Large scale gene expression profiling data of the model moss *Physcomitrella patens* help to understand developmental progression, culture and stress conditions

*Resource Article*

Hiss M.0,1,2, Laule O.4, Meskauskiene R.M.4, Arif M.A.5,6, Decker E.L.5, Erxleben A.5,12, Frank W.6, Hanke S.T.0,1, Lang D.5, Martin A.5,7, Neu C.1, Reski R.2,3,5,13, Richardt S.1,9, Schallenberg-Rüdinger M.0, Szövényi P.1,11, Tiko T.1, Wiedemann G.5, Wolf L.1,10, Zimmermann P.4, Rensing S.A.0,1,2,3

0 Plant Cell Biology, Faculty of Biology, University of Marburg, Karl-von-Frisch-Str. 8, 35043 Marburg, Germany

1 Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104 Freiburg, Germany

2 FRISYS Freiburg Initiative for Systems Biology, University of Freiburg, 79104 Freiburg, Germany

3 BIOSS Centre for Biological Signalling Studies, University of Freiburg, 79104 Freiburg, Germany

4 NEBION AG, Hohlstrasse 515, 8048 Zurich, Switzerland

5 Plant Biotechnology, Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104 Freiburg, Germany
6 Plant Molecular Cell Biology, Department Biology I, Ludwig-
Maximilians-University Munich, LMU Biocenter,
Grosshadernerstr. 2-4, 82152 Planegg-Martinsried, Germany

7 Present address: Thermo Fisher, Opelstraße 9, 68789 St. Leon-
Rot, Germany

9 Present address: Roche Diagnostics GmbH, Nonnenwald 2,
82377 Penzberg, Germany

10 Present address: Biozentrum, University of Basel,
Klingelbergstrasse 50/70, 4056 Basel, Switzerland

11 Present address: Institute of Evolutionary Biology and
Environmental Studies, University of Zurich, Winterthurstr. 190,
8057 Zurich, Switzerland

12 Present address: Pharmaceutical Bioinformatics, Faculty of
Chemistry, Pharmacy, and Earth Sciences, University of Freiburg,
Hermann-Herder-Str. 9, 79104 Freiburg, Germany

13 FRIAS - Freiburg Institute for Advanced Studies, University of
Freiburg, 79104 Freiburg, Germany
Authors for correspondence:

Stefan A. Rensing
Tel. +49 6241 28 21940
Fax +49 6421 28 22190
stefan.rensing@biologie.uni-marburg.de

Philip Zimmermann
Tel. +41 44 500 44 00
phz@nebion.com

Word Count

Main body 3856
Introduction 260
Results 2177
Discussion 323
Materials and Methods 1018
Acknowledgements 78
Number of figures 5
Number of tables 3
Summary

The moss *Physcomitrella patens* is an important model organism to study 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 and NAC 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](https://www.genevestigator.com). By providing large scale expression profiles, the usability of this model organism is further enhanced, e.g. 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 \textit{in silico} and to generate hypotheses to subsequently put to test. Through the availability of expression data in compliance with the MIAME standard (Brazma \textit{et al.} 2001, Zimmermann \textit{et al.} 2006) in repositories such as ArrayExpress (Rocca-Serra \textit{et al.} 2003), incorporation into meta-analysis tools such as Genevestigator (Hruz \textit{et al.} 2008) becomes feasible. The availability of data in such a tool allows for the end user to browse through experiments conducted by different labs or using different technology platforms with ease. Moreover, the use of anatomy, development and treatment ontologies allows to analyze e.g. developmental progression or to compare expression data across taxonomic boundaries.

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 (Fig. S1). In the following, we have conducted some example analyses to demonstrate the usefulness of the data to analyze development, culture conditions and stress.

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 vs. solid medium, while 247 genes are differentially expressed...
between protonemata and gametophores on solid medium. Partial least squares analysis confirms that the majority of expression differences are covariant with the two principal growth stages, protonemata and gametophores (Fig. 1), which is therefore dominant over culture form with regard to alteration of the gene expression profile.

**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 conducted during liquid culture regime does not seem to constitute a stress condition, unlike e.g. darkness or dehydration (see below), where stress terms are found to be over-represented among the differentially regulated genes.

