DR PIERRE-FRANÇOIS PERROUD (Orcid ID : 0000-0001-7607-3618)

DR FABIAN B. HAAS (Orcid ID : 0000-0002-7711-5282)

DR J. C. COATES (Orcid ID : 0000-0002-2381-0298)

PROFESSOR STEFAN A RENSING (Orcid ID : 0000-0002-0225-873X)

Article type : Resource

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

Pierre-François Perroud[1,*], Fabian B. Haas[1], Manuel Hiss[1], Kristian K. Ullrich[1, 18], Alessandro Alboresi[2, 19], Mojgan Amirebrahimi[3], Kerrie Barry[3], Roberto Bassi[2], Sandrine Bonhomme[4], Haodong Chen[5], Juliet Coates[6], Tomomichi Fujita[7], Anouchka Guyon-Debast[4], Daniel Lang[8], Junyan Lin[3], Anna Lipzen[3], Fabien Nogué[4], Melvin J. Oliver[9], Inés Ponce de León[10], Ralph S. Quatrano[11], Catherine Rameau[4], Bernd Reiss[12], Ralf Reski[13,17], Mariana Ricca[14], Younousse Saidi[6,20], Ning Sun[5], Peter Szövényi[14], Avinash Sreedasyam[15], Jane Grimwood[15], Gary Stacey[16], Jeremy Schmutz[3,15], Stefan A. Rensing[1, 17,*]

[1] Plant Cell Biology, Faculty of Biology, University of Marburg, Marburg, Germany.

[2] Dipartimento di Biotecnologie, Università di Verona, Verona, Italy.

[3] DOE Joint Genome Institute, Walnut Creek, CA, USA.

[4] Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, RD10, Versailles, France.

[5] School of Advanced Agriculture Sciences, and School of Life Sciences, Peking University, Beijing, China.

[6] School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK.

[7] Department of Biological Sciences, Hokkaido University, Hokkaido, Japan.

[8] Helmholtz Zentrum München, Munich, Germany.

[9] USDA-ARS-MWA, Plant Genetics Research Unit, University of Missouri, Columbia MO, USA.

[10] Department of Molecular Biology, Clemente Estable Biological Research Institute, Montevideo, Uruguay.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tpj.13940

[11] Department of Biology, Washington University in St Louis, St Louis, USA
[12] Max Planck Institute for Plant Breeding Research, Köln, Germany.
[13] Plant Biotechnology, Faculty of Biology, University of Freiburg, Freiburg, Germany.
[14]Department of Systematic and Evolutionary Botany, University of Zurich, Switzerland.
[15] HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA.
[16] Divisions of Plant Science and Biochemistry, National Center for Soybean Biotechnology, University of Missouri, Columbia, MO, USA.
[17] BIOSS Centre for Biological Signaling Studies, University of Freiburg, Germany.
[18] Present address: Max Planck Institute for Evolutionary Biology, Ploen, Germany.
[19] Present address: Dipartimento di Biologia, Università di Padova, Padova, Italy.
[20] Present address: Bayer Crop Science, Technologiepark, Gent (Zwijnaarde), Belgium.

* Authors for correspondence:

pierre-francois.perroud@biologie.uni-marburg.de

stefan.rensing@biologie.uni-marburg.de

Running head: A RNA-seq transcriptome for Physcomitrella patens

Keywords:

Physcomitrella patens, RNA-seq, transcriptome analysis, developmental stage, stress, differential expression

Abstract

High throughput RNA sequencing (RNA-seq) has recently become the method of choice to define and analyse transcriptomes. For the model moss *Physcomitrella patens*, although this method has been used to help the analysis of specific perturbations, no overall reference dataset has been established yet. In the framework of its 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 analyse inter-laboratory variation, as 13 different laboratories around the world provided samples. We compare the differences between single-laboratory experiment replication with the comparison of the same experiment between different laboratories.

Introduction

Since the discovery of its intrinsically efficient gene targeting (Schaefer and Zrÿd, 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 establishment of tri-dimensional growth in plants (for review (Harrison, 2017), both fields of study integrating detailed functional cell biology investigation with kingdom-wide gene and genome comparison. As these multiple research fields grew, so did their associated technical approaches. Amongst them, RNA–seq (deep sequencing of cDNA) is now dominating the field of RNA detection and quantification at the transcriptome level, replacing the 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 (Eveland *et al.*, 2010) and rice (Zhang *et al.*, 2010), for which a complete genome sequence was available. Furthermore, the improvement of *de novo* assembly of transcriptomes allowed use of this approach to characterize transcriptomes in organisms without genome sequence available (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 transcriptome. The most wide-spread *de novo* transcriptome

assembly effort so far is the 1KP project, covering more than 1,300 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 to genome sequencing projects and via *de novo* analysis. Besides *P. patens* (see below), other mosses have published transcriptomes: *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). Due to their large genome size, ferns form a group reporting 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 replication (for 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 JGI Gene Atlas Project (http://jgi.doe.gov/doe-jgi-plant-flagship-gene-atlas/). After reviewing the dataset we focus on experiment 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. 13 laboratories actively working with *P. patens* around the world contributed to the samples described in the Materials and Methods section. The detailed description of all samples and primary sequencing statistics are presented in Table S1 and Table 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. Since 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 (see Table S2 for detailed harvesting times for each experiment). With this time criteria, the samples VII, XI, XVIII, XIX and XXI to 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 to 150 million raw reads. 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 to 26,012 out of 34,361 gene models of the *P. patens* v3.3 genome annotation, i.e. 65.8 to 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 (RPKM) counts were calculated for each individual gene model (see Data S1 for the full RPKM dataset).

