the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2020) **102,** 165–177 doi: 10.1111/tpj.14607

RESOURCE

PEATmoss (*Physcomitrella* Expression Atlas Tool): a unified gene expression atlas for the model plant *Physcomitrella patens*

Noe Fernandez-Pozo[1], Fabian B. Haas[1], Rabea Meyberg[1], Kristian K. Ullrich[1,2], Manuel Hiss[1], Pierre-François Perroud[1], Sebastian Hanke[1], Viktor Kratz[1], Adrian F. Powell[3], Eleanor F. Vesty[4], Christopher G. Daum[5], Matthew Zane[5], Anna Lipzen[5], Avinash Sreedasyam[6], Jane Grimwood[6], Juliet C. Coates[4], Kerrie Barry[5], Jeremy Schmutz[5,6], Lukas A. Mueller[3] and Stefan A. Rensing[1,7,8,*]

[1]*Plant Cell Biology, Faculty of Biology, University of Marburg, Marburg, Germany,*
[2]*Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Ploen, Germany,*
[3]*Boyce Thompson Institute, Ithaca, NY, USA,*
[4]*School of Biosciences, University of Birmingham, Birmingham, UK,*
[5]*US Department of Energy (DOE) Joint Genome Institute, Walnut Creek, CA 94598, USA,*
[6]*HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA,*
[7]*BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany, and*
[8]*LOEWE Center for Synthetic Microbiology (SYNMIKRO), University of Marburg, Germany*

Received 26 March 2019; revised 14 October 2019; accepted 30 October 2019; published online 12 November 2019.
*For correspondence: (e-mail: noefp@uni-marburg.de; stefan.rensing@biologie.uni-marburg.de).

SUMMARY

Physcomitrella patens **is a bryophyte model plant that is often used to study plant evolution and development. Its resources are of great importance for comparative genomics and evo-devo approaches. However, expression data from *Physcomitrella patens* were so far generated using different gene annotation versions and three different platforms: CombiMatrix and NimbleGen expression microarrays and RNA sequencing. The currently available *P. patens* expression data are distributed across three tools with different visualization methods to access the data. Here, we introduce an interactive expression atlas, *Physcomitrella* Expression Atlas Tool (PEATmoss), that unifies publicly available expression data for *P. patens* and provides multiple visualization methods to query the data in a single web-based tool. Moreover, PEATmoss includes 35 expression experiments not previously available in any other expression atlas. To facilitate gene expression queries across different gene annotation versions, and to access *P. patens* annotations and related resources, a lookup database and web tool linked to PEATmoss was implemented. PEATmoss can be accessed at https://peatmoss.online.uni-marburg.de**

Keywords: *Physcomitrella patens*, expression atlas, RNA-seq, bioinformatics, evolution, development, plant hormones, light, annotation.

INTRODUCTION

Physcomitrella patens is a bryophyte model plant essential for the study of plant evolution. As a bryophyte, its sister position to vascular plants is of special interest to perform evolutionary developmental (evo-devo) approaches (Rensing, 2017, 2018). The *P. patens* life cycle is predominantly haploid (Figure 1), facilitating functional genomics studies via homologous recombination or genome editing to understand gene function in the plant.

The available genomic and transcriptomic resources for bryophytes and streptophyte algae, key species to understand the transition from water to land of plant ancestors, are under-represented in comparison to angiosperms (Rensing, 2017). The V1 draft genome sequence and annotation of *P. patens* have been available since 2008 (Rensing *et al.*, 2008). Continuous work for 10 years to improve the resources and knowledge about *P. patens* resulted in the V3 chromosome-scale genome assembly and the most

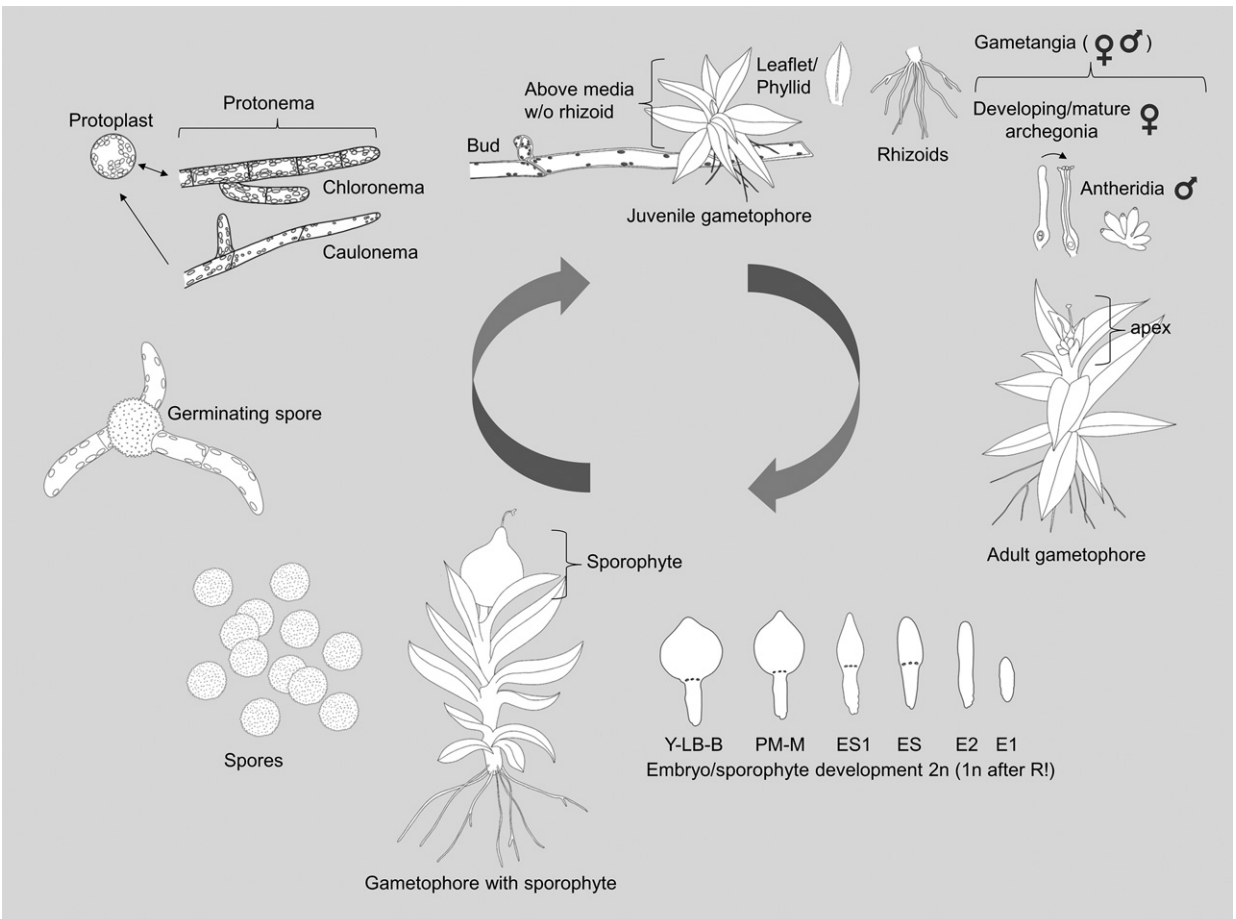

Figure 1. *Physcomitrella patens* life cycle. All stages available in PEATmoss are included. All developmental stages are haploid except the sporophyte. The embryo/sporophyte development is divided in the stages Embryo 1 (E1), Embryo 2, Early sporophyte (ES), Early sporophyte 1 (ES1), Pre-meiotic – Meiotic sporophyte (PM-M) and Yellow, Light Brown and Brown sporophyte (Y-LB-B) based on developmental stages defined in (Hiss *et al.*, 2017). R! in the sporophyte development stands for reduction cell division after meiosis, when haploid spores are formed.

recent gene annotation v3.3 (Lang *et al.*, 2018). These resources make *P. patens* one of the best studied non-seed plants and a good reference for comparative genomics, especially in cross-lineage studies. As such, it is one of the plant flagship species tackled in the US DOE gene atlas project (Perroud *et al.*, 2018). However, the available gene expression data for *P. patens* are fragmented in three web tools/databases, comprising 77 experiments to date in the eFP Browser (Winter *et al.*, 2007; Ortiz-Ramirez *et al.*, 2016), Genevestigator (Hruz *et al.*, 2008) and Phytozome (Goodstein *et al.*, 2012), each with different advantages and limitations. We use the term experiment for all replicates generated from the same experimental condition.

The eFP (electronic fluorescent pictograph) Browser for *P. patens* (http://bar.utoronto.ca/efp_physcomitrella/cgi-bin/efpWeb.cgi) displays the expression values as coloured cartoons representing the plant or plant parts, in which different colours represent different expression values. Expression values can also be displayed as bar plots. However, in the eFP Browser, only one gene can be queried at a time or the relative expression of two genes can be compared using a different colour scale. No information about the expression values of replicates is shown. Only one data set is currently available at the eFP Browser. This data set contains 11 experiments (three replicates each) based on the gene annotation v1.6 and NimbleGen microarray (Ortiz-Ramirez *et al.*, 2016), representing most of the developmental stages of the *P. patens* life cycle (Figure 1).

