Comparative transcriptomics reveals molecular components associated with differential lipid accumulation between microalgal sp., *Scenedesmus dimorphus* and *Scenedesmus quadricauda*

Tamanna Sharma, Rajinder Singh Chauhan *

Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan 173215, H.P., India

ABSTRACT

Scenedesmus is considered as a potential oil-producing green microalgae having higher lipid content with suitable fatty acid profile (high oleic acid) for biodiesel production. Comparative transcriptome analysis of two *Scenedesmus* sp., viz. *Scenedesmus dimorphus* (26%) and *Scenedesmus quadricauda* (14%) having equivalent biomass and variable lipid content was performed to uncover molecular mechanisms controlling differential lipid production. 76,969 and 40,979 CDSs were predicted from the transcriptomes of *S. dimorphus* and *S. quadricauda* respectively, which were subsequently mapped to metabolic pathways. Overall up-regulation of metabolic pathways contributing precursors to storage lipid biosynthesis was observed in *S. dimorphus*. Glyceraldehyde 3-phosphate dehydrogenase, enolase, acetyl-CoA synthetase, pyruvate dehydrogenase, ATP citrate lyase, glycerol kinase, citrate synthase were identified as major regulators of high lipid content in *S. dimorphus*. Further, WRINKLED1 transcription factor significantly correlated with high lipid accumulation as revealed by reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) in three growth stages of two species. This study provides a broad view of storage lipid production in *Scenedesmus* species with potential implications in designing suitable genetic interventions towards increase in lipid content vis-à-vis central carbon metabolism.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Biomass from different feedstocks is considered a major component for the production of biofuels such as biodiesel, bioethanol, biohydrogen, bio-oil, and biogas [1]. Each feedstock has its own advantages in terms of lipid content, biomass yield, fatty acid composition, and geographical distribution [2]. Microalgal feedstock has recently emerged as the only renewable approach to produce biofuels as it offers various advantages over other conventional sources for biofuel production, such as rapid growth rate, high lipid content and no competition with other agricultural crops for land [3]. Microalgal species with high biomass productivities and robust lipid yield are preferred for the production of biodiesel [4]. However, only few microalgal strains are available, which have the combination of traits of high biomass productivity

with considerable lipid content [5]. Lipid content of microalgae is species/strain dependent as different strains and species have shown variation for oil content and fatty acid composition. Usually, microalgae synthesize low lipids under favourable conditions where lipid content gets increased up to several folds under stress conditions [2]. There are various reports in which molecular mechanisms behind the increase of lipid content under stress conditions were examined. However, the mechanism behind interspecies variation among microalgae is unknown. As storage lipid biosynthesis is regulated by multiple pathways, transcriptomic analysis aids in deciphering the regulatory mechanisms leading to high lipid production in microalgae [6]. Transcriptomic analysis is a powerful approach for the molecular dissection of phenotypic traits in contrasting conditions [7]. A number of transcriptomes have been sequenced and annotated for the model as well as oleaginous microalgae, which provided the broad view of different metabolic pathways involved in biofuel production from microalgae. However, comparative transcriptome analysis is required to gain preliminary information about the differentially expressed pathways and genes, that can be targeted for metabolic engineering of microalgae to have desired strains with enhanced lipid production [8].

Transcriptome of oleaginous microalga *Neochloris oleoabundans* was assembled where quantitative gene expression analysis under nitrogen replete and nitrogen limiting conditions demonstrated the up-regulation of pathways leading to storage lipid accumulation, such as fatty

acid and TAG biosynthesis and the pentose phosphate pathway [6]. Also, differentially expressed genes under nitrogen depletion conditions were examined in the model organism *Chlamydomonas reinhardtii* using 454 and Illumina next-generation sequencing techniques which revealed repressed β-oxidation pathway and up-regulated TAG biosynthesis [9]. In another report, transcriptomes of *Dunaliella tertiolecta* under nitrogen repletion, nitrogen depletion, and high salinity conditions were analysed for identification of genes involved in starch and lipid biosynthesis [10]. Also, in model green microalga *Chlamydomonas reinhardtii*, time-course transcriptome sequencing was carried out during lipid biosynthesis and accumulation processes using Illumina RNA-seq platform, which subsequently revealed higher expression of triacylglycerol biosynthetic genes in lipid accumulating phase [11]. Further, differential transcriptome analysis of *Chlorella protothecoides* in contrasting growth conditions of nitrogen repletion and depletion showed up-regulation of genes in the lipid accumulating condition of nitrogen depletion [12]. In a recent study, transcriptome sequencing of halophilic green alga *Dunaliella parva* was performed in nitrogen limited and nitrogen sufficient conditions to identify important pathways and genes for biofuel production [13]. All of the previous transcriptome based studies were focused on the changes in metabolic flux due to stress conditions or during lipid accumulating growth phases, but no report exists as of today on the comparative transcriptomic analysis between species having contrasting lipid contents. It is, therefore, essential to reveal the pathways associated with high lipid content in microalgae.

Scenedesmus is a genus of algae belonging to Chlorophyceae class, which has gained attention for its use in biodiesel production due to its high biomass productivity with high lipid content and desired fatty acid profile among green algae [14–16]. Various studies have reported the optimization of biomass productivity as well as lipid content by varying the concentration of supplemental nutrients. *Scenedesmus* sp. has shown highest biomass (4 mg/ml) and lipid content (49%) in comparison to *Botryococcus* and *Chlorella* [17]. Also the fatty acid composition of *Scenedesmus* was found to be suitable with high percentage of oleic acid, which is a prerequisite for biodiesel production [18].

Variation in the lipid content has been reported among different species of *Scenedesmus* [19]. Different isolates of two *Scenedesmus* species, viz. *Scenedesmus dimorphus* and *Scenedesmus quadricauda* were collected from various geographical locations of Himachal Pradesh, India [20]. Also, it has been reported that an increase in the lipid content under stress conditions in microalgae is due to the reduction of biomass of the cell, rather than enhanced lipid accumulation [21]. In the current study, two isolates of *Scenedesmus* species, viz. *S. dimorphus* (26% lipid content) and *S. quadricauda* (14% lipid content) with equivalent biomass yields were taken. Fatty acid profiling of these two species revealed higher proportion of oleic acid in *S. dimorphus* [22]. Transcriptomes of the two species at lipid accumulating phase were sequenced and assembled de novo and further comparative analysis of the gene ontology categories, pathways and genes was performed. The present study not only provides molecular basis for the inter-species variation of lipid content, but also presents novel insights into metabolism and transcriptional regulation leading to differential lipid accumulation, which provided the molecular cues to enhance and regulate lipid content in microalgae through metabolic engineering approaches.

2. Materials and methods

2.1. Culturing of Scenedesmus species

Cultures of *S. dimorphus* and *S. quadricauda* were maintained in the glass house of the Jaypee University of Information Technology, Waknaghat, H.P., India, under optimal conditions of temperature and light and grown in the BG11 media. Axenic cultures of both the species were obtained by treating with antibiotics: penicillin, streptomycin and chloramphenicol following the protocol by Guillard [23]. These cultures were incubated in 250 ml flasks containing 100 ml BG11 media, at a

light intensity of 42 μmol photons m^{-2} s^{-1} with a diurnal cycle of 16 h light and 8 h dark at temperature of 25 ± 2 °C. Cultures were continuously aerated with 0.22 μm filtered air through a mechanical pump. Cultures were harvested after 20 days and stored at −80 °C for further use.

2.2. Estimation of lipid content and biomass in S. dimorphus and S. quadricauda

Lipid extraction from microalgal strains was carried out as described by Bligh and Dyer with minor modifications [24]. Equal amount of biomass was freeze-dried using lyophilizer and the lyophilized cells were suspended in chloroform: methanol solution (1:2) and vortexed immediately. The mixture was sonicated for 5 min and kept on a shaker overnight. Next day, equal amount of chloroform: distilled water (1:1) was added to the mixture, vortexed and then centrifuged at 6000 ×*g* for 10 min. Lipids are soluble in chloroform, hence form a dense layer at the bottom of the centrifuge tube, while methanol and water create a uniform top layer, and cell debris creates a middle layer. Lipid with chloroform was taken with the help of micropipette by applying gentle, positive pressure and passed through a 2.5 cm thick layer of anhydrous sodium sulfate using Whatman filter paper in a funnel into a pre-weighed container suitable for rotary evaporation. The solvent was removed using a rotary evaporator under reduced pressure at 60 °C and weight of remaining lipids was recorded. Total lipid content was calculated as percentage of the total biomass (in % dry weight). Fatty acid methyl esters (FAMEs) were estimated for the determination of triacylglycerol [22].

Microalgal cells of both the species were stained with the fluorescent dye, BODIPY 505/515 (Invitrogen) according to the protocol described by Cooper et al. [25]. Fluorescence analysis of the green BODIPY stained micoalgal cells was determined by fluorescence microscope (Olympus BX53).

2.3. RNA extraction

Total RNA was isolated using RaFlex RNA isolation kit (GeNei™) following manufacturer's instructions. The quality of RNA was checked in 1% (w/v) ethidium bromide-stained denaturing agarose gel at 100 V for 30 mins. Further, total RNA was quantified through the absorbance spectrum at wavelengths 260 nm and 280 nm using NanoDrop 8000 spectrophotometer (Thermo scientific).

2.4. Library preparation for transcriptome analysis

Paired-end cDNA sequencing libraries were prepared using IlluminaTruSeq stranded total RNA library preparation kit according to the protocol described by the manufacturer. rRNA was depleted from total RNA followed by fragmentation. The fragmented rRNA depleted RNA was converted into first-strand cDNA, followed by second-strand cDNA synthesis, A-tailing, adapter ligation and finally ended by index PCR amplification of adaptor-ligated library. Library quantification and validation was performed using Qubit dsDNA HS kit and High Sensitivity Assay Kit, respectively.

2.5. De novo transcriptome assembly of S. dimorphus and S. quadricauda

The raw data of two microalgal species were generated on NextSeq. The raw reads generated were filtered using Trimmomatic (v 0.30) with quality value QV > 20 and other contaminants, such as adapters were also trimmed. Parameters considered for filtration were adapter trimming, sliding window, leading and trailing. Minimum length was taken as 40 bp.

QC passed reads of the two species were subjected to de novo assembly with Trinity and final assembled transcripts were generated. CD-HIT-EST was run on the assembled transcripts to get unigenes.

CDS were predicted from these unigenes using ORFPredictor (http://proteomics.ysu.edu/tools/OrfPredictor.html) with default parameters.

2.6. Functional annotation of predicted CDSs

Functional annotation of predicted CDSs from both the species, i.e. *S. dimorphus* and *S. quadricauda* was performed by aligning CDSs to non-redundant database of NCBI using BLASTx with E-value cut-off of 10^{-6} for both the species.

2.7. GO mapping and CDS distribution

GO assignments were used to classify the functions of predicted CDSs. GO mapping provides the ontology of defined terms representing gene product properties, which are grouped into three main domains: biological process, molecular function and cellular component. BLASTx result accession IDs were searched directly in the gene product table of GO database. Gene ontology analysis helps in specifying all the annotated nodes comprising of GO functional groups such as cellular component, biological process and molecular function.

