Royal Netherlands Institute for Sea Research

This is a preprint of:

Balzano, S.; Villanueva, L.; de Bar, M.; Sahonero Canavesi, D.X; Yildiz, C.; Engelmann, J.C.; Maréchal, E.; Lupette, J.; Sinninghe Damsté, J.S. & Schouten, S (2019). Biosynthesis of long chain alkyl diols and long chain alkenols in *Nannochloropsis* spp. (Eustigmatophyceae). *Plant and cell physiology*, 60, 1666-1682

Published version: https://dx.doi.org/10.1093/pcp/pcz078

NIOZ Repository: http://www.vliz.be/nl/imis?module=ref&refid=310966

[Article begins on next page]

The NIOZ Repository gives free access to the digital collection of the work of the Royal Netherlands Institute for Sea Research. This archive is managed according to the principles of the Open Access Movement, and the Open Archive Initiative. Each publication should be cited to its original source - please use the reference as presented.
When using parts of, or whole publications in your own work, permission from the author(s) or copyright holder(s) is always needed.

Title: **Biosynthesis of Long Chain Alkyl Diols and Long Chain Alkenols in *Nannochloropsis* spp. (Eustigmatophyceae)**

Short title: Diol and Alkenol Biosynthesis in *Nannochloropsis*

Corresponding author: Sergio Balzano, Department of Marine Microbiology and Biogeochemistry (MMB), NIOZ Royal Netherlands Institute for Sea Research, and Utrecht University, P.O. Box 59, 1790 AB Den Burg, The Netherlands

Present address: Stazione Zoologica Anton Dohrn, Istituto Nazionale di Biologia Ecologia e Biotecnologie Marine, Villa Comunale Napoli, 80121, Italy

email: sergio.balzano@szn.it

Subject areas: (1) proteins, enzymes and metabolism and (2) regulation of gene expression

Number of black and white figures: 3
Number of colour figures: 2
Number of tables: 2

Supplemental Material:

1) *Text:* Supplemental Information
2) *Number of supplemental tables:* 13
3) *Number of supplemental figures:* 9

For Peer Review

Title: **Biosynthesis of Long Chain Alkyl Diols and Long Chain Alkenols in *Nannochloropsis* spp. (Eustigmatophyceae)**

Short title: Diol and Alkenol Biosynthesis in *Nannochloropsis*

Authors: Sergio Balzano[1,2*], Laura Villanueva[1,3], Marijke de Bar[1], Diana X Sahonero Canavesi[1], Caglar Yildiz[1], Julia C. Engelmann[1], Eric Marechal[4], Josselin Lupette[4,5], Jaap S. Sinninghe Damsté[1,3], and Stefan Schouten[1,3]

[1]Department of Marine Microbiology and Biogeochemistry (MMB), NIOZ Royal Netherlands Institute for Sea Research, and Utrecht University, P.O. Box 59, 1790 AB Den Burg, The Netherlands

[2] Present address: Stazione Zoologica Anton Dohrn, Istituto Nazionale di Biologia Ecologia e Biotecnologie Marine, Villa Comunale Napoli, 80121, Italy

[3]Utrecht University, Faculty of Geosciences, Department of Earth Sciences, P.O. Box 80.021, 3508 TA Utrecht, The Netherlands

[4]Laboratoire de Physiologie Cellulaire et Végétale, Unité mixte de recherche 5168 CNRS, CEA, INRA, Univ. Grenoble Alpes, CEA Grenoble, 17 avenue des Martyrs, 38000 Grenoble, France

[5] Present address: MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824, USA

*corresponding author: sergio.balzano@szn.it

For Peer Review

Abstract

We investigated potential biosynthetic pathways of long chain alkenols (LCAs), long chain alkyl diols (LCDs), and long chain hydroxy fatty acids (LCHFAs) in *Nannochloropsis oceanica* and *Nannochloropsis gaditana,* by combining culturing experiments with genomic and transcriptomic analyses. Incubation of *Nannochloropsis* spp. in the dark for one week led to significant increases in the cellular concentrations of LCAs and LCDs in both species. Consistently, ^{13}C-labeled substrate experiments confirmed that both LCA and LCD were actively produced in the dark from $C_{14–18}$ fatty acids by either condensation or elongation/hydroxylation, although no enzymatic evidence was found for the former pathway. *Nannochloropsis* spp. did, however, contain (1) multiple polyketide synthases (PKSs) including one type (PKS Clade II) that might catalyse incomplete fatty acid elongations leading to the formation of 3-OH-fatty acids, (2) 3-hydroxyacyl dehydratases (HADs), which can possibly form Δ^2/Δ^3 monounsaturated fatty acids, and (3) fatty acid elongases (FAEs) that could elongate 3-OH-fatty acids and Δ^2/Δ^3 monounsaturated fatty acids to longer products. The enzymes responsible for reduction of the long chain fatty acids to LCDs and LCAs are, however, unclear. A putative wax ester synthase/acyl coenzyme A (acyl-CoA):diacylglycerol acyltransferase (WS-DGAT) is likely to be involved in the esterification of LCAs and LCDs in the cell wall. Our data thus provide useful insights in predicting the biosynthetic pathways of LCAs and LCDs in phytoplankton suggesting a key role of FAE and PKS enzymes.

Keywords: *Nannochloropsis,* diols, alkenols, hydroxylated fatty acids, polyketide synthase, bioproduct

For Peer Review

INTRODUCTION

Since phytoplankton do not require clean water for their growth and can encompass high levels of biomass productivities per area compared to terrestrial plants (Chisti, 2007), microalgal mass culturing could contribute to the sustainable production of chemical products of interest for the biotechnological industry. *Nannochloropsis* species (Eustigmatophyceae) are considered among the most suitable candidates for biofuel development because of their high growth rate and lipid content with respect to other phytoplankters (Rodolfi et al. 2009). Specifically, both free and ester-bound fatty acids from *Nannochloropsis* spp. are currently considered as potential candidates for biodiesel production (Chen et al. 2012). The genomes of several *Nannochloropsis* species have been sequenced allowing the identification of major metabolic pathways for lipid biosynthesis (Radakovits et al. 2012, Vieler et al. 2012, Corteggiani Carpinelli et al. 2014, Wang et al. 2014). Transcriptomic analyses of *Nannochloropsis* cultures have contributed to the identification of the genes potentially involved in different lipid pathways including the biosynthesis of polyunsaturated fatty acids (PUFAs) (Vieler et al. 2012) and triacylglycerols (Radakovits et al. 2012, Li et al. 2014).

Besides regular fatty acids and PUFAs, Eustigmatophyceae also produce long chain hydroxy fatty acids (LCHFAs), with an alkyl chain of 28–32 carbon and a hydroxyl group at a mid-chain position (Volkman et al. 1992, Gelin et al. 1997a). Plant hydroxy fatty acids, such as the 12-OH $C_{18:1}$ from *Ricinus communis* and 14-OH $C_{20:1}$ from *Physaria fendleri* were previously shown to act as lubricants when added to reference diesel fuel (Goodrum and Geller 2005). This suggests that the quality of *Nannochloropsis* oils for biodiesel development might also be improved if tiny amounts of LCHFAs ($\leq$ 1%) are present in the lipid extract to be used for methanol transesterification. LCHFAs possess a combustion enthalpy slightly higher than that of $C_{14\text{-}18}$ fatty acids (Table 1) and thus such addition would

For Peer Review

not affect the energy yield. Furthermore, hydroxylated aliphatic compounds are also under investigation for polymer development (Sharma and Kundu 2006, Mutlu and Meier 2010), hence their diversity and biosynthetic pathways have been partially elucidated (Buschhaus et al. 2013, Busta and Jetter 2018, Li et al. 2018).

Nannochloropsis spp. produce two other classes of hydroxylated compounds related to LCHFAs, in which the terminal carboxylic group is replaced with an alcohol group, i.e. long chain alkyl diols (LCDs) and long chain alkenols (LCAs). LCAs differ from LCDs because of an intermediate double bond instead of the secondary alcohol group. Similar to bifunctional aliphatic compounds from plants, LCDs might also be of interest to the chemical and polymer industries. For example polyricinoleate diol, prepared from 12-OH-C_{18} fatty acid (ricinoleic acid) was tested for the synthesis of polyurethane, revealing faster degradation times than petrochemical polyurethanes (Petrovic et al. 2010). Polyurethane synthesis requires highly hydroxylated compounds as starters and there is a common interest in introducing new hydroxyl groups in naturally occurring fatty acids as well as in screening natural products with a high number of hydroxyl groups. (Petrovic 2008). LCDs in *Nannochloropsis* might thus be of interest for the polymer industry. However, it is crucial to identify the culturing conditions affecting the cellular concentrations of LCHFAs, LCAs, and LCDs as well as their biosynthetic pathways.

While the total lipid content of microalgae typically increases during the stationary phase of their growth (Dunstan et al. 1993), as well as under high salinity (Martinez-Roldan et al. 2014) or nitrogen deprivation (Pal et al. 2011), such culture manipulations do not increase the cellular concentration of LCAs, LCDs, and LCHFAs significantly (Balzano et al. 2017), suggesting that these compounds are unlikely to serve as storage lipids. Instead, their decrease under hydrogen peroxide-driven oxidative stress suggests a protective role for LCAs, LCDs, and LCHFAs in *Nannochloropsis* cells (Balzano et al. 2017). Finally, LCAs and LCDs are

For Peer Review

thought to occur in the outer layer of the *Nannochloropsis* cell wall (Gelin et al. 1997b, Scholz et al. 2014, Zhang and Volkman 2017, Volkman 2018) as part of a polymer, termed algaenan.

Since LCAs, LCDs, and LCHFAs are structurally related among each other in terms of carbon number and position of the functional groups, common biosynthetic pathways have been long hypothesised for *Nannochloropsis* species (Volkman et al. 1992, Versteegh et al. 1997, Gelin et al. 1997a). LCHFAs were suggested to originate from the elongation or condensation of $C_{14–18}$ fatty acids (Gelin et al. 1997a, Scholz et al. 2014) and this was confirmed by the positive correlation recently found between the cellular concentrations of $C_{14–16}$ fatty acids and two LCHFAs (13-hydroxy $C_{30:0}$ and 15-hydroxy $C_{32:0}$ fatty acids), respectively, in three *Nannochloropsis* spp. (Balzano et al. 2017). However, the biosynthetic pathways of LCHFAs, LCAs, and LCDs are not fully understood and the enzymes potentially involved in the process are unknown.

Here, (1) we analysed the genomes from different *Nannochloropsis* spp. to identify the enzymes potentially involved in the biosynthesis of LCAs and LCDs, (2) searched for culturing conditions promoting the accumulation of LCAs, LCDs, and LCHFAs in *Nannochloropsis oceanica* and *Nannochloropsis gaditana,* and (3) performed transcriptomic analyses to identify genes potentially involved in their biosynthesis.

For Peer Review

RESULTS AND DISCUSSION

Dark incubation enhances LCA and LCD concentrations

To identify potential genes for LCA, LCD and LCHFA synthesis we first investigated conditions which stimulated the production of these compounds. A previous study showed that manipulations that typically promote the accumulation of storage lipids, such as nitrogen deprivation (< 1 µM nitrate), exposure to high light irradiance (300 µE m^{-2} s^{-1}) and culturing at high (50 g kg^{-1}) salinity, did not increase the cellular concentrations of LCAs, LCDs, and LCHFAs significantly (Balzano et al. 2017). Unfavourable environmental conditions in the marine water column, such as prolonged exposure to dark conditions, are known to trigger the formation of resting stages in phytoplankton (McQuoid and Hobson 1996) for cell protection purposes. Resting forms of *Nannochloropsis limnetica* exhibit a thicker cell wall compared to active cells (Fietz et al. 2005), and might thus contain higher amounts of LCAs and LCDs. We therefore attempted to enhance the production of these lipids by incubating living cultures of *N. oceanica* and *N. gaditana* in the dark for one week.

Before the incubation in the dark, cells from both *N. oceanica* and *N. gaditana* exhibited growth rates (Fig. 1A-1B) comparable to those reported previously (Balzano et al. 2017). Cell abundance decreased by ca. 50% in both species over one week under dark conditions. Cells were observed under transmission electron microscopy (TEM) and the outer cell wall, which contains LCAs and LCDs (Scholz et al. 2014), was extremely thin (approximately 10 nm) for the resolution of the instrument used and its thickness could not be measured reliably (Supplemental Fig. S1).

For Peer Review

Both *N. oceanica* and *N. gaditana* exhibited increased cellular concentrations of LCAs and LCDs after dark incubation. Cells were enumerated by flow cytometry before filtration and we did not observe dead cells or large debris (i.e. particles with comparable forward scatter but lower chlorophyll fluorescence compared to ordinary *Nannochloropsis* cells) in the cytograms of cells harvested from the dark treatment. This suggests that most of the material filtered contained viable cells, or at least dead cells with intact chloroplasts, while debris were likely to be smaller in size (i.e. with a forward scatter comparable to the background noise of the instrument) and were not retained by the filters. In spite of the significant decline in viable cells, observed in our dark incubations, the LCAs and LCDs analysed here are thus likely derived from intact cells. The cellular concentration of $C_{32:1}$ alkenol increased by almost one order of magnitude (3.5 ± 0.5 to 22 ± 3 fg cell^{-1}), while $C_{32:1}$ and $C_{32:0}$ diols tripled in concentration reaching 26 ± 4, 37 ± 9, and 35 ± 10 fg cell^{-1}, respectively (Fig. 1C), and the $C_{30:0}$ diol nearly doubled in *N. oceanica* at the end of the dark incubation. Similarly, in *N. gaditana*, $C_{32:2}$ and $C_{32:1}$ alkenols increased from 2.9 ± 0.8 to 7.8 ± 1.7 fg cell^{-1} and from 4.7 ± 1.0 to 14 ± 2.0 fg cell^{-1}, respectively, the $C_{30:0}$ diols doubled (7.4 ± 1.4 to 16 ± 2.5 fg cell^{-1}) and the $C_{32:1}$ and $C_{32:0}$ diols almost tripled (3.3 ± 1.3 to 7.0 ± 2.5 and 9.3 ± 2.9 to 20 ± 4, respectively, Fig. 1D) in concentration. In contrast with LCAs and LCDs, the concentration of LCHFAs dropped dramatically, with the 13-OH-$C_{30:0}$ fatty acid decreasing from 7.8 ± 1.0 and 5.5 ± 0.4 fg cell^{-1} for *N. oceanica* and *N. gaditana,* respectively, to values below the detection limit. Similarly, the 15-OH-$C_{32:0}$ fatty acid decreased from 2.5 ± 0.4 and 1.7 ± 0.4 fg cell^{-1}, for *N. oceanica* and *N. gaditana,* respectively, to below the detection limit (Figs 1C-1D). Furthermore, the concentration of $C_{16:0}$ fatty acid decreased by nearly half for both species (Figs. 1E-F). In contrast, the other $C_{14–20}$ fatty acids followed different dynamics with the concentration of $C_{16:1}$ fatty acid decreasing under dark conditions and that of the $C_{20:5}$ PUFA increasing for *N. gaditana*, whereas no significant changes were observed in *N. oceanica* (Fig.

