

ARTICLE

Received 27 Oct 2011 | Accepted 17 Jan 2012 | Published 21 Feb 2012

DOI: 10.1038/ncomms1688

Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*

Randor Radakovits^{1,*}, Robert E. Jinkerson^{1,*}, Susan I. Fuerstenberg², Hongseok Tae³, Robert E. Settlage³, Jeffrey L. Boore^{2,4} & Matthew C. Posewitz¹

The potential use of algae in biofuels applications is receiving significant attention. However, none of the current algal model species are competitive production strains. Here we present a draft genome sequence and a genetic transformation method for the marine microalga *Nannochloropsis gaditana* CCMP526. We show that *N. gaditana* has highly favourable lipid yields, and is a promising production organism. The genome assembly includes nuclear (~29 Mb) and organellar genomes, and contains 9,052 gene models. We define the genes required for glycerolipid biogenesis and detail the differential regulation of genes during nitrogen-limited lipid biosynthesis. Phylogenomic analysis identifies genetic attributes of this organism, including unique stramenopile photosynthesis genes and gene expansions that may explain the distinguishing photoautotrophic phenotypes observed. The availability of a genome sequence and transformation methods will facilitate investigations into *N. gaditana* lipid biosynthesis and permit genetic engineering strategies to further improve this naturally productive alga.

¹ Department of Chemistry and Geochemistry, Colorado School of Mines, Golden, Colorado 80401, USA. ² Genome Project Solutions, 1024 Promenade Street, Hercules, California 94547, USA. ³ Data Analysis Core, Virginia Bioinformatics Institute, Virginia Tech, 1 Washington Street, Blacksburg, Virginia 24060, USA. ⁴ Department of Integrative Biology, University of California, Berkeley, California 94720, USA. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.C.P. (email: mposewit@mines.edu).

In recent years, a detailed understanding of the many biosynthetic pathways that can be used for the production of biofuel feedstocks or higher value bioproducts has emerged, and novel pathways for the production of specific bioenergy carriers are continuously being discovered in a variety of organisms^{1–4}. These advances, in combination with the development of reliable genetic transformation protocols for photosynthetic organisms with high innate biomass accumulation rates will enable the engineering of improved strains that not only have high production rates but also produce tailored precursors, or even finished products of biotechnological interest^{2,5–7}.

Several species of *Nannochloropsis* have attracted sustained interest from algal biofuels researchers owing to their high photoautotrophic biomass accumulation rates, high lipid content^{8–12}, and their successful cultivation at large scale using natural sunlight in either open ponds or enclosed systems by companies such as Solix Biofuels, Aurora Algae, Seambiotic, Hairong Electric Company/Seambiotic and Proviron. Further improvements in strain productivity have been hampered by the lack of a genetically tractable model system for these highly productive oleaginous algae. Currently, the most developed algal model species are the green alga *Chlamydomonas reinhardtii* and the diatom *Phaeodactylum tricorutum*, both of which have genome sequences and established transformation methods^{13–18}. Genetic engineering approaches have been successfully used to improve biofuel phenotypes in both of these organisms^{19–22}; unfortunately, neither of these algae is a natively exceptional producer of biomass or lipids, and, as such, extensive genetic modifications will be needed before their use in biofuel applications.

An alternative alga that has inherently desirable biomass production characteristics and has been successfully cultivated outdoors at commercial scale is *Nannochloropsis gaditana*, a stramenopile alga in the Eustigmatophyceae, which is oleaginous and stores relatively large amounts of lipid, in the form of triacylglycerides (TAG), even during logarithmic growth. Various strains of *Nannochloropsis* have been investigated for their biomass and lipid production characteristics, and several isolates have been grown for aquaculture purposes. *N. salina*, *N. oculata* and *N. gaditana* have received the most attention because of their exceptional lipid production characteristics^{9,23,24}. *N. gaditana* has high photoautotrophic biomass and lipid production rates and can grow to high densities ($>10\text{ g l}^{-1}$) while tolerating a wide range of conditions with regards to pH, temperature and salinity. *N. gaditana* is therefore a good candidate for development into a model organism for algal biofuel production, and the availability of a genome sequence and reliable transformation protocols are required advances in this direction. In addition, there are reports that homologous recombination is tractable in the eustigmatophyte *Nannochloropsis* W2]3B^{25,26}.

In this study, to transform a natively robust and oleaginous alga into a model system for biofuel production, we sequenced the genome and developed a genetic transformation method for *N. gaditana* CCMP526. We also investigated the *N. gaditana* lipid metabolic pathways on the genome and transcriptome levels, quantifying gene transcript levels during a relatively low lipid production stage, (logarithmic growth), and a high lipid production stage, (stationary phase) after nitrate depletion. Finally, we conducted comparative phylogenomic analyses among other algal lineages to determine genes unique to *N. gaditana* and to identify sets of conserved proteins across photosynthetic stramenopiles. The genome sequence, its analysis, and the development of genetic transformation in *N. gaditana* are important first steps in improving this industrially proven, oleaginous algal for biofuel production.

Results

High lipid yields from high-density cultures of *N. gaditana*.

N. gaditana is a robust producer of both biomass and lipids under

a wide array of culture conditions, including minimal f/2 seawater medium and artificial seawater (10–120‰ seawater salinity, pH 7–10) supplemented with nitrate, phosphate and CO₂. The yields from *N. gaditana* cultures grown in f/2 medium at 50‰ seawater salinity are shown in Figure 1a,b. Yields of $0.65\text{ g l}^{-1}\text{ d}^{-1}$ biomass and $0.31\text{ g l}^{-1}\text{ d}^{-1}$ total lipids were achieved over a period of 3 months in 11 Roux Flasks sparged with air/2% CO₂, when half the cultures were exchanged for fresh medium every week. Lipid body accumulation can be triggered/enhanced in most algae by nitrogen deprivation or other stress conditions²⁷, and the high lipid content (47.5%) in actively growing cultures of *N. gaditana* is likely facilitated by the rapid depletion of nitrate in dense cultures ($3\text{--}8\text{ g l}^{-1}$) during growth. Optimal lipid yields were obtained with a starting culture density of $\sim 3.6\text{ g l}^{-1}$. It is likely that lack of light penetration due to self-shading is the main limiting factor for cultures at higher starting densities. Low-density cultures ($<0.5\text{ g l}^{-1}$) can be growth inhibited by high light ($>200\text{ }\mu\text{E}$) but the higher density cultures have good production between $1,000\text{ }\mu\text{E}$ and $2,000\text{ }\mu\text{E}$. For medium to high-density cultures ($3\text{--}10\text{ g l}^{-1}$), no substantial increase in productivity is observed on increasing the light from $1,000\text{ }\mu\text{E}$ to $2,000\text{ }\mu\text{E}$, supporting the hypothesis that self shading becomes the limiting factor at these densities. The laboratory productivity numbers have been extrapolated to calculate potential lipid yields in comparison with other algae (Fig. 1c) and to other biofuel production platforms (Fig. 1d). In Figure 1d, the green bars indicate our extrapolations based on data from Chisti *et al.*²⁸ and Chen *et al.*²⁹, whereas grey bars indicate estimations originally provided by Atsumi *et al.*³⁰ It is important to note that some of the values represent actual production yields from large-scale cultivation (Soy, Palm)^{28,31}, whereas other values are extrapolated from small scale cultures with 24 h light (*Synechococcus elongatus* Isobutyraldehyde and Isobutanol). The *N. gaditana* lipid production yields have been derived from small scale cultures with 12 h light/12 h dark cycles and therefore provide a more realistic estimation relative to *S. elongatus*. Robust lipid yields from *Nannochloropsis* scale from 25 ml cultures to 8 l cultures under laboratory conditions, to 10-hectare outdoor ponds where it is grown on a commercial scale (Hairong Electric Company and Seambiotic). The high lipid content of *N. gaditana* cells is apparent on fluorescent labelling of algal triglycerides with the lipophilic dye, BODIPY. Actively growing cells have a constitutive lipid droplet that expands within cells in stationary phase or during nitrogen deprivation (Supplementary Fig. S1a,b respectively). The large majority of lipids in *N. gaditana* are composed of palmitic and palmitoleic acid with a minor content of myristic and oleic acid (Supplementary Fig. S1c), resulting in a relatively simple fatty acid profile, and these fatty acids can be used for the production of biodiesel or biopetrol.

