

## **Annotation of the microalgae genomes: fatty acid biosynthesis gene discovery and pathway description for production of next-generation biofuels**

Deepak Kumar Sharma\*, Neelam Singh, R. K. Trivedi  
Department of Oil and Paint Technology, HBTI Kanpur, UP, India; \*Corresponding Author;  
Deepak Kumar  
Sharma, email: [deepakshrm89@gmail.com](mailto:deepakshrm89@gmail.com), Ph.: 7985523200

### **ABSTRACT**

In the present study, genome-wide identification of 333 putative genes involved in fatty acids biosynthesis in genome of *Asterochloris sp.* Cgr/DA1pho v1.0, *Bigelowiella natans* CCMP2755 v1.0, *Chlamydomonas reinhardtii* v4.0, *Chlorella variabilis* NC64A, *Coccomyxa subellipsoidea* C-169 v2.0,

*Fragilariopsis cylindrus*, *Micromonas pusilla* CCMP1545 v3.0, *Ostreococcus sp.* RCC809 2.0, *Ostreococcus tauri* v2.0, *Pseudo-nitzschia multiseriis* CLN-47 v1.0, *Thalassiosira pseudonana* and *Volvox carteri f. nagariensis* were identified. Physico-chemical parameters of identified protein were determined using ProtParam tool. The results indicated that the fatty acids metabolic pathways in all the species are carried out by a comparable number of orthologous proteins. These studies will improve our understanding of the fatty metabolic pathway and contribute to the engineering of regulatory networks of algal strains for higher accumulation of biofuel production.

**Keywords:** Fatty acids biosynthetic genes; microalgae; biofuel; *Arabidopsis*

## 1.INTRODUCTION

Increasing levels of air pollution, high energy demand and the incessant rise in crude oil prices are the main issues in this time and mounting pressure on global research in biofuel production. Algae-based biofuel is a new energy source that attracts researchers due to their rapid growth rate coupled with high carbon dioxide uptake, high lipid content and comparatively low, marginal land usage rates [1]. Notwithstanding the many advantages of biofuels and their technical feasibility has led to renewed interest in microalgae as potential feedstocks and rising research activity focused on the basic biology of algae so that better strain develop and improved post-harvest process engineering [2]. Microalgae cells under controlled stress conditions and engineering metabolic pathways can accumulate large quantities of oils (triacylglycerols) and carbohydrates, particularly when nutrient-deprived [3-4]. Dismukes et al. 2008; Wijffels and Barbosa 2010). Over the past few years various studies have been carried out concerning alteration of fatty acid composition in plants through genetic engineering approaches, along with the development and deployment of a number of plant lipid-related genomics databases[5-8]. (Misra et al., 2012; Napier 2007; Topfer et al. 1995; Mao et al. 2009).

The present work was designed to identify the genes involved in lipid metabolic pathway from the genomes of microalgae (including *Asterochloris sp.* Cgr/DA1pho v1.0, *Aureococcus anophagefferens*, *Bigeloviella natans* CCMP2755 v1.0, *Chlamydomonas reinhardtii* v4.0, *Chlorella variabilis* NC64A, *Coccomyxa subellipsoidea* C-169 v2.0, *Emiliana huxleyi* CCMP1516 main genome assembly v1.0, *Fragilariopsis cylindrus*, *Guillardia theta* CCMP2712 v1.0, *Micromonas pusilla* CCMP1545 v3.0, *Micromonas sp.* RCC299 v3.0, *Ostreococcus lucimarinus* v2.0, *Ostreococcus sp.* RCC809 2.0, *Ostreococcus tauri* v2.0, *Phaeodactylum tricornutum* v2.0, *Pseudo-nitzschia multiseris* CLN-47 v1.0, *Thalassiosira pseudonana* and *Volvox carteri f. nagariensis*) using sequence similarity search with *Arabidopsis thaliana* homologs. The present study contributes to the biochemical and molecular information needed for augmentation of fatty acid biosynthesis in microalgae.

## 2. MATERIALS AND METHODS

### 2.1. Gene Retrieval and Annotation

Lipid genes were obtained from the *Arabidopsis thaliana* lipid gene database (<http://lipids.plantbiology.msu.edu/>) to construct a query protein set. The *Arabidopsis* fatty acids gene database is a convenient and reliable source of genes covering all the major biochemical events responsible for biosynthesis of plant. Subsequently, each protein in the query dataset was used to identify homologs in microalgae by subjecting it to BLAST search with e-value inclusion threshold set to  $1.0E-5$  against microalgal genome databases provided by Joint Genome Institute. These include *Asterochloris sp.* Cgr/DA1pho v1.0, *Aureococcus anophagefferens*, *Bigeloviella natans* CCMP2755 v1.0, *Chlamydomonas reinhardtii* v4.0, *Chlorella variabilis* NC64A, *Coccomyxa subellipsoidea* C-169 v2.0, *Emiliana huxleyi* CCMP1516 main genome assembly v1.0, *Fragilariopsis cylindrus*, *Guillardia theta* CCMP2712 v1.0, *Micromonas pusilla* CCMP1545 v3.0, *Micromonas sp.* RCC299 v3.0, *Ostreococcus lucimarinus* v2.0, *Ostreococcus sp.* RCC809 2.0, *Ostreococcus tauri* v2.0, *Phaeodactylum tricornutum* v2.0, *Pseudo-nitzschia multiseris* CLN-47 v1.0, *Thalassiosira pseudonana* and *Volvox carteri f. nagariensis*. Based on blast score some sequences “hits” were then discarded.

## 2.2 Metabolic Pathway Study

Metabolic pathways were analyzed using the KEGG pathway database [9](Kanehisa et al. 2004) to analyze fatty acid biosynthesis pathways in diverse organisms. To annotation pathway, sequences were submitted to the KEGG Automatic Annotation Server (KAAS) to identify the orthologous gene groups [10].(Moriya et al. 2000). KAAS annotates every submitted sequence with a KEGG ortholog (KO) identifier that allows identification of orthologous and paralogous relationships between the genes of interest. Further, fatty acid biosynthesis pathways were downloaded from the KEGG database. This dataset contains a complete biochemical description of the pathway of fatty acid biosynthesis observed in different organisms.

