New Insights into the Nature and Phylogeny of Prasinophyte Antenna Proteins: *Ostreococcus tauri*, a Case Study

Christophe Six,[1] *Alexandra Z. Worden,*†[1] *Francisco Rodríguez,** *Hervé Moreau,*‡ *and Frédéric Partensky**

*Station Biologique, Roscoff cedex, France; †Marine Biology and Fisheries Division, Rosenstiel School of Marine and Atmospheric Science, University of Miami; and ‡Laboratoire Arago, Banyuls sur mer cedex, France

The basal position of the Mamiellales (Prasinophyceae) within the green lineage makes these unicellular organisms key to elucidating early stages in the evolution of chlorophyll *a*/*b*–binding light-harvesting complexes (LHCs). Here, we unveil the complete and unexpected diversity of Lhc proteins in *Ostreococcus tauri*, a member of the Mamiellales order, based on results from complete genome sequencing. Like *Mantoniella squamata*, *O. tauri* possesses a number of genes encoding an unusual prasinophyte-specific Lhc protein type herein designated "Lhcp". Biochemical characterization of the complexes revealed that these polypeptides, which bind chlorophylls *a*, *b*, and a chlorophyll *c*–like pigment (Mg-2,4-divinyl-phaeoporphyrin a_5 monomethyl ester) as well as a number of unusual carotenoids, are likely predominant. They are retrieved to some extent in both reaction center I (RCI)– and RCII-enriched fractions, suggesting a possible association to both photosystems. However, in sharp contrast to previous reports on LHCs of *M. squamata*, *O. tauri* also possesses other LHC subpopulations, including LHCI proteins (encoded by five distinct *Lhca* genes) and the minor LHCII polypeptides, CP26 and CP29. Using an antibody against plant Lhca2, we unambiguously show that LHCI proteins are present not only in *O. tauri*, in which they are likely associated to RCI, but also in other Mamiellales, including *M. squamata*. With the exception of *Lhcp* genes, all the identified *Lhc* genes are present in single copy only. Overall, the discovery of LHCI proteins in these prasinophytes, combined with the lack of the major LHCII polypeptides found in higher plants or other green algae, supports the hypothesis that the latter proteins appeared subsequent to LHCI proteins. The major LHC of prasinophytes might have arisen prior to the LHCII of other chlorophyll *a*/*b*–containing organisms, possibly by divergence of a LHCI gene precursor. However, the discovery in *O. tauri* of CP26-like proteins, phylogenetically placed at the base of the major LHCII protein clades, yields new insight to the origin of these antenna proteins, which have evolved separately in higher plants and green algae. Its diverse but numerically limited suite of *Lhc* genes renders *O. tauri* an exceptional model system for future research on the evolution and function of LHC components.

Introduction

Photosynthetic organisms use light-harvesting antennae to collect and transfer solar energy to photosystem (PS) reaction centers (RC) where it is transformed into life-supporting chemical energy. The most common antenna systems found in oxyphototrophic eukaryotes, including both higher plants and green algae, are light-harvesting complexes (LHCs) consisting of assemblages of several different, but structurally similar, Lhc proteins binding chlorophylls (chls) and carotenoids. These pigments are bound at specific sites along Lhc proteins, each of which possesses three membrane-spanning helices (MSHs) and one or two short amphipathic helices (Kühlbrandt, Wang, and Fujiyoshi 1994; Ben Shem, Frolow, and Nelson 2003; Liu et al. 2004). The first and third MSHs of Lhc proteins show a high level of similarity to one another and display a twofold symmetry, whereas their second MSH has a distinct sequence and is more variable among members of the Lhc protein family (Green and Kühlbrandt 1995). Despite being located in the thylakoid membranes, all Lhc proteins are encoded by nuclear genes, in contrast to RCI and RCII proteins, as well as the PSII core antenna polypeptides (CP43 and CP47) which are chloroplast encoded (Green, Pichersky, and Kloppstech 1991; Elrad and Grossman 2004).

In higher plants, the major antenna complexes have been shown to be associated either specifically with PSI or predominantly with PSII and have been named LHCI and LHCII, respectively. Studies resolving the atomic structure of the LHCII of plants (Kühlbrandt, Wang, and Fujiyoshi 1994; Liu et al. 2004) have led to a comprehensive understanding of the overall organization of this antenna system. Each PSII possesses a permanent inner antenna, in the form of three monomeric minor proteins, CP29, CP26, and CP24, encoded respectively, by *Lhcb4*, *Lhcb5*, *and Lhcb6* genes, and a peripheral and potentially mobile major antenna, constituted of homologous or heterologous trimers of three LHCII protein types: Lhcb1, Lhcb2, or Lhcb3 (Jansson 1999), often referred to as the "major" LHCII polypeptides. In *Arabidopsis thaliana*, Lhcb1 is encoded by five genes (*Lhcb1.1* through *Lhcb1.5*), Lhcb2 by four genes (*Lhcb2.1* through *Lhcb2.4*), and Lhcb3 by a single gene (*Lhcb3*) (Jansson 1999). The inner PSII antenna proteins in this organism are encoded by three genes in the case of CP29 (*Lhcb4.1* through *Lhcb4.3*) and single genes in the case of CP26 (*Lhcb5*) and CP24 (*Lhcb6*). Both CP29 and CP24 are only distantly related to any of the other LHCII proteins, sharing higher similarity with LHCI proteins (Durnford et al. 2003). The structural organization of LHCI, which was found to be located on one side of the PSI core, with the other side likely serving for the docking of mobile LHCII trimers, has been resolved only recently (Ben Shem, Frolow, and Nelson 2003). Overall, the LHCI antenna complex is composed of four major proteins, assembled into two dimers, Lhca1/Lhca4 (referred to as LHCI types I and IV) and Lhca2/Lhca3 (referred to as LHCI types II and III; Green and Durnford 1996; Jansson

[1] These authors contributed equally to this work.

Key words: *Ostreococcus tauri*, Prasinophyceae, light-harvesting complex, evolution, phylogeny.

E-mail: partensky@sb-roscoff.fr.

Mol. Biol. Evol. 22(11):2217–2230. 2005
doi:10.1093/molbev/msi220
Advance Access publication July 27, 2005

1999; Ben Shem, Frolow, and Nelson 2003), responsible for the specific long-wavelength fluorescence properties of the PSI-LHCI assemblage (Croce et al. 2002). The two dimers are arranged in a series strongly anchored to the PSI core protein PsaG by the third MSH of Lhca1 (Ben Shem, Frolow, and Nelson 2003). Two other LHCI genes (*Lhca5* and *Lhca6*), showing a low expression level, have been found in *Arabidopsis* but the role and function of their products, possibly isoforms of some LHCI proteins, await further biochemical characterization (Jansson 1999).

In the case of green algae, LHCs have also been partially characterized in several species, such as *Scenedesmus obliquus*, *Dunaliella tertiolecta*, and *Dunaliella salina* and most comprehensively in *Chlamydomonas reinhardtii* (for review see e.g., Larkum and Howe 1997; Durnford et al. 2003; Elrad and Grossman 2004; Nield, Redding, and Hippler 2004). In *C. reinhardtii,* genome analysis (Elrad and Grossman 2004) has revealed nine genes (*Lhcbm1-6/8-9/11*) encoding major LHCII polypeptides and only two genes (*Lhcb4* and *Lhcb5*) encoding minor antenna polypeptides of PSII, CP26, and CP29, with the third, CP24, seemingly being absent. In terms of the PSI-associated antennae these authors identified nine LHCI genes (*Lhca1* through *Lhca9*). This composition has been largely confirmed by biochemical and proteomic approaches (Stauber et al. 2003; Takahashi et al. 2004; Tokutsu et al. 2004). None of the individual Lhcbm polypeptides of *C. reinhardtii* or other Chlorophyta can be clearly ascribed to any of the different major LHCII types (Lhcb1–3) conserved among higher plants (Teramoto, Ono, and Minagawa 2001; Durnford et al. 2003; Elrad and Grossman 2004). In contrast, some LHCI protein types (Lhca1 and Lhca3) do appear to be highly conserved between green algae and higher plants, while others (Lhca2 and Lhca4) are less so (see e.g., Tokutsu et al. 2004). This suggests that gene multiplication and divergence among the major LHCII types (and possibly also among the LHCI types Lhca2 and Lhca4) have occurred independently for the green algae and vascular plants, after the separation of these lineages (Teramoto, Ono, and Minagawa 2001; Elrad and Grossman 2004; Tokutsu et al. 2004). It is worth noting in this respect that, based on compositional aspects, the LHCI antenna of *C. reinhardtii* has recently been proposed to be larger than that of vascular plants because it has recently been proposed to comprise four dimers and three monomers of Lhca proteins (Kargul, Nield, and Barber 2003).

The Prasinophyceae constitute a particularly interesting algal class, holding a basal position in the evolution of the extant green lineage (Fawley, Yun, and Qin 2000; Guillou et al. 2004). Ecologically, prasinophytes belonging to the order Mamiellales are widespread and often abundant marine phytoplankton, playing an important role in oceanic biogeochemical cycling and primary production (Diez, Pedros Alio, and Massana 2001; Guillou et al. 2004; Not et al. 2004; Worden, Nolan, and Palenik 2004). Mamiellales, including *M. squamata*, *Micromonas pusilla*, *Bathycoccus prasinos*, and *Ostreococcus tauri*, display a complex pigment composition relative to other green algae, including carotenoids from the uriolide and prasinoxanthin series (Egeland, Guillard, and Liaaen-Jensen 1997) and a chl *c*–like pigment, Mg-2,4-divinyl-phaeoporphyrin a_5 monomethyl ester (Mg-DVP; Jeffrey and Vesk 1997). To date, only *M. squamata* has been studied in detail with respect to pigment-protein complexes (Fawley, Stewart, and Mattox 1986; Rhiel, Lange, and Morschel 1993; Schmitt et al. 1993; Jiao and Fawley 1994; Schmitt et al. 1994). These studies indicated the presence of a single, unique LHC type in *M. squamata* (hereafter called Lhcp, with the "p" referring to prasinophyte), associated to both PSI and PSII. These findings have stimulated much thought regarding LHC evolution and the primary endosymbiosis event (see e.g., Paulsen 1995; Durnford et al. 1999; Tokutsu et al. 2004).

