Algal Research 54 (2021) 102181

Contents lists available at ScienceDirect

Algal Research

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

Biomolecule composition and draft genome of a novel, high-lipid producing Scenedesmaceae microalga

Vilmos Soós [a,*], Prateek Shetty [b,c], Gergely Maróti [b,c,d], Norbert Incze [a], Eszter Badics [a], Péter Bálint [e], Vince Ördög [e,f], Ervin Balázs [a,f]

[a] Department of Applied Genomics, Centre for Agricultural Research, 2 Brunszvik str, 2462 Martonvásár, Hungary
[b] Plant Biology Institute of the Biological Research Center, 62 Temesvári krt., 6726 Szeged, Hungary
[c] Seqomics Biotechnology Ltd., 7 Vállalkozók str., 6782 Mórahalom, Hungary
[d] Faculty of Water Sciences, National University of Public Service, 12-14 Bajcsy-Zs. str., 6050 Baja, Hungary
[e] Institute of Plant Science, Faculty of Agricultural and Food Sciences, Széchenyi István University, 2 Lucsony str., 9200 Mosonmagyaróvár, Hungary
[f] Research Centre for Plant Growth and Development, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, Private Bag X01, Scottsville 3209, South Africa

ARTICLE INFO

Keywords:
Scenedesmaceae
Genome sequencing
Transcriptome
Lipid accumulation
N-starvation

ABSTRACT

Lipid biosynthesis in microalgae can be stimulated by cultivation in low nitrogen medium. MACC-401 was isolated from the soil surface in Tres Marias (MG-Brazil). The strain shows the morphological characteristics of the Scenedesmaceae green algae. The daily biomass and lipid production of MACC-401 is remarkable, 0.36 g L^{-1} and 110 mg L^{-1}, respectively. Exploration of the genetic background of this promising strain not only allows the utilization of this species for industrial-scale lipid production, but also provides genetic targets to select lipid-producing strains from microalgae collections. We conducted physiological experiments by cultivating MACC-401 in complete and N-limited media and performed genome sequencing as well as transcriptome analysis. The estimated nuclear genome size of MACC-401 is 99.503 Mbp and the chloroplast genome is 0.15 Mbp. The phylogenetic analysis confirmed that the MACC-401 belongs to the Scenedesmaceae family, and represents a genetically distinct accession in this family. A basic comparative transcriptome analysis resulted in the identification of N-starvation responsive genes, which could serve as markers to monitor the onset of lipid accumulation in algal cultures.

1. Introduction

Biofuel or "energy" crops compete with major food crops for resources, such as arable land, freshwater and natural landscapes with biodiversity and have limited abilities to be targets for biofuel production [1]. Thus, the idea of sustainable production of renewable energy is heavily disputed and these concerns paved the way for the development of second and third generation biofuels such as microalgae, which can be cultivated on degraded land and consume lesser resources [2].

The increasing dependence on energy, primarily from fossil fuel has resulted in unprecedented levels of CO_2, leading to global warming and climate change and threatening global food security, social welfare and biodiversity. CO_2 is converted into high energy organic compounds by photosynthesis providing a renewable source for fuel production. Very recently, the increasing petroleum prices and need to reduce greenhouse gas emission stirred up interest in large-scale biodiesel production [3]. Microalgae biodiesel has been identified as a sustainable, renewable and environment friendly alternative fuel having the ability to replace conventional fossil diesel fuel [3–5]. Lipids produced by microalgae can be structural lipids, storage lipids, or might have signaling functions. Storage lipids are mainly in the form of triacylglycerols (TAG) consisting of predominately saturated and unsaturated fatty acids which can be easily catabolized to support the energy needs of growing cells [4]. TAGs are mostly synthesized in the light, stored in cytosolic lipid bodies, and then reutilized for polar lipid synthesis in the dark [6].

Microalgae can set or modify their metabolic processes very efficiently to survive adverse and suboptimal environmental conditions. Under optimal growth conditions, green algae produce high biomass with relatively low lipid content, which constitute about 5–20% of dry cell weight [7]. Biomass production and TAGs biosynthesis compete for

* Corresponding author.
E-mail address: soos.vilmos@atk.hu (V. Soós).

https://doi.org/10.1016/j.algal.2020.102181
Received 2 October 2020; Received in revised form 24 November 2020; Accepted 20 December 2020
Available online 9 January 2021

assimilates. Thus, a physiological reprogramming is a prerequisite to induce lipid biosynthesis. Adverse environmental or stress conditions trigger a shift of lipid biosynthetic pathways towards the formation and accumulation of neutral lipids (20–50% dry cell weight) [7], predominantly in the form of TAG, enabling algal cells to sustain these suboptimal conditions [8]. Alternatively, these lipids are suitable precursors for biodiesel production after transesterification [9]. The amount and composition of TAGs produced is species/strain-specific and controlled by the genetic background of microalgae strains [10]. Nutrient starvation, especially the most critical nitrogen depletion is one of the most extensively used lipid induction methods in microalgal TAG production and has been observed in many species [11].

Although the physiological response to macronutrient starvation is extensively studied, our knowledge about the proper molecular mechanisms driving these metabolic shift is still scarce [12]. A substantial insight into the early transcriptional reprogramming upon nitrogen depletion revealed that *Chlamydomonas reinhardtii* cells respond within minutes, and subsequently, core metabolism and carbon fixation are remodeled quickly to induce TAG accumulation [13,14]. In the early and mid-stages of induction (0–60 min) genes responsible for sensing N deficiency and metabolic reorganization are affected. Between 1 and 8 h post depletion, the expression of gene sets included in the lipid accumulation is altered. However, the complex molecular choreography of the phenotypic readjustment and growing parameters of lipid accumulating (hyperaccumulator) strains over longer time course is still elusive. Stresses lower the productivity and growth rate of the system, which hinders the production of biofuels and by-products on commercial scales. An inverse correlation between biomass production and TAG content has been observed, as the expense of TAG synthesis is not compensated during adverse environmental conditions [12]. The slower growth rate negatively affects lipid productivity as reflected by the variations in biomass and total lipid content [15,16]. As a result, microalgae cultures growing under nitrogen-stressed conditions increased in lipid content and decreased in protein content to compensate the energy needs for TAG synthesis [17,18]. It was shown that TAG accumulation can be triggered by moderately decreasing the N levels to 10%N to provide improved biomass accumulation and preferential lipid productivity [16,19]. The MACC-401 microalgae strain, a putative Scenedesmaceae sp. has drawn attention due to its unique lipid production capabilities even under less stressful conditions to maintain a moderate biomass production [16]. The strain was isolated from a green spot growing on soil surface of a fish farm in Três Marias, Minas Gerais, Brazil and maintained in the Mosonmagyaróvár Algal Culture Collection (MACC). The MACC-401 has not been extensively characterised previously. However, it shows a high daily biomass (0.36 g L^{-1}) and lipid (110 mg L^{-1}) production upon extended nitrogen starvation [16]. In this paper we further characterize the biomolecule and major lipid composition of this promising strain as well as we provide details of its genome and transcriptome and finally, we present N depletion responsive genes which might serve as markers to monitor the onset of lipid accumulation even in commercial scale N-starving cultures.

2. Material and methods

2.1. Microalgae cultures and growth conditions

The axenic *Scenedesmus* strain MACC-401 from the Mosonmagyaróvár Algal Culture Collection (MACC), Hungary was inoculated from agar stock cultures into Tamiya medium [16,20]. Culture conditions were as follows: 25 ± 2 °C and a 14:10-h light/dark cycle at a light intensity of 130 μmol photons m^{-2} s^{-1}, aerated during the light period with 20 L h^{-1} 1.5% CO_2-enriched sterile humidified air [21]. Cultures were checked for potential bacterial contamination after 7 days by spraying diluted algal suspension onto solidified nutrient medium (1% glucose, 1% peptone, and 1% yeast extract). Cultures were considered axenic if no bacterial colonies developed on the plates 7 days after incubation at 25 ± 2 °C.