For Peer Review

1C-D). The decrease in $C_{16:0}$ fatty acid under dark conditions is likely due to the consumption of storage lipids necessary to sustain cell metabolism. Storage lipids such as triacylglycerols are typically dominated by the $C_{16:0}$ fatty acid in *Nannochloropsis* spp. (Alboresi et al. 2016).

The incubation under dark conditions for one week thus promoted a substantial increase of the cellular concentrations of LCAs and LCDs (Fig. 1C-D), which is the first culture condition ever described shown to trigger an increase of the LCAs and LCDs content in *Nannochloropsis* spp.. Seemingly, prolonged light deprivation affects the biosynthetic pathways of LCHFAs, LCAs and LCDs resulting in the complete removal of LCHFAs and an accumulation of both LCAs and LCDs. Since dark conditions are thought to promote the formation of resting stages, which can result in thicker cell walls as shown for *N. limnetica* (Fietz et al., 2005), the dramatic decline in LCHFAs under dark conditions strongly suggests that these lipids are unlikely to be present in the cell wall, but rather form the precursors of LCAs and LCDs. In turn, the LCHFAs might derive from $C_{14–18}$ fatty acids (Volkman et al. 1992, Gelin et al. 1997b, Balzano et al. 2017).

Further clues were obtained by determination of the double bond positions of unsaturated LCAs and LCDs in a replicate from the dark treatment of *N. oceanica* CCMP1779 using dimethyl disulfide derivatisation. Consistent with previous findings (Gelin et al. 1997b), the double bond in LCAs occurs at the same position as that of the mid-chain alcohol group in the corresponding LCDs and LCHFAs (Supplemental Fig. S2). For example, the $C_{32:1}$ alkenol mostly consists of two isomers with double bonds at Δ^{14} and Δ^{15} which correspond to the position of the mid-chain alcohol group in the 15-OH $C_{32:0}$ fatty acid and the $C_{32:0}$ 1,15 diol. Moreover, the $C_{32:2}$ alkenol has a second double bond at the same position (Δ^{27}) as that of the $C_{32:1}$ diol (Supplemental Fig. S2). The excellent correspondence between the double bond position of monounsaturated alkenols and the position of the intermediate hydroxyl group in diols and LCHFAs with the same carbon number, as well as the presence of

For Peer Review

a double bond at Δ^{27} in both diunsaturated alkenols and monounsaturated diols, clearly indicates common biosynthetic pathways for LCAs, LCDs, and LCHFAs (Supplemental Fig. S2).

^{13}C-labelling indicates active biosynthesis of LCAs and LCDs in the dark

We incubated *N. oceanica* CCMP1779 with ^{13}C-[2]-acetate under alternating 12/12 dark/light conditions to investigate the biosynthetic relationships among $C_{14\text{-}18}$ fatty acids, LCHFAs, LCAs, and LCDs (Figs 2A-2B). All these lipids were significantly labelled with ^{13}C (Fig. 2C), with atomic ^{13}C percentages ranging from 4.2% (cholesterol) to 15.2% ($C_{18:1}$ fatty acid). In a second experiment, cells were initially cultured under alternating 12/12 dark/light conditions, and during exponential growth, ^{13}C-[2]-acetate was then added to the culture which was subsequently incubated under dark conditions for one week (Fig. 2B). As expected, the enrichment levels of lipids observed under dark conditions were significantly lower than those found under dark/light conditions due to the absence of growth. The $C_{16:1}$ fatty acid showed no incorporation of ^{13}C and the $C_{14:0}$, $C_{16:0}$ and the $C_{20:5}$ were only slightly labelled (1.3%, 1.2%, and 1.2%, respectively). However, the $C_{18:1}$ fatty acid (3.2%), the $C_{32:0}$ diols (3.0%) and to a lesser extent the other LCDs and the LCAs (1.7–1.8%) showed substantial incorporation of ^{13}C label (Fig. 2D). Thus, our labelling experiments show that during regular growth under alternating dark/light conditions, the ^{13}C-[2]-acetate was taken for *de novo* synthesis of $C_{16:0}$ fatty acids, as well as for the formation of LCHFAs, LCAs, and LCDs. In contrast, when labelled sodium acetate was supplied prior to incubation in the dark, *de novo* fatty acid synthesis was likely to be insignificant as the cell growth in the dark was nearly negligible (Fig. 2B); however, the lack of detection of $C_{18:0}$ fatty acid along with the

For Peer Review

high atomic ^{13}C percentage measured for the $C_{18:1}$ fatty acid (Fig. 2D) suggests that an active synthesis of $C_{18:1}$, probably via $C_{16:0}$ elongation to $C_{18:0}$ followed by desaturation to $C_{18:1}$, was taking place. Importantly, both LCAs and LCDs were actively synthesised under dark conditions. The greater ^{13}C content of LCAs and LCDs compared to $C_{14\text{-}16}$ under dark conditions suggests that if the biosynthesis of LCD and LCA took place by condensation of two $C_{14\text{-}16}$ fatty acids, such process would have rapidly taken up all ^{13}C labelled $C_{14\text{-}16}$ fatty acids. Alternatively, and perhaps more likely, unlabelled $C_{14\text{-}16}$ fatty acids were elongated with ^{13}C labelled sodium acetate to LCHFAs and subsequent reduction may have resulted in the formation of ^{13}C labelled LCAs and LCDs (Fig. 2D).

Transcriptomic analyses and hypothetical biosynthetic pathway

To determine which genes were up-regulated in the dark incubation experiments, and thus potentially involved in LCD and LCA biosynthesis, we extracted RNA and sequenced the transcriptomes of *Nannochloropsis* cultures harvested at the end of the experiments. We compared the gene expression level of *Nannochloropsis* spp. from the dark treatment (i.e. treatment leading to high concentrations of LCAs and LCDs) with the dark/light control. Overall we mapped 10,043 genes from *N. oceanica* CCMP1779 against the reference genome from the same strain (Vieler et al. 2012) and 11,222 genes from *N. gaditana* CCMP526 (Table 2) against the reference genome from the strain *N. gaditana* B31 (Corteggiani Carpinelli et al. 2014). In addition, we also carried out a functional analyses of all the proteins predicted from the genomes of *N. oceanica* CCMP1779 and *N. gaditana* B31 using Interproscan (Jones et al. 2014) to identify putative catalytic domains for the elongation and

For Peer Review

hydroxylation of fatty acids, the dehydration of secondary alcohols, the reduction of their carboxylic groups, and the formation of ethers and esters.

About 60% of the genes from *N. oceanica* and 74% of the genes from *N. gaditana* did not change significantly in expression during the dark incubation (Table 2). *Nannochloropsis oceanica* exhibited a comparable number of up-regulated and down-regulated genes (1,950 and 2,067, respectively) whereas 1,855 genes from *N. gaditana* were down-regulated and only 955 up-regulated (Table 2). The expression changes of all the genes from both *Nannochloropsis* species are shown in details in Supplemental Table 1. The results of our dark incubation experiments with and without stable isotope labelling (Figs 1–2), as well as previous findings (Volkman et al. 1992, Gelin et al. 1997a, Balzano et al. 2017, Volkman 2018), suggest that LCHFAs originate from $C_{14–18}$ fatty acids either by condensation of two fatty acids or elongation/in-chain hydroxylation. Subsequently LCHFAs are likely to be reduced to form LCDs. Similarly, LCAs might derive from the dehydration of the secondary alcohol groups in LCDs or LCHFAs.

We thus focused on genes potentially coding enzymes that catalyse (1) the condensation of two C_{14-18} fatty acids, (2) the elongation and (3) the in-chain hydroxylation of fatty acids ≥ C_{16}, (4) the reduction of fatty acids to alcohols, the (5) dehydration of secondary alcohol groups, and (6) the formation of esters. We searched for these genes in *Nannochloropsis* genomes and compared their expression levels in the dark treatment with those found for the control treatment. We also searched publicly available genomes from other *Nannochloropsis* spp. to identify genes homologous to those potentially involved in the biosynthetic processes hypothesised here for *N. gaditana* and *N. oceanica*.

Condensation of two C_{14-18} fatty acids

For Peer Review

The condensation of two fatty acids to form longer products has been rarely reported in literature. γ-Proteobacteria from the genera *Xanthomonas* and *Photorabdus* can perform head-to-head condensation of fatty acids mediated by oleA and photopyrone synthase enzymes, respectively (Kresovic et al. 2015, Christenson et al. 2017). Similarity analyses of oleA and photopyrone synthase sequences against *N. oceanica* and *N. gaditana* proteins did not yield significant results (Supplemental Table S2) suggesting *Nannochloropsis* spp. do not contain oleA or photopyrone homologues. Moreover, a head-to-head condensation would produce a mid-chain functionalised intermediate which would still require an additional ω-functionalisation to yield a primary/secondary aliphatic compound such as LCDs, LCAs, or LCHFAs.

Acidobacteria are known to produce a C_{30} 13,16-dimethyl dicarboxylic acid from a tail-to-tail condensation of two C15 *iso* fatty acids (Sinninghe Damsté et al. 2011) but the enzymes involved in such process are unknown and a tail-to-tail condensation would yield an intermediate functionalised (i.e. carboxylic acids) on both ends and one of these ends would therefore require to be fully reduced to a methyl group. Both head-to-head and tail-to-tail condensations would thus form intermediates which need a further functionalisation or reduction step to yield the LHCFAs observed in *Nannochloropsis* spp.. Long chain aliphatic compounds resulting from head-to-head (i.e. mid-chain functionalised) or tail-to-tail (functionalised on both ends) condensation have never been detected in *Nannochloropsis*.

The condensation of the carboxylic end of a fatty acid with the aliphatic end of another fatty acid (head-to-tail condensation) would instead require fewer reaction steps and the resulting biosynthetic pathway appears thus to be less energy demanding compared to both head-to-head and tail-to-tail condensations. For example, the reaction between the carboxylic end of a C_{14} fatty acid with the methyl end of a C_{18} fatty acid would yield a C_{32} product

For Peer Review

functionalised on the first and the 15th carbon such as the 15-OH-C_{32} fatty acid, the C_{32} 1,15 diol, the 15-ene-$C_{32:1}$ alkenol and the 15-27-ene-$C_{32:2}$ alkenol. Similarly, the condensation between a C_{12} and a C_{18} fatty acids would lead to C_{30} compounds with a secondary functionalisation on the 13th carbon as well as a terminal carboxylic group. Although this pathway cannot be fully discarded we did not find any evidence reported in literature for such a biosynthetic process.

Enzymes responsible for chain elongation in *Nannochloropsis*

Fatty acid elongation is based on stepwise additions of 2 carbon units to the growing acyl-coenzyme A (CoA) chain (Leonard et al. 2004), with each addition consisting of the (1) condensation of the acyl CoA with a malonyl group to form a 3-ketoacyl-CoA, (2) reduction of 3-ketoacyl-CoA to 3-hydroxyacyl-CoA, (3) dehydration to enoyl-CoA and (4) reduction to an elongated acyl chain (Leonard et al. 2004). While 3-ketoacyl-CoA synthases (KCS) are typically substrate specific (Leonard et al. 2004, Haslam and Kunst 2013), the other 3 enzymes required for the elongation are known to have a broad substrate preference being able to accept 3-ketoacyl, 3-hydroxyacyl, or 3-enoyl units of different lengths. Enzymes belonging to two different families, the elongation proteins (ELO) and the fatty acid elongases (FAE) possess the KCS domain (Leonard et al. 2004, Haslam and Kunst 2013). In addition, the polyketide synthases family (PKSs) consists in proteins known to contain ketoacyl acyl carrier protein (ACP) synthase (KAS) and can also accept $C_{16–18}$ fatty acids as substrates for elongation (Staunton and Weissman 2001).

Δ0-ELOs as elongators of fatty acids

For Peer Review

Seven ELOs that accept monounsaturated or saturated fatty acids as substrates (Δ0-ELO) have been previously identified in *N. gaditana* CCMP526 (Dolch et al. 2017), and we found eight Δ0-ELOs in *N. oceanica* by similarity searches. Four Δ0-ELOs from *N. gaditana* and two Δ0-ELOs from *N. oceanica* can potentially accept fatty acids containing up to 28 carbons as substrates (Supplemental Information) as predicted by comparing their secondary structure (Supplemental Fig. S3) with that of known Δ0-ELOs from yeasts (Denic and Weissman 2007). Only two Δ0-ELOs, both from *N. gaditana* (*Naga_100083g23* and *Naga_100017g49*) were up-regulated under dark conditions (Fig. 3). One of these enzymes (*Naga_100083g23*) has been proven experimentally, by heterologous expression in yeasts, to catalyse the formation of saturated fatty acids containing up to 28 carbons (Dolch et al. 2017). However, analysis of a mutant of *N. gaditana* in which the gene coding for the Δ0-ELO *Naga_100083g23* has been knocked-out (Dolch et al. 2017) exhibited a distribution of LCAs, LCDs, and LCHFAs very similar to that of the wild type (CCMP526, Supplemental Fig. S4), indicating that *Naga_100083g23* is not involved in the biosynthesis of these compounds in *N. gaditana*. This, along with the lack of up-regulated Δ0-ELOs in *N. oceanica* under dark conditions (Fig. 3), suggests that Δ0-ELOs are not involved in the biosynthesis of LCHFAs in *Nannochloropsis* spp.. The intermediates required for the biosynthesis of LCAs, LCDs, and LCHFAS might thus be formed by other enzymes.

FAE enzymes as elongators of fatty acids

FAE enzymes are known to be involved in the biosynthesis of saturated and monounsaturated $C_{20–28}$ fatty acids in plants (Joubes et al. 2008, Haslam and Kunst 2013). *Nannochloropsis gaditana* contains one gene coding for FAE (*Naga_100174g4*) and we found in *N. oceanica* two amino acid (AA) sequences (*CCMP1779|2397* and *CCMP1779|4026*) that align with two different regions of the gene product of

For Peer Review

Naga_100174g4 (Supplemental Fig. S5). The alignment of *CCMP1779|2397* and *CCMP1779|4026* with *Naga_100174g4* as well as the similar expression level exhibited by *CCMP1779|2397* and *CCMP1779|4026* (increase by 18 to 22 fold in the dark treatment, Fig. 3), strongly suggest that these two AA sequences are two contiguous parts of the same protein. Phylogenetic analyses indicate that FAEs from *Nannochloropsis* spp. cluster with proteins from diatoms and Pelagophyceae forming a well-supported clade (Supplemental Fig. S6), which groups with known FAEs from higher plants (Joubes et al. 2008). *Nannochloropsis* FAEs possess two trans membrane helices (TMHs), two domains for KCSs and a domain for chalcone/stilbene synthase (Supplemental Fig. S6).

