Comparative transcriptome analysis in the moss *Physcomitrella patens* and the genetic basis of key reproductive innovations

Marcela Hernández Coronado

Dissertation presented to obtain the Ph.D degree in Biological Sciences
Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Research work coordinated by

INSTITUTO GULBENKIAN DE CIÊNCIA

Oeiras, July, 2015

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA /UNL
Knowledge Creation

UNIVERSIDADE NOVA DE LISBOA

This dissertation was sponsored by the Marie Skłodowska-Curie ITN actions through the project “PLANTORIGINS-Plant developmental biology: discovering the origins of form” (FP7-PEOPLE-ITN-2008).

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all the people that directly or indirectly had contribute for this work. First and foremost, I would like to thank to my supervisor Jörg Becker. It has been a privilege to be his student and to have learned from him. Through this years he encouraged me to follow my ideas and to gain independence, been always a valued source of advice and suggestions. He continuously challenged me, pushing me to give the next step. This helped me grow and improve and I am very grateful for that. It has been a rewarding process and I feel proud for the project we built together.

All former and present members of the Plant Genomics Lab had contributed immensely with ideas and observations, but also with their example. I would like to thank in particular to Patricia Pereira and Filipe Borges who welcomed me warmly and taught me several techniques, to Leonor Boavida for all the scientific input and help throughout these years, and to Anna Thamm for her great collaboration and friendship. I would also like to thank José Feijó and the Cell Biophysics and Development Lab. All of them were a second lab for me and were very supportive and kind. And to Paulo Navarro for his useful comments and suggestions. I feel very fortunate to have been part of the Marie Curie Plant Origins Initial Training Network and I would like to acknowledge the entire group that were part of it. I learned enormously from that experience.

Also, I would like to thank Thomas Tam for teaching me the techniques to work with *Physcomitrella patens*, and to Vera Nunes for her assistance with the growing chambers and plant material. The microarrays presented in this work would not have been possible

without the excellent work of João Sobral at the Gene Expression Unit of the IGC, to whom I am very grateful. Also, I would like to acknowledge all members of the IGC Unit of Imaging, whose assistance was crucial for the image acquisition and antherozoid collection conducted in this study.

During my PhD, the IGC has been a source of friendship. I would like to thank to Ewa Chrostek and Krzysztof Kus for the scientific and personal support, to Raquel Carvalho and Nadine Pires for sharing their climbing passion with me, and to Oscar Ruíz, Karen Berman, Fatuel Tecuapetala and Jazmín Ramiro, who were an example in many ways.

I would like to thank my parents who raised me with the love for knowledge and for nature, and have inspired me to pursuit my goals. And to my brother and sister who are always there for me, for their endless love and all they had taught me. Lastly, I would like to thank to Carlos for his great support and encouragement both in the lab and outside it. He has been a guide and a companion.

SUMMARY

During land plant evolution plants underwent extensive transformations. For the successful adaptation to the terrestrial environment the vegetative tissues were modified to cope with low water availability while new reproductive organs and strategies emerged to ensure plant dispersal and survival. Importantly, plant life cycles were dramatically modified, evolving from gametophyte dominant in early land plants such as Bryophytes to sporophyte dominance as it is observed in flowering plants. The genetic changes underlying plant evolution can be approached by genetic comparative studies, and although available information of gene expression in Angiosperms exists we know relatively little regarding early land plants. Therefore we generated a gene expression atlas for all major tissues and developmental stages of the Bryophyte *Physcomitrella patens*. Through comparative transcriptomic analysis genes with possible roles for land plant evolution are proposed. Among those we focus on the increase of sporophyte complexity and male gamete evolution.

In our work, a set of transcription factors specifically expressed in the moss sporophyte was identified, suggesting possible functions during development and evolution. The loss-of-function of one of those genes, *PpTCP5*, causes the development of supernumerary sporangia attached to a single seta, indicating a conserved role of class II TCP genes in sporophyte architecture. Based on the study of specific gene families expressed in the gametophyte, such as the OVATE transcription factors, the possible “recruitment” of gametophyte genes for sporophyte functions is suggested. Moreover,

the study of tip growth related genes revealed a process of specification during evolution, with genes being expressed in all tissues with polarized growth in *P. patens* but restricted to either pollen tubes or root hairs in *Arabidopsis*.

We developed a methodology to isolate *P. patens* antherozoids. The transcriptional profile of antherozoids in conjunction with data from two developmental stages of antheridia constitutes the first study of male gametogenesis in Bryophytes. During moss spermatogenesis a complex transcriptome was observed, characterized by a high number of enriched and preferentially expressed genes. Clustering analysis of enriched transcripts revealed two major functional groups: one with putative roles in sperm identity and fertilization (e.g. sperm motility), and a second with transcripts that could play a role in embryo development after their delivery upon fertilization. Based on a phylostratigraphic analysis we could show that transcripts corresponding to evolutionarily younger genes were preferentially expressed in the antherozoids, thus extending similar observations in angiosperms and animals, in which male reproductive cells act as a source of evolutionary gene innovation, to extant early land plants.

SUMÁRIO

Durante a evolução ao meio terrestre, as plantas passaram por transformações extensivas que permitiram uma adaptação bem sucedida a um novo ambiente. Neste processo, os tecidos vegetativos sofreram modificações que lhes permitiram lidar com a baixa disponibilidade de água, enquanto novos órgãos e novas estratégias reprodutivas surgiram para assegurar a dispersão e sobrevivência das plantas.

É de salientar que os ciclos de vida sofreram alterações drásticas, evoluindo de um gametófito dominante nas plantas terrestres primitivas, tais como as Briófitas, para a dominância do esporófito como é observado em plantas com flores. As alterações genéticas que deram lugar à evolução das plantas podem ser abordadas por estudos comparativos genéticos, e embora exista informação disponível da expressão génica em Angiospérmicas, pouco sabemos relativamente a plantas terrestres primitivas. Neste trabalho gerámos um atlas de expressão génica, para todos os principais tecidos e estádios de desenvolvimento da Briófita *Physcomitrella patens*. Através de estudos comparativos transcricionais identificámos genes com possível importância na evolução das plantas. Nesta análise focámos o nosso interesse no aumento da complexidade do esporófito e na evolução do gameta masculino.

No nosso trabalho, foram identificados um conjunto de factores de transcrição especificamente expressos no esporófito do musgo, que sugerem possíveis funções durante o desenvolvimento e evolução. A perda de função de um desses genes, *PpTCP5*, produz cápsulas supernumerárias ligadas a uma única haste, indicando a

conservação dos genes TCP classe II na regulação da arquitectura esporofítica. Com base no estudo de famílias de genes expressas especificamente no gametófito, tais como os factores de transcrição *OVATE*, é sugerido o possível "recrutamento" de genes do gametófito para funções no esporófito. Além disso, o estudo de genes relacionados com crescimento apical revelou um processo de especificação durante a evolução, em que genes expressos em todos os tecidos com crescimento apical em *P. patens* aparecem restritos a tubos polínicos ou pelos radiculares em *Arabidopsis*.

Neste trabalho desenvolvemos uma metodologia para isolar os anterozóides de *P. patens*. O perfil transcricional dos anterozóides em conjunto com dados de dois estádios de desenvolvimento do anterídio constitui o primeiro estudo da gametogénese masculina em Briófitas. A espermatogénese do musgo envolve um transcriptoma complexo, caracterizado por um elevado número de genes enriquecidos e preferencialmente expressos. A análise do agrupamento de transcritos enriquecidos revelou dois grandes grupos funcionais: um com funções putativas na identidade das células espermáticas e fertilização (por exemplo, na mobilidade do esperma), e um segundo com transcritos com potenciais funções no desenvolvimento embrionário após a fertilização. Com base numa análise filoestratigráfica podémos demonstrar que transcritos correspondentes a genes evolutivamente mais jovens são preferencialmente expressos nos anterozóides, alargando assim observações semelhantes em angiospérmicas e animais a plantas terrestres primitivas ainda existentes, em que células reprodutivas masculinas funcionam como uma fonte evolucionária de inovação genética.

ABBREVIATION LIST

AF	After Fertilization
Anth10d	Antheridia collected at day 10 after induction
Anth14d	Antheridia collected at day 14 after induction
BAR	Bio-analytic Resource for Plant Biology
DAVID	The Database for Annotation, Visualization and Integrated Discovery
eFP	electronic Fluorescent Pictograph
FDR	False Discovery Rate
GFP	Green Florescent Protein
GO	Gene Ontology
GRN	Gene Regulatory Network
K.O	Knock Out
LCB FC	Lower-Confidence Bound Fold-Change
MGU	Male Germ Unit
PCA	Principal Component Analysis
RFP	Red Fluorescent Protein
RMA	Robust Multichip Average
SCgrn	Sperm cells from the *Gransden* strain
SCvsx	Sperm cells from *Villersexel* strain
S1	Sporophyte Stage 1
S2	Sporophyte Stage 2
S3	Sporophyte Stage 3
SM-1	Sporophyte Stage Mature -1
SM	Sporophyte Mature Stage
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TCP	Teosinte Branched1/Cycloidea/Proliferating Cell Factor 1
TF	Transcription Factor
WT	Wild Type

TABLE OF CONTENTS

CHAPTER I. General Introduction

1. The origins of land plants

The land plant colonization process that took place between 480 and 360 million of years (Mya) ago represented a decisive event in the life history. During the Paleozoic, plant species diversification resulted in significant environmental changes that led to the development of terrestrial ecosystems (Graham et al., 2000; Kenrick and Crane, 1997; Niklas and Kutschera, 2010). It is believed that green plants originated from a flagellated algal ancestor, probably over a billion years ago (Lewis and McCourt, 2004). Early during evolution, the split of algae into two main lineages gave rise to the Chlorophyta and the Streptophyta. While the first is comprised of the marine plankton (Parasinophytes) and the Chlorophytes (including genera such as Chlamydomonas, Volvox, and Ulva), the Streptophytes were the lineage that gave rise to the majority of freshwater algae (Charophytes) and land plants (Embryophytes) (Leliaert et al., 2011) (Figure 1).

Based on these phylogenetic relationships, a single origin of land plants from Chareophytes (such as the Charales, Coleochaetales or Zygnematales) has been widely accepted, however the specific phylogenetic relationship among those algae groups and the early land plants (Bryophytes) is still controversial (Leliaert et al., 2011; Lewis and McCourt, 2004; McCourt et al., 2004; Qiu et al., 2007). Based on fossil evidences and molecular phylogenetic studies it has been indicated that liverworts are the closest relatives to early land plants (Kenrick and Crane, 1997; Qiu et al., 2007, 2006), and in the most accepted scenario, mosses are the sister group to hornworts and Tracheophytes (vascular plants) (Kenrick and Crane, 1997; Ligrone et al., 2012; Qiu et al., 2006). Evidences from the fossil record indicate a great increase of vascular plant diversity in the Early

Devonian (398 Mya), from where the appearance of important living groups has been highlighted (Kenrick and Crane, 1997). According to the fossil record, Angiosperms had their origin in the Early Cretaceous (100-145 Mya) becoming the most diverse and widespread plant group to date.

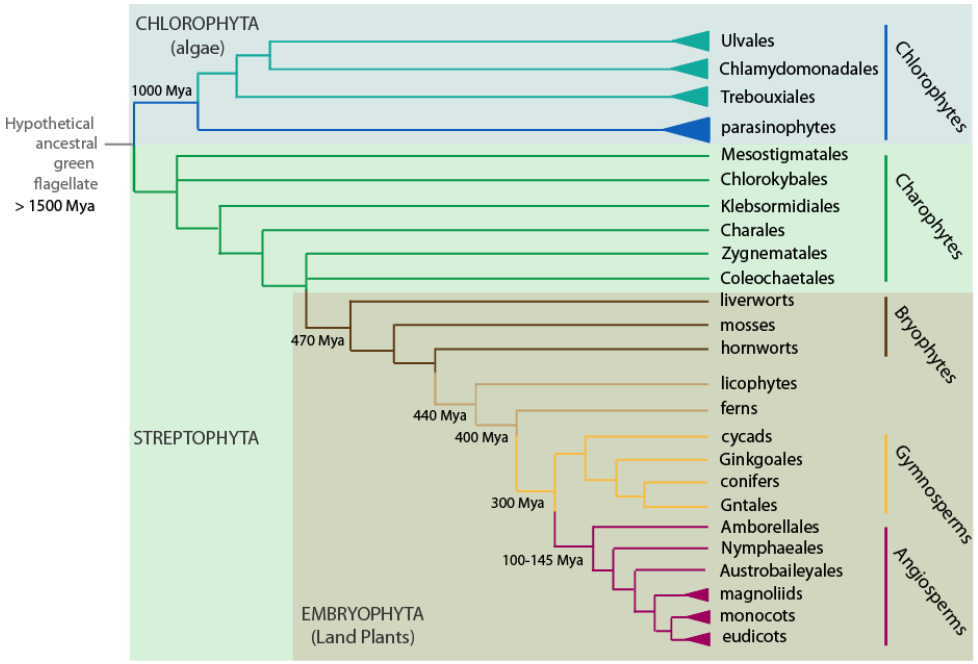


Figure 1. Phylogenetic relationships among main lineages of green plants.
The main lineages of green plants and their phylogenetic relationships are displayed. The estimated divergence time in Millions of years (Mya) is shown next to the branches. Drawing based on published phylogenetic trees depicting the relationships among green plants (Leliaert et al., 2011; Pires and Dolan, 2012).

1.1 Innovations for a life on land: a diploid multicellular sporophyte

The process of land plant colonization required several morphological and physiological transformations. In the early phases of plant evolution, innovations in the vegetative plant body would have been

required for the drier and fluctuating environment. And subsequently, innovations in the reproductive system would have arisen culminating in the development of seeds and flowers in Angiosperms (review in Crane and Leslie; Langdale and Harrison, 2008; Pires and Dolan, 2012). Land plants differ from their green algae ancestors by possessing specialized mechanisms to overcome the shortage of water, such as sporopollenin-spore walls, waxy cuticles and stomata (review in Ligrone et al., 2012). But perhaps the strongest difference that characterizes land plants is the presence of a multicellular diploid sporophyte generation that increased in complexity during evolution (Graham et al., 2000).

The life cycle of green plants can consist of haploid and diploid phases. While extant charophycean algae possess a life cycle with only a haploid multicellular generation (haplobiontic-haploid life cycle), the land plants show a dimorphic life cycle with both of the phases being multicellular (diplobiontic life cycle) (Niklas and Kutschera, 2010). In accordance, it has been suggested that land plants may have inherited a multicellular gametophore from their algal ancestors, but that the multicellularity in the sporophyte would have evolved in the terrestrial environment giving rise to the characteristic alternation of generations observed in embryophytes (Graham et al., 2000; Kenrick and Crane, 1997). The process by which the dimorphic alternation of generations originated is the subject of two conflicting models: the anthitetic and the homologous theory (reviewed in Bennici, 2008; Blackwell, 2003). In the less accepted homologous theory, an algal ancestor with two equal life phases (or isomorphic) is assumed (Fritsch, 1921), and although fossil evidences of organisms with isomorphic life cycles support the idea, they do not correspond to groups phylogenetically close to embryophytes. The antithetic theory,

on the other hand, postulates that the sporophyte generation gained complexity by “recruiting” the expression of gametophytic genes. In this case, the zygote would have been retained in the gametophyte due to meiosis delay and repeated mitotic divisions (Bower, 1890), producing a dependent sporophyte as the one observed in Bryophytes. According to this model, it is considered that the sporophyte would have increased in size and complexity during evolution to become the dominant phase in the life cycle of vascular plants. The gametophyte on the other hand would have been dramatically reduced as found in Angiosperms where it only comprises the embryo sac and the pollen grain.

1.2 Gene families involved in the evolution of the life cycle in plants

The genetic changes underlying the transitions in the life cycle of plants are not fully understood. However, a number of genes involved in the identity and maintenance of meristems in Angiosperms have been proposed to be important for sporophyte evolution such as the *KNOTTED-LIKE HOMEOBOX* (*KNOX*) genes, the *MADS–box* transcription factors, *LEAFY* (*LFY*) and *TEOSINTE BRANCHED 1 / CYCLOIDEA / PROLIFERATING CELL FACTOR 1* (*TCP*) (Bowman et al., 2007; Floyd and Bowman, 2007; Graham et al., 2000).

Among those, the class 1 *KNOX* genes are one of the best studied examples. The *KNOX* genes are involved in the regulation of the indeterminate Somatic Apical Meristem (SAM) in Angiosperms (Hake et al., 2004). They are members of the *TALE* (three amino acid loop extension) homeodomain containing transcription factors and can also interact with other *TALE* members, such as the *BELLRINGER* (*BELL*) proteins for functions in SAM maintenance, inflorescence and fruit patterning (review in Hay and Tsiantis, 2010). In the moss

Physcomitrella patens, mutations disrupting *class 1 KNOX* genes (*MKN2*, *MKN4* and *MKN5*) or *class 2 KNOX* genes (*MKN1* and *MKN3*) result in defects during sporophyte development (Sakakibara et al., 2008; Singer and Ashton, 2007). And in accordance, absence of *KNOX* gene expression in gametophytic meristems of Bryophytes has been reported (Frank and Scanlon, 2014). The ancient, conserved role of *KNOX* genes in the diploid generation is further evidenced in the algae *Chlamydomonas reinhardtii*. There, the *KNOX* ortholog *GAMETE-SPECIFIC MINUS1* (*GSM1*) interacts with the *BELL*-related *GAMETE-SPECIFIC PLUS1* (*GSP1*) to trigger diploid development (Lee et al., 2008). The expansion of the *KNOX* transcription factor family observed in Angiosperms with respect to early land plants (Hake et al., 2004; Hay and Tsiantis, 2010) correlates with an increase in sporophyte complexity, implying a possible relation of these genes with the evolution of the sporophyte.

More evidences indicating that sporophyte specific genes could have acquired new roles during evolution are given by the *MADS-box* and *LFY* transcription factor families. For the *MADS* genes expression was observed to be restricted to the sporophyte in the fern *Ceratopteris richardii* (Hasebe et al., 1998), while the *LFY* genes *PpLFY1* and *PpLFY2* in *P. patens* were shown to regulate zygote development (Tanahashi et al., 2005). In Angiosperms, members from both gene families have attributed functions in controlling floral meristem identity (Theissen et al., 2000; Weigel et al., 1992), suggesting emerging roles for those transcription factors that are already expressed in non-flowering plants. The *TCP* gene family represents a particularly interesting case, since its role as modulator of sporophyte architecture has been shown (review in Martín-Trillo and Cubas, 2010). *Class I TCP* genes in Angiosperms are expressed

in meristems, and it has been proposed that they promote plant growth and proliferation (Li et al., 2005), while Class II genes have been identified to repress the growth of axillar branches both in maize (Doebley et al., 1997) and in *Arabidopsis* (Aguilar-Martínez et al., 2007; Finlayson, 2007). In the moss *P. patens*, genes belonging to the two classes have been identified (Martín-Trillo and Cubas, 2010) and although expression in the sporophyte has been shown recently (Frank and Scanlon, 2014), functional characterization is missing.

Besides this scenario in which sporophytic genes would have gained roles during evolution to increase the complexity and diversity of the sporophytes in land plants, the recruitment of genes from the gametophyte to perform roles in the sporophyte has also been indicated (Dolan, 2009; Menand et al., 2007; Niklas and Kutschera, 2010). The *ROOT HAIR DEFECTIVE SIX LIKE1* (*RSL1*) gene constitutes a well-studied example. This basic helix–loop–helix transcription factor regulates root hair differentiation in *Arabidopsis*. In *P. patens*, the ortholog genes *PpRSL1* and *PpRSL2* are involved in rhizoid development, a tip growing gametophytic tissue with rooting functions in the moss. The conservation in gene function was further evidenced by complementation assays, indicating the enrolment of a gametophytic gene in sporophyte functions during plant evolution (Menand et al., 2007).

1.3 Reproductive novelties acquired by land plants

The changes in the reproductive systems and reproductive strategies during plant evolution were decisive, as exemplified by the evolutionary success of flowering plants (Friis et al., 2006; Kenrick and Crane, 1997). During plant adaptation to dryer environments the independence of water for sexual reproduction was achieved,

allowing plants to disperse far from water sources. In the first land plants such as Bryophytes and ferns, multicellular sex organs (the antheridia and archegonia) are comprised by a sterile jacket of cells surrounding the gametes. They depend on water for the opening of the archegonial neck that conducts to the egg cell, as well as for the release of flagellated sperm cells from the antheridia. Upon release, the motile sperm swim through water films to enter the archegonia and fertilize the egg (Campbell, 1971).

By contrast sexual reproduction in seed plants involves sperm cells contained in the pollen grain. This novel structure not only offers protection from desiccation since the cell wall formed by sporopollenin is very resistant (Boavida et al., 2005) but it also marks the loss of motility in the male gametes. Some Gymnosperms, such as *Ginkgo* and *Cycads* still possess motile sperm. However, they are produced in the pollen grain and depend on the liquid inside the pollen chamber to swim into the egg (Gifford and Larson, 1980). In other Gymnosperm groups, as well as in Angiosperms, sperm cells have lost the motile apparatus and depend on the pollen tube to be transported to the female gametophyte (Boavida et al., 2005; Crane and Leslie, 2013; McCormick, 1993). The evolution of seeds and flowers during the Cretaceus had a major impact on the success of land plant reproductive strategies. The angiosperm flower comprised by the reproductive structures (the pistils and the stamens) surrounded by the sterile, and often colorful, perianth was a key trait for the evolution of complex pollination strategies. In addition the double fertilization process that gave rise to a nourishing endosperm in Angiosperms increased the survival probabilities of the embryo (Crane and Leslie, 2013; Friedman and Williams, 2004; Friis et al., 2006). Although significant progress has been made unraveling the

genetic components involved in the differentiation and development of reproductive organs in flowering plants little is known about their counterparts in early land plants.

1.4 The genetic toolkit controlling sexual organ development in flowering plants

The developmental transitions required for sexual reproduction in Angiosperms are controlled by very complex gene networks acting at different stages. Functional studies of the model organism *Arabidopsis thaliana* have allowed the identification of key players for floral development, such as the *KANADI*, *APETALA* and *MADS-box* genes (review in Specht and Bartlett, 2009) for which homologs have been identified in the genomes of early land plants (Floyd and Bowman, 2007).The transitional changes from a vegetative to a reproductive phase in the apical meristem are controlled by the *MADS-box* transcription factor *FLOWERING LOCUS C* (*FLC*) that integrates several developmental and environmental cues to ensure flowering at the proper time. Inactivation of *FLC* by Polycomb proteins in response to vernalization leads to the activation of their target genes *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), promoting the transition to a reproductive phase (Crevillén and Dean, 2011). Through the interaction of the previously mentioned *LFY* with the floral homeotic protein *APETALA 1* (*AP1*), the floral meristem identity is established (Weigel et al., 1992). As opposed to animals, the germ cell lineage specification in plants occurs late in development. On a second reprogramming event, the floral somatic cells acquire a germ cell identity by a series of meiotic and mitotic cell divisions (review in Berger and Twell, 2011).

1.4.1 Female gametophyte development

In the female gametophyte of *Arabidopsis*, the functional megaspore undergoes three rounds of mitotic divisions (M I to III) to generate an eight nucleated cell (FG5). During cellularization, the different nuclei adopt distinct fates according to their position: two of the nuclei migrate to the center and fuse to produce the diploid central cell (CC); three of them migrate to the micropylar position where the egg cell (EC) and the two synergids (SY) are formed, being important for pollen tube attraction. The remaining three nuclei migrate to the chalazal pole to generate the antipodals (AP) that in *Arabidopsis* undergo cell death (Figure 2).

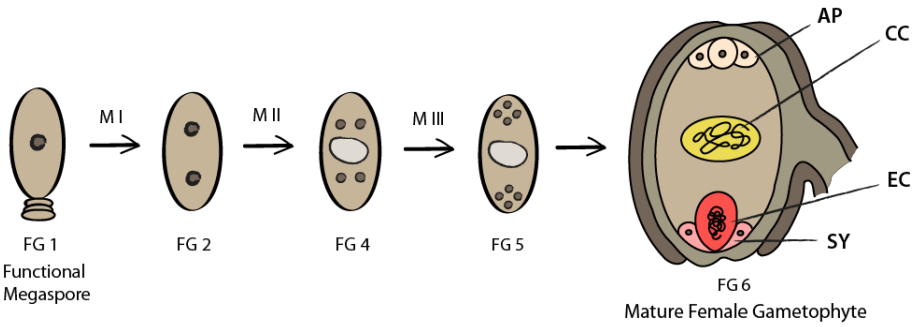


Figure 2. Female gametophyte development.
During female gametogenesis the functional megaspore undergoes three consecutive mitotic divisions (M I to III) to generate an eight nucleated cell (FG5). Subsequent cellularization and differentiation results in a seven-celled embryo sac consisting of two female gametes, the diploid central cell (CC) and the haploid egg cell (EC) with two types of accessory cells, three antipodals (AP) and two synergids (SY).

The genetic mechanisms controlling cell fate in the female gametophyte remain largely unknown, and only very few genes have been identified to be important for this process. The splicing factor *LACHESIS* (*LIS*) has been shown to prevent accessory cells from adopting a central cell fate (Groß-Hardt et al., 2007), while in the

eostre mutant the cell identity in the embryo sac is affected as a result of expression changes in the *BEL1-like* gene (Pagnussat et al., 2007). Finally, evidences indicate that the plant hormone auxin may play a central role for female gamete specification (Pagnussat et al., 2009).

1.4.2 Male gametophyte development

During male gametophyte development, the haploid microspores (MS) undergo an asymmetric division called Pollen Mitosis I (PM I) to produce the vegetative cell and the generative cell. During a second mitotic division (PMII), the generative cell divides symmetrically to generate the two sperm cells required for double fertilization. At the end of this division a tri-nucleate pollen is generated in *Arabidopsis* (Figure 3A) (Boavida et al., 2005; McCormick, 1993). The two sperm cells and the vegetative nucleus constitute the Male Germ Unit (MGU) that travels within the pollen tube to reach the female gametophyte. During the double fertilization process, the sperm cells fuse with the egg cell and the central cell to generate the embryo and the endosperm, respectively. The vegetative cell constitutes a companion cell and does not contribute with genetic material to the next generation (Boavida et al., 2005; Yadegari and Drews, 2004).

In recent years, great efforts have been made to identify the genetic components regulating male gametogenesis. This includes transcriptomic studies on maize (Engel et al., 2003; Ma et al., 2008), *Arabidopsis* (Becker et al., 2003; Borges et al., 2008; Honys and Twell, 2004; Pina et al., 2005), tobacco (Hafidh et al., 2012) and rice (Russell et al., 2012; Wei et al., 2010), comprising different cells and developmental stages (review in Rutley and Twell, 2015). To date,

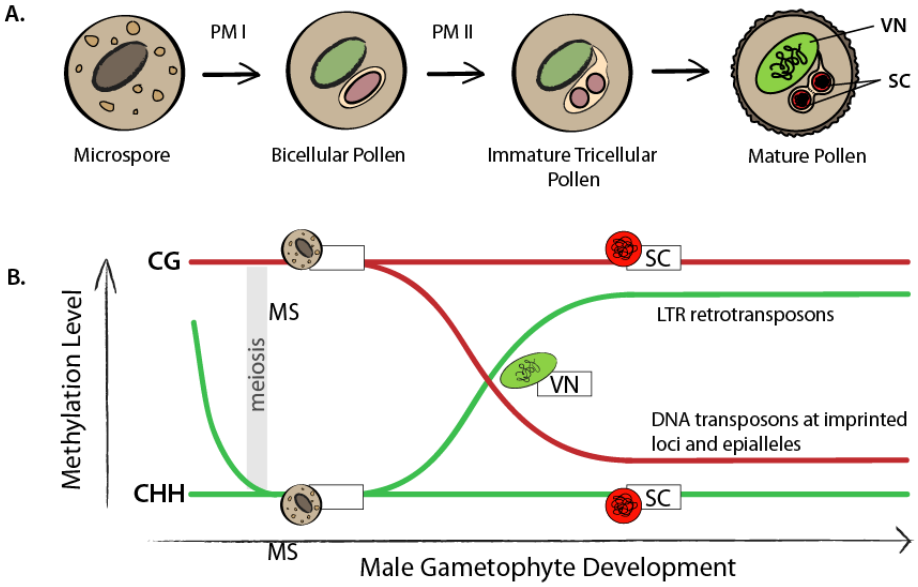


Figure 3. Male gametophyte development.
A. During male gametophyte development haploid microspores (MS) divide asymmetrically (PM I) to give rise to bicellular pollen, consisting of a vegetative cell harboring a smaller generative cell. The latter will undergo a second mitosis (PM II), originating two sperm cells (SC). **B.** Differences in the expression of methylation related genes result in distinct epigenetic landscape in the sperm cells (SC) and the vegetative nucleus (VN).

several genes important for the distinct steps of pollen development have been identified. This includes *STUD* (*STD*), involved in tetrad proper cytokinesis to generate haploid microspores (Hülskamp et al., 1997), and the *SIDECAR POLLEN* (*SCP*) protein required for the asymmetric division during PMI (Oh et al., 2010). Additionally, it has been shown that several *MADS-box MIKC** genes such as *AGAMOUS-LIKE 66* (*AGL66*) and *AGAMOUS-LIKE 104* (*AGL104*) have functions in pollen maturation and pollen tube growth (Adamczyk and Fernandez, 2009).

The discovery of the male gamete master regulator *DUO POLLEN 1* (*DUO1*) provided important insights of gene regulatory networks during male gametogenesis. The *R2R3 MYB* transcription factor *DUO1* is specifically expressed in the male germ line. It regulates PMII progression (Durbarry et al., 2005; Rotman et al., 2005) and activates several male germ line enriched genes for correct sperm cell specification (Borg et al., 2011; Brownfield et al., 2009). Among the targets of *DUO1* are the fusogenic protein *GENERATIVE CELL SPECIFIC 1* (*GCS1*) / *HAPLESS 2* (*HAP2*) with high conservation among eukaryotes (Liu et al., 2008), the *GAMETE EXPRESSED PROTEIN 1* (*GEX1*) shown to be important for male gametophyte development (Alandete-Saez et al., 2011; von Besser et al., 2006), and the male specific histone *MALE GAMETE HISTONE 3* (*MGH3*) / *HISTONE RELATED 10 variant* (*HTR10*) (Brownfield et al., 2009).

Functional studies in gametes of non-vascular plants are restricted to few examples, such as the *HAP2* homolog in *Chlamydomonas*, required for gamete fusion (Liu et al., 2008) or the previously mentioned *GSP1* and *GSM1* genes, expressed in the plus and minus gametes of the algae, respectively (Lee et al., 2008). In the moss *Physcomitrella*, the use of a gene-trap mutagenesis system allowed the identification of a number of genes expressed in the gametangia (Kofuji et al., 2009), however functional characterization is missing. In addition the role of auxin in the development of the reproductive organs of *P. patens* has been demonstrated (Landberg et al., 2013). Still, the unavailability of transcriptomic data from gametes in non-vascular plants hinders direct comparative studies from which evolutionary relationships could be established.

1.5 Epigenetics and imprinting in the context of male gametogenesis

The development of protocols to isolate the cells comprising the MGU as well as the microspores (Borges et al., 2012a), has allowed studying the transcriptomic and epigenetic changes occurring during male gamete development. Both in plants and in animals DNA methylation is associated with silencing of genes and transposable elements (TEs). In plants, maintenance methylation at symmetrical CG and CHG motifs, where H represents any residue but G, is accomplished by *DNA METHYLTRANSFERASE 1* (*MET1*) and *CHROMOMETHYLASE 3* (*CMT3*) (Jullien et al., 2012). *De novo* methylation in a CHH context relies on the production of 24-nt small interfering RNA (siRNA) and in *Arabidopsis* it is performed by the *DOMAIN REARRANGED METHYLTRANSFERASE 2* (*DRM2*), ortholog of the animal *DNMT3* enzymes (Cao et al., 2000; Daxinger et al., 2009). During male gametogenesis, epigenetic reprogramming involves dynamic changes in DNA methylation, chromatin remodeling and siRNA activity (reviewed in Gutierrez-Marcos and Dickinson, 2012) impacting in the development and function of the male gametes.

At early stages of male gamete differentiation, activation of TEs has been observed in the precursor cells or meiocytes (Borges et al., 2012b; Calarco et al., 2012), probably as result of reprogramming events. After meiosis, restoration of the repressive marks occurs in microspores where TEs are poorly represented (Calarco et al., 2012). However, the levels of CHH methylation decrease and during the following stages CHH methylation marks are only reestablished in the vegetative nucleus (VN), where *DRM2* is expressed (Figure 3B). In the sperm cells (SCs) loss of CHH methylation occurs preferentially at

class I retrotransposons suggesting the requirement of maternal siRNAs to reestablish silencing in the embryo, where restored CHH levels can be observed (Calarco et al., 2012).

Both in the sperm cells and in the embryo the levels of symmetric methylation (CG and CHG) are maintained. And in accordance, expression of *MET1* and *CMT3* in SCs has been detected (Borges et al., 2008; Jullien et al., 2012). In those cells, the maintenance of CG methylation marks on genes corresponding to maternally expressed loci generates the paternal allele repression (Calarco et al., 2012; Ibarra et al., 2012) required for imprinting. Differently, in the VN specific expression of the DNA glycosylase *DEMETER* (*DME*) and its homologs *DEMETER-LIKE 1* (*DML1*) / *REPRESSOR OF SILENCING* (*ROS1*), *DEMETER-LIKE 2* (*DML2*) and *DEMETER-LIKE 3* (*DML3*) produces active demethylation in CG specific sites. Those sites correspond to class II DNA transposons from MuDR and Helitron classes (Calarco et al., 2012; Ibarra et al., 2012), which have been shown to flank imprinting genes (Hsieh et al., 2011; McKeown et al., 2011; Wolff et al., 2011). In the SCs, the accumulation of corresponding 24nt siRNAs contributes to silence maternally expressed genes via RNA-directed DNA methylation (RdDM). The activation of TEs in the VN as a result of DNA glycosilase activity has been hypothesized to originate those 24nt siRNAs. Similarly, activation of *Athila* retrotransposons in the VN triggers production of corresponding 21nt siRNAs, which accumulate in the SCs, indicating a possible contribution to reinforce silencing in the germ cells (Slotkin et al., 2009).

Overall, the different mechanisms described indicate a complex relationship between the SCs and the VN, with profound

consequences for genome integrity and epigenetic inheritance. If similar mechanism exist in early land plants and if so, how are they are achieved is not known. And although it has been anticipated that imprinting may already occur in zygotes of early land plants (Haig and Wilczek, 2006), experimental evidences to unravel the specific imprinting mechanisms are missing.

2. The enigmatic RNA of sperm cells

Both in plants and in animals, the presence of messenger RNA (mRNA) and small non-coding RNA (nc RNA) in mature male gametes of diverse species such as *Lilium longiflorum* (Blomstedt et al., 1996), *Arabidopsis thaliana* (Borges et al., 2008; Pina et al., 2005), *Zea mays* (Engel et al., 2003), *Homo sapiens* (Krawetz et al., 2011; Ostermeier et al., 2002), *Mus musculus* (Grivna et al., 2006) and *Drosophila melanogaster* (Fischer et al., 2012) have been documented, challenging the previous idea of mature gametes being transcriptionally silent (reviewed in Miller et al., 2005). According to the great amount and diversity of transcripts detected, functional and non-functional roles have been proposed such as roles in fertilization and zygote development (see Dadoune, 2009; Hosken and Hodgson, 2014; Miller and Ostermeier, 2006). To date, the transmission of paternal elements besides its genome has been indicated, and the effect of some sperm produced transcripts in the resulting offspring has been shown. In mammals sperm centromeres are required for the first cell divisions in the zygote (Saunders et al., 2007, 2002), while sperm produced transcripts such as the *PREGNANCY SPECIFIC B-1-GLYCOPROTEIN 1* (*PSG1*) and the *HUMAN LEUKOCYTE ANTIGEN-E* (*HLA-E*) are demonstrated to be required for zygote implantation (Avendaño et al., 2009). Moreover, the

delivery of the paternal *FORKHEAD BOX G1B* (*FOXG1B*) mRNA necessary for embryo patterning has been demonstrated (Ostermeier et al., 2004). In *Caenorhabditis elegans*, paternal mRNAs were detected in one cell embryos and disruption of those transcripts by cross-linking was shown to cause embryo lethality (Stoeckius et al., 2014). In plants, however, the *SHORT SUSPENSOR* (*SSP*) gene constitutes the only example of paternal mRNA roles in embryo development. Whereas transcripts are specifically produced in pollen, accumulation of *SSP* protein was only observed in the embryo where it is required for proper embryo elongation through the YODA-MAP kinase pathway (Bayer et al., 2009).

Interestingly, transcripts corresponding to evolutionary new genes have been observed to be preferentially expressed in the male gametes. In animals numerous examples of evolutionary new genes showing testis-specific or testis-biased expression have been documented, such as the *PIPSL* in hominids generated by fused transcripts (Babushok et al., 2007), the *de novo* originated *JINGWEI* gene in *Drosophila* (Long and Langley, 1993) and the long non-coding *Pidi* originated by transcriptional activation of a specific region in mouse post-meiotic testis cells (Heinen et al., 2009). Based on those studies it has been postulated that the male organs such as testis in animals (Kaessmann, 2010) and pollen in angiosperms (Cui et al., 2014; Wu et al., 2014) may function as source of evolutionary innovation where new genes are preferentially expressed. According to the “out of the testis / out of the pollen” model it has been suggested that newly generated genes would initially be expressed in the male gametes, and later a more diverse expression pattern would evolve (Cui et al., 2014).

3. *Physcomitrella patens* as a model organism for evo-devo studies

In recent years, the moss *Physcomitrella patens* has emerged as a model organism for evo-devo studies. As a member of the Bryophytes it is characterized by the lack of a vascular system and alternation of generations with a gametophytic dominant life cycle (Graham et al., 2000; Qiu et al., 2006). Additionally, the availability of the complete genome sequence (Rensing et al., 2008), the high frequency of homologous recombination (Schaefer, 2001) and the possibility to perform targeted mutagenesis represents a great advantage for functional and evolutionary studies (Cove, 2005; Cove and Knight, 1993; Prigge and Bezanilla, 2010).

In the haploid phase of the moss life cycle (Figure 4), the spores germinate to produce the protonemal tissue comprised by two different cell types: the chloronema and the caulonema. The chloronemal cells are the first filaments to be generated, and from those the caulonema will differentiate. While the first are richer in chloroplasts and have a proposed assimilatory role, the caulonema present less chloroplasts and faster growth rate in accordance with possible adventitious roles. Some caulonema branches develop into leafy shoots called gametophores. At the tip of the gametophores the reproductive structures form. The archegonia (female structure) and the antheridia (male structure) develop together covered by leafy structures. After being released from the antheridia, motile sperm cells (antherozoids) swim into the archegonia to fertilize the egg. The diploid zygote, product of fertilization, constitutes the sporophyte phase of the moss. It is comprised by a short seta and the spore capsule, where the haploid spores are produced (Cove, 2005; Prigge and Bezanilla, 2010). Interestingly, the development of the

gametangia requires temperatures below 18ºC and short day lengths (Hohe et al., 2002) to trigger reproductive fate. Upon fertilization the diploid sporophytic structure develops and reaches maturity after 3-4 weeks. Spores are produced as a result of meiosis in cells inside the structure. Upon release of the spores the cycle starts again.

