Identification of Major Lipid Droplet Protein in a Marine Diatom *Phaeodactylum tricornutum*

Lipid Droplet Proteins in diatom *P. tricornutum*

Corresponding author: Masaki Yoshida

Faculty of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan

Tel.: +81-29-853-4301; Fax: +81-29-853-4301; E-mail: yoshida.masaki.gb@u.tsukuba.ac.jp

Subject areas: (4) proteins, enzymes and metabolism, (6) structure and function of cells

Number of black and white figures: three figures

Number of color figures: two figures

Number of tables: two tables

Number of supplemental materials: one table and one color figure

Identification of Major Lipid Droplet Protein in a Marine Diatom *Phaeodactylum tricornutum*

Lipid Droplet Protein in diatom *P. tricornutum*

Kohei Yoneda[1], Masaki Yoshida[2], Iwane Suzuki[2] and Makoto M. Watanabe[2]

[1]Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki 305-8572, Japan

[2]Faculty of Life and Environmental Sciences, University of Tsukuba, Ibaraki 305-8572, Japan

Abbreviations

BLAST, basic logical alignment search tool; DOAP1, diatom-oleosome-associated protein 1; ER, endoplasmic reticulum; HOGP, *Haematococcus* oil globule protein; Hsp70, heat shock protein 70; JGI, Joint Genome Institute; MLDP, major lipid droplet protein; NCBI, National Center for Biotechnology Information; OLE3, oleosin3; qRT-PCR, real-time quantitative reverse-transcriptase polymerase chain reaction; SLDP, *Symbiodinium* lipid droplet protein; StLDP, Stramenopile-type lipid droplet protein; ESI-Q/TOF, electrospray ionization quadrupole time-of-flight; rbcL, Rubisco large subunit; TAG, triacylglycerol; TLC, thin layer chromatography.

Abstract

Various kinds of organisms, including microalgae, accumulate neutral lipids in distinct intracellular compartments called lipid droplets. Generally, lipid droplets are generated from the endoplasmic reticulum and particular proteins localize on their surface. Some of these proteins function as structural proteins to prevent fusion between the lipid droplets, and the others could have an enzymatic role or might be involved in intracellular membrane trafficking. However, information about lipid droplet proteins in microalgae is scarce as compared with that in animals and land plants. We focused on the oil-producing, marine, pennate diatom *Phaeodactylum tricornutum* that forms lipid droplets during nitrogen deprivation and we investigated the proteins located on the lipid droplets. After 6 days of cultivation in a nitrate-deficient medium, the mature lipid droplets were isolated by sucrose density gradient centrifugation. Proteomic analyses revealed five proteins, with Stramenopile-type lipid droplet protein (StLDP) being the most abundant protein in the lipid droplet fraction. Though the primary sequence of StLDP did not have homology to any known lipid droplet proteins, StLDP had a central hydrophobic domain. This structural feature is also detected in oleosin of the land plant and lipid droplet surface protein (LDSP) of the *Nannochloropsis*. As a proline knot motif of oleosin, conservative proline residues existed in the hydrophobic domain. StLDP was upregulated during nitrate deprivation and fluctuations of StLDP expression levels corresponded with the size of the lipid droplets.

Keywords

Diatom, Lipid droplet protein, Nitrogen deprivation, Oleosin, *Phaeodactylum tricornutum*, Proteomic analysis

Introduction

Microalgae have great potential as a sustainable feedstock for biofuel and oils because of their high biomass productivity per area or per time (Chisti 2007). Many microalgae have the ability to store triacylglycerol (TAG) in cell compartments called lipid droplets (Goold et al. 2015). The marine pennate diatom *Phaeodactylum tricornutum* also accumulates TAG in lipid droplets under nitrogen (N)-limited conditions (Yang et al. 2013). The biomass of this microalga is expected to be used as an oil resource, and mass cultivation trials have been performed for oil production (Acién Fernández et al. 2003, Fernández Sevilla et al. 2004). Moreover, molecular biological tools such as genetic modifications, real-time quantitative reverse-transcription polymerase chain reaction (qRT-PCR) (Siaut et al. 2007), and complete genome information of *P. tricornutum* (Bowler et al. 2008) are currently available. Thus, *P. tricornutum* is a suitable model organism in molecular and applied phycology.

Lipid droplets are intracellular compartments that pool lipophilic molecular species. Various kinds of organisms including animals, land plants, yeasts, algae, and bacteria can form lipid droplets in their cells (Murphy and Vance 1999, Murphy 2012). Common features of lipid droplets are as follows: they arise from the endoplasmic reticulum (ER) and are released into the cytosol, they are mainly filled with TAG, and they are surrounded by a phospholipid monolayer derived from the ER membrane and contain specific proteins on their surface (Ohsaki et al. 2014, Pol et al. 2014). These proteins play a role as structural proteins to stabilize the lipid droplet, as enzymes for lipid metabolism, and as intracellular membrane trafficking (Murphy and Vance 1999, Martin and Parton 2006, Murphy 2012). Proteomic analyses of the lipid droplets in various organisms have recently been reported (Yang et al. 2012). Although lipid droplets are ubiquitous organelles, their surface proteins are quite diverse; for example, the PAT family proteins on mammalian cytosolic lipid droplets include perilipin, adipophilin, TIP47, S3-12, and OXPAT (Wolins et al. 2006, Brasaemle 2007). In spermatophyta (seed plants), the major proteins on lipid droplets are oleosin family proteins, and they include oleosin, caleosin, and steroleosin (Frandsen et al. 2001, Chapman et al. 2012). Although lipid droplets in mammalian adipocytes and plant seeds have similar roles, the sequences of these two protein families located on the lipid droplet are very low. On the other hand, information about algal lipid droplets and their surface proteins is scarce, and only a few cases have been reported for the model algae. For example, proteins that regulate lipid droplet size, i.e., major lipid droplet protein (MLDP), were identified in

Chlamydomonas reinhardtii (Moellering and Benning 2010, Nguyen et al. 2011) and *Dunaliella salina* (Davidi et al. 2012). Not only cytoplasmic but also plastidic β-carotene lipid body proteins were analyzed in *Dunaliella bardawil* (Davidi et al. 2014). In an astaxanthin-accumulating alga, *Haematococcus pluvialis*, the lipid droplet surface protein termed *Haematococcus* oil globule protein (HOGP) was revealed to be an ortholog of MLDP (Peled et al. 2011). In addition, putative caleosin was discovered to be the main protein in the lipid droplet fraction of *Chlorella* sp. (Lin et al. 2012), as well as caleosin-related *Symbiodinium* lipid droplet protein (SLDP) identified in the endosymbiotic dinoflagellates, *Symbiodinium* (Pasaribu et al. 2014). Moreover, alkenone body-associated proteins were analyzed in the haptophyte alga *Tisochrysis lutea* (Shi et al. 2015).

Lipid droplet proteins in the Stramenopiles have only been analyzed in two species, *Nannochloropsis oceanica* (Vieler et al. 2012) and *Fistulifera* sp. JPCC DA0580, named *F. solaris* (Nojima et al. 2013). Surface proteins of lipid droplets were identified in *N. oceanica* and were found to have a central hydrophobic domain (Vieler et al. 2012). Nojima et al. (2013) identified five candidates for lipid droplet proteins in the diatom *F. solaris*.

We focused on the oil-producing diatom *P. tricornutum* and investigated its lipid droplet proteins to understand the mechanism of lipid accumulation. In this study, we firstly isolated the lipid droplets by sucrose gradient centrifugation, and then we identified the proteins by performing SDS-PAGE followed by ESI-Q/TOF mass spectrometry of target peptide fragments. Obtained m/z data were processed using the *P. tricornutum* genome-based database search for identification. The sequence of the lipid droplet protein identified was then confirmed by GenBank database search and the presence of similar sequences of lipid droplet-associated proteins in a variety of organisms were suggested through a BLAST search.

Results

Isolation of lipid droplets and evaluation of contaminants

The isolated lipid droplet fraction was analyzed by microscopy, spectrophotometry, thin layer chromatography (TLC), and immunological methods (Figure 1). Using light microscopy, we did not detect any contaminant debris in the isolated lipid droplet fraction (Figure 1A, B). Lipids in the isolated lipid droplet fraction were first extracted with acetone and then with ethyl acetate. Figure 1C shows the UV-visible absorption spectra of acetone extracts from the isolated lipid droplets, chloroplasts, and

whole cells. The absorption peaks of the lipid droplet fraction were observed at 448 nm and 474 nm, which correspond to carotenoids. Although an absorption peak corresponding to chlorophyll *a* was observed at 663 nm in the extracts from chloroplasts and whole cells, the peak of A_{663} was well suppressed in the extract from the isolated lipid droplets. These results indicate that chloroplast contamination in the isolated fraction was minimal.