The suspension cultures were subcultured into flasks containing 250 mL either N-replete (complete, 100%N) or N-deplete (limited, 10%N) Tamiya nutrient medium (0.5 g L^{-1} KNO_3 700 mg L^{-1} N) with a starting algal density of 10 mg L^{-1} DW. Algal biomass was harvested by centrifugation after 0, 3, 6, 9, 12 and 15 days in three biological replicates in each time point. The harvest yielded 3 × 6 samples from 100%N cultures (0, 3, 6, 9, 12 and 15 d) and 3 × 5 samples from 10%N cultures (3, 6, 9, 12 and 15 d). Dry weight (DW) quantification was performed according to Ördög et al. [16]. Briefly, algal cultures were filtered using dried and pre-weighed Whatman GF/C glass fiber filters, then the filters were dried for 2 h at 105 °C and weighed. Algal suspension density was calculated as g·L^{-1} DW.

For the chemical analysis of lipid composition, samples from each harvest were combined into one replicate per day to obtain sufficient biomass. DNA samples for genome sequencing has been collected from 15 d old samples. For RNA sequencing, three samples were harvested on day 6 and 15 from each experimental setup and replicates.

2.2. Carbohydrate quantification

The carbohydrate content of the samples was measured using a method based on the MSZ 6830/26 Hungarian standard (Determination of feed nutritive value: Determination of sugar content) as described by Aremu et al. [18].

2.3. Crude protein quantification

Crude protein content, as a derivative of nitrogen content was quantified using the Kjeldahl method (MSZ EN ISO 5983-2:2005 as described by Ördög et al. [16]). In short, dried algal samples (5–250 mg DW) were digested using a Tecator Digestion System 20 (1015 Digester, TECATOR, Sweden) as per manufacturer's recommendations. Steam distillation and titration of the digested samples was performed in a 2400 Kjeltec Unit automated instrument (FOSS Analytical AB, Sweden) to quantify the total nitrogen content in algal biomass. 0.1 M HCl was used as the standard volumetric solution. Crude protein content (g 100 g^{-1} DW; % DW) was calculated using the EN ISO 5983-2 standard as follows: WP 0 WN × F, where WP is the crude protein content (% DW), WN is the nitrogen content of the sample (% DW), and F is the Kjeldahl factor (F06.25).

2.4. Lipid quantification

Lipids were quantified as described in Ördög et al. [16]. In brief, three parallel algal preparations (100–150 mg DW) were hydrolyzed (10 mL 3 M HCl at 95–100 °C for 90 min) and the lipids were extracted from the lysate with a subsequent addition of 10 mL methanol, 10 mL hexane and 10 mL diethyl ether to the sample. After vortexing, samples were centrifuged at 2534 ×*g* for 3 min at 20 °C. Then, the extraction step was repeated with the supernatant mixed with 10 mL hexane and 10 mL diethyl ether. The collected supernatants were evaporated under vacuum at 45 °C and dried at 60 °C for 90 min. The resulting extract was considered as a measure of lipid yield, calculated as g 100 g^{-1} DW (% DW).

2.5. Fatty acid analysis

Fatty acid composition has been determined according to the method described in Ördög et al. [16]. Briefly, the dried residue from lipid yield quantification was dissolved in 2 mL hexane then, a further 2 mL hexane/dimethyl carbonate (50:50 *v*/v) and 1 mL sodium methylate (30% solution in methanol) were added and the sample has been vortexed. After adding 10 mL deionized water and 15 μL ethyl tridecanoate solution (Merck) as internal standard, the sample was extensively vortexed and the upper hexane layer was removed and dried. The sample was

analyzed by a Varian CP-3800, Varian Corp. USA gas chromatographic system equipped with a split/splitless injector and FID detector fitted with a FAMEWAX column (30 m, 0.25 mm ID, 0.25 μm df, Restek Corp. USA). Individual FAME peaks were identified by their retention times in comparison to the reference solution (Supelco 37 Component FAME Mix, SigmaAldrich Corp.). FA content in each sample was calculated as g 100 g^{-1} lipid (% lipid) using the internal standard method according to EN 5508 standard and then converted to g 100 g^{-1} DW (% DW).

2.6. DNA extraction

For DNA extraction, algal cells were harvested by centrifugation and immediately frozen using liquid N_2. Subsequently, cells were allowed to thaw out and were frozen again to properly disrupt cell walls. This freeze-thaw cycle was repeated rapidly twice. DNA was extracted by a classic, modified version of CTAB method [22], then samples were cleaned up with DNEasy Plant Mini Kit (Qiagen).

2.7. Genome sequencing, whole genome assembly and annotation

A mate-pair library was generated using Illumina Nextera Mate-Pair Kit with insert sizes ranging between 7 and 11 kb. Also, a paired-end library was generated using the NEBNext Ultra II DNA Library Prep kit. DNA sequencing was carried out on an Illumina MiSeq machine using V2 sequencing chemistry resulting in 2 × 250 nt reads (for both libraries). Primary data analysis (base-calling) was carried out with MiSeq Reporter software provided by Illumina. For *de novo* assembly, paired-end and mate-pair Illumina reads were adapter and quality trimmed using Adapterremoval2 (v. 2.1.7) [23] with the switches –trimns and –trimqualities. Illumina reads were used as input for Spades (version 3.11.1) [24] which was used for genome assembly by utilizing paired-end reads and mate-paired reads (Table 1). Metrics for assembled reads was calculated using "stats.sh" tool from the BBmap utilities (version 38.34) Genome assembly completion was evaluated using BUSCO v3 [27] against the Chlorophyta marker gene database.

2.8. Chloroplast assembly and annotation

NOVOplasty (v2.7.2), an organelle specific assembler was used to assemble the chloroplast genome [28]. The assembled genome was annotated and visualized using GeSeq [29].

2.9. Phylogenetic analysis

Phylogenetic analysis was carried out by using translated protein sequences of six genes from the chloroplast genome, namely *psbD, psaC, psbB, rbcL, atpA,* and *psaB*. The sequences were concatenated into a single, 2684 amino acid long sequence and aligned using MUSCLE. The MUSCLE alignment was used as input to build phylogenetic trees using iqTree (v1.6.12) [30] with 2000 ultrafast bootstrap replicates and 468 different models compatible with RaxML [31]. 337 parsimony-informative sites, 535 singleton sites and 1812 constant sites were identified by iqTree. Generated tree along with bootstrap values were visualized in iTOL [32].

Table 1
Metrics of reads and assembled scaffolds.

No of MP reads	2,239,520
No. of PE reads	17,311,956
No. of scaffolds > 1 kb	5091
Total number of bases in assembled genome	99.53 MB
N50	58.742 KB
N90	8.369 KB
GC percentage	57,49
Max scaffold length	456.32 KB
Percentage of main genome in scaffolds > 50 KB	55.23%

2.10. RNA sequencing

The same rapid freeze–thaw cycles used for DNA extraction protocol was applied to extract RNA. Disrupted samples were extracted with Trizol reagent (Invitrogen) and the resulting RNA was cleaned with RNEasy Mini Kit (Qiagen) according to the manufacturers' protocol. RNA quality was checked with TapeStation according to the manufacturers instructions (Agilent). Only high quality samples were processed. RNA samples of three technical repeats in each biological repeats were pooled in equal quantities (700 ng each). Paired-end libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina. An Illumina NextSeq instrument was used to generate 2 × 150 nt reads by using NextSeq 500/550 High Output Kit v2.5 (300 Cycles) for sequencing.

2.11. Raw read correction and adapter trimming

Raw reads were downloaded from NCBI using SRA IDs. The downloaded datasets were first read corrected with rCorrector [33]. Adapter trimming was carried out using Trim Galore! (v0.6.4) [34].

2.12. Transcriptome assembly

Reads from all conditions were used to assemble the transcriptome data. This was carried out individually for all datasets using Trinity (v2.1.1) [35]. CD-Hit (4.7) [36] was used to reduce the high number of isoforms that are generated when using Trinity for transcriptome assembly. Transcripts that had an identity of greater than 98% were marked as duplicates and removed from the assembly. This protocol was carried out for studies which had no transcriptome assembly available online.

2.13. Read mapping to transcriptome assembly

Reads were mapped using Hisat2 (v2.1.0) [37]. Mapped reads were quantified using eXpress [38]. This was carried out for all datasets with *de novo* transcriptome assembly. Transcriptome analysis with read counts was carried out in Rstudio (v1.3) [39] and differential expression analysis was performed using EdgeR [40]. Up and downregulated genes were identified keeping the N-repleted condition as control and N-depleted condition as treatment and with a *p*-value of greater than 0.05.