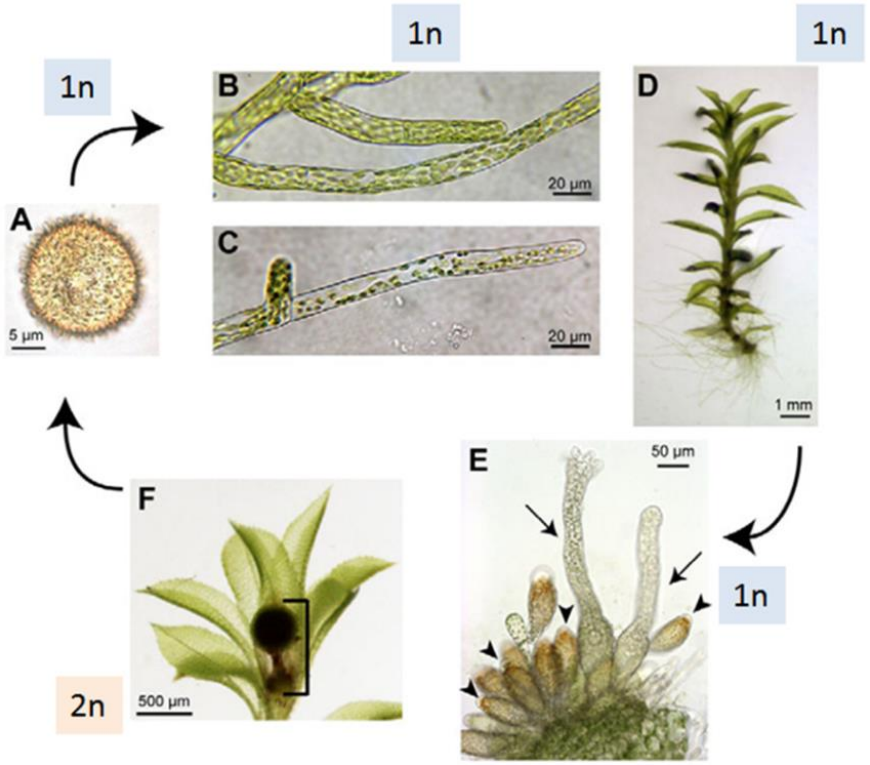


Figure 4. *Physcomitrella patens* life cycle.
The haploid dominant life cycle of the moss is shown. (A) Spores germinate to produce protonemal tissue (B and C) that will differentiate further into gametophores (D). At the tip of those, the gametangia develop (E). , Sperm cells (antherozoids) released from the antheridia swim to the archegonia and fertilize the egg. The product of fertilization is the sporophyte, representing the diploid generation (F). Inside, cells undergo meiosis to generate new spores. Image from Prigge and Bezanilla, 2010.

3.1 Gene expression in *Physcomitrella*

The availability of the *P. patens* genome has allowed comparative studies, revealing the evolution of gene families as well as the emergence of important traits for land plant adaptation. Based on comparisons with both algae and flowering plant genomes, the acquisition of more complex signaling systems, as seen by the increase of G protein coding genes, efficient mechanism for DNA repair and desiccation response genes are considered to have evolved in the moss. Hormone signaling pathways on the other hand, seem to have been acquired later by vascular plants, together with the expansion of several gene families (Rensing et al., 2008).

In *Physcomitrella*, the generation of mutants has been facilitated by the high frequency of homologous recombination, the available protocols for transformation and the haploid dominant life cycle. This has allowed to characterize numerous genes involved in different aspects, such as tip growth (Finka et al., 2007; Furt et al., 2013), hormone pathways (Landberg et al., 2013; Minami et al., 2003; Prigge et al., 2010), development (Sakakibara et al., 2008; Singer and Ashton, 2007; Sugiyama et al., 2012), and cell differentiation (Mosquna et al., 2009; Sakakibara et al., 2014) among many more. In addition, miRNAs pathways in the moss have been described (Arazi, 2012; Arif et al., 2012; Sung et al., 2008). With these reverse genetic approaches, conserved gene functions and evolutionary relationships have been established. However, large scale comparative studies are hindered by the lack of suitable expression data. In the last years great efforts have been undertaken to generate transcriptomic data for *P. patens*, so that available information to date includes different tissues and tissues subjected to distinct conditions.

For the gametophytic generation the transcriptome of protonema at different developmental stages, as well as isolated chloronema and caulonema cells was generated, showing dynamic expression changes on regulatory genes such as protein kinases and members of the *bHLH*, *MYB-like*, *GATA* and *bZIP* transcription factor families (Xiao et al., 2011). In addition transcriptomic responses in gametophores and protonema exposed to several biotic and abiotic stresses were recently reported (Hiss et al., 2014). In the case of the sporophytic generation studies are limited due to difficulties to obtain the samples from the *Gransden* strain. This strain has been maintained in the laboratory for long and routinely asexual propagation in conducted. Consistently, low fertility rates have been reported (Perroud et al., 2011). By using the *Villersexel* strain, for which high rates of fertilization are observed, the transcriptome of mature and immature sporophytes was generated (O'Donoghue et al., 2013). The study showed that major transcriptional changes occurred during the transition from the gametophytic to the sporophytic generations. Moreover, in a recent study the transcriptome of *Gransden* sporophytes at an early stage of development was generated (Frank and Scanlon, 2014). According to the authors the observed expression of meiotic genes in the sporophytes (harvested at 10 days after fertilization) may account for the maintenance of meristem indeterminate fate. However, detailed information concerning sporophyte developmental progression is still missing and we lack expression data from isolated reproductive organs.

The identification of genes involved in the transition at specific developmental points of sporophyte development, together with data from the gametes is essential for conducting reproductive studies in

the moss. With the determination of important genetic components for reproduction and embryo development in Bryophytes comparative studies could be established, providing insights on the evolution of reproductive traits.

4. Scope of this thesis

With the availability of diverse plant genomes, evolutionary comparative studies have been possible and important changes underlying plant evolution have been identified. Changes in gene expression are considered to be essential for this process. And although data from flowering plants exist, our knowledge in algae and early land plants is limited. In this work the transcriptome of all major tissues of the Bryophyte *P. patens* was generated. Through comparative transcriptomic analysis enriched and preferentially expressed genes were identified for each tissue, and putative functions are discussed.

In Chapter II, the identification of important genes for sporophyte development is validated through the characterization of the *Pptcp5* mutant, and gene expression changes required for the evolution of the sporophyte are hypothesized. Moreover, comparison with the flowering plant *A. thaliana* allowed the establishment of conserved cores of gene expression, including genes involved in tip growth processes.

In Chapter III, gene expression data from immature and mature antheridia as well as from isolated antherozoids is presented, presenting for the first time data from male gametogenesis in Bryophytes. During moss spermatogenesis high numbers of enriched

genes were identified, displaying also specific expression patterns. The putative roles for those transcripts are presented, including functions in gamete processes and embryo development. Also, the tendency to express evolutionary new genes in the mature male gametes is documented.

This transcriptome ATLAS aims to provide data for evolutionary comparative studies, including samples for which expression data had not been generated before. The identification of important genes for male gametogenesis, sporophyte development and tip growth hopefully should improve our understanding of the genetic mechanisms underlying the adaptation of plants to drier and more challenging environments. Such adaptive traits are relevant for improving crop production.

5. 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*, 1713–1723.

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.

Alandete-Saez, M., Ron, M., Leiboff, S., and McCormick, S. (2011). Arabidopsis thaliana GEX1 has dual functions in gametophyte development and early embryogenesis. Plant J. *68*, 620–632.

Arazi, T. (2012). MicroRNAs in the moss Physcomitrella patens. Plant Mol. Biol. *80*, 55–65.

Arif, M.A., Fattash, I., Ma, Z., Cho, S.H., Beike, A.K., Reski, R., Axtell, M.J., and Frank, W. (2012). DICER-LIKE3 activity in Physcomitrella

patens DICER-LIKE4 mutants causes severe developmental dysfunction and sterility. Mol. Plant *5*, 1281–1294.

Avendaño, C., Franchi, A., Jones, E., and Oehninger, S. (2009). Pregnancy-specific β-1-glycoprotein 1 and human leukocyte antigen-E mRNA in human sperm: Differential expression in fertile and infertile men and evidence of a possible functional role during early development. Hum. Reprod. *24*, 270–277.

Babushok, D. V., Ohshima, K., Ostertag, E.M., Chen, X., Wang, Y., Mandal, P.K., Okada, N., Abrams, C.S., and Kazazian, H.H. (2007). A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. Genome Res. *17*, 1129–1138.

Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinnel, T., and Lukowitz, W. (2009). Paternal control of embryonic patterning in Arabidopsis thaliana. Science *323*, 1485–1488.

Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature *480*, 245–249.

Becker, D., Boavida, L.C., Carneiro, J., Haury, M., and Feijo, a (2003). Transcriptional Profiling of Arabidopsis Tissues. Society *133*, 713–725.

Bennici, A. (2008). Origin and early evolution of land plants: Problems and considerations. Commun. Integr. Biol. *1*, 212–218.

Berger, F., and Twell, D. (2011). Germline specification and function in plants. Annu. Rev. Plant Biol. *62*, 461–484.

Von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). Arabidopsis HAP2 (GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. Development *133*, 4761–4769.

Blackwell, W.H. (2003). Two Theories of Origin of the Land-Plant Sporophyte: Which Is Left Standing? Bot. Rev. *69*, 125–148.

Blomstedt, C.K., Knox, R.B., and Singh, M.B. (1996). Generative cells of Lilium longiflorum possess translatable mRNA and functional protein synthesis machinery. Plant Mol. Biol. *31*, 1083–1086.

Boavida, L.C., Becker, J.D., and Feijó, J. a. (2005). The making of gametes in higher plants. Int. J. Dev. Biol. *49*, 595–614.

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M., and Twell, D. (2011). The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. Plant Cell *23*, 534–549.

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.

Borges, F., Gardner, R., Lopes, T., Calarco, J.P., Boavida, L.C., Slotkin, R.K., Martienssen, R. a, and Becker, J.D. (2012a). FACS-based purification of Arabidopsis microspores, sperm cells and vegetative nuclei. Plant Methods *8*, 44.

Borges, F., Calarco, J.P., and Martienssen, R. a. (2012b). Reprogramming the epigenome in Arabidopsis pollen. Cold Spring Harb. Symp. Quant. Biol. *77*, 1–5.

Bower, F.O. (1890). The origin of a land flora: a theory based upon the facts of alternation.

Bowman, J.L., Floyd, S.K., and Sakakibara, K. (2007). Green Genes-Comparative Genomics of the Green Branch of Life. Cell *129*, 229–234.

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T., and Twell, D. (2009). A plant germline-specific integrator of sperm specification and cell cycle progression. PLoS Genet. *5*.

Calarco, J.P., Borges, F., Donoghue, M.T. a, Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J. a., Becker, J.D., et al. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell *151*, 194–205.

Campbell, E.O. (1971). Problems in the origin and classification of bryophytes with particular reference to liverworts. New Zeal. J. Bot. *9*, 678–688.

Cao, X., Springer, N.M., Muszynski, M.G., Phillips, R.L., Kaeppler, S., and Jacobsen, S.E. (2000). Conserved plant genes with similarity to

mammalian de novo DNA methyltransferases. Proc. Natl. Acad. Sci. U. S. A. *97*, 4979–4984.

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

Cove, D.J., and Knight, C.D. (1993). The Moss Physcomitrella-Patens, a Model System with Potential for the Study of Plant Reproduction. Plant Cell *5*, 1483–1488.

Crane, P.R., and Leslie, A.B. (2013). Major Events in the Evolution of Land Plants. In The Princeton Guide to Evolution, Losos Jhonathan B., ed. (Princeton New Jersey), pp. 143–151.

Crevillén, P., and Dean, C. (2011). Regulation of the floral repressor gene FLC: The complexity of transcription in a chromatin context. Curr. Opin. Plant Biol. *14*, 38–44.

Cui, X., Lv, Y., Chen, M., Nikoloski, Z., Twell, D., and Zhang, D. (2014). Young Genes out of the Male: An Insight from Evolutionary Age Analysis of the Pollen Transcriptome. Mol. Plant 1–11.

Dadoune, J.P. (2009). Spermatozoal RNAs: What about their functions? Microsc. Res. Tech. *72*, 536–551.

Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A.J.M., and Matzke, M. (2009). A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. EMBO J. *28*, 48–57.

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.

Durbarry, A., Vizir, I., and Twell, D. (2005). Male germ line development in Arabidopsis. duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. Plant Physiol. *137*, 297–307.

Engel, M.L., Chaboud, A., Dumas, C., and McCormick, S. (2003). Sperm cells of Zea mays have a complex complement of mRNAs. Plant J. *34*, 697–707.

Finka, A., Schaefer, D.G., Saidi, Y., Goloubinoff, P., and Zrÿd, J.P. (2007). In vivo visualization of F-actin structures during the development of the moss Physcomitrella patens. New Phytol. *174*, 63–76.

Finlayson, S.A. (2007). Arabidopsis TEOSINTE BRANCHED1-LIKE 1 regulates axillary bud outgrowth and is homologous to monocot teosinte branched. Plant Cell Physiol. *48*, 667–677.

Fischer, B.E., Wasbrough, E., Meadows, L. a, Randlet, O., Dorus, S., Karr, T.L., and Russell, S. (2012). Conserved properties of Drosophila and human spermatozoal mRNA repertoires. Proc. R. Soc. B Biol. Sci.

Floyd, S.K., and Bowman, J.L. (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.

Friedman, W.E., and Williams, J.H. (2004). Developmental evolution of the sexual process in ancient flowering plant lineages. Plant Cell *16 Suppl*, S119–S132.

Friis, E.M., Pedersen, K.R., and Crane, P.R. (2006). Cretaceous angiosperm flowers: Innovation and evolution in plant reproduction. Palaeogeogr. Palaeoclimatol. Palaeoecol. *232*, 251–293.

Fritsch, F.E. (1921). Thalassiophyta and the algal ancestry of the higher plants. New Phytol. *20*, 165–178.

Furt, F., Liu, Y.C., Bibeau, J.P., Tüzel, E., and Vidali, L. (2013). Apical myosin XI anticipates F-actin during polarized growth of Physcomitrella patens cells. Plant J. *73*, 417–428.

Gifford, E.M., and Larson, S. (1980). Developmental Features of the Spermatogenous Cell in Ginkgo biloba. Am. J. Bot. *67*, 119.

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

Grivna, S.T., Beyret, E., Wang, Z., and Lin, H. (2006). RESEARCH COMMUNICATION A novel class of small RNAs in mouse spermatogenic cells. Genes Dev. *20*, 1709–1714.

Groß-Hardt, R., Kägi, C., Baumann, N., Moore, J.M., Baskar, R., Gagliano, W.B., Jürgens, G., and Grossniklaus, U. (2007). LACHESIS restricts gametic cell fate in the female gametophyte of Arabidopsis. PLoS Biol. *5*, 0494–0500.

Gutierrez-Marcos, J.F., and Dickinson, H.G. (2012). Epigenetic reprogramming in plant reproductive lineages. Plant Cell Physiol. *53*, 817–823.

Hafidh, S., Breznenová, K., Růžička, P., Feciková, J., Čapková, V., and Honys, D. (2012). Comprehensive analysis of tobacco pollen transcriptome unveils common pathways in polar cell expansion and underlying heterochronic shift during spermatogenesis. BMC Plant Biol. *12*, 24.

Haig, D., and Wilczek, A. (2006). Sexual conflict and the alternation of haploid and diploid generations. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *361*, 335–343.

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.

Hasebe, M., Wen, C.K., Kato, M., and Banks, J. a (1998). Characterization of MADS homeotic genes in the fern Ceratopteris richardii. Proc. Natl. Acad. Sci. U. S. A. *95*, 6222–6227.

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

Heinen, T.J.A.J., Staubach, F., Häming, D., and Tautz, D. (2009). Emergence of a New Gene from an Intergenic Region. Curr. Biol. *19*, 1527–1531.

Hiss, M., Laule, O., Meskauskiene, R.M., Arif, M.A., Decker, E.L., Erxleben, A., Frank, W., Hanke, S.T., Lang, D., Martin, A., 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*, 530–539.

Hohe, a, Rensing, S., and Mildner, M. (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.

Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biol. *5*, R85.

Hosken, D.J., and Hodgson, D.J. (2014). Why do sperm carry RNA? Relatedness, conflict, and control. Trends Ecol. Evol. *29*, 451–455.

Hsieh, T.-F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., et al. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. Proc. Natl. Acad. Sci. U. S. A. *108*, 1755–1762.

Hülskamp, M., Parekh, N.S., Grini, P., Schneitz, K., Zimmermann, I., Lolle, S.J., and Pruitt, R.E. (1997). The STUD Gene Is Required for Male-Specific Cytokinesis after Telophase II of Meiosis inArabidopsis thaliana. Dev. Biol. *187*, 114124.

Ibarra, C. a., Feng, X., Schoft, V.K., Hsieh, T.-F., Uzawa, R., Rodrigues, J. a., Zemach, a., Chumak, N., Machlicova, a., Nishimura, T., et al. (2012). Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. Science (80-.). *337*, 1360–1364.

Jullien, P.E., Susaki, D., Yelagandula, R., Higashiyama, T., and Berger, F. (2012). DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr. Biol. *22*, 1825–1830.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Res. *20*, 1313–1326.

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

Kofuji, R., Yoshimura, T., Inoue, H., Sakakibara, K., Hiwatashi, Y., Kurata, T., Aoyama, T., Ueda, K., and Hasebe, M. (2009). Gametangia Development in the Moss Physcomitrella patens. In The Moss Physcomitrella Patens, C.D. Knight, P.-F. Perroud, and D. Cove, eds. (Oxford, UK.: Wiley-Blackwell), pp. 167–181.

Krawetz, S. a, Kruger, A., Lalancette, C., Tagett, R., Anton, E., Draghici, S., and Diamond, M.P. (2011). A survey of small RNAs in human sperm. Hum. Reprod. *26*, 3401–3412.

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*, 1406–1419.

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

Leliaert, F., Verbruggen, H., and Zechman, F.W. (2011). Into the deep: New discoveries at the base of the green plant phylogeny. BioEssays *33*, 683–692.

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

Li, C., Potuschak, T., Colón-Carmona, A., Gutiérrez, R.A., and Doerner, P. (2005). Arabidopsis TCP20 links regulation of growth and cell division control pathways. Proc. Natl. Acad. Sci. *102*, 12978–12983.

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

Liu, Y., Tewari, R., Ning, J., Blagborough, A.., Garbom, S., Pei, J., Grishin, N.., Steele, R.., Sinden, R.., Snell, W.., et al. (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in Chlamydomonas and Plasmodium gametes. Genes Dev. *22*, 1051–1068.

Long, M., and Langley, C.H. (1993). Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila. Science *260*, 91–95.

Ma, J., Skibbe, D.S., John, F., and Walbot, V. (2008). Male reproductive development: gene expression profiling of maize anther and pollen ontogeny. Genome Biol. *9*, R181.

Martín-Trillo, M., and Cubas, P. (2010). TCP genes: a family snapshot ten years later. Trends Plant Sci. *15*, 31–39.

McCormick, S. (1993). Male Gametophyte Development. Plant Cell *5*, 1265–1275.

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

McKeown, P.C., Laouielle-Duprat, S., Prins, P., Wolff, P., Schmid, M.W., Donoghue, M.T., Fort, A., Duszynska, D., Comte, A., Lao, N.T., et al. (2011). Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thaliana seeds. *11*, 113.

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

Miller, D., and Ostermeier, G.C. (2006). Spermatozoal RNA: Why is it there and what does it do? Gynecol. Obstet. Fertil. *34*, 840–846.

Miller, D., Ostermeier, G.C., and Krawetz, S. a. (2005). The controversy, potential and roles of spermatozoal RNA. Trends Mol. Med. *11*, 156–163.

Minami, A., Nagao, M., Arakawa, K., Fujikawa, S., and Takezawa, D. (2003). Abscisic acid-induced freezing tolerance in the moss Physcomitrella patens is accompanied by increased expression of stress-related genes. J. Plant Physiol. *160*, 475–483.

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

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*, 3567–3581.

Oh, S.A., Park, K.S., Twell, D., and Park, S.K. (2010). The SIDECAR POLLEN gene encodes a microspore-specific LOB/AS2 domain protein required for the correct timing and orientation of asymmetric cell division. Plant J. *64*, 839–850.

Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., and Krawetz, S. a. (2002). Spermatozoal RNA profiles of normal fertile men. Lancet *360*, 772–777.

Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamond, M.P., and Krawetz, S.A. (2004). Reproductive biology: delivering spermatozoan RNA to the oocyte. Nature *429*, 154.

Pagnussat, G.C., Yu, H.-J., and Sundaresan, V. (2007). Cell-fate switch of synergid to egg cell in Arabidopsis eostre mutant embryo sacs arises from misexpression of the BEL1-like homeodomain gene BLH1. Plant Cell *19*, 3578–3592.

Pagnussat, G.C., Alandete-Saez, M., Bowman, J.L., and Sundaresan, V. (2009). Auxin-dependent patterning and gamete specification in the Arabidopsis female gametophyte. Science *324*, 1684–1689.

Perroud, P.F., Cove, D.J., Quatrano, R.S., and Mcdaniel, S.F. (2011). An experimental method to facilitate the identification of hybrid sporophytes in the moss Physcomitrella patens using fluorescent tagged lines. New Phytol. *191*, 301–306.

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. Plant Physiol. *138*, 744–756.

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

Prigge, M.J., and Bezanilla, M. (2010). Evolutionary crossroads in developmental biology: Physcomitrella patens. Development *137*, 3535–3543.

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

Qiu, Y., Li, L., Wang, B., Chen, Z., Dombrovska, O., Lee, J., Kent, L., Li, R., Jobson, R.W., Hendry, T. a., et al. (2007). A Nonflowering Land Plant Phylogeny Inferred from Nucleotide Sequences of Seven Chloroplast, Mitochondrial, and Nuclear Genes. Int. J. Plant Sci. *168*, 691–708.

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

Rensing, S. a, Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E. a, Kamisugi, Y., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science *319*, 64–69.

Rotman, N., Durbarry, A., Wardle, A., Yang, W.C., Chaboud, A., Faure, J.E., Berger, F., and Twell, D. (2005). A novel class of MYB factors controls sperm-cell formation in plants. Curr. Biol. *15*, 244–248.

Russell, S.D., Gou, X., Wong, C.E., Wang, X., Yuan, T., Wei, X., Bhalla, P.L., and Singh, M.B. (2012). Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. New Phytol. *195*, 560–573.

Rutley, N., and Twell, D. (2015). A decade of pollen transcriptomics. Plant Reprod. 1–17.

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.

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

Saunders, C., Swann, K., and Lai, F. (2007). PLCzeta, a sperm-specific PLC and its potential role in fertilization. Biochem. Soc. Symp. *74*, 23–36.

Saunders, C.M., Larman, M.G., Parrington, J., Cox, L.J., Royse, J., Blayney, L.M., Swann, K., and Lai, F.A. (2002). PLC zeta: a sperm-specific trigger of Ca(2+) oscillations in eggs and embryo development. Development *129*, 3533–3544.

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

Schmitz, R.J., Schultz, M.D., Lewsey, M.G., O'Malley, R.C., Urich, M.A., Libiger, O., Schork, N.J., and Ecker, J.R. (2011). Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. Science (80-.). *334*, 369–373.

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.

Slotkin, R.K., Slotkin, R.K., Vaughn, M., Vaughn, M., Borges, F., Borges, F., Tanurdzić, M., Tanurdzić, M., Becker, J.D., Becker, J.D., et al. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell *136*, 461–472.

Specht, C.D., and Bartlett, M.E. (2009). Flower Evolution: The Origin and Subsequent Diversification of the Angiosperm Flower. Annu. Rev. Ecol. Evol. Syst. *40*, 217–243.

Stoeckius, M., Grün, D., and Rajewsky, N. (2014). Paternal RNA contributions in the Caenorhabditis elegans zygote. EMBO J. *33*, 1740–1750.

Sugiyama, T., Ishida, T., Tabei, N., Shigyo, M., and Konishi, M. (2012). Involvement of PpDof1 transcriptional repressor in the nutrient

condition-dependent growth control of protonemal filaments in Physcomitrella patens. *63*, 3185–3197.

Sung, H.C., Addo-Quaye, C., Coruh, C., Arif, M.A., Ma, Z., Frank, W., and Axtell, M.J. (2008). Physcomitrella patens DCL3 is required for 22-24 nt siRNA accumulation, suppression of retrotransposon-derived transcripts, and normal development. PLoS Genet. *4*.

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

Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J.T., Münster, T., Winter, K.U., and Saedler, H. (2000). A short history of MADS-box genes in plants. Plant Mol. Biol. *42*, 115–149.

Wei, L.Q., Xu, W.Y., Deng, Z.Y., Su, Z., Xue, Y., and Wang, T. (2010). Genome-scale analysis and comparison of gene expression profiles in developing and germinated pollen in Oryza sativa. BMC Genomics *11*, 338.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., and Meyerowitz, E.M. (1992). LEAFY controls floral meristem identity in Arabidopsis. Cell *69*, 843–859.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T. a, Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C. (2011). High-Resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm. PLoS Genet. *7*.

Wu, D.-D., Wang, X., Li, Y., Zeng, L., Irwin, D.M., and Zhang, Y.-P. (2014). “Out of Pollen” Hypothesis for Origin of New Genes in Flowering Plants: Study from Arabidopsis thaliana. Genome Biol. Evol. *6*, 2822–2829.

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.

Yadegari, R., and Drews, G.N. (2004). Female Gametophyte Development. PLANT CELL ONLINE *16*, S133–S141.

CHAPTER II. A *Physcomitrella patens* Comprehensive Transcriptome Atlas Provides Insights into The Evolution and Development of Land Plants

Reformatted version of the manuscript submitted to *Molecular Plant*:

A *Physcomitrella patens* Comprehensive Transcriptome Atlas Provides Insights into The Evolution and Development of Land Plants (2015). Carlos Ortiz-Ramírez[a,1], Marcela Hernandez-Coronado[a,1], Anna Thamm[a], Bruno Catarino[b], Mingyi Wang[c], Liam Dolan[b], José A. Feijó[a,d], and Jörg D. Becker[a,2]

[a] Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal
[b] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK
[c] Division of Plant Biology, The Samuel Roberts Noble Foundation, Ardmore, Oklahoma, USA
[d] University of Maryland, Dept of Cell Biology and Molecular Genetics. 0118 BioScience Research Bldg, College Park, MD 20742-5815, USA

[1] These authors contributed equally to this work.
[2] Corresponding author: J.D.B. (jbecker@igc.gulbenkian.pt); phone +351-214464526; fax +351-214407970

Expression data is provided in the CD included in this Thesis, as well as the list of primers used. Supplementary Tables and Figures are included in this Chapter

Author contributions: The characterization of the gametangia and sporophyte development was conducted by me. Tissue isolation, RNA extraction and sample preparation was executed together with Carlos Ramirez, as well as qPCR to assess cDNA quality. Bruno Catarino collected and isolated the RNA from rhizoids. Anna Thamm performed the *PpTCP5* cloning and transformation. The phenotyping of mutant lines was done by Anna Thamm, Carlos Ramirez and me. Experimental design was conceived by Jörg Becker, Carlos Ramírez. Bruno Catarino and me. Writing of the manuscript was mainly done by Carlos Ramirez, Jörg Becker and José Feijó.

1. SUMMARY

Identifying the genetic mechanisms that underpins 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 through land plant evolution. Here, we report the generation of a transcriptome atlas covering most phases in the life cycle of the model bryophyte *Physcomitrella patens (P. patens),* including rhizoids and several sporophyte developmental stages. We identified the most complete set of sporophyte specific transcription factors reported so far, and detailed temporal resolution allowed us to establish a genetic basis of sporophyte developmental progression in mosses. Furthermore, we discovered that many transcription factors have homologs in angiosperms that function in developmental processes such as flowering or shoot branching. Deletion of one of these genes, *PpTCP5*, results in development of supernumerary sporangia attached to a single seta, suggesting that this gene negatively regulates branching in the sporophyte. Given that *TCP* genes repress branching in angiosperms, we suggest that this activity is ancient. Finally, by comparing gene expression profiles between *P. patens* and *Arabidopsis thaliana* (*A. thaliana*) we found that tip growing cells express a conserved core of transcription factors common to both species. We identified modifications in the expression patterns of genes comprising such core that could account for developmental differences between *P. patens* tip growing cells and *A. thaliana* pollen tubes and root hairs.

2. INTRODUCTION

It is believed that the first plants that colonized the terrestrial environment resembled to extant bryophytes (mosses, liverworts and hornworts) (Kenrick and Crane, 1997). Subsequent diversification of land plants and their adaptation to very diverse environmental conditions required a series of anatomical innovations, as well as alterations in their life cycles, which were accompanied by specific genetic modifications (Dolan, 2009). It has been proposed that the emergence of new gene families, together with horizontal gene transfer events could account for the appearance of some of the traits that were essential for plants to thrive on land (Yue et al., 2012). These include a more elaborated plant body architecture, the development of water conducting tissues, the specialization of reproductive organs, and the diversification of growing habits (Graham et al., 2000). However, genome-wide comparisons show that most of the gene families encoding basic developmental functions in flowering plants are already present in the genome of the bryophyte *P. patens* (Pires and Dolan, 2012; Rensing et al., 2008). Hence, it seems that the expansion of pre-existing families by gene duplications, as well as re-arrangements in their expression patterns leading to the emergence of new gene regulatory networks (GRNs), as proposed to animal evo-devo (Wilkins, 2002), might have been the most relevant processes responsible for the emergence of such traits.

As a result of the adaptation to land, the sporophyte and the gametophyte underwent considerable anatomical and morphological changes giving rise to the characteristic dimorphic alternation of generations present in bryophytes and vascular plants (Kenrick,

1994; Kenrick and Crane, 1997). The origin of a multicellular sporophyte attached to the gametophyte has been explained mainly by two different models: the antithetic and the homologous theories, both reviewed in (Blackwell, 2003). According to the more generally accepted antithetic theory, the retention of the zygote on the gametophyte is postulated as an emerging trait, product of meiosis delay and repeated mitotic divisions in the zygote (Bower, 1890). Under this scenario the sporophyte would have evolved from being a small and simplified structure, physiologically dependent on the gametophyte (like it is in mosses) to the complex, free-living, and sometimes huge organisms that we see today (Graham et al., 2000).

Several transcription factor families have been identified that probably were 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 their function in *P. patens* has been restricted to the sporophyte (Sakakibara et al., 2008; Singer and Ashton, 2007). Moreover, it was shown recently that the expression of the *KNOX* genes in non-vascular plants (including *P. patens*) is not related to gametophytic meristems (Frank and Scanlon, 2014) suggesting the recruitment of already present sporophytic transcription factors for new functions during land plant evolution. Whether this is the case for other important transcription factor families such as the *BELLRINGER 1-LIKE HOMEOBOX (BELL1)*, *MADS-box*, *GRAS*, and *TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR1 (TCP)* is still unknown. 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.

While it is thought that the sporophyte became more morphologically complex during land plant evolution, the opposite has been proposed for the gametophyte (Kenrick, 1994; Taylor et al., 2005). It would have evolved from being the dominant form (as it is in bryophytes) to a reduced phase in angiosperms, where it only comprises the embryo sac and the pollen grain. 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 (Dolan, 2009; Menand et al., 2007a; Niklas and Kutschera, 2010), as exemplified by the type II *MADS-box* transcription factors *MIKC^c* (Zobell et al., 2010) or the *RHD SIX-LIKE1* (*RSL)* genes (Menand et al., 2007a; Pires et al., 2013), whose functions are restricted to the gametophyte of *P. patens* but have a role in the sporophyte of the angiosperm *A. thaliana*. This is also true for specific processes such as apical growth, which is characterized by a restricted growth of the tip of the cell by exocytosis of specific cell wall precursors (Boavida et al., 2005; Campanoni and Blatt, 2007). Interestingly, apical growth can be found both in the gametophytic pollen tube and in the sporophytic root hairs in *A. thaliana*, but this is restricted to the gametophyte of *P. patens* where caulonema, chloronema and rhizoids can be considered as analogous structures (Rounds and Bezanilla, 2013). The existence of a conserved core of genes expressed in *A. thaliana* tip growing cells

suggests the possibility of a genetic apical growth signature (Becker et al., 2014), that may be evolutionary conserved. To further elucidate how the genetic basis underlying apical growth evolved, we could benefit from comparative gene expression analyses between tip growing cells from bryophytes and angiosperms.

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 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). The authors suggested that the expression of specific meiotic related genes early in sporophyte development may be the cause for its inability to achieve indeterminate growth. However, the lack of developmental resolution

makes it difficult to assess if those meiotic related genes are involved in sporophyte meristem control, or if their expression is rather temporarily restricted with the purpose of sporogenesis induction. In any case, detailed information concerning sporophyte developmental progression is still missing. Such information is essential in order to identify and characterize the set of genes that control development of the diploid generation in bryophytes. 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, 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/~asher/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.

3. RESULTS AND DISCUSSION

3.1 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. As a prerequisite for sample collection, the development of gametangia and sporophytes under our laboratory conditions was assessed (see material and methods). 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. 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 (Supplemental fig 1). We also performed qRT-PCRs on genes we identified as preferentially expressed in some of the tissues (see below), and saw a high correlation between the microarray and the qRT-PCRs expression data (Supplemental fig 2). Finally, we compared our microarray data with published experimental data, finding a high correlation on the detected tissue expression (Supplemental fig 3). Overall these observations validate our microarray results.

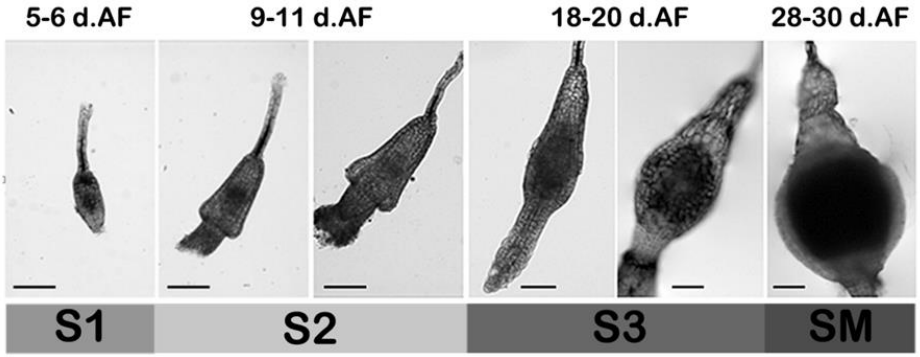


Figure 1. Sporophyte developmental stages. Four different developmental time points were defined and sporophytes collected based on the number of days that passed since the fertilization event (d.AF – days After Fertilization) until sporophyte maturation. Stages were denominated as Sporophyte 1 (S1),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

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/~asher/efp_physcomitrella/cgi-bin/efpWeb.cgi), for which diagrams representing the different *P. patens* tissues were generated (Fig 2). 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).

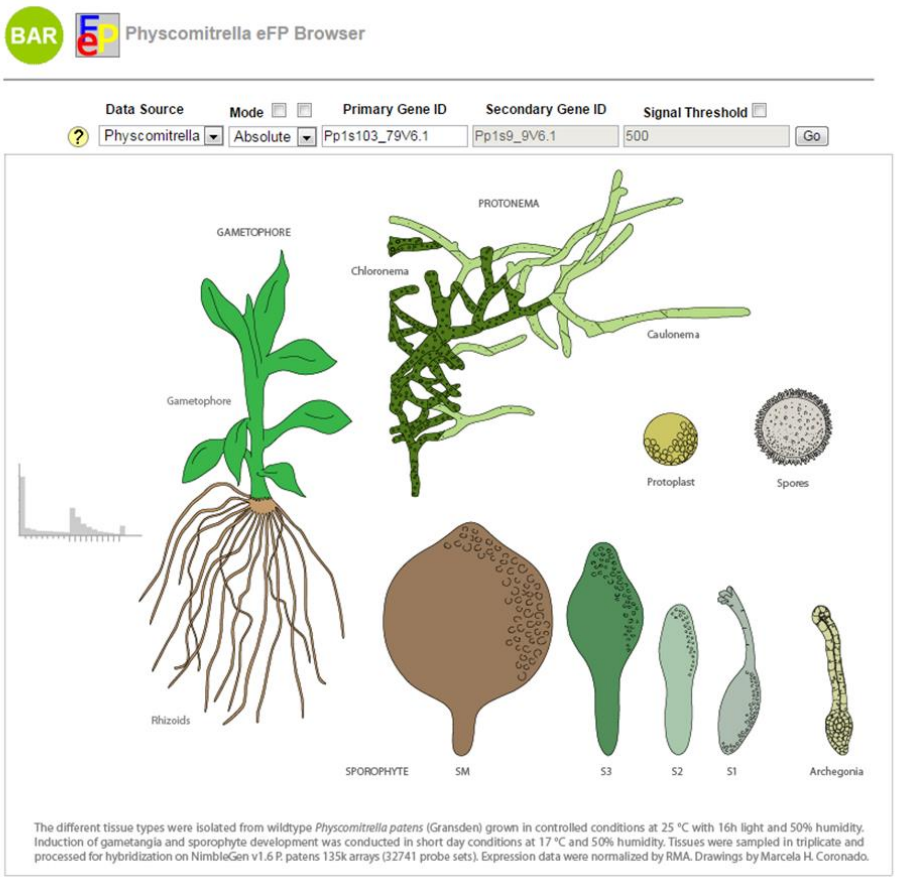


Figure 2. *Physcomitrella* eFP browser. Cartoon representation of *P. patens* tissues integrated into the Bio-analytical Resource for Plant Biology (BAR). Images, together with all expression data generated, are available in BAR for public consultation.

3.2 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). 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). It is interesting to note that the sporophyte transcriptome seems to be similar until the SM stage when the spores become mature, suggesting that the maturation of the sporophyte requires a global change in gene expression compared with the expression during earlier stages. Isolated spores also have a very different transcriptome than the rest of the tissues. Despite the fact that spores will become tip growing chloronema cells upon germination, they never appear close to tip growing cells in our analysis as we would expect if transcripts for rapid germination were stored in a similar fashion to the storage of transcripts in pollen grains (Becker et al., 2003). A comparison with available expression data from germinating spores (Hiss et al., 2014) could help to determine if the transcripts needed upon germination are actually stored, but this was beyond the scope of this study. Finally, the separation of caulonema from chloronema and rhizoid samples can be due to differences in tip growth physiology, reflecting

their different physiology and biological roles. 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.

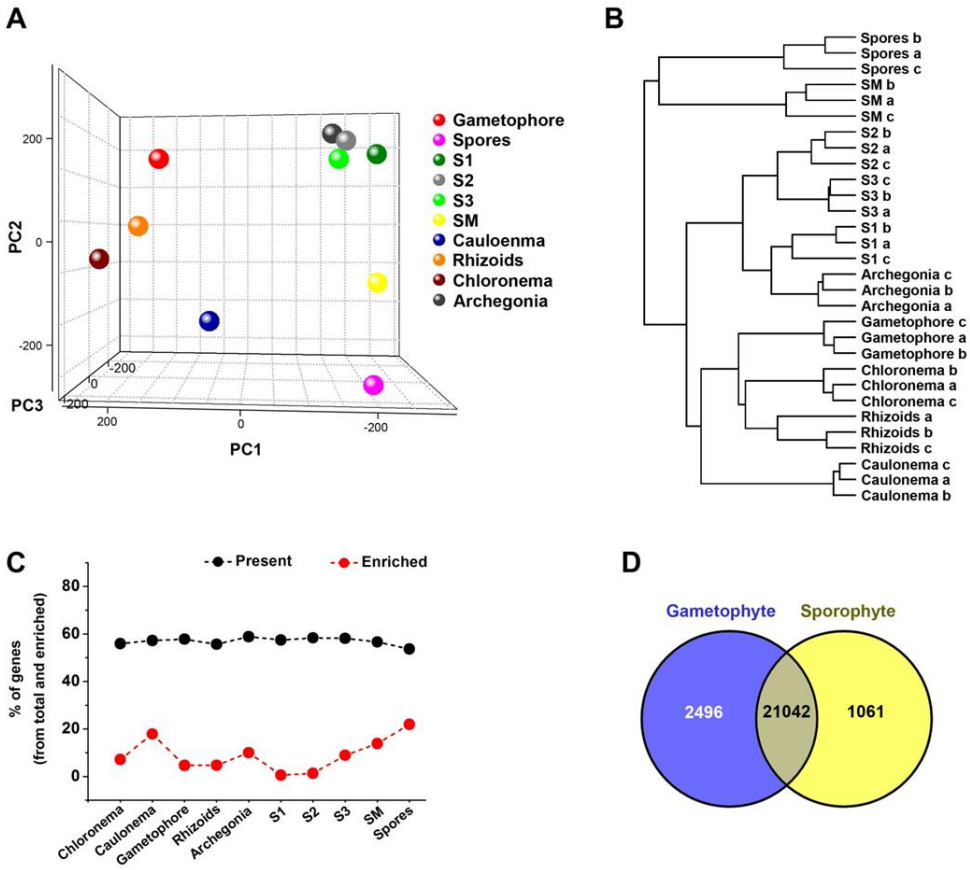


Figure 3. Global and tissue specific expression patterns. (a) PCA applied to the average gene expression values of 3 replicates per tissue. (b) Hierarchical clustering using Pearson´s correlation coefficient. (c) Percentage of tissue expressed genes (present) in relation to total number of genes represented 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.

