

Identification and overexpression of endogenous transcription factors to enhance lipid accumulation in the biotechnologically relevant species *Chlamydomonas pacifica*[☆]

Abhishek Gupta ^{a,g}, Aaron Oliver ^b, João Vitor Dutra Molino ^a, Kathryn M.J. Wnuk-Fink ^d, Marissa Tessman ^e, Kalisa Kang ^a, Évellin do Espírito Santo ^{a,f}, Yasin Torres-Tiji ^a, Michael D. Burkart ^d, Stephen Mayfield ^{a,c,e,*}

^a Department of Molecular Biology, School of Biological Sciences, University of California San Diego, La Jolla, CA, United States

^b Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, United States

^c California Center for Algae Biotechnology, University of California, San Diego, La Jolla, CA, United States

^d Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, United States

^e Algenesis Inc., 1238 Sea Village Dr., Cardiff, CA, United States

^f Department of Biochemical and Pharmaceutical Technology, Faculty of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil

^g Microbial and Environmental Genomics Department, J. Craig Venter Institute, La Jolla, CA, United States

ARTICLE INFO

Keywords:
Sustainability
Biotechnology
Microalgae
Transcription factors
Computational modeling
Biofuels
Sustainable Aviation Fuel

ABSTRACT

Sustainable low-carbon energy solutions are critical to mitigating global carbon emissions. Algae-based platforms offer potential by converting carbon dioxide into valuable products while aiding carbon sequestration. However, scaling algae cultivation faces challenges like contamination in outdoor systems. Previously, our lab evolved *Chlamydomonas pacifica*, an extremophile green alga, which tolerates high temperature, pH, salinity, and light, making it ideal for large-scale bioproduct production, including biodiesel. Here, we enhanced lipid accumulation in evolved *C. pacifica* by identifying and overexpressing key endogenous transcription factors through genome-wide in-silico analysis and in-vivo testing. These factors include Lipid Remodeling Regulator 1 (CpaLRL1), Nitrogen Response Regulator 1 (CpaNRR1), Compromised Hydrolysis of Triacylglycerols 7 (CpaCHT7), and Phosphorus Starvation Response 1 (CpaPSR1). Under nitrogen deprivation, CpaLRL1, CpaNRR1, and CpaCHT7 overexpression enhanced lipid accumulation compared to wild-type. However, CpaPSR1 increased lipid accumulation compared to wild-type in normal media and did not increase further under nitrogen deprivation, highlighting the difference in function based on media conditions. Notably, lipid analysis of CpaPSR1 under normal media conditions revealed a 2.4-fold increase in triglycerides (TAGs) compared to the wild-type, highlighting its potential for biodiesel production. This approach provides a framework for transcription factor-focused metabolic engineering in algae, advancing bioenergy and biomaterial production.

1. Introduction

Since the dawn of industrialization, the extensive utilization of fossil fuels has been a cornerstone of economic growth and technological advancement [1]. However, this heavy reliance has come at a significant cost to the environment, contributing to the current threat of climate change brought about by the emissions of greenhouse gases [2]. The resultant global warming, characterized by rising temperatures, melting ice caps, and extreme weather phenomena, underscores a pressing need

to transition towards sustainable energy sources [3]. The shift towards sustainable energy is not merely a response to the environmental crisis but also a prudent approach to preparing for a future where fossil fuel resources are bound to become scarcer and more expensive [4].

Microalgae hold enormous potential as a sustainable source of food, feed, and energy, presenting several advantages over traditional plant-based sources [5,6]. Their biomass is nutritionally rich, containing essential components such as lipids, proteins, and carbohydrates crucial for both human and animal nutrition. Some strains can accumulate up to

[☆] This article is part of a Special issue entitled: 'APEX' published in Algal Research.

* Corresponding author at: Department of Molecular Biology, School of Biological Sciences, University of California San Diego, La Jolla, CA, United States.

E-mail address: smayfield@ucsd.edu (S. Mayfield).

65–70 % protein, making them an excellent source of nutritional protein, while others are rich in lipids, providing a sustainable source of omega-3 fatty acids [7]. Microalgae exhibit faster growth rates than conventional plants and can be harvested throughout the year without the need to compete for arable land [8]. Moreover, microalgae are extremely efficient in absorbing carbon dioxide and converting it to useful bioproducts much faster than most traditional crop plants [9]. These traits make microalgae an attractive feedstock for producing biodiesel and Sustainable Aviation Fuel [10]. In spite of this potential, the economic viability of microalgal products is challenged by the substantial costs tied to biomass production, encompassing expensive cultivation systems, harvesting, and processing, all of which inhibit its commercial practicality [11,12].

To minimize the costs associated with biomass production and promote the adoption of microalgae-based technologies, it is essential to boost productivity both at the cellular and cultivation scales [13,14]. Increasing cellular productivity necessitates strain improvement, either through molecular modifications or breeding and selection, while rising biomass productivity on large-scale cultivation systems can be achieved through the use of strains with enhanced productivity as well as resilience to pond crashes [6,15,16]. Transcription factors (TFs) have been demonstrated to play a pivotal role in enhancing productivity in terrestrial crops like wheat and corn [17,18]. Transcription factors, proteins that oversee gene expression, can be leveraged to steer metabolic flux in a desired direction, thereby elevating biomass quality. Utilizing transcription factor-based metabolic engineering has been shown to significantly improve the quality of biomass production in microalgae [19]. For instance, several studies have employed native or plant transcription factors to augment lipids production in various microalgae strains [20–23]. Extremophile strains have evolved to endure harsh conditions and have an advantage in reduced contamination risks due to limited resource competition and predator competition in extreme environments [24,25]. Hence, implementing transcription-factor-based metabolic engineering on an extremophile alga could be a robust approach to make microalgae a viable platform for advancing a sustainable economy [26–28].

Recently, our group identified a unique green microalga, *Chlamydomonas pacifica*, capable of thriving in high-pH, high-temperature, and high-salinity conditions (CC-5697 & CC-5699) [29]. Moreover, we have also evolved this strain to have a high tolerance to light (CC-6190, 402wt x 403wt) [30]. The newly available genomic resources for this strain make it possible to conduct computational studies of transcription factors pertinent to genetic engineering. In this study, we enhanced lipid accumulation in the evolved *C. pacifica* by identifying and over-expressing endogenous transcription factors. Through a comprehensive genome-wide in-silico analysis and by in-vivo testing, we identified key endogenous putative transcription factors, including CpaLRL1, CpaNRR1, CpaCHT7, and CpaPSR1 in *C. pacifica*. Overexpression of these transcription factors induced lipid accumulation under normal or nitrogen-deplete minimal media for all four transgenic lines. Additionally, we converted the extracted lipids into biodiesel, underscoring the commercial relevance of these transgenic high-lipid strains. This approach not only positions these strains as prime candidates for large-scale biodiesel and biomaterial production but also provides a framework for applying transcription factor-focused metabolic engineering to other commercially important microalgae species.

2. Results

2.1. Overview and comparative analysis of transcription factors in *C. pacifica* with other algae, and higher plants

The identification and classification of transcription factors depend significantly upon their unique features. Essential attributes, like DNA-binding domains, are crucial for recognizing TFs [31,32]. These elements are fundamental in DNA interaction and gene regulation, which

are the primary functions of TFs. Additionally, TFs are grouped into various families based on these characteristics, with each family having distinct domains that aid in distinguishing between them and shedding light on their specific roles and regulatory processes. The transcription factor database PlantTFDB has established rules for family classification by extensively reviewing existing literature [33]. This database contains transcription factors for 165 species across 9 taxonomic groups encompassing plants and algae.

To predict all transcription factors, we employed the transcription factor prediction tool from PlantTFDB, leveraging its methodology to detect TFs in the *C. pacifica* genome. Given the abundance of chlorophyte genomes within the PlantTFDB database (16 in total), this resource is apt for forecasting TFs in the green alga *C. pacifica*. This initial method enabled us to successfully predict a total of 162 transcription factors in *C. pacifica* (Supplementary Fig. 1; Supplementary Table S1). These potential transcription factors were additionally annotated using the Clusters of Orthologous Genes (COG) database to identify broad functional potential (Fig. 1a). This analysis assigned 48 % of transcription factors to a functional category. The most common COG categories for our transcription factors are associated with post-translational modifications and signal transduction. Regulation of these pathways will define *C. pacifica* response to abiotic stresses. Few transcription factors associated with lipid and starch metabolism could be identified with this initial analysis, requiring additional techniques to predict which transcription factors would be most relevant for exploiting *C. pacifica* as a source of food and fuel.

Our findings particularly highlight the prevalence of certain TF families—namely Myeloblastosis (MYB; 11 %), MYB-related (11 %), SQUAMOSA promoter binding protein (SBP; ~10 %), Cysteine3Histidine (C3H; ~10 %), and Basic leucine zipper (bZIP; ~9 %)—which together represent over half (51.23 %) of all identified transcription factors (Fig. 1b). MYB, SBP, and bZIP TFs have been identified as key players in lipid metabolism and stress responses, instrumental in enhancing microalgae's potential for biofuel production [22,34–39]. MYB, MYB-related, and C3H transcription factors are known to be involved in the transcription regulation of astaxanthin synthesis-related genes in microalgae *Haematococcus pluvialis* [40]. Moreover, C3H TFs are known to be involved in biotic and abiotic stress [41,42].

Grasping the evolutionary and functional landscapes of transcription factors within microalgae, and their comparison with other algae species and higher plant species, is fundamental to deciphering their relational and adaptive intricacies. Here, we performed Principal Component Analysis (PCA) to compare the transcription factor profiles of *C. pacifica*, a newly identified green alga, with those of other algae such as *Nannochloropsis* and Rhodophyta (red algae), as well as a range of higher plant species, encompassing a total of 36 species and 58 transcription factor families. PCA results demonstrated that *C. pacifica* clusters closely with other Chlorophyta (green algae) species (Fig. 1c). A statistical Mann-Whitney *U* test to assess the segregation between algae and higher plant species revealing a significant difference between these two groups (PC1: statistic: 0.0, *p*-value: ~2.39e-07; PC2: statistic: 49.0, *p*-value: ~2.4e-04). These findings suggest a distinct segregation between the alga group and the higher plant species in our PCA analysis. An analysis of Chlorophyta highlighted *C. pacifica*'s transcription factor profile as closely related to other green algae, particularly *C. reinhardtii* (Fig. 1d). Spearman's rank correlation (*r* = 0.947, *p*-value = 3e-29) confirmed their strong TF profile similarity, suggesting both a shared evolutionary history as well as shared regulatory mechanisms.