Lipids extracted from the isolated lipid droplets and the whole cells were analyzed by silica-gel TLC (Figure 1D). The TAG spot was observed in extracts from the isolated lipid droplets.

We performed a Western blot analysis using rabbit antiserum against the large subunit of Rubisco (As-RbcL) in order to confirm the purity of the lipid droplets (Figure 1E, F). The As-RbcL reacted with the protein extracted from the whole cells but did not react with the protein extracted from the isolated lipid droplets (Figure 1E, F).

SDS-PAGE and protein identification

We fractionated proteins that were extracted from whole cells and from isolated lipid droplets with SDS-PAGE (R1–R3) (Figure 2). The molecular mass of the major protein in the lipid droplets of *P. tricornutum* was 49 kDa (gel fraction 4).

Proteins identified in the isolated lipid droplet fraction are shown in Table 1. The protein Phatr48859 was identified in all three replicates in gel fraction 4 and it corresponded with the 49-kDa major band. Thus, Phatr48859 may be one of the abundant proteins on lipid droplets in *P. tricornutum*. As shown in Supplemental figure 1, the orthologs of Phatr48859 were conserved mainly in Stramenopiles; hence we named this protein “Stramenopile-type lipid droplet protein (StLDP)”. The results of BLAST search indicated that StLDP did not share any known functional domain with those registered in the National Center for Biotechnology Information (NCBI) database and did not have sequence homology to any known lipid droplet proteins.

Phatr48778, which includes an acyl-CoA binding site, was identified in all three replicates, suggesting that Phatr48778 plays an important role on the lipid droplets, whereby the acyl-CoA binding site relates to fatty acid metabolism. The other identified proteins, presented in Table 1, may be of minor importance because they were only detected in a single experiment. Phatr54019, which is similar to the heat shock protein 70 (Hsp70), may be a molecular chaperone for proteins located on the lipid droplets. Phatr45894 and Phatr49981 seemed to undergo redox reactions, but we will not speculate on

the substrates of these reactions in this work.

Molecular characteristics of StLDP

Figure 3 shows hydropathy plots of StLDP and representative proteins from lipid droplets; i.e., oleosin of *Arabidopsis thaliana* and LDSP of *Nannochloropsis* sp. Although StLDP and two other comparative proteins were composed of different numbers of amino acids, all these proteins had hydrophobic domains in the central region (222–275 aa of StLDP, 54–128 aa of oleosin, and 72–133 aa of LDSP), despite low similarities in their amino acid sequences.

As a result of a BLAST search on databases in NCBI and the Joint Genome Institute (JGI), we found homologs of StLDP conserved in four diatoms (*Fragilariopsis cylindrus*, *Pseudo-nitzschia multiseries*, *Thalassiosira pseudonana*, *and T. oceanica*) and two other heterokontophytes (a Eustigmatophyte; *Nannochloropsis gaditana*; and a brown alga*, Ectocarpus siliculosus*). Figure 4 shows a multiple alignment of amino acid sequences of StLDP and its orthologs. The hydrophobic domains are shown in the red dotted square (222–275 aa in *P. tricornutum* StLDP). This domain consists of 54 amino acid residues, which are mainly hydrophobic, except for proline. Four distinctive proline residues (red arrows) were recognized and we suggest that they make a characteristic motif ($PX_9PX_{10}PX_3P$) in the hydrophobic domain. In addition, there were two conserved domains at the N-terminal and C-terminal regions from the central hydrophobic domains (indicated by black and red two-headed arrows, respectively).

Expression levels of StLDP and sizes of lipid droplets during N-deficient cultivation

To examine whether StLDP expression is induced under N-deficient conditions, we determined the levels of StLDP transcripts by qRT-PCR along with the sizes of lipid droplets (Figure 5). We speculated that if the StLDP served as major protein, lipid droplet surface should be filled with it; therefore, the change of surface area of lipid droplet should be accompanied with the expression of the StLDP. The StLDP expression levels were upregulated during the N-deficient condition and the extent of its induction ranged from 2.8-folds (1 d after N-deprivation) to 7.1-folds (3 d after N-deprivation) than during N-sufficient conditions. The expression levels reached a peak at 3 d after N-deprivation and then decreased to a steady-state level that was approximately 3- to 4-folds higher than the control. Fluctuations in the expression levels corresponded with the size of the lipid droplets, as we expected.

Figure 5 presents the average sizes of the lipid droplets during N-deficiency. The average size of the lipid droplets increased up to 4 d after N-deprivation and reached plateau (Figure 5).

Discussion

Evaluation of contaminants and lipid contents of the lipid droplet fraction

We disrupted the cells as gently as possible to prevent unnecessary breakage of the lipid droplets and other organelles. If the disruption forces are too intense, the risk of contamination by other organelles increases, particularly because debris and vesicles from various organelles are generated. Moreover, cells of common diatoms possess distinct, hard siliceous frustules; thus, it is difficult to break their cells with mild disruption. In previous studies of the diatom *Fistulifera solaris*., Nojima et al. (2013) adopted bead beating as a disruption method, and they reported that debris of chloroplasts could not be excluded. We disrupted the diatom cells using a French press and then the lipid droplets were isolated using sucrose density gradient centrifugation, based on the method of Ding et al. (2013). Vieler et al. (2012) also used a French press, although at 20 kpsi, to disrupt *Nannochloropsis* cells in order to isolate lipid droplets, whereas we employed 1 kpsi. After fractionation using sucrose density gradient centrifugation, we washed the lipid droplets with a weak detergent to remove small amount of debris and to purify the lipid droplets.

Evaluation of purity is an important step in organelle isolation as it guarantees the quality of the fraction. We performed a Western blot analysis using As-RbcL because it is highly sensitive at detecting the target proteins. Nevertheless, contaminants from organelles other than chloroplast stroma may have been present in the isolated lipid droplet fractions. Better evaluation could be accomplished using antibodies against various other organelle markers. We did not detect any contamination by Rubisco, one of the most abundant proteins in cells.

Acetone extracts of the isolated lipid droplets contained certain carotenoids (Figure 1C). According to previous reports, purified lipid droplets from other microalgae also contained some carotenoids (Moellering and Benning 2010, Peled et al. 2011, Davidi et al. 2012, Vieler et al. 2012). Whether carotenoids are contaminants or native compounds in isolated lipid droplets has not been concluded. In agreement with previous reports of other microalgal lipid droplets (Moellering and Benning 2010, Davidi et al. 2012, Vieler et al. 2012), isolated lipid droplets in *P. tricornutum* mainly contained TAG (Figure 1D).

Characterization of the identified lipid droplet proteins

In this study, we identified five proteins from the isolated lipid droplets, and StLDP was the major protein. The orthologs of StLDP were conserved in the four other diatom genomes that have been determined and published: Phatr48778 with an acyl-CoA binding site, and three other proteins Phatr54019, Phatr45894, and Phatr49981. Previously five other proteins were identified as candidates for lipid droplet proteins from the diatom *F. solaris*. (Nojima et al. 2013). More recently, Maeda et al. (2014) identified the protein g4301 (similar to g12504) in *F. solaris* as a diatom-oleosome-associated protein 1 (DOAP1) and they discussed an ER-targeting signal in DOAP1. We searched for orthologous proteins among the proteins encoded by the genome of *P. tricornutum*. Except for protein g6705, orthologs of the other four proteins were also conserved in the genome of *P. tricornutum* (Table 2); however, none of them was found in our isolated lipid droplets.

Shi et al. (2015) reported that coccolith scale associated protein as a result of alkenone body proteomics and the protein was also upregulated under nitrogen deprivation in *P. tricornutum* in the report of Valenzuela et al. (2012). However the reported protein Phatr55010, was not detected in our experiment.

Chlamydomonas MLDP was more hydrophobic protein than *Arabidopsis* oleosin, mouse perilipin, and ADRP (Moellering and Benning 2010). *Phaeodactylum* StLDP showed higher hydrophobic score on GRAVY index at 0.26 than those of *Chlamydomonas* MLDP at 0.11. Ranking the GRAVY index score on some known lipid droplet proteins is as follows: *Nannochloropsis* LDSP (0.71, AFB75402), *Phaeodactylum* StLDP (0.26, XP_002183367), *Chlamydomonas* MLDP (0.11, XP_001697668), avocado LDAP-1 (-0.10, AGQ04593), *Haematococcus* HOGP (-0.13, ADN95182), *Arabidopsis* oleosin (-0.14, AAA87295), mouse ADRP (-0.28, AEB77763), mouse perilipin (-0.40, NP_783571), and *Auxenochlorella* caleosin (-0.59, AEB77763). Thus, especially the lipid droplet proteins in Stramenopiles, namely LDSP and StLDP seem to have higher hydrophobicity than those of the others.