Alternatively, the commercial distribution of Genevestigator (https://genevestigator.com/) contains many tools for expression visualization of multiple genes. However, the free version is limited to only few basic visualization plots and only one gene per query. It contains 34 experiments (varying levels of replication), based on the gene annotation v1.2 and CombiMatrix microarray, under different conditions of light, pH, hydration/dehydration and biotic stress as well as hormone treatments (Busch *et al.*, 2013; Hiss *et al.*, 2014).

The third database with *P. patens* expression data is Phytozome (https://phytozome.jgi.doe.gov/). Here, the expression values of a single gene can be visualized in a table, together with the experiment name and a tag for high or low expressed genes. A very useful feature on the Phytozome expression tab is to show the list of co-expressed genes and their correlation value for the query gene. This feature is helpful for finding genes that might be involved in similar biological processes to the query gene (Ruprecht *et al.*, 2017). Phytozome includes 32 *P. patens* RNA-seq experiments (three replicates each), based on the most recent v3.3 gene annotation, and representing multiple conditions such as developmental stages, light perturbation, dehydration and hormone application (Perroud *et al.*, 2018).

Here we introduce PEATmoss (*Physcomitrella* Expression Atlas Tool), a gene expression atlas to unify the expression data of *P. patens*. This web tool is based on the well accepted Tomato Expression Atlas (Fernandez-Pozo *et al.*, 2017), is comprised of 109 experiments, provides multiple visualization methods and is available at https://peatmoss.online.uni-marburg.de.

A current limitation of *P. patens* data is that published resources are based on different gene annotation versions. Using the lookup table (Supporting Information Data S1) from Perroud *et al.* (2018) it is possible, but awkward, to look up genes across annotation versions. For that reason, a database to easily convert between gene annotation versions and to access to *P. patens* annotations and sequences was developed and is accessible through PEATmoss.

Table 1 Expression data sets for *Physcomitrella patens* in PEATmoss

Data set name	Experiments	Publication
RNA-seq developmental stages	15	Perroud *et al.* (2018) and new data[†1]
RNA-seq gametophore treatments	4	Perroud *et al.* (2018)
RNA-seq protonema treatments	25	Perroud *et al.* (2018) and new data[†2]
NimbleGen major developmental stages including sexual reproduction	11	Ortiz-Ramirez *et al.* (2016)
NimbleGen Reute development and mycorrhiza	9	Hiss *et al.* (2017)[†3], and Hanke *et al.*, in preparation[†4]
CombiMatrix major developmental stages	9	Hiss *et al.* (2014)
CombiMatrix gametophore treatments	18	Hiss *et al.* (2014), Beike *et al.* (2015)[†3], and Possart *et al.* (2017)[†3]
CombiMatrix detached leaflet development	11	Busch *et al.* (2013) and Hiss *et al.* (2014)
CombiMatrix protonema treatments	7	Hiss *et al.* (2014) and Arif *et al.*, (2019)[†3]

[†1]Includes spores for Gransden ecotype and protonema, juvenile and adult gametophores for Reute ecotype.
[†2]Time series in control conditions and phosphate deficiency.
[†3]Published data but not previously included in any other expression atlas.
[†4]Mycorrhizal fungi interaction experiment including *Rhizophagus irregularis* and *Gigaspora margerita* exudates.

RESULTS

Data sets available in PEATmoss

PEATmoss unifies in one single tool *P. patens* expression data from CombiMatrix and NimbleGen expression microarrays and RNA-seq, containing 109 experiments organized into nine data sets (Table 1). PEATmoss includes 35 experiments (replicate sets) in addition to the 74 experiments also available in the eFP Browser, Genevestigator, and Phytozome.

The data sets contain experiments from the ecotypes Gransden and Reute, and 17 different tissues including protoplasts and most of the *P. patens* life cycle developmental stages (Figure 1). Among these are dry spores, imbibed spores, germinating spores, protonema, caulonema, chloronema, juvenile gametophores, rhizoids, leaflets (phyllids), adult gametophores (including sexual organs), archegonia (female reproductive organs) and sporophyte development stages (Figure 1).

In PEATmoss, many experimental conditions can be found, such as hormone treatments (abscisic acid (ABA), auxin, GA9, strigolactone, OPDA), light perturbations (darkness, low light, high light, continuous light, UV-B light, blue light, red light, far red light, sunlight), abiotic stresses (PO_4 deficiency, ammonium, cold stress, heat stress, dehydration, rehydration, pH 4.5) and biotic stresses or biotic interactions (*Botrytis cinerea*, *Rhizophagus irregularis* exudate, *Gigaspora margerita* exudate).

Raw data from other expression experiments for *P. patens* are available in public repositories such as the Sequence Read Archive (SRA) (Leinonen *et al.*, 2011), but it is not the present goal of PEATmoss to include all of these. Experiments for multiple treatments, developmental stages, and media were selected, knock-out experiments were not considered. In the future more experiments including additional treatments or tissues will be included.

Expression data included in PEATmoss and not available in other tools

PEATmoss includes 35 expression experiments not available in any other expression atlas. RNA-seq experiments for developmental stages from Gransden and Reute ecotypes (Table 1, note [†1]) and for a phosphate deficiency time series were made available with PEATmoss (Table 1, note [†2]). For expression microarrays, gene expression data are included for mycorrhiza experiments (Table 1, note [†4]) as well as

published data not available in any other expression tool (Table 1, note [†3]). These data are from the early response to ABA (Arif *et al.*, 2019), response to red light experiments (Possart *et al.*, 2017), response to cold stress (Beike *et al.*, 2015) and from Reute ecotype gametophore and sporophyte developmental stages (Hiss *et al.*, 2017).

Developmental stage experiments

Five experiments not available in other expression tools were included in the RNA-seq developmental stages data set (Table 1, note [†1]). Among these there are dry spores (two replicates) and imbibed spores (one replicate) from Gransden ecotype, and protonema in Knop liquid, juvenile gametophores and adult gametophores on Knop solid from Reute ecotype (three replicates each). The Reute ecotype has recently been introduced as a tool to study sexual reproduction, since many laboratories report fertility problems of their Gransden strains (Hiss *et al.*, 2017; Perroud *et al.*, 2019).

ABA early response experiment

A time series experiment to study early molecular response to the phytohormone ABA at 30, 60 and 180 min is included in PEATmoss (Arif *et al.*, 2019). These experiments using the CombiMatrix expression microarray were generated to study cell wall thickening and related morphological changes in response to ABA (Arif *et al.*, 2019). Of note, at 180 min the developmental decision to form brachycytes or brood cells (vegetative diaspores) has already been made. The ABA differentially expressed genes (DEGs) show a high level of overlap with those differentially expressed upon stresses such as UV-B, drought and cold (Arif *et al.*, 2019).

Arbuscular mycorrhizal fungi (AMF) interaction experiments

Experiments for the AMF *Gigaspora margerita* and *Rhizophagus irregularis* (Hanke *et al.*, in preparation) are available exclusively in PEATmoss (Table 1, note [†4]). The co-evolution between AMF and bryophytes was probably instrumental in the transition of plants to land (Rensing, 2018), the symbiosis probably evolved in the last common ancestor of land plants. Although *P. patens* is not known to mutualistically interact with these fungi, it possesses the conserved signalling cascade (Delaux *et al.*, 2015), which prompted to analyze the molecular response to fungal exudates. For this aim, experiments of a control exudate without AMF, with *R. irregularis* exudate after 1 and 24 h, and with *G. margerita* exudate after 24 h were generated using NimbleGen expression microarrays (Hanke *et al.*, in preparation).

The Expression Atlas Web Tool

PEATmoss not only unifies the expression data from *P. patens* available in several other tools, but also provides the user with multiple visualization methods and other features that facilitate finding expression patterns in the data. PEATmoss is based on the Tomato Expression Atlas (Fernandez-Pozo *et al.*, 2017), hence similar tools and features can be used. Three different input types are possible in PEATmoss: using a gene ID, a BLAST search (Altschul *et al.*, 1997) or a list of genes. Additionally, for better visualization, experiments from the data sets can be filtered by the user to display only experiments for the selected treatments, ecotypes, media, stages or tissues. The expression colour scale can be changed to get higher resolution in specific ranges of expression, for example to better observe differences for genes in similar ranges of expression or for very low or very highly expressed genes where the default colour scale does not provide enough resolution. PEATmoss offers six visualization methods to explore expression data that are detailed below. Furthermore, PEATmoss is able to automatically convert the input gene name to the gene version needed to query any data set by connecting in the background to the *Physcomitrella patens* gene model lookup database (PpGML DB, see below).

PEATmoss expression cube

The main output view is a 3D cube where the expression of multiple genes can be compared simultaneously (Figure 2). This is an advantage over other expression visualization tools, since the commercial version of Genevestigator is the only one of the tools mentioned before where multiple gene comparison is possible and several visualization methods are available. The expression cube shows the expression for the query gene and co-expressed genes when searching by gene ID. In the case of using BLAST or the custom list, a selection of the best hits from BLAST or the list of genes provided in the custom list will be displayed as the layers of the cube, respectively.