Subsequently, a network plot was constructed to examine the clustering patterns of *C. pacifica* transcription factors alongside those from other Chlorophyte genomes, such as *C. reinhardtii* (Fig. 1e). Of the 162 *C. pacifica* transcription factors predicted by PlantTFDB, just 17 % of them were found in a cluster containing a transcription factor from *C. reinhardtii*. An additional Sankey plot was constructed to elucidate the transcription factor profile similarities and differences between *C. pacifica* and *C. reinhardtii* (Supplementary Fig. 2). Notably, the Sankey

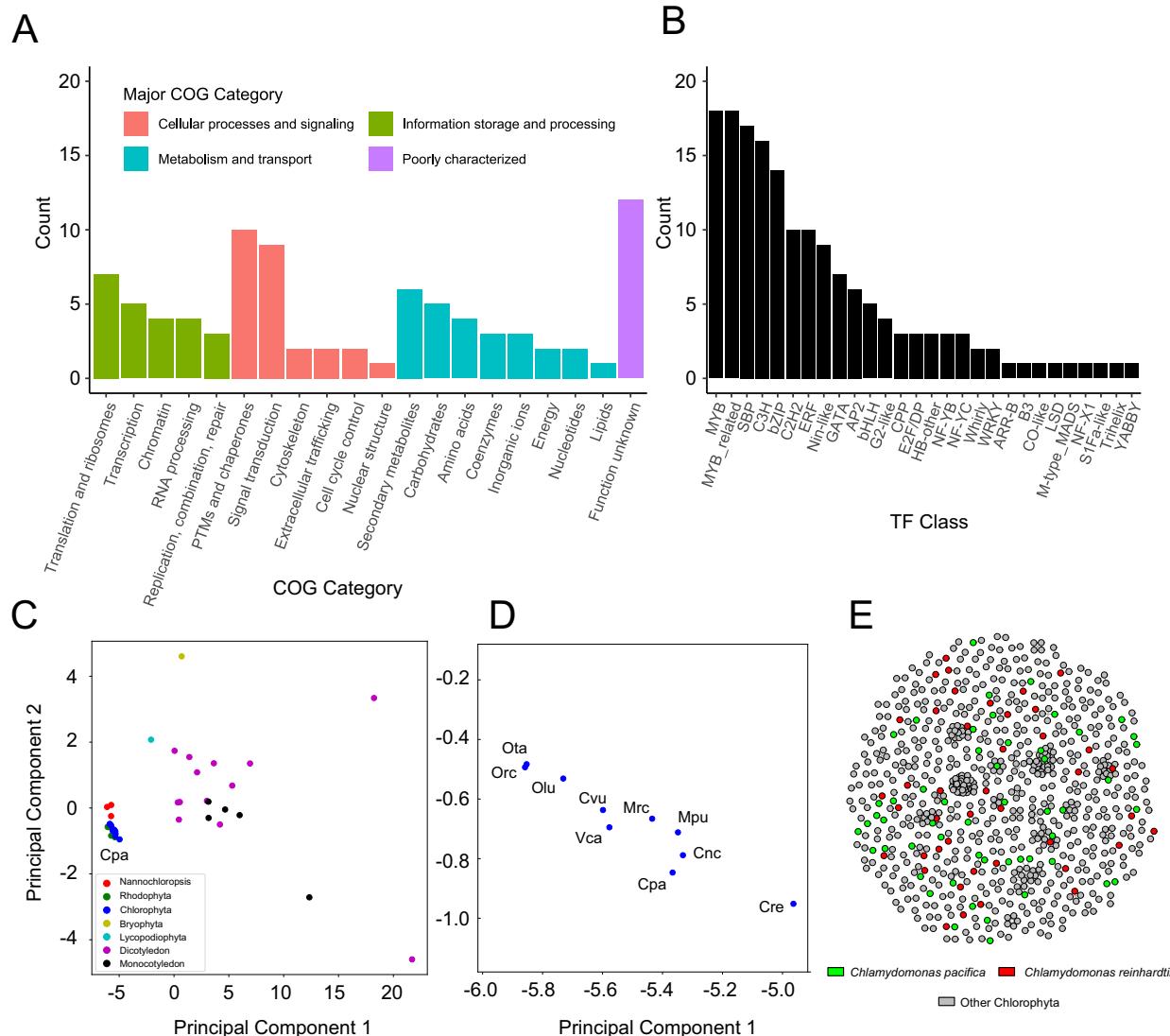


Fig. 1. Overview and comparative analysis of transcription factors in *Chlamydomonas pacifica* with other algae and higher plant species. (A) Distribution of transcription factors across different Clusters of Orthologous Genes (COG) categories, indicating their potential functions in *Chlamydomonas pacifica*. (B) Bar chart showing the distribution of transcription factors across various families in *Chlamydomonas pacifica*, highlighting the abundance of different types within the species. (C) PCA plot illustrating the transcription factor profiles of *Chlamydomonas pacifica* compared with those of various other algae and higher plant species. (D) A zoomed-in PCA plot focusing on the Chlorophyta group, highlighting the close relationship of *Chlamydomonas pacifica* with other green algae species. (E) Network plot showing the interactions between transcription factors of *Chlamydomonas pacifica*, *Chlamydomonas reinhardtii*, and other Chlorophyta species.

The labels for species in panel (C,D) are as follows: *Chlamydomonas reinhardtii* (Cre), *Chlamydomonas pacifica* (Cpa), *Volvox carteri* (Vca), *Chlorella sp. NC64A* (Cnc), *Coccomyxa sp. C-169* (Cvu), *Micromonas pusilla CCMP1545* (Mpu), *Micromonas sp. RCC299* (Mrc), *Ostreococcus lucimarinus CCE9901* (Olu), *Ostreococcus sp. RCC809* (Orc), and *Ostreococcus tauri* (Ota). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

plot revealed the absence of predicted Heat Shock Factors (HSF) and DNA binding with one finger (Dof) TF families in *C. pacifica*, which are present in *C. reinhardtii*. The observed discrepancies could arise from limitations in the prediction technique, inaccuracies in genome assembly or sequencing, or the actual absence of these transcription factor families, underscoring the need for further investigation.

2.2. High-throughput approach to identify lipid enhancing transcription factors

To pinpoint TFs with the potential to boost lipid production, we synthesized data from multiple transcription factor prediction tools (Fig. 2, see Methods). We started by identifying all TFs using two additional databases, iTAK and TAPscan, alongside PlantTFDB. iTAK and TAPscan identified 182 and 183 TFs, respectively (Supplementary

Tables S2-S3). Subsequently, we selected only those TFs consistently predicted by all three databases, yielding 139 common TFs (Supplementary Table S4). Next, these commonly predicted TFs were compared via phylogenetic trees and active domain alignment with a list of known TFs involved in lipid metabolism, as reported in the existing literature [21]. This identified 10 transcription factors in *C. pacifica* that are orthologs of TFs found in *Arabidopsis thaliana*, *Auxenochlorella protothecoides*, *C. reinhardtii*, *Dunaliella parva*, and *Scenedesmus obliquus* (Fig. 2, Table 1). These TFs are categorized into the AP2, NF-YB, MYB, bZIP, CPP, and SBP families.

To enhance confidence in our prediction, we investigated several of the predicted transcription factors in-depth. For instance, our analysis identified a transcription factor in *C. pacifica* with significant resemblance to *C. reinhardtii*'s PSR1 and *Arabidopsis thaliana*'s PHR1, characterized by MYB and coiled-coil domains (Fig. 3a; percentage identity =

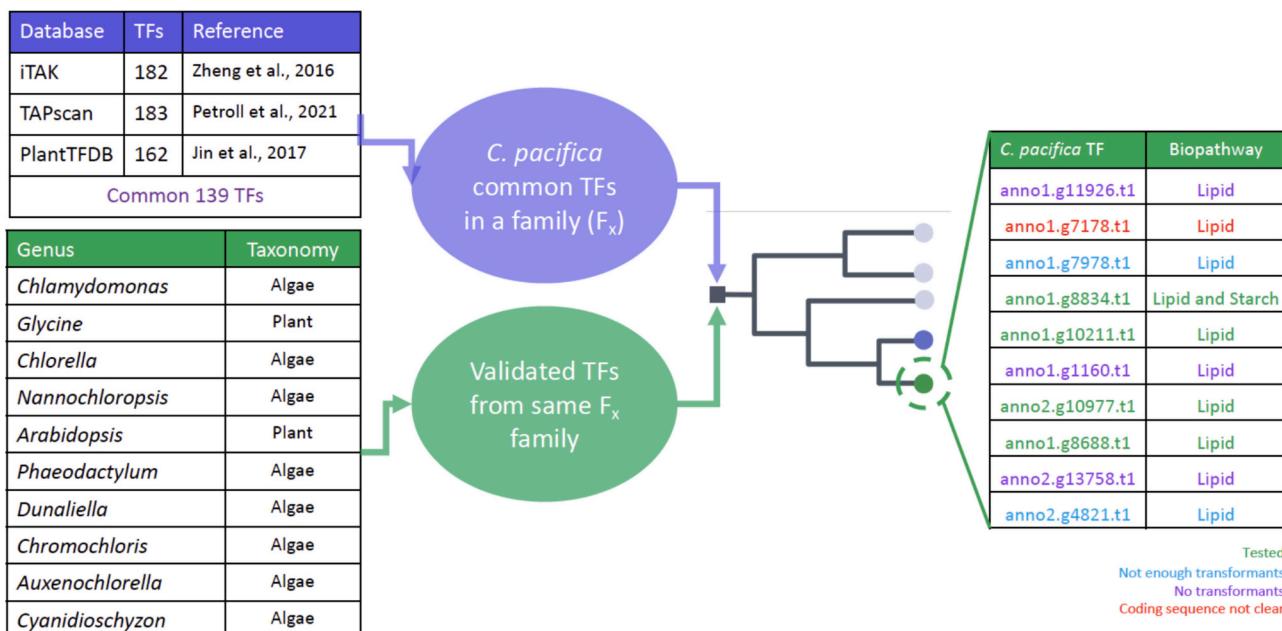


Fig. 2. Pipeline for predicting transcription factors involved in starch and lipid biosynthesis in *C. pacifica*.

The schematic illustrates the prediction of transcription factors involved in starch and lipid biosynthesis in *C. pacifica*. In Step 1, transcription factors were predicted using three different tools, and a comprehensive database of validated transcription factors was created from 10 different genera based on literature. In Step 2, a phylogenetic tree analysis was conducted on the transcription factors from *C. pacifica* and the curated validated database within the same transcription factor family. This analysis resulted in the identification of 10 predicted transcription factors involved in the biosynthetic pathways. The colour coding of various transcription factors signifies whether they have been experimentally tested in this study.

Table 1

Predicted transcription factors associated with lipid and starch synthesis enhancement in *Chlamydomonas pacifica*.

Microalgae species	Source species	Transcription factor	Family	C. pacifica Candidate	Biopathway	Reference
<i>Nannochloropsis salina</i>	<i>Arabidopsis thaliana</i>	WRINKLED1 (AtWRI1)	AP2	anno1.g11926.t1	Lipid	Kang et al., 2017
<i>Chlorella ellipoidea</i>	<i>Arabidopsis thaliana</i>	Leafy Cotyledon1 (AtLEC1)	NF-YB	anno1.g7178.t1	Lipid	Liu et al., 2021
<i>Auxenochlorella protothecoides</i>	<i>Auxenochlorella protothecoides</i>	Myeloblastosis 6 (ApMYB6)	MYB	anno1.g7978.t1	Lipid	Xing et al., 2021
<i>Chlamydomonas reinhardtii</i>	<i>Chlamydomonas reinhardtii</i>	Phosphorus Starvation Response 1 (PSR1)	MYB	anno1.g8834.t1	Lipid and Starch	Bajhaiya et al., 2016; Ngan et al., 2015
<i>Chlamydomonas reinhardtii</i>	<i>Chlamydomonas reinhardtii</i>	Lipid Remodeling Regulator 1 (LRL1)	MYB	anno1.g10211.t1	Lipid	Hidayati et al., 2019
<i>Chlamydomonas reinhardtii</i>	<i>Chlamydomonas reinhardtii</i>	Basic Leucine Zipper 1 (CrbZIP1)	bZIP	anno1.g1160.t1	Lipid	Yamaoka et al., 2019
<i>Chlamydomonas reinhardtii</i>	<i>Chlamydomonas reinhardtii</i>	Compromised Hydrolysis of Triacylglycerols 7 (CHT7)	CPP	anno2.g10977.t1	Lipid	Tsai 2014
<i>Chlamydomonas reinhardtii</i>	<i>Chlamydomonas reinhardtii</i>	Nitrogen Response Regulator 1 (NRR1)	SBP	anno1.g8688.t1	Lipid	Boyle et al., 2012
<i>Dunaliella parva</i>	<i>Dunaliella parva</i>	WRINKLED1 (DpWRI1)	AP2	anno2.g13758.t1	Lipid	Shang et al., 2016
<i>Scenedesmus obliquus</i>	<i>Scenedesmus obliquus</i>	SQUAMOSA-promoter binding protein-like 12 (SoSPL12)	SBP	anno2.g4821.t1	Lipid	Calhoun et al., 2023

~63 %, $p < 1e-05$, permutation-test). PSR1 is known for its regulatory function in lipid and starch biosynthesis during nitrogen and phosphorus deficiency in *C. reinhardtii*, and PHR1 is implicated in triacylglycerol accumulation under phosphorus starvation in *Arabidopsis thaliana* [43–45]. Furthermore, we have pinpointed another TF that bears resemblance to the lipid-enhancing WRINKLED1 (WRI1) transcription factor, as observed in *A. thaliana* and *Helianthus annuus* (Supplementary Fig. 3; percentage identity = ~56 %, $p < 1e-05$, permutation-test) [46,47]. This TF maintains the key characteristics of the AP2 DNA binding domain, the VYL transcriptional activation motif, and two phosphorylation sites essential for protein stability [48].

