Proteomics Reveals Plastid- and Periplastid-Targeted Proteins in the Chlorarachniophyte Alga *Bigelowiella natans*

Julia F. Hopkins[1], David F. Spencer[1], Sylvie Laboissiere[2], Jonathan A.D. Neilson[3], Robert J.M. Eveleigh[1], Dion G. Durnford[3], Michael W. Gray[1], and John M. Archibald[1,*]

[1]Department of Biochemistry and Molecular Biology, Dalhousie University, Nova Scotia, Canada

[2]Proteomics Unit, McGill University and Génome Québec Innovation Centre, Quebec, Canada

[3]Department of Biology, University of New Brunswick, New Brunswick, Canada

*Corresponding author: E-mail: jmarchib@dal.ca; john.archibald@dal.ca.

Accepted: November 29, 2012

Abstract

Chlorarachniophytes are unicellular marine algae with plastids (chloroplasts) of secondary endosymbiotic origin. Chlorarachniophyte cells retain the remnant nucleus (nucleomorph) and cytoplasm (periplastidial compartment, PPC) of the green algal endosymbiont from which their plastid was derived. To characterize the diversity of nucleus-encoded proteins targeted to the chlorarachniophyte plastid, nucleomorph, and PPC, we isolated plastid–nucleomorph complexes from the model chlorarachniophyte *Bigelowiella natans* and subjected them to high-pressure liquid chromatography-tandem mass spectrometry. Our proteomic analysis, the first of its kind for a nucleomorph-bearing alga, resulted in the identification of 324 proteins with 95% confidence. Approximately 50% of these proteins have predicted bipartite leader sequences at their amino termini. Nucleus-encoded proteins make up >90% of the proteins identified. With respect to biological function, plastid-localized light-harvesting proteins were well represented, as were proteins involved in chlorophyll biosynthesis. Phylogenetic analyses revealed that many, but by no means all, of the proteins identified in our proteomic screen are of apparent green algal ancestry, consistent with the inferred evolutionary origin of the plastid and nucleomorph in chlorarachniophytes.

Key words: chlorarachniophyte algae, plastids, nucleomorphs, proteome, mass spectrometry, evolution.

Introduction

More than a billion years ago, the engulfment and retention of a cyanobacterium by a eukaryotic host cell resulted in the evolution of plastids (chloroplasts), the light-harvesting organelles of algae and plants (Reyes-Prieto et al. 2007). The "primary" plastids of red and green algae subsequently spread across the eukaryotic tree by "secondary" endosymbiosis, that is, the assimilation of an alga by an unrelated eukaryote (Reyes-Prieto et al. 2007). Of known secondary plastid-bearing algae, two lineages, the cryptophytes and chlorarachniophytes, are atypical in that the remnant nucleus of the algal endosymbiont persists in a miniaturized form: the "nucleomorph" (Archibald 2007; Moore and Archibald 2009). These organisms are among the most complex eukaryotic cells investigated thus far, with four genomes (nuclear and nucleomorph as well as mitochondrial and plastid) and two distinct cytosolic compartments in which core cellular processes occur.

Nucleomorph genomes are the smallest nuclear genomes known. The nucleomorph genome of the chlorarachniophyte *Bigelowiella natans* (fig. 1*A*) is a mere 380 kilobases (kb) in size (Gilson et al. 2006) and encodes 284 proteins primarily involved in "housekeeping" functions (e.g., transcription, translation, and protein degradation). Genes encoding important—and assumed to be essential—proteins such as DNA polymerases, proteasome components, and many ribosomal proteins are apparently absent. Interestingly, the *B. natans* nucleomorph genome has only 17 genes known to encode plastid-targeted proteins (Gilson et al. 2006). The remaining nucleomorph-encoded proteins with predictable functions are thought to work in the nucleomorph itself or within the remnant cytosol of the endosymbiont, referred to as the periplastidial compartment (PPC) (fig. 1*A* and *B*). It is not known how many nucleomorph genes have been lost during evolution or have migrated to the host nuclear genome by endosymbiotic gene transfer (Timmis et al. 2004).

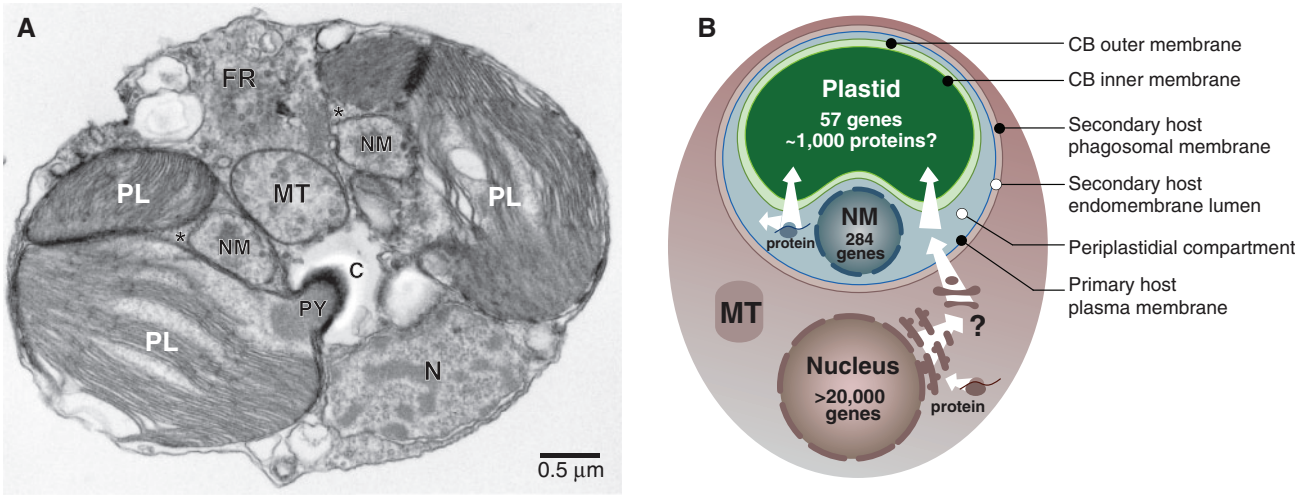

Fig. 1.—The chlorarachniophyte alga *Bigelowiella natans*. (*A*) Transmission electron micrograph showing ultrastructural features of *B. natans*. Asterisks (*) highlight the PPC, that is, the residual cytosol of the engulfed algal cell. (*B*) Simplified diagram showing main subcellular compartments and the protein synthesis/trafficking in chlorarachniophytes relevant to this study (highlighted by cartoon ribosomes and white arrows, respectively). The "?" indicates uncertainty with respect to protein trafficking from the ER to the outermost plastid membrane (see text). The numbers of protein genes present in the plastid, nucleomorph, and host nucleus genomes are taken from Rogers et al. (2007), Gilson et al. (2006), and Curtis et al. (2012), respectively. Details surrounding the mitochondrion have been omitted for simplicity. PL, plastid; N, nucleus; NM, nucleomorph; MT, mitochondrion; PY, pyrenoid; C, pyrenoid cap; FR, flagellar root; CB, cyanobacterium derived.

Because of the extremely limited coding capacity of nucleomorph genomes in both cryptophytes and chlorarachniophytes (Douglas et al. 2001; Gilson et al. 2006; Lane et al. 2007; Moore and Archibald 2009; Tanifuji et al. 2011), most of the proteins needed for proper functioning of the plastid, nucleomorph and PPC must be synthesized in the host cytoplasm and imported co- or posttranslationally. Genome sequence data make it possible to predict proteins that are targeted to the plastid, but current in silico prediction methods are limited in their utility, particularly when applied to organisms that are distantly related to the species whose protein sequences have been used to "train" the prediction algorithms (Foth et al. 2003; Patron and Waller 2007). Because of the multiple locations to which proteins must be targeted within a secondary plastid-bearing organism—that is, the PPC, the plastid and, in the case of cryptophytes and chlorarachniophytes, the nucleomorph—a complex protein-targeting system has evolved. To be transported across the four membranes surrounding the *B. natans* plastid (two cyanobacterium-derived plastid membranes, the former plasma membrane of the engulfed alga [i.e., the primary host], and the outermost host-derived membrane) (Rogers et al. 2004; Archibald 2009), nucleus-encoded proteins must have a bipartite leader sequence. This sequence is composed of a signal peptide (SP) and a transit peptide (TP)-like sequence, the biochemical characteristics of which determine the ultimate destination of the protein within the plastid complex (McFadden 1999; Gould et al. 2006; Patron and Waller 2007). How the PPC/plastid preproteins are transported to the outermost membrane of the chlorarachniophyte plastid is not well understood. It is believed that they are brought to the membrane via endoplasmic reticulum (ER)-derived vesicles (Hempel et al. 2007; Hirakawa et al. 2012). Once inside the PPC, another unknown mechanism selectively transports preproteins destined for the plastid to the Toc–Tic (*t*ranslocon on the *o*uter/*i*nner envelope membrane of *c*hloroplasts) import apparatus (Nassoury and Morse 2005). It is not yet clear how plastid preproteins are distinguished from PPC or NM preproteins, because such proteins also possess TP-like sequences (Hirakawa et al. 2012).

Toward the goal of characterizing the full spectrum of nucleus-encoded proteins targeted to the plastid, PPC and nucleomorph of the model chlorarachniophyte *B. natans*, and inferring the evolutionary origin(s) of the genes that encode them, we performed a proteomics investigation of purified plastid–nucleomorph (PL–NM) complexes, the first of its kind for a nucleomorph-containing alga. Specifically, we isolated PL–NM complexes using sucrose gradient ultracentrifugation and performed high-pressure liquid chromatography-tandem mass spectrometry (HPLC-MS/MS). The resulting peptides were compared with proteomes inferred from all four of the completely sequenced *B. natans* genomes. More than 300 proteins were identified, corresponding to a mix of plastid, nucleomorph, and nuclear gene products. The putative functions and cellular locations of the nucleus-encoded proteins are diverse in nature, as are their evolutionary histories.

