**Chloroplast-Targeting Protein Expression in the Oleaginous Diatom *Fistulifera solaris* JPCC
DA0580 toward Metabolic Engineering**

[Running title] Protein expression in the diatom chloroplast

**Yoshihiko Sunaga^{1,2}, Yoshiaki Maeda¹, Takashi Yabuuchi¹, Masaki Muto^{1,2}, Tomoko Yoshino¹
and Tsuyoshi Tanaka^{1,2,*}**

¹ Division of Biotechnology and Life Science, Institute of Engineering, Tokyo University of
Agriculture and Technology, 2-24-16, Naka-cho, Koganei, Tokyo 184-8588, Japan

² Japan Science and Technology Agency (JST), Core Research for Evolutionary Science and
Technology (CREST), 5, Sanbancho, Chiyoda-ku, Tokyo 102-0075, Japan

*Author to whom correspondence should be addressed

2-24-16, Naka-cho, Koganei, Tokyo 184-8588, Japan

E-mail; tsuyo@cc.tuat.ac.jp

Tel.: +81-42-388-7401; Fax: +81-42-385-7713.

1   **Abstract:**

2   The chloroplast plays critical roles in lipid metabolism of microalgae, thus it is recognized as an
3   attractive target of metabolic engineering to enhance biofuel production. It has been well known that
4   recombinant protein expression in microalgal chloroplasts needs specific signal sequence which
5   governs the transition manner of nuclear-encoded polypeptides within the subcellular compartments.
6   However certain microalgae, including diatoms, have complex membrane systems surrounding the
7   chloroplast, and thus chloroplast-targeting protein expression with the signal sequence has rarely
8   been demonstrated except for a few model non-oleaginous diatoms. In this study, we performed
9   recombinant green fluorescence protein (GFP) expression and transportation into the chloroplast of
10   the oleaginous marine diatom, *Fistulifera solaris* JPCC DA0580. The signal sequence of ATP
11   synthetase gamma subunit, which was predicted to localize in the chloroplast according to a
12   bioinformatics analysis pipeline, was employed as a key factor of this technique. As a result, specific
13   localization of GFP in the chloroplast was observed. It would be useful to engineer the lipid synthesis
14   pathways existing in the chloroplast. Furthermore, intensive gathering of GFP in the rod-like
15   structure was also detected, which has not been observed in model diatom studies. As comparing
16   with electron microscopic observation, the structure was estimated to be a pyrenoid.

17

Keywords: chloroplast; signal sequence; pyrenoid; *Fistulifera solaris* JPCC DA0580;
marine oleaginous diatom

Introduction

With an increased demand for sustainable energy supply, biofuel production has attracted much attention. Microalgal biodiesel production has been researched to meet such demand due to its advantageous features (e.g., global carbon dioxide fixation, no competition with foods, much higher biomass yield than higher plants, and oil accumulation at a high level inside the cells) (1). Several eukaryotic microalgae can accumulate triacylglycerol (TAG) in high level as a form of oleosomes (also known as oil bodies), and some promising oil producers have intensively been studied to understand the TAG biosynthesis (2, 3). We recently discovered an oleaginous diatom, *Fistulifera solaris* JPCC DA0580 from our marine microalgal culture collection (4, 5). A beneficial feature of this diatom for biodiesel production is high total lipid content (~60%, w/w) (6), leading to high oil productivity (7). Bench-scale outdoor mass-cultivation was already studied using race-way and column bioreactors (8).

TAG accumulation in microalgae can be promoted by environmental stimuli including depletion of nutrition, while excess depletion tends to inhibit cell growth. Thus the oil production process must be carefully designed in order to evaluate the maximal potential of the microalgal strains of interest (7). Another possible strategy is metabolic engineering. Overexpression of the native or heterogeneous enzymes could allow us to exploit the further potential of microalgae. To apply the metabolic engineering to *F. solaris*, whole genome analysis of (Tanaka, T. et al, manuscript submitted) and establishment of genetic transformation system (9) should be helpful.

Chloroplasts are responsible for or take part in lipid biosynthesis including primary synthesis, elongation, desaturation of fatty acids, and also phospholipids and TAG assembly in eukaryotic autotrophic organisms including higher plants and microalgae (10). These important reactions heavily influence the productivity and quality of biofuel; e.g., desaturation degree of acyl chains

1 determines the oxidative stability of biodiesel fuel (11). Given this fact, enzymes working in the
2 chloroplasts can be primary targets of metabolic engineering in order to evaluate the capacity of oil
3 accumulation in the cells, or redirect the carbon flux into biosynthesis of the metabolites of interests.
4 For this purpose, in higher plants, methodology for chloroplast-targeting recombinant protein
5 expression have been established with chloroplast-specific signal sequences, and engineering of lipid
6 biosynthesis pathways has been successfully demonstrated (12-14).

