



## RESOURCE

# The *Physcomitrella patens* gene atlas project: large-scale RNA-seq based expression data

Pierre-François Perroud<sup>1,\*</sup> , Fabian B. Haas<sup>1</sup> , Manuel Hiss<sup>1</sup>, Kristian K. Ullrich<sup>1,†</sup> , Alessandro Alboresi<sup>2,‡</sup>, Mojgan Amirebrahimi<sup>3</sup>, Kerrie Barry<sup>3</sup>, Roberto Bassi<sup>2</sup>, Sandrine Bonhomme<sup>4</sup>, Haodong Chen<sup>5</sup>, Juliet C. Coates<sup>6</sup> , Tomomichi Fujita<sup>7</sup>, Anouchka Guyon-Debast<sup>4</sup>, Daniel Lang<sup>8</sup>, Junyan Lin<sup>3</sup>, Anna Lipzen<sup>3</sup>, Fabien Nogu  <sup>4</sup>, Melvin J. Oliver<sup>9</sup>, In  s Ponce de Le  n<sup>10</sup>, Ralph S. Quatrano<sup>11</sup>, Catherine Rameau<sup>4</sup>, Bernd Reiss<sup>12</sup>, Ralf Reski<sup>13,17</sup> , Mariana Ricca<sup>14</sup>, Younouss Saidi<sup>6, </sup>, Ning Sun<sup>5</sup>, P  ter Sz  v  nyi<sup>14</sup>, Avinash Sreedasyam<sup>15</sup>, Jane Grimwood<sup>15</sup>, Gary Stacey<sup>16</sup>, Jeremy Schmutz<sup>3,15</sup> and Stefan A. Rensing<sup>1,17,\*</sup> 

<sup>1</sup>Plant Cell Biology, Faculty of Biology, University of Marburg, Karl-von-Frisch-Str. 8, 35043, Marburg, Germany,

<sup>2</sup>Dipartimento di Biotechnologie, Universit   di Verona, C   Vignal 1, Strada Le Grazie 15, 37134, Verona, Italy,

<sup>3</sup>US Department of Energy (DOE) Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, 94598, USA,

<sup>4</sup>Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Universit   Paris-Saclay, Route de St-Cyr RD10, 78026, Versailles Cedex, France,

<sup>5</sup>School of Advanced Agriculture Sciences and School of Life Sciences, Peking University, Beijing, China,

<sup>6</sup>School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK,

<sup>7</sup>Department of Biological Sciences, Faculty of Science, Hokkaido University, Kita 10 Nishi 8, Kita-ku, Sapporo 060-0810, Japan,

<sup>8</sup>Helmholtz Zentrum M  nchen, Ingolst  dter Landstr. 1, 85764, Neuherberg, Germany,

<sup>9</sup>USDA-ARS-MWA, Plant Genetics Research Unit, University of Missouri, Columbia, MO, 652117, USA,

<sup>10</sup>Department of Molecular Biology, Clemente Estable Biological Research Institute, Avenida Italia 3318, CP 11600, Montevideo, Uruguay,

<sup>11</sup>Department of Biology, Washington University in St Louis, One Brookings Drive, St Louis, MO, 63130, USA,

<sup>12</sup>Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, 50829, K  ln, Germany,

<sup>13</sup>Plant Biotechnology, Faculty of Biology, University of Freiburg, Sch  nzlestr. 1, 79104, Freiburg, Germany,

<sup>14</sup>Department of Systematic and Evolutionary Botany, University of Zurich, Zollikerstr. 107, 8008 Z  rich, Switzerland,

<sup>15</sup>HudsonAlpha Institute for Biotechnology, 601 Genome Way Northwest, Huntsville, AL, 35806, USA,

<sup>16</sup>Divisions of Plant Science and Biochemistry, National Center for Soybean Biotechnology, University of Missouri, Columbia, MO, 65211, USA,

<sup>17</sup>BIOSS Centre for Biological Signalling Studies, University of Freiburg, Sch  nzlestr. 18, 79104, Freiburg, Germany,

<sup>†</sup> Present address: Max Planck Institute for Evolutionary Biology, August-Thienemann-Stra  e 2, 24306, Ploen, Germany,

<sup>‡</sup> Present address: Dipartimento di Biologia, Universit   di Padova, Viale Giuseppe Colombo, 3, 35131, Padova, Italy, and

<sup> </sup> Present address: Bayer Crop Science, Technologiepark, Zwijnaarde 38, 9052, Gent, Belgium

Received 29 November 2017; revised 2 April 2018; accepted 5 April 2018; published online 22 April 2018.

\*For correspondence (e-mail pierre-francois.perroud@biologie.uni-marburg.de; stefan.rensing@biologie.uni-marburg.de).

## SUMMARY

High-throughput RNA sequencing (RNA-seq) has recently become the method of choice to define and analyze transcriptomes. For the model moss *Physcomitrella patens*, although this method has been used to help analyze specific perturbations, no overall reference dataset has yet been established. In the framework of the Gene Atlas project, the Joint Genome Institute selected *P. patens* as a flagship genome, opening the way to generate the first comprehensive transcriptome dataset for this moss. The first round of sequencing described here is composed of 99 independent libraries spanning 34 different developmental stages and conditions. Upon dataset quality control and processing through read mapping, 28 509 of the 34 361 v3.3 gene models (83%) were detected to be expressed across the samples. Differentially expressed genes (DEGs) were calculated across the dataset to permit perturbation comparisons between conditions. The analysis of the three most distinct and abundant *P. patens* growth stages – protonema, gametophore and

sporophyte – allowed us to define both general transcriptional patterns and stage-specific transcripts. As an example of variation of physico-chemical growth conditions, we detail here the impact of ammonium supplementation under standard growth conditions on the protonemal transcriptome. Finally, the cooperative nature of this project allowed us to analyze inter-laboratory variation, as 13 different laboratories around the world provided samples. We compare differences in the replication of experiments in a single laboratory and between different laboratories.

**Keywords:** developmental stage, differential expression, *Physcomitrella patens*, RNA-seq, stress, transcriptome analysis.

## INTRODUCTION

Since the discovery of its intrinsically efficient gene targeting (Schaefer and Zryd, 1997), followed by its genome sequencing (Rensing *et al.*, 2008; Lang *et al.*, 2018), the moss *Physcomitrella patens* has become the leading reference non-seed plant model. It is now notably integrated in studies focused on the water-to-land plant transition (e.g. Renault *et al.*, 2017) or the establishment of tri-dimensional growth in plants (for a review, see Harrison, 2017), with both fields of study integrating detailed investigation of functional cell biology with kingdom-wide gene and genome comparison. As these multiple research fields grew, so did their associated technical approaches. Amongst them, RNA-seq (the deep sequencing of cDNA) is now dominating the field of RNA detection and quantification at the transcriptome level, replacing probe-based microarray technology. Detailed knowledge of the transcriptome of a given organism is being used to improve genome assemblies (Song *et al.*, 2016), to better understand and describe RNA splicing pattern (Gaidatzis *et al.*, 2015) and to characterize spatiotemporal transcriptome variation (expression profiling), both with respect to development and environmental perturbations.

In the green lineage, RNA-seq approaches to assess transcriptome-wide patterns were initially used in models such as *Arabidopsis thaliana* (Lister *et al.*, 2008) or *Chlamydomonas reinhardtii* (González-Ballester *et al.*, 2010), followed by major crop plants such as maize (maize; Eveland *et al.*, 2010) and *Oryza sativa* (rice; Zhang *et al.*, 2010), for which a complete genome sequence was available. Furthermore, the improvement of the *de novo* assembly of transcriptomes allowed the use of this approach to characterize transcriptomes in organisms without available genome sequences (e.g. *Cassia angustifolia*; Rama Reddy *et al.*, 2015), gaining knowledge of both the raw sequence information and about biological processes in species previously limited by the lack of a genome. The most wide-spread *de novo* transcriptome assembly effort so far is the 1KP project, covering more than 1300 species of the green algae and the land plant lineage with low RNA-seq sequencing coverage (Matasci *et al.*, 2014).

In non-seed plants, RNA-seq based transcriptomic studies have been reported in multiple species, both in parallel with genome sequencing projects and via *de novo* analysis. Besides *P. patens* (see below), transcriptomes of other mosses have been published: *Sphagnum* spp. (Devos *et al.*, 2016), *Bryum argenteum* (Gao *et al.*, 2014), *Ceratodon purpureus* (Szövényi *et al.*, 2015), *Funaria hygrometrica* (Szövényi *et al.*, 2011) or *Syntrichia caninervis* (Gao *et al.*, 2015). Published datasets are also available for liverworts, including *Pellia endiviifolia* (Alaba *et al.*, 2015) and *Marchantia polymorpha* (Sharma *et al.*, 2014), the genome of which has recently been published (Bowman *et al.*, 2017). The transcriptome of the lycophyte *Selaginella moellendorffii*, for which a draft genome is available (Banks *et al.*, 2011), has also been subjected to extensive RNA-seq study (Zhu *et al.*, 2017). With large genomes, ferns form a group with exclusively *de novo* transcriptome datasets so far: e.g. *Acrostichum* spp., *Ceratopteris thalictroides* (Zhang *et al.*, 2016), *Ceratopteris richardii* (Bushart *et al.*, 2013) and *Lygopodium japonicum* (Aya *et al.*, 2015).

In *P. patens*, RNA-seq datasets have been released in multiple experimental contexts, unfortunately with no systematic multiple experimental replications (for a review, see Hiss *et al.*, 2017). For example, the *P. patens* transcriptome profile has been studied with respect to developmental stage (Xiao *et al.*, 2011), in addition to stress treatments including bleomycin (Kamisugi *et al.*, 2016). Analysis of heat-stress impact on alternative splicing has also benefited from the RNA-seq approach (Chang *et al.*, 2014). Recently, transcriptomic responses to plant hormone treatments with abscisic acid (Stevenson *et al.*, 2016) and auxin (Lavy *et al.*, 2016) have been studied. Additionally, comparative transcriptomic approaches have been applied to mutant analysis (Chen *et al.*, 2012; Demko *et al.*, 2014), and to analyze and catalogue small RNAs in *P. patens* (e.g. Coruh *et al.*, 2015; Lang *et al.*, 2018).

The present study describes the first part of the *P. patens* dataset of the US Department of Energy (DOE) Joint Genome Institute (JGI) Gene Atlas Project (<http://jgi.doe.gov/doe-jgi-plant-flagship-gene-atlas/>). After reviewing the dataset we focus on experimental comparisons,

underscoring different aspects of such large-scale projects. In terms of tissue-specific expression profiling, we show the possibility of defining specific transcripts for the three dominant life stages of *P. patens*: protonema, gametophore and sporophyte. We also tackle two aspects of transcriptome comparison experiments. We evaluate the impact of nitrogen supplementation in single-laboratory settings and show here the power of such an approach. Moreover, the diversity of the sample sources permitted us to compare two experimental replica sets of the same growing conditions performed by two different laboratories to evaluate inter-laboratory replication.