Interestingly, the genes coding for FAEs in both *N. oceanica* and *N. gaditana* are up-regulated by >10 fold in the dark treatment (Fig. 3) suggesting an enhanced enzymatic activity of FAEs under dark conditions. FAE enzymes are reported to elongate functionalized fatty acids at an intermediate position and also can accept substrates of variable length including $C_{24–28}$ fatty acids (Haslam & Kunst, 2013). For example a FAE from the higher plant *Physaria fendlerii* is known to catalyse the elongation of 12-OH-$C_{18:1}$ to 14-OH-$C_{20:1}$ fatty acid (Moon et al. 2001) and the moss *Funaria hygrometrica* contains $C_{32:0}$ 1,7 diols, which have been suggested to originate from the elongation of 3-hydroxyacyl intermediates, catalyzed by FAE enzymes (Busta et al. 2016). Furthermore, the 7-18-$(OH)_2$-$C_{24:1}$ fatty acid from *Orychophragmus violaceus* (Brassicaceae) has also been shown to derive from FAE-catalysed elongation of a 3-OH-intermediate of the 12-OH-$C_{18:1}$ fatty acid (Li et al. 2018). Thus, the high expression level of genes coding FAEs in both *N. oceanica* and *N. gaditana*, along with the potential enzymatic capability of these proteins to elongate in-chain functionalised fatty acids, suggest that FAE enzymes might play a role in the formation of LCHFAs from $C_{14–20}$ fatty acids.

Role of PKSs in fatty acid hydroxylation

For Peer Review

Since FAE enzymes can elongate not only fatty acids but also hydroxy fatty acids, the hydroxylation process required for the formation of mid-chain hydroxy groups might occur before chain elongation takes place. The formation of both 13-OH-$C_{30:0}$ and 15-OH-$C_{32:0}$ fatty acids by FAE-based elongation would then require a 3-OH-$C_{20:0}$ fatty acid as a starter (Fig. 4). Since 3-OH-$C_{20:0}$ fatty acid has not been detected in *Nannochloropsis* spp. as well as other Eustigmatophyceae, it might be an intermediate in the chain elongation. The mid chain functionalisation of fatty acids can be catalysed by PKS enzymes since they possess acyl transferase (AT) and KAS domains but might lack any or all of the other catalytic sites required to complete a fatty acid elongation cycle (Staunton and Weissman 2001, Jenke-Kodama et al. 2005). Type I PKSs consist of single multifunctional enzymes possessing several catalytic domains and their distribution is scattered among different lineages since genes coding PKSs have not been found in ciliates and Rhizaria (Shelest et al. 2015, Kohli et al. 2016). Three genes from *N. oceanica* have been previously suggested to code for PKSs (Vieler et al. 2012, Poliner et al. 2015, Alboresi et al. 2016) and two genetically distinct PKS clades were previously detected in *N. oceanica* and *N. gaditana* (Shelest et al. 2015).

We identified 22 genes coding for PKSs in the different *Nannochloropsis* spp. (Supplemental Table S3) and built a phylogenetic tree of the KAS domain (KAS-PKS). KAS-PKS phylogeny indicates that five gene products (PKS-Clade I) correspond to the iterative type I PKSs previously identified by Shelest et al. (2015) and are closely related to two other PKSs from *N. gaditana* (*Naga_100093g21*) and *N. oceanica* (*CCMP1779|289*), respectively (Fig. 5A). Sequences from PKS-Clade I cluster with type I FAS/PKS from fungi and Metazoa, whereas 15 other gene products show only weak similarities with KAS-PKS from other species and form three distinct clades: PKS-Clade II, PKS-Clade III and PKS-Clade IV (Fig. 5A). Transcriptomic data from the dark incubation experiments of *N. oceanica* and *N. gaditana* indicate that the genes coding for PKS-Clade I and PKS-Clade III enzymes were

For Peer Review

down-regulated or did not exhibit significant changes under dark conditions, while those coding for PKS-Clade II were up-regulated (Fig. 3), and we did not detect genes coding for PKS-Clade IV in our transcriptomes. PKS-Clade II enzymes *CCMP1779|6720* and *Naga_100086g4* increased their expression in the dark treatment by 3.2 and 2.5 fold, respectively (Fig. 3) suggesting they can be potentially involved in the hydroxylation of fatty acids.

Interestingly PKS-Clade II enzymes possess domains for PKS-KAS (IPR020841), AT (IPR020801), as well as an adenosine monophosphate (AMP) binding domain (IPR000873), a phosphopantetheine-binding ACP domain (PPT, IPR009081), and a ketoacyl-ACP-reductase (KAR, IPR013968) domain (Fig. 5B). The presence of catalytic domains for both KAS and KAR in PKS-Clade II enzymes and the lack of hydroxyacyl dehydratase (HAD) and enoyl reductase (ER) domains suggest that PKS-Clade II enzymes might catalyse an incomplete fatty acid elongation leading to the formation of 3-OH-fatty acids. $C_{14\text{-}18}$ fatty acids might thus be elongated to form 3-OH-$C_{16–20}$ fatty acid intermediates by PKS-Clade II enzymes. The incomplete elongation of the $C_{18:0}$ fatty acid might lead to the formation of a 3-OH-$C_{20:0}$ intermediate which, after 5 or 6 full elongation cycles, potentially catalyzed by the FAE enzymes, would form the 13-OH-$C_{30:0}$ and 15-OH-$C_{32:0}$ fatty acids, respectively, the two LCHFAs present in *Nannochloropsis* spp. (Gelin et al. 1997a, Balzano et al. 2017).

Reduction of LCHFAs to LCDs

LCDs and LCAs are likely formed from LCHFAs as evidenced by the depletion of LCHFAs and increase in LCAs and LCDs in the dark incubation experiments (Fig. 1). Furthermore, the presence of $C_{14–24}$ alkanols (Volkman et al. 1999) as well as $C_{15–17}$ alkanes

For Peer Review

and the $C_{15–31}$ alkenes (Gelin et al. 1997b, Sorigue et al. 2016, Zhou et al. 2016), also suggest the occurrence of fatty acid reduction activities in *Nannochloropsis* spp.. Odd-numbered alkanes and alkenes are typically formed from the reduction of even-numbered fatty acids to aldehydes followed by a decarbonylation step, as described in *Arabidopsis thaliana* (Bernard et al. 2012) and *Chlamydomonas reinhardii* (Sorigue et al. 2016). Similarly, fatty alcohols are also formed from the reduction of fatty acids catalysed by alcohol-forming fatty acid reductases (FAR) as shown in *A. thaliana* (Li-Beisson et al. 2010), *Apis mellifera* (Teerawanichpan et al. 2010), *Calanus finmarchicus* (Teerawanichpan and Qiu 2012) and *Euglena gracilis* (Teerawanichpan and Qiu 2010). However we could not find any protein sequence containing the conserved motif [IVF]X[ILV]TGXTGF[MLV][GA] which corresponds to the FAR catalytic site (Hofvander et al. 2011, Teerawanichpan and Qiu 2012), and none of the *Nannochloropsis* protein belongs to any FAR family (IPR026055, IPR008670, IPR016836, IPR003157). Furthermore, similarity searches (blastp) of known FARs against the deduced amino acid sequences of *Nannochloropsis* genomes produced hits with low (bit score < 50) similarity (data not shown). Indeed a recent study also failed to detect genes coding for fatty acyl-CoA reductases (FARs) in *Nannochloropsis* genomes (Sorigue et al. 2016). This indicates that the enzymes involved in fatty acid reduction in *Nannochloropsis* are either unrelated or greatly divergent from known FARs.

We then searched within the genomes of *N. oceanica* and *N. gaditana* for genes coding for domains involved in the reduction of carboxylic acids. We found 44 genes coding for the short chain dehydrogenase/reductase (SDR) and eight genes that can code for the male sterility 2 (MS2) domain (Supplemental Table S4). Proteins with the male sterility (MS2) domain can catalyse the reduction of fatty acids in *A. thaliana* (Aarts et al. 1997), and 5 of these proteins are annotated as PKS-Clade I (Supplemental Table S5, Fig. 5) since they also possess the catalytic domains for fatty acid elongation. PKS-Clade I enzymes were previously

For Peer Review

485 suggested to be involved in the reduction of fatty acids in *N. gaditana* (Scholz et al. 2014).
486 Since the genes coding for PKS-Clade I as well as the other genes coding for the MS domain
487 are not up-regulated under dark conditions (Fig. 3, Supplemental Table S4) their products are
488 unlikely to be involved in the reduction of LCHFAs to LCDs, although a role in other
489 reduction processes cannot be discarded. Genes coding for SDR were also mostly down-
490 regulated under dark condition (Supplemental Table S5); only six of them were up-regulated
491 but blastp analyses revealed similarities with proteins from other species with very different
492 functions (Supplemental Table S5).
493 Thus, although *Nannochloropsis* spp. contain a range of compounds (LCAs, LCDs,
494 alkanes, alkenes, and alkanols) that are very likely to originate from the reduction of fatty
495 acids, we could not find any enzyme potentially involved in these reductive processes.
496
497
498 **Δ^2- and Δ^3-$C_{20:1}$ fatty acids as potential LCA precursors**
499
500 LCAs contain a double bond at the same position where LCDs have the hydroxy group
501 (Supplemental Fig. S2, Gelin et al. 1997b) suggesting that LCAs might be formed from the
502 dehydration of the mid-chain alcohol group of LCDs. Thus, we searched for dehydratase
503 domains and found 14 genes coding for different lipid dehydratase domains (Supplemental
504 Table S7), and one of them (*Naga_100113g71*) was up-regulated under dark conditions
505 (Supplemental Table S8). *Naga_100113g71,* and its *N. oceanica* homologue
506 (*CCMP1779|4800*) code for proteins possessing a HAD domain and cluster with HADs from
507 other species in our phylogenetic analyses (Supplemental Fig. S7).
508 Alternatively, the dehydration of the secondary alcohol group may occur at an earlier
509 stage, e.g. as a result from an incomplete fatty acid elongation process in which ER activity is

For Peer Review

missing, followed by several further elongation processes (Fig. 4). The higher expression levels of *Naga_100113g71* under dark conditions (Fig. 3), in spite of a decrease in $C_{14–16}$ fatty acids (Fig. 1F), suggest that these enzymes might have been catalysing the dehydration of other compounds such as longer fatty acids. Since HAD enzymes are thought to have a broad substrate specificity (Heath and Rock 1996, Leonard et al. 2004), potentially catalysing the dehydration of 3-hydroxyacyl chains of different lengths, they might also accept 3-OH-$C_{20:0}$ fatty acids as substrates to form the Δ^2- and Δ^3-$C_{20:1}$ fatty acid, which, if further elongated and reduced, might lead to the formation of the C_{30-32} alkenols typically found in *Nannochloropsis* spp. (Fig. 4).

Incorporation of LCAs and LCDs in *Nannochloropsis* cell wall biopolymer

The presence of ether- and ester-bound LCAs and LCDs within the cell wall of *Nannochloropsis* spp. has been long hypothesised (Volkman et al. 1992, Gelin et al. 1996, Gelin et al. 1997b, Volkman 2018). Fourier transform infrared spectroscopy (FTIR) on the cell wall of *N. gaditana* demonstrated the presence of ether bonds and also found some C=O stretches, but whether the latter are related to carboxylic, aldehyde, ketone, or ester functional groups is not clear (Scholz et al. 2014). Although the core of the cell wall polymer may be ether-bound, as they are resistant against base and acid hydrolysis (Gelin et al. 1997b), some of the LCAs and LCDs present in *Nannochloropsis* spp. likely occur as ester-bound moieties to polymeric carboxyl groups (Volkman 2018). The formation of esters from alcohols and fatty acids is typically catalysed by bifunctional wax ester synthase/acyl coenzyme A (acyl-CoA):diacylglycerol acyltransferase (WS-DGAT) (Kalscheuer and Steinbuchel 2003). A gene coding for WS-DGAT (*Naga_100114g61*) was previously predicted in *N. gaditana* (Cui et al.

For Peer Review

2018), and our phylogenetic analyses (Supplemental Fig. S8) indicate that the proteins encoded by *Naga_100114g61* and its *N. oceanica* homologue *(CCMP1779|236*) are closely related to a WS-DGAT from *Phaeodactylum tricornutum* (PtWS-DGAT). PtWS-DGAT has been recently shown to catalyse the formation of esters from alcohols and fatty acids (Cui et al. 2018) and has a domain structure (Fig. 1 in Cui et al. 2018) similar to that found here for *Nannochloropsis* WS-DGATs (*Naga_100114g61* and *CCMP1779|236,* Supplemental Fig. S8). Interestingly, putative WS-DGAT from *N. gaditana* and *N. oceanica* increased in expression by 6 and 1.5 fold, respectively (Fig. 3), under dark conditions, suggesting that an active production of esters was likely to take place during dark incubation. Thus, WS-DGATs in *Nannochloropsis* spp. might be involved in esterification of LCAs/LCDs to carboxyl groups to form the ester-bound structures which have been previously detected in the cell wall (Scholz et al. 2014).

In contrast, we could not find any gene potentially catalysing the formation of ether-bonds within the cell wall biopolymers. Similarity (blastp) analyses of known ether synthases such as the 9-divinyl ether synthase from *Solanum lycopersicum* (tomato plant), the corvol ether synthase from *Kisatasospora setae* (bacteria), against the predicted proteins of *N. oceanica* CCMP1779 and *N. gaditana* B31 did not yield significant hits (Supplemental Table S9). The polymerisation of LCAs and LCDs to form ether-bound structures in algaenans remains thus unclear.

Potential biosynthetic pathways for LCA, LCD, and LCHFAs in *Nannochloropsis* spp.

For Peer Review

As previously suggested (Gelin et al. 1997a, Scholz et al. 2014, Balzano et al. 2017, Volkman 2018), our results from stable isotope experiments confirm that LCAs and LCDs derive from $C_{14–18}$ fatty acids by either condensation or elongation.

If biosynthesis occurs via condensation the dominant pathway involves the interaction between the aliphatic end of a C_{14} fatty acid with the carboxylic end of a C_{18} fatty acid to produce a C_{32} compound functionalised on the 15th carbon which would be a precursor of 15-OH-C_{32} fatty acids, C_{32} 1,15 diol and C_{32} alkenols. Although we could not find evidence for such a pathway in literature nor could find genes potentially coding such biosynthetic processes within *Nannochloropsis* genomes, this pathway cannot be fully discarded.

