sharing their kinematic models and GPS data for comparison with the dynamic models, L. Wen for help with using his code, and three anonymous reviewers for their comments that substantially improved the manuscript. We are grateful to UNAVCO, Incorporated Research Institutions for Seismology, NSF-EarthScope, International GNSS Service, NASA, and countless researchers for facilitation and availability of geodetic and seismic data. The codes are available as a zip file as part of the supporting online material.

Supporting Online Material
www.sciencemag.org/cgi/content/full/335/6070/838/DC1
Methods
SOM Text
Figs. S1 and S2
Table S1
References (*29*–*37*)
Computer Codes

20 September 2011; accepted 24 January 2012
10.1126/science.1214209

Cyanophora paradoxa Genome Elucidates Origin of Photosynthesis in Algae and Plants

Dana C. Price,[1] Cheong Xin Chan,[1]* Hwan Su Yoon,[2,3] Eun Chan Yang,[2] Huan Qiu,[2] Andreas P. M. Weber,[4] Rainer Schwacke,[5] Jeferson Gross,[1] Nicolas A. Blouin,[6] Chris Lane,[6] Adrián Reyes-Prieto,[7] Dion G. Durnford,[8] Jonathan A. D. Neilson,[8] B. Franz Lang,[9] Gertraud Burger,[9] Jürgen M. Steiner,[10] Wolfgang Löffelhardt,[11] Jonathan E. Meuser,[12] Matthew C. Posewitz,[13] Steven Ball,[14] Maria Cecilia Arias,[14] Bernard Henrissat,[15] Pedro M. Coutinho,[15] Stefan A. Rensing,[16,17,18] Aikaterini Symeonidi,[16,17] Harshavardhan Doddapaneni,[19] Beverley R. Green,[20] Veeran D. Rajah,[1] Jeffrey Boore,[21,22] Debashish Bhattacharya[1]†

The primary endosymbiotic origin of the plastid in eukaryotes more than 1 billion years ago led to the evolution of algae and plants. We analyzed draft genome and transcriptome data from the basally diverging alga *Cyanophora paradoxa* and provide evidence for a single origin of the primary plastid in the eukaryote supergroup Plantae. *C. paradoxa* retains ancestral features of starch biosynthesis, fermentation, and plastid protein translocation common to plants and algae but lacks typical eukaryotic light-harvesting complex proteins. Traces of an ancient link to parasites such as Chlamydiae were found in the genomes of *C. paradoxa* and other Plantae. Apparently, *Chlamydia*-like bacteria donated genes that allow export of photosynthate from the plastid and its polymerization into storage polysaccharide in the cytosol.

Eukaryote evolution has largely been shaped by the process of primary endosymbiosis, whereby bacterial cells were taken up and over time evolved into double membrane–bound organelles, the plastid and the mitochondrion [e.g., (*1*, *2*)]. The cyanobacterium-derived plastid is found in diverse photosynthetic organisms, including Glaucophyta, Rhodophyta, and green algae and their land plant descendants (the Viridiplantae). These three lineages are postulated to form the monophyletic group Plantae (or Archaeplastida) (*3*–*6*), a hypothesis that suggests the primary cyanobacterial endosymbiosis occurred exclusively in their single common ancestor. Plastid gene trees demonstrate a single origin of the Plantae (*5*, *7*); however, many nuclear, multiprotein phylogenies provide little (*8*) or no support (*9*, *10*) for their monophyly. These latter results may reflect a reticulate ancestry among genes that can mislead phylogenetic inference (*11*). Furthermore, glaucophytes retain ancestral cyanobacterial features not found in other Plantae (*12*)—such as the presence of peptidoglycan between the two bounding membranes of the plastid (*13*)—that cast doubt on their evolutionary history. It is therefore unclear whether the Plantae host and its plastid, with its associated complex machinery (e.g., for plastid protein import and solute transport) (*14*, *15*), had a single origin or multiple origins. To elucidate the evolutionary history of key algal and land plant traits and to test Plantae monophyly, we have generated a draft assembly of the ≈70 Mbp nuclear genome from the glaucophyte *Cyanophora paradoxa* CCMP329 (Pringsheim strain) (Fig. 1A).

