Title

Multi-omics profiling of the cold tolerant Monoraphidium minutum 26B-AM in response to abiotic stress

Authors

Calhoun, Sara

Kamel, Bishoy

Bell, Tisza AS

et al.

Publication Date

2022-07-01

DOI

10.1016/j.algal.2022.102794

Title: Multi-omics profiling of the cold tolerant *Monoraphidium minutum* 26B-AM in response to abiotic stress

Sara Calhoun[1,4], Bishoy Kamel[1], Tisza A.S. Bell[3], Colin P.S. Kruse[2], Robert Riley[1], Vasanth Singan[1], Yuliya Kunde[2], Cheryl D. Gleasner[2], Mansi Chovatia[1], Laura Sandor[1], Christopher Daum[1], Daniel Treen[1], Benjamin P. Bowen[1,4], Katherine B. Louie[1,4], Trent R. Northen[1,4], Shawn R. Starkenburg[2], Igor V. Grigoriev[1,4,5]

Affiliations

[1] US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

[2] Applied Genomics Team, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545

[3] Division of Biological Sciences, Genome Core, University of Montana, Missoula, MT 59801

[4] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720

[5] Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA 94720

Abstract

Microalgae that are of interest for applications in biofuel production must be able to tolerate environmental changes that occur in outdoor cultivation systems. While algal cultures may experience daily temperature fluctuations and seasonal environmental changes, the underlying mechanisms that control and regulate physiological responses and adaptation to environmental pressures are largely unknown. Systems-level characterization enabled by functional genomics can help identify biochemical pathways that promote stability and productivity of algae in various environmental conditions. *Monoraphidium minutum* 26B-AM, a freshwater green microalga, was identified as a top performer in biomass production in winter season screens. We sequenced the genome of *M. minutum* 26B-AM and applied our multi-omics pipeline to profile this high potential strain under high salt and cold temperature perturbations. Through comparative analysis, including other green algae in the class Chlorophyceae, we identified gene families unique to the genus *Monoraphidium*, including a desaturase that has been linked to cold tolerance in plants. We observed that osmolytes, such as trehalose, proline and betaine, accumulate under salt stress, coinciding with upregulation of genes involved in biosynthesis of these metabolites. From the genome annotation, we reconstructed a metabolic model to provide a detailed map of the metabolic pathways and can be used to simulate growth and reaction fluxes. This multi-omics analysis provides a foundation to explore algal strain potential for biofuel applications, guides strain engineering, and expands our understanding of metabolic and regulatory mechanisms of algae in applied systems.

1. Introduction

Marine microalgae are being developed for a myriad of industrial applications due to their ability to sustainably convert carbon dioxide into renewable products. Recently, large-scale screening efforts in photobioreactors and outdoor ponds have evaluated strains from public, academic, and industrial culture collections with suitable characteristics for industrial deployment [1]. Through an integrated screening platform, it was found that *Monoraphidium minutum* 26B-AM exhibited the highest biomass productivity in winter season simulations [1,2] and in outdoor field trials over the winter season (December through February) [3].

Originally collected as a pond contaminant of *Chlorella sorokiniana* as part of the National Association for the Advancement of Bioproducts and Biofuels (NAABB) program, *M. minutum* 26B-AM was first identified as *Kirchneriella cornuta* by morphology and was re-designated as a close relative of *M. minutum* based on high sequence similarity (97-99%) of the 18S small subunit ribosomal RNA genes (NCBI accession: AY846380) [2]. *M. minutum* is a member of the order Sphaeropleales and family Selenastraceae. Many species within the Sphaeropleales, including other *Monoraphidium* strains [4–6], have previously been investigated for their potential as biofuel feedstocks due to their high lipid production. Yet, published genomic data for members of the family Selenastraceae is sparse, with published sequences for only two other nuclear genomes (*Monoraphidium neglectum* [4] and *Messastrum gracile* [7]) and several organellar genomes, including *M. neglectum* [4], *Kirchneriella aperta*, and *Ourococcus multisporus* [8]. An additional genome sequence of *Monoraphidium* sp. 549 is publicly available without an annotation (Genbank accession: NMUB00000000.1). Thus, additional genome

sequencing of these understudied algal lineages, including annotation of novel and extremophilic strains, is needed to help fill existing knowledge gaps.

Microalgal strains that exhibit halotolerance are of particular interest, given that these strains can be cultivated in sea or brackish water, eliminating the need for freshwater-based cultivation. However, increasing salinity for many strains results in biomass productivity losses and can alter lipid composition [9]. Similarly, changes in temperature inherent to outdoor cultivation may also impact productivity and biomass composition [10,11]. A detailed understanding of how algal strains can adapt to abiotic stress on both the molecular and genetic level can help identify pathways that can be modified for enhanced stability and productivity under changing environmental conditions.

Here, we present the genome sequence and gene annotations of *M. minutum*, as well as the transcriptomic and metabolomic profiling for two different perturbations, salt and low temperature. Our multi-omics investigation and metabolic modeling identified transcriptional changes in stress response genes and metabolic changes in the levels of common osmolytes under high salinity. In addition, unique and expanded gene families reflect differences in genomic content compared to other green algae that may be related to cold tolerance. By generating these omics datasets for an industrially relevant strain like *M. minutum*, we've identified specific genes, pathways, and metabolites affected by stress to guide future strain improvement studies.

2. Materials and Methods

2.1 Salinity Challenge

M. minutum 26B-AM was grown in Erlenmeyer flasks of DISCOVR medium [1] in temperature-controlled chambers and stirred at a constant 120 rpm. The chambers maintained a constant temperature of 27 °C, a 14:10 light:dark cycle, and 3000 ppm of CO_2. In brief, three flasks containing unamended DISCOVR medium with a salinity of 5 ppt (control) and three flasks containing amended DISCOVR medium with a salinity of 20 ppt (salt) were inoculated with *M. minutum*. Sampling for sequencing and metabolomic analysis began 24 hours post inoculation and occurred every 24 hours for 6 days. Growth was monitored with optical density for 8 days (Supplementary Figure 1A).

2.2 Temperature Challenge

M. minutum 26B-AM was grown in Erlenmeyer flasks of DISCOVR medium [1] in temperature-controlled chambers and stirred at a constant 120 rpm. Sixteen flasks containing DISCOVR medium were inoculated with *M. minutum* and kept at a constant temperature of 27 °C, a 14:10 light:dark cycle, and 3000 ppm of CO_2 for 24-hours. After 24 hours, 8 flasks were moved to a chamber with a constant temperature of 17 °C, a 14:10 light:dark cycle, and 3000 ppm of CO_2 for an additional 24 hours. Sampling (in quadruplicate) for metabolomic analysis occurred at 6, 12, 24, and 48 hours and sampling (in triplicate) for RNA extraction and sequencing occurred at 12, 24, and 48 hours. Growth was determined by optical density and monitored for 8 days (Supplementary Figure 1B).

2.3 DNA Extraction

Genomic DNA (gDNA) was extracted as previously described [12]. In brief, algal cells digested in agarose plugs to enable extraction of high molecular weight DNA. The gDNA extract was purified using a High Salt-Phenol:Chloroform:IsoAmyl Alcohol protocol recommended by Pacific Biosciences (Pacific Biosciences of California, USA, https://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Guidelines-for-Using-a-Salt-Chloroform-Wash-to-Clean-Up-gDNA.pdf), followed by precipitation of the gDNA with ethanol. Purified gDNA was concentrated using AMPure PB beads (Pacific Biosciences, Menlo Park, USA). Qubit 1 Fluorometer (Invitrogen, USA) was used to measure the DNA concentration. The DNA purity ratios were assessed using the NanoDrop 1000 instrument (Thermo Scientific, USA). The DNA fragment size was determined on TapeStation 4200 (Agilent, Germany).

2.4 Genome assembly

5 µg of genomic DNA was sheared to >10kb using Covaris g-Tubes. The sheared DNA was treated with exonuclease to remove single-stranded ends and DNA damage repair mix followed by end repair and ligation of blunt adapters using SMRTbell Template Prep Kit 1.0 (Pacific Biosciences, Menlo Park, CA). The library was purified with AMPure PB beads. PacBio Sequencing primer was then annealed to the SMRTbell template library and sequencing polymerase was bound to them using Sequel Binding kit 3.0. The prepared SMRTbell template libraries were then sequenced on a Pacific Biosystem's Sequel sequencer using v3 sequencing primer, 1M v3 SMRT cells, and Version 3.0 sequencing chemistry with 600-minute sequencing movie run times.

Filtered subread data was assembled together with Falcon version (pb-assembly 0.0.2, falcon-kit 1.2.3, pypeflow 2.1.0) (https://github.com/PacificBiosciences/FALCON) to generate an initial assembly. Substantial contamination by several bacterial species was detected using BLAST vs. NCBI refseq, analysis of GC/coverage, and principal components analysis of tetranucleotide frequencies (Supplementary Table 1). Two of the bacteria (*Aquimonas* and *Pseudomonas*) were observed at substantially higher coverage than the algal genome. These bacteria were assembled individually by subsampling preads to target their respective coverages using BBtools version 38.47 [reformat.sh samplerate={0.1,0.2}] targeting a coverage of 100x, and assembled with Flye version 2.4 [--pacbio-corr --genome-size {5m,6m} --asm- coverage 50]. Reads that did not recruit to either of these bacteria were then assembled with Flye version 2.4.1 [-pacbio-corr --genome-size 75m --asm-coverage 50], resulting in a main assembly with a third bacterial contaminant (Microbacteriaceae). This third bacterium was assembled by recruiting preads with BBtools version 38.47 [bbduk.sh k=31 mm=f mkf=0.05] to the identified contaminant contigs and a related reference genome downloaded from NCBI (NZ_CP031423.1), and subsequently assembled with Flye version 2.4.1 [--pacbio-corr --genome-size 3.5m --asm-coverage 50]. Finally, preads were recruited to the three assembled bacteria using BBtools version 38.47 [bbduk.sh k=31 mm=f mkf=0.05] and those preads that did not match any contaminant bacterium (47.59% of the original preads) were assembled with Flye version 2.4.1 [flye --pacbio-corr --genome-size 75m --asm-coverage 50]. Residual bacterial contamination was then identified using BLAST vs. NCBI refseq, GC/coverage, and tetranucleotide frequency, and removed from the main assembly. This round of contaminant removal resulted in a fourth nearly-complete bacterial genome (*Nitratireductor*). The algal nuclear genome and organelle contigs were polished with Arrow version SMRTLink v6.0.0.47841

(https://github.com/PacificBiosciences/GenomicConsensus). Contigs less than 1000 bp were excluded.

2.5 Genome annotation and analysis

The nuclear genome assembly was annotated using the Joint Genome Institute (JGI) annotation pipeline [13,14]. The annotation pipeline starts with masking the assembly for repeats using RepeatMasker [15] with the RepBase library [16] and the most frequent (>150) repeats recognized by RepeatScout [17]. Next, protein-coding gene models were predicted using ab initio Fgenesh [18], homology-based Fgenesh+, homology-based GeneWise [19] seeded by BLASTx alignments against the NCBI NR database and transcriptome-based Fgenesh. Automated filtering selected the best model at each genomic locus based on homology and transcriptome support. Predicted proteins were functionally annotated using SignalP [20] for signal sequences, TMHMM [21] for transmembrane domains, InterProScan [22], and protein alignments to NCBI NR, SwissProt [23], KEGG [24], KOG [25] and TCDB for transporter classifications [26]. Hits from InterPro and SwissProt were used to map Gene Ontology terms [27]. Completeness of the genome annotation was evaluated using BUSCO v 4.1.4 in protein mode on the Chlorophyta ortholog dataset (chlorophyta_obd10; 11-20-19) [28]. The isoelectric point was computed using the Biopython package v1.79 [29] based on the methods of Bjellqvist [30,31]. Orthologous gene families were predicted using the OrthoFinder software package v 2.3.8 [32].

