the plant journal

SEB Society for Experimental Biology

The Plant Journal (2014) **79,** 530–539 doi: 10.1111/tpj.12572

RESOURCE

Large-scale gene expression profiling data for the model moss *Physcomitrella patens* aid understanding of developmental progression, culture and stress conditions

Manuel Hiss[1,2,3], Oliver Laule[4,†], Rasa M. Meskauskiene[4], Muhammad A. Arif[5,6], Eva L. Decker[5], Anika Erxleben[5,‡], Wolfgang Frank[6], Sebastian T. Hanke[1,2], Daniel Lang[5], Anja Martin[5,§], Christina Neu[2], Ralf Reski[3,5,7,8], Sandra Richardt[2,¶], Mareike Schallenberg-Rüdinger[1], Peter Szövényi[2,], Theodhor Tiko[2], Gertrud Wiedemann[5], Luise Wolf[2,**,††], Philip Zimmermann[4,*] and Stefan A. Rensing[1,2,3,7,*]**

[1]*Plant Cell Biology, Faculty of Biology, University of Marburg, Karl-von-Frisch-Strasse 8, 35043 Marburg, Germany,*
[2]*Faculty of Biology, University of Freiburg, Schänzlestrasse 1, 79104 Freiburg, Germany,*
[3]*FRISYS Freiburg Initiative for Systems Biology, University of Freiburg, 79104 Freiburg, Germany,*
[4]*NEBION AG, Hohlstrasse 515, 8048 Zurich, Switzerland,*
[5]*Department of Plant Biotechnology, Faculty of Biology, University of Freiburg, Schänzlestrasse 1, 79104 Freiburg, Germany,*
[6]*Plant Molecular Cell Biology, Department Biology I, Ludwig-Maximilians-University Munich, LMU Biocenter, Grosshadernerstrasse 2-4, 82152 Planegg-Martinsried, Germany,*
[7]*BIOSS Centre for Biological Signalling Studies, University of Freiburg, 79104 Freiburg, Germany and*
[8]*FRIAS – Freiburg Institute for Advanced Studies, University of Freiburg, 79104 Freiburg, Germany*

Received 1 August 2013; revised 22 May 2014; accepted 27 May 2014; published online 6 June 2014.
*For correspondence (e-mails stefan.rensing@biologie.uni-marburg.de or phz@nebion.com).
‡Present address: Department of Pharmaceutical Bioinformatics, Faculty of Chemistry, Pharmacy, and Earth Sciences, University of Freiburg, Hermann-Herder-Strasse 9, 79104 Freiburg, Germany.
§Present address: Thermo Fisher, Opelstrasse 9, 68789 St Leon-Rot, Germany.
¶Present address: Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany.
**Present address: Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
††Present address: Department of Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland.
†deceased.

SUMMARY

The moss *Physcomitrella patens* is an important model organism for studying plant evolution, development, physiology and biotechnology. Here we have generated microarray gene expression data covering the principal developmental stages, culture forms and some environmental/stress conditions. Example analyses of developmental stages and growth conditions as well as abiotic stress treatments demonstrate that (i) growth stage is dominant over culture conditions, (ii) liquid culture is not stressful for the plant, (iii) low pH might aid protoplastation by reduced expression of cell wall structure genes, (iv) largely the same gene pool mediates response to dehydration and rehydration, and (v) AP2/EREBP transcription factors play important roles in stress response reactions. With regard to the AP2 gene family, phylogenetic analysis and comparison with *Arabidopsis thaliana* shows commonalities as well as uniquely expressed family members under drought, light perturbations and protoplastation. Gene expression profiles for *P. patens* are available for the scientific community via the easy-to-use tool at https://www.genevestigator.com. By providing large-scale expression profiles, the usability of this model organism is further enhanced, for example by enabling selection of control genes for quantitative real-time PCR. Now, gene expression levels across a broad range of conditions can be accessed online for *P. patens*.

Keywords: *Physcomitrella patens*, moss, gene expression, microarray, genevestigator, transcriptomics, development, culture, stress.

INTRODUCTION

Gene expression profiles are a valuable community resource. They allow researchers interested in certain sets of genes or conditions (tissues, developmental stages, stress treatments, etc.) to investigate transcription levels *in silico* and to generate hypotheses to subsequently put to the test. Through the availability of expression data in compliance with the minimum information about a microarray experiment (MIAME) standard (Brazma *et al.*, 2001; Zimmermann *et al.*, 2006) in repositories such as ARRAYEXPRESS (Rocca-Serra *et al.*, 2003), incorporation into meta-analysis tools such as GENEVESTIGATOR (Hruz *et al.*, 2008) becomes feasible. The availability of data in such a tool allows for the end user to browse with ease through experiments conducted by different labs or using different technology platforms. Moreover, the use of anatomy, development and treatment ontologies allows users to analyze, for example, developmental progression or to compare expression data across taxonomic boundaries.

The moss *Physcomitrella patens* has been developed over the last decade as a plant model organism for which a large set of experimental tools is available (Reski and Cove, 2004; Frank *et al.*, 2005; Quatrano *et al.*, 2007; Kamisugi *et al.*, 2008; Lang *et al.*, 2008; Prigge and Bezanilla, 2010; Mueller *et al.*, 2014). The sequencing of the genome (Rensing *et al.*, 2008) allowed the development of a gene expression microarray covering all predicted protein-coding genes (Wolf *et al.*, 2010).

We decided to generate expression profiles from a set of principal tissues/developmental stages and environmental/stress treatments that we consider useful for the community. Here we present the generation of large-scale gene expression data for *P. patens* as well as their integration and availability via GENEVESTIGATOR.

RESULTS

The initial set of expression profile data represents a wide range of conditions, including various tissue types, stages of development and perturbations (Table 1). In general, three biological replicates were generated per condition. Exceptions are the leaflet time series where each time point is represented by a single biological replicate (Busch *et al.*, 2013) and the developmental stage 'gametophore formed' with four biological replicates. To verify the level of standardization of experimental conditions we analyzed all samples using hierarchical clustering. The results show a close clustering of biological replicates relative to other samples, confirming a high level of reproducibility (Figure S1). In the following we have conducted some example analyses to demonstrate the usefulness of the data to analyze development, culture conditions and stress.

