Integrative multi-omic and phenotypic analysis of open raceway pond production of *Monoraphidium minutum* 26B-AM reveals distinct stress signatures for scale-up and infection.

Georgios Kepesidis [a], Jenna Schambach [b], Daniel Yang [a], Elise Wilbourn [c], Thomas Sheffield [a], Tyler Eckles [c], Olivia Watt [c], Matthew P. Hirakawa [a], Todd W. Lane [c], Raga Krishnakumar [a+]

a Biosecurity and Bioassurance, Sandia National Laboratories, Livermore, 94550, U.S.A.

b Environmental Systems Biology, Sandia National Laboratories, Albuquerque, 87123, U.S.A.

c Bioresource & Environmental Security, Sandia National Laboratories, Livermore, 94550, U.S.A.

+ Corresponding author

Abstract

Background: Green microalgae, such as *Monoraphidium minutum* 26B-AM, have garnered significant commercial interest due to their high biomass production and lipid yield, providing promising candidates for various bioprocessing applications. However, the economic viability of large-scale algal cultivation in open raceway ponds is limited by biocontamination and environmental stressors, necessitating deeper understanding of the molecular mechanisms that underpin resilience and productivity in these systems. We hypothesized that the molecular signature associated with the cellular responses of *M. minutum* to environmental stressors will reveal critical information for the timely prediction of resilience and productivity in algal cultures within open pond systems.

Results: To test this hypothesis, we conducted a longitudinal multi-omic study, integrating transcriptomics, proteomics, metabolomics, and phenomics, to monitor the acclimation, growth dynamics, and pathogen responses of algal cultures in two 1000 L raceway ponds, before and after the introduction of a pathogen as a stressor. We identified a number of molecular patterns that correlate with changes in the algal environment, and we can track these changes within the ponds per time. Furthermore, we identify scale-up and infection-specific molecular pathways through integrated multi-omics, showing that most patterns are unique to each studied stressor/transition.

Conclusions: Ultimately, this study demonstrates the utility of multi-omics observations at scale, revealing unique signatures and laying the groundwork for developing molecular detection techniques and predictive models that can improve the sustainability and efficiency of large-scale algae biomass production.

Keywords

Multi-omics, algae, *Monoraphidium minutum*, ponds, transcriptomics, metabolomics, proteomics, biomarkers

1. Background

Microalgae are a polyphyletic group of unicellular photosynthetic aquatic organisms that have attracted commercial interest since the middle of the past century because of their potential to advance biotechnology and biomanufacturing processes. Compared to land plants, their biomass yield is greater [1], and their high content of useful metabolites, such as lipids, has increased their presence in multiple industrial bioprocesses, including the production of food, pharmaceuticals, cosmetics, bioplastics, and biofuels. The high production costs, however, curtail most of these applications, especially for low-cost and high-volume products such as biofuels, making the use of algae currently unfeasible and unsustainable and restricting their potential [2] [3] [4].

A low-energy and low-cost method for large-scale production of algal biomass, given the financial restrictions for algal cultivation mentioned earlier, is the paddlewheel-based open raceway pond cultivation system [5] [6]. These widely used systems allow the cultures to grow in direct relation to the

surrounding environment and the local weather conditions. Additionally, these ponds can be established on land that is not suitable for terrestrial agriculture, thereby utilizing a resource that would otherwise remain unused.

Despite the aforementioned advantages of these open systems, a number of barriers preclude the overall economic viability of bioprocesses within these systems. One crucial barrier is the susceptibility of the cultures to environmental "intruders", which threaten both the quality of the final product and the overall survival of the pond cultures [7]. These events lead to pond crashes and a significant increase in operational costs, strongly establishing the need for their timely prediction.

In combination with the growth of undesired microbial organisms, open pond systems are susceptible to unpredictable changes in the weather [8], as are other site-specific environmental factors, such as sand particles or dust, which change the composition of the culture media [9]. All these factors, which are directly related to the immediate exposure of the cultures to their biotic and abiotic surroundings, can lead to unpredictable outcomes in terms of biomass productivity [10]. In addition, the growth and productivity are not always aligned with the observations made at smaller scales [11]. Understanding the effects that all these variables have on culture outcomes is critical to optimally predict growth in industrial settings. Consequently, optimal prediction can guide the design of resilient culture systems, ultimately reducing the cost of each operation. Identification of exposure- and threat-agnostic biomarkers of stress as expressed by the microalgae themselves would allow for detailed monitoring and remediation of open cultures in real-time and understanding what molecular pathways might be involved in resilience.

In the current industrial landscape, enhancing the resilience (i.e., sustained productivity and adaptability in the face of potential stressors) within expansive open pond systems and optimizing our predictive tools are essential for maintaining the stability of algal production systems. By utilizing the algal cells as sensors of their environment, we identified key genetic and molecular pathways that can be engineered to bolster their resilience. Here, we performed an in-depth multi-omics analysis (transcriptomics, proteomics, metabolomics and phenomics) in order to further understand the molecular dynamics behind the acclimation, growth, and response to infection by a parasitoid amoeboaphelid species (PL3) that are known to cause pond crashes in two 1000 L open raceway ponds under environmentally-simulated conditions. For this study, we used the freshwater green microalga *Monoraphidium minutum* 26B-AM, a species currently studied as a promising winter species for biofuel feedstock because of its elevated lipid yield [12] [13] [14] [15] [16]. The genome of *M. minutum* has also been recently annotated [17]. Samples were collected for all the different analyses during a 5-day pond run, followed by an infection with the parasitoid amoeboaphelid species FDO1 [18] [19], which is known to cause pond crashes in real-life scenarios. In addition to the samples used for the molecular observations, daily measurements of internal conditions, as well as certain phenotypical outcomes, were also monitored to obtain a holistic view of the dynamics between environmental signaling, genetic regulation and phenotypical outcomes in a nonextensively studied species, such as *M. minutum*. By leveraging this rich dataset, we aimed to perform in-depth molecular observations on the dynamic resilience of algal cultures over time, adding to the ongoing efforts to increase stability and productivity in large-scale applications. This work reveals a number of molecular signatures and regulatory elements of interest, and lays the groundwork for proactive interventions that can sustain algal health and performance in the face of environmental challenges. In this study, in addition to phenotypical observations, molecular markers evaluated via multi-omics are indicative of distinct stages of open pond culture. We verified this method by demonstrating a known link between light and redox signaling and photosynthesis and translation-related genes in *M.*

minutum, in accordance with similar observations in closely related green algal species, confirming the applicability of the markers identified here across algal culture systems. Finally, we focused on the molecular pathways associated with the scale-up acclimation and the introduction of the pathogen into the system through integrated multi-omic analysis. While specific molecular networks of genes, proteins and metabolites were identified for each transition, we discovered unique signatures in response to the two stressors, revealing distinct signaling outcomes across stressors.

2. Methods

2.1 Strain and conditions

The *M. minutum* 26B-AM strain was initially collected and isolated as described previously [20], and it has been utilized in our laboratory as part of the DISCOVR strain library. Seed cultures of *M. minutum* (100 ml) were maintained in sterile 250 ml Erlenmeyer flasks, containing DISCOVR medium [20] under a 16:8 h light/dark cycle with continuous shaking at 110 rpm in a closed incubator with LED lighting. The light intensity during the light cycle was 100 µmol photons m^{-2} s^{-1} and the temperature was set between a maximum of 30 °C during the light period and a minimum of 23 °C during the dark period. The seed cultures were diluted regularly to maintain exponential phase growth.

2.2 Scale-up

The culture scale-up is graphically summarized in Fig.1. The seed cultures of *M. minutum* described above were used as inocula for 500 ml cultures grown in 1 L Erlenmeyer flasks. Progressively, the cultures were transitioned to 2 L bottles, 8 L carboys with stirring and air bubbling and 20 L carboys with bubbling. Each step lasted 10–15 days, until the inoculum was dense enough (approximately 2 million cells/ml) to be diluted to the next scale. Up to the 2 L scale, the cultures were grown under the same conditions as the initial seed cultures detailed in the previous section. For the 20 L carboys, the cultures were acclimated to relatively high light intensities (1500 µmol photons m^{-2} s^{-1}). The 20 L carboys were thermally acclimated between 22-28 °C under a 12/12 h dark/light cycle in a water bath, to match the conditions of the next scale.

Fig. 1: M.minutum scale-up. *A) Graphic representation of the scale-up. B) Picture of the 1000L open raceway pond on Day 0 of the culture run, after the inoculation of the media with the precultures.*

2.3 Cultivation and sample collection at the 1000 L opened raceway ponds

The cultures from the 20 L carboys (3 from each pond) were used as inocula for the 1000 L open raceway ponds (manuscript in preparation). The ponds were filled with tap water and the nutrients and salts were added via the modified 10% BG-11 recipe, as described previously [20]. A total of 60 L of the inoculum was added to each pond. The ponds were operated under a 12 h/12 h dark/light cycle, implemented with LED light panels, placed over the ponds, reaching 1500 µmol photons m^{-2} s^{-1} during the light period, under a sinusoidal light intensity program (measured in closest proximity to the source). The temperature was maintained within a temperature range of 22-28 °C, with the highest temperature being reached in the middle of the light period and the lowest in the middle of the dark period. Temperature control was achieved through a heat-exchanger lining on the walls of the ponds. The cultures were constantly mixed with a mechanical paddlewheel at 20 rpm. The pH was held constant at 7.2 via CO_2 sparging. The cells were grown under these conditions for 5 days and samples for the different omic analyses were collected

162 on a daily basis one hour after the lights were turned on. The phenotypic and conditional measurements
163 were taken directly after that.

164 On the 5th day of the cultivation, 2/3 of each pond was drained and fresh water with new nutrients was
165 added. In parallel, an inoculum of the parasitoid amoeboaphelid species FDO1 [19] [18] was added at a
166 0.05% volume concentration, in order to study the dynamics of a pond crash caused by this strain. This
167 specific strain has been isolated by the Arizona Center for Algae Technology and Innovation as the cause
168 of *Monoraphdium minutum* pond crashes, and it has not yet been fully characterized. Under laboratory
169 conditions, it caused *M. minutum* cultures to crash in 4-5 days. The determination of culture crashes was
170 visual, when the ponds had completely transitioned from green to brown.

171

172 *2.4 Phenotypic and conditional measurements*

173

174 All phenotypic and conditional measurements were taken on a daily basis in technical triplicates unless
175 otherwise stated. The cell count and cell size were determined with the Countess™ 3 FL Automated Cell
176 Counter (Thermo Fisher Scientific, Waltham, MA, USA) on a daily basis, from a sample collected from the
177 ponds one hour before the lights were turned on.

178 To measure chlorophyll fluorescence, 200 µl pond samples (in technical triplicates) in 96-well transparent
179 plates were measured at an excitation wavelength of 450 nm and an emission wavelength of 685 nm with
180 a Spark® Multimode Microplate Reader (TECAN, Männedorf, Switzerland). The chlorophyll content of the
181 ponds is presented in relative units.

182 For the ash-free dry weight (AFDW) measurements, the steps described in [21] were followed.

183 The total carbohydrate concentration was determined via the Total Carbohydrate Quantification Assay
184 (Abcam, Cambridge, UK), following the manufacturer's instructions. Samples in technical triplicates were
185 collected from each pond on a daily basis, alongside the samples for the omic analyses.

186 Samples were collected in triplicate on a daily basis from each pond for the quantification of total H3
187 histones. This measurement was performed with the Histone H3 Total Quantification Kit (Abcam,
188 Cambridge, UK), following the manufacturer’s protocol, and the colorimetric measurements were
189 conducted using the Spark® Multimode Microplate Reader (TECAN, Männedorf, Switzerland).

190 Samples for the determination of caspase activity were also collected in triplicate on a daily basis. For the
191 estimation of cell lysis and caspase activity, the EnzChek Caspase-3 Assay Kit #1 (Thermo Fisher Scientific,
192 Waltham, MA, USA) was used, with Z-DEVD-AMC as the substrate, according to the manufacturer’s
193 protocol. Dissolved oxygen (%DO), pH, conductivity, temperature and light intensity were measured in
194 each pond daily, immediately after sample collection. For each variable, we took the average of 9 collected
195 data points. The %DO and temperature were measured with the Waterproof Dissolved Oxygen Meter Pen
196 (Sper Scientific Direct, Scottsdale, AZ, USA), the pH and conductivity were measured with an Ohaus Starter
197 AB33M1-F pH & Conductivity Bench Meter (Parsippany, NJ, USA), and the light intensity was measured
198 with an LI-1500 Light Sensor Logger (Lincoln, NE, USA).

199

200 *2.5 Transcriptomics/ RNA extraction and cDNA library preparation*

201

202 The samples collected for the transcriptomic analyses were centrifuged after collection at 5000 rpm for 2
203 minutes at 4 °C, and the resulting cell pellets were stored at -20 °C until RNA extraction was performed.

