J. Phycol. **46,** 602–615 (2010)

DOI: 10.1111/j.1529-8817.2010.00841x

ANALYSES OF THE COMPLETE CHLOROPLAST GENOME SEQUENCES OF TWO MEMBERS OF THE PELAGOPHYCEAE: *AUREOCOCCUS ANOPHAGEFFERENS* CCMP1984 AND *AUREOUMBRA LAGUNENSIS* CCMP1507[1]

Han Chuan Ong[2]

Department of Biology, University of Washington, Box 355325, Seattle, Washington 98195, USA
School of Oceanography, University of Washington, Box 357940, Seattle, Washington 98195, USA

Steven W. Wilhelm

Department of Microbiology, The University of Tennessee, Knoxville, Tennessee 37996, USA

Christopher J. Gobler

School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York 11794-5000, USA

George Bullerjahn

Department of Biological Sciences, Bowling Green State University, Bowling Green, Ohio 43403-0208, USA

Michael A. Jacobs

Department of Medicine, University of Washington, Box 352145, Seattle, Washington 98195, USA

John McKay

School of Oceanography, University of Washington, Box 357940, Seattle, Washington 98195, USA

Elizabeth H. Sims, Will G. Gillett, Yang Zhou, Eric Haugen

Department of Medicine, University of Washington, Box 352145, Seattle, Washington 98195, USA

Gabrielle Rocap

School of Oceanography, University of Washington, Box 357940, Seattle, Washington 98195, USA

and Rose Ann Cattolico[3]

Department of Biology, University of Washington, Box 355325, Seattle, Washington 98195, USA
School of Oceanography, University of Washington, Box 357940, Seattle, Washington 98195, USA

Heterokont members of the Pelagophyceae form the massive brown tides that have continually plagued the coastal regions of the eastern U.S. seaboard and the Gulf of Mexico. To gain a better understanding of the photosynthetic competence that may be linked to their success in forming massive blooms, we sequenced the chloroplast genomes of two pelagophytes: *Aureococcus anophagefferens* Hargraves et Sieburth and *Aureoumbra lagunensis* D. A. Stockw., DeYoe, Hargraves et P. W. Johnson. The chloroplast genomes of *A. anophagefferens* (89,599 bp) and *Ar. lagunensis* (94,346 bp) are significantly smaller than those of six other stramenopiles sequenced to date. The structure (or configuration) is partially due to the absence of the large inverted repeats common in chloroplast genomes. Eight of 10 small and tandem repeats from the *A. anophagefferens* and *Ar. lagunensis* genomes are adjacent to genes coding for photosynthetic or energy production functions, implying that these domains may have functional constraints. High genomic synteny, a multigene phylogenetic analysis, and a synapomorphic change in the form of an attenuated *psb*A gene confirm that *A. anophagefferens* and *Ar. lagunensis* are closely related taxa. Finally, the presence of three light-independent chl-biosynthesis genes in the chloroplast of *Ar. lagunensis*, but absence in the chloroplast and nuclear genomes of *A. anophagefferens*, suggests the persistence of a more ancient (i.e., dark-adaptive) potential in *Ar. lagunensis* but not in *A. anophagefferens*. Whether the presence of both

[1]Received 2 July 2009. Accepted 7 December 2009.
[2]Present address: Division of Science, Lyon College, 2300 Highland Road, Batesville, Arkansas 72501, USA.
[3]Author for correspondence: e-mail racat@u.washington.edu.

chl-biosynthesis pathways in *Ar. lagunensis* contributes to the ability of this organism to sustain prolonged bloom (continuously for ~8 years) under reduced light conditions, but not *A. anophagefferens* (a few months), remains an open question.

Key index words: **chlorophyll; chloroplast; comparative genomics; evolution; genome; harmful algae; protochlorophyllide; stramenopile**

Abbreviations: **cpDNA, chloroplast DNA; LIPOR, light-independent protochlorophyllide reductase; POR, light-dependent protochlorophyllide oxidoreductase; tmRNA, transfer-messenger RNA; UWGC, University of Washington Genome Center; WGS, whole-genome shotgun**

Brown tides are caused by the pelagophytes *A. anophagefferens* and *Ar. lagunensis*. Destructive blooms of *A. anophagefferens* have been reported in northeastern and mid-Atlantic U.S. estuaries for two decades (Gobler et al. 2005) and, more recently, in South Africa (Probyn et al. 2001). *Ar. lagunensis* has been responsible for similar blooms in coastal bays in and around the Laguna Madre, Texas, for the past 15 years (Buskey et al. 1997, Villareal et al. 2004). Both species contain the unique pigment 19′-butanoyloxyfucoxanthin and are small (4–5 μm for *Ar. lagunensis* and 2–3 μm for *A. anophagefferens*), spherical, nonmotile cells with simple life cycles (DeYoe et al. 1997). These algae were originally assigned to the class Chrysophyceae (Sieburth et al. 1988, Buskey and Stockwell 1993), but later examination of their pigments, physiology, 18S rRNA sequences, and morphology led to their formal placement in the class Pelagophyceae (DeYoe et al. 1997). Although *A. anophagefferens* and *Ar. lagunensis* are sufficiently different genetically to be placed in separate genera, the genetic diversity within each species appears to be small (Bailey and Andersen 1999).

The first *A. anophagefferens* blooms occurred simultaneously in the summer of 1985 in several estuaries in the northeastern U.S., including Narragansett Bay, Rhode Island; Great South Bay and the Peconic Estuary on Long Island, New York; and putatively in Barnegat Bay, New Jersey (Sieburth et al. 1988, Cosper et al. 1989, Olsen 1989). During the past decade, blooms have expanded south along the U.S. East Coast into bays in New Jersey (Barnegat Bay; Gastrich et al. 2004), Delaware (Little Assawoman Bay; Popels et al. 2003), Maryland, and Virginia (Chincoteague Bay; Trice et al. 2004). *A. anophagefferens* blooms have also occurred recently in Saldanha Bay, South Africa (Probyn et al. 2001). Low abundances of *A. anophagefferens* have been observed along the entire eastern seaboard of the U.S. from Maine to Florida, indicating the potential for these blooms to continue to spread north and south of their current range (Anderson et al. 1993, Popels et al. 2003). Brown tides in Long Island bays have been associated with high cell concentrations ($>10^6 \cdot mL^{-1}$) and substantially increased light attenuation, which apparently caused a large-scale die-off of *Zostera marina* seagrass beds (Dennison 1987). Blooms have also caused mass mortality and recruitment failure in shellfish populations (*Argopecten irradians*, *Mercenaria mercenaria* [Bricelj et al. 1989, 2001, Greenfield and Lonsdale 2002, Padilla et al. 2006]). Subsequent research has established that *A. anophagefferens* also adversely impacts growth and grazing by zooplankton (Lonsdale et al. 1996, Caron et al. 2004, Gobler et al. 2004, Deonarine et al. 2006).

The first recorded *Ar. lagunensis* bloom in the Laguna Madre and Baffin Bay, Texas, began in January 1990 and persisted for nearly 8 years, representing the longest continuous HAB event ever recorded (Buskey et al. 2001). During this time, *Ar. lagunensis* comprised most of the algal biomass and generally maintained cell densities of 0.5–$5 \times 10^6 \cdot mL^{-1}$ (Buskey and Stockwell 1993, Buskey et al. 1997). The bloom terminated in the fall of 1997, but reemerged during the summer and fall of 1998 (Buskey et al. 2001). Since then it has occurred intermittently in the Laguna Madre system. Low concentrations of *Ar. lagunensis* cells have also been observed in coastal bays across Florida, Texas, and Mexico (Villareal et al. 2004). As in *A. anophagefferens*, the *Ar. lagunensis* bloom caused a substantial increase in light attenuation, resulting in seafloor-level shading that has decreased the abundance of once extensive seagrass beds (Onuf 1996). There has also been a decrease in the biomass and diversity of benthic invertebrates in the Laguna Madre (Ward et al. 2000). The dominant clam, *Mulinia lateralis*, virtually disappeared after the onset of the brown-tide bloom (Montagna et al. 1993), and the dominant polychaete, *Streblospio benedicti,* has declined in abundance by two orders of magnitude (Buskey et al. 1997). The brown-tide bloom also was associated with substantial decreases in the grazing activity, growth, and egg release rates in mesozooplankton (e.g., *Acartia tonsa*), and decreases in the abundance and grazing rates of microzooplankton (Buskey and Stockwell 1993).

A. anophagefferens and *Ar. lagunensis* blooms are associated with high levels of algal biomass and thus severe light attenuation. *A. anophagefferens* is able to maintain a near-maximum growth rate (at 20°C) at a light intensity of 100 μmol photons $\cdot m^{-2} \cdot s^{-1}$ (~4% of noon solar irradiance; Milligan and Cosper 1997, MacIntyre et al. 2004), while *Ar. lagunensis* is able to do the same at a slightly higher irradiance (150 μmol photons $\cdot m^{-2} \cdot s^{-1}$; Buskey et al. 1998). Curiously, the accessory pigments in *A. anophagefferens* make it well adapted for growth at low levels of blue light, as found in the deep chl maximum in the open ocean where pelagophytes commonly occur, and less well adapted for growth at low inten-

sities of longer wavelength visible light that occur during brown-tide blooms (Yentsch et al. 1989). Based on these observations, Yentsch et al. suggested that *A. anophagefferens* might be an expatriate open-ocean species. The ability of the two brown-tide species to grow maximally at low light intensities likely contributes to their ability to compete well with non-HAB species during blooms (Milligan and Cosper 1997, MacIntyre et al. 2004).

In summary, there is no doubt that these two pelagophytes have a destructive impact on ecosystem health and aquaculture endeavors. To gain further insight to the biology of these algae, we sequenced the chloroplast genomes of *A. anophagefferens* CCMP1984 and *Ar. lagunensis* CCMP1507. The present study utilizes a fosmid cloning approach that has significantly simplified the process of sequencing chloroplast genomes from these minute cells, because it eliminates the need to obtain purified organellar DNA (Cattolico et al. 2008). The information generated in this study contributes to our understanding of stramenopile chloroplast function, evolution, and phylogeny.