As the *C. reinhardtii* PSR1 TF has been shown to affect starch metabolism, we extended our investigation to include orthologs of genes

involved in the *C. reinhardtii* starch pathway [43]. In this pursuit, we successfully identified orthologs for eight genes reported previously within the genome of *C. pacifica* (Fig. 3b, Supplementary Table S5), whose functions were confirmed with EggNOG [43]. A noteworthy discovery emerged from examining the native promoters of these genes in *C. pacifica*. Two of these gene promoters exhibited motifs that are strikingly similar to the consensus motif of PSR1 as identified in the PlantTFDB database (Fig. 3b). This similarity in promoter Transcription factor binding sites (TFBSs) suggests a potential regulatory relationship with PSR1 in *C. pacifica* akin to that observed in *C. reinhardtii*.

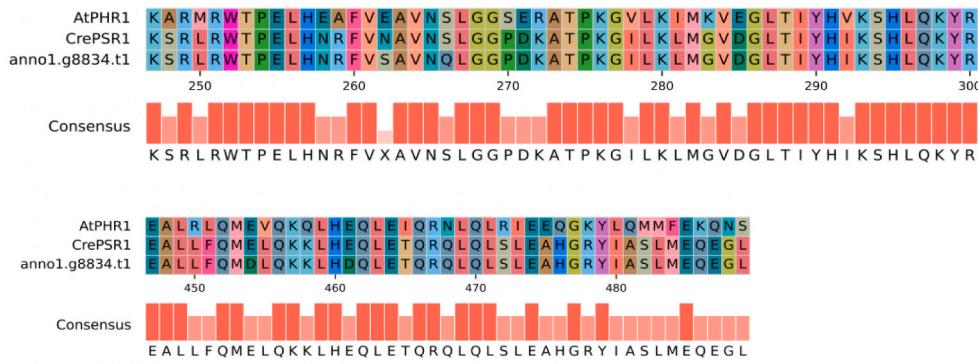
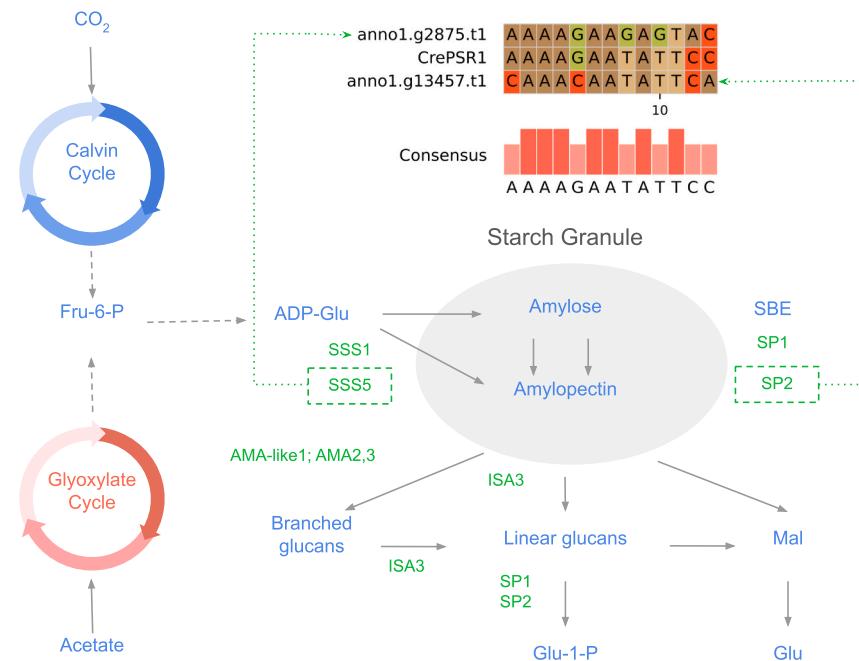
A**B**

Fig. 3. Conservation of transcription factor domain and genes in starch metabolism.

(A) Alignment showing MYB and coiled-coil domains from *Chlamydomonas reinhardtii*'s PSR1 (CrePSR1), *Arabidopsis thaliana*'s PHR1 (AtPHR1), and an orthologous predicted CpaPSR1 transcription factor from *Chlamydomonas pacifica* (anno1.g8834.t1), highlighting amino acid conservation. (B) An illustration of orthologous genes related to starch synthesis and breakdown is marked in green text. It features the enriched motifs within the promoter sequences of two of these genes that are similar to the consensus motif sequence of CrePSR1. The schematic is adapted from Bajhaiya et al., 2016. The enzyme label SBE represents Starch Branching Enzyme, and the orthologous gene definitions are described in Supplementary Table S5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. Overexpression of predicted endogenous transcription factors

To validate our predictions, we experimentally tested them in the evolved *C. pacifica* strain to determine if overexpression could enhance lipid accumulation. We cloned 9 out of 10 transcription factors under the β 2 tubulin promoter sequence with hygromycin selection (see Methods). One transcription factor gene (anno1.g7178.t1) was not well annotated and thus was not cloned (Fig. 2). Among the 9 TFs, we successfully obtained transformants for 6, and only 4 provided enough transformants for screening. The absence of transformants in some TF lines may be attributed to using the upstream β 2 tubulin promoter, whose strong and constitutive expression may prove to be lethal. To address this issue, employing an inducible promoter for the overexpression of TFs could provide a more controlled and effective approach. We proceeded with these 4 TFs (anno1.g10211: CpaLRL1, anno1.g8688: CpaNRR1, anno2.g10977: CpaCHT7, anno1.g8834: CpaPSR1) for further evaluation. We first confirmed the transformants through colony PCR and sequencing

(see Methods; Supplementary Fig. 5). Subsequently, RT-qPCR analysis revealed that transcript levels of CpaLRL1, CpaNRR1, CpaCHT7, and CpaPSR1 were elevated by 3.13-fold, 3.41-fold, 6.41-fold, and 10.74-fold, respectively, in their corresponding overexpression lines compared to the wild-type under normal growth conditions.

First, we compared the growth of the transgenic lines to the wild-type evolved strain, as overexpression of a transcription factor can impede the growth of the cells. However, we observed no significant differences in growth ($p > 0.05$ ANOVA for all pairwise differences), as measured by culture density and chlorophyll fluorescence, between the transgenic lines and wild-type cells (Fig. 4). Next, we evaluated lipid accumulation in the transgenic lines using Nile Red lipid-binding dye and flow cytometry in the phycoerythrin (PE) channel under normal and nitrogen-deprived minimal media conditions (Fig. 5a, b; see Methods). A higher response in this channel corresponds to greater lipid content in the stained cells. We compared the population percentage above a threshold of 7.7×10^4 RFU (Relative Fluorescence Units). This threshold

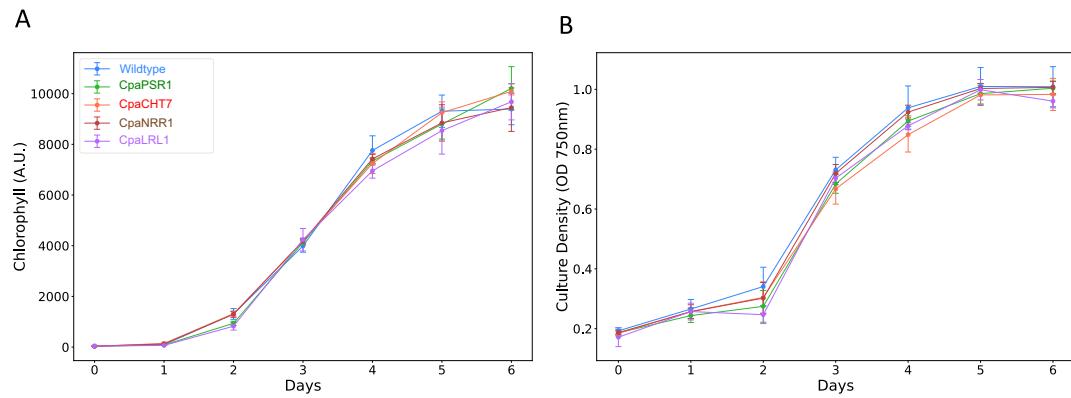


Fig. 4. Cell growth measured as chlorophyll fluorescence and cell density in transgenic lines.

(A), (B) Growth curves depicting chlorophyll levels and culture density of transcription factor overexpressed lines compared to the wildtype strain, respectively. Each line represents three replicates. A.U. stands for arbitrary fluorescence units. O.D. stands for optical density.

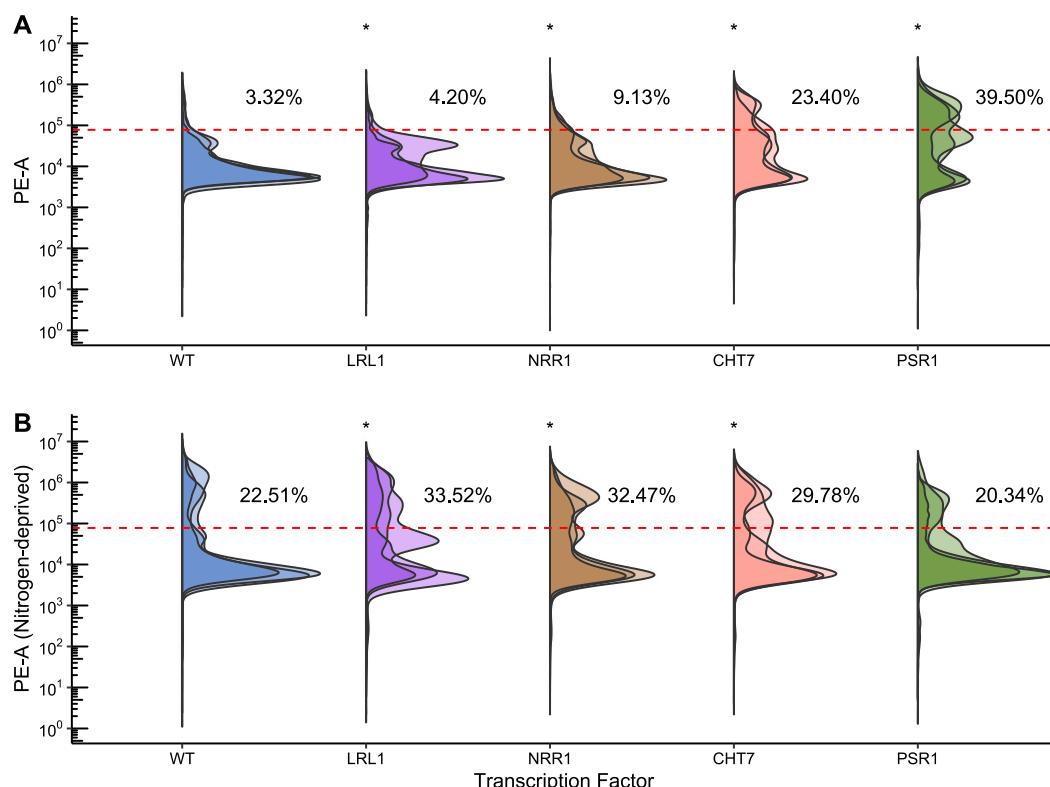


Fig. 5. Flow cytometry data showing enhanced lipid production.