2.6 Phylogenetic analysis

The species tree was estimated from putative single-copy orthologs from Chlorophyta genomes identified using OrthoFinder software package v 2.3.8 [32]. The genomes annotations were accessed from the JGI Algal Genome Portal PhycoCosm [33]: *Chlamydomonas reinhardtii* (Chlre5_6), *Chlamydomonas eustigma* (Chleu1), *C. zofingiensis* (Chrzof1), *Dunaliella salina* (Dunsal1_1), *Gonium pectorale* (Gonpec1), *M. neglectum* (Monneg1), *R. subcapitata* (Rapsub1), *Scenedesmus obliquus* UTEX 393 (Sobl393_1), *Scenedesmus obliquus* DOE0152z (Sceobl1), *Scenedesmus* sp. NREL 46B-D3 (Scesp_1), *Tetrabaena socialis* NIES-571 (Tetso1) and *Volvox carteri* (Volca2_1). Using single-copy orthologs, the trimmed alignments were concatenated into a single alignment, and the phylogenetic tree was inferred under maximum-likelihood (RAxML v 8.2.8) using automated bootstrapping, which converged after 100 bootstrap replicates (arguments: -f a -m PROTGAMMAWAGF) [34]. The tree was visualized using FigTree v 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

The phylogenetic tree of the putative sphingolipid delta-4 desaturase gene was built using protein sequences identified by Blastp search of the UniprotRep90 database [35] and PhycoCosm [33] with the *M. minutum* DES4 sequence. Protein sequences were aligned using MAFFT [36], and the tree was inferred under maximum-likelihood with IQ-TREE v 1.6.12 [37]. Support values were estimated using ultrafast bootstrapping (1,000 bootstraps) [38], and the tree was visualized using the iTOL web server [39].

2.7 RNA extraction and sequencing

RNA was extracted and sequenced as previously described [12]. In brief, RNA was extracted from 54 samples (36 salt and 18 cold) using the Omega E.Z.N.A. Plant RNA Kit (Omega Bio-

Tek, Norcross, GA) following manufacturer's protocols with the addition of 3 rounds of liquid nitrogen freeze, thaw, and grinding for cell lysis [40]. The concentration of total RNA was obtained using the Qubit RNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA) and the quality of the RNA was determined by the Agilent RNA 6000 Pico Kit (Agilent Technologies, Santa Clara, CA). 5 μg of total RNA was depleted of ribosomal RNA using the Ribo-Zero H/M/R kit (Illumina Inc., San Diego, CA). The ribosomal RNA depleted samples were then processed using the NEBNext Poly(A) mRNA Magnetic Kit (New England Biolabs Inc., Ipswich, MA) to select for polyadenylated RNA. The resulting poly(A) selected RNA was prepared for sequencing using the ScriptSeq v2 Library Preparation Kit (Illumina Inc., San Diego, CA). All libraries were multiplexed and sequenced on a NextSeq 500 to generate paired-end 151 bp reads using the NextSeq 500/550 High Output Kit v2.5 Kit (300 cycles) (Illumina, Cat. #20024908). RNA-seq data is available in the Sequence Read Archive under Bioproject PRJNA812585.

2.8 Transcriptomic analysis

Following Illumina sequencing, resulting demultiplexed forward and reverse fastq files were filtered and trimmed for quality and artifacts. Using BBDuk, raw reads were evaluated for artifact sequence by kmer matching (kmer = 25), allowing 1 mismatch and detected artifact was trimmed from the 3' end of the reads. RNA spike-in reads, PhiX reads and reads containing any Ns were removed. Quality trimming was performed using the phred trimming method set at Q6. Finally, following trimming, reads under the length threshold were removed (minimum length 25 bases or 1/3 of the original read length - whichever is longer). Filtered reads from each library were aligned to the reference genome using HISAT2 v 2.1.0 [41]. Strand-specific coverage

bigWig files (fwd and rev) were generated using deepTools v 3.1 [42], and featureCounts [43] was used to generate the raw gene counts. DESeq2 (v 1.18.1) [44] was subsequently used to determine which genes were differentially expressed between treatment and control samples at equivalent timepoints. A gene is considered differentially expressed between conditions when the adjusted p-value < 0.05 and $|Log_2FC| > 1$. Functional enrichment analysis was performed on the differentially expressed genes at each time point for both upregulated and downregulated groups. Enrichment for GO terms and KOG domains was calculated and visualized using clusterProfiler v 3.18.1 [45].

2.9 Metabolomics

2.9.1 Metabolite extraction

Prior to extraction, algae biomass (~20 mg) was first lyophilized dry (FreeZone 2.5 Plus, Labconco) then powderized by bead-beating with a 3.2 mm stainless steel bead in a bead-beater (Mini-Beadbeater-96, BioSpec Products) for 5 seconds (2x).

2.9.2 Lipid extraction

To extract lipids, a chloroform-based lipid extraction was performed using a modified Bligh-Dyer approach [46]. Here, 300 µL methanol (MeOH), 150 µL CH_3Cl and 120 µL water was added to the powderized pellet (final ratio of 2:1:0.8 MeOH:CH_3Cl:H_2O), vortexed and sonicated in a water bath for 10 minutes, then an additional 150 µL each of CH_3Cl and MeOH was added (final ratio of 1:1:0.9 MeOH:CH_3Cl:H_2O), which was followed by a brief vortex and 10 minute incubation in a sonic water bath. Samples were centrifuged for 5 minutes at 5000 rpm, and the bottom lipid-enriched chloroform phase was transferred to a new tube. An addition of 300 µL of

chloroform was added, followed by another vortex, sonication, and centrifugation. The bottom chloroform phase was then combined with the previously collected extract. Lastly, chloroform extracts of lipid were dried in a SpeedVac (SPD111V, Thermo Scientific) and stored at -20 ºC.

Polar metabolite extraction. To extract polar metabolites, 450 µL methanol was added to powderized pellets, vortexed and sonicated in a water bath for 10 minutes, stored at -20 ºC for 2 hours, and finally vortexed and sonicated for 10 minutes. After centrifuging for 5 minutes at 5000 rpm, 200 µL of supernatant was transferred to a separate tube. Extracts were then dried in a SpeedVac (SPD111V, Thermo Scientific) and stored at -80 ºC.

2.9.3 Liquid chromatography tandem mass spectrometry

Liquid chromatography tandem mass spectrometry (LC-MS/MS) was performed on extracts of algae biomass using both normal and reverse phase chromatography for polar and lipid metabolites, respectively. Chromatography was performed using an Agilent 1290 LC stack, with MS and MS/MS data collected using a Q Exactive Orbitrap MS (Thermo Scientific, San Jose, CA).

2.9.4 Reverse phase chromatography

Extracted lipids from 20 mg wet algae biomass were resuspended in 240 µL 3:3:4 isopropanol:acetonitrile:methanol (IPA:ACN:MeOH) containing a 4 µM internal standard mixture of deuterium-labeled lipids (Lysophosphatidylcholine (17:1) - #110905, Phosphatidylglycerol-d5 (16:0/18:1) - #110899, Diacylglyceryl-trimethylhomoserine-d9 (16:0/16:0) - #857463P, oleic acid (18:1-d9) - #861809O, Phosphatidylserine (16:0-d31/18:1) -

#110922, Phosphatidic acid (16:0-d31/18:1) - #110920, Phosphatidylethanolamine (16:0-d31/18:1) - #110921, Phosphatidylcholine (16:0-d31/18:1) - #110918, Diacylglycerol-d5 (1,3-16:1) - #110579, Triacylglycerol-d5 (17:0/17:1/17:0) - #110544; Avanti Polar Lipids) and 1 µg/mL 2-Amino-3-bromo-5-methylbenzoic acid (Sigma, M9508). One hundred twenty microliters of this were centrifuge-filtered through a 0.22 µm hydrophilic PVDF membrane (#UFC30GV00, Millipore) and transferred to a glass LC-MS vial. Full MS spectra were collected from *m/z* 132-1500 at 70,000 resolution in both positive and negative ion mode, with MS/MS fragmentation data acquired using stepped 10, 20, and 40 eV collision energies at 17,500 resolution. Chromatography was performed using a C18 column (Agilent ZORBAX Eclipse Plus C18, Rapid Resolution HD, 2.1 x 50 mm, 1.8 µm) at a flow rate of 0.4 mL/min with a 2 µL injection volume. To detect lipids, samples were run on the C18 column at 55 ºC equilibrated with 100% buffer A (60:40 H_2O:ACN with 5 mM ammonium acetate and 0.1% formic acid) for 1 minute, diluting buffer A down to 45% with buffer B (90:10 IPA:ACN w/ 5 mM ammonium acetate and 0.1% formic acid) over 2 minutes, down to 20% A over 8 minutes, then down to 0% A over 1.5 minutes, followed by isocratic elution in 100% buffer B for 1.5 minutes. Samples consisted of 3 biological replicates each and 3 extraction controls, with sample injection order randomized and an injection blank (2 µL of 3:3:4 IPA:ACN:MeOH) run between each sample.

2.9.5 Normal phase chromatography

Extracted polar metabolites from 25-105 mg wet algae biomass were resuspended in 100% methanol containing 10 µM internal standard (5-50 µM of ^{13}C,^{15}N Cell Free Amino Acid Mixture, #767964, Sigma), with the resuspension volume varied for each sample to normalize by biomass weight. One hundred fifty microliters of this were centrifuge-filtered through a 0.22 µm

hydrophilic PVDF membrane (#UFC30GV00, Millipore) and transferred to a glass LC-MS vial. Full MS spectra were collected from *m/z* 70-1050 at 70,000 resolution in both positive and negative mode, with MS/MS fragmentation data acquired using stepped 10, 20 and 40 eV collision energies at 17,500 resolution. Chromatography was performed using a hydrophilic interaction liquid chromatography (HILIC) column (Agilent InfinityLab Poroshell 120 HILIC-Z, 2.1 x 150 mm, 2.7 µm, #673775-924) at a flow rate of 0.45 mL/minute with a 2 µL injection volume. To detect metabolites, samples were run on the HILIC column at 40 ºC equilibrated with 100% buffer B (95:5 ACN:H_2O with 5 mM ammonium acetate) for 1 minute, diluting buffer B down to 89% with buffer A (100% H_2O with 5 mM ammonium acetate and 5 µM methylenediphosphonic acid) over 10 minutes, down to 70% B over 4.75 minutes, then down to 20% B over 0.5 minutes, followed by isocratic elution in 80% buffer A for 2.25 minutes. Samples consisted of 3 biological replicates each and 3 extraction controls, with sample injection order randomized and an injection blank (2 µL MeOH) run between each sample.

2.9.6 Metabolomic identification

Metabolites were identified based on exact mass and comparing retention time (RT) and MS/MS fragmentation spectra to that of standards run using the same chromatography and MS/MS method. LC-MS data was analyzed using custom Python code [47]. Each detected feature (unique m/z coupled with RT) was assigned a score (0 to 3) representing the level of confidence in the metabolite identification. Metabolites given a positive identification had detected m/z $\leq$ 5 ppm or 0.001 Da from theoretical as well as RT $\leq$ 0.5 minutes compared to a pure standard run using the same LC-MS method. A compound with the highest level of positive identification (score of 3) also had matching MS/MS fragmentation spectra compared to either an outside

database (METLIN) [48] or internal database generated from standards run and collected on a Q Exactive Orbitrap MS. MS/MS mismatches to the standard invalidated an identification.

For a detected lipid compound, lipid class was determined based on characteristic fragment ions or neutral loss, and coupled with exact mass to determine specific lipid identity (number of carbons in fatty acid tails and degree of unsaturation). In POS mode, MGDG and DGDG lipids ionized as $[M+NH_4]^+$ with a neutral loss of 179 and 341 [49], respectively, and TAGs as $[M+NH_4]^+$ with fatty acid tails detected in the MS/MS spectra [50]. Lipid standards run separately (DGDG (18:3/18:3) - #840524P, MGDG (18:3/16:3) - #840523P; Avanti Polar Lipids, Inc.) and deuterated internal standards of DG and TAG were used to verify fragmentation pattern and elution times of each lipid class. Owing to lipid degradation over total sample runtime, detected intensity for peak height was normalized to the deuterated DG internal standard for MGDG and DGDG lipids, and to the deuterated TAG internal standard for TAG lipids.

2.10 Metabolic model construction

Enzyme commission numbers (EC numbers) were generated using various homology based tools, specifically: PRIAM [51], eggNOGmapper [52], KOFAM [53] and the BioCyc database [54]. The EC numbers were examined and curated manually to generate a comprehensive consensus gene to EC number table. Model construction was performed using the COBRApy toolbox [55] following the general procedure outlined in [56]. Gap filling of the metabolic model was attempted using MENECO [57] and using the approach described in [58]. Iterations of manual curation were performed to fill remaining gaps and to ensure biosynthesis of core

metabolic precursors using information about biomass composition from other published models for algae and plants [59]. The model was exported in SBML format [60] and subsequent analysis such as flux balance analysis, flux variability analysis and phenotypic phase plane analysis were conducted using COBRApy and the Sybil [61] package. The MEMOTE package was used to detect errors and aid in the curation of the model [62].

