Molecular Biology and Evolution

Generation-Biased Gene Expression in a Bryophyte Model System

Péter Szövényi,*[,1,2] Stefan A. Rensing,[3,4] Daniel Lang,[5] Gregory A. Wray,[1] and A. Jonathan Shaw[1]

[1]Department of Biology and Institute for Genome Sciences and Policy, Duke University

[2]Faculty of Biology, University of Freiburg, Freiburg, Germany

[3]Freiburg Initiative for Systems Biology (FRISY), Faculty of Biology, University of Freiburg, Freiburg, Germany

[4]Centre for Biological Signalling Studies (BIOSS), University of Freiburg, Freiburg, Germany

[5]Department of Plant Biotechnology, Faculty of Biology, University of Freiburg, Freiburg, Germany

*Corresponding author: E-mail: pis@duke.edu.

Associate editor: Neelima Sinha

Research article

Abstract

The evolution of land plants is tightly linked to the evolution of the alternation of generations. Because alternating ploidal generations share their genomes, investigating generation-biased gene expression can give insight into the evolution of life cycles in land plants. Toward this end, we describe gene expression differences associated with the alternation of isogenic sporophyte and gametophyte generations in bryophytes, extant representatives of early diverging land plants, using a moss model system (*Funaria hygrometrica*). We found that differentiation in gene expression between the sporophyte and gametophyte generations is weaker in the bryophyte model system than in *Arabidopsis thaliana*. This is in line with the basal phylogenetic position of bryophytes and with the origin of alternating generations from a purely haplontic life cycle. Comparative analysis of *F. hygrometrica* and *A. thaliana* gene expression data shows that there is limited conservation of generation-biased gene expression across land plants. However, genes showing shared sporophyte-biased expression in both *F. hygrometrica* and *A. thaliana* appear to be enriched for biological pathways representing critical molecular adaptations to terrestrial life. Comparative analyses of the expression of *F. hygrometrica* and *A. thaliana* regulatory genes suggest that conserved regulatory networks may be involved in growth and reproductive tissue development of the angiosperm and bryophyte sporophyte generations despite their morphological divergence. This study represents the first attempt to describe generation-biased gene expression in a plant with a well-developed sporophyte and gametophyte generations, and as such it lays the foundation for future targeted research on the developmental mechanisms underlying evolutionary diversification of plant sporophytes.

Key words: high throughput sequencing, gametophyte, early land plant evolution, gene expression, *Funaria hygrometrica*, sporophyte.

Introduction

The alternation of haploid and diploid life history stages separated by syngamy and meiosis is the fundamental feature of plant sexual life cycles. Phyla of embryophytes vary in the relative dominance of the haploid (gametophyte) and diploid (sporophyte) generation, and their evolution is tightly linked to modifications in this life cycle (Goebel 1905; Haig 2008).

It is well supported that the monophyletic clade of embryophytes evolved from a green alga characterized by a life cycles in which the haploid gametophyte dominates, and the diploid phase is represented only by the unicellular zygote (Qiu et al. 2006; Haig 2008; Qiu 2008). Hence, land plants originated from a haplontic ancestor with a progressive elaboration of the sporophyte and eventual reduction of the gametophyte generation. Phylogenetic, biochemical, and morphological evidences also suggest a single evolutionary origin of the embryophyte sporophyte (Qiu et al. 2006; Qiu 2008). This phylogenetic insight implies that the multicellular sporophyte generation originated by mitotic divisions of the zygote prior to meiosis and that genetic and regulatory networks of early embryophyte sporophytes were recruited from those of the gametophyte (Nishiyama et al. 2003; Menand et al. 2007; Dolan 2009; Niklas and Kutschera 2009).

Because multicellular sporophyte and gametophyte generations of the same organism share genomes, and generation-specific morphologies and functions can be induced experimentally without gamete fusion or even changes in ploidy, phenotypic and functional divergence arises by differential usage of the same genetic material via modifications in gene expression (Langdale 2008; Dolan 2009; Mosquna et al. 2009; Okano et al. 2009). Therefore, comparative analysis of generation-biased gene expression across land plants can provide critical insights into the origin of embryophytes and the evolution of plant life cycles.

In this study, we sought to investigate genome-wide generation-biased gene expression patterns in extant early diverging land plants using a moss model system. Fossil evidence suggests that the earliest colonizers of the terrestrial environment were bryophyte-like organisms with sporic meiosis, rendering bryophytes an appropriate model

group for studying early evolution of the embryophyte sporophyte (Wellman and Gray 2000; Wellman et al. 2003; Gensel 2008; Steemans et al. 2009). *Physcomitrella patens*, the model moss, represents an inappropriate system for our study owing to its highly reduced sporophyte generation, exhibiting an atypical development (cleistocarpous) among mosses. Therefore, in this study, *F. hygrometrica*, a close relative of the model moss *P. patens*, with a well-developed and typical sporophyte generation was used. Gene expression in the gametophyte and premeiotic sporophyte generation of *F. hygrometrica*, a moss, was assessed using 454 pyrosequencing. Isogenic gametophyte and sporophyte generations were analyzed assuring that differential gene expression is solely a function of intergeneration transition. In particular, we aimed at addressing the following questions: 1) What is the extent of differential gene expression between sporophytes and gametophytes? 2) Which genes and genetic pathways are involved in generation-specific gene expression? and 3) What are the common patterns in generation-specific gene expression in bryophytes and angiosperms, and what are the implications for the evolution of the land plant sporophyte?

Materials and Methods

Plant Material

Isogenic sporophyte and gametophyte generations were obtained by intragametophytic selfing of a single haploid strain of *F. hygrometrica*. The genetic clone was established from one single-spore isolate and further replicated clonally as described in the Supplementary text (Supplementary Material online). Sporophyte and gametophyte clones were grown next to one another in a greenhouse at Duke University, and water was provided with an automatic misting system.