2.8. Pathway analysis

KAAS was used to functionally annotate the CDSs of *S. dimorphus* and *S. quadricauda* by BLAST comparisons against KEGG GENES database. The BBH (bi-directional best hit) option was used to assign KO terms. The KEGG Orthology database (http://www.genome.jp/kegg/ko.html) was used for pathway mapping.

2.9. Transcript abundance estimation

Quantification of transcript abundance for the de novo assembled sequences was carried out using RSEM approach, which assesses the transcript abundances based on the mapping of RNA-Seq reads to the assembled transcriptome. RSEM performs transcript abundance estimation in two steps. Firstly, a set of reference transcript sequences was generated and preprocessed for the second step. Second, a set of RNA-Seq reads was aligned to the reference transcripts and the resulting alignments were used to estimate abundances and their credibility intervals [26]. Command rsem-prepare-reference was used to prepare the reference sequences and rsem-calculate-expression was used to calculate the transcript abundances as TPM and FPKM values, which is a sensitive approach to detect the expression level and measure expression of even poorly expressed transcripts using fragment count.

2.10. Identification of SSRs from transcriptome data

SSRs were identified using Perl script MISA (http://pgrc.ipkgatersleben.de/misa/misa.html). The frequency of SSRs according to the size and type of constituting SSRs were determined. The minimum repeat unit was defined as ten for mononucleotides, six for dinucleotides and five for all the higher order motifs including tri-, tetra-, penta-nucleotides.

2.11. Identification of transcription factors from transcriptomes of S. dimorphus and S. quadricauda

Identified CDSs of *S. dimorphus* and *S. quadricauda* were subjected to BLAST search against PlantTFDB (http://planttfdb.cbi.pku.edu.cn) with E-value cut-off of 10^{-5}. The conserved domains in transcription factors were identified by using conserved domain database available at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Only the top bit-scoring significant hit for each sequence was considered. Further, list of lipid related transcription factors from previous reports was used for the identification of lipid related transcription factors in *S. dimorphus* and *S. quadricauda*.

2.12. Gene expression validation through RT-qPCR

To study gene expression, RNA verification was done by measuring the amount of RNA with a spectrophotometer. First-strand cDNA synthesis was done using Verso cDNA synthesis kit (Thermo Scientific, USA) from total RNA (2 μg) template as per manufacturer's instructions. RT-qPCR was performed using gene specific primers in triplicate on a RT-qPCR system (CFX96, Bio-Rad Laboratories; Hercules CA) with the iScript RT PCR kit (Bio-rad). The PCR protocol was as follows: denaturation for 5 min at 94 °C, followed by 35 cycles each of denaturation for 20s at 94 °C, annealing for 30s at 50–55 °C, followed by one elongation step for 20s at 72 °C.

18S rRNA and actin [27,28] were used as an internal reference to normalize the expression data as described by Vandesompele [29]. For comparative analysis between *Scenedesmus* species, relative fold changes were determined from Cq values obtained in *S. dimorphus* and *S. quadricauda* samples in stationary phase. While for calculating relative fold expression of CDS encoding WRI1 transcription factor in three growth stages, exponential phase (EXP) having minimum lipid accumulation was kept as a calibrator condition and relative fold changes in other two stages were determined with respect to EXP from Cq values using the comparative Ct (ΔΔCt) method as described by Schmittgen and Livak [30]. All experiments were performed in triplicates and repeated twice. Standard error was applied for expression values.

2.13. Heat map analysis

A complete linkage hierarchical cluster analysis was performed on top 100 differentially expressed genes using multiple experiment viewer (MEV v4.9). Heat map was constructed using log-transformed and normalized value of genes based on Pearson's uncentered correlation distance as well as based on the complete linkage method. Top 50 up-regulated and 50 down- regulated genes were considered for the analysis.

3. Results and discussion

Two species of *Scenedesmus* viz. *Scenedesmus dimorphus* and *Scenedesmus quadricauda*, having maximum lipid content of $26 \pm 1.3\%$ and $14 \pm 0.7\%$ respectively, were selected for transcriptomic analysis to dissect the molecular basis of cross species lipid content variation (Fig. 1). Microalgal growth has been classified into different developmental phases, viz. lag phase, exponential phase and stationary phase. When a microalgal culture is transferred to a fresh media, it starts adapting according to the new environmental conditions, which is considered as lag phase. Next is exponential phase, in which cell density of the culture increases as a function of time. When microalgal growth ceases and remains to be constant, it enters stationary phase [31]. Optical density of cultures of both the species started to increase after second day of incubation and reached maximum at 18th day and then became constant (Fig. 2). Cultures of both the species were harvested at 20th day for further analysis.

3.1. De novo sequence assembly

RNA-Seq libraries were prepared using Illumina sequencing. The mean size of the fragment distribution ranged from 550 to 700 bp. After quality filtering of raw reads, 17,907,034 and 11,290,041 high quality (HQ) reads were obtained from *S. dimorphus* and *S. quadricauda* respectively. These HQ reads were subjected to de novo assembly with Trinity resulting in 60,826 transcripts for *S. dimorphus* and 145,112 transcripts for *S. quadricauda* (Table 1). CD-HIT-EST was applied on the trinity assembled transcripts, which resulted in 42,979 unigenes in *S. dimorphus* generating 38 Mb data (N50 = 1148) and 78,386 unigenes for *S. quadricauda,* generating 110 Mb data (N50 = 1736).
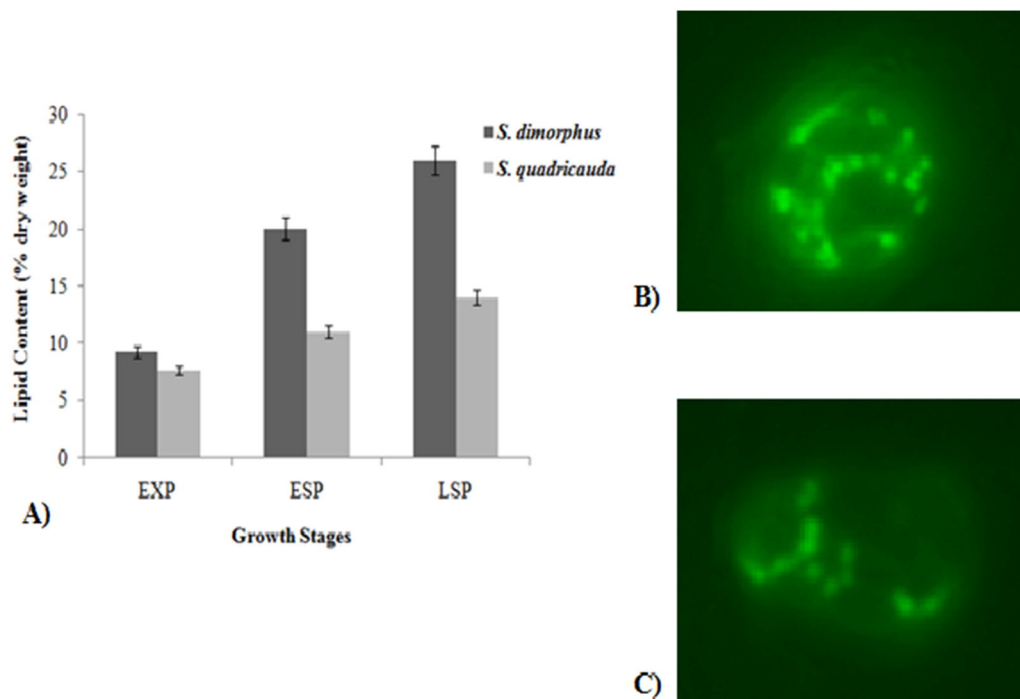
Fig. 1. A) Lipid content of *S. dimorphus* and *S. quadricauda* in three growth stages i.e. EXP, ESP, LSP [31]. Error bars indicate 5% standard error. *N* = 3 microalgal cultures. Abbreviations: EXP-Exponential Phase; ESP-Early Stationary Phase; LSP- Late Stationary Phase. B) Fluorescence microscopy image of *S. dimorphus*. C) Fluorescence microscopy image of *S. quadricauda*. The brightness indicate amount of triacylglycerol content.

Further, CDSs were predicted from these unigenes using ORFPredictor. Parameters of ORFPredictor were set as default. 40,979 and 76,969 unigenes were successfully mapped into CDSs in *S. dimorphus* and *S. quadricauda* respectively. The size of CDSs ranged from 200 to 6000 bp in both the species, with an average length of 300–400 bp in *S. dimorphus* and 400–500 bp in *S. quadricauda*. The size distribution of CDSs in both the species is shown in Fig. 3.

3.2. Functional annotation of predicted CDSs

Annotation of predicted CDSs in both the species was performed by aligning them with non-redundant (nr) protein database of NCBI using BLASTx algorithm with an E-value cut-off of 10^{-6}. Best hits of the database were taken as the sequence annotations. Total of 16,678 CDSs showed significant hits in *S. dimorphus*, out of which 10,194 CDSs were annotated with green algal species. In case of *S. quadricauda*, 11,917 CDSs were considered as the best hits, out of which 11,917 CDSs were annotated with green algal species. It was found that the majority of significant hits were contributed from *Volvox carteri* followed by *Chlamydomonas reinhardtii* and *Chlorella variabilis* in both the species (Fig. 4).

Further, GO assignments were used to classify the functions of predicted CDSs. GO mapping provides ontology of defined terms representing gene product properties which are grouped into three main domains, i.e. biological process, molecular function and cellular component. Accession IDs from BLASTx annotation were directly searched in the gene product table of GO database. The analysis resulted in majority of assignments in molecular function category in both the species (6551 in *S. dimorphus*; 7591 in *S. quadricauda*) followed by biological process (6123 in *S. dimorphus*; 6090 in *S. quadricauda*) and cellular component (3839 in *S. dimorphus*; 4430 in *S. quadricauda*) (Fig. 5).

3.3. Identification of SSRs from transcriptome data

SSRs are small arrays of tandem repeats of one to six nucleotides, which are widely detected in both mRNAs and genomes. SSRs are extremely variable as the numbers of repeat units in an array are different in different species. SSRs are locus-specific, co-dominant, PCR-based and highly polymorphic; hence they act as powerful genetic markers [32]. SSRs were mined from the transcriptomes of *S. dimorphus* and *S. quadricauda* with di- to penta-nucleotide motifs having minimum repeat unit as ten for mononucleotides, six for dinucleotides and five for
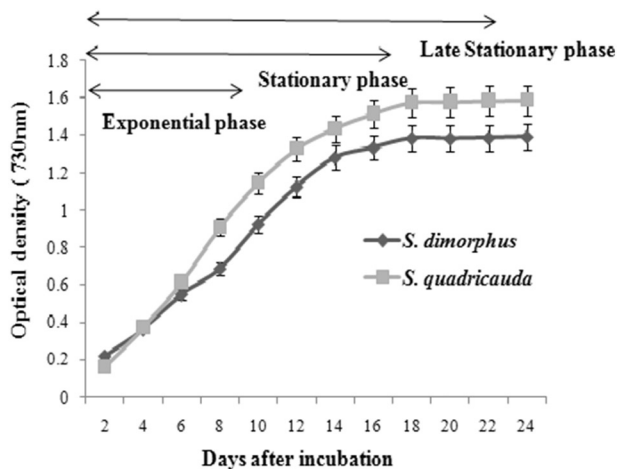

Fig. 2. Biomass growth of *S. dimorphus* and *S. quadricauda* reflected through optical density at 730 nm [31]. Error bars indicate 5% standard error.