Table 1 Experimental conditions. List of experimental conditions that are available in GENEVESTIGATOR as microarray data sets. The ARRAYEXPRESS bulk accession numbers for the experiments are given. Each experiment consists of three biological replicates with the exception of the time series of detached leaflets (one replicate for each time point) and the developmental stage 'gametophore formed' (four replicates)

	Tissue/treatment	Bulk accession number
Tissues/cells (Anatomy)	Spores	E-MTAB-916
	Protoplasts	E-MTAB-976
	Protonemata	E-MTAB-976, E-MTAB-917
	Gametophores	E-MTAB-917, E-MTAB-916
Developmental stages (Development)	Germination (protonemal development)	
	Germinated spores	E-MTAB-916
	Primarily chloronemata	E-MTAB-976, E-MTAB-917
	Gametophore growth	
	Gametophore formed	E-MTAB-917
	Gametangia development	
	Mature antheridia/archegonia	E-MTAB-917
	Different genotype (Reute, Villersexel)	E-MTAB-916
Perturbations		
Biotic	Botrytiscinerea inoculation	E-MTAB-919
Light intensity	Darkness	E-MTAB-913
	Strong light	E-MTAB-913
	Sunlight	E-MTAB-913
Light quality	UV-B (303 nm) supplementation, different intensities	E-MEXP-2508
	Shift long day to short day	E-MTAB-917
Water stress	Dehydration	E-MTAB-914
	Rehydration	E-MTAB-914
Other	Different growth media	E-MTAB-976
	Shift from pH 5.8 to pH 4.5	E-MTAB-976
	Timeseries of detached leaflets	E-MTAB-915

Growth stage is dominant over culture form

Principal component analysis across all experiments from the initial set of published data shows a deep cleft between liquid culture/filamentous (protonemal) growth stage and culture on solid medium/gametophore stage. There are 229 genes differentially expressed (Cyber-T, $q < 0.05$) between protonema on liquid versus solid medium, while 247 genes are differentially expressed between protonema and gametophores on solid medium. Partial least squares analysis confirms that the majority of differences in expression are covariant with the two principal growth stages, protonema and gametophores (Figure 1), which are therefore dominant over culture form with regard to alteration of the gene expression profile.

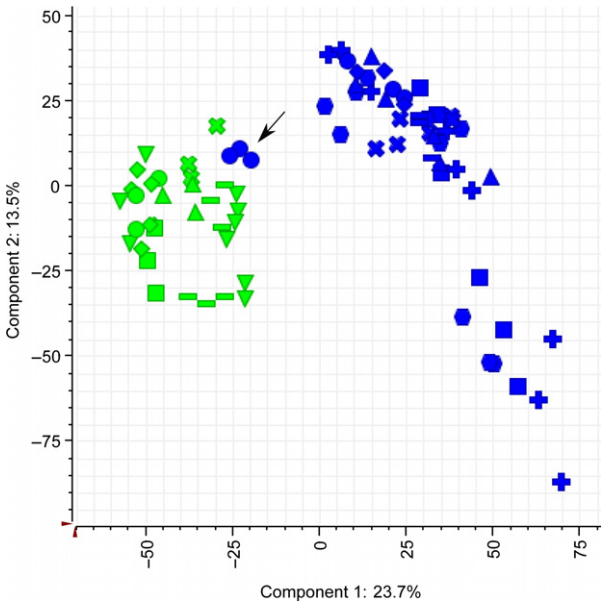

Figure 1. Partial least squares analysis across experiments.
The tissue type is used as the covariant (activity) factor. The plot shows in blue the experiments on a solid medium (mainly gametophores) and in green the liquid culture experiments (protonemata). Arrow shows the three replicate experiments for protonema on solid medium. Growth stage (tissue type) is clearly dominant over culture form, since these replicates cluster with the other protonema experiments, although those were conducted in liquid culture. Axes show the first two components and the covariance explained by these components.

Liquid culture does not represent a stress condition

Although many genes are differentially regulated between protonema in liquid culture and on solid medium, stress-related Gene Ontology categories are not over-represented among these. Therefore, the regular shearing (fragmentation) of protonemal filaments done during the liquid culture regime does not seem to constitute a stress condition, unlike, for example, darkness or dehydration (see below), where stress terms are found to be over-represented among the differentially regulated genes.

An acidic medium might aid protoplastation by downregulation of cell wall structure genes

It is recommended that plant material which will be used for protoplast preparation is grown at pH 4.5 prior to enzymatic digestion of the cell wall (Hohe and Reski, 2002). Plants growing at pH 4.5 apparently have a different cell wall composition from plants growing at pH 5.8 (the cell wall can be more easily digested to produce protoplasts, as measured by a higher rate of released protoplasts). Comparing the gene expression data for tissue growing at pH 5.8 and at pH 4.5 should give insights in the activation of genes responsible for the change in cell wall composition.

Six hundred and seventy-one genes are found to be upregulated by the shift to pH 4.5, and 270 downregulated (Cyber-T, $q < 0.05$). Among the upregulated genes there are only three that are annotated as involved in 'cell wall catabolism', but no expansins (Schipper *et al.* 2002) or other genes associated with cell wall loosening, like xyloglucan endotransglucosylases or pectinesterases (Lagaert *et al.*, 2009), are found. In contrast, the downregulated genes include five expansin genes and three cellulose synthase genes, resulting in an overrepresentation of the Gene Ontology (GO) term 'plant-type cell wall organization' (Table 2). This observation seems to clash with previous studies in cucumber (*Cucumis sativus*), where expansin proteins were shown to increase activity after the external pH was changed to acidic conditions (McQueen-Mason *et al.*, 1992). However, reaction to acidic conditions might vary between different plant species, and the control treatment was growth at pH 5.8, i.e. mildly acidic conditions. In the case of *P. patens* the downregulation of specific cell wall structure genes might influence cell wall composition, leading to better digestibility of the cell wall. Another explanation might be that only newly formed cells start to change their cell wall composition at pH 4.5 (Hohe and Reski, 2002).