2.14. Functional annotation

Functional annotation was carried out in 2 parts. The first part was finding orthologs across different studies. This was carried out by using a reciprocal bidirectional blast method along with Transcriptologs to filter out bad matches. Orthologs were identified by using the up and down-regulated IDs from *Scenedesmus* [41] as reference. After finding similar matches across studies, functional annotation was provided by matching the *Scenedesmus* transcripts against algal proteins downloaded from Uniprot (uniprot.org).

2.15. Nucleotide sequence accession number

The complete genome sequence and the transcriptome data of MACC-401 have been deposited with links to BioProject accession number PRJNA663804 (genome) and PRJNA663868 (transcriptome) in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/).

3. Results and discussion

3.1. Major biomolecule composition differs between the N-repleted and N-starving MACC-401 cultures

A previous study revealed that the MACC-401 strain displays unique growth characteristics and a favourable lipid composition upon N depletion. N starving cultures accumulated lipids, such as saturated fatty acids and monounsaturated fatty acids after a period of 8 and 14 days. To further characterize the phenotype and biomolecule composition of lipid accumulating MACC-401, the strain was batch-cultivated under nitrogen repleted/control (100%N) and depleted (10%N) conditions. Samples were taken from cultures 0, 3, 6, 9, 12 and 15 days after inoculation. The total dry mass has been gradually increased in both conditions (Fig. 1A), and growth was not restricted under nitrogen-deprived conditions. A logarithmic growth phase was clearly observed (Fig. 1A), reaching a maximum of 3 g L^{-1} biomass on day 15. However, a lower protein content was observed after 9 days in cultures under conditions of nitrogen deprivation (Fig. 1D), clearly indicating that nitrogen reserves were depleted and protein synthesis has been adjusted to low N condition. The protein content slightly increased in 100%N cultures, and in contrast, slowly declined under 10%N conditions, denoting a trade-off between N depletion and protein synthesis. As nitrogen reserves were depleted, protein synthesis has been adjusted to N starvation as a result of the reprogramming of algal metabolism. In line with this phenomena, the total carbohydrate production increased in 10%N, and gradually decreased in 100%N cultures (Fig. 1B).

The fatty acid composition of MACC-401 strain over the course of growth experiment under nitrogen depletion has been investigated in detail. Contrary to the biomass production, there was a remarkable difference in total lipid content between ±N cultures. As expected, the total lipid content was *ca.* 3-fold higher on day 15 in 10%N cultures (Fig. 1C). The proportional amount of FAs whithin the total lipid content gradually increased in 10%N cultures (Fig. 2A) and the average daily FA production has been the highest on day 12 under nitrogen depleted conditions reaching 10-fold higher FA content compared to nitrogen repleted cultures (Fig. 2A and B). In 10%N cultures, FAs made up 62% of the total lipid content on day 12, while in 100%N cultures, FAs represented only 19% of total lipids (Fig. 2A and B).

The detailed analysis of FA composition revealed that uncharacterized FAs (other FAs) attributed largely to the total FA build up in the 0–15 days after N depletion (Fig. 2A and B; Supplementary Table 1.).

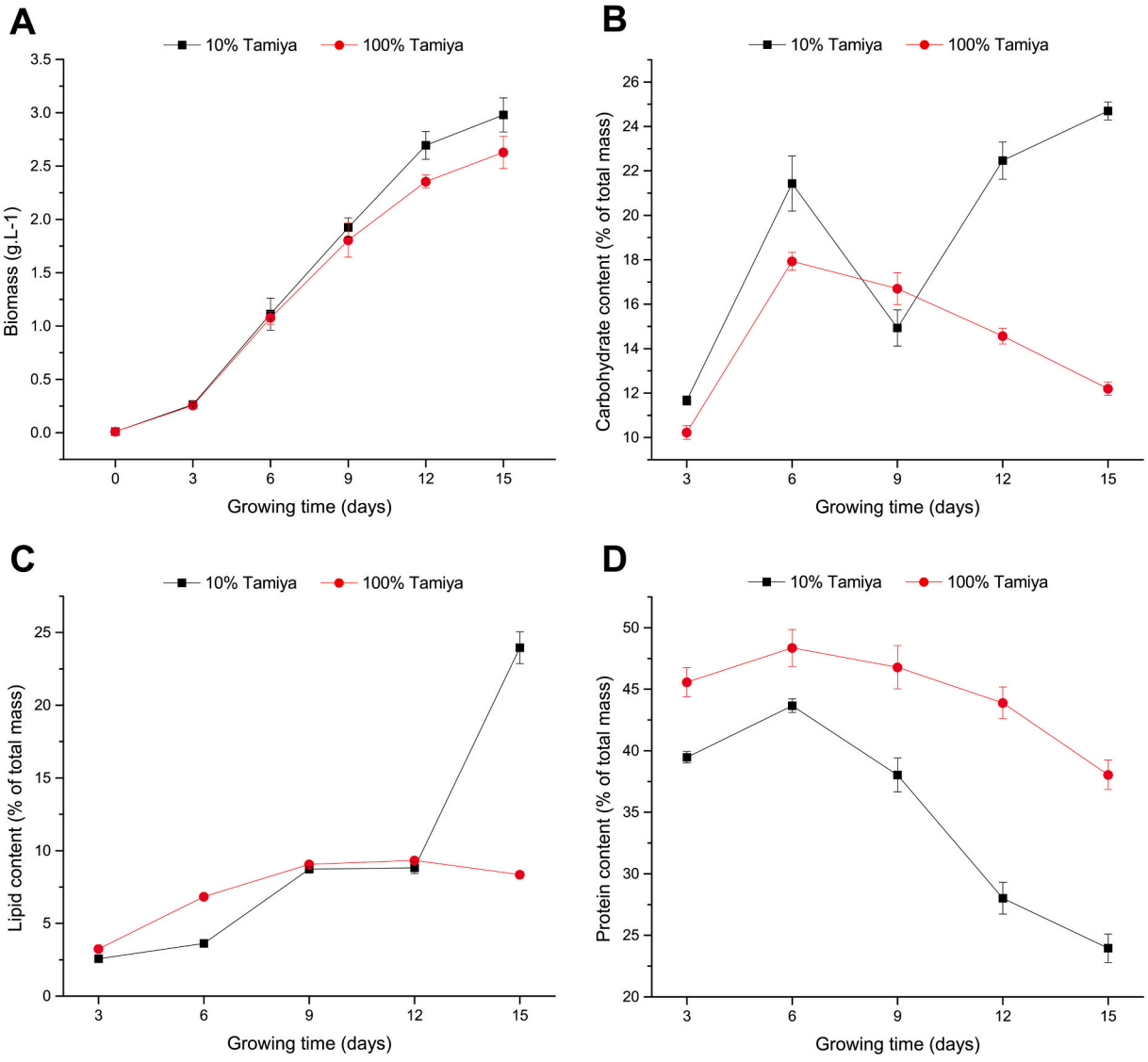

Fig. 1. Physiological parameters and major biomolecule composition of control (100%N Tamiya) and N-depleted (10%N Tamiya) growing cultures of MACC-401 strain ($n = 3$). Samples were harvested from cultures 0, 3, 6, 9, 12 and 15 days after inoculation.

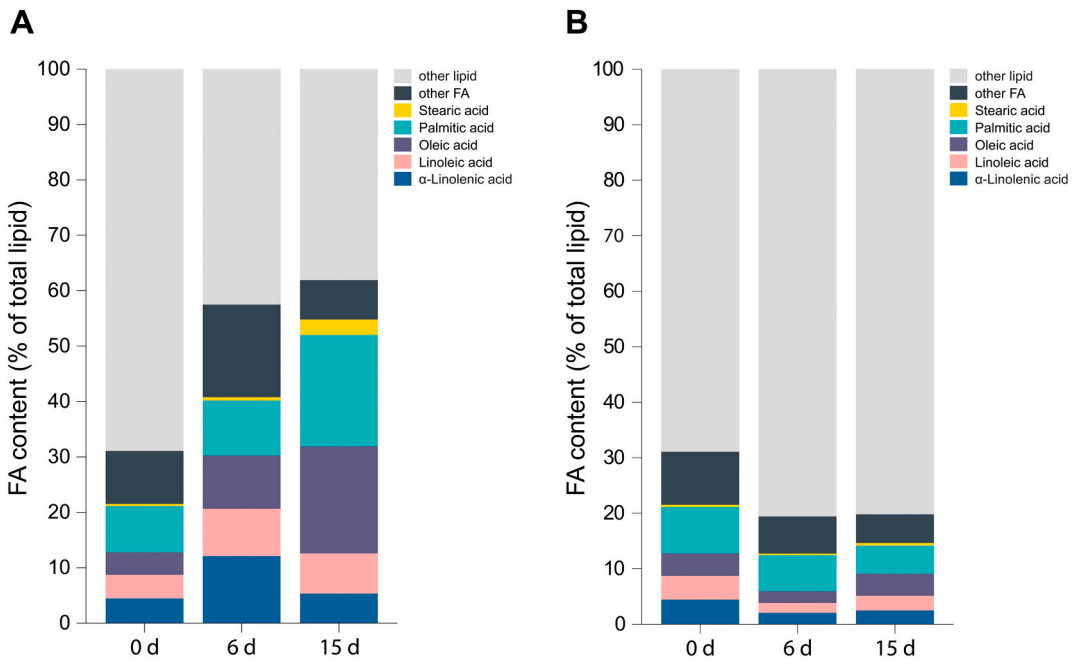

Fig. 2. The fatty acid composition of N-depleted (10%N, A) and N-repleted (100%N, B) cultures of MACC-401 strain. Samples were harvested from cultures 0, 6 and 15 days after inoculation ($n = 3$). Stearic acid, palmitic acid, oleic acid, linoleic acid and other FA content have been assessed by gas chromatography and related to the total lipid content.