A Principal Component Analysis (PCA) performed with the RPKM normalized counts of all libraries (Figure 2) allows 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 hormonal signal integrator, ABA. Biological replicates are expected to be tightly grouped and for most of them this is the case (see for example triplicate of exp. XIX in Figure 2, dotted ellipse *a*). Yet, it is to note that several triplicates are more scattered than expected (see for example triplicate of exp. XVI in Figure 2, dotted ellipse *b*), potentially making the comparison between experiments challenging, particularly within some of the protonemal treatments (Figure 2 red ellipse). 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 replicas 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 (see Material and Methods for computational details). Overall, 50 relevant experiment comparisons (Table S3 for general overview) were generated. The complete list of detected DEGs are 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 to dehydrated gametophore (exp. XIII compared to XVII) with 9,305 DEGs, protoplast compared to protonema with 7,746 DEGs (exp. VIII compared to IX) or protonema compared to ABA treated protonema with 6,940 DEGs (exp. XIX compared to XVIIII). 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 (exp. XIII compared to XII). The treatment itself is not lethal and the gametophores restart to grow after the treatment. However, the two hours of rehydration prior to harvesting are probably too short to generate significant transcriptional changes. More puzzling is the detection of a single DEG between the tissue treated with the strigolactone analogue G24 and its solvent control (exp. V compared to XXXVIII). This 24 h treatment has been shown to affect transcript accumulation as well as 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) or in 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 size compared to 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 three-week old tissue could be insensitive to the strigolactone treatment. In comparison with most of the other comparisons with higher DEG number, 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 experiment VII, XVII and XV (Table 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 seven days to ensure an absence of early gametophore development. Finally, the protonemata were visually checked for the absence of gametophores. To get a bird's eye overview of the differences between the three most abundant *P. patens* tissues we performed a Gene Ontology (GO) term enrichment analysis on the pair-wise up- or down-regulated 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 to both protonema (Figure S2B) and gametophore (Figure S2D) is easily observable. The two most the two most abundant enriched GO terms among the DEGs of lower abundance in sporophytic tissue compared to both gametophytic tissues are identical: photosynthesis and photosynthesis light reaction. Together with other terms directly linked to photosynthesis, such as photosystem assembly, generation of precursor metabolites and energy or plastid organization, they are dominating the term list associated with down-regulated 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 a sporophyte-specific carbon consumption-related pathway up-regulation, which has been previously described in *P. patens* (O'Donoghue *et al.*, 2013). Compared to protonemata and gametophore, carbohydrate metabolism is the most over-represented term in the up-regulated 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), while terms associated with coumarin (biosynthesis, metabolism) are abundant in the over-accumulated transcripts in sporophyte as compared to 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, de Vries *et al.*, 2017, Niklas *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 too. Carbon metabolism is the most frequent GO term associated with the up-regulated transcripts in protonema (Figure S2E). The nature of this seven 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. Concurrently, GO terms associated with lipid, amino acid and nucleic acid biosynthesis are associated with up-regulated protonemal DEGs, all marking actively growing tissue. Overall a similar signal was detected previously (Xiao *et al.*, 2011) between three days old and 14 days old protonemal culture. In contrast, GO terms associated with photosynthesis are dominating the protonemal low / gametophore high abundance transcript list (Figure S2F). These GO terms reflect the fact that on the contrary to protonema, the leafy gametophore is a mature structure dedicated to photosynthesis as principal function. Photosynthates are used not only to maintain viability of the tissue, but will be used to feed the sporophyte during its 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 > 1,500) 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 level than in the sporophyte (> 3,000 RPKM in both cases). Hence, we used two criteria to define stage-specific transcripts: 1) presence in the given tissue (RPKM > 2) and the absence in the other two (RPKM = 0-2) and 2) 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 a reduced protection to light, specific expression of high turnover proteins of the photosystem 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 in bryophytes by Phytozome gene ancestry list. Its specific presence in protonema probably explains the lack of data about 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 pattern, but the only one indicating a complete specificity to gametophore tissue. Pp3c7_9490V3.1 encodes a carbonic anhydrase:dioscorin precursor protein and is accumulating specifically in gametophores. Finally, transcripts specific to sporophyte tissue are more abundant, with more than 150 transcripts found only in that tissue. General morphology (capsule and seta are the only enclosed multilayer tissues in moss), the unique presence of meiosis and 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 chosen sprorophyte-specific transcript, reflects a potential metabolic need, the transport of carbohydrate across the sporophyte as it encode 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 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 analyzed from a tissue specific perspective (Hiss *et al.*, 2014, Ortiz-Ramirez *et al.*, 2016, Hiss *et al.*, 2017). While they all were confirmed to be 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 pattern.

Impact of ammonium supplementation on the protonemal liquid culture transcriptome

The comparison between *P. patens* liquid cultures grown under identical conditions but for the source of nitrogen (exp. XXIII and XXIV, respectively supplemented Knop medium containing 5 mM ammonium tartrate and 4.2 mM nitrate compared to standard Knop medium containing only 4.2 mM nitrate) yielded 357 DEGs with more than a two-fold change, with 289 down-regulated DEGs and 68 up-regulated DEGs by the 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 (Figure 4A and B respectively for the 15 most abundant GO terms present in the down- and up regulated DEGs induced by the ammonium supplementation, Figure S3A and B for complete set of enriched GO terms).

More specifically, the effect of the addition of 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 down-regulation upon ammonium treatment (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 analyzed plants (Hachiya and Sakakibara, 2016). This pattern is as well detected in the present dataset where both nitrate reductase coding genes (Pp3c10_9670V3.1, Pp3c10_9540V3.1, Pp3c14_9410V3.1) and nitrite reductase coding genes (Pp3c27_6610V3.1, Pp3c16_15880V3.1) are strongly inhibited by the ammonium supplementation (Figure S4A).

On the other hand, we also observe loss of transcript abundance for genes involved in ammonium assimilation: both ammonium transport genes, *AMT2*s (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 ammonium treatment (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 an overall regulatory repression of the 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 a Na P-type ATPase protein demonstrated to be necessary for active Na^+ cell export in *P. patens* (Lunde *et al.*, 2007). The repression of nitrate import under ammonium supplementation affects K^+ import (Coskun *et al.*, 2015), hence the cytoplasmic Na^+/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 involved notably in amino acid transport (Ladwig *et al.*, 2012). Thus, Pp3c20_1770V3.1 abundance increase hints at amino acid relocation upon ammonium treatment.

In parallel to nitrate and ammonium related processes, a cluster of genes associated with cell wall modification appears down-regulated under ammonium supplementation. This repression may reflect the morphological change observed upon ammonium addition: in 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 cell mostly comprise of 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 methylesterases (Pp3c3_30560V3.1, Pp3c3_35240V3.1, Pp3c4_22420V3.1) and extensins (Pp3c16_3130V3.1, Pp3c27_3570V3.1) (Figure S3B), all involved in cell elongation and modification of cell shape (Lamport *et al.*, 2011, Cosgrove, 2016).

The addition of ammonium promotes transcript over-accumulation as well. 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, Pp3c3_12530V3.1) or carbonate dehydratase (Pp3c26_6810V3.1, 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 presence of ammonium, in which smaller cells are filled with numerous chloroplasts, while in nitrate-only condition 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 to compare identical experiments performed within a single laboratory and between two different laboratories. Experiments VII and XXI and XXII (Table S1 and S2) are true replicas. Unfortunately, experiment XXI is only formed of two replica as mentioned previously. Hence the comparison of results needs to 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 seven 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. Experiments VII and XXI samples were isolated using the same standardized protocol (see Methods), experiment XXII the Trizol step was omitted. RNA samples passed a 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 total number of sequenced reads (see Figure S5 for the read length profile and number for each libraries), indicating that detected differences should be mostly attributed to laboratory (culture) variation.