Dehydration and rehydration responses are mediated by the same gene pool

Statistical analysis (see Experimental Procedures) of the gene expression data for all expressed genes from dehydrated (50% fresh weight loss), rehydrated and untreated gametophores finds 690 genes upregulated and 1231 genes downregulated after an hour of dehydration out of a total of 26 853 genes on the microarray. Genes important in regulation of transcription, protein modification and

Table 2 Gene Ontology (GO) term enrichment analysis – pH shift. List of significantly overrepresented GO terms (biological process ontology) and corresponding *q*-values from the enrichment analysis for genes downregulated after a pH shift from 5.8 to 4.5

GO-term	Number of genes	*q*-Value
Unsaturated fatty acid biosynthetic process	5	1.7×10^{-4}
Glucose metabolic process	9	5.0×10^{-3}
Plant-type cell wall organization	4	9.1×10^{-3}
Fatty acid biosynthetic process	5	2.3×10^{-2}
Spermidine biosynthetic process	2	2.5×10^{-2}
Carbon fixation	3	2.5×10^{-2}
NAD biosynthetic process	2	3.1×10^{-2}
Pyridine nucleotide biosynthetic process	2	3.8×10^{-2}
Glucose catabolic process	6	4.1×10^{-2}
Lipid biosynthetic process	7	4.1×10^{-2}
Fatty acid metabolic process	5	4.4×10^{-2}
Polyamine metabolic process	2	4.6×10^{-2}

Table 3 Gene Ontology (GO) term enrichment analysis – dehydration and rehydration. List of significantly overrepresented GO terms (biological process ontology) and corresponding *q*-values from the enrichment analysis for genes upregulated after dehydration

GO term	Number of genes	*q*-Value
Regulation of transcription, DNA-dependent	37	4.2×10^{-4}
Regulation of transcription	38	4.2×10^{-4}
Protein modification process	57	3.7×10^{-3}
Protein ubiquitination	18	3.7×10^{-3}
Metal ion transport	12	1.3×10^{-2}
Response to water	2	2.0×10^{-2}
Protein amino acid dephosphorylation	5	3.2×10^{-2}
SRP-dependent cotranslational protein targeting to membrane	2	3.2×10^{-2}
Lipid metabolic process	14	3.2×10^{-2}
Cation transport	16	3.2×10^{-2}
Mitochondrial transport	3	3.2×10^{-2}
Exocytosis	3	3.2×10^{-2}
Cell adhesion	3	3.2×10^{-2}
Oligosaccharide metabolic process	3	3.2×10^{-2}
Guanosine tetraphosphate metabolic process	2	3.2×10^{-2}
Lipid glycosylation	2	3.2×10^{-2}
Ion transport	17	3.8×10^{-2}
Steroid metabolic process	2	4.0×10^{-2}

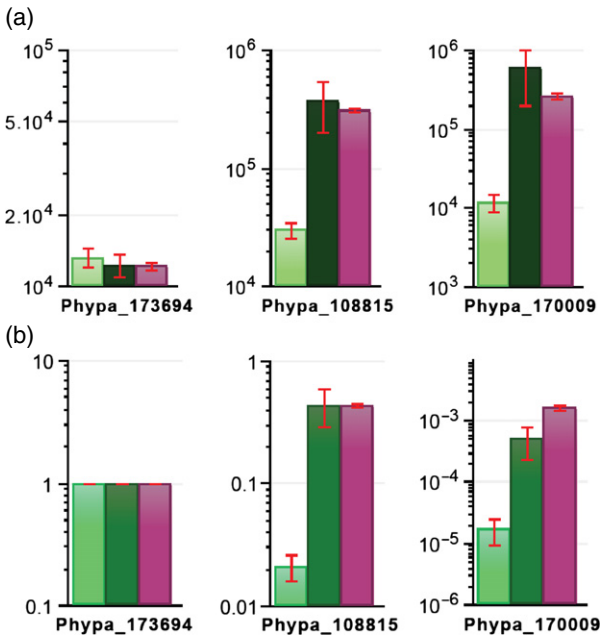

Figure 2. Bar charts of normalized expression strength.
Normalized expression level of the *lea* genes Phypa_108815 and Phypa_170009 and of the reference gene Phypa_173694 (encoding a thioredoxin) according to microarray (a) and quantitative real time PCR (b). Light green, control sample; dark green, dehydration sample; lilac, rehydration sample. Error bars show the standard error of the mean of two to four biological replicates. The *y*-axis in (a) is in arbitrary fluorescence units scaled to a median of 10 000, and in (b) arbitrary fluorescence units normalized to the reference gene Phypa_173694.

response to water are significantly overrepresented among the upregulated genes after dehydration (q-value < 0.05; Table 3). As previously shown for protonema (Cuming *et al.*, 2007), effector genes like *lea* (late embryogenesis abundant; also known as dehydrins), were found to be upregulated after dehydration. An association of *lea* proteins with osmotic stress and response to ABA has also been described for mosses and seed plants in different studies (Kamisugi and Cuming, 2005; Olvera-Carrillo *et al.*, 2010). The two *lea* genes Phypa_108815 and Phypa_170009 are strongly expressed after dehydration and remain expressed after 1 h of rehydration (Figure 2a). They were used in this study to validate the microarray gene expression by quantitative real-time PCR (Figure 2b). As shown before for this platform (Busch *et al.* 2013) the data are in very good agreement.

The large-scale expression data enabled us to find reference genes on a more global scale than previously possible. We selected a new reference gene (Phypa_173694, a thioredoxin gene; Figure 2), which shows stable expression over all conducted microarray expression analyses (Figure S2). In contrast, out of 12 reference genes used in previous studies, and selected for phytohormone treatments in a recently published study (Le Bail *et al.*, 2013), only ARC34 (Phypa_146870) shows a globally stable expression over all microarray experiments (Figure S2). Due to the diverse set of conditions the present data are therefore well suited for the selection of reference genes allowing us to investigate a wide range of stages and perturbations.

Besides the two *lea* genes, 619 of the 690 genes upregulated after dehydration remain activated after rehydration. The same key stress regulator and effector genes activated during dehydration and rehydration might be explained by the comparable stress situations during those two treatments that represent the changes in the water regime that poikilohydric plants have to cope with. Activation during dehydration and rehydration was also seen for some *lea* genes in *Tortula* (*Syntrichia*) *ruralis* (Oliver *et al.*, 2005). The 71 genes which are upregulated after dehydration and downregulated after rehydration are linked to 'lipid biosynthetic processes' and acting in 'cytoskeleton organization' and 'lipid metabolic processes' as shown in GO term enrichment analyses (q-value < 0.05). Genes within these groups mainly encode for membrane repair proteins. This leads to the suggestion that the main damage to the membranes has already been repaired during dehydration, whereas most processes, like the downregulation of photosynthesis-related genes, are active during dehydration and

rehydration (Table S1). Our dehydration and rehydration data on gametophores extend the existing data on dehydration of protonemata (Cuming *et al.*, 2007) by adding rehydration and by detecting more genes as being differentially expressed under drought conditions.