3. Results and Discussion

3.1 Genome analysis

The *M. minutum* 26B-AM genome was sequenced using Pacific Biosciences technology. The genome was assembled into 512 contigs with a size of 68.17 Mbp, which is of similar size to other public Selenestraceae genome assemblies (51 Mbp to 75 Mbp, Supplementary Figure 2, Supplementary Table 2). Assessment of genome completeness based on BUSCO suggests that the genome assembly and annotation are high-quality and nearly complete [28]. BUSCO analysis based on the Chlorophyta set (chlorophyta_odb10; 11-20-19) estimated the completeness of the annotation at 96.8%, and over 90% of 26B-AM RNA-seq reads were mapped to the genome. The phylogenetic tree based on single-copy orthologous genes places *M. minutum* among the other sequenced Sphaeropleales, including *M. neglectum*, *Raphidocelis subcapitata* [63], *Chromochloris zofingiensis* [64] and *Scenedesmus* species (Figure 1). The predicted 15,341 genes are comparable in number to other Sphaeropleales genomes but fewer than in other *Scenedesmus* strains which also have larger genome assemblies (22,378 in UTEX 3031 estimated haplotype [65] and 17,406 in NREL 46B-D3 estimated haplotype [12]). Previously sequenced *Monoraphidium* genomes were both produced using Illumina short-read technology, and only *M. neglectum* has a published genome annotation. Although the number of genes in the

M. minutum annotation is fewer than the number of genes in the *M. neglectum* annotation, the gene models are more complete and longer in sequence on average; 95% of genes have complete CDS in *M. minutum* compared to 85% of genes with complete CDS in *M. neglectum*. In addition, the average protein length in the *M. minutum* annotation is 485 amino acids, as compared to an average of 348 amino acids in *M. neglectum,* which likely reflects the difference in genome assembly quality and not true divergence in amino acid length within *Monoraphidium.* The GC content of the *M. minutum* genome is relatively high at 72.1%, which is higher than most Chlorophyta genomes, including *M. neglectum*, and closest to the GC content in *Raphidocelis subcapitata* (71.6%).

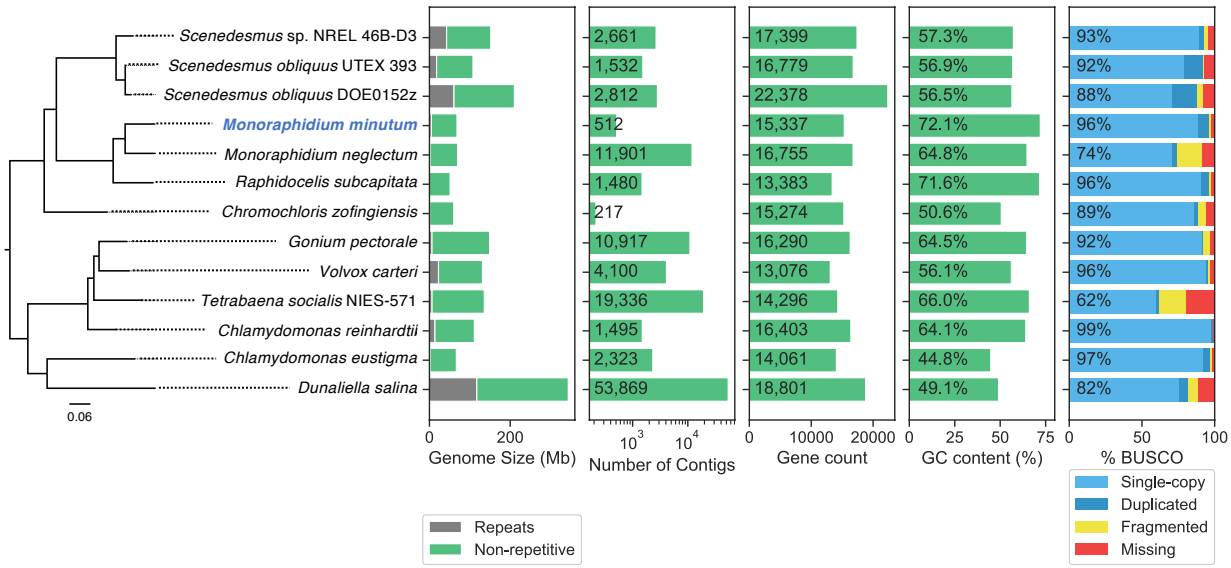

Figure 1. Phylogenetic tree based on single-copy ortholog genes (left) and comparison of genome statistics (right) showing assembly size, number of contigs, number of genes in a single haplotype, GC content, and BUSCO scores (Chlorophyta_odb10).

We analyzed the amino acid composition of the *M. minutum* protein sequences (Supplementary Figure 3A). Overall, the frequencies of the 20 standard amino acids in *M. minutum* proteins were similar to the frequencies observed across the *M. neglectum* and *R. subcapitata* proteins, but the

frequency of glutamine was lower in the proteome of *M. minutum* than in *M. neglectum*. Compared to *Scenedesmus*, the amino acid composition observed in *M. minutum* proteins had lower frequencies of alanine, glycine, arginine and proline. In addition to amino acid composition, we calculated the isoelectric point (pI) of *M. minutum* proteins and compared the distribution to other Chlorophyta species (Supplementary Figure 3C, Supplementary File 1). The distribution of pI across the *M. minutum* proteome has a significantly larger number of proteins with high pIs compared to other species, including *M. neglectum*. There are 2,152 *M. minutum* proteins with a pI greater than 11.5, representing over 14% of its proteome and four times as many as found in the *M. neglectum* proteome. Most of the high pI proteins are hypothetical proteins, which lack known protein domains (fewer than 4% have Pfam domains) and do not have detectable orthologs. These proteins are typically short in length (an average length of 225 amino acids) yet are transcribed according to the RNAseq data. The high pI proteins also have a distinct amino acid composition characterized by high frequencies of arginine and proline compared to the rest of the proteome (Supplementary Figure 3B). There are several studies showing examples of highly basic proteins linked to stress response, including cold acclimation, in other organisms [66–69]. While most of these high pI proteins in *M. minutum* have no functional annotation and no known function, they could be important for cold temperature adaptation. Further investigation is needed to functionally characterize these hypothetical proteins and explore the link to stress tolerance.

We compared orthologs that were annotated as salt-responsive and cold-responsive in the *Arabidopsis thaliana* genome to identify genes that show conserved function in responding to abiotic stress. We downloaded the sequences of 738 protein-coding genes annotated as salt stress

responsive and 830 protein-coding genes annotated as cold stress responsive from The Arabidopsis Information Resource (TAIR) database [70]. Orthologs to the *Arabidopsis* salt stress and cold stress responsive genes were identified in the *M. minutum* genome using Orthofinder. 244 salt stress orthologs and 331 cold stress orthologs were found in *M. minutum*, with 60 genes belonging to both sets (Supplementary File 2).

To identify patterns of conserved genes across green algae, we performed a comparative analysis of the *M. minutum* genome with 12 other Chlorophyta genomes (*Chlamydomonas reinhardtii*, *Chlamydomonas eustigma*, *C. zofingiensis*, *Dunaliella salina*, *Gonium pectorale*, *M. neglectum*, *R. subcapitata*, *Scenedesmus obliquus* UTEX 393, *Scenedesmus obliquus* DOE0152z, *Scenedesmus* sp. NREL 46B-D3, *Tetrabaena socialis* NIES-571 and *Volvox carteri*) by clustering genes into homologous gene families using the OrthoFinder software [32]. Clustering of these genes detected a total of 8,731 families identified in *M. minutum*, including 2,848 gene families conserved across all the Chlorophyta genomes included in this analysis (Figure 2A). 3,251 gene families were identified as unique to *M. minutum*, representing over a third of all gene families and consisting of 3,576 individual genes. The majority of the species-specific gene families were single-copy genes (3,110), and the remaining 141 gene families had multiple gene copies. Of the species-specific gene families, less than 4% were annotated with Pfam domains (114 out of 3,251) (Figure 2B), but many of these gene models were supported by RNA-seq evidence, which could reflect the large number of new gene functions to be discovered. Of the species-specific gene families with annotations (Supplementary Table 3), some of these unique families contain Pfam domains that are associated with regulation, such as protein kinase domains (PF00069, PF007714) and DNA-binding domains (PF00097, PF01753) and domains of

unknown function (PF04765, PF005336, PF03235). There were also a few species-specific genes annotated with domains with metabolic functions, including a taurine catabolism dioxygenase (PF02668), a type III polyketide synthase-like protein (PF08392), an acyl transferase (PF00698) and a citrate synthase (PF00285). Most of the high pI proteins were encoded by single-copy species-specific genes (2,005 out of 2,152 genes). While only 179 gene families were specific to the genus *Monoraphidium*, nearly a third of the genus-specific families contained Pfam domains (57 out of 179). Similar to the species-specific gene families, some of genus-specific families contains domains that may be related to regulation and signaling, such as the AP2 domain (PF00847), MYND finger (PF01753), and protein kinase domains (PF00069, PF07714). Other domains were associated with transport, including a ZIP Zinc transporter (PF02535), an ABC transporter (PF00005, PF01061), and a sugar transporter (PF00083). Related to lipid metabolism, there were domains in the genus-specific gene families annotated with a sphingolipid delta-4-desaturase (PF08557) and a methyltransferase (PF13489).

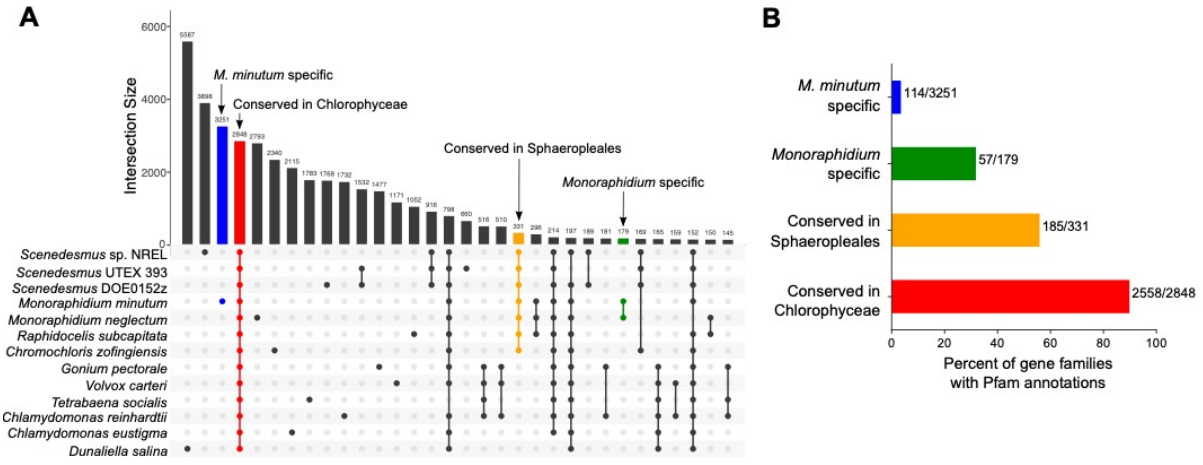

Figure 2. Homologous gene families across Chlorophyceae genomes. A) Upset plot of gene families of shared gene families. Top of the plot shows the number of gene families present, and the bottom shows intersections where gene families are present in the Chlorophyceae genomes, indicated by the filled circle. B) Percent of gene families with Pfam annotations for gene families specific to *M. minutum*, gene families specific to *Monoraphidium* genomes, gene families conserved in Sphaeropleales genomes, and gene families conserved in Chlorophyceae genomes. The numbers show the fraction of gene families with Pfam annotations.

3.2 Sphingolipid desaturase domains in Monoraphidium

One of the *Monoraphidium*-specific gene families includes a gene present in *M. minutum* (Protein ID: 384353) that is annotated with a sphingolipid delta-4 desaturase domain (PF08557) and a fatty acid desaturase domain (PF00487) (Figure 3A). The ortholog in *M. neglectum* shares 78% sequence identity and contains the fatty acid desaturase domain, but lacks the sphingolipid delta-4 desaturase (DES4) domain. However, the *M. neglectum* gene is located at the end of a short scaffold and is likely a fragmented model. Desaturases catalyze the introduction of double bonds into fatty acid chains or long chain bases, and different desaturases can be highly specific for the position at which they modify the chemical bonds [71]. Other sphingolipid desaturase genes are present in the *M. minutum* genome, including two copies annotated as sphingolipid delta-8 desaturase (DES8). The DES8 proteins also contain two Pfam domains, a cytochrome b5-like domain (PF00173) and a fatty acid desaturase domain (PF00487) (Figure 3A). Comparing the pattern of the sphingolipid desaturase genes in other Chlorophyta genomes based on the presence of these domains, the DES8 gene is conserved in many other Chlorophyta and across the class Chlorophyceae (Figure 3B). In contrast, the DES4 gene is present in genomes outside of the Chlorophyceae, but within the Chlorophyceae, it is only found in the two *Monoraphidium* genomes.