Materials and Methods

Cell Culturing, Subcellular Fractionation, and Organelle Enrichment

Bigelowiella natans CCMP 2755 was grown in stirred 1 l cultures of f/2 medium at room temperature with a 12 h light/12 h dark cycle to an average density of 2.8×10^6 cells/ml. Cells were harvested, centrifuged for 15 min at 3,000 × g, then resuspended in 4 ml of isolation buffer (IB: 600 mM sorbitol; 5 mM EDTA; 5 mM $MgCl_2$; 10 mM KCl; 1 mM $MnCl_2$; 50 mM HEPES pH 8.0) (modified from Wittpoth et al. 1998) and 500 µl glycerol and then frozen. Cells were thawed and lysed using a French press at 1,500–2,000 psi. A Teflon homogenizer was used to break up large clumps in the cell homogenate before layering it on a three-step sucrose gradient. The gradient was composed of 3.5 ml each of 1.55, 1.35, and 1.05 M sucrose in IB. Gradients were centrifuged at 60,000 × g for 45 min in a Beckman L7-55 ultracentrifuge. Discrete fractions were collected and concentrated in IB and stored at −20°C in IB plus 20% glycerol.

DAPI Staining and Light Microscopy

Cells and cell fractions were fixed with 1% acetate-buffered formaldehyde overnight at 4°C. Aliquots (1 ml) of the fixed cells/cell fractions were collected on 0.2 µm Isopore membrane filters (Millipore). Filters were then washed with 1 ml of 70% ethanol, followed by an additional 100% ethanol wash. The filters were then air dried and stained with 2.5 µg/ml 4′,6-diamidino-2-phenylindole (DAPI) solution for approximately 5 min in the dark, washed with water for 5 min and then with 80% ethanol for 2 min. Filter sections were mounted on a microscope slide using AF1 (Citifluor Ltd.) mounting medium and viewed in a Zeiss Axiovert 200 M microscope (Germany). The overlay images were produced using the DAPI channel to detect the cell nuclei and nucleomorphs while the autofluorescence of the chlorophyll was detected using the Rhodamine channel.

Transmission Electron Microscopy

Whole cells were fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer. Isolated PL–NM samples were fixed in 2.5% glutaraldehyde in IB. Whole cell and PL–NM samples were then fixed for 2 h in 1% osmium tetroxide and placed in 0.25% uranyl acetate at 4°C overnight. The samples were embedded in Epon Araldite Resin, and sections were cut with an LKB Huxley Ultramicrotome and stained with 2% aqueous uranyl acetate and lead citrate. The samples were viewed using a JEOL JEM 1230 Transmission Electron Microscope at 80 kV and images were captured with a Hamamatsu ORCA-HR digital camera.

Proteomics

Bigelowiella natans subcellular fractions isolated from sucrose gradients (specifically those enriched in PL–NM complexes) were prepared for proteomic analysis by diluting 1:1 with 2X Laemmli buffer and heated at 95°C for 5 min. The PL-NM fraction was analyzed at the McGill University and Génome Québec Innovation Centre. The sample was fractionated on a 2.4 cm 1D sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (7–15% acrylamide gradient) using a Hoefer SE 600 Rubys system (GE Health Care). After electrophoresis, the gel was stained with Coomassie Brilliant Blue G (Sigma). The full lane was then subjected to automated band excision using a Protein Picking Workstation ProXCISION (Perkin Elmer). Fourteen bands were excised from the gel. Proteins were digested with trypsin (sequencing-grade modified trypsin; Promega) in a MassPrep Workstation (Micromass, Manchester, UK) as previously described (Wasiak et al. 2002).

Peptides were subjected to mass spectrometry analysis as follows. Extracted peptides were analyzed using a Nanopump series 1000 (Agilent Technologies) coupled to a Q-Tof Micro mass spectrometer (Waters Micromass) equipped with a Nanosource modified with a nanospray adapter (New Objective) to hold the PicoFrit column spraying tip near the sampling cone. Peptide solution (20 µL) was injected at 15 µL/min using an isocratic pump and trapped on a precolumn (Zorbax 300SB-C18, 5 mm × 0.3 mm, 5 µm). Samples were washed with aqueous solution (3% acetonitrile, 0.1% formic acid) before connecting the precolumn to the PicoFrit analytical column (75 µm i.d. [New Objective], which was filled with BioBasic C18 packing [10-cm bed length, 5 µm, 300 Å]). Solvent A was 0.1% formic acid in water and solvent B was acetonitrile:water:formic acid (97:3:0.1). The linear gradient was started after the washing step. At time 0 min, solvent B was 10.5%. The instrument was set to reach 42% B at 45 min, 73.5% B at 53 min, and 100% B at 58 min. Solvent B was then kept at 100% for 2 min and brought back to 10.5% between 60 and 72 min.

During the MS run, the capillary voltage was adjusted to obtain the best spraying plume at 35% solvent B. The MS Survey scan was set to 1 s (0.1 s interscan) and recorded from 350 to 1,600 *m/z*. In a given MS Survey scan, all doubly and triply charged ions with intensities higher than 25 counts were considered candidates for MS/MS fragmentation. From these, the strongest were selected. MS/MS acquisition stopped as soon as the total ion current reached 2,800 counts/scan or after a maximum time of 4 s. The MS/MS scan was acquired between 50 and 1,990 *m/z*, the scan time was 1.35 s and the interscan was 0.15 s. The next precursor ion was selected from the subsequent MS Survey scan. The doubly and triply charged selected ions were fragmented with the following preprogrammed collision energies: 1) For doubly charged ions, the collision energies were 25 eV for

400–653 *m/z*, 26 eV for 653–740 *m/z*, 28 eV for 740–820 *m/z*, 32 eV for 820–1,200 *m/z*, and 55 eV for 1,200–1,600 *m/z*. 2) For triply charged ions, collision energies were 14 eV for 435–547 *m/z*, 19 eV for 547–605 *m/z*, 24 eV for 605–950 *m/z*, and 35 eV for *m/z* >950 *m/z*.

MS data were acquired using the data-directed analysis (DDA) feature of the MassLynx operating software (Micromass) with a 1,1,4 duty cycle (1 s in MS Survey mode, 1 peptide selected for fragmentation, maximum of 4 s in MS/MS acquisition mode). Analyses were done in two steps. For the first step, data acquisition involved allowing the MassLynx software to select the most intense precursor from the MS Survey scan. From the data for each injected sample, a list of precursor ions with their chromatographic retention times was then prepared and used as an "exclusion list" for a second analysis. Each sample was injected again and DDA was repeated, using the information contained in the exclusion list. This process enabled the mass spectrometer to collect data from less intense precursors that were not acquired during the first analysis. DDA settings for the inclusion set runs were identical to those described above, except that the precursor *m/z* was excluded within ±1,900 mDa of the entries on the exclusion list. The *m/z* was also chromatographically excluded for a time window of ±75 s of the retention time registered during the first analysis.

Peaklists were generated from the MS/MS raw data using Distiller software (http://www.matrixscience.com/distiller.html, last accessed December 13, 2012) with peak-picking parameters set at 5 for signal noise ratio and 0.4 for Correlation Threshold (CT). The peak-listed data were then searched against a database containing predicted protein sequences from the *B. natans* nuclear, mitochondrial, plastid (Rogers et al. 2007), and nucleomorph (Gilson et al. 2006) genomes using Mascot (http://www.matrixscience.com, last accessed December 13, 2012) and X! Tandem (http://www.thegpm.org, last accessed December 13, 2012), restricting the search to allow up to one missed (trypsin) cleavage, fixed carbamidomethyl alkylation of cysteines, variable oxidation of methionine, a 0.5 Da mass unit tolerance on parent and fragment ions, and monoisotopic.

Scaffold (Proteome Software Inc.) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm (Keller et al. 2002). Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al. 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

Data Analysis

Proteins identified by mass spectrometry were subjected to BLASTP analysis (Altschul et al. 1990) and the first 200 amino acids of the N-termini of all identified proteins were analyzed for the presence of signal and TPs by SignalP version 3.0 (Bendtsen et al. 2004), iPSORT (Bannai et al. 2002), ChloroP 1.1 (Emanuelsson et al. 1999, 2000, 2007), and PredSL (Petsalaki et al. 2006). Sequence similarity to proteins known to function in a particular subcellular compartment in other organisms (in particular the plastid) was also taken into consideration. Protein ID numbers refer to proteins in the *B. natans* gene catalog (http://genome.jgi-psf.org/Bigna1/Bigna1.home.html, last accessed December 13, 2012) as described by Curtis et al. (2012). Proteins were also analyzed using the Pfam database to identify the various protein domains.

Phylogenetic Analyses

The nucleus-encoded proteins identified by HPLC-MS/MS in this study were matched against 5002 maximum likelihood protein trees for *B. natans* generated by Curtis et al. (2012). Trees were manually sorted into four categories: red algal or green algal (with >75% bootstrap support), ambiguous (if support was weak but was nonetheless consistent with a red or green origin), or "other" (if too few taxa were present or the phylogenetic groupings were unclear).

Additional maximum likelihood RAxML (Stamatakis et al. 2008) phylogenetic trees were produced for *B. natans* proteins with the following IDs: 236790; 91954; 86287; and 50246. Expanded multiple sequence alignments were generated as necessary using the MEGA5 program (Tamura et al. 2011) and MUSCLE (Edgar 2004) after retrieval of additional sequences from the database used by Curtis et al. (2012) and from the NCBI GenBank nr (nonredundant) database. Masked alignments were produced by trimALv1.3 (Capella-Gutiérrez et al. 2009) using the web resource Phylemon 2.0 (Sánchez et al. 2011), with the exception of 91954 and 50246, which were done by hand.