Here, we use a combination of genomic, phylogenetic, and biochemical approaches to characterize the LHCs of *O. tauri*, a widespread unicellular marine prasinophyte which is also the smallest free-living eukaryote (ca. 1 μm diameter) known to date (Courties et al. 1994; Chrétiennot-Dinet et al. 1995; Courties et al. 1998). From an evolutionary perspective, based on its phylogenetic position (Guillou et al. 2004), genome compaction (12.64 Mb total genome size; Derelle et al. 2002; Derelle et al., personal communication), as well as its widespread environmental distribution, *Ostreococcus* is a fascinating case study organism. Our work yields new insight into the structure and function of prasinophyte LHCs and should help elucidate evolutionary processes that have led to such a diversified *Lhc* gene family within the green lineage.

Materials and Methods

Strains and Culture Conditions

Ostreococcus tauri strain OTH95 (aka RCC745), *M. squamata* strain CCMP480 (aka RCC417), *B. prasinos* strain CCMP1898 (aka RCC113), and *M. pusilla* strain CCMP490 (aka RCC114) were grown in K medium (Keller et al. 1987). All strains were grown in 10 liter polycarbonate carboys at 20°C under 10 μmol photons m^{-2} s^{-1} constant blue-green light, provided by fluorescent bulbs (Daylight, Sylvania, Danvers, Mass.) covered by blue-green filters (LEE Filters, Panavision, Woodland Hills, Calif.). Frozen spinach (*Spinacia oleracea*) was used as control.

Spectrometric Measurements

In vivo absorption spectra were recorded on cell suspensions using a double wavelength spectrophotometer (Aminco Chance DW2) equipped with an integrating sphere and 77 K fluorescence spectra on glass fiber filters (GF/F, Whatman) within a spectrofluorometer (Hitachi F4500) with 2.5/5 nm excitation/emission slit widths.

Pigment Analyses

Analytical and semipreparative pigment separations were performed using Waters Symmetry C_8 columns (150 × 4.6 mm, 3.5 μm particle size and 150 × 7.8 mm, 7 μm particle size, respectively) according to published procedures (Zapata, Rodriguez, and Garrido 2000), optimized as follows. The mobile phase A was methanol: acetonitrile:aqueous pyridine solution (0.1 M pyridine; 45:35:20 v/v/v) and the mobile phase B was acetonitrile: methanol:acetone (60:20:20 v/v/v). Solvent gradient (time,

% B): 0 min, 0%; 5 min, 4%; 18 min, 30%; 21 min, 30%; 26 min, 95%; 36 min, 95%; 38 min, 0%. Flow rate was 1 and 3 ml min^{-1} for analytical and semipreparative separations, respectively. Cultures and protein extracts were filtered onto 25-mm-diameter GF/F, immediately frozen, and stored at −20° C until further analysis. Frozen filters were thawed in 1 ml of 95% methanol and refiltered through 25-mm-diameter GF/F to remove cell and filter debris. A total of 200 µl of the pigment extracts were mixed with 40 µl Milli-Q water to avoid peak distortion (Zapata and Garrido 1991) and immediately injected into a Hewlett-Packard HPLC 1100 Series System equipped with a quaternary pump and diode array detector. All sample preparations were made under subdued light. Chlorophylls and carotenoids were detected by absorbance at 440 nm and identified by diode array spectroscopy (wavelength range: 350–750 nm; 1 nm spectral resolution). Pigments were identified and quantified using standard curves that employed known concentrations of both commercial standards (Sigma-Aldrich, St. Louis, Mo.) and standards derived from macroalgae and phytoplankton cultures by preparative HPLC (Repeta and Bjørnland 1997), applying previously compiled extinction coefficients (Jeffrey, Mantoura, and Wright 1997).

Thylakoid Extraction

Thylakoids were obtained using a modification of the protocol described by Schmitt and colleagues (Schmitt et al. 1993). Cells were harvested by centrifugation, washed, and resuspended in 10 ml N-Z-hydroxyethyl piperazine-N′-Z-ethanesulfonic acid (HEPES) buffer pH 7.5 (10 mM HEPES, 2 mM $MgCl_2$, 10 mM KCl, 1 mM ethylenediaminetetraacetic acid) containing a protease inhibitor cocktail composed of 1 mM benzamidin, 1 mM phenylmethylsulfonyl fluorure and 1 mM ε-amino-*n*-caproic acid (Sigma-Aldrich). This suspension was passed through a French pressure cell twice, at 20,000 psi and 4°C. All subsequent steps were carried out under subdued light at 4°C. Unbroken cells and debris were pelleted by centrifugation; the supernatant was removed and loaded onto 2 ml HEPES buffer containing 60% sucrose (m/v). After centrifugation at 130,000 × g for 1 h in a SW41Ti rotor (Beckman), the dark band floating at the sucrose interface was collected carefully and resuspended in 12 ml of HEPES buffer, supplemented with protease inhibitors. Subsequently, this suspension was centrifuged for 2 h at 150,000 × g, the supernatant was discarded, and the pellet resuspended in 2 ml HEPES buffer containing 15% glycerol and protease inhibitors.

Thylakoid Dissociation and LHC Purification

The thylakoid suspension was solubilized with *n*-dodecyl β-D-maltoside (DDM; Sigma-Aldrich) with a DDM:chl ratio of 60:1. Solubilization was made in dark for 2 h at 4°C with strong vortexing. After centrifugation at 6,000 × g for 5 min, the supernatant was loaded onto a 10 ml continuous sucrose gradient (5%–50%) in HEPES buffer containing 0.5% DDM (m/v) and protease inhibitors. The gradient was then ultracentrifuged for 15 h at 130,000 × g, the colored bands were collected, and their pigment content was analyzed by HPLC.

To further purify the LHC polypeptides, aliquots of sucrose gradient fractions were loaded onto a 7% acrylamide nondenaturing isoelectric focusing (IEF) gel with no stacking gel and containing ampholyte carriers pH 4–6.5 (Amersham Biosciences, Piscataway, N.J.) and 0.5% DDM. The electrophoresis system was a mini-Protean II where internal glass plates were replaced by alumina plates (Hoefer) in order to facilitate heat dissipation. The cathode tank was filled with 20 mM NaOH 0.5% DDM and the anode tank contained 20 mM orthophosphoric acid. The complexes were allowed to focus at 4°C in steps increasing by 50 V every 10 min up to 400 V, then 1 h at 500 V. The resulting bands were excised and crushed in HEPES buffer containing 0.5% DDM and a protease inhibitor cocktail (see above) using an electric grinder. Acrylamide remnants were pelleted by centrifugation at 20,000 × g for 15 min and the supernatant characterized at room temperature by fluorescence, absorption, and its pigment content.

Immunoreactions

Aliquots containing an equal chl concentration of either sucrose gradient fractions from *O. tauri* or thylakoids extracted from spinach and the four Mamiellales species (see above) were precipitated by 10% trichloroacetic acid (TCA) and 0.5 M β-mercaptoethanol, then resuspended in 4% sodium dodecyl sulfate (SDS) Laemmli denaturation buffer, and loaded onto a 15% acrylamide SDS–polyacrylamide gel electrophoresis (PAGE) minigel (Biorad, Hercules, Calif.). After ca. 1.5 h electrophoretic run, proteins were transferred overnight onto a nitrocellulose membrane in a Laemmli Tris-glycine buffer, pH 8.3, containing 10% methanol and 0.5% acetic acid, at 0.5 mA gel cm^{-2}. The membranes were then colored with Ponceau red, saturated during 1 h with 50% (m/v) milk tween-tris-boratesodium (TTBS) buffer (0.1% Tween, 20 mM NaCl, 500 mM Tris), and incubated in 0.5% milk TTBS containing rabbit primary antibodies directed against either *M. squamata* major antenna proteins (courtesy of E. Rhiel, Philipps-Universität, Marburg, Germany) or Lhca2 (AgriSera, Vännäs, Sweden) polypeptides. For characterization of sucrose gradient fractions, antibodies directed against PsaA and PsbA polypeptides, kindly provided by P. R Chitnis (Iowa State University, Ames, Iowa) and B. Green (University of British Columbia, Vancouver, Canada), were also used. After rinsing with 0.5% milk TTBS, a secondary anti-rabbit antibody coupled to a horseradish peroxidase (Biorad) was applied for 2 h. The chromogene reaction was allowed by the addition of chloronaphthol and hydrogen peroxide in TBS buffer containing 10% methanol.

Genome Sequencing

Purified genomic DNA fragments broken by sonication and ranging from 1 to 5 kb were inserted into pBluescript II KS (Stratagene, La Jolla, Calif.). Plasmidic DNA was extracted according to the TempliPhi method (Amersham Biosciences) and inserts were sequenced on both strands using universal forward and reverse M13 primers and the ET DYEnamic terminator kit (Amersham

Biosciences). Data were analyzed and contigs were assembled using Phred-Phrap and Consed software packages (http://depts.washington.edu/ventures/UW_Technology/Express_Licenses/Phred_Phrap.php). The whole *O. tauri* genome as well as the open reading frames database can be blasted at the following publicly available Web site: http://bioinformatics.psb.ugent.be/blast/public/?project=ostreococcus.

cDNA Libraries

Two cDNA libraries were also generated from different *O. tauri* cultures to improve the representation of expressed sequences. One cDNA library was created in the Lambda-zap vector (Stratagene) and the second in the gateway system (Invitrogen, Carlsbad, Calif.) according to the manufacturers' instructions. The average insert size was 1.5 Kbp. Sequences were obtained using forward and reverse primers, and single reads were assembled in contigs using the Phred-Phrap software. The expressed sequence tag (EST) database was subsequently screened using TBlastN for each identified Lhc proteins (see below).