Interestingly enough, on day 15 palmitic acid and oleic acid became more prevalent, making up more than a half of total FAs. Although linoleic acid, α-linolenic acid and stearic acid represented a smaller fraction of total FAs, the amount of these also increased dramatically by day 15. γ-linolenic acid made up only a fraction of total FAs (Fig. 2A and B).

3.2. The draft genome of MACC-401

A total of about 17 million paired-end reads and about 2 million mate-paired reads were generated using Illumina MiSeq. The assembly of the draft genome of MACC-401 resulted in 18,549 contigs and 5094 scaffolds (Table 1). The scaffolds and contigs represent nuclear, plastid and mitochondria genomes. Only scaffolds greater than 1 KB in size were retained. The estimated total nuclear genome size of MACC-401 is 99.503 Mbp. To test the completeness of the assembly, we compared the assembly against BUSCO marker genes for the Chlorophyta lineage revealing that the genome is 87.2% complete (Supplementary Table 2). The sequencing confirmed that the MACC-401 microalgae isolate belongs to the Scenedesmaceae family, and the phylogenetically closest, sequenced species is the *Scenedesmus armatus* var. *subalternans* [42].

The assembled chloroplast genome consists of 150,832 nucleotides, 83 annotated protein coding genes and 23 tRNA genes (Fig. 3 and Table 2.). The inverted repeat regions were 9567 kb in size, and very similar to the other closely related organisms of *Acutodesmus obliquus* (NCBI Reference Sequence: NC_008101.1) and *Pectinodesmus pectinatus* (GenBank ID: KU847995.1) (Supplementary Figure 1).

3.3. The phylogenetic status of MACC-401

To further elucidate the phylogenetic relationships of the MACC-401 isolate, we performed a detailed analysis using *psbD, psaC, psbB, rbcL, atpA,* and *psaB* genes from the assembled chloroplast genome. Selected green algae representatives with available complete chloroplast genome sequences are included in the phylogenetic tree. *Ulva fasciata* served as the outgroup species. The phylogenetic tree demonstrates that the MACC-401 strain clusters with 100% bootstrap value within the Scenedesmaceae clade (Fig. 4). However, it clusters away from *Acutodesmus obliquus* (*Scenedesmus obliquus/ Tetradesmus obliquus*) and *Pectinodesmus pectinatus*. Considering the obvious genetic distances from known species in the Scenedesmaceae family, we assume that MACC-401 represents a divergent genus within this family.

3.4. Selection of N-starvation responsive genes in MACC-401

To obtain a comprehensive list of N-repletion related genes in microalgae, we compared the published transcriptomic data of widely studied oleaginous and non-oleaginous species to the transcriptome of N-starving MACC-401 cultures. In the first hours of induction, when cultures are in the exponential growth stage, acyl chains are recycled from membrane lipids for early TAG build up, starch is accumulated and the metabolism of N-depleted cells is being remodelled [43]. The time resolved expression of early N-starvation responsive genes in various algae species and strains have been extensively investigated in the recent years. In *Chlamydomonas reinhardtii*, genes whose expression has been affected 0, 0.5, 1, 2, 4, 6, 12 and 24 h after N deprivation have been characterised [13]. Similarly, in the promising oleaginous algae *Monoraphidium neglectum* [44], *Synechococcus elongatus* [45], or in *Nannochloropsis oceanica* [46] early stage responsive genes have been identified which were induced after 2, 4, 8 and 24 h, 24 and 48 h, or 3, 4, 6, 12, 24 and 48 h of N depletion, respectively. The gene expression landscape of N starving algal cells in the latter growth phase has also been widely characterised. For example, cultures of lipid hyper-accumulating *Monoraphidium neglectum* [44], *Nannochloropsis oceanica* [46], *Dunaliella tertiolecta* [47] or *Neochloris oleoabundans* [48] were investigated in detail. Sampling time and period varied between the experiments, such as 48, 56 and 96 h in *Monoraphidium neglectum*, 2 days in *Nannochloropsis oceanica*, 3 and 5 days in *Dunaliella tertiolecta* or 11 days in *Neochloris oleoabundans*. The transcriptome analysis and gene expression profiling of N-starving, 2d old *Scenedesmus acutus* culture has been published by Sirikhachornkit et al. [41].

These valuable transcriptome resources provided the basis to select universal N-starvation responsive genes which expression is likely altered in lipid accumulating MACC-401 cultures as well as in their

