Algal Research 22 (2017) 187–215

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Review of the algal biology program within the National Alliance for Advanced Biofuels and Bioproducts

Clifford J. Unkefer [a], Richard T. Sayre [a,b], Jon K. Magnuson [c], Daniel B. Anderson [c], Ivan Baxter [d], Ian K. Blaby [e], Judith K. Brown [f], Michael Carleton [g,1], Rose Ann Cattolico [h], Taraka Dale [a], Timothy P. Devarenne [i], C. Meghan Downes [j], Susan K. Dutcher [k], David T. Fox [a], Ursula Goodenough [k], Jan Jaworski [l], Jonathan E. Holladay [c], David M. Kramer [m], Andrew T. Koppisch [a,2], Mary S. Lipton [c], Babetta L. Marrone [a], Margaret McCormick [g], István Molnár [f], John B. Mott [a], Kimberly L. Ogden [f], Ellen A. Panisko [c], Matteo Pellegrini [e], Juergen Polle [n], James W. Richardson [i], Martin Sabarsky [o], Shawn R. Starkenburg [a], Gary D. Stormo [k], Munehiro Teshima [a], Scott N. Twary [a], Pat J. Unkefer [a], Joshua S. Yuan [i], José A. Olivares [a,*]

[a] Los Alamos National Laboratory, P.O. Box 1663, MS-M888, Los Alamos, NM 80745, USA
[b] New Mexico Consortium, 100 Entrada Drive, Los Alamos, NM 87544, USA
[c] Pacific Northwest National Laboratory, P.O. Box 999, MSIN P8-60, Richland, WA 99352, USA
[d] USDA at Donald Danforth Plant Science Center, 975 North Warson Road St. Louis, MO 63132, USA
[e] Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA
[f] University of Arizona, Tucson, AZ 85721, USA
[g] Matrix Genetics, LLC, 1600 Fairview Ave E Ste 300. Seattle, WA 98102, USA
[h] University of Washington, Seattle, WA 98195, USA
[i] Texas A&M University, 400 Bizzell St, College Station, TX 77840, USA
[j] New Mexico State University, 1780 E University Ave, Las Cruces, NM 88003
[k] Washington University in St. Louis, 1 Brookings Dr, St. Louis, MO 63130, USA
[l] Donald Danforth Plant Science Center, 975 North Warson Road St. Louis, MO 63132, USA
[m] Michigan State University, 220 Trowbridge Rd, East Lansing, MI 48824, USA
[n] Brooklyn College of the University of New York, 2900 Bedford Ave, Brooklyn, NY 11210, USA
[o] Cellana Inc., Kailua-Kona, HI 96740, USA

ARTICLE INFO

Article history:
Received 21 November 2014
Received in revised form 22 April 2016
Accepted 1 June 2016
Available online 21 June 2016

Keywords:
Algal biology
Genomics
Proteomics
Lipid biosynthesis
NAABB
National Alliance for advanced biofuels and Bioproducts

ABSTRACT

In 2010, when the National Alliance for Advanced Biofuels and Bioproducts (NAABB) consortium began, little was known about the molecular basis of algal biomass or oil production. Very few algal genome sequences were available and efforts to identify the best-producing wild species through bioprospecting approaches had largely stalled after the U.S. Department of Energy's Aquatic Species Program. This lack of knowledge included how reduced carbon was partitioned into storage products like triglycerides or starch and the role played by metabolite remodeling in the accumulation of energy-dense storage products. Furthermore, genetic transformation and metabolic engineering approaches to improve algal biomass and oil yields were in their infancy. Genome sequencing and transcriptional profiling were becoming less expensive, however; and the tools to annotate gene expression profiles under various growth and engineered conditions were just starting to be developed for algae. It was in this context that an integrated algal biology program was introduced in the NAABB to address the greatest constraints limiting algal biomass yield. This review describes the NAABB algal biology program, including hypotheses, research objectives, and strategies to move algal biology research into the twenty-first century and to realize the greatest potential of algae biomass systems to produce biofuels.

1. Introduction

1.1. Preface

Unlike competing renewable energy production systems (wind, solar, or hydroelectric), biomass has the advantage that it can be

* Corresponding author.
E-mail address: Olivares@lanl.gov (J.A. Olivares).
[1] Currently at Presage Biosciences, 530 Fairview Ave N, Seattle, WA 98109, USA
[2] Currently at Department of Chemistry, Northern Arizona University, S San Francisco St, Flagstaff, AZ 86001, USA

http://dx.doi.org/10.1016/j.algal.2016.06.002

converted into an energy-dense, liquid fuel feedstock (biocrude) that is compatible with current petroleum refinery technologies. In addition, biological systems provide their own chemical energy storage systems, reducing potential inefficiencies from the conversion and/or storage of electrical energy generated by wind or photovoltaic systems [1]. Recently, there has been substantial interest in the use of microalgae for sustainable production of biofuels, which can be attributed to the ability of algae to produce two- to ten-fold greater annual aerial biomass than that produced by terrestrial crops [2–6]. In addition, algae have the potential to capture inorganic carbon injected into ponds as CO_2 and hydrate it to nongaseous bicarbonate, substantially reducing the potential residence time of CO_2 emitted from point sources in the atmosphere. With respect to fuel conversion technologies, algae are particularly attractive since many are facultatively capable of accumulating between 4% and 60% lipids by dry weight, making them among the most efficient biofuel feedstock production systems [7]. Various estimates indicate that potential oil and biomass yields from algal ponds range from 20,000 to 60,500 L/ha/year and 50,000 to 15,000 kgdw/ha/year, respectively [5]. Emerging developments in genetically modified algae promise to increase production yields by as much as two- to three-fold [8–10].

Before the start of the National Alliance for Advanced Biofuels and Bioproducts (NAABB) consortium, there had been efforts to characterize growth and lipid-generation potential of a variety of algal species. It was known that under nitrogen limitation, many algae stored reduced carbon in the form of triglycerides, terpenoids, or polysaccharides that could be used for high-energy density fuel production. However, little was known about the fundamental physiology involved in partitioning carbon or the role played by metabolite remodeling in the accumulation of energy-dense storage products. At the outset of the NAABB, very few algal genome sequences were available and efforts to identify the best-producing wild species through bioprospecting approaches had largely stalled after the U.S. Department of Energy's Aquatic Species Program. Furthermore, algal genetic transformation required for metabolic engineering approaches was only possible in a few model species. At that same time commercial next-generation DNA sequencers and the computational tools to assemble to annotate complex genomes and expression profiles were becoming widely available, making rapid algal genome sequencing and transcriptional profiling possible. It was in this context that an integrated algal biology program was introduced in NAABB to address the greatest constraints limiting algal biomass yield. The NAABB final report to the Department of Energy provides details on the consortium and all of the technical and scientific efforts undertaken [11]. In the following sections, we describe the hypotheses, research objectives, and strategies of this program to move algal biology research into the twenty-first century and to realize the greatest potential of algae biomass systems to produce biofuels.

1.2. Approach

At the outset of NAABB, our techno-economic analysis showed that increasing the productivity of algae would have the greatest impact on the economic viability and sustainability of an algal biofuels industry; this observation remains true today. The overall objective of the algal biology teams was to increase the overall productivity of algal biomass accumulation and lipid/hydrocarbon content. The specific targets were to increase by more than 50% the total lipid dry weight and to improve pond cultivation by more than 20 gdw/m^2/day. Our approach to characterizing the biological potential of algal biofuels is outlined in Fig. 1 and the Algal Biology Work Breakdown Structure (WBS) (Table 1). Two potentially complementary paths were explored to achieve the program's objectives, including bioprospecting algal diversity, and/or genetic modification of algae to increase both yield and robustness while reducing energy and material inputs. Clearly, given the extensive biological diversity of algae (>100,000 species), tremendous advantages can be achieved by identifying the most robust and productive strains of algae that could be further improved through genetic engineering. Thus, our bioprospecting efforts focused on surveying the natural diversity of wild algal strains for production-relevant traits. Independently, substantial efforts were also focused on targeted transformations and directed engineering of high-performance algae, which required a greater understanding of the molecular basis of the observed traits. Our systems biology and bioinformatics efforts included genome sequencing along with transcriptomic, proteomic, and phenomic analyses. At each step in the process, it was often necessary to develop new technologies or tools to address the unique biology of algae. To develop and

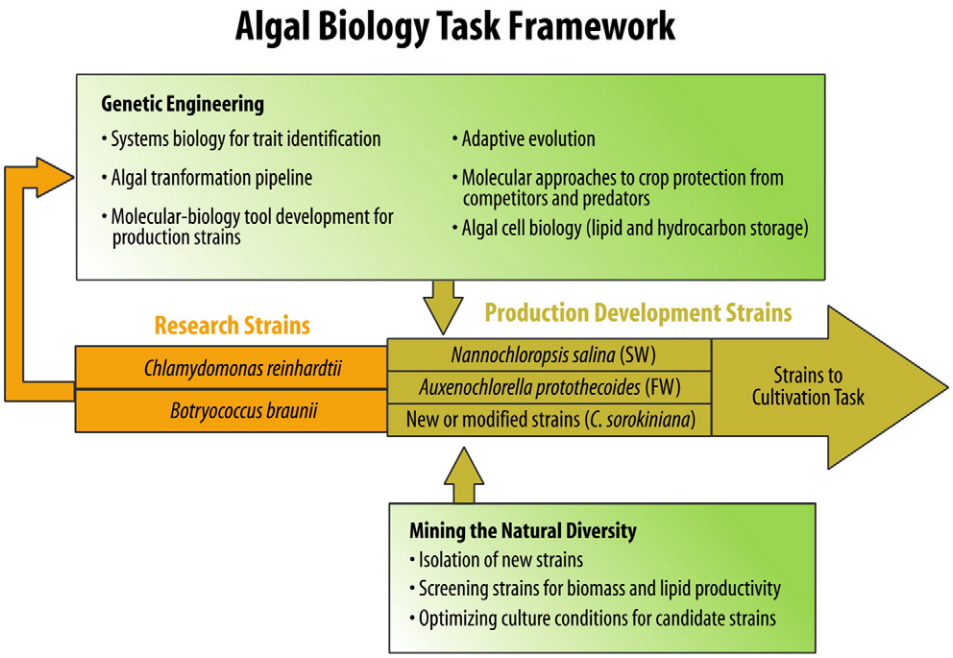

Fig. 1. Algal biology task framework.

Table 1
Algal biology WBS with deliverables.

1.1	Screen biodiversity to maximize yield and lipid production by
1.1.1	Rapid screening for lipid production and growth
1.1.2	Archiving promising strains
	M/D1.1: 1500 algal isolates screened, ≥30 best strains verified and deposited to UTEX 36
	M/D 1.1 Best strains demonstrated in production
1.2	Systems biology - to identify bottlenecks in lipid biosynthesis
1.2.1	Transcriptomics with NexGen sequencing methods
1.2.2	Proteomics - comprehensive proteome analysis
1.2.3	Metabolomics- isotope labeling for metabolic flux analysis
1.2.4	Deep annotation for genome and transcriptome analysis
	M/D1.2: Genes for increased yield, productivity, and nutrient utilization cataloged.
	D1.2 Transgenic tools demonstrated for *C. reinhardtii*, *Chlorella* and *N. salina*.
1.3	Crop protection techniques to create competitive, robust strains
1.3.1	Rational engineering and directed evolution
	M/D1.3 Genes for crop protection cataloged
1.4	Maximize algal growth, lipid production and light harvesting using:
1.4.1	Rational metabolic engineering
1.4.2	Directed evolution, including screening
1.4.3	Optimized growth conditions and micronutrient control
1.4.4	Engineered growth regulators to optimize production
	M/D1.4: Transgenic strains incorporating best trait(s) demonstrated in culture.
	M/D1.4: Best evolved strain of *Nanochloropsis* & *Chlorella* demonstrated in culture.

M/D – Milestone/deliverable.

test these tools, we relied the readily transformable model alga *Chlamydomonas reinhardtii*.

2. Technical accomplishments

2.1. Mining algae's natural diversity

The most intensive NAABB bioprospecting effort focused on a relatively broad temporal and geographic sampling approach. The general flow of the approach is shown in Fig. 2. Approximately 400 samples were collected across the continental United States. From these samples, over 2200 independent strains of algae were isolated, and over 1500 of those were subjected to a preliminary screen for oil

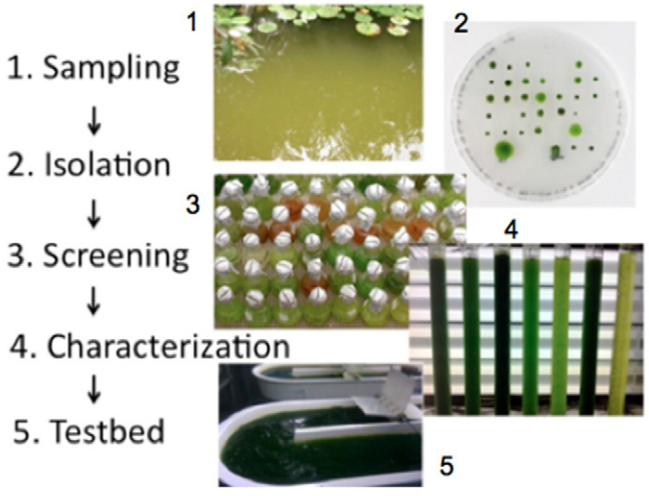

Fig. 2. Flow diagram of the broad temporal, climatic, and geographic survey approach to isolation and characterization of algal biofuel candidate strains. Panel 1, an example sampling site; Panel 2, isolates from fluorescence activated cell sorting (FACS) on an agar plate; Panel 3, first-tier screening in traditional flask cultures; Panel 4, 300 mL bubble columns for characterizing the most promising candidates; Panel 5, 200 L NAABB testbeds.

accumulation. Samples were collected from the summer of 2010, through the spring of 2012, from various habitats, including nonaqueous (e.g., soil) substrates, freshwater, brackish water, marine, and hyper-saline environments. Freshwater habitats included man-made and natural waterways, while saline habitats included estuaries, inland salt lakes, shrimp ponds, algal raceways, and the coastal waters of Texas, California, New York, and Connecticut. The tactic of sampling across seasons from the same geographical sites was intended to isolate strains suitable for winter and summer cultivation.

Strains were isolated by traditional culture methods [12] using a variety of growth media for initial plating and by high-throughput fluorescence-activated cell sorting (FACS). Since any isolation approach introduces biases, obtaining algal isolates using multiple independent procedures increased the likelihood of recovering a broader diversity of isolates. Previously, it was observed that many strains of microalgae did not survive the cell-sorting process; relying on this approach alone would therefore limit the diversity of the strains to be isolated. On the other hand, traditional cell culture techniques using petri dishes and multi-well plates are much slower and would not have provided the large numbers of strains necessary to meet NAABB goals and milestones. Combining these approaches provided wide diversity and large numbers of unialgal (not axenic) isolates.

Once we isolated the strains to apparent homogeneity, we carried out high-throughput screening using shaker flasks in sets of 120 to identify strains that grew well autotrophically and that accumulated triacylglycerides (TAGs). Growth was monitored by optical density at 750 nm and the lipophilic dye Nile Red was used for fluorescence measurements of relative TAG content [13]. We note that because of the low permeability of Nile Red into some algal species with robust cell walls, this screen could have overlooked some potential high-lipid producers. On the other hand, we identified more lipid-producing strains than we had resources to characterize.

Further characterization of the best strains from the high-throughput screen was performed in shake flasks with a varied media differing by 10-fold in the concentrations of N, P, and Fe to identify the best growth media for the highest oil yield. The best performing strains were then grown in bubble columns using growth media that allowed for the greatest oil yields (Panel 3, of Fig. 2). Strains were grown in the columns bubbled with 1% CO_2 under constant light (200 μE/m^2/s), and at constant temperature (28 °C). Biomass productivity was measured by determining the dry weight at a set harvest time and comparing the result with the biomass productivity of *N. salina CCMP1776*, which is a marine alga that over-produces lipids under nitrogen stress. This strain has been cultured in ponds year around in many locations in the world, including at three NAABB cultivation sites, and was adopted by the NAABB and the DOE as a benchmark strain.

Hundreds of algal strains were assembled into a cataloged culture collection. Unialgal strains that approximated or exceeded the biomass productivity of *N. salina* were deposited at the University of Texas (UTEX) and considered for large-scale cultivation. The 30 best-producing strains from this task are listed in Table 2 and were forwarded to UTEX to ensure that they would be available to the algal research community at large, thus meeting a NAABB final deliverable. This combination of new strain isolation and initial testing in the laboratory combined with later testing in outdoor settings using multiple cultivation systems proved successful. Not all strains that were selected after laboratory screening and testing performed well in the testbed cultivation systems, but several algal strains could be successfully grown in the testbed facilities. Several of these strains were tested on the 4–8 L scale in a greenhouse before final selection for cultivation in the test beds. Listed in Table 3, several strains were examined extensively by other consortium members and cultured in one or more of the NAABB testbed facilities. For example, *Chlorella sorokiniana* (DOE1412) was cultured by the NAABB cultivation team in both traditional and the ARID raceways. It performed well outdoors, producing up to 30 g/m^2/d biomass with a lipid content of 25%. Cultivation results are discussed in detail by

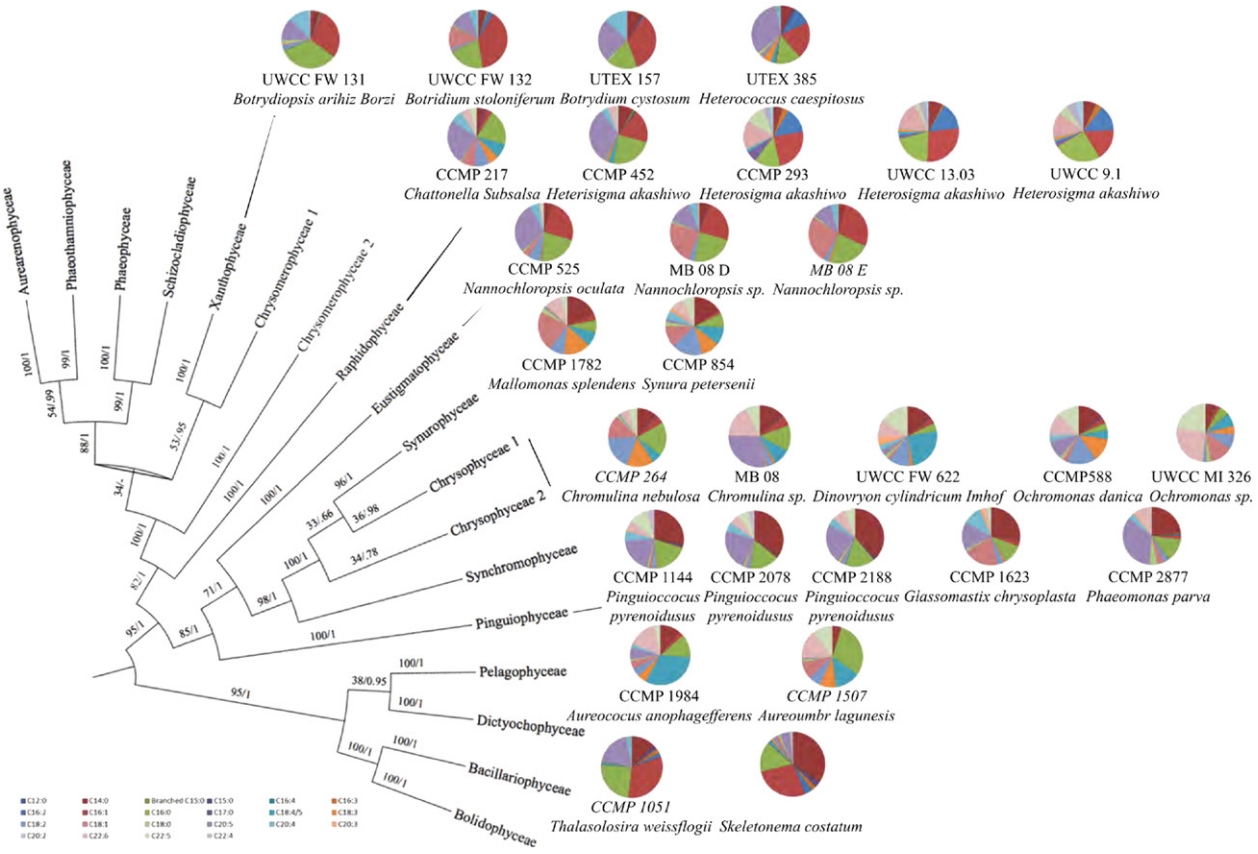

Fig. 3. Lipid profiles determined using the micro-GC–MS technique of representative taxa appended to a modified stramenopile phylogeny based on a five-gene phylogeny [15,18].

Lammers and coworkers [14]. This positive result contrasted the previously published Aquatic Species Program report from 1998 [7], which indicated that it would be impossible to match a successful combination of indoor isolation/screening/testing with outdoor strain performance.

2.1.1. Chromophyta bioprospecting

A complementary approach to the previously described high-throughput strain isolation protocol was the in-depth characterization of fatty acids in a sampling of 27 oleaginous phototrophs from the under-studied but phylogenetically broad group of stramenopiles that includes *N. salina*. A simple, relatively high-throughput analytical method for the analysis of cellular fatty acid profiles was developed for this screening as well. To facilitate lipid analysis from small algal cultures, a sub-microscale in situ (SMIS) method for gas chromatography-mass spectrometry (GC-MS_ analysis of fatty acids was developed [15,16]. This method can be performed with less than 250 μg of dry algal biomass. It employs boron trifluoride catalyzed transesterification of lipids in methanol, followed by a two-step phase separation to recover the fatty acid methyl esters (FAMEs). GC-MS analysis is then performed relative to internal standards by single-ion quantitation.

The utility of this method was demonstrated by generating lipid profiles of 27 marginally studied algal representatives within the stramenopile clade Table 2. Each of these algal species were cultured in their optimal medium at 20 °C under fluorescent lights (100 μE/m^2/s) with a 12 h day/night cycle and harvested in the early stationary growth phase by centrifugation. The modified five-gene (*rn*A, *rbc*L, *psb*A, *psa*A, and *psb*C) phylogeny of Yang and coworkers [17] was used to generate Fig. 3. Most revealing is the fact that no distinct fatty acid pattern is immediately obvious at almost any taxonomic level. For example, *Chattonella* and *Heterosigma*, members of the small class Raphidophyceae, exhibit markedly different lipid profiles. Similarly, five members of the Pinguiophyceae class have different lipid profiles. Although the cultures were grown identically and harvested at the stationary phase, the lack of fatty acid pattern replication is also seen at the strain level. When phyla outside the stramenopile cluster (Fig. 3) are examined, their fatty acid profiles (with rare exception) have little taxonomic affinity. Comparative analysis does indicate that algal members of the chromaphyta (Stramenopiles, Primnesiophuceae, Dinophyceae, and Cryptophyceae) contain a much broader and often higher polyunsaturated fatty acid complement than do green algae (Chlorophyta). As genetic modification becomes more routine, those organisms with unique products or with high productivity of a desirable product could potentially provide targets for molecular harvesting or engineering.

2.2. Genome sequencing and systems biology[3]

For the long-term sustainability of an algal biofuels industry, it is essential to revolutionize algal fuel production by engineering the best available algal strains to increase their productivity, control and optimize their lipid/hydrocarbon biosynthesis pathways, and make them deployable by engineering “crop protection” mechanisms. At the time the project was initiated, there had been significant efforts to characterize growth phenotypes and the lipid-generation potential of a variety of algal species. Furthermore, it was known that under certain growth conditions including stress, nutrient limitation, or heterotrophic boost, many algae accumulate large concentrations of lipids, terpenoids, or carbohydrates that could be used for high-energy density fuel production. However, very little was known about the fundamental physiology, genomics, metabolism, and regulation of the metabolic pathways in the previously characterized phenotypes. Only a few genomes of model organisms, most notably *Chlamydomonas reinhardtii*, had been

[3] Genomic and transcriptomic data are publically available at https://greenhouse.lanl.gov. Questions should be directed to Dr. Shawn R. Starkenburg, shawns@lanl.gov.

Table 2
Biomass productivity measured in a bubbling column.