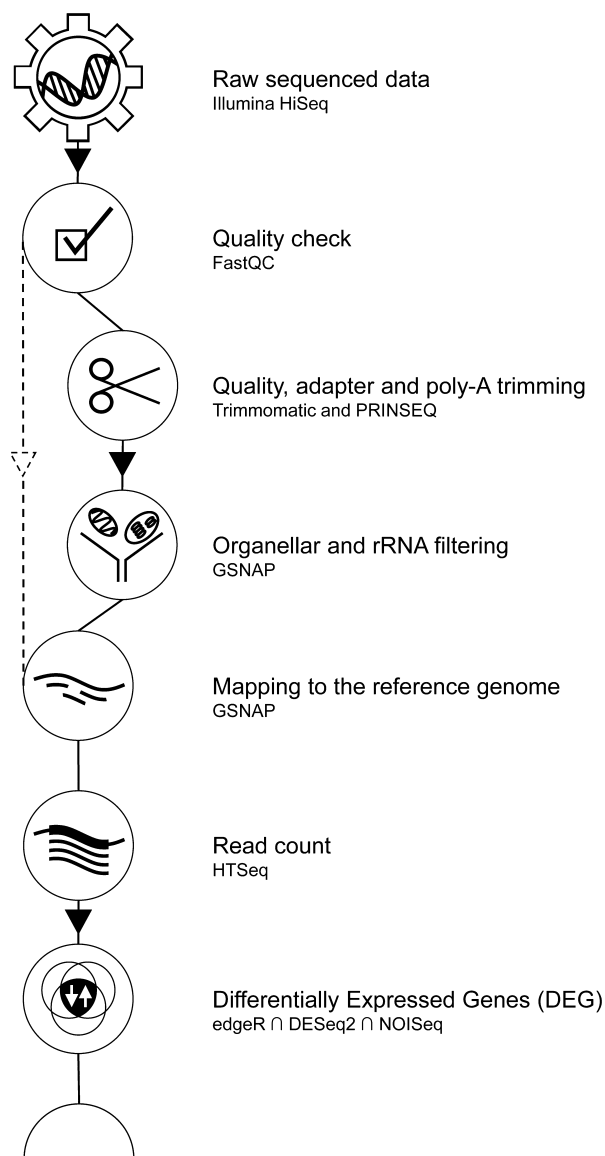
## RESULTS AND DISCUSSION

### Overview of the dataset

The *P. patens* Gene Atlas dataset comprises 99 sequenced libraries of 34 different experiments. All but three experiments are composed of three biological replicates. For experiments XIV, XVIII and XXII, one of the libraries failed for technical reasons, hence they are formed of biological duplicates. Thirteen laboratories actively working with *P. patens* around the world contributed to the samples described in Experimental procedures. The detailed description of all samples and primary sequencing statistics are presented in Tables S1 and S2. The sampling covers the three dominant *P. patens* stages, protonema (the gametophytic two-dimensional filamentous stage emerging from the spore), gametophore (the gametophytic tridimensional leafy shoot stage) and sporophyte (the sporophytic tissue developing after sexual reproduction that forms spores by meiosis). It must be noted that the age of the protonema at harvest varies from 7 to 21 days. As gametophore buds typically start to emerge after 7 days of growth, most of the protonemal samples are a mixture of protonemal cells and gametophore cells (for detailed harvesting times for each experiment, see Table S2). With this time criteria, the samples VII, XI, XVIII, XIX and XXI–XXIV are the only samples that are potentially pure protonema. The sequencing output (raw sequenced reads) was analyzed *in silico* using the standardized procedure schematized in Figure 1. Overall, 4.2 billion raw reads were generated, with each condition represented by 76–150 million raw reads. A total of 99.02% of the reads were mapped successfully to the genomes of *P. patens* (nuclear, chloroplast and mitochondrial). Furthermore, 90.04% of the reads mapped uniquely to the *P. patens* nuclear genome V3, and were used for further data analyses. After mapping, 22 610–26 012 out of 34 361 gene models of the *P. patens* v3.3 genome annotation, i.e. 65.8–75.7% of the gene models, are observed with more than one read. All conditions considered, more than 80% of all predicted gene models are detected with more than one read. Subsequently, normalized counts (reads per kilobase of transcript

per million mapped reads, RPKM) were calculated for each individual gene model (for the full RPKM dataset, see Data S1).

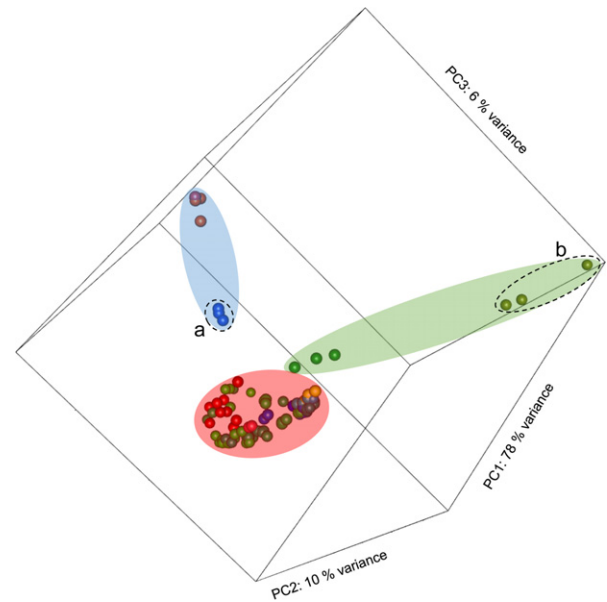
A principal component analysis (PCA) performed with the RPKM normalized counts of all libraries (Figure 2) allows the detection of three major sample clusters. The largest cluster (circled in red) is formed by the protonema and gametophore samples, regardless of the perturbation (except ABA/drought). The second distinct cluster comprises the six sporophytic samples (circled in green). Finally, both ABA treatment and dehydrated/rehydrated gametophore samples form a third cluster (circled in blue), probably linked to water stress and its



**Figure 1.** RNA-seq data analysis. Diagram illustrating the sequential RNA-seq data treatment from raw read to differentially expressed genes (DEGs). For details, see Experimental procedures.

hormonal signal integrator, ABA. Biological replicates should be tightly grouped, and for most replicates this is the case (for example, see the triplicate of experiment XIX in Figure 2, indicated by dotted ellipse a). Yet, note that several triplicates are more scattered than expected (for example, see the triplicate of experiment XVI in Figure 2, indicated by dotted ellipse b), potentially making the comparison between experiments challenging, particularly within some of the protonemal treatments (red ellipse in Figure 2). Finally, to complement and confirm the expected experimental sample clustering, we performed a hierarchical clustering of all 99 RNA-seq samples (Figure S1). Here, 95% of the replicates grouped properly. The exception are restricted to two groups of closely related samples (V and VIII; XII and XIII) that form two clusters of six libraries, but do not group by experiment. Also, the clustering of experiment XI is scattered, suggesting a potential problem with these samples.

The last computing step of our pipeline (Figure 1) was the detection of differentially expressed genes (DEGs) between experiments. The DEGs were called using a strict consensus approach of three callers (for computational details, see Experimental procedures). Overall, 50 relevant experiment comparisons were generated (for a general overview, see Table S3). The complete list of detected DEGs is shown in Data S1, next to the individual RPKM library counts. The highest number of DEGs were detected in experiments associated with very strong perturbations, such as: gametophore compared with dehydrated gametophore, with 9305 DEGs (experiment XIII compared with XVII); protoplast compared with protonema, with 7746 DEGs (experiment VIII compared with IX); or protonema compared with ABA-treated protonema, with 6940 DEGs (experiment XIX compared with XVIII). At the other extreme, a few comparisons displayed a very limited number of DEGs. Of note, the comparison between dehydrated and rehydrated gametophore showed only 10 DEGs (experiment XIII compared with XII). The treatment itself is not lethal and the gametophores begin to grow again after the treatment; however, the 2 h of rehydration prior to harvesting is probably too short a time to generate significant transcriptional changes. More puzzling is the detection of a single DEG between the tissue treated with the strigolactone analog G24 and its solvent control (experiment V compared with XXXVIII). This 24-h treatment has been shown to affect transcript accumulation and tissue morphology in *P. patens* (Hoffmann *et al.*, 2014; Decker *et al.*, 2017), as well as in angiosperms (in *A. thaliana*; Mashiguchi *et al.*, 2009) and in *Solanum lycopersicum* (tomato; Mayzlish-Gati *et al.*, 2010), for example, indicating that the assay may not have worked properly for this specific treatment. On the other hand, the actual *P. patens* regulatory network under strigolactone influence appears reduced in



**Figure 2.** Principal component analysis (PCA) of reads per kilobase of transcript per million mapped reads (RPKM) values for the 99 libraries of this study.

Each dot represents one library. Dots for each experiment have the same color. The red ellipse highlights most of the gametophore experiments. The green ellipse highlights the sporophyte experiments. The blue ellipse indicates strong stress experiments. Dotted ellipse a highlights an experiment with tightly grouped triplicate results. Dotted ellipse b highlights an experiment with more loosely grouped triplicate results.

size compared with those of other hormones (Waldie *et al.*, 2014), and the detection of specific transcript accumulation variation upon strigolactone treatment is dependent on light conditions during growth as well as on the endogenous level of strigolactone (Lopez-Obando *et al.*, 2016). In this context it is possible that 3-week-old tissue could be insensitive to strigolactone treatment. Compared with most of the other comparisons with higher numbers of DEGs, these two cases with almost no detected DEGs show that near-perfect replication can be achieved with such comparative experiments.

### Stage-specific transcriptome

Protonema, gametophore and sporophyte are the three dominant life stages of *P. patens*. We choose experiments VII, XX and XV (Tables S1 and S2) as representative of these tissues based on three criteria. First, all cultures were performed on Knop medium. Second, the timing was strictly controlled, particularly the harvesting time for protonema, which was at 7 days to ensure an absence of early gametophore development. Finally, the protonemata were visually checked for the absence of gametophores. To gain an overview of the differences between the three most abundant *P. patens* tissues, we performed a gene ontology (GO) term enrichment analysis on the pairwise up- or downregulated DEGs between these tissues (Data S1 for



DEG; Data S2 for GO term list). The most obvious detectable signals are differences in metabolism, as illustrated in the word cloud in Figure S2. Foremost, the reduction in photosynthetic activity in the sporophyte compared with both protonema (Figure S2b) and gametophore (Figure S2d) is easily observable. The two most abundantly enriched GO terms among DEGs of lower abundance in sporophytic tissue, compared with both gametophytic tissues, are identical: photosynthesis and photosynthesis light reaction. Together with other terms directly linked to photosynthesis, such as photosystem assembly, the generation of precursor metabolites and energy, or plastid organization, they dominate the term list associated with downregulated transcripts in sporophytic tissue. The protonemata–sporophyte comparison complements and validates the previously observed pattern between gametophore and sporophyte in *P. patens* (O'Donoghue *et al.*, 2013), and in another moss, *Funaria hygrometrica* (Szövényi *et al.*, 2011). This trend is in line with the known nutritional dependency of the sporophyte on the gametophore.

