Integrated systems biology and imaging of the smallest free-living eukaryote *Ostreococcus tauri*

Authors: Chuck R. Smallwood[1], Jian-Hua Chen[2], Neeraj Kumar[1], William Chrisler[1], Samuel O. Purvine[1], Jennifer E. Kyle[1], Carrie D. Nicora[1], Rosanne Boudreau[2], Axel Ekman[2], Kim K. Hixson[1], Ronald J. Moore[1], Gerry McDermott[2], William R. Cannon[1] & James E. Evans[1*]

Affiliations:

1) Earth and Biological Sciences, Pacific Northwest National Laboratory, 902 Battelle Blvd., Richland, WA 99354, USA

2) Department of Anatomy, School of Medicine, UCSF, San Francisco, CA 94110, USA; Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Corresponding Author: james.evans@pnnl.gov

Abstract:

Ostreococcus tauri is an ancient phototrophic microalgae that possesses favorable genetic and cellular characteristics for reductionist studies probing biosystem design and dynamics. Here multimodal bioimaging and multi-omics techniques were combined to interrogate *O. tauri* cellular changes in response to variations in bioavailable nitrogen and carbon ratios. Confocal microscopy, stimulated Raman scattering, and cryo-soft x-ray tomography revealed whole cell ultrastructural dynamics and composition while proteomic and lipidomic profiling captured changes at the molecular and macromolecular scale.

Despite several energy dense long-chain triacylglycerol lipids showing more than 40-fold higher abundance under N deprivation, only a few proteins directly associated with lipid biogenesis showed significant expression changes. However, the entire pathway for starch granule biosynthesis was highly upregulated suggesting much of the cellular energy is preferentially directed towards starch over lipid accumulation. Additionally, three of the five most downregulated and five of the ten most upregulated proteins during severe nitrogen depletion were unnamed protein products that warrant additional biochemical analysis and functional annotation to control carbon transformation dynamics in this smallest eukaryote.

Keywords:

systems biology, proteomics, lipidomics, biofuel, microalgae, photosynthesis, oleaginous, triacylglycerol, lipid, starch, carbon transformation, lipid droplet, nitrogen scavenging, soft x-ray tomography, stimulated raman scattering microscopy, fluorescence microscopy

Introduction

Microalgae are ubiquitous in oceans and maintain a major carbon sink in the complex world-wide ecosystem. Spanning more than one-billion years of evolution, microalgae are

phylogenetically diverse and exhibit varying cellular phenotypes that naturally produce high value metabolites, proteins, carbohydrates and energy dense lipids that can be exploited for a wide array of industrial applications [1-3]. Due to their high photosynthetic efficiency for energy conversion, and minimal growth requirements consisting of sustainable resources such as marine or brackish media, light, CO_2 and trace vitamins, microalgae are prime bioproduction platforms [4,5].

Triacylglycerol (TAG) lipids are significantly enhanced when microalgae are subjected to cellular stressors such as light, temperature, and nutrient deprivation [6]. TAG lipids possess nonpolar character and are stored in anhydrous, high-density organelles called lipid bodies, which are desirable for industrial lipid feedstock applications [7]. For other oleaginous algae, TAG production can be triggered by nutrient deprivation of iron, sulfur, nitrogen, phosphate, or silicon [6]. In most eukaryotes, combinatorial reduction of these nutrients results in altered levels of growth-associated structural lipids (phospholipids) and energy storage lipids (TAG) products [8,9]. Unfortunately, in most cases starvation or deprivation can be detrimental to cell viability and overall growth capacity thereby limiting cell biomass yields needed for viable lipid feedstock industrial applications [10]. While several reports have shown that supplementing additional C sources when combined with N depletion for certain organisms can yield increased growth and lipid accumulation rates compared to strict starvation[2,11,12], a detailed understanding of the interplay between carbon and nitrogen bioavailability is needed to advance bioproduction applications.

Studying phototrophic metabolism in primitive species, such as the prasinophyte *Ostreococcus tauri* can provide convenient opportunities to define minimal and critical metabolic pathways for C transformation[13, 14]. *O. tauri* is the smallest known free-living eukaryote (~0.8μm in thickness), lacks a cell wall, and thrives in varying photic, toxic and thermal ecosystems[15,16]. It also has a highly condensed genome with only ~8,000 genes[17] so most reactions are governed

by a single enzyme (rather than multiple duplicating function enzymes) which simplifies engineering requirements and interpretation. For example, the canonical model green alga *Chlamydomonas* has 4 copies of Acetyl-CoA Carboxylase (E.C. 6.4.1.2) whereas *Ostreococcus* has only 1 enzyme of that class. In addition, a recent study[18] reported the genetic diversity associated with large phenotypic differences between *Ostreococcus* strains highlighting the uncharted abundance of genetic biodiversity. These characteristics make *O. tauri* a potential candidate for future industrial applications. However, considerable biodesign efforts will likely be needed to develop an efficient cell factory for controlled and cost-effective bioproduction of lipid feedstocks or other high value metabolites. Here, an integrated analysis of bioimaging, proteomic, and lipidomic characterization was used to investigate *O. tauri* cellular response to varying C:N ratios. Our results provide additional understanding of C storage and energy transformation pathways within the microalgae *O. tauri* and identify new proteins to target for future engineering efforts.

Results

Other than the bioavailable C:N ratio, all other experimental parameters (e.g., temperature, duration, diurnal cycling, light fluence, etc.) were kept constant throughout this study. The four different media conditions used herein are designated as K6CN, K2CN, K2C and K6C which are listed in decreasing C:N ratio. The normal growth media for *O. tauri* is Keller Media (referenced herein as K2CN media) and contains ~2.5mM bicarbonate and 0.9mM total nitrogen. To streamline interpretation, bicarbonate was kept as the sole media carbon source. For K6CN media the nitrogen content was kept equivalent, but the bicarbonate increased to 6mM providing elevated carbon but normal nitrogen conditions. The use of 6mM bicarbonate as the elevated carbon set point was chosen following an initial growth screen of *O. tauri* in K media supplemented with various levels of bicarbonate from 0-10mM where 6mM showed

highest growth response with healthy chlorophyll ratios per cell. The final two media conditions K2C and K6C share the same composition as K2CN and K6CN but are depleted of all nitrogen sources. Thus all subsequent experiments consisted of 4 conditions of decreasing C:N bioavailable ratios from elevated carbon with normal nitrogen (K6CN), to normal carbon with normal nitrogen (K2CN), to normal carbon with no nitrogen (K2C), and elevated carbon with no nitrogen (K6C).

Diurnal cycling with 12:12 hour light:dark cycles was used to synchronize cellular division and growth. Cell growth was monitored by absorbance at 750nm (measure of particulates) and 680nm (measure of chlorophyll a) for up to 144 hours (Fig. 1B). Decreases in absorbance at 680nm were typical of K2C and K6C cell cultures relative to K2CN and K6CN. Absorbance at 750nm exhibited similar decreases for N deprived cultures versus N replete. However, when comparing normal C to excess C for either N replete or N deprived conditions, cultures with excess C consistently displayed higher A680 and A750 values. Confocal fluorescence microscopy was used to compare phenotypes for a couple dozen cells from each condition whereas fluorescence activated cell sorting (FACS) allowed quantitative analysis of larger population dynamics every 24 hours. In both case, dramatic increases of neutral lipid (NL) content were detected for K2C and K6C conditions (Fig. 1 C–F). Interestingly K6CN (Fig. 1C) cell cultures only exhibited subtle differences between NL and some increases in phospholipid (PL) intensity indicating some photosynthetic lipid metabolism difference to K2CN (Fig. 1D). Since confocal microscopy can only track dynamics for fluorescently labelled components, label-free cell ultrastructure and composition changes were also evaluated.

Cryogenic soft x-ray nanotomography (CSXT) was performed on 72-hour cell cultures to increase resolution and highlight native cellular features of intracellular lipid accumulation (Fig. 1E). Dark subcellular features, assigned as vacuoles due to their linear absorption coefficient (LAC), were observed in the intracellular cytosolic space of cells in K6CN and K2CN culture

conditions, but no vacuoles were observed for K2C and K6C cultured cells. The vacuoles were much larger in K6CN compared to K2CN conditions. Based on studies conducted in other algae and plants starch is likely present in the chloroplast[17]. However, intracellular starch was undetected in *O. tauri* cells via X-ray nanotomography likely due to the density appearing similar to or masked by the surrounding tissues with a similar LAC - possibly due to its small cell size or chloroplast packing density. Therefore, stimulated Raman scattering (SRS) microscopy was used as a second label-free imaging method to identify relative abundance of intracellular starch accumulation. Interestingly, starch was detected in the chloroplast for all tested conditions. However, the starch seen for K2C was minimal (Supplemental Figure 1). Considering that K6C represents an even more deprived nitrogen to carbon ratio compared to K2C, it was anticipated that K6C would show the lowest levels of starch since both conditions also show significant lipid accumulation. However, the increase in starch content for K6C suggests that these cells still have abundant carbon available for transformation into carbohydrates, and potential future engineering efforts could focus on knocking out genes associated with starch accumulation to divert this excess energy toward lipid feedstock production instead.

The confocal imaging also happened to capture a possible lipid droplet secretion event. During live-cell imaging, a single cell was captured over a few minutes showing the formation of a cellular bleb containing a single lipid droplet that was released in subsequent scans (Supplemental Figure 2). We have previously reported that this organism does not appear to have any canonical proteins associated with lipid droplet secretion from other organisms[19]. That study also captured static images showing what was described as blebbing intermediates. The current image series reported here is the first case in which the process was observed live thereby lending additional support to the theory that *O. tauri* is capable of lipid droplet secretion although the detailed mechanism remains elusive.