A total of 27,921 *C. paradoxa* proteins were predicted from the genome data, and 4628 had significant BLASTp hits ($e \leq 10^{-10}$) to prokaryote and eukaryote genome data in our comprehensive local database (table S1). Using phylogenomics (*16*), we generated 4445 maximum likelihood trees from the *C. paradoxa* proteins and found that >60% support a sister-group relationship between glaucophytes and red and/or green algae with a bootstrap value ≥90% (Fig. 1B and fig. S1). The Plantae clade in many of these trees is, however, interrupted by chlorophyll $a + c$ containing "chromalveolates." An example of this type of tree is fructose-1,6-bisphosphatase (Fig. 1C and fig. S2), which has cytosolic and plastidic isoforms. The gene for this enzyme, found in stramenopiles (e.g., diatoms) and haptophytes, originated from the red algal secondary endosymbiont that gave rise to the plastid in these taxa (*2*, *9*). This sort of intracellular gene transfer associated with endosymbiosis (EGT) has greatly enriched algal and land plant genomes (*17*, *18*).

We estimated the "footprint" of cyanobacterium-derived EGT in Plantae genomes. The proportion of cyanobacterium-derived nuclear genes varies from 18% in *Arabidopsis thaliana* (*19*) to ~7% in mesophilic red algae and 6% in *Chlamydomonas reinhardtii* (*20*, *21*). Phylogenomic analysis of the predicted *C. paradoxa* proteins showed 274 to be of cyanobacterial provenance (*22*). This constitutes ~6% of proteins in the glaucophyte that have significant BLASTp hits (i.e., 274 out of 4628), as found in other algae (*20*, *21*). BLASTp analysis identified 2029 proteins that are putatively destined for the plastid, of which 293 contain the transit sequence for plastid import [identified by the presence of phenylalanine (F) within the first four amino acids: MF, MAF, MNAF, MSAF, and MAAF] (*23*, *24*) (fig. S4B). Of these 293 proteins, 80% are derived from Cyanobacteria.

Another source of foreign genes in Plantae is horizontal gene transfer (HGT), which is not associated with endosymbiosis. Using 35,126 bacterial sequences as a query, we found 444 noncyanobacterial gene families with a common origin shared amongst Bacteria and Plantae. Among them, 15 genes are present in all three Plantae phyla. An example of a gene derived from Bacteria after an ancient HGT event that is shared by

[1]Department of Ecology, Evolution, and Natural Resources and Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ 08901, USA. [2]Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME 04575, USA. [3]Department of Biological Sciences, Sungkyunkwan University, Suwon 440-746, Korea. [4]Institute for Plant Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf, Germany. [5]Botanical Institute, Biocenter Cologne, University of Cologne, Zülpicher Strasse 47b, 50674 Cologne, Germany. [6]Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881, USA. [7]Canadian Institute for Advanced Research, Department of Biology, University of New Brunswick, Fredericton, New Brunswick E3B 5A3, Canada. [8]Department of Biology, University of New Brunswick, Fredericton, New Brunswick E3B 5A3, Canada. [9]Centre Robert Cedergren, Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada. [10]Department of Biology/Plant Physiology, Martin-Luther-University Halle-Wittenberg, Germany. [11]Max F. Perutz Laboratories, Department of Biochemistry and Cell Biology, University of Vienna, Vienna, Austria. [12]Department of Civil and Environmental Engineering, Colorado School of Mines, Golden, CO 80401, USA. [13]Department of Chemistry and Geochemistry, Colorado School of Mines, Golden, CO 80401, USA. [14]Unité de Glycobiologie Structurale et Fonctionnelle, UMR 8576 CNRS-USTL, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq Cedex, France. [15]Universités Aix-Marseille I & II, Case 932, 163 Avenue de Luminy, 13288 Marseille Cedex 9, France. [16]Faculty of Biology, University of Freiburg, Freiburg, Germany. [17]FRISYS Freiburg Initiative in Systems Biology, University of Freiburg, Freiburg, Germany. [18]BIOSS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany. [19]Roy J. Carver Center for Genomics, Department of Biology, University of Iowa, Iowa City, IA, 52242, USA. [20]Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [21]Genome Project Solutions, Hercules, CA 94547, USA. [22]Department of Integrative Biology, University of California, Berkeley, CA 94720, USA.

*Present address: Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia.
†To whom correspondence should be addressed. E-mail: bhattacharya@aesop.rutgers.edu

Downloaded from www.sciencemag.org on February 21, 2012

Plantae is that encoding a thiamine pyrophosphate–dependent pyruvate decarboxylase family protein involved in alcohol fermentation (Fig. 1D). Another 60 genes were present in only two of the tree phyla (i.e., 24, 10, and 26 genes in Glaucophyta-Viridiplantae, Glaucophyta-Rhodophyta, and Rhodophyta-Viridiplantae, respectively) (*22*).