The GO term analysis also detects the sporophyte-specific upregulation of a carbon consumption-related pathway, which has been described previously in *P. patens* (O'Donoghue *et al.*, 2013). Compared with protonemata and gametophore, carbohydrate metabolism is the most over-represented term in the upregulated transcripts of the sporophyte; however, the type of carbon use appears to differ between the two comparisons. Terms associated with fatty acid (metabolism, biosynthesis or general lipid metabolism) characterize the difference between sporophyte and protonema (Figure S2c), whereas terms associated with coumarin (biosynthesis and metabolism) are abundant in the over-accumulated transcripts in sporophyte as compared with gametophore (Figure S2d). Furthermore, the term coumarin covers the biosynthesis of the phenylpropanoids, a large group of secondary metabolites with protective functions such as lignin precursors or sporopollenin (Colpitts *et al.*, 2011; Daku *et al.*, 2016; Niklas *et al.*, 2017; de Vries *et al.*, 2017), all of which are enriched in or specific to sporophytes.

The dominant GO term enrichment between protonemata and gametophore is linked to carbon fixation and use. Carbon metabolism is the most frequent GO term associated with the upregulated transcripts in protonema (Figure S2e). The nature of this 7-day-old tissue, young cells dividing and expanding constantly, requires a carbon conversion to cell wall compounds (e.g. the GO term external encapsulating structure organization) that is not present in most of the more mature cells of gametophores. At the same time, GO terms associated with lipid, amino acid and nucleic acid biosynthesis are linked to upregulated protonemal DEGs, all indicating actively growing tissue. Overall, a similar signal was detected previously (Xiao

*et al.*, 2011) between 3-day-old and 14-day-old protonemal culture. In contrast, GO terms associated with photosynthesis dominate the list of low-protonemal/high-gametophore abundance transcripts (Figure S2f). These GO terms reflect the fact that in contrast to protonema, the leafy gametophore is a mature structure dedicated to photosynthesis as a principal function. Photosynthates are not only used to maintain the viability of the tissue, but will also be used to feed the sporophyte during development (Hiss *et al.*, 2014; Regmi *et al.*, 2017).

From the GO term enrichment analysis to the single transcript level, the challenge to define stage-specific transcripts resides in the fact that even a transcript that is highly abundant in a given stage, for example the sporophyte, and absent in others, e.g. protonema or gametophore grown in standard growth conditions, can be induced by a variation of the growth conditions. For example, the transcript Pp3c7\_6750V3.1, which encodes a Ferritin-like domain-containing protein, displays a very high accumulation in sporophyte (RPKM > 1500), and is below detection level in protonema and gametophore under standard growing conditions; however, this transcript can be induced in protonema treated with ABA or in dehydrated gametophores to even higher levels than in the sporophyte (>3000 RPKM in both cases). Hence, we used two criteria to define stage-specific transcripts: (i) presence in the given tissue (RPKM > 2) and absence in the other two tissues (RPKM = 0–2); and (ii) absence in all other samples across the dataset that do not contain the specific tissue. Figure 3 displays the six transcripts selected to represent protonema, gametophore and sporophyte stages (two each) using the present dataset. Pp3c2\_4100V3.1 encodes one copy (out of 25 in *P. patens*) of the ribulose-bisphosphate carboxylase small chain (rbcS) protein that appears to be specifically expressed in protonemal tissue. The protonemal cell wall is essentially formed of primary cell wall that provides reduced protection to light, and the specific high expression of photosynthesis proteins with high turnover may be a way to cope with this higher light exposure. The other protonema-specific selected transcript, Pp3c1\_10720V3.1, encodes a protein without known annotation and only detected in bryophytes by the Phytozome gene ancestry list. Its specific presence in protonema probably explains the lack of data for it, but makes it a good marker for such tissue. Pp3c26\_11940V3.1, specifically expressed in gametophores, encodes a putative SF7 – FASCICLIN-LIKE ARABINOGALACTAN PROTEIN 11, a cell-wall component in a group that appears to be specific to bryophytes too. Pp3c26\_11940V3.1 is one of the four homologs coding for such proteins showing similar accumulation patterns, but the only one indicating a complete specificity to gametophore tissue. Pp3c7\_9490V3.1 encodes a carbonic anhydrase:dioscorin precursor protein, and accumulates specifically in gametophores. Finally, transcripts

specific to sporophyte tissue are more abundant, with more than 150 transcripts found in that tissue only. General morphology (capsule and seta are the only enclosed multilayer tissues in moss), the unique presence of meiosis and the generation of secondary metabolites, such as oleosin and sporopollenin (Daku *et al.*, 2016; Hiss *et al.*, 2017), may explain this observation. Pp3c6\_15559V3.1, the first sporophyte-specific transcript chosen, reflects a potential metabolic need (the transport of carbohydrate across the sporophyte), as it encodes for a member of the Nodulin-like protein family (Denancé *et al.*, 2014). These integral proteins are known to transport carbohydrates such as sucrose across membranes, and thus allow the optimal allocation of reserve between cells. The second transcript selected for sporophyte identity is Pp3c5\_26440V3.1, which encodes for the MKN1-3 protein, a class-II knotted1-like homeobox transcription factor (Champagne and Ashton, 2001). This gene has been extensively studied in *P. patens* (Singer and Ashton, 2007; Sakakibara *et al.*, 2008) and is involved in sporophyte patterning, a developmental network specific to this organ. These six tissue markers were part of previously conducted microarray experiments and were analyzed from a tissue-specific perspective (Hiss *et al.*, 2014, 2017; Ortiz-Ramirez *et al.*, 2016). Although they all were confirmed as expressed in the respective tissue, the tissue specificity does not match perfectly with the present dataset, except for the sporophyte-specific genes (Table S4). Notably, non-tissue-specific expression was detected for the four gametophytic markers (in sporophytic tissue). The differences in both tissue preparations and technologies (in particular the higher sensitivity of RNA-seq) may be the cause of these varied expression patterns.

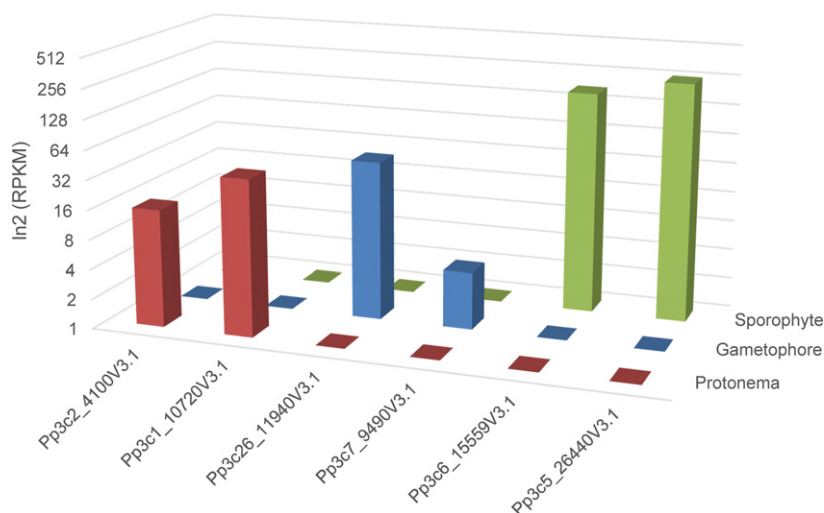
#### Impact of ammonium supplementation on the protonemal liquid culture transcriptome

Comparison between *P. patens* liquid cultures grown under near-identical conditions except for the source of nitrogen (experiment XXIII, with Knop medium supplemented with 5 mM ammonium tartrate and 4.2 mM nitrate, compared with experiment XXIV, with standard Knop medium containing only 4.2 mM nitrate) yielded 357 DEGs with a greater than twofold change, with 289 DEGs downregulated and 68 DEGs upregulated by ammonium supplementation (Data S1). The GO term enrichment analysis performed on each subset concurs with the well-known plant response to ammonium supplementation (Data S3). The addition of ammonium in the medium is clearly reflected by the repression of genes involved in nitrate assimilation and metabolism, as it is generating an accumulation of transcripts related to primary carbon metabolism (see Figure 4a and b for the 15 most abundant GO terms present in the down- and upregulated DEGs induced by ammonium supplementation; for the complete set of enriched GO terms, see Figure S3a and b).

More specifically, the effect of the addition of  $\text{NH}_4$  to the medium corresponds with previous studies: the gene expression associated with nitrate cell import, nitrate primary metabolism and some associated genes is strongly reduced, in some cases to the absence of detectable transcripts. Tsujimoto *et al.* (2007) analyzed nitrate transporter transcript accumulation under different nitrogen sources in *P. patens*, and their results are recapitulated in the present dataset: *NRT2* and the *Nar2* nitrate transporter family members show strong downregulation upon treatment with ammonium (Figure 4c; Tsujimoto *et al.*, 2007). Plant nitrate to ammonium conversion is an energetically costly process, hence upon ammonium supplementation both necessary enzymes are transcriptionally repressed in all plants analyzed (Hachiya and Sakakibara, 2017). This pattern is also detected in the present dataset where both nitrate reductase coding genes (Pp3c10\_9670V3.1, Pp3c10\_9540V3.1, and Pp3c14\_9410V3.1) and nitrite reductase coding genes (Pp3c27\_6610V3.1 and Pp3c16\_15880V3.1) are strongly inhibited by ammonium supplementation (Figure S4a).

On the other hand, we also observe the loss of transcript abundance for genes involved in ammonium assimilation: both ammonium transport genes, *AMT2s* (Pp3c18\_18460V3.1, Pp3s397\_40V3.1, Pp3c16\_12080V3.1 and Pp3c18\_18460V3.1) and primary ammonium assimilation genes, glutamate synthase (Pp3c14\_8740V3.1), glutamine synthetase (Pp3c18\_10780V3.1) and asparagine synthetase (Pp3c20\_17620V3.1) display a reduction of transcript abundance upon treatment with ammonium (Figure S4a). This reduction may be the result of the ammonium concentration used in the experimental setting (5 mM ammonium tartrate), a concentration high enough that it may require the overall regulatory repression of ammonium metabolism.

Indirect effects of different trophic conditions are also detected in this dataset. The two genes most induced by the ammonium treatment (>30-fold), Pp3c20\_19940V3.1 and Pp3c20\_1770V3.1, belong to transporter gene families involved in salt and metabolite homeostasis. Pp3c20\_19940V3.1 encodes a gene coding for an  $\text{Na}^+$  P-type ATPase protein, demonstrated to be necessary for active  $\text{Na}^+$  cell export in *P. patens* (Lunde *et al.*, 2007). The repression of nitrate import under ammonium supplementation affects  $\text{K}^+$  import (Coskun *et al.*, 2015), and hence the cytoplasmic  $\text{Na}^+/\text{K}^+$  ratio may be adjusted as a result of this specific transcript increase. Pp3c20\_1770V3.1 encodes for a member of the nodulin family. The solute specificity is not well established for all members of this family (Denancé *et al.*, 2014), but specific homologs of nodulin genes in angiosperm are notably involved in amino acid transport (Ladwig *et al.*, 2012). Thus, an increase in Pp3c20\_1770V3.1 abundance hints at amino acid relocation upon treatment with ammonium.