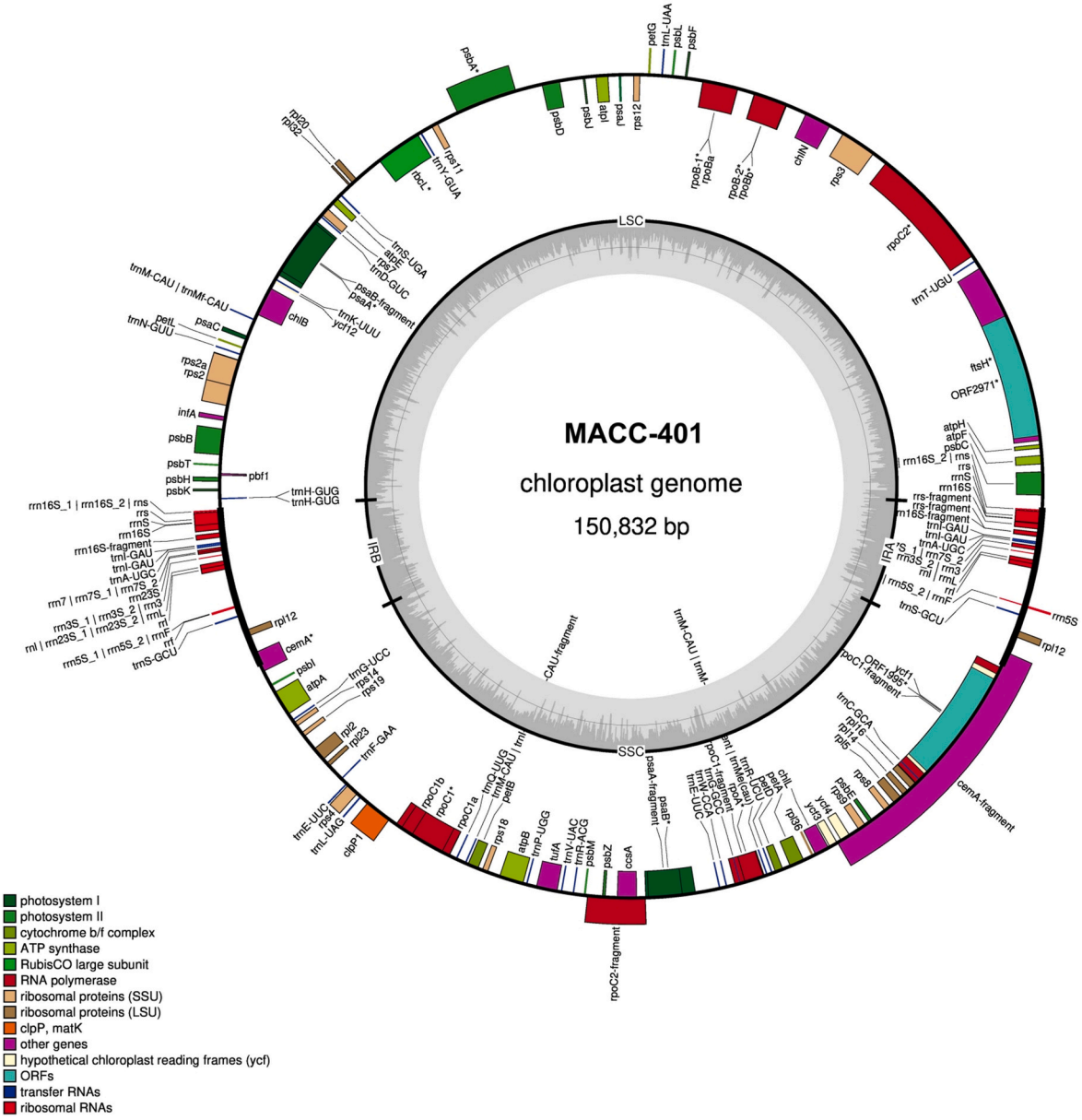

Fig. 3. The assembled chloroplast genome of MACC-401. Genes drawn inside are transcribed clockwise, and genes drawn outside are counterclockwise directions. Genes attributed to different functional groups are color-coded. Inner circle represents the quadripartite structure of the chloroplast: small single copy (SSC), large single copy (LSC) and a pair of inverted repeats (IRa and IRb). The innermost (gray) ring marks the GC content. Asterisks denote intron containing genes.

Table 2
Chloroplast genome metrics: IR stands for inverted repeat region.

	MACC-401	Acutodesmus obliquus	Pectinodesmus pectinatus
No. of bases	150,832	161,452	196,809
IRA region	150,783–9517	149,430–161,452	183,316–196,809
IRB region	87,438–97,005	72,440–84,462	99,157–112,650
IR size	9567	12,022	13,493
No. of genes	83	87	84
No. of tRNA	29	30	30

stationary growth phase. This shared set of genes could be used to easily monitor the onset of lipid production in N starving algal cultures, even on the industrial scale. To generate a list of potentially N-depletion related genes across various algae taxons we collected and extracted the set of transcripts which were up, or downregulated in N-starving cultures in seven published experiments [13,41,43–47] (Supplementary Table 3.). Then, we filtered this gene set with N depletion-related transcriptome datasets obtained from 6 and 15 days old N-deprived MACC-401 cultures in stationary growth phases.

A total of 61 genes were either up (33 genes) or downregulated (28 genes) with FC$\geq \pm$ 1.5 across different studies compared to the *Scenedesmus acutus* transcriptome which we used as a reference [41]. Hereafter the term "upregulated" refers to genes which expression were switched on under nitrogen deprivation conditions. Correspondingly, "downregulated" refers to genes that were repressed in N starving cultures.

We found 28 genes in lipid accumulating *Scenedesmus* cultures which were downregulated and exhibited a similar expression pattern in other algal species, too (Table 3). The observed downregulation of LHC protein genes correlates well with the finding that the chlorophyll content

Fig. 4. Phylogenetic analysis of MACC-401 isolate. MUSCLE alignment was used as input to build phylogenetic tree, *Ulva fasciata* was used as outgroup.

Table 3
List of overlapping, downregulated genes between the *Scenedesmus acutus* reference transcriptome (N-starving cells), MACC-401 and six other transcriptomes of N-depleted, lipid producing microalgal cell cultures. Accessions which have the corresponding orthologous gene in the reference dataset, are marked with „+"; „-" represents no hit.

Subject (GenBank ID)	*Chlamydomonas reinhardtii*	MACC-401	*Dunaliella tertiolecta*	*Monoraphidium neglectum*	*Nannochloropsis oceanica*	*Neochloris oleoabundans*	*Synechococcus elongatus*	Annotation (product)
GFUP01003434.1	+	–	–	–	–	–	–	Citrate Synthase
GFUP01001804.1	–	+	+	–	–	–	–	Uncharacterised protein
GFUP01007113.1	–	+	–	–	–	–	–	Lipase_3 domain-containing protein
GFUP01047063.1	–	+	+	–	–	+	+	Protein kinase domain-containing protein
GFUP01001811.1	–	–	+	–	–	–	+	bZIP domain-containing protein
GFUP01001812.1	–	–	+	–	–	–	+	bZIP domain-containing protein
GFUP01035136.1	–	–	–	–	–	+	–	Laminin EGF-like domain-containing protein
GFUP01039412.1	–	–	+	–	+	–	–	Aurora protein
GFUP01017235.1	–	–	–	+	–	–	+	NAD(P)-bd_dom domain-containing protein
GFUP01022383.1	–	–	–	+	–	–	–	Putative rubredoxin-like protein
GFUP01027000.1	–	–	–	+	+	–	–	Phosphoglycerate kinase
GFUP01036337.1	–	–	–	+	–	+	–	Light-harvesting protein of photosystem II
GFUP01038441.1	–	–	–	+	–	+	–	Light-harvesting protein of photosystem I
GFUP01038446.1	–	–	–	+	–	–	–	Copper response defect 1 protein
GFUP01042291.1	–	–	–	+	+	–	+	Oxygen-evolving enhancer protein 1, chloroplastic
GFUP01001906.1	–	–	–	–	+	–	–	Guanylate cyclase domain-containing protein
GFUP01033041.1	–	–	–	–	+	–	–	SBP-type domain-containing protein
GFUP01038976.1	–	–	–	–	+	–	–	Aquaporin, glycerol transport
GFUP01041422.1	–	–	–	–	+	–	–	Uncharacterised protein
GFUP01043568.1	–	–	–	–	–	+	–	Uncharacterised protein
GFUP01049969.1	–	–	–	–	–	+	–	C3H1-type domain-containing protein
GFUP01050656.1	–	–	–	–	–	+	–	Uncharacterised protein
GFUP01022116.1	–	–	–	–	–	–	+	Uncharacterised protein
GFUP01023025.1	–	–	–	–	–	–	+	Uncharacterised protein
GFUP01025502.1	–	–	–	–	–	–	+	CULLIN_2 domain-containing protein
GFUP01033829.1	–	–	–	–	–	–	+	Uncharacterised protein
GFUP01046902.1	–	–	–	–	–	–	+	Uncharacterised protein
GFUP01047096.1	–	–	–	–	–	–	+	Uncharacterised protein

and photosynthetic activity is decreasing in N-starving cells [49,50] in a controlled and orderly manner, even under phototrophic conditions [51]. Downregulation of genes related to the photosynthetic apparatus (Light-harvesting proteins of PSI and PSII, oxygen-evolving enhancer protein 1 of PSII) lead to a decrease in C assimilation, and resulting in a complex metabolic and physiological response in which both N and C availabilities are altered [51]. In accordance with the readjustment of the photosynthetic apparatus, genes related to glycolysis (phosphoglycerate kinase and glycerate transporter) are downregulated. Very interestingly, guanylate cyclase gene was found to be extensively regulated by different N sources in *Chlamydomonas reinhardtii* [52].

Three out of 28 universal N-repletion downregulated genes, an uncharacterised protein, lipase_3 domain-containing protein and protein kinase domain-containing protein were consistently repressed in N-starving MACC-401 cultures. Notably, a class 3 lipase domain-containing protein gene has been downregulated in MACC-401, presumably to compensate the decrease in lipogenic activity to allow an effective lipid accumulation [50]. Lipases, which catalyse the release of free FAs from lipids, are ideal targets to increase lipid accumulation. The knockdown of a specific lipase from *Thalassiosira pseudonana* resulted in a notable lipid accumulation without compromising growth characteristics [53]. A protein kinase domain-containing gene was remarkably downregulated in MACC-401. Protein kinases are extensively studied signaling mediators and modulators that are involved in numerous physiological processes, including the metabolic network shifts following nitrogen deprivation [54], for example, in *Phaeodactylum tricornutum* [55] or *Chlamydomonas reinhardtii* [56]. In the 2Xc1 mutant of *Tisochrysis lutea*, a steady repression of a set of calcium-dependent protein kinases are suspected to be responsible for the constitutive lipid accumulation [57].