Phylogenetic and comparative analysis of stress-mediating AP2/EREBP transcription factors

Of the 690 genes upregulated after dehydration, 126 were predicted to be responsive to ABA (Timmerhaus *et al.*, 2011), including members of the AP2/EREBP transcription factor family (Lang *et al.*, 2010). The AP2/EREBP family is involved in both salt stress and ABA responses in *P. patens* (Richardt *et al.*, 2010). Members of the AP2/EREBP family are also detected by ANOVA as upregulated (Figure 3, Table S2) in all of the available non-standard light conditions (strong light, sunlight and UV light). This evidence strengthens the suggestion that AP2/EREBP factors have a central regulatory role during stress conditions.

To detect evolutionarily conserved expression patterns in the AP2/EREBP family the differentially expressed genes (DEGs) were annotated in a phylogenetic tree of the gene family based on members from *Arabidopsis thaliana* (167 sequences), *Chlamydomonas reinhardtii* (14 sequences) and *P. patens* (156 sequences). Of 156 *Physcomitrella* sequences, 39 were detected by ANOVA with post-hoc testing as DEGs under the tested conditions (sunlight, strong light, darkness, UV-B, drought, pH shift, protonema in liquid culture and protoplastation). Interestingly, we noticed that one subclade within the tree contains most (69%) of the *P. patens* AP2-DEGs. These DEGs show diverse expression patterns under the different tested conditions (Figure 4). Only a few Arabidopsis sequences (10% of all sequences) and no *Chlamydomonas* sequences are found within this subclade. This particular subclade shows *P. patens* genes that are activated under several stress conditions, like Pp1s373_18V6.1 and Pp1s60_269V6.1 which are activated under protoplastation as well as drought, UV-B and sunlight (Figure 4). Such genes potentially represent upstream mediators that integrate different stress response pathways. In that regard they are similar to *A. thaliana* genes from the same subclade of the AP2 family which are activated under many or all of these stresses, most prominent among them being At3g50260.1 [encoding COOPERATIVELY REGULATED BY ETHYLENE AND JASMONATE 1 (CEJ1), also known as DREB AND EAR MOTIF PROTEIN 1 (DEAR 1)], which has been described as being involved in several stress pathways, namely cold, drought and defense to bacteria (Lamesch *et al.*, 2012). Based on its presence in the same subclade, and having a similar activation profile, we suggest that Pp1s60_269V6.1 might play a similar role in *P. patens*.

Within this subclade there are also more specialized genes that are induced only under specific conditions, like Pp1s60_228V6.1 after UV-B exposure or Pp1s199_50V6.1 during protoplastation (Figure 4). Again in the lower part of the tree there are also *A. thaliana* genes that show such a specific profile, for example At5g67190.1 (*DEAR2*), a close paralog of *DEAR1*, that is transcriptionally activated in protoplasts (Figure 4). Two *P. patens* genes, Pp1s199_50V6.1 and Pp1s240_84V6.1, show similar activation. Such genes are candidates for regulators that act downstream and thus mediate more specific responses, in this case stresses involved in protoplastation.

Integration into and availability in the GENEVESTIGATOR tool

In order to integrate the *P. patens* data into GENEVESTIGATOR (https://www.genevestigator.com), application ontologies were developed for tissue types, developmental stages and experimental factors (see Experimental Procedures) by adapting standard ontologies developed in collaboration with Plant Ontology (Walls *et al.*, 2012; Cooper *et al.*, 2013). GENEVESTIGATOR is not a microarray data analysis tool per se, but is a gene expression search engine that focuses on integrating the complete content and comparing results between experiments. As a starting point *P. patens* gene IDs (look-up between release versions and cross-links to other databases are provided online) can be selected, or sets of experiments. As an illustration, we used the Perturbations tool from the Gene Search toolset to identify the top five genes that are most strongly upregulated in individual perturbations but show minimal regulation in all other conditions. We then clustered these genes according to their expression profile in the perturbation and

	Strong light				Climate chamber				UV light				Dark			
	TAP		AP2/EREBP		TAP		AP2/EREBP		TAP		AP2/EREBP		TAP		AP2/EREBP	
	up	down	up	down	up	down	up	down	up	down	up	down	up	down	up	down
Sun light	13	7	4	0	53	5	8	1	111	48	18	11	162	11	17	1
Strong light					55	1	7	0	116	45	18	9	161	8	19	1
Climate chamber									54	94	11	15	139	11	10	2
UV light													118	47	15	14

Low High

Figure 3. Number of up- and downregulated transcriptional regulator genes and AP2/EREBP family members.
As found by ANOVA using the factor light intensity. Red colors show a low number of genes and green colors a high number of genes. Transcription associated proteins (TAPs) comprise transcription factors and general transcriptional regulators (Lang *et al.*, 2010). Genes are shown as up- and downregulated compared with the treatment in the first column (e.g. in strong light there are 13 TAPs upregulated and seven TAPs downregulated compared with in sunlight).