Sphingolipid desaturases have been previously associated with cold tolerance in plants [72–74] and similarly, could be important for cold stress tolerance in *Monoraphidium*. Sphingolipids are often associated with signaling and regulatory roles that could be essential for stress responses. Modifications of these molecules could impact how this alga responds to environmental

stressors, including cold temperature. Using the *M. minutum* DES4 gene as a query sequence, we searched for similar protein sequences in the public databases PhycoCosm [33] and UniProt [23] and constructed a phylogenetic tree of related desaturase genes from diverse organisms (Figure 3C). The *M. minutum* DES4 is placed in a clade separate from all other green algal sequences, aside from the *M. neglectum* ortholog. Instead, the clade contains Rhodophyta genes from *Porphyra umbilicalis* and *Gracilariopsis chorda* and a gene from a *Sphaeroforma arctica*, a cold-adapted marine Ichthyosporea protist. While the functions of the putative proteins in this cluster have not been characterized, the annotation of functional domains support a role in lipid desaturation. The placement of the *Monoraphidium* DES4 gene in the tree suggests that it may have been acquired through a horizontal gene transfer event and could confer traits for stress tolerance that distinguish it from other green algae. Convergent evolution of sphingolipid desaturases associated with cold tolerance may have occurred in plants [73]. Similarly, our phylogenetic analysis suggests that the DES4 gene may have originated from outside of the green algae evolutionary lineage. Furthermore, the DES4 gene in the arctic *Chlorella vulgaris* NJ-7 was found to be under positive selection compared to its mesophilic relative, suggesting that this enzyme plays a role in adaptation to cold environments [75].

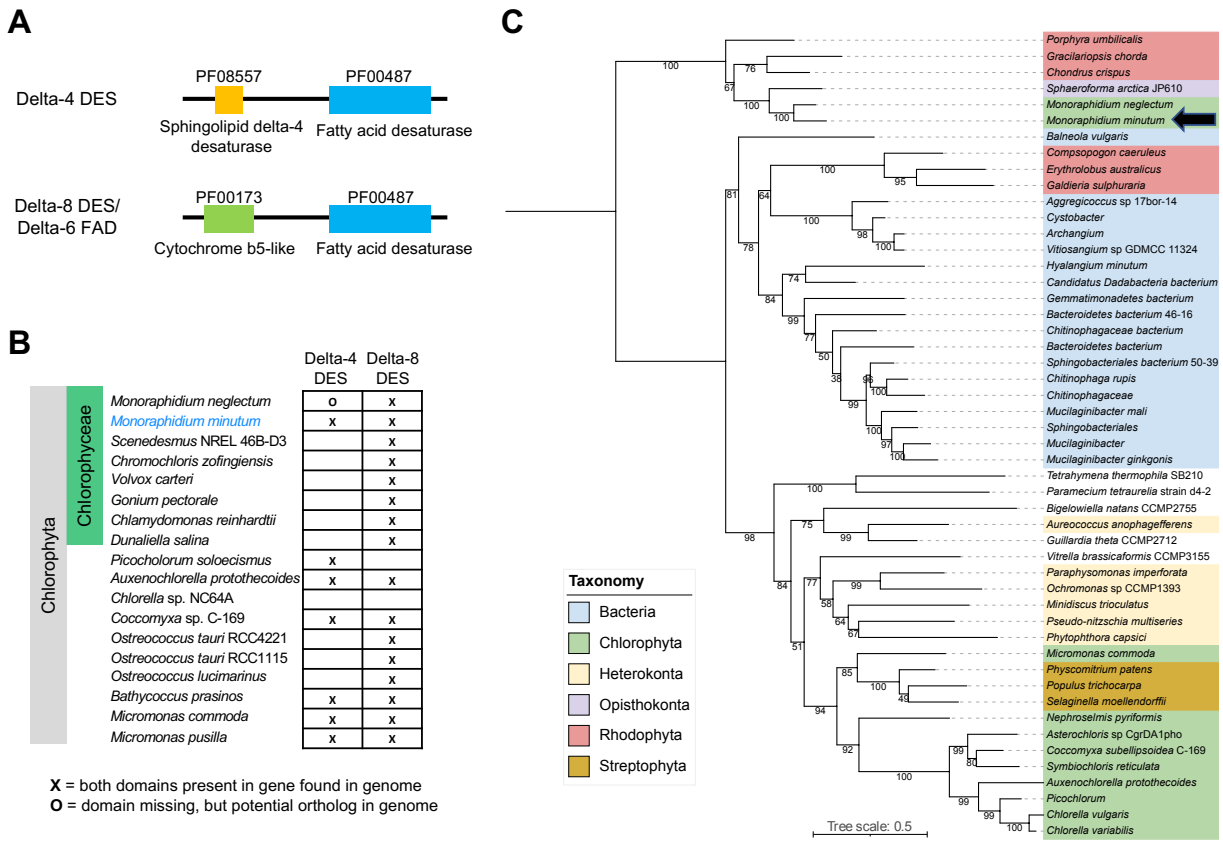

Figure 3. Phylogeny of sphingolipid delta-4 desaturase and sphingolipid delta-8 desaturase genes. A) Pfam domain content in genes annotated as a putative sphingolipid delta-4 desaturase (DES) and a putative sphingolipid delta-8 desaturase/delta-6 fatty acid desaturase. B) Presence of predicted sphingolipid delta-4 desaturase and sphingolipid delta-8 desaturase genes in Chlorophyta genomes based on Pfam domains. C) Phylogenetic tree of putative sphingolipid delta-4 desaturase genes determined by maximum likelihood. Leaves are labeled by organism name and colored by taxonomy. The black arrow shows where the *Monoraphidium minutum* delta-4 DES gene is located on the tree.

3.3 Perturbation studies under high salt and cold temperature conditions

In addition to the genomic analysis, we profiled the transcriptomic and metabolic response of *M. minutum* under two perturbation conditions to understand how this algal strain responds to different abiotic stressors. Cultures of *M. minutum* were grown in photobioreactors under salt stress (5 vs. 20 ppt) and cold temperature stress (17°C vs 27°C, approximately 10 degrees lower than temperature that induced the optimum maximum specific growth rate [1]). Growth was monitored by optical density during the sampling period (Supplementary Figure 1). At 20 ppt

salt concentration, decreased growth relative to the control was observed starting at day 2. The cultures at 17°C did not show a significant difference in growth compared to the control samples grown at 27°C over the period monitored. Samples for transcriptomic and metabolomic profiling were taken once a day over the course of six days for the salt perturbation and over the course of a 48-hr period for the cold temperature perturbation, along with control samples (salinity of 5ppt and 27°C).

3.4 Differentially expressed genes identified by transcriptomics

Using principal component analysis (PCA) performed on the transcriptome counts, the high salt samples were clustered separately from the control samples. Over the time course, there were 2,860 (18.6%) differentially expressed genes (DEGs) in at least one of the high salt samples compared to the control samples using a $|Log_2FC|$ greater than 1 and an adjusted p-value < 0.05 (Supplementary File 3). However, fewer than 100 genes were differentially expressed at all time points, suggesting variation in expression over the six-day time period.

The expression levels of the *Arabidopsis* salt stress and cold stress responsive ortholog genes were compared in the perturbation experiments to control samples (Figure 4). While only a subset of the orthologs of *Arabidopsis* stress responsive genes are differentially expressed in *M. minutum*, upregulation of multiple transcription factors, including the cold shock DNA-binding domain protein (Protein ID: 115409) an SRF-type transcription factor (Protein ID: 456447), and other regulatory proteins, some suggest that there may be shared regulatory pathways across distant lineages. Orthologs that may be involved in signaling pathways were also differentially

expressed including a histidine kinase (Protein ID: 492346), a Ser/Thr protein kinase (Protein ID: 458027), and a protein phosphatase (Protein ID: 387870).

A membrane-associated gene (Protein ID: 443680) showed consistent upregulation in both the salt stress and cold stress samples and encodes for a 59-amino acid long protein annotated as a proteolipid membrane potential (PMP) modulator that is a homolog of the stress responsive RCI2A gene in *Arabidopsis thaliana* (AT3G05880) [76]. The *Arabidopsis* RCI2A gene is induced under multiple different stress conditions, including low temperature, dehydration, salt stress, and abscisic acid (ABA) [76,77]. Differential expression of the PMP gene in *M. minutum* suggests that it acts through a conserved mechanism response to abiotic stress. Genes annotated as chlorophyll A/B binding proteins associated with photosystem I showed upregulation in salt stress. Although the chlorophyll binding proteins were annotated as cold-responsive in *Arabidopsis*, these genes were upregulated across multiple time points in the salt stress and early in the cold stress response.

Functional enrichment analysis using KOG and GO annotations was performed on the upregulated and downregulated DEGs at each time point. No KOG terms were enriched across all time points in the salt stress samples, but KOG terms for Ca^{2+}-dependent lipid-binding protein CLB1 (KOG1012), Ca^{2+}-transporting ATPase (KOG0202), and dyneins (KOG3595) were enriched at multiple time points in upregulated genes (Figure 5A-B). Genes annotated as fasciclin and related adhesion glycoproteins (KOG1437) were also found to be enriched in the set of upregulated DEGs across all three time points in the cold samples. GO terms that were enriched in the upregulated DEGs under high salt are associated with photosynthesis and

chlorophyll biosynthesis, and downregulated DEGs under high salt were enriched with GO terms associated with general transporter activity and chitin metabolism (Supplementary Figure 4). Both the high salt and cold temperature samples showed downregulated DEGs enriched with GO terms associated with the ribosome and translation, indicative of slowing protein production under stress.

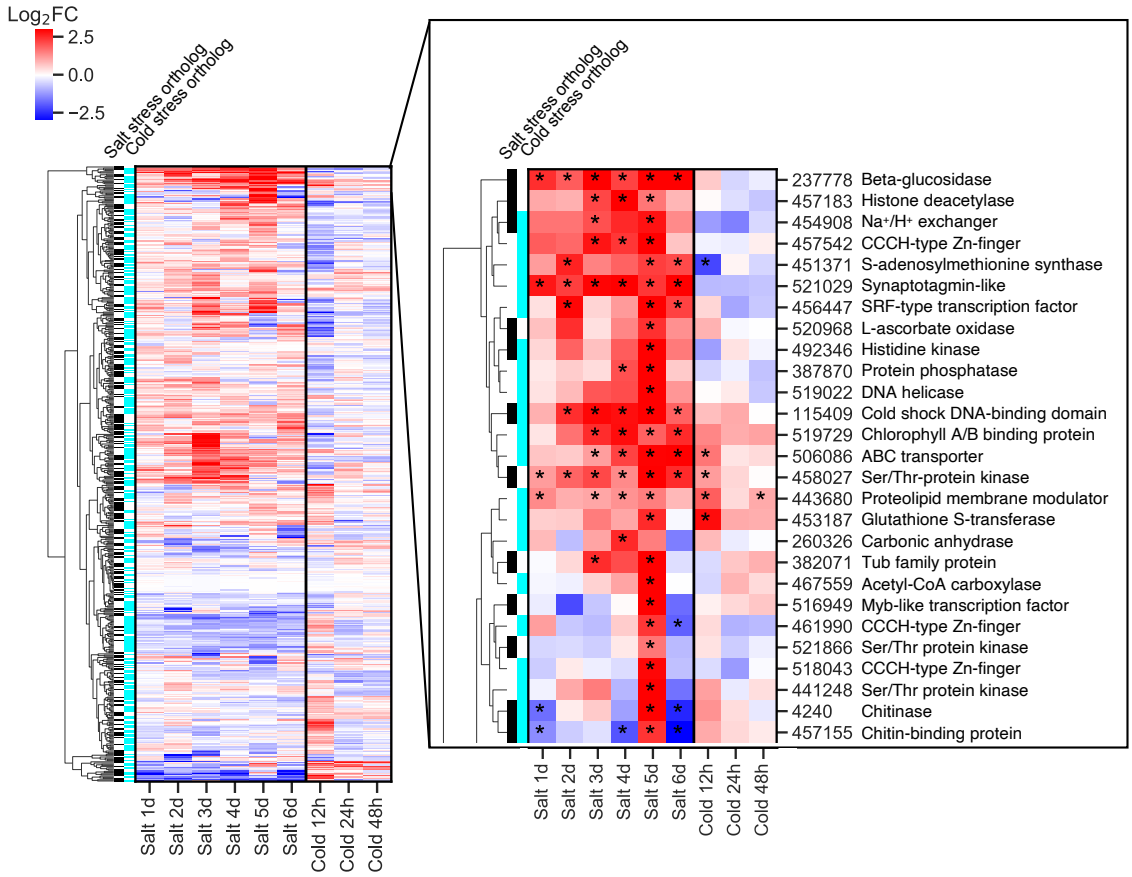

Figure 4. Fold change in expression of genes with stress-responsive orthologs in *Arabidopsis thaliana*. Heatmap rows are ordered by hierarchical clustering of Log_2FC. Inset: Subset of genes that show upregulation across salt stress samples. Colored bars to the left of the heatmap indicate genes with a salt stress ortholog (black) or cold stress ortholog (cyan) in *Arabidopsis*.