In contrast the elongation pathway is more likely to occur. The lack of elongation intermediates such as aliphatic compounds with a number of carbons comprised between 20 and 26 in *Nannochloropsis* spp. might be due to a rapid uptake of such compounds for the following steps of the pathway. Nevertheless, results from our genomic and transcriptomic analyses, combined with comparisons with biosynthetic pathways in plants, more likely suggest that LCHFAs are formed from $C_{14–18}$ fatty acids via elongation. Specifically we found two key enzymes, PKS-Clade II and FAE, potentially involved in the elongation process. PKS-Clade II are likely to elongate and hydroxylate the $C_{18:0}$ fatty acid and, to a lesser extent, the $C_{16:0}$ fatty acid, to form the 3-OH-$C_{20:0}$ and 3-OH-$C_{18:0}$ fatty acids, respectively (Fig. 4). Subsequently, FAE enzymes can potentially catalyse the multiple elongation of the 3-OH-$C_{20:0}$ fatty acid to 13-OH-$C_{30:0}$ and 15-OH-$C_{32:0}$ fatty acids after 5 or 6 complete elongation cycles, respectively (Fig. 4). Six complete elongation cycles of 3-OH-$C_{18:0}$ fatty acid and subsequent reduction might form the $C_{30:0}$ 1,15 diol which is also present in *Nannochloropsis* spp. as well as other eustigmatophycean representatives (Rampen et al. 2014).

Since saturated LCDs are functionalised at the same mid-chain position as their corresponding monounsaturated LCAs (Supplemental Fig. S2), both lipid classes are very

For Peer Review

583 likely to share a similar biosynthetic pathway and to originate from the same precursors, the
584 3-OH-$C_{20:0}$ fatty acid and, to a lesser extent the 3-OH-$C_{18:0}$ fatty acid. The pathway leading to
585 the formation of LCAs would start from the dehydration of 3-OH-$C_{20:0}$ fatty acids to both Δ^2
586 and Δ^3 $C_{20:1}$ fatty acid catalysed by an HADs (Fig. 4). The Δ^2 and $\Delta^3C_{20:1}$ fatty acids would
587 then undergo 5 complete FAE-catalysed elongations and a reduction to form the Δ^{12} and Δ^{13}
588 $C_{30:1}$ alkenols or 6 elongations and a reduction to form Δ^{14} and Δ^{15} $C_{32:1}$ alkenols (Fig. 4). The
589 Δ^{12} and Δ^{13} $C_{30:1}$ alkenols as well as Δ^{14} and Δ^{15} $C_{32:1}$ alkenols have been detected here in both
590 *N. gaditana* and *N. oceanica* (Supplemental Fig. S2) and were also found previously in
591 *Nannochloropsis salina* and an unidentified *Nannochloropsis* strain (Gelin et al. 1997b). The
592 Δ^{14} and Δ^{15} $C_{30:1}$ alkenols would instead derive from the dehydration of 3-OH-$C_{18:0}$ fatty acid
593 to Δ^2 and $\Delta^3C_{18:1}$ fatty acids followed by 6 complete elongation cycles and the reduction of
594 the carboxylic group to alcohol.

595 The formation of a double bond in LCDs and a second double bond in LCAs would
596 originate at an early stage of the pathway, before the 3-OH $C_{20:0}$ fatty acid is either elongated
597 to form LCHFAs, or dehydrated to form LCA precursors (Fig. 4). A double bond on a Δ^{27}
598 position, for both the $C_{32:2}$ alkenol and the $C_{32:1}$ diol, might originate from a desaturation of
599 the 13th carbon in $C_{18:0}$ fatty acid or a desaturation of the 15th carbon in 3-OH-$C_{20:1}$ fatty acid.
600 The formation of Δ^{13} $C_{18:1}$ or a Δ^{15} 3-OH-$C_{20:1}$ would potentially involve the activity of a
601 stereospecific desaturase such as a Δ^{13}stearoyl desaturase. *Nannochloropsis* spp. contains 29
602 proteins with domains for fatty acid desaturation and some of them are up-regulated under
603 dark conditions (Supplemental Table S9), it is unclear whether any of these enzymes exhibits
604 Δ^{13}stearoyl desaturase activity.
605
606
607 **LCD production in other species**

For Peer Review

LCDs can also be produced by other phytoplankters (Sinninghe Damsté et al., 2003; Rampen et al., 2011) as well as some plants (Buschhaus et al. 2013). To evaluate the presence of FAEs and PKSs in LCD-producers other than Eustigmatophyceae we analysed the proteins predicted from genomes or transcriptomes available to date. The diatom *Proboscia alata* can code for three putative PKSs as well as a FAE (Supplemental Fig. S9). Similarly to PKS-Clade II enzymes from *Nannochloropsis* spp. (Fig. 5B), PKSs from *P. alata* possess both KAS and KAR domains (Supplemental Fig. S9) being thus potentially able to catalyse the formation of hydroxylated products. Indeed, *Proboscia* species contain $C_{28–30}$ 1,14 diols and 12-OH $C_{27–29}$ methyl alkanoates which were previously suggested to be formed from 12-OH-C_{26-28} fatty acids (Sinninghe Damsté et al. 2003). The 12-OH-$C_{26–28}$ fatty acids might originate after 5 full elongation cycles of 2-OH-$C_{16–18}$ fatty acids, which would in turn derive from an incomplete elongation (and thus hydroxylation) of $C_{14–16}$ fatty acids, with FAEs catalysing the former reaction and PKSs the latter. The hydroxylation of $C_{14–16}$ fatty acids should thus occur, in this case, on the 2nd, rather than on the 3rd carbon to eventually produce the LCD detected in *Proboscia* spp..

$C_{26–32}$ aliphatic diols with a primary and a secondary alcohol group can also be present in the epicuticular waxes of aquatic ferns (Speelman et al. 2009, Mao et al. 2017), terrestrial ferns (Jetter and Riederer 1999) as well as other land plants such as mosses (Busta et al. 2016), conifers (Wen and Jetter 2007), and flowering plants (Wen et al. 2006, Racovita and Jetter 2016). Similarly to the biosynthetic pathways proposed here for *Nannochloropsis* spp., LCDs from plants could start with the formation of 3-hydroxyacyl compounds mediated by P450 hydroxylases or PKS enzymes, followed by FAE-catalysed elongation of 3-hydroxyacyl intermediates as suggested for plants (Wen and Jetter 2007, Busta et al. 2016).

For Peer Review

CONCLUSIONS

LCAs and LCDs are likely to originate from C_{14-18} fatty acids after either condensation of C_{14-18} fatty acids or an incomplete fatty acid elongation, forming 3-OH-fatty acids, followed by a further elongation to LCHFAs and a final reduction of the terminal carboxylic group. Enzymes potentially involved in such a condensation of two fatty acids are not known to date. We identified instead two enzymes (PKS-Clade II and FAE) likely to be involved in the elongation of C_{14-18} fatty acids to larger products. HAD enzymes might play a role in the dehydration of secondary alcohols forming the double bonds present in LCAs whereas WS-DGAT enzymes are potentially involved in the formation of the ester-bound structures present in the *Nannochloropsis* cell wall. Although the biosynthetic pathways for LCAs and LCDs have not been fully elucidated and the formation of ether bonds within cell wall polymers is still unclear, our work identifies a potential mechanism, similar to biosynthetic processes described in higher plants, for the formation of mid-chain functionalised aliphatic compounds in phytoplankton. Future challenges include the biochemical and functional characterization of the candidate enzymes predicted here. Eventually, if long chain aliphatic compounds are formed from the elongation of C_{14-18} fatty acids, genetic manipulations of PKS-Clade II and FAE enzymes might contribute to increase the productivity of both LCHFAs and LCDs in *Nannochloropsis* species.

For Peer Review

MATERIALS AND METHODS

Culturing and dark incubations

Nannochloropsis oceanica CCMP1779 and *N. gaditana* CCMP526 were cultured in batch using f/2 medium (Guillard 1975) under 12:12 dark/light conditions at 20 °C and algal growth was regularly monitored using flow cytometry (Marie et al. 2001). For the experiments each strain was grown in 6 replicate 1.5 L Erlenmeyer flasks (Supplemental information) and cells were harvested from 3 flasks to assess the initial concentration of lipids and the background gene expression (Control 1). From each of the remaining 6 flasks (3 per species) an aliquot (20 mL equals approximately 7 x 10^7 cells) was transferred into new Erlenmeyer flasks pre-filled with medium, incubated under dark/light conditions and used as positive control (Control 2), whereas the initial flasks with the remaining volume (780 mL) were instead transferred under dark conditions. Both Control 2 and dark treatment were incubated for 1 week at 20 °C. Cells were harvested from their culturing flasks by filtration through 0.7 μm GF/F filters (Whatman, Maidstone, UK). Cells were enumerated by flow cytometry before filtration and we did not observe dead cells or debris (i.e. particles with comparable forward scatter and lower chlorophyll fluorescence than ordinary *Nannochloropsis* cells) in our cytograms in any of the samples filtered. This suggests that most of the material filtered contained viable cells, or at worst dead cells with intact chloroplasts. Filters for lipid analyses were immediately rinsed in demineralised water and stored at -80 °C, whereas filters for RNA extraction and further transcriptomic analyses were flash frozen in liquid nitrogen immediately and then stored at -80 °C until analyses.

For Peer Review

In addition we analysed the composition of LCAs, LCDs, and LCHFAs in mutant strains of *N. gaditana* CCMP526 in which the gene *Naga_100083g23* coding for a Δ0-ELOs has been silenced (Dolch et al. 2017). Three mutants (Clone 5, Clone 13, and Clone 15) were obtained from the Cell and Plant Physiology Laboratory (Grenoble, France), cultured under the same conditions as above (f/2 medium, 12/12 dark light cycle) along with the wild type (*N. gaditana* CCMP526) and harvested during the exponential phase of their growth.

Stable isotope labelling

To assess whether LCDs are formed from $C_{14–16}$ fatty acids, we incubated *N. oceanica* CCMP1779 with sodium ^{13}C-[2]-acetate (Sigma-Aldrich, 279315-1G, Zwijndrecht, Netherlands) under both dark/light and dark conditions. We used an axenic culture of *N. oceanica* CCMP1779 to avoid any consumption of ^{13}C-labelled acetate by heterotrophic bacteria typically present in phytoplankton cultures. The strain was cultured at 20 °C under 12:12 dark/light conditions in a 5 L glass carboy and continuous air bubbling was provided by an aquarium pump connected through0.2 µm pore size filters (Sartorius, Göttingen Germany) to dilute any formed ^{13}C labelled CO_2 A volume of 10 mL of 0.5 M sodium ^{13}C-[2]-acetate was added to the cultures when they reached cellular densities of approx. 1.5 x 10^7 cell mL^{-1}. The dark/light treatment was then incubated at the same conditions for 2 days, whereas the dark treatment was incubated for 1 week in the dark. At the end of the experiments cells were harvested as described above.

Lipid extraction, GC-MS and IR-MS

For Peer Review

For the non-isotopically labelled dark incubation experiment we extracted 18 filters in total, i.e. three replicates per treatment for two strains (*N. oceanica* CCMP1779 and *N. gaditana* CCMP526). The filters were freeze-dried using a Lyoquest (Telstart, Life Sciences) freeze-drier and then they were saponified and acid hydrolysed as described previously (Rodrigo-Gámiz et al. 2015, Balzano et al. 2017). The total lipid extract (TLE) was then dissolved in dichloromethane (DCM) and dried through anhydrous sodium acetate, and subsequently under nitrogen. After extraction, we added 20.4 µg $C_{22:0}$ 7,16-diol as internal standard to the TLE, for quantification purposes. Subsequently, the extracts were methylated with diazomethane and cleaned over a small silica gel column using ethyl acetate as eluent. Prior to analysis, the TLEs were silylated by the addition of N,O-bis-(trimethylsilyl)-trifluoroacetamide (BSTFA) and pyridine, and heating at 60 °C for 20 min.

An aliquot of one lipid extract from *N. oceanica* CCMP1779 (dark treatment) was used to determine the position of the double bond in alkenols and unsaturated LCDs, lipids were derivitised by adding 50 µL dimethyl disulfide and 5 µL iodine solution (60 mg iodine in 1mL diethyl ether), eluted then in DCM and iodine was removed using 50–100 mL 5% sodium thiosulfate solutions.

Compounds were quantified using gas chromatography flame ionization detection using an Agilent 7890B Gas Chromatograph (GC) with a 25 m fused silica column diameter 0.32 mm, coated with CP Sil-5 (thickness 0.12 µm). Identification of the lipids was achieved using GC-mass spectrometry (GC-MS) with an Agilent 7890B GC coupled to an Agilent 5977 mass spectrometer. Identification of the LCDs, LCAs and LCHFAs was based on the fragmentation mass spectra obtained in full scan (m/z 50–800) as described by Versteegh et al. (1997). To discriminate between C30:0 diols and the C32:2 alkenols, which co-elute on the GC, we determined the relative contributions of each compound to the total peak area in the GC-MS

For Peer Review

chromatogram (MassHunter software, Agilent Technologies) in which the two compounds are partially separated.

For the stable isotope experiment, we analysed two filters corresponding to the two treatments of *N. oceanica*. The extraction set-up consisted in saponification followed by acid hydrolysis as described above. Subsequently, we added 19.2 µg $C_{22:0}$ 7,16-diol as internal standard to the extracts and we methylated the TLEs using boron trifluoride in methanol. The extracts were separated by column chromatography, using aluminium oxide as stationary phase, which was activated for 2 h at 150 °C. Lipids were then extracted in three different solutions: hexane:DCM (9:1, vol./vol.), hexane:DCM (1:1, vol./vol.) and methanol:DCM (1:1, vol./vol.). Fatty acids were mostly dissolved in the second hexane:DCM solution whereas LCAs, LCDs and LCHFAs were dissolved in the methanol:DCM solutions. For the silylation and the methylation we used BSTFA and BF_3/methanol, respectively, with known $\delta^{13}C$ values of -32.2‰ and -25.7‰, respectively. Compounds were quantified by GC-FID chromatograms as described above, whereas identification was achieved using an Agilent 7890A GC coupled to an Agilent 5975C MS. All GC-MS parameters were also identical, but the total run time was 60 min. The isotopic composition of the different compounds was analysed using gas chromatography-isotope ratio mass spectrometry, using an Agilent 6890 GC coupled to a Thermo Delta Plus isotope ratio mass spectrometer (IR-MS, ThermoFisher). Separation was achieved on a ZB-5MS column with a length of 60 m, a column diameter of 0.32 mm and a film thickness of 0.25 µm. Oven program was identical as that described for the GC-FID and GC-MS, but the end temperature of 320 °C was held for 30 min (total run time of 80 min). The injection volume was 1 µL and the four fractions were all analyzed in triplicate, and the reported data represent averaged values. The isotopic compositions are reported in units of atom percent (At%). Values were corrected considering the $\delta^{13}C$ values of both BSTFA and methanol.