PEATmoss bar plots

Clicking on a layer of the cube will open a bar plot with the expression values of the gene that the layer refers to (Figure 3). Several bar plots from multiple genes can be opened simultaneously to facilitate gene expression comparison. Bar plots contain standard error (SE) bars to show replicate variability from each experiment. Moreover, a transpose button allows the user to transpose the bar plots to change treatments and stages from the x-axis to coloured categories and *vice versa*. Gene names in the bar plots are linked to multiple annotations via the PpGML DB (see below).

PEATmoss expression data downloading

The button 'Download Expression Data' on the top-left of the expression cube (Figure 2) will create a tabular text file with the expression data from all the genes and experiments in the cube, including all co-expressed genes from all pages and correlation values and functional descriptions. This format can be easily imported in a spreadsheet

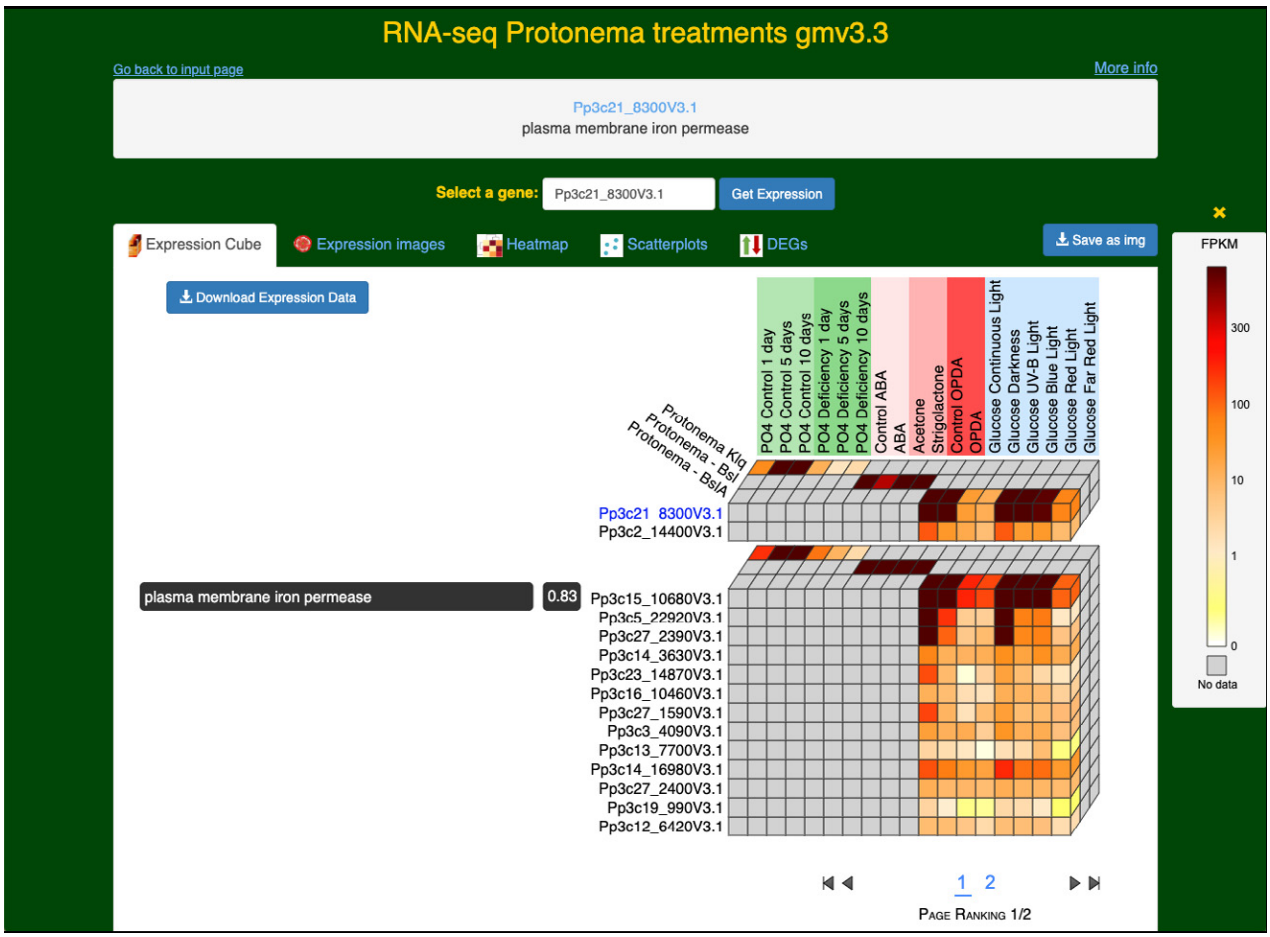

Figure 2. PEATmoss Expression Cube. On top of the cube the experiment treatments are displayed. On the diagonal top-left of the cube, experiment stages/tissues and media are displayed. On the left of the cube gene names are displayed. On top of these, in blue, is the query gene and below are all the co-expressed genes sorted by correlation value. The layers of the cube can be split and merged by clicking on gene names. The gene description and correlation value are displayed when moving the cursor over the gene name. At the bottom of the cube the pagination menu allows access to more co-expressed genes.

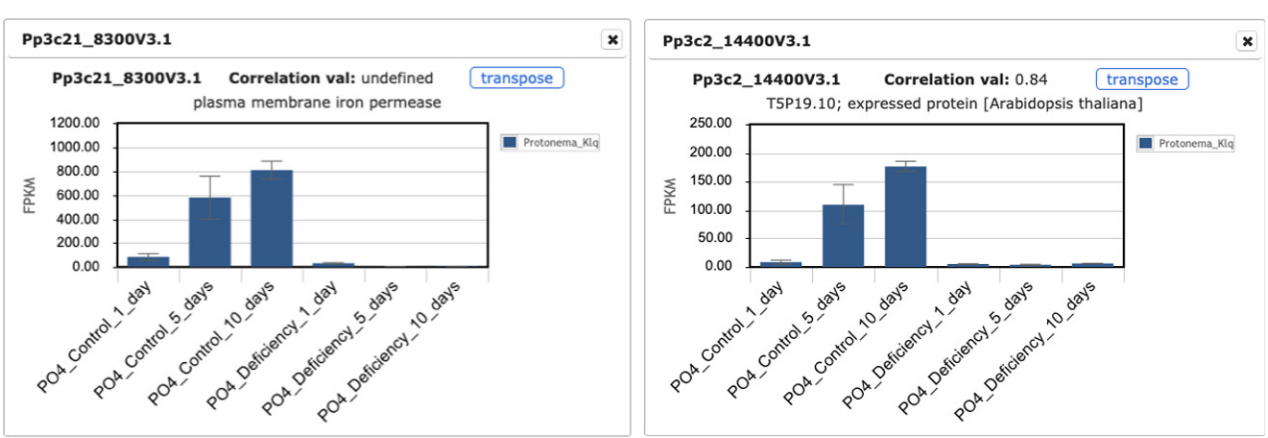

Figure 3. PEATmoss bar plots. Gene expression of gene Pp3c21_8300V3.1 on the left and the most correlated gene (0.84), with similar expression profile, Pp3c2_14400V3.1, on the right.

to visualize genes as rows and expression values from each experiment as columns and is easy to parse for bioinformatic analyses.

PEATmoss expression images

A fourth option to visualize gene expression is the expression images (Figure 4). As in the main visualization from the eFP Browser, each tissue from the data set is represented by a drawing and coloured based on the expression value of the query gene in that tissue. Expression images were designed for PEATmoss, to represent all included tissues and developmental stages of the *P. patens* life cycle (Figure 1).

The expression image system is very flexible when adding additional data sets, since the figures to represent expression data are reusable and independent for each tissue. Figures can represent the expression of one single tissue or can be formed by combination of several drawings to display the expression of multiple tissues in one single image. This system makes the addition of additional data straight forward.

PEATmoss heatmap

The hierarchical clustering heatmap will cluster all experiments of the genes from the first page of the cube, that is the query gene and the 14 co-expressed genes with highest correlation value (Figure 5). When accessing more pages from expression cube pagination menu, the next sets of 14 genes together with the query gene will be displayed. This output provides another way to visualize expression data showing patterns by clustering the most similar genes and experiments. Moving the cursor over the heatmap will show gene, experiment and expression value for each rectangle. Genes and/or experiments can be highlighted for a better visualization, and any region can be selected to zoom in to see it in more detail.