## 2.3 Physico-chemical characterization prediction

Physico-chemical properties like length, molecular weight, isoelectric point (pI), total number of positive and negative residues, Instability Index (Guruprasad et al 1990), Aliphatic Index (Ikai 1980) and Grand Average hydropathy (GRAVY)(Doolittle 1982) were computed using the Expasy's ProtParam server (Gasteiger et al. 1999).[11-14].

## 3.RESULTS AND DISCUSSION

### 3.1 Comparative genomic analyses of fatty acid genes in microalgal species

Under optimal conditions, microalgae synthesize fatty acids used primarily for esterification into polar glycerol based membrane lipids like glycosylglycerides and phosphoglycerides, whereas under stress conditions, many microalgae tend to accumulate storage lipids called triacylglycerol (TAGs)[15] (Reyes-Prieto et al. 2006). Although global fatty acid biosynthetic mechanisms are known in higher plants [16](Coleman and Lee 2004), pathways responsible for fatty acid accumulation in microalgae are not well studied. Hence, in order to bridge our existing knowledge gap regarding algal fatty acid metabolism, comparative metabolic pathway analyses have been performed across 18 microalgal genomes, using homologous plant genes as reference with an objective of functional characterization of predicted genes. In the present study, using the *Arabidopsis* annotation data as the BLAST input query set, a total of 333 orthologous genes present in *Asterochloris sp. Cgr/DA1pho v1.0*, *Aureococcus anophagefferens*, *Bigelowiella natans CCMP2755 v1.0*, *Chlamydomonas reinhardtii v4.0*, *Chlorella variabilis NC64A*, *Coccomyxa subellipsoidea C-169 v2.0*, *Emiliana huxleyi CCMP1516 main genome assembly v1.0*, *Fragilariopsis cylindrus*, *Guillardia theta CCMP2712 v1.0*, *Micromonas pusilla CCMP1545 v3.0*, *Micromonas sp. RCC299 v3.0*, *Ostreococcus lucimarinus v2.0*, *Ostreococcus sp. RCC809 2.0*, *Ostreococcus tauri v2.0*, *Phaeodactylum tricorutum v2.0*, *Pseudo-nitzschia multiseriis CLN-47 v1.0*, *Thalassiosira pseudonana* and *Volvox carteri f. nagariensis* genomes were identified.

The above approach to identify candidate genes involved in biosynthesis and accumulation of storage oil has been successfully demonstrated in plants by Sharma and Chauhan[17] (Sharma and Chauhan 2012). In these 333 genes includes 20, 17, 15, 22, 25, 19, 21, 16, 21, 17, 0, 18, 18, 20, 17, 20 and 24 genes from *Asterochloris sp. Cgr/DA1pho v1.0*, *Aureococcus anophagefferens*, *Bigelowiella natans CCMP2755 v1.0*, *Chlamydomonas reinhardtii v4.0*, *Chlorella variabilis NC64A*, *Coccomyxa subellipsoidea C-169 v2.0*, *Emiliana huxleyi CCMP1516 main genome assembly v1.0*, *Fragilariopsis cylindrus*, *Guillardia theta CCMP2712 v1.0*, *Micromonas pusilla CCMP1545 v3.0*, *Micromonas sp. RCC299 v3.0*, *Ostreococcus lucimarinus v2.0*, *Ostreococcus*

*sp. RCC809 2.0, Ostreococcus tauri v2.0, Phaeodactylum tricornutum v2.0, Pseudo-nitzschia multiseriata CLN-47 v1.0, Thalassiosira pseudonana and Volvox carterii f. nagariensis* genomes, respectively (Table 1&2).

To investigate metabolic processes responsible for the synthesis of microalgal biofuel precursors, KO identifiers were assigned to the predicted 333 genes, which can subsequently used to study metabolic pathway maps available in KEGG pathway database. KEGG is considered one of the most important bioinformatics resources for understanding higher-order functional meaning and the utilities of the organism from its genome information. It hosts information on the majority of well-known metabolic pathways, including fatty acid pathways for several organisms such as higher plants, bacteria and algae.

### 3.2 Physico-chemical parameters prediction

Various physico-chemical parameters were computed using ExPASy's ProtParam tool [18](Gasteiger et al. 2005) is shown in table 3. Molecular weight was observed between the ranges of 7859.7– 263166.3 for all fatty acid biosynthetic proteins in microalgae. The majority of the predicted proteins were found to have a pI less than 7, indicating that proteins involved in fatty acid biosynthesis are generally acidic in nature. Instability Index analyses reveals the presence of certain dipeptides occurring at significantly different frequencies between stable and unstable proteins. Proteins with an instability index less than 40 are predicted to be stable while those with a value greater than 40 are assumed to be unstable. In the present study the high occurrence frequency of stable proteins exist. The aliphatic index refers to the relative volume of a protein that is occupied by aliphatic side chains (eg, alanine, isoleucine, leucine and valine) and contributes to the increased thermal stability observed for globular proteins. Aliphatic index for the screened proteins ranged from 66.77 to 112.50. The very high aliphatic index for all sequences indicated that their structures are more stable over a wide range of temperature. The GRAVY index indicates the solubility of the protein. The fatty acid biosynthetic proteins which showed large negative values indicated that these proteins are relatively more hydrophobic when compared to proteins with less negative values.

## 4. CONCLUSION

A comprehensive computational analysis of the predicted genes of microalgae against *Arabidopsis* was performed through gene annotation and physico-chemical characterization studies. The results revealed that although each of the algal species maintains the basic genomic repertoire required for fatty acid biosynthesis, they possess additional lineage-specific gene groups. The in-depth broad investigation of each individual gene and their encoded products across the microalgal genome will certainly facilitate metabolic engineering of microalga for biofuel production.

## 5. REFERENCES

1. Chisti Y (2007) Biodiesel from microalgae. *Biotechnol Adv.* 25:294–306.
2. Pienkos PT, Darzins A (2009) the promise and challenges of microalgal-derived biofuels. *Biofuels* 3:431–40.
3. Dismukes GC, Carrieri D, Bennette N, Ananyev GM, Posewitz MC (2008) Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. *Curr Opin Biotechnol* 19: 235–240.
4. Wijffels RH, Barbosa MJ (2010) An outlook on microalgal biofuels. *Science* 329: 796-99

log5. Misra N, Panda PK, Parida BK, Mishra BK (2012). Phylogenomic Study of Lipid Genes Involved in Microalgal Biofuel Production-Candidate Gene Mining and Metabolic Pathway Analyses. *Evolutionary Bioinformatics* 8: 545–564.