Gene Identification

Lhc gene sequences from *O. tauri* were initially retrieved by systematically Blasting (TBlastN) the complete genome using Lhc protein sequences from a variety of other chl *a*/*b*–containing organisms. After completion of gene modeling and determination of intron/exon structure, the whole *O. tauri* data bank of derived amino acid sequences was further screened via BlastP searches to identify potentially missed Lhc sequences. cDNAs were used to confirm derived amino acid sequences as well as intron/exon structure. N-terminal regions of the amino acid sequences were analyzed for putative transit peptides (TP) using ChloroP, TargetP, and homology to known TP and mature polypeptides. All genes were hand curated and deposited in GenBank under accession numbers AY954729–AY954742 (see *Results*).

Protein Alignments and Phylogenetic Analysis

A 254-position amino acid alignment, representing predicted mature regions of the proteins, was generated using a combination of ClustalW (Thompson, Higgins, and Gibson 1994) on Biology Workbench, San Diego Supercomputer Center (http://workbench.sdsc.edu/) and manual editing using BioEdit (terminal regions excluded; Hall 1999). The alignment was composed of 74 polypeptide sequences, including all known major and minor LHCI and LHCII proteins from both *A. thaliana* and *C. reinhardtii*, as described by Jansson (1999) and Elrad and Grossman (2004), respectively. Also included were *C. reinhardtii* LI818r-1 (Savard, Richard, and Guertin 1996) and Lhcq (as in Elrad and Grossman 2004). Representative sequences from *Lycopersicon esculentum* (Cab9, CAA43590; Lhca1-3: P12360, P10708, P27522; CP24, P27524), *Volvox carterae* (Lhca, AAD55569; Lhca2, P10708), the red alga *Porphyridium cruentum* (Lhca1, U58679; Lhca2, U58680), and the chl *a*/*c*–containing algae *Isochrysis galbana* (Fcp1, X77333), *Laminaria saccharina* (Fcp1, U73510), *Macrocystis pyrifera* (FcpA, Q40297), *Odontella sinensis* (Fcp1, X81054), and *Phaeodactylum tricornutum* (FcpD, S42133; FcpA, Q08584) were also incorporated. For *O. tauri,* one (Lhcp2.4) of the 15 identified *Lhc* genes was excluded from the phylogenetic analysis due to potential sequencing/assembly problems, however, three other Lhcp2 genes were included. Members of the carotene biosynthesis–related (CBR) protein/early light-induced protein (ELIP)/light-harvesting–like protein (Lil) family (see *Results*) were also excluded. From this first alignment (see Supplementary Material online), 141 amino acid positions (falling largely within the membrane-spanning regions and immediately preceding regions) of the putative mature peptides from selected organisms were used for subsequent phylogenetic analysis, while nonhomologous positions and regions of ambiguous alignment were omitted.

Phylogenetic trees were inferred from the amino acid sequence alignment (described above) using parsimony, neighbor-joining (NJ)–distance and maximum likelihood approaches and the PHYLIP-3.62 package (http://evolution.genetics.washington.edu/phylip.html). For parsimony analysis (PROTPARS), sequence input order was randomized (10 jumbles). Distance matrices (PROTDIST) used the Dayhoff PAM matrix and randomized sequence input. Subsequent determination of tree topology used NJ (NEIGHBOR) and randomly added distances. Maximum likelihood (PROML) used the PAM probability model, global rearrangements, again with randomized sequence input order. For all analyses, support for trees was obtained by bootstrapping 100 data sets (generated in SEQBOOT) and consensus trees were generated using CONSENSE.

Results

In Vivo Measurements

The in vivo absorption spectrum of whole *O. tauri* cells was typical of chl *a*/*b*–containing organisms, with blue-green maxima at 440 and 470 nm (chl and carotenoids) and red maxima at 678 and 648 nm from chls *a* and *b*,

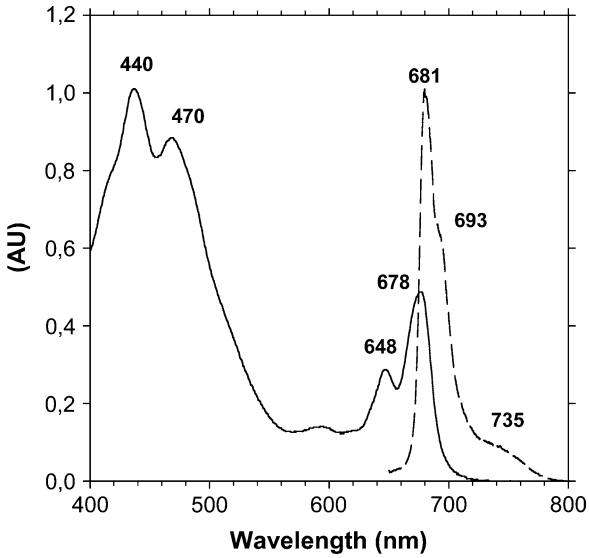

FIG. 1.—In vivo absorption and 77 K fluorescence emission spectra of *Ostreococcus tauri* cells. Numbers indicate the exact position of maxima in nanometers.

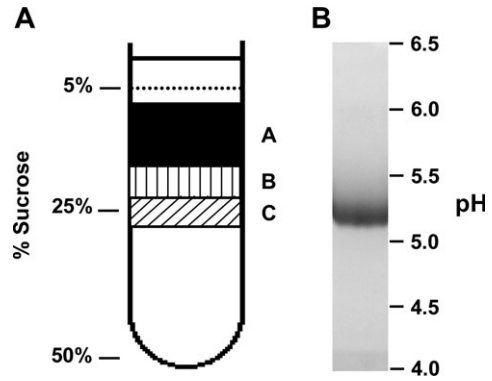

FIG. 2.—Isolation of pigment-protein complexes from dissociated thylakoids and purification of the major LHC. (*A*) Distribution of pigment-protein complexes in the sucrose density gradient (5%–50%) obtained after ultracentrifugation for 15 h at 130,000 × g and (*B*) IEF gel, showing the purified LHCs.

respectively (fig. 1). The chl *b*:chl *a* ratio was estimated as 0.60. No clear contribution from the chl *c*–like pigment Mg-DVP could be identified in this spectrum. The 77 K fluorescence emission spectrum of whole cells with excitation at 440 nm presented a major peak at 681 nm and only small shoulders at 693 and 735 nm.

Thylakoid Dissociation and LHC Characterization

Three major bands (fig. 2*A*) were obtained after dodecyl maltoside–mediated dissociation of the thylakoid membranes on a 5%–50% sucrose gradient. These bands were located in the upper section of the gradient, corresponding to 10%–30% sucrose. The major fraction, containing the bulk of LHCs (see below), was a dark brown band located at the uppermost region of the gradient (band A; fig. 2*A*). This band harbored approximately 60% of the total chlorophyll. Bands B and C exhibited light green and brown colors and represented ca. 21% and 9% of the total chlorophyll, respectively.

HPLC analysis of band A revealed that it contained the same array of pigments identified in the whole thylakoid fraction with a general enrichment of all accessory pigments but β,β-carotene (table 1). In contrast, band B exhibited a low chl *b*:chl *a* ratio (ca. 0.332 vs. ca. 0.868 in thylakoids) and a significant enrichment in β,β-carotene (its ratio to chl *a* was ca. 0.051 vs. ca. 0.019 in thylakoids), suggesting the presence of RCs in this fraction. B and C had an intermediate chl *b*:chl *a* ratio (ca. 0.858) compared to bands A and B and, more generally, had a pigment composition very similar to that of thylakoids, indicating that band C was composed of a mixture of LHCs and RCs.

Further purification of LHCs from sucrose gradient bands A and C by IEF led to the isolation of a broad brown band focusing at pH 5.2 (fig. 2*B*). The biochemical and biophysical properties of purified LHCs were identical for both samples. Compared to those in band A, purified complexes were impoverished in violaxanthin but enriched in chl *b* and Mg-DVP (table 1 and fig. 3*A*). The absorption spectrum of purified LHCs showed blueshifted maxima compared to the in vivo spectrum and a higher chl *b*:chl *a* ratio (0.69; fig. 3*B*). The fluorescence emission spectrum at room

Table 1
Pigment Ratios Relative to chl *a* of *Ostreococcus tauri* Thylakoids, Sucrose Gradient Fractions A Through C, and the IEF-Purified LHCs (see Fig. 2)

	Chl *b*	Mg-DVP	Pra	Neo	Uri	DLu	Unk car	Mic	Vio	β-car
Thylakoids	0.868 ± 0.141	0.078 ± 0.015	0.127 ± 0.017	0.068 ± 0.003	0.055 ± 0.010	0.052 ± 0.005	0.044 ± 0.002	0.041 ± 0.007	0.021 ± 0.004	0.019 ± 0.011
B and A	1.033 ± 0.303	0.086 ± 0.011	0.151 ± 0.024	0.082 ± 0.003	0.065 ± 0.012	0.067 ± 0.010	0.050 ± 0.011	0.051 ± 0.014	0.024 ± 0.021	0.007 ± 0.002
B and B	0.332 ± 0.054	0.022 ± 0.010	0.050 ± 0.014	0.021 ± 0.009	0.021 ± 0.007	0.015 ± 0.005	0.010 ± 0.002	0.012 ± 0.004	0.018 ± 0.004	0.051 ± 0.012
B and C	0.858 ± 0.074	0.079 ± 0.019	0.118 ± 0.009	0.071 ± 0.007	0.053 ± 0.009	0.055 ± 0.009	0.030 ± 0.010	0.043 ± 0.007	0.017 ± 0.009	0.019 ± 0.014
LHC	1.331 ± 0.146	0.131 ± 0.049	0.206 ± 0.081	0.087 ± 0.017	0.097 ± 0.021	0.071 ± 0.009	0.064 ± 0.009	0.056 ± 0.013	0.006 ± 0.004	0.002 ± 0.001

NOTE.—β-car, β,β-carotene; Dlu, dihydrolutein; Mg-DVP, Mg-divinyl-phaeoporphyrin a_5; Mic, micromonal; Neo, neoxanthin; Pra, prasinoxanthin; Unk car, unknown carotenoid; Uri, uriolide; Vio, violaxanthin.