7 In contrast, chloroplast-targeting metabolic engineering in microalgae, particularly in diatoms, is
8 still rare. Although one of the earliest efforts toward lipid metabolic engineering was overexpressing
9 a native plastidal enzyme, acetyl-CoA carboxylase in a diatom, *Cyclotella cryptica* (15), subsequent
10 research have not been intensively continued. Difficulty in specific protein expression in the diatom
11 chloroplast is attributable to its complex membrane structure, in which 4 bilayer membranes
12 envelope the chloroplasts (16). The innermost and 2nd innermost membranes actually correspond to
13 the 2 membranes of the primary chloroplast. The space between the 2nd outermost membrane and the
14 outermost membrane represents the ER (also known as chloroplast-ER; CER). Nuclear-encoded
15 chloroplast-targeting polypeptides must pass through these 4 membrane barriers. Although the
16 transition mechanism in the diatom chloroplast has not fully been elucidated (12, 17), it is known
17 that bipartite signal sequences consisting of an ER-signal and a transit domain play a critical role in
18 the transition (18-21). The ER-signal sequence directs proteins to ER, followed by translocation into
19 the inner chloroplast with the aid of the transit domain. However, thus far, methodology for
20 chloroplast-targeting recombinant protein expression with nuclear-encoded polypeptides has been
21 demonstrated only in a model, non-oleaginous diatom, *Phaeodactylum tricornutum* (20). Other
22 “non-model” diatoms have not been studied in such research despite extremely vast diversity in
23 diatom species (more than 200,000 different species (14), and considered to be responsible for around
24 one-fifth of the primary productivity of the Earth (22, 23)). One of the reasons why only a few model
25 diatoms have been intensively investigated is due to the fact that there are only two diatoms whose

genomes were sequenced (24, 25). This has long hindered investigation in most other “non-model” diatoms possessing unique and attractive phenotypes. Toward engineering biofuel production, the organelle-specific targeting manner must be tested in oleaginous species.

In this study, chloroplast-targeting protein expression was studied in the oleaginous diatom *F. solaris* using the signal sequence of an endogenous chloroplast protein. Prior to the experiment, we determined the pipeline of a series of *in silico* analyses to predict the subcellular localization of proteins of interest. It is helpful to assess the selected sequence as a chloroplast-targeting signal. Then, as a proof of concept, the fusion protein of green fluorescence protein (GFP) and the selected signal sequence was constructed, and specific chloroplast targeting was demonstrated. However its distribution was uneven in the chloroplast; GFP gathering was observed in the rod-like structure which was estimated to be a pyrenoid. Because this observation was different from the results in a model, non-oleaginous diatom, the chloroplast-targeting manner in the present study might be a unique feature of *F. solaris*. To our best knowledge, this is the first report to show recombinant protein-targeting into the complex chloroplast in the oleaginous diatom.

Materials and methods

Prediction Pipeline

Localization of proteins was predicted using several bioinformatics programs. TargetP (26) and HECTAR (27) were used for predicting general protein localization. Screening for endoplasmic reticulum (ER) or chloroplast-targeting peptides was performed by using SignalP (28). When cleavage sites were predicted by SignalP, the presence of an ER retention signal (K(/D)-D(/E)-E-L) in the protein C-terminus was checked manually (29). If proteins possessed no ER retention signals and contained F, W, Y, or L at the +1 position of cleavage sites, they were considered to be chloroplast-targeted proteins (21). When proteins were predicted to be localized in the mitochondria by both TargetP and HECTAR, they were considered to be transported into mitochondria. In addition,

Mitoprot (30) was used to detect mitochondrial-targeting peptides. If (i) the score of Mitoprot was >0.9 or (ii) the score of Mitoprot was >0.8, and mitochondrial localization was predicted by either TargetP or HECTAR, the protein was also considered to be transported into mitochondria. For proteins without signal, chloroplast-targeting, or mitochondrial-targeting peptides, peroxisome-targeting signals at the C-terminus (S(/A/C)-K(/R/H)-L(/M) or S-S-L) were checked manually (31). A transmembrane prediction program (TMHMM) (32) was used to detect the presence of transmembrane regions. If transmembrane regions were predicted in proteins not containing other targeting peptides, the possibility that the proteins might be localized in the ER could not be excluded. Proteins that conformed to none of the rules described above were predicted to be localized in the cytoplasm.