We sequenced the mitochondrial genome from *C. paradoxa* and from the distantly related glaucophyte *Glaucocystis nostochinearum* and generated a near-complete plastid genome sequence from *G. nostochinearum* that was added to the existing plastid genome data from *C. paradoxa* (GenBank NC_001675). The mitochondrial DNAs (mtDNAs) of *C. paradoxa* and *G. nostochinearum* share similar characteristics, including a large gene content, but differ markedly in size [Fig. 3A, tables S3 and S4, and discussion in the supporting online material (SOM)]. A concatenated multiprotein (17,049 aligned amino acid positions) phylogeny of the plastid data shows the expected monophyly of Plantae plastids (*2*, *5*, *6*) and places glaucophytes very close to the divergence point of red and green algae (fig. S5A).

Building on the support for Plantae monophyly provided by the phylogenomic analyses, we analyzed the evolution of landmark traits that are associated with plastid endosymbiosis. Previous work suggests that in red and green algae, the initial link for carbon metabolism between the

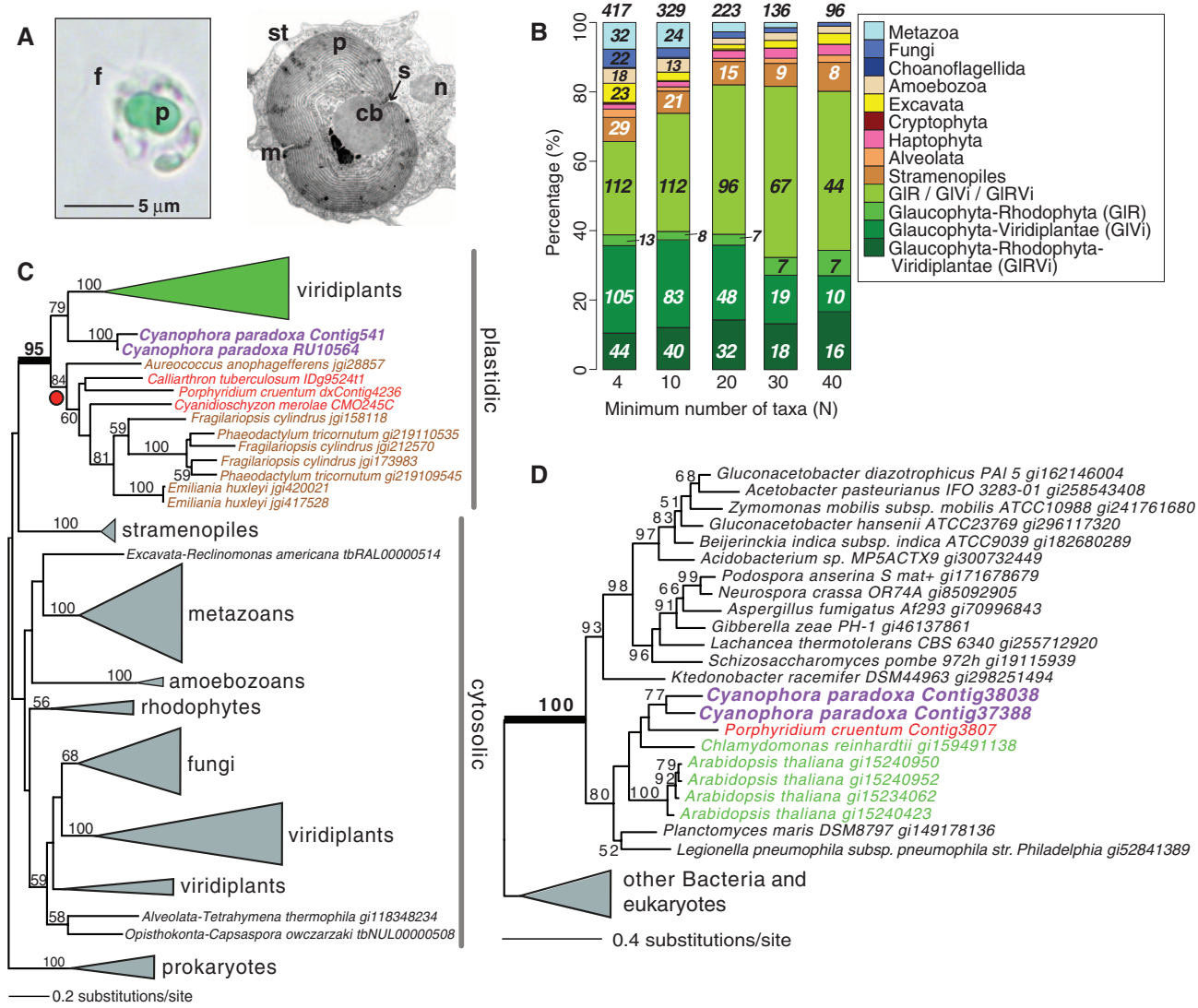

Fig. 1. (**A**) Schematic (left) and transmission electron micrograph (right) images of *C. paradoxa*. This freshwater microalga has two flagella (f) of unequal lengths, and the plastid (p) contains an electron-dense central body (cb). The nucleus (n), mitochondrion (m), starch granules in the cytoplasm (st), and peptidoglycan septa (s) formed during plastid division are also shown. (**B**) Percentage of single-protein Randomized Axelerated Maximum Likelihood (RaxML) trees (raw numbers shown in the bars) that support the monophyly of Glaucophyta (bootstrap ≥ 90%) solely with other Plantae members, or in combination with non-Plantae taxa that interrupt this clade. These latter groups of trees are primarily explained by red/green algal EGT into the nuclear genome of chromalveolates and euglenids. For each of these algal lineages, the set of trees with different numbers of taxa (*N*) ≥4, ≥10, ≥20, ≥30, and ≥40 and distinct phyla ≥3 in a tree are shown. The Plantae-only groups are Glaucophyta-Rhodophyta (GlR), Glaucophyta-Viridiplantae (GlVi), and Glaucophyta-Rhodophyta-Viridiplantae (GlRVi). Trees with evidence of EGT are shown as the single group, GlR/GlVi/GlRVi. An expanded analysis that shows the results when red or green algae are used as the query to search for support for Plantae monophyly is summarized in fig. S1. (**C**) Schematic ML phylogeny of fructose-1,6-bisphosphatase, an enzyme with cytosolic and plastidic isoforms that unites Plantae (plastid-targeted protein) and shows an example of a protein affected by EGT. The plastidic gene has been transferred from red algae to chromalveolates that contain a red algal–derived plastid, presumably through EGT (marked by the filled red circle). The full tree is shown in fig. S2. (**D**) Schematic ML phylogeny of a gene encoding a thiamine pyrophosphate (TPP)–dependent pyruvate decarboxylase family protein involved in alcohol fermentation. RAxML bootstrap support values are shown at the nodes of the trees in panels (C) and (D), in which glaucophytes, red algae, green algae, and chromalveolates are in purple, red, green, and brown, respectively.