204 To initiate cell lysis, the frozen pellets were thawed on ice and resuspended in 750 µL of Qiagen RLT Buffer
205 supplemented with 1% β-mercaptoethanol from the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany).
206 The resuspended cells were then transferred to 2 mL tubes containing Lysing Matrix Y zirconium oxide
207 beads (MP Biomedicals, Santa Ana, CA, USA). Following a 10-second vortex, the samples were flash-frozen
208 in liquid nitrogen. After thawing on ice, the cells were lysed via a FastPrep-24 bead beater (MP
209 Biomedicals, Santa Ana, CA, USA) for 25 seconds at 40 Hz. The samples were subsequently centrifuged at
210 9000 rpm for 10 minutes at 4 °C, after which the clear supernatant was collected for RNA extraction. RNA
211 was extracted via the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's
212 protocol. The quantity of RNA was assessed via a NanoDrop™ One (Thermo Scientific, Waltham, MA, USA),
213 whereas its integrity was evaluated via the 4200 TapeStation System (Agilent Technologies, Santa Clara,
214 CA, USA). RNA samples with a RIN value of 6.5 or higher were deemed to have acceptable integrity and
215 were subsequently processed for the cDNA library construction.

216 For the library preparation for RNA sequencing, the KAPA RNA HyperPrep Kit (Roche, Basel, Switzerland)
217 was utilized, following the manufacturer's instructions. cDNAs were ligated with Unique Dual-Indexed
218 Adapters from the KAPA Kit (Roche, Basel, Switzerland). The concentration of each library was determined
219 via a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), and quality was assessed via the
220 4200 TapeStation System (Agilent Technologies, Santa Clara, CA, USA). The libraries were pooled in equal
221 quantities, and the pooled samples were sent for sequencing at Azenta Genewiz, South Plainfield, NJ, USA.
222 The sequencing was performed on an Illumina NextSeq 500/550 (Illumina, Inc., San Diego, USA) using a
223 75-cycle run for paired-end reads.

224 Genes were filtered for at least 75% of the values to be above zero, run through principal component
225 analysis (PCA, with 10 components), and a standard scalar was applied to the PCs.

226

227 *2.6 Proteomics*

228

229 For the proteomic analysis (protein extraction, sample preparation, mass spectrometry and early data
230 processing), the pond-collected samples were sent to MS Bioworks, Ann Arbor, MI, USA.

231 For the sample preparation, the cells were washed twice with 1X PBS and lysed in a modified RIPA buffer
232 containing 50 mM Tris HCl (pH 8.0), 150 mM NaCl, 2.0% SDS, 0.1% TX100, and 1X Roche Complete Protease
233 Inhibitor (Roche, Basel, Switzerland). Lysis was performed using 1.4 mm stainless steel beads in a Next
234 Advance Bullet Blender (Next Advance, Troy, NY, USA) for two cycles of 3 minutes each. The samples were
235 then heated to 60 °C for 30 minutes and clarified by centrifugation at 13,200 rpm. Protein was precipitated
236 via trichloroacetic acid (TCA) overnight at -20 °C, after which the pellet was washed and resuspended in a
237 solution of 8 M urea, 50 mM Tris HCl (pH 8.0), and 1X Roche Complete Protease Inhibitor (Roche, Basel,
238 Switzerland). Protein quantification was performed via a Qubit fluorometer (Thermo Fisher Scientific,

239 Waltham, MA, USA), and a pooled sample was created from all the individual samples. For digestion, 2.5
240 µg of protein was treated overnight with trypsin. The samples were first reduced at room temperature
241 for 1 hour in 12 mM DTT, followed by alkylation for 1 hour at room temperature in 15 mM iodoacetamide.
242 Trypsin was added at an enzyme-to-substrate ratio of 1:20. Each sample was then acidified to 0.3% TFA
243 and subjected to solid-phase extraction (SPE) using a Waters µHLB (Waters, Pleasenton, CA, USA). The SPE
244 process involved activating the matrix with four additions of 500 µL of 70% acetonitrile, equilibrating with
245 four additions of 500 µL of 0.3% TFA, adding the sample, washing the wells with three additions of 500 µL
246 of 0.3% TFA, and eluting in 200 µL followed by an additional 400 µL of 60% acetonitrile in 0.3% TFA. The
247 samples were then frozen at -80 °C and lyophilized overnight.

248 For the mass spectrometry analysis, 500 ng of each sample was analyzed in triplicate via an LC-MS/MS
249 setup consisting of a Thermo Fisher Vanquish Neo UPLC system (Thermo Fisher Scientific, Waltham, MA,
250 USA) interfaced with a Thermo Fisher Orbitrap Astral (Thermo Fisher Scientific, Waltham, MA, USA).
251 Peptides were loaded onto a trapping column and eluted through a 75 µm analytical column at a flow rate
252 of 350 nL/min, with the column temperature maintained at 55 °C. A 30-minute gradient was used for
253 elution. The mass spectrometer was operated in data-independent acquisition (DIA) mode, capturing full
254 scan MS data at a resolution of 240,000 FWHM over the m/z range of 380-980, followed by 300 sequential
255 2 m/z precursor isolation windows, with product ions acquired at a resolution of 40,000 FWHM. The
256 maximum ion injection time (IIT) was set to 3.5 ms for DIA, and the normalized collision energy (NCE) was
257 set to 25. An injection of the pooled sample was included at both the beginning and end of the analysis
258 batch.

259 The initial data processing for the DIA results was performed via DIA-NN (v1.9.1) [22] , which facilitated
260 several key functions: conversion of RAW files to the QUANT format, alignment on the basis of retention
261 times, searching data via an in-silico spectral library generated from the FASTA file and iteratively from
262 the spectral library derived from RAW data, filtering of database search results at a 1% false discovery rate
263 (FDR) for both peptides and proteins, calculation of peak areas for detected peptides, and normalization
264 of the data. The data were subjected to percentile normalization for the PCA plots.

265

266 *2.7 Metabolomics*

267

268 For the metabolomics analysis, 1 ml from the samples taken from the ponds (containing both the media
269 and cells) was sent to Creative Proteomics, Shirley, NY, USA, for untargeted metabolomics analysis. For
270 the blank, samples with only media were taken from the ponds before the algal inoculation. For the
271 sample preparation, the samples were transferred into a tube for lyophilization. Following transfer, 500
272 µL of 80% methanol was added, and the mixture was vortexed for 30 seconds and then subjected to
273 sonication in an ice-water bath for 30 minutes. The samples were subsequently stored at -20 °C for 1 hour,
274 after which they were vortexed for an additional 30 seconds and centrifuged at 12,000 rpm for 10 minutes
275 at 4 °C, to lyse and remove the cells. From the resulting supernatant, 200 µL was transferred into a vial,
276 and 5 µL of a 0.14 mg/mL solution of DL-o-chlorophenylalanine was added. The final solution was filtered
277 through a 0.22 µm filter in preparation for LC-MS analysis. The separation of the compounds was achieved
278 via the Vanquish Flex UPLC system in conjunction with a Q Exactive Plus mass spectrometer (Thermo
279 Scientific, Waltham, MA, USA), which featured a heated electrospray ionization (ESI) source. The LC

separation utilized a Hypersil Gold C18 column (100 × 2.0 mm, 1.9 µm), with the mobile phase consisting of solvent A (0.05% formic acid in water) and solvent B (acetonitrile) under a gradient elution profile: from 0 to 1 minute at 5% B, from 1 to 12.5 minutes increasing from 5% to 95% B, holding at 95% B from 12.5 to 13.5 minutes, returning to 95% B from 13.5 to 13.6 minutes, and finally returning to 5% B from 13.6 to 16 minutes. The flow rate was maintained at 0.3 mL/min, with the column temperature set at 40 °C and the sample manager temperature at 4 °C. The mass spectrometry parameters for both ESI positive (ESI+) and negative (ESI-) modes were configured as follows: for ESI+, the heater temperature was set to 300 °C, the sheath gas flow rate was 45 arb, the auxiliary gas flow rate was 15 arb, the sweep gas flow rate was 1 arb, the spray voltage was 3.0 kV, the capillary temperature was 350 °C, and the S-Lens RF level was 30%. For ESI-, the heater temperature remained at 300 °C, with the same sheath and auxiliary gas flow rates as those used for ESI+, but the spray voltage was adjusted to 3.2 kV, and the S-Lens RF level was increased to 60%.

For the bioinformatic analysis (see section 2.8), we considered the metabolites whose abundance was more than 10x greater than the abundance of the metabolites in the blank to be significant.

2.8 Bioinformatics

Principal component analysis (PCA) was performed in R with the factoextra package [23]. For 3D PCA analysis, the first 3 components were plotted for each replicate per day, and the centroid point was also plotted on the same axes. Day-specific standard deviation from the centroid (based on replicates) is shown by a shaded sphere.

For RNA sequencing analysis, the following sequence processing steps and associated packages were used: the fastq file was quality-checked with FastQC, and adapter trimming was performed with Trimmomatic [24], t

The pseudoalignment of the transcripts to the transcriptome of *M. minutum* (as it was published and publicly available by [17]) and the quantification of their abundance were performed with the Kallisto package [25]. The integration of the Kallisto data and the differential gene expression analysis, including the time course analysis, was performed with Sleuth [26]. The GEMMA package [27] was used for the correlation of the gene expression with specified quantitative phenotypes.

Additional tools for the multi-omic analyses include ggplot2 [28] for graphic representation and pheatmap [29] for the heatmap.

De novo motif enrichment analysis of the promoter regions (200 bps upstream of ATG) was performed with the Homer Motif Discovery and Analysis software [30].

We used the Data Integration Analysis for Biomarker Discovery using Latent variable approaches for Omics studies (DIABLO) framework from the R software package mixOmics (version 6.28.0) to combine the transcriptomics, proteomics, and metabolomics datasets [31]. Specifically, we compared these data across Day 0 and Day 1 and subsequently Day 4 and Day 5 to identify key features of culture scale up and infection, respectively. The metabolomics and proteomics datasets were log transformed prior to this analysis, and the transcriptomics data were normalized to transcripts per million (TPM) values. The

proteomic and transcriptomic datasets were further filtered via the nearZeroVar function from the caret R package (version 7.0- 1), with a unique cutoff of 20 for the proteomics dataset and selecting transcripts with variances of greater than or equal to 200 were selected. In the DIABLO design matrix, we prioritized correlations between the three omics datasets by choosing a 0.6 weighted model. The features from latent component 1, from the Day 0 and Day 1 comparison and from the Day 4 and Day 5 comparison were then compiled for comparison across all four days via the R package pheatmap (version). The values were scaled and centered prior to plotting, and both rows and columns were clustered via Euclidean distance.

3. Results

3.1 Phenotypic and conditional measurements reveal distinct acclimation, growth and infection dynamics in algae in open pond systems

We performed a longitudinal analysis of algal growth in two 1000 L open raceway ponds, and introduced an amoeboaphelid pathogen on day 4 of the experiment, tracking a total of 12 variables. Phenotypic measurements included cell count, AFDW, chlorophyll concentration, total carbohydrate content, total H3 histone concentration, cell size, and caspase activity (Additional File 1). In addition, we measured five environmental variables within the ponds, including %DO, pH, conductivity, temperature, and light intensity.

Across the time-course, we observed 3 growth phases of interest (color-coded in Fig. 2A):

a) The acclimation or lag phase (cyan). This phase occurred immediately after the inoculation in the open system with refreshed nutrients. This period lasted 2 days, and growth during this phase was slower than that during the other phases.

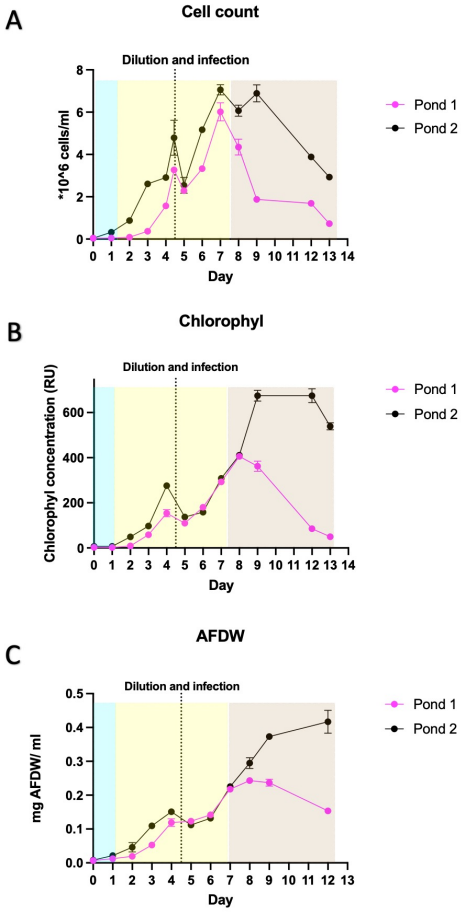
b). The log phase (yellow). The cells were fully acclimated to the open-pond conditions and grown exponentially. This phase lasted for 4 days before the dilution and the infection and 3 days after dilution.

c) Crash phase (brown). This phase is caused by algal death, which leads to the complete overtake of the ponds by the pathogen and other opportunistic species.

The ponds behave largely similarly, although we do observe a divergence of phenotypes at later stages of infection, as documented in Fig. 2.

