Title: A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants

Running Title: Physcomitrella Transcriptome Atlas

Carlos Ortiz-Ramírez[1,§], Marcela Hernandez-Coronado[1,§], Anna Thamm[1], Bruno Catarino[2], Mingyi Wang[3], Liam Dolan[2], José A. Feijó[1,4], and Jörg D. Becker[1,*]

[1] Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal

[2] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

[3] Division of Plant Biology, The Samuel Roberts Noble Foundation, Ardmore, Oklahoma, USA

[4] University of Maryland, Dept of Cell Biology and Molecular Genetics. 0118 BioScience Research Bldg, College Park, MD 20742-5815, USA

[§] These authors contributed equally to this work.

[*] Corresponding author: Jörg D. Becker (jbecker@igc.gulbenkian.pt); phone +351-214464526; fax +351-214407970

Short Summary

The first land plants evolved approximately 480 million years ago and by doing so they greatly modified the terrestrial environment and global climate. Important innovations in the plant body were necessary to make the transition from an aquatic to a terrestrial environment. New tissues for water conduction, gas exchange, isolation, structural support, nutrient uptake, and to allow reproduction far from water sources were necessary. To have a better idea of which genes were responsible for the emergence of such innovations, and which still control important developmental processes in modern plants, the authors performed a

large scale gene expression analysis in the moss *Physcomitrella* patens. Mosses are believed to resemble the earliest land plants in terms of anatomy and life cycle. The authors identified and compared groups of tissue specific genes in *Physcomitrella* and in the thale cress *Arabidopsis thaliana*, a member of the most recently evolved group of flowering plants. Furthermore, they identified conserved candidate genes for the control of important developmental processes shared by all land plants. By disrupting one of these genes referred to as TCP, the authors showed that it controls branching in mosses, suggesting that this gene had a role in the evolutionary transition to a more complex plant body architecture.

Abstract

Identifying the genetic mechanisms that underpin the evolution of new organ and tissue systems is an aim of evolutionary developmental biology. Comparative functional genetic studies between angiosperms and bryophytes can define those genetic changes that were responsible for developmental innovations. Here, we report the generation of a transcriptome atlas covering most phases in the life cycle of the model bryophyte *Physcomitrella patens*, including detailed sporophyte developmental progression. We identified a comprehensive set of sporophyte specific transcription factors, and found that many of these genes have homologs in angiosperms that function in developmental processes such as flowering and shoot branching. Deletion of the *PpTCP5* transcription factor results in development of supernumerary sporangia attached to a single seta, suggesting that it negatively regulates branching in the moss sporophyte. Given that *TCP* genes repress branching in angiosperms, we suggest that this activity is ancient. Finally, comparison of *P. patens* and *Arabidopsis thaliana* transcriptomes led us to the identification of a conserved core of transcription factors expressed in tip growing cells. We identified modifications in the expression patterns of these genes that could account for developmental differences between *P. patens* tip growing cells and *A. thaliana* pollen tubes and root hairs.

Introduction

During the adaptation to life on land the gametophytic and sporophytic generations of early land plants underwent considerable anatomical and morphological changes giving rise to the characteristic dimorphic alternation of generations present in bryophytes and vascular plants today (Kenrick, 1994; Kenrick and Crane, 1997). For example, It is believed that the sporophyte evolved from being a small and simplified structure, physiologically dependent on the gametophyte (like it is in mosses) to become the complex, free-living, and sometimes huge organisms that dominate most terrestrial ecosystems today (Graham et al., 2000). This transition involved the generation of new organs and tissue systems, and identification of the genetic basis of such innovations is relevant for understanding basic questions of plant development. Several transcription factor families have been identified that were probably important for achieving such transformation (Floyd and Bowman, 2007). The class 1 *KNOTTED-LIKE HOMEOBOX (KNOX*) genes, for example, are responsible for the regulation of the indeterminate shoot apical meristem (SAM) in flowering plants (Hake et al., 2004), and since their function in *P. patens* was found to be restricted to the sporophyte (Singer and Ashton, 2007; Sakakibara et al., 2008) they probably have an ancient role regulating sporophyte meristem growth. Other important families regulating sporophyte development that probably had important roles during land plant evolution are *BELLRINGER 1-LIKE HOMEOBOX (BELL1)*, *MADS-box*, *GRAS*, and *TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR1 (TCP)*. In particular, members of the *TCP* gene family remain as a very interesting case study due to their major role on sporophyte architecture control (Martín-Trillo and Cubas, 2010), and although their expression in the sporophyte of *P. patens* has been pointed out (Frank and Scanlon, 2014), functional studies are still missing.

On the other hand, it is believed that the gametophyte followed the opposite evolutionary pathway, since in angiosperms it only comprises the embryo sac and the pollen grain (Kenrick, 1994; Taylor et al., 2005). One of the outcomes that can be predicted from such a shift in dominance is a reduction in complexity of the

gametophytic transcriptome, and at the same time, the “enlargement” of the sporophytic transcriptional program. Indeed, it has been proposed that during land plant evolution part of the genetic program controlling gametophyte development may have been co-opted by the sporophyte (Menand et al., 2007b; Dolan, 2009; Niklas and Kutschera, 2010), as exemplified by the type II *MADS-box* transcription factors *MIKC^c* (Zobell et al., 2010) and the *RHD SIX-LIKE1* (*RSL)* genes. *RSL* genes represent a remarkable example because they control tip growth both in mosses and angiosperms, however, their activity is restricted to the gametophyte of *P. patens* while in *A. thaliana* they control root hair growth, a sporophytic tissue (Pires et al., 2013; Menand et al., 2007b).

Large-scale comparative transcriptome analyses have been employed successfully to identify candidate genes according to their temporal and spatial expression profiles. This can be applied to evo-devo studies by identifying GRNs in early land plants and assessing how they have changed during evolution. In the case of *A. thaliana,* a great amount of transcriptomic data has been generated, and it is accessible for users through several meta-analysis tools. Valuable efforts have been made to create this kind of transcriptomic data sets for *P. patens*. For example, the transcriptome of protonema in several developmental conditions was generated (Xiao et al., 2011), and transcriptomic responses in gametophores and protonema exposed to several biotic and abiotic stresses were recently reported (Hiss et al., 2014). Studies regarding the sporophytic generation are limited due to the small size of the structure and to the low rate of sporophyte production in the *P. patens Gransden* strain. However, major transcriptional changes during the transition from the gametophytic to the sporophytic generations have been reported for *Villersexel* strain (O’Donoghue et al., 2013). On the other hand, the only sporophyte transcriptomic data available for *P. patens Gransden* strain was generated from immature sporophytes harvested 10 days after fertilization (Frank and Scanlon, 2014). Therefore, detailed information concerning sporophyte developmental progression is still missing. In addition, there is currently no studies including rhizoids transcriptomic data or individual reproductive organs, and no

comprehensive transcriptomic analysis has been carried out which includes data from most of the tissues that comprise *P. patens* life cycle in detail. This represents a problem when identifying differentially expressed genes (DEGs), because comparison between a limited number of tissues and developmental stages can result in lack of resolution.

In this study, we describe the transcriptional profile of most of the life cycle phases of *P. patens*, including chloronema, caulonema, rhizoids, gametophores, spores, archegonia and 4 different sporophyte developmental stages. For easy, visual access of the full data set an eFP browser was created (http://bar.utoronto.ca/efp_physcomitrella/cgi-bin/efpWeb.cgi). We compared our *P. patens* transcriptome data with available microarray data sets from *A. thaliana* to extract evolutionarily conserved gene expression signatures underlying analogous tissues in early diverging groups of land plants. We found that important gene homologs believed to control sporophyte development in angiosperms are also enriched in the sporophyte of *P. patens*, such as genes involved in shoot organ morphogenesis and floral organ development. Moreover, functional characterization of the sporophyte specific *TCP* transcription factor *PpTCP5* allowed us to propose that this gene has an evolutionary conserved function in controlling sporophyte architecture, and thus, it probably played a major role on the acquisition of sporophyte complexity during land plant evolution. Regarding the gametophyte generation, our analysis suggests that a similar core of transcription factors underlies tip growth in *A. thaliana* and *P. patens*, and that specific modifications in the expression patterns of these TFs could account for developmental differences between the tip growing cells from these distantly related species. We finally propose that this *P. patens* transcriptome atlas can be used for further genetic and physiological dissection of the early evo-devo mechanisms underlying the development of specialized organs in the subsequent lineages of land plants.

Results and Discussion

Generation of a *P. patens* transcriptome atlas and eFP browser integration

Microarray data sets from a wide variety of *P. patens* tissues representing most steps of its life cycle were generated using custom designed NimbleGene microarrays (Supplemental table 1). Tissues analysed included chloronema, caulonema, gametophore, rhizoids, protoplasts, archegonia, four different developmental sporophytic stages and spores (Fig 1). As a prerequisite for sample collection, the development of gametangia and sporophytes under our laboratory conditions was assessed (see material and methods). We found some discrepancies on the later gametangia developmental time points compared to a previous characterization done by Landberg et al. (2013) (e.g. antheridia and archegonia develop faster under our conditions) which might be attributed to the different temperatures under which gametangia formation was induced (e.g. 17 ^{o}C vs 15^{o}C). In the case of the sporophyte, development was divided into stages according to morphological characteristics (e.g. size, shape and degree of maturation). As a guide, dates for sporophyte collection were established based on the putative time of the fertilization event as follows: sporophyte 1 (S1), comprising sporophytes collected 5-6 days after fertilization (AF); sporophyte 2 (S2), 9-11 days AF; sporophyte 3 (S3), 18-20 days AF; and sporophyte M (SM), 28-33 days AF. Pictures of sporophytes representative for each stage are shown in figure 1 (bottom). Only sporophytes that morphologically resembled to the ones shown were manually dissected for each stage. To rule out the possibility of cross-tissue contamination during dissection and extraction, the expression of several genes whose transcription profile has already been determined experimentally was verified by qRT-PCR (see supplemental table 2 for full list of genes tested). Candidate genes include *RM09*, expressed in protonemal tissues (Ishikawa et al., 2011) and the *PIP2;1* aquaporin, reported to be expressed in the gametophore but not in protonema (Liénard et al., 2008). In all the cases analysed our data showed the expected pattern. Moreover, we could determine differences in expression among closely related tissue/stages such as caulonema and chloronema or distinct

sporophyte stages (Fig 2a & b). We also performed qRT-PCRs on genes we identified as preferentially expressed in several tissues (see below), and saw a high correlation between the microarray and the qRT-PCRs expression data (Fig 2c). Finally, we compared the expression profile of several genes families as predicted from our microarray and compared it with published experimental data, finding a high correlation between them (Fig 2d and supplemental fig 1). Overall these observations validate our microarray results.

To provide easy access and visualization of the dataset created in this study, expression data was uploaded to the free Bio-analytic Resource for Plant Biology (BAR) server, where it is available to the community through the electronic Fluorescent Pictograph (eFP) browser (http://bar.utoronto.ca/efp_physcomitrella/cgi-bin/efpWeb.cgi). This tool provides a very intuitive and graphical representation of large-scale microarray data sets, enabling users with no bio-informatic experience to quickly explore the data (Winter et al., 2007).

Global transcriptome trends correspond to major developmental transitions and physiological functions

Correlative relationships of the data sets can be approached by principal component analysis (PCA) and hierarchical clustering (HC) using global expression values as input. In the PCA analysis of all samples, a preferential grouping related to development or physiological functions over gametophytic or sporophytic identity is observed (Fig. 3a). In the first principal component of the PCA, samples are separated into two groups: a first group composed of tip growing cells together with gametophore, and a second group composed of sporophytic tissues together with archegonia and spores. In the second principal component (PC2) archegonia, S1, S2 and S3 samples are grouped in a tight cluster, while SM and spores are shown separately. In addition, PC2 indicates that caulonema is different from rhizoids and chloronema (Fig 3a). However, since caulonema was grown under dark conditions to allow its physical separation from chloronema, it

cannot be ruled out that some of the variation comes from differences in experimental growth conditions. Our global analysis though is suggestive that no such major impacts may exist. The HC dendogram confirms the separation of samples that was previously observed in the PCA. Importantly, biological replicates are grouped together, underlining our successful separation of developmental stages, in particular during sporophyte development (Fig 3b).

Gene expression in *P. patens* suggests a lower generation-biased specialization than that observed for *A. thaliana*

To obtain an overview of the transcriptional diversity between all tissues, we determined the number of genes that were expressed (present), enriched and preferentially expressed (those which are only expressed in a particular tissue and are absent from the rest) in each tissue. To calculate the present/absent expression threshold we performed qRT-PCR experiments on a set of genes with low expression values (close to random probe values). We correlated the detection of the transcript of a gene with its corresponding expression value reported by the array, establishing the minimum value for calling a gene present (Supplemental fig 2). We found that there is a high and constant number of genes being expressed in all tissues (Table 1), ranging from 17,468 in spores to 19,169 in archegonia, corresponding to 54% and 59% of the total number of genes represented in the array, respectively (Fig 3c). Thus, in accordance with previous observations in *Funaria hygrometrica* (Szövényi et al., 2011), differentiation in gene expression between the two generations is weaker in *P. patens* than in the angiosperm *A. thaliana,* where the gametophyte expresses only half the number of genes expressed in the sporophyte (Pina et al., 2005). Furthermore, using our present/absent threshold we performed additional analysis to determine the number of preferentially expressed genes in each generation. We found that in *P. patens* the gametophyte has more than twice the number of “uniquely” expressed genes than the sporophyte (Fig 3d). In this case, this observation is in accordance with the higher diversity of tissues that comprises the gametophytic generation in mosses and its dominant status over the sporophyte.

At a tissue level, DEGs are potentially very important for defining its identity and development. We took advantage of our large data set and performed pairwise comparisons of each sample against all the rest. This allowed us to identify genes that were consistently differentially up-regulated across all comparisons, generating a list of tissue enriched genes (TEGs). The tissue with the highest number of TEGs was the spores with 635. This was followed by caulonema and archegonia with 517 and 289 TEGs, respectively (Table 1 and fig 3c). It should be noted that most of the genes enriched in archegonia were also preferentially expressed, suggesting this tissue has the most unique transcriptional program in the gametophytic generation, which might be related to its role in reproduction. Regarding the sporophyte, we observed that the number of enriched and preferentially expressed genes increased substantially as the structure developed and became mature (Fig 3c and table 1).

Transcription factors controlling diploid development and general plant architecture in angiosperms are expressed in several sporophytic stages

The dominance of the sporophytic life cycle in vascular plants is marked by an increased complexity in body plan architecture that is controlled by a set of transcription factors from which *KNOX*, *BELL1*, and *TCP* families are particularly important. *KNOX2* transcription factors, for example, have been shown to be essential for the alternation of generations in mosses, since they prevent the haploid genetic program to be expressed (Sakakibara et al., 2013). Moreover, in the sporophyte of vascular plants interactions between *KNOX* and *BELL1* are known to occur. Such interactions modify the way in which these transcription factors control shoot apical meristem activity and maintenance, which ultimately generates a huge diversity of plant body forms and sizes. The expression data indicates that *KNOX2* class proteins *MKN1*, *MKN6* and one *BELL1-like* homolog (Pp1s258_6v6) are expressed at similar levels throughout sporophyte development in *P. patens* (Fig 4), consistent with the hypothesis that these proteins can interact in mosses. In addition, we found two *TCP* homologs (Pp1s332_35v6 & Pp1s207_110v6) expressed in all sporophytic stages at very high levels, with a

peak of expression in SM. In angiosperms, *TCP* genes are also important regulators of sporophyte architecture because they control axillary bud development, branching, and shoot symmetry (Doebley et al., 1997; Aguilar-Martínez et al., 2007; Koyama et al., 2007).

An abundantly expressed TCP family member represses sporophyte branching

The abundance of *PpTCP* transcripts were the greater than any other transcription factor transcript expressed in the sporophyte of *P. patens*; they were almost as abundant as constitutively expressed *α-tubulin*. Phylogenetic analyses have shown that TCP genes are present both in the genomes of basal land plant groups and in their algal relatives (Navaud et al., 2007). They are grouped into two major subfamilies: class I, also known as *TCP-P*, and class II, also known as *TCP-C*. In *A. thaliana,* class I genes promote general cell growth and proliferation, and class II genes control shoot branching, lateral organ development and flower symmetry through the regulation of tissue proliferation patterns (Navaud et al., 2007; Martín-Trillo and Cubas, 2010). Furthermore, there is an inverse relationship between the amount of *TEOCINTE BRANCHED 1 (TB1)* gene product (a class II *TCP*) and the degree of maize sporophyte branching (Doebley et al., 1997). Using *A. thaliana* and *P. patens TCP* amino acid sequences we performed further phylogenetic analysis and determined that *P. patens* has two *TCP* class II homologues, and four class I homologues (Supplemental fig 3). The class II genes, *PpTCP5* and *PpTCP6*, are transcriptionally enriched in the sporophyte (Fig 5A), while class I genes are expressed significantly less and they tend to be more expressed in the gametophyte.

To investigate the function of class II genes in *P. patens,* we generated five independent knock out mutant lines of *PpTCP5* (Pp1s332_35V6) by homologous recombination. Replacement of the PpTCP5 coding region and integration of the resistance cassette by homologous recombination was verified by PCR (Supplemental fig 9). Three lines were selected for further phenotypic analysis

(designated as *Pptcp5(5), Pptcp5(8) and Pptcp5(27)*). To confirm the absence of *PpTCP5* transcript in mutant lines, we isolated RNA from *WT* and mutant sporophytes and performed RT-PCR experiments. *PpTCP5* and *PpTCP6* transcripts were amplified in *WT* sporophytes, but no *PpTCP5* transcript was detected in mutant lines (Fig 5C, supplemental figure 9). Importantly, *PpTCP6* transcript levels were roughly the same between *Pptcp5(5), Pptcp5(8)* and *WT*, i.e. *PpTCP6* transcript levels did not increase in the mutant lines to compensate for the loss of *PpTCP5* function. Finally, class I genes were never expressed in *WT* or mutant sporophytes at significant levels (Fig 5C).