Downloaded from www.sciencemag.org on February 21, 2012

host cell and plastid relied on sugar-phosphate transporters that evolved from existing host endomembrane nucleotide sugar transporters (NSTs) (*15*, *25*). Unexpectedly, we found that although six endomembrane-type NST genes exist in *C. paradoxa* (fig. S3), this alga lacks plastidial phosphate-translocator (PT) genes (Fig. 2A). We searched the genome for alternative candidate genes encoding homologs of bacterial carbon exporters and found two UhpC-type hexose-phosphate transporters (Fig. 2B) that are also present in red and green algae. One of these putative transporters (contig 37408) encodes a plastid-targeting signal containing a conserved phenylalanine (F) and proline (P) at the N terminus typical for *C. paradoxa* plastid-targeted proteins [e.g., (*23*)] (Fig. 2, B and C). The second UhpC homolog contains an F and a P at the N terminus and may also be plastid-targeted. These transporters are related to sequences in the parasites Chlamydiae and *Legionella*. We suggest that the UhpC gene originated in Plantae and *Legionella* species through independent HGT events from Chlamydiae. This direction of gene transfer for the prokaryotes is supported by the nested position of a single proteobacterial clade (i.e., *Legionella*) within Chlamydiae. Chlamydiae have contributed a substantial number of other genes to Plantae that are involved in plastid functions (*26*, *27*).

A second landmark trait of photosynthetic eukaryotes is the presence of protein-conducting channels (translocons) in the outer and inner envelope membranes of plastids (Toc and Tic, respectively) [e.g., (*14*)], for which there is biochemical evidence in the *C. paradoxa* plastid (*28*). We identified nuclear-encoded homologs of the major Toc75 and Tic110 translocon proteins, two Toc34-like receptors, homologs of the plastid Hsp70 and Hsp93 chaperones, and a stromal processing peptidase (table S2). This minimal set of components likely formed the ancestral plastid protein translocation system in Plantae (*29*). Candidates for additional translocon subunits Tic20 to Tic22, Tic32, Tic55, and Tic62 were also found in the glaucophyte. The presence of a conserved core of cyanobacterium-derived, homologous translocon subunits shared by *C. paradoxa* and other Plantae (i.e., Toc75, Tic20, and Tic22) or those apparently evolved de novo in the host (i.e., Toc34 and Tic110) (table S2 and fig. S4A) provides strong evidence that the primary plastid was established in a single ancestor of Plantae (*14*, *29*, *30*).