33 upregulated genes were shown to share a similar expression pattern with N-repleted *Scenedesmus acutus* cultures (Table 4). Two of the 33 universal N-repletion upregulated genes, an unknown protein and the urea active transporter which were induced in N-starving MACC-401 cultures as well as in the *D. tertiolecta*, *M. neglectum*, *N. oceanica*, *N. oleoabundans* strains which points out it may have a key role across

Table 4
List of overlapping, upregulated genes between the *Scenedesmus acutus* reference transcriptome (N-starving cells), MACC-401 and six other transcriptomes of N-depleted, lipid producing microalgal cell cultures. Accessions which have the corresponding orthologous gene in the reference dataset, are marked with „+"; „-" represents no hit.

Subject (GenBank ID)	*Chlamydomonas reinhardtii*	MACC-401	*Dunaliella tertiolecta*	*Monoraphidium neglectum*	*Nannochloropsis oceanica*	*Neochloris oleoabundans*	*Synechococcus elongatus*	Annotation (product)
GFUP01007317.1	+	−	+	+	+	+	−	Urea carboxylase/ allophanate hydrolase
GFUP01007324.1	+	−	−	−	−	+	−	Ammonium transporter
GFUP01010145.1	+	−	−	+	−	+	−	Nitrate/nitrite transporter
GFUP01017121.1	+	−	−	+	+	+	−	Ferredoxin-dependent glutamate synthase
GFUP01028335.1	+	−	+	+	−	+	−	Nitrate reductase
GFUP01028335.1	+	−	+	+	−	+	−	Nitrate reductase
GFUP01037273.1	+	−	−	+	−	+	−	Nitrite reductase
GFUP01044612.1	+	−	−	−	−	−	−	Allantoinase
GFUP01008482.1	−	+	−	+	−	−	−	Uncharacterised protein
GFUP01013467.1	−	+	+	+	+	+	−	Urea active transporter
GFUP01000159.1	−	−	+	−	−	+	−	Ribosome biogenesis protein WDR12 homolog
GFUP01000299.1	−	−	+	−	−	−	−	Fe2OG dioxygenase domain-containing protein
GFUP01012270.1	−	−	+	−	+	+	−	CN hydrolase domain-containing protein
GFUP01012909.1	−	−	+	−	+	+	−	Protein kinase domain-containing protein
GFUP01017336.1	−	−	−	−	+	−	−	Cellulase domain-containing protein
GFUP01026601.1	−	−	+	−	−	−	−	Uncharacterised protein
GFUP01027041.1	−	−	+	+	+	+	−	Xanthine/uracil/ vitamin C permease-like
GFUP01027072.1	−	−	+	−	−	−	−	Nitrate reductase
GFUP01032876.1	−	−	−	−	+	−	−	Cyclopropane-fatty-acyl-phospholipid synthase
GFUP01039670.1	−	−	+	−	−	−	−	Uncharacterised protein
GFUP01008640.1	−	−	−	−	+	−	−	Uncharacterised protein
GFUP01008962.1	−	−	−	−	+	−	−	Protein DETOXIFICATION
GFUP01019333.1	−	−	−	−	+	−	−	Uncharacterised protein
GFUP01022454.1	−	−	−	−	+	−	−	Uncharacterised protein
GFUP01023441.1	−	−	−	−	+	−	−	Uncharacterised protein
GFUP01000174.1	−	−	−	−	−	+	−	ComE operon protein 3
GFUP01000174.1	−	−	−	−	−	+	−	ComE operon protein 3
GFUP01002328.1	−	−	−	−	−	+	−	Agmatine deiminase
GFUP01002328.1	−	−	−	−	−	+	−	Agmatine deiminase
GFUP01002628.1	−	−	−	−	−	+	−	Agmatine deiminase
GFUP01046611.1	−	−	−	−	−	+	−	Uncharacterized protein
GFUP01050342.1	−	−	−	−	−	+	−	Unsaturated glucuronyl hydrolase
GFUP01050342.1	−	−	−	−	−	+	−	Unsaturated glucuronyl hydrolase

the species. Notably, this gene set predominantly consists of known N-related genes which are mainly involved in mobilising N reserves and readjustment of N metabolism. Under N-depleted conditions, adapting to N starvation involves short-term N salvage from the catabolism of amino acids and/or nucleic acids to maintain basic cell functions, including the regulation of lipid metabolism [58]. Allantoinase and agmatine deiminase are key genes of the metabolism of heterocyclic nitrogen compounds such as polyamines, purines and their degradation products [59]. Allantoinase gene was shown to be a good marker to monitor N status in rice [60] as well as in nitrogen deprived *C. reinhardtii* cultures [13]. Glutamate synthase, a key enzyme of N assimilation and a part of the Glu-Gln (GS/GOGAT) system was also shown to be upregulated in N-starving *C. reinhardtii* cells [50] or Chlorella strains [58]. Nitrate and nitrite reductases are well-known key enzymes in nitrogen assimilation [61]. Nitrate reductase activity increased substantially in N-starving *Chlorella vulgaris* [62]. Low and high affinity transporters, such as ammonia, nitrate/nitrite and urea transporters are responsible for N-uptake and intracellular allocation of N and induced under N-starvation regime [50]. In N-starving MACC-401, a urea active transporter gene was remarkably upregulated. In *C. reinhardtii*, urea assimilation is notably preferred when other nitrogen sources are limited [13]. Hence, urea active transporter genes, such as DUR3A are upregulated to various extents in N-deprived *C. reinhardtii* cultures, likely to facilitate urea availability [13]. A similar expression pattern has been reported from several other N-starving cultures of microalgae species *Prorocentrum donghaiense* [63] or *Prymnesium parvum* [64], indicating that urea transporters are associated with processing extracellular nitrogen or reallocating intracellular nitrogen when N status is perturbed.

It can be concluded that MACC-401 transcriptome shares a limited overlap with the "universal" N-starvation response genes from seven related datasets. This might be either due to that our culturing conditions and sampling periods differ from the referenced experimental conditions, or a low similarity between MACC-401 and *Scenedesmus* transcripts could be another reason why a low number of orthologues have been found. Alternatively, MACC-401 uses diverse gene sets to mitigate N-deprivation and to promote lipid accumulation. Clearly, a class 3 lipase gene is repressed, supposedly to decrease lipolytic activity in the highly N-limiting environment. On the contrast, a urea active transporter gene is upregulated to facilitate the uptake or allocation of this nitrogen source. As the urea active transport gene has been reported in six datasets as N-starvation responsive gene, it holds a potential to be a good marker to follow lipid accumulation under N-limited conditions to assess N status [63]. Finally, we identified a core set of N – starvation related genes which might represent an evolutionally conservative response across different algal species. A deeper functional analysis of these candidate genes in the future would definitely reveal their importance in N – deprivation response.

4. Conclusions

In this study, the physiological characteristics, the genome and transcriptome, as well as the phylogenetic relationships of MACC-401 microalgae isolate have been investigated. We found that this strain displayed a remarkably high production of daily biomass, lipid content and fatty acid composition upon N-starvation. The phylogenetic analysis suggested that MACC-401 likely represents a genetically distinct accession in the Scenedesmaceae family. Finally, by a comparative analysis of N-starvation related green algae transcriptomes, we identified genes which eventually could serve as markers to monitor lipid production in N-starving algal cultures.

CRediT authorship contribution statement

Vilmos Soós: Conceptualization, Methodology, Investigation, Formal Analysis, Data Curation, Writing – Original Draft, Project Administration, Supervision **Prateek Shetty:** Software, Formal analysis, Resources, Writing – Original Draft **Gergely Maróti:** Software, Formal analysis, Resources, Writing – Original Draft **Norbert Incze:** Investigation, Formal Analysis, Writing – Original Draft **Eszter Badics:** Investigation, Formal Analysis, Writing – Original Draft **Péter Bálint:** Investigation, Formal Analysis **Vince Ördög:** Conceptualization, Supervision, Funding Acquisition **Ervin Balázs:** Conceptualization, Supervision, Funding Acquisition

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the TÁMOP-4.2.2.A-11/1/KONV-2012-0003 project, GM received support from the Hungarian NKFIH (FK123899) and through the Lendület-Programme of the Hungarian Academy of Sciences (LP2020-5/2020).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.algal.2020.102181.