Results and Discussion

Organelle Enrichment and Proteomic Analysis

The plastid–nucleomorph (PL–NM) complex of *B. natans* is an intricate multimembrane structure (fig. 1). To isolate the plastid and nucleomorph, along with the contents of the PPC, we attempted to disrupt the outer cell membrane while leaving the two membranes surrounding the PL–NM complex intact. The most effective approach involved passing cells through a French press twice at a pressure of 1,500–2,000 psi, layering the resulting material on a three-step sucrose gradient, and performing high-speed centrifugation. Standard light microscopy, DAPI staining/fluorescence microscopy, and transmission electron microscopy (TEM) were used to monitor the nature of the disrupted cellular material before and after centrifugation. A typical gradient with three distinct bands is shown in figure 2*A*. Intact cells were found to pellet to the

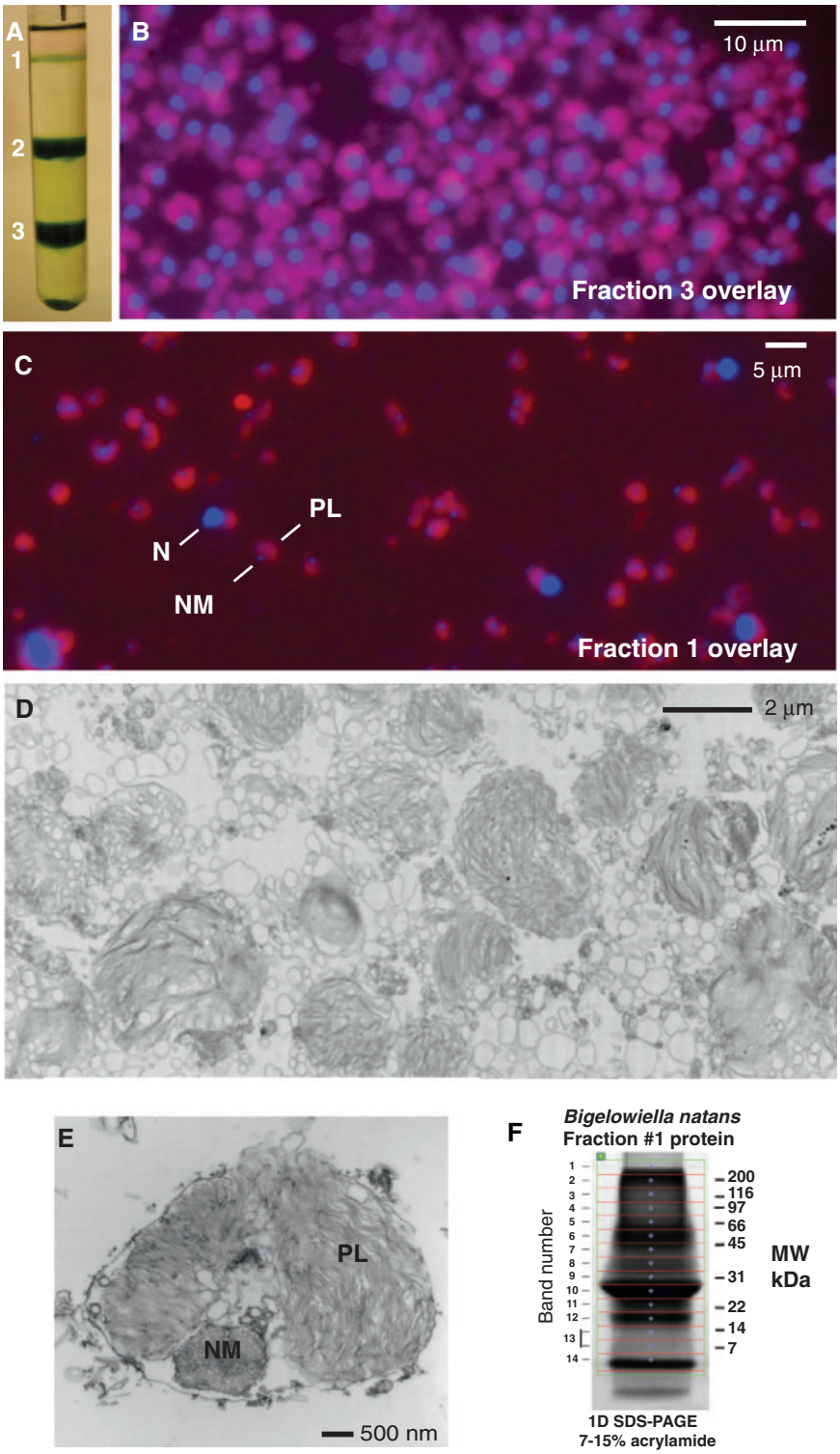

Fig. 2.—*Bigelowiella natans* subcellular fractionation and protein isolation. (*A*) Three-step sucrose density gradient of French press-disrupted *B. natans* cells. Three distinct bands are apparent. (*B*) Fluorescence microscopy of material present in "fraction 3." The image is an overlay of two different channels showing DAPI-stained nuclei (blue) and chlorophyll autofluorescence from the plastids under a rhodamine filter (red). (*C*) Fluorescence microscopy showing material in fraction 1, which was enriched in plastid–nucleomorph complexes and largely devoid of host cell nuclei. Faint blue spots corresponding to nucleomorphs (NM) can be seen in close association with plastids (PL). A large blue spot corresponding to a nucleus (N) can also be seen. (*D*) Transmission electron micrograph of isolated *B. natans* plastids obtained from fraction 1 of the sucrose gradients after subcellular fractionation. (*E*) TEM of an isolated plastid–nucleomorph complex from *B. natans*. (*F*) One-dimensional SDS-PAGE of *B. natans* proteins isolated from fraction 1. Fourteen discrete bands were excised from the gel, as shown, for proteomic analysis (refer to text for further details).

bottom of the gradient (data not shown) and "fraction 3" also comprised mainly intact cells (fig. 2*B*). In contrast, fractions 1 and 2 were enriched in subcellular material. Fraction 1, a faint band at the interface between the buffer and the first layer of the sucrose gradient, was found to contain the least amount of host nucleus contamination as compared with fraction 2. Fluorescence microscopy (fig. 2*C* and supplementary fig. S1, Supplementary Material online) and TEM (figs. 2*D* and *E*) showed fraction 1 to be composed mainly of bean-shaped autofluorescent plastids, many of which cupped a faintly staining (by DAPI) structure, presumably the nucleomorph. The relative abundance of PL–NM complexes and nucleomorph-free plastids in fraction 1 could not be determined.

Fraction 1 proteins were separated into 14 subfractions by 1D SDS-PAGE (fig. 2*F*), digested with trypsin and the resulting peptides subjected to HPLC-MS/MS. The resulting data were analyzed against a total of four different reference proteomes: 1) conceptual translations of expressed sequence tags (ESTs) generated by Sanger sequencing; 2) an initial set of proteins predicted from the *B. natans* nuclear genome sequence using the ab initio gene prediction tool Augustus (Stanke and Waack 2003); 3) a more refined set of 21,708 proteins based on multiple gene-finding programs used by the JGI annotation pipeline (http://genome.jgi-psf.org/Bigna1/Bigna1.home.html, last accessed December 13, 2012); and 4) proteins inferred from Illumina "RNA-seq" transcriptome data generated from total *B. natans* RNA (proteins encoded by the mitochondrial, plastid and nucleomorph genomes were included in each case). Because of differences among these four "whole-cell" reference proteomes, some proteins were confidently identified in our proteomic data using one data set but not the others. Combining the results of these four analyses, mass spectrometry identified a total of 324 proteins with >95% confidence (supplementary tables S1 and S2, Supplementary Material online). Of the 324 proteins, 18 were identified as being encoded by the plastid genome, three were nucleomorph-encoded, and one was encoded by the *B. natans* mitochondrial genome. The remaining 302 proteins (93%) were nucleus encoded.

Plastid-, Nucleomorph-, and Mitochondrion-Encoded Proteins

HPLC-MS/MS analysis identified 18 of 57 proteins encoded in the *B. natans* plastid genome (Rogers et al. 2007) (supplementary table S2, Supplementary Material online). Only two of these 18 proteins, tufA (elongation factor Tu) and chlI (Mg-protoporphyrin IX, involved in chlorophyll biosynthesis), are not directly implicated in photosynthesis. Of the three nucleomorph-encoded proteins identified, two are plastid targeted: the molecular chaperone DnaK and an ATP-binding subunit of Clp protease, ClpC (the *B. natans* nucleomorph genome encodes a total of 17 plastid-targeted proteins [Gilson et al. 2006]). The third nucleomorph-encoded protein, a cytosol-type HSP70, is predicted to be PPC localized. The sole mitochondrion-encoded protein identified, the F1 alpha subunit of ATP synthase, presumably represents a contaminant (discussed later). The retrieval of very few mitochondrial proteins (either mitochondrion or nucleus encoded) is consistent with a high degree of purity of our starting material.

Signal and TP Predictions

Nucleus-encoded preproteins destined for the *B. natans* plastid, nucleomorph, and PPC are thought to require a bipartite topogenic signal composed of an SP and a TP (McFadden 1999; Gould et al. 2006; Patron and Waller 2007). The N-termini of the proteins identified by HPLC-MS/MS analysis were examined for the presence of such sequences using SignalP version 3.0, iPSORT, ChloroP 1.1, and PredSL (Emanuelsson et al. 1999, 2000; Bannai et al. 2002; Bendtsen et al. 2004; Petsalaki et al. 2006). Relatively little is known about the TP-like sequences in *B. natans* and at present it is unclear precisely how PPC proteins are distinguished from plastid or nucleomorph proteins as they are imported from the cytoplasm (Hirakawa et al. 2009). For this reason, any TP-containing proteins that also contained an SP were retained for further consideration, including those with predicted mitochondrial targeting peptides. If a given protein had a predicted plastid or mitochondrial targeting peptide that was preceded by an SP, it was included in our initial list of potential PL–NM–PPC proteins, because mitochondrion-targeted preproteins in *B. natans* do not (to our knowledge) require an SP. Of the 302 nucleus-encoded proteins identified, 153 (51%) were found to have a predicted signal and TP-like sequence at their N-terminus. An additional 18 (6%) were predicted to have a TP (9 proteins) or an SP (9 proteins) but not both (supplementary table S1, Supplementary Material online). The large number of proteins with signal sequences is consistent with what we would predict for a proteomic analysis of a highly enriched subcellular fraction, that is, the majority of these proteins being targeted to the nucleomorph, plastid, or PPC. However, our results also speak to the possibility of alternate targeting pathways to these compartments (discussed later) (Nassoury and Morse 2005; Hempel et al. 2009).

Nucleus-Encoded Cytosolic and Mitochondrial Proteins

Bigelowiella natans possesses two distinct cytosolic compartments, corresponding to the "host" and "endosymbiont" components of the cell (fig. 1) (McFadden et al. 1994). The biochemical complexity of the endosymbiont-derived cytosol, the PPC, is unknown. We attempted to determine whether the 302 nucleus-encoded proteins (supplementary table S1, Supplementary Material online) we identified are likely to represent bona fide PPC-targeted proteins or host cytoplasmic contaminants. Six of the eight known protein subunits of the cytosolic chaperonin complex TCP/CCT (Archibald et al.

