Multi-omic and phenotypic analysis of growth and resilience of open raceway pond production of *Monoraphidium minutum* 26B-AM

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 Labortatories, Livermore, 94550, U.S.A.

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

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

Abstract

Green microalgae, such as *Monoraphidium minutum* 26B-AM, have garnered significant commercial interest due to 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 challenged 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. To test this hypothesis, we performed a pilot study involving a multi-omic time-course approach, integrating transcriptomics, proteomics, metabolomics, and phenomics, to monitor the acclimation, growth dynamics, and pathogen responses of algal cultures in two 1000L 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 longitudinally across the ponds. Furthermore, we propose a way to focus on transition-specific molecular pathways through integrated multi-omics, showing that most patterns are unique to each studies stressor and growth transition. Ultimately, this study suggests the further use of multi-omics observations at scale, 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

Highlights

- Multi-omics can be indicative of phase transitions in open-aired pond algal systems.
- Light and gas are correlated with expression changes in translation-related genes.
- Proteomic stability is linked to higher algal growth and pathogen defense.
- Hormones, such as estriol, have altered production upon infection.
- Multi-omics showed distinct networks related to scale-up and pathogen defense.

1. Introduction

Microalgae are a polyphyletic group of unicellular photosynthetic aquatic organisms, which have attracted commercial interest since the middle of the past century, due to their potential to advance biotechnology and biomanufacturing processes. Their higher biomass production yield compared to land plants [1], as well as their high content of useful metabolites such as lipids, have elevated 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 restriction 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 to the 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 to the weather [8], as well as to other site-specific environmental factors, such as sand particles or dust changing the composition of the culture media [9]. All these factors, directly related to the immediate exposure of the cultures to their biotic and abiotic surroundings, can lead to unpredictable outcomes when it comes to biomass productivity [10]. In addition, the growth and productivity is not always aligned to 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 bringing down 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 is essential for maintaining the stability of algal production systems. By utilizing the algal cells as sensors of their environment, we can identify 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 1000L open raceway ponds in environmentally-simulated conditions. For this study we used the freshwater green microalgae *Monoraphidium minutum* 26B-AM, a species currently studied as a promising winter species for biofuel feedstock due to its elevated lipid yield [12] [13] [14] [15] [16]. The genome of *M. minutum* has been also recently annotated [17]. Samples have been 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. Alongside the samples for the molecular observations, daily measurements of internal conditions, as well as certain phenotypical outcomes were also being monitored, in order to have a more holistic view on the dynamics between environmental signaling, genetic regulation and phenotypical outcomes in a non-extensively 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 on the ongoing effort for the enhancement of stability and productivity in large-scale applications. This approach aims to lay the groundwork for proactive interventions that can sustain algal health and performance in the face of environmental challenges. In this study we found that, beyond the phenotypical observations, molecular markers as evaluated by multi-omics can be indicative of distinct stages of open pond cultures. We also verified a link between light and redox signaling with photosynthesis and translation-related genes in *M. minutum*, in accordance with similar observations in closely related green algae species, suggesting applicability of the markers identified here across algal culture systems. Finally, we focused on the molecular pathways which correlate with the scale-up acclimation and the introduction of the pathogen in the system through integrated multi-omic analysis. While specific molecular networks of genes, proteins and metabolites were able to be identified for each transition, we discovered unique reactions to those two stressors, indicating separated signaling pathways.

2. Materials and methods

2.1 Strain and conditions

The *M. minutum* 26B-AM strain was initially collected and isolated as described in [20] and it has been utilized in our lab as part of the DISCOVR strain library. Seed cultures of *M. minutum* (100ml) were maintained in sterile 250ml Erlenmeyer flasks, containing DISCOVR medium [20] under a 16:8h light/dark cycle and continuous shaking at 110rpm 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 30°C maximum during the light period and 23°C minimum during the dark period. The seed cultures were diluted regularly to maintain exponential phase growth.

2.2 Scale-up

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

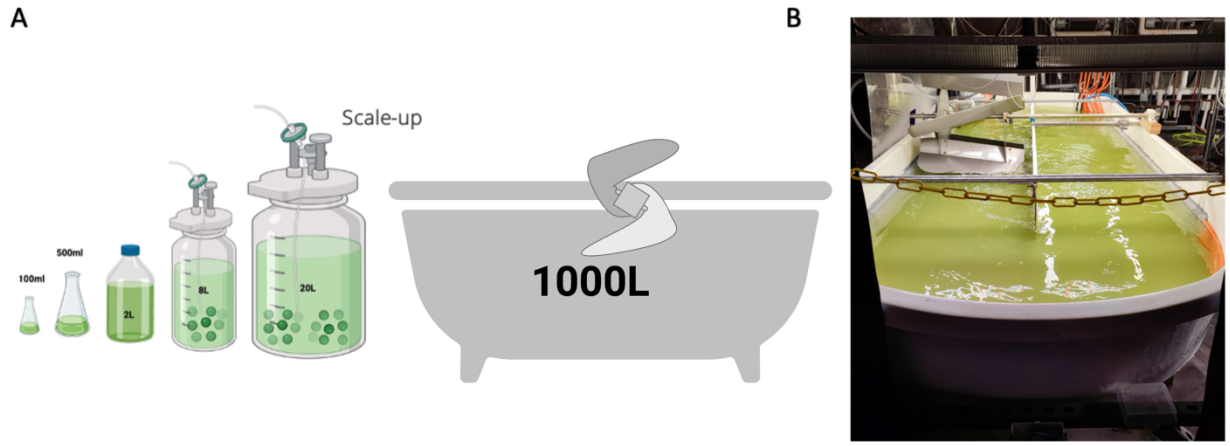


Fig. 1: A) Graphic representation of the scale-up. B) Picture of the 1000L open raceway pond on Day 0 of the culture run, after the inoculation of the media with the precultures

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

The cultures from the 20L carboys (3 for each pond) were used as inoculum for the 1000L open raceways ponds (manuscript in preparation). The ponds were filled with tap water and the nutrients and salts were added using the Modified 10% BG-11 recipe, as described in [20]. 60L from the inoculum were added to each pond. The ponds were operating under a 12h/12h dark/light cycle, implemented by 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 kept at a temperature range between 22-28°C, with the highest being reached in the middle of the light period and the lowest in the middle of the dark period. The temperature control was achieved through a heat-exchanger lining on the walls of the ponds. The cultures were under constant mixing by a mechanical paddlewheel at 20rpm. The pH was kept 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 on a daily basis one hour after the lights turned on. The phenotypic and conditional measurements were taken directly after that.

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

2.4 Phenotypic and conditional measurements

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

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

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

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

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

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

2.5 Transcriptomics/ RNA extraction and cDNA libraries preparation

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

To initiate cell lysis, the frozen pellets were thawed on ice and resuspended in 750 µL of Qiagen RLT Buffer supplemented with 1% β-mercaptoethanol from the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The resuspended cells were then transferred to 2 mL tubes containing Lysing Matrix Y zirconium oxide beads (MP Biomedicals, Santa Ana, CA, USA). Following a 10-second vortex, the samples were flash-frozen in liquid nitrogen. After thawing on ice, the cells were lysed using a FastPrep-24 bead beater (MP Biomedicals, Santa Ana, CA, USA) for 25 seconds at 40 Hz. The samples were centrifuged at 9000 rpm for 10 minutes at 4 °C, and the clear supernatant was collected for RNA extraction. RNA was extracted using

217 the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The
218 quantity of RNA was assessed using a NanoDrop™ One (Thermo Scientific, Waltham, MA, USA), while its
219 integrity was evaluated with the 4200 TapeStation System (Agilent Technologies, Santa Clara, CA, USA).
220 RNA samples with a RINe value of 6.5 or higher were deemed to have acceptable integrity and were
221 subsequently processed for the cDNA library construction.

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

230

231 *2.6 Proteomics*

232

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

235 For the sample preparation, the cells were washed twice with 1X PBS and lysed in a modified RIPA buffer
236 containing 50 mM Tris HCl (pH 8.0), 150 mM NaCl, 2.0% SDS, 0.1% TX100, and 1X Roche Complete Protease
237 Inhibitor (Roche, Basel, Switzerland). Lysis was performed using 1.4 mm stainless steel beads in a Next
238 Advance Bullet Blender (Next Advance, Troy, NY, USA) for two cycles of 3 minutes each. The samples were
239 then heated to 60 °C for 30 minutes and clarified by centrifugation at 13,200 rpm. Protein was precipitated
240 using trichloroacetic acid (TCA) overnight at -20 °C, after which the pellet was washed and resuspended
241 in a solution of 8 M urea, 50 mM Tris HCl (pH 8.0), and 1X Roche Complete Protease Inhibitor (Roche,
242 Basel, Switzerland). Protein quantification was performed using a Qubit fluorometer (ThermoFisher
243 Scientific, Waltham, MA, USA), and a pooled sample was created from all the individual samples. For
244 digestion, 2.5 µg of protein was treated overnight with trypsin. Samples were first reduced at room
245 temperature for 1 hour in 12 mM DTT, followed by alkylation for 1 hour at room temperature in 15 mM
246 iodoacetamide. Trypsin was added at an enzyme-to-substrate ratio of 1:20. Each sample was then
247 acidified to 0.3% TFA and subjected to solid-phase extraction (SPE) using Waters µHLB (Waters,
248 Pleasenton, CA, USA). The SPE process involved activating the matrix with four additions of 500 µL of 70%
249 acetonitrile, equilibrating with four additions of 500 µL of 0.3% TFA, adding the sample, washing the wells
250 with three additions of 500 µL of 0.3% TFA, and eluting in 200 µL followed by an additional 400 µL of 60%
251 acetonitrile in 0.3% TFA. The samples were then frozen at -80 °C and lyophilized overnight.

252 For the mass spectrometry analysis, 500 ng of each sample was analyzed in triplicate using an LC/MS/MS
253 setup consisting of a ThermoFisher Vanquish Neo UPLC system (ThermoFisher Scientific, Waltham, MA,
254 USA) interfaced with a ThermoFisher Orbitrap Astral (ThermoFisher Scientific, Waltham, MA, USA).
255 Peptides were loaded onto a trapping column and eluted through a 75 µm analytical column at a flow rate
256 of 350 nL/min, with the column temperature maintained at 55 °C. A 30-minute gradient was utilized for

elution. The mass spectrometer operated in data-independent acquisition (DIA) mode, capturing full scan MS data at a resolution of 240,000 FWHM over the m/z range of 380-980, followed by 300 sequential 2 m/z precursor isolation windows, with product ions acquired at a resolution of 40,000 FWHM. The maximum ion injection time (IIT) was set to 3.5 ms for DIA, and the normalized collision energy (NCE) was set to 25. An injection of the pooled sample was included at both the beginning and end of the analysis batch.