The within-laboratory comparison, experiment XXII compared to XXI, displays a limited number of DEGs, 42 with a fold change > 0 - 2 equally distributed between up (23) and down (19) DEGs (Data S1). This low number could be attributed at least partially to the absence of biological triplicate for experiment XXI or to the Trizol omission in the experiment XXII. Nevertheless, the amplitude of the variation remains in a single order of magnitude, contrary to most of the other comparison in which the calculated fold change can span up to five orders of magnitude. The GO term enrichment analysis (Data S4, Figure S6) performed on these DEGs point to a potential trophic variation source: experiment XXII displays an increase of GO terms associated with response to external nitrogen processes (nitrate transport, 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 down regulated DEGs. Together, these terms point to a potential nitrogen source depletion leading to metabolic and metal homeostasis redirection. Since 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 good reproducibility, given the sensitivity of RNA-seq to detect minor changes in transcript abundance.

In contrast, 1,262 DEGs were detected in total (727 up, 535 down) with a fold change > 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 difference: a contamination with other *P. patens* tissue, and a variation in growing condition generating a stress response. To assess these two possibilities, we compared the DEG list between the two experiments and the DEGs obtained by comparing them to protonemata compared to gametophore (exp. VII compared to XX) to test the tissue hypothesis. We also compared them to three different stress conditions: the effect of ABA on protonemata (exp. XIX compared to XVIII), the effect of high light on protonemata + gametophores (exp. II compared to I) and the effect of elevated temperature (heat stress) to protonemata + gametophores (exp. XXV compared to XXVI). Figure 5 illustrates that the observed differences can indeed come from different sources. 1,101 out of 1,262 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 in the protonema to gametophore comparison. A clear example of gametophore presence 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 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 develops gametophore buds after seven days and thus the culture needs to be blended regularly to reset the protonema to day one 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 lab 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-lab comparison thus demonstrates the sensitivity of RNA-seq and hence the fact that minor fluctuations in growth conditions can result in well 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 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 pointed out by the laboratory replica case. As more datasets are published, the body of data will permit 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 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 plants' response 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). It will be a valuable future goal 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* 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 atop 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 µmol m^{-2} s^{-1} white light (long-day conditions). All harvests were performed in the middle of the light photoperiod (+8 hours 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 condition.

Sample description

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

Light treatments

Light quality:

Prior to treatment, *P. patens* protonemata were cultivated for one week on solidified supplemented BCD medium (BCD supplemented with 5mM 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 following light conditions:

Dark treated samples were grown in darkness for 1-2 weeks (exp. XXIX).

Far red light samples were grown under continuous 2 µmol m^{-2} s^{-1} far red light at 720-740 nm for 1-2 weeks and harvested (exp. XXXIII).

Red light samples were grown under continuous 10 µmol m^{-2} s^{-1} red light at 660-680 nm for 1-2 weeks and harvested (exp. XXX).

Blue light samples were grown under continuous 5 µmol m^{-2} s^{-1} blue light at 460-480 nm for 1-2 weeks and harvested (exp. XXXI).

UV-B light samples were grown under continuous 4 µmol m^{-2} s^{-1} white light supplemented with 1 µmol m^{-2} s^{-1} UV-B light at 300-320 nm for 1-2 week and harvested (exp. 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 µmol m^{-2} s^{-1} white light for 2 hours and harvested (exp. II).

Low light samples were transferred to 10 µmol m^{-2} s^{-1} white light for 2 hours and harvested (exp. III).

Control light samples were maintained at 70 µmol m^{-2} s^{-1} white light for 2 hours and harvested (exp. I).

Tissues:

Germinating spores

P. 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 harvested (exp. IV).

Protonemata grown on solid medium

Protonema was cultivated on solidified BCD medium, in Petri dishes sealed with 3M Micropore tape under standard conditions. Protonema was cultivated for 7 days and harvested (exp. XVIII).

Protonemata grown in liquid medium

Protonema was cultivated in liquid Knop medium in flasks with continuous shaking under standard conditions. Protonema was cultivated for 7 days and harvested by filtering with a 100µm sieve (exp. VII, XXI, 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 part of the plant was harvested (exp. XVII).

Gametophores were cultivated on solidified Knop medium without cellophane, under standard conditions. Gametophores were grown for 5 weeks and the aerial part of the plant was harvested (exp. 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 into RNA*later* solution (Qiagen) before liquid nitrogen flash freeze as a single sample (exp. XIV).

Sporophytes

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

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

Brown sporophytes with a round capsule shape and brown colour were harvested after 7 weeks of short day growth (mature stage stage B (Hiss *et al.*, 2017)) and stored into RNA*later* solution (Qiagen) before liquid nitrogen flash freeze as a single sample (exp. XVI)

Hormones

Auxin

Gametophores were cultivated on a sieve above liquid Knop medium in Magenta boxes (Sigma-Aldrich, St. Louis, US) sealed with parafilm under standard conditions for 10 months. At that time point, auxin samples were treated with 10 µM naphthaleneacetic acid (NAA) and cultivated for 10 days before harvesting. Samples with and without NAA were generated (exp. 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 discs containing tissue were transferred to BCD plates containing either 1 µM racemic GR24 (synthetic strigolactone) or acetone without GR24 as control, incubated for 24 hours and harvested (exp. V and XXXVIII).

Gibberellin

Protonemata was cultivated on solidified BCD medium, supplemented with 20µM GA_9-methylester under standard conditions. GA_9-methylester was synthesized and donated by Peter Hedden's group at Rothamstead Research, Harpenden, UK. Protonema was cultivated for 7 days and harvested (exp. XI).

Abscisic acid

Protonemata were cultivated on solidified BCD medium, in Petri dishes sealed with 3M Micropore tape under standard conditions. Protonema was cultivated for 6 d. Cellophane disk containing tissue were transferred to BCD plates containing 50 µM abscisic acid (ABA) and incubated for 24 h. before harvesting (exp. XIX).

OPDA

Protonemata were cultivated on solidified BCD medium, supplemented with 5 mM ammonium tartrate with or without 50 µM 12-oxo-phytodienoic acid (OPDA) in Petri dishes sealed with 3M Micropore tape under standard conditions. Protonemata were cultivated for 14 days. Cellophane discs containing tissue were transferred to ammonium tartrate-containing BCD plates with or without 50 µM 12-oxo-phytodienoic acid (OPDA) and incubated for 6 h. before harvesting (exp. 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 harvested (exp. 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. Protonema was cultivated for 7 days and harvested 2 hours after the lights were turned on (exp. XXIV without ammonium, exp. XXIII, with ammonium).

De- and Rehydration

Gametophores were cultivated on cellophane discs on solidified BCD medium under standard conditions for 5 weeks prior to the dehydration treatment. The gametophores on the cellophane discs were placed in empty Petri dishes that were sealed in chambers containing an atmosphere of 91% RH generated by a saturated solution of $MgSO_4$ 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 150h (Koster *et al.*, 2010) and gametophores were sampled at 180h (exp. XII). The water potential of the gametophore tissue at equilibrium was -13 MPa.