Sequencing and assembly. DNA sequencing reads obtained, using both Roche and Illumina (including both unpaired and LIPES protocols) technologies, were trimmed for quality, and then assembled separately. These assemblies were merged, followed by removal of scaffolds of bacterial contaminant(s), producing a genome assembly of 2,087 scaffolds, with an N50 of 257 and an L50 of 37,693 nucleotides (nts) (Table 1). There are 35 scaffolds longer than 100 kb, a total of 561 longer than 20 kb, and a total of 1,447 that are longer than 2 kb. Table 1 also includes statistics on the contigs before assembly into the final scaffolds.

In addition to the nuclear genome, the plastid and mitochondrial genomes were also sequenced, assembled and annotated (Supplementary Figs S2 and S3). Relative to the organellar genomes of *P. tricorutum* and *Thalassiosira pseudonana*, significant conservation of gene content and gene organization was observed, with some notable exceptions³². See Supplementary Note 1 and Supplementary Table S1 for a detailed description.

RNA was isolated from a variety of culturing conditions and growth phases, converted into complementary DNA (cDNA), then

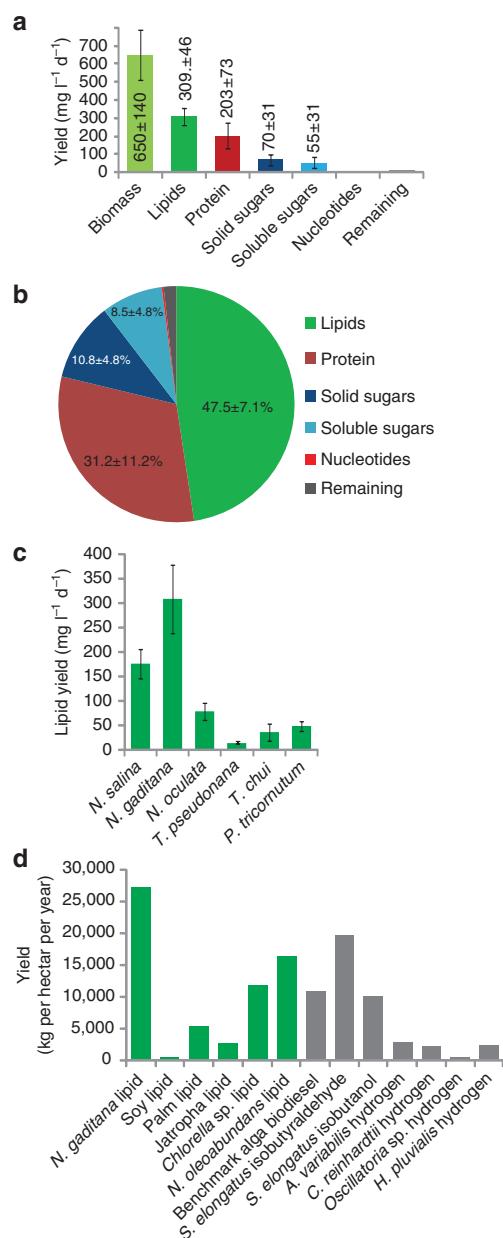


Figure 1 | Biomass production by *N. gaditana*. (a) *N. gaditana* production of biomass, lipids, protein and sugars quantified during continuous growth over a period of 3 months in 50% salinity seawater medium supplemented with nitrate, phosphate and CO₂ with continuous 1,000 μE light. Every week, half of the culture was collected and replaced with fresh medium. Inset values show the yield in mg l⁻¹ per day. Values are from 12 measurements and error bars show the standard deviation. (b) Chart illustrating collected biomass compositions, the majority of which consists of lipids even under nutrient replete conditions. Inset values show percentage of total biomass. (c) Comparison of *N. gaditana* lipid production rates with other algae examined in this work. Values are from at least three separate experiments and error bars show the standard deviation. (d) Comparison of *N. gaditana* large-scale production rates with other biofuel production platforms. Bars in green indicate our estimations; bars in grey indicate estimations by Atsumi *et al.*³⁰ The values for *N. gaditana* have been extrapolated from 11 cultures and adjusted for our observed productivity in 12 h/12 h light/dark cycles. The *S. elongatus* production values are for 24 h light and would presumably be lower in 12 h/12 h light/dark cycles.

sequenced using the Illumina SIPES protocol, followed by assembly of these reads using the commercial package from CLC Bio (Katrinebjerg) into 37,055 contigs.

Genome annotation. A variety of methods were used, including *ab initio* predictions, homology detection, and RNAseq matching to the genome assembly, and then these were reconciled into a single gene set using Maker³³. Contigs from the transcript assembly that had strong homology support, but were otherwise not part of the Maker gene set, were added to form gene set version 1.1 with 9,052 members (Table 1 and Supplementary Table S2).

We identified several functional gene clusters, including a uniquely arranged cluster of four genes involved in hydrogenase function (*HYDA1*, *HYDE*, *HYDF* and *HYDG*) and a cluster of three genes involved in nitrogen assimilation (nitrate reductase, nitrite reductase, and a nitrate transporter) (Supplementary Fig. S4). Similar clusters of nitrogen assimilation genes can be observed in prasinophytes and *C. reinhardtii*^{34,35}. However, despite the presence of other functional clusters, no nitrogen assimilation gene cluster can be observed in the more closely related *Ectocarpus siliculosus*³⁶. An expanded analysis of functional gene clusters can be found in Supplementary Note 2 and Supplementary Table S3.

Comparative genomics. *N. gaditana* is a eustigmatophyte alga that is closely related to the Phaeophyceae (brown algae), with the most closely related organism having a fully sequenced genome being the multicellular brown alga, *E. siliculosus* (Fig. 2a)³⁶. Among other species of *Nannochloropsis*, *N. gaditana* is most closely related to *N. salina* (Fig. 2b). To identify novel features of the *N. gaditana* genome, we determined which *N. gaditana* genes have homologues found in brown algae³⁶ and the pelagophyte *Aureococcus anophagefferens*³⁷, green algae (*Chlorella variabilis* NC64A³⁸ and *C. reinhardtii*¹³), red algae (*Cyanidioschyzon merolae*³⁹), and diatoms (*T. pseudonana*⁴⁰ and *P. tricornutum*¹⁴). This analysis confirms the close evolutionary proximity between the Eustigmatophyceae and Phaeophyceae (Fig. 2c), and provides us with 2,733 genes that