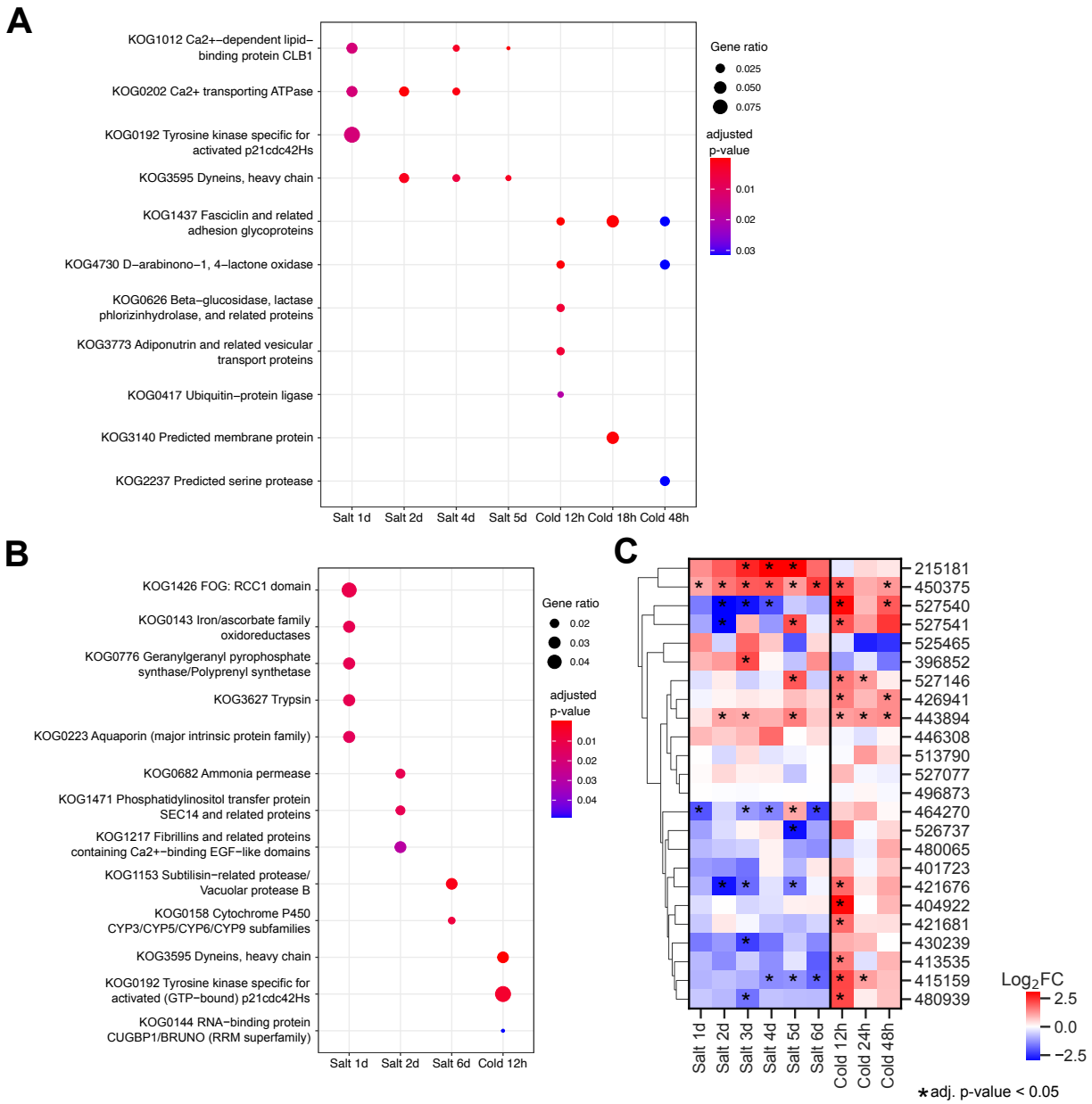

Figure 5. Functional enrichment of protein domains in differentially expressed genes under salt and cold stresses. A) Enrichment of KOG domains in upregulated genes in each treatment compared to control at the equivalent time point. B) Enrichment of KOG domains in downregulated genes in each treatment compared to control. C) Log_2FC in gene expression of genes annotated with fasciclin-like domains (Pfam domain PF02469) in *M. minutum* in salt stress and cold stress samples relative to controls at respective time points.

3.5 Differential expression of fasciclin-like domains in response to stress

Many *M. minutum* genes that contain fasciclin-like domains were differentially expressed in response to both salt and cold stress. Fasciclin-like domains are functionally diverse with known

roles in cell adhesion and various stress responses in other algae and plants [78–81]. There are 35 genes with the Pfam annotation (PF02469) present in the *M. minutum* genome. The number and proportion of fasciclin-annotated genes in Sphaeropleales genomes is larger than in other Chlorophyceae genomes (Supplementary Figure 5). Genes annotated with the KOG annotation for fasciclin-like domains were statistically enriched in upregulated genes under cold stress. Fourteen of the fasciclin genes were upregulated at the initial 12-hr time point under cold stress (Figure 5C). Several of the fasciclin genes continue to exhibit increased gene expression through the rest of the 48-hr sampling period. Three fasciclin genes were upregulated in the salt stress relative to control at multiple time points (Protein IDs: 215181, 450375, 443894), and two fasciclin genes decreased in expression (Protein IDs: 464270, 415159). One of these genes (Protein ID: 450375) shares sequence similarity with an astaxanthin-binding fasciclin protein in *Scenedesmus* sp. strain Ki-4 that is involved in protection from photooxidative stress [78]. Given that fasciclin domains are associated with flocculation [82], polysaccharide binding [83] and other extracellular interactions [80], protective effects are likely derived from extracellular polysaccharide matrix (EPM) and/or palmelloid (or palmelloid-like cell clustering events) formation known to be protective in high salinity environments [84,85]. While most well studied in salinity response, EPM or palmelloid formation could also be protective in cold environments by reducing convective heat loss. Therefore, the transcriptional changes of genes containing fasciclin domains may play an important role in adaptation to stress in *M. minutum*.

3.6 Changes in metabolites levels in response to stress

Metabolomic profiling was performed under the perturbation conditions to investigate changes in carbon, nitrogen, and amino acid metabolism in *M. minutum*. Over a hundred unique metabolites were identified based on their exact mass and comparison with the retention time and spectra of

internal standards. Out of the detected metabolites, 45 were significantly different in abundance, measured by ion intensity (peak height), across multiple perturbation samples relative to the control (Figure 6A). The metabolites in the salt stress samples showed a more consistent pattern over time, whereas the changes in metabolite abundance were more transient in the cold stress sample during the period monitored. Under salt stress, there was an increase in levels of common osmolytes, betaine, proline, and trehalose. In addition, some amino acids, like arginine, orthinine, asparagine, glutamine, and aspartic acid showed an increase in abundance in the salt stress samples. A metabolite putatively identified as ophthalmic acid, a tripeptide analogue of glutathione, was also increased in abundance in the salt stress samples. Ophthalmic acid has been associated with oxidative stress in plant species, which is linked to salt stress due to accumulation of reactive oxygen species [86].

Accumulation of osmolytes coincided with increased expression of some genes directly involved in the synthesis of these metabolites. Several genes in proline biosynthesis were upregulated under the salt stress, including pyrroline 5-carboxylate reductase (Protein Id: 450819) with an average Log_2FC of 0.69 and pyrroline 5-carboxylate synthase (Protein Id: 222064) with an average Log_2FC of 1.19. Choline dehydrogenase (Protein id: 420590), which catalyzes the conversion of choline to betaine aldehyde, was also upregulated under salt stress with an average Log_2FC of 1.27.

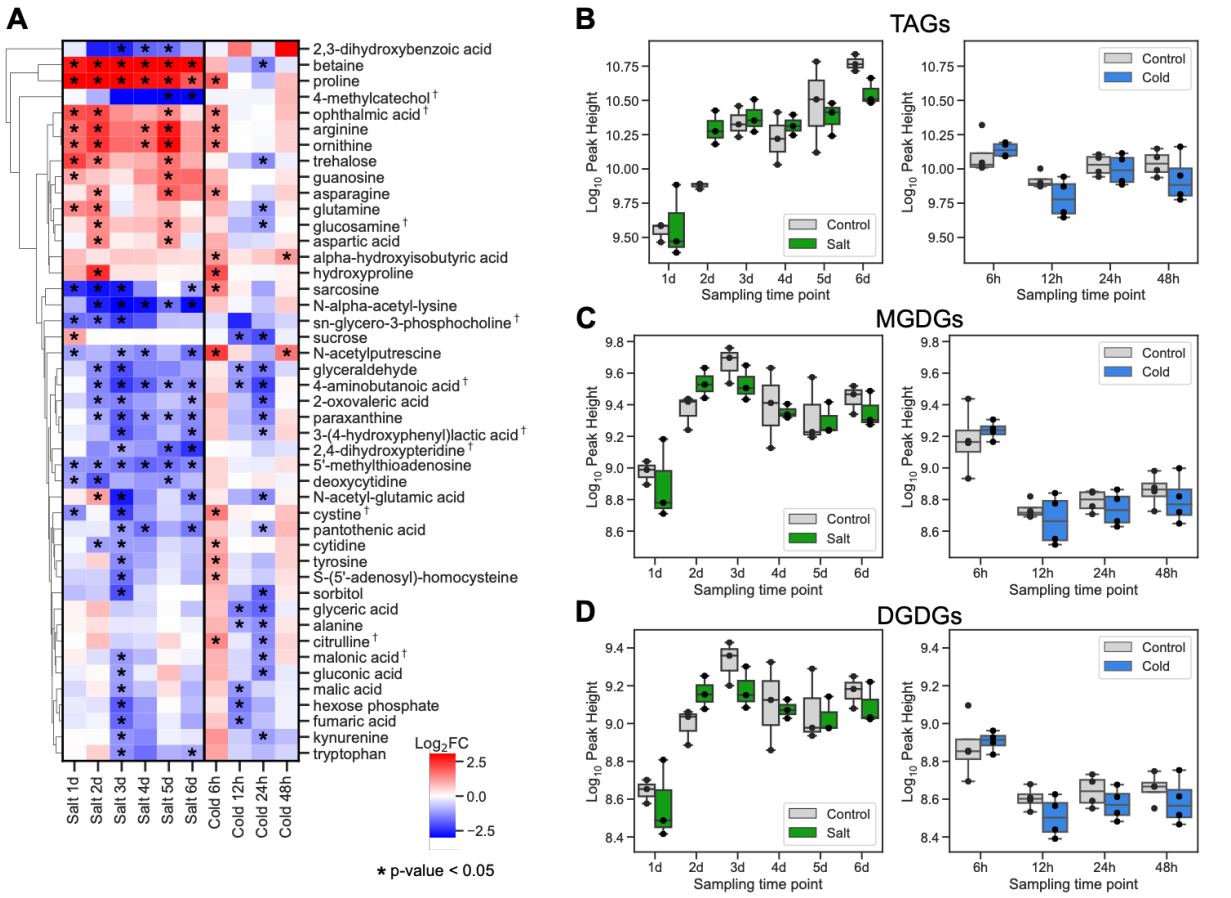

Figure 6. Metabolomic profiling by liquid chromatography tandem mass spectrometry under salt stress and cold stress over time. A) Fold change of metabolite levels comparing treatment samples to controls at the corresponding time points. † indicates that metabolites were putatively identified. B) Abundance measured by ion intensity (peak height) of triacylglycerols (TAGs) at each sampling time point under salt perturbation (left) and cold temperature perturbation (right). C) Abundance measured by ion intensity (peak height) of monogalactosyl diacylglycerols (MGDGs) at each sampling time point. D) Abundance measured by ion intensity (peak height) of digalactosyl diacylglycerols (DGDGs) at each sampling time point. Midline of box plots show the median, and individual points represent 3-4 biological replicates.