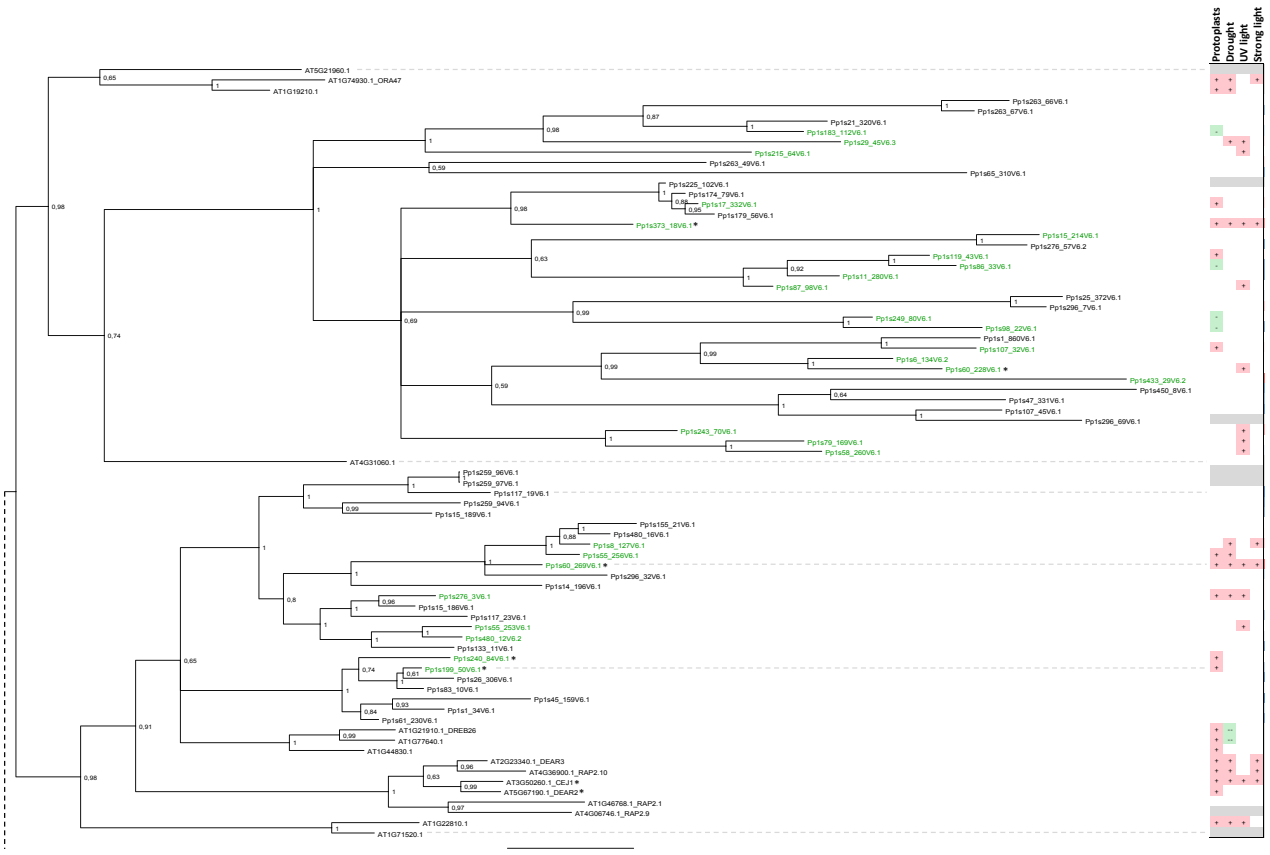

Figure 4. Rooted phylogenetic tree of AP2/EREBP family proteins.
Phylogenetic tree of part of the *Arabidopsis thaliana* (15 sequences) and *Physcomitrella patens* (60 sequences) AP2/EREBP family members; numbers at the nodes are support values (posterior probabilities from MRBAYES). The heatmap shows if the genes were detected via microarray analysis as upregulated (+/red) or downregulated (−/green) in the corresponding condition (see Experimental Procedures and Table S4 for *A. thaliana*) as compared with the control (see Experimental Procedures). No expression data were available for genes with gray bars. Green-colored identifiers are genes that were detected as differentially expressed in the above or one of the following additional conditions: sunlight, darkness, protonema in liquid culture (supplemented medium) and protonema in liquid culture at pH 4.5. Genes that are marked with an asterisk are mentioned in the Results: Pp1s373_18V6.1, Pp1s60_629V6.1, Pp1s60_228V6.1, Pp1s199_50V6.1, Pp1s240_84V6.1, At3g50260.1 (*DEAR1*/*CEJ1*) and At5g67190.1 (*DEAR2*).

development matrices using the Hierarchical Clustering tool. The results show a clustering of conditions that share gene-specificity profiles (Figure 5a), i.e. genes that were specifically upregulated in the chosen conditions but are unchanged in all other conditions. The clusters represented in Figure 5 are responsive to sunlight, protoplastation, photoperiod, UV-B, biotic stress and dehydration/rehydration, respectively. If these genes are plotted against tissue types/developmental stages (Figure 5b), their clustering reveals several distinct groups of genes that have quite different expression domains. The Gene Search tools in GENEVESTIGATOR further allow the identification of genes that have properties as defined by the user, for example being specifically expressed in a tissue, at a particular stage of development or in response to a perturbation. The search is performed by comparing the average expression in a target category (e.g. a chosen tissue type) with the sum of average expressions from a baseline set of categories (e.g. all tissue types). This approach allows us to look for genes that are generally specific for a chosen category (i.e. as compared with all other categories) or relatively specific (as compared with only a subset of categories). Due to the nature of the underlying data, such comparisons can only be performed between categories of the same type, such as tissues against tissues or perturbations against perturbations. Intuitive interfaces with checkboxes to choose categories of interest make it very straightforward for users to run this type of analysis.

DISCUSSION

Physcomitrella patens has been established as an important model for plant evolutionary developmental biology (e.g. Tanahashi *et al.*, 2005; Menand *et al.*, 2007; Mosquna *et al.*, 2009; Okano *et al.*, 2009; Khandelwal *et al.*, 2010; Khraiwesh *et al.*, 2010; Sakakibara *et al.*, 2013) and comparative genomics (e.g. O'Toole *et al.*, 2008; Rensing *et al.*, 2008; Peers *et al.*, 2009; Cutler *et al.*, 2010; Perez-Rodriguez *et al.*, 2010; Richardt *et al.*, 2010). In collaboration between

