

PhytoNet: comparative co-expression network analyses across phytoplankton and land plants

Camilla Ferrari¹, Sebastian Proost¹, Colin Ruprecht² and Marek Mutwil^{1,3,*}

¹Max-Planck Institute for Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam, Germany, ²Max-Planck Institute of Colloids and Interfaces, Am Muehlenberg 1, 14476 Potsdam, Germany and ³School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore

Received January 26, 2018; Revised March 26, 2018; Editorial Decision April 06, 2018; Accepted April 11, 2018

ABSTRACT

Phytoplankton consists of autotrophic, photosynthesizing microorganisms that are a crucial component of freshwater and ocean ecosystems. However, despite being the major primary producers of organic compounds, accounting for half of the photosynthetic activity worldwide and serving as the entry point to the food chain, functions of most of the genes of the model phytoplankton organisms remain unknown. To remedy this, we have gathered publicly available expression data for one chlorophyte, one rhodophyte, one haptophyte, two heterokonts and four cyanobacteria and integrated it into our PlaNet (Plant Networks) database, which now allows mining gene expression profiles and identification of co-expressed genes of 19 species. We exemplify how the co-expressed gene networks can be used to reveal functionally related genes and how the comparative features of PhytoNet allow detection of conserved transcriptional programs between cyanobacteria, green algae, and land plants. Additionally, we illustrate how the database allows detection of duplicated transcriptional programs within an organism, as exemplified by two putative DNA repair programs within *Chlamydomonas reinhardtii*. PhytoNet is available from www.gene2function.de.

INTRODUCTION

While phytoplankton (used here in the broadest sense that includes cyanobacteria) is often not visible by unaided eye, it plays a crucial role in marine environments. These photosynthesis-capable organisms form the foundation of aquatic ecosystems and contribute to ~50% of the global carbon assimilation (1–3). Cyanobacteria are extant relatives of the organisms that gave rise to plant chloroplasts

through endosymbiosis ~1.5 billion years ago, and therefore are vital to understand the origin of plants (4). Furthermore, the phylogenetic position of green algae makes them interesting outgroups for evolutionary studies focussed on higher plants (5). Having often a simpler genetic makeup than higher plants, various single-celled algae have been hailed as a ‘green yeasts’, with potential applications ranging from the production of biofuels to food supplements (6,7). Finally, diatoms and coccolithophores, which construct intricate shells from silica and calcium carbonate respectively (8,9), have value for biotechnology and nanotechnology (10).

Despite their importance, the function of many algal genes remains unknown, and gene function prediction approaches for algae rely mainly on sequence similarity analysis. These approaches thus lag behind when compared to gene function prediction methods applied to model organisms in other branches of the tree-of-life (11). Online resources focussed on algae are rare, and while a comparative genomics platform was established recently (12), tools to study gene expression in algae are still lacking. As genes with similar expression profiles across developmental stages, environmental perturbations and organs tend to be functionally related, identification of co-expressed genes has been used to reveal functionally related genes (13–17). These co-expression relationships can be visualized as networks, where nodes (or vertices) correspond to genes and edges (or links) connect genes that display similar expression profiles (18). Thus, by identifying groups of connected genes in these networks, functional gene modules can be detected and used to predict gene function. Furthermore, gene modules conserved over large phylogenetic distances can reveal core components of shared processes, while duplicated gene modules can reflect recent adaptations combined with an increase in complexity (19–21).

Here, we present PhytoNet (www.gene2function.de), a freely available extension to PlaNet, which allows studying co-expression networks in nine algae through an online interface.

*To whom correspondence should be addressed. Tel: +6569047503; Email: mutwil@ntu.edu.sg

MATERIALS AND METHODS

Integration of nine phytoplankton species into the PlaNet database

To make PhytoNet a complete platform to study gene expression of phytoplankton, we screened ArrayExpress database for expression data of photosynthesizing microorganisms (22), with criteria: (i) the candidate organism must be sequenced and (ii) at least 20 expression datasets must be available (23,24). We arrived at nine species (Table 1, Figure 1A, Supplementary Table S1), which include four cyanobacteria (*Synechocystis* sp. PCC 6803, *Nostoc punctiforme* PCC 73102, *Prochlorococcus marinus* subsp. *pastoris* str. CCMP1986, *Cyanothece* sp. ATCC 51142), two heterokonts (*Ectocarpus siliculosus*, *Phaeodactylum tricoratum*), one rhodophyte (*Cyanidioschyzon merolae*), one haptophyte (*Emiliana huxleyi*) and one chlorophyte (*Chlamydomonas reinhardtii*). *Synechocystis*, *Cyanothece*, *Phaeodactylum* and *Emiliana* raw microarray data were downloaded and processed using the R-package *limma* (25), *Nostoc* and *Ectocarpus* raw data were normalized using *Deva* software (Roche, Nimblegen), *Chlamydomonas* raw RNA sequencing data were downloaded and transcript per million (TPM)-normalized using *LSTrAP* software (26), while *Prochlorococcus* and *Phaeodactylum* processed data were downloaded from ArrayExpress (Supplementary Table S1). *LSTrAP* was used to detect and discard samples that showed either (i) low mapping to the genome (< 65%), (ii) low mapping to coding sequences (<40%) or (iii) with too few useful reads (less than 8M reads mapping to the genome). The remaining samples were used to construct expression matrices and co-expression networks. Pfam domain labels were assigned to genes by blasting CDS sequences against Pfam database version 27 with an e-value cutoff of 10^{-5} (21,27), while Gene Ontology terms and gene families were obtained by using the CDS sequences in TRAPID (28).