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

2.7 Metabolomics

For the metabolomics analysis, 1ml from the samples taken from the ponds (containing both the media and cells) was sent to Creative Proteomics, Shirley, NY, USA for Untargeted Metabolomics Analysis. For the blank, samples with just the media were taken from the ponds before the algal inoculation. For the sample preparation, the samples were transferred into a tube for lyophilization. Following transfer, 500 µL of 80% methanol was added, and the mixture was vortexed for 30 seconds, then subjected to sonication in an ice-water bath for 30 minutes. The samples were subsequently stored at -20 °C for 1 hour, after which they were vortexed for an additional 30 seconds and centrifuged at 12,000 rpm for 10 minutes at 4 °C, to lyse and remove the cells. From the resulting supernatant, 200 µL was transferred into a vial, and 5 µL of a 0.14 mg/mL solution of DL-o-Chlorophenylalanine was added. The final solution was filtered through a 0.22 µm filter in preparation for LC-MS analysis. The separation of compounds was achieved using the Vanquish Flex UPLC system in conjunction with the Q Exactive Plus mass spectrometer (Thermo 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. 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, sheath gas flow rate to 45 arb, auxiliary gas flow rate to 15 arb, sweep gas flow rate to 1 arb, spray voltage to 3.0 kV, capillary temperature to 350 °C, and S-Lens RF level to 30%. For ESI-, the heater temperature remained at 300 °C, with the same sheath and auxiliary gas flow rates as 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 as significant the metabolites whose abundance was more that 10x higher than the abundance of the metabolites in the blank.

297

298 *2.8 Bioinformatics*

299

300 Principal component analysis (PCA) was performed in R with the factoextra package [23].

301 For RNA sequencing analysis, the following sequence processing steps and associated packages were
302 used: The fastq file was quality-checked with FastQC, adapter trimming was performed with Trimmomatic
303 [24], t

304 The pseudoalignment of the transcripts to the transcriptome of *M. minutum* (as it was published and
305 publicly available by [17]) and the quantification of their abundance was done with the Kallisto package
306 [25]. The integration of the Kallisto data and the differential gene expression analysis, including the time
307 course analysis, was performed with Sleuth [26]. The GEMMA package [27] was used for the correlation
308 of the gene expression with specified quantitative phenotypes.

309 Additional tools for the multi-omic analyses include ggplot2 [28] for graphic representation and pheatmap
310 [29] for the heatmap.

311 The de novo motif enrichment analysis among the promoter regions (200bps upstream of ATG) was done
312 with the Homer Motif Discovery and Analysis software [30].

313 We used the Data Integration Analysis for Biomarker discovery using Latent variable approaches
314 for Omics studies (DIABLO) framework from the R software package mixOmics (version 6.28.0) to combine
315 the transcriptomics, proteomics, and metabolomics datasets [31]. Specifically, we compared these data
316 across Day 0 and Day 1 and subsequently Day 4 and Day 5 to identify key features of culture scale up and
317 infection, respectively. Metabolomics and proteomics dataset were log transformed prior to this analysis
318 and transcriptomics data was normalized using transcripts per million (TPM). The proteomic and
319 transcriptomic datasets were further filtered using the nearZeroVar function from the caret R package
320 (version 7.0-1) with a unique cutoff of 20 for the proteomics and selecting transcripts with variances of
321 greater than or equal to 200. In the DIABLO design matrix, we prioritized correlation between the three
322 omics datasets by choosing a 0.6 weighted model. The features from latent component 1, from the Day 0
323 and Day 1 comparison and Day 4 and Day 5 comparison were then compiled for comparison across all
324 four days using the R package pheatmap (version). Values were scaled and centered prior to plotting, and
325 both rows and columns were clustered using Euclidean distance.

326

327 **3. Results**

328

329 *3.1 Phenotypic and conditional measurements reveal distinct acclimation, growth and infection dynamics*
330 *in algae the open pond systems*

331

332 We performed a longitudinal analysis of algal growth in two 1000L open raceway ponds, and introduced
333 an amoeboaphelid pathogen on day 4 of the experiment, tracking a total of 12 variables. Phenotypic

334 measurements included cell count, AFDW, chlorophyl concentration, total carbohydrates, total H3 histone
335 concentration, cell size, caspase activity (Sup. Dataset 1). In addition, we measured five environmental
336 variables within the ponds including %DO, pH, conductivity, temperature, light intensity.

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

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

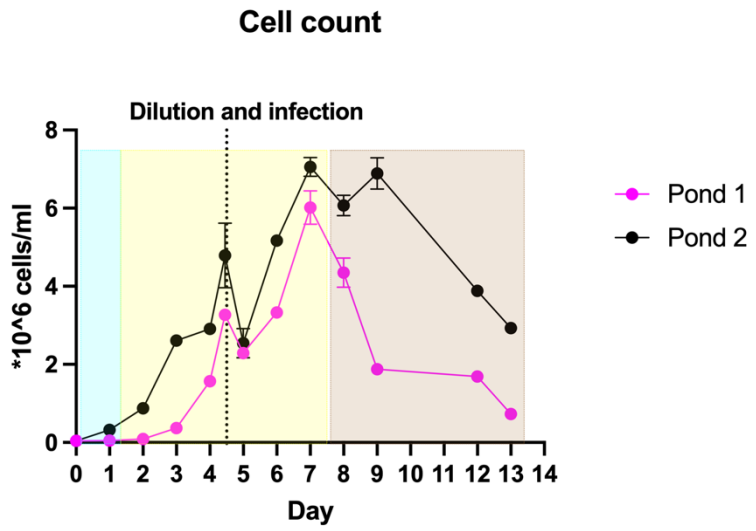
341 b). The log phase (yellow). The cells were fully acclimated to the open pond conditions and growing
342 exponentially. This phase lasted for 4 days before the dilution and the infection and 3 days after.

343 c) The crash phase (brown). This phase is caused by algal death leading to the complete overtake
344 of the ponds by the pathogen and other opportunistic species.

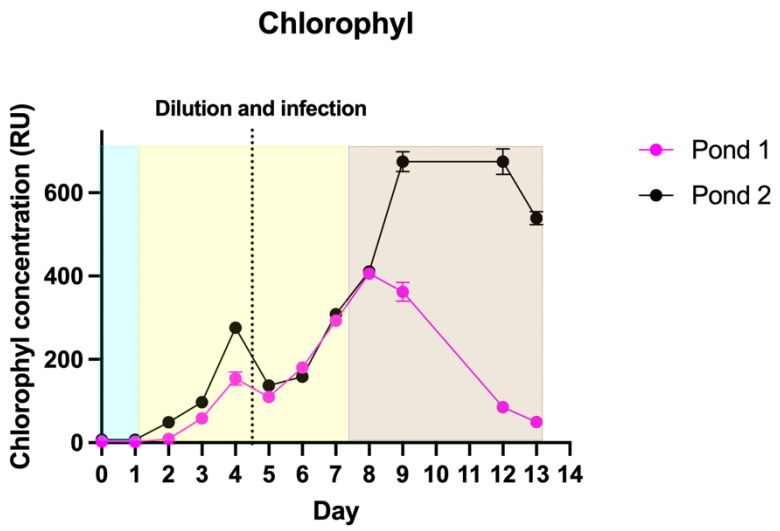
345

346 The ponds behave largely similarly, though we do see divergence of phenotypes at later stages of
347 infection, as documented in Fig. 2.

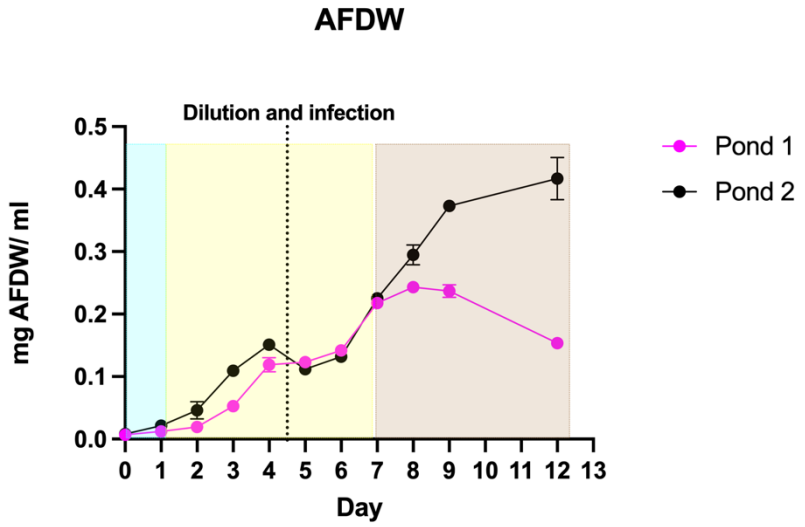
A


Cell count
Dilution and infection
*10^6 cells/ml
0
2
4
6
8
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Day
Pond 1
Pond 2

B


Chlorophyl
Dilution and infection
Chlorophyl concentration (RU)
0
200
400
600
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Day
Pond 1
Pond 2

C


AFDW
Dilution and infection
mg AFDW/ ml
0.0
0.1
0.2
0.3
0.4
0.5
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Day
Pond 1
Pond 2

348

Fig. 2: Growth tracking (as represented by the cell count, chlorophyl count and AFDW) and internal conditional measurements (%DO and light intensity) of the M. minutum cultures at the 1000L opened raceway ponds. Pond 1 is indicated with pink and Pond 2 with black. Day 0 refers to the point immediately after the inoculation of the ponds with the precultures growing in the 20L bubbling carboys. Between day 4 and 5 a 1/3 volume dilution was performed and each culture was infected with antagonistic amoeboaphelids, as indicated. Cyan shows the acclimation period, yellow the log phase and brown the crash period. The error bars represent the variation between the three technical triplicates. A) The cell count between the three ponds. B) Chlorophyl concentration in relative units (RU) from florescence measurements. C) The Ash-free dry weight (AFDW) of algae cells in mg per ml of culture.

To understand whether stress conditions trigger programmed cell death, we measured caspase activity throughout the run. Caspases are proteases that catalyze the breakdown of proteins during programmed cell death [32]. We observed that caspase activity was higher initially, after the pond inoculation, and gradually decreased as the cells acclimated to their new environment (Sup. Fig 2). Of note, we did not see an increase of caspase activity post-infection with the amoeboaphelid, suggesting that the reduction of algal cell count during the post-infection phase is caused by caspase-independent cell death [33] [34].

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

We observe that the three distinct phases (acclimation, log, and crash) are resolved away from each other in the PCA bi-plot shown in Figure 3B, and we can discern a trajectory in PC space as the longitudinal experiment progresses, which is shown by connecting temporally adjacent points. While the largest adjacent transition occurs during scale-up (Day 0 to Day 1), we also observe a significant shift upon 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 transcriptomics, proteomics and metabolomics across the first 6 days of this experiment.