MATERIALS AND METHODS

Strains and growth conditions. One-liter cultures of *A. anophagefferens* CCMP1984, isolated as a single cell from Great South Bay, New York, USA, in 1986, were grown axenically in 2 L flasks at 21°C on a 14:10 light:dark (L:D) cycle at 100 microeinsteins $\cdot$ m^{-2} in F/2 seawater media (Guillard 1975) supplemented with selenium. A culture of *Ar. lagunensis* CCMP 1507, isolated from Laguna Madre, Texas, USA, in 1992, was obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (West Boothbay Harbor, ME, USA). One-liter cultures were grown axenically in 2.8 L flasks at 22°C on a 12:12 L:D cycle at 60 microeinsteins $\cdot$ m^{-2} cool-white light without shaking in F/2 seawater media supplemented with silica.

Genomic DNA purification. To obtain *A. anophagefferens* DNA, eight 1 L cultures in exponential growth phase were combined and harvested using a Sharples continuous flow centrifuge (model T-1P, Pennwalt Corp., Warminster, PA, USA). Cells were concentrated with a tabletop centrifuge (IEC CL31R multispeed; Thermo Scientific, Wilmington, DE, USA), and the tube was temporarily stored in liquid nitrogen vapors. The tube was briefly placed in 68°C water before the pellet was ground using a pestle and placed into 50 mL plastic centrifuge tubes. Then 2X cetyltrimethylammonium bromide and β-mercaptoethanol (1% final volume) were added to each tube, and the tubes were incubated for 3 h at 68°C. A 1:1 ratio of phenol with chloroform-isoamyl alcohol (IAA) was added to the cell suspension. After partial inversion, the cell suspensions were centrifuged at 5,000*g* for 5 min. The aqueous fraction was removed, and the IAA treatment was repeated. The aqueous phase was transferred to a new tube and precipitated with 0.6 volumes of ice-cold isopropanol. The precipitant was dissolved in Tris-EDTA buffer and then quantified with a Nanodrop-1000 UV-spectrophotometer (Thermo Scientific) by agarose gel electrophoresis.

Total high-molecular-weight DNA (>500 kb in size) was extracted from *Ar. lagunensis* for fosmid library construction using the Qiagen Genomic-Tip 500G kit according to the manufacturer's directions (Qiagen, Valencia, CA, USA). A brief description of the extraction process is as follows: *Ar. lagunensis* CCMP 1507 cells were harvested by centrifugation at 1,000*g* and 4°C for 5 min and then resuspended in cold lysis buffer (20 mM EDTA, 10 mM TrisCl, pH 8, 1% Triton X, 500 mM guanidine-HCl, and 200 mM NaCl). The lysed cell suspension was incubated at 37°C for 1 h with gentle agitation. The DNA was further treated with RNAse (20 μg $\cdot$ mL^{-1}) for 30 min at 37°C followed by Proteinase K (0.8 mg $\cdot$ mL^{-1}) treatment for 2 h at 50°C with gentle agitation. Cell debris was removed from the lysed cell suspension by centrifugation at 7,000*g* for 20 min at 4°C, and 3 mL of the lysate was added to each Qiagen genomic tip. The columns were washed twice, and DNA was eluted from the genomic tip. DNA precipitation used 0.7 volume of room temperature isopropanol. The DNA was pelleted by centrifugation at 7,000*g* for 20 min at 4°C and washed with 4 mL of cold 70% ethanol, before being recentrifuged at 7,000*g* for 10 min at 4°C. The supernatant was removed, and the pellet air-dried for 10 min. The pellet was resuspended in 1 mL of TE buffer and concentrated with a speed-vacuum (Savant SPD131DDA SpeedVac Concentrator; Thermo Scientific). DNA was quantitated using fluorometry following Hoescht dye binding (Gallagher 2001).

A. anophagefferens *chloroplast genome sequencing and assembly.* The genome of *A. anophagefferens* was sequenced using whole-genome shotgun (WGS) strategy. Three libraries with an insert size of 2–3 Kbp, 6–8 Kbp, and 35–40 Kbp were used. The sequenced reads were screened for vector using cross_match (Gordon et al. 1998), trimmed for vector and quality, and filtered to remove reads shorter than 100 bases, which resulted in the following data set: 306,657 of 2–3 Kbp reads containing 215 Mbp of sequence; 301,713 of 6–8 Kbp reads containing 215 Mbp of sequence; and 37,362 of 35–40 Kbp reads containing 18 Mbp of sequence. The data were assembled using release 2.10.6 of Jazz (Aparicio et al. 2002), a WGS assembler developed at the DOE Joint Genome Institute (JGI). A word size of 13 was used for seeding alignments between reads. The unhashability threshold was set to 40, preventing words present in the data set in more than 40 copies from being used to seed alignments. A mismatch penalty of −30.0 was used, which tended to assemble together sequences that were more than ~97% identical. The final assembly contained 1,202 scaffolds, with a scaffold of N/L50 of 13/1.3 Mbp, and a contig N/L50 of 405/34 Kbp. The sequence depth derived from the assembly was 7.00 ± 0.13.

The majority of the chloroplast genome of *A. anophagefferens* assembled to a single 88,471 bp scaffold (number 67). This scaffold was obtained from the JGI Web site (http://genome.jgi-psf.org/Auran1/Auran1.download.html) and contained one gap ~1,200 bp in length between the end of *rpoC2* and the beginning of *rps2*. To span this gap, primers designed to the gap-bordering regions were used in a standard PCR reaction containing genomic *A. anophagefferens* DNA as template.

Ar. lagunensis *chloroplast genome sequencing and assembly.* The genome of *Ar. lagunensis* was sequenced using a fosmid cloning approach (Cattolico et al. 2008). Large-insert fosmid clones were prepared from high-molecular-weight DNA using the method previously described in Raymond et al. (2005) and subsequently adapted in Cattolico et al. (2008). Briefly, sheared (45 kb) total cellular DNA was size-selected by agarose gel-electrophoresis using a DRIII CHEF gel apparatus (Bio-Rad, Hercules, CA, USA), followed by end repair and packaging into the PCC1Fos Vector, using the Epicentre CopyControl Fosmid Library Production Kit (Cat CCFOS110; Epicentre Biotechnologies, Madison, WI, USA). Clones were plated using 12 μg $\cdot$ mL^{-1} chloramphenicol selection, picked using the Q-pix automated colony picker (Genetix Ltd., New Milton, Hampshire, UK) and inoculated into 384-well freezing plates that contained UWGC freezing medium [10 g $\cdot$ L^{-1} tryptone, 5 g $\cdot$ L^{-1} yeast extract, 0.17 M NaCl, 36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM $Na_3C_6H_5O_7 \cdot 2H_2O$, 6.8 mM $(NH_4)_2SO_4$, 4.4%

glycerol, 12 $\mu g \cdot mL^{-1}$ chloramphenicol]. Reactions were inoculated using a 384-pin plastic plate replicator (ISC BioExpress, Kaysville, UT, USA) directly from the 384-well fosmid glycerol stock (see above).

Fosmid DNA was recovered using a standard alkaline-lysis protocol and sequenced according to ABI manufacturer's directions in an 8 μL reaction using 0.5 μL Big Dye Terminator reagent (BDT) version 3.1 (Applied Biosystems Inc., Foster City, CA, USA), 5 pmol of vector end-sequencing primers, and 100 ng DNA per reaction. Cycle sequencing was accomplished using standard thermocycling conditions (3 min denaturation at 94°C, 99 cycles of the following regime: 94°C for 30 s, 50°C for 20 s, 60°C for 4 min) and analyzed on an ABI 3730 automated sequencer (Applied Biosystems Inc.). To process the sequence reads, vector sequences were removed, and sequences were further trimmed from both ends until a window of 12 bp with 90% of positions having a Phred score of Q20 or greater was reached. Sequences were compared using BLASTX (Altschul et al. 1997) to the GenBank nonredundant database and to a custom database consisting of published chloroplast genomes. Fosmids in which both end sequences had high-quality matches (*E*-value < 10^{-4}) to a chloroplast gene as judged by both BLAST analyses were identified as chloroplast derived. A total of 768 fosmids were end sequenced. Of those, 18 had end sequences with chloroplast signatures. All fosmid end sequences are available on our Web site: http://chloroplast.ocean.washington.edu.

Fosmid restriction analysis and tiling. The 18 *Ar. lagunensis* clones with chloroplast end sequences were analyzed by multiple complete digest (MCD) analysis as described by Raymond et al. (2005) and Hayden et al. (2008). Briefly, fosmid clone DNA was isolated by alkaline lysis, resuspended in 40 μL TE buffer (pH 8.0) that contained 5 $U \cdot \mu L^{-1}$ RNAse A and 0.05 $U \cdot \mu L^{-1}$ RNAse T1. The resuspended fosmid DNA was separated into four separate 6 μL aliquots that were digested with 5 $U \cdot \mu L^{-1}$ *Hin*dIII, *Bgl*II, *Nsi*I, and *Eco*RI, respectively. Digests were subjected to electrophoretic separation on a 0.8% agarose gel, using 4°C circulating 2X GGB buffer [80 mM Tris base, 40 mM sodium acetate, 4 mM EDTA, 52 mM glacial acetic acid, pH 8.0–8.3 (Wong et al. 1997)] at 8 $V \cdot cm^{-1}$ for 18 h. Bands were visualized by staining with SYBR Green I (Invitrogen, Carlsbad, CA, USA), followed by image capture using a Typhoon 8600 Variable Mode Imager (Amersham Biosciences, Piscataway, NJ, USA). Automated band calling was performed using the Quantitative Gel Analysis Program (Wong et al. 1997, updated in Hayden et al. 2008).

