1g26945[a]	represses light signal transduction; interacts and negatively regulates HFR1	Hyun and Lee, 2006

Orphans

AMS	AtbHLH021		Sorensen et al. 2003
DYT1	AtbHLH022	required for correct anther development, particularly tapetum development	Zhang et al. 2006
TDR	OsbHLH005		Li et al. 2006a
Udt1	OsbHLH164		Jung et al. 2005
MEE8	AtbHLH108	required for early embryo development	Pagnussat et al. 2005
Fer	Tomato[a]	controls iron-uptake responses in roots	Ling et al. 2002
Gmyc1	Gerbera[a]	regulates the expression of an anthocyanin pathway enzyme	Elooma et al. 1998
delila	Antirrhinum[a]	regulates the pattern of anthocyanin pigmentation	Goodrich et al. 1992
JAF13	Petunia[a]	regulates the anthocyanin biosynthetic pathway	Quattrocchio et al. 1998
PAR1	At2g42870[a]	negatively control growth and metabolic shade avoidance responses	Roig-Villanova et al. 2007
PAR2	At3g58850[a]		

[a] These proteins were not included in the phylogenetic analysis; their classification was based on pHMM scores to subfamily-specific pHMMs.

Plant transcription factor families usually have high expansion rates compared to metazoan families, caused by elevated rates of retention of duplicated genes (Shiu et al. 2005). Accordingly, there are usually many (1-12) proteins per species in each of the 26 plant bHLH subfamilies (Supplementary File 4), in contrast with the small number (1-4) of genes found in each of the 44 metazoan subfamilies (Ledent and Vervoort 2001; Simionato et al. 2007). Members of the same plant bHLH subfamily are frequently involved in the same biological process (Table 2.1). Usually the functions of these proteins overlap, causing them to be partially or totally redundant (e.g. HEC or BEE proteins). A striking exception comes from three *A. thaliana* subfamily Ia proteins, MUTE, SPEECHLESS and FAMA: they play non-overlapping roles in controlling sequential cell fate specification during stomatal differentiation, in a pathway surprisingly similar to metazoan bHLH proteins controlling muscle and neural development (Nadeau 2009; Serna 2009). Interestingly, the function of these proteins seems to be mostly conserved in rice and maize homologs, despite these species having considerably different stomata morphology and differentiation patterns (Liu et al. 2009). Other examples of members of the same bHLH subfamily regulating similar processes in different species are currently known (Table 2.1). A new challenge will be to understand how the function of bHLH proteins has changed during plant evolution. An interesting glimpse comes from subfamily VIIIc(1), where the *P. patens* proteins PpRSL1 and PpRSL2 – the only moss bHLH proteins that have been characterized so far – were shown to be required for the development of rhizoids (Menand et al. 2007). Rhizoids were lost during vascular plant evolution, but the two representatives of subfamily VIIIc(1) in *A. thaliana* (AtRHD6 and AtRSL1) are required for the formation of root hairs, analogous structures to rhizoids with a similar rooting function (Menand et al. 2007). This suggests that these proteins were independently recruited to fulfil similar functions during land plant evolution.

The presence of highly conserved motifs (such as the APB motif in PIF proteins) in the different plant bHLH subfamilies (Fig. 2.6) indicates that the partners of molecular interactions are also conserved. This is particularly exciting, because it suggests that protein interactions that are at the base of gene regulatory networks have been conserved in land plants for over 400 million years. Several plant bHLH proteins are known to form transcription complexes with MYB proteins (Ramsay and Glover 2005). Although the early evolution of MYB proteins in plants has not been characterized, over 30 MYB sequences were found in the genome of *C. reinhardtii* and more than 150

sequences in *P. patens* (data not shown). Given the large number of both bHLH and MYB proteins in mosses, it is appealing to hypothesize that the bHLH-MYB complex had evolved early in land plant evolution.

The body plan of plants has dramatically changed since the colonization of land, with the increase in complexity of the sporophyte generation and the invention of innovative structures such as vasculature, leaves, seeds and flowers. Since a similar set of developmental regulators have been continually used throughout land plant history, it can be concluded that the evolution of land plants occurred largely through the reusing and recycling of very ancient gene regulatory networks.

Today's availability of whole genome sequences from different plant species opens many doors for understanding the evolution of gene regulatory networks in land plants. The picture that is starting to emerge is that much of the complex regulatory machinery that is currently being dissected in 'higher' plants was actually invented by very 'simple' ones, early in land plant evolution. The recent reappraisal of algae, bryophytes and lycophytes as experimental organisms will be an excellent tool to clarify the molecular and biological foundations of many of these processes.

Chapter 3

Evolution of *RSL* class II genes

3.1. Introduction

Root hairs are tubular projections from root epidermal cells that increase the area available for nutrient absorption and help roots to penetrate into the substrate. In *Arabidopsis*, the initiation of root hair development is controlled by a bHLH transcription factor called AtRHD6: in *Atrhd6* mutant plants very few root hairs develop, and these often have polarity defects (Masucci and Schiefelbein 1994; Menand et al. 2007b). The *Arabidopsis* genome also encodes a duplicated gene (*AtRSL1*), which is redundantly required for root hair development with *AtRHD6*: root hair development in *Atrhd6 Atrsl1* double mutant plants is completely abolished (Menand et al. 2007b). AtRHD6 and AtRSL1 are the only two *Arabidopsis* members of the bHLH subfamily VIIIc(1) and will hereafter be denominated *RSL* class I genes.

The mechanism of root hair initiation controlled by RSL class I proteins is a key step in root hair development. The *Atrhd6* mutation is epistatic to several mutations affecting root hair development (Parker et al. 2000), indicating that AtRHD6 acts upstream of major events that mediate root hair development. Accordingly, *AtRHD6* and *AtRSL1* are expressed in root hair cells between the root meristem and elongation zone, and the expression disappears before the emergence of root hairs. Furthermore, the expression of *AtRHD6* and *AtRSL1* is controlled by the epidermal patterning genes *GL2*, *TTG*, *WER* and *CPC* (Menand et al. 2007b). This indicates that *RSL* class I genes form a link between the epidermal patterning mechanism and the formation of root hairs.

RSL class I proteins positively regulate the transcription of all four members of their sister bHLH subfamily: the VIIIc(2) bHLH subfamily (Yi 2008). This subfamily (composed of *AtRSL2*, *AtRSL3*, *AtRSL4* and *AtRSL5*) will hereafter be denominated as *RSL* class II. In WT plants, *RSL* class II genes are expressed in hair cells around the initiation zone and during root hair elongation. In *Atrhd6 Atrsl1* double mutant plants, however, the expression level of these genes is highly decreased. Analyses of single, double and triple mutants and of plants overexpressing these genes showed that all *RSL* class II proteins are involved in root hair tip growth (Yi 2008; Yi et al. 2010). *Atrsl2* and *Atrsl4*, in particular, show clear defects in root hair growth, and an *Atrsl2 Atrsl4* double mutant is completely hairless (although it can form very small bulges, indicating that root hair initiation is not defective).
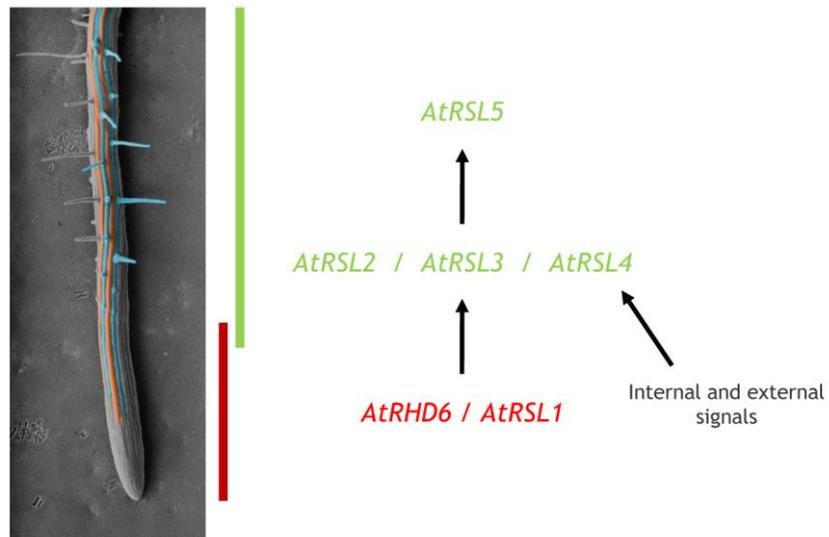
Figure 3.1 – Model of the *RSL* network that regulates root hair development in *Arabidopsis*.

The left picture is an SEM picture of the *Arabidopsis* primary root, with root hair cells marked in blue and non-hair cells marked in orange. The red and green bar indicates the region where *RSL* class I and *RSL* class II genes are expressed, respectively. The overlap between the two bars indicates an overlapping in expression. The diagram on the right indicates the genetic relationships between the different genes. Adapted from Yi (2008).

Auxin and ethylene, which are important regulators of root hair development, can bypass *RSL* class I genes and regulate root hair initiation by positively regulating *AtRSL4* (Yi et al. 2010). Auxin and ethylene can also modulate the expression of the other RSL class II genes, suggesting the existence of a transcriptional network with positive and negative feedback mechanisms (Yi 2008).

These observations suggested the existence of a multi-level transcriptional regulatory cascade that controls the development of root hairs in *Arabidopsis* (Yi 2008; Fig. 3.1). RSL class I proteins, whose expression is controlled by epidermal patterning genes, control the initiation of root hair development and positively regulate the expression of RSL class II genes: *AtRSL4* is directly regulated by AtRHD6, *AtRSL2* and *AtRSL3* are indirectly regulated by AtRHD6, and *AtRSL5* is regulated by the other RSL class II proteins. Although AtRSL4 is also involved in root hair initiation, the overarching function of RSL class II proteins is to modulate root hair tip growth. Endogenous and exogenous factors, (auxin, ethylene and phosphate stress) modulate this pathway primarily by regulating the transcription of *RSL* class II genes.

When and how has the RSL network that control the development of root hairs appeared during plant evolution? The moss *Physcomitrella patens* has *RSL* class I and *RSL* class II genes indicating that these bHLH subfamilies were present in early land plants (see Chapter 2). A functional analysis of the *Physcomitrella RSL* class I genes showed that they are redundant and required for the development of caulonema and rhizoid filaments (Menand et al. 2007b). Root hairs, rhizoids and caulonemata are structures that fulfil similar rooting functions. There are, however, fundamental differences between these cell types: root hairs are tubular projections from epidermal cells of the root in the sporophytic life cycle stage, whereas rhizoids and caulonemata are filamentous structures that grow in the gametophytic life cycle stage. This suggests that *RSL* class I genes controlled the development of rooting filaments in the gametophytes of early land plants and that, during the explosive radiation of vascular plants during the Devonian Period, *RSL* class I genes were recruited to the sporophyte generation, where they control the development of rhizoids and root hairs.

In summary, RSL class I and RSL class II proteins are two ancient subfamilies of bHLH transcriptions factors. In the moss *Physcomitrella patens*, RSL class I proteins regulate the development of caulonema and rhizoid cells. In *Arabidopsis thaliana*, *RSL* class I genes form a transcriptional network with *RSL* class II genes that regulate the development of root hairs. Was the RSL class I and class II network that controls root

hair formation in *Arabidopsis* also present in the early land plants? If so, it would be expected that RSL class I and class II proteins would form a mechanism that controls caulonema and rhizoid development in mosses. The aim of the work described in this chapter is to test this hypothesis, define the function of RSL class II proteins in *Physcomitrella* and understand the evolution of the RSL mechanisms of control of rooting cells development in land plants.

3.2. Methods

3.2.1. Sequence retrieval and phylogenetic analyses

An initial alignment of conserved amino acid sequences from RSL class I and RSL class II proteins (bHLH subfamilies VIIIc(1) and VIIIc(2), respectively) was used to build an RSL specific pHMM (Supplementary File 7) with the program hmmbuild (Eddy 1998). This pHMM was used to identify *RSL* coding sequences with the program hmmsearch (Eddy 1998) in the gene model databases of *Brachypodium distachyon* (The International Brachypodium Initiative 2010), *Mimulus guttatus* v1.0, *Physcomitrella patens* (Rensing et al. 2008), *Populus trichocarpa* v1.1 (Tuskan et al. 2006), *Sorghum bicolor* (Paterson et al. 2009) and *Selaginella moellendorffii* v1.0 from the DOE Joint Genome Institute (http://www.jgi.doe.gov/); *Arabidopsis lyrata*, *Cucumis sativus*, *Glycine max* (Schmutz et al. 2010) and *Manihot esculenta* v1.1 from Phytozome 5.0 (http://www.phytozome.net/); *Arabidopsis thaliana* from The Arabidopsis Information Resource (http://www.arabidopsis.org/); *Oryza sativa* 6.1 from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/) and *Zea mays* from the Maize Genome Sequencing Project (http://www.maizesequence.org/). BLAST searches were made against the EST database of GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

The amino acid sequences were manually aligned and conserved domains used for phylogenetic analyses. ML analyses were done with the program PhyML version 3.0.1 (Guindon and Gascuel 2003) using the JTT model of amino acid substitution, an estimated gamma distribution parameter and a SH-like aLRT. The Bayesian analysis was performed with MrBayes version 3.1.2 (http://mrbayes.csit.fsu.edu/): 2 independent runs were computed until the standard deviation of split frequencies was less than 0.01; one tree was saved every 1000 generations and 75% of the trees from each run were summarized to give rise to the final tree. Trees were visualised using the program Figtree (http://tree.bio.ed.ac.uk/software/figtree/). A logo of the alignment of amino acid sequences was created with WebLogo (http://weblogo.berkeley.edu/). The similarities between amino acid alignments were calculated using a BLOSUM62 matrix.

3.2.2. Plant growth conditions

The Gransden wild type strain of *Physcomitrella patens* (Hedw.) Bruch and Schimp (Ashton and Cove 1977) was used in this study. Moss sporophytes were surface sterilised in 5% (v/v) sodium hypochlorite for 10 minutes, rinsed five times with sterile distilled water and crushed to release the spores into the water. Sterile spore suspensions were kept at 4°C in darkness for several months. The minimal medium used for spore germination and phenotypical analysis (based in the recipe described in Ashton et al. (1979)) contained 800mg $Ca(NO_3)_2$, 250mg $MgSO_4.7H_2O$, 12.5mg $FeSO_4.7H_2O$, 1ml KH_2PO_4 buffer pH 7, 1ml trace element solution and 8g agar (Formedium cat#AGA03) per litre. The KH_2PO_4/KOH buffer contained 25g KH_2PO_4 per 100ml; pH 7.0 was obtained by titrating with 4M KOH. Trace element solution contained 55mg $CuSO_4.5H_2O$, 55mg $ZnSO_4.7H_2O$, 614mg H_3BO_3, 389mg $MnCl_2.4H_2O$, 55mg $CoCl_2.6H_2O$, 28mg KI, 25mg $Na_2MoO_4.2H_2O$ per litre. 35ml media was poured in 90mm plastic plates and overlaid with autoclaved cellophane disks (AA packaging, UK). 1ml of sterile spore suspension was inoculated onto each plate; plates were closed with Micropore tape. Plants were grown at 25°C in plant growth cabinets (Sanyo MLR-351), illuminated with a light regime of 16h light / 8h darkness. For the production of chloronema-rich tissue for DNA extractions and protoplast isolation, minimal medium was supplemented with 500mg ammonium tartrate and 5g glucose; protonema tissue was blended with a homogenizer (PowerGen 500, Fisher Scientific). Chloronema cultures were subcultured every 5-7 days. For the NAA treatments, spores were grown in minimal medium overlaid with cellophane disks for two weeks. The cellophane disks with protonema colonies were then transferred to fresh minimal media supplemented with NAA and incubated for a further week.

Arabidopsis thaliana seeds were surface sterilised in 5% (v/v) sodium hypochlorite for 10 minutes and rinsed four times with sterile distilled water. Seeds were then stratified at 4°C in darkness for three days and sown on Murashige and Skoog medium (pH 5.8) supplemented with 1% (w/v) sucrose and 0.5% (w/v) phytagel. The plants were grown vertically ay 25°C and illuminated with a light regime of 16h light / 8h darkness. The roots were observed 4-6 days after sowing.

3.2.3. Constructs for plant transformation

For the PpRSL4-ko construct, a 734bp fragment upstream and a 987bp fragment downstream of the *PpRSL4* coding sequence were amplified from genomic DNA by PCR and cloned into the XbaI-XhoI and SpeI-MluI sites of pBNRF (confers resistance to G418)(Menand et al. 2007b), respectively, yielding a 6515bp plasmid. For the PpRSL5-ko construct, a 698bp fragment upstream and a 710bp fragment downstream of the *PpRSL5* coding sequence were amplified from genomic DNA by PCR and cloned into the SalI-BamHI and SpeI-MluI sites of pBHSNR (confers resistance to hygromycin) (Menand et al. 2007b), respectively, yielding a 6567bp plasmid. For the PpRSL6-ko construct, a 712bp fragment upstream and a 602bp fragment downstream of the *PpRSL6* coding sequence were amplified from genomic DNA by PCR and cloned into the SalI-HindIII and MluI-NsiI sites of pBZRF (confers resistance to zeocin) (gift from Fabien Nogue, Versailles), respectively, yielding a 6393bp plasmid. *Pprsl3* knockout plants had been previously generated by Benoît Menand (John Innes Centre, UK), using a pBHRF-based plasmid that confers resistance to hygromycin.

For the overexpression of genes in *Arabidopsis*, the coding sequences of *PpRSL3*, *PpRSL4*, *PpRSL5* and *PpRSL6* were amplified by RT-PCR (from RNA extracted from 25 days old plants grown in minimal medium) and subcloned in pGEM®-T Easy (Promega cat#A1360), generating pPpRSL3, pPpRSL4, pPpRSL5 and pPpRSL6, respectively. A SmaI-XhoI pPpRSL3 fragment was cloned into the EcoICRI-SalI sites of a modified pCambia1300 plasmid (Yi et al. 2010), generating the binary vector p35S::PpRSL3. A KpnI-SalI pPpRSL4 fragment, a BamHI-SalI PpRSL5 fragment and a KpnI-PstI pPpRSL6 fragment were cloned into the respective sites in the modified pCambia1300, generating the binary vectors p35S::PpRSL4, p35S::PpRSL5 and p35S::PpRSL6, respectively.

For the overexpression of genes in *Physcomitrella*, a 2973bp fragment from pGWB2 (Tsuyoshi Nakagawa, Japan), carrying a 35S promoter-attR1-CmR-ccdB-attR2 cassette, was amplified by PCR and cloned into the NotI-SpeI sites of p108-lox-35Snpt-lox, a plasmid based in pBNRF that carries the '108 locus' of *Physcomitrella* (gift from Prof. Pierre Goloubinoff, Lausanne, Switzerland), generating the moss transformation vector p108GW35S (9976bp). The coding sequences of *PpRSL3*, *PpRSL4*, *PpRSL5* and *PpRSL6* were amplified from the pPpRSL(3-6) plasmids and subcloned into pCR8/GW/TOPO-TA (Invitrogen cat#K250020). The resulting clones were cloned by LR reaction using the Gateway LR Clonase II enzyme mix (Invitrogen cat#11791-020) into

p108GW35S, generating the moss transformation constructs p108oxRSL3, p108oxRSL4, p108oxRSL5 and p108oxRSL6.

The two artificial microRNA precursors were generated according to the protocol described in http://wmd2.weigelworld.org/themes/amiRNA/pics/Cloning_of_artificial_ microRNAs.pdf, using the plasmid pRS300 as a template. The PCR products with the two amiRNA precursor fragments were subcloned into pCR8/GW/TOPO-TA (Invitrogen cat#K2500-20), and then cloned by LR reaction (Invitrogen cat#11791-020) into p108GW35S, generating the moss transformation constructs pAmiRNAa and pAmiRNAb.

3.2.4. Plant transformation

All the binary vectors were transferred into the *Agrobacterium* strain GV3101. *Arabidopsis* plants were transformed using a modified version of the floral dip method (Clough and Bent 1998): *Agrobacterium* transformed with a binary vector were grown in 50 ml liquid LB medium with kanamycin (50 µg/ml) and incubated overnight at 28°C; 40 ml of this culture was used to inoculate 500 ml of fresh LB media (with kanamycin) and incubated overnight at 28°C; cells were pelleted by centrifugation and resuspended in 200 ml 5% (w/v) sucrose supplemented with 0.05% (v/v) Silwet L-77 (Vac-in-Stuff, Lehle Seeds); *Arabidopsis* plants were dipped into the bacterial suspension for 10-20 seconds, laid on their sides and covered with a dome for 24 hours to maintain humidity. Plants were returned to normal greenhouse conditions for seed harvesting. Transformants were selected on 25 µg.ml^{-1} hygromycin. The expression of the construct was confirmed by RT-PCR, pictures were taken from T2 plants and the phenotypes were confirmed in the T3 generation.

Physcomitrella was transformed by PEG-mediated direct DNA transfer into protoplasts, as described in Schaefer and Zrÿd (1997). Selection for antibiotic-resistant colonies was done using 50 µg.ml^{-1} G418 disulfate, 25 µg.ml^{-1} hygromycin B or 50 µg.ml^{-1} zeocin (Invitrogen cat#R250-01). Stable transformants were confirmed by PCR and Southern blot.

3.2.5. DNA extractions and Southern blots

For large scale DNA extraction from *Physcomitrella*, the Nucleon Phytopure Genomic DNA Extraction Kit (GE Healthcare cat#RPN8511) was used; three to four 6

days old chloronema cultures were used for each extraction. For small scale DNA extraction, a modified CTAB protocol was used (http://biology4.wustl.edu/moss /DNA.pdf). DNA extraction from *Arabidopsis* was performed as described previously (Edwards et al. 1991).

For Southern blots analysis, 1µg genomic DNA was digested overnight with 100U of the appropriate restriction enzyme. After electrophoresis, the DNA was transferred to a positively charged nylon membrane (Roche cat#11666657001). Hybridization and detection were performed as described in the Roche DIG Application Manual using DIG Easy Hyb Granules (Roche cat#11796895001), DIG Luminescent Detection Kit (Roche cat#11363514910) and the Lumi-Film Chemiluminescent Detection Film (Roche cat#11666657001). DIG-labeled hybridization probes for detection of the selection cassette were prepared by PCR labelling using the PCR DIG Probe Synthesis Kit (Roche cat#11636090910).

3.2.6. RNA extraction, cDNA synthesis and qRT-PCR analysis

Total RNA was isolated from frozen plant tissue with the RNeasy Plant Mini Kit (Qiagen cat#74904), with a DNase (Qiagen cat#79254) treatment step on the RNeasy spin column. Total RNA from moss sporophytes was first isolated using Trizol (Invitrogen cat#15596-018) and then treated with DNase on an RNeasy spin column. RNA was quantified on a NanoDrop ND1000 spectrophotometer (Thermo Fisher Scientific, USA). For cDNA synthesis, RNA was reverse transcribed with the SuperScript III First-Strand Synthesis System for RT-PCR using oligo(dT) (Invitrogen cat#18080-051). qRT-PCR analysis was performed with the SYBR Green PCR Master Mix (Applied Biosystems cat#4364344) in the Applied Biosystems 7300 Real-Time PCR System. Cycle conditions were as follows: 10 minutes incubation at 95°C followed by 40 cycles of 15 seconds incubation at 95°C and 1 minute at 60°C; a data collection step was performed at the end of each cycle. A dissociation stage was performed at the end of the run to confirm the amplification of specific amplicons. Relative expression levels were calculated using the $\Delta\Delta$Ct method, using the GAPDH (PHYPADRAFT_226280) and Elongation Factor 1α (PHYPADRAFT_158916) transcripts for normalization.

3.2.7. Microscopy and statistical analysis

Protonema, gametophores and roots were imaged with a Leica DFC310 FX camera mounted on a Leica M165 FC stereo microscope. For the visualization of individual protonema cells, protonema filaments were dissected from 21 days old protonema colonies (growing on minimal media overlaid with cellophane disks) and mounted on a 100 µg.ml^{-1} Calcofluor White solution (Sigma cat#F3543); epifluorescent imaging was performed with a Retiga EXi CCD camera (Qimaging) mounted on a Olympus BX50 microscope, with an excitation filter of 330-385nm and a barrier filter of 420-385nm. Images were post-processed with ImageJ (http://rsb.info.nih.gov/ij/) and Adobe Photoshop.

All measurements were performed in ImageJ; protonema colony diameter was determined as the mean of the end-to-end distance in the X-axis and in the Y-axis. Microsoft Excel and GraphPad Prism were used for statistical analysis; multiple comparison tests were calculated by ANOVA using Dunnett's post-test, as implemented in GraphPad Prism.

3.3. Results

3.3.1. *RSL* class I and class II genes are conserved in land plants

RSL class I and RSL class II genes are present in Physcomitrella

In the previous chapter it was shown that *RSL* class I (bHLH subfamily VIIIc(1)) and *RSL* class II (bHLH subfamily VIIIc(2)) genes are present in *Physcomitrella*: searches using pHMM on the predicted gene models of *P. patens* found two *RSL* class I genes (*PpRSL1/PpbHLH043* and *PpRSL2/PpbHLH033*) and four *RSL* class II genes (*PpRSL3/PpbHLH028, PpRSL4/PpbHLH083, PpRSL6/PpbHLH095* and *PpRSL7/PpbHLH063*). However, when the raw *P. patens* genomic sequences were searched using tBLASTn, a fifth *RSL* class II gene was found[4] (*PpRSL5*) (Appendix 2).

To confirm the expression and predicted coding sequences of the *Physcomitrella RSL* class II genes, RNA was extracted from protonema tissue. *PpRSL3*, *PpRSL4*, *PpRSL5* and *PpRSL6* expression was detected by RT-PCR, but the putative *PpRSL7* transcript could not be detected. The full predicted coding sequence plus the 5' and 3' UTR of the *PpRSL3*, *PpRSL4*, *PpRSL5* and *PpRSL6* transcripts were amplified by RT-PCR and sequenced. The correct start codon of the proteins was identified as the first ATG codon after a STOP codon (in-frame with the bHLH coding sequence). Likewise, the correct STOP codon was identified as the first in-frame STOP codon after the bHLH sequence. This way an incorrect splicing junction of the gene model of *PpbHLH083* (*PpRSL4*) was found and corrected. The aligned full coding sequences of PpRSL3-6 are represented in Appendix 3. The coding sequence of *PpRSL1* and *PpRSL2* had been determined previously (Menand et al. 2007b).

All *Arabidopsis* and *Physcomitrella RSL* class I and class II genes have two conserved intron positions in the bHLH coding region (Fig. 3.2B). A third intron located in the bHLH region is present in roughly half of the *RSL* genes, but this does not correlate with the phylogenetic position of the respective genes. Apart from 1) this intron, 2) a

---

[4] This gene was not added to the bHLH sequence dataset used in Chapter 2: that was done based only in published gene models.

fourth intron located C-terminal to the bHLH region in *PpRSL3* and *PpRSL4* and 3) a truncation in the N-terminus of *PpRSL2*, the intron-exon structure of the *Arabidopsis* and *Physcomitrella RSL* genes is well conserved and quite characteristic. In contrast, *AtIND*, a member of the VIIIb bHLH subfamily (which is the sister clade to the *RSL* class I + *RSL* class II clade), has no introns.

RSL genes are conserved across land plants

The extended alignment of the RSL amino acid sequences (bHLH + RSL conserved domains; Fig. 3.2A) was used to build a pHMM specific to the *RSL* genes (Supplementary File 7). This pHMM was then used to retrieve *RSL* genes from genomic databases of species for which a whole genome sequence is available. In addition, tBLASTn was used to identify *RSL* genes from EST sequences. A total of 92 RSL sequences were identified this way (Appendix 2; Supplementary File 8). These sequences were then sorted according to the characteristic conserved RSL domain C-terminal to the bHLH into 34 RSL class I and 58 RSL class II proteins. Different phylogenetic analyses confirmed the classification of each protein as RSL class I or II (data not shown).

The genomes of these different species have a similar number of *RSL* genes: usually 1-3 *RSL* class I genes and 4-7 *RSL* class II genes per genome. The conservation of the number of *RSL* genes, despite the occurrence of several gene duplications and losses, suggests that there is an evolutionary constraint to maintain a constant number of *RSL* genes per genome. Interestingly, the number of *RSL* class II genes is usually twice the number of *RSL* class I in any given species. It is tempting to speculate that this ratio is important for the biological function of these proteins.

