PREFERENTIAL SYNTHESIS OF VERY LONG CHAIN POLYUNSATURATED FATTY ACIDS IN *EUTREPTIELLA* SP. (EUGELNOZOA) REVEALED BY CHROMATOGRAPHIC AND TRANSCRIPTOMIC ANALYSES[1]

Rita C. Kuo[2]

Department of Marine Sciences, University of Connecticut, Groton, CT 06340, USA

Huan Zhang

Department of Marine Sciences, University of Connecticut, Groton, CT 06340, USA

James D. Stuart

Department of Chemistry, University of Connecticut, Storrs, CT 06269, USA

Anthony A. Provatas

Center of Environmental Sciences and Engineering, University of Connecticut, Storrs, CT 06269, USA

Linda Hannick

SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Rockville, MD 20852, USA

Senjie Lin[3]

Department of Marine Sciences, University of Connecticut, Groton, CT 06340, USA

[1]Received May 15, 2020

[2]Current address: Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA 94598, USA. The work is part of her dissertation.

[3]Address for correspondence: Phone, 860-405-9168; Email, senjie.lin@uconn.edu

Running Title: Fatty acid and lipid synthesis in alga *Eutreptiella*

Abstract

Algal lipids are a fundamental energy source in the ecosystem and material for biofuel production, but their production and regulation remains unexplored in many species including the common coastal phytoplankton *Eutreptiella* spp. Using gas chromatography-tandem mass spectrometry (GC/MS/MS), we discovered 24 types of fatty acids (FAs) in *Eutreptiella* sp. with unsaturated C24 FAs and a relatively high proportion of long chain unsaturated FAs. The abundances of C16, C18 and saturated FAs decreased when phosphate in the culture medium was depleted. Among the 24 FAs, docosahexaenoic acid (22:6) and eicosapentaenoic acid (20:5) were the most abundant, suggesting that *Eutreptiella* sp. preferentially invests in the synthesis of very long chain polyunsaturated fatty acids (VLCPFA), making this alga nutritionally valuable to grazers. Further transcriptomic analysis revealed that *Eutreptiella* sp. likely synthesizes VLCPFA via Δ8 pathway and uses type I and II fatty acid synthases. Using RT-qPCR, we found that some of the lipid production genes, such as β –ketoacyl-ACP reductase, fatty acid desaturase, acetyl-CoA carboxylase, acyl carrier protein, Δ8 desaturase, and Acyl-ACP thioesterase, were more actively expressed during light period. In addition, were more highly expressed in the high lipid-producing light period, suggesting a linkage between photosynthesis and lipid production.

Keywords: *Eutreptiella,* fatty acid biosynthesis, lipid biosynthesis, unsaturated fatty acids

Introduction

Eutreptiella is a genus of the euglenoid algae, which have excellent nutritive value (*i.e.* high vitamin and lipid content, Takeyama et al., 1996; Yamane et al., 2001) and are rather abundant in some marine ecosystems (Henriksen et al. 2002). Some *Eutreptiella* spp. can seasonally be a dominant group of phytoplankton (Álvarez-Góngora and Herrera-Silveira 2006, Bates and Strain 2006, Olli et al. 1996, Rodríguez-Graña et al. 2008, Seong et al. 2006), forming blooms in nutrient-rich coastal or brackish waters (Anderson et al. 2000, Lindholm 1993, Olli et al. 1996, Stonik & Selina 200, Stonik 2007). In some areas, *Eutreptiella braarudii* alone can make up to 46% of the phytoplankton population (Stonik, 2007). As primary producers, these algae are important in energy flow and nutrient cycling in the coastal marine ecosystem. Meanwhile, as solar energy converters, these algae are potentially candidates of biofuel species.

Lipid content is one important indicator of quality of an alga as prey and as source of biofuel. In general, lipids are essential for structural constituents of cellular membranes in all organisms and protection of delicate internal organs and hormones in animals (Singh 2002). In the marine ecosystem, fatty acids are a fundamental energy source for growth and cell membrane fluidity (especially in cold water) in animals such as zooplankton and fish (*e.g.* Arzel et al. 1994, Cossins et al. 1977, Klein Breteler 2004, Rainuzzo et al. 1997, Tang and Taal 2005), which consume phytoplankton as food. To gain better understanding on the energy flow in marine food chains, it is important to study fatty acid and lipid biosynthesis pathways in dominant phytoplankton (prey) species. As the type and abundance of lipids in the cells also determine the potential of algae as source of biofuel, understanding the regulatory pathways of lipid production in algae is also important for evaluating or genetically enhancing algal species for biofuel production.

71 In land plants, photosynthesis produces small precursors for lipid production in
72 chloroplasts. The small precursors are later converted into long chain fatty acids via two enzyme
73 systems, acetyl-CoA carboxylase (ACCase) and fatty acid synthase (FAS). ACCase is essential
74 for lipid metabolism as it catalyzes the first step of fatty acid synthesis and is found in both
75 bacteria and eukaryotes (Cronan and Waldrop 2002, Harwood 1988). FAS consists of 6 enzyme
76 activities and an acyl-carrier protein (CA) to catalyze a 2 carbon elongation process (Bloch and
77 Vance 1977, Nelson and Cox 2008). It produces palmitate and stearate and these fatty acids are
78 subject to elongation, desaturation or further modification by elongase or desaturase in
79 chloroplasts (Gunstone et al. 2007). There are two major types of fatty acid synthases. Fatty acid
80 synthase I (FASI), found in vertebrates and fungi, consists of a multi-enzyme complex contained
81 in a single polypeptide chain (Nelson and Cox 2008, Schweizer and Hofmann 2004). In contrast,
82 fatty acid synthase II (FASII), found in plants and bacteria, contains discrete enzymes (Nelson
83 and Cox 2008, White et al. 2005). Among euglenoid algae, *Euglena gracilis* composes both
84 FASI and plastid FASII for *de novo* fatty acid synthesis (Delo et al. 1971, Worsham et al. 1993).
85 In addition to FASII, Goldberg and Bloch (1972) reported another ACP-dependent fatty acid
86 synthase (FASIII) that elongates acyl-CoA derivatives from C10 to C18 to longer chain ACP
87 thioesters in plastids.
88 Similar to land plants, microalgae initiate fatty acid synthesis in chloroplasts (Guschina
89 and Hardwood, 2006) and lipid biosynthesis is probably linked directly to photosynthesis.
90 Studies have shown that the activities of ACCase are related to the light reactions of
91 photosynthesis, as NADPH and ATP produced in the photochemical reaction are required for the
92 synthesis of palmitic acid (Sasaki et al. 1997). Furthermore, fatty acid synthesis in photosynthetic
93 organisms obviously relies on carbon fixation for carbon precursors for fatty acid synthesis (Bao

94 et al. 2000). The synthesized fatty acids are in turn used for triacylglycerol (TAG) synthesis
95 (Thelen and Ohlrogge 2002). In addition, dihydroxyacetone phosphate, the precursor for glycerol
96 3-phosphate needed in the synthesis of lipid TAG, can be produced from Calvin-Benson Cycle
97 (Nelson and Cox 2008).

98 It is well understood that lipid production in microalgae is influenced by nutrient and
99 other environmental conditions (e.g. Gouveia and Oliveira 2009, Griffiths and Harrison 2009, Hu
100 et al. 2008, Illman et al. 2000, Liu et al. 2008). Nutrient limitation can enhance the production of
101 lipid in algal cells. For example, the lipid content of some *Chlorella* species (*e.g. C. emersonil*
102 *and C. pyrenoldosa*) can increase to more than 60% of the dry weight under nitrogen-deprived
103 conditions (Griffiths and Harrison 2009, Illman et al. 2000). However, our knowledge of the
104 genetic regulatory mechanisms of lipid biosynthesis in marine microalgae is limited and
105 fragmentary (Guchina and Harwood 2006, Hu et al., 2008). Current understanding of lipid
106 biosynthesis pathways is mainly based on land plants.

107 In this study, we used 454 high-throughput sequencing to study transcriptomic profiles
108 and gas chromatography-tandem mass spectrometry (GC/MS/MS) to identify fatty acids
109 produced under phosphate-depleted and phosphate-replete conditions, and conducted reverse-
110 transcription qPCR (RT-PCR) to quantify expression of several key enzyme coding genes to
111 investigate molecular mechanisms that regulate lipid biosynthesis in *Eutreptiella* sp.
112

113 **Materials and Methods**

114 ***Culture preparation and RNA isolation.*** Two sets of cultures were maintained in phosphate-
115 depleted medium (to produce high-lipid culture) and phosphate-replete f/2 medium (to produce

116 low-lipid culture); cells were harvested, total RNA and mRNA were isolated as reported
117 previously (Kuo et al. 2013)

118 ***Lipid measurement.*** Nile Red (9-diethylamino-5H-benzo-α-phenoxazine-5-one, a lipid-soluble
119 fluorescent probe) staining was used to measure relative abundance of neutral lipids as reported
120 (Lee et al., 1998, Hu *et al*., 2008). Forty µl of Nile Red solution in acetone (250 mg/l) were
121 added to 10 ml of algal suspension at room temperature for 10 min. A spectrophotometer
122 (HITACHI, Tokyo, Japan) was then used to measure the fluorescence with excitation at the
123 wavelength of 490 nm and emission in the wavelength band of 580 to 590 nm, as reported in
124 Kuo and Lin (2012).

125 ***Lipid extraction and fatty acid identification by gas chromatography-tandem mass***
126 ***spectrometry (GC/MS/MS).*** A gravimetric method was applied to determine actual lipid contents
127 by using chloroform-methanol method according to Bligh and Dyer (1959). In short, lipids were
128 extracted from freeze-dried cells with chloroform/methanol (1:2, v/v) and the residue was
129 extracted one more time with 1:1 chloroform/methanol. The chloroform layer was collected and
130 evaporated under a gentle flow of nitrogen gas. Derivatization for GC/MS/MS analysis followed
131 the method summarized by (Carvalho and Malcata 2005). Briefly, samples were dissolved in 1
132 ml of a freshly prepared mixture of dry acetyl chloride and methanol, at a ratio of 5:100 (v/v),
133 and kept at 100 °C under pure nitrogen for 1 h. After cooling, 1 ml of hexane wad added and
134 mixed by vortexing. Purification of the solution was achieved by adding 1 ml of saturated
135 sodium chloride solution. The prepared solution of fatty acid methyl esters was filtered using
136 0.45 µm Millipore filter. Analysis of the solution was performed on a Waters Quattro Micro
137 GC/MS/MS system equipped with a Rxi-5Sil MS column (30 m x 0.25 mm, 0.25 µm film

138 thickness, Restek Co., PA). Helium was used as the carrier gas with a flow rate of 1 ml/min.
139 Initial oven temperature was 70 °C for 1 min and the temperature gradient was 5 °C/min from
140 70 °C to 270 °C with no holding time. Sample was introduced in splitless injection mode.
141 Injector temperature was 270 °C and purged for 1 min with a purge flow of 25 ml/min. Total run
142 time was 41 min. The MS source temperature was set at 250 °C and the GC interface was 275 °C.
143 A fatty acids methyl ester mixture (FAMQ-005 FAME Reference Standard, AccuStandard Inc.,
144 CT, USA) was used as external standards. The mass spectrum and the corresponding retention
145 time of each component were exported to NIST MS Search 2.0 software to identify compounds
146 by means of library and standard comparisons.