(A), (B) Flow cytometry analysis of Nile red-stained cells was conducted to quantify lipid content in transcription factor overexpressed lines compared to the wild-type strain under normal or nitrogen-deprived minimal media conditions. The statistics displayed in the top right corner of each plot correspond to populations above the threshold of 7.7×10^4 . PE-A refers to the phycoerythrin-area channel. Each plot shows curves corresponding to three biological replicates. ** represents significant upward differences of the entire distribution from wild-type (Wilcoxon ranked sum test, $p < 0.001$) and each media is compared separately. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

value was selected to more effectively distinguish a specific subpopulation of cells (see Methods). Approximately 3.32 % and 22.5 % of the wild-type population exceeded this threshold under normal and nitrogen-deprived conditions, respectively (Fig. 5a, b). All the transcription factors resulted in statistically significant (Wilcoxon ranked-sum, $p < 0.05$) increase of lipid accumulation in minimal media. In addition, CpaLRL1, CpaNRR1, CpaCHT7, and CpaPSR1 respectively displayed 1.3, 2.8, 7.0, and 11.9 times more observations above the nile red cutoff threshold compared to the wildtype (Fig. 5a). Under nitrogen-deprived conditions, lipid accumulation also increased in CpaLRL1,

CpaNRR1, and CpaCHT7 but showed no statistically significant increase for CpaPSR1 compared to wild-type, suggesting different roles for these transcription factors under nitrogen stress (Fig. 5b). Confocal microscopy corroborated our flow cytometry results, with visual inspection showing more lipid-producing cells in the transgenic strains compared to the wild-type under both media conditions (Fig. 6).

To further examine the differences in lipid accumulation between the wild-type and the promising CpaPSR1 under normal conditions, semi-quantitative lipidomic analysis was conducted on the two strains. Through this analysis, the relative abundance of triglycerides (TAGs)

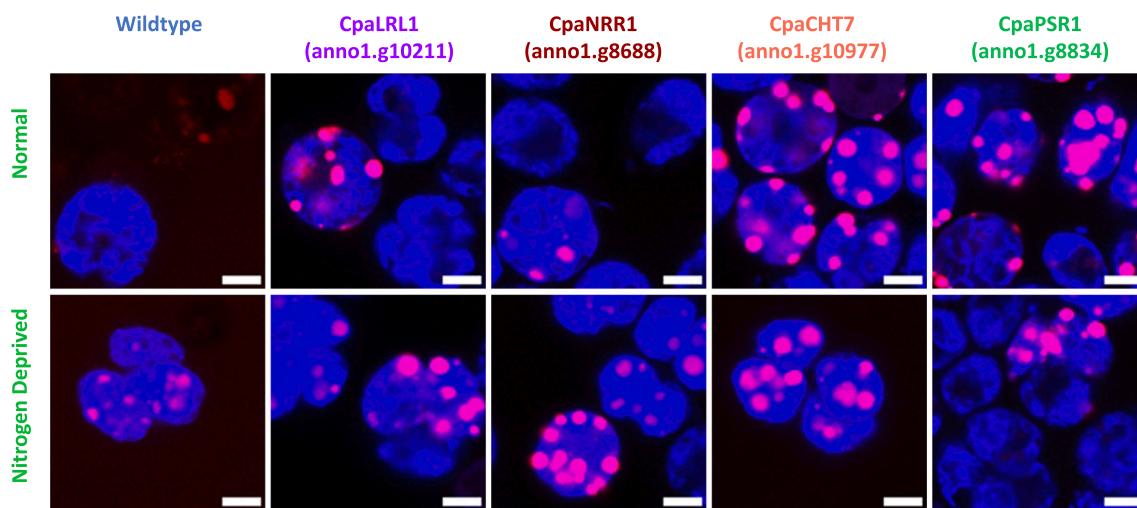


Fig. 6. Confocal microscopy of cells in different media conditions.

Cells in normal minimal and nitrogen-deprived media, respectively. The scale bar represents 5 μ m. Blue and magenta correspond to Chlorophyll and triacylglycerol, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and diglycerides (DAGs) were determined for three replicates (see Method). For the wild-type, the total TAG content was found to be 152 ± 21 relative abundance/mg biomass, while the CpaPSR1 contained 369 ± 70 relative abundance/mg biomass. The diglyceride analysis showed values of 868 ± 128 relative abundance/mg biomass for the wild-type while the CpaPSR1 samples had DAG content of 241 ± 102 relative abundance/mg biomass. The observed increased accumulation of neutral TAGs in the CpaPSR1 sample has favorable implications for the cellular function and industrial relevance of this strain as TAGs provide more fatty acid chains per molecule compared to DAGs that can be used for biodiesel production.

Overall, the flow cytometry and microscopy, and lipidomics data validated our predictions and provided new insights into the functioning of these transcription factors under both normal and nitrogen-stress conditions. Notably, CpaPSR1 exhibited the highest lipid accumulation under normal media conditions, eliminating the need for time-

consuming and resource-intensive nitrogen deprivation setups. Hence, CpaPSR1 is an excellent candidate for large-scale biodiesel production because it enables a continuous growth cycle for biomass harvesting.

2.4. Biodiesel production

To demonstrate commercial applications and understand potential alterations in fatty acids produced under varied growth conditions and transcription factor overexpression, lipids from *C. pacifica* were extracted and neutral TAGs were converted into fatty acid methyl ester (FAME) biodiesel and analyzed using Gas Chromatography-Mass Spectrometry (GC-MS). Following lipid extraction and esterification, the components of each final biodiesel were identified, and the FAMEs were sorted based on their chain length and degree of unsaturation. All final samples contained trace amounts of “non-biodiesel” nonpolar lipid components such as cholesteryl esters. However, purity remained high

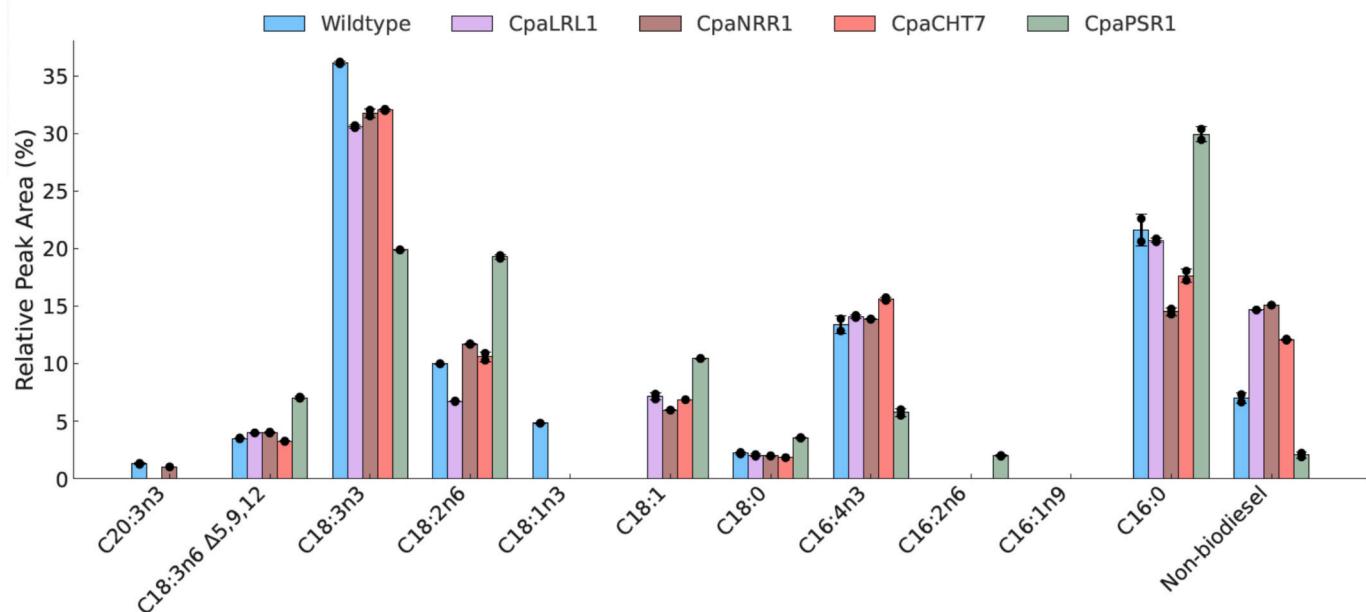


Fig. 7. GC-MS analysis of alga-derived FAME biodiesel.

Comparison of biodiesel composition of overexpressed lines with evolved wildtype *C. pacifica* in normal minimal media through relative peak area integration percentages in gas chromatography-mass spectrometry analysis ($n = 2$).

across all samples, ranging from 85 to 97 % without any added purification steps (Fig. 7, Supplementary Fig. 6). Samples grown under nitrogen-starvation conditions exhibited a reduced proportion of “non-biodiesel” components in the final samples compared to those grown in normal media with percent purity of biodiesel FAME components ranging from 95 to 98 % (Fig. 7, Supplementary Fig. 6). This observation aligns with existing reports on common microalgal responses to nitrogen-deprivation; growth of microalgae in nitrogen-deplete media has been observed to induce neutral lipid TAG production due to potential shifts in cellular biosynthetic pathways that normally rely on the presence of nitrogen [49–51]. Though the increased TAG accumulation in nitrogen-starved samples correlated with higher overall FAME biodiesel purity, the high purity of biodiesel from all samples with over-expressed transcription factors is a promising indicator of their potential for biodiesel production. A major challenge in use of wildtype microalgae as feedstocks for biodiesel is the lack of TAG accumulation compared to other plant-based oil feedstocks [52]. This was observed in previous experiments to produce biodiesel from wild-type evolved *C. pacifica*, where fatty acids needed to initially be isolated from total lipid extracts to produce biodiesel at this level of purity [30]. The high purity of biodiesel combined with the observed increased TAG accumulation under different conditions in this experiment suggests that these newly modified strains of *C. pacifica* could be a promising source of microalgal TAGs for biodiesel production.

For initial analysis, the FAME profile of the original evolved *C. pacifica* strain developed by Gupta & Molino et al. (CC-6190 402wt x 403wt) was compared to CpaLRL1, CpaNRR1, and CpaCHT7 grown in minimal media (Fig. 7). While no significant differences in fatty acid profiles were seen between these samples, a secondary comparison with biodiesel from CpaLRL1, CpaNRR1, and CpaCHT7 grown under nitrogen-deplete media revealed promising changes in the fatty acid composition (Supplementary Fig. 6). Notably, a decrease in the production of C18:3n3 and C16:4n3 polyunsaturated biodiesel in favor of the production of C16:0 and C18:1 biodiesel was seen. The biodiesel composition of the original evolved *C. pacifica* strain was also compared to FAME profiles from CpaPSR1 grown under normal media conditions (Fig. 7). A similar reduction in polyunsaturated fatty acids was seen in the CpaPSR1 compared to the evolved strain, and the C16:0 FAME production in the CpaPSR1 grown in normal media was the highest across all samples (Fig. 7).

3. Discussion

Transcription factor-based metabolic engineering has proven to be highly effective in enhancing biomass quality in microalgae [19]. Notably, numerous studies have employed both exogenous and endogenous transcription factors to successfully increase lipid production across various microalgal strains [36,43,47,53]. In this study, we have demonstrated a high-throughput pipeline for identifying and validating lipid-enhancing transcription factors in a new microalga. We used *C. pacifica* as an example species due to its great potential as a commercially relevant green extremophile alga. Previously, we showed that overexpression of an exogenous soybean-based *Dof* transcription factor led to increase in lipid accumulation in *C. pacifica* under nitrogen-deprived conditions compared to the wild-type strain [30]. Here, we identified more potent endogenous transcription factors that result in higher lipid accumulation in minimal media compared to the wild-type. This discovery not only avoids cell starvation which can lead to lower biomass productivity, but also saves time and resources needed to create nitrogen-deprived conditions for large-scale cultivation.

This study successfully presented an in-silico high-throughput pipeline designed to identify lipid-enhancing transcription factors in microalgae (Fig. 2). Our approach involved predicting all transcription factors within the genome of the target species and subsequently comparing these predictions with a curated database of validated transcription factors from ten different genera, as reported in the literature

[21,34,43,54–59]. The novelty of this approach lies in its ability to streamline the identification process, significantly reducing the time and resources typically required for experimental validation. By applying and validating this pipeline in the newly discovered species *C. pacifica*, we demonstrated its robustness and reliability. The successful identification and experimental validation of lipid-enhancing transcription factors in *C. pacifica* not only showcase the efficacy of our approach but also instill confidence in its application to other commercially relevant microalgae species. Given the increasing demand for sustainable biofuels and bioproducts, this pipeline could play a pivotal role in advancing microalgal biotechnology by accelerating the discovery of key genetic regulators that enhance lipid production.