Rehydration was achieved by floating the cellophane discs containing the dehydrated gametophores on sterile water in a Petri dish for 5 minutes to ensure full rehydration. Once fully rehydrated the discs were placed on solid BCD media and incubated under standard conditions in the light for 2h before harvest (exp. XIII).

Heat

Protonemata was 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 for 5 days with heat shock cycles of 6 hours repeating 5 hours at 22°C and 1 hour at 37° before harvesting (exp. XXVI treatment and XXVII control).

RNA extraction, RNA processing and sequencing.

Frozen samples were pulverized with mortar and pestle and total RNA was extracted in two steps: 1) total RNA was extracted using Trizol reagent (Invitrogen), using the manufacturers' instruction (maximum of 100 mg of tissue per ml of Trizol reagent) 2) 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 manufacturer's instruction (Agilent technologies, Santa Clara, CA, USA). Plate-based RNA sample preparation was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Illumina's TruSeq Stranded mRNA HT sample prep kit utilizing 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, and with the following conditions: total RNA starting material was 1 μg per sample and 8 cycles of PCR was used for library amplification. The prepared libraries were then quantified by qPCR using the Kapa SYBR Fast Illumina Library Quantification Kit (Kapa Biosystems) and run on a Roche LightCycler 480 real-time PCR instrument. 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 flowcell for sequencing. Sequencing of the flowcell was performed on the Illumina HiSeq2500 sequencer using HiSeq TruSeq SBS sequencing kits, v4, following a 2x150 indexed run recipe.

RNA-seq processing

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

Quality trimming and adapter removal

Each library was initially checked with FastQC version 0.11.2 to evaluate read quality (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Subsequently, lower quality bases and sequencing adapters were removed using Trimmomatic version 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

Due to the nature of RNA-seq data, poly-A tails are expected. With PRINSEQ version 0.20.4 (Schmieder and Edwards, 2011) poly-A tails with a minimum length of 12 were identified and removed.

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 are overlapping each other. Such overlapping reads were merged with the help of COPEread version 1.2.5 (Liu *et al.*, 2012).

Mapping

The read mapping was performed using GSNAP version 2015-12-31.v5 (Wu and Nacu, 2010) using 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, 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 kept.

File converting

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

Read count

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

Differential expression analysis

DEG calling approaches can generate different results (Zhang *et al.*, 2014, Schurch *et al.*, 2016). Hence, in order improve confidence in the DEGs used here, several algorithms were tested. DEG analysis was performed in R version 3.2.0 using three R packages, egdeR v3.14.0 (Robinson *et al.*, 2010), DESeq2 v1.12.3 (Love *et al.*, 2014) and NOISeq v2.12.0 (Tarazona *et al.*, 2011). P-value cutoffs for edgeR and DESeq2 were 0.001, 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 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 (see Figure S7 for the Venn diagram representation of the four ways DEG call comparison), which are characterized by a particularly low average expression level (3.5 FPKM, as compared to 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, the 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. 620 DEGs were called by the microarray data (Hiss et al. 2014). RNA-seq due to its higher sensitivity was calling 3,309 DEGs. We found that 69 % of the microarray DEGs are overlapping 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 be sharing 95.5 % of their associated terms with the GO terms associated with the RNA-seq DEGs (Data S5). 758 of 762 GO terms were not showing significant differences between the two DEG sets in terms of their number of associated genes (Fisher's exact test, p.adjust < 0.05).

GO term enrichment analysis

GO enrichment analyses were conducted as described previously (Widiez *et al.*, 2014). Visualization of the GO terms was implemented using word clouds *via* the www.wordle.net application. Word size is proportional to the –log10 (q-value) and over-represented GO terms were colored dark green if –log10 (q-value) ≤ 4 and light green if –log10 (q-value) > 4.

Data visualization

Principal component analysis (PCA) was performed in R version 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 version 1.1.1 (https://CRAN.R-project.org/package=plot3D). Hierarchical clustering was calculated and visualized using R (version 3.4.3) and the package "ComplexHeatmap" version 1.17.1 (Gu *et al.*, 2016) with the euclidean distance method. 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 of the University of Gent (http://bioinformatics.psb.ugent.be/webtools/Venn/) with the symmetric and colored options.

Acknowledgments

We would like to thank Christine Glockner (University of Freiburg), Marco Göttig (University of Marburg) and Mitsuhiro Ishibashi (Hokkaido University) for their technical assistance. The work conducted by the U.S. Department of Energy Joint Genome Institute 902 is supported by the Office of Science of the U.S. 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.'s 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-methylester. 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).

The authors declare no conflicts of interest.

Short Supporting Information Legends:

Figure S1: Hierarchical clustering of all 99 RNA-seq samples, RPKM normalized.

Figure S2: GO term enrichment of developmental stage DEGs.

Figure S3: Impact of ammonium supplementation on protonemata transcriptome.

Figure S4: Specific transcript accumulation variation induced by ammonium supplementation.

Figure S5: Libraries quality evaluation in the laboratory comparison dataset.

Figure S6: Comparison between replica experiments performed in a single laboratory

Figure S7: Comparison of the DEGs called by the NOISeq, DESeq2 and edgeR packages with RNA dataset and by microarray approach.

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

Table S2: Harvesting time point for each experiment in this study 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.

Data S1: RPKM count for each library described in this study and DEG comparisons performed between 50 experiments.

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

Data S3: GO terms enrichment lists associated with stage-specific comparisons.

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., Pacak, A.M., Sierocka, I., Nuc, P.W., Singh, K., Plewka, P., Sulkowska, A., Jarmolowski, A., Karlowski, W.M. and Szweykowska-Kulinska, Z. (2015) The liverwort Pellia endiviifolia shares microtranscriptomic traits that are common to green algae and land plants. *New Phytologist*, **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, Ljtrans DB. *Plant and Cell Physiology*, **56**, e5-e5.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., Ashton, N.W., Axtell, M.J., Barker, E., Barker, M.S., Bennetzen, J.L., Bonawitz, N.D., Chapple, C., Cheng, C., Correa, L.G.G., Dacre, M., DeBarry, J., Dreyer, I., Elias, M., Engstrom, E.M., Estelle, M., Feng, L., Finet, C., Floyd, S.K., Frommer, W.B., Fujita, T., Gramzow, L., Gutensohn, M., Harholt, J., Hattori, M., Heyl, A., Hirai, T., Hiwatashi, Y., Ishikawa, M., Iwata, M., Karol, K.G., Koehler, B., Kolukisaoglu, U., Kubo, M., Kurata, T., Lalonde, S., Li, K., Li, Y., Litt, A., Lyons, E., Manning, G., Maruyama, T., Michael, T.P., Mikami, K., Miyazaki, S., Morinaga, S.-i., Murata, T., Mueller-