Generation and Analysis of the Transcriptomic Data

Gametophyte and sporophyte tissues were harvested simultaneously. Gametophytes had no gametangia, and sporophytes were all premeiotic with either swollen or unswollen capsules. The gametophyte sample consisted of whole gametophores with varying size and developmental stages including leaves, stems, and rhizoids. Similarly, the sporophyte sample comprised elongating and elongated sporophytes with swollen capsules (all sporophytes were longer than 2 cm) and included tissues of both the seta and the growing tip/capsule. Therefore, both gametophyte and sporophyte samples include a wide range of developmental stages, and gene expression estimates presented here refer to large-scale differences between the two generations. Total RNA was extracted from each clonal replicate of sporophyte and gametophyte samples separately and then pooled in equimolar ratio for cDNA amplification. Amplified cDNA of sporophyte and gametophyte libraries were sequenced on a 454 GS-20 and GS-FLX sequencer. Two 454 runs were conducted on two batches of RNA extractions originating from the same genetic clone to investigate the reliability of the expression estimates. Sequence data have been deposited in the ArrayExpress database under the study accession number E-MTAB-284. Details of sampling, RNA extraction, cDNA preparation, and 454 sequencing are given in Supplementary text (Supplementary Material online).

Raw 454 reads were processed using custom scripts written in R (R Development Core Team 2006), including adapter trimming, quality, and length clipping/filtering, according to the criteria provided in Supplementary text (Supplementary Material online). In all bioinformatics analyses, the genomic resources available for *P. patens* (v1.2; www.cosmoss.org) were used to assign reads to gene models and estimate differential expression. Filtered reads were mapped both to the *P. patens* transcriptome and to the genomic sequence using the PanGEA-BlastN code (Kofler et al. 2009) and the spliced alignment algorithm implemented in Program to Assemble Spliced Alignments (PASA) (Campbell et al. 2006; Haas 2008), respectively. PASA generated alignments are available as a separate track in the genome browser at www.cosmoss.org. After confirming the consistency of mapping between the two strategies, results generated by the PanGEA-BlastN algorithm was used in all further analyses. Gene expression was estimated as the number of reads mapped to each particular gene model and normalized by the total number of unambiguously mapped reads. Details of the bioinformatics analyses are described in Supplementary text (Supplementary Material online).

Differential Gene Expression and Functional Analyses of Genes

Significance of differential gene expression between generations was obtained using Fisher's exact test for each gene model separately. Local false discovery rate (q value) and a suitable cutoff maximizing the ratio of true positives and false nondiscoveries were obtained using the fdrtool package (Strimmer 2008). Transcripts showing a minimum of 2-fold expression change and a q value of 0.1 or less were considered differentially expressed.

Gene Ontology (GO) annotations were obtained from the relational databases available at www.cosmoss.org. Differential representation of GO terms of the biological process and cellular component ontologies were tested using the algorithm provided in the GO-diff (Chen et al. 2006) and FUNC (Prüfer et al. 2007) packages.

To compare generation-specific expression in *F. hygrometrica* to that of *Arabidopsis thaliana*, the AtGenExpress data set was retrieved from the botany array website (http://bar.utoronto.ca). Gene set enrichment analysis was conducted using the GSEA-R package (Subramanian et al. 2005) to test whether bryophyte genes with significant sporophyte- or gametophyte-biased expression are significantly overrepresented in the gametophyte- or sporophyte-biased gene sets of *A. thaliana*. Enrichment of GO terms in the core set of sporophyte- or gametophyte-characteristic gene sets was obtained using the FUNC package and the refinement option. Similar analysis was conducted to compare the

expression of transcription factor/regulator genes. Details of the statistical procedures and data sets used are described in Supplementary text (Supplementary Material online).

To investigate the differential expression of *F. hygrometrica* transcripts with putative transcription factor or transcription regulator function, *P. patens* gene models were grouped into transcription factor/regulator families, applying a unified set of classification rules (Lang et al. 2010). To compare the expression of *A. thaliana* and *F. hygrometrica* transcription factors, orthologous relationship among *A. thaliana* and *P. patens* transcription factors/regulators was established using phylogenetic analysis. Normalized expression values of transcription factor genes were retrieved from the AtGenExpress data set from the Botany array database (http://bar.utoronto.ca/). Details of all statistical analyses are provided in the Supplementary text (Supplementary Material online).

Results

454 Sequencing and Mapping Efficiency

Isogenic sporophytes were produced by intergametophytic selfing and are thus diploid and homozygous for the alleles present in the gametophyte. cDNA libraries were prepared from premeiotic sporophyte and gametophyte generations of a single genetic individual of the moss *F. hygrometrica* and sequenced using 454 pyrosequencing. After adapter trimming, quality filtering, and short read removal, 257,469 and 336,806 high-quality reads were retained for the gametophyte and sporophyte libraries, respectively.

Reads were mapped to the *P. patens* genome v1.2 (http://www.cosmoss.org/). To correct for ambiguity arising because of using the genome of a different species for mapping reads, two strategies were applied: 1) mapping against the gene models of the *P. patens* transcriptome and 2) mapping against the *P. patens* genomic sequence (supplementary table S1, Supplementary Material online). Mapping the reads against the genomic sequence using spliced alignment resulted in 67% mapped reads, and unambiguous mapping could be achieved for 34% of the sequences. Similarly, 69% of the reads could be mapped against the *P. patens* transcriptome but on average only 30% could be uniquely assigned to a particular gene model. Spliced alignment of sequence reads against the genomic sequence resulted in 16,781 assembly subclusters (genes), although they matched 13,291 gene models in the transcriptome. In both mapping strategies, average similarity (nucleotide) between *F. hygrometrica* and *P. patens* sequences was high (average: 95% and 94%; range: 77–100%), indicating close genetic relatedness of the two species. Therefore, genomic resources available for the model moss *P. patens* represent an appropriate reference for studying gene expression in *F. hygrometrica*. Approximately 2% of the transcripts showed alternative splicing, but relatively low coverage of the assemblies prohibited further analyses (data not shown). For all downstream analyses, gene expression was estimated by counting the number of reads unambiguously mapped to each gene model and normalizing with the total number of reads per sample. The two strategies differed slightly in their mapping efficiency; hence, differential gene expression was separately assessed from each approach and consistency verified. In spite of the different mapping efficiency, the pattern of differential gene expression was largely congruent between the two mapping algorithms. Gene expression estimates obtained by the first mapping strategy (mapping to the transcriptome) were used in all further analyses.