Table 1
De novo assembly statistics.

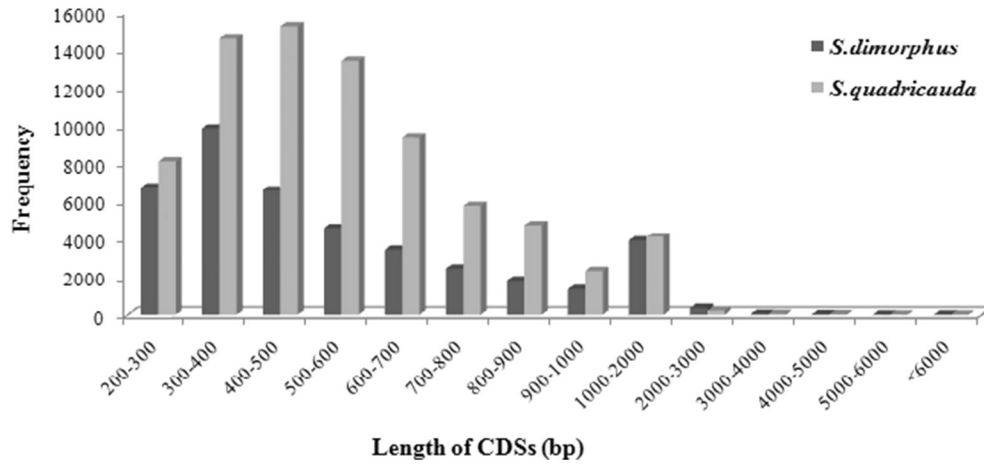| Parameter	S. dimorphus	S. quadricauda
Number of reads	17,907,034	11,290,041
Number of assembled transcripts	60,826	145,112
Total number of bases	45.5 Mb	140.7 Mb
N50	1099	1513

Fig. 3. Length distribution of CDSs for *S. dimorphus* and *S. quadricauda*.

all higher order motifs including tri-, tetra-, penta-nucleotides. Total of 6800 SSRs were identified in 40,979 sequences of *S. dimorphus* whereas 12,169 SSRs were identified in 76,969 sequences of *S. quadricauda*. 785 and 1173 sequences contained more than one SSR in *S. dimorphus* and *S. quadricauda*, respectively. Tri-nucleotide repeats were found to be most prominent repeat patterns (97% in *S. dimorphus*, 85% in *S.*

(A)
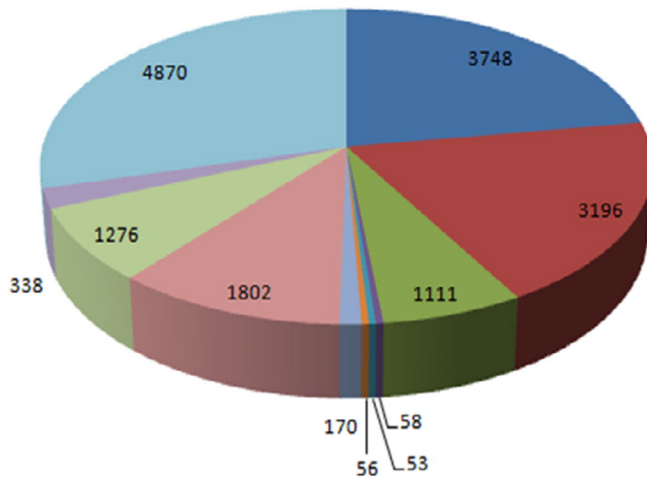
■ *Volvox carteri*
■ *Chlamydomonas reinhardtii*
■ *Chlorella variabilis*
■ *Acutodesmus obliquus*
■ *Ostreococcus tauri*
■ *Guillardia theta*
■ *Micromonas sp.*
■ *Coccomyxa subellipsoidea*
■ *Aspergillus nidulans*
■ *Batrachochytrium denrobatidis*
■ *Others*

(B)
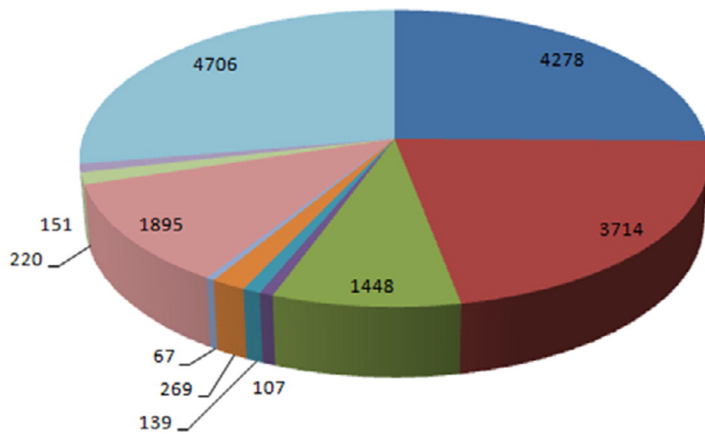
■ *Volvox carteri*
■ *Chlamydomonas reinhardtii*
■ *Chlorella variabilis*
■ *Acutodesmus obliquus*
■ *Ostreococcus sp.*
■ *Micromonas sp.*
■ *Guillardia theta*
■ *Coccomyxa subellipsoidea*
■ *Physcomitrella patens*
■ *Selaginella moellendorffii*
■ *Others*

Fig. 4. A Top Blast species distribution for BLASTx matches in *S. dimorphus* B Top Blast species distribution for BLASTx matches in *S. quadricauda*.
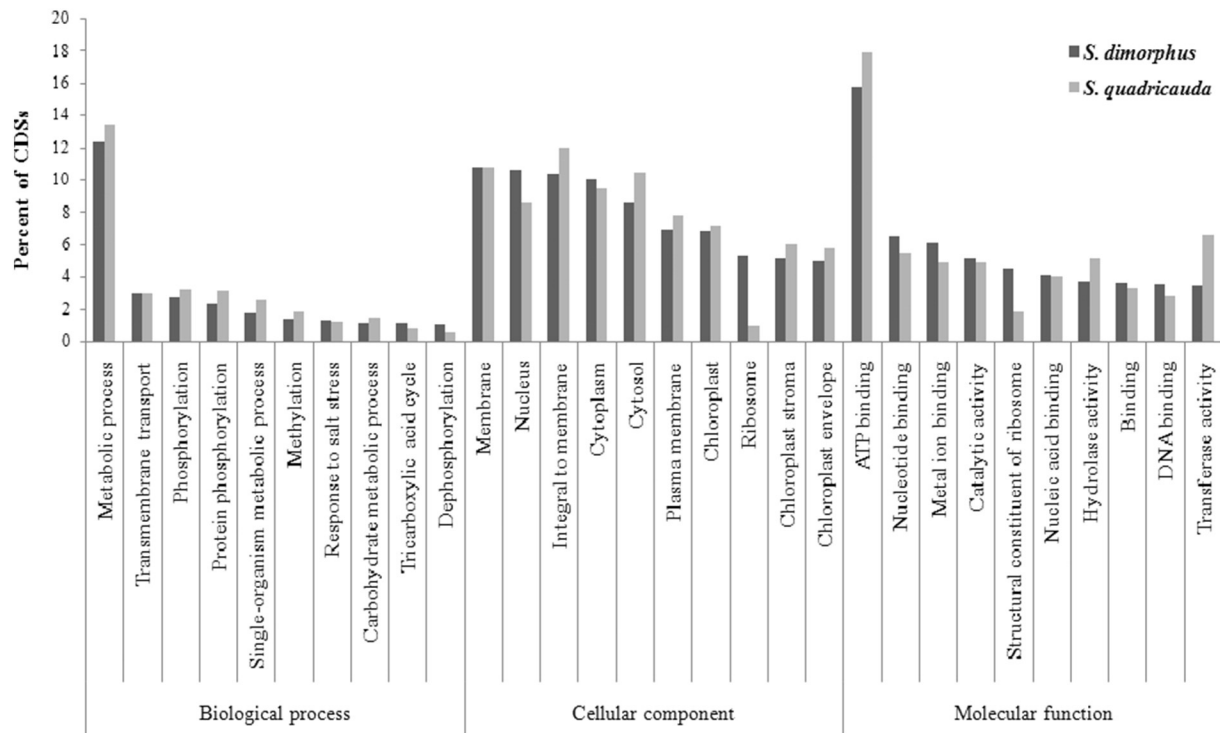
Fig. 5. GO categories in *S. dimorphus* and *S. quadricauda*.

quadricauda) with CAG, GCA, AGC, CTG, and GCT as most frequent repeats, followed by di-, tetra- and penta-nucleotides. Frequency of repeats was observed to be in reciprocal relationship with the length of repeat pattern. The repeat unit of potential SSRs mostly represented was five, followed by six and then others. Based on these identified SSRs, it has been demonstrated that the pattern of repeats was almost similar in both the species. However, differences lie in their abundance. Identification of SSRs is important to reveal polymorphism in the contrasting species (Table 2).

3.4. Pathway classification by KEGG

KAAS was used for assigning KO identifiers to the predicted CDSs and subsequently mapping them to generate KEGG pathways. KAAS first computes the BLAST scores between a query sequence and KEGG GENES database and then homologs are identified on the basis of score. Bi-directional best hit method was used for the analysis and orthologs were divided into KO groups according to the annotation of the KEGG GENES database. Finally, the assignment score was calculated for the assignment of K numbers to the CDSs. The KO entry represents

Table 2
SSRs identified in transcriptomes of *S. dimorphus* and *S. quadricauda*.

Description	*S. dimorphus*	*S. quadricauda*
Total number of sequences examined	40,979	76,969
Total size of examined sequences (bp)	23,585,019	42,992,676
Total number of identified SSRs	6800	12,169
Number of SSR containing sequences	5877	10,866
Number of sequences containing more than one SSR	785	1173
Number of SSRs present in compound formation	9	8
Di-nucleotide	154	1587
Tri-nucleotide	6621	10,395
Tetra-nucleotide	23	181
Penta-nucleotide	2	6

an ortholog group, that is linked to a gene product in the KEGG pathway, hence the assignment of KO identifiers to CDSs enables the reconstruction of KEGG pathways [33]. We have identified enzymes involved in major metabolic pathways, genetic information processing, environmental information processing, cellular processes and organism systems. Distribution of CDS of *S. dimorphus* and *S. quadricauda* in KEGG pathways has been shown in Fig. 6.