Roeber, B., Nelson, D.R., Obara, M., Oguri, Y., Olmstead, R.G., Onodera, N., Petersen, B.L., Pils, B., Prigge, M., Rensing, S.A., Riaño-Pachón, D.M., Roberts, A.W., Sato, Y., Scheller, H.V., Schulz, B., Schulz, C., Shakirov, E.V., Shibagaki, N., Shinohara, N., Shippen, D.E., Sørensen, I., Sotooka, R., Sugimoto, N., Sugita, M., Sumikawa, N., Tanurdzic, M., Theißen, G., Ulvskov, P., Wakazuki, S., Weng, J.-K., Willats, W.W.G.T., Wipf, D., Wolf, P.G., Yang, L., Zimmer, A.D., Zhu, Q., Mitros, T., Hellsten, U., Loqué, D., Otillar, R., Salamov, A., Schmutz, J., Shapiro, H., Lindquist, E., Lucas, S., Rokhsar, D. and Grigoriev, I.V. (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., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., Adam, C., Aki, S.S., Althoff, F., Araki, T., Arteaga-Vazquez, M.A., Balasubrmanian, S., Barry, K., Bauer, D., Boehm, C.R., Briginshaw, L., Caballero-Perez, J., Catarino, B., Chen, F., Chiyoda, S., Chovatia, M., Davies, K.M., Delmans, M., Demura, T., Dierschke, T., Dolan, L., Dorantes-Acosta, A.E., Eklund, D.M., Florent, S.N., Flores-Sandoval, E., Fujiyama, A., Fukuzawa, H., Galik, B., Grimanelli, D., Grimwood, J., Grossniklaus, U., Hamada, T., Haseloff, J., Hetherington, A.J., Higo, A., Hirakawa, Y., Hundley, H.N., Ikeda, Y., Inoue, K., Inoue, S.I., Ishida, S., Jia, Q., Kakita, M., Kanazawa, T., Kawai, Y., Kawashima, T., Kennedy, M., Kinose, K., Kinoshita, T., Kohara, Y., Koide, E., Komatsu, K., Kopischke, S., Kubo, M., Kyozuka, J., Lagercrantz, U., Lin, S.S., Lindquist, E., Lipzen, A.M., Lu, C.W., De Luna, E., Martienssen, R.A., Minamino, N., Mizutani, M., Mizutani, M., Mochizuki, N., Monte, I., Mosher, R., Nagasaki, H., Nakagami, H., Naramoto, S., Nishitani, K., Ohtani, M., Okamoto, T., Okumura, M., Phillips, J., Pollak, B., Reinders, A., Rovekamp, M., Sano, R., Sawa, S., Schmid, M.W., Shirakawa, M., Solano, R., Spunde, A., Suetsugu, N., Sugano, S., Sugiyama, A., Sun, R., Suzuki, Y., Takenaka, M., Takezawa, D., Tomogane, H., Tsuzuki, M., Ueda, T., Umeda, M., Ward, J.M., Watanabe, Y., Yazaki, K., Yokoyama, R., Yoshitake, Y., Yotsui, I., Zachgo, S. and Schmutz, J. (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. *American Journal of Botany*, **100**, 161-174.

Champagne, C.E.M. and Ashton, N.W. (2001) Ancestry of KNOX genes revealed by bryophyte (Physcomitrella patens) homologs. *New Phytologist*, **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 Physiology*, **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 phytobilin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **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 Phytologist*, **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. *The Plant Cell*, **27**, 2148-2162.

Cosgrove, D.J. (2016) Catalysts of plant cell wall loosening. *F1000Res*, **5**.

Coskun, D., Britto, D.T. and Kronzucker, H.J. (2015) The nitrogen-potassium intersection: Membranes, metabolism, and mechanism. *Plant Cell Environment*.

Coskun, D., Britto, D.T. and Kronzucker, H.J. (2016) Nutrient constraints on terrestrial carbon fixation: The role of nitrogen. *Journal of Plant Physiology*, **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.

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.

Decker, E.L., Alder, A., Hunn, S., Ferguson, J., Lehtonen, M.T., Scheler, B., Kerres, K.L., Wiedemann, G., Safavi-Rizi, V., Nordzieke, S., Balakrishna, A., Baz, L., Avalos, J., Valkonen, J.P.T., Reski, R. and Al-Babili, S. (2017) Strigolactone biosynthesis is evolutionarily conserved, regulated by phosphate starvation and

contributes to resistance against phytopathogenic fungi in a moss, Physcomitrella patens. *New Phytologist*, **216**, 455-468.

Demko, V., Perroud, P.F., Johansen, W., Delwiche, C.F., Cooper, E.D., Remme, P., Ako, A.E., Kugler, K.G., Mayer, K.F., Quatrano, R. and Olsen, O.A. (2014) Genetic analysis of DEFECTIVE KERNEL1 loop function in three-dimensional body patterning in Physcomitrella patens. *Plant Physiology*, **166**, 903-919.

Denancé, N., Szurek, B. and Noël, L.D. (2014) Emerging Functions of Nodulin-Like Proteins in Non-Nodulating Plant Species. *Plant and Cell Physiology*, **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 Phytologist*, **211**, 300-318.

Engel, P.P. (1968) The Induction of Biochemical and Morphological Mutants in the Moss Physcomitrella patens. *American Journal of Botany*, **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 Physiology*, **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. *Nature Biotechnology*, **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 Research Notes*, **7**, 490-490.

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 Genomics*, **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. *The 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. (2016) Interactions between nitrate and ammonium in their uptake, allocation, assimilation, and signaling in plants. *Journal of Experimental Botany*.

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., Arif, M.A., Decker, E.L., Erxleben, A., Frank, W., Hanke, S.T., Lang, D., Martin, A., Neu, C., Reski, R., Richardt, S., Schallenberg-Rudinger, M., Szovenyi, P., Tiko, T., Wiedemann, G., Wolf, L., Zimmermann, P. and Rensing, S.A. (2014) Large-scale gene expression profiling data for the model moss Physcomitrella patens aid understanding of developmental progression, culture and stress conditions. *The Plant Journal*, **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. *The Plant Journal*, **90**, 606-620.

Hoffmann, B., Proust, H., Belcram, K., Labrune, 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 Biology*, **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 Regulation*, **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 Physiology*, **158**, 1643-1655.

Lamport, D.T., Kieliszewski, M.J., Chen, Y. and Cannon, M.C. (2011) Role of the extensin superfamily in primary cell wall architecture. *Plant Physiology*, **156**, 11-19.