To understand whether stress conditions trigger programmed cell death, we measured caspase activity throughout the experiment. Caspases are proteases that catalyze the breakdown of proteins during programmed cell death [32]. We observed that caspase activity initially increased after pond inoculation and then gradually decreased as the cells acclimated to their new environment (Additional File. 2). Notably, we did not observe an increase in caspase activity after infection with amoeboaphelid, suggesting

356 that the reduction in the algal cell count during the postinfection phase is caused by caspase-independent
357 cell death [33] [34].



358

359 ***Fig. 2: Growth tracking of the M. minutum pond cultures.*** *A) The cell count between the three ponds. B) Chlorophyl concentration*
360 *in relative units (RU) from florescence measurements. C) The Ash-free dry weight (AFDW) of algae cells in mg per ml of culture.*

361

362 For an integrated view of the phenotypic and conditional profile of the longitudinal experiment, principal
363 component analysis (PCA) was performed with the 12 phenotypic and environmental variables.
364 Dimension 1 represented 49.4% of the explained variance (Fig. 3A) and separated the variables that are
365 indicative of growth, from those indicative of stress. This provides insight into the competing pressures
366 that the algae face during cultivation, and the contributions of these pressures to outcomes.

367 We observe that the three distinct phases (acclimation, log, and crash) are resolved away from each other
368 in the PCA biplot shown in Figure 3B, and we can discern a trajectory in PC space as the longitudinal
369 experiment progresses, which is shown by connecting temporally adjacent points. While the largest
370 adjacent transition occurred during scale-up (Day 0 to Day 1), we also observed a significant shift upon
371 infection into a distinct PC space highlighted in brown in Figure 3B. This visualization further emphasizes

the distinction between the conditions and outcomes during these phases and suggests the presence of distinct molecular profiles within the algae that likely precede outcomes. To this end, we performed transcriptomic, proteomic and metabolomic analyses across the first 6 days of this experiment.

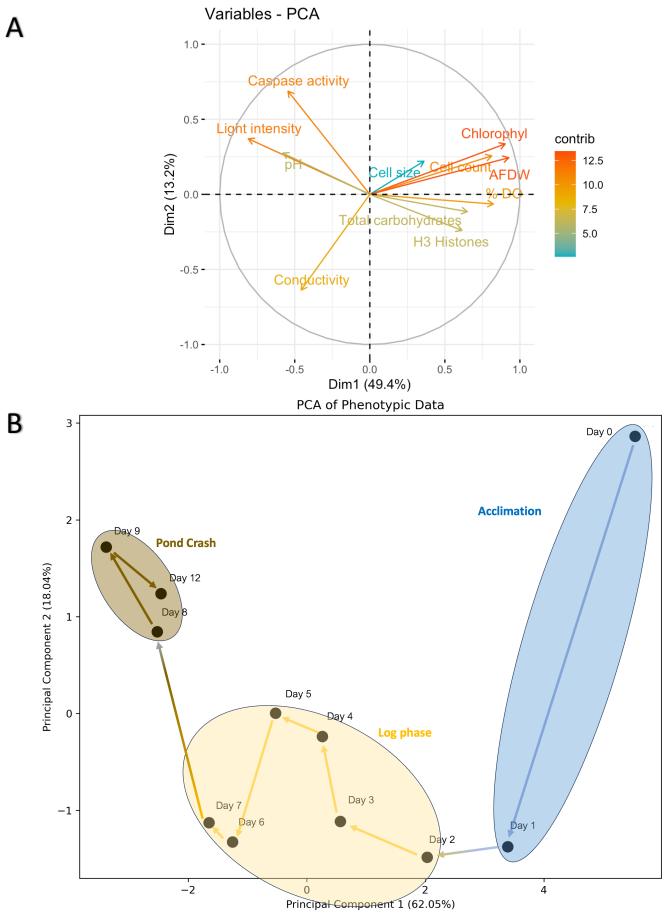


Fig. 3: PCA for 12 variables as measured at the two ponds. *A) PCA for the 12 variables used. The color of each line represents the contribution of each variable to the principal axes. Created with factoextra in R. B) Each day of the cultivation, plotted as individuals. The arrows are indicating the trajectory of each day during the culture run.*

3.2 Transcriptomic stability tracks with optimal growth and culture resiliency

With the goal of identifying molecular patterns within the algae that precede the observed phenotypic outcomes, we performed transcriptomic analysis from Day 0 to Day 6 (Additional File 2). The samples thus

385 represented the acclimation period (day 0 and day 1), the preinfection log phase (day 2- day 4) and the
386 early infection days (days 5 and 6).

387 A considerable percentage of transcripts (12.3%) were significantly differentially expressed from day 0
388 (the inoculation day) to day 1, reflecting the transcriptional acclimation to the scale transition from the
389 20 L carboys to the open 1000 L ponds (Fig.4A). This difference in expression was multiple times greater
390 than any daily difference observed during the following days. This included days after infection where
391 only 0.7% of the transcripts were significantly differentially expressed.

392 We performed PCA on the transcriptomic data to determine whether the longitudinal trajectory was
393 comparable to that of our phenotypic and environmental variables (Fig. 4B and 4C). The transition from
394 Day 0 to Day 1 was the most striking, reflecting the rapid and large transcriptomic response to the 1000 L
395 pond conditions. The distributions of the days corresponding to the pre-infection log phase, on the other
396 hand, are closer in PC space, reflecting transcriptional stability during that growing stage. A clear
397 distinction between the distribution of the preinfection and postinfection days was also observed, as a
398 result of both the culture dilution and the introduction of the pathogen. In fact, unlike in Fig. 3B where
399 the measured environmental and phenotypic variables postinfection were in a distinct cluster away from
400 the scale-up, the early postinfection transcriptomic profile trends back toward the acclimation phase
401 profile but still remains distinct from them.

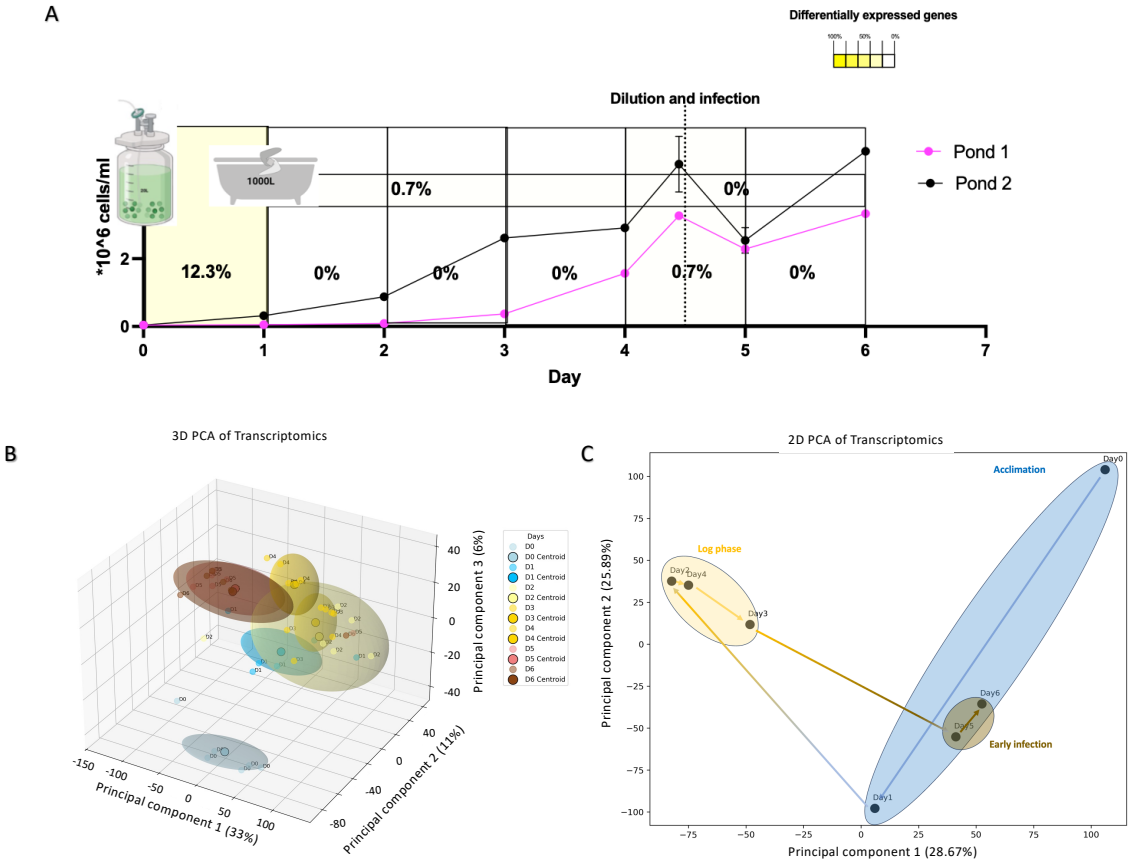


Fig. 4: Comparative analysis of the transcript abundances overlaid with cell growth results. *A) Growth curve of both ponds over the time course, with transcriptomic data collected on the same day as growth was measured (as shown by points on the curve). Percentages of differentially expressed genes between specific daily points within the cultivation are shown by the numbers. B) 3D PCA analysis of the transcriptomics, showing the distribution of all the technical replicates for each day from the two ponds. Centroids are indicated with a darker point, and standard deviation is indicated by the shaded circle. C) 2D representation of the PCA showing the average points for each day. The arrows indicate the trajectory of the time course.*

The annotation of *M. minutum* is incomplete, making traditional pathway analysis complicated. To further investigate the regulation behind those transcriptional changes we performed a motif analysis among the promoter regions of the genes that were significantly differentially expressed to identify regions that could be bound by conserved transcription factors. We focused on the transitions between day 0 and day 1 (Table 1), and between day 4 and day 5 (Table 2). The ten most enriched motifs that emerged from each group are shown in Table 1 and Table 2, alongside their best matches from other organisms. Most identified motifs are unique to each transition.

Among the acclimation phase gene promoters, we also observed two motifs (labeled with HOMER, such as ZNF549 and ZNF415) that are likely bound by zinc finger proteins. In addition, 40% of the motifs

correspond to a promoter motif from the diatom *Thalassiosira pseudonana*. A notably high percentage (66.66%) of the promoter regions of the genes that were significantly differentially expressed after the infection contained a motif that can be recognized by Auxin Response Factor 1 (ARF1), a transcription factor identified in *Arabidopsis thaliana.*

Table 1: Enriched promoter motifs at the scale adaptation-related genes.

	Motif	% of Targets	P- value	Best match/ Details
1		43.38%	1e-12	PB0188.1_Tcf7l2_2/Jaspar (0.841)
2		40.40%	1e-9	Tp_0225(RRM)/Thalassiosira_pseudonana-RNCMPT00225-PBM/HughesRNA (0.730)
3		25.17%	1e-9	RTG3/Literature (Harbison)/Yeast(0.896)
4		32.78%	1e-8	ABF2/MA0266.2/Jaspar (0.781)
5		41.72%	1e-8	ZNF549/MA1728.2/Jaspar (0.753)
6		27.48%	1e-8	STB4/STB4_YPD/[] (Harbison)/Yeast(0.723)
7		34.44%	1e-8	Syncrip(RRM)/Xenopus_tropicalis-RNCMPT00281-PBM/HughesRNA(0.754)
8		21.85%	1e-7	ZBED4/MA2328.1/Jaspar (0.742)
9		18.54%	1e-7	RAP211(AP2EREBP)/colamp-RAP211-DAP-Seq (GSE60143)/Homer (0.722)
10		34.44%	1e-7	ZNF415(Zf)/HEK293-ZNF415.GFP-ChIP-Seq (GSE58341)/Homer (0.721)

Table 2: Enriched promoter motifs at the infection-related genes.

	Motif	% of Targets	P- value	Best match/ Details
1		66.67%	1e-9	ARF1(ABI3/VP1)/Arabidopsis thaliana/AthaMap (0.719)
2		64.41%	1e-9	gcm2/MA0917.1/Jaspar (0.770)
3		35.59%	1e-8	Abd-B/dmmpmm(Bergman)/fly(0.775)

4		32.33%	1e-8	RTG3/Literature (Harbison)/Yeast (0.882)
5		50.38%	1e-8	RGT1/Literature (Harbison)/Yeast (0.775)
6		28.57%	1e-8	Sp1(Zf)/Promoter/Homer (0.747)
7		40.85%	1e-7	bHLH78/MA0963.1/Jaspar (0.804)
8		26.07%	1e-7	HES2/MA0616.3/Jaspar (0.840)
9		23.06%	1e-7	ZmHOX2a (2) (HD-HOX)/Zea mays/AthaMap (0.795)
10		41.10%	1e-7	PBX1/MA0070.2/Jaspar (0.812)

3.3 Transcriptomic and phenotypic data reveal the light and dissolved oxygen (DO) regulation of protein translation and photosynthesis

One of the advantages of performing time-course transcriptomics on large open cultures in a poorly studied species such as *M. minutum* is that it allows for the extraction of reliable and useful information about specific environmentally triggered expressed genes. Light intensity and DO are two environmental factors known to influence outcomes in algae. To understand their role in the outcome of our cultures, we used a program called GEMMA that applies mixed linear models to connect phenotypes with large-scale molecular data [27]. We identified transcripts whose expression was positively or negatively correlated with light intensity and %DO across samples (Figs. 5 and 6).