**Figure 3.** Stage-specific transcript markers.

Two gene models for protonema, gametophore and sporophyte tissue were selected for their specific presence at the stage and for their absence in all experiments not containing exclusively the specific tissue. Reads per kilobase of transcript per million mapped reads (RPKM) values from experiments VII (protonema), XX (gametophore) and XV (sporophyte).

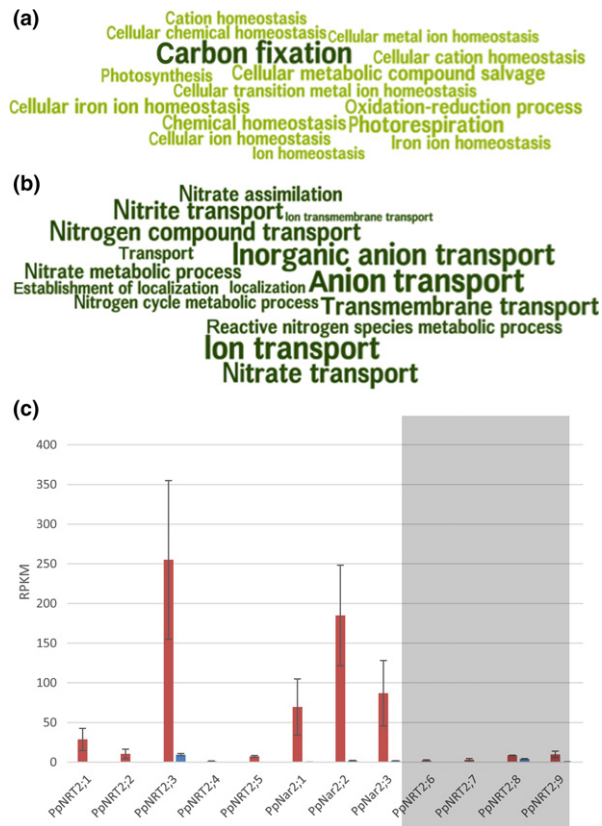
In parallel with nitrate and ammonium-related processes, a cluster of genes associated with cell wall modification appears downregulated under ammonium supplementation. This repression may reflect the morphological change observed upon the addition of ammonium: in the presence of nitrate only, *P. patens* tip cells rapidly differentiate into caulonemal cells, the faster elongating and longer protonemal cell type (Figure S3d). In the presence of ammonium, the tip cells mostly comprise chloronemal cells of shorter size (Figure S3c). Indeed, we observed a reduction of transcript accumulation of known cell wall loosening genes, such as xyloglucan endotransglucosylase hydrolase (Pp3c16\_20960V3.1), pectin methyl esterases (Pp3c3\_30560V3.1, Pp3c3\_35240V3.1 and Pp3c4\_22420V3.1) and extensins (Pp3c16\_3130V3.1 and Pp3c27\_3570V3.1) (Figure S3b), all involved in cell elongation and the modification of cell shape (Lampert *et al.*, 2011; Cosgrove, 2016).

The addition of ammonium also promotes transcript over-accumulation. Although this list is much shorter, we can detect a trend to primary carbon metabolism genes such as Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBISCO) small subunit (Pp3c15\_22730V3.1, Pp3c12\_7010V3.1 and Pp3c3\_12530V3.1) or carbonate dehydratase (Pp3c26\_6810V3.1 and Pp3c26\_6990V3.1) (Figure S3c). Carbon and nitrogen metabolism are closely linked (Coskun *et al.*, 2016). The increase of primary carbon metabolism-associated transcripts in *P. patens* is similar to what is observed in angiosperms. This expression bias is also reflected in the difference in appearance of the *P. patens* tissue between the two conditions. The chloronemal cells are dominant under the presence of ammonium, in which smaller cells are filled with numerous chloroplasts, whereas in nitrate-only conditions the tissue is dominated by caulonemal cells displaying a reduced number of chloroplasts (Figure S3d and c, respectively).

### Intra- and inter-laboratory comparison

The present dataset allows us to compare identical experiments performed within a single laboratory and between two different laboratories. Experiments VII, XXI and XXII (Tables S1 and S2) are true replicates. Unfortunately, experiment XXI only has two replicates, as mentioned previously. Hence the comparison of results must be regarded as indicative of a trend only. Each of the two laboratories involved generated protonemal liquid culture in Knop medium, and sampling was performed at day 7 after tissue homogenization. The RNA was subsequently extracted by the laboratory and sent to the JGI for uniform library construction and sequencing. Hence, the two main sources of variation are the cultures themselves and the RNA extraction. Samples from experiments VII and XXI were isolated using the same standardized protocol (see Experimental procedures); in experiment XXII the Trizol step was omitted. RNA samples passed rigorous quality control (carried out by a single lab) prior to library construction. Post-sequencing quality control indicates no major differences in read length, GC content and the total number of sequenced reads (for the read length profile and number for each libraries, see Figure S5), indicating that differences detected should be mostly attributed to laboratory (culture) variation.

The within-laboratory comparison, experiment XXII compared with XXI, displays a limited number of DEGs: 42 with a fold change of  $>0-2$ , equally distributed between up- (23) and downregulated (19) DEGs (Data S1). This low number could be attributed at least partially to the absence of a biological triplicate for experiment XXI or to the omission of Trizol in experiment XXII. Nevertheless, the amplitude of the variation remains in a single order of magnitude, contrary to most of the other comparisons in which the calculated fold change can span up to five



**Figure 4.** Impact of ammonium supplementation on transcriptome. Gene ontology (GO) term analysis representation of the 15 most over-represented GO terms in the up- (a) and downregulated (b) differentially expressed genes (DEGs) in ammonium-supplemented liquid protonemal culture (experiment XXIII), compared with ammonium-free liquid protonemal culture (experiment XXIV). The size of the words is proportional to the  $-\log_{10}$  (q value), and over-represented GO terms were colored dark green if  $-\log_{10}$  (q value)  $\leq 4$  and were colored light green if  $-\log_{10}$  (q value)  $> 4$ . For the GO term identities and their respective over-representation values, see Data S3. (c) RPKM values for nitrate and nitrite transporter gene models in the absence of ammonium (red bar) and in the presence of ammonium (blue bar). Gene models not identified by Tsujimoto *et al.* (2007) are shaded in gray. Identifiers of the different nitrate transporter genes described by the *P. patens* v3.3 genome from Tsujimoto *et al.* (2007) read as follows: PpNRT2:1/Pp3c22\_21990V3.1, PpNRT2:2/Pp3c22\_21970V3.1, PpNRT2:3/Pp3c7\_13340V3.1, PpNRT2:4/Pp3c19\_10950V3.1, PpNRT2:5/Pp3c22\_9060V3.1, PpNar2:1/Pp3c21\_13230V3.1, PpNar2:2/Pp3c18\_3270V3.1, PpNar2:3/Pp3c22\_21950V3.1, PpNRT2:6/Pp3c22\_5710V3.1, PpNRT2:7/Pp3c19\_10820V3.1, PpNRT2:8/Pp3c19\_21550V3.1 and PpNRT2:9/Pp3c16\_10420V3.1.

orders of magnitude. The GO term enrichment analysis (Data S4; Figure S6) performed on these DEGs point to a potential source of trophic variation: experiment XXII displays an increase of GO terms associated with a response to external nitrogen processes (nitrate transport and response to nitrate) as well as to cell death and protein recycling (regulation of cell death, positive regulation of cell death or regulation of cellular processes). On the other hand, iron import-related terms dominate the downregulated DEGs. Together, these terms point to a potential

nitrogen source depletion leading to metabolic and metal homeostasis redirection. As both iron and nitrate are added separately in the medium, a slight variation in the media could potentially explain these DEGs; however, the low number of observed DEGs and their low fold change indicates high reproducibility, given the sensitivity of RNA-seq to detect minor changes in transcript abundance.

In contrast, 1262 DEGs were detected in total (727 up- and 535 downregulated) with a fold change of  $>2$  between experiments VII and XXII (Data S1). The number and the amplitude of the DEGs (up to 500) suggest a clear difference between these samples. Two major sources of variation could generate such a difference: contamination with other *P. patens* tissue and variation in growing conditions, generating a stress response. To assess these two possibilities, we compared the DEG list between the two experiments and the DEGs for protonemata compared with gametophore (experiment VII compared with XX), to test the tissue hypothesis. We also compared them with three different stress conditions: the effect of ABA on protonemata (experiment XIX compared with XXVIII); the effect of high light on protonemata plus gametophores (experiment II compared with I); and the effect of elevated temperature (heat stress) on protonemata plus gametophores (experiment XXV compared with XXVI). Figure 5 illustrates that the observed differences can indeed come from different sources: 1101 of 1262 DEGs detected in the inter-laboratory comparison can be seen in other experiments. Focusing on the comparison with developmental stage only, 680 (54%) of these DEGs are also seen in the DEGs identified by comparing protonema against gametophore. A clear example of the presence of gametophore are transcripts for Pp3c27\_3570V3.1, a gene coding for a putative extensin precursor that displays a 250-fold increase between protonemal and gametophore tissues, and shows a more than 64-fold difference between experiment XXI and VII. The source of such gametophore contamination could be explained in differences of weekly grinding of the tissue to maintain a pure protonemata culture. Continuously cultured protonema in liquid culture sometimes develop gametophore buds after 7 days, and thus the culture needs to be blended regularly to reset the protonema to day 1 of the culture cycle. Yet, the difference between experiments VII and XXII cannot be attributed to this kind of tissue contamination alone. The three stresses compared in the same figure, ABA treatment, heat and high light, also display DEG overlap. Each stress displays a specific signature as well as overlap with other experiments, but we can identify 32 DEGs between the three stress conditions and the laboratory comparison that should reflect a general stress response. It is difficult to evaluate exactly the cause of these stresses, but differences between laboratories such as temperature and humidity regime of the growth chamber and type/age of



the white light system used can potentially generate the stresses detected in this comparison. It should be noted that in a within-laboratory comparison it was previously shown that liquid culture (with regular blending) as such does not seem to represent a stress condition for *P. patens* (Hiss *et al.*, 2014). The relatively large number of DEGs detected in the inter-laboratory comparison thus demonstrates the sensitivity of RNA-seq and hence the fact that minor fluctuations in growth conditions can result in clearly detectable changes to the transcriptome.