Hydrophobic region and functional domain of StLDP

StLDP has a hydrophobic domain in the protein central region as well as in oleosin of the plant *A. thaliana* and in LDSP of *Nannochloropsis*. On the other hand, other microalgal lipid droplet proteins, such as MLDP in *Chlamydomonas* and *Dunaliella* and DOAP1 in the *F. solaris*, do not have a

hydrophobic domain. In addition, the typical motif ($PX_9PX_{10}PX_3P$), which was conserved among Heterokontophyta, exists in this hydrophobic region (Figure 4). In the hydrophobic domain of oleosin, there is also a conserved proline knot motif (PX_5SPX_3P, where *X* consists of hydrophobic amino acids) (Tzen et al. 1992). The hydrophobic domain in oleosin makes a hairpin-like loop structure, and the proline knot motif (Abell et al. 2004) is located at the tip of the hairpin loop. The folded-hydrophobic domain can be anchored into the hydrophobic core of the lipid droplet and this structure enables oleosin to localize on the surface of the lipid droplets. We predicted the presence of a transmembrane region of StLDP using TMHMM Server v. 2.0 (Krogh et al. 2001). There were two predicted transmembrane helices in StLDP of *P. tricornutum*; both were 23 amino acid residues (223–245 aa and 252–274 aa). Both of these predicted helices were located in the central hydrophobic region in StLDP and there is a probability that these two helices make a hairpin structure. However, positions for all of the conservative proline residues did not exist between the helices (i.e., at the tip of the hypothetical hairpin loop), unlike the proline knot in oleosin; therefore, the conservative proline residues may have other roles.

When the gene for the oleosin1 protein was knocked out in *Arabidopsis*, the sizes of the lipid droplets increased, and the efficiency of lipid hydrolysis decreased because of a reduced surface area per volume, and germination was also delayed (Siloto et al. 2006). Furthermore, when LDSP was expressed in the oleosin1 knocked-out mutant, the sizes of the lipid droplets recovered but the TAG degradation rate did not recover completely compared with those in the wild type (Vieler et al. 2012). These results indicate that oleosin not only has a structural function but also plays an important role in lipid metabolism. Concerning the enzymatic activities of oleosin, Parthibane et al. (2012) demonstrated that oleosin3 (OLE3) in peanut has both monoacylglycerol acyltransferase and phospholipase activities. OLE3 has *GXSXG* lipase and HX_4D motifs, and these motifs are important for the enzymatic activities in OLE3. Thus, that oleosin itself is related to the biosynthesis and degradation of plant lipids. In the case of StLDP, *GXSXG* or HX_4D motifs did not exist in the entire amino acid sequence. The N-terminal region from the central hydrophobic domain of StLDP was more variable than the C-terminal region when we compared StLDP orthologs in Heterokontophyta (Figure 4). A short sequence in the N-terminal region, which ranged from 164 to 197 amino acid residues in StLDP of *P. tricornutum*, was conserved among the orthologs (black two-headed arrow). On the other hand, approximately 80 amino acid residues located at the C-terminal region, from the next hydrophobic

domain to amino acid 403 in the StLDP were conserved. Interestingly, we noticed that the sequence from 329 to 402 amino acid residues in the C-terminal region of StLDP (red two-headed arrow) was highly conserved among various microorganisms. The proteins that contained these 73 amino acids as a homologous domain were basically conserved in the Stramenopiles (Heterokonta), and they were also found in the Chlorophyta, *Coccomyxa subellipsoidea* C-169 (NCBI ID, XP_005651106), *Micromonas* sp. RCC-299 (XP_002500436), *Micromonas pusilla* CCMP1545 (XP_003062838), the Rhodophyta, *Chondrus crispus* (XP_005715278), the Cryptophyta, *Guillardia theta* CCMP2712 (XP_005838121), the Haptophyta, *Emiliania huxleyi* (XP_005760304), and in bacterium such as Flavobacterium species. It seems that this domain has some kind of function.

Relationship between the expression level of StLDP and the surface area of lipid droplets

The qRT-PCR results indicate that the expression level of StLDP reached a peak at 3 d after N-deprivation, which is very similar to the pattern of growth of the lipid droplets (Figure 5). This was, however, a relatively late response compared with HOGP or MLDP, the green algal lipid droplet proteins. Expression of HOGP reached a maximum level at after 12 h of cultivation in an N-deficient medium (Peled et al. 2011). In the case of *C. reinhardtii*, a maximum expression of MLDP was observed after 24 h of N-deprivation (Moellering and Benning 2010). After 3 d in our experiment, StLDP expression decreased and maintained a steady-state level. Observations with a microscope reveal that the average sizes of the lipid droplets were maximal at 4 d after N-deprivation and maintained thus until 6 d (Figure 5). These results provide indirect evidence that StLDP is a main surface protein on the lipid droplets because the change in expression level corresponded to the change in size of the lipid droplets. A difference in the expression pattern between StLDP and green algal lipid droplet proteins may be driven by a different regulatory mechanism.

Protein transition hypothesis during lipid accumulation

The results presented in Figure 5 indicate that there were at least two stages of formation of the lipid droplets. The first was the early stage, which ranged from inoculation (0 d) to 3 d or 4 d on transcription level or actual lipid droplet growth level, respectively, when the lipid droplets accumulated oils inside. The second was at the late stage after these periods, when the lipid droplets retained their oils. We speculate that the composition or the state of the surface proteins on the lipid droplets changed

between these two stages because the proteins that are required for lipid droplets in each stage were different.

A compositional change in surface proteins of lipid droplets was observed in a study of lipid droplets in mouse adipocytes (Wolins et al. 2003, 2005). According to these reports, TIP47 and S3-12 moved from the cytosol to the surface of nascent lipid droplets during initial fat accumulation. In contrast, perilipin and adipophilin constitutively exist on large or middle-sized lipid droplets and relate to sustaining fat storage and lipolysis.

In this report, we isolated lipid droplets from cells that were cultured for 6 d in N-deficient medium, in other words, we prepared lipid droplet samples at their late stage of formation. Accordingly, we speculate that StLDP has a function for maintenance, distribution, and degradation of the lipids as well as perilipin. In addition, Phatr48778 with an acyl-CoA binding protein, and Phatr54019, which is similar to Hsp70, may play a role to assist the distribution or degradation of TAG in lipid droplets.

Materials and methods

Strain and culture condition

Phaeodactylum tricornutum CCAP 1055/6 was used for all experiments. The cells were cultured in a modified Mann and Myers medium (Mann and Myers 1968) as the normal medium. The composition of the medium is described in Table S1. In the N-deficient culture, we used a N-free medium that did not contain sodium nitrate. The cells were washed with N-free medium three times before inoculation into N-free culture. Cultivation was conducted at 20°C under 200 µmol photons m^{-2} s^{-1} from a continuous white fluorescent lamp. The culture was aerated with filtered air containing 1% (v/v) CO_2. For lipid droplet isolation, we used two 1 L Erlenmeyer flasks with 800 mL of medium for cultivation, and 1.6 L of culture broth was used for each experiment. The cells cultured in the normal medium were transferred into the N-free medium and were then cultured for 6 days to induce the formation of lipid droplets.

Lipid droplet isolation

The cells were harvested by centrifugation at 3000 × *g* for 5 min at room temperature and washed using Tris buffer (10 mM Tris-HCl, pH 7.6) with 2% (w/v) NaCl solution. The following procedures were performed on ice. The harvested cells were re-suspended with sucrose buffer (0.25 M sucrose and protease inhibitor cocktail (cOmplete, Roche Diagnostics) in the Tris buffer), then disrupted using a

French press at 1000 psi. After the disruption, the unbroken cells and other organelles were removed as a pellet by centrifugation at 50,000 × *g* for 5 min at 4°C. The surface layer of the supernatant containing the lipid droplets was collected into new tubes. Then, 1 mL of 2.5 M sucrose solution was added to 4 mL of the collected supernatant to adjust the sucrose concentration to 0.7 M. We gently poured the Tris buffer on the 0.7 M sucrose solution to make a discontinuous sucrose gradient layer. The tubes were centrifuged again at 50,000 × *g* for 20 min at 4°C. Lipid droplets on the surface of the solution were collected into ten 1.5 mL tubes and then centrifuged at 20,000 × *g* for 10 min at 4°C to remove any remnant buffer at the bottom. The remaining lipid droplets in the ten tubes were gathered into one tube and then re-centrifuged and as much buffer solution was discarded as possible. After the concentration of lipid droplets, 1 mL of weak detergent buffer (0.2% (v/v) Triton X-100 in the Tris buffer) was added to the tube of lipid droplets and incubated on ice for 10 min to remove the debris. Then, the tube was centrifuged at 20,000 × *g* for 10 min at 4°C and the weak detergent buffer was discarded. After the detergent treatment, the lipid droplets were washed twice with the Tris buffer.