Lang, D., Ullrich, K.K., Murat, F., Fuchs, J., Jenkins, J., Haas, F.B., Piednoel, M., Gundlach, H., Van Bel, M., Meyberg, R., Vives, C., Morata, J., Symeonidi, A., Hiss, M., Muchero, W., Kamisugi, Y., Saleh, O., Blanc, G., Decker, E.L., van Gessel, N., Grimwood, J., Hayes, R.D., Graham, S.W., Gunter, L.E., McDaniel, S., Hoernstein, S.N.W., Larsson, A., Li, F.W., Perroud, P.F., Phillips, J., Ranjan,

P., Rokshar, D.S., Rothfels, C.J., Schneider, L., Shu, S., Stevenson, D.W., Thummler, F., Tillich, M., Villarreal, A.J., Widiez, T., Wong, G.K., Wymore, A., Zhang, Y., Zimmer, A.D., Quatrano, R.S., Mayer, K.F.X., Goodstein, D., Casacuberta, J.M., Vandepoele, K., Reski, R., Cuming, A.C., Tuskan, J., Maumus, F., Salse, J., Schmutz, J. and Rensing, S.A. (2018) The P. patens chromosome-scale assembly reveals moss genome structure and evolution. *The Plant Journal*, **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., 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., Li, Z., Xie, Y., Chen, Y., Shi, Y., Zhang, H., Li, Y., Lam, T.W. and Luo, R. (2012) COPE: an accurate k-mer-based pair-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 Biology*, **15**, 550-550.

Lunde, C., Drew, D.P., Jacobs, A.K. and Tester, M. (2007) Exclusion of Na+ via sodium ATPase (PpENA1) ensures normal growth of Physcomitrella patens under moderate salt stress. *Plant Physiology*, **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., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., Burleigh, J.G., Gitzendanner, M.A., Wafula, E., Der, J.P., dePamphilis, C.W., Roure, B., Philippe, H., Ruhfel, B.R., Miles, N.W., Graham, S.W., Mathews, S., Surek, B., Melkonian, M., Soltis, D.E., Soltis, P.S., Rothfels, C., Pokorny, L., Shaw, J.A., DeGironimo, L., Stevenson,

D.W., Villarreal, J.C., Chen, T., Kutchan, T.M., Rolf, M., Baucom, R.S., Deyholos, M.K., Samudrala, R., Tian, Z., Wu, X., Sun, X., Zhang, Y., Wang, J., Leebens-Mack, J. and Wong, G.K.-S. (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. *Journal of Experimental Botany*, **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. *Journal of Experimental Botany*.

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. *Molecular 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. *Frontiers in Plant Science*, **8**.

Renault, H., Alber, A., Horst, N.A., Lopes, A.B., Fich, E.A., Kriegshauser, L., Wiedemann, G., Ullmann, P., Herrgott, L., Erhardt, M., Pineau, E., Ehlting, J., Schmitt, M., Rose, J.K.C., Reski, R. and Werck-Reichhart, D. (2017) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. *Nature Communications*, **8**, 14713.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin, I.T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W.B., Barker, E., Bennetzen, J.L., Blankenship, R., Cho, S.H., Dutcher, S.K., Estelle, M., Fawcett, J.A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K.A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D.R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P.J., Sanderfoot, A., Schween, G., Shiu, S.H., Stueber, K., Theodoulou, F.L., Tu, H., Van de Peer, Y., Verrier, P.J., Waters, E., Wood, A., Yang, L., Cove, D.,

Cuming, A.C., Hasebe, M., Lucas, S., Mishler, B.D., Reski, R., Grigoriev, I.V., Quatrano, R.S. and Boore, J.L. (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. *Evolution and Development*, **10**, 555-566.

Schaefer, D.G. and Zrÿd, J.P. (1997) Efficient gene targeting in the moss Physcomitrella patens. *The Plant Journal*, **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., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., Blaxter, M. and Barton, G.J. (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 Reporter*, **26**, 2039-2054.

Song, L., Shankar, D.S. and Florea, L. (2016) Rascaf: Improving Genome Assembly with RNA Sequencing Data. *The Plant Genome*, **9**.

Stevenson, S.R., Kamisugi, Y., Trinh, C.H., Schmutz, J., Jenkins, J.W., Grimwood, J., Muchero, W., Tuskan, G.A., Rensing, S.A., Lang, D., Reski, R., Melkonian, M., Rothfels, C.J., Li, F.-W., Larsson, A., Wong, G.K.S., Edwards, T.A. and Cuming, A.C. (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. *The Plant Cell*, **28**, 1310-1327.

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.

Szövényi, P., Rensing, S.A., Lang, D., Wray, G.A. and Shaw, A.J. (2011) Generation-biased gene expression in a bryophyte model system. *Molecular Biology and Evolution*, **28**, 803-812.

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. and Conesa, A. (2011) Differential expression in RNA-seq: a matter of depth. *Genome research*, **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 and Cell Physiology*, **48**, 484-497.

Waldie, T., McCulloch, H. and Leyser, O. (2014) Strigolactones and the control of plant development: lessons from shoot branching. *The Plant Journal*, **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. *The Plant Journal*, **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 Biology*, **11**, 177-177.

Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., Zhuang, R., Lu, Z., He, Z., Fang, X., Chen, L., Tian, W., Tao, Y., Kristiansen, K., Zhang, X., Li, S., Yang, H., Wang, J. and Wang, J. (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Research*, **20**, 646-654.

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.

Zhang, Z.H., Jhaveri, D.J., Marshall, V.M., Bauer, D.C., Edson, J., Narayanan, R.K., Robinson, G.J., Lundberg, A.E., Bartlett, P.F., Wray, N.R. and Zhao, Q.Y. (2014) A comparative study of techniques for differential expression analysis on RNA-Seq data. *PLoS One*, **9**, e103207.

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 Genomics*, **18**, 1042.

Figure Legends

Figure 1: RNA-seq data analysis

Diagram illustrating the sequential RNA-seq data treatment from raw read to differentially express genes (DEG). See Methods for details.

Figure 2: Principal components analysis (PCA) of RPKM values for the 99 libraries of this study.

Each dot represents one library. Dots for each experiment have the same color. Red ellipse highlights most gametophytic experiments. Green ellipse highlights the sporophytes experiments. Blue ellipse indicates strong stress experiments. Dotted ellipse *a* points to an experiment with tightly grouped triplicate. Dotted ellipse *b* points to an experiment with somehow loosely grouped triplicate.

Figure 3: Stage-specific transcript markers.

Two gene models each for protonema, gametophore and sporophyte are selected for their specific presence for the stage and their absence in all experiments not containing exclusively the specific tissue. RPKM values from experiment VII (protonema), XX (gametophore) and XV (sporophyte).