The phenotype of the *Pptcp5* mutants was compared to *WT* upon transfer to sporophyte-inducing conditions. Since some mutant sporophytes branched - sporophyte branching was defined as a sporophyte bearing two or more capsules connected to the gametophyte by the same foot (Fig 5B, middle panels) - we compared sporophyte number and the percentage of branched sporophytes of Pptcp5 mutants and *WT* (and at least four biological replicates for each line were included). All mutant lines developed more than twice as many branched sporophytes as *WT*. Branching frequencies in *WT* were approximately 5%, while branching in *Pptcp5(27)* was close to 15% (Fig 5D). Furthermore, sporophytes with two or more branches were always more frequent in *Pptcp5* lines than in wild type, with some individuals producing as many as 5 capsules on a single structure (Fig 5B, lower row, middle panel).

Branching originated from the seta region once the first sporangium was almost mature in each *Pptcp5* mutants sporophyte (Fig 5B, upper left panel). The second sporangium developed from the seta region close to the initial sporangium (Fig 5B, lower left panel), suggesting that areas with meristematic competence exist within the seta which can give rise to several capsules once *TCP* expression is suppressed. These data suggest that their role in repressing lateral organ primordia proliferation has been conserved across the plant kingdom, as reported for *Zea mays* (Doebley et al., 1997), *A. thaliana* (Koyama et al., 2007; Aguilar-

Martínez et al., 2007) and now *P. patens*. Since meristem proliferation was necessary for shoot branching we hypothesize that *TCP* transcription factors might have been involved in the control of and the evolution of body plan architecture in land plants.

Treatment of *P. patens* with the auxin transport inhibitor 1-N-naphthylphthalamic acid (NPA) also results in the development of branched sporophytes (Fujita et al., 2008). Furthermore, Bennett et al. (2014) showed that more branches form on sporophytes in *PppinA* and *PppinB* loss of function mutants which lack functional PIN auxin transport proteins than on *WT*. Furthermore, the branching frequency is similar in the *PppinA*, *PppinB* and *Pptcp5* mutants, suggesting a possible relationship between auxin transport and PpTCP function. To determine if there is correlation between *PpPINA*, *PpPINB* and *PpTCP5* transcritpion, we searched for the expression pattern of PpPINA and PpPINB using the *P. patens* eFP browser. Both genes are highly expressed from early stages of sporophyte development, reaching a peak of expression during S2 and S3 (Supplemental fig 4). PpTCP5 transcript abundance increases sharply from the S2 stage onwards, reaching a maximum level during SM. Therefore, *PpPINA* and *PpPINB* are expressed just before *PpTCP5*. These data are consistent with the hypothesis that auxin acts upstream of *PpTCP5* activity in the control of sporophyte branching.

GO analysis of sporophyte development reflects the major transitions leading to spore maturation

To assess the functional role of genes with enriched expression in the tissues analyzed, we performed a gene ontology (GO) enrichment analysis using DAVID functional annotation tool (Huang et al., 2009). This tool clusters genes based on GOs terms, and statistically assigns significant functional categories to related groups of genes. During the S1 stage, regulation of transcription is the only functional category identified. This was due to the fact that there are only 17 genes with enriched expression at this stage; however, most of them are transcription factors and they might be responsible, at least in part, for initiating the sporophytic

transcriptional program. In S2 stage relevant functional categories are lipid catabolism, shoot morphogenesis and to a lesser extent, transcription factor activity (Fig 6). During S3 and SM stages the number of enriched genes and the diversity of functional categories increase substantially (table 1). In S3 GO terms can be related with the biogenesis of spores, since two of the most highly enriched categories are gametophyte development and meiosis I. This finding is relevant and suggests that during this specific stage sporogenesis is initiated. Furthermore, lipid biosynthetic processes are highly represented in S3, indicating that lipid metabolism is also important for spore biogenesis (Fig 6). Finally, associated to SM are categories like fatty acid biosynthesis, response to ABA, oxylipin biosynthesis, iron ion binding, carbohydrate metabolism, and sugar transmembrane transport (Fig 6). Taken from this angle, this developmental stage seems to be committed to the transport and storage of energy reserves and to the growth and maturation of spores.

Homologs of transcription factors controlling vernalization and stomata development in *A. thaliana* are expressed early in sporophyte development

During S1 all annotated genes identified are related with regulation of transcription. Of special interest is one transcription factor homolog to the *A. thaliana REDUCED VERNALIZATION RESPONSE 1 (VRN1)* gene (Pp1s434_19V6). This transcript seems to be important for early sporophyte development since it is highly enriched during this stage and expressed at very low levels during subsequent stages (Fig 4). In *A. thaliana*, *VRN* genes mediate the vernalization response, a process by which flowering is accelerated after plants are exposed to cold conditions for a period of time. In cereals however, expression of *VRN1* can increase without vernalization in response to long days, meaning it has a role in regulating reproductive meristem identity that is not limited to vernalization (Trevaskis et al., 2007). The expression of one putative ortholog in the early sporophyte stages of *P. patens* suggests that it is also important for maintaining meristem activity/identity in this species. In this respect it is important to note that many mosses require

exposure to cold temperatures for the formation of reproductive structures and that consequently their young sporophytes develop under such cold conditions.

Other transcription factors enriched at this stage seem to be involved in the differentiation of water transport specialized cells, like *PpVNS6* and *PpVNS4*, two *NAC* transcription factors that have been shown to be involved in the formation of hydroid and stereid cells in the gametophore leaves, and in the differentiation of central and transfer cells in the sporophyte foot respectively (Xu et al., 2014). In addition, during the S2 stage we found that the transcription factor *PpSMF1*, an ortholog of the *bHLH FAMA*, was highly expressed. In *A. thaliana*, *FAMA* and *MUTE* transcription factors are known for controlling guard cell differentiation. In accordance with our data, mosses are known for producing stomata only during the sporophytic generation. Furthermore, in an independent study, *PpSMF1* transcript was also detected early in sporophyte development (O'Donoghue et al., 2013).

Expression of pollen development homologs mark the onset of sporogenesis in stage S3

Sexual reproduction in plants relies on meiotic cell divisions producing the haploid cells that will give rise to the gametophyte generation. In *P. patens,* meiosis occurs at some point during sporophyte development, giving rise to the spores. We identified a transcription factor homolog to *MALE MEIOTIC DEATH 1 (MMD1)* (Pp1s271_42V6), which is preferentially expressed during S2 and S3 stages of sporophyte development. In *A. thalaina, MMD1* mutants exhibit alterations in meiosis that result in male meiocyte arrest and cell death (Yang et al., 2003). Our data shows that this transcription factor starts to be expressed during S2 and reaches a very high expression level during S3 (Fig 4). Importantly, the transcript could not be detected in the subsequent SM stage or in any other tissue, with the exception of archegonia, were it seems to be expressed at very low levels. In addition, homologs of the meiotic recombination protein *DOSAGE SUPPRESSOR OF MCK1 (DMC1)* (Pp1s9_248V6), which catalyzes the DNA strand exchange

reactions during meiosis in several organisms (Pradillo et al., 2014), and of *SPORULATION 11 (SPO11)* (Pp1s248_24V6), which is responsible for the double strand breaks that initiate meiotic recombination (Bergerat et al., 1997; Keeney, 2007) are also preferentially expressed during the S3 stage. Finally, we also found two homologs of *MALE STERILITY 1 (MS1)* (Pp1s271_42V6) being enriched during this stage. In *A. thaliana, MS1* controls the formation of pollen cytoplasmic components and belongs to the same transcription factor family as MMD1 (Ito et al., 2007). These observations provide strong evidence for meiosis control, and therefore suggest that sporogenesis starts during the S3 developmental stage. Furthermore, it seems that some genes controlling meiosis in *P. patens* are homologous to the ones controlling meiosis in pollen mother cells.

The transcriptome of tip growing cells reveals gametophyte developmental processes

Since tip growing cells represent a major stage of the haploid generation, we analyzed their transcriptome in order to provide insights into important gametophytic developmental processes. The set of genes exclusively enriched in each tip growing cell type was determined to assess, if classes related to specific functions could be identified. Enriched GO categories for rhizoids were specially revealing because they were all related with ion transport, including cation and anion transport, potassium transport, and phosphorous metabolism (Fig 7), suggesting rhizoids are involved in the uptake and exchange of different ions from the surrounding media, probably with the purpose of supplying the gametophores with inorganic nutrients to sustain growth. Transcripts of genes related with potassium, phosphate and iron transport were especially abundant. In contrast, the most enriched categories in chloronema are ethylene mediated signaling pathway, tetrapyrrole binding, metal ion binding, and response to light, reflecting the role of this tissue in the generation of energy through photosynthesis and carbon assimilation (Fig 7). According to this analysis ethylene signaling is likely to be involved in chloronema growth and development. Finally, in caulonema we found that physiological processes related with tip growth were more active, with the

most enriched categories being: cell wall modification, regulation of cell size, glucan metabolic process, and carbohydrate catabolism (Fig 7). In accordance with this observation, it has been shown that caulonema grow significantly faster than chloronema (Duckett et al., 1998; Menand et al., 2007a), reinforcing the idea that this enrichment in tip growth related processes would support its faster growth rate. Expression data for caulonema and chloronema has also been compared by Xiao et al. (2011), and functional annotation of significantly enriched transcripts in caulonema suggested alternative roles like ion binding, protein binding and RNA binding. In contrast to our results most of these categories are not representative of tip growing processes and no cell wall-related functions were identified. It is also noteworthy that we could identify 2602 genes with enriched expression in caulonema versus chloronema samples, while in the study by Xiao et al. (2011) 223 genes were identified, and that the overlap between the two lists was only 24 genes (Supplemental table 3). Apart from technical and data analysis differences in these two studies, we believe that the discrepancies might be explained by the different conditions in which caulonema were grown. In this study caulonema was induced under dark and RNA was isolated from highly differentiated cells, as evidenced by the number of cells with caulonemal identity (at least four) observed after the last chloronemal cell of each filament (Supplemental fig 7). Indeed, it has been established that darkness and low nutrient availability strongly promote caulonema differentiation and growth, since its role is mainly related to the expansion of the protonemal colony into new, possibly more favorable environments (Cove et al., 1978; Cove et al., 2006). In contrast, samples from Xiao et al. (2011) were grown in full nutrient medium and under 16/8 hour (light/dark) photoperiod, which resulted in caulonemal cells that appear considerably less elongated, and contain more chloroplasts. Therefore, we propose that the two studies represent different stages in caulonemal differentiation, and that they should be seen as complementing datasets which can provide relevant insights into caulonemal developmental transition.

Homologs of highly expressed genes in tip growing cells became pollen tube or root hair specific in *A. thaliana*

To identify candidate transcription factors controlling general tip growth functions in *P. patens,* we searched for genes that were highly expressed in all three types of tip growing cells: caulonema, chloronema and rhizoids. Most of the tip growing cell enriched (TGE) transcription factors (TF) belonged either to the *bHLH* or to the *MICK** class of the MICK *MADS-box* subfamily. Several members of these families are known to control root hair and pollen tube development, respectively, in *A. thaliana* (Menand et al., 2007b; Karas et al., 2009; Adamczyk and Fernandez, 2009; Yi et al., 2010). TGE-TFs found in *P. patens* belonging to the *bHLH* family include *PpRSL1* and *PpRSL2*, orthologs of *ROOT HAIR DEFECTIVE 6 (RHD6)* and *ROOT HAIR DEFECTIVE SIX LIKE 1 (RSL1)*, respectively (Menand et al., 2007b), and two homologs of the *LJ-ROOT HAIRLESS LIKE (LRL-like)* transcription factors. According to our data the expression pattern of these genes is different in *P. patens* compared to *A. thaliana,* since in the latter their expression is limited to only one kind of tip growing cell: the root hairs. In terms of biological function, *PpRSL1* and *PpRSL2* are essential for caulonema and rhizoid development in *P. patens* (Jang and Dolan, 2011), while orthologs in *A. thaliana* are necessary for root hair development with no role on pollen tube growth, meaning their activity is restricted to only one type of tip growing cell (Menand et al., 2007b; Pires et al., 2013).

A similar example can be found in the MICK* transcription factors, since in *A. thaliana* all MICK* genes except one, are exclusively expressed in just one type of tip growing cell: the pollen tube (Verelst et al., 2007). Again, we found that 10 out of the 11 *MICK** proteins encoded in the genome of *P. patens* are expressed in all tip growing cells. Moreover, while three were preferentially expressed in caulonema, chloronema and rhizoids, the remaining 8 were also expressed in the spores and in mature sporophyte at lower levels (Supplemental table 4). Importantly, the same genes detected as expressed in mature sporophyte were also detected in isolated spores at higher levels, thus the presence of *MICK** mRNA in our mature sporophyte samples might be a consequence of the high

number of spores being contained inside the sporangium from which RNA was isolated. This expression profiles are in according with previous studies done in the moss *Funaria hygrometrica*, the lycophyte *Selaginella moellendorffii* and the fern *Ceratopteris richardii* where *MICK** proteins have been found to be predominantly expressed in the gametophyte (Zobell et al., 2010; Kwantes et al., 2012; Hohe et al., 2002). We performed a phylogenetic analysis using the aligned amino acid sequences from *P. patens* and *A. thaliana MICK** proteins. We found that all *P. patens* proteins formed a single clade that is sister to the *A. thaliana* S-class protein clade comprising *AGL66, AGL67* and *AGL104* (Supplemental fig 5). This is of particular interest, because loss-of-function alleles from the S-class in *A. thaliana* produce the strongest pollen defective phenotype, with delayed germination and aberrant pollen tube growth (Adamczyk and Fernandez, 2009). Functional studies in *P. patens* are lacking, but it is likely that *MICK** genes have a general function in controlling tip growth related processes in the moss, that has become more specific in vascular plants.

Taken together, these observations suggest that the transcriptomes of chloronema, caulonema and rhizoids are more similar to each other (i.e., the three tissues express the same core of TGE-TFs) than angiosperm tip growing cells. We found further evidence supporting this hypothesis by comparing our total list of TGE genes in *P. patens* with a similar list generated in a separate study for *A. thaliana* root hairs and pollen (Becker et al., 2014). In the case of *P. patens* we found 142 gene transcripts that were both expressed and enriched in all tip growing cells (Supplemental table 5), while in *A. thaliana* a total of 49 were found (Becker et al., 2014). The lower number of TGE genes reported for the later can be explained by the fact that pollen and root hairs do not express as many common genes as *P. patens* tip growing cells do, however, tissue diversity is higher in *A. thaliana* which can also influence the final number of TGE genes. On the other hand, when pollen and root hairs are individually compared against the rest of the tissues, larger numbers of enriched genes can be obtained, supporting our aforementioned hypothesis.

Since it has been proposed that the expression of part of the genetic network controlling root hair growth was recruited from the gametophyte generation through evolution (Menand et al., 2007b), it is possible to imagine a scenario that would explain our observations, in which root hairs only partially acquired the expression of an ancestral tip growing cell developmental gene network. This would explain why they do not express the same genes as pollen. On the other hand, pollen tubes probably suffered a “shut down” of the expression of several genes that comprised such core, including for example the *bHLH* transcription factors *PpRSL1* and *PpRSL2* mentioned earlier (assuming these genes were also expressed in all tip growing cells from ancestral bryophyte plants). These observations are relevant because they support the antithetic theory of land plant evolution.

The transcriptome of spores indicates extensive transcriptional and epigenetic reprogramming

Spores and pollen grains can produce tip growing cells upon germination and can be considered analogous structures to some degree. However, while the transcriptomes of pollen grains and pollen tubes show only moderate alterations in *A. thaliana* (Qin et al., 2009), we found that the transcriptome of mature spores is very different from that of *P. patens* tip growing cells. In fact, spores were transcriptionally unique, since we found several families of transcription factors uniquely enriched at this stage (Fig 8). Highly enriched gene transcripts in spores were mainly involved in RNA metabolism. Some relevant categories were non-coding RNA (ncRNA) processing, ATP-dependent helicase activity, RNA splicing and DNA repair. In particular, several genes involved in both mRNA and sRNA processing were detected as highly enriched. These include a homolog of *HEN1 SUPRESSOR1 (HESO1)* (Pp1s131_38v6) that uridilates miRNAs, small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs) in *A. thaliana* (Zhao et al., 2012; Ren et al., 2012); several RNA methytranferases and helicases, as well as several splicing factor homologs (see supplemental table 1). Furthermore, we found transcripts encoding proteins involved in DNA chromatin modification and methylation, suggesting that such processes are active in the spores. Some

examples are the homolog of *CHROMATIN REMODELING 18 (CHR18)* (Pp1s197_19v6) which belongs to the *SNF2* chromatin remodeling family, an homolog of the *SSRP1* (Pp1s199_36V6) subunit from the *FACILITATES CHROMATIN TRANSCRIPTION (FACT)* complex required for DNA demethylation by *DEMETER (DME)*, and the homolog of the *RNA DIRECTED DNA METHYLATION 4 (RDM4)* gene (Pp1s70_268v6), which plays a role in transposon silencing in *A. thaliana* (Knizewski et al., 2008; Ikeda et al., 2011; He et al., 2009). Transcription seems to be highly active in plastids as well. Several mitochondrial transcription terminator factors (*mTERF*) important for initiation and termination of transcription, both in chloroplast and mitochondria (Kleine, 2012), are highly expressed (Supplemental table 1). Taken together, these observations suggest that processes involved in the modification of gene expression are highly active and dynamic, including epigenetic modifications. It is puzzling to observe this pattern of gene expression in the mature spores, but not in any other tissue nor in the transcriptome of pollen (Table 2). One hypothesis would be that extensive changes in the transcriptional program are required to achieve a transition between an active and a dormant metabolic state. In angiosperms seeds also have to go through a similar transition, and it has been established that genes involved in chromatin remodeling processes, including methylation, play a role in such rearrangements (van Zanten M., Liu Y., 2013).