Differential Gene Expression

After normalization, 558 of the 13,291 genes were found to be differentially expressed at a 2-fold change and a q value threshold of 0.1, maximizing the ratio of true positives and false nondiscoveries (supplementary table S1, Supplementary Material online). The differentially expressed domain represents approximately 4.2% of the expressed genes and should be regarded as a minimum estimate. The average log2 fold change of genes showing differential gene expression, excluding generation-specific transcripts, was 2.25 (highest value 6.05) with a similar average value for gametophyte- and sporophyte-biased transcripts (2.01 and 2.53, respectively). Genes fell into three natural groups (five subgroups reflecting directionality) based on their differential expression patterns: 1) genes with highly biased "generation-specific" expression, 2) genes with "generation-biased" expression, and 3) genes with "similar expression" in both generations (fig. 1). From the 558 differentially expressed genes, similar numbers were differentially expressed toward the two generations; there were 277 gametophyte- and 281 sporophyte-characteristic transcripts (characteristic transcripts: generation-specific and generation-biased genes). Overall, approximately 70% and 68% of the generation-characteristic genes fell into the biased category. Most of the biased genes (91% and 96%) passing the q value threshold of 0.1 also showed an expression fold change of ≥ 2. Approximately 70% of the differentially expressed genes were generation biased, and 25% were generation-specific in their gene expression (supplementary table S1, Supplementary Material online).

Functional Analysis of Differentially Expressed Domains

In order to assess the biological meaning of generation-characteristic gene usage, differential representation of GO categories (biological process and cellular localization) among the annotated gene models was investigated. From the 13,291 *P. patens* gene models, 7,551 had GO annotation, with 5,397 and 3,158 annotated genes in the biological process and cellular component ontologies, respectively. Among the 558 significantly (fold change $\geq$ 2 and q value $\leq$ 0.1) sporophyte- and gametophyte-biased gene models, 178 and 219 had GO annotation. Considerable difference among the generations was revealed by comparing both the biological processes and the cellular component categories (fig. 2; supplementary table S2, Supplementary Material online). However, the difference was mainly due to a shift in the abundance of particular GO terms

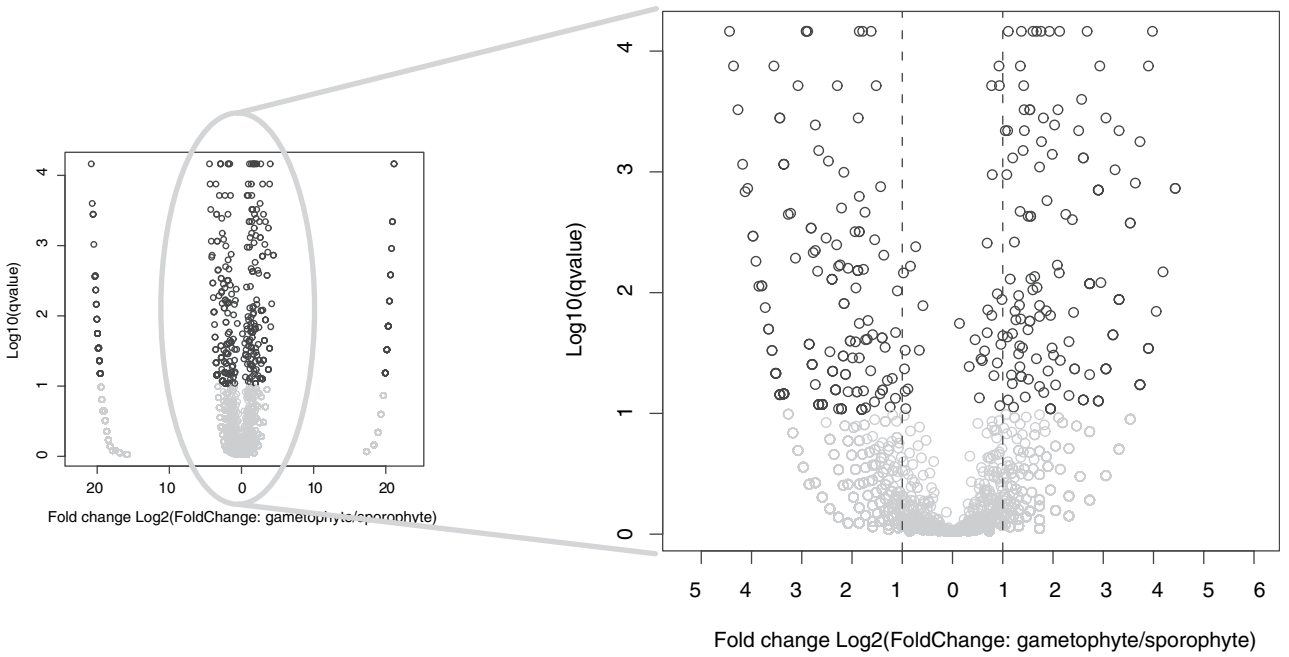

Fig. 1. Volcano plot showing the significance and fold change of differential expression of genes between isogenic gametophyte and sporophyte generations in the moss *Funaria hygrometrica*. Nondifferentially expressed transcripts (q value > 0.1) are in gray, whereas significant genes are in black. Dashed lines indicate the 2-fold expression change threshold.

rather than to the absolute occurrence of generation-specific terms.

In the gametophyte library, GO terms such as photosystem and membrane-related terms including chloroplast, plastid stroma, thylakoid, (cellular localization terms), photosynthesis, and carbon utilization-related terms (biological process terms) were highly overrepresented. Transcripts of genes involved in the light reaction of photosynthesis, such as Ribulose-1,5-bisphosphate-carboxylase/-oxigenase, the oxygen-evolving complex, and the two photosystems with their antenna proteins were abundant. Also, transcripts coding genes for glycolysis were highly differentially expressed.

