Gene expression analysis of *Cyanophora paradoxa* reveals conserved abiotic stress responses between basal algae and flowering plants

Camilla Ferrari\(^1\) and Marek Mutwil\(^{1,2}\)

\(^{1}\)Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476, Potsdam, Germany; \(^{2}\)School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore

**Summary**

- The glaucophyte *Cyanophora paradoxa* represents the most basal member of the kingdom Archaeplastida, but the function and expression of most of its genes are unknown. This information is needed to uncover how functional gene modules, that is groups of genes performing a given function, evolved in the plant kingdom.
- We have generated a gene expression atlas capturing responses of *Cyanophora* to various abiotic stresses. The data were included in the CoNekT-Plants database, enabling comparative transcriptomic analyses across two algae and six land plants.
- We demonstrate how the database can be used to study gene expression, co-expression networks and gene function in *Cyanophora*, and how conserved transcriptional programs can be identified. We identified gene modules involved in phycobilisome biosynthesis, response to high light and cell division. While we observed no correlation between the number of differentially expressed genes and the impact on growth of *Cyanophora*, we found that the response to stress involves a conserved, kingdom-wide transcriptional reprogramming, which is activated upon most stresses in algae and land plants.
- The *Cyanophora* stress gene expression atlas and the tools found in the https://conekt.plant.nt.tools database thus provide a useful resource to reveal functionally related genes and stress responses in the plant kingdom.

**Introduction**

The glaucophyte algae are a basally diverging group of unicellular taxa with four described genera and an estimated 15 species (Kies & Kremer, 1986). Glaucophytes represent one branch of the Archaeplastida (Plantae) kingdom, where the other branches are represented by the red (Rhodophyta) and green algae and plants (Viridiplantae) (Adl et al., 2012). The common ancestor of these taxa captured the cyanobacterial endosymbiont c. 1.6 billion years ago, giving rise to the chloroplast found in glaucophytes, red and green algae, and land plants (Thomas, 1982; Bhattacharya et al., 2004; Yoon et al., 2004; Blank, 2013). The plastids of red and green algae participated in several other endosymbioses, leading to the appearance of diatoms, dinoflagellates, euglenids and haptophytes (Reyes-Prieto et al., 2007). Glaucophytes, and their representative *Cyanophora paradoxa*, retain traits from the ancestral cyanobacterial endosymbiont, such as phycobilisomes, peptidoglycan (PG), an ancient, primitive RNA interference pathway (Gross et al., 2013), lack of Chlb (Löffelhardt, 2013) and presence of a bacterial-derived UhpC-type hexose-phosphate transporter used to translocate sugars from the plastid to the host cytosol, that are not found in land plants (Price et al., 2012). Therefore, due to these unique traits, *Cyanophora* can provide invaluable insights into the ancestral state of the Archaeplastida host, its photosynthetic organelle and the evolution of the functional gene modules found in the plant kingdom.

Our ability to extract useful knowledge from the *Cyanophora* genome, and to understand how its genes work together to form these traits, relies on our ability to correctly assign biological functions to gene products (Rhee & Mutwil, 2014; Hansen et al., 2018). Whereas the experimental characterization of protein function is resource-intensive, the *in silico* inference of gene function through gene function prediction methods can be fast and readily available. Therefore, with the increasing availability of assembled genomes, gene function prediction is currently one of the most active areas in bioinformatics. Of the many available methods, co-expression analysis is the most successful and reliable way of predicting gene function (Lee et al., 2010; Hansen et al., 2014, 2018; Rhee & Mutwil, 2014; Proost & Mutwil, 2016; Ruprecht et al., 2017b). Co-expression is based on the guilt-by-association principle, which states that genes involved in the same or closely related biological processes tend to have similar expression patterns across organs, developmental stages, and biotic as
well as abiotic perturbations (Usadel et al., 2009). Co-expression analyses are widely used (Radivojac et al., 2013; Rhee & Mutwil, 2014; Jiang et al., 2016), and have been applied to successfully identify genes involved in plant viability (Mutwil et al., 2010), seed germination (Bassel et al., 2011), shade avoidance (Jiménez-Gómez et al., 2010), cyclic electron flow (Takahashi et al., 2009), cell division (Takahashi et al., 2008), drought sensitivity and lateral root development (Lee et al., 2010), and others (Stuart et al., 2003; Yu et al., 2003; Persson et al., 2005; Itkin et al., 2013; Proost & Mutwil, 2016; Sibout et al., 2017).

Classical comparative genomic approaches that are based on gene sequences are very useful, although they present certain shortcomings as genes often operate as functional gene modules (Hartwell et al., 1999). This means that gene products frequently form, for example, enzymatic pathways or protein complexes that usually require multiple genes working together to perform a given task (e.g. photosystem II complex and ribosomes). While genomic analyses can reveal which gene families have appeared and expanded in a given organism, they might not show which of the underlying genes are functionally related. Because plant gene families can be large and functionally divergent (Shiu & Bleecker, 2001; Shiu et al., 2005), sequence-based analyses could result in incorrectly predicted gene functions (Lynch & Katju, 2004). Therefore, a more rewarding approach to study the evolution of new traits needs to integrate the classical genomic approaches with functional gene modules, which can be identified by expression and co-expression analysis (Ruprecht et al., 2017b; Lampugnani et al., 2019).

To establish these enhanced, co-expression-driven comparative analyses for Cyanophora, we first investigated the changes in gene expression as a response to 10 different abiotic stresses. Together with publicly available data that capture changes in gene expression during 13 time points of the diurnal cycle, we provide a comprehensive expression atlas of this basally diverging member of the Archaeplastida kingdom. To make the data easily accessible we have uploaded it to the CoNekT-Plants database, which provides advanced comparative transcriptomic analyses for algae and land plants. We then exemplify how these analyses can be used to study phycobilisome formation, response to high light and cell division. Finally, we perform a meta-analysis of stress responses in the Archaeplastida kingdom and find evidence of a kingdom-wide transcriptional reprogramming mechanism as a response to a wide variety of stresses.

Materials and Methods

Experimental growth conditions and sampling

Cyanophora paradoxa UTEX555 (SAG 29.80, CCMP329) was obtained from Provasoli-Guillard National Center for Marine Algae and Microbiota (NCMA, Bigelow Laboratory for Ocean Sciences). The cultures were grown in C medium (Supporting Information Table S1) at 24°C under continuous light (40 μmol photons m⁻² s⁻¹) and aeration. We have modulated the magnitude of stresses to inhibit but not stop the cell division of the cultures. Growth of the cultures (cells ml⁻¹) was measured with a Z2 Coulter Counter (Beckman Coulter, Brea, CA, USA).