3.3 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 4). 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 did additional analysis to determine the number of preferentially expressed genes in each generation. Interestingly, 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.

Differentially expressed genes at a tissue level 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: 635 were found; 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 as the structure developed and became mature (Fig 3c and table 1). The S1 and S2 pre-meiotic stages had the lowest number of enriched genes in the sporophytic set, probably as a consequence of not being fully differentiated in terms of morphology and development. This could be analogous to the maternal transcriptional dependence in early stages of embryo development in higher plants and animals (Baroux et al., 2008), meaning the sporophyte transcriptional program would have just started to be expressed at this point and only a limited number of genes would be detected as differentially expressed. It is interesting, however, that most of the enriched genes expressed at these initial stages were also preferentially expressed, suggesting very specific roles in early sporophyte development. Since most of them are transcription factors, it is tempting to hypothesize they are responsible for activating the expression of the sporophyte transcriptional program. As the sporophyte grew there was a significant increase in the number of enriched genes, from 38 genes

in S2 to 258 in S3 and 400 in the SM stage (Table 1), probably reflecting the fact that from the S3 stage onwards the sporophyte already acquired its full identity, both anatomically and transcriptionally. 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.

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

3.4 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 which is controlled by a set of transcription factors from which *KNOX*, *BELL1*, and *SCARECROW* 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. These 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. Finally, homologs of these families are also known to interact in the green algae *Chlamydomonas reinhardtii* to activate the expression of the diploid genetic program (Lee et al., 2008). Interestingly, our 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), suggesting that interaction between these proteins may be ancient. In addition, other important transcription factor transcripts detected in the sporophyte were a *SCR* homolog (Pp1s85_139v6), which is expressed during S2 and SM, and 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

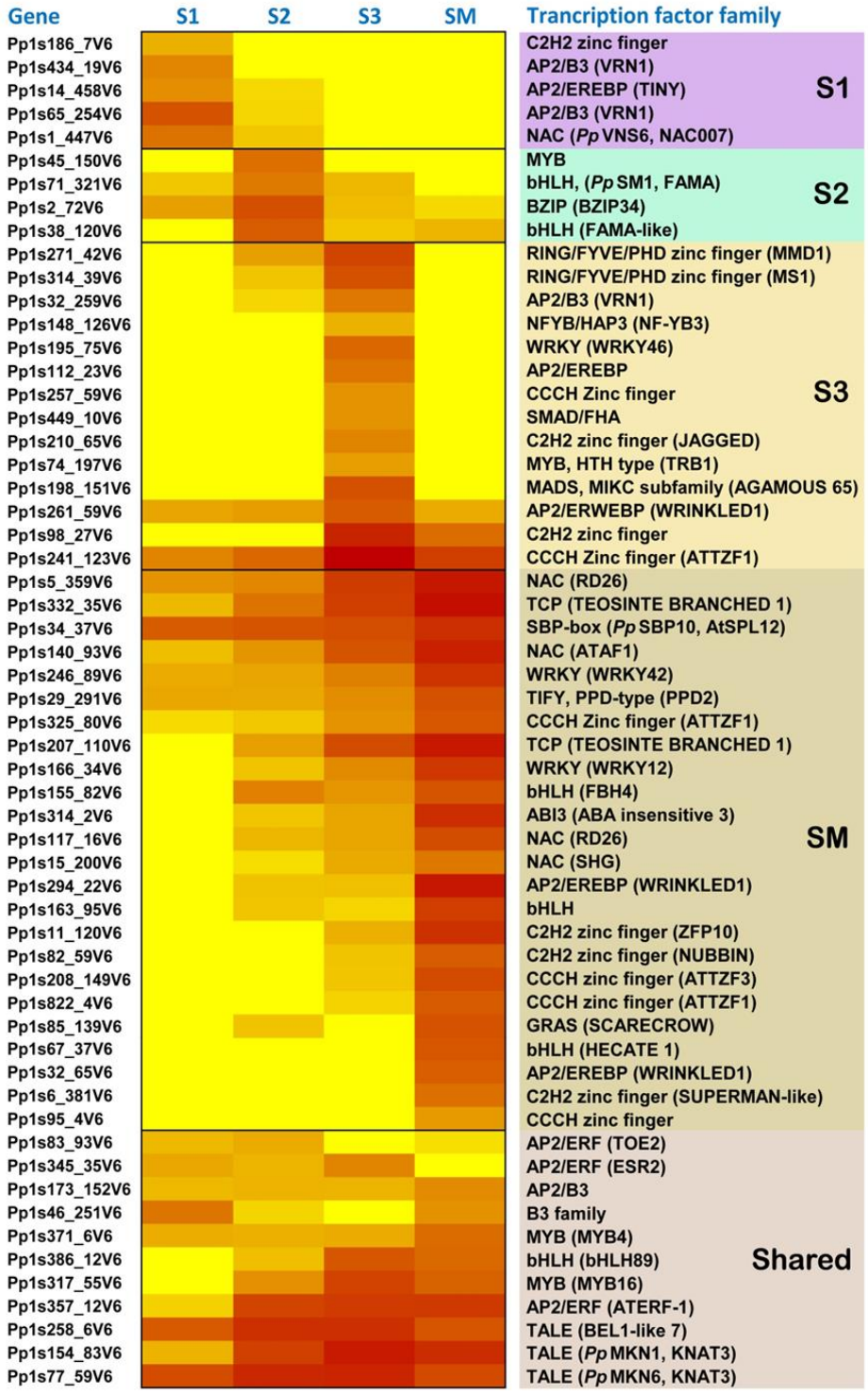

Gene
S1
S2
S3
SM
Trancription factor family
Pp1s186_7V6
Pp1s434_19V6
Pp1s14_458V6
Pp1s65_254V6
Pp1s1_447V6
Pp1s45_150V6
Pp1s71_321V6
Pp1s2_72V6
Pp1s38_120V6
Pp1s271_42V6
Pp1s314_39V6
Pp1s32_259V6
Pp1s148_126V6
Pp1s195_75V6
Pp1s112_23V6
Pp1s257_59V6
Pp1s449_10V6
Pp1s210_65V6
Pp1s74_197V6
Pp1s198_151V6
Pp1s261_59V6
Pp1s98_27V6
Pp1s241_123V6
Pp1s5_359V6
Pp1s332_35V6
Pp1s34_37V6
Pp1s140_93V6
Pp1s246_89V6
Pp1s29_291V6
Pp1s325_80V6
Pp1s207_110V6
Pp1s166_34V6
Pp1s155_82V6
Pp1s314_2V6
Pp1s117_16V6
Pp1s15_200V6
Pp1s294_22V6
Pp1s163_95V6
Pp1s11_120V6
Pp1s82_59V6
Pp1s208_149V6
Pp1s822_4V6
Pp1s85_139V6
Pp1s67_37V6
Pp1s32_65V6
Pp1s6_381V6
Pp1s95_4V6
Pp1s83_93V6
Pp1s345_35V6
Pp1s173_152V6
Pp1s46_251V6
Pp1s371_6V6
Pp1s386_12V6
Pp1s317_55V6
Pp1s357_12V6
Pp1s258_6V6
Pp1s154_83V6
Pp1s77_59V6
C2H2 zinc finger
AP2/B3 (VRN1)
AP2/EREBP (TINY)
AP2/B3 (VRN1)
NAC (*Pp* VNS6, NAC007)
S1
MYB
bHLH, (*Pp* SM1, FAMA)
BZIP (BZIP34)
bHLH (FAMA-like)
S2
RING/FYVE/PHD zinc finger (MMD1)
RING/FYVE/PHD zinc finger (MS1)
AP2/B3 (VRN1)
NFYB/HAP3 (NF-YB3)
WRKY (WRKY46)
AP2/EREBP
CCCH Zinc finger
SMAD/FHA
C2H2 zinc finger (JAGGED)
MYB, HTH type (TRB1)
MADS, MIKC subfamily (AGAMOUS 65)
AP2/ERWEBP (WRINKLED1)
C2H2 zinc finger
CCCH Zinc finger (ATTZF1)
S3
NAC (RD26)
TCP (TEOSINTE BRANCHED 1)
SBP-box (*Pp* SBP10, AtSPL12)
NAC (ATAF1)
WRKY (WRKY42)
TIFY, PPD-type (PPD2)
CCCH Zinc finger (ATTZF1)
TCP (TEOSINTE BRANCHED 1)
WRKY (WRKY12)
bHLH (FBH4)
ABI3 (ABA insensitive 3)
NAC (RD26)
NAC (SHG)
AP2/EREBP (WRINKLED1)
bHLH
C2H2 zinc finger (ZFP10)
C2H2 zinc finger (NUBBIN)
CCCH zinc finger (ATTZF3)
CCCH zinc finger (ATTZF1)
GRAS (SCARECROW)
bHLH (HECATE 1)
AP2/EREBP (WRINKLED1)
C2H2 zinc finger (SUPERMAN-like)
CCCH zinc finger
SM
AP2/ERF (TOE2)
AP2/ERF (ESR2)
AP2/B3
B3 family
MYB (MYB4)
bHLH (bHLH89)
MYB (MYB16)
AP2/ERF (ATERF-1)
TALE (BEL1-like 7)
TALE (*Pp* MKN1, KNAT3)
TALE (*Pp* MKN6, KNAT3)
Shared

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 *A. thaliana* 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.

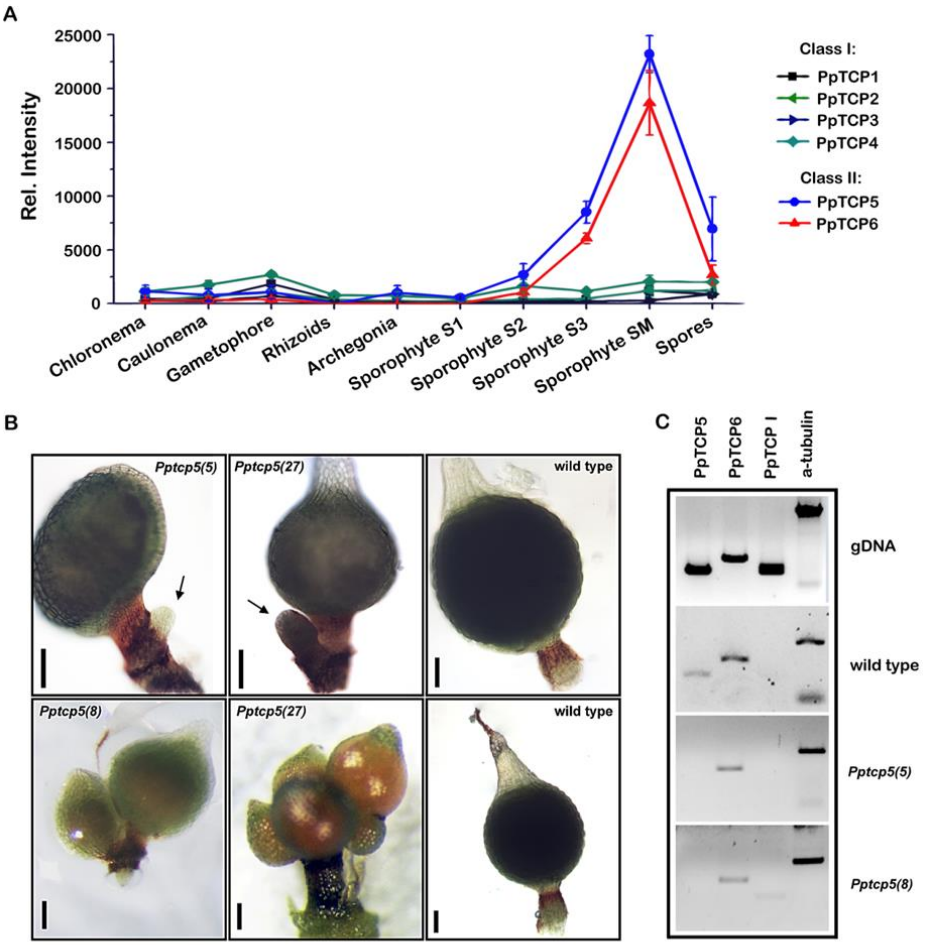
development, branching, and shoot symmetry (Aguilar-Martínez et al., 2007; Doebley et al., 1997; Koyama et al., 2007), and *SCR* transcription factors affect radial patterning in both shoot and roots (Wysocka-diller et al., 2000). Our data suggests that these major regulators of sporophyte development have similar expression profiles in both bryophytes and angiosperms. However, *KNOX*, *BELL1* and *TCP* are regarded as generators of sporophyte body plan diversity only in the latter group. This suggests that it was probably the evolutionary modifications in the temporal regulation of their expression and the reshaping of specific gene interactions networks that actually produced such diversity in angiosperms, and not merely their expression in the sporophytic generation.

3.5 *PpTCP5* represses branching during development of the *P. patens* sporophyte

TCP transcripts were abundant in the sporophyte of *P. patens,* almost as abundant as constitutively expressed *α-tubulin*. The control of shoot branching might be an ancient function of these transcription factors, since phylogenetic analyses have shown that they are present both in the genomes of basal land plant groups and in their algal relatives (Navaud et al., 2007). All *TCP* genes 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 (Martín-Trillo and Cubas, 2010; Navaud et al., 2007). Furthermore, studies in maize have shown there is an inverse relationship between the amount of *TEOCINTE BRANCHED 1 (TB1)* gene product (a class II *TCP*) and the degree of 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 5). 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. This pattern is consistent with the reported function for *TCP* class II as regulators of sporophyte development.

To investigate the function of class II genes in *P. patens,* we generated three independent knock out mutant lines of *PpTCP5* (Pp1s332_35V6) by homologous recombination (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). 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



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.

*Pptcp5*mutants was compared to *WT* upon transfer to sporophyte-inducing conditions. Since some mutant sporophytes exhibited branching, we compared sporophyte number and the percentage of branched sporophytes per biological replicate (one container with four gametophore-covered peat pellets), and at least four biological replicates for each line were included. All mutant lines developed more than twice as many branched sporophytes than *WT*. In this case, sporophyte branching was defined as a sporophyte bearing two or more capsules connected to the gametophyte by the same foot (Fig 5B, middle panels). 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, with some individuals producing as many as 5 capsules on a single structure (Fig 5B, lower row, middle panel).

Branching in mutant sporophytes originated from the seta region once the first sporangium was almost mature (Fig 5B, upper left panel). The second sporangium developed in close proximity to the previous with no clear seta or foot regions of its own (Fig 5B, lower left panel). This developmental pattern suggests that areas with meristematic competence exist in the region of the seta which can give rise to several sporangia once *TCP* expression is suppressed. Furthermore, it seems that the meristematic area giving rise to the new organ quickly acquires a reproductive developmental program, since the new structure develops as a sporangium without a seta, or alternatively, if the seta exists it is much reduced (Fig 5B, lower row,

middle panel). In turn, this further suggests that the meristem giving rise to the seta and foot region has a different identity and that these tissues probably originate when the meristem is going through vegetative growth.

These findings have implications for the understanding of polysporangiophytes evolution, i.e. plants with branched sporophytes and multiple sporangia (Tomescu et al., 2014). That is, the transition from a simple determinate sporophyte, like that of mosses, to the complex branched sporophytes of tracheophytes required not only the acquisition of a sporophytic apical meristem that allowed indeterminate growth, but also the proliferation of meristems that could give rise to lateral organs and branching patterns (Graham et al., 2000). Hence, meristem proliferation was necessary for shoot branching and the development of complex body plan architecture. In line with our results, we hypothesize that *TCP* transcription factors might have been involved in the control of this latter trait, since their role in repressing lateral organ primordia proliferation seems to be ancient, as reported for *Zea mays* (Doebley et al., 1997), *A. thaliana* (Aguilar-Martínez et al., 2007; Koyama et al., 2007) and now *P. patens*.

3.6 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 perform 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 the S2 stage the number of enriched genes increased to 38 and the 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. Interestingly, in the S3 stage 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

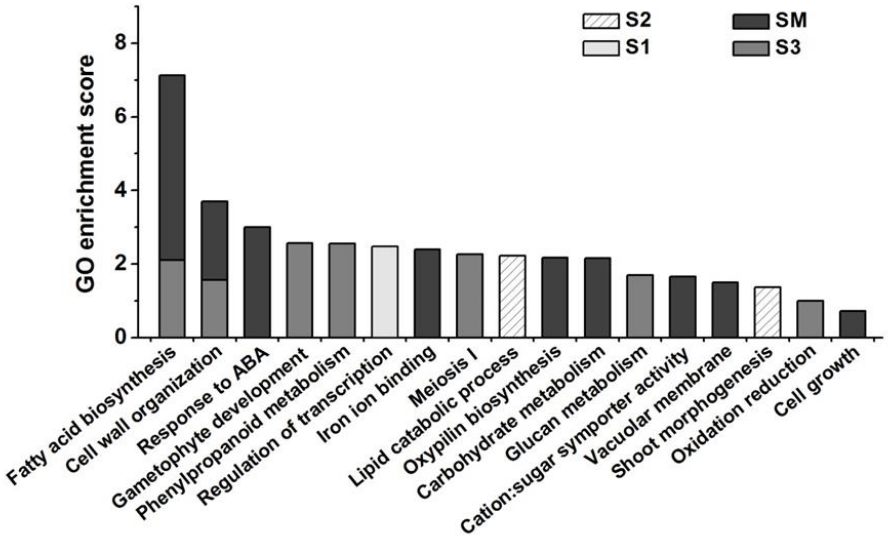


Figure 6. Gene Ontology enrichment analysis for sporophyte samples. GO terms for enriched genes in 4 different sporophytic developmental stages were clustered using DAVID tools and the terms with the highest enrichment scores ($p < 0.05$) for each developmental stage were selected.

during this specific stage sporogenesis is initiated. Furthermore, lipid biosynthetic processes are highly represented in S3, indicating that lipid metabolism is very important for sporophyte development (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. These processes are probably regulated through the action of hormones like oxylipin derivatives, since they have a role in the maturation of the angiosperm male gametophyte (pollen) (Park et al., 2002), and through ABA, which may also play a role in the inhibition of spore germination, although no studies on the effect of this hormone have been performed on the moss sporophyte.

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

During the S1 stage, we only detected 17 enriched genes from which 9 could not be annotated. However, all annotated genes are related with regulation of transcription. Of special interest are two transcription factors with high sequence similarity to the *A. thaliana REDUCED VERNALIZATION RESPONSE 1 (VRN1)* gene (Pp1s434_19V6 & Pp1s65_254V6). These transcripts seem to be important for early sporophyte development since they are 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 this species, the expression of *VRN1* increases during cold, and in turn it

can reduce the RNA levels of *FLOWERING LOCUS C (FLC)*, a key repressor of flowering, and increase the levels of *FLOWERING LOCUS T (FT)* and *SUPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)*, which promote flowering (Levy et al., 2002). 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 two putative orthologs in the early sporophyte stages of *P. patens* suggest that they are also important for maintaining meristem activity/identity in this species. This expression profile is also interesting because many mosses require exposure to cold temperatures for the formation of reproductive structures; hence, after fertilization young sporophytes develop under such 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 the transcription factor *PpSMF1*, an ortholog of the *bHLH FAMA*, being 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). Importantly, the function of this transcription factor in guard cell differentiation seems to be conserved, because transformation of *A.*

thaliana fama and *mute* loss of function mutants with a construct overexpressing *PpSMF1* was able to partially complement the stomata developmental defect (Macalister and Bergmann, 2011). Finally, we hypothesize that stomata development starts during the S2 phase of sporophyte development and that it seems to be restricted to this phase alone.

3.8 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 the *MALE MEIOCYTE DEATH 1 (MMD1-like)* gene (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.

3.9 Transcription factor homologs important for shoot organ morphogenesis in angiosperms are expressed in mature sporophytes

Unlike the S3 stage, most of the transcription factors enriched during SM are not unique to this stage. They are expressed at lower levels throughout the development of the sporophyte (fig 4), meaning they are likely to be important for the development of the immature sporophyte as well. The majority of these genes are homologs of transcription factors that in *A. thaliana* are important for the development of the aerial organs, including some flower tissue structures. Interesting examples are the *NUBBIN (NUB)* and *JAGGED* C2H2-type zinc finger transcription factor homologs (Pp1s82_59V6 and Pp1s210_65V6). In *A. thaliana, NUB* and *JAG* act together to control morphogenesis and late events of histogenesis in lateral organs. They affect leaf blade growth and petal cell numbers as well as anther development by controlling the formation of the adaxial microsporangia tissue, which is essential for pollen development (Dinneny et al., 2006; Ohno et al., 2004). In addition, we also detected the expression of several *SQUAMOSA (SPL)*

transcription factor orthologs which have similar roles to those of *JAG* and *NUB* in angiosperms: they are involved in leaf morphogenesis, phase transition, sporogenesis and flowering (Preston and Hileman, 2010; Stone et al., 2005; Zhang and Li, 2013). In particular, we found an ortholog of AtSPL12 (*PpSBP10* in *P. patens*), and two orthologs of *AtSPL8* (*PpSBP1* and *PpSBP8*), which belong to the group II and III of SPLs, respectively (Riese et al., 2007). In *A. thaliana,* mutants in *AtSPL8* have smaller anthers and produce less pollen due to a defect in the formation of the cell layers giving rise to the tapetum, which suggests that *PpSBP1* and *PpSBP8* might be important for spore development and maturation. These findings are relevant, because genes like *NUB, JAG* and *SPL*s, for which orthologs exist in *P. patens*, are necessary for the organ tissues to grow once the organ primordium is initiated (both in leaves and flowers), but they cannot create organ patterns by themselves. They rather translate information into cell types and tissue structures (Dinneny et al., 2006). Therefore, it was probably the co-option of some of these genes, which already support sporophyte tissue growth in mosses, together with the appearance of new genetic regulators of tissue patterning that lead to the formation of more complex vegetative and reproductive organs in the sporophyte, like those of vascular plants.

3.10 GO analysis assigns rapid tip growth, energy production and ion transport as specific functions to tip growing cell types

To get information about the biological functions important for protonema and rhizoids development, we carried out the same GO analysis as before (Fig 7). First we analyzed the set of genes commonly enriched in all tip growing cells: caulonema, chloronema and rhizoids. The most significant categories were associated, as expected , with tip growth physiology ; these included: cell wall

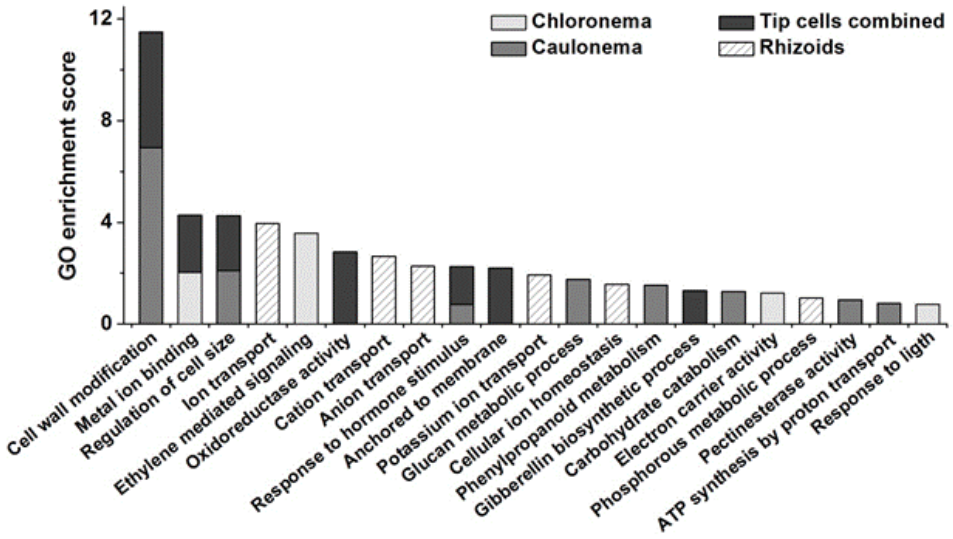


Figure 7. Gene Ontology enrichment analysis of tip growing cells. GO terms from enriched genes in tip growing cells were clustered using DAVID tools and the terms with the highest enrichment scores ($p < 0.05$) for each tissue, and all three combined, respectively, were selected.

modification, regulation of cell size, oxidoreductase activity, and metal ion binding (Fig 7). Previous analyses in *A. thaliana* have shown that similar GO categories are also over represented in pollen and root hairs (Becker et al., 2014; Qin et al., 2009). However, to assess if classes related to specific functions can be identified, we also analyzed the set of genes exclusively enriched in each tip growing cell type. 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). This suggests that 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. 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 physiological processes related with tip growth seem to be more active, with the most enriched categories being: cell wall modification, regulation of cell size, glucan metabolic process, and carbohydrate catabolism (Fig 7). Those categories are not enriched at the same levels in rhizoids or chloronema. In addition, it has been shown that caulonema grows significantly faster than chloronema (Duckett et al., 1998; Menand et al., 2007b), reinforcing the idea that the enrichment in tip growth related processes would support its faster growth rate.

3.11 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* MADS-box* gene families. Interestingly, several members of these families are known to control root hair and pollen tube development, respectively, in *A. thaliana* (Adamczyk and Fernandez, 2009; Karas et al., 2009; Menand et al., 2007a; 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., 2007a), 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., 2007a; Pires et al., 2013).

Similarly, in *A. thaliana* all transcription factors from the *MICK** subfamily, except one, are exclusively expressed in just one type of tip growing cell: the pollen tube (Verelst et al., 2007). In contrast, we found that 10 out of the 11 *MICKs** 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 3). 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 *MICKs** have been found to be predominantly expressed in the gametophyte (Hohe et al., 2002; Kwantes et al., 2012; Zobell et al., 2010).We performed a phylogenetic analysis using the aligned amino acid sequences from *P. patens* and the *A. thaliana MICKs**. 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 6). 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 *MICKs** have an ancient function in controlling general tip growth processes that later became more specific as vascular plants evolved.

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 4), while in *A. thaliana* a total of 49 were found (Becker et al., 2014). The lower number of TGE genes reported for 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., 2007a), 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 would have to have suffered a “shut down” of the expression of several genes that comprised such core, for example the *bHLH* transcription factors mentioned earlier, since apparently they were expressed in all tip growing cells from ancestral bryophyte plants. These observations are relevant because they support the antithetic theory of land plant evolution.

3.12 Expressed gene profiles define specific functions of rhizoids in *P. patens*

Analysis of preferentially expressed genes in rhizoids revealed that most of them encode proteins involved with ion transport and homeostasis, and protein phosphorylation. Examples are the homologs of *POTASSIUM TRANSPORTER 1 (KT1)* (Pp1s25_346V6), *PHOSPHATE TRANSPORTER 1;3 (PHT1;3)* (Pp1s302_62V6), and indirectly related with iron transport, *FERRIC CHELATE REDUCTASE* (FRO) (Pp1s90_3V6). Potassium, phosphate and iron are important inorganic macronutrients essential for plant growth, which in higher plants are obtained mainly from the soil through the roots, a process in which root hairs play a major role (Gilroy and Jones, 2000). Interestingly, our data suggests that the rhizoids are the cells involved in the uptake and exchange of essential inorganic nutrients in *P. patens*. Moreover, potassium is also necessary for the maintenance of cell turgor pressure and therefore it is involved in cell expansion and enlargement. *A. thaliana* mutants deficient in *AtKT1* potassium transporter activity produce shorter root hairs (Desbrosses et al., 2003), highlighting its role in cell

elongation and suggesting a possible role in rhizoid cell elongation. Finally, two of the preferentially expressed genes in rhizoids had no homologs in *A. thaliana*: a chromate transporter gene (Pp1s24_14V6), which is only present in moss and green algae genomes (Díaz-Pérez et al., 2007) and a putative alpha 2 delta subunit of the VOLTAGE GATED CALCIUM CHANNEL (*VGCC)* (Pp1s2_334V6), known to be expressed in animal excitable cells. This protein contains a MIDAS and a *VGCC-like* vWFA domain, which is the domain linking it with the animal calcium channel proteins. It is striking that we found 8 putative homologue genes with this domain in *P. patens* through BLAST (Supplemental table 5), but no homologous genes in any other available plant genomes, with the exception of *Selaginella moellendorffii,* which contains five. This finding suggests that voltage gated calcium channels were substituted in their functions by other calcium channels, namely *GLUTAMATE-RECEPTOR-LIKE (GLR)* which bear only 2 copies in *P. patens*, but have diversified and expanded dramatically in angiosperms (Lacombe et al., 2001).

3.13 The *OVATE* transcription factor family seems to have a major role in caulonema development

OVATE (OFP) family of transcription factors were exclusively detected in tip growing cells, especially in caulonema. From the 11 *OFP* genes identified in the genome of *P. patens* (Liu et al., 2014), one was never detected in any tissue, five were preferentially expressed in caulonema, three in tip growing cells and spores, and only two were significantly expressed in the sporophyte (Supplemental table 6). This finding was unexpected because *OFP* transcription factors control different aspects of sporophyte development in *A. thaliana* and tomato, and there are no reports

implicating these genes in tip growing cell developmental processes. For example, disruption of *AtOFP1* gene affects the length of several aerial organs by controlling cell elongation (Wang et al., 2007), while *AtOFP4* controls secondary cell wall formation in vessel elements and xylary fiber cells (Li et al., 2011). On the other hand *AtOFP5* has a role in the development and cell specification of the embryo sac (Pagnussat et al., 2007), and finally, overexpression of several *AtOFP*s produces morphological defects in hypocotyls, leaves and siliques (Wang et al., 2011). Because of their expression pattern in *P. patens,* it is tempting to speculate that OFP transcription factors may have a role in caulonema cell elongation, and in cell wall modification of tip growing cells. No functional characterizations of *OFPs* have been performed in *P. patens* yet, but their action seems to be mainly restricted to the gametophyte generation and tip growing cells since 8 out of the 10 *OFP* genes detected in the microarray are only expressed in the haploid generation. In contrast, expression of all *OFP* transcription factors in *A. thaliana* has been detected in the sporophytic generation (Wang et al., 2011). Moreover, genes like *AtOFP4* and *AtOFP5* have been reported to modulate the activity of TALE homeodomain transcription factors from the *KNOX* and *BELL* families, which are master regulators of sporophyte development (Li et al., 2011; Pagnussat et al., 2007). This strongly suggests that in the ancestors of vascular plants expression of the *OFP* transcription factors played a role in controlling the elongation of polarized cells mainly in the gametophyte generation, while in angiosperms they gained important new functions in the sporophyte as this structure became more conspicuous and complex. Importantly, this implies that the expression of these genes was acquired by the sporophyte during evolution, again supporting the antithetic theory.

3.14 The transcriptome of spores indicates extensive transcriptional and epigenetic reprogramming

Spores and pollen grains can be considered analogous structures to some degree. They are both haploid gametophytes surrounded by a thick protecting wall which are produced from diploid sporogenous cells that undergo meiosis. Their cell walls have a similar composition and their biosynthesis is controlled, in some cases, by the same genes. Examples are the *GAMYB* transcription factors *OsGAMYB* and *PpGAMYB* which control the biosynthesis of sporopollenin both in pollen and spores, respectively (Aya et al., 2011). In addition, they both produce tip growing cells upon germination. However, while the transcriptomes of pollen grains and pollen tubes show moderate alterations in *A. thaliana* (Qin et al., 2009), we found that the transcriptome of the mature spores is very different from that of *P. patens* tip growing cells. Moreover, the list of genes with enriched expression in spores differs a lot from those genes with enriched expression in pollen grains. Functional analysis performed on such genes revealed that not only the genes are different, but their function diverges significantly as well. In spores, highly enriched genes were involved mainly in RNA metabolism (Table 2). 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* (Ren et al., 2012; Zhao 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

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

Enriched gene list for pollen grains was obtained by performing pairwise comparisons of gene expression values from mature pollen against several *A. thaliana* tissue samples like leaves, siliques, seedlings, root hairs, ovaries, hydrated pollen tubes, sperm cells and flowers. Already available *A. thaliana* 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.

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* (He et al., 2009; Ikeda et al., 2011; Knizewski et al., 2008).

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 only in the mature spore transcriptome but not in the other tissues analyzed. One hypothesis would be that extensive changes in the transcriptional program are required to achieve a transition between the active metabolic state in which production and accumulation of nutrients for spore maturation are necessary, and the dormant state in which the spores have to remain until conditions are favorable to germinate. Encountering such conditions could in turn trigger another transition to a metabolically active growing state. In angiosperms, the seeds also have to go through these kinds of transitions, and it has been established that genes involved in chromatin remodeling processes , including

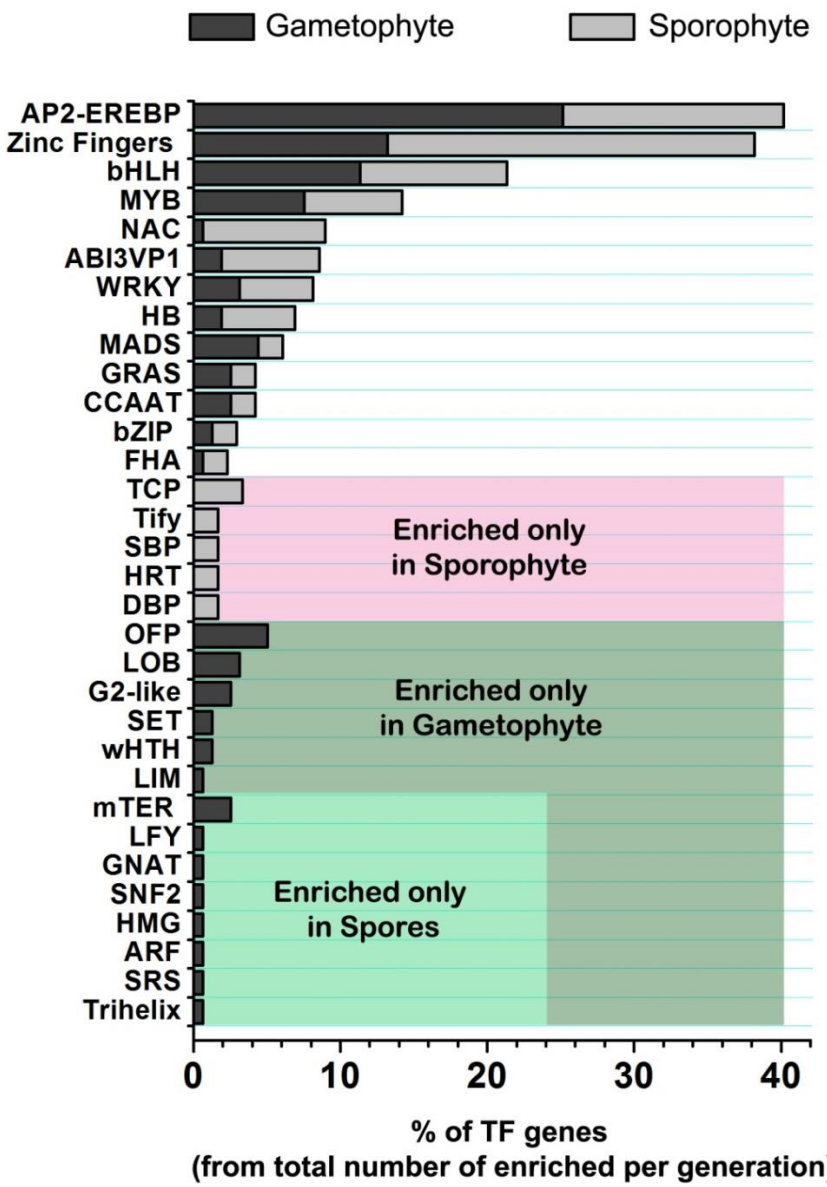


Figure 8. Transcription factor families in gametophytic and sporophytic generations. From the total number of enriched transcription factor genes in each generation, the percentage of genes belonging to a specific transcription factor family was calculated. Several transcription factor families are exclusively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

methylation, play a role in such rearrangements (van Zanten et al., 2013). Finally, we found several families of transcription factors that are only enriched in spores (Fig 8). Specific enrichment of transcription factor families was not observed in any other individual tissue, which further reflects the uniqueness of the spore transcriptome.

4. MATERIALS AND METHODS

4.1 Plant Material and Growth Conditions

The *Gransden* wild type strain from *Physcomitrella patens* Bruch & Schimp (Ashton and Cove, 1977) was used for this study. The strain was subjected to selfing at least three times a year. 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.

4.2 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, suggesting that fertilization events take place from day 15 on. 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).

4.3 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: using as a guide the developmental stages previously defined, 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 7), 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 7B). No tmema was included. Tissue was directly placed in Trizol until RNA extraction. Rhizoids were harvested from gametophores grown under normal conditions in Magenta boxes (Sigma-Aldrich) containing KNOPS minimal medium (Reski and Abel, 1985). Gametophore: 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.

4.4 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.

4.5 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.).

4.6 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. 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 (Becker et al., 2014; Borges et al., 2008; Pina et al., 2005; Qin et al., 2009; Swanson et al., 2005). 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 (Huang et al., 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 7. 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.

4.7 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 1). 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.

4.8 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.

4.9 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 8). Polyethylene glycol-mediated transformation was performed (Schaefer, 2001; Schaefer et al., 1991) 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 and the correct integration of the transgene on both 5' and 3' ends (using the primers oAT07 / oAT08 and oAT09 / oAT10, respectively) (Supplemental fig 8). 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.

4.10 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 due to excessive evaporation of water. Pptcp5 lines showed normal fertility rates in comparison with *Gransden* wild type, therefore samples from both genotypes with fertility rates below 43% were discarded for statistical analysis, considered as stress affected. The number of sporophytes growing from the same foot was quantified, being more than one considered as branching. Percentages of branching were calculated from the total number of sporophytes. Several independent biological replicates were counted per genotype. The observed sporophyte phenotype was confirmed in the three independent *Pptcp5* knockout lines that were phenotypically characterized and its statistical significance was evaluated by t-test.

4.11 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. 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.

4.12. Accession Number

Microarray data are available in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-3069. (Reviewers' account details are as follows: Username: Reviewer_E-MTAB-3069 Password: xoigesf7)

5. 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. 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. JAF acknowledges additional funding by grants PTDC/BEX-BCM/0376/2012 and PTDC/BIA-PLA/4018/2012 from Fundacão para a Ciência e a Tecnologia-FCT, Portugal. We thank 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.

6. 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*, 1713–1723.

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). Isolation and Preliminary Characterization of Auxotrophic and Analog Resistant Mutants of Moss, Physcomitrella-Patens. Mol. Gen. Genet. *154*, 87–95.

Aya, K., Hiwatashi, Y., Kojima, M., Sakakibara, H., Ueguchi-Tanaka, M., Hasebe, M., and Matsuoka, M. (2011). The Gibberellin perception system evolved to regulate a pre-existing GAMYB-mediated system during land plant evolution. Nat. Commun. *2*, 544.

Baroux, C., Autran, D., Gillmor, C.S., Grimanelli, D., and Grossniklaus, U. (2008). The maternal to zygotic transition in animals and plants. Cold Spring Harb. Symp. Quant. Biol. *73*, 89–100.

Becker, D., Boavida, L.C., Carneiro, J., Haury, M., and Feijo, J.A. (2003). Transcriptional Profiling of Arabidopsis Tissues. Society *133*, 713–725.

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.

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.

Blackwell, W.H. (2003). Two Theories of Origin of the Land-Plant Sporophyte: Which Is Left Standing? Bot. Rev. *69*, 125–148.

Boavida, L.C., Vieira, A.M., Becker, J.D., and Feijó, J.A. (2005). Gametophyte interaction and sexual reproduction: how plants make a zygote. Int. J. Dev. Biol. *49*, 615–632.