Culture conditions

The marine diatom, *F. solaris* JPCC DA0580 was isolated from a mangrove swamp located in Sumiyo Bay below the mouths of the Sumiyo and Yakugachi Rivers in Amami-Oshima Kagoshima, Japan (28°15'N, 129°24'E) (4, 5). The culture was maintained in the f/2 medium (33) dissolved in artificial seawater. Transformant cells were incubated in the f/2 medium with antibiotics G418 (500 µg/mL). Cultures were aerated with sterile air at 25°C under 140 µmol/m²/s of continuous illumination.

Vector construction and transformation

The ATP synthetase γ -subunit (AtpC) gene (DDBJ/EMBL/GenBank database accession number AB938081) was screened from the draft genome database of the *F. solaris* JPCC DA0580 (Tanaka, T. et al, manuscript submitted) based on the Basic Local Alignment Search Tool Protein (BLASTP) search. Amino acid sequences of AtpC of *F. solaris* and *P. tricornutum* were shown in Figure S1. The GFP expression vector, pSP-GFP/GAPDH was constructed in our previous study (34), in which *gfp* gene was inserted between the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) promoter

derived from *F. solaris* and fucoxanthin chlorophyll a/c-binding protein A (fcpA) terminator derived from *P. tricornutum* (9). In this study, the DNA fragment encoding the fusion protein of GFP and signal sequence of AtpC (S_{AtpC}-GFP) was amplified by polymerase chain reaction (PCR) using the primer pair (5'-TAA TTG AAT TCA TGA GAT TCC TTG TCA CGG CCG CTC TTC TTG CAG TTG TAT CTG CTT TCA CAA CGC AAC AAC CTG GCT TGG CGC ACC GCA CGC GTC AAG CGA CCG TCG TTT ATG ACA TGG TGA GCA AGG GCG AGG AG-3' and 5'-TTG TTC TGC AGT TAC TTG TAC AGC TCG TCC ATG C-3') and pSP-GFP/GAPDH as a template. The amplified fragment was inserted between the GAPDH promoter and fcpA terminator using *EcoRI* and *PstI* sites. The constructed plasmid was referred to as pSP-S_{AtpC}GFP/GAPDH.

Transformation of *F. solaris* was performed by microparticle bombardment using the Biolistic PDS-1000/He Particle Delivery System (Bio-Rad Laboratories, Inc., Hercules, CA, USA) as described previously (9, 34).

SDS-PAGE and Western blotting

F. solaris transformants (5×10^7 cells) were collected by centrifugation, washed with water, suspended in 100 μ l of 1% (w/v) sodium dodecyl sulfate (SDS) in aqueous solution and boiled for 10 min. After centrifugation, supernatant was collected, and SDS sample buffer was added (final concentration of 62.5 mM Tris-HCl, pH 6.8, 5% 2-mercaptoethanol, 2% SDS, 5% sucrose, and 0.002% bromophenol blue). Denatured proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using 12.5% (w/v) gels. Bio-Safe Coomassie G-250 Stain (Bio-Rad laboratories, Inc., Hercules, CA, USA) was used for gel staining. For Western blotting, the separated proteins were transferred to a polyvinylidene difluoride membrane. GFP was then detected using alkaline phosphatase (ALP)-labeled anti-GFP antibody (Rockland immunochemicals inc., Gilbertsville, PA, USA, 1/5,000 dilution from stock in PBS containing 0.05 % Tween 20). BCIP/NBT-Blue (Sigma, St. Louis, Mo. USA) was used as the ALP substrate for visualization.

Fluorescent microscopy and image analysis

Transformant cells were observed using a fluorescent microscope BX51 (Olympus Corporation, Tokyo, Japan); a U-MGFPPHQ filter set visualized GFP, a U-MWIG2 filter set did chlorophyll and a U-MWU2 filter set did GFP, chlorophyll and Hoechst 33342 fluorescence (Hoechst emitted stronger fluorescence than GFP and Chlorophyll). Hoechst staining was conducted by adding Hoechst 33342 (Invitrogen, Eugene, OR, USA) to the cell culture at 1:50 volume ratio (final concentration = 200 µg/ml). Confocal microscopy was performed with Fluoview FV1000 (Olympus Corporation, Tokyo, Japan).