Sample preparation for lipid and protein analysis

Cooled-acetone was added to the lipid droplet fraction and incubated at −20°C overnight. The sample tube was centrifuged at 20,000 × *g* for 10 min at 0°C. The lipid-containing acetone supernatant was collected for UV-visible spectrophotometry and lipid analysis. Furthermore, cooled-ethyl acetate was added to the sample to eliminate any residual oil component from the protein precipitate and incubated at −20 °C for 2 h. The tube was centrifuged and the ethyl acetate supernatant was removed. The precipitated protein fraction was dissolved in 6 μL of lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS, 3% Triton X-100, and 2% SDS). For the protein preparation from whole cells, lysis buffer was directly added to the harvested cells, incubated on ice for 30 min and centrifuged. The supernatant was used as the protein extract from whole cells.

UV-visible spectrophotometry and silica-gel thin layer chromatography

The crude lipids from whole cells were also extracted by acetone with sonicator treatment. These acetone extracts of lipid droplets and whole cells were analyzed using an UV-visible spectrophotometer (UV-1800, UV spectrophotometer, Shimadzu). The absorbance of the 350–750 nm wavelength was measured to evaluate the contamination level, especially of chloroplasts. After the

spectrophotometric analysis, the acetone and ethyl acetate extracts described above were evaporated under a N_2 stream and weighed gravimetrically. The lipid extracts were used for silica-gel TLC.

The lipid extracts were re-dissolved in the measured amounts of solvent to adjust the concentration and following volumes of them were taken to TLC plate with micro syringe. For each lipid extract, 25 μg of crude lipid from whole cells, 10 μg of lipid droplet fraction, and 10 μg of TAG standard (triolein) were spotted onto a silica-gel TLC plate (HPTLC Silica gel 60 F_{254}, Merch). Hexane:chloroform 1:1 (v/v) was used as a developing solvent. After development, 20% sulfuric acid was sprayed onto the plate and then the plate was heated to visualize the lipid spots.

SDS-PAGE and Western blotting

A 6-μL aliquot of the protein sample solution from the lipid droplets was mixed with 2 μL of 4× SDS sample buffer (0.25 M Tris-HCl (pH 6.8), 8% SDS, 20% sucrose, and 0.008% bromophenol blue), then 0.8 μL of 500 mM dithiothreitol was added to the mixture and incubated at room temperature for 1 h to denature the proteins. Half of the prepared solution (4.4 μL) was used for proteomic analysis and the other half was used for Western blotting against As-RbcL. A protein sample from the whole cell was also prepared in the same manner.

For the proteomic analysis, protein electrophoresis was performed in Novex 12% Tris-Glycine gel (Invitrogen). The gel was then fixed with a solution that consisted of 40% MeOH and 10% acetic acid for 15 min and stained using GelCode Blue Stain Reagent (Thermo Scientific) for 30 min.

For the Western blotting, protein electrophoresis was performed in c-PAGEL C-12.5L minigel (Atto Corporation) as described previously (Tsuji et al. 2012). Universal rbcL antibody (Agrisera) which diluted with blocking buffer at 1:20,000 was used as primary antibody.

Peptide preparation for mass spectrometry

The proteins separated by SDS-PAGE were sliced into approximately 1-mm^3 pieces. The gel slices were destained and digested with sequence-grade modified trypsin (Promega, Madison, WI, USA) as described previously (Katayama et al. 2001) with minor modifications. After digestion, the peptides were extracted from the gel pieces with acetonitrile:5 % (v/v) formic acid aqueous solution 1:1 (v/v). The extracted solution was recovered into a vial.

MS analysis and database search

The digested peptides were separated by HPLC with a capillary pump (Agilent 1200 series) equipped with ZORBAX 300SB-C18 (0.3 mm × 150 mm, Agilent) column. For the elution, mobile phase A consisted of H_2O:acetonitrile 95:5 (v/v) containing 0.1% formic acid, and mobile phase B consisted of H_2O:acetonitrile 10:90 (v/v) containing 0.1 % formic acid. The peptide samples were eluted at 5 µL min^{-1} under the following gradient: ratio of mobile phase B started at 5% and increased to 50% for 60 min, then B ratio elevated quickly to 95% for 1 min and was maintained at 95% for 14 min. The eluted peptides were applied to an electrospray ionization quadrupole time-of-flight (ESI-Q/TOF) system (Agilent 6520 Accurate-Mass QTOF LC/MS). The MS scan range was set to m/z 105–3000 and multi-charged ions (+2, +3, and >+3) were preferentially subjected to MS/MS analysis. The obtained data were exported as Mascot generic files and then each corresponding protein was searched for in the genome database of *P. tricornutum* obtained from JGI (http://genome.jgipsf.org/Phatr2/Phatr2.home.html; *Phaeodactylum tricornutum* v2.0) using Mascot Server (version 2.2.06, Matrix Science). A BLASTp search for the identified proteins was performed at NCBI (http://www.ncbi.nlm.nih.gov/) and JGI (http://genome.jgi.doe.g.,ov/) websites. Prediction of the transmembrane helix was performed at TMHMM Server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/, Krogh et al. 2001).

Real-time qRT-PCR

After harvesting the cells, respective samples were immediately frozen with liquid N_2 and kept at −80°C until RNA extraction. A bead beater was used for cell disruption. TRIzol reagent and PureLink RNA Mini Kit (Invitrogen) were used for the RNA extraction. The qualities of total RNA extracts were checked by MOPS-agarose gel electrophoresis.

We performed qRT-PCR and then conducted the analysis using the comparative C_t method (Livak and Schmittgen 2001). The actin12 gene was used as the housekeeping gene. We used the primer set that was previously designed for actin12 (Siaut et al. 2007). For the detection of mRNA of DLDP, we used the following primers: (F-) 5'-GCCTGGTTTCGTTTCGTTG-3' and (R-) 5'-AAGACGGCGACAATCGGTA-3'. SuperscriptIII Platinum SYBR Green qRT-PCR Kit (Invitrogen) was used for the preparation of the reaction mixture. StepOnePlus (Applied Biosynthesis) was used for executing qRT-PCR.

Funding

This study was supported by the Maekawa Houonkai Foundation and the Japan Society for the Promotion of Science (JSPS) Grant No.23770234 to M. Y.

Disclosures

Conflicts of interest: No conflicts of interest declared.

Acknowledgments

We are grateful to Drs. Y. Tsuji, H. Araie and M. Baba for experimental supports and helpful suggestions and to colleagues in Laboratory of Plant Physiology and Metabolism in University of Tsukuba for having active discussions. And we also appreciate Prof. T. Kuroiwa and research fellows in Laboratory of cell biology in Rikkyo University for giving us permission to use MALDI-TOF mass spectrometer for proteomic analysis on preliminary experiment. This study was financially supported by JSPS KAKENHI Grant Number 23770234 and Mayekawa Houonkai Foundation to MY.

References

Acién Fernández, F. G., Hall, D. O., Cañizares Guerrero, E., Krishna Rao, K and Molina Grima, E. (2003) Outdoor production of *Phaeodactylum tricornutum* biomass in a helical reactor. J. Biothechnol. 103: 137-152.

Abell, B. M., Hahn, M., Holbrook, L. A. and Moloney, M. M. (2004) Membrane topology and sequence requirements for oil body targeting of oleosin. Plant J. 37: 461-470.

Brasaemle D. L. (2007) The perilipin family of structural lipid droplet proteins: stabilization of lipid droplets and control of lipolysis. J. Lipid. *Res.* 48: 2547-2559.

Bréhélin, C., Kessler, F. and Wijk, K. J. V. (2007) Plastoglobules: versatile lipoprotein particles in plastids. Trends. Plant. Sci. 12: 260-266.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456: 239-244.

Chapman, K. D., Dyer, J. M. and Mullen, R. T. (2012) Biogenesis and functions of lipid droplets in plants. J. Lipid. Res. 53: 215-216.

Chisti Y. (2007) Biodiesel from microalgae. Biotechnol. *Adv.*25: 294-306.

Davidi, L., Katz, A. and Pick, U. (2012) Characterization of major lipid droplet proteins from *Dunaliella*. Planta 236:19-33.

Davidi, L., Levin, Y., Ben-Dor, S., and Pick, U. (2015) Proteomic analysis of cytoplasmatic and plastidic β-carotene lipid droplets in Dunaliella bardawil. Plant Physiol. 167: 60-79.