To assess the quality of the expression data, we first analyzed the sample similarity dendrogram, with the assumption that replicates and samples representing similar experiments should group together. The sample similarity dendrograms showed an expected clustering of replicates and related samples (i.e. separate clustering of *Synechocystis* day and night samples, Supplementary Figures S1–S9). Furthermore, co-expression networks tend to show a scale-free topology (also called power law behavior), where most nodes have few connections, and few nodes have many connections (29). The expression data of the nine species produced the expected pattern on a line with a negative slope (recommended Pearson Correlation Coefficient cutoff >0.7, (30), Supplementary Figure S10), which is indicative of scale-freeness of the co-expression networks (24,31,32). Taken together, this indicates that the data is of sufficient quantity and quality to produce co-expression networks with correct topology.

The expression data were used to construct Highest Reciprocal Rank (HRR) co-expression networks, and the clusters of co-expressed genes were identified with Heuristic Cluster Chiseling Algorithm (HCCA, (32)). To detect conserved and duplicated modules, we applied the FamNet

pipeline, which identifies co-expression neighborhoods that contain same gene families and Pfam domains (21). Briefly, when two neighborhoods of sizes A and B are compared, the number of Pfam domains and gene families (labels) in common between the two neighborhoods is counted and compared to an expected number of random neighborhoods of these sizes (21). Finally, the data were uploaded to the PlaNet database (33).

RESULTS

Identification of functionally related gene clusters

In order to have a clear overview of the biological processes present in the co-expression networks, the networks were clustered with the HCCA algorithm, which is optimized to cluster HRR-based co-expression networks (32). For the unicellular green algae *Chlamydomonas reinhardtii*, 173 clusters were identified (the number of clusters identified for each species is shown in Table 1). In order to elucidate the biological functions of these clusters, we computed enriched MapMan gene ontology bins present in the clusters (34). Sixty three of the clusters showed significant enrichment in at least one of the 35 MapMan bins, and 14 of them showed enrichment in more than one process (false discovery rate-adjusted *P*-value < 0.05, Figure 1B). Processes such as minor carbohydrate metabolism, fermentation, oxidative pentose pathway, nitrogen metabolism, hormone metabolism, polyamine metabolism, micro RNA and mineral nutrition were not enriched in any of the *Chlamydomonas* clusters, while the bin ‘protein’ was one of the terms with highest enrichment (nine clusters), indicating the presence of numerous bins involved in synthesis, modification and degradation of proteins.

We demonstrate how these clusters can be used to identify relevant genes with a stress-related cluster from *Chlamydomonas*. Abiotic stress can negatively affect biomass production in all living organisms, but when applied to microalgae it can specifically lead to an increase of lipid content, which is of great interest for biofuel production (35–37). Numerous studies have investigated the transcription profiles and physiological changes in *Chlamydomonas* caused by abiotic stresses (38–40). Consequently, we decided to explore cluster 72 (<http://aranet.mpimp-golm.mpg.de/responder.py?name=gene!cre!c72>), which showed significant enrichment for the MapMan bin ‘stress’ (Figure 1B). Within the cluster, we found three groups of genes involved in three different processes (Figure 1C). The first and largest group comprised chaperones belonging to Pfams DNAJ (brown circle), HSP90 (beige square), HSP70 (violet square), chaperone activators AHSA1 (yellow circle), activating HSP90 (41) and GRPe (gray square, activating HSP70) (42). The second group contained metalloproteinases M16 (brown square), M20 (red circle), M32 (blue triangle) and a Clp ATPase involved in protein unfolding (ClpB, purple circle) (43). The third group included proteins involved in mitochondrial protein import, such as Translocases s17 (red hexagon), 23 (purple triangle), s44 (purple diamond), inner membrane translocases belonging to TIM17/22/23 families (yellow diamond, turquoise hexagon), mitochondrial carriers translocating proteins across membranes (red triangle) (44) and OXA1, which