Levels of triacylglycerols (TAGs), monogalactosyl diacylglycerols (MGDGs), and digalactosyl diacylglycerols (DGDGs) were measured using reverse phase liquid chromatography and MS/MS (Figure 6B). In the salt stress experiment, lipids increased in overall abundance after the first day of the sampling period in both high salt and control samples. In the cold stress experiment, lipids dropped in abundance after 6h. The cold stress samples and control samples

showed minimal differences in abundance at corresponding time points. The distribution of abundance for each detected lipid species under the stress condition compared to control is shown in the Supplementary Figures 6-9. No noticeable pattern was observed comparing individual lipid levels in the samples grown under salt and cold stress conditions to the control samples. While these conditions did not result in accumulation of lipids in *M. minutum*, it is possible that more extreme conditions or different types of stress could induce changes in lipid metabolism.

TAG composition varies across different microalgae strains [87], and the fatty acid profile can affect the biofuel properties, such as oxidative stability and viscosity [88]. Previous work investigating lipid levels in *Scenedesmus* sp. NREL 46B-D3 showed an increase in abundance of some TAGs in response to temperature stress [12]. The TAG, MGDG, and DGDG composition of *M. minutum* differed from the NREL 46B-D3 strain (Supplementary Figure 10A-C). In *M. minutum*, several of the most abundant TAG species, such as TAG(54:2), TAG(54:3), TAG(54:4), TAG(54:5), TAG(54:6), and TAG(56:4), had a higher proportion of the total TAG composition compared to the NREL 46B-D3 strain (Supplementary Figure 10A). Other TAG species, TAG(50:1), TAG(52:3), and TAG(52:4), had a higher proportion of the total TAG composition in the 46B-D3. For some MGDGs and DGDGs, the proportions in 46B-D3 under cold stress appeared more similar to the proportion observed in *M. minutum* (Supplementary Figure 10B-C). The most abundant MGDG in *M. minutum*, MGDG(34:7) at 57% of total MGDGs on average, made up a higher percentage of total MGDGs than the percentage observed in the 46B-D3 strain. However, the percentage of MGDG(34:7) out of the total MGDGs in the 46B-D3 strain was 38% in the cold samples, which was increased relative to the percentage in

the control samples at 10% of total MGDGs. MGDG(34:4), MGDG(34:5), and DGDG(34:4) were lower in proportion in the cold samples than in the control samples in the NREL 46B-D3 strain, which is more similar to the relative proportions observed in the cold and control samples in *M. minutum*. While *M. minutum* showed minimal changes in lipid composition in response to cold stress, *Scenedesmus* sp. NREL 46B-D3 had an altered composition when exposed to colder temperatures. As lipid composition has been shown to be important for cold tolerance [89,90], the composition of *M. minutum* may be better suited for adaptation to cold temperature

3.7 Metabolic modeling

The functional annotations, transcriptomics and metabolomics information were utilized to construct a metabolic model for *M. minutum.* A total of 3,738 proteins were annotated with a particular enzymatic activity, spanning a total of 1,496 unique Enzyme Commission numbers (Supplementary File 4). The metabolic model contained 2,661 metabolic reactions and 3,048 metabolites (Supplementary Figure 11). Flux Balance Analysis (FBA), commonly utilized to model the behavior of metabolic models under various scenarios, assumes conservation of mass and that all reactions are under steady state (at equilibrium) [91]. We used FBA to investigate the completeness of metabolic reconstruction to identify important reactions linked to metabolites of interest. FBA solutions using linear programming showed that the model is able to simulate algal biomass synthesis. To understand the metabolism of the osmolytes detected in the metabolomics analysis, choline and proline, using phenotypic phase analysis [92] we explored the production envelopes of key reactions involving these metabolites considering nitrogen as a limiting factor and their impact on the biomass objective function (Supplementary Figure 12). The feasible

solution space observed shows that the production of these metabolites is likely to be impacted under nitrogen limitation with slightly different outcomes for the optimal growth of the alga.

3.8 in silico gene deletion analysis

To detect proteins that are crucial for growth (or limited the growth function), flux modeling was conducted after gene knockout simulations via the Linear Minimization of Metabolic Adjustment (LMOMA) algorithm [93,94]. 354 essential genes were detected which drastically impaired growth. When compared with the stress-responsive ortholog genes (Figure 4) and metabolites (Figure 6), several key metabolic enzymes were identified, such as sucrose synthase (EC: 2.4.1.13; Protein ID: 454733), arginine decarboxylase (EC: 4.1.1.19; Protein ID 452877), and trehalose-phosphatase (EC: 3.1.3.12, Protein ID: 385325). While some of these enzymes did not show a high fold change difference across conditions (Supplementary File 5), when combined with the metabolomics evidence (Figure 6) and metabolic modeling, it is likely that they play an important role in the survival and growth of the algae under salt stress. Furthermore, metabolic modeling suggests that membrane transporters play an important role in the salinity stress, including the glucose/fructose:H^+ symporter, STP13, AtMRP2 glutathione transporter, and the AKT1 potassium channel. These transporters have documented roles in salinity and stress responses in plants [95–98]**.** Another key enzyme identified through flux modeling that has a well-known role in salinity homeostasis in *Arabidopsis* is the salt-sensitive 3'-phosphoadenosine-5'-phosphatase HAL2/SAL1 (EC: 3.1.3.7; Protein ID 399370). This enzyme is responsible for counteracting the toxic levels of sodium and potassium inside the cell and has been shown to play multiple roles in salinity regulation in plants [99]. Taken together, metabolic modeling (while still at its infancy in non-model algal species) offers an opportunity to integrate gene

expression and metabolomics data beyond standard analysis of differential expression and was therefore useful to pinpoint candidate genes involved in conferring stress resistance or increasing biomass yields under certain conditions, especially given that many of the metabolic pathways may only show subtle gene expression differences or complex regulatory behavior [100].

4. Conclusions

By sequencing and annotating the genome of *M. minutum* and performing global transcriptomic and metabolomic profiling to determine genes and pathways implicated in stress response, we identified over 3,000 strain-specific gene families and functional domains related to gene regulation unique to *Monoraphidium*. Genes like the sphingolipid delta-4 desaturase, which are absent in other related Chlorophyceae outside of this genus, and fasciclin-like proteins, which are upregulated under stress, may be important for its stress tolerant phenotype. Differential expression of *M. minutum* genes with stress-responsive orthologs in *Arabidopsis* identified some conserved responses under salt stress. The accumulation of osmolytes in response to salt stress coincided with increased expression of genes involved in proline and betaine biosynthesis. Metabolic modeling and flux balance analysis revealed a set of essential genes for the growth and survival of *Monoraphidium minutum,* corroborating findings of gene expression and metabolomics analyses. This multi-omics characterization of *M. minutum* along with a genome-scale metabolic modeling can facilitate strain improvement, leading to the development of robust algal strains for deployment in large-scale outdoor cultivation systems.

Data availability

The *Monoraphidium minutum* 26B-AM strain is available from the UTEX culture collection. The current genome assembly and annotations are publicly available at the JGI Algal Genome Portal PhycoCosm: https://phycocosm.jgi.doe.gov/Monmin1. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JALPTG000000000. The version described in this paper is version JALPTG010000000. Transcriptomics data have been deposited to SRA under the accession PRJNA812585.

Declaration of competing interests

The authors declare no competing interests.

Acknowledgements

We thank Michael Huesemann at PNNL for providing the *Monoraphidium minutum* 26B-AM strain. This research was supported by the Bioenergy Technology Office (BETO) within the Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Agreements NL0032355 and NL0032266. The work (proposal 10.46936/10.25585/60001010) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] M. Huesemann, S. Edmundson, S. Gao, S. Negi, H. Daligaut, T. Dale, C. Gleasner, S. Starkenburg, A. Gutknecht, J. Freeman, W. Louie, S. McGuire, DISCOVR Strain Pipeline Tier I Screening: Maximum Specific Growth Rate as a Function of Temperature and Salinity for 38 Candidate Microalgae for Biofuels Production, Algal Research. (2022).

[2] K. Ogden, D. Anderson, S. Simpson, W. Van Voorheis, J. Brown, M. Huesemann, M. Kacira, R. Skaggs, P. Waller, Regional Algal Feedstock Testebed (RAFT) Final Report, Univ. of Arizona, Tucson, AZ (United States), 2019. https://doi.org/10.2172/1492217.

[3] R. Davis, B. Klein, Algal Biomass Production via Open Pond Algae Farm Cultivation: 2020 State of Technology and Future Research, National Renewable Energy Laboratory, 2021.

[4] C. Bogen, A. Al-Dilaimi, A. Albersmeier, J. Wichmann, M. Grundmann, O. Rupp, K.J. Lauersen, O. Blifernez-Klassen, J. Kalinowski, A. Goesmann, J.H. Mussgnug, O. Kruse, Reconstruction of the lipid metabolism for the microalga Monoraphidium neglectum from its genome sequence reveals characteristics suitable for biofuel production, BMC Genomics. 14 (2013) 926. https://doi.org/10.1186/1471-2164-14-926.

[5] S. Chaichalerm, P. Pokethitiyook, W. Yuan, M. Meetam, K. Sritong, W. Pugkaew, K. Kungvansaichol, M. Kruatrachue, P. Damrongphol, Culture of microalgal strains isolated from natural habitats in Thailand in various enriched media, Appl. Energy. 89 (2012) 296–302. https://doi.org/10.1016/j.apenergy.2011.07.028.

[6] X. Yu, P. Zhao, C. He, J. Li, X. Tang, J. Zhou, Z. Huang, Isolation of a novel strain of Monoraphidium sp. and characterization of its potential application as biodiesel feedstock, Bioresour. Technol. 121 (2012) 256–262. https://doi.org/10.1016/j.biortech.2012.07.002.

[7] K.Y. Teh, C.L.W. Afifudeen, A. Aziz, L.L. Wong, S.H. Loh, T.S. Cha, De novo whole genome sequencing data of two mangrove-isolated microalgae from Terengganu coastal waters, Data Brief. 27 (2019) 104680. https://doi.org/10.1016/j.dib.2019.104680.

[8] K. Fučíková, P.O. Lewis, L.A. Lewis, Chloroplast phylogenomic data from the green algal order Sphaeropleales (Chlorophyceae, Chlorophyta) reveal complex patterns of sequence evolution, Mol. Phylogenet. Evol. 98 (2016) 176–183. https://doi.org/10.1016/j.ympev.2016.01.022.

[9] A. Ben-Amotz, T.G. Tornabene, W.H. Thomas, Chemical profile of selected species of microalgae with emphasis on lipids1, J. Phycol. 21 (2004) 72–81. https://doi.org/10.1111/j.0022-3646.1985.00072.x.

[10] J.O. Nalley, D.R. O'Donnell, E. Litchman, Temperature effects on growth rates and fatty acid content in freshwater algae and cyanobacteria, Algal Research. 35 (2018) 500–507. https://doi.org/10.1016/j.algal.2018.09.018.

[11] S.P. Singh, P. Singh, Effect of temperature and light on the growth of algae species: A review, Renewable Sustainable Energy Rev. 50 (2015) 431–444. https://doi.org/10.1016/j.rser.2015.05.024.

[12] S. Calhoun, T.A.S. Bell, L.R. Dahlin, Y. Kunde, K. LaButti, K.B. Louie, A. Kuftin, D. Treen, D. Dilworth, S. Mihaltcheva, C. Daum, B.P. Bowen, T.R. Northen, M.T. Guarnieri, S.R. Starkenburg, I.V. Grigoriev, A multi-omic characterization of temperature stress in a halotolerant Scenedesmus strain for algal biotechnology, Commun Biol. 4 (2021) 333. https://doi.org/10.1038/s42003-021-01859-y.