Dramatic changes observed for lipid staining profiles prompted the exploration of underlying global proteomic expression profiles for each experimental culture condition. Cultures were harvested at 24- and 48-hour time points for LC-MS/MS proteomics to obtain a measure of global proteomic expression. These time-points were chosen for comparison since they showed the biggest relative 24-hour changes via FACS and confocal analysis. The proteomics data was mapped into individual metabolic groups (synthesis, degradation, energy, other and non-metabolic pathways) to interpret relative changes of cells cultured in K6CN, K2CN, K2C and K6C conditions for each metabolic pathway (Fig. 2). Of particular note was the increased changes in abundance for both K2C and K6C in the carbohydrate synthesis pathway compared to K2CN and K6CN. However, very few proteins within the fatty acid (FA)/lipid synthesis pathways showed significant, if any, change in abundance despite bioimaging observations of cellular lipid increases up to 60% the volume of cells. For example, while DGAT, the canonical enzyme representing the last committed step of TAG synthesis was detected with global proteomics, it's abundance remained effectively equivalent across all 4 sample conditions.

High C to N ratio of K6C displayed consistently higher proteomic responses compared to K2C for upregulated proteins related to C storage. The complete starch pathway was detected with ascending upregulation from alpha amylase to granule-bound starch synthetase (GBSSI) for N depletion conditions. GBSSI was the third highest upregulated protein overall highlighting the reliance on carbon storage under these conditions. Previous studies have found similar results for GBSSI[20]. N deprivation also caused downregulation of proteins involved in N acquisition, such as nitrate transporters, nitrate reductase with concomitant upregulation of N scavenging proteins for glutamine, asparagine, and urea. Much of the downregulated proteins detected were ribosome based or were proteins localized to the chloroplast.

Overall 471 unnamed protein products were found to be upregulated or downregulated with several exhibiting some of the most extreme abundance changes (Fig. 4). N deplete conditions K2C and K6C had similar distributions of upregulated and downregulated protein trends, with more increased upregulation of proteins for excess carbon cultures. For K6C, 5 of the 10 most upregulated and 3 of the 10 most downregulated were UPP (Fig. 3). Interesting, several of these proteins show inversion of abundance for replete versus depleted N conditions. XP_003075209, XP_003081059, XP_003078347 all exhibit significant upregulation under K6CN and K2CN conditions at 48 hours compared to K2CN at 24 hours, however, these proteins show significant downregulation for K2C and K6C conditions at 48 hours. The biggest change was seen for XP_003078347 which was had a log2 value change of -0.54 for K6CN but a value of 1.96 for K6C. Similarly, other unnamed proteins XP_003084215, XP_003082140 and XP_003082699 were all downregulated for K6CN and K2CN at 48 hours but upregulated for K2C and K6C. Clearly these unnamed proteins are dramatically affected by the bioavailable ratio of C and N and represent interesting targets for future in-depth functional annotation.

Conducting a BLAST alignment analysis on the UPP identified in these runs resolved a predicted membrane protein (XP_003080099) upregulated for N depleted conditions K2C and K6C, which was identified as having a domain with homology to TMEM14, an uncharacterized superfamily believed to be involved in membrane transport of lipids in higher eukaryotes. This protein could possibly play a role in the lipid droplet secretion from *O. tauri*. Other potentially interesting unknown protein products were also uncovered during BLAST alignments with two notable probable identifications: A putative Acyl-CoA N-acyltransferase (XP_003084085.1) slightly upregulated for K2CN and K2C but not K6CN or K6C; and a putative Zinc finger (XP_003080741.1) slightly downregulated for K6CN, K6C, and K2C conditions, which have been known to participate in a number of eukaryotic cellular mechanisms including lipid binding. Additional proteins such as sarcosine-dimethyltransferase (SDMT), an enzyme found in the

174 betaine biosynthesis pathway in other algae and higher plants related to osmoprotection during
175 cellular stress[21,22], were found to be elevated only for K2C and K6C conditions, which may
176 reinforce cellular stability during N stress and lipid accumulation.

177 Despite observing limited changes in the proteome related to lipid metabolism, the visual
178 confocal and FACS analysis provided ample evidence of significant lipid accumulation that
179 warranted further characterization to understand cellular lipid composition and relative
180 abundance. Thus, the same K6CN, K2CN, K2C, and K6C cultures were surveyed at the same
181 24- and 48-hour time-points using LC-MS/MS global lipidomics analysis. Quantitative lipid
182 profiles were collected for more than 280 lipids (Supplemental Figures 3 – 9). The N depletion
183 conditions (K2C and K6C) resulted in significant increases of energy dense TAG lipids with
184 more than 10 TAGs showing >40-fold increase in abundance already at 48-hours. The FACS
185 and confocal analysis showed that the NL content continues to accumulate well beyond 48
186 hours. In comparison, TAG lipid abundance was flat or decreased for K6CN and K2CN
187 conditioned cells, indicating that cells were not diverting C into lipid energy storage during
188 nutrient rich conditions although they were still accumulating starch as seen from SRS
189 (Supplemental Figure 1). The observed FA profiles provide additional support for recently
190 reported[23] unique long chain FAs despite lacking annotated enzymes known to synthesize
191 them, indicating that *O. tauri* is at the very least an intriguing oleaginous organism for lipid
192 feedstock development. In addition, the number of long chain FAs in TAGs were significantly
193 enhanced during K6C conditions, demonstrating this organism's capacity to uptake and
194 transform excess C into long chain energy dense lipids without genetic modification
195 (Supplemental Figures 7–9).

196 Changes in structural lipids were also detected (Supplemental Figures 3 – 6).
197 Thylakoidal membranes are composed of monogalactosyldiacylglycerol (MGDG) and
198 digalactosyldiacylglycerol (DGDG), which stabilize the thylakoid for maximal photosynthetic

efficiency [24]. In addition, algal thylakoid membranes contain abundant amounts of sulphoquinovosyldiacylglycerol (SQDG) which contributes to increased stability in the photosynthetic harvesting complexes and accommodates membrane protein associations unique to microalgae [25]. Increased abundances for MGDG and 3- to 7-fold increases in DGDG were detected for K2C and K6C conditions, with C16 chain lengths exhibiting the most abundant and dramatic changes during N depletion. C16 chain length SQDG lipids were also slightly upregulated for K6C, suggesting they may be adding stability to light harvesting complexes or membrane protein expression unique to excess C exposure and N depletion.

Discussion

Linkages between N and C metabolism related to lipid biogenesis were interrogated for the microalgae *Ostreococcus tauri*. Single cell and population imaging experiments combined with global proteomic and lipidomic experiments on the same cultures demonstrated that even the simplified cell architecture and genome of *O. tauri* displays complicated regulatory linkages as a function of bioavailable carbon to nitrogen ratios. Cryogenic soft x-ray nanotomography revealed distinct lipid droplet distributions of varying sizes confirming initial fluorescence microscopy results. K2C conditioned cells had uniform sized lipid droplets whereas cells in K6C conditions had lipid droplets of varying size including very large lipid droplets that swelled and deformed the cells.

Numerous photosynthetic, structural, and energy storage fatty acid (FA)/lipids were verified through LC–MS/MS lipidomic analysis that included detection of long chain TAGs, ideal for lipid feedstocks. Proteomic data combined with physiological cell responses to varying C and N revealed that although FA/lipid synthesis pathways had little proteomic changes, carbohydrate synthesis proteomics was independently upregulated. Both ribulose 1,5-biphosphate carboxylase/oxygenase large and small subunits (Rubisco) were slightly

downregulated for K2C and K6C relative to K2CN suggesting Rubisco expression in *Ostreococcus* contains minimum concentrations of Rubisco to support normal growth as reported for other organisms[26]. Restricting protein identifications to subsets related to carbohydrate conversion, N scavenging, and energy regulation allowed for the simplification of biological interpretations related to lipid feedstock optimization targets. The most upregulated protein was a protein kinase involved in serine/threonine phosphorylation and was equally upregulated for K2C, K6C, and K6CN conditions, which could possibly play a role in regulating diurnal cycling or stress response since K6CN could also be seen as another N deprived state relative to available C. GBSSI was the third most upregulated protein in all and was more upregulated in K6C than K2C. Knock-down of GBSSI could potentially provide a redirection of *O. tauri* metabolism to less starch accumulation and more lipid accumulation under K6C conditions.

In addition to the proteins related to energy storage, increased abundances of proteins related to lipid viability, lipid production, and osmotic shock were detected to varying degrees. Finally, a surprising number of proteins of unknown function and identity were revealed to be part of the 10 most significant proteomic increases and decreases overall and each is a potential target for future engineering efforts. Furthermore, the wealth of undefined proteins emphasizes the extent of unexplored opportunities related to N, C, and energy storage pathways, and highlights the peculiar genetic diversity within natural populations of *Ostreococcus* and possibly other primitive marine species.

Methods:

Strain maintenance, culture growth media, and nutrient starvation conditions

O. tauri cell cultures were obtained from the Roscoff Culture Collection (RCC745); strain name: OTTH0595, which has been fully sequenced[17]. Cultures of RCC745 were grown in defined

Keller (K) media[27] with normal or depleted N and HCO^{3-}: K2CN contained normal N and C, K6CN with normal N and 6mM HCO^{3-}, K2C with depleted N and 2.5mM HCO^{3-}, and K6C with depleted N and 6mM HCO^{3-}. All Keller media-based culture conditions were prepared in fresh artificial seawater (ASW) with defined amounts of nutrients analytically prepared fresh and sterile filtered prior to each experiment. To monitor growth and lipid accumulation over time absorbance at 680nm and 750nm was measured to obtain both values for chlorophyll content and particulate matter, respectively, for each culture. Graphing the ratio of 750nm/680nm provided a measure of chlorophyll functional efficiency as well as possible lipid particulates in solution. Graphical analysis of each growth and starvation curve required minimal normalization due to our consistent efforts in capturing cells during mid-log stages of growth. For growth and lipid accumulation studies cells were grown initially in normal K media to approximately 0.03 at OD680 then gently centrifuging cultures at 1200XG for 10 minutes in a swing bucket rotor, washed with respective defined K media, and resuspending cells into defined K media at a target 0.03 at OD680 and continued diurnal light entrainment for specified time courses in sealable CytoOne non-treated cell culture flasks (USA Scientific, USA) with mixing of cultures once per 24 hours. To prepare cell cultures for starvation surveys they were gently centrifuged fresh cultures at 2200xG for 10 mins with swing bucket rotor centrifugation and washed cell pellets once with defined K media of interest then suspended cells in defined media conditions and continued diurnal light entrainment for specified time courses in sealable CytoOne non-treated cell culture flasks (USA Scientific, USA) with mixing of cultures once per 24 hours.