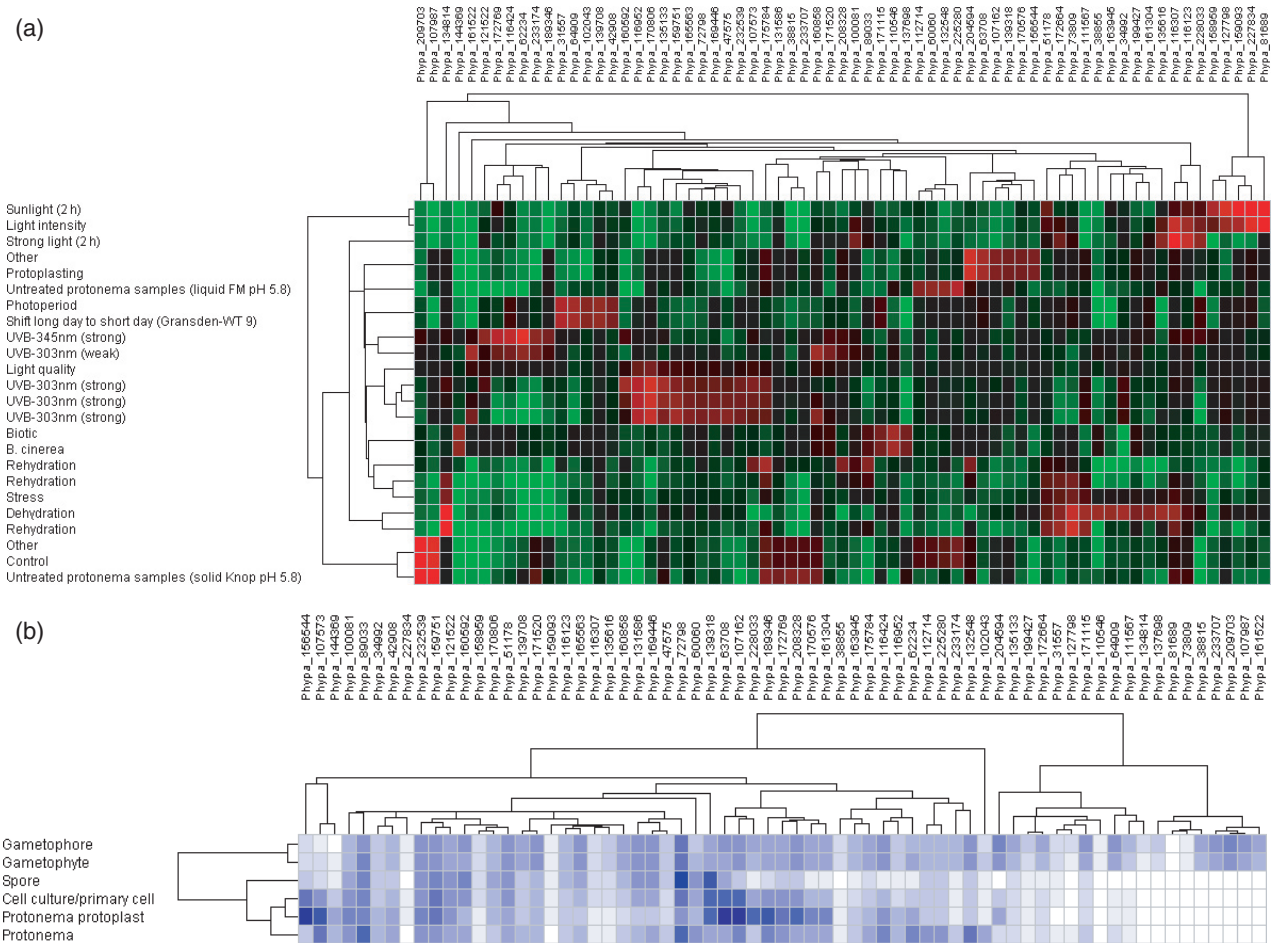

Figure 5. Snapshots of *Physcomitrella patens* data clustered in GENEVESTIGATOR.
(a) The Gene Search Perturbations tool was used to identify genes that are specifically upregulated in individual experimental conditions. The expression matrix from the combined list of most specific genes was clustered by genes and by perturbation type. Upregulation is shown in green, downregulation in red.
(b) The same list of genes was clustered by absolute expression across different stages of *P. patens* development. The more intense the blue color, the higher the gene expression.

the US Department of Energy and the International *Physcomitrella* Genome Consortium, the V3 plant flagship genome representation is a work in progress (pre-release available at http://phytozome.org/). Given the large and ever-increasing interest in this plant model, it was high time to create and make available expression profiling data for *P. patens*. We are confident that *P. patens* is a valuable addition to GENEVESTIGATOR that will not only represent a resource for research on *P. patens* as such but will also enhance cross-species comparisons of gene expression among photosynthetically active species. As shown, for example, for the dehydration and rehydration and culture condition analyses in this study, genes can be identified with specific functions under selected conditions. Also, expression profiles of transcription factors and their phylogenetic comparison with other plants (as shown for the example of AP2 here) can be used to find candidate regulators that coordinate responses to several stimuli or regulate specific pathways. The data can also be used to identify promoters that act under discrete conditions or to derive global control genes, for example for quantitative real-time PCR. The currently released data are soon to be complemented with more experimental conditions from the presented *P. patens* microarray platform (manuscripts in preparation). Moreover, a new publicly available microarray design has been generated based on the improved V1.6 gene annotation (Zimmer *et al.*, 2013) and *P. patens* is part of an RNA-seq pilot study conducted with the US Department of Energy to derive further expression profiles of development, stress and metabolic perturbations.

EXPERIMENTAL PROCEDURES

Plant material and growth conditions

The *P. patens* laboratory strain 'Gransden 2004' (Rensing *et al.*, 2008) was used for the majority of experiments. Since this strain

was subjected to selfing approximately once a year, an experiment using a parental strain, six selfing cycles away, was also conducted. In addition, experiments were carried out using the *P. patens* isolates 'Reute' and 'Villersexel' (McDaniel *et al.*, 2010), both displaying more sexual vigor than the Gransden strain. The former is genetically very close to Gransden (McDaniel *et al.*, 2010) while the latter exhibits a significant amount of polymorphism (von Stackelberg *et al.*, 2006) and has been used as the second parental strain for genetic mapping (Kamisugi *et al.*, 2008).

Plants were grown under long-day conditions (16 h white light, 8 h dark) in Knop medium at 20–25°C as previously described (Wolf *et al.*, 2010). Exceptions are listed in Table 1, summarizing all the xperimental conditions. Isolation of RNA and microarray processing were carried out as previously described, including array platform and design information (Wolf *et al.*, 2010).

Mapping gene IDs

The *P. patens* microarray expression data are based on V1.2 gene models. Phylogenetic analyses were conducted with the V1.6 gene models. Conversion between the identifiers was done using the mapping information available on http://cosmoss.org/.

Statistical testing

Microarray data processing was carried out as previously described (Wolf *et al.*, 2010). To detect differentially expressed genes the Cyber-T test was performed (Long *et al.*, 2001). All false discovery rate (FDR) corrections were carried out as described by Benjamini and Hochberg (1995). One-way analysis of variance (one-way ANOVA) used the hydration state or light intensity as factor and effect, respectively. States were defined as high (rehydration), low (dehydration) or normal (untreated control). For light intensities photosynthetically active radiation (UV-B) was used (see Table S3). An ANOVA post-hoc test was used to correct for multiple testing and calculated the FDR-corrected *P*-values (*q*-values) for all possible state combinations. The GO term enrichment analyses used Fisher's exact test to calculate *P*-values. Multiple testing corrected (Benjamini and Hochberg, 1995) *q*-values were calculated in R with the function p.adjust (R Development Core Team, 2008). Partial least squares (PLS) analysis used culture condition and tissue type as potential responses and analyzed their covariance with the activity factor tissue type. ANOVA with a post-hoc test, Cyber-T, hierarchical clustering and PLS were carried out with ANALYST 7.5 (Genedata, https://www.genedata.com/). The GO term enrichment analyses were conducted using in-house scripts. Clustering and visualization was carried out using GENEVESTIGATOR or ANALYST.