PEATmoss scatterplots

The 'scatterplots' feature enables users to visualize and compare the expression of all genes in the database for any two selected experiments (squares in blue in Figure 6a). The scatterplot's x-axis and y-axis represent the gene expression levels in each of the two experiments, and each gene is plotted as a point in relation to these axes (Figure 6b). When plotting many genes, the density of points can make it difficult to clearly distinguish individual points. Here, the user has the ability to zoom in on a portion of the plot by clicking and dragging the mouse to highlight an area for closer examination (Figure 6c). Resetting the zoom is achieved by a single click anywhere on the plot. By hovering the mouse over an individual point, the user can see the gene ID and the exact expression value in each experiment (Figure 6d). This allows the user to readily identify, for example, genes that have high

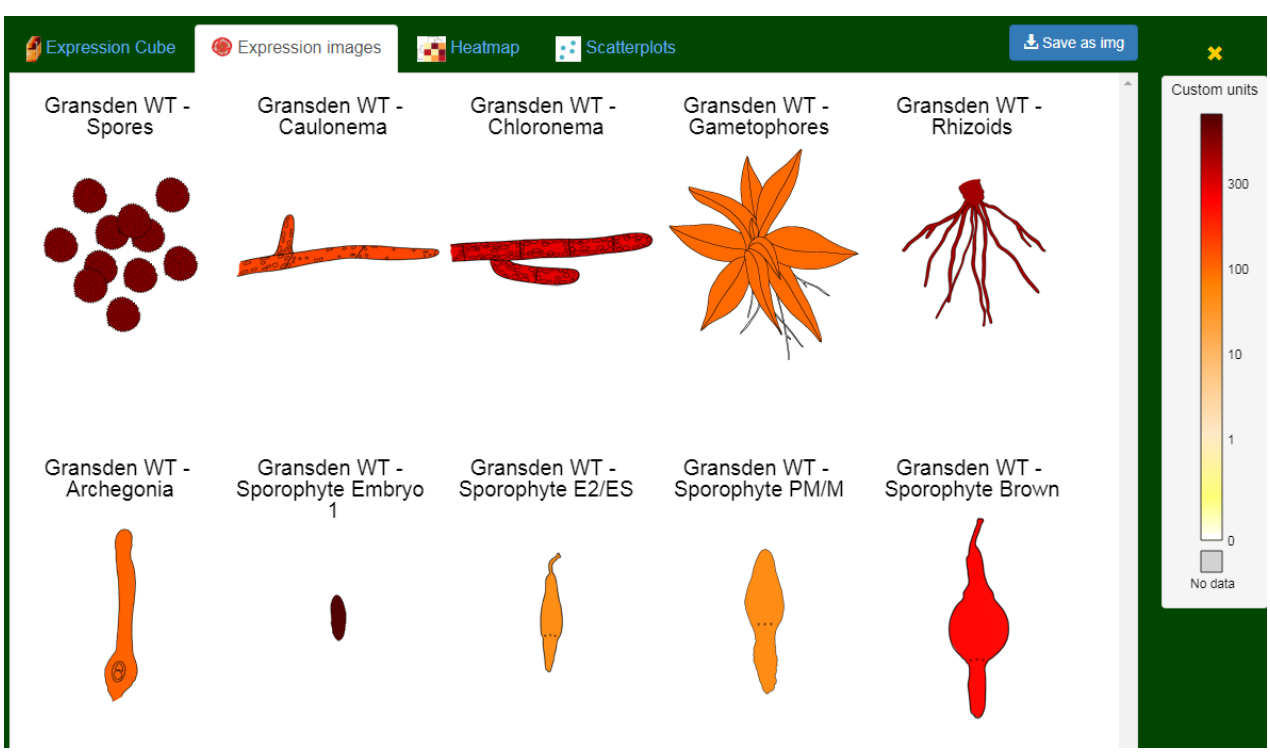

Figure 4. PEATmoss Expression images. Example from the data set NimbleGen Ortiz-Ramirez *et al.* (2016) showing different expression values represented by different colours in 10 tissues/developmental stages.

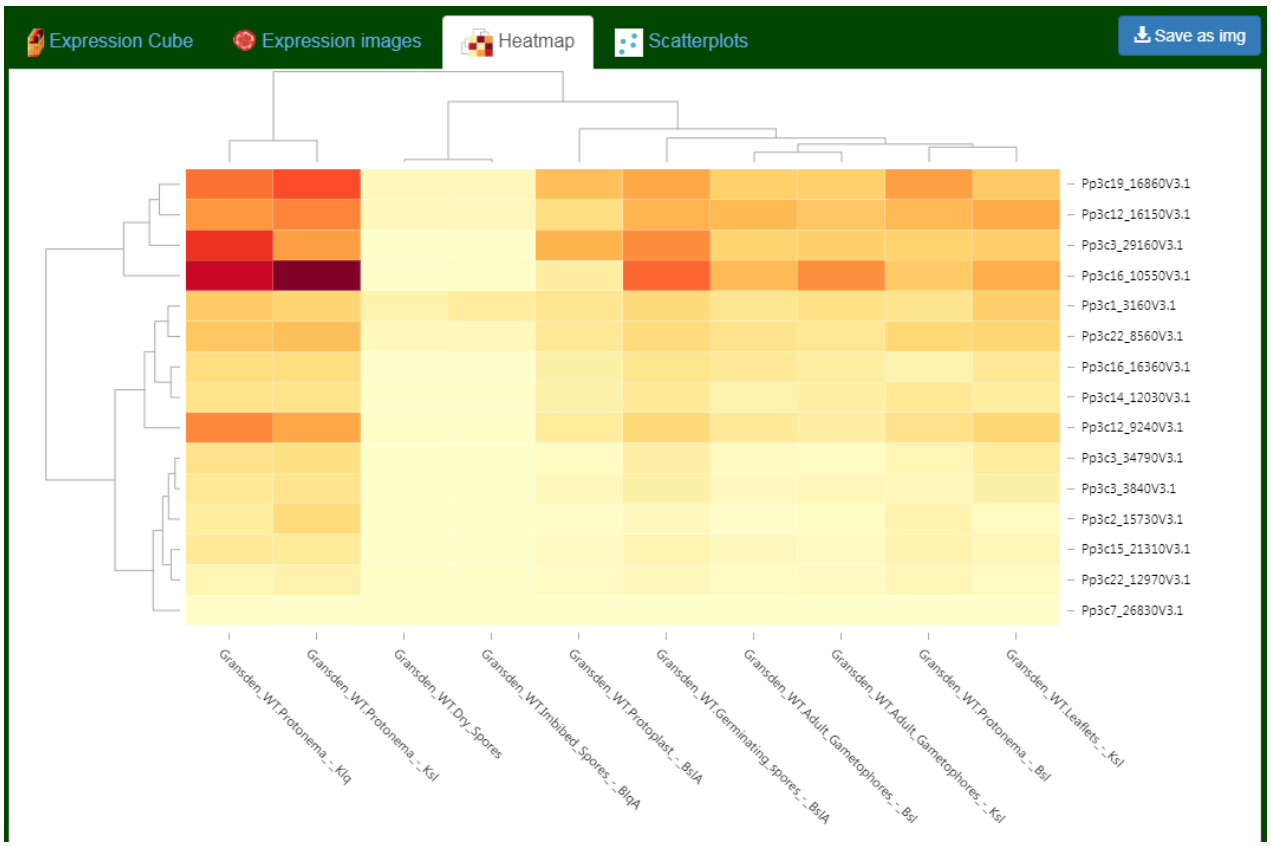

Figure 5. PEATmoss heatmap clustering. The query gene and the top 14 co-expressed genes are displayed on the right. Experiment names are listed at the bottom and hierarchical clustering trees are shown at the top, for experiments, and on the left for genes.

expression in one experiment and low expression in the other. Therefore, users can visually explore the relative expression of genes in any pair of experiments.

PEATmoss DEGs

Using the PEATmoss DEGs tab it is possible to select two experiments from a data set to calculate differentially expressed genes between them. To perform this statistical test, the R package NOISeq (Tarazona *et al.*, 2015) is used with a probability threshold of $q > 0.9$ and biological replicate normalized expression values; FPKM or microarray normalized expression data (stored in PEATmoss), as in Perroud *et al.* (2018), in which NOISeq best represented the overlap of DEGs from EdgeR (Robinson *et al.*, 2010), DESeq2 (Love *et al.*, 2014) and NOISeq. If DEGs are calculated, results are stored in PEATmoss to make them available for future queries, thereby it is possible to display results in a fast way for all comparisons previously calculated.

P. patens gene model lookup database (PpGML DB)

There are six microarray data sets available in PEATmoss, and some of them are also available in the eFP browser and Genevestigator. However, these data used probes based on gene models and gene names from previous genome annotation versions. The data from NimbleGen microarray technology are based on the annotation version 1.6 and the data from CombiMatrix microarray technology are based on version 1.2. When gene versions are different between expression data sets, it is hard to find information and make comparisons, for example, if we are interested in expression values from microarray platforms for a query gene from the current genome version and annotation, or vice versa. For that reason, PEATmoss is connected to the PpGML DB to automatically lookup genes from different versions. Any gene from any gene version can be used to query any of the PEATmoss data sets. The tool will search for the equivalent gene version needed for the selected data set in the PpGML DB and will display the results for the correct gene version. In case of multiple matches, a list of all the matched genes is displayed.