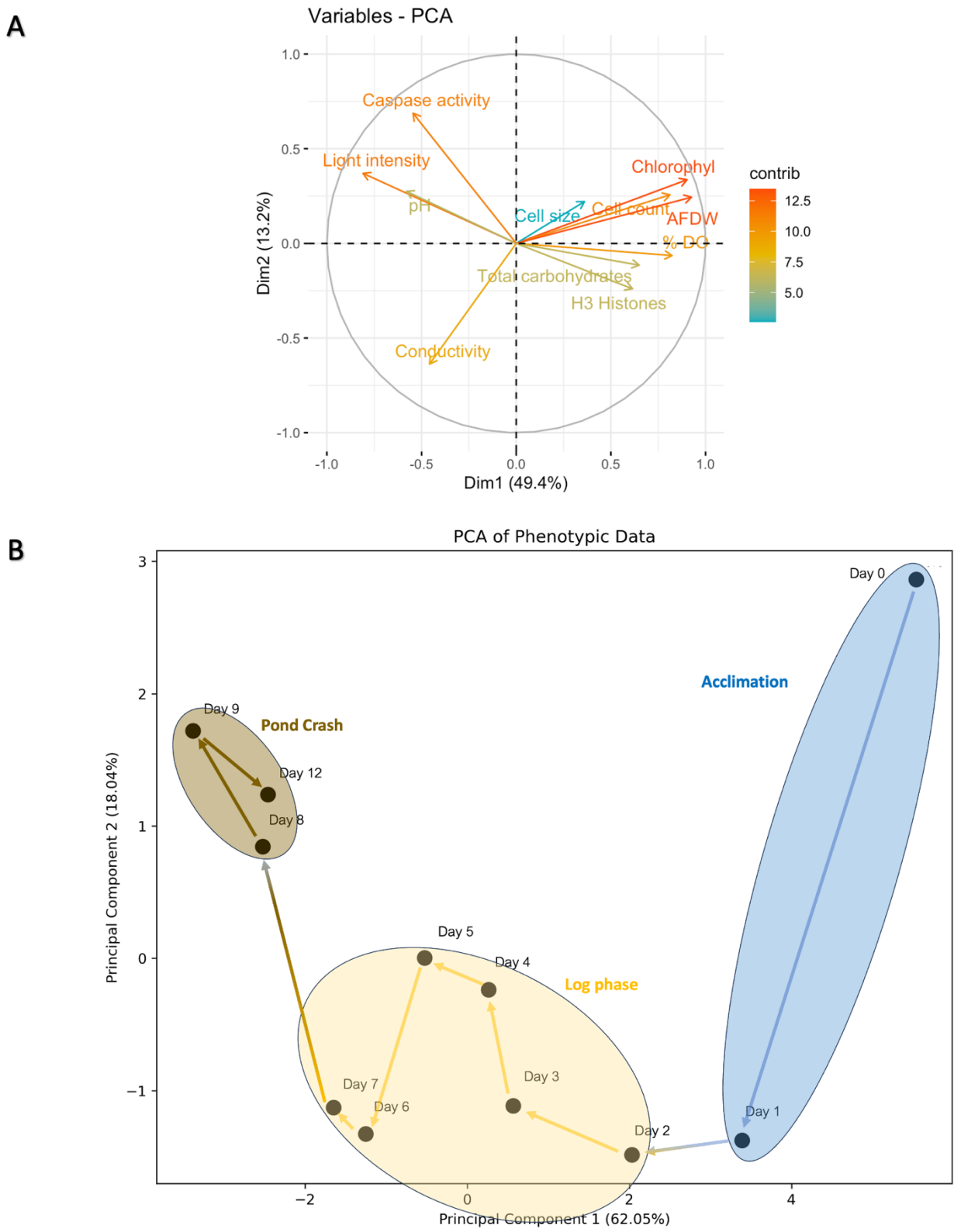


Fig. 3: PCA for 12 variables (cell count, AFDW, chlorophyll concentration, total carbohydrates, total H3 histone concentration, cell size, caspase activity, %DO, pH, conductivity, temperature, light intensity) as measured at the two ponds. A) PCA for the 12 variables used. The color of each line represents 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 (Sup. Dataset 2). The samples thus represented the acclimation period (day 0 and day 1), the pre-infection log phase (day 2- day 4) and the early infection days (day 5 and 6).

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

We performed PCA on the transcriptomic data to determine whether the longitudinal trajectory was comparable to that of our phenotypic and environmental variables (Fig. 4B). As with Fig 3B, the transition from Day 0 to Day 1 was the most striking, showcasing the rapid and large transcriptomic response to the 1000 L pond conditions. The distribution of the days corresponding to the pre-infection log phase, on the other hand, are closer in PC space, reflecting transcriptional stability during that growing stage. A clear distinction between the distribution of the pre-infection and post-infection days was also seen, as a result of both the culture dilution and the introduction of the pathogen. In fact, unlike in Fig 3B where the measured environmental and phenotypic variables post-infection were in a distinct cluster away from the scale-up, the early post-infection transcriptomic profile trends back towards the acclimation phase profile, while still remaining distinct from them.

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 region 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 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 genes promoters, we also observed two motifs (labeled by HOMER 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 contain a motif which can be recognized by Auxin Response Factor 1 (ARF1), a transcription factor identified in *Arabidopsis thaliana.*

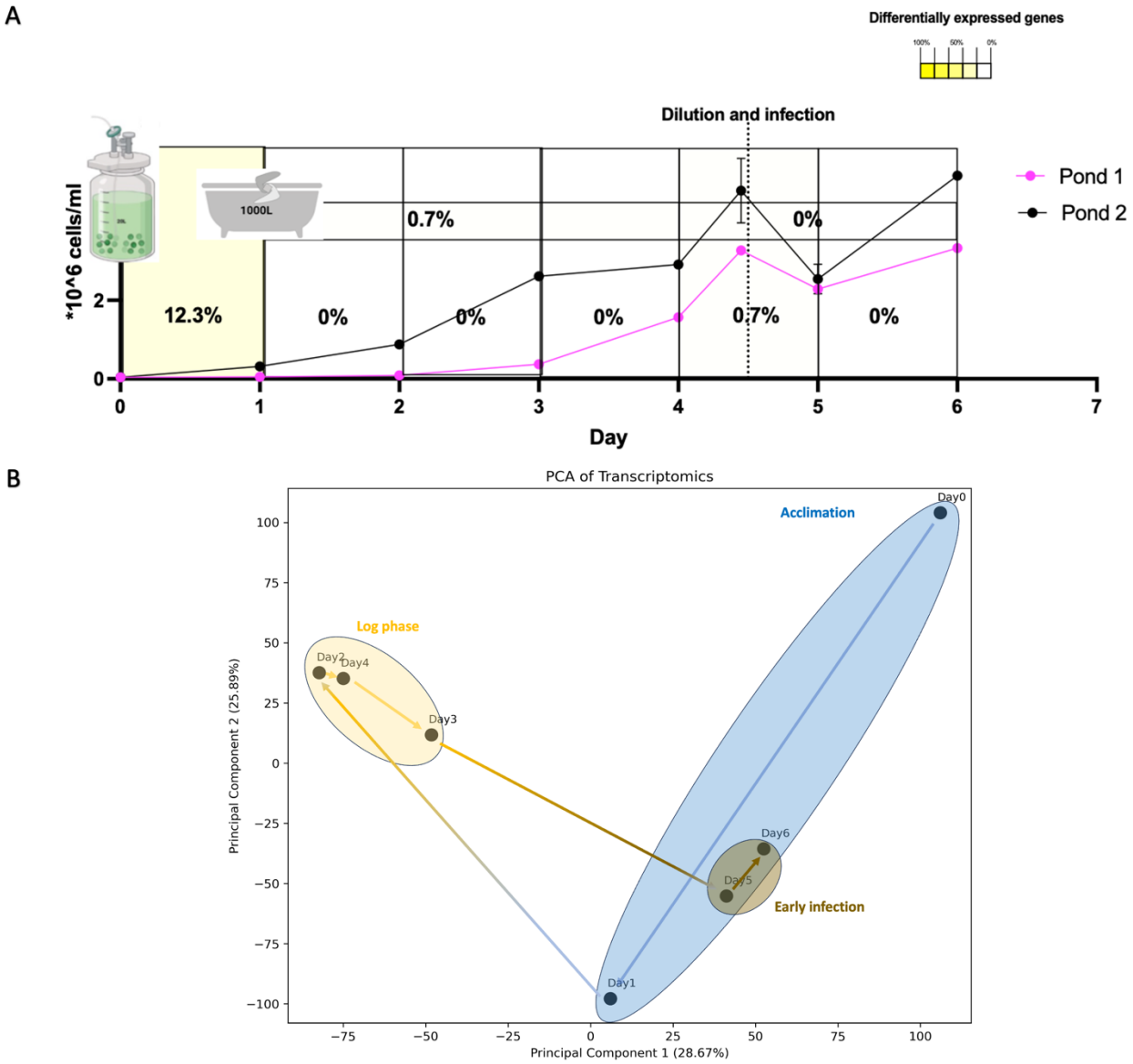


Fig. 4: Comparative analysis of the transcript abundances overlaid with cell growth results. A) Growth curve of both ponds over the time course, with transcriptomic data collected on the same day as growth was measured (as shown by points on the curve). Percentages of differentially expressed genes between specific daily points within the cultivation are shown by the numbers. B) PCA of the transcriptomics. Each day was individually plotted in principal component space. The arrows *indicate the trajectory of the time course.*

	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 1: The most significantly enriched motifs within the promoter regions (200bp upstream of ATG) of the significantly up- or down-regulated genes during the transition from day 0 to day 1.