Materials and Methods

Plant Material and Growth Conditions

The *Gransden* wild type strain from *Physcomitrella patens* Bruch & Schimp (Ashton and Cove, 1977) was used for this study. Plant material was routinely grown on Petri dishes containing KNOPS media (Reski and Abel, 1985) supplemented with 0,5g/L ammonium tartrate dibasic (Sigma-Aldrich Co) and 5 g/L glucose (Sigma-Aldrich Co) at 25ºC, 50% humidity and 16h light (light intensity 90-100 µmol m-2 s-1). Vegetative propagation was maintained by subculture every 6-7 days by mechanical disruption of plant tissue using a tissue disruptor (TissueRuptor,

Quiagen) leading to predominant growth of chloronema. For gametangia and sporophyte development, protonema was cultured on sterile peat pellets (Jiffy-7, Jiffy Products International B.V.) in plant culture boxes during 3-4 weeks (Sakakibara et al., 2008). Water was supplied to the bottom of each box containing 4 pellets and samples were transferred to 17ºC, 8h light and 50% humidity (light intensity 80-85 µmol m-2 s-1) to induce the development of reproductive structures (Hohe et al., 2002). Further development of the sporophyte was conducted at these conditions.

Assessment of gametangia and sporophyte development

The development of gametangia after cold induction was monitored daily. At day 14 after induction mature gametangia were detected. However, the opening of the archegonia and the release of antherozoids was preferentially observed 24 hours later (Supplemental fig 6), suggesting that fertilization events take place from day 15 onwards. Therefore, we considered 15 days after induction as the moment in which fertilization occurs. However, the occurrence of more “fertilization waves” cannot be ruled out due to the presence of gametangia in clusters of differently staged organs (Landberg et al., 2013). The development of the sporophyte under our laboratory conditions was further determined based on morphological characteristics and divided into discrete stages defined by days After Fertilization (AF): S1 (7 days AF), S2 (15 days AF), S3 (20 days AF), SM-1 (28 days AF) and SM (33 days AF) (see Figure 1).

Tissue Isolation

Triplicates for the different tissue samples were isolated from wild-type *Gransden* strain according with the particularities of each tissue. Archegonia: archegonia at 15 days after short day conditions were isolated manually under the stereoscope (Nikon, SMZ800) and placed into Trizol. At this point maturity was ensured and only closed organs were selected. 300 archegonia were used per replicate to isolate RNA. Sporophyte: sporophytes from the different stages S1, S2, S3 and SM were manually collected. Induced gametophores at the specific time points

described above were dissected under the stereoscope and only the sporophytes fitting to the desired morphological stage were isolated (Figure 1). At stage S1 the separation from archegonia tissue is not possible with our isolation methods and in order to maintain the comparative capabilities of our set, the residual archegonium was attached to all sporophytes collected. Samples were stored directly in Trizol. Approximately 25 sporophytes from each developmental stage were used for RNA extraction. Spores: spores from 3-5 fully mature spore capsules were released by resuspension on 1 mL dH2O and filtering through 41µm nylon mesh by centrifugation 5 min at 13,400 rcf in eppendorf tubes. A subsequent filter-resuspension step was done on 27 µm nylon mesh and two washing steps were performed by centrifugation for 5 min at 13,400 rcf. A final centrifugation to pellet the samples was done 5 min at 16,100 rcf and spores were resuspended in 150 µm MilliQ water. Purity of the sample was confirmed by microscopy before proceeding to RNA extraction. Tip growing tissues: Caulonema development was induced by growing protonema under dark conditions in vertical KNOPS plates for 5 days (Supplemental fig 7a), while chloronema was produced under normal conditions (16h light/8h dark). Both filaments were cut manually under a stereoscope (Nikon, SMZ800) and identity was confirmed by looking at the cell division plates (Supplemental fig 7c). No tmema was included. Tissue was directly placed in Trizol until RNA extraction. Rhizoids were harvested from gametophores grown in Magenta boxes (Sigma-Aldrich) containing KNOPS minimal medium (Reski and Abel, 1985) under normal conditions (16h light/8h dark). Around 100 gametophores per replicate were isolated and the entire rhizoids were manually dissected with a razor blade under a stereomicroscope (Supplemental fig 7b). Isolated rhizoids were frozen in liquid nitrogen right after dissection. Gametophores: Gametophores grown on sterile peat pellets (Jiffy-7, Jiffy Products International B.V.) for 3 weeks were isolated manually and placed on Trizol for subsequent RNA extraction. Protoplast: Protoplasts were isolated following an existing protocol (Schaefer et al., 1991) with small modifications. Briefly, 5 days old protonema tissue grown on KNOPS media was digested in 1 % (w/v) Driselase (Sigma-Aldrich Co.) in 0.36M Mannitol for 30 min. Tissue was filtered through a 80

µm stainless steel sieve and transfered to a new 50 ml tube for centrifugation (85 rcf, 5 min). The supernatant was carefully removed and the pellet was gently resuspended in 10 mL Mannitol (0.36M) solution. Centrifugation and resuspension were repeated and protoplast number was estimated using a hemocytometer. 300,000 cells were used per replicate for RNA extraction.

RNA Isolation and cDNA synthesis

RNA from all samples was isolated using Direct-zol® columns (Zymo Research Co.) following manufacturer's instructions. Samples were treated with TURBO DNase at 37°C for 30 min (Ambion, Life Technologies). RNA integrity and quantity were assessed on an Agilent 2100 Bioanalyzer using a 6000 Pico Assay (Agilent Technologies). On average 7.5 ng of RNA were used to synthetize cDNA with an Ovation Pico WTA System V2 amplification kit (NuGen Technologies, Inc.). The cDNA concentration obtained from all samples were in the same range (180-260 ng/µl) and further quality controls were performed both by Bioanalyzer and by qPCR using several housekeeping and tissue specific probes.

cDNA labeling and hybridization of Nimblegen arrays

750 ng of cDNA was used for labeling and hybridization on custom Nimblegen 12x135K arrays (Roche NimbleGen, Inc.) following manufacturer's instructions in the IRB Barcelona Functional Genomics Core Facility (FGC). Arrays were scanned and raw data was obtained using the DEVA software, applying Robust Multichip Average (RMA) normalization to all arrays (Roche NimbleGen, Inc.).

Data Analysis and Functional Annotation

Differential expression analysis was conducted using dChip software (Li and Wong, 2001). Normalized data was imported as External Data in tab-delimited text format, including expression data and standard error as well as presence/absence calls (see below). In order to determine enriched and preferentially expressed genes pair-wise comparisons were conducted. A lower-confidence bound fold-change (LCB FC) cut-off was estimated for each comparison, ranging from 1.2 to 4

and maintaining a false discovery rate (FDR) below 10%. Transcripts that were called present in a tissue and identified as significantly up-regulated in every single comparison of this tissue against the rest were identified as enriched, while transcripts with a present call in only one tissue and absent call in the rest were defined as preferentially expressed. It is important to note that comparisons among all sporophytic stages were performed with fold change values ranging from 1.3 to 2.8 and with false discovery rates (FDR) below 10%, demonstrating that it was possible to obtain discrete developmental stage units. Due to high similarity between protoplast and chloronema transcriptomes, protoplast samples were excluded from all subsequent analyses. However, protoplast data is available in the eFP browser.

Present and absent calls were generated experimentally after determining signal thresholds for each tissue sample. In order to establish this threshold, primers for transcript models with expression values of 150, 200 and 260 were designed for each tissue type and amplification was tested by qPCR (ABI7900HT, Applied Biosystems), using 1 ng of cDNA per reaction. Physcomitrella alpha tubulin (Pp1s215_51V6) was used as housekeeping gene. Since a similar trend was observed, an average value from all tissues was applied to the microarray set. A criterion of majority vote was applied, meaning that at least 2 out of 3 replicates (>66% presence calls) should be above the limit for a transcript to be called present. Our comparisons indicated that the protoplast transcriptomes were very similar to the chloronema ones.

In addition, present, enriched, and preferentially expressed gene lists for several *A. thaliana* tissues were also generated using dChip for comparing tissue expression profiles with those of *P. patens* analogous tissues. Previously available *A. thaliana* microarray data sets were used for this analysis (Swanson et al., 2005; Qin et al., 2009; Borges et al., 2008; Pina et al., 2005; Becker et al., 2014). Hierarchical clustering and PCA were calculated using Chipster (Kallio et al., 2011). *P. patens* genome annotation v1.6 release 2012.3 was used (cosmoss.org), combined with information from several alignments against non-plant organisms and STRING

information (Franceschini, et al. 2013). Functional enrichment analysis was done using DAVID Bioinformatics Resources 6.7 (Da Wei Huang, Brad T Sherman, 2009) based on *A. thaliana* homologs. Genes were clustered according to their GO terms, including biological process, molecular function and cellular component, and enrichment scores were calculated based on the EASE scores (modified Fisher exact P-value) of each cluster term members. A P-value<0.05 (EASE score) was established as a cutoff for the selection of annotation clustering terms. Default parameters were used for the generation of enriched functional clusters for all samples except S2 and SM. Lower GO term classification stringency gave better results in S2 due to the reduced number of DEGs obtained for this sample, while the opposite was true for SM, the sample analysed with the highest number of DEGs. Detailed information about GO clusters is summarized in supplemental table 6. As a background, we uploaded a custom gene list containing all *P. patens* genes found in the v1.6 genome annotation which had an *A. thaliana* homolog assigned.

Quantitative real-time PCR

To corroborate gene expression of reported tissue specific genes and housekeeping genes, previously synthesised cDNA was used and each sample replicate was tested independently. qPCR reactions were prepared in triplicates with SYBR Green FastMix (Quanta, BioSciences), 300 nM primers and 1 ng cDNA to a final volume of 20 µL. Gene-specific primers were designed (Supplemental table 2) and assessed using OligoCalc (Kibbe, 2007). Alpha tubulin (Pp1s341_23V6) was used as housekeeping reference since low variation in tissue specific expression was observed (Supplemental fig 8). qPCRs were conducted using an Applied Biosystems 7900HT with an initial 10 min incubation at 95ºC followed by 40 cycles of 30s at 95ºC, 30s at 60ºC and 45s at 72ºC, and a final extension of 10min at 72ºC. Thermal ramping stage was included after each run to determine dissociation curve properties. For the determination of presence/absence and selective tissue expression cDNA from the tissue replicates

was pooled and 1ng of the mix was used. qPCRs were conducted using the specifications listed above.

Phylogenetic analysis

Coding sequences of gene sequences were retrieved from the Phytozome v9.1 database (www.phytozome.net). An alignment of the amino acid sequences of *MICK** and *TCP* genes present in *A. thaliana*, *P. patens* and *Z. mays* was constructed using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) program. The UPGMB clustering method was selected. Phylogenetic tree was constructed using Maximum Likelihood analysis and the Jones-Taylor-Thornton (JTT) amino acid substitution model available in the MEGA 6 suite (www.megasoftware.net). Nearest neighbor analyses was used for tree inference. The reliability of the inferred tree was assessed using bootstrap method (with at least 250 replications). The resulting phylogenetic tree was visualized with MEGA 6. For the rest of the gene families mentioned in the manuscript, the term homolog or putative ortholog is used when two genes share a common ancestor gene (according to phytozome phylogenomics resource), but it is unclear whether the genes are paralogs or orthologs. Genes are referred as orthologs exclusively when published phylogenetic analyses support such relationship, and references to appropriate publications can be found in the text.

Generation of knockout *TCP5* mutants in *P. patens*

P. patens TCP5 knockout mutants were generated by homologous recombination (Schaefer et al., 1991), using the G418 resistance cassette to replace the gene of interest. Genomic sequences of 600 bp upstream and 561 bp downstream the *TCP5* (Pp332_35V6) gene, respectively, were amplified by PCR using specific primers with restriction sites added (oAT01 / oAT02 and oAT03 / oAT04). Fragments were cloned into the transformation vector pBNRF by sticky end ligation into the BamHI/XhoI and NdeI/AscI sites (Supplemental fig 9). Polyethylene glycol-mediated transformation was performed (Schaefer et al., 1991; Schaefer, 2001) using 15 µg of the linearized pBNRF-tcp5 DNA. Selection of stable transformants

was conducted on G418 (50 µg/mL) antibiotic plates and confirmed by PCR. Positive transformants showed absence of the *TCP5* gene (primers oAT05/ oAT06; product size 749bp) and the correct integration of the transgene on both 5' and 3' ends (using the primers oAT07 / oAT08 and oAT09 / oAT10, respectively) (Supplemental fig 9). Five independent lines were identified as stable knockout mutants for the *TCP5* gene, named *Pptcp5(3), Pptcp5(5), Pptcp5(8) Pptcp5(22) and Pptcp5(27).* Three of them were phenotypically characterized.

Phenotyping *Pptcp5* lines

The *Pptcp5* lines were maintained and grown in the same conditions as described for wild type. Induction of gametangia was conducted after 28 days of growth on sterile peat pellets (Jiffy-7, Jiffy Products International B.V.). The phenotyping of the lines was done once the sporophytes reached the mature stage (6 weeks after gametangia induction). Fertility rates (percentage of gametophores with sporophyte) were assessed by counting successfully developed sporophytes in at least 100 randomly collected gametophores. Previous characterization of fertility rates in *WT* showed that under our conditions sporophyte production rates were between 43-60%. Furthermore, we observed that samples with lower production rates always showed defects like fungus or bacteria contamination, or low humidity. Therefore samples from both genotypes with fertility rates below 43% were discarded for statistical analysis, considered as stress affected.

RT-PCR of *PpTCP* genes

Wild type and mutant sporophytes from S3 and mature stages were collected directly into Trizol for RNA extraction using Direct-zol® columns (Zymo Research Co.) and DNase treatment was performed following manufacturer's specifications. cDNA was synthesized with SuperScript III first strand kit (Invitrogen) using oligodT primers. For PCR amplification, specific primers for genes *TCP5* (Pp1s332_35V6) and *TCP6* (Pp1s207_110V6) were designed with product sizes of 331 bp and 511 bp respectively. For genes *TCP1-4* (Pp1s67_183V6, Pp1s446_21V6, Pp1s356_40V6, Pp1s348_6V6) primers on a conserved domain were used giving

product sizes ranging from 316 bp to 343 bp. As control alpha tubulin (Pp1s341_23V6) was amplified with a product size of 1570 bp in genomic DNA and 773 bp in cDNA for mutant lines *Pptcp(5)* and *Pptcp5(8)* and 106 bp for mutant lines *Pptcp5(3)* and *Pptcp5(27)*, respectively. PCR amplification cycles varied among samples according with the starting material used for cDNA synthesis. In all cases unsaturated PCR reactions were conducted after testing cycle number.

Accession Number

Microarray data are available in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-3069.

Author Contributions

C.O-R., M.H-C., J.A.F., L.D. and J.D.B. conceived and designed the experiments. C.O-R., M.H-C., A.T. and B.C. performed the experiments. C.O-R., M.H-C., M.W. and J.D.B. analyzed the data. C.O-R., M.H-C., J.A.F. and J.D.B. wrote the article.

Acknowledgments

We gratefully acknowledge funding by FP7-PEOPLE-ITN-2008 “PLANT developmental biology: Discovering the ORIGINS of form (PLANTORIGINS)” to C.O-R., M.H-C., A.T., B.C., L.D., J.A.F. and J.D.B. Part of the results have been achieved within the framework of the 2nd call ERA-NET for Coordinating Plant Sciences, with funding from Fundacão para a Ciência e a Tecnologia-FCT, (Portugal) through grant ERA-CAPS/0001/2014 to JDB. JAF acknowledges funding from FCT through grants PTDC/BEX-BCM/0376/2012 and PTDC/BIA-PLA/4018/2012. We thank Stefan Rensing (University of Marburg, Germany) for allowing us to use the Nimblegen_Ppat_SR_exp_HX12 array design and the IRB Functional Genomics Core (Barcelona, Spain) for microarray processing. Krzysztof Kus (IGC, Portugal) is acknowledged for his help with GO annotations and Thomas Tam (University of Oxford, UK) for his introduction to *Physcomitrella* methods. We thank Manuel Hiss and Mareike Schallenberg-Rüdinger (University of Marburg) for fruitful discussions, and Nicholas Provart and Asher Pasher (University of Toronto,

Canada) for their help setting up the eFP browser. Joao Sobral (IGC, Portugal) is acknowledged for his excellent technical assistance.

References

Adamczyk, B.J. and Fernandez, D.E. (2009). MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in Arabidopsis. Plant Physiol. **149**: 1713–23.

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

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

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

Bennett, T.A. et al. (2014). Plasma membrane-targeted PIN proteins drive shoot development in a moss. Curr. Biol. **24**: 2776–2785.

Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P.C., Nicolas, A., and Forterre, P. (1997). An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature **386**: 414–417.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J.A., and Becker, J.D. (2008). Comparative transcriptomics of Arabidopsis sperm cells. Plant Physiol. **148**: 1168–1181.