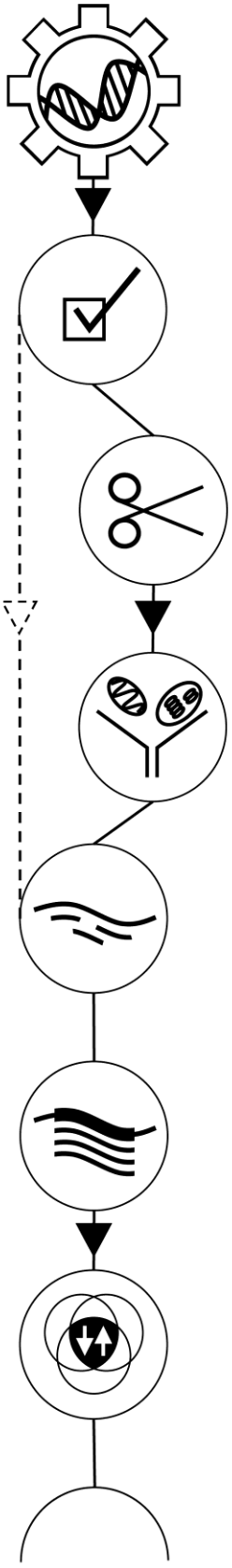
Figure 4: Ammonium supplementation impact on transcriptome.

GO term analysis representation of the 15 top over-represented GO terms in the up- (**A**) and under accumulated down-regulated (**B**) DEGs in ammonium-supplemented liquid protonemal culture (exp. XXIII) compared to ammonium-free liquid protonemal culture (exp. XXIV). Word size is proportional to the –log10 (q-value) and over-represented GO terms were colored dark green if –log10 (q-value) ≤ 4 and light green if –log10 (q-value) > 4. For the GO term IDs and their respective over-representation values see Data S3. (**C**) RPKM values for nitrate and nitrite transporter gene models in absence of ammonium (red bar) and presence of ammonium (blue bar). In grey shade are the gene models not identified by Tsujimoto et al. (2007). Identifiers of the different nitrate transporter genes described by

Tsujimoto et al. (2007) *P. patens* V3.3 genome read as follow: 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, PpNRT2;9/Pp3c16_10420V3.1.

Figure 5: Contrasting DEGs across experiments.

Venn diagram analysis of the differentially expressed genes (DEGs) of replica experiments between different laboratories (Lab A, exp. VII compared to Lab B, exp. XXII) and with DEGs between protonema and gametophore (protonema, exp. VII compared to gametophore, exp. XX), between ABA treatment (Control, exp. XVIII compared to ABA, exp. XIX), heat treatment (Control, exp. XXV compared to Heat, exp. XXVI) and high light treatment (control, exp. I compared to high light, exp. II). The number of transcripts meeting the cutoff (> 2-fold up/down- regulated, adjusted p-value < .05) are contained within each section of the labeled circle.



Raw sequenced data
Illumina HiSeq

Quality check
FastQC

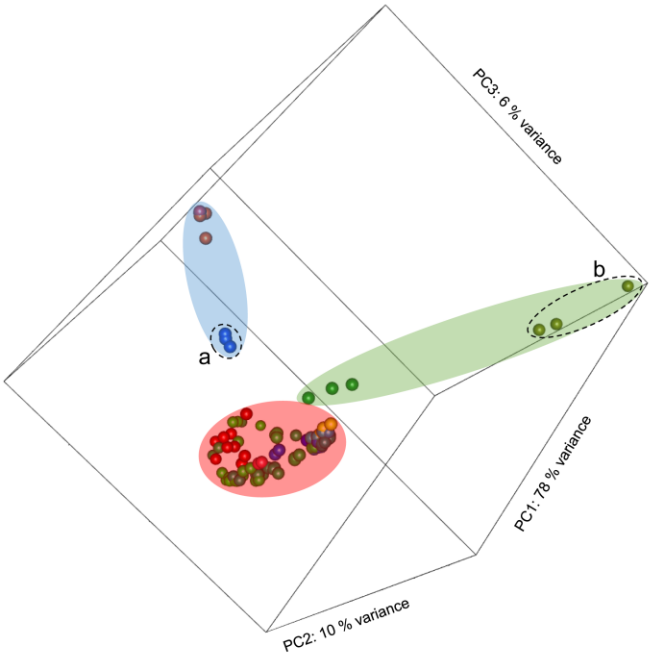
Quality, adapter and poly-A trimming
Trimmomatic and PRINSEQ