For Peer Review

753

754

755 **Genomic analyses**

756

757 To identify genes potentially involved in the biosynthesis of LCAs, LCDs, and LCHFAs
758 in Eustigmatophyceae we constructed a local blast database (Altschul et al. 1990, Camacho et
759 al. 2009) using the predicted proteins from the genomes of 8 *Nannochloropsis* strains
760 (*Nannochloropsis* sp. CCMP531, *N. gaditana* B-31, *N. gaditana* CCMP526, *Nannochloropsis*
761 *granulata* CCMP529, *N. oceanica* CCMP1779, *N. oceanica* IMET1, *Nannochloropsis*
762 *oculata* CCMP525, and *N. salina* CCMP537, Supplemental Information). Conserved protein
763 domains were searched using Interproscan (Jones et al. 2014) or by manually enquiring for
764 specific AA motifs. Specifically we searched for proteins containing the motifs HWYHH,
765 GMGCSAG, and [D/E]TACSSS or H[G/A]TGT, which correspond to highly conserved
766 regions of Δ0-ELOs (Hashimoto et al. 2008), FAEs (Millar et al. 1999), and PKSs (Shelest et
767 al. 2015) enzymes, respectively. Moreover we searched for genes potentially coding for the
768 conserved motif [I/V/F]X[I/L/V]TGXTGF[M/L/V][G/A] which corresponds to the catalytic
769 site of FARs (Hofvander et al. 2011, Teerawanichpan and Qiu 2012). The presence and
770 position of trans membrane helices (TMHs), in Δ0-ELOs, FAEs, and PKSs proteins was
771 assessed using TMHMM (Krogh et al. 2001).

772 We carried out similarity searches of known proteins from other species against the
773 locally built *Nannochloropsis* database, as well as similarity searches of *Nannochloropsis*
774 proteins potentially involved in the biosynthetic processes against the non-redundant (NR)
775 (Pruitt et al. 2005) and the Swissprot (The UniProt Consortium, 2017) databases, using blastp
776 (Camacho et al. 2009).

For Peer Review

To compare putative *Nannochloropsis* enzymes involved in the biosynthetic processes with known and unknown proteins from other species we performed phylogenetic analyses on four protein families: FAEs, PKSs, the HADs, and the WS-DGATs. We aligned putative *Nannochloropsis* FAEs with known FAE proteins from *Arabidopsis thaliana, Brassica napus* (Joubes et al. 2008), as well as a FAE known to elongate hydroxy fatty acids from *Physaria fendlerii* (Moon et al. 2001). Sequences were aligned using MAFFT-linsy (Katoh and Standley 2013) and poorly aligned regions (regions containing > 50% gaps) were trimmed from the alignment which finally consisted of 50 sequences and 195 unambiguously aligned positions. We analysed the KAS domain of PKSs (KAS-PKS) using *Nannochloropsis* proteins previously identified as PKSs (Shelest et al. 2015, Alboresi et al. 2016) as well as other proteins containing the KAS-PKS domain (IPR020841) and/or containing the conserved motifs [D/E]TACSSS and H[G/A]TGT. Sequences were then aligned to a pre-existing alignment of 92 KAS-PKS sequences (Kohli et al. 2016) and trimmed as described above. The final alignment consisted of 138 sequences and 173 AA positions. For HADs we extracted two AA sequences containing a domain for HAD (IPR010084) from *N. oceanica* and *N. gaditana*, respectively. We then searched for homologues in the *Nannochloropsis* and Swissprot databases, aligned and trimmed the sequences as described above and the final alignment included 50 AA sequences and 147 positions. For the WS-DGAT phylogeny we downloaded known AA sequences from the Swissprot database, searched for homologues in *Nannochloropsis*, aligned the sequences and trimmed the alignment as described above. The alignment included 43 AA sequences and 245 positions. Phylogenetic trees were constructed using both Maximum Likelihood (ML) and Neighbour Joining (NJ) algorithms based on 1,000 bootstraps. ML phylogeny was inferred using RAxML with 1000 bootstraps (Stamatakis 2014) and was used to build the phylogenetic trees, whereas NJ bootstrap support values were calculated using MEGA (Tamura et al. 2007).

For Peer Review

To evaluate the occurrence of similar biosynthetic processes in other LCD-producers (Balzano et al. 2018), we downloaded a number of predicted proteins, obtained from transcriptomes of phytoplankton cultures (Keeling et al. 2014) from iMicrobe (https://www.imicrobe.us/). The species used were *Florenciella parvula* (MMETSP1323), *Florenciella* sp. (MMETSP1324), *Heterosigma akashiwo*(MMETSP0292, MMETSP0294, MMETSP0295, MMETSP0296, MMETSP0409, MMETSP0410, MMETSP0411, MMETSP0414, MMETSP0415, MMETSP0416, MMETSP0894, MMETSP0895, MMETSP0896, MMETSP0897), *Phaeomonas parva* (MMETSP1163), *Florenciella parvula* (MMETSP1323), *Florenciella* sp. (MMETSP1324), *Proboscia alata* (MMETSP0174, MMETSP0176), *Proboscia inermis* (MMETSP0816), and *Sarcinochrysi* sp. (MMETSP1170). In addition, we downloaded the genomes of the plants *Azolla filliculoides* (www.fernbase.org)and *Triticum aestivum* (Kersey et al. 2018) which are also known to produce LCDs (Speelman et al. 2009, Racovita and Jetter 2016). We then analysed the domain structure of these proteins using Interproscan (Jones et al. 2014) and searched for PKSs coding KAS and KAR domains.

RNA extraction

To prevent RNA degradation, the extractions were carried out under sterile and cold (approx. 10 °C) conditions in a clean laboratory; samples, tubing, and all other equipment used were kept in ice unless otherwise stated. RNA was extracted from each of 3 replicates of each of the three treatments (initial biomass, positive control, and dark-incubated biomass) from both *N. oceanica* CCMP1779 and *N. gaditana* CCMP526 for a total of 18 samples. Cells were disrupted using a combination of thermal, chemical, and mechanical lyses: from each

For Peer Review

sample about ½ of a GF/F filter was cut in many small pieces using sterile tweezers and scissors, drilled using disposable pellet pestles, and then transferred into 12 mL falcon tubes prefilled with 0.1 and 0.5 µm glass beads (Biospec, Bartlesville, Canada). Tubes were then rapidly submerged several times into liquid nitrogen to promote thermal cell lysis. One mL of RLT buffer (Qiagen, Venlo, Netherlands), 10 µL mercaptoethanol (Sigma-Aldrich,), and 50 µL plant RNA isolation aid (Thermo Fisher Scientific, Waltham, US) were then added to the tubes which were vortexed for 5 minutes, incubated for 5 min in ice, vortexed again for 5 min and finally centrifuged at 4,500 g. The supernatant was transferred into 2 mL tubes which were centrifuged again at 16,000 g and the supernatant removed. 35 µL lysozyme (Qiagen), 20 µL proteinase-K (Qiagen), and 100 µL sodium-dodecyl-sulfate (Ambion, Bleiswijk, Netherlands) were then added to the samples which were incubated at 37 °C for 10 min. Tubes were then centrifuged for 15 min and the supernatant transferred into DNA spin column (DNAeasy blood and tissue kit, Qiagen) and centrifuged to remove most of the DNA. The lysate was transferred into RNAeasy spin columns (RNAeasy mini-kit, Qiagen) and the RNA was then isolated following the instructions provided by the supplier. Traces of DNA were removed from the RNA extract using Turbo DNAse (Thermo Fisher Scientific). RNA concentration and integrity were assessed using Qubit Fluorometric Quantitation (Thermo Fisher Scientific) and a Bioanalyser (Agilent, Santa Cruz, US) whereas a PCR using universal eukaryote primers (Stoeck et al. 2010) was carried out to confirm the absence of DNA contamination within the RNA extracts.

RNA extracts were sent to Utrecht Sequencing Facility (www.useq.nl), where cDNA was generated, sequencing libraries prepared, and sequencing carried out with two runs on a NextSeq500 with reads of 75 bp.

For Peer Review

852 **Gene expression analyses**

853

854 A total of 614,537,691 raw fastq reads were obtained and processed locally on a
855 bioinformatic cluster. Low quality reads were trimmed or removed using Trimmomatic
856 (Bolger et al. 2014) with the maxinfo method (MAXINFO:40:0.6) and a minimal length of
857 acceptable reads of 36 bp. Adapters were also removed using Trimmomatic with the
858 Illuminaclip option. The quality of the trimmed reads was controlled with fastqc
859 (www.bioinformatics.babraham.ac.uk/projects/fastqc/) and reads were sorted and counted
860 using the R library Rsamtools (Morgan et al. 2017). Reads were mapped against the
861 previously sequenced genomes of *N. oceanica* CCMP1779 (Vieler et al. 2012) available at
862 Joint Genome Institute (https://genome.jgi.doe.gov/pages/search-for-
863 genes.jsf?organism=Nanoce1779) and *N. gaditana* B-31 (Corteggiani Carpinelli et al. 2014)
864 downloaded from www.nannochloropsis.org. Mapping was performed using the R library
865 GenomicFeatures (Lawrence et al. 2013), data were normalised using Deseq2 (Love et al.
866 2014) and a gene expression table as well as log2 fold changes with corresponding p-values
867 corrected with the false discovery rate (Benjamini and Hochberg 1995) was obtained.
868 Expression changes were considered significant for p-values < 0.01.

For Peer Review

FUNDING INFORMATION

This research was funded by the European Research Council (ERC) under the European Union's Seventh Framework Program (FP7/2007-2013) ERC grant agreement [339206]. Financial support from the Netherlands Earth System Science Centre (NESSC) is provided to S. Schouten and J. S. Sinninghe Damsté. The French National Research Agency supports J. Lupette and E. Marechal.

ACKNOWLEDGEMENTS

We thank M. Verweij for help in lipid analyses, M. Grego for support in algal culturing, S. Vreugdenhil for help in RNA extraction, A. Abdala for assistance in bioinformatic analyses and Rob Langelaan from Naturalis for help in electron microscopy. We thank Utrecht Sequencing Facility (USEQ) for providing sequencing service and data. USEQ is subsidized by the University Medical Center Utrecht, Hubrecht Institute and Utrecht University.

DISCLOSURE

The authors have no conflict of interest to declare.

For Peer Review

REFERENCES

Aarts, M. G. M., R. Hodge, K. Kalantidis, D. Florack, Z. A. Wilson, B. J. Mulligan, W. J. Stiekema, R. Scott and A. Pereira (1997). "The *Arabidopsis* MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes." Plant Journal **12**(3): 615-623.

Alboresi, A., G. Perin, N. Vitulo, G. Diretto, M. Block, J. Jouhet, A. Meneghesso, G. Valle, G. Giuliano, E. Marechal and T. Morosinotto (2016). "Light Remodels Lipid Biosynthesis in *Nannochloropsis gaditana* by Modulating Carbon Partitioning between Organelles." Plant Physiology **171**(4): 2468-2482.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." Journal of Molecular Biology **215**(3): 403-410.

Balzano, S., J. Lattaud, L. Villanueva, S. W. Rampen, C. P. D. Brussaard, J. van Bleijswijk, N. Bale, J. S. Sinninghe Damste and S. Schouten (2018). "A quest for the biological sources of long chain alkyl diols in the western tropical North Atlantic Ocean." Biogeosciences **15**(19): 5951-5968.

Balzano, S., L. Villanueva, M. de Bar, J. S. Sinninghe Damsté and S. Schouten (2017). "Impact of culturing conditions on the abundance and composition of long chain alkyl diols in species of the genus *Nannochloropsis*." Organic Geochemistry **108**: 9-17.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the false discovery rate. A pratical and powerful approach to multiple testing." Journal of the Royal Statistical Society Series B-Methodological **57**(1): 289-300.

Bernard, A., F. Domergue, S. Pascal, R. Jetter, C. Renne, J. D. Faure, R. P. Haslam, J. A. Napier, R. Lessire and J. Joubes (2012). "Reconstitution of Plant Alkane Biosynthesis in Yeast Demonstrates That *Arabidopsis* ECERIFERUM1 and ECERIFERUM3 Are Core Components of a Very-Long-Chain Alkane Synthesis Complex." Plant Cell **24**(7): 3106-3118.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Buschhaus, C., C. Peng and R. Jetter (2013). "Very-long-chain 1,2-and 1,3-bifunctional compounds from the cuticular wax of *Cosmos bipinnatus* petals." Phytochemistry **91**: 249-256.

Busta, L., J. M. Budke and R. Jetter (2016). "Identification of beta-hydroxy fatty acid esters and primary, secondary-alkanediol esters in cuticular waxes of the moss *Funaria hygrometrica*." Phytochemistry **121**: 38-49.

Busta, L. and R. Jetter (2018). "Moving beyond the ubiquitous: the diversity and biosynthesis of specialty compounds in plant cuticular waxes." Phytochemistry Reviews **17**(6): 1275-1304.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden (2009). "BLAST plus : architecture and applications." Bmc Bioinformatics **10**: 9.

Chen, L., T. Z. Liu, W. Zhang, X. L. Chen and J. F. Wang (2012). "Biodiesel production from algae oil high in free fatty acids by two-step catalytic conversion." Bioresource Technology **111**: 208-214.

Christenson, J. K., M. R. Jensen, B. R. Goblirsch, F. Mohamed, W. Zhang, C. M. Wilmot and L. P. Wackett (2017). "Active Multienzyme Assemblies for Long-Chain Olefinic Hydrocarbon Biosynthesis." Journal of Bacteriology **199**(9).

Corteggiani Carpinelli, E., A. Telatin, N. Vitulo, C. Forcato, M. D'Angelo, R. Schiavon, A. Vezzi, G. M. Giacometti, T. Morosinotto and G. Valle (2014). "Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion." Molecular Plant **7**(2): 323-335.

Cui, Y. L., J. L. Zhao, Y. C. Wang, S. Qin and Y. D. Lu (2018). "Characterization and engineering of a dual-function diacylglycerol acyltransferase in the oleaginous marine diatom *Phaeodactylum tricornutum*." Biotechnology for Biofuels **11**.