3.5. Differentially expressed genes in both the species

Cross species variation in lipid content may be attributed to the differential expression of genes. Our study aimed to compare the transcriptomes of *S. dimorphus* and *S. quadricuada* at stationary phase for identifying differentially expressed genes leading to differences in lipid content to understand the molecular basis for variation. Distribution of genes expressed in both the species has been shown in Fig. 7. About 3893 genes were expressed in both species, whereas 1530 genes were found exclusively in *S. dimorphus* and 412 genes were present exclusively in *S. quadricauda*. Out of the commonly occurring genes in two species, 2537 genes exhibited up- regulation in high lipid content containing strain *S. dimorphus* whereas 1356 genes were found to be down-regulated in *S. dimorphus*. Genes showing up-regulation and down-regulation were found to be involved in various metabolic pathways, including photosynthesis, carbon metabolism, fatty acid and triacylglycerol metabolism, and energy metabolism (Table 3).

3.5.1. Enhanced photosynthesis in S. dimorphus

Photosynthesis is a tightly controlled process, in which light energy is captured and converted into ATP and reducing power (NADPH) [34]. Regulation of photosynthetic reactions is essential for other cellular processes and metabolic reactions. About twenty six genes involved in photosynthesis were found common between both the species. Majority of the photosynthetic genes, including PSI, PSII, cytochrome b6 and light-harvesting complex proteins exhibited enhanced expression in *S. dimorphus* compared to *S. quadricauda*. Last electron transfer from photosystem I is catalyzed by Ferredoxin-NADP$^+$ reductase [35] and gene
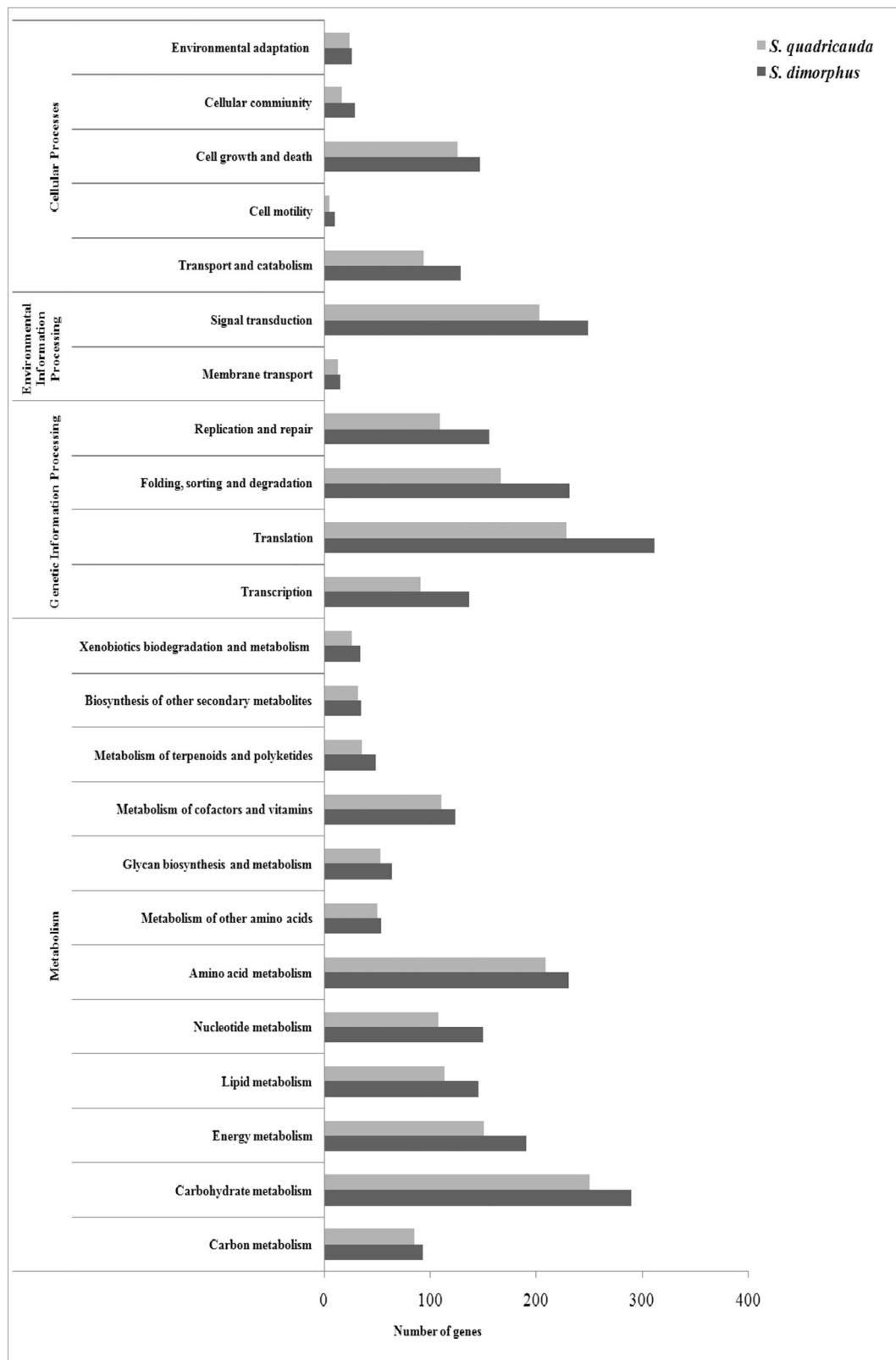
Fig. 6. Functional characterization and abundance of KEGG pathways in *S. dimorphus* and *S. quadricauda*.

encoding the particular enzyme was also found to be up-regulated in *S. dimorphus*. This suggests that photosynthesis is more active in *S. dimorphus* than in *S. quadricauda*, which is consistent with the earlier report, in which lipid accumulation was found to be improved with increase in photosynthetic light reactions in *Chlorella vulgaris* [36].

3.5.2. Carbon metabolism provides precursors for storage lipid production

Light reactions in photosynthesis feed NADPH and ATP towards carbon fixation pathway (Calvin cycle). Fixed carbon provides ATP, reducing power and carbon skeletons for metabolism [34]. Further, glycolysis, pyruvate metabolism and TCA cycle provide precursors for fatty acid

Fig. 7. A) Distribution of genes expressed in *S. dimorphus* and *S. quadricauda* B) Up-regulated and down-regulated genes among the common genes of *S. dimorphus* and *S. quadricauda*.
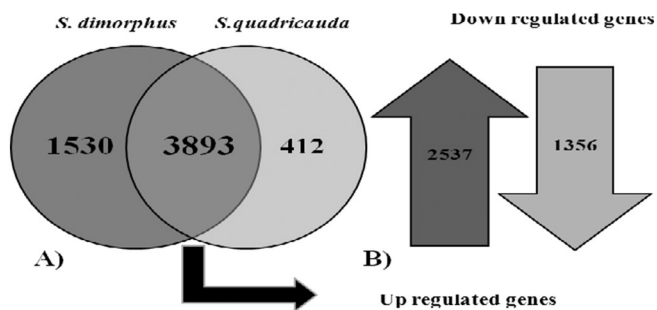
and triacylglycerol biosynthesis in algae and the pentose phosphate pathway is important for glucose consumption coupled with NADPH generation, which is required for fatty acid synthesis [37].

Almost all genes involved in carbon fixation showed enhanced expression in *S. dimorphus* than in *S. quadricauda*. These included, ribulose-bisphosphate carboxylase large chain (RBCL), phosphoglycerate kinase (PGK), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), fructose-bisphosphate aldolase (ALDO), fructose-1,6-bisphosphatase (FBP), sedoheptulose-bisphosphatase (SBP), transketolase (TKT), ribose 5-phosphate isomerase A (RPIA) and phosphoribulokinase (PRK). Increased expression of these genes in *S. dimorphus* suggests the enhancement of carbon fixation process, thus excess carbon supply may have contributed in promoting the lipid biosynthesis and storage.

Further, gene encoding enzyme phosphoenolpyruvate carboxylase (PEPC) was found to be downregulated in *S. dimorphus*. PEPC catalyzes the irreversible β-carboxylation of phosphoenolpyruvate (PEP) to produce inorganic phosphate and oxaloacetate [38]. Decreased expression of PEPC resulted in an increase in the triacylglycerol content in *Chlamydomonas reinhartii* [39,40]. Therefore, down-regulation of PEPC in *S. dimorphus* signifies its vital role for high lipid content of *S. dimorphus*. Another gene of the glycolytic pathway, enolase (ENO) showed considerable increase of expression in *S. dimorphus* (Fig. 8B). ENO is the enzyme that catalyzes the production of PEP, which is the precursor of pyruvate [41]. Increased expression of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and enolase (ENO) was reported in nitrogen deprived cultures of *Isochrysis galbana* leading to high lipid accumulation [42]. High expression of enolase was correlated with the high oil accumulation in sunflower seeds, supported with its important role in carbon partitioning [43].

Also, NADP$^+$ dependent glyceraldehyde-3-phosphate dehydrogenase (GAPA), which catalyzes the irreversible oxidation of glyceraldehyde-3-phosphate and NADP$^+$ into 3-phosphoglycerate and NADPH in glycolysis, was significantly expressed in *S. dimorphus*. 3-phosphoglycerate is the precursor for triacylglycerol biosynthesis, and NADPH is utilized in fatty acid biosynthesis as two NADPH molecules are required to incorporate a carbon unit during elongation of the fatty acid [44]. It has been reported earlier that glyceraldehyde-3-phosphate dehydrogenase (GAPA) expression resulted in promoting the NADPH supply and enhanced the productivity of fatty acid-derived biofuels [45].

Another gene encoding enzyme citrate synthase (CS) was found to be down-regulated in *S. dimorphus*. CS is a rate-limiting enzyme of the TCA cycle, that catalyzes the first step in which condensation of acetyl-CoA with oxaloacetate takes place resulting in the formation of citrate. It has been reported that the suppression of citrate synthase resulted in increase of storage lipid content in *Chlamydomonas reinhardtii* [46]. Down-regulation of CS in *S. dimorphus*, represents the shifting of carbon flux towards lipid production. It can be considered as a significant control point for lipid content enhancement in microalgae.