Strain classification	Initial strain ID	[a]NAABB strain ID	[b]Biomass productivity (g/L/d)	[c]Lipid content (%)
[d]*N. salina* CCMP1776	N/A	N/A	0.63	31
Scenedesmus obliquus	DOE0013	NAABB1013	0.67	
Coelastrella sp.	DOE0088	NAABB1088	0.65	10[f]
Chlamydomonadales	DOE0101	NAABB1101	0.73	20
[e]*Scenedesmus obliquus*	DOE0111	NAABB1111	0.81	28
[e]*Scenedesmus obliquus*	DOE0152	NAABB1152	0.89	27
Coelastrella sp.	DOE0155	NAABB1155	0.7	
Coelastrella sp.	DOE0202	NAABB1202	0.86	18[f]
Chlorophyceae	DOE0222	NAABB1222	0.70	30
Ankistrodesmus sp.	DOE0259	NAABB1259	0.56	38
Trebouxiophyceae	DOE0314	NAABB1314	0.61	10
Coelastrella sp.	DOE0369	NAABB1369	1.0	
Chlorophyta	DOE0488	NAABB1488	0.81	38
Chlorellaceae	DOE0623	NAABB1623	0.69	24
Chlorellaceae	DOE0686	NAABB1686	0.78	36
Chlorophyta	DOE0700	NAABB1700	0.74	
Chlorellaceae	DOE0717	NAABB1717	0.78	31
Eustigmophyta	DOE0899	NAABB1899	0.55	34
Eustigmophyta	DOE0907	NAABB1907	0.56	32
Chlorellaceae	DOE1041	NAABB2041	0.75	
Chlorellaceae	DOE1044	NAABB2044	0.81	27
Desmodesmus sp.	DOE1051	NAABB2051	0.65	
Chlorellaceae	DOE1070	NAABB2070	0.7	
Chlorellaceae	DOE1095	NAABB2095	1.00	32
Chlorellaceae	DOE1116	NAABB2116	0.94	25
Chlorellaceae	DOE1135	NAABB2135	0.95	
Desmodesmus sp.	DOE1357	NAABB2357	0.81	23
[e]*Chlorella sorokiniana*	DOE1412	NAABB2412	0.65	25
Desmodesmus sp.	DOE1418	NAABB1418	0.62	25
Chlorophyceae	DOE1426	NAABB2426	0.84	40
Chlorophyta	DOE1727	NAABB2727	0.72	14

[a] To date these strains have not been assigned a UTEX strain ID number. They can be obtained using the DOE strain ID number. Reported data are based on a single determination. Questions about these strains should be directed to Dr. Juergen Polle, jpolle@brooklyn.cuny.edu.
[b] Biomass productivity was measured as the ash-free dry weight produced per liter of culture medium per day and is expressed as g/L/d. Cells were grown at 28 °C under constant light (200 μE/m^2/s) provided by daylight fluorescent lamps using 1%CO_2.
[c] After extraction, total lipid content was determined gravimetrically.
[d] For comparison, the biomass productivity and lipid content of *Nannochloropsis salina* CCMP1776 were determined.
[e] Species were confirmed by subsequent rRNA sequencing.
[f] Lipid extraction not successful due to tough cell wall.

sequenced and annotated. We set a goal of experimentally improving the performance of the best available lipid-producing algal strains including *Nannochloropsis salina* and *Picochlorum* sp., for brackish/salt water production and *Auxenochlorella protothecoides* for fresh water production. In addition, a team of investigators focused on understanding the physiology associated with terpenoid overproduction in *Botryococcus braunii*. Due to the high degree of evolutionary diversity in algae, it was not clear that it would be possible to extrapolate genetic, physiologic, and biochemical properties observed in *Chlamydomonas* to produce commercializable strains of algae. To gain a quick understanding of the physiology of lipid over-production, it would be necessary to obtain high-quality genome sequences for these organisms. In addition, high-quality genome sequences of production strains would facilitate the transfer of genetic tools (vectors, promoters, terminators, codon optimizations, etc.) developed for *C. reinhardtii* to production strains.

Because genome sequences for production strains were not available at the outset of the NAABB, we initially carried out an extensive systems biology studies correlating gene expression changes with triglyceride production in the starchless mutant of *C. reinhardtii*. It had been demonstrated that the starchless mutant of *C. reinhardtii* rapidly produced large quantities of triglycerides when shifted to nitrogen-free culture medium that contained acetate as a carbon source. We recognized that because it is a natural starch accumulator that required acetate to produce high quantities of lipids, the starchless mutant of C. reinhardtii had limitations as a model for lipid accumulation in naturally oleaginous species like *Nannochloropsis*. However, model studies with *C. reinhardtii* provided a baseline for comparing expression profiles obtained later in the production strains *N. salina*, *Picochlorum*, and *A. protothecoides*. Although there were significant differences, we did observe some similarities in the lipid induction expression profiles of *C. reinhardtii* and *N. salina*.

2.2.1. Genome sequencing

For these reasons, NAABB took advantage of new genome sequencing technologies including the Illumina [18,19], 454, and Pacific Biosciences [20] platforms, which were complemented by the development of novel computational tools [21–29], to sequence, assemble, and annotate high-quality algal genomes and transcriptomes quickly. A major accomplishment of the NAABB consortium was the sequencing and assembly of eight high-quality algal genomes from three independent phyla, the greatest biodiversity of algae sequences at that time (Table 4). To develop gene models and correlate algal gene expression with phylogenetic profiling, 220 RNAseq transcriptome samples were sequenced. Many of the gene expression studies were completed under nitrogen deprivation or other stress or growth conditions to monitor changes in gene expression during lipid induction.

Discussed in detail below, we provided extensive transcriptome sequences to analyze genes involved in lipid production in *C. reinhardtii.* The RNAseq data were also used to generate gene models and functional annotations for *N. salina*, *Picochlorum* sp., *A. protothecoides* and *Chrysochromulina* sp. This information enabled construction of metabolic pathways. Simultaneously, the NAABB team used genomic and RNAseq transcriptome data produced by the U.S. Department of Energy Joint Genome Institute to analyze the production of hydrocarbons by *B. braunii*. Although annotation and comparative genomics analyses of the data continue, the sequence data developed by NAABB have thus far allowed us to conduct a pangenomic comparison of the chloroplast and mitochondrial genomes of *Nannochloropsis and Chrysochromulina* with other algal taxa [30,31]. Many of the organellar gene modifications observed in these algae are novel and deviate from conserved orthologs found across the tree of life; moreover, many such modifications have been aquired via lateral gene transfer. The observed differences affect a broad range of metabolic genes/systems, including those contributing to the regulation of branched-chain amino synthesis (acetohydroxyacid synthase), carbon fixation (RuBiscCO activase), energy conservation

Table 3
Strains on which NAABB partners obtained significant data.

Classification	Strain ID	Culture Testing[a]	Sequencing[b]
Chlorophycea	DOE0101	L, CS, G, P	$G^{(LANL)}$
S. obliquus	DOE0152	L, CS, G, P	$G^{(UCLA)}$ & $T^{(UCLA)}$
C. sorokiniana.	DOE1412	L, CS, G, P	$G^{(LANL)}$ & $T^{(LANL)}$

[a] Culture Testing: L–Laboratory testing in bubbling columns; CS–Climate-simulated laboratory culturing; G–intermediate scale growth in the green house; P–Growth in algal raceways at the Texas ArgiLife Research, Pecos, Texas site.
[b] Sequencing G–Genome sequencing; T–Transcriptome sequencing.

Table 4
NAABB genome projects.

[a]Genome	Assembly quality	Size, Mbp
Picochlorum sp.	Improved high quality draft	15.2
Auxenochlorella protothecoides UTEX25	Improved high quality draft	21.4
Chrysochromulina sp.	High quality draft	59
Nannochloropsis salina 1776	Improved high quality draft	29.4
Tetraselmis sp. LANL1001	Standard draft	220
Chlorococcum sp. DOE0101	Standard draft	120
Chlorella sorokiniana DOE1412	Standard draft	56
Chlorella sorokiniana Phycal1228	Standard draft	56

[a] Genomic and transcriptomic data is publically available at https://greenhouse.lanl.gov. Questions should be directed to Dr. Shawn R. Starkenburg, shawns@lanl.gov.

(ATP synthase), protein synthesis and homeostasis (Clp protease), and regulatory activities (ribosomal protein zinc finger domains, two-component signal transduction arrays, and nitrate transcriptional regulation). Assessment of internal chloroplast repeat structures provided new insights into phylogenetically determined gene regulatory mechanisms.

These observations and further discovery of currently unidentified genetic and structural modifications to critical cellular components and pathways will be required to help explain (and exploit) the unique physiological properties found in candidate production strains. It is also worth noting that a high degree of divergence in the amino acid sequences of many algal proteins has led to false annotations when based strictly on gene homology. Thus, further optimization of gene annotation tools and curation of draft annotations are needed. Finally, analysis of the *N. salina* and *N. gaditana* organellar genomes suggested that these two isolates should be reclassified as different strains of the same species; the availability and consequent comparative analyses of the nuclear genomes of both isolates are required to provide verification to support this reclassification.

2.2.2. Systems biology

To gain further understanding of the molecular control of oil accumulation, we used a combination of RNAseq transcriptomics, proteomics, and metabolomics studies to characterize changes in gene expression associated with nitrogen deprivation and lipid production in *Chlamydomonas*, *Picochlorum*, and *Nannochloropsis*. As discussed above, *C. reinhardtii* was chosen as a model organism because its annotated genome was available and there were significant data on the induction of lipid accumulation in the starchless mutant of *C. reinhardtii*. *Picochlorum* is a marine microalgae of industrial interest due to its high lipid accumulation and its ability to grow under non-ideal conditions. The *Picochlorum* genome was the first produced by the NAABB and is relatively small at 15.2 Mbp. *N. salina* is also of industrial interest due to its high lipid accumulation and it is also the NAABB and DOE/EERE baseline organism. Because the genomes of the strains isolated during the NAABB project were not available until the end of the three-year project, it was necessary to develop additional systematic procedures for analysis. Toward this goal, we constructed an in-house pipeline for algal genomics/transcriptomics [32], as depicted in Fig. 4.

2.2.3. Transcriptome analysis of lipid production in C. reinhardtii

C. reinhardtii is capable of accumulating intracellular lipids that can be converted to biofuel under conditions in which the cells experience physiological stress such as nutrient deprivation [33,34] or salt stress [35]. Furthermore, starchless mutants, such as *sta6*, hyperaccumulate lipids when stressed [36]. The generation of "obese" cells containing higher TAG content per dry weight in *sta6* suggests either remodeling of existing lipids and/or channeling of fixed carbon into lipid production [37–39]. In the context of NAABB's objectives, we sought to gain better understanding of the global response to N deprivation in the starchless mutant, *sta6*, using RNA-Seq transcriptomics [40]. Our intention was to compare gene expression in wild-type *C. reinhardtii* and in the *sta6* mutant to deduce the specific metabolic pathways that enable the *sta6* mutant to produce more lipids, potentially identifying genes that may regulate oil accumulation and paving the way for future metabolic engineering projects.

2.2.3.1. Comparative phenotype of wild-type (cw15) C. reinhardtii and its mutant (sta6) under nitrogen deprivation. The onset of nitrogen deprivation induces starch and lipid accumulation, in addition to provoking gametogenesis, chlorophyll degradation, down-regulation of photosystems, and ribosomal proteins. To investigate in more depth the physiological changes in wild-type (available as CC-4349 and CC-4348 from the Chlamydmonas Resource Center) and *sta6* following nitrogen deprivation, we assayed the two strains and the *STA6*-complemented strains, *STA6* — C2, *STA6* — C4 and *STA6* — C6, (CC-4565, CC-4566 and CC-4567) [37,41]. The complemented strains were obtained by transformation of *sta6* with a plasmid carrying a genomic copy of the wild-type *STA6* gene and a paromomycin resistance cassette [36]. All five strains were assayed for cell counts and for chlorophyll, starch, and lipid contents. Upon nitrogen limitation, the cell counts of all strains remained unchanged, indicating that cell division had been arrested. The chlorophyll content per cell decreased in all strains, as has previously been noted. Near zero starch was accumulated in the *sta6* mutant strain, but rapidly accumulated in the parent and complemented strains. Although no significant difference between the TAG content in CC-4349 (a wild-type presumed parental strain of *sta6*, but see below) and *sta6* was observed in the first 48 h of nitrogen deprivation, the TAG content of *sta6* greatly exceeded that of CC-4349 and was N-deprivation dependent at 96 h.[4]

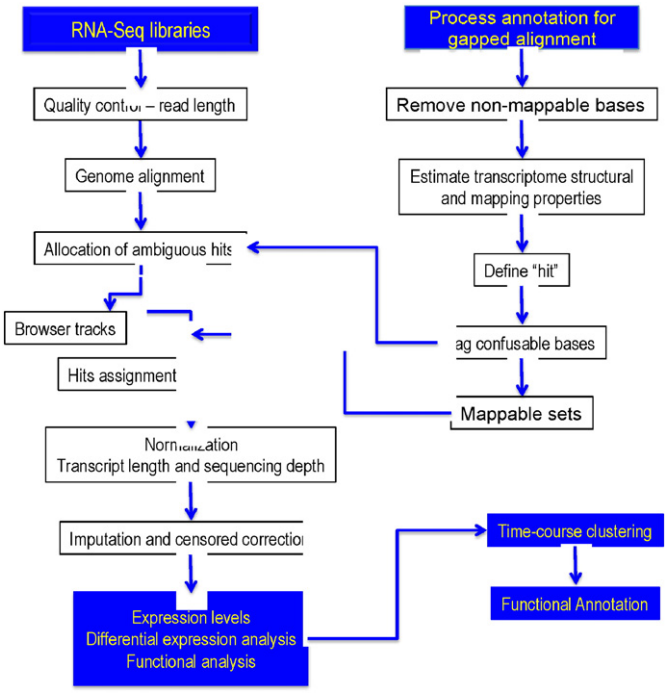

Fig. 4. Schematic of the algal genome/transcriptome analysis pipeline for *C. reinhartii* [33]. Given the large number of regulated genes in nitrogen-deprived cells and the size of the datasets, we classified the genes according to their temporal expression profiles across individual or combined experiments (time-course clustering). We also performed different model-based clustering analyses by combining responses at early and late time points. Finally, we used our pathways annotation tool to couple genes to function.

Throughout the course of this work, we noted several unexpected phenotypes in the *sta6* mutant relative to CC-4349, including a mating-type discrepancy and a small cell size that was not rescued by restoring *STA6*. To investigate further, we sequenced the genomes of CC-4349, *sta6* (CC-4348), and *STA6*-C6. In addition to precisely defining the *sta6* locus, we identified a large number of single nucleotide variants (SNVs) between CC-4349 and sta6, indicating that this was not the parental strain. We therefore concentrated our subsequent analysis on comparing *sta6* with the three *STA6*-complemented strains.

2.2.3.2. TAG biosynthesis. Given the ability of *sta6* to amass a greater quantity of lipids per cell than CC-4349 could, we expected that results from the transcriptome analyses might demonstrate increased transcript abundance of fatty acid and TAG biosynthesis-related genes in *sta6* versus CC-4349. However for most genes associated with lipid biosynthesis, there was no significant difference in reads per kilobase per

[4] Statistics: at 96 h, *N*-deprived CC-4349 contained 24 ± 3.2 fmol TAG per cell; sta6 contained 52 ± 4.8 fmol TAG per cell. These values represent the mean of three biological repeats.

million (RPKM) between the two strains. Chlamydomonas stores lipids as TAGs, which are synthesized by covalent attachment of fatty acids to glycerol by acyltransferases. We isolated three genes related to TAG accumulation, MLDP1, DGTT2 and DGAT3. The exact function of MLDP1 is not known, but a knock-down mutant in *Arabidopsis* exhibited increased lipid body size, suggesting that it stabilizes lipid droplets [42]. Acyltransferase activity has been postulated to be the rate-limiting step in TAG synthesis. If the concentration of acyl-CoA is assumed to be non-limiting, then the increased expression of DGTT2 and/or DGAT3-*like* suggesting that they are key mediators of TAG accumulation in *sta6* (Fig. 5). Unlike many of the genes induced in *sta6* versus *cw15*, there is high-level expression of DGTT2 and DGAT3-*like* following nutrient stress (Fig. 5). Algal production strains engineered to overexpress these genes early may have the potential to produce TAGs more rapidly following nutrient stress.

Our transcript analysis also showed that six key genes encoding enzymes involved in pathways associated with central carbon metabolism were up-regulated in *sta6* compared to CC-4349 and the three strains complemented for *STA6* (Fig. 6). These genes included *ACS3*, which encodes acetyl-CoA synthase, which catalyzes the first step in acetate metabolism, *MAS1* and *ICL1*, which encode malate synthase and isocitrate lyase, respectively, in the glyoxylate fatty acid B-oxidation pathway, and *FBP1* and *PCK1*, which encode fructose-1,6-bisphosphatase and phosphenolpyruvate carboxykinase involved in primary carbon metabolism. Finally, *TAL1*, *which* encodes transaldolase and is important for the balance of metabolites in the pentose phosphate pathway, was also overexpressed during N-deprivation in *sta6* compared with in the *STA6* strains. All six genes displayed highly uniform elevated expression profiles, suggesting metabolic linkage and implying increased metabolic flux through these pathways in the *sta6* mutant (Fig. 6). These expression patterns are consistent with increased turnover of fatty acids and enhanced carbon scavenging during N-deprivation, processes that may be associated with lipid remodeling.

Although comparison of the transcriptome data from all five strains showed that the *STA6* deletion was the cause of many of the differentially expressed genes, we sought to independently verify this by determining whether we could observed correlated changes in the at the protein and metabolite levels. We determine the effect of nitrogen deprivation on the concentrations of key metabolites ADP-Glc, citrate, aconitate isocitrate, succinate, malate, FDP and Glucose-6-Phosphate in sta6 and the complimented strains. These data are available [41] and are represented on the metabolic chart in Fig. 7. Malate synthase and isocitrate lyase enzyme activities were assessed in cell extracts from N-deprived cultures. Our assays showed that *sta6* had a two- to three-fold increase in malate synthase and isocitrate lyase activities compared to the *STA6*-complemented strains after 48 h of N deprivation. Similarly, intracellular concentrations of key metabolites were determined at 48 h with and without nitrogen. As expected, we found that in *sta6* there was a reduction in ADP-glucose levels relative to those in *STA6* strains. In addition, and in agreement with our transcriptome analysis, *sta6* exhibited increases in isocitrate, succinate, malate, and fructose-1,6-bisphoshate levels in N-deprived cultures compared with those in *STA6* strains. These data confirm that the increased mRNA abundance translated into increased enzyme activities and subsequently into increased flux through the glyoxylate cycle and potentially into carbon scavenging via the Calvin-Benson cycle. Significantly, by short-circuiting the Krebs cycle via increased glyoxylate cycle activity, potentially less carbon is lost during respiration.

To realize the potential of algal-derived lipids as transportation fuel, a complete understanding of the molecular circuitry of algae is essential. We have taken strides toward this objective by applying transcriptomics and metabolomics approaches to examining nitrogen-deprivation–induced lipid production in the model organism *C. reinhardtii*. We identified a large number of significantly differentially expressed genes between *sta6* and *STA6*, including fatty acid acyltransferases involved in the biosynthesis of TAGs. Genes whose expression are substantially altered during oil accumulation are potential targets for genetic regulation to improve lipid production in production strains.

2.2.4. Transcriptome analysis of lipid production in A. protothecoides

To determine the patterns of gene expression associated with oil accumulation induced by various processes, we carried out RNAseq analyses on *A. protothecoides* at various time intervals after a variety of treatments. These treatments included those that induce oil accumulation (glucose or decane addition), increased temperature (37 °C) and decreased temperature (4 °C), and controls for daylight and at midnight. The transcriptomes were collectively analyzed using the Sequedex program for rapid gene annotation based on kmers of 10 amino acids [43, 44]. Fig. 8 shows that the transcript abundance of genes encoding enzymes functional in fatty acid synthesis increased with glucose addition, which increased TAG levels from 5–45% (dry weight). In contrast, short exposure (1 h) to decane treatments, which are associated with lipid remodeling and similar levels of TAG accumulation, was associated with increased expression levels of genes associated with stress tolerance and at longer time periods (3 h) with the expression of genes involved in protein metabolism and repair. These results demonstrate that de novo TAG synthesis induced by glucose differs substantially from stress-induced (decane) accumulation of TAGs in terms of the patterns

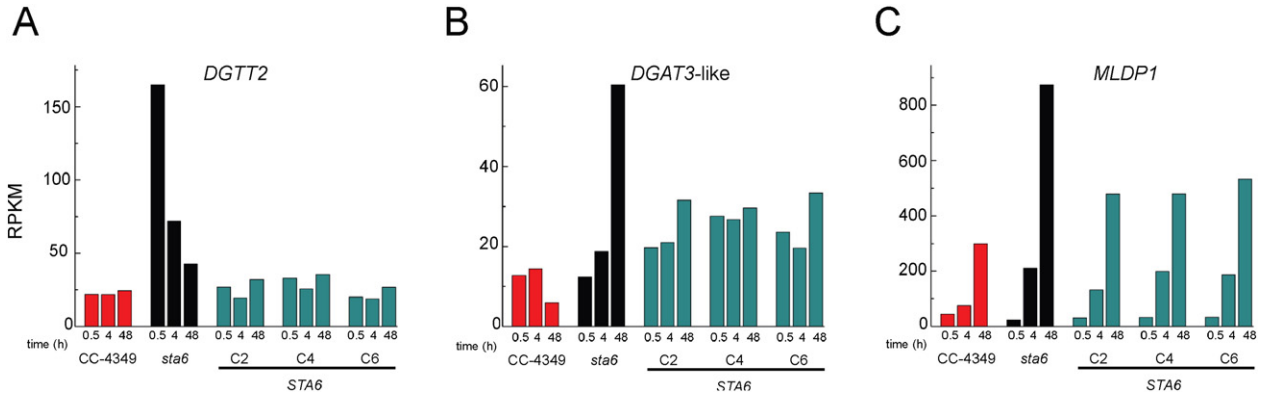

Fig. 5. Expression profiles of genes related to TAG accumulation. (A) DGTT2, encoding a type 2 diacylglycerolacyltransferase, (B) DGAT3-like, encoding a putative soluble diacylglycerolacyltransferase, and (C) MLDP1 encoding major lipid droplet protein. Expression level is expressed as reads per kilobase of exons in each gene model per million aligned reads (RPKM) for each mRNA at 0.5, 4 and 48 h. CC-4349 is shown in red, sta6 in black and STA6 in blue.

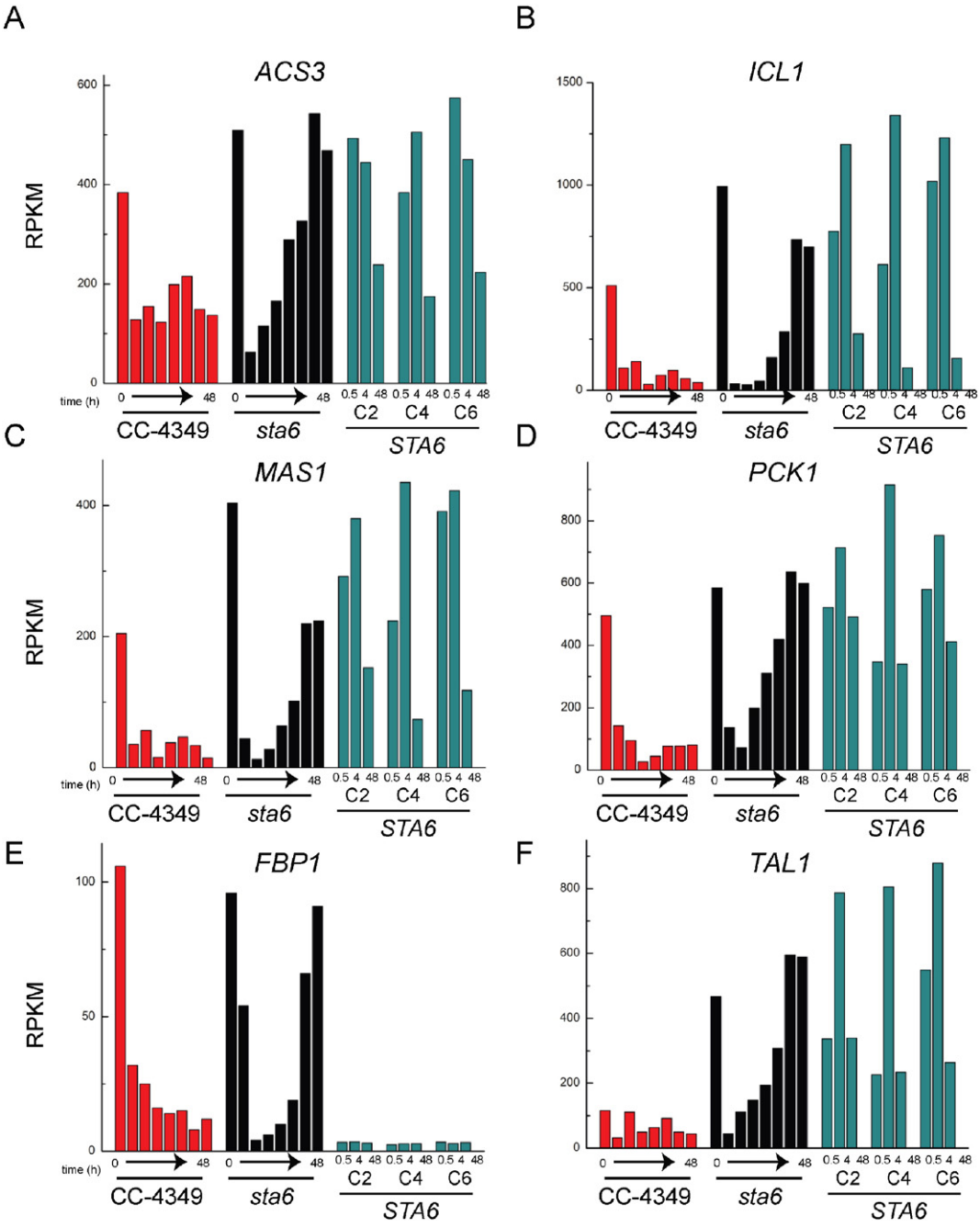

Fig. 6. Greater up-regulation of genes encoding enzymes of acetate metabolism, gluconeogenesis and the oxidative pentose phosphate pathway in sta6 vs. STA6. Expression level is expressed as reads per kilobase of exons in each gene model per million of aligned reads (RPKM). (A) acetyl CoA synthase (ACS3), (B) isocitrate lyase (ICL1), (C) malate synthase (MAS1), (D) phosphoenolpyruvate carboxykinase (PCK1), (E) fructose-1,6-bisphosphatase (FBP1), and (F) transaldolase (TAL1); in CC-4349 (red); sta6 (black) at 0, 0.5 2, 4, 8, 12, 24 and 48 h; and STA6-C2, —C4 and —C6 (blue) at 0.5, 4 and 48 h after transfer to N-free medium.

of gene expression. These analyses were made possible due to the sequencing and annotation of the *A. protothecoides* genome by NAABB.