The conserved amino acid regions of the RSL proteins (in the C-terminus) were manually aligned (Appendix 4 and 5; Supplementary Files 9 and 10). A feature of RSL class I and RSL class II proteins (also shared with the bHLH subfamilies VIIIa and VIIIc) is the presence of the conserved amino acid motif Q_5-A_9-R_{13} in the DNA-binding region of the bHLH domain. Most bHLH proteins have a H_5-E_9-R_{13} motif instead, which has been shown to be crucial for the recognition and binding of E-box hexanucleotide sequences (see Chapter 2). This raises three alternative scenarios: 1) Q_5-A_9-R_{13} containing proteins bind a novel DNA sequence; 2) act as negative regulators of DNA transcription; 3) or modulate the DNA-binding activity of other proteins, but do not bind DNA themselves.
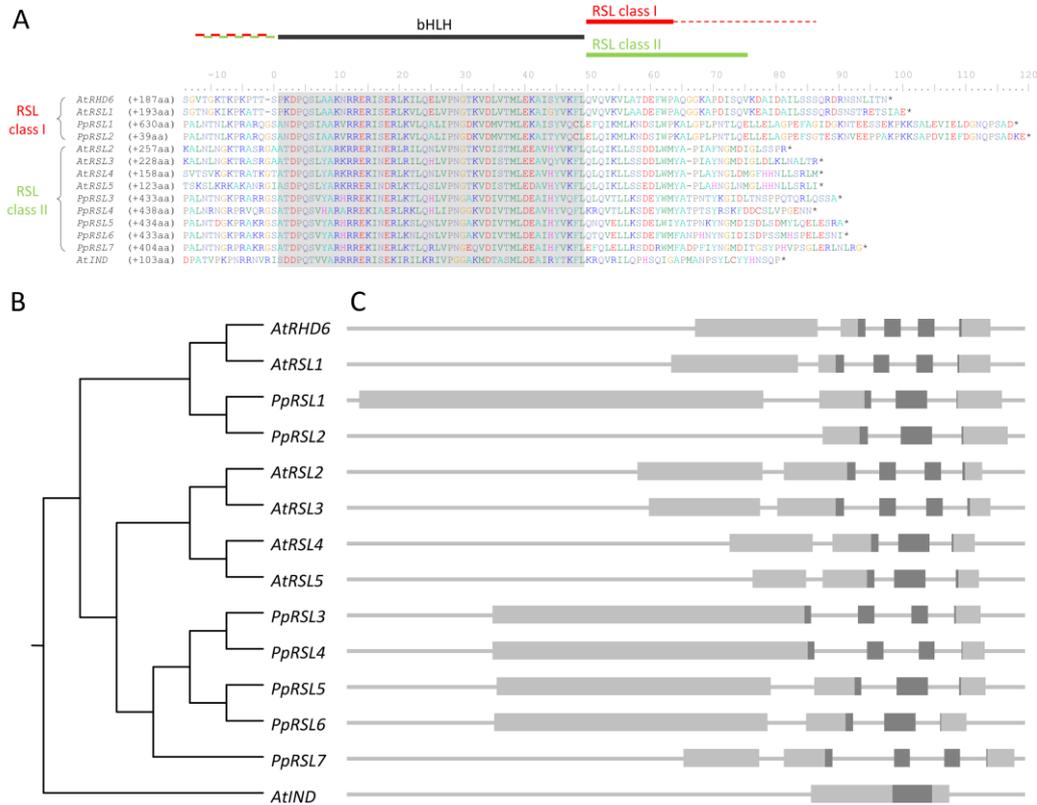
Figure 3.2 – *RSL* class I and class II genes in *Arabidopsis* and *Physcomitrella.*

A Alignment of the amino acid residues of the *Arabidopsis* and *Physcomitrella* N-terminus of RSL proteins. The bHLH domain is highlighted in grey. Immediately N-terminal to the bHLH there is a stretch of amino acids that is highly characteristic of RSL class I or RSL class II proteins (red and green lines, respectively). There is also a conserved region common to both RSL classes immediately C-terminal to the bHLH (shown in a red and green dashed line). **B** ML tree showing the phylogenetic relationship of the *Arabidopsis* and *Physcomitrella RSL* genes; the tree was rooted with AtIND (bHLH subfamily VIIIb). **C** Representation of the intron-exon structures of the RSL genes; the bHLH region is shown in dark.

The RSL class II protein AtRSL4 has a D-box (Yi 2008), an amino acid motif that has been implicated in the degradation of a human bHLH protein involved in axon growth (Lasorella et al. 2006). The canonical D-box motif R**L***N was found in SmRSL3, SmRSL4, SbRSL6, CsRSL3, AlRSL4 and AtRSL4. A slightly different version of the motif, R**L****N/D/E (Wei et al. 2004), was found in PpRSL6, SmRSL1, SmRSL2, ZmRSL2 and ZmRSL5. The lack of consistent phylogenetic prevalence of the D-box and the relative simplicity of its amino acid motif suggest that the AtRSL4 D-box motif, although conserved in *Arabidopsis lyrata*, has evolved several times and is not a conserved feature derived from ancient RSL proteins.

Several residues are characteristic of each of the two RSL subfamilies and can be used to easily distinguish them. The most conspicuous is the presence, in the C-terminal RSL conserved region, of the residues WP in RSL class I and WMYA in RSL class II proteins (position 123 in Appendix 4; position 78 in Appendix 5). Another obvious distinction of the two RSL classes is the presence, in the 7th amino acid of the second helix, of a basic amino acid (usually K) in class I proteins and an acidic amino acid (usually E) in RSL class II proteins (position 102 in Appendix 4; position 52 in Appendix 5). A few other residues are characteristic to each of the two RSL classes.

To determine the evolutionary relationships between the different RSL proteins, the alignments shown in Appendix 4 and Appendix 5 were used for phylogenetic analyses. Fig. 3.3 shows a ML tree of the RSL class I proteins. The two *Physcomitrella* genes are clustered at the base of the RSL class I subfamily tree, indicating that they derived from a recent genome duplication (the ancestor of *Physcomitrella* underwent a whole-genome duplication 30-60 million years ago (Rensing et al. 2007). Three pairs of highly similar *Selaginella moellendorffii* proteins separate the mosses from the angiosperms. The high similarity of the amino acid sequences makes it difficult to unambiguously define the evolutionary relationships of the angiosperm *RSL* genes. All the monocot RSL class I proteins seem to derive from a common ancestral protein; AtRHD6 and AtRSL1 separated before the divergence of *Arabidopsis lyrata* and *Arabidopsis thaliana*. There is a high degree of duplications of these proteins – this is clearly manifested in *Glycine max*, whose four RSL class I proteins shared a recent common ancestral.
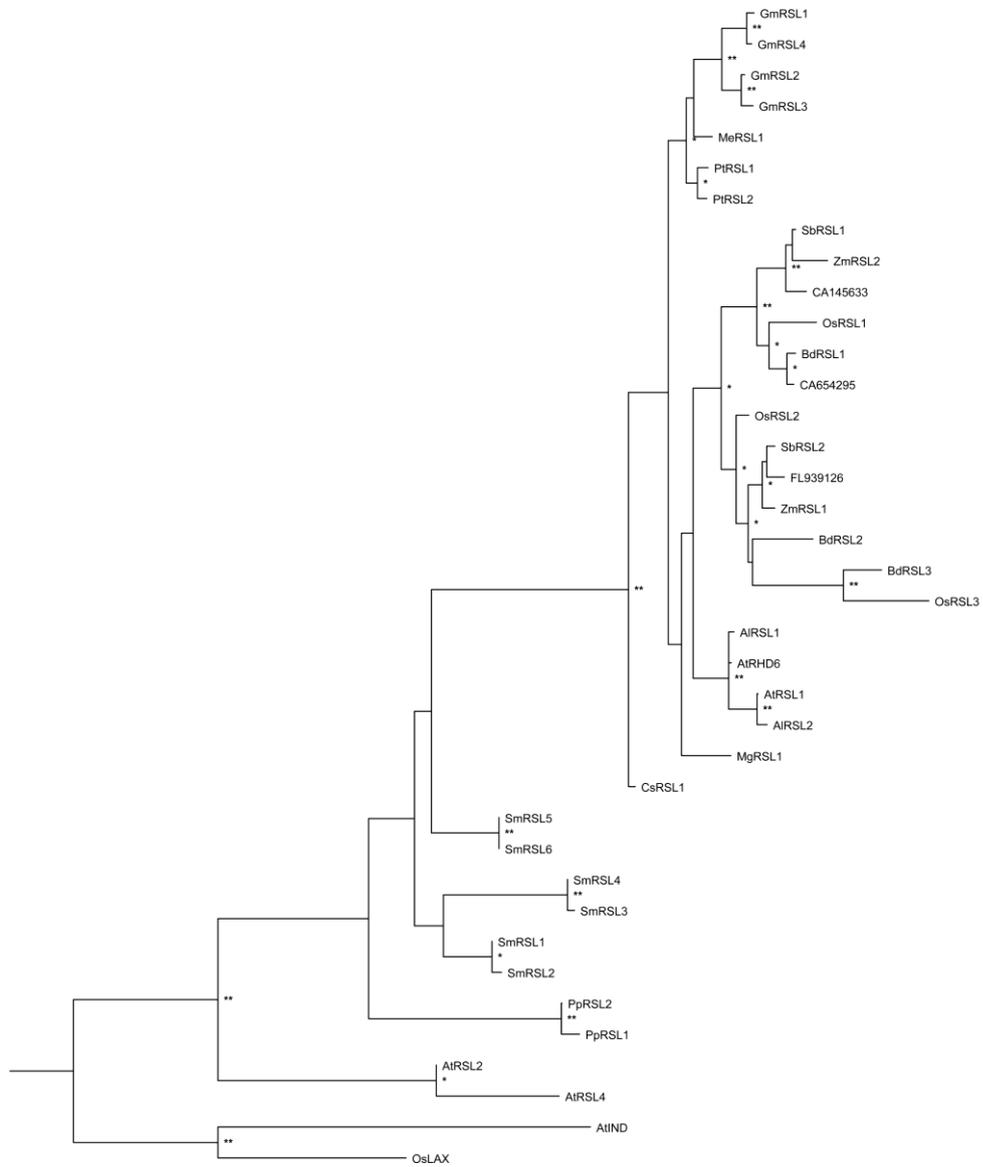
Figure 3.3 – ML phylogenetic tree of RSL class I proteins.

The tree was calculated using the alignment shown in Appendix 4 and rooted with the bHLH subfamily VIIIb clade (AtIND and OsLAX). Two RSL class II proteins (AtRSL2 and AtRSL4) are also shown. aLRT support values are indicated in the nodes (* > 0.8; ** > 0.9).
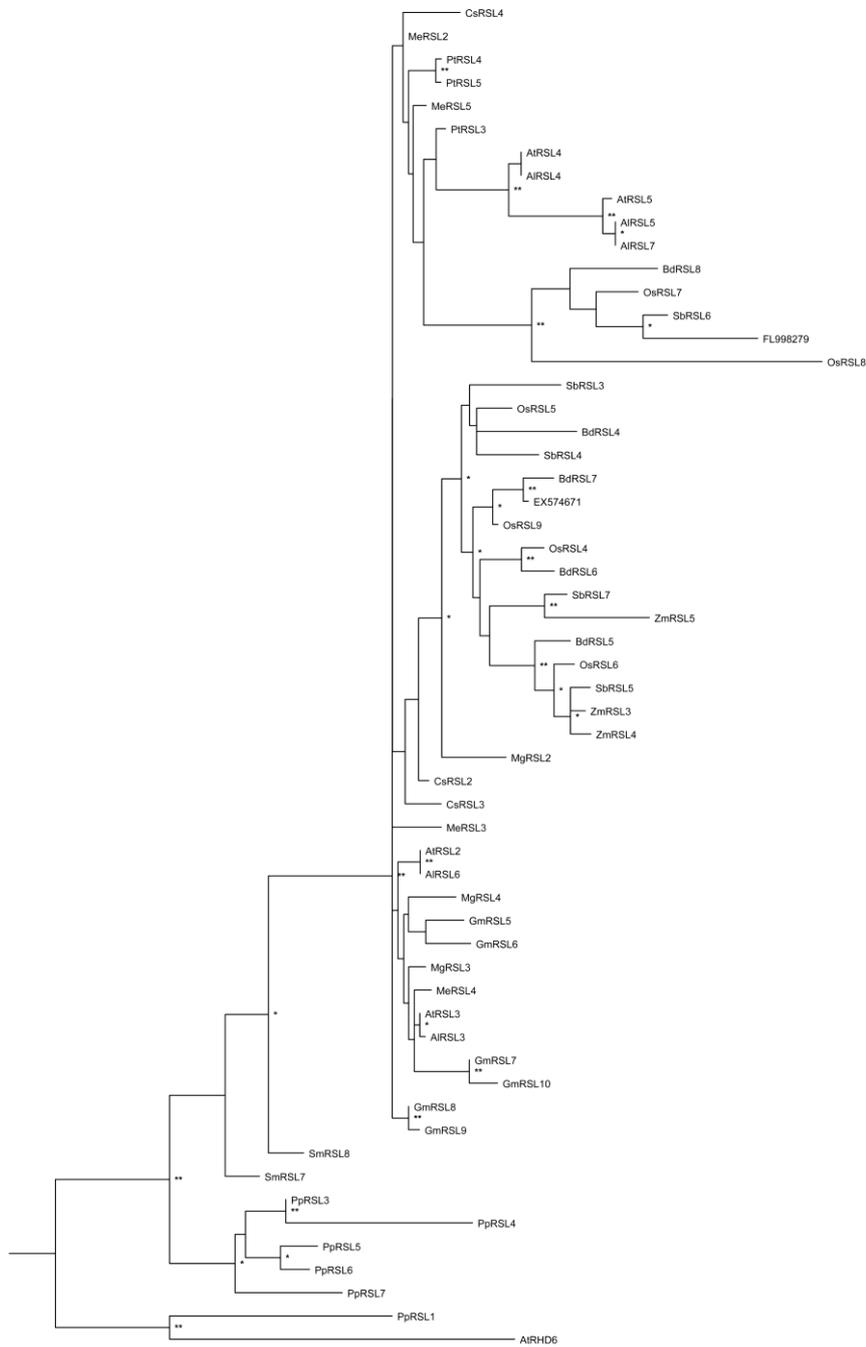
Figure 3.4 – ML phylogenetic tree of RSL class II proteins.

The tree was calculated using the alignment shown in Appendix 5 and rooted with two RSL class I proteins (PpRSL1 and AtRHD6). aLRT support values are indicated in the nodes (* > 0.8; ** > 0.9).

Figure 3.5 – Bayesian phylogenetic tree of RSL class II proteins.

The tree was calculated using the alignment shown in Appendix 5 and rooted with two RSL class I proteins (PpRSL1 and AtRHD6). Posterior probability values are indicated in the nodes (* > 0.8; ** > 0.9).

The RSL class II alignment was used to calculate ML (Fig. 3.4) and Bayesian trees (Fig. 3.5). Again, the five *Physcomitrella* RSL class II proteins cluster together at the base of the tree, and are clearly separated from the angiosperm RSLs by the *Selaginella moellendorffii* proteins. The resolution of the phylogenetic relationships of the angiosperm RSL class II proteins is problematic: branches have low support values in the ML tree and are highly unresolved in the Bayesian tree. However, it is clear that the four *Arabidopsis thaliana* genes have evolved by genome duplications before the divergence of *Arabidopsis thaliana* and *Arabidopsis lyrata*. There are two monophyletic groups of monocot RSL class II proteins. Unfortunately, the existence of very long branches in one of these complicates the phylogenetic analysis, and it is not possible to determine if they diverged after of before the divergence of monocots from other angiosperms.

In conclusion, RSL class I and class II are distinct bHLH subfamilies that evolved before the divergence of mosses from other land plants, over 443 million years ago. There are many similarities between the two classes, and an RSL class I + RSL class II clade is clearly monophyletic. These proteins are present throughout land plants, and the number of *RSL* genes per genome is quite constant. However, apart from the very high degree of conservation in the bHLH domain and adjacent regions, most of the protein primary structure is not well conserved.

3.3.2. The molecular function of RSL class II proteins is conserved between mosses and angiosperms

Despite the low similarity of the amino acid sequences of RSL proteins outside the highly conserved bHLH and adjacent amino acids, the function of RSL class I proteins is highly conserved between mosses and angiosperms. This was demonstrated by a cross-species complementation experiment: expression of *PpRSL1* under the control of the CaMV 35S promoter in *Atrhd6* plants rescues the development of root hairs (Menand et al. 2007b). This shows that the *Physcomitrella* PpRSL1 protein can substitute for the *Arabidopsis* AtRHD6, demonstrating that the molecular function of *RSL* class I genes is conserved despite having last shared an ancestral over 440 million years ago.

To investigate if the molecular function of *RSL* class II genes is also conserved between mosses and angiosperms, the coding sequences of the four expressed *Physcomitrella* RSL class II genes were expressed in *Atrsl2-1 Atrsl4-1* plants, under the control of the CaMV 35S promoter. *Atrsl2 Atrsl4* plants do not develop root hairs (Yi

2008; Yi et al. 2010; Fig.3.6): the root epidermis develops small bulges, but root hair tip growth fails to initiate. This is caused by an additive effect of the *Atrsl2* and *Atrsl4* mutations, which individually cause subtler defects in root hair initiation (*Atrsl4*) and tip growth (*Atrsl2* and *Atrsl4*) (Yi et al. 2010; Fig. 3.6).

All *Atrsl2 Atrsl4* double mutants expressing *Physcomitrella RSL* class II genes show some degree of complementation (Fig. 3.6). The roots of *Atrsl2 Atrsl4* plants expressing the *PpRSL3* gene have well defined bulges, clearly more pronounced than the loose bulges of *Atrsl2 Atrsl4* plants. These bulges are uniformly present across the whole root epidermis, but occasionally some elongate to form very short root hairs. In the collet (the hypocotyl-root junction) there are many short hairs, as opposed to the totally hairless collet of *Atrsl2 Atrsl4* double mutants (Fig. 3.6B). Plants expressing *PpRSL4* have hairless roots with some bulges but, unlike in *Atrsl2 Atrsl4* double mutants, the collet region of these plants develop well defined bulges. Plants expressing *PpRSL5* have very short hairs or bulges all over the root epidermis. Many of these hairs have a tip with a rounded hook shape, indicating a defect in the polarization of tip growth. Plants expressing *PpRSL6* show the strongest complementation of the *Atrsl2 Atrsl4* hairless phenotype: the root epidermis forms root hairs that are larger (66 ±29 µm)[5] than in plants expressing *PpRSL5* (36 ±12 µm), although much smaller than in the wild-type Col-0 (325 ±161 µm). Some of these root hairs have an abnormal shape, again indicating a defect in growth polarization. The collet of *Atrsl2 Atrsl4* double mutants expressing *PpRSL5* or *PpRSL6* shows many short root hairs.

Interestingly, the overexpression of the native *AtRSL5* gene under the control of the CaMV35S promoter in *Atrsl2 Atrsl4* double mutants results in the formation of small hairs and bulges (Yi 2008) very similar to the ones present in *Atrsl2 Atrsl4* double mutants overexpressing *PpRSL5* and *PpRSL6*. This indicates that the divergence in function between the *Physcomitrella* and *Arabidopsis* RSL class II proteins is not greater than the divergence between the RSL class II paralogs in *Arabidopsis*. Together, these results indicate that the molecular function of *RSL* class II proteins is partially conserved between mosses and angiosperms.

---

[5] mean ±SD n=30
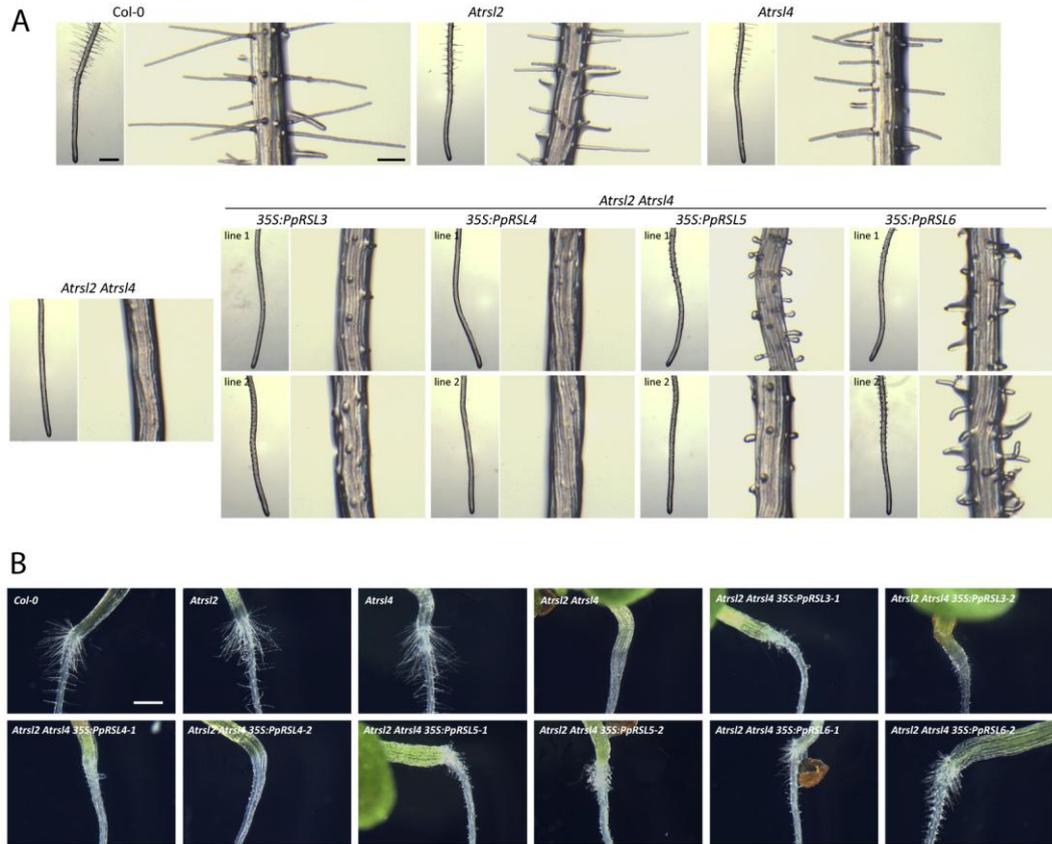
Figure 3.6 – Expression of *Physcomitrella RSL* class II genes in *Atrsl2 Atrsl4 Arabidopsis* plants.

A Primary roots and root epidermis of the wild type Col-0, *Atrsl2-1* and *Atrsl4c* single mutants, *Atrsl2-1 Atrsl4c* double mutants and *Atrsl2-1 Atrsl4c* plants expressing the *Physcomitrella RSL* class II genes *PpRSL3*, *PpRSL4*, *PpRSL5* and *PpRSL6* (two independent lines are shown). **B** Collet region in the same plants. Scale bars indicate 500μm in the root and collet pictures and 100μm in the close-ups of the root epidermis.

3.3.3. *Physcomitrella RSL* class II genes are expressed in protonema and gametophores

To define the expression patterns of the different *RSL* genes during the life cycle of *Physcomitrella*, RNA was extracted at four different stages of development: 1) 14 days old protonema colonies, which consist only of chloronema and small caulonema cells (Fig. 3.7A); 2) 21 days old protonema colonies, which consists mostly of chloronema and highly differentiated caulonema cells (Fig. 3.7B); 3) isolated gametophores (Fig. 3.7C); 4) and green, immature sporophytes (Fig. 3.7D).

RT-PCR analysis shows that *RSL* class I genes are expressed in all the four stages of development (Fig. 3.7E). This came as a slight surprise, since a previous analysis of *PpRSL1* and *PpRSL2* promoter-GUS fusions failed to detect GUS staining in young sporophytes (Benoît Menand, data not published). A possible explanation for this inconsistency is that the steady state levels of *RSL* class I mRNAs are lower in sporophytes than in gametophores and protonema. RT-PCR is more sensitive to low amounts of transcript than GUS staining, so it could detect a weak expression of *PpRSL1* and *PpRSL2* in the sporophytes. Supporting this observation is the fact that *PpRSL1* and *PpRSL2* transcripts were recently found in a cDNA library derived from green sporophytes (NCBI Accessions DC929769 and DC944304).

Unlike *RSL* class I, the expression of *RSL* class II genes is restricted to the protonema and gametophores (Fig. 3.7E). *PpRSL3* is expressed in mature protonema and gametophores; *PpRSL4* is predominantly expressed in mature protonema colonies; while *PpRSL5* and *PpRSL6* are expressed in gametophores and throughout protonema development. The transcript of *PpRSL7* was not detected in any of the four stages.

Although RT-PCR should not be strictly considered as semi-quantitative (the 35 cycles used result in a saturation of amplification for some of the reactions), it is possible to draw some conclusions regarding the expression levels for some of the genes. *PpRSL3* has higher steady state levels in gametophores than in 21 days old protonema; *PpRSL4* has a very faint expression in gametophores; *PpRSL5* and *PpRSL6* have a peak of expression in 21 days old protonema. These results also suggest that the steady state levels of *PpRSL3* and *PpRSL6* are higher than other *RSL* class II genes.
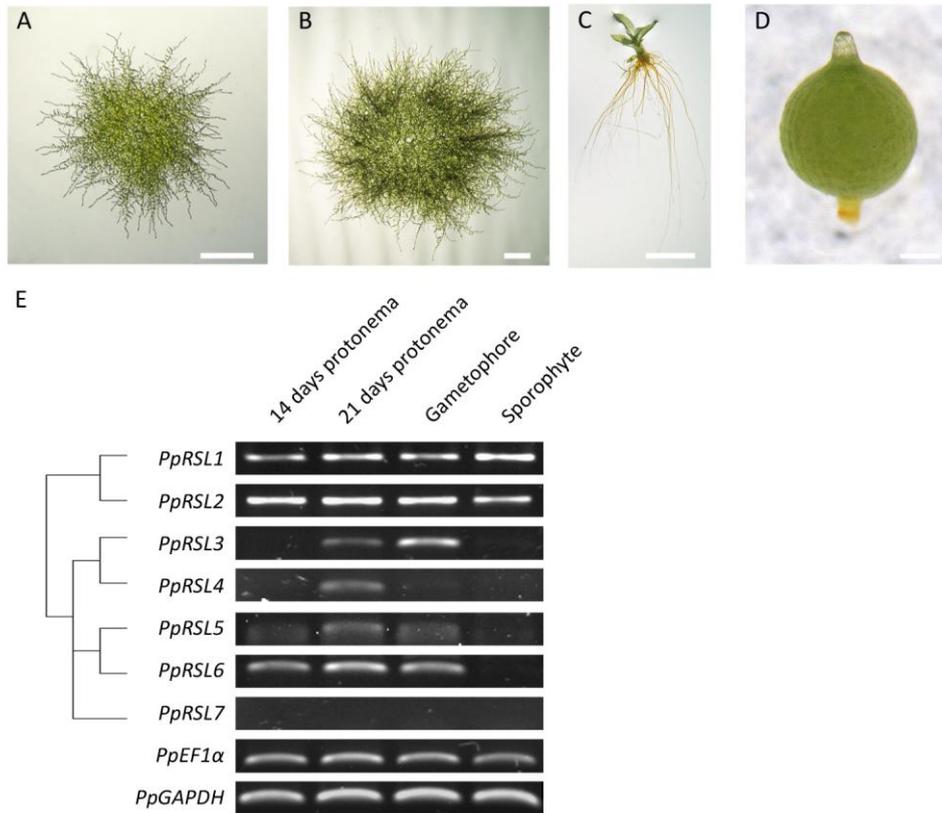
Figure 3.7 – Expression of *RSL* class I and *RSL* class II genes throughout the life cycle of *Physcomitrella*. RNA was extracted from 14 days old (**A**) and 21 days old (**B**) protonema colonies growing on cellophane disks, gametophores isolated from 28 days old plants (**C**) and immature sporophytes, isolated from plants 32 days after the beginning of the 3-week 17°C gametangia induction period (**D**). 200ng total RNA from each sample was reverse transcribed and used for PCR amplification using 35 cycles (**E**). Scale bars indicate 1mm in A-C and 200μm in D.

The expression of *RSL* class I genes is not only more widespread throughout the life cycle, but it also seems to occur at a higher level that RSL class II genes. A non-orthodox corroboration of this hypothesis is that the NCBI database contains more ESTs of *RSL* class I than *RSL* class II genes. In other words, the transcripts of *PpRSL1* and *PpRSL2* have been detected in more cDNA libraries (about 10 EST sequences each) than those of *PpRSL3* and *PpRSL5* (only one EST) or *PpRSL4*, *PpRSL6* and *PpRSL7* (no EST sequences).

In summary, *RSL* class II genes are expressed only in the gametophyte of *Physcomitrella*; their peak of expression occurs either in mature protonema or in gametophores. In contrast, *RSL* class I genes are expressed throughout the different stages of *Physcomitrella* life cycle, and at a higher level.

3.3.4. RSL class II proteins redundantly control protonema development

Strategy and generation of mutants

To determine the function of *RSL* class II genes in *Physcomitrella*, single and double mutants were generated using double homologous recombination. This technique allows the precise replacement of a genomic locus with a foreign DNA sequence that confers resistance to an antibiotic (Schaefer 2001). The constructs for *Physcomitrella* transformation were built using a genomic fragment upstream of the start codon and a genomic fragment downstream of the stop codon as the target sequences for the homologous recombination. Each transformed line was initially screened (twice) for antibiotic resistance, then by a series of PCR reactions (Fig. 3.8) that verified: 1) the absence of the targeted locus 2) the correct integration of the 5' border of the antibiotic resistance construct 3) and the correct integration of the 3' border of the antibiotic resistance construct. Finally, the number of unspecific integration events was examined by Southern blot (Fig. 3.8).