GO terms characteristic of the sporophyte library were distinct from that of the gametophyte (fig. 2; supplementary table S2, Supplementary Material online). This

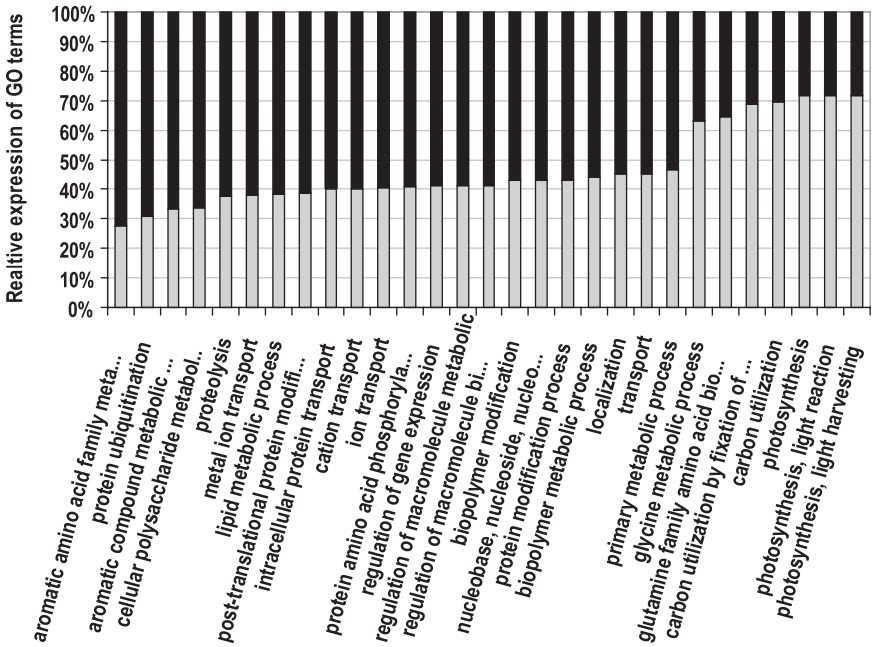

Fig. 2. Most specific significant (q value ≤ 0.1) GO terms (biological process ontology) and their relative expression level in the sporophyte and gametophyte generations of *Funaria hygrometrica*. Relative expression levels refer to the relative proportion of reads mapped to gene models with a particular GO term (black: sporophyte, gray: gametophyte).

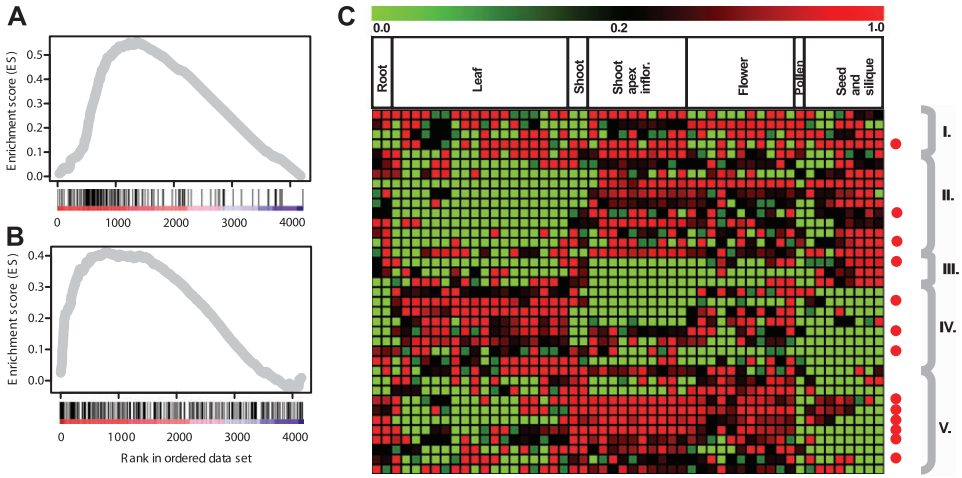

Fig. 3. Gene set enrichment analysis (GSEA) of whole-genome gene expression data (*A*, *B*) and expression heat map (*C*) of regulatory genes (*Funaria hygrometrica* and *Arabidopsis thaliana* orthologs). (*A*, *B*) GSEAs of *F. hygrometrica* and *A. thaliana* gene sets showing gametophyte-biased (*A*) or sporophyte-biased (*B*) gene expression. *x* axis: rank order of *A. thaliana* genes based on their generation specificity from sporophyte (red) to gametophyte (blue); bar code at the bottom of the graphs: position of the appropriate bryophyte orthologs in the ordered list of *A. thaliana* genes; gray curves: running enrichment score (ES) of the bryophyte genes. (*C*) *Arabidopsis thaliana* orthologs of transcription factors with gametophyte-biased expression in the bryophyte system are marked with red dots. All other orthologs show sporophyte-biased expression. Groups of transcription factors/regulators are numbered (see see I-V in the artwork.). Colors refer to log2 (fold change) values relative to the appropriate control (see color bar at the top).

transcriptome was found to be overrepresented for protein modification/degradation, gene expression regulation, lipid metabolism, transport, polysaccharide, and aromatic compound metabolism-related processes. The detailed list of overrepresented terms shows that the sporophyte transcripts are biased toward ubiquitin-regulated protein degradation processes, DNA replication and RNA splicing, fatty acid beta oxidation, both vesicle and transmembrane transport, cytoskeleton-related movement, cell wall synthesis, and stress-related processes.

Core Transcripts in the Bryophyte and Angiosperm Generation-Characteristic Domains

In order to identify generation-biased gene sets shared by the bryophyte system and angiosperms, orthology relationships (one to one and one to many) of *P. patens* and *A. thaliana* genes were extracted from the inparanoid database (http://inparanoid.sbc.su.se/cgi-bin/index.cgi). Gene expression data for *A. thaliana* were retrieved from the botany array resource (www.bar.utoronto.ca) in the form of the AtGenExpress data set (Schmid et al. 2005). We used gene set enrichment analysis (Subramanian et al. 2005) to test whether *A. thaliana* orthologs of bryophyte generation-biased genes are enriched at either end of the list of *A. thaliana* genes ranked according to their generation specificity.

Overall, *A. thaliana* orthologs of both the gametophyte- and sporophyte-biased gene sets of *F. hygrometrica* were significantly enriched in the *A. thaliana* sporophyte generation (randomization tests $P < 0.001$). Genes of the *F. hygrometrica* gametophyte-biased domain exhibited a larger number of significantly enriched genes (70% and 29%, respectively) with a higher enrichment score than those with sporophyte bias (normalized enrichment scores, 1.99 and 1.50, respectively, for *F. hygrometrica* gametophytes and sporophytes) (fig. 3). *Arabidopsis thaliana* orthologs of neither the bryophyte gametophyte-biased nor the sporophyte-biased gene sets were significantly enriched in the group of genes preferentially expressed in the *A. thaliana* gametophyte generation (randomization tests $P > 0.25$ in both cases). These patterns were independent of the mode of the orthology relationship established (one to one or one to many).