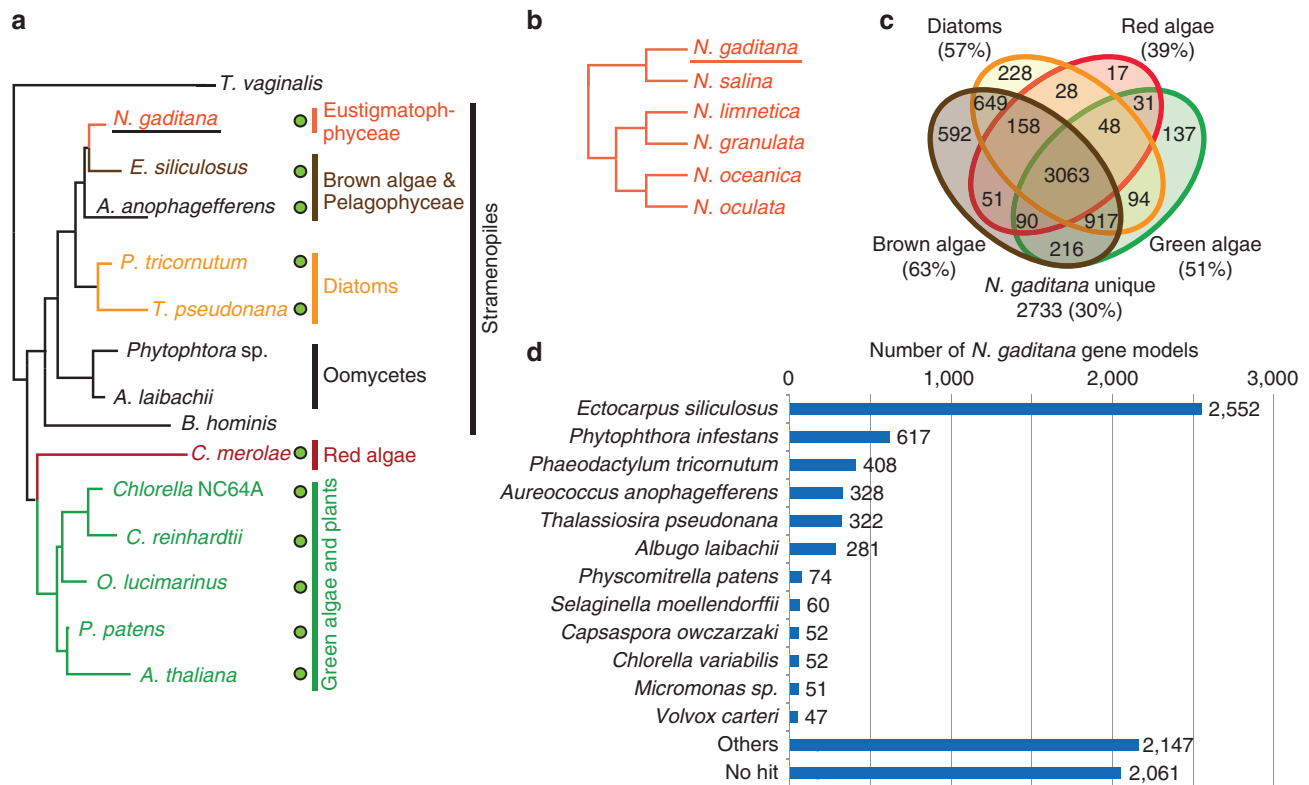


Figure 2 | Phylogenetic analysis of the *N. gaditana* genome. (a) Schematic phylogenetic tree of stramenopiles and photosynthetic algae. The tree is adapted from Eisenreich *et al.*⁴³ and Tyler *et al.*⁶⁰ Filled green circles on the right indicate photosynthetic species. (b) The tree indicates the relationship between different strains of *Nannochloropsis* based on 18S ribosomal RNA gene sequences. (c) Venn diagram representation of shared/unique genes of *N. gaditana* in comparison with brown algae, diatoms, red algae and green algae. (d) *N. gaditana* gene models were compared with all previously sequenced genomes in the non-redundant protein database using BLASTp. The number of times an organism was the top BLASTp hits (E-value less than 1E-3) of a *N. gaditana* gene model is indicated.

are exclusive to *N. gaditana*, not found in the other algal genomes queried. This corresponds to 30.2% of the total gene repertoire in *N. gaditana*, which is similar to the fraction of unique genes found in *T. pseudonana*⁴⁰, *E. siliculosus*³⁶ and *P. tricornutum*¹⁴. Comparison of *N. gaditana* gene models to the non-redundant protein database (BLASTp) yielded top hits from a variety of organisms, the most frequent being stramenopiles (Fig. 2d), which was expected on the basis of the phylogeny of *N. gaditana*.

Previous attempts have been made at establishing the minimal essential set of genes needed for photosynthesis, the 'GreenCut' of photosynthetic genes, which is a set of 597 orthologues that are conserved in plant and green algal lineages, but not in non-photosynthetic organisms^{13,41}. We decided to take advantage of the fact that there are both photosynthetic and non-photosynthetic stramenopiles to generate an analogous set of genes conserved in photosynthetic stramenopiles. To establish this 'Stramenopile-PhotoCut' of photosynthetic genes, orthologues common to *N. gaditana* and four photosynthetic stramenopiles (*E. siliculosus*, *A. anophagefferens*, *T. pseudonana* and *P. tricornutum*), but not present in non-photosynthetic stramenopiles (*Phytophthora sojae*, *Phytophthora ramorum*, *Phytophthora infestans*, *Albugo laibachii* or *Blastocystis hominis*), were selected, resulting in a list of 363 genes. (Fig. 3a and Supplementary Data 1). The majority of these genes have orthologues in the green and red algal lineages and 115 are found in the 'GreenCut'⁴¹. However, 39 genes with homologues only found in photosynthetic stramenopiles are present in the genome (Supplementary Data 1). Similar to many genes found in the 'GreenCut', some of the 39 stramenopile-specific 'StramenopilePhotoCut' genes are of completely unknown function,

but several of the genes have known domains, including several peptidases/proteases, DNA-binding proteins/transcription factors, and transport proteins, as well as genes that are thought to directly interact with the photosystems (Fig. 3b). Because of the high photoautotrophic growth rates exhibited by *N. gaditana*, we also characterized the complete pathways for synthesis of chlorophyll and accessory pigments (Supplementary Table S4). All expected genes could be identified except for those encoding the mevalonate (MVA) pathway for isopentenyl-pyrophosphate biosynthesis (see the analysis of bioenergy metabolic pathways).

Bioenergy metabolic pathways. To investigate metabolic pathways of interest for biofuel production, functional annotations were assigned to *N. gaditana* gene models. Gene ontology terms (GO-terms) were assigned to 3,838 gene models, from which 2,766 genes were identified as performing enzyme-catalysed reactions representing 700 unique EC numbers that were in turn used to populate metabolic pathway maps (Fig. 4). Some of the most frequent GO-terms, aside from housekeeping functions, are terms involved in auxin biosynthesis, photosynthesis, and lipid biosynthesis (Supplementary Fig. S5). Because of the exemplary lipid production by *N. gaditana* cultures, we focused on characterizing lipid metabolic pathway genes, including those involved in fatty acid biosynthesis, TAG assembly and lipid activation/degradation (Supplementary Table S5). BLASTp was used to identify homologues of the *N. gaditana* lipid metabolic genes in red/green/brown algae and diatoms. Comparison of the number of genes in each step of the lipid metabolic pathways suggests that *N. gaditana* has an expanded repertoire of genes involved in both TAG assembly and