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

Cove, D.J., Schild, A., Ashton, N.W., and Hartmann E. (1978). Genetic and physiological studies of the effect of light on the development of the moss, Physcomitralla patens. Photochem. Photobiol. **27**: 249–254.

Doebley, J., Stec, A., and Hubbard, L. (1997). The evolution of apical dominance in maize. Nature **386**: 485–488.

Dolan, L. (2009). Body building on land: morphological evolution of land plants. Curr. Opin. Plant Biol. **12**: 4–8.

Duckett, J.G., Schmid, A.M., and Ligrone, R. (1998). Protonemal morphogenesis. In Bryology for the twenty-first century, D.J.G. Bates, J.W., Ashton, N.W., ed (British Bryological Society: Leeds, UK), pp. 233–246.

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

Frank, M.H. and Scanlon, M.J. (2014). Transcriptomic evidence for the evolution of shoot meristem function in sporophyte-dominant land plants through concerted selection of ancestral gametophytic and sporophytic genetic programs. Mol. Biol. Evol. **32**: 355–367.

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

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

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

He, X., Hsu, Y., Zhu, S., Liu, H., Pontes, O., Zhu, J., Cui, X., and Wang, C. (2009). A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development service A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. Genes and Development 23: 2717–2722.

Hiss, M. et al. (2014). Large-scale gene expression profiling data for the model moss Physcomitrella patens aid understanding of developmental progression, culture and stress conditions. Plant J. **79**: 530–9.

Hohe, A., Rensing, S.A., Mildner, M., Lang D. and Reski, R. (2002). Day length and temperature strongly influence sexual reproduction and expression of a novel MADS-Box gene in the moss Physcomitrella patens. Plant Biol. **4**: 595–602.

Ikeda, Y., Kinoshita, Y., Susaki, D., Ikeda, Y., Iwano, M., Takayama, S., Higashiyama, T., Kakutani, T., and Kinoshita, T. (2011). HMG Domain containing SSRP1 is required for DNA demethylation and genomic imprinting in Arabidopsis. Dev. Cell **21**: 589–596.

Ishikawa, M. et al. (2011). Physcomitrella cyclin-dependent kinase A links cell cycle reactivation to other cellular changes during reprogramming of leaf cells. Plant Cell **23**: 2924–2938.

Ito, T., Nagata, N., Yoshiba, Y., Ohme-Takagi, M., Ma, H., and Shinozaki, K. (2007). Arabidopsis MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. Plant Cell **19**: 3549–3562.

Jang, G. and Dolan, L. (2011). Auxin promotes the transition from chloronema to caulonema in moss protonema by positively regulating PpRSL1and PpRSL2 in Physcomitrella patens. New Phytol. **192**: 319-27.

Kallio, M.A., Tuimala, J.T., Hupponen, T., Klemelä, P., Gentile, M., Scheinin, I., Koski, M., Käki, J., and Korpelainen, E.I. (2011). Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics **12**: 507.

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

Keeney, S. (2007). Spo11 and the formation of DNA double-strand breaks in meiosis. Genome Dyn. Stab.: 27–68.

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

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

Kleine, T. (2012). Arabidopsis thaliana mTERF proteins: evolution and functional classification. Front. Plant Sci. **3**: 1–15.

Knizewski, L., Ginalski, K., and Jerzmanowski, A. (2008). Snf2 proteins in plants: gene silencing and beyond. Trends Plant Sci. **13**: 557–565.

Koyama, T., Furutani, M., Tasaka, M., and Ohme-Takagi, M. (2007). TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in Arabidopsis. Plant Cell **19**: 473–484.

Kwantes, M., Liebsch, D., and Verelst, W. (2012). How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. Mol. Biol. Evol. **29**: 293–302.

Landberg, K., Pederson, E.R.A., Viaene, T., Bozorg, B., Friml, J., Jönsson, H., Thelander, M., and Sundberg, E. (2013). The moss Physcomitrella patens reproductive organ development is highly organized, affected by the two SHI/STY genes and by the level of active auxin in the SHI/STY expression domain. Plant Physiol. **162**: 1406–19.

Li, C. and Wong, W.H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc. Natl. Acad. Sci. U. S. A. **98**: 31–36.

Liénard, D., Durambur, G., Kiefer-Meyer, M.-C., Nogué, F., Menu-Bouaouiche, L., Charlot, F., Gomord, V., and Lassalles, J.-P. (2008). Water transport by aquaporins in the extant plant Physcomitrella patens. Plant Physiol. **146**: 1207–1218.

Martín-Trillo, M. and Cubas, P. (2010). TCP genes: a family snapshot ten years later. Trends Plant Sci. **15**: 31–9.

Menand, B., Calder, G., and 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, D.G., and Dolan, L. (2007b). An ancient mechanism controls the development of cells with a rooting function in land plants. Science **316**: 1477–80.

Navaud, O., Dabos, P., Carnus, E., Tremousaygue, D., and Hervé, C. (2007). TCP transcription factors predate the emergence of land plants. J. Mol. Evol. **65**: 23–33.

Niklas, K.J. and Kutschera, U. (2010). The evolution of the land plant life cycle. New Phytol. **185**: 27–41.

O'Donoghue, M.-T., Chater, C., Wallace, S., Gray, J.E., Beerling, D.J., and Fleming, A.J. (2013). Genome-wide transcriptomic analysis of the sporophyte of the moss Physcomitrella patens. J. Exp. Bot. **64**: 3567–81.

Pina, C., Pinto, F., Feijó, J.A., and Becker, J.D. (2005). Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol. **138**: 744–756.

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

Pradillo, M., Varas, J., Oliver, C., and Santos, J.L. (2014). On the role of AtDMC1, AtRAD51 and its paralogs during Arabidopsis meiosis. Front. Plant Sci. **5**: 23.

Qin, Y., Leydon, A.R., Manziello, A., Pandey, R., Mount, D., Denic, S., Vasic, B., Johnson, M.A., and Palanivelu, R. (2009). Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. PLoS Genet. **5**.

Ren, G., Chen, X., and Yu, B. (2012). Uridylation of miRNAs by HEN1 SUPPRESSOR1 in Arabidopsis. Curr. Biol. **22**: 695–700.

Reski, R. and Abel, W.O. (1985). Induction of budding on chloronemata and caulonemata of the moss, Physcomitrella patens, using isopentenyladenine. Planta **165**: 354–358.

Sakakibara, K., Ando, S., Yip, H.K., Tamada, Y., Hiwatashi, Y., Murata, T., Deguchi, H., Hasebe, M., and Bowman, J.L. (2013). KNOX2 genes regulate the haploid-to-diploid morphological transition in land plants. Science **339**: 1067–70.

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

Schaefer, D., Zryd, J., Knight, C., and Cove, D.J. (1991). Stable transformation of the moss Physcomitrella patens. Mol Gen Genet. 226: 418–424.

Schaefer, D.G. (2001). Gene targeting in Physcomitrella patens. Curr. Opin. Plant Biol. **4**: 143–150.

Singer, S.D. and Ashton, N.W. (2007). Revelation of ancestral roles of KNOX genes by a functional analysis of Physcomitrella homologues. Plant Cell Rep. **26**: 2039–2054.

Swanson, R., Clark, T., and Preuss, D. (2005). Expression profiling of Arabidopsis stigma tissue identifies stigma-specific genes. Sex. Plant Reprod. **18**: 163–171.

Szövényi, P., Rensing, S.A., Lang, D., Wray, G.A., and Shaw, A.J. (2011). Generation-biased gene expression in a bryophyte model system. Mol. Biol. Evol. **28**: 803–12.

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

Trevaskis, B., Hemming, M.N., Dennis, E.S., and Peacock, W.J. (2007). The molecular basis of vernalization-induced flowering in cereals. Trends Plant Sci. **12**: 352–357.

Verelst, W., Twell, D., de Folter, S., Immink, R., Saedler, H., and Münster, T. (2007). MADS-complexes regulate transcriptome dynamics during pollen maturation. Genome Biol. **8**: R249.

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

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., and Provart, N.J. (2007). An “electronic fluorescent pictograph” browser for exploring and analyzing large-scale biological data sets. PLoS One **2**: 1–12.

Xiao, L., Wang, H., Wan, P., Kuang, T., and He, Y. (2011). Genome-wide transcriptome analysis of gametophyte development in Physcomitrella patens. BMC Plant Biol. **11**: 177.

Xu, B. et al. (2014). Contribution of NAC transcription factors to plant adaptation to land. Science **343**: 1505–8.

Yang, X., Makaroff, C.A., and Ma, H. (2003). The Arabidopsis MALE MEIOCYTE DEATH1 gene encodes a PHD-finger protein that is required for male meiosis. Plant Cell **15**: 1281–1295.

Yi, K., Menand, B., Bell, E., and Dolan, L. (2010). A basic helix-loop-helix transcription factor controls cell growth and size in root hairs. Nat. Genet. **42**: 264–7.

Van Zanten M., Liu Y., and Soppe, W.J. (2013). Epigenetic signalling during the life of seeds. In Epigenetic Memory and Control in Plants, O.N. Grafi G., ed (Springer: New York), pp. 127–154.

Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T.T., Meyers, B.C., Mo, B., and Chen, X. (2012). The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. Curr. Biol. **22**: 689–694.

Zobell, O., Faigl, W., Saedler, H., and Münster, T. (2010). MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. Mol. Biol. Evol. **27**: 1201–11.

Figure Legends

Figure 1. Developmental stages of the Physcomitrella life cycle analysed by microarrays. A cartoon representation of *P. patens* tissues as integrated into a Physcomitrella eFP browser at the Bio-analytical Resource for Plant Biology (BAR) is shown. In the case of sporophyte developmental stages, representative images of tissue used for mRNA isolation are included (bottom). Stages were denominated as Sporophyte 1 (S1), Sporophyte 2 (S2), Sporophyte 3 (S3) and mature (brown) sporophytes (SM) based on the number of days that passed since the fertilization event (d.AF – days After Fertilization). Early sporophytes stages retain some gametophytic tissue since originally the embryo develops as part of the archegonia. Scale bars = 100 µm.

Figure 2. Assessment of microarray data quality. Quality of tissue sample replicates was tested by qRT-PCR using marker genes. (a) The expression of the gene RM09 (Pp1s407_31V6.1) reported to be present in protonemal tissues was tested. Expression is shown to be enriched in protonemal tissues (caulonema, chloronema, protoplasts and rhizoids), where differences in expression levels among them can also be observed. (b) In the same way, the PIP2;1 aquaporin (Pp1s8_151V6.1) was tested and the opposite pattern was obtained. In this case gametophytic and sporophytic tissues showed expression while the transcript was not detected in any protonemal tissue. Relative expression ($2^{-\Delta\Delta Ct}$) was calculated using alpha-tubulin (Pp215_51V6.1) as housekeeping gene and caulonema or gametophore as the reference tissue. The standard error range (±) is depicted next to the bars. (c) Genes with tissue preferential expression were selected for pattern corroboration by qRT-PCR. cDNA from the three biological replicates was pooled. Data is shown as fold change ($2^{-\Delta\Delta ct}$) relative to gametophore expression in all cases except when the gene is preferentially expressed in gametophore, in which case rhizoids were used. (d) Correlation of gene expression values as reported by our microarrays with experimental evidence of transcript presence for the PpMKN family of transcription factors. Filled circles above the tissue cartoons indicate

which PpMKN genes are expressed at the respective stage according to the literature.

Figure 3. Global and tissue specific expression patterns. (a) PCA applied to the average gene expression values of 3 replicates per tissue. (b) Hierarchical clustering using Pearson´s correlation coefficient. (c) Percentage of tissue expressed genes (present) in relation to total number of genes represented on the array (black), and percentage of tissue enriched genes in relation to total number of genes detected as enriched (red). (d) Venn diagram showing the number of common and uniquely expressed genes between the gametophyte and sporophyte generations.

Figure 4. Heat map showing the expression dynamics of transcription factors enriched in the sporophyte. Gene identifiers (left) and gene descriptions (right), including family, gene name and Arabidopsis homologs (in brackets) are shown when available. Transcription factors were selected from lists of enriched genes in S1, S2, S3, SM and “shared”. To generate the “shared” enriched list, we looked for genes that were enriched in at least 3 sporophytic developmental stages when compared against non-sporophytic tissues. Red represents high expression while yellow represents low expression values. Two TCP genes were selected from this list for further analysis (highlighted with an asterisk).

Figure 5. PpTCP gene expression patterns and Pptcp5 mutant phenotype. (a) Normalized microarray intensity values for the six PpTCP gene homologs were plotted to visualize their expression pattern in the different tissues analyzed. (b) Knockout mutation of the PpTCP5 gene shows several capsules emerging from the same foot (arrow) at different developmental stages. Scale bars = 100 µm. (c) Expression of TCP class II and class I (PpTCPI) genes in mutant and *WT* lines is shown by RT-PCR amplification, and (d) average percentages of sporophyte branching for *WT* and K.O. lines presented.

Figure 6. Gene Ontology enrichment analysis for sporophyte samples. GO terms for enriched genes in 4 different sporophytic developmental stages were clustered using DAVID tools and the terms with the highest enrichment scores ($p < 0.05$) for each developmental stage were selected.

Figure 7. Gene Ontology enrichment analysis of tip growing cells. GO terms from enriched genes in tip growing cells were clustered using DAVID tools and the terms with the highest enrichment scores ($p < 0.05$) for each tissue, and all three combined, respectively, were selected.

Figure 8. Transcription factor families in gametophytic and sporophytic generations. From the total number of enriched transcription factor genes in each generation, the percentage of genes belonging to a specific transcription factor family was calculated. Several transcription factor families are exclusively enriched in spores.

Tables

Table 1. Number of present, enriched and preferentially expressed transcripts per tissue.

Table 2. Gene Ontology enrichment analysis of spores and mature pollen grain.

Figure 1

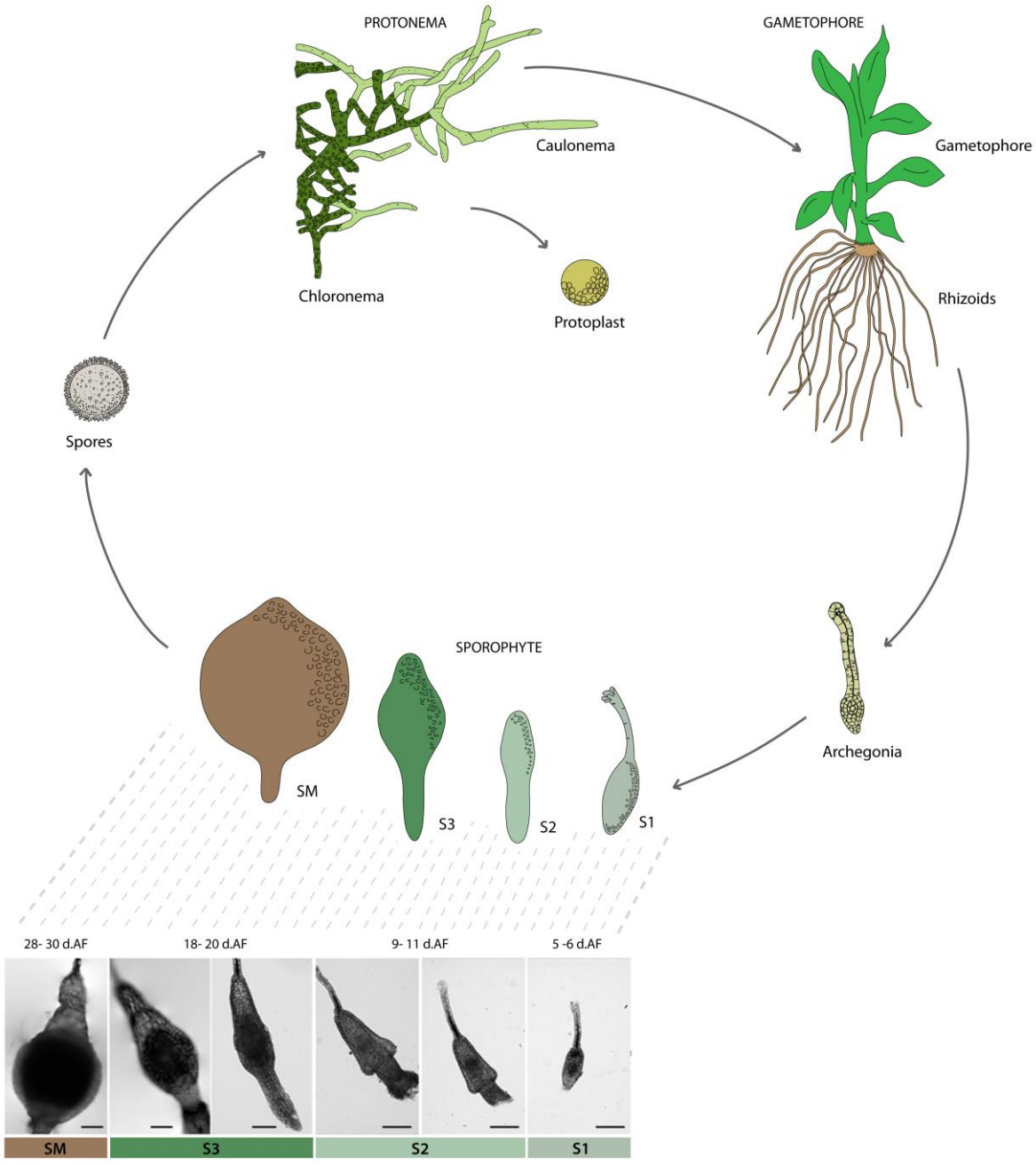


Figure 1. Sporophyte developmental stages. Four different developmental time points were defined and sporophytes collected based on the number of days that passed since the fertilization event (d.AF – days After Fertilization) until sporophyte maturation. Stages were denominated as Sporophyte 1 (S1), Sporophyte 2 (S2), Sporophyte 3 (S3) and mature (brown) sporophytes (SM). Scale bars = 100 µm.