We extracted the GO terms enriched in the core set of genes showing sporophyte- or gametophyte-biased expression in both *F. hygrometrica* and *A. thaliana* (hereafter referred to as sporophyte/gametophyte-biased core set) (table 1). The sporophyte-biased core set contained multiple stress response–related genes including those important in desiccation tolerance, UV protection, osmotic stress, and homeostasis (e.g., genes such as *ERD15*, *LEAv14*, *MDAR3*). Another group of genes indicates hormonal signaling related to auxin and putatively gibberellin/DELLA protein–regulated processes (*RGA1*, polyubiquitin, *UBC5*, *AFB2*, and *PTA1*). Genes under the term hexose synthetic processes are involved in glyconeogenesis and cell wall synthesis (*PEPCK* and *RHM1/ROL1*). Interestingly, multiple sporophyte development-related genes are also in the common domain and thus suggest similar developmental functions across land plants. Examples include genes responsible for vegetative–reproductive transition (*ATH1*), repression of seed maturation genes (*ASIL1*), arrest of embryo in seed stage (*MEE14*), and receptor-like kinases involved in meristem organization and maintenance (receptor-like kinase, AT3G56370).

We also identified the core set of genes showing gametophyte-biased expression in *F. hygrometrica* and sporophyte-biased expression in *A. thaliana* (table 1). In

Table 1. Enriched GO Terms in the Core Set of Generation-Biased Genes. GO Enrichment of the Core Set of Genes, Sporophyte–Sporophyte, and Sporophyte–Gametophyte Domains, in *Arabidopsis thaliana* and in *Funaria hygrometrica*, Respectively.

			Proportion of Genes with a Particular GO Annotation			
Domain	Node Name	Node ID	In the Total Data Set	In the Core Set	log2 (Overrepresentation)	Significance of Overrepresentation After Refinement
GO terms enriched in the gene set preferentially expressed in both the *Arabidopsis thaliana* and the *Funaria hygrometrica* sporophyte	Hexose biosynthetic process	GO:0019319	0.00293	0.04615	3.97968	0.00074
	Phenylpropanoid biosynthetic process	GO:0009699	0.00479	0.04615	3.26919	0.00334
	Response to wounding	GO:0009611	0.00771	0.07692	3.31810	0.00011
	Response to hormone stimulus	GO:0009725	0.02819	0.12308	2.12623	0.00041
	Carboxylic acid metabolic process	GO:0019752	0.06410	0.16923	1.40069	0.00242
	Response to abiotic stimulus	GO:0009628	0.07793	0.20000	1.35983	0.00121
	Response to stress	GO:0006950	0.09574	0.24615	1.36230	0.02466
	Response to chemical stimulus	GO:0042221	0.08856	0.30769	1.79670	0.00040
GO terms enriched in the gene set preferentially expressed in both the *A. thaliana* sporophyte and the *F. hygrometrica* gametophyte generation	Photosynthesis	GO:0015979	0.01702	0.21818	3.68012	0.00000
	Photosynthesis, light reaction	GO:0019684	0.01037	0.12727	3.61711	0.00252
	Tetrapyrrole biosynthetic process	GO:0033014	0.00984	0.08182	3.05563	0.00858
	Photosynthesis electron transport chain	GO:0009767	0.00452	0.06364	3.81505	0.00013
	Chlorophyll biosynthetic process	GO:0015995	0.00585	0.05455	3.22069	0.00003
	Photosynthesis, light harvesting	GO:0009765	0.00239	0.03636	3.92523	0.00008
	Carbon utilization	GO:0015976	0.00186	0.03636	4.28780	0.00002
	Regulation of photosynthesis, light reaction	GO:0042548	0.00213	0.03636	4.09516	0.00004
	Photosynthetic electro transport in photosystem I	GO:0009773	0.00186	0.02727	3.87276	0.00078
	Response to light stimulus	GO:0009416	0.03245	0.11818	1.86486	0.01450
	Response to cold	GO:0009409	0.01676	0.07273	2.11788	0.00042
	Response to reactive oxygen species	GO:0000302	0.00346	0.02727	2.97968	0.00563
	Respose to blue light	GO:0009637	0.00319	0.02727	3.09516	0.00442
	Response to high light intensity	GO:0009644	0.00346	0.02727	2.97968	0.00563
	Response to red light	GO:0010114	0.00319	0.02727	3.09516	0.00442
	Response to far red light	GO:0010218	0.00319	0.02727	3.09516	0.00442
	Translation	GO:0006412	0.04468	0.11818	1.40328	0.00106
	Ribosome biogenesis	GO:0042254	0.01436	0.05455	1.92523	0.00450
	Peptidyl-cysteine S-nitrosylation	GO:0018119	0.00213	0.03636	4.09516	0.00004

general, this set is mainly enriched for photosynthesis-related and carbon fixation–related genes including light harvesting, electron transport, chlorophyll synthesis, photosynthesis-regulating protein products, and enzymes of the Calvin cycle.

Regulatory Genes and their Generation-Biased Expression

Differential gene expression in the two generations prompted us to examine the transcription factors and other regulatory genes present in our data set first by investigating the abundance of each major family. Gene models with putative transcription factor or transcription regulator activity were identified by applying a comprehensive set of classification rules (Lang et al. 2010). From the 95 transcription factor families present in the *P. patens* genome, members of 92 families (ca., 97% of the total) were detected in the total expression data set representing 736 *P. patens* gene models (supplementary table S3, Supplementary Material online). In the total data set, neither the proportion of families shared by all land plants but missing from sequenced green algae (further on referred to as SBL families) nor the proportion of green plant–specific families (further on referred to as GS families) were significantly different from the ratio of these families in the *P. patens* genome (23% vs. 27%, $\chi^2_1=0.1813$, $P = 0.67$ and 17% vs. 28%, $\chi^2_1=2.75$, $P = 0.097$). Also, families known to have expanded in the common ancestor of land plants were not preferentially represented in the data set (47% vs. 47%, $\chi^2_1=0.019$, $P=0.89$). Therefore, transcription factors associated with the origin of embryophytes were not preferentially expressed in our data set. Abundance of most of the transcription factor families (73 in total) could be well explained by their corresponding gene number in the *P. patens* genome. However, 8 among the most abundantly expressed families were preferentially expressed, including the families PHD, GRAS, Aux/IAA, Zinc finger, C2C2, FHA, and TCP, whereas 11 were less abundant than expected

(supplementary table S3, Supplementary Material online). These overrepresented or underrepresented families were not enriched for SBL nor for GS or for those expanded in the last common ancestor of land plants (further on referred to as ELP families) (38% vs. 23%, $\chi_1^2=0.25$, $P = 0.62$; 63% vs. 28%, $\chi_1^2=2.59$, $P = 0.11$; 40% vs. 47%, $\chi_1^2=0.037$, $P = 0.85$). Only 13 of the 92 expressed families showed significant generation-biased gene expression (q value ≤ 0.1), and these were enriched neither for SBL nor for ELP families ($\chi_1^2=0.08$, $P = 0.78$; $\chi_1^2=0.45$, $P = 0.48$). Interestingly, only one of the abundantly expressed families (OFP) was restricted to one generation (sporophyte).