Fluorescence activated cell sorting analysis of intracellular lipid content

O. tauri cells were cultured to mid log phase and gently centrifuged to concentration then stained with Nile red lipid stain for exactly 10 mins before each experimental measurement on the BD INFLUX flow cytometer (BD Biosciences, San Jose, CA, USA). FSC and SSC were

used to gate out any non-specific cellular debris. Specific gating in the range of known cell size of *O. tauri* was used to determine the fluorescence from stained neutral lipid (488/542±13.5 nm), phospholipid (561/615±12 nm), and natural chlorophyll autofluorescence (640/670±15 nm) for defined populations of cells. Each individual FACS experiment was calibrated to 3.6 ide scatter 10 mins before running our sample measurements in defined media cultures. The fluorescence intensity of neutral lipid and phospholipid fluorescence intensity at specific time points was compared in scatter plots to demonstrate population dynamics for each sample condition. K2CN at 24 hours was used as the baseline for normal conditions to detect changes due to varying C:N bioavailability.

Fluorescence and SRS confocal microscopy

Confocal images were obtained on Zeiss LSM 710 (Carl Zeiss AG, Germany) confocal microscope with a 100x oil immersion objective. An InTune Laser with 505nm and 535nm light was used to maximize the separation of the triglyceride (585nm) and phospholipid (638nm) emission peaks while diminishing crosstalk of the Nile Red Stained cells. In addition, chlorophyll autofluorescence was excited with 405nm light and monitored the emission profile at 680nm. Nile Red stained cells were immobilized on glass slides with poly-L-lysine and imaged immediately with z-scan slicing of 0.43μm to survey whole cell fluorescence labeling distribution. All fluorescence channels were set with identical gain and laser power settings to provide relative levels of fluorescent intensity. No adjustments of contrast or gain were applied during post processing. Microalgal cells were also imaged on a Leica DMi8 Confocal Microscope with HyD Detector using a 40× 1.10 NA water immersion objective with 1× and 4× zoom to satisfy Nyquist frequency. Auto-fluorescence of chlorophyll (red) was excited at 552 nm, and emissions were monitored at 674–684 nm and then displayed over bright-field channel.

299 **Cryogenic Soft X-ray Tomography for Intracellular Distribution of Organelles**

300 Cell cultures were grown on-site at the Advanced Light Source (ALS) at Lawrence
301 Berkeley National Laboratory with a homemade 470nm light source with a measured intensity of
302 20μE. Cell cultures were incubated with 12-hour diurnal light at RT and harvested by
303 centrifugation at 48, 72 and 96-hour time points. Cells were gently centrifuged at 1000xG for
304 10mins to pellet. Pelleted cells had all but ~5μL of supernatant removed to remove a viscous
305 cell biomass. Wet cell biomass was inserted into 5μm micro capillaries and plunge frozen in cold
306 liquid propane. In some cases, 6μm polystyrene beads were added to cell suspensions prior to
307 centrifugation to minimize the impact of freezing on large lipid containing cells. Frozen
308 capillaries were stored in liquid nitrogen until imaging by the soft x-ray light source. Soft x-ray
309 data acquisition was carried out on beamline 2.1, a soft x-ray microscope in the National Center
310 for X-Ray Tomography (NCXT) located at the Advanced Light Source in Berkeley, California[28].
311 The microscope soft x-ray illumination was generated by a bend-magnet in the synchrotron
312 lattice and focused onto the specimen by a Fresnel Zone Plate (FZP) condenser. Specimen
313 illumination was order-sorted by a pinhole positioned just in front of the specimen. A second
314 zone plate, located downstream of the specimen, magnified and focused an image of the
315 specimen on a CCD detector. During data collection, the cells were maintained in a stream of
316 helium gas that had been cooled to liquid nitrogen temperatures. Each tomographic dataset
317 (i.e., 90 projection images spanning a range of 180°) was collected using Fresnel zone plate
318 based objective lens with a resolution of 50 nm. Exposure times for each projection image
319 ranged from 150 to 300 msec. The software suite AREC3D was used to align the projection
320 images calculate tomographic reconstructions[29].
321
322 **Lipid Extraction and Lipidomics Using Liquid Chromatography Tandem Mass**
323 **Spectrometry**

Each *O. tauri* culture was harvested at specific time points and ultimately spun down at 3000 x g for 10 min to form a pellet in chloroform compatible 2 mL Safe Seal microfuge tubes (Sorenson Bioscience, Inc, Salt Lake City, UT). The supernatant was removed, and the wet pellet was weighed for an estimate of biomass, then a mixture of 1 mL of methanol, 0.5 mL of chloroform and 0.4 mL of nanopure water was added to each pellet. The pellets were vortexed and sonicated for about 10 sec in a bath sonicator and the mixture was allowed to stand at room temperature for 10 min. Then an additional 0.5 mL of chloroform and 0.5 mL of water were added, and the mixtures were shaken vigorously into an emulsion. Each sample was centrifuged at 10,000 x g for 10 min. The bottom chloroform layer was carefully removed from the separated solvent layers using a Pasteur pipette and taking care not to take up any of the protein disc or the top polar layer and was placed in a pre-weighed glass auto sampler vial. The bottom lipid layer was dried down overnight under nitrogen and the vials with the total lipid extract (TLEs) were weighed to obtain the total mass. 5μL of chloroform was added to each dried sample and then the sample diluted to 50 μg/μL with methanol. The samples were then stored at –20 °C until ready for mass spectrometric analysis.

Extracted lipids were dried down then reconstituted in methanol. The TLEs were analyzed by LC-MS/MS using a Waters NanoAquity UPLC system interfaced with a Velos Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) as outlined in Dautel et al. 2017[30]. LC-MS/MS raw data files were imported into the in-house developed software LIQUID[31] for identification of lipid molecular species. Manual validation of the lipid identifications were determined by examining the tandem mass spectra for diagnostic ion fragments along with associated chain fragment information. In addition, the isotopic profile, extracted ion chromatogram (XIC), and mass error of measured precursor ion were examined for lipid

Cellular Protein Extraction and Digestion

To the remaining protein/debris pellet obtained from the lipid extraction, 1 mL of ice cold methanol was added 3 x to wash the majority of residual metabolites away from the pellet. A Filter-Aided Sample Prep (FASP) protein extraction and trypsin digestion was then performed using the FASP protein digestion kit (Expedeon, San Diego CA) using manufacturer's suggested protocol (Erde J et al, 2014, J Proteome Res 13(4):1885-1895). Briefly, methanol washed pellets were air dried for 2 hrs and were resuspended in solubilization buffer of 12 mM deoxycholate, 12 mM N-lauroyl sarcosine with 10 mM TCEP, and 200 mM ammonium bicarbonate, pH 8.0 at an approximate concentration of 13 μg/μL. 30 μL of this protein solution was then mixed with 200 μL of urea sample solution (kit provided), the sample was centrifuged on the kit provided spin filter at 14,000 x g for 15 min. Washes with urea and ammonium bicarbonate along with trypsin digestion and alkylation with iodoacetamide were carried out as the kit specifies. Peptides were then suspended in nanopure water and peptides were then quantified using a BCA assay (Pierce, Rockford IL) with a bovine serum albumin standard.

iTRAQ Peptide Labeling

Peptides were labeled with 8-plex iTRAQ (AB Sciex, Redwood City, CA) reagents as described below. 100 μg of each peptide sample was placed in a new tube and dried down. Chanel designations are as follows: 43 μg of dissolution buffer (iTRAQ buffer kit) was added to each sample, these being vortexed into solution and centrifuged briefly to draw sample to the bottom of each tube. The iTRAQ reagent (30 μL) was diluted further with isopropanol (115 μL) and this was then added to each sample. Each reaction was carried out at RT for 2 hrs, with 50 mM ammonium bicarbonate (200 μL) added to quench each reaction tube. After 1 hr, the contents from all iTRAQ channel reactions were added to one tube and then the sample was vortexed and dried down in a speed vac. The labeled peptides were cleaned up using C-18 SPE columns (SUPELCO Discovery) were then employed to remove the salts, using a 0.1%

TFA in nanopure water to wash the peptides and 80% acetonitrile, 0.1% TFA in water to elute the peptides.

Offline Fractionation of Peptides and Preparation of Proteome Samples

400 μg of iTRAQ labeled peptides were separated using an off-line high pH (pH 10) reversed-phase (RP) separation with a Waters XBridge C18 column (250 mm x 4.6 mm column containing 5 μm particles and a 4.6 mm x 20 mm guard column) using an Agilent 1200 HPLC System. The sample loaded onto the C18 column was washed for 15 min with Solvent A (10 mM ammonium formate, adjusted to pH 10 with ammonium hydroxide). The LC gradient started with a linear increase of Solvent B (10 mM ammonium formate, pH 10, 90% acetonitrile in water) to: 5% over 10 min, 45% Solvent B over 65 min, and then a linear increase to 100% Solvent B over 15 min. Solvent B was held at 100% for 10 min, and then was changed to 100% Solvent A, this being held for 20 min to recondition the column. The flow rate was 0.5 mL/min. A total of 96 fractions were collected into a 96 well plate throughout the LC gradient. The high pH RP fractions were then combined into 12 fractions using the concatenation strategy previously reported[32]. Peptide fractions were dried down and re-suspended in nanopure water at a concentration of 0.075 μg/ μL for mass spectrometry analysis using a Q Exactive HF Hybrid Quadrupole-Orbitrap MS (Thermo Scientific) system as described below.