	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)

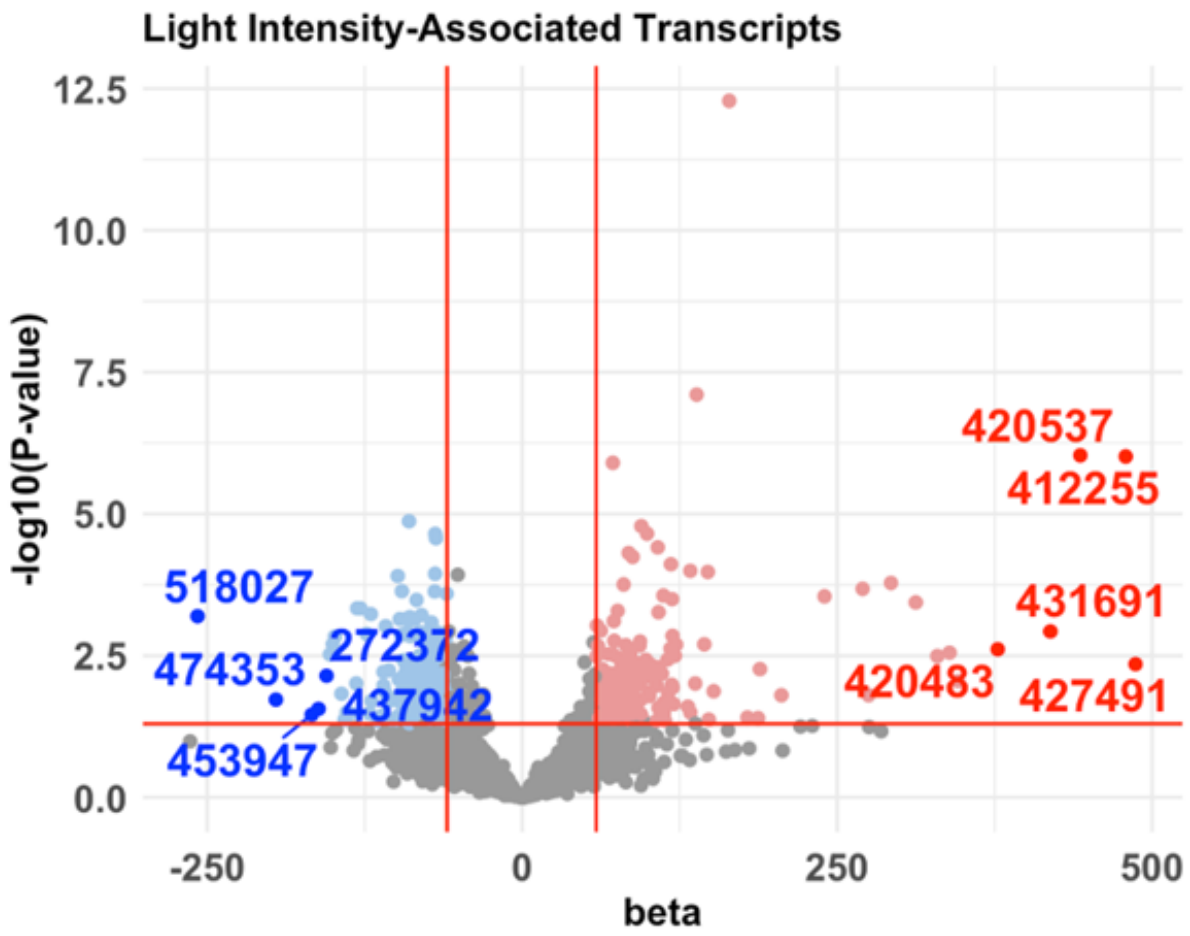
Table 2: The most significantly enriched motifs within the promoter regions (200bp upstream of ATG) of the significantly up- or down-regulated genes during the transition from day 4 to day 5.

3.3 Transcriptomic and phenotypic data reveal 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 like *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 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 is positively or negatively correlated with light intensity and %DO across samples (Fig 5 and 6).

From observing the top up- and down- regulated genes correlated with light intensity (Tables 3 and 4), it was evident that there is 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 that are down-regulated in response to light, which is consistent with previous literature showing that photosynthetic organisms down-regulate photosystem components in response to light stress [35].

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



454

455 *Fig. 5: GEMMA analysis, revealing the genes which are correlated with light intensity. Volcano plot showing the significance*
456 *and the extent of the correlation between the light intensity and the transcript abundance. The top 5 up- and down- regulated*
457 *genes by light are indicated.*

458

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 3: The top 5 upregulated genes with light with their annotations.

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 4: The top 5 downregulated genes with light with their annotations.

	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 5: The most significantly enriched motifs within the promoter regions (200bp upstream of ATG) of the genes which got significantly upregulated by light. Indicated next to each motif are the % of targets that each motif is present, the p-value and the best match with known binding motifs of studied transcription factors. Analysis done by Homer.

	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)

Table 6: The most significantly enriched motifs within the promoter regions (200bp upstream of ATG) of the genes which got significantly downregulated by light. Indicated next to each motif are the % of targets that each motif is present, the p-value and the best match with known binding motifs of studied transcription factors. Analysis done by Homer.

A similar GEMMA analysis was repeated with the genes most correlated with %DO. Among the ten most up-regulated and down-regulated 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 up-regulated genes correlated with DO. We performed motif analysis on the promoter regions of all the genes correlated with DO (Tables 9 and 10). We saw enrichment of the binding site for the same transcriptional factor from the diatom *Thalassiosira pseudonana* at both the up- and down- regulated genes from the DO.

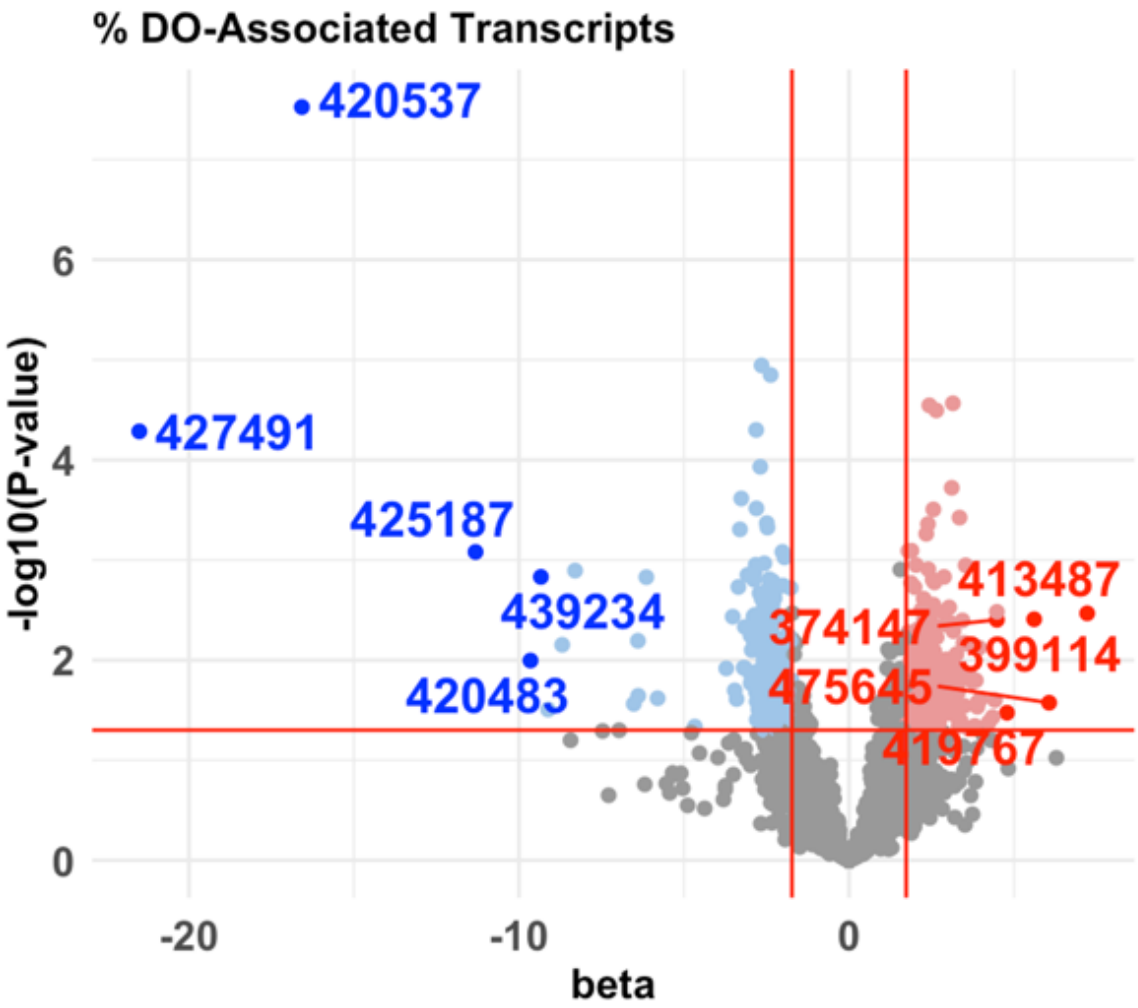


Fig. 6: GEMMA analysis, revealing genes correlated with %DO. Volcano plot showing the significance and the extent of the correlation between the light intensity and the transcript abundance. The top 5 up- and down- regulated genes by %DO are indicated.

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

484 *Table 7: The top 5 upregulated genes with %DO with their annotations.*

485

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

486 *Table 8: The top 5 downregulated genes with %DO with their annotations.*

487

	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 9: The most significantly enriched motifs within the promoter regions (200bp upstream of ATG) of the genes which got significantly upregulated by %DO. Indicated next to each motif are the % of targets that each motif is present, the p-value and the best match with known binding motifs of studied transcription factors. Analysis done by Homer.

	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 10: The most significantly enriched motifs within the promoter regions (200bp upstream of ATG) of the genes which got significantly downregulated by %DO. Indicated next to each motif are the % of targets that each motif is present, the p-value and the best match with known binding motifs of studied transcription factors. Analysis done by Homer.

3.4 The 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 (Sup. Dataset 3) revealed a highly dynamic phase during the first two days of the culture (Fig.7). After entering the log phase, however, the proteome gets progressively more stable, while at the same time maintaining a distinct profile between the pre- and post-infection periods. This was in contrast to the more dynamic changes of 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 on the correlations between specific transcripts and proteins will be discussed later in this article.