147 ***454 sequencing of the SL-based transcriptomes.*** mRNA samples isolated as described above
148 from the four culture conditions were used to synthesize cDNA, which was subsequently used
149 for 454 sequencing as reported previously (Kuo et al., under peer review process). Briefly, a
150 modified random oligo named 454AT$_7$N$_9$ (5'-
151 CGTATCGCCTCCCTCGCGCCATCAGTAATACGACTCACTATAGGGAGNNNNNNNNN-
152 3', where N is any of the 4 nucleotides) was used to synthesize the 1st strand cDNAs. The
153 cDNAs were then used as templates for PCR amplification of the 5'-end of the cDNAs of the
154 *trans*-spliced transcripts using ExTaq with a *Eutreptiella* spliced leader-based primer 454BEutSL,
155 5'-
156 GAGACTATGCGCCTTGCCAGCCCGCTCAGACACTTTCTGAGTGTCTATTTCTTTTCG-
157 3'), paired with 454AT$_7$ (5'-
158 CGTATCGCCTCCCTCGCGCCATCAGTAATACGACTCACTATAGGGAG-3'). After PCR
159 we selected the amplicons from agarose gel in the size range of 300-700 bp as the template for
160 emulsion PCR (emPCR) using GS Titanium SV emPCR Kit. Sequencing was carried out using

161 GS Titanium Sequencing Kit on the GS FLX System at the Center for Applied Genetics and
162 Technology, University of Connecticut.

163 ***Sequence processing and annotation.*** Sequence reads were processed and analyzed as recently
164 reported (Kuo et al. 2012). In short, raw sequencing reads from all of the 4 samples were pooled
165 together and trimmed using CLC Genomics Workbench (CLC Bio, Aarhus, Demark). After
166 quality and primer trimming, sequences shorter than 150 nt were discarded. In order to filter off
167 454 sequencing errors and create reduced-redundancy sequence dataset, USEARCH (Edgar 2010)
168 was used for sequence clustering. The resulting unique transcripts were then annotated using
169 Blast2Go V.2.5.0 (Götz et al. 2008) against NCBI's non-redundant (nr) database using BLASTx
170 algorithm (Altschul et al. 1990), with a cut-off-E-value $\leq 10^{-3}$.

171 ***Reverse-transcription quantitative PCR (RT-qPCR).*** In order to further investigate the
172 expression patterns of genes potentially regulating lipid production in *Eutreptiella*, some of the
173 genes identified from the transcriptomic data as related to carbon fixation and lipid synthesis
174 were selected and further analyzed by RT-qPCR. First strand cDNA libraries were prepared by
175 the methods mentioned earlier. It is suggested that at least two reference genes are needed for
176 proper normalization of gene expression levels (Guo and Ki 2012, Vandesompele et al. 2002).
177 Tubulin, actin, and elongation factor 1α were selected as candidates, as they had been shown to
178 be good reference genes in some microalgae and plants (Guo and Ki 2012, Le Bail et al. 2008).
179 We used the geometric average of the expression levels of these reference genes to normalize the
180 measured expression levels of target genes as reported (Vandesompele et al. 2002). For each
181 gene investigated, two sets of primers were designed (Table 3). To prepare standards, the primer
182 combination that specifies longer gene fragment was used to PCR amplify each target gene from

183 *Eutreptiella* cDNA. The amplicons of the target genes were checked with gel electrophoresis to
184 assure the absence of primer dimmers and then purified using DNA Clean & Concentrate™
185 column (Zymo Research, Orange, CA). The purified DNA was serially diluted to 10^2, 10^3, 10^4,
186 10^5, and 10^6 gene copies to generate standard curves to analyze the amplification efficiency and
187 primer specificity for every primer pair. qPCR was performed using the iCycler iQ™ real-time
188 PCR Detection System (Bio-Rad Laboratories, Inc., Hercules, CA) with SYBR Green supermix.
189 The qPCR reactions included a single denaturation cycle of 95 °C for 3 min, 40 cycles of 95 °C
190 for 20 sec, 58 °C for 30 sec, and 72 °C for 15 sec, followed by a melt curve analysis from 55 to
191 100 °C.

192 **Results**

193 ***Physiological conditions of the cultures.*** Two sets of cultures were grown under phosphate-
194 depleted (to produce high-lipid culture) and phosphate-replete (to produce low-lipid culture)
195 conditions. On the 11th day, the phosphate-replete cultures were in exponential growth phase
196 whereas phosphate-depleted cultures already reached stationary phase. When normalized to per
197 cell basis, the lipid contents of the cultures grown in phosphate-depleted medium were about 2-
198 fold higher than that of the cultures grown in phosphate-replete medium (Figure 1). This
199 observed difference in lipid content prompted cell harvesting from both cultures in the light and
200 dark periods, which yielded four samples referred to as high-lipid-light, high-lipid-dark, low-
201 lipid-light, and low-lipid-dark cultures.

202 ***Comparison of gene diversity among samples.*** The four samples were subjected to RNA
203 extraction, cDNA library construction, and 454 sequencing (Kuo et al. 2012). The Venn diagram
204 in Figure 2 summarizes the numbers of unique genes, total genes and shared genes among the 4

205 cDNA libraries. The analysis revealed that each library contained different numbers of total
206 genes as well as different unique genes that were not shared by any other libraries. Between the
207 two light cultures, the high-lipid cultures had more unique genes and total genes. Between the
208 cultures in the same lipid content categories, cultures harvested under the light condition
209 displayed more unique genes (i.e. light-specific) and total genes. Overall, high-lipid-light
210 cultures expressed the highest number of unique and total genes among the 4 samples,
211 suggesting that a more complex system was involved in lipid production and light reaction in
212 *Eutreptiella* sp.

213 ***Fatty acid synthesis.*** Based on the functional annotation, we have identified genes encoding key
214 enzymes involved in the lipid biosynthesis and catabolism (Table 1). Fatty acid biosynthesis
215 from acetyl-CoA involved in two enzymes: ACCase and FAS. Eukaryotic type ACCase (72%
216 identical to ACCase of *Thalassiosira pseudonana*; GenBank accession number: XP_002296083)
217 and type I FAS (67% identical to FASI of *Caenorhabditis elegans*; NP_492417) were found in
218 our dataset. Figure 3 illustrates a potential lipid biosynthesis pathway of *Eutreptiella* sp. ACCase
219 catalyzes carboxylation of acetyl-CoA to produce malonyl-CoA, which is catalyzed to malonyl-
220 ACP by malonyl-CoA transacylase. Next, long-chain saturated fatty acids are synthesized by
221 type I fatty acid synthase, which have 7 active sites within a single large polypeptide complex
222 (Nelson and Cox, 2008). The active enzyme domains include β-ketoacyl-ACP synthase (KAS),
223 malonyl/acetyl-CoA-ACP transferase (MAT), β –hydroxyacyl-ACP dehydratase (HAD), enoyl-
224 ACP reductase (EAR), β –ketoacyl-ACP reductase (KAR), and Acyl-ACP thioesterase (TE).
225 Synthesized palmitate (C16:0) and stearate (C18:0) via the activity of FASI are further subjected
226 to elongation, desaturation, or other modifications (see next section for more details). Besides
227 FASI, individual genes encoding β –ketoacyl-ACP reductase and Acyl-ACP thioesterase related

228 to type II fatty acid synthase (FASII) were also identified (Table 1). Malonyl-CoA transacylase,
229 which functions in the fatty acid synthesis pathway in other plants, was missing in our dataset.

230 ***Fatty acid identification and biosynthesis pathways.*** The stepwise desaturation and elongation
231 of C18 acid lead to the extension of fatty acids and synthesis of unsaturated fatty acids (Nelson
232 and Cox, 2008). The major fatty acids found in *Eutreptiella* sp. were the very long chain
233 polyunsaturated fatty acids, docosahexaenoic acid (DHA, 22:6n-3). Various desaturases and
234 elongases were found in our study (Table 1), including Δ5 desaturase, Δ6 desaturase, Δ8
235 desaturase, Δ9 desaturase, Δ12 desaturase and Δ15 desaturase. The elongases were: Δ5 elongase,
236 Δ6 elongase, and Δ9 elongase.

237 In order to link the identified genes to their expected roles in *Eutreptiella* sp., the fatty
238 acid (FA) profiles of *Eutreptiella* sp. were analyzed by screening C16-C24 fatty acids via
239 GC/MS/MS. We identified 24 fatty acids (including C16, C20, C22, and C24 fatty acids) from
240 the cultures of *Eutreptiella* sp. (Figure 4 and Table 2). In both of the phosphate-depleted and
241 phosphate-replete cultures, *Eutreptiella* sp. was relatively rich in polyunsaturated fatty acids
242 (PUFAs), especially very long chain polyunsaturated ω3 fatty acids (VLCPFA),
243 docosahexaenoic acid (DHA, 22:6) and eicosapentaenoic acid (EPA, 20:5). DHA had the highest
244 amount in both cultures (Figure 4), so we used it as a standard to estimate relative abundance of
245 other fatty acids (Table 2). Trace amount of tetracosahexaenoic acid, a C24 VLCPFA, was found.
246 Three saturated fatty acids, palmitic acid (16:0), stearic acid (18:0), and behenic acid (20:0),
247 were also identified. The abundance of saturated fatty acids was relatively low in the phosphate-
248 depleted cultures. The analysis also revealed the presence of C16 unsaturated fatty acids: 7-
249 hexadeceonoid acid (16:1), 9-hexadecenoid acid (16:1), 7,10-hexadecadienoic acid (16:2), 7, 10,

250 13-hecadecatrienoic acid (16:3), and 4, 7, 10, 13-hexadecatetraenoic acid (16:4) in both of the
251 cultures. In the phosphate-depleted cultures, the relative abundance of C16 and C18 fatty acids
252 was lower than the fatty acids in the phosphate-replete cultures (Figure 4 and Table 2),
253 suggesting that phosphate availability in the culture medium affected the lipid composition in
254 *Eutreptiella* sp.

255 Based on the transcriptome and GC/MS/MS results, *Eutreptiella* sp. is predicted to be
256 capable of producing various intermediates via ω3 and ω6 pathways, as well as producing
257 VLCPFA EPA and DHA by using the Δ8 alternative pathway, as reported in *Euglena* (Meyer et
258 al. 2003; Qi et al. 2004; Wallis and Browse 1999). Figure 5 shows the potential fatty acid
259 elongation pathways for *Eutreptiella* sp. as constructed from the transcriptome results. The first
260 desaturation for unsaturated fatty acids synthesis is catalyzed by Δ9 desaturase to introduce a
261 double bond into stearic acid resulting in 18:1 –ACP. Then a series of elongation and
262 desaturation takes place to produce the 24 types of unsaturated fatty acid mentioned before
263 (Figure 5). The elongation procedure is terminated when the acyl group is removed by acyl-ACP
264 thioesterase (TE) or oleoyl-ACP hydrolase (OAH), or when the fatty acids are transferred to the
265 biosynthesis of triacylglycerol.

266 ***Triacylglycerol (TAG) biosynthesis.*** Microalgal TAG biosynthesis is proposed to take place via
267 transfer of fatty acids from CoA to glycerol-3-phosphate using the direct glycerol pathway
268 (Ratledge 1988). However, the existing knowledge on the pathways and enzymes involved in
269 TAG biosynthesis in most of the microalgae (e.g. euglenophytes) are limited. In this study, four
270 transcripts coding for the enzymes involved in TAG biosynthesis were found in our *Eutreptiella*
271 sp. transcriptomic dataset. They were glycerol 3-phosphate dehydrogenase (GPD), glycerol 3-

272 phsphate acyltransferase (GPAT), and acylglycerophosphate acyltransferase (AGPAT) (Table 1).
273 For TAG biosynthesis, the precursor glycerol 3-phosphate is produced during glycolysis from
274 dihydroxyacetone phosphate by the action of glycerol 3-phosphate dehydrogenase.

275 Based on the data, a potential pathway for *Eutreptiella* sp. TAG biosynthesis is proposed
276 (Figure 3). The first step is the acylation of the two hydroxyl groups of glycerol3-phosphate by
277 two fatty acyl-CoA to yield phosphatidic acid, which is hydrolyzed by phosphatidate
278 phosphatase to form 1,2 diacylglycerol. Diacylglycerols are then converted to triacylglycerols by
279 transferring a third fatty acyl-CoA. Phosphatidic acid phosphohydrolase (PAP) and
280 diacylglycerol acyltransferase (DGAT) were missing in our cDNA libraries.