Restriction data were analyzed using GenVal software (Hayden et al. 2008), which compares DNA fingerprints and aligns end-sequence data for multiple clones, either against a reference genome or de novo. The program uses the restriction data (or Gel Fragments: GFs) and attempts to align clones based upon the shared GFs. With some refinements, this classical contig-building approach uses the specific overlap metric, mutual overlap statistic (Coulson et al. 1986, Olson et al. 1986), from several independent experiments to generate a composite mutual overlap statistic. This metric uses all restriction fragments from all digests and aligns clones based upon the shared fragments and extends them based upon their unique fragments. Bias in the resulting alignment creates a contig that is roughly 1.5 times the actual contig size when the clones are sequenced (W. Gillett, personal communication).

Fosmid DNA was purified using a standard alkaline-lysis protocol, sheared to 3–5 kb fragments using a Hydra-Shear (GeneMachines Inc., San Carlos, CA, USA), and transformed into a blunt-ended pUC19 library using 100 $\mu g \cdot mL^{-1}$ carbenicillin and X-Gal/IPTG (US Biochemicals, Cleveland, OH, USA) for selection on solid agar bioassay plates (VWR, West Chester, PA, USA). White colonies were picked using a Q-pix automated colony picker (Genetix Ltd.) and inoculated into 384-well freezing plates (Cat# X7001; Genetix Ltd.) containing UWGC freezing medium (as described earlier but with the antibiotic selection changed to include only 100 $\mu g \cdot mL^{-1}$ carbenicillin). Templates were amplified using TempliPhi (Amersham Biosciences) and sequenced according to standard protocols using BDT v3.1 (0.25 μL per reaction). Sequencing reactions were analyzed using ABI 3730 automated sequencers (Applied Biosystems). Sequencing reads were processed using the Phred/Phrap/Consed package of base-calling, sequence assembly, and finishing/editing software (Ewing and Green 1998, Ewing et al. 1998, Gordon et al. 1998). Final finishing of the sequence was performed using experiments designed by Autofinish (Gordon et al. 2001).

Genome annotation. For both genomes, open reading frames (ORFs) were initially predicted using Glimmer 3.0 (Delcher et al. 2007) and then refined manually. Ribosomal RNA genes were identified with RNAmmer (Lagesen et al. 2007) and refined with the aid of the comparative RNA Database (Cannone et al. 2002). Genes for tRNAs, tmRNA, and the signal recognition particle RNA were identified using tRNASCAN-SE (Lowe and Eddy 1997), ARAGORN (Laslett and Canback 2004), and SRPscan (Regalia et al. 2002). Predicted gene functions were initially assigned using a BLASTP search of a custom chloroplast genome database and refined manually with the aid of conserved protein motifs identified using the PFAM (Finn et al. 2006) database. Tandem repeats were found with Tandem Repeat Finder (Benson 1999) using default settings. Inverted repeats were found with E-inverted from the EMBOSS package (Rice et al. 2000) using the default settings and the additional constraint that repeats had to be >80% similar, and the length of the loop shorter than the stem. Artemis, the Artemis Comparison Tool, and MAUVE were used to visualize the comparative genome architecture (Rutherford et al. 2000, Darling et al. 2004, Carver et al. 2005). Circular genome maps were created with OGDRAW (Lohse et al. 2007). Sequences have been deposited in GenBank under accession numbers GQ231541 (*A. anophagefferens* CCMP1984) and GQ231542 (*Ar. lagunensis* CCMP 1507).

Reconstruction of genome phylogenies. Genes were sorted by function and reduced to 160 genes that were divided into basic functional groups. Of the 160 genes within the functional groups, 41 were found in all chloroplasts and removed for the generation of the dendrogram (yielding 119 different characters for the analyses). A 42nd gene (*tufA*) was also common to all genomes but did not group within our functional assignments. Heat map generation was completed using the conditional formatting setting in Excel (Microsoft, Redmond, WA, USA). A hierarchical clustering approach (Ward's minimum variance, parameters: distance type, Euclidean; scale type, standard deviation) was used to generate a dendrogram based on presence or absence.

RESULTS AND DISCUSSION

Ar. lagunensis *de novo contig building and sequencing of clones.* Of the 768 end-sequenced *Ar. lagunensis* clones, 18 were assessed as having a chloroplast-like signature (see Materials and Methods). These clones were digested with four different restriction enzymes: *Hin*dIII, *Bgl*III, *Nsi*I, and *Eco*RI. The patterns from these digests were used to create a de novo tiling path (Fig. S1, panel A, in the supplementary material), which showed the overlap and order of the clones. A total of 12 clones fell into a single contig. A minimal tiling path of five clones based on the de novo contig building was

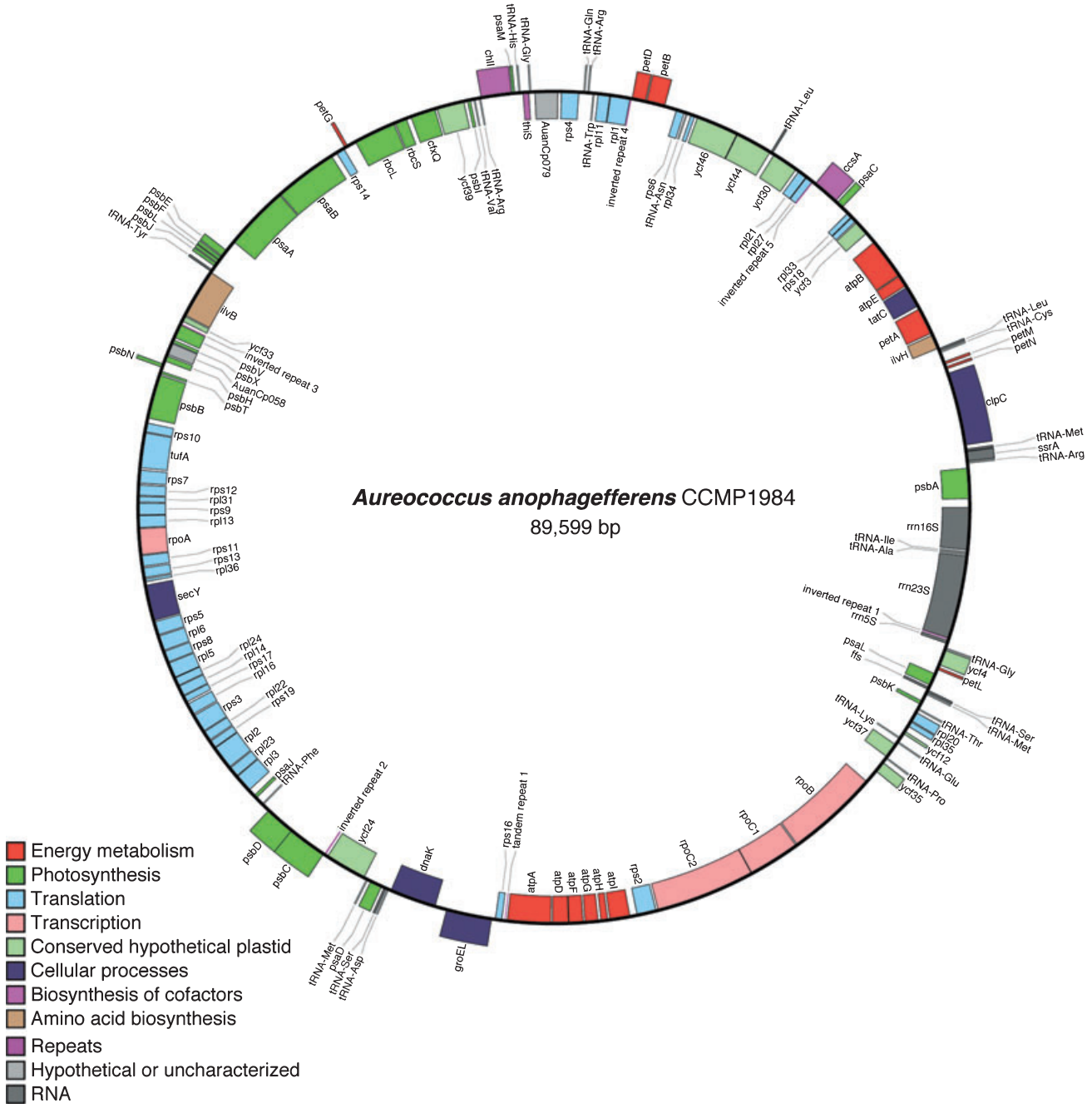

FIG. 1. *Aureococcus anophagefferens* and *Aureoumbra lagunensis* chloroplast genome maps. Both genomes begin at the 16S rRNA and proceed clockwise. Genes are color coded according to their functional category. Small inverted repeats (IR) and tandem repeats (TR) are also shown.

sequenced, which included the ''anchor'' clone and four others (numbers 2–5, Fig. S1, panel A). Once these clones were sequenced and finished separately, they were aligned using Sequencher (Gene Codes Inc. Ann Arbor, MI, USA) and formed a 94 Kbp contig (Fig. S1, panel B). A sixth clone (marked ''fin.,'' Fig. S1, panel B) overlapped both the left and right ends of the five-clone contig, and custom primers were designed to sequence the portion of the fosmid that closed the circle.