Analysis of transcription factor distributions across species (Fig. 1) supports existing data that multicellular land plants possess a significantly larger TF repertoire compared to unicellular algae like *Chlamydomonas reinhardtii* [60]. The clear demarcation between algae and higher plants in the PCA plot underscores the evolutionary divergence in the TF complexity and possibly reflects the distinct regulatory needs of unicellular and multicellular organisms. In contrast to other green algae, *C. pacifica*'s alignment with *C. reinhardtii* underscores potential evolutionary conservation in their TF composition. This relationship implies that the breadth of prior work involving *C. reinhardtii* can inform studies on *C. pacifica*, aiding green algae biotechnological advancements. Conversely, *C. pacifica* exhibited the presence of TFs from the Trihelix family, a group not as extensively studied in green microalgae. In higher plants, Trihelix TFs are known to coordinate genes related to salicylic acid metabolism, a pathway integral to plant stress responses [61,62]. Identifying Trihelix TFs in *C. pacifica* opens new avenues for research, particularly in understanding their functional roles in microalgae. This finding could potentially bridge gaps in knowledge regarding the regulatory mechanisms in green algae, especially those linked to environmental stress responses.

Next, to evaluate the full potential of the predicted TFs, we overexpressed them in *C. pacifica* using the β 2 tubulin promoter sequence with hygromycin selection. We successfully obtained transformants for four out of the nine transcription factors that were cloned for overexpression. The partial success might be due to the β 2 tubulin promoter being too strong, resulting in excessive expression of the transcription factor and potential lethality for cell growth [63]. As *C. pacifica* is a newly discovered species with limited molecular tools, this study did not prioritize testing alternative promoters to explore potential variation among different transformation clones. However, the identified TFs are promising candidates for further testing under inducible promoters [64,65]. Nonetheless, we did not observe any negative impact on the transgenic lines of the four tested TFs (Fig. 4). Using cytometry and confocal microscopy, we observed improved or unchanged lipid accumulation in all four-transcription factor overexpressing lines under normal or nitrogen-deprived media (Figs. 5, 6). To further boost lipid accumulation, future efforts can employ mutagenesis and breeding techniques on these transgenic lines [66].

Previously, the LRL1 mutant was known to suppress growth and lipid accumulation under phosphorus stress due to its role in lipid remodeling [58]. In our study, we demonstrated that overexpression of CpaLRL1 does not affect cell growth and leads to higher lipid accumulation under nitrogen-deprived conditions, indicating its role in managing nitrogen stress. Similarly, the NRR1 mutant in *C. reinhardtii* was shown to reduce TAG levels by 50 % under nitrogen starvation [59]. In accordance, our study found that overexpressing this transcription factor increases lipid accumulation under nitrogen-starvation conditions. Moreover, the CHT7 mutant is known to impact lipid degradation and cell regrowth when nitrogen is resupplied [57,67]. In our study, we observed that overexpression of CpaCHT7 increases lipid production under normal media conditions. Additionally, the PSR1 knockdown line has been shown to inhibit lipid accumulation and play a role in starch metabolism. Overexpression of PSR1 in a cell-wall-deficient strain increases starch content without changing or reducing neutral lipid levels [43].

This also aligns with findings from another study showing that PSR1 overexpression in normal wild-type strains of *C. reinhardtii* also results in higher intracellular lipid content [44]. We found that overexpressing PSR1 in *C. pacifica* increases lipid accumulation when nutrients were not limiting, but did not improve lipid accumulation when nitrogen is limited. Collectively, these results suggest the role of PSR1 in lipid biosynthesis is strain-dependent. It is unclear whether these differences are due to the presence of a cell wall, or if they result from other variations in the genetic backgrounds of the strains used.

To fully demonstrate the potential of the engineered lines, we proceeded to use their biomass for biodiesel production. A current challenge in producing high-quality diesel from biobased resources lies in feedstocks' natural fatty acid composition compared to the ideal chemical composition of high-performance diesel [68]. An ideal biodiesel formulation should have a high cetane number (CN), indicating a faster rate of fuel ignition upon injection into the engine and complete combustion while maintaining high oxidative stability and moderate viscosity [69]. Per the regulations set in ASTM standard D6751-24 for bio-based middle distillate fuels (jet fuel and diesel fuel), biodiesel fuel blend stocks should possess a minimum cetane value of 45, kinematic viscosity between 1.9 and 6 mm²/s at 40 °C, and a minimum oxidative stability of 3 h at 110 °C [70]. The FAME analysis of the evolved *C. pacifica* in this study is consistent with our previous findings and closely resembles the profile of *C. reinhardtii*, with the highest compositions being C16:0 and C18:3n3 [30,71]. Moreover, the observed shifts in fatty acid composition in samples with overexpressed lipid-associated transcription factors are relevant towards improved properties of biodiesel. Comparing above-mentioned metrics to the fatty acids produced by *C. pacifica* in this study, the increased production of C16:0 and C18:1 in the CpaPSR1 samples and nitrogen-starved CpaLRL1, CpaNRR1, and CpaCHT7 have favorable implications for the production of high-quality algal biodiesel. The metrics for C16:0 methyl palmitate (CN = 86, 4.38 mm²/s, >24 h stability) and C18:1 methyl oleate (CN = 57, 4.51 mm²/s, 2.79 h stability) demonstrate improvements towards the ASTM standard for biofuel compared to the C18:3n3 methyl linolenate (CN = 23, 3.14 mm²/s, 0.00 h stability) that comprised the majority of the biodiesel from the initial evolved *C. pacifica* strain [69,72]. Paired with the high lipid accumulation observed under normal media conditions, the high abundance of C16:0 methyl ester biodiesel from the CpaPSR1 sample indicates that this strain of *C. pacifica* could be a useful tool for

industrial-scale production of sustainable fuel. However, the finding differs from the fatty acid composition observed in PSR1 overexpression lines in *C. reinhardtii*, which could be attributed to species-specific differences or variability among clones [44,73,74]. Future studies will evaluate the properties of *C. pacifica* biodiesel extracts with respect to ASTM D6751 standards for biodiesel fuel blends, including kinematic viscosity, specific gravity, cetane number (CN), cloud point, iodine value (IV), long-chain saturated factor (LCSF), and cold filter plugging point (CFPP) [70,75]. This will allow for a full assessment of transgenic *C. pacifica* strains for industrial-level biodiesel production.

Nonetheless, leveraging these prior findings of the functionality of PSR1, CHT7, NRR1, and LRL1 in *C. reinhardtii* in the context of new genomic resources for *C. pacifica* reveals how these transcription factors can be used in tandem to influence the metabolic processes of a new algal strain. Complete pathways for TAG and starch biosynthesis were found in the *C. pacifica* genome based on homology to other *Chlamydomonas* genomes and KEGG orthologs (Fig. 8). *C. pacifica* genes relevant to TAG biosynthesis are identified in Supplementary Table S6. PSR1 and CHT7 appear to improve starch and TAG accumulation, respectively, by downregulating genes associated with their utilization [43]. Our confocal microscopy supports prior observations that CHT7 improves lipid accumulation even in the presence of a nitrogen source (Fig. 6). Prior transcriptomic work on NRR1, LRL1, and PSR1 suggests that these transcription factors drive TAG accumulation primarily through the upregulation of the Kennedy pathway rather than the upstream production of free fatty acids [43,58,59]. Thus, additional transcription factors driving type II fatty acid synthase systems could be used with the transcription factors in this study to improve TAG biosynthesis further. CHT7 preventing utilization of TAGs may explain why it appears to be more effective in nutrient rich conditions (Fig. 5a) compared to the other cell lines, as additional TAGs produced from upregulation by PSR1, NRR1, or LRL1 might cause additional TAG hydrolysis and lessen TAG buildup. Consistency of transcription factor behavior between *Chlamydomonas* species is encouraging for industrial exploitation, as novel strains with valuable characteristics such as light and halotolerance can undergo genetic engineering using the wealth of knowledge gained from *Chlamydomonas reinhardtii*.

In this study, we have tested and demonstrated several transcription factors that enhance lipid production within the same species. Future studies could focus on generating additional lines for these transcription

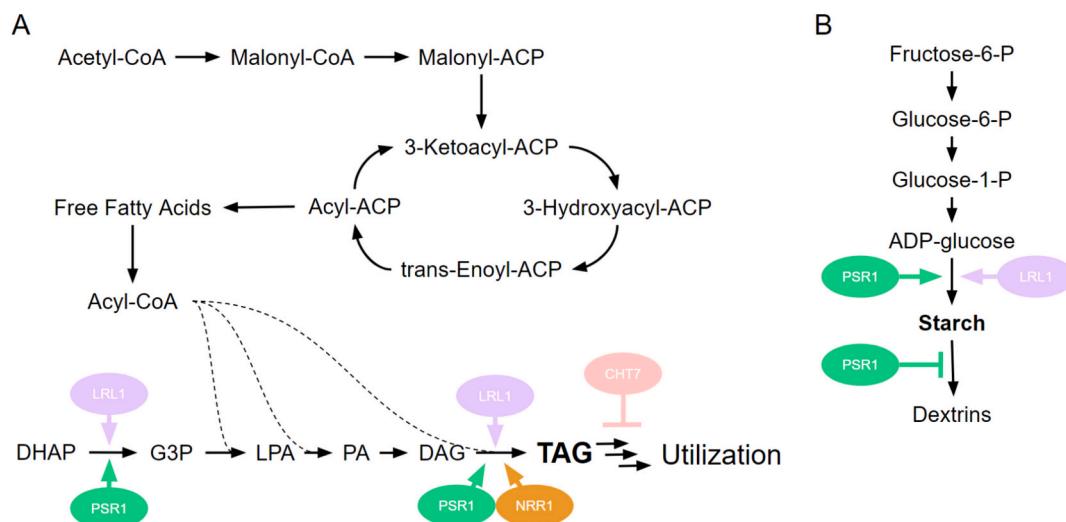


Fig. 8. Proposed model showing how transcription factors shape triglyceride and starch accumulation in *Chlamydomonas pacifica* through increased production and decreased degradation.

The metabolic pathways for (A) TAG and (B) starch biosynthesis and initial degradation in *C. pacifica*. Arrows from TFs represent putative upregulation, and perpendicular tipped lines represent putative downregulation. Abbreviations: ACP, acyl carrier protein; DHAP, dihydroxyacetone phosphate; G3P, glycerol-3-phosphate; LPA, lysophosphatidic acid; PA, phosphatidic acid; DAG, diacylglycerol; TAG, triacylglyceride.

factors and performing comprehensive lipidomic analyses to assess whether they can be leveraged to synthesize specific fatty acid lipid chains. If specificity is found, transcription factor-based tuning of different fatty acid chains could have significant applications for industries other than biofuels, including food, cosmetic, and nutraceutical [7,76–78].

4. Conclusion

This study underscores the potential of utilizing endogenous transcription factors to enhance lipid accumulation in *Chlamydomonas pacifica*, positioning it as a prime candidate for large-scale biodiesel and biomaterial production. By identifying and overexpressing key transcription factors—CpaLRL1, CpaNRR1, CpaCHT7, and CpaPSR1—we achieved substantial increases in lipid biosynthesis compared to the wild-type strain, with differential roles observed under nitrogen-deprived conditions. The significant increase in lipid accumulation in CpaPSR1 overexpressing strains highlights its commercial relevance, and the successful conversion of these lipids into biodiesel demonstrates their practical applicability. This approach not only advances our understanding of lipid biosynthesis in extremophile algae but also offers a robust framework for transcription factor-focused metabolic engineering in other commercially important microalgae species, contributing to sustainable bioenergy solutions and a circular bioeconomy. Future studies could explore the impact of combinatorial overexpression of the tested transcription factor on lipid production and other commercially valuable metabolites.