2000; Valpuesta et al. 2002) were identified, none of which possessed obvious signal and/or TPs; moreover, four of the six identified subunits have an obvious orthologous counterpart encoded in the *B. natans* nucleomorph genome (Gilson et al. 2006). Twelve cytosol-type ribosomal proteins were also identified, as were several 14-3-3 family proteins. None of these proteins has detectable SP and/or TP motifs, suggesting that they represent cytosolic contaminants. However, it is interesting that in plants 14-3-3 proteins function together with HSP70 to act as a guidance complex for plastid preproteins, targeting them to the Tic/Toc protein import apparatus (May and Soll 2000). Within the *B. natans* PPC, the precise biochemical mechanisms involved in targeting preproteins to the plastid are unknown, but it is conceivable that at least some of the 14-3-3 proteins identified in our screen could be involved. As mentioned earlier, only one mitochondrial genome-encoded protein (ATP synthase alpha subunit) was recovered in our proteomic screen. Two additional putative mitochondrial proteins were found amongst the nucleus-encoded proteins, fumarate hydratase and isocitrate dehydrogenase, the latter having a predicted mitochondrial TP (supplementary table S1, Supplementary Material online). In addition, two potential ER proteins were identified, including the evolutionarily conserved ER isoform of heat shock protein 90.

There are several possible explanations for the existence in our data set of nucleus-encoded proteins for which N-terminal topogenic signals were not identified. First, although each of the 302 *B. natans* gene models was manually curated (where possible, taking into account transcriptomic data), it is possible that in some cases the correct N-terminus was not identified and analyzed. Second, the search tools employed herein may be insufficiently well trained to reliably detect topogenic signals in a poorly studied organism such as *B. natans*. Third, we cannot exclude the possibility that one or more alternate targeting pathways exist in this organism, ones that do not require canonical N-terminal SPs and TPs. Finally, such proteins might represent contaminants from other subcellular compartments, most notably the cytoplasm and mitochondrion. Similar levels of experimental contamination have been seen in organellar proteome studies of other organisms such as the green alga *Ostreococcus* (Le Bihan et al. 2011) and the ciliate *Tetrahymena* (Smith et al. 2007). We return to this issue in a later section.

Functional Categories of Nucleus-Encoded Proteins

The nucleus-encoded proteins identified in our proteomic screen were divided into the following functional categories for further analysis: photosynthesis related, chlorophyll biosynthesis, "plastid miscellaneous," cytoskeleton, transcription/translation, posttranslational modification/protein turnover/chaperones, metabolism, "transport proteins," signal transduction, miscellaneous, and "unknown" (table 1 and supplementary table S1, Supplementary Material online). The proteins classified as "unknown" show only weak sequence similarity to proteins in the NCBI database or share significant similarity with proteins of unknown function in other organisms.

Table 1

Fraction of Identified Nucleus-Encoded *Bigelowiella natans* Proteins with SPs and TPs

	No. of Proteins	SP^+/TP^+	Total (%)
Photosynthesis related	32	31	97
Chlorophyll biosynthesis	10	10	100
Plastid miscellaneous	6	6	100
Transport proteins	8	5	62
Cytoskeleton	19	0	0
Transcription/translation	34	7	21
Posttranslational modification	49	10	20
Metabolism	48	27	56
Intracellular transport	5	0	0
Signal transduction	12	2	17
Miscellaneous	5	1	20
Unknown	74	54	73
Total	*302*	*153*	*51*

Photosynthesis Related

Of 32 nucleus-encoded proteins (out of 302) predicted to be involved in photosynthesis, 31 have predicted SPs and TPs (table 1). These include the light-harvesting proteins, photosystem proteins, and components of plastid ATP synthase (supplementary table S1, Supplementary Material online). Not surprisingly some of the most abundant proteins identified in our screen function in photosynthesis. Of the 25 proteins with the largest number of unique peptides, 12 are light-harvesting complex (LHC) and photosystem I and II proteins: seven are plastid encoded and five are nucleus encoded. For the alpha subunit of the plastid-encoded ATP synthase, 20 unique peptides were mapped to its sequence (48% coverage). Interestingly, five of the 25 most abundant proteins are of unknown function, including several ORFans.

LHC proteins are amongst the most abundant proteins in our preparations and are primarily located in bands 9–12 (32–20 kDa) (fig. 2*F*). From the complete *B. natans* nuclear genome sequence and subsequent phylogenetic analyses (below), we infer that there are 19 LHC genes, 9 of which are part of the LHCII clade and likely make up the PSII antenna. There is evidence that all nine of these individual LHCII-related proteins are expressed under the growth conditions used in this study. These include proteins encoded by genes *Lhcbm1-8* plus CP29, a minor PSII antenna encoded by the *Lhcb4* gene. Other antennae proteins include those encoded by the three novel LHC genes *Lhcy1-3* described by Koziol et al. (2007) as being distantly related to the family of stress-related proteins in green algae (LHCSR). Although the function of these proteins is unknown, we hypothesize, based on their abundance

and the lack of genes orthologous to the LHCI antenna protein genes in *B. natans* (Koziol et al. 2007), that these proteins function as the PSI antenna system. Of particular interest is the lack of any peptides corresponding to the LHCX and LHCZ-class of proteins. In *B. natans*, there are five Lhcx and two Lhcz genes. The LHCX proteins are related to the LHCSR proteins of green algae that function in photoprotection (Peers et al. 2009). The absence of any of these proteins in this study suggests that these too are stress-induced and may function in photoprotection. The LHCZ proteins form a strongly supported clade (Koziol et al. 2007), but the function of this group of LHCs is unknown. They may only be induced under specific conditions or are present in low abundance. For both Lhcx and Lhcz genes, only a small number of ESTs, stemming from two different Sanger-based transcriptome sequencing projects, were recorded for each, emphasizing the low abundance of these transcripts under the nonstressed growth conditions used.

Chlorophyll Biosynthesis

We identified 10 nucleus-encoded, plastid-targeted *B. natans* proteins predicted to function in chlorophyll biosynthesis (table 1). Of these proteins, seven are predicted to function as part of the multistep tetrapyrrole biosynthesis pathway starting at glutamyl-tRNA^{Glu} and leading to the synthesis of chlorophyll *a* or *b* (Tanaka and Tanaka 2007; Stenbaek and Jensen 2010). We also identified chlorophyll *b* reductase, an enzyme that can convert chlorophyll *b* to chlorophyll *a* and is involved in chlorophyll degradation (Tanaka and Tanaka 2007). A chloroplast fluorescent-in-blue-light (FLU)-like protein was identified as a potential negative regulator of the tetrapyrrole pathway via interactions with glutamyl-tRNA reductase (Meskauskiene et al. 2001; Kauss et al. 2012). The latter protein was not found in the proteomic screen but the corresponding gene was identified in the nuclear genome. Geranylgeranyl reductase and a LIL3-like protein, two proteins also known to function in chlorophyll biosynthesis, were identified (LIL3 is thought to stabilize geranylgeranyl reductase) (Tanaka et al. 2010). All 10 of these proteins have a predicted SP and cTP (supplementary table S1, Supplementary Material online). Another identified protein involved in chlorophyll biosynthesis is Mg-protoporphyrin IX, which is encoded in the plastid genome.

Plastid-Miscellaneous

Of the six proteins categorized as "plastid-miscellaneous," three are thylakoid proteins as inferred from their closest matches in BLAST searches and considering their subcellular localizations in other organisms. Of particular interest is the organelle division protein FtsZ, which is notable in that it appears to be of red algal origin (discussed later). Only one member of the plastid protein import system (Tic–Toc), Tic62, was identified. Tic62 is an inner envelope protein involved in the regulation of preprotein import through sensing the $NADP^+$/NADPH ratio in the stroma via its NADP-binding domain (Balsera et al. 2007; Benz et al. 2009). Other proteins identified that are known to be involved in the protein import pathway include a plastid-targeted Cpn60 beta subunit and ClpC, as well as cytosolic Hsp70, Hsp90, and 14-3-3 proteins (Benz et al. 2009; Strittmatter et al. 2010). In primary plastids, preproteins are bound by Hsp70 in the cytosol and together with 14-3-3 proteins form a guidance complex that directs them to the plastid protein import system (Strittmatter et al. 2010). In the stroma, ClpC is involved in preprotein translocation across the inner envelope membrane and Cpn60 is believed to be involved in folding the proteins after they are imported (Li and Chiu 2010). How these processes are altered in the context of a secondary plastid surrounded by two additional membranes is not well understood. As noted above, Hsp70 is one of the three nucleomorph-encoded proteins identified in our proteomic screen, suggesting that Hsp70 and the 14-3-3 proteins might carry out the same functions in the PPC as they do in primary plastid-containing organisms: transporting plastid-targeted preproteins to the Tic–Toc complex (May and Soll 2000). However, recent work by Hirakawa et al. (2012) did not identify genes for Toc159 or Toc34 in the completely sequenced *B. natans* genome (Hirakawa et al. 2012); these proteins are believed to initially interact with 14-3-3-/Hsp70-bound proteins before import into the plastid (Strittmatter et al. 2010). It is not clear how this process would occur in the absence of Toc159 or Toc34.

Transport Proteins

Eight *B. natans* proteins were identified as being homologous to metabolite transporter proteins, including a Na^+/K^+ symporter, an adenine nucleotide translocator and a Na^+-dependent bile acid symporter. Five of these proteins are predicted to be SP^+/TP^+ and two others SP^+/TP^-. It is not possible to infer the precise cellular locations and roles of these transporters in *B. natans* based on the information currently available. However, Na^+-dependent bile acid transporters are known to be involved in glucosinolate metabolism within the plastid in other organisms (Linka and Weber 2010; Weber and Linka 2011). A protein belonging to the mitochondrial carrier family (SP^+/TP^+) was also identified in our proteomic screen. Proteins in this family have previously been identified in plastids, peroxisomes, and the ER, in addition to mitochondria (Weber and Linka 2011). An ABC transporter-like protein, belonging to a large protein superfamily of transporters, was also identified.