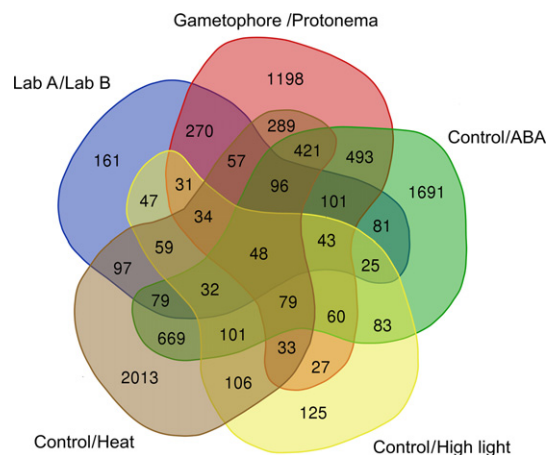
### Conclusions and outlook

The present transcriptome dataset represents an important addition to the existing expression profiling data for the moss *P. patens*. By its sample size and sequencing depth, covering more than 80% of the v3.3 *P. patens* gene models, the dataset will, along with others, permit the improvement of future gene annotation versions. The RPKM values for all individual v3.3 *P. patens* gene models in addition to 50 DEG sets are published with this study, representing a valuable benchmark reference for future RNA-seq studies. Cross-comparison across large datasets is an important approach to confirm transcript specificity to any biological phenomenon, be it developmental, as exemplified in the present study by the stage comparisons, or environmental, as indicated by the laboratory replicates. As more datasets are published, the body of data will permit a better understanding of variable parameters, but it is clear that the RNA-seq approach is sensitive enough to detect differences in growth conditions, qualitative or quantitative, that can escape careful laboratory observation. Therefore, experimental replica conditions should be very carefully controlled and documented to allow for comparison within and between laboratories. More *P. patens* Gene Atlas data are forthcoming, representing, for example, additional developmental stages such as non-germinated spores and gametophores bearing gametangia (sexual organs), as well as further perturbation experiments looking into the response of the plant to variation in phosphate concentration in the medium, which will further enhance the usefulness of the present set of expression profiling data. The data presented here are currently available as a supplement to this paper (Data S1). Moreover, expression values assigned to genes can be accessed at Phytozome (<https://phytozome.jgi.doe.gov/>). Other large-scale *P. patens* expression data are available at Genevestigator (Hiss *et al.*, 2014) and the eFP browser (Ortiz-Ramirez *et al.*, 2016). A valuable future goal is to unify these data into a single resource.

## EXPERIMENTAL PROCEDURES

### Plant material

*Physcomitrella patens* Gransden (Engel, 1968) was used for all samples apart from the two sporophyte sets, for which *P. patens*



**Figure 5.** Contrasting differentially expressed genes (DEGs) across experiments.

Venn diagram analysis of the DEGs of replica experiments between different laboratories (Laboratory A, experiment VII, compared with Laboratory B, experiment XXII), and with DEGs between protonema and gametophore (protonema, experiment VII, compared with gametophore, experiment XX), between ABA treatment (Control, experiment XVIII, compared with ABA, experiment XIX), heat treatment (Control, experiment XXV, compared with Heat, experiment XXVI) and high light treatment (Control, experiment I, compared with High light, experiment II). The number of transcripts meeting the cut-off values are contained within each section of the labeled circle (up- or downregulated by more than twofold; adjusted  $P < 0.05$ ).

Reute was used (Hiss *et al.*, 2017). The protonemata cultures were systematically entrained by two successive weeks of culture prior to treatment, in order to obtain a homogeneous culture. Medium referred to as BCD uses the composition described by Cove *et al.* (2009), and medium referred to as Knop uses the composition described by Reski and Abel (1985), based on Knop's medium (Knop, 1868). Solid medium [medium with 1% (w/v) agar] protonemal cultures were grown on top of a cellophane film to allow tissue transfer for specific treatments (e.g. with hormones), and for easy harvesting. If not otherwise mentioned, Petri dishes were sealed with parafilm during the growing period and plates and flasks were cultivated at 22°C with a 16-h light/8-h dark regime under 60–80  $\mu\text{mol m}^{-2} \text{s}^{-1}$  white light (long-day conditions). All harvests were performed in the middle of the light photoperiod (+8 h of light in long-day conditions), followed by immediate flash-freezing in liquid nitrogen, unless otherwise stated. All experiments, referred to by Roman numerals, consist of biological triplicates of the given conditions.

### Sample description

Table S1 presents a simplified version of the experiments with the associated repository references for the libraries.

### Light treatments

**Light quality.** Prior to treatment, *P. patens* protonemata were cultivated for 1 week on solidified supplemented BCD medium (BCD supplemented with 5 mM ammonium tartrate and 0.5% sucrose, for dark-treated samples, or 0.5% glucose, for light treatments). Plants were cultivated in long-day conditions. Subsequently, the cultures were transferred into the light conditions described below.

- Dark-treated samples were grown in darkness for 1–2 weeks (experiment XXIX).
- Far-red light samples were grown under continuous  $2 \mu\text{mol m}^{-2} \text{s}^{-1}$  far-red light at 720–740 nm for 1–2 weeks and then harvested (experiment XXXIII).
- Red-light samples were grown under continuous  $10 \mu\text{mol m}^{-2} \text{s}^{-1}$  red light at 660–680 nm for 1–2 weeks and then harvested (experiment XXX).
- Blue-light samples were grown under continuous  $5 \mu\text{mol m}^{-2} \text{s}^{-1}$  blue light at 460–480 nm for 1–2 weeks and then harvested (experiment XXXI).
- UV-B light samples were grown under continuous  $4 \mu\text{mol m}^{-2} \text{s}^{-1}$  white light supplemented with  $1 \mu\text{mol m}^{-2} \text{s}^{-1}$  UV-B light at 300–320 nm for 1–2 weeks and then harvested (experiment XXVIII).

**Light quantity.** Prior to treatment, *P. patens* protonemata were cultivated for 10 days under standard long-day conditions on solidified Knop medium.

- High-light samples were transferred to  $850 \mu\text{mol m}^{-2} \text{s}^{-1}$  white light for 2 h and then harvested (experiment II).
- Low-light samples were transferred to  $10 \mu\text{mol m}^{-2} \text{s}^{-1}$  white light for 2 h and then harvested (experiment III).
- Control light samples were maintained at  $70 \mu\text{mol m}^{-2} \text{s}^{-1}$  white light for 2 h and then harvested (experiment I).

## Tissues

**Germinating spores—*Physcomitrella patens*** mature sporophytes were harvested, opened and spores suspended in sterile water. Spores were distributed on Petri dishes containing solid Knop medium supplemented with 5 mM ammonium tartrate. Spores were germinated for 4 days at 24°C under continuous light and then harvested (experiment IV).

**Protonemata grown on solid medium—**Protonemata were cultivated on solidified BCD medium, in Petri dishes sealed with 3M Micropore tape under standard conditions. Protonemata were cultivated for 7 days and then harvested (experiment XVIII).

**Protonemata grown in liquid medium—**Protonemata were cultivated in liquid Knop medium in flasks with continuous shaking under standard conditions. Protonemata were cultivated for 7 days and then harvested by filtering with a 100- $\mu\text{m}$  sieve (experiments VII, XXI and XXII).

**Gametophores—**Gametophores were cultivated on solidified BCD medium without cellophane, in Petri dishes sealed with 3M Micropore tape under standard conditions. Gametophores were cultivated for 5 weeks and the aerial parts of the plants were harvested (experiment XVII).

Gametophores were cultivated on solidified Knop medium without cellophane, under standard conditions. Gametophores were grown for 5 weeks and the aerial parts of the plants were harvested (experiment XX).

Leaflets (phyllids, non-vascular leaves of the gametophore): Gametophores were cultivated on solidified Knop medium without cellophane, under standard conditions. After 5 weeks of growth, leaflets were separated from the gametophore stem using forceps, harvested and stored in RNA*later* solution (Qiagen, Hilden, Germany) before flash-freezing in liquid nitrogen as a single sample (experiment XIV).

**Sporophytes—*Physcomitrella patens*** Reute sporophytes were induced as initially described by Hohe and collaborators, and later modified by Hiss and collaborators (Hohe *et al.*, 2002; Hiss *et al.*, 2017). Briefly, gametophytic tissue was grown for 5 weeks on solidified mineral Knop medium under standard conditions. Gametangia production was induced by transferring the plates to 16°C under an 8-h light/16-h dark regime, with  $20 \mu\text{mol m}^{-2} \text{s}^{-1}$  white light (short-day conditions). After 3 weeks in this regime, plants were watered regularly to promote efficient fertilization and allowed to develop under short-day conditions.

Green sporophytes with a round capsule shape and green color were harvested after 5 weeks of short-day growth conditions (immature post meiotic stage M; Hiss *et al.*, 2017) and stored in RNA*later* solution (Qiagen) before flash-freezing in liquid nitrogen as a single sample (experiment XV).

Brown sporophytes with a round capsule shape and brown color were harvested after 7 weeks of short-day growth conditions (mature stage stage B; Hiss *et al.*, 2017), and stored in RNA*later* solution (Qiagen) before flash-freezing in liquid nitrogen as a single sample (experiment XVI).

## Hormones

**Auxin—**Gametophores were cultivated on a sieve above liquid Knop medium in Magenta boxes (Sigma-Aldrich, Merck KGaA, Darmstadt, Germany) sealed with parafilm under standard conditions for 10 months. At that time point, auxin samples were treated with 10  $\mu\text{M}$  naphthaleneacetic acid (NAA) and cultivated for 10 days before harvesting. Samples with and without NAA were generated (experiment XXXIII and XXXIV).

**Strigolactone—**Protonemata were cultivated on solidified BCD medium in Petri dishes sealed with 3M Micropore tape under standard conditions. The tissue was cultivated for 21 days. Cellophane disks containing tissue were transferred to BCD plates containing either 1  $\mu\text{M}$  racemic GR24 (synthetic strigolactone) or acetone without GR24, as a control, incubated for 24 h and then harvested (experiment V and XXXVIII).

**Gibberellin—**Protonemata were cultivated on solidified BCD medium, supplemented with 20  $\mu\text{M}$  GA<sub>3</sub>-methyl ester under standard conditions. GA<sub>3</sub>-methyl ester was synthesized and donated by Peter Hedden's group at Rothamsted Research (<https://www.rothamsted.ac.uk>). Protonemata were cultivated for 7 days and then harvested (experiment XI).

**Abscisic acid—**Protonemata were cultivated on solidified BCD medium, in Petri dishes sealed with 3M Micropore tape under standard conditions. Protonemata were cultivated for 6 days. Cellophane disks containing tissue were transferred to BCD plates containing 50  $\mu\text{M}$  abscisic acid (ABA) and incubated for 24 h before harvesting (experiment XIX).

**OPDA—**Protonemata were cultivated on solidified BCD medium, supplemented with 5 mM ammonium tartrate, with or without 50  $\mu\text{M}$  12-oxo-phytodienoic acid (OPDA), in Petri dishes sealed with 3M Micropore tape under standard conditions. Protonemata were cultivated for 14 days. Cellophane disks containing tissue were transferred to ammonium tartrate-containing BCD plates, with or without 50  $\mu\text{M}$  OPDA and incubated for 6 h before harvesting (experiments X and IX, respectively).