Genome architecture. The chloroplast genomes of *A. anophagefferens* and *Ar. lagunensis* are 89,599 and 94,346 bp in size, respectively (Fig. 1; Table 1). The two genomes display a high degree of gene synteny, broken by only a few major large-scale rearrangements (Fig. 2). Transfer RNA genes border many, but not all, of these rearrangement sites. Architecturally, neither genome contains a large inverted repeat, and both genomes display a paucity of small, internal repeats. The *A. anophagefferens* genome

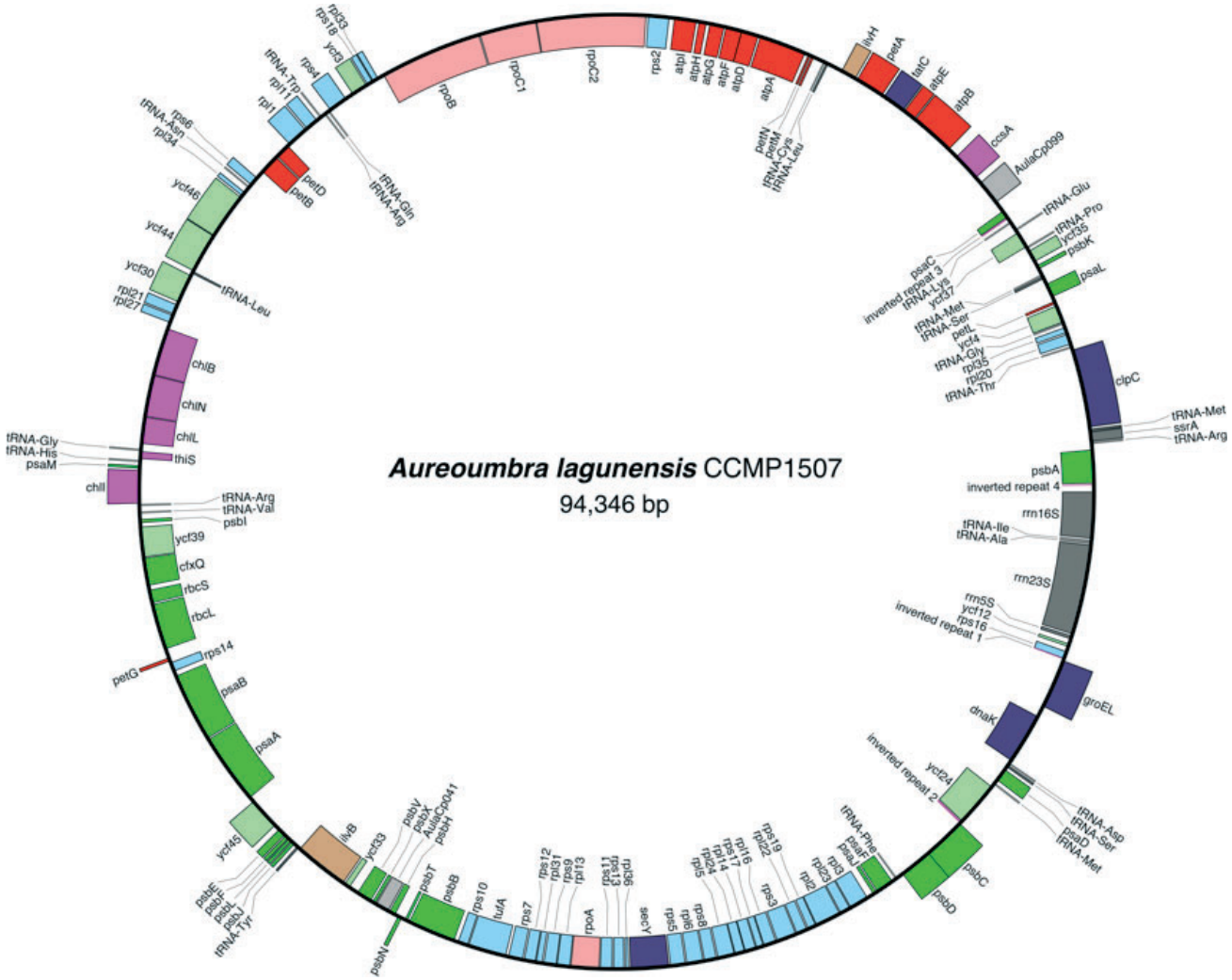

FIG. 1. Continued.

contains six inverted repeats, 20 ± 2.7 bp in length with an average loop domain of 2.6 ± 1.3 bp. Including an additional single tandem repeat (17 bp; 1.9 copies), these small repeats represent 0.29% of the total *A. anophagefferens* genome. The four inverted repeats of *Ar. lagunensis*, averaging 24.5 ± 7.5 bp in size with a loop domain of 2.6 ± 1.3 bp, represent 0.22% of this cpDNA. Similar to bacterial genomes (Lillo et al. 2002), the inverted repeats found in *A. anophagefferens* and *Ar. lagunensis* cpDNAs are located 3′ to two genes that are transcribed toward one another on opposite coding strands. Only one repeat element, located between *ycf4* (ABC transporter) and *psb*D (PSII D2 protein and the 44 kDa apoprotein) genes, is common to both genomes. There is no sequence homology between these two common held inverted repeats. Notably, eight of the 10 repeats found in *A. anophagefferens* and *Ar. lagunensis* cpDNAs serve genes related to photosynthesis or energy production (Table 2). As repeat elements are often located adjacent to these same genes in those stramenopiles that have been sequenced (Cattolico et al. 2008), the placement conservation of such an element may imply a functional constraint.

TABLE 1. Overview of *Aureococcus anophagefferens* and *Aureoumbra lagunensis* chloroplast genomes.

	A. anophagefferens	*Ar. lagunensis*
Length (bp)	89,599	93,436
G + C content (%)	36.0	33.6
Total number of genes	137	141
Protein-coding genes	105	110
With assigned function	92	96
Conserved hypothetical (*ycf*)	11	12
Hypothetical (orf)	2	2
Ribosomal RNA operons	1	1
Transfer RNA	27	27
Transfer-messenger RNA (tmRNA)	1	1
Signal recognition particle RNA (ffsRNA)	1	0

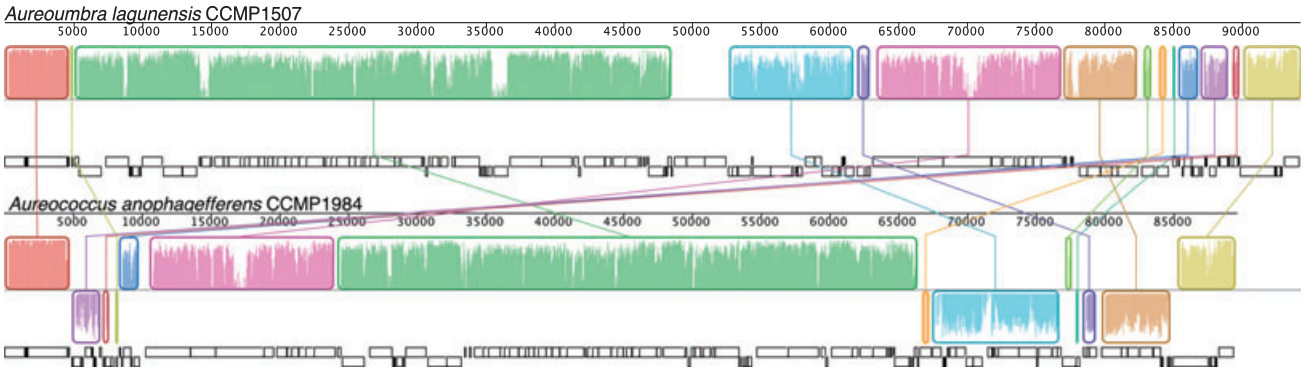

FIG. 2. Linear comparison of rearrangements between *Aureococcus anophagefferens* and *Aureoumbra lagunensis* chloroplast genomes. Each colored block indicates a region of synteny between the two genomes, and shading within each block represents the degree of sequence similarity. Rearrangements that have resulted in a strand change in *A. anophagefferens* relative to *Ar. lagunensis* are shown on the second line of *A. anophagefferens*. An outline of the location of coding genes is shown below each genome.

TABLE 2. Total and locations of tandem repeats (T), inverted repeats (IR), and absence of a repeat (O) immediately adjacent to coding regions in eight stramenopile chloroplast genomes sequenced. Some genomes contain multiple repeats adjacent to the same gene.

Organism	Genome size (kb)	Total repeats	*atp*A	*psa*C	*psb*A	*psb*D/C	*psb*V	*pet*B/D	*css*A
Aureococcus anophagefferens	89,599	6	T	O	O	IR	IR	IR	IR
Aureoumbra lagunensis	93,436	4	O	IR	IR	IR	O	O	O
Heterosigma akashiwo (CCMP452)	160,149	65	T	IR	T, T, T, IR	T, T, T, IR	O	IR	IR
Heterosigma akashiwo (NIES293)	159,370	56	T	IR	T, T	T, T	O	IR	IR
Pinguiococcus pyrenoidosus	114,670	29	O	IR	IR, IR	IR	O	IR	O
Odontella sinensis	119,704	23	IR	O	O	IR	IR	O	O
Phaeodactylum tricornutum	117,369	18	IR	IR	IR	O	IR	IR	O
Thalassiosira pseudonana	128,814	31	IR, T	IR	IR	O	O	O	O
Vaucheria litorea	115,341	49	IR	IR	IR, IR	O	O	IR	O

Gene complement. *A. anophagefferens* and *Ar. lagunensis* genomes contain genes that encode 105 and 110 proteins, respectively. The five genes present in *Ar. lagunensis* and absent in *A. anophagefferens* are *chl*L, *chl*N, *chl*B, *psa*F, and *ycf45*. Both genomes contain one ribosomal operon containing 16S, 23S and 5S rRNAs, as well as 27 transfer RNAs (Table 1). Though each algal chloroplast genome has a tmRNA, the encoded tag peptides for that of *A. anophagefferens* (ANNILKLFTKRPVLAFA) and *Ar. lagunensis* (AKFFNSIFPTRPQLAFA) differ. *A. anophagefferens* also encodes a signal recognition particle RNA (ffs), but we were unable to detect this small RNA in *Ar. lagunensis*, suggesting it is either absent or extremely divergent. Approximately 10% of the two cpDNAs encode hypothetical *ycf* genes. Eleven genes are common to *A. anophagefferens* and *Ar. lagunensis*, including *ycf3, 4, 12, 19, 24, 30, 33, 35, 37, 39,* and *46*. *Ar. lagunensis* also contains *ycf45*. Five of these ORFs have been assigned a putative function: *ycf3* and *4* (stable accumulation of PSI complex), *ycf24* (ABC transporter), *ycf30* (LysR-like transcriptional regulator), and *ycf46* (ATPase AAA protein) family.