A key component of plastids is light-harvesting complex (LHC) proteins that increase the capacity to capture incoming light. Cyanobacteria, red algae, and glaucophytes use phycobilisomes (protein complexes anchored to the thylakoid) for light harvesting, but red algae also have nuclear-encoded pigment-binding proteins homologous to the typical green algal LHCs but specifically associated with photosystem I (*31*). Unexpectedly, no candidate genes with three membrane-spanning regions, characteristic of LHCs in all other photosynthetic eukaryotes, were found in *C. paradoxa*. This alga does, however, encode several members of the LHC-like (LIL) protein family that contain a single chlorophyll-binding transmembrane helix, including two copies of the "one-helix proteins" (OHPs) (see fig. S8 and SOM).

The anaerobic capabilities of *C. paradoxa* provide parallels to the well-characterized anoxic metabolism of the green alga *Chlamydomonas reinhardtii*

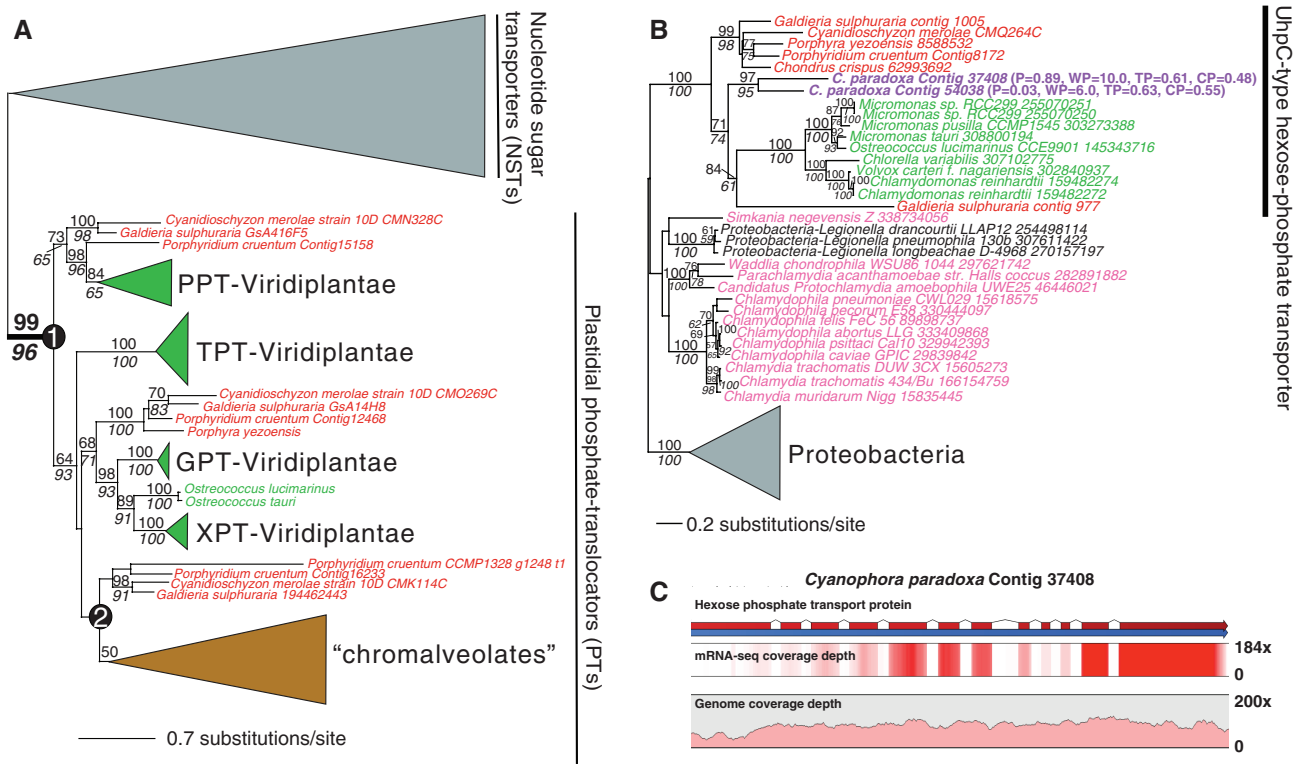

Fig. 2. (**A**) Maximum likelihood (PhyML) phylogeny of endomembrane (NST) and plastidial phosphate-translocators (PTs). PhyML bootstrap values are shown above the branches, and RAxML bootstrap values are shown below the branches in italics. Only bootstrap values ≥50% are shown. Red algae, Viridiplantae, and chromalveolates are shown in red, green, and brown, respectively. The different plastid-targeted transporters are glucose 6-phosphate translocator (GPT), phosphoenolpyruvate translocator (PPT), triose phosphate translocator (TPT), and xylulose 5-phosphate translocator (XPT). The numbers in the filled circles indicate transporters shown to the right that (1) resulted from primary endosymbiosis and (2) were transferred from the red algal secondary endosymbiont to the chromalveolate host. The full tree is shown in fig. S3. (**B**) RAxML phylogeny of UhpC-type hexose-phosphate transporters