Transmission electron microscopy

Wild type cells were collected by centrifugation. The collected cells were washed with 0.85% NaCl solution and subsequently re-suspended in cell fixation solution (2% glutaraldehyde/30 mM HEPES, pH7.0) for 2 h at 4°C, followed by washing with 30 mM HEPES (pH7.0) at 4°C. The cells were treated with 30% and 100% ethanol for 15 min each for anhydration. After the treatment using propylene oxide, the cells were embedded in epoxy resin Quetol 651. Thermal polymerization process was carried out at 60°C for 24 h. Then observation specimens were prepared using an ultramicrotome (thickness = 80-90 nm). After uranyl acetate/lead double-staining, TEM observation was performed using JEOL JEM1200EM with an acceleration voltage of 80 kV.

Results

Pipeline for prediction of protein localization in diatoms

1 In previous studies, *in silico* analyses were conducted to predict the protein subcellular localization
2 in a diatom (25, 29). However, pipelines for comprehensive analysis with well-defined criteria have
3 not been described in detail. Therefore, we attempted to determine the analysis pipeline with several
4 bioinformatics programs (TargetP, SignalP, HECTAR, Mitoprot, and TMHMM) by reference to the
5 previous studies (Figure 1, see also materials and methods section).

6 To evaluate the prediction accuracy of our pipeline, subcellular localizations of 12 proteins, whose
7 subcellular localizations were experimentally validated in *P. tricornutum* (Table S1), were predicted
8 using the pipeline. As a result, 11 prediction results by our pipeline coincided with experimental
9 validations (~92% accuracy, Table S1). In contrast, prediction results by each program such as
10 TargetP and HECTAR coincided with experimental validations in only 5 and 7 of the proteins tested
11 (~42% and ~58% accuracy), respectively. It should be noted that the only incorrect prediction by our
12 pipeline was for the mitochondrial carbamoyl phosphate synthase, and it could be caused by the
13 incomplete sequence registered in databases (Table S1). These results indicate that use of our
14 pipeline determined here would be reasonable to assess the chloroplast-targeting signal sequence
15 employed below.

17 ***Signal peptide of ATP synthetase of *F. solaris****

18 In order to achieve transportation of recombinant proteins into complex diatom chloroplast,
19 identification of the signal sequence targeting the chloroplast is the first step to be addressed. Among
20 proteins expected to be transported into chloroplast by our pipeline, we decided to utilize the signal
21 sequence of a chloroplast ATP synthetase γ -subunit (AtpC) because of reasons explained below. It
22 has been widely known that ATP synthesis concomitant with photosynthesis was catalyzed by
23 plastid (chloroplast) F_0F_1 -ATP synthetase complexes composed of multi-subunits at thylakoid

membranes. The F_0 unit is responsible for proton translocation while the F_1 unit, in which AtpC is participated, possesses catalytic activity. Thus it would be assured that AtpC is a chloroplast protein. Furthermore strong support for use of AtpC signal sequence is that previous research succeeded in recombinant protein transportation into the chloroplast of a pennate model non-oleaginous diatom, *P. tricornutum* (20).

When AtpC sequences of *F. solaris* and *P. tricornutum* (PHATRDRAFT_20657) were aligned, the sequence identity was 80% (Figure S1). Despite the high identity throughout the full sequences, there is a region with low sequence similarity at their N-termini downstream the typical ER-signal peptide cleavage sites (so-called ASAF motif) (19). In this study, we employed N-terminal 35 amino acids as a chloroplast-targeting signal of *Fistulifera* AtpC (S_{AtpC}) with reference to the previous truncation trial where N-terminal 38 amino acids of *Phaeodactylum* AtpC was capable of chloroplast-targeting (Figure S1) (20).

Expression of the fusion protein in transformants

We constructed the expression vector for the fusion protein of S_{AtpC} and GFP (S_{AtpC} -GFP), and introduced it into *F. solaris* cells in order to demonstrate our concept. As-prepared transformants were subjected to SDS-PAGE and Western blotting to confirm whether S_{AtpC} -GFP was produced in the cells. As a result of SDS-PAGE with Coomassie staining, significant difference did not appear between the wild type and transformants; it could be due to low sensitivity of the Coomassie method (Figure 2A). Subsequently, specific detection was attempted with Western blotting using anti-GFP antibody, and bands were visualized at the lane of transformants expressing GFP and S_{AtpC} -GFP, respectively (Figure 2B). The size of the detected S_{AtpC} -GFP protein was larger than the GFP without S_{AtpC} produced in the distinct *F. solaris* transformants, suggesting the successful expression of the target fusion protein. Furthermore, its size was slightly smaller than the intact protein encoded

(approximately 31 kDa, Figure S2). As a negative control experiment, wild type cells were also examined, and no signal was detected.