Both *A. anophagefferens* (Auan Cp 079) and *Ar. lagunensis* (Aula Cp 099) also encode a unique protein of 267 amino acids that has not been previously reported. Though the 5′ third of this protein has diverged between the two algae, the remainder of the protein remains highly conserved at the amino-acid level (overall 40% identical, 76% similar). This sequence similarity is maintained despite the fact that this ORF is located differently in one genome relative to the other. AuanCp079 lies between *rps4* and *thi*S, while AulaCp099 is flanked by *psa*C and *ccs*A. Neither protein contains identifiable functional domains, but both have at least five transmembrane domains. Although these proteins show no homology to those from previously sequenced organisms, they have a high degree of similarity in the 3′ half to two environmental sequences from the Indian Ocean (JCVI_PEP_1112708879268 and JCVI_PEP_1112740957822) (Rusch et al. 2007). The second hypothetical ORF encoded by both genomes is smaller: 140 aa (AulaCp41) or 150 aa (AuanCp58). These genes are located at syntenic positions on the two genomes, both flanked by *psb*X and *psb*H, and are 40% identical (73% similar) to each other at the amino-acid level.

Among the proteins coded by the two newly sequenced pelagophyte chloroplast genomes, the D_1 protein, encoded by *psb*A, is of particular interest. This protein, which is also present in cyanobacteria,

cyanophages, and in the chloroplast genomes of photosynthetic eukaryotes (Hardison et al. 1995, Chénard and Suttle 2008), is a core component of PSII. As the D_1 protein is damaged by electron transfer reactions, it is rapidly degraded and replaced by newly synthesized D_1 product. The immature D_1 protein predominantly has a C-terminal extension that is cleaved specifically at alanine 344 by the protease CtpA after the D_1 protein is imbedded in the thylakoid membrane (Ivleva et al. 2000). Recent studies show that C-terminal removal at alanine 344 is absolutely necessary to insure both proper association of the mature (and intrinsic) D_1 protein with functional tetranuclear magnesium cluster and the proper interaction of D_1 with the extrinsic PSII subunits: *psb*O, *psb*U, and *psb*V (Roose and Pakrasi 2004). Though D_1 proteins are highly conserved among all taxa (Hardison et al. 1995, Raymond and Blankenship 2004), these proteins often differ in the sequence of their C-terminus domain (Fig. 3). The D_1 protein extension of cyanobacteria, cyanophage, *Cyanophora paradoxa, Emiliania huxleyi,* and seven rhodophytes is 16 amino acids long, whereas in Viridiplantae and chlorophytic algae, it is only eight to nine amino acids in length. Extension size difference among these organisms is determined by the insertion/deletion of a seven-amino-acid region (Yamamoto et al. 2001). In contrast to previous reports (Hardison et al. 1995, Satoh and Yamamoto 2007), we find that different stramenopile taxa are not uniform in C-terminus length, having either the extended or a shortened C-terminal sequence; *A. anophagefferens, Ar. lagunensis, Bumilleriopsis filiformis,* and *Synura petersenii* lack the seven-amino-acid region, while the longer 16-amino-acid domain is present in 10 other stramenopiles (Fig. 3).

Why the C-terminus itself is retained (with or without the seven-amino-acid addition) remains an open question. Five *Euglena gracilis* species and several dinoflagellate genera lack the C-terminal extension completely, though at least one dinoflagellate representative (*Gymnodinium mikimotoi*) has a 13-amino-acid C-terminus extension (perhaps reflecting different ancestral endosymbionts). Mutagenesis studies using *Synechocystis* PCC 6803 have been used to probe the function of this enigmatic D_1 addition. Data showed that omission of the D_1 extension compromised optimal photosynthetic fitness, especially when this cyanobacterium was maintained under high-light growth conditions (Yamamoto et al. 2001, Kuvikova et al. 2005).

Phylogenetic assessment. To examine the phylogenetic structure of the genomes with respect to intact genetic elements, a matrix was created with the 232 different genes from 25 available chloroplast genomes. Genes common to all 25 genomes were not used in the generation of the phylogeny. A hierarchical clustering approach based on the presence and absence of genes (Ward's minimum variance) yielded the dendrogram shown in Figure 4. The complete list of genes analyzed and a more detailed version of Figure 4 are provided, respectively, in Table S1 and Figure S2 in the supplementary material. Supporting these data, a nearest-neighbor clustering model revealed similar dendrogram structure (not shown). Indeed, both analyses confirm that *A. anophagefferens* and *Ar. lagunensis* are sister taxa, a finding consistent with

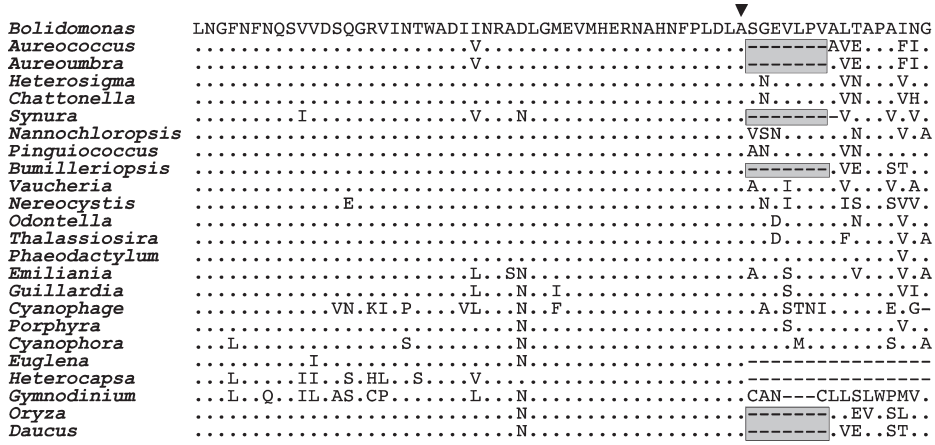

FIG. 3. Analysis of the *psb*A gene conservation/evolution. The alignment shown is of the C-termini of PsbA amino-acid sequences from 14 stramenopiles (*Aureococcus anophagefferens, Aureoumbra lagunensis, Bolidomonas pacifica, Bumilleriopsis filiformis, Chattonella subsalsa, Heterosigma akashiwo, Nannochloropsis oculata, Nereocystis luetkeana, Odontella sinensis, Phaeodactylum tricornutum, Pinguiococcus pyrenoidosus, Synura petersenii, Thalassiosira pseudonana,* and *Vaucheria litorea*), one haptophyte (*Emiliania huxleyi*), one cryptophyte (*Guillardia theta*), one cyanophage (*Prochlorococcus* phage P-SSP7), one rhodophyte (*Porphyra purpurea*), one glaucocystophyte (*Cyanophora paradoxa*), one euglenozoa (*Euglena gracilis*), two dinoflagellates (*Heterocapsa niei* and *Gymnodinium mikimotoi*), and two flowering plants (*Oryza sativa* and *Daucus carota*). A conserved seven-amino-acid residue deletion at the C-termini (gray box) is shared between three stramenopiles (*Aureoumbra, Aureococcus, Bumilleriopsis,* and *Synura* [with one additional residue deletion]) and two terrestrial plants (*Oryza* and *Daucus*). Alanine 344, the cleavage site for the CtpA protease processing of the immature transcript, is indicated with a wedge. Sequences are aligned relative to *Bolidomonas psb*A, with identical amino acids indicated by dots, and gaps by dashes.

FIG. 4. Phylogenetic reconstruction of genomes based on the presence (black boxes) or absence (white boxes) of genes within the genome of 25 chloroplasts. Genes found to be in every genome were not considered, leaving a total of 232 potential genes. A complete and alphabetical list can be found in Table S1 (see supplementary material), and a more detailed figure with the gene names listed next to the heat map can be found in Figure S2 (in the supplementary material).

prior comparisons of rRNA sequences (DeYoe et al. 1997). Thereafter, the chloroplast genomes of these species display a close similarity to *E. huxleyi* and *G. theta*. The overall dendrogram arrangement is also retained in a nearest-neighbor phylogenetic analysis of *rbc*L sequences (S. W. Wilhelm, unpublished data). Amino-acid sequence similarity between proteins encoded by orthologous genes

on the two genomes ranges from 32% to 100%, with an average of 78%. Interestingly, this degree of divergence is quite similar to that between two other stramenopiles, the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (range 31%–100%, average 77%). Both pairs of species share much lower average amino-acid similarity than two strains of the same species, *Heterosigma akashiwo* CCMP452 and NIES293 (average 99% identity).

The two pelagophyte chloroplast sequences have yielded more precise understanding of the core stramenopile chloroplast genome. Twenty-one genes are absent from both pelagophyte chloroplast genomes that are present in all previously reported stramenopile chloroplast genomes [those of the diatoms *Odontella sinensis* (Kowallik et al. 1995), *P. tricornutum*, and *T. pseudonana* (Oudot-Le Secq et al. 2007); the raphidophyte *H. akashiwo* (Cattolico et al. 2008); and the xanthophyte *Vaucheria litorea* (Rumpho et al. 2008)]. These genes (*dna*B, *fts*H, *pet*F, *psa*E, *psa*I, *psb28, psb*Y, *psb*Z, *rpl12, rpl18, rpl19, rpl29, rpl32, rpl4, rps20, sec*A, *thi*G, *ycf16, ycf41, ycf42, ycf66*) account for ~12.5 Kbp and encode functions related to photosynthesis/energy transduction, transcription/translation, and cellular processes.