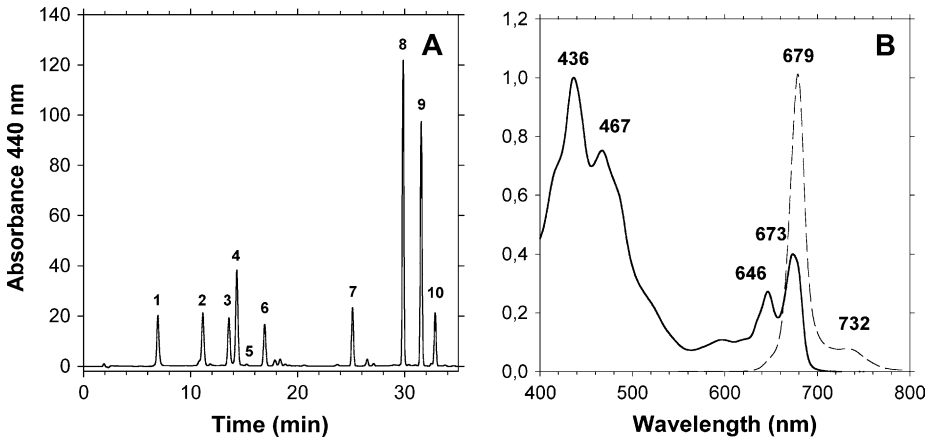

FIG. 3.—HPLC pigment and spectrometric analyses of IEF-purified LHCs from *Ostreococcus tauri*. (*A*) Chromatogram by order of elution: 1, Mg-DVP; 2, uriolide; 3, neoxanthin; 4, prasinoxanthin; 5, violaxanthin; 6, micromonal; 7, dihydrolutein; 8, chl *b*; 9, chl *a* and 10, unknown carotenoid. (*B*) Absorption (continuous line) and room temperature fluorescence emission (dashed line) spectra of purified LHCs. Numbers indicate the exact position of maxima in nanometers.

temperature exhibited a major peak at 679 nm and a shoulder at 732 nm.

SDS-PAGE and Immunoblots Analyses

SDS-PAGE analysis of sucrose gradient bands A, B, and C revealed distinct protein patterns (fig. 4, upper part). The polypeptidic profile from band A confirmed its enrichment in Lhc proteins, forming a major band at 21 kDa. B and B contained few antenna proteins, but both fractions B and C exhibited prominent bands at about 65 kDa, corresponding to the PSI core proteins, PsaA and PsaB. This band was the strongest in fraction C, indicating a specific RCI enrichment.

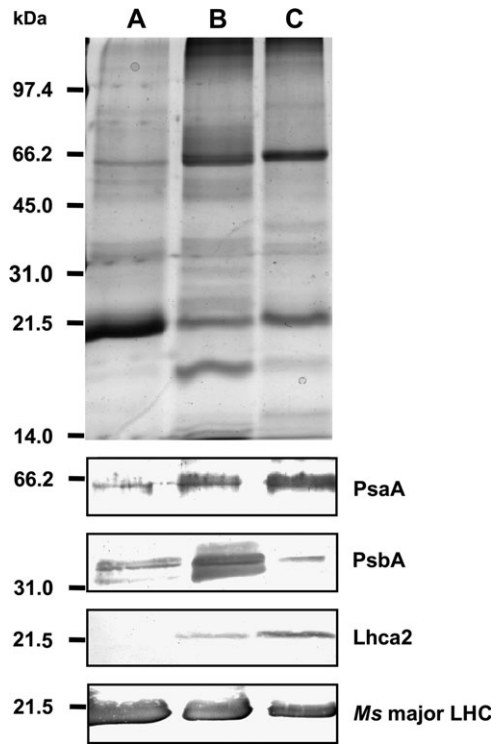

FIG. 4.—SDS-PAGE of sucrose gradient fractions and immunoblots with antibodies directed to PsaA, PsbA, Lhca2, and *Mantoniella squamata* (*Ms*) major LHC.

Immunoreactions with anti-PsaA and anti-PsbA antibodies provided some indication of the relative abundance of RCI and RCII, respectively, in the different fractions (fig. 4). Cross-reaction between these antibodies and proteins from band A was weak, supporting the finding that this band contains few PS core proteins. In contrast, strong cross-reactions were observed with anti-PsbA for band B (at ca. 35 kDa) and with anti-PsaA for band C (at 65 kDa), confirming their relative enrichment in RCII and RCI, respectively.

An anti-Lhca2–directed antibody, raised against a synthetic polypeptide sequence matching a part of the Lhca2 from various plants as well as from *C. reinhardtii*, was also tested. This Lhca2 antibody recognized a protein at approximately 22 kDa in bands B and C (fig. 4), whereas no cross-reaction was observed with band A. The cross-reaction was the strongest with the RCI-enriched fraction (band C), suggesting a specific association between the identified proteins and PSI.

Finally, a polyclonal antibody that was raised against a purified major antenna fraction from *M. squamata* (Rhiel, Lange, and Morschel 1993) cross-reacted at ca. 20 kDa in all *O. tauri* fractions. As expected, the strongest cross-reaction was to the main LHC fraction (band A), but this antibody also reacted with bands B and C, suggesting the presence of the prasinophyte-specific Lhc proteins of *O. tauri* in both RCII- and RCI-enriched fractions.

To assess general commonalities between antenna complexes of different members of the Mamiellales order, immunoreactions were conducted with whole thylakoid proteins from several species, using spinach as a negative control. Immunoreactions conducted with the antibody directed against *M. squamata* major antenna (functionally targeting Lhcp proteins) confirmed the presence of a protein of ca. 22 kDa in all the species tested, namely in *O. tauri*, *M. pusilla*, *B. prasinos*, and, expectedly, *M. squamata* (fig. 5). Surprisingly, the profile of *B. prasinos* thylakoids exhibited a second band at a relatively low apparent

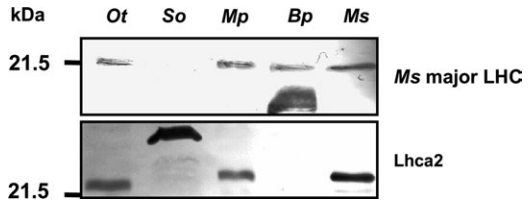

FIG. 5.—Immunoblots with antibodies directed to Lhca2 and *Mantoniella squamata* major LHC performed on thylakoid membranes from *Ostreococcus tauri* (*Ot*), *Spinacia oleracea* (*So*), *Micromonas pusilla* (*Mp*), *Bathycoccus prasinos* (*Bp*), and *M. squamata* (*Ms*).

molecular weight (MW of ca. 17 kDa) showing a strong cross-reaction with the anti-*M. squamata* Lhcp antibody. As expected, no cross-reaction of this antibody was observed with spinach proteins. A clear cross-reaction of the anti-Lhca2 antibody was observed with the Lhca2 proteins of both spinach, at ca. 24 kDa, and with polypeptides of significantly lower apparent MW (ca. 22 kDa) in all prasinophyte strains tested but *B. prasinos* (fig. 5). Although it was beyond the scope of this paper to characterize further the Lhc proteins of the latter species, these immunological characterizations strongly suggest that they have a number of unique characteristics with regard to other Mamiellales. It must be noted that although we show only partial immunoblots in figures 4 and 5, there were no or only minor contaminant bands in the rest of the blots.

Characterization of the Lhc Gene Family in *O. tauri*

Analysis of the completely sequenced *O. tauri* genome revealed a total of 19 genes encoding precursor proteins of the Lhc superfamily (table 2). These genes possessed very few exons (table 2), especially relative to *C. reinhardtii* (Elrad and Grossman 2004). Among these genes, five encoded proteins clearly related to the major polypeptides of PSI (Lhca), with most having best hits to plant sequences. Two genes encoded proteins related to the minor PSII antenna polypeptides CP29 (Lhcb4) and CP26 (Lhcb5). No homologs of genes coding for the major polypeptides of PSII (Lhcb1–3 types in vascular plants, Lhcbm1-6/8/9/11 in *C. reinhardtii*; Elrad and Grossman 2004) or of *Lhcb6* (encoding the PSII core antenna protein CP24 in plants), or *psbS*, (encoding a four-helix protein of the Lhc family involved in nonphotochemical quenching; Kim et al. 1992) could be identified in the *O. tauri* genome. However, an apparently prasinophyte-specific *Lhc* gene family, herein termed *Lhcp* (five genes), was present. The only other members of this family known so far were described in *M. squamata* and previously referred to as chlorophyll *a*/*b*/*c* LHCs (Rhiel and Morschel 1993). Four of the *Lhcp* genes found in *O. tauri* (*Lhcp2.1-4*) were almost identical (96%–100% nucleotide identity). Two other *Lhc* genes, related to genes previously found in *C. reinhardtii* but of still unknown function, *LI818* (Savard, Richard, and Guertin 1996) and *Lhcq* (Elrad and Grossman 2004), were each represented in single copy in *O. tauri*. Although not the focus of this study, the presence of five other distant members of the Lhc gene superfamily should be noted, which encode members of the CBR protein or ELIP family. Such genes are thought to be involved in photoprotection processes and transiently expressed during shifts from low to high light in *C. reinhardtii* (for a review, see e.g., Montane and Kloppstech 2000). Following Jansson's (1999) nomenclature, we called these genes *Lil1* through *Lil5* (for light-harvesting–like proteins; table 2).