Members of a particular transcription factor family are expected to regulate divergent processes and may show different generation/tissue specificity. Therefore, we looked at generation-biased gene expression at the level of individual transcription factor genes. To overcome the low statistical power associated with the generally low expression level of transcription factors, their differential expression was investigated at multiple statistical thresholds. From the 736 putative transcription factors/regulators 51, 23, or 15 were significantly biased toward one of the generations depending on the statistical threshold used (corresponding to q value thresholds of 0.3, 0.2, and 0.1, respectively; supplementary table S4, Supplementary Material online). At the most stringent significance level, all but one transcription factor gene was sporophyte biased, with seven showing specific expression. The most significantly sporophyte-biased transcription factors were members of the AP2-EREBP, GRAS, zinc finger, TRAF, bHLH, C3H, HSF, OFP, C2H2, and HB families, whereas one gene of the PHD family showed gametophyte-biased expression. Several other genes, namely members of the C2C2-Dof, C2C2-GATA, ABI3/VP1, Aux/IAA, LIM, ARID, TIFY, Sigma70-like, BSD, WRKY, MYB, SBP, PLATZ, TCP, AFR, bZIP, and Trihelix families showed less significant generation-biased expression (q value thresholds of 0.2 and 0.3). At both the most stringent and the most relaxed significance levels, several families were represented by multiple genes showing generation-biased expression. Some families had members showing both sporophyte- and gametophyte-biased patterns, indicating functional divergence of the expression domain. Importantly, SBL families did not show significantly more abundant generation-biased expression than expected based on their frequency in the *P. patens* genome (14% vs. 27%, $\chi_1^2=2.48$, $P = 0.115$). In contrast, ELP families were significantly overrepresented among the differentially expressed families (80% vs. 45%, $\chi_1^2=14.02$, $P < 0.001$). This suggests that members of these families have played a prominent role in the development and diversification of the embryophyte generations.

To quantify the similarity in generation-biased expression of transcription factors across land plants, we tested whether *A. thaliana* orthologs of the *F. hygrometrica* generation-biased transcription factors are enriched in the sporophyte- or gametophyte-characteristic gene set of *A. thaliana* using gene set enrichment analysis. Orthologous relationships were established by phylogenetic analyses of the conserved domain alignments of *A. thaliana* and *P. patens* gene models (supplementary table S4, Supplementary Material online). Genes with both gametophyte- and sporophyte-biased expression were biased toward the sporophyte-characteristic gene set of *A. thaliana*. However, genes with gametophyte-biased expression in *F. hygrometrica* had a higher normalized enrichment score and a higher proportion of genes in the core set (1.36% and 64%) than genes with sporophyte-biased expression (0.95% and 43%).

Finally, we used the AtGenExpress (http://bar.utoronto.ca/) data set to investigate the expression pattern of *A. thaliana* orthologs of the differentially expressed transcription factors found in our study. Based on their tissue-specific expression pattern, *A. thaliana* orthologs of the differentially expressed *F. hygrometrica* transcription factors fall into five large groups: 1) genes showing no obvious generation or tissue preference; 2) genes mainly expressed in the shoot apex, flower, and seed development; 3) genes expressed mainly in seeds; 4) genes expressed mainly in leaves and flowers; and 5) genes mainly expressed in the shoot apex, flower buds, and in the root (fig. 3). Clearly, *A. thaliana* orthologs of differentially expressed *F. hygrometrica* genes are expressed and have important functions in the developmental processes of all major plant organs. Especially interesting is the high proportion of *A. thaliana* orthologs preferentially expressed in the shoot apex and in developing reproductive tissues, suggesting similar molecular regulatory mechanisms in *F. hygrometrica* and *A. thaliana* sporophytes (supplementary table S4, Supplementary Material online).

Discussion

Here we describe gene expression differences associated with the alternation of isogenic sporophyte and gametophyte generations in a moss, an extant representative of early diverging land plants. Our findings are based on the model species, *F. hygrometrica*, which has a well-developed sporophyte, in contrast to the atypical and highly reduced cleistocarpous sporophytes of the closely related model moss *P. patens*. In our study, isogenic generations could be compared directly: Thus, gene expression differences reflect only intergeneration changes in the transcriptome. Although our study relies on a vast amount of sequence data, only approximately 30% of the *Funaria* reads could be unambiguously mapped to the *P. patens* genome. Therefore, resolution of the current data set is limited and provides reliable information about the large-scale pattern of generation-biased gene expression in bryophytes.

Both phylogenetic and morphological analyses as well as fossil data suggest that mosses, liverworts, and hornworts are early branching groups of land plants (Wellman and Gray 2000; Wellman et al. 2003; Graham et al. 2004; Qiu et al. 2006; Qiu 2008). Bryophytes are thus believed to represent an early stage in the evolution of the multicellular sporophyte generation (Gensel 2008). The origin of