Table 3
Differentially expressed genes of carbon metabolism and fatty acid and triacylglycerol metabolism in *S. dimorphus* and *S. quadricauda*.

| Gene symbol	Gene name	EC number	Fold change
Carbon metabolism			
HK	Hexokinase	2.7.1.1	− 0.25
GPI	Glucose-6-Phosphate isomerase	5.3.1.9	+ 0.62
PFK	Phosphofructokinase 1	2.7.1.11	− 0.50
ALDO	Fructose-bisphosphate Aldolase	4.1.2.13	+ 13.65
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase	1.2.1.12	+ 15.67
PGK	Phosphoglycerate kinase	2.7.2.3	+ 107.67
GPMI	Bisphosphoglycerate-independent phosphoglycerate mutase	5.4.2.12	+ 1.08
ENO	Enolase	4.2.1.11	+ 20.75
PK	Pyruvate kinase	2.7.1.40	− 1.88
FBP	Fructose-1,6-bisphosphatase	3.1.3.11	− 4.57
RBCL	Ribulose-bisphosphate carboxylase large chain	4.1.1.39	+ 1.48
TKT	Transketolase	2.2.1.1	+ 9.30
TAL	Transaldolase	2.2.1.2	+ 6.10
PPDK	Pyruvate, orthophosphate dikinase	2.7.9.1	+ 2.06
PEPC	Phosphoenolpyruvate carboxylase	4.1.1.31	− 0.14
PEPCK	Phosphoenolpyruvate carboxykinase (ATP)	4.1.1.49	+ 7.04
MDH	Malate dehydrogenase	1.1.1.37	+ 8.27
ACOS	Acetyl-CoA synthetase	6.2.1.1	+ 1.86
CS	Citrate synthase	2.3.3.1	− 0.63
ACL	ATP citrate lyase	2.3.3.8	+ 1.01
PDH	Pyruvate dehydrogenase E1 component alpha subunit	1.2.4.1	+ 2.79
DLD	Dihydrolipoamide dehydrogenase	1.8.1.4	+ 1.53
G6PD	Glucose-6-phosphate 1-dehydrogenase	1.1.1.49	+ 41.96
PGD	Phosphogluconate dehydrogenase	1.1.1.343	+ 67.81
GAPN	Glyceraldehyde-3-phosphate dehydrogenase (NADP+)	1.2.1.9	+ 1.56
GAPA	Glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating)	1.2.1.13	+ 19.30
Fatty acid biosynthesis			
ACCA	Acetyl-CoA carboxylase carboxyl transferase subunit alpha	6.4.1.2	− 0.20
ACCC	Acetyl-CoA carboxylase, biotin carboxylase subunit	6.3.4.14	+ 2.77
ACCD	Acetyl-CoA carboxylase carboxyl transferase subunit beta	6.4.1.2	− 0.21
KAS III	3-oxoacyl-ACP synthase III	2.3.1.180	− 0.16
KAS II	3-oxoacyl-ACP synthase II	2.3.1.179	− 1.14
KAR	3-oxoacyl-ACP reductase	1.1.1.100	+ 1.14
HAD	3-hydroxyacyl-ACP dehydratase	4.2.1.59	+ 4.00
FATA	Fatty acyl-ACP thioesterase	3.1.2.14	+ 5.45
EAR	Enoyl-ACP reductase	1.3.1.10	− 0.53
SAD	Stearoyl-ACP desaturase	1.14.19.2	+ 5.05
FAD	Omega-6 fatty acid desaturase	1.14.19.	− 0.64
Triacylglycerol biosynthesis			
GPAT	Glycerol-3-phosphate acyltransferase	2.3.1.15	− 0.35
PAP	Phosphatidate phosphatase	3.1.3.4	+ 41.94
DGAT	Diacylglycerol acyltransferase	2.3.1.20	+ 14.07
GK	Glycerol kinase	2.7.1.30	+ 1.70
Fatty acid and triacylglycerol degradation			
ACSL	Long-chain acyl-CoA synthetase	6.2.1.3	− 0.47
MFP	Enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase	4.2.1.17	− 0.61
ACOX	Acyl-CoA oxidase	1.3.3.6	− 0.50
ACD	Acyl-CoA dehydrogenase	1.3.8.7	− 0.22
ACAT	Acetyl-CoA acetyltransferase	2.3.1.9	+ 2.93
TGL	TAG lipase	3.1.1.3	− 0.27

(+: Up-regulation; −: Down-regulation).
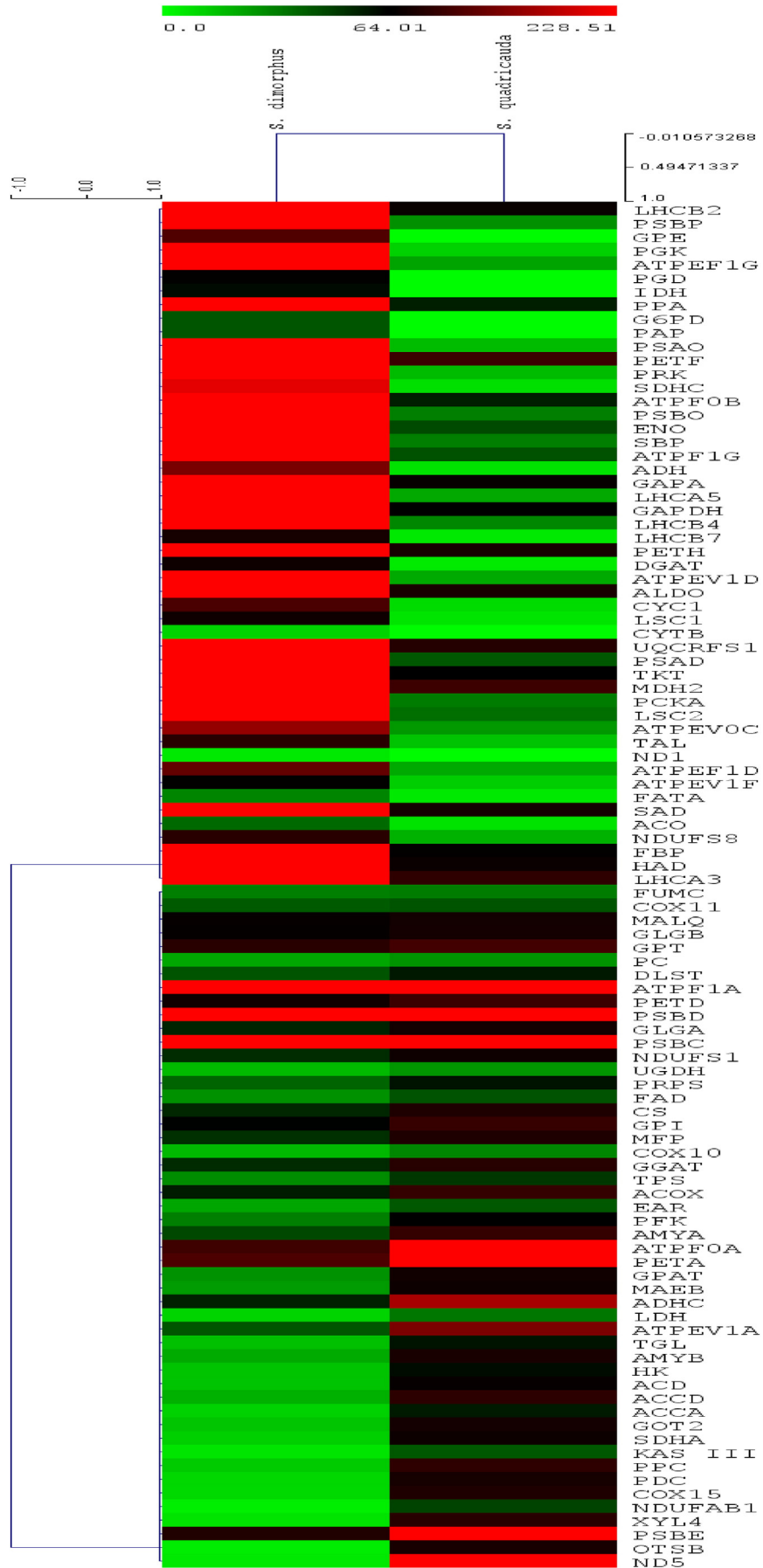
Fig. 8. Heat map representing top 50 up-regulated and top 50 down-regulated genes in *S. dimorphus*.

Acetyl-CoA synthetase (ACOS), pyruvate dehydrogenase (PDH) and ATP citrate lyase (ACL) are three enzymes involved in the production to acetyl-CoA, which is the main precursor of fatty acid synthesis, by alternative routes [47]. Genes encoding these three enzymes showed high transcript abundance in *S. dimorphus*, suggesting their putative role in the regulation of storage lipid biosynthesis. PDH is the enzyme, which links glycolysis with lipid biosynthesis pathway by directly converting pyruvate from glycolysis to acetyl-CoA. Up- regulation of PDH has been reported in oleaginous microalga *Nannochloropsis oceanica* grown under nitrogen depleted conditions [48]. Also, suppression of pyruvate dehydrogenase kinase (PDK), an enzyme that functions to deactivate pyruvate dehydrogenase, resulted in the enhancement of neutral lipid content in *Phaeodactylum tricornutum* [49]. ACOS catalyzes the conversion of acetate to acetyl-CoA and its up-regulation also contributes for the enhancement of storage lipid biosynthesis. Over expression of acetyl-CoA synthetase (ACOS) in microalga *Schizochytrium* resulted in increased fatty acid proportion [50]. Third route for acetyl-CoA biosynthesis involves enzyme ACL which catalyzes the cleavage of citrate for the generation of acetyl-CoA and has been considered as the rate limiting enzyme of lipid biosynthesis [51]. Comparative study of three algal species, i.e. *Chlorella desiccata*, *Dunaliella tertiolecta* and *Chlamydomonas reinhardtii*, that have different amounts of triacylglycerol accumulation revealed up-regulation of ACL in high TAG accumulator *Chlorella desiccata* [47].

Hence, overall up-regulation of genes providing precursors for fatty acid and triacylglycerol biosynthesis from different routes was observed in *S. dimorphus*. These results revealed the importance of pathways supplying carbon precursors for high lipid accumulation in microalgae.

3.5.3. Comparative analysis of fatty acid and triacylglycerol biosynthetic pathways in S. dimorphus and S. quadricauda

Storage lipid biosynthesis in microalgae occurs in two steps involving fatty acid biosynthesis and triacylglycerol biosynthesis pathway. As described above, oxidative decarboxylation of pyruvate provides the main precursor, acetyl-CoA, for fatty acid biosynthesis pathway. In microalgae, fatty acid biosynthesis occurs in the chloroplast. Acetyl-CoA acts as the substrate for acetyl CoA carboxylase (ACC), which is the rate limiting enzyme of the pathway. Fatty acids produced from this pathway were transported to the cytosol and used as precursors for triacylglycerol biosynthesis in endoplasmic reticulum [52].

About eleven genes from the fatty acid biosynthetic pathway were identified in both the species. Majority of the genes, such as fatty acyl-ACP thioesterase A (FATA), oxoacyl-ACP synthase II (KASII), oxoacyl-ACP reductase (KAR), stearoyl-ACP desaturase (SAD) and hydroxyacyl-ACP dehydratase (HAD) showed up-regulation in *S. dimorphus*, which is in accordance with our previous report [31]. Production of monounsaturated fatty acids is critical for microalgal prospective of biodiesel production. Stearoyl-ACP desaturase (SAD) synthesizes monounsaturated fatty acids from saturated fatty acids by introducing double bond into stearic acid [53]. Also fatty acyl-ACP thioesterase A (FATA) performs the termination of fatty acid chain elongation by hydrolyzing the newly formed monounsaturated acyl-ACP into free fatty acids and ACP [54]. These two genes correspond to the production of oleic acid, which is the most desired fatty acid for biodiesel production.

Fatty acids produced in the chloroplast are utilized by enzymes for triacylglycerol accumulation for the sequential acylation of glycerol-3-phosphate [55]. Glycerol kinase (GK), the enzyme that catalyzes the formation of glycerol-3-phosphate from glycerol and provides substrate for triacylglycerol biosynthesis pathway [56] was found to be up-regulated in *S. dimorphus*. This is supported by an earlier report, in which over expression of glycerol kinase resulted in improved lipid content and biomass productivity in *Fistulifera solaris* [57].