When subjected to nutrient stress (deprivation of nitrogen: 10% of recommended KNO₃, sulphur: 2% of recommended MgSO₄·7H₂O, phosphate: no glycerophosphateNa₂·5H₂O and trace metals: no trace metal solution (TMS), Table S1), the culture was washed twice with new media and the growth rate was monitored daily for 6 d. On the fifth day, part of the culture was harvested for isolation of RNA, while the remaining part was left to monitor the growth rate. The control culture contained unmodified C medium.

When subjected to environmental stress (high light, heat, cold, prolonged darkness), the cultures were grown in C medium without aeration for 7 d and subjected to the stresses for 12 h, except for prolonged darkness, which lasted for 72 h. The cultures treated for cold (4°C for 12 h), heat (37°C for 12 h) and the control (24°C for 12 h) were kept in the dark for the duration of the experiment. The high light treatment (150 μmol photons m⁻² s⁻¹, 24°C for 12 h) and its control culture (40 μmol photons m⁻² s⁻¹, 24°C for 12 h) were performed simultaneously with the other environmental stress samples. After harvesting, the remaining culture was transferred to normal growth conditions (24°C, under continuous light, 40 μmol photons m⁻² s⁻¹, aeration with ordinary air) and growth was monitored for the following 3 d. The reference time 0 was assigned to the moment the stress was first applied.

Determination of starch levels

Cultures with ~10⁶ cells ml⁻¹ were split into two cultures. One culture was incubated in the dark for 72 h, while the control culture was incubated in continuous light at 40 μmol m⁻² s⁻¹. For visualizing starch levels, the cells were stained with 20% Lugol solution and analyzed by microscopy (Fig. 1c). For OD measurements, 20 ml of culture was harvested (300 g for 10 min at 4°C) and the cell pellet was incubated with 200 µl dimethyl sulfoxide at 70°C for 5 min, followed by incubation with 3 ml MetOH at 4°C for 12 h. The cells were collected by centrifugation and resuspended with 200 µl of 70% EtOH, 400 µl 2 M NaOH and 400 µl H₂O and incubated at 25°C for 2.5 h. A solution of 400 µl 2 M HCl, 1 ml 0.5 M sodium acetate, 7 ml H₂O and 200 µl iodine reagent (1% KI + 0.1% I₂) was added immediately before measuring the absorbance at 680 nm (Avidan et al., 2015). Three independent cultures with different cell densities were used for this analysis: (1) 0.5 × 10⁶ cells ml⁻¹, (2) 10⁶ cells ml⁻¹ and (3) 2 × 10⁶ cells ml⁻¹. Three technical replicates were performed.

RNA preparation and sequencing

The RNA was extracted using a Spectrum Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer’s instructions. The integrity of the RNA was assessed using an RNA nano chip on an Agilent Bioanalyzer 2100. The libraries were prepared from total RNA using polyA enrichment and sequenced using Illumina-HiSeq2500/4000 at Beijing Genomics Institute and...
Max Planck-Genome-centre in Cologne. Three RNA isolations were done for each sample.

Analysis of RNAseq data and microarrays data

The reads were trimmed, mapped, counted and TPM (transcripts per million)-normalized using the LSTrAP pipeline (Proost et al., 2017). The genome used for mapping the reads was *C. paradoxa* v.1.0 (Price et al., 2019). More than 81% of the reads mapped to the genome, and on average 88% of the reads mapped to the coding sequences (Table S2). Principal component analysis revealed that one of the three replicates of salt stress did not cluster with the other two. We therefore excluded the outlier sample and sequenced two additional biological replicates.

The publicly available data were processed as follows: raw expression data for *Arabidopsis thaliana*, *Chlamydomonas reinhardii* and *Synechocystis* sp. PCC 6803 (Table S3) were downloaded from ArrayExpress. The raw fastq files were processed using the LSTrAP pipeline and the reads were mapped to *A. thaliana* TAIR10 and *C. reinhardii* v.5.5. Raw counts from the RNA-
sequencing (RNA-seq) experiments were used with the R package DESeq2 (Love & Anders Somin, 2016) to identify differentially expressed genes (DEGs). The various types of microarrays were processed based on the manufacturer. Affymetrix cell files were processed with the R package affy (Gautier et al., 2004). Agilent raw files were processed with the R package limma (Ritchie et al., 2015). DEGs for all microarray experiments were identified by using the R package limma. For further analyses, we only considered genes which showed an adjusted \( P < 0.05 \) and an \( -1 > \log_2 \text{fold} > 1 \) as DEGs (Tables 4–7).

Adding Cyanophora to the CoNekT-Plants database

The normalized TPM matrix containing expression values from diurnal cycle samples (Ferrari et al., 2019) and the expression matrix of the newly generated stress samples (Table S8) were uploaded in CoNekT-Plants together with coding sequences (CDS) (Price et al., 2019), annotation (Mercator4 v.1.0) (Schwacke et al., 2019) and protein domain information (InterProScan 5.32–71.0) (Jones et al., 2014). Gene families were detected with Orthofinder 1.1.8 (Emms & Kelly, 2015) or using Diamond with default settings (Buchfink et al., 2014) and the MCL clustering algorithm (Enright, 2002), according to PLAZA 4.0 settings (Van Bel et al., 2018). For the functional description of the genes, we used Mercator4 v.1.0 with standard settings. To investigate the effect of the various stresses on the main biological processes, we analyzed the expression of genes assigned to the first level MapMan bins obtained from Mercator. By comparing the percentage of second- and third-level bins for Synechocystis, Cyanophora, Chlamydomonas and Arabidopsis, we observed that the quality of functional annotation for Cyanophora and Chlamydomonas is comparable (Fig. S1).

Functional enrichment analysis of clusters

The Heuristic Cluster Chiseling Algorithm (HCCA) clusters of C. paradoxa were downloaded from CoNekT-Plants (Table S9). For each cluster we counted the number of times a specific biological process was represented. This observed distribution of biological processes was compared to a permuted distribution obtained by sampling an equal number of genes from the total pool of C. paradoxa genes for 10 000 permutations. Empirical \( P \)-values obtained were corrected for false-discovery rate (FDR) by the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995).

Gene function analysis of abiotic stress

To investigate how the stress conditions perturbed certain biological processes (MapMan bins), we first analyzed the fraction of DEGs assigned to each bin. We assigned the fractions into strongly affected \((\geq 50\% \text{ of genes in the bin are DEGs})\), mildly affected \((20\% \leq \text{DEGs} < 50\%)\) or weakly affected \((\text{DEGs} < 20\%)\). To visualize if the DEGs found in the bin are preferentially up- or down-regulated we used the formula:

\[
\text{No. of upregulated genes} - \text{no. of downregulated genes}
\]

\[
\text{No. of genes in the bin}
\]

This formula returns 1 when all genes in the bin are upregulated, 0 when there is an equal number of up- and down-regulated genes, and \(-1\) when all genes in the bin are down-regulated.