281 ***Gene expression levels as measured using reverse transcription quantitative PCR.*** We further
282 confirmed the differential expression patterns of several genes using reverse transcription
283 quantitative PCR (RT-qPCR). Nine genes involved in lipid synthesis were chosen for the
284 analysis, including type I fatty acid synthase (FASI), 3 different types of β –ketoacyl-ACP
285 reductase (KAR1, KAR2, KAR3, 12-33% identity at amino acid level), acetyl-CoA carboxylase
286 (ACCase), acyl carrier protein (ACP), thioesterase (TE), Δ8 desaturase, and an unclassified fatty
287 acid desaturase (FAD). Seven of the nine genes, KAR2, KAR3, FAD, ACCase, ACP, Δ8, and
288 TE had higher expression levels in the high-lipid-light cultures (Figure 6 and Figure 7). The
289 expression of FASI showed an opposite trend between high-lipid and low-lipid cultures. In high-
290 lipid cultures, FASI was expressed at a higher level in the light period, whereas in the low-lipid
291 cultures, FASI was expressed at a lower level in the light period (Figure 6). As for genes
292 involved in triacylglycerol synthesis, glycerol 3-phosphate dehydrogenase (GPD) and 1-acyl-sn-

293 glycero3-phosphate o-acyltransferase (AGPAT) exhibited slightly higher expression (transcript
294 abundance) in high-lipid-light cultures (Figure 7).

295 We also examined the expression of genes potentially involved in carbon fixation in
296 *Eutreptiella* sp.: phosphoenolpyruvate carboxylase (PEPCase), 2 variants of
297 phosphoenolpyruvate carboxykinase (PEPCK1 and PEPCK2, 54% identical at amino acid level),
298 pyruvate-phosphate dikinase (PPDK), and phosphoribulokinase (PRK). Three out of the five
299 genes (i.e. PPDK, PRK, and PEPCK1) showed positive correlation of transcript abundance with
300 lipid production. In particular, PEPCK1 and PRK expression levels in the high-lipid-light
301 cultures were over 6- and 4-fold higher than in the low-lipid-light cultures, respectively (Figure
302 8).

303

304 **Discussion**

305 ***Unique lipid and fatty acid production patterns.*** While the effects of depleting nitrate on lipid
306 production has been studied in many algae (Griffiths and Harrison 2009, Hu et al. 2008), effects
307 of phosphate are less documented. Our preliminary experiments showed that phosphate depletion
308 more strongly promotes lipid content in *Eutreptiella* sp. cells than nitrate depletion (data not
309 shown). The study reported here also showed that phosphate stress markedly increased
310 *Eutreptiella* lipid content. Among the cultures grown in parallel, lipid content was 2 fold as
311 much in the phosphate-depleted cultures as that in the phosphate-replete, f/2 medium-grown,
312 cultures (Figure 1). The type of main energy reserve (carbohydrates or lipids) in algae depends
313 on which carbon fixation product is utilized in the Calvin-Benson cycle. Among those products,
314 hexose phosphates are used in the synthesis of carbohydrates, while 3-phophoglycerate is used

315 for lipid synthesis (Raven 1974). In addition, the availability of inorganic nitrogen influences
316 carbon partition (Huppe and Turpin 1994, Raven 1974). The presence of inorganic nitrogen
317 source favors 3-phosphoglycerate and increases the flux of photosynthetically fixed carbon into
318 citrate cycle (Raven 1974), which is one of the sources for acetyl CoA (Nelson and Cox 2008).

319 Interestingly, *Eutreptiella* sp. has abundant polyunsaturated fatty acids. DHA was the
320 most abundant fatty acid from both the phosphate-depleted and phosphate-replete cultures.
321 Polyunsaturated fatty acids are valuable for cold adaptation (Wallis et al. 2002 and references
322 therein). The relatively high abundance of unsaturated fatty acids in *Eutreptiella* sp. may be due
323 to the adaptation to their native environment in Long Island Sound, where the temperature can be
324 below 0 °C in the winter (McCarthy 2008). All the fatty acids that we identified in this study
325 were also present in *Euglena gracilis* (Hulanicka et al. 1964, Korn 1964, Meyer et al. 2003),
326 except C24:1 and C24:6 fatty acids. Different from *E. gracilis*, *Eutreptiella* sp. has a very high
327 proportion of very long chain unsaturated fatty acids (C22 and longer, i.e. DHA and EPA),
328 indicating that this species may be a nutritional food resource for zooplankton and other
329 heterotrophic organisms in the marine ecosystem. Moreover, the proportions of saturated fatty
330 acids (i.e. C16:0, C18:0, and C20:0) and unsaturated C16 and C18 fatty acids decreased when
331 phosphate was depleted (Figure 4 and Table 2), suggesting that it is possible to enhance the
332 production of desired components of fatty acids from *Eutreptiella* sp. by manipulating nutrient
333 composition in the culture medium. Although it is not clear why the abundance of C16:0 and
334 C16: 4 fatty acids were much lower in the phosphate-depleted cultures, phosphate stress might
335 alter the composition of fatty acids in the algal cells, as plants and algae can respond to stress by
336 adjusting membrane fluidity and by releasing unsaturated fatty acids from the membrane
337 (Maréchal et al. 1997, Upchurch 2008).

338

339 ***Expressed genes linked to lipid synthesis in* Eutreptiella *sp.*** Fatty acids are synthesized via two
340 enzyme systems, ACCase and FAS (Nelson and Cox 2008). ACCase plays a critical role in
341 regulating fatty acid synthesis and is essential for lipid metabolism as it catalyzes the first step of
342 fatty acid synthesis and is found in all kingdoms of life except archaea (Cronan Jr and Waldrop
343 2002, Harwood 1988). It comprises three functional constituents: biotin carrier protein, biotin
344 carboxylase, and transcarboxylase. There are two distinct types of ACCase in nature: one
345 consists of three separate subunits (prokaryote-type), and the other is a single multi-domain
346 multifunctional polypeptide (eukaryote-type) (Sasaki et al. 1995). The eukaryote-type occurs in
347 mammals, fungi or yeast, while both the prokaryote- and the eukaryote- type can be found in
348 plants (Huang et al. 2002, Sasaki et al. 1995). Nine 454 sequencing reads encoding an ACCase
349 were found in our dataset, which shared ~72% amino acid identity to an eukaryotic type found in
350 the diatom *T. pseudonana*, suggesting *Eutreptiella* sp. ACCase to be of the eukaryotic type. In
351 the qPCR analysis, ACCase showed highest expression level in the high-lipid-light cultures,
352 indicating that this gene is related to photosynthetic lipid production in *Eutreptiella* sp.

353 Our results suggest that *Eutreptiella* sp. may have both type I fatty acid synthase (FASI)
354 and type II fatty acid synthase (FASII), which are the two major types of fatty acid synthases that
355 naturally occur in different groups of organisms (Nelson and Cox 2008). FASI, found in
356 vertebrate and fungi, is a single multifunctional polypeptide with 7 active sites, whereas FASII,
357 found in plants and bacteria, compose discrete and freely diffusible enzymes. Interestingly,
358 *Euglena gracilis*, which is another euglenoid species, has different types of fatty acid synthesis
359 pathways and enzymes to synthesize lipids under different environmental conditions (Delo et al.

360 1971, Goldberg and Bloch 1972, Inui et al. 1984). ACP-dependent FASI is located in the cytosol
361 and mainly produces C16 fatty acids (Goldberg and Bloch 1972). When grown in the light, *E.*
362 *gracilis* utilizes ACP-dependent FAS II in plastids to produce fatty acids (Ernst-Fonberg and
363 Bloch 1971). Inui et al. (1984) also reported a malonyl-CoA independent fatty acid synthesis
364 system to form wax ester from paramylon as energy storages in *E. gracilis* under anaerobic
365 conditions. We found genes coding for FASI and for β –ketoacyl-ACP reductase and Acyl-ACP
366 thioesterase of FASII in *Eutreptiella* sp. Our failure to retrieve the rest of FASII gene complex in
367 our cDNA libraries could be due to their being plastid-coded genes, which would not contain
368 Eut-SL that we used to construct the cDNA libraries, or just simply were not expressed. Whether
369 *Eutreptiella* sp. contains the other type of ACP-dependent FAS similar to that in *E. gracilis*
370 remains to be further investigated in the future.

371 Based on the fatty acid composition and possession of genes coding for enzymes
372 involved in fatty acid elongation and desaturation in *Eutreptiella* sp., it is evident that the ω3, ω6
373 and alternative Δ8 pathways occur in the euglenoid algae. Different from most of plants and most
374 algae, we found that *Eutreptiella* is similar to *Euglena* in using Δ8 desaturase to introduce the
375 third double bond on 20:2 (11, 14) to form 20:3 (8, 11, 14) (Meyer et al. 2003, Qi et al. 2004).
376 The 20:3 fatty acid is then converted to 22:5 EPA via stepwise Δ5 desaturase, Δ5 elongase and
377 Δ4 desaturase (Figure 5). In order to produce DHA, *Eutreptiella* may also utilize another Δ8
378 pathway to introduce the 4th double bound on 20:3 (11,14,17), resulting in 20:4 (8,11,14,17) for
379 later stepwise process of elongation and desaturation. Although Δ4 desaturase was not identified
380 in our dataset, it is very likely that *Eutreptiella* sp. has this gene since it was identified and
381 involved in DHA synthesis in *E. gracilis* (Meyer et al. 2003). In our study, both EPA would need
382 Δ4 desaturase to form their last double bond. We also detected trace amount of 15-tetracosenoic

383 acid (24:1) and 6,9,12,15,18,21-tetracosahexaenoic acid (24:6). The 24:6 fatty acid can be found
384 in some algae (Mansour et al. 2005), but the synthesis pathway is not well studied. We proposed
385 that the 24:6 fatty acid could potentially be synthesized by elongation from DHA or elongation
386 and desaturation from other C22 fatty acids.

387 The 16:4 (4,7,10,13) fatty acid was also abundant in phosphate-replete cultures of
388 *Eutreptiella* sp. Highly unsaturated stearidonic acid (18:4) and hexadecatetraenoic acid (16:4) are
389 produced in some macroalgae (Dembitsky et al. 1991, Ishihara et al. 2000) and microalgae
390 (Wallis and Browse 1999). Unsaturated C16 fatty acids have been identified in *E. gracilis*
391 (Constantopoulos and Bloch 1967). Increasing light intensity can enhance the production of 16:4
392 acid and 18:3 α-linolenic acid in *E. gracilis* (Constantopoulos and Bloch 1967). Korn (1964)
393 suggested a possible C16 pathway for *E. gracilis*: 16:1 (7) → 16:2 (7, 10) → 16:3 (7, 10, 13)
394 →16:4 (4, 7, 10, 13). Although we identified 16:4 (4, 7, 10, 13) in *Eutreptiella* sp., the genes
395 responsible for the synthesis of this abundant fatty acid remain uncertain.

396

397 ***Regulation of genes associated with lipid production.*** Our RT-qPCR revealed that many of the
398 genes examined had different expression levels between the dark and light periods. Genes
399 involved in high lipid biosynthesis showed higher expression levels under light condition,
400 suggesting that these genes promoted lipid production in *Eutreptiella* sp. during photosynthesis.
401 Those genes were ACCase, ACP, KAR2, KAR3, FAD, Δ8 desaturase, TE, AGPAT, and GPD.
402 The results of RT-qPCR verified our physiological observation from a 24h diel experiment,
403 showing that lipids in *Eutreptiella* sp. cells start to accumulate during the light period after cells
404 have actively divided (Kuo and Lin, 2012). Other evidence showed that the activity of ACCase is

405 correlated to that of the Calvin-Benson Cycle enzymes, with the greatest activities under light
406 (Sasaki et al. 1997) and negligible in the dark (Bao et al. 2000).