Next, sequential esterification of acyl chains from acyl-CoA to positions 1 and 2 of glycerol-3-phosphate takes place, resulting in the formation of phosphatidic acid, catalyzed by enzymes glycerol-3-

phosphate acyltransferase (GPAT) and 1-acyl-sn-glycerol-3-phosphate acyltransferase (AGPAT) [10]. Gene encoding glycerol-3-phosphate acyltransferase (GPAT) was down-regulated in *S. dimorphus*, also revealed in our previous study [31]. Other genes catalyzing the last steps of biosynthesis of triacylglycerol, i.e. phosphatidate phosphatase (PAP) and diacylglycerol acyltransferase (DGAT) were highly up-regulated in *S. dimorphus*.

Further, phospholipid:diacylglycerol acyltransferase (PDAT) catalyzes the recycling of fatty acids by extracting them from membrane lipids and consequently reacylating to form triacylglycerol. Recycling of fatty acids from the membrane lipids has been reported in *C. reinhardtii* which resulted in high storage lipid accumulation [58]. Gene encoding enzyme PDAT has been identified in *S. dimorphus* with a remarkable expression while it has not been found to be expressed in *S. quadricauda*, indicating another control point for high storage lipid accumulation in *S. dimorphus*.

3.5.4. Catabolism of fatty acids and triacylglycerol

Catabolism of triacylglycerol takes place by the action of enzyme triacylglycerol lipase that hydrolyzes the ester bonds, linking fatty acids with glycerol backbone in triacylglycerol and diacylglycerol and thereby releasing free fatty acids [59]. TAG lipase (TGL) showed reduced expression in *S. dimorphus* compared to *S. quadricauda*. Further, breakdown of fatty acids takes place by β-oxidation, subsequently forming the end product acetyl CoA, which is utilized by the TCA cycle [60]. Long-chain acyl-CoA synthetase (ACSL), acyl-CoA oxidase (ACOX) and enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase (MFP) were found to be down-regulated whereas acetyl-CoA acyltransferase (ACAT) was up-regulated in *S. dimorphus*. Overall it has been observed that the degradation of triacylglycerol and fatty acids was more active in the low lipid content isolate *S. quadricauda*.

3.5.5. Graphical representation of differentially expressed genes by heat map

Differentially expressed genes in *S. dimorphus* and *S. quadricauda* were analysed by hierarchical clustering. A heat map was constructed using log-transformed and normalized value of genes based on Pearson's uncentered correlation distance as well as based on complete linkage method (Fig. 8).

3.6. Transcription factor WRI1 correlates positively with high lipid production in S. dimorphus

For identification of transcription factors in both the species, CDSs predicted in *S. dimorphus* and *S. quadricauda* were subjected to BLAST search against PlantTFDB with an E-value cut-off of 10^{-5}. Total of 613 and 454 CDSs (encoding transcription factors) belonging to 25 TF families were found in *S. dimorphus* and *S. quadricauda*, respectively. The most abundant TFs in both the species were found to be of MYB TF family, including MYB and MYB-related families. MYB transcription factors are widely distributed in all eukaryotic organisms consisting of a conserved DNA binding domain at the N-terminus [61]. MYB group of TFs represents 32% and 27% of total CDSs encoding transcription factors in *S. dimorphus* and *S. quadricauda* respectively. These transcription factors are involved in the regulation of gene expression, secondary metabolism and responses to environmental stresses [62]. Also, over expression of soybean MYB in transgenic *Arabidopsis* plants resulted in increased lipid content [63]. Further, Nin-like TF family was observed as an abundant TF family in *S. dimorphus* in comparison to *S. quadricauda*. Nin-like transcription factors have their role in nitrate signaling [64] and hence involved in higher triacylglycerol accumulation because biosynthesis of storage carbohydrates and lipids in microalgae also require nitrogen for its growth and biomass production [65].

Further, literature based mining was performed to identify CDSs encoding transcription factors associated with lipid metabolism in both the species. Hu et al. identified thirty transcription factors involved

in lipid metabolism by *in-silico* approach that belong to various TF families, viz. AP2, HB-other, C3H, MYB_related, E2F/DP, MYB, NF-YC, CPP, HSF, bZIP, LFY [66]. Out of these, twenty transcription factors were identified in *S. dimorphus*, whereas only eight were identified in *S. quadricauda*. Seven transcription factors were common, out of which five were found to be up-regulated belonging to TF families i.e. AP2, MYB, LFY and NF-YC (Table 4). It was observed that one up-regulated transcription factor of AP2 family has shown considerable fold increase of expression in *S. dimorphus* than in *S. quadricauda*. This transcription factor was found to be WRINKLED1 (WRI1) ortholog, which governs the fatty acid synthesis and triacylglycerol accumulation in various plant species [67–69]. Its expression was further examined in three growth stages, viz. EXP, ESP, and LSP of both the species of *Scenedesmus*, i.e. *S. dimorphus* and *S. quadricauda*.

Another CDS encoding transcription factor of NF-YB family was found to be highly expressed in *S. dimorphus*. Nuclear Factor Y is a multimeric transcription factor family consisting of three subunits: NF-YA (CBF-B/HAP2), NF-YB (CBF-A/HAP3) and NF-YC (CBF-C/HAP5). These transcription factors are also called Heme-associated proteins (HAPs) and CCAAT box binding factors (CBFs) and act as important regulators of numerous plant developmental processes and stress-induced responses [70]. LEC1 gene belongs to NF-YB family, a HAP3 subunit of the CCAAT-binding transcription factors [71] and its over expression in various plant species has resulted in considerable increase of seed oil content [72]. Hence, higher transcript abundance of NF-YB suggests that it might be associated with high lipid accumulation in *S. dimorphus*.

3.7. Experimental validation of in-silico transcript abundance by RT-qPCR

Seven key genes identified to be involved in supplying the precursors for fatty acid and triacylglycerol biosynthesis and expected to be the control points of lipid biosynthesis were further validated in both high lipid content and low lipid content species of *Scenedesmus* using RT-qPCR. All genes showed positive correlation with *in-silico* transcript abundance analysis, i.e. genes glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and enolase (ENO), acetyl-CoA synthetase (ACOS), pyruvate dehydrogenase (PDH), ATP citrate lyase (ACL), glycerol kinase (GK) exhibited higher expression in high lipid content *S. dimorphus* whereas citrate synthase (CS) was down-regulated in *S. dimorphus* (Fig. 9A).

Further, the expression of CDS encoding transcription factor WRI1 was analysed by RT-qPCR in both the species at three growth stages.

Table 4
Lipid related transcription factor families in *S. dimorphus* and *S. quadricauda*.

TF family	*S. dimorphus*	*S. quadricauda*	Up-regulation(↑) / Down-regulation (↓)
NF-YC	+	+	↑
C3H	+	−	
C3H	+	−	
MYB_related	+	−	
NF-YC	+	+	↑
MYB_related	+	−	
E2F/DP	+	−	
C3H	+	+	↓
AP2	+	−	
MYB	+	+	↑
NF-YC	+	+	↓
MYB_related	+	−	
MYB_related	+	−	
E2F/DP	+	−	
NF-YC	+	−	
LFY	+	+	↑↑
CPP	+	−	
AP2	+	−	
C3H	+	−	
AP2	+	+	↑↑
E2F/DP	−	+	

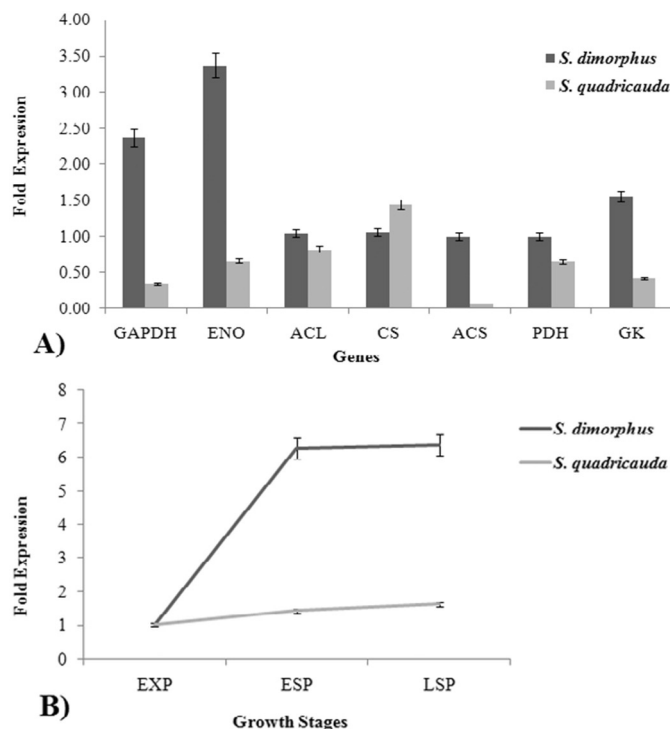(+: Present; −: Absent; ↑↑: Significantly up-regulated).

Fig. 9. A) Expression analysis of seven identified key genes in *S. dimorphus* and *S. quadricauda* at stationary phase by RT-qPCR. Abbreviations: GAPDH -Glyceraldehyde 3-phosphate dehydrogenase; ENO- Enolase; ACOS- Acetyl-CoA synthetase; PDH - Pyruvate dehydrogenase; ACL- ATP citrate lyase; GK- Glycerol kinase; CS- Citrate synthase B) Relative expression analysis of CDS encoding transcription factor WRINKLED 1 in *S. dimorphus* and *S. quadricauda* in three growth stages. Error bars indicate 5% standard error.

There is a considerable fold difference in the expression of WRI1 in *S. dimorphus* and *S. quadricauda* as confirmed by RT-qPCR. In three growth stages, expression of WRI1 showed a dramatic increase at the start of stationary phase in *S. dimorphus* whereas in *S. quadricauda*, slight increase in the expression was observed at stationary phase (Fig. 9B). RT-qPCR results confirmed the reliability of computational analysis of RNA—Seq data and identified genes governing differential lipid accumulation in both the species.

Overall, our study unveiled up-regulation of photosynthesis, carbon metabolism and fatty acid and triacylglycerol biosynthetic pathways in *S. dimorphus* (with high lipid content) compared to *S. quadricauda* (Fig. 10). Transcript abundance of genes of respective pathways suggested the potential role of photosynthesis and carbon fixation in high storage lipid biosynthesis in *S. dimorphus*. Further, genes that provide precursors for triacylglycrol biosynthesis were identified and showed increased expression in *S. dimophus*. Up-regulation of genes involved in the production of acetyl-CoA, which is the main substrate for fatty acid synthesis provided a clue for high lipid content in *S. dimorphus* (Table 5). Two genes involved in triacylglycerol biosynthesis i.e. PAP and DGAT exerted more influence for high storage lipid production. Elevated β-oxidation pathway was found in low lipid accumulating isolate *S. quadricauda*, which is responsible for metabolizing free fatty acids, which are further utilized by TCA cycle. Further, identification of transcription factors revealed higher abundance of lipid-related transcription factors in high lipid content isolate *S. dimorphus*. Expression analysis of WRI1 in both the species with its temporal profile provided a strong indication about its significant role in the high lipid accumulation.