Mass-Spectrometry Based Analysis of Samples

All peptide samples were analyzed using an automated home-built constant flow nano LC system (Agilent) coupled to an Q Exactive HF Hybrid Quadrupole-Orbitrap MS (Thermo Fisher Scientific). Electrospray emitters were custom made using 150 μm o.d. x 20 μm o.d. x 20 μm i.d. chemically etched fused silica. An on-line 4-cm x 360 μm o.d. x 150 μm i.d. fused-silica capillary analytical column (3 μm Jupiter C18) was used. Mobile phases consisted of

399 0.1% formic acid in water (A) and 0.1% formic acid acetonitrile (B) operated at 300 nL/min with
400 a gradient profile as follows (min:%B); 0:5, 2:8, 20:12, 75:35, 97:60, 100:85.
401
402 **Peptide Identification, Quantification and Analysis**
403 For peptide identification, MS/MS spectra were searched against a decoy O. tauri
404 database using the algorithm SEQUEST. An approach to correlate tandem mass spectral data
405 of peptides with amino acid sequences in a protein database[33]. Search parameters included: no
406 enzyme specificity for proteome data and trypsin enzyme specificity with a maximum of two
407 missed cleaves, ± 50 ppm precursor mass tolerance, ± 0.05 Da product mass tolerance, and
408 carbamidomethylation of cysteines and iTRAQ labeling of lysines and peptide N-termini as fixed
409 modifications. Allowed variable modifications were oxidation of methionine. MSGF+ spectra
410 probability values were also calculated for peptides identified from SEQUEST searches[34].
411 Measured mass accuracy and MSGF spectra probability were used to filter identified peptides to
412 <0.4% false discovery rate (FDR) at spectrum level and <1% FDR at the peptide level using the
413 decoy approach. iTRAQ reporter ions were extracted using the MASIC software[35] for fast
414 quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS)
415 features with a 10-ppm mass tolerance for each expected iTRAQ reporter ion as determined
416 from each MS/MS spectrum.
417 Relative abundances of peptides were determined using iTRAQ reporter ion intensity
418 ratios from each MS/MS spectrum. Individual peptide intensity values were determined by
419 dividing the base peak intensity by the relative ratio associated with each reporter ion. All
420 peptide values were then transformed into log2 values for comparison between conditions.

Acknowledgements:

This research was performed using the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at PNNL. This research also used resources of the Advanced Light Source, which is a DOE Office of Science User Facility under contract no. DE-AC02-05CH11231.

Funding

This work was supported by DOE-BER Mesoscale to Molecules Bioimaging Project FWP# 66382.

Competing Interests: We declare no competing interests.

Contributions

JEE devised and managed all experiments with input from CRS. CRS conducted confocal and Raman imaging experiments and coordinated integrated omics analysis. KKH and CDN performed lipid and proteomics sample prep. RM collected proteomics data that SP analyzed, and NK and WRC worked into the cell model. JEK performed lipidomic data collection and lipidomic analysis. Cryogenic soft X-ray sample preparation was conducted by JHC, RB and CRS with JHC and GM acquiring tilt series for subsequent reconstruction by AE. FACS analysis was conducted by WC. JEE and CRS wrote initial manuscript and all authors edited and approved final text.

References

1 Gimpel, J. A., Henriquez, V. & Mayfield, S. P. In Metabolic Engineering of Eukaryotic Microalgae: Potential and Challenges Come with Great Diversity. *Front Microbiol* **6**, 1376, doi:10.3389/fmicb.2015.01376 (2015).

2 Wase, N., Tu, B., Allen, J. W., Black, P. N. & DiRusso, C. C. Identification and Metabolite Profiling of Chemical Activators of Lipid Accumulation in Green Algae. *Plant Physiol* **174**, 2146-2165, doi:10.1104/pp.17.00433 (2017).

3 Longworth, J., Wu, D., Huete-Ortega, M., Wright, P. C. & Vaidyanathan, S. Proteome response of Phaeodactylum tricornutum, during lipid accumulation induced by nitrogen depletion. *Algal Res* **18**, 213-224, doi:10.1016/j.algal.2016.06.015 (2016).

4 Levering, J., Broddrick, J. & Zengler, K. Engineering of oleaginous organisms for lipid production. *Curr Opin Biotechnol* **36**, 32-39, doi:10.1016/j.copbio.2015.08.001 (2015).

5 Liao, J. C., Mi, L., Pontrelli, S. & Luo, S. Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nat Rev Microbiol* **14**, 288-304, doi:10.1038/nrmicro.2016.32 (2016).

6 Goncalves, E. C., Wilkie, A. C., Kirst, M. & Rathinasabapathi, B. Metabolic regulation of triacylglycerol accumulation in the green algae: identification of potential targets for engineering to improve oil yield. *Plant Biotechnol J* **14**, 1649-1660, doi:10.1111/pbi.12523 (2016).

7 Klok, A. J., Martens, D. E., Wijffels, R. H. & Lamers, P. P. Simultaneous growth and neutral lipid accumulation in microalgae. *Bioresour Technol* **134**, 233-243, doi:10.1016/j.biortech.2013.02.006 (2013).

8 Griffiths, M. J., van Hille, R. P. & Harrison, S. T. L. Lipid productivity, settling potential and fatty acid profile of 11 microalgal species grown under nitrogen replete and limited conditions. *J Appl Phycol* **24**, 989-1001 (2012).

9 Guarnieri, M. T., Nag, A., Yang, S. & Pienkos, P. T. Proteomic analysis of Chlorella vulgaris: potential targets for enhanced lipid accumulation. *J Proteomics* **93**, 245-253, doi:10.1016/j.jprot.2013.05.025 (2013).

10 Minhas, A. K., Hodgson, P., Barrow, C. J. & Adholeya, A. A Review on the Assessment of Stress Conditions for Simultaneous Production of Microalgal Lipids and Carotenoids. *Front Microbiol* **7**, 546, doi:10.3389/fmicb.2016.00546 (2016).

11 Lohman, E. J. *et al.* Optimized inorganic carbon regime for enhanced growth and lipid accumulation in Chlorella vulgaris. *Biotechnol Biofuels* **8**, 82, doi:10.1186/s13068-015-0265-4 (2015).

12 Levering, J., Dupont, C. L., Allen, A. E., Palsson, B. O. & Zengler, K. Integrated Regulatory and Metabolic Networks of the Marine Diatom Phaeodactylum tricornutum Predict the Response to Rising CO2 Levels. *mSystems* **2**, doi:10.1128/mSystems.00142-16 (2017).

13 Chen, Y., Xu, C. & Vaidyanathan, S. Microalgae: a robust "green bio-bridge" between energy and environment. *Crit Rev Biotechnol*, 1-18, doi:10.1080/07388551.2017.1355774 (2017).

14 Leliaert, F., Verbruggen, H. & Zechman, F. W. Into the deep: new discoveries at the base of the green plant phylogeny. *Bioessays* **33**, 683-692, doi:10.1002/bies.201100035 (2011).

15 Courties, C. *et al.* Phylogenetic analysis and genome size of Ostreococcus tauri (Chlorophyta, Prasinophyceae). *J Phycol* **34**, 844-849, doi:DOI 10.1046/j.1529-8817.1998.340844.x (1998).

16 Guillou, L. *et al.* Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist* **155**, 193-214, doi:10.1078/143446104774199592 (2004).

17 Derelle, E. *et al.* Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. *Proc Natl Acad Sci U S A* **103**, 11647-11652, doi:10.1073/pnas.0604795103 (2006).

18 Blanc-Mathieu, R. *et al.* Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Sci Adv* **3**, e1700239, doi:10.1126/sciadv.1700239 (2017).

19 Chuck R Smallwood, W. C., Jian-Hua Chen, Emma Patello, Mathew Thomas, Rosanne Boudreau, Axel Ekman, Hongfei Wang, Gerry McDermott, James E Evans. *Ostreococcus tauri* is a high-lipid content green algae that extrudes clustered lipid droplets. *BioRXiv* **249052**, doi:https://doi.org/10.1101/249052 (**2018**).

20 Le Bihan, T. *et al.* Label-free quantitative analysis of the casein kinase 2-responsive phosphoproteome of the marine minimal model species Ostreococcus tauri. *Proteomics* **15**, 4135-4144, doi:10.1002/pmic.201500086 (2015).

21 McCoy, J. G. *et al.* Discovery of sarcosine dimethylglycine methyltransferase from Galdieria sulphuraria. *Proteins* **74**, 368-377, doi:10.1002/prot.22147 (2009).

22 Luo, G. Z., Blanco, M. A., Greer, E. L., He, C. & Shi, Y. DNA N(6)-methyladenine: a new epigenetic mark in eukaryotes? *Nat Rev Mol Cell Biol* **16**, 705-710, doi:10.1038/nrm4076 (2015).

23 Degraeve-Guilbault, C. *et al.* Glycerolipid Characterization and Nutrient Deprivation-Associated Changes in the Green Picoalga Ostreococcus tauri. *Plant Physiol* **173**, 2060-2080, doi:10.1104/pp.16.01467 (2017).

24 Shimojima, M. & Ohta, H. Critical regulation of galactolipid synthesis controls membrane differentiation and remodeling in distinct plant organs and following environmental changes. *Prog Lipid Res* **50**, 258-266, doi:10.1016/j.plipres.2011.03.001 (2011).

25 Schaller-Laudel, S. *et al.* Influence of thylakoid membrane lipids on the structure of aggregated light-harvesting complexes of the diatom Thalassiosira pseudonana and the green alga Mantoniella squamata. *Physiol Plant* **160**, 339-358, doi:10.1111/ppl.12565 (2017).

26 Losh, J. L., Young, J. N. & Morel, F. M. Rubisco is a small fraction of total protein in marine phytoplankton. *The New phytologist* **198**, 52-58, doi:10.1111/nph.12143 (2013).

27 Keller, M. D., Selvin, R. C., Claus, W. & Guillard, R. R. L. Media for the Culture of Oceanic Ultraphytoplankton. *J Phycol* **23**, 633-638 (1987).

28 Le Gros, M. A. *et al.* Biological soft X-ray tomography on beamline 2.1 at the Advanced Light Source. *J Synchrotron Radiat* **21**, 1370-1377, doi:10.1107/S1600577514015033 (2014).

29 Parkinson, D. Y., Knoechel, C., Yang, C., Larabell, C. A. & Le Gros, M. A. Automatic alignment and reconstruction of images for soft X-ray tomography. *J Struct Biol* **177**, 259-266, doi:10.1016/j.jsb.2011.11.027 (2012).