Figure 2

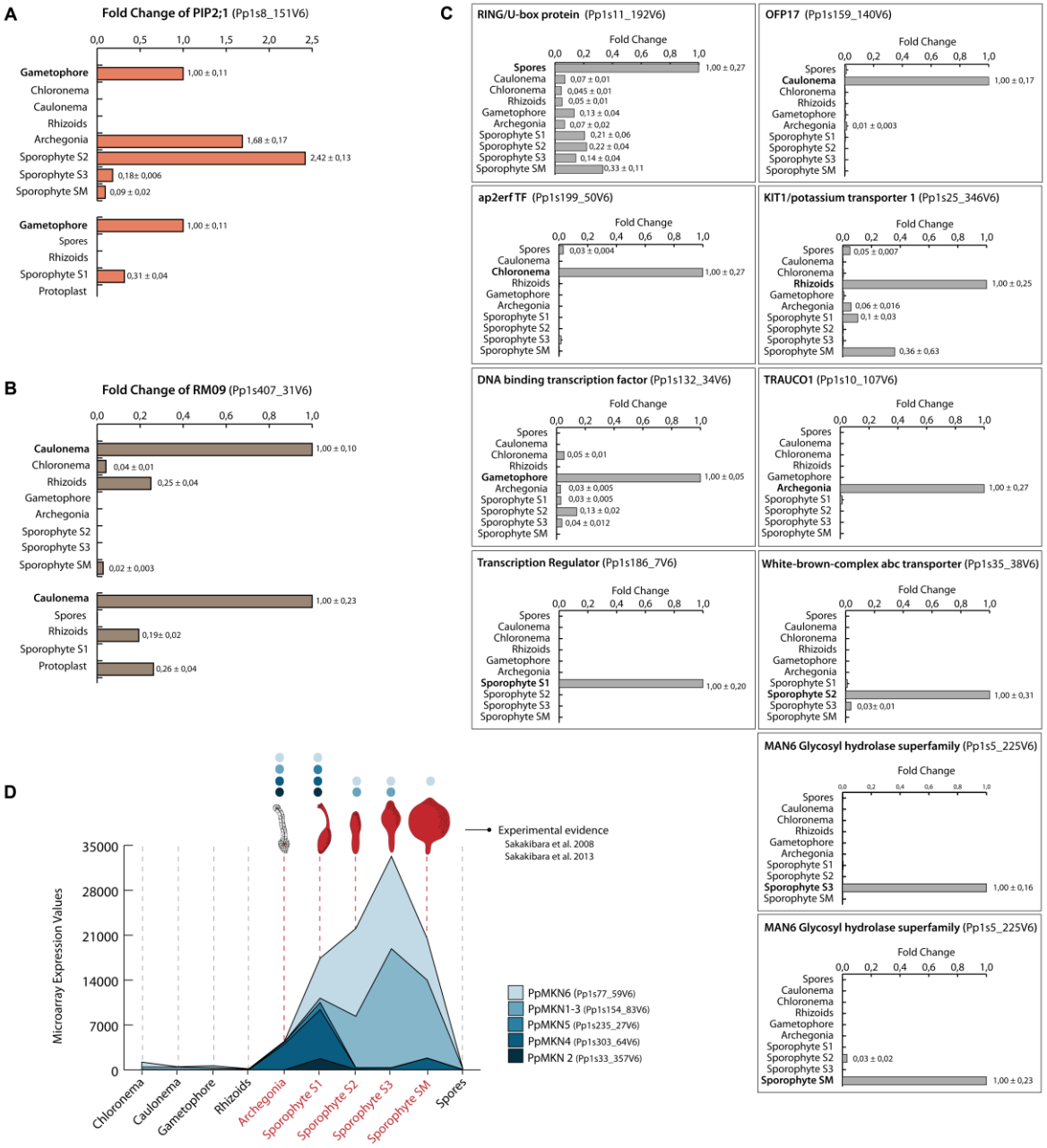


Figure 2. Assessment of microarray data quality. Quality of tissue sample replicates was tested by qRT-PCR using marker genes. (a) The expression of the gene RM09 (Pp1s407_31V6.1) reported to be present in protonemal tissues was tested. Expression is shown to be enriched in protonemal tissues (caulonema, chloronema, protoplasts and rhizoids), where differences in expression levels among them can also be observed. (b) In the same way, the PIP2;1 aquaporin

(Pp1s8_151V6.1) was tested and the opposite pattern was obtained. In this case gametophytic and sporophytic tissues showed expression while the transcript was not detected in any protonemal tissue. Relative expression ($2^{-\Delta\Delta Ct}$) was calculated using alpha-tubulin (Pp215_51V6.1) as housekeeping gene and caulonema or gametophore as the reference tissue. The standard error range (±) is depicted next to the bars. (c) Genes with tissue preferential expression were selected for pattern corroboration by qRT-PCR. cDNA from the three biological replicates was pooled. Data is shown as fold change ($2^{-\Delta\Delta ct}$) relative to gametophore expression in all cases except when the gene is preferentially expressed in gametophore, in which case rhizoids were used. (d) Correlation of gene expression values as reported by our microarrays with experimental evidence of transcript presence for the PpMKN family of transcription factors. Filled circles above the tissue cartoons indicate which PpMKN genes are expressed at the respective stage according to the literature.

Figure 3

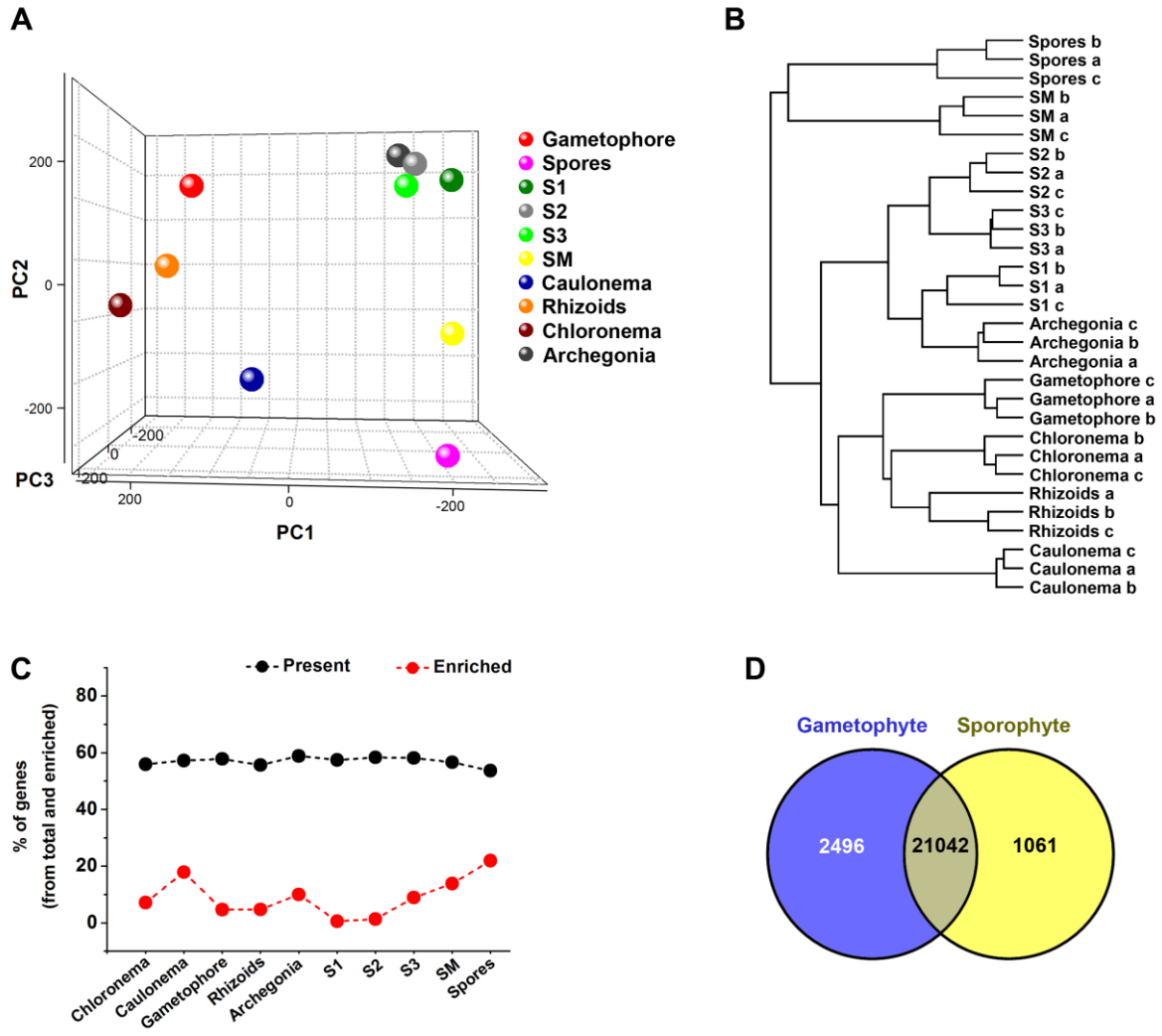


Figure 3. Global and tissue specific expression patterns. (a) PCA applied to the average gene expression values of 3 replicates per tissue. (b) Hierarchical clustering using Pearson´s correlation coefficient. (c) Percentage of tissue expressed genes (present) in relation to total number of genes represented on the array (black), and percentage of tissue enriched genes in relation to total number of genes detected as enriched (red). (d) Venn diagram showing the number of common and uniquely expressed genes between the gametophyte and sporophyte generations.

Figure 4

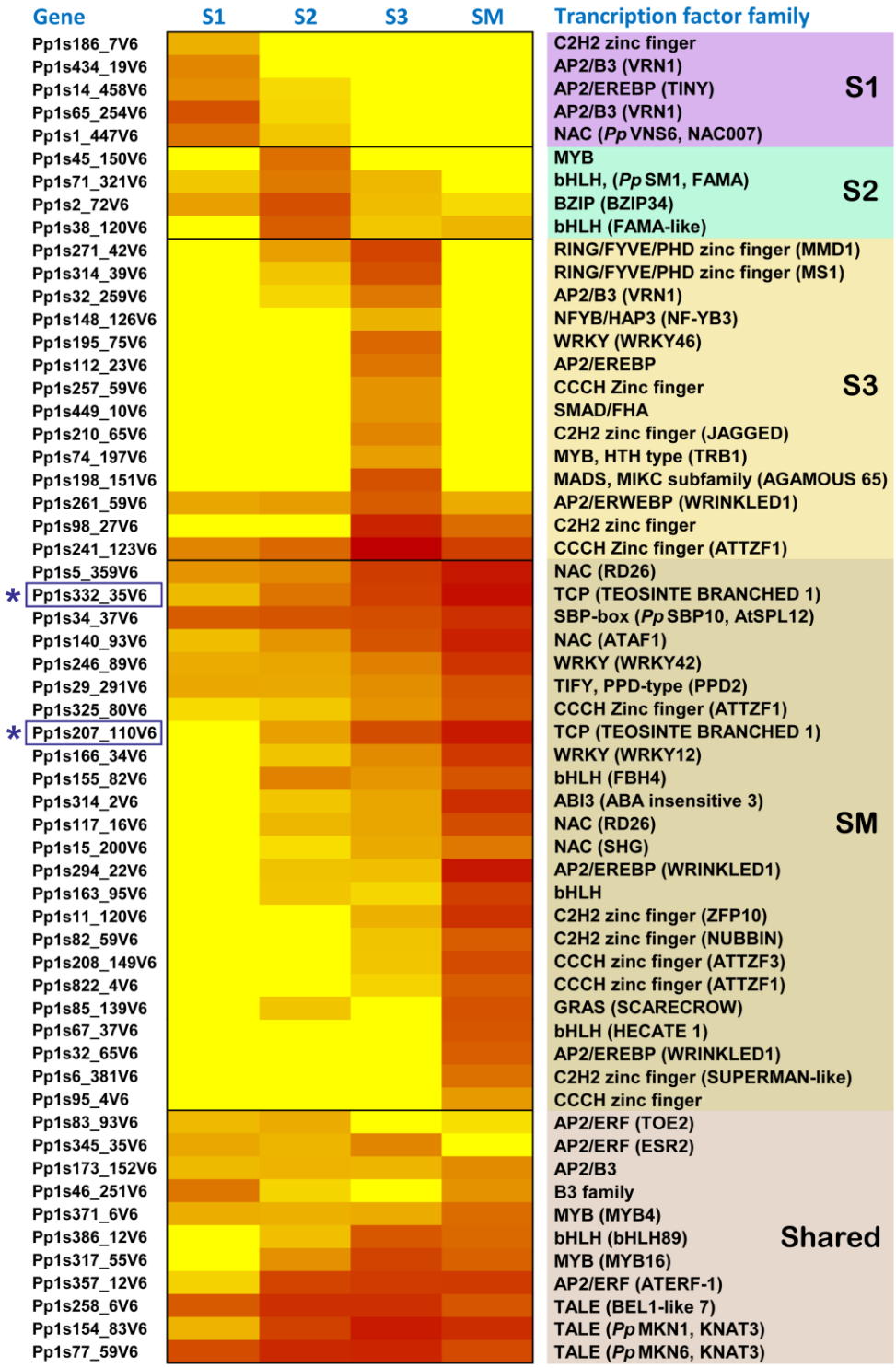
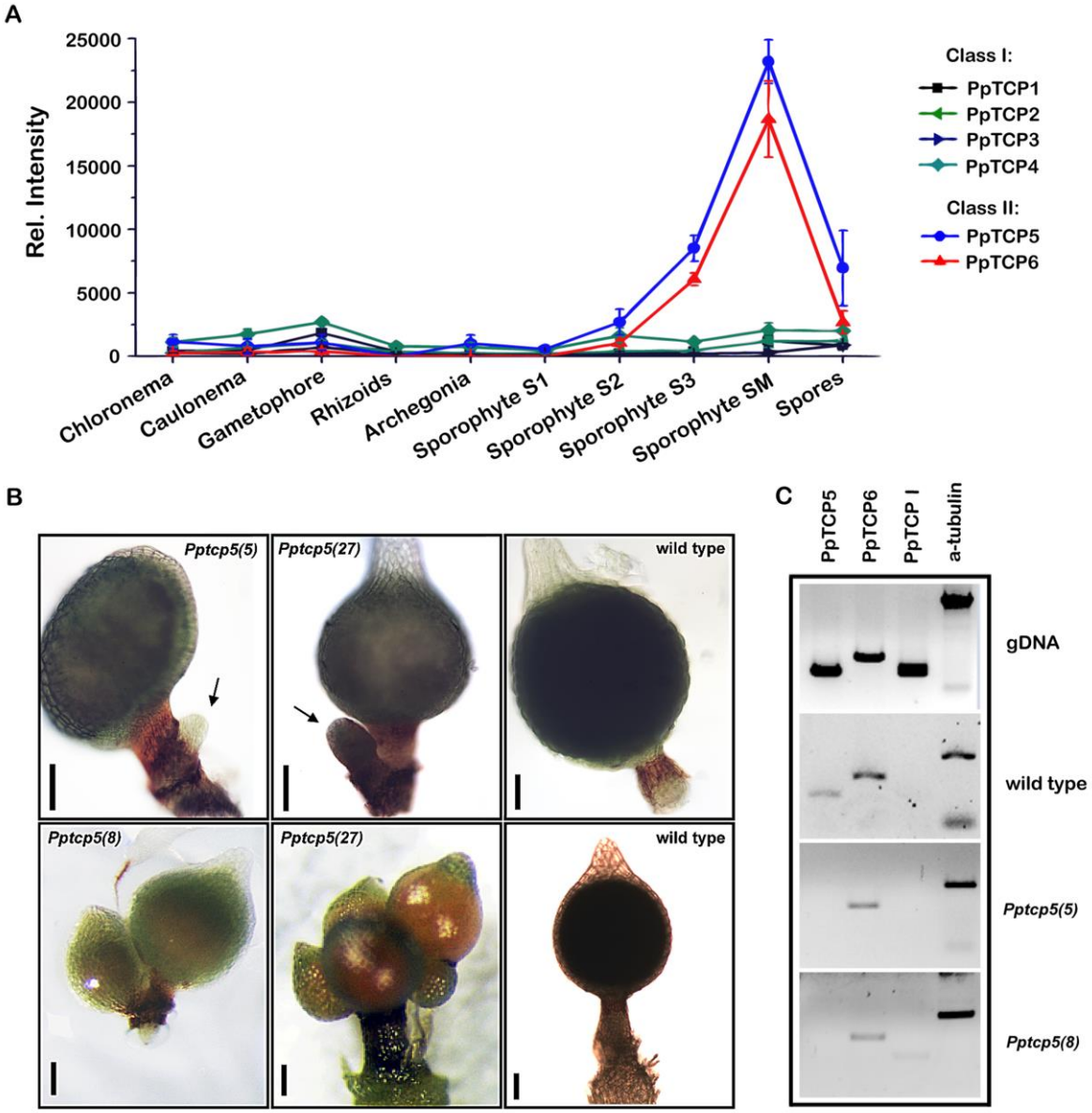


Figure 4. Heat map showing the expression dynamics of transcription factors enriched in the sporophyte. Gene identifiers (left) and gene descriptions (right), including family, gene name and Arabidopsis homologs (in brackets) are shown

when available. Transcription factors were selected from lists of enriched genes in S1, S2, S3, SM and “shared”. To generate the “shared” enriched list, we looked for genes that were enriched in at least 3 sporophytic developmental stages when compared against non-sporophytic tissues. Red represents high expression while yellow represents low expression values. Two TCP genes were selected from this list for further analysis (highlighted with an asterisk).