**Acidic medium might aid protoplastation by downregulation of cell wall structure genes**

It is recommended to grow plant material which will be used for protoplast preparation 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 than plants
growing at pH 5.8 (the cell wall can more easily be digested to produce protoplasts, as measured from a higher rate of released protoplasts). Comparing the gene expression data of tissue growing on pH 5.8 and on pH 4.5 should give insights in the activation of genes responsible for the change in cell wall composition.

By the shift to pH 4.5, 671 genes are found to be upregulated and 270 are downregulated (Cyber-T, q < 0.05). Among the upregulated genes are only three that are annotated as involved in “cell wall catabolism”, however 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 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 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 already grown at pH 5.8, i.e. mildly acidic. In case of P. patens the downregulation of specific cell wall structure genes might influence the cell wall composition, leading to a better digestibility
of the cell wall. Another explanation might be that only newly formed cells start changing their cell wall composition at pH 4.5 (Hohe and Reski 2002).

De- and rehydration response are mediated by the same gene pool

Statistical analysis (cf. Methods) on the gene expression of all expressed genes from dehydrated (50% fresh weight loss), rehydrated and untreated gametophores finds 690 genes upregulated and 1,231 genes downregulated after one hour of dehydration out of overall 26,853 genes on the microarray. Genes important in regulation of transcription, protein modification and 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), resp. dehydrins, were found to be upregulated after dehydration. 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 one hour of rehydration (Fig. 2A). They were used in this study to validate the microarray gene expression by quantitative Real Time PCR (Fig. 2B). As
shown before for this platform (Busch et al. 2010) the data are in very good congruence.

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; Fig. 2), which shows stable expression over all conducted microarray expression analyses (Fig. S2). In contrast, out of twelve reference genes used in previous studies, resp. 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 (Fig. S2). The present data, due to the diverse set of conditions, are therefore well suited to select reference genes that allow to tackle 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 with comparable stress situations during those two treatments that represent changes of water regime 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 de- and downregulated after
rehydration are linked to “lipid biosynthetic process”, acting in “cytoskeleton organization” and “lipid metabolic process” 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 is already repaired during dehydration, whereas most processes, like the downregulation of photosynthesis related genes, are active during de- and rehydration (Table S1). Our de- and rehydration data on gametophores extend the existing data on dehydration of protonemata (Cuming et al. 2007) by adding rehydration and by detection of more genes as being differentially expressed under drought.

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

Of the 690 genes upregulated after dehydration, 126 were predicted to be ABA responsive (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 (Fig. 3, Table S2) in all of the available non-standard light conditions (strong light, sun light and UV light).
These evidences strengthen the suggestion of the central regulatory role of AP2/EREBP factors during stress conditions.

To detect evolutionary 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 *Physcomitrella patens* (156 sequences). Of 156 *Physcomitrella* sequences, 39 were detected by ANOVA with posthoc testing as DEGs under the tested conditions (sun light, 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 (Fig. 4). Within this subclade only a few *Arabidopsis* sequences (10% of all sequences) and no *Chlamydomonas* sequences are found. This particular subclade shows *P. patens* genes that are activated under several stress conditions, like Pp1s373_18V6.1 and Pp1s60_269V6.1 that are activated under protoplastation as well as drought, UV-B and sun light (Fig. 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 that are activated.
under many or all of these stresses, most prominent among them

At3g50260.1 (COOPERATIVELY REGULATED BY ETHYLENE AND JASMONATE 1 / CEJ1, also known as DREB AND EAR MOTIF PROTEIN 1 / DEAR 1) that has been described to be involved in several stress pathways, namely cold, drought and defense to bacteria (Lamesch et al. 2012). Based on being present 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 (Fig. 4). Again in the lower part of the tree there are also *A. thaliana* genes that show such a specific profile, e.g. At5g67190.1 (DEAR2), a close paralog of DEAR1, that is transcriptionally activated in protoplasts (Fig. 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 that 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 (cf. Methods) by adapting standard ontologies developed in collaboration with the Plant Ontology (Cooper *et al.* 2013, Walls *et al.* 2012). 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 start point, *P. patens* gene ids (lookup 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 up-regulated 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 development matrices using the Hierarchical Clustering tool. The results show a clustering of conditions that share gene-specificity profiles (Fig. 5A), i.e. genes that were specifically upregulated in the chosen conditions but are unchanged in all other conditions. The clusters represented in Fig. 5 are responsive to sunlight, protoplastation, photoperiod, UV-B, biotic stress and de-
If these genes are plotted against tissue types/developmental stages (Fig. 5B), their clustering reveals several distinct groups of genes that have quite different expression domains. The Gene Search tools in Genevestigator further allow identifying genes that have properties as defined by the user, for example being specifically expressed in a tissue, at a 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 looking for genes that are generally specific for a chosen category (i.e. as compared to all other categories), or relatively specific (as compared to 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