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.

Bower, F. (1890). On anthithetic as distinct from homologous alternation of generations in plants. Ann. Bot. *4*, 347–370.

Campanoni, P., and Blatt, M.R. (2007). Membrane trafficking and polar growth in root hairs and pollen tubes. J. Exp. Bot. *58*, 65–74.

Pina C., Pinto F., Feijó J.A., 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.

Desbrosses, G., Josefsson, C., Rigas, S., Hatzopoulos, P., and Dolan, L. (2003). AKT1 and TRH1 are required during root hair elongation in Arabidopsis. J. Exp. Bot. *54*, 781–788.

Díaz-Pérez, C., Cervantes, C., Campos-García, J., Julián-Sánchez, A., and Riveros-Rosas, H. (2007). Phylogenetic analysis of the chromate ion transporter (CHR) superfamily. FEBS J. *274*, 6215–6227.

Dinneny, J.R., Weigel, D., and Yanofsky, M.F. (2006). NUBBIN and JAGGED define stamen and carpel shape in Arabidopsis. Development *133*, 1645–1655.

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., Ligrone R. (1998). Protonemal morphogenesis. In Bryology for the Twenty-First Century, D.J.G. Bates, J.W., Ashton, N.W., ed. (Leeds, UK: British Bryological Society), 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.

Gilroy, S., and Jones, D.L. (2000). Through form to function: Root hair development and nutrient uptake. Trends Plant Sci. *5*, 56–60.

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. 2717–2722.

Hiss, M., Laule, O., Meskauskiene, R.M., Arif, M. a, Decker, E.L., Erxleben, A., Frank, W., Hanke, S.T., Lang, D., Martin, A., 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*, 530–539.

Hohe, A., Rensing, S., and Mildner, M. (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., Murata, T., Sato, Y., Nishiyama, T., Hiwatashi, Y., Imai, a., Kimura, M., Sugimoto, N., Akita, a., Oguri, Y., 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. *doi: 10.11*.

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.

Lacombe B., Becker D., Hedrich R., DeSalle R., Hollmann M., Kwak J.M., Schroeder J.I., Le Novère N., Nam H.G., Spalding E.P., Tester M., Turano F.J., Chiu J., C.G. (2001). The identity of plant glutamate receptors. Science (80-.). *292*, 1486–1487.

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*, 1406–1419.

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

Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R., and Dean, C. (2002). Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. Science *297*, 243–246.

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.

Li, E., Wang, S., Liu, Y., Chen, J.G., and Douglas, C.J. (2011). OVATE FAMILY PROTEIN4 (OFP4) interaction with KNAT7 regulates secondary cell wall formation in Arabidopsis thaliana. Plant J. *67*, 328–341.

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.

Liu, D., Sun, W., Yuan, Y., Zhang, N., Hayward, A., Liu, Y., and Wang, Y. (2014). Phylogenetic analyses provide the first insights into the evolution of OVATE family proteins in land plants. Ann. Bot. *113*, 1219–1233.

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

Martín-Trillo, M., and Cubas, P. (2010). TCP genes: a family snapshot ten years later. Trends Plant Sci. *15*, 31–39.

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

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

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*, 3567–3581.

Ohno, C.K., Reddy, G.V., Heisler, M.G.B., and Meyerowitz, E.M. (2004). The Arabidopsis JAGGED gene encodes a zinc finger protein that promotes leaf tissue development. Development *131*, 1111–1122.

Pagnussat, G.C., Yu, H.-J., and Sundaresan, V. (2007). Cell-fate switch of synergid to egg cell in Arabidopsis eostre mutant embryo sacs arises from misexpression of the BEL1-like homeodomain gene BLH1. Plant Cell *19*, 3578–3592.

Park, J.H., Halitschke, R., Kim, H.B., Baldwin, I.T., Feldmann, K. a., and Feyereisen, R. (2002). A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in Arabidopsis due to a block in jasmonic acid biosynthesis. Plant J. *31*, 1–12.

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

Pires, N.D., Yi, K., Breuninger, H., Catarino, B., Menand, B., 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*, 9571–9576.

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.

Preston, J.C., and Hileman, L.C. (2010). SQUAMOSA-PROMOTER BINDING PROTEIN 1 initiates flowering in Antirrhinum majus through the activation of meristem identity genes. Plant J. *62*, 704–712.

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.

Rensing, S. a, Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E. a, Kamisugi,

Y., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science *319*, 64–69.

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.

Riese, M., Höhmann, S., Saedler, H., Münster, T., and Huijser, P. (2007). Comparative analysis of the SBP-box gene families in P. patens and seed plants. Gene *401*, 28–37.

Rounds, C.M., and Bezanilla, M. (2013). Growth mechanisms in tip-growing plant cells. Annu. Rev. Plant Biol. *64*, 243–265.

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.

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.

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

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

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.

Stone, J.M., Liang, X., Nekl, E.R., and Stiers, J.J. (2005). Arabidopsis AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. Plant J. *41*, 744–754.

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*, 803–812.

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.

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

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.

Wang, S., Chang, Y., Guo, J., and Chen, J.-G. (2007). Arabidopsis Ovate Family Protein 1 is a transcriptional repressor that suppresses cell elongation. Plant J. *50*, 858–872.

Wang, S., Chang, Y., Guo, J., Zeng, Q., Ellis, B.E., and Chen, J.G. (2011). Arabidopsis Ovate family proteins, a novel transcriptional repressor family, control multiple aspects of plant growth and development. PLoS One *6*.

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

Wilkins, A.S. (2002). The evolution of developmental pathways (Sunderland, Massachusetts, USA: Sinauer Associates).

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.

Wysocka-diller, J.W., Helariutta, Y., Fukaki, H., Malamy, J.E., and Benfey, P.N. (2000). Molecular analysis of SCARECROW function

reveals a radial patterning mechanism common to root and shoot. *603*, 595–603.

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

Yang, 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*, 264–267.

Yue, J., Hu, X., Sun, H., Yang, Y., and Huang, J. (2012). Widespread impact of horizontal gene transfer on plant colonization of land. Nat. Commun. *3*, 1152–1159.

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

Zhang, H., and Li, L. (2013). SQUAMOSA promoter binding protein-like7 regulated microRNA408 is required for vegetative development in Arabidopsis. Plant J. *74*, 98–109.

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*, 1201–1211.

7. SUPPLEMENTAL DATA

Supplemental Figure 1. qRT-PCR to assess variation between replicates and stable expression of housekeeping genes.

Supplemental Figure 2. qRT-PCR of preferentially expressed genes.

Supplemental Figure 3. Correlation of microarray expression values and published evidence.

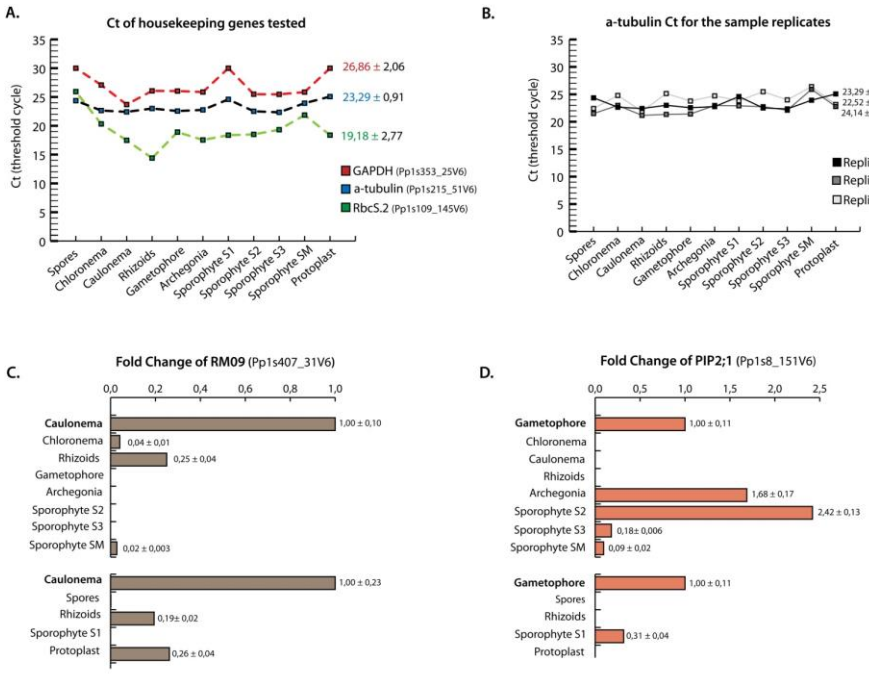
Supplemental Figure 4. Threshold determination for presence/absence detection calls.

Supplemental Figure 5. Phylogenetic analysis of TCP genes.

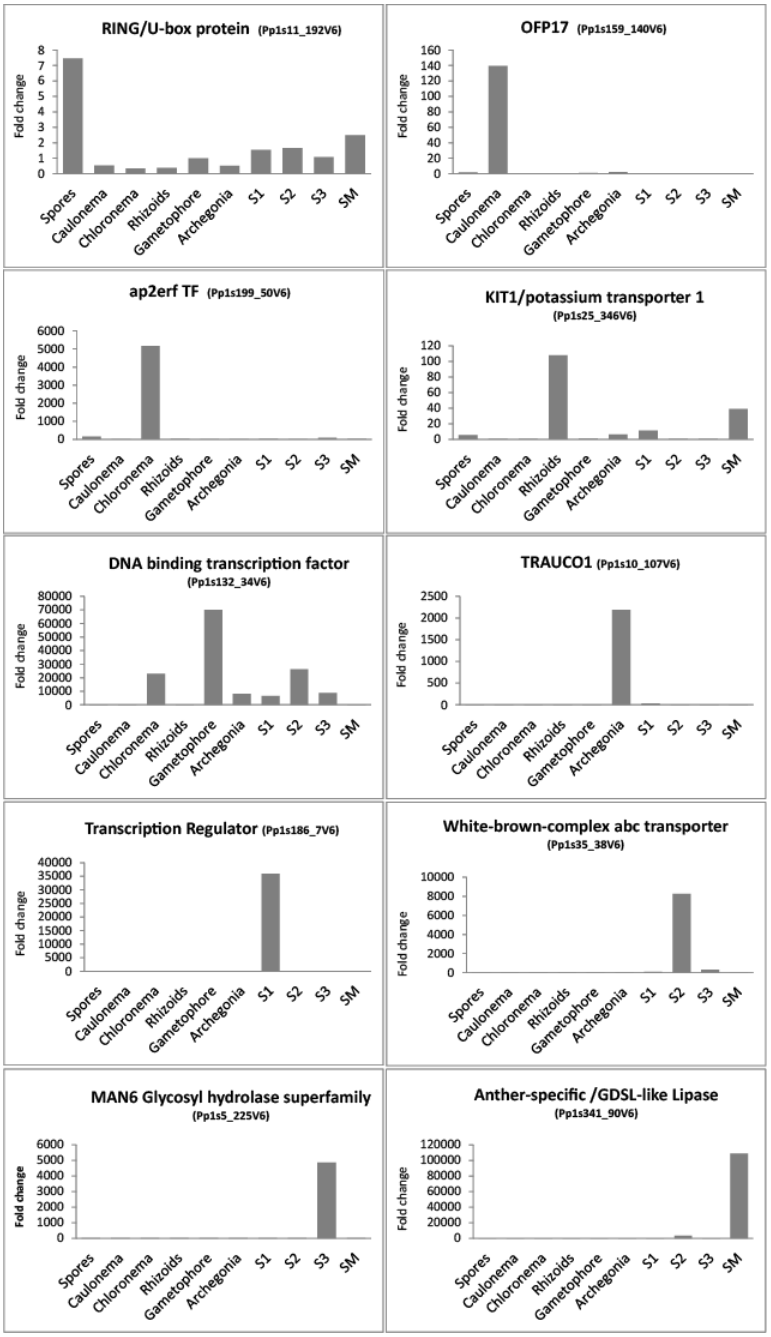
Supplemental Figure 6. Phylogenetic analysis of MICK* genes.

Supplemental Figure 7. Induction and isolation of caulonemal filaments.

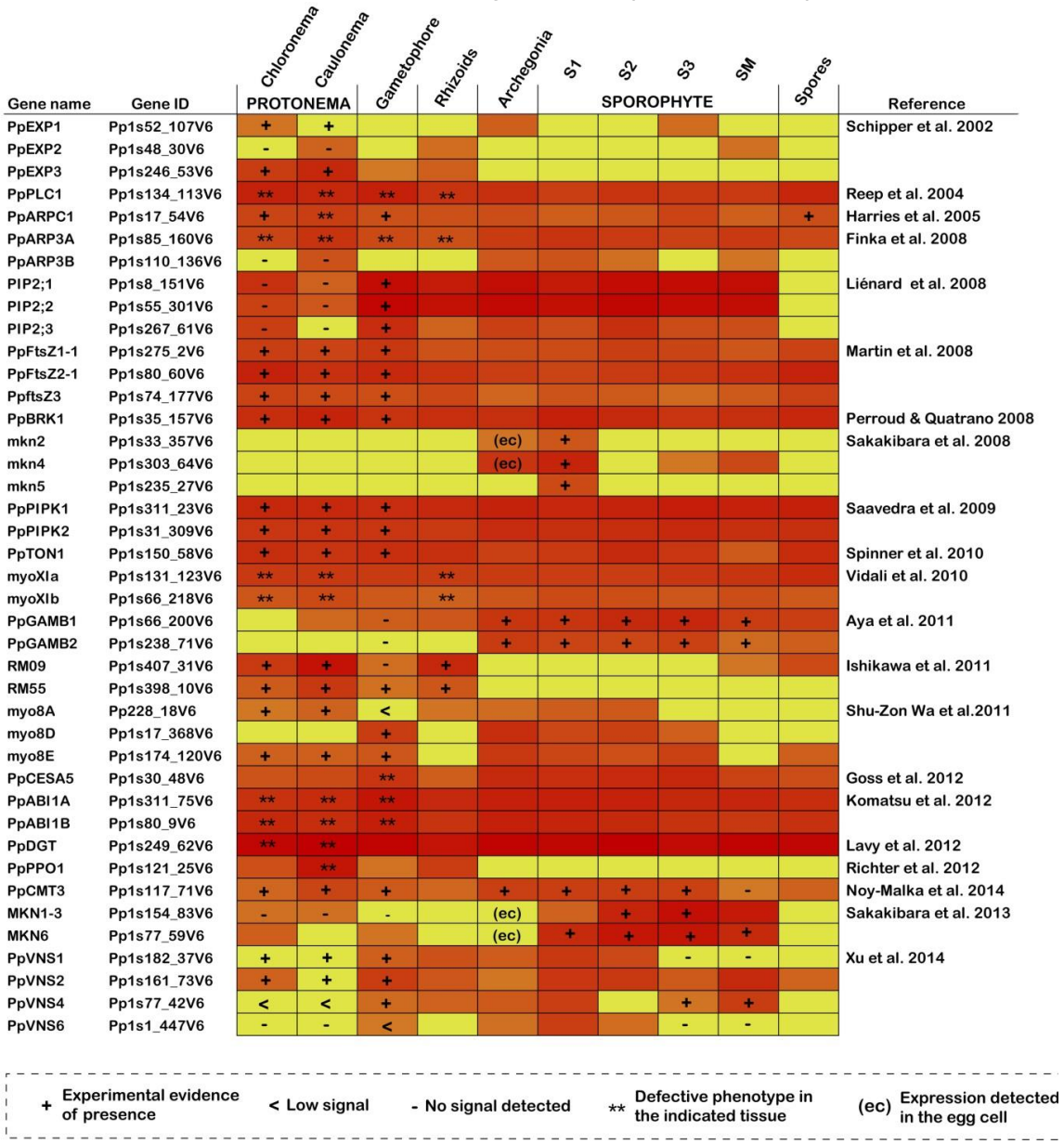
Supplemental Figure 8. Schematic representation of the construct used for PpTCP5 homologous recombination.



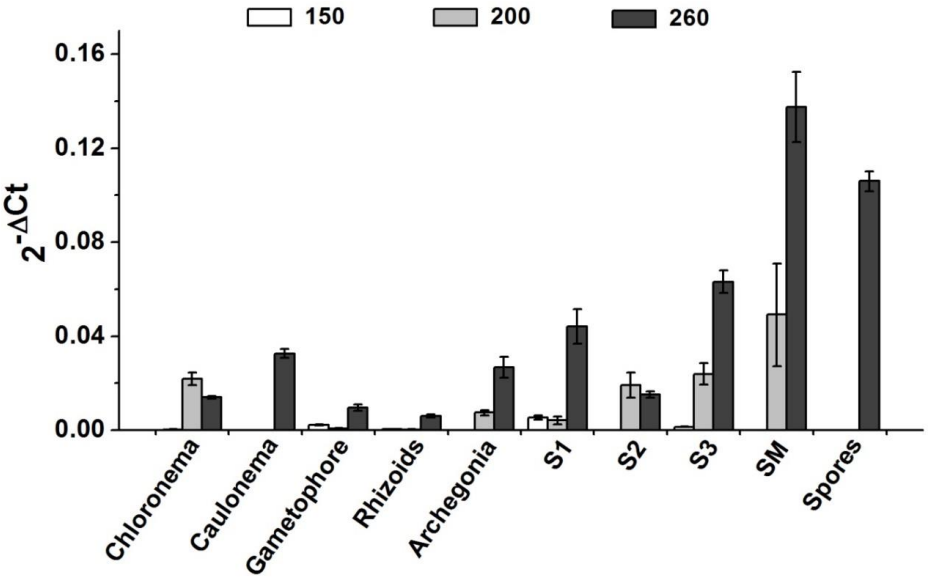
Supplemental figure 1. qRT-PCR to assess sample replicate quality. Quality of tissue sample replicates was tested by qRT-PCR using marker genes. (a) The comparison among reported housekeeping genes: Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), alpha-tubulin (a-tubulin) and Rubisco chloroplast encoded subunit (RbcS.2) shows a more stable profile for alpha-tubulin. (b) The housekeeping gene alpha-tubulin shows also low variation among sample replicates and tissue samples. (c) The expression of the RM09 gene (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. (d) In the same way, the PIP2;1 aquaporin (Pp1s8_151V6.1) was tested and the opposite pattern was obtain. 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 gen and caulonema or gametophore as the reference tissue. The standard error range (±) is depicted next to the bars.



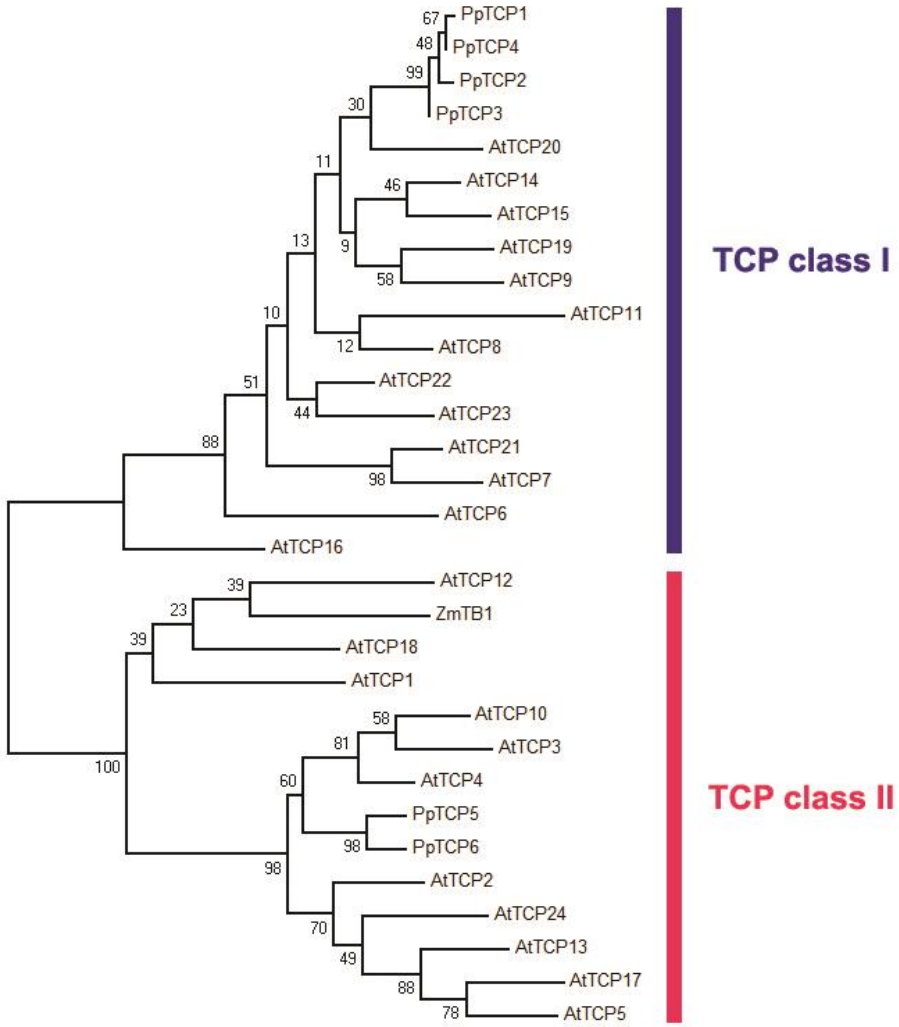
Supplemental figure 2. Preferential expression of genes. Genes with tissue preferential expression were selected for pattern corroboration by qPCR. cDNA from the three biological replicates was pooled. Data is shown as fold change ($2^{-\Delta\Delta ct}$) relative to gametophore expression in all cases except when the gene is preferentially expressed in gametophore, in which case rhizoids were used.



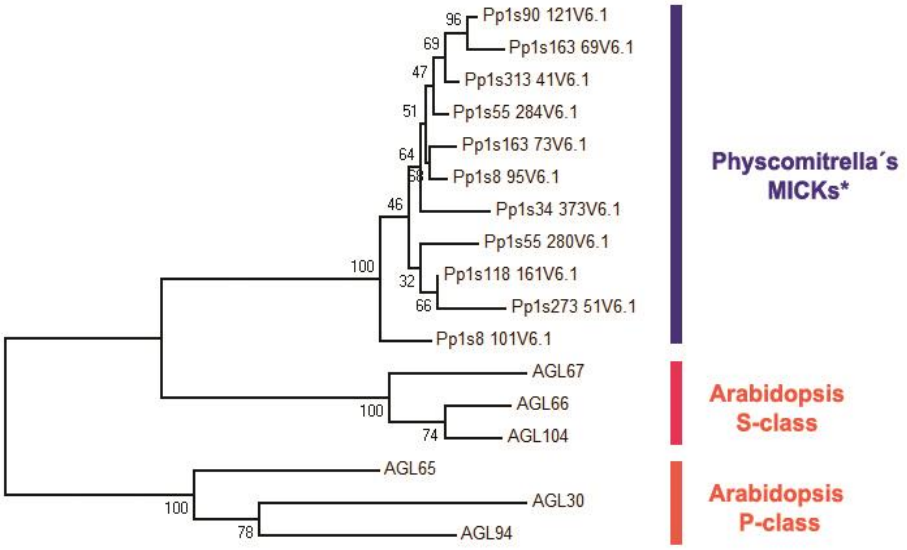
Supplemental figure 3. Correlation of microarray expression values and published evidence. A bibliographical search was performed for published data on some of the genes present in our microarray with the purpose of correlating the expression values with experimental evidence. Gene name and gene identifiers are displayed on the left, while references in which such genes are mentioned are displayed on the right. In 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.



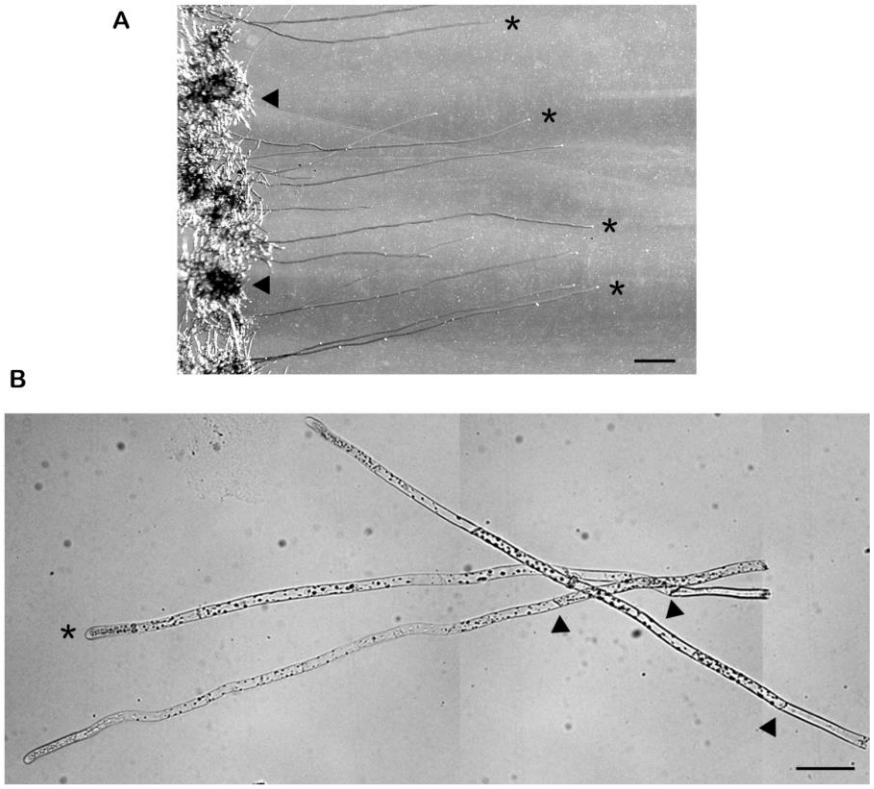
Supplemental figure 4. Threshold determination for presence/absence detection calls. qRT-PCR experiments were conducted using excess cDNA synthetized for microarray hybridization from all *Physcomitrella* tissues. After a preliminary analysis we expected the detection threshold to be between the relative expression values of 150 to 260 as reported in the mircroarray 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.



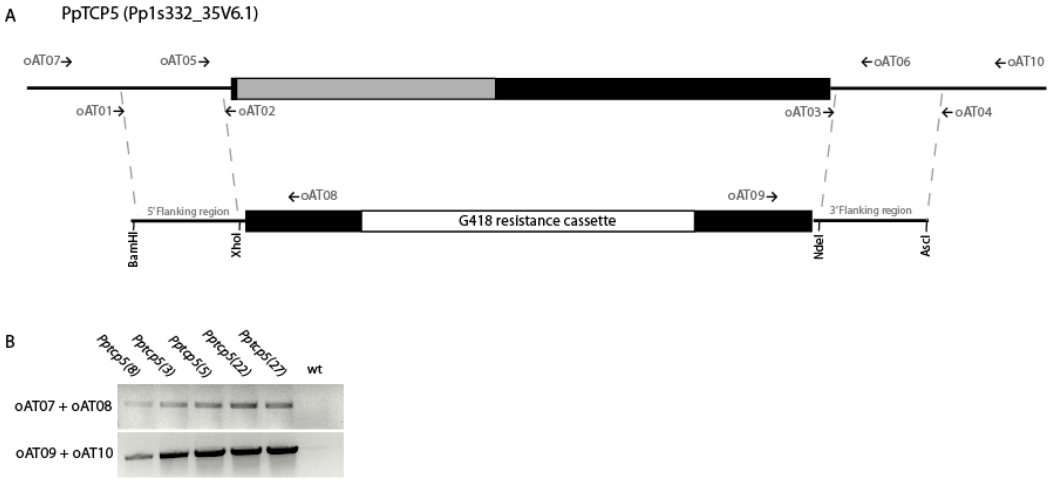
Supplemental figure 6. Phylogenetic analysis of MICK* genes. Phylogenetic tree constructed using maximum likelihood statistical method from aligned amino-acid sequences of MICK* transcription factors from *Arabidopsis thaliana (AGL)* and *Physcomitrella patens (Pp).* Posterior probability values are shown at the nodes of the tree.



Supplemental figure 5. 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).* Posterior probability values are shown at the nodes of the tree.



Supplemental figure 7. Induction and isolation of caulonemal filaments. 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) magnification of dissected caulonema showing the oblique cell plates (arrowheads) and clear tip region (asterisk) characteristic of caulonema. Scale bar = 50µM.



Supplemental figure 8. Schematic representation of the construct used for *Pp*TCP5 homologous recombination. (a) The localization of the primers used for the cloning and characterization of the knockout *Pptcp5* lines are represented. (b) 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.

Supplemental figure 3 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 moss Physcomitrella patens. Cell Motil. Cytoskeleton. 65: 769–784.

Goss, C.A., Brockmann D.J., Bushoven, J.T., Roberts, 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., Quatrano, R.S. (2005) Actin-related protein2/3 complex component ARPC1 is required for proper cell morphogenesis and polarized cell growth in Physcomitrella patens. The Plant Cell. 17:2327-2339.

Ishikawa, M., Murata, T., Sato,Y., Nishiyama, T., Hiwatashi, Y., Imai, A., Kimura, M., Sugimoto, N., Akita, A., Oguri, Y., 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., 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 cyclophilin 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., 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., 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., 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., Hartmann, E. (2004). Phosphoinositide-specific phospholipase C isinvolved in cytokinin and gravity responses inthe mossPhyscomitrella patens. Plant J 40:250–259.

Richter H., Lieberei R., Strnad M., Novák O., Gruz J., Rensing S.A., Schwartzenberg K. (2012). Polyphenol oxidases in *Physcomitrella*: functional PPO1 knockout modulates cytokinin-dependent developmentin the moss*Physcomitrella patens. JXP.* doi:10.1093/jxb/ers169.

Saavedra, L., Balbi, V., Lerche, J., Mikami, K., Heilmann, I., 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., 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., Bowman, J.L. (2013). KNOX2 genes regulate the haploid-to-dipliod morphological transition in land plants. Science. 339: 1067-1070.

Schipper, O., Schaefer, D., Reski R., 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., 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., 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., Bezanilla M. (2011). Myosin VIII Regulates Protonemal Patterning and Developmental Timing in the Moss Physcomitrella patens. Molecular Plant. 4 (5): 909-921.

Vidali, L., G.M. Burkart, R.C. Augustine, E. Kerdavid, E. Tüzel, M. Bezanilla. (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., Demura T. (2014). Contribution of NAC transcription factors to plant adaptation to land. Science. 343: 1505-1508.

CHAPTER III. Transcriptome Profiling of Moss Spermatogenesis

Author contributions: All experimental work presented in this chapter was performed by me. Experimental design was conceived by Jörg Becker and me. The phylostratigraphic analysis and the TAI estimation was performed by Marek Mutwil (Max Planck Institute of Molecular Plant Physiology).

Expression data is provided in the CD included in this Thesis, as well as the list of primers used. Supplementary Tables and Figures are included at the end of this Chapter

1. SUMMARY

Changes in the reproductive systems of plants contributed significantly to their success on land. Among those, substantial modifications in the male gametophyte occurred, evolving from the motile and flagellated sperm cells in Bryophytes to the sperm cells contained in the pollen grain observed in Angiosperms. In recent years, several studies have been conducted on male gametes in flowering plants, and genes important for their development and function have been identified. However, little was known about the genetic changes underlying the evolution of plant male gametes, mostly due to a lack of studies on plant species with free-swimming sperm. In this work we generated transcriptomic data from three developmental stages of male gametogenesis in the moss *Physcomitrella patens*, comprising the first data set of plant motile sperm. We detected highly dynamic processes of gene expression during spermatogenesis, result in a complex transcriptome with a striking increment of enriched and preferentially expressed transcripts. Based on gene expression patterns we identified two major functional groups with putative roles in gamete identity and function as well as embryo development. Consequently a temporal separation of transcription and translation during male gametogenesis is suggested. Moreover, the preferential expression of evolutionarily young genes in mature antherozoids was observed. The detection of those transcripts expressed at high levels may indicate an evolutionary mechanism to preferentially transcribe and select new genes in the male gametophyte, similarly to observations in animals and Angiosperms.

2. INTRODUCTION

During spermatogenesis, a series of cell divisions both in plants and in animals originate the population of male gametes that will contribute to the next generation. In animals the germ cell line is established early in embryogenesis and maintained throughout the entire life of the organisms by a subtle balance between differentiation, reprogramming and apoptosis (Zhao and Garbers, 2002). In contrast plants retain meristematic cells that differentiate late in development to generate the sex organs (Huala and Sussex, 1993), and that after a series of meiotic and mitotic divisions give rise to the functional male gametes. In angiosperms these divisions originate the sperm cells and the vegetative nucleus contained in the pollen grain (McCormick, 1993). Whereas in early land plants the motile sperm cells or antherozoids and the sterile surrounding cells will be formed as result of gametogenesis (Renzaglia and Garbary, 2001). Despite differences in germ line establishment and cell identity acquisition, common processes of chromatin remodeling and organelle structural modifications during spermatogenesis have been shown in eukaryotes (Balhorn, 2007; Ekici and Dane, 2004), and in all cases the final product of male gametogenesis represents a unique example of cell differentiation whose complexity we are starting to recognize.

For long, mature sperm cells were considered to be transcriptionally and translationally quiescent cells, with their only role in reproduction being the contribution of the paternal genome. However, increasing evidence of messenger RNA (mRNA) and small non-coding RNA (nc RNA) in mature male gametes of diverse species such as *Lilium longiflorum* (Blomstedt et al., 1996), *Arabidopsis thaliana* (Borges et al., 2008; Pina et al., 2005), *Zea mays* (Engel et al., 2003), *Homo*

sapiens (Krawetz et al., 2011; Ostermeier et al., 2002), *Mus musculus* (Grivna et al., 2006) and *Drosophila melanogaster* (Fischer et al., 2012) have strongly challenged this view. Moreover, observations of newly synthesized RNA in *Drosophila* post-meiotic cells (Vibranovski et al., 2010) and examples of absent or incomplete chromatin packaging in mature sperm of different plant species (Rejon et al., 1988; Renzaglia and Garbary, 2001), as well as the histone retention in mammals (Gatewood et al., 1987) have offered the possibility of active transcription taking place in the mature gametes. Yet, the indication of reduced translational capabilities of mature sperm (Gur and Breitbart, 2006; Johnson et al., 2011; Ostermeier et al., 2004) led to the view that there might be no functional role for the produced transcripts. In accordance with the diversity of transcripts observed in mature sperm a range of possible functional and non-functional roles have been attributed (Martins and Krawetz, 2005), including being leftovers from previous developmental stages, roles in imprinting and chromatin repacking and functions in fertilization and embryogenesis (reviewed in Boerke et al., 2007; Dadoune, 2009; Miller, 2014).

The requirement of paternal centromeres for zygote cell divisions (Palermo et al., 1997; Terada et al., 2004) and the role of the sperm derived Ca^{2+} channel *PLCz* in egg activation (Saunders et al., 2007, 2002) suggested that male gametes could have a wider contribution in early embryogenesis than previously considered. And with the detection of intact sperm mRNA in the zygote of mammals (Ostermeier et al., 2004) and *Drosophila* (Fischer et al., 2012) sperm transcript functions in fertilization and early embryo development have been suggested. To date, experimental confirmation of functional paternally delivered transcripts in zygote development are restricted to few examples in human such as the *PREGNANCY*

SPECIFIC B-1-GLYCOPROTEIN 1 (*PSG1*) and the *HUMAN LEUKOCYTE ANTIGEN-E* (*HLA-E*) transcripts important for zygote implantation (Avendaño et al., 2009), as well as the *FORKHEAD BOX G1B* (*FOXG1B*) with functions in embryo patterning (Ostermeier et al., 2004). In *Caenorhabditis elegans*, the general importance of sperm derived transcripts for embryo development has been determined by specifically radiolabel paternal RNAs (Stoeckius et al., 2014), whereas in plants the evidence of paternal contribution is limited to the *SHORT SUSPENSOR* (*SSP*), involved in the YODA-MAP kinase pathway to regulate embryo elongation (Bayer et al., 2009). Complementary to roles in early embryo development, the indications of poly(A) length reduction on specific sperm mRNAs (Gu et al., 1996; Sendler et al., 2013) and the observation of transcript intron retention in mouse spermatozoa (Soumillon et al., 2013) similar to stored transcripts during *Marsilea vestita* gametogenesis (Boothby et al., 2013), suggest the existence of mechanisms for translation regulation of stored transcripts. Although the shortening of the poly(A) tail can be linked to degradation, short 3'UTR in transcripts at early stages of embryogenesis (Ji et al., 2009) and during mammalian T cell activation (Sandberg et al., 2008) indicate a role in posttranscriptional regulation during rapid developmental processes. Overall, accumulating evidence suggests that transcript storage and delivery upon fertilization by sperm cells may be a more common phenomenon; but the extent to which sperm transcripts are poised for this function remains to be elucidated.

The intense selective pressures to which the male gamete are subjected, such as sperm competition and sexual conflict (Snook, 2005; Wedell et al., 2002), together with the permissive state of chromatin during reprogramming in spermatogenesis (review in

Sassone-Corsi, 2002) may shape genome expression on a different level. And although sperm retain great importance as transmitters of the genomic information to the next generation, the fact that they do not impose consequences to the organism that generated them together with the great numbers that a male can produce in most of the species allows in theory certain flexibility for evolutive selection (Calvel et al., 2010). In accordance numerous examples of evolutionary new genes showing testis-specific or testis-biased expression have been observed, such as the *PIPSL* in hominids generated by fusion between the *PHOSPHATIDYLINOSITOL- 4-PHOSPHATE 5-KINASE* (*PIP5K1A*) and the 26S proteasome subunit (*PSMD4*) transcripts (Babushok et al., 2007), the *de novo* originated *JINGWEI* gene in *Drosophila* (Long and Langley, 1993) and the long non-coding *Pidi* originated by transcriptional activation of a specific region in mouse post-meiotic testis cells (Heinen et al., 2009). Based on those studies, it has been postulated that the male organs such as testis in animals (Kaessmann, 2010) and pollen in angiosperms (Cui et al., 2014; Wu et al., 2014) may function as source of evolutionary innovation where new genes are preferentially expressed.

In this study we used the moss *Physcomitrella patens*, a bryophyte with motile sperm cells (Cove, 2005), as model organism to assess the transcriptomic complexity of male gametes in an extant early land plant. Differentiation of meristematic cells gives rise to the sexual organs: the archegonia (female structure) and the antheridia (male structure), where the sperm cells (or antherozoids) are produced. Retention of flagellated cells parallels the generation of a motile apparatus during animal spermatogenesis and a similar process of cytoplasm reduction in mature sperm cells has been observed in mosses (Renzaglia and Garbary, 2001). Due to its evolutionary

position and biological characteristics (e.g. dominant haploid life cycle), the availability of the complete genome sequence (Rensing et al., 2008) and the possibility to perform targeted mutagenesis, *P. patens* represents a very suitable model for genetic evolutionary studies (Cove and Knight, 1993). Through a comparative transcriptome analysis using three stages of male gametogenesis in *P. patens* we were able to establish gene expression dynamics. The complexity of transcripts identified as enriched and preferentially expressed during spermatogenesis suggest very diverse functions, including important processes for gamete identity and fusion and possible roles in early embryo development. Moreover, our data indicate that it is in the male organs were new genes are preferentially expressed and analysis of methylation pathways points to a relaxed epigenetic state favoring this process. With the study of antherozoids we had been able to assess genetic evolutionary changes during land plant evolution, but also to establish general conservation among eukaryotes showing a similar tendency to express evolutionary new genes.

3. RESULTS

3.1 Generation of *P. patens* male gametogenesis transcriptome

To assess the developmental transcriptome of *Physcomitrella patens* male gametogenesis we isolated samples corresponding to three developmental stages, including antheridia collected at an early stage of development (enriched in immature antheridia), antheridia collected at a later stage (enriched in mature antheridia) and released antherozoids from the two distinct *Gransden* and *Villersexel* strains (see Material and Methods). The developmental time points suitable for our study were defined through daily observations of the organs after induction by cold. Clusters of antheridia were detected from day 6 onwards, in accordance with Landberg et al., (2013). At day 15 fully mature gametangia were observed (Figure 1): at this point, antheridia located at the center of the cluster turned yellow and the tip of the organ swelled until it broke and approximately 100 antherozoids were released as clusters of cells embedded in a matrix. Although releases could occur from day 14 after induction onwards for several days, it was at day 15 when more events were observed in both strains studied. We detected differences between strains in the number of releasing events and in the number of days in which releases occurred (Table 1), and consistent with published data regarding fertility rates (O'Donoghue et al., 2013; Perroud et al., 2011) the *Villersexel* strain showed higher numbers in both parameters with more actively swimming sperm cells.