Cytoskeletal Proteins

Nineteen *B. natans* nucleus-encoded proteins classified as having functions related to the cytoskeleton were identified, none of which yields robust SP- or TP-like predictions. Eleven proteins were initially identified; upon closer inspection,

however, it became clear that in some cases these 11 proteins had been matched with peptides that also identified other proteins in the proteome, usually duplicate copies. For example, peptides mapping to four distinct beta tubulin proteins were present in the data set. Because we cannot determine which individual protein(s) is (are) the true match, we retained all four in our list. Cytoskeletal proteins containing peptides that could be matched to other proteins are alpha and beta tubulin, actin, the actin-related protein complex subunit ARPC2, and the actin-binding proteins coronin and cofilin. An additional hypothetical protein in this category was identified by Pfam analysis (protein ID: 60593) (supplementary table S1, Supplementary Material online).

In sum, these data are most consistent with the inference that most or all of the cytoskeletal proteins identified in our screen represent cytosolic contamination, perhaps due to physical interactions between the cytoskeleton and the PL–NM complexes. Such interactions can be used for positioning and movement of the plastid within the cell (Wada and Suetsugu 2004; Jouhet and Gray 2009). Notably, no obvious cytoskeleton proteins are encoded in the *B. natans* nucleomorph genome (Gilson et al. 2006).

Posttranslational Modification, Protein Turnover, and Chaperones

Bigelowiella natans proteins assigned to the posttranslational modification category consist of seven peptidyl-prolyl *cis*-trans isomerases (PPIases), six of eight subunits of the cytosolic chaperonin complex CCT, and several heat shock proteins. As noted earlier, the CCT proteins are likely contaminants. Based on the fact that the *B. natans* nucleomorph genome encodes six of the eight CCT subunits (Gilson et al. 2006), one might expect that the "missing" two subunits (CCT1 and CCT6) would be nucleus-encoded. However, there appears to be only a single gene for each of the eight CCT subunits in the nuclear genome, and none of the products of these genes possesses any obvious signaling or targeting information. Of the peptidyl-prolyl *cis*-trans isomerases, two were identified as being FKBP-type; the other five are cyclophilin type. These proteins are known to have many roles within the cell including functioning as molecular chaperones (Kurek et al. 2002). They may therefore be involved in chloroplast biogenesis (Gollan and Bhave 2010) and photosystem assembly (Gollan et al. 2011). Of the seven isomerases, we identified, five possess a bipartite leader sequence (SP^+/TP^+). In *Arabidopsis,* at least 16 PPIases have been identified in the thylakoid lumen (Peltier et al. 2002; Schubert et al. 2002), but only two actually possess PPIase activity (Shapiguzov et al. 2006). Also binned in this category are the five SP^-/TP^- 14-3-3 proteins discussed earlier.

A total of seven heat shock proteins were identified: two Hsp70, two Hsp90, two Hsp101, and a protein with weak sequence similarity to Hsp20 (ID: 88939). The two Hsp90 and Hsp70 proteins each have one isoform (IDs: 50519 and 92743) that is devoid of targeting signals and appears to be cytosolic. The other isoforms are predicted to have SP^+/TP^- sequences (IDs: 88281 and 155596). The SP^+/TP^- Hsp70 and Hsp90 proteins also possess, respectively, an HDEL and KDEL ER-retention sequence at their C-terminus, strongly suggesting they are localized to the ER. At least one of the heat shock proteins, Hsp101 (protein ID: 125406; SP^+/TP^+), is predicted to be PPC-targeted. Heat shock proteins of the Hsp70 and Hsp90 families are encoded in the *B. natans* nucleomorph genome (Gilson et al. 2006) as well as in cryptophyte nucleomorph genomes (Douglas et al. 2001; Lane et al. 2007; Tanifuji et al. 2011). Such proteins are presumably involved in the folding of proteins synthesized in the PPC as well as those being imported from the host cytosol. The Hsp20-like protein also possesses an SP^+/TP^+ sequence although we are unable to predict whether it is PPC- or plastid-localized. Cpn60 was also identified as discussed earlier.

We identified an FtsH protease, a protein that is a member of the AAA+ ATPases involved in the maintenance of the photosystem II complex through removal of damaged proteins (Chi et al. 2012). Noticeably absent were any proteasome subunits with evident SPs and/or TPs. The *B. natans* nucleomorph genome does not encode any obvious proteasome subunits (a prsA6 locus was annotated [Gilson et al. 2006] but it lacks any identifiable sequence similarity to proteasome subunits, suggesting it was mis-identified). We would therefore predict that genes for proteasome subunits mediating protein degradation in the PPC might be found in the host nuclear genome. However, only one gene copy of each of the 20S proteasome subunits, as well as the 26S RPT and RPN subunits, were found, none of which is obviously SP^+/TP^+. This observation raises the question of how protein degradation is carried out within the *B. natans* PPC. It is formally possible that the proteasome subunits are transported to the PPC by an alternative mechanism or that they are in fact among the subset of divergent nucleomorph genes with no identifiable function (Gilson et al. 2006). As mentioned earlier, ubiquitin was identified in our proteomic screen and there is a host nuclear gene encoding a ubiquitin monomer with an SP^+/TP^+ sequence; however, ubiquitin is involved in many processes in addition to protein degradation, including DNA repair, signaling, and protein trafficking (Dikic et al. 2009; Liu and Walters 2010). Nevertheless, although no proteasome subunits were identified in our proteomic screen, other proteins that have roles in protein degradation were found: these include the ubiquitin-conjugating enzyme E2, ubiquitin-activating enzyme E1, and a protein with weak similarity to a ubiquitin carboxy-terminal hydrolase (Reyes Turcu et al. 2009). None of these proteins has obvious SP^+/TP^+ sequences, suggesting they are cytosolic. Also identified was P97/Cdc48 (SP^-/TP^-), a highly abundant AAA-type ATPase involved in numerous processes, including the ubiquitin-proteasome system, ER-associated degradation (ERAD), autophagy, ER and Golgi

membrane fusion, among others (Uchiyama and Kondo 2005; Stolz et al. 2011; Wolf and Stolz 2012). It has been suggested that in other nonchlorarachniophyte secondary plastid-bearing organisms, the proteins of the ERAD machinery may be involved in the protein import process (Sommer et al. 2007; Moog et al. 2011). Proteins such as Cdc48 and ubiquitin-activating E1 were identified as being involved in this process (Moog et al. 2011). However, the *B. natans* nucleomorph genome encodes a cdc48-like protein (Gilson et al. 2006), consistent with the notion that the nucleus-encoded P97/Cdc48 is not PPC-targeted.

Transcription, Translation, and DNA-Binding Proteins

Of the 35 proteins categorized as transcription- or translation-related, 21 are ribosomal proteins. Sixteen of these are predicted to be cytosolic ribosomal proteins and the remaining five appear to be targeted to the plastid. All of the plastid ribosomal proteins have SP^+/TP^+ sequences. Three of the cytosolic ribosomal proteins appear to be ubiquitin-fusion proteins (Archibald, Teh, et al. 2003), all of which are SP^-/TP^-. The remaining proteins in this category include histones, translation elongation factors, and transcription factors. With the exception of chloroplast elongation factor G and plastid protein ycf 65, all these proteins lack obvious targeting information and appear to be localized to the cytosol or, in the case of the histones, the host nucleus. The proteins with SP^+/TP^+ sequences are most likely all plastid localized.

Metabolism

A wide variety of metabolic processes are represented by the proteins identified in our proteomic screen. These include glycolysis and the Calvin cycle (phosphoglycerate kinase [PGK], glyceraldehyde 3-phosphate dehydrogenase [GAPDH], and transketolase), as well as fatty acid biosynthesis [proteins such as acetyl-CoA carboxylase and 3-oxoacyl-(acyl-carrier-protein) reductase]. Of the proteins in this category, 56% are SP^+/TP^+ (table 1).

Carbohydrate metabolism is a major function of plastids in photosynthetic organisms, and we identified a number of enzymes that are involved in various carbohydrate biosynthetic pathways. Indeed, 33% of all identified metabolic enzymes are inferred to function in carbohydrate biosynthesis, and the majority are predicted to be localized to the plastid (table 2, supplementary table S1, Supplementary Material online). In organisms with primary plastids, part of the pathway takes place in the cytosol. However, in organisms with plastids of secondary origin and two cytosolic compartments, plastids may contain a different distribution of enzymes in various metabolic pathways.

The nucleomorph genome of *B. natans* encodes mainly "housekeeping" proteins, and the metabolic complexity of its PPC is unknown. Of the 16 glycolytic/Calvin cycle enzymes detected by proteomics, 11 have predicted SP^+/TP^+ motifs.

Table 2

Bigelowiella natans Nucleus-Encoded Metabolic Proteins Identified by Proteomic Screen

	No. of Proteins	Total (%)	SP^+/TP^+	Total (%)
Carbohydrate metabolism	16	33	11	69
Amino acid metabolism	7	15	3	43
Nucleotide metabolism	3	6	1	33
Energy production	6	12	4	67
Lipid metabolism	7	15	4	57
Other	9	19	4	44

The four Calvin cycle enzymes identified are two copies of the ribulose-1,5-*bis*phosphate carboxylase/oxygenase small subunit, PGK, transketolase and GAPDH. Two copies of GAPDH were identified, a plastid-targeted homolog (ID: 50264) and a cytosolic version (ID: 92554). Peptides matching the cytosolic GAPDH could also be assigned to a third GAPDH (ID: 225526) (SP^+/TP^+), suggesting that this version is a PPC-targeted protein. The carbohydrate metabolism proteins without obvious SP/TP sequences are cytosolic fructose-1, 6-bisphosphate aldolase and pyruvate phosphate dikinase (PPDK) (supplementary table S1, Supplementary Material online). There are two PPDK genes in the *B. natans* nuclear genome. Neither of these enzymes has an obvious bipartite leader sequence, which suggests they function in the host cytosol.

Two UDP-glucose 4,6-dehydratase/rhamnose biosynthetic enzyme proteins (ID: 54005 and 52540) were identified. An SP was identified by more than one prediction program (iPSORT and PredSL) and iPSORT additionally recognized a cTP; however, these proteins do not exhibit an N-terminal extension as compared with their homologs in other organisms, so they may be localized in the cytosol. These proteins function in the conversion of D-glucose to L-rhamnose (Reiter 2008). Three other *B. natans* proteins predicted to function in the saccharide biosynthetic pathway were identified: nucleoside-diphosphate sugar epimerase (SP^+/TP^+), and two GDP-mannose epimerases. Given that in *Arabidopsis* rhamnose biosynthesis occurs in the cytosol (Seifert 2004), it is possible that these enzymes are PPC-localized in *B. natans*. Other carbohydrate metabolic proteins identified that were predicted to be plastid- or PPC-targeted are phosphogluconate dehydrogenase (ID: 144847) and a NADH-ubiquinone oxidoreductase complex I-like protein (ID: 155984).