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 enhance cross-species comparisons of gene expression among photosynthetically active species. As shown for e.g. the de- 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 exemplarily shown for 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, e.g. 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 U.S. Department of Energy to derive further expression profiles of development, stress and metabolic perturbations.
Materials and Methods

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 polymorphisms (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 (16h white light, 8h dark) in Knop medium at 20-25 Centigrade as previously described (Wolf et al. 2010). Exceptions are listed in Table 1, summarizing all experimental conditions. Isolation of RNA and microarray processing was 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. The conversion between the identifiers was done using the mapping information available on 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 PAR, resp. UV-B was (cf. Table S3). An ANOVA posthoc test was used to correct for multiple testing and calculates 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 posthoc test, Cyber-T, hierarchical clustering and PLS were carried out with Analyst 7.5 (Genedata, Basel, Switzerland). 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 based on 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 Physcomitrella patens were annotated in the tree and the subclade containing most of the P. patens DEGs was selected for further analyses.
For this subtree (that did not contain C. reinhardtii sequences), the corresponding P. patens V1.6 protein sequences were retrieved from cosmoss.org, the A. thaliana sequences from TAIR 10 and aligned with MAFFT-LINSI (v7.037b). The alignment was manually curated with Jalview (v 2.8). Prottest (v 3.3) 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 x64) 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 experiment showed an up- or downregulation it was marked with an + or – respectively in the heatmap visualization (Fig. 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](http://www.invitrogen.com)) and random hexamer primers (Fermentas, [http://www.thermoscientificbio.com/fermentas/](http://www.thermoscientificbio.com/fermentas/)).
(Untergasser *et al.* 2012) was used for 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](http://www.thermoscientific.com)). Concentration of cDNA was normalized to a thioredoxin transcript (Phypa_173694), showing expression level and treatment-independent expression over all microarray analyses (*cf.* Methods, Fig. 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, Basel, Switzerland).

**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 www.plantontology.org, http://wiki.plantontology.org:8080/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 relieved by a new design (Nimblegen_Ppat_SR_exp_HX12: Nimblegen 12x135k chip, four 60mer probes per gene, v1.6 gene models) that is publicly available. The Genevestigator tool and supporting documentation is available at www.genevestigator.com.
Acknowledgements

Funding by the German Research Foundation (RE 837/10-2 to R.R. and S.A.R.) and by the German Federal Ministry of Education and Research (FKZ 0315057B to A.M. and G.W. and Freiburg Initiative for Systems Biology, FKZ 0313921, to R.R. and S.A.R.) is gratefully acknowledged. Financial support of the DAAD to P.S. and M.A.A., and of the Alexander von Humboldt Foundation and the SNSF to P.S., is also acknowledged. We are grateful to Stefanie Pilz for excellent technical assistance.

Conflict of interest

O.L., R.M.M. and P.Z. are/were employed by Nebion, the company providing Genevestigator.
Literature cited


R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*.
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,
evolutionary insights into the conquest of land by plants.
*Science*, 319, 64-69.


Richardt, S., Timmerhaus, G., Lang, D., Qudeimat, E., Correa,
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., Sarkans, U.,
Shojatalab, M., Contrino, S., Vilo, J.,
Abeygunawardena, N., Mukherjee, G., Holloway, E.,
Kapushesky, M., Kemmeren, P., Lara, G.G., Oezcimen,
database of gene expression data at EBI. *C R Biol*, 326,
1075-1078.