5. Methods

5.1. Algae strain and genomic data

The wild-type strain of *Chlamydomonas pacifica* has been submitted to the Chlamydomonas Resource Center under the identifiers CC-5697 and CC-5699 [23]. The datasets employed in our research, encompassing genome assembly, protein sequences, coding DNA sequences for all proteins, and predicted transcript data, are available at the following URL: <https://doi.org/10.5281/zenodo.10811182> or were sourced from Molino et al. [29]. For all experiments, we utilized the evolved *C. pacifica* strain developed by Gupta & Molino et al. (CC-6190 402wt x 403wt) and cultivated in the conditions as outlined [30]. Specifically, we used High-Salt Media with acetate (HA) and urea as the nitrogen source at a pH of 10.5. The culture media recipe was adapted from the work of Molino et al. [29]. The cells were grown under continuous 24-h light at approximately 25 °C, with a photon flux of 125 μE/m²/s, on shaken tables rotating at 125 RPM. Biomass for lipidomics and biodiesel analysis was obtained via settling and centrifugation after algal growth in 10 L media carboys. Prior to lipid extraction, biomass samples of *C. pacifica* were flash frozen in a dry ice/isopropanol bath for lyophilization to remove excess aqueous media (LabConco FreeZone 2.5, 0.6 mbar reduced 498 pressure, –52 °C).

5.2. Prediction of transcription factors

Our methodology for predicting transcription factors involved the use of three key bioinformatics tools: PlantTFDB (version 5), TAPscan, and iTAK [33,79,80]. These tools were selected to leverage their specific strengths in TF identification and classification. A notable distinction among these databases is the quantity of Chlorophyte genomes each contains. Specifically, PlantTFDB (version 5) and TAPscan (version 3) include 16 and 13 Chlorophyte genomes, respectively, in contrast to iTAK, which houses just six alga genomes. Our lipids and starch TFs predictions were cross-referenced with a list compiled by Gupta et al., to ensure comprehensiveness and accuracy [21]. For multiple sequence alignment, we utilized EMBL-EBI Clustal Omega, chosen for its efficiency and reliability in aligning protein sequences [81,82]. The

alignments facilitated a comparative analysis of the predicted TFs [83]. Subsequently, phylogenetic trees were generated using the Simple Phylogeny tool, providing insight into the evolutionary relationships of these TFs [81,83]. 10 *C. pacifica* transcription factors' closest sister node was previously experimentally validated starch or lipid, and these TFs were selected for further experimental validation. General functionalities for transcription factors were annotated based on categories from the March 2022 version of the Database of Clusters of Orthologous Genes (COGs) predicted using eggNOG-mapper (version 2) [84,85]. Bar plots for transcription factor classes and COG categories were visualized using R version 4.2.1 and R package ggplot2 (version 3.4.4.) [86].

5.3. Principal component analysis and network plot

Transcription factor distributions for algal species visualized via PCA (Fig. 1C) were sourced from Hu et al. [32] and PlantTFDB (version 5) [33]. The PCA was generated using the sklearn package in Python (version 3.10.9), with additional preprocessing using the "StandardScaler" function to normalize TF abundance between species. A sequence similarity network of TFs from 16 chlorophyte genomes in PlantTFDB [33] (Fig. 1E) was constructed using all-vs-all blastp with a minimum percent identity cutoff of 50% [87]. Transcription factors that did not reach this similarity cutoff to another TF were not pictured. The resulting network was visualized in R version 4.2.3 using the graphopt layout of igraph v. layout [88].

5.4. Conserved domain analysis

Our study's conserved domain analysis entailed comparing predicted transcription factors with established ones from the literature. We utilized the protein sequences of *Arabidopsis* WRINKLED1 (AtWRI1) and sunflower WRINKLED1 (HaWRI1) as presented in Sánchez et al. [48]. Additionally, we referenced conserved domain information for PHOSPHORUS STARVATION RESPONSE 1 (PSR1) TF in *C. reinhardtii* and *A. thaliana* from Rubio et al. [89]. The multiple sequence alignments were visualized using the pymsaviz (version 0.4.2) package in Python, available at <https://pypi.org/project/pyMSAviz/>. This approach enabled the efficient alignment and visualization of similarities and conserved domains between our predicted TFs and those from *Arabidopsis* and sunflower.

In our analysis, we performed a multiple sequence alignment of transcription factors from *C. pacifica* and related species, focusing on the key functional regions (Supplementary Fig. 4a, b). In the case of CrePSR or AtPHR1, the key function region corresponds to the alignment position of 247–300 and 447–489 based on *A. thaliana*'s PHR1. Whereas in the case of AtWRI1 and HaWRI1, it corresponds to alignment positions of 107–177 and 212–271. We concatenated these regions to emphasize conserved domains and calculated percentage identity from the concatenated sequences. We used permutation testing with numerous iterations ($n = 100,000$) to determine statistical significance, randomly shuffling amino acids within the regions of interest across all sequences. This generated a distribution of alignment scores under the null hypothesis, from which we derived a *p*-value, indicating the probability of the observed alignment occurring by chance.

5.5. Identification of lipid and starch metabolism gene orthologs and promoter motif analysis

In our analysis, we focused on identifying and validating the orthologs of genes involved in starch and triacylglyceride metabolism in *Chlamydomonas reinhardtii*, as characterized by Bajhaiya et al. and Shin et al. [43,90]. Utilizing the gene sequences from this study, we conducted local NCBI BLASTP searches to find corresponding orthologs in *C. pacifica* using Biopython (version 1.81) package in Python with a cutoff e-value = 0.001 [91]. The functional annotation of the identified orthologs was conducted using the EggNOG tool, with the e-value and

score as calculated by EggNOG [92]. Further annotation was performed using the KEGG Automatic Annotation Server to confirm that *C. pacifica* contains a complete pathway for starch and sucrose metabolism (pathway: ko000500) and fatty acid biosynthesis (pathway: ko00061) [93,94]. For promoter sequences, the analysis focused on the 1000 base pairs upstream of the transcription start site. During this promoter analysis, we searched for motifs that were similar to the consensus motif of the PSR1 transcription factor, as identified in the PlantTFDB database (TF ID: Cre12.g495100.t1.2) [33]. Transcription factor binding site analyses for PSR1 and CHT7 used PlantTFDB-provided transcription factor binding motifs, scanned against the *C. pacifica* genome using FIMO version 5.5.5 [95].

5.6. Vector design, algae transformation, and Reverse Transcription-Quantitative Polymerase Chain (RT-qPCR) Reaction

The vector assembly utilized genetic components from the genome of *C. pacifica*. The β-Tubulin A 2 gene promoter was employed to regulate the expression of hygromycin (used as a selection marker) and the endogenous transcription factor genes. All the vectors (pJPCHx1_CpaDpWRI1, pJPCHx1_CpaAtWRI1, pJPCHx1_CpaMYB6, pJPCHx1_CpabZIP1, pJPCHx1_CpaSPL12, pJPCHx1_CpaPSR1, pJPCHx1_CpaCHT7, pJPCHx1_CpaNRR1, pJPCHx1_CpaLRL1) have been submitted to the Chlamydomonas Resource Center at the University of Minnesota in St. Paul, Minnesota and the plasmid sequences can be accessed in the following URL: <https://doi.org/10.5281/zenodo.13287080>.

For the algal transformation, plasmid DNA was first digested with *Kpn*I and *Xba*I enzymes (New England Biolabs, Ipswich, MA, USA). Subsequently, the DNA fragments were purified using the Wizard SV Gel and PCR Clean-up System (Promega Corporation, Madison, WI, USA) without fragment separation. The DNA concentration was then determined using the Qubit dsDNA High Sensitivity Kit (Thermo Fisher Scientific, Waltham, MA, USA), and the DNA amount for each transformation was adjusted to ensure 1 µg of payload DNA, based on the payload-to-backbone DNA ratio. The transformation was carried out via electroporation following the protocol by Molino et al. (2018) [96]. Transformed cells were selected on HA agar plates containing 30 µg/mL hygromycin.

To confirm successful transformation, Colony-PCR was conducted using primers from Gupta & Molino et al., targeting a 277 base pair segment of the hygromycin gene (forward primer: TGATTCCCTACGC-GAGCCTGC; reverse primer: AACAGCTTGATCACCGGGGCC) (Supplementary Fig. 5) [30]. The housekeeping gene primers target the ATP2 gene, amplifying an 86 bp segment using the following sequences: forward primer – ACGGTGGTTCTCTGTGTC, and reverse primer – CACACCCGACTCAATCATCTC. The results were then verified through sequencing.

To evaluate the abundance of transcription factor transcripts, total RNA was extracted from the samples using the Direct-zol RNA Miniprep Plus Kit (Zymo Research). Complementary DNA (cDNA) was subsequently synthesized using the AzuraFlex cDNA Synthesis Kit (Azura Genomics) with a combination of random hexamers and oligo(dT) primers, ensuring broad coverage of the RNA transcriptome. Quantitative PCR (qPCR) was performed using the SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) on a Bio-Rad thermocycler. The qPCR protocol included an initial denaturation and polymerase activation step at 98 °C for 3 min, followed by 40 amplification cycles of 15 s at 95 °C for denaturation and 1 min at 60 °C for annealing and extension. A melt curve analysis was conducted from 65 °C to 98 °C, increasing by 0.5 °C every 5 s, to confirm the specificity of the amplified products. ATP2 served as the housekeeping gene for normalization, and the relative expression levels of target genes were calculated using the $2^{-\Delta\Delta CT}$ method [97]. The primers used for amplification were: ATP2 (housekeeping gene)-Forward: 5'-ACGGTGGTTCTCTGTGTC-3', Reverse: 5'-CACACCCGACTCAATCATCTC-3'; CpaLRL1-Forward: 5'-GAGCGGTT-CAATCACCATTG-3', Reverse: 5'-CGGTTGCCGAAACGTAGA-3';

CpaNRR1-Forward: 5'-TGCAACATGGACCACGAA-3', Reverse: 5'-GCATGTGAACGCCTTACAAC-3'; CpaCHT7-Forward: 5'-CTTCCAAG-CAACGACAACAC-3', Reverse: 5'-GGGAATTCCACCTGGAAAGA-3'; CpaPSR1-Forward: 5'-CAGATGGACCTGCAGAAGAAG-3', Reverse: 5'-CATGAGGCTGGCGATGTAG-3'.

5.7. Lipid staining and flow cytometry

We followed the method outlined by Gupta and Molino et al. [30] In brief, a 50 µL sample of the culture was combined with 110 µL of buffer to make a final volume of 160 µL. Next, 1.6 µL of 1 mM Nile red in DMSO was added to this mixture. Three replicates were used for each transgenic line (or wildtype) per nitrogen condition. The cells were incubated at 40 °C for 10 min using a Tecan Infinite M200 Pro plate reader (Tecan Group Ltd., Männedorf, Zurich, Switzerland). After incubation, the samples were analyzed with a Beckman Coulter CytoFLEX Cytometer. Intracellular lipids were detected by measuring Nile Red fluorescence using the PE channel (585/42 bandpass filter) with an excitation of 561 nm (yellow/green) and a gain setting of 1 [97]. For each sample, 10,000 events were recorded at a flow rate of 30 µL/min. The visualized RFU threshold was selected to be one standard deviation above average wild-type nitrogen-replete RFU (approx. 7.7×10^4), surpassed by 3.32 % of control cells. A Wilcoxon ranked-sum test with $p < 0.05$ was used to determine increased lipid production in transgenic lines compared to wild-type controls.

5.8. Confocal microscopy

The sample slides were prepared using Frame-Seal™ Slide Chambers filled with agar enhanced with media. Cells were positioned on the agar and covered with a cover slip to seal them. The confocal microscopy was conducted at the UCSD School of Medicine Microscopy Core using a Leica STED microscope. Chlorophyll was detected with an excitation wavelength of 405 nm and an emission range of 630–680 nm. For Nile red, the excitation wavelength was 470 nm, and the emission range was 475–575 nm. An ANOVA with $p < 0.05$ was used to determine if cell growth or chlorophyll levels were significantly different between cell lines.