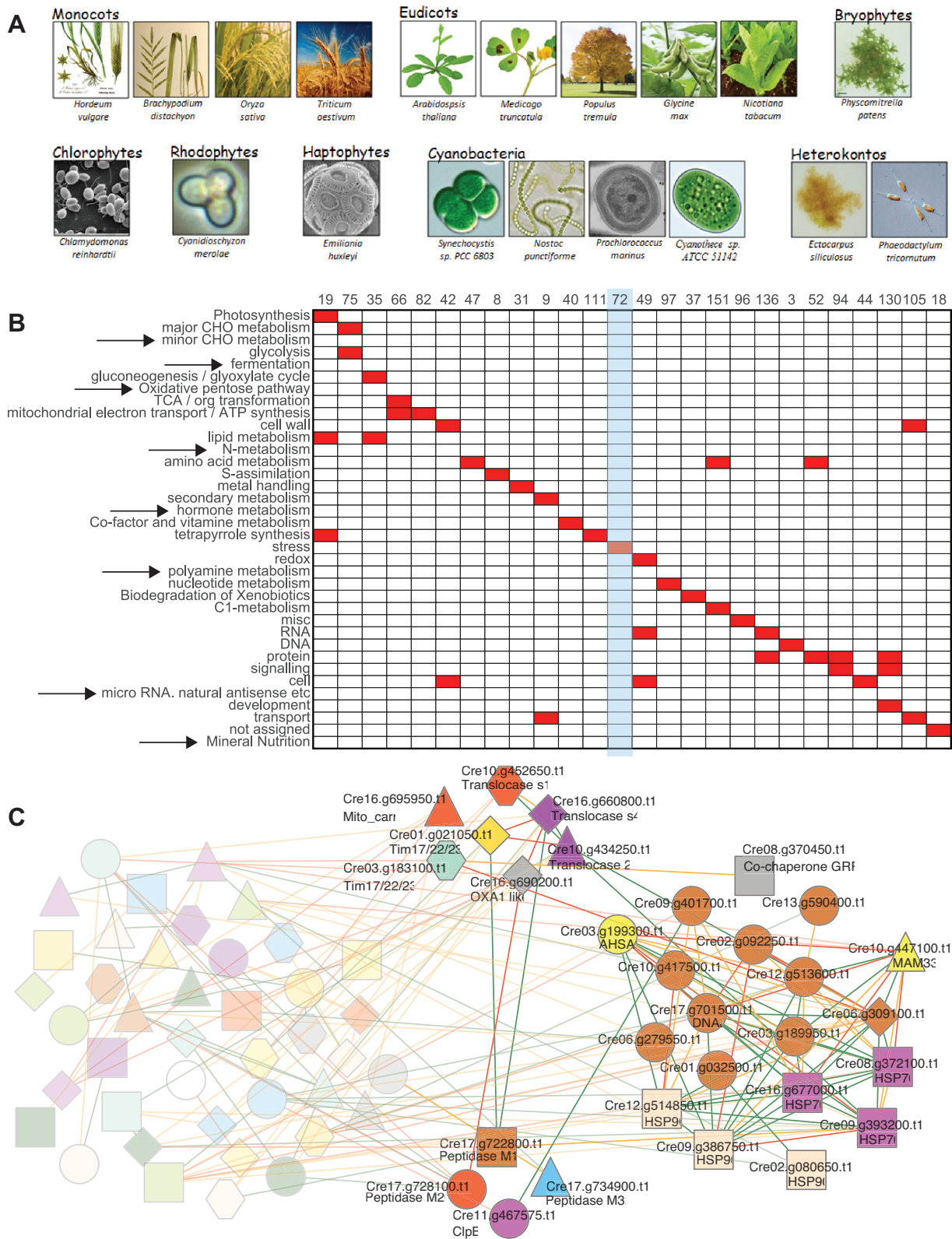


Figure 1. MapMan cluster analysis of *Chlamydomonas reinhardtii*. (A) An overview of species found in PhytoNet. (B) MapMan analysis of *Chlamydomonas* clusters. Rows represent MapMan bins (i.e. functional categories) while columns represent clusters. Red squares indicate that the specific term is significantly enriched (FDR-adjusted P -value < 0.05) in a given cluster. Arrows indicate bins with no significant enrichment. The light blue rectangle indicates the stress cluster 72. (C) Co-expression network representation of cluster 72 (stress). Nodes represent genes while edges connect genes which are coexpressed. Colored shapes are used to indicate genes that belong to the same gene family and/or contain same Pfam domains. Colored edges indicate the degree of co-expression between genes with green, orange and red edges denoting Highest Reciprocal Rank (HRR) of 10, 20 and 30, respectively. Note that certain gene IDs can appear several times in the network, as several probes can exist for one gene. For clarity, only discussed genes are highlighted.

Table 1. Description of the utilized expression data. The table indicates the number of expression measurements (samples), fraction of total protein-coding genes captured by the expression measurement platform (microarray or RNA sequencing), fraction of interrogated protein-coding genes with assigned Pfam domain and/or PLAZA gene family, types of experiments and the number of HCCA clusters

Organism	Phylum	# Experiments	# Interrogated protein coding genes / # in genome	# Genes with label / # interrogated genes	Description of experiments	# HCCA clusters
<i>Synechocystis</i> sp. PCC 6803	Cyanobacteria	138	3412/3725	2827/3412 (82.8%)	Depletion (C, Cu, As, Mn, Fe), diurnal and clock	35
<i>Nostoc punctiforme</i>	Cyanobacteria	90	7151/7664	5586/7151 (78.1%)	Heterocyst and hormogonium development	66
<i>Prochlorococcus marinus</i>	Cyanobacteria	40	1942/2275	1482/1942 (76.3%)	CO ₂ , O ₂ limitation time course	17
<i>Cyanothece</i> sp. ATCC 51142	Cyanobacteria	69	5251/5304	3719/5251 (70.8%)	N ₂ and CO ₂ treatment time series	99
<i>Cyanidioschyzon merolae</i>	Rhodophyta	75	4586/5331	3670/4586 (80%)	Diurnal and cell cycle	46
<i>Ectocarpus siliculosus</i>	Heterokontophyta	64	10658/16256	10263/10658 (96.3%)	Cu toxicity, hyposaline, hypersaline, oxidative stress	156
<i>Phaeodactylum tricorutum</i>	Heterokontophyta	58	10288/10402	7416/10288 (72.1%)	Blue, green, red light response, light/dark cycles, Cd toxicity	497
<i>Emiliana huxleyi</i>	Haptophyta	42	27108/ 0569	9944/27108 (36.7%)	Haploid and diploid growing in N and CO ₂ limitation, low / high light	268
<i>Chlamydomonas reinhardtii</i>	Chlorophyta	605	17741/17741		-N, -S, -Cu, -Fe, UV-B, H ₂ O ₂ and rapamycin treatment, Diurnal expression	173

is involved in protein insertion into the inner mitochondrial membrane (grey diamond) (45). These results revealed known components of protein folding, degradation and import into mitochondria, and suggest that these processes might be coordinated in *Chlamydomonas*. Taken together, the functionally enriched clusters can serve as a rapid entry point to identify genes involved in the biological process of interest.