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 timeseries of detached leaflets (one replicate for each timepoint) and the developmental stage “gametophore formed” (four replicates).

<table>
<thead>
<tr>
<th>Tissue/Treatment</th>
<th>Bulk accession nr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spores</td>
<td>E-MTAB-916</td>
</tr>
<tr>
<td>Protoplasts</td>
<td>E-MTAB-976</td>
</tr>
<tr>
<td>Protonemata</td>
<td>E-MTAB-976, E-MTAB-917</td>
</tr>
<tr>
<td>Gametophores</td>
<td>E-MTAB-916, E-MTAB-917</td>
</tr>
<tr>
<td>Germination (protonemal development): germinated spores</td>
<td>E-MTAB-916</td>
</tr>
<tr>
<td>primarily chloronemata</td>
<td>E-MTAB-976, E-MTAB-917</td>
</tr>
<tr>
<td>Gametophore growth: gametophore formed</td>
<td>E-MTAB-917</td>
</tr>
<tr>
<td>Gametangia development: mature antheridia/archegonia</td>
<td>E-MTAB-917</td>
</tr>
<tr>
<td>different genotype (Reute, Villersexel)</td>
<td>E-MTAB-916</td>
</tr>
</tbody>
</table>

Perturbations:

<table>
<thead>
<tr>
<th>Biotic</th>
<th>Bulk accession nr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botrytis cinerea inoculation</td>
<td>E-MTAB-919</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Light intensity</th>
<th>Bulk accession nr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Darkness</td>
<td>E-MTAB-913</td>
</tr>
<tr>
<td>Strong light</td>
<td>E-MTAB-913</td>
</tr>
<tr>
<td>Sunlight</td>
<td>E-MTAB-913</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Light quality</th>
<th>Bulk accession nr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>UV-B (303nm) supplementation, different intensities</td>
<td>E-MEXP-2508</td>
</tr>
<tr>
<td>Shift long day to short day</td>
<td>E-MTAB-917</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Water stress</th>
<th>Bulk accession nr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dehydration</td>
<td>E-MTAB-914</td>
</tr>
<tr>
<td>Rehydration</td>
<td>E-MTAB-914</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other</th>
<th>Bulk accession nr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different growth media</td>
<td>E-MTAB-976</td>
</tr>
<tr>
<td>Shift from pH 5.8 to pH 4.5</td>
<td>E-MTAB-976</td>
</tr>
<tr>
<td>Timeseries of detached leaflets</td>
<td>E-MTAB-915</td>
</tr>
</tbody>
</table>
**Table 2: 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 pH-shift from 5.8 to 4.5.

<table>
<thead>
<tr>
<th>GO-Term</th>
<th>number</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>unsaturated fatty acid biosynthetic process</td>
<td>5</td>
<td>1.7E-04</td>
</tr>
<tr>
<td>glucose metabolic process</td>
<td>9</td>
<td>5.0E-03</td>
</tr>
<tr>
<td>plant-type cell wall organization</td>
<td>4</td>
<td>9.1E-03</td>
</tr>
<tr>
<td>fatty acid biosynthetic process</td>
<td>5</td>
<td>2.3E-02</td>
</tr>
<tr>
<td>spermidine biosynthetic process</td>
<td>2</td>
<td>2.5E-02</td>
</tr>
<tr>
<td>carbon fixation</td>
<td>3</td>
<td>2.5E-02</td>
</tr>
<tr>
<td>NAD biosynthetic process</td>
<td>2</td>
<td>3.1E-02</td>
</tr>
<tr>
<td>pyridine nucleotide biosynthetic process</td>
<td>2</td>
<td>3.8E-02</td>
</tr>
<tr>
<td>glucose catabolic process</td>
<td>6</td>
<td>4.1E-02</td>
</tr>
<tr>
<td>lipid biosynthetic process</td>
<td>7</td>
<td>4.1E-02</td>
</tr>
<tr>
<td>fatty acid metabolic process</td>
<td>5</td>
<td>4.4E-02</td>
</tr>
<tr>
<td>polyamine metabolic process</td>
<td>2</td>
<td>4.6E-02</td>
</tr>
</tbody>
</table>
**Table 3: GO term enrichment analysis - de- and rehydration**

List of significantly overrepresented GO terms (biological process ontology) and corresponding q-values from the enrichment analysis for genes upregulated after dehydration.