Phylogenetic analysis

The selection of sequences was done using an existing nucleic acid sequence-based phylogeny of all genes detected to be AP2 transcription factors (from *A. thaliana*, *C. reinhardtii* and *P. patens*) based on classification rules previously described (Lang *et al.*, 2010). Genes detected as differentially expressed in *P. patens* were annotated in the tree and the subclade containing most of the *P. patens* DEGs was selected for further analyses.

For this subtree (which did not contain *C. reinhardtii* sequences), the corresponding *P. patens* V1.6 protein sequences were retrieved from http://cosmoss.org/ and the *A. thaliana* sequences from TAIR 10 and aligned with MAFFT-LINSI (v.7.037b, http://mafft.cbrc.jp/alignment/software/). The alignment was manually curated with JALVIEW (v.2.8, http://www.jalview.org/). PROTTEST (v.3.3, http://code.google.com/p/prottest3/) was used to select the most suitable substitution model (JTT+I+G+F). The phylogenetic tree was constructed with the MRBAYES (v.3.2.2 × 64, http://mrbayes.sourceforge.net/) parallelized version using the above-mentioned model with eight gamma distributed rates, two hot and two cold chains and 50 burn-in trees. The run was stopped after the standard deviation of split frequencies dropped below 0.01 (1.4 million generations and with no remaining observable trend detectable in the overlay plot). The protein sequence subtree was rooted based on the outgroup information from the nucleotide tree. The curated alignment is available upon request.

Expression data and fold-change matrices for *A. thaliana* and *P. patens* were retrieved from GENEVESTIGATOR. For *A. thaliana* several studies existed for each of the conditions, and if one or more experiments showed an up- or downregulation it was marked with a + or −, respectively, in the heatmap visualization (Figure 4). Control experiments for *P. patens* fold changes were protonemata at pH 5.8 for the protoplasts and gametophores in the developmental stage 'gametophore formed' for drought, UV light and strong light.

Quantitative real-time PCR

For quantitative real-time PCR, RNA was reverse transcribed using SuperScript III (Invitrogen, http://www.invitrogen.com) and random hexamer primers (Fermentas, http://www.thermoscientific-bio.com/fermentas/). PRIMER3 (Untergasser *et al.*, 2012) was used for the design of specific oligonucleotides. Primer sequences used for amplification of the respective gene models are available upon request. For each 20-μl reaction, 20 ng of reverse-transcribed RNA was used and the reaction was carried out using SensiMix dT and SYBRGreen (Invitrogen) on a PicoReal Real-Time-PCR System (Thermo Scientific, http://www.thermoscientific.com). The concentration of cDNA was normalized to a thioredoxin transcript (Phypa_173694), showing expression-level and treatment-independent expression over all microarray analyses (cf. Experimental Procedures; Figure S2). The thioredoxin transcript was selected using a coefficient of variance filtering of the normalized mean expression values. Triplicate measurements were performed for each of two to three biological replicates. Analyses were performed with EXPRESSIONIST ANALYST 7.5 (Genedata).

Development of application ontologies for GENEVESTIGATOR integration

Data available in GENEVESTIGATOR are manually curated using a controlled vocabulary from sample description ontologies. In order to integrate the *P. patens* experimental data into GENEVESTIGATOR, application ontologies were developed by adapting standard ontologies developed in collaboration with the Plant Ontology Consortium (POC, released May 2011 on http://www.plantontology.org/; http://wiki.plantontology.org/index.php/Summary_of_Changes_to_PO_May_2011) and with the aid of expert knowledge. The ontologies, in particular the perturbation ontologies, are highly dynamic and will be adapted according to the growing database content. The current ontologies comprise 57 anatomical categories, 18 developmental stages and 33 defined perturbation-related comparisons. Many of those are already available as experimental data (Table 1).

Availability

All data have been made available in the public repository ARRAYEXPRESS (http://www.ebi.ac.uk/arrayexpress/) under the accession numbers shown in Table 1. The array design described here has recently been replaced by a new design (Nimblegen_Ppat_

SR_exp_HX12: Nimblegen 12 × 135 k chip, four 60mer probes per gene, V1.6 gene models) that is publicly available. The GENEVESTIGATOR tool and supporting documentation is available at https://www.genevestigator.com/gv/.

ACKNOWLEDGEMENTS

Funding by the German Research Foundation (RE 837/10-2 to RR and SAR) and by the German Federal Ministry of Education and Research (FKZ 0315057B to AM and GW and Freiburg Initiative for Systems Biology, FKZ 0313921, to RR and SAR) is gratefully acknowledged. Financial support of the DAAD to PS and MAA, and of the Alexander von Humboldt Foundation and the SNSF to PS, is also acknowledged. We are grateful to Stefanie Pilz for excellent technical assistance.

CONFLICT OF INTEREST

OL, RMM and PZ are/were employed by Nebion, the company providing GENEVESTIGATOR.

SUPPORTING INFORMATION

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

Table S1. Photosynthesis related genes (according to Gene Ontology annotation) downregulated during dehydration and rehydration.

Table S2. Members of the AP2/EREBP transcription factor family up- and downregulated by different light treatments.

Table S3. Experimental conditions of sun light, strong light, UV light, dehydration and rehydration and darkness.

Table S4. *Arabidopsis thaliana* studies chosen from GENEVESTIGATOR.

Figure S1. Hierarchical clustering of microarray experiments.

Figure S2. Bar charts of microarray expression values from published reference genes and Phypa_173694 (thioredoxin) in various conditions.

REFERENCES

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Methodol.* **57**, 289–300.

Brazma, A., Hingamp, P., Quackenbush, J. *et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29**, 365–371.

Busch, H., Boerries, M., Bao, J., Hanke, S.T., Hiss, M., Tiko, T. and Rensing, S.A. (2013) Network theory inspired analysis of time-resolved expression data reveals key players guiding P. patens stem cell development. *PLoS One*, **8**, e60494.

Cooper, L., Walls, R.L., Elser, J. ***et al.*** (2013) The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol.* **54**, e1.

Cuming, A.C., Cho, S.H., Kamisugi, Y., Graham, H. and Quatrano, R.S. (2007) Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss, *Physcomitrella patens*. *New Phytol.* **176**, 275–287.

Cutler, S.R., Rodriguez, P.L., Finkelstein, R.R. and Abrams, S.R. (2010) Abscisic acid: emergence of a core signaling network. *Annu. Rev. Plant Biol.* **61**, 651–679.