5. Napier JA (2007) the production of unusual fatty acids in transgenic plants. *Annu Rev Plant Biol.* 58:295–319.
6. Topfer R, Martini N, Schell J (1995) Modification of Plant Lipid synthesis. *Science* 268:681–6.
7. Mao F, Yin Y, Zhou FF, et al (2009) pDAWG: an integrated database for plant cell wall genes. *Bioenerg Res.* 2:209–16.
8. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M(2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32:277–80.
9. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2000) KAAS: an automatic genome annotation and pathway reconstruction server. *NucleicAcids Res.* 35:W182–5.
10. Guruprasad K, Reddy BVP, Pandit MV (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Prot Eng.* 4:155–64.
11. Ikai AJ (1980) Thermo stability and aliphatic index of globular proteins. *J Biochem.* 88:1895–1898.
12. Doolittle RFK (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 157:105–32.
13. Gasteiger E, Hoogland C, Gattiker A, et al (1999) Protein identification and analyses tools on the Expasy's server. *Methods Mol Biol.* 112:531–52.
14. Reyes-Prieto A, Yoon HS, Bhattacharya D (2006) Phylogenomics: its growing impact on Algal Phylogeny and evolution. *Algae* 21:1–10.
15. Coleman RA, Lee DP (2004) Enzymes of triacylglycerol synthesis and their regulation. *Prog Lipid Res.* 43:134–76.
16. Sharma A, Chauhan R (2012) In-silico Identification and Comparative Genomics of Candidate Genes Involved in Biosynthesis and Accumulation of Seed Oil in Plants. *Comp Funct Genom.* Article ID914843:1–14.





**Table 1:** Candidate genes involved in Fatty Acid biosynthesis in Plastids of *Arabidopsis thaliana*, *Asterochloris* sp. Cgr/DA1pho v1.0, *Aureococcus anophagefferens*, *Bigeloviella natans* CCMP2755 v1.0, *Chlamydomonas reinhardtii* v4.0, *Chlorella variabilis* NC64A, *Coccomyxa subellipsoidea* C-169 v2.0.

| Enzyme Name   | Gene / Protein name                     | KOG no.                    | Protein code | Corresponding homologous protein predicted in algal species(SwissProt accession ID) |  |                                    |  |                                       |                                   |
|---|---|----------------------------|--------------|---|--|------------------------------------|--|---------------------------------------|-----------------------------------|
|   |   |                            |              | <i>Arabidopsis thaliana</i>   | <i>Asterochloris</i> sp. Cgr/DA1pho v1.0 | <i>Aureococcus anophagefferens</i> | <i>Bigeloviella natans</i> CCMP2755 v1.0 | <i>Chlamydomonas reinhardtii</i> v4.0 | <i>Chlorella variabilis</i> NC64A |
| Homomeric Acetyl-CoA Carboxylase (ACCase) (EC 6.4.1.2)      | acetyl-CoA carboxylase, putative (ACC2) | K01961<br>K11263<br>K11262 | At1g36180    | HG001   | F0YJA4                                   | HG002                              |  | E1Z4T6                                | HG003                             |
| Carboxyltransferase of Heteromeric ACCase                   | -                                       | K01962<br>K01958           | At2g38040    | HG007   | F0XWB2                                   |                                    | E1ZJK9                                   | A8J646                                | I0Z5U2                            |
| Carboxyltransferase of Heteromeric ACCase (plastid encoded) | accD                                    | K01963<br>K01958<br>K01966 | AtCg00500    | HG008   | F0XWB2                                   | HG009                              | A8JHU1                                   |                                       | HG010                             |
| Biotin Carboxyl Carrier Protein of Heteromeric ACCase       | BCCP1 (CAC1A)                           | K02160<br>K01965           | At5g16390    |   |  |                                    | E1Z723                                   | A8JDA7                                | I0Z3K5                            |
|   | BCCP2                                   | K02160<br>K01965           | At5g15530    |   |  |                                    | E1Z723                                   | A8JDA7                                | I0Z3K5                            |
| Biotin Carboxylase of                                       | CAC2                                    | K01961                     | At5g35360    | HG014   | F0XXN5                                   | HG015                              | E1Z5P4                                   | A8JGF4                                | I0Z7E6                            |

|   |   |                                      |           |       |        |       |        |                      |                      |
|---|---|--------------------------------------|-----------|-------|--------|-------|--------|----------------------|----------------------|
| Heteromeric ACCase  |   | K01965<br>K01958<br>K01968<br>K11263 |           |       |        |       | E1ZNZ6 |                      |                      |
| Malonyl-CoA : ACP<br>Malonyltransferase<br>(EC<br>2.3.1.39) | putative malonyl-CoA<br>:<br>acyl carrier protein<br>transacylase   | K00645                               | At2g30200 | HG019 | F0YJC8 | HG020 | E1Z713 | HG021                | I0Z821               |
| Ketoacyl-ACP<br>Synthase<br>I (KAS I) (EC<br>2.3.1.41)      | 3-oxoacyl-[acyl]-<br>carrier-<br><br>protein] synthase I<br>precursor (beta-<br>ketoacyl-<br>acp synthase I) (KAS<br>I) | K09458<br><br>K01897                 | At5g46290 | HG026 | F0Y6D5 | HG027 | E1ZF89 | A8JEF7<br><br>A8JCK1 | I0YKJ8<br><br>I0YJ46 |
| KAS II (EC<br>2.3.1.41)                                     | putative 3-keto-acyl-<br>ACP<br>synthase  | K09458                               | At1g74960 | HG031 | F0Y6D5 | HG032 | A8JCK1 | E1ZF89               | I0YJ46               |
| KAS III (EC<br>2.3.1.41)                                    | 3-ketoacyl-acyl carrier   | K00648                               | At1g62640 | HG036 | F0XWH7 | HG037 | HG038  | E1ZFD0               | I0YIV0               |