Three independent *Pprsl3* plants had previously been generated (Benoît Menand). Two independent *Pprsl4* lines, one *Pprsl5* line and three *Pprsl6* lines were then generated and selected for phenotypical analysis. Since the full transcript of *PpRSL7* could never be amplified (due to its very low/absent expression), no knockout construct was generated. Given the possibility that *PpRSL3* and *PpRSL4* are redundant duplicated genes (see Figs. 3.4, 3.5 and Appendix 3; PpRSL3 and PpRSL4 have 92% similarity in the
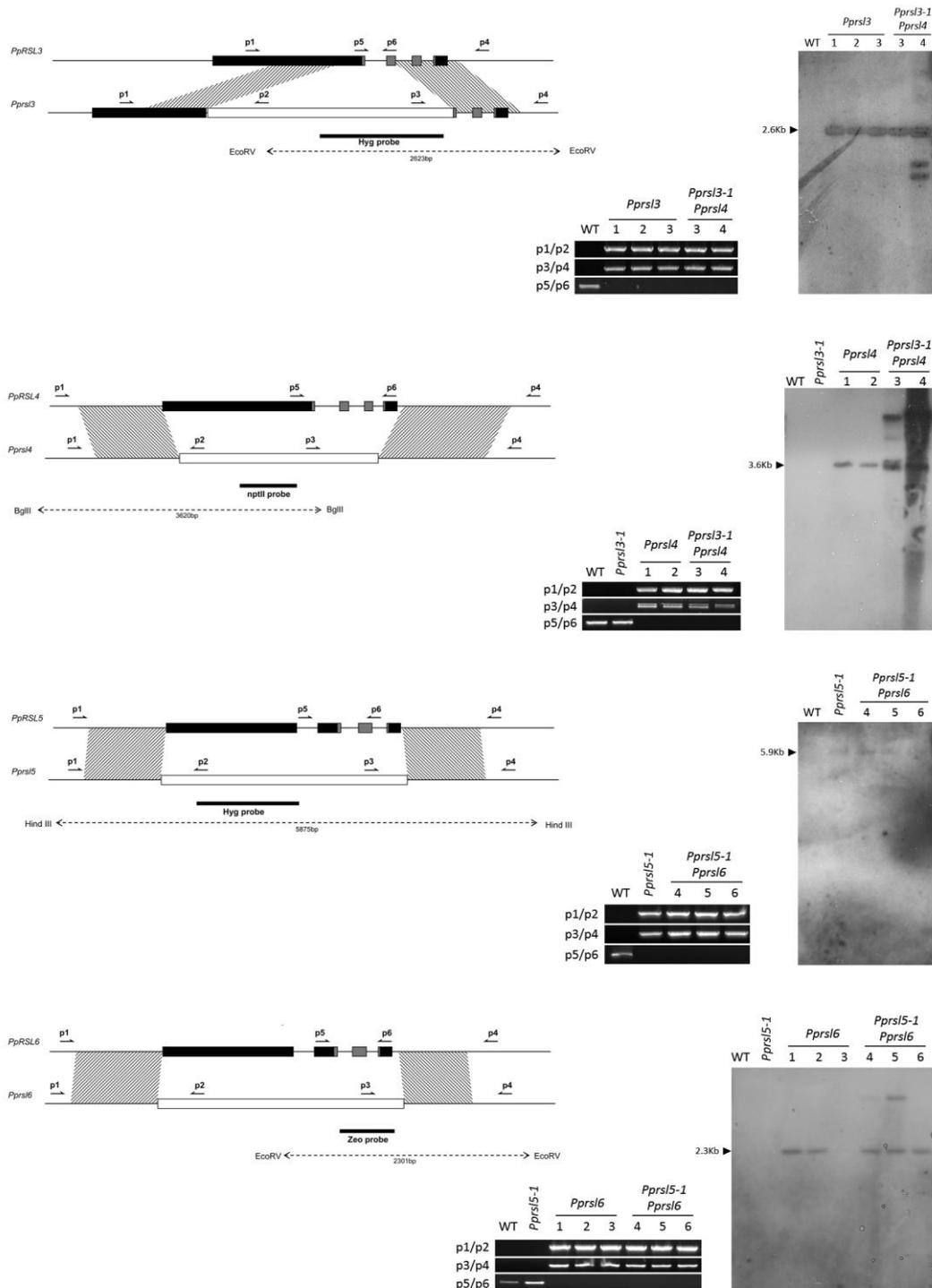
Figure 3.8 – Construction of knockout lines by double homologous recombination.

The structure of each locus is indicated. The dark boxes correspond to exons, with the grey boxes indicating the position of the bHLH domain; the white box indicates the cassette that confers resistance to hygromycin, G418 or zeocin. The dashed region indicates the position of the genomic region used for the homologous recombination. The harpoons indicate the location of the primers (p1-p6) used in the PCRs to confirm the recombination events. The black bar indicates the location of the probe used in the Southern blots to confirm the number of integrations in the genome. The location of the corresponding restriction sites is indicated by a dashed arrow.
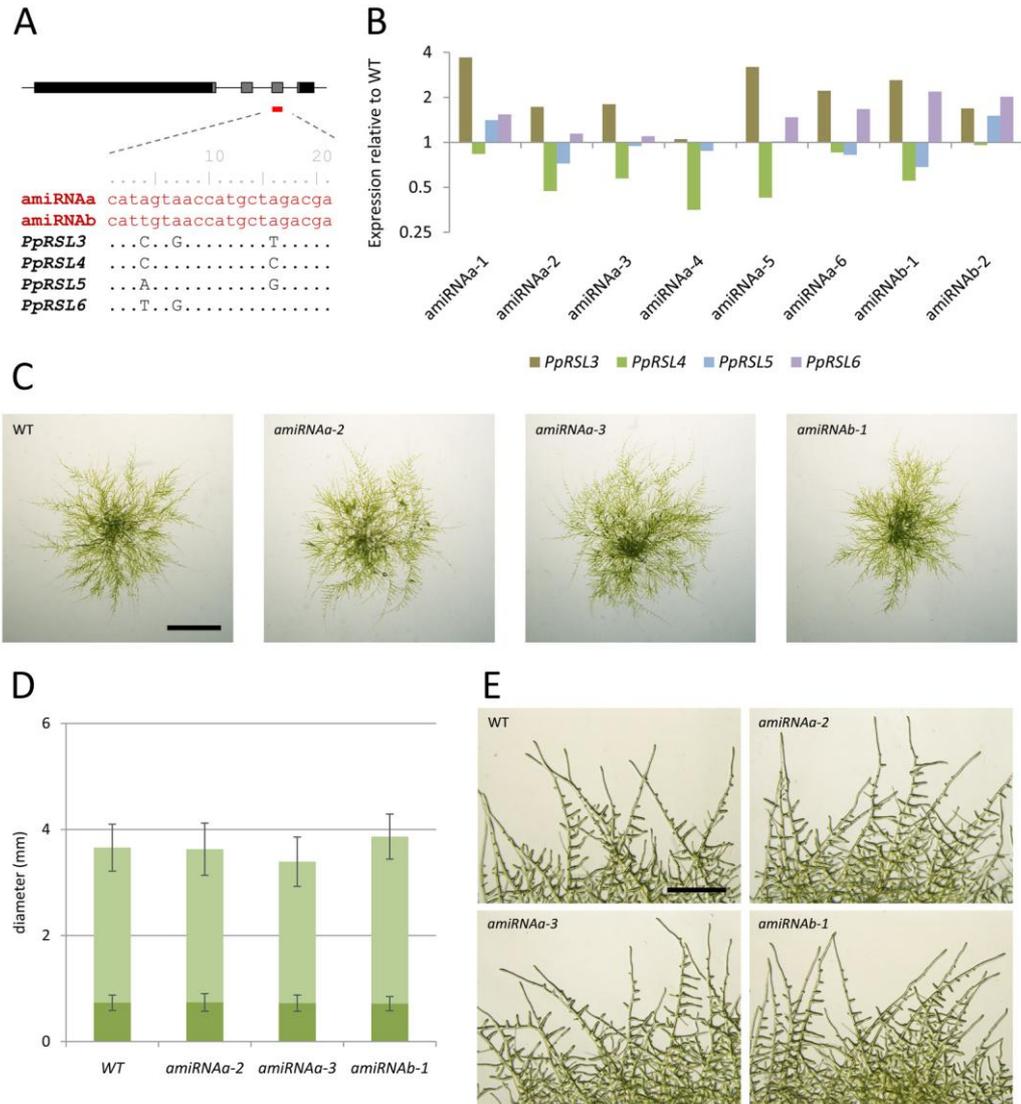
Figure 3.9 – Generation and characterisation of amiRNA lines.

A Design of two 21bp constructs targeting a conserved nucleotide sequence in the bHLH coding region of *RSL* class II transcripts. The grey boxes in the intron-exon diagram indicate the position of the bHLH coding region; the dots in the alignment represent identical nucleotides. **B** Expression of RSL class II genes in the amiRNA lines generated, determined by qRT-PCR. The Y-axis represents fold change in expression relative to WT. **C** 21 days old protonema colonies grown on minimal media without cellophane disks. **D** Diameter of 21 days old colonies (mean ±SD, n=23); the dark green bars indicate the inner denser protonema and the light green bars indicate the full diameter (see Fig 3.11). Means are not significantly different from WT (p<0.01). **E** Micrograph of the edge of 21 days old colonies. Scale bars indicate 1mm in C and 500µm in E.

bHLH domain and 61% in the whole protein), *Pprsl3 Pprsl4* double mutants were generated. The *Pprsl3-1 Pprsl4-4* and *Pprsl3-1 Pprsl4-5* lines show multiple integrations of the G418 resistance cassette, probably caused by a concatenation of the targeting construct prior to the integration (Kamisugi et al. 2006). The Hyg probe also detected non-specific bands in the *Pprsl3-1 Pprsl4-5* line, possibly the result of an occurrence of homologous recombination between the neomycin resistance construct and the hygromycin resistance cassette already present in the *Pprsl3-1* line (the hygromycin and G418 resistant cassettes have high similarity over 60% of the sequence). Nevertheless, the phenotype of the two independent *Pprsl3 Pprsl4* lines is identical. *Pprsl5 Pprsl6* mutants were also generated (PpRSL5 and PpRSL6 have 100% similarity in the bHLH domain and 74% in the whole protein). One of the lines (*Pprsl5-1 Pprsl6-5*) shows a non-specific integration of the hygromycin resistance construct, but these plants are undistinguishable from *Pprsl5-1 Pprsl6-4* and *Pprsl5-1 Pprsl6-6* double mutants.

In order to try to down-regulate the expression levels of more than two *RSL* genes at the same time, two amiRNA constructs targeting a conserved nucleotide region of four class II *RSL* genes were designed (Fig. 3.9A). Several plants transformed with the amiRNA precursor were generated. However, a qRT-PCR analysis showed that the expression levels of *PpRSL3*, *PpRSL4*, *PpRSL5* and *PpRSL6* were not significantly affected (Fig. 3.9B). Accordingly, these plants were undistinguishable from WT (Fig. 3.9C-E).

All the knockout lines generated were grown until completion of the life cycle, so that spores could be collected and used to start cultures for the phenotypical analyses. Protonema, gametophores, sporophytes and viable spores could be easily obtained for all the different lines, indicating that no critical defects were caused by the knockout of these genes.

Pprsl3 Pprsl4 plants have smaller and denser protonema colonies

In order to characterise protonema development in the mutant lines, spores were germinated on minimal media and the development of the protonema colonies was followed for 4 weeks. WT plants develop an inner region of denser protonema, composed predominantly of cells with chloronema characteristics (small and filled with many chloroplasts), growing predominantly on the surface of the media (Fig. 3.10).
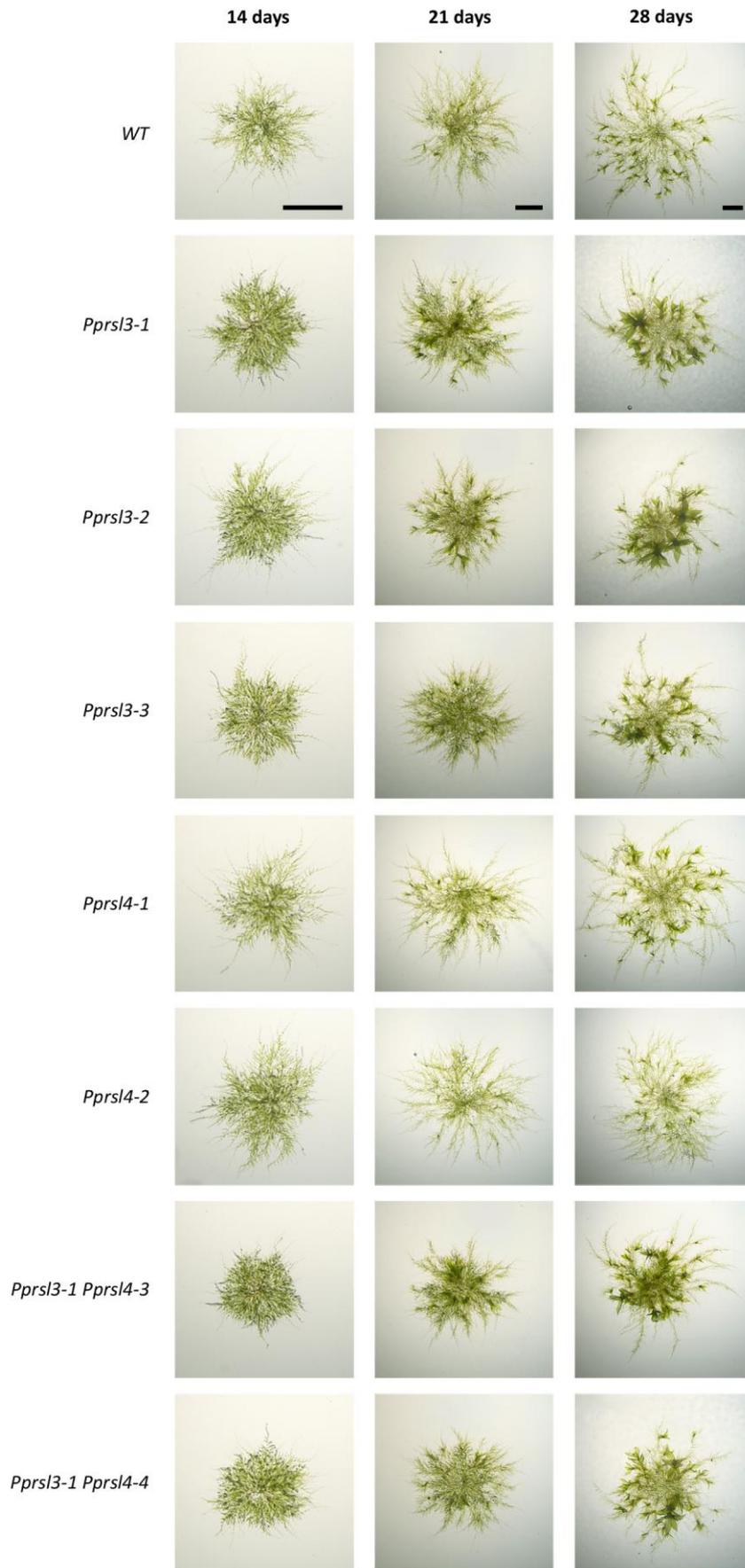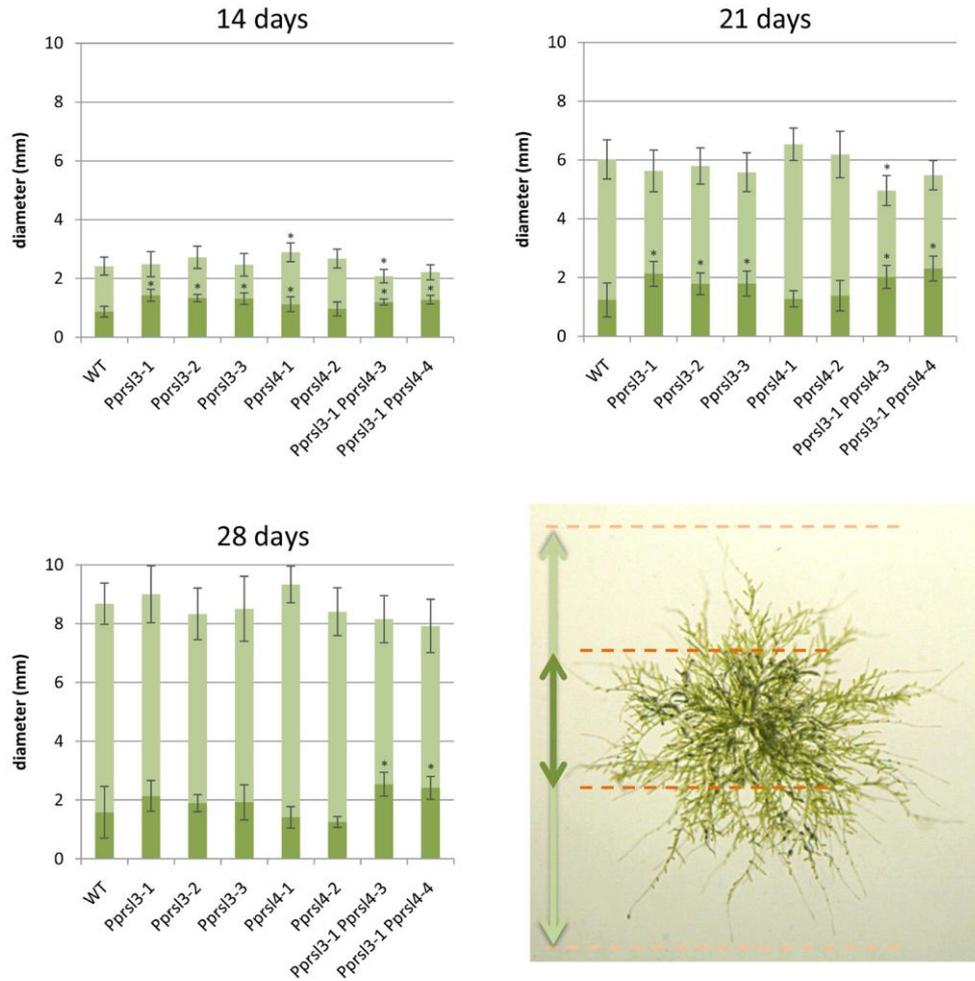
Figure 3.11 – Diameter of *Pprsl3*, *Pprsl4* and *Pprsl3 Pprsl4* protonema colonies.

Spores were germinated on minimal media without cellophane disks; the diameters of the protonema colonies (light green bars) and of the inner denser protonema region (dark green bars) were measured after 14, 21 and 28 days. Error bars indicate SD, n=23 (14 days), n=21 (21 days), n=11 (28 days). Asterisks indicate values that are statistically significantly different from WT (p<0.001).

_____

(previous page)

Figure 3.10 – *Pprsl3*, *Pprsl4* and *Pprsl3 Pprsl4* protonema colonies.

Spores were germinated on minimal media without cellophane disks. Pictures were taken after 14, 21 and 28 days. Scale bars indicate 1mm.

	14 days	**21 days**	**28 days**
WT			
Pprsl5-1			
Pprsl6-1			
Pprsl6-2			
Pprsl6-3			
Pprsl5 Pprsl6-4			
Pprsl5 Pprsl6-5			
Pprsl5 Pprsl6-6			
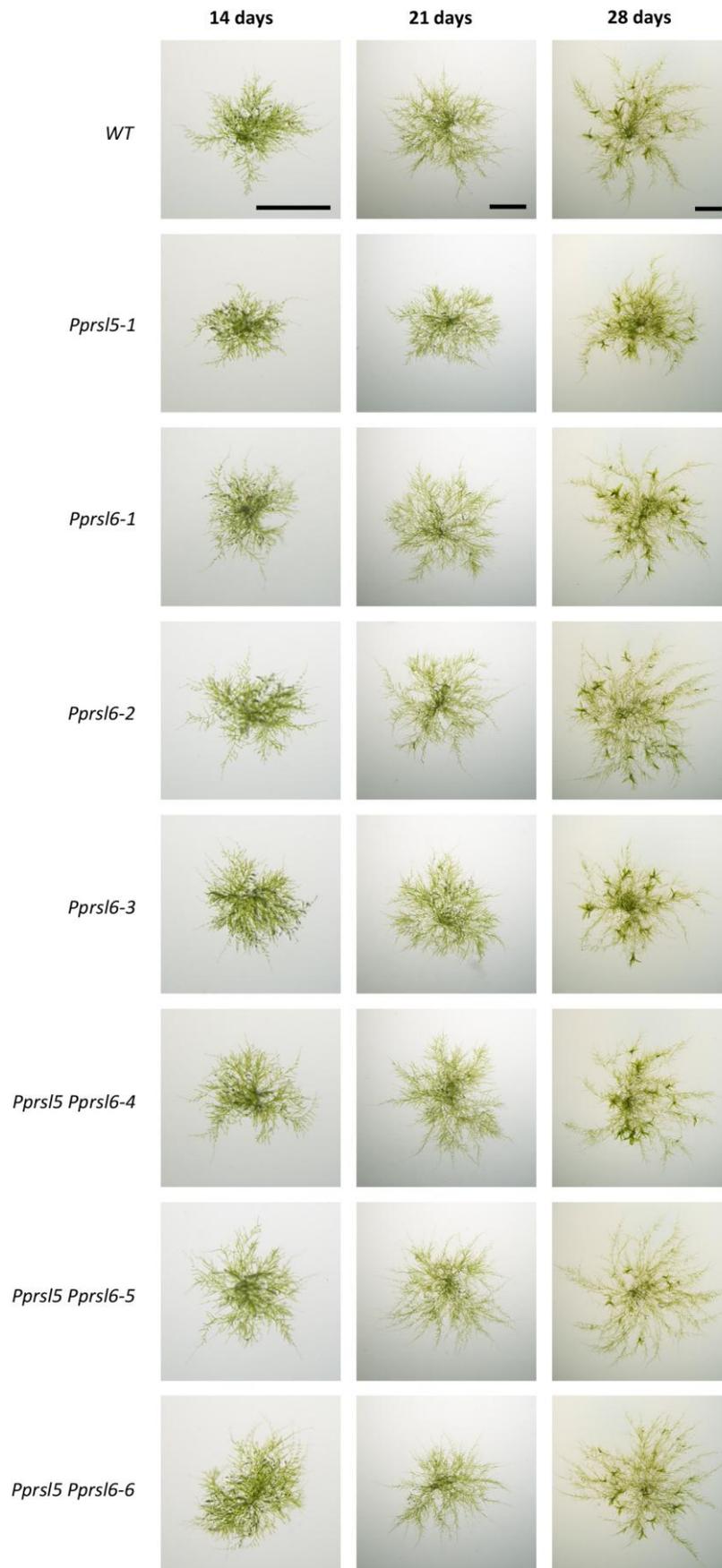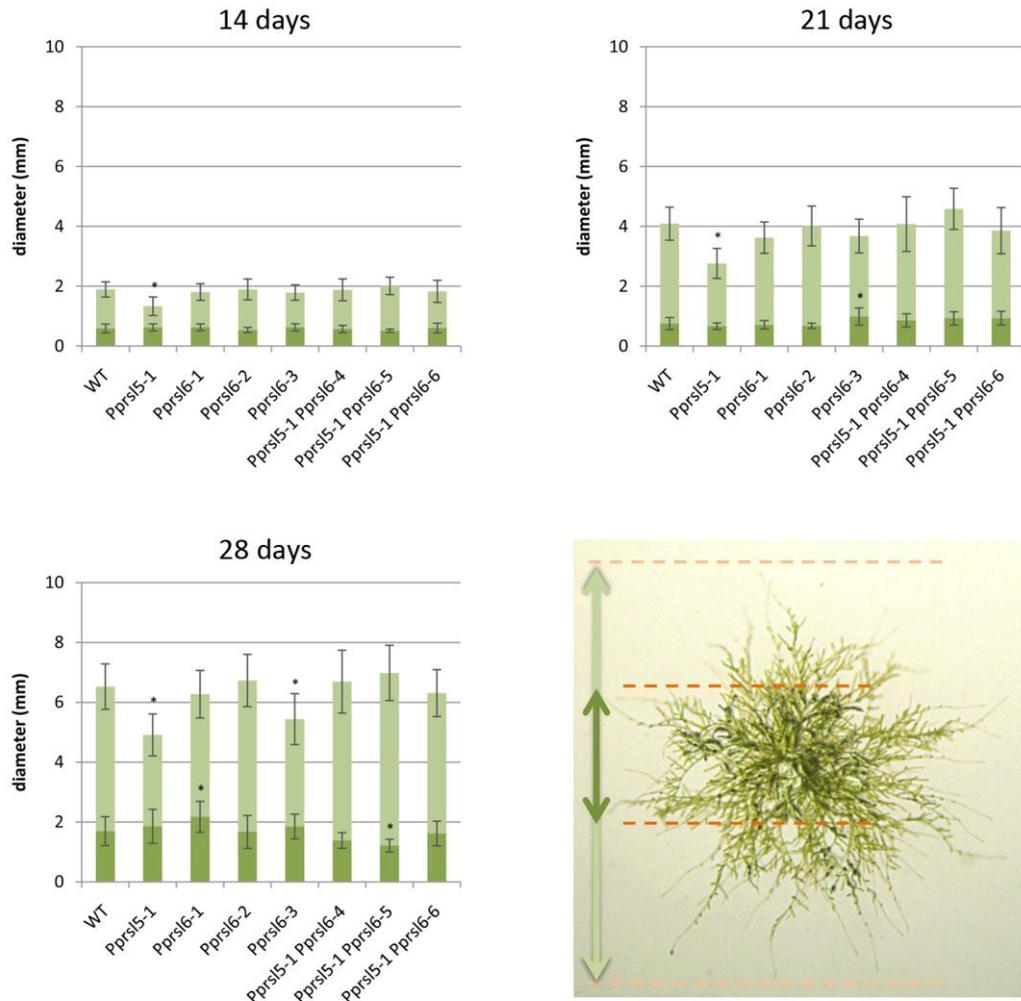
Figure 3.13 – Diameter of *Pprsl5*, *Pprsl6* and *Pprsl5 Pprsl6* protonema colonies.

Spores were germinated on minimal media without cellophane disks; the diameters of the protonema colonies (light green bars) and of the inner denser protonema region (dark green bars) were measured after 14, 21 and 28 days. Error bars indicate SD, n=25 (14 days), n=27 (21 days), n=27 (28 days). Asterisks indicate values that are statistically significantly different from WT (p<0.001).

(previous page)

Figure 3.12 – *Pprsl5*, *Pprsl6* and *Pprsl5 Pprsl6* protonema colonies.

Spores were germinated on minimal media without cellophane disks. Pictures were taken after 14, 21 and 28 days. Scale bars indicate 1mm.

At the edges of the colony, longer filaments grow away from the centre, often penetrating into the agar; these filaments are composed of cells with predominantly caulonema characteristics (long and with few chloroplasts). In *Pprsl3* and *Pprsl3 Pprsl4* plants (and occasionally in *Pprsl4*), the inner denser protonema region is larger and more conspicuous than in WT colonies (Fig. 3.10 and Fig. 3.11). This is already apparent in 2 week-old plants, and it continues to be a visible trait in older plants. The total diameter of *Pprsl3 Pprsl4* colonies is also consistently reduced when compared with either WT or the single mutant lines.

Due to the large number of lines involved, *Pprsl5*, *Pprsl6* and *Pprsl5 Pprsl6* plants were grown separately from *Pprsl3*, *Pprsl4* and *Pprsl3 Pprsl4*. Fig. 3.12 shows the phenotypes of these typical protonema colonies at 14, 21 and 28 days. Unlike the *Pprsl3* and *Pprsl3 Pprsl4* double mutants, the *Pprsl5* single, *Pprsl6* single and *Pprsl5 Pprsl6* double mutants are undistinguishable from WT colonies (Fig. 3.12 and Fig. 3.13). *Pprsl5-1* mutant colonies were smaller than WT colonies, but this phenotype was not reproducible. Unfortunately, another *Pprsl5* line (*Pprsl5-2*) was lost due to an irreversible bacterial contamination, so there is no independent way to verify the *Pprsl5* phenotype. However, the most logical interpretation is that *Pprsl5* has no defects in protonema development and the differences observed in Fig. 3.12 occurred by chance.

Protonema branching

The edges of *Pprsl3*, *Pprsl4* and *Pprsl3 Pprsl4* protonema colonies were indistinguishable from WT: a main axis of long (caulonema) cells with side-branches developing behind the first subapical cells (Fig. 3.14A-H). In WT, the different lengths and cell cycle duration of caulonema and chloronema cells result in different branching patterns in these two cell types: the first branch of a caulonema filament typically initiates further away from the filament tip than does the first branch of a chloronema filament. To determine if there is a different branching pattern in any of the knockout lines, the distance between the tip of filaments protruding from the edge of protonema colonies and their first side-branch was measured (Fig. 3.14I). Interestingly, the first side-branch develops further away from the tip in WT than in any of the *Pprsl3* single, *Pprsl4* single and *Pprsl3 Pprsl4* double mutants. This suggests that the branching in these lines is less caulonema-like than in WT, supporting the hypothesis that PpRSL3 and PpRSL4 positively regulate caulonema development. *Pprsl5*, *Pprsl6* and *Pprsl5 Pprsl6*

filaments, however, have a similar branching pattern to WT (Fig. 3.15) and the edge of these protonema colonies appears normal.

Cell size is not controlled by RSL class II genes

The *Pprsl3 Pprsl4* phenotype suggests that there is a defect in the transition from chloronema to caulonema. To confirm that the different knockout lines differentiate both caulonema and chloronema cells, protruding filaments were isolated from 21 days-old plants and stained with Calcofluor White, a fluorescent dye that stains cellulose in the cell walls.