4. Conclusion

Transcriptomes of two lipid content contrasting species of *Scenedesmus* viz. *S. dimorphus* and *S. quadricauda* were assembled and
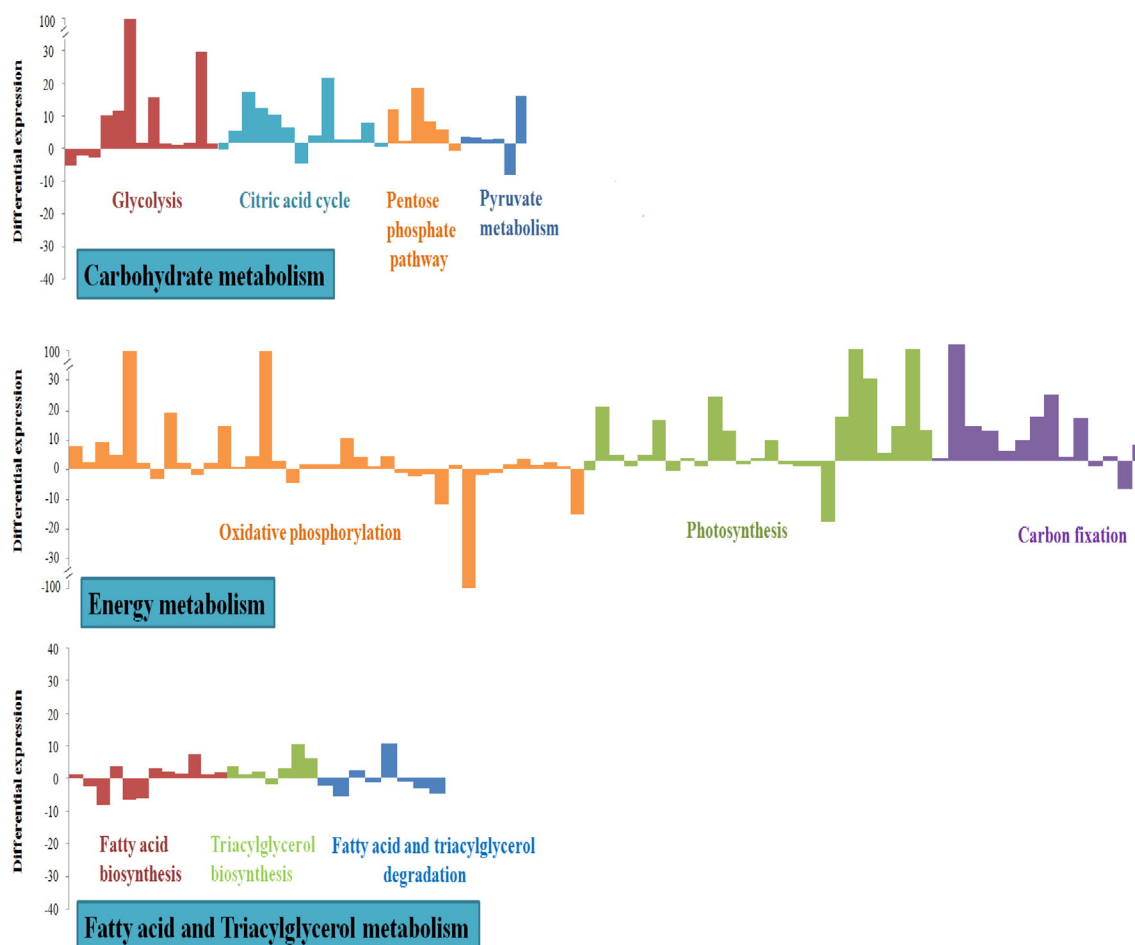
Fig. 10. Up-regulation/Down-regulation of pathways and genes associated with storage lipid accumulation in *S. dimorphus* versus *S. quadricauda*.

analysed, providing comparative investigation of various metabolic pathways, genes and transcriptional regulators leading to variation in the lipid content. Both *S. dimorphus* and *S. quadricauda* showed high similarity with *Volvox carteri* and *Chlamydomonas reinhardtii* and shared same core functional pathways. Pathway analysis and differential expression analysis of the respective pathways unveiled molecular mechanisms that exert a significant role in the enhancement of triacylglycerol accumulation in *Scenedesmus* species. High lipid accumulation was found to be correlated with certain pathways including, photosynthesis, carbon fixation, glycolysis, pyruvate metabolism, TCA cycle, fatty acid and triacylglycerol biosynthesis. Key genes of carbon metabolism, glyceraldehyde 3-phosphate dehydrogenase, enolase, acetyl-CoA synthetase, pyruvate dehydrogenase, ATP citrate lyase, glycerol kinase, citrate synthase were identified as major control points for high lipid content in *S. dimorphus.* These genes provide carbon precursors for

triacylglycerol biosynthesis and therefore, can be targeted by metabolic engineering approaches for enhanced lipid production. Moreover, identification of transcription factors and their expression analysis provided insights towards differential lipid accumulation in two species. Correlation of WRI1 transcription factor with high lipid accumulation was inferred from the comparative analysis between *S. dimorphus* and *S. quadricauda.* Current study provided valuable data source for future research on these species.

Conflict of interest

The authors declare that they have no conflict of interest.

Authors contributions

TS and RSC conceived and designed the experiments. TS conducted the laboratory experiments and *in-silico* analysis. TS and RSC analysed and interpreted the results. RSC reviewed the manuscript. Both authors approved the final manuscript.

Acknowledgements

Table 5
Identified key genes for overall lipid content enhancement in various algal species.

Gene	Algal species	Reference
Glyceraldehyde3-phosphate dehydrogenase	*Isochrysis galbana*	[41]
Enolase	*Isochrysis galbana, Chlamydomonas reinhardtii*	[11,41]
Acetyl-CoA synthetase	*Chlorella desiccata, Schizochytrium* sp.	[46,49]
Pyruvate dehydrogenase	*Nannochloropsis oceanic, Phaeodactylum tricornutum*	[47,48]
ATP citrate lyase	*Chlorella desiccata*	[46]
Citrate synthase	*Chlamydomonas reinhardtii*	[45]
Glycerol kinase	*Fistulifera* sp.	[56]

Appendix A. Supplementary data

Transcript abundance data of metabolic pathways, SSR information, transcription factor data and figures: flow chart and pathways are made available in supplementary file. Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.algal.2016.07.020.

References

[1] Y. Li, M. Horsman, N. Wu, C.Q. Lan, N. Dubois-Calero, Biofuels from microalgae, Biotechnol. Prog. 24 (2008) 815–820.
[2] J. Liu, F. Chen, J. Huang, Microalgae as Feedstocks for Biodiesel Production, Biodiesel – Feedstocks and Processing Technologies, INTECH Open Access Publisher, 2011.
[3] P. Schenk, S. Thomas-Hall, E. Stephens, U. Marx, J. Mussgnug, C. Posten, O. Kruse, B. Hankamer, Second generation biofuels: high-efficiency microalgae for biodiesel production, Bioenergy Res. 1 (2008) 20–43.
[4] Y. Chisti, Biodiesel from microalgae, Biotechnol. Adv. 25 (2007) 294–306.
[5] F.W. Bai, C.G. Liu, H. Huang, G.T. Tsao, D. Wang, Y. Lu, J. Xu, Establishing Oleaginous Microalgae Research Models for Consolidated Bioprocessing of Solar Energy, Biotechnology in China III: Biofuels and Bioenergy, Springer, Berlin Heidelberg, 2011 69–84.
[6] B.H. Hamid Rismani-Yazdi, C. Hsin, J. Peccia, Transcriptomic analysis of the oleaginous microalga Neochloris oleoabundans reveals metabolic insights into triacylglyceride accumulation, Biotechnol. Biofuels 5 (2012).
[7] C. Soneson, M. Delorenzi, A comparison of methods for differential expression analysis of RNA-seq data, BMC Bioinf. 14 (2013) 91.
[8] M.T. Guarnieri, A. Nag, S.L. Smolinski, A. Darzins, M. Seibert, P.T. Pienkos, Examination of triacylglycerol biosynthetic pathways via de novo transcriptomic and proteomic analyses in an unsequenced microalga, PLoS One 6 (2011), e25851.
[9] R. Miller, G. Wu, R.R. Deshpande, A. Vieler, K. Gartner, X. Li, E.R. Moellering, S. Zauner, A.J. Cornish, B. Liu, Changes in transcript abundance in Chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism, Plant Physiol. 154 (2010) 1737–1752.
[10] H. Rismani-Yazdi, B.Z. Haznedaroglu, K. Bibby, J. Peccia, Transcriptome sequencing and annotation of the microalgae Dunaliella tertiolecta: pathway description and gene discovery for production of next-generation biofuels, BMC Genomics 12 (2011) 148.
[11] H. Lv, G. Qu, X. Qi, L. Lu, C. Tian, Y. Ma, Transcriptome analysis of Chlamydomonas reinhardtii during the process of lipid accumulation, Genomics 101 (2013) 229–237.
[12] J.O. Davidson, J. Overton, R. Waikel, Transcriptome analysis of Chlorella protothecoides to identify novel pro-lipid genes for biofuel production, FASEB J. 26 (2012) 790.712.
[13] C. Shang, G. Bi, Z. Yuan, Z. Wang, M.A. Alam, J. Xie, Discovery of genes for production of biofuels through transcriptome sequencing of Dunaliella parva, Algal Res. 13 (2016) 318–326.
[14] L. Rodolfi, G. Chini 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.
[15] G. Ahlgren, L. Lundstedt, M. Brett, C. Forsberg, Lipid composition and food quality of some freshwater phytoplankton for cladoceran zooplankters, J. Plankton Res. 12 (1990) 809–818.
[16] L. Xin, H. Hong-ying, Y. Jia, Lipid accumulation and nutrient removal properties of a newly isolated freshwater microalga, Scenedesmus sp. LX1, growing in secondary effluent, New Biotechnol. 27 (2010) 59–63.
[17] S.T. Girisha, K. Ravikumar, V. Girish, B.R. Mrunalini, Growing of Chlorella, Scenedesmus and Botryococus in sewage water for biodiesel production, Arch. Appl. Sci. Res. 6 (2014) 131–138.
[18] P. Prabakaran, A.D. Ravindran, Scenedesmus as a potential source of biodiesel among selected microalgae, Curr. Sci. 102 (2012) 616–620.
[19] E.B. Sydney, W. Sturm, J.C. de Carvalho, V. Thomaz-Soccol, C. Larroche, A. Pandey, C.R. Soccol, Potential carbon dioxide fixation by industrially important microalgae, Bioresour. Technol. 101 (2010) 5892–5896.
[20] R.S. Gour, A. Kant, R.S. Chauhan, Screening of microalgae for growth and lipid accumulation properties, J. Algal Biomass Utln. 5 (1) (2014) 38–46.
[21] J. Sheehan, T. Dunahay, J. Benemann, P. Roessler, A Look back at the US Department of energy's Aquatic Species Program: Biodiesel from Algae, National Renewable Energy Laboratory Golden, 1998.
[22] R.S. Gour, A. Chawla, H. Singh, R.S. Chauhan, A. Kant, Characterization and screening of native Scenedesmus sp. isolate suitable for biofuel feedstocks, PLoS One 11 (5) (2015) e0155321.
[23] R.R.L. Guillard, Methods for Microflagellates and Nannoplankton, 1973 16.
[24] E.G. Bligh, W.J. Dyer, A rapid method of total lipid extraction and purification, Can. J. Biochem. Physiol. 37 (1959) 911–917.
[25] M.S. Cooper, W.R. Hardin, T.W. Petersen, R.A. Cattolico, Visualizing "green oil" in live algal cells, J. Biosci. Bioeng. 109 (2010) 198–201.
[26] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome, BMC Bioinf. 12 (2011) 323.
[27] N. Ji, L. Li, L. Lin, S. Lin, Screening for suitable reference genes for quantitative real-time PCR in Heterosigma akashiwo (Raphidophyceae), PLoS One 10 (7) (2015), e0132183.