|  |   |                            |           |       |        |        |        |                  |        |
|--|---|----------------------------|-----------|-------|--------|--------|--------|------------------|--------|
|  | protein synthase III (KAS III)  |                            |           |       |        |        |        |                  |        |
| Ketoacyl-ACP Reductase (EC 1.1.1.100)  | putative 3-oxoacyl-[acyl-carrier protein] reductase                   | K00059                     | At1g24360 | HG042 | F0Y3V7 | HG043  | Q84X75 | E1ZFR4           | I0YXT1 |
| Hydroxyacyl-ACP Dehydrase (EC 4.2.1.*) | putative beta-hydroxyacyl-ACP dehydratase                             | K02372                     | At2g22230 | HG047 | F0YQ60 | HG048  | A8IX17 | E1Z8J0           | I0YWZ4 |
|  | [3R]-hydroxymyristoyl-[acyl carrier protein] dehydratase-like protein | K02372                     | At5g10160 | HG053 | F0YQ60 | HG054  | A8IX17 | E1Z8J0           | I0YWZ4 |
| Enoyl-ACP Reductase (EC 1.3.1.9)       | enoyl-ACP reductase (enr-A)   | K00208                     | At2g05990 | HG058 | F0XZA4 | HG059  | HG060  | E1Z2Y2           | I0Z0U2 |
| Stearoyl-ACP Desaturase (EC 1.14.19.2) | stearoyl-ACP desaturase (Fab 2)                                       | K03921                     | At2g43710 | HG064 |        |        | A8IQB8 | E1ZCJ2<br>E1Z8P9 | I0YV07 |
|  | putative stearoyl-acyl carrier protein desaturase                     | K03921                     | At3g02620 | HG070 |        |        | A8IQB8 | E1ZCJ2           | I0YV07 |
|  | putative stearoyl-acyl carrier protein desaturase                     | K03921                     | At3g02630 | HG075 |        |        | A8IQB8 | E1ZCJ2           | I0YV07 |
|  | stearoyl-acyl carrier protein desaturase                              | K03921                     | At5g16230 | HG080 |        |        | A8IQB8 | E1ZCJ2           | I0YV07 |
| Acyl Carrier Protein (ACP)             | acyl carrier protein 1 precursor (ACP)                                | K01961<br>K02078<br>K03955 | At3g05020 | HG085 | 0YQ15  | Q7XYK4 | Q6UKY5 | E1Z5W8<br>E1Z6T8 | I0Z6E0 |
|  | ACP2  | K02078                     | At1g54580 | HG089 | F0YQ15 | Q7XYK4 | Q6UKY5 | E1Z5W8           | I0Z6E0 |

|  |                           |                  |           |       |                  |        |        |                  |                  |
|--|---------------------------|------------------|-----------|-------|------------------|--------|--------|------------------|------------------|
|  |                           | K03955           |           |       | F0Y827           |        |        |                  | I0YSD5           |
|  | hypothetical protein      | K02078<br>K03955 | At1g54630 | HG094 |                  | Q7XYK4 | Q6UKY5 | E1Z5W8           | I0Z6E0           |
|  | acyl carrier-like protein | K03955<br>K02078 | At4g25050 | HG100 | F0YQ15<br>F0Y827 | Q7XYK4 | Q6UKY5 | E1Z5W8<br>E1Z6T8 | I0Z6E0<br>I0YSD5 |

**HG:** Hypothetical gene.

**Table 2:** Candidate genes involved in Fatty Acid biosynthesis in Plastids of *Arabidopsis thaliana*, *Emiliana huxleyi* CCMP1516 main genome assembly v1.0, *Fragilariopsis cylindrus*, *Guillardia theta* CCMP2712 v1.0., *Micromonas pusilla* CCMP1545 v3.0, *Micromonas sp.* RCC299 v3.0, *Ostreococcus lucimarinus* v2.0, *Ostreococcus sp.* RCC809 2.0, *Ostreococcus tauri* v2.0, *Phaeodactylum tricornutum* v2.0, *Pseudo-nitzschia multiseriata* CLN-47 v1.0, *Thalassiosira pseudonana* and *Volvox carteri f. nagariensis*.

| Protein Code                | Corresponding homologous predicted protein in algal species(SwissProt accession ID) |                                 |                                       |   |                                   |                                      |                                    |                                |                                       |  |                                 |                                      |
|-----------------------------|---|---------------------------------|---------------------------------------|---|-----------------------------------|--------------------------------------|------------------------------------|--------------------------------|---------------------------------------|--|---------------------------------|--------------------------------------|
| <i>Arabidopsis thaliana</i> | <i>Emiliana huxleyi</i> CCMP1516 main genome assembly v1.0                          | <i>Fragilariopsis cylindrus</i> | <i>Guillardia theta</i> CCMP2712 v1.0 | <i>Micromonas pusilla</i> CCMP1545 v3.0 | <i>Micromonas sp.</i> RCC299 v3.0 | <i>Ostreococcus lucimarinus</i> v2.0 | <i>Ostreococcus sp.</i> RCC809 2.0 | <i>Ostreococcus tauri</i> v2.0 | <i>Phaeodactylum tricornutum</i> v2.0 | <i>Pseudo-nitzschia multiseriata</i> CLN-47 v1.0 | <i>Thalassiosira pseudonana</i> | <i>Volvox carteri f. nagariensis</i> |
| At1g36180                   | R1FFT9  | HG004                           |                                       | C1ML75                                  |                                   |                                      | HG005                              |                                | B7GEB5                                | HG006  | B8BVD1                          | D8UA31                               |
| At2g38040                   |   |                                 |                                       |   |                                   |                                      |                                    |                                |                                       |  |                                 | D8TNY0                               |
| AtCg00500                   | R1DCW3  | HG011                           | L1K4J6                                |   |                                   |                                      |                                    |                                | B7FZ17                                | HG012  | B8BWH0                          | D8U455                               |
| At5g16390                   |   | HG013                           |                                       |   |                                   |                                      |                                    |                                |                                       |  |                                 | D8U256                               |
| At5g15530                   |   |                                 | L1JIA2                                |   |                                   |                                      |                                    |                                |                                       |  |                                 | D8U256                               |
| At5g3536                    | R1DFR1  | HG016                           | L1JIA2                                | C1MSW                                   |                                   |                                      | HG017                              |                                | B5Y589                                | HG018  | B8LCZ8                          | D8UF54                               |