References

[1] J. Valentine, J. Clifton-Brown, A. Hastings, P. Robson, G. Allison, P. Smith, Food vs. fuel: The use of land for lignocellulosic "next generation" energy crops that minimize competition with primary food production, GCB Bioenergy. 4 (2012) 1–19, https://doi.org/10.1111/j.1757-1707.2011.01111.x.
[2] S. Behera, R. Singh, R. Arora, N.K. Sharma, M. Shukla, S. Kumar, Scope of algae as third generation biofuels, Frontiers in Bioengineering and Biotechnology. 2 (2015) 90, https://doi.org/10.3389/fbioe.2014.00090.
[3] E. Stephens, I.L. Ross, J.H. Mussgnug, L.D. Wagner, M.A. Borowitzka, C. Posten, O. Kruse, B. Hankamer, Future prospects of microalgal biofuel production systems, Trends Plant Sci. 15 (2010) 554–564, https://doi.org/10.1016/j.tplants.2010.06.003.
[4] M. Hannon, J. Gimpel, M. Tran, B. Rasala, S. Mayfield, Biofuels from algae: challenges and potential, Biofuels. 1 (2010) 763–784, https://doi.org/10.4155/bfs.10.44.
[5] M.I. Khan, J.H. Shin, J.D. Kim, The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products, Microb. Cell Factories 17 (2018) 36, https://doi.org/10.1186/s12934-018-0879-x.
[6] G.A. Thompson, Lipids and membrane function in green algae, Biochimica et Biophysica Acta - Lipids and Lipid Metabolism. 1302 (1996) 17–45, https://doi.org/10.1016/0005-2760(96)00045-8.
[7] K.K. Sharma, H. Schuhmann, P.M. Schenk, High lipid induction in microalgae for biodiesel production, Energies. 5 (2012) 1532–1553, https://doi.org/10.3390/en5051532.
[8] R. Radakovits, R.E. Jinkerson, A. Darzins, M.C. Posewitz, Genetic engineering of algae for enhanced biofuel production, Eukaryot. Cell 9 (2010) 486–501, https://doi.org/10.1128/EC.00364-09.
[9] V.C. Akubude, K.N. Nwaigwe, E. Dintwa, Production of biodiesel from microalgae via nanocatalyzed transesterification process: a review, Materials Science for Energy Technologies. 2 (2019) 216–225, https://doi.org/10.1016/j.mset.2018.12.006.
[10] S.E. Shin, H.G. Koh, N.K. Kang, W.I. Suh, B.R. Jeong, B. Lee, Y.K. Chang, Isolation, phenotypic characterization and genome wide analysis of a Chlamydomonas reinhardtii strain naturally modified under laboratory conditions: towards enhanced microalgal biomass and lipid production for biofuels, Biotechnology for Biofuels. 10 (2017) 1–15, https://doi.org/10.1186/s13068-017-1000-0.
[11] L.D. Zhu, Z.H. Li, E. Hiltunen, Strategies for lipid production improvement in microalgae as a biodiesel feedstock, Biomed. Res. Int. 2016 (2016), https://doi.org/10.1155/2016/8792548.
[12] K.W.M. Tan, Y.K. Lee, The dilemma for lipid productivity in green microalgae: importance of substrate provision in improving oil yield without sacrificing growth, Biotechnology for Biofuels. 9 (2016) 255, https://doi.org/10.1186/s13068-016-0671-2.
[13] J.J. Park, H. Wang, M. Gargouri, R.R. Deshpande, J.N. Skepper, F.O. Holguin, M. T. Juergens, Y. Shachar-Hill, L.M. Hicks, D.R. Gang, The response of Chlamydomonas reinhardtii to nitrogen deprivation: a systems biology analysis, Plant J. 81 (2015) 611–624, https://doi.org/10.1111/tpj.12747.
[14] A.L.G. de Lomana, S. Schäuble, J. Valenzuela, S. Imam, W. Carter, D.D. Bilgin, C. B. Yohn, S. Turkarslan, D.J. Reiss, M.V. Orellana, N.D. Price, N.S. Baliga,