Phylogenetic analyses allowed more complete assessment of the relatedness of chlorophyll *a*/*b*–binding proteins subpopulations found in *O. tauri* to those of *A. thaliana*, representing higher plants and *C. reinhardtii*, representing evolved green algae (fig. 6). For clarification purposes, it should be emphasized that (1) we have designated *C. reinhardtii* genes utilizing the nomenclature of Elrad and Grossman (2004), which differs from that of e.g., Tokutsu et al. (2004) and (2) the nomenclature used by Elrad and Grossman (2004) does not necessarily correspond with that adopted for higher plants (e.g., by Jansson 1999). As in previous studies, we found that CP26 (Lhcb5) and the major LHCII polypeptides of higher plants and green algae were monophyletic (fig. 6). Two *O. tauri* sequences fell at the base of this group, although bootstrap values were low for several nodes in this region. One of these two sequences shows similarity to Lhcq of *C. reinhardtii* (Elrad and Grossman 2004), although lacking the divergence seen in the N-terminal region of the putative mature *C. reinhardtii* polypeptide. In analyses where Lhcq was omitted (not shown), the other *O. tauri* protein sequence (Lhcb5) in this region forms a single, bootstrap-supported clade with CP26 (Lhcb5) of higher plants and *C. reinhardtii*. However, in the presented tree (including Lhcq), this branching pattern did not retain significant support (fig. 6). Analyses omitting Lhcq also yielded increased bootstrap support for the major LHCII protein group (86%–96% with NJ-distance methods). Bootstrap analysis strongly supported the exclusion of Lhcp, the prasinophyte-specific Lhc proteins, from the major LHCII polypeptides found in all other members of the green lineage studied thus far.

Consistent with previous studies (see e.g., Durnford et al. 1999), we found that CP29 (Lhcb4) and CP24 (Lhcb6) formed a supported group within LHCI proteins (fig. 6). This included *O. tauri* CP29 that branched with higher plant rather than *C. reinhardtii* CP29 sequences. LHCI proteins were largely divergent, making relationships difficult to resolve. However, significant support was found for a green algal–"LHCI type III" group that included both *C. reinhardtii* and *O. tauri* sequences, distinct from, but closely related to Lhca3 protein sequences of the higher plants *A. thaliana* and *L. esculentum*. A second likely LHCI type common to both green algae and higher plants appears to be "LHCI type I", a group gathering Lhca1 of *A. thaliana* and *L. esculentum* and *C. reinhardtii* Lhca4. In all our analyses, *O. tauri* Lhca1 formed a basal branch (unsupported) to this group. *O. tauri* Lhca5 fell within a green algal–only Lhca group, which branched close to *A. thaliana* Lhca5. Two *O. tauri* sequences, here termed Lhca2 and Lhca4, branched with *A. thaliana* Lhca4 and *C. reinhardtii* Lhca5, although without significant support. Again, this grouping was consistent through multiple tree analyses regardless of the phylogenetic approach employed. Lastly, in trees constructed with fucoxanthin-chlorophyll proteins (FCP) sequences (not shown), LI818 consistently branched

Table 2
Some Characteristics of *Ostreococcus tauri Lhc* Gene Family Members As Well As Putative *Lil* (Light-Harvesting–Like) Genes, Including Proposed Designation, Potential Pigment-Protein Complex Association, Exon Number, and GenBank Accession Number (or Gene Model Number in the Public *O. tauri* Database: http://bioinformatics.psb.ugent.be/blast/public/?project=ostreococcus) and BlastP best hit results (March 3, 2005)

Gene Designation	Potential Complex Association	Exons	GenBank Accession	Lowest e Value (ExPasy BlastP)
Lhca1	LHCI type I	2	AY954737	*Hordeum vulgare* Q9SDM1 chl *a*/*b*–binding protein (lhca1), 2×10^{-37}, 45% I, 55% +
Lhca2	LHCI	1	AY954734	*Pisum sativum* Q41038 type II lhaB, 4×10^{-59}, 52% I, 66% +
Lhca3	LHCI type III	2	AY954735	*Pinus sylvestris* Q02071 type III chl *a*/*b*–binding protein (LhcaC1), 2×10^{-73}, 63% I, 75% +
Lhca4	LHCI	2	AY954729	*Lolium temulentum* Q40252 type II LHCI (lhaB), 5×10^{-63}, 54% I, 66% +
Lhca5	LHCI	2	AY954736	*Volvox carteri* Q9SBM9, Light-harvesting complex a protein (lhca), 2×10^{-50}, 55% I, 69% +
Lhcq	Unknown	1	AY954740	*Oryza sativa* Q69TB3 putative chl *a*/*b*–binding protein, 2×10^{-53}, 48% I, 56%+
Lhcb4	CP29	4	AY954732	*Chlorarachnion* CCMP621 Q7XYQ5 chl a/b–binding protein CP29, 1×10^{-53}, 47% I, 60% +
Lhcb5	CP26	2	AY954738	*Chlamydomonas reinhardtii* Q9FEK6 CP26 (lhcb5), 3×10^{-67}, 54% I, 69% +
LI818	Unknown	1	AY954739	*Cyclotella cryptica* O65892 fucoxanthin cac (fcp08), 9×10^{-42}, 56% I, 66% +
Lhcp1	Lhcp	1[a]	AY954733	*Mantoniella squamata* Q40351 chl *a*/*b*/*c*–binding protein (Lhc1*4), 10^{-105}, 84% I, 90% +
Lhcp2.1	Lhcp	2	AY954731	*M. squamata* Q40351 chl *a*/*b*/*c*–binding protein (Lhc1*4), 10^{-121}, 90% I, 93% +
Lhcp2.2	Lhcp	2	AY954741	*M. squamata* Q40351 chl *a*/*b*/*c*–binding protein (Lhc1*4), 10^{-117}, 87% I, 90% +
Lhcp2.3	Lhcp	2	AY954742	*M. squamata* Q40351 chl *a*/*b*/*c*–binding protein (Lhc1*4), 10^{-120}, 90% I, 93% +
Lhcp2.4	Lhcp	1[a]	AY954730	*M. squamata* Q40351 chl *a*/*b*/*c*–binding protein (Lhc1*4) 10^{-115}, 93% I, 96% +
Lil1	Unknown	1	Ot13g00030	*Arabidopsis thaliana* P93735 early light-induced protein, 4×10^{-10}, 40% I, 54% +
Lil2	Unknown	1	Ot05g04640	*O. sativa* Q7XIT4 light-induced protein–like protein, 2×10^{-32}, 67% I, 82% +
Lil3	Unknown	1	Ot04g01810	*Dunaliella bardawil* P27516 carotene biosynthesis–related protein CBR, 3×10^{-14}, 38% I, 51% +
Lil4	Unknown	1	Ot03g01530	*C. reinhardtii* Q5W9T2 Lhc-like protein (Lhl3), 4×10^{-56}, 50% I, 64% +
Lil5	Unknown	1	Ot00g07320	*D. bardawil* P27516 carotene biosynthesis–related protein CBR, 8×10^{-12}, 40% I, 52% +

NOTE.—I, identity; +, similarity.
[a] Sequence incomplete at N-terminus (within putative TP).

with Fcp1 of the Haptophyte *I. galbana*, forming a group neighbored by other FCP sequences.

Deduced structural elements of *O. tauri* Lhc proteins were similar to plants and other green algae. This was true for the LHCI- and LHCII-related protein sequences both with respect to transmembrane domains and known chlorophyll ligands, as originally illustrated for LHCII of *Pisum sativum* and *S. oleracea* (Kühlbrandt, Wang, and Fujiyoshi 1994; Liu et al. 2004). In terms of the chl ligands, conservative substitutions (based on PAM and Blosum45 matrices) were observed in several *O. tauri* Lhc proteins. In almost all cases, identical substitutions are found in LHCI or LHCII of other organisms, for example, *A. thaliana* and *C. reinhardtii*. Lhca1 and Lhca3 were exceptions, with a conservative substitution for the chl *b6* ligand (asparagine for histidine) apparently not present in other organisms. In Lhcp proteins of both *O. tauri* and *M. squamata* two substitutions were found: (1) at the chl *a2* site (His for Asp), a conservative "substitution" also seen in minor LHCII polypeptides (e.g., CP29 and CP26 of *A. thaliana* and *L. esculentum*; CP29 of *C. reinhardtii*; and CP29 and Lhcq of *O. tauri*) and (2) at the chl *b6* site (glutamic acid for glutamine), a common substitution.

Screening of EST libraries provided preliminary assessment of potential utilization of the encoded proteins, addressing the question: are these genes expressed? All of the identified genes were expressed, with the exception of some *Lil* genes, for which no corresponding EST were found under the standard growth conditions employed. Further studies on high–light-stressed *O. tauri* cells are required to determine whether the latter genes are expressed.

Discussion

Specific Pigment Composition of Mamiellales Antenna Proteins

The pigment suite of *O. tauri* is unique to the order Mamiellales (Latasa et al. 2004), which possesses more primitive characters than any other group of Chlorophyta (Melkonian 2001) and constitutes a well-defined