30 Dautel, S. E. *et al.* Lipidomics reveals dramatic lipid compositional changes in the maturing postnatal lung. *Sci Rep* **7**, 40555, doi:10.1038/srep40555 (2017).

31 Kyle, J. E. *et al.* LIQUID: an-open source software for identifying lipids in LC-MS/MS-based lipidomics data. *Bioinformatics* **33**, 1744-1746, doi:10.1093/bioinformatics/btx046 (2017).

32 Wang, Y. *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019-2026, doi:10.1002/pmic.201000722 (2011).

33 Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989, doi:10.1016/1044-0305(94)80016-2 (1994).

34 Kim, S., Gupta, N. & Pevzner, P. A. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J Proteome Res* **7**, 3354-3363, doi:10.1021/pr8001244 (2008).

35 Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput Biol Chem* **32**, 215-217, doi:10.1016/j.compbiolchem.2008.02.006 (2008).

Main Figures

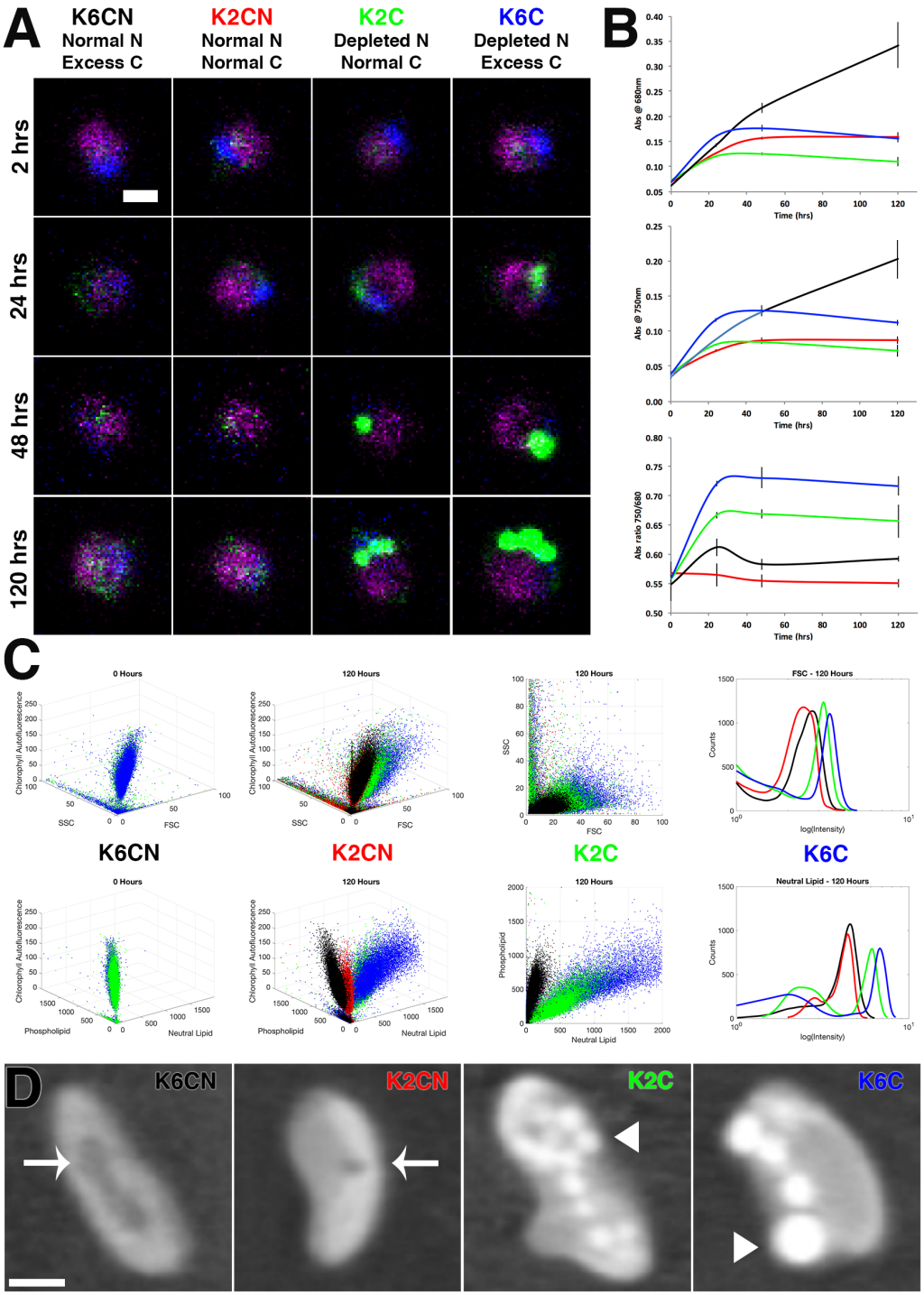


Figure 1 Single cell and population imaging and growth under varying C:N ratios

(A) Confocal microscopy of K6CN, K2CN, K2C, and K6C cultures to observe chlorophyll autofluorescence (magenta), nucleic acid fluorescent staining (blue), neutral lipid fluorescent staining (green) over time. Scale bar represents 1µm scale. (B) Corresponding growth plots for the same cultures monitored over time (error bars represent a replicate of 5) at 750nm and 680nm, and the ratio of the 750/680 as a measure of photosynthetic efficiency. (C) FACS time course study of same cultures comparing SSC (side-scattering) versus FSC (forward scattering) and chlorophyll autofluorescence in 3D plots at 0 hours and 120 hours, then 2D plot of SSC versus FSC at 120 hours, and histogram of FSC at 120 hours. (D) FACS time course study of same cultures comparing phospholipid versus neutral lipid and chlorophyll autofluorescence in 3D plots at 0 hours and 120 hours, then 2D plot of phospholipid versus neutral lipid at 120 hours, and histogram of neutral lipid at 120 hours. (E) Label-free cryogenic soft x-ray nanotomography images of cells cryogenically frozen in microcapillaries. Central slices (1µm scale bar)

display intracellular structures common to each culture condition. Chloroplasts and cytoplasm seen in all images. Vacuoles in K6CN and K2CN appear as dark objects indicated with white arrow. Lipid droplets appear as bright white circular objects indicated by white arrowhead.

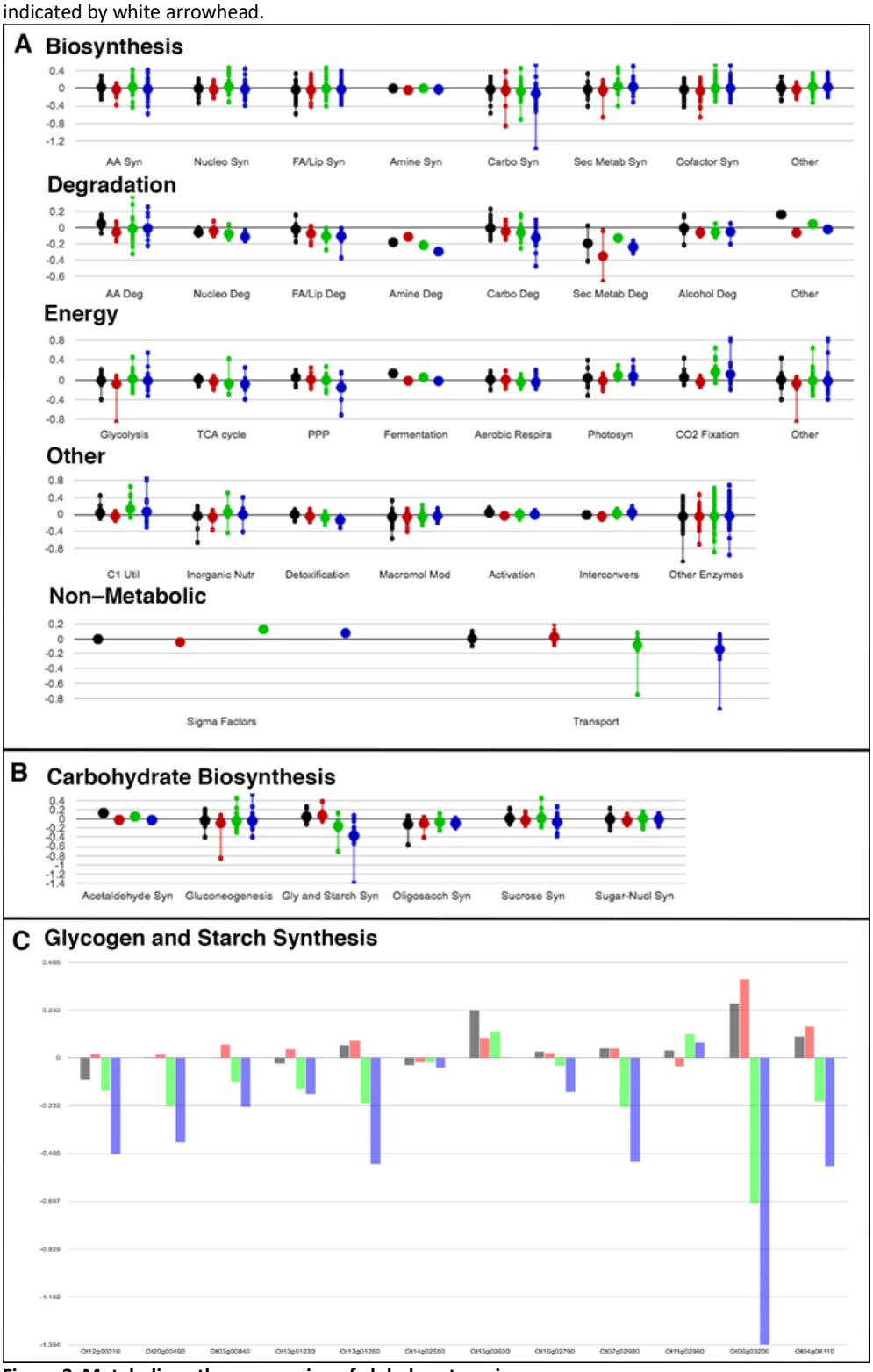


Figure 2 Metabolic pathway mapping of global proteomics

O. tauri cells cultured in K6CN (black), K2CN (red), K2C (green), K6C (blue) at 48 hours relative to cultures from K2CN at 24 hours were grouped and mapped by metabolic pathway(A), subpathway (B), and individual proteins (C). Each large circle represents

an average level of abundance on a log2 scale where negative values are increases in abundance relative to cell conditions K2CN at 24 hours. Smaller circles represent individual proteins in the respective metabolic pathway.