Transcriptional program for nitrogen starvation-induced lipid accumulation in Chlamydomonas reinhardtii, Biotechnology for Biofuels. 8 (2015) 1–18, https://doi.org/10.1186/s13068-015-0391-z.
[15] A.M. Illman, A.H. Scragg, S.W. Shales, Increase in Chlorella strains calorific values when grown in low nitrogen medium, Enzym. Microb. Technol. 27 (2000) 631–635, https://doi.org/10.1016/S0141-0229(00)00266-0.
[16] V. Ördög, W.A. Stirk, P. Bálint, C. Lovász, O. Pulz, J. van Staden, Lipid productivity and fatty acid composition in Chlorella and Scenepdesmus strains grown in nitrogen-stressed conditions, J. Appl. Phycol. 25 (2013) 233–243, https://doi.org/10.1007/s10811-012-9857-6.
[17] L. Rodolfi, G.C. Zittelli, N. Bassi, G. Padovani, N. Biondi, G. Bonini, M.R. Tredici, Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor, Biotechnol. Bioeng. 102 (2009) 100–112, https://doi.org/10.1002/bit.22033.
[18] A.O. Aremu, M. Neményi, W.A. Stirk, V. Ördög, J. van Staden, Manipulation of nitrogen levels and mode of cultivation are viable methods to improve the lipid, fatty acids, phytochemical content, and bioactivities in Chlorella minutissima, J. Phycol. 51 (2015) 659–669, https://doi.org/10.1111/jpy.12308.
[19] N. Hempel, I. Petrick, F. Behrendt, Biomass productivity and productivity of fatty acids and amino acids of microalgae strains as key characteristics of suitability for biodiesel production, J. Appl. Phycol. 24 (2012) 1407–1418, https://doi.org/10.1007/s10811-012-9795-3.
[20] D.E. Kuznjecov, M.G. Vladimirova, Iron as a limiting factor of Chlorella growth in Tamiya nutrient medium, Plant Physiol. 11 (1964) 615–619.
[21] V. Ördög, Apparatus for laboratory algal bioassay, Int. Rev. Gesamten Hydrobiol. 67 (1982) 127–136.
[22] J.J. Doyle, J.L. Doyle, A rapid DNA isolation procedure for small quantities of fresh leaf tissue, Phytochemical Bulletin. 19 (1987) 11–15. https://worldveg.tind.io/record/33886 (accessed August 26, 2020).
[23] M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging, BMC Research Notes. 9 (2016) 88, https://doi.org/10.1186/s13104-016-1900-2.
[24] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V. M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, https://doi.org/10.1089/cmb.2012.0021.
[27] R.M. Waterhouse, M. Seppey, F.A. Simao, M. Manni, P. Ioannidis, G. Klioutchnikov, E.V. Kriventseva, E.M. Zdobnov, BUSCO applications from quality assessments to gene prediction and phylogenomics, Mol. Biol. Evol. 35 (2018) 543–548, https://doi.org/10.1093/molbev/msx319.
[28] N. Dierckxsens, P. Mardulyn, G. Smits, NOVOPlasty: De novo assembly of organelle genomes from whole genome data, Nucleic Acids Research. 45 (2017), https://doi.org/10.1093/nar/gkw955.
[29] M. Tillich, P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, S. Greiner, GeSeq - versatile and accurate annotation of organelle genomes, Nucleic Acids Res. 45 (2017) W6–W11, https://doi.org/10.1093/nar/gkx391.
[30] L.T. Nguyen, H.A. Schmidt, A. Von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274, https://doi.org/10.1093/molbev/msu300.
[31] A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models, Bioinformatics. 22 (2006) 2688–2690, https://doi.org/10.1093/bioinformatics/btl446.
[32] I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, Nucleic Acids Res. 44 (2016) W242–W245, https://doi.org/10.1093/nar/gkw290.
[33] L. Song, L. Florea, Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads, GigaScience. 4 (2015) 48, https://doi.org/10.1186/s13742-015-0089-y.
[34] F. Krueger, Trim Galore. A Wrapper Tool Around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files, 2015, pp. 516–517.
[35] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. 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) 644–652, https://doi.org/10.1038/nbt.1883.
[36] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics. 22 (2006) 1658–1659, https://doi.org/10.1093/bioinformatics/btl158.
[37] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (2019) 907–915, https://doi.org/10.1038/s41587-019-0201-4.
[38] V. Hower, R. Starfield, A. Roberts, L. Pachter, Quantifying uniformity of mapped reads, Bioinformatics. 28 (2012) 2680–2682, https://doi.org/10.1093/bioinformatics/bts451.
[39] RStudio Team, RStudio: integrated Development for R., RStudio, Inc., Boston, MA. vol. 42, 2015, p. 14. http://www.rstudio.com (accessed September 24, 2020).
[40] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics. 26 (2009) 139–140, https://doi.org/10.1093/bioinformatics/btp616.
[41] A. Sirikhachornkit, A. Suttangkakul, S. Vuttipongchaikij, P. Juntawong, De novo transcriptome analysis and gene expression profiling of an oleaginous microalga Scenedesmus acutus TISTR8540 during nitrogen deprivation-induced lipid accumulation, Sci. Rep. 8 (2018) 1–12, https://doi.org/10.1038/s41598-018-22080-8.
[42] I.V. Grigoriev, R.D. Hayes, S. Calhoun, B. Kamel, A. Wang, S. Ahrendt, S. Dusheyko, R. Nikitin, S.J. Mondo, A. Salamov, I. Shabalov, A. Kuo, Lawrence Berkeley National Laboratory Recent Work Title PhycoCosm, a comparative algal genomics resource, Permalink, https://escholarship.org/uc/item/5v63934m.
[43] E.C. Goncalves, J.V. Johnson, B. Rathinasabapathi, Conversion of membrane lipid acyl groups to triacylglycerol and formation of lipid bodies upon nitrogen starvation in biofuel green algae Chlorella UTEX29, Planta. 238 (2013) 895–906, https://doi.org/10.1007/s00425-013-1946-5.
[44] D. Jaeger, A. Winkler, J.H. Mussgnug, J. Kalinowski, A. Goesmann, O. Kruse, Time-resolved transcriptome analysis and lipid pathway reconstruction of the oleaginous green microalga Monoraphidium neglectum reveal a model for triacylglycerol and lipid hyperaccumulation, Biotechnology for Biofuels. 10 (2017) 1–34, https://doi.org/10.1186/s13068-017-0882-1.
[45] S.Y. Choi, B. Park, I.G. Choi, S.J. Sim, S.M. Lee, Y. Um, H.M. Woo, Transcriptome landscape of Synechococcus elongatus PCC 7942 for nitrogen starvation responses using RNA-seq, Sci. Rep. 6 (2016) 1–10, https://doi.org/10.1038/srep30584.
[46] J. Li, D. Han, D. Wang, K. Ning, J. Jia, L. Wei, X. Jing, S. Huang, J. Chen, Y. Li, Q. Hu, J. Xu, Choreography of transcriptomes and lipidomes of Nannochloropsis reveals the mechanisms of oil synthesis in microalgae, Plant Cell 26 (2014) 1645–1665, https://doi.org/10.1105/tpc.113.121418.
[47] K.W.M. Tan, H. Lin, H. Shen, Y.K. Lee, Nitrogen-induced metabolic changes and molecular determinants of carbon allocation in Dunaliella tertiolecta, Sci. Rep. 6 (2016) 1–13, https://doi.org/10.1038/srep37235.
[48] H. Rismani-Yazdi, B.Z. Haznedaroglu, C. Hsin, J. Peccia, Transcriptomic analysis of the oleaginous microalga Neochloris oleoabundans reveals metabolic insights into triacylglyceride accumulation, Biotechnology for Biofuels. 5 (2012) 1–16, https://doi.org/10.1186/1754-6834-5-74.
[49] W. Majeran, F.-A. Wollman, O. Vallon, Evidence for a role of ClpP in the degradation of the chloroplast cytochrome b 6 f complex, Plant Cell 12 (2000) 137–149, https://doi.org/10.1105/tpc.12.1.137.
[50] L. Valledor, T. Furuhashi, L. Recuenco-Muñoz, S. Wienkoop, W. Weckwerth, System-level network analysis of nitrogen starvation and recovery in Chlamydomonas reinhardtii reveals potential new targets for increased lipid accumulation, Biotechnology for Biofuels. 7 (2014) 171, https://doi.org/10.1186/s13068-014-0171-1.
[51] M.T. Juergens, R.R. Deshpande, B.F. Lucker, J.J. Park, H. Wang, M. Gargouri, F. Omar Holguin, B. Disbrow, T. Schaub, J.N. Skepper, D.M. Kramer, D.R. Gang, L. M. Hicks, Y. Shachar-Hill, The regulation of photosynthetic structure and function during nitrogen deprivation in Chlamydomonas reinhardtii, Plant Physiol. 167 (2015) 558–573, https://doi.org/10.1104/pp.114.250530.
[52] A. de Montaigu, E. Sanz-Luque, A.G. Galván, E. Fernández, A soluble guanylate cyclase mediates negative signaling by ammonium on expression of nitrate reductase in Chlamydomonas, Plant Cell 22 (2010) 1532–1548, https://doi.org/10.1105/tpc.108.062380.
[53] E.M. Trentacoste, R.P. Shrestha, S.R. Smith, C. Glé, A.C. Hartmann, M. Hildebrand, W.H. Gerwick, Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 19748–19753, https://doi.org/10.1073/pnas.1309299110.
[54] T. Takeuchi, C. Benning, Nitrogen-dependent coordination of cell cycle, quiescence and TAG accumulation in Chlamydomonas, Biotechnology for Biofuels. 12 (2019) 1–20, https://doi.org/10.1186/s13068-019-1635-0.
[55] Z.K. Yang, Y.H. Ma, J.W. Zheng, W.D. Yang, J.S. Liu, H.Y. Li, Proteomics to reveal metabolic network shifts towards lipid accumulation following nitrogen deprivation in the diatom Phaeodactylum tricornutum, J. Appl. Phycol. 26 (2014) 73–82, https://doi.org/10.1007/s10811-013-0050-3.
[56] Y. Li, X. Fei, H. Dai, J. Li, W. Zhu, X. Deng, Genome-wide identification of calcium-dependent protein kinases in Chlamydomonas reinhardtii and functional analyses in nitrogen deficiency-induced oil accumulation, Front. Plant Sci. 10 (2019) 1147, https://doi.org/10.3389/fpls.2019.01147.
[57] M. Garnier, G. Bougaran, M. Pavlovic, J.B. Berard, G. Carrier, A. Charrier, F. Le Grand, E. Lukomska, C. Rouxel, N. Schreiber, J.P. Cadoret, H. Rogniaux, B. Saint-Jean, Use of a lipid rich strain reveals mechanisms of nitrogen limitation and carbon partitioning in the haptophyte Tisochrysis lutea, Algal Res. 20 (2016) 229–248, https://doi.org/10.1016/j.algal.2016.10.017.
[58] H. Chen, Y. Zheng, J. Zhan, C. He, Q. Wang, Comparative metabolic profiling of the lipid-producing green microalga Chlorella reveals that nitrogen and carbon metabolic pathways contribute to lipid metabolism, Biotechnology for Biofuels. 1 (2017) 1–20, https://doi.org/10.1186/s13068-017-0839-4.
[59] J. Yang, K.H. Han, Functional characterization of allantoinase genes from Arabidopsis and a nonureide-type legume black locust, Plant Physiol. 134 (2004) 1039–1049, https://doi.org/10.1104/pp.103.034637.
[60] D.-K. Lee, M.C.F.R. Redillas, H. Jung, S. Choi, Y.S. Kim, J.-K. Kim, A nitrogen molecular sensing system, comprised of the allantoinase and ureide permease 1 genes, can be used to monitor N status in rice, Front. Plant Sci. 9 (2018) 444, https://doi.org/10.3389/fpls.2018.00444.
[61] E. Sanz-Luque, A. Chamizo-Ampudia, A. Llamas, A. Galvan, E. Fernandez, Understanding nitrate assimilation and its regulation in microalgae, Front. Plant Sci. 6 (2015) 899, https://doi.org/10.3389/fpls.2015.00899.

[62] I. Morris, P.J. Syrett, The effect of nitrogen starvation on the activity of nitrate Reductase and other enzymes in Chlorella, J. Gen. Microbiol. 38 (1965) 21–28, https://doi.org/10.1099/00221287-38-1-21.
[63] X. Jing, S. Lin, H. Zhang, C. Koerting, Z. Yu, Utilization of urea and expression profiles of related genes in the dinoflagellate Prorocentrum donghaiense, PLoS One 12 (2017), e0187837, https://doi.org/10.1371/journal.pone.0187837.
[64] Z. Liu, A.E. Koid, R. Terrado, V. Campbell, D.A. Caron, K.B. Heidelberg, Changes in gene expression of Prymnesium parvum induced by nitrogen and phosphorus limitation, Front. Microbiol. 6 (2015) 631, https://doi.org/10.3389/fmicb.2015.00631.