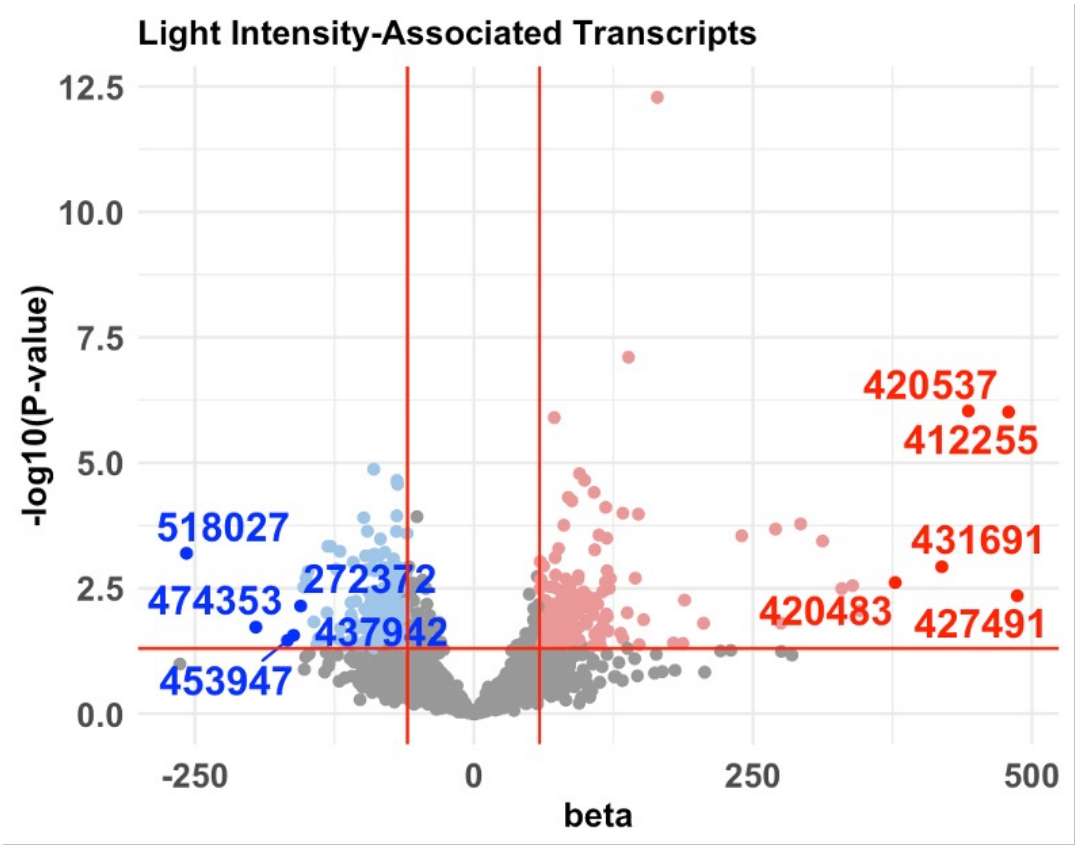


Fig. 5: GEMMA analysis, revealing the genes which are correlated with light intensity. *The volcano plot shows the significance and the extent of the correlation between the light intensity and the transcript abundance. The top 5 up- and downregulated genes by light are indicated.*

The observation of the top up- and downregulated genes correlated with light intensity (Tables 3 and 4), revealed a strong correlation between high light intensity and the expression of genes related to structural components of ribosomes and other translation-related proteins. In addition, we noted a structural component of Photosystem I among the genes whose expression was downregulated in response to light, which is consistent with previous literature showing that photosynthetic organisms downregulate photosystem components in response to light stress [35].

To further understand the regulatory genomic elements that might contribute to light-dependent changes in gene expression, we performed a motif analysis on the promoter regions of all the genes whose expression was significantly up- or downregulated by light (Tables 5 and 6). Among the most significant motifs from the upregulated genes, two corresponded to transcription factors from two higher plant species, *Arabidopsis* and *Zea mays*.

Table 3: The annotations of the top 5 genes whose expression was upregulated with light.

JGI ID	beta	Gene product
427491	486.8095	ribosomal protein S9 component of cytosolic 80S ribosome and 40S small subunit
412255	479.0458	40S ribosomal protein S29-like
420537	442.9904	putative 40S ribosomal protein S3a
431691	419.257	Translation protein SH3-like domain-containing protein
420483	377.4611	component of cytosolic 80S ribosome and 60S large subunit

Table 4: The annotations of the top 5 genes whose expression was downregulated with light.

JGI ID	beta	Gene product
518027	-257.7188	heat shock cognate 70 kDa protein 4-like
474353	-195.5336	hypothetical protein
453947	-167.1721	eukaryotic translation initiation factor 5
437942	-161.7026	photosystem I protein P30
272372	-155.3597	hypothetical protein

Table 5: Enriched promoter motifs at the light upregulated genes.

	Motif	% of Targets	P- value	Best match/ Details
1		21.74%	1e-8	Unknown3/Arabidopsis-Promoters/Homer (0.710)
2		33.33%	1e-8	CAT8/MA0280.2/Jaspar (0.749)

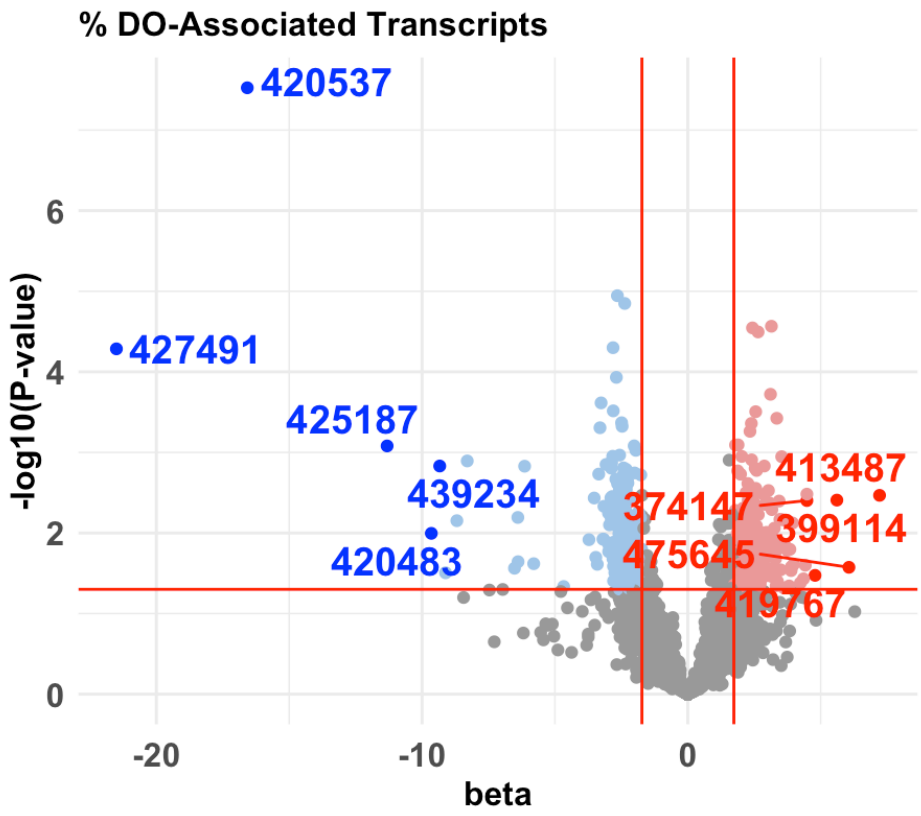
3		23.91%	1e-7	ZmHOX2a (2) (HD-HOX)/Zea mays/AthaMap (0.823)
4		23.91%	1e-7	AT5G04390/MA1736.2/Jaspar (0.752)
5		23.91%	1e-7	lsy-2/MA2181.1/Jaspar (0.874)
6		47.46%	1e-7	Sp1(Zf)/Promoter/Homer (0.783)
7		30.80%	1e-6	RGT1/Literature (Harbison)/Yeast (0.793)
8		30.43%	1e-6	insv/MA2315.1/Jaspar (0.821)
9		16.67%	1e-6	ZNF549/MA1728.2/Jaspar (0.779)
10		30.80%	1e-5	ZNF610/MA1713.2/Jaspar (0.721)

Table 6: Enriched promoter motifs at the light downregulated genes.

	Motif	% of Targets	P- value	Best match/ Details
1		36.93%	1e-10	br/MA0011.2/Jaspar (0.725)
2		41.83%	1e-9	PB0188.1_Tcf7l2_2/Jaspar (0.842)
3		32.35%	1e-8	ZNF331/MA1726.2/Jaspar (0.870)
4		23.53%	1e-8	MSN4(MacIsaac)/Yeast (0.777)
5		27.12%	1e-7	MET31(MacIsaac)/Yeast (0.683)

6		26.47%	1e-7	PHD1(MacIsaac)/Yeast (0.856)
7		17.97%	1e-6	PUF68(RRM)/Drosophila_melanogaster-RNCMPT00141-PBM/HughesRNA (0.805)
8		27.45%	1e-6	ZmHOX2a(2)(HD-HOX)/Zea mays/AthaMap (0.799)
9		17.32%	1e-6	WRKY1/MA0589.2/Jaspar (0.869)
10		41.50%	1e-6	TEAD2/MA1121.2/Jaspar (0.808)

A similar GEMMA analysis was repeated with the genes most correlated with %DO. Among the ten most upregulated and downregulated genes correlated with %DO (Fig. 6A, Tables 7 and 8), multiple genes related to photosynthesis and carbon fixation were identified. A phosphoribulokinase/uridine kinase, an enzyme responsible for a key step in the Calvin-Benson cycle, a light-harvesting protein of Photosystem I, and a protein involved in programmed cell death were among the most upregulated genes correlated with DO. We performed motif analysis on the promoter regions of all the genes correlated with DO (Tables 9 and 10). We observed enrichment of the binding site for the same transcription factor from the diatom *Thalassiosira pseudonana* at both the up- and downregulated genes from the DO.



472

473 ***Fig. 6: GEMMA analysis, revealing genes correlated with %DO.*** *The volcano plot shows the significance and the extent of the*
474 *correlation between the light intensity and the transcript abundance. The top 5 up- and downregulated genes by %DO are indicated.*

475

476 *Table 7: The annotations of the top 5 genes whose expression was upregulated with %DO.*

JGI ID	beta	Gene product
413487	7.216919	Phosphoribulokinase / Uridine kinase family-domain containing protein
475645	6.064846	hypothetical protein
399114	5.610648	expressed protein
419767	4.78923	light-harvesting chlorophyll-a/b protein of photosystem I

374147	4.481988	Programmed cell death protein 2, C-terminal putative domain-domain containing protein

Table 8: The annotations of the top 5 genes whose expression was downregulated with %DO.

JGI ID	beta	Gene product
427491	-21.50261	ribosomal protein S9 component of cytosolic 80S ribosome and 40S small subunit
420537	-16.57716	putative 40S ribosomal protein S3a
425187	-11.31527	ribosomal protein L3 component of cytosolic 80S ribosome and 60S large subunit
420483	-9.651194	component of cytosolic 80S ribosome and 60S large subunit
439234	-9.335318	P-loop containing nucleoside triphosphate hydrolase protein

Table 9: Enriched promoter motifs at the %DO upregulated genes.

	Motif	% of Targets	P- value	Best match/ Details
1		23.02%	1e-8	GATA15(C2C2gata)/col-GATA15-DAP-Seq (GSE60143)/Homer (0.833)
2		28.68%	1e-8	ARR1/Literature (Harbison)/Yeast (0.743)
3		41.51%	1e-7	NKX2-3/MA0672.2/Jaspar (0.771)
4		21.89%	1e-6	ABF2/MA0266.2/Jaspar (0.803)
5		37.74%	1e-6	ZNF549/MA1728.2/Jaspar (0.751)
6		38.87%	1e-6	Tp_0225(RRM)/Thalassiosira_pseudonana-RNCMPT00225-PBM/HughesRNA (0.796)
7		57.36%	1e-6	ZNF415(Zf)/HEK293-ZNF415.GFP-ChIP-Seq (GSE58341)/Homer (0.760)

8		17.74%	1e-6	Vdr/MA0693.4/Jaspar (0.846)
9		16.98%	1e-6	EFL-1(E2F)/cElegans-L1-EFL1-ChIP-Seq(modEncode)/Homer (0.719)
10		28.30%	1e-6	REB1/SacCer-Promoters/Homer (0.815)

Table 10: Enriched promoter motifs at the %DO downregulated genes.