In WT plants, a typical filament is formed by a tip-growing caulonema apical cell that divides regularly (and obliquely to the transverse plane), leaving behind it an axis of long caulonema cells with characteristic oblique cross cell walls. The third / fourth cell of a filament then forms a lateral side-branch cell that also starts growing by tip-growth; regular spaced divisions leave behind it a secondary axis of short chloronema cells with characteristic perpendicular cross cell walls. As the filament grows larger, the side-branch apical cell starts to differentiate caulonema characteristics, and the newly formed subapical cells start to display typical caulonema oblique cross cell walls (Fig. 3.16B). All the *Pprsl3*, *Pprsl4*, *Pprsl5*, *Pprsl6*, *Pprsl3 Pprsl4* and *Pprsl5 Pprsl6* lines follow the same development pattern as a WT filament (Fig. 3.16Aand Fig. 3.17A). To confirm that there are no differences in the length of caulonema cells and chloronema cells, 1) the length of the first subapical cell of protruding filaments and 2) the length of the second proximal cell of side-branches with more than three cells, respectively, were measured (Fig. 3.16Band Fig. 3.17B). No significant differences between the WT and any of the knockout mutants were found. Together, this data indicates that *Pprsl3*, *Pprsl4*, *Pprsl5*, *Pprsl6*, *Pprsl3 Pprsl4* and *Pprsl5 Pprsl6* plants are able to differentiate caulonema and chloronema cells.

Single and double mutants of RSL class II genes differentiate rhizoids

Protonema colonies growing on minimal media start to differentiate gametophores 3-4 weeks after spore germination. The number and size of gametophores is similar in the different lines. Unlike *Pprsl1 Pprsl2* plants, which have few and small rhizoids (Menand et al. 2010b), all single and double mutants of *RSL* class
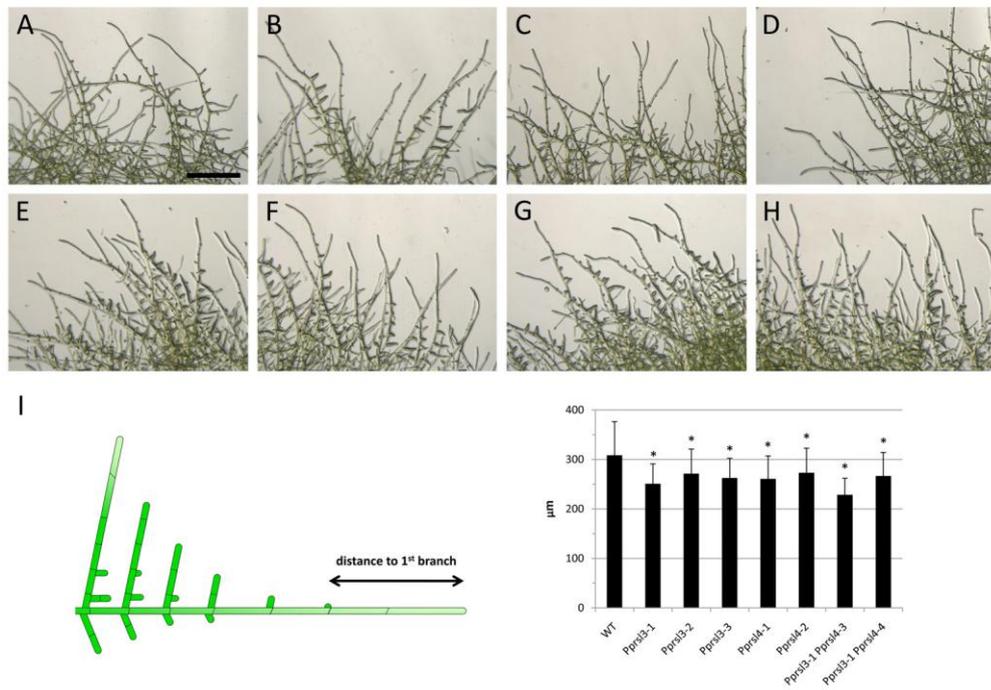
Figure 3.14 – Close up of the edges of *Pprsl3*, *Pprsl4* and *Pprsl3 Pprsl4* protonema colonies.

A WT; **B** *Pprsl3-1*; **C** *Pprsl3-2*; **D** *Pprsl3-3*; **E** *Pprsl4-1*; **F** *Pprsl4-2*; **G** *Pprsl3-1 Pprsl4-3*; **H** *Pprsl3-1 Pprsl4-4*. The plants were grown for 21 days in minimal media overlaid with cellophane disks. The scale bar indicates 500μm. **I** Distance from the tip of a filament to the first branch; error bars indicate SD, n=52. Asterisks indicate values that are statistically significantly different from WT (p<0.001).
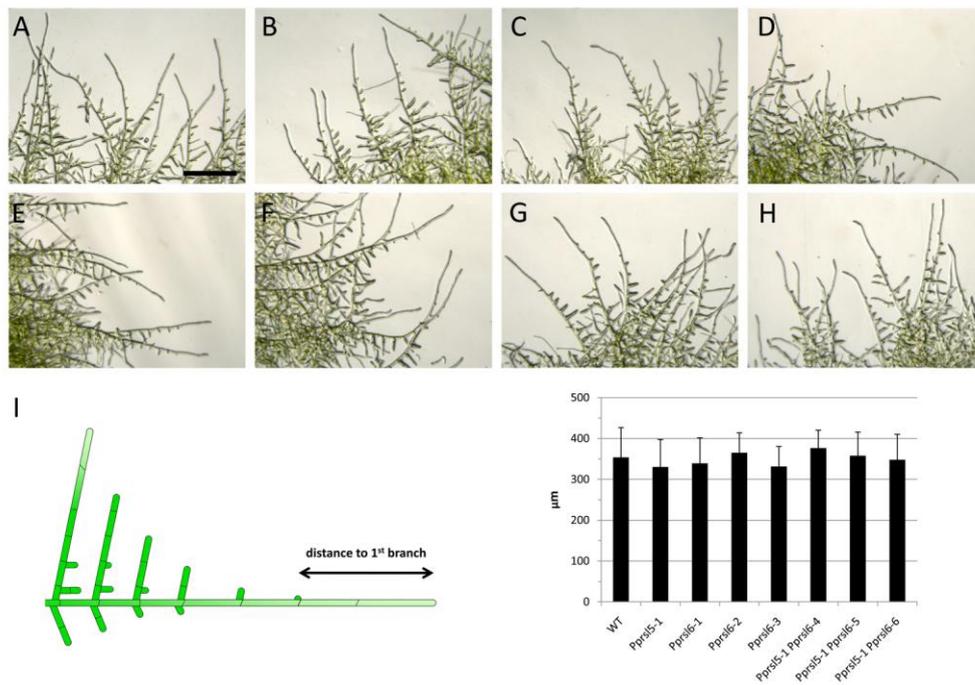
Figure 3.15 – Close up of the edges of *Pprsl5*, *Pprsl6* and *Pprsl5 Pprsl6* protonema colonies.

A WT; **B** *Pprsl5-1*; **C** *Pprsl6-1*; **D** *Pprsl6-2*; **E** *Pprsl6-3*; **F** *Pprsl5-1 Pprsl6-4*; **G** *Pprsl5-1 Pprsl6-5*; **H** *Pprsl5-1 Pprsl6-6*. The plants were grown for 21 days in minimal media overlaid with cellophane disks. The scale bar indicates 500μm. **I** Distance from the tip of a filament to the first branch; error bars indicate SD, n=47. None of the lines is significantly different from WT (p<0.05).
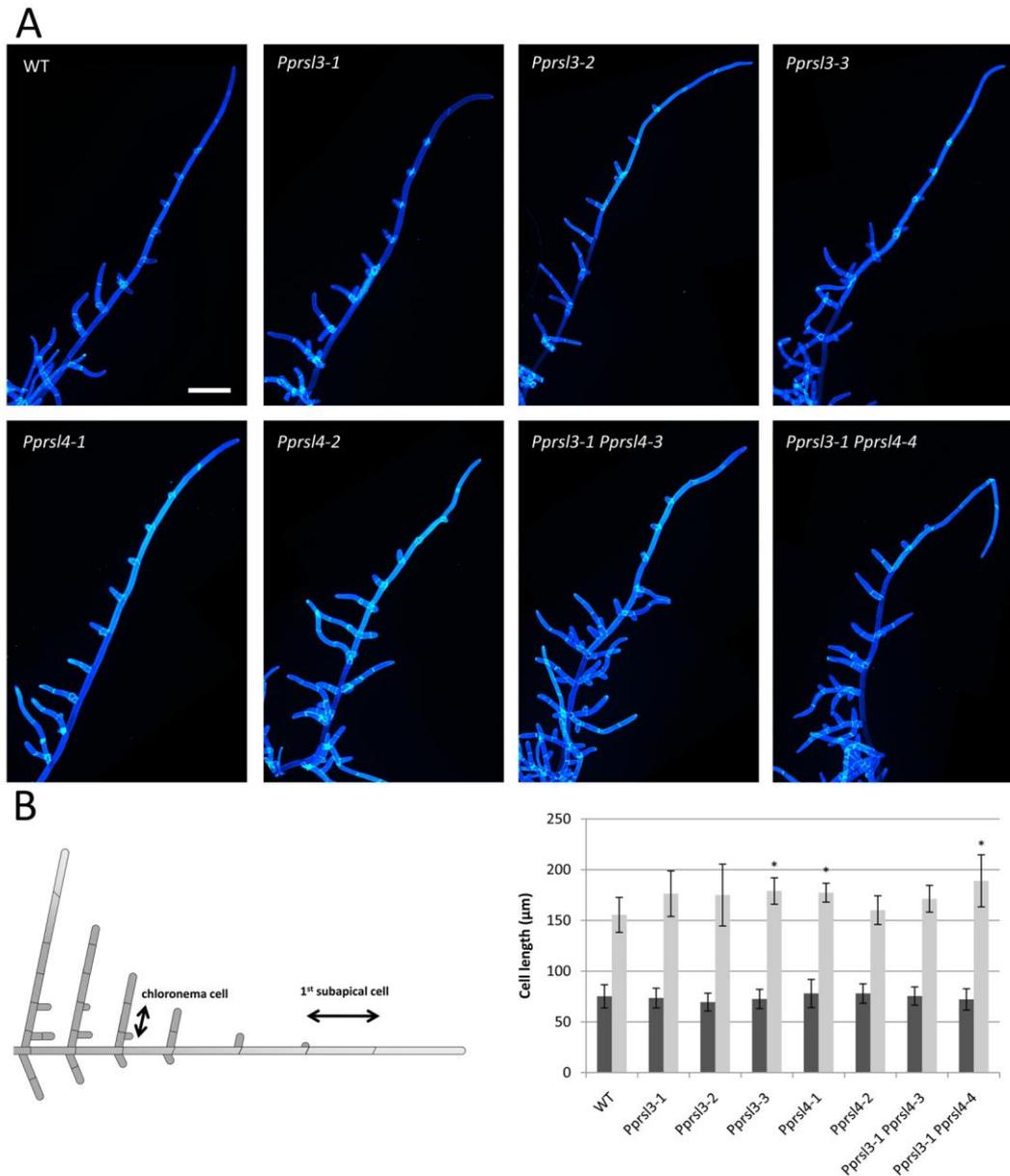
Figure 3.16 – Protonema cell size in *Pprsl3*, *Pprsl4* and *Pprsl3 Pprsl4* plants.

A Filaments protruding from 21 days old protonema colonies grown on minimal media overlaid with cellophane disks were isolated, stained with Calcofluor White and observed in an epifluorescence microscope. **B** Lengths of the first subapical cell of a long (>10 cells) filament (light grey bars) and the length of the second proximal cell of a side-branch with ≥3 cells (dark grey bars). Error bars indicate SD, n=15 (subapical), n=37 (side-branch). Asterisks indicate values that are statistically significantly different from WT (p<0.01). Scale bars indicate 200μm.
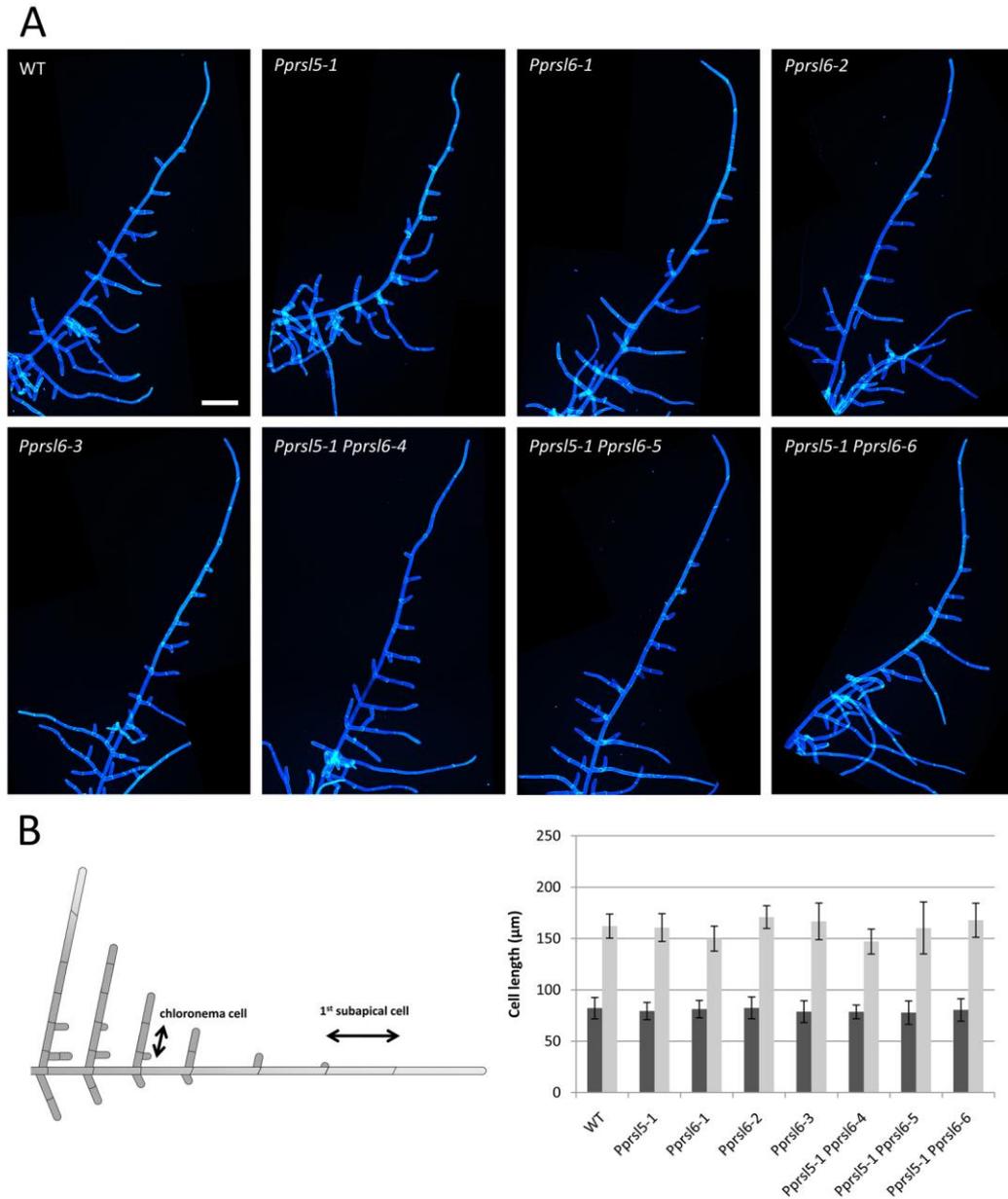
Figure 3.17 – Protonema cell size in *Pprsl5*, *Pprsl6* and *Pprsl5 Pprsl6* plants.

A Filaments protruding from 21 days old protonema colonies grown on minimal media overlaid with cellophane disks were isolated, stained with Calcofluor White and observed in an epifluorescence microscope. **B** Lengths of the first subapical cell of a long (>10 cells) filament (light grey bars) and the length of the second proximal cell of a side-branch with ≥3 cells (dark grey bars). Error bars indicate SD, n=8 (subapical), n=48 (side-branch). Asterisks indicate values that are statistically significantly different from WT (p<0.01). Scale bars indicate 200µm.

II genes differentiate normal rhizoids (Fig. 3.18A and Fig. 3.19A). To further visualise the rhizoids, 4 week-old protonema colonies were grown for a further 2 weeks in a vertical orientation (Fig. 3.18B and Fig. 3.19B). This causes the rhizoids to grow downwards and away from the protonema centre. Colonies grown this way again confirm that rhizoid development in *Pprsl3*, *Pprsl4*, *Pprsl5*, *Pprsl6*, *Pprsl3 Pprsl4* and *Pprsl5 Pprsl6* plants is similar to WT. A similar experiment where the vertical incubation was done in the dark (which causes the rhizoids to grow upwards) also showed no difference between these lines (data not shown).

Phenotypical variability

The growth of protonema filaments is extremely sensitive to environmental factors such as light and nutrient conditions. Additionally, stochastic differences play a major role in the patterning of protonema filaments. This creates a very large variability of shapes and sizes in protonema colonies that can easily mask quantitative differences in the differentiation of caulonema and chloronema cells.

The mean of colony diameters and cell lengths (and its statistical significance) changed considerably between individual experiments (e.g. Fig. 3.11 vs. Fig. 3.13).This was probably caused by minute differences in the media dryness or composition. Another important source of variation comes from individual plates: again, this may be caused by differences in the media dryness. There is also, occasionally, small variation between colonies growing in different regions of a same 90mm plate, perhaps explained by differential access to the light sources. These external sources were superimposed on endogenous natural protonema variability, further complicating the phenotypical and statistical analyses.

Nevertheless, the denser smaller protonema colonies of *Pprsl3 Pprsl4* were a clear and highly reproducible phenotype. It remains a possibility that some of the other knockout lines also have subtler defects in protonema development, but these could not be detected using the experimental systems described here.
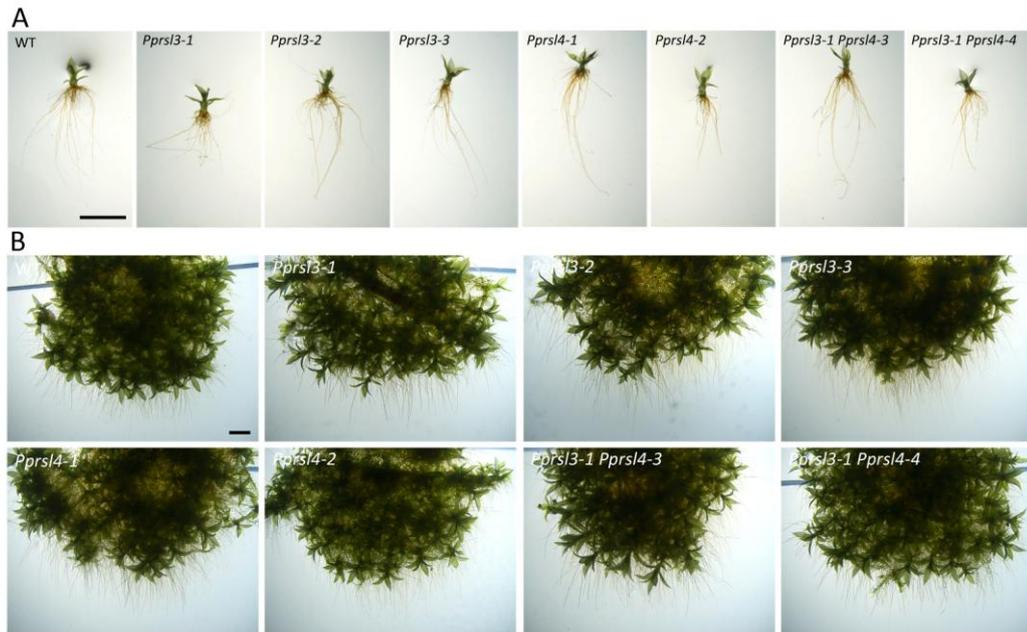
Figure 3.18 – Rhizoids with gametophores in *Pprsl3*, *Pprsl4* and *Pprsl3 Pprsl4* plants.

A Gametophores isolated from 28 days old protonema colonies growing in minimal media overlaid with cellophane disks. **B** 6-week old colonies, first grown for 28 days horizontally in minimal media overlaid with cellophane disks, and then transferred into fresh plates (minimal media without cellophane disks) and incubated vertically for further 14 days. Scale bars indicate 1mm.
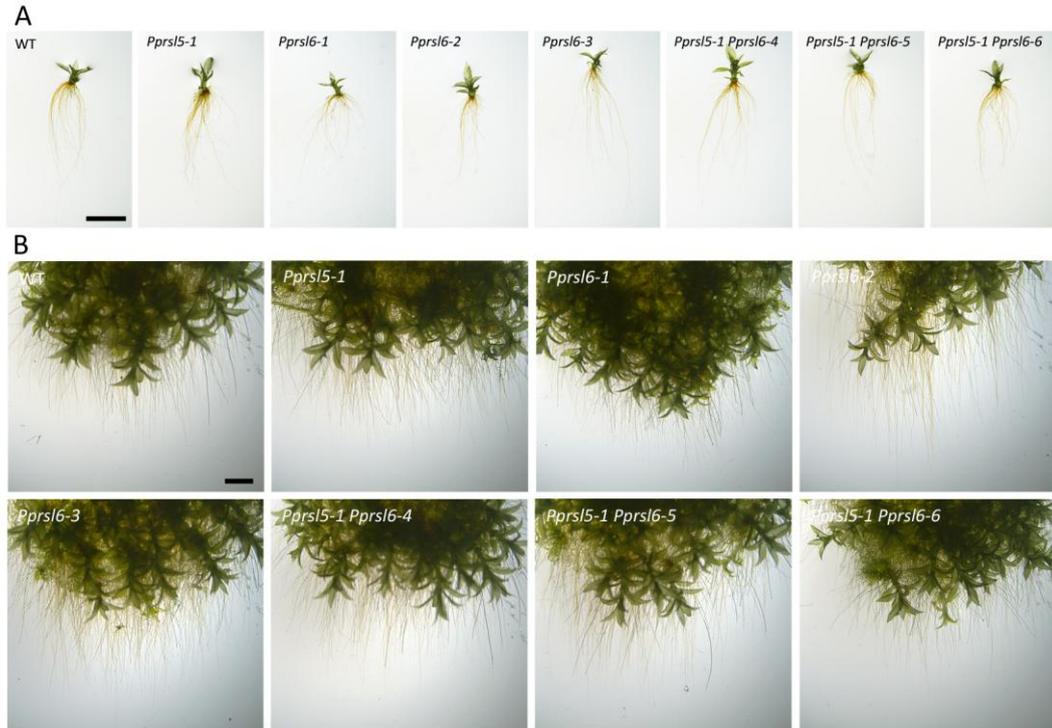
Figure 3.19 – Rhizoids with gametophores in *Pprsl5*, *Pprsl6* and *Pprsl5 Pprsl6* plants.

A Gametophores isolated from 28 days old protonema colonies growing in minimal media overlaid with cellophane disks. **B** 7-week old colonies, first grown for 28 days horizontally in minimal media overlaid with cellophane disks, and then transferred into fresh plates (minimal media without cellophane disks) and incubated vertically for further 20 days. Scale bars indicate 1mm.

3.3.5. Overexpression of *PpRSL4* causes defective caulonema differentiation

To investigate the effect of constitutively expressing the *Physcomitrella RSL* class II genes, the coding sequences of *PpRSL3* and *PpRSL4* were cloned in front of the CaMV 35S promoter and these constructs were used to transform WT plants. Two *35S::PpRSL3* and six *35S::PpRSL4* lines were obtained, and the respective increase in the steady state levels of *PpRSL3* and *PpRSL4* transcripts were confirmed by qRT-PCR (Fig. 3.20A).

The two *35S::PpRSL3* lines showed no differences from WT plants: protonema colonies (Fig. 3.20B-C), filaments (Fig. 3.20D), gametophores (Fig. 3.21A) and rhizoids (Fig. 3.21B) were all undistinguishable from WT. The *35S::PpRSL4* lines, in contrast, showed considerable differences from WT plants. The three lines examined had smaller and denser protonema colonies (Fig. 3.20B-C). This was more pronounced in the *35S::PpRSL4-1 and 35S::PpRSL4-3* lines, which is consistent with these lines having a higher increase in the levels of the *PpRSL4* transcript (Fig. 3.20A). These plants have a very dense and green protonema composed mostly of chloronema cells and are slightly reminiscent of *Pprsl1 Pprsl2* double mutant colonies (which do not differentiate caulonema cells) (Menand et al. 2010b). However, unlike these *RSL* class I double mutants, *35S::PpRSL4* plants develop occasional WT-like caulonema filaments (Fig. 3.20D). Rhizoid development in *35S::PpRSL4* was indistinguishable from WT plants (Fig. 3.21).

These results show that disrupting the pattern and level of *PpRSL4* expression causes defects in caulonema differentiation. Interestingly, both the loss of function *Pprsl3 Pprsl4* mutant and the overexpression of *PpRSL4* suppress caulonema differentiation. This suggests that a mechanism requiring a correct balance of the PpRSL4 protein level is required for caulonema differentiation.
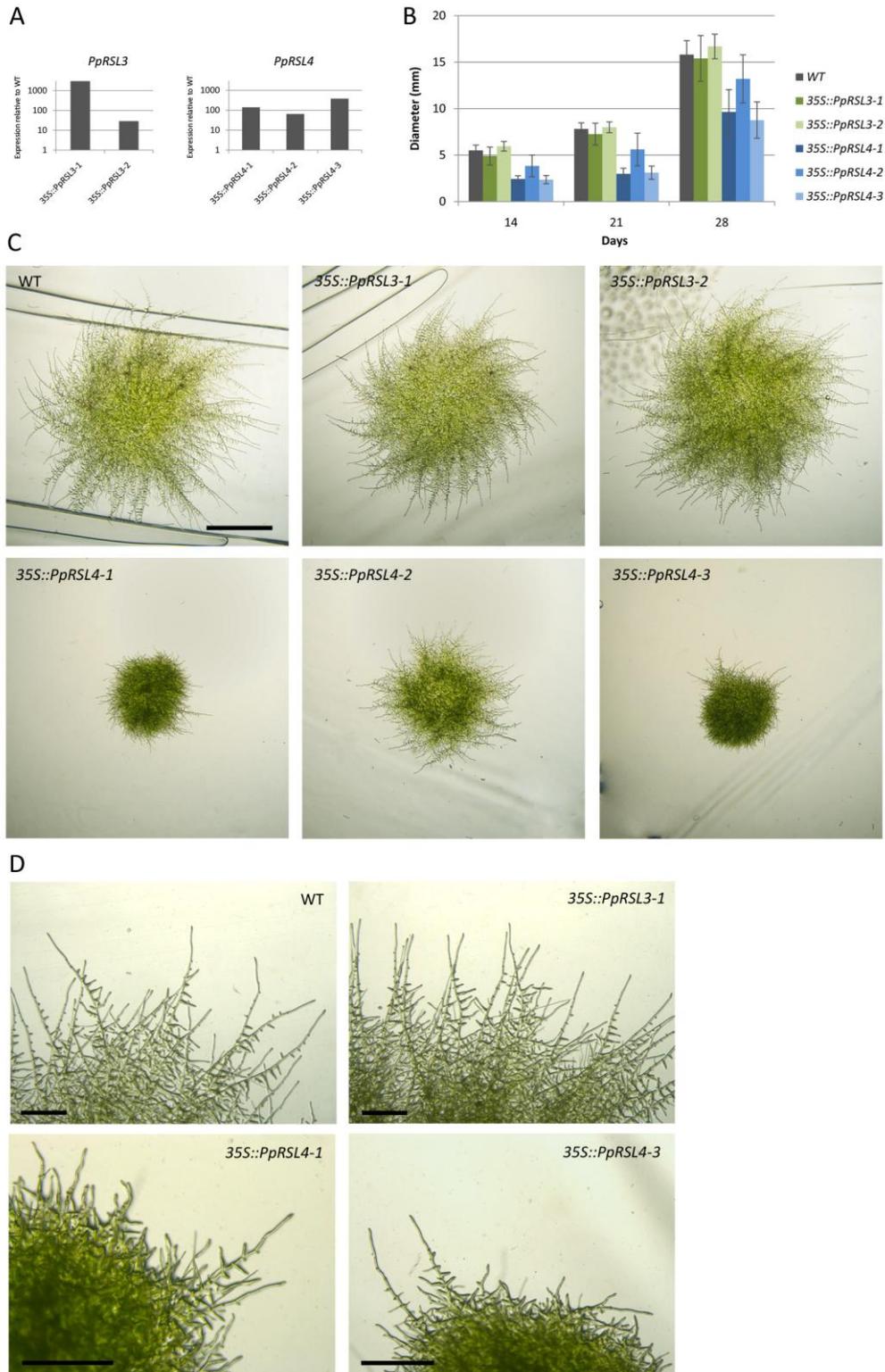
Figure 3.20 – Protonema colonies of plants constitutively expressing *PpRSL3* and *PpRSL4*.

A qPCR showing the expression levels of *PpRSL3* and *PpRSL4* in the overexpression lines, relative to the expression level of WT plants. **B** Protonema colony diameters at 14, 21 and 28 days. Error bars indicate SD, n=25 (14 days), n=17 (21 days), n=15 (28 days). **C** 21 days old protonema colonies grown on minimal media overlaid with cellophane. **D** Close-ups of the protonema edges. Scale bars indicate 2mm in C and 500μm in D.

Figure 3.21 – Rhizoids with gametophores in *35S::PpRSL3* and *35S::Pprsl4* plants

A Gametophores isolated from 28 days old protonema colonies growing in minimal media overlaid with cellophane disks. **B** 6-week old colonies, first grown for 28 days horizontally in minimal media overlaid with cellophane disks, and then transferred into fresh plates (minimal media without cellophane disks) and incubated vertically for further 14 days. Scale bars indicate 1mm.

3.3.6. The transcriptional network of *RSL* genes is different in *Physcomitrella* and in *Arabidopsis*

Physcomitrella RSL class II genes are not regulated by RSL class I genes.