Chloroplast-targeting of S_{AtpC} -GFP

To examine whether the AtpC signal sequence directs proteins to specific organelles, the cells expressing S_{AtpC} -GFP were observed using a fluorescent microscope at low magnification. In all cells observed, spatial distribution of GFP and chlorophyll overlapped (Figure 3A). It was also revealed that GFP did **not** evenly exist in the chloroplast. For further study, the cells expressing GFP and S_{AtpC} -GFP were compared using a confocal fluorescent microscope. The GFP-expressing cell exhibited distinct fluorescence distribution of GFP and chlorophyll (Figure 3B-a); GFP is estimated to localize at the cytoplasm as observed previously (34). In contrast, the S_{AtpC} -GFP-expressing cell showed the intense green fluorescence from GFP which overlapped on the red fluorescence from chlorophyll (Figure 3B-a). When GFP and chlorophyll distributions were spatially profiled, it was demonstrated that the fluorescent regions derived from GFP and chlorophyll exactly overlapped (Figure S3). These results suggest that S_{AtpC} -GFP and chlorophyll co-localized in the chloroplast. Note that wild type cells not expressing GFP showed no green fluorescence under the same observation condition.

Significant discovery from this observation is that GFP fluorescence was not evenly distributed in the chloroplast, and some regions showed significant intensity. First the edge of chloroplast tended to emit stronger fluorescence than inner part (Figure S3). Furthermore, a rod-like structure was visible which showed significant fluorescence at the constricted part of the chloroplast (Figure 3B-c). A lateral image of the transformant cell demonstrated that the fluorescent rod-like structure was connected to the chloroplast, but looked a different object from the chloroplast (Figure 3B-d). This structure was observable throughout the cultivation period, suggesting that this localization manner is independent on the cell growth stages (Figure S4). When the cells were stained with the Hoechst

dye, the DNA was demonstrated to exist at a different location from neither chloroplast nor the rod-like structure (Figure 3C), suggesting that the rod-like structure was not a nucleus. Subsequently TEM observation of *F. solaris* was performed for further elucidation. As a result, a lenticular pyrenoid crossed by a single lamella was observed in the chloroplast, and its arrangement was found to closely resemble that of the rod-like structure (Figure 4).

Discussion

Before the demonstration of chloroplast-targeting protein expression, we determined the bioinformatics pipeline for prediction of protein localization in diatoms. This pipeline is useful to assess the chloroplast-targeting ability of the signal sequence employed in this study. According to our pipeline, the signal sequence of AtpC (gamma subunit of chloroplast ATP synthetase) was predicted to target the chloroplast.

Subsequently, we constructed the expression vector of S_{AtpC}-GFP, where the ER-signal sequence and a part of transit sequence (M1-D35 in the Figure S1) were together fused at the N-terminus of GFP. Western blotting for the *F. solaris* transformants harboring the S_{AtpC}-GFP expression gene confirmed substantial expression of the target protein. Furthermore the relative migration analysis using the Western blotting data revealed that the detected S_{AtpC}-GFP was slightly smaller than expected (Figure S2), implying the cleavage of signal peptide during transportation (12, 17). The cleavage site is estimated to be the point between A15 and F16 in the typical ASAF motif (Figure S1) (19, 21). It would be reasonable to consider this cleavage manner because the detected protein size on Western blotting was coincide with that calculated from the amino acid sequence (Figure S2), while further study would be required to determine the exact cleavage site. It should be noted that chloroplast-targeting protein expression can also be achieved by modifying the chloroplast genome (23). This method is also worth to try in the near future.

1 *In vivo* localization of S_{AtpC}-GFP in *F. solaris* was observed with fluorescent microscopy. As a result,
2 green fluorescence was found to overlap with that from chlorophyll, strongly indicating the success in
3 chloroplast-targeting protein expression (Figure 3). This is not surprising because the signal sequence
4 of AtpC ortholog in other pennate diatom *P. tricornutum* have previously achieved
5 chloroplast-targeting protein expression (20). However unexpected results also appeared in this study
6 that S_{AtpC}-GFP was not evenly distributed in the chloroplast, and some region showed relatively
7 stronger fluorescence.

8 Relatively strong emission was observed at inner edges of the chloroplast (Figure S3). This might
9 be caused by stacking of S_{AtpC}-GFP at intermembrane spaces in the complex chloroplast membrane
10 structures. More surprisingly, the rod-like structure showing significant fluorescence was found,
11 which the previous study using *P. tricornutum* did not report (20). For further elucidation, these
12 confocal images were compared with TEM observation. Spatial distribution of the rod-like structure
13 was similar to the subcellular arrangement of the lenticular object containing single lamella (Figure
14 4). Given the shape and ultrastructure observed, this structure could be a pyrenoid, whose
15 morphology is reported to be typical in some diatoms (35-37). Guided by these findings, we
16 estimated that S_{AtpC}-GFP was directed and distributed into the entire chloroplast, especially in the
17 pyrenoid.