2.2.5. Transcriptome analysis of lipid production in Picochlorum sp. and N. salina

To study lipid production in *Picochlorum* and *N. salina*, we cultured these algae strains in five-liter cylindrical photobioreactors in which the temperature, pH, dissolved [O_2], and rate of addition of CO_2, O_2, and N_2 were controlled. The algae were grown under saturating light intensities (400 μE/m^2/s) for 16 h/day at pH 8.2. During growth, the pH was used to control the addition of CO_2, which ensured that both the pH and the concentration of the carbon substrate ([CO_2] + [HCO_3^-] + [CO_3^{-2}]) were constant throughout the experiment. Cell numbers, (A_{750}), total lipid as FAMEs, and nitrate in the medium were monitored. Samples were also analyzed by flow cytometry to measure cell number, cell size, and autofluorescence. The fluorescent dye BODIPY was used to measure total neutral lipids.

Representative data from photobioreactor growth of *Picochlorum* and *N. salina* are shown in Fig. 9. In the experiments shown, the initial concentration of sodium nitrate was 4 mM for *Picochlorum* and 8 mM for *N. salina*. As the algae grew, nitrate in the growth medium became depleted. Each of these marine algae increased their lipid accumulation with the depletion of nitrogen from the growth medium. Under nitrogen depletion, the rate of lipid accumulation increased three-fold in

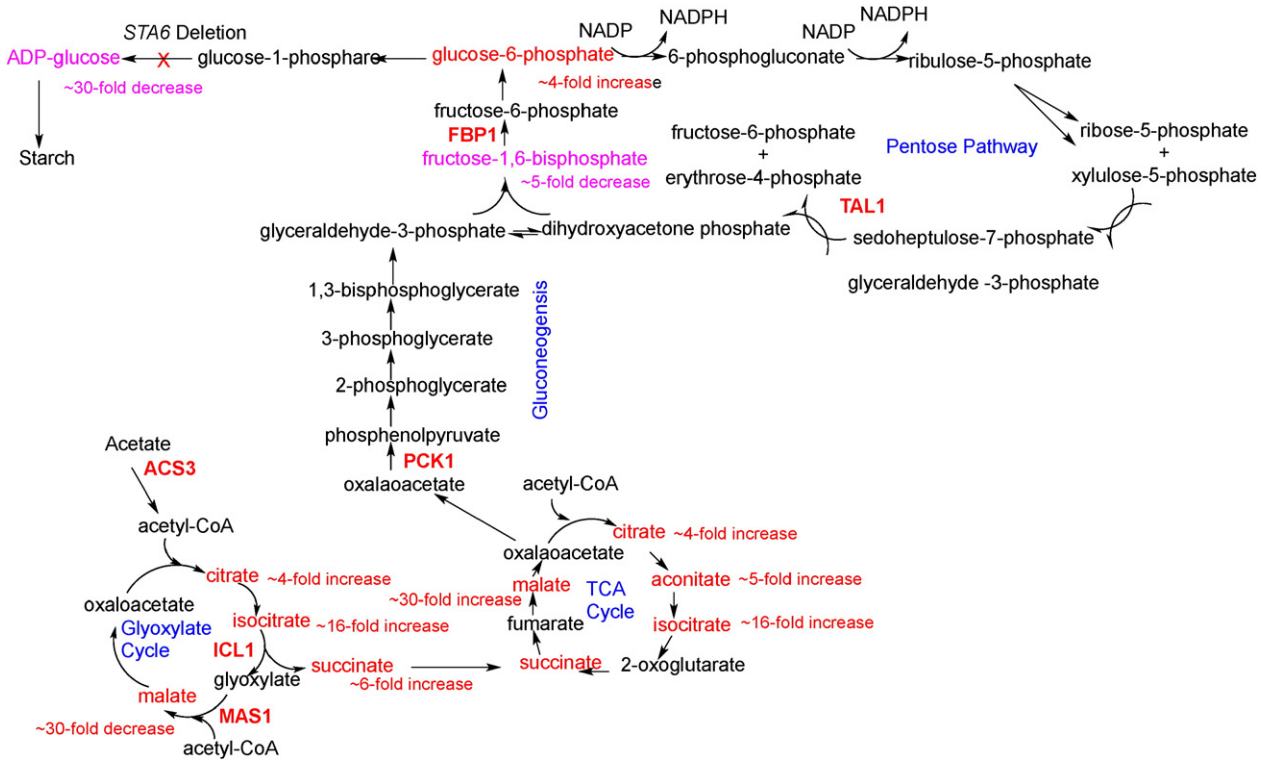

Fig. 7. Altered metabolism in sta6 vs. STA6. The genes ACS3, ICL1, MAS1, PCK1, FBP1, and TAL1 (labeled in red) are upregulated during N-deprivation. Enzyme activity measurements confirmed a 2–3 fold increase in the expression of ICL1 and MAS1. In agreement with our transcriptome cell extracts show increased activity for ICL1 and MAS1. The concentration of the metabolites shown in red increases in sta6 compared to the STA6-C2, —C4 and —C6 cells during N-deprivation. The concentration of fructose-1,5-biphosphate is decreased during N-deprivation.

the *Picochlorum* cultures (Fig. 9A). More significantly, when nitrate became depleted from the growth medium of *N. salina* cultures, the lipid accumulation rate increased 6.3-fold from 24 mg/L/d to 151 mg/L/d (Fig. 9B). The flow cytometry results indicated that the cells stopped dividing some time after the nitrate became depleted from the growth medium. However, the biomass continued to increase as the algae accumulated lipids, causing the cells to increase in volume, which resulted in an increased optical density at 750 nm. Under these conditions, *N. salina* cultures accumulated lipids to approximately 50–60% of their biomass.

2.2.6. Transcriptome analysis of Picochlorum sp.

To examine the changes in gene expression associated with the depletion of nitrate (nitrogen) and the subsequent increased lipid accumulation rate, samples before, during, and after nitrogen depletion were subjected to RNAseq transcriptome analysis. RNA was isolated from samples obtained at four time points: (1) 4 days before nitrate was depleted from the growth medium, (2) the day the nitrate concentration fell to zero, (3) 1 day after nitrogen depletion, and (4) 4 days after nitrogen depletion.

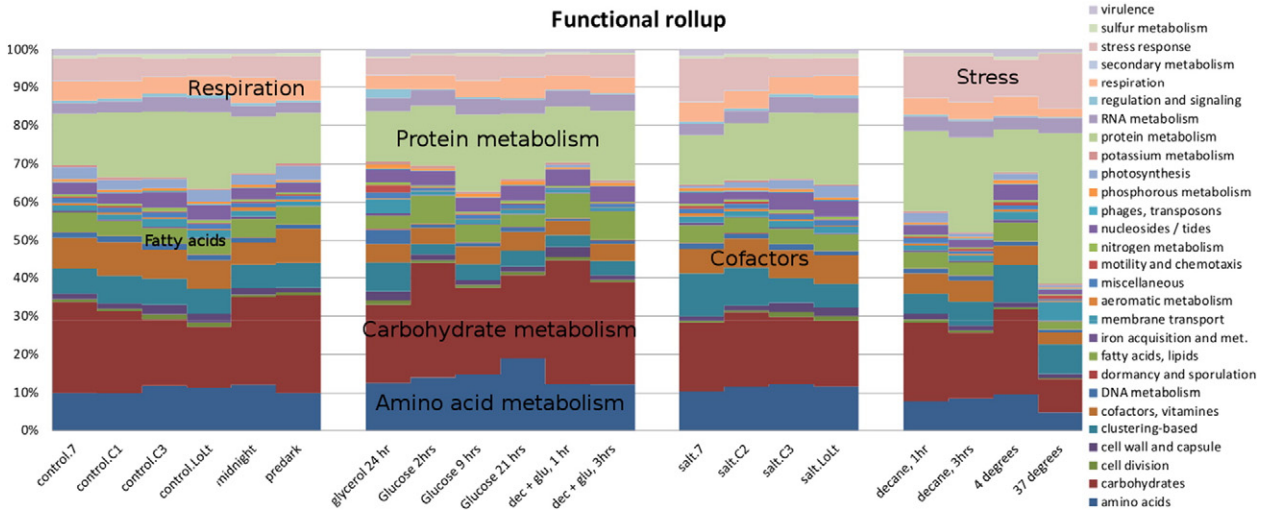

Fig. 8. Relative transcriptional abundane of various functional gene assignments as influenced by various stresses as well as carbohydrate-induced oil accumulation.

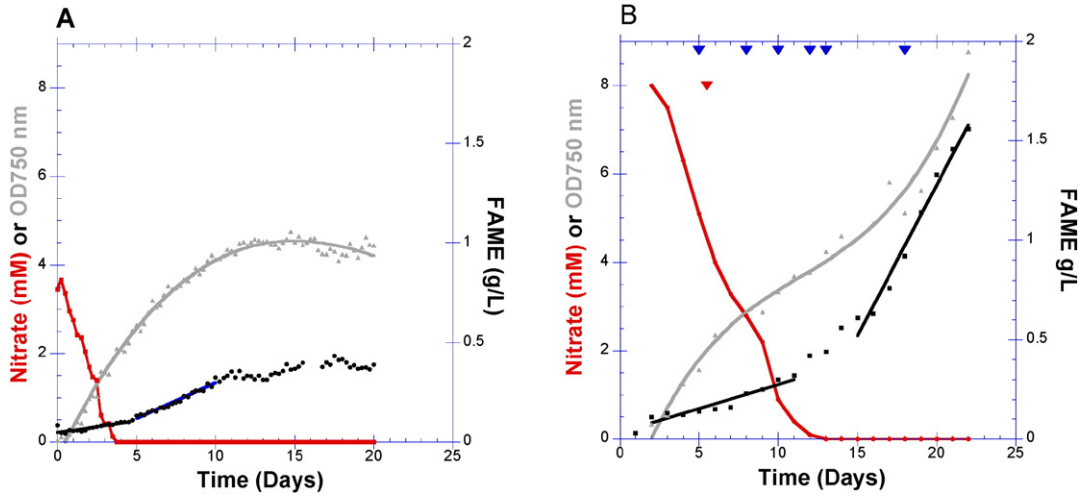

Fig. 9. (A) Growth of *Picochlorum* sp. in our photobioreactor. When nitrate was depleted from the growth medium, the lipid accumulation rate increased from 12 mg/L/d to 36 mg/L/d. (B) Growth of *N. salina* in our photobioreactor. When nitrate was depleted from the growth medium, the lipid accumulation rate increased from 24 mg/L/d to 151 mg/L/d. The cells stopped dividing sometime after the nitrate was depleted from the growth medium, but the OD_{750} continued to increase during rapid lipid accumulation because the cells increased in volume. Transcriptomes were analyzed on samples collected on days marked ▼ and nights marked ▼.

2.2.6.1. Lipid and TAG biosynthesis. We annotated genes involved in fatty acid biosynthesis including the genes encoding; acetylCoA carboxylase 1, AccABCD; acetylCoA carboxylase 3, Acc; malonyl CoA ACP transferase, FabD; 3-ketoacylACP reductase, FabG; 3-ketoacylACP synthase 1, FabB; 3-ketoacylACP synthase 2, FabH; enoylACP reductase, FabI, K, or L; acylACP desaturase; and acylACP thioesterase. The expression of each of these fatty acid biosynthesis genes increased at least two fold as a consequence of nitrogen depletion. We also observed increased expression of genes encoding for enzymes involved in TAG synthesis. Similar to the observation reported above for *Chlamydomonas*, we found that nitrogen depletion induced the expression of acyltransferases including DGAT in *Picochlorum*.

These data in Fig. 10 are presented as a metabolic pathway in which enzymes encoded by genes that were induced upon nitrogen depletion are shown in red and those whose expression was repressed are shown in blue. Genes encoding for enzymes shown in black were not identified in our annotation. Further analysis and experimentation are necessary to identify which of these enzymes are expressed in the chloroplast,

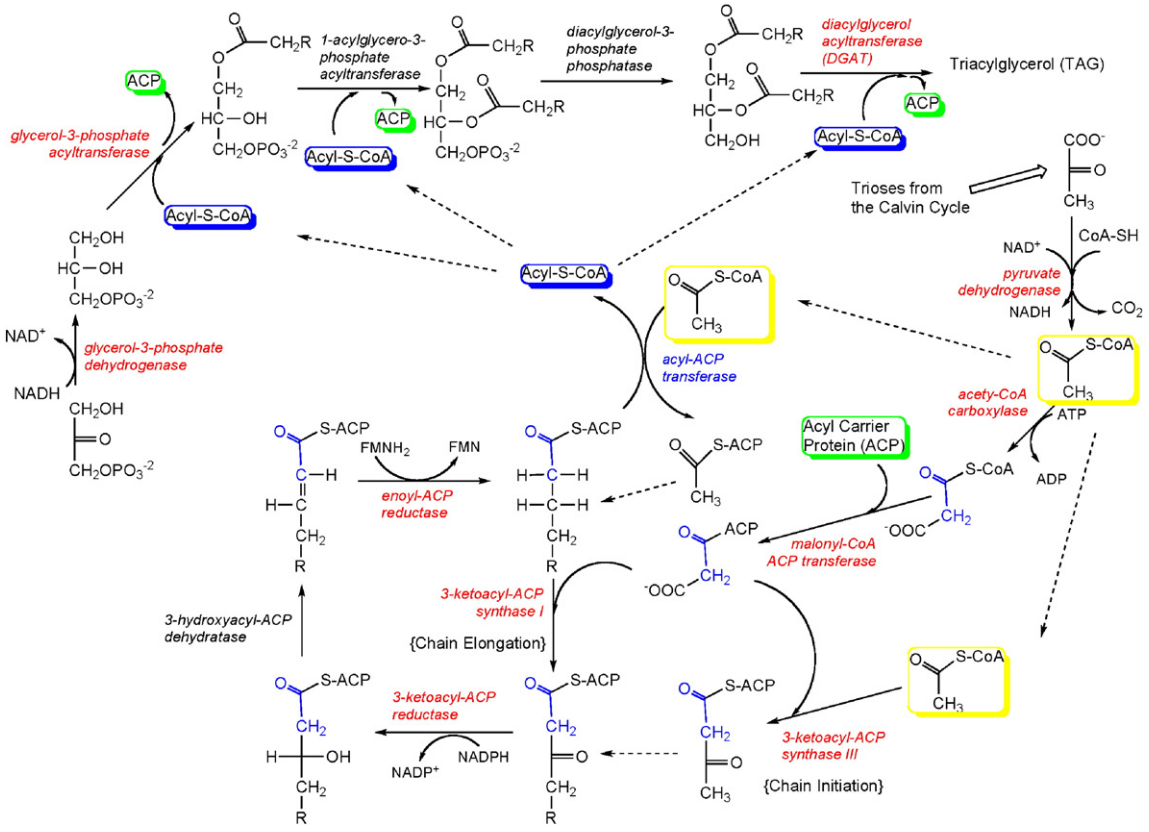

Fig. 10. Metabolic map of lipid and TAG biosynthesis. Enzymes encoded by genes that are induced during the high lipid production phase of growth are lettered in red, repressed genes in blue. No genes have been annotated to encode for 2-hydroxyacyl-ACP hydratase, 1-acylglycerol-3P acyltransferase or diacylglycero-3P phosphatase.

the site of fatty acid and triglyceride biosynthesis. However, the overall conclusion is that the increased rate of lipid production results from the induction of the entire lipid biosynthesis pathway. Acetyl-CoA carboxylase (ACCase) catalyzes the conversion of acetyl CoA to malonyl-CoA, the first and possibly rate-limiting step in lipid biosynthesis. It was observed that one of the ACCase gene family members was induced to a greater degree than other gene family members during lipid accumulation.

2.2.6.2. Starch biosynthesis. Nitrogen depletion causes induction of most of the enzymes required for starch biosynthesis. Genes encoding both subunits of the ADP-glucose pyrophosphorylase were induced two- to 16-fold more after nitrogen depletion. One of the isozymes of starch synthase was also induced four-fold more at every time point after nitrogen was depleted. Several enzymes required for production of hexoses from the triose-phosphates produced by the Calvin-Benson Cycle were also induced by nitrogen depletion. This transcriptome profile suggests that upon nitrogen depletion from the growth medium, *Picochlorum* stores energy and carbon transiently as starch, which may provide the substrates for TAG accumulation.

2.2.7. Transcriptome analysis of N. salina

RNA isolated from samples of *N. salina* cultures taken on Days 5, 8, 10, 12, 13, and 18 (Fig. 9B) were subjected to RNAseq analysis using deep sequencing with a HiSeq sequencer. Proteome analysis was carried out on parallel samples from the *N. salina* culture. RNA was isolated both from the light and mid-dark growth periods (Fig. 9B).

2.2.7.1. Expression of tricarboxylic acid (TCA) cycle enzymes during day versus night. During the day, algae derive their energy and reducing power from photosynthesis and the TCA cycle. At night, algae are thought to derive energy primarily via the TCA cycle. Thus, the overall rate of lipid accumulation is the difference between the rate of biosynthesis of lipids during the day and the rate of oxidation of lipids during the day and night. The enzymes of the TCA cycle also provide essential functions by producing 4- and 5-carbon skeletons required for the biosynthesis of amino acids. We examined the day versus night expression of the mitochondrial TCA cycle enzymes. Consistent with their roles in biosynthesis, all of the TCA cycle enzymes are highly expressed during the day and night. Importantly, 2-oxoglutarate dehydrogenase is induced more then two-fold at night, consistent with the energy generation role of the TCA cycle and potentially enhanced fatty acid oxidation and lipid turnover at night.

2.2.7.2. Expression of the lipid droplet protein. The gene encoding for the *Nannochloropsis*-specific lipid droplet protein identified by Benning and coworkers [45] is one of the highest expressed genes identified in our transcriptome experiment. The gene encoding for the lipid droplet protein was induced 4.6-fold more after nitrogen was depleted from the culture medium (Day 12), consistent with the increase in lipid storage vesicles during lipid accumulation.

2.2.7.3. Fatty acid biosynthesis and TAG biosynthesis. The expression of the genes encoding for fatty acid biosynthesis in *N. salina* are shown in Fig. 11. All of the fatty acid biosynthesis genes are expressed in all six transcriptome samples. The only gene in fatty acid biosynthesis that is induced following an initial reduction in expression during nitrogen depletion is *fabH*, which encodes one of the 3-ketoacylACP synthase gene family members [46,47]. *FabH* is induced two- to three-fold more after the eighth day of nitrogen deprivation. Similar to results for *Chlamydomonas* and *Picochlorum*, the expression of the gene encoding DGAT was induced when nitrate was depleted from the growth medium of *N. salina* (data not shown). DGAT catalyzes the last step in TAG biosynthesis.

2.2.7.4. Expression of genes encoding for acetyl-CoA carboxylase (ACCase). In bacteria, ACCase is encoded by four genes, *accA*, *accB*, *accC*, and *accD*, that give rise to a four-subunit enzyme. In eukaryotes, ACCase is encoded by a single gene that produces a large single-subunit (1400 amino acids) enzyme. Plants have both bacterial and eukaryote ACCase genes [48]. The bacterial-like ACCase is expressed in the plant chloroplast and thought to be responsible for *de novo* lipid biosynthesis. The eukaryote-type ACCase is expressed in the cytoplasm and is involved in the biosynthesis of specialized long-chain lipids. Algae have two eukaryotic-type ACCase genes and the genes for a four-gene bacterial type ACCase. In *N. salina*, the bacterial ACCase was expressed at very low levels throughout the 18-day nitrogen deprivation period. However, we observed that one of the cytoplasmic ACCase's gene family members was induced 3.3- to 5.6-fold more upon nitrogen depletion during the onset of TAG accumulation (Fig. 12).

2.2.8. Proteomics of N. salina

The goal of our proteomics analysis was to utilize LC-MS [49] to identify proteins whose abundance changed during the transition to lipid accumulation. For this task, proteomic surveys were carried out on *Nannochloropsis* grown in a laboratory photobioreactor and in an outdoor photobioreactor.

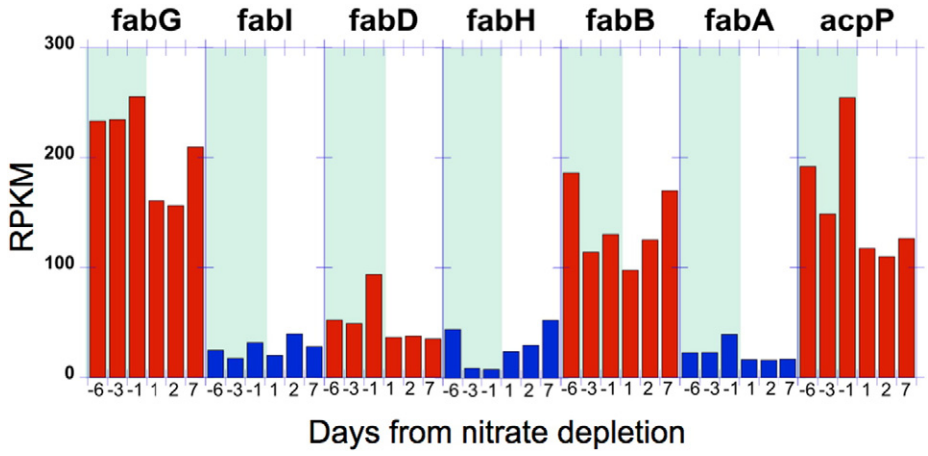

Fig. 11. Expression of genes encoding for fatty acids biosynthesis in N. salina. The genes encode for fatty acid biosynthesis enzymes as follows: fabA—3-hydroxyacyl-ACP dehydratase, fabB—3-ketoacyl-ACP synthase, fabD—malonyl-CoA:ACP transacylase, fabG—3-ketoacyl-ACP reductase, fabH—3-ketoacyl-ACP synthase, fabI—enoyl-ACP reductase, acpP—acyl carrier protein. Blue background identifies data taken prior to nitrogen depletion and a white background identifies data taken after nitrogen depletion.

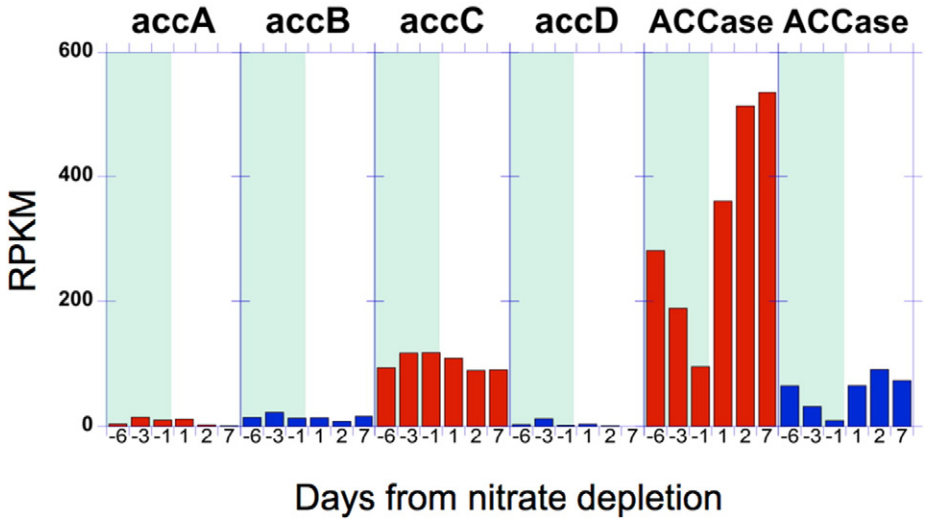

Fig. 12. Expression of acetyl-CoA carboxylase (ACCase) genes in *N. salina*. Genes labeled accA, accB, accC, and accD encode for the four-subunit bacterial enzyme, which is thought to be expressed in the chloroplast. One of the single-subunit eukaryotic ACCase genes is highly expressed and induced upon nitrogen depletion from the growth medium. Blue background identifies data taken prior to nitrogen depletion and a white background identifies data taken after nitrogen depletion.