This database also enables the search for genes or annotations, conversion of gene IDs to multiple versions, and retrieval of annotation and sequences for a custom list of genes. Each gene contains descriptions and links for best BLAST hits in NCBI Nr (https://blast.ncbi.nlm.nih.gov/Blast.cgi), SwissProt (UniProt Consortium, 2018) and TAIR

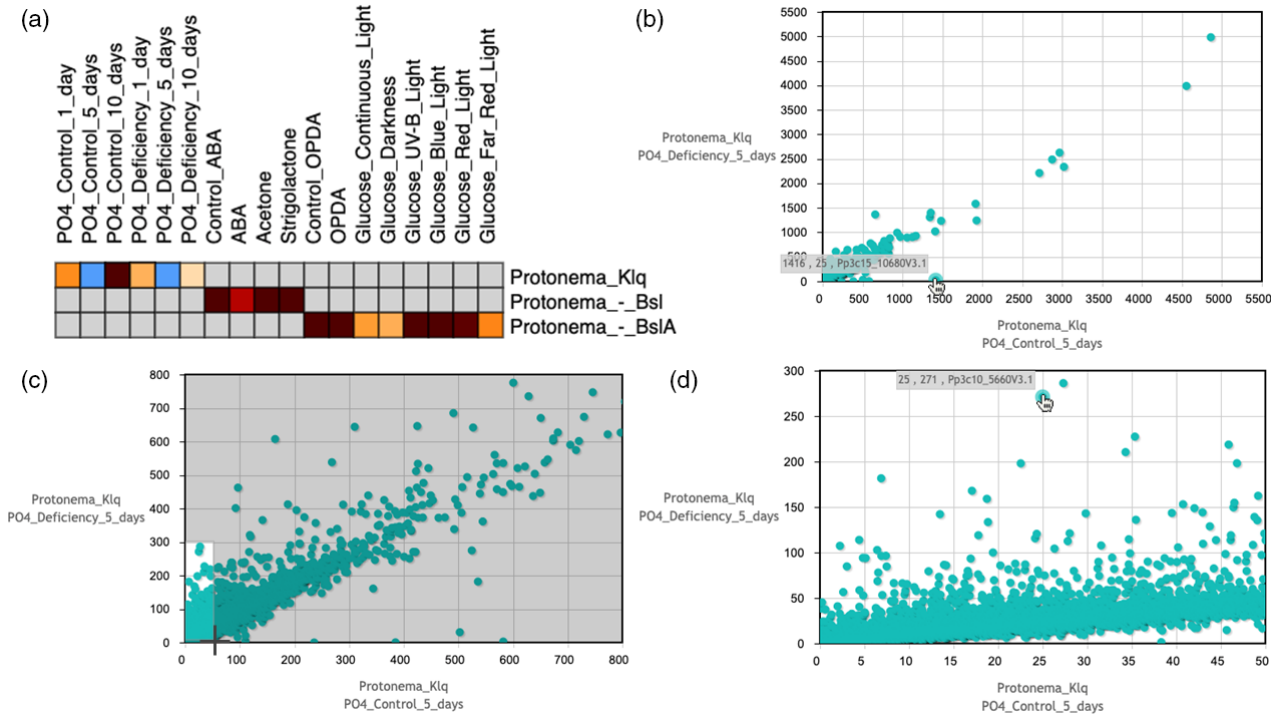

Figure 6. PEATmoss scatterplots. (a) Scatterplot experiment selection. (b) Scatterplot with gene expression values of all genes from two selected experiments. (c) An area of the scatterplot highlighted in order to zoom in. (d) Zoomed in view of area near the y-axis, showing expression values and name for one gene with high expression in one experiment and low expression in the other.

(Berardini *et al.*, 2015). Genes are also linked to their information in Phytozome (Goodstein *et al.*, 2012), PLAZA (Van Bel *et al.*, 2018), Ensembl Plants (Kersey *et al.*, 2016), CoGe (Lyons and Freeling, 2008), TAPscan (Wilhelmsson *et al.*, 2017), Phytozome and Ensembl genome browsers and the expression data in PEATmoss. In addition, many *P. patens* resources from multiple versions such as genome, protein, CDS and transcript sequences and GFF files are available for downloading at the PpGML DB. A BLAST tool is implemented linking the output to the gene annotation pages and with the possibility of downloading the results in BLAST tabular format.

PEATmoss customization

As mentioned before, PEATmoss is based on the Tomato Expression Atlas (TEA), so changes were made in the TEA source code, and they are applied in TEA, PEATmoss and all the tools based on TEA that regularly update the code from GitHub. For example, the application to calculate DEGs was developed for PEATmoss and it is available for TEA just updating the code. PEATmoss was customized to adapt TEA code and style to the needs of the *P. patens* data. The website contains detailed descriptions of the data sets with links to their publications and help pages to learn how to use the tool, including a set of videos about PEATmoss and the PpGML DB (https://peatmoss.online.uni-marburg.de/help).

Phosphate deficiency time series – an example data analysis based on PEATmoss

Phosphorus is a limiting nutrient for plants because of its low availability and mobility in soils (Abel *et al.*, 2002). Therefore, phosphate (PO_4) deficiency is a common abiotic stress in plants which contributes to reduced crop productivity (Schachtman *et al.*, 1998; Lynch and Brown, 2008) and alters global transcriptome profiles in vascular plants (Misson *et al.*, 2005; Zheng *et al.*, 2009; Takehisa *et al.*, 2013). Inorganic phosphate (Pi) deficiency has been intensively studied in seed plants unravelling genes that contribute to the Pi deficiency signalling cascade (Schachtman and Shin, 2007; Rubio *et al.*, 2009; Lin *et al.*, 2011). Here, we investigate Pi deficiency of filamentous, tip-growing protonemata across three time points [1, 5 and 10 days post transfer (dpt) into Pi-deficient medium] and analyze their global molecular responses with the PEATmoss software tool (Table 1, note [†2]).

First, DEGs were calculated in PEATmoss comparing Pi deficiency condition versus the corresponding control for the three time points. Subsequently, DEGs found were queried in the 'Find gene versions and annotations for a list of genes' function of the PpGML DB, selecting the check boxes for gene annotation version 3.3 and 'show annotations' (resulting in descriptions to the closest orthologues in SwissProt, Phytozome, NCBI Nr and TAIR being displayed and downloaded in one click). To facilitate

discussions about gene function, annotations were added to the DEGs (Table S1).

The gene Pp3c21_8300V3.1, found as a DEG downregulated in all time points and annotated as a 'plasma membrane iron permease', was queried in PEATmoss (data set RNA-seq protonema treatments). In PEATmoss Cube (Figure 2), this gene is shown on the top, in blue. Below it, the top 14 co-expressed genes with most similar expression profiles are displayed. In total, 25 genes were found to be correlated to the query gene. There are 14 genes with correlation values over 0.75, of those 10 genes were DEGs in 5 and 10 days of Pi deficiency, demonstrating the usefulness of the co-expression data. Among the correlated genes (with similar expression profile to the query gene) many annotations related to the effects of Pi starvation were found. Some examples are a plasma membrane iron permease, a chlorophyll *a*–*b* binding protein of LHCII, a ferric reductase, and a glutaredoxin s17, which are related to functions altered in phosphate deficiency such as iron homeostasis, photosynthesis, redox balance and reactive oxygen species (ROS) (Hirsch *et al.*, 2006; Bournier *et al.*, 2013; Hernandez and Munne-Bosch, 2015; Carstensen *et al.*, 2018).

The top ranked gene Pp3c2_14400V3.1 (Figure 2), with a correlation value of 0.84, shows a very similar expression profile to the query gene (Figure 3). This gene is clearly downregulated under Pi deficiency conditions but there are no annotations available and its function is unknown, even though it shows similarities in many other species in NCBI nr database. This gene could be a good candidate for further studies to understand the function of genes related to Pi starvation and is a good example of how PEATmoss can assist researchers to visualize expression data for finding genes of interest when no annotations are available.

Looking at the 5 days Pi time point using the PEATmoss scatterplot function (Figure 6a), it is easy to spot some genes with high expression differences between treatment and control (Figure 6b). For example, with 1,416 FPKM in the control and 25 FPKM at 5 dpt, the gene Pp3c15_10680V3.1 (Figure 6b) is found, annotated as a plasma membrane iron permease, previously found as correlated with Pp3c21_8300V3.1 (Figure 2) and identified as a downregulated DEG at 5 dpt (Table S1). To find genes with high expression after 5 days of Pi deficiency and low expression in control conditions, the scatterplot was zoomed in to 50×300 (Figure 6c, d). Some examples found are Pp3c10_5660V3.1 (Figure 6d) and Pp3c12_20160V3.1, annotated as a LEA protein and an ABC transporter in the PpGML DB, and both identified as upregulated DEGs at 5 dpt (Table S1).