On days 15 to 17 after induction antherozoids from the two strains were manually collected immediately after release using

micromanipulators. We generated four independent pooled samples of approximately 300 antherozoid clusters from both *Gransden* and

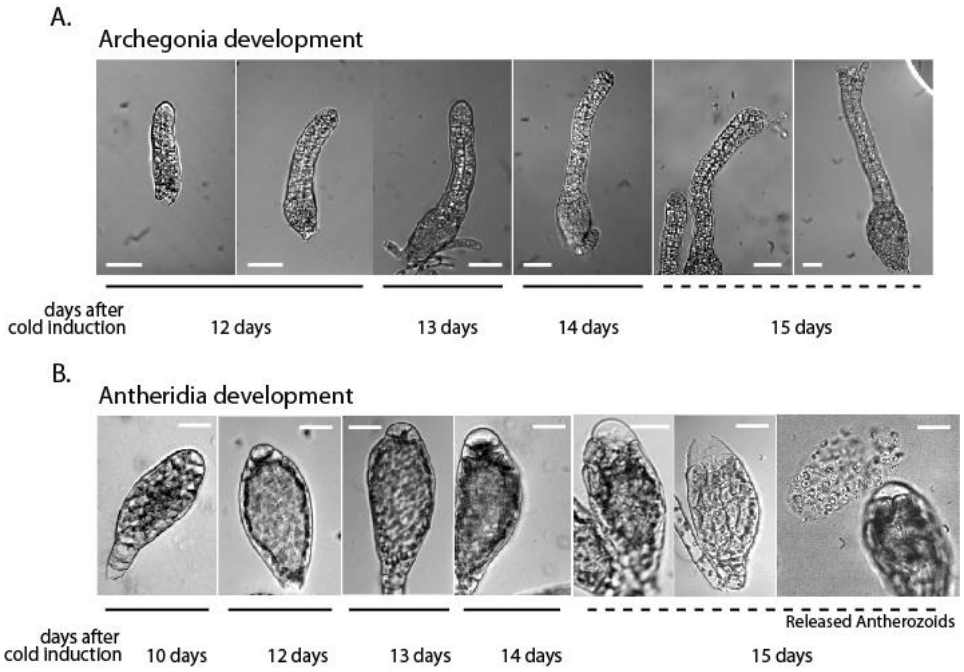


Figure 1. Characterization of *Physcomitrella patens* gametangia development. The developmental stages of *P. patens* gametangia under our laboratory conditions were characterized. Induction of gametangia was conducted by exposing 3-4 week old gametophores to short day conditions. Gametophores were manually dissected under the stereoscope to obtain the gametangia, which was then observed in detail under the microscope. Observations were conducted every day after the day of 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, including the moment of their opening. (B) In the same way the antheridia development was followed. For this organ the most dramatic changes, besides growth, includes the change from green to yellow and the formation of a cap. This cap brakes and the antherozoids are released. Releases occur from day 14 and continue until day 17, being day 15 when more releases are observed. Both antheridia and archegonia form clusters of organs at different developmental stages and data represents the moment in which each stage is visible for the first time.

Villersexel strains, giving total RNA yields of 6.25 -10 ng each. In accordance with observed profiles in high quality human sperm samples (Ostermeier et al., 2002) no ribosomal RNA peaks were detected in Bioanalyzer profiles (Supplemental figure 1). In contrast to the rest of the samples, none of the DNAse treatments performed were successful, affecting both cDNA quantity and quality parameters as well as gene expression tested by qPCR. Since no genomic DNA traces were observed in the RNA profiles and preferential amplification of RNA by the Ovation WT system (NuGen Technologies, Inc) with no co-amplification of contaminating genomic DNA has been reported (Vermeulen et al., 2009), untreated samples were used for further steps. For the antheridia samples (derived only from the *Gransden* strain) we preferentially collected the central antheridium of each antheridia cluster at days 10 and 14 after induction, corresponding to early and late stages of organ development. Three biological replicates for each antheridia sample were generated and the quality and quantity of the RNA obtained was tested both by Bioanalyzer profile and by qRT-PCR using housekeeping genes and tissue specific probes (Supplemental figure 2).

The three stages of *P. patens* male gametogenesis isolated were analyzed on custom designed NimbleGene microarrays, covering 91.1% of *P. patens* genome version 1.6. Making use of data previously generated (see Chapter II) we were able to compare them with the rest of the tissues that represent most stages of the life cycle of the moss. In this way we were able to assess the sequential gene expression necessary for male gamete development and function.

DAYS AFTER COLD INDUCTION (16h dark/8h light, 17°C)	*Villersexel*		*Gransden*	
	General observations	Male gametangia	General observations	Male gametangia
11	No gametangia fully developed		No gametangia fully developed	
13	Some gametangia developed	No antherozoids	Some gametangia developed	No antherozoids
14	Gametangia developed	No antherozoids	Gametangia developed	No antherozoids
15	Mature archegonia	Antherozoid release	Mature archegonia	Antherozoid release
16	Mature archegonia	Antherozoid release	Mature archegonia	Antherozoid release
17	Fertilization	Antherozoid release	Fertilized archegonia (brown neck)	Few antherozoid release events
18	Fertilization	Antherozoid release	Archegonia abortion events	Few antherozoid release events
19	Fertilized archegonia (brown neck)	Antherozoid release	Archegonia abortion events	Very few antherozoid release events
20	Fertilized archegonia (brown neck)	Antherozoid release	Some fertilized archegonia	No antherozoids
21	Early stage of sporophyte observed		Early stage of sporophyte observed	

Table 1. Characterization of antherozoid release in *P. patens* Gransden and Villersexel strains. The development of the gametangia was induced by exposing 4 week old gametophores to cold day conditions (16h dark/8h light, 17ºC). Observations of antheridia and archegonia were performed daily to assess the developmental process of the organs. In particular, the maturation of antheridia and the release of antherozoids were followed carefully for both strains studied. Fully mature gametangia were observed at day 15 after cold induction, when antherozoid release occurred. For the *Villersexel* strain, more releases were observed per preparation (containing approximately 40 antheridia clusters) and release events continued to be numerous for several days. Maturity of the archegonia was evidenced by the opening of the channel that leads to the egg cell. Few days after sperm release, a brown coloration was observed at the neck of the archegonia. Many archegonia acquired darker neck coloration and ceased further development (abortion events).

3.2 Antherozoids show a very distinct transcriptomic profile

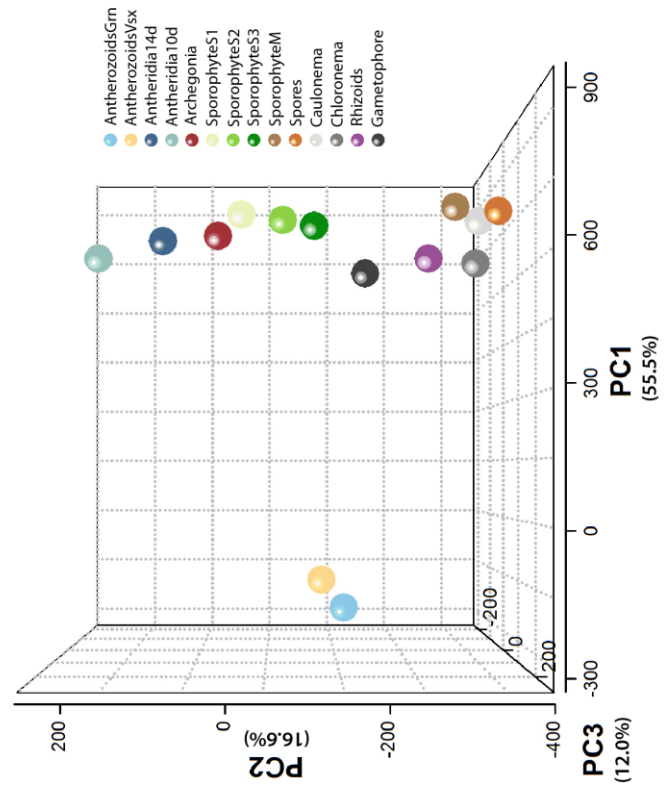
The transcriptome relationships among tissue samples were approached by Hierarchical Clustering (HC) and Principal Component Analysis (PCA). In the HC the sample replicates grouped by tissue identity (Figure 2A) separating both antheridia stages as discrete units. In accordance, a Pearson's correlation of r>0.929 among sample replicates of each stage was observed (Table 2). The HC analysis grouped the antheridia samples closer to the archegonia and sporophyte, while antherozoids were group far from all other tissues establishing a separate clade or simplicifolious. The antherozoid replicates clustered indifferently from the strain they belong to, and

Antherozoid Sample replicates	Pearson's Correlation Coefficient							
	Gransden a	**Gransden b**	**Gransden c**	**Gransden d**	Villersexel a	Villersexel b	Villersexel c	Villersexel d
Gransden a	**1.000**							
Gransden b	0.810	**1.000**						
Gransden c	0.882	0.813	**1.000**					
Gransden d	0.864	0.842	0.869	**1.000**				
Villersexel a	0.806	0.760	0.807	0.813	**1,000**			
Villersexel b	0.836	0.837	0.841	0.860	0.777	**1,000**		
Villersexel c	0.871	0.836	0.878	0.878	0.828	0.861	**1.000**	
Villersexel d	0.777	0.752	0.780	0.788	0.755	0.771	0.790	**1.000**

Antheridia Sample replicates	Pearson's Correlation Coefficient					
	Antheridia14d a	**Antheridia14d b**	**Antheridia14d c**	Antheridia10d a	Antheridia10d b	Antheridia10d c
Antheridia14d a	**1.000**					
Antheridia14d b	0.924	**1.000**				
Antheridia14d c	0.951	0.954	**1.000**			
Antheridia10d a	0.924	0.860	0.899	**1.000**		
Antheridia10d b	0.897	0.827	0.873	0.953	**1.000**	
Antheridia10d c	0.880	0.812	0.856	0.937	0.929	**1.000**

Table 2. Pearson's correlation coefficient for sample replicates of antheridia and antherozoids. In the upper distance matrix table correlation coefficients for all antherozoid samples are shown. Each sample replicate contained approximately 20,000 pooled cells from independent collections. The Pearson's average correlation for *Gransden* and *Villersexel* replicates is r=0.862 (0.832-0.895) and r=0.817 (0.798-0.875), respectively. Between strains a correlation value of r=0.827 (0.77-0.89) suggests a high degree of variation among replicates regardless of the strain. The bottom table shows the correlation among replicates of the antheridia samples, were values >0.92 can be observed among samples from the same developmental stage.

the comparison of all replicates yielded a Pearson's correlation average of r=0.827 (0.77-0.89) (Table 2). In the PCA analysis the proximity of samples is related to global gene expression similarities and overall variation can be ranked according to the Principal Components (PCs). In accordance with the tissue identity separation observed in the HC, the PCA shows antherozoids clearly separated from antheridia samples by Principal Component (PC) 1, as well as from the rest of the tissues, suggesting a very distinct transcriptome identity (Figure 2B). When we plotted samples excluding the antherozoids (Supplemental figure 3) the antheridia and archegonia are displaced to the right separated by PC1 and PC3 from the other tissues. In the same trend the immature antheridium shows a separation from the non-reproductive tissues by PC1, situated between the mature gametes and the stage 1 sporophyte. Overall, these results revealed a clear and distinct transcriptomic identity for the reproductive tissues analyzed. Intriguingly, the antherozoids showed a very particular transcriptome separated even from the mature antheridium. One possible explanation is that the adjacent sterile cells included in the antheridia, could contribute with their RNA to the transcriptomic profile. And though we cannot discard problems with the samples due to the low amount of RNA obtained, the consistency of all independently collected replicates suggests actual differences.


AntherozoidsGrn
AntherozoidsVsx
Antheridia14d
Antheridia10d
Archegonia
SporophyteS1
SporophyteS2
SporophyteS3
SporophyteM
Spores
Caulonema
Chloronema
Rhizoids
Gametophore
PC2
(16.6%)
PC1
(55.5%)
PC3
(12.0%)

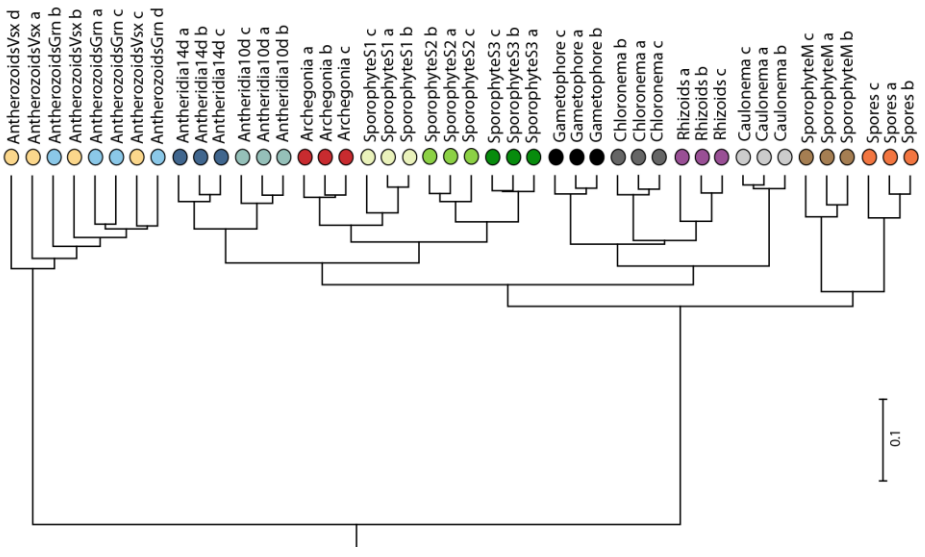
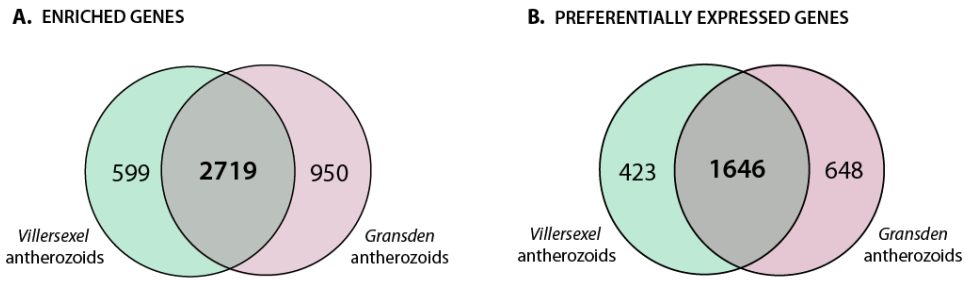

AntherozoidsVsx d
AntherozoidsVsx a
AntherozoidsGrn b
AntherozoidsVsx b
AntherozoidsGrn a
AntherozoidsGrn c
AntherozoidsVsx c
AntherozoidsGrn d
Antheridia14d a
Antheridia14d b
Antheridia14d c
Antheridia10d c
Antheridia10d a
Antheridia10d b
Archegonia a
Archegonia b
Archegonia c
SporophyteS1 c
SporophyteS1 a
SporophyteS1 b
SporophyteS2 b
SporophyteS2 a
SporophyteS2 c
SporophyteS3 c
SporophyteS3 b
SporophyteS3 a
Gametophore c
Gametophore a
Gametophore b
Chloronema b
Chloronema a
Chloronema c
Rhizoids a
Rhizoids b
Rhizoids c
Caulonema c
Caulonema a
Caulonema b
SporophyteM c
SporophyteM a
SporophyteM b
Spores c
Spores a
Spores b
0.1

Figure 2. Hierarchical clustering and Principal Component Analysis of *Physcomitrella patens* tissues.
A. In the Hierarchical Clustering (HC) dendogram the replicates of the analyzed tissues group closely together as discrete developmental units. Antheridia for the two stages are well separated, forming a closer clade to the archegonia and the early stages of the sporophyte. In the case of the antherozoids replicates group indifferently from the strain they belong, establishing a separated clade from the rest of the analyzed tissues. **B.** Similarly, the Principal Component Analysis displays the antherozoid samples separated far from the rest of the tissues by the first Principal Component (PC), suggesting a very different transcriptome. The presence of the antherozoids constricts the rest of the tissues. In PC 2 both antheridia stages, but in particular the mature stage, are closer to the archegonia.

3.3 The transcriptomes of *Gransden* and *Villersexel* antherozoids show little differences

The inability to separate the replicates of sperm cell samples from *Gransden* and *Villersexel* strains by hierarchical clustering was intriguing. Observations from several laboratories regarding low sporophyte production indicate a loss of sexual vigor in *Gransden* (O'Donoghue et al., 2013; Perroud et al., 2011). On the contrary, the more recently isolated strain *Villersexel* shows very high fertilization rates (Perroud et al., 2011). The observations in our laboratory confirm those differences both in fertilization rate and antherozoid activity (Table 1). And although we anticipated that dissimilarities on sexual vigor could be explained by differences in gene expression, the analysis by hierarchical clustering as well as the correlation matrix shows the same level of variation among replicates as between strains. Furthermore, pair-wise comparisons between strains were hampered by their similarity. In our analysis we could not determine a cut-off value that would result in a False Discovery Rate (FDR) below 10%, and no statistically significant differences in gene expression were identified.

To validate gene expression consistency of antherozoid samples and sustain our findings on a more robust base, we estimated the similarities of enriched and preferentially expressed genes in the two studied strains when compared with the rest of the tissues. As a prerequisite for differential expression analysis, the signal threshold for presence and absence calls of all *Physcomitrella* tissue samples was experimentally estimated by qRT-PCR (Supplemental figure 4).



C.

	Villersexel	*Gransden*
Total enriched genes	3318	3669
% of overlap	**81.94**	**74.1**
Total preferentially exp.genes	2069	2294
% of overlap	**79.55**	**71.75**

Figure 3. Common enriched and preferentially expressed genes in *Gransden* and *Villersexel* antherozoids.
A. Venn diagram showing number of specific and common enriched genes in antherozoids of *Gransden* and *Villersexel*. **B.** Venn diagram showing number of specific and common preferentially expressed genes in antherozoids of *Gransden* and *Villersexel.* **C.** The total number of enriched and preferentially expressed genes in each strain studied is shown, as well as the percentage of overlap among them. In all cases, common sequences represent more than 70%.

In the differential expression analysis independently performed for each strain, we observed similar numbers of enriched and preferentially expressed genes (Figure 3A) with a high percentage of overlap. From the total number of enriched genes in *Gransden* (3669) and *Villersexel* (3318) 2719 were common, representing more than

70% overlap (Figure 3B). And in the case of the preferentially expressed genes, 2294 in *Gransden* and 2069 in *Villersexel*, 1646 were common (Figure 3C).

When we tried to analyze the gene identity of those 2719 common enriched genes we were faced with lack of annotation in 64% of them. And despite efforts conducted to find similar sequences by Basic Local Alignment Search Tool (BLAST) performed against non-plant databases (see methods), no further improvement was achieved. This percentage of unannotated sequences is higher than the one observed for all the sequences represented in the array for which 42% miss annotation. The preferential lack of annotation for sperm cell enriched transcripts may be explained by their unique transcriptomic profile, as well as by the absence of transcriptomic data that could be used to curate the genome annotation of *P. patens*. Importantly, in the common enriched list generated we identified a *DUO1* ortholog (Pp1s16_281V6.1) (Supplemental figure 5), which in *A. thaliana* has been recognized as a master regulator of male gamete differentiation expressed during pollen development (Borg et al., 2011; Durbarry et al., 2005).

3.4 Extensive reprogramming during male gametogenesis

To assess the transcriptional changes occurring during male gametogenesis in *P. patens*, the samples corresponding to the three male gamete stages were individually compared with the rest of the tissues that comprise our transcriptome atlas (Chapter II). By excluding comparisons among each other, the enriched genes for all stages of male gametogenesis with respect to other tissues were identified. For the antherozoids independent comparisons for each strain were performed and only the common elements in both were

included in the analysis. In addition, the progression of sporophyte development starting from the archegonia was assessed in a similar way avoiding comparisons among the tissues comprising this particular developmental set. *Villersexel* antherozoid data were not used in these particular pair-wise comparisons.

Total numbers of enriched and preferentially expressed genes identified for each stage are shown in Figure 4A, taking into account that due to problems with gene annotation the microarray contains duplicated probe sets for some genes, meaning that in those cases more than one probe set is measuring the expression of the same protein encoding gene. Indicating complexity gain during developmental progression in the two cases analyzed (male gametogenesis and sporophyte development) the number of enriched and preferentially expressed genes increased during each time-course. However, the overall numbers observed during male gametogenesis were dramatically higher than those observed for the sporophyte development. At immature stage, antheridia showed 495 enriched genes, increasing to 1231 in mature antheridia and 2719 in antherozoids (number of genes corresponding to the overlap between strains). The same trend was observed for the preferentially expressed genes. Interestingly, the immature stage of antheridia shared 91.1% of the enriched genes with the other two stages. This percentage dropped to less than half in the mature antheridium (41.2%), and dramatically decreased in the antherozoids (3.5%) (Supplemental figure 6). This trend may reflect the probable transcriptomic changes occurring during sperm maturation, as well as differences in the contribution from adjacent sterile cells.

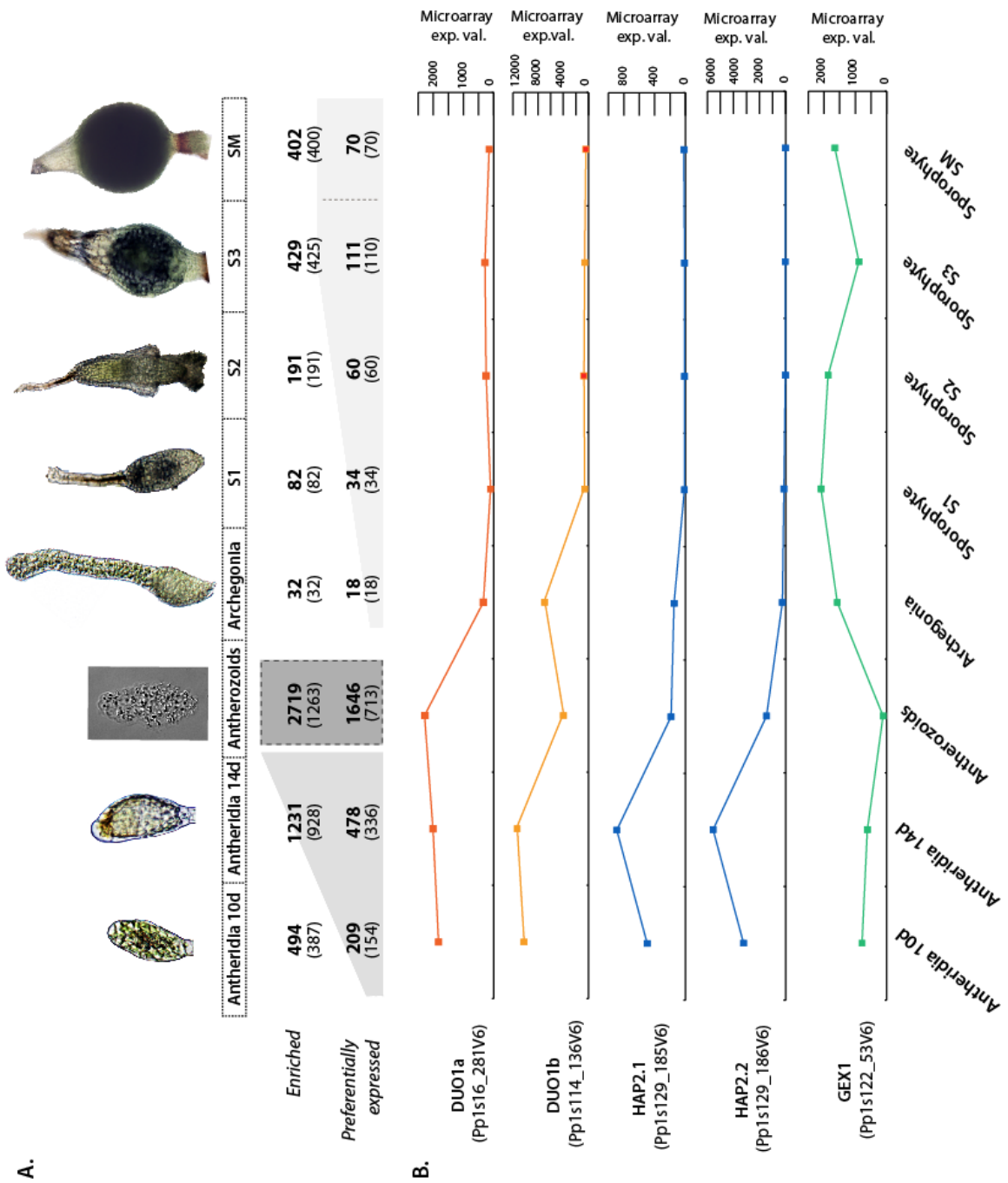

A.
Antheridia 10d
Antheridia 14d
Antherozoids
Archegonia
S1
S2
S3
SM
Enriched
Preferentially expressed
494 (387)
1231 (928)
2719 (1263)
32 (32)
82 (82)
191 (191)
429 (425)
402 (400)
209 (154)
478 (336)
1646 (713)
18 (18)
34 (34)
60 (60)
111 (110)
70 (70)
B.
DUO1a (Pp1s16_281V6)
DUO1b (Pp1s114_136V6)
HAP2.1 (Pp1s129_185V6)
HAP2.2 (Pp1s129_186V6)
GEX1 (Pp1s122_53V6)
Microarray exp. val.
Antheridia 10d
Antheridia 14d
Antherozoids
Archegonia
Sporophyte S1
Sporophyte S2
Sporophyte S3
Sporophyte SM

Figure 4. Gene expression changes during spermatogenesis and sporophyte development.
The transcriptional changes during male gametogenesis were assessed in comparison with sporophyte development. **A.** Enriched and preferentially expressed genes during male gametogenesis were identified by statistical comparisons with the rest of tissues for which we generated transcriptomic data. For antherozoids only common elements between *Gransden* and *Villersexel* strains were considered. Bold numbers indicate total number of probe sets identified as enriched or preferentially expressed, while unique protein encoding genes are depicted between brackets. As development progressed, the number of enriched and preferentially expressed genes increased dramatically. In the same way, changes of gene expression during sporophyte development were assessed. Archegonia, pre-meiotic (S1 and S2), meiotic (S3) and post-meiotic (SM) sporophyte stages were included. The numbers of enriched and preferentially expressed genes also increased during development. However, total numbers are much lower than those observed for male gametogenesis. **B.** Expression patterns of particular genes potentially important for male gametogenesis or embryo development are shown. *DUO1a*, *DUO1b*, *HAP2.1*, *HAP2.2* and *GEX1* homologs in *P. patens* were determined by sequence homology. In the case of *DUO1*, the phylogenetic analysis indicates them as orthologs of *AtDUO1* (Supplemental figure 5).

The expression patterns of genes identified as important for male gametes displayed interesting trends (Figure 4B). The two *DUO1* orthologs, *PpDUO1a* (Pp1s16_281V6.1) and *PpDUO1b* (Pp1s114_136V6.1), were expressed in the male gametophyte. But while *DUO1a* was detected at similar levels in the three stages of male gametogenesis, DUO1b showed very high expression in the antheridia and a decreased expression in antherozoids. Moreover, *DUO1a* was called present only in the male organs whereas *DUO1b* was also present in the archegonia. In the case of the *HAP2/GCS1* homolog in *P. patens* the protein is annotated by two adjacent transcript models (Pp1s129_185V6.1 and Pp1s129_186V6.1 named *HAP2.1* and *HAP2.2*, respectively) represented as separate probe sets on the microarray. Although the similarities in the expression pattern exhibited by the two probe sets suggest only one transcript, the existence of an alternative transcript cannot be ruled out. *HAP2* is

a highly conserved sperm cell protein necessary for gamete fusion (von Besser et al., 2006; Liu et al., 2008) and accordingly, in *P. patens* its expression seems to be restricted to the male gametophyte. At the immature stage of antheridia the transcript was detected at high levels, increasing even further at maturity and decreasing in the antherozoids (Figure 4B). In the case of *HAP2.1* the expression of the transcript dropped below the presence call threshold in antherozoids, while *HAP2.2* remained as called present. Finally, the expression pattern of the *PpGEX1* homolog (Pp1s122_53V6.1) is shown in figure 4B. *GEX1* is a conserved plasma membrane protein, expressed in the sperm and egg cell membrane as well as during early embryo development in *Arabidopsis* (Alandete-Saez et al., 2011; Engel et al., 2003). In *P. patens* the transcript showed an enriched expression in the archegonia and during pre-meiotic stages (sporophyte stage 1 and 2), while being expressed at low levels in the antheridia and called absent in the antherozoids.

The expression of *DUO1b* and *HAP2* transcripts is intriguing. In both cases evidence of diminished expression in the antherozoids is observed. However, due to their reported activity and importance for mature sperm cells their absence at the final stage of male gametogenesis in *P. patens* would be unexpected. *GEX1* on the other hand has been shown to have a role in embryo progression in *A. thaliana* (Alandete-Saez et al., 2011), a compatible situation with the expression pattern in the pre-meiotic sporophyte of *P. patens*.

3.5 Male gametogenesis gene clustering reveals two major trends

To further investigate the function of the diverse transcripts expressed during male gametogenesis average linkage hierarchical clustering of the data was performed (Figure 5A). Based on gene expression pattern eight clusters with overall average homogeneity of 0.994 were generated (Figure 5B). The lists of annotated genes included on each cluster were used for Gene Onthology (GO) Enrichment Analysis using DAVID software (Huang et al., 2009a, 2009b). Although this analysis can be hampered due to the high percentage of unnanotated sequences for antherozoids (see above), it offers some insights on the transcript function by linking the available information from *A. thaliana* homologs.

Cluster I contained 2203 transcripts characterized by increased expression in the antherozoids compared with both antheridia stages. This comprises the transcripts preferentially expressed at the later gametogenesis stage. Interestingly, the GO enrichment analysis indicated processes related to transcript production and processing, such as nucleotide binding, DNA directed RNA polymerase complex and mRNA catabolism debating the long-maintained idea of transcriptional silencing in male mature gametes (Kierszenbaum and Tres, 1975). Furthermore, the presence of post-embryonic and foliar development suggests an even more intriguing function for some of those transcripts. The existence of important transcripts for embryo development delivered by the sperm cells upon fertilization has been shown both for plants and animals (Bayer et al., 2009; Ostermeier et al., 2004) and the presence of the highly conserved *PUMILIO* (*PUM*) (Pp1s96_109V6.1, Pp1s96_110V6.1 and Pp1s35_161V6.1) homolog genes involved in mRNA translational repression important for germ cell identity and embryo development (Abbasi et al., 2011; Crittenden

et al., 2002; Moore et al., 2003; Murata and Wharton, 1995) supports this idea. Also, the *CLAVATA3/ESR-related 10* (*CLE10*) (Pp1s292_57V6.1) homolog important for multicellular development (Sharma et al., 2003) and the *AINTEGUMENTA-LIKE 5* (*AIL5*) (Pp1s66_255V6.1) homolog expressed in actively dividing tissues in *A. thaliana* (Klucher et al., 1996; Mizukami and Fischer, 2000) are candidate transcripts for delivery.

Cluster II (including 639 transcripts) displayed the opposite trend with transcripts showing increased expression from the immature to the mature antheridia stage, followed by a drop in expression below presence threshold in the antherozoids. In this case the GO enrichment showed categories that would be expected for flagella based movement activity, such as microtubule based movement, protein assembly and transport and energy production. *HAP2.1* and *HAP2.2* transcripts were also found in this cluster. Cluster III contained 336 transcripts with enriched expression in the mature antheridium. The GO enrichment for this group included mitotic cell cycle; exemplified by the *CELL DIVISION CYCLE 48* (*CDC48*) and the *WUSCHEL RELATED HOMEOBOX 14* (*WOX14*) homolog genes (Pp1s44_265V6.1 and Pp1s217_9V6.1, respectively) involved in cell cycle control in *Arabidopsis thaliana*. Consistent with the antheridia's function as a reproductive organ we also found reproductive developmental process, chromosome organization and pollen tube development enriched categories, represented by the *TONSOKU* (*TSK*) / *BRUSHY 1* (*BRU1*) homolog gene (Pp1s1184V6.1) involved in meristem organization and chromatin modifications and the *HISTONE H1.2* (Pp1s 25_2V6.1).

In Cluster IV (with 235 transcripts), a diminished expression is observed in the mature antheridium with respect to the other two

stages. The GO enrichment revealed also post-embryonic development together with macromolecular complex assembly, secondary metabolism and phosphorylation. Since the increment of expression in the antherozoids with respect to the previous stage resembled the profile of Cluster I it is tempting to speculate that a sub list of transcripts included here could also be delivered upon fertilization for functions on later developmental stages. However, the presence of genes related to metabolism and vesicle trafficking propose other functions as well.

The Cluster V (185 transcripts included) represent transcripts which expression increased as the development of male gametes progressed. This included duo1a (Pp1s16_281V6.1), a *MYB-like* transcription factor similar to *DUO1* (Pp1s58_187V6.1) and a *SU(VAR)3-9 homolog 2* (*SUVH2*) (Pp1s44_197V6.1) possibly involved in a reprogramming process necessary for gamete identity. Although most of the GO enriched categories showed no statistical significance (*n.s.*), it was interesting to notice the presence of hormone-mediated signaling category.

Finally, Clusters VI, VII and VIII were the smallest with 155, 111 and 27 transcripts, respectively. Due to the short gene lists, the analysis of GO enrichment was not statistical significant. Cluster VI contained transcripts that showed maintained levels in mature antheridium and antherozoids, with increased expression in respect to the immature stage. Within the genes that appeared in this cluster, *UNUSUAL FLORAL ORGANS* (*UFO*) (Pp1s1_769V6.1) and *HIGH PLOIDY 2* (*HPY2*) (Pp1s405_28V6.1) homolog genes may account for relevant functions in meristem maintenance. Cluster VII was similar to Cluster II, but in this case the higher expression occurred in immature antheridia.

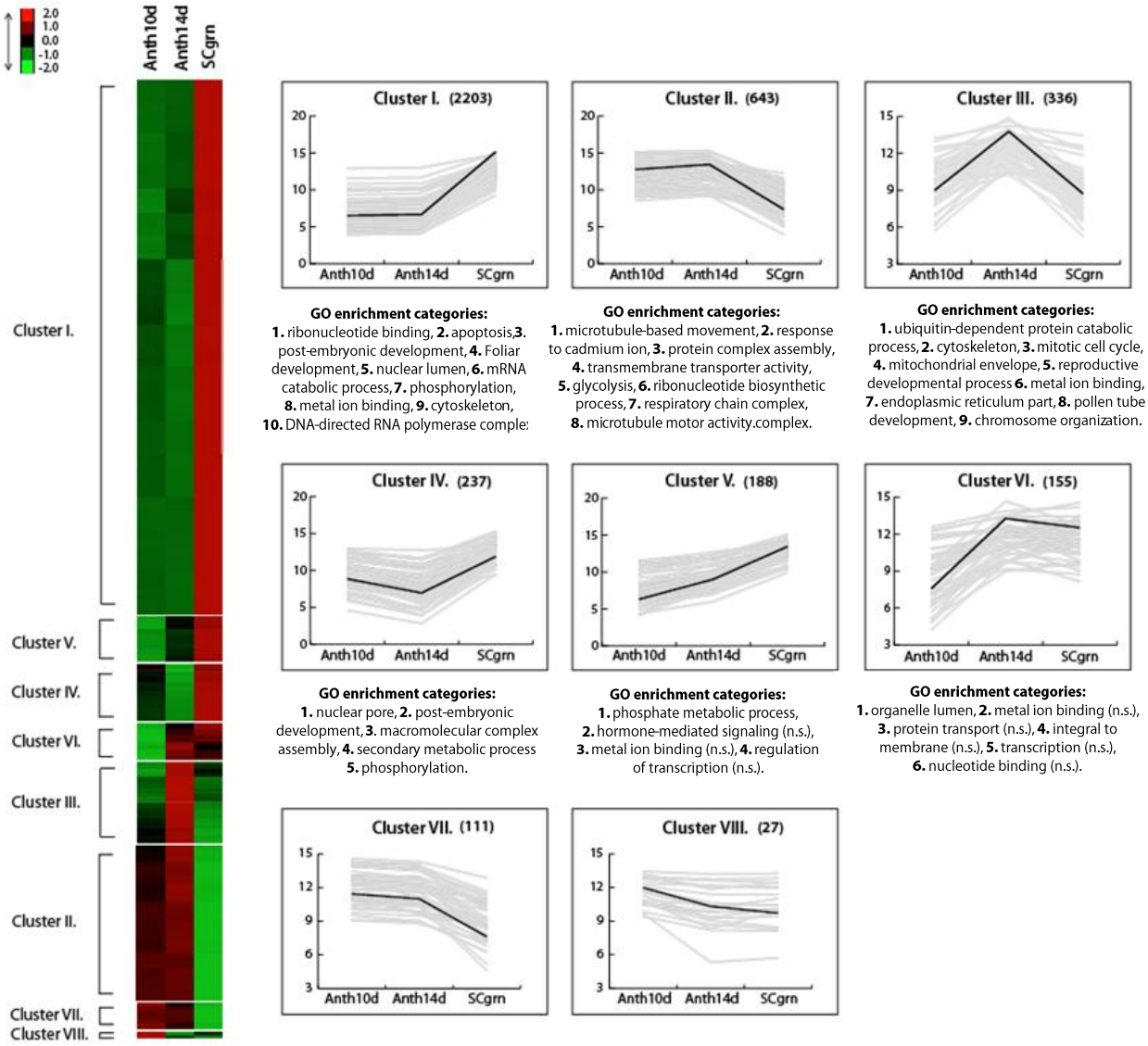


Figure 5. Hierarchical clustering of *P. patens* male gametogenesis enriched genes.

Male gametogenesis enriched genes identified were grouped by average linkage hierarchical clustering (HC). **A.** The HC dendogram shows the eight clusters grouped according to expression in immature antheridia (Anth10d), mature antheridia (Anth14d) and *Gransden* antherozoids (SCgrn). **B.** Average expression pattern of each cluster. In the graph, the number of transcripts in this cluster is shown in brackets. Below the graph GO enriched categories are listed. Not statistically significant terms (n.s.) are also included for clusters V and IV. Analysis of clusters VII and VIII did not retrieve results due to the small number of transcripts.

Interestingly, the homolog gene of *MALE GAMETOPHYTE DEFECTIVE 2* (*MGP2*) (Pp1s209_60V6.1), identified to be required for pollen germination in *A. thaliana,* was present. Finally, cluster VIII represented the smallest cluster, containing the genes enriched in mature antheridia.

3.6 Transcription Factor enrichment in male gametogenesis clustered genes

A further characterization of transcriptional factors (TF) present on each cluster was performed (Figure 6) in order to add evidence to the possible role of transcripts contained. We focused our analysis on clusters I and II due to their interesting profiles. Annotation related to TF families in *P. patens* was downloaded from cosmoss (www.cosmoss.org, last accessed on 20th of March, 2015), and confirmed by Pfam domain search (pfam.xfam.org).