In *Arabidopsis*, the *RSL* genes form a transcriptional network in which RSL class I proteins positively regulate the transcription of *RSL* class II genes. Accordingly, *RSL* class II genes are not expressed on *Atrhd6 Atrsl1* plants (Yi 2008). In order to verify if a similar transcriptional network is present in *Physcomitrella*, the expression level of *RSL* class II genes in 3 weeks old *Pprsl1 Pprsl2* plants was determined by qRT-PCR (Fig. 3.22A). *PpRSL3*, *PpRSL4* and *PpRSL5* expression levels in *Pprsl1 Pprsl2* plants fall within 0.6-1.6x of their expression levels in WT, indicating that they are not regulated by *RSL* class I genes. However, *PpRSL6* is slightly upregulated (4x the expression level in the WT), suggesting that its transcription in WT is partially inhibited (directly or indirectly) by *RSL* class I genes.

The expression of the different RSL genes is independent of RSL class II genes

Although RSL class I proteins do not play a major role in regulating the expression of *RSL* class II genes in *Physcomitrella*, it is possible that RSL class II genes themselves regulate the transcription of other *RSL*s. In order to further dissect these putative transcriptional regulations, the expression level of *PpRSL1-PpRSL6* in the *Pprsl3 Pprsl4* and in the *Pprsl5 Pprsl6* double mutant lines was examined by qRT-PCR (Fig. 3.22B-C). The expression level of *PpRSL1-PpRSL6* in these plants falls within 0.8-1.7x of the WT levels, indicating that their transcription is independent of *RSL* class II genes.

These results indicate that in *Physcomitrella*, unlike *Arabidopsis*, the expression of *RSL* genes is mostly independent of the other RSL proteins. The only exception seems to be *PpRSL6*, whose expression is slightly inhibited by RSL class I proteins in WT plants. This indicates that the topology of the transcriptional interactions between *RSL* class I and *RSL* class II genes is substantially different between *Arabidopsis* and *Physcomitrella*.
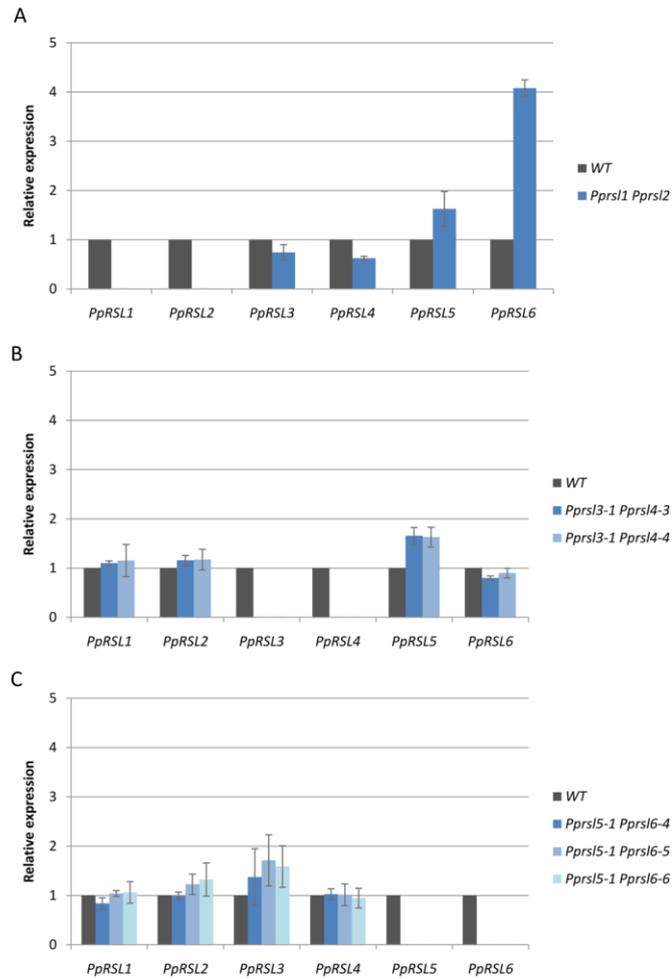
Figure 3.22 – Expression level of *RSL* genes in different *rsl* double mutants.

RNA was extracted from 21 days old *Pprsl1 Pprsl2* (A), *Pprsl3 Pprsl4* (B) and *Pprsl5 Pprsl6* (C) colonies growing on minimal media overlaid with cellophane disks. The expression levels were determined by qRT-PCR: each value corresponds to the expression level relative to WT. The mean and SD of three biological replicates are indicated.

3.3.7. RSL class I proteins negatively regulate the response of *RSL* class II genes to auxin

In *Arabidopsis*, a key player in the *RSL* transcriptional network is auxin: exogenous NAA positively regulates the expression of *AtRSL4* and *AtRSL5* and negatively regulates *AtRSL2* and *AtRSL3*, although it does not play a role in the regulation of *RSL* class I genes (Yi 2008). Interestingly, treatment of the hairless *Atrhd6 Atrsl1* plants with NAA can rescue the development of root hairs by bypassing *RSL* class I genes and directly activating *AtRSL4* (Yi et al. 2010). In *Physcomitrella*, auxin positively regulates the differentiation of caulonema and rhizoid filaments (Johri and Desai 1973; Ashton et al. 1979; Sakakibara et al. 2003). Together, this strongly suggests that auxin may play an important role in the transcriptional regulation of *RSL* genes in *Physcomitrella*.

RSL class II genes are differentially regulated by NAA

In order to this hypothesis, two-week old WT *Physcomitrella* plants were incubated for one week on minimal media supplemented with different concentrations of NAA. The expression level of *RSL* class II genes was then determined by qRT-PCR (Fig. 3.23). The results show that *PpRSL3* is negatively regulated by NAA, *PpRSL6* is positively regulated, and the expression levels of *PpRSL4* and *PpRSL5* are not significantly affected. Furthermore, the increase in the transcription of these RSL class II genes caused by NAA is concentration-dependent. Despite the fact that *Physcomitrella* and *Arabidopsis RSL* class II family members have radiated independently through numerous gene duplication and gene losses (Fig. 3.4 and Fig. 3.5), the differential regulation of *RSL* class II genes by NAA (i.e. both positive and negative regulation) in both species suggests that there was an evolutionary convergence of their regulatory mechanisms.

The auxin regulation of RSL class II transcription is repressed by RSL class I proteins

The transcription of the two *Physcomitrella RSL* class I genes (*PpRSL1* and *PpRSL2*) is positively regulated by NAA (Geupil Jang and Liam Dolan, unpublished). This raises the possibility that the effects of NAA in *RSL* class II transcription are mediated by PpRSL1 and PpRSL2. In order to test this hypothesis, the expression level of *RSL* class II

genes in *Pprsl1 Pprsl2* double mutants incubated with NAA was determined by qRT-PCR (Fig. 3.23).

In *Pprsl1 Pprsl2* plants, *PpRSL5* and *PpRSL6* are up-regulated, *PpRSL3* was down-regulated, and *PpRSL4* was not affected. This means that, with the exception of *PpRSL5*, the qualitative effects of a treatment with NAA are similar in WT and in *RSL* class I mutant plants. However, *RSL* class II genes are much more sensitive to NAA in *Pprsl1 Pprsl2* plants than in WT: significant changes in the expression levels (both positive or negative regulations) are quite clear at low NAA concentrations (0.1 and 1µM), in contrast with WT plants, where only 10µM NAA causes clear changes in expression levels. These results suggest that PpRSL1 and PpRSL2 negatively regulate the response of *RSL* class II genes to NAA. In other words, the effect of NAA on the transcription of *RSL* class II genes is inhibited, directly or indirectly, by RSL class I proteins. In a *Pprsl1 Pprsl2* background, the inhibition of the NAA response is removed and the transcription of *RSL* class II genes respond more readily to NAA (either positively or negatively).
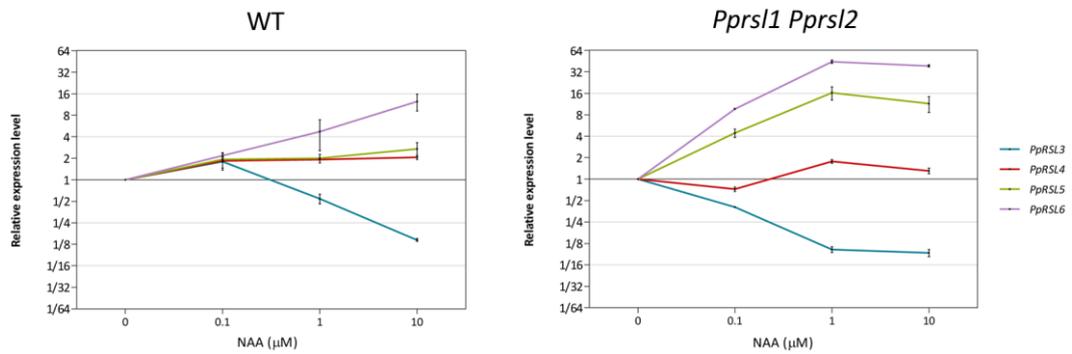
Figure 3.23 – Expression level of *RSL* class II genes in NAA treated plants.

Protonema colonies grown on minimal media for two weeks were transferred to fresh minimal media supplemented with different concentrations of NAA and incubated for one week. The expression levels were determined by qRT-PCR: each value corresponds to the expression level relative to plants grown without NAA. The mean and SE of three replicates are indicated.

3.4. Discussion

In *Arabidopsis*, *RSL* class I and *RSL* class II genes form a transcriptional network that controls the development of root hairs (Yi 2008). Root hairs are epidermal projections from the roots of vascular plants that evolved in the Devonian or Silurian Period (Raven and Edwards 2001), but the *RSL* genes originated in or before the late Ordovician Period (Chapter 2). This implies that the *RSL* developmental mechanism evolved long before the appearance of root hairs. A functional characterization of *RSL* class I genes showed that they control the differentiation of rhizoid and caulonema filaments in mosses (Menand et al. 2007b). This suggests that the development of rooting cells in early land plants was controlled by a network of *RSL* class I and class II genes, and that this network has been used to control the development of cells with a rooting function in land plants. The analyses presented in this chapter support this hypothesis, and show that the transcriptional regulatory interactions of the RSL regulatory network have changed considerably during land plant evolution.

The *RSL* class I and *RSL* class II families are highly conserved in land plants. All the land plant genomes examined have *RSL* class I and class II genes and each of these genomes has a similar number of paralogues (Appendix 2). The bHLH and adjacent RSL-specific domain of the 92 identified RSL proteins is extremely well conserved (Appendix 4 and 5). This suggests that the molecular function of the RSL proteins is conserved, a hypothesis that is further supported by cross-species complementation experiments in which the *Physcomitrella RSL* class I and *RSL* class II genes were shown to rescue (or partially rescue) the development of root hairs in RSL class I (Menand et al. 2007b) and RSL class II *Arabidopsis* mutants (Fig. 3.6), respectively.

If a differentiation mechanism involving *RSL* class I and *RSL* class II genes controlled the development of rooting cells in early land plants, then it would be expected that RSL class II proteins would (as do RSL class I proteins) also control the development of rhizoids and caulonemata in mosses. *RSL* class II genes are preferentially expressed in mature protonema and gametophores of *Physcomitrella* (Fig. 3.7), giving support to this hypothesis. Intriguingly, *RSL* class I genes (but not RSL class II) were found to be expressed in the sporophyte, although *Pprsl1 Pprsl2* plants do not show defects in sporophyte development and their expression cannot be detected using promoter:GUS constructs (Benoît Menand and Liam Dolan, unpublished). A possible explanation for

this discrepancy is that there is only a residual expression of *PpRSL1* and *PpRSL2* in the sporophyte. However, a role of *RSL* class I genes in the *Physcomitrella* sporophyte cannot be completely excluded.

The analyses of knockout mutants of *RSL* class II genes shows that there is a high degree of redundancy between them. As with many other genes in *Physcomitrella* (including *PpRSL1* and *PpRSL2*), single mutants are undistinguishable from WT plants. The main reason for this is probably the fact that there was a whole-genome duplication in *Physcomitrella* 30-60 million years ago (Rensing et al. 2007). However, *Pprsl5 Pprsl6* double mutants are also undistinguishable from WT. Only *Pprsl3 Pprsl4* plants appear consistently distinct from WT plants, showing an inhibition of the differentiation of caulonema filaments (Figs. 3.10, 3.11 and 3.14). However, no defects in rhizoids development could be detected for any of the RSL class II mutants (Figs. 3.18 and 3.19). The high protein sequence similarities and the subtle / absent phenotypes of *Pprsl3 Pprsl4* and *Pprsl5 Pprsl6* double mutants suggest that there is a high degree of redundancy between the RSL class II proteins; it would be interesting to examine the phenotype of a quadruple *Pprsl3 Pprsl4 Pprsl5 Pprsl6* mutant (no *RSL* class II function), but this was not available at the time of the writing of this thesis.

The defective caulonema differentiation caused by the overexpression of *PpRSL4* (Fig. 3.20) supports its hypothesised role in protonema development. Interestingly, both the *Pprsl3 Pprsl4* double mutant and plants overexpressing *PpRSL4* show an inhibition of caulonema development. A simple scenario that can explain these apparently contradictory results is that chloronema development is the basic 'state' during protonemal growth; for caulonema caulonema differentiation to occur, a balanced level of RSL class I, PpRSL3 and PpRSL4 proteins would be required. Interfering with this balance would cause disrupt the development of caulonema cells.

In *Arabidopsis*, a regulatory cascade involving a positive regulation of *RSL* class II genes by RSL class I proteins forms a backbone of the RSL network that controls root hair development (Yi 2008; Fig. 3.24). However, such a topology of transcriptional interactions is not present in *Physcomitrella*: apart from a subtle inhibition of *PpRSL6* expression by RSL class I proteins, the transcription of *RSL* class I and class II genes is independent of each other (Fig. 3.22 and 3.24). However, *Physcomitrella* RSL class I proteins negatively regulate the response of *RSL* class II genes to auxin (Fig. 3.23 and 3.24). This suggests that a regulatory mechanism involving auxin, *RSL* class I and *RSL* class II genes is present in *Physcomitrella*. Auxin is a key regulator of the differentiation
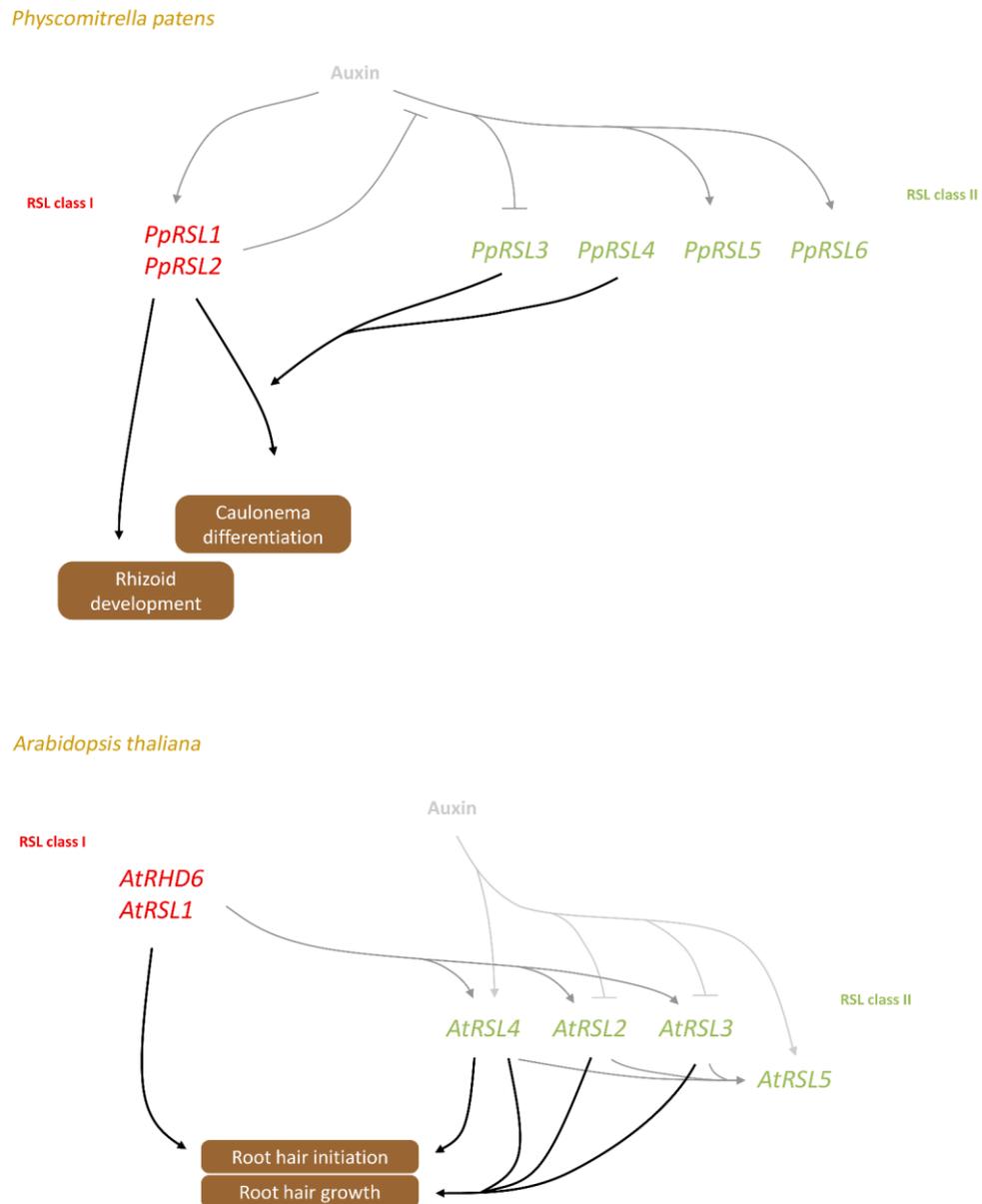
Figure 3.24 – Models of the RSL regulatory network in *Physcomitrella* and in *Arabidopsis*.

The transcriptional interactions between auxin, *RSL* class I and *RSL* class II genes are indicated. [→]: positive regulations; [⊣]: negative regulations.

of rooting cells in both mosses and angiosperms. In *Arabidopsis*, auxin positively regulates the *RSL* class II gene *AtRSL4*, bypassing the requirement for *RSL* class I genes for root hair development (Yi et al. 2010; Fig. 3.24). Importantly, auxin is not a regulator of *RSL* class I expression in *Arabidopsis.* The topology of these interactions is substantially different in *Physcomitrella*: auxin is a positive regulator of *RSL* class I expression (Geupil Jang and Liam Dolan, unpublished), and the application of exogenous auxin cannot rescue the development of caulonemata and rhizoids in the *Pprsl1 Pprsl2* mutant, indicating that auxin acts upstream of *RSL* class I genes. The expression of the *RSL* class II genes *PpRSL3*, *PpRSL5* and *PpRSL6* is also regulated by auxin (Fig. 3.23), suggesting that auxin regulates caulonema and rhizoid formation by modulating the levels of both *RSL* class I and *RSL* class II genes (Fig. 3.24). Accordingly, preliminary results (data not shown) suggest that the induction of caulonema and rhizoid development by exogenous auxin is incomplete in the *Pprsl3 Pprsl4* mutant, indicating that *RSL* class II genes are required to mediate the auxin response.

Together, these results suggest that an ancient differentiation mechanism involving auxin, *RSL* class I and *RSL* class II genes controlled the differentiation of rooting structures in the first land plants. During land plant evolution, this regulatory network was recruited to control the development of root hairs. Significantly, the topology of the transcriptional interactions between *RSL* class I and *RSL* class II genes have changed during plant evolution. The recruitment of the *RSL* regulatory network to organise the differentiation of root hairs points to a possibly widespread evolutionary mechanism: the reutilisation of ancient gene regulatory networks that evolved in green algal or in the very first land plants. However, the fine tuning of such networks (and their spatial and temporal location) will have changed considerably during plant evolution. These smaller regulatory differences probably underlie the huge morphological and physiological diversity found in land plants.

Chapter 4

General discussion

Biologists say that two structures are 'homologous' if they are derived by a series of evolutionary transformations from the same structure in a common ancestor. In this sense, human arms and bird wings are homologous, but insect and vertebrate eyes are usually thought not to be. Two genes are said to be homologous if they are derived by replication from the same gene in a common ancestor: homology of genes is deduced from similarity of base sequence. Can we deduce that two structures are homologous if they are induced by homologous genes? The example of eyes suggests that we cannot. But surely a common genetic mechanism must count as evidence in favour of homology? This is one of the many unanswered question raised by recent work in developmental genetics.

John Maynard Smith (1999) *Shaping Life: genes, embryos and evolution*

The evolution of plants on land was characterised by a complex series of morphological transformations that included a transition from a gametophyte-dominant life cycle to a diploid-dominant one. The marked differences between the bodies of modern angiosperm plants and the bodies of early land plants indicates that virtually every morphological structure that occurs in angiosperms had its origin sometime during the last 450 million years. Very little was known regarding the evolution of the molecular mechanisms that directed these changes until different genome sequencing projects revealed that most families of developmental regulators were present in lycophytes, mosses and even in unicellular chlorophyte algae. These surprising discoveries imply that a conserved set of proteins is responsible for the development of extremely divergent body plans.

How could such plasticity be achieved? One possibility is that ancient gene regulatory networks have been recruited multiple times during plant evolution to control the development of different structures. This was shown to be the case for the RSL class I bHLH transcription factors: they control the development of filamentous cells with a rooting function in mosses (and probably in early land plants), but were also recruited to control the development of root hairs in the sporophyte of vascular plants (Menand et al. 2007b). This suggests that an ancient regulatory mechanism was recruited from the gametophyte to the sporophyte generation during the explosion of sporophyte morphologies that occurred in the Devonian period. In *Arabidopsis*, RSL class I proteins regulate the transcription of a related group of bHLH transcription factors, the *RSL* class II genes. Since homologues of *RSL* class II genes are also present in

Physcomitrella, this suggested that the development of rooting cells in early land plants was also controlled by a network of *RSL* class I and *RSL* class II genes.

Testing this hypothesis requires a correct delineation of the phylogenetic relationships of bHLH proteins in land plants. The phylogenetic analyses presented in Chapter 2 demonstrate that bHLH proteins underwent a radiation in the streptophyte lineage. Furthermore, the existence of 20 bHLH subfamilies in the common ancestral of mosses and vascular plants, and at least six subfamilies in the common ancestor of Zygnematales and more derived streptophytes, indicates that there was a large radiation of the bHLH family in charophytes, before the appearance of the first land plants. This suggests that a main driver of the bHLH radiation was the evolution of multicellularity. Significantly, the bHLH families formed this early in plant evolution have been highly conserved ever since. Several small amino acid domains (some of which previously shown to be involved in protein-protein interactions) are conserved across the plant bHLH subfamilies. This suggests that many present-day bHLH interactions already occurred in early land plants and have been conserved in the major plant groups. Since these protein-protein interactions are central to the activity of complex regulatory networks, it is likely that these networks have been partially conserved for more than 450 million years.

Was a regulatory network involving *RSL* class I and class II genes present in early land plants? This hypothesis was explored in Chapter 3 and supported by the finding that the *Physcomitrella* loss-of-function *Pprs3 Pprsl4* double mutants and plants overexpressing *PpRSL4* show defects in the differentiation of caulonema cells. Despite a high degree of redundancy (which perhaps explains the lack of defects in rhizoid development) these phenotypes implicate these *RSL* class II genes in the control of the same developmental process as *RSL* class I genes: the chloronema to caulonema transition. Furthermore, *RSL* class I genes were found to modulate the response of *RSL* class II genes to auxin. *Physcomitrella RSL* class I and *RSL* class II genes can complement the respective homologues in *Arabidopsis*, indicating that the molecular function of these proteins is conserved, despite the more than 440 million years since they last shared an ancestral. These results suggests that an ancient differentiation mechanism involving auxin, *RSL* class I and *RSL* class II genes controlled the differentiation of rooting structures in the first land plants. During plant evolution, this regulatory network was recruited to the sporophyte generation to control the development of root hairs.

Does this imply that root hairs and rhizoids are homologous structures? The excerpt that opened this chapter was taken written a few years after the discovery that transgenic expression of the mouse transcription factor *Pax6* in *Drosophila* caused the development of ectopic eyes in the fly. *Pax6* controls the differentiation of vertebrate camera-type eyes in mouse, whereas its homologue in *Drosophila*, *eyeless*, had been found to control the development of its compound-type eyes (Halder et al. 1995). The functional conservation of Pax6/eyeless implied that the gene regulatory mechanisms driving the differentiation of these analogous eye types were homologous. In other words, that there was a clear dissociation between morphological and genetic homology.

There are now many examples in which the development of seemingly unrelated structures in distantly related species is controlled by homologous regulatory genes. Some of these now 'classic' examples include the discovery that the development of a variety of non-homologous animal appendages is controlled by Distal-less/Dlx homeoproteins (Panganiban et al. 1997), the development of the different types of eyes is controlled by an array of homologous transcription factors including Pax6 (reviewed in Vopalensky and Kozmik 2009), a late phase of *Hoxd* expression occurs in the distal fin buds of fishes and during digit specification in tetrapods (reviewed in Shubin et al. 2009) and the development of different types of hearts is governed by a set of homologous transcription factors in different animals (reviewed in Olson 2006). This has led to the proposal of the concept of 'deep homology': morphologically disparate organs whose formation (and evolution) depends on homologous genetic regulatory circuits (Shubin et al. 1997, 2009). Despite its powerful meaning, the concept of deep homology is actually quite subtle: it reflects the fact that homology depends on the hierarchical level which is being compared (for example, the wings of birds and bats are not homologous as wings, because wings evolved independently in the two lineages, but they are homologous as tetrapod forelimbs). Similarly, two structures may be analogous in a morphological context, but (deeply) homologous at a genetic level (Shubin et al. 1997; Scotland 2010).

The *RSL* regulatory network represents a clear example of deep homology. The root hair cells of *Arabidopsis* and the rhizoids/caulonema filaments of *Physcomitrella* are analogous structures with no structural correspondence. However, the *RSL* regulatory network that controls the development of these rooting structures is homologous, since it likely controlled the development of rooting structures in the common ancestral of mosses and vascular plants. There are other known cases of deep homologies in plants.

Two clear examples are the formation of boundary domains that delimit the leaflets of compound leaves being promoted by NAM/CUC3 genes in several eudicots, despite the multiple origins of compound leaves (Blein et al. 2008), and the KNOX-ARP mechanism that operates during both microphyll development in lycophytes and macrophyll development in angiosperms (Chapter 1; Harrison et al. 2005). Another example is the control of petal differentiation by *AP3/PI* genes, given the many independent origins of a differentiated perianth in angiosperms (Irish 2009); however, Scotland (2010) argues that the function of *AP3/PI* is not homologous, and therefore this example does not represent a true deep homology.

These examples point to a recurring reutilisation of ancient genetic networks to control the development of novel structures during evolution. This has been one of the major insights of the field of evolutionary developmental biology: form evolves largely by altering the expression of functionally conserved proteins, usually through mutations in the *cis*-regulatory regions of developmental regulatory genes and their target genes (Carroll et al. 2008). This is clearly exemplified in plants by the RSL network, an ancient and originally gametophytic mechanism that was later recruited to a different life cycle context (roots in the sporophyte generation), concomitantly with a change in the direction of the regulatory interactions between auxin, *RSL* class I and *RSL* class II genes.