Insights into chloroplast to nucleus transfers. Five genes encoded in the *Ar. lagunensis* chloroplast genome are absent in *A. anophagefferens* (*chl*L, *chl*N, *chl*B, *psa*F, and *ycf45*). Given the close relationship of these two organisms, this difference must reflect a relatively recent event, either gene loss or functional transfer to the nucleus. The availability of the sequenced *A. anophagefferens* nuclear genome gave us opportunity to search for these missing genes using the *Ar. lagunensis* chloroplast-encoded proteins. Though we failed to find *chl*L, *chl*N, or *chl*B, both *psa*F and *ycf45* were identified using this approach.

Of the two traditionally chloroplast-encoded genes that are nuclear encoded in *A. anophagefferens*, the *psa*F gene (subunit III of PSI) appears to be indispensable to the maintenance of photosynthesis. The protein product encoded by this gene mediates electron transfer from plastocyanin and cytochrome to PSI (Hippler et al. 1998). The nuclear-encoded *A. anophagefferens* gene encodes a protein that is 22 amino acids (MSLKPETAAALGKAAGAAALGA) longer at its N terminus than the *Ar. lagunensis* chloroplast-localized polypeptide. Given that the extended domain shows significant hydrophobic character using the Kite-Doolittle scale (http://www.vivo.colostate.edu/molkit/hydropathy/index.html), it undoubtedly serves both as a signal sequence to target the chloroplast and a transit function, since the *psa*F protein displays both transmembrane and luminal potential via Phobius prediction (http://phobius.sbc.su.se/). The conserved, adjacent alanine residues in both chloroplast and nuclear *psa*F proteins most likely serve as the transit peptide-processing site (Smeekens et al. 1990). The *psa*F gene is chloroplast encoded in the glaucophyte *Cyanophora paradoxa*, the haptophyte *Emiliania huxleyi*, all five rhodophytes sequenced to date, and 13 stramenopiles for which we now have data (not shown). To our knowledge, this report is the first example of a *psa*F gene to be nuclear encoded in an alga that does not contain chl *b*. The *psa*F gene is nuclear encoded in Viridiplantae and chlorophytic algae and in the chlorarachniophyte *Bigelowiella natans,* which has the smallest chloroplast genome (69.2 kb) on record (Rogers et al. 2007).

Significant differences were seen when the product of the *A. anophagefferens ycf45* nuclear-encoded gene was compared to its *Ar. lagunensis* chloroplast-encoded ortholog, both in amino-acid sequence and in the presence of insertion/deletions. In a generalized BLAST search, the putative *A. anophagefferens ycf45* has best identity to proteins found in Viridiplantae and cyanobacteria, whereas *Ar. lagunensis* chloroplast *ycf45* has closer affinity to orthologs in stramenopiles, rhodophytes, and cyanobacteria. Both proteins are members of the large AAA–ATPase superfamily of proteins, which are responsible for multiple cellular functions (http://aaa-proteins.uni-graz.at/AAA/List.html), arguing that it would be premature to assign the resident nuclear gene as an evolutionary lateral gene product.

Within the framework of gene retention and genome shuffling, the maintenance of the three light-independent chl-biosynthesis genes encoding protochlorophyllide oxidoreductase subunits L, N, and B (*chl*L, *chl*N, and *chl*B) in the chloroplast genome of *Ar. lagunensis* is particularly interesting. The formation of chl in oxygenic prototrophs can be achieved by two distinct enzymes—the light-dependent (POR) or the light-independent (LIPOR) protochlorophyllide oxidoreductase (Shi and Shi 2006). Both processes involve the same initial molecule, magnesium protoporphyrin IX, which is first methylated to generate protochlorophyllide. Then, through either the LIPOR or POR pathways, protochlorophyllide is reduced to chlorophyllide *a*, which in turn becomes chl *a* following a reduction of the 4-vinyl group and esterification of the C_{17}-propionate group with phytol (Senge 1993, Griffiths et al. 1996).

It is known that several cyanobacteria encode both LIPOR and POR chl-biosynthetic pathways (Shui et al. 2009). However, evolutionary selection appears to have driven the maintenance of either LIPOR or of POR (but not both) pathways in several prokaryotic and eukaryotic lineages. Syanapomorphies to the cyanobacterial LIPOR *chl*L, *chl*N and *chl*B genes are seen (Boivin et al. 1996, Armstrong 1998, Mulkidjanian et al. 2006, Yoon et al. 2009) in the anoxygenic phototrophic bacteria *Rhodomonas paulustris* (proteobacterium), *Chorobium tepdium* (green sulfur bacterium), and *Chloroflexus aurantiacus* (green nonsulfur bacterium), and in *Paulinella chromatophora*—an amoeba that contains

cyanobacterial-like inclusions. Similarly, at least 16 chlorophytic algae (e.g., *Chlorokybus atmophyticus*, *Chara vulgaris, Oedogonium cardiacum*, *Stigeoclonium helveticum*), as well as nonseed plants (e.g., bryophytes, lycophytes, hornworts, and ferns), encode LIPOR genes in their chloroplasts. In contrast, all flowering plants, including the basal representative *Amborella trichopoda* (Schoefs and Franck 2003, Soltis et al. 2008), appear to utilize the POR pathway exclusively in light-dependent biosynthesis of chl (Reinbothe and Reinbothe 1996, Armstrong 1998, the Stramenopile Chloroplast Genomics Project Web site: http://chloroplast.ocean.washington.edu/home).

Unfortunately, the distribution of LIPOR and POR modes of chl production is less obvious in several important eukaryotic taxonomic lineages. For example, among gymnosperms (Kusumi et al. 2006, the Stramenopile Chloroplast Genomics Project Web site), the occurrence of both LIPOR and POR pathways varies at the species level. For several algal taxa, the issue of chl-biosynthetic pathway identity is particularly complex (Fong and Archibald 2008, the Stramenopile Chloroplast Genomics Project Web site). Within the rhodophytic algae, LIPOR genes are both present (e.g., *Porphyra yezoensis*) or absent (e.g., *Cyanidium caldarium*). Similarly, cryptomonads may have (*Hemiselmis andersenii*) or lack (*Guillardia theta*) the *chl*L, *chl*N, and *chl*B LIPOR genes (the Stramenopile Chloroplast Genomics Project Web site). Our ongoing stramenopile chloroplast-sequencing studies show LIPOR genes to be present in many algal classes including *Ar. lagunensis* (Pelagophyceae), *Vaucheria litorea* (Xanthophyceae), *Botrydium cystosum* (Xanthophyceae), *Chattonella subsalsa* (Raphidophyceae), *Nereocystis luetkeana* (Phaeophyceae), *Pinguiococcus pyrenoidosus* (Pinguiophyceae), and *Rhizochromulina marina* (Dictyochophyceae), but absent in *A. anophagefferens* (Pelagophyceae), *Odontella sinensis*, *P. tricornutum* and *T. pseudonana* (Bacillariophyceae), as well as *H. akashiwo* (Raphidophyceae).

As seen from the discussion above, the absence/presence distribution of LIPOR genes can be punctuated among members of closely related organisms. The significance of the selection of one biosynthetic pathway over another, as it may affect cell function, remains an open question. The fact that all three LIPOR genes are missing in both the chloroplast and nuclear genomes of *A. anophagefferens,* but are present in *Ar. lagunensis* chloroplasts, suggests that *Ar. lagunensis* may be wholly capable of light-independent chl biosynthesis, but that *A. anophagefferens* is not. This observation would suggest that *Ar. lagunensis* may be better adapted (Shi and Shi 2006, Shui et al. 2009) to extended periods of low or no light—a situation common in the shallow lagoons in which it occurs (Buskey et al. 1997). The presence of LIPOR in *Ar. lagunensis* may account for the ability of this alga to bloom without disruption for many years. By comparison, *A. anophagefferens*, which is dependent on the POR biosynthetic pathway, ceases bloom formation every fall when annual light levels reach a minimum (Gobler et al. 2005).

We would like to thank R. Andersen of Bigelow Laboratories for providing the algal strains. We would additionally like to acknowledge R. Kaul, M. V. Olson, K. A. Phelps, H. Hayden, and L. Benson for bioinformatics and/or technical support; and the two reviewers for helpful comments. *A. anophagefferens* sequence data were produced by the U.S. Department of Energy Joint Genome Institute (http://www.jgi.doe.gov) in collaboration with the user community. This study was supported by Washington Sea Grants NA16RG1044 AM09 and NA040AR170032 to R. A. C.; NSF 0523756 to R. A. C. and G. R.; New York Sea Grant to C. J. G.; and an NIH/NHGRI Genome Training Grant T32 HG00035 award to H. C. O. Finally, we would like to thank B. Green and R. Andersen for suggesting the *Aureoumbra-Aureococcus* collaboration.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–402.

Anderson, D. M., Keafer, B. A., Kulis, D. M., Waters, R. M. & Nuzzi, R. 1993. An immunofluorescent survey of the brown tide chrysophyte *Aureococcus anophagefferens* along the northeast coast of the United States. *J. Plankton Res.* 15:563–80.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., Christoffels, A., et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301–10.

Armstrong, G. 1998. Greening in the dark:light-independent chlorophyll biosynthesis from anoxygenic photosynthetic bacteria to gymnosperms. *J. Photochem. Photobiol. B Biol.* 43: 87–100.

Bailey, J. C. & Andersen, R. A. 1999. Analysis of clonal cultures of the brown tide algae *Aureococcus* and *Aureoumbra* (Pelagophyceae) using 18S rRNA, *rbc*L, and RUBISCO spacer sequences. *J. Phycol.* 35:570–4.

Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–80.

Boivin, R., Richard, M., Beauseigle, D., Bousquet, J. & Bellemare, G. 1996. Phylogenetic inferences from chloroplast *chl*B gene sequences of *Nephrolepis exaltata* (Filicopsida), *Ephedra altissima* (Gnetopsida), and diverse land plants. *Mol. Phylogenet. Evol.* 6:19–29.

Bricelj, V. M., Fisher, N. S., Guckert, J. B. & Chu, F.-L. E. 1989. Lipid composition and nutritional value of the brown tide alga *Aureococcus anophagefferens*. *Coast. Estuar. Stud.* 35:85–100.