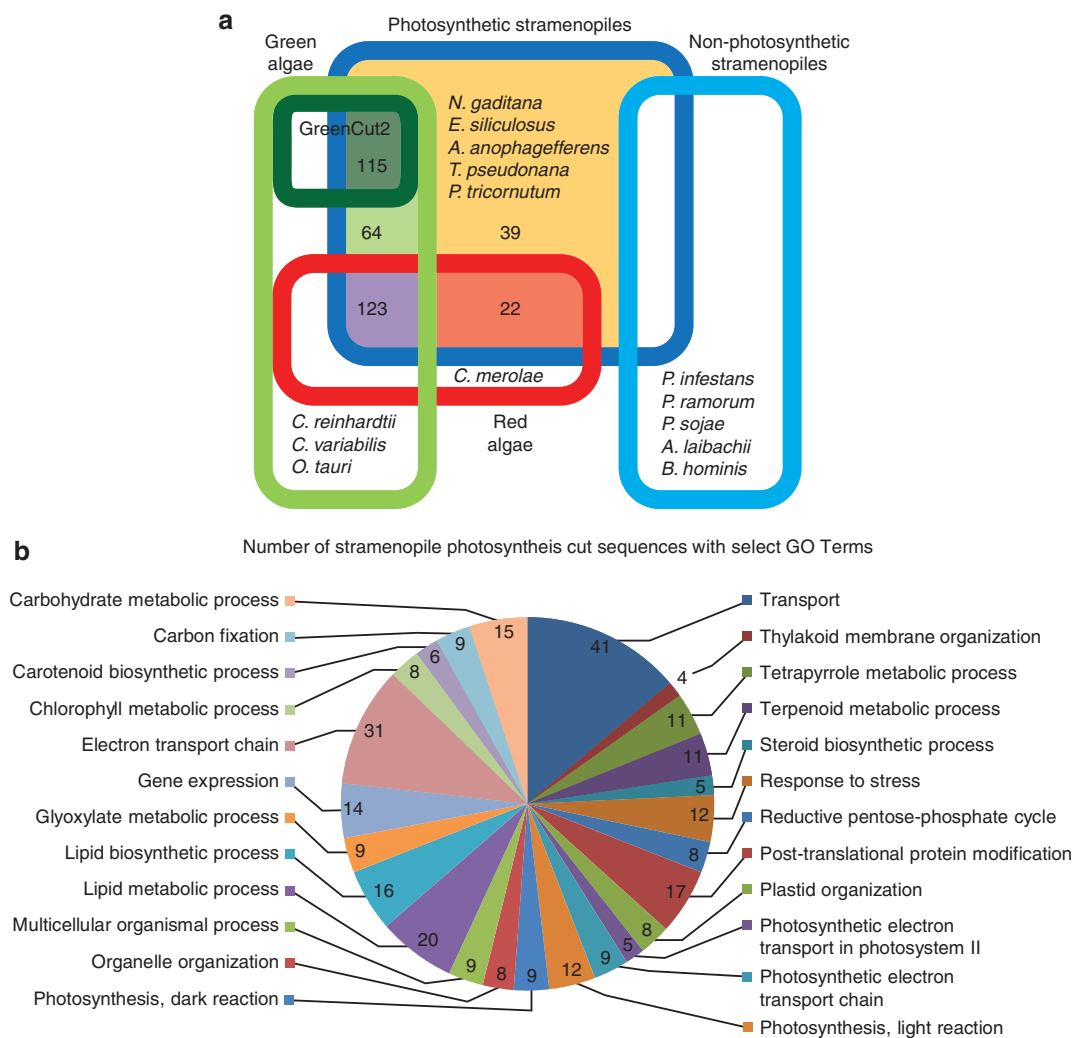


Figure 3 | Genes conserved in photosynthetic stramenopiles. (a) Euler diagram showing the 363 genes that make up the StramenopilePhotoCut genes common to photosynthetic and absent in non-photosynthetic stramenopiles. Number of genes found in each sector is indicated. The centre yellow sector indicates genes unique for the photosynthetic Stramenopiles (not found in green or red lineages). (b) Chart showing the number of 'StramenopilePhotoCut' genes with select GO terms. 'StramenopilePhotoCut' genes with no GO terms are not indicated.

lipid degradation, including glycerol 3-phosphate dehydrogenase, glycerol 3-phosphate acyltransferase, diacylglycerol acyltransferase, long-chain acyl-CoA ligase and acyl-CoA oxidase (Fig. 5 and Supplementary Table S6). This increased number of lipid metabolic pathway genes is probably significant considering that *N. gaditana* has fewer total genes than all other algae used for this comparison, with the exception of *C. merolae*. To further examine the expansion of gene families in *N. gaditana*, we compared the prevalence of GO-terms with *P. tricornutum* and *C. reinhardtii* using the Fisher's exact test. A selected list of over- and under-represented terms is shown in Supplementary Table S7. This analysis confirms the overrepresentation of the GO-term for acyl-carrier protein biosynthetic processes and also indicates the expansion of several other gene families that may be of importance for the biomass production phenotype of *N. gaditana*. These include genes involved in auxin biosynthetic processes, carbon utilization, response to stress (including chemical, temperature and salt), and pyruvate metabolic processes. See Supplementary Note 3 for a more detailed analysis of these gene expansions/reductions.

To assist in the identification of genes and to improve metabolic pathway maps of *N. gaditana*, we sequenced the transcriptome (RNAseq) under a variety of physiological conditions. Additionally,

transcriptome sequencing was conducted during logarithmic growth (low lipid production) and during stationary phase due to nitrate deprivation (high lipid production) to discover how transcriptional changes in *N. gaditana* modulate increased metabolic flux into lipid biosynthesis during nutrient deprivation. Genes that are most strongly regulated during these different conditions are shown in Supplementary Data 2. Similar to the findings in *C. reinhardtii*⁴², many of the genes that are most strongly upregulated during nitrogen deprivation are genes involved in nitrogen assimilation and protein degradation/recycling, whereas many of the most downregulated genes are involved in photosynthesis. In addition, we annotated the most highly regulated pathways on the metabolic pathway map (Fig. 4). This map highlights the decreased expression of genes involved in photosynthesis, carbon fixation, and oxidative phosphorylation that would be expected during stationary phase owing to nutrient deprivation. Surprisingly, few genes that are directly involved in lipid biosynthesis are transcriptionally upregulated to a significant extent. Because *N. gaditana* constitutively produces TAG even during logarithmic growth, a possible explanation for this low amount of differential transcript accumulation is that the lipid production machinery may already be abundant within the cell, and existing levels can manage increased metabolic flux. In support of this hypothesis, we