For Peer Review

938 Denic, V. and J. S. Weissman (2007). "A molecular caliper mechanism for determining very long-chain
939 fatty acid length." Cell **130**(4): 663-677.
940 Dolch, L. J., C. Rak, G. Perin, G. Tourcier, R. Broughton, M. Leterrier, T. Morosinotto, F. Tellier, J. D.
941 Faure, D. Falconet, J. Jouhet, O. Sayanova, F. Beaudoin and E. Marechal (2017). "A palmitic
942 acid elongase affects eicosapentaenoic acid and plastidial monogalactosyldiacylglycerol
943 levels in *Nannochloropsis*." Plant Physiology **173**(1): 742-759.
944 Dunstan, G. A., J. K. Volkman, S. M. Barrett and C. D. Garland (1993). "Changes in the lipid
945 composition and maximization of the polyunsaturated fatty acid content of 3 microalgae
946 grown in mass culture." Journal of Applied Phycology **5**(1): 71-83.
947 Fietz, S., W. Bleiss, D. Hepperle, H. Koppitz, L. Krienitz and A. Nicklisch (2005). "First record of
948 *Nannochloropsis limnetica* (Eustigmatophyceae) in the autotrophic picoplankton from Lake
949 Baikal." Journal of Phycology **41**(4): 780-790.
950 Gelin, F., I. Boogers, A. A. M. Noordeloos, J. S. Sinninghe Damsté, P. G. Hatcher and J. W. de Leeuw
951 (1996). "Novel, resistant microalgal polyethers: An important sink of organic carbon in the
952 marine environment?" Geochimica Et Cosmochimica Acta **60**(7): 1275-1280.
953 Gelin, F., I. Boogers, A. A. M. Noordeloos, J. S. Sinninghe Damsté, R. Riegman and J. W. De Leeuw
954 (1997b). "Resistant biomacromolecules in marine microalgae of the classes
955 Eustigmatophyceae and Chlorophyceae: geochemical implications." Organic Geochemistry
956 **26**(11-12): 659-675.
957 Gelin, F., J. K. Volkman, J. W. de Leeuw and J. S. Sinninghe Damsté (1997a). "Mid-chain hydroxy long-
958 chain fatty acids in microalgae from the genus *Nannochloropsis*." Phytochemistry **45**(4): 641-
959 646.
960 Goodrum, J. W. and D. P. Geller (2005). "Influence of fatty acid mehyl esters from hydroxylated
961 vegetable oils in diesel fuel lubricity." Bioresource Technology **96**(7): 851-855.
962 Guillard, R. R. L. (1975). Culture of phytoplankton for feeding marine invertebrates. Culture of marine
963 invertebrate animals. W. L. Smith and M. H. Chanley. New York, Plenum Book Publication
964 Corporation**:** 29-60.
965 Hashimoto, K., A. C. Yoshizawa, S. Okuda, K. Kuma, S. Goto and M. Kanehisa (2008). "The repertoire
966 of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes." Journal
967 of Lipid Research **49**(1): 183-191.
968 Haslam, T. M. and L. Kunst (2013). "Extending the story of very-long-chain fatty acid elongation."
969 Plant Science **210**: 93-107.
970 Heath, R. J. and C. O. Rock (1996). "Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein
971 dehydratases in *Escherichia coli* fatty acid biosynthesis." Journal of Biological Chemistry
972 **271**(44): 27795-27801.
973 Hofvander, P., T. T. P. Doan and M. Hamberg (2011). "A prokaryotic acyl-CoA reductase performing
974 reduction of fatty acyl-CoA to fatty alcohol." Febs Letters **585**(22): 3538-3543.
975 Jenke-Kodama, H., A. Sandmann, R. Muller and E. Dittmann (2005). "Evolutionary implications of
976 bacterial polyketide synthases." Molecular Biology and Evolution **22**(10): 2027-2039.
977 Jetter, R. and M. Riederer (1999). "Long-chain alkanediols, ketoaldehydes, ketoalcohols and ketoalkyl
978 esters in the cuticular waxes of *Osmunda regalis* fronds." Phytochemistry **52**(5): 907-915.
979 Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Z. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell,
980 G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez
981 and S. Hunter (2014). "InterProScan 5: genome-scale protein function classification."
982 Bioinformatics **30**(9): 1236-1240.
983 Joubes, J., S. Raffaele, B. Bourdenx, C. Garcia, J. Laroche-Traineau, P. Moreau, F. Domergue and R.
984 Lessire (2008). "The VLCFA elongase gene family in *Arabidopsis thaliana*: phylogenetic
985 analysis, 3D modelling and expression profiling." Plant Molecular Biology **67**(5): 547-566.
986 Kalscheuer, R. and A. Steinbuchel (2003). "A novel bifunctional wax ester synthase/acyl-CoA :
987 diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in
988 *Acinetobacter calcoaceticus* ADP1." Journal of Biological Chemistry **278**(10): 8075-8082.

For Peer Review

Katoh, K. and D. M. Standley (2013). "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." Molecular Biology and Evolution **30**(4): 772-780.

Keeling, P. J., F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. B. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. Twary, P. E. Umale, D. Vaulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone and A. Z. Worden (2014). "The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing." Plos Biology **12**(6).

Kersey, P. J., J. E. Allen, A. Allot, M. Barba, S. Boddu, B. J. Bolt, D. Carvalho-Silva, M. Christensen, P. Davis, C. Grabmueller, N. Kumar, Z. Liu, T. Maurel, B. Moore, M. D. McDowall, U. Maheswari, G. Naamati, V. Newman, C. K. Ong, M. Paulini, H. Pedro, E. Perry, M. Russell, H. Sparrow, E. Tapanari, K. Taylor, A. Vullo, G. Williams, A. Zadissia, A. Olson, J. Stein, S. Wei, M. Tello-Ruiz, D. Ware, A. Luciani, S. Potter, R. D. Finn, M. Urban, K. E. Hammond-Kosack, D. M. Bolser, N. De Silva, K. L. Howe, N. Langridge, G. Maslen, D. M. Staines and A. Yates (2018). "Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species." Nucleic Acids Research **46**(D1): D802-D808.

Kohli, G. S., U. John, F. M. Van Dolah and S. A. Murray (2016). "Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes." ISME Journal **10**(8): 1877-1890.

Kresovic, D., F. Schempp, Z. Cheikh-Ali and H. B. Bode (2015). "A novel and widespread class of ketosynthase is responsible for the head-to-head condensation of two acyl moieties in bacterial pyrone biosynthesis." Beilstein Journal of Organic Chemistry **11**: 1412-1417.

Krogh, A., B. Larsson, G. von Heijne and E. L. L. Sonnhammer (2001). "Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes." Journal of Molecular Biology **305**(3): 567-580.

Lawrence, M., W. Huber, H. Pages, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan and V. J. Carey (2013). "Software for Computing and Annotating Genomic Ranges." Plos Computational Biology **9**(8).

Leonard, A. E., S. L. Pereira, H. Sprecher and Y. S. Huang (2004). "Elongation of long-chain fatty acids." Progress in Lipid Research **43**(1): 36-54.

Li-Beisson, Y., B. Shorrosh, F. Beisson, M. X. Andersson, V. Arondel, P. D. Bates, S. Baud, D. Bird, A. Debono, T. P. Durrett, R. B. Franke, I. A. Graham, K. Katayama, A. A. Kelly, T. Larson, J. E. Markham, M. Miquel, I. Molina, I. Nishida, O. Rowland, L. Samuels, K. M. Schmid, H. Wada, R. Welti, C. Xu, R. Zallot and J. Ohlrogge (2010). "Acyl-lipid metabolism." The Arabidopsis Book **8**: e0133-e0133.

Li, J., D. Han, D. Wang, K. Ning, J. Jia, L. Wei, X. Jing, S. Huang, J. Chen, Y. Li, Q. Hu and J. Xu (2014). "Choreography of transcriptomes and lipidomes of *Nannochloropsis* reveals the mechanisms of oil synthesis in microalgae." Plant Cell **26**(4): 1645-1665.

Li, X. J., A. M. Teitgen, A. Shirani, J. Ling, L. Busta, R. E. Cahoon, W. Zhang, Z. Y. Li, K. D. Chapman, D. Berman, C. Y. Zhang, R. E. Minto and E. B. Cahoon (2018). "Discontinuous fatty acid elongation yields hydroxylated seed oil with improved function." Nature Plants **4**(9): 711-720.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biology **15**(12).

For Peer Review

Mao, S. Y., X. W. Zhu, N. Y. Wu, G. D. Jia, Y. G. Sun, H. X. Guan and D. D. Wu (2017). "Alcohol compounds in *Azolla imbricata* and potential source implication for marine sediments." Science China-Earth Sciences **60**(2): 348-359.

Marie, D., F. Partensky, D. Vaulot and C. Brussaard (2001). Enumeration of phytoplankton, bacteria, and viruses in marine samples. Current protocols in cytometry New York, John Wiley & Sons. **Supplement 10**.

Martinez-Roldan, A. J., H. V. Perales-Vela, R. O. Canizares-Villanueva and G. Torzillo (2014). "Physiological response of *Nannochloropsis* sp. to saline stress in laboratory batch cultures." Journal of Applied Phycology **26**(1): 115-121.

McQuoid, M. R. and L. A. Hobson (1996). "Diatom resting stages." Journal of Phycology **32**(6): 889-902.

Millar, A. A., S. Clemens, S. Zachgo, E. M. Giblin, D. C. Taylor and L. Kunst (1999). "CUT1, an *Arabidopsis* gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme." Plant Cell **11**(5): 825-838.

Moon, H., M. A. Smith and L. Kunst (2001). "A condensing enzyme from the seeds of *Lesquerella fendleri* that specifically elongates hydroxy fatty acids." Plant Physiology **127**(4): 1635-1643.

Morgan, M., H. Pagès, V. Obenchain and N. Hayden (2017). "Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 1.30.0. http://bioconductor.org/packages/release/bioc/html/Rsamtools.html. ."

Mutlu, H. and M. A. R. Meier (2010). "Castor oil as a renewable resource for the chemical industry." European Journal of Lipid Science and Technology **112**(1): 10-30.

Pal, D., I. Khozin-Goldberg, Z. Cohen and S. Boussiba (2011). "The effect of light, salinity, and nitrogen availability on lipid production by *Nannochloropsis* sp." Applied Microbiology and Biotechnology **90**(4): 1429-1441.

Petrovic, Z. S. (2008). "Polyurethanes from vegetable oils." Polymer Reviews **48**(1): 109-155.

Petrovic, Z. S., Y. J. Xu, J. Milic, G. Glenn and A. Klamczynski (2010). "Biodegradation of Thermoplastic Polyurethanes from Vegetable Oils." Journal of Polymers and the Environment **18**(2): 94-97.

Poliner, E., N. Panchy, L. Newton, G. X. Wu, A. Lapinsky, B. Bullard, A. Zienkiewicz, C. Benning, S. H. Shiu and E. M. Farre (2015). "Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light/dark cycles." Plant Journal **83**(6): 1097-1113.

Pruitt, K. D., T. Tatusova and D. R. Maglott (2005). "NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins." Nucleic Acids Research **33**(Database Issue): D501-D504.

Racovita, R. C. and R. Jetter (2016). "Identification of in-chain-functionalized compounds and methyl-branched alkanes in cuticular waxes of *Triticum aestivum* cv. Bethlehem." Plos One **11**(11): 25.

Radakovits, R., R. E. Jinkerson, S. I. Fuerstenberg, H. Tae, R. E. Settlage, J. L. Boore and M. C. Posewitz (2012). "Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*." Nature Communications **3**.

Rampen, S. W., V. Willmott, J.-H. Kim, M. Rodrigo-Gamiz, E. Uliana, G. Mollenhauer, E. Schefuss, J. S. Sinninghe Damsté and S. Schouten (2014). "Evaluation of long chain 1,14-alkyl diols in marine sediments as indicators for upwelling and temperature." Organic Geochemistry **76**: 39-47.

Rodolfi, L., G. C. Zittelli, N. Bassi, G. Padovani, N. Biondi, G. Bonini and M. R. Tredici (2009). "Microalgae for Oil: Strain Selection, Induction of Lipid Synthesis and Outdoor Mass Cultivation in a Low-Cost Photobioreactor." Biotechnology and Bioengineering **102**(1): 100-112.

Rodrigo-Gámiz, M., S. W. Rampen, H. de Haas, M. Baas, S. Schouten and J. S. Sinninghe Damsté (2015). "Constraints on the applicability of the organic temperature proxies UK037, TEX86 and LDI in the subpolar region around Iceland." Biogeosciences **12**: 6573-6590.

For Peer Review

Scholz, M. J., T. L. Weiss, R. E. Jinkerson, J. Jing, R. Roth, U. Goodenough, M. C. Posewitz and H. G. Gerken (2014). "Ultrastructure and composition of the *Nannochloropsis gaditana* cell wall." Eukaryotic Cell **13**(11): 1450-1464.

Sharma, V. and P. P. Kundu (2006). "Addition polymers from natural oils - A review." Progress in Polymer Science **31**(11): 983-1008.

Shelest, E., N. Heimerl, M. Fichtner and S. Sasso (2015). "Multimodular type I polyketide synthases in algae evolve by module duplications and displacement of AT domains in trans." Bmc Genomics **16**.

Sinninghe Damsté, J. S., S. Rampen, W. Irene, C. Rupstra, B. Abbas, G. Muyzer and S. Schouten (2003). "A diatomaceous origin for long-chain diols and mid-chain hydroxy methyl alkanoates widely occurring in Quaternary marine sediments: Indicators for high-nutrient conditions." Geochimica Et Cosmochimica Acta **67**(7): 1339-1348.

Sinninghe Damsté, J. S., W. I. C. Rijpstra, E. C. Hopmans, J. W. H. Weijers, B. U. Foesel, J. Overmann and S. N. Dedysh (2011). "13,16-Dimethyl Octacosanedioic Acid (iso-Diabolic Acid), a Common Membrane-Spanning Lipid of Acidobacteria Subdivisions 1 and 3." Applied and Environmental Microbiology **77**(12): 4147-4154.

Sorigue, D., B. Legeret, S. Cuine, P. Morales, B. Mirabella, G. Guedeney, Y. Li-Beisson, R. Jetter, G. Peltier and F. Beisson (2016). "Microalgae Synthesize Hydrocarbons from Long-Chain Fatty Acids via a Light-Dependent Pathway." Plant Physiology **171**(4): 2393-2405.

Speelman, E. N., G. J. Reichart, J. W. de Leeuw, W. I. C. Rijpstra and J. S. Sinninghe Damsté (2009). "Biomarker lipids of the freshwater fern *Azolla* and its fossil counterpart from the Eocene Arctic Ocean." Organic Geochemistry **40**(5): 628-637.

Stamatakis, A. (2014). "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies." Bioinformatics **30**(9): 1312-1313.

Staunton, J. and K. J. Weissman (2001). "Polyketide biosynthesis: a millennium review." Natural Product Reports **18**(4): 380-416.

Stoeck, T., D. Bass, M. Nebel, R. Christen, M. D. M. Jones, H. W. Breiner and T. A. Richards (2010). "Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water." Molecular Ecology **19**: 21-31.

Tamura, K., J. Dudley, M. Nei and S. Kumar (2007). "MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0." Molecular Biology and Evolution **24**(8): 1596-1599.

Teerawanichpan, P. and X. Qiu (2010). "Fatty Acyl-CoA Reductase and Wax Synthase from *Euglena gracilis* in the Biosynthesis of Medium-Chain Wax Esters." Lipids **45**(3): 263-273.

Teerawanichpan, P. and X. Qiu (2012). "Molecular and Functional Analysis of Three Fatty Acyl-CoA Reductases with Distinct Substrate Specificities in Copepod *Calanus finmarchicus*." Marine Biotechnology **14**(2): 227-236.