Looking more closely into the DEGs, Pi deficiency for one day (1 dpt) is apparently causing only minor stress to the plant, which differentially changes the expression of a few genes (11 were over-expressed and eight under-expressed: Table S1). Four DEGs at 1 dpt are annotated as transporters. One of them is upregulated (heavy-metal transporter, Pp3c25_11360V3.1) and three downregulated (iron transporter, Pp3c21_8300V3.1; inorganic phosphate transporter, Pp3c6_26510V3.1 and ABC transporter, Pp3c14_2470V3.1). These alterations in transporter transcription might prepare the plant for altered ion flux and to keep the homeostasis of the cells after the disequilibrium caused by the Pi deficiency. The upregulation of an alpha-amylase (Pp3c13_20160V3.1) might be linked to an intrinsic metabolic change to mobilize carbon from the starch reservoir to reduce the needs of carbon produced by photosynthesis. Gene Ontology (GO) bias analyses were performed to support the DEG results found in PEATmoss. This analyses showed 'amylase activity' as significant Molecular Function (MF) term, whereas the downregulation of the transporters led to a significant GO term enrichment in the GO domain MF for 'transmembrane transporter activity' (Table S2).

At 5 dpt the highest number of DEGs were found (337 up; 34 down; Table S1). The highly significant enriched GO terms (significance level < 0.0001) in the GO domain Biological Process (BP) indicate a severe stress ('response to stress', 'response to external stimulus'; Table S3). Multiple DEGs annotated as calmodulin and calcium binding proteins might be related to signalling as a response to the Pi deficiency stress, as observed in other abiotic stresses (Zeng *et al.*, 2015; Wang *et al.*, 2018). This signal could be associated with a mobilization of Pi from the vacuole (Liu *et al.*, 2011; Chien *et al.*, 2018; Xu *et al.*, 2019), and is supported by five DEGs associated with the vacuole (Table S1) and the upregulated cellular component (CC) GO term 'endomembrane system' (Table S3). The abundance of the GO term 'ion transport' for the up- and downregulated DEGs shows the plant adaptation to the altered ion status. One member of the PHO1-like proteins (Pp3c3_10280V3.1) was found among the upregulated genes, and has been shown earlier to be important for the adaptation to Pi deficiency (Wang *et al.*, 2008). Pp3c7_22280V3.1 (ADP-Glc pyrophosphorylase), an orthologue of *A. thaliana* APS1 reported to be important for the plant under Pi deficiency conditions (Wang *et al.*, 2008), might contribute to starch synthesis. Many transcription factors (TF) from the AP2 and RING finger families are among the DEGs at 5 dpt and might play a role in ubiquitination of proteins for degradation (Mizoi *et al.*, 2012; Rodriguez-Celma *et al.*, 2019). Many of these TFs contain GO terms associated with 'cell death' and 'immune response'. Many chaperones and LEA proteins are also activated (Table S1), which could be involved in the degradation and stability of proteins. Furthermore, many DEGs and GO terms are related to phospholipids and cell wall, which might indicate the degradation of multiple components of the cell to recycle phosphate and carbon (Liao *et al.*, 2011; Nakamura, 2013; Pant *et al.*, 2015;

Zhang *et al.*, 2016). The significantly enriched GO terms for CC indicate that the ongoing adaptation at 5 dpt takes primarily place at the plasma membrane, the endomembrane and the Golgi apparatus (Table S3). Four ethylene responsive TF DEGs and GO terms associated with jasmonic acid (JA) and salicylic acid (SA) response were found to be upregulated at 5dpt. These plant hormones are known to be associated with phosphate deficiency and transport (Wang *et al.*, 2014; Song and Liu, 2015; Khan *et al.*, 2016).

After 10 days of Pi deficiency, the plant seems to be more adapted to the stress and the number of DEGs is reduced to 144 (70 up; 74 down; Table S1). Among the three time points, 5 dpt and 10 dpt share the highest number of upregulated and downregulated genes (Figure S1), which might indicate the slowdown of the adaptation process to the constant external stimulus of Pi deficiency. The same is true for the comparison of shared significantly enriched GO terms. As for dpt 5, at dpt 10 the enriched GO terms for BP indicate a response to external stimulus, whereas the terms 'phosphate' and 'starvation' now appear as individual categories. These findings are in accordance with previous results (Wang *et al.*, 2008) that some genes can respond more slowly to Pi starvation conditions (See Supporting results 1). As compared with 5 dpt, at 10 dpt the GO CC terms show a shift towards the chloroplasts as the main actors for DEGs (Table S4). The biggest changes found for 10 dpt were an alteration in the photosynthesis and electron transport chain in the chloroplast (Tables S1 and S4), also observed in previous studies in other plants (Hernandez and Munne-Bosch, 2015; Zhang *et al.*, 2016; Carstensen *et al.*, 2018). The chloroplast as the predominant organelle for DEGs at 10 dpt is also reflected by the GO domain MF which highlights photosynthesis related categories such as 'electron carrier activity', 'xanthophyll binding' and 'chlorophyll binding'. Many DEGs contain annotations related to the photosynthesis and the electron transport chain such as cytochrome b6, photosystems I and II proteins, chlorophyll A/B binding protein, ribulose-bisphosphate carboxylase, ribulose-phosphate 3-epimerase, photosynthetic NDH subcomplex B3, ascorbate ferrireductase, oxidase and peroxidase, ferredoxin (2Fe-2S), ferroxidase and ferritin (Table S1).

Iron and phosphate homeostasis are tightly connected (Hirsch *et al.*, 2006). Iron plays an important role in many of the processes that appeared in the DEGs and GO results, such as photosynthesis, respiration, redox balance and ROS production (Bournier *et al.*, 2013). In consequence, DEGs and GO terms related to keep the redox state of molecules and prevent damage from ROS were found (10 dpt significant GO term 'response to hydrogen peroxide'; Table S4). At all time points, downregulated iron transporters (Table S1) and GO terms related to iron transport are found (Tables S2 and S4), especially at 5 and 10 dpt. To keep iron homeostasis and probably avoid the production of ROS that could be produced by an unbalanced accumulation of iron in the cell, it seems that iron transporter gene transcription is downregulated to cope with the decreasing internal Pi pool.

In summary, PEATmoss tools and the PpGML DB can be employed to find meaningful biological results. These results are supported by previous publications and can be very useful to explore expression data and support analyses, using the many available applications, such as downloading data, adding annotations, converting gene model versions or finding candidates of interest.

EXPERIMENTAL PROCEDURES

Data sets

Expression data from the three RNA-seq data sets available in PEATmoss were published in Perroud *et al.* (2018). DNA library preparation, cDNA sequencing and RNA-seq analysis for the not previously published RNA-seq experiments, the phosphate deficiency time series and the developmental stages, were carried out as in Perroud *et al.* (2018). The CombiMatrix expression data sets were published in Busch *et al.* (2013); Hiss *et al.* (2014); Beike *et al.* (2015); Possart *et al.* (2017) and Arif *et al.* (2019). The NimbleGen expression data sets were published in Hiss *et al.* (2017) and Ortiz-Ramirez *et al.* (2016). NimbleGen microarray data for the arbuscular mycorrhiza (AM) interactions (Hanke *et al.*, in preparation) were processed as described in the next paragraphs.

AM interaction experiments

RNA was isolated from plant material using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. RNA concentration and size distribution was tested on the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA) with the Agilent RNA 6000 Nano Kit to determine quantity and quality. About 200 ng of total RNA was reverse transcribed and amplified with the WTA Kit (Sigma-Aldrich, St. Louis, CO, USA). Here, 1 g cDNA was labelled with Cy3 according to the NimbleGen One-Colour DNA Labelling Kit (Roche, Basel, Switzerland). Concentration and quality of the labelled cDNA was monitored. 4 μg of labelled cDNA were used for the hybridization on a NimbleGen 12 × 135 k DNA microarray, probe design OID33087 (Roche, Basel, Switzerland) according to the manufacturer's protocol using the NimbleGen Hybridization Kit (Roche, Basel, Switzerland). The NimbleGen Wash Buffer Kit (Roche) was used to prepare the slide for scanning. The arrays were imaged using a laser scanner Agilent G2565CA Microarray Scanner System (Agilent Technologies, Santa Clara, CA, USA). The image of the arrays was cut into single array images using the NimbleScan Software 2.5 (Roche, Basel, Switzerland) and the pixel intensities were extracted with the same software. Microarray expression data were analyzed with Analyst 7.5 (Genedata, Basel, Switzerland). Data were processed as in Hiss *et al.* (2014).

Pi deficiency time series experiments

Protonemata were pre-cultured in liquid medium as described earlier (Perroud *et al.*, 2018) and transferred to either Pi-deficient liquid growth medium (K_2SO_4 was used instead of KH_2PO_4) as described in Wang *et al.* (2008) or cultivated in liquid Knop medium (Knop, 1868). Three replicates of protonemata were harvested 1 day post transfer (1 dpt), 5 days (5 dpt) and 10 days (10 dpt) and RNA was extracted. Correlation matrices of the replicates from the

PO_4 deficiency experiment are shown in Figure S4. RNA extraction, DNA library preparation, cDNA sequencing and RNA-seq processing and normalization were done as in Perroud *et al.* (2018) with DEG calculation performed in PEATmoss.

GO analysis was conducted in Cytoscape v3.5 with the BiNGO plugin (Maere *et al.*, 2005) using gene annotations for *P. patens* v3.3. To find significantly over-represented or under-represented GO categories, hypergeometric statistical tests were applied at a significance level of 0.05 after correcting for multiple testing according to (Benjamini and Hochberg, 1995).