Transcription factors (TFs) were identified by retrieving A. thaliana and C. reinhardtii TFs from PlantTFDB (Jin et al., 2017). Orthologous genes for C. paradoxa were used for this analysis.

Inferring similarities of stress responses

We used the Jaccard Index (JI) to assess the similarity of stress responses between two experiments. To this end, we first identified the gene families corresponding to the DEGs, which were used to calculate the JI for two compared stresses, X and Y. The observed JI between stresses X and Y is:

\[
\text{JI}_{\text{obs}}(X,Y) = \frac{\text{No.ofDEgenefamiliesinX} \cap \text{no.ofDEgenefamiliesinY}}{\text{No.ofDEgenefamiliesinX} \cup \text{no.ofDEgenefamiliesinY}}
\]

To test whether a given \(\text{JI}_{\text{obs}}(X,Y)\) is larger than expected by chance (i.e. two stresses show similar transcriptomic response), we performed a permutation test where we drew a number of genes equal to the amount of DEGs present in a specific experiment (DEGs_r) and calculated the expected JI values for each experimental pair (\(\text{JI}_{\text{perm}}(X,Y)\)). The procedure was performed 10 000 times, and the empirical \( P \)-values obtained, calculated independently for up- and down-regulated families, were FDR corrected. The Orthofinder gene families used in this analysis were obtained from Ferrari et al. (2019).

Identification of stress-responsive gene families

We counted the number of experiments when a given family contained DEGs, and calculated the empirical \( P \)-value by estimating if this number was larger than the corresponding counts obtained from shuffled gene–DEG assignments. The shuffling was done 10 000 times, and the empirical \( P \)-values were FDR-corrected.

Phylostratigraphic enrichment analysis of DEGs

To test whether there is a correlation between the age (phylostrata) (Domazet-Lošo et al., 2007) of a gene family and the transcriptomic responses to stress, we calculated the fraction of DEGs found in each phylostratum. Every fraction was then tested for significance by estimating whether the observed fraction was larger than fractions of shuffled gene–phylostrata assignments. The shuffling was done 10 000 times, and the empirical \( P \)-values were FDR-corrected. The gene–phylostrata assignment was retrieved from Ferrari et al. (2019).

Data availability statement

The raw sequencing data are available from EBI under accession number E-MTAB-7822.
Results

Generation of *C. paradoxa* expression atlas

To be able to assign gene functions via co-expression analyses we induced broad changes in gene expression in *C. paradoxa* by subjecting the liquid culture to four nutritional, one osmotic, one salt and four environmental stresses. The intensities of the different stresses were modulated to reduce the growth rate without stopping it entirely. To this end, culture growth was monitored by cell counting from the time when the stress was first applied (time 0 h) up to several days after the harvest time for gene expression analyses (Fig. 1a,b,d). The nutritional stress category involves the depletion of macro- or micronutrients to the growth medium (nitrogen, sulfate, phosphate, micronutrient solution), while the salt and osmotic stresses were induced by addition of 75 mM NaCl or 100 mM mannitol, respectively. Harvest time for all nutritional, salt and osmotic stress conditions was on day 4 when a pronounced reduction of growth was observed for most stresses compared to the control (Fig. 1a). The presence of NaCl in the medium had the strongest negative effect on growth, followed by the presence of mannitol, depletion of phosphate and depletion of Trace metal solution (TMS, micronutrient solution). The effect of sulfate depletion was only visible on day 5 (Fig. 1a). For the environmental stresses we subjected the cultures to heat (37°C), cold (4°C) or high light (150 μE) with a harvest time at 12 h for each stress condition (Fig. 1b,d). Cultures grown for 72 h in darkness showed an expected absence of starch granules as compared to the control (Figs 1c, S2). The environmental stresses resulted in less pronounced negative growth effects (Fig. 1b,d), although on days 2.5 and ay 3.5 after the environmental stress was applied, the effects were more clearly visible at the level of growth rate (Fig. 1b,d).

To evaluate how gene expression in *Cyanophora* was affected by the different stresses, RNA was isolated from the harvested cultures and sequenced in triplicate to produce at least 15 million reads per sample (Table S2). Principal component analysis (PCA) revealed an expected clustering of replicates and related stresses (Fig. S3). While the separation of the different treatments was quite distinct for most stresses, we did not observe a clear separation between the high light treatment and its specific control (Fig. S3b), suggesting a milder effect of increasing light intensity to 150 μE.

To identify the significantly differentially expressed genes (DEGs, adjusted \( P < 0.05, |\log_2 \text{fold} > 1\)) we compared the raw counts of the stress conditions to the respective control. The impact of the stress on the transcriptome was highly variable, ranging from 270 DEGs detected after 12 h of high light treatment to 7628 DEGs after 72 h of darkness (Table S5). Surprisingly, we did not observe any significant correlation between the number of DEGs and the magnitude of effect the stress had on the relative growth of the culture (Fig. 1e). Cold and sulfate deprivation (gray triangle and red dot, respectively) showed mild inhibition of growth, but a much higher proportion of DEGs compared to treatments with a severe growth phenotype, such as salt and mannitol treatments (Fig. 1e, turquoise and orange dots, respectively). Consequently, we observed no significant correlation between DEGs and relative growth rate in *Cyanophora* (\( r = 0.278, P = 0.437 \)).

Integration of *Cyanophora* transcriptome into the CoNekT database

Biological networks are characterized by their scale-free topology, which results in few genes being connected (correlated) to many genes, while the majority of genes have only a few connections (Barabási & Bonabeau, 2003). This architecture is presumed to ensure stability in the case of perturbation, as the network topology remains unaffected when certain genes mutate (Barabási & Oltvai, 2004). To test if the *C. paradoxa* data also follow an expected scale-free network, we calculated the Pearson correlation coefficient (PCC) for every gene pair, with a threshold of 0.9 (Usadel et al., 2009), and counted the number of times a given gene is co-expressed with other genes at this threshold (node degree). The points of the resulting power law plot formed a line with a negative slope, showing a negative correlation between node frequency (i.e. number of genes with a certain number of connections) and node degree (i.e. number of connections per gene). Thus, this result confirms the scale-free topology of the network and indicates that our expression data produce a biologically relevant topology (Fig. 2a).