Detecting conserved gene modules

The combination of gene families, Pfam domains, and co-expression modules allows detection and analysis of gene modules that are present in multiple species (13,46,47), and this property can be used to transfer functional knowledge from one species to another (48–51). Such conserved modules indicate that the biological process captured is present in the investigated species, and provides additional confidence that the homologous genes present in these modules are needed for the process and perform the same function (48).

To illustrate how conserved modules can be detected and analyzed, we used *Synechocystis* threonine synthase gene *sll1172* (<http://aranet.mpimp-golm.mpg.de/responder.py?name=genesyn!3362>). The expression profile of *sll1172* showed that the gene is expressed in all conditions, and co-expressed with other genes involved in amino acid biosynthesis, glycolysis and nucleotide metabolism (the co-expression network and gene ontology enrichment analysis is found on the gene page). The database visualizes conserved modules by using a gene module network (Figure 2A). The nodes in the network represent modules (groups of co-expressed genes), blue edges indicate conserved modules (i.e. modules that contain same gene families and Pfam

domains), while the edge style indicates the number of gene families and Pfam domains in common between two modules (52). Finally, modules that are overlapping (i.e. contain same co-expressed genes) are connected by orange edges, while green edges indicate duplicated modules (discussed in the next section).

The gene module network for *sll1172* indicated that conserved modules were found in flowering plants (green and orange nodes), cyanobacteria (blue nodes), early land plants (moss *Physcomitrella*, yellow nodes), red alga (*Cyanidioschyzon*, red nodes), green algae (*Chlamydomonas*, brown nodes), heterokonts (purple nodes) and haptophytes (*Emiliana*, gray nodes), suggesting that the module is conserved over large evolutionary distance. To exemplify a typical analysis and to gain insight into the function of these conserved modules, we selected *Synechocystis sll1172*, *Arabidopsis At3g59760* and *Physcomitrella Pp1s370_61v6.1* from the table below the module network and clicked on ‘Compare’ button. The output of the analysis indicated which common gene families and Pfam domains were found in the three modules, by using colored shapes to show presence or absence of a family (Figure 2B). The database also provides a more detailed view of the modules, where individual genes are shown (Figure 2C). Closer inspection of the *Arabidopsis* module revealed numerous genes involved in ATP synthesis (indicated by a black box), amino acid synthesis (blue box), steroid biosynthesis (red box), glycolysis (green box) and adenylate salvage (purple box, Figure 2C legend). To conclude, these results show that the observed primary metabolic processes are transcriptionally coordinated in cyanobacteria, plantae and species that arose via secondary endosymbiosis, indicating that the observed modules are ancient.