in algae, land plants, and Bacteria. RAxML bootstrap values are shown above the branches, and PhyML bootstrap values are shown below the branches in italics. Only bootstrap values ≥50% are shown. Red algae, Viridiplantae, glaucophytes, and Chlamydiae are shown in red, green, magenta, and pink, respectively. The numbers after the *C. paradoxa* contig names are bioinformatic predictions for plastid targeting of the encoded proteins with Predotar (P), WoLF PSORT (WP), TargetP (TP), and ChloroP (CP). These data suggest that contig 37408 (and perhaps also contig 54308) is destined for the *C. paradoxa* plastid. (**C**) Intron distribution, genome contig coverage, and transcriptome coverage (mRNA-seq) of *C. paradoxa* genome contig 37408 that encodes a putative plastid-targeted UhpC-type hexose phosphate transporter.

Downloaded from www.sciencemag.org on February 21, 2012

(*32*). The predicted pathways in *C. paradoxa* suggest a complex heterofermentative state in the Plantae ancestor (Fig. 3B), with retention in Chlorophyta (e.g., Chlorophyceae and Trebouxiophyceae) and losses in Streptophyta. Orthologs of fermentation enzymes encoded by the *C. paradoxa* genome can also be found in Cyanobacteria (for lactate, formate, acetate, and ethanol) and hydrogenosomal eukaryotes (for H_2, lactate, formate, and ethanol). Therefore, the complex fermentative capabilities conserved between the distantly related *C. paradoxa* and green algae (*33*) likely represent an evolutionarily advantageous combination of anoxic enzymes from the eukaryote host and the cyanobacterial endosymbiont. Genome evidence for H_2 metabolism in *C. paradoxa* is demonstrated by the presence of a [FeFe]-hydrogenase and associated maturases (HydE, HydF, and HydG).

Putative carbohydrate metabolism enzymes in *C. paradoxa* were identified using the Carbohydrate-Active enZymes (CAZy) database (*34*) annotation