To easily browse the expression profiles of genes under different stress conditions, to analyze the *Cyanophora* co-expression network (Table S10) and functional gene modules, and to compare the gene modules to higher species of the Archaeplastida kingdom, we added our gene expression data, together with publicly available diurnal samples for *Cyanophora* (Ferrari et al., 2019), to the CoNekT-Plants platform (Proost & Mutwil, 2018). CoNekT-Plants is a user-friendly web-tool containing functional and expression data for eight species including a glaucophyte (the newly added *C. paradoxa*), chlorophyte (*C. reinhardtii*), a gymnosperm (*Picea abies*), two monocots (*Oryza sativa, Zea mays*) and three dicots (*Vitis vinifera, A. thaliana* and *Solanium lycopersicum*), and allows comparative genomic and transcriptomic network analysis for these species. CoNekT-Plants allows users to view expression profiles and co-expression networks of their genes of interest, but also performs more sophisticated analyses, such as identification of functional gene modules involved in a biological process of interest, comparison of gene expression across different species, and identification of genes highly expressed in a given stress or organ (Proost & Mutwil, 2018).

To exemplify the applicability of the tool for *Cyanophora*, we explored the expression profile and co-expression network of *CpcK2* (this was done by entering the *Cpa* evm.model.tig00000361.75 gene identifier in the search box), a gene encoding a linker protein of phycobilisomes (https://conekt.sbs.ntu.edu.sg/sequence/view/4475). *C. paradoxa* as a glaucohyte lacks the light-harvesting Chl antenna complex present in green algae and higher plants but, similarly to cyanobacteria and red algae, possesses a phycobilin-based light-harvesting antenna complex known as phycobilisomes (PBS) (Watanabe et al.,...
PBS consist of phycobiliproteins associated with linker proteins, which capture and transfer light energy to photosystems I and II (PSI and PSII) (Chang et al., 2015). Analysis of the expression profile of CpcK2 (Cpa|evm.model.tig00000361.75) (Fig. 2b) showed an oscillatory pattern with high expression during the light period and decreased expression in the dark. Interestingly, we also observed an anticipatory behavior (Whitehead et al., 2009), where expression of the gene decreased towards the end of the day (anticipating the night) and increased towards the end of the night (anticipating the day). This anticipatory behavior is typical for genes under circadian control and exists to optimize the expression and activity of genes during the day/night cycle (Bell-Pedersen et al., 2005).

We explored the co-expression network associated with the CpcK2 (by clicking on the graph icon under the 'Co-expression networks/Neighborhood' panel), which showed genes directly co-expressed (direct neighbors) with the query gene. In the network we identified additional linker protein encoded genes (CpcK1, CpcG1, CpcG2, CpcD, ApcC1, ApcC2) (Watanabe et al., 2012), together with several genes coding for proteins associated

Fig. 2 Co-expression network of Cyanophora paradoxa. (a) Power law plot based on the Cyanophora gene expression data. The x-axis indicates the node degree (number of connections a given gene has), while the y-axis shows the frequency of a given degree. Both axes were log10-transformed. (b) Expression profile of the phycobilisome linker protein CpcK2 (Cpa|evm.model.tig00000361.75). The x-axis shows the zeitgeber time (hours after the light was switched on), bars indicate the average expression at a given time, dots show the maximum and minimum expression of each sample, and y-axis shows the expression, as transcripts per million (TPM). The color of the bars shows the samples taken in light (yellow) and dark (blue). (c) Simplified co-expression network of CpcK2 (turquoise square). Nodes represent genes, edges connect co-expressed genes (green, orange and red correspond to strong, medium and weak co-expression, respectively), while node colors/shapes indicate the gene family and Pfam domains that the genes have in common. Consequently, genes with the same colored shapes belong to the same family and have the same Pfam domains.

2012).
with PSII (psbO, psbW, psbU) (Nickelsen & Rengstl, 2013), PSI (psaK, psaL) (Fromme et al., 2001), cytochrome b6/f complex (petC) (Hasan et al., 2013) and photosynthetic electron transport (petf, Fig. 2c, Fig. S4) (Durán et al., 2004). Because co-expressed genes are functionally related, the other genes in the network are prime candidates for being involved in photosynthesis. Thus, the database allows us to rapidly identify functionally related genes in *Cyanophora*.

**Functional analysis of network clusters**

Functionally related genes form highly connected clusters in the co-expression network, and these clusters can be identified and studied to unravel the functional gene modules of an organism (Mutwil et al., 2010; Rhee & Mutwil, 2014; Aoki et al., 2016). To identify the functional gene modules of *Cyanophora*, we used the HCCA (Mutwil et al., 2010), and obtained 319 clusters of co-expressed genes (Table S9). To reveal their function, we performed an enrichment analysis and identified 114 clusters significantly enriched for at least one biological process (Fig. 3a, Fig. S5). Benjamini–Hochberg-adjusted empirical *P*<0.05, as defined by MarMan bin (Schwacke et al., 2019). All the bins were enriched in at least one cluster, except Polyamine metabolism. Other bins such as Redox homeostasis, Chromatin organization, DNA damage response, Protein degradation, Protein translocation, Nutrient uptake, Multi-process regulation, Carbohydrate metabolism and Nucleotide metabolism were enriched in multiple clusters, suggesting a more complex regulation of these processes involving several gene modules (Fig. S5).

CoNekT-Plants shows the average expression of all genes belonging to a given cluster. As an example, we explored cluster 54 (https://conekt.sbs.ntu.edu.sg/cluster/view/788), which was significantly enriched for the process ‘Photosynthesis’ (Fig. 3a), indicated by a gray rectangle, and observed that the cluster expression is the highest for the sample ‘Environmental stress − High light − 150 µE’. By contrast, ‘Environmental stress − Prolonged darkness − 72 h’ showed the lowest expression in this cluster, supporting the evidence of an association with photosynthesis-related processes (Fig. 3b). This observation suggests that cluster 54 contains genes that are highly responsive to the amount of light perceived by *Cyanophora*.

We further explored the network associated with cluster 54 and by looking at the labels obtained from InterPro annotation, we identified five major classes of genes: oxidoreductases, transporters, photosynthesis machinery-related proteins, peptidases, and DNA/RNA repair and modification proteins (Figs 3c, S6). Additionally, Gene Ontology (GO) enrichment found on the cluster page for cluster 54 (GO enrichment is found under the cluster page for cluster 54 (GO enrichment is found under the expression profile) showed a significant enrichment of terms related to oxidoreduction (‘oxidation-reduction process’ GO:0055114, ‘oxidoreductase activity’ GO:0016491, ‘oxidoreductase activity, acting on single donors with incorporation of molecular oxygen’ GO:0016701). These results suggest that this light responsive cluster is likely to be active during high light stress and is involved in the cellular repair of biological components damaged by high light. Thus, this example underlines the applicability of co-expression analyses on uncovering functional gene modules.