Cluster I contained the highest number of TFs (44), with some families of TFs represented by several members. Remarkably, two *TEOSINTE BRANCHED 1/CYCLOIDEA/ PROLIFERATING CELL FACTORS 1* (*TCP*) *Class I* ortholog genes (Pp1s446_21V6.1 and Pp1s348_6V6.1) (Chapter II, Supplemental figure 5) known to act on cell proliferation of developing tissues (Martín-Trillo and Cubas, 2010) were found in this cluster. As well as members of the *JUMONJI* family described to be important for reprogramming and developmental processes in eukaryotes (Pp1s235_45V6.1 and Pp1s235_2V6.1) (Chen et al., 2011; Takeuchi et al., 2006) and of the *LEAFY* (*LFY*) TF family essential for floral meristem identity (Pp1s258_46V6.1) (Schultz and Haughn, 1991). Among the 11 TFs present in Cluster II we detected one gene of the *ZINC FINGER-HOMEODOMAIN* (*ZF-*

CLUSTER I.
CLUSTER II.
CLUSTER III.
CLUSTER IV.
CLUSTER V.
CLUSTER VI.
CLUSTER VII.
TF Family
probe_set
Anth 10d
Anth 14d
SCgrn
AP2 Pp1s66_255V6
B3 Pp1s57_383V6
B3 Pp1s179_63V6
bZIP Pp1s20_118V6
bZIP Pp1s49_161V6
bZIP Pp1s71_364V6
bZIP Pp1s87_156V6
bZIP Pp1s355_22V6
C2H2 Pp1s577_1V6
C2H2 Pp1s13_348V6
C3H Pp1s339_18V6
C3H Pp1s36_324V6
C3H Pp1s6_315V6
C3H Pp1s102_155V6
C3H Pp1s258_57V6
C3H Pp1s297_45V6
C3H Pp1s177_93V6
Dof Pp1s200_73V6
ERF Pp1s126_30V6
ERF Pp1s45_159V6
ERF Pp1s117_19V6
G2-like Pp1s29_23V6
G2-like Pp1s104_175V6
HD-ZIP Pp1s31_269V6
HD-ZIP Pp1s59_146V6
JUMONJI Pp1s235_45V6
JUMONJI Pp1s235_2V6
LFY Pp1s258_46V6
MYB_related Pp1s128_8V6
MYB_related Pp1s132_120V6
MYB_related Pp1s66_90V6
MYB_related Pp1s471_2V6
MYB_related Pp1s966_1V6
MYB_related Pp1s226_2V6
MYB_related Pp1s208_2V6
MYB_related Pp1s176_5V6
MYB_related Pp1s426_45V6
Nin-like Pp1s109_79V6
TCP class I Pp1s446_21V6
TCP class I Pp1s348_6V6
Trihelix Pp1s132_67V6
Trihelix Pp1s60_28V6
Trihelix Pp1s60_31V6
WRKY Pp1s2_539V6
AP2 Pp1s72_220V6
bHLH Pp1s89_172V6
LBD Pp1s7_261V6
MYB Pp1s65_233V6
MYB_related Pp1s87_134V6
MYB_related Pp1s80_153V6
MYB_related Pp1s98_6V6
Nin-like Pp1s9_112V6
Nin-like Pp1s250_38V6
Trihelix Pp1s148_18V6
WOX Pp1s29_82V6
ZF-HD Pp1s31_206V6
B3 Pp1s344_29V6
bZIP Pp1s25_341V6
ERF Pp1s340_16V6
LBD Pp1s69_130V6
MYB Pp1s206_27V6
Trihelix Pp1s44_57V6
WOX Pp1s217_9V6
ERF Pp1s115_154V6
G2-like Pp1s73_109V6
HD-ZIP Pp1s42_217V6
HD-ZIP Pp1s209_10V6
M-type Pp1s68_293V6
Trihelix Pp1s5_14V6
ERF Pp1s11_198V6
MYB (DUO1a) Pp1s16_281V6
Dof Pp1s341_82V6
GRAS Pp1s280_61V6
M-type Pp1s55_284V6
PLATZ Pp1s35_284V6
C3H Pp1s66_196V6
37400.0
7900.0
450.0
0.0

Figure 6. Heatmap of transcription factors present in the male gametogenesis gene clusters.
The presence and distribution of Transcription Factor (TF) gene families was determined for the male gametogenesis gene clusters. The expression level of TF's in immature antheridia (Anth 10d), mature antheridia (Anth 14d) and *Gransden* antherozoids (SCgrn) is displayed as a heatmap, where darker blue represents higher expression. Proportionally with the number of genes contained, cluster I has the major number of TFs represented (44), with some families such as the *JUMONJI*, *LFY* and *TCP* being exclusively detected in this cluster.

HD) transcription factor family (Pp1s31_206V6.1). In *A. thaliana*, genes of this family were shown to have preferential expression in floral tissues and roles during flower development have been proposed (Tan and Irish, 2006). Moreover, homologs of the *WUSCHEL-RELATED HOMEOBOX* (*WOX*) genes described to maintain undifferentiated stem cells (Laux et al., 1996; Mayer et al., 1998) were found in Clusters II (Pp1s29_82V6.1) and Cluster III (Pp1s217_9V6.1), where preferential expression in the antheridia stages was observed. Consistently, the *PpWOX* proteins (Phypa_7235 and Phypa_91889, respectively) are situated in the more ancient clade of the phylogenetic tree, close to *Arabidopsis* genes (*AtWOX10, AtWOX13, AtWOX14*) described to prevent premature differentiation in roots and flowers (Deveaux et al., 2008).

In general, we observed high frequency of *MYB* and *MYB-like* TFs distributed along the clusters, including the previously mentioned *DUO1a* and several highly similar genes (Supplemental figure 5) for which their role in *P. patens* remains to be addressed. Additionally, in almost all the clusters we found members of the *TRIHELIX* (*HELIX-LOOP-HELIX-LOOP-HELIX*) and/or the *ETHYLENE-RESPONSIVE FACTOR* (*ERF*) families, implicated in responses to external stimulus and stress (Fujimoto et al., 2000; Qin et al., 2014).

3.7 Translational reporters of Cluster I genes PpPERK1 and PpZOITO failed to show protein expression in antherozoids

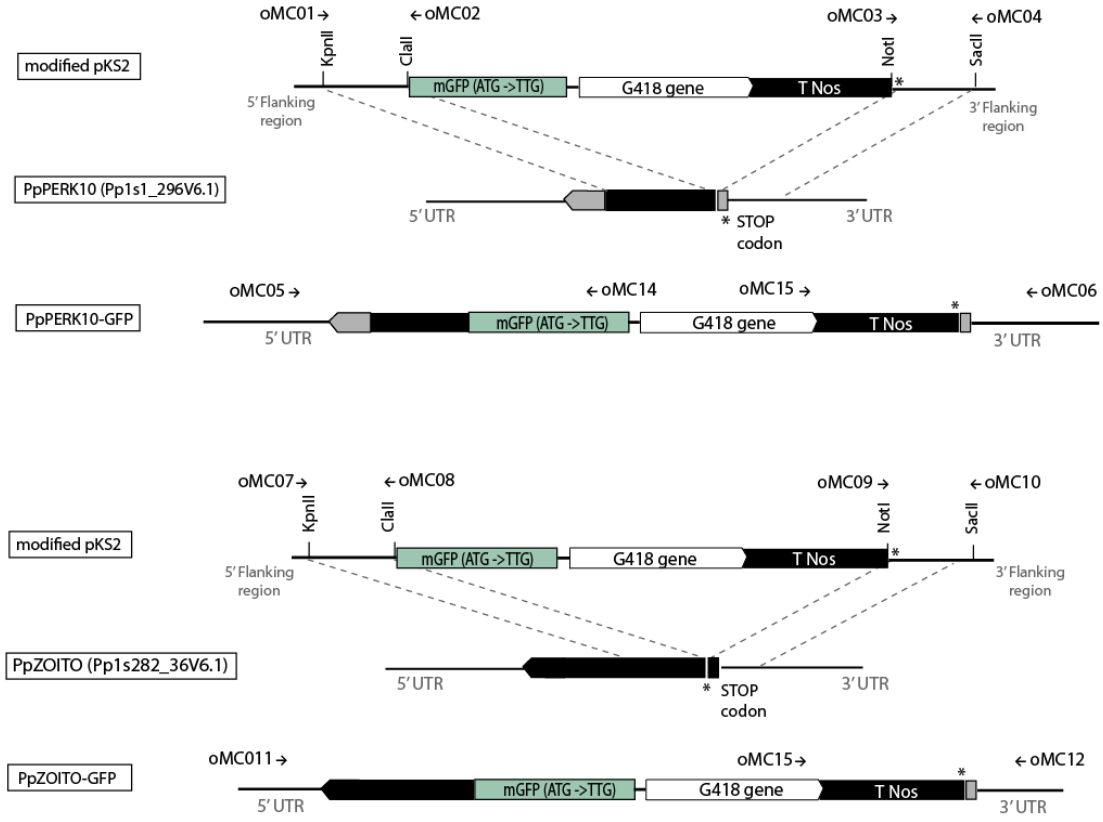
Based on their predicted function and expression levels, two antherozoid preferentially expressed genes with differing expression levels were selected to generate GFP translational reporters. The homolog of the *Arabidopsis PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASE 10* (*PERK10*) (Pp1s1_296V6.1) showed very high expression in antherozoids (21766 in *Gransden* and 21704 in

Villersexel, respectively), while the putative *CIRCUMSPOROZITE* (*ZOITO*) gene (Pp1s282_36V6.1), annotated by BLAST (E-val= 0.00009 and Bit score=52.8) showed expression of 2019 in *Gransden* and 1812 in *Villersexel*. In both cases, the stop codon of the protein was removed and GFP was fused in frame (Figure 7A) by homologous recombination (Schaefer et al., 1991). After three rounds of antibiotic selection the correct insertion of the transgene was corroborated by PCR, and positive stable transformants were selected. In the case of *PpPERK10-GFP*, fourteen independent lines were generated while *for PpZOITO-GFP* two lines were obtained. However, none of the eight *PpPERK10-GFP* transformants observed, or *the PpZOITO-GFP* lines, showed expression in the antherozoids or any gametophytic tissue observed (Figure 7B).

3.8 The male gametophyte of *P. patens* exhibits expression bias for new genes

The observation of a high number of genes being enriched and preferentially expressed in the antherozoids was intriguing, especially since the transcriptional capabilities of spermatozoa are highly controversial (for a review see Miller et al., 2005). Furthermore, the low overlap of antherozoid enriched transcripts with antheridia stages (Supplemental figure 5) suggested active transcription to take place in

A.


modified pKS2
oMC01→
←oMC02
oMC03→
←oMC04
KpnII
ClaII
NotI
SacII
mGFP (ATG->TTG)
G418 gene
T Nos
5' Flanking region
3' Flanking region
PpPERK10 (Pp1s1_296V6.1)
5' UTR
* STOP codon
3' UTR
PpPERK10-GFP
oMC05→
←oMC14
oMC15→
←oMC06
5' UTR
3' UTR
modified pKS2
oMC07→
←oMC08
oMC09→
←oMC10
PpZOITO (Pp1s282_36V6.1)
PpZOITO-GFP
oMC011→
oMC15→
←oMC12

B.

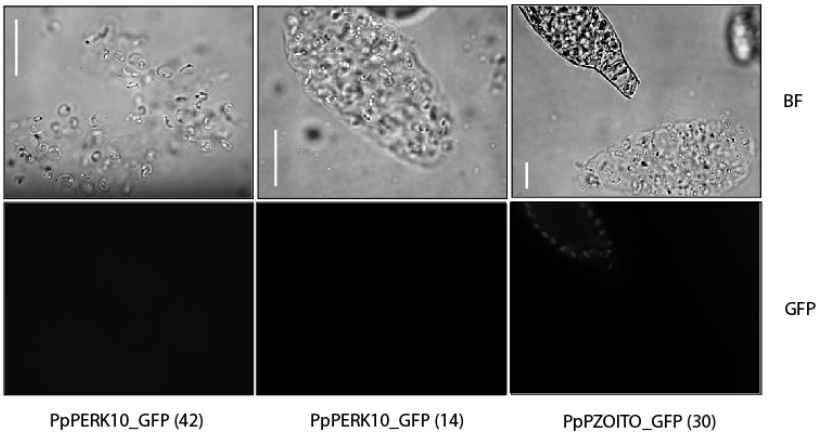

BF
GFP
PpPERK10_GFP (42)
PpPERK10_GFP (14)
PpPZOITO_GFP (30)

Figure 7. GFP reporter lines with no detectable fluorescence in the antherozoids.
The two antherozoid preferentially expressed genes *PpPERK10* (Pp1s1_296V6.1) and *PpZOITO* (Pp1s282_36V6.1) belonging to cluster I were selected to generate fluorescent translational reporters. **A.** Scheme representing constructs generated for homologous recombination in *P. patens*. Using a modified pKS2 vector, gene sequences were fused in frame with GFP. **B.** The released antherozoids were observed for GFP signal under the microscope. However, no fluorescence was detected in any of the four *PpPERK10-GFP* transformants observed, nor the *PpZOITO-GFP* lines. Representative pictures of lines PpPERK10-GFP (14), PpPERK10-GFP (42) and PpZOITO-GFP (30) are shown. In the upper panel images in bright field (BF) are shown. Scale bars = 25µm.

the antherozoids. In order to evaluate, if those differentially expressed transcripts were originated by a general enhanced transcription or product of specific regions being transcriptionally active, we determined the total number of genes expressed in the male gametogenesis related samples and compared them with the previously analyzed samples (Chapter II). For comparative reasons only tissue samples corresponding to the *Gransden* strain were analyzed.

In the comparison of total number of expressed genes among *Physcomitrella* tissues (Figure 8A), antherozoids showed the lowest number to 16226 transcripts representing nearly 50% of the microarray included genes. In opposition, the mature antheridium was the tissue with the highest number of expressed genes (19101) corresponding to 58.7% of the array. The female gametophyte exhibited a similar percentage (58.3%) with 18974 expressed transcripts, followed very closely by the immature antheridium with 18921 (58.1%). This results demonstrated that the antherozoids are not more transcriptionally active than other tissues, but that the transcribed genes are highly expressed and with a tendency to be expressed only in this tissue. Interestingly, the previously described

propensity of antherozoid enriched transcripts to lack annotation indicated that these highly expressed genes may correspond to evolutionary new sequences for which no homologs were found by BLAST. Similarly, expression bias towards new genes being expressed in the male reproductive organs have been reported both for plants (Cui et al., 2014; Wu et al., 2014) and animals of diverse taxa (for a review see Kaessmann, 2010), leading to the "out of testis/out of pollen" hypotheses that postulate the male tissue as a source of gene innovation.

In order to determine if a similar process is taking place in *Physcomitrella* the proportion of young genes expressed in each tissue was estimated. Gene evolutionary innovations were traced using a phylostratigraphic approach (Domazet-loso and Tautz, 2003), where each gene was assigned to a phylostratum according to the emergence of their founders and based on gene age estimation and microarray expression data the transcriptome age index (TAI) of each tissue was determined. The significance of the TAI values observed was estimated by permuting gene-phylostratum assignments (see Material and Methods). In Figure 8B the TAI values of each tissue are shown, where a higher TAI corresponds to a major proportion of young genes being expressed. The observed TAI values ranged from 1.85 to 1.38 with antherozoids showing the highest value, followed by the mature and immature antheridia with TAI values of 1.45 and 1.42, respectively. Interestingly, archegonia and mature sporophyte had also values of 1.42 while spores and the tip growing tissues showed the lowest TAI with values ranging from 1.38 to 1.39. The observed pattern corroborates the expression bias for new genes in the male reproductive organs of *Physcomitrella*, in particular in the antherozoids.

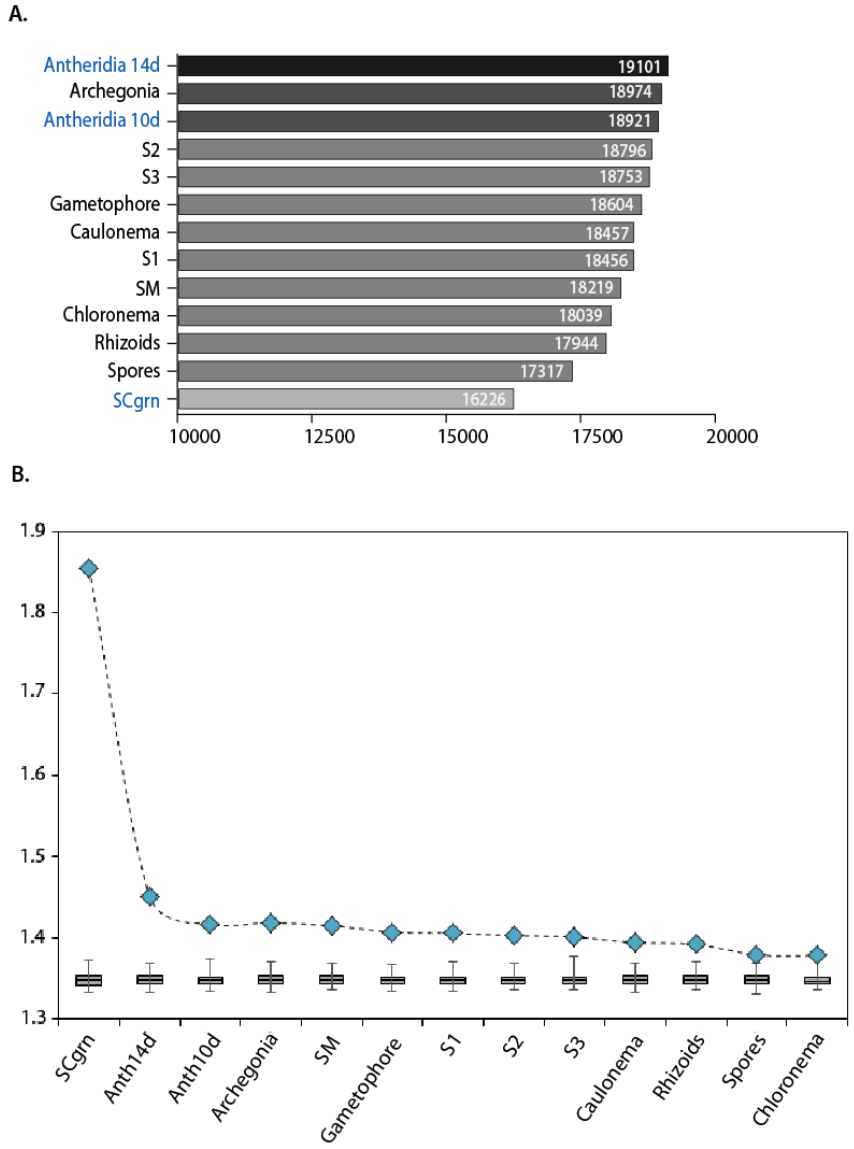


Figure 8. Transcriptome complexity of *P. patens* male gametangia.
The number of expressed transcripts and the proportion of young genes expressed in male gametes indicate high complexity. **A.** Making use of the experimentally determined presence/absence threshold, the total number of transcripts called present in each tissue was determined. Antherozoids (SCgrn) showed the smallest number of genes expressed when compared with the rest of the tissues. **B.** The transcriptome age index (TAI) was estimated using a phylostratigraphic approach. Green bars indicate observed TAI patterns, while grey boxplots indicate TAI profiles obtained by permuting gene-phylostratum assignments. Observed TAI values were significantly higher ($P<0.001$) for all tissues. In the analysis, the antherozoids and the mature antheridia displayed the highest TAI, suggesting a major proportion of young genes being expressed in these sample types.

3.9 A relaxed epigenetic state in the antherozoids?

It has been hypothesized that the particular relaxed chromatin state observed at early stages of spermatogenesis (Kaessmann, 2010; Soumillon et al., 2013) or in the pollen vegetative nucleus of *Arabidopsis* (Wu et al., 2014) favors the preferential expression of new genes. Similarly antherozoids may display a relaxed epigenetic state. We analyzed the expression patterns of methylation and chromatin remodeling related genes in *Physcomitrella.* The homolog genes were obtained from the literature when reported or identified by BLAST using *Arabidopsis* protein sequences. Additionally, comparisons with the *Arabidopsis* sperm transcriptome (Borges et al., 2008) were conducted.

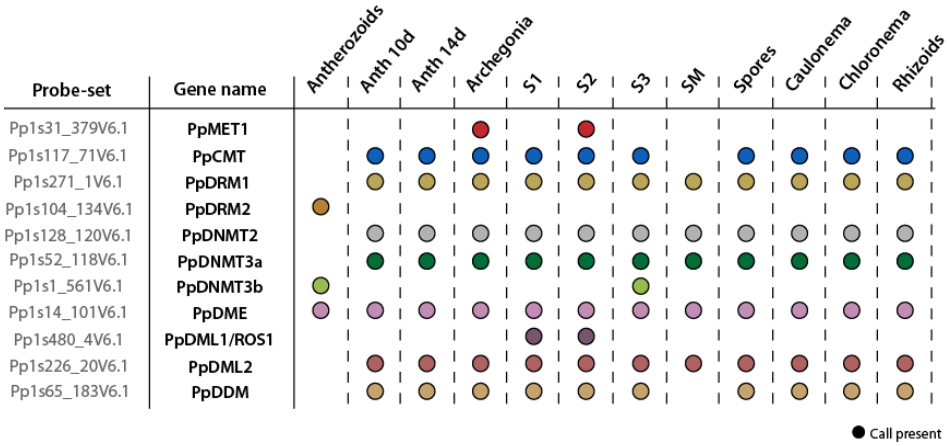


Figure 9. Expression of methylation related genes in *P. patens* tissues.
The expression patterns of methylation and chromatin remodeling related genes in *P. patens* tissues was analysed. *Physcomitrella* homolog genes were obtained from the literature when reported or identified by BLAST using *Arabidopsis* protein sequences. No expression of the canonical CG methyltransferases *MET1* (Pp1s31_379V6.1) and *CMT3* (Pp1s117_71V6.1) was detected in the antherozoids. Expression of the *de novo* CHH methyltransferases *DRM2* (Pp1s104_134V6.1), *DNMT3a* (Pp1s52_118V6.1) and *DNMT3b* (Pp1s1_561V6.1) was observed as well as expression of the DNA glycosylase *DME* homolog gene (Pp1s14_101V6.1).

Remarkably, the antherozoids showed no expression of the *DNA METHYLTRANSFERASE 1* (*MET1*) (Pp1s31_379V6.1) and the *CHROMOMETHYLASE 3* (*CMT3*) (Pp1s117_71V6.1) genes (Malik et al., 2012), identified as the canonical CG and CHG methylation enzymes, respectively, important for repressing gene expression in *Arabidopsis* (Bartee et al., 2001; Kankel et al., 2003). Only the expression of *de novo* CHH methyltransferases was observed (Figure 9). The *DOMAINS REARRANGED METHYLTRANSFERASE 2* (*DRM2*) (Pp1s104_134V6.1) homolog was preferentially expressed in *Physcomitrella* sperm cells. Moreover, expression of the closer orthologs to animal (cytosine-5-)-methyltransferase *DNMT3b* (Pp1s1_561V6.1) (Malik et al., 2012) was observed in the antherozoids, in accordance with expression of the mammalian *DNMT3b* in testis (Chen et al., 2002; Watanabe et al., 2004). Finally, absence of the *DECREASE IN DNA METHYLATION* (*DDM1*) (Pp1s65_183V6.1) helicase, shown to be involved in maintaining methylation patterns (Vongs et al., 1993), and the expression of the DNA glycosylase *DEMETER* (*DME*) homolog gene (Pp1s14_101V6.1) (Table 3) was observed in the antherozoids.

4. DISCUSSION

4.1 Gene expression during male gametogenesis in *P. patens* reveals the uniqueness of the antherozoid transcriptome

We present gene expression data of three distinct stages of spermatogenesis in *P. patens* including immature antheridia, mature antheridia and released antherozoids. To our knowledge this represents the first transcriptome study of male gametogenesis in non-vascular plants, making also available for the first time whole-genome expression data of plant motile sperm. As a prerequisite to isolate the samples we characterized the developmental stages of gametangia under our laboratory conditions. In agreement with previous observations (Landberg et al., 2013) we observed specific morphological transitions during the development with small discrepancies on the specific dates that could perhaps be explained by differing growth conditions. During the characterization conducted we also observed differences in sperm performance between *Physcomitrella Gransden* and *Villersexel* strains, for which variation in fertility rates had consistently been reported (O'Donoghue et al., 2013; Perroud et al., 2011).

The limitations to generate these transcriptomic data, imposed by the difficulties to isolate the samples, were overcome by combining the collection methods with a highly efficient cDNA synthesis and amplification protocol, allowing the generation of microarray data from very small RNA quantities. During previous work in our lab, the efficiency of the Ovation WT system (NuGen Technologies, Inc) was evaluated and high quality microarray data were generated. In the case of antherozoids no DNAse treatment was effective. And although this represents a potential problem and caution must be

taken when evaluating the results obtained from those samples, we performed qPCR controls that increase our confidence in the data. The particular absence of the 28S and 18S ribosomal peaks in the RNA profile of antherozoid samples is in agreement with published results and observations in human sperm, for which the lack of rRNA transcripts are considered markers of quality (Ostermeier et al., 2002). This not only demonstrates sample quality, but also the absence of tissue contamination. The overlap of transcripts in eight independent biological replicates corresponding with two *P. patens* strains (see below) and the appearance of specific transcripts with sperm related functions are further evidence of non-biased amplification.

The transcriptome relationships among *P. patens* tissue samples were approached by HC and PCA analysis, showing a strong separation of antherozoid samples from other reproductive and somatic tissues. The demonstration of a very different global gene expression from somatic and female reproductive tissues in pollen (Pina et al., 2005) and later in sperm cells of *Arabidopsis* (Borges et al., 2008) correlates with our findings, underlining the uniqueness of sperm transcriptomes. In our analysis the transcriptomic differences of antherozoids additionally isolate the sperm cells from the mature antheridia that contain them. Although this is an unexpected result for us, the PCA analysis generated by Borges et al., (2008) separates as well the sperm from pollen. We hypothesize that contribution of surrounding cells may account for this difference in global expression patterns. In *Arabidopsis* the strong differences of both pollen and sperm from the rest of the tissues is evidenced by their separation in PC1. But a further separation by PC2 between sperm and pollen is also observed. This weaker separation of the two tissues in

Arabidopsis in contrast with the observed separation by PC1 in *Physcomitrella* correlates with antheridia having more contribution of non-sperm cells than pollen. Consistently, the mature antheridium grouped with the other reproductive organs and early sporophyte tissues while pollen is located far from the rest of the analyzed tissues. Interestingly, in both pollen and antheridia input from other sources than the RNA from sperm are present, namely somatic cells in antheridia and the vegetative nuclei and pollen cytoplasm in pollen. And even when that RNA contribution would be very different in the two mentioned examples in both cases is enough to create a separation from sperm. Under this scenario, the observed separation between antherozoids and mature antheridia is validated, and an attribution of gene expression differences to sample problems or amplification bias caused by the low RNA quantity obtained from antherozoids seems less plausible. Additionally, the highly similar transcriptome in both strains studied shows consistency in the antherozoid samples. The great proportion of enriched and preferentially expressed genes common among *Gransden* and *Villersexel* (>70%) adds further support to our data and validates the transcripts identified in moss sperm.

4.2 Active transcription in mature antherozoids is indicated by gene expression changes during male gametogenesis

According to our results very dynamic processes of gene expression may be occurring during spermatogenesis, where an increase of enriched and preferentially expressed transcripts is observed to occur from each spermatogenesis stage to the next. For long, the idea that transcription and translation was not occurring in animal sperm cells was widely accepted. During the 70s, evidence of reduced transcription in post-meiotic murine sperm (MacLaughlin and Terner,

1973) and transcription restricted to mitochondrial products in bovine gametes (Premkumar and Bhargava, 1972) provided support for this concept. Although some doubts were raised regarding a complete transcriptional silence, indications of absent nuclear translation were solid on those works. The lack of rRNAs in mature sperm (Ostermeier et al., 2002) not only supported the idea of translationally silenced gametes, but was also taken as evidence of transcription not being necessary and therefore absent in mature sperm cells. With the emergence of high-throughput expression technologies, allowing the analysis of thousands of transcripts in parallel, the study of sperm RNA has regained attention in the last decade. To date, the presence of mRNA and sncRNA in mature sperm of several plant and animal species (for example see Borges et al., 2008; Engel et al., 2003; Fischer et al., 2012; Krawetz et al., 2011; Ostermeier et al., 2002; Russell et al., 2012) is well documented. However, the highly compacted chromatin state observed in sperm nuclei, mainly due to histone replacement by protamines and transition proteins (Balhorn, 2007; Reynolds and Wolfe, 1984), has led many to believe that mature sperm is not transcriptionally active. Under that reasoning, it has been proposed that residual mRNA from previous spermatogenesis stages is kept in mature sperm (Hecht, 1998), implying a strong correlation of transcripts among stages. The striking increase of expressed genes during *P. patens* gametogenesis stages, as well as the low overlap of transcripts between them, challenges this idea. According with our data, new synthesis of RNA must be occurring at mature sperm stage. Despite evidence of inactive transcription in human sperm derived from the failure to detect incorporation of radioactive labeled uridine triphosphate (^{33}P-UTP) (Grunewald et al., 2005), several pieces of evidence support our

findings. Similar indication against a passive retention of RNA came from *Drosophila* microarrays showing low correspondence between highly expressed transcripts of testis and those in sperm (Fischer et al., 2012). Interestingly, newly synthesized RNA in post-meiotic *Drosophila* cells was detected by BromoUridine (BrU) incorporation (Vibranovski et al., 2010) and maintenance of euchromatin regions through histone retention in fern (Rejon et al., 1988) suggest that a selective compaction of the nuclei may allow certain regions to be transcribed. In accordance, the lower proportion of expressed genes in *Physcomitrella* antherozoids with respect to the rest of the tissues indicates a restricted transcriptional activity on those cells. This further suggests that the great amount of enriched genes observed in our analysis is not product of a general transcriptional activity but a more directed process. However, the mechanisms allowing selective nuclear compaction, if is the case in *P. patens* antherozoids, requires further clarification.

4.3 Expression patterns of *PpDUO1* and *PpHAP2* suggest a role during male gametogenesis

Based on sequence similarity, we identified two close orthologs of the male germline specific *R2R3 MYB* transcription factor *DUO POLLEN1 (DUO1)*, recognized as a master regulator of male gamete differentiation in *A. thaliana*. By binding to a conserved MYB motif, *AtDUO1* activates several specific transcripts required during germ-line differentiation (Borg et al., 2011). When mutated, the generative cell is unable to undergo the second pollen mitosis resulting in a single non-functional sperm cell (Durbarry et al., 2005; Rotman et al., 2005). According to our data, both *DUO1a* (Pp1s16_281V6.1) and *DUO1b* (Pp1s114_136V6.1) are expressed in the male gametophyte. But while *DUO1a* is only expressed in the male, *DUO1b* shows

expression also in the female. Interestingly, *AtDUO1* is only acting in the male gamete suggesting that during evolution the genetic programs for gamete development diverged until a strong differentiation arose. Among the genes *DUO1* regulates during sperm cell specification in *Arabidopsis*, several male germ-line enriched or preferentially expressed genes have been identified, such as *HAPLESS 2* / *GENERATIVE CELL SPECIFIC 1* (*HAP2/GCS1*) and *GAMETE EXPRESSED PROTEIN 1* (*GEX1*) (Brownfield et al., 2009). And interestingly in *Physcomitrella* the identified homologs of those genes showed expression during male gametogenesis. *HAP2/GCS1* is a transmembranal protein essential for gamete membrane fusion that has been identified in diverse eukaryotes with a restricted expression to the male gamete (see Wong and Johnson, 2010). Disruption of *HAP2/GCS1* function in *Plasmodium* and *Chlamydomonas* blocks gamete fusion despite morphologically normal gametes (Liu et al., 2008), while in *Arabidopsis* the two sperm cells of mutant *hap2/gcs1* were unable to fuse and problems in pollen tube guidance were observed (von Besser et al., 2006). In *Physcomitrella*, we identified one homolog gene. However, the predicted protein (Phypa_231342) is separated into two transcripts represented in the microarray by two probe sets. Although we hypothesize that they correspond to only one functional transcript, since a similar expression pattern is observed, the slight differences in the expression values of both probe sets (Pp1s129_185V6.1 and Pp1s129_186V6.1) may account for different transcript isoforms. In our analysis we also included the homolog of *GEX1*, a versatile protein with roles in gametophyte development as well as during early embryogenesis. In mutant Arabidopsis *gex1* plants where the extracellular domain is disrupted, effects in both male and female gametogenesis are observed while mutations affecting the

cytoplasmic domain result in embryo arrest at early stages (Alandete-Saez et al., 2011). In *Physcomitrella*, expression of the *GEX1* homolog (Pp1s122_53V6.1) is observed in the two antheridia stages, in the archegonia and in the different sporophyte stages. However, we observed an increase of transcript expression from archegonia to pre-meiotic sporophyte tissues and elevated expression at the mature sporophyte stage.

4.4 Male gametogenesis gene clusters reveal possible transcript functions.

The observation of RNAs in sperm cells of a wide variety of organisms is not only changing previous conceptions, but it also raises new questions regarding their functions. Although residual, non-functional roles had been proposed (see above), increasing evidence suggests complex functions for sperm transcripts (Miller, 2014; Miller et al., 2005). The generated clusters containing particular signatures of gene expression offer a gene functional classification. Remarkably, comparison of our data with data generated for mouse spermatogenesis shows similar GO enriched categories. In the published work of Soumillon et al., (2013), four clusters with particular expression patterns during spermatogenesis were generated using spermatogonia, spermatocytes, spermatids and spermatozoa samples. Consistently with our observations for cluster II, enrichment of flagella and motility processes is reported for a cluster with a similar strong downregulation in late spermatogenesis. The presence of those transcripts involved in motility, and the inclusion of *PpHAP2* in cluster II, supports the idea that important proteins for sperm function are synthesized previous to antherozoid release. Although mitochondrial mediated translation was shown by incorporation of labeled amino acids during sperm capacitation in mammals (Gur and

Breitbart, 2006), we hypothesize that translation of important proteins for antherozoid function may preferentially occur prior to release. In the mouse spermatogenesis cluster where up regulation at the late stage occurs, enrichment of apoptotic related processes is observed. In accordance, *Physcomitrella* cluster I shows apoptosis as one of the enriched categories. In animals, the important roles of this process during spermatogenesis have been recognized (Shaha et al., 2010), supporting a functional role during moss male gametogenesis.

Based on the eight clusters generated, the high complexity and gene diversity contained in *Physcomitrella* male gametes is displayed. For some clusters very specific processes were identified and supported by conservation among evolutionary distant organism, yet experimental evidence will be necessary.

4.5 Antherozoid transcripts may play a role during early embryogenesis.

Among the complex pool of transcripts generated during *Physcomitrella* spermatogenesis, we report possible functions for early embryo development. The presence of post-embryonic development related genes in clusters I and IV, characterized by an increased expression in the antherozoids with respect to antheridia stages, was found based on our GO enrichment analysis. In the first cluster, specific genes such as the *CLAVATA* related gene *CLE10* (Pp1s292_57V6.1) predicted to be involved in meristem cell proliferation (Sharma et al., 2003), the cell cycle regulator *ANT* (Pp1s66_255V6.1), shown to be important during ovule and shoot development (Klucher et al., 1996; Mizukami and Fischer, 2000), as well as two *TCP class I* TF genes (Pp1s446_21V6.1 and Pp1s348_6V6.1), known to regulate cell proliferation (Martín-Trillo and Cubas, 2010) were identified. The evidence of transcript storage

during spermatogenesis (Hecht, 1998) and the reduced translational capabilities evidenced by absence of intact rRNA in mature sperm (Johnson et al., 2011; Miller et al., 2005), indicates a temporal separation of transcription and translation programs in the sperm. And the lack of correspondence among the described functions with immediate tasks of mature sperm support further a role for those transcripts during fertilization or early embryogenesis.

The presence of the *PUMILIO* homolog genes (Pp1s96_109V6.1, Pp1s96_110V6.1 and Pp1s35_161V6.1) in cluster I represent additional examples of possible paternal contribution, especially since very high expression of those genes is observed in the antherozoids. Both in plants and animals, *PUMILIO* or *PUMILIO-like* proteins (*PUMILIO-FBF* or *PUF* proteins) are known to be involved in transcriptional control by several mechanisms including the direct binding to RNA (Abbasi et al., 2011; Chritton and Wickens, 2010; Murata and Wharton, 1995), the control of poly-A tail length mediated by the *CCR4-NOT* complex (Goldstrohm et al., 2007; Wreden et al., 1997) and the blockage of transcription initiation factor binding (Cao et al., 2010). In *Arabidopsis* the involvement of *PUM* proteins in the regulation of developmental processes in plants has been proposed in accordance with protein expression assays revealing diverse localizations, such as stem cells, seedlings and reproductive tissues (Abbasi et al., 2011; Aggarwal et al., 2010). Although functional characterization is missing, the *in vitro* interaction of some *PUM* proteins with the DNA sequences of *CLAVATA 1* (*CLV-1*), *WUSCHE*L (*WUS*), *ZWILLE* (*ZLL*) and *FASCIATA 2* (*FAS-2*) genes suggest roles in meristem maintenance and differentiation (Francischini and Quaggio, 2009). Interestingly, *ATPUM16* (At5g59280) and *ATPUM17* (At1g35850) are specifically expressed in sperm cells (Borges et al., 2008). In animals, the *PUM* genes are implicated in the maintenance

of germ cell identity (Crittenden et al., 2002; Forbes and Lehmann, 1998; Moore et al., 2003) and roles in *Drosophila* embryo patterning have been shown (Parisi and Lin, 1999). But even if roles in gamete development cannot be discarded, particular evidence in *Drosophila* embryos offers an interesting scenario. In *Drosophila PUM* is regulating the transcription of the maternally derived *HUNCHBACK* (*HB*) gene, ensuring the transition to active zygote transcription for proper segmentation (Murata and Wharton, 1995). Although there is no evidence of this transcript being of paternal origin in *Drosophila*, the expression of its homolog in *Physcomitrella* antherozoids offers this possibility. Upon translation, the *PUM* transcripts found in antherozoids may play a role in the processing of other paternally produced transcripts and/or in preventing the expression of the maternally derived ones. With the delivery of these cues upon fusion, successful fertilization could be signaled and zygote development would progress.

Although it was postulated that all components required for embryogenesis, such as the organelles and cytosol, are exclusively provided by the oocyte (Theodor Boveri, reviewed by Laubichler and Davidson, 2008), the observation of sperm centromeres being required for zygote cell progression (Palermo et al., 1997; Terada et al., 2004) and the later identification of the sperm delivered oocyte activation factor *PLCz* (Saunders et al., 2007, 2002) demonstrate that paternal contribution is not restricted to its genome. The delivery of sperm messenger RNA into the oocyte during fertilization was first demonstrated in mammals where intact human sperm transcripts were detected in wild type hamster oocytes, demonstrating also that RNA remained intact for some hours after fertilization (Ostermeier et al., 2004). Similar evidences were generated in *Drosophila* where paternal transcripts were detected in the oocyte after fertilization prior zygote transcription activation (Fischer et al., 2012). The observation

of transgenerational behavioral and metabolic inheritance in mice using sperm RNA from traumatized individuals to fertilize wild type oocytes suggests that environmental cues can also be transmitted via paternal contribution (Gapp et al., 2014). The detected changes in microRNA (miRNA) and Piwi-interacting RNA (piRNA) in the sperm of the stressed male progenitors used in this experiment suggest further that delivery of sperm small non-coding (sncRNA) may account for such effect. To date, in vivo characterization of functions for sperm delivered mRNAs is restricted to an assay in *C. elegans* showing that selective mutagenesis of paternal transcripts affects embryo viability (Stoeckius et al., 2014). In humans, the importance of the sperm produced *PREGNANCY SPECIFIC B-1-GLYCOPROTEIN 1* (*PSG1*) and *HUMAN LEUKOCYTE ANTIGEN-E* (*HLA-E*) transcripts for embryogenesis and implantation has been documented (Avendaño et al., 2009). And examples of paternal delivery include the *FORKHEAD BOX G1B* (*FOXG1B*) and *WINGLESS-type MMTV INTEGRATION SITE FAMILY MEMBER 5A* (*WNT5A*) transcripts, important for embryo patterning (Ostermeier et al., 2004). In plants evidence for paternal contribution is restricted to the *SHORT SUSPENSOR* (*SSP*) gene, involved in the YODA-MAP kinase pathway to regulate embryo elongation (Bayer et al., 2009). Interestingly, expression of the *SSP* gene is confined to sperm cells. However using promoter-reporter fusions the authors were unable to detect GFP signal in sperm or pollen and fluorescence was only observed briefly in the zygote. The failure to detect GFP expression in the antherozoids of our knock-in translational reporters (*PpPERK10-GFP* and *PpZOITO-GFP*) resembles this case. Since this also suggests that expression may take place in the early stages of the sporophyte, observations at that developmental point are needed to clarify, if those transcripts might have a role in the early embryo.