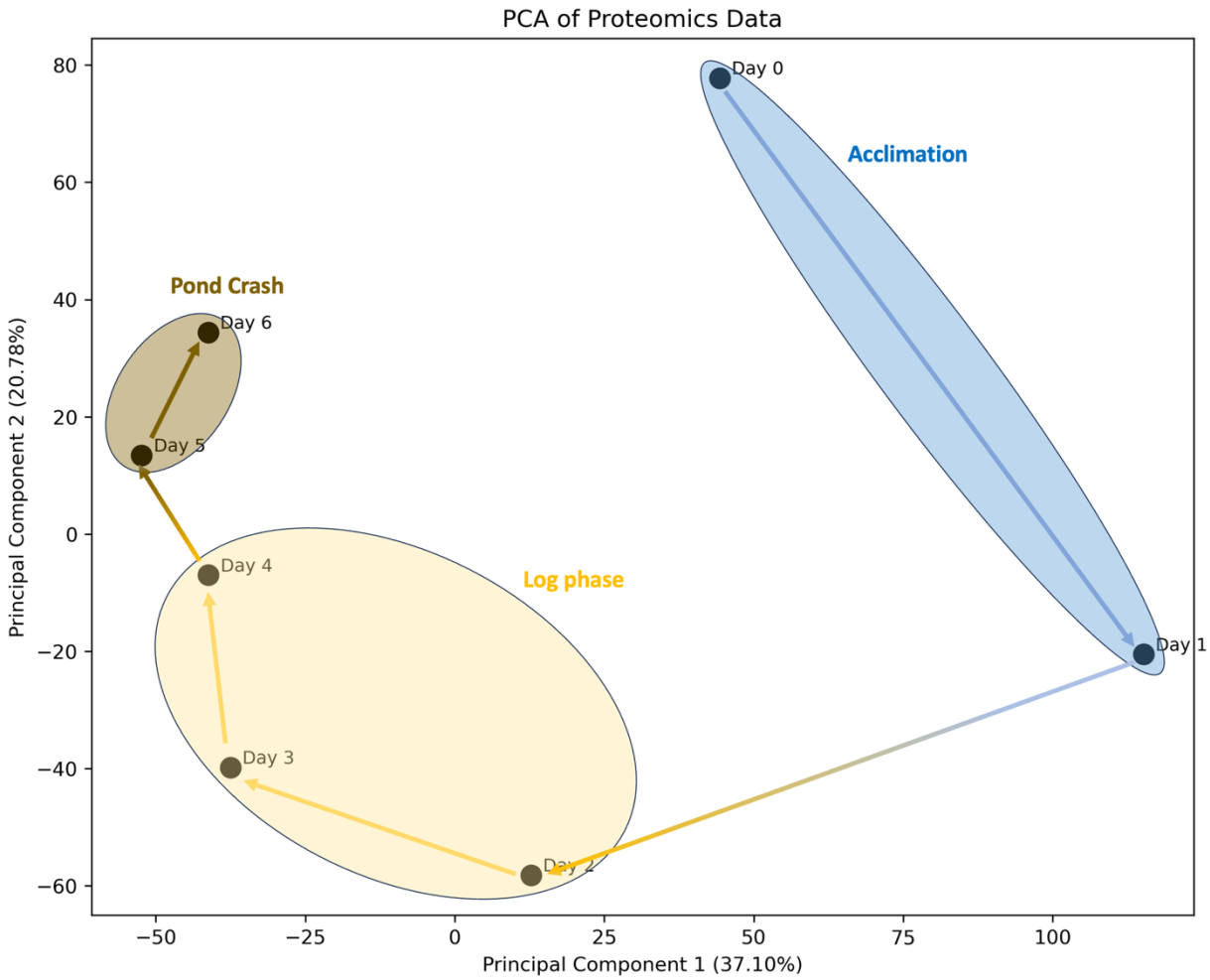


Fig 7: PCA of the proteomics. Each day was plotted as individual point in principal component space. The arrows indicate the trajectory of the time course.

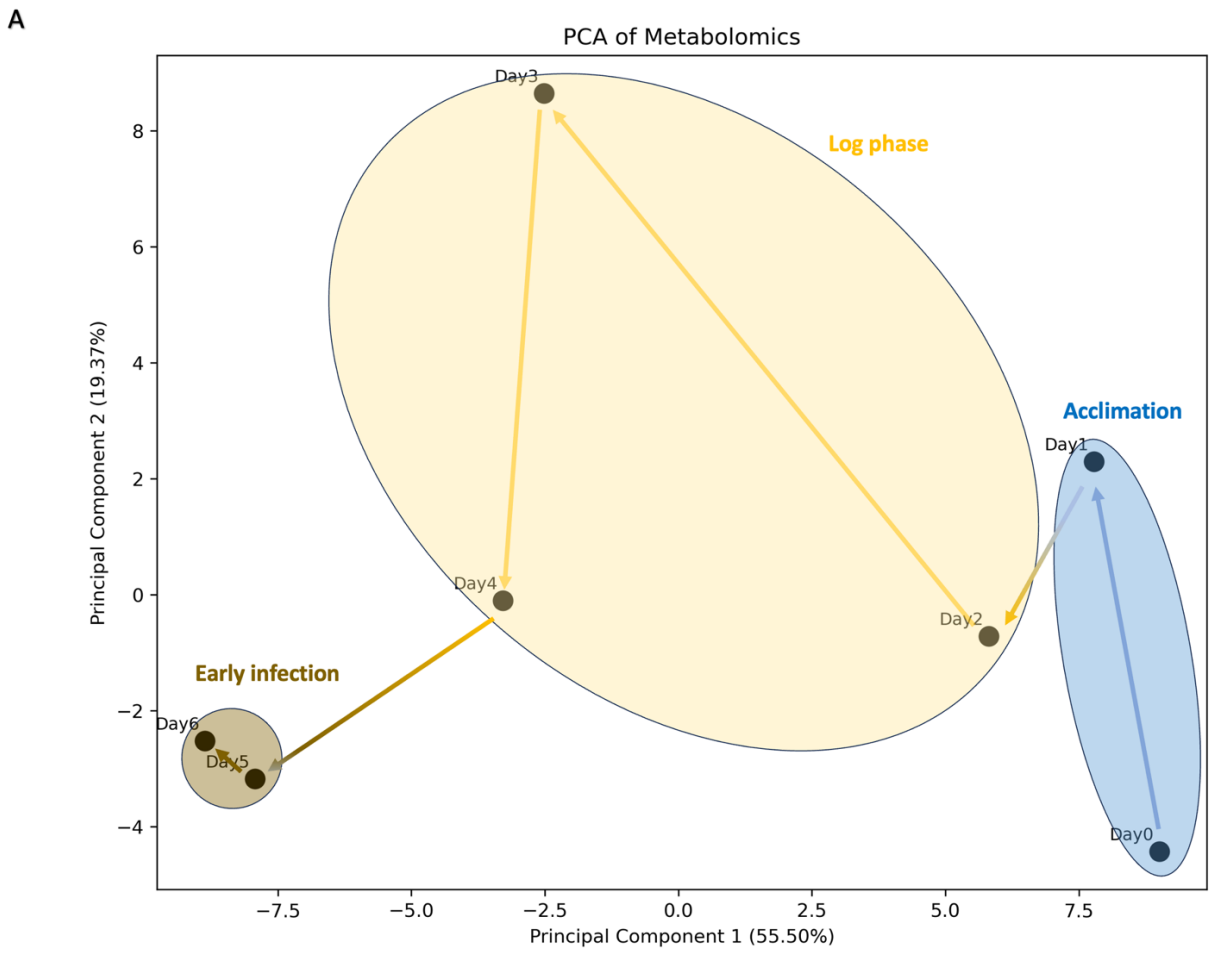
3.5 Metabolomics analysis reveals potential indicators of stress response and pathogen infection.

PCA of the metabolites (Sup. Dataset 4) shows distinct profiles during the three phases of cultivation similar to the transcriptomic and proteomic PCAs (Fig. 8A). 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. Where log phase was linked with transcriptomic stability, the metabolic profile was very dynamic during the growth phase.

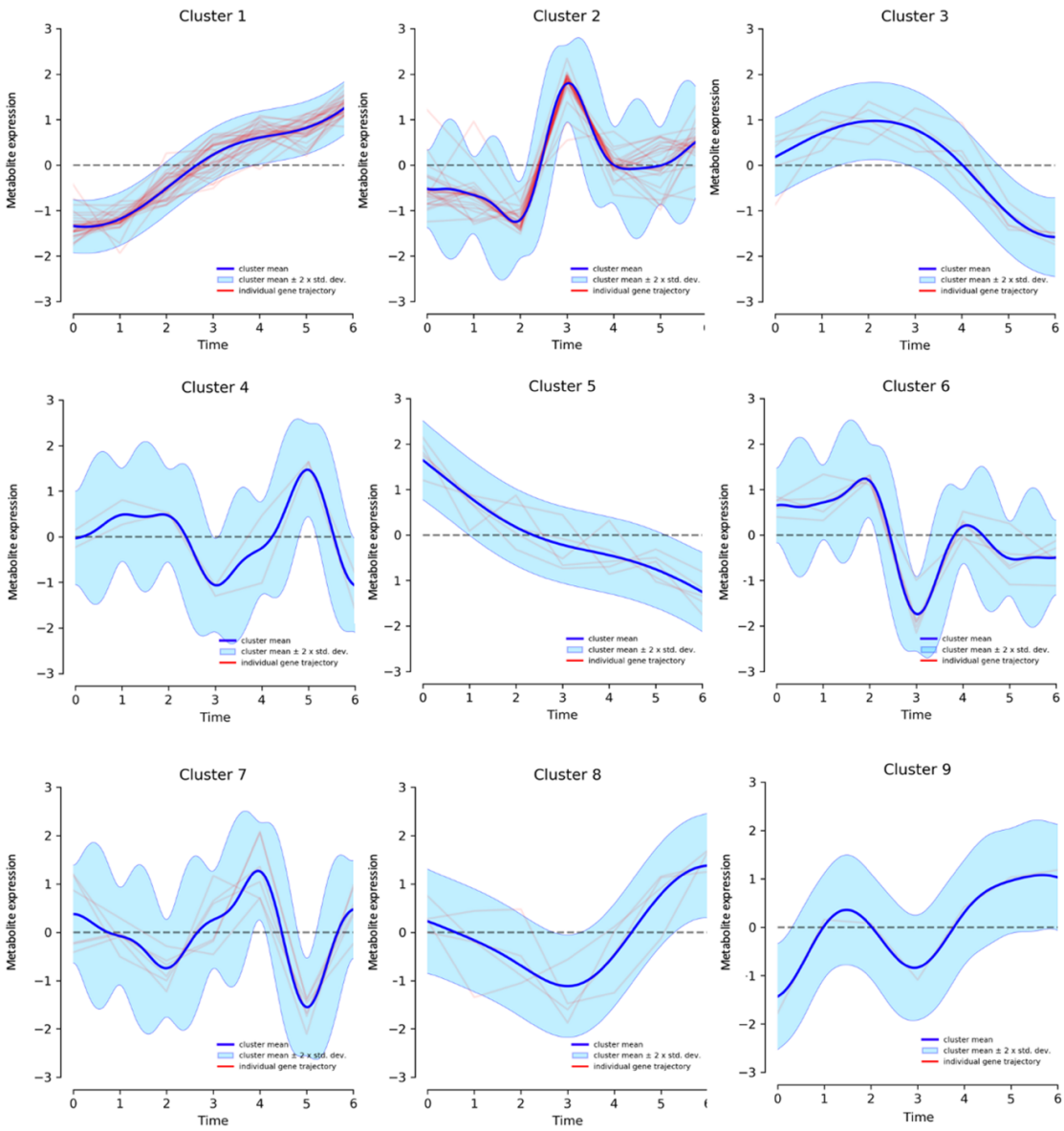
We performed clustering analysis of the longitudinal metabolomics data to further understand metabolites that co-varied. We used DPGP [36], a non-parametric clustering algorithm that can factor in longitudinal information, to separate metabolites into clusters based on their patterns in abundance. We found may metabolites either had their peak or trough on day 3 or 4, the mid-point of the log phase (clusters 2, 4, 6 and 8 and 9). The most abundant cluster (Cluster 1) shows metabolites that 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. 8C, highlights the emergence of three groupings of metabolites, and are in relative accordance with the DPGP clustering. The first group consists of the metabolites whose abundance peaked in lag phase. The second group peaked in the log phase. The last group, which was the largest by number, contains the metabolites that reach their peak abundance after the infection. Exclusively in the last group we found hormones and steroids, such as 5-Dihydrotestosterone, 7-Hydroxytestosterone, Estrone, Estriol and 19-Norandrostenedion.