Ding, Y., Zhang, S., Yang, Li., Na H., Zhang, P., Zhang, H., and Wang, Y. et al. (2013) Isolating lipid droplets from multiple species. Nat. Protoc. 8: 43-51.

Fernández Sevilla, J. M., Cerón García, M. C., Sánchez Mirón, A., El Hassan, B., García Camacho, F. and Molina Grima, E. (2004) Pilot-plant-scale outdoor mixotrophic cultures of *Phaeodactylum tricornutum* using glycerol in vertical bubble column and airlift photobioreactors: studies in fed-batch mode. Biotechnol.Progr. 20: 728-736.

Frandsen, G. I., Mundy, J. and Tzen, J. T. C. (2001) Oil bodies and their associated proteins, oleosin and caleosin. Physiol.Plantarum 112: 301-307.

Goodson, C., Roth, R., Wang, Z. T. and Goodenough, U. (2011) Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas reinhardtii* and stimulation of lipid body

production with acetate boost. Eukaryot. Cell 10: 1592-1606.

Goold, H., Beisson, F., Peltier, G., and Li-Beisson, Y. (2015) Microalgal lipid droplets: composition, diversity, biogenesis and functions. Plant Cell Rep. 34: 545-555.

Katayama, H., Nagasu, T. and Oda, Y. (2001) Improvement of in-gel digestion protocol for peptide mass fingerprinting by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun.Mass Spectrom. 15:1416-1421.

Krogh, A., Larsson, B., Heijne, G. V. and Sonnhammer, E. L. L. (2001) Predicting transmembrane protein topology with a Hidden Markov Model: Application to complete genomes. J. Mol. Biol. 305: 567-580.

Lin, I., Jiang, P., Chen, C. and Tzen, J. T. C. (2012) A unique caleosin serving as the major integral protein in oil bodies isolated from *Chlorella* sp. cells cultured with limited nitrogen. Plant Physiol.Biochem. 61:80-87

Livak, K. J. and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-CΔΔ(T)). Methods 25: 402-408.

Maeda, Y., Sunaga, Y., Yoshino, T. and Tanaka, T. (2014) Oleosome-associated protein of the oleaginous diatom *Fistulifera solaris* contains an endoplasmic reticulum-targeting signal sequence. Mar. Drugs 12: 3892-3903.

Mann, J. E. and Myers, J. (1968) On pigments, growth, and photosynthesis of *Phaeodactylum tricornutum*. J. Phycol. 4: 349-355.

Martin, S. and Parton, R. G. (2006) Lipid droplets: a unified view of a dynamic organelle. Nat.Rev. Mol. Cell.Bio. 7: 373-378.

Murphy, D. J. and Vance, J. (1999) Mechanisms of lipid-body formation. Trends. Biochem. Sci. 24: 109-115.

Murphy, D. J. (2012) The dynamic roles of intracellular lipid droplets: from archaea to mammals. Protoplasma 249: 541-585.

Moellering, E. R. and Benning, C. (2010) RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii*. Eukaryot. Cell 9: 97-106.

Nguyen, H. M., Baudet, M., Cuine, S., Adriano, J. M., Barthe, D., Billon, E. et al. (2011) Proteomic profiling of oil bodies isolated from the unicellular green microalga *Chlamydomonas reinhardtii*: With focus on proteins involved in lipid metabolism. Proteomics 11:4266-4273.

Nojima, D., Yoshino, T., Maeda, Y., Tanaka, M., Nemoto, M. and Tanaka, T. (2013) Proteomics analysis of oil body associated proteins in the oleaginous diatom. J. Proteome res.12: 5293-5301.

Ohsaki, Y., Suzuki, M., and Fujimoto, T. (2014) Open questions in lipid droplet biology. Chem. Biol. 21: 86-96.

Parthibane, V., Rajakumari, S., Venkateshwari, V., Iyappan, R. and Rajasekharan, R. (2012) Oleosin is bifunctional enzyme that has both monoacyltransferase and phospholipase activities. J. Biol.Chem. 287: 1946-1954.

Pasaribu, B., Lin, I. P., Tzen, J. T. C., Jauh, G. Y., Fan, T. Y., Ju, Y. M., et al. (2014) SLDP: a novel protein related to caleosin is associated with the endosymbiotic *Symbiodinium* lipid droplets from *Euphyllia glabrescens*. Mar. Biotechnol. 16: 560-571.

Peled, E., Leu, S., Zarka, A., Weiss, M., Pick, U., Khozin-Goldberg, I., et al. (2011) Isolation of a novel oil globule protein from the green alga *Haematococcus pluvialis* (Chlorophyceae). Lipids 46: 851-861.

Riisberg, I., Orr, R. J. S., Kluge, R., Shalchian-Tabrizi, K., Bowers, H. A., Patil, V., et al. (2009) Seven Gene Phylogeny of Heterokonts. Protist 160: 191-204.Pol, A., Gross, S. P., and Parton, R. G. (2014) Biogenesis of the multifunctional lipid droplet: lipids, proteins, and sites. J. Cell Biol. 204: 635-646.

Shi, Q., Araie, H., Bakku, R. K., Fukao, Y., Rakwal, R., Suzuki, I., and Shiraiwa, Y. (2015) Proteomic analysis of lipid body from the alkenone-producing marine haptophyte alga *Tisochrysis lutea*. Proteomics. *in press*.

Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A. et al. (2007) Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. Gene 406: 23-35.

Siloto, R. M., Findlay, K., Lopez-Villalobos, A., Yeung, E. C., NykiforukC, L. and Moloney, M. M. (2006) The accumulation of oleosins determines the size of seed oilbodies in *Arabidopsis*. Plant Cell 18: 1961-1974

Spanova, M., Czabany, T., Zellnig, G., Leitner, E., Hapala, I. and Daum, G. (2010) Effect of lipid particle biogenesis on the subcellular distribution of squalene in the yeast *Saccharomyces cerevisiae*. J. Biol. Chem. 285: 6127-6133

Tsuji, Y., Suzuki, I., and Shiraiwa, Y. (2012) Enzymological evidence for the function of plastid-located pyruvate carboxylase in the haptophyte alga *Emiliania huxleyi*: a novel pathway for the production of C_4 compounds. Plant Cell Physiol. 53: 1043-1052.

Tzen, J. T. H., Lie, G. C., and Huang, A. H. C. (1992) Characterization of the charged components and

their topology on the surface of plant seed oil bodies. J Biol. Chem. 267: 15626-15634.

Valenzuela, J., Mazurie, A., Carlson, R. P., Gerlach, R., Cooksey, K. E., Peyton, B. M., et al. (2012) Potential role of multiple carbon fixation pathway during lipid accumulation in *Phaeodactylum tricornutum*. Biotechnol. Biofuels 5: 40.

Vieler, A., Brubaker, S. B., Vick, B. and Benning, C. (2012) A Lipid droplet protein on *Nannochloropsis* with functions partially analogous to plant oleosins. Plant Physiol. 158: 1562-1569.

Wang, H., Gilham, D. and Lehner, R. (2007) Proteomic and lipid characterization of Apolipoprotein B-free luminal lipid droplets from mouse liver microsomes. J. Biol. Chem. 282: 33218-33226.

Wolins, N. E., Skinner, J. R., Schoenfish, M. J., Tzekov, A., Bensch, K. G. and Bickel, P. E. (2003) Adipocyte protein S3-12 coats nascent lipid droplets. J. Biol. Chem. 278: 37713-37721.

Wolins, N. E., Quaynor, B. K., Skinner, J. R., Schoenfish, M. J., Tzekov, A. and Bickel, P. E. (2005) S3-12, Adipophilin, and TIP47 package lipid in Adipocytes. J. Biol. Chem. 280: 19146-19155.

Wolins, N. E., Brasaemle, D. L. and Bickel, P. E. (2006) A proposed model of fat packaging by exchangeable lipid droplet proteins. FEBS Lett. 580: 5484-5491

Yang, L., Ding, Y., Chen, Y., Zhang, S., Huo, C., Wang, Y., et al. (2012) The proteomics of lipid droplets: structure, dynamics, and functions of organelle conserved from bacteria to humans. J. Lipid. Res. 53: 1245-1253

Yang, Z., Niu, Y., Ma, Y., Xue, J., Zhang, M., Yang, W., et al. (2013) Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. Biotechnol. for biofuels 6:67

Legends to figures

Figure 1. Quality evaluation of the isolated lipid droplet fraction. (A, B) Microscopic image of a Nile-red stained lipid droplet. Scale bars indicate 20 μm. (A) Differential interference contrast image. (B) Fluorescence image. (C) Absorption spectra of acetone extracts obtained from nitrate (N) deprived whole cells (black line), chloroplast (green line), and lipid droplet (yellow line). (D) Silica-gel thin layer chromatography of lipid extracts obtained from whole cells (+N, nitrate sufficient; −N, nitrate deficient); lipid droplets from –N cells (LD); and triglycerol standard, triolein (std). P/S indicates peak of developing solvent. (E) Western blotting results against rbcL antibody. Lane 1–5: proteins from the whole cell. Lane 6–8: proteins from the lipid droplet fraction (replicate 1–3, respectively). Lane 9–11: proteins from the whole cell extract washed with acetone in the same manner as the lipid droplet fraction. M indicates the marker lane. (F) Magnified image of the white dotted area in E. R1–3 indicates samples from the three biological replicates.