[13] I.V. Grigoriev, R. Nikitin, S. Haridas, A. Kuo, R. Ohm, R. Otillar, R. Riley, A. Salamov, X. Zhao, F. Korzeniewski, T. Smirnova, H. Nordberg, I. Dubchak, I. Shabalov, MycoCosm portal: gearing up for 1000 fungal genomes, Nucleic Acids Res. 42 (2014) D699-704.

https://doi.org/10.1093/nar/gkt1183.
[14] A. Kuo, B. Bushnell, I.V. Grigoriev, Fungal Genomics: Sequencing and Annotation, in: F.M. Martin (Ed.), Advances in Botanical Research, Academic Press, 2014: pp. 1–52. https://doi.org/10.1016/B978-0-12-397940-7.00001-X.
[15] A.F.A. Smit, R. Hubley, P. Green, RepeatMasker Open-3.0, 1996-2010. https://www.repeatmasker.org.
[16] J. Jurka, V.V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase Update, a database of eukaryotic repetitive elements, Cytogenet. Genome Res. 110 (2005) 462–467. https://doi.org/10.1159/000084979.
[17] A.L. Price, N.C. Jones, P.A. Pevzner, De novo identification of repeat families in large genomes, Bioinformatics. 21 Suppl 1 (2005) i351-8. https://doi.org/10.1093/bioinformatics/bti1018.
[18] A.A. Salamov, V.V. Solovyev, Ab initio gene finding in Drosophila genomic DNA, Genome Res. 10 (2000) 516–522. https://doi.org/10.1101/gr.10.4.516.
[19] E. Birney, R. Durbin, Using GeneWise in the Drosophila annotation experiment, Genome Res. 10 (2000) 547–548. https://doi.org/10.1101/gr.10.4.547.
[20] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, Protein Eng. 10 (1997) 1–6.
[21] K. Melen, A. Krogh, G. von Heijne, Reliability measures for membrane protein topology prediction algorithms, J. Mol. Biol. 327 (2003) 735–744. https://doi.org/10.1016/s0022-2836(03)00182-7.
[22] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, InterProScan: protein domains identifier, Nucleic Acids Res. 33 (2005) W116-20. https://doi.org/10.1093/nar/gki442.
[23] UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Res. 49 (2021) D480–D489. https://academic.oup.com/nar/article-abstract/49/D1/D480/6006196.
[24] M. Kanehisa, Y. Sato, M. Furumichi, K. Morishima, M. Tanabe, New approach for understanding genome variations in KEGG, Nucleic Acids Res. 47 (2018) D590–D595. https://doi.org/10.1093/nar/gky962.
[25] E.V. Koonin, N.D. Fedorova, J.D. Jackson, A.R. Jacobs, D.M. Krylov, K.S. Makarova, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya, B.S. Rao, I.B. Rogozin, S. Smirnov, A.V. Sorokin, A.V. Sverdlov, S. Vasudevan, Y.I. Wolf, J.J. Yin, D.A. Natale, A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes, Genome Biol. 5 (2004) R7. https://doi.org/10.1186/gb-2004-5-2-r7.
[26] M.H. Saier Jr, V.S. Reddy, B.V. Tsu, M.S. Ahmed, C. Li, G. Moreno-Hagelsieb, The Transporter Classification Database (TCDB): recent advances, Nucleic Acids Res. 44 (2016) D372-9. https://doi.org/10.1093/nar/gkv1103.
[27] Gene Ontology Consortium, Gene Ontology Consortium: going forward, Nucleic Acids Res. 43 (2015) D1049-56. https://doi.org/10.1093/nar/gku1179.
[28] M. Seppey, M. Manni, E.M. Zdobnov, BUSCO: Assessing Genome Assembly and Annotation Completeness, in: M. Kollmar (Ed.), Gene Prediction: Methods and Protocols, Springer New York, New York, NY, 2019: pp. 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.
[29] P.J.A. Cock, T. Antao, J.T. Chang, B.A. Chapman, C.J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M.J.L. de Hoon, Biopython: freely available Python tools for computational molecular biology and bioinformatics, Bioinformatics. 25 (2009)

1422–1423. https://doi.org/10.1093/bioinformatics/btp163.
[30] B. Bjellqvist, B. Basse, E. Olsen, J.E. Celis, Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions, Electrophoresis. 15 (1994) 529–539. https://doi.org/10.1002/elps.1150150171.
[31] B. Bjellqvist, G.J. Hughes, C. Pasquali, N. Paquet, F. Ravier, J.C. Sanchez, S. Frutiger, D. Hochstrasser, The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences, Electrophoresis. 14 (1993) 1023–1031. https://doi.org/10.1002/elps.11501401163.
[32] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (2019) 238. https://doi.org/10.1186/s13059-019-1832-y.
[33] I.V. Grigoriev, R.D. Hayes, S. Calhoun, B. Kamel, A. Wang, S. Ahrendt, S. Dusheyko, R. Nikitin, S.J. Mondo, A. Salamov, I. Shabalov, A. Kuo, PhycoCosm, a comparative algal genomics resource, Nucleic Acids Res. 49 (2021) D1004–D1011. https://doi.org/10.1093/nar/gkaa898.
[34] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics. 30 (2014) 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
[35] B.E. Suzek, Y. Wang, H. Huang, P.B. McGarvey, C.H. Wu, UniProt Consortium, UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches, Bioinformatics. 31 (2015) 926–932. https://doi.org/10.1093/bioinformatics/btu739.
[36] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780. https://doi.org/10.1093/molbev/mst010.
[37] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274. https://doi.org/10.1093/molbev/msu300.
[38] B.Q. Minh, M.A.T. Nguyen, A. von Haeseler, Ultrafast approximation for phylogenetic bootstrap, Mol. Biol. Evol. 30 (2013) 1188–1195. https://doi.org/10.1093/molbev/mst024.
[39] I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, Nucleic Acids Res. 49 (2021) W293–W296. https://doi.org/10.1093/nar/gkab301.
[40] T.A.S. Bell, L. Doig, B.M. Peyton, R. Gerlach, M.W. Fields, Contributions of the microbial community to algal biomass and biofuel productivity in a wastewater treatment lagoon system, Algal Research. 39 (2019) 101461. https://doi.org/10.1016/j.algal.2019.101461.
[41] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (2019) 907–915. https://doi.org/10.1038/s41587-019-0201-4.
[42] F. Ramírez, D.P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis, Nucleic Acids Res. 44 (2016) W160-5. https://doi.org/10.1093/nar/gkw257.
[43] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics. 30 (2014) 923–930. https://doi.org/10.1093/bioinformatics/btt656.
[44] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550. https://doi.org/10.1186/s13059-

014-0550-8.
[45] G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS. 16 (2012) 284–287. https://doi.org/10.1089/omi.2011.0118.
[46] E.G. Bligh, W.J. Dyer, A rapid method of total lipid extraction and purification, Can. J. Biochem. Physiol. 37 (1959) 911–917. https://doi.org/10.1139/o59-099.
[47] Y. Yao, T. Sun, T. Wang, O. Ruebel, T. Northen, B.P. Bowen, Analysis of metabolomics datasets with high-performance computing and metabolite atlases, Metabolites. 5 (2015) 431–442. https://doi.org/10.3390/metabo5030431.
[48] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: a metabolite mass spectral database, Ther. Drug Monit. 27 (2005) 747–751. https://doi.org/10.1097/01.ftd.0000179845.53213.39.
[49] C.Y. Botte, Y. Yamaryo-Botte, J. Janouskovec, T. Rupasinghe, P.J. Keeling, P. Crellin, R.L. Coppel, E. Marechal, M.J. McConville, G.I. McFadden, Identification of plant-like galactolipids in Chromera velia, a photosynthetic relative of malaria parasites, J. Biol. Chem. 286 (2011) 29893–29903. https://doi.org/10.1074/jbc.M111.254979.
[50] A.M. McAnoy, C.C. Wu, R.C. Murphy, Direct qualitative analysis of triacylglycerols by electrospray mass spectrometry using a linear ion trap, J. Am. Soc. Mass Spectrom. 16 (2005) 1498–1509. https://doi.org/10.1016/j.jasms.2005.04.017.
[51] C. Claudel-Renard, C. Chevalet, T. Faraut, D. Kahn, Enzyme-specific profiles for genome annotation: PRIAM, Nucleic Acids Res. 31 (2003) 6633–6639. https://doi.org/10.1093/nar/gkg847.
[52] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (2019) D309–D314. https://doi.org/10.1093/nar/gky1085.
[53] M. Kanehisa, Y. Sato, KEGG Mapper for inferring cellular functions from protein sequences, Protein Sci. 29 (2020) 28–35. https://doi.org/10.1002/pro.3711.
[54] P.D. Karp, R. Billington, R. Caspi, C.A. Fulcher, M. Latendresse, A. Kothari, I.M. Keseler, M. Krummenacker, P.E. Midford, Q. Ong, W.K. Ong, S.M. Paley, P. Subhraveti, The BioCyc collection of microbial genomes and metabolic pathways, Brief. Bioinform. 20 (2019) 1085–1093. https://doi.org/10.1093/bib/bbx085.
[55] A. Ebrahim, J.A. Lerman, B.O. Palsson, D.R. Hyduke, COBRApy: COnstraints-Based Reconstruction and Analysis for Python, BMC Syst. Biol. 7 (2013) 74. https://doi.org/10.1186/1752-0509-7-74.
[56] I. Thiele, B.Ø. Palsson, A protocol for generating a high-quality genome-scale metabolic reconstruction, Nat. Protoc. 5 (2010) 93–121. https://doi.org/10.1038/nprot.2009.203.
[57] S. Prigent, C. Frioux, S.M. Dittami, S. Thiele, A. Larhlimi, G. Collet, F. Gutknecht, J. Got, D. Eveillard, J. Bourdon, F. Plewniak, T. Tonon, A. Siegel, Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks, PLoS Comput. Biol. 13 (2017) e1005276. https://doi.org/10.1371/journal.pcbi.1005276.
[58] V. Satish Kumar, M.S. Dasika, C.D. Maranas, Optimization based automated curation of metabolic reconstructions, BMC Bioinformatics. 8 (2007) 212. https://doi.org/10.1186/1471-2105-8-212.
[59] D. Sarkar, M. Landa, A. Bandyopadhyay, H.B. Pakrasi, J.P. Zehr, C.D. Maranas,

Elucidation of trophic interactions in an unusual single-cell nitrogen-fixing symbiosis using metabolic modeling, PLoS Comput. Biol. 17 (2021) e1008983. https://doi.org/10.1371/journal.pcbi.1008983.
[60] M. Hucka, A. Finney, H.M. Sauro, H. Bolouri, J.C. Doyle, H. Kitano, A.P. Arkin, B.J. Bornstein, D. Bray, A. Cornish-Bowden, A.A. Cuellar, S. Dronov, E.D. Gilles, M. Ginkel, V. Gor, I.I. Goryanin, W.J. Hedley, T.C. Hodgman, J.-H. Hofmeyr, P.J. Hunter, N.S. Juty, J.L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L.M. Loew, D. Lucio, P. Mendes, E. Minch, E.D. Mjolsness, Y. Nakayama, M.R. Nelson, P.F. Nielsen, T. Sakurada, J.C. Schaff, B.E. Shapiro, T.S. Shimizu, H.D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, SBML Forum, The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, Bioinformatics. 19 (2003) 524–531. https://doi.org/10.1093/bioinformatics/btg015.
[61] G. Gelius-Dietrich, A.A. Desouki, C.J. Fritzemeier, M.J. Lercher, Sybil--efficient constraint-based modelling in R, BMC Syst. Biol. 7 (2013) 125. https://doi.org/10.1186/1752-0509-7-125.
[62] C. Lieven, M.E. Beber, B.G. Olivier, F.T. Bergmann, M. Ataman, P. Babaei, J.A. Bartell, L.M. Blank, S. Chauhan, K. Correia, C. Diener, A. Dräger, B.E. Ebert, J.N. Edirisinghe, J.P. Faria, A.M. Feist, G. Fengos, R.M.T. Fleming, B. García-Jiménez, V. Hatzimanikatis, W. van Helvoirt, C.S. Henry, H. Hermjakob, M.J. Herrgård, A. Kaafarani, H.U. Kim, Z. King, S. Klamt, E. Klipp, J.J. Koehorst, M. König, M. Lakshmanan, D.-Y. Lee, S.Y. Lee, S. Lee, N.E. Lewis, F. Liu, H. Ma, D. Machado, R. Mahadevan, P. Maia, A. Mardinoglu, G.L. Medlock, J.M. Monk, J. Nielsen, L.K. Nielsen, J. Nogales, I. Nookaew, B.O. Palsson, J.A. Papin, K.R. Patil, M. Poolman, N.D. Price, O. Resendis-Antonio, A. Richelle, I. Rocha, B.J. Sánchez, P.J. Schaap, R.S. Malik Sheriff, S. Shoaie, N. Sonnenschein, B. Teusink, P. Vilaça, J.O. Vik, J.A.H. Wodke, J.C. Xavier, Q. Yuan, M. Zakhartsev, C. Zhang, MEMOTE for standardized genome-scale metabolic model testing, Nat. Biotechnol. 38 (2020) 272–276. https://doi.org/10.1038/s41587-020-0446-y.
[63] S. Suzuki, H. Yamaguchi, N. Nakajima, M. Kawachi, Raphidocelis subcapitata (= Pseudokirchneriella subcapitata) provides an insight into genome evolution and environmental adaptations in the Sphaeropleales, Sci. Rep. 8 (2018) 1–13. https://www.nature.com/articles/s41598-018-26331-6.
[64] M.S. Roth, S.J. Cokus, S.D. Gallaher, A. Walter, D. Lopez, E. Erickson, B. Endelman, D. Westcott, C.A. Larabell, S.S. Merchant, M. Pellegrini, K.K. Niyogi, Chromosome-level genome assembly and transcriptome of the green alga Chromochloris zofingiensis illuminates astaxanthin production, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) E4296–E4305. https://doi.org/10.1073/pnas.1619928114.
[65] S.R. Starkenburg, J.E.W. Polle, B. Hovde, H.E. Daligault, K.W. Davenport, A. Huang, P. Neofotis, Z. McKie-Krisberg, Draft Nuclear Genome, Complete Chloroplast Genome, and Complete Mitochondrial Genome for the Biofuel/Bioproduct Feedstock Species Scenedesmus obliquus Strain DOE0152z, Genome Announc. 5 (2017). https://doi.org/10.1128/genomeA.00617-17.
[66] S. Dharshini, V.M. Manoj, G.S. Suresha, J.A. Narayan, T.S.S. Padmanabhan, R. Kumar, M.R. Meena, M. Manickavasagam, B. Ram, C. Appunu, Isolation and Characterization of Nuclear Localized Abiotic Stress Responsive Cold Regulated Gene 413 (SsCor413) from Saccharum spontaneum, Plant Mol. Biol. Rep. 38 (2020) 628–640. https://doi.org/10.1007/s11105-020-01224-z.