2.2.8.1. Laboratory photobioreactor. The proteomic survey was carried out in parallel with the *N. salina* transcriptome analysis described above (Fig. 9B) where cultures were grown in a photobioreactor supplied initially with 8 mM nitrate. By Day 12, the nitrogen was depleted from the growth medium, which led to a 6.3-fold increase in the rate of lipid accumulation. The overall goal was to identify proteins differentially expressed under lipid and nonlipid productive growth states. In the proteomics analysis of the *Nannochloropsis* time course, a total of 6 LC-MS injections were performed on each of six samples of *N. salina* cultures taken on Days 5, 8, 10, 12, 13, and 18. Over 900 proteins and 2150 peptides were identified. In statistical modeling of relative protein abundance levels, 250 proteins were analyzed by spectral counting methods and 375 proteins by mixed-effect statistical modeling of mass spectral abundances. Four proteins associated with stress response, including glutathione reductase, peroxiredoxin, and superoxide dismutase, increased in abundance during nitrogen deprivation.

Peptidyl-prolyl *cis*-trans isomerase levels also increased in abundance. These enzymes have been shown to have increased abundance in plants under stres [50] and influence gene expression in eukaryotes [51]. A majority of the proteins with reduced abundance at later harvest time points are involved in protein synthesis machinery or regulation of transcription. Interestingly, adenosyl-homocysteinase levels decreased during N-deprivation. In yeast, the reduced expression of this enzyme correlates with an increase in TAGs [52]. An open reading frame with homology to ketol acid reductoisomerase also has reduced expression during N deprivation. This protein is a member of the branched chain amino acid biosynthetic pathway.

2.2.8.2. Outdoor photobioreactor. N. salina cultured in the outdoor reactor were provided replete nitrogen and grown for 22 days. Samples were taken for proteomic surveys on Days 1, 8, 15, and 22. Over 1700 proteins and 4000 peptides were identified. In all, 445 proteins were identified by spectral counting methods and 510 proteins by mixed-effect statistical modeling of mass spectral abundances. Due to the limited amount of genome similarity of *Nannochloropsis* to other annotated model organisms, a significant number of open reading frames that were identified in the proteome analysis did not have a homolog in the UniProt database. The majority of the identified proteins that had increased expression over the three-week growth period were those involved in glycolysis, including glyceraldehyde-3-phosphate dehydrogenase, fructose bisphosphate aldolase (also involved in gluconeogenesis and the Calvin cycle), and enolase. The dihydrolipoamide acetyltransferase (E2) subunit of the pyruvate dehydrogenase complex, which converts pyruvate to acetyl-CoA also increased in expression in comparison to the Day 1 sample. Glutamate decarboxylase, which also increased in the Day 22 sample, is a component of the γ–aminobutyrate (GABA) shunt, which in plants has been associated with responses to stress [53]. The majority of proteins with decreased abundance levels were involved in photosynthesis, including photosystem II D1 protein (psbA), chlorophyll *a*-b binding protein, photosystem II reaction center protein, fucoxanthin-chlorophyll *a*-c binding protein, photosystem II 47 kDa protein (psbB), and chlorophyll *a*-b binding protein. Ribulose bisphosphate carboxylase oxygenase (RuBisCO) expression was also reduced in comparison to Day 1. On the whole, these results suggest that *Nannochloropsis* metabolism at Day 22 is drastically shifted from early-stage growth as the photosynthesis machinery appears to be down-regulated and enzymes in glycolysis are up-regulated.

2.2.9. Transcriptome analysis of B. braunii

Among the oleaginous algae, *B. braunii* is unique in producing large amounts of liquid hydrocarbons (polyterpenoids and alkenes), which are comparable to fossil crude oil [54]. Biosynthetic engineering of hydrocarbon biocrude production requires identification of genes, reconstruction of metabolic pathways responsible for the production of these hydrocarbons, and mapping the biosynthesis of other metabolites that compete for photosynthetic carbon and energy. Our goal was to use transcriptomic analyses to identify the genes encoding for hydrocarbon biosynthesis.

The three chemical races of *B. braunii* are defined based on the type of hydrocarbon they produce [55,56]. The A race produces fatty acid derived alkadienes and alkatrienes; the B race produces the isoprenoid derived triterpenes known as botryococcenes; and the L race produces the isoprenoid derived tetraterpene known as lycopadiene. All three oils have been found to be major constituents of petroleum deposits, suggesting that *B. braunii* was a large contributor to the formation of these deposits. We used next-generation sequencing to assemble essentially complete *de novo* transcriptomes for all three *B. braunii* races. Next, we used a large array of bioinformatics methods to annotate the functions of these transcripts and to reconstruct biosynthetic pathways and biosynthetic networks. Using the dataset for the B strain that produces triterpenoid hydrocarbons, we manually curated pathways that affect hydrocarbon biosynthesis and export. In particular, we were interested in the biosynthetic pathways that:

- Yield the generalized terpenoid precursors DMAPP and IPP;
- Provide C_{30} botryococcenes and C_{40} lycopadienes;

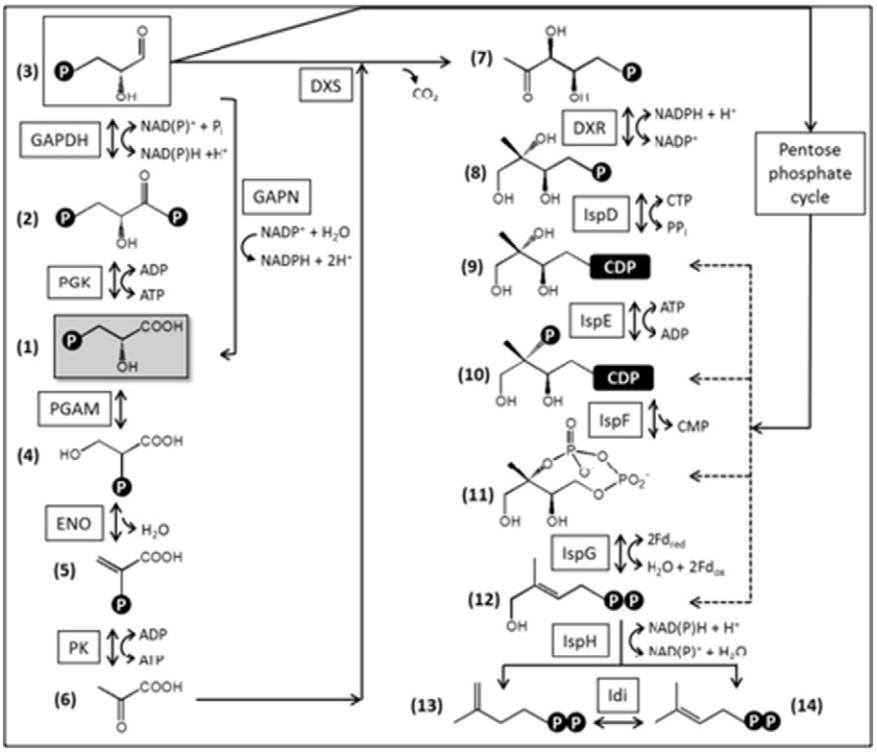

Fig. 13. The *B. braunii* race B genome encodes for a complete MEP/DOXP pathway for terpene biosynthesis. Significantly, three isozymes for 1-deoxy-D-xylulose 5-phosphate synthase (DXS, E.C. 2.2.1.7), the committed step in the pathway, are expressed.

- Provide C_{23}-C_{33} alkadienes and alkatrienes;
- Contribute to the extracellular localization of hydrocarbons; and
- Channel photosynthetic carbon and energy into non-hydrocarbon storage compounds.

2.2.9.1. Botryococcene biosynthesis. For the B race, we used the transcriptome to reconstruct the entire pathway for botryococcene production through the isoprenoid pathway [57]. Terpenes are biosynthesized from the universal C_5 building blocks of isopentenyl diphosphate (IPP, **13**) and dimethylallyl diphosphate (DMAPP, **14**) (Fig. 13). These precursors originate from the mevalonate pathway in the cytosol of animal, fungal, archaeal, and higher plant cells, while the methylerythritol 4-phosphate/deoxyxylulose phosphate (MEP/DOXP) pathway is operational in plant plastids and many Gram-positive and Gram-negative Eubacteria [58]. Experimental evidence from *B. braunii* and other algae also argues for the exclusive utilization of the MEP/DOXP pathway by the green algae.

lineage. Fittingly, exhaustive searches of the assembled race B transcriptome identified expressed sequence tags (ESTs) only for the first two of the six enzymes of the mevalonic acid (MVA) pathway. Because these enzymes are also involved in various catabolic processes, their presence in the race B transcriptome is not an indication of a functional MVA pathway. In contrast, a complete contingent of deduced enzymes for the MEP/DOXP pathway is well represented in the *B. braunii* race B transcriptome [58]. The MEP/DOXP pathway uses D-glyceraldehyde 3-phosphate and pyruvate as its metabolic input. Multiple isoforms of the key enzymes for the biosynthesis of these precursors from photosynthetic 3-phospho-D-glycerate (**1**) were identified in the race B transcriptome [59]. Some of these transcripts are present at very high abundance (>250 reads/kb), suggesting a high metabolic flux in *B. braunii* race type B.

Of particular interest, genes were identified for three isozymes of the first enzyme of the MEP/DOXP pathway, 1-deoxy-D-xylulose 5-phosphate synthase (DXS). While multiple isozymes of DXS are routinely found in plants [60], genomic evidence shows that strains of green algae harbor only a single DXS each. Recently, biochemical characterization of three isozymes of DXS from race B demonstrated that all three isozymes are active and have similar kinetic parameters [61]. Our data also indicate similar, moderate transcript abundances for all three DXS isozymes of the race B strain. DXS has been described as one of the rate-limiting steps of the MEP/DOXP pathway in plants [58]. The expression of three isoforms of this enzyme in *B. braunii* race B thus might provide an increased metabolic flux for the production of terpenoid precursors.

Curated contigs were also identified in all subsequent steps of the MEP/DOXP pathway. Each of the predicted enzymes for the downstream half of the pathway (IspF and onwards) are encoded by single genes with high to very high sequence coverage, indicating vigorous transcription and perhaps robust metabolic flow through these enzymes.

2.2.9.2. Terpenoid backbones. Terpenoid backbones were generated by the stepwise addition of IPP (**13**) with allylic polyprenyl diphosphates catalyzed by prenyl diphosphate synthetases. We identified genes encoding for synthetases for the production of C_{10} geranyl diphosphate, C_{15} farnesyl diphosphate, and C_{20} geranylgeranyl diphosphate. Interestingly, two genes encoding for putative isozymes of farnesyl diphosphate synthase (FDPS) with 72% amino acid identity were identified in the race B transcriptome, both with moderate sequence coverage. We also identified in the race B transcriptome the 3-squalene synthase-like genes that catalyze the condensation of two C_{15} farnesyl diphosphate molecules to produce the C_{30} squalene and botryococcene backbones. Finally, we identified six methyl transferases in the transcriptome that are used to mature the methylated $C_{31–37}$ botryococcenes and the methylated $C_{31–34}$ squalenes.

2.2.9.3. Alkadiene and alkatriene biosynthesis. The A race transcriptome was used to identify potential genes that are involved in alkadiene/

triene biosynthesis. The first step in alkadiene/triene production is the elongation of the fatty acid oleic acid (18:1 *cis*-Δ9) and/or its isomer elidaic acid (18:1 *trans*-Δ9). This elongation is similar to that which takes place in the production of waxes for the leaf cuticle in terrestial plants. Thus, we searched the A race transcriptome for contigs similar to the fatty acid elongation genes from terrestial plants and identified six candidate genes. These genes are currently being cloned for future characterization.

2.2.9.4. Lycopadiene biosynthesis. In the L race the biosynthesis of lycopadiene is predicted to require a gene similar to the squalene synthase (*SS*) gene. Thus, the L race transcriptome was screened for contigs that are similar to SS. Two L-race contigs were found that have significant homology to SS and they have been cloned for characterization of the enzyme activity of the encoded protein. In these enzyme characterization studies, an in vitro cell free enzyme assay was developed for the production of lycopadiene. This assay entails using an L-race protein extract with NADPH, the substrate geranylgeranyl diphosphate (GGPP), and analyzing the reaction products by GC–MS. Two molecules of GGPP are expected to be condensed to produce lycopadiene or a precursor with a higher nubmer of double bonds. This assay will be used with bacterial protein extracts from cells expressing the cloned *SS*-like genes to test the activity of the encoded proteins.

2.2.10. Summary of systems biology experiments

2.2.10.1. Lipid-producing strains. Analysis of the transcriptome was essential to build gene models and correlate gene expression with each phenotype. We used transcriptomics to examine nitrogen-deprivation–induced production of lipids by three green algae, *C. reinhardtii*, *A. protothecoides*, *Picochlorum* sp., and a stramenopile *N. salina*. The available annotation of the *Chlamydomonas* genome has facilitated a more complete analysis of the transcriptomics data. While the experimental protocols were distinct for each of the organisms, several trends emerged from comparing these three transcriptomes. First, in all three organisms, several of the acyl transferases, particularly DGAT, are induced during the lipid production phase. Similarly, in *Auxenochlorella* glucose-induced TAG accumulation also was correlated with the expression of genes involved in TAG biosynthesis. In *Picochlorum*, glycerol-3-phosphate dehydrogenase and glycerol-3-phosphate acyltransferase are induced during the lipid production phase. Although structurally unrelated, genes encoding for lipid-droplet–associated proteins were induced in the nitrogen-depleted medium in both *Chlamydomonas and Nannochloropsis.* Genes encoding for TAG biosynthesis are important targets for metabolic engineering, which we are testing. In *Auxenochlorella* it was evident that stress genes were also induced during nitrogen deprivation or decane treatments, which rapidly (<48 h) induced oil accumulation to maximal levels. We also observed the induction of some stress genes in *Picochlorum*.

In all three organisms, recognizable genes involved in lipid biosynthesis are all expressed, but only in *Picochlorum* are they modestly induced (~about 2-fold) after nitrogen depletion from the growth medium. In *Picochlorum* and particularly in *Nannochloropsis*, the cytoplasmic acetyl-CoA carboxylase gene was induced during lipid production. In *Chlamydomonas* ACCase, expression in *sta6* and *cw15* showed no induction during N deprivation. However, genes predicted to encode biotin synthase genes were induced in *sta6*.

Interestingly in *Chlamydomonas*, genes encoding for enzymes in the central carbon metabolism pathways are induced during lipid biosynthesis. Enzymes of the pathway for acetate and acetyl-CoA assimilation are induced, including acetyl-CoA synthase, isocitrate lyase, and malate synthase. These genes were all expressed to modest levels in *N. salina* at all tested time points. The isocitrate lyase gene is induced three-fold more during the dark cycle and modestly (1.4-fold) during high lipid production.

In addition, *Chlamydomonas* genes encoding for steps in gluconeogenesis and the pentose pathway, including phosphoenolpyruvate carboxykinase, fructose-1,6-bisphophate phosphatase, and transaldolase, were all highly induced during lipid production. We have postulated that increasing the carbon flux through these pathways is necessary to generate reducing equivalents in the form of NADPH for lipid biosynthesis. Alternatively, over-expression of these genes may facilitate carbon scavenging. Engineering approaches that increase the flux through gluconeogenesis and pentose pathway flux may be necessary to increase fatty acid biosynthesis.

2.2.10.2. Hydrocarbon-producing strains. Our study reconstituted metabolic pathways related to the biosynthesis of terpenoid hydrocarbons. We followed the fate of photosynthetic carbon from 3-phosphoglycerate to the general terpenoid precursors IPP and DMAPP, then onwards to the production of linear polyprenyl backbones and the biosynthesis of triterpenoids, botryococcene, and squalene to yield liquid hydrocarbon compounds, matrix structural materials, and possible routes for the extracellular localization of these compounds. Metabolic pathways leading to other terpenoids have also been reconstructed, and anabolic pathways for competing storage compounds (TAG and polysaccharides) were similarly mapped.

A recurrent theme in the terpenome biosynthesis in *B. braunii* was the expansion of particular gene families. This allows the adaptation of the paralogs to structurally orthogonal substrates (botryococcene methyltransferases) and permits neofunctionalization to support novel biochemical reactions (botryococcene synthesis). Paralogs may enhance increased metabolic flux, or they may provide additional flexibility in terms of regulation, compartmentalization, and biochemical properties (deoxyxylulose phosphate synthase and farnesyl diphosphate synthase).

The reconstructed metabolic networks, their participating enzymes and the corresponding cDNA sequences provide a genetic and metabolic framework that should empower biosynthetic engineering approaches targeting the increased production of hydrocarbons in *B. braunii*. Even more relevant to the objectives of NAABB, these pathways/genes may be mobilized into genetically tractable photosynthetic (bacterial, algal, or terrestial plant) hosts or heterotrophic microbial strains.

2.3. Algal transformation

2.3.1. Photobioreactor array for phenotype characterization of transformants

Because light is the main energy source for algal growth, cultivation of microalgae poses unique challenges. This is particularly true in the laboratory where it is very difficult to reproduce solar light intensity. Previously available laboratory-scale photobioreactors did not attempt to imitate the dynamic environmental conditions found under production conditions in outdoor ponds. This project had three major goals: (1) to produce a new type of environmental photobioreactor (ePBR) that simulated key environmental parameters and could be used in the laboratory to predict the productivity of algal strains under production pond conditions; (2) to produce an ePBR that was small and relatively inexpensive so that it could be arrayed in a laboratory to rapidly test algal strains or growth conditions in parallel; and (3) to use the ePBR to explore the importance of environmental light fluctuations in controlling the efficiency of light capture in algal and cyanobacterial water columns. For the first goal, we designed the ePBR to simulate the abiotic features of a pond that have the greatest influence on algal photosynthesis and growth: light intensity and quality, temperature, gas exchange, and natural dynamics. Shown in Fig. 14 is our environmental photobioreactor, a laboratory-scale platform for growing algae under simulated natural environments [62]. We have designed a columnar vessel to mimic a water column in an algal production pond by using collimated white light from a high-power LED to reasonably reproduce both the intensity of sunlight and the light gradient throughout the

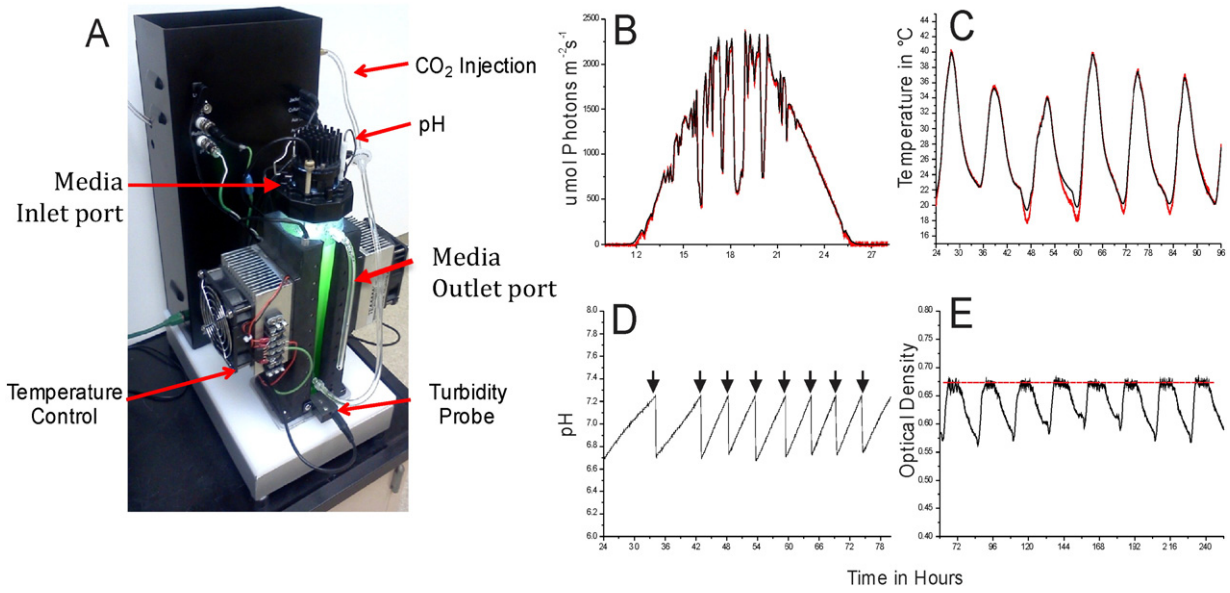

Fig. 14. The environmental photobioreactor (ePBR) in A is a modular unit capable of simulating an outdoor water column. The ePBR can simulate dynamic environmental conditions within the culture vessel. Solar data from Altus, Oklahoma, was programed into the ePBR at 5 minute intervals (black line) and actual light intensity output from the reactor was recorded with a data logger connected to a PAR meter (B). Culture temperature as programmed into the reactor (black line) and temperature as recorded by the ePBR (red line) (C). The ePBR can control pH via CO_2 injection, as marked by arrows (D). When run in turbidistat mode with a 12 h day/night cycle, the reactors can maintain a constant (or target) optical density during the day the as denoted by the red line (E).

water column. The ePBR provides programmable control over light, mixing, temperature, and gas flow while autonomously measuring optical density and pH. Scripting allows the ePBR to create complex environments and react to its own data. Additionally, the system is scalable for parallel and matrix experiments. We designed, prototyped, and collaborated with Phenometrics Inc. to produce ePBR arrays.

2.3.2. Algal transformation pipeline: engineering Chlamydomonas for greater lipid production

A major deliverable of the NAABB program was to demonstrate proof-of-concept for advanced biomass and oil accumulation in genetically engineered algae. Two approaches were used for proof-of-concept including: (1) engineering the model fresh-water algae *Chlamydomonas reinhardtii* and (2) engineering new potential production strains of algae that have greater biomass production potential than *Chlamydomonas*. Because these engineering efforts were largely built on the genome sequencing and transcriptomics experiments previously described in this report, the engineering efforts were initiated in the last 14 months of the NAABB program by developing an algal transformation pipeline. Essential to this pipeline was the development of a robust bioreactor array to rapidly test the phenotype of the engineered strains.

2.3.3. Transformation pipeline

2.3.3.1. Selection of genes of interest. The Systems Biology Advisory Committee of NAABB evaluated transcriptomics, proteomics, and metabolomics data from the algal biology team partners to identify gene targets that may increase biomass yield, increase oil content, enhance harvestability, or increase the extraction efficiency for harvesting algae or lipids. Candidate genes were prioritized for transgenic expression in targeted algae based primarily on evaluation of their potential impact on biomass productivity and oil yield.

2.3.3.2. Vector construction. The team designed and verified the construction of all vectors and plasmids containing the genes of interest. All genes of interest were codon-optimized for the relevant algal expression system. For siRNA constructs, an algal intron was often included between the ~200 bp (5′ to 3′) and ~200 bp (3′ to 5′) fold-back elements to enhance μRNA expression. All proposed fold-back constructs were screened against genomic databases to determine if there were potential nontarget hits for any 21- to 24-mer siRNA products. No more than 15 sequential nucleotides can be conserved between the fold-back construct and a nontarget gene. All DNA constructs were also sequenced for quality control purposes. This verification included: (1) sequencing all constructs to confirm that there were no mutations or frame shifts and that all essential elements (5′ and 3″ UTR, transit peptides if required, etc.) were present in the construct design; (2) in silico translation of all sequenced constructs to ensure proper protein synthesis and no false start sites in the constructs; and (3) cataloguing all constructs with complete restriction and gene maps, DNA sequences, and diagnostic polymerase chain reaction (PCR) primer sets.

2.3.3.3. Algal transformation. We developed an algal transformation pipeline in which all transformations were repeated at least three times for each DNA construct. At least 10 independent transgenes were isolated for each construct. For chloroplast transformation in *Chlamydomonas*, we used a particle gun. For nuclear transformation of *Chlamydomonas*, the glass bead system was used to minimize copy number integration using cell wall-less strains. Molecular characterization of transformants included PCR verification of the presence of an intact transgene in all putative transgenes. Transformants were enriched by antibiotic selection during phototrophic growth on agar plates.

2.3.3.4. Transformation vectors. For *Chlamydomonas* transformation, two vector systems were used. For nuclear transformation, we used the PSL18 vector in which the transgenes were driven by the strong psaD promoter/terminator pair [63]. These genes were introduced into *Chlamydomonas* using the glass bead procedure for wall-less strains and the particle for walled or chloroplast transformation events. For chloroplast transformation, we linked the transgene of interest to the appropriate promoter-terminator pair or linked the transgene to the psbA gene to complement the psbA deletion mutants.

2.3.3.5. Algal phenomics. Using the PBR array developed by NAABB, we characterized the phenotypes of the transgenes. To demonstrate gene expression, RT-PCR or Q-PCR analyses were carried out and normalized to expression of actin or B-tubulin. All transgenes were to be screened for (1) growth rate by flow cytometry and absorbance at 750 nm, (2)

terminal dry weight at the stationary phase, (3) high throughput lipid content by Nile Red fluorescence flow cytometry and gas chromatography–mass spectrometry flame ionization detection (GC–MS-FID). More focused assays were carried out as appropriate including western blots, enzyme assays, metabolomics, etc.