18 Pyrenoid is an organelle which exists inside the chloroplast and acts as a center of inorganic carbon
19 fixation (38). Ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) (36) and carbonic
20 anhydrase (CA) (37) were proved to localize in diatom pyrenoid. Pyrenoid membrane structure (if
21 any), detail of inclusions, and their differences between diatom species remain unclear, but these
22 factors might be attributed to the difference between the results in this study and previous one where
23 GFP showed completely even distribution in the *P. tricornutum* chloroplast. Pyrenoid tends to
24 contain densely-packed soluble proteins including RubisCO and CA (38), and this feature might
25 induce the S_{AtpC}-GFP accumulation observed in this study.

1 In conclusion, the fusion experiment with GFP and the signal sequence of the chloroplast protein
2 AtpC was transported into the chloroplast of the candidate biodiesel producer, *F. solaris* JPCC
3 DA0580. This result indicates the establishment of the technique for chloroplast-targeting protein
4 expression in the oleaginous “non-model” diatom. This technique would be useful for future attempt
5 to engineer the lipid metabolisms.

6

7 **Acknowledgments**

8 This study was supported by the Japan Science and Technology Agency (JST) and CREST.

9

1   **References**

- 2   1. Smith, V. H., Sturm, B. S., Denoyelles, F. J., Billings, S. A.: The ecology of algal biodiesel
3   production, *Trends Ecol Evol*, 25, 301-309 (2010).
- 4   2. Matsunaga, T., Matsumoto, M., Maeda, Y., Sugiyama, H., Sato, R., Tanaka, T.: Characterization of
5   marine microalga, *Scenedesmus* sp. strain JPCC GA0024 toward biofuel production, *Biotechnol Lett*,
6   31, 1367-1372 (2009).
- 7   3. Rodolfi, L., Chini Zittelli, G., Bassi, N., Padovani, G., Biondi, N., Bonini, G., Tredici, M. R.:
8   Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a
9   low-cost photobioreactor, *Biotechnol Bioeng*, 102, 100-112 (2009).
- 10   4. Matsumoto, M., Sugiyama, H., Maeda, Y., Sato, R., Tanaka, T., Matsunaga, T.: Marine diatom,
11   *Navicula* sp. strain JPCC DA0580 and marine green alga, *Chlorella* sp. strain NKG400014 as
12   potential sources for biodiesel production, *Appl Biochem Biotechnol*, 161, 483-490 (2010).
- 13   5. Matsumoto, M., Mayama, S., Nemoto, M., Fukuda, Y., Muto, M., Yoshino, T., Matsunaga, T.,
14   Tanaka, T.: Morphological and molecular phylogenetic analysis of the high triglyceride-producing
15   marine diatom, *Fistulifera solaris* sp. nov. (Bacillariophyceae), *Phycol Res*, (Just accepted).
- 16   6. Liang, Y., Maeda, Y., Yoshino, T., Matsumoto, M., Tanaka, T.: Profiling of fatty acid methyl esters
17   from the oleaginous diatom *Fistulifera* sp. strain JPCC DA0580 under nutrition-sufficient and

- 1 -deficient conditions, J Appl Phycol, DOI: 10.1007/s10811-014-0265-y (In press).
- 2 7. Satoh, A., Ichii, K., Matsumoto, M., Kubota, C., Nemoto, M., Tanaka, M., Yoshino, T., Matsunaga,
3 T., Tanaka, T.: A process design and productivity evaluation for oil production by indoor mass
4 cultivation of a marine diatom, *Fistulifera* sp. JPCC DA0580, Bioresour Technol, 137, 132-138
5 (2013).
- 6 8. Sato, R., Maeda, Y., Yoshino, T., Tanaka, T., Matsumoto, M.: Seasonal variation of biomass and oil
7 production of the oleaginous diatom *Fistulifera* sp. in outdoor vertical bubble column and
8 raceway-type bioreactors, J Biosci Bioeng, 117, 720-724 (2014).
- 9 9. Muto, M., Fukuda, Y., Nemoto, M., Yoshino, T., Matsunaga, T., Tanaka, T.: Establishment of a
10 genetic transformation system for the marine pennate diatom *Fistulifera* sp. strain JPCC DA0580--a
11 high triglyceride producer, Mar Biotechnol, 15, 48-55 (2013).
- 12 10. Sandelius, A. S., Aronsson, H.: The chloroplast, Springer (2008).
- 13 11. Chisti, Y.: Biodiesel from microalgae, Biotechnol Adv, 25, 294-306 (2007).
- 14 12. Roesler, K., Shintani, D., Savage, L., Boddupalli, S., Ohlrogge, J.: Targeting of the *Arabidopsis*
15 homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds, Plant Physiol, 113, 75-81
16 (1997).
- 17 13. Rangasamy, D., Ratledge, C.: Genetic enhancement of fatty acid synthesis by targeting rat liver