Figure 2. SDS-PAGE gel image of proteins from the whole cell extract and the lipid droplet fraction. All lanes of the lipid droplet fraction (R1–3) were sliced into numbered gel pieces for proteomic analysis, as indicated at the right side.

Figure 3. Hydropathy plots of StLDP, oleosin, and LDSP. For this analysis we used an amino acid sequence of StLDP (*Phaeodactylum tricornutum*, NCBI accession no. XP_002183367), Oleosin (*Arabidopsis thaliana,* GenBank: AAA87295), and LDSP (*Nannochloropsis* sp. CCMP1779, GenBank: AFB75402).

Figure 4. Multiple sequence alignment of StLDP and homologs. The red dotted square indicates the hydrophobic domain conserved in the StLDP homologs. Small red arrows indicate the conservative proline residue in the hydrophobic region. The black two-headed arrow indicates a short conserved domain in the N-terminal side and the red two-headed arrow indicates a widely conserved domain in the C-terminal side. Fracyl167885 (*Fragilariopsis cylindrus,* JGI protein ID: 167885), Psemu21664 (*Pseudo-nitzschia multiseries,* JGI protein ID: 21664), Phatr48859 (*Phaeodactylum tricornutum,* JGI protein ID: 48859), Thaps23929 (*Thalassiosira pseudonana*, JGI protein ID: 23929), Thaoc30415 (*T. oceanica,* JGI protein ID: 30415), Naga100005g121 (*Nannochloropsis gaditana,* GenBank: EWM25464), and EctsiCBN75389 (*Ectocarpus siliculosus*, GenBank: CBN75389).

Figure 5. Changes in StLDP mRNA expression level and lipid droplet diameter during nitrate deprivation. Expression level was normalized using the housekeeping gene actin12 and the comparative C_t method; the error bars indicate S.D. (n = 3). Lipid droplet diameters were determined using the Nile-red stained cell image and the scale bars indicate S.E. values (n > 80).

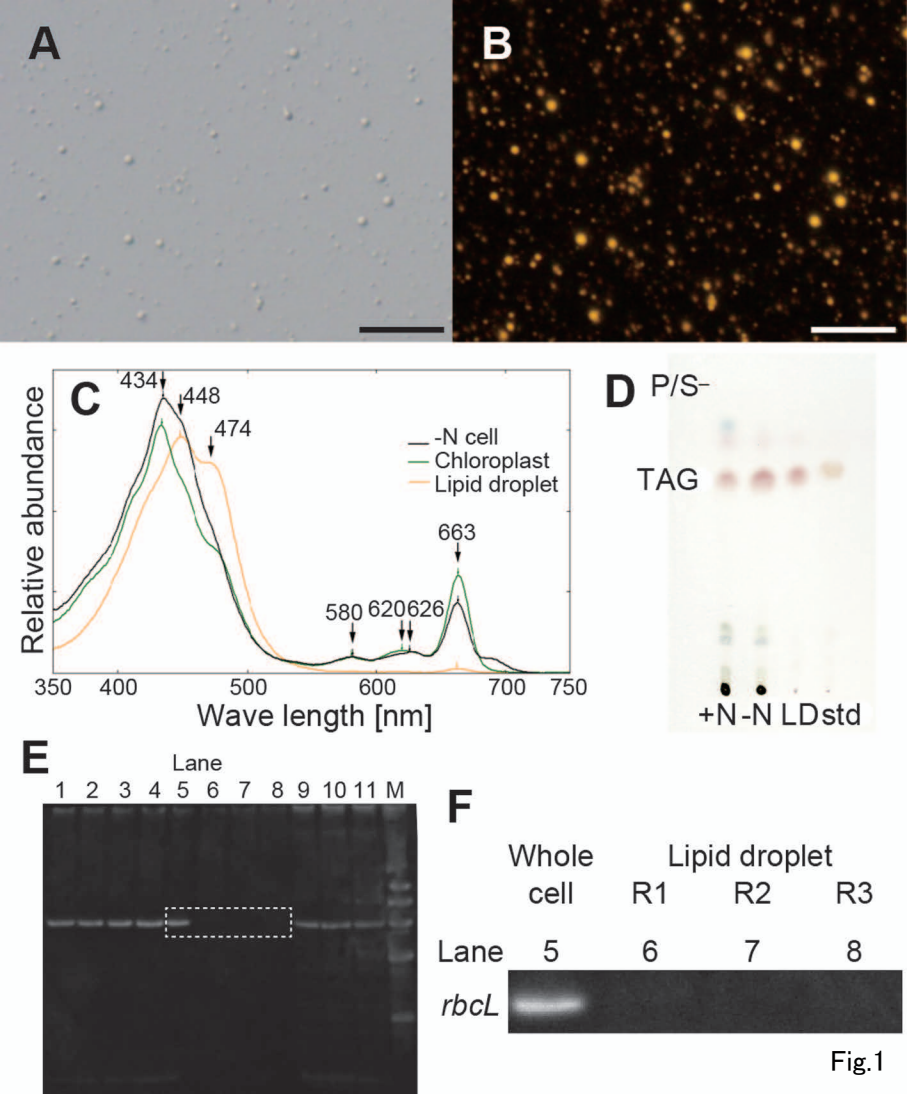
A
B
C
434
448
474
663
580
620
626
-N cell
Chloroplast
Lipid droplet
Relative abundance
350
400
500
600
700
750
Wave length [nm]
D
P/S-
TAG
+N
-N
LD
std
E
Lane
1 2 3 4 5 6 7 8 9 10 11 M
F
Whole cell
Lipid droplet
R1
R2
R3
Lane
5
6
7
8
rbcL

Fig.1

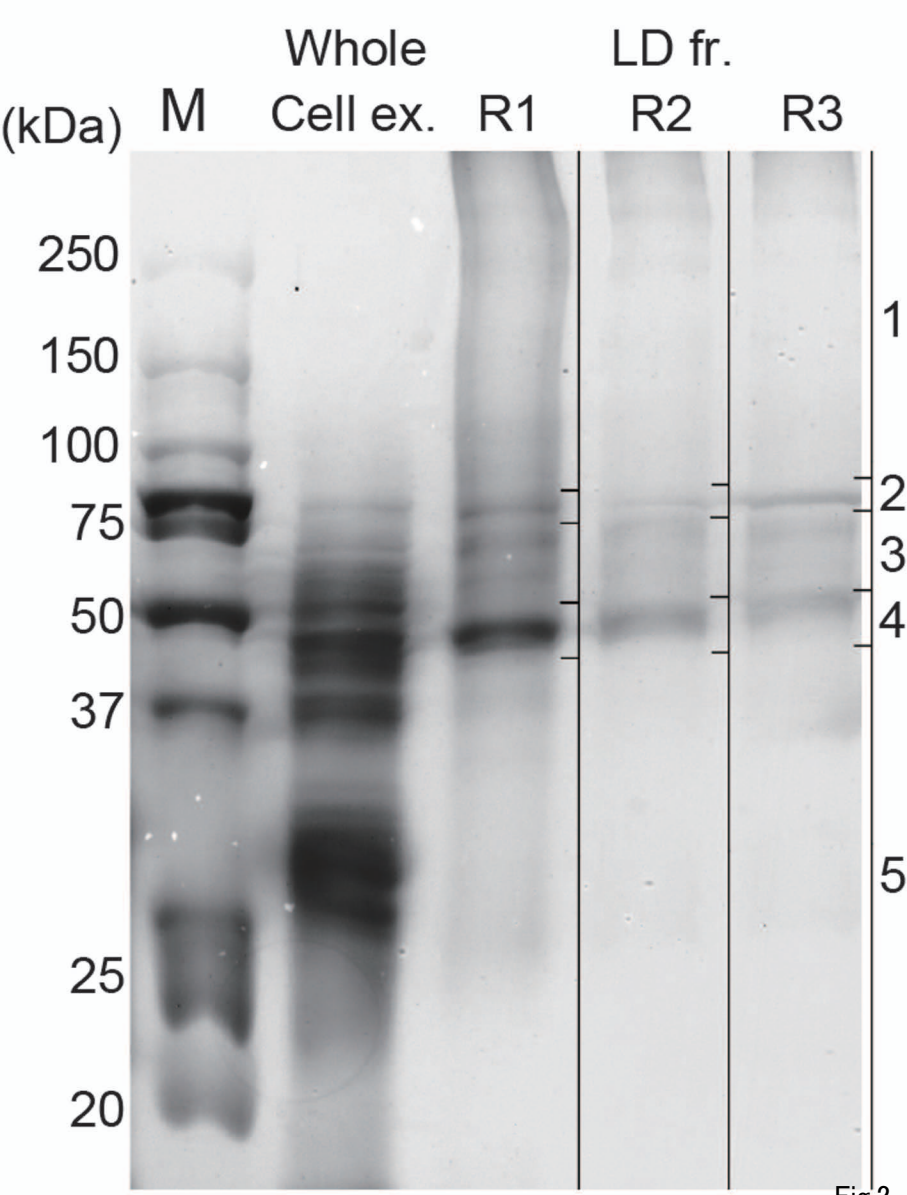

Fig.2

Tables

Table 1. List of the proteins identified from the lipid droplet fraction.