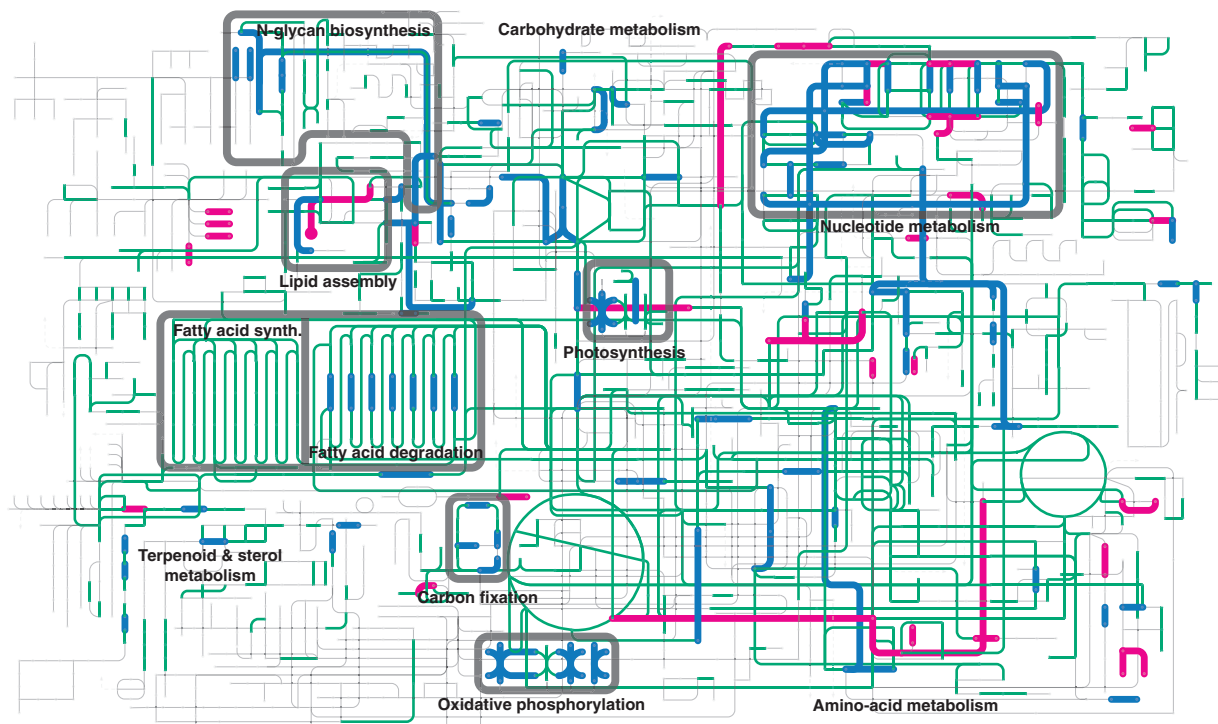


Figure 4 | *N. gaditana* metabolic pathway map. Light grey background traces indicate KEGG pathways not encoded by the *N. gaditana* genome. KEGG pathways in green, magenta or blue are present in the *N. gaditana* genome. Genes that are up- or down-regulated during nitrogen deprivation are labelled in magenta and blue, respectively.

found that genes assigned with the GO-term for post-transcriptional regulation of gene expression were overrepresented in *N. gaditana* in comparison with *P. tricornutum* and *C. reinhardtii*, whereas the GO-term for transcription factors was underrepresented (Supplementary Table S7). Interestingly, genes involved in gluconeogenesis (fructose-1,6-bisphosphatase, fructose-1,6-bisphosphate aldolase and phosphoglycerate kinase) are downregulated, which could help direct carbon flux away from carbohydrate biosynthesis into lipid biosynthesis. To determine the exact mechanisms of lipid accumulation during nutrient deprivation, further transcriptomic, proteomic and metabolomic investigations are needed.

Other pathways that are of interest for bioenergy applications are the two isoprenoid biosynthesis pathways, the mevalonate and the non-mevalonate pathways (DXP). Ancestral eukaryotes generally have only the MVA pathway although many photosynthetic organisms have acquired the DXP pathway, most probably through a cyanobacterial endosymbiont or secondarily through a red algal symbiont⁴³. Interestingly, most higher plants retained both the MVA and DXP pathways, whereas the green and red algae (for example, *C. reinhardtii*, *Ostreococcus lucimarinus*, *C. merolae*) have kept only the more recently acquired DXP pathway and eliminated the more ancestral MVA pathway. Similarly, *N. gaditana* and *A. anophagefferens* have only the DXP pathway (Supplementary Fig. S6 and Supplementary Table S4), whereas other stramenopiles in the case of diatoms and brown algae (*P. tricornutum*, *T. pseudonana*, *Ectocarpus siliculosus*) have kept both the MVA and DXP pathways. Parasitic chromalveolates, including stramenopiles, seem to differ in their isoprenoid biosynthesis capacity depending on whether they have kept at least a remnant plastid. Both *P. marinus* (has a functional plastid) and *P. falciparum* (has a remnant plastid) have kept the DXP pathway, whereas *P. sojiae*, *P. ramorum* and *A. laibachii* (no plastid) have lost both the MVA and DXP pathways.

Carbon-concentrating mechanisms. Simple mechanisms for carbon concentration and carbon assimilation have been described in

many algae⁴⁴. These mechanisms typically rely on carbonic anhydrases, which catalyse the reversible conversion of CO₂ to bicarbonate (HCO₃⁻). The physiological function of a carbonic anhydrase is dictated by its compartmentalization. Bicarbonate cannot passively cross membranes and has to either be transported into the cell by a bicarbonate transporter or be converted to freely diffusible CO₂ by extracellular carbonic anhydrases. Cytosolic carbonic anhydrases can produce HCO₃⁻ either for transport by a bicarbonate transporter into the chloroplast or for use in C₄-like carbon-concentrating mechanisms. Chloroplastic carbonic anhydrases produces CO₂ in the vicinity of Rubisco from actively transported HCO₃⁻.

We were able to identify all components necessary for inorganic carbon assimilation, including putatively targeted extracellular, mitochondrial, chloroplast, and cytosolic carbonic anhydrases and bicarbonate transporters localized in the plasma membrane and in the chloroplast (Supplementary Table S8). The total number of carbonic anhydrases found in *N. gaditana* is fewer than that found in *P. tricornutum* or *E. siliculosus*. The use of a carbonic anhydrase-type carbon-concentrating mechanism is in part supported by previous studies by Huertas and Lubian that suggested the presence of an active uptake of bicarbonate and at least intracellular carbonic anhydrases^{45–47}.

C₄-like carbon-concentrating mechanisms have been suggested for other photosynthetic stramenopiles, including *E. siliculosus*, *Thalassiosira weissflogii* and *P. tricornutum*^{36,48,49}. Typical C₄ metabolism relies on spatial separation between the sequestration of CO₂ into C₄ acids and the release of CO₂ from C₄ acids to Rubisco, either by differentiated cell types or the presence of specialized organelles. Single-cell C₄-like mechanisms have recently been described in land plants, such as *Bienertia sinuspersici*, *Bienertia cycloptera* and *Borsowia aralocaspiensis*^{50,51}, but an actual mechanism for a C₄-like metabolism in single-cell alga has not been fully characterized.

We were able to identify the genes needed for both C₃- and C₄-type carbon assimilation (Supplementary Table S9). TargetP⁵² and HECTAR⁵³ were used to establish the targeting of these proteins, and a possible model for carbon-concentrating mechanisms is shown

Table 2 | The efficiency of transformation is affected by voltage field strength and the strength of the native promoter used for expression.

Promoter construct*	9,000 V cm ⁻¹ †	10,500 V cm ⁻¹ †	12,000 V cm ⁻¹ †	Native gene expression +N‡	Native gene expression -N‡
No plasmid§	0	0	0	N/A	N/A
TUB	0	166.7×10 ⁻⁶	12.5×10 ⁻⁶	6,992	7,149
UEP	0	62.5×10 ⁻⁶	27.8×10 ⁻⁶	1,491	1,341
HSP	0	166.7×10 ⁻⁶	166.7×10 ⁻⁶	5	2
pPha-T1-fcpB	0	0	0	N/A	N/A

* The plasmid used for transformation.

† Colonies generated per electroporated cell using different field strengths and promoter constructs.

‡ RNAseq quantification of the expression of the native β -tubulin (TUB), ubiquitin extension protein (UEP) and heat shock protein 70 (HSP). Measured in number of reads per kb of the corresponding genes during normal and nitrogen deprived growth.

§ Negative control went through entire electroporation protocol without any plasmid DNA. Survival appeared unaffected on positive control plates without zeocin.

|| pPha-T1-fcpB indicates use of the *P. tricornutum* fcpB promoter.

promoters from the genes encoding β -tubulin (TUB, Nga00092), heat shock protein 70 (HSP, Nga07210) and the ubiquitin extension protein (UEP, Nga02115.1). The efficiency of the transformations was strongly affected by the promoter used (Table 2) and the most efficient transformation was achieved using the TUB promoter that resulted in an efficiency of 12.5×10^{-6} . This was achieved using a very high $12,000 \text{ V cm}^{-1}$ field strength during the electroporation. Use of lower field strength ($10,500 \text{ V cm}^{-1}$) resulted in fivefold lower transformation efficiency (60×10^{-6}). Interestingly, the transformation efficiency using the chosen promoters approximately corresponds with the RNAseq quantification values for these genes (Table 2). We also attempted using the fucoxanthin-binding protein B (FcpB) promoter from *P. tricornutum* without success. The highest efficiency achieved, (12.5×10^{-6}) is comparable to the efficiency (10×10^{-6}) observed with transformations of *P. tricornutum*⁵⁴. While this manuscript was in review, a report of stable transformation in *Nannochloropsis* W2J3B using similar electroporation parameters was published by Kilian *et al.*²⁶

Confirmation of successful *N. gaditana* transformation was done after 4–5 months of growth with antibiotic selection. Genomic PCR confirmed the presence of the transgene in selected colonies, and Southern blot analysis confirmed successful incorporation of the transgene into the nuclear genomes of the mutant colonies (Supplementary Fig. S7). The Southern blots also indicated that multiple insertions of the transgene occurred in some cases, and that integration into the genome with the construct used is random. Our results demonstrate a straightforward approach to the genetic modification of this oleaginous alga, and we anticipate that the ability to further engineer *N. gaditana* will allow this organism to emerge as an important model species for algal biofuel production.