Appendices

Appendix 1 – Primers sequences

qRT-PCR primers

PpRSL1
GTGTCCTATTCCGAGGACCA
CAACCTTGAGGTCCCAAAAA
PpRSL2
AGGATAAGTGAGCGGCTGAA
CCTTGGGCCAGATAGAATCA
PpRSL3
AACGAGCGCTTGAAGACATT
TTCAACAGCGTCACTTGGAG
PpRSL4
CGACTGATCCGCAGAGTGTA
GGTTACGATGTCCACCTTCC
PpRSL5
GCAACCGATCCTCAGAGTGT
TCAACCTTGGCTCCATTAGG
PpRSL6
AAATCTCGTGCCAAATGGAG
CATCCAGAACTCGTCGGATT
PpGAPDH
CTTGAGAAGCCTGCCTCCTA
TGCTGTCGGTAATGAAGTCG
PpEF1α
GGATCTTGTCGGGGTTGTAA
TTTCACCTTGGGAGTGAAGC

Construction of plasmids for knockouts

PpRSL4 – 5' homologous region
gcctctagaTCGTGGCTTTCTTTCAGGTG
ggctcgagTCTTCTCAACGGGTGCTTCA
PpRSL4 – 3' homologous region
cggactagtGGTACCGCCGAAATCTACCA
ccgacgcgtTCAGCAGTGCAAACTTTGGTT
PpRSL5 – 5' homologous region
cgccgtcgacATTCCATTTTCTGGCGCTTG
ccggatccTGCTGTCACATGAAATGAGTTG
PpRSL5 – 3' homologous region
gcctctagaAACCACGCCGCTCATAATTT
ccgacgcgtTGTTGCCTCTTGTCGTGTGT
PpRSL6 – 5' homologous region
cttGGTAACGTGGACAGCTCGAT
gacGCTACTCTGCGGTTAGTCAGG
PpRSL6 – 3' homologous region
cgtGTGTAACAGCCCCACCAG
catCAACCAAAATGTATTTCAATGG

Genotyping *Physcomitrella* knockouts

Pprsl3
p1   AGTCGCCTTCCTCTCCTCTC
p2   CGGTGAGTTCAGGCTTTTTC
p3   TCCGAGGGCAAAGAAATAGA
p4   TCAGTTGCCTTCTTGTGTGC
p5   CGACTGATCCGCAGAGTGTA
p6   AATGTCTTCAAGCGCTCGTT

Pprsl4
p1   ATGGATGGCTGAGGTGTTGT
p2   TTGCTTTGAAGACGTGGTTG
p3   CTTCGACGGATCTCGACCT

p4   GGCATGAGCTACCAAAAGGTG
p5   TCTTCGGGGATCTAGCTGTC
p6   TGGCGTACATCCAATACTCG

Pprsl5
p1   CAAAGAATCTGAACGGCTCAA
p2   GGTGGAGCTCGGTACCATAA
p3   CCGGCCAGATCTATAACTTCG
p4   GTGGAGCTAGCCGCAGATG
p5   TCATGCATCGAAACCTCGTC
p6   TCTCCTCAAGTTCAAGAGGGTGT

Pprsl6
p1   ACAGCTTCGGCCTTTCACTA
p2   CGTGGGATCCTCTAGAGTCG
p3   GCCGGCCAGATCTATAACTTC
p4   TTTTATCTCCGATTCTTATGTCTAAGT
p5   GCGGTCCTACTTCCATTCTG
p6   ATGCCGTTGTAGTTGTGTGG

Construction of DIG probes

Hyg probe
ATCCGGTCGGCATCTACTCT
TGTAGGAGGGCGTGGATATG

NptII probe
TGAATGAACTGCAGGACGAG
AATATCACGGGTAGCCAACG

Zeo probe
GACTAAACCTGGAGCCCAGAC
GAACTAGTGGATCCCCGTCA

Cloning coding sequences

PpRSL3
cccgggATGACCGATCTAAATTCGAGC
ctcgagCTACGCACTGGACTGCAGTCTCT

PpRSL4
ggtaccATGACCGATCTGATTTCGATCT
gtcgacTCAGTTGTTCTCGCCGGGA

PpRSL5
ggatccATGGTGCAGTTATACATGTCCTC
gtcgacTCACGCACGTGACTCCAG

PpRSL6
ggtaccATGGTGCGGTTTAACTACATG
ctgcagCTAAATATTCGACTCCAGCTCCG

Amplification of pGWB2 fragment
ccgcggccgcGTTGAATGTCGCCCTTTTGT
gcactagtCGGAAATTCCTCTCCTGTCA

Construction of amiRNA precursors

amiRNAa
I   GATCGTCTAGCATGGTTACTATGTCTCTCTTTTGTATTCC
II   GACATAGTAACCATGCTAGACGATCAAAGAGAATCAATGA
III   GACACAGTAACCATGGTAGACGTTCACAGGTCGTGATATG
IV   GAACGTCTACCATGGTTACTGTGTCTACATATATATTCCT

amiRNAb
I   GATCGTCTAGCATGGTTACAATGTCTCTCTTTTGTATTCC
II   GACATTGTAACCATGCTAGACGATCAAAGAGAATCAATGA
III   GACACTGTAACCATGGTAGACGTTCACAGGTCGTGATATG
IV   GAACGTCTACCATGGTTACAGTGTCTACATATATATTCCT

Appendix 2 – *RSL* genes identified in different plant species

Species	Name	RSL Class	Retrieved from	Name in original database	bHLH number[f]
Arabidopsis lyrata	AlRSL1	I	Phytozome 5.0	jgi|Araly1|894298|scaffold_201372.1	
	AlRSL2	I		jgi|Araly1|493759|fgenesh2_kg.7__3192__AT5G37800.1	
	AlRSL3	II		jgi|Araly1|480366|fgenesh2_kg.3__3067__AT2G14760.1	
	AlRSL4	II		jgi|Araly1|472979|fgenesh2_kg.1__2933__AT1G27740.1	
	AlRSL5	II		jgi|Araly1|494687|fgenesh2_kg.8__503__AT5G43175.1	
	AlRSL6	II		jgi|Araly1|491291|fgenesh2_kg.7__724__AT4G33880.1	
	AlRSL7	II		jgi|Araly1|855794|Al_scaffold_0797_1	
Arabidopsis thaliana	AtIND	not RSL[d]	The Arabidopsis Information Resource (TAIR)	AT4G00120	AtbHLH040
	AtRHD6	I		AT1G66470	AtbHLH083
	AtRSL1	I		AT5G37800	AtbHLH086
	AtRSL2	II		AT4G33880	AtbHLH085
	AtRSL3	II		AT2G14760	AtbHLH084
	AtRSL4	II		AT1G27740	AtbHLH054
	AtRSL5	II		AT5G43175	AtbHLH139
Brachypodium distachyon	BdRSL1	I	DOE Joint Genome Institute	Bradi2g01000.1 chr02_pseudomolecule brac version0 634132-632405 BestGuessCds	
	BdRSL2	I		Bradi3g53060.1 chr03_pseudomolecule brac version0 53840035-53841056 BestGuessCds	
	BdRSL3	I		Bradi1g42440.1 chr01_pseudomolecule brac version0 39524292-39525456 BestGuessCds	
	BdRSL4	II		Bradi1g14600.1 chr01_pseudomolecule brac version0 11561477-11562837 BestGuessCds	
	BdRSL5	II		Bradi1g22960.1 chr01_pseudomolecule brac version0 18442319-18443680 BestGuessCds	
	BdRSL6	II		Bradi1g70860.1 chr01_pseudomolecule brac version0 69006335-69007839 BestGuessCds	
	BdRSL7	II		Bradi4g03070.1 chr04_pseudomolecule brac version0 2417289-2418632 BestGuessCds	
	BdRSL8	II		*not annotated* [e]	
Saccharum sp.	CA145633	I	NCBI	SCSGRT2066E06.g RT2 Saccharum hybrid cultivar SP80-3280 cDNA clone	
Triticum aestivum	CA654295	I	NCBI	wre1n.pk164.d12 wre1n Triticum aestivum cDNA clone wre1n.pk164.d12	
Cucumis sativus	CsRSL1	I	Phytozome 5.0	Cucsa.006340.1	
	CsRSL2	II		Cucsa.097490.1	
	CsRSL3	II		Cucsa.219180.1	
	CsRSL4	II		Cucsa.352750.1	
Hordeum vulgare	EX574671	II	NCBI	HDP18P23w HDP Hordeum vulgare subsp. vulgare cDNA clone HDP18P23,	
Panicum virgatum	FL939126	I	NCBI	CCGP9527.g1 CCGP Panicum virgatum root (L) Panicum virgatum cDNA	
	FL998279	II		CCHY25227.b1 CCHY Panicum virgatum callus (N) Panicum virgatum cDNA	
Glycine max	GmRSL1	I	Phytozome 5.0	Glyma02g41370.1	
	GmRSL2	I		Glyma18g04420.1	
	GmRSL3	I		Glyma11g33840.1	

	GmRSL4	I		Glyma14g07590.1	
	GmRSL5	II		Glyma04g04800.1	
	GmRSL6	II		Glyma06g04880.1	
	GmRSL7	II		Glyma10g40360.1	
	GmRSL8	II		Glyma14g09770.1	
	GmRSL9	II		Glyma17g35420.1	
	GmRSL10	II		Glyma20g26980.1	
Manihot esculenta	**MeRSL1**	I	Phytozome 5.0	cassava10199.m1	
	MeRSL2	II		cassava38950.m1	
	MeRSL3	II		cassava25589.m1	
	MeRSL4	II		cassava20440.m1	
	MeRSL5	II		cassava14986.m1	
Mimulus guttatus	**MgRSL1**	I	DOE Joint Genome Institute	mgf025678m	
	MgRSL2	II		mgf019332m	
	MgRSL3	II		mgf024863m	
	MgRSL4	II		mgf018282m	
	MgRSL5	II[c]		mgf022238m	
Oryza sativa	**OsLAX**	not RSL[d]	Rice Genome Annotation Project	LOC_Os01g61480	OsbHLH123
	OsRSL1	I		LOC_Os01g02110	OsbHLH125
	OsRSL2	I		LOC_Os02g48060	OsbHLH126
	OsRSL3	I		LOC_Os06g30090	OsbHLH127
	OsRSL4	II		LOC_Os03g10770	OsbHLH129
	OsRSL5	II		LOC_Os03g42100	OsbHLH131
	OsRSL6	II		LOC_Os07g39940	OsbHLH128
	OsRSL7	II		LOC_Os11g41640	OsbHLH132
	OsRSL8	II		LOC_Os12g32400	OsbHLH133
	OsRSL9	II		LOC_Os12g39850	OsbHLH130
	OsRSL10	II[c]		LOC_Os03g55550	OsbHLH134
Physcomitrella patens	**PpRSL1**[a]	I	DOE Joint Genome Institute	jgi|Phypa1_1|167487|estExt_fgenesh1_pg.C_1470079	PpbHLH043
	PpRSL2[a]	I		jgi|Phypa1_1|165193|estExt_fgenesh1_pg.C_840011	PpbHLH033
	PpRSL3	II		jgi|Phypa1_1|163809|estExt_fgenesh1_pg.C_600113	PpbHLH028
	PpRSL4[b]	II		jgi|Phypa1_1|65387|fgenesh1_pg.scaffold_6000069	PpbHLH083
	PpRSL5	II		*not annotated* [e]	
	PpRSL6	II		jgi|Phypa1_1|88041|fgenesh1_pg.scaffold_164000047	PpbHLH095
	PpRSL7	II		jgi|Phypa1_1|231841|fgenesh2_pg.scaffold_140000014	PpbHLH063
Populus trichocarpa	**PtRSL1**	I	DOE Joint Genome Institute	jgi|Poptr1_1|756089|fgenesh4_pg.C_LG_II002269	
	PtRSL2	I		jgi|Poptr1_1|759048|fgenesh4_pg.C_LG_IV000676	
	PtRSL3	II		jgi|Poptr1_1|754916|fgenesh4_pg.C_LG_II001096	

	PtRSL4	II		jgi|Poptr1_1|767688|fgenesh4_pg.C_LG_IX000739	
	PtRSL5	II		jgi|Poptr1_1|798077|fgenesh4_pm.C_LG_I001080	
Sorghum bicolor	**SbRSL1**	I	DOE Joint Genome Institute	jgi|Sorbi1|5035419|Sb03g008290	
	SbRSL2	I		jgi|Sorbi1|5038869|Sb04g030230	
	SbRSL3	II		jgi|Sorbi1|5028258|Sb01g007160	
	SbRSL4	II		jgi|Sorbi1|5028668|Sb01g014580	
	SbRSL5	II		jgi|Sorbi1|5034588|Sb02g037990	
	SbRSL6	II		jgi|Sorbi1|5040921|Sb05g025230	
	SbRSL7	II		jgi|Sorbi1|5045388|Sb08g019780	
Selaginella moellendorffii	**SmRSL1**	I	DOE Joint Genome Institute	jgi|Selmo1|413232|fgenesh2_pg.C_scaffold_20000093	SmbHLH003
	SmRSL2	I		jgi|Selmo1|426794|fgenesh2_pg.C_scaffold_86000063	SmbHLH004
	SmRSL3	I		jgi|Selmo1|409967|fgenesh2_pg.C_scaffold_11000350	SmbHLH005
	SmRSL4	I		jgi|Selmo1|414414|fgenesh2_pg.C_scaffold_23000278	SmbHLH006
	SmRSL5	I		jgi|Selmo1|418491|fgenesh2_pg.C_scaffold_38000089	SmbHLH101
	SmRSL6	I		jgi|Selmo1|423809|fgenesh2_pg.C_scaffold_63000123	SmbHLH103
	SmRSL7	II		jgi|Selmo1|404767|fgenesh2_pg.C_scaffold_2000755	SmbHLH018
	SmRSL8	II		jgi|Selmo1|419388|fgenesh2_pg.C_scaffold_42000088	SmbHLH042
Zea mays	**ZmRSL1**	I	Maize Genome Sequencing Project	AC216731.3_FGT001_seq=cdna;_coord=5:201012621..201013822:1;_parent_gene=AC216731.3_FG001	
	ZmRSL2	I		GRMZM2G066057_T01_seq=cdna;_coord=8:8532281..8534158:-1;_parent_gene=GRMZM2G066057	
	ZmRSL3	II		GRMZM2G057260_T01_seq=cdna;_coord=7:156447257..156452969:-1;_parent_gene=GRMZM2G057260	
	ZmRSL4	II		GRMZM2G383841_T01_seq=cdna;_coord=2:205690389..205691689:-1;_parent_gene=GRMZM2G383841	
	ZmRSL5	II		AC198518.3_FGT005_seq=cdna;_coord=3:87821662..87823007:1;_parent_gene=AC198518.3_FG005	

[a] The JGI annotated version of PpRSL1 and PpRSL2 have a longer N-terminus than the published sequences (Menand et al. 2007b).

[b] The JGI annotated version of PpRSL4 has an incorrect splicing prediction in the C-terminus.

[c] The sequences of MgRSL5 and OsRSL10 appear to have incorrect splicing predictions. They were not used in the alignments and phylogenetic analyses.

[d] AtIND and OxLAX were used as outgroups in the phylogenetic analyses.

[e] These sequences are not annotated. They were found by a BLAST search on the available genomic sequences.

[f] See Chapter 2.

Appendix 3 – *Physcomitrella* RSL class II proteins

```
               10        20        30        40        50        60        70        80
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
PpRSL3  ----------------MTDLNSSLESPGSSVEETP--AAANSVATSCEAMMWEAWSTQPSTGDEASTSKLD--LLPELV
PpRSL4  ----------------MTDLISILESSGSSREEMCPVAVPSSVASSCERLIWEGWTAQPSPVEESTTSKLLPKLLPELE
PpRSL5  MVQL-YMSSVEEQRETMVQPYVSSMDSGSTSGRQTPSCVVQQGSNTFETSNLWEEWTQASNGDDTVSTSNFL----PEIS
PpRSL6  MVRFNYMYPVQEQLEAMTDQHTPSMDSVSSAGEKTSSCIVQQGGNASETSNLWEEWTQGSNGDDSVSTSNFL----PELN

               90       100       110       120       130       140       150       160
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
PpRSL3  SSSNSRLSFQQSDLLSNMLSSFHPLSQHSSAGFELSHNRGGSEHSPEFLQEGSSEAD---TETSSFGDLYINRQTSRNSF
PpRSL4  TSSYSALTLQQPDALSSILSVLHPFSHYSSASLELARNPDWSLKSSNPLRESSSEAG---IRTSSFEGLYSGQHTTKKIH
PpRSL5  SFTSSRLSFQQSDSLTTWMSGFPPLSQ-TALSPDLSHSSDPVDHPPAFMQEGLGPGDSILDYSPALTEMYP-KSSSKHNS
PpRSL6  SSTSSRLAFHQSDILSTWISGYHPLSQ-SSLSSEFSHTSDRENHPPAFMQEGLIPSGLILDSDPALTDIYT-RSSS----

              170       180       190       200       210       220       230       240
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
PpRSL3  LGSIACPLPSNHSDSGKNIRREDLSNQLGAKSSAPLQLWQSLGPESPSSPLAYHNIGYRHSHGEKWETGSQSHSPWPTVS
PpRSL4  LGVIPYHLSEDQRQCAVSPP-ENECRLLSANSSGSLHWWHSIGPESPSSTLAFHNIGIQHSTFEKCEPRGQSHSSWPAAS
PpRSL5  SDCLPYPAASAPDKKMTDHELGSAISLAYDRGTVSRQLLRALGPLSPSSPLALQN-GLQNPLGDPWDASP-SAMPWPMAT
PpRSL6  SDSLPYPTARIMDKALTDHELESAVPLAYEKGCVPPQVLRNLGPLSPSSPLAFQN-GLLNPLRDPWDSCP-SALPWSNVT

              250       260       270       280       290       300       310       320
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
PpRSL3  NTSSTIQLLGGRAAENEVIQVLKSNDSEISKSLATLQQYGDHGRQLNLN-HSSTTNHPEVIYSSKFGPKPSASSHTDVLM
PpRSL4  GTSPTVQYFHAHSADNEGVEVVKQDDSQISKALATYQPHGDHSLVLNSDRIASTTSHSEDPCGPKPGRRPAASYDTEMIL
PpRSL5  TG-HAYGPGATRTSIPDHLANAINHLEGIAPSSASH---------------ASKPRHTDIFIAPNGTFDSTPGGWTPQYY
PpRSL6  TASQTYGQVTTRTFIPDHSASAIDKLEAVATITAGYG--------------ASKPQHTDVFIEPNGTFQSTPAGWAPQFY

              330       340       350       360       370       380       390       400
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
PpRSL3  SSTNSSFLSIPTAWSTPEYSMSGPSTRSEQFMNFVRIAQEQNSVPISGPSPILGSYVG-CSNRSKSGISRVVSQETPTAK
PpRSL4  SPS-ESFLTTPNMLSTLECVISGASNISDQYMNFVREPQEQRLSSISDLSLIPDSHADPHSIGFISGTFRTDSHGTGIRK
PpRSL5  DGSVTTDESVKAMKLIASLREAG---HAEATIGFCTESKPSFLRGGDRTTSPVDSFFG-KCVGAKTSIKQACSGKHPLEL
PpRSL6  DGSEATGLLVKPMRAIASLGEAG---CGEATSEFCTKTKPGLLKGGDTITSPVGSLLG-DCKKAESSMKQVWPGKHRLEL

              410       420       430       440       450       460       470       480
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
PpRSL3  NRLLACEGSSGPAPKRPS-YAAHSDSHADQAAAMWSSPNLRRSSFPSIL---TSQAMEIYAIGPALNTGKPRARRGSAT
PpRSL4  NRIFLSDEESDFLPKKRSKYTVRGDFQMDRFDAVWGNTGLRGSSCPGNS---VSQMMAIYEFGPALNRNGRPRVQRGSAT
PpRSL5  EEIVDSENSELNPTQLKRSKLFENHPNALWSDQSMNGRELRSYSHLVGSSLTASQPMDIIAIGPALNTDGKPRAKRGSAT
PpRSL6  VELVDGEDTKSSPTQLKRPKHSTDYANVLLSDHILKGAELRSYFHSGDVGLNASQAMDIIVIGPALNTNGKPRAKRGSAT

              490       500       510       520       530       540       550       560
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
PpRSL3  DPQSVYARHRREKINERLKTLQHLVPNGAKVDIVTMLDEAIHYVQFLQLQVTLLKSDEYWMYATPNTYKG-IDLTNSPPQ
PpRSL4  DPQSVHARARREKIAERLRKLQHLIPNGGKVDIVTMLDEAVHYVQFLKRQVTLLKSDEYWMYATPTSYRSKFDDCSLVPG
PpRSL5  DPQSVYARHRREKINERLKSLQNLVPNGAKVDIVTMLDEAIHYVKFLQNQVELLKSDELWIYATPNKYNG-MDISDLSDM
PpRSL6  DPQSVYARHRREKINERLKNLQNLVPNGAKVDIVTMLDEAIHYVKFLQTQVELLKSDEFWMFANPHNYNG-IDISDPSSM

              570
        ....|....|
PpRSL3  TQRLQSSA*-
PpRSL4  ENN*------
PpRSL5  YLQELESRA*
PpRSL6  HSPELESNI*
```

Appendix 4 – Alignment of the C-terminus of RSL class I proteins

```
                 10        20        30        40        50        60        70        80        90       100       110       120       130       140
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
AlRSL1    SPKSAGNKRSQT-GESTQ---------PS-KKPNSGVTGKA-KPK-----------PTTSPKD-PQSLAAKNRRERISERLKILQELVPNGTKVDLVTMLEKAISYVKFLQVQVKVLATDEFWPAQGGKAPDISQVKDAIDAILSSS
AlRSL2    SPKLAGNKRPFT-GDNTH---------LS-KKPSSGTNGEA-KPK-----------ATTSPKD-PQSLAAKNRRERISERLKVLQELVPNGTKVDLVTMLEKAIGYVKFLQVQVKVLAADEFWPAQGGKAPDISQVKEAIDAILSST
AtRHD6    SPKSAGNKRSHT-GESTQ---------PS-KKLSSGVTGKT-KPK-----------PTTSPKD-PQSLAAKNRRERISERLKILQELVPNGTKVDLVTMLEKAISYVKFLQVQVKVLATDEFWPAQGGKAPDISQVKDAIDAILSSS
AtRSL1    SPKLAGNKRPFT-GENTQ---------LS-KKPSSGTNGKI-KPK-----------ATTSPKD-PQSLAAKNRRERISERLKVLQELVPNGTKVDLVTMLEKAIGYVKFLQVQVKVLAADEFWPAQGGKAPDISQVKEAIDAILSSS
BdRSL1    PPAKMAQKRACQGGETQAAA---------KKQCGGSKK-S-KAK-----------AAPAKD-PQSAVAKVRRERISERLKVLQDLVPNGTKVDMVTMLEKAITYVKFLQLQVKVLATDDFWPVQGGKAPELSQVKDALDAILSSQ
BdRSL2    RPFRAPVPERVRDAEPR-AS----------KKQCGASRKTTAKAKSPAPA-------ITSPKD-PQSLAAKNRREKISERLRTLQEMVPNGTKVDMVTMLEKAISYVKFPLQVRKVLATDEFWPAQGGMAPEISQVKEALDAILSSQ
BdRSL3    ATVPASYKRPRAHVQPQQEAEEQESITPNPKKQCGDGKVVI-KSSAAATG------TSPRKE-PQSQAAKSRRERIGERLRALQELVPNGSKVDMVTMLDKAITYVKFMQLQLTVLETDAFWPAQGGAAPEISQVKAALDAIILSS
CA145633  --------------------------------GGARK-S-KANAA----------PAPTKD-PHMVTAKVRREKIAEKLKVLQDLVPNGTKVDLVTMLEKAITYVKFLQLQVKVLADEFWPAQGGKAPELSQVKDALDAILSSQ
CA654295  ----------------------------------------------------------VRRERISERLKVLQDLVPNGTKVDLVTMLEKAITYVKFLQLQVKVLATDEFWPVQGGKAPELSQVKTALDAILSSQ
CsRSL1    QEIVPNHKRSHTTGESSGSVC---------KKQCTAAPKKQ-KPK-----------SATAKD-PQSIAAKNRRERISERLKILQELVPNGRKVDLVTMLEKAISYVKFLQLQVKILATDEFWPVQGGKAPDISQVKEAIDVILSSQ
FL939126  APVRASQKRTYVGVESPAAV-------SP-KKHCGAGRKATSKAKSA---------PTVPTKD-PQSLAAKNRRERISERLRTLQELVPNGTKVDLVTMLEKAISYVKFLQLQVKVLATDEFWPAQGGKAPEISQVREALDAILSSA
GmRSL1    AQELVLTKRSSM-GENMQAT--------NAKKPCTSASKAA-KPK-----------LNPFKD-PQSVAAKNRRERISERLKILQELVPNGSKVDLVTMLEKAISYVKFLQLQVKVLAADEFWPVQGGKAPDISQVRQAIDAILSSQ
GmRSL2    AQESVLQKRPSM-GENMKAA---------KKQCSTESKTP-KHK-----------SSPSKD-PQSVAAKNRRERISERLKILQELVPNGSKVDLVTMLEKAISYVKFLQLQVKVLATDEFWPVQGGKPPEISQVKEVIDVILSSQ
GmRSL3    AQESVLQKRPFM-GESMKAA---------KKQCSIESKTT-KHN-----------SSPSKD-PQSVAAKNRRERISERLKILQELVPNGSKVDLVTMLEKAISYVKFLQLQVKVLATDEFWPVQGGKPPDISQVKEVIDTILSSQ
GmRSL4    AQELVLKKRSFM-GENMQ-V-------TNAKKPCTSASKAA-KPK-----------SNPSQD-PQSVAAKNRRERISERLKILQELVPNGSKVDLVTMLEKAISYVKFLQLQVKVLAADEFWPVQGGKAPDISQVRQAIDAILSSQ
MeRSL1    TQDVNFHKRPNL-GESMQAL---------KKPCNGATRKP-KPK-----------SSPSKD-PQSLAAKNRRERISERLKVLQDLVPNGTKVDLVTMLEKAITYVKFLQLQVKVLAADEFWPAQGGKAPDISQVKEAIDAILSSQ
MgRSL1    SPLSLYH-----GENVQAM---------KKQCVGSSSDSKKP-----------------D-PQSLAAKNRRERISERLKILQELVPNGTKVDLVTMLEKAITYVKFLQLQVKVLAADEFWPAQGGKAPDLSQVREAIDAILASQ
OsRSL1    PPPPAAKKRACPSGEARAAG---------KKQCRKGSKPN-KAASASSPSPSPSPSPSPSPNKEQPQSAAAKVRRERISERLKVLQDLVPNGTKVDLVTMLEKAISYVKFLQLQVKVLATDEFWPAQGGKAPELSQVKDALDAILSSQ
OsRSL2    PAPRGSQKRAH--AESSQAM------SPS-KKQCGAGRKAG-KAKSA---------PTTPTKD-PQSLAAKNRRERISERLRILQELVPNGTKVDLVTMLEKAISYVKFLQLQVKVLATDEFWPAQGGKAPEISQVKEALDAILSSS
OsRSL3    AAASGSQRRARPPPSPLQGSELHEY---S-KKQ----RANN--------------------KE-TQSSAAKSRRERISERLRALQELVPSGGKVDMVTMLDRAISYVKFMQMQLRVLETDAFWPASDGATPDISRVKDALDAIILSS
PpRSL1    ----------------------------------NLKPRAR---------QGSAND-PQSIAARVRRERISERLKVLQALIPNGDKVDMVTMLEKAISYVQCLEFQIKMLKNDSLWPKALGPLPNTLQELLELAGPEFAG
PpRSL2    ---------------------------GRALGPALNTNLKPRAR---------QGSAND-PQSIAARVRRERISERLKVLQALIPNGDKVDMVTMLEKAITYVQCLELQIKMLKNDSIWPKALGPLPNTLQELLELAGPEFSG
PtRSL1    PDASSFHKRPNM-GESMQAL---------KKQRDSATKKP-KPKSA----------GPAKD-PQSIAAKNRRERISERLKMLQDLVPNGSKVDLVTMLEKAISYVKFLQLQVKVLATDEFWPVQGGKAPDISQVKGAIDATLSSQ
PtRSL2    PDASSFHKRPNM-GESMQAL---------KKQCNNATKKP-KPKSA---------AGPAKD-LQSIAAKNRRERISERLKVLQDLVPNGSKVDLVTMLEKAISYVKFLQLQVKVLATDELWPVQGGKAPDISQVKEAIDALLSSQ
SbRSL1    LPAKPPHKRARRDGE-VQAAAA--------KKQCGGGARKS-KAKAA---------PAPTKD-PQSVAAKVRREKIAEKLKVLQDLVPNGTKVDLVTMLEKAITYVKFLQLQVKVLAADEFWPAQGGKAPELSQVKDALDAILSSQ
SbRSL2    PVRRAPQKRTYVSAEPQ-AV-------SP-KKHCGAGRKAS-KAKSP--------STTPTKD-PQSLAAKNRRERISERLRTLQELVPNGTKVDLVTMLEKAISYVKFLQLQVKVLATDEFWPAQGGKAPEISQVREALDAILSSA
SmRSL1    -------------------------KAPRVPALNTNFKPRAR---------QGSAND-PQSIAARHRRERISDRLKILQELVPNSTKVDLVTMLEKAINYVKFLQLQVKVLTSDDYWP-SGATWQNSSKADTAL-------
SmRSL2    -------------------------KAPRVPALNTNFKPRAR---------QGSAND-PQSIAARHRRERISDRLKILQELVPNSTKVDLVTMLEKAINYVKFLQLQVKVGAS-------------------------
SmRSL3    -------------------------SHKREPALNTNLKPRAK---------QGCAND-PQSIAARQRRERISDRLKILQELIPNGSKVDLVTMLEKAINYVKFLQLQVKVLMNDEYWPPKGDGEEDYPVSQKYLFLSKIS-
SmRSL4    -------------------------SHKREPALNTNLKPRAK---------QGCAND-PQSIAARQRRERISDRLKILQELIPNGSKVDLVTMLEKAINYVKFLQLQVKVLMNDEYWPPKGDGEEDYPMSQKYLFMSKIS-
SmRSL5    ------------------------CTTALNTNLKPRSR---------QGTAND-PQSIAARQRRERISQRLKILQDLVPNGSKVDLVTMLEKAINYVKFMQLQLQASVVD-----------------------
SmRSL6    ------------------------CTTALNTNLKPRSR---------QGTAND-PQSIAARQRRERISQRLKILQDLVPNGSKVDLVTMLEKAINYVKFMQLQLQASVVD-----------------------
ZmRSL1    APVRVPQKRTYLSAEPQ-AV------SPN-KKHCGAGRKAS-KAKLA--------STAPTKD-PQSLAAKNRRERISERLRALQELVPNGTKVDLVTMLEKAISYVKFLQLQVKVLATDEFWPAQGGKAPEISQVREALDAILSSA
ZmRSL2    PAKAPPHKRARRDGDQVQAAAAA-------KKQCGGVGARM-KSKQAKLAA------PAPTKD-PQSVAAKVRREKIAEKLKVLQDLVPNGTKVDLVTMLEKAITYVKFLQLQVKVLAADEFWPAQGGKAPDLSQVKDALDAILSSS
AtRSL2    ------------------------PSKALNLNGKTRAS---------RGAATD-PQSLYARKRRERINERLRILQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPIAFNGMDIGLSSPR------
AtRSL4    ------------------------KASVTSVKGKTRAT----------KGTATD-PQSLYARKRREKINERLKTLQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPLAYNGLDMGFHHNL------
AtIND     -------------------PATVPKPNRRN---------VRISDD-PQTVVARRRRERISEKIRILKRIVPGGAKMDTASMLDEAIRYTKFLKRQVRILQ-------------------------
OsLAX     ------------------RMRGGGRRRPG---------AKLSTD-PQSVAARERRHRISDRFRVLRSLVPGGSKMDTVSMLEQAIHYVKFLKAQVTLHQ------------------------
```

(previous page)

Alignment used for the phylogenetic analysis of RSL class I proteins.