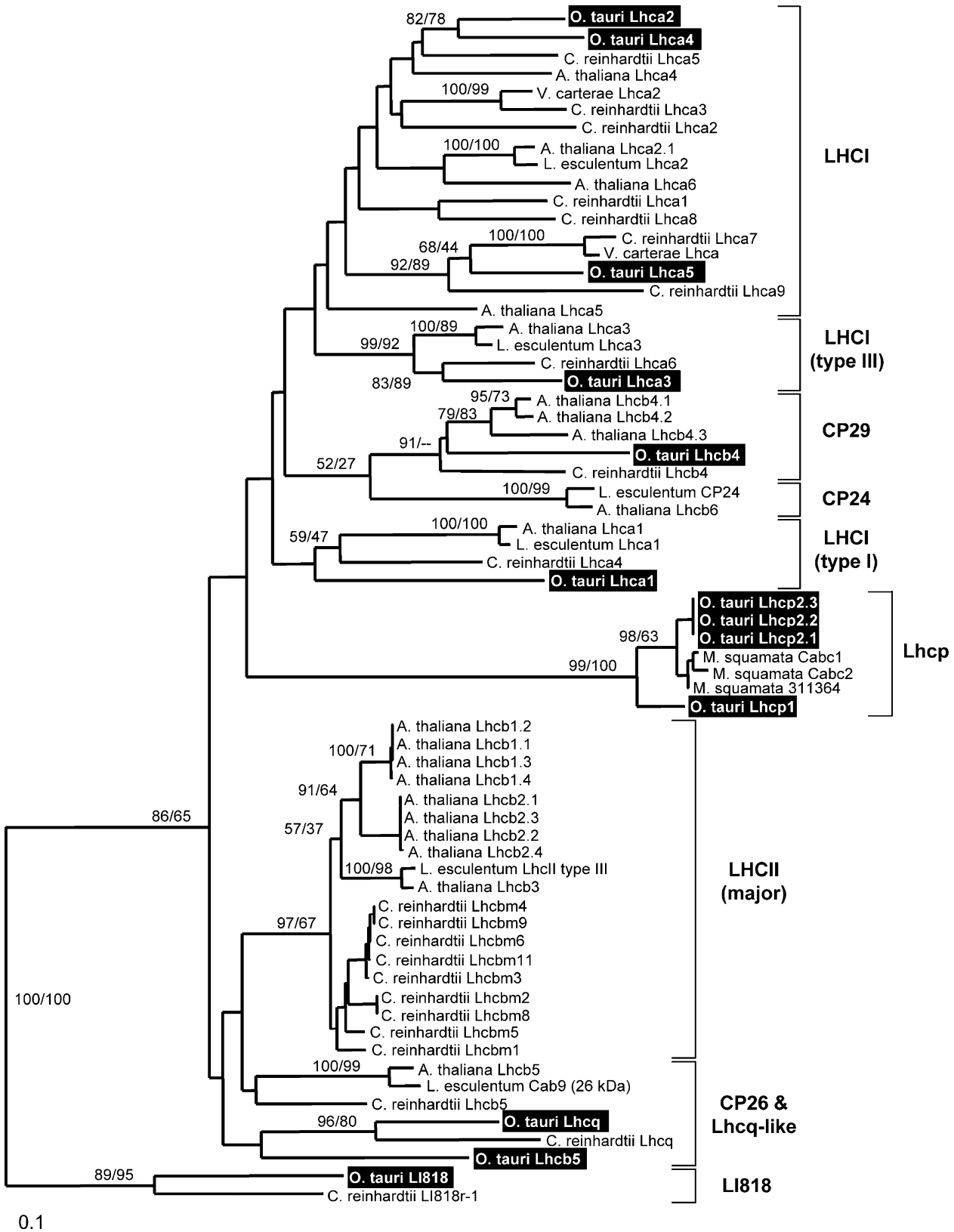

FIG. 6.—Phylogenetic analysis of Lhc protein sequences from *Ostreococcus tauri* (indicated with a black background), higher plants, and selected chlorophyll *a*/*b*–containing algae (green algae) inferred by NJ and maximum likelihood methods. The analysis includes both PSI (Lhca)- and PSII (Lhcb)-associated antennae complexes. The distance tree is shown with corresponding bootstrap values on the internal branches obtained from both maximum likelihood and NJ methods. Bootstrap values (percentage of 100 replicates) are shown in the order NJ-distance/ML. Sequence alignment details are given as Supplementary Material online.

monophylogenetic clade within the paraphyletic group of prasinophytes (clade II; Fawley, Yun, and Qin 2000; Guillou et al. 2004).

Analysis of the pigment composition of the *O. tauri* major LHC fraction (table 1 and fig. 3), which (as discussed below) is mainly constituted of prasinophyte-specific Lhcp protein-pigment complexes, shows that this fraction is significantly enriched in chl *b* and Mg-DVP. This strongly suggests that the latter chl *c*–like pigment, which was misidentified as a true chl *c* in early studies on *Micromonas* sp. (Wilhelm et al. 1986) and *M. squamata* (Herold et al. 1991; Schmitt et al. 1993), has a clear role in light harvesting. A similar function was also proposed for Mg-DVP found in the green oxyphotobacteria, *Prochloron* and *Prochlorococcus* (Larkum et al. 1994; Helfrich et al. 1999; Partensky and Garczarek 2003). It must be remembered that, in spite of being a chl *c*–like pigment, the occurrence of Mg-DVP in prasinophytes has no real significance in terms of a phylogenetic relatedness to chl *c*–containing organisms. This conclusion is largely confirmed by the fact that the Lhcp proteins of *O. tauri* and *M. squamata* are much more related to Lhc proteins from chl *a*/*b*–containing organisms than to chl *a*/*c* antenna proteins (Green and Durnford 1996; Durnford et al. 1999; A. Z. Worden, unpublished data).

The semipurified LHC fraction of *O. tauri* was also found to contain notable amounts of unusual carotenoids, namely micromonal, neoxanthin, prasinoxanthin, uriolide, and an unknown carotenoid (table 1). The role of such a variety of α- and β-carotene derivatives was assessed in the Mamiellales species *M. squamata* (Böhme, Wilhelm, and Goss 2002). According to this work, these pigments would be linked through light-dependent interconversions and are therefore involved in either light harvesting or photoprotection of the LHC supercomplexes. A few minor xanthophylls, including uriolide and micromonal, have been proposed to participate in the process of LHC trimerisation (Böhme, Wilhelm, and Goss 2002). The fact that violaxanthin, a carotenoid also involved in the xanthophyll cycle (Masojidek et al. 2004), was specifically lost during purification of fraction A (table 1) might indicate that it is more loosely bound to the LHCs than are the other carotenoids (Böhme, Wilhelm, and Goss 2002). Comparison of the β,β-carotene content of the different fractions (table 1) clearly indicates that this pigment is specifically associated with PS.

First Evidence of LHCI Complexes in a Prasinophyte

Based on biochemical, immunological, and genomic analyses of *O. tauri*, we demonstrate here for the first time that Lhca proteins are present in a species belonging to the prasinophytes. This strongly contrasts with previous literature on *M. squamata* LHCs (see e.g., Rhiel, Lange, and Morschel 1993; Schmitt et al. 1993), which reported that this species (and, by extension, prasinophytes in general) possessed only one type of antenna, corresponding to the group of Lhc proteins herein referred to as "Lhcp". It appears that methodological limitations are likely responsible for overlooking other Lhc subpopulations in *M. squamata*. For instance, Schmitt et al. (1993) based their conclusion that there was a sole type of antenna complex associated to both PSs in *M. squamata* on the observation that semipurified PSII and PSI antenna fractions had similar biochemical and biophysical properties. This erroneous interpretation might be explained by the fact that Lhcp-pigment complexes are most likely the predominant LHC type in Mamiellales. Indeed, in *O. tauri*, Lhcp proteins seemingly constituted the bulk (if not the totality) of antenna proteins in fraction A (see figs. 3 and 4), a fraction which alone gathered more than 60% of total chlorophyll. Furthermore, Lhcp proteins were retrieved at significant abundance in all fractions of the sucrose gradient, including the PSI-enriched fraction, as confirmed by immunological analyses with an anti-Lhcp antibody (fig. 4, bottom panel). Whether this implies that Lhcp-pigment complexes are bound to (and serve) both PSs would require more refined analyses. Nevertheless, Lhca-like proteins are also present in thylakoids and are found preferentially associated to the PSI-enriched fraction. This is clearly demonstrated here by immunological characterization of thylakoid proteins from this fraction with an anti-Lhca2 antibody. Based on genome-derived data, cross-reaction between this anti-Lhca2 synthetic peptide sequence (AADPDRPIWFPGS) and *O. tauri* Lhca proteins appears likely to only be with the *Lhca2* gene product (AEGAERPVWYPGK). Cross-reaction to proteins encoded by the other four *Lhca* genes of *O. tauri* is improbable due to lack of homology. We also tested an antibody to a synthetic peptide sequence from plant Lhca1 (also from AgriSera) but, likely for the same reason (no apparent homology), no cross-reaction with any Lhca proteins from *O. tauri* was obtained (data not shown). The anti-Lhca2 antibody cross-reacted with several other Mamiellales as well, including *M. squamata* (fig. 6). The absence of cross-reaction with *B. prasinos* thylakoid proteins is, however, notable and should foster further studies on the nature and diversity of Lhc proteins in this organism. Our data do not allow a clear evaluation of the relative importance of LHCI protein subpopulations in *O. tauri*. However, assuming that all LHCI proteins remained in the RCI-enriched fraction (like Lhca2), it can be reasonably assessed from the relatively weak intensity on SDS-PAGE of the ca. 22-kDa band in this fraction that LHCI proteins are much less abundant in thylakoids than are Lhcp proteins. This would explain in large part why LHCI proteins have been previously totally overlooked in *M. squamata*.

Although it is very likely that Lhcp polypeptides constitute the major PSII antenna in Mamiellales (and possibly prasinophytes at large), more studies are needed to understand which of the Lhcp-pigment complexes and the more classical LHCI (or a combination of those) constitute the major PSI antenna. The originality of the organization of LHCI complexes in *O. tauri* with regard to that in other green algae and higher plants (see e.g., Kargul, Nield, and Barber 2003) is emphasized by the weak PSI long-wavelength fluorescence emission at 77 K (fig. 1) compared to other Chlorophyta (Wollman and Bennoun 1982).

Phylogeny of the Lhc Protein Family

Complete analysis of *O. tauri* genome revealed a large diversity of genes belonging to the *Lhc* gene superfamily (table 2 and *Results*). The expression of these genes was

confirmed by ESTs. Although these results do not provide evidence of protein synthesis, they suggest that all of the identified Lhc proteins are synthesized and utilized in the photosynthetic antennae of *O. tauri*. As mentioned above, this was confirmed by immunological approaches for Lhca2 and Lhcp proteins.

For phylogenetic analyses of the whole chl *a*/*b*–binding Lhc protein family, we analyzed only homologous positions of unambiguously aligned sequences, corresponding to the membrane-spanning regions and immediately preceding regions of the putative mature peptides. Our results present similar tree topologies to other published reports using the same approach (e.g., Durnford et al. 1999; Durnford 2003). However, it differs to some extent from several other recent analyses (Jansson 1999; Elrad and Grossman 2004; Tokutsu et al. 2004) focusing on smaller groups of sequences (generally within species) and utilizing the entire gene sequence, including nonhomologous positions and transit peptides.