Discussion

Here we present the annotated draft genome and a method for genetic transformation of a biofuel relevant alga, the eustigmatophyte *N. gaditana*. Photosynthetic algae have long been considered a possible renewable feedstock for biofuel production and have recently experienced intense interest owing to diminishing petroleum reserves and increasing atmospheric levels of CO₂. One of the main challenges has been the lack of a genetically tractable model alga capable of industrial biofuels production. The availability of such an alga could eventually permit the sort of comprehensive systems-biology approaches that have been applied towards the development of highly productive strains of industrial bacteria. The characterization of the genome of *N. gaditana* and the identification of the genes and pathways that are involved in lipid production in this alga, in combination with the establishment of a method for genetic transformation, allows for further analysis of potential bottle necks in the TAG biosynthesis pathway and the discovery of suitable targets for gene overexpression and/or knockout. Several of the genes

involved in TAG assembly, including PAP (Nga21116) and PDAT (Nga02737), have only one homologue and represent obvious targets for either overexpression for potential increased TAG production, or knockout for studying the physiological effects of attenuated TAG assembly. Several other identified gene families, such as TAG lipases (Nga30958, Nga30749) and acyl-CoA oxidases (Nga03053, Nga04370.1, Nga30819), as well as genes involved in gluconeogenesis are also interesting targets for gene knockout or knockdown for the purpose of increasing lipid production. In addition, both the carbonic anhydrases (Nga01240, Nga01717, Nga03728, Nga30848, Nga10007, Nga21222) and the putative bicarbonate transporters (Nga00165.01, Nga06584) are excellent targets for gene overexpression for improved carbon assimilation. The continued development of *N. gaditana* into a model for oleaginous algal biofuel production is a step towards the cost competitive photoautotrophic production of biofuels.

Methods

Growth of *N. gaditana*. *N. gaditana* CCMP526 (National Center for Marine Algae and Microbiota) was cultivated in a defined artificial seawater medium (ASW). The ASW was prepared as follows: 15 g l^{-1} NaCl, 6.6 g l^{-1} MgSO₄·7H₂O, 5.6 g l^{-1} MgCl₂·6H₂O, 0.5 g l^{-1} CaCl₂·2H₂O, 1.45 g l^{-1} KNO₃, 0.12 g l^{-1} KH₂PO₄, 0.04 g l^{-1} NaHCO₃, 0.01 g l^{-1} FeCl₃·6H₂O, 0.035 g l^{-1} Na₂-EDTA, 0.25 ml l^{-1} 3.64 mM MnCl₂·4H₂O, and 0.5 ml l^{-1} trace metal mix (20 mg l^{-1} CoCl₂·6H₂O, 12 mg l^{-1} Na₂MoO₄·2H₂O, 44 mg l^{-1} ZnSO₄·7H₂O, 20 mg l^{-1} CuSO₄·5H₂O, 7.8 g l^{-1} Na₂-EDTA). The pH of the trace metal mix was adjusted to 7.5 and the final pH of the ASW was adjusted to 7.3. Low-density starting cultures were grown in low light ($50 \mu\text{mol m}^{-2} \text{ s}^{-1}$) without CO₂ supplementation. The light intensity was gradually increased and maximum biomass production was achieved from medium-density starting cultures ($> 3 \text{ g l}^{-1}$) bubbled with 2% CO₂/air at high light ($200\text{--}2,000 \mu\text{mol m}^{-2} \text{ s}^{-1}$).

Estimation of biomass and lipid production yields. Dry biomass yields were determined via filtration of algal cultures. The biomass from 5–10 ml of culture was collected by vacuum filtration using $0.7 \mu\text{m}$ glass fibre filters (Pall Corporation), which were then washed twice with 20 ml of diH₂O before being dried overnight at 80 °C and then weighed.

Total lipids were extracted and derivatized from liquid cultures as described previously^{19,20}. Briefly, 1.0 ml of 1 M NaOH in 95% methanol was added to 0.5 ml of algal culture and then heated in tightly sealed vials at 100 °C for 2 h, which resulted in cell lysis and lipid saponification. Acid-catalysed methylation was accomplished by adding 1.5 ml 12 N HCl:MeOH, 1:16 (v/v) and incubating at 80 °C for 5 h. Fatty acid methyl esters (FAMES) were extracted into 1.25 ml hexane through gentle inversion for 20 min. Extracts were washed with distilled water and analysed directly by GC-FID using an Agilent 7890A gas chromatograph with a DB-5ms column. FAMES were quantified against a standard 37-component FAME mix (Sigma-Aldrich). Tridecanoic acid was also spiked into representative samples, and recovery of this internal standard converted to FAME was above 95%. We also verified that the conversion efficiency of TAG, free fatty acids and phospholipids was above 95% by converting standards into FAMES.

Several sources were used to determine comparative yields from different large-scale biofuel production platforms. For the production of lipids from *N. gaditana*, we extrapolated our small-scale yield values to large-scale production assuming 36 l per square metre. These values were corrected for 12 h:12 h light/dark cycles using a correction factor of 0.66, which we deducted from previous lipid production

experiments using light/dark cycles. For soy, palm and jatropha lipid production, we used the values cited by Chisti *et al.*²⁸ For *Chlorella* sp. and *Neochloris oleoabundans* lipid production, we used values from Chen *et al.*²⁹ For the yields shown by the grey bars in Figure 1d, we utilized values derived from Atsumi *et al.*³⁰ These include *S. elongatus* isobutyraldehyde and isobutanol production values from their results during cultivation in 24 h light; ethanol production from *S. elongatus*; hydrogen production from *Anabaena variabilis*, *C. reinhardtii* and *Oscillatoria* sp.; and lipid production from *Haematococcus pluvialis*.

DNA and RNA extraction for sequencing. DNA was extracted from separate cultures as described previously¹⁹. Briefly, 10 ml of *N. gaditana* culture was pelleted and nucleic acid was extracted using phenol–chloroform according to the Newman protocol⁵⁷ and treated with RNase to degrade RNA. Nuclear, mitochondrial and plastid DNA were isolated in this fraction.

RNA was extracted as previously described¹⁹ from several different conditions and growth phases to increase the number of expressed genes. These conditions included, +/– nitrate, logarithmic phase, stationary phase, heat-shocked culture (2 h at 37°C), cold-treated culture (2 h at 4°C), 12-h dark acclimation, and +/– supplemental CO₂. Aliquots of RNA from each condition were then pooled for conversion into cDNA and sequenced to obtain transcriptome data. The quality of the DNA and RNA used for sequencing was assessed by agarose-gel analysis and an Agilent Bioanalyzer, and satisfied all quality-control metrics.

DNA and RNA sequencing. To capitalize on their differing strengths, we employed both Roche (454) and Illumina sequencing. The relatively long sequencing reads produced by the Roche technologies are especially useful for resolving short repeated sequences during assembly. Further, we performed sequencing using an Illumina protocol called LIPEs (Long insert paired-end sequencing), which pairs the sequencing reads at ~4 kb separation.