**Comparative co-expression analysis reveals a conserved cell division program**

For another case study illustrating comparative co-expression analyses in *Cyanophora* we selected the cell cycle and its regulation, which have been extensively studied in many model organisms. Notably, the many components involved in DNA replication and chromosome separation are highly conserved across clades and kingdoms of life (Stuart et al., 2003; Yu et al., 2003). However, given the complexity of multicellular organisms, certain aspects of cell division have evolved to accommodate the different morphologies and lifestyles of single and multicellular plants. For example, the coupling of cell division with chloroplast division is found in many unicellular algae but is absent in multicellular plants (Miyagishima et al., 2012). Furthermore, while cell division tends to be regulated by the light/dark cycle in algae, multicellular land plants are unlikely to show this behavior (Ferrari et al., 2019).

To study the conservation of cell division in *Cyanophora* in relation to the plant kingdom, we examined the gene CpaMCM2 (https://conekt.sbs.ntu.edu.sg/sequence/view/13961), which is annotated as ‘DNA replication licensing factor MCM2’ and shows the typical cyclical expression of cell division genes, peaking towards the end of the light period during a diurnal cycle (Fig. S7) (Ferrari et al., 2019). MCM proteins contain a mini-chromosome maintenance (MCM) domain and are well known to be involved in DNA replication during both initiation and elongation (Bell & Dutta, 2002). Phylogenetic and expression analysis of the ortholog of the MCM2 gene (https://conekt.sbs.ntu.edu.sg/tree/view/5112) revealed that the gene has orthologous genes in *C. reinhardtii*, *S. moellendorfii*, *P. abies*, *O. sativa*, *Z. mays*, *V. vinifera*, *A. thaliana* and *S. lycopersicum*, and that these orthologs are expressed in various tissues in land plants (Fig. 4a).

CoNekT uses the Expression Context Conservation (ECC) and label co-occurrence values to identify conserved co-expressed gene modules (Movahedi et al., 2011; Proost & Mutwil, 2018). The ECC network showed that the *C. paradoxa* cell division module is conserved across the Archaeplastida kingdom (Fig. 4b, https://conekt.sbs.ntu.edu.sg/ecc/graph/13961/3/1). We then further explored the conserved corresponding module in *C. reinhardtii*, the species which showed the highest ECC value (0.11) to the *Cyanophora* MCM2 gene. The ECC pair network (https://conekt.sbs.ntu.edu.sg/ecc/graph_pair/758667) showed the presence of several orthologs (dashed lines) and, among them, multiple genes with an MCM domain (light blue square). Several other genes with domains related to DNA replication and cell division are found in the network, for example DNA ligase (Moriyama & Sato, 2014), SRA_YD (Citterio et al., 2004), DNA primase (Moriyama & Sato, 2014), Cdc6 (Mar Castellano et al., 2003), tubulins (Cross & Umen, 2015), DNA polymerase and MinD (Adams et al., 2003; Miyagishima et al., 2012). Closer inspection of the genes present in the network revealed the
ellipse), peptidases (green ellipse), DNA repair (black ellipse) and transporters (yellow ellipse) are indicated. Gene expression levels are indicated on the color-coded according to the type of the experiment, where, for example, nutrient stress samples are colored light blue. Gene expression levels are most for one bin are shown. (b) Average expression profile of the genes found in cluster 54. The different experiments are indicated on the x-axis and are defined by MAPMAN bins. Clusters enriched for a MAPMAN bin (FDR adjusted $P < 0.05$) are indicated by red squares. For brevity, only 79 clusters enriched at most for one bin are shown. (b) Average expression profile of the genes found in cluster 54. The different experiments are indicated on the x-axis and are defined by MAPMAN bins. Clusters enriched for a MAPMAN bin (FDR adjusted $P < 0.05$) are indicated by red squares. For brevity, only 79 clusters enriched at most for one bin are shown. (c) Simplified co-expression network of cluster 54. The genes involved in photosynthesis (red ellipse), oxidoreductases (blue ellipse), peptidases (green ellipse), DNA repair (black ellipse) and transporters (yellow ellipse) are indicated.

Fig. 3 Functional analysis of co-expression clusters of Cyanophora paradoxa. (a) Columns represent cluster IDs, while rows indicate biological processes, as defined by MAPMAN bins. Clusters enriched for a MAPMAN bin (FDR adjusted $P < 0.05$) are indicated by red squares. For brevity, only 79 clusters enriched at most for one bin are shown. (b) Average expression profile of the genes found in cluster 54. The different experiments are indicated on the x-axis and are defined by MAPMAN bins. Clusters enriched for a MAPMAN bin (FDR adjusted $P < 0.05$) are indicated by red squares. For brevity, only 79 clusters enriched at most for one bin are shown. (c) Simplified co-expression network of cluster 54. The genes involved in photosynthesis (red ellipse), oxidoreductases (blue ellipse), peptidases (green ellipse), DNA repair (black ellipse) and transporters (yellow ellipse) are indicated.
presence of two orthologs for DNA polymerase A (\textit{Cpa|evm.model.tig00020554.77}, \textit{Cpa|evm.model.tig00000342.9}) and one for DNA polymerase D (\textit{Cpa|evm.model.tig0001545.7}) (Zones \textit{et al.}, 2015) and four orthologs of the MCM protein necessary to form the heterohexamer active during DNA replication (\textit{Cpa|evm.model.tig00020554.77}, \textit{Cpa|evm.model.tig00000144.164}, \textit{Cpa|evm.model.tig00000241.38}, \textit{Cpa|evm.model.tig00000241.38}, \textit{MCM2}; \textit{Cpa|evm.model.tig00000144.164}; \textit{MCM4}; \textit{Cpa|evm.model.tig00020918.5}, \textit{MCM6}; \textit{Cpa|evm.model.tig0000241.38}, \textit{MCM7}) (Tuteja \textit{et al.}, 2011). Together, these results show how the core components of the cell division machinery are conserved over a great evolutionary distance and showcase how the expression data and the updated CoNekT-Plants can be mined to study the conservation of biological processes.
Comparative analysis of stress responses in algae and flowering plants

Comparative transcriptomic studies across species have helped us considerably to understand which biological pathways are conserved and how they have evolved over time (Mutwil et al., 2011; Movahedi et al., 2012; Hansen et al., 2014; Ruprecht et al., 2017a). For example, we showed that diurnal transcriptional programs are conserved over a huge time span of 1 billion years (Ferrari et al., 2019). Co-expression networks are also conserved across several species (Ruprecht et al., 2011, 2016, 2017a; Mutwil et al., 2011). In the present study we aimed to investigate if the same is true for transcriptomic responses to a wide range of abiotic stresses. To this end, we retrieved publicly available gene expression data capturing similar stresses as studied in C. paradoxa (Table S3) for Synechocystis sp. PCC 6803, C. reinhardii and A. thaliana. Gene expression data for high light, N deprivation and micronutrient stress are available for all four species. Data for sulfate deprivation have been published for all species except for C. reinhardii, and dark, heat, cold, phosphate deficiency and salt stress data are available for C. paradoxa and A. thaliana. We also included our data on mannitol stress for C. paradoxa.