2.3.4. Engineering Chlamydomonas reinhardtii with enhanced biomass and oil yield

Chlamydomonas was chosen for initial surveys of gene targets for several reasons. *Chlamydomonas* was the first algal species that was genetically transformed. It was also the first organism in which the chloroplast and mitochondrial genomes were engineered. In addition, *Chlamydomonas* has two mating types and so it is possible to introgress transgenes into the progeny resulting from sexual crosses. Importantly, the *Chlamydomonas* genome has been sequenced and a robust array of mutants is available, including deletion mutants that can be functionally complemented, thereby greatly facilitating the selection of transgenics. One challenge for *Chlamydomonas*, however, is that it stores energy predominantly in the form of starch rather than oil. Mutants (*sta6*) that are impaired in starch production, however, store TAG as an alternative. Here, we describe the outcomes of the transformation events that increased either biomass or oil yield. Over 50 independent gene constructs were tested either by expression in the nuclear or chloroplast genome. The greatest yield increases observed were two-fold in biomass and five-fold in oil levels.

Our initial targets for overexpression were identified in the transcriptomics experiments described previously, including: (1) overexpression of DGAT to increase the rate of triglyceride biosynthesis; (2) overexpression of the glyoxylate cycle enzymes isocitrate lyase and malate synthase; and (3) overexpression of the lipid storage droplet protein. In addition, we developed a strategy to increase biomass production by increasing the rate of carbon assimilation.

2.3.4.1. Overexpression of DGAT. The enzyme DGAT catalyzes the last step in triglyceride biosynthesis (Fig. 15). In the transcriptome data we analyzed for *C. reinhardtii*, *Picochlorum* sp., and *N. salina*, there was a correlation between increased lipid production and increased transcription of the genes encoding for DGAT. Therefore, we over-expressed the gene DGAT from *Arabidopsis*. The DGAT gene was codon optimized for *Chlamydomonas* and cloned into the PSL18 vector under control of the psaD promoter and terminator. We isolated 89 transformant colonies and confirmed 24 as PCR positive for the DGAT gene. We characterized the phenotype of 12 of the PCR positive transformants. We observed a 2.5-fold increase in accumulated lipids in transgenic *Chlamydomonas* overexpressing the plant DGAT relative to wild-type algae. This occurred with or without nitrogen, indicating that nitrogen stress was not required for the additional TAG accumulation and that TAG levels could be further elevated above those achieved by withholding nitrogen alone.

2.3.4.2. Glyoxylate cycle enzymes. Our transcriptomics studies of *Chlamydomonas* indicated that under nitrogen deprivation conditions, there is a strong induction of the expression of genes encoding for glyoxylate cycle enzymes. This is somewhat surprising because the glyoxylate cycle is thought to play a primary role in the turnover of fatty acids by assimilating acetate (2-carbon) units produced by β-oxidation of lipids (Fig. 7). However, our hypothesis is that the glyoxylate cycle is required to generate reducing equivalents under nitrogen-deprivation conditions. To test the effects of glyoxylate enzymes on lipid production, we both over- and underexpressed the glyoxylate cycle enzymes isocitrate lyase (ICL) and malate synthase (MS). Independently, we constructed interference RNAs (RNAi) designed to block the expression of ICL (ICL-RNAi) and MS (MS-RNAi). We cloned these RNAi's under control of the psaD promoter/terminator pair. Independently, we transformed *Chlamydomonas* with the ICL-RNAi plasmid or the MS-RNAi plasmid. In addition, we cotransformed using both plasmids. Transformants were isolated and confirmed by PCR to harbor the ICL-RNAi, the MS-RNAi, or both ICL-RNAi and MS-RNAi. Interestingly, post-translational gene silencing of isocitrate lyase by ICL-RNAi caused an increase in lipid accumulation in nitrogen-replete growth. RNAi-based repression of both isocitrate lyase and malate synthase caused a five-fold enhancement of oil accumulation in *Chlamydomonas* grown with nitrogen and two-fold enhancement after N deprivation to induce oil accumulation. While the increased lipid accumulation was encouraging, RNAi-based repression of both ICL and MS caused the growth rates to be reduced by 30%.

In contrast, overexpression of a *Euglena* bifunctional ICL/MS gave a 2.6-fold increase in oil content with no impact on the growth rate with or without nitrogen [64]. While this result is consistent with our transcriptomics studies, it seems inconsistent with the data presented above because the repression of the *Chlamydomonas* ICL and MS also induced lipid biosynthesis. This observation could be explained, however, by the fact that the bifunctional gene product (ICL/MS) is targeted to the mitochondria in *Euglena*. In *Chlamydomonas*, mistargeting of the ICL/MS protein to the mitochondria rather than the gyloxyzome would lead to a competition for the fate of isocitrate [65]. If isocitrate were consumed by the TCA cycle, two carbons would be lost through decarboxylation events. In contrast, the two carbons would not be lost if isocitrate were metabolized by the ICL/MS enzyme, thus increasing the efficiency of carbon utilization from acetyl CoA entering the mitochondria and perhaps allowing for elevated acetyl CoA pools for fatty acid synthesis. Currently, we are confirming the localization of the ICL/MS in *Chlamydomonas* mitochondria and its impact on acetyl CoA pools.

2.3.4.3. Overexpression of the lipid storage droplet protein. An additional strategy for enhancing end-product (oil) accumulation is to enhance oil storage by overexpressing proteins or enzymes involved in controlling lipid storage or turnover. Previous studies demonstrated that during glucose-induced oil storage in *A. protothecoides*, the levels of the

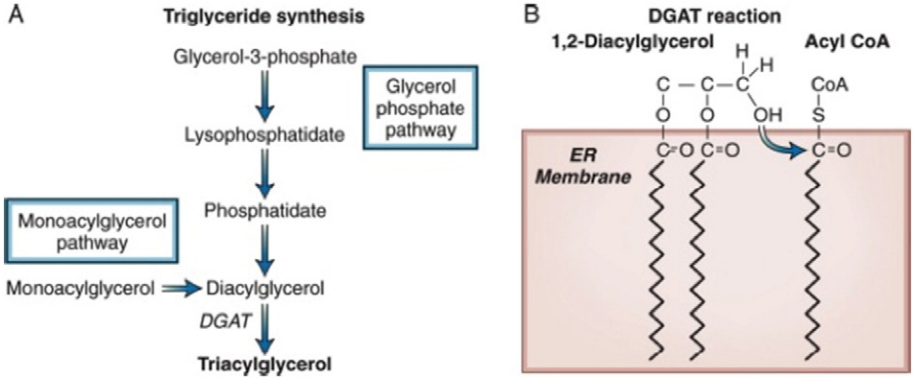

Fig. 15. TAG biosynthesis (A) and the reaction catalyzed by DGAT (B). By overexpressing DGAT, we increase the metabolic flux toward triglycerides and the overall lipid yield in algae.

lipid storage droplet protein, caleosin, increased substantially. We made a similar observation in our transcriptomics studies, which demonstrated a strong induction of the lipid storage droplet protein during high lipid biosynthesis in *Nannochloropsis*. To mimic this effect, we overexpressed the plant lipid storage droplet protein, oleosin, from *Arabidopsis* in *Chlamydomonas*. We observed a five-fold increase in lipid accumulation in nitrogen-replete growth with no loss of biomass yield. This was the single largest increase in lipid accumulation observed for any single transgene.

2.3.4.4. Increasing the rate of carbon assimilation. In cyanobacteria and many eukaryotic algae with active inorganic carbon concentrating mechanisms (CCM), the enzyme carbonic anhydrase (CA) is colocalized or concentrated around RuBisCO either in carboxysomes or pyrenoids. CA accelerates the interconversion of bicarbonate and carbon dioxide with a relatively low equilibrium constant. Since bicarbonate is the major form of inorganic carbon that is pumped into cells and since RuBisCO fixes only carbon dioxide, bicarbonate must be converted back to CO_2 to be fixed by RuBisCO [66]. We hypothesized that linking CA, in this case human carbonic anhydrase II, one of the fastest enzymes in nature, to the C-terminus of the RuBisCO large subunit with linkers of various lengths (3–43 amino acids) would effectively localize CA with RuBisCO, thereby increasing the local concentration of CO_2 and accelerating carbon fixation. These constructs were introduced into an *RbcL* deletion mutant of *Chlamydomonas* and restored autotrophic growth, indicating that the gene fusions were functional. We are currently evaluating the relative photosynthetic activity of these various constructs.

2.3.4.5. Optimizing light-harvesting antenna size. Previous studies had demonstrated that intermediate-sized peripheral (Chl a/b binding) antenna sizes were optimal for growth in *Chlamydomonas* [67]. In this project, we compared the growth rates of algae in which the accumulation of chlorophyll *b* was light regulated so that it decreased at high light levels. This regulation was achieved by controlling the expression and binding of the NAB1 protein to a 5′ light responsive element of the mRNA fused to the *Cao* gene. As shown in Fig. 16, as much as a two-fold increase in biomass was achieved with the best performing transgenics when grown in ePBRs mimicking a typical summer day. This gene (trait) conferred the greatest increase in biomass productivity of any tested by the NAABB consortium.

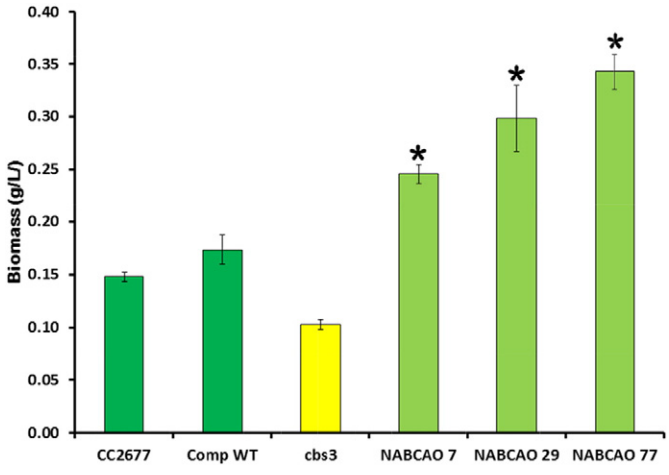

Fig. 16. Cellular and dry weight productivity of wild-type and transgenic Chlamydomonas algae with self-adjusting light-harvesting antenna sizes. The NABCAO lines (green bars) have been engineered to self-adjust their Chl a/b ratios and hence peripheral light-harvesting antenna size in response to changing light levels or culture densities. Strain CC2677 is the parental wild-type strain for all other strains tested; strain cbs3 is a mutant in which the chlorophyll *a* oxygenase (Cao) gene, which converts Chl *a* to Chl b, has been inactivated and is the parent strain for the NABCAO lines; strain Comp WT is the cbs3 line complemented with the Cao gene driven under the control of a Cab1 gene promoter and terminator. Chlorophyll *a*/b ratios ranged from 4.2 (Day 6) to 3.4 (Day 12) for the best performing strain (NABCAO 77). Chlorophyll *a*/b ratios (2.5) did not change over the time course in the Comp WT strain. The Chl *a*/b ratio for CC2677 was 2.3.

2.3.4.6. Transformation of Nannochloropsis and Picochlorum. As discussed above, we have characterized the growth and lipid production of *Picochlorum* transformed with *BIC A* and *ACCase* genes. *Picochlorum* transformed with *ACCase* show normal growth, but a significant 27% increase in accumulated lipids. Similarly, transformation of *Picochlorum* with the gene encoding for the cyanobacterial bicarbonate transporter gene, *BIC A*, show normal growth, but a 38% increase in accumulated lipids. We are in the process of characterizing *Nannochloropsis* transformed with plasmids designed to express *BIC A* and *ACCase*.

2.3.4.7. Hydrocarbon production in engineered Chlamydomonas. While much work by NAABB focused on species that store carbon as lipids, we also examined the potential for developing an autotrophic production system for hydrocarbon-like terpenoids to satisfy the unique requirements for aviation fuels. For example, the monoterpene limonene (Fig. 17) could be an important component of aviation fuels. Although algae are not known to produce monoterpenes, they make larger terpenoids, such as carotenoids. Thus, they express the biochemical machinery to produce the C5 isoprene precursors as well as the prenyl transferases necessary to produce the C10 geranyl diphosphate in the chloroplast. All that is necessary to produce limonene is to engineer a monoterpene synthase like limonene synthase (Fig. 17) so that it is expressed in the chloroplast. This study demonstrated that the hydrocarbon limonene can be produced by engineering the eukaryotic green algae *Chlamydomonas* to express the rice monoterpene synthase gene (*OsTPS26*). Using a modified nuclear transformation strategy, we engineered stable transformants expressing a functional rice limonene synthase (*OsTPS23*) targeted to the plastid. Our vector was constructed so that expression of limonene synthase is driven by the strong inducible/constitutive Hsp70A/RbcS2 promoter and targeted to the chloroplast using a stroma cTP signal in place of the rice cTP signal. In total, 10 transformant lines were evaluated for limonene production; mRNA abundance was quantified by qPCR and limonene analyte in the headspace of the photobioreactor was quantified by GC–MS analysis. *Chlamydomonas* mutant lines were cultured in enclosed vessels, allowing a purge-and-trap method of trapped volatile limonene in the headspace of the photobioreactor. Limonene was adsorbed onto volatile collection traps, which were subjected to GC–MS for product identification and quantification analyses. We demonstrated limonene production from a single transformant line at ~1 μg/g dry biomass in 72 h. We also established a baseline capability of monoterpene production in algae, which will pave the way for advanced autotrophic hydrocarbon production. Additionally, we showed that limonene is emitted from the algal biomass into the headspace region of enclosed photobioreactors, eliminating the requirement of cell harvesting, dewatering, and biomass processing.

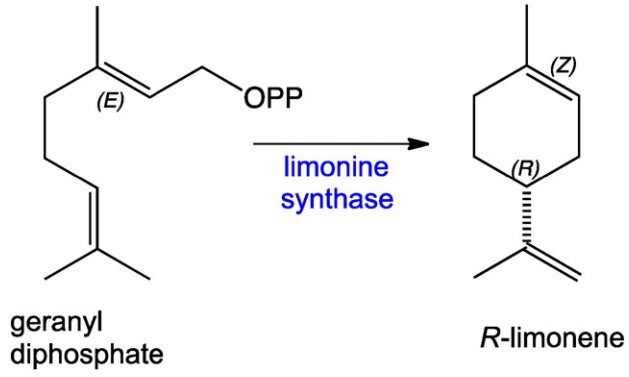

Fig. 17. The monoterpene synthase from rice catalyzes the cyclization of geranyl diphosphate to yield limonene.

2.3.5. Advanced autotrophic hydrocarbon production

While we have engineered strains that can produce limonene, our demonstrated production levels are lower by a factor ~ 10^5 than the natural lipid production systems discussed in this report. Reported in the systems biology section of this report, we have carried out extensive transcriptomics studies on the colonial microgreen alga, *B braunii* race B, which is known to be a prolific producer of the highly branched, unsaturated hydrocarbons known as botryococcenes. This freshwater alga has been reported to produce greater than 50% of its dry cell weight in botryococcenes. Although this alga is attractive as a source of biofuels, the scientific community generally recognizes that this alga grows far too slowly for practical use as a production strain. Our primary goal was therefore to unravel the molecular details associated with hydrocarbon production in *B. braunii* race B and, ultimately, engineer an algal production strain that is fast growing and produces large amounts of hydrocarbons for fuels. Discussed above careful annotation of the *B. braunii* race B transcriptome identified all the genes encoding for the biosynthesis of the enzyme isoprenoid and narrowed in on likely candidates to address bottlenecks for increasing carbon flux through the MEP/DOXP pathway to isoprenoids in a biofuel production strain. Briefly, gene candidates included *dxs I*, *II*, *III* (deoxyxylulose phosphate synthase), *dxr* (MEP synthase), *ispH* (hydroxydimethylallyl diphosphate synthase), *fps* (farnesyl diphosphate synthase) and, importantly, *sslI*, *II*, *III* (squalene synthase-like), which were found to be responsible for the head-to-head coupling of two farnesyl units to provide the C30 botryococcene, followed by *tmt* (triterpene methyltransferase), which iteratively adds a methyl group to the carbon backbone to provide the high-energy density methylated triterpenoid. The complete list genes that are likely limit carbon flux to downstream isoprenoids is found in Table 5.

However, in order to maximize carbon flux to downstream isoprenoids, the central metabolism will likely need to be altered in order to increase intracellular concentrations of the primary metabolites, pyruvate and glyceraldehyde-3-phosphate. One logical target for overexpression studies is pyruvate kinase (*pyk*) since this enzyme converts PEP to pyruvate and, importantly, was found to have at least six enzyme isoforms present in the transcriptome data from *B. braunii*. This observation could be one of the contributing factors to high botryococcene concentrations—more pyruvate (substrate to MEP pathway) *and* a source for ATP regeneration (12 mol of ATP equivalents consumed per mole of botryococcene synthesized).

Based on these findings, we are engineering isoprenoid-specific genes into the proof-of-concept cyanobacteria, *Synechococcus elongatus PCC7942* [68], and PCC7002 to determine the extent of increased hydrocarbon production as a result of overexpression of the respective enzyme. Transformations into Synechococcus 7002 are accomplished using the recombinant plasmid, pAQ1Ex-cpc, which was kindly donated by Donald Bryant (Penn State University) [69]. Gene insertion into the essential extrachromosomal plasmid in Synechococcus 7002, pAQ1, occurs through a double homologous recombination event and, in our hands, we confirmed high-levels of protein expression, working off the cpc promoter, with Yellow Fluorescent Protein (YFP) by both SDS-PAGE and flow cytometry. This construct will serve as our first generation plasmid to transport and overexpress our genes of interest for insertion into pAQ1 of 7002. We also have demonstrated a successful double homologous recombination event with a heterologously-derived deoxyxylulose synthse (dxs) through PCR. Confirmation of enzyme activity and changes in isoprenoid distributions are in progress. In addition, we recently constructed a novel transformation vector, pZ0, for gene expression studies into Synechococcus 7942. A four piece Gibson Assembly [70] utilizing the up- and downstream sequences of the neutral site (NS1) in the Synechococcus 7942 genome and the *Escherichia coli*-based backbone from pUC19 provided the core construct (Fig. 18a). The cpc-YFP-spectinomycin core was adopted from pAQ1Ex-cpc and inserted into the core construct to provide pZ0. The plasmid design enables facile exchange of both the promoter and for insertion of multiple gene casettes through standard molecular biology techniques into unique endonuclease restriction sites. YFP expression was verified through both SDS-PAGE and flow cytometry (Fig. 18b and c). In addition, the endogenous dxs, dxr, and idi genes in Synechococcus 7942 were synthesized, subcloned into pZ0 and enzyme expression verified (Fig. 18d, heterologous IDI expression by SDS-PAGE analysis).

Table 5
Genes limiting carbon flux to isoprenoid.

Gene	Enzyme	Function
dxs	Deoxyxylulose phosphate (DXP) synthase	First step in MEP pathway—condenses pyruvate and glyceraldehyde-3-phosphate to generate DXP
dxr	Methylerythritol phosphate synthase	First committed step in MEP pathway—converts DXP to MEP in an NADPH-dependent reaction
ispG	cMEPP reductase	NADPH-dependent ring opening of cMEPP to HDMAPP
ispH	HDMAPP reductase	NADPH-dependent reductive elimination of water to provide both IPP and DMAPP
idi	Isopentenyl diphosphate (IPP) isomerase	Interconverts IPP and DMAPP
sslI, II, III	Squalene synthase-like (SSL) enzyme	Depending on the combination of I, II and III, provides the C30 triterpenoids, squalene and botryococcene in an NADPH-dependent reaction
tmt1, 2, 3	Triterpenoid methyltransferase	Methylates the squalene or botryococcene carbon backbone (*S*-adenosylmethionine (SAM)-dependent)
gps	Geranyl diphosphate (GPP) synthase	Condenses one molecule of IPP to one molecule of DMAPP to provide the C_{10}-framework
fps	Farnesyl diphosphate (FPP) synthase	Condenses two molecules of IPP to one molecule of DMAPP to provide the C_{15}-framework, FPP
pyk	Pyruvate kinase	Converts PEP and ADP to pyruvate + ATP

Concomitant to our metabolic engineering research, our team was able to design, synthesize and assay a panel of ionic liquids as extractive agents for isolation of botryococcenes from both the media and directly from *B. braunii* Race B [71]. We identified several choline, triflate and phosphonium-based ILs as initial starting poinst for selective extraction of biocatalytically-derived botryococcenes should our bioengineering platforms in cyanobacteria prove successful.

2.3.6. Summary

Our metabolic engineering studies in *Chlamydomonas* are summarized in Table 6. Overall, we have demonstrated improvement in oil accumulation without a deficit in biomass accumulation using a variety of metabolic engineering strategies in *Chlamydomonas.* Oil accumulation levels increased as much as five-fold without affecting growth rates. We have also engineered transformants that overexpressed fructose bisphosphatase had significantly increased growth. In addition, engineering self-adjusting photosynthetic antenna into *Chlamydomonas* results in a significant two-fold increase in biomass accumulation. We also demonstrated that engineered *Chlamydomonas* to express rice limonene synthase in the chloroplast of *Chlamydomonas* produces small amounts of limonene.

Overall, we have demonstrated that systems biology studies can be used to direct metabolic engineering in complex algal systems and that *Chlamydomonas* can be a robust platform for testing novel gene constructs. While our *Chlamydomonas* model is likely to have the greatest implications for engineering closely related production strains such as *Chlorella*, our transcriptomics studies encourage us to test these engineering strategies in other production strains. The increases in lipid production reported here are significant. While it is not known if these improvements are additive, even a five-fold increase in the rate of lipid biosynthesis caused, for example, by overexpression of caleosin coupled with the increase in biomass yields resulting from engineering overexpression of fructose bisphosphatase and expression of self-adjusting antenna into our best production strains could lead to

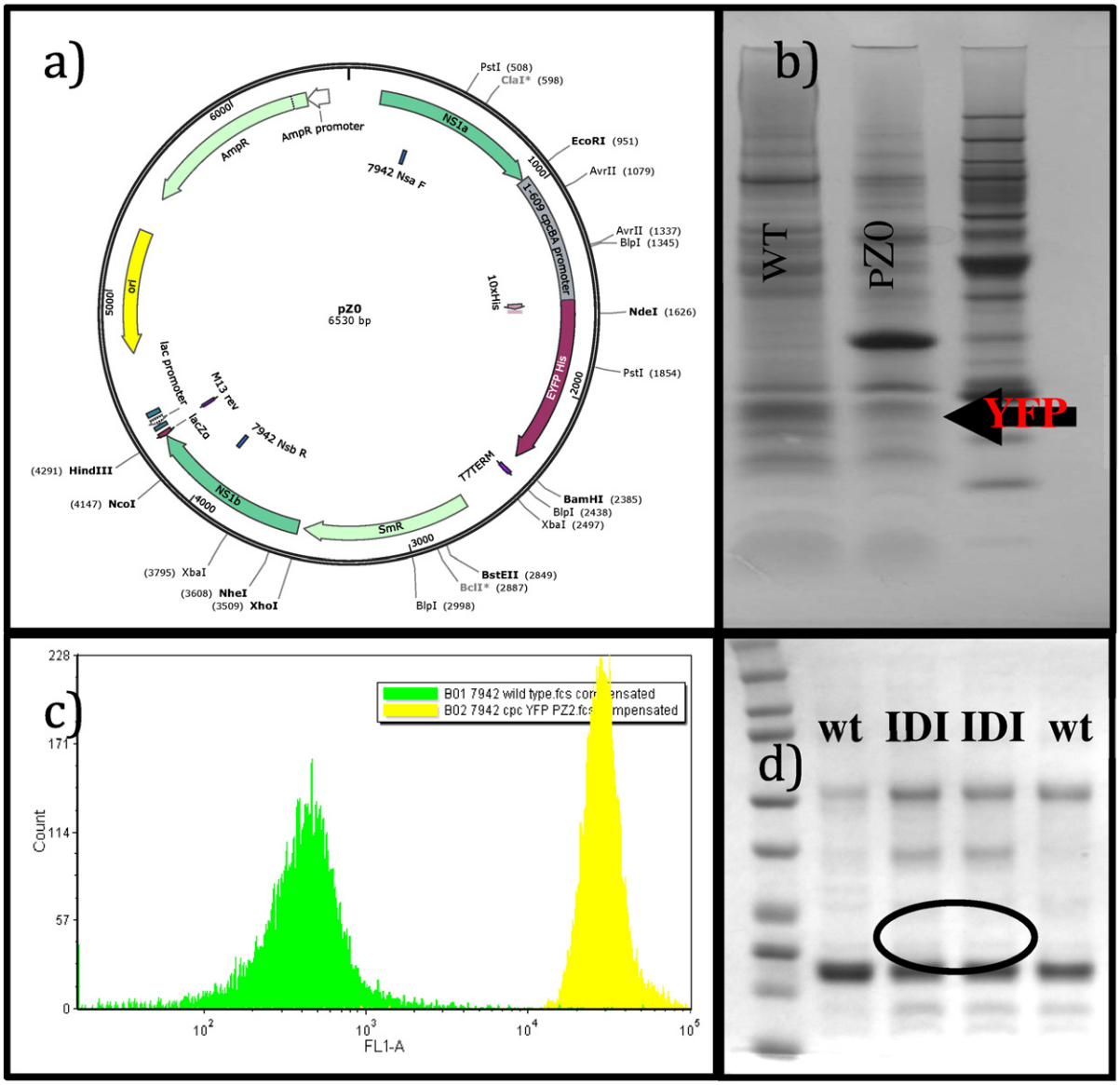

Fig. 18. a) pZ0 plasmid map. b) SDS-PAGE demonstrating high levels of YFP expression in Synechococcus 7942. c) Flow cytometry histogram on wt (left) and YFP-expressing 7942 (right). d) SDS-PAGE analysis of IDI overexpression off the pZ0-idi 7942 transformant.

economically viable algal transportation fuels. As discussed in this report, economic models show that with the 2.5-fold increase in biomass yield predicted here from combining the overexpression of fructose bisphosphatase with expression of self-adjusting antenna could enable significant progress toward a sustainable biofuels industry. As discussed elsewhere in this report, we have made significant progress in developing the genetic tools necessary to engineer production strains and we are using these tools to engineer the production strains *C. sorokiniana*, *N. salina* and *Picochlorum* sp. We have produced transformants in *Picochlorum* and *N. salina* with increased lipid production.