	Motif	% of Targets	P- value	Best match/ Details
1		43.38%	1e-12	PB0188.1_Tcf7l2_2/Jaspar (0.841)
2		40.40%	1e-9	Tp_0225(RRM)/Thalassiosira_pseudonana-RNCMPT00225-PBM/HughesRNA (0.730)
3		25.17%	1e-9	RTG3/Literature (Harbison)/Yeast (0.896)
4		32.78%	1e-8	ABF2/MA0266.2/Jaspar (0.781)
5		41.72%	1e-8	ZNF549/MA1728.2/Jaspar (0.753)
6		27.48%	1e-8	STB4/STB4_YPD/ [] (Harbison)/Yeast (0.723)
7		34.44%	1e-8	Syncrip (RRM)/Xenopus_tropicalis-RNCMPT00281-PBM/HughesRNA (0.754)
8		21.85%	1e-7	ZBED4/MA2328.1/Jaspar (0.742)
9		18.54%	1e-7	RAP211 (AP2EREBP)/colamp-RAP211-DAP-Seq (GSE60143)/Homer (0.722)
10		34.44%	1e-7	ZNF415(Zf)/HEK293-ZNF415.GFP-ChIP-Seq (GSE58341)/Homer (0.721)

3.4 Proteome composition shifts during algal growth and stress response.

To gain additional molecular information from the algae as they sensed transitions in their environment, a comprehensive proteomic analysis was conducted on the two pond populations longitudinally. Similar to the transcriptomics data, PCA of the proteomics data (Additional File 3) revealed a highly dynamic phase during the first two days of the culture (Fig. 7). After entering the log phase, however, the proteome becomes progressively more stable, while at the same time maintaining a distinct profile between the pre- and postinfection periods. This was in contrast to the more dynamic changes in the transcriptome after infection, highlighted in the plasticity of the transcriptional regulation in relation to the environmental triggers in comparison with the stability of the protein expression. Further analysis of the correlations between specific transcripts and proteins will be discussed later in this article.

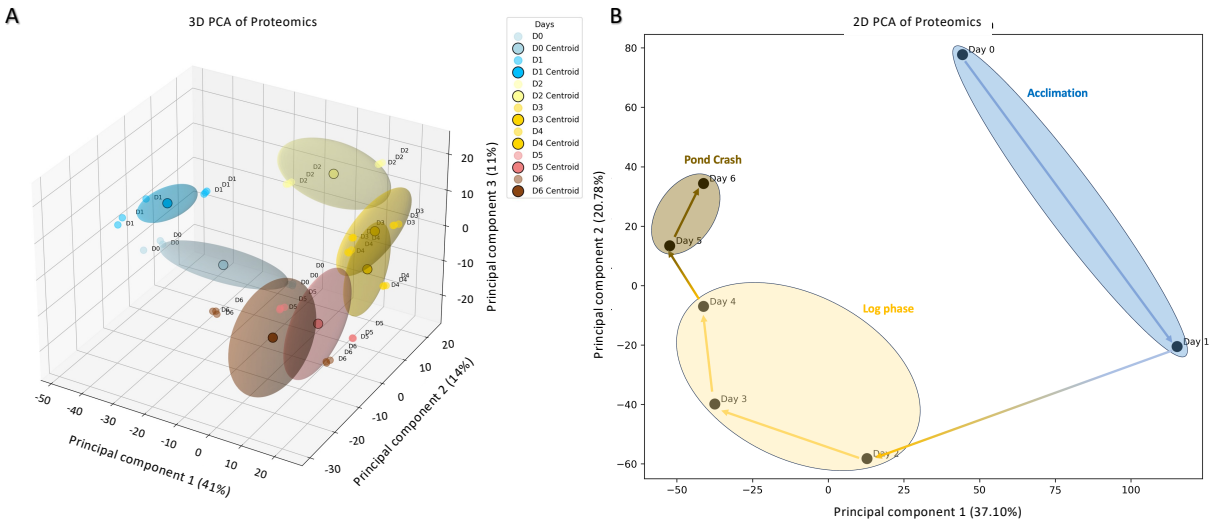


Fig. 7: PCA of the proteomics. *A) 3D PCA analysis of the proteomics, showing the distribution of all the technical replicates for each day from the two ponds. Centroids are indicated with a darker point, and standard deviation is indicated by the shaded circle. B) 2D representation of the PCA showing the average points for each day. The arrows indicate the trajectory of the time course.*

3.5 Metabolomic analysis reveals potential indicators of the stress response and pathogen infection.

PCA of the metabolites (Sup. Dataset 4) revealed distinct profiles during the three phases of cultivation similar to the transcriptomic and proteomic PCAs (Fig. 8A and 8B). Interestingly, day 3, which is in the middle of the log phase, stood out in the PCA. This contrasted with the observation made from transcriptomics. While log phase was linked with transcriptomic stability, the metabolic profile is very dynamic during the growth phase.

We performed clustering analysis of the longitudinal metabolomics data to further understand the metabolites that covaried (Fig. 8C). We used DPGP [36], a nonparametric clustering algorithm that can factor in longitudinal information, to separate metabolites into clusters on the basis of their patterns in abundance. We found that many metabolites either had their peak or trough on days 3 or 4, the midpoint of the log phase (clusters 2, 4, 6 and 8 and 9). The most abundant cluster (Cluster 1) included metabolites whose abundance steadily increased over time. On the other hand, Cluster 5 showed the opposite trend, starting with high abundance and slowly decreasing in abundance over time.

Hierarchical clustering of the scaled and centered metabolite abundances in the heatmap, shown in Fig. 8D, highlights the emergence of three groupings of metabolites, and is in relative accordance with the DPGP clustering. The first group consists of the metabolites whose abundance peaks in the lag phase. The second group peaked in the log phase. The last group, which was the largest by number, contained the metabolites that reached their peak abundance after the infection. The last group included hormones and steroids, such as 5-Dihydrotestosterone, 7-Hydroxytestosterone, Estrone, Estriol and 19-Norandrostenedion.

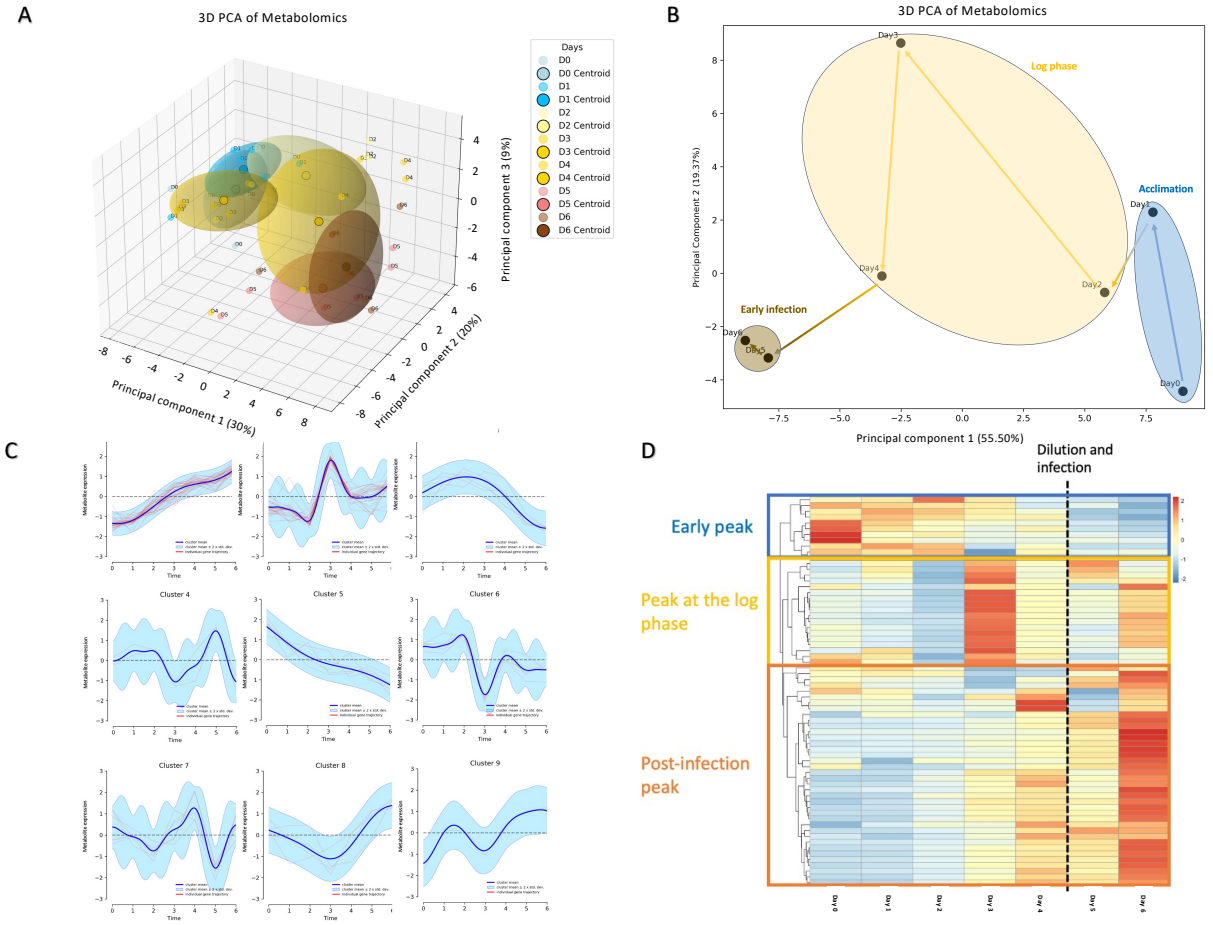


Fig. 8: The metabolomic analysis from the pond cultures. *A) 3D PCA analysis of the metabolomics, showing the distribution of all the technical replicates for each day from the two ponds. Centroids are indicated with a darker point, and standard deviation is indicated by the shaded circle. B) 2D representation of the PCA showing the average points for each day. The arrows indicate the trajectory of the time course. C) Clustering analysis of the abundance of the metabolites throughout the studied period, via DPGP. D) Heatmap with the hierarchical clustering of the significantly produced metabolites throughout the run of the cultivation. It includes the metabolites which had more that 10x higher signal than the metabolites in the blank.*

3.6 Integrated multi-omic analysis revealed distinct molecular features related to scale transition and infection defense

To assess the global relationships among proteins, metabolites and transcripts across our longitudinal experiment, we performed an integrative analysis via the DIABLO framework in the mixOmics R package [31]. We performed this analysis using all omics data from days 0 and 1 to explore the molecular features of the scale-up stress, and separately used days 4 and 5 data to explore that of the infection stress. While we found that there was no overlap in the key proteins or transcripts output by the model between the

scale up and infection stress, here were two metabolites, guanine and 12-oxo-Phytodienoic acid, were common across the two stressors, suggesting more convergence at the metabolite level than at the molecular pathway level.

We took the molecular features from the first latent component of both analyses and combined them into one heatmap to comprehensively visualize these results (Fig. 9). As expected, the columns clustered by day, and we observed that the two transitions had almost entirely unique patterns. We observed a cluster of transcripts (all significantly differentially expressed between days 0 and 1), proteins, and myristic acid elevated on day 0 relative to the other days. Similarly, we observed another distinct cluster of transcripts and proteins, as well as 1,5-hexadiene and 2,3-Diaminosalicylic acid, which were elevated on Day 4 compared with the other days. However, a decrease in a different cluster of proteins, as well as maltose and hypogeic acid, distinguished day 1 relative to the other days. These results suggest distinct molecular and metabolic profiles in response to different stressors, identifying potential unique biomarkers and target pathways for detection and intervention.