We first investigated how many genes change their expression profile upon exposure to different stresses. The fraction of stress-specific DEGs was overall below 20% across species, except for prolonged darkness, which resulted in about 26% and 31% of DEGs in Arabidopsis and Cyanophora, respectively (Fig. 5a, gray bars; Tables S4–S7). Other stresses that caused strong transcriptional responses were nitrogen deprivation and temperature stresses. Interestingly, Synechocystis shows an overall large transcriptomic response with many DEGs found for most stresses, suggesting that cyanobacteria might respond more strongly to stresses. As TFs play an essential role in coordinating transcriptional responses to abiotic stresses, we explored whether they are more differentially expressed in certain stresses. Interestingly, we observed that the percentage of differentially expressed TFs is comparable across the different stresses, but higher in Arabidopsis (an average of 8% of differentially expressed TFs, Fig. 5b) than the algae (1%). Finally, we observed that in most cases the proportion of up- and down-regulated genes was equal (Fig. 5c, blue/red bars). The few exceptions included light stress in C. paradoxa and phosphate and iron deprivation in A. thaliana, showing a high proportion of up-regulated genes, and sulfate deprivation in Synechocystis and salt stress in A. thaliana instead showing predominantly down-regulated genes.

We then explored how the different biological processes were affected by these stresses. We analyzed the fraction of DEGs in each of the MAPMAN ontology bins (Fig. 5d). For brevity, we only discuss the general MAPMAN terms. Prolonged darkness had the strongest impact on the transcriptome, especially for C. paradoxa, where more than one-third of all biological processes (11 out of 29) showed more than 50% DEGs (Fig. 5d). The majority of these processes shows a mild prevalence of down-regulated genes, suggesting a response to carbon starvation. Interestingly, we observed a general tendency for down-regulation of genes involved in photosynthesis for all four species and across all stress conditions.

To test whether stress responses are conserved across species, we measured the similarity of responses of gene families in the different experiments. As a measure of similarity we determined whether a pair of stresses contains more common up- or down-regulated gene families than expected by chance. We found that the down-regulatory responses (blue cells) were more frequent than the up-regulatory responses (red cells, Fig. 6a). The high light treatment tends not to be similar to the other stresses, especially in C. reinhardii and C. paradoxa, while in A. thaliana it shows a rather similar response to other stress within A. thaliana (blue cells, Fig. 6a; see Fig. S8 where the samples are grouped according to the stresses). Conversely, other conditions showed significantly similar down-regulatory responses across stress types and species. For example, the nitrogen depletion response is similar across the four analyzed species, but also similar to the other nutrient depletion conditions, heat, and cold responses (blue cells, Figs 6a, S8). The significant conservation of responses was observed for most stresses (Fisher’s method P<0.05), with the exception of cold and salt treatments (P>0.05), indicating that the transcriptional responses to the two latter stresses are not conserved across species (Fig. 6b). Interestingly, we also observed that the transcriptomic responses to multiple stresses within one species are also conserved. For example, the response of C. paradoxa to all stresses was significantly similar, regardless of the stress (P<0.01, Fig. 6c), indicating a general transcriptional program that is activated upon stress. This is further supported by the significant conservation of both up- and down-regulated responses in all species and stresses (P<0.001), suggesting that the transcriptomic responses are conserved. We conclude that the down-regulatory responses are highly conserved, regardless of the type of stress and species.

Next, investigated which gene families tend to significantly respond to the stresses (Figs 6d, S9). We observed that gene families involved in photosynthesis, tetrpyrrole biosynthesis and glycolysis showed a predominant down-regulation response. Other processes, such as nutrient uptake, cellular respiration, amino acid production, and protein biosynthesis and modification showed a mixed response to the stresses. Together, these results show that the conserved responses across the Archaeplastida kingdom involve down-regulation of numerous transcripts, especially photosynthesis, during stress.

Phylostratigraphic analysis of stress responses

It has been previously shown that the age (phylostrata) of a gene family (Domazet-Lošo et al., 2007) is correlated with the expression levels of the corresponding genes in the plant kingdom (Ruprecht et al., 2017a; Ferrari et al., 2019) and the number of publications dedicated to the gene family (Ruprecht et al., 2017a; Hansen et al., 2018). The latter finding suggests that older gene families produce more pronounced mutant phenotypes and are involved in more fundamental biological processes. It is unclear, however, whether gene families that appeared at a specific time in plant evolution are more stress responsive.
To investigate whether the age of a gene family is correlated to its stress response, we arranged the phylostrata from oldest (prokaryotic) to youngest (species-specific gene families) and observed that the oldest phylostrata (prokaryotic and Archaeplastida I) showed the highest enrichment for the down-regulation response, being significantly enriched in >45% of the experiments (Fig. 7). Furthermore, the Gymnospermae phylostratum, representing gene families that appeared in the ancestor of seed plants, tends to be enriched only in high light and heat (Fig. 7). Conversely, species-specific genes showed the least enrichment, suggesting that newly formed genes are less involved in the stress response (Fig. 7). We conclude that stress responses tend to be regulated independently of the age of the gene families, except the oldest (prokaryotic, more responsive than expected) and the youngest (species-specific, less responsive than expected).

**Discussion**

*C. paradoxa* is the only glaucophyte representative with a sequenced genome, and due to the importance of this species for evolutionary studies we generated a comprehensive expression atlas comprising several abiotic stresses. We focused the atlas on a wide range of abiotic stresses for two reasons. The first is that...
Fig. 6 Similarity of the transcriptomic responses to stresses. (a) Conservation of transcriptomic responses of the different stresses and species. The rows and columns indicate the stresses (e.g. dark, heat, cold) and species (e.g. Arabidopsis: purple, Chlamydomonas: yellow), while the lower left and upper right part of the heatmap indicates the conservation of down- and up-regulated responses, respectively. The intensity of the cell colors indicates the FDR-corrected P-value of conserved responses. (b) Similarity of transcriptomic responses to a given stress across the different species. Fisher’s method P-values indicating significant up-regulation (red), down-regulation (blue), and up- and down-regulation are indicated in the key. (c) Similarity of transcriptomic responses of a species to different stresses. Fisher’s method P-values are color-coded as above. (d) Transcriptomic responses of the gene families that respond significantly to the stresses included in this study. Rows indicate the function of the gene families and columns represent stresses and species. The color of the cells indicates whether the genes in a given family are up-regulated (red), down-regulated (blue), up- and down-regulated (green) or unchanged (gray) for a given stress. Trace metal solution (TMS) refers to medium depleted of micronutrient solution.
expression networks require gene expression data capturing as many different observations as possible (Usadel et al., 2009). The second is that all organisms have developed transcriptional programs to cope with various abiotic stresses (Zhu, 2016), but currently no study is available on gene expression comparison at the plant kingdom-wide scale.