Organellar and rRNA filtering
GSNAP

Mapping to the reference genome
GSNAP

Read count
HTSeq

Differentially Expressed Genes (DEG)
edgeR ∩ DESeq2 ∩ NOISeq



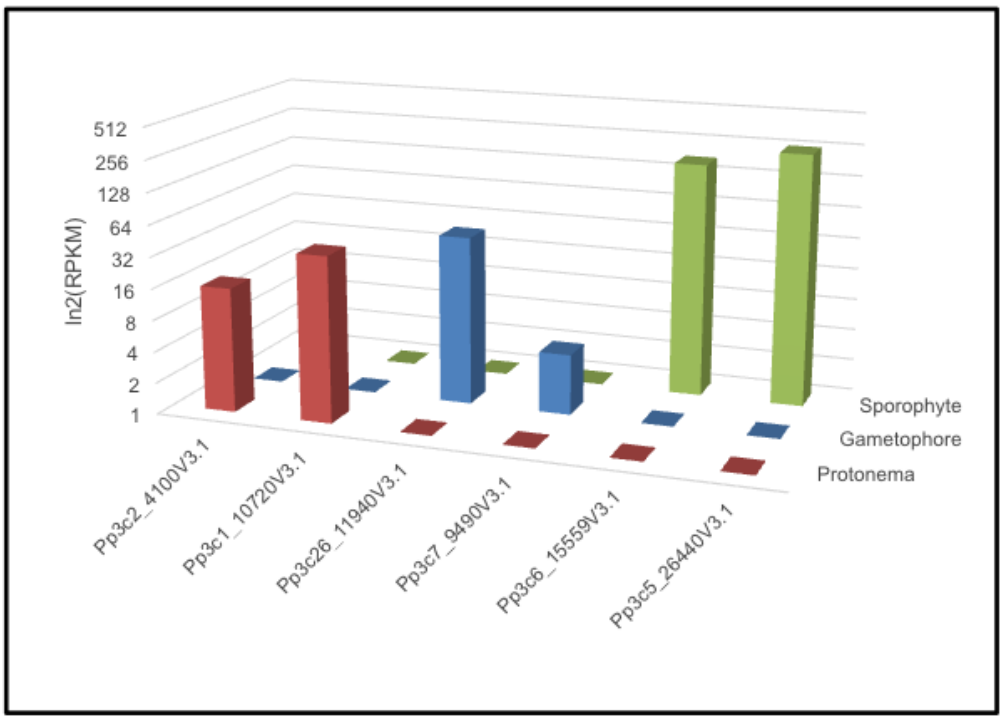


Figure 3

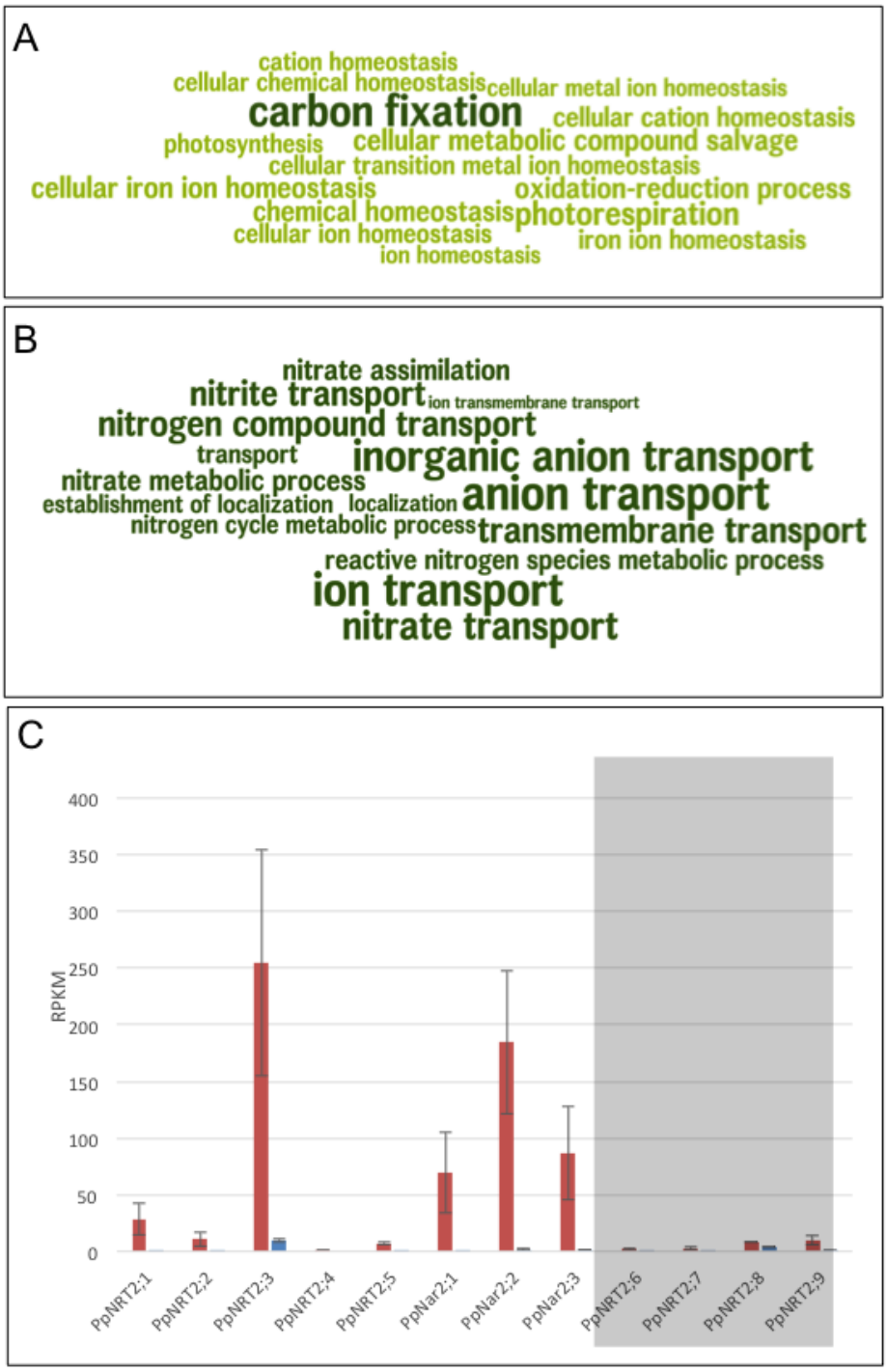


Figure 4

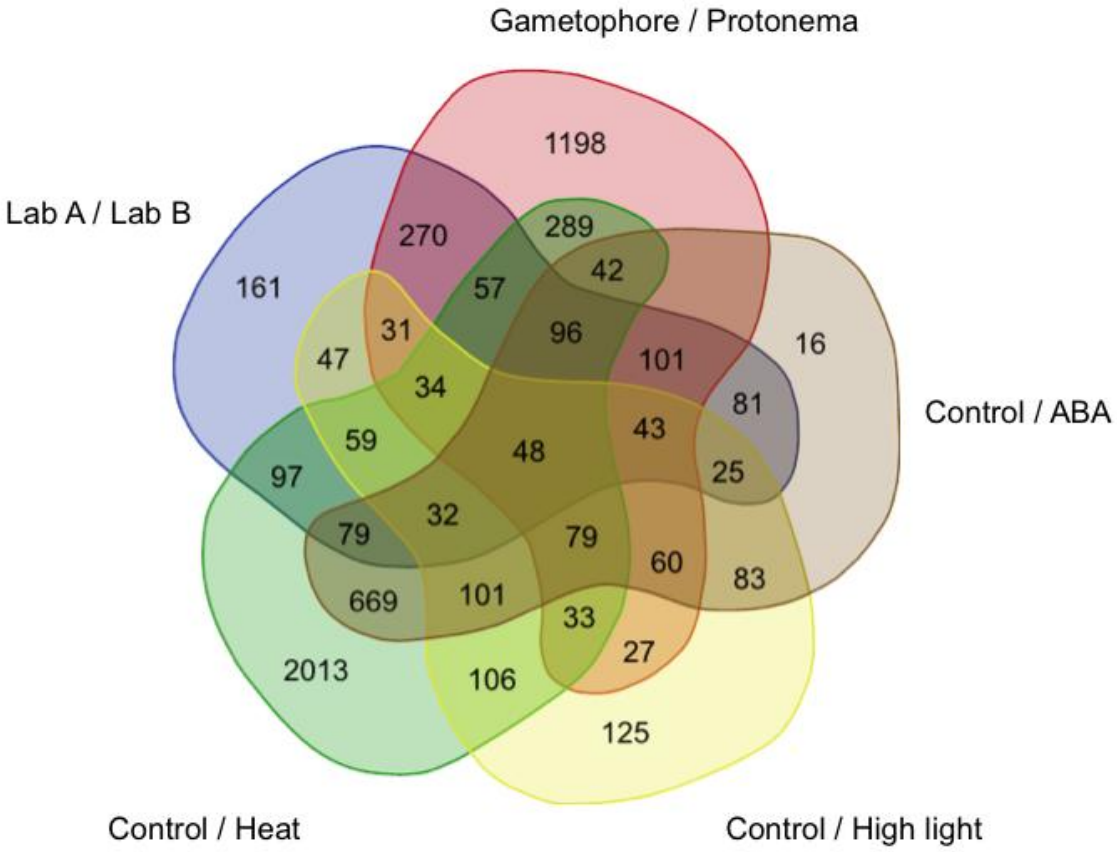


Figure 5