- 1   ATP: citrate lyase into plastids of tobacco, *Plant Physiol*, 122, 1231-1238 (2000).
- 2   14. Klaus, D., Ohlrogge, J. B., Neuhaus, H. E., Dormann, P.: Increased fatty acid production in
3   potato by engineering of acetyl-CoA carboxylase, *Planta*, 219, 389-396 (2004).
- 4   15. Dunahay, T. G., Jarvis, E. E., Dais, S. S., Roessler, P. G.: Manipulation of microalgal lipid
5   production using genetic engineering, *Appl Biochem Biotechnol*, 57-58, 223-231 (1996).
- 6   16. Gibbs, S. P.: The chloroplast endoplasmic reticulum: structure, function, and evolutionary
7   significance, *Int Rev Cytol*, 72, 49-99 (1981).
- 8   17. Huang, A.: Oleosins and oil bodies in seeds and other organs, *Plant Physiol*, 110, 1055 (1996).
- 9   18. Peschke, M., Moog, D., Klingl, A., Maier, U. G., Hempel, F.: Evidence for glycoprotein transport
10   into complex plastids, *Proc Natl Acad Sci U S A*, 110, 10860-10865 (2013).
- 11   19. Kilian, O., Kroth, P. G.: Identification and characterization of a new conserved motif within the
12   presequence of proteins targeted into complex diatom plastids, *Plant J*, 41, 175-183 (2005).
- 13   20. Apt, K. E., Zaslavkaia, L., Lippmeier, J. C., Lang, M., Kilian, O., Wetherbee, R., Grossman, A. R.,
14   Kroth, P. G.: *In vivo* characterization of diatom multipartite plastid targeting signals, *J Cell Sci*, 115,
15   4061-4069 (2002).
- 16   21. Gruber, A., Vugrinec, S., Hempel, F., Gould, S. B., Maier, U. G., Kroth, P. G.: Protein targeting
17   into complex diatom plastids: functional characterisation of a specific targeting motif, *Plant Mol Biol*,

- 1 64, 519-530 (2007).
- 2 22. Field, C. B., Behrenfeld, M. J., Randerson, J. T., Falkowski, P.: Primary production of the
3 biosphere: Integrating terrestrial and oceanic components, *Science*, 281, 237-240 (1998).
- 4 23. Tran, M., Van, C., Barrera, D. J., Pettersson, P. L., Peinado, C. D., Bui, J., Mayfield, S. P.:
5 Production of unique immunotoxin cancer therapeutics in algal chloroplasts, *Proc Natl Acad Sci U S*
6 *A*, 110, E15-22 (2013).
- 7 24. Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U.,
8 Martens, C., Maumus, F., Otiillar, R. P., and other 67 authors: The *Phaeodactylum* genome reveals the
9 evolutionary history of diatom genomes, *Nature*, 456, 239-244 (2008).
- 10 25. Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S.,
11 Allen, A. E., Apt, K. E., Bechner, and other 35 authors: The genome of the diatom *Thalassiosira*
12 *pseudonana*: ecology, evolution, and metabolism, *Science*, 306, 79-86 (2004).
- 13 26. Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G.: Predicting subcellular localization of
14 proteins based on their N-terminal amino acid sequence, *J Mol Biol*, 300, 1005-1016 (2000).
- 15 27. Gschloessl, B., Guermeur, Y., Cock, J. M.: HECTAR: a method to predict subcellular targeting in
16 heterokonts, *BMC Bioinformatics*, 9, 393 (2008).
- 17 28. Bendtsen, J. D., Nielsen, H., von Heijne, G., Brunak, S.: Improved prediction of signal peptides:

- 1   SignalP 3.0, J Mol Biol, 340, 783-795 (2004).
- 2   29. Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M. S., Stanley, M.
- 3   S., Kaplan, A., Caron, L., Weber, T., Maheswari, U., Armbrust, E. V., Bowler, C.: A model for
- 4   carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative
- 5   whole genome analysis, PLoS One, 3, e1426 (2008).
- 6   30. Claros, M. G., Vincens, P.: Computational method to predict mitochondrially imported proteins
- 7   and their targeting sequences, Eur J Biochem, 241, 779-786 (1996).
- 8   31. Gonzalez, N. H., Felsner, G., Schramm, F. D., Klingl, A., Maier, U. G., Bolte, K.: A single
- 9   peroxisomal targeting signal mediates matrix protein import in diatoms, PLoS One, 6, e25316
- 10   (2011).
- 11   32. Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E. L.: Predicting transmembrane protein
- 12   topology with a hidden Markov model: application to complete genomes, J Mol Biol, 305, 567-580
- 13   (2001).
- 14   33. Guillard, R. R., Ryther, J. H.: Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt,
- 15   and *Detonula confervacea* (Cleve) Gran, Can J Microbiol, 8, 229-239 (1962).
- 16   34. Nojima, D., Yoshino, T., Maeda, Y., Tanaka, M., Nemoto, M., Tanaka, T.: Proteomics analysis of
- 17   oil body-associated proteins in the oleaginous diatom, J Proteome Res, 12, 5293-5301 (2013).

- 1 35. Bedoshvili, Y. D., Poplova, T. P., Likhoshway, Y., V: Chloroplast structure of diatoms of different
2 classes, *Cell Tissue Biol*, 3, 297-310 (2009).
- 3 36. Mayama, S., Mayama, N., Shihira-Ishikawa, I.: Characterization of linear-oblong pyrenoids with
4 cp-DNA along their sides in *Nitzschia sigmoidea* (Bacillariophyceae), *Phycol Res*, 52, 129-139
5 (2004).
- 6 37. Tachibana, M., Allen, A. E., Kikutani, S., Endo, Y., Bowler, C., Matsuda, Y.: Localization of
7 putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira*
8 *pseudonana*, *Photosynth Res*, 109, 205-221 (2011).
- 9 38. Roberts, K., Granum, E., Leegood, R. C., Raven, J. A.: Carbon acquisition by diatoms,
10 *Photosynth Res*, 93, 79-88 (2007).
11
12
13

1 **Figure legend**

2 **Figure 1** Scheme for prediction of protein localization in diatoms. (Abbreviations: ER,
3 endoplasmic reticulum; Chloro, chloroplast; Mito, mitochondrion; Peroxi, peroxisome; Cyto,
4 cytoplasm)

5 **Figure 2** SDS-PAGE (A) and Western blotting (B) analyses of the proteins extracted from *F.*
6 *solaris* JPCC DA0580 transformants expressing GFP (Lane 1), S_{AtpC}-GFP (Lane 2), and
7 wild-type cells (Lane 3). M represents molecular markers.

8 **Figure 3** Microscopy studies on *F. solaris* JPCC DA0580 transformants expressing S_{AtpC}-GFP.
9 (A) Fluorescent microscopic images of the transformant cells expressing S_{AtpC}-GFP. GFP, WIG
10 filter sets were used for visualization of GFP and chlorophyll, respectively. The WU filter set
11 was used for GFP and chlorophyll simultaneously (scale bar = 10 μm). (B) Fluorescent
12 microscopy of the transformant cells expressing S_{AtpC}-GFP (the merged fluorescence images
13 were computer-generated); (a) the transformant cell expressing GFP as a control (scale bar = 1.5
14 μm); (b) The typical transformant cell showing overlapping fluorescence derived from GFP and
15 chlorophyll (scale bar = 1.3 μm); (c) The rod-like structure emitting intensive fluorescence is
16 shown by a white arrow (scale bar = 1.5 μm); (d) A lateral image of the cell clearly showing
17 the rod-like structure (scale bar = 1.5 μm). (C) The transformant expressing S_{AtpC}-GFP with
18 Hoechst 33342 staining (scale bar = 5 μm, the merged fluorescence image was
19 computer-generated).

20 **Figure 4** Transmission electron microscopic (TEM) images of *F. solaris* JPCC DA0580 wild
21 type, and the confocal fluorescent microscopic images of the transformant cells expressing
22 S_{AtpC}-GFP. Arrows on the TEM images show pyrenoid, and arrows on the fluorescent
23 microscopic images shows rod-like structures emitting intensive green fluorescence.

5. Figure 1

*C-terminal is K(D)-D(E)-E-L

**C-terminal is S(A/C)-K(R/H)-L(M) or S-S-L

TEM

Fluorescence microscopy