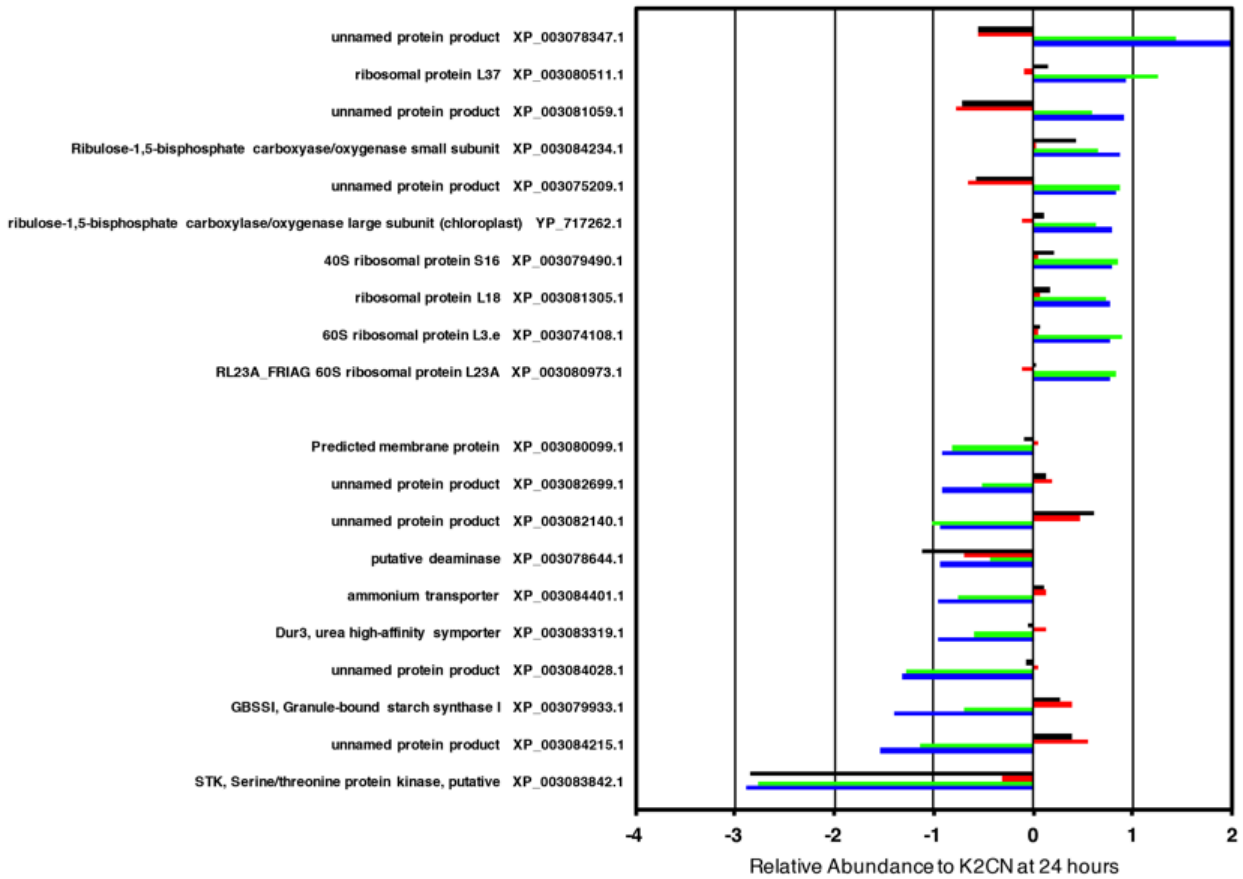


Figure 3 Selection differentially expressed proteins from global proteomics of varying N and C conditions
Abundance profiles for 10 most downregulated (positive values) and upregulated (negative values) proteins comparing K2CN at 24 hours versus K6CN at 48 hours (black), K2CN at 48 hours (red), K2C at 48 hours (green), and K6C at 48 hours (blue) cell cultures.

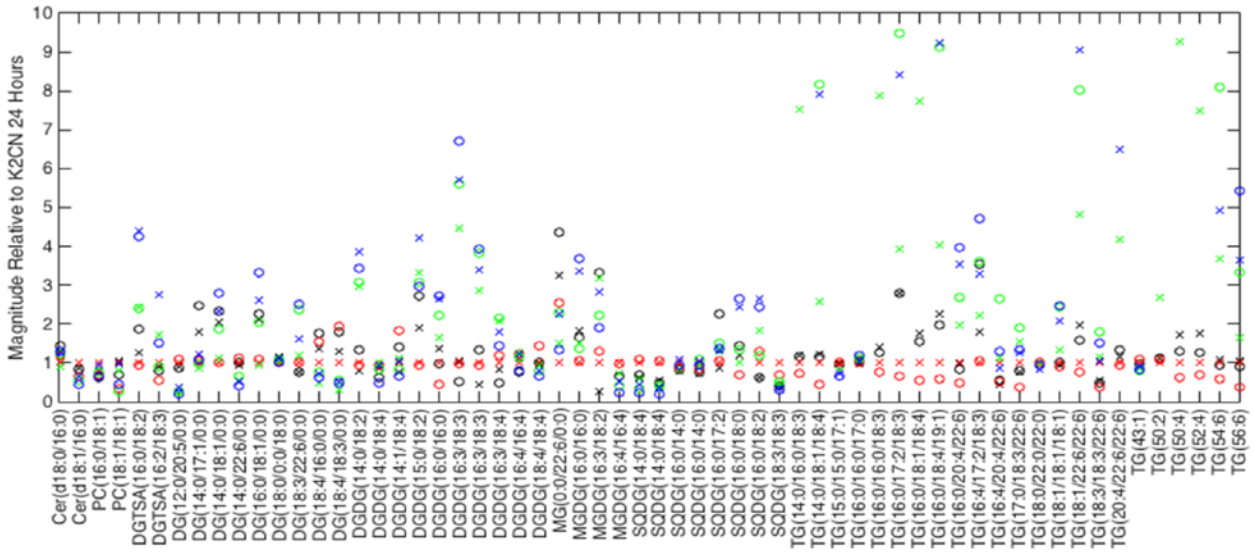
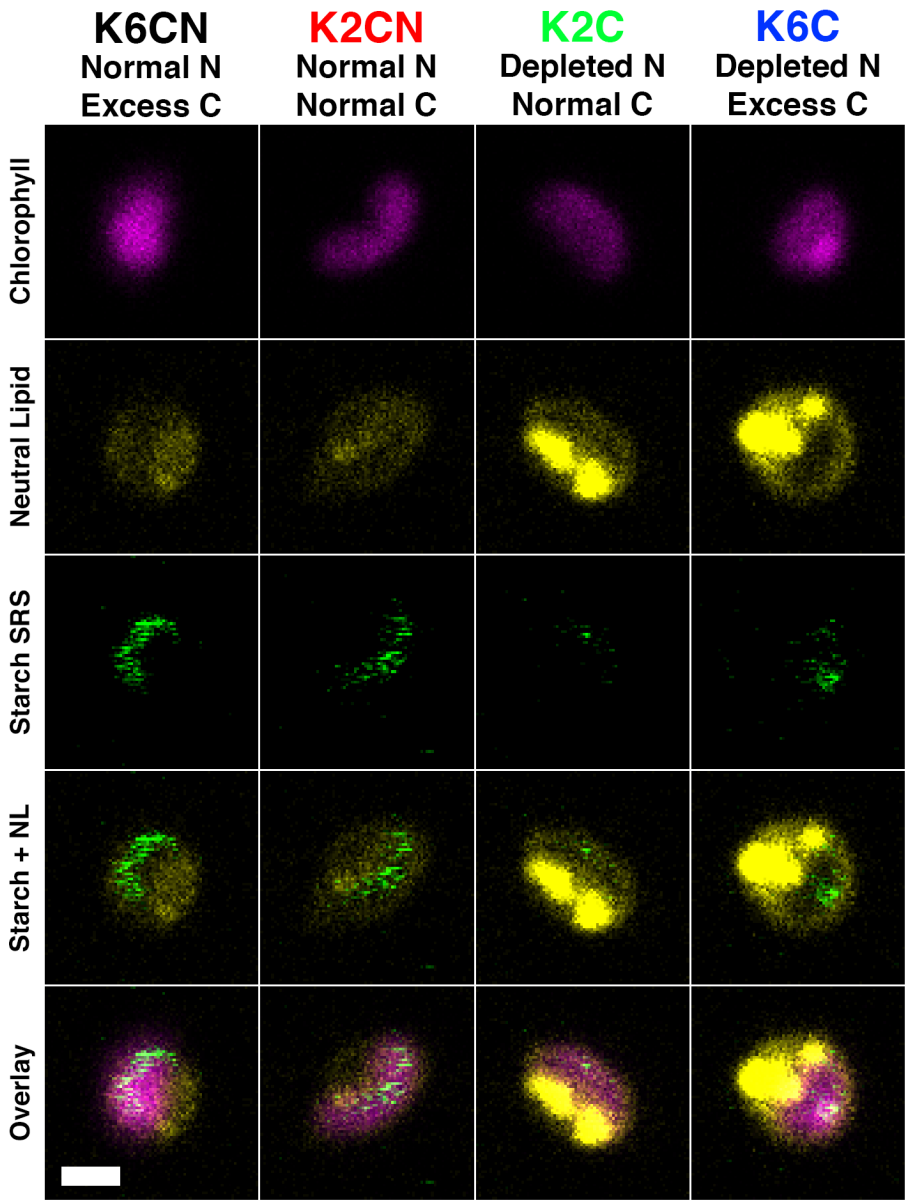