B


Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Metabolite expression
Time
cluster mean
cluster mean ± 2 x std. dev.
individual gene trajectory

535

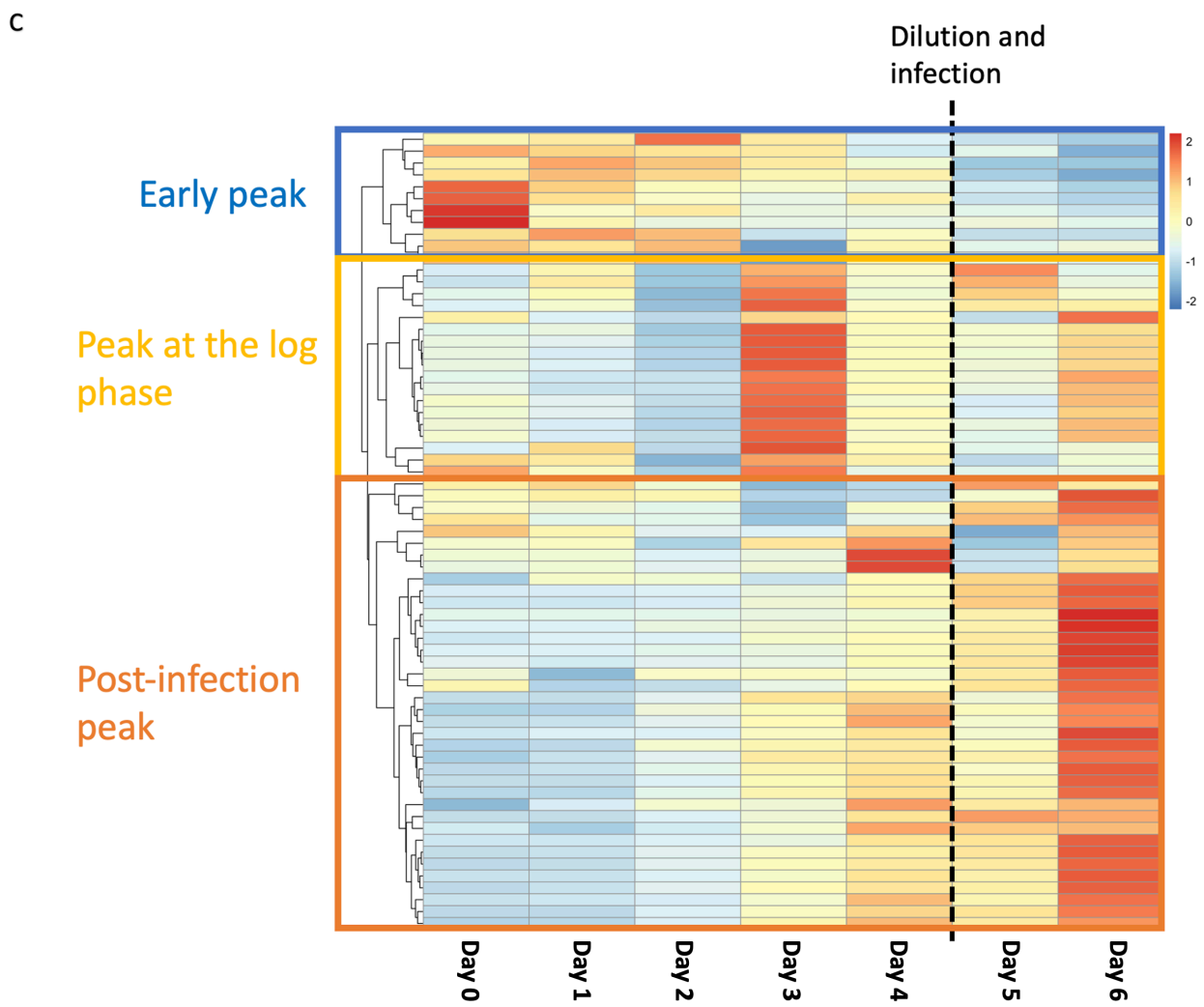


Fig. 8: A) PCA of the metabolomics. Each day was plotted as an individual point in 2-dimensional principal component space. The arrows indicate the trajectory of the time course. B) Clustering analysis of the abundance of the metabolites throughout the studied period, via DPGP C) Heatmap with the hierarchical 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 global relationships in proteins, metabolites and transcripts across our longitudinal experiment, we performed an integrative analysis using the DIABLO framework in the mixOmics R package [31]. We performed this analysis using all omics data from day 0 and 1 to explore the molecular features of the scale up stress, and separately, using day 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, in common

across the two stressors, suggesting more convergence at the metabolite level than 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 saw that the two transitions had almost entirely unique patterns. We saw a cluster of transcripts (all significantly differentially expressed between day 0 and 1), proteins, as well as myristic acid elevated on day 0 relative to the other days. Similarly, we saw another distinct cluster of transcripts and proteins, as well as 1,5-hexadiene and 2,3-Diaminosalicylic acid, elevated on Day 4 compared to the other days. Whereas a decrease in a different cluster of proteins as well as maltose and hypogeic acid distinguished day 1 relative to the other days. This result suggests distinct molecular and metabolic profiles in response to different stressors, pointing to potential unique biomarkers and target pathways for detection and intervention.

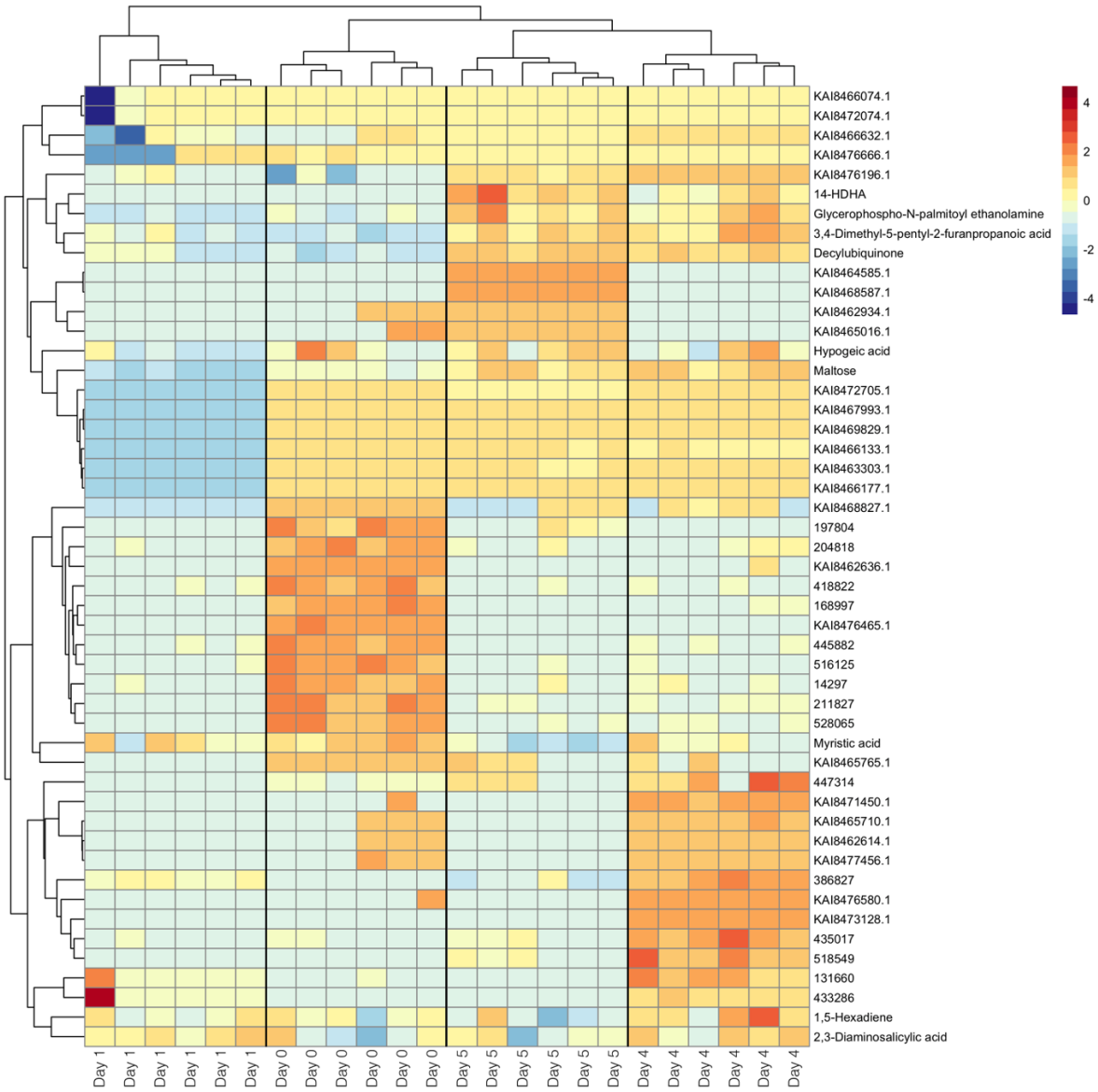


Fig. 9: Heatmap containing the molecular features (genes, proteins and metabolites) which changes significantly their abundance during the transition from day 0 to day 1 and/or from day 4 to day 5. Their abundance is being indicated by a scaled blue-to-red scheme. Each of the 6 biological replicates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

4. Discussion

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

576 We performed a multi-omic analysis of *M. minutum* 26B-AM cultivated over 12 days in 1000 L open
577 raceways ponds to explore the molecular signatures that manifest in response to stressors often
578 encountered at large scale. The three distinct growth phases we observed in the 1000 L ponds were in
579 line with what is typically seen in large scale, batch cultivations of algae [37] [38]. Culture scale-up is often
580 evidenced by a lag phase as we saw in our cultures, especially when moving from well-controlled
581 laboratory cultures into open raceway ponds [37] [39] [40]. Change in scale represents a significant
582 stressor for the cultures, introducing a combination of different environment signals, including changes in
583 light, radiation, heat, gas distribution, as well as microbial and environmental particle composition that
584 requires an acclimation period, which is reflected in our growth curves over day 0 and 1.