Although functional evidence of paternal contribution in the zygote especially in plants is still limited, the available data from animals indicate a widespread phenomenon. In this work we generated a list of genes with possible roles after delivery, providing testable hypothesis that may offer further evidence for a still controversial theme. Understanding the mechanisms that allow transcript stability and selective degradation or translation to occur would help to clarify the potential role of paternal transcripts during embryogenesis.

4.6 *Physcomitrella* antherozoids may act as a source of evolutive innovation

The origin of new genes has been widely studied since those are considered a major source for evolutive adaptation. Surprisingly, in animals several examples of testis-biased expression in newly originated genes, both *de novo* or by duplication, have been shown (Begun et al., 2007; Emerson et al., 2004; Levine et al., 2006). In hominids, for example, the *PIPSL* ubiquitin-binding gene was originated as result of transcript fusion and is transcribed specifically in the testis of both humans and chimpanzees (Babushok et al., 2007), while in *Drosophila* the testis expressed *JGW* gene was originated by fusion of two DNA fragments (Long and Langley, 1993). Based on existing literature showing this tendency it has been postulated that testis may function as a source of evolutionary innovation (Kaessmann, 2010). Recently, the preferential expression of new genes in pollen of Arabidopsis and rice has been shown, extending this model to plants (Cui et al., 2014; Wu et al., 2014). According to the general “out of the male” model it has been suggested that newly generated genes would initially be expressed in the male gametes, and later a more diverse expression pattern would evolve (Cui et al., 2014). In *P. patens*, our data indicate that

reproductive organs express more evolutionary young genes than somatic tissues. Despite antherozoids expressing lower percentage of genes than the other tissues, it displayed the highest TAI value. This reveals a greater proportion of new genes being expressed but also at higher levels. Moreover, the preferential lack of annotation in antherozoid enriched genes may account for recently originated genes for which no alignment with known sequences was possible. In accordance, both in rice and in *Arabidopsis* young genes showed a more common lack of annotation with only 3% and 33% genes respectively, being functionally annotated (Cui et al., 2014).

Although the mechanism allowing this preferential expression of younger evolutionary genes in the male organs remains elusive, it has been hypothesized that chromatin relaxed states occurring during spermatogenesis in animals (reviewed in Sassone-Corsi, 2002) or in the vegetative nucleus in pollen (Calarco et al., 2012) may represent the opportunity for those genes to be expressed. Both in plants and in animals, DNA methylation is associated with silencing the expression of genes and transposable elements (TEs). In plants, maintenance methylation at symmetrical CG and CHG motifs, where H represents any residue but G, is accomplished respectively by the *DNA METHYLTRANSFERASE 1* (*MET1*) and the *CHROMOMETHYLASE 3* (*CMT3*) enzymes (Jullien et al., 2012). *De novo* methylation in the CHH context relies on the production of 24-nt small interfering RNA (siRNA) and in *Arabidopsis* it is performed by the *DOMAIN REARRANGED METHYLTRANSFERASE 2* (*DRM2*), ortholog of the animal *DNMT3* enzymes (Cao et al., 2000; Daxinger et al., 2009). In *Physcomitrella*, homologs of the different *Arabidopsis* methyltransferases have been identified (Dangwal et al., 2014; Malik et al., 2012; Noy-Malka et al., 2014). Additionally, two DNA methyltransferase genes showing retention of Ubiquitin Associated

(UBA) domain, similar to the *DNMT3* in animals, have been reported (Pp1s52_118V6.1 and Pp1s1_561V6.1) (Malik et al., 2012).

According to our expression data, expression of *PpMET1* (Pp1s31_379V6.1) and *PpCMT3* (Pp1s117_71V6.1) genes was not detected in the Antherozoids, and only expression of *de novo* methyltransferases *PpDRM2* (Pp1s104_134V6.1) and *PpDNMT3b* (Pp1s1_561V6.1) was observed. These expression patterns are different from the ones observed for *Arabidopsis* sperm cells and more similar to the vegetative nuclei, where methyltransferases associated to maintenance of methylation in the CG and CHG context are not expressed, and only CHH methylation is maintained (Borges et al., 2012; Calarco et al., 2012). Among other cases, the absence of the *DECREASE IN DNA METHYLATION* (*DDM1*) (Pp1s65_183V6.1) helicase, shown to be involved in maintaining methylation patterns (Brzeski and Jerzmanowski, 2003; Vongs et al., 1993; Zemach et al., 2013), and the presence of the DNA glycosylase *DEMETER* (*DME*) homolog gene (Pp1s14_101V6.1), suggest further a more relaxed chromatin state than in *Arabidopsis* sperm where *DDM1* (and not *DME*) is expressed (Borges et al., 2008; Schoft et al., 2011). Although the presence of other enzymes undertaking such functions cannot be discarded, it is tempting to speculate that the described expression patterns may account for a chromatin permissive state that promotes the preferential expression of new genes. How this state would be balanced with maintaining the genomic information and safe-guarding it from transposable elements awaits clarification.

5. MATERIAL AND METHODS

5.1 Plant Material and Growth Conditions

Wild-type *Physcomitrella patens* Bruch & Schimp. subsp. Patens from *Gransden* Wood and *Villersexel* K3 strains were grown routinely on Petri dishes with 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) under previously described light and temperature conditions (see Chapter II). Vegetative propagation was maintained by subculture every 6-7 days by mechanical disruption, and replacement with tissue resulting from sexual reproduction was conducted at least three times a year. The development of reproductive structures was induced by exposure to short day conditions (17ºC, 8h light and 50% humidity), as detailed in Chapter II Material and Methods.

5.2 Gametangia development

Daily observations after induction of the gametangia were conducted under the microscope (Leica DM LB2) after manually dissecting the organs from the gametophore apex. Special focus was put on the male gametangia, for which morphological changes such as size and colour were recorded. From day 6 after induction, antheridia were observed as small clusters and in the subsequent days the size and number of organs per cluster increased and changes in coloration were observed. When immature, the male organ is very green and full of chloroplasts and as development progresses becomes yellow or brownish. At day 14 after induction, mature antheridia were detected, but only few sperm releases occurred. We observed that releases take place preferentially 24 hours later, when more antheridia from each cluster are mature. At this point we also observed dead

antheridia since the tissue seems to degrade quickly after antherozoid release.

5.3 Tissue Isolation

The developmental stages of antheridia were collected from the *Gransden* strain. At days 10 (early stage) and 14 (late stage) after the induction of gametangia, the tip of the gametophores was dissected and the antheridia where manually isolated under the stereoscope (Nikon, SMZ800) directly into Trizol tubes. Antheridia located at the top of each cluster were preferentially dissected so that enrichment in early and late stages of development, respectively, was achieved. Three biological replicates for each developmental stage were generated by pooling 200 antheridia for every RNA extraction. For the isolation of antherozoids, preparations with manually dissected antheridia from more than 50 gametophores were generated by placing them on a cover slip with 50 µl of sterile MilliQ water, where sperm release occurred naturally after few minutes. Using micromanipulators (Eppendorf CellTram®) attached to an inverted microscope (Leica, DMIRE2) the clusters of cells were collected with microcapillaries. The collected sperm was placed directly into Trizol, frozen in liquid nitrogen and stored at -80ºC until use. For RNA extraction, 200-400 clusters were used per replicate and four replicates of each strain were generated. Samples from *Gransden* and *Villersexel* were independently prepared.

5.4 RNA Isolation and cDNA synthesis

Isolation of RNA from collected samples was performed using Direct-zol® columns (Zymo Research Co.) according to manufacturer's instructions. Quantity and sample integrity were measured on an

Agilent 2100 Bioanalyzer using a 6000 Pico Assay (Agilent Technologies). Antheridia samples were subjected to DNase treatment using TURBO enzyme (Ambion, Life Technologies) for 30 min at 37ºC. On average 7.5 ng of total RNA were used for cDNA synthesis using the Ovation Pico WTA System V2 amplification kit (NuGen Technologies, Inc.). All DNase procedures tested on antherozoids resulted in degraded samples that, when used for cDNA synthesis showed a reduction in yield (of 66% in average) and quality. Since preferential amplification of RNA by the Ovation WT system and no co-amplification of contaminating genomic DNA has been reported (Vermeulen et al. 2009), untreated RNA ranging from 1.5 to 3.0 ng were used for cDNA synthesis. qPCRs were performed on all samples using housekeeping genes and tissue specific probes (Supplementary figure 2).

5.5 cDNA labeling and hybridization to 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 (Roche NimbleGen, Inc.).

5.6 Data Analysis and Functional Annotation

The raw expression data from the male gametogenesis related tissues was normalized together with the previously generated transcriptome data (Chapter II) by applying Robust Multi-Array Average (RMA) (Roche NimbleGen, Inc.). The differences in global gene expression among microarrays were assessed using chipster software (Kallio et al., 2011). For that, data was imported as external

file and the pre-normalized option was used. *Physcomitrella patens* genome annotation v1.6 2012.3 (Zimmer et al., 2013) was downloaded from (https://www.cosmoss.org/physcome_project/wiki/Downloads), and complemented with information from the PlantGDB database www.plantgdb.org/XGDB/phplib/download.php?GDB=Pp (*P. patens* genome annotation last modified on May, 2011). In this way 17576 gene models were annotated, corresponding to 54% of all transcripts represented on the array. Additionally, STRING information (Franceschini et al., 2013) and information from several alignments against non-plant databases were included (E value > 0.001) increasing the annotated sequences to 18824 (58%). Transcription factor families were attributed according to the Plant Transcription Factor Database (PTFDB) and confirmed by Pfam domain search (pfam.xfam.org).

Differential expression analysis was conducted using dChip software (Li and Wong, 2001), for which the normalized data was imported as described in Chapter I. In order to determine enriched and preferentially expressed genes pair-wise comparisons were conducted. Transcripts that were called present in a given tissue and identified as significantly up-regulated in every single comparison of this tissue against the rest were named enriched, while transcripts with a present call in only one tissue and absent call in the rest were defined as preferentially expressed. To identify common enriched genes among *Gransden* and *Villersexel* antherozoids, independent analyses of each strain's data against all other tissues were performed using a lower-confidence bound fold-change (LCB-FC) cut-off ranging from 1.5 to 1.8 and a false discovery rate (FDR) below 10%. The common elements on both comparisons were identified

using Venn diagrams (http://bioinfogp.cnb.csic.es/tools/venny/). In a similar way common preferentially expressed genes were identified. For the identification of enriched and preferentially expressed genes during development time-courses (male gametogenesis and sporophyte development) a LCB-FC cut-off ranging from 1.5 to 4.0 was used in pair-wise comparisons, maintaining a FDR below 10% and excluding comparisons with tissues from the same developmental set. In the identification of enriched and preferentially expressed genes during sporophyte development only antherozoids from *Gransden* were included for the comparisons.

5.7 Threshold determination for detection calls

The signal threshold for presence/absence detection calls of already tested tissues (Chapter I) were adjusted according with the new normalization performed including male gametogenesis samples. For those new samples, present and absent calls were generated experimentally in a similar way as previously described. Primers for transcript models with expression values of 350, 400 and 450 were designed for each tissue and amplification was tested by qPCR (ABI7900HT, Applied Biosystems), using 1 ng of cDNA per reaction. Relative expression values were analysed using the housekeeping gene *ALPHA-TUBULIN* (Pp1s215_51V6.1). A similar signal threshold value of 450 was observed for the three male gametogenesis samples. The expression threshold values previously determined (Chapter I) were adjusted accordingly with the values obtained when normalizing all samples together. In all cases, a value of 450 was identified for presence threshold. In order to consider a transcript present a criterion of majority vote was used for replicates, meaning

that a given transcript must have >66% of presence calls in the replicates to consider it present in the tissue.

5.8 Gene clustering and functional enrichment analysis

Log2 expression data of identified gametogenesis enriched genes was analyzed with EXPANDER (EXPression ANalyzer and DisplayER) (Ulitsky et al., 2010). Average linkage hierarchical clustering was performed using Pearson's correlation similarity measurement and manually selected subtrees were exported to generate the groups. Functional enrichment analysis of clustered genes was performed using DAVID Bioinformatics Resources 6.7 (Huang et al., 2009a, 2009b) based on *Arabidopsis* homologs. The enrichment scores of GO terms grouped by biological process, molecular function and cellular component were calculated. EASE scores >0.01 were considered significant.

5.9 Phylogenetic analysis

P. patens protein sequences were retrieved from Phytozome v9.1 database (`www.phytozome.net`), while Arabidopsis sequences for *DUO1* (At3g60460), *HAP2* (At4g11720), and *GEX2* (At5g49150) were obtained from TAIR database (`www.arabidopsis.org`). Alignment of the *DUO1* amino acid sequences was constructed using ClustalW program using default parameters. 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 300 replications). The resulting phylogenetic tree was visualized with MEGA 6.

5.10 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 1). 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.

5.11 Generation of knock-in GFP translational reporters in *P. patens*

The translational GFP reporters for *PpPERK10* (Pp1s1_296V6.1) and *PpZOITO* (Pp1s282_36V6.1) were generated by homologous recombination (Schaefer et al., 1991), using a modified pKS2 plasmid (Mitsuyasu Hasebe Lab) from which the rice actin promoter was excised. The 5' genomic fragments upstream the stop codon (1616bp for *PpPERK10* and 1119bp for *PpZOITO*) were amplified using specific primers with restriction sites added (oMC01 / oMC02 and oMC007 / oMC08) and cloned by sticky end ligation into the KpnII/ClaI site in frame with the mGFP sequence. 3' sequences,

beginning with the stop codon of the gene (1620bp and 1163pb, respectively), were amplified using the primers oMC03 / oMC04 for *PERK10* and oMC09 / oMC010 for *ZOITO* and inserted at the NotI/SacII site. Polyethylene glycol-mediated transformation was performed (Schaefer and Zrÿd, 1997; Schaefer et al., 1991) using 15 µg of the linearized vector DNA and selection of stable transformants was conducted on G418 (50 µg/mL) antibiotic plates. The selected transformants were verified by PCR amplification using a specific primer for the 5' UTR gene sequence (oMC05 for *PERK10* and oMC11 for *ZOITO*) paired with a primer in the GFP sequence (oMC14), and the obtained fragment was sequenced to confirm in frame fusion. As control for tandem insertions, primers oMC14 and oMC15 were paired on a separate reaction (Supplemental figure 7). Two transgenic lines was obtained for *PpZOITO-GFP* (30) and (7), while eight lines were obtained for *PpPERK10-GFP* (2), (14), (42), (45), (53), (55), (66), and (71). Four of these were phenotypically characterized.

5.12 Phenotyping of GFP translational reporter lines

The translational reporter lines *PpZOITO-GFP* (30) and (7), as well as *PpPERK10-GFP* (2), (14), (42) and (53) were maintained and grown in the same conditions as described for the wild type. Induction of gametangia was conducted after 28 days of growth on sterile peat pellets (Jiffy-7, Jiffy Products International B.V.). At day 15 after cold induction observations of antheridia and sperm cells were conducted under a fluorescent microscope (Nikon Eclipse TE2000-S).

5.13 Transcriptome Age Index (TAI) estimation

The TAI calculation was done following a phylostratigraphic approach as in Domazet-loso and Tautz, (2003). Phylostratum ranks used to date the genes were estimated with PLAZA 2.5 gene families (Van Bel et al., 2012)and the gene families specific for green plants, land plants and *P. patens* were assigned to phylostratum rank of 1, 2 and 3, respectively. According with the formula used, the TAI of a tissue *s* is defined as weighted mean of phylostratum rank ps_i of gene *i* by the expression value e_{is} in the transcriptome of sample *s*:

$$TAI_s = \frac{\sum_{i=0}^{n} ps_i * e_{is}}{\sum_{i=0}^{n} e_{is}}$$

To estimate significance of observed TAI values for each tissue, the ps_i ranks of genes were permuted 1000 times. TAI values from the permutations ($TAI_{s(permuted)}$) were used to generate the boxplots shown on Figure 7 and to estimate the empirical P-value for each tissue *s* (P-value(TAI_s)), given by formula below:

$$\text{P-value}(TAI_s) = \frac{\sum_{i=0}^{1000} I\left(TAI_{s\,(observed)} \geq TAI_{s\,(permuted)}\right)}{1000}$$

where I(.) is indicator function.

6. ACKNOWLEDGEMENTS

We gratefully acknowledge Marek Mutwil for his valuable contribution for this work. João Sobral (IGC, Portugal) is acknowledged for his excellent technical assistance.

7. REFERENCES

Abbasi, N., Park, Y.-I., and Choi, S.-B. (2011). Pumilio Puf domain RNA-binding proteins in Arabidopsis. Plant Signal. Behav. *6*, 364–368.

Aggarwal, P., Yadav, R.K., and Reddy, G.V. (2010). Identification of novel markers for stem-cell niche of Arabidopsis shoot apex. Gene Expr. Patterns *10*, 259–264.

Alandete-Saez, M., Ron, M., Leiboff, S., and McCormick, S. (2011). Arabidopsis thaliana GEX1 has dual functions in gametophyte development and early embryogenesis. Plant J. *68*, 620–632.

Avendaño, C., Franchi, A., Jones, E., and Oehninger, S. (2009). Pregnancy-specific β-1-glycoprotein 1 and human leukocyte antigen-E mRNA in human sperm: Differential expression in fertile and infertile men and evidence of a possible functional role during early development. Hum. Reprod. *24*, 270–277.

Babushok, D. V., Ohshima, K., Ostertag, E.M., Chen, X., Wang, Y., Mandal, P.K., Okada, N., Abrams, C.S., and Kazazian, H.H. (2007). A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids. Genome Res. *17*, 1129–1138.

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. Genome Biol. *8*, 227.

Bartee, L., Malagnac, F., and Bender, J. (2001). Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev. *15*, 1753–1758.

Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinnel, T., and Lukowitz, W. (2009). Paternal control of embryonic patterning in Arabidopsis thaliana. Science *323*, 1485–1488.

Begun, D.J., Lindfors, H. a., Kern, A.D., and Jones, C.D. (2007). Evidence for de novo evolution of testis-expressed genes in the Drosophila yakuba/Drosophila erecta clade. Genetics *176*, 1131–1137.

Van Bel, M., Proost, S., Wischnitzki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., and Vandepoele, K. (2012). Dissecting Plant Genomes with the PLAZA Comparative Genomics Platform. PLANT Physiol. *158*, 590–600.

Von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). Arabidopsis HAP2 (GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. Development *133*, 4761–4769.

Blomstedt, C.K., Knox, R.B., and Singh, M.B. (1996). Generative cells of Lilium longiflorum possess translatable mRNA and functional protein synthesis machinery. Plant Mol. Biol. *31*, 1083–1086.

Boerke, a., Dieleman, S.J., and Gadella, B.M. (2007). A possible role for sperm RNA in early embryo development. Theriogenology *68*, 147–155.

Boothby, T.C., Zipper, R.S., Van der Weele, C.M., and Wolniak, S.M. (2013). Removal of Retained Introns Regulates Translation in the Rapidly Developing Gametophyte of Marsilea vestita. Dev. Cell *24*, 517–529.

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M., and Twell, D. (2011). The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. Plant Cell *23*, 534–549.

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.

Borges, F., Calarco, J.P., and Martienssen, R. a. (2012). Reprogramming the epigenome in Arabidopsis pollen. Cold Spring Harb. Symp. Quant. Biol. *77*, 1–5.

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T., and Twell, D. (2009). A plant germline-specific integrator of sperm specification and cell cycle progression. PLoS Genet. *5*.

Brzeski, J., and Jerzmanowski, A. (2003). Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. J. Biol. Chem. *278*, 823–828.

Calarco, J.P., Borges, F., Donoghue, M.T. a, Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J. a., Becker, J.D., et al. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell *151*, 194–205.

Calvel, P., Rolland, A.D., Jégou, B., and Pineau, C. (2010). Testicular postgenomics: targeting the regulation of spermatogenesis. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *365*, 1481–1500.

Cao, Q., Padmanabhan, K., and Richter, J.D. (2010). Pumilio 2 controls translation by competing with eIF4E for 7-methyl guanosine cap recognition. RNA *16*, 221–227.

Cao, X., Springer, N.M., Muszynski, M.G., Phillips, R.L., Kaeppler, S., and Jacobsen, S.E. (2000). Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. Proc. Natl. Acad. Sci. U. S. A. *97*, 4979–4984.

Chen, T., Ueda, Y., Xie, S., and Li, E. (2002). A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J. Biol. Chem. *277*, 38746–38754.

Chen, X., Lu, C., Prado, J.R.M., Eun, S.H., and Fuller, M.T. (2011). Sequential changes at differentiation gene promoters as they become active in a stem cell lineage. Development *138*, 2441–2450.

Chritton, J.J., and Wickens, M. (2010). Translational repression by PUF proteins in vitro. RNA *16*, 1217–1225.

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

Cove, D.J., and Knight, C.D. (1993). The Moss Physcomitrella-Patens, a Model System with Potential for the Study of Plant Reproduction. Plant Cell *5*, 1483–1488.

Crittenden, S.L., Bernstein, D.S., Bachorik, J.L., Thompson, B.E., Gallegos, M., Petcherski, A.G., Moulder, G., Barstead, R., Wickens, M., and Kimble, J. (2002). A conserved RNA-binding protein controls germline stem cells in Caenorhabditis elegans. Nature *417*, 660–663.

Cui, X., Lv, Y., Chen, M., Nikoloski, Z., Twell, D., and Zhang, D. (2014). Young Genes out of the Male: An Insight from Evolutionary Age Analysis of the Pollen Transcriptome. Mol. Plant 1–11.

Dadoune, J.P. (2009). Spermatozoal RNAs: What about their functions? Microsc. Res. Tech. *72*, 536–551.

Dangwal, M., Kapoor, S., and Kapoor, M. (2014). The PpCMT chromomethylase affects cell growth and interacts with the homolog of LIKE HETEROCHROMATIN PROTEIN 1 in the moss Physcomitrella patens. Plant J. *77*, 589–603.

Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A.J.M., and Matzke, M. (2009). A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. EMBO J. *28*, 48–57.

Deveaux, Y., Toffano-Nioche, C., Claisse, G., Thareau, V., Morin, H., Laufs, P., Moreau, H., Kreis, M., and Lecharny, A. (2008). Genes of the most conserved WOX clade in plants affect root and flower development in Arabidopsis. BMC Evol. Biol. *8*, 291.

Domazet-loso, T., and Tautz, D. (2003). An Evolutionary Analysis of Orphan Genes in Drosophila An Evolutionary Analysis of Orphan Genes in Drosophila. Genome Res. 2213–2219.

Durbarry, A., Vizir, I., and Twell, D. (2005). Male germ line development in Arabidopsis. duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. Plant Physiol. *137*, 297–307.

Ekici, N., and Dane, F. (2004). Polarity during sporogenesis and gametogenesis in plants. PoLAR 687–696.

Emerson, J.J., Kaessmann, H., Betrán, E., and Long, M. (2004). Extensive gene traffic on the mammalian X chromosome. Science *303*, 537–540.

Engel, M.L., Chaboud, A., Dumas, C., and McCormick, S. (2003). Sperm cells of Zea mays have a complex complement of mRNAs. Plant J. *34*, 697–707.

Fischer, B.E., Wasbrough, E., Meadows, L. a, Randlet, O., Dorus, S., Karr, T.L., and Russell, S. (2012). Conserved properties of Drosophila and human spermatozoal mRNA repertoires. Proc. R. Soc. B Biol. Sci.

Forbes, a, and Lehmann, R. (1998). Nanos and Pumilio have critical roles in the development and function of Drosophila germline stem cells. Development *125*, 679–690.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C., et al. (2013). STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. *41*.

Francischini, C.W., and Quaggio, R.B. (2009). Molecular characterization of Arabidopsis thaliana PUF proteins--binding specificity and target candidates. FEBS J. *276*, 5456–5470.

Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H., and Ohme-Takagi, M. (2000). Arabidopsis Ethylene-Responsive Element Binding Factors Act as Transcriptional Activators or Repressors of GCC Box–Mediated Gene Expression. Plant Cell *12*, 393–405.

Gapp, K., Jawaid, A., Sarkies, P., Bohacek, J., Pelczar, P., Prados, J., Farinelli, L., Miska, E., and Mansuy, I.M. (2014). Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. Nat. Neurosci. *17*, 667–669.

Gatewood, J., Cook, G., Balhorn, R., Bradbury, E., and Schmid, C. (1987). Sequence-specific packaging of DNA in human sperm chromatin. Science (80-.). *236*, 962–964.

Goldstrohm, A.C., Seay, D.J., Hook, B.A., and Wickens, M. (2007). PUF protein-mediated deadenylation is catalyzed by Ccr4p. J. Biol. Chem. *282*, 109–114.

Grivna, S.T., Beyret, E., Wang, Z., and Lin, H. (2006). RESEARCH COMMUNICATION A novel class of small RNAs in mouse spermatogenic cells. Genes Dev. *20*, 1709–1714.

Grunewald, S., Paasch, U., Glander, H.J., and Anderegg, U. (2005). Mature human spermatozoa do not transcribe novel RNA. Andrologia *37*, 69–71.

Gu, W., Kwon, Y.K., and Hecht, N.B. (1996). In postmeiotic male germ cells poly (A) shortening accompanies translation of mRNA encoding ?? enteric actin but not cytoplasmic ?? and ?? actin mRNAs. Mol. Reprod. Dev. *44*, 141–145.

Gur, Y., and Breitbart, H. (2006). Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes. Genes Dev. *20*, 411–416.

Hecht, N.B. (1998). Molecular mechanisms of male germ cell differentiation. Bioessays *20*, 555–561.

Heinen, T.J.A.J., Staubach, F., Häming, D., and Tautz, D. (2009). Emergence of a New Gene from an Intergenic Region. Curr. Biol. *19*, 1527–1531.

Huala, E., and Sussex, I. (1993). Determination and Cell Interactions in Reproductive Meristems. Plant Cell *5*, 1157–1165.

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

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Ji, Z., Lee, J.Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3’ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc. Natl. Acad. Sci. U. S. A. *106*, 7028–7033.

Johnson, G.D., Sendler, E., Lalancette, C., Hauser, R., Diamond, M.P., and Krawetz, S. a. (2011). Cleavage of rRNA ensures translational cessation in sperm at fertilization. Mol. Hum. Reprod. *17*, 721–726.

Jullien, P.E., Susaki, D., Yelagandula, R., Higashiyama, T., and Berger, F. (2012). DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr. Biol. *22*, 1825–1830.

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Res. *20*, 1313–1326.

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.

Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L., and Richards, E.J. (2003). Arabidopsis MET1 cytosine methyltransferase mutants. Genetics *163*, 1109–1122.

Kierszenbaum, A., and Tres, L.L. (1975). Structural and transcriptional features of the mouse spermatid genome. J. Cell Biol. *65*, 258–270.

Klucher, K.M., Chow, H., Reiser, L., and Fischer, R.L. (1996). The AINTEGUMENTA gene of Arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. Plant Cell *8*, 137–153.

Krawetz, S. a, Kruger, A., Lalancette, C., Tagett, R., Anton, E., Draghici, S., and Diamond, M.P. (2011). A survey of small RNAs in human sperm. Hum. Reprod. *26*, 3401–3412.

Lalancette, C., Platts, A.E., Johnson, G.D., Emery, B.R., Carrell, D.T., and Krawetz, S. a. (2009). Identification of human sperm transcripts as candidate markers of male fertility. J. Mol. Med. *87*, 735–748.

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*, 1406–1419.

Laubichler, M.D., and Davidson, E.H. (2008). Boveri's long experiment: Sea urchin merogones and the establishment of the role of nuclear chromosomes in development. Dev. Biol. *314*, 1–11.

Laux, T., Mayer, K.F., Berger, J., and Jürgens, G. (1996). The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. Development *122*, 87–96.

Levine, M.T., Jones, C.D., Kern, A.D., Lindfors, H. a, and Begun, D.J. (2006). Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression. Proc. Natl. Acad. Sci. U. S. A. *103*, 9935–9939.

Lewis, S.E.M. (2007). Is sperm evaluation useful in predicting human fertility? Reproduction *134*, 31–40.

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.

Liu, Y., Tewari, R., Ning, J., Blagborough, A.., Garbom, S., Pei, J., Grishin, N.., Steele, R.., Sinden, R.., Snell, W.., et al. (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in Chlamydomonas and Plasmodium gametes. Genes Dev. *22*, 1051–1068.

Long, M., and Langley, C.H. (1993). Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila. Science *260*, 91–95.

MacLaughlin, J., and Terner, C. (1973). Ribonucleic acid synthesis by spermatozoa from the rat and hamster. Biochem. J. *133*, 635–639.

Malik, G., Dangwal, M., Kapoor, S., and Kapoor, M. (2012). Role of DNA methylation in growth and differentiation in Physcomitrella patens and characterization of cytosine DNA methyltransferases. FEBS J. *279*, 4081–4094.

Martins, R.P., and Krawetz, S. a. (2005). RNA in human sperm. Asian J. Androl. *7*, 115–120.

Martín-Trillo, M., and Cubas, P. (2010). TCP genes: a family snapshot ten years later. Trends Plant Sci. *15*, 31–39.

Mayer, K.F.X., Schoof, H., Haecker, A., Lenhard, M., Jürgens, G., and Laux, T. (1998). Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem. Cell *95*, 805–815.

McCormick, S. (1993). Male Gametophyte Development. Plant Cell *5*, 1265–1275.

Miller, D. (2014). Sperm RNA as a Mediator of Genomic Plasticity. *2014*.

Miller, D., Ostermeier, G.C., and Krawetz, S. a. (2005). The controversy, potential and roles of spermatozoal RNA. Trends Mol. Med. *11*, 156–163.

Mizukami, Y., and Fischer, R.L. (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. Proc. Natl. Acad. Sci. U. S. A. *97*, 942–947.

Moore, F.L., Jaruzelska, J., Fox, M.S., Urano, J., Firpo, M.T., Turek, P.J., Dorfman, D.M., and Pera, R.A.R. (2003). Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. Proc. Natl. Acad. Sci. U. S. A. *100*, 538–543.

Murata, Y., and Wharton, R.P. (1995). Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in Drosophila embryos. Cell *80*, 747–756.

Noy-Malka, C., Yaari, R., Itzhaki, R., Mosquna, A., Auerbach Gershovitz, N., 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*, 719–735.

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*, 3567–3581.

Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., and Krawetz, S. a. (2002). Spermatozoal RNA profiles of normal fertile men. Lancet *360*, 772–777.

Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamond, M.P., and Krawetz, S.A. (2004). Reproductive biology: delivering spermatozoan RNA to the oocyte. Nature *429*, 154.

Palermo, G.D., Colombero, L.T., and Rosenwaks, Z. (1997). The human sperm centrosome is responsible for normal syngamy and early embryonic development. Rev. Reprod. *2*, 19–27.

Parisi, M., and Lin, H. (1999). The Drosophila pumilio gene encodes two functional protein isoforms that play multiple roles in germline development, gonadogenesis, oogenesis and embryogenesis. Genetics *153*, 235–250.

Perroud, P.F., Cove, D.J., Quatrano, R.S., and Mcdaniel, S.F. (2011). An experimental method to facilitate the identification of hybrid sporophytes in the moss Physcomitrella patens using fluorescent tagged lines. New Phytol. *191*, 301–306.

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. Plant Physiol. *138*, 744–756.

Premkumar, E., and Bhargava, P.M. (1972). Transcription and translation in bovine spermatozoa. Nat. New Biol. *240*, 139–143.

Qin, Y., Ma, X., Yu, G., Wang, Q., Wang, L., Kong, L., Kim, W., and Wang, H.W. (2014). Evolutionary History of Trihelix Family and Their Functional Diversification. DNA Res. 1–12.

Rejon, E., Bajon, C., Blaize, A., and Robert, D. (1988). RNA in the nucleus of a motile plant spermatozoid: characterization by enzyme-gold cytochemistry and in situ hybridization. Mol. Reprod. Dev. *1*, 49–56.

Rensing, S. a, Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E. a, Kamisugi, Y., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science *319*, 64–69.

Renzaglia, K.S., and Garbary, D.J. (2001). Motile Gametes of Land Plants: Diversity, Development, and Evolution. CRC. Crit. Rev. Plant Sci. *20*, 107–213.

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.

Reynolds, W.F., and Wolfe, S.L. (1984). Protamines in plant sperm. Exp. Cell Res. *152*, 443–448.

Rotman, N., Durbarry, A., Wardle, A., Yang, W.C., Chaboud, A., Faure, J.E., Berger, F., and Twell, D. (2005). A novel class of MYB factors controls sperm-cell formation in plants. Curr. Biol. *15*, 244–248.

Russell, S.D., Gou, X., Wong, C.E., Wang, X., Yuan, T., Wei, X., Bhalla, P.L., and Singh, M.B. (2012). Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. New Phytol. *195*, 560–573.

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science *320*, 1643–1647.

Sassone-Corsi, P. (2002). Unique chromatin remodeling and transcriptional regulation in spermatogenesis. Science *296*, 2176–2178.

Saunders, C., Swann, K., and Lai, F. (2007). PLCzeta, a sperm-specific PLC and its potential role in fertilization. Biochem. Soc. Symp. *74*, 23–36.

Saunders, C.M., Larman, M.G., Parrington, J., Cox, L.J., Royse, J., Blayney, L.M., Swann, K., and Lai, F.A. (2002). PLC zeta: a sperm-specific trigger of Ca(2+) oscillations in eggs and embryo development. Development *129*, 3533–3544.

Schaefer, D.G., and Zrÿd, J.P. (1997). Efficient gene targeting in the moss Physcomitrella patens. Plant J. *11*, 1195–1206.

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

Schoft, V.K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J.-S., Park, G.T., et al. (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. Proc. Natl. Acad. Sci. U. S. A. *108*, 8042–8047.

Schultz, E., and Haughn, G. (1991). LEAFY, a Homeotic Gene That Regulates Inflorescence Development in Arabidopsis. Plant Cell *3*, 771–781.

Sendler, E., Johnson, G.D., Mao, S., Goodrich, R.J., Diamond, M.P., Hauser, R., and Krawetz, S.A. (2013). Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Res. *41*, 4104–4117.

Shaha, C., Tripathi, R., and Mishra, D.P. (2010). Male germ cell apoptosis: regulation and biology. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *365*, 1501–1515.

Sharma, V.K., Ramirez, J., and Fletcher, J.C. (2003). The Arabidopsis CLV3-like (CLE) genes are expressed in diverse tissues and encode secreted proteins. Plant Mol. Biol. *51*, 415–425.

Snook, R.R. (2005). Sperm in competition: Not playing by the numbers. Trends Ecol. Evol. *20*, 46–53.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., et al. (2013). Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. Cell Rep. *3*, 2179–2190.

Stoeckius, M., Grün, D., and Rajewsky, N. (2014). Paternal RNA contributions in the Caenorhabditis elegans zygote. EMBO J. *33*, 1740–1750.

Takeuchi, T., Watanabe, Y., Takano-Shimizu, T., and Kondo, S. (2006). Roles of jumonji and jumonji family genes in chromatin regulation and development. Dev. Dyn. *235*, 2449–2459.

Tan, Q.K., and Irish, V.F. (2006). The Arabidopsis Zinc Finger-Homeodomain Genes Encode Proteins with Unique Biochemical Properties That Are Coordinately Expressed during Floral Development 1. *140*, 1095–1108.

Terada, Y., Nakamura, S.I., Simerly, C., Hewitson, L., Murakami, T., Yaegashi, N., Okamura, K., and Schatten, G. (2004). Centrosomal Function Assessment in Human Sperm Using Heterologous ICSI with Rabbit Eggs: A New Male Factor Infertility Assay. Mol. Reprod. Dev. *67*, 360–365.

Ulitsky, I., Maron-Katz, A., Shavit, S., Sagir, D., Linhart, C., Elkon, R., Tanay, A., Sharan, R., Shiloh, Y., and Shamir, R. (2010). Expander: from expression microarrays to networks and functions. Nat. Protoc. *5*, 303–322.

Vermeulen, J., Derveaux, S., Lefever, S., De Smet, E., De Preter, K., Yigit, N., De Paepe, A., Pattyn, F., Speleman, F., and Vandesompele, J. (2009). RNA pre-amplification enables large-scale RT-qPCR gene-expression studies on limiting sample amounts. BMC Res. Notes *2*, 235.

Vibranovski, M.D., Chalopin, D.S., Lopes, H.F., Long, M., and Karr, T.L. (2010). Direct evidence for postmeiotic transcription during Drosophila melanogaster spermatogenesis. Genetics *186*, 431–433.

Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. (1993). Arabidopsis thaliana DNA methylation mutants. Science *260*, 1926–1928.

Watanabe, D., Suetake, I., Tajima, S., and Hanaoka, K. (2004). Expression of Dnmt3b in mouse hematopoietic progenitor cells and spermatogonia at specific stages. Gene Expr. Patterns *5*, 43–49.

Wedell, N., Gage, M.J.G., and Parker, G. a. (2002). Sperm competition, male prudence and sperm-limited females. Trends Ecol. Evol. *17*, 313–320.

Wong, J.L., and Johnson, M.A. (2010). Is HAP2-GCS1 an ancestral gamete fusogen? Trends Cell Biol. *20*, 134–141.

Wreden, C., Verrotti, a C., Schisa, J. a, Lieberfarb, M.E., and Strickland, S. (1997). Nanos and pumilio establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. Development *124*, 3015–3023.

Wu, D.-D., Wang, X., Li, Y., Zeng, L., Irwin, D.M., and Zhang, Y.-P. (2014). “Out of Pollen” Hypothesis for Origin of New Genes in Flowering Plants: Study from Arabidopsis thaliana. Genome Biol. Evol. *6*, 2822–2829.

Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The arabidopsis nucleosome remodeler DDM1 allows DNA

methyltransferases to access H1-containing heterochromatin. Cell *153*, 193–205.

Zhao, G.Q., and Garbers, D.L. (2002). Male germ cell specification and differentiation. Dev. Cell *2*, 537–547.

Zhao, Y., Li, Q., Yao, C., Wang, Z., Zhou, Y., Wang, Y., Liu, L., Wang, Y., Wang, L., and Qiao, Z. (2006). Characterization and quantification of mRNA transcripts in ejaculated spermatozoa of fertile men by serial analysis of gene expression. Hum. Reprod. *21*, 1583–1590.

Zimmer, A.D., Lang, D., Buchta, K., Rombauts, S., Nishiyama, T., Hasebe, M., Van de Peer, Y., Rensing, S. a, and Reski, R. (2013). Reannotation and extended community resources for the genome of the non-seed plant Physcomitrella patens provide insights into the evolution of plant gene structures and functions. BMC Genomics *14*, 498.

8. SUPPLEMENTAL DATA

Supplemental Figure 1. RNA profile of antherozoid and antheridia samples.

Supplemental Figure 2. qRT-PCR to assess male gametogenesis sample quality.

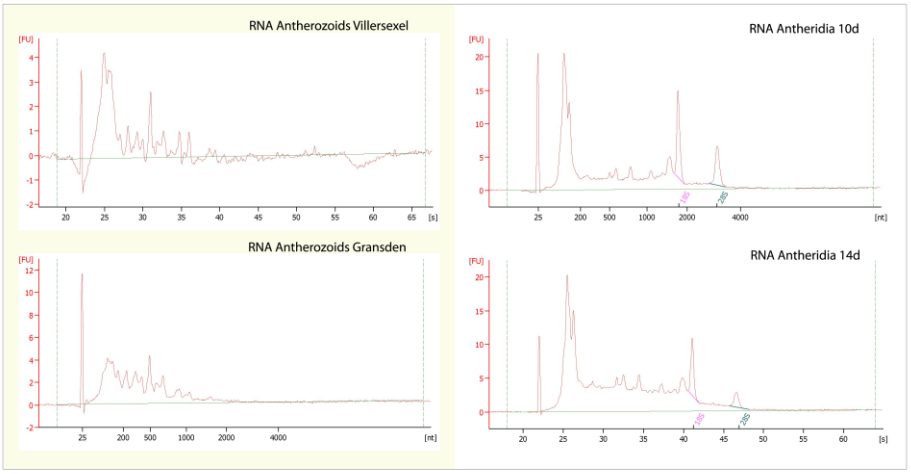
Supplemental Figure 3. Principal Component Analysis of Physcomitrella patens tissues excluding the antherozoids.