[28] S. Wu, J. Zhou, X. Cao, S. Xue, Determination of internal controls for quantitative gene expression of Isochrysis zhangjiangensis at nitrogen stress condition, J. Ocean Univ. China 15 (1) (2016) 137–144.
[29] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (7) (2002) research0034.1–research0034.11.
[30] T.D. Schmittgen, K.J. Livak, Analyzing real-time PCR data by the comparative C(T) method, Nat. Protoc. 3 (6) (2008) 1101–1108.
[31] T. Sharma, R.S. Gour, A. Kant, R.S. Chauhan, Lipid content in Scenedesmus species correlates with multiple genes of fatty acid and triacylglycerol biosynthetic pathways, Algal Res. 12 (2015) 341–349.
[32] H. Kuntal, V. Sharma, H. Daniell, Microsatellite analysis in organelle genomes of chlorophyta, Bioinformation 8 (2012) 255–259.
[33] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome annotation and pathway reconstruction server, Nucleic Acids Res. 35 (2007) W182–W185.
[34] X. Johnson, J. Alric, Central carbon metabolism and electron transport in Chlamydomonas reinhardtii: metabolic constraints for carbon partitioning between oil and starch, Eukaryot. Cell 12 (2013) 776–793.
[35] 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.
[36] B.D. Woodworth, R.L. Mead, C.N. Nichols, D.R.J. Kolling, Photosynthetic light reactions increase total lipid accumulation in carbon-supplemented batch cultures of Chlorella vulgaris, Bioresour. Technol. 179 (2015) 159–164.
[37] C. Gao, Y. Wang, Y. Shen, D. Yan, X. He, J. Dai, Q. Wu, Oil accumulation mechanisms of the oleaginous microalga Chlorella protothecoides revealed through its genome, transcriptomes, and proteomes, BMC Genomics 15 (2014) 582.
[38] Z.K. Yang, Y.F. Niu, Y.H. Ma, J. Xue, M.H. Zhang, W.D. Yang, H.Y. Li, Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation, Biotechnol. Biofuels 6 (1) (2013) 1.
[39] X. Deng, J. Cai, Y. Li, X. Fei, Expression and knockdown of the PEPC1 gene affect carbon flux in the biosynthesis of triacylglycerols by the green alga Chlamydomonas reinhardtii, Biotechnol. Lett. 36 (11) (2014) 2199–2208.
[40] Q.L. Tian, D.J. Shi, X.H. Jia, H.L. Mi, X.W. Huang, P.M. He, Recombinant expression and functional analysis of a Chlamydomonas reinhardtii bacterial-type phosphoenolpyruvate carboxylase gene fragment, Biotechnol. Lett. 36 (4) (2014) 821–827.
[41] J.E. Polle, P. Neofotis, A. Huang, W. Chang, K. Sury, E.M. Wiech, Carbon partitioning in green algae (chlorophyta) and the enolase enzyme, Metabolites 4 (3) (2014) 612–628.
[42] P. Song, L. Li, J. Liu, Proteomic analysis in nitrogen-deprived Isochrysis galbana during lipid accumulation, PLoS One 8 (2013), e82188.
[43] M.A. Troncoso-Ponce, R. Garces, E. Martinez-Force, Glycolytic enzymatic activities in developing seeds involved in the differences between standard and low oil content sunflowers (Helianthus annuus L.), Plant Physiol. Biochem. 48 (2010) 961–965.
[44] X. Tang, J. Lee, W.N. Zhen, Engineering the fatty acid metabolic pathway in Saccharomyces cerevisiae for advanced biofuel production, Metab. Eng. Commun. 2 (2015) 58–66.
[45] J. Sheng, X. Feng, Metabolic engineering of yeast to produce fatty acid-derived biofuels: bottlenecks and solutions, Front. Microbiol. 6 (2015).
[46] X. Deng, J. Cai, X. Fei, Effect of the expression and knockdown of citrate synthase gene on carbon flux during triacylglycerol biosynthesis by green algae Chlamydomonas reinhardtii, BMC Biochem. 14 (2013) 38.
[47] O. Avidan, U. Pick, Acetyl-CoA synthetase is activated as part of the PDH-bypass in the oleaginous green alga Chlorella desiccata, J. Exp. Bot. 66 (2015) 7287–7298.
[48] 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.
[49] Y.H. Ma, X. Wang, Y.F. Niu, Z.-K. Yang, M.H. Zhang, Z.M. Wang, W.D. Yang, J.S. Liu, H.Y. Li, Antisense knockdown of pyruvate dehydrogenase kinase promotes the neutral lipid accumulation in the diatom Phaeodactylum tricornutum, Microb. Cell Factories 13 (2014) 100.
[50] J. Yan, R. Cheng, X. Lin, S. You, K. Li, H. Rong, Y. Ma, Overexpression of acetyl-CoA synthetase increased the biomass and fatty acid proportion in microalga Schizochytrium, Appl. Microbiol. Biotechnol. 97 (2013) 1933–1939.
[51] Z. Ma, C.H. Chu, D. Cheng, A novel direct homogeneous assay for ATP citrate lyase, J. Lipid Res. 50 (2009) 2131–2135.
[52] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, A. Darzins, Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances, Plant J. 54 (2008) 621–639.
[53] J.L. Watts, J. Browse, A palmitoyl-CoA- specific delta9 fatty acid desaturase from Caenorhabditis elegans, Biochem. Biophys. Res. Commun. 272 (1) (2000) 263–269.
[54] A. Jones, H.M. Davies, T.A. Voelker, Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases, Plant Cell 7 (3) (1995) 359–371.
[55] J. Fan, C. Andre, C. Xu, A chloroplast pathway for the de novo biosynthesis of triacylglycerol in Chlamydomonas reinhardtii, FEBS Lett. 585 (12) (2011) 1985–1991.
[56] A. Cagliari, R. Margis, F. dos Santos Maraschin, A.C. Turchetto-Zolet, G. Loss, M. Margis-Pinheiro, Biosynthesis of triacylglycerols in plants and algae, Int. J. Plant Biol. 2 (1) (2011) 10.
[57] M. Muto, Y. Fukuda, M. Nemoto, T. Yoshino, T. Matsunaga, T. Tanaka, Establishment of a genetic transformation system for the marine pennate diatom Fistulifera sp. strain JPCC DA0580 - a high triglyceride producer, Mar. Biotechnol. 15 (2014).
[58] K. Yoon, D. Han, Y. Li, M. Sommerfeld, Q. Hu, Phospholipid:diacylglycerol acyltransferase is a multifunctional enzyme involved in membrane lipid turnover and

degradation while synthesizing triacylglycerol in the unicellular green microalga *Chlamydomonas reinhardtii*, Plant Cell 24 (2012) 3708–3724.

[59] J. Chen, R.K. Tan, X.J. Guo, Z.L. Fu, Z. Wang, Z.-Y. Zhang, X.-L. Tan, Transcriptome analysis comparison of lipid biosynthesis in the leaves and developing seeds of *Brassica napus*, PLoS One 10 (2015).

[60] A. Mühlroth, K. Li, G. Røkke, P. Winge, Y. Olsen, M.F. Hohmann-Marriott, O. Vadstein, A.M. Bones, Pathways of lipid metabolism in marine algae, co-expression network, bottlenecks and candidate genes for enhanced production of EPA and DHA in species of *Chromista*, Mar. Drugs 11 (11) (2013) 4662–4697.

[61] H. Du, B.R. Feng, S.S. Yang, Y.B. Huang, Y.X. Tang, The R2R3-MYB transcription factor gene family in maize, PLoS One 7 (6) (2012), e37463.

[62] S. Ambawat, P. Sharma, N.R. Yadav, R.C. Yadav, MYB transcription factor genes as regulators for plant responses: an overview, Physiol. Mol. Biol. Plants 19 (3) (2013) 307–321.

[63] Y.F. Liu, Q.T.X. Li, Q.X. LuSong, S.M. Lam, W.K. Zhang, B. Ma, Q. Lin, W.Q. Man, W.G. Du, G.H. Shui, Soybean GmMYB73 promotes lipid accumulation in transgenic plants, BMC Plant Biol. 14 (1) (2014) 1.

[64] M. Konishi, S. Yanagisawa, *Arabidopsis* NIN-like transcription factors have a central role in nitrate signalling, Nat. Commun. 4 (2013) 1617.

[65] C. Remacle, G. Eppe, N. Coosemans, E. Fernandez, H. Vigeolas, Combined intracellular nitrate and NIT2 effects on storage carbohydrate metabolism in *Chlamydomonas*, J. Exp. Bot. 65 (1) (2014) 23–33.

[66] J. Hu, D. Wang, J. Li, G. Jing, K. Ning, J. Xu, Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae *Nannochloropsis*, Sci. Rep. 4 (2014).

[67] Y. Yang, J. Munz, C. Cass, A. Zienkiewicz, Q. Kong, W. Ma, S. Sanjaya, J.C. Sedbrook, C. Benning, Ectopic expression of WRI1 affects fatty acid homeostasis in *Brachypodium distachyon* vegetative tissues, Plant Physiol. 01236 (2015).

[68] W. Ma, Q. Kong, V. Arondel, A. Kilaru, P.D. Bates, N.A. Thrower, C. Benning, J.B. Ohlrogge, WRINKLED1, a ubiquitous regulator in oil accumulating tissues from *Arabidopsis* embryos to oil palm mesocarp, PLoS One 8 (7) (2013), e68887.

[69] D. Tajima, A. Kaneko, M. Sakamoto, Y. Ito, N.T. Hue, M. Miyazaki, Y. Ishibashi, T. Yuasa, M. Iwaya-Inoue, WRINKLED 1 (WRI1) homologs, AP2-type transcription factors involving master regulation of seed storage oil synthesis in castor bean (*Ricinus communis*L.), Am. J. Plant Sci. (2013).

[70] F. Zhang, M. Han, Q. Lv, F. Bao, Y. He, Identification and expression profile analysis of NUCLEAR FACTOR-Y families in *Physcomitrella patens*, Front. Plant Sci. 6 (2015) 642.

[71] M. Suzuki, D.R. McCarty, Functional symmetry of the B3 network controlling seed development, Curr. Opin. Plant Biol. 11 (2008) 548–553.

[72] B. Shen, W.B. Allen, P. Zheng, C. Li, K. Glassman, J. Ranch, D. Nubel, M.C. Tarczynski, Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize, Plant Physiol. 153 (2010) 980–987.