585 Similarly, algal pests are a common stressor experienced in open raceway ponds and an active area of
586 study. Changes in phenotypic measurements in response to infection or even visual indications of
587 infection stress, typically appear sometime after the initial infection. In our cultures, we introduced the
588 amoebaphelid on day 4 but didn't begin to see the decline in growth, emblematic of a culture crash, until
589 day 8, accompanied of a color change in the culture from light green to brown. Pond crashes like this,
590 driven by algal pathogens and often damaging crops overnight, account for almost 30% of algal crop losses
591 annually [41], driving up the overall cost of algal applications. Despite the dramatic effects of those
592 infections on ponds, the cellular mechanism of the infection which includes the insertion of the parasite
593 in the host cell [19] allows a certain time of incubation before a noticeable crash [42]. Cellular and
594 molecular alterations are already in place and if they can be detected in a timely manner, the irreversible
595 effects of a pond crash can potentially be avoided, or at least resource allocation decisions can be made
596 much earlier.

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

604 With respect to infection, current strategies for early indication of contamination include visual detection
605 strategies such as flow cytometry [43], in situ microscopy [44], multispectral image analysis [45] and
606 photosynthetic measurements [46] [47]. Those methodologies, though relatively quick, require a certain
607 pathogen titer in the culture. Biomolecular approaches include 16S and 18S sequencing [48], and the
608 detection of Volatile Organic Compounds (VOC), as indirect early indicators of an infection [49]. Here we
609 show that the algae can also be used as early indicators of infection, as demonstrated by the changes in
610 transcripts.

611 Because many of the transcripts in *M. minutum* are poorly annotated, we used motif analysis to further
612 explore the significantly differentially expressed genes. Most identified motifs were unique to each
613 transition, potentially indicating two distinct regulatory pathways between the scale adaptation and
614 pathogen stress. However, common elements between the two datasets exist, implying that there may
615 be factors that could be used to co-regulate responses to different stressors (either endogenously, or
616 using synthetic biology). An example of a potential point of convergence are the motifs recognized by the

transcriptional factors mitochondrial retrograde 1 (RTG1) and 3 (RTG3), both of which are originally identified in yeast and are involved in mitochondrial metabolic regulation.

Among the promoters of genes differentially expressed during scale up, we also observed two motifs labeled by HOMER as ZNF549 and ZNF415, that 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 algae 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 which 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 specific molecular responses.

A survey of the metabolites released by the algae upon encountering various stressors can also provide valuable insights into predictive markers, pond health, 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. Where log phase was linked with transcriptomic stability, the metabolic profile was very dynamic during the growth phase.

These metabolites have a potential importance, 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 found 14-HDHA, a hydroxy docosahexaenoic acid oxylipin which is derived from docosahexaenoic acid (DHA). Production of oxylipins from algae is linked with oxidative stress [54], suggesting a possibly similarly regulation 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 of those compounds playing a role in the molecular signaling or progression of the infection is one that deserves a closer investigation for many reasons, including the discovery of early biomarkers for an upcoming pond crash.

4.2 Transcriptomics and phenotypical observations reflect light and gas-regulated responses in M. minutum

Light and gas constitute the two cornerstone environmental factors for aquatic organisms, especially the photosynthetic organisms such as green microalgae. Indeed, light and CO_2 availability are essential not only for photosynthesis, the primary metabolic function of microalgae, but also photoprotection, lipid

656 composition, and carbon concentration [60] [61] [62] [63] [64]. The significance of these two factors led
657 us to exploring which *M. minutum* genes corelate with them.

658 We observed that translation-related proteins were strongly correlated with high light intensity. Protein
659 translation is an essential regulatory step for phenotypic acclimation, and its control by light has been
660 observed in other algae, including the Chlorophyte *Ulva prolifera* [65], the Charophyte *Klebsormidium*
661 *nitens* [66], and *Chlorella sp.* [67]. In addition, we noted a structural component of Photosystem I among
662 the genes that are down-regulated in response to light. Photosystem I is a crucial component of the
663 electron transport chain, responsible for the light-mediated transfer of electrons across the thylakoid
664 membrane. Excess light can cause oxidative stress and subsequent photoinhibition of Photosystem I, and
665 in many photosynthetic organisms, including in *Chlamydomonas reinhardtii*, light mediates the
666 suppression of the expression of structural proteins of Photosystem I to protect the machinery from this
667 type of damage [68]. Future work to test this hypothesis across algal species will be critical.

668 Motif analysis of the promoter regions of genes correlated with light highlighted two motifs related
669 homeobox proteins in higher plants. The *Zea mays* homeobox 2a (ZmHOX2a) protein, whose expression
670 in maize is restricted to meristems, is involved in in plant development and growth [69]. In addition, this
671 motif might more broadly represent homeobox transcription factors, which are known to regulate
672 environmental responses in plants [70] [71], suggesting a possible evolutionary conservation of light-
673 induced gene regulation through homeobox transcriptional factors across the plants and algae. A motif
674 corresponding to the same transcription factor is also found among the motifs from the promoters of the
675 down-regulated genes by light, showcasing a possible involvement of these motifs on both the up- and
676 down-regulation responses to light. Interestingly, 30.80% of the targets of the promoter regions from the
677 up-regulated genes contained a motif corresponding to the transcriptional factor RGT1 (Restores Glucose
678 Transport) from *Saccharomyces cerevisiae*. The protein regulates glucose transport, though the
679 expression of glucose transporters [72], and the connection between light-induced stress and glucose
680 transport has been documented in other organisms [73] [74]. The motifs of the down-regulated genes
681 contain one which is recognized by the *Saccharomyces cerevisiae* stress-related multicopy suppressor of
682 SNF1 mutation protein 4 (Msn4), which controls metabolic response under cellular stress conditions [75].

683 Of the genes correlated with DO, multiple genes relate to photosynthesis and carbon fixation. A
684 phosphoribulokinase/uridine kinase, an enzyme responsible for a key step in the Calvin-Benson cycle, and
685 a light-harvesting protein of Photosystem I are among the most up-regulated genes. Genes related to
686 programmed cell death were positively correlated with DO, an observation that aligns with results from
687 other green algae species linking ROS signaling and programmed cell death [76] [77] [78] [79] [80]. These
688 observations highlight specific molecular changes experienced by large-scale cultures that are likely
689 directly related to environmental fluctuations. Our results also showed that increased DO causes more
690 gene down-regulation than up-regulation overall, especially for the genes indicated in Table 8. Most of
691 these genes are involved with structural components of ribosomes, indicating a strong suppressive
692 relationship between O_2 levels and translational regulation, as was also observed in the case of genes up-
693 regulated with light. In fact, the most down-regulated gene from the DO is the most up-regulated one
694 from light. Since high light is a stress factor, as is low oxygen, this reciprocal relationship between their
695 regulated genes aligns with expectations. The list of the top five down-regulated genes is completed with
696 a P-loop containing nucleoside triphosphate hydrolase, which in higher plants like *Arabidopsis* acts as a
697 signaling molecule involved in ribosome formation and attenuates ROS formation and oxidative stress
698 [81] [82].

Motif analysis on the promoter regions of all the genes correlated with DO revealed enrichment of the binding site for the same transcriptional factor from the diatom *Thalassiosira pseudonana* that we saw 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 intervention. The second most enriched motif among the promoters of the up-regulated genes corresponds to the transcriptional factor Arrestin 1 (ARR1), which in higher plants is responsive to the phytohormones Cytokinins and plays a fundamental role to 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 promoters of down-regulated 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 gene 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 note that we identified overlap in the genes correlated with light and DO and those related with 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 that have expression levels coordinated with both (annotation # 421681, 446308). This result suggests that there are algal biomarkers that can transcend both scale and study and can potentially provide us with predictive markers for tracking large-scale pond culture resiliency.

4.3 Integrated multi-omics and culture dynamics?

Integrative multi-omic analysis can provide a holistic understanding of molecular pathways and outcomes. The mixOmics framework allowed us to explore coordinated movement between molecular features and outcomes. The distinct clusters of features output by DIABLO not only highlighted that the two stressors elicited different molecular responses, but provided distinguishable molecular profiles for each individual day. In addition, this tool aided us in pinpointing correlated features, proposing potential causal relationships for further investigation [85]. Unfortunately, many of the key features highlighted here are poorly annotated and/or their physiological role within *M. minutum* in the context of stress is currently unknown. Future studies repeating this analysis under multiple stress conditions are critical to determine if these molecular signatures and relationships hold true and if their dynamics can be used as reliable predictors of broad-scale stress.

5. Conclusion

In this study we used multi-omics to investigate dynamic responses of the green algae *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]. Our research aimed to utilize 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. At first, we demonstrated the ability to obtain datasets of genes, proteins and metabolites responding to specific stressors or corresponding to certain culture stages and transitions, opening up possibilities for follow-up studies in the fields of strain engineering, biomarkers' development, bioprocessing, as well as in basic cellular studies. We, thus, showcased that potential biomarker information can be extracted from multi-omic data (transcripts, proteins, or metabolites) that can indicate resilient growth and upcoming stressful events, such as prolonged acclimation periods during scale transitions and potential crashes due to the presence of pathogens. Specific genes related to those transitions are shown if Fig.9, alongside proteins and metabolites which are positively correlated with them, showcasing more holistic molecular pathways related to those responses. When it comes particularly to metabolites, specific hormones and steroids, such as 5-Dihydrotestosterone, 7-Hydroxytestosterone, Estrone, Estriol and 19-Norandrostenedion have been identified as highly expressed ones at the early infection days. Hormones and steroids act as signaling molecules, regulating, not only growth development, but also stress adaptation [87] [88] [58]. Given that their abundance increased after the introduction of the pathogen, it can be hypothesized that those compounds, produced by either the pathogen or the algae cells as a 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 can provide pre-emptive information on impending phenotypic changes, which can be useful both for pond monitoring and eventual intervention and mitigation strategies. In addition, if some of these molecules might be acting in defense of the algae, microbiomes can be engineered to produce the molecules as a prophylactic during large-scale culture.

We also see a number of conserved transcription factor binding sites at regulated genes in *Monoraphidium*. While there isn't an abundance of information on transcription factor binding in *Monoraphidium* and related species, the presence of conserved sequence suggests not only that the binding of specific factors is likely to be pivotal in stress-related regulatory mechanisms, but that these mechanisms are likely conserved across algal species. This suggests that the facors identified here are likely to be regulating stress-response in other algal species, and mitigation and intervention strategies can be broadly applied. Future work will focus on testing across species to confirm this finding.

On a broader scale, the collection of molecular data from such expansive cultivation systems can pave the way for analysis aimed at developing predictive power with respect to culture resiliency, as well as potential targets for engineering this resiliency in the future. Indeed, we see transcripts shared in 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

782 bioprocessing, beyond algae cultures, and this study provides a roadmap in that direction, using omic data
783 to optimize large-scale applications.