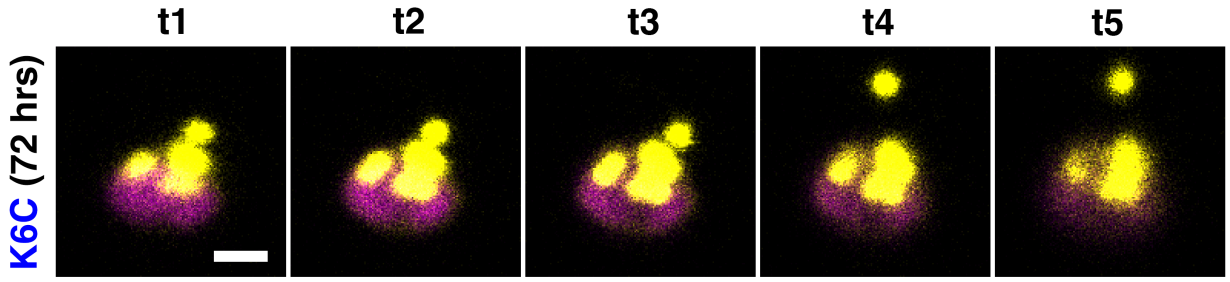
Figure 4 Time resolved differential expression of global lipids

Selection of lipids (out of 287) detected via LC-MS/MS for cell cultures at 24 (X symbols) and 48 hours (O symbols) for varying C:N ratio conditions K6CN (black), K2CN (red), K2C (green) and K6C (blue). *Cer: Ceramide; PC: phosphatidylcholine; MGDG: monogalactosyldiacylglycerol; DG: diacylglycerol; DGDG: digalactosyldiacylglycerol; DGTSA: diacylglyceryl trimethylhomoserine or diacylglyceryl trimethyl-beta-alanine; SQDG: sulfoquinovosyldiacylglycerol; and TG: triacylglycerols*

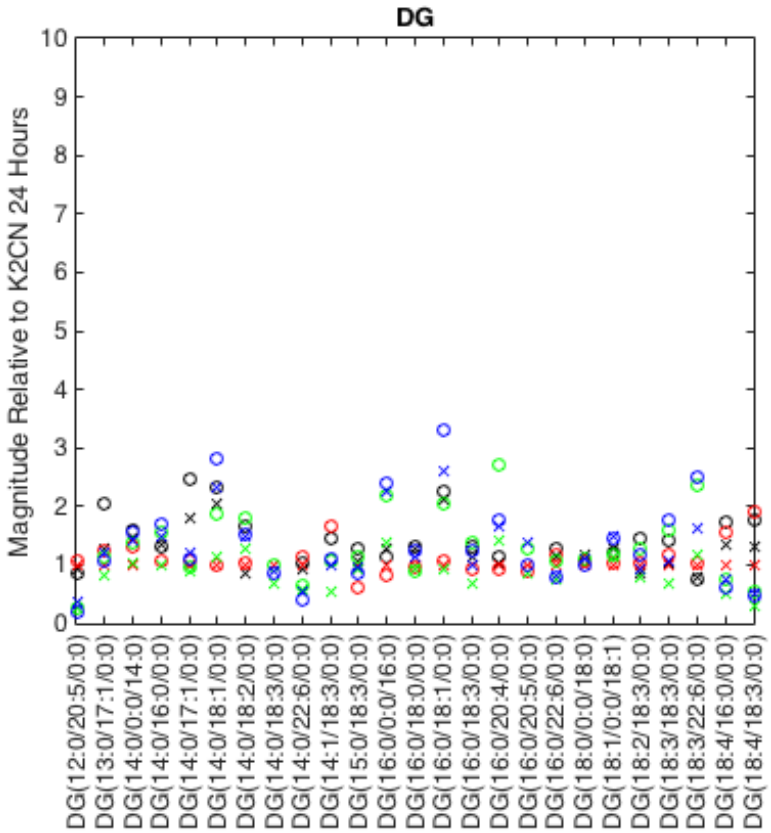
Supplemental Figures



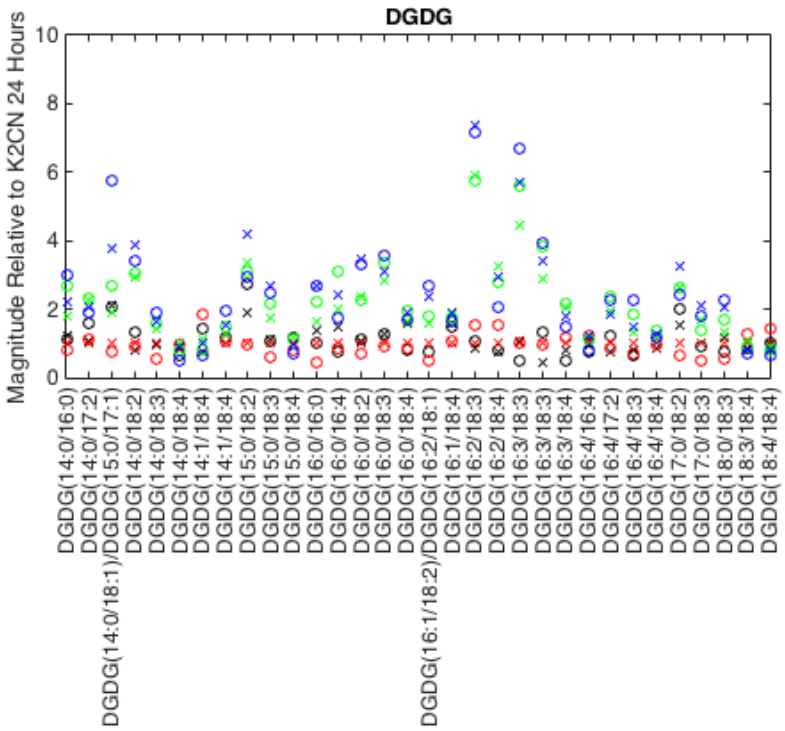
Supplemental Figure 1 Fluorescence and SRS verification of neutral lipid and starch accumulation
Fluorescence and SRS microscopy were conducted on *O. tauri* cells stained with Nile red from various C:N conditions after 96 hours of starvation. Chlorophyll autofluorescence (pink) was observed in all cells with neutral lipid (yellow) and SRS starch signal (green). Elevated neutral lipid was observed in K2C and K6C cells localized outside the chloroplast. Starch was detected and localized to the chloroplast for K6CN, K2CN, and to a lesser extent K6C, with little to no starch signal found in K2C cells. Scale bar represents 1µm.



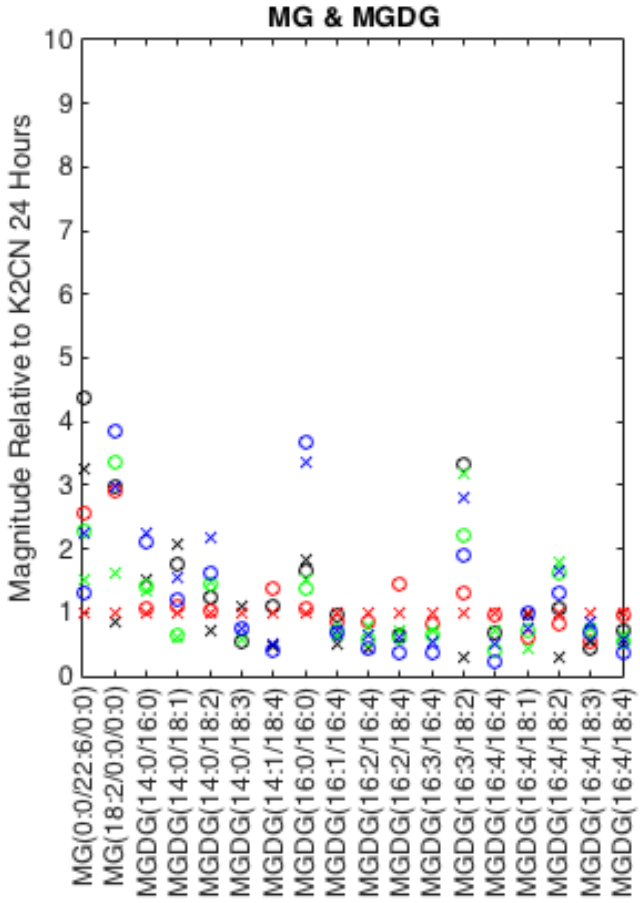
Supplemental Figure 2 Time course of lipid release from cells into intercellular space
O. tauri cells cultured in K6C media for 72 hours were harvested and stained with Nile red and then imaged by confocal fluorescence microscopy. Chlorophyll autofluorescence (pink) and neutral lipid (yellow) revealed many large lipid bodies during z-slice imaging and a single lipid body release was tracked over time (t) in intercellular milieu while other lipids remained inside the intracellular space.



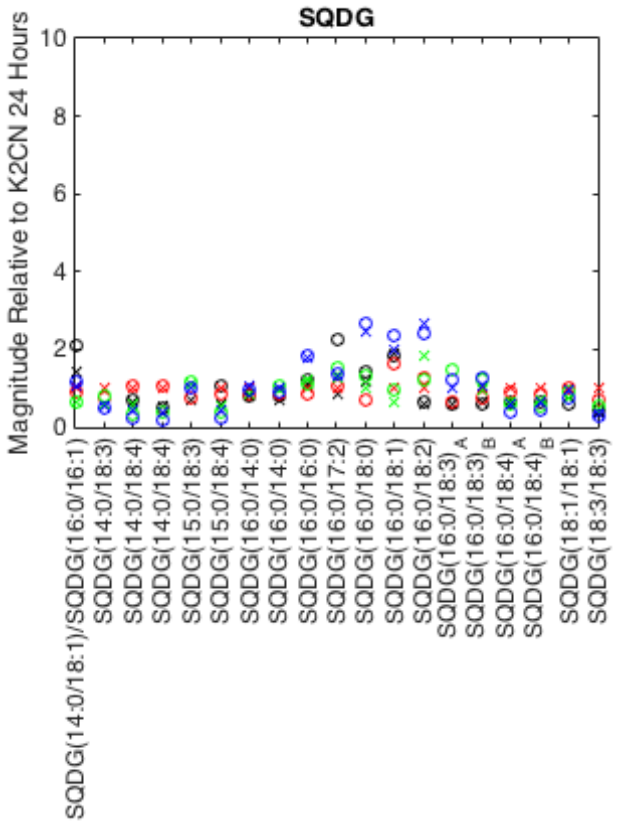
Supplemental Figure 3 Time-resolved lipidomics of diacylglycerol lipids
Diacylglycerol (DG) lipids detected via LC-MS/MS for cell cultures at 24 (X symbols) and 48 hours (O symbols) for varying C:N ratio conditions K6CN (black), K2CN (red), K2C (green) and K6C (blue) relative to K2CN at 24 hours (y-axis).