|           |        |       |                  |                            |                  |  |       |                  |        |       |                  |                  |
|-----------|--------|-------|------------------|----------------------------|------------------|--|-------|------------------|--------|-------|------------------|------------------|
| 0         | R1FUG6 |       |                  | 8                          |                  |  |       |                  | B7GCL6 |       | B8C7R3           | D8THK5           |
| At2g30200 | R1FRD8 | HG022 | L1JFY1           | C1MGN5<br>C1FD34           |                  |  | HG023 | HG024            | B7G3D4 | HG025 | B8C2A9           | D8TTQ7           |
| At5g46290 | R1DM58 | HG028 | L1K1Y4           | C1MVN2<br>C1MYJ6<br>C1N0D7 | C1EC96<br>C1EET5 |  | HG029 | Q01EI4<br>Q00V56 | B7G2W2 | HG030 | B8C1R1           | D8UDW0<br>D8TXC7 |
| At1g74960 | R1DM58 | HG033 | L1K1Y4           | C1MYJ6                     | C1EET5           |  | HG034 | Q01EI4<br>Q00V56 | B7G2W2 | HG035 | B8C1R1           | D8TXC7           |
| At1g62640 | R1G8H9 | HG039 | L1JJ78           | C1MYF8                     | C1EEV4           |  | HG040 | Q00V15           |        | HG041 | B8CGL6           | D8TXF1           |
| At1g24360 | R1CUW5 | HG044 | L1JTF3<br>L1JZG9 | C1MKF9                     | C1FDQ2           |  | HG045 | Q01GL3           | B7G1R8 | HG046 | B8BTK3<br>B8BYY5 | B8BYY5           |
| At2g22230 | R1CAN9 | HG049 | L1ITP1           | C1MJC2                     | C1DZI2           |  | HG050 | HG051            | B7S3L6 | HG052 | B8C490           | D8TV61           |
| At5g10160 | R1CAN9 | HG055 | L1ITP1           | C1MJC2                     | C1DZI2           |  | HG056 | HG057            | B7S3L6 |       | B8C490           | D8TV61           |
| At2g05990 | R1DPI9 | HG061 | L1IM03           | C1MZE1                     | C1E829           |  | HG062 | Q014N2           | B7FS72 | HG063 | B8BXA1           | D8UC03           |
| At2g43710 |        | HG065 |                  | HG066                      | C1EDY7           |  | HG067 | HG068            | B7FQK1 | HG069 |                  | D8TTF7           |

|           |                            |       |        |        |                  |  |        |        |                  |       |                  |        |
|-----------|----------------------------|-------|--------|--------|------------------|--|--------|--------|------------------|-------|------------------|--------|
| At3g02620 |                            | HG071 |        | C1MXN3 | C1EDY7           |  | HG072  | HG073  | B7FQK1           | HG074 |                  | D8TTF7 |
| At3g02630 |                            | HG076 |        | C1MXN3 | C1EDY7           |  | HG077  | HG078  | B7FQK1           | HG079 |                  | D8TTF7 |
| At5g16230 |                            | HG081 | L1J501 | C1MXN3 | C1EDY7           |  | HG082  | HG083  | B7FQK1           | HG084 |                  | D8TTF7 |
| At3g05020 | R1FZ54<br>R1D0Y2           | HG086 | L1J501 | HG087  | C1E7T0           |  | A4S5H6 | HG088  | B7FRX6           |       | B8C103<br>B8C5G6 | D8U1L7 |
| At1g54580 | R1FZ54                     | HG090 | L1J501 | HG091  | C1E7T0           |  | HG092  | Q00SL4 | B7FRC2           | HG093 | B8C5G6           | D8U1L7 |
| At1g54630 | R1EKZ4                     | HG095 | L1J501 | HG096  | C1E7T0           |  | HG097  | HG098  | B7FRX6           | HG099 | B8C5G6<br>B8C103 | D8U1L7 |
| At4g25050 | R1FZ54<br>R1D0Y2<br>R1EKZ4 | HG101 |        | C1N9A6 | C1E7T0<br>C1E2X1 |  | A4S5H6 | Q00SL4 | B7FRX6<br>B7FRC2 | HG102 | B8C5G6<br>B8C103 | D8U1L7 |

**HG:** Hypothetical gene.

**Table 3:** Various physico-chemical charactersexhibited by putative proteins encoded by genes involved in fatty acid biosynthesis pathway in *A. thaliana*, *Asterochloris sp. Cgr/DA1pho* v1.0, *Aureococcus anophagefferens*, *Bigelowiella natans* CCMP2755 v1.0, *Chlamydomonas reinhardtii* v4.0, *Chlorella variabilis* NC64A, *Coccomyxa subellipsoidea* C-169 v2.0, *Emiliana huxleyi* CCMP1516 main genome assembly v1.0, *Fragilariopsis cylindrus*, *Guillardia theta* CCMP2712 v1.0, *Micromonas pusilla* CCMP1545 v3.0, *Micromonas sp. RCC299* v3.0, *Ostreococcus lucimarinus* v2.0, *Ostreococcus sp. RCC809* 2.0, *Ostreococcus tauri* v2.0, *Phaeodactylum tricornutum* v2.0, *Pseudo-nitzschia multiseriis* CLN-47 v1.0, *Thalassiosira pseudonana* and *Volvox carteri f. nagariensis*, as calculated by ProtParam server.

| Enzyme   | Microalgae Genome                     | Protein code | No of amino acids | Mol wt   | PI   | Total no of +ve residues | Total no of -ve residues | Instability Index | Aliphatic Index | GRAVY  |
|--|---------------------------------------|--------------|-------------------|----------|------|--------------------------|--------------------------|-------------------|-----------------|--------|
| Homomeric Acetyl-CoA Carboxylase (ACCase) (EC 6.4.1.2) | <i>Asterochloris sp. Cgr/DA1pho</i>   | HG001        | 2140              | 232769.3 | 6.06 | 199                      | 227                      | 47.71             | 91.15           | -0.083 |
|  | <i>Bigelowiella natans</i> CCMP2755   | HG002        | 2053              | 227513.1 | 8.63 | 251                      | 234                      | 40.43             | 86.80           | -0.249 |
|  | <i>Coccomyxa subellipsoidea</i> C-169 | HG003        | 2366              | 254795.2 | 5.92 | 235                      | 272                      | 42.65             | 93.77           | -0.071 |
|  | <i>Fragilariopsis cylindrus</i>       | HG004        | 2302              | 253210.7 | 5.39 | 241                      | 298                      | 37.10             | 90.70           | -0.183 |
|  | <i>Ostreococcus sp. RCC809</i>        | HG005        | 2013              | 222317.8 | 5.16 | 223                      | 287                      | 38.48             | 86.83           | -0.244 |
|  | <i>Pseudo-nitzschia multiseriis</i>   | HG006        | 2297              | 252797.5 | 5.13 | 245                      | 313                      | 37.09             | 87.89           | -0.211 |