Figure 5



D

	Gametophores with sporophyte (%)	SE	Sporophytes branched (%)	SE	Biological replicates *
wt	52,5	2,6	5,1	1,1	9
Pptcp5(5)	47,1	2,4	13,2**	1,9	6
Pptcp5(8)	50,3	3,2	12,9**	3,2	6
Pptcp5(27)	47,9	2,9	14,9**	3,9	4

*At least 100 gametophores were counted per biological replicate for each genotype
**Statistical significant difference compared with wt

Figure 5. PpTCP gene expression patterns and Pptcp5 mutant phenotype. (a) Normalized microarray intensity values for the six PpTCP gene homologs were plotted to visualize their expression pattern in the different tissues analyzed. (b) Knockout mutation of the PpTCP5 gene shows several capsules emerging from the same foot (arrow) at different developmental stages. Scale bars = 100 µm. (c) Expression of TCP class II and class I (PpTCPI) genes in mutant and *WT* lines is shown by RT-PCR amplification, and (d) average percentages of sporophyte branching for *WT* and K.O. lines presented.

Figure 6

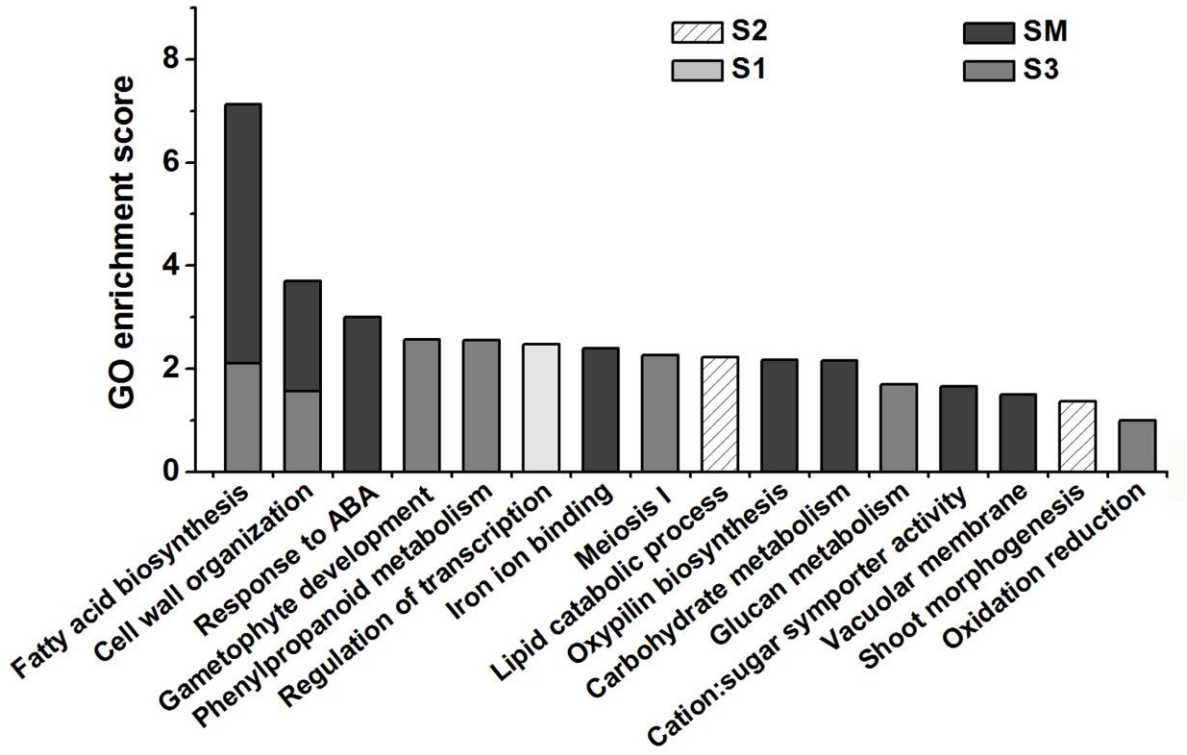


Figure 6. Gene Ontology enrichment analysis for sporophyte samples. GO terms for enriched genes in 4 different sporophytic developmental stages were clustered using DAVID tools and the terms with the highest enrichment scores ($p < 0.05$) for each developmental stage were selected.

Figure 7

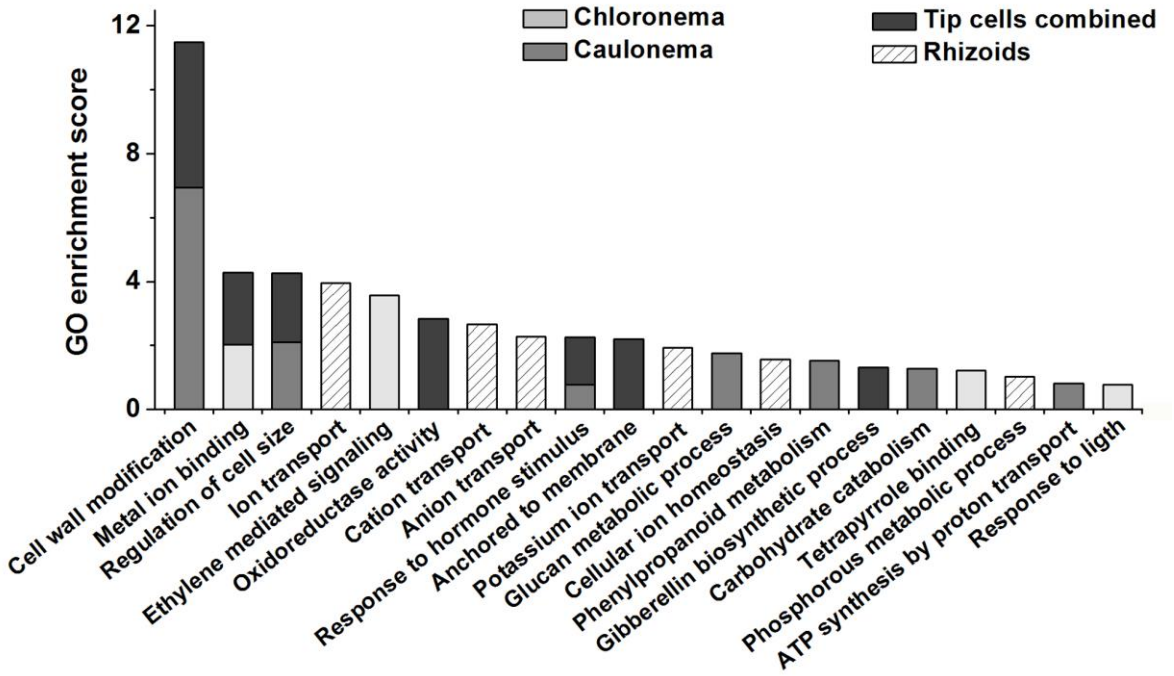


Figure 7. Gene Ontology enrichment analysis of tip growing cells. GO terms from enriched genes in tip growing cells were clustered using DAVID tools and the terms with the highest enrichment scores ($p < 0.05$) for each tissue, and all three combined, respectively, were selected.

Figure 8

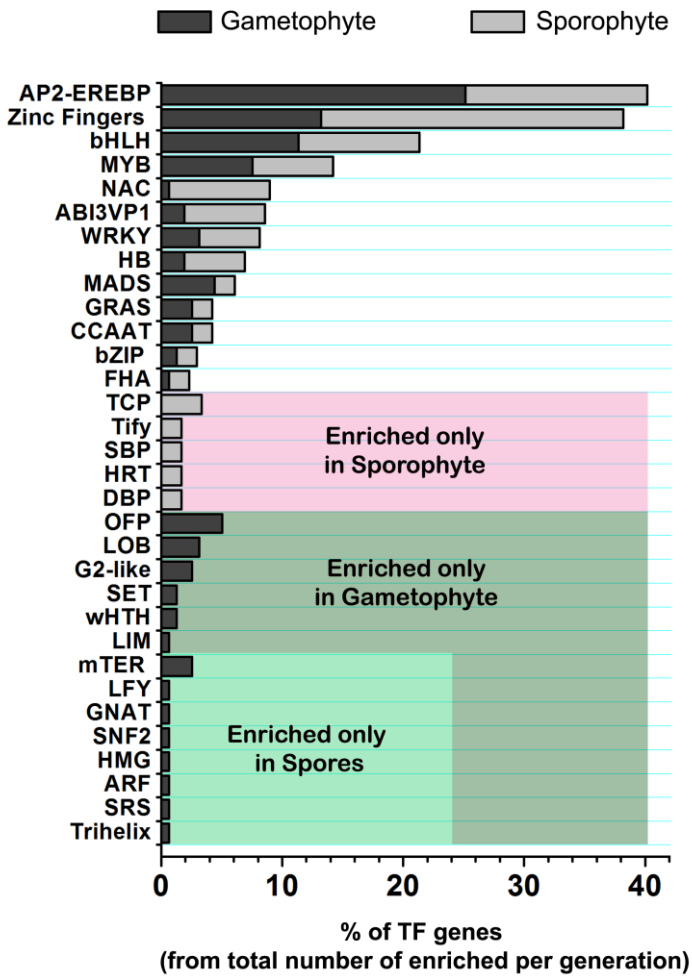


Figure 8. Transcription factor families in gametophytic and sporophytic generations. From the total number of enriched transcription factor genes in each generation, the percentage of genes belonging to a specific transcription factor family was calculated. Several transcription factor families are exclusively enriched in spores.

Table 1. Number of present, enriched and preferentially expressed transcripts per tissue.

Tissue	Present	Enriched	Preferential
Chloronema	18209	206	30
Caulonema	18638	517	84
Gametophore	18819	135	11
Rhizoids	18123	136	51
Archegonia	19169	289	197
Sporophyte 1	18696	17	10
Sporophyte 2	18995	38	15
Sporophyte 3	18929	258	104
Sporophyte M	18436	400	80
Spores	17468	635	70

Table 2. Gene Ontology enrichment analysis of spores and mature pollen grain.

Tissue	Gene ontology term	ES
Mature spores	Nuclear lumen	9.95
	ncRNA processing	7.23
	ATP-dependent helicase activity	6.99
	CUL4 RING ubiquitin ligase complex	4.06
	ATPase activity	3.11
	RNA splicing	2.27
	post-embryonic development	2.03
	Mitochondrial lumen	1.93
	DNA repair	1.89
Mature pollen grain	Cation/proton antiporter activity	5.23
	Hexokinase activity	2.63
	Phosphatase activity	2.59
	Alditol metabolic process	1.69
	Peroxisomal membrane	1.51
	Beta-galactosidase complex	1.42
	Metal ion transporter activity	1.40
	vacuolar membrane	1.37
	Secretion/exocytosis	1.35

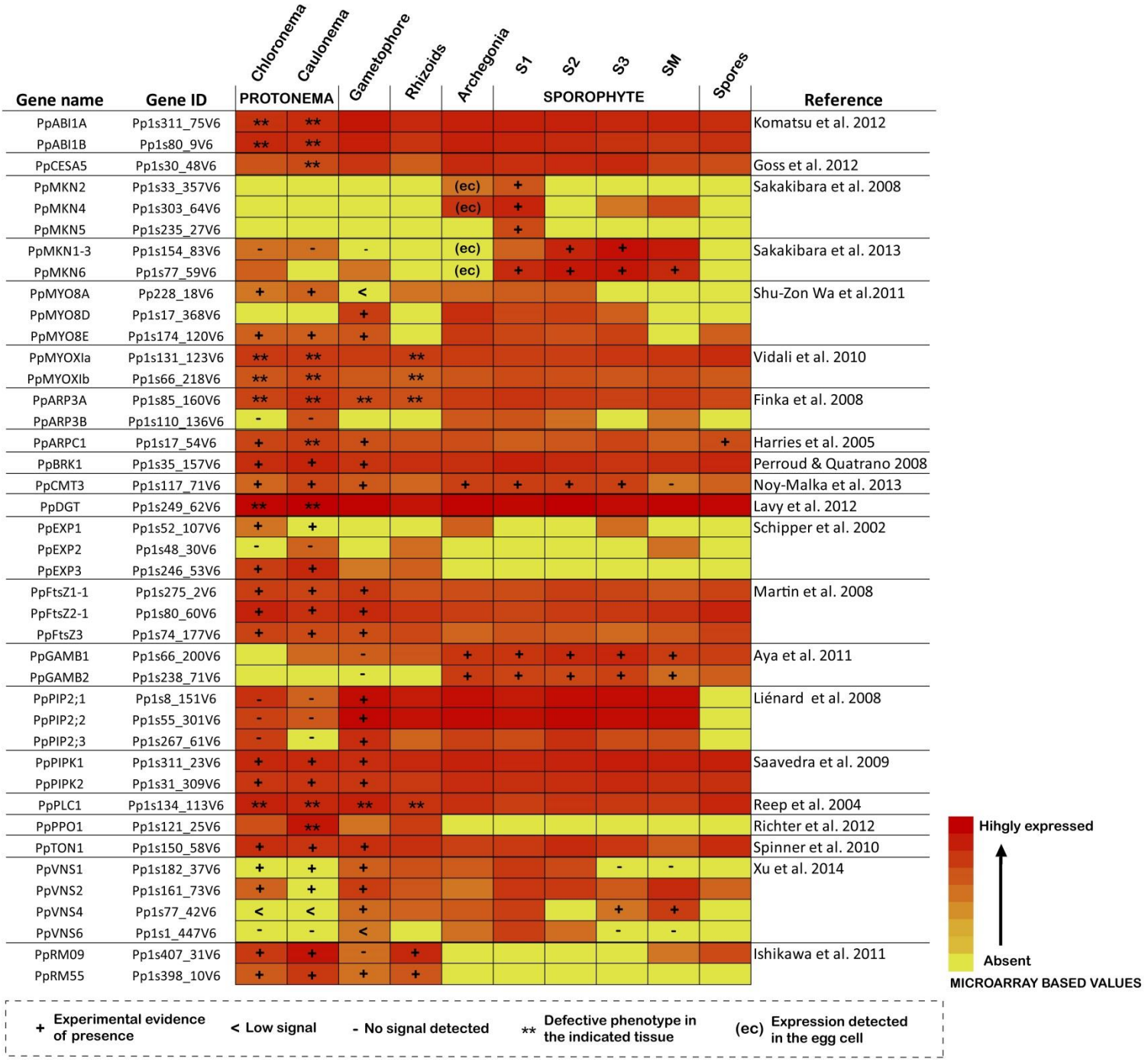
Table 2. Gene Ontology enrichment analysis of spores and mature pollen grain. Enriched gene list for pollen grains was obtained by performing pairwise comparisons of gene expression values from mature pollen against several *Arabidopsis* tissue samples like leaves, siliques, seedlings, root hairs, ovaries, hydrated pollen tubes, sperm cells and flowers. Already available *Arabidopsis* microarray data was used for this analysis. GO terms were clustered using DAVID tools and the terms with the highest enrichment scores for each tissue were selected. ES = enrichment score.

Supplemental Data

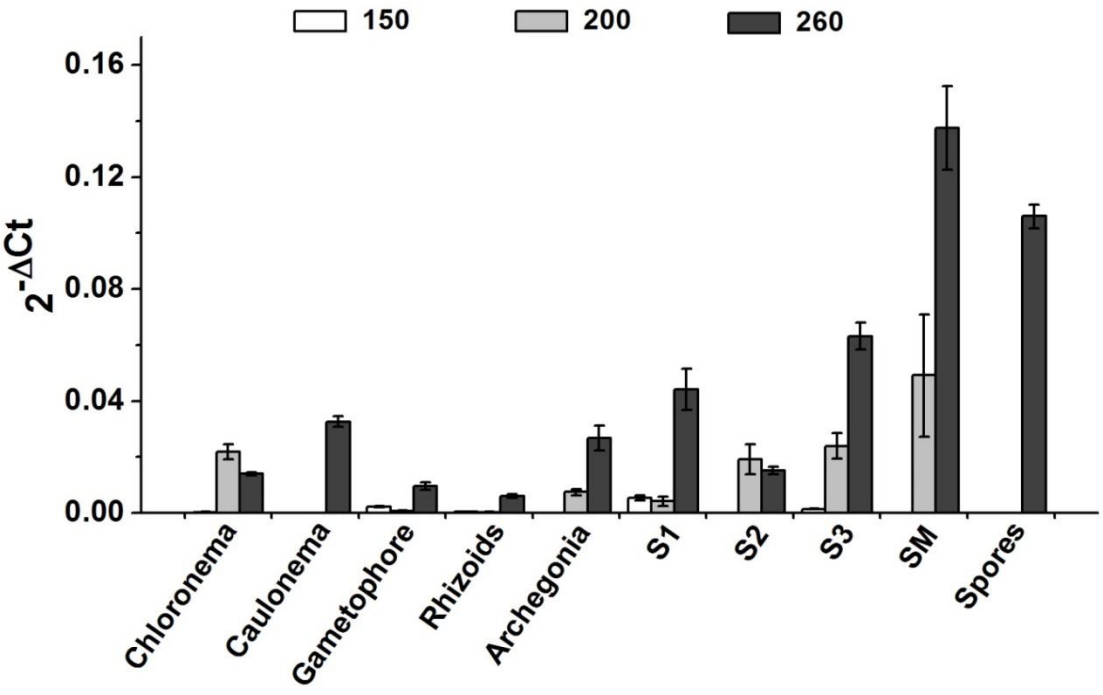
The following materials are available in the online version of this article.