407 Three genes encoding carbon fixation enzymes showed higher expression levels under
408 the high-lipid-light condition, indicating that those enzymes are related to the biosynthesis of
409 precursors for lipid production. Phosphoribulokinase (PRK) and one copy of
410 phosphoenolpyruvate carboxykinase (i.e. PEPCK1) exhibited the highest level of relative gene
411 expression among these genes. PEPCK activity is linked to CO_2 fixation in algae (Reiskind and
412 Bowes 1991). It converts oxaloacetate into CO_2 and phosphoenolpyruvate, bridging C4 to C3
413 pathway in providing CO_2 for photosynthesis. It is noteworthy that C4 enzymes have been
414 identified in this species (Kuo et al. 2012). The phosphoenolpyruvate can later be converted into
415 pyruvate, which is an important precursor for Acetyl-CoA. PRK catalyzes the reaction of ATP
416 and ribulose 5-phosphate to produce ADP and ribulose 1,5-bisphosphate, the substrate needed to
417 capture CO_2 in photosynthesis. The concomitant up-regulation of the fatty acid synthesis and
418 photosynthesis genes indicates the linkage of the two carbon metabolic processes.

419

420 **Acknowledgements**

421 We thank the Center of Environmental Science and Engineering at UConn for providing the
422 facility of GC/MS/MS, and Yunyun Zhuang for helping with 454 sequencing. The research was
423 supported by the Predoctoral Awards of the Department of Marine Sciences at UConn (to RCK),
424 Multidisciplinary Environmental Research Award from the Center of Environmental Science and
425 Engineering at UConn (to RCK), and National Science Foundation "Assembling the Tree of Life
426 (AToL)" grant EF-0629624 (to SL and collaborators).

427

428 **References**

429 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment
430 search tool. *J. Mol. Biol.* 215: 403-410
431 Álvarez-Góngora, C. & Herrera-Silveira J. A. 2006. Variations of phytoplankton community
432 structure related to water quality trends in a tropical karstic coastal zone. *Mar. Pollut.*
433 *Bull.* 52: 48-60
434 Anderson, D. M., Hoagland, P., Kaoru, Y. & White, A. W. 2000. Estimated annual economic
435 impacts from harmful algal blooms (HABs) in the Unites States. Woods Hole, MA, USA:
436 Woods Hole Oceanographic Isntitution
437 Arzel J., Martinez Lopez, F. X., Métailler, R., Stéphan, G., Viau, M., Gandemer, G. & Guillaume,
438 J. 1994. Effect of dietary lipid on growth performance and body composition of brown
439 trout (*Salmo trutta*) reared in seawater. Aquaculture. 123: 361-375
440 Bao, X., Focke, M., Pollard, M. & Ohlrogge, J. 2000. Understanding in vivo carbon precursor
441 supply for fatty acid synthesis in leaf tissue. The Plant Journal 22: 39-50
442 Bates, S. S., Strain & P. M. 2006. Nutrients and phytoplankton in Prince Edward Island Inlets
443 during late summer to fall: 2001-2003. Canadian Technical Report of Fisheries and
444 Aquatic Sceicnes 2668. Canada: Fisheries and Oceans
445 Bloch, K. & Vance, D. 1977. Control mechanisms in the synthesis of saturated fatty acids. Annu.
446 Rev. Biochem. 46: 263-298
447 Carvalho, A. P. & Malcata F. X. 2005. Preparation of fatty acid methyl esters for gas-
448 chromatographic analysis of marine lipids: insight studies. J. Agric. Food Chem. 53:
449 5049-5059
450 Constantopoulos, G. & Bloch, K. 1967. Effect of Light Intensity on the Lipid Composition of
451 Euglena gracilis. J. Biol. Chem. 242: 3538-3542
452 Cossins, A. R., Friedlander, M. J. & Prosser, C. L. 1977. Correlations between behavioral
453 temperature adaptations of goldfish and the viscosity and fatty acid composition of their
454 synaptic membranes. J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol. 120:
455 109-121
456 Cronan, J. E. & Waldrop, G. L. 2002. Multi-subunit acetyl-CoA carboxylases. Prog. Lipid Res.
457 41: 407-435
458 Delo, J., Ernst-Fonberg, M. L. & Bloch, K. 1971. Fatty acid synthetases from Euglena gracilis.
459 Arch. Biochem. Biophys. 143: 384-391
460 Dembitsky, V. M., Pechenkina-Shubina, E. E. & Rozentsvet, O. A. 1991. Glycolipids and fatty
461 acids of some seaweeds and marine grasses from the black sea. Phytochemistry 30: 2279-
462 2283
463 Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics
464 26: 2460-2461
465 Ernst-Fonberg, M. L & Bloch, K. 1971. A chloroplast-associated fatty acid synthetase system in
466 Euglena. Arch. Biochem. Biophys. 143: 392-400
467 Goldberg, I. & Bloch, K. 1972. Fatty acid synthetases in *Euglena gracilis*. J. Biol. Chem. 247:
468 7349-7357

469 Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles,
470 M., Talón, M., Dopazo, J. & Conesa, A. 2008. High-throughput functional annotation and
471 data mining with the Blast2GO suite. Nucleic Acids Res 36: 3420-3435
472 Gouveia, L. & Oliveira, A. 2009. Microalgae as a raw material for biofuels production. J. Ind.
473 Microbiol. Biotechnol. 36: 269-274
474 Griffiths, M. & Harrison, S. 2009. Lipid productivity as a key characteristic for choosing algal
475 species for biodiesel production. J. Appl. Phycol. 21: 493-507
476 Gunstone, F. D., Harwood, J. L. & Dijkstra A. J. 2007. *The Lipid Handbook ,* 2nd ed. Taylor &
477 Francis, FL, USA
478 Guo, R.& Ki, J. S. 2012. Evaluation and validation of internal control genes for studying gene
479 expression in the dinoflagellate Prorocentrum minimum using real-time PCR. Eur. J.
480 Protistol. 48(3): 199-206
481 Guschina, I. A & Harwood, J. L. 2006. Lipids and lipid metabolism in eukaryotic algae. Prog
482 Lipid Res 45: 160-186
483 Harwood, J. L. 1988. Fatty acid metabolism. Annu. Rev. Plant Physiol. Plant Mol. Biol. 39: 101-
484 138
485 Henriksen, P., Riemann, B., Kaas, H., Sørensen, H. M. & Sørensen, H. L. 2002. Effects of
486 nutrient-limitation and irradiance on marine phytoplankton pigments. J. Plankton Res. 24:
487 835-858
488 Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M. & Darzins, A. 2008.
489 Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and
490 advances. The Plant J. 54: 621-639
491 Huang, S., Sirikhachornkit. A., Faris, J. D., Su, X., Gill, B. S., Haselkorn, R. & Gornicki, P. 2002.
492 Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci
493 in wheat and other grasses. Plant Mol. Biol. 48: 805-820
494 Hulanicka, D., Erwin, J & Bloch, K. 1964. Lipid metabolism of Euglena gracilis. J. Biol. Chem.
495 239: 2778-2786
496 Huppe, H. C.& Turpin, D. H. 1994. Intergration of carbon and nitrogen metabolism in plant and
497 algal cells. Annu. Rev. Plant Physiol. Plant Mol. Biol. 45: 577-607
498 Illman, A. M., Scragg, A. H. & Shales, S. W. 2000. Increase in Chlorella strains calorific values
499 when grown in low nitrogen medium. Enzyme Microb. Technol. 27: 631-635
500 Inui, H., Miyatake, K., Nakano, Y. & Kitaoka, S. 1984. Fatty acid synthesis in mitochondria of
501 Euglena gracilis. Eur. J. Biochem. 142: 121-126
502 Ishihara, K., Murata, M., Kaneniwa, M., Saito, H., Komatsu, W. & Shinohara, K. 2000.
503 Purification of stearidonic Acid (18:4(n-3)) and hexadecatetraenoic Acid (16:4(n-3))
504 from algal fatty acid with lipase and medium pressure liquid chromatography. Biosci.
505 Biotechnol. Biochem. 64: 2454-2457
506 Klein Breteler, W. C. M., Koski, M. & Rampen, S. 2004. Role of essential lipids in copepod
507 nutrition: no evidence for trophic upgrading of food quality by a marine ciliate. Mar. Ecol.
508 Prog. Ser. (Halstenbek) 274: 199
509 Korn, E. D. 1964. The fatty acids of Euglena gracilis. J. Lipid Res. 5: 352-362
510 Kuo, R. C & Lin, S. 2012. Ectobiotic and endobiotic bacteira associated with *Eutreptiella* sp.
511 isolated from Long Island Sound. Protist 164 (1): 60-74
512 Kuo, R. C., Zhang, H., Zhuang, Y., Hannick, L.& Lin, S. 2013 Transcriptomic study reveals
513 widespread spliced leader *trans*-splicing, short 5'-UTRs and potential complex carbon

514 fixation mechanisms in the euglenoid alga *Eutreptiella* sp. PLoS ONE. 8(4): e60826.
515 doi:10.1371/journal.pone.0060826.
516 Le Bail, A., Dittami, S., De Franco, P. O., Rousvoal, S., Cock, M., Tonon, T.& Charrier, B. 2008.
517 Normalisation genes for expression analyses in the brown alga model Ectocarpus
518 siliculosus. BMC Mol. Biol. 9: 75
519 Lee S., Yoon, B. D. & Oh, H. M. 1998. Rapid method for the determination of lipid from the
520 green alga Botryococcus braunii. Biotechnol Tech. 12: 553-556
521 Lindholm, T. 1995. Green water caused by *Eutreptiella gymnastica* (Euglenophyceae) in a
522 stratified Baltic Sea inlet. In *Lassus, P., Arzul G., Denn EE-L., Gentien, P., Baut CM-L.,*
523 *[Eds.] Harmful Marine Alagae Blooms (Proceedings of the Sixth International*
524 *Conference on Toxic Marine Phytoplankton).* Nantes, France.
525 Liu, Z. Y., Wang, G. C.& Zhou, B. C. 2008. Effect of iron on growth and lipid accumulation in
526 Chlorella vulgaris. Bioresour. Technol. 99: 4717-4722
527 Mansour, M. P., Frampton, D. M. F., Nichols, P. D., Volkman, J. K.& Blackburn, S. I. 2005.
528 Lipid and fatty acid yield of nine stationary-phase microalgae: Applications and unusual
529 C24–C28; polyunsaturated fatty acids. J. Appl. Phycol. 17: 287-300
530 Maréchal, E., Block, M. A., Dorne, A. J., Douce, R.& Joyard, J. 1997. Lipid synthesis and
531 metabolism in the plastid envelope. Physiol. Plant 100: 65-77
532 Mccarthy, G. 2008. Monitoring Long Island Sound 2008. Connecticut Department of
533 Environmental Protection, Connecticut
534 Meyer, A., Cirpus, P., Ott, C., Schlecker, R., Zähringer, U. & Heinz, E. 2003. Biosynthesis of
535 docosahexaenoic acid in *Euglena gracilis*: biochemical and molecular evidence for the
536 involvement of a Δ4-fatty acyl group desaturase. Biochemistry 42: 9779-9788
537 Nelson, D. L. &Cox, M. M. 2008. Lipid biosynthesis. In *Nelson, D. L. & Cox, M. M. [Eds.]*
538 *Lehninger Principles of Biochemistry*. W. H. Freeman and Company, New York, pp 805-
539 845
540 Olli, K., Heiskanen, A. S. & Seppälä, J. 1996. Development and fate of *Eutreptiella gymnastica*
541 bloom in nutrient-enriched enclosures in the coastal Baltic Sea. J. Plankton Res. 18:
542 1587-1604
543 Qi, B., Fraser, T., Mugford, S., Dobson, G., Sayanova, O., Butler, J., Napier, J. A., Stobart, A. K.
544 & Lazarus, C. M. 2004. Production of very long chain polyunsaturated omega-3 and
545 omega-6 fatty acids in plants. Nat. Biotech. 22: 739-745
546 Rainuzzo, J. R., Reitan, K. I. & Olsen, Y. 1997. The significance of lipids at early stages of
547 marine fish: a review. Aquaculture 155: 103-115
548 Ratledge, C. 1988. An overview of microbial lipids. In C Ratledge, SG Wilkinson, eds,
549 Microbial Lipids. Academic Press, New York, pp 3-21
550 Raven, J. A. 1974. Carbon dioxide fixation. In *Stewart, W. D. P. [Ed.] Algal Physiology and*
551 *Biochemistry: Biotenical Monographs, Vol 10*. University of California Press, California,
552 pp 434-455
553 Reiskind, J. B. & Bowes, G. 1991. The role of phosphoenolpyruvate carboxykinase in a marine
554 macroalga with C4-like photosynthetic characteristics. PNAS 88: 2883-2887
555 Rodríguez-Graña, L., Calliari, D., Conde, D., Sellanes, J. & Urrutia, R. 2008. Food web of a SW
556 Atlantic shallow coastal lagoon: spatial environmental variability does not impose
557 substantial changes in the trophic structure. Mar. Ecol. Prog. Ser. 362: 69-83
558 Sasaki, Y., Konishi, T. & Ngano, Y. 1995. The compartmentation of Acetyl-Coenzyme A
559 carboxylase in plants. Plant Physiol. 108: 445-449