|  |   |       |     |         |      |    |    |       |        |        |
|--|---|-------|-----|---------|------|----|----|-------|--------|--------|
|  | CLN-47  |       |     |         |      |    |    |       |        |        |
| Carboxyltrans<br>fe<br>rase of<br>Heteromeric<br>ACCcase | <i>Asterochloris sp.</i><br>Cgr/DA1pho        | HG007 | 350 | 38823.6 | 6.73 | 44 | 45 | 35.18 | 90.63  | -0.207 |
| Carboxyltrans<br>fe<br>rase of                           | <i>Asterochloris sp.</i><br>Cgr/DA1pho        | HG008 | 343 | 38745.7 | 9.44 | 47 | 34 | 38.38 | 89.53  | -0.317 |
| Heteromeric<br>ACCcase                                   | <i>Bigelowiella natans</i><br>CCMP2755        | HG009 | 495 | 54404.3 | 6.22 | 57 | 62 | 42.69 | 87.23  | -0.196 |
| (plastid<br>encoded)                                     | <i>Coccomyxa subellipsoidea</i><br>C-169      | HG010 | 72  | 8204.5  | 4.88 | 7  | 11 | 49.46 | 112.50 | 0.065  |
|  | <i>Fragilariopsis cylindrus</i>               | HG011 | 575 | 63326.0 | 5.87 | 63 | 72 | 33.63 | 83.93  | -0.226 |
|  | <i>Pseudo-nitzschia multiseriis</i><br>CLN-47 | HG012 | 135 | 14674.2 | 9.11 | 12 | 9  | 33.90 | 109.85 | 0.295  |

|   |                                     |       |      |          |      |     |     |       |       |        |
|---|-------------------------------------|-------|------|----------|------|-----|-----|-------|-------|--------|
| Biotin Carboxyl Carrier Protein of Heteromeric ACCase | <i>Fragilariopsis cylindrus</i>     | HG013 | 731  | 81184.3  | 5.00 | 69  | 102 | 44.50 | 89.23 | -0.410 |
| Biotin Carboxylase of Heteromeric ACCase              | <i>Asterochloris sp. Cgr/DA1pho</i> | HG014 | 514  | 56716.6  | 5.72 | 57  | 70  | 36.48 | 87.70 | -0.299 |
|   | <i>Bigelowiella natans</i> CCMP2755 | HG015 | 623  | 68654.6  | 4.84 | 57  | 93  | 47.20 | 83.15 | -0.308 |
|   | <i>Fragilariopsis cylindrus</i>     | HG016 | 1159 | 126659.7 | 5.01 | 121 | 164 | 33.25 | 85.82 | -0.214 |
|   | <i>Ostreococcus sp. RCC809</i>      | HG017 | 1132 | 123211.9 | 5.26 | 121 | 155 | 34.07 | 87.42 | -0.177 |
|   | <i>Ostreococcus tauri</i>           | HG018 | 1163 | 126361.0 | 5.50 | 128 | 153 | 32.60 | 90.21 | -0.142 |
| Malonyl-CoA : ACP Malonyltransferase (EC 2.3.1.39)    | <i>Asterochloris sp. Cgr/DA1pho</i> | HG019 | 335  | 35619.1  | 8.46 | 37  | 34  | 35.51 | 94.45 | 0.019  |
|   | <i>Bigelowiella natans</i> CCMP2755 | HG020 | 421  | 45194.9  | 5.60 | 44  | 50  | 36.18 | 85.80 | -0.013 |
|   | <i>Chlorella variabilis</i> NC64A   | HG021 | 225  | 23458.0  | 9.69 | 31  | 20  | 36.82 | 85.47 | -0.113 |
|   | <i>Fragilariopsis cylindrus</i>     | HG022 | 300  | 31573.3  | 4.81 | 30  | 42  | 27.25 | 88.50 | 0.109  |
|   | <i>Ostreococcus sp. RCC809</i>      | HG023 | 294  | 30745.2  | 4.83 | 29  | 40  | 24.11 | 89.12 | 0.068  |

|  |   |       |     |         |      |    |    |       |       |        |
|--|---|-------|-----|---------|------|----|----|-------|-------|--------|
|  | <i>Ostreococcus tauri</i>                     | HG024 | 290 | 30404.9 | 4.84 | 28 | 39 | 23.03 | 90.34 | 0.096  |
|  | <i>Pseudo-nitzschia multiseriis</i><br>CLN-47 | HG025 | 300 | 31407.1 | 4.75 | 28 | 41 | 27.43 | 91.47 | 0.167  |
| Ketoacyl-ACP<br>Synthase I<br><br>(KAS I) (EC<br>2.3.1.41) | <i>Asterochloris sp.</i><br>Cgr/DA1pho        | HG026 | 411 | 43674.6 | 5.82 | 36 | 43 | 33.54 | 81.00 | -0.072 |
|  | <i>Bigelowiella natans</i><br>CCMP2755        | HG027 | 449 | 47196.6 | 5.56 | 39 | 52 | 40.59 | 91.00 | -0.008 |
|  | <i>Fragilariopsis cylindrus</i>               | HG028 | 463 | 48432.7 | 5.26 | 39 | 50 | 34.85 | 71.30 | -0.098 |
|  | <i>Ostreococcus sp.</i> RCC809                | HG029 | 422 | 44725.8 | 5.59 | 39 | 49 | 32.67 | 81.00 | -0.073 |
|  | <i>Pseudo-nitzschia multiseriis</i><br>CLN-47 | HG030 | 443 | 46875.1 | 5.23 | 42 | 55 | 34.41 | 72.10 | -0.128 |