In the case of Lhcp proteins, determining the true associations with other known groups is complicated by the long-branch effect (see fig. 6). Still, these proteins are clearly outside the clade containing the major LHCII polypeptides of plants and green algae. As in previous studies, no relationship was resolved between *C. reinhardtii* Lhcbm polypeptides and the three major Lhcb subfamilies of higher plants (Elrad and Grossman 2004; Tokutsu et al. 2004). Instead, two well-supported clades were identified, one with all the Lhcbm's of green algae and the other with all of the higher plant major Lhcb polypeptides, highlighting independent evolution within these large LHCII protein families (Durnford 2003). Given that no *O. tauri* sequence is found in this cluster, the finding of two genes encoding putative proteins clustering with the CP26/Lhcq family appears as a potential key to understanding the origin of the LHCII superfamily. Indeed, these proteins apparently branch, together with CP26/Lhcq sequences from more evolved organisms, at the base of the LHCII clade. However, the hypothesis that the different LHCII families may be derived from an ancestral CP26-like gene should be confirmed by sequencing additional CP26/Lhcq-like genes in order to improve the bootstrap values and better resolve its branching point within the LHCII family.

Ostreococcus tauri possesses two genes clustering with previously identified CP29 (Lhcb4) and CP26 (Lhcb5) populations, indicating the potential presence of PSII core antennae. Both proteins have been shown to be critical for the stability of LHCII-PSII supercomplexes in plants (Boekema et al. 1999; Yakushevska et al. 2003), and one can assume that they have a similar role in maintaining the integrity of Lhcp-PSII complexes in *O. tauri*. Still, *Ostreococcus* shares with *C. reinhardtii* (and possibly all other green algae) the absence of CP24 (Lhcb6), a protein which is located close to CP29 in *A. thaliana* PSII internal antenna (Yakushevska et al. 2003) and is thought to play a key role in the xanthophyll cycle (Pagano, Cinque, and Bassi 1998).

Five clearly distinct *Lhca* genes were retrieved in the *O. tauri* genome, again suggesting utilization of a differentiated LHCI antenna. As in *C. reinhardtii* (Tokutsu et al. 2004), *O. tauri* has single copies of genes putatively corresponding to LHCI types I and III of higher plants. However, *O. tauri* differs dramatically from *C. reinhardtii* in terms of other LHCI populations, having just three other Lhca genes, loosely related to higher plant LHCI types II and/or IV, while *C. reinhardtii* has a multiplicity of such genes (seven). Whether this implies that *O. tauri* LHCI is structurally more closely related to that of higher plants (e.g., Ben Shem, Frolow, and Nelson 2003) than of *C. reinhardtii* (e.g, Kargul, Nield, and Barber 2003) would require further structural analyses. Unlike the *Lhca* genes of *O. tauri*, those of *C. reinhardtii* are also particularly exon rich (an average of one and seven exons, respectively). In this regard too, *O. tauri* seems more similar to higher plants. Lastly, in analyses where *Lhca* genes from the red alga *P. cruentum* were included (data not shown), all green algal (including *O. tauri*) and plant *Lhc* genes except *LI818* (grouping with FCP genes) were more closely related to one another than to those of the red alga (Tokutsu et al. 2004; this study).

Structural Characteristics of Lhcp Proteins

In addition to the overall phylogenetic placement of the LHCI and LHCII subpopulations, deduced structural elements appear to be similar to those in other members of the green lineage. For example, all substitutions for the putative chlorophyll ligands in these subpopulations were conservative, likely able to maintain the side-arm function of individual ligands. Of note is the conservative His substitution for Asp at chl *a*2 in all Lhcp proteins. Although this "substitution" is not unique because it is also present in minor LHCII polypeptides of higher plants, *O. tauri* and *C. reinhardtii*, it is interesting from the view point that, for known bacteriochlorophyll proteins, all chl ligands are His. This situation has changed dramatically in the evolution of eukaryotic Lhc proteins (in which substitutions for His are found at the majority of ligand sites). Thus, Lhcp proteins may reflect a more ancestral state. Overall, however, other ligands in the Lhcp proteins were highly conserved with only one other deviation from LHCII ligands identified by Kühlbrandt, Wang, and Fujiyoshi (1994) and Liu et al. (2004), an extremely common and conservative substitution at chl *b*6. The overall high degree of conservation emphasizes the strong potential for the newly identified genes to encode proteins with similar function to known classical LHCs.

Conclusions

We demonstrate here that *O. tauri* contains differentiated subpopulations of Lhc polypeptide-encoding genes, although its prasinophyte-specific Lhcp-pigment complexes probably constitute the large majority of LHCs in thylakoid membranes. Furthermore, we show that, in contrast to previous literature, several Mamiellales, including *M. squamata*, do contain classical Lhca protein(s) likely associated with PSI. Whether these conclusions can be extended to all prasinophytes requires further systematic screening for LHC diversity in various other members of this paraphyletic class. For instance, the apparent absence of Lhca2-like proteins in *B. prasinos*, though a member of the Mamiellales order, shows that generalization might be inappropriate.

A considerable research effort is still needed to better elucidate the structure and function of *O. tauri* (and other prasinophyte) LHCs, especially with regard to the true nature and structure of its PSI antennae. A salient feature of the *O. tauri* system is the fact that *Lhc* genes are single copy and appear to roughly correspond with higher plant groupings, with the exception of *Lhcp* with are present in five copies (including four almost identical), a state possibly reflecting the high demand of the cell for these particular proteins. Finally, the complete genome sequence of *O. tauri* is close to publication. This provides a valuable foundation for comprehensive, integrative approaches to Lhc polypeptides, their relationship with other components of the photosynthetic apparatus, their molecular evolution, and potential role in the adaptation of these algae to their marine habitat.

Supplementary Material

Supplementary Material is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank D. G. Durnford for kindly providing Lhc protein alignments (as in Durnford et al. 1999). We acknowledge R. Cooke, M. Delseni, J. Demaille, E. Derelle, S. Eychénié, and C. Ferraz from the Génopole Languedoc-Roussillon for giving us access to yet unpublished genome data. We are also indebted to F. Y. Bouget, B. Khadaroo, and C. Schwartz for cDNA libraries preparation, as well as to S. Rombauts and Y. Van de Peer for providing us access to the *O. tauri* annotation site. This work was supported by funding from the Région Bretagne (PHOSTREO) to F.P. and A.Z.W. and the Gordon and Betty Moore Foundation and National Science Foundation (MCB-0429359) to A.Z.W. and was conducted within the framework of the "Marine Genomics Europe" Network of Excellence (2004–2006). C.S. was supported by a doctoral fellowship from the Ministère de la Recherche et des Nouvelles Technologies (France) and F.R. by a postdoctoral fellowship from Fundación Caixanova (Spain).

Literature Cited

Ben Shem, A., F. Frolow, and N. Nelson. 2003. Crystal structure of plant photosystem I. Nature **426**:630–635.

Boekema, E. J., H. Van Roon, J. F. Van Breemen, and J. P. Dekker. 1999. Supramolecular organization of photosystem II and its light-harvesting antenna in partially solubilized photosystem II membranes. Eur. J. Biochem. **266**:444–452.

Böhme, K., C. Wilhelm, and R. Goss. 2002. Light regulation of carotenoid biosynthesis in the prasinophycean alga *Mantoniella squamata*. Photochem. Photobiol. Sci. **1**:619–628.

Chrétiennot-Dinet, M. J., C. Courties, A. Vaquer, J. Neveux, H. Claustre, J. Lautier, and M. C. Machado. 1995. A new marine picoeucaryote: *Ostreococcus tauri* gen. *et* sp. nov. (Chlorophyta, Prasinophyceae). Phycologia **34**:285–292.

Courties, C., R. Perasso, M. J. Chretiennot-Dinet, M. Gouy, L. Guillou, and M. Troussellier. 1998. Phylogenetic analysis and genome size of *Ostreococcus tauri* (Chlorophyta, Prasinophyceae). J. Phycol. **34**:844–849.

Courties, C., A. Vaquer, M. Trousselier, J. Lautier, M.-J. Chrétiennot-Dinet, J. Neveux, C. Machado, and H. Claustre. 1994. Smallest eukaryotic organism. Nature **370**:255.

Croce, R., T. Morosinotto, S. Castelletti, J. Breton, and R. Bassi. 2002. The Lhca antenna complexes of higher plants photosystem I. Biochim. Biophys. Acta **1556**:29–40.

Derelle, E., C. Ferraz, P. Lagoda et al. (12 co-authors). 2002. DNA libraries for sequencing the genome of *Ostreococcus tauri* (Chlorophyta, Prasinophyceae): the smallest free-living eukaryotic cell. J. Phycol. **38**:1150–1156.

Diez, B., C. Pedros-Alio, and R. Massana. 2001. Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. Appl. Environ. Microbiol. **67**:2932–2941.

Durnford, D. G. 2003. Structure and regulation of algal light-harvesting complex genes. Pp. 63–82 *in* A. W. D. Larkum, S. E Douglas, and J. A. Raven, eds. Photosynthesis in algae, Vol. 14. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Durnford, D. G., J. A. Deane, S. Tan, G. I. McFadden, E. Gantt, and B. R. Green. 1999. A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastids evolution. J. Mol. Evol. **48**:59–68.

Durnford, D. G., J. A. Price, S. M. McKim, and M. L. Sarchfield. 2003. Light-harvesting complex gene expression is controlled by both transcriptional and post-transcriptional mechanisms during photoacclimation in *Chlamydomonas reinhardtii*. Physiol. Plant. **118**:193–205.

Egeland, E. S., R. R. L. Guillard, and S. Liaaen-Jensen. 1997. Algal carotenoids. 63. Carotenoids from Prasinophyceae. 7. Additional carotenoid prototype representatives and a general chemosystematic evaluation of carotenoids in Prasinophyceae (Chlorophyta). Phytochemistry **44**:1087–1097.

Elrad, D., and A. R. Grossman. 2004. A genome's-eye view of the light-harvesting polypeptides of *Chlamydomonas reinhardtii*. Curr. Genet. **45**:61–75.