560 Sasaki, Y., Kozaki, A. & Hatano, M. 1997. Link between light and fatty acid synthesis:
561 Thioredoxin-linked reductive activation of plastidic acetyl-CoA carboxylase. PNAS 94:
562 11096-11101
563 Schweizer, E. & Hofmann, J. 2004. Microbial type I fatty acid synthases (FAS): major players in
564 a network of cellular FAS systems. Microbiol Mol. Biol. Rev. 68: 501-517
565 Seong, K. A., Jeong, H. J., Kim, S., Kim, G. H. & Kang, J. H. 2006. Bacterivory by co-
566 occurring red-tide algae, heterotrophic nanoflagellates, and ciliates. Mar. Ecol. Prog. Ser.
567 322: 85-97
568 Singh, S. C. 2002. Role of lipids and fatty acids in stress tolerance in cyanobacteria. Acta.
569 Protozool. 41: 297
570 Stonik, I. 2007. Species of the genus *Eutreptiella* (Euglenophyceae) from Russian waters of
571 East/Japan Sea. Ocean Sci. J. 42: 81-88
572 Stonik, I. V. & Selina, M. S. 2001. Species composition and seasonal dynamics of density and
573 biomass of euglenoids in Peter the Great Bay, Sea of Japan. Russ. J. Mar. Biol. 27: 174-
574 176
575 Takeyama, H., Kanamaru, A., Yoshino, Y., Kakuta, H., Kawamura, Y. & Matsunaga, T. 1997.
576 Production of antioxidant vitamins, β-carotene, vitamin C, and vitamin E, by two-step
577 culture of *Euglena gracilis* Z. Biotechnol. Bioeng. 53: 185-190
578 Tang, K. W. & Taal, M. 2005. Trophic modification of food quality by heterotrophic protists:
579 species-specific effects on copepod egg production and egg hatching. J. Exp. Mar. Biol.
580 Ecol. 318: 85-98
581 Thelen, J. J. & Ohlrogge, J. B. 2002. Metabolic engineering of fatty acid biosynthesis in plants.
582 Metab. Eng. 4: 12-21
583 Upchurch, R. 2008. Fatty acid unsaturation, mobilization, and regulation in the response of
584 plants to stress. Biotechnol. Lett. 30: 967-977
585 Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. & Speleman,
586 F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric
587 averaging of multiple internal control genes. Genome Biology 3: research0034.0031 -
588 research0034.0011
589 Wallis, J. G. & Browse, J. 1999. The Δ8-Desaturase of *Euglena gracilis*: an alternate pathway
590 for synthesis of 20-carbon polyunsaturated fatty acids. Arch. Biochem. Biophys. 365:
591 307-316
592 Wallis, J. G, Watts, J. L. & Browse, J. 2002. Polyunsaturated fatty acid synthesis: what will they
593 think of next? Trends Biochem. Sci. 27: 467-473
594 White, S. W., Jie, Z., Yong-Mei, Z. & Rock, C. O. 2005. The structural biology of type II fatty
595 acid biosynthesis. Annu. Rev. Biochem. 74: 791-831
596 Worsham, L. M. S., Williams, S. G. & Ernst-Fonberg, M. L. 1993. Early catalytic steps of
597 *Euglena gracilis* chloroplast type II fatty acid synthase. Biochimica et Biophysica Acta
598 (BBA) - Lipids and Lipid Metabolism 1170: 62-71
599 Yamane, Y. I., Utsunomiya, T., Watanabe, M. & Sasaki, K. 2001. Biomass production in
600 mixotrophic culture of Euglena gracilis under acidic condition and its growth energetics.
601 Biotechnol. Lett. 23: 1223-1228

602

603

604

605 **Figure legends**

606 Figure 1. Growth and lipid production of *Eutreptiella* cultures under phosphate-replete and
607 phosphate-depleted conditions. (A) Growth curves of *Eutreptiella* cultures. (B) Growth rates and
608 lipid fluorescence levels at the time when cultures were harvested.

609 Figure 2. Common and unique expressed genes between different culture conditions. HL: high-
610 lipid-light cultures. HD: high-lipid-dark cultures. LL: low-lipid-light cultures. LD: low-lipid-dark
611 cultures. Numbers outside the diagram represent total number of genes found in each library.

612 Figure 3. Putative lipid biosynthesis pathway in *Eutreptiella* sp. inferred from our transcriptomic
613 data. GPD: glycerol 3-phosphate dehydrogenase (EC: 1.1.1.53); GPAT: glycerol-3-phosphate
614 acyltransferase (EC: 2.3.1.15); AGPAT: acylglycerophosphate acyltransferase(EC: 2.3.1.51); PP:
615 phosphatidate phosphatase (EC: 3.1.3.4); DGAT: diacylglycerol acyltransferase (EC: 2.3.1.20);
616 FA: fatty acids; FA-CoA: fatty acyl-CoAs. FA: fatty acids. Dashed boxes indicate enzymes that
617 were not found in our dataset. The process of FA elongation and desaturation is shown in Figure
618 5.

619 Figure 4. Chromatograms of total fatty acid esters from *Eutreptiella* sp. Note that DHA (peak 19)
620 was the most abundant fatty acid. Corresponding fatty acids (labeled by length of carbon, e.g.
621 C16) were listed in Table 3. (A) Chromatogram of fatty acid esters from the phosphate–replete
622 cultures. (B) Chromatogram of fatty acid esters from the phosphate-depleted cultures. Insets are
623 closer-ups of some of the small peaks of esters. The first number (bold-typed) by the peak is
624 peak ID, the second number retention time, and the third number peak area (only trace elements
625 are shown). Note that retention times between the two samples are slightly different.

626 Figure 5. Putative pathways of fatty acids elongation and desaturation inferred from *Eutreptiella*
627 sp. transcriptome annotation and GC/MS/MS analyses. Boxes with dashed lines indicate the
628 enzymes or substrates that were not found in our study.

629 Figure 6. Expression levels of genes involved in fatty acids biosynthesis in *Eutreptiella* sp. HD:
630 high-lipid-dark cultures. HL: high-lipid-light cultures. LL: low-lipid-light cultures. LD: low-
631 lipid-dark cultures.

632 Figure 7. Expression levels of genes involved in fatty acids biosynthesis (FAD, Δ 8, and TE) and
633 triacylglycerol biosynthesis (GPD, GPAT, and AGPT). HD: high-lipid-dark cultures. HL: high-
634 lipid-light cultures. LL: low-lipid-light cultures. LD: low-lipid-dark cultures.

635 Figure 8. Expression levels of genes involved in carbon fixation. HD: high-lipid-dark cultures.
636 HL: high-lipid-light cultures. LL: low-lipid-light cultures. LD: low-lipid-dark cultures.

637

638

639

640

641 Table 1. Genes potentially involved in fatty lipid biosynthesis in *Eutreptiella* sp. Sequences and BLAST
642 results were listed in the supporting file "lipid_biosynthesis_genes".

Gene	EC code	Number of transcripts
Fatty acid biosynthesis		
Acetyl-CoA carboxylase	EC 6.4.1.2	2
Fatty acid synthase (animal type)	EC 2.3.1.85	1
Acyl-ACP thioesterase	EC 3.1.2.14	1
Acyl carrier protein		1
β-ketoacyl-ACP reductase	EC 1.1.1.100	3
Fatty acid desaturation		
Δ^5 desaturase	EC 1.14.19.-	2
Δ^6 desaturase	EC 1.14.19.-	5
Δ^8 desaturase	EC 1.14.19.-	2
Δ^9 desaturase (Acyl-ACP desaturase)	EC 1.14.19.2	1
Δ^{12} desaturase	EC 1.14.19.-	1
Δ^{15} desaturase	EC 1.14.19.-	1
Fatty acid elongation		
Δ^5 elongase		1
Δ^6 elongase	EC 1.14.19.-	1
Δ^9 elongase	EC 1.14.19.2	1
Palmitoyl-CoA hydrolase	EC 3.1.2.2	1
Palmitoyl hydrolase	EC 3.1.2.22	1
acetyl-CoA C-acyltransferase	EC 2.3.1.16	1
3-hydroxyacyl-CoA dehydrogenase	EC 1.1.1.35	2
Triacylglycerol biosynthesis		
Glyerol-3-phosphate dehydrogenase	EC 1.1.5.3	1
Glycerol3-phosphate acyltransferase	EC 2.3.1.15	2
Acylglycerophosphate acyltransferase	EC 2.3.1.51	2
Fatty acid metabolism		
Acyl-CoA dehydrogenase	EC 1.3.99.3	3
Long-chain-fatty-acid-CoA ligase	EC 6.2.1.3	2
Acetyl-CoA C-acyltransferase	EC 2.3.1.16	1
3-hydroxyacyl-CoA dehydrogenase	EC 1.1.1.35	2
Carnitine o-palmitoyltransferase	EC 2.3.1.21	1
Enoyl-CoA hydratase	EC 4.2.1.17	1
Acetyl-CoA thiolase	EC 2.3.1.9	1
Aldehyde dehydrogenase (AND+)	EC 1.2.1.3	2

643 Table 2. Types and relative abundance of fatty acids detected in *Eutreptiella* sp.

Peak ID	Retention time	Systematic name	Abbreviation	Relative abundance (%)F/2*	Relative abundance (%)F/2-P*
1	26.19	4,7,10,13-Hexadecatetraenoic acid	16:4(n-3)	98.04	11.44
2	26.37	7,10-Hexadecadienoic acid	16:2(n-6)	2.59	0.08
3	26.47	7,10,13-Hexadecatrienoic acid	16:3(n-3)	12.63	0.54
4	26.62	7-Hexadecenoid acid	16:1(n-9)	11.89	0.27
5	26.97	9-Hexadecenoid acid	16:1(n-7)	13.54	0.79
6	27.1	Hexadecanoic acid	16:0	90.45	11.28
7	29.11	6,9,12-Octadecatrienoic acid	18:3(n-6)	3.90	0.13
8	30.04	6.9.12.15-Octadecatetraenoic acid	18:4(n-3)	41.00	12.40
9	30.23	9,12-Octadecadienoic acid	18:2(n-6)	3.20	0.53
10	30.36	9,12,15-Octadecatrienoic acid	18:3(n-3)	72.63	23.58
11	30.47	9-Octadecenoic acid	18:1(n-9)	7.97	0.58
12	30.85	Octadecanoic acid	18:0	8.43	2.12
13	33.1	5,8,11,14-Eicosatetraenoic acid	20:4(n-6)	6.44	1.88
14	33.22	5,8,11,14,17-Eicosapentaenoic acid	20:5(n-3)	78.69	40.02
15	33.53	8,11,14,17-Eicosatetraenoic acid	20:4(n-3)	8.39	4.02
16	33.77	8,11-Eicosadienoic acid	20:2(n-9)	1.13	0.71
17	33.87	11,14,17-Eicosatrienoic acid	20:3(n-3)	3.16	1.02
18	36.17	4,7,10,13,16-Docosapentaenoic acid	22:5(n-6)	68.78	54.93
19	36.3	4,7,10,13,16,19-Docosahexaenoic acid	22:6(n-3)	100.00	100.00
20	36.4	7,10,13,16-Docosatetraenoic acid	22:4(n-6)	43.11	47.18
21	36.49	7,10,13,16,19-Docosapentaenoic acid	22:5(n-3)	3.56	6.03
22	37.54	Docosanoic acid	22:0	5.94	1.01
23	38.79	6,9,12,15,18,21-Tetracosahexaenoic acid	24:6 (n-3)	5.32	0.85
24	40.14	15-Tetracosenoic acid	24:1(n-9)	0.46	-

644

645 *The numbers of relative abundance were obtained by dividing the peak area of the fatty acid to the peak
646 area of DHA; F/2 and F/2-P depict phosphate-replete and phosphate-depleted F/2 medium, respectively.