The *Cyanophora* gene expression atlas revealed that there was no correlation between the severity of growth inhibition caused by the stress and the number of transcriptional changes (given as the number of DEGs) \((r=0.278, P=0.437, \text{Fig. 1e})\), even though both culture growth and the transcriptome were affected by the applied stresses. For instance, stresses that cause a mild growth retardation (e.g. nitrogen and sulfate deprivation) showed a high number of DEGs, while stresses causing a severe growth inhibition (salt and mannitol) showed the lowest number of DEGs. The overall low correlation between growth inhibition and transcriptome changes suggests highly diverse strategies and regulatory programs that control stress acclimation.

We expanded the CoNekT-Plants database with the newly generated expression dataset to allow easy access to gene expression and co-expression networks through the use of various comparative tools. These tools allow us to overcome the paucity of knowledge related to gene function in *C. paradoxa*. We exemplify these tools by studying phycobilisome formation (Figs 2c, S4) and high light response (Figs 3c, S6), demonstrating how the co-expression network can be used to predict the function of unknown genes. Furthermore, the updated CoNekT-Plants can be used for cross-species analyses, as we have demonstrated by identifying a conserved module involved in cell division in *C. paradoxa* and *C. reinhardtii* (Fig. 4c). These tools thus allow us to uncover the similarities and differences in gene expression and functional modules across evolutionary distances, making CoNekT-Plants a unique resource to study the evolution of functional modules.

Moreover, we analyzed whether the responses to abiotic stresses are conserved across conditions and organisms, and we observed a conserved down-regulation response related to genes involved in photosynthesis and primary metabolism (Fig. 5d). This observation is also supported by the analysis of the most conserved families operating upon stress (Figs 6d, S9), in which we find several, mostly down-regulated photosynthesis and primary metabolism-related families. Similar trends have been found in grasses, where the various members of the family Pooidae showed highly diverse transcriptomic responses to cold stress, but a common down-regulation of transcripts involved in photosynthesis and metabolism (Schubert *et al.*, 2019). Analysis of photosynthesis parameters during salt stress revealed that this is reflected by a decrease in the total performance index of photosynthesis in multiple species (Pavlović *et al.*, 2019), showing a correlation between the transcript responses and plant phenotype. Our findings suggest that photosynthesis is down-regulated to prevent photoinhibition and cellular damage not only in cold and salt stress but for a wider variety of stresses. Furthermore, because this response is seen in algae and land plants, we propose that down-regulation of photosynthesis and metabolism is a kingdom-wide, stress-responsive program that is activated for various types of stresses. Because transcriptomic responses of a species are significantly similar regardless of the type of stress (Figs 6c, S8), this finding should caution researchers to claim that transcriptomic responses of the same type of stress are conserved across species. Indeed, rather than detecting conserved, specific responses to a given stress, the researcher might be observing the conserved, kingdom-wide stress response program.

In contrast to the conserved down-regulation responses, we observed only a modest conservation of up-regulation responses to various stresses across different species (Fig. 6a). This is also reported for salt stress in six *Lotus* accessions, where only 1% of genes showed a conserved response (Sanchez *et al.*, 2011), in two
strawberry cultivars, which displayed a modest conservation of DEGs to the same pathogen (Wang et al., 2017), in Poaceae, which revealed a poor expression conservation of orthologs across same tissue types (Davidson et al., 2012), and in seven Arabidopsis accessions, which showed a divergent response of genes to exogenous salicylic acid (Van Leeuwen et al., 2007). We propose two nonequivalent hypotheses for this surprising lack of conservation in transcriptomic responses. First, assuming that orthologous sets of genes are still needed for acclimation to a given stress, investigating only one layer of the gene activity regulation (transcript levels) might be insufficient. Second, a high rate of rewiring, divergence and the presence of multiple signal transduction pathways might confound the signal. To provide a deeper understanding of stress responses, we suggest that snapshots of additional regulatory layers, preferably closer to the end-product of gene regulation (i.e. the protein), should be included in future studies.

Finally, we have demonstrated that the oldest gene families, in contrast to the youngest species-specific families, are actively responding to abiotic stress (Fig. 7). This suggests that the young gene families are not typically involved in stress acclimation, which is in line with the fact that the investigated stresses, and the corresponding coping mechanisms, are ancient.

Thus, our study on abiotic stress responses in *C. paradoxa* provides new insights into stress responses, gene function prediction, and the conservation and evolution of abiotic stress acclimation.

Acknowledgements

We thank Dr Daniela Mutwil-Anderwald for proofreading the manuscript.

Author contributions

MM conceived the project. CF and MM designed the study. CF performed the experiments and data analysis with input from MM. CF and MM wrote the manuscript.

ORCID

Camilla Ferrari https://orcid.org/0000-0002-5843-0500
Marek Mutwil https://orcid.org/0000-0002-7848-0126

References

Adams KL, Cronn R, Percifield R, Wendent JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of Sciences, USA* 100: 4649–4654.

Adl SM, Simpson AGB, Lane CE, Lukes J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampshire V et al. 2012. The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology* 59: 429–493.

Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T. 2016. ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression. *Plant and Cell Physiology* 57: 65.

Avidan O, Brandis A, Rogachev I, Pick U. 2015. Enhanced acetyl-CoA production is associated with increased triglyceride accumulation in the green alga *Chlorella* desiccata. *Journal of Experimental Botany* 66: 3725–3735.

Barabási A-L, Bonabeau E. 2003. Scale-free networks. *Scientific American* 288: 60–69.

Barabási AL, Oltvai ZN. 2004. Network biology: understanding the cell’s functional organization. *Nature Reviews Genetics* 5: 101–113.

Bassel GW, Lan H, Glazeb E, Gibbs DJ, Gerjets T, Krasnogor N, Bonner AJ, Holdsworth MJ, Provant NJ. 2011. Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proceedings of the National Academy of Sciences, USA* 108: 9709–9714.