5.9. Lipidomics analysis

Lipids analysis were conducted at UCSD Lipidomics Core facility. Lipids were extracted following protocols described by Murawska et al., Hartler et al. and Quehenberger et al. [98–101] 8–10 mg of algae biomass was homogenized into 200µL 10 % methanol. A mix of deuterated lipid standards (Equisplash, Avanti) was added to 150µL biomass homogenate and extracted via BUME. Briefly, samples were extracted using 250 µL of butanol/methanol (BuOH/MeOH), 250 µL of heptane/acylacetate, and 250 µL of 1 % acetic acid. The lipid-containing top layer was collected, solvents were removed, and the lipids were reconstituted in 50 µL of buffer (18:1:1 IPA/DCM/MeOH). Chromatographic separation was performed using a Thermo Vanquish UPLC system with a Cortecs T3 (C18) column (2.1 mm × 150 mm, 1.8 µm particle size), using a binary solvent system. Five microliters (about 7.5 % of total biomass) of each sample were injected. Lipid identification and quantification were conducted using a Thermo Q Exactive mass spectrometer operating in data-dependent MS/MS acquisition mode. LipidSearch and LDA were used for data analysis. The relative abundance of TAG metabolites was analyzed through co-injection with an EQUISPLASH LipidoMIX 15:0–18:1-d7–15:0 TG quantitative mass spec internal standard. The relative abundance is calculated by analyte intensity/internal standard intensity.

5.10. Biodiesel

Algal biodiesel was synthesized from total lipid extracts using acidic

Fischer esterification to produce FAMEs. The masses of dried samples were recorded and two technical replicates per transcription factor and growth condition were used for subsequent lipid extraction and biodiesel synthesis. For each replicate, 0.25 g of dry biomass was resuspended into 1 mL of deionized water and the total lipids were extracted using the Bligh-Dyer extraction protocol [102,103]. Following lipid extraction into chloroform, the solvent was evaporated under air before proceeding with biodiesel synthesis. Each dried lipid sample was treated with an excess of 1 M hydrochloric acid in methanol with stirring at 65 °C for 2 h. Biodiesel was recovered from this mixture via liquid-liquid extraction into hexane. The solvent was removed under ambient evaporation and the final liquid biodiesel was diluted to 2 mg/mL in hexane for GC-MS characterization. GC-MS analysis of biodiesel was performed on an Agilent 7890 A GC system connected to an Agilent 5205977C GC/MSD. Samples were separated on an Agilent HP-5MS UI 30 m × 0.25 mm × 0.25um GCMS column with hydrogen as carrier gas and a gradient of 3 °C /min from 70 °C to 250 °C over 60 min. The mass-to-charge ratio and NIST23 library were used to identify peaks associated with methyl ester biodiesels. Integration values from associated peaks were used to determine the abundance of different biodiesel FAMEs within each sample. The plots were generated using GraphPad Prism.

5.11. Statistics and reproducibility

Comprehensive details of the statistical analysis are provided in the respective section, with significance determined at p -values <0.05. To ensure reproducibility, the experimental design has been thoroughly described in the Methods section, and data has been made available wherever feasible.

CRediT authorship contribution statement

Abhishek Gupta: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Aaron Oliver:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **João Vitor Dutra Molino:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kathryn M.J. Wnuk-Fink:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation. **Marissa Tessman:** Writing – review & editing, Methodology, Data curation. **Kalisa Kang:** Writing – review & editing, Methodology, Data curation. **Évellin do Espírito Santo:** Writing – review & editing, Methodology, Data curation. **Yasin Torres-Tiji:** Writing – review & editing. **Michael D. Burkart:** Writing – review & editing, Funding acquisition. **Stephen Mayfield:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition.

Funding

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the APEX award number DE-EE0009671.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Stephen Mayfield reports financial support was provided by US Department of Energy. Stephen Mayfield reports a relationship with Algenesis Inc. that includes: board membership and equity or stocks. Michael D. Burkart reports a relationship with Algenesis Inc. that includes: board membership and equity or stocks. Marissa Tessman reports a relationship with Algenesis Inc. that includes: equity or stocks.

Acknowledgments

The authors would like to extend their sincere thanks to Prof. Stefan Rensing (University of Freiburg) and Romy Petroll (Max Planck Institute for Biology Tübingen) for their assistance in conducting the TAPscan prediction on our dataset. The authors express their gratitude to Jennifer Santini and Marcy Erb from the UCSD School of Medicine Microscopy Core, supported by the NINDS P30NS047101 grant, for their assistance with confocal microscopy. The authors extend their gratitude to Dr. Oswald Quehenberger, Aaron Armando, and Milda Simonaitis from the UCSD Lipidomics Core for their invaluable assistance with lipidomics analysis. Additionally, the authors are grateful for the support provided by the United States Department of Energy, the California Center for Algae Biotechnology, and the University of California San Diego.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.algal.2025.104192>.

Data availability

We have thoroughly included links to the deposited data throughout the Methods section.

References

- [1] D.I. Stern, The role of energy in economic growth, *Ann. N. Y. Acad. Sci.* 1219 (2011) 26–51.
- [2] F. Johnsson, J. Kjärstad, J. Rootzén, The threat to climate change mitigation posed by the abundance of fossil fuels, *Clim. Pol.* 19 (2019) 258–274.
- [3] A.G. Olabi, M.A. Abdelkareem, Renewable energy and climate change, *Renew. Sust. Energ. Rev.* 158 (2022) 112111.
- [4] I. Capellán-Pérez, M. Mediavilla, C. de Castro, Ó. Carpintero, L.J. Miguel, Fossil fuel depletion and socio-economic scenarios: an integrated approach, *Energy* 77 (2014) 641–666.
- [5] M.J. Barbosa, M. Janssen, C. Südfeld, S. D'Adamo, R.H. Wijffels, Hypes, hopes, and the way forward for microalgal biotechnology, *Trends Biotechnol.* 41 (2023) 452–471.
- [6] S. Smaoui, et al., Microalgae Xanthophylls: from biosynthesis pathway and production techniques to encapsulation development, *Foods* 10 (2021) 2835.
- [7] Y. Torres-Tiji, F.J. Fields, S.P. Mayfield, Microalgae as a future food source, *Biotechnol. Adv.* 41 (2020) 107536.
- [8] M. Benedetti, V. Vecchi, S. Barera, Dall'Osto, L., Biomass from microalgae: the potential of domestication towards sustainable biofactories, *Microb. Cell Factories* 17 (2018) 173.
- [9] J. Hu, W. Meng, Y. Su, C. Qian, W. Fu, Emerging technologies for advancing microalgal photosynthesis and metabolism toward sustainable production, *Front. Mar. Sci.* 10 (2023).
- [10] Z.I. Rony, et al., Unanswered issues on decarbonizing the aviation industry through the development of sustainable aviation fuel from microalgae, *Fuel* 334 (2023) 126553.
- [11] M. Wiatrowski, et al., Techno-economic assessment for the production of algal fuels and value-added products: opportunities for high-protein microalgae conversion, *Biotechnol. Biofuels* 15 (2022) 8.
- [12] A. Difusa, K. Mohanty, V.V. Goud, Advancement and challenges in harvesting techniques for recovery of microalgae biomass, in: P. Thangavel, G. Sridevi (Eds.), Environmental Sustainability: Role of Green Technologies, Springer India, New Delhi, 2015, pp. 159–169, https://doi.org/10.1007/978-81-322-2056-5_9.
- [13] T.-Y. Zhang, et al., Promising solutions to solve the bottlenecks in the large-scale cultivation of microalgae for biomass/bioenergy production, *Renew. Sust. Energ. Rev.* 60 (2016) 1602–1614.
- [14] H. Sun, et al., Powerful tools for productivity improvements in microalgal production, *Renew. Sust. Energ. Rev.* 152 (2021) 111609.
- [15] A.E. Sproles, et al., Recent advancements in the genetic engineering of microalgae, *Algal Res.* 53 (2021) 102158.
- [16] E. Molina-Grima, et al., Pathogens and predators impacting commercial production of microalgae and cyanobacteria, *Biotechnol. Adv.* 55 (2022) 107884.
- [17] J. Liu, A.R. Fernie, J. Yan, The past, present, and future of maize improvement: domestication, genomics, and functional genomic routes toward crop enhancement, *Plant Commun.* 1 (2020) 100010.
- [18] G. Yang, Y. Zhang, X. Wei, L. Cui, X. Nie, Genetic diversity of transcription factor genes in *triticum* and mining for promising haplotypes for beneficial agronomic traits, *Front. Plant Sci.* 13 (2022) 899292.
- [19] I. Ajjawi, et al., Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator, *Nat. Biotechnol.* 35 (2017) 647–652.