Gel fraction	Mascot score			Sequence coverage [%]			Protein ID	Protein name	Functional domain (amino acid region)	Amino acid number	mol wt [kDa]	NCBI reference no	Similar protein in *Thalassiosira pseudonana*
	R1	R2	R3	R1	R2	R3							
4	1113	378	190	43	41	34	Phatr48859	Stramenopile-type Lipid Droplet Protein, StLDP	none	456	48.774	XP_002183367	hypothetical protein [XP_002292405]
5	50	45	41	13	5	9	Phatr48778	acyl-CoA binding protein	Acyl-CoA binding region (14-96)	351	38.195	XP_002183443	acyl-CoA binding protein, partial

													[XP_002289611]
2	60	-	-	33	-	-	Phatr54019	heat shock protein, Hsp70	HspA1-2,6-8 like nucleotide-binding domain (7-382)	653	70.831	XP_002177351	heat shock protein 70 [XP_002291508]
4	40	-	-	20	-	-	Phatr45894	hypothetical protein	Thioredoxin-like (171-260)	354	37.97	XP_002180271	hypothetical protein [XP_002286222]
1	-	-	37	-	-	5	Phatr49981	hypothetical protein	2OG-Fe(II) oxygenase superfamily (102-323), SAD/SRA domain	544	60.786	XP_002184813	hypothetical protein [XP_002294418]

(363-476)

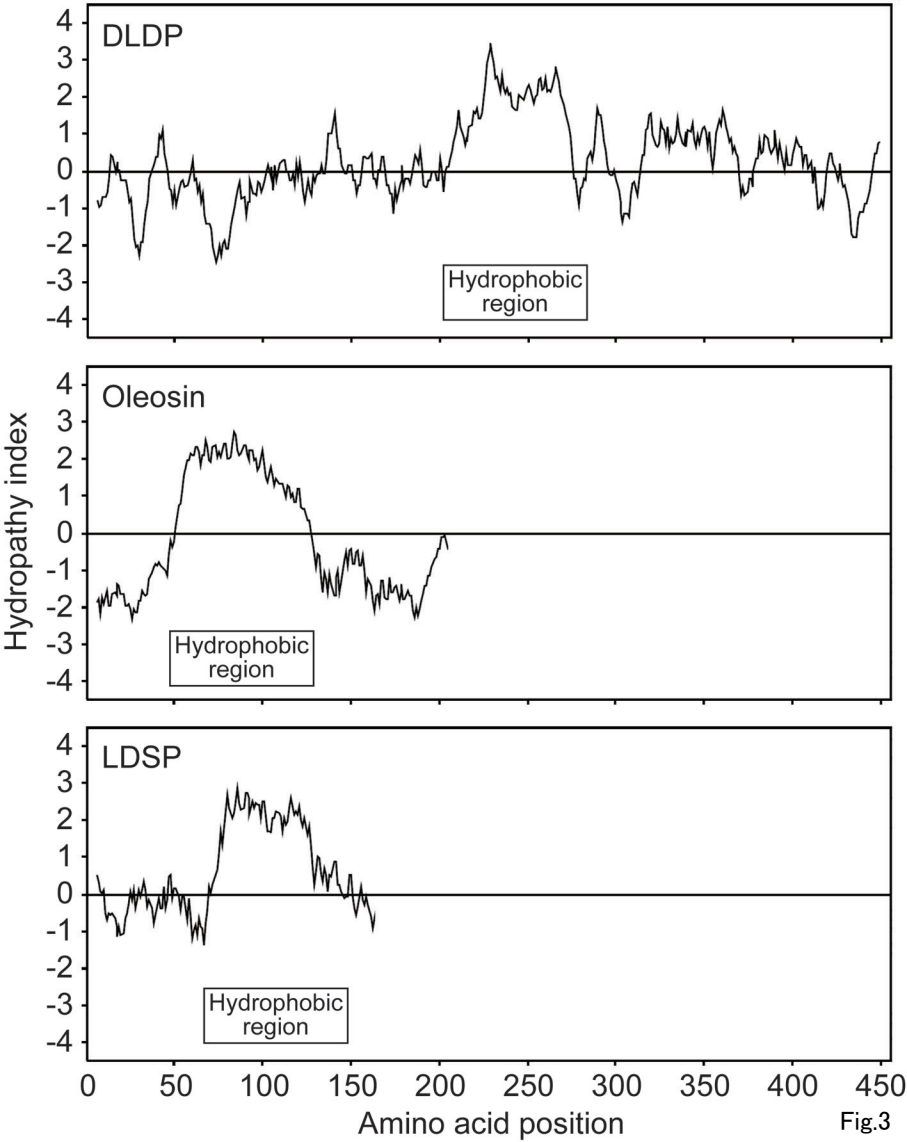
DLDP
Hydrophobic region
Oleosin
Hydrophobic region
LDSP
Hydrophobic region
Hydropathy index
4
3
2
1
0
-1
-2
-3
-4
0
50
100
150
200
250
300
350
400
450
Amino acid position

Fig.3

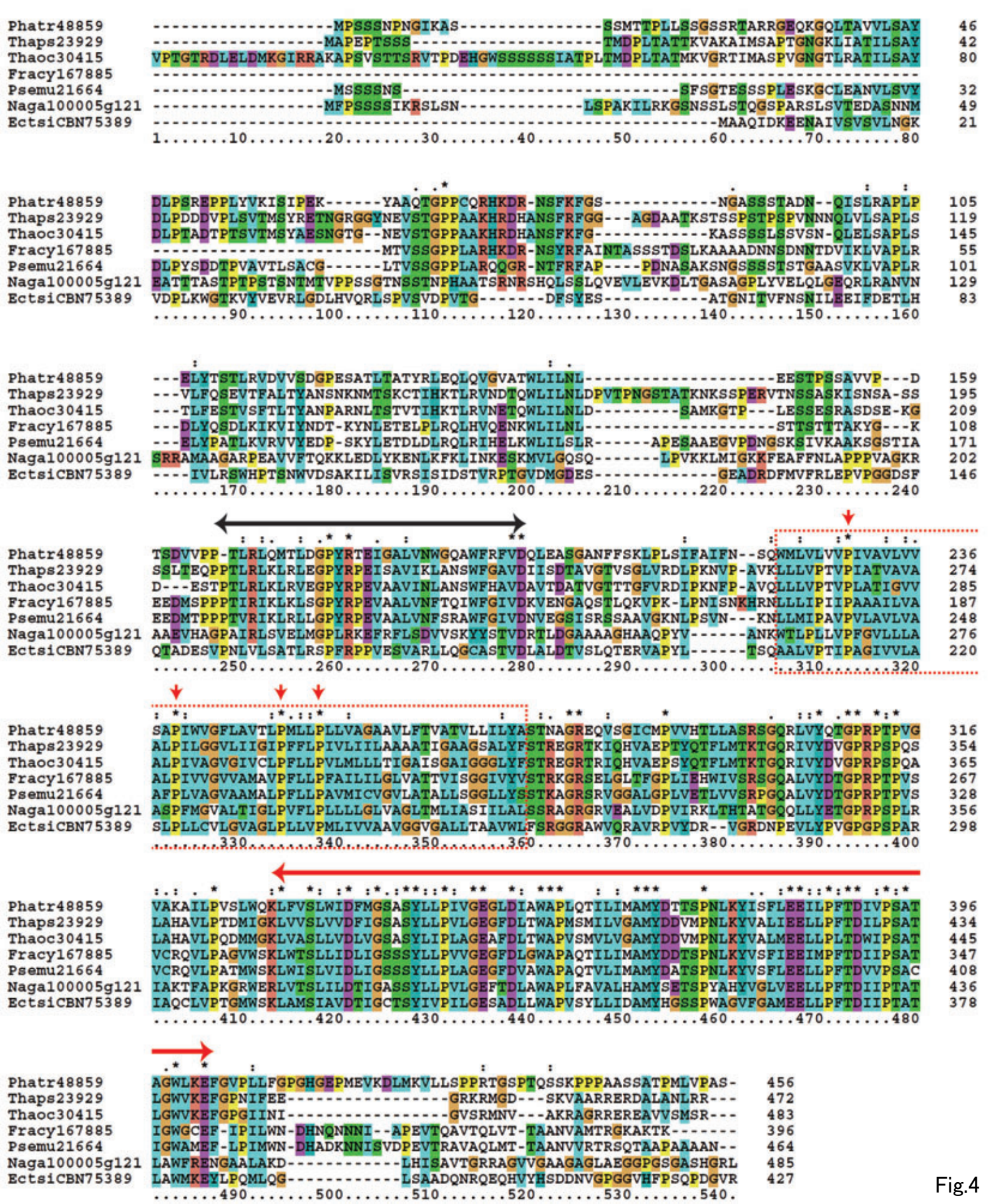
Phatr48859
Thaps23929
Thaoc30415
Fracy167885
Psemu21664
Naga100005g121
EctsiCBN75389