784

785 *5.1 Limitations of our study and future directions*

786

787 We acknowledge that while this work aims to reveal insights into the molecular dynamics of open raceway
788 pond algal culture, there are a number of outstanding points and questions that we and others can
789 address in future studies. First, we note that this study focuses on a single run of just two 1000 L ponds,
790 and as such is a pilot study whose goal is to reveal potential pathways and patterns for follow-up work. In
791 addition, we collected samples over a limited period of time. Finally, we exposed the cultures to a single
792 antagonist/stressor. While we collect samples in multiple replicates, in order to provide a robust analysis
793 of long-term reproducibility, we would need to conduct multiple large-scale experiments across
794 conditions and over a period of time. However, even with a 2-pond analysis and controlled variables, we
795 have gained considerable insight into both stable and variable phenotypic and molecular mechanisms
796 over time, and we note that this study was necessary to motivate larger, more comprehensive studies in
797 the future. Our goal is to continue collecting data from the ponds, using a range of conditions and algal
798 species, to demonstrate broad applicability of our pipelines and conclusions.

799

800 **Glossary**

801

802 14-HDHA: 14-hydroxy-docosahexaenoic acid

803 AFDW: Ash-free dry weight

804 ARF1: ADP-ribosylation factor 1

805 ARR1: Arabidopsis response regulator 1

806 CCM: Carbon concentration mechanism

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

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

809 DPGP: Dirichlet Process Gaussian Process mixture model

810 Msn4: multicopy suppressor of SNF1 mutation protein 4

811 PCA: principal component analysis

812 ROS: reactive oxygen species

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

814 RTG1: mitochondrial retrograde 1

RTG3: mitochondrial retrograde 3

SNF1: sucrose non-fermenting 1

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.

Author 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 run and maintenance, sampling and phenotyping, omics preparation, 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 review and edited the manuscript.

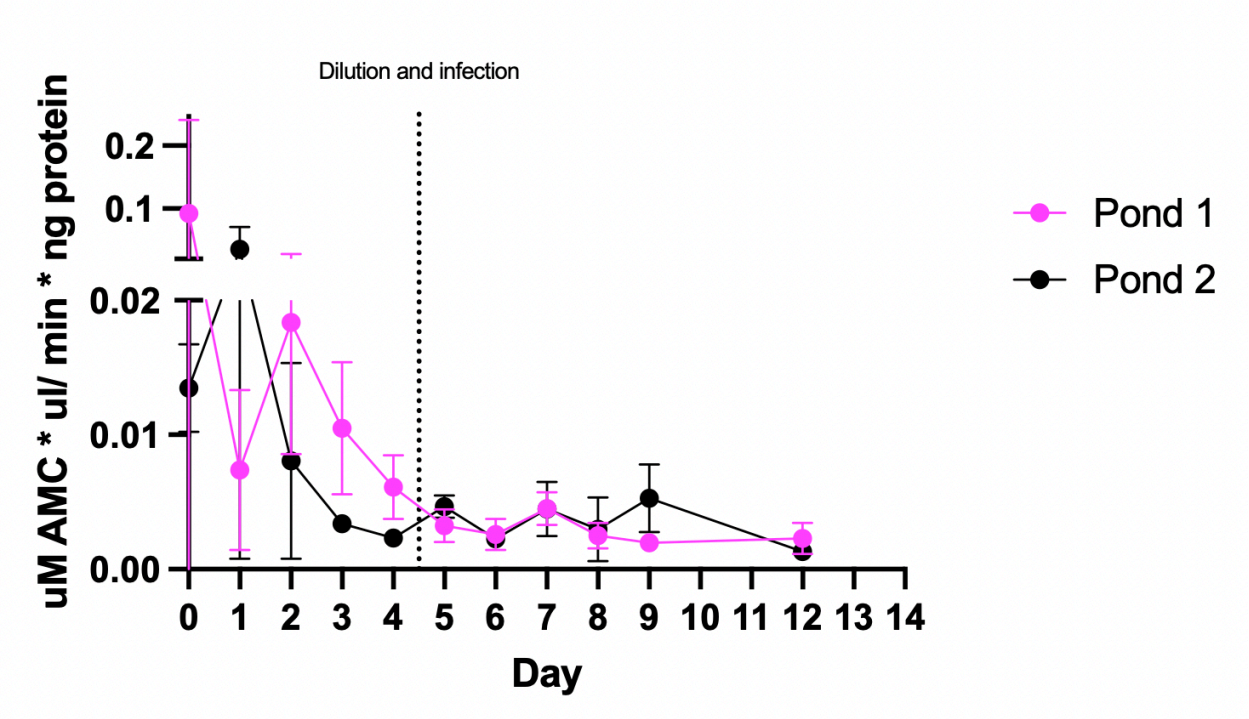
Funding sources

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.

Data availability

Transcriptomic sequencing files can be accessed on NCBI SRA under BioProject PRJNA1307767.

Supplementary data



Sup.Fig. 1: The caspase activity for every day of the culture run in pond 1 (pink) and pond 2 (black)

Sup. Dataset 1: Phenomics

Sup. Dataset 2: Transcriptomics

Sup. Dataset 3: Proteomics

Sup. 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.

1010 66. Serrano-Pérez, E., et al., *Transcriptomic and Metabolomic Response to High Light in the*
1011 *Charophyte Alga Klebsormidium nitens.* Frontiers in Plant Science, 2022. **13**.
1012 67. Wang, B., et al., *Cell damage repair mechanism in a desert green algae Chlorella sp. against UV-B*
1013 *radiation.* Ecotoxicology and Environmental Safety, 2022. **242**: p. 113916.
1014 68. Nama, S., et al., *High light induced changes in organization, protein profile and function of*
1015 *photosynthetic machinery in Chlamydomonas reinhardtii.* Journal of Photochemistry and
1016 Photobiology B: Biology, 2015. **152**: p. 367-376.
1017 69. Kirch, T., et al., *The two homeodomains of the ZmHox2a gene from maize originated as an internal*
1018 *gene duplication and have evolved different target site specificities.* Nucleic Acids Res, 1998.
1019 **26**(20): p. 4714-20.
1020 70. Bondino, H.G. and E.M. Valle, *A small intergenic region drives exclusive tissue-specific expression*
1021 *of the adjacent genes in Arabidopsis thaliana.* BMC Mol Biol, 2009. **10**: p. 95.
1022 71. Ejaz, M., et al., *ARABIDOPSIS THALIANA HOMEOBOX GENE 1 controls plant architecture by locally*
1023 *restricting environmental responses.* Proc Natl Acad Sci U S A, 2021. **118**(17).
1024 72. Polish, J.A., J.-H. Kim, and M. Johnston, *How the Rgt1 Transcription Factor of Saccharomyces*
1025 *cerevisiae Is Regulated by Glucose.* Genetics, 2005. **169**(2): p. 583-594.
1026 73. Valifard, M., et al., *The novel chloroplast glucose transporter pGlcT2 affects adaptation to*
1027 *extended light periods.* Journal of Biological Chemistry, 2023. **299**(6): p. 104741.
1028 74. Powner, M.B. and G. Jeffery, *Light stimulation of mitochondria reduces blood glucose levels.*
1029 Journal of Biophotonics, 2024. **17**(5): p. e202300521.
1030 75. Rajvanshi, P.K., M. Arya, and R. Rajasekharan, *The stress-regulatory transcription factors Msn2*
1031 *and Msn4 regulate fatty acid oxidation in budding yeast.* J Biol Chem, 2017. **292**(45): p. 18628-
1032 18643.
1033 76. Al-Azab, A.J.S., et al., *Chronological transcriptome changes induced by exposure to cyanoacrylate*
1034 *resin nanoparticles in Chlamydomonas reinhardtii with a focus on ROS development and cell wall*
1035 *lysis-related genes.* Algal Research, 2022. **68**: p. 102884.
1036 77. Gao, Y. and D.L. Erdner, *Cell death responses to acute high light mediated by non-photochemical*
1037 *quenching in the dinoflagellate Karenia brevis.* Scientific Reports, 2022. **12**(1): p. 14081.
1038 78. Xu, C., et al., *Programmed cell death process in freshwater Microcystis aeruginosa and marine*
1039 *Phaeocystis globosa induced by a plant derived allelochemical.* Science of The Total Environment,
1040 2022. **838**: p. 156055.
1041 79. Bai, F., et al., *Paraquat induces different programmed cell death patterns in Microcystis*
1042 *aeruginosa and Chlorella luteoviridis.* Ecotoxicology and Environmental Safety, 2023. **249**: p.
1043 114429.
1044 80. Zhao, Y., et al., *ROS-mediated programmed cell death (PCD) of Thalassiosira pseudonana under*
1045 *the stress of BDE-47.* Environmental Pollution, 2020. **262**: p. 114342.
1046 81. Zhao, M.G., Q.Y. Tian, and W.H. Zhang, *Nitric oxide synthase-dependent nitric oxide production is*
1047 *associated with salt tolerance in Arabidopsis.* Plant Physiol, 2007. **144**(1): p. 206-17.
1048 82. Moreau, M., et al., *AtNOS/AtNOA1 is a functional Arabidopsis thaliana cGTPase and not a nitric-*
1049 *oxide synthase.* J Biol Chem, 2008. **283**(47): p. 32957-67.
1050 83. Sakai, H., et al., *ARR1, a Transcription Factor for Genes Immediately Responsive to Cytokinins.*
1051 Science, 2001. **294**(5546): p. 1519-1521.
1052 84. Hashim, Z., et al., *Metabolic profiling of retrograde pathway transcription factors rtg1 and rtg3*
1053 *knockout yeast.* Metabolites, 2014. **4**(3): p. 580-98.
1054 85. Singh, A., et al., *DIABLO: an integrative approach for identifying key molecular drivers from multi-*
1055 *omics assays.* Bioinformatics, 2019. **35**(17): p. 3055-3062.
1056 86. Harun, R., et al., *Bioprocess engineering of microalgae to produce a variety of consumer products.*
1057 Renewable and Sustainable Energy Reviews, 2010. **14**(3): p. 1037-1047.

1058 87. Bajguz, A., *Brassinosteroids in Microalgae: Application for Growth Improvement and Protection*
1059 *Against Abiotic Stresses*, in *Brassinosteroids: Plant Growth and Development*, S. Hayat, et al.,
1060 Editors. 2019, Springer Singapore: Singapore. p. 45-58.
1061 88. Stirk, W.A., et al., *Endogenous brassinosteroids in microalgae exposed to salt and low temperature*
1062 *stress.* European Journal of Phycology, 2018. **53**(3): p. 273-279.

1063