Data processing and normalization

Experiments previously published were not processed and replication was not evaluated. For RNA-seq experiments not published before, they were analyzed as in Perroud *et al.* (2018). Experiment replicates were checked by PCA, hierarchical clustering, and correlation of the replicates. No minimum expression value was used to filter the data, so users can set their own cut-off values. Experiments of expression microarrays not published before were analyzed as in Hiss *et al.* (2014) for CombiMatrix and as in Hiss *et al.* (2017) for NimbleGen. Background was subtracted and replicates were examined using hierarchical clustering and correlation matrices, removing in some cases replicates not clustering properly or with low correlation in comparison with the other replicates from the same experiment. NimbleGen data set from Ortiz-Ramirez *et al.* (2016) were included as in the publication. Published data from expression microarrays were downloaded from ArrayExpress (Kolesnikov *et al.*, 2015).

All experiments from CombiMatrix and NimbleGen were normalized using Genedata Analyst v. 9.5.2 (https://www.genedata.com/products/analyst/) to adjust the distribution to a median value of 12. This is the median value found for the distribution of all RNA-seq experiments from Perroud *et al.* (2018), which adapts to a similar scale range and makes data comparison in the visualization tools easier.

The Expression Atlas

PEATmoss is based on the TEA (Fernandez-Pozo *et al.*, 2017), a tool developed and hosted at the Sol Genomics Network (Fernandez-Pozo *et al.*, 2015). The code was cloned from the Solgenomics account in GitHub (https://github.com/solgenomics/Tea). Perl scripts included with the tool were used to format and import the data to the database and expression indexes. The code of TEA, PEATmoss, and other tools based on TEA is in continuous development. The code used to customize PEATmoss style and page structure is available on GitHub (https://github.com/noefp/ppatens_expr). Functional annotations for the *P. patens* genes displayed in PEATmoss were obtained from Cosmoss (Zimmer *et al.*, 2013), genes with unknown description were then subsequently annotated with SwissProt, TAIR, or the NCBI Nr database following that order. The heatmap output is implemented using the R package d3heatmap. It uses hierarchical clustering complete agglomeration method and euclidean distances.

P. patens gene model lookup DB (PpGML DB)

The GML DB is implemented as a Postgres v10 relational database with a very simple schema (Figure S3) to store gene names, versions and annotations, and the relation between them. The module pg_trgm is used to support non-exact text search based on trigram matching. The website is written in PHP (v7.0.30-0 + deb9u1) and uses Bootstrap 3 libraries (https://getbootstrap.com/) for style and interactive elements. JQuery DataTables (https://datatables.net/) are used to display search outputs and tables with gene annotations. Gene version and annotations stored in the database were extracted from a modified version of the lookup table (Data S1) from Perroud *et al.* (2018). This lookup table was based on the Cosmoss annotation (Zimmer *et al.*, 2013) and was improved using the GMAP version 2015-12-31 (Wu and Nacu, 2010) to map nucleotide transcript sequences against the *P. patens* V3 genome. The intersection with the *P. patens* gene model v3.3 was obtained using bedtools intersect version 2.26.0 (Quinlan and Hall, 2010). BLAST+ (Camacho *et al.*, 2009) and the blastdbcmd script are integrated in the website to search sequences by similarity and to extract list of sequences from BLAST databases. BLAST output interface was modified from the code from the Sol Genomics Network (Fernandez-Pozo *et al.*, 2015) (https://github.com/solgenomics/sgn).

ACCESSION NUMBERS

Raw data for the RNA-seq experiments included in PEATmoss and not available in other tools are stored at the SRA (Leinonen *et al.*, 2011) with the library names BBTWT, BBTWU, BBTWS, BBTWW, BBTWY, BBTWX, BAOBH, BAOBP, BAOBS, BBTXA, BBTXB, BBTWZ, BAOBT, BAOBN, BBTXG, BAOAW, BAOAX, BAOBA, BAOAZ, BAOBC, BAOAU, BAOBG, BAOBU, BAOBO, BAOBB, BAOAY, BAOBW, BAOAT, BBTXC, and BBTWO. ArrayExpress accession numbers for the CombiMatrix data are E-MTAB-2165 (Beike *et al.*, 2015), E-MTAB-913, E-MTAB-914, E-MTAB-915, E-MTAB-916, E-MTAB-917, E-MTAB-919, E-MTAB-976, E-MEXP-2508 (Hiss *et al.*, 2014), and E-MTAB-2227 (Possart *et al.*, 2017). ArrayExpress accession numbers for the NimbleGen data are E-MTAB-3069 (Ortiz-Ramirez *et al.*, 2016) and E-MTAB-4630 (Hiss *et al.*, 2017). NimbleGen microarray data from the arbuscular mycorrhiza (AM) interaction experiments (Hanke *et al.*, in preparation) can be found in ArrayExpress with the accession number E-MTAB-3081.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg, Germany, through bwHPC and the German Research Foundation (DFG) through grant number INST 37/935-1 FUGG. The authors also acknowledge the support of the DFG through grant number DFG RE 1697/6-1. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the US Department of Energy under contract number DE-AC02-05CH11231. The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTION

NFP, KU and SAR wrote the manuscript with help from all authors. NFP implemented and customized the tool, formatted the data and imported the data to index files and the database. FH calculated FPKMs for all RNA-seq experiments and created the lookup table. MH conducted the microarray experiments and assisted with microarray metadata and data processing. RM, KU, SH, JC, and EV conducted RNA-seq experiments. RM drew the plant figures and organized sporophyte development experiments metadata. PFP organized and supervised experiment metadata. NFP and VK developed the PpGML DB. AP and LAM developed the scatterplot application in PEATmoss. NFP and FH set up the PEATmoss server and SAR supervised the whole project.

SUPPORTING INFORMATION

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

Figure S1. Common differentially expressed genes among three Pi-deficient time points.

Figure S2. Protonemata expression values for known phosphate deficiency responsive genes.

Figure S3. PpGML DB schema.

Figure S4. Correlation matrices from the phosphate deficiency experiment.

Table S1. PEATmoss DEG output for each phosphate deficiency time point.

Table S2. Gene ontology analysis for Pi deficiency time point dpt 1.

Table S3. Gene ontology analysis for Pi deficiency time point dpt 5.

Table S4. Gene ontology analysis for Pi deficiency time point dpt 10.

Table S5. Expression of known Pi deficiency candidate genes (FPKM).

Supporting results 1

REFERENCES

Abel, S., Ticconi, C.A. and Delatorre, C.A. (2002) Phosphate sensing in higher plants. *Physiol. Plant*, **115**, 1–8.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

Arif, M.A., Hiss, M., Tomek, M., Busch, H.S., Tintelnot, S., Meyberg, R., Reski, R., Rensing, S.A. and Frank, W. (2019) ABA-induced vegetative diaspore formation in *Physcomitrella* patens. *Front. Plant Sci.* **10**, 315.

Beike, A.K., Lang, D., Zimmer, A.D., Wust, F., Trautmann, D., Wiedemann, G., Beyer, P., Decker, E.L. and Reski, R. (2015) Insights from the cold transcriptome of *Physcomitrella* patens: global specialization pattern of conserved transcriptional regulators and identification of orphan genes involved in cold acclimation. *New Phytol.* **205**, 869–881.

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., Coppens, F. and Vandepoele, K. (2018) PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **46**, D1190–D1196.

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

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The Arabidopsis information resource: making and mining the 'gold standard' annotated reference plant genome. *Genesis*, **53**, 474–485.

Bournier, M., Tissot, N., Mari, S., Boucherez, J., Lacombe, E., Briat, J.F. and Gaymard, F. (2013) Arabidopsis ferritin 1 (AtFer1) gene regulation by the phosphate starvation response 1 (AtPHR1) transcription factor reveals a direct molecular link between iron and phosphate homeostasis. *J. Biol. Chem.* **288**, 22670–22680.

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.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421.

Carstensen, A., Herdean, A., Schmidt, S.B., Sharma, A., Spetea, C., Pribil, M. and Husted, S. (2018) The impacts of phosphorus deficiency on the photosynthetic electron transport chain. *Plant Physiol.* **177**, 271–284.

Chien, P.S., Chiang, C.P., Leong, S.J. and Chiou, T.J. (2018) Sensing and signaling of phosphate starvation - from local to long distance. *Plant Cell Physiol.* **59**, 1714–1722.

Delaux, P.M., Radhakrishnan, G.V., Jayaraman, D. *et al.* (2015) Algal ancestor of land plants was preadapted for symbiosis. *Proc. Natl. Acad. Sci. USA*, **112**, 13390–13395.

Fernandez-Pozo, N., Menda, N., Edwards, J.D. *et al.* (2015) The Sol Genomics Network (SGN)–from genotype to phenotype to breeding. *Nucleic Acids Res.* **43**, D1036–1041.

Fernandez-Pozo, N., Zheng, Y., Snyder, S.I., Nicolas, P., Shinozaki, Y., Fei, Z., Catala, C., Giovannoni, J.J., Rose, J.K.C. and Mueller, L.A. (2017) The Tomato Expression Atlas. *Bioinformatics*, **33**, 2397–2398.