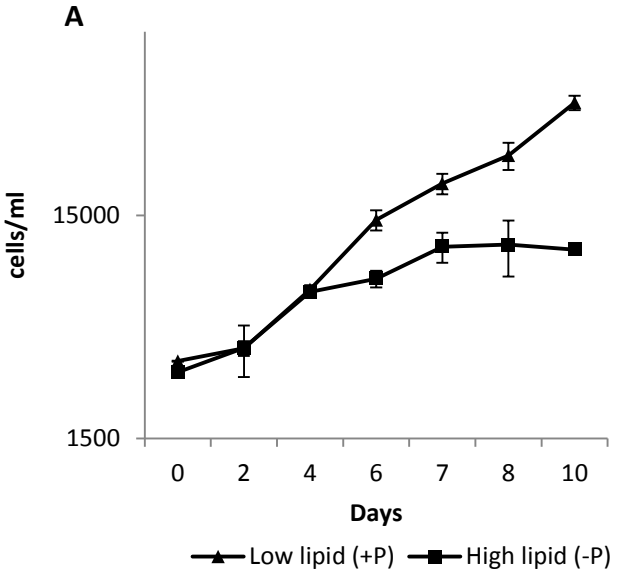
647

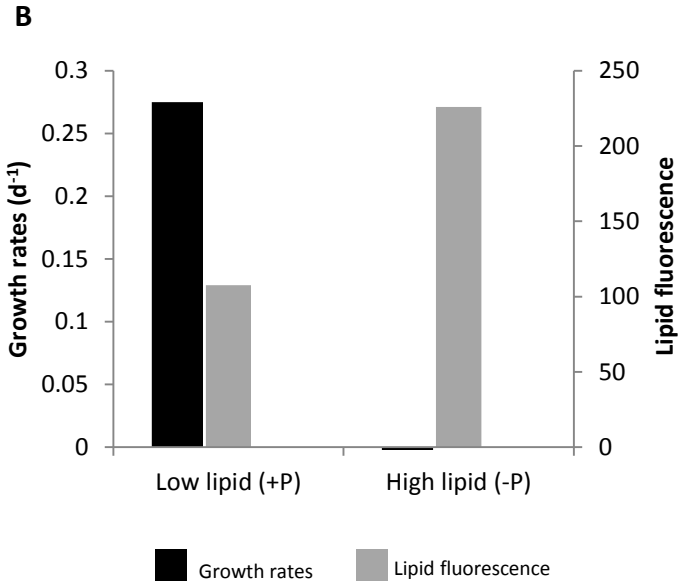
648 Table 3. Primers used in reverse-transcription quantitative PCR (RT-qPCR) analysis in this study.

Primer name	Sequences (5´-3´)	RT-qPCR for *Eutreptiella* sp. gene
ACPF1	GCAGCAAAGCAAATCAAACCA	acyl carrier protein forward
ACPF2	CAACCCAACTCTCACTTCTACAT	acyl carrier protein forward
ACPR	GCATCAGCGACCGATACAATA	acyl carrier protein reverse
AGPCF	TTACGAACTTGCGATGCTTCA	1-acyl-sn-glycerol-3-phosphate acyltransferase forward
AGPCR1	GGAGGTATGATTGGAGATGATGAG	1-acyl-sn-glycerol-3-phosphate acyltransferase reverse
AGPCR2	CGATGACTTGACACCACCAA	1-acyl-sn-glycerol-3-phosphate acyltransferase reverse
ACCASE-F	GTGTTGGCATGTTGTCTCTTG	ACCase forward
ACCASE-R1	GGATACCGCCATTGGAACC	ACCase reverse
ACCASE-R2	GAAGTCGGCTTGAATCTGTGA	ACCase reverse
DELTA8FADF	GACATCATTGAGAATTATCGGAACC	delta-8 fatty acid desaturase forward
DELTA8FADR1	GCGTCACCATCAGCATAGC	delta-8 fatty acid desaturase reverse
DELTA8FADR2	CGGAAGTCCTCAGCAACCT	delta-8 fatty acid desaturase reverse
PRKF	TTCCAAGCCAATCTACAACCAT	Phosphoribulokinase forward
PRKR1	CGAATCTGAGACGGAAGAACTT	Phosphoribulokinase reverse
PRKR2	GCAACACGGTCATCCAACA	Phosphoribulokinase reverse
ThioesteraseF	AGCAATCGGCAGGAGCAA	Thioesterase forward
ThioesteraseR1	ACAGGCAAGCGGCATCT	Thioesterase reverse
ThioesteraseR2	TGGAACAACGGCAGCAATG	Thioesterase reverse
KAR1-F	AAGCCATCGCTGTTGGATTG	β-Ketoacyl-ACP reductase1 forward
KAR1-R1	ACCGCATCCCAATCAGCAA	β-Ketoacyl-ACP reductase1 reverse
KAR1-R2	CTGATTCGTAATGTCTGCCTGAA	β-Ketoacyl-ACP reductase1 forward
KAR2-F1	TTCAACGACGCAGCCTGTA	β-Ketoacyl-ACP reductase2 forward
KAR2-F2	CCGTCAACTCGCTCATTGTT	β-Ketoacyl-ACP reductase2 forward
KAR2-R	GCCTCTTCTGTCATCTTGTCAA	β-Ketoacyl-ACP reductase2 reverse
KAR3-F1	CGGTGCTTCCAGAGGTGTT	β-Ketoacyl-ACP reductase3 forward
KAR3-F2	GGAGCCTCTGTGGTTGTCA	β-Ketoacyl-ACP reductase3 forward
KAR3-R1	TTCATCTGCTTGTCCATCATCTC	β-Ketoacyl-ACP reductase3 reverse
FAD-F	AACATTCTAACGCCACACCAT	FAD family gene forward
FAD-R1	ACTGCCACATTGAGACTCCT	FAD family gene reverse
FAD-R2	ACAGTCGGTGCGTTGTAATC	FAD family gene reverse
GPD-F	CTTGGCAGGCTTCTGTAATCG	glycerol-3-phosphate dehydrogenase forward
GPD-R1	GCTTCGTGGATCGTGAGGA	glycerol-3-phosphate dehydrogenase reverse

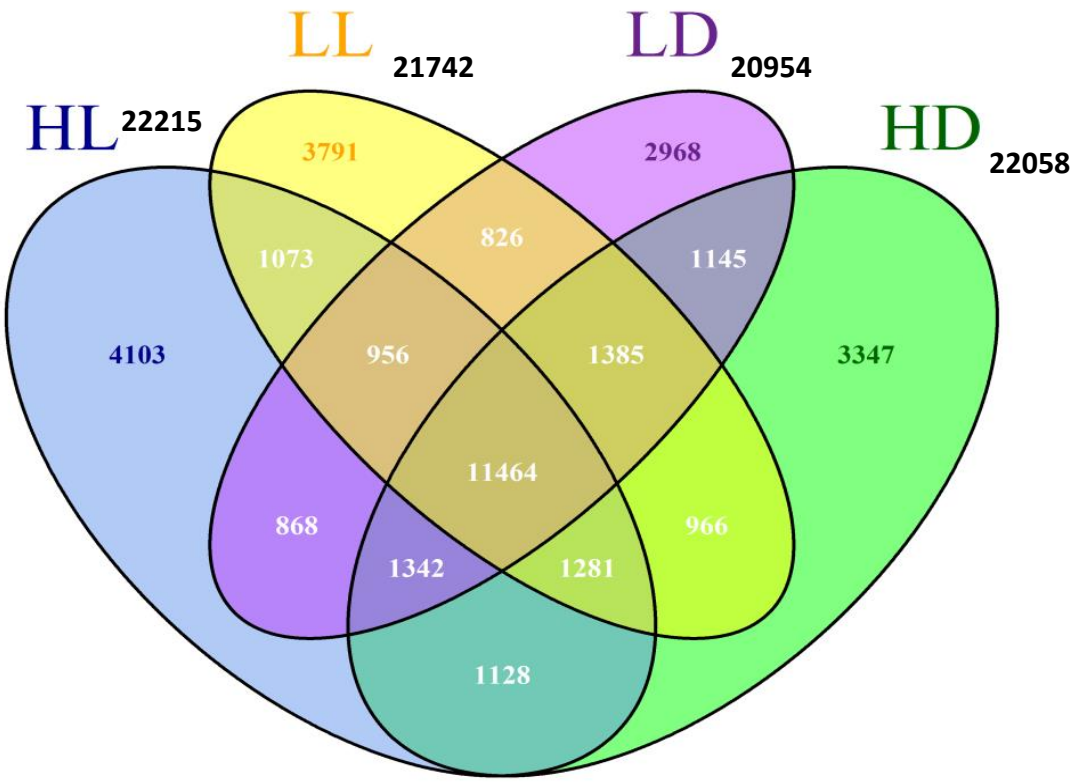
GPD-R2	ATCACCACAACATCGTACTCCT	glycerol-3-phosphate dehydrogenase reverse
GPAT-F	CCAGATGAGGCTTCTTCTACAA	glycerol-3-phosphate acyltransferase forward
GPAT-R1	CAACAGCGATGGCGTTCAT	glycerol-3-phosphate acyltransferase reverse
GPAT-R2	TCACCGCATCCACGAACT	glycerol-3-phosphate acyltransferase reverse
PEPCASE-F	GCCTGTCTTGAATGACTCCTT	Phosphoenolpyruvate carboxylase forward
PEPCASE-R1	GCATACTTGACCAGCAGTGA	Phosphoenolpyruvate carboxylase reverse
PEPCASE-R2	CCGTGGTTCTGGATCAGGTT	Phosphoenolpyruvate carboxylase reverse
PEPCK1-F1	AAGGCTCTGGCGACATTAAG	Phosphoenolpyruvate carboxykinase1 forward
PEPCK1-F2	TGGCGACATTAAGAACAGCATA	Phosphoenolpyruvate carboxykinase1 forward
PEPCK1-R	TCTGGTTGAGAGGAATGAGTGT	Phosphoenolpyruvate carboxykinase1 reverse
PEPCK2-F	TGGTATTGATTGAGCACAACAATG	Phosphoenolpyruvate carboxykinase2 forward
PEPCK2-R1	GAGGTCTTGATTGTGGATTGGAAT	Phosphoenolpyruvate carboxykinase2 reverse
PEPCK2-R2	TACTTGCGGATCTCGGTGATA	Phosphoenolpyruvate carboxykinase2 reverse
PEPCK3-F	GCAACGGACGATGTGAACC	Phosphoenolpyruvate carboxykinase3 forward
PEPCK3-R1	CGCATGTGAGGAACCAACTG	Phosphoenolpyruvate carboxykinase2 reverse
PEPCK3-R2	GTCTGATTGCGGTCCAACTTAT	Phosphoenolpyruvate carboxykinase2 reverse
PPDK-F1	CTTATCTCCCACAAATGACTACCC	phosphate dikinase forward
PPDK-F2	ACTGGTTCATCTTCGCCTCAT	phosphate dikinase forward
PPDK-R	TCAGCCAAGTTCGCACCTT	phosphate dikinase reverse
EF-1aF	GCACCTCTCACCATCTTACAAT	elongation factor 1 alpha forward
EF-1aR1	GTTCCGCCTTCAGCTTGTC	elongation factor 1 alpha reverse
EF-1aR2	TTTCCTTTCGCTTTCCCTTCTT	elongation factor 1 alpha reverse
TubulinF	CCTTGGYCAGGCTGGTATCC	tubulin forward
TubulinR1	CGGCGTTGAAGCACAAGAA	tubulin reverse
TubulinR2	AGTTGTTGGCGGCATCCT	tubulin reverse
ActinF	AAGCGAAGCGTGGTATCTTG	actin forward
ActinR1	ACATAATCTGCGTCATCTTCTCC	actin reverse
ActinR2	GGTTCATAGGAGCCTCAGTGA	actin reverse