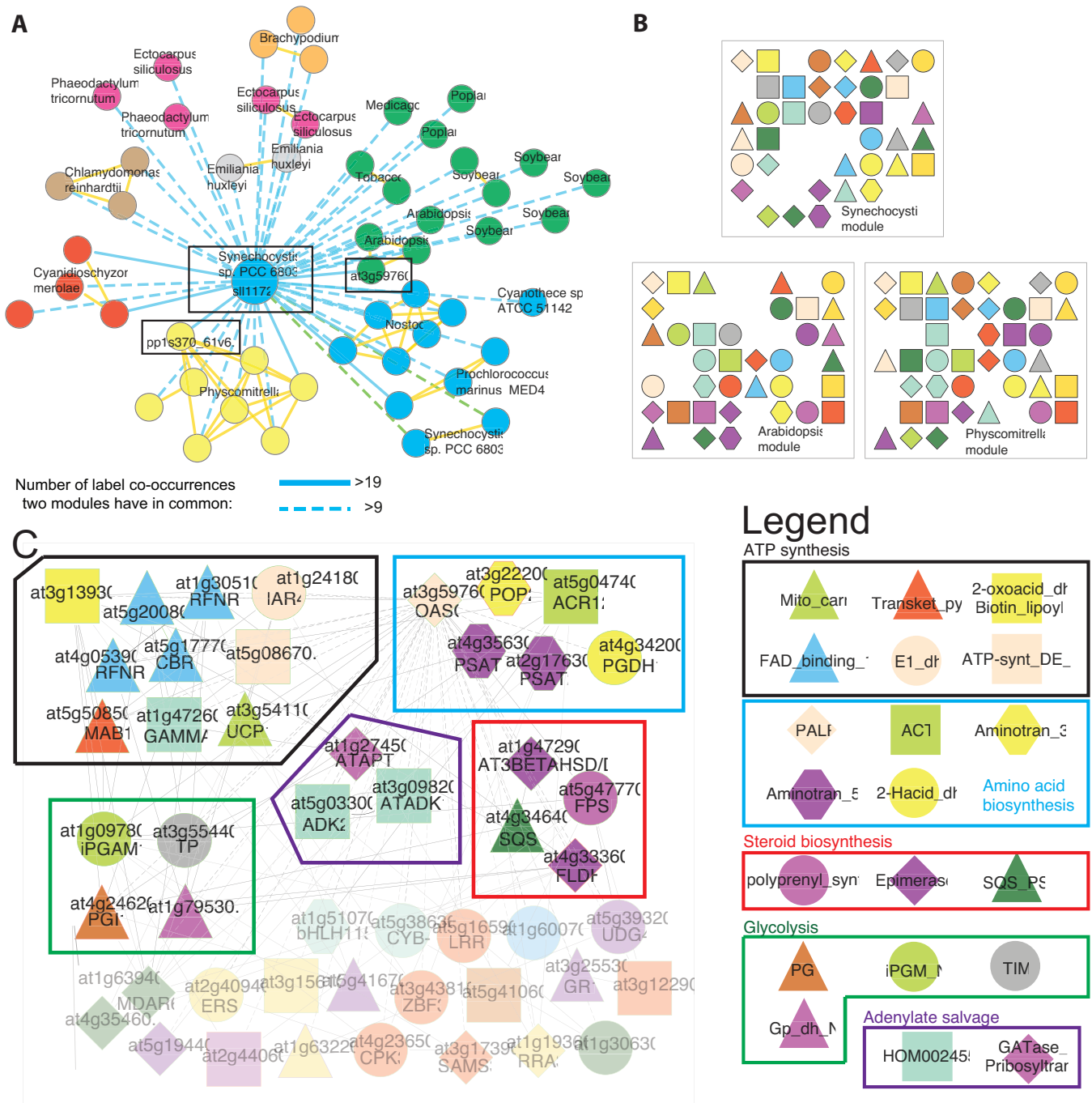


Figure 2. Detailed gene module analysis of the algae. (A) Gene module network of *sll1172* showing only modules with at least 10 gene families and Pfam domains (= label co-occurrences) in common with the bait module. The central node (large, blue circle) represents *sll1172* gene module, while similar modules are connected by blue edges, where edges with heavier weight (solid edges) connect gene modules with at least 20 label co-occurrences in common. Yellow and green edges indicate overlapping modules and similar modules within the same species. The different phyla present in the network are indicated by the different node colors. For example, cyanobacteria, dicots, and bryophytes are indicated by blue, green and yellow nodes. The three selected modules are indicated by black boxes. (B) Gene families and Pfam domains indicated by colored shapes that the three modules have in common. (C) Co-expression network of the Arabidopsis gene module. Nodes represent genes, colored shapes represent gene families and Pfam domains, while co-expressed genes are connected by gray edges. The discussed genes and gene families are highlighted by colored boxes.