[67] B. Karpinska, M. Karlsson, H. Schinkel, S. Streller, K.H. Süss, M. Melzer, G. Wingsle, A novel superoxide dismutase with a high isoelectric point in higher plants. expression, regulation, and protein localization, Plant Physiol. 126 (2001) 1668–1677. https://doi.org/10.1104/pp.126.4.1668.
[68] V. Srivastava, H. Schinkel, J. Witzell, M. Hertzberg, M. Torp, M.K. Srivastava, B. Karpinska, M. Melzer, G. Wingsle, Downregulation of high-isoelectric-point extracellular superoxide dismutase mediates alterations in the metabolism of reactive oxygen species and developmental disturbances in hybrid aspen, Plant J. 49 (2007) 135–148. https://doi.org/10.1111/j.1365-313X.2006.02943.x.
[69] S. Chakrabortee, F. Meersman, G.S. Kaminski Schierle, C.W. Bertoncini, B. McGee, C.F. Kaminski, A. Tunnacliffe, Catalytic and chaperone-like functions in an intrinsically disordered protein associated with desiccation tolerance, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 16084–16089. https://doi.org/10.1073/pnas.1006276107.
[70] T.Z. Berardini, L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, E. Huala, The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome, Genesis. 53 (2015) 474–485. https://doi.org/10.1002/dvg.22877.
[71] J. Shanklin, E.B. Cahoon, Desaturation and related modifications of fatty acids, Annu. Rev. Plant Physiol. Plant Mol. Biol. 49 (1998) 611–641. https://doi.org/10.1146/annurev.arplant.49.1.611.
[72] M. Chen, J.E. Markham, E.B. Cahoon, Sphingolipid Δ8 unsaturation is important for glucosylceramide biosynthesis and low-temperature performance in Arabidopsis, Plant J. 69 (2012) 769–781. https://doi.org/10.1111/j.1365-313X.2011.04829.x.
[73] H.C. Resemann, C. Herrfurth, K. Feussner, E. Hornung, A.K. Ostendorf, J. Gömann, J. Mittag, N. van Gessel, J. de Vries, J. Ludwig-Müller, J. Markham, R. Reski, I. Feussner, Convergence of sphingolipid desaturation across over 500 million years of plant evolution, Nat Plants. 7 (2021) 219–232. https://doi.org/10.1038/s41477-020-00844-3.
[74] Y. Zhou, L. Zeng, X. Fu, X. Mei, S. Cheng, Y. Liao, R. Deng, X. Xu, Y. Jiang, X. Duan, S. Baldermann, Z. Yang, The sphingolipid biosynthetic enzyme Sphingolipid delta8 desaturase is important for chilling resistance of tomato, Sci. Rep. 6 (2016) 38742. https://doi.org/10.1038/srep38742.
[75] Y. Wang, X. Liu, H. Gao, H.-M. Zhang, A.-Y. Guo, J. Xu, X. Xu, Early Stage Adaptation of a Mesophilic Green Alga to Antarctica: Systematic Increases in Abundance of Enzymes and LEA Proteins, Mol. Biol. Evol. 37 (2020) 849–863. https://doi.org/10.1093/molbev/msz273.
[76] J. Capel, J.A. Jarillo, J. Salinas, J.M. Martínez-Zapater, Two homologous low-temperature-inducible genes from Arabidopsis encode highly hydrophobic proteins, Plant Physiol. 115 (1997) 569–576. https://doi.org/10.1104/pp.115.2.569.
[77] J. Medina, R. Catalá, J. Salinas, Developmental and stress regulation of RCI2A and RCI2B, two cold-inducible genes of arabidopsis encoding highly conserved hydrophobic proteins, Plant Physiol. 125 (2001) 1655–1666. https://doi.org/10.1104/pp.125.4.1655.
[78] S. Kawasaki, K. Mizuguchi, M. Sato, T. Kono, H. Shimizu, A novel astaxanthin-binding photooxidative stress-inducible aqueous carotenoprotein from a eukaryotic microalga isolated from asphalt in midsummer, Plant Cell Physiol. 54 (2013) 1027–1040. https://doi.org/10.1093/pcp/pct080.
[79] H. Shi, Y. Kim, Y. Guo, B. Stevenson, J.-K. Zhu, The Arabidopsis SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion, Plant Cell. 15 (2003) 19–32. https://doi.org/10.1105/tpc.007872.

[80] O. Huber, M. Sumper, Algal-CAMs: isoforms of a cell adhesion molecule in embryos of the alga Volvox with homology to Drosophila fasciclin I, EMBO J. 13 (1994) 4212–4222. https://www.ncbi.nlm.nih.gov/pubmed/7925267.
[81] K.L. Johnson, B.J. Jones, A. Bacic, C.J. Schultz, The fasciclin-like arabinogalactan proteins of Arabidopsis. A multigene family of putative cell adhesion molecules, Plant Physiol. 133 (2003) 1911–1925. https://doi.org/10.1104/pp.103.031237.
[82] B.-L. Chen, W. Mhuantong, S.-H. Ho, J.-S. Chang, X.-Q. Zhao, F.-W. Bai, Genome sequencing, assembly, and annotation of the self-flocculating microalga Scenedesmus obliquus AS-6-11, BMC Genomics. 21 (2020) 743. https://doi.org/10.1186/s12864-020-07142-4.
[83] E. Liu, C.P. MacMillan, T. Shafee, Y. Ma, J. Ratcliffe, A. van de Meene, A. Bacic, J. Humphries, K.L. Johnson, Fasciclin-Like Arabinogalactan-Protein 16 (FLA16) Is Required for Stem Development in Arabidopsis, Front. Plant Sci. 11 (2020) 615392. https://doi.org/10.3389/fpls.2020.615392.
[84] P. Shetty, M.M. Gitau, G. Maróti, Salinity Stress Responses and Adaptation Mechanisms in Eukaryotic Green Microalgae, Cells. 8 (2019). https://doi.org/10.3390/cells8121657.
[85] F. de Carpentier, S.D. Lemaire, A. Danon, When Unity Is Strength: The Strategies Used by Chlamydomonas to Survive Environmental Stresses, Cells. 8 (2019). https://doi.org/10.3390/cells8111307.
[86] L. Servillo, D. Castaldo, A. Giovane, R. Casale, N. D'Onofrio, D. Cautela, M.L. Balestrieri, Ophthalmic acid is a marker of oxidative stress in plants as in animals, Biochim. Biophys. Acta Gen. Subj. 1862 (2018) 991–998. https://doi.org/10.1016/j.bbagen.2018.01.015.
[87] K.M. MacDougall, J. McNichol, P.J. McGinn, S.J.B. O'Leary, J.E. Melanson, Triacylglycerol profiling of microalgae strains for biofuel feedstock by liquid chromatography-high-resolution mass spectrometry, Anal. Bioanal. Chem. 401 (2011) 2609–2616. https://doi.org/10.1007/s00216-011-5376-6.
[88] T.P. Durrett, C. Benning, J. Ohlrogge, Plant triacylglycerols as feedstocks for the production of biofuels, Plant J. 54 (2008) 593–607. https://doi.org/10.1111/j.1365-313X.2008.03442.x.
[89] D.L. Falcone, J.P. Ogas, C.R. Somerville, Regulation of membrane fatty acid composition by temperature in mutants of Arabidopsis with alterations in membrane lipid composition, BMC Plant Biol. 4 (2004) 17. https://doi.org/10.1186/1471-2229-4-17.
[90] C. Barrero-Sicilia, S. Silvestre, R.P. Haslam, L.V. Michaelson, Lipid remodelling: Unravelling the response to cold stress in Arabidopsis and its extremophile relative Eutrema salsugineum, Plant Sci. 263 (2017) 194–200. https://doi.org/10.1016/j.plantsci.2017.07.017.
[91] J.D. Orth, I. Thiele, B.Ø. Palsson, What is flux balance analysis?, Nat. Biotechnol. 28 (2010) 245–248. https://doi.org/10.1038/nbt.1614.
[92] J.S. Edwards, R. Ramakrishna, B.O. Palsson, Characterizing the metabolic phenotype: a phenotype phase plane analysis, Biotechnol. Bioeng. 77 (2002) 27–36. https://doi.org/10.1002/bit.10047.
[93] D. Segrè, D. Vitkup, G.M. Church, Analysis of optimality in natural and perturbed metabolic networks, Proceedings of the National Academy of Sciences. 99 (2002) 15112–15117. https://doi.org/10.1073/pnas.232349399.
[94] D. Alzoubi, A.A. Desouki, M.J. Lercher, Flux balance analysis with or without molecular crowding fails to predict two thirds of experimentally observed epistasis in yeast, Sci. Rep. 9 (2019) 11837. https://doi.org/10.1038/s41598-019-47935-6.
[95] K. Yamada, M. Kanai, Y. Osakabe, H. Ohiraki, K. Shinozaki, K. Yamaguchi-Shinozaki,

Monosaccharide absorption activity of Arabidopsis roots depends on expression profiles of transporter genes under high salinity conditions, J. Biol. Chem. 286 (2011) 43577–43586. https://doi.org/10.1074/jbc.M111.269712.
[96] N. Chakraborty, S. Mukherjee, A. Sarkar, P. Shaw, K. Acharya, 18 - Role of glutathione transporter in plants under stress, in: A. Roychoudhury, D.K. Tripathi, R. Deshmukh (Eds.), Transporters and Plant Osmotic Stress, Academic Press, 2021: pp. 345–364. https://doi.org/10.1016/B978-0-12-817958-1.00021-9.
[97] Y. Yang, Y. Guo, Unraveling salt stress signaling in plants, J. Integr. Plant Biol. 60 (2018) 796–804. https://doi.org/10.1111/jipb.12689.
[98] D. Golldack, F. Quigley, C.B. Michalowski, U.R. Kamasani, H.J. Bohnert, Salinity stress-tolerant and -sensitive rice (Oryza sativa L.) regulate AKT1-type potassium channel transcripts differently, Plant Mol. Biol. 51 (2003) 71–81. https://doi.org/10.1023/a:1020763218045.
[99] R. Gil-Mascarell, J.M. López-Coronado, J.M. Bellés, R. Serrano, P.L. Rodríguez, The Arabidopsis HAL2-like gene family includes a novel sodium-sensitive phosphatase, Plant J. 17 (1999) 373–383. https://doi.org/10.1046/j.1365-313x.1999.00385.x.
[100] D.L. Rothman, S.C. Stearns, R.G. Shulman, Gene expression regulates metabolite homeostasis during the Crabtree effect: Implications for the adaptation and evolution of Metabolism, Proceedings of the National Academy of Sciences. 118 (2021) e2014013118. https://doi.org/10.1073/pnas.2014013118.