alternating multicellular generations from a purely haplontic life cycle suggests that gene expression differentiation of sporophyte and gametophyte generations may have been weak in the early diverging lineages of land plants such as mosses. Our findings show that a large proportion of genes are expressed in both the gametophyte and the sporophyte generation in bryophytes. In particular, we found that the proportion of transcripts with generation-biased gene expression (2.5–2.5%, gametophyte, sporophyte) is lower in the bryophyte studied here as compared with the angiosperm *A. thaliana* (ca., 5% and ca., 25%, gametophyte and sporophyte, respectively) (Honys and Twell 2003; Pina et al. 2005; Ma et al. 2008; Haerizadeh et al. 2009). Furthermore, the similar proportion of sporophyte- and gametophyte-biased genes suggests less gene expression specialization between generations in the bryophyte lineage compared with angiosperms, where distribution of generation-specific gene products is unequal. The extensive overlap and weak specialization in gene expression between the two generations is in agreement with the putative origin of alternation of generations from an algal ancestor with purely haplontic life cycle (Karol et al. 2001; Qiu et al. 2006). Therefore, our data suggest that extensive sharing in gene expression between generations may have been the rule rather than the exception in the earliest land plants. Furthermore, the finding that no generation-specific GO terms were discovered, but differences were rather due to their differential abundance, suggests that most biological pathways may have been shared between gametophytic and sporophytic tissues. Later during seed plant evolution, the elaboration of the sporophyte and the reduction of the gametophyte generation resulted in less overlapping gene expression and extensive functional specialization of the generations. Although the resolution of our data set is limited (ca., 30% of the reads could be unambiguously assigned to *Physcomitrella* gene models), the weak gene expression differentiation between generations in *Funaria* is unlikely to be due to limited tissue sampling because both sporophyte and gametophyte samples consisted of a wide range of developmental stages. Therefore, we believe that our conclusions are well established; however, we acknowledge that additional data would improve our estimates concerning the extent of generation-biased gene expression in bryophytes.

One major finding of our comparative analysis is that gene expression in the *F. hygrometrica* sporophyte and gametophyte generations is more similar to those of the *A. thaliana* sporophyte. Moreover, the overlap in gene expression is considerably greater between gametophyte and sporophyte than sporophyte and sporophyte generations of *F. hygrometrica* and *A. thaliana*, respectively. Because embryophytes are monophyletic with a putative haplontic ancestor, it is hypothesized that genetic networks of the early sporophyte were mainly recruited from that of the gametophyte generation (Karol et al. 2001; Nishiyama et al. 2003; Qiu et al. 2006; Menand et al. 2007; Dolan 2009; Niklas and Kutschera 2009). Therefore, more overlap in gene expression between bryophyte gametophyte and angiosperm sporophyte generations implies that their morphological and functional similarity relies on the shared usage of a common set of orthologous genes and gene networks that may have already been in place in the common ancestor's gametophyte generation. Moreover, the observation that orthologs of transcription factors with gametophyte-biased expression in bryophytes are enriched in the sporophyte-characteristic domain of *A. thaliana* supports the hypothesis that multiple regulatory processes involved in the angiosperm sporophyte were likely recruited from the gametophyte (Langdale 2008; Dolan 2009; Niklas and Kutschera 2009). Although angiosperm sporophytes and bryophyte gametophytes are morphologically similar (stem, leaf, and root-like structures), their macromorphological developmental processes are highly divergent (Harrison et al. 2009). Therefore, it would be interesting to see how massive genomic transfer of regulatory programs between generations could take place in face of highly divergent macromorphological developmental processes. Remarkably, there was little similarity in gene expression between either of the bryophyte generations and that of the angiosperm gametophyte. Therefore, gene networks of the angiosperm gametophyte may have been either recruited from the filamentous phase of the ancestral life cycle (protonemata in bryophytes, not sampled here) or secondarily acquired from the angiosperm sporophyte program.

Gene expression conservation in the sporophyte generation across land plants and function of the shared set of sporophyte-characteristic genes provides hints to the complexity of the common ancestor's sporophyte generation. Poor gene expression conservation across the sporophyte generation of land plants suggests that bryophyte and angiosperm sporophytes must have diverged in an early stage of sporophyte evolution. Recruitment of gametophytic programs into the common ancestor's sporophyte may have been rudimentary, resulting in the poor conservation of sporophytic gene expression. However, as our analysis indicates, the core set of sporophyte-expressed transcripts includes genes likely necessary for terrestrial life (enhanced osmotic regulation, stress, desiccation and UV tolerance, transporters, and intercellular communication via hormones). Therefore, biological processes underlying molecular adaptation to terrestrial life may have already been used by early sporophytes (Graham et al. 2000; Rensing et al. 2008). Overall, our data suggest that the last common ancestor of embryophytes may have possessed a simple sporophyte nevertheless molecularly adapted to terrestrial life.

Early divergence of the bryophyte and angiosperm sporophyte is well mirrored by their contrasting morphology and developmental biology (Renzaglia et al. 2007). However, single origin of the embryophyte sporophyte implies the presence of shared basic developmental mechanisms. Previously, class I KNOX, LEAFY, MADS box, and Polycomb group genes have been implicated as important transcription factors involved in the development of both the bryophyte and the angiosperm sporophyte generation (Sano

et al. 2005; Quodt et al. 2007; Singer and Ashton 2007; Langdale 2008; Mosquna et al. 2009; Okano et al. 2009; Singer and Ashton 2009). Our data suggest the presence of other conserved regulatory processes. We show that a significant proportion of *A. thaliana* orthologs of the preferentially sporophyte-expressed transcription factors in *F. hygrometrica* are known to function in the shoot apical meristem and in development of reproductive structures. Examples include OFP, BELL-like homeobox, SBP, TCP, C2H2, and GRAS proteins (orthologous to *A. thaliana* SCL- and DELLA-like genes) (Bolle 2004; Hackbusch et al. 2005; Gomez-Mena and Sablowski 2008; Martín-Trillo and Cubas 2009; Poethig 2009). It will be interesting to investigate whether these genes have parallel functions in the developing moss sporophyte, with its subterminal/apical meristem and sporogenous tissue. The presence of orthologous transcription factors with putatively shared functions in the developmental processes of both reproductive and vegetative sporophytic tissues/organs implies that both domains may have been already present in the common ancestor's sporophyte (Kato and Akiyama 2005). Overall, our data suggest that in spite of considerable morphological and developmental divergence, similar basic regulatory networks may be involved in the growth and reproductive tissue development of the bryophyte and angiosperm sporophyte generation. This observation is in accordance with the single evolutionary origin of the land plant sporophyte (Qiu et al. 2006; Qiu 2008).

In this study, we provide the first comprehensive analysis of generation-biased gene expression in a moss model system and discuss its evolutionary implications. It is envisioned that our findings will provide an invaluable tool for future research on the developmental mechanisms underlying diversification of land plant generations.

Supplementary Material

Supplementary text and supplementary tables S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This research was financially supported by Duke University Institute for Genome Sciences and Policy (postdoctoral fellowship in evolutionary genomics to P.S.) and by the Deutscher Akademischer Austausch Dienst A/09/90062 (postdoctoral fellowship to P.S.). Critical comments of two anonymous reviewers are greatly acknowledged.