Teerawanichpan, P., A. J. Robertson and X. A. Qiu (2010). "A fatty acyl-CoA reductase highly expressed in the head of honey bee (*Apis mellifera*) involves biosynthesis of a wide range of aliphatic fatty alcohols." Insect Biochemistry and Molecular Biology **40**(9): 641-649.

Versteegh, G. J. M., H. J. Bosch and J. W. De Leeuw (1997). "Potential palaeoenvironmental information of C-24 to C-36 mid-chain diols, keto-ols and mid-chain hydroxy fatty acids; a critical review." Organic Geochemistry **27**(1-2): 1-13.

Vieler, A., G. Wu, C.-H. Tsai, B. Bullard, A. J. Cornish, C. Harvey, I.-B. Reca, C. Thornburg, R. Achawanantakun, C. J. Buehl, M. S. Campbell, D. Cavalier, K. L. Childs, T. J. Clark, R. Deshpande, E. Erickson, A. A. Ferguson, W. Handee, Q. Kong, X. Li, B. Liu, S. Lundback, C. Peng, R. L. Roston, Sanjaya, J. P. Simpson, A. TerBush, J. Warakanont, S. Zaeuner, E. M. Farre, E. L. Hegg, N. Jiang, M.-H. Kuo, Y. Lu, K. K. Niyogi, J. Ohlrogge, K. W. Osteryoung, Y. Shachar-Hill, B. B. Sears, Y. Sun, H. Takahashi, M. Yandell, S.-H. Shiu and C. Benning (2012). "Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779." Plos Genetics **8**(11).

For Peer Review

Volkman, J. K. (2018). Lipids of Geochemical Interest in Microalgae. Hydrocarbons, Oils and Lipids: Diversity, Origin, Chemistry and Fate. H. Wilkes. Cham, Springer International Publishing**:** 1-34.

Volkman, J. K., S. M. Barrett and S. I. Blackburn (1999). "Eustigmatophyte microalgae are potential sources of C-29 sterols, C-22-C-28 *n*-alcohols and C-28-C-32 *n*-alkyl diols in freshwater environments." Organic Geochemistry **30**(5): 307-318.

Volkman, J. K., S. M. Barrett, G. A. Dunstan and S. W. Jeffrey (1992). "C-30-C-32 Alkyl diols and unsaturated alcohols in microalgae of the class Eustigmatophyceae." Organic Geochemistry **18**(1): 131-138.

Wang, D., K. Ning, J. Li, J. Hu, D. Han, H. Wang, X. Zeng, X. Jing, Q. Zhou, X. Su, X. Chang, A. Wang, W. Wang, J. Jia, L. Wei, Y. Xin, Y. Qiao, R. Huang, J. Chen, B. Han, K. Yoon, R. T. Hill, Y. Zohar, F. Chen, Q. Hu and J. Xu (2014). "*Nannochloropsis* Genomes Reveal Evolution of Microalgal Oleaginous Traits." Plos Genetics **10**(1).

Wen, M., J. Au, F. Gniwotta and R. Jetter (2006). "Very-long-chain secondary alcohols and alkanediols in cuticular waxes of *Pisum sativum* leaves." Phytochemistry **67**(22): 2494-2502.

Wen, M. and R. Jetter (2007). "Very-long-chain hydroxyaldehydes from the cuticular wax of *Taxus baccata* needles." Phytochemistry **68**(20): 2563-2569.

Zhang, Z. and J. K. Volkman (2017). "Algaenan structure in the microalga *Nannochloropsis oculata* characterized from stepwise pyrolysis." Organic Geochemistry **104**: 1-7.

Zhou, L., J. Chen, J. L. Xu, Y. Li, C. X. Zhou and X. J. Yan (2016). "Change of volatile components in six microalgae with different growth phases." Journal of the Science of Food and Agriculture **97**(3): 761-769.

For Peer Review

1163 **TABLES**

1164

1165 **Table 1.** Combustion enthalpies of the main fatty acids present in *Nannochloropsis* species[a]

1166

Compound	Reaction	Combustion enthalpy KJ.mol^{-1}	KJ.g^{-1}
$C_{14:0}$ FA	$C_{14:0}$ FA + 20 O_2 → 14 H_2O + 14 CO_2	8 300	36.0
$C_{16:1}$ FA	$C_{16:1}$ FA + 22.5 O_2 → 15 H_2O +16 CO_2	9 400	37.0
$C_{16:0}$	$C_{16:0}$ FA + 23 O_2→ 16 H_2O + 16 CO_2	9 500	37.0
$C_{18:1}$	$C_{18:1}$ FA + 25.5 O_2 → 17 H_2O + 18 CO_2	11 160	39.5
$C_{18:0}$	$C_{18:0}$ FA + 26 O_2 → 18 H_2O + 18 CO_2	10 800	38.0
$C_{20:5}$ FA	$C_{20:5}$ FA + 26.5 O_2→ 15 H_2O + 20 CO_2	11 400	38.0
$C_{30:0}$ OH-FA	$C_{30:0}$ OH-FA+ 43.5 O_2 → 30 H_2O + 30 CO_2	18 000	39.0
$C_{32:0}$ OH-FA	$C_{32:2}$ OH-FA + 46.5 O_2 → 32 H_2O + 32 CO_2	19 200	39.0

1167

1168

1169 [a] Source:
1170 http://chemwiki.ucdavis.edu/Theoretical_Chemistry/Chemical_Bonding/General_Principles_of_Chemical_Bond
1171 ing/Bond_Energies

1172

For Peer Review

1173 **Table 2**. Overview of the transcriptomic analyses of the strains analysed in the present study[a]

	N. gaditana CCMP526	*N. oceanica* CCMP1779
Mapped genes	10,043	11,222
Up-regulated[b]	1,950	955
Highly up-regulated[c]	440	292
Down-regulated[b]	2,067	1,855
Highly down-regulated[c]	612	1,133
Not significant	6,026	8,412

1174

1175 [a]Gene expression of the dark treatment was compared with that of the dark/light controls.

1176 [b]A gene is considered up or down-regulated if its expression level changes by at least 2-fold
1177 in the dark treatment compared to the dark/light control and the change is associated with an
1178 FDR corrected p-value < 0.01.

1179 [c]Number of down- and up-regulated genes exhibiting an expression change of at least 8 fold.

1180

For Peer Review

Figure legend

Figure 1. Growth curves depicting the cell abundance of *N. oceanica* CCMP1779 (**A**) and *N. gaditana* CCMP526 (**B**) during the experiments. The shaded areas correspond to the time interval during which the cultures were incubated under dark conditions. Cellular levels of LCAs, LCDs and LCHFAs (**C**-**D**) and major fatty acids (**E**-**F**) for *N. oceanica* CCMP1779 (**C**, **E**) and *N. Gaditana* CCMP526 (**D**, **F**).

Figure 2. Incorporation of ^{13}C-labelled sodium acetate in the biomass of *N. oceanica* CCMP1779 under dark/light (**A**, **C**), and dark (**B**, **D**) conditions. Growth curves (**A**, **B**) depicting the cell abundance of the culture during the experiment. The grey area denotes the time interval of dark incubation whereas arrows indicate the supply of ^{13}C-labelled sodium acetate. Atomic ^{13}C percentage (of measured ^{13}C+^{12}C) (**C**, **D**) measured by GC-IR-MS for selected lipids at the end of the experiment. Error bars represent the standard deviation calculated from three replicate measurements on the lipid extract (note that some error bars are too small to be visible in the graph). The straight line indicates the natural atomic ^{13}C percentage.

Figure 3. Expression level of the genes potentially coding the enzymes catalysing the different reactions involved in the biosynthesis of saturated $C_{14–18}$ fatty acids, LCHFAs, LCAs and LCDs as well as the formation of ester-bonds within the cell wall biopolymer. The horizontal axis indicates the $\log_2$ fold change in gene expression, between the dark treatment and the light/dark control. Significant differences (p-value < 0.01) are indicated with an asterisk. The prefixes *Naga* and *CCMP1779* denote transcripts from *N. gaditana* CCMP526 and *N. oceanica* CCMP1779, respectively. FAS, fatty acid synthase; Δ0-ELO, Elongation enzymes; PKS, polyketide synthase; HAD, hydroxyacyl-acyl carrier protein-dehydratase;

For Peer Review

FAE, fatty acid elongation enzyme; WS-DGAT, bifunctional wax ester synthase/diacylglycerol acyltransferase.

Figure 4. Putative pathways for the biosynthesis of LCDs and LCAs. For the biosynthetic steps in which the enzymes potentially involved have been predicted, the enzyme name is indicated next to the arrow, underlined. The lipids detected in *Nannochloropsis* spp. in this study or reported in previous studies are written in bold, whereas those that have not been observed are written in red.

Figure 5. Phylogenetic and functional analyses of *Nannochloropsis* genes potentially coding for polyketide synthases (PKS) and potentially involved in the formation of 3-OH-intermediates. Phylogeny (**A**) of the deduced amino acid sequences from the ketoacyl ACP synthase (KAS) domain of the PKSs. Sequences from different *Nannochloropsis* spp. were compared with a pre-existing alignment from Kohli et al. (2016) and the phylogenetic tree was constructed using the Maximum Likelihood (ML) algorithm by RAxML. Sequences from *N. oceanica* CCMP1779 and *N. gaditana* B31 are coloured according to their expression levels in the transcriptome and numbers in brackets after the sequences denote the expression levels of these genes in the transcriptome, expressed as $\log_2$ fold changes. Node labels indicate the bootstrap support based on ML and neighbour joining (NJ) algorithms, respectively, support values < 50% are omitted. (**B**) Structural analyses of two putative PKS from *N. oceanica* and *N. gaditana* affiliated to Clade II. The domain structure of all putative PKSs from *Nannochloropsis* is shown in details in Supplemental Table S1. The domains likely to correspond to the catalytic sites for ketoacyl-acyl carrier protein (ACP)-synthase (KAS) and ketoacyl-ACP-reductase (KAR) are in red, and green, respectively. Trans-membrane helices (TMH) are in grey, other domains in black. AT, acyl transferase; HAD,

For Peer Review

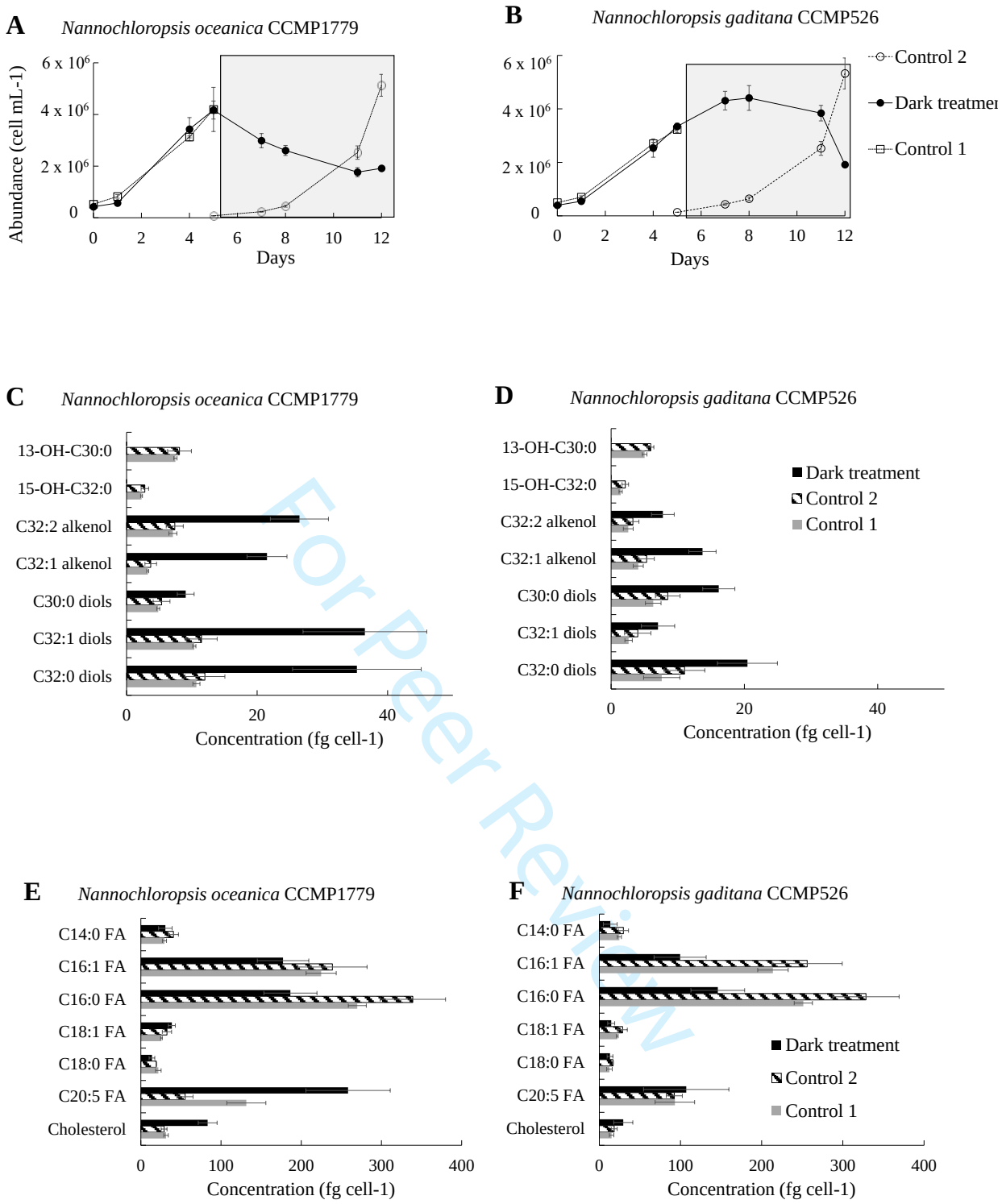
1232 hydroxyacyl dehydratase; NAD, Nicotinamide adenine dinucleotide-binding domain; ER,
1233 enoyl reductase; PPT, phosphopantetheine-binding domain; AMP, adenosine monophosphate-
1234 dependent synthetase/ligase.