Table 6
Summary of metabolic engineering studies in *Chlamydomonas*.[a]

Gene[b]	Growth	Lipid accumulation + nitrogen	Lipid accumulation − nitrogen
DGAT	No effect	2.5-fold increase	2.5-fold increase
ICL-RNAi	30% decrease	2.5-fold increase	
ICL-RNAi and MS-RNAi	30% decrease	5-fold increase	2-fold increase
Bifunctional ICL/MS	No Effect	2.6 fold increase	2.6 fold increase
Oleosin	No Effect	5 fold increase	
FBPase	1.4X	ND	ND
Self-adjusting Antenna	2.0X	ND	ND

[a] For questions about strains and vectors, please contact Dr. Richard Sayre, rsayre@newmexicoconsortium.org

[b] DGAT, diacylglycerol acyl transferase; ICL-RNAi, RNAi for isocitrate lyase; MS-RNAi, RNAi for malate synthase; bifunctional ICL/MS, bifunctional isocitrate lyase/malate synthase, FBPase, fructose bisphosphatase.

2.4. Molecular biology and genetic tool development in algae

2.4.1. Gene identification strategies for the model alga C. reinhardtii

Similar to strategies that have been used for the model plant *Arabidopsis*, developing a library of algal mutants in which particular genes have been silenced using RNA interference in independent isolates could be useful in elucidating the functions of unknown genes and DNA elements or gene products that regulate the expression of other genes. RNA interference (RNAi) is an effective method in most eukaryotic organisms to reduce the expression of specifically targeted genes [72]. Commonly, RNAi is performed by substituting designed hairpins into natural pre-miRNA genes, where they are converted into microRNAs (miRNAs) by the normal cellular machinery and can lead

to down-regulation of the targeted genes [73–75]. A whole genome library of miRNA that targets every gene would provide a tool to identify those genes of interest through an efficient screening procedure. The goal of this project was to develop an inducible RNAi system that contains RNAi targeted against every gene. This was accomplished by generating a random library of genomic DNA that is converted into a precursor, which can be converted by the cell into miRNA. A genomic miRNA library was cloned into the artificial miRNA vector and transformed into *E. coli* where more than 1 million transformants were obtained. That library was then sequenced. Over 6 million reads had the expected length and matched exactly to an expected location in the *Chlamydomonas* genome. After removing redundancy, there were about 850,000 unique matches to the *Chlamydomonas* genome. Nearly every gene (more than 95%) had at least one miRNA targeted to it. The library was copied and moved into the *Nit1* promoter vector and transformed into *E. coli*. Again, we obtained over 1 million transformants. That library is now being sequenced to determine its coverage of the *Chlamydomonas* genome and the fraction of genes with at least miRNA targeted to it. We expect the coverage to be the same as for the original vector.

The library exists in two vectors, one in which the miRNA is constitutively expressed and another in which it is driven by the *Nit1* promoter and inducible by nitrogen starvation. We tested the artificial miRNA vector (amiRNA) by designing an amiRNA against the *Chlamydomonas sta6* gene and demonstrated that it effectively reduced expression of that gene, giving rise to the same phenotypes that are observed in strains in which the *sta6* gene is mutated. We then tested the new *Nit1* promoter vector using the same amiRNA against *sta6* and demonstrated that it showed the *sta6* phenotype only after being placed in media lacking ammonia. The results show that the two different vector systems we have for expressing amiRNA genes work and can be used to knock down expression of *Chlamydomonas* genes, in one case under specific inducible conditions. Screens are ongoing to select strains with desired phenotypes from which the miRNA inserts can be sequenced *en masse* to identify the gene knockdowns responsible for those phenotypes. The ultimate goal is to use this system to identify genes that up- or down-regulate the production of TAGs and other lipids.

2.4.2. Gene mapping in C. reinhardtii

Recombinant populations that have been stabilized by single-seed descent are invaluable tools in quantitative genetics in plants. *Chlamydomonas*, which is haploid in the vegetative stage, can produce segregating populations with fixed recombination events in the first generation following a cross [76,77]. In some ways, this makes it ideal for quantitative genetics studies. Once the population is genotyped, the individual lines can be phenotyped for a variety of different traits. Each trait can then be mapped for the underlying genetic architecture. The quality of the mapping improves with greater numbers of lines and greater densities of markers. We set out to create a mapping resource for the NAABB consortium and the broader algal biology community.

We created over 800 individual lines, from which we randomly sampled 384 to be the core population. We distributed the population to the Purugganan lab at NYU and the Niyogi lab at UC Berkeley. We grew 192 of the lines and made DNA samples and libraries for sequencing. These strains were then screened in various nutrient media for their relative growth rates and mineral composition to identify strains with minimal nutrient requirements as well as to characterize genes involved in nutrient metabolism. We observed a large variation in the phenotypes among the progeny. We are sequencing the libraries we made and the Purugganan lab is sequencing the other 192 in order to generate a genetic map of the mapping traits.

2.4.3. Chloroplast transformation of A. protothecoides

The unicellular green alga *A. protothecoides* is a photosynthetic aquatic species that is a member of the Scotielloideae subfamily under the Chlorellaceae family. Initially this alga was selected as a freshwater production strain due to its high lipid production. *A. protothecoides*, cultured autotrophically, rapidly produce TAGs under heterotrophic boost or stress conditions. Although TAGs are assembled in the endoplasmic reticulum, their main constituents—fatty acids—are synthesized in the chloroplast.

We sequenced the algal chloroplast genomes and used that information to design chloroplast transformation vectors for *A. protothecoides*. Intact chloroplasts from the green oleaginous microalga were successfully enriched. Quantitative PCR analyses revealed that the intact chloroplasts were effectively enriched by 2.36-fold from *A. protothecoides*, compared to the number in control samples. DNA was isolated from the enriched chloroplast fraction and seqenced using the Illumina platform to yield a total of 3,032,536 reads. The chloroplast genome sequence was determined to be 84.5 kb in size and was found to reveal 114 annotated open reading frames (putative coding regions). The genes included 32 tRNAs, 26 rRNAs, 21 photosystem subunit genes, 6 ATP synthesis genes, 6 transcription/translation related genes, 5 cytochrome genes, 4 chlorophyll genes, 2 chloroplast division relatedgenes, 1 gene encoding RuBisCO, a gene encoding acetyl CoA reductase (*accD*) involved in the fatty acid biosynthetesis as well as 7 others. The size of the cpDNA of *A. protothecoides* was relatively smaller than the cpDNA of *C. variabilis* (124,579 bp) and *C. vulgaris* (150,613 bp) due to the loss of noncoding regions. Not only were some genes missing, but the cpDNA of *A. protothecoides* showed high gene compactness (only 19% of entire genome included noncoding sequences) while *C. variabilis* cpDNA ad 46% and *C. vulgaris* cpDNA had 53%. Overall, the three *Chlorella* genomes shared numerous highly conserved genes, while they also contained unique coding and noncoding regions. Gene synteny was primarily conserved, but was variable to some extent among the three genomes. Lastly, phylogenetic analysis revealed a distinct *Chlorella* sp. clade, in relation to the more divergent species included in the analysis.

The presence of introns within target gene sequences can confound the use of gene sequences obtained from one species to transform another species. It was therefore important that we examined these genomes for the presence of introns. Group I/II introns are sporadically distributed in the nuclear, chloroplast, and mitochondrial genomes of a broad range of organisms. Most introns found in the chloroplasts of higher plants and algae belong to group II and are made up of a catalytic RNA (ribozyme; domain I–VI) and intron encoded protein (IEP) composed of four ORFs encoding reverse transcriptase, maturase, DNA binding domain, and endonuclease. The catalytic domain enables introns to self-splice and is promoted by IEP, which allows it to move the fragment into other locations in the genome. Group I and II introns were found in *Chlorella* spp. in two databases, Group I Intron Sequence and Structure Database (GISSD) and Database for Bacterial Group II introns. The GISSD contains 1789 known group-I introns, which are classified into 14 subgroups based on the structure. Most group-I introns (>95%) are found in the chloroplast *tRNA-leu* gene. To examine the presence of group I introns, the cpDNA of three *Chlorella* spp. were blasted against the GISSD database. We found only one group-I intron within the *trnL-UUA* gene in *C. variabilis* and two introns within *rrn23* and *trnL-UUA* genes in *C. vulgaris*, while none were found in *A. protothecoides*. In contrast, only partial fragments of group-II introns (16–65 nt in size, AT rich) are found in all three species (8 in *A. protothecoides*, 19 in *C. variabilis*, and 45 in *C. vulgaris*) when blasted against a bacterial Group II intron database.

The complete chloroplast genome sequences enabled us to construct three over-expression plasmid vector cassettes flanked with the chloroplast regulatory sequences, *atpA*, *psbD*/*C*, and *rbcL* and the fluorescent protein, *iLOV* reporter. Each cassette was flanked with chloroplast-derived homologous sequences for chloroplast recombination. Two chloroplast target vector systems were used. The first system used the chloroplast homologous sequences required for transgene integration and three sets of regulatory sequences (5′UTR promoter and 3′UTR terminator) from *atpA*, *psbD*/*C*, and *rbcL* were selected from resultant data

Table 7
Potential selection agents of *N. salina* and *Picochlorum*.

Selection agent	Concentration required for inhibition of growth	
	N. Salina	*Picochlorum* sp.[a]
Blasticidin	50 μg/ml	50 μg/ml
Puromycin	40 μg/ml	N.D.
Zeocin	5 μg/ml	25 μg/ml
Chloramphenicol	No inhibition	N.D.
Spectinomycin	No inhibition	N.D.
Streptomycin	No inhibition	N.D.

[a] N.D. Not determined.

from *A. protothecoides* cpDNA sequences (Fig. 19A). As a selectable marker, a paromomycin resistant gene (*aphVIII*; 1.8 kb) under regulation of *hsp70*/*rbc2* promoter was engineered in the *T*-DNA region of binary vector pGreen plasmid DNA to validate the glass bead nuclear DNA transformation method. In addition, a novel reporter gene, *iLOV*, was conjugated using the over-expression vector system for rapid screening. The second system uses a chloroplast signal peptide (ImapctVctor™1.4-tag) for chloroplast expression of genes of interest (Fig. 19B). The plasmid vector employs the RuBisCO small subunit (*RbcS1*) promoter from *Asteraceous chrysanthemum* and 1 kb of the *RbcS1* terminator sequence. In addition to the chloroplast-targeting signal peptide, the vector has a *c*-Myc-tag allowing identification of expressed proteins using commercially available monoclonal antibodies and six histidines for protein purification using a nickel column. Chloroplast target vector systems were used for *A. protothecoides* and *DOE1412* (a NAABB field isolate of *Chlorella* sp.) transformation.

The various constructions were tested using glass bead, electroporation, and particle bombardment transformation. A method that employs algal protoplasts and glass beads was used successfully for nuclear transformation of *A. protothecoides* and *Chlorella* sp. *DOE1412* using the pGreen plasmid vector bearing the *aphVIII* gene encoding paromomycin resistance. PCR confirmed transformation efficiencies for the foreign gene of interest were 4/114 clones or 3.5% and 2/96 or 2% for antibiotic resistant colonies of *A. protothecoides* and *Chlorella* sp. *DOE1412*, respectively.

2.4.4. Transformation of algal production strains

N. salina CCMP1776 and *Picochlorum* sp. were chosen for transformation studies for several reasons. *N. salina* had persisted and performed well in outdoor ponds during the testing carried out by the DOE Aquatic Species Program [7] and more recently in additional tests in outdoor ponds. This marine alga has performed well in high-salt-content waters; such waters are not suitable for agricultural or municipal uses. When grown photosynthetically with nitrogen depletion, *N. salina* accumulates triglycerides suitable for fuel to 50–60% of its biomass. No transformation method had been reported for *N. salina* nor had it yielded to transformation methods successfully used to transform a close relative, *N. Gaditana* [78]. We initiated a focused effort to transform *N. salina.* Based on our success with *N. salina*, we also transformed another production alga, *Picochlorum* sp. This second alga was selected because it has grown vigorously and accumulated significant amounts of lipid (approximately 25% of its dry weight as lipid).

2.4.4.1. Vector development. The literature suggests that strain-specific sequences in promoters and terminators are necessary for good expression in algae [63,79]. NAABB obtained the genome sequence of *N. salina*; this information allowed us to design *N. salina*-specific vectors utilizing homologous promoters and codon-optimized sequences. Our goal included incorporation of multiple target genes. We therefore needed to identify and test multiple promoters and selection agents.

2.4.4.2. Vector backbone. A suitable compact vector backbone was built by modifying the shuttle vector pPha-*T1* (GenBank Accession # AF219942). This shuttle vector had been used successfully in the transformation of the diatom *Phaeodactlyum tricornutum* [80].

2.4.4.3. Selection agents. Suitable selection agents were identified by testing a set of potential selection agents for their ability to kill *N. salina 1776* (Table 7). Three selection agents were identified as suitable and genes for resistance to these agents were included in the vector construction strategy. The most suitable selection agents for *N. salina 1776* were blasticidin, puromycin, and zeocin. Blasticidin was the easiest to use of the three because the molecule itself was stable and robust and was very effective at killing *N. salina* at a concentration of 50 μg/ml. Zeocin was also very effective but it is not as stable as blasticidin and more care must be taken to ensure its performance. Puromycin was somewhat leaky because it allowed some nontransformed cultures to grow, but it can be used if this limitation is kept in mind.

2.4.4.4. Promoters. Potential suitable promoters were chosen from our results of the transcriptome studies of *N. salina*. Candidate promoters were chosen from those regulating expression of the more highly expressed genes and reports from other algal transformations. The lipid droplet protein was the most highly expressed gene and strongly induced during lipid production (nitrogen depletion); tubulin and the photosystem I subunit protein (*psaD*) genes were also well expressed. Promoters for tubulin and psaD protein have also worked well in other transformed alga [63]. These three promoters were tested with the selection agent resistance genes used as the reporter genes. The lipid droplet protein and psaD promoters drive strong constitutive expression. The tublin promoter drives moderate constitutive expression.

2.4.4.5. Terminator. Some algal transformation efforts have reported the need to pair the promoter with a terminator unique to the promoter. We found that the *N. salina* terminator for fcpa, the fucoxanthin chlorophyll protein, works well when paired with each of our three promoters. The fcpa terminator has been used in each of our vectors. Several vectors were constructed utilizing these components.

2.4.4.6. Vectors constructed. Four vectors were initially constructed to allow for multiple gene promoters and selectable markers in different combinations to aid the gene stacking activity. In addition, the PTY1000 vector was constructed with the viral linker FMDV 2 A so that one promoter would drive expression of both the gene of interest and the resistance marker gene [81,82]. This coupling resulted in more uniform selection of transformants since the selection agent, blasticidin

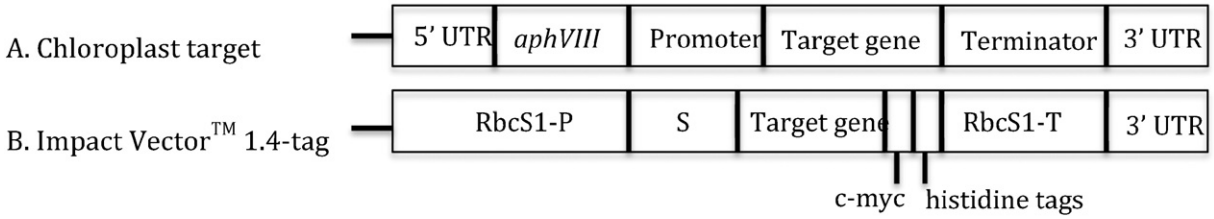

Fig. 19. Chloroplast target vector systems.

resistance, cannot be expressed unless the gene of interest is expressed also. Specifically, the vectors we constructed were as follows:

- PTY100: The gene of interest is under control of the under lipid droplet promoter and fcpa terminator and the Zeocin resistance gene is controlled by the tubulin promoter driving and fcpa terminator (Fig. 20A).
- PTY120: The gene of interest is controlled by the lipid droplet promoter and psaD (photosystem I subunit protein) terminator separately and the Zeocin resistance gene is under control of a second lipid droplet promoter and fcpa terminator pair.
- PTY1000: The gene of interest and blasticidin resistance gene are linked by the FMDV 2 A viral linker [82] and both genes are controlled by the lipid droplet protein promoter and the fcpa terminator (Fig. 20B).
- PTY423: The gene of interest and hygromycin resistance gene (hph) are linked by the FMDV 2 A viral linker, and both genes are driven by the lipid droplet protein promoter and the fcpa terminator.

Although we have prepared versions of these vectors using *Picochlorium*-specific promotors and terminators, we have demonstrated transformation of *Picochlorum* and expression of foreign genes using PTY100, and PTY1000.

2.4.4.7. Transformation method. The preparation of competent cells began with *N. salina* or *Picochlorum* grown with 100 μg/ml kanamycin to ensure the absence of the bacterial contamination that is frequently a problem in algal cultures. Cells were grown in f/2 media with 5% (v/v) CO2 and protoplasts were generated using macerozyme and cellulose for electroporation with the plasmid DNA. Cells were then cultivated in the dark overnight followed by growth in flasks in media supplemented with sorbitol and mannitol. Live cultures were then grown phototrophically (~4 weeks) on antibiotic selection plates to identify transformants followed by PCR confirmation of the transgene. Using this procedure and the plasmids discussed above, we successfully transformed *N. salina* and *Picochlorum* with a transformation efficiency of ~10^{-7}. To date, we have validated the transformation of both *Picochlorum* and *N. salina* by PCR (Fig. 21) with a cyanobacterial bicarbonate transporter gene, BIC A. Using western blots, we have confirmed the expression of BIC A in both *Picochlorum* and *N. salina*. In addition we have PCR-verified transformation with the gene for acetyl-CoA carboxylase, ACCase, the gene encoding the acyl-carrier protein, the gene encoding the lipid droplet protein, and the gene for diacylglycerol acyltransferase (DGAT) in *Picochlorum.*

We compared the growth (as biomass) and total lipid production (measured as FAMEs) of transformants with wildtype *Picochlorum* or *N. salina* cutured under identical conditions. *Picochlorum* transformed with ACCase shows normal growth, but a significant 27% increase in accumulated lipids. Similarly, transformation of *Picochlorum* with the gene encoding for the cyanobacterial bicarbonate transporter gene, *BIC A*, shows normal growth, but a 38% increase in accumulated lipids. Our initial characterization of the *BIC A* transgene in *N. Salina* shows a 36% increase in biomass productivity compared to wild-type controls. The lipid content of wildtype and the BIC A transformants was identical; thus, the overall lipid productivity was increased by 36%.

2.5. Adaptive evolution

Algal lipid production must be improved for algae to become a cost-effective fuel feedstock. Reducing the cost per barrel of algal lipids can be achieved via two strategies: increase the production/unit area and reduce the inputs. Our adaptive evolution efforts considered each of these two strategies. Greater lipid production on a per cell basis was achieved by flow cytometry sorting to isolate stable algal lines with greater lipid production while other efforts worked to reduce phosphate requirements.

2.5.1. Selecting high-lipid content Picochlorum using lipid staining and flow cytometry

Picochlorum is a marine microalgae of industrial interest due to its high lipid accumulation and its ability to grow under nonideal conditions. We used flow cytometry techniques to isolate and characterize algal strains and subpopulations of interest (Fig. 22). The flow cytometry assays were validated using more traditional monitoring methods, such as optical density for culture density and GC–MS for lipid content. Algal cultures with varying levels of neutral lipids showed distinct separation in the stained samples. Sample work-ups involved a simple dilution.

This rapid flow cytometry-based assay was used to isolate a hyperperforming subpopulation of *Picochlorum* sp. This *Picochlorum* culture was starved of nitrogen, stained with BODIPY, and subjected to multiple rounds of fluorescence-activated cell sorting (FACS). Hyperperforming and hypoperforming cells (as defined by BODIPY fluorescence) were isolated and cultured. After characterization, the parent and hyperperforming ("sorted-high") cultures were subjected to genome and transcriptome sequencing to identify genes related to hyperperformance.

2.5.2. Basic characterization

Characterization was first conducted using nitrogen-depletion experiments, where lipid accumulation was expected to be the highest. Histograms of BODIPY fluorescence of the parent and sorted populations showed that the sorted-high lipid population had a distinctly improved level of BODIPY fluorescence. When monitored over time, the sorted-high population showed consistently better performance than that of the parent. Interestingly, the sorted-low population showed

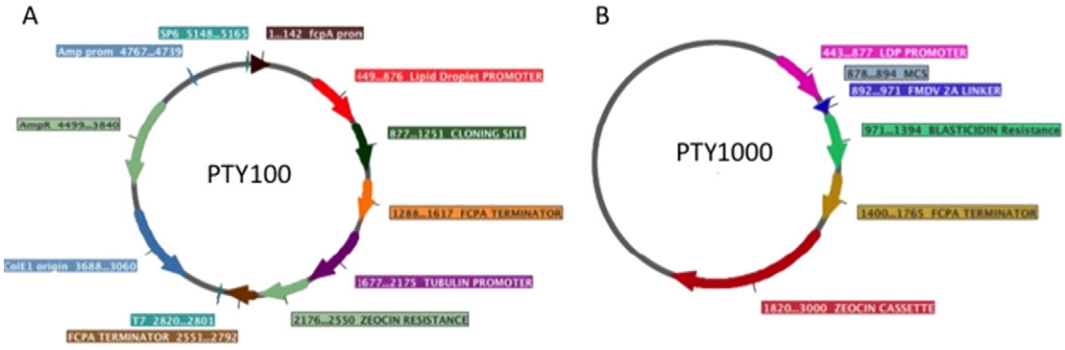

Fig. 20. Panel A: Vector pTY100, the basic early generation of an *N. salina*-specific vector. These PTY vectors (100, 110, 120, and 320) provide different combinations of genes for resistance to zeocin, puromycin, or blasticidin in combinations with strong constitutive psaD, tubulin, or lipid droplet gene promoters. Panel B: Vector pTY1000, a later generation of our *N. salina*-specific vector, utilizes the FMDV 2A linker to allow easier insertion of target genes.

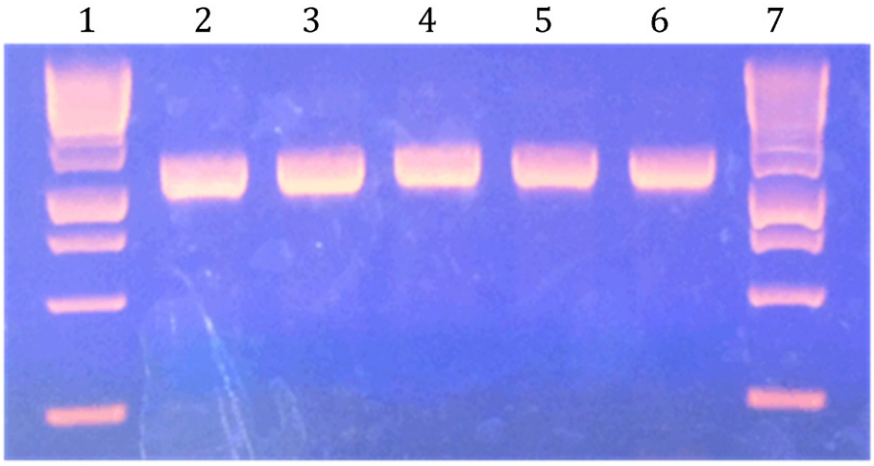

Fig. 21. Single-colony PCR analysis showing transformation of *N. salina* with a cyanobacterial bicarbonate transporter gene. Lanes 1 and 7 are DNA size ladders. Lanes 2 through 6 are single colony templates showing the BicA PCR product at 1100 bp.

the same performance as the parent. Overall, the lipid accumulation was 100% greater than in the parent strain, on a per cell basis using BODIPY 505/515. Closer examination revealed that not only was the level of lipid content higher in the sorted-high population, but the rate of accumulation during nitrogen depletion was increased by nearly two-fold relative to the parent line. An improved rate of accumulation could reduce culture turnaround time in an industrial setting. Further, the cell size as determined by forward light scatter (a correlate of cell size in flow cytometry) was increased in the sorted-high population (Fig. 23).