- [20] N. Arora, H.-W. Yen, G.P. Philippidis, Harnessing the power of mutagenesis and adaptive laboratory evolution for high lipid production by oleaginous microalgae and yeasts, *Sustainability* 12 (2020) 5125.
- [21] A. Gupta, et al., Harnessing genetic engineering to drive economic bioproduct production in algae, *Front. Bioeng. Biotechnol.* 12 (2024).
- [22] J. Zhao, et al., Overexpression of a MYB1 transcription factor enhances triacylglycerol and starch accumulation and biomass production in the green microalga *Chlamydomonas reinhardtii*, *J. Agric. Food Chem.* 71 (2023) 17833–17841.
- [23] Y. Song, F. Wang, L. Chen, W. Zhang, Engineering fatty acid biosynthesis in microalgae: recent Progress and perspectives, *Mar. Drugs* 22 (2024) 216.
- [24] P. Varshney, P. Mikulic, A. Vonshak, J. Beardall, P.P. Wangikar, Extremophilic micro-algae and their potential contribution in biotechnology, *Bioresour. Technol.* 184 (2015) 363–372.
- [25] E.B. Sydney, et al., Biomolecules from extremophile microalgae: from genetics to bioprocessing of a new candidate for large-scale production, *Process Biochem.* 87 (2019) 37–44.
- [26] D. Rojas-Villalta, et al., Exploring extremotolerant and extremophilic microalgae: new frontiers in sustainable biotechnological applications, *Biology* 13 (2024) 712.
- [27] V. Malavasi, S. Soru, G. Cao, Extremophile microalgae: the potential for biotechnological application, *J. Phycol.* 56 (2020) 559–573.
- [28] M. Singh, N. Trivedi, N. Mal, S. Meharipa, Extremophilic microalgae as a potential source of high-value bioproducts, in: *Extremophiles*, CRC Press, 2023.
- [29] J.V.D. Molino, et al., Description of a novel extremophile green algae, *Chlamydomonas pacifica*, and its potential as a biotechnology host, *Algal Res.* 88 (2025) 1–15.
- [30] A. Gupta, et al., Engineering the novel extremophile alga *Chlamydomonas pacifica* for high lipid and high starch production as a path to developing commercially relevant strains, *ACS EST Eng.* 5 (2025) 36–49.
- [31] S.C. Harrison, A structural taxonomy of DNA-binding domains, *Nature* 353 (1991) 715–719.
- [32] J. Hu, et al., Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae *Nannochloropsis*, *Sci. Rep.* 4 (2014) 5454.
- [33] J. Jin, et al., PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants, *Nucleic Acids Res.* 45 (2017) D1040–D1045.
- [34] M. Shi, L. Yu, J. Shi, J. Liu, A conserved MYB transcription factor is involved in regulating lipid metabolic pathways for oil biosynthesis in green algae, *New Phytol.* 235 (2022) 576–594.
- [35] Y. Yamaoka, et al., The bZIP1 transcription factor regulates lipid remodeling and contributes to ER stress management in *Chlamydomonas reinhardtii*, *Plant Cell* 31 (2019) 1127–1140.
- [36] S. Kwon, et al., Enhancement of biomass and lipid productivity by overexpression of a bZIP transcription factor in *Nannochloropsis salina*, *Biotechnol. Bioeng.* 115 (2018) 331–340.
- [37] P. Zhang, et al., Exploring a blue-light-sensing transcription factor to double the peak productivity of oil in *Nannochloropsis oceanica*, *Nat. Commun.* 13 (2022) 1664.
- [38] R. Miller, et al., Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen deprivation predict diversion of metabolism, *Plant Physiol.* 154 (2010) 1737–1752.
- [39] H. Lv, et al., Transcriptome analysis of *Chlamydomonas reinhardtii* during the process of lipid accumulation, *Genomics* 101 (2013) 229–237.
- [40] C. Wang, et al., Transcription factors from *Haematococcus pluvialis* involved in the regulation of astaxanthin biosynthesis under high light-sodium acetate stress, *Front. Bioeng. Biotechnol.* 9 (2021).
- [41] S. Thiriet-Rupert, et al., Transcription factors in microalgae: genome-wide prediction and comparative analysis, *BMC Genomics* 17 (2016) 282.
- [42] X. Peng, et al., CCHC-type zinc finger family in maize: genome-wide identification, classification and expression profiling under abscisic acid and drought treatments, *PLoS One* 7 (2012) e40120.
- [43] A.K. Bajhiya, A.P. Dean, L.A.H. Zeef, R.E. Webster, J.K. Pittman, PSR1 is a global transcriptional regulator of phosphorus deficiency responses and carbon storage metabolism in *Chlamydomonas reinhardtii*, *Plant Physiol.* 170 (2016) 1216–1234.
- [44] C.Y. Ngan, et al., Lineage-specific chromatin signatures reveal a regulator of lipid metabolism in microalgae, *Nat. Plants* 1 (2015) 1–12.
- [45] B.D. Pant, et al., The transcription factor PHR1 regulates lipid remodeling and triacylglycerol accumulation in *Arabidopsis thaliana* during phosphorus starvation, *J. Exp. Bot.* 66 (2015) 1907–1918.
- [46] S. Baud, et al., WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*, *Plant J. Cell Mol. Biol.* 50 (2007) 825–838.
- [47] N.K. Kang, et al., Increased lipid production by heterologous expression of AtWR11 transcription factor in *Nannochloropsis salina*, *Biotechnol. Biofuels* 10 (2017) 231.
- [48] R. Sánchez, et al., The sunflower WRINKLED1 transcription factor regulates fatty acid biosynthesis genes through an AW box binding sequence with a particular base bias, *Plants* 11 (2022) 972.
- [49] E.M. Fakhry, D.M. El Maghraby, Lipid accumulation in response to nitrogen limitation and variation of temperature in *Nannochloropsis salina*, *Bot. Stud.* 56 (1) (2015) 6.
- [50] E.E. Andeden, S. Ozturk, B. Aslim, Effect of alkaline pH and nitrogen starvation on the triacylglycerol (TAG) content, growth, biochemical composition, and fatty acid profile of *Auxenochlorella protothecoides* KP7, *J. Appl. Phycol.* 33 (2021) 211–225.
- [51] X.-F. Shen, et al., Combining nitrogen starvation with sufficient phosphorus supply for enhanced biodiesel productivity of *Chlorella vulgaris* fed on acetate, *Algal Res.* 17 (2016) 261–267.
- [52] R. Halim, M.K. Danquah, P.A. Webley, Extraction of oil from microalgae for biodiesel production: a review, *Biotechnol. Adv.* 30 (2012) 709–732.
- [53] A. Ibáñez-Salazar, et al., Over-expression of Dof-type transcription factor increases lipid production in *Chlamydomonas reinhardtii*, *J. Biotechnol.* 184 (2014) 27–38.
- [54] B. Jia, et al., Increased lipids in *Chlamydomonas reinhardtii* by multiple regulations of DOF, LACS2, and CIS1, *Int. J. Mol. Sci.* 23 (2022) 10176.
- [55] S. Takahashi, et al., Identification of transcription factors and the regulatory genes involved in triacylglycerol accumulation in the unicellular red alga *Cyanidioschyzon merolae*, *Plants* 10 (2021) 971.
- [56] C. Shang, B. Pang, H. Yu, S. Gan, Y. Li, Identification of targets of transcription factor WRINKLED1-like related to lipid biosynthesis from marine microalga *Dunaliella parva*, *Front. Mar. Sci.* 8 (2022).
- [57] C.-H. Tsai, et al., The protein compromised hydrolysis of triacylglycerols 7 (CHT7) acts as a repressor of cellular quiescence in *Chlamydomonas*, *Proc. Natl. Acad. Sci. USA* 111 (2014) 15833–15838.
- [58] N.A. Hidayati, et al., Lipid remodeling regulator 1 (LRL1) is differently involved in the phosphorus-depletion response from PSR1 in *Chlamydomonas reinhardtii*, *Plant J. Cell Mol. Biol.* 100 (2019) 610–626.
- [59] N.R. Boyle, et al., Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*, *J. Biol. Chem.* 287 (2012) 15811–15825.
- [60] S.-H. Shiu, M.-C. Shih, W.-H. Li, Transcription factor families have much higher expansion rates in plants than in animals, *Plant Physiol.* 139 (2005) 18–26.
- [61] R. Völz, et al., The Trihelix transcription factor GT2-like 1 (GTL1) promotes salicylic acid metabolism, and regulates bacterial-triggered immunity, *PLoS Genet.* 14 (2018) e1007708.
- [62] M.I.R. Khan, M. Fatma, T.S. Per, N.A. Anjum, N.A. Khan, Salicylic acid-induced abiotic stress tolerance and underlying mechanisms in plants, *Front. Plant Sci.* 6 (2015).
- [63] J.P. Davies, D.P. Weeks, A.R. Grossman, Expression of the arylsulfatase gene from the beta 2-tubulin promoter in *Chlamydomonas reinhardtii*, *Nucleic Acids Res.* 20 (1992) 2959–2965.
- [64] Y. Torres-Tijer, et al., Bioinformatic Prediction and High Throughput In Vivo Screening to Identify cis-regulatory Elements for the Development of Algal Synthetic Promoters, 2024.03.20.585996 Preprint at, 2024, <https://doi.org/10.1101/2024.03.20.585996>.
- [65] A. Milito, M. Aschner, J.L. McQuillan, J.-S. Yang, Challenges and advances towards the rational design of microalgal synthetic promoters in *Chlamydomonas reinhardtii*, *J. Exp. Bot.* 74 (2023) 3833–3850.
- [66] M. Trovão, et al., Random mutagenesis as a promising tool for microalgal strain improvement towards industrial production, *Mar. Drugs* 20 (2022) 440.
- [67] T. Takeuchi, et al., *Chlamydomonas CHT7* is required for an effective quiescent state by regulating nutrient-responsive cell cycle gene expression, *Plant Cell* 32 (2020) 1240–1269.
- [68] P.X. Pham, K.T. Nguyen, T.V. Pham, V.H. Nguyen, Biodiesels manufactured from different feedstock: from fuel properties to fuel atomization and evaporation, *ACS Omega* 5 (2020) 20842–20853.
- [69] G. Knothe, Improving biodiesel fuel properties by modifying fatty ester composition, *Energy Environ. Sci.* 2 (2009) 759–766.
- [70] ASTM International, Standard Specification for Biodiesel Fuel Blendstock (B100) for Middle Distillate Fuels (ASTM D6751-24), 2024.
- [71] T. Matthew, et al., The metabolome of *Chlamydomonas reinhardtii* following induction of anaerobic H2 production by sulfur depletion *, *J. Biol. Chem.* 284 (2009) 23415–23425.
- [72] G. Knothe, "Designer" biodiesels: optimizing fatty Ester composition to improve fuel properties, *Energy Fuel* 22 (2008) 1358–1364.
- [73] Y. Liang, K. Mai, S. Sun, Differences in growth, total lipid content and fatty acid composition among 60 clones of *Cylindrotheca fusiformis*, *J. Appl. Phycol.* 17 (2005) 61–65.
- [74] N.V. Zhukova, N.A. Aizdaicher, Fatty acid composition of 15 species of marine microalgae, *Phytochemistry* 39 (1995) 351–356.
- [75] R. Selvarajan, et al., Screening and evaluation of some green algal strains (Chlorophyceae) isolated from freshwater and soda lakes for biofuel production, *Energies* 8 (2015) 7502–7521.
- [76] M. De Luca, et al., Lipids from microalgae for cosmetic applications, *Cosmetics* 8 (2021) 52.
- [77] M. Hawrot-Paw, et al., Fatty acid profile of microalgal oils as a criterion for selection of the best feedstock for biodiesel production, *Energies* 14 (2021) 7334.
- [78] A. Udayan, et al., Production of microalgae with high lipid content and their potential as sources of nutraceuticals, *Phytochem. Rev.* 22 (2023) 833–860.
- [79] R. Petroll, et al., Signatures of transcription factor evolution and the secondary gain of red algae complexity, *Genes* 12 (2021) 1055.
- [80] Y. Zheng, et al., iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases, *Mol. Plant* 9 (2016) 1667–1670.
- [81] W. Li, et al., The EMBL-EBI bioinformatics web and programmatic tools framework, *Nucleic Acids Res.* 43 (2015) W580–W584.
- [82] F. Sievers, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega, *Mol. Syst. Biol.* 7 (2011) 539.

- [83] F. Madeira, et al., Search and sequence analysis tools services from EMBL-EBI in 2022, *Nucleic Acids Res.* 50 (2022) W276–W279.
- [84] M.Y. Galperin, et al., COG database update: focus on microbial diversity, model organisms, and widespread pathogens, *Nucleic Acids Res.* 49 (2021) D274–D281.
- [85] C.P. Cantalapiedra, A. Hernández-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale, *Mol. Biol. Evol.* 38 (2021) 5825–5829.
- [86] H. Wickham, ggplot2. WIREs Comput. Stat. 3, 2011, pp. 180–185.
- [87] H.J. Atkinson, J.H. Morris, T.E. Ferrin, P.C. Babbitt, Using sequence similarity networks for visualization of relationships across diverse protein superfamilies, *PLoS One* 4 (2009) e4345.
- [88] Csárdi, G. & Nepusz, T. The igraph software package for complex network research, *InterJournal, Complex Systems*.
- [89] V. Rubio, et al., A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae, *Genes Dev.* 15 (2001) 2122–2133.
- [90] S.-E. Shin, et al., Increasing lipid production in *Chlamydomonas reinhardtii* through genetic introduction for the overexpression of glyceraldehyde-3-phosphate dehydrogenase, *Front. Bioeng. Biotechnol.* 12 (2024).
- [91] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410.
- [92] J. Huerta-Cepas, et al., eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, *Nucleic Acids Res.* 47 (2019) D309–D314.
- [93] M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, M. Ishiguro-Watanabe, KEGG for taxonomy-based analysis of pathways and genomes, *Nucleic Acids Res.* 51 (2023) D587–D592.
- [94] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome annotation and pathway reconstruction server, *Nucleic Acids Res.* 35 (2007) W182–W185.
- [95] C.E. Grant, T.L. Bailey, W.S. Noble, FIMO: scanning for occurrences of a given motif, *Bioinformatics* 27 (2011) 1017–1018.
- [96] J.V.D. Molino, J.C.M. de Carvalho, S.P. Mayfield, Comparison of secretory signal peptides for heterologous protein expression in microalgae: expanding the secretion portfolio for *Chlamydomonas reinhardtii*, *PLoS One* 13 (2018) e0192433.
- [97] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method, *Methods* 25 (2001) 402–408.
- [98] J. Hartler, et al., Automated annotation of sphingolipids including accurate identification of hydroxylation sites using MS_n data, *Anal. Chem.* 92 (2020) 14054–14062.
- [99] J. Hartler, et al., Deciphering lipid structures based on platform-independent decision rules, *Nat. Methods* 14 (2017) 1171–1174.
- [100] O. Quehenberger, et al., Lipidomics reveals a remarkable diversity of lipids in human plasma, *J. Lipid Res.* 51 (2010) 3299–3305.
- [101] G.M. Murawska, A.M. Armando, E.A. Dennis, Lipidomics of phospholipase A2 reveals exquisite specificity in macrophages, *J. Lipid Res.* 65 (2024) 100571.
- [102] E.G. Bligh, W.J. Dyer, A rapid method of total lipid extraction and purification, *Can. J. Biochem. Physiol.* 37 (1959) 911–917.
- [103] P. Manirakiza, A. Covaci, P. Schepens, Comparative study on total lipid determination using Soxhlet, Roes-Gottlieb, Bligh & Dyer, and Modified Bligh & Dyer extraction methods, *J. Food Compos. Anal.* 14 (2001) 93–100.