Van Bel M, Diels T, Vancaestre E, Kreft I, Botzki A, Van De Peer Y, Coppens F, Vandepoele K. 2018. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research* 46: D1190–D1196.

Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL, Zoran MJ. 2005. Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature Reviews Genetics* 6: 544–556.

Bell SP, Dutta A. 2002. DNA replication in eukaryotic cells. *Annual Review of Biochemistry* 71: 333–374.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57: 289–300.

Bhattacharya D, Yoon HS, Hackett JD. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *BioEssays* 26: 50–60.

Blank CE. 2013. Origin and early evolution of photosynthetic eukaryotes in freshwater environments: reinterpreting Proterozoic paleobiology and biogeochemical processes in light of trait evolution. *Journal of Phycology* 49: 1040–1055.

Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12: 59–60.

Chang I, Liu X, Li Y, Liu CC, Yang F, Zhao J, Sui SF. 2015. Structural organization of an intact phycobilisome and its association with photosystem II. *Cell Research* 25: 726–737.

Citterio E, Papait R, Nicassio F, Vecchi M, Gomiero P, Mantovani R, Di Fiore PP, Bonapace IM. 2004. Np95 is a histone-bending protein endowed with ubiquitin ligase activity. *Molecular and Cellular Biology* 24: 2526–2535.

Cross FR, Umen J. 2015. The *Cblaminus* gene family. *The Plant Journal* 82: 370–392.

Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu SH, Jiang N, Robin Buell C. 2012. Comparative transcriptomics of three *Poaceae*-species reveals patterns of gene expression evolution. *The Plant Journal* 71: 492–502.

Doman-Losi T, Brjakovic I, Tautz D. 2007. A phylogenigraph approach uncovers the genomic history of major adaptations in metazoan lineages. *Trends in Genetics* 23: 533–539.

Durán RV, Hervás M, De La Rosa MA, Navarro JA. 2004. The efficient functioning of photosynthesis and respiration in *Synechocystis* sp. PCC 6803 strictly requires the presence of either cytochrome c6 or plastocyanin. *Journal of the Royal Society Interface* 279: 7229–7233.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16: 157.

Enright AJ. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* 30: 1575–1584.

Ferrari C, Proost S, Janowski M, Becker J, Nikolski Z, Bhattacharya D, Price D, Tohge T, Bar-Even A, Fernie A et al. 2019. Kingdom-wide comparison reveals the evolution of diurnal gene expression in *Archaeplastida*. *Nature Communications* 10: 737.

Fromme P, Jordan P, Krauß N. 2001. Structure of photosystem I. *Biochimica et Biophysica Acta – Bioenergetics* 1507: 5–31.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. Affy – Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20: 307–315.

Gross J, Wajid S, Price DC, Zelzion E, Chan CX, Bhattarcharya D. 2013. Evidence for widespread exonic small RNAs in the glaucophyte alga *Cyanophora paradoxa*. *PLoS ONE* 8: e67660.

Hansen BO, Meyer EH, Ferrari C, Vaid N, Movahedi S, Vandepoele K, Nikolski Z, Mutwil M. 2018. Ensemble gene function prediction database reveals genes important for complex I formation in *Arachidopsis thaliana*. *New Phytologist* 217: 1521–1534.

Bell SP, Dutta A. 2002. DNA replication in eukaryotic cells. *Annual Review of Biochemistry* 71: 333–374.
Takabayashi A, Ishikawa N, Obayashi T, Ishida S, Obokata J, Endo T, Sato F. 2009. Three novel subunits of Arabidopsis chloroplastic NAD(P)H dehydrogenase identified by bioinformatic and reverse genetic approaches. The Plant Journal 57: 207–219.

Takahashi N, Lammens T, Boudolf V, Maes S, Yoshizumi T, De Jaeger G, Witters E, Inzé D, De Veylder L. 2008. The DNA replication checkpoint aids survival of plants deficient in the novel replisome factor ETG1. EMBO Journal 27: 1840–1851.

Thomas L. 1982. The lives of eukaryotic cells symbiosis in cell evolution. Life and its environment on the early earth. Lynn Margulis. The Quarterly Review of Biology 57: 167–169.

Tuteja N, Tran NQ, Dang HQ, Tuteja R. 2011. Plant MCM proteins: role in DNA replication and beyond. Plant Molecular Biology 77: 537–545.

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ. 2009. Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant, Cell & Environment 32: 1633–1651.

Wang F, Zhang F, Chen M, Liu Z, Zhang Z, Fu J, Ma Y. 2017. Comparative transcriptomics reveals differential gene expression related to Colletotrichum gloeosporioides resistance in the octoploid strawberry. Frontiers in Plant Science 8: 779.

Watanabe M, Sato M, Kondo K, Narikawa R, Ikeuchi M. 2012. Phycobilisome model with novel skeleton-like structures in a glaucocystophyte Cyanophora paradoxa. Biochimica et Biophysica Acta – Bioenergetics. 1428–1435.

Whitehead K, Pan M, Masumura KI, Bonneau R, Baliga NS. 2009. Diurnally entrained anticipatory behavior in archaea. PLoS ONE 4: e5485.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Molecular Biology and Evolution 21: 809–818.

Yu H, Luscombe NM, Qian J, Gerstein M. 2003. Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends in Genetics 19: 422–427.

Zhu JK. 2016. Abiotic stress signaling and responses in plants. Cell 167: 313–324.

Zones JM, Blaby IK, Merchant SS, Umen JG. 2015. High-resolution profiling of a synchronized diurnal transcriptome from Chlamydomonas reinhardtii reveals continuous cell and metabolic differentiation. Plant Cell 27: 2743–2769.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Coverage of MapMan bins for Synechocystis, C. paradoxa, C. reinhardtii and A. thaliana.

Fig. S2 Quantification of starch in prolonged darkness.

Fig. S3 PCA of C. paradoxa.

Fig. S4 Co-expression network of the CpcK2 gene.

Fig. S5 Significantly enriched clusters of C. paradoxa.

Fig. S6 Co-expression network of cluster 54.

Fig. S7 Diurnal expression profile of MCM2 (Cpa|evm.model.tig00020554.77).

Fig. S8 Similarity of the transcriptomic responses to stresses.

Fig. S9 Commonly enriched families across abiotic stresses and species.

Table S1 Compositions of the C media.

Table S2 LSTRAP mapping statistics.

Table S3 Gene expression data used in this study.

Tables S4–S7 Up- and down-regulated genes in all stresses and species.

Table S8 Cyanophora paradoxa TPM-normalized gene expression matrix.

Table S9 HCCA clusters of C. paradoxa.

Table S10 Co-expression network of C. paradoxa.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.