|  |   |       |      |          |      |    |     |       |       |        |
|--|---|-------|------|----------|------|----|-----|-------|-------|--------|
| KAS II (EC<br>2.3.1.41)                        | <i>Asterochloris sp.</i><br>Cgr/DA1pho        | HG031 | 1360 | 139494.7 | 6.19 | 93 | 106 | 46.13 | 97.54 | 0.296  |
|  | <i>Bigelowiella natans</i><br>CCMP2755        | HG032 | 449  | 47196.6  | 5.56 | 39 | 52  | 40.59 | 91.00 | -0.008 |
|  | <i>Fragilariopsis cylindrus</i>               | HG033 | 463  | 48432.7  | 5.26 | 39 | 50  | 34.85 | 71.30 | -0.098 |
|  | <i>Ostreococcus sp.</i> RCC809                | HG034 | 412  | 43395.3  | 5.20 | 35 | 49  | 27.56 | 84.17 | 0.024  |
|  | <i>Pseudo-nitzschia multiseriis</i><br>CLN-47 | HG035 | 443  | 46875.1  | 5.23 | 42 | 55  | 34.41 | 72.10 | -0.128 |
| KAS III (EC<br>2.3.1.41)                       | <i>Asterochloris sp.</i><br>Cgr/DA1pho        | HG036 | 342  | 36296.2  | 5.45 | 30 | 38  | 29.75 | 93.39 | 0.010  |
|  | <i>Bigelowiella natans</i><br>CCMP2755        | HG037 | 418  | 43710.3  | 5.94 | 38 | 43  | 30.55 | 84.55 | -0.034 |
|  | <i>Chlamydomonas reinhardtii</i>              | HG038 | 374  | 38909.3  | 8.46 | 36 | 33  | 34.80 | 86.42 | 0.055  |
|  | <i>Fragilariopsis cylindrus</i>               | HG039 | 367  | 38801.6  | 4.64 | 28 | 45  | 29.72 | 88.34 | -0.065 |
|  | <i>Ostreococcus sp.</i> RCC809                | HG040 | 378  | 39822.1  | 7.68 | 40 | 39  | 27.48 | 89.66 | 0.046  |
|  | <i>Pseudo-nitzschia multiseriis</i><br>CLN-47 | HG041 | 339  | 35990.6  | 4.57 | 27 | 44  | 25.02 | 92.68 | -0.031 |
| Ketoacyl-ACP<br>Reductase<br>(EC<br>1.1.1.100) | <i>Asterochloris sp.</i><br>Cgr/DA1pho        | HG042 | 273  | 28398.5  | 9.06 | 29 | 24  | 34.57 | 87.95 | 0.033  |
|  | <i>Bigelowiella natans</i><br>CCMP2755        | HG043 | 268  | 27622.6  | 7.76 | 26 | 25  | 29.25 | 92.91 | 0.179  |
|  | <i>Fragilariopsis cylindrus</i>               | HG044 | 253  | 26628.5  | 5.96 | 26 | 28  | 22.67 | 91.03 | 0.106  |
|  | <i>Ostreococcus sp.</i> RCC809                | HG045 | 273  | 28084.6  | 9.18 | 29 | 22  | 23.81 | 97.33 | 0.251  |

|                                  |  |       |     |         |      |    |    |       |        |        |
|----------------------------------|--|-------|-----|---------|------|----|----|-------|--------|--------|
|                                  |  |       |     |         |      |    |    |       |        |        |
|                                  | <i>Pseudo-nitzschia multiseriata</i><br>CLN-47 | HG046 | 253 | 26516.4 | 5.35 | 26 | 29 | 22.00 | 92.61  | 0.142  |
| Hydroxyacyl-<br>ACP<br>Dehydrase | <i>Asterochloris sp.</i><br>Cgr/DA1pho         | HG047 | 205 | 22933.6 | 9.69 | 29 | 18 | 41.81 | 75.66  | -0.311 |
|                                  |  | HG053 | 205 | 22933.6 | 9.69 | 29 | 18 | 41.81 | 75.66  | -0.311 |
| (EC 4.2.1.*)                     | <i>Bigeloviella natans</i><br>CCMP2755         | HG048 | 146 | 15969.7 | 6.90 | 15 | 15 | 32.89 | 95.41  | 0.140  |
|                                  |  | HG054 | 146 | 15969.7 | 6.90 | 15 | 15 | 32.89 | 95.41  | 0.140  |
|                                  | <i>Fragilariopsis cylindrus</i>                | HG049 | 183 | 19913.3 | 6.16 | 22 | 23 | 25.04 | 104.26 | 0.117  |
|                                  |  | HG055 | 183 | 19913.3 | 6.16 | 22 | 23 | 25.04 | 104.26 | 0.117  |