For Peer Review

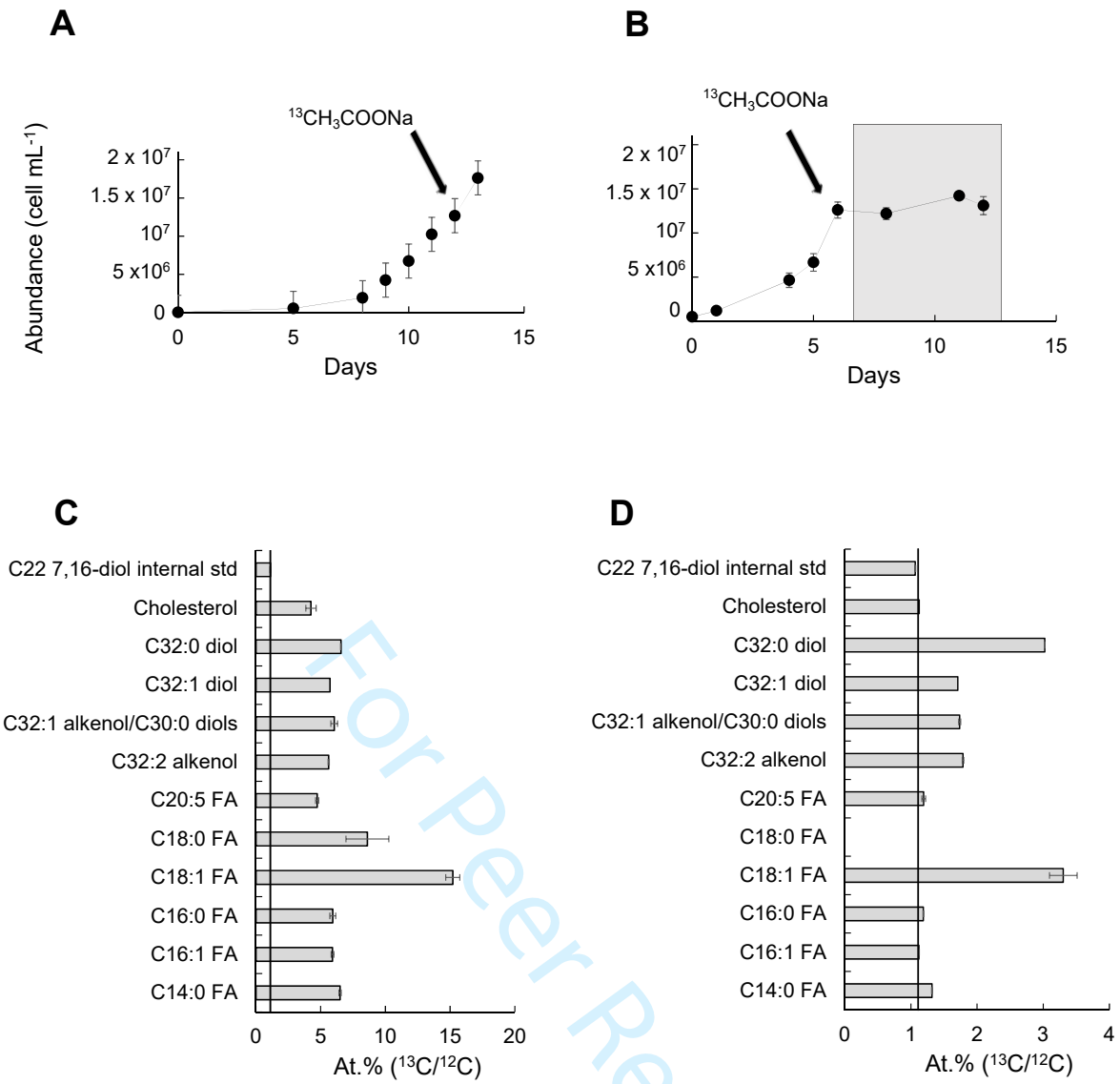
SUPPLEMENTAL MATERIAL

1) Supplemental Information: Supplemental methods and further discussions on genes/enzymes unlikely to be involved in the biosynthetic processes
2) 13 supplemental tables with legends
3) 9 supplemental figures with legends

For Peer Review



For Peer Review



For Peer Review

Figure 3

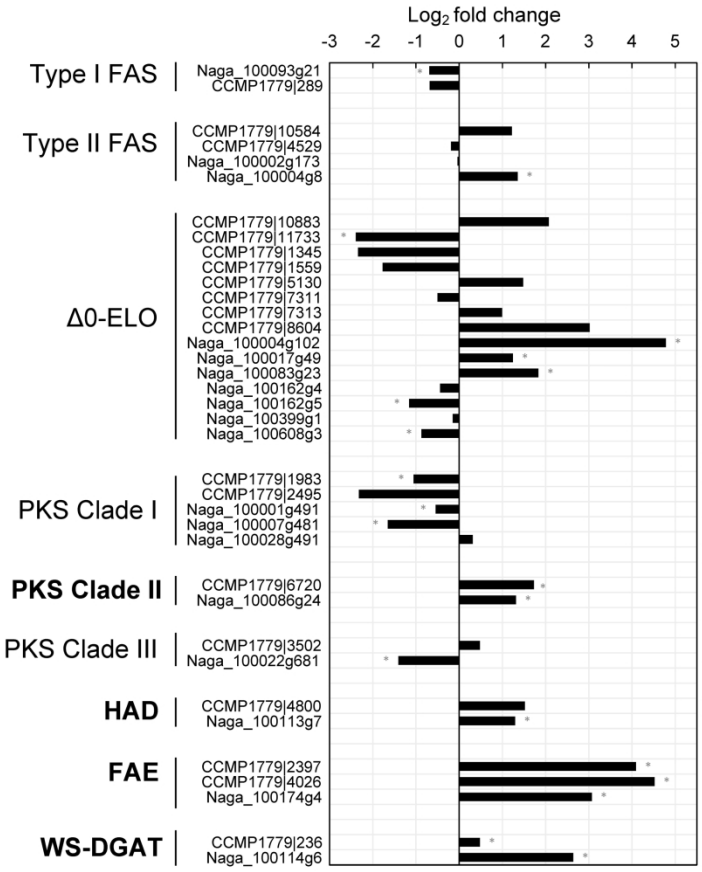
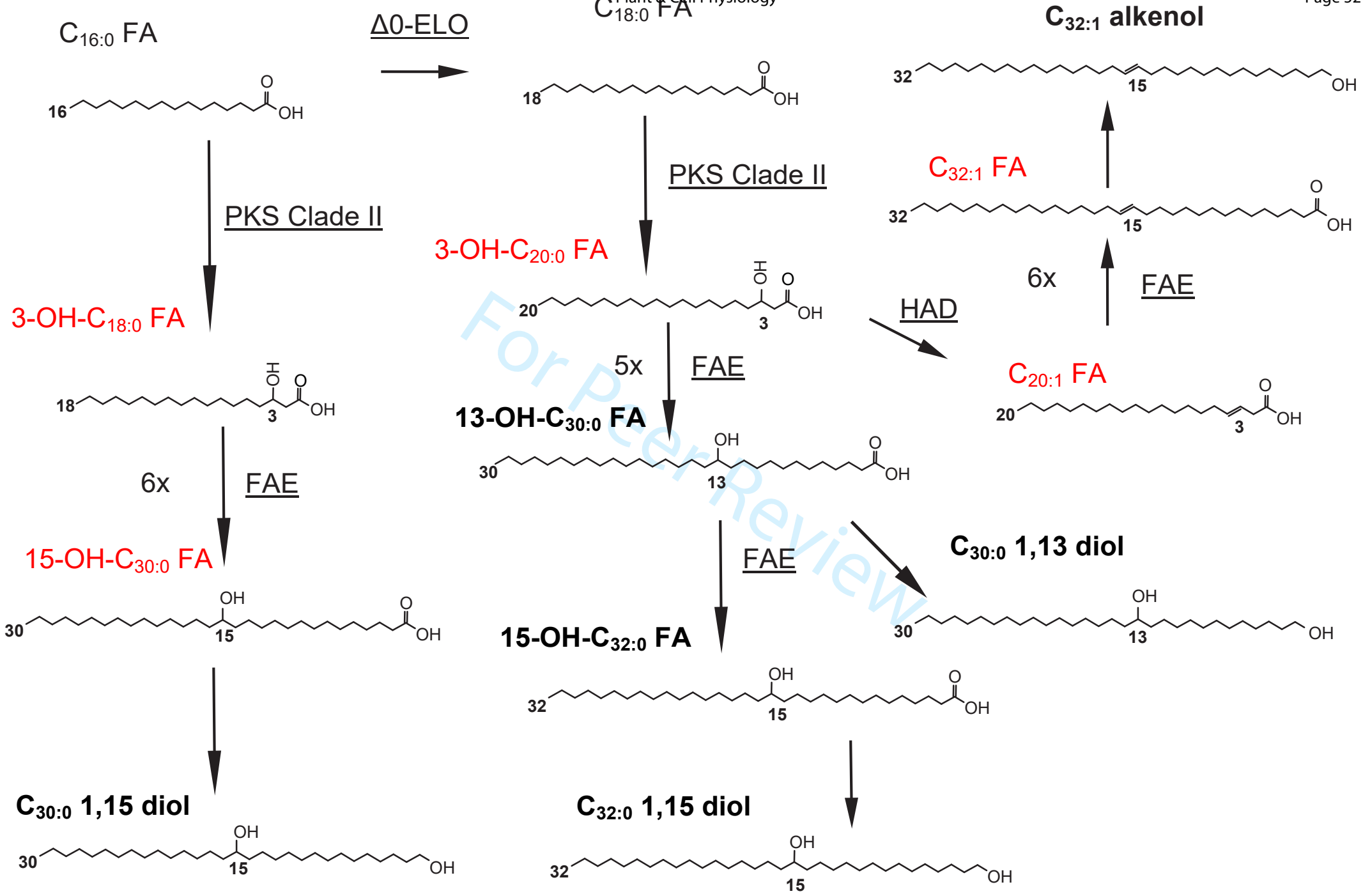


Figure 3. Expression level of the genes potentially coding the enzymes catalysing the different reactions involved in the biosynthesis of saturated C14–18 fatty acids, LCHFAs, LCAs and LCDs as well as the formation of ester-bonds within the cell wall biopolymer. The horizontal axis indicates the log2 fold change in gene expression, between the dark treatment and the light/dark control. Significant differences (p-value < 0.01) are indicated with an asterisk. The prefixes Naga and CCMP1779 denote transcripts from N. gaditana CCMP526 and N. oceanica CCMP1779, respectively. FAS, fatty acid synthase; Δ0-ELO, Elongation enzymes; PKS, polyketide synthase; HAD, hydroxyacyl-acyl carrier protein-dehydratase; FAE, fatty acid elongation enzyme; WS-DGAT, bifunctional wax ester synthase/diacylglycerol acyltransferase.

209x297mm (300 x 300 DPI)



For Peer Review

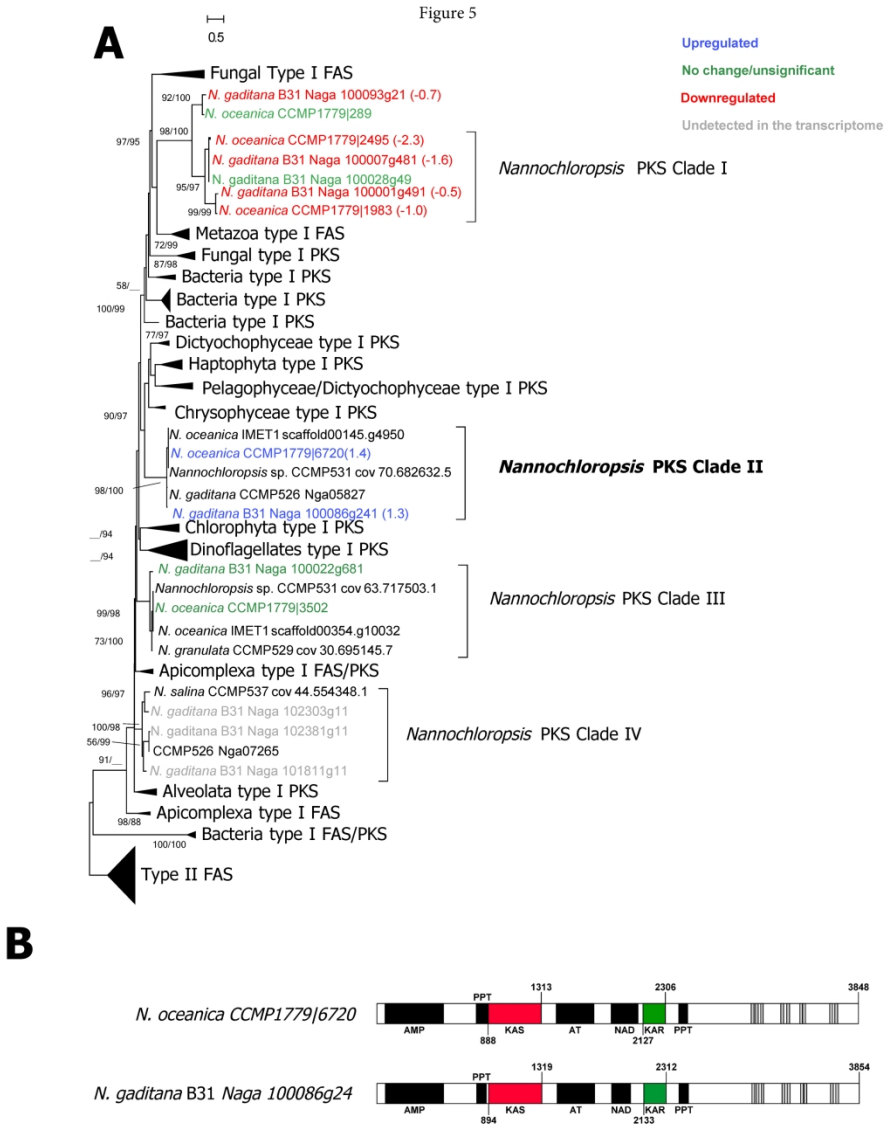


Figure 5. Phylogenetic and functional analyses of Nannochloropsis genes potentially coding for polyketide synthases (PKS) and potentially involved in the formation of 3-OH- intermediates. Phylogeny (A) of the deduced amino acid sequences from the ketoacyl ACP synthase (KAS) domain of the PKSs. Sequences from different Nannochloropsis spp. were compared with a pre-existing alignment from Kohli et al. (2016) and the phylogenetic tree was constructed using the Maximum Likelihood (ML) algorithm by RAxML. Sequences from N. oceanica CCMP1779 and N. gaditana B31 are coloured according to their expression levels in the transcriptome and numbers in brackets after the sequences denote the expression levels of these genes in the transcriptome, expressed as log 2 fold changes. Node labels indicate the bootstrap support based on ML and neighbour joining (NJ) algorithms, respectively, support values < 50% are omitted. (B) Structural analyses of two putative PKS from N. oceanica and N. gaditana affiliated to Clade II. The domain structure of all putative PKSs from Nannochloropsis is shown in details in Supplemental Table S1. The domains likely to correspond to the catalytic sites for ketoacyl-acyl carrier protein (ACP)-synthase (KAS) and ketoacyl-ACP-reductase (KAR) are in red, and green, respectively. Trans-membrane helices (TMH) are in grey, other domains in black. AT, acyl transferase; HAD, hydroxyacyl dehydratase; NAD, Nicotinamide adenine

dinucleotide-binding domain; ER, enoyl reductase; PPT, phosphopantetheine-binding domain; AMP, adenosine monophosphate-dependent synthetase/ligase.

238x289mm (300 x 300 DPI)

View publication stats