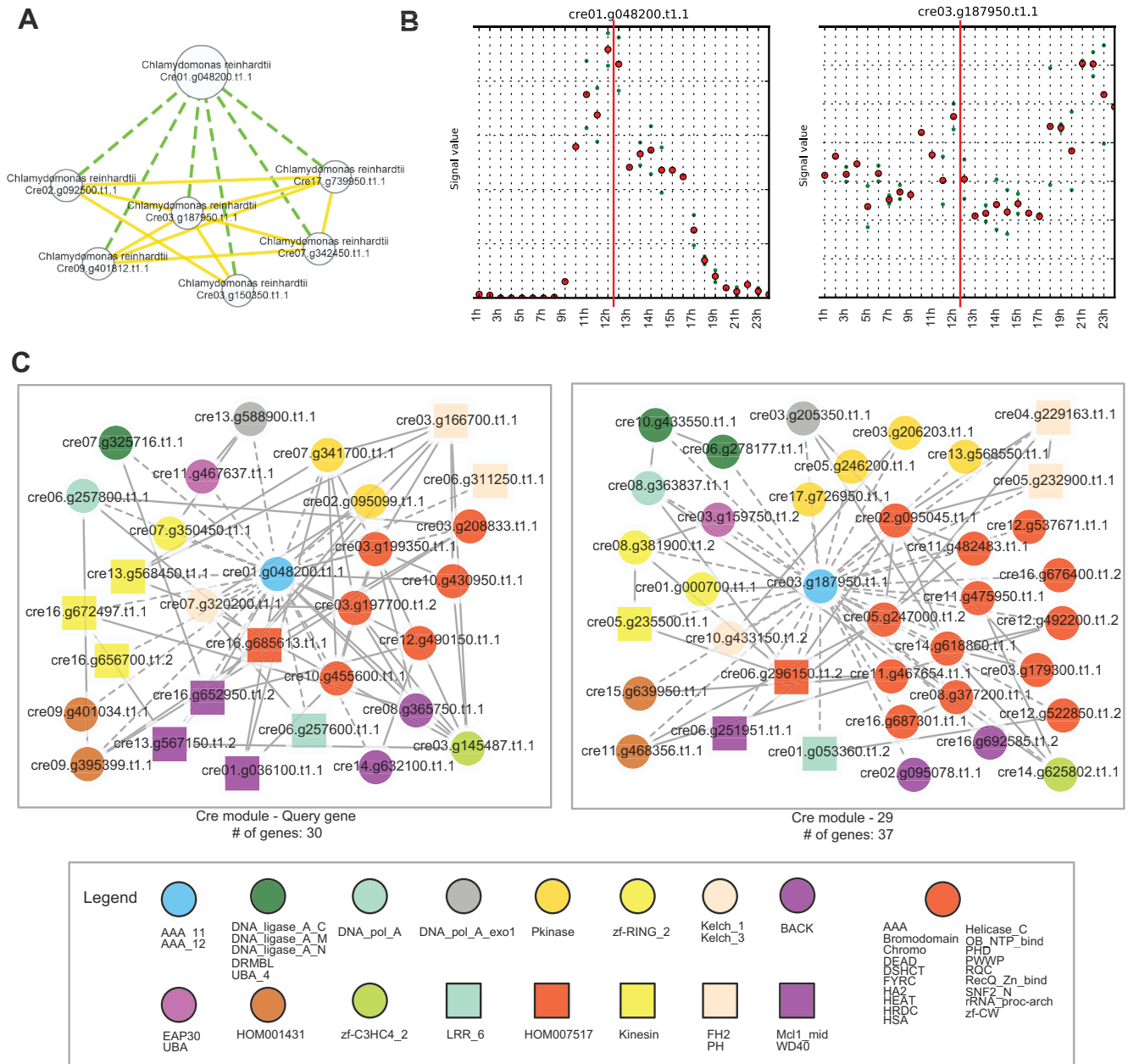


Figure 3. Duplicated modules involved in DNA repair in *Chlamydomonas reinhardtii*. (A) Gene module network of *Cre01.g048200* showing only *Chlamydomonas* modules sharing at least 10 label co-occurrences. Green edges indicate similarity of modules within the same species, whereas yellow edges indicate overlapping modules. (B) Expression profiles of *Cre01.g048200* and *Cre03.g187950* during diurnal cycle (12h light/12h dark). The red line indicates the shift from light to dark. Full expression profiles can be viewed on the respective gene pages. (C) Network comparator result showing the similarity of *Cre01.g048200* and *Cre03.g187950* networks, with the query genes nodes enlarged. The legend provides label names and their corresponding functions in brackets. The corresponding full modules can be viewed at the respective gene pages at PhytoNet.

Duplicated gene modules

We have previously shown that duplication of co-expression modules is a widespread phenomenon in flowering plants and in early land plants, such as mosses (20,48,53). We exemplify such duplicated transcriptional modules in algae with *Chlamydomonas reinhardtii* gene *Cre01.g048200.t1.1* as a bait gene (<http://aranet.mpimp-golm.mpg.de/responder.py?name=gene!Cre!10456>). This gene is annotated as a putative RNA helicase and contains

'AAA' Pfam domains, which correspond to 'ATPases associated with diverse cellular activities'. The gene module network for this gene revealed six highly similar modules that were overlapping with each other (Figure 3A). Of these six modules, we selected *Cre03.g187950* for comparison with *Cre01.g048200*, as this showed the highest number of gene families and Pfam domains (expressed as label co-occurrences) (21) in common with our bait gene.

The expression patterns of these two genes are distinct, as *Cre03.g187950* is expressed during the whole diurnal cycle,

whereas *Cre01.g048200* is induced at the end of the day with again decreasing expression during the night (Figure 3B). The network comparison of these two genes revealed common Pfam domains, such as 'DNA_ligase_A', 'DNA_pol_A', 'Helicase_C', and 'DNA_pol_A.exo' (Figure 3C), which are associated to DNA repair, replication and proofreading, as well as several DNA/chromatin-associated Pfam domains, such as 'Bromodomain', and 'SNF_2'. To discover the putative function of the two modules, we further analyzed the GO term enrichment of their respective individual gene networks. Interestingly, we found significant enrichment for genes involved in DNA repair in both modules and for DNA replication in the module that is induced at the end of the day (Supplementary Tables S2 and S3, highlighted in yellow). In synchronized cell cultures under a 12 h light/dark regime, *Chlamydomonas reinhardtii* cells divide with the onset of the night (54). As the DNA replication occurs shortly before, the module that is induced at the end of the day might be involved in DNA mismatch repair during DNA replication. Conversely, the constitutively expressed module might function as a general DNA repair module. Taken together, our analyses identified putatively duplicated DNA repair modules with distinct expression patterns and specialized functions. It is important to note that further experimental testing is required to verify the involvement of these modules in DNA repair. This example thus illustrates how hypotheses on functions of modules can be generated as a starting point for additional analyses.

CONCLUSION

To remedy the lack of transcriptome analysis tools for phytoplankton, we introduce PhytoNet, an extension of the PlaNet database which adds cyanobacteria and eukaryotic algae to the web server, thus increasing the number of species to 19. The web server contains gene expression profile plots, gene co-expression networks and Gene Ontology enrichment analyses for gene neighborhoods and clusters, which can be used to predict gene functions. Furthermore, the comparative features of PlaNet allow detection of conserved and duplicated gene modules, allowing rapid identification of functionally related gene modules across and within phytoplankton and land plants. Therefore, PhytoNet is a versatile and easy to use hypothesis generation server for phytoplankton researchers who study gene functions in these organisms.

DATA AVAILABILITY

The expression matrices, co-expression networks, functionally enriched clusters and conserved/duplicated modules are found at <http://aranet.mpimp-golm.mpg.de/download.html>.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Max Planck Society (to M.M., C.R., C.F.); ERA-CAPS grant EVOREPRO (to S.P.). Funding for open access charge: Nanyang Technological University.

Conflict of interest statement. None declared.