|   |   |       |     |         |      |    |    |       |       |        |
|---|---|-------|-----|---------|------|----|----|-------|-------|--------|
|   | <i>Ostreococcus sp.</i> RCC809                | HG050 | 146 | 15912.7 | 5.37 | 14 | 17 | 36.45 | 96.71 | 0.236  |
|   |   | HG056 | 146 | 15912.7 | 5.37 | 14 | 17 | 36.45 | 36.45 | 0.236  |
|   | <i>Ostreococcus tauri</i>                     | HG051 | 146 | 15955.9 | 6.73 | 16 | 16 | 33.31 | 96.03 | 0.233  |
|   |   | HG057 | 146 | 15955.9 | 6.73 | 16 | 16 | 33.31 | 96.03 | 0.233  |
|   | <i>Pseudo-nitzschia multiseriis</i><br>CLN-47 | HG052 | 207 | 21983.4 | 5.27 | 20 | 25 | 26.94 | 93.72 | 0.125  |
| Enoyl-ACP<br>Reductase<br>(EC<br>1.3.1.9)       | <i>Asterochloris sp.</i><br>Cgr/DA1pho        | HG058 | 300 | 31636.0 | 6.85 | 32 | 32 | 26.20 | 89.93 | 0.015  |
|   | <i>Bigelowiella natans</i><br>CCMP2755        | HG059 | 397 | 42328.3 | 8.15 | 41 | 39 | 36.27 | 86.78 | -0.012 |
|   | <i>Chlamydo-monas reinhardtii</i>             | HG060 | 374 | 39347.0 | 9.49 | 44 | 34 | 29.93 | 91.66 | 0.057  |
|   | <i>Fragilariopsis cylindrus</i>               | HG061 | 363 | 38307.6 | 5.42 | 34 | 38 | 29.15 | 84.24 | -0.010 |
|   | <i>Ostreococcus sp.</i> RCC809                | HG062 | 335 | 35550.5 | 8.85 | 39 | 35 | 27.15 | 85.73 | -0.079 |
|   | <i>Pseudo-nitzschia multiseriis</i><br>CLN-47 | HG063 | 363 | 38176.2 | 5.29 | 34 | 40 | 30.88 | 85.81 | -0.022 |
| Stearoyl-ACP<br>Desaturase<br>(EC<br>1.14.19.2) | <i>Asterochloris sp.</i><br>Cgr/DA1pho        | HG064 | 431 | 47664.0 | 6.51 | 50 | 53 | 33.94 | 76.13 | -0.471 |
|   |   | HG070 | 431 | 47664.0 | 6.51 | 50 | 53 | 33.94 | 76.13 | -0.471 |
|   |   | HG075 | 431 | 47664.0 | 6.51 | 50 | 53 | 33.94 | 76.13 | -0.471 |
|   |   | HG080 | 431 | 47664.0 | 6.51 | 50 | 53 | 33.94 | 76.13 | -0.471 |
|   | <i>Fragilariopsis cylindrus</i>               | HG065 | 431 | 48924.4 | 5.05 | 49 | 70 | 43.44 | 76.71 | -0.435 |
|   |   | HG071 | 431 | 48924.4 | 5.05 | 49 | 70 | 43.44 | 76.71 | -0.435 |
|   |   | HG076 | 431 | 48924.4 | 5.05 | 49 | 70 | 43.44 | 76.71 | -0.435 |
|   |   | HG081 | 431 | 48924.4 | 5.05 | 49 | 70 | 43.44 | 76.71 | -0.435 |
|   | <i>Micromon-as pusilla</i><br>CCMP1545        | HG066 | 342 | 38486.7 | 5.41 | 42 | 53 | 33.50 | 75.70 | -0.496 |
|   | <i>Ostreococcus sp.</i> RCC809                | HG067 | 329 | 37052.8 | 5.45 | 39 | 49 | 33.13 | 70.12 | -0.540 |

|   |       |     |         |      |    |    |       |       |        |
|---|-------|-----|---------|------|----|----|-------|-------|--------|
|   | HG072 | 329 | 37052.8 | 5.45 | 39 | 49 | 33.13 | 70.12 | -0.540 |
|   | HG077 | 329 | 37052.8 | 5.45 | 39 | 49 | 33.13 | 70.12 | -0.540 |
|   | HG082 | 329 | 37052.8 | 5.45 | 39 | 49 | 33.13 | 70.12 | -0.540 |
| <i>Ostreococcus tauri</i>                     | HG068 | 345 | 38623.8 | 5.32 | 42 | 53 | 34.54 | 77.01 | -0.460 |
|   | HG073 | 345 | 38623.8 | 5.32 | 42 | 53 | 34.54 | 77.01 | -0.460 |
|   | HG078 | 345 | 38623.8 | 5.32 | 42 | 53 | 34.54 | 77.01 | -0.460 |
|   | HG083 | 345 | 38623.8 | 5.32 | 42 | 53 | 34.54 | 77.01 | -0.460 |
| <i>Pseudo-nitzschia multiseriis</i><br>CLN-47 | HG069 | 345 | 39375.4 | 4.83 | 41 | 62 | 42.00 | 76.90 | -0.550 |
|   | HG074 | 345 | 39375.4 | 4.83 | 41 | 62 | 42.00 | 76.90 | -0.550 |
|   | HG079 | 345 | 39375.4 | 4.83 | 41 | 62 | 42.00 | 76.90 | -0.550 |
|   | HG084 | 345 | 39375.4 | 4.83 | 41 | 62 | 42.00 | 76.90 | -0.550 |

|                            |   |       |     |         |      |    |    |       |        |        |
|----------------------------|---|-------|-----|---------|------|----|----|-------|--------|--------|
| Acyl Carrier Protein (ACP) | <i>Asterochloris sp.</i><br>Cgr/DA1pho        | HG085 | 161 | 17997.3 | 4.72 | 17 | 30 | 38.57 | 95.16  | -0.357 |
|                            |   | HG089 | 81  | 9016.1  | 4.09 | 8  | 20 | 34.85 | 92.84  | -0.422 |
|                            |   | HG094 | 81  | 9016.1  | 4.09 | 8  | 20 | 34.85 | 92.84  | -0.422 |
|                            |   | HG100 | 81  | 9016.1  | 4.09 | 8  | 20 | 34.85 | 92.84  | -0.422 |
|                            | <i>Fragilariopsis cylindrus</i>               | HG086 | 121 | 13126.9 | 4.60 | 12 | 19 | 29.85 | 105.54 | 0.183  |
|                            |   | HG090 | 121 | 13126.9 | 4.60 | 12 | 19 | 29.85 | 105.54 | 0.183  |
|                            |   | HG095 | 102 | 10829.2 | 3.81 | 5  | 20 | 44.17 | 119.51 | 0.505  |
|                            |   | HG101 | 121 | 13126.9 | 4.60 | 12 | 19 | 29.85 | 105.54 | 0.183  |
|                            | <i>Micromonas pusilla</i><br>CCMP1545         | HG087 | 123 | 13081.0 | 4.52 | 13 | 20 | 29.83 | 100.16 | 0.115  |
|                            |   | HG091 | 123 | 13081.0 | 4.52 | 13 | 20 | 29.83 | 100.16 | 0.115  |
|                            |   | HG096 | 123 | 13081.0 | 4.52 | 13 | 20 | 29.83 | 100.16 | 0.115  |
|                            | <i>Ostreococcus tauri</i>                     | HG088 | 89  | 9445.6  | 4.35 | 9  | 16 | 20.97 | 95.39  | 0.090  |
|                            |   | HG098 | 89  | 9445.6  | 4.35 | 9  | 16 | 20.97 | 95.39  | 0.090  |
|                            | <i>Pseudo-nitzschia multiseriis</i><br>CLN-47 | HG093 | 121 | 12890.6 | 4.55 | 12 | 19 | 20.81 | 99.17  | 0.226  |
|                            |   | HG099 | 121 | 12890.6 | 4.55 | 12 | 19 | 20.81 | 99.17  | 0.226  |
|                            |   | HG102 | 106 | 11211.6 | 3.72 | 4  | 20 | 41.70 | 110.47 | 0.435  |

**HG:** Hypothetical gene.