The Roche 454 sequences were processed to trim off primer sequences. All sequencing reads were trimmed to an error probability of ~1:100 and to contain no ambiguous nucleotide identities. Reads shorter than 30 nucleotides were removed. Supplementary Table S10 shows the performance metrics for each type of sequencing method and Supplementary Figure S8 shows the read length distribution before and after trimming (There was little change in this for the Illumina sequencing reads.)

To generate an assembled transcriptome and to map transcriptome reads to the genome for the most accurate possible gene models (using what is commonly called ‘RNA-seq’ methods), we generated Illumina sequencing on pooled samples of cDNA that had been made from polyA-selected RNA isolated from varying conditions (above). This cDNA was sequenced using the Illumina SIPES (Short insert paired-end sequencing) protocol, which pairs the sequencing reads at ~200 bp separation. Sequencing reads were trimmed as for the genomic sequencing data, except that all reads as long as 20 nts were retained. A total of 17,823,072 raw reads were determined and 17,723,662 survived the trimming, with a mean trimmed read length of 50.7 nts. If we assume a total transcriptome size of 11.8 MB with an average transcript length of 1,180 nts (estimated from the inferred gene content in version 1.1; see below), this corresponds to 76× sequencing coverage on a transcript represented at the mean level in these samples.

Assembly of nuclear genome. Numerous assemblies were performed using a variety of genome assembly programs with varying parameters followed by measuring and comparing their quality. We then used software, developed in-house, for merging of the results of the best of these various attempts into a single, well-reconciled assembly that capitalizes on the relative strengths of each type of assembly software while minimizing the chance of creating redundancy. After many trials, the best result was obtained by using Newbler version 2.3 (Roche) to assemble the Roche data and convey graph constructor (CGC; Convey Computer Corporation) to assemble the Illumina data followed by scaffold creation (based on paired-end reads that fall into adjacent contigs) using Velvet version 1.0.19 (refs 58,59).

To perform the merging step, all contig/scaffold sequences from the Newbler and Velvet assemblies were first aligned all-by-all using BLASTn. Aligned pairs with a minimum reported BLAST expect value (*E*-value) of 1E-7 were then screened by requiring a minimum overlapping length of 40 bp and minimum identity score of 95. The identity score was calculated as the sum of +1 for base matches, –2 penalty for mismatches, and –1 penalty for insertions or deletions (indels). As homopolymer errors are common issues for reads generated by the Roche platform, a lower indel penalty was used. A graph was then built with contigs as nodes and their pairings as edges, then continuously overlapping contigs were constructed in a greedy fashion, that is, the longest path wins in the case of conflict. In the case of conflict where one path is not significantly longer than the other (that is, differing in length by < 40 nt), neither path was created. To reduce homopolymer errors from merged sequences, the consensus used the portion of the overlap taken from only the Illumina sequencing reads.

We noticed that some of the scaffolds from the first assembly were likely to be bacterial. We also found that some of these scaffolds had been created from only the Illumina sequencing reads with few or no aligning Roche reads. Considering this, and suspecting that there had been some bacterial contamination specifically in the DNA preparation used for Illumina sequencing, we conducted a

systematic search based on sequence matching to either bacterial or stramenopile genomic sequences and careful manual examination. This identified 363 scaffolds that contain a total of 7 Mb that were concluded to be bacterial, which were then removed along with the five scaffolds that constituted the mitochondrial and plastid genomes, leaving 2,087 scaffolds in what we designated assembly version 1.1. To estimate genome size, we subtracted duplicated regions with 100% homology from the 2,087 scaffolds. Following these corrections we assume a genome size of ~29 Mb. It takes 257 scaffolds from the assembly to contain half of the total scaffold size (the N50 statistic), and the length of the 257th longest scaffold is 37,693 nts (the L50 statistic). There are 35 scaffolds longer than 100 kb, a total of 561 longer than 20 kb, and a total of 1,447 that are longer than 2 kb. Supplementary Figure S9 shows the distribution of the paired-end separation distance for the Illumina LIPEs pairs. Additional methods are found in the Supplementary Information.

References

- Steen, E. J. *et al.* Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559–562 (2010).
- Radakovits, R., Jinkerson, R. E., Darzins, A. & Posewitz, M. C. Genetic engineering of algae for enhanced biofuel production. *Eukaryot. Cell* **9**, 486–501 (2010).
- Jang, Y.-S. *et al.* Engineering of microorganisms for the production of biofuels and perspectives based on systems metabolic engineering approaches. *Biotech. Adv.* <http://dx.doi.org/10.1016/j.biotechadv.2011.08.015> (2011).
- Li, H., Cann, A. F. & Liao, J. C. Biofuels: biomolecular engineering fundamentals and advances. *Annu. Rev. Chem. Biomol. Eng.* **1**, 19–36 (2010).
- Lawrence, P. W. Engineering microbes to produce biofuels. *Curr. Opin. Biotechnol.* **22**, 388–393 (2011).
- Ducat, D. C., Way, J. C. & Silver, P. A. Engineering cyanobacteria to generate high-value products. *Trends Biotechnol.* **29**, 95–103 (2011).
- Jinkerson, R. E., Subramanian, V. & Posewitz, M. C. Improving biofuel production in phototrophic microorganisms with systems biology. *Biofuels* **2**, 125–144 (2011).
- Rodolfi, L. *et al.* Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol. Bioeng.* **102**, 100–112 (2009).
- Converti, A., Casazza, A. A., Ortiz, E. Y., Perego, P. & Del Borghi, M. Effect of temperature and nitrogen concentration on the growth and lipid content of *Nannochloropsis oculata* and *Chlorella vulgaris* for biodiesel production. *Chem. Eng. Process.* **48**, 1146–1151 (2009).
- Gouveia, L. & Oliveira, A. Microalgae as a raw material for biofuels production. *J. Ind. Microbiol. Biotechnol.* **36**, 269–274 (2009).
- Pal, D., Khozin-Goldberg, I., Cohen, Z. & Boushiba, S. The effect of light, salinity, and nitrogen availability on lipid production by *Nannochloropsis* sp. *Appl. Microbiol. Biotechnol.* **90**, 1429–1441 (2011).
- Zou, N., Zhang, C., Cohen, Z. & Richmond, A. Production of cell mass and eicosapentaenoic acid (EPA) in ultrahigh cell density cultures of *Nannochloropsis* sp. (Eustigmatophyceae). *Eur. J. Phycol.* **35**, 127–133 (2000).
- Merchant, S. S. *et al.* The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318**, 245–250 (2007).
- Bowler, C. *et al.* The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008).
- Siaut, M. *et al.* Molecular toolbox for studying diatom biology in *Phaeodactylum tricorutum*. *Gene* **406**, 23–35 (2007).
- Zaslavskaja, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R. & Apt, K. E. Transformation of the diatom *Phaeodactylum tricorutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *J. Phycol.* **36**, 379–386 (2000).
- Boynton, J. *et al.* Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. *Science* **240**, 1534–1538 (1988).
- Kindle, K. L. High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* **87**, 1228–1232 (1990).
- Radakovits, R., Eduafo, P. M. & Posewitz, M. C. Genetic engineering of fatty acid chain length in *Phaeodactylum tricorutum*. *Metab. Eng.* **13**, 89–95 (2011).
- Work, V. H. *et al.* Increased lipid accumulation in the *Chlamydomonas reinhardtii* sta7–10 starchless isoamylase mutant and increased carbohydrate synthesis in complemented strains. *Eukaryot. Cell* **9**, 1251–1261 (2010).
- Wang, Z. T., Ullrich, N., Joo, S., Waffenschmidt, S. & Goodenough, U. Algal Lipid Bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*. *Eukaryot. Cell* **8**, 1856–1868 (2009).
- Li, Y. *et al.* *Chlamydomonas* starchless mutant defective in ADP-glucose pyrophosphorylase hyper-accumulates triacylglycerol. *Metab. Eng.* **12**, 387–391 (2010).
- Simionato, D. *et al.* Acclimation of *Nannochloropsis gaditana* to different illumination regimes: effects on lipids accumulation. *Bioresour. Technol.* **102**, 6026–6032 (2011).
- Boushiba, S., Vonshak, A., Cohen, Z., Avissar, Y. & Richmond, A. Lipid and biomass production by the halotolerant microalga *Nannochloropsis salina*. *Biomass* **12**, 37–47 (1987).