Fig.4

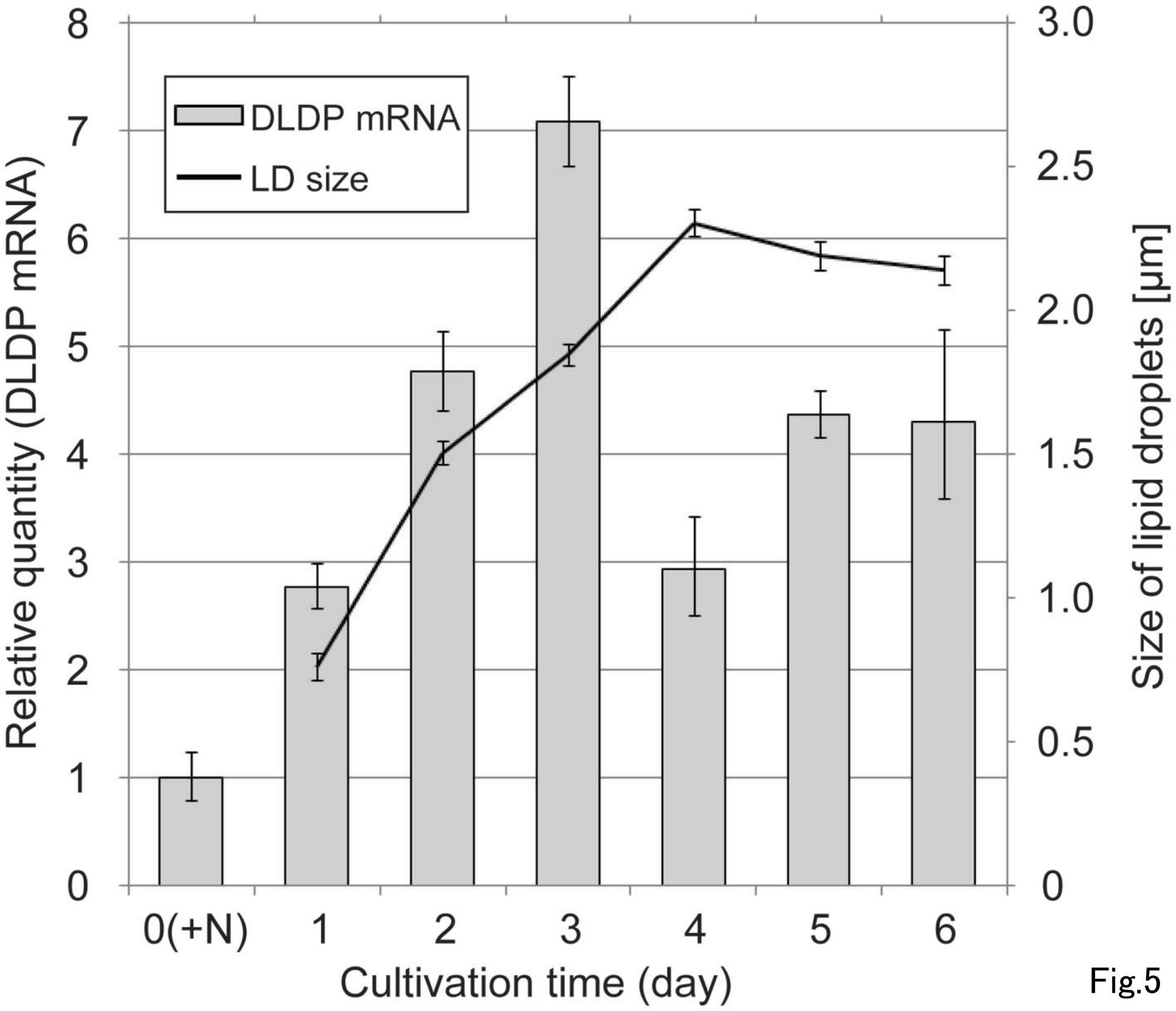
DLDP mRNA
LD size
Relative quantity (DLDP mRNA)
Size of lipid droplets [μm]
Cultivation time (day)
0(+N)
1
2
3
4
5
6
0
1
2
3
4
5
6
7
8
0
0.5
1.0
1.5
2.0
2.5
3.0

Fig.5

Table 2. List of the proteins identified in a previous report (Nojima et al. 2013) and its homologs in *P. tricornutum*.

ProteinID	Putative function	mol wt [kDa]	Homolog in *Phaeodactylum*
g4796	transmembrane protein	85	Phatr44488/50592
g6705	ABC transporter, partial	149	-
g6574	potassium channel	58	Phatr13578
g4301	unknown, DOAP1	53	Phatr48876
g5708	unknown	49	Phatr45146

Supplementary Figure S1. Multiple sequence alignment of StLDP orthologs. The red dotted square indicates the hydrophobic domain in the StLDP homologs. Small red arrows indicate conservative proline residues in the hydrophobic region. The black two-headed arrow indicates a short conserved domain in the N-terminal side and the red two-headed arrow indicates a widely conserved domain in the C-terminal side.

Amino acid sequences used in this analysis are as follows; Phatr48859 (*Phaeodactylum tricornutum*,JGI ID: 48859), Fracyl167885 (*Fragilariopsis cylindrus*, JGI ID: 167885), Psemu21664 (*Pseudo-nitzschia multiseries*, JGI ID: 21664), Thaps23929 (*Thalassiosira pseudonana*, JGI ID: 23929), Thaoc30415 (*Thalassiosira oceanica*, JGI ID: 30415), Naga100005g121 (*Nannochloropsis gaditana*, GenBank: EWM25464), EctsiCBN75389 (*Ectocarpus siliculosus*, GenBank: CBN75389), Aplke97617 (*Aplanochytrium kerguelense*, JGI ID: 97617), Aurli151081 (*Aurantiochytrium limacinum*, JGI ID: 151081), Schag94104 (*Schizochytrium aggregatum*, JGI ID: 94104), Phyci422862 (*Phytophthora cinnamomi*, JGI ID: 422862), Physo532873 (*Phytophthora sojae*, JGI ID: 532873), Guith161621 (*Guillardia theta*, JGI ID: 161621), and Emihu310878 (*Emiliania huxleyi*, JGI ID: 310878).

Supplementary Table S1. Composition of the modified Mann and Myers medium.

Compound	
$NaNO_3$	50 mg
K_2HPO_4	5 mg
Na_2SiO_3 $9H_2O$	30 mg
Vitamin B_{12}	0.05 µg
Biotin	0.05 µg
Thiamine HCl	10 µg
NaCl	1.5 g
$MgSO_4$ $7H_2O$	360 mg
KCl	180 mg
$CaCl_2$ $2H_2O$	120 mg
Na_2EDTA $2H_2O$	30 mg
H_3BO_3	6 mg
$FeSO_4$ $7H_2O$	2 mg
$MnCl_2$ $4H_2O$	1.4 mg
$ZnSO_4$ $7H_2O$	33 µg
$Co(NO_3)_2$ $6H_2O$	7 µg
$CuSO_4$ $6H_2O$	2 µg
Tris (hydroxymethyl) aminomethane	100 mg
pH was adjusted with HCl at 8.0	
Distilled water	100 mL