649

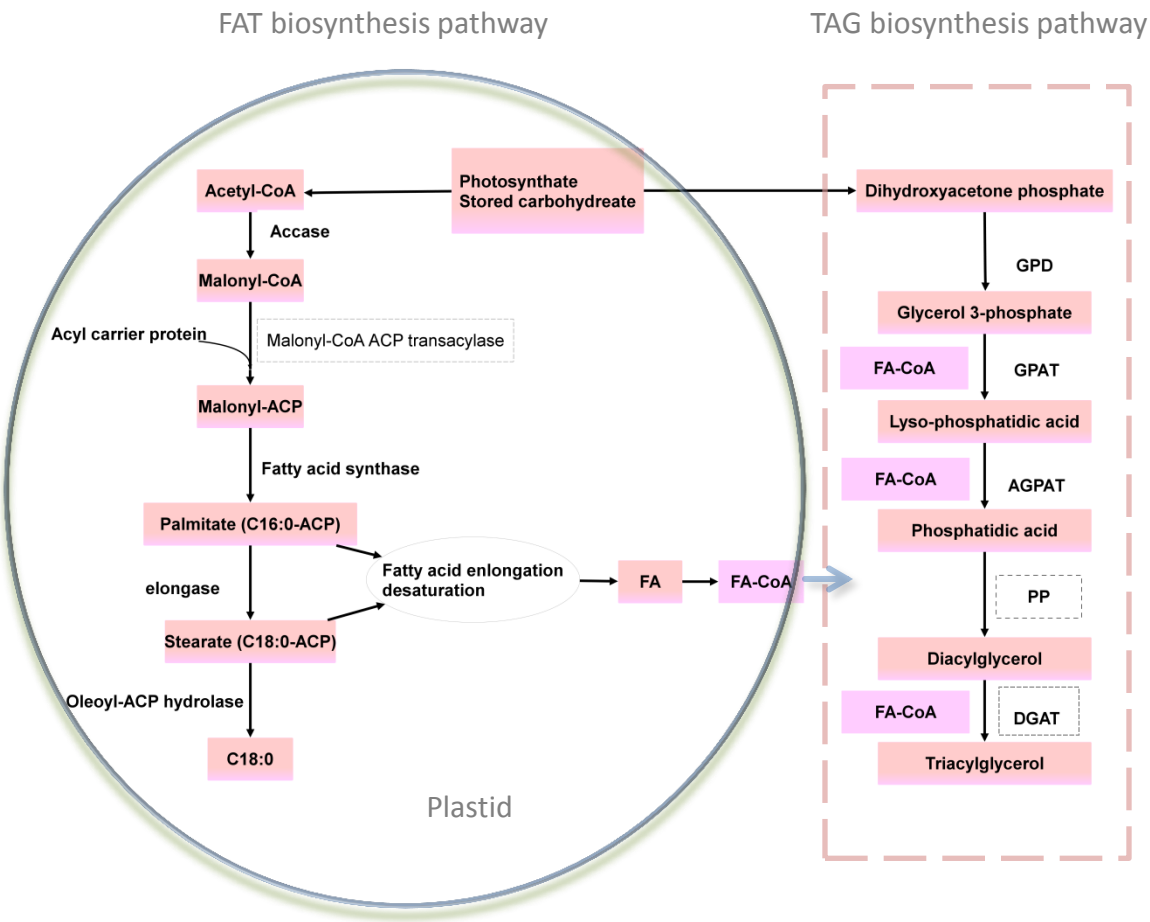




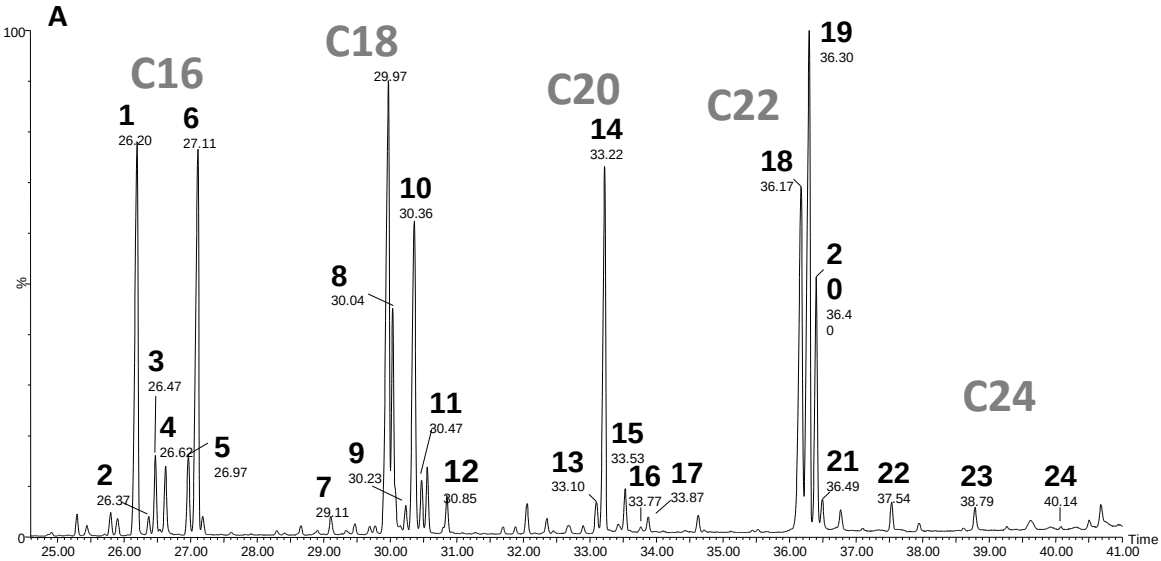
Kuo et al. Figure 1

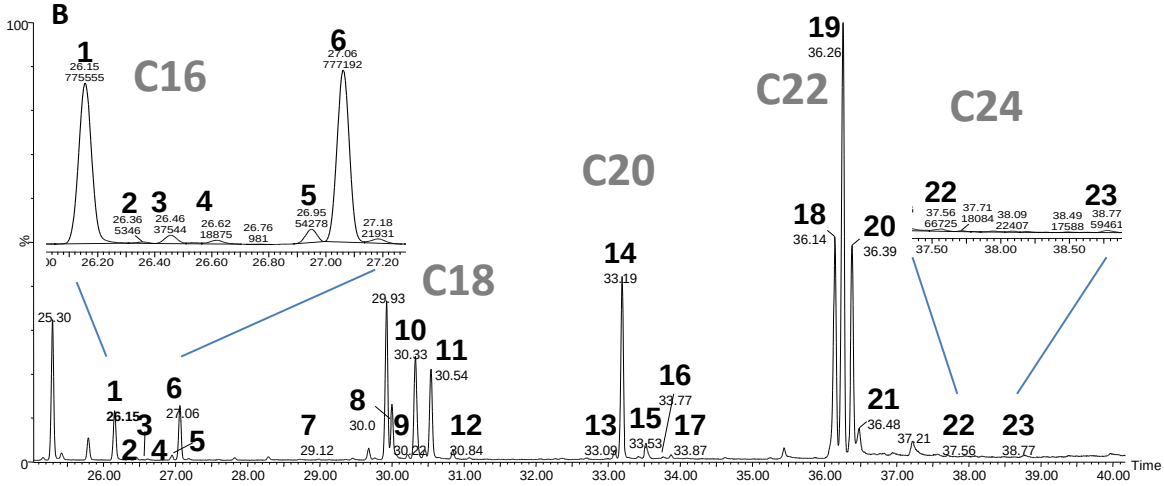


Kuo et al. Figure 2

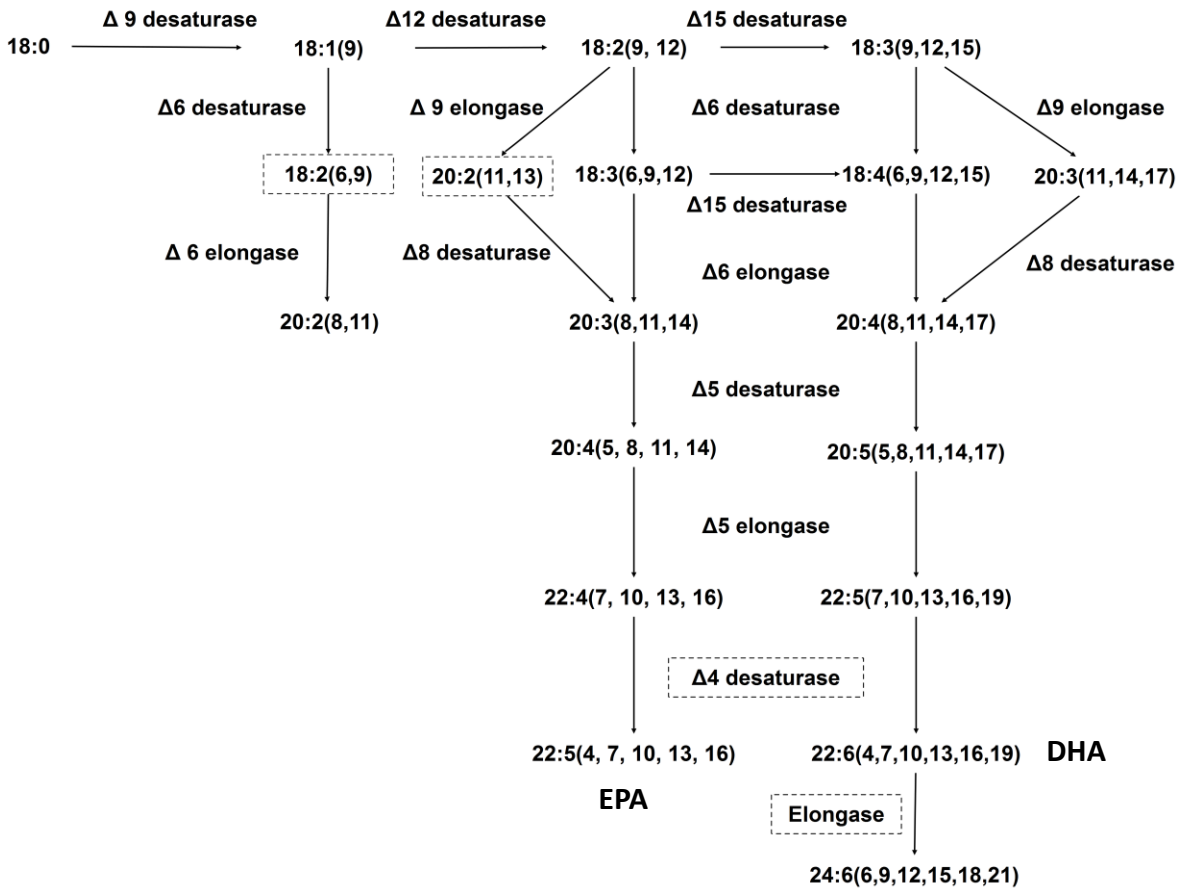


Kuo et al. Figure 3