## Perturbations

**Protoplasts**—Protonemata were cultivated for 6 days on solidified mineral medium BCD, supplemented with ammonium tartrate (2.7 mM) and glucose (0.5%), in Petri dishes sealed with 3M Micropore tape under standard conditions. Protoplasts were released using driselase treatment (Cove *et al.*, 2009) and then harvested (experiment VIII).

**Ammonium treatment**—Protonemata cultivated in liquid Knop medium was used to inoculate two parallel cultures, one with Knop medium and a second with Knop medium supplemented with 5 mM ammonium tartrate, in flasks with continuous shaking under standard conditions. Protonemata were cultivated for 7 days and then harvested 2 h after the lights were turned on (experiment XXIV, without ammonium; experiment XXIII, with ammonium).

**De- and rehydration**—Gametophores were cultivated on cellophane disks on solidified BCD medium under standard conditions for 5 weeks prior to dehydration treatment. The gametophores on the cellophane disks were placed in empty Petri dishes that were sealed in chambers containing an atmosphere of 91% relative humidity (RH) generated by a saturated solution of MgSO<sub>4</sub> in an incubator at 17°C, with a 16-h light/8-h dark cycle. Gametophores were exposed to the dehydrating atmosphere until they reached a constant weight (equilibrium). Equilibrium was reached at approximately 150 h (Koster *et al.*, 2010), and gametophores were sampled at 180 h (experiment XII). The water potential of the gametophore tissue at equilibrium was −13 MPa. Rehydration was achieved by floating the cellophane disks containing the dehydrated gametophores on sterile water in a Petri dish for 5 min to ensure full rehydration. Once fully rehydrated the disks were placed on solid BCD media and incubated under standard conditions in the light for 2 h before harvest (experiment XIII).

**Heat**—Protonemata were cultivated on solidified BCD medium supplemented with 5 mM ammonium tartrate under continuous light for the duration of the treatment. The heat treatment was applied after 5 days of pre-growth and lasted for 5 days, with repeated heat-shock cycles of 5 h at 22°C followed by 1 h at 37°C, before harvesting (experiment XXVI, treatment; experiment XXVII, control).

## RNA extraction, RNA processing and sequencing

Frozen samples were pulverized with a mortar and pestle and total RNA was extracted in two steps: (i) total RNA was extracted using Trizol reagent (Invitrogen, now ThermoFisher Scientific, <https://www.thermofisher.com>), using the manufacturer's instructions (maximum of 100 mg of tissue per ml of Trizol reagent); and (ii) total RNA was purified using the RNeasy Plant Mini Kit (Qiagen), omitting the shredding step of the kit. Total RNA was checked for integrity using a BioAnalyzer with an Agilent RNA 6000 Nano Chip, following the manufacturer's instructions (Agilent, <https://www.agilent.com>). Plate-based RNA sample preparation was performed on the PerkinElmer Sciclone NGS robotic liquid handling system (<http://www.perkinelmer.com>), using Illumina's TruSeq Stranded mRNA HT sample prep kit (Illumina, <https://www.illumina.com>) with poly-A selection of mRNA following the protocol outlined by Illumina in their user guide ([http://support.illumina.com/sequencing/sequencing\\_kits/truseq\\_stranded\\_mrna\\_ht\\_sample\\_prep\\_kit.html](http://support.illumina.com/sequencing/sequencing_kits/truseq_stranded_mrna_ht_sample_prep_kit.html)), and with the following conditions: total RNA starting material was 1 µg per

sample and eight cycles of PCR were used for library amplification. The prepared libraries were then quantified by qPCR using the Kapa SYBR Fast Illumina Library Quantification Kit (Kapa Biosystems, <https://www.kapabiosystems.com>), and run on a Roche LightCycler 480 real-time PCR instrument (Roche, <https://www.roche.com>). The quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform using a TruSeq paired-end cluster kit, v4, and Illumina's cBot instrument to generate a clustered flow cell for sequencing. Sequencing of the flow cell was performed on the Illumina HiSeq2500 sequencer using HiSeq TruSeq SBS sequencing kits, v4, following a 2 × 150 indexed run recipe.

## RNA-seq processing

The RNA-seq processing steps described below are presented in a condensed view in the illustrated pipeline presented in Figure 1.

**Quality trimming and adapter removal.** Each library was initially checked with FASTQC 0.11.2 to evaluate their read quality (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Subsequently, lower quality bases and sequencing adapters were removed using TRIMMOMATIC 0.33 (Bolger *et al.*, 2014) with the following parameters: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 SLIDINGWINDOW:4:15 HEADCROP:12 MINLEN:50. Finally, a read length of minimum 50 nt per read was required for further processing.

**Poly-A tail trimming.** As a result of the nature of RNA-seq data, poly-A tails are expected: poly-A tails with a minimum length of 12 were identified and removed with PRINSEQ 0.20.4 (Schmieder and Edwards, 2011).

**Paired-end read merging.** During an Illumina paired-end sequencing, fragmented RNA will be sequenced from both sides. If the fragments are smaller than the double read length, the reads overlap each other. Such overlapping reads were merged with the help of COPEREAD 1.2.5 (Liu *et al.*, 2012).

**Mapping.** The read mapping was performed using GSNAP 2015-12-31.v5 (Wu and Nacu, 2010), with the options -A sam -N 1 -split-output -failed-input. The read mapping was performed in two steps: all reads were mapped first against *P. patens* organellar genomes and rRNA sequences (mitochondrial NC\_007945.1; chloroplast NC\_005087.1; ribosomal HM751653.1, X80986.1 and X98013.1). The remaining reads were then mapped against *P. patens* genome v3 (Lang *et al.*, 2018; <https://phytozome.jgi.doe.gov/pz/portal.html>) and concordant unique mapped read pairs were retained.

**File converting.** The conversion of the mapping output files from SAM to BAM format and the sorting by positions was performed using SAMTOOLS 1.2 (Li *et al.*, 2009).

**Read count.** For read counting, HTSEQ-COUNT 0.6.1p1 (Anders *et al.*, 2015), in combination with the *P. patens* gene model v3.3 (Lang *et al.*, 2018), was applied. Additionally for default options, the following parameters were set: -s reverse -r pos -t exon -i Parent.

## Differential expression analysis

Differentially expressed gene (DEG) calling approaches can generate different results (Zhang *et al.*, 2014; Schurch *et al.*, 2016).



Hence, in order to improve confidence in the DEGs used here, several algorithms were tested. DEG analysis was performed in R 3.2.0 using three R packages: EDGER 3.14.0 (Robinson *et al.*, 2010); DESEQ2 1.12.3 (Love *et al.*, 2014); and NOISEQ 2.12.0 (Tarazona *et al.*, 2011). *P*-value cut-offs for EDGER and DESEQ2 were 0.001, and for NOISEQ the probability of differential expression ('prob') was >0.9. Genes with zero counts in all replicates were removed. The previously detected array DEGs are known to be of high quality (Hiss *et al.*, 2014). The higher sensitivity of RNA-seq based approaches often leads to the calling of DEGs that exhibit very low expression levels, the biological significance of which might be questionable. In order to rely on a trustworthy set of DEGs we decided to use the NOISEQ RPKM-normalized DEGs because they capture the majority of the array DEGs, overlap with a high fraction of DEGs also detected by EDGER and DESEQ2, but exclude a high number of DEGs detected only by the latter two tools (for a Venn diagram representation of the four-way DEG call comparison, see Figure S7), which are characterized by a particularly low average expression level (3.5 FPKM, as compared with the 51.7 FPKM average for the overlap of array and all three RNA-seq DEG callers). The number of DEGs called exclusively by NOISEQ (not overlapping with other approaches) is the lowest one of all approaches. Thus, for further analysis NOISEQ RPKM-normalized DEGs were used.

To further confirm our DEG procedure using the NOISEQ approach, experiments XX and XXI were compared with previously published microarray data (Hiss *et al.*, 2014). Both experiments on both platforms, gametophores on solid Knop medium (XX) and protonemata in liquid Knop medium (XXI), were performed in the same laboratory. A total of 620 DEGs were called by the microarray data (Hiss *et al.*, 2014). With its higher sensitivity RNA-seq called 3309 DEGs. We found that 69% of the microarray DEGs overlap with the RNA-seq data, providing evidence that the RNA-seq based DEGs coincide well with previous approaches. An even better overlap can be found by comparing the GO terms associated with both DEG sets. The microarray GO terms were found to share 95.5% of their associated terms with the GO terms associated with the RNA-seq DEGs (Data S5). A total of 758 of 762 GO terms did not show significant differences between the two DEG sets in terms of their number of associated genes (Fisher's exact test,  $P_{\text{adjust}} < 0.05$ ).

### GO term enrichment analysis

The GO enrichment analyses were conducted as described previously (Widiez *et al.*, 2014). Visualization of the GO terms was implemented using word clouds using the <http://www.wordle.net> application. Word size is proportional to the  $-\log_{10}$  (q value), and over-represented GO terms were colored dark green if  $-\log_{10}$  (q value)  $\leq 4$  and light green if  $-\log_{10}$  (q value)  $> 4$ .

### Data visualization

Principal component analysis (PCA) was performed in R 3.2.0 using the R function PRINCOMP (<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/princomp.html>). PCA visualization was generated using the R package PLOT3D 1.1.1 (<https://CRAN.R-project.org/package=plot3d>). Hierarchical clustering was calculated and visualized using R 3.4.3 and the package COMPLEXHEATMAP 1.17.1 (Gu *et al.*, 2016), with the Euclidean distance method. The calculation was performed with all gene models, with at least three samples with RPKM  $> 2$ . Venn diagrams were created using the online tool VENN DIAGRAM developed by the University of Gent (<http://bioinformatics.psb.ugent.be/webtools/Venn/>), with the symmetric and colored options.

### ACKNOWLEDGEMENTS

We would like to thank Christine Glockner (University of Freiburg), Marco Göttig (University of Marburg) and Mitsuhiro Ishibashi (Hokkaido University) for technical assistance. The work conducted by the US Department of Energy Joint Genome Institute 902 is supported by the Office of Science of the US Department of Energy under Contract No. DE-903 AC02-05CH11231. H.C. and N.S.'s work was supported by grants from the National Natural Science Foundation of China (31621001). R.B. and A.A. acknowledge the support of the EEC project S2B (675006). R.R. and S.A.R. acknowledge financial support from the Excellence Initiative of the German Federal and State Governments (EXC 294). J.C. and Y.S.'s work was supported by the UK Leverhulme Trust research grant F/0094/BA. J.C. thanks Peter Hedden for his generous gift of the GA9-methyl ester. P.S.'s work was supported by the Swiss National Science Foundation (160004 and 131726) and the URPP Evolution in Action. M.R.'s work was supported by the Plant Fellows program and the 'Forschungskredit der Universität Zurich'. S.B., A.G., F.N., and C.R. acknowledge the support of the LabEx Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

### CONFLICT OF INTEREST

The authors declare no conflicts of interest.

### SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

**Figure S1.** Hierarchical clustering of all 99 RNA-seq samples, RPKM normalized. The upper colored line represents the experiments, the lower one the corresponding tissues. This analysis confirms and illustrates that the replicates for each experiment cluster as expected with each other in most of the cases, as for example experiment XVI (brown sporophyte). The exceptions are the cases in which consequently few DEGs were detected. For example, experiments V and XXXVIII libraries group together, indication of the closeness of these samples, but the triplicates are not resolved between the two experiments. The clustering provides independent confirmation of the absence of effect of the strigolactone treatments in this specific experiment.

**Figure S2.** GO term enrichment of developmental stage DEGs. GO term analysis representation of all over-represented GO terms associated with DEGs between specific stages. (A): Over-represented GO terms associated with the up-DEGs in sporophyte (exp. XV) compared to gametophore (exp. XVII). (B): Over-represented GO terms associated with the down-DEGs in sporophyte (exp. XV) compared to gametophore (exp. XVII). (C): GO terms associated with the up-DEGs in protonema (exp. VII) compared to sporophyte (exp. XV). (D): GO terms associated with the down-DEGs in protonema (exp. VII) compared to sporophyte (exp. XV). (E): GO terms associated with the up-DEGs in protonema (exp. VII) compared to gametophore (exp. XV). (F): GO terms associated with the down-DEGs in protonema (exp. VII) compared to gametophore (exp. XV). Word size is proportional to the  $-\log_{10}$  (q-value) and over-represented GO terms were colored dark green if  $-\log_{10}$  (q-value)  $\leq 4$  and light green if  $-\log_{10}$  (q-value)  $> 4$ . For the GO term IDs and their respective over-representation values see Data S2.

**Figure S3.** Impact of ammonium supplementation on protonemata transcriptome. GO term analysis representation of all over-represented GO terms associated with DEGs between ammonium-supplemented liquid protonemal culture (exp. XXIII) and ammonium-free liquid protonemal culture (exp. XXIV). (A): Over-represented GO terms associated with the up-DEGs in ammonium-



supplemented liquid protonemal culture (exp. XXIII) compared to ammonium-free liquid protonemal culture (exp. XXIV). (B): Over-represented GO terms associated with the down-DEGs in ammonium-supplemented liquid protonemal culture (exp. XXIII) compared to ammonium-free liquid protonemal culture (exp. XXIV). Word size is proportional to the  $-\log_{10}$  (q-value) and over-represented GO terms were colored dark green if  $-\log_{10}$  (q-value)  $\leq 4$  and light green if  $-\log_{10}$  (q-value)  $> 4$ . (C): Typical aspect of protonemal cells in presence of ammonium. (D): Typical aspect of protonemal cells in absence of ammonium. For the GO term IDs and their respective over-representation values see Data S3.

**Figure S4.** Specific transcripts affected by ammonium supplementation in protonemata. (A): Repressed transcripts upon ammonium supplementation pertaining directly to nitrate metabolism. (B): Repressed transcripts upon ammonium supplementation related to cell wall remodeling. (C): Induced gene upon ammonium supplementation related to primary carbon fixation metabolism. Error bar: standard deviation,  $n=3$ .

**Figure S5.** Library quality evaluation of the laboratory comparison dataset. Read length distribution of for the different libraries. (A): 7805.5.84013.TCGGCA, (B): 7806.3.84002.ACAAA, (C): 7806.5.84005.CGAGAA, (D): 7806.6.83999.TGAATG, (E): 7805.1.84009.CTTGTA, (F): 7805.1.84009.CCGTCC, (G): 7805.2.84012.ATGAGC, (H): 7806.6.83999.TTCGAA, (I): Overview of the library primary statistic for experiments VII, XXI and XXII.

**Figure S6.** Comparison between replica experiments performed in a single laboratory. GO term analysis representation of all over-represented GO terms associated with DEGs between the protonemal replica experiment XXII and XXI. (A): Over-represented GO terms associated with up-DEGs in experiment XXII compared to experiment XXI. (B): Over-represented GO terms associated with down-DEGs in experiment XXII compared to experiment XXI. Word size is proportional to the  $-\log_{10}$  (q-value) and over-represented GO terms were colored dark green if  $-\log_{10}$  (q-value)  $\leq 4$  and light green if  $-\log_{10}$  (q-value)  $> 4$ . For the GO terms IDs and their respective overrepresentation values see Data S4.

**Figure S7.** Comparison of the DEGs called by the NOISeq, DESeq2 and edgeR packages with RNA dataset and by microarray approach. Venn diagram comparing the DEGs called by NOISeq (in blue), DESeq2 (in yellow) and edgeR (in green) between the Experiments XXI (gametophore) and XX (protonemal liquid culture) and the DEGs called in a microarray experiment performed on the same tissues (Hiss et al. 2014).

**Table S1.** Overview of the experiments and their primary library data presented in this study.

**Table S2.** Harvesting time point after initiation of the specific culture and experimental location for each experiment in this study.

**Table S3.** Overview of the experiment pairs for which DEGs have been calculated in the present study.

**Table S4: Tissue markers detection in *P. patens* microarray studies.** nt: not tested; - no specificity; +/-: gene model displays in specific tissue enrichment but present somewhere else; +: tissue specificity confirmed with no other tested tissue showing any transcript accumulation.

**Data S1.** Calculated RPKM for all *Physcomitrella patens* gene models V3.3 in all libraries described in the present study and calculated DEG for 50 relevant comparisons.

**Data S2.** GO terms enrichment lists associated with ammonium supplementation.

**Data S3.** GO terms enrichment lists associated with stage specific comparison.

**Data S4.** GO terms enrichment lists associated lab specific replication.

**Data S5.** GO terms comparison between the RNA-seq experiment XX and XXI and the similar treatment performed with microarray approach (data from Hiss et al. 2014).

## REFERENCES

- Alaba, S., Piszczalka, P., Pietrykowska, H. et al. (2015) The liverwort *Pellia endiviifolia* shares microtranscriptomic traits that are common to green algae and land plants. *New Phytol.* **206**, 352–367.
- Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
- Aya, K., Kobayashi, M., Tanaka, J., Ohyanagi, H., Suzuki, T., Yano, K., Takano, T., Yano, K. and Matsuoka, M. (2015) De novo transcriptome assembly of a fern, *Lygodium japonicum*, and a web resource database, ljtans DB. *Plant Cell Physiol.* **56**, e5.
- Banks, J.A., Nishiyama, T., Hasebe, M. et al. (2011) The compact *Selaginella genome* identifies changes in gene content associated with the evolution of vascular plants. *Science (New York, N.Y.)*, **332**, 960–963.
- Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
- Bowman, J.L., Kohchi, T., Yamato, K.T. et al. (2017) Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell*, **171** (287–304), e215.
- Bushart, T.J., Cannon, A.E., ul Haque, A., San Miguel, P., Mostajeran, K., Clark, G.B., Porterfield, D.M. and Roux, S.J. (2013) RNA-seq analysis identifies potential modulators of gravity response in spores of *Ceratopteris* (Parkeriaceae): evidence for modulation by calcium pumps and apyrase activity. *Am. J. Bot.* **100**, 161–174.
- Champagne, C.E.M. and Ashton, N.W. (2001) Ancestry of KNOX genes revealed by bryophyte (*Physcomitrella patens*) homologs. *New Phytol.* **150**, 23–36.
- Chang, C.-Y., Lin, W.-D. and Tu, S.-L. (2014) Genome-wide analysis of heat-sensitive alternative splicing in *Physcomitrella patens*. *Plant Physiol.* **165**, 826–840.
- Chen, Y.-R., Su, Y.-S. and Tu, S.-L. (2012) Distinct phytochrome actions in nonvascular plants revealed by targeted inactivation of phytylbin biosynthesis. *Proc. Natl Acad. Sci. USA*, **109**, 8310–8315.
- Colpitts, C.C., Kim, S.S., Posehn, S.E., Jepson, C., Kim, S.Y., Wiedemann, G., Reski, R., Wee, A.G.H., Douglas, C.J. and Suh, D.-Y. (2011) PpASCL, a moss ortholog of anther-specific chalcone synthase-like enzymes, is a hydroxyalkylpyrone synthase involved in an evolutionarily conserved sporopollenin biosynthesis pathway. *New Phytol.* **192**, 855–868.
- Coruh, C., Cho, S.H., Shahid, S., Liu, Q., Wierzbicki, A. and Axtell, M.J. (2015) Comprehensive annotation of *Physcomitrella patens* small RNA loci reveals that the heterochromatic short interfering RNA pathway is largely conserved in land plants. *Plant Cell*, **27**, 2148–2162.
- Cosgrove, D.J. (2016) Catalysts of plant cell wall loosening. *F1000Res*, **5**, 119. <https://doi.org/10.12688/f1000research.7180.1>
- Coskun, D., Britto, D.T. and Kronzucker, H.J. (2015) The nitrogen-potassium intersection: membranes, metabolism, and mechanism. *Plant, Cell Environ.* **40**, 2029–2041.
- Coskun, D., Britto, D.T. and Kronzucker, H.J. (2016) Nutrient constraints on terrestrial carbon fixation: the role of nitrogen. *J. Plant Physiol.* **203**, 95–109.
- Cove, D.J., Perroud, P.F., Charron, A.J., McDaniel, S.F., Khandelwal, A. and Quatrano, R.S. (2009) The moss *Physcomitrella patens*: a novel model system for plant development and genomic studies. *Cold Spring Harbor Protocol*, **2009**, pdb emo115.
- Daku, R.M., Rabbi, F., Buttigieg, J., Coulson, I.M., Horne, D., Martens, G., Ashton, N.W. and Suh, D.Y. (2016) PpASCL, the *Physcomitrella patens* anther-specific chalcone synthase-like enzyme implicated in sporopollenin biosynthesis, is needed for integrity of the moss spore wall and spore viability. *PLoS ONE*, **11**, e0146817.
- Decker, E.L., Alder, A., Hunn, S. et al. (2017) Strigolactone biosynthesis is evolutionarily conserved, regulated by phosphate starvation and contributes to resistance against phytopathogenic fungi in a moss, *Physcomitrella patens*. *New Phytol.* **216**, 455–468.
- Demko, V., Perroud, P.F., Johansen, W. et al. (2014) Genetic analysis of DEFECTIVE KERNEL1 loop function in three-dimensional body patterning in *Physcomitrella patens*. *Plant Physiol.* **166**, 903–919.