Conserved amino acid regions of 34 RSL class I proteins were manually aligned. The RSL class II proteins AtRSL2 and AtRSL4 and the bHLH subfamily VIIIb proteins AtIND and OsLAX were used as outgroups. The sequence logo shown below the alignment was generated using WebLogo; the red line indicates the location of the bHLH domain. Sequence names are indicated in Appendix 2; a FASTA format version of this alignment is given in Supplementary File 9.

Appendix 5 – Alignment of the C-terminus of RSL class II proteins


```
                    10        20        30        40        50        60        70        80        90
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AlRSL3    ALNLNGKTRASRGAATD-PQSLYAR-KRRERINERLRILQHLVPNGTKVHISTMLEEAVQYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-LDL
AlRSL4    VTSVKGKTRATKGTATD-PQSLYAR-KRREKINERLKTLQNLVPNGTKVDISTMLEEAHYVKFLQLQIKLLSSDDLWMYAPLAYNGLDMG-FHH
AlRSL5    SKSLKRKAKSNKGIASD-PQSLYAR-KRRERINDRLKTLQSLVPNGTKVDISTMLEDAVHYVKFLQLQIKLLSSDDLWMYALLAHNGLNMG-LHH
AlRSL6    ALNLNGKTRASRGAATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPIAFNGMDIG-LSS
AlRSL7    SKSLKRKAKSNKGIASD-PQSLYAR-KRRERINDRLKTLQSLVPNGTKVDISTMLEDAVHYVKFLQLQIKLLSSDDLWMYALLAHNGLNMG-LHH
AtRSL2    ALNLNGKTRASRGAATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPIAFNGMDIG-LSS
AtRSL3    ALNLNGKTRASRGAATD-PQSLYAR-KRRERINERLRILQHLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-LDL
AtRSL4    VTSVKGKTRATKGTATD-PQSLYAR-KRREKINERLKTLQNLVPNGTKVDISTMLEEAHYVKFLQLQIKLLSSDDLWMYAPLAYNGLDMG-FHH
AtRSL5    SKSLKRKAKANRGIASD-PQSLYAR-KRRERINDRLKTLQSLVPNGTKVDISTMLEDAVHYVKFLQLQIKLLSSEDLWMYAPLAHNGLNMG-LHH
BdRSL4    SPVAANPSGKGRQSTTD-PQSLYAR-KRRERINERLKVLQNLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDDMWMYAPIAYNGMNIG-VDL
BdRSL5    SCSKKTGTRASRGAATD-PQSLYAR-KRRERINERLKTLQNLVPNGTKVDISTMLEEAVEYVKFMQLQIKLLSSDDTWMYAPLAYNGMNVGSFDL
BdRSL6    SSRGGAKARAGRGAATD-PQSLYAR-KRREKINERLKVLQNLVPNGTKVDISTMLEEAVHYIKFMQLQIKLLSSDDMWMFAPIAYNGFNVG-LDL
BdRSL7    GVCPKGKARAARGASTD-PQSLYARQKRRERINERLKTLQTLVPNGTKVDMSTMLEEAVHYVKFLQLQIKVLSSDDMWMYAPLAYNGMNIG-LDL
BdRSL8    VANQSDNTSGCKRPSKN-MQSLYAK-KRRERINEKLRVLQQLIPNGTKVDISTMLEEAVQYVKFLQLQIKVLSSDETWMYAPLAYNGMDIG-LTL
CsRSL2    SSTSNGKPRASRGAATD-PQSLYAR-KRRERINERLRILQSLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-LNL
CsRSL3    SLNGANKSRASRGSATD-PQSLYAR-KRRERINERLRILQTLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-LNP
CsRSL4    NASPKPKTRATRGSATD-PQSLYAR-KRRERINERLRILQKLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-LQQ
EX574671  GVTPKGKARAGRGAATD-PQSLYAR-KRRERINERLKTLQTLVPNGTKVDMSTMLEEAVQYVKFLQLQIKVLSSDEMWMYAPIAYNGMNIG-LDL
FL998279  -------------SSKN-SQSLYAK-KRRERINERLKTLQQLIPNGTKVDMSTMLEEAVQYVKFLQLQIKLLSSDETWMYAPQPLEH--------
GmRSL10   PPNLDRKSRATTSAAAD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSEDLWMYAPIVYNGINIG-LDL
GmRSL5    ALNFKGKTKASKGSATD-PQSLYAR-KRRERIDDRLRILQNLVPNGTKVDISTMLEEAVQYVKFLQLQNKLLSSDDLWMYAPIAYNGLDLV-LNL
GmRSL6    -------TKVSRGSATD-PQSLYAR-KRRLRINERLRILQNLVPNGTKVDRSSMLEEAVQYMKFLQLQIKLLSSDDLWMYAPIAYNGLDL-----
GmRSL7    PPNLHRKSRATTGAATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDDLWMYAPIAYNGINIG-LDL
GmRSL8    ALNSNGKTRASRGSATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVNYVKFLQLQIKLLSSDDLWMYAPLAYNGLDIG-LNL
GmRSL9    ALNSNGKTRASRGSATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVNYVKFLQLQIKLLSSDDLWMYAPFAHNGLDIG-LNL
MeRSL2    TLNSNGKTRASRGSATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-LNQ
MeRSL3    AQKSNGKARTSRGSATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIS-KLI
MeRSL4    ALNLNGRTRASRGSATD-PQSLYAR-KRRERINERLKILQNLVPNGTKVDISTMLEEAVNYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-LDL
MeRSL5    ILNSNGRTRASRGAATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-LT-
MgRSL2    KSASNGKAKVGKGTATD-PQSLYAR-KRREKINERLRILQNLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDEMWMYAPIAYNGMDMG-LYQ
MgRSL3    ALNLDGKTRASRGSATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVEYVKFLQLQIKLLSSDDLWMYSPIAYNGMDIG-LDL
MgRSL4    LKSLNGKAKANRGSATD-PQSLYAR-KRRERINERLKILQNLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-I--
OsRSL4    TGHGGAKARAGRGAATD-PQSLYAR-KRRERINERLKILQNLIPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDMWMFAPIAYNGVNVG-LDL
OsRSL5    AAKSNGKAQSGHRSATD-PQSLYAR-KRRERINERLKILQNLVPNGTKVDISTMLEEAMHYVKFLQLQIKLLSSDEMWMYAPIAYNGMNIG-IDL
OsRSL6    PSSKKMGTRANRGAATD-PQSLYAR-KRRERINERLKILQNLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDDTWMYAPIAYNGVNISNIDL
OsRSL7    VLKQSDNSRGHKQCSKD-TQSLYAR-KRRERINERLRILQQLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDDTWMFAPLAYNGMNMD-LGH
OsRSL8    RTKMSVSKACKHSVSAESSQSYYAK-NRRQRINERLKILQELIPNGTKVDISTMLEEAIQYVKFLHLQIKLLSSDEMWMYAPLAFDSGNNR-LYQ
OsRSL9    GATSKGKSRAGRGAATD-PQSLYAR-KRRERINERLKTLQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDEMWMYAPIAYNGMNIG-LDL
PpRSL3    ALNTNGKPRARRGSATD-PQSVYAR-HRREKINERLKTLQHLVPNGAKVDIVTMLDEAIHYVQFLQLQVTLLKSDEYWMYATPNTYKGIDL-TNS
PpRSL4    ALNRNGRPRVQRGSATD-PQSVHAR-ARREKIAERLRKLQHLIPNGGKVDIVTMLDEAVHYVQFLKRQVTLLKSDEYWMYATPTSYRSKFD-DCS
PpRSL5    ALNTDGKPRAKRGSATD-PQSVYAR-HRREKINERLKSLQNLVPNGAKVDIVTMLDEAIHYVKFLQNQVELLKSDELWIYATPNKYNGMDI-SDL
PpRSL6    ALNTNGKPRAKRGSATD-PQSVYAR-HRREKINERLKNLQNLVPNGAKVDIVTMLDEAIHYVKFLQTQVELLKSDEFWMFANPHNYNGIDI-SDP
PpRSL7    ALNTNGRPRAKRGSATD-PQSVYAR-HRREKINERLKTLQRLVPNGEQVDIVTMLEEAIHFVKFLEFQLELLRSDDRWMFADPFIYNGMDI-TGS
PtRSL3    VLNSNGKTRATRGAATD-PQSLYAR-KRRERINERLKILQNLVPNGTKVDISTMLEEAVHYNFLQLQIKLLSSDDLWMYAPLAYNGIDIG-LNQ
PtRSL4    TLNSSGKTRASRGAATD-PQSLYAR-KRRERINERLRILQTLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSDDLWMYAPIAYNGMDIG-LDH
PtRSL5    TLNSSGKTRASKGAATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVQYVKFLQLQIKLLSSEDLWMYAPIAYNGMDIG-LDH
SbRSL3    GSGGNAKAQAAKGSATD-PQSLYAR-RRRERINERLKILQKLVPNGTKVDISTMLEEAVHYVRFLQQQIKMLSSDEMWMYAPIAYNGMSLG-IDM
SbRSL4    SDNLGGKAKADRRSATE-SQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAVEYVKFLQLQIKLLSSDEMWMYAPIAYNGMNIG-LDL
SbRSL5    SSTKKAYTRASRGAATD-PQSLYAR-KRRERINERLRILQKLVPNGTKVDISTMLEEAAQYVKFLQLQIKLLSSDDTWMYAPIAYNGININSNVDL
SbRSL6    VLKQSTSSRGRSRSSKD-LQSLYAK-RRRERINERLRTLQQLIPNGTKVDMSTMLEEAVQYVKFLQLQIKLLSSEDTWMYAPLAYNHMSMD-VSQ
SbRSL7    ARRPKGKGRAGRSATTE-PQSLYAR-KRRERINERLKILQNLVPNGTKVDISTMLEEAVHYVKFLQLQIRLLSSDDTWMYAPIAYNGMNIG-IGI
SmRSL7    ALNTNGKPRAKRGSATD-PQSVYAR-HRREKINERLKTLQHLVPNGAKVDIVTMLEEAIHYVKFLQLQVNMLSSDEYWTYAPTTYNGPETP-LGL
SmRSL8    ALNTDGKPRAKRGSATD-PQSIYAR-QRRERINERLRALQGLVPNGAKVDIVTMLEEAINYVKFLQ--LLSSDEYWMYAPTNYNGMNIS-LGM
ZmRSL3    SSAKKTCTRASRGSATD-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAAQYVKFLQLQIKLLSCDDTWMYAPIAYNGININSNVDL
ZmRSL4    SSSKKTCARASRGAATE-PQSLYAR-KRRERINERLRILQNLVPNGTKVDISTMLEEAAQYVKFLQLQIKLLCDDTWMYAPIAYNGININGNVDL
ZmRSL5    ARAAGRGAAAATTTTAE-PQSLYAR-KRRERINERLKVLQSLVPNGTKVDMSTMLEEAVHYVKFLQLQIRLLSSDDTWMYAPIAYNGMGIG-IDL
AtRHD6    GVTGKTKPKPTT-SPKD-PQSLAAK-NRRERISERLKILQELVPNGTKVDLVTMLEKAISYVKFLQVQVKVLATDEFWPAQGGKAPDISQV-KDA
PpRSL1    ALNTNLKPRARQGSAND-PQSIAAR-VRRERISERLKVLQALIPNGDKVDMVTMLEKAISYVQCLEFQIKMLKNDSLWPKALGPLPNTLQELLEL
```

Alignment used for the phylogenetic analyses of RSL class II proteins.

Conserved amino acid regions of 56 RSL class II proteins were manually aligned. The RSL class I proteins AtRHD6 and PpRSL1 were used as outgroups. The sequence logo shown below the alignment was generated using WebLogo; the red line indicates the location of the bHLH domain. Sequence names are indicated in Appendix 2 (MgRSL5 and OsbHLH10 are not shown due to presumably incorrect splicing predictions); a FASTA format version of this alignment is given in Supplementary File 10.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21:** 2104-2105

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell **15:** 63-78

Algeo TJ, Scheckler SE (1998) Terrestrial-marine teleconnections in the Devonian: links between the evolution of land plants, weathering processes, and marine anoxic events. Philos Trans R Soc Lond B Biol Sci **353:** 113-130

Anantharaman V, Koonin EV, Aravind L (2001) Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol **307:** 1271-1292

Anterola A, Shanle E (2008) Genomic insights in moss gibberellin biosynthesis. The Bryologist **111:** 218-230

Anterola A, Shanle E, Mansouri K, Schuette S, Renzaglia K (2009) Gibberellin precursor is involved in spore germination in the moss *Physcomitrella patens*. Planta **229:** 1003-1007

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism. Science **306:** 79-86

Ashton NW, Cove DJ (1977) The Isolation and Preliminary Characterisation of Auxotrophic and Analogue Resistant Mutants of the Moss, *Physcomitrella patens*. Molec Gen Genet **154:** 87-95

Ashton NW, Grimsley NH, Cove DJ (1979) Analysis of gametophytic development in the moss, *Physcomitrella patens*, using auxin and cytokinin resistant mutants. Planta **144:** 427-435

Atchley WR, Fitch WM (1997) A natural classification of the basic helix-loop-helix class of transcription factors. Proc Natl Acad Sci USA **94:** 5172-5176

Atchley WR, Terhalle W, Dress A (1999) Positional Dependence, Cliques, and Predictive Motifs in the bHLH Protein Domain. J Mol Evol **48:** 501-516

Atchley WR, Zhao J (2007) Molecular Architecture of the DNA-Binding Region and Its Relationship to Classification of Basic Helix-Loop-Helix Proteins. Mol Biol Evol **24:** 192-202

Atzorn R, Bopp M, Merdes U (1989) The physiological role of indole acetic acid in the moss *Funaria hygrometrica* Hedw. II. Mutants of *Funaria hygrometrica* which exhibit enhanced catabolism of indole-3-acetic acid. J. Plant Physiol. **135:** 526-530

Augustine RC, Vidali L, Kleinman KP, Bezanilla M (2008) Actin depolymerizing factor is essential for viability in plants, and its phosphoregulation is important for tip growth. Plant J **54:** 863-875

Axtell MJ, Bowman JL (2008) Evolution of plant microRNAs and their targets. Trends Plant Sci **13:** 343-349

Axtell MJ, Snyder JA, Bartel DP (2007) Common functions for diverse small RNAs of land plants. Plant Cell **19:** 1750-1769

Badawi M, Reddy YV, Agharbaoui Z, Tominaga Y, Danyluk J, Sarhan F, Houde M (2008) Structure and Functional Analysis of Wheat *ICE* (Inducer of CBF Expression) Genes. Plant Cell Physiol **49:** 1237-1249

Bailey PC, Martin C, Toledo-Ortiz G, Quail PH, Huq E, Heim MA, Jakoby M, Werber M, Weisshaar B (2003) Update on the Basic Helix-Loop-Helix Transcription Factor Gene Family in *Arabidopsis thaliana*. Plant Cell **15:** 2497-2502

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol **2:** 28-36

Baldauf SL (2003) The Deep Roots of Eukaryotes. Science **300:** 1703-1706

Bateman RM, Crane PR, DiMichele WA, Kenrick PR, Rowe NP, Speck T, Stein WE (1998) EARLY EVOLUTION OF LAND PLANTS: Phylogeny, Physiology, and Ecology of the Primary Terrestrial Radiation. Annu Rev Ecol Syst **29:** 263-292

Bauer P, Ling H-Q, Guerinot ML (2007) *FIT*, the *FER-LIKE IRON DEFICIENCY INDUCED TRANSCRIPTION FACTOR* in *Arabidopsis*. Plant Physiol Bioch **45:** 260-261

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

Bell PR (1992) Apospory and Apogamy: Implications for Understanding the Plant Life Cycle. Int J Plant Sci **153:** S123-S136

Bernhardt C, Lee MM, Gonzalez A, Zhang F, Lloyd A, 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

Bharathan G, Goliber TE, Moore C, Kessler S, Pham T, Sinha NR (2002) Homologies in Leaf Form Inferred from *KNOXI* Gene Expression During Development. Science **296:** 1858-1860

Bierfreund NM, Reski R, Decker EL (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

Blackwell W (2003) Two theories of origin of the land-plant sporophyte: Which is left standing? Bot Rev **69:** 125-148

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell **16:** 1679-1691

Blein T, Pulido A, Vialette-Guiraud A, Nikovics K, Morin H, Hay A, Johansen IE, Tsiantis M, Laufs P (2008) A Conserved Molecular Framework for Compound Leaf Development. Science **322:** 1835-1839

Bopp M, Atzorn R (1992) The morphogenetic system of the moss protonema. Crypt Bot **3:** 3-10

Bornberg-Bauer E, Rivals E, Vingron M (1998) Computational approaches to identify leucine zippers. Nucl Acids Res **26:** 2740-2746

Bower FO (1908) The origin of a land flora. MacMillan & Co., London

Bowman J (1997) Evolutionary conservation of angiosperm flower development at the molecular and genetic levels. J Biosciences **22:** 515-527

Boyce CK (2010) The evolution of plant development in a paleontological context. Curr Opin Plant Biol **13:** 102-107

Buck M, Atchley W (2003) Phylogenetic Analysis of Plant Basic Helix-Loop-Helix Proteins. J Mol Evol **56:** 742-750

Burr FA, Burr B, Scheffler BE, Blewitt M, Wienand U, Matz EC (1996) The maize repressor-like gene intensifier1 shares homology with the r1/b1 multigene family of transcription factors and exhibits missplicing. Plant Cell **8:** 1249-1259

Buzgo M, Soltis PS, Soltis DE (2004) Floral Developmental Morphology of *Amborella trichopoda* (Amborellaceae). Int J Plant Sci **165:** 925-947

Carroll SB (2008) Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell **134:** 25-36

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

Chinnusamy V, Ohta M, Kanrar S, Lee B-h, Hong X, Agarwal M, Zhu J-K (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Gene Dev **17:** 1043-1054

Chipman DM, Shaanan B (2001) The ACT domain family. Curr Opin Struc Biol **11:** 694-700

Cho SH, Addo-Quaye C, Coruh C, Arif MA, Ma Z, Frank W, Axtell MJ (2008) *Physcomitrella patens DCL3* is required for 22-24 nt siRNA accumulation, suppression of retrotransposon-derived transcripts, and normal development. PLoS Genet **4:** e1000314

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J **16:** 735-743

Conrad PA, Hepler PK (1988) The effect of 1,4-dihydropyridines on the initiation and development of gametophore buds in the moss *Funaria*. Plant Physiol **86**: 684-687

Cove D (2005) The moss Physcomitrella patens. Annu Rev Genet **39**: 339-358

Cove D, Bezanilla M, Harries P, Quatrano R (2006) Mosses as model systems for the study of metabolism and development. Annu Rev Plant Biol **57**: 497-520

Cove D, Knight C (1987) Gravitropism and phototropism in the moss, *Physcomitrella patens*. *In* H Thomas, D Grierson, eds, Developmental Mutants in Higher Plants. Cambridge Univ. Press, London, pp 181-196

Crane PR, Kenrick P (1997) Diverted development of reproductive organs: A source of morphological innovation in land plants. Plant Syst Evol **206**: 161-174

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW (2006) Widespread genome duplications throughout the history of flowering plants. Genome res **16**: 738-749

Degnan BM, Vervoort M, Larroux C, Richards GS (2009) Early evolution of metazoan transcription factors. Curr Opin Genet Dev **19**: 591-599

de Lucas M, Daviere J-M, Rodriguez-Falcon M, Pontin M, Iglesias-Pedraz JM, Lorrain S, Fankhauser C, Blazquez MA, Titarenko E, Prat S (2008) A molecular framework for light and gibberellin control of cell elongation. Nature **451**: 480-484

Doonan JH (1991) The cytoskeleton and moss morphogenesis. *In* The Cytoskeletal Basis of Plant Growth and Form. Academic Press Limited

Doonan JH, Cove DJ, Corke FMK, Lloyd CW (1987) Pre-prophase band of microtubules, absent from tip-growing moss filaments, arises in leafy shoots during transition to intercalary growth. Cell Motil Cytoskeleton **7:** 138-153

Doonan JH, Cove DJ, Lloyd CW (1988) Microtubules and microfilaments in tip growth: evidence that microtubules impose polarity on protonemal growth in *Physcomitrella patens*. Journal of Cell Science **89:** 533-540

Duckett JG, Schmid AM, Ligrone R (1998) Protonemal morphogenesis. *In* Bryology for the Twenty-first Century. British Bryological Society, pp 223-246

Duek PD, Fankhauser C (2003) HFR1, a putative bHLH transcription factor, mediates both phytochrome A and cryptochrome signalling. Plant J **34**: 827-836

Eddy SR (1998) Profile Hidden Markov Models. Bioinformatics **14**: 755 - 763

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

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res **19**: 1349

Eklund DM, Thelander M, Landberg K, Ståldal V, Nilsson A, Johansson M, Valsecchi I, Pederson ERA, Kowalczyk M, Ljung K, Ronne H, Sundberg E (2010) Homologues of the Arabidopsis thaliana *SHI/STY/LRP1* genes control auxin biosynthesis and affect growth and development in the moss *Physcomitrella patens*. Development **137**: 1275-1284

Elomaa P, Mehto M, Kotilainen M, Helariutta Y, Nevalainen L, Teeri TH (1998) A bHLH transcription factor mediates organ, region and flower type specific signals on *dihydroflavonol-4-reductase* (*dfr*) gene expression in the inflorescence of *Gerbera hybrida* (Asteraceae). Plant J **16**:93-99.

Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc Natl Acad Sci USA **106**: 5737-5742

Feller A, Hernandez JM, 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

Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics **5:** 164-166

Ferré-D'Amaré A, Prendergast G, Ziff E, Burley S (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature **363**: 38-45

Finka A, Saidi Y, Goloubinoff P, Neuhaus JM, Zryd JP, Schaefer DG (2008) The knock-out of ARP3a gene affects F-actin cytoskeleton organization altering cellular tip growth, morphology and development in moss *Physcomitrella patens*. Cell Motil Cytoskeleton

Floyd SK, Bowman JL (2006) Distinct Developmental Mechanisms Reflect the Independent Origins of Leaves in Vascular Plants. Curr Biol **16:** 1911-1917

Floyd SK, Bowman JL (2007) The ancestral developmental tool kit of land plants. Int J Plant Sci **168:** 1-35

French JC, Paolillo DJ, Jr (1975) Intercalary meristematic activity in the sporophyte of *Funaria* (*Musci*). Am J Bot **62:** 86-96

Friedman WE, Moore RC, Purugganan MD (2004) The evolution of plant development. Am J Bot **91:** 1726-1741

Friedrichsen DM, Nemhauser J, Muramitsu T, Maloof JN, Alonso J, Ecker JR, Furuya M, Chory J (2002) Three Redundant Brassinosteroid Early Response Genes Encode Putative bHLH Transcription Factors Required for Normal Growth. Genetics **162:** 1445-1456

Friis EM, Pedersen KR, Crane PR (2006) Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. Palaeog Palaeoclim Palaeoecol **232:** 251-293

Fujita T, Sakaguchi H, Hiwatashi Y, Wagstaff SJ, Ito M, Deguchi H, Sato T, Hasebe M (2008) Convergent evolution of shoots in land plants: lack of auxin polar transport in moss shoots. Evol Dev **10:** 176-186

Fursova OV, Pogorelko GV, Tarasov VA (2009) Identification of *ICE2*, a gene involved in cold acclimation which determines freezing tolerance in *Arabidopsis thaliana*. Gene **429:** 98-103

Gensel P, Kotyk M, Basinger J (2001) Morphology of above- and below-ground structures in early Devonian (Pragian-Emsian). *In* P Gensel, D Edwards, eds, Plants invade the land: evolutionary and environmental perspectives. Columbia University Press, New York, pp 83-102

Gensel PG (2008) The Earliest Land Plants. Annu Rev Ecol Evol Syst **39:** 459-477

Giuliano G, Pichersky E, Malik VS, Timko MP, Scolnik PA, Cashmore AR (1988) An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. Proc Natl Acad Sci U S A **85:** 7089-7093

Goodrich J, Carpenter R, Coen ES (1992) A common gene regulates pigmentation pattern in diverse plant species. Cell **68:** 955-964

Graham LE, Cook ME, Busse JS (2000) The origin of plants: Body plan changes contributing to a major evolutionary radiation. Proc Natl Acad Sci USA **97:** 4535-4540

Gremski K, Ditta G, Yanofsky MF (2007) The *HECATE* genes regulate female reproductive tract development in *Arabidopsis thaliana*. Development **134:** 3593-3601

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:** 696-704

Guo A-Y, Chen X, Gao G, Zhang H, Zhu Q-H, Liu X-C, Zhong Y-F, Gu X, He K, Luo J (2008) PlantTFDB: a comprehensive plant transcription factor database. Nucl Acids Res **36:** D966-969

Hake S, Smith HMS, Holtan H, Magnani E, Mele G, Ramirez J (2004) The role of KNOX genes in plant development. Annu Rev Cell Dev Biol **20:** 125-151

Halder G, Callaerts P, Gehring W (1995) Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila*. Science **267:** 1788-1792

Handa AK, Johri MM (1976) Cell differentiation by 3',5'-cyclic AMP in a lower plant. Nature **259:** 480-482

Handa AK, Johri MM (1979) Involvement of cyclic adenosine-3',5'-monophosphate in chlronema differentiation in protonema cultures of *Funaria hygrometrica*. Planta **144:** 317-324

Harries PA, Pan A, Quatrano RS (2005) Actin-related protein2/3 complex component ARPC1 is required for proper cell morphogenesis and polarized cell growth in Physcomitrella patens. Plant Cell **17:** 2327-2339

Harrison CJ, Corley SB, Moylan EC, Alexander DL, Scotland RW, Langdale JA (2005) Independent recruitment of a conserved developmental mechanism during leaf evolution. Nature **434:** 509-514

Harrison CJ, Roeder AHK, Meyerowitz EM, Langdale JA (2009) Local Cues and Asymmetric Cell Divisions Underpin Body Plan Transitions in the Moss *Physcomitrella patens*. Curr Biol **19:** 461-471

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

Hayashi K, Tan X, Zheng N, Hatate T, Kimura Y, Kepinski S, Nozaki H (2008) Small-molecule agonists and antagonists of F-box protein-substrate interactions in auxin perception and signaling. Proc Natl Acad Sci USA **105:** 5632-5637

Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB (2001) Molecular Evidence for the Early Colonization of Land by Fungi and Plants. Science **293:** 1129-1133

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003) The Basic Helix-Loop-Helix Transcription Factor Family in Plants: A Genome-Wide Study of Protein Structure and Functional Diversity. Mol Biol Evol **20:** 735-747

Heisler MG, Atkinson A, Bylstra YH, Walsh R, Smyth DR (2001) SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. Development **128:** 1089-1098

Henschel K, Kofuji R, Hasebe M, Saedler H, Munster T, Theißen G (2002) Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. Mol Biol Evol **19:** 801-814

Hernick LV, Landing E, Bartowski KE (2008) Earth's oldest liverworts—*Metzgeriothallus sharonae* sp. nov. from the Middle Devonian (Givetian) of eastern New York, USA. Rev Paleobot Palyno **148:** 154-162

Hirano K, Nakajima M, Asano K, Nishiyama T, Sakakibara H, Kojima M, Katoh E, Xiang H, Tanahashi T, Hasebe M, Banks JA, Ashikari M, Kitano H, Ueguchi-Tanaka M, Matsuoka M (2007) The GID1-mediated gibberellin perception mechanism is conserved in the lycophyte *Selaginella moellendorffii* but not in the bryophyte *Physcomitrella patens*. Plant Cell 19: 3058-3079

Hu J, Reddy VS, Wessler SR (2000) The rice R gene family: two distinct subfamilies containing several miniature inverted-repeat transposable elements. Plant Mol Biol **42:** 667-678

Hyun Y, Lee I (2006) *KIDARI*, Encoding a Non-DNA Binding bHLH Protein, Represses Light Signal Transduction in *Arabidopsis thaliana*. Plant Mol Biol **61:** 283-296

Imai A, Hanzawa Y, Komura M, Yamamoto KT, Komeda Y, Takahashi T (2006) The dwarf phenotype of the *Arabidopsis acl5* mutant is suppressed by a mutation in an upstream ORF of a bHLH gene. Development **133:** 3575-3585

Imaizumi T, Kadota A, Hasebe M, Wada M (2002) Cryptochrome light signals control development to suppress auxin sensitivity in the moss *Physcomitrella patens*. Plant Cell **14:** 373-386

Irish VF (2009) Evolution of petal identity. J Exp Bot **60:** 2517-2527

Jensen L (1981) Division, growth, and branch formation in protonema of the moss *Physcomitrium turbinatum*: studies of sequential cytological changes in living cells. Protoplasma **107:** 301-317

John BDM (1994) Alternation of generations in algae: its complexity, maintenance and evolution. Biol Rev **69:** 275-291