Supplemental Figure 4. Transcript presence / absence threshold determination by qRT-PCR.

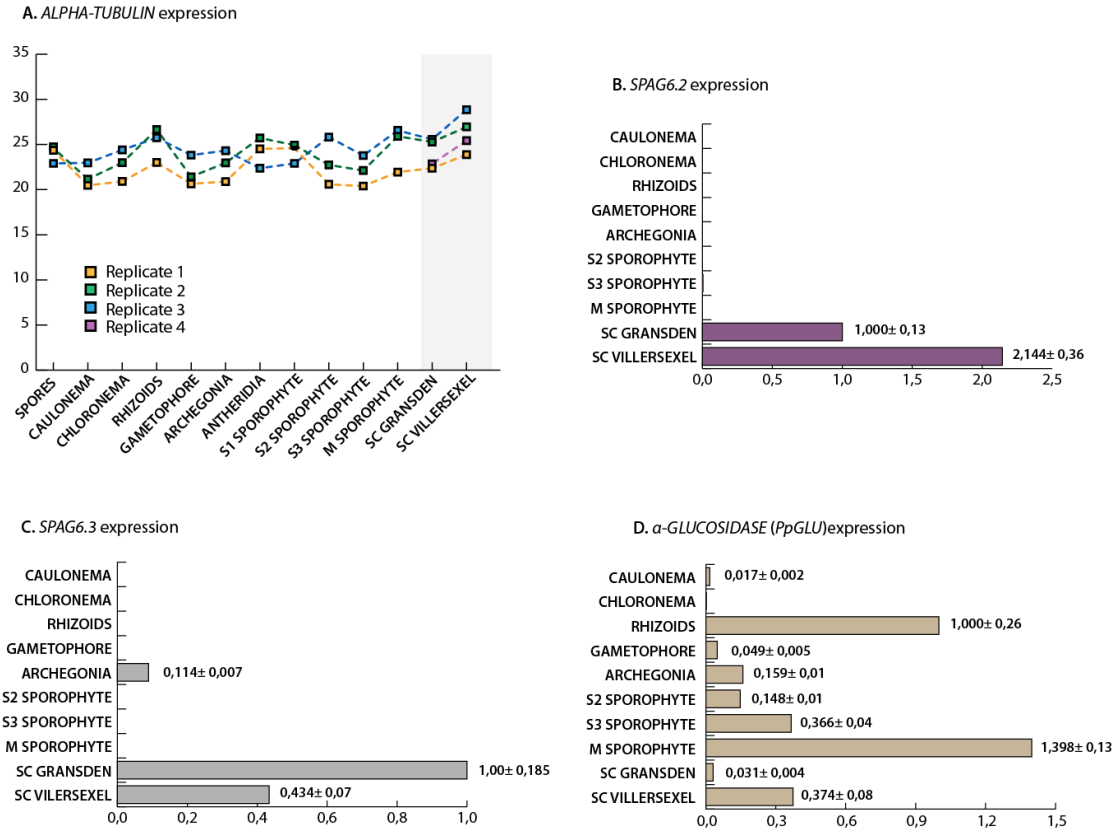
Supplemental Figure 5. PpDUO1 phylogenetic analysis.

Supplemental Figure 6. Enriched genes shared among male gametogenesis stages.

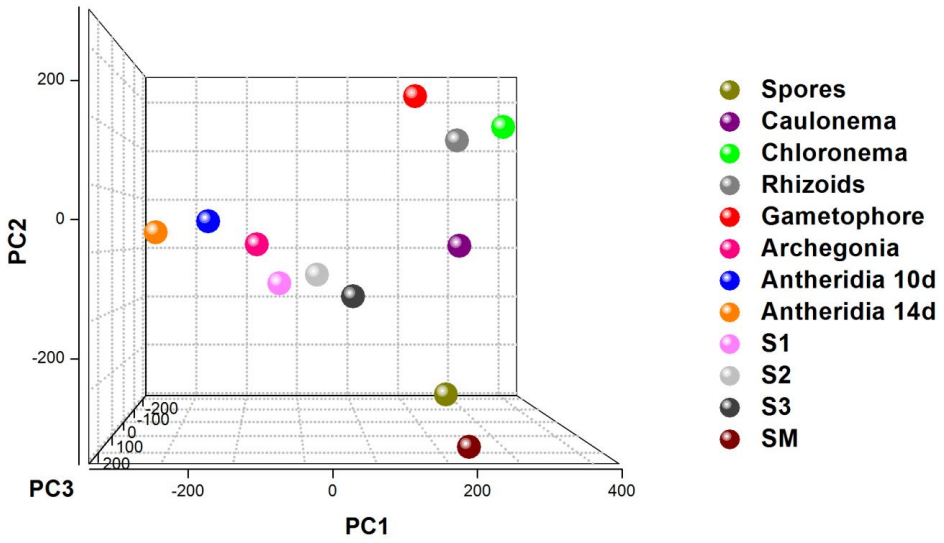
Supplemental Figure 7. Genotyping of the GFP translational reporter lines.



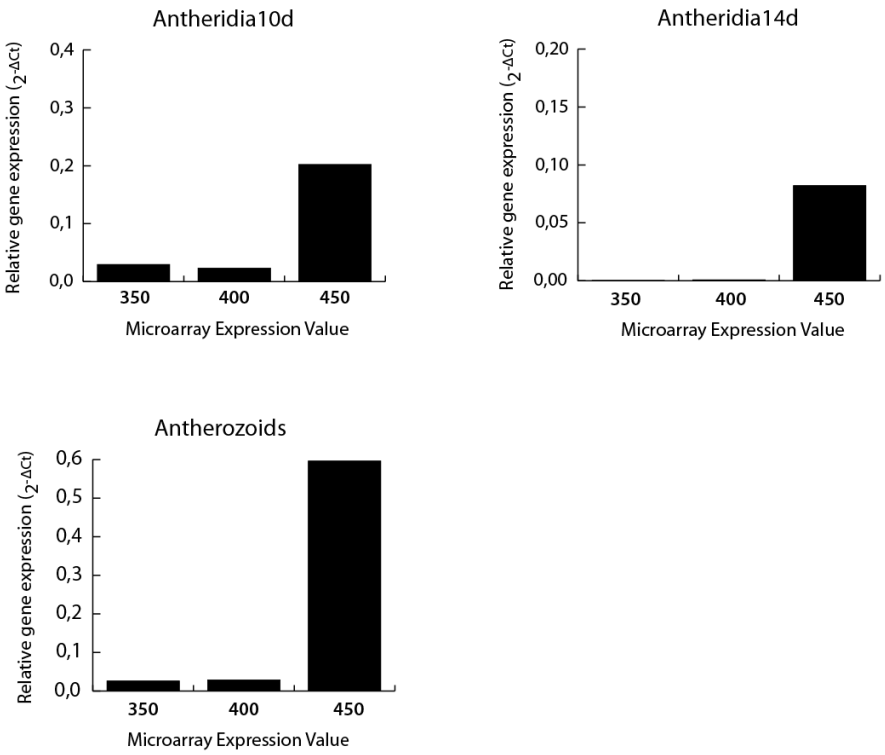
Supplemental Figure 1. RNA profile of antherozoid and antheridia samples.
The electrophoretic profile of total RNA from antherozoids, immature antheridia (10d) and mature antheridia (14d) was assessed on an Agilent Bioanalyzer. The profiles of antherozoid samples miss the two peaks corresponding to 18S and 28S rRNA, while in the antheridia samples these peaks are clearly identified.



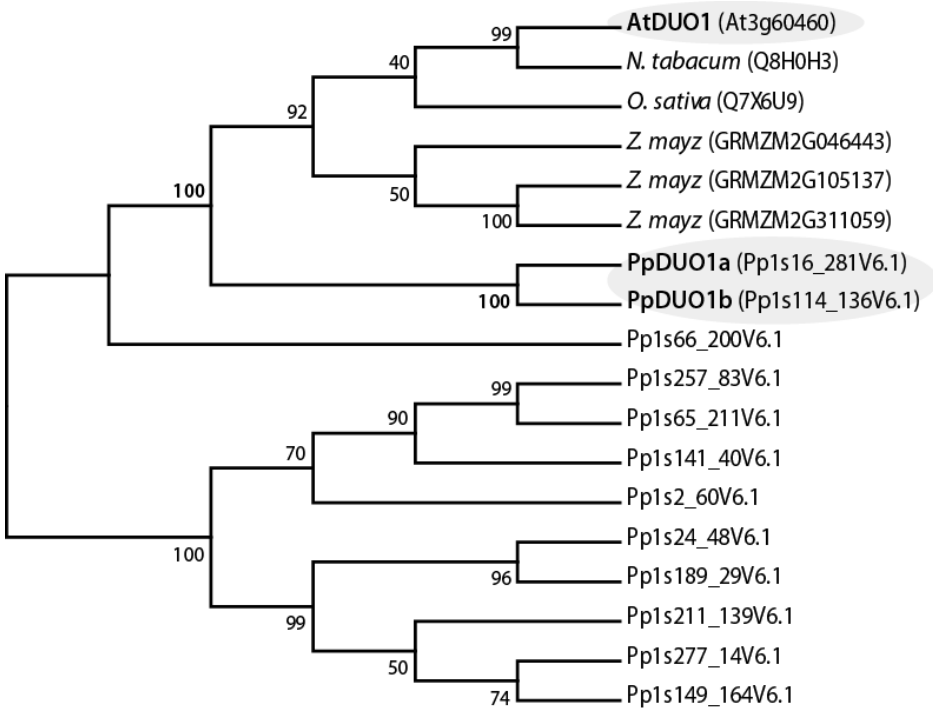
Supplemental figure 2. qRT-PCR to assess male gametogenesis sample quality. Quality of tissue sample replicates was tested by qRT-PCR using marker genes. **A.** The expression of the housekeeping gene *ALPHA-TUBULIN* (Pp1s215_51V6.1) in the male gametogenesis tissues was evaluated. In order to confirm that similar quantities were obtained during cDNA synthesis, the expression levels were compared with other previously assessed *P. patens* tissues. The sample replicates show low variation in *ALPHA-TUBULIN* expression levels. **B.** The expression of the *SPERM ASSOCIATED ANTIGEN 6* (*SPAG6*) homolog genes in *P. patens* was assessed. The highly conserved *SPAGs* are associated to the flagella of male gametes. In accordance, the *PpSPAG6.2* (Pp1s344_19V6.1) is restricted to the antherozoids **C.** Similarly, *PpSPAG6.3* (Pp1s344_16V6.1) expression was detected in antherozoids, but in this case expression was also observed in the archegonia. **D.** The *α-GLUCOSIDASE* gene *PpGLU* (Pp1s25_97V6.1), preferentially expressed in rhizoids according to Ishikawa et al. (2011), was tested. In *Gransden* antherozoids low levels were detected, while in *Villersexel* a higher expression was observed. The standard error range (±) is depicted next to the bars.



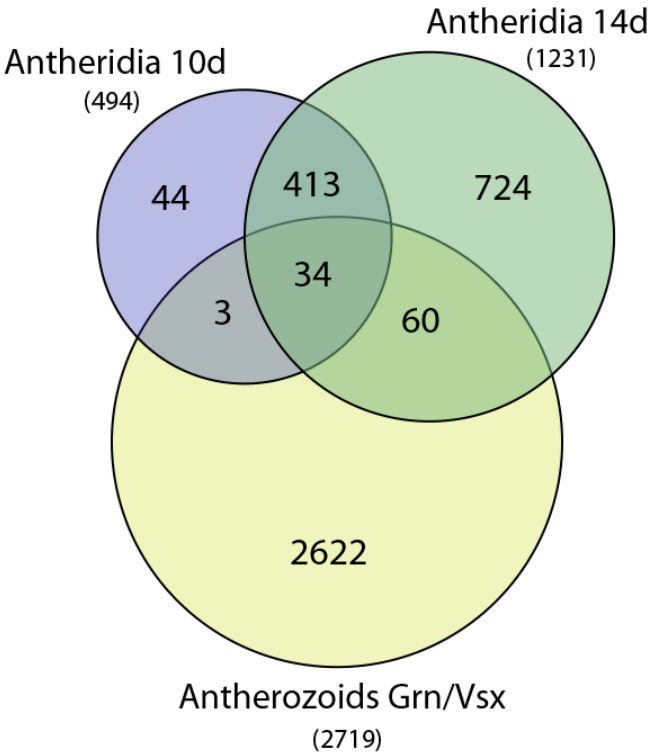
Supplemental Figure 3. Principal Component Analysis of *Physcomitrella patens* tissues excluding the antherozoids.
When *Physcomitrella* samples were plotted without the antherozoids the constriction imposed by those samples was removed, allowing a better separation of the other reproductive organs. The antheridia samples and archegonia are displaced from the rest of the tissues by PC1 and PC3, but while the archegonia is closer to the pre-meiotic sporophyte samples, the antheridia showed a stronger separation from other tissues.



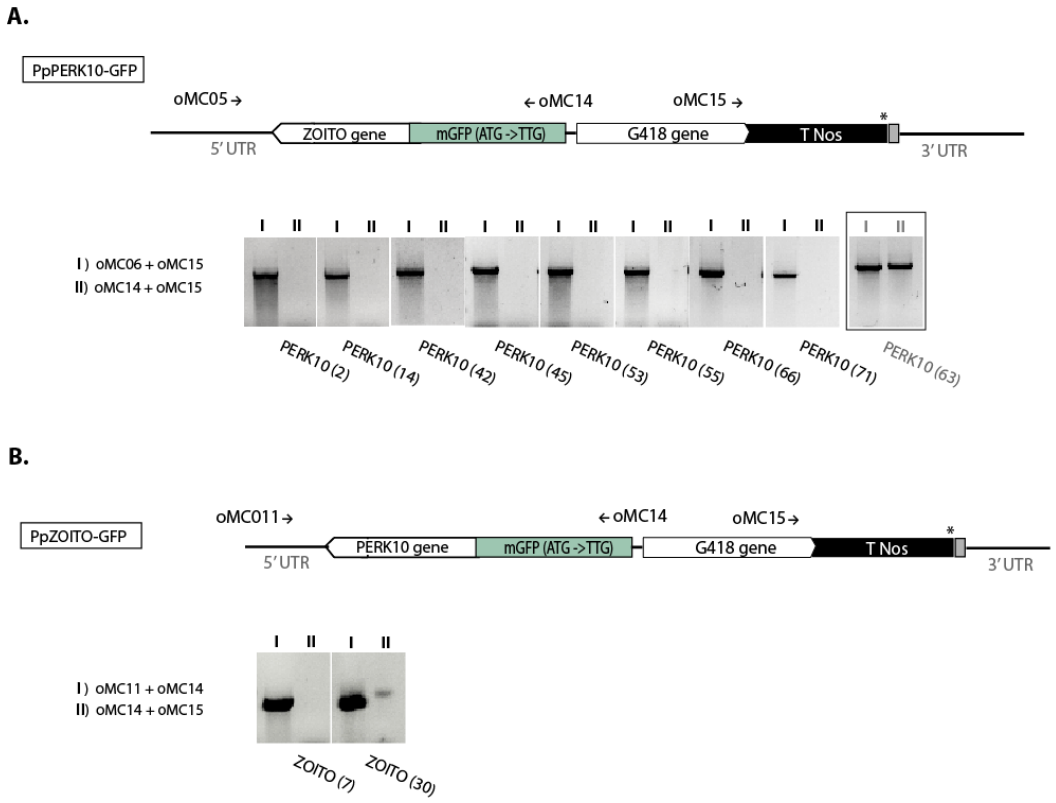
Supplemental figure 4. Transcript presence / absence threshold determination by qRT-PCR. The experiments to determine transcript presence / absence signal thresholds were conducted using excess cDNA synthetized for microarray hybridization from the three male gametogenesis related tissues. Based on previous threshold determinations for the rest of *P. patens* tissues (Chapter II), a threshold value between 300 and 450 was predicted. Three genes per tissue with relative expression values corresponding to 350, 400 and 450 were tested. From these, the ones with a relative expression of 450 were always detected by qRT-PCR.



Supplemental Figure 5. PpDUO1 phylogenetic analysis.
The phylogenic relationships of *PpDUO1* putative orthologs were inferred using the Maximum Likelihood method. The protein sequence of *A. thaliana* (*AtDUO1*) and highly similar sequences from *N. tabacum*, *O. sativa* and *Z. mays* were aligned to the twelve sequences from *P. patens*. In the unrooted tree shown, Pp1s16_281V6.1 and Pp1s114_136V6.1 (named *PpDUO1a* and *PpDUO1b*) display a close relationship to the angiosperm sequences.



Supplemental figure 6. Enriched genes shared among male gametogenesis stages.
From the 3900 enriched genes identified for the male gametogenesis related tissues, the common elements among them are shown as a Venn diagram.



Supplemental Figure 7. Genotyping of the GFP translational reporter lines. The selected transformants for *ZOITO-GFP* and *PERK10-GFP* were verified by PCR amplification. **A.** Specific primers corresponding to the 5'UTR of the PERK10 gene (oMC05) and a primer in the GFP sequence (oMC14) were paired on a PCR reaction, and the obtained fragment was sequenced to confirm in frame fusion. As control for tandem insertions, primers oMC14 and oMC15 were used on a separate reaction. Eight lines were obtained for *PpPERK10-GFP* (2), (14), (42), (45), (53), (55), (66), and (71). Line (63) is presented as an example of more than one insertion. **B.** Similarly, for the *ZOITO-GFP* lines, the oMC11primer was paired with the oMC14 to amplify the 5'UTR of the gene together with the GFP sequence. Verification of more than one insertion was done using primers oMC14 and oMC15. Two transgenic lines were obtained for *PpZOITO-GFP,* (30) and (7).

GENERAL DISCUSSION

1. Evolution in the life cycle of land plants

As a result of the adaptation to land, the sporophyte and the gametophyte underwent anatomical and morphological changes, generating the dimorphic alternation of generations observed in Embryophytes (Graham et al., 2000; Kenrick and Crane, 1997). The transition from haploid dominant to diploid dominant life cycles involved genetic changes that we are just starting to unravel. The acquisition of new gene functions, as evidenced by the *KNOX* transcription factors (Sakakibara et al., 2008; Singer and Ashton, 2007), and the co-option of genes from the gametophyte generation exemplified by the *RSL* genes (Menand et al., 2007) are postulated to be key during this transition.

The comparative analysis of gene expression in the bryophyte *Physcomitrella patens* generated in this work offers insights on the genetic processes that may have been required for the evolution and development of land plants. Based on our data, the total number of genes expressed in the distinct tissues analyzed indicates a relatively constant number (ranging from 50% to 59%) despite gametophytic or sporophytic sample identity. Similarly, in *Funaria hygrometrica* (Szövényi et al., 2011) differences in total number of expressed genes among the two phases are smaller than in the angiosperm *Arabidopsis thaliana* ,where the gametophyte generation expresses less than half the genes than the sporophyte (Loraine et al., 2013). This supports the idea that during evolution generation-biased gene expression was acquired, and that the sporophyte gained major complexity while the gametophyte was reduced. From our work, two lines of evidence for the gain of complexity in the sporophyte during land plant evolution are emerging.

1.1 Genes important for sporophyte development in Angiosperms show conserved expression patterns in *P. patens*

In Angiosperms, several transcription factor gene families have been identified to be important for the regulation of sporophyte development such as *KNOX*, *BELL1*, *SCARECROW* and *TCP* (Bowman et al., 2007; Floyd and Bowman, 2007; Graham et al., 2000). Interestingly, in *P. patens* several homolog genes of those showed preferential expression in the sporophyte. In our analysis, expression of the *KNOX2* genes *MKN*1 (Pp1s154_83V6.1) and *MKN*6 (Pp1s77_59V6.1) was observed throughout sporophyte development. In accordance, loss-of-function mutants on *KNOX2* genes result in defects during sporophyte development (Sakakibara et al., 2008). The identification of a similar expression pattern for the *BELL1* homolog (Pp1s258_6V6.1) suggests a possible interaction among the two *TALE* genes. Both in *Arabidopsis* and in *Chlamydomonas* dimers of *KNOX* and *BELL1* have roles in the diploid generation (Hay and Tsiantis, 2010; Lee et al., 2008). Interestingly, our data indicated that the two *TCP class II* orthologs identified (Pp1s332_35v6.1 & Pp1s207_110v6.1) have restricted expression in the sporophyte, with higher levels occurring at the mature stage. And with the experimental evidences generated for the *PpTCP5* gene, the conserved roles of *class II TCP* genes in auxiliary branch repression are further demonstrated. In the three different independent *Pptcp5* mutant lines, an increased frequency of “branched” sporophytes was observed (>12%) with respect to the wild type. On those, more than one capsule was observed to be attached to the same seta, probably as result of defective meristem repression. Overall, these findings indicate that genes important for the control of sporophyte development and for the regulation of its architecture may have been present already in the early land plants.

Subsequent acquisition of more complex roles may at least in part account for the evolution of the sporophyte body plan.

The expression of the *REDUCED VERNALIZATION RESPONSE 1* (*VRN1*) (Pp1s434_19V6.1 & Pp1s65_254V6) gene homologs early during sporophyte development, may indicate a complex regulation of sporophyte meristematic cells in the moss. In *Arabidopsis, VRN1* levels increase in response to cold conditions negatively regulating the levels of the floral repressor *FLOWERING LOCUS C* (*FLC*) (Levy et al., 2002). Low levels of *FLC* allow the expression of the downstream targets *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF CONSTANS OVEREXPRESSION 1* (*SOC1*), triggering the change from a vegetative to a reproductive meristem (Crevillén and Dean, 2011). Since this gene is expressed prior spore formation in the sporophyte of *P. patens*, it is possible that *VRN1* genes play a role in the meristematic transition to generate the cells that will give rise to the spores. During evolution, *VRN1* could have acquired functions in meristem reproductive fate transitions. Since mosses require cold and short day conditions to trigger the formation of reproductive organs, a conserved function in the apical meristem cannot be discarded.

The recruitment of genes present in the moss sporophyte, such as the *SQUAMOSA* (*SPL*), *NUBBIN* (*NUB*) and *JAGGED* (*JAG*) transcription factors, for roles in the regulation of the sporophyte body plan represents another layer in sporophyte gene evolution. Those genes are important for organ shape in Angiosperms once the organ is formed (Dinneny et al., 2006; Preston and Hileman, 2010; Zhang and Li, 2013). Since they cannot trigger the formation of the organ, it is hypothesized that co-option of those together with the evolution of

new gene networks can account for the increase in diversity of sporophyte shape.

1.2 Tip growth related genes indicate the recruitment of gametophyte specific transcripts for sporophyte functions

The reduction of the gametophyte generation during land plant evolution may have been accompanied by a loss of gene functions. Correspondingly, the sporophyte generation would have acquired some of those functions. The *OVATE* (*OFP*) transcription factor family represents an example of this process. In *P. patens*, the genes belonging to this family showed restricted gene expression in tip growing cells. In *A. thaliana*, however, members of that family have been shown to control sporophyte developmental processes, including the growth of aerial organs (Wang et al., 2007) vascular tissue formation (Li et al., 2011) and the development of hypocotyls, leaves and siliques (Wang et al., 2011). Moreover, it has been shown that *OVATE* members can regulate *KNOX* and *BELL1* families, emphasizing their roles in the diploid generation (Li et al., 2011; Wang et al., 2011).

The study of gametophytic expressed genes in *P. patens* revealed possible examples of generation-biased gene function acquisition. In our analysis, genes identified to be important for tip growth processes showed common expression among the three tissues that present polarized tip growth in *Physcomitrella*: the caulonema, the chloronema and the rhizoids. Among those, members of the *bHLH* and the *MICK* MADS-box* transcription factor families were identified. In *Arabidopsis*, the homologous genes showed expression restricted to one of the two tip growing cell types. While the *bHLH* genes *RSL1* and *RSL2* are only expressed in root hairs (Menand et al., 2007), the

*MICK** genes are predominantly expressed in pollen tubes (Adamczyk and Fernandez, 2009; Verelst et al., 2007). This suggests that during evolution, the functions of those genes became restricted to a specific type of cells.

2. Evolution of the male gametophyte

Besides the changes in the vegetative tissues that contributed to modifications in the plant life cycle, changes in the reproductive strategies and in the organs specialized for those functions represented a key innovation for the successful terrestrial adaptation. The male gametes underwent a significant transformation during this process, from motile and flagellated cells as found in Bryophytes to the sperm cells contained in the pollen grains of flowering plants.

In this work, gene expression data of three developmental stages of male gametogenesis of the moss *P. patens* was generated, including an immature and a mature stage of the antheridia, as well as isolated antherozoids from two different strains. Through comparative analysis, transcriptome dynamics underlying the development and identity of the male gametes were revealed. According to global gene expression, the transcriptome of *Physcomitrella* antherozoids showed strong differences with respect to the rest of analyzed tissues, indicating a very unique profile. In a similar manner, the gene expression of sperm cells in *Arabidopsis* was revealed to be different from somatic and female reproductive tissues (Borges et al., 2008), underlining the occurrence of particular gene expression dynamics in the male gamete.

2.1 Complex gene expression patterns during moss spermatogenesis

The large amount of genes identified to be enriched at the distinct stages during spermatogenesis in *Physcomitrella* is in accordance with a pronounced transcriptomic complexity in the sperm cells. The lower number of genes identified when a similar analysis was performed for the development of the sporophyte adds to the significance of this observation. Moreover, the striking increase of enriched transcripts during the spermatogenesis stages, increasing more than 2 fold at each stage, as well as the low overlap of expressed genes between the stages strongly suggests active transcription to occur in the mature antherozoids. The transcriptional and translational capabilities of mature male gametes have been very controversial (review in Miller et al., 2005). However, increasing studies indicating the presence of RNA in mature sperm cell from both plants and animals (Borges et al., 2008; Engel et al., 2003; Fischer et al., 2012; Ostermeier et al., 2002; Russell et al., 2012) suggest functional roles for those transcripts.

2.2 Possible functions for sperm transcripts

The generation of clusters containing particular signatures of expression provided gene functional classification of the transcripts detected during *P. patens* spermatogenesis. According to our data, very specific processes take place at distinct stages of male gametogenesis. Genes related to flagella and motility, for example, showed higher expression in the antheridia stages relatively to antherozoids, suggesting that important transcripts for gamete function are produced before antherozoid release. In accordance, similar genes were observed to be preferentially expressed prior sperm maturation in mammals (Soumillon et al., 2013). Under that

scenario, a temporal separation of transcription and translation is proposed to be required in order to ensure the presence of important components for gamete function and fertilization in mature sperm. The observation of *HAP2* following this expression profile further supports this notion. Both in *Chlamydomanas* and in *Arabidopsis HAP2* is expressed in the mature gametes and is required for gamete fusion (von Besser et al., 2006; Liu et al., 2008). In the case of the *DUO1* ortholog genes expression was detected at different stages during male gametogenesis. In *Arabidopsis*, *DUO1* is expressed from microspore to mature pollen (Rotman et al., 2005) ensuring male gamete specification through the activation of sperm enriched genes (Borg et al., 2011). Similarly, *PpDUO1a* (Pp1s16_281V6.1) is restricted to male gametogenesis related tissues, where steady expression levels are maintained.

The paternal contribution to the next generation was for long considered to be restricted to its genome. Nevertheless increasing evidence suggests that paternally originated transcripts may have a role for fertilization and embryo development. In animals several examples have been described (reviewed in Martins and Krawetz, 2005; Miller and Ostermeier, 2006; Sendler et al., 2013), while in plants evidences are restricted to the *SHORT SUSPENSOR* (*SSP*) gene, involved in the YODA-MAP kinase pathway to regulate embryo elongation (Bayer et al., 2009). Based on our cluster analysis, several genes with possible roles for embryo development were identified, providing interesting candidates for functional characterization. Two *TCP class I* genes (Pp1s332_35v6 and 15 Pp1s207_110v6) with higher expression in the antherozoids with respect to antheridia stages are among these candidates. In *Arabidopsis class I TCP* genes are involved in meristem differentiation and maintenance

(Martín-Trillo and Cubas, 2010). Additionally, members of the *JUMONJI* family (Pp1s235_45V6.1 and Pp1s235_2V6.1) involved in development and chromatin reprogramming (Chen et al., 2011; Takeuchi et al., 2006) and the homolog of *LEAFY* (*LFY*) (Pp1s258_46V6.1) important for floral meristem identity (Schultz and Haughn, 1991) were identified as belonging to this cluster.

With the aim to establish possible delivery roles, the characterization of two antherozoid preferentially expressed genes (*PpPERK1* and *PpZOITO*) with different expression levels is being conducted. Indications obtained from future experiments may document paternal RNA contribution in early land plants, adding support to a still controversial topic.

2.3 The origin of new genes in antherozoids

New genes are considered the major source of adaptive evolution. And interestingly, both in plants and in animals expression bias of new genes in the male gametes has been documented (Begun et al., 2007; Emerson et al., 2004; Levine et al., 2006). Based on those findings it was first postulated in animals (Kaessmann, 2010) and then in plants (Cui et al., 2014; Wu et al., 2014), that testis and pollen, respectively, may act as a source of evolutionary innovation. By correlating the estimated phylogenetic origin of each *P. patens* gene with their expression value on a particular tissue, we calculated the Transcriptomic Age Index (TAI) for all major tissues of the moss. According with the obtained high TAI value, the antherozoids present greater proportion of new genes being also expressed at higher levels. In accordance, lack of annotation was preferentially observed for antherozoid enriched genes, evidencing the lack of

correspondence with known genes in other species. According with these observations, the role of the male gametes for gene innovation seems to be a phylogenetically more widespread phenomenon.

Even though the mechanism allowing the preferential expression of new genes remains elusive, in animals it has been proposed to derive from relaxed chromatin states occurring in reprogramming events during early stages of gametogenesis (Kaessmann, 2010). In plants, high TAI values attributed to pollen have been considered to arise from the vegetative nucleus (VN) (Wu et al., 2014), where the activity of transposable elements (TEs) indicates a more relaxed epigenetic control (Borges et al., 2012; Slotkin et al., 2009). In *Physcomitrella*, the expression pattern of methylation related genes suggests a particular epigenetic state in mature antherozoids. On those cells, absence of transcripts encoding the maintenance methyltransferases *MET1* (Pp1s31_379V6.1) and *CMT3* (Pp1s117_71V6.1) is similar to the pattern observed in the VN, where absence of expression of those enzymes affects CG methylation preferentially in pericentromeric regions (Borges et al., 2008; Calarco et al., 2012). And only selective expression of the *de novo* methyltransferase *PpDRM2* (Pp1s104_134V6.1) and presence of the *PpDNMT3b* (Pp1s1_561V6.1) was observed, the last being highly similar to the animal *DNMT3* (Malik et al., 2012). Moreover, the expression of the DNA glycosylase *DME* (Pp1s14_101V6.1) in the antherozoids, shown to be active in the VN but not in the sperm cells of *Arabidopsis*, was identified in the antherozoids. Lacking expression of the *DECREASE IN DNA METHYLATION* (*DDM1*) (Pp1s65_183V6.1) helicase, of which its homolog in *Arabidopsis* is involved in maintaining methylation patterns (Brzeski and Jerzmanowski, 2003; Vongs et al., 1993; Zemach et al., 2013), suggests further that a relaxed chromatin

state occurring in the antherozoids may account for the expression bias of new genes. The occurrence of the described gene expression profile in the antherozoids generates questions regarding the mechanisms to achieve genomic stability and TE control in the mature male gametes. Bisulfite sequencing data and mutant phenotypic characterization will be required to clarify both the epigenetic landscape in the antherozoids as well as the possible roles of the above mentioned genes for the activation of new genes.

In brief this work provides insights on the evolution of genes controlling important traits for land plant evolution. Among those indications of the gain of functions in sporophytic genes are provided, and the functional characterization of a *class II TCP* gene validates the predictive power of the generated transcriptomic atlas. For the first time gene expression data from male gametes of early land plants was generated, offering insights on the complex functions displayed. Finally, indications of possible evolutive forces shaping male gametes in Bryophytes are presented. Future studies revealing the methylation landscape of moss male gametes, as well as functional characterization of candidate genes will allow deciphering the mechanisms involved in early land plant reproduction.

3. 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*, 1713–1723.

Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinnel, T., and Lukowitz, W. (2009). Paternal control of embryonic patterning in Arabidopsis thaliana. Science *323*, 1485–1488.

Begun, D.J., Lindfors, H. a., Kern, A.D., and Jones, C.D. (2007). Evidence for de novo evolution of testis-expressed genes in the Drosophila yakuba/Drosophila erecta clade. Genetics *176*, 1131–1137.

Von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). Arabidopsis HAP2 (GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. Development *133*, 4761–4769.

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M., and Twell, D. (2011). The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. Plant Cell *23*, 534–549.

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.

Borges, F., Calarco, J.P., and Martienssen, R. a. (2012). Reprogramming the epigenome in Arabidopsis pollen. Cold Spring Harb. Symp. Quant. Biol. *77*, 1–5.

Bowman, J.L., Floyd, S.K., and Sakakibara, K. (2007). Green Genes-Comparative Genomics of the Green Branch of Life. Cell *129*, 229–234.

Brzeski, J., and Jerzmanowski, A. (2003). Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. J. Biol. Chem. *278*, 823–828.

Calarco, J.P., Borges, F., Donoghue, M.T. a, Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J. a., Becker, J.D., et al.

(2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell *151*, 194–205.

Chen, X., Hu, Y., and Zhou, D.X. (2011). Epigenetic gene regulation by plant Jumonji group of histone demethylase. Biochim. Biophys. Acta - Gene Regul. Mech. *1809*, 421–426.

Crevillén, P., and Dean, C. (2011). Regulation of the floral repressor gene FLC: The complexity of transcription in a chromatin context. Curr. Opin. Plant Biol. *14*, 38–44.

Cui, X., Lv, Y., Chen, M., Nikoloski, Z., Twell, D., and Zhang, D. (2014). Young Genes out of the Male: An Insight from Evolutionary Age Analysis of the Pollen Transcriptome. Mol. Plant 1–11.

Dinneny, J.R., Weigel, D., and Yanofsky, M.F. (2006). NUBBIN and JAGGED define stamen and carpel shape in Arabidopsis. Development *133*, 1645–1655.

Emerson, J.J., Kaessmann, H., Betrán, E., and Long, M. (2004). Extensive gene traffic on the mammalian X chromosome. Science *303*, 537–540.

Engel, M.L., Chaboud, A., Dumas, C., and McCormick, S. (2003). Sperm cells of Zea mays have a complex complement of mRNAs. Plant J. *34*, 697–707.

Fischer, B.E., Wasbrough, E., Meadows, L. a, Randlet, O., Dorus, S., Karr, T.L., and Russell, S. (2012). Conserved properties of Drosophila and human spermatozoal mRNA repertoires. Proc. R. Soc. B Biol. Sci.

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

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

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

Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Res. *20*, 1313–1326.

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

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

Levine, M.T., Jones, C.D., Kern, A.D., Lindfors, H. a, and Begun, D.J. (2006). Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression. Proc. Natl. Acad. Sci. U. S. A. *103*, 9935–9939.

Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R., and Dean, C. (2002). Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. Science *297*, 243–246.

Li, E., Wang, S., Liu, Y., Chen, J.G., and Douglas, C.J. (2011). OVATE FAMILY PROTEIN4 (OFP4) interaction with KNAT7 regulates secondary cell wall formation in Arabidopsis thaliana. Plant J. *67*, 328–341.

Liu, Y., Tewari, R., Ning, J., Blagborough, A.., Garbom, S., Pei, J., Grishin, N.., Steele, R.., Sinden, R.., Snell, W.., et al. (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in Chlamydomonas and Plasmodium gametes. Genes Dev. *22*, 1051–1068.

Loraine, a. E., McCormick, S., Estrada, a., Patel, K., and Qin, P. (2013). RNA-Seq of Arabidopsis Pollen Uncovers Novel Transcription and Alternative Splicing. Plant Physiol. *162*, 1092–1109.

Malik, G., Dangwal, M., Kapoor, S., and Kapoor, M. (2012). Role of DNA methylation in growth and differentiation in Physcomitrella patens and characterization of cytosine DNA methyltransferases. FEBS J. *279*, 4081–4094.

Martins, R.P., and Krawetz, S. a. (2005). RNA in human sperm. Asian J. Androl. *7*, 115–120.

Martín-Trillo, M., and Cubas, P. (2010). TCP genes: a family snapshot ten years later. Trends Plant Sci. *15*, 31–39.

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

Miller, D., and Ostermeier, G.C. (2006). Spermatozoal RNA: Why is it there and what does it do? Gynecol. Obstet. Fertil. *34*, 840–846.

Miller, D., Ostermeier, G.C., and Krawetz, S. a. (2005). The controversy, potential and roles of spermatozoal RNA. Trends Mol. Med. *11*, 156–163.

Ostermeier, G.C., Dix, D.J., Miller, D., Khatri, P., and Krawetz, S. a. (2002). Spermatozoal RNA profiles of normal fertile men. Lancet *360*, 772–777.

Preston, J.C., and Hileman, L.C. (2010). SQUAMOSA-PROMOTER BINDING PROTEIN 1 initiates flowering in Antirrhinum majus through the activation of meristem identity genes. Plant J. *62*, 704–712.

Rotman, N., Durbarry, A., Wardle, A., Yang, W.C., Chaboud, A., Faure, J.E., Berger, F., and Twell, D. (2005). A novel class of MYB factors controls sperm-cell formation in plants. Curr. Biol. *15*, 244–248.

Russell, S.D., Gou, X., Wong, C.E., Wang, X., Yuan, T., Wei, X., Bhalla, P.L., and Singh, M.B. (2012). Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. New Phytol. *195*, 560–573.

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.

Schultz, E., and Haughn, G. (1991). LEAFY, a Homeotic Gene That Regulates Inflorescence Development in Arabidopsis. Plant Cell *3*, 771–781.

Sendler, E., Johnson, G.D., Mao, S., Goodrich, R.J., Diamond, M.P., Hauser, R., and Krawetz, S.A. (2013). Stability, delivery and functions of human sperm RNAs at fertilization. Nucleic Acids Res. *41*, 4104–4117.

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.

Slotkin, R.K., Slotkin, R.K., Vaughn, M., Vaughn, M., Borges, F., Borges, F., Tanurdzić, M., Tanurdzić, M., Becker, J.D., Becker, J.D., et al. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell *136*, 461–472.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., et al. (2013). Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. Cell Rep. *3*, 2179–2190.

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*, 803–812.

Takeuchi, T., Watanabe, Y., Takano-Shimizu, T., and Kondo, S. (2006). Roles of jumonji and jumonji family genes in chromatin regulation and development. Dev. Dyn. *235*, 2449–2459.

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.

Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. (1993). Arabidopsis thaliana DNA methylation mutants. Science *260*, 1926–1928.

Wang, S., Chang, Y., Guo, J., and Chen, J.-G. (2007). Arabidopsis Ovate Family Protein 1 is a transcriptional repressor that suppresses cell elongation. Plant J. *50*, 858–872.

Wang, S., Chang, Y., Guo, J., Zeng, Q., Ellis, B.E., and Chen, J.G. (2011). Arabidopsis Ovate family proteins, a novel transcriptional repressor family, control multiple aspects of plant growth and development. PLoS One *6*.

Wu, D.-D., Wang, X., Li, Y., Zeng, L., Irwin, D.M., and Zhang, Y.-P. (2014). “Out of Pollen” Hypothesis for Origin of New Genes in Flowering Plants: Study from Arabidopsis thaliana. Genome Biol. Evol. *6*, 2822–2829.

Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell *153*, 193–205.

Zhang, H., and Li, L. (2013). SQUAMOSA promoter binding protein-like7 regulated microRNA408 is required for vegetative development in Arabidopsis. Plant J. *74*, 98–109.