Fawley, M. W., K. D. Stewart, and K. R. Mattox. 1986. The novel light-harvesting pigment-protein complex of *Mantoniella squamata* (Chlorophyta): phylogenetic implications. J. Mol. Biol. **23**:168–176.

Fawley, M. W., Y. Yun, and M. Qin. 2000. Phylogenetic analyses of 18S rDNA sequences reveal a new coccoid lineage of the Prasinophyceae (Chlorophyta). J. Phycol. **36**:387–393.

Green, B. R., and D. G. Durnford. 1996. The chlorophyll-carotenoid proteins of oxygenic photosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47**:685–714.

Green, B. R., and W. Kühlbrandt. 1995. Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. Photosynth. Res. **44**:139–148.

Green, B. R., E. Pichersky, and K. Kloppstech. 1991. Chlorophyll *a*/*b*-binding proteins: an extended family. Trends Biochem. Sci. **16**:181–186.

Guillou, L., W. Eikrem, M. J. Chretiennot-Dinet, F. Le Gall, R. Massana, K. Romari, C. Pedros-Alio, and D. Vaulot. 2004. Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. Protist **155**:193–214.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41**:95–98.

Helfrich, M., A. Ross, G. C. King, A. G. Turner, and A. W. Larkum. 1999. Identification of [8-vinyl]-protochlorophyllide *a* in phototrophic prokaryotes and algae: chemical and spectroscopic properties. Biochim. Biophys. Acta **1410**: 262–272.

Herold, A., A. Schmitt, C. Wilhelm, and A. Wild. 1991. Isolation and purification of an intact light-harvesting protein from the green alga *Mantoniella squamata* (Prasinophyceae). Photosynthetica **25**:645–653.

Jansson, S. 1999. A guide to the *Lhc* genes and their relatives in *Arabidopsis*. Trends Plant Sci. **4**:236–240.

Jeffrey, S. W., R. F. C. Mantoura, and S. W. Wright. 1997. Phytoplankton pigments in oceanography: guidelines to modern methods. Unesco Publishing, Paris.

Jeffrey, S. W., and M. Vesk. 1997. Introduction to marine phytoplankton and their pigment signatures. Pp. 37–84 *in* S. W. Jeffrey, R. F. C. Mantoura, and S. W. Wright, eds. Phytoplankton pigments in oceanography: guidelines to modern methods. Unesco Publishing, Paris.

Jiao, S., and M. W. Fawley. 1994. A cDNA clone encoding a light-harvesting protein from *Mantoniella squamata*. Plant Physiol. **104**:797–798.

Kargul, J., J. Nield, and J. Barber. 2003. Three-dimensional reconstruction of a light-harvesting complex I-photosystem I (LHCI-PSI) supercomplex from the green alga *Chlamydomonas reinhardtii* —insights into light harvesting for PSI. J. Biol. Chem. **278**:16135–16141.

Keller, M. D., R. C. Selvin, W. Claus, and R. R. L. Guillard. 1987. Media for the culture of oceanic ultraphytoplankton. J. Phycol. **23**:633–638.

Kim, S., P. Sandusky, N. R. Bowlby, R. Aebersold, B. R. Green, S. Vlahakis, C. F. Yocum, and E. Pichersky. 1992. Characterization of a spinach *psbS* cDNA encoding the 22 kDa protein of photosystem II. FEBS Lett. **314**:67–71.

Kühlbrandt, W., D. N. Wang, and Y. Fujiyoshi. 1994. Atomic model of plant light-harvesting complex by electron crystallography. Nature **367**:614–621.

Larkum, A. W. D., and C. J. Howe. 1997. Molecular aspects of light-harvesting processes in algae. Adv. Bot. Res. **27**:257–330.

Larkum, A. W. D., C. Scaramuzzi, G. C. Cox, R. G. Hiller, and A.G. Turner. 1994. Light-harvesting chlorophyll *c*-like pigment in *Prochloron*. Proc. Natl Acad. Sci. USA **91**: 679–683.

Latasa, M., R. Scharek, F. Le Gall, and L. Guillou. 2004. Pigment suites and taxonomic groups in Prasinophyceae. J. Phycol. **40**:1149–1155.

Liu, Z. F., H. C. Yan, K. B. Wang, T. Y. Kuang, J. P. Zhang, L. L. Gui, X. M. An, and W. R. Chang. 2004. Crystal structure of spinach major light-harvesting complex at 2.72 angstrom resolution. Nature **428**:287–292.

Masojidek, J., J. Kopecky, M. Koblizek, and G. Torzillo. 2004. The xanthophyll cycle in green algae (Chlorophyta): its role in the photosynthetic apparatus. Plant Biol. **6**:342–349.

Melkonian, M. 2001. Systematics and evolution of the algae. I. Genomics meets phylogeny. Prog. Bot. **62**:340–382.

Montane, M. H., and K. Kloppstech. 2000. The family of light-harvesting-related proteins (LHCs, ELIPs, HLIPs): was the harvesting of light their primary function? Gene **258**:1–8.

Nield, J., K. Redding, and M. Hippler. 2004. Remodeling of light-harvesting protein complexes in *Chlamydomonas* in response to environmental changes. Eukaryot. Cell **3**:1370–1380.

Not, F., M. Latasa, D. Marie, T. Cariou, D. Vaulot, and N. Simon. 2004. A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the western English channel. Appl. Environ. Microbiol. **70**:4064–4072.

Pagano, A., G. Cinque, and R. Bassi. 1998. In vitro reconstitution of the recombinant photosystem II light-harvesting complex CP24 and its spectroscopic characterization. J. Biol. Chem. **273**:17154–17165.

Partensky, F., and L. Garczarek. 2003. The photosynthetic apparatus of chlorophyll *b*- and *d*-containing Oxychlorobacteria. Pp. 29–62 *in* A. W. D. Larkum, S. E Douglas, and J. A. Raven, eds. Photosynthesis in algae, Vol. 14. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Paulsen, H. 1995. Chlorophyll a/*b*-binding proteins. Photochem. Photobiol. **62**:367–382.

Repeta, D. J., and T. Bjørnland. 1997. Preparation of carotenoid standards. Pp. 239–260 *in* S. W. Jeffrey, R. F. C. Mantoura, and S. W. Wright, eds. Phytoplankton pigments in oceanography: guidelines to modern methods. Unesco Publishing, Paris.

Rhiel, E., W. Lange, and E. Morschel. 1993. The unusual light-harvesting complex of *Mantoniella squamata*: supramolecular composition and assembly. Biochim. Biophys. Acta **1143**: 163–172.

Rhiel, E., and E. Morschel. 1993. The atypical chlorophyll *a*/*b*/*c* light-harvesting complex of *Mantoniella squamata*: molecular cloning and sequence analysis. Mol. Gen. Genet. **240**: 403–413.

Savard, F., C. Richard, and M. Guertin. 1996. The *Chlamydomonas reinhardtii* LI818 gene represents a distant relative of the *cabI*/*II* genes that is regulated during the cell cycle and in response to illumination. Plant Mol. Biol. **32**: 461–473.

Schmitt, A., G. Frank, P. James, W. Staudenmann, H. Zuber, and C. Wilhelm. 1994. Polypeptide sequence of the chlorophyll *a*/*b*/*c*-binding protein *of* the Prasinophycean *Mantoniella squamata*. Photosynth. Res. **40**:269–277.

Schmitt, A., A. Herold, C. Welte, A. Wild, and C. Wilhelm. 1993. The light-harvesting system of the unicellular alga *Mantoniella squamata* (Prasinophyceae): evidence for the lack of a photosystem I-specific antenna complex. Photochem. Photobiol. **57**:132–138.

Stauber, E. J., A. Fink, C. Markert, O. Kruse, U. Johanningmeier, and M. Hippler. 2003. Proteomics of *Chlamydomonas reinhardtii* light-harvesting proteins. Eukaryot. Cell **2**:978–994.

Takahashi, Y., T. Yasui, E. J. Stauber, and M. Hippler. 2004. Comparison of the subunit compositions of the PSI-LHCI supercomplex and the LHCI in the green alga *Chlamydomonas reinhardtii*. Biochemistry **43**:7816–7823.

Teramoto, H., T. Ono, and J. Minagawa. 2001. Identification of Lhcb gene family encoding the light-harvesting chlorophyll-*a*/*b* proteins of photosystem II in *Chlamydomonas reinhardtii*. Plant Cell Physiol. **42**:849–856.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

Tokutsu, R., H. Teramoto, Y. Takahashi, T. A. Ono, and J. Minagawa. 2004. The light-harvesting complex of photosystem I in *Chlamydomonas reinhardtii*: protein composition, gene structures and phylogenic implications. Plant Cell Physiol. **45**:138–145.

Wilhelm, C., I. Lenartz-Weiler, I. Wiedemann, and A. Wild. 1986. The light-harvesting system of *Micromonas* spec., Butcher (Prasinophyceae): the combination of three different chlorophyll species in one single chlorophyll-protein complex. Phycologia **25**:304–312.

Wollman, F., and P. Bennoun. 1982. A new chlorophyll-protein complex related to photosystem I in *Chlamydomonas reinhardtii*. Biochim. Biophys. Acta **680**:352–360.

Worden, A. Z., J. K. Nolan, and B. Palenik. 2004. Assessing the dynamics and ecology of marine picophytoplankton: the

importance of the eukaryotic component. Limnol. Oceanogr. **49**:168–179.

Yakushevska, A. E., W. Keegstra, E. J. Boekema, J. P. Dekker, J. Andersson, S. Jansson, A. V. Ruban, and P. Horton. 2003. The structure of photosystem II in *Arabidopsis*: localization of the CP26 and CP29 antenna complexes. Biochemistry **42**:608–613.

Zapata, M., and J. L. Garrido. 1991. Influence of injection conditions in reversed phase high-performance liquid chromatography of chlorophylls and carotenoids. Chromatographia **31**:589–594.

Zapata, M., F. Rodriguez, and J. L. Garrido. 2000. Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using a reversed phase C-8 column and pyridine-containing mobile phases. Mar. Ecol. Prog. Ser. **195**:29–45.

Charles Delwiche, Associate Editor

Accepted July 7, 2005