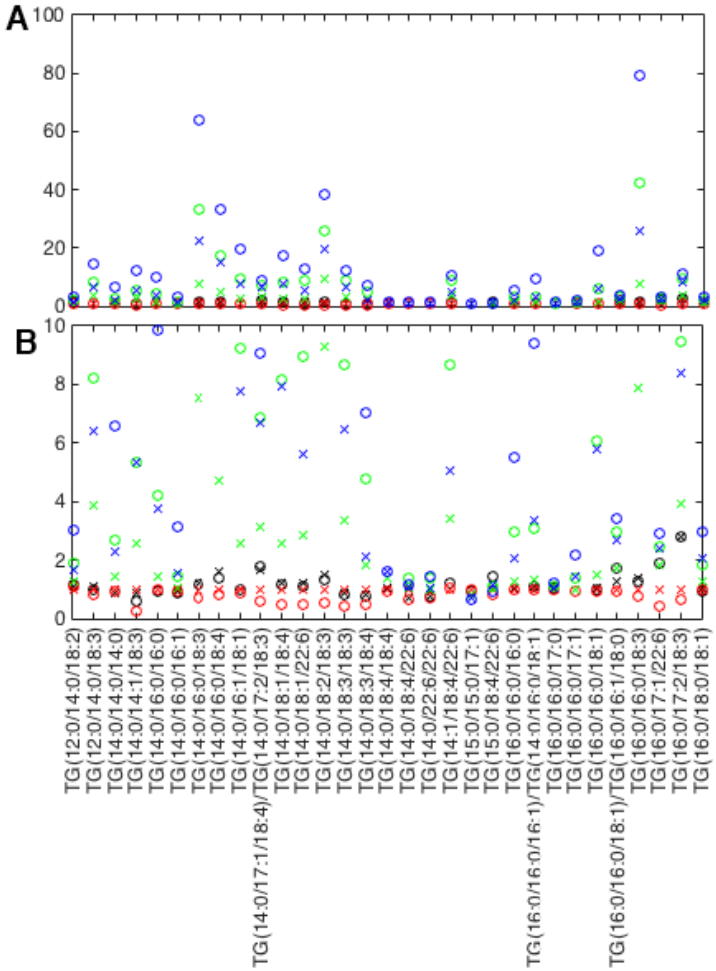
Supplemental Figure 4 Time resolved differential expression of digalactosyldiacylglycerol lipids
Digalactosyldiacylglycerol (DGDG) lipids detected for cell cultures at 24 (X symbols) and 48 hours (O symbols) for varying C:N ratio conditions K6CN (black), K2CN (red), K2C (green) and K6C (blue) relative to K2CN at 24 hours (y-axis).



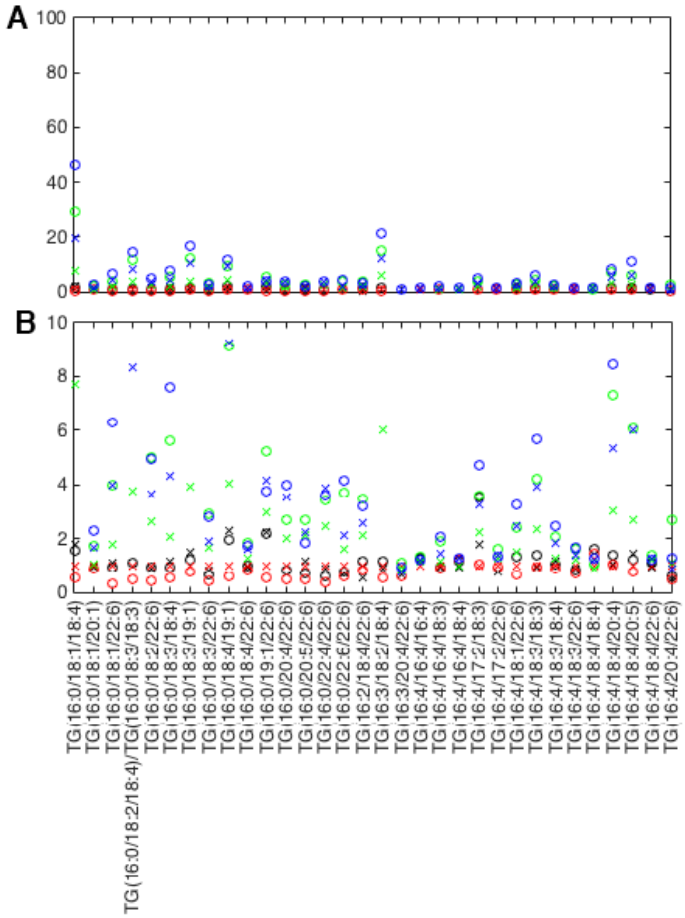
Supplemental Figure 5 Time resolved differential expression of monoacylglycerol and monogalactosyldiacylglycerol lipids
Monoacylglycerol (MG) and monogalactosyldiacylglycerol (MGDG) lipids detected via LC-MS/MS for cell cultures at 24 (X symbols) and 48 hours (O symbols) for varying C:N ratio conditions K6CN (black), K2CN (red), K2C (green) and K6C (blue) relative to K2CN at 24 hours (y-axis).



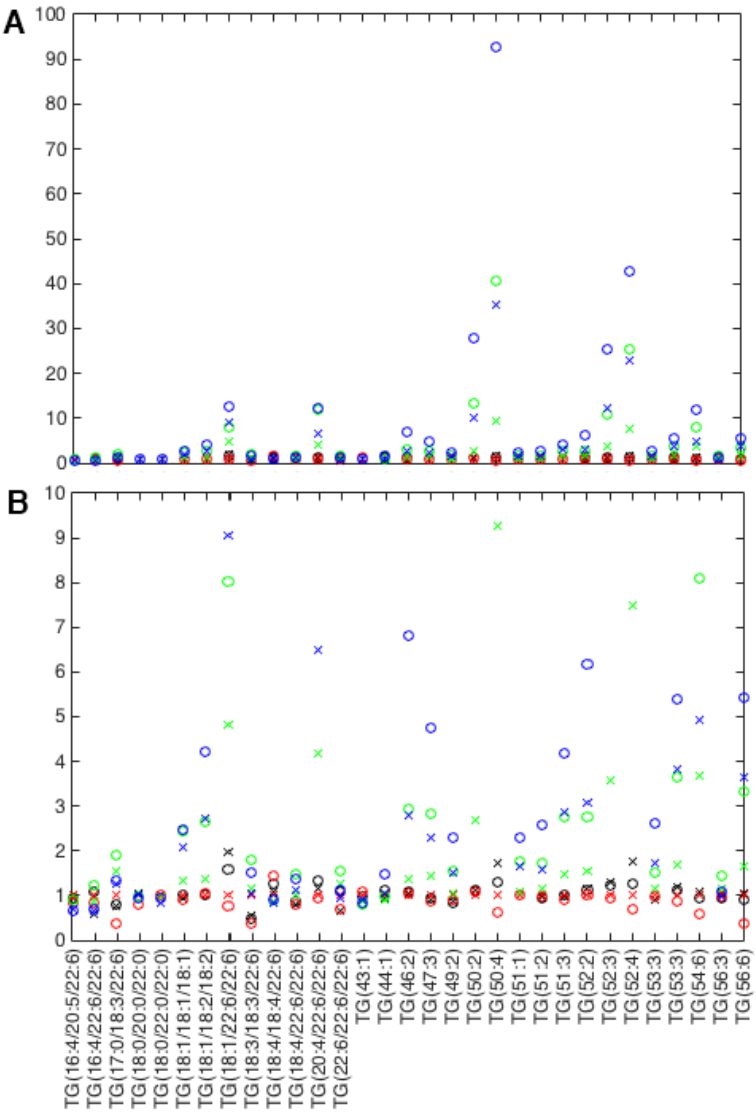
Supplemental Figure 6 Time resolved differential expression of lipids
Sulfoquinovosyldiacylglycerol (SQDG) lipids detected via LC-MS/MS for cell cultures at 24 (X symbols) and 48 hours (O symbols) for varying C:N ratio conditions K6CN (black), K2CN (red), K2C (green) and K6C (blue) relative to K2CN at 24 hours (y-axis).



Supplemental Figure 7 Time resolved differential abundance of triacylglycerol lipids set 1
Triacylglycerol (TG) lipidomics for cell cultures at 24 (X symbols) and 48 hours (O symbols) for varying C:N ratio conditions K6CN (black), K2CN (red), K2C (green) and K6C (blue) relative to K2CN at 24 hours (y-axis).



Supplemental Figure 8 Time resolved differential abundance of triacylglycerol lipids, set 2
Triacylglycerol (TG) lipidomics detected via LC-MS/MS for cell cultures at 24 (X symbols) and 48 hours (O symbols) for varying C:N ratio conditions K6CN (black), K2CN (red), K2C (green) and K6C (blue) relative to K2CN at 24 hours (y-axis).



Supplemental Figure 9 Time resolved differential abundance of long chain triacylglycerol lipids, scale 10
Long chain triacylglycerol (TG) lipidomics for cell cultures at 24 (X symbols) and 48 hours (O symbols) for varying C:N ratio conditions K6CN (black), K2CN (red), K2C (green) and K6C (blue) relative to K2CN at 24 hours (y-axis).