Supplemental Figure 1. Correlation of microarray expression values and published evidence.
Supplemental Figure 2. Threshold determination for presence/absence detection calls.
Supplemental Figure 3. Phylogenetic analysis of TCP genes.
Supplemental Figure 4. PpPINA and PpPINB expression patterns as reported by the *P. patens* eFP browser.
Supplemental Figure 5. Phylogenetic analysis of MICK* genes.
Supplemental Figure 6. Developmental characterization of gametangia.
Supplemental Figure 7. Induction and isolation of caulonemal filaments and rhizoids.
Supplemental Figure 8. Analysis of several reported housekeeping genes to determine consistency in gene expression across different samples.
Supplemental Figure 9. Characterization of Pptcp5 knockout plants.

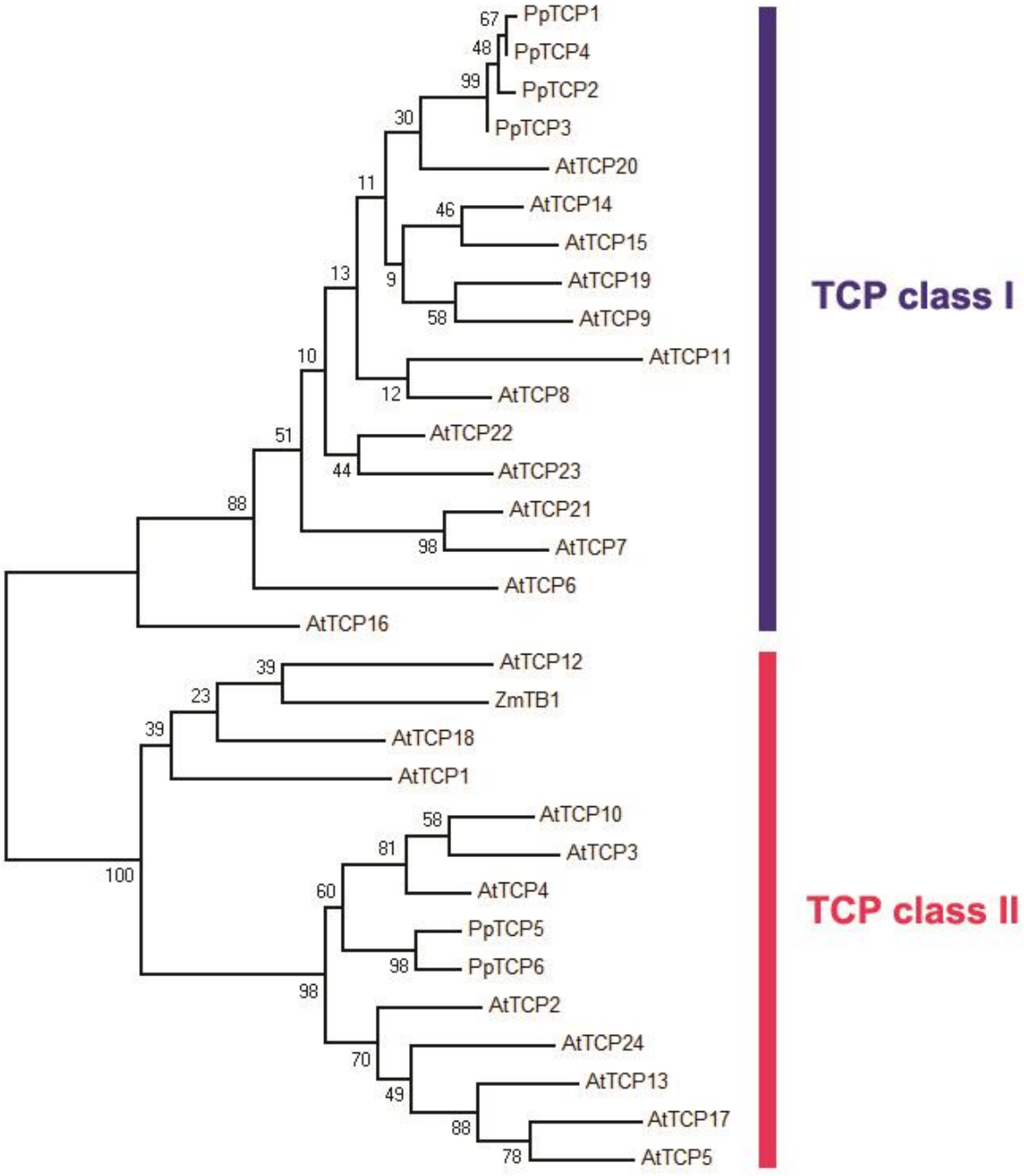
Supplemental Table 1. Gene expression data and tissue comparisons.
Supplemental Table 2. List of primers used.
Supplemental Table 3. Comparison of caulonema enriched transcripts identified in Xiao et al. (2011) and in the present study.
Supplemental Table 4. Expression profile of MICK* transcription factors.
Supplemental Table 5. List of preferentially expressed genes in tip growing cells from *P. patens*.
Supplemental Table 6. DAVID GO clusters for tip growing cells and sporophyte DEGs.



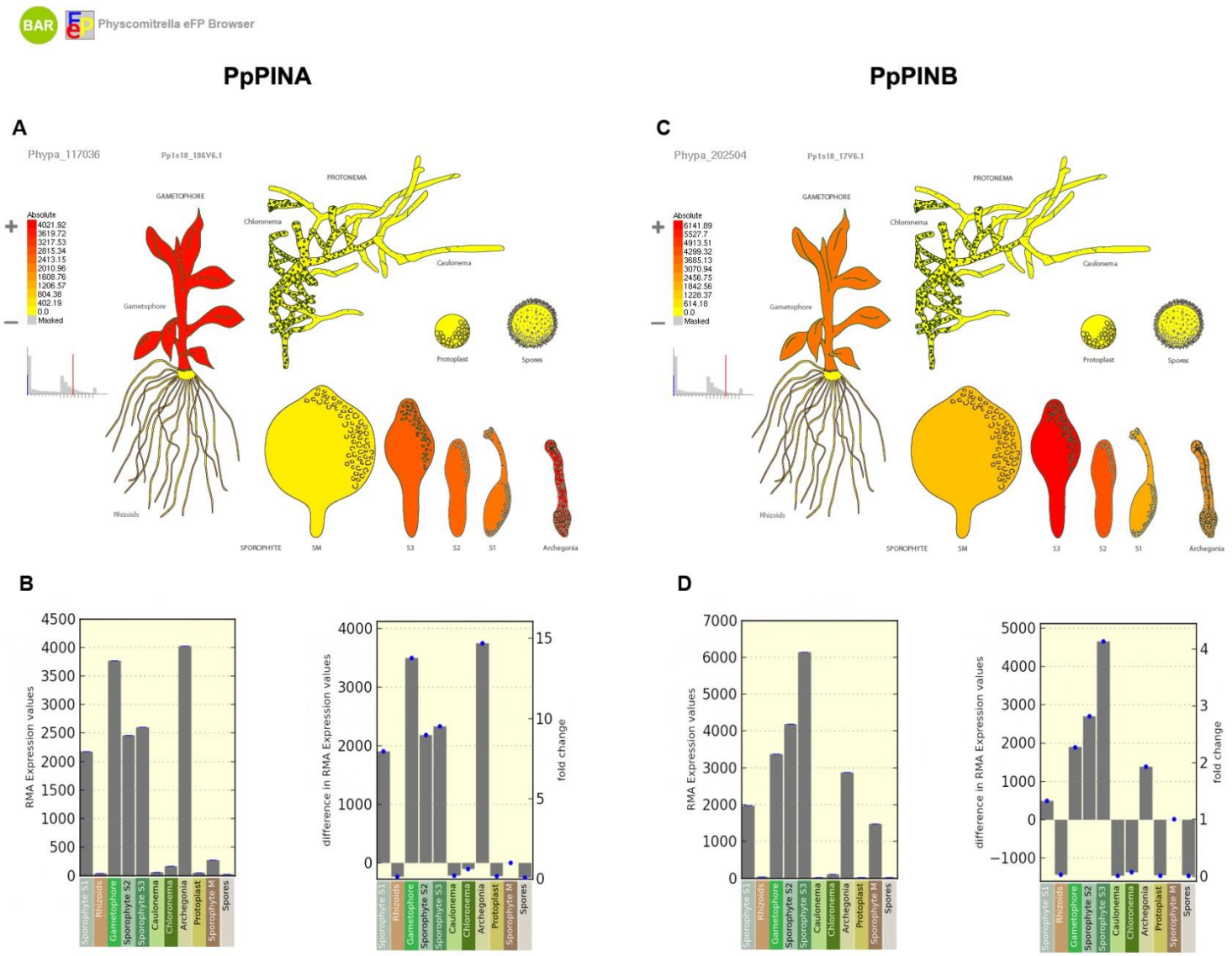
Supplemental Figure 1. Correlation of microarray expression values and published evidence. A bibliographical search was performed for published data on some of the genes present in our microarray with the purpose of correlating the expression values with experimental evidence. Gene name and gene identifiers are displayed on the left, while references in which such genes are mentioned are displayed on the right. In many cases several genes were evaluated in the same study. In the absence of tissue evidence (not studied or not shown), no symbol is displayed."+" means experimental evidence of presence, "<" indicates low GUS or GFP signal was detected by the authors, "-" no signal was detected, "**" indicates a defective phenotype was observed in corresponding tissues after gene K.O., and "(ec)" means detection in egg cell. Red represents high expression while yellow represents low expression values.



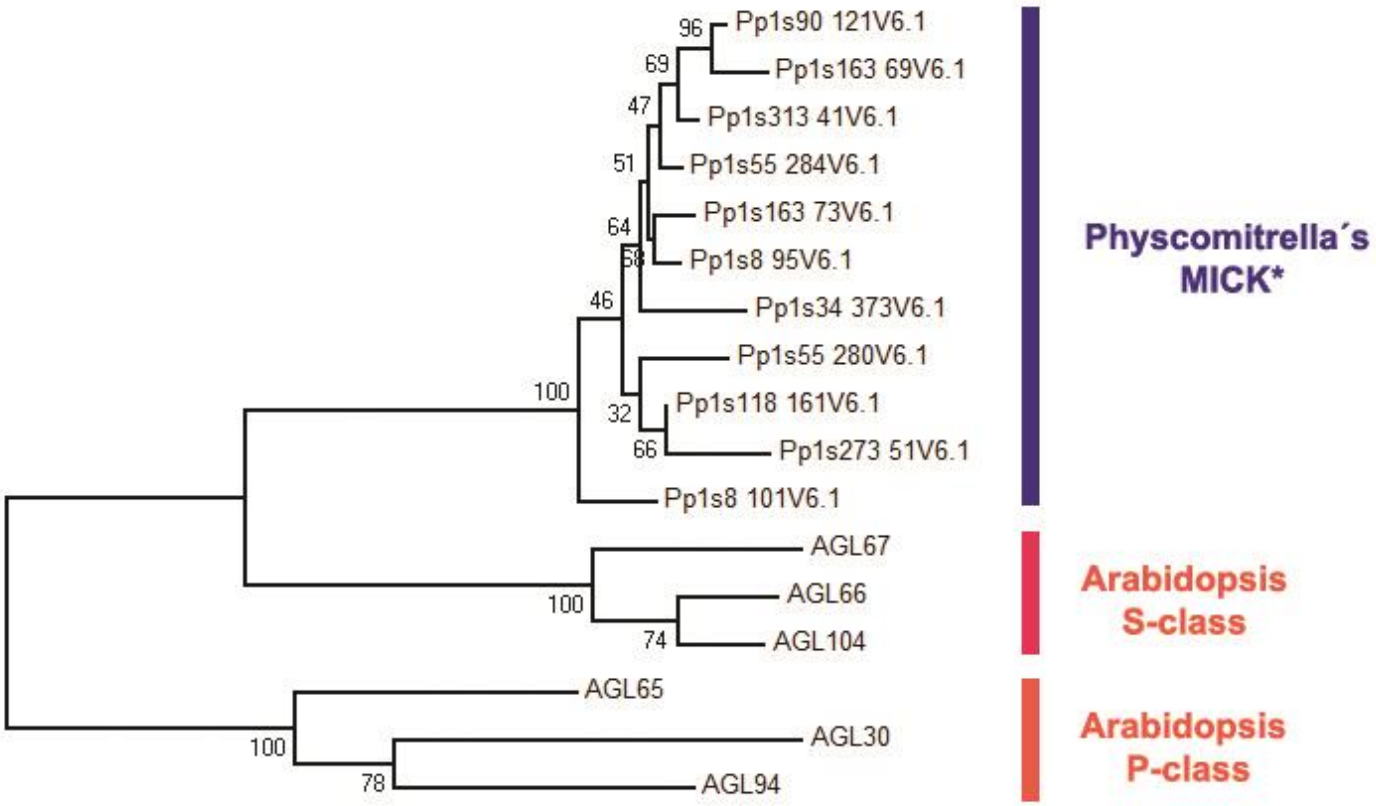
Supplemental Figure 2. Threshold determination for presence/absence detection calls. qRT-PCR experiments were conducted using excess cDNA synthetized for microarray hybridization from all *Physcomitrella* tissues. After a preliminary analysis we expected the detection threshold to be between the relative expression values of 150 to 260 as reported in the microarray data. Three genes per tissue with a relative expression values corresponding to 150, 200 and 260 were tested. From these, only the ones with a relative expression of 260 were always detected (amplified) by qRT-PCR.



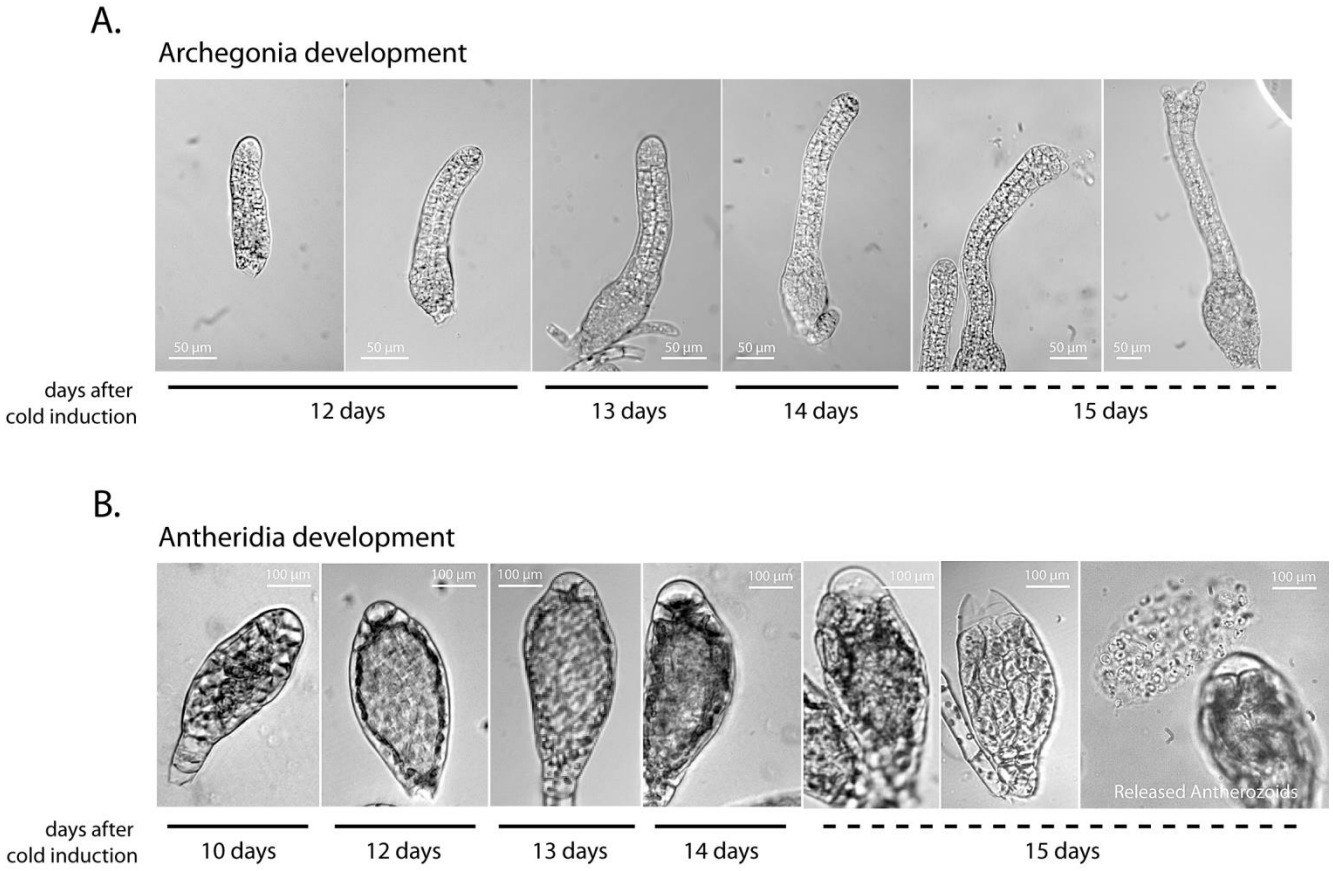
Supplemental Figure 3. Phylogenetic analysis of TCP genes. Phylogenetic tree constructed using maximum likelihood statistical method from aligned amino-acid sequences of TCP genes from *Arabidopsis thaliana (At), Physcomitrella patens (Pp)* and *Zea mays (Zm).* Bootstrap support values are shown at the nodes of the tree.