- Denancé, N., Szurek, B. and Noël, L.D. (2014) Emerging functions of nodulin-like proteins in non-nodulating plant species. *Plant Cell Physiol.* **55**, 469–474.
- Devos, N., Szövényi, P., Weston, D.J., Rothfels, C.J., Johnson, M.G. and Shaw, A.J. (2016) Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta). *New Phytol.* **211**, 300–318.
- Engel, P.P. (1968) The induction of biochemical and morphological mutants in the moss *Physcomitrella patens*. *Am. J. Bot.* **55**, 438–446.
- Eveland, A.L., Satoh-Nagasawa, N., Goldshmidt, A., Meyer, S., Beatty, M., Sakai, H., Ware, D. and Jackson, D. (2010) Digital gene expression signatures for maize development. *Plant Physiol.* **154**, 1024–1039.
- Gaidatzis, D., Burger, L., Florescu, M. and Stadler, M.B. (2015) Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat. Biotechnol.* **33**, 722–729.
- Gao, B., Zhang, D., Li, X., Yang, H. and Wood, A.J. (2014) De novo assembly and characterization of the transcriptome in the desiccation-tolerant moss *Syntrichia caninervis*. *BMC Res. Notes* **7**, 490.
- Gao, B., Zhang, D., Li, X., Yang, H., Zhang, Y. and Wood, A.J. (2015) De novo transcriptome characterization and gene expression profiling of the desiccation tolerant moss *Bryum argenteum* following rehydration. *BMC Genom.* **16**, 416.
- González-Ballester, D., Casero, D., Cokus, S., Pellegrini, M., Merchant, S.S. and Grossman, A.R. (2010) RNA-seq analysis of sulfur-deprived *Chlamydomonas* cells reveals aspects of acclimation critical for cell survival. *Plant Cell*, **22**, 2058.
- Gu, Z., Eils, R. and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, **32**, 2847–2849.
- Hachiya, T. and Sakakibara, H. (2017) Interactions between nitrate and ammonium in their uptake, allocation, assimilation, and signaling in plants. *J. Exp. Bot.* **68**, 2501–2512.
- Harrison, J.C. (2017) Development and genetics in the evolution of land plant body plans. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20150490.
- Hiss, M., Laule, O., Meskauskiene, R.M. et al. (2014) Large-scale gene expression profiling data for the model moss *Physcomitrella patens* aid understanding of developmental progression, culture and stress conditions. *Plant J.* **79**, 530–539.
- Hiss, M., Meyberg, R., Westermann, J., Haas, F.B., Schneider, L., Schallenberg-Rudinger, M., Ullrich, K.K. and Rensing, S.A. (2017) Sexual reproduction, sporophyte development and molecular variation in the model moss *Physcomitrella patens*: introducing the ecotype Reute. *Plant J.* **90**, 606–620.
- Hoffmann, B., Proust, H., Belcram, K., Labruno, C., Boyer, F.D., Rameau, C. and Bonhomme, S. (2014) Strigolactones inhibit caulonema elongation and cell division in the moss *Physcomitrella patens*. *PLoS ONE*, **9**, e99206.
- Hohe, A., Rensing, S.A., Mildner, M., Lang, D. and Reski, R. (2002) Day length and temperature strongly influence sexual reproduction and expression of a novel MADS-box gene in the moss *Physcomitrella patens*. *Plant Biol.* **4**, 595–602.
- Kamisugi, Y., Whitaker, J.W. and Cuming, A.C. (2016) The transcriptional response to DNA-double-strand breaks in *Physcomitrella patens*. *PLoS ONE*, **11**, e0161204.
- Knop, W. (1868) *Der Kreislauf des Stoffs: Lehrbuch der Agricultur-Chemie Leipzig*. Germany: Haessel, H.
- Koster, K.L., Balsamo, R.A., Espinoza, C. and Oliver, M.J. (2010) Desiccation sensitivity and tolerance in the moss *Physcomitrella patens*: assessing limits and damage. *Plant Growth Regul.* **62**, 293–302.
- Ladwig, F., Stahl, M., Ludewig, U., Hirner, A.A., Hammes, U.Z., Stadler, R., Harter, K. and Koch, W. (2012) Siliques are Red1 from arabidopsis acts as a bidirectional amino acid transporter That is crucial for the amino acid homeostasis of siliques. *Plant Physiol.* **158**, 1643–1655.
- Lampert, D.T., Kieliszewski, M.J., Chen, Y. and Cannon, M.C. (2011) Role of the extensin superfamily in primary cell wall architecture. *Plant Physiol.* **156**, 11–19.
- Lang, D., Ullrich, K.K., Murat, F. et al. (2018) The *P. patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533.
- Lavy, M., Prigge, M.J., Tao, S., Shain, S., Kuo, A., Kirchsteiger, K. and Estelle, M. (2016) Constitutive auxin response in *Physcomitrella* reveals complex interactions between Aux/IAA and ARF proteins. *eLife*, **5**, e13325.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. and Ecker, J.R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, **133**, 523–536.
- Liu, B., Yuan, J., Yiu, S.M. et al. (2012) COPE: an accurate k-mer-based paired-end reads connection tool to facilitate genome assembly. *Bioinformatics*, **28**, 2870–2874.
- Lopez-Obando, M., Conn, C.E., Hoffmann, B., Bythell-Douglas, R., Nelson, D.C., Rameau, C. and Bonhomme, S. (2016) Structural modelling and transcriptional responses highlight a clade of PpKAI2-LIKE genes as candidate receptors for strigolactones in *Physcomitrella patens*. *Planta*, **243**, 1441–1453.
- Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Lunde, C., Drew, D.P., Jacobs, A.K. and Tester, M. (2007) Exclusion of Na<sup>+</sup> via sodium ATPase (PpENA1) ensures normal growth of *Physcomitrella patens* under moderate salt stress. *Plant Physiol.* **144**, 1786–1796.
- Mashiguchi, K., Sasaki, E., Shimada, Y., Nagae, M., Ueno, K., Nakano, T., Yoneyama, K., Suzuki, Y. and Asami, T. (2009) Feedback-regulation of strigolactone biosynthetic genes and strigolactone-regulated genes in Arabidopsis. *Biosci. Biotechnol. Biochem.* **73**, 2460–2465.
- Matasci, N., Hung, L.-H., Yan, Z. et al. (2014) Data access for the 1,000 Plants (1KP) project. *GigaScience*, **3**, 17.
- Mayzlish-Gati, E., LekKala, S.P., Resnick, N., Wininger, S., Bhattacharya, C., Lemcoff, J.H., Kapulnik, Y. and Koltai, H. (2010) Strigolactones are positive regulators of light-harvesting genes in tomato. *J. Exp. Bot.* **61**, 3129–3136.
- Niklas, K.J., Cobb, E.D. and Matas, A.J. (2017) The evolution of hydrophobic cell wall biopolymers: from algae to angiosperms. *J. Exp. Bot.* **68**, 5261–5269.
- 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.
- Ortiz-Ramirez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., Feijo, J.A. and Becker, J.D. (2016) A Transcriptome Atlas of *Physcomitrella patens* Provides Insights into the Evolution and Development of Land Plants. *Mol. Plant*, **9**, 205–220.
- Rama Reddy, N.R., Mehta, R.H., Soni, P.H., Makasana, J., Gajbhiye, N.A., Ponnuchamy, M. and Kumar, J. (2015) Next generation sequencing and transcriptome analysis predicts biosynthetic pathway of sennosides from senna (*Cassia angustifolia* Vahl.), a non-model plant with potent laxative properties. *PLoS ONE*, **10**, e0129422.
- Regmi, K.C., Li, L. and Gaxiola, R.A. (2017) Alternate modes of photosynthate transport in the alternating generations of *Physcomitrella patens*. *Fron. Plant Sci.* **8**, 1956.
- Renault, H., Alber, A., Horst, N.A. et al. (2017) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. *Nat. Commun.* **8**, 14713.
- Rensing, S.A., Lang, D., Zimmer, A.D. 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.
- Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
- 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. Devel.* **10**, 555–566.
- Schaefer, D.G. and Zryd, J.P. (1997) Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J.* **11**, 1195–1206.
- Schmieder, R. and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**, 863–864.

- Schurch, N.J., Schofield, P., Gierlinski, M. *et al.* (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, **22**, 839–851.
- Sharma, N., Jung, C.-H., Bhalla, P.L. and Singh, M.B. (2014) RNA Sequencing Analysis of the Gametophyte Transcriptome from the Liverwort, *Marchantia polymorpha*. *PLoS ONE*, **9**, e97497.
- 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.
- Song, L., Shankar, D.S. and Florea, L. (2016) Rascaf: improving Genome Assembly with RNA Sequencing Data. *Plant Genome*, **9**, <https://doi.org/10.3835/plantgenome2016.03.0027>.
- Stevenson, S.R., Kamisugi, Y., Trinh, C.H. *et al.* (2016) Genetic analysis of *Physcomitrella patens* identifies ABSCISIC ACID NON-RESPONSIVE, a regulator of ABA responses unique to basal land plants and required for desiccation tolerance. *Plant Cell*, **28**, 1310–1327.
- 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.
- Szövényi, P., Perroud, P.F., Symeonidi, A., Stevenson, S., Quatrano, R.S., Rensing, S.A., Cuming, A.C. and McDaniel, S.F. (2015) De novo assembly and comparative analysis of the *Ceratodon purpureus* transcriptome. *Mol. Ecol. Resour.* **15**, 203–215.
- Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. and Conesa, A. (2011) Differential expression in RNA-seq: a matter of depth. *Genome Res.* **21**, 2213–2223.
- Tsujimoto, R., Yamazaki, H., Maeda, S.-I. and Omata, T. (2007) Distinct roles of nitrate and nitrite in regulation of expression of the nitrate transport genes in the moss *Physcomitrella patens*. *Plant Cell Physiol.* **48**, 484–497.
- de Vries, J., de Vries, S., Slamovits, C.H., Rose, L.E. and Archibald, J.M. (2017) How embryophytic is the biosynthesis of phenylpropanoids and their derivatives in streptophyte algae? *Plant Cell Physiol.* **58**, 934–945.
- Waldie, T., McCulloch, H. and Leyser, O. (2014) Strigolactones and the control of plant development: lessons from shoot branching. *Plant J.* **79**, 607–622.
- Widiez, T., Symeonidi, A., Luo, C., Lam, E., Lawton, M. and Rensing, S.A. (2014) The chromatin landscape of the moss *Physcomitrella patens* and its dynamics during development and drought stress. *Plant J.* **79**, 67–81.
- Wu, T.D. and Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.
- 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.
- Zhang, G., Guo, G., Hu, X. *et al.* (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* **20**, 646–654.
- Zhang, Z.H., Jhaveri, D.J., Marshall, V.M. *et al.* (2014) A comparative study of techniques for differential expression analysis on RNA-Seq data. *PLoS ONE*, **9**, e103207.
- Zhang, Z., He, Z., Xu, S., Li, X., Guo, W., Yang, Y., Zhong, C., Zhou, R. and Shi, S. (2016) Transcriptome analyses provide insights into the phylogeny and adaptive evolution of the mangrove fern genus *Acrostichum*. *Sci. Rep.* **6**, 35634.
- Zhu, Y., Chen, L., Zhang, C., Hao, P., Jing, X. and Li, X. (2017) Global transcriptome analysis reveals extensive gene remodeling, alternative splicing and differential transcription profiles in non-seed vascular plant *Selaginella moellendorffii*. *BMC Genom.* **18**, 1042.