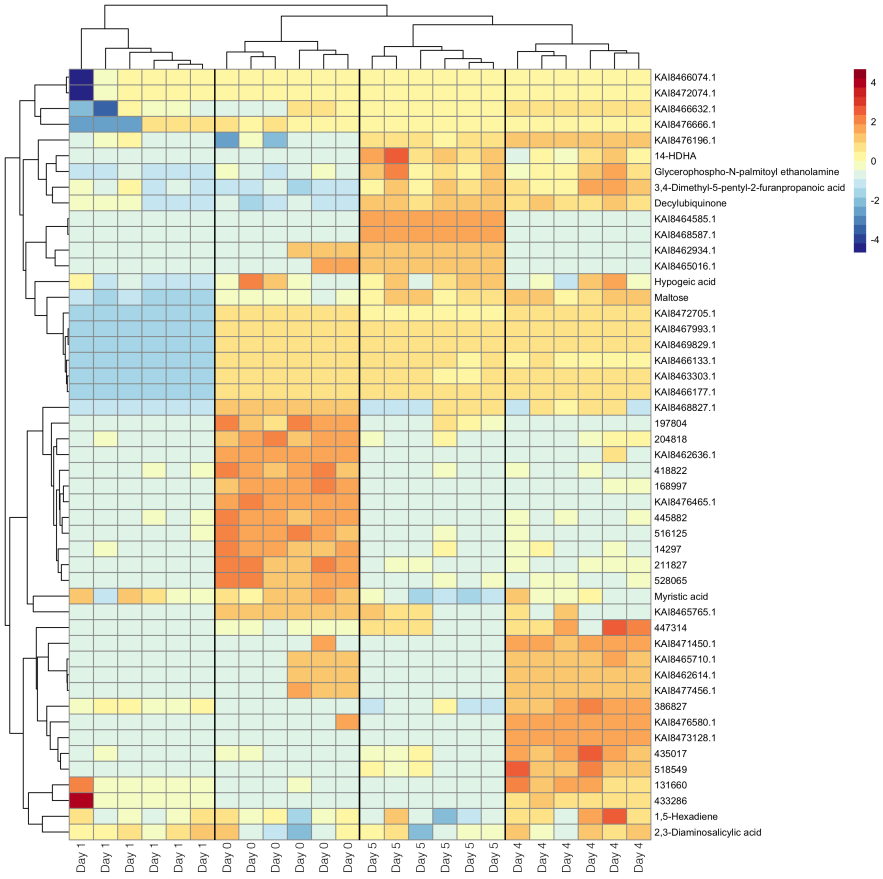


Fig. 9: Indicative signatures of scale adaptation and infection. *Heatmap containing the molecular features (genes, proteins and metabolites) which change significantly their abundance during the transition from day 0 to day 1 and/or from day 4 to day 5. Their abundance is being indicated by a scaled blue-to-red scheme. Each of the 6 biological replicates for each condition are indicated separately.*

555

556 ## 4. Discussion

557

558 *4.1 Transcriptomics and metabolomics reveal the commonalities and differences between the molecular*
559 *mechanisms governing scale-up stress and infection response*

560

561 We performed a multi-omic analysis of *M. minutum* 26B-AM cultivated over 12 days in 1000 L open
562 raceway ponds to explore the molecular signatures that manifest in response to stressors often
563 encountered at a large scale. The three distinct growth phases we observed in the 1000 L ponds were in
564 line with what is typically observed at a large scale, i.e., batch cultivations of algae [37] [38]. Culture scale-
565 up is often evidenced by a lag phase, as we observed in our cultures, especially when moving from well-
566 controlled laboratory cultures into open raceway ponds [37] [39] [40]. A change in scale represents a
567 significant stressor for the cultures, introducing a combination of different environmental signals,
568 including changes in light, radiation, heat, gas distribution, as well as microbial and environmental particle
569 composition that requires an acclimation period, which is reflected in our growth curves over days 0 and
570 1.

571 Similarly, algal pests are common stressors experienced in open raceway ponds and active areas of study.
572 Changes in phenotypic measurements in response to infection or even visual indications of infection
573 stress, typically appear sometime after the initial infection. In our cultures, we introduced the
574 amoebaphelid on day 4 but did not begin to observe a decline in growth, which was emblematic of a
575 culture crash, until day 8, accompanied by a color change in the culture from light green to brown. Pond
576 crashes, which are driven by algal pathogens and often damage crops overnight, account for almost 30%
577 of algal crop losses annually [41], increasing the overall cost of algal applications. Despite the dramatic
578 effects of those infections on ponds, the cellular mechanism of the infection which includes the insertion
579 of the parasite in the host cell [19], allows a certain amount of time of incubation before a noticeable
580 crash [42]. Cellular and molecular alterations are already in place and if they can be detected in a timely
581 manner, the irreversible effects of a pond crash can potentially be avoided, or at least resource allocation
582 decisions can be made much earlier.

583 Differential expression analysis revealed that culture scale up has a substantial effect on the
584 transcriptome, as evidenced by the much larger portion of significantly differentially expressed transcripts
585 between days 0 and 1 compared with the other days, including those impacted by infection. In addition,
586 infection is not expected to trigger a wide immune response from the algal cells, reflected by
587 transcriptional adaptation. Characterizing these pathways and adjusting conditions to prevent their
588 activation could help acclimate the culture in a less stressful way and potentially make it more resilient to
589 future stressors.

590 With respect to infection, current strategies for early indication of contamination include visual detection
591 strategies such as flow cytometry [43], in situ microscopy [44], multispectral image analysis [45] and
592 photosynthetic measurements [46] [47]. Those methodologies, although relatively quick, require a certain
593 pathogen titer in the culture. Biomolecular approaches include 16S and 18S sequencing [48], and the
594 detection of Volatile Organic Compounds (VOC), as indirect early indicators of an infection [49]. Here we

show that the algae can also be used as early indicators of infection, as demonstrated by the changes in transcripts.

Because many of the transcripts in *M. minutum* are poorly annotated, we used motif analysis to further explore the significantly differentially expressed genes. Most identified motifs were unique to each transition, potentially indicating two distinct regulatory pathways between the scale adaptation and pathogen stress. However, common elements between the two datasets exist, implying that there are factors that could be used to coregulate responses to different stressors (either endogenously or via synthetic biology). An example of a potential point of convergence are the motifs recognized by the transcription factors mitochondrial retrograde 1 (RTG1) and 3 (RTG3), both of which were originally identified in yeast and are involved in mitochondrial metabolic regulation.

Among the promoters of genes differentially expressed during scaling up, we also identified two motifs labeled by HOMER, ZNF549 and ZNF415, which are likely bound by zinc finger proteins, a family of transcription factors that have been shown to regulate stress responses in other algal species [50] [51]. In addition, 40% of the motifs correspond to a promoter motif from the diatom *Thalassiosira pseudonana* suggesting possible functional importance across diverse algal species. A notably high percentage (66.66%) of the promoter regions of the genes that were significantly differentially expressed after the infection contained a motif that can be recognized by Auxin Response Factor 1 (ARF1), a transcription factor identified in *Arabidopsis thaliana.* ARF1 is responsive to the phytohormone auxin, which is responsible for a wide range of developmental and stress responses in plants [52] [53]. Given the evolutionary proximity between higher plants and green algae, it is likely that ARF1 may have similar functions in algal responses post-infection and might suggest a possible defense mechanism. The identification of these factors points to potential targets for synthetic biology in *M. minutum* in the future, to either boost or tamper with specific molecular responses.

A survey of the metabolites released by the algae upon encountering various stressors also provides valuable insights into predictive markers, pond health indicators, and infection indicators, while also explaining the natural mechanisms of growth and defense in algae. We used PCA and DPGP to look at global trends in metabolites over time. While log phase was linked with transcriptomic stability, the metabolic profile is very dynamic during the growth phase.

These metabolites are highly important, both as early indicators of an upcoming crash, multiple days before the effects of the infections can be visible, and as guides for the development of prophylactic strategies for algal cultures. Among those metabolites, we identified 14-HDHA, a hydroxy docosahexaenoic acid oxylipin which is derived from docosahexaenoic acid (DHA). The production of oxylipins from algae is linked with oxidative stress [54], suggesting a similar regulatory pathway here for pathogen defense. In this group, there are also multiple hormones and steroids (5-Dihydrotestosterone, 7-Hydroxytestosterone, Estrone, Estriol and 19-Norandrostenedion). While such compounds originate in the aquatic environments from sources other than algae, many algal species are known to biotrasform them or completely remove them from the system [55] [56] [57], with effects that can be either toxic or growth promoting for the cells [58] [59]. The hypothesis that these compounds play a role in the molecular signaling or progression of the infection is one that deserves further investigation for many reasons, including the discovery of early biomarkers for an upcoming pond crash.

636 *4.2 Transcriptomic and phenotypical observations reflect light and gas-regulated responses in M. minutum*

637

638 Light and gas constitute the two cornerstone environmental factors for aquatic organisms, especially the
639 photosynthetic organisms such as green microalgae. Indeed, light and CO_2 availability are essential not
640 only for photosynthesis, the primary metabolic function of microalgae but also for photoprotection, lipid
641 composition, and carbon concentration [60] [61] [62] [63] [64]. The conserved and well-studied role of
642 light and CO_2 availability led us to explore which *M. minutum* genes are corelated with these genes, as a
643 mechanism by which to verify that our method identifies valid molecular signatures.

644 We observed that translation-related proteins were strongly correlated with high light intensity. Protein
645 translation is an essential regulatory step for phenotypic acclimation, and its control by light has been
646 observed in other algae, including the Chlorophyte *Ulva prolifera* [65], the Charophyte *Klebsormidium*
647 *nitens* [66], and *Chlorella sp.* [67]. In addition, we noted a structural component of Photosystem I among
648 the genes whose expression was downregulated in response to light. Photosystem I is a crucial component
649 of the electron transport chain and is responsible for the light-mediated transfer of electrons across the
650 thylakoid membrane. Excess light can cause oxidative stress and subsequent photoinhibition in
651 Photosystem I, and in many photosynthetic organisms, including in *Chlamydomonas reinhardtii*, light can
652 suppress the expression of structural proteins in Photosystem I to protect the machinery from this type
653 of damage [68]. Future work to test this hypothesis across algal species will be critical.

654 Motif analysis of the promoter regions of genes correlated with light highlighted two motifs related to
655 homeobox proteins in higher plants. The *Zea mays* homeobox 2a (ZmHOX2a) protein, whose expression
656 in maize is restricted to meristems, is involved in in plant development and growth [69]. In addition, this
657 motif might more broadly represent homeobox transcription factors, which are known to regulate
658 environmental responses in plants [70] [71], suggesting evolutionary conservation of light-induced gene
659 regulation through homeobox transcription factors across the plants and algae. A motif corresponding to
660 the same transcription factor is also found among the motifs from the promoters of the downregulated
661 genes by light, highlighting the involvement of these motifs in both the up- and downregulation responses
662 to light. Interestingly, 30.80% of the targets of the promoter regions from the up-regulated genes
663 contained a motif corresponding to the transcription factor RGT1 (Restores Glucose Transport) from
664 *Saccharomyces cerevisiae*. The protein regulates glucose transport though the expression of glucose
665 transporters [72], and the connection between light-induced stress and glucose transport has been
666 documented in other organisms [73] [74]. The motifs of the downregulated genes include one that is
667 recognized by the *Saccharomyces cerevisiae* stress-related multicopy suppressor of SNF1 mutation
668 protein 4 (Msn4), which controls the metabolic response under cellular stress conditions [75].

669 Among the genes correlated with DO, multiple genes are related to photosynthesis and carbon fixation.
670 A phosphoribulokinase/uridine kinase, an enzyme responsible for a key step in the Calvin-Benson cycle,
671 and a light-harvesting protein of Photosystem I are among the most upregulated genes. Genes related to
672 programmed cell death were positively correlated with DO, an observation that aligns with results from
673 other green algal species linking ROS signaling and programmed cell death [76] [77] [78] [79] [80]. These
674 observations highlight specific molecular changes experienced by large-scale cultures that are likely
675 directly related to environmental fluctuations. Our results also revealed that increased DO caused more
676 gene downregulation than gene upregulation overall, especially for the genes indicated in Table 8. Most
677 of these genes are involved with structural components of ribosomes, indicating a strong suppressive

relationship between O_2 levels and translational regulation, as was also observed in the case of genes upregulated with light. In fact, the most downregulated gene from the DO was the most upregulated one from light. Since high light is a stress factor, as is low oxygen, this reciprocal relationship between their regulated genes aligns with expectations. The list of the top five downregulated genes includes a P-loop containing nucleoside triphosphate hydrolase, which, in higher plants such as *Arabidopsis,* acts as a signaling molecule involved in ribosome formation and attenuates ROS formation and oxidative stress [81] [82].

Motif analysis of the promoter regions of all the genes correlated with DO revealed enrichment of the binding site for the same transcription factor from the diatom *Thalassiosira pseudonana* that we observed during scale acclimation. This putative transcription factor motif is found in 40.40% of the promoter regions of the genes differentially expressed during scale acclimation (Table 1), which is a sign of a strong contribution of the DO differences to the scale transition stress. Future efforts to identify this transcription factor that is conserved across algae might be fruitful for understanding stress mechanisms and broad-scale synthetic biology-based interventions. The second most enriched motif among the promoters of the upregulated genes corresponds to the transcription factor Arrestin 1 (ARR1), which in higher plants is responsive to the phytohormones Cytokinins and plays a fundamental role in numerous developmental stages of the plant [83] .The motif bound by the yeast transcription factor Regulation of the Mitochondrial Retrograde 3 (RTG3) is also found at the promoters of downregulated genes. This transcription factor regulates in multiple metabolic pathways, including amino acid and nucleotide synthesis [84] indicating a possible role of oxygen-regulated metabolic responses in *Monoraphidium*. Interestingly, motifs corresponding to the same transcription factor were found among the photometers of the differentially expressed genes during the acclimation period and after the infection with the amoeboapheids (Tables 1 and 2), which implies a role for transcription factors that bind this motif across stressors and environmental changes.

We also noted that we identified overlap in the genes correlated with light and DO and those related to stress responses (specifically responses to cold and salt), as determined by [17]. Specifically, of the 24 genes they outline as being annotated with a fasciclin-like domain, a domain with known roles in algal and plant stress responses, we note 2 with expression coordinated with dissolved oxygen (annotation # 450375, 525465), 1 with expression coordinated with light (annotation # 480939) and 2 with expression levels coordinated with both (annotation # 421681, 446308). These results support the validity of our signature identification through multi-omic data integration in algae, and point towards the need for follow-up on our stress signatures for identification of predictive markers for tracking large-scale pond culture resiliency, and molecules for therapeutic intervention.