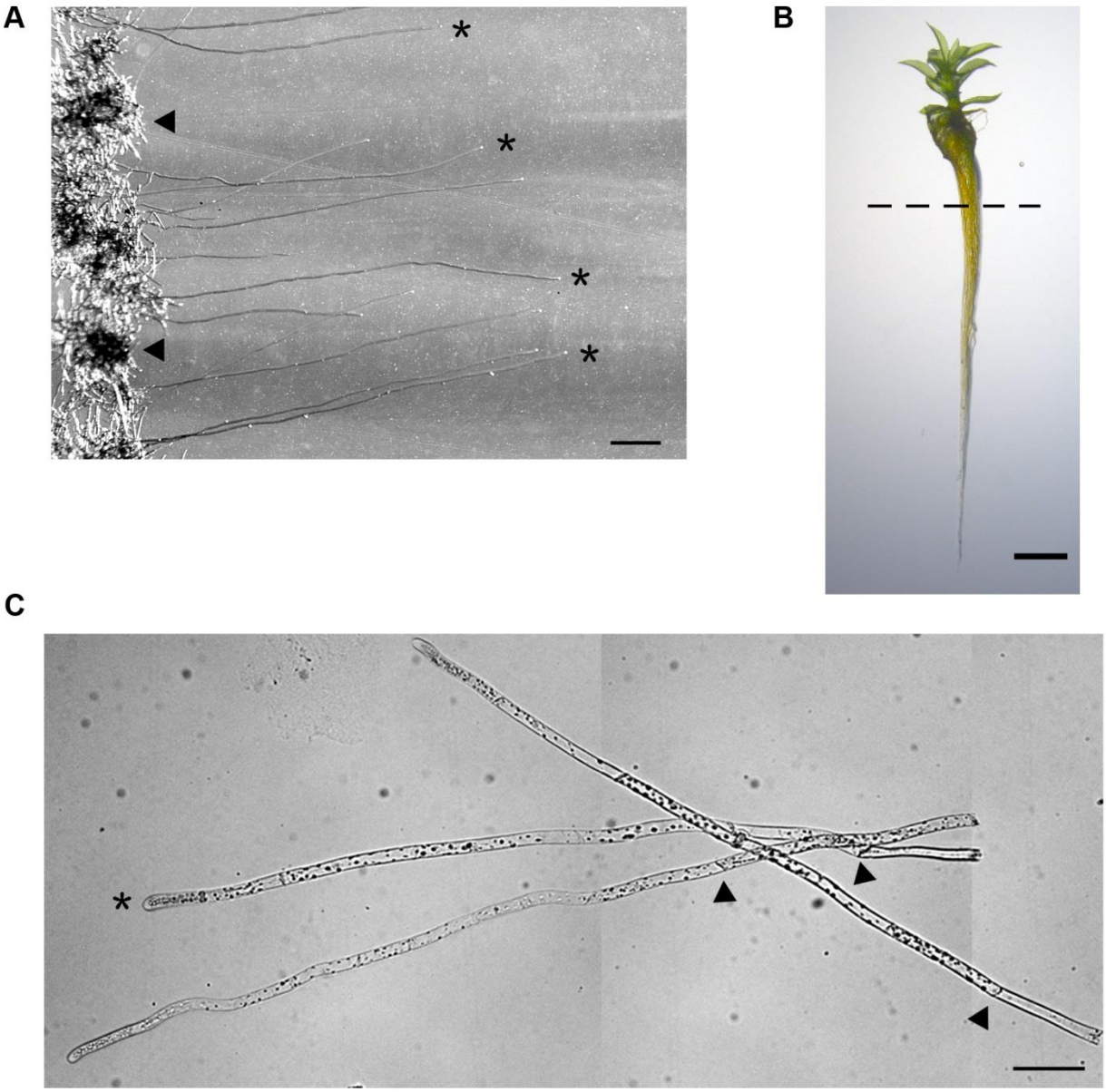
Supplemental Figure 4. PpPINA and PpPINB expression patterns as reported by the Physcomitrella eFP browser. A search was performed in the Physcomitrella eFP browser using the gene identifiers for PpPINA (Pp1s10_17V6.1) and PpPINB (Pp1s18_186V6.1). (a) Pictograph of *P. patens* tissues showing the expression profile of PpPINA in color code (red represents high expression while yellow represents low expression values). (b) Charts generated by the eFP browser showing PpPINA absolute (left) and relative (right) expression values for each tissue. (c) Pictograph of *P. patens* tissues showing the expression profile of PpPINB in color code. (d) Charts generated by the eFP browser showing PpPINB absolute (left) and relative (right) expression values for each tissue.



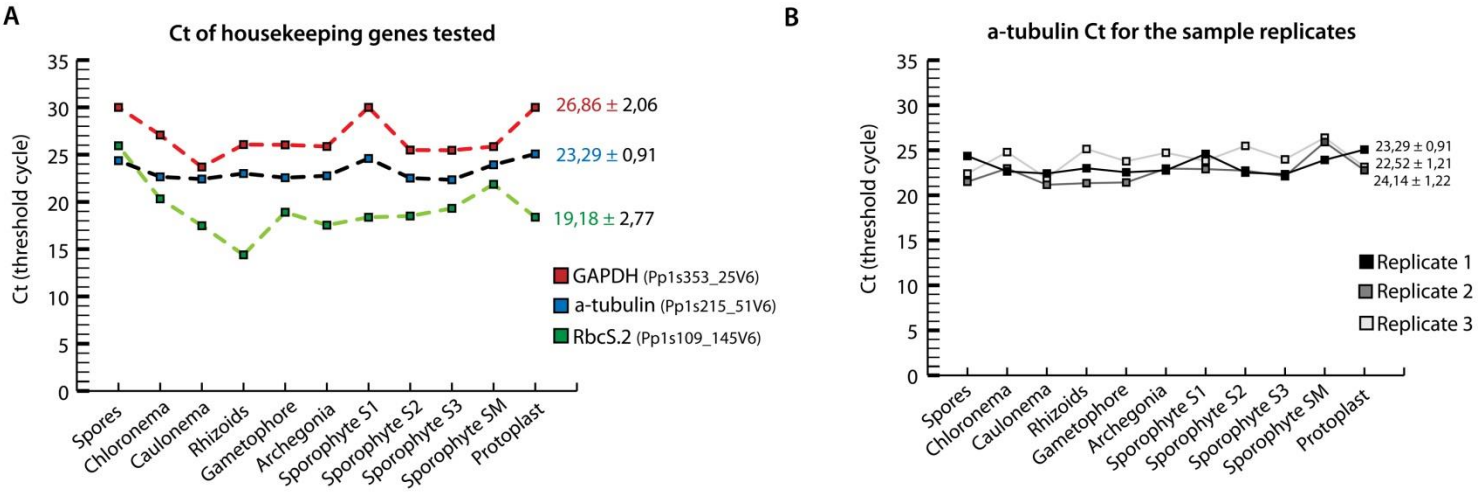
Supplemental Figure 5. Phylogenetic analysis of MICK* genes. Phylogenetic tree constructed using maximum likelihood statistical method from aligned amino-acid sequences of MICK* transcription factors from *Arabidopsis thaliana (AGL)* and *Physcomitrella patens (Pp).* Bootstrap support values are shown at the nodes of the tree.



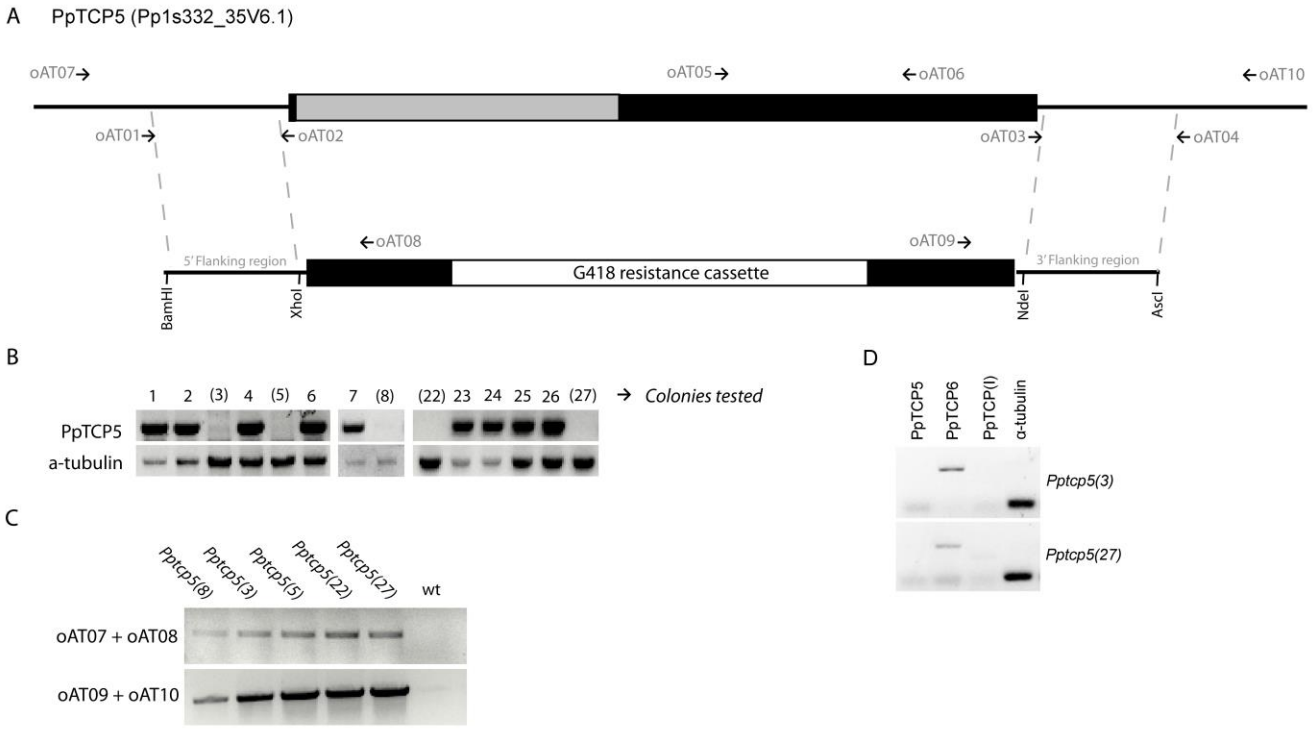
Supplemental Figure 6. Developmental characterization of gametangia. Induction of gametangia was conducted by exposing 3-4 week old gametophores to colder, short day conditions (17ºC, 8h light and 50% humidity). Observations were conducted almost every day after cold induction. (a) The developmental stages of archegonia from day 12 to 15 after cold induction are shown. At day 15 fully developed archegonia are observed. In many of them the channel leading to the egg cell appears open. (b) Antheridia development was followed in the same way. In this case, the most dramatic changes, besides growth, includes the change from green to yellow and the formation of a transparent “cap” at the anterior part of the organ. This cap brakes and the antherozoids are released. Sperm cell release events occur from day 14 and continue until day 17; however, it is during day 15 when we observed highest number of events. Both antheridia and archegonia can form organ clusters at different developmental stages. Reported dates represent the moment in which the depicted developmental stage was first seen.



Supplemental Figure 7. Induction and isolation of caulonemal filaments and rhizoids. Caulonema was induced by exposing 4-5 day old protonema to dark conditions in KNOPS medium containing glucose and covered with cellophane (25ºC and 50% humidity) for 5 days. (a) Caulonemal filaments at the moment of isolation for RNA extraction (asterisks) 5 days after induction. At this stage they can be clearly distinguished and separated from the chloronemal filaments at the base (arrowhead). Scale bar = 300μM. (b) Representative gametophore with rhizoids at the moment of rhizoid dissection. Rhizoids were cut well below the base of the gametophores to avoid tissue contamination. Scale bar = 1.25 mm. (c) magnification of dissected caulonema showing the oblique cell plates (arrowheads) and clear tip region (asterisk) characteristic of caulonema. Scale bar = 50μM.



Supplemental Figure 8. Analysis of several reported housekeeping genes to determine consistency in gene expression across different samples. (a) Comparison among reported housekeeping genes Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), alpha-tubulin (a-tubulin) and Rubisco chloroplast encoded subunit (RbcS.2) shows a more stable profile for alpha-tubulin. (b) Expression of the housekeeping gene alpha-tubulin was tested showing low variation among sample replicates and tissue samples.



Supplemental Figure 9. Characterization of *Pptcp5* knockout plants. (a) Schematic representation of the construct used for PpTCP5 homologous recombination. The localization of the primers used for the cloning and characterization of the knockout *Pptcp5* lines are represented. (b) Genotyping of potentially positive *Pptcp5* knockout colonies after three rounds of G418 selection using primers oAT05 and oAT06. Lines 3,5,8,22 and 27 show a disruption of the WT locus. (c) Corroboration of the insertion points in the genome was performed by PCR using the indicated primers. The amplification for both 5' (oAT07 + oAT08) and 3' (oAT09 + oAT10) insertion points are shown for the five stable knockout lines identified. (d) Expression of TCP class II and class I (PpTCPI) genes in mutant lines Pptcp5(3) and Pptcp5(27) measured by RT-PCR.

Supplemental figure 1 references:

Finka, A., Saidi, Y., Goloubinoff, P., Neuhaus, J.M., Zryd, J.P., and Schaefer, D.G. (2008). The knock-out of ARP3a gene affects F-actin cytoskeleton organization altering cellular tip growth, morphology and development in the moss Physcomitrella patens. Cell Motil. Cytoskeleton 65: 769–784.

Goss, C.A., Brockmann D.J., Bushoven, J.T., Roberts, and A.W. (2012). A cellulose synthase (CESA) gene essential for gametophore morphogenesis in the moss Physcomitrella patens. Planta 235: 1355–1367.

Harries, P.A., Pan, A., and Quatrano, R.S. (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.

Ishikawa, M., et al. (2011). Physcomitrella cyclin-dependent kinase A links cell cycle reactivation to other cellular changes during reprogramming of leaf cells. Plant Cell 23: 2924-2938.

Komatsu K., Suzuki N., Kuwamura M., Nishikawa Y., Nakatani M., Hitomi O., Takezawa D., Seki M., Tanaka M., Taji T., Hayashi T., and Sakata Y. (2012). Group A PP2Cs evolved in land plants as key regulators of intrinsic desiccation tolerance. Nature Comm. 4: 2219

Lavy, M., Prigge, M.J., Tigyi, K., and Estelle, M. (2012). The cyclo-philin DIAGEOTROPICA has a conserved role in auxin signaling. Development. 139: 1115–1124.

Liénard, D., Durambur, G., Kiefer-Meyer, M., Nogué, F., Menu-Bouaouiche, L., Charlot, F., Gomord, V., and Lassalles, P. (2008) Water transport by aquaporins in the extant plant Physcomitrella patens. Plant Physiology 146: 1207-1218.

Martin, A., Lang, D., Heckmann, J., Zimmer, A.D., Vervliet-Scheebaum, M., and Reski, R. (2009). A uniquely high number of *ftsZ* genes in the moss *Physcomitrella patens.* Plant Biol.11 (5): 744–750.

Noy-Malka C., Yaari R., Itzhaki R., Mosquna A., Gershovitz N.A., Katz A., and Ohad N. (2014). A single CMT methyltransferase homolog is involved in CHG DNA methylation and development of Physcomitrella patens. Plant Mol Biol. 84 (6): 719-735.

Perroud, P.F. and Quatrano, R.S. (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.

Repp, A., Mikami, K., Mittmann, F., and Hartmann, E. (2004). Phosphoinositide-specific phospholipase C is involved in cytokinin and gravity responses in the moss Physcomitrella patens. Plant J. 40: 250–259.

Richter H., Lieberei R., Strnad M., Novák O., Gruz J., Rensing S.A., and Schwartzenberg K. (2012). Polyphenol oxidases in *Physcomitrella*: functional PPO1 knockout modulates cytokinin-dependent development in the moss *Physcomitrella patens.* J Exp Bot 63(14): 5121-35.

Saavedra, L., Balbi, V., Lerche, J., Mikami, K., Heilmann, I., and Sommarin. M. (2011). PIPKs are essential for rhizoid elongation and caulonemal cell development in the moss Physcomitrella patens. Plant J. 67: 635–647.

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

Sakakibara K., Ando, S., Yip, H.K., Tamada, Y., Hiwatashi, Y., Murata, T., Deguchi, H., Hasebe, M., and Bowman, J.L. (2013). KNOX2 genes regulate the haploid-to-diploid morphological transition in land plants. Science. 339: 1067-1070.

Schipper, O., Schaefer, D., Reski R., and Flemin, A. (2002). Expansins in the bryophyte Physcomitrella patens. Plant Molecular Biology 50: 789–802.

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

Sugiyama, T., Ishida, T., Tabei, N., Shigyo, M., Konishi, M., Yoneyama, T., and Yanagisawa, S. (2012). Involvement of PpDof1 transcriptional repressor in the nutrient condition-dependent growth control of protonemal filaments in Physcomitrella patens. J. Exp. Bot. 63: 3185–3197.

Shu-Zon W., Ritchie J.A., Ai-Hong P., Quatrano R.S., and Bezanilla M. (2011). Myosin VIII regulates protonemal patterning and developmental timing in the moss Physcomitrella patens. Molecular Plant 4 (5): 909-921.

Vidali, L., Burkart G.M., Augustine R.C., Kerdavid E., Tüzel E., and Bezanilla M. (2010). Myosin XI is essential for tip growth in Physcomitrella patens. Plant Cell 22: 1868–1882.

Xu B., Ohtani M., Yamaguchi M., Toyooka K., Wakazaki M., Sato M., Kubo M., Nakano Y., Sano R., Hiwatashi Y., Murata T., Kurata T., Yoneda A., Kato K., Hasebe M., and Demura T. (2014). Contribution of NAC transcription factors to plant adaptation to land. Science 343: 1505-1508.