To ensure that the isolated subpopulation did not just have improved dye uptake and therefore show an increase in BODIPY fluorescence, sorted and parent populations were examined by microscopy and GC/MS. Microscopy validated that the parent population had smaller cells with an average of one (sometimes two) lipid bodies per cell. The sorted-high population cells were larger and had an average of two (sometimes three) lipids bodies per cell. These lipid bodies were also larger in size. Analysis by GC/MS confirmed an increase in lipid content in the sorted-high population over the parent population. The intracellular lipid content (μg lipid/cell) was increased 70% in the sorted-high population. Total biomass production in the sorted-high population was slightly reduced; however, the overall lipid productivity on a per volume basis was increased from 300 μg/ml in the parent culture to 450 μg/ml in the sorted-high culture, a 50% increase. Therefore, this sorted population shows a substantially improved lipid accumulation phenotype, in spite of a slightly reduced growth rate. Our isolated population has been stable in its greater lipid production phenotype for more than 100 generations (over a year).

2.5.3. Genomics and transcriptomics

We sequenced the genome and transcriptome of the sorted and parent populations under nitrogen replete and deplete conditions. Very few changes in protein coding genes were observed between the parent and sorted cultures. Many changes in the nearly 7000 gene transcriptomes were observed, however. Our analysis focused largely on gene IDs that have an associated KEGG assignment (891 genes). The largest number of genes showing a greater than two-fold change in the sorted-high population relative to the parent population was under nutrient replete conditions. Fewer changes were observed during depletion. This result may suggest that the improved performance under depletion is a property achieved by significant changes in expression level prior to depletion; i.e., changes in baseline expression may have a greater effect on lipid accumulation response than changes during the actual response.

Initial analysis of metabolic pathways showing differences in expression between the sorted-high and parent populations is underway. We observed increases in gene expression (sorted strain relative to parent) in the lipid biosynthesis, glycolysis, and TCA cycle pathways over the time course as the algae moved from the replete nitrogen growth phase and well into the depleted nitrogen lipid production phase. Changes were also observed in the following pathways, among others:

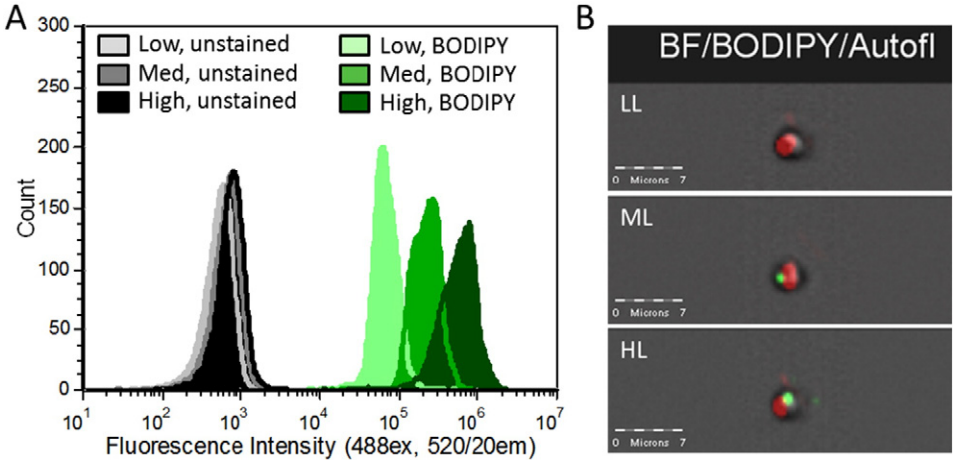

Fig. 22. (A) Fluorescence intensities of unstained and BODIPY-stained aliquots of low, medium, and high lipid (100, 180, and 400 μg/ml FAMEs) Picochlorum sp. samples. (B) Representative single cell images of low (LL), medium (ML), and high (HL) lipid algae taken on an Amnis Imagestream X imaging flow cytometer. BF/BODIPY/Autofl = brightfield, BODIPY, and autofluorescence channels are overlaid[1]

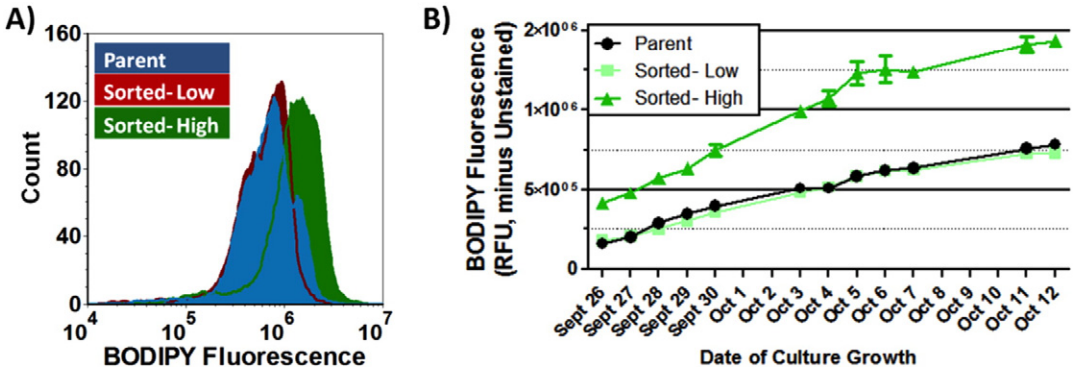

Fig. 23. Panel A: Histograms of BODIPY fluorescence of Picochlorum parent and sorted populations. Panel B: During nitrogen starvation, all cultures accumulate lipids, with the sorted-high population outperforming the parent on all days (average 2× improvement).

amino acid metabolism, starch and sucrose metabolism, purine metabolism, pyrimidine metabolism, and carbon fixation in photosynthetic pathways. Much work remains in better understanding the differences in regulation between the hyperperforming strain and the parent. Future work will include mining the genome and transcriptome data to generate a list of gene candidates for genetic modification. These genetically modified organism (GMO) strains will ideally (1) also be hyperproducers and (2) validate proposed mechanisms of hyperproduction suggested by the sequencing data.

2.5.4. Adaptation for low-phosphorous-requiring strains

Growing algae under suboptimal conditions can induce adaptation and therefore create optimized strains that will outperform the native algal strain under the selected target conditions. A prime target for adaptation is low-media-phosphate concentrations because of anticipated future limitations in phosphate supply that could be expected to impact large-scale algal production. Prolonged growth (>2 months) of *A. protothecoides* in a chemostat with a constant selective pressure of low phosphate shows initial signs of adaptation. The adapted strain appears to outperform the wild type under low-phosphate growth with a ~25% higher biomass yield. The robustness of this potentially adapted strain is currently being tested. Populations that show significant, robust improvement will be used for follow-on proteomics and transcriptomics to identify the biochemical basis for adaptation.

2.6. Crop protection

Cultivation of algae in algal raceways and open ponds is envisioned as the most economical route for algal biomass and biofuels production. In the coming decades, if the projected scales of biofuels are to be generated, algae have to be cultured on thousands of hectares of land for the necessary biomass yields. Like most plant crops, algal cultures in open ponds are susceptible to a number of environmental factors, including biological agents that gain entry into ponds. Our goal was to examine and understand the nature of biotic factors such as bacteria, viruses, invasive algal species, fungi, and herbivores in algal ponds that impact the algal biomass yields. It is broadly accepted that the biological agents that are present in the algal ponds could be (1) beneficial to algal growth, (2) harmful to algal culture, or (3) unobtrusive to the cultivated algal culture. Entry of invasive species into algal cultures can have deleterious effects on algal cultivation and could result in lowered biomass yields and even cause pond crashes, resulting in crippling losses to biofuel producers. For instance, it has been documented that certain bacteria, viruses, fungi, or herbivores such as rotifers can cause demonstrable losses to an algal crop. One of the major and imminent challenges in cultivating algal crops in an open environment is protection of algal crops from invaders. Furthermore, our investigations are also important because we hope to determine if any potential human, livestock, or plant pathogens can propagate in an algal pond ecosystem that could cause serious problems to cultivators and the environment. If this is not addressed actively, in the long run, pathogens could pose serious sustainability problems for the algae production industry. Other challenges faced included identifying economically viable strategies to reduce and eliminate contaminants without incurring additional cost in mitigating biomass losses. Furthermore, we proposed to investigate molecular approaches to contain genetically modified algae propagation outside the algal ponds.

As part of the algal biology team in the NAABB consortium, we carried out census analysis on organisms that enter and possibly thrive in algal ponds. With this knowledge, it was also our aim to determine economically viable alternatives for algal crop protection from harmful contaminants through bioengineering approaches. To understand the biological agents that co-inhabit algal ponds, we designed and conducted experiments with both laboratory-grown cultures of *C. reinhardtii* and *A. protothecoides UTEX25* (representative fresh water cultures grown in nutrient-rich media) and raceway-grown cultures of *N. salina* (a representative marine species grown at Texas AgriLife Research, Corpus Christi, provided by Dr. Tzachi Samocha). Based on the ribosomal RNA (rRNA) gene-based identification of species of contaminates in a *Chlorella UTEX25* culture, we identified more than 50 different bacterial and fungal species with bacterial populations reaching one-tenth of the algal culture at the end of seven days. Importantly, many representative bacteria belonged to the genus *Pseudomonas*, with few of them being opportunistic pathogens. For the *N. salina* contaminant survey, both rRNA sequencing and more extensive chip-based analysis of species distribution (PhyloChip–Lawrence Berkeley National Laboratory) were undertaken. Our surveys indicated that an array of over 40 identifiable organisms ranging among bacteria to fungi to viruses inhabit these raceways. Significantly, in our analyses, only one known species of human pathogen was identified. It is unclear if this strain was actively propagating in the open ponds or happened to have entered the pond at the time of sampling.

2.6.1. Agents to minimize bacterial contamination and eliminate rotifers in open ponds

Our pond survey studies conclusively proved that large numbers of bacteria can coexist with algae in open pond cultures and the occurrence of algae-invading opportunistic pathogens is certainly a looming problem. In addition, NAABB observed pond crashes caused by rotifer infestation. To address this, we explored the use of antimicrobial peptides [83,84] (AMPs) and other biomolecule production in algae for improving their innate defense against bacteria and rotifers. We successfully screened and identified one or more AMPs that kill bacteria and eliminate rotifers but not our algae of choice for biomass production. Sample results on rotifer control are presented here.

Two freshwater rotifers, *Philodina acuticornis* and *Adineta vaga*, were continuously cultured and maintained under laboratory conditions. Bioassays with rotifers were successfully developed in a multi-well plate

format to analyze the effect of tested molecules on these fresh water rotifers. As detailed in Table 8, we determined the efficacy of selected molecules on rotifers (at 0.5 mg/ml concentrations) and also on three species of green algae as detailed previously for AMP assays with algae. Based on the results, we identified several candidates that are well tolerated by algae but have detrimental/lethal effects on rotifers (Fig. 24). We are currently in the process of identifying viable bioengineering routes for developing transgenic algae for studying in vivo expression and applicability of these strains in rotifer elimination applications.

To achieve economically viable means to employ AMPs for crop protection, we have successfully developed bioengineering tools for expressing these agents in algae. Various lines of algae expressing these biomolecules have been generated to test the viability of these transgenic algae lines in the presence of deleterious bacterial and rotifer species. Based on a similar approach, new biomolecules or gene products can be identified to test their applicability in eliminating other invasive contaminants. Furthermore, we have also undertaken the task of characterizing gene-switch-mediated regulation of chloroplast genes in the model organism *C. reinhardtii* to determine its applicability in genetically modified (GM) algae containment. The regulation of the gene-switch is currently coupled to a small molecule effector that regulates the expression and/or translation of the gene coupled to this element. If this strategy is determined to be successful, we hope to extend this idea into other green algae that are genetically engineered for biofuels or other value-added product production, thus averting their proliferation in the natural environment.

2.7. Algal cell biology

The ability to visualize cell structures and correlate changes in those structures with genetic, species or culture differences is a powerful approach to making advances in our understanding of algal growth and lipid production. Two complex physiological traits of great interest in regard to the cell biology of algal biofuel production are lipid body formation and cell wall structure. Clearly, lipid body formation is of interest with respect to maximizing the amount of biofuel precursors. The structure of cell walls potentially impacts the robustness of the algae in cultivation and the extraction of the lipid during post-culture processing. Being able to observe these processes and correlate them with genetic traits provides valuable information for fundamental understanding of the underlying biology and the potential to implement genetic engineering or process engineering to elicit desirable traits.

2.7.1. Lipid bodies in algae

Quick-freeze deep-etch electron microscopy (QFDEEM) [85] was used to follow lipid body formation in three strains of special interest to NAABB, *C. reinhardtii* [38,86], *B. braunii*, and *Nannochloropsis* spp., with the goal of uncovering unique and common characteristics in lipid body formation between these diverse species. The three species come from two different families of the green algae and a stramenopile genus (*Nannochloropsis*), so it is not surprising that there are considerable differences in lipid body biogenesis. *C. reinhardtii* normally stores its excess carbon as starch. In the *sta6* mutant strain (which also has lesions in genes other than *sta6*), *C. reinhardtii* accumulates more lipid than do its starch-forming strains.

C. reinhardtii cells in N-replete media contain small cytoplasmic lipid bodies (α-cyto-LBs) and small chloroplast plastoglobules, as observed by QFDEEM. When starved for N, β-cyto-LB formation is greatly stimulated, apparently resulting in an enlargement of existing α-cyto lipid bodies (LBs) rather than the formation of new LBs. β-Cyto-LBs are in intimate association with the endoplasmic reticulum (ER) and the outer membrane of the chloroplast envelope (Fig. 25, cyto-LB), suggesting the active participation of both organelles in β-cyto-LB production/packaging. When *sta6* mutant cells, blocked in starch biosynthesis, are N starved, they produce β-cyto-LBs and also chloroplast LBs (cpst-LBs) that are much larger than plastoglobules and eventually engorge the chloroplast stroma (Fig. 25, cpst-LB). Production of β-cyto-LBs and cpst-LBs is dependent on exogenous acetate under our culture conditions and inhibited by darkness. We proposed that the greater LB yield reported for N-starved *sta6* cells can be attributed, at least in part, to its ability to produce cpst-LBs, a capacity lost when the mutant is complemented by *STA6* transgenes.

2.7.2. Lipid-bodies in stramenopiles

In association with external and internal collaborators, NAABB used QFDEEM to investigate four species of the marine stramenopile *Nannochloropsis*: *N. gaditana*, *N. salina*, *N. oceanica*, and *N. oculata*. Compared with the other algae examined in this QFDEEM survey, all *Nannochloropsis* species have many LBs even under N-replete conditions, but like other species, the LBs increase in size upon N-starvation. The log-phase LBs lie in the cell periphery and are oblong in shape, as contrasted with the central and circular LBs of the green algae. They are invariably associated with ER. The direct association between N-stressed LBs and the chloroplast envelope observed in *Chlamydomonas* is not found in *Nannochloropsis*. However, since the ER in stramenopiles completely encloses the chloroplast (the so-called "chloroplast ER") and makes direct contact with the chloroplast envelope, indirect envelope/LB associations are maintained and possibly have functional significance.

2.7.3. Lipid-body formation in Chrysochromulina

The haptophyte *Chrysochromulina* sp. is a fascinating alga in many regards. This wall-less alga that can be maintained in both fresh and brakish water is enclosed only by a simple plasma membrane. The organism has a pair of lipid bodies, each intimately associated with a chloroplast. The ability of NAABB investigators to grow the alga in synchronous culture induced by a 12 h light/12 h dark photoperiod facilitates observation of the cyclical nature of lipid body volume changes. The lipid bodies reach a maximum volume at the end of the light period and a minimum at the end of the dark period (Fig. 26). The ability to grow this alga reproducibly on a completely defined artificial media,

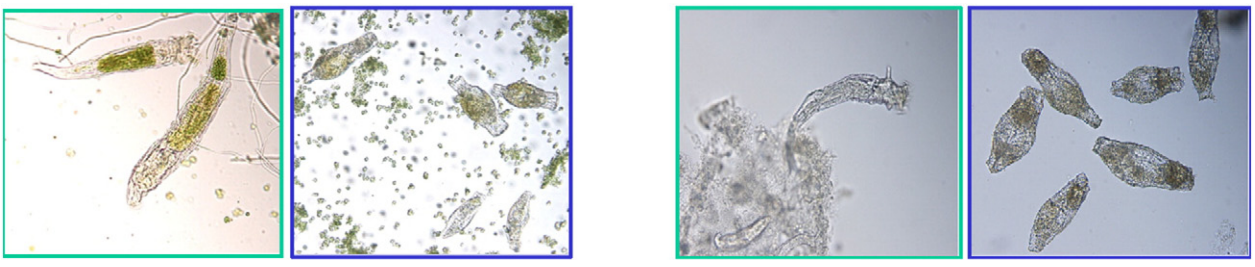

Fig. 24. Microscopic images of rotifers that are treated with molecules of choice. The left two images are *A. vaga* and the right two images are *P. acuticornis*. Live rotifers can be seen in the green-boxed images; dead ones are visible in the blue-boxed images.

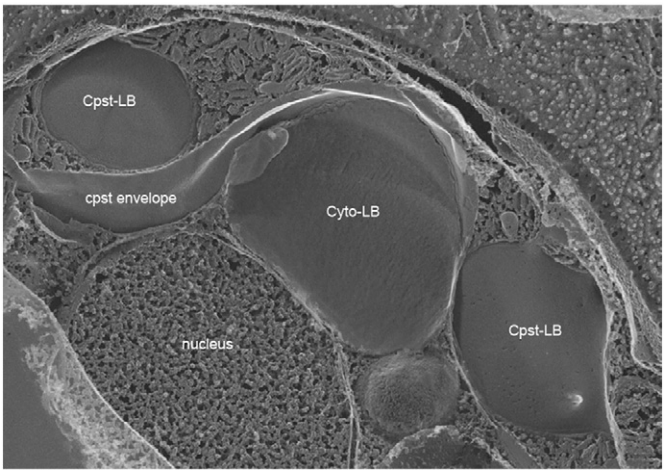

Fig. 25. Cytoplasmic and chloroplast LBs in the sta6 strain of *C. reinhardtii*.

the broad profile of fatty acids produced by the organism, and the ease and reproducibilty of fatty acid extraction suggest this alga might be a good reference standard for fatty acid recovery [87].

Extensive transcriptomic analysis of genes critical to fatty acid biosynthesis, elongation and saturation in *Chrysochromulina* cells maintained on a 12 h light:12 h dark photoperiod has been completed. Two distict gene expression patterns (lipid metabolism vs. lipid processing) occur over the 24 h cycle. These fatty acid gene expression profiles parallel the observations of the augmentation and decline in lipid quatities that influence the size of a lipid body. Besides metabolically serving as structural and storage components for a cell, many simple fatty acids are made into polyketides, which are biologically active secondary metabolites that include pharmaceuticals, antibiotics and toxins. Insight into the organization and expression of genes critical to *Chrysochromulina* synthesis of these fatty acid products during the circadian photoperiod has also been established. Most notably is the presence of a hybrid polyketide synthase – nonribosomal peptide synthetase complex. This hybrid arrangement has never been previously seen in any algal species.

2.7.4. B. braunii

Given that *B. braunii* is known for producing various hydrocarbons from terpene or fatty acid elongation pathways (depending on the type) it was perhaps surprising that it appears to produce lipid bodies that are similar to other green algae. However, *B. braunii* has unique characteristics in regard to its structural organization and composition of the cells, extracellular matrix and shell. By utilizing fluorescent staining followed by imaging and biochemical analyses, we made major advances in the understanding of this organization and the components involved. The B type of *B. braunii* was used for this study and it produces the triterpenoid botryococcenes, which, like TAGs, can be stained with Nile Red. The polysaccharides were stained with a modified periodic acid Schiff (mPAS) reagent system (utilizing the fluorescent dye propidium iodide as the Schiff reagent) for in situ visualization. The carbohydrate was isolated and its composition analyzed by derivatization GC–MS, revealing that the extracellular matrix was largely composed of an arabinogalactan with a mixture of unusual linkages. Individual and overlay images of the chloroplasts, polysaccharides, and hydrophobic botryococcenes or algaenans are shown in Fig. 27.

This summary covering some of the cell biology studies of various algae conducted as part of NAABB reveals the power of combining advanced imaging and systems biology methods with biochemical analyses to understand algal composition, structural organization, diurnal or developmental variation in fatty acid composition. Clearly a greater array of tools is now available and being utilized for meaningful studies of algae that have potential for biofuel production or have characteristics of interest in an engineered biofuel production algal strain. It is equally clear that much remains to be learned from various algal species across the taxonomic breadth of biology.

2.8. Genetic modification of algae

Life cycle and techno-economic analyses indicate that biomass yields will need to be increased at least three-fold to approach economic parity with petroleum. Thermodynamically it is feasible to achieve 11% efficiency for photon to biomass energy conversion using solar energy. Yet our estimates are that best biomass producing algae identified through our bio-prospecting efforts achieve only 3% thermodynamic efficiency. To achieve increased yields will require substantial improvements in the efficiency of light capture, energy conversion and biomass accumulation [10]. It is our hypothesis that these enhanced yields can only be achieved through genetic modification of algae. The risks associated with GM algae have been considered carefully [8]. There are two major concerns regarding GM algae; the first is whether GM algae will be evolutionarily or ecologically more fit than wild-type algae and the second is whether the GM strains will produce toxins or other molecules that are harmful to the environment. Each of these concerns has to be addressed on a case-by-case basis for the particular trait expressed. The initial risk assessment of genetically modified (GM) algae for commodity-scale cultivation concluded that most GM traits will actually reduce fitness since they will commit the engineered alga to a particular metabolic fate that reduces environmental flexibility

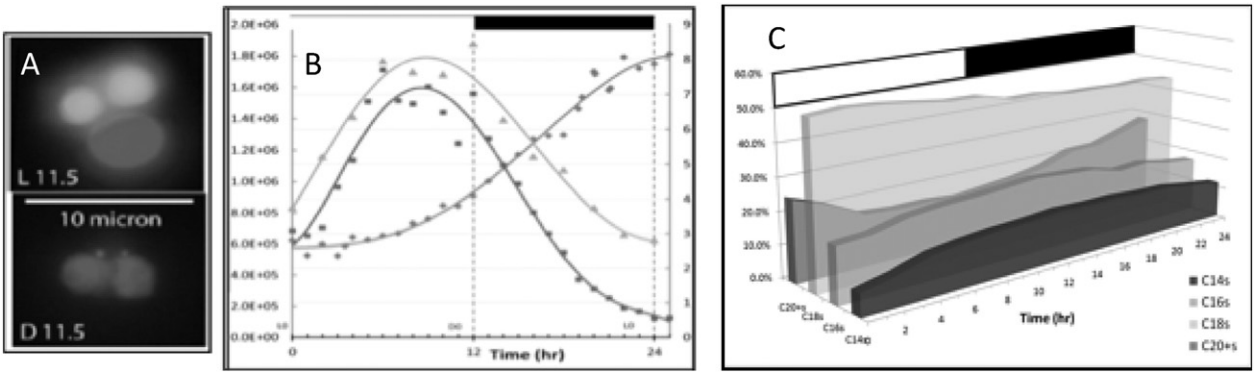

Fig. 26. Observations of lipid body formation in *Chrysochromulina* sp. during a photoperiod (light: 0 12 h; dark: 12–24 h). (A) Fluorescent micrographs of cells harvested at the end of the light period (upper panel, L11.5) and end of the dark period (lower panel, D11.5) where red is chlorophyll autofluorescence in chloroplasts and green is from lipid bodies stained with BODIPY 505/515. (B) Left axis: culture density in cells/ml (blue curve). Right axis: Neutral lipid/BODIPY 505/515 (relative fluorescent units, red) and total fatty acids (pg/cell, green). (C) Diurnal variations in fatty acid profiles of the lipids.

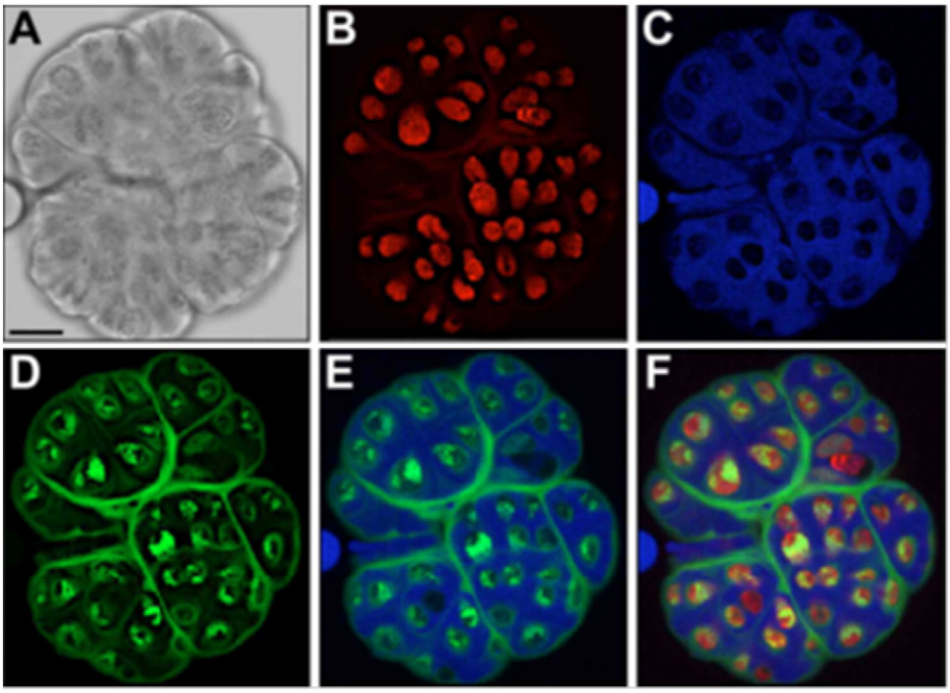

Fig. 27. AB-V-3. Imaging *B. braunii* type B colonial organization. All images were taken from a single z axis of the same colony. A. DIC image of a single *B. braunii* colony that was stained with Nile Red and mPAS. Bar, 10 μm. B. Chlorophyll autofluorescence channel. C. Nile red channel, false colored blue. D. mPAS channel, false colored green. E. Merge of Nile Red and mPAS images. F. Merge of Nile Red, mPAS, and chlorophyll autofluorescence images.