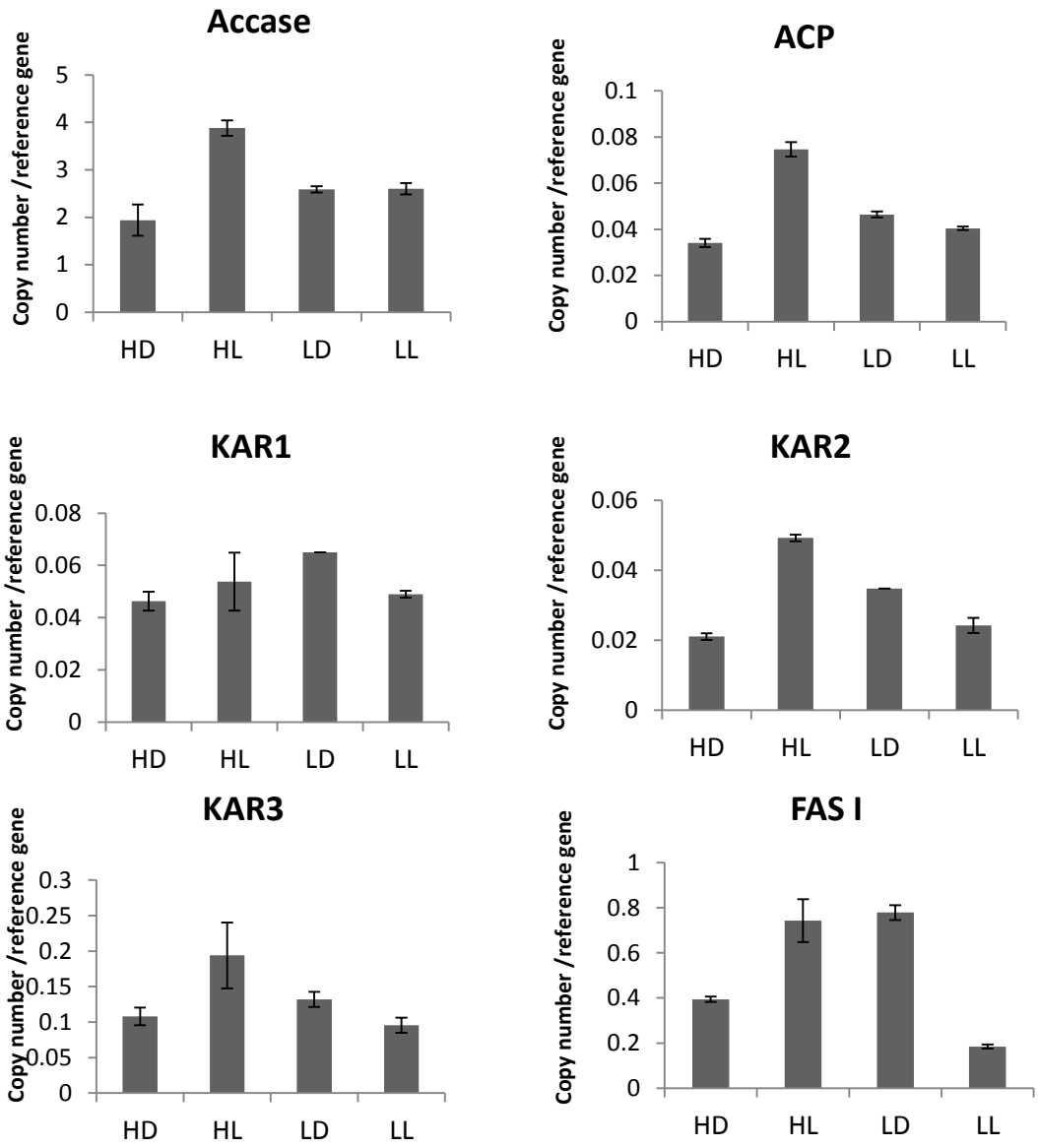




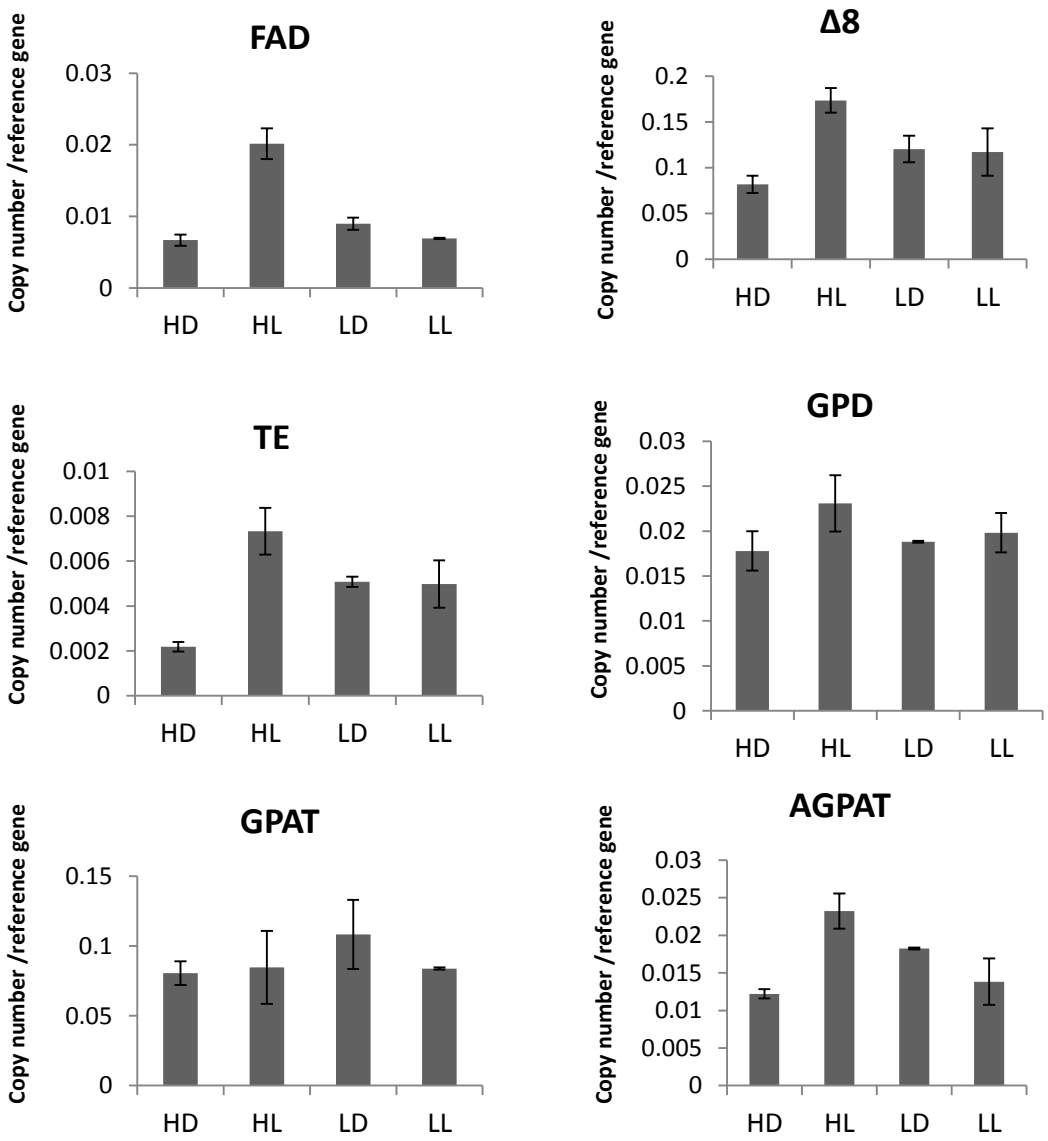
Kuo et al. Figure 4



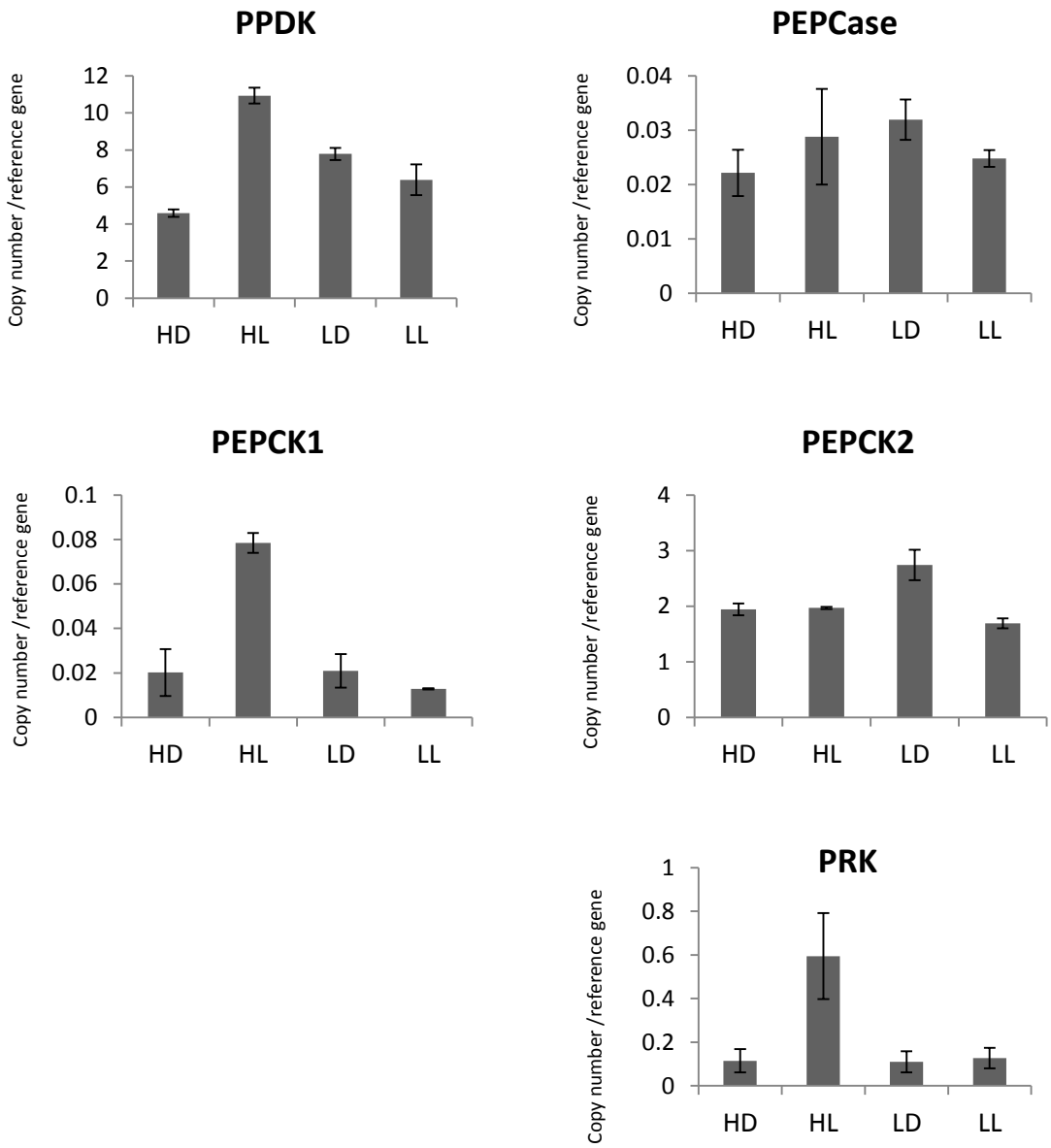
Kuo et al. Figure 5



Kuo et al. Figure 6



Kuo et al. Figure 7



Kuo et al. Figure 8