5. Conclusions

In this study, we used multi-omics to investigate the dynamic responses of the green alga *M. minutum* to various environmental changes and stressors. A study of multi-omic analyses of M. *minutum* was recently published, providing a molecular profile of this industrially significant species under high salt and cold temperature stress [17]. We utilized a multi-omic approach in larger open and, thus, more heterogeneous culture systems in order to provide a roadmap for the application of molecular observations at the cellular

level and data processing to address the significant challenges faced by the large-scale bioprocessing industry. Specifically, in the context of algae open pond systems, issues of unpredictability, inconsistency, and often deadly infections lead to crashes and severe crop losses [86] [42]. Our approach aims to address these issues while gaining information about the cellular responses of the species. We demonstrated the ability to integrate datasets of genes, proteins and metabolites that respond to specific stressors or correspond to certain culture stages and transitions, opening up possibilities for follow-up studies in the fields of strain engineering, biomarker development, bioprocessing, and basic cellular studies. We show that biomarker information can be extracted from this integrated data (transcripts, proteins, or metabolites) that indicate resilient growth and upcoming stressful events, such as prolonged acclimation periods during scale transitions and potential crashes due to the presence of pathogens. With respect to metabolites in particular, which are the most likely candidates for biomarker detection and immediate therapeutic intervention, specific hormones and steroids, such as 5-Dihydrotestosterone, 7-Hydroxytestosterone, Estrone, Estriol and 19-Norandrostenedion, are highly expressed during early infection. Hormones and steroids act as signaling molecules that regulating not only growth development but also stress adaptation [87] [88] [58]. Given that their abundance increased after the introduction of the pathogen, showcases that those compounds, produced by either the pathogen or the algae cells in response to their interaction with the pathogen, act as early regulators of a stress response and can be utilized as timely biomarkers for upcoming pond crashes. The detection of these markers provides preemptive information on impending phenotypic changes, which can be useful both for pond monitoring and eventual intervention and mitigation strategies. In the long term, if some of these molecules act in defense against algae, microbiomes can be engineered to produce these molecules as a prophylactic strategy during large-scale culture. The mechanism for this can be gleaned through the associated transcripts and proteins that are co-regulated.

In addition, when observing changes in transcription, proximal genomic elements can provide clues as to how algae can be engineered to optimize gene expression profiles and cell states. While there is not an abundance of information on transcription factor binding in *Monoraphidium* and related species, the presence of conserved sequences suggests not only that the binding of specific factors is likely to be pivotal in stress-related regulatory mechanisms but also that these mechanisms are likely conserved across algal species. These findings suggest that the factors identified here are likely to regulate stress response in other algal species and that mitigation and intervention strategies can be broadly applied. Future work will focus on testing across species to confirm these findings.

On a broader scale, the collection of molecular data from such expansive cultivation systems can pave the way for analyses aimed at developing predictive power with respect to culture resiliency, as well as potential targets for engineering this resiliency in the future. Indeed, we observed transcripts shared in both this study and the study conducted by Calhoun et al (2022), which indicates that more data generation in small- and large-scale systems, coupled with integrated analysis, could reveal novel health monitoring molecules in algal culture. The ultimate goal is for these models to be adaptable to multiple types of bioprocessing, beyond algae cultures, and this study provides a roadmap in that direction, using omic data to optimize large-scale applications.

Abbreviations

761 14-HDHA: 14-hydroxy-docosahexaenoic acid

762 AFDW: Ash-free dry weight

763 ARF1: ADP-ribosylation factor 1

764 ARR1: Arabidopsis response regulator 1

765 CCM: Carbon concentration mechanism

766 DIABLO: Data Integration Analysis for Biomarker discovery using Latent Components

767 DISCOVR: Development of integrated screening, cultivar optimization, and verification research

768 DPGP: Dirichlet Process Gaussian Process mixture model

769 Msn4: multicopy suppressor of SNF1 mutation protein 4

770 PCA: principal component analysis

771 ROS: reactive oxygen species

772 RuBisCO: Ribulose-1,5-bisphosphate carboxylase/oxygenase

773 RTG1: mitochondrial retrograde 1

774 RTG3: mitochondrial retrograde 3

775 SNF1: sucrose non-fermenting 1

776

777 **Declarations**

778

779 *Ethics approval and consent to participate*

780 N/A

781

782 *Consent for publication*

783 N/A

784

785 *Availability of data and materials*

786 The datasets supporting the conclusions of this article are available in the NCBI SRA repository, under
787 BioProject PRJNA1307767 and are included within the article and its additional files.

788

789 *Competing interests*

The authors declare that they have no competing interests.

Funding

This work was supported by the DOE Bioenergy Technology Office (BETO) and the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

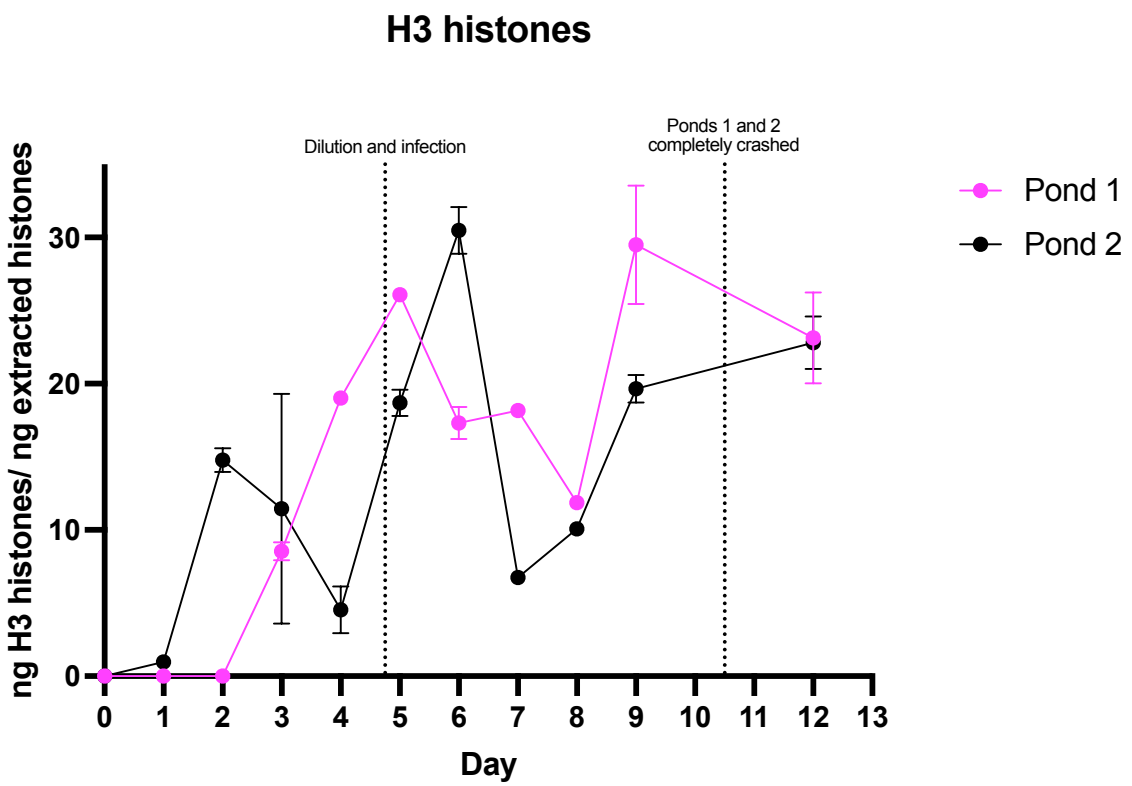
Authors' contributions

R.K. led the funding acquisition efforts and had the logistic and programmatic lead of the project. G.K and R.K. conceptualized the project and designed the methodology and the experimental procedure. G.K., D.Y., E.W., T.E., O.W. and M.P.H. worked on the different aspects of the experimental procedure (scale-up, pond running and maintenance, sampling and phenotyping, omics preparation and sequencing). G.K., J.S., D.Y., T.S. and R.K. worked on the computational and bioinformatic analysis. T.W.L. advised on the experimental design and algal physiology. G.K, J.S. and R.K. wrote the original manuscript. All the authors reviewed and edited the manuscript.

Acknowledgements

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. We thank Ryan W. Davis, Catherine Mageeney, Pamela Lane, and Carolyn Fisher, Sharon Nademanee, and Damian Carrieri for their insightful comments on an earlier draft of this manuscript.

Additional data



Additional Figure. 1: The caspase activity during the pond run. *The caspase activity for every day of the culture run in pond 1 (pink) and pond 2 (black).*

Additional Dataset 1: Phenomics

Additional Dataset 2: Transcriptomics

Additional Dataset 3: Proteomics

Additional Dataset 4: Metabolomics

References

1. Dismukes, G.C., et al., *Aquatic phototrophs: efficient alternatives to land-based crops for biofuels.* Current Opinion in Biotechnology, 2008. **19**(3): p. 235-240.
2. Cruce, J.R. and J.C. Quinn, *Economic viability of multiple algal biorefining pathways and the impact of public policies.* Applied Energy, 2019. **233-234**: p. 735-746.
3. Thanigaivel, S., et al., *The urge of algal biomass-based fuels for environmental sustainability against a steady tide of biofuel conflict analysis: Is third-generation algal biorefinery a boon?* Fuel, 2022. **317**: p. 123494.
4. Wiatrowski, M., et al., *Techno-economic assessment for the production of algal fuels and value-added products: opportunities for high-protein microalgae conversion.* Biotechnology for Biofuels and Bioproducts, 2022. **15**(1): p. 8.

5. Costa, J.A.V., et al., *Chapter 9 - Open pond systems for microalgal culture*, in *Biofuels from Algae (Second Edition)*, A. Pandey, et al., Editors. 2019, Elsevier. p. 199-223.
6. Jerney, J. and K. Spilling, *Large Scale Cultivation of Microalgae: Open and Closed Systems*, in *Biofuels from Algae: Methods and Protocols*, K. Spilling, Editor. 2020, Springer New York: New York, NY. p. 1-8.
7. Lane, T.W., *Barriers to microalgal mass cultivation.* Current Opinion in Biotechnology, 2022. **73**: p. 323-328.
8. Quiroz, D., et al., *Geographical assessment of open pond algal productivity and evaporation losses across the United States.* Algal Research, 2021. **60**: p. 102483.
9. Echenique-Subiabre, I., et al., *Site-specific factors override local climatic conditions in determining microalgae productivity in open raceway ponds.* Algal Research, 2023. **74**: p. 103235.
10. Kleiman, R.M., et al., *Characterizing weather-related biophysical and financial risks in algal biofuel production.* Applied Energy, 2021. **294**: p. 116960.
11. Borowitzka, M. and A. Vonshak, *Scaling up microalgal cultures to commercial scale.* European Journal of Phycology, 2017. **52**: p. 407-418.
12. Patidar, S.K., et al., *Potential of Monoraphidium minutum for carbon sequestration and lipid production in response to varying growth mode.* Bioresource Technology, 2014. **172**: p. 32-40.
13. Dommange, X., P.A. Tanguy, and M. Jolicoeur, *Feasibility of lipid mechanical extraction from viable Monoraphidium minutum.* Microalgae Biotechnol, 2015. **1**(1): p. 12-19.
14. Bogen, C., et al., *Reconstruction of the lipid metabolism for the microalga Monoraphidium neglectum from its genome sequence reveals characteristics suitable for biofuel production.* BMC genomics, 2013. **14**: p. 1-18.
15. Yu, X., et al., *Isolation of a novel strain of Monoraphidium sp. and characterization of its potential application as biodiesel feedstock.* Bioresource Technology, 2012. **121**: p. 256-262.
16. Pikoli, M.R., et al., *Characteristics of tropical freshwater microalgae Micractinium conductrix, Monoraphidium sp. and Choricystis parasitica, and their potency as biodiesel feedstock.* Heliyon, 2019. **5**(12).
17. Calhoun, S., et al., *Multi-omics profiling of the cold tolerant Monoraphidium minutum 26B-AM in response to abiotic stress.* Algal Research, 2022. **66**: p. 102794.
18. Mikhailov, K.V., et al., *Genomic analysis reveals cryptic diversity in aphelids and sheds light on the emergence of Fungi.* Current Biology, 2022. **32**(21): p. 4607-4619.e7.
19. Letcher, P.M., et al., *Characterization of Amoeboaphelidium protococcarum, an Algal Parasite New to the Cryptomycota Isolated from an Outdoor Algal Pond Used for the Production of Biofuel.* PLOS ONE, 2013. **8**(2): p. e56232.
20. Huesemann, M., et al., *DISCOVR strain pipeline screening – Part I: Maximum specific growth rate as a function of temperature and salinity for 38 candidate microalgae for biofuels production.* Algal Research, 2023. **71**: p. 102996.
21. Van Wychen, S. and L.M.L. Laurens, *Determination of Total Solids and Ash in Algal Biomass: Laboratory Analytical Procedure (LAP)*. 2016, National Renewable Energy Lab. (NREL), Golden, CO (United States): United States. p. Medium: ED; Size: 415 KB.
22. Demichev, V., et al., *DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput.* Nature Methods, 2020. **17**(1): p. 41-44.
23. Kassambara, A. and F. Mundt. *Extract and Visualize the Results of Multivariate Data Analyses [R package factoextra version 1.0.7]*. 2020.
24. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.
25. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nature Biotechnology, 2016. **34**(5): p. 525-527.