References

Bolle C. 2004. The role of GRAS proteins in plant signal transduction and development. *Planta* 218:683–692.

Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR. 2006. Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. *BMC Genomics*. 7:327.

Chen Z, Wang W, Ling XB, Liu JJ, Chen L. 2006. GO-Diff: mining functional differentiation between EST-based transcriptomes. *BMC Bioinformatics*. 7:72.

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

Gensel PG. 2008. The earliest land plants. *Annu Rev Ecol Syst*. 39:459–477.

Goebel K. 1905. Organography of plants, especially of the archegoniatae and spermatophyta. Part II. Special organography. Oxford: Clarendon Press.

Gomez-Mena C, Sablowski R. 2008. *Arabidopsis thaliana* homeobox gene 1 establishes the basal boundaries of shoot organs and controls stem growth. *Plant Cell*. 20:2059–2072.

Graham LE, Cook ME, Busse JS. 2000. The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proc Natl Acad Sci U S A*. 97:4535–4540.

Graham LE, Wilcox LW, Cook ME, Gensel PG. 2004. Resistant tissues of modern marchantioid liverworts resemble enigmatic Early Paleozoic microfossils. *Proc Natl Acad Sci U S A*. 101:11025–11029.

Haas BJ. 2008. Nuclear pre-mRNA processing in plants. In: Reddy ASN, Golovkin M, editors. Current topics in microbiology and immunology. Berlin (Germany): Springer. p. 17–37.

Hackbusch J, Rischter K, Müller J, Salamini F, Uhrig JF. 2005. A central role of Arabidopsis thaliana ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. *Proc Natl Acad Sci U S A*. 102:4908–4912.

Haerizadeh F, Wong CE, Bhalla PL, Gresshoff PM, Singh MB. 2009. Genomic expression profiling of mature soybean (*Glycine max*) pollen. *BMC Plant Biol*. 9:25.

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

Harrison CJ, Roeder AHK, Meyerowitz EM, Langdale JA. 2009. Local cues and asymmetric cell divisions underpin body plan transitions in the moss *Physcomitrella patens*. *Curr Biol*. 19:461–471.

Honys D, Twell D. 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol*. 132:640–652.

Karol KG, McCourt RM, Cimino MT, Delwiche CF. 2001. The closest living relatives of land plants. *Science* 294:2351–2353.

Kato M, Akiyama H. 2005. Interpolation hypothesis for origin of the vegetative sporophyte of land plants. *Taxon* 54:443–450.

Kofler R, Torres TT, Lelley T, Schlötterer C. 2009. PanGEA: identification of allele specific gene expression using the 454 technology. *BMC Bioinformatics*. 10:143.

Lang D, Weiche B, Timmerhaus G, Richardt S, Riaño-Pachón DM, Corrêa LG, Reski R, Mueller-Roeber B, Rensing SA. 2010. Genome-wide classification and phylogenetic comparative analysis of transcriptional regulation in plants: timeline for a story of loss, gain, expansion and evolutionary correlation with morphological complexity. *Genome Biol Evol*. 2:488–503.

Langdale JA. 2008. Evolution of developmental mechanisms in plants. *Curr Opin Genet Dev*. 18:368–373.

Ma J, Skibbe DS, Fernandes J, Walbot V. 2008. Male reproductive development: gene expression profiling of maize anther and pollen ontogeny. *Genome Biol*. 9:R181.

Martín-Trillo M, Cubas P. 2009. 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 AG, Dolan L. 2007. An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 316:1477–1480.

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

Niklas KJ, Kutschera U. 2009. The evolution of the plant life cycle. *New Phytol*. 185:27–41.

Nishiyama T, Fujita T, Shin-I T, et al. (12 co-authors). 2003. Comparative genomics of Physcomitrella patens gametophytic

transcriptome and Arabidopsis thaliana: implications for land plant evolution. *Proc Natl Acad Sci U S A*. 13:8007–8012.

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

Pina C, Pinto F, Feijo JA, Becker JD. 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.

Poethig RS. 2009. Small RNAs and developmental timing in plants. *Curr Opin Genet Dev*. 19:374–378.

Prüfer K, Muetzel B, Do HH, Weiss G, Khaitovich P, Rahm E, Pääbo S, Lachmann M, Enard W. 2007. FUNC: a package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics*. 8:41.

Qiu YL. 2008. Phylogeny and evolution of charophytic algae and land plants. *J Syst Evol*. 46:287–306.

Qiu YL, Li L, Wang B, et al. (21 co-authors). 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci U S A*. 103:15511–15516.

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

R Development Core Team. 2006. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.ISBN 3-900051-07-0. [cited 2010 July 10] Available from: http://www.R-project.org

Rensing SA, Lang D, Zimmer AD, et al. (70 co-authors). 2008. The genome of the moss *Physcomitrella patens* reveals evolutionary insights into the conquest of land by plants. *Science* 319:64–69.

Renzaglia KS, Schuette S, Duff RJ, Ligrone R, Shaw AJ, Mishler BD, Duckett J. 2007. Bryophyte phylogeny: advancing the molecular and morphological frontiers. *Bryologist* 110:179–213.

Sano R, Juárez CM, Hass B, Sakakibara K, Ito M, Banks JA, Hasabe M. 2005. KNOX homeobox genes potentially have similar function in both diploid unicellular and multicellular meristems, but not in haploid meristems. *Evol Dev*. 7:69–78.

Schmid M, Davison TS, Henz SR, et al. (9 co-authors). 2005. A gene expression map of *Arabidopsis* development. *Nat Genet*. 37:501–506.

Singer SD, Ashton NW. 2007. Revelation of ancestral roles of KNOX genes by a functional analysis of *Physcomitrella* homologues. *Plant Cell Rep*. 26:2039–2054.

Singer SD, Ashton NW. 2009. MADS about moss. *Plant Signal Behav*. 4:111–112.

Steemans P, Hérissé AL, Melvin J, Miller MA, Paris F, Verniers J, Wellman CH. 2009. Origin and radiation of the earliest vascular land plants. *Science* 324:353.

Strimmer K. 2008. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* 24:1461–1462.

Subramanian A, Tamayo P, Mootha VK, et al. (11 co-authors). 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 102:15545–15550.

Wellman CH, Gray J. 2000. The microfossil record of early land plants. *Philos Trans R Soc London Ser B Biol Sci*. 355:717–732.

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