REFERENCES

1. Azam, F. and Malfatti, F. (2007) Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.*, **5**, 782–791.
2. Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E. and Keeling, P.J. (2015) Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science*, **347**, 1257594.
3. Field, C.B., Behrenfeld, M.J., Randerson, J.T. and Falkowski, P. (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, **281**, 237–240.
4. Archibald, J.M. (2015) Endosymbiosis and eukaryotic cell evolution. *Curr. Biol.*, **25**, R911–R921.
5. Chang, C., Bowman, J.L. and Meyerowitz, E.M. (2016) Field guide to plant model systems. *Cell*, **167**, 325–339.
6. Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, J., Witman, G.B., Terry, A., Salamov, A., Fritz-laylin, L.K., Maréchal-drouard, L. *et al.* (2010) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*, **318**, 245–250.
7. Graham, L.E., Wilcox, L.W. and Knack, J.J. (2015) Why we need more algal metagenomes. *J. Phycol.*, **51**, 1029–1036.
8. Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R.P. *et al.* (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*, **456**, 239–244.
9. Read, B.A., Kegel, J., Klute, M.J., Kuo, A., Lefebvre, S.C., Maumus, F., Mayer, C., Miller, J., Monier, A., Salamov, A. *et al.* (2013) Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature*, **499**, 209–213.
10. Skeffington, A.W. and Scheffel, A. (2018) Exploiting algal mineralization for nanotechnology: bringing coccoliths to the fore. *Curr. Opin. Biotechnol.*, **49**, 57–63.
11. Rhee, S.Y. and Mutwil, M. (2014) Towards revealing the functions of all genes in plants. *Trends Plant Sci.*, **19**, 212–221.
12. Vandepoele, K., Van Bel, M., Richard, G., Van Landeghem, S., Verhelst, B., Moreau, H., Van de Peer, Y., Grimsley, N. and Pignoneau, G. (2013) pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. *Environ. Microbiol.*, **15**, 2147–2153.
13. Stuart, J.M., Segal, E., Koller, D. and Kim, S.K. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **302**, 249–55.
14. Yu, H., Luscombe, N.M., Qian, J. and Gerstein, M. (2003) Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet.*, **19**, 422–427.
15. Persson, S., Wei, H., Milne, J., Page, G.P. and Somerville, C.R. (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 8633–8638.
16. Itkin, M., Heinig, U., Tzfadia, O., Bhidi, A.J., Shinde, B., Cardenas, P.D., Bocobza, S.E., Unger, T., Malitsky, S., Finkers, R. *et al.* (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*, **341**, 175–179.
17. Proost, S. and Mutwil, M. (2016) Tools of the trade: Studying molecular networks in plants. *Curr. Opin. Plant Biol.*, **30**, 143–150.
18. Lee, T., Yang, S., Kim, E., Ko, Y., Hwang, S., Shin, J., Shim, J.E., Shim, H., Kim, H., Kim, C. *et al.* (2015) AraNet v2: an improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species. *Nucleic Acids Res.*, **43**, D996–D1002.
19. 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.
20. Ruprecht, C., Vaid, N., Proost, S., Persson, S. and Mutwil, M. (2017) Beyond genomics: studying evolution with gene coexpression networks. *Trends Plant Sci.*, **22**, 298–307.
21. Ruprecht, C., Mendrinna, A., Tohge, T., Sampathkumar, A., Klie, S., Fernie, A.R., Nikoloski, Z., Persson, S. and Mutwil, M. (2016) FamNet: a framework to identify multiplied modules driving pathway expansion in plants. *Plant Physiol.*, **170**, 1878–1894.

22. Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T. *et al.* (2015) ArrayExpress update-simplifying data submissions. *Nucleic Acids Res.*, **43**, D1113–D1116.
23. Ballouz, S., Verleyen, W. and Gillis, J. (2015) Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics*, **31**, 2123–2130.
24. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted gene co-expression network analysis. *BMC Bioinformatics*, **9**, 559.
25. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
26. Proost, S., Krawczyk, A. and Mutwil, M. (2017) LSTrAP: efficiently combining RNA sequencing data into co-expression networks. *BMC Bioinformatics*, **18**, 444.
27. Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A. *et al.* (2015) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
28. Van Bel, M., Proost, S., Van Neste, C., Deforce, D., Van de Peer, Y. and Vandepoele, K. (2013) TRAPID: An efficient online tool for the functional and comparative analysis of de novo RNA-Seq transcriptomes. *Genome Biol.*, **14**, R134.
29. Barabási, A.-L. and Bonabeau, E. (2003) Scale-free networks. *Sci. Am.*, **288**, 60–69.
30. Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S. and Provart, N.J. (2009) Co-expression tools for plant biology: Opportunities for hypothesis generation and caveats. *Plant Cell Environ.*, **32**, 1633–1651.
31. Langfelder, P., Horvath, S., Fisher, R., Zhou, X., Kao, M., Wong, W., Steffen, M., Petti, A., Aach, J., D'haeseleer, P. *et al.* (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.
32. Mutwil, M., Usadel, B., Schütte, M., Loraine, A., Ebenhöf, O. and Persson, S. (2010) Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. *Plant Physiol.*, **152**, 29–43.
33. Proost, S. and Mutwil, M. (2017) PlaNet: Comparative Co-Expression network analyses for plants. In: *Methods in Molecular Biology (Clifton, N.J.)*. Vol. **1533**, pp. 213–227.
34. Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.Y. and Stitt, M. (2004) MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.*, **37**, 914–939.
35. Kwak, M., Park, W.K., Shin, S.E., Koh, H.G., Lee, B., Jeong, B., Ryoo, L. and Chang, Y.K. (2017) Improvement of biomass and lipid yield under stress conditions by using diploid strains of *Chlamydomonas reinhardtii*. *Algal Res.*, **26**, 180–189.
36. Rodolfi, L., Zittelli, G.C., Bassi, N., Padovani, G., Biondi, N., Bonini, G. and Tredici, M.R. (2009) Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol. Bioeng.*, **102**, 100–112.
37. Ördög, V., Stirk, W.A., Bálint, P., van Staden, J. and Lovász, C. (2012) Changes in lipid, protein and pigment concentrations in nitrogen-stressed *Chlorella minutissima* cultures. *J. Appl. Phycol.*, **24**, 907–914.
38. Gao, X., Zhang, F., Hu, J., Cai, W., Shan, G., Dai, D., Huang, K. and Wang, G. (2016) MicroRNAs modulate adaption to multiple abiotic stresses in *Chlamydomonas reinhardtii*. *Sci. Rep.*, **6**, 38228.
39. Gargouri, M., Park, J.J., Holguin, F.O., Kim, M.J., Wang, H., Deshpande, R.R., Shachar-Hill, Y., Hicks, L.M. and Gang, D.R. (2015) Identification of regulatory network hubs that control lipid metabolism in *Chlamydomonas reinhardtii*. *J. Exp. Bot.*, **66**, 4551–4566.
40. Ho, S.H., Nakanishi, A., Kato, Y., Yamasaki, H., Chang, J.S., Misawa, N., Hirose, Y., Minagawa, J., Hasunuma, T. and Kondo, A. (2017) Dynamic metabolic profiling together with transcription analysis reveals salinity-induced starch-To-lipid biosynthesis in alga *Chlamydomonas* sp. *JSC4. Sci. Rep.*, **7**, 45471.
41. Meyer, P., Prodromou, C., Liao, C., Hu, B., Roe, S.M., Vaughan, C.K., Vlastic, I., Panaretou, B., Piper, P.W. and Pear, L.H. (2004) Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. *EMBO J.*, **23**, 511–519.
42. Harrison, C.J., Hayer-Hartl, M., Di Liberto, M., Hartl, F.U. and Kuriyan, J. (1997) Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science (80-)*, **276**, 431–435.
43. Zolkiewski, M. (2006) A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. *Mol. Microbiol.*, **61**, 1094–1100.
44. Palmieri, F. (1994) Mitochondrial carrier proteins. *FEBS Lett.*, **346**, 48–54.
45. Lührink, J., Samuelsson, T. and De Gier, J.W. (2001) YidC/Oxa1p/Alb3: Evolutionarily conserved mediators of membrane protein assembly. *FEBS Lett.*, **501**, 1–5.
46. Zarrineh, P., Sánchez-Rodríguez, A., Hosseinkhan, N., Narimani, Z., Marchal, K. and Masoudi-Nejad, A. (2014) Genome-scale co-expression network comparison across *Escherichia coli* and *Salmonella enterica* serovar Typhimurium reveals significant conservation at the regulon level of local regulators despite their dissimilar lifestyles. *PLoS One*, **9**, e102871.
47. Gerstein, M.B., Rozowsky, J., Yan, K.-K., Wang, D., Cheng, C., Brown, J.B., Davis, C.A., Hillier, L., Sisu, C., Li, J.J. *et al.* (2014) Comparative analysis of the transcriptome across distant species. *Nature*, **512**, 445–448.
48. Ruprecht, C., Mutwil, M., Saxe, F., Eder, M., Nikoloski, Z. and Persson, S. (2011) Large-Scale Co-Expression approach to dissect secondary cell wall formation across plant species. *Front. Plant Sci.*, **2**, 1–13.
49. Mutwil, M., Klie, S., Tohge, T., Giorgi, F.M., Wilkins, O., Campbell, M.M., Fernie, A.R., Usadel, B., Nikoloski, Z. and Persson, S. (2011) PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *Plant Cell*, **23**, 895–910.
50. Park, C.Y., Wong, A.K., Greene, C.S., Rowland, J., Guan, Y., Bongo, L.A., Burdine, R.D. and Troyanskaya, O.G. (2013) Functional knowledge transfer for high-accuracy prediction of under-studied biological processes. *PLoS Comput. Biol.*, **9**, e1002957.
51. Tzfadia, O., Diels, T., De Meyer, S., Vandepoele, K., Aharoni, A. and Van de Peer, Y. (2016) CoExpNetViz: comparative co-expression networks construction and visualization tool. *Front. Plant Sci.*, **6**, 1194.
52. Ruprecht, C., Mendrinna, A., Tohge, T., Sampathkumar, A., Klie, S., Fernie, A.R., Nikoloski, Z., Persson, S. and Mutwil, M. (2016) Famnet: a framework to identify multiplied modules driving pathway expansion in plants. *Plant Physiol.*, **170**, 1878–1894.
53. Sibout, R., Proost, S., Hansen, B.O., Vaid, N., Giorgi, F.M., Ho-Yue-Kuang, S., Legée, F., Cézart, L., Bouchabké-Coussa, O., Soulhat, C. *et al.* (2017) Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in *Brachypodium distachyon*. *New Phytol.*, **215**, 1009–1025.
54. Spudich, J.L. and Sager, R. (1980) Regulation of the *Chlamydomonas* cell cycle by light and dark. *J. Cell Biol.*, **85**, 136–145.