Goodstein, D.M., Shu, S., Howson, R. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–1186.

Hernandez, I. and Munne-Bosch, S. (2015) Linking phosphorus availability with photo-oxidative stress in plants. *J. Exp. Bot.* **66**, 2889–2900.

Hirsch, J., Marin, E., Floriani, M., Chiarenza, S., Richaud, P., Nussaume, L. and Thibaud, M.C. (2006) Phosphate deficiency promotes modification of iron distribution in Arabidopsis plants. *Biochimie*, **88**, 1767–1771.

Hiss, M., Laule, O., Meskauskiene, R.M. *et al.* (2014) Large-scale gene expression profiling data for the model moss *Physcomitrella* patens aid understanding of developmental progression, culture and stress conditions. *Plant J.* **79**, 530–539.

Hiss, M., Meyberg, R., Westermann, J., Haas, F.B., Schneider, L., Schallenberg-Rudinger, M., Ullrich, K.K. and Rensing, S.A. (2017) Sexual reproduction, sporophyte development and molecular variation in the model moss *Physcomitrella* patens: introducing the ecotype Reute. *Plant J.* **90**, 606–620.

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. Bioinfor.* **3**, 1–5.

Kersey, P.J., Allen, J.E., Armean, I. *et al.* (2016) Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.* **44**, D574–580.

Khan, G.A., Vogiatzaki, E., Glauser, G. and Poirier, Y. (2016) Phosphate deficiency induces the Jasmonate Pathway and enhances resistance to insect Herbivory. *Plant Physiol.* **171**, 632–644.

Knop, W. (1868) *Der Kreislauf des Stoffs.* Рипол Классик.

Kolesnikov, N., Hastings, E., Keays, M. *et al.* (2015) ArrayExpress update–simplifying data submissions. *Nucleic Acids Res.* **43**, D1113–1116.

Lang, D., Ullrich, K.K., Murat, F. *et al.* (2018) The *Physcomitrella* patens chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533.

Leinonen, R., Sugawara, H. and Shumway, M. (2011) The sequence read archive. *Nucleic Acids Res.* **39**, D19–21.

Liao, Y.Y., Buckhout, T.J. and Schmidt, W. (2011) Phosphate deficiency-induced cell wall remodeling: linking gene networks with polysaccharide meshworks. *Plant Signal. Behav.* **6**, 700–702.

Lin, W.D., Liao, Y.Y., Yang, T.J., Pan, C.Y., Buckhout, T.J. and Schmidt, W. (2011) Coexpression-based clustering of Arabidopsis root genes predicts functional modules in early phosphate deficiency signaling. *Plant Physiol.* **155**, 1383–1402.

Liu, T.Y., Aung, K., Tseng, C.Y., Chang, T.Y., Chen, Y.S. and Chiou, T.J. (2011) Vacuolar Ca2+/H+ transport activity is required for systemic phosphate homeostasis involving shoot-to-root signaling in Arabidopsis. *Plant Physiol.* **156**, 1176–1189.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Lynch, J.P. and Brown, K.M. (2008) Root strategies for phosphorus acquisition. In *The Ecophysiology of Plant-Phosphorus Interactions* (White, P.J. and Hammond, J.P. eds). Dordrecht: Springer, Netherlands, pp. 83–116.

Lyons, E. and Freeling, M. (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* **53**, 661–673.

Maere, S., Heymans, K. and Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448–3449.

Misson, J., Raghothama, K.G., Jain, A. *et al.* (2005) A genome-wide transcriptional analysis using Arabidopsis thaliana Affymetrix gene chips determined plant responses to phosphate deprivation. *Proc. Natl. Acad. Sci. USA*, **102**, 11934–11939.

Mizoi, J., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2012) AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta.* **1819**, 86–96.

Nakamura, Y. (2013) Phosphate starvation and membrane lipid remodeling in seed plants. *Prog. Lipid Res.* **52**, 43–50.

Ortiz-Ramirez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., Feijo, J.A. and Becker, J.D. (2016) A Transcriptome Atlas of *Physcomitrella* patens Provides Insights into the Evolution and Development of Land Plants. *Mol. Plant*, **9**, 205–220.

Pant, B.D., Burgos, A., Pant, P., Cuadros-Inostroza, A., Willmitzer, L. and Scheible, W.R. (2015) The transcription factor PHR1 regulates lipid remodeling and triacylglycerol accumulation in Arabidopsis thaliana during phosphorus starvation. *J. Exp. Bot.* **66**, 1907–1918.

Perroud, P.F., Haas, F.B., Hiss, M. *et al.* (2018) The *Physcomitrella* patens gene atlas project: large-scale RNA-seq based expression data. *Plant J.* **95**, 168–182.

Perroud, P.F., Meyberg, R. and Rensing, S.A. (2019) *Physcomitrella* patens Reute mCherry as a tool for efficient crossing within and between ecotypes. *Plant Biol. (Stuttg)* **21**(Suppl 1), 143–149.

Possart, A., Xu, T., Paik, I. *et al.* (2017) Characterization of phytochrome interacting factors from the Moss *Physcomitrella* patens illustrates conservation of phytochrome signaling modules in land plants. *Plant Cell*, **29**, 310–330.

Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Rensing, S.A. (2017) Why we need more non-seed plant models. *New Phytol.* **216**, 355–360.

Rensing, S.A. (2018) Great moments in evolution: the conquest of land by plants. *Curr. Opin. Plant Biol.* **42**, 49–54.

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.

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.

Rodriguez-Celma, J., Chou, H., Kobayashi, T., Long, T.A. and Balk, J. (2019) Hemerythrin E3 ubiquitin ligases as negative regulators of iron homeostasis in plants. *Front. Plant Sci.* **10**, 98.

Rubio, V., Bustos, R., Irigoyen, M.L., Cardona-Lopez, X., Rojas-Triana, M. and Paz-Ares, J. (2009) Plant hormones and nutrient signaling. *Plant Mol. Biol.* **69**, 361–373.

Ruprecht, C., Proost, S., Hernandez-Coronado, M., Ortiz-Ramirez, C., Lang, D., Rensing, S.A., Becker, J.D., Vandepoele, K. and Mutwil, M. (2017) Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. *Plant J.* **90**, 447–465.

Schachtman, D.P. and Shin, R. (2007) Nutrient sensing and signaling: NPKS. *Annu. Rev. Plant Biol.* **58**, 47–69.

Schachtman, D.P., Reid, R.J. and Ayling, S.M. (1998) Phosphorus uptake by plants: from soil to cell. *Plant Physiol.* **116**, 447–453.

Song, L. and Liu, D. (2015) Ethylene and plant responses to phosphate deficiency. *Front. Plant Sci.* **6**, 796.

Takehisa, H., Sato, Y., Antonio, B.A. and Nagamura, Y. (2013) Global transcriptome profile of rice root in response to essential macronutrient deficiency. *Plant Signal. Behav.* **8**, e24409.

Tarazona, S., Furio-Tari, P., Turra, D., Pietro, A.D., Nueda, M.J., Ferrer, A. and Conesa, A. (2015) Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* **43**, e140.

UniProt Consortium, T. (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res*, **46**, 2699.

Wang, Y., Secco, D. and Poirier, Y. (2008) Characterization of the PHO1 gene family and the responses to phosphate deficiency of *Physcomitrella* patens. *Plant Physiol.* **146**, 646–656.

Wang, G., Zhang, C., Battle, S. and Lu, H. (2014) The phosphate transporter PHT4;1 is a salicylic acid regulator likely controlled by the circadian clock protein CCA1. *Front. Plant Sci.* **5**, 701.

Wang, X., Hao, L., Zhu, B. and Jiang, Z. (2018) Plant Calcium Signaling in Response to Potassium Deficiency. *Int. J. Mol. Sci.* **19**.

Wilhelmsson, P.K.I., Muhlich, C., Ullrich, K.K. and Rensing, S.A. (2017) Comprehensive Genome-Wide Classification Reveals That Many Plant-Specific Transcription Factors Evolved in Streptophyte Algae. *Genome Biol. Evol.* **9**, 3384–3397.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS ONE*, **2**, e718.

Wu, T.D. and Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.

Xu, L., Zhao, H., Wan, R. *et al.* (2019) Identification of vacuolar phosphate efflux transporters in land plants. *Nat. Plants*, **5**, 84–94.

Zeng, H., Xu, L., Singh, A., Wang, H., Du, L. and Poovaiah, B.W. (2015) Involvement of calmodulin and calmodulin-like proteins in plant responses to abiotic stresses. *Front. Plant Sci.* **6**, 600.

Zhang, K., Liu, H., Song, J., Wu, W., Li, K. and Zhang, J. (2016) Physiological and comparative proteome analyses reveal low-phosphate tolerance and enhanced photosynthesis in a maize mutant owing to reinforced inorganic phosphate recycling. *BMC Plant Biol.* **16**, 129.

Zheng, L., Huang, F., Narsai, R. *et al.* (2009) Physiological and transcriptome analysis of iron and phosphorus interaction in rice seedlings. *Plant Physiol.* **151**, 262–274.

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 Genom.* **14**, 498.