[8]. Currently, the EPA regulates GM algae and some permits have been issued.

3. Conclusions and recommendations

In the period between the Aquatic Species Program and NAABB, the age of genomics began. This enabled NAABB to take new approaches to studying biofuel production in algae, such as systems biology, and led to the development of new genetic tools. NAABB researchers generated a valuable reservoir of information pertaining to potential targets to improve algal growth and lipid productivity that has been tapped but by no means exhausted. The biology of algae for biofuels production is like every other area of biology in the genomics age—there is a wealth of information in the form of systems biology data and resultant gene targets but the hard work of gene deletion and expression analysis to understand and verify the efficacy of individual or stacked genetic variations remains a relatively time-consuming process. NAABB researchers began and continue those studies and have therefore obtained exciting and promising results on strains that have improved growth or lipid productivity. Yet, many interesting gene candidates for improving productivity in various ways (including crop protection, growth enhancement, and lipid productivity) remain in need of thorough characterization with regard to the impacts of those genetic changes. Further analyses of the systems biology data already collected would be beneficial for understanding the effects of nutrient and physical parameters on growth and lipid production in terms of gene expression and would lead to the identification of additional interesting candidates. The enormity of data sets requires increasingly sophisticated tools for analysis and

Table 8
Bioassays of three green algae and *Adineta vaga* and *Philodina acuticornis* with molecules to determine their effect on algal growth and rotifer-killing efficiency.

Compound	Inhibition of aglal growth[a]			Rotifer sensitivity[b]	
	A. protothecoides	*Chlorella sorokiniana*	*C. reinhardtii*	*Philodina acuticornis*	*Adineta vaga*
1	16 μM	66 μM	33 μM	+++	+++
2	10 μM	20 μM	10 μM	+++	+++
3	44 μM	88 μM	22 μM	+++	+++
4	78 μM	155 μM	20 μM	+++	+++
5	37 μM	37 μM	10 μM	+++	+++
6	158 μM	79 μM	20 μM	+++/++	+++/++
7	275 μM	275 μM	17 μM	+++/++	+++/++
8	48 μM	97 μM	12 μM	+++	+++
9	23 μM	46 μM	11 μM	+++	+++
10	22 μM	87 μM	5 μM	+++	+++
11	44 μM	87 μM	11 μM	+++	+++
12	12 μM	24 μM	12 μM	+++	+++
13	NK	492 μM	62 μM	++	++

[a] Concentration of compound that inhibits algal growth (μM). Algal growth was assayed for 6 days.
[b] Rotifer sensitivity: 0.5 mg/ml of each molecule was incubated with the rotifer in spring water and assayed after 24 h; +++ complete kill efficiency (100%); ++ moderate kill efficiency (50–75% death).

higher throughput methods for genetic modifications and stacking traits (multiple or recyclable markers in existing and novel systems). The tools for effective analysis of new traits have been developed, including more sophisticated pond-mimicking photobioreactors, many new algal genomes, transcriptomics data sets and analysis tools, proteomics, and lipidomics. Importantly, the team structure of NAABB allowed the development of an efficient conduit for transferring lead algal strains or genetic variants to the algal cultivation testbeds for analysis in contained or open pond environments. Former members of the NAABB cultivation team have reported a significant biomass accumulation rate in ponds (30 g/m^2/d) for *C. sorokiniana* (DOE1412), which was discovered by the NAABB algal biology team [14]. Furthermore, it has been determined that the energy content of *C. sorokiniana* biomass can increase by 50% with no loss in biomass yield during cultivation under nitrogen-free conditions [88]. It is critical going forward that this conduit remains open for analysis of genetic variants that show promise in the lab. The size of the issues and the inter-disciplinary nature of the solutions for algal biofuels production necessitate that the collaborative approach afforded by consortia, such as NAABB, have continued encouragement and support. Overall, the algal strains identified, the greatly expanded genomic resources, the systems biology approaches, and genetic tools developed should enable the broader algal biofuels community to accelerate progress in understanding and implementing existing or developing algal strains for more efficient algal biofuels production.

Acknowledgements

This work was funded by U.S. DOE-EERE Bioenergy Technologies Office, under contract DE-EE0003046 to the National Alliance for Advanced Biofuels and Bioproducts.

References

[1] R.E. Blankenship, D.M. Tiede, J. Barber, G.W. Brudvig, G. Fleming, M. Ghirardi, M.R. Gunner, W. Junge, D.M. Kramer, A. Melis, T.A. Moore, C.C. Moser, D.G. Nocera, A.J. Nozik, D.R. Ort, W.W. Parson, R.C. Prince, R.T. Sayre, Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement, Science 332 (2011) 805–809.
[2] Y. Chisti, Biodiesel from microalgae beats bioethanol, Trends Biotechnol. 26 (2008) 126–131.
[3] T.M. Mata, A.A. Martins, N.S. Caetano, Microalgae for biodiesel production and other applications: a review, Renew. Sust. Energ. Rev. 14 (2010) 217–232.
[4] R.T. Sayre, Microalgae: the potential for carbon capture, Bioscience 60 (2010) 722–727.
[5] K.M. Weyer, D.R. Bush, A. Darzins, B.D. Willson, Theoretical maximum algal oil production, Bioenergy Res. 3 (2010) 204–213.
[6] M.S. Wigmosta, A.M. Coleman, R.J. Skaggs, M.H. Huesemann, L.J. Lane, National microalgae biofuel production potential and resource demand, Water Resour. Res. 47 (2011).
[7] J. Sheehan, T. Dunahay, J.R. Benemann, P. Roessler, A look back at the U.S. Department of Energy's aquatic species program: biodiesel from algae, in: D.o. Energy (Ed.), NREL/TP-580-24190, 1998.
[8] W.J. Henley, R.W. Litaker, L. Noyoyeska, C.S. Duke, H.D. Quemada, R.T. Sayre, Initial risk assessment of genetically modified (GM) microalgae for commodity-scale biofuel cultivation, Algal Res. 2 (2013) 66–77.
[9] J.A. Olivares, R.H. Wijffels, Responsible approaches to genetically modified microalgae production, Algal Res. 2 (2013) 1.
[10] S. Subramanian, A.N. Barry, S. Pieris, R.T. Sayre, Comparative energetics and kinetics of autotrophic lipid and starch metabolism in chlorophytic microalgae: implications for biomass and biofuel production, Biotechnol. Biofuels 6 (2013).
[11] National Alliance for Advanced Biofuels and Bioproducts Synopsis (NAABB) FINAL REPORT, U.S. Department of Energy, 2014 (http://energy.gov/eere/bioenergy/downloads/national-alliance-advanced-biofuels-and-bioproducts-synopsis-naabb-final).
[12] R.A. Andersen, M. Kawachi, Traditional microalgae isolation techniques, in: R.A. Andersen (Ed.), Algal Culturing Techniques, Elsevier, Academic Press, Burlington, MA 2005, pp. 83–100.
[13] K.E. Cooksey, J.B. Guckert, S.A. Williams, P.R. Callis, Fluorometric-determination of the neutral lipid-content of microalgal cells using nile red, J. Microbiol. Methods 6 (1987) 333–345.
[14] P.J. Lammers, et al., Review of the cultivation program within the National Alliance for Advanced Biofuels and Bioproducts, Algal Res. (2015) (under review).
[15] N. Bigelow, B. Hardin, J. Barker, C. Deodato, S. Ryken, A. MacRae, R.A. Cattolico, A novel, rapid, sub-microscale in-situ fatty acid assay and applications to aquatic microorganisms, J. Phycol. 47 (2011) S58.
[16] N.W. Bigelow, W.R. Hardin, J.P. Barker, S.A. Ryken, A.C. MacRae, R.A. Cattolico, A comprehensive GC–MS sub-microscale assay for fatty acids and its applications, J. Am. Oil Chem. Soc. 88 (2011) 1329–1338.
[17] E.C. Yang, G.H. Boo, H.J. Kim, S.M. Cho, S.M. Boo, R.A. Andersen, H.S. Yoon, Supermatrix data highlight the phylogenetic relationships of photosynthetic Stramenopiles, Protist 163 (2012) 217–231.
[18] S. Bennett, Solexa ltd, Pharmacogenomics 5 (2004) 433–438.
[19] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z.T. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L.I. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P.G. Yu, R.F. Begley, J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, Nature 437 (2005) 376–380.
[20] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X.X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C.C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, Real-time DNA sequencing from single polymerase molecules, Science 323 (2009) 133–138.
[21] B.L. Cantarel, I. Korf, S.M.C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A.S. Alvarado, M. Yandell, MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes, Genome Res. 18 (2008) 188–196.
[22] A.L. Delcher, K.A. Bratke, E.C. Powers, S.L. Salzberg, Identifying bacterial genes and endosymbiont DNA with glimmer, Bioinformatics 23 (2007) 673–679.
[23] R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, S.J. Sammut, H.R. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L.L. Sonnhammer, A. Bateman, The Pfam protein families database, Nucleic Acids Res. 36 (2008) D281–D288.
[24] S. Gotz, J.M. Garcia-Gomez, J. Terol, T.D. Williams, S.H. Nagaraj, M.J. Nueda, M. Robles, M. Talon, J. Dopazo, A. Conesa, High-throughput functional annotation and data mining with the Blast2GO suite, Nucleic Acids Res. 36 (2008) 3420–3435.
[25] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q.D. Zeng, Z.H. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) (644-U130).
[26] S. Koren, M.C. Schatz, B.P. Walenz, J. Martin, J.T. Howard, G. Ganapathy, Z. Wang, D.A. Rasko, W.R. McCombie, E.D. Jarvis, A.M. Phillippy, Hybrid error correction and de novo assembly of single-molecule sequencing reads, Nat. Biotechnol. 30 (2012) (692 − +).
[27] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinforma. 12 (2011).
[28] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
[29] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, Genome Res. 18 (2008) 821–829.
[30] B.T. Hovde, S.R. Starkenburg, H.M. Hunsperger, L.D. Mercer, C.R. Deodato, R.K. Jha, O. Chertkov, R.J. Monnat, R.A. Cattolico, The mitochondrial and chloroplast genomes of the haptophyte *Chrysochromulina tobin* contain unique repeat structures and gene profiles, BMC Genomics 15 (2014).
[31] S.R. Starkenburg, K.J. Kwon, R.J. Jha, C. McKay, M. Jacobs, O. Chertkov, S. Tway, G. Rocap, R.A. Cattolico, A pangenomic analysis of the *Nannochloropsis* organellar genomes reveals novel genetic variations in key metabolic genes, BMC Genomics 15 (2014) 212.
[32] D. Lopez, D. Casero, S.J. Cokus, S.S. Merchant, M. Pellegrini, Algal functional annotation tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data, BMC Bioinforma. 12 (2011) 212.
[33] J. Kropat, A. Hong-Hermesdorf, D. Casero, P. Ent, M. Castruita, M. Pellegrini, S.S. Merchant, D. Malasarn, A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*, Plant J. 66 (2011) 770–780.
[34] T. Matthew, W.X. Zhou, J. Rupprecht, L. Lim, S.R. Thomas-Hall, A. Doebbe, O. Kruse, B. Hankamer, U.C. Marx, S.M. Smith, P.M. Schenk, The metabolome of *Chlamydomonas reinhardtii* following induction of anaerobic H-2 production by sulfur depletion, J. Biol. Chem. 284 (2009) 23415–23425.
[35] M. Siaut, S. Cuine, C. Cagnon, B. Fessler, M. Nguyen, P. Carrier, A. Beyly, F. Beisson, C. Triantaphylides, Y.H. Li-Beisson, G. Peltier, Oil accumulation in the model green alga *Chlamydomonas reinhardtii*: characterization, variability between common laboratory strains and relationship with starch reserves, BMC Biotechnol. 11 (2011).
[36] Y.T. Li, D.X. Han, G.R. Hu, D. Dauvillee, M. Sommerfeld, S. Ball, Q. Hua, *Chlamydomonas* starchless mutant defective in ADP-glucose pyrophosphorylase hyper-accumulates triacylglycerol, Metab. Eng. 12 (2010) 387–391.
[37] U. Goodenough, I. Blaby, D. Casero, S.D. Gallaher, C. Goodson, S. Johnson, J.H. Lee, S.S. Merchant, M. Pellegrini, R. Roth, J. Rusch, M. Singh, J.G. Umen, T.L. Weiss, T. Wulan, The path to triacylglyceride obesity in the sta6 strain of Chlamydomonas Reinhardtii, Eukaryot. Cell 13 (2014) 591–613.
[38] C. Goodson, R. Roth, Z.T. Wang, U. Goodenough, Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas reinhardtii* and stimulation of lipid body production with acetate boost, Eukaryot. Cell 10 (2011) 1592–1606.

[39] S.S. Merchant, J. Kropat, B.S. Liu, J. Shaw, J. Warakanont, TAG, You're it! *Chlamydomonas* as a reference organism for understanding algal triacylglycerol accumulation, Curr. Opin. Biotechnol. 23 (2012) 352–363.
[40] N.R. Boyle, M.D. Page, B.S. Liu, I.K. Blaby, D. Casero, J. Kropat, S.J. Cokus, A. Hong-Hermesdorf, J. Shaw, S.J. Karpowicz, S.D. Gallaher, S. Johnson, C. Benning, M. Pellegrini, A. Grossman, S.S. Merchant, Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*, J. Biol. Chem. 287 (2012) 15811–15825.
[41] I.K. Blaby, A.G. Glaesener, T. Mettler, S.T. Fitz-Gibbon, S.D. Gallaher, B.S. Liu, N.R. Boyle, J. Kropat, M. Stitt, S. Johnson, C. Benning, M. Pellegrini, D. Casero, S.S. Merchant, Systems-level analysis of nitrogen starvation-induced modifications of carbon metabolism in a *Chlamydomonas reinhardtii* starchless mutant, Plant Cell 25 (2013) 4305–4323.
[42] E.R. Moellering, C. Benning, Phosphate regulation of lipid biosynthesis in *Arabidopsis* is independent of the mitochondrial outer membrane DGS1 complex, Plant Physiol. 152 (2010) 1951–1959.
[43] U.K. Aryal, S.J. Callister, B.H. McMahon, L.-A. McCue, J. Brown, J. Stockel, M. Liberton, S. Mishra, X. Zhang, C.D. Nicora, T.E. Angel, D.W. Koppenaal, R.D. Smith, H.B. Pakrasi, L.A. Sherman, Proteomic profiles of five strains of oxygenic photosynthetic cyanobacteria of the genus Cyanothece, J. Proteome Res. 13 (2014) 3262–3276.
[44] J. Berendzen, W.J. Bruno, J.D. Cohn, N.W. Hengartner, C.R. Kuske, B.H. McMahon, M.A. Wolinsky, G. Xie, Rapid phylogenetic and functional classification of short genomic fragments with signature peptides, BMC Res. Notes 5 (2012) 460.
[45] A. Vieler, S.B. Brubaker, B. Vick, C. Benning, A lipid droplet protein of Nannochloropsis with functions partially analogous to plant oleosins, Plant Physiol. 158 (2012) 1562–1569.
[46] C.O. Rock, J.E. Cronan, *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis, Bba-Lipids Lipid Metab. 1302 (1996) 1–16.
[47] C.O. Rock, S. Jackowski, Forty years of bacterial fatty acid synthesis, Biochem. Biophys. Res. Commun. 292 (2002) 1155–1166.
[48] T. Konishi, K. Shinohara, K. Yamada, Y. Sasaki, Acetyl-CoA carboxylase in higher plants: most plants other than gramineae have both the prokaryotic and the eukaryotic forms of this enzyme, Plant Cell Physiol. 37 (1996) 117–122.
[49] F. Xie, T. Liu, W.J. Qian, V.A. Petyuk, R.D. Smith, Liquid chromatography–mass spectrometry-based quantitative proteomics, J. Biol. Chem. 286 (2011) 25443–25449.
[50] D. Gupta, N. Tuteja, Chaperones and foldases in endoplasmic reticulum stress signaling in plants, Plant Signal. Behav. 6 (2011) 232–236.
[51] D. Dilworth, G. Gudavicius, A. Leung, J. Nelson Christopher, The roles of peptidyl-proline isomerases in gene regulation, Biochem. Cell Biol. 90 (2012) 55–69.
[52] N. Malanovic, I. Streith, H. Wolinski, G. Rechberger, D. Kohlwein Sepp, O. Tehlivets, S-adenosyl-L-homocysteine hydrolase, key enzyme of methylation metabolism, regulates phosphatidylcholine synthesis and triacylglycerol homeostasis in yeast: implications for homocysteine as a risk factor of atherosclerosis, J. Biol. Chem. 283 (2008) 23989–23999.
[53] A. Fait, H. Fromm, D. Walter, G. Galili, A.R. Fernie, Highway or byway: the metabolic role of the GABA shunt in plants, Trends Plant Sci. 13 (2008) 14–19.
[54] A. Banerjee, R. Sharma, Y. Chisti, U.C. Banerjee, *Botryococcus braunii*: a renewable source of hydrocarbons and other chemicals, Crit. Rev. Biotechnol. 22 (2002) 245–279.
[55] A.C. Brown, B.A. Knights, E. Conway, Hydrocarbon content and its relationship to physiological state in green alga *Botryococcus braunii*, Phytochemistry 8 (1969) (543-&).
[56] P. Metzger, C. Largeau, *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids, Appl. Microbiol. Biotechnol. 66 (2005) 486–496.
[57] S.K. Jarchow-Choy, A.T. Koppisch, D.T. Fox, Synthetic routes to methylerythritol phosphate pathway intermediates and downstream isoprenoids, Curr. Org. Chem. 18 (2014) 1050–1072.
[58] M. Lohr, J. Schwender, J.E.W. Polle, Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae, Plant Sci. 185 (2012) 9–22.
[59] I. Molnar, D. Lopez, J.H. Wisecaver, T.P. Devarenne, T.L. Weiss, M. Pellegrini, J.D. Hackett, Bio-crude transcriptomics: gene discovery and metabolic network reconstruction for the biosynthesis of the terpenome of the hydrocarbon oil-producing green alga, *Botryococcus braunii* race B (Showa), BMC Genomics 13 (2012).
[60] E. Cordoba, H. Porta, A. Arroyo, C.S. Roman, L. Medina, M. Rodriguez-Concepcion, P. Leon, Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize, J. Exp. Bot. 62 (2011) 2023–2038.
[61] D. Matsushima, H. Jenke-Kodama, Y. Sato, Y. Fukunaga, K. Sumimoto, T. Kuzuyama, S. Matsunaga, S. Okada, The single cellular green microalga *Botryococcus braunii*, race B possesses three distinct 1-deoxy-D-xylulose 5-phosphate synthases, Plant Sci. 185 (2012) 309–320.
[62] B.F. Lucker, C.C. Hall, R. Zegarac, D.M. Kramer, The environmental photobioreactor (ePBR): an algal culturing platform for simulating dynamic natural environments, Algal Res. 6 (2014) 242–249.
[63] A. Kumar, V.R. Falcao, R.T. Sayre, Evaluating nuclear transgene expression systems in *Chlamydomonas reinhardtii*, Algal Res. 2 (2013) 321–332.
[64] M. Nakazawa, T. Minami, K. Teramura, S. Kumamoto, S. Hanato, S. Takenaka, M. Ueda, H. Inui, Y. Nakano, K. Miyatake, Molecular characterization of a bifunctional glyoxylate cycle enzyme, malate synthase/isocitrate lyase, in *Euglena gracilis*, Comp. Biochem. Physiol. B Biochem. Mol. Biol. 141B (2005) 445–452.
[65] K. Ono, M. Kondo, T. Osafune, K. Miyatake, H. Inui, S. Kitaoka, M. Nishimura, Y. Nakano, Presence of glyoxylate cycle enzymes in the mitochondria of *Euglena gracilis*, J. Eukaryot. Microbiol. 50 (2003) 92–96.
[66] M. Hildebrand, R.M. Abbriano, J.E.W. Polle, J.C. Traller, E.M. Trentacoste, S.R. Smith, A.K. Davis, Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production, Curr. Opin. Chem. Biol. 17 (2013) 506–514.
[67] Z. Perrine, S. Negi, R.T. Sayre, Optimization of photosynthetic light energy utilization by microalgae, Algal Res. 1 (2012) 134–142.
[68] J.W. Hickman, K.M. Kotovic, C. Miller, P. Warrener, B. Kaiser, T. Jurista, M. Budde, F. Cross, J.M. Roberts, M. Carleton, Glycogen synthesis is a required component of the nitrogen stress response in *Synechococcus elongatus* PCC 7942, Algal Res. 2 (2013) 98–106.
[69] Y. Xu, R.M. Alvey, P.O. Byrne, J.E. Graham, G. Shen, D.A. Bryant, Expression of genes in cyanobacteria: adaptation of endogenous plasmids as platforms for high-level gene expression in *Synechococcus* sp. PCC 7002, Methods Mol. Biol. 684 (2011) 273–293.
[70] C. Merryman, D.G. Gibson, Methods and applications for assembling large DNA constructs, Metab. Eng. 14 (2012) 196–204.
[71] K.S. Lovejoy, L.E. Davis, L.M. McClellan, A.M. Lillo, J.D. Welsh, E.N. Schmidt, C.K. Sanders, A.J. Lou, D.T. Fox, A.T. Koppisch, R.E. Del Sesto, Evaluation of ionic liquids on phototrophic microbes and their use in biofuel extraction and isolation, J. Appl. Phycol. 25 (2013) 973–981.
[72] S. Mohr, C. Bakal, N. Perrimon, Genomic screening with RNAi: results and challenges, Annu. Rev. Biochem. 79 (2010) 37–64.
[73] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297.
[74] D.P. Bartel, C.Z. Chen, Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs, Nat. Rev. Genet. 5 (2004) 396–400.
[75] M.W. Jones-Rhoades, D.P. Bartel, B. Bartel, MicroRNAs and their regulatory roles in plants, Annu. Rev. Plant Biol. 57 (2006) 19–53.
[76] E.H. Harris, The *Chlamydomonas* Sourcebook: Second Edition Introduction to *Chlamydomonas* and Its Laboratory Use, Academic Press, Burlington, MA, 2009.
[77] R. Sager, S. Granick, Nutritional control of sexuality in *Chlamydomonas reinhardi*, J. Gen. Physiol. 37 (1954) 729–742.
[78] R. Radakovits, R.E. Jinkerson, S.I. Fuerstenberg, H. Tae, R.E. Settlage, J.L. Boore, M.C. Posewitz, Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*, Nat. Commun. 3 (2012).
[79] N. Fischer, J.D. Rochaix, The flanking regions of PsaD drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*, Mol. Gen. Genomics. 265 (2001) 888–894.
[80] L.A. Zaslavskaia, J.C. Lippmeier, P.G. Kroth, A.R. Grossman, K.E. Apt, Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes, J. Phycol. 36 (2000) 379–386.
[81] B.A. Rasala, P.A. Lee, Z.X. Shen, S.P. Briggs, M. Mendez, S.P. Mayfield, Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide, PLoS One 7 (2012).
[82] M. Sun, K.X. Qian, N. Su, H.Y. Chang, J.X. Liu, G.F. Chen, Foot-and-mouth disease virus VP1 protein fused with cholera toxin B subunit expressed in *Chlamydomonas reinhardtii* chloroplast, Biotechnol. Lett. 25 (2003) 1087–1092.
[83] H. Jenssen, P. Hamill, R.E.W. Hancock, Peptide antimicrobial agents, Clin. Microbiol. Rev. 19 (2006) (491 − +).
[84] G.S. Wang, X. Li, Z. Wang, APD2: the updated antimicrobial peptide database and its application in peptide design, Nucleic Acids Res. 37 (2009) D933–D937.
[85] J.E. Heuser, The origins and evolution of freeze-etch electron microscopy, J. Electron Microsc. 60 (2011) S3–S29.
[86] Z.T. Wang, N. Ullrich, S. Joo, S. Waffenschmidt, U. Goodenough, Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*, Eukaryot. Cell 8 (2009) 1856–1868.
[87] N. Bigelow, J. Barker, S. Ryken, J. Patterson, W. Hardin, S. Barlow, C. Deodato, R.A. Cattolico, *Chrysochromulina* sp.: a proposed lipid standard for the algal biofuel industry and its application to diverse taxa for screening lipid content, Algal Res. 2 (2013) 385–393.
[88] S. Negi, A. Barry, N. Friedland, N. Sudasinghe, S. Subramanian, S. Pieris, F.O. Holguin, B. Dungan, T. Schaub, R. Sayre, Impact of nitrogen limitation on biomass, photosynthesis, and lipid accumulation in *Chlorella sorokiniana*, J. Appl. Phycol. (2015) 1–10.