Fatty acid biosynthesis is another major biochemical process occurring in the plastid (Joyard et al. 2010), and we identified seven *B. natans* enzymes in this category. Peptides matching acetyl-CoA carboxylase, 3-oxoacyl-(acyl-carrier-protein) reductase, AMP-dependent synthetase, and a long-chain acyl-CoA synthetase were found; all four enzymes have a predicted SP^+/TP^+ and are presumably plastid localized. Two long-chain acyl-CoA synthetases were also recovered but they lack obvious SP^+/TP^+ sequences (the genome sequence contains a gap

in the 5′ region of the gene for one of the two long-chain acyl-CoA synthetases [ID: 155970], so the precise gene structure is uncertain). In addition, an SP^-/TP^- acetyl-CoA synthetase was detected (so is perhaps a cytosolic version). However, other acetyl-CoA synthetase genes in the *B. natans* genome are predicted to have SP^+/TP^+ coding regions.

Three *B. natans* proteins were identified as being involved in nucleotide metabolism: adenylate kinase (SP^+/TP^+), carbamoyl phosphate synthetase protein (SP^-/TP^-), and phosphoribosylaminoimidazole carboxylase. The latter enzyme is involved in purine biosynthesis, but its coding sequence could not be examined for the presence of an SP/TP due to a gap in the genome sequence near the 5′ end of the gene. Adenylate kinase catalyzes the transfer of a phosphate from ATP to AMP, whereas carbamoyl phosphate synthetase II functions in pyrimidine biosynthesis and is expected to have a cytosolic function. The apparent absence of an SP^+/TP^+ suggests that this protein is localized to the host cytosol, not the PPC.

Seven enzymes were predicted to function in amino acid metabolism, three of which exhibit SP^+/TP^+ sequences (table 2). The three proteins with SP^+/TP^+ sequences are cysteine synthase, arginine kinase and a glycine/serine hydroxymethyltransferase. Cysteine synthase is expected to be plastid-localized. An NADH-dependent glutamate synthase (SP^-/TP^+) was also identified; however, the ChloroP TP prediction program predicted a cTP sequence, whereas other TP prediction programs (IPSort; PredSL) indicated an mTP. This example illustrates the difficulty in assigning localizations based on TP predictions. The other three proteins—glutamine synthetase, a second isoform of glycine/serine hydroxymethyltransferase, and alanine-glyoxylate aminotransferase (AGT)—do not have identifiable SP/TP sequences. In addition to participating in alanine biosynthesis, AGT is known to function together with glycine/serine hydroxymethyltransferases in serine and glycine metabolism. AGT catalyzes the conversion of L-alanine and glyoxylate to pyruvate and glycine. Interestingly, serine metabolism is predicted to occur in the mitochondria of diatoms (Kroth et al. 2008). Glutamine synthetase functions in glutamine biosynthesis as well as in the nitrogen assimilation pathway; whereas in some organisms, isoforms are localized to the cytosol, others can also be localized in the plastid (Bernard et al. 2008). However, none of the five glutamine synthetase genes found in the *B. natans* genome has an identifiable SP^+/TP^+ sequence. One isoform has an mTP and another has a TP, although whether cTP or mTP is unclear. We also identified an SP^+/TP^+ arginine kinase, consistent with the possibility that it is plastid or PPC localized.

With respect to energy production, seven *B. natans* proteins were recovered in our proteomic screen; all had SP^+/TP^+ sequences except for fumarate hydratase and isocitrate dehydrogenase, the latter being a mitochondrial protein functioning in the TCA/Krebs cycle. Indeed, isocitrate dehydrogenase is predicted to possess a mitochondrial TP, whereas fumarate hydratase does not have any obvious targeting sequences. Ferredoxin-NADP oxidoreductase, quinone oxidoreductase-like protein, chloroplast ascorbate peroxidase, and a predicted thylakoid kinase protein are all predicted to have SP^+/TP^+ sequences. We identified an SP^+/TP^+ quinone oxidoreductase-like protein, which was bioinformatically predicted to be localized to the PPC (Curtis et al. 2012). The subcellular localization of an identified fumarate reductase (SP^+/TP^-) is unknown. In addition, there are two other fumarate reductase-like proteins in the genome, one of which also has an SP^+/TP^+ sequence.

Nine additional proteins do not obviously fit into any of the above metabolic categories. Four of these have SP^+/TP^+ sequences: a short-chain dehydrogenase (ID: 89296), a ferredoxin nitrite reductase, a nitroreductase-like protein, and cytochrome c peroxidase. Two proteins, fumarate reductase and 2-nitropropane dioxygenase, have only an SP (supplementary table S1, Supplementary Material online).

Signal Transduction and Intracellular Transport

We identified 12 proteins classified as signal transduction proteins, only two of which (IDs: 89051 and 237242) have predicted bipartite leader sequences; both share similarity to GTPase-like proteins in the C-terminal half of the protein. It is unknown whether these proteins are plastid-targeted or localized in the PPC. The remaining proteins consist of Ras-related GTPases (all SP^-/TP^-), two Rab8/RabE-related proteins (SP^+/TP^-), a guanine nucleotide-binding protein beta subunit (SP^-/TP^-), and an adenylate cyclase-associated protein (SP^-/TP^-) (supplementary table S1, Supplementary Material online).

Five proteins fall into the intracellular transport category: two clathrin heavy chain proteins, an adaptor protein complex subunit (AP2A1), a coat protein complex subunit (COPG-2), and an ADP-ribosylation factor. These proteins are expected to be localized to the cytosol, as they do not have any predicted SP/TP leader sequences.

Miscellaneous Proteins and Proteins of Unknown Function

The functions of 73 *B. natans* proteins identified in our screen (24% of the total) are "unknown," making this the largest functional category of proteins. This group includes conserved proteins of unknown function (24 in total) as well as true ORFans (21 proteins), that is, proteins with no significant sequence similarity to any known sequence. More than 70% of these proteins are predicted to have bipartite leader sequences. One such unknown protein (ID: 86118) is among the highest-coverage proteins in our data set, with 29 unique peptides spanning 39% of the protein (326 of 825 amino acids).

Interestingly, 11 proteins within the unknown category display sequence similarity to other *B. natans* proteins of

unknown function that are also present in our proteomic data set (which comprises both conserved unknown proteins and ORFans). Another unknown protein (ID: 86287) (SP^+/TP^+) is highly similar to hypothetical proteins in a wide range of bacteria and a restricted set of unrelated algae, including dinoflagellates (e.g., *Symbiodinium* sp.), the green alga *Micromonas* sp., and the heterokont *Aureococcus* sp. (discussed later). Unknown protein 86287 is also homologous to two other proteins identified in our screen (86292 and 50950) and an additional homolog encoded by the *B. natans* genome. Three of these four proteins have predicted SP^+/TP^+ sequences, whereas the fourth protein (ID: 50950) has only a TP.

The "miscellaneous" category comprises five proteins (table 1 and supplementary table S1, Supplementary Material online), only one of which (ID: 90881) (SP^+/TP^+) is predicted to be plastid-localized. The others include a cell division control 2-like protein and a protein kinase.

Phylogenetic Origins of *B. natans* Plastid, Nucleomorph, and PPC Proteins

The *B. natans* plastid and nucleomorph evolved from a green alga by secondary endosymbiosis (Ishida et al. 1997). It would be expected that most of the plastid-, PPC-, and nucleomorph-targeted proteins would be green algal in origin and that this expectation would be reflected in their phylogenies. A previous phylogenetic study of 78 plastid-targeted proteins in *B. natans* (Archibald, Rogers, et al. 2003) revealed that while the majority of these proteins were indeed of green algal origin, a significant fraction appeared to be the result of lateral gene transfer. Approximately 21% of genes examined had either red algal, streptophyte algal or bacterial origin (Archibald, Rogers, et al. 2003). This result was attributed to the mixotrophic nature of chlorarachniophytes (Archibald, Rogers, et al. 2003), whereby prey organisms ingested as food could be a source of "foreign" genes for the organism.

We revisited the issue of the mosaic nature of the *B. natans* plastid using the expanded set of proteins identified herein by proteomics (excluding those that were deemed to be host cytosolic contaminants) and the >5,000 phylogenetic trees generated by Curtis et al. (2012). Plastid-, PPC-, and nucleomorph-targeted proteins were binned according to whether the *B. natans* protein 1) grouped with "green" plastid-bearing organisms (i.e., chlorophyte green algae and streptophytes), 2) with red algae or organisms with red algal-derived plastids such as heterokonts and haptophytes, 3) was "ambiguous" but of probable algal origin, or 4) was of uncertain origin ("other"). Proteins were assigned to either "red" or "green" if they were grouped within those categories with a bootstrap support of >75%.

Of the 302 nucleus-encoded *B. natans* proteins identified in the proteomic screen, 89 did not yield RAxML trees in the phylogenomic pipeline of Curtis et al. (2012), due to a lack of detectable sequence similarity with other proteins and/or with homology to only a few known proteins. As expected, the majority of these proteins reside in our "unknown" category. Of the 74 unknown proteins, 57 did not have a corresponding phylogenetic tree. Of the 213 phylogenetic trees that we could consider, only 47 showed a robust "red" or "green" signal for the *B. natans* homolog. Interestingly, these proteins were evenly split between the red and green categories. Twenty-two proteins showed a clear green evolutionary origin and 24 proteins grouped with red algae and/or organisms with red algal-derived plastids. Of the "green" proteins, 21 out of 22 had a predicted SP and TP. In contrast, only 15/24 of the "red" proteins were SP^+/TP^+, and an additional 3 proteins were SP^-/TP^+. The proteins showing a strong "green" signal were predominantly in photosynthesis-related functional categories (e.g., light-harvesting proteins and chlorophyll biosynthesis), whereas proteins showing a strong red signal largely fell into the metabolism (8/24) and unknown (5/24) categories.