Bricelj, V. M., MacQuarrie, S. P. & Schaffner, R. A. 2001. Differential effects of *Aureococcus anophagefferens* isolates (''brown tide'') in unialgal and mixed suspensions on bivalve feeding. *Mar. Biol.* 139:605–15.

Buskey, E. J., Liu, H. B., Collumb, C. & Bersano, J. G. F. 2001. The decline and recovery of a persistent Texas brown tide algal bloom in the Laguna Madre (Texas, USA). *Estuaries* 24: 337–46.

Buskey, E. J., Montagna, P. A., Amos, A. F. & Whitledge, T. E. 1997. Disruption of grazer populations as a contributing factor to the initiation of the Texas brown tide algal bloom. *Limnol. Oceanogr.* 42:1215–22.

Buskey, E. J. & Stockwell, D. A. 1993. Effects of a persistent ''brown tide'' on zooplankton populations in the Laguna Madre of south Texas. *In* Smayda, T. J. & Shimizu, Y. [Eds.] *Toxic Phytoplankton Blooms in the Sea.* Elsevier, Amsterdam, pp. 659–66.

Buskey, E. J., Wysor, B. & Hyatt, C. 1998. The role of hypersalinity in the persistence of the Texas 'brown tide' in the Laguna Madre. *J. Plankton Res.* 20:1553–65.

Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., et al. 2002. The Comparative

RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2.

Caron, D. A., Gobler, C. J., Lonsdale, D. J., Buck, N. J., Cerrato, R. M., Schaffner, R. A., Rose, J. M., Taylor, G. T., Boissonneault, K. R. & Mehran, R. 2004. Microbial herbivory on the brown tide alga, *Aureococcus anophagefferens*: results from natural ecosystems, mesocosms and laboratory experiments. *Harmful Algae* 3:439–57.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. 2005. ACT: the Artemis comparison tool. *Bioinformatics* 21:3422–3.

Cattolico, R. A., Jacobs, M. A., Zhou, Y., Chang, J., Duplessis, M., Lybrand, T., McKay, J., Ong, H. C., Sims, E. & Rocap, G. 2008. Chloroplast genome sequence analysis using a fosmid cloning approach: analysis of *Heterosigma akashiwo* CCMP452 (West Atlantic) and NIES 293 (West Pacific) strains. *BMC Genomics* 9:211.

Chénard, C. & Suttle, C. A. 2008. Phylogenetic diversity of sequences of cyanophage photosynthetic genes *psbA* in marine and freshwaters. *Appl. Environ. Microbiol.* 74:5317–24.

Cosper, E. M., Dennison, W. & Milligan, A. 1989. An examination of environmental factors important to initiating and sustaining ''brown tide'' blooms. *Coast. Estuar. Sci.* 35:317–40.

Coulson, A., Sulston, J., Brenner, S. & Karn, J. 1986. Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U. S. A.* 83:7821–5.

Darling, A. C. E., Mau, B., Blatter, F. R. & Perna, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–403.

Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–9.

Dennison, W. C. 1987. Effects of light on a seagrass photosynthesis, growth and depth distribution. *Aquat. Bot.* 27:15–26.

Deonarine, S. N., Gobler, C. J., Lonsdale, D. J. & Caron, D. A. 2006. The role of zooplankton in the occurrence of harmful brown tide blooms (*Aureococcus anophagefferens*) in US mid-Atlantic estuaries. *Aquat. Microb. Ecol.* 44:181–95.

DeYoe, H. R., Stockwell, D. A., Bidigare, R. R., Latasa, M., Johnson, P. W., Hargraves, P. E. & Suttle, C. A. 1997. Description and characterization of the algal species *Aureoumbra lagunensis* gen. et sp. nov. and referral of *Aureoumbra* and *Aureococcus* to the Pelagophyceae. *J. Phycol.* 33:1042–8.

Ewing, B. & Green, P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* 8: 186–94.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* 8:175–85.

Finn, R. D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., et al. 2006. Pfam: clans, web tools and services. *Nucleic Acids Res.* 34(Database issue):D247–51.

Fong, A. & Archibald, J. M. 2008. Evolutionary dynamics of light-independent protochlorophyllide oxidoreductase genes in the secondary plastids of cryptophyte algae. *Eukaryot. Cell* 7:550–3.

Gallagher, S. R. 2004. Quantitation of DNA and RNA with absorption and fluorescence spectroscopy. *In* Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. [Eds.] *Current Protocols in Molecular Biology*. John Wiley & Sons, New York, pp. A.3D.1–12.

Gastrich, M. D., Bell, J. L., Gobler, C. J., Anderson, O. R. & Wilhelm, S. W. 2004. Viruses as potential regulators of regional brown tide blooms caused by the alga, *Aureococcus anophagefferens:* a comparison of bloom years 1999–2000 and 2002. *Estuaries* 27:112–9.

Gobler, C. J., Deonarine, S., Leigh-Bell, J., Gastrich, M. D., Anderson, O. R. & Wilhelm, S. W. 2004. Ecology of phytoplankton communities dominated by *Aureococcus anophagefferens*: the role of viruses, nutrients, and microzooplankton grazing. *Harmful Algae* 3:471–83.

Gobler, C. J., Lonsdale, D. J. & Boyer, G. L. 2005. A synthesis and review of causes and impact of harmful brown tide blooms caused by the alga, *Aureococcus anophagefferens*. *Estuaries* 28: 726–49.

Gordon, D., Abajian, C. & Green, P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8:195–202.

Gordon, D., Desmarais, C. & Green, P. 2001. Automated finishing with autofinish. *Genome Res.* 11:614–25.

Greenfield, D. I. & Lonsdale, D. J. 2002. Mortality and growth of juvenile hard clams *Mercenaria mercenaria* during brown tide. *Mar. Biol.* 141:1045–50.

Griffiths, W. T., McHugh, T. & Blankenship, R. R. 1996. The light intensity dependence of protochlorophyllide photoconversion and its significance to the catalytic mechanism of protochlorophyllide reductase. *FEBS Lett.* 398:235–8.

Guillard, R. R. L. 1975. Culture of phytoplankton for feeding marine invertebrates. *In* Smith, W. L. & Chanley, M. H. [Eds.] *Culture of Marine Invertebrate Animals*. Plenum Press, New York, pp. 26–60.

Hardison, L., Boczar, B. & Cattolico, R. A. 1995. *psb*A in the marine chromophyte *Heterosigma carterae*: evolutionary analysis and comparative structure of the D1 carboxyl terminus. *Am. J. Bot.* 82:893–902.

Hayden, H. S., Gillett, W., Saenphimmachak, C., Lim, R., Zhou, Y., Jacobs, M. A., Chang, J., et al. 2008. Large-insert genome analysis technology detects structural variation in *Pseudomonas aeruginosa* clinical strains from cystic fibrosis patients. *Genomics* 91:530–7.

Hippler, M., Drepper, F., Haehnel, W. & Rochaix, J.-D. 1998. The N-terminal domain of PsaF: precise recognition site for binding and fast electron transfer from cytochrome C6 and plastocyanin to photosystem I of *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U. S. A.* 95:7339–44.

Ivleva, N., Shestakov, S. & Pakrasi, H. 2000. The carboxy-terminal extension of the precursor D1 protein of photosystem II is required for optimal photosynthetic performance of the cyanobacterium *Synechocystis sp.* PCC 6803. *Plant Physiol.* 12: 1403–11.

Kowallik, K. V., Stroebe, B., Schaffran, I. & Freir, U. 1995. The chloroplast genome of a chlorophyll a+c-containing alga, *Odontella sinensis*. *Plant Mol. Biol. Rep.* 13:336–42.

Kusumi, J., Sato, A. & Tachida, H. 2006. Relaxation of functional constraint on light-independent protochlorophyllide oxidoreductase in *Thuja*. *Mol. Biol. Evol.* 23:941–8.

Kuvikova, S., Tichy, M. & Komenda, J. 2005. A role for the C-terminal extension of the photosystem II D1 protein in sensitivity of the cyanobacterium *Synechocystis* PCC6803 to photoinhibition. *Photochem. Photobiol. Sci.* 4:1044–8.

Lagesen, K., Hallin, P. F., Rødland, E., Stærfeldt, H. H., Rognes, T. & Ussery, D. W. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–8.

Laslett, D. & Canback, B. 2004. ARAGORN, a program to detect tRNA and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32:11–6.

Lillo, F., Basile, S. & Mantegna, R. N. 2002. Comparative genomics study of inverted repeats in bacteria. *Bioinformatics* 18:971–9.

Lohse, M., Drechsel, O. & Bock, R. 2007. OrganellarGenomeDRAW (OGDRAW) – a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52:267–74.

Lonsdale, D. J., Cosper, E. M., Kim, W.-S., Doall, M. H., Divadeenam, A. & Jonasdottir, S. H. 1996. Food web interactions in the plankton of Long Island bays, with preliminary observations on brown tide effects. *Mar. Ecol. Prog. Ser.* 134:247–63.

Lowe, T. M. & Eddy, S. R. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–64.

MacIntyre, H. L., Lomas, M. W., Cornwell, J., Suggett, D. J., Gobler, C. J., Koch, E. W. & Kana, T. M. 2004. Mediation of benthic-pelagic coupling by microphytobenthos: an energy- and material-based model for initiation of blooms of *Aureococcus anophagefferens*. *Harmful Algae* 3:403–37.

Milligan, A. J. & Cosper, E. M. 1997. Growth and photosynthesis of the 'brown tide' microalga *Aureococcus anophagefferens* in subsaturating constant and fluctuating irradiance. *Mar. Ecol. Prog. Ser.* 153:67–75.

Montagna, P. A., Stockwell, D. A. & Kalke, R. D. 1993. Dwarf surf clam *Mulinea lateralis* (Say, 1822) populations and feeding during the Texas brown tide. *J. Shellfish Res.* 12:833–42.