Johri MM, Desai S (1973) Auxin regulation of caulonema formation in moss protonema. Nature New Biol **245:** 223-224

Jones S (2004) An overview of the basic helix-loop-helix proteins. Genome Biol **5:** 226

Jung K-H, Han M-J, Lee Y-S, Kim Y-W, Hwang I, Kim M-J, Kim Y-K, Nahm BH, An G (2005) Rice *Undeveloped Tapetum1* Is a Major Regulator of Early Tapetum Development. Plant Cell **17:** 2705-2722

Kamisugi Y, Schlink K, Rensing SA, Schween G, von Stackelberg M, Cuming AC, Reski R, Cove DJ (2006) The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. Nucleic Acids Res **34**: 6205-6214

Kanaoka MM, Pillitteri LJ, Fujii H, Yoshida Y, Bogenschutz NL, Takabayashi J, Zhu J-K, Torii KU (2008) *SCREAM/ICE1* and *SCREAM2* Specify Three Cell-State Transitional Steps Leading to *Arabidopsis* Stomatal Differentiation. Plant Cell **20**: 1775-1785

Karol KG, McCourt RM, Cimino MT, Delwiche CF (2001) The Closest Living Relatives of Land Plants. Science **294**: 2351-2353

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

Khanna R, Huq E, Kikis EA, Al-Sady B, Lanzatella C, Quail PH (2004) A Novel Molecular Recognition Motif Necessary for Targeting Photoactivated Phytochrome Signaling to Specific Basic Helix-Loop-Helix Transcription Factors. Plant Cell **16**: 3033-3044

Kimura S, Koenig D, Kang J, Yoong FY, Sinha N (2008) Natural variation in leaf morphology results from mutation of a novel *KNOX* gene. Curr Biol **18**: 672-677

Kiribuchi K, Sugimori M, Takeda M, Otani T, Okada K, Onodera H, Ugaki M, Tanaka Y, Tomiyama-Akimoto C, Yamaguchi T, Minami E, Shibuya N, Omori T, Nishiyama M, Nojiri H, Yamane H (2004) *RERJ1*, a jasmonic acid-responsive gene from rice, encodes a basic helix-loop-helix protein. Biochem Bioph Res Co **325**: 857-863

Knight CD, Sehgal A, Atwal K, Wallace JC, Cove DJ, Coates D, Quatrano RS, Bahadur S, Stockley PG, Cuming AC (1995) Molecular Responses to Abscisic Acid and Stress Are Conserved between Moss and Cereals. Plant Cell **7**: 499-506

Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda K, Ito M, Hasebe M (2003) Evolution and Divergence of the MADS-Box Gene Family Based on Genome-Wide Expression Analyses. Mol Biol Evol **20**: 1963-1977

Koini MA, Alvey L, Allen T, Tilley CA, Harberd NP, Whitelam GC, Franklin KA (2009) High Temperature-Mediated Adaptations in Plant Architecture Require the bHLH Transcription Factor PIF4. Curr Biol **19**: 408-413

Komatsu K, Maekawa M, Ujiie S, Satake Y, Furutani I, Okamoto H, Shimamoto K, Kyozuka J (2003) *LAX* and *SPA*: Major regulators of shoot branching in rice. Proc Natl Acad Sci USA **100**: 11765-11770

Komatsu K, Nishikawa Y, Ohtsuka T, Taji T, Quatrano R, Tanaka S, Sakata Y (2009) Functional analyses of the ABI1-related protein phosphatase type 2C reveal evolutionarily conserved regulation of abscisic acid signaling between Arabidopsis and the moss *Physcomitrella patens*. Plant Mol Biol **70**: 327-340

Kondou Y, Nakazawa M, Kawashima M, Ichikawa T, Yoshizumi T, Suzuki K, Ishikawa A, Koshi T, Matsui R, Muto S, Matsui M (2008) RETARDED GROWTH OF EMBRYO1, a New Basic Helix-Loop-Helix Protein, Expresses in Endosperm to Control Embryo Growth. Plant Physiol **147**: 1924-1935

Lasorella A, Stegmuller J, Guardavaccaro D, Liu G, Carro MS, Rothschild G, de la Torre-Ubieta L, Pagano M, Bonni A, Iavarone A (2006) Degradation of Id2 by the anaphase-promoting complex couples cell cycle exit and axonal growth. Nature **442**: 471-474

Lau S, Shao N, Bock R, Jürgens G, De Smet I (2009) Auxin signaling in algal lineages: fact or myth? Trends Plant Sci **14**: 182-188

Ledent V, Paquet O, Vervoort M (2002) Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol **3**: research0030.0031–0030.0018

Ledent V, Vervoort M (2001) The Basic Helix-Loop-Helix Protein Family: Comparative Genomics and Phylogenetic Analysis. Genome Res **11**: 754-770

Lee J-H, Lin H, Joo S, Goodenough U (2008) Early Sexual Origins of Homeoprotein Heterodimerization and Evolution of the Plant KNOX/BELL Family. Cell **133**: 829-840

Lee S, Lee S, Yang K-Y, Kim Y-M, Park S-Y, Kim SY, Soh M-S (2006) Overexpression of *PRE1* and its Homologous Genes Activates Gibberellin-dependent Responses in *Arabidopsis thaliana*. Plant Cell Physiol **47**: 591-600

Leivar P, Monte E, Al-Sady B, Carle C, Storer A, Alonso JM, Ecker JR, Quail PH (2008) The *Arabidopsis* Phytochrome-Interacting Factor PIF7, Together with PIF3 and PIF4, Regulates Responses to Prolonged Red Light by Modulating phyB Levels. Plant Cell **20:** 337-352

Lenton TM (2001) The role of land plants, phosphorus weathering and fire in the rise and regulation of atmospheric oxygen. Glob Change Biol **7:** 613-629

Lewis LA, McCourt RM (2004) Green algae and the origin of land plants. Am J Bot **91:** 1535-1556

Li H, Sun J, Xu Y, Jiang H, Wu X, Li C (2007) The bHLH-type transcription factor AtAIB positively regulates ABA response in *Arabidopsis*. Plant Mol Biol **65:** 655-665

Li N, Zhang D-S, Liu H-S, Yin C-S, Li X-x, Liang W-q, Yuan Z, Xu B, Chu H-W, Wang J, Wen T-Q, Huang H, Luo D, Ma H, Zhang D-B (2006a) The Rice *Tapetum Degeneration Retardation* Gene Is Required for Tapetum Degradation and Anther Development. Plant Cell **18**: 2999-3014

Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J, Ma H, Wang J, Zhang D (2006b) Genome-Wide Analysis of Basic/Helix-Loop-Helix Transcription Factor Family in Rice and Arabidopsis. Plant Physiol **141:** 1167-1184

Liljegren SJ, Roeder AHK, Kempin SA, Gremski K, Østergaard L, Guimil S, Reyes DK, Yanofsky MF (2004) Control of Fruit Patterning in *Arabidopsis* by INDEHISCENT. Cell **116**: 843-853

Ling H-Q, Bauer P, Bereczky Z, Keller B, Ganal M (2002) The tomato *fer* gene encoding a bHLH protein controls iron-uptake responses in roots. Proc Natl Acad Sci USA **99:** 13938-13943

Liu H, Yu X, Li K, Klejnot J, Yang H, Lisiero D, Lin C (2008) Photoexcited CRY2 Interacts with CIB1 to Regulate Transcription and Floral Initiation in *Arabidopsis*. Science **322:** 1535-1539

Liu T, Ohashi-Ito K, Bergmann DC (2009) Orthologs of *Arabidopsis thaliana* stomatal bHLH genes and regulation of stomatal development in grasses. Development **136:** 2265-2276

Lorenzo O, Chico JM, Sanchez-Serrano JJ, Solano R (2004) *JASMONATE-INSENSITIVE1* Encodes a MYC Transcription Factor Essential to Discriminate between Different Jasmonate-Regulated Defense Responses in Arabidopsis. Plant Cell **16:** 1938-1950

Ludwig SR, Habera LF, Dellaporta SL, Wessler SR (1989) Lc, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proc Natl Acad Sci USA **86:** 7092-7096

Ludwig-Müller J, Jülke S, Bierfreund NM, Decker EL, Reski R (2009) Moss (*Physcomitrella patens*) GH3 proteins act in auxin homeostasis. New Phytol **181:** 323-338

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci USA **102:** 5454-5459

Martínez-García JF, Huq E, Quail PH (2000) Direct Targeting of Light Signals to a Promoter Element-Bound Transcription Factor. Science **288:** 859-863

Massari ME, Murre C (2000) Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. Mol Cell Biol **20:** 429-440

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

Matsushima R, Fukao Y, Nishimura M, Hara-Nishimura I (2004) NAI1 Gene Encodes a Basic-Helix-Loop-Helix-Type Putative Transcription Factor That Regulates the Formation of an Endoplasmic Reticulum-Derived Structure, the ER Body. Plant Cell **16:** 1536-1549

Matsuzaki M, Misumi O, Shin-i T, Maruyama S, Takahara M, Miyagishima S-y, Mori T, Nishida K, Yagisawa F, Nishida K, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T, Minoda A, Sano M, Nomoto H, Oishi K, Hayashi H, Ohta F, Nishizaka S, Haga S, Miura S, Morishita T, Kabeya Y, Terasawa K, Suzuki Y, Ishii Y, Asakawa S, Takano H, Ohta N, Kuroiwa H, Tanaka K, Shimizu N, Sugano S, Sato N, Nozaki H, Ogasawara N, Kohara Y, Kuroiwa T (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature **428**: 653-657

Melzer R, Wang Y-Q, Theißen G (2010) The naked and the dead: The ABCs of gymnosperm reproduction and the origin of the angiosperm flower. Semin Cell Dev Biol **21:** 118-128

Menand B, Calder G, Dolan L (2007a) Both chloronemal and caulonemal cells expand by tip growth in the moss *Physcomitrella patens*. J Exp Bot **58**: 1843-1849

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

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L, Marshall WF, Qu L-H, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen C-L, Cognat V, Croft MT, Dent R, Dutcher S, Fernandez E, Fukuzawa H, Gonzalez-Ballester D, Gonzalez-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanon M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV, Lohr M, Manuell A, Meier I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV, Paulsen IT, Pazour G, Purton S, Ral J-P, Riano-Pachon DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen C-J, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang P, Ball S, Bowler C, Dieckmann CL, Gladyshev VN, Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martinez D, Ngau WCA, Otillar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR (2007) The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. Science **318**: 245-250

Miller DH, Miller HA (2006) *Funaria*. *In* Bryophyte Flora of North America, provisional publication (http://www.mobot.org/plantscience/bfna/V1/FunaFunaria.htm). Missouri Botanical Garden, USA

Mittmann F, Brucker G, Zeidler M, Repp A, Abts T, Hartmann E, Hughes J (2004) Targeted knockout in Physcomitrella reveals direct actions of phytochrome in the cytoplasm. Proc Natl Acad Sci USA **101**: 13939-13944

Molnár A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC (2007) miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. Nature **447**: 1126-1129

Morgenstern B, Atchley WR (1999) Evolution of bHLH transcription factors: modular evolution by domain shuffling? Mol Biol Evol **16**: 1654-1663

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

Mravec J, Skupa P, Bailly A, Hoyerova K, Krecek P, Bielach A, Petrasek J, Zhang J, Gaykova V, Stierhof Y-D, Dobrev PI, Schwarzerova K, Rolcik J, Seifertova D, Luschnig C, Benkova E, Zazimalova E, Geisler M, Friml J (2009) Subcellular homeostasis of phytohormone auxin is mediated by the ER-localized PIN5 transporter. Nature **459**: 1136-1140

Mukherjee K, Brocchieri L, Burglin TR (2009) A Comprehensive Classification and Evolutionary Analysis of Plant Homeobox Genes. Mol Biol Evol **26**: 2775-2794

Munster T, Pahnke J, Di Rosa A, Kim JT, Martin W, Saedler H, Theißen G (1997) Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. Proc Natl Acad Sci USA **94**: 2415-2420

Murre C, McCaw PS, Baltimore D (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. Cell **56**: 777-783

Nadeau JA (2009) Stomatal development: new signals and fate determinants. Curr Opin Plant Biol **12**: 29-35

Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L (2000) The *TT8* Gene Encodes a Basic Helix-Loop-Helix Domain Protein Required for Expression of *DFR* and *BAN* Genes in Arabidopsis Siliques. Plant Cell **12**: 1863-1878

Niklas KJ, Kutschera U (2009) The evolutionary development of plant body plans. Funct Plant Biol **36**: 682-695

Niklas KJ, Kutschera U (2010) The evolution of the land plant life cycle. New Phytol **185**: 27-41

Ohashi-Ito K, Bergmann DC (2007) Regulation of the *Arabidopsis* root vascular initial population by *LONESOME HIGHWAY*. Development **134**: 2959-2968

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

Olson EN (2006) Gene Regulatory Networks in the Evolution and Development of the Heart. Science **313**: 1922-1927

Pagnussat GC, Yu H-J, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie L-F, Ye D, Sundaresan V (2005) Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. Development **132**: 603-614

Palenik B, Grimwood J, Aerts A, Rouzé P, Salamov A, Putnam N, Dupont C, Jorgensen R, Derelle E, Rombauts S, Zhou K, Otillar R, Merchant SS, Podell S, Gaasterland T, Napoli C, Gendler K, Manuell A, Tai V, Vallon O, Piganeau G, Jancek S, Heijde M, Jabbari K, Bowler C, Lohr M, Robbens S, Werner G, Dubchak I, Pazour GJ, Ren Q, Paulsen I, Delwiche C, Schmutz J, Rokhsar D, Van de Peer Y, Moreau H, Grigoriev IV (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. Proc Natl Acad Sci USA **104**: 7705-7710

Panganiban G, Irvine SM, Lowe C, Roehl H, Corley LS, Sherbon B, Grenier JK, Fallon JF, Kimble J, Walker M, Wray GA, Swalla BJ, Martindale MQ, Carroll SB (1997) The origin and evolution of animal appendages. Proc Natl Acad Sci USA **94**: 5162-5166

Parker JS, Cavell AC, Dolan L, Roberts K, Grierson CS (2000) Genetic Interactions during Root Hair Morphogenesis in Arabidopsis. Plant Cell **12**: 1961-1974

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature **457**: 551-556

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

Penfield S, Josse E-M, Kannangara R, Gilday AD, Halliday KJ, Graham IA (2005) Cold and Light Control Seed Germination through the bHLH Transcription Factor SPATULA. Curr Biol **15**: 1998-2006

Perroud PF, Quatrano RS (2006) The role of ARPC4 in tip growth and alignment of the polar axis in filaments of Physcomitrella patens. Cell Motil Cytoskeleton **63**: 162-171

Perroud PF, Quatrano RS (2008) BRICK1 is required for apical cell growth in filaments of the moss Physcomitrella patens but not for gametophore morphology. Plant Cell **20**: 411-422

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

Poli D, Jacobs M, Cooke TJ (2003) Auxin regulation of axial growth in bryophyte sporophytes: its potential significance for the evolution of early land plants. Am J Bot **90**: 1405-1415

Preston JC, Hileman LC (2009) Developmental genetics of floral symmetry evolution. Trends Plant Sci **14**: 147-154

Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet **3**: 827-837

Qian W, Tan G, Liu H, He S, Gao Y, An C (2007) Identification of a bHLH-type G-box binding factor and its regulation activity with G-box and Box I elements of the *PsCHS1* promoter. Plant Cell Rep **26**: 85-93

Qiu Y-L, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ,

Stech M, Frey W, Quandt D, Davis CC (2006) The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci USA **103:** 15511-15516

Qiu Y-L (2008) Phylogeny and evolution of charophytic algae and land plants. J Syst Evol **43:** 287-306

Quattrocchio F, Wing JF, van der Woude K, Mol JN, Koes R (1998) Analysis of bHLH and MYB domain proteins: species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. Plant J **13:**475-488.

Quodt V, Faigl W, Saedler H, Munster T (2007) The MADS-domain protein PPM2 preferentially occurs in gametangia and sporophytes of the moss *Physcomitrella patens*. Gene **400:** 25-34

Rajani S, Sundaresan V (2001) The *Arabidopsis* myc/bHLH gene *ALCATRAZ* enables cell separation in fruit dehiscence. Curr Biol **11:** 1914-1922

Rampey RA, Woodward AW, Hobbs BN, Tierney MP, Lahner B, Salt DE, Bartel B (2006) An Arabidopsis Basic Helix-Loop-Helix Leucine Zipper Protein Modulates Metal Homeostasis and Auxin Conjugate Responsiveness. Genetics **174:** 1841-1857

Ramsay NA, Glover BJ (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci **10:** 63-70

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

Rensing SA, Ick J, Fawcett JA, Lang D, Zimmer A, Van de Peer Y, Reski R (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. BMC Evol Biol **7:** 130

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin IT, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science **319**: 64-69

Riaño-Pachon DM, Correa LGG, Trejos-Espinosa R, Mueller-Roeber B (2008) Green Transcription Factors: A Chlamydomonas Overview. Genetics **179:** 31-39

Richardt S, Lang D, Reski R, Frank W, Rensing SA (2007) PlanTAPDB, a Phylogeny-Based Resource of Plant Transcription-Associated Proteins. Plant Physiol. **143:** 1452-1466

Riese M, Zobell O, Saedler H, Huijser P (2008) SBP-domain transcription factors as possible effectors of cryptochrome-mediated blue light signalling in the moss *Physcomitrella patens*. Planta **227:** 505-515

Robinson KA, Lopes JM (2000) *Saccharomyces cerevisiae* basic helix-loop-helix proteins regulate diverse biological processes. Nucl Acids Res **28:** 1499-1505

Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF (2005) Monophyly of Primary Photosynthetic Eukaryotes: Green Plants, Red Algae, and Glaucophytes. Curr Biol **15:** 1325-1330

Roig-Villanova I, Bou-Torrent J, Galstyan A, Carretero-Paulet L, Portoles S, Rodriguez-Concepcion M, Martinez-Garcia JF (2007) Interaction of shade avoidance and auxin responses: a role for two novel atypical bHLH proteins. EMBO J **26:** 4756-4767

Rokas A (2008) The Origins of Multicellularity and the Early History of the Genetic Toolkit For Animal Development. Annu Rev Genet **42:** 235-251

Rose S, Bopp M (1983) Uptake and polar transport of indoleacetic acid in moss rhizoids. Physiol Plant **58:** 57-61

Sack F, Paolillo DJ, Jr (1983) Structure and development of walls in *Funaria* stomata. Am J Bot **70:** 1019-1030

Sakakibara K, Nishiyama T, Deguchi H, 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 Dev **10:** 555-566

Sakakibara K, Nishiyama T, Sumikawa N, Kofuji R, Murata T, Hasebe M (2003) Involvement of auxin and a homeodomain-leucine zipper I gene in rhizoid development of the moss *Physcomitrella patens*. Development **130:** 4835-4846

Sakamoto W, Ohmori T, Kageyama K, Miyazaki C, Saito A, Murata M, Noda K, Maekawa M (2001) The *Purple leaf* (*Pl*) Locus of Rice: the *Pl^{w}* Allele has a Complex Organization and Includes Two Genes Encoding Basic Helix-Loop-Helix Proteins Involved in Anthocyanin Biosynthesis. Plant Cell Physiol **42:** 982-991

Saunders MJ, Hepler PK (1981) Localization of membrane-associated calcium following cytokinin treatment in *Funaria* using chlorotetracycline. Planta **152:** 272-281

Saunders MJ, Hepler PK (1982) Calcium ionophore A23187 stimulates cytokinin-like mitosis in *Funaria*. Science **217:** 943-945

Schaefer DG (2001) Gene targeting in *Physcomitrella patens*. Curr Opin Plant Biol **4:** 143-150

Schaefer DG, Zrÿd JP (1997) Efficient gene targeting in the moss *Physcomitrella patens*. Plant J **11:** 1195-1206

Schmiedel G, Schnepf E (1979) Side branch formation and orientation in the caulonema of the moss, *Funaria hygrometrica*: normal development and fine structure. Protoplasma **100:** 367-383

Schmiedel G, Schnepf E (1980) Polarity and growth of caulonema tip cells of the moss *Funaria hygrometrica*. Planta **147:** 405-413

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X-C, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature **463**: 178-183

Schofield WB (1985) Introduction to Bryology. Macmillan Publishing Company, New York

Schumaker KS, Dietrich MA (1997) Programmed changes in form during moss development. Plant Cell **9:** 1099-1107

Schween G, Gorr G, Hohe A, Reski R (2003) Unique Tissue-Specific Cell Cycle in *Physcomitrella*. Plant Biol **5:** 50-58

Scotland RW (2010) Deep homology: A view from systematics. BioEssays **32:** 438-449

Serna L (2009) Emerging Parallels between Stomatal and Muscle Cell Lineages. Plant Physiol **149:** 1625-1631

Sharma S, Johri MM (1982) Uptake and degradation of cyclic AMP by chloronema cells. Plant Physiol **69:** 1401-1403

Shiu S-H, Shih M-C, Li W-H (2005) Transcription Factor Families Have Much Higher Expansion Rates in Plants than in Animals. Plant Physiol **139:** 18-26

Shubin N, Tabin C, Carroll S (1997) Fossils, genes and the evolution of animal limbs. Nature **388:** 639-648

Shubin N, Tabin C, Carroll S (2009) Deep homology and the origins of evolutionary novelty. Nature **457:** 818-823

Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan B, 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

Singer SD, Krogan NT, Ashton NW (2007) Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. Plant Cell Rep **26:** 1155-1169

Smith JM (1999) Shaping Life: genes, embryos and evolution. Yale University Press, Yale

Sorensen A-M, Kröber S, Unte US, Huijser P, Dekker K, Saedler H (2003) The *Arabidopsis ABORTED MICROSPORES* (*AMS*) gene encodes a MYC class transcription factor. Plant J **33:** 413-423

Specht CD, Bartlett ME (2009) Flower Evolution: The Origin and Subsequent Diversification of the Angiosperm Flower. Annu Rev Ecol Evol S **40:** 217-243

Spelt C, Quattrocchio F, Mol JNM, Koes R (2000) *anthocyanin1* of Petunia Encodes a Basic Helix-Loop-Helix Protein That Directly Activates Transcription of Structural Anthocyanin Genes. Plant Cell **12:** 1619-1632

Spinner L, Pastuglia M, Belcram K, Pegoraro M, Goussot M, Bouchez D, Schaefer DG (2010) The function of TONNEAU1 in moss reveals ancient mechanisms of division plane specification and cell elongation in land plants. Development **137:** 2733-2742

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

Steemans P, Herisse AL, Melvin J, Miller MA, Paris F, Verniers J, Wellman CH (2009) Origin and Radiation of the Earliest Vascular Land Plants. Science **324**: 353

Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught Red-Handed: *Rc* Encodes a Basic Helix-Loop-Helix Protein Conditioning Red Pericarp in Rice. Plant Cell **18:** 283-294

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) *In*, Ed Version 4. Sinauer Associates, Sunderland, Massachusetts.

Szecsi J, Joly C, Bordji K, Varaud E, Cock JM, Dumas C, Bendahmane M (2006) *BIGPETALp*, a *bHLH* transcription factor is involved in the control of *Arabidopsis* petal size. EMBO J **25:** 3912-3920

Taiz L, Zeiger E (2002) Plant Physiology, Ed Third. Sinauer Associates, Inc, Sunderland, USA

Talmor-Neiman M, Stav R, Frank W, Voss B, Arazi T (2006) Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss. Plant J **47:** 25-37

Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, Münster T, Theißen G, Nozaki H, Ito M (2005) Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. Proc Natl Acad Sci USA **102:** 2436-2441

Tanahashi T, Sumikawa N, Kato M, 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

Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome res **18:** 1944-1954

Tanzer A, Riester M, Hertel J, Bermudez-Santana CI, Gorodkin J, Hofacker IL, Stadler PF (2010) Evolutionary genomics of microRNAs and their relatives. *In* G Caetano-Anollés, ed, Evolutionary Genomics and Systems Biology. Wiley-Blackwell, Hoboken

Taylor TN, Taylor EL, Krings M (2009) Paleobotany: The Biology and Evolution of Fossil Plants, Ed Second. Academic Press

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463:** 763-768

Theißen C, Becker A, Winter K-U, Munster T, Kirchner C, Saedler H (2002) How the land plants learned their ABCs: the role of MADS-box genes in the evolutionary origin of flowers. *In* QCB Cronk, R Bateman, JA Hawkins, eds, Development genetics and plant evolution. Taylor & Francis, London, pp 173-205

Thelander M, Olsson T, Ronne H (2005) Effect of the energy supply on filamentous growth and development in *Physcomitrella patens*. J Exp Bot **56:** 653-662

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

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J,

Degroeve S, Dejardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313**: 1596-1604

Uenaka H, Wada M, Kadota A (2005) Four distinct photoreceptors contribute to light-induced side branch formation in the moss Physcomitrella patens. Planta **222:** 623-631

Vandenbussche F, Fierro AC, Wiedemann G, Reski R, Van Der Straeten D (2007) Evolutionary conservation of plant gibberellin signalling pathway components. BMC Plant Biol **7:** 65

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009a) The flowering world: a tale of duplications. Trends Plant Sci **14:** 680-688

Van de Peer Y, Maere S, Meyer A (2009b) The evolutionary significance of ancient genome duplications. Nat Rev Genet **10:** 725-732

Vernot B, Stolzer M, Goldman A, Durand D (2008) Reconciliation with non-binary species trees. Journal of computational biology **15:** 981-1006

Vidali L, Augustine RC, Kleinman KP, Bezanilla M (2007) Profilin is essential for tip growth in the moss *Physcomitrella patens*. Plant Cell **19:** 3705-3722

Vidali L, Burkart GM, Augustine RC, Kerdavid E, Tuzel E, Bezanilla M (2009) Myosin XI Is Essential for Tip Growth in *Physcomitrella patens*. Plant Cell **22:** 1868-1882

Vopalensky P, Kozmik Z (2009) Eye evolution: common use and independent recruitment of genetic components. Philos Trans R Soc Lond B Biol Sci **364:** 2819-2832

Wei W, Ayad NG, Wan Y, Zhang G-J, Kirschner MW, Kaelin WG (2004) Degradation of the SCF component Skp2 in cell-cycle phase G1 by the anaphase-promoting complex. Nature **428:** 194-198

Wellman CH, Osterloff PL, Mohiuddin U (2003) Fragments of the earliest land plants. Nature **425:** 282-285

Wiedemann G, Hermsen C, Melzer M, Büttner-Mainik A, Rennenberg H, Reski R, Kopriva S (2010) Targeted knock-out of a gene encoding sulfite reductase in the moss Physcomitrella patens affects gametophytic and sporophytic development. FEBS Lett **584:** 2271-2278

Willis KJ, McElwain JC (2002) The Evolution of Plants. Oxford University Press, Oxford, UK

Xiong Y, Liu T, Tian C, Sun S, Li J, Chen M (2005) Transcription Factors in Rice: A Genome-wide Comparative Analysis between Monocots and Eudicots. Plant Mol Biol **59:** 191-203

Yadav V, Mallappa C, Gangappa SN, Bhatia S, Chattopadhyay S (2005) A basic helix-loop-helix transcription factor in Arabidopsis, MYC2, acts as a repressor of blue light-mediated photomorphogenic growth. Plant Cell **17:** 1953-1966

Yang S, Johnston N, Talideh E, Mitchell S, Jeffree C, Goodrich J, Ingram G (2008) The endosperm-specific *ZHOUPI* gene of *Arabidopsis thaliana* regulates endosperm breakdown and embryonic epidermal development. Development **135:** 3501-3509

Yasumura Y, Crumpton-Taylor M, Fuentes S, Harberd NP (2007) Step-by-step acquisition of the gibberellin-DELLA growth-regulatory mechanism during land-plant evolution. Curr Biol **17:** 1225-1230

Yi K (2008) Temporal regulation of root hair development by *RHD6* family genes. PhD thesis. University of East Anglia, Norwich, UK

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

Yi K, Wu Z, Zhou J, Du L, Guo L, Wu Y, Wu P (2005) *OsPTF1*, a Novel Transcription Factor Involved in Tolerance to Phosphate Starvation in Rice. Plant Physiol **138:** 2087-2096

Yin Y, Vafeados D, Tao Y, Yoshida S, Asami T, Chory J (2005) A New Class of Transcription Factors Mediates Brassinosteroid-Regulated Gene Expression in Arabidopsis. Cell **120:** 249-259

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004) A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. Mol Biol Evol **21:** 809-818

Yuko O, Reiko Nakanishi I, Hiromi N, Takanori K, Michiko T, Satoshi M, Naoko KN (2007) The rice bHLH protein OsIRO2 is an essential regulator of the genes involved in Fe uptake under Fe-deficient conditions. Plant J **51:** 366-377

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

Zhang W, Sun Y, Timofejeva L, Chen C, Grossniklaus U, Ma H (2006) Regulation of *Arabidopsis* tapetum development and function by *DYSFUNCTIONAL TAPETUM1* (*DYT1*) encoding a putative bHLH transcription factor. Development **133:** 3085-3095

Zhao T, Li G, Mi S, Li S, Hannon GJ, Wang X-J, Qi Y (2007) A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. Gene Dev **21:** 1190-1203

Zhu Y, Cai X-L, Wang Z-Y, Hong M-M (2003) An Interaction between a MYC Protein and an EREBP Protein Is Involved in Transcriptional Regulation of the Rice *Wx* Gene. J Biol Chem **278:** 47803-47811

Zobell O, Faigl W, Saedler H, Munster T (2010) MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. Mol Biol Evol **27:** 1201-1211