Representative phylogenetic trees are shown in figure 3. The phylogeny of the *B. natans* plastid-targeted protein cytochrome C6 (50246) is indicative of a green algal ancestry for this protein (fig. 3*A*). Two red algal-type *B. natans* proteins are the hcf136 (236790) (fig. 3*B*) and organelle division protein FtsZ (91954) (supplementary fig. S2, Supplementary Material online). Both proteins are found in a variety of green algae, red algae, and red secondary plastid-bearing organisms and, significantly, are encoded in the red algal-derived nucleomorph genomes of several cryptophytes (Tanifuji et al. 2011). *Bigelowiella natans* has two highly similar FtsZ genes/proteins, both of which are red algal in nature and appear to be targeted to the plastid. The phylogenetic tree of unknown protein 86287 (supplementary fig. S3, Supplementary Material online) is even more complex: four *B. natans* homologs branch with an eclectic mix of algae, including the green alga *Micromonas* sp., the heterokont *Aureococcus* sp., and the dinoflagellate *Symbiodinium* sp., within a tree otherwise made up exclusively of prokaryotic sequences. No simple model of endosymbiotic or lateral gene transfer can explain the existence of this protein gene in *B. natans* and such a small yet diverse set of eukaryotic phototrophs. In sum, our results are consistent with those of Archibald, Rogers, et al. (2003): a significant proportion of the plastid/PPC proteome in *B. natans* is of nongreen algal origin, presumably the result of repeated lateral gene transfers involving red algal plastid-bearing donors. Interestingly, a roughly equal "red" and "green" phylogenetic signal was recently observed by Woehle et al. (2011) in an analysis of nucleus-encoded plastid proteins in *Chromera*, a newly discovered alga with a red secondary plastid (Moore et al. 2008). In trying to make sense of these conflicting signals, these authors emphasize the current challenges associated with accurately inferring the evolutionary histories of genes with incomplete taxonomic sampling

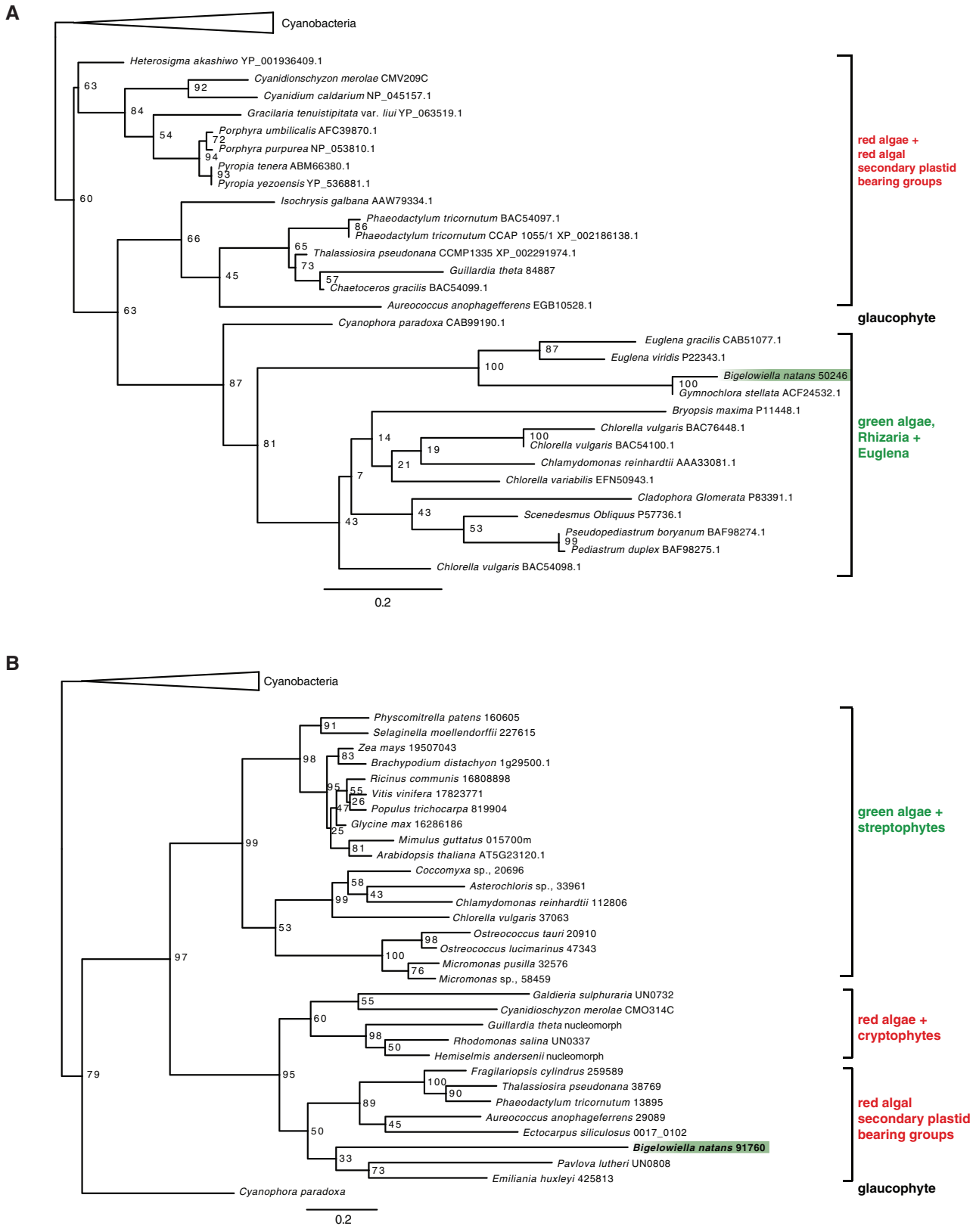

Fig. 3.—Maximum likelihood phylogenetic trees of proteins (*A*) cytochrome C6 (50246) and (*B*) Hcf136 (236790). The *Bigelowiella natans* cytochrome C6 protein shows a strong grouping (>75% bootstrap support) with green algae, whereas the tree of Hcf136 suggests a red algal origin.

and the possibility of phylogenetic artifacts. These uncertainties apply to the *B. natans* data set analyzed herein as well.

Conclusions

The goal of our proteomic screen was to gain insight into the proteomes of three distinct subcellular compartments in *B. natans*: the plastid, the PPC, and the nucleomorph. Of 302 nucleus-encoded proteins identified in this screen, only a handful can confidently be assigned to the PPC. Of the proteins believed to be targeted to the PL–NM complex, the majority are predicted to be plastid-targeted. This result is not entirely unexpected, for two main reasons. First, the PPC and nucleomorph are tiny compartments relative to the plastid; their proteins thus presumably constituted only a small fraction of the total sample analyzed by HPLC-MS/MS. Indeed, only three nucleomorph-encoded proteins were identified in this study. Second, from a functional perspective, the PPC is largely an "unknown quantity," unlike plastids whose proteomes have been extensively characterized in other systems. More than 70% of the "unknown" proteins recovered in our screen have bipartite leader sequences, so a significant fraction of these unknowns could be PPC localized.

One of the major challenges of this study was determining whether proteins were cytosolic contaminants or *bona fide* PPC proteins. Curtis et al. (2012) used a bioinformatics approach to predict PPC-targeted proteins in *B. natans* and the cryptophyte *Guillardia theta*, using the presence of SP and TP sequences as the first step of their screening procedure. We also relied on the presence or absence of such sequences in deciding whether a protein identified in our proteomic screen with a predicted "cytosolic" function is most likely to be cytosolic or PPC targeted. Nevertheless, our data provide an important starting point for investigating the possibility of alternative, signal and/or TP-independent, protein trafficking pathways in *B. natans*. We compared the predictions of protein localization of our data set to the bioinformatically predicted proteomes of the plastid, PPC, mitochondria, and ER/Golgi inferred by Curtis et al. (2012). Only 81 of 302 proteins identified in this study were included in the Curtis et al. (2012) proteomes. The identification of over 200 additional possible PL/NM complex proteins effectively emphasizes the importance of proteomic studies in identifying those proteins whose localization cannot easily be predicted bioinformatically.

Unlike the situation in cryptophytes, the plastid/PPC of chlorarachniophytes such as *B. natans* does not reside within the ER lumen, and precisely how proteins are transported to the outermost plastid membrane is not understood. Advances in this area will require detailed lab-based experimentation, including subcellular localizations. Fortunately, a transfection system for chlorarachniophytes has been developed (Hirakawa et al. 2008) and has already been used to determine the subcellular locations of a handful of proteins (Hirakawa et al. 2009, 2010), as well as to better understand the biochemical features of the N-terminal topogenic signals important for targeting specificity. The proteins of known and unknown function identified in this study should prove useful for future investigations, including those with and without obvious bipartite leader sequences.

Supplementary Material

Supplementary figures S1–S3 and tables S1 and S2 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Mary Ann Trevors and Stephen Whitefield for technical support with microscopy. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant awarded to J.M.A. and an NSERC Special Research Opportunities Grant awarded to J.M.A. and M.W.G. D.G.D. also acknowledges NSERC Discovery grant support. J.M.A. and M.W.G. acknowledge support from the Canadian Institute for Advanced Research and the Centre for Comparative Genomics and Evolutionary Bioinformatics at Dalhousie University. J.M.A. holds a New Investigator Award from the Canadian Institutes of Health Research.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Archibald JM. 2007. Nucleomorph genomes: structure, function, origin and evolution. Bioessays 29:392–402.

Archibald JM. 2009. The puzzle of plastid evolution. Curr Biol. 19: R81–R88.

Archibald JM, Logsdon JM Jr, Doolittle WF. 2000. Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. Mol Biol Evol. 17:1456–1466.

Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ. 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. Proc Natl Acad Sci U S A. 100:7678–7683.

Archibald JM, Teh EM, Keeling PJ. 2003. Novel ubiquitin fusion proteins: ribosomal protein P1 and actin. J Mol Biol. 328:771–778.

Balsera M, Stengel A, Soll J, Bölter B. 2007. Tic62: a protein family from metabolism to protein translocation. BMC Evol Biol. 7:43.

Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S. 2002. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics 18:298–305.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 340:783–795.

Benz JP, Soll J, Bolter B. 2009. Protein transport in organelles: the composition, function and regulation of the Tic complex in chloroplast protein import. FEBS J. 276:1166–1176.