26. Pimentel, H., et al., *Differential analysis of RNA-seq incorporating quantification uncertainty.* Nature Methods, 2017. **14**(7): p. 687-690.
27. Zhou, X. and M. Stephens, *Genome-wide efficient mixed-model analysis for association studies.* Nat Genet, 2012. **44**(7): p. 821-4.
28. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, 2016.
29. Kolde, R., *pheatmap: Pretty Heatmaps.* 2025.
30. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.
31. Rohart, F., et al., *mixOmics: An R package for 'omics feature selection and multiple data integration.* PLOS Computational Biology, 2017. **13**(11): p. e1005752.
32. Dho, S.H., et al., *Caspases as master regulators of programmed cell death: apoptosis, pyroptosis and beyond.* Experimental & Molecular Medicine, 2025. **57**(6): p. 1121-1132.
33. Vavilala, S.L., et al., *KCl induces a caspase-independent programmed cell death in the unicellular green chlorophyte Chlamydomonas reinhardtii (Chlorophyceae).* Phycologia, 2016. **55**(4): p. 378-392.
34. Zuppini, A., et al., *Chlorella saccharophila cytochrome f and its involvement in the heat shock response.* Journal of Experimental Botany, 2009. **60**(14): p. 4189-4200.
35. Teramoto, H., et al., *Light-intensity-dependent expression of Lhc gene family encoding light-harvesting chlorophyll-a/b proteins of photosystem II in Chlamydomonas reinhardtii.* Plant Physiol, 2002. **130**(1): p. 325-33.
36. McDowell, I.C., et al., *Clustering gene expression time series data using an infinite Gaussian process mixture model.* PLOS Computational Biology, 2018. **14**(1): p. e1005896.
37. Gao, S., et al., *DISCOVR strain screening pipeline – Part III: Strain evaluation in outdoor raceway ponds.* Algal Research, 2023. **70**: p. 102990.
38. Yang, H., Q. He, and C. Hu, *Feasibility of biodiesel production and CO2 emission reduction by Monoraphidium dybowskii LB50 under semi-continuous culture with open raceway ponds in the desert area.* Biotechnology for Biofuels, 2018. **11**(1): p. 82.
39. Huesemann, M., et al., *A validated model to predict microalgae growth in outdoor pond cultures subjected to fluctuating light intensities and water temperatures.* Algal Research, 2016. **13**: p. 195-206.
40. Matula, E.E. and D.J.A. Nabity, *Metabolic response of Chlorella vulgaris to a transient thermal environment for supporting simultaneous air revitalization and thermal control in a crewed habitat.* Acta Astronautica, 2021. **187**: p. 406-415.
41. McBride, R.C., et al., *Contamination Management in Low Cost Open Algae Ponds for Biofuels Production.* Industrial Biotechnology, 2014. **10**(3): p. 221-227.
42. Harmon, V.L., et al., *Reliability metrics and their management implications for open pond algae cultivation.* Algal Research, 2021. **55**: p. 102249.
43. Ioc Coi Coi, M., et al., *Microscopic and molecular methods for quantitative phytoplankton analysis*

IOC. Manuals and guides.
44. Havlik, I., et al., *Monitoring of microalgal cultivations with on-line, flow-through microscopy.* Algal Research, 2013. **2**(3): p. 253-257.
45. Murphy, T.E., K. Macon, and H. Berberoglu, *Rapid algal culture diagnostics for open ponds using multispectral image analysis.* Biotechnology Progress, 2014. **30**(1): p. 233-240.
46. Deore, P., et al., *Non-photochemical quenching, a non-invasive probe for monitoring microalgal grazing: an early indicator of predation by Oxyrrhis marina and Euplotes sp.* Applied Phycology, 2020. **1**(1): p. 20-31.

47. Fundel, B., et al., *Can phaeopigments be used as markers for Daphnia grazing in Lake Constance?* Journal of Plankton Research, 1998. **20**(8): p. 1449-1462.
48. Bartley, M.L., et al., *Effects of salinity on growth and lipid accumulation of biofuel microalga Nannochloropsis salina and invading organisms.* Biomass and Bioenergy, 2013. **54**: p. 83-88.
49. Reese, K.L., et al., *Chemical Profiling of Volatile Organic Compounds in the Headspace of Algal Cultures as Early Biomarkers of Algal Pond Crashes.* Scientific Reports, 2019. **9**(1): p. 13866.
50. Xie, J., et al., *Genome-Wide Identification and Functional Analysis of C2H2 Zinc Finger Transcription Factor Genes in the Intertidal Macroalga Pyropia haitanensis.* Int J Mol Sci, 2025. **26**(9).
51. Thiriet-Rupert, S., et al., *Transcription factors in microalgae: genome-wide prediction and comparative analysis.* BMC Genomics, 2016. **17**(1): p. 282.
52. Liu, L., et al., *Enigmatic role of auxin response factors in plant growth and stress tolerance.* Front Plant Sci, 2024. **15**: p. 1398818.
53. Ulmasov, T., G. Hagen, and T.J. Guilfoyle, *ARF1, a Transcription Factor That Binds to Auxin Response Elements.* Science, 1997. **276**(5320): p. 1865-1868.
54. Linares-Maurizi, A., et al., *Stress-Induced Production of Bioactive Oxylipins in Marine Microalgae.* Mar Drugs, 2024. **22**(9).
55. Lai, K.M., M.D. Scrimshaw, and J.N. Lester, *Biotransformation and bioconcentration of steroid estrogens by Chlorella vulgaris.* Appl Environ Microbiol, 2002. **68**(2): p. 859-64.
56. Shi, W., et al., *Removal of estrone, 17alpha-ethinylestradiol, and 17beta-estradiol in algae and duckweed-based wastewater treatment systems.* Environ Sci Pollut Res Int, 2010. **17**(4): p. 824-33.
57. Liu, D., et al., *Biochemical responses of freshwater microalgae Chlorella sorokiniana to combined exposure of Zn(Ⅱ) and estrone with simultaneous pollutants removal.* J Environ Manage, 2023. **348**: p. 119392.
58. Yu, H., et al., *Exploring the molecular mechanism of Chlorella vulgaris in response to androstenedione exposure based on genes continuously up-regulated in transcription analysis.* Ecotoxicology and Environmental Safety, 2024. **271**: p. 115996.
59. Czarny, K., et al., *Toxicity of single steroid hormones and their mixtures toward the cyanobacterium Microcystis aeruginosa.* Journal of Applied Phycology, 2019. **31**(6): p. 3537-3544.
60. Yamaoka, Y., et al., *Light, CO2, and carbon storage in microalgae.* Current Opinion in Plant Biology, 2025. **84**: p. 102696.
61. Tokutsu, R., et al., *Isolation of photoprotective signal transduction mutants by systematic bioluminescence screening in Chlamydomonas reinhardtii.* Sci Rep, 2019. **9**(1): p. 2820.
62. Ruiz-Sola, M.Á., et al., *Light-independent regulation of algal photoprotection by CO2 availability.* Nature Communications, 2023. **14**(1): p. 1977.
63. Maltsev, Y., et al., *Influence of Light Conditions on Microalgae Growth and Content of Lipids, Carotenoids, and Fatty Acid Composition.* Biology (Basel), 2021. **10**(10).
64. Dao, O., et al., *The green algae CO2 concentrating mechanism and photorespiration jointly operate during acclimation to low CO2.* Nature Communications, 2025. **16**(1): p. 5296.
65. Jahan, K., et al., *Transcriptomic Analysis Provides New Insights into the Tolerance Mechanisms of Green Macroalgae Ulva prolifera to High Temperature and Light Stress.* Biology, 2024. **13**(9): p. 725.
66. Serrano-Pérez, E., et al., *Transcriptomic and Metabolomic Response to High Light in the Charophyte Alga Klebsormidium nitens.* Frontiers in Plant Science, 2022. **13**.
67. Wang, B., et al., *Cell damage repair mechanism in a desert green algae Chlorella sp. against UV-B radiation.* Ecotoxicology and Environmental Safety, 2022. **242**: p. 113916.

979 68. Nama, S., et al., *High light induced changes in organization, protein profile and function of*
980 *photosynthetic machinery in Chlamydomonas reinhardtii.* Journal of Photochemistry and
981 Photobiology B: Biology, 2015. **152**: p. 367-376.
982 69. Kirch, T., et al., *The two homeodomains of the ZmHox2a gene from maize originated as an internal*
983 *gene duplication and have evolved different target site specificities.* Nucleic Acids Res, 1998.
984 **26**(20): p. 4714-20.
985 70. Bondino, H.G. and E.M. Valle, *A small intergenic region drives exclusive tissue-specific expression*
986 *of the adjacent genes in Arabidopsis thaliana.* BMC Mol Biol, 2009. **10**: p. 95.
987 71. Ejaz, M., et al., *ARABIDOPSIS THALIANA HOMEOBOX GENE 1 controls plant architecture by locally*
988 *restricting environmental responses.* Proc Natl Acad Sci U S A, 2021. **118**(17).
989 72. Polish, J.A., J.-H. Kim, and M. Johnston, *How the Rgt1 Transcription Factor of Saccharomyces*
990 *cerevisiae Is Regulated by Glucose.* Genetics, 2005. **169**(2): p. 583-594.
991 73. Valifard, M., et al., *The novel chloroplast glucose transporter pGlcT2 affects adaptation to*
992 *extended light periods.* Journal of Biological Chemistry, 2023. **299**(6): p. 104741.
993 74. Powner, M.B. and G. Jeffery, *Light stimulation of mitochondria reduces blood glucose levels.*
994 Journal of Biophotonics, 2024. **17**(5): p. e202300521.
995 75. Rajvanshi, P.K., M. Arya, and R. Rajasekharan, *The stress-regulatory transcription factors Msn2*
996 *and Msn4 regulate fatty acid oxidation in budding yeast.* J Biol Chem, 2017. **292**(45): p. 18628-
997 18643.
998 76. Al-Azab, A.J.S., et al., *Chronological transcriptome changes induced by exposure to cyanoacrylate*
999 *resin nanoparticles in Chlamydomonas reinhardtii with a focus on ROS development and cell wall*
1000 *lysis-related genes.* Algal Research, 2022. **68**: p. 102884.
1001 77. Gao, Y. and D.L. Erdner, *Cell death responses to acute high light mediated by non-photochemical*
1002 *quenching in the dinoflagellate Karenia brevis.* Scientific Reports, 2022. **12**(1): p. 14081.
1003 78. Xu, C., et al., *Programmed cell death process in freshwater Microcystis aeruginosa and marine*
1004 *Phaeocystis globosa induced by a plant derived allelochemical.* Science of The Total Environment,
1005 2022. **838**: p. 156055.
1006 79. Bai, F., et al., *Paraquat induces different programmed cell death patterns in Microcystis*
1007 *aeruginosa and Chlorella luteoviridis.* Ecotoxicology and Environmental Safety, 2023. **249**: p.
1008 114429.
1009 80. Zhao, Y., et al., *ROS-mediated programmed cell death (PCD) of Thalassiosira pseudonana under*
1010 *the stress of BDE-47.* Environmental Pollution, 2020. **262**: p. 114342.
1011 81. Zhao, M.G., Q.Y. Tian, and W.H. Zhang, *Nitric oxide synthase-dependent nitric oxide production is*
1012 *associated with salt tolerance in Arabidopsis.* Plant Physiol, 2007. **144**(1): p. 206-17.
1013 82. Moreau, M., et al., *AtNOS/AtNOA1 is a functional Arabidopsis thaliana cGTPase and not a nitric-*
1014 *oxide synthase.* J Biol Chem, 2008. **283**(47): p. 32957-67.
1015 83. Sakai, H., et al., *ARR1, a Transcription Factor for Genes Immediately Responsive to Cytokinins.*
1016 Science, 2001. **294**(5546): p. 1519-1521.
1017 84. Hashim, Z., et al., *Metabolic profiling of retrograde pathway transcription factors rtg1 and rtg3*
1018 *knockout yeast.* Metabolites, 2014. **4**(3): p. 580-98.
1019 85. Singh, A., et al., *DIABLO: an integrative approach for identifying key molecular drivers from multi-*
1020 *omics assays.* Bioinformatics, 2019. **35**(17): p. 3055-3062.
1021 86. Harun, R., et al., *Bioprocess engineering of microalgae to produce a variety of consumer products.*
1022 Renewable and Sustainable Energy Reviews, 2010. **14**(3): p. 1037-1047.
1023 87. Bajguz, A., *Brassinosteroids in Microalgae: Application for Growth Improvement and Protection*
1024 *Against Abiotic Stresses*, in *Brassinosteroids: Plant Growth and Development*, S. Hayat, et al.,
1025 Editors. 2019, Springer Singapore: Singapore. p. 45-58.

1026 88. Stirk, W.A., et al., *Endogenous brassinosteroids in microalgae exposed to salt and low temperature*
1027 *stress.* European Journal of Phycology, 2018. **53**(3): p. 273-279.

1028