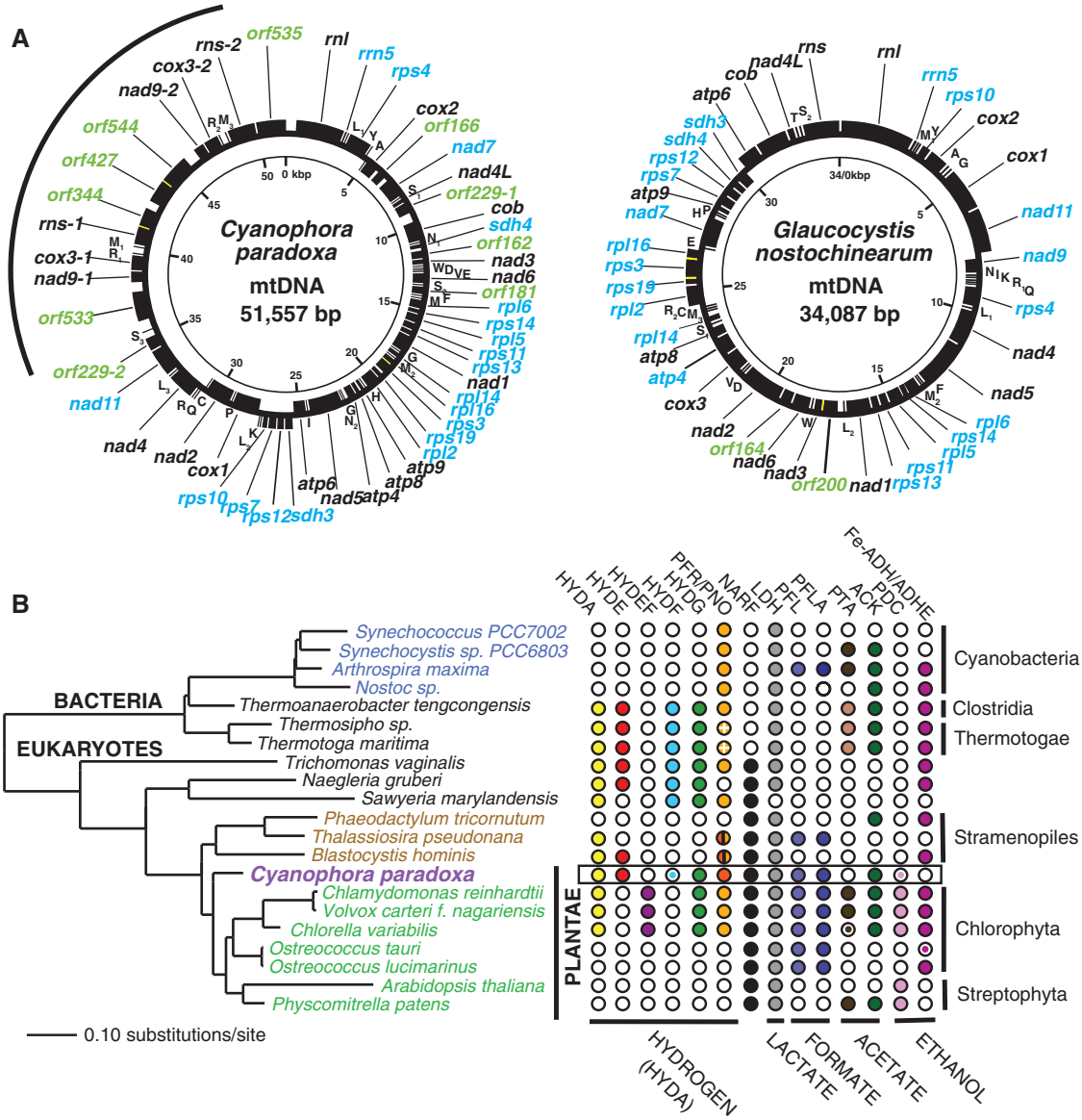

Fig. 3. (**A**) Gene maps of glaucophyte mitochondrial genomes. Black blocks indicate genes that are arranged outside and inside the circle, indicating they are transcribed clockwise and counterclockwise, respectively. tRNA gene names are abbreviated in single–amino acid letter code; subscript numbers distinguish isoacceptor tRNAs. The color-coding identifies genes that are typically present in fungal and animal mtDNAs (black); additional genes found in land plants and protists are marked in blue and unidentified ORFs in green. The arc marks the genome region with long segmental duplications in *C. paradoxa* mtDNA. (**B**) Phylogeny of small subunit rRNA of organisms with sequenced genomes that share predicted fermentative orthologs encoded in the *C. paradoxa* genome. Predicted orthologs encoded by *C. paradoxa* are shown boxed. Enzyme abbreviations are as follows: Fe-ADH, iron-containing alcohol dehydrogenase; ADHE, iron-containing alcohol/aldehyde dehydrogenase; ACK, acetate kinase; HYDA, [FeFe]-hydrogenase; HYDE, radical S-adenosylmethionine HYDA maturase; HYDF, GTPase-domain HYDA maturase; HYDEF, fused HYDE and HYDF maturase; HYDG, radical S-adenosylmethionine HYDA maturase; LDH, lactate dehydrogenase; NARF, nuclear prelamin A recognition factor; PTA, phosphotransacetylase; PDC, pyruvate decarboxylase; PFL, pyruvate:formate lyase; PFLA, pyruvate:formate lyase activase; PFR, pyruvate:ferredoxin oxidoreductase; PNO, pyruvate:NADP oxidoreductase composed of a PFR domain fused to a C-terminal NADPH-cytochrome P450 reductase domain. Filled circles indicate the presence of the indicated enzyme(s), whereas incompletely filled circles indicate support from a partial *C. paradoxa* sequence. In the case of PFR/PNO, the darker filled circle represents PNO, a divided fill represents evidence for both enzymes, and white crosses indicate the heteromeric PFRs present in the Thermotogae. Shading is also used to discriminate the single-domain type 1 (light) and multidomain type II (dark) PTA isoforms. Single-letter abbreviations for the amino acids are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

Downloaded from www.sciencemag.org on February 21, 2012

Table 1. List of CAZymes (*34*) present in the *C. paradoxa* genome. GH, glycoside hydrolase; GT, glycosyl transferase; PL, polysaccharide lyase; CE, carbohydrate esterase; CBM, carbohydrate-binding module. Photosynthetic taxa are shown in the green field and nonphotosynthetic taxa in the blue field.