Mulkidjanian, A., Koonin, E., Makarova, K., Sorokin, A., Wolf, Y., Dufresne, A., Partensky, F., et al. 2006. The cyanobacterial genome core and the origin of photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 103:13126–31.

Olsen, P. S. 1989. Development and distribution of a brown-water algal bloom in Barnegat Bay, New Jersey with perspective on resources and other red tides in the region. *Coast. Estuar. Stud.* 35:189–212.

Olson, M. V., Dutchnik, J. E., Graham, M. Y., Brodeur, G. M., Helms, C., Frank, M., McCollin, M., Scheinman, R. & Frank, T. 1986. Random-clone strategy for genomic restriction mapping in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 83:7826–30.

Onuf, C. P. 1996. Seagrass responses to long-term light reduction by brown tide in upper Laguna Madre, Texas: distribution and biomass patterns. *Mar. Ecol. Prog. Ser.* 138:219–31.

Oudot-Le Secq, M. P., Grimwood, J., Shapiro, H., Armbrust, E. V., Bowler, C. & Green, B. R. 2007. Chloroplast genomes of the diatom *Phaeodactylum tricornutum* and *Thalassiosira pseudonana:* comparison with other plastid genomes of the red lineage. *Mol. Genet. Genomics* 277:427–39.

Padilla, D. K., Doall, M. H., Gobler, C. J., Hartson, A. & O'Boyle, K. 2006. Brown tide alga, *Aureococcus anophagefferens,* can affect growth but not survivorship of *Mercenaria mercenaria* larvae. *Harmful Algae* 5:736–48.

Popels, L. C., Coyne, K. J., Forbes, R., Pustizzi, F., Gobler, C. J., Cary, S. C. & Hutchins, D. A. 2003. The use of quantitative polymerase chain reaction for the detection and enumeration of the harmful alga *Aureococcus anophagefferens* in environmental samples along the United States East Coast. *Limnol. Oceanogr. Methods* 1:92–102.

Probyn, T., Pitcher, G., Pienaar, R. & Nuzzi, R. 2001. Brown tides and mariculture in Saldanha Bay, South Africa. *Mar. Poll. Bull.* 42:405–8.

Raymond, J. & Blankenship, R. 2004. The evolutionary development of the protein complement of photosystem 2. *Biochim. Biophys. Acta* 1655:133–9.

Raymond, C. K., Subramanian, S., Paddock, M., Qiu, R., Deodato, C., Palmieri, A., Chang, J., et al. 2005. Targeted, haplotype-resolved resequencing of long segments of the human genome. *Genomics* 86:759–66.

Regalia, M., Rosenblad, M. A. & Samuelsson, T. 2002. Prediction of signal recognition particle RNA genes. *Nucleic Acids Res.* 30:3368–77.

Reinbothe, S. & Reinbothe, C. 1996. Regulation of chlorophyll biosynthesis in angiosperms. *Plant Physiol.* 111:1–7.

Rice, P., Longden, I. & Bleasby, A. 2000. The European molecular biology open software suite. *Trends Genet.* 16:276–7.

Rogers, M. B., Gilson, P. R., Su, V., McFadden, G. I. & Keeling, P. J. 2007. The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol. Biol. Evol.* 24:54–62.

Roose, J. L. & Pakrasi, H. B. 2004. Evidence that D1 processing is required for manganese binding and extrinsic protein assembly into photosystem II. *J. Biol. Chem.* 29:45417–22.

Rumpho, M. E., Worful, J. M., Lee, J., Kannan, K., Tyler, M. S., Bhattacharya, D., Moustafa, A. & Manhart, J. R. 2008. Horizontal gene transfer of the algal nuclear gene *psbO* to the photosynthetic sea slug *Elysia chlorotica. Proc. Natl. Acad. Sci. U. S. A.* 105:17867–71.

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Wu, D., et al. 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 5:e77.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–5.

Satoh, K. & Yamamoto, Y. 2007. The carboxyl-terminal processing of precursor D1 protein of the photosystem II reaction center. *Photosynth. Res.* 94:203–15.

Schoefs, B. & Franck, F. 2003. Protochlorophyllide reduction: mechanisms and evolution. *Photochem. Photobiol.* 78:543–57.

Senge, M. O. 1993. Recent advances in the biosynthesis and chemistry of the chorophylls. *Photochem. Photobiol.* 57:189–206.

Shi, C. & Shi, X. 2006. Expression switching three genes encoding light-independent protochlorophyllideoxidoreductase in *Chlorella protothecoides. Biotechnol. Lett.* 28:261–5.

Shui, J., Saunders, E., Needleman, R., Nappi, M., Cooper, J., Hall, L., Kehoe, D. & Stowe-Evans, E. 2009. Light-dependent and light-independent protochlorophyllide oxidoreductases in the chromatically adapting cyanobacterium *Fremyella diplosiphon* UTEX 481. *Plant Cell Physiol.* 50:1507–21.

Sieburth, J. M., Johnson, P. W. & Hargraves, P. E. 1988. Ultrastructure and ecology of *Aureococcus anophagefferens* gen. et sp. nov. (Chrysophyceae): the dominant picoplankter during a bloom in Narragansett Bay, Rhode Island, summer 1985. *J. Phycol.* 24:416–25.

Smeekens, S., Weisbeek, P. & Robinson, C. 1990. Protein transport into and within chloroplasts. *Trends Biochem. Sci.* 15:73–6.

Soltis, D., Albert, V., Leebens-Mack, J., Palmers, J., Wing, R., dePamphlilis, C., Ma, H., et al. 2008. The *Amborella* genome: an evolutionary reference for plant biology. *Genome Biol.* 9:402.

Trice, T. M., Glibert, P. M., Lea, C. & Van Heukelem, L. 2004. HPLC pigment records provide evidence of past blooms of *Aureococcus anophagefferens* in the Coastal Bays of Maryland and Virginia, USA. *Harmful Algae* 3:295–304.

Villareal, T. A., Chirichella, T. & Buskey, E. J. 2004. Regional distribution of the Texas brown tide (*Aureoumbra lagunensis*) in the Gulf of Mexico. *In* Steidinger, K. A., Landsberg, J. H., Tomas, C. R. & Vargo, G. A. [Eds.] *Harmful Algae 2002.* Florida Fish and Wildlife Conservation Commission, Florida Institute Oceanography, IOC-UNESCO, St. Petersburg, Florida, pp. 374–6.

Ward, L. A., Montagna, P. A., Kalke, R. D. & Buskey, E. J. 2000. Sublethal effects of Texas brown tide on *Streblospio benedicti* (Polychaeta) larvae. *J. Exp. Mar. Biol. Ecol.* 248:121–9.

Wong, G. K.-S., Yu, J., Thayer, E. C. & Olson, M. V. 1997. Multiple-complete-digest restriction fragment mapping: generating sequence-ready maps for large-scale DNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 94:5225–30.

Yamamoto, Y., Inagaki, N. & Satoh, K. 2001. Overexpression and characterization of carboxyl-terminal processing protease for precursor D1 protein: regulation of enzyme-substrate interaction by molecular environments. *J. Biol. Chem.* 276:7518–25.

Yentsch, C. S., Phinney, D. A. & Shapiro, L. P. 1989. Absorption and fluorescent characteristics of the brown tide chrysophyte: its role on light reduction in coastal marine environments. *Coast. Estuar. Stud.* 35:77–83.

Yoon, H. S., Nakayama, T., Reyes-Prieto, A., Andersent, R., Boo, S., Ishida, K. & Bhattacharya, D. 2009. A single origin of the photosynthetic organelle in different *Paulinella* lineages. *BMC Evol. Biol.* 9:98.

Supplementary Material

The following supplementary material is available for this article:

Figure S1. De novo contig building and tiling of *Aureoumbra lagunensis* chloroplast fosmid clones. (A) Eighteen clones were subjected to multiple complete digest (MCD) analysis, and the overlap of the resulting fingerprints generated a preliminary contig of 12 clones. The top candidate clone (by bioinformatic analysis of their end sequences) was sequenced as the ''anchor'' clone, and four additional clones (numbered 2–5) were selected for sequencing. (B) The complete sequences of the five clones were aligned into a 94 Kbp sequence contig, and tiling of end-sequenced and MCD data from all other clones (only nine clones aligned once the sequence-based contig was used) revealed a sixth clone (labeled ''fin.''), which was partially sequenced using custom primers to complete the genome sequence.

Figure S2. A larger and more detailed version of Figure 4 with gene names listed to the right of the heat map.

Table S1. Gene presence (black boxes) and absence (white boxes) among 25 genomes analyzed in Figure 4. Genes and taxa are listed alphabetically for *Aa* (*Aureococcus anophagefferens*), *Al* (*Aureoumbra lagunensis*), *Bn* (*Bigelowiella natans*), *Cr* (*Chlamydomonas reinhardtii*), *Cm* (*Cyanodioschyzon merolae*), *Cc* (*Cyanidium caldarium*), *Eh* (*Emiliania huxleyi*), *Eg* (*Euglena gracilis*), *Gte* (*Gracilaria tenuistipitata*), *Gth* (*Guillardia theta*), *Ha90* (*Heterosigma akashiwo* NIES 293), *Ha91* (*Heterosigma akashiwo* CCMP452), *Lt* (*Leptosira terrestris*), *Ms* (*Micromonas* sp.), *Os* (*Odontella sinensis*), *Ot* (*Ostreococcus tauri*), *Pt* (*Phaeodactylum tricornutum*), *Ppy* (*Pinguiococcus pyrenoidosus*), *Pp* (*Porphyra purpurea*), *Py* (*Porphyra yezoensis*), *Rs* (*Rhodomonas salina*), *So* (*Scenedesmus obliquus*), *Sh* (*Stigeoclonium helveticum*), *Tp* (*Thalassiosira pseudonana*), and *Vt* (*Vaucheria litorea*).

This material is available as part of the online article.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.