25. Kilian, O. & Vick, B. Homologous recombination in an algal nuclear genome, Patent number: US 2011/0091977 A1 (2011).
26. Kilian, O., Benemann, C. S. E., Niyogi, K. K. & Vick, B. High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proc. Natl Acad. Sci. USA* **108**, 20859–20860 (2011).
27. Hu, Q. *et al.* Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J.* **54**, 621–639 (2008).
28. Chisti, Y. Biodiesel from microalgae. *Biotech. Adv.* **25**, 294–306 (2007).
29. Chen, C.-Y., Yeh, K.-L., Aisyah, R., Lee, D.-J. & Chang, J.-S. Cultivation, photobioreactor design and harvesting of microalgae for biodiesel production: a critical review. *Bioresour. Technol.* **102**, 71–81 (2010).
30. Atsumi, S., Higashide, W. & Liao, J. C. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat. Biotechnol.* **27**, 1177–1180 (2009).
31. Sumathi, S., Chai, S. P. & Mohamed, A. R. Utilization of oil palm as a source of renewable energy in Malaysia. *Renew. Sust. Energ. Rev.* **12**, 2404–2421 (2008).
32. Oudot-Le Secq, M.-P. *et al.* Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* comparison with other plastid genomes of the red lineage. *Mol. Genet. Genomics* **277**, 427–439 (2007).
33. Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
34. Fernandez, E. & Galvan, A. Inorganic nitrogen assimilation in *Chlamydomonas*. *J. Exp. Bot.* **58**, 2279–2287 (2007).
35. Quesada, A. & Fernández, E. Expression of nitrate assimilation related genes in *Chlamydomonas reinhardtii*. *Plant Mol. Biol.* **24**, 185–194 (1994).
36. Cock, J. M. *et al.* The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621 (2010).
37. Gobler, C. J. *et al.* Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc. Natl Acad. Sci. USA* **108**, 4352–4357 (2011).
38. Blanc, G. *et al.* The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* **22**, 2943–2955 (2010).
39. Matsuzaki, M. *et al.* Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657 (2004).
40. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).
41. Karpowicz, S. J., Prochnik, S. E., Grossman, A. R. & Merchant, S. S. The GreenCut2 resource, a phylogenomically derived inventory of proteins specific to the plant lineage. *J. Biol. Chem.* **286**, 21427–21439 (2011).
42. Miller, R. *et al.* Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen-deprivation predict diversion of metabolism. *Plant Physiol.* **154**, 1737–1752 (2010).
43. Eisenreich, W., Bacher, A., Arigoni, D. & Rohdich, F. Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cell. Mol. Life Sci.* **61**, 1401–1426 (2004).
44. Reinfelder, J. R. Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Annu. Rev. Mar. Sci.* **3**, 291–315 (2011).
45. Huertas, I. E., Colman, B. & Espie, G. S. Inorganic carbon acquisition and its energization in eustigmatophyte algae. *Funct. Plant Biol.* **29**, 271–277 (2002).
46. Huertas, I. E., Espie, G. S., Colman, B. & Lubian, L. M. Light-dependent bicarbonate uptake and CO₂ efflux in the marine microalga *Nannochloropsis gaditana*. *Planta* **211**, 43–49 (2000).
47. Huertas, I. E. & Lubian, L. M. Comparative study of dissolved inorganic carbon utilization and photosynthetic responses in *Nannochloris* (Chlorophyceae) and *Nannochloropsis* (Eustigmatophyceae) species. *Can. J. Bot.* **76**, 1104–1108 (1998).
48. Roberts, K., Granum, E., Leegood, R. C. & Raven, J. A. C₃ and C₄ pathways of photosynthetic carbon assimilation in marine diatoms are under genetic, not environmental, control. *Plant Physiol.* **145**, 230–235 (2007).
49. Kroth, P. G. *et al.* A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* **3**, e1426 (2008).
50. Edwards, G. E., Franceschi, V. R. & Voznesenskaya, E. V. Single-cell C₄ photosynthesis versus the dual-cell (Kranz) paradigm. *Annu. Rev. Plant Biol.* **55**, 173–196 (2004).
51. Offermann, S., Okita, T. W. & Edwards, G. E. Resolving the compartmentation and function of C₄ photosynthesis in the single-cell C₄ species *Bienertia sinuspersici*. *Plant Physiol.* **155**, 1612–1628 (2011).
52. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953–971 (2007).
53. Gschloessl, B., Guermeur, Y. & Cock, J. M. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* **9**, 393 (2008).
54. Apt, K. E., Grossman, A. R. & Kroth-Pancic, P. G. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet.* **252**, 572–579 (1996).
55. Chen, H. L., Li, S. S., Huang, R. & Tsai, H.-J. Conditional production of a functional fish growth hormone in the transgenic line of *Nannochloropsis oculata* (Eustigmatophyceae). *J. Phycol.* **44**, 768–776 (2008).
56. Li, S.-S. & Tsai, H.-J. Transgenic microalgae as a non-antibiotic bactericide producer to defend against bacterial pathogen infection in the fish digestive tract. *Fish Shellfish Immunol.* **26**, 316–325 (2009).
57. Newman, S. M. *et al.* Transformation of chloroplast ribosomal RNA genes in *Chlamydomonas*: molecular and genetic characterization of integration events. *Genetics* **126**, 875–888 (1990).
58. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
59. Zerbino, D. R., McEwen, G. K., Margulies, E. H. & Birney, E. Pebble and Rock Band: heuristic resolution of repeats and scaffolding in the Velvet short-read *de novo* assembler. *PLoS ONE* **4**, e8407 (2009).
60. Tyler, B. M. *et al.* Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**, 1261–1266 (2006).

Acknowledgements

We would like to thank Anis Karimpour-Fard and Jason Shao for their help and advice. R.E.J. was supported by a Graduate Research Fellowship from the National Science Foundation. R.R. and this research were supported with funding provided by Conoco-Phillips through a grant to the Colorado Center for Biofuels and Biorefining (C2B2) and the Air Force Office of Scientific Research (FA9550-11-1-0211).

Author contributions

M.C.P. and R.R. conceived and planned the project. R.R. wrote the manuscript, extracted RNA and genomic DNA, developed the transformation protocol, and assisted with annotation and metabolic pathway analysis. R.E.J. performed comparative genome analysis and metabolic pathway analysis, assisted with annotation and writing of the manuscript. S.I.F. and J.L.B. assembled the scaffolds, enabled and implemented the automated annotation of the genome and performed the RNA quantification. H.T. and R.E.S. developed and implemented the method for merging Illumina and Roche sequencing assemblies. M.C.P., S.I.F. and J.L.B. also assisted with the manuscript writing.

Additional information

Accession codes: This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AGNI00000000. The version described in this paper is the first version, AGNI01000000. The data can also be freely accessed through the project's website, <http://Nannochloropsis.genomeprojectsolutions-databases.com/>.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Radakovits, R. *et al.* Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nat. Commun.* **3**:686 doi: 10.1038/ncomms1688 (2012).

License: This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>