Bernard SM, et al. 2008. Gene expression, cellular localisation and function of glutamine synthetase isozymes in wheat (*Triticum aestivum* L.). Plant Mol Biol. 67:89–105.

Capella-Gutiérrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.
Chi W, Sun X, Zhang L. 2012. The roles of chloroplast proteases in the biogenesis and maintenance of photosystem II. Biochim Biophys Acta. 1817:239–246.
Curtis BA, et al. 2012. Algal nuclear genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492:59–65.
Dikic I, Wakatsuki S, Walters KJ. 2009. Ubiquitin-binding domains—from structures to functions. Nat Rev Mol Cell Biol. 10:659–671.
Douglas S, et al. 2001. The highly reduced genome of an enslaved algal nucleus. Nature 410:1091–1096.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.
Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2: 953–971.
Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 300:1005–1016.
Emanuelsson O, Nielsen H, von Heijne G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. 8:978–984.
Foth BJ, et al. 2003. Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. Science 299:705–708.
Gilson PR, et al. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Proc Natl Acad Sci U S A. 103:9566–9571.
Gollan PJ, Bhave M. 2010. A thylakoid-localised FK506-binding protein in wheat may be linked to chloroplast biogenesis. Plant Physiol Biochem. 48:655–662.
Gollan PJ, Ziemann M, Bhave M. 2011. PPIase activities and interaction partners of FK506-binding proteins in the wheat thylakoid. Physiol Plant. 143:385–395.
Gould SB, et al. 2006. Protein targeting into the complex plastid of cryptophytes. J Mol Evol. 62:674–681.
Hempel F, et al. 2007. Transport of nuclear-encoded proteins into secondarily evolved plastids. Biol Chem. 388:899–906.
Hempel F, Bullmann L, Lau J, Zauner S, Maier U-G. 2009. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. Mol Biol Evol. 26:1781–1790.
Hirakawa Y, Burki F, Keeling PJ. 2012. Genome-based reconstruction of the protein import machinery in the secondary plastid of a chlorarachniophyte alga. Eukaryot Cell. 11:324–333.
Hirakawa Y, Gile GH, Ota S, Keeling PJ, Ishida K. 2010. Characterization of periplastidal compartment-targeting signals in chlorarachniophytes. Mol Biol Evol. 27:1538–1545.
Hirakawa Y, Kofuji K, Ishida K. 2008. Transient transformation of a chlorarachniophyte alga, *Lotharella amoebiformis* (Chlorarachniophyceae), with *uid*A and *egfp* reporter genes. J Phycol. 44:814–820.
Hirakawa Y, Nagamune K, Ishida K. 2009. Protein targeting into secondary plastids of chlorarachniophytes. Proc Natl Acad Sci U S A. 106: 12820–12825.
Ishida K, Cao Y, Hasegawa M, Okada N, Hara Y. 1997. The origin of chlorarachniophyte plastids, as inferred from phylogenetic comparisons of amino acid sequences of EF-Tu. J Mol Evol. 45:682–687.
Jouhet J, Gray JC. 2009. Interaction of actin and the chloroplast protein import apparatus. J Biol Chem. 284:19132–19141.
Joyard J, et al. 2010. Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. Prog Lipid Res. 49:128–158.
Kauss D, Bischof S, Steiner S, Apel K, Meskauskiene R. 2012. FLU, a negative feedback regulator of tetrapyrrole biosynthesis, is physically linked to the final steps of the Mg^{++}-branch of this pathway. FEBS Lett. 586: 211–216.
Keller A, Nesvizhskii AI, Kolker E, Aebersold R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem. 74:5383–5392.
Koziol AG, et al. 2007. Tracing the evolution of the light-harvesting antennae in chlorophyll *a/b*-containing organisms. Plant Physiol. 143: 1802–1816.
Kroth PG, et al. 2008. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. PLoS One 3:e1426.
Kurek I, Pirkl F, Fischer E, Buchner J, Breiman A. 2002. Wheat FKBP73 functions in vitro as a molecular chaperone independently of its peptidyl prolyl *cis-trans* isomerase activity. Planta 215:119–126.
Lane CE, et al. 2007. Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. Proc Natl Acad Sci U S A. 104:19908–19913.
LeBihan T, et al. 2011. Shotgun proteomic analysis of the unicellular alga *Ostreococcus tauri*. J Proteomics. 74:2060–2070.
Li HM, Chiu C-C. 2010. Protein transport into chloroplasts. Annu Rev Plant Biol. 61:157–180.
Linka N, Weber APM. 2010. Intracellular metabolite transporters in plants. Mol Plant. 3:21–53.
Liu F, Walters KJ. 2010. Multitasking with ubiquitin through multivalent interactions. Trends Biochem Sci. 35:352–360.
May T, Soll J. 2000. 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. Plant Cell 12:53–63.
McFadden GI. 1999. Plastids and protein targeting. J Eukaryot Microbiol. 46:339–346.
McFadden GI, Gilson PR, Hofmann CJ, Adcock GJ, Maier U-G. 1994. Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. Proc Natl Acad Sci U S A. 91:3690–3694.
Meskauskiene R, et al. 2001. FLU: a negative regulator of chlorophyll biosynthesis in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A. 98: 12826–12831.
Moog D, Stork S, Zauner S, Maier U-G. 2011. In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. Genome Biol Evol. 3:375–382.
Moore CE, Archibald JM. 2009. Nucleomorph genomes. Annu Rev Genet. 43:251–264.
Moore RB, et al. 2008. A photosynthetic alveolate closely related to apicomplexan parasites. Nature 451:959–963.
Nassoury N, Morse D. 2005. Protein targeting to the chloroplasts of photosynthetic eukaryotes: getting there is half the fun. Biochim Biophys Acta. 1743:5–19.
Nesvizhskii AI, Keller A, Kolker E, Aebersold R. 2003. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 75: 4646–4658.
Patron NJ, Waller RF. 2007. Transit peptide diversity and divergence: a global analysis of plastid targeting signals. Bioessays 29:1048–1058.
Peers G, et al. 2009. An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. Nature 462:518–521.
Peltier J-B, et al. 2002. Central functions of the lumenal and peripheral thylakoid proteome of *Arabidopsis* determined by experimentation and genome-wide prediction. Plant Cell 14:211–236.
Petsalaki EI, Bagos PG, Litou ZI, Hamodrakas SJ. 2006. PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. Genomics Proteomics Bioinformatics 4:48–55.
Reiter W-D. 2008. Biochemical genetics of nucleotide sugar interconversion reactions. Curr Opin Plant Biol. 11:236–243.
Reyes-Prieto A, Weber APM, Bhattacharya D. 2007. The origin and establishment of the plastid in algae and plants. Annu Rev Genet. 41: 147–168.
Reyes Turcu FE, Ventii KH, Wilkinson KD. 2009. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. Annu Rev Biochem. 78:363–397.

Rogers MB, et al. 2004. Plastid-targeting peptides from the chlorarachniophyte *Bigelowiella natans*. J Eukaryot Microbiol. 51:529–535.
Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 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.
Sánchez R, et al. 2011. Phylemon 2.0: a suite of Web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucleic Acids Res. 39:W470–W474.
Schubert M, et al. 2002. Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. J Biol Chem. 277:8354–8365.
Seifert GJ. 2004. Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. Curr Opin Plant Biol. 7: 277–284.
Shapiguzov A, Edvardsson A, Vener AV. 2006. Profound redox sensitivity of peptidyl-prolyl isomerase activity in Arabidopsis thylakoid lumen. FEBS Lett. 580:3671–3676.
Smith DGS, et al. 2007. Exploring the mitochondrial proteome of the ciliate protozoon *Tetrahymena thermophila*: direct analysis by tandem mass spectrometry. J Mol Biol. 374:837–863.
Sommer MS, et al. 2007. Der1-mediated preprotein import into the periplastid compartment of chromalveolates? Mol Biol Evol. 24:918–928.
Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol. 57:758–771.
Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19(Suppl 2):ii215–ii225.
Stenbaek A, Jensen PE. 2010. Redox regulation of chlorophyll biosynthesis. Phytochemistry 71:853–859.
Stolz A, Hilt W, Buchberger A, Wolf DH. 2011. Cdc48: a power machine in protein degradation. Trends Biochem Sci. 36:515–523.
Strittmatter P, Soll J, Bölter B. 2010. The chloroplast protein import machinery: a review. Methods Mol Biol. 619:307–321.
Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.
Tanaka R, Tanaka A. 2007. Tetrapyrrole biosynthesis in higher plants. Annu Rev Plant Biol. 58:321–346.
Tanaka R, et al. 2010. LIL3, a light-harvesting-like protein, plays an essential role in chlorophyll and tocopherol biosynthesis. Proc Natl Acad Sci U S A. 107:16721–16725.
Tanifuji G, et al. 2011. Complete nucleomorph genome sequence of the nonphotosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. Genome Biol Evol. 3:44–54.
Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 5:123–135.
Uchiyama K, Kondo H. 2005. p97/p47-mediated biogenesis of Golgi and ER. J Biochem. 137:115–119.
Valpuesta JM, Martín-Benito J, Gómez-Puertas P, Carrascosa JL, Willison KR. 2002. Structure and function of a protein folding machine: the eukaryotic cytosolic chaperonin CCT. FEBS Lett. 529:11–16.
Wada M, Suetsugu N. 2004. Plant organelle positioning. Curr Opin Plant Biol. 7:626–631.
Wasiak S, et al. 2002. Enthoprotin: a novel clathrin-associated protein identified through subcellular proteomics. J Cell Biol. 158: 855–862.
Weber APM, Linka N. 2011. Connecting the plastid: transporters of the plastid envelope and their role in linking plastidial with cytosolic metabolism. Annu Rev Plant Biol. 62:53–77.
Wittpoth C, Kroth PG, Weyrauch K, Kowallik KV, Strotmann H. 1998. Functional characterization of isolated plastids from two marine diatoms. Planta 206:79–85.
Woehle C, Dagan T, Martin WF, Gould SB. 2011. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related *Chromera velia*. Genome Biol Evol. 3:1220–1230.
Wolf DH, Stolz A. 2012. The Cdc48 machine in endoplasmic reticulum associated protein degradation. Biochim Biophys Acta. 1823:117–124.

Associate editor: Bill Martin