<table>
<thead>
<tr>
<th>GO Term</th>
<th>number of genes</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>regulation of transcription, DNA-dependent</td>
<td>37</td>
<td>4.2E-04</td>
</tr>
<tr>
<td>regulation of transcription</td>
<td>38</td>
<td>4.2E-04</td>
</tr>
<tr>
<td>protein modification process</td>
<td>57</td>
<td>3.7E-03</td>
</tr>
<tr>
<td>protein ubiquitination</td>
<td>18</td>
<td>3.7E-03</td>
</tr>
<tr>
<td>metal ion transport</td>
<td>12</td>
<td>1.3E-02</td>
</tr>
<tr>
<td>response to water</td>
<td>2</td>
<td>2.0E-02</td>
</tr>
<tr>
<td>protein amino acid dephosphorylation</td>
<td>5</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>SRP-dependent cotranslational protein targeting to membrane</td>
<td>2</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>lipid metabolic process</td>
<td>14</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>cation transport</td>
<td>16</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>mitochondrial transport</td>
<td>3</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>exocytosis</td>
<td>3</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>cell adhesion</td>
<td>3</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>oligosaccharide metabolic process</td>
<td>3</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>guanosine tetraphosphate metabolic process</td>
<td>2</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>lipid glycosylation</td>
<td>2</td>
<td>3.2E-02</td>
</tr>
<tr>
<td>ion transport</td>
<td>17</td>
<td>3.8E-02</td>
</tr>
<tr>
<td>steroid metabolic process</td>
<td>2</td>
<td>4.0E-02</td>
</tr>
</tbody>
</table>
Figure Legends

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

Fig. 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 (thioredoxin) according to microarray (A) and quantitative RealTime PCR (B). Light green: control sample, dark green: dehydration sample and lilac: rehydration sample. Error bars show 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 arbitrary fluorescence units normalized to the reference gene Phypa_173694 in (B).

**Fig. 3:** Number of up- and downregulated transcriptional regulator genes and AP2/EREBP family members

As found by ANOVA using the factor light intensity – Red color shows low number of genes, green color shows high number of genes. Transcription associated proteins (TAPs) comprise transcription factors and general transcriptional regulators (Lang, Weiche, Timmerhaus, Richardt, Riano-Pachon, Correa, Reski, Mueller-Roeber and Rensing 2010). Genes are shown as up- and downregulated as compared to the treatment in the first column (e.g. in strong light there are 13 TAPs up- and seven TAPs down/regulated as compared to sun light).

**Fig 4:** Rooted phylogenetic tree of AP2/EREBP family proteins.

Phylogenetic tree of part of the *A. thaliana* (15 sequences) and *P. 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 up- (+/red) or downregulated (-/green) in the corresponding condition (cf. Methods/Table S4 for A. thaliana) as compared to the control (cf. Methods). For genes with grey bars no expression data was available. Green colored identifiers are genes that were detected as differentially expressed in the above or one of the following additional conditions: sun light, 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 part: Pp1s373_18V6.1, Pp1s60_629V6.1, Pp1s60_228V6.1, Pp1s199_50V6.1, Pp1s240_84V6.1, At3g50260.1 (DEAR1/CEJ1) and At5g67190.1 (DEAR2).

Fig 5: Snapshots of P. 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.
Figures

Fig. 1

Fig. 2
<table>
<thead>
<tr>
<th></th>
<th>strong light</th>
<th>climate chamber</th>
<th>UV light</th>
<th>dark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>up</td>
<td>down</td>
<td>up</td>
<td>down</td>
</tr>
<tr>
<td>sun light</td>
<td>13</td>
<td>7</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>strong light</td>
<td>55</td>
<td>1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>climate chamber</td>
<td>54</td>
<td>54</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>UV light</td>
<td>118</td>
<td>47</td>
<td>15</td>
<td>14</td>
</tr>
</tbody>
</table>

**Fig. 3**

**Fig. 4**
Fig. 5