Species	GH	GT	PL	CE	CBM	GH+GT
Cyanophora paradoxa (glaucophyte)	84	128	3	2	24	212
Cyanidioschyzon merolae (red alga)	21	61	0	2	16	82
Ostreococcus lucimarinus CCE9901 (green alga)	30	69	0	3	23	99
Arabidopsis thaliana (land plant)	400	468	34	110	126	868
Crocosphaera watsonii WH8501 (cyanobacterium)	18	92	0	4	7	110
Dictyostelium discoideum (amoebozoan)	75	69	0	3	34	144
Saccharomyces cerevisiae S288C (yeast)	56	68	0	3	12	124

pipeline (*22*). The genome of this alga encodes ~84 glycoside hydrolases (GHs) and 128 glycosyl transferases (GTs). This is far greater than in the marine green microalga *Ostreococcus lucimarinus* CCE9901 and the extremophilic red alga *Cyanidioschyzon merolae* but less than in *Arabidopsis thaliana* (Table 1). Consistent with these results, inspection of the number of CAZy families (*22*) shows that *C. paradoxa* contains twice the number of GH families and 20% more GT families than *O. lucimarinus* and *C. merolae*. The smaller number of CAZy families in these unicellular algal lineages when compared to *A. thaliana* suggests less functional redundancy.

Many *C. paradoxa* CAZymes are involved in starch metabolism (*22*). Synthesis of the polysaccharide within Viridiplantae plastids relies on enzymes of the GT5 CAZy (*34*) family associated with glycogen synthesis in Bacteria such as adenosine diphosphate (ADP)–glucose pyrophosphorylases and ADP-glucose using starch synthases (SSs). Eukaryotes synthesize glycogen from uridine diphosphate (UDP)–glucose using either a GT3 (all fungi, some Amoebozoa, and animals) or a GT5 (alveolates, parabasalids, and Amoebozoa) type of transferase. The major *C. paradoxa* enzyme is phylogenetically related to the GT5 UDP-glucose–specific enzyme of heterotrophic eukaryotes (*35*) and has been partially purified from this alga (*36*). This suggests the absence of ADP-glucose pyrophosphorylase in *C. paradoxa*. Therefore, it was unexpected to find a second gene in the glaucophyte genome whose gene product is related to the SSIII-SSIV (GT5) type of starch synthases in Viridiplantae. This gene is phylogenetically related to glucan synthase in Chlamydiae, Cyanobacteria, and some Proteobacteria (fig. S5B) and likely played a key role in linking the biochemistry of the host and the endosymbiont. The SSIII-SSIV enzyme uses ADP-glucose in Bacteria and land plants (*35*), suggesting that *C. paradoxa* or, alternatively, the common ancestor of Viridiplantae and glaucophytes may have used both types of nucleotide sugars for starch synthesis.

Analysis of the gene-rich *C. paradoxa* genome unambiguously supports Plantae monophyly (*2*, *5*, *6*, *14*, *15*) (see discussion of Plantae branching order in the SOM), laying to rest a long-standing issue in eukaryote evolution. Plantae share many genes with an EGT or HGT origin that have essential functions such as photosynthesis, starch biosynthesis, plastid protein import, plastid solute transport, and alcohol fermentation. The alternative explanation of a polyphyletic Plantae would require the unlikely combination of a large number of independent HGT events in its major phyla followed by gene loss in all (or many) other eukaryotes. Consolidation of the Plantae allows insights into the gene inventory of their common ancestor. It is now clear that the Plantae ancestor contained many of the key innovations that characterize land plant and algal genomes, including extensive EGT from the cyanobacterial endosymbiont and retargeting of plastid-destined proteins, the minimal machinery required for plastid protein translocation, and complex pathways for fermentation and starch biosynthesis. In spite of the fact that glaucophytes retain peptidoglycan, an ancestral trait lost by all other algae (except the primary plastid in *Paulinella*) (*37*) and land plants, glaucophytes are not a lineage of "living fossils." Rather, the *C. paradoxa* genome contains a unique combination of ancestral, novel, and "borrowed" (e.g., via HGT) genes, similar to the genomes of other Plantae (*38*).

References and Notes

1. L. Margulis, *Origin of Eukaryotic Cells* (Yale University Press, New Haven, CT, 1970).
2. A. Reyes-Prieto, A. P. Weber, D. Bhattacharya, *Annu. Rev. Genet.* **41**, 147 (2007).
3. T. Cavalier-Smith, *Biosystems* **14**, 461 (1981).
4. S. M. Adl *et al.*, *J. Eukaryot. Microbiol.* **52**, 399 (2005).
5. N. Rodríguez-Ezpeleta *et al.*, *Curr. Biol.* **15**, 1325 (2005).
6. C. X. Chan *et al.*, *Curr. Biol.* **21**, 328 (2011).
7. H. S. Yