bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Proteomic changes orchestrate metabolic acclimation of a unicellular diazotrophic cyanobacterium during light-dark cycle and nitrogen fixation states

Punyatoya Panda[1], Swagarika J. Giri[2], Louis A. Sherman[3], Daisuke Kihara[2,3*], and Uma K. Aryal[1,4*]

[1]Department of Comparative Pathobiology, Purdue University, West Lafayette, IN 47907

[2]Department of Computer Science, Purdue University, West Lafayette, IN 47907

[3]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907

[4]Purdue Proteomics Facility, Bindley Bioscience Center, Purdue University, West Lafayette, IN 47907

*Correspondence should be addressed to Daisuke Kihara: dkihara@purdue.edu and Uma K Aryal: uaryal@purdue.edu

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Cyanobacteria have developed an impressive array of proteins and pathways, each tailored for specific metabolic attributes, to execute photosynthesis and biological nitrogen (N_2)-fixation. An understanding of these biologically incompatible processes provides important insights into how they can be optimized for renewable energy. To expand upon our current knowledge, we performed label-free quantitative proteomic analysis of the unicellular diazotrophic cyanobacterium *Crocosphaera subtropica* ATCC 51142 grown with and without nitrate under 12-hour light-dark cycles. Results showed significant shift in metabolic activities including photosynthesis, respiration, biological nitrogen fixation (BNF), and proteostasis to different growth conditions. We identified 14 nitrogenase enzymes which were among the most highly expressed proteins in the dark under nitrogen-fixing conditions, emphasizing their importance in BNF. Nitrogenase enzymes were not expressed under non nitrogen fixing conditions, suggesting a regulatory mechanism based on nitrogen availability. The synthesis of key respiratory enzymes and uptake hydrogenase (HupSL) synchronized with the synthesis of nitrogenase indicating a coordinated regulation of processes involved in energy production and BNF. Data suggests alternative pathways that cells utilize, such as oxidative pentose phosphate (OPP) and 2-oxoglutarate (2-OG) pathways, to produce ATP and support bioenergetic BNF. Data also indicates the important role of uptake hydrogenase for the removal of O_2 to support BNF. Overall, this study expands upon our knowledge regarding molecular responses of *Crocosphaera* 51142 to nitrogen and light-dark phases, shedding light on potential applications and optimization for renewable energy.

Introduction

Cyanobacteria are diverse photosynthetic organisms that play a vital role in harvesting solar energy (1). Their ability to produce oxygen contributed to the oxygenation of the Earth's atmosphere that not only made the Earth hospitable for other life forms that rely on oxygen but also allowed for more complex organisms to evolve (2) (3). The ability to consume CO_2, harvest solar energy, and convert it into chemical energy allowed them to succeed in an otherwise hostile environment. Consequently, cyanobacteria provide a powerful solution for carbon-neutral energy production, carbon sequestration, and a wealth of targets for metabolic engineering of energy-rich biomolecules (4). Among many unicellular diazotrophs, *Crocosphaera subtropica* (formerly known as *Cyanothece* sp. ATCC 51142) have evolved diurnal rhythms and biological nitrogen fixation (BNF), a process sensitive to oxygen (5-7), making it interesting and metabolically versatile cyanobacteria yet studied. It can maintain an anoxic cytoplasmic environment to prevent

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

oxygen toxicity of the nitrogenase enzymes which are responsible for BNF. The genome indicates a wealth of metabolic potential, in addition to highly active photosynthesis and CO_2-uptake mechanisms (8). Importantly, it can perform BNF either in the dark or during a prolonged continuous light phase after entrainment of light-dark growth (9, 10) and has evolved as an attractive model to study processes such as photosynthesis, BNF, and carbon sequestration as well as for understanding their circadian and diurnal rhythms (11, 12).

The presence or absence of light and nutrition (carbon and nitrogen), trigger a cascade of gene expression, protein translation, signaling, and shifts in membrane organization that lead to the functional changeover from photosynthesis to BNF, and *vice-versa*. In tandem with these metabolically exclusive processes are a host of other time-dependent processes, which the cell switches based on its internal mechanisms as if anticipating the light-dark cycle, rather than reacting to it (9, 13). Under nitrogen-fixing condition, the cells become filled with large granules between the photosynthetic membranes (9, 10). These granules, known as glycogen granules, contain semi-amylopectin, essential to generate energy through their degradation during BNF (14). The composition and changes of enzyme activity between light-dark transitions for these glycogen granules are still outstanding. An understanding of the mechanistic processes that link between sensing light-dark transitions, protein translation, membrane organization, and signaling provides key insights into how *Crocosphaera* 51142 might be optimized for enhanced renewable energy production. The expression of nitrogenase is limited to nitrate-depleted growth, and in the absence of an external carbon source, nitrogenase expression is strictly restricted to the dark cycle (15-17). However, when cells are grown in the presence of sufficient carbon, such as glycerol, nitrogenase expression occurs during both the light and dark cycles, along with other enzymes related to respiration, glycogen metabolism, and glycolytic/pentose phosphate pathways (15). These observations suggest that enhanced respiration during the light cycle, attributed to excess carbon sources, may create an anoxic cellular condition, and this condition, in turn, facilitates the expression of nitrogenase.

Nitrogenase enzymes are rapidly inactivated or degraded upon exposure to O_2 (18, 19). *Crocosphaera* 51142 metabolic machinery is cycling daily between O_2 production in photosynthesis during the day and oxygen scavenging during respiration and BNF at night. Under subjective light, cells also show distinct circadian rhythms of photosynthesis and BNF with peaks every 24 hours. Interestingly, under subjective light, the capacity for photosynthesis is reduced during the period of BNF (20), even though light energy is available. It is possible that the circadian rhythms and transitions between photosynthesis and BNF are correlated not only with intense respiration and nitrogenase activity but also with redox state of the cell through the action of

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

thioredoxin (21) and activation or inhibition of PSI and PSII complexes (20). While previous transcriptomic and proteomic studies have shed light on certain aspects of the regulation of this desperate metabolic process, there is still much to be unmasked about the intricate interplay between nitrate, light-dark cycles, and metabolic processes in cyanobacteria. Proteomic technologies are also constantly improving for greater sensitivity, resolution, and accuracy, allowing in-depth analysis. Taking advantage of such new concepts and improved technologies, we performed quantitative mass spectrometry analysis of *Crocosphaera subtropica* ATCC51142 to determine mechanisms underlying shift in cellular metabolism in response to BNF and light-dark transitions. Particularly, we sought to link changes in protein abundances and pathways to these conditions to dig deeper into the complex relationship between CO_2 fixation, respiration, and glycogen metabolism that influence temporal separation of photosynthesis and BNF in cyanobacteria.

Materials and Methods

Cell growth and experimental conditions

Crocosphaera sp. subtropica ATCC 51142 stocks were maintained in ASP2 medium with 17.6 mM $NaNO_3$ and 30 µmol photons m^{-2} s^{-1} of continuous light. One-twentieth (5ml) of the *Crocosphaera* 51142 stock was first inoculated to each 250-ml flask containing 100 mL ASP2 medium with or without 17.6mM $NaNO_3$ and allowed to grow for 7 days on a shaker at 125 rpm, 30°C, and 30 µmol photons m^{-2} s^{-1} of continuous light. After seven days, cultures were transitioned to alternative 12 h-light/12 h-dark (30 µmol photons m^{-2} s^{-1}) diurnal cycles and grown for additional seven days before harvesting. Cells were harvested at 6 h into the light period (L) or 6 h into the dark period (D).

Fifteen (15) ml of cell cultures were collected by centrifugation at 3,220 × g in 15 ml tubes (Corning), washed with 1 ml of 50 mM HEPES-KOH buffer (pH7.5), and then centrifuged again at 10,000 × g at 4°C for 15 min to collect cell pellets. The pellets were resuspended in 200 ul of HEPES/KOH (pH7.8) buffer, supplemented with 1 mM phenylmethylsulfonylfluoride (PMSF) protease inhibitor and homogenized in Precellys VK 0.5 tubes (Bertin Corp., Rockville, MD, USA), 3× at 6000 rpm for 3 x 20 s in each cycle followed by probe sonication. Protein concentration was determined by bicinchoninic acid (BCA) assay (Pierce Chemical Co., Rockford, IL, USA). Following BCA assay, cell lysates corresponding to 200 µg of total protein (equivalent volumes) were ultracentrifuged at 150,000 × g for 20 min to divide proteins into soluble and insoluble fractions. The soluble fractions were acetone precipitated with four volumes of cold acetone and incubated overnight at -20°C, while the insoluble pellets were resuspended in 200µl of

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

HEPES/KOH (pH 7.5), bath sonicated for 5 mins and then acetone precipitated with four volumes of cold (-20°C) acetone. The soluble and insoluble protein pellets were collected by centrifuging at 17,200 × g for 20 min at 4°C, washed 3× with 80% cold (-20°C) acetone and prepared for LC-MS/MS analysis as described below.

Protein extraction and proteolysis

Both soluble and insoluble pellets were resuspended in 20µl buffer of 8M urea and solubilized by incubating at room temperature with continuous vortexing for an hour, and 0.1% Rapigest (Waters, MA) was added to the insoluble pellets only with the urea solution. The samples were reduced with 10mM dithiothreitol (DTT) at 37°C for 45 min and then cysteines alkylated with iodoethanol mix (195 uL acetonitrile, 4 uL iodoethanol and 1 ul triethyl phosphine) for 45min at 37°C in a dark. After reduction and alkylation, samples were dried in a vacuum centrifuge (Vacufuge Plus, Eppendorf, Enfield, CT) at 45°C, reconstituted in 150 ul of 50 mM ammonium bicarbonate and then digested with trypsin at a 1:25 enzyme to substrate ratio. High pressure digestion was performed using a Barocycler (Pressure Bioscience INC., Easton, MS, USA) at 50°C with 60 cycles, each cycle consisting of 50s at 20,000 PSI and 10s at 1 atm) as described before (22, 23). Digested peptides were cleaned using Pierce Peptide Desalting Spin Columns (Thermo Fisher Scientific, Waltham, MA, USA). Eluted clean peptides were dried in a vacuum centrifuge and reconstituted in 20 µL 0.1% formic acid (FA) in 3% acetonitrile. The peptide concentration was measured using a nanodrop spectrophotometer (ThermoFisher Scientific), adjusted the final concentration in each sample to 1 µg/µl, and 1 µg (1µl) was used for proteomics analysis of the highest concentrated fraction of each sample (either soluble or insoluble), and 0.5ug was injected for the lower concentrated fraction.

Liquid chromatography tandem mass spectrometry analysis

One µg (1 µl) and 0.5 µg (of the other fraction) of clean peptides were analyzed by reverse-phase HPLC separation using a Dionex UltiMate 3000 RSLC nano system, coupled to a Orbitrap Fusion Lumos mass spectrometer via Nanospray Flex™ electrospray ionization source (Thermo Fisher Scientific) as described previously (23, 24). Briefly, peptides were first loaded into a PepMap C18 trap column (3µm × 75µm ID × 2 cm) (Thermo Fisher Scientific, Waltham, MA, USA), and then separated using a reverse phase 1.7 µm 120 Å IonOptics Aurora Ultimate C18 column (75 µm x, 25cm). The column was maintained at 50 °C, mobile phase solvent A was 0.1% FA in water, solvent B was 0.1% FA in 80% ACN. The loading buffer was 0.1% FA in 2% ACN. Peptides were loaded into the trap column for 5 min at 5 µl/min, then separated with a flow rate of 400 nl/min

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

using a 130 min linear gradient. The concentration of mobile phase B was increased linearly to 8% in five minutes, 27% B in 80 min, and then 45% B at 100 min. After 100 min, it was subsequentially increased to 100% of B at 105 min and held constant for another 7 min before reverting to 2% of B in 112.1 min and maintained at 2% B until the end of the run. The mass spectrometer was operated in positive ion and standard data dependent acquisition (DDA) mode. The spray voltage was set at 2.8 kV, the capillary temperature was 320 °C and the S-lens RF was set at 50. The resolution of Orbitrap mass analyzer was set to 60,000 and 15,000 at 200 *m/z* for MS1 and MS2, respectively, with a maximum injection time of 100 ms for MS1 and 20 ms for MS2. The full scan MS1 spectra were collected in the mass range of 350–1600 *m/z* and the MS2 first fixed mass was 100 *m/z*. The automatic gain control (ACG) target was set to 3×10^6 for MS1 and 1×10^5 for MS2. The fragmentation of precursor ions was accomplished by higher energy C-trap collision dissociation (HCD) at a normalized collision energy setting of 27% and an isolation window of 1.2 *m/z*. The DDA settings were for a minimum intensity threshold of 5×10^4 and a minimum AGC target of 1×10^3. The dynamic exclusion was set at 15 s and accepted charge states were selected from 2 to 7 with 2 as a default charge. The exclude isotope function was activated.

LC-MS/MS data analysis

LC–MS/MS data were processed with MaxQuant software (Ver 2.0.3.0) (25, 26). Raw spectra were searched against the *Crocosphaera* 51142 protein sequence database obtained from the UniProt (downloaded in July 2022) containing 5403 protein sequences, for protein identification and MS1 based label-free quantitation. The minimum length of the peptides was set at six AA residues in the database search. The following parameters were edited for the searches: precursor mass tolerance was set at 10 ppm, MS/MS mass tolerance was set at 20 ppm, enzyme specificity of trypsin/Lys-C enzyme allowing up to 2 missed cleavages, oxidation of methionine (M) as a variable modification and iodoethanol of cysteine as a fixed modification. The decoy reverse database was considered for data analysis and to control false discovery rate (FDR) and was set at 0.01 (1%) both for peptide spectral match (PSM) and protein identification. The unique plus razor peptides (non-redundant, non-unique peptides assigned to the protein group with most other peptides) were used for peptide quantitation. Only proteins detected with at least one unique peptide and MS/MS ≥ 2 (spectral counts) were considered valid identifications. Label-free quantitation (LFQ) intensity values were used for relative protein abundance comparisons.

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bioinformatics data analysis

We mainly performed data analysis in Perseus (version 1.6.0.9) (27), Microsoft Excel and data visualized using OriginPro (Version 2022, OriginLab Corporation, Northampton, MA, USA), InteractiVenn (28), Morpheus (https://software.broadinstitute.org/morpheus). Log2-transformed LFQ intensities (protein intensities) were used for further analysis. Coefficients of variations were calculated for raw protein intensities of Hela digest using triplicate runs to determine the reproducibility of LC-MS/MS analysis and label-free quantitation. Data sets were filtered to make sure that identified proteins showed expression in at least three out of four biological replicates of at least one treatment group and the missing values were subsequently replaced by imputation in Perseus that were drawn from a normal distribution. Principal component analysis of treatment effects and biological replicates was performed as described in (29). Multi-sample test (ANOVA) for determining if any of the means of differentiation stages were significantly different from each other was applied to protein data set. For hierarchical clustering and heatmap generation of significant proteins, mean protein abundances of biological replicates were z-scored and clustered using Euclidean as a distance measure for row clustering. Significantly upregulated or downregulated proteins between the treatment groups (± nitrate, L/D, and nitrate × L/D) were determined by ANOVA and a two tallied student's *t*-test. Differentially expressed proteins were determined with p-value <0.05.

Gene Ontology (GO) analysis

We used three sequence-based function prediction methods: PFP(30), Phylo-PFP(31), and Extended Similarity Group method (ESG) (32) to assign Gene Ontology (GO) terms (33) to protein-coding genes. The PFP algorithm scores GO terms based on Expect (E)-values of sequences with those GO terms retrieved from the UniProt sequence database by PSI-BLAST (34). It then propagates the scores to parental terms on the GO Directed Acyclic Graph (DAG) according to the number of database sequences annotated with parent and child terms. Additionally, based on validation results over a set of benchmark sequences, it assigns a confidence score to GO term predictions. Phylo-PFP, which is an improvement over PFP, improves the performance by incorporating phylogenetic information. The ESG method performs iterative sequence database searches and annotates a query sequence with GO terms. Each annotation is given a probability based on how similar it is to other sequences in the protein similarity graph. To ensure the inclusion of meaningful GO term annotations, we focused exclusively on predictions characterized by high confidence, and all GO terms exceeding confidence score cutoffs of 20,000 for PFP, 0.7 for Phylo-PFP, and 0.7 for ESG were incorporated

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

into our analysis. To improve high-confident prediction, we combined results from three prediction methods and presented the consolidated GO term annotations. Each result file has information about Protein ID, GO ID, Depth, Class, and GO Description. The depth refers to the depth of GO ID in the GO DAG, and class refers to GO functional category (f - molecular function, p- Biological process, c- Cellular Component), and GO Description describes the predicted GO term. The Gene Ontology release 2021-11-16 was used for this analysis.

Correlation of proteomics and transcriptomics data

Transcriptomics data was obtained from Stockel et al. (35) and matched with our proteomics data. Like the transcriptomics "pooled control", the proteomics "pooled control" was the average of the intensities of all nitrate-depleted samples. The ratio of D- to the pooled control was then used to calculate the fold change (FC) of proteins. The $\log_2$(FC) of these values were plotted for both the proteomics and transcriptomics data using scatterplots in OriginPro (Version 2022, OriginLab Corporation, Northampton, MA, USA). A threshold of ±1.5 (or log (foldchange = ±0.58) was used to decide if the fold change was significant enough. The data was divided into various categories from the plots, and colored accordingly, such as those indicated in both proteomics and transcriptomics data with a threshold fold change of ±1.5, or in any one of these, or none. Then, the GO enrichment was performed for these various categories to provide insights into the molecular regulation of several proteins of interest, such as those of the nif gene cluster, proteases, and ribosomal proteins.

Results and Discussion

Overview of the proteomic results

The minima and the maxima of O_2 production occurs around 6-8 hours of the dark and the light periods, respectively (20). Therefore, we choose to examine the proteome at six hours into both the light and dark cycles, to capture the proteome responses during the peak or near-peak of the organism's metabolic activities related to photosynthesis and nitrogen fixation.

Cells were grown as described previously (36) and as outlined in Figure 1 (see materials and methods for details). Cells were harvested six hours into the light or into the dark cycles and four biological replicates per treatment group were divided into soluble and insoluble preparations by ultracentrifugation before LC-MS/MS analysis. Tryptic peptides were analyzed in a single injection using a 130-min LC method in DDA mode (see Methods for details). MaxQuant search

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

resulted in the identification of 29,946 peptides (Supplementary Table S1) and 2688 proteins (Supplementary Table S2), representing ~40% of the proteome. Data were filtered to retain a high confidence set of proteins (proteins identified at minimum three of the four replicates in at least one treatment group with LFQ intensity >0 and MS/MS counts >2). This resulted in the identification of 2050 proteins, which were used for downstream bioinformatic and statistical analysis. Of the 2050 proteins, 1321 proteins were shared across all groups and the highest number of proteins were identified under nitrogen-fixing condition in the dark (Figure 2A). Principal component analysis (PCA) showed proteins clustering separately based on treatments with very high concordance between replicates (Figure 2B). In ANOVA comparison, 1370 proteins (*adj. p* value ≤ 0.05) changed due to nitrate, 557 proteins changed due to light and dark, and 361 proteins changed due to the combination of both nitrate and light (Figure 2C, Supplementary Table S3). Hierarchical clustering of significant proteins (ANOVA, *p*-value < 0.05) revealed largest proteome differences in response to nitrate (Supplementary Figure 1A). Nitrogenases were among the most upregulated proteins and the flavoproteins (cce_3835; cce_3833) and heme oxygenase (Ho1; cce_2573), were among the most downregulated proteins in the dark under nitrogen-fixing conditions. The data represented the key proteins in each of the major functional categories that provided important insights about the major metabolic changes during L/D cycles and nitrogen-fixation. The largest number of differentially abundant proteins represented cellular processes such as nitrogen-fixation, photosynthesis and respiration, CO_2 fixation, carbohydrate metabolism, transport, folding, protein translation, and protein degradation. The predicted proteome of *Crocosphaera* 51142 consists of 5269 open reading frames, of which only 34% are of known function, (8). About 20% of proteins identified in this study hypothetical, uncharacterized or unknown function proteins, highlighting the challenges and need to determine the functions of these hypothetical proteins to better understand the cellular metabolism of *Crocosphaera* 51142 under various growth conditions.

To contextualize the differentially regulated proteins across growth conditions, we performed GO enrichment analysis. Figure 3 shows the top enriched GO terms for each treatment condition. Due to the interaction between nitrate and light-dark conditions, nitrogen fixation, molybdenum cofactor binding, tryptophan synthase activity, oxidoreductase activity, and phycobilisome were among the most enriched GO terms. Under the influence of light and dark conditions alone, cellular macromolecule biosynthetic processes, gene expression, primary metabolic processes, and photosynthesis, etc. were enriched. Because of nitrate alone, cellular nitrogen compound biosynthetic process, gene expression, cellular and macromolecular biosynthetic processes were enriched (Supplementary Table S11).

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To determine how the cells grown under nitrogen-fixing growth condition differed from non-fixing growth under L/D cycles, we displayed all quantified proteins as a volcano plots, highlighting the proteins with the highest log transformed fold-change and lowest q-values (Figure 4, A-D). Nitrogenase enzymes together with HupL and PetH were among the top upregulated proteins and CmpA, CmpC, and CmpD were among the top-down regulated proteins in the dark under nitrogen-fixing growth condition compared to non-nitrogen fixing conditions (Figure 4A, Supplementary table S4). Similarly, in the light, PgI, GlpD, MetE and NifH were upregulated in the nitrogen fixing conditions compared to non-nitrogen fixing conditions (Figure 4B, Supplementary Table S5). Comparing the effect of light and dark under nitrogen fixing conditions only, we found that NifK, NifB, Gap were upregulated while Ho1, Ugd and CmpA were downregulated in the dark (D^-) compared to the light (L^-) (Figure 4C; Supplementary table S6). On the other hand, under non-nitrogen fixing conditions, psbA4, aphA, kaiC2, TCA Cycle enzymes like talA and glgA1 were upregulated in the dark, compared to acsF, murA and cmpC that were downregulated in the dark compared to light (Figure 4D, Supplementary table S7).

Differential regulation of proteins involved in respiration and glycogen metabolism.

The cytochrome c oxidase (CoxB1; cce_1977) showed a significant 13-fold higher abundance in the dark under nitrogen-fixing conditions compared to non-fixing conditions (Table 1). In the light cycle, it was 3.8 times. The expression of the nitrogenase enzyme cluster was limited to nitrogen-fixing conditions only (Figure 5A, Table 5B). Nitrogenase expression mirrored with the expression of CoxB1 supporting the potential role of CoxB1 in active respiration to create anoxic condition for nitrogenase activity (15).

Cyanobacteria convert CO_2 and water into carbohydrates, primarily glucose, during photosynthesis to supporting cellular activities during periods of darkness (37). Enzymes responsible for glycogen synthesis; GlgA1 and GlgA2 were more abundant in the light whereas enzymes for glycogen metabolism; GlgP1 and GlgP2 were more abundant in the dark under nitrogen-fixing condition (Table 1). Active glycogen metabolism generates ATP and creates a low oxygen environment for BNF (6, 9). There are two glycogen debranching enzymes, *glgP*, which is found in the β-cyanobacterial clade (the common ancestor of all cyanobacteria that possessed a β-carboxysome) and *glgX*, which is found in the α-cyanobacteria (which possess α-carboxysome) (7, 8, 38, 39). We identified both enzymes but observed different responses to L/D cycles and nitrate. Among the three GlgP isoforms, GlgP1 (cce_1629) was upregulated whereas GlgP2 (cce_5186) and GlgP3 (cce_1603) were downregulated under nitrogen-fixing condition independent of L/D cycles. The GlgX, on the other hand, was upregulated under nitrogen-fixing

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

condition. These data suggest that glycogen accumulation and degradation is carefully regulated in *Crocosphaera* during BNF.

Differential regulation of glycolysis, TCA cycle and OPP pathway enzymes

Glycolytic enzymes showed similar abundances between L/D cycles but exhibited stronger response to nitrate (Table 1, Supplementary Figure 2A, Supplementary Table S3). Glucose-6-phosphate dehydrogenase (*Zwf;* cce_2536) directs carbon to OPP pathway and initiates glucose oxidation concomitantly with the generation of NADPH to provide reductants for BNF (40). This enzyme was upregulated in the dark under nitrogen-fixing condition (Table 1). OPP pathway enzymes, including 6-phosphogluconate dehydrogenase (Gnd, cce_3746), OxPPCycle protein (OpcA, cce_2535), transaldolase (TalA, cce_4686), TalC (cce_4208), and 6-phosphogluconolactonase (Pgl, cce_4743) were also upregulated in the dark under nitrogen-fixing condition compared to other growth conditions (Table1). TalA had higher abundance in the dark than in the light but was unaffected by nitrate. TalC was unaffected by the L/D cycle but showed higher abundance without nitrate than with nitrate (Table 1).

TCA cycle enzymes were also in general, more abundant in the dark than in the light, but the effect of nitrate was more pronounced (Table 1, Supplementary Figure 2A). TCA cycle enzyme, isocitrate dehydrogenase (Icd, cce_3202) which converts isocitrate to 2-oxaloacetate (2-OG), which is utilized for the biosynthesis of glutamate via GS-GOGAT cycle (41), was upregulated under nitrogen-fixing growth (Table 1). The succinate semialdehyde dehydrogenase NADP+ (GabD, cce_4228), was lower in nitrogen-fixing cells than non-fixing cells and was also more abundant in the dark than in the light. GabD participates in glutamate metabolism for the conversion of 2-OG to succinate (42). The expression patterns of both Icd and GabD are in agreement with our previous report (16). Increased expression of *icd* transcripts under nitrogen-depleted condition has been reported in other cyanobacteria (41).

Among other TCA cycle enzymes, expression of aconitase (AcnB, cce_3280) was affected by both nitrate as well as L/D cycles independent to each other (Table 1). Under nitrogen-fixing growth, cyanobacteria preferentially utilize alternative nitrogen sources (43). The ammonia (NH_4^+) produced during BNF can be trapped by GlnA (cce_4432) and converted to glutamate via GS (GlnA)-GOGAT (GlsF) cycle and incorporated into other amino acids through the activities of aminotransferase or transaminase (44). Increased synthesis of GlnA, GlnB (cce_1775), and GlsA (cce_2244) in the dark cycle under nitrogen-fixation suggest utilization of ammonia produced by BNF for protein synthesis. Indeed, upregulation of carbamoyl-phosphate synthase (CarA, 0902) and CarB (cce_2038) in the dark under nitrogen-fixing conditions suggest the conversion of

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

glutamine into carbamoyl-P, which can be then directed as a substrate for purines and pyrimidine synthesis and for urea cycle. This is supported by the observation that orotate phosphoribosyltransferase (PyrFE, cce_0502) was >2.8 times higher and CTP synthase (PyrG, cce_2923) was >9.3 times higher in the dark under nitrogen-fixing conditions compared to non-fixing conditions (Table 1). ArgB (cce_3224), ArgG (cce_4370), and ArgF (cce_3251), which catalyze reactions to convert carbamoyl-P to citrulline and argininosuccinate through urea cycle were also upregulated in the dark cycle. These results suggest that *Crocosphaera* 51142 not only utilize traditional glycolysis and TCA cycle but also various alternative or modified pathways to optimize its metabolic activities.

Expression of nitrogenase and hydrogenase enzyme clusters

The acquisition of the nitrogenase complex is a significant evolutionary event that evolved and shaped nitrogen cycling in the ecosystem. *Crocosphaera* 51142 consists of a 35 *nif* genes cluster within a single contiguous region of DNA separated by no more than 3 kb (7). We focused on their expression six hours into the L/D cycles and identified 14 annotated nitrogenase enzymes within the *nif* gene clusters that were differentially regulated and were all exclusively expressed under nitrogen-fixing condition, with actual loss of their identification under non-fixing conditions (Figure 5A, Table 5B). Under nitrogen-fixing growth, while all the nitrogenase enzymes were identified in both the light and the dark cycles, their abundances were significantly higher in the dark than in the light. The level of abundance changes was remarkable with fold changes ranging from 7-fold for NifN to 450-fold for NifH (Table 4B). Such a high level of dynamic range of nitrogenase expression has not been seen in previous proteomic studies. The structural proteins NifHDK were among the most upregulated proteins followed by NifT (cce_0547), and NifW (cce_0568). The 2Fe-2S putative N_2-fixation related protein (cce_0571) was also one of most highly expressed proteins in the dark with 134-fold higher abundance without nitrate than with nitrate (Figure 5A, Table 5B).

The significant accumulation of NifT and NifW is interesting. While specific function of NifT is unknown (45, 46), this protein is believed to be involved in transporting ammonia produced by nitrogenase within the cell. Previous studies in other strains has revealed that deletion of NifT did not affect diazotrophic growth (47, 48), however nitrogenase activity was higher in the mutant lines than that of the wild type (46), indicating some suppressive effect of NifT on nitrogenase activity. More biochemical studies are needed to determine the functions of NifT in cyanobacteria. NifW is a small nitrogenase stabilizing and protective protein with 116 amino acid residues (13.6 kDa), located within the 35 nitrogenase gene cluster, was 148-fold higher abundance in the dark

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cycle without nitrate compared to with nitrate. Previous studies have shown higher level of expression of this enzyme towards the end of the dark transition, thus its high abundance six hours into the dark period in the current study is in conformity with its late expression pattern. The FeMo cofactor biosynthesis protein NifB (cce_0554) is a sensitive target for assessing the regulation of BNF (49) and was more than 43-times higher abundance in the dark without nitrate compared to with nitrate. However, its expression was not different in the light cycle between nitrate and non-nitrate growth. The *hesA* (cce_0569) and *hesB* (cce_0570) genes are conserved in diazotrophic cyanobacteria and are within the *nif* gene cluster (46). Proteins encoded by these genes were 3-fold and 30-fold more abundant in the dark under nitrogen-fixing condition compared to N_2-sufficient conditions (Figure 5A, Table 5B). While their actual function in nitrogenase activity is unknown, recent study indicates their involvement in the efficient production of the MoFe protein (46).

Crocosphaera 51142 contains *hupSL* genes for an uptake hydrogenase, and these genes are induced in the dark under nitrogen-fixing growth (15, 50). The small subunit HupS mediates electron transport from the active site of the large subunit (HupL) to redox partners and downstream reactions through a set of Fe-S clusters and the HupL harbors the active site that contains the Ni-Fe (37). The uptake hydrogenase recycle the H_2 produced as a byproduct during BNF by the nitrogenase activity (51), but it may have other functions such as removing O_2 from nitrogenase, thereby protecting it from inactivation; and supply reducing equivalents (electrons) to nitrogenase and other enzymes (51, 52). The HupSL enzymes were indeed highly expressed in the dark with HupL showing 76 times higher abundance and HupS showing 8 times higher abundance in the dark under nitrogen-fixing condition than with non-fixing conditions (Figure 5B). These results indicate that the regulation of uptake hydrogenase is influenced by both the absence of light and the absence of nitrate. The gene cluster *kaiABC* encodes circadian clock proteins. Among them, KaiA, KaiC1 and KaiC2 were upregulated under nitrogen-fixing growth compared to non-fixng growth in both the L/D cycles. Noticeably, KaiC2 expression was almost 9-fold higher during the light without nitrate than with nitrate (Supplementary Figure 2B).

Photosynthesis and CO_2 fixation

The CO_2 fixation primarily involves the enzyme Rubisco during the Calvin cycle that results in the synthesis of carbohydrates. As the name goes, Rubisco can also catalyze a reaction with oxygen (O_2), leading to a process known as photorespiration. Reaction with O_2 rather than CO_2 results in loss of some fixed carbon as CO_2, loss of NH_3, as well as consumption of ATP, thus decreasing

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

photosynthesis efficiency (53). Catalytic activity of Rubisco is influenced by factors such as the concentration of CO_2 and the competing oxygenation reaction (54).

Carbonic anhydrase (IcfA1, cce_2257) was upregulated in the light cycle under nitrogen-fixing condition. Bicarbonate transport system substrate binding proteins (CmpA, CmpB, CmpC, and CmpD) were also upregulated in the light cycle, however, unlike IcfA1, they were downregulated in the absence of nitrate (Figure 4C, Table 1). Sodium dependent bicarbonate transporter (SbtA, cce_2939) was downregulated in the dark compared to the light in both nitrogen-fixing as well as non-fixing conditions without nitrate as well as with nitrate. These observations agree with our previous results except SbtA, which was shown to have higher abundance in the dark (16, 55). The Rubisco large subunit (CbbL) was affected by L/D cycles but not by nitrate, however, the small subunit (RbcS) was affected by both the L/D and nitrate. RbcS was down under nitrogen-fixing condition. The expression of glycolate oxidase (cce_3708), on the other hand, was dependent on nitrate not on the L/D cycles. These observations suggest active carboxysome and that activity was influenced by the presence or absence of nitrate.

PSI and PSII proteins were down in the dark compared to the light, and again down without nitrate compared to its presence (Supplementary Table S3). D1 proteins, PsbA1/ PsbA2 were grouped together and showed > 8-fold higher in the dark under nitrogen-fixing conditions than under non-fixing conditions (Table 1). They were >6-fold more abundant in the light under the same nitrate conditions. The effect of L/D cycle on these D1 proteins was minimal. Another PsbA2 protein (cce_3411) and PsbA4 (cce_3477), were higher in the dark compared to the light under both the nitrogen-fixing and non-fixing conditions (Table 1). Higher PsbA4 levels in the dark period have been reported in a previous study (55). The phycobilisome complex was upregulated mostly in the light cycle but was also strongly influenced by nitrate with significant downregulation under nitrogen-fixing conditions. The cytochrome b_6f complex involved in electron transport chain, linking the PSII and PSI (56) was upregulated in the dark under nitrogen-fixing conditions but was constant between LD cycle under non-fixing conditions.

Translation, folding, degradation and cellular homeostasis.

Multiple metabolic forces including protein synthesis, chaperones, detoxifying proteins, and proteases play crucial roles in maintaining proteome homeostasis in cells (57). Unlike ubiquitin proteasome system in eukaryotes, proteostasis in cyanobacteria is regulated by protein quality control (PQC) system depending on AAA+ proteolytic machines (58). We were interested to determine any changes in proteins involved in these biological processes under L/D cycles and nitrogen. The majority of the identified 30S and 50S ribosomal proteins were downregulated under

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nitrogen-fixing conditions (Figure 6A, B). Protein homeostasis is tightly controlled by chaperones and proteases (59). The small heat shock protein, HSP20 (HspA3; cce_5270), was 3.5-fold more abundant in the dark and 3.0-fold more abundant in the light when compared between nitrogen-fixing and non-fixing conditions. The chaperone DnaK2 (cce_4004) also showed a significant >7.5-fold higher abundance in the dark and >4.5-fold higher abundance in the light when compared between nitrogen-fixing and non-fixing conditions (Table 1). Nitrogen, not the L/D cycle, had a significant impact on the chaperonin GroEL (cce_1344 and cce_3314). While specific roles of chaperones and heat shock proteins in cyanobacteria have not been thoroughly investigated, increased abundances of HSP20 and GroEL under nitrogen-fixing conditions may suggest cellular stress responses.

Protease activities must be tightly regulated in a biological system, particularly in unicellular diazotrophic cyanobacteria, because of their constant shift in metabolic activities between L/D transitions. We identified different classes of proteases including ATP-dependent Clp proteases (ClpX, ClpP, ClpP2, ClpP4, ClpC1), ATP-dependent zinc metalloproteases (Ftsh, FtsH2, FtsH4), CAAX protease, and carboxy-terminal proteases (CtpA, CtpB) (Figure 6D, Table 1). These proteases were, in general, more abundant under nitrogen-fixing conditions than the non-fixing condition in both the L/D cycles. The CtpA (cce_3991), was >5-fold higher under nitrogen-fixing cells compared to non-fixing cells but its expression changed only by 20% between L/D (Table 1). The Clp proteases also showed significant upregulation under nitrogen-fixing condition with ClpP2 showing almost 5-fold higher abundance (Table 1). Among the four zinc metalloproteases, FtsH4 (cce_1593) showed a 2-fold increase in the dark under nitrogen-fixing compared to non-fixing conditions. Our data indicates potential contribution of chaperones, heat shock proteins and diverse classes of proteases in BNF to maintain cellular proteostasis.

Comparison between transcripts and proteins expressions

It has been demonstrated that in eukaryotic cells, changes in gene transcripts alone can define only one-third of the cell phenotypes, while about 40% can be defined by proteins alone (60), suggesting that both techniques are complementary. There is often a poor correlation between transcripts levels and protein abundances which is attributed to factors such as differences in timing between peak transcript and protein expressions, post-translational modifications, stability, translational efficiency, and protein turnover. Comparing the relationship between mRNA transcript and protein levels is crucial for understanding regulatory processes and functions within cells. While such comparisons have been widely reported in plants (61-63) and animals (64-67), they are less documented in cyanobacteria. Using previously published transcriptomic data (35),

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which represents transcript level changes at three time points during the dark cycle compared to the light cycle under nitrogen-fixing growth conditions (1h, 5h, and 9h). Our proteomic data represents a single time point at 6 hours into the dark cycle. We identified 868 proteins in our data sets that matched the transcriptomics data. We also identified 158 proteins from those that were significantly changing between D/L cycles under nitrogen-fixing condition that matched the transcriptomics data.

The comparison revealed good correlations between proteomic data and transcriptomic data at one and five hours into the dark cycle, with the strongest correlation observed between one-hour transcript and six-hour protein data (Figure 7, Supplementary Table S8). As expected, there was no correlation between proteomic data and nine-hour transcriptomic data. When comparing significant proteins with corresponding transcripts, a similar correlation was observed under nitrogen-fixing growth conditions. Interestingly, there was a good correlation between transcripts of nitrogen fixation (*nif)* genes and proteins at all time points, suggesting that *nif* transcripts are stable and present throughout the dark cycle (Figure 7).

To focus on the nitrogenase enzymes, we separately plotted them with *nif* transcripts (Supplementary Figure 3). The proteomic data showed good correlation with the five hours transcriptomic data. However, as expected one and nine hours of *nif* transcripts showed poor to no correlation. The comparison, though highlights timing differences between the peak of transcripts and proteins in *Crocosphaera* 51142, many transcripts peak at early dark phases and remain stable several hours into the phase, suggesting that the relationship between transcripts and proteins depends on individual or groups of transcripts and proteins involved in specific pathways. This comparison sheds light on the regulation of gene expression and protein abundance to coordinate cellular processes, particularly under nitrogen-fixing growth conditions.

Conclusions

Our analysis provides a global view of how nitrogen and light-dark phases impact the *Crocosphaera* 51142 proteome. Results demonstrate significant modulation of metabolic pathways by cells to adapt to transitions between light-dark phases under nitrogen-fixing or non-fixing conditions. Regulation of CoxB1 and GlgP1 in synchronization with nitrogenase and uptake hydrogenase suggests that cells maintained suboxic condition for nitrogenase activities by higher levels of respiration and glycogen metabolism. Previous transcriptomics and proteomics studies in *Crocosphaera* 51142 and other strains have also reported significant upregulation of *nif* genes and proteins ranging from 10-40-fold during the dark cycle under nitrogen-fixing conditions.

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

However, our current data reveals a dynamic range of expression levels for Nif proteins, ranging from a 7-fold to a dramatic 450-fold increase, specifically after six hours into the dark cycle under nitrogen-fixing conditions, is remarkable, and has not been observed before. Such a remarkable and previously unreported dynamic range of nitrogenase expression in response to nitrogen-fixing conditions during the dark cycle is important and may contribute to the understanding of the regulatory mechanisms governing BNF in cyanobacteria and may have broader implications for the optimization of nitrogen-fixing cyanobacteria in various applications, including biotechnology and environmental management.

Data Availability:

All the raw LC-MS/MS data are deposited in MassIVE data repository (massive.ucsd.edu) with MASSIVE-ID: MSV000094471

All the mapped spectra for the annotated peptides in the global dataset are available in the MS Viewer repository and can be accessed using the search key **q5zmiw9ebd** and by using the following URL:
https://msviewer.ucsf.edu/prospector/cgi-bin/mssearch.cgi?report_title=MS-Viewer&search_key=q5zmiw9ebd&search_name=msviewer

Acknowledgements

All LC-MS/MS experiments were performed at the Purdue Proteomics Facility in the Bindley Bioscience Center of Purdue University. We thank Rodrigo Mohallem and other members of the Aryal lab for discussion and feedback in data analysis and interpretation.

Authors contributions

Punyatoya Panda: Culture growth, treatments, proteomics sample preparation, LC-MS/MS data acquisition, data analysis, writing original draft, review, and editing. **Swagarika J. Giri**: GO annotation, data analysis, writing original draft, review, and editing. **Louis Sherman**: conceptualization, project supervision, data interpretation, review, and editing. **Daisuke Kihara**: conceptualization, data analysis, interpretation, manuscript editing, fund acquisition. **Uma K. Aryal**: conceptualization, methodology, data collection, data analysis and interpretation, writing original manuscript, review, and editing, fund acquisition, project supervision.

Funding

This work was partly supported by funding from the National Science Foundation – DBI2003635. DK also acknowledges support from National Science Foundation (DBI2146026,

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

IIS2211598, DMS2151678, CMMI1825941, and MCB1925643) and by the National Institutes of Health (R01GM133840).

Notes:

The authors declare no competing financial interests.

Abbreviations

LC-MS/MS	Liquid Chromatography-Tandem Mass Spectrometry
RSLC	Rapid Separation Liquid Chromatography
HEPES-KOH	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid potassium salt
ATCC	American Type Culture Collection
L^-	Cells harvested six hours into the light cycle under nitrogen fixing condition.
L^+	Cells harvested six hours into the light cycle under nitrogen non-fixing condition.
D^-	Cells harvested six hours into the dark cycle under nitrogen fixing condition.
D^+	Cells harvested six hours into the dark cycle under nitrogen non-fixing condition.
BNF	Biological Nitrogen Fixation
PSI and PSII	Photosystem I and Photosystem II
GO	Gene Ontology
CO_2	Carbon dioxide
N_2	Dinitrogen
Nif	Nitrogenase enzymes
O_2	Oxygen
PQC	Protein Quality Control
ATP	Adenosine Triphosphate
AAA+ Protease	ATP Associated Protease
NADPH	Nicotinamide Adenine Dinucleotide Phosphate

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Representative list of proteins exhibiting significant changes in at least one pairwise comparison among different experimental conditions (nitrate, light-dark cycles, or interaction between nitrate and light-dark cycles).

Figure legends:

Figure 1: Experimental workflow. (A). Culture growth conditions and sample collection timeline. *Crocosphaera subtropica* ATCC51142 cultures were grown in ASP2 medium without $NaNO_3$ (NO_3^-) and with $NaNO_3$(NO_3^+) at 30ºC under continuous light for seven days. Cultures were then grown for seven days at 12h light/dark (L/D) cycles before harvesting cells at 6 hours into the light (L) and 6 hours into the dark (D). (B). After lysis, cell lysates were separated into soluble and insoluble fractions by differential centrifugation. Soluble lysates were acetone precipitated and insoluble fractions were solubilized with 0.1% Rapigest. Both fractions were digested with trypsin and analyzed by LC-MS/MS with four biological replicates per group. Data processing and statistical analysis were performed using MaxQuant and Perseus respectively. Sequence-based function prediction of the identified proteins was done using PFP, Phylo-PFP and ESG to assign GO terms to the identified proteins.

Figure 2: Differentially regulated proteins during diurnal cycles and nitrogen fixation states. *(*A). Venn Diagram showing protein overlap between D^-, D^+, L^- and L^+ conditions. (B). PCA Plot of all the analyzed replicates of D^-, D^+, L^- and L^+ samples. The explained variances are indicated in the brackets. (C). Heatmap shows the hierarchical clustering of 361 proteins changing due to interaction of effect of light and dark as well as nitrate. Both rows and columns were clustered by Euclidean distance and average linkage method, column clustering indicates the reproducibility within the replicates for each condition.

Figure 3: GO classification of significantly changing proteins. Circular bar plot shows the gene ontology classification for the significantly changing proteins due to differences in light and dark (blue), effect of nitrate (green) and interaction of both these conditions (red) (Supplementary Table S11). BSP: Biosynthetic process, MP: Metabolic process, EC 3.5.4.26: diaminohydroxyphosphoribosylaminopyrimidine deaminase, EC 1.1.1.193: 5-amino-6-(5-phosphoribosylamino)uracil reductase.

Figure 4. Proteins exhibiting the highest fold change dynamics. Volcano Plots indicate the differential regulation of proteins because of nitrate (A, B) and effect of light and dark (C, D).

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Significantly changing proteins determined by T-tests with a q-value ≤ 0.05 and threshold log_2(fold change) =±0.58 are indicated in red (significantly up) and blue (significantly down) in their respective conditions. Proteins exhibiting the highest fold change dynamics in each comparison are specified. Proteins with q=0 were replaced with the lowest possible q-value (or the highest possible –log10(q-value) for the respective dataset.

Figure 5: Expression of nitrogenase enzyme clusters. A. Spider chart depicting the differential regulation of enzymes involved in the nitrogenase cluster under various conditions ($D^+/D^-/L^+/L^-$). The indicated the log (LFQ intensity) in increasing order. B. List of all the identified nitrogenase clusters and associated genes that play an important role in nitrogen-fixation. All the proteins (except the highlighted ones in grey) are significantly changing in at least one condition.

Figure 6. Response of proteins involved in ribosomal subunits, carbon concentrating mechanism and protease degradation. A. Heatmap of proteins involved in 30s (A) and 50s ribosomal subunit (B). Heatmap of proteins involved in carbon concentrating mechanism (C) and proteasomal degradation pathway(D). Hierarchical clustering of the proteins was performed by Euclidean distance measurement and average linkage. The average intensities of all the four replicates for each condition was used to plot these heatmaps.

Figure 7: Comparison of proteomics and transcriptomics data. Scatterplots show the comparison of proteomics and transcriptomics data. The left panel shows the correlation of all the identified proteins in D- condition that were common between the proteomic and transcriptomic data, and the right panel shows the significantly changing proteins in D- compared to the pooled control. Proteomics data at 6hours into the dark was compared with transcriptomics data 1 hour, 5 hours and 9 hours into the dark respectively. The color of the data points indicates if the proteins were identified with a threshold fold change of ±1.5 (or log (foldchange= ±0.58) under both conditions (purple), protein only(blue), transcript only(green) or none(grey).

Supplementary Figure 1. A. Heatmap represents significantly changing proteins because of nitrate. Hierarchical clustering of 1370 proteins was performed by Euclidean distance, with average linkage method for both rows and columns**.** B. Correlation scatterplot. Correlational analysis of all the replicates of all conditions by Pearson's correlation method.

Supplementary Figure 2. **Differential regulation of proteins under light-dark cycles and nitrogen-fixation states.** A. Heatmap indicating the response of proteins involved in various

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

glycolysis, TCA cycle and OPP pathway. B. Heatmap indicating the regulation of proteins in the kai cluster of genes involved in maintenance of circadian rhythm. Hierarchical clustering of the proteins was performed by Euclidean distance measurement and average linkage. All the four replicates for each condition are indicated.

Supplementary Figure 3. Comparison of proteomic and transcriptomic expression of genes involved in the Nif cluster. Proteomics data at 6hours into the dark was compared with transcriptomics data 1 hour, 5 hours and 9 hours into the dark respectively. The color of the data points indicates if the proteins were identified with a threshold fold change of ±1.5 (or log (foldchange= ±0.58) under both conditions (purple), protein only(blue), transcript only(green) or none(grey).

References

1. Reddy, K. J., Haskell, J. B., Sherman, D. M., and Sherman, L. A. (1993) Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. *J Bacteriol* 175, 1284-1292
2. Bar-On, Y. M., Phillips, R., and Milo, R. (2018) The biomass distribution on Earth. *Proc Natl Acad Sci U S A* 115, 6506-6511
3. Szeinbaum, N., Toporek, Y. J., Reinhard, C. T., and Glass, J. B. (2021) Microbial helpers allow cyanobacteria to thrive in ferruginous waters. *Geobiology* 19, 510-520
4. Satta, A., Esquirol, L., and Ebert, B. E. (2023) Current Metabolic Engineering Strategies for Photosynthetic Bioproduction in Cyanobacteria. *Microorganisms* 11
5. Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omoregie, E., Steward, G. F., Hansen, A., and Karl, D. M. (2001) Unicellular cyanobacteria fix N2 in the subtropical North Pacific Ocean. *Nature* 412, 635-638
6. Sherman, L. A., Meunier, P., and Colón-López, M. S. (1998) Diurnal rhythms in metabolism:: A day in the life of a unicellular, diazotrophic cyanobacterium. *Photosynth Res* 58, 25-42
7. Sherman, L. A., Min, H., Toepel, J., and Pakrasi, H. B. (2010) Better living through cyanothece - unicellular diazotrophic cyanobacteria with highly versatile metabolic systems. *Adv Exp Med Biol* 675, 275-290
8. Welsh, E. A., Liberton, M., Stockel, J., Loh, T., Elvitigala, T., Wang, C., Wollam, A., Fulton, R. S., Clifton, S. W., Jacobs, J. M., Aurora, R., Ghosh, B. K., Sherman, L. A., Smith, R. D., Wilson, R. K., and Pakrasi, H. B. (2008) The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc Natl Acad Sci U S A* 105, 15094-15099
9. Schneegurt, M. A., Sherman, D. M., Nayar, S., and Sherman, L. A. (1994) Oscillating behavior of carbohydrate granule formation and dinitrogen fixation in the cyanobacterium Cyanothece sp. strain ATCC 51142. *J Bacteriol* 176, 1586-1597
10. Schneegurt, M. A., Sherman, D. M., and Sherman, L. A. (1997) Composition of the carbohydrate granules of the cyanobacterium, Cyanothece sp. strain ATCC 51142. *Arch Microbiol* 167, 89-98
11. Liberton, M., Austin, J. R., 2nd, Berg, R. H., and Pakrasi, H. B. (2011) Unique thylakoid membrane architecture of a unicellular N2-fixing cyanobacterium revealed by electron tomography. *Plant Physiol* 155, 1656-1666

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12. Toepel, J., Welsh, E., Summerfield, T. C., Pakrasi, H. B., and Sherman, L. A. (2008) Differential transcriptional analysis of the cyanobacterium Cyanothece sp. strain ATCC 51142 during light-dark and continuous-light growth. *J Bacteriol* 190, 3904-3913
13. Kondo, T., Strayer, C. A., Kulkarni, R. D., Taylor, W., Ishiura, M., Golden, S. S., and Johnson, C. H. (1993) Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. *Proc Natl Acad Sci U S A* 90, 5672-5676
14. Deschamps, P., Colleoni, C., Nakamura, Y., Suzuki, E., Putaux, J. L., Buleon, A., Haebel, S., Ritte, G., Steup, M., Falcon, L. I., Moreira, D., Loffelhardt, W., Raj, J. N., Plancke, C., d'Hulst, C., Dauvillee, D., and Ball, S. (2008) Metabolic symbiosis and the birth of the plant kingdom. *Mol Biol Evol* 25, 536-548
15. Aryal, U. K., Callister, S. J., Mishra, S., Zhang, X., Shutthanandan, J. I., Angel, T. E., Shukla, A. K., Monroe, M. E., Moore, R. J., Koppenaal, D. W., Smith, R. D., and Sherman, L. (2013) Proteome analyses of strains ATCC 51142 and PCC 7822 of the diazotrophic cyanobacterium Cyanothece sp. under culture conditions resulting in enhanced H(2) production. *Appl Environ Microbiol* 79, 1070-1077
16. Aryal, U. K., Stockel, J., Krovvidi, R. K., Gritsenko, M. A., Monroe, M. E., Moore, R. J., Koppenaal, D. W., Smith, R. D., Pakrasi, H. B., and Jacobs, J. M. (2011) Dynamic proteomic profiling of a unicellular cyanobacterium Cyanothece ATCC51142 across light-dark diurnal cycles. *BMC Syst Biol* 5, 194
17. Aryal, U. K., Stockel, J., Welsh, E. A., Gritsenko, M. A., Nicora, C. D., Koppenaal, D. W., Smith, R. D., Pakrasi, H. B., and Jacobs, J. M. (2012) Dynamic proteome analysis of Cyanothece sp. ATCC 51142 under constant light. *J Proteome Res* 11, 609-619
18. Fay, P., and Cox, R. M. (1967) Oxygen inhibition of nitrogen fixation in cell-free preparations of blue-green algae. *Biochim Biophys Acta* 143, 562-569
19. Wong, P. P., and Burris, R. H. (1972) Nature of oxygen inhibition of nitrogenase from Azotobacter vinelandii. *Proc Natl Acad Sci U S A* 69, 672-675
20. Meunier, P. C., Colon-Lopez, M. S., and Sherman, L. A. (1997) Temporal Changes in State Transitions and Photosystem Organization in the Unicellular, Diazotrophic Cyanobacterium Cyanothece sp. ATCC 51142. *Plant Physiol* 115, 991-1000
21. Mallen-Ponce, M. J., Huertas, M. J., and Florencio, F. J. (2022) Exploring the Diversity of the Thioredoxin Systems in Cyanobacteria. *Antioxidants (Basel)* 11
22. Kim, S. Q., Mohallem, R., Franco, J., Buhman, K. K., Kim, K. H., and Aryal, U. K. (2022) Global landscape of protein complexes in postprandial-state livers from diet-induced obese and lean mice. *Biochem Biophys Res Commun* 629, 40-46
23. Kim, S. Q., Mohallem, R., Franco, J., Buhman, K. K., Kim, K. H., and Aryal, U. K. (2022) Multi-Omics Approach Reveals Dysregulation of Protein Phosphorylation Correlated with Lipid Metabolism in Mouse Non-Alcoholic Fatty Liver. *Cells* 11
24. Mohallem, R., and Aryal, U. K. (2020) Regulators of TNFalpha mediated insulin resistance elucidated by quantitative proteomics. *Sci Rep* 10, 20878
25. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26, 1367-1372
26. Tyanova, S., and Cox, J. (2018) Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Methods Mol Biol* 1711, 133-148
27. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13, 731-740
28. Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P., and Minghim, R. (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16, 169
29. Deeb, S. J., Tyanova, S., Hummel, M., Schmidt-Supprian, M., Cox, J., and Mann, M. (2015) Machine Learning-based Classification of Diffuse Large B-cell Lymphoma Patients by Their Protein Expression Profiles. *Mol Cell Proteomics* 14, 2947-2960

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

30. Hawkins, T., Chitale, M., Luban, S., and Kihara, D. (2009) PFP: Automated prediction of gene ontology functional annotations with confidence scores using protein sequence data. *Proteins* 74, 566-582

31. Jain, A., and Kihara, D. (2019) Phylo-PFP: improved automated protein function prediction using phylogenetic distance of distantly related sequences. *Bioinformatics* 35, 753-759

32. Chitale, M., Hawkins, T., Park, C., and Kihara, D. (2009) ESG: extended similarity group method for automated protein function prediction. *Bioinformatics* 25, 1739-1745

33. Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J. M., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S., Fisk, D. G., Hirschman, J. E., Hong, E. L., Nash, R. S., Sethuraman, A., Theesfeld, C. L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S. Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E. M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., White, R., and Gene Ontology, C. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32, D258-261

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

35. Stockel, J., Welsh, E. A., Liberton, M., Kunnvakkam, R., Aurora, R., and Pakrasi, H. B. (2008) Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc Natl Acad Sci U S A* 105, 6156-6161

36. Aryal, U. K., Ding, Z., Hedrick, V., Sobreira, T. J. P., Kihara, D., and Sherman, L. A. (2018) Analysis of Protein Complexes in the Unicellular Cyanobacterium Cyanothece ATCC 51142. *J Proteome Res* 17, 3628-3643

37. Zhang, X., Sherman, D. M., and Sherman, L. A. (2014) The uptake hydrogenase in the unicellular diazotrophic cyanobacterium Cyanothece sp. strain PCC 7822 protects nitrogenase from oxygen toxicity. *J Bacteriol* 196, 840-849

38. Badger, M. R., Hanson, D., and Price, G. D. (2002) Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. *Funct Plant Biol* 29, 161-173

39. Whitehead, L., Long, B. M., Price, G. D., and Badger, M. R. (2014) Comparing the in vivo function of alpha-carboxysomes and beta-carboxysomes in two model cyanobacteria. *Plant Physiol* 165, 398-411

40. Summers, M. L., Wallis, J. G., Campbell, E. L., and Meeks, J. C. (1995) Genetic evidence of a major role for glucose-6-phosphate dehydrogenase in nitrogen fixation and dark growth of the cyanobacterium Nostoc sp. strain ATCC 29133. *J Bacteriol* 177, 6184-6194

41. Muro-Pastor, M. I., Reyes, J. C., and Florencio, F. J. (2001) Cyanobacteria perceive nitrogen status by sensing intracellular 2-oxoglutarate levels. *J Biol Chem* 276, 38320-38328

42. Bach, B., Meudec, E., Lepoutre, J. P., Rossignol, T., Blondin, B., Dequin, S., and Camarasa, C. (2009) New insights into gamma-aminobutyric acid catabolism: Evidence for gamma-hydroxybutyric acid and polyhydroxybutyrate synthesis in Saccharomyces cerevisiae. *Appl Environ Microbiol* 75, 4231-4239

43. Singh, A. K., Elvitigala, T., Bhattacharyya-Pakrasi, M., Aurora, R., Ghosh, B., and Pakrasi, H. B. (2008) Integration of carbon and nitrogen metabolism with energy production is crucial to light acclimation in the cyanobacterium Synechocystis. *Plant Physiol* 148, 467-478

44. Forde, B. G., and Lea, P. J. (2007) Glutamate in plants: metabolism, regulation, and signalling. *J Exp Bot* 58, 2339-2358

45. Buren, S., Jimenez-Vicente, E., Echavarri-Erasun, C., and Rubio, L. M. (2020) Biosynthesis of Nitrogenase Cofactors. *Chem Rev* 120, 4921-4968

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

46. Nonaka, A., Yamamoto, H., Kamiya, N., Kotani, H., Yamakawa, H., Tsujimoto, R., and Fujita, Y. (2019) Accessory Proteins of the Nitrogenase Assembly, NifW, NifX/NafY, and NifZ, Are Essential for Diazotrophic Growth in the Nonheterocystous Cyanobacterium Leptolyngbya boryana. *Front Microbiol* 10, 495
47. Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E., and Dean, D. R. (1989) Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii. *J Bacteriol* 171, 1017-1027
48. Simon, H. M., Homer, M. J., and Roberts, G. P. (1996) Perturbation of nifT expression in Klebsiella pneumoniae has limited effect on nitrogen fixation. *J Bacteriol* 178, 2975-2977
49. Shi, T., Ilikchyan, I., Rabouille, S., and Zehr, J. P. (2010) Genome-wide analysis of diel gene expression in the unicellular N(2)-fixing cyanobacterium Crocosphaera watsonii WH 8501. *ISME J* 4, 621-632
50. Toepel, J. R., McDermott, J. E., Summerfield, T. C., and Sherman, L. A. (2009) Transcriptional Analysis of the Unicellular, Diazotrophic Cyanobacterium Cyanothece Sp. Atcc 51142 Grown under Short Day/Night Cycles(1). *J Phycol* 45, 610-620
51. Tamagnini, P., Axelsson, R., Lindberg, P., Oxelfelt, F., Wunschiers, R., and Lindblad, P. (2002) Hydrogenases and hydrogen metabolism of cyanobacteria. *Microbiol Mol Biol Rev* 66, 1-20, table of contents
52. Bothe, H., Schmitz, O., Yates, M. G., and Newton, W. E. (2010) Nitrogen fixation and hydrogen metabolism in cyanobacteria. *Microbiol Mol Biol Rev* 74, 529-551
53. Parry, M. A., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H., and Whitney, S. M. (2013) Rubisco activity and regulation as targets for crop improvement. *J Exp Bot* 64, 717-730
54. Hanson, M. R., Lin, M. T., Carmo-Silva, A. E., and Parry, M. A. (2016) Towards engineering carboxysomes into C3 plants. *Plant J* 87, 38-50
55. Welkie, D., Zhang, X., Markillie, M. L., Taylor, R., Orr, G., Jacobs, J., Bhide, K., Thimmapuram, J., Gritsenko, M., Mitchell, H., Smith, R. D., and Sherman, L. A. (2014) Transcriptomic and proteomic dynamics in the metabolism of a diazotrophic cyanobacterium, Cyanothece sp. PCC 7822 during a diurnal light-dark cycle. *BMC Genomics* 15, 1185
56. Tikhonov, A. N. (2018) The Cytochrome b (6) f Complex: Biophysical Aspects of Its Functioning in Chloroplasts. *Subcell Biochem* 87, 287-328
57. Niforou, K., Cheimonidou, C., and Trougakos, I. P. (2014) Molecular chaperones and proteostasis regulation during redox imbalance. *Redox Biol* 2, 323-332
58. He, Z., and Mi, H. (2016) Functional Characterization of the Subunits N, H, J, and O of the NAD(P)H Dehydrogenase Complexes in Synechocystis sp. Strain PCC 6803. *Plant Physiol* 171, 1320-1332
59. Bouchnak, I., and van Wijk, K. J. (2021) Structure, function, and substrates of Clp AAA+ protease systems in cyanobacteria, plastids, and apicoplasts: A comparative analysis. *J Biol Chem* 296, 100338
60. Plotkin, J. B. (2010) Transcriptional regulation is only half the story. *Mol Syst Biol* 6, 406
61. Olanrewaju, G. O., Haveman, N. J., Naldrett, M. J., Paul, A. L., Ferl, R. J., and Wyatt, S. E. (2023) Integrative transcriptomics and proteomics profiling of Arabidopsis thaliana elucidates novel mechanisms underlying spaceflight adaptation. *Front Plant Sci* 14, 1260429
62. Wu, Y., Liu, H., Bing, J., and Zhang, G. (2022) Integrative transcriptomic and TMT-based proteomic analysis reveals the mechanism by which AtENO2 affects seed germination under salt stress. *Front Plant Sci* 13, 1035750
63. Kirk, P., Amsbury, S., German, L., Gaudioso-Pedraza, R., and Benitez-Alfonso, Y. (2022) A comparative meta-proteomic pipeline for the identification of plasmodesmata proteins and regulatory conditions in diverse plant species. *BMC Biol* 20, 128

bioRxiv preprint doi: https://doi.org/10.1101/2024.07.30.605809; this version posted July 30, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

64. Lasut-Szyszka, B., Gdowicz-Klosok, A., Malachowska, B., Krzesniak, M., Bedzinska, A., Gawin, M., Pietrowska, M., and Rusin, M. (2024) Transcriptomic and proteomic study of cancer cell lines exposed to actinomycin D and nutlin-3a reveals numerous, novel candidates for p53-regulated genes. *Chem Biol Interact* 392, 110946

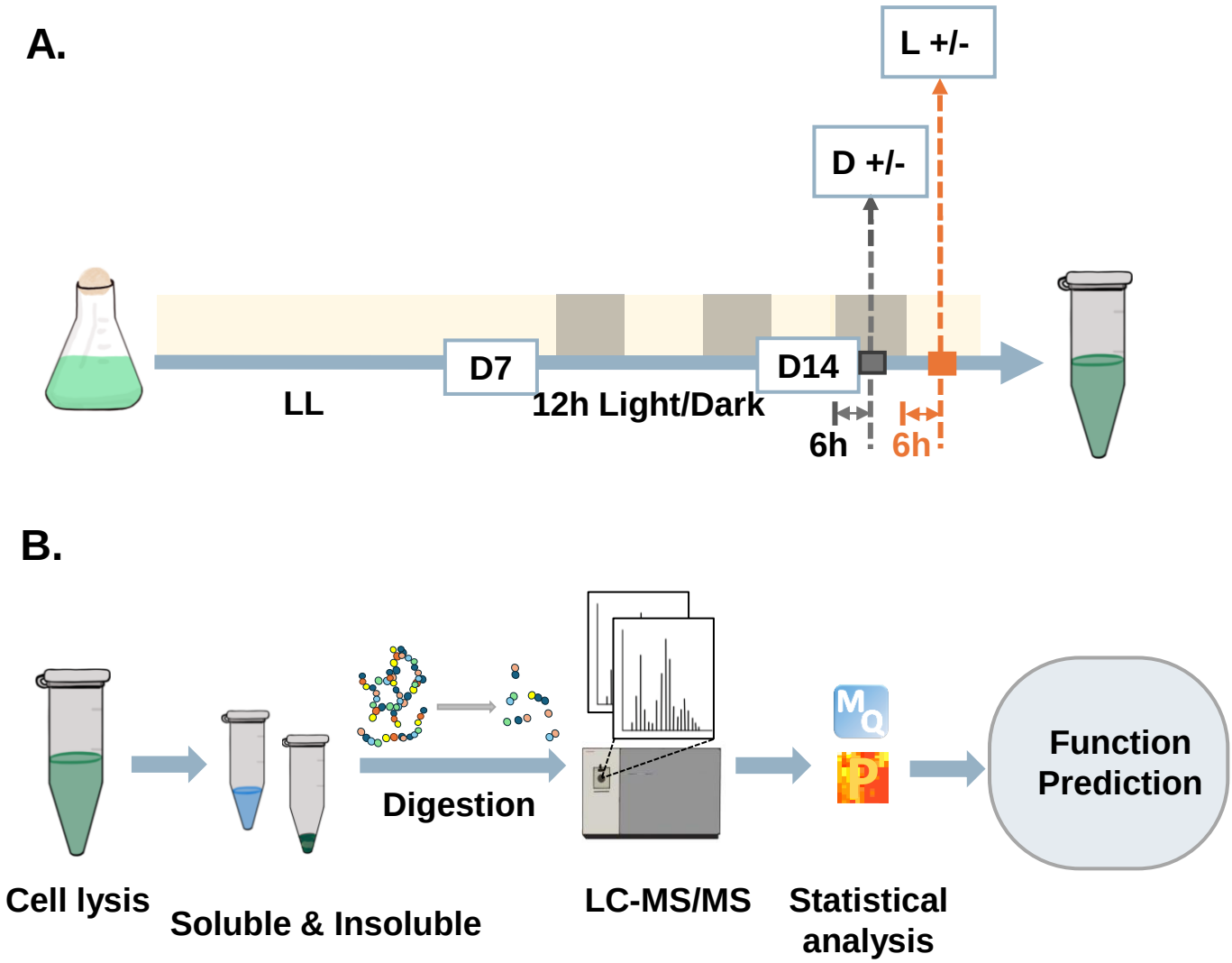
65. Wang, C., Pan, Z., Sun, L., and Li, Q. (2024) Integrative transcriptomic and proteomic profile revealed inhibition of oxidative phosphorylation and peroxisomes during renal interstitial fibrosis. *J Proteomics* 298, 105144

66. Qu, J. H., Chakir, K., Tarasov, K. V., Riordon, D. R., Perino, M. G., Silvester, A. J., and Lakatta, E. G. (2024) Reprogramming of cardiac phosphoproteome, proteome, and transcriptome confers resilience to chronic adenylyl cyclase-driven stress. *Elife* 12

67. Li, C., Zhang, Y., Xia, Q., Hao, B., Hong, Y., Yue, L., Zheng, T., Li, M., and Fan, L. (2023) Multi-omics analysis revealed the mitochondrial-targeted drug combination to suppress the development of lung cancer. *J Cancer Res Clin Oncol* 149, 17159-17174

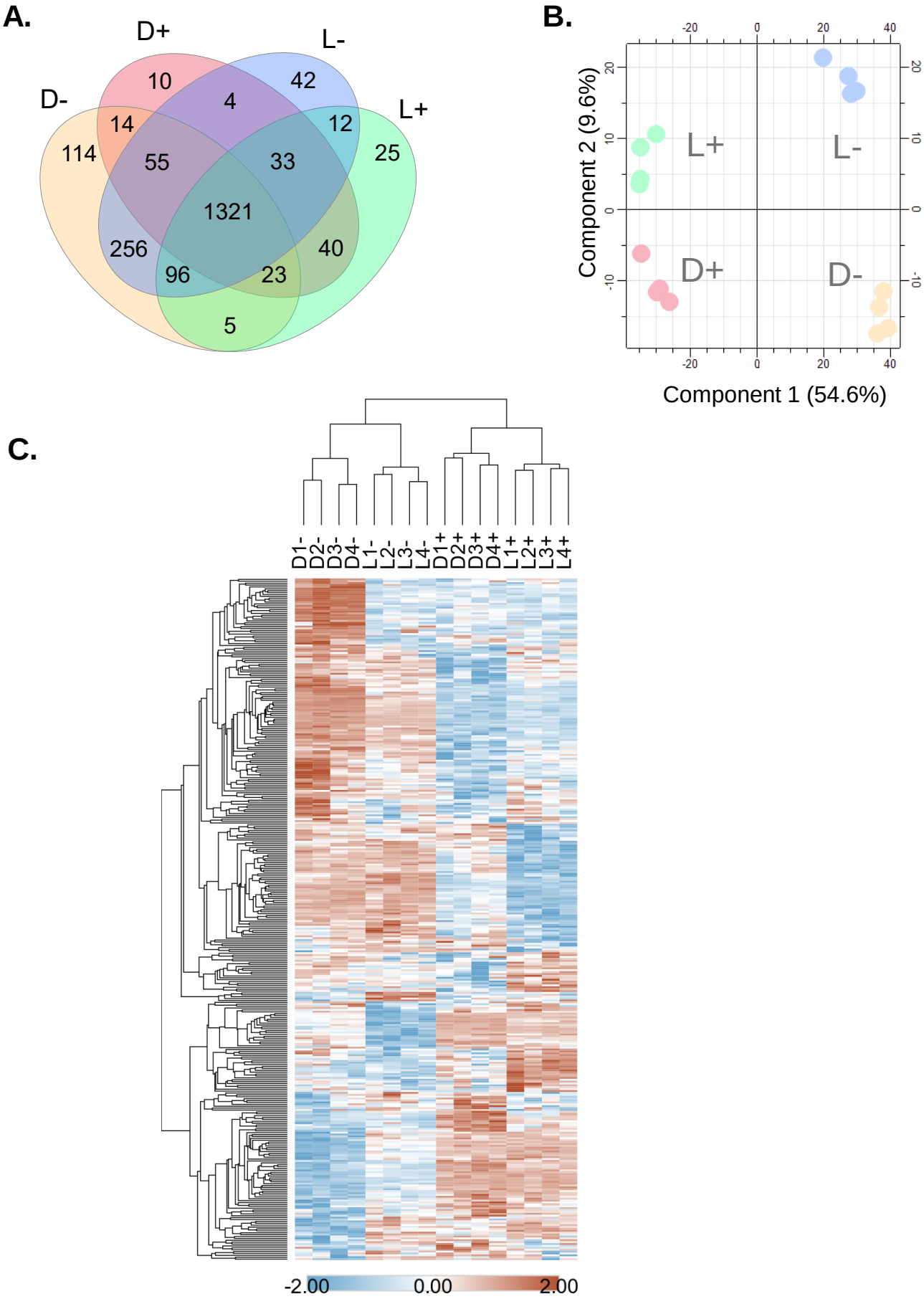
was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1



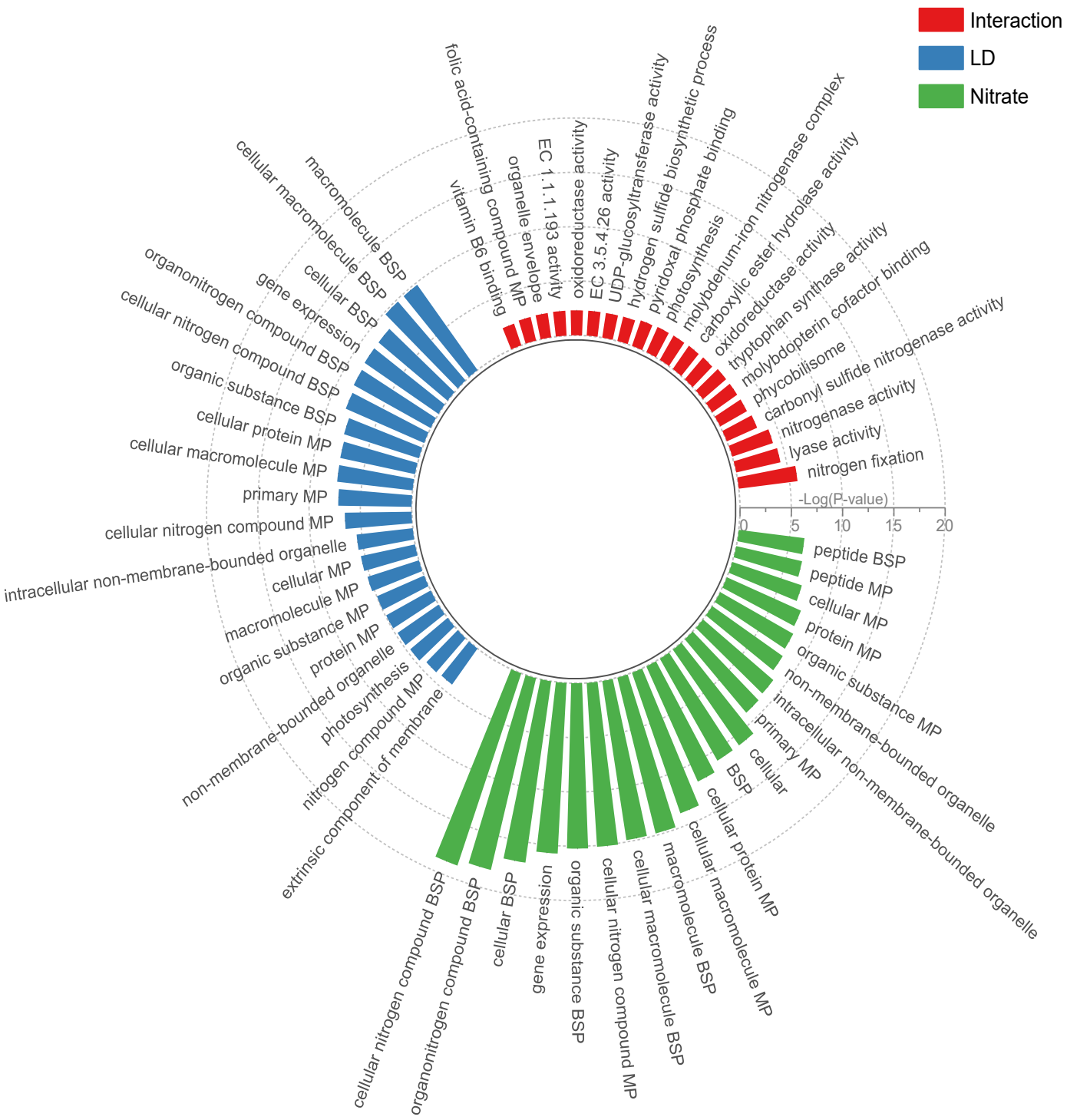
was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2



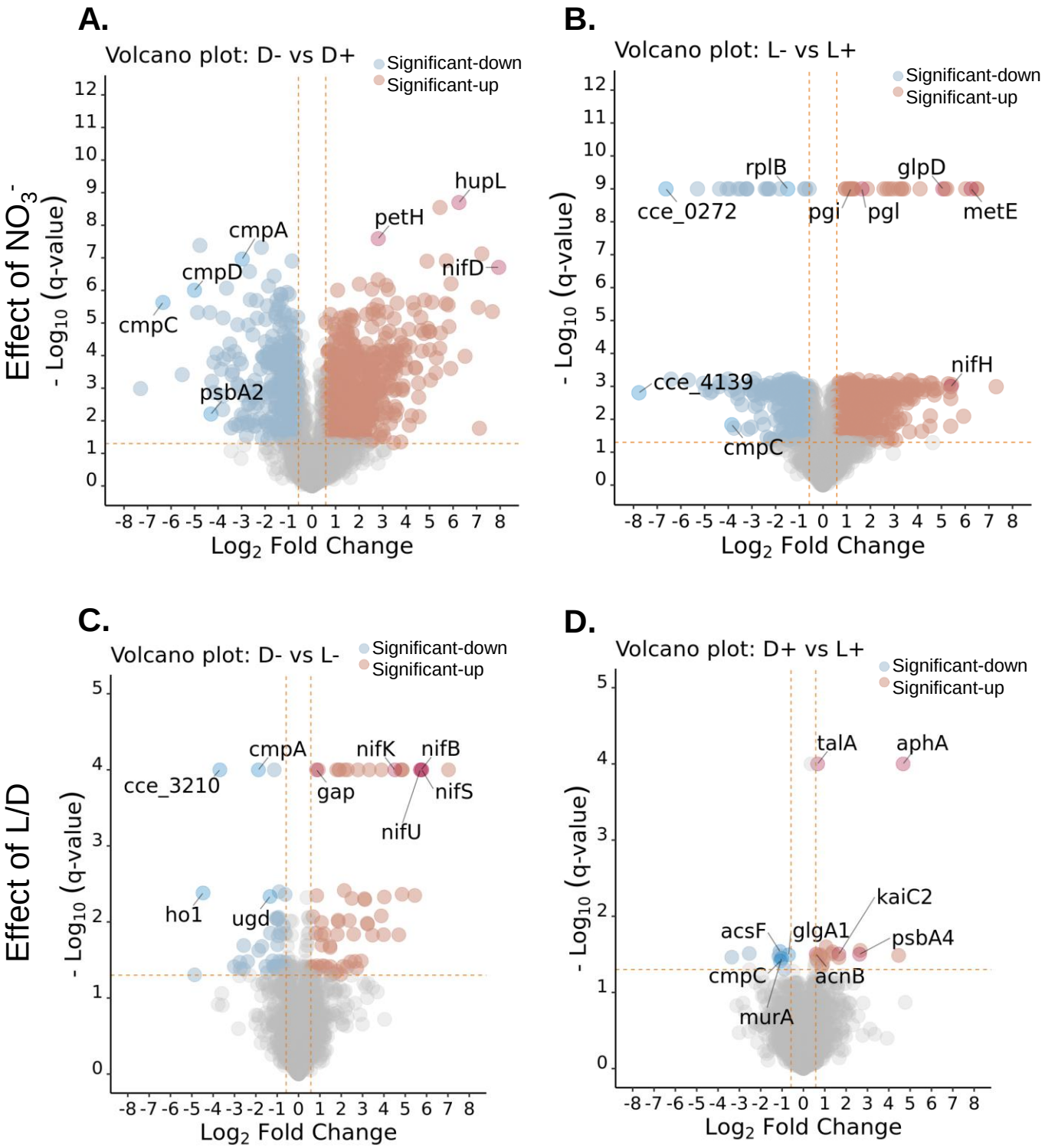
was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3



was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

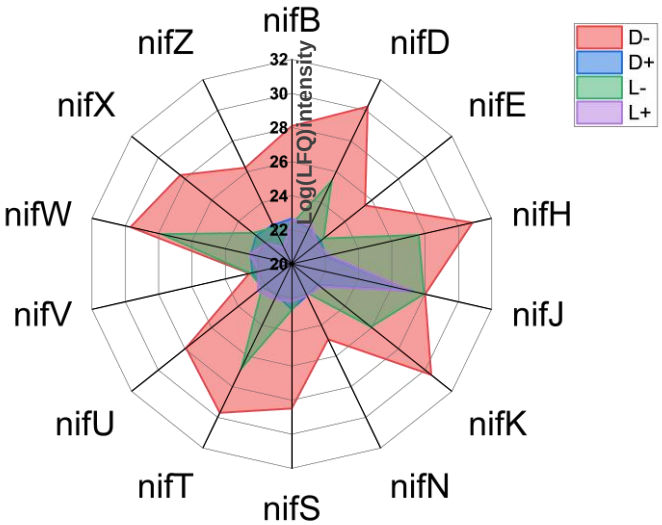
Figure 4



was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5

A.

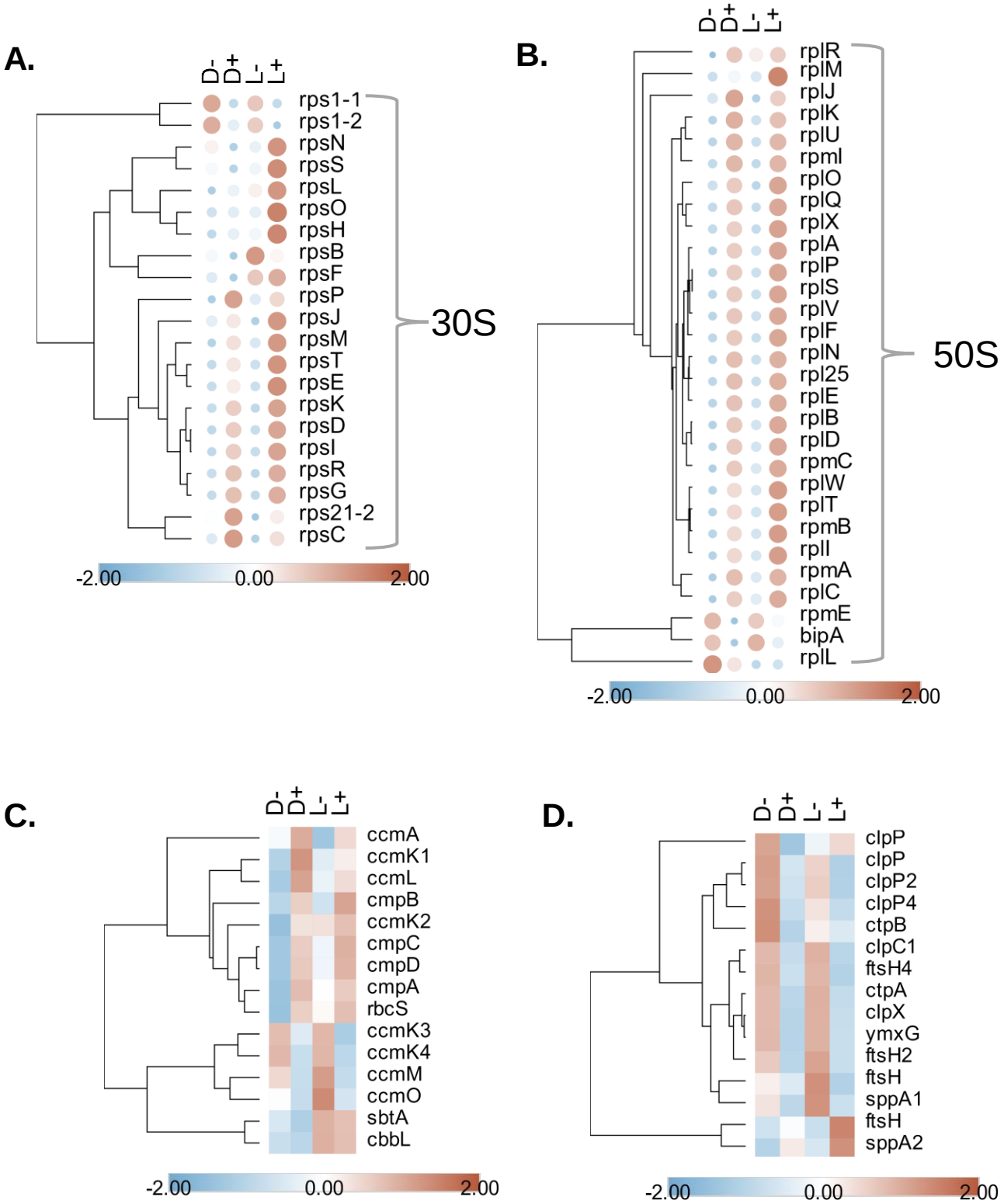


B.

Gene Symbol	Locus tag	Protein name	Fold change		
			Dark	Light	-NO_3
			D-_D+	L-_L+	D-_L-
hesA	cce_0569	HesA protein	3.3	0.8	3.4
hesB	cce_0570	Fe-S cluster biosynthesis, putative nitrogen fixation protein	30.4	2.4	16.2
hupL	cce_1064	[NiFe] uptake hydrogenase large subunit	76.0	9.0	3.8
hupS	cce_1063	HupS	7.9	1.4	9.3
nifB	cce_0554	FeMo cofactor biosynthesis protein NifB	43.2	0.8	53.9
nifD	cce_0560	Nitrogenase molybdenum-iron protein alpha chain	245.0	7.0	28.9
nifE	cce_0563	Nitrogenase iron-molybdenum cofactor biosynthesis protein	10.0	1.4	8.6
nifH	cce_0559	Nitrogenase iron protein	451.0	42.3	9.4
nifJ	cce_0953	Pyruvate-flavodoxin oxidoreductase	1.7	1.8	1.1
nifK	cce_0561	Nitrogenase molybdenum-iron protein beta chain	362.0	13.9	22.8
nifN	cce_0564	Nitrogenase iron-molybdenum cofactor biosynthesis protein	7.2	0.8	8.5
nifS	cce_0557	Cysteine desulfurase	56.6	1.4	55.0
nifT	cce_0547	NifT/fixU	204.0	23.1	7.4
nifU	cce_0558	Nitrogen fixation protein	52.5	0.8	52.1
nifV	cce_0549	Homocitrate synthase	1.1	1.4	1.1
nifW	cce_0568	Nitrogenase-stabilizing/protective protein	149.0	39.0	3.7
nifX	cce_0565	Nitrogen fixation protein X	51.8	2.0	43.6
nifZ	cce_0548	Iron-sulfur cofactor synthesis protein	13.3	2.5	14.8

was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 6



was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

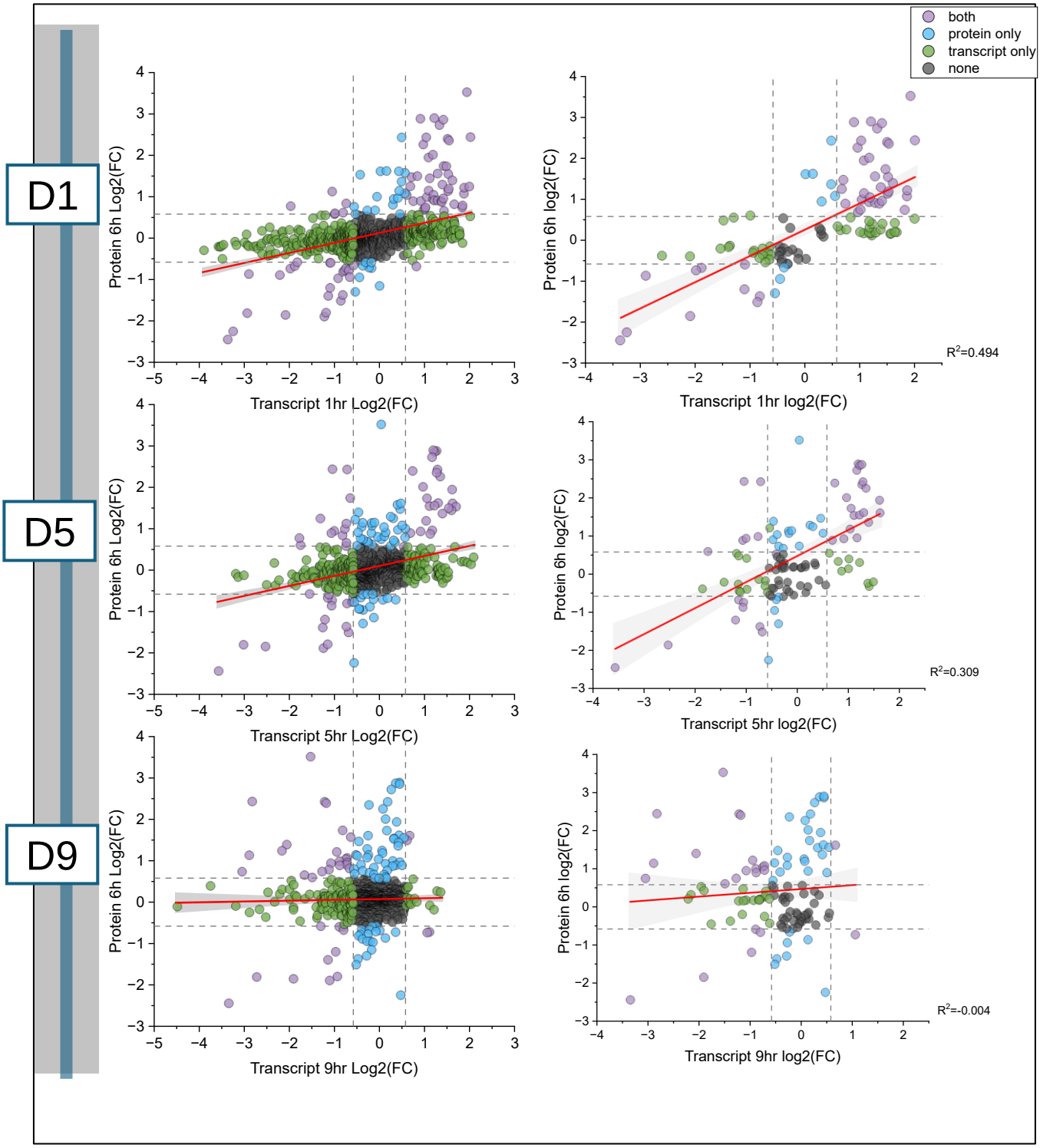



Figure 7