Frank, W., Decker, E.L. and Reski, R. (2005) Molecular tools to study *Physcomitrella patens*. *Plant Biol. (Stuttg)*, **7**, 220–227.

Hohe, A. and Reski, R. (2002) Optimisation of a bioreactor culture of the moss *Physcomitrella patens* for mass production of protoplasts. *Plant Sci.* **163**, 69–74.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W. and Zimmermann, P. (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinformatics*, **2008**, 420747.

Kamisugi, Y. and Cuming, A.C. (2005) The evolution of the abscisic acid-response in land plants: comparative analysis of group 1 LEA gene expression in moss and cereals. *Plant Mol. Biol.* **59**, 723–737.

Kamisugi, Y., von Stackelberg, M., Lang, D., Care, M., Reski, R., Rensing, S.A. and Cuming, A.C. (2008) A sequence-anchored genetic linkage map for the moss, *Physcomitrella patens*. *Plant J.* **56**, 855–866.

Khandelwal, A., Cho, S.H., Marella, H., Sakata, Y., Perroud, P.F., Pan, A. and Quatrano, R.S. (2010) Role of ABA and ABI3 in desiccation tolerance. *Science*, **327**, 546.

Khraiwesh, B., Arif, M.A., Seumel, G.I., Ossowski, S., Weigel, D., Reski, R. and Frank, W. (2010) Transcriptional control of gene expression by microRNAs. *Cell*, **140**, 111–122.

Lagaert, S., Belien, T. and Volckaert, G. (2009) Plant cell walls: protecting the barrier from degradation by microbial enzymes. *Semin. Cell Dev. Biol.* **20**, 1064–1073.

Lamesch, P., Berardini, T.Z., Li, D. ***et al.*** (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210.

Lang, D., Zimmer, A.D., Rensing, S.A. and Reski, R. (2008) Exploring plant biodiversity: the *Physcomitrella* genome and beyond. *Trends Plant Sci.* **13**, 542–549.

Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riano-Pachon, D.M., Correa, L.G., Reski, R., Mueller-Roeber, B. and Rensing, S.A. (2010) Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. *Genome Biol. Evol.* **2**, 488–503.

Le Bail, A., Scholz, S. and Kost, B. (2013) Evaluation of reference genes for RT qPCR analyses of structure-specific and hormone regulated gene expression in *Physcomitrella patens* gametophytes. *PLoS One*, **8**, e70998.

Long, A.D., Mangalam, H.J., Chan, B.Y., Tolleri, L., Hatfield, G.W. and Baldi, P. (2001) Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in *Escherichia coli* K12. *J. Biol. Chem.* **276**, 19937–19944.

McDaniel, S.F., von Stackelberg, M., Richardt, S., Quatrano, R.S., Reski, R. and Rensing, S.A. (2010) The speciation history of the Physcomitrium-Physcomitrella species complex. *Evolution*, **64**, 217–231.

McQueen-Mason, S., Durachko, D.M. and Cosgrove, D.J. (1992) Two endogenous proteins that induce cell wall extension in plants. *Plant Cell*, **4**, 1425–1433.

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

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

Mueller, S.J., Lang, D., Hoernstein, S.N. ***et al.*** (2014) Quantitative analysis of the mitochondrial and plastid proteomes of the moss *Physcomitrella patens* reveals protein macrocompartmentation and microcompartmentation. *Plant Physiol.* **164**, 2081–2095.

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

Oliver, M.J., Velten, J. and Mishler, B.D. (2005) Desiccation tolerance in bryophytes: a reflection of the primitive strategy for plant survival in dehydrating habitats? *Integr. Comp. Biol.* **45**, 788–799.

Olvera-Carrillo, Y., Campos, F., Reyes, J.L., Garciarrubio, A. and Covarrubias, A.A. (2010) Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in *Arabidopsis*. *Plant Physiol.* **154**, 373–390.

O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M. and Small, I. (2008) On the expansion of the pentatricopeptide repeat gene family in plants. *Mol. Biol. Evol.* **25**, 1120–1128.

Peers, G., Truong, T.B., Ostendorf, E., Busch, A., Elrad, D., Grossman, A.R., Hippler, M. and Niyogi, K.K. (2009) An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature*, **462**, 518–521.

Perez-Rodriguez, P., Riano-Pachon, D.M., Correa, L.G., Rensing, S.A., Kersten, B. and Mueller-Roeber, B. (2010) PlnTFDB: updated content and

new features of the plant transcription factor database. *Nucleic Acids Res.* **38**, D822–D827.

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

Quatrano, R.S., McDaniel, S.F., Khandelwal, A., Perroud, P.F. and Cove, D.J. (2007) *Physcomitrella patens*: mosses enter the genomic age. *Curr. Opin. Plant Biol.* **10**, 182–189.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

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 Cove, D.J. (2004) Quick guide: *Physcomitrella patens*. *Curr. Biol.* **14**, R261–R262.

Richardt, S., Timmerhaus, G., Lang, D., Qudeimat, E., Correa, L.G., Reski, R., Rensing, S.A. and Frank, W. (2010) Microarray analysis of the moss *Physcomitrella patens* reveals evolutionarily conserved transcriptional regulation of salt stress and abscisic acid signalling. *Plant Mol. Biol.* **72**, 27–45.

Rocca-Serra, P., Brazma, A., Parkinson, H. *et al.* (2003) ArrayExpress: a public database of gene expression data at EBI. *C R Biol.* **326**, 1075–1078.

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

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

von Stackelberg, M., Rensing, S.A. and Reski, R. (2006) Identification of genic moss SSR markers and a comparative analysis of twenty-four algal and plant gene indices reveal species-specific rather than group-specific characteristics of microsatellites. *BMC Plant Biol.* **6**, 9.

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

Timmerhaus, G., Hanke, S.T., Buchta, K. and Rensing, S.A. (2011) Prediction and validation of promoters involved in the abscisic acid response in *Physcomitrella patens*. *Mol. Plant*, **11**, 11.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. (2012) Primer3–new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115.

Walls, R.L., Athreya, B., Cooper, L. *et al.* (2012) Ontologies as integrative tools for plant science. *Am. J. Bot.* **99**, 1263–1275.

Wolf, L., Rizzini, L., Stracke, R., Ulm, R. and Rensing, S.A. (2010) The molecular and physiological response of *Physcomitrella patens* to UV-B radiation. *Plant Physiol.* **153**, 1123–1134.

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

Zimmermann, P., Schildknecht, B., Craigon, D. *et al.* (2006) MIAME/Plant – adding value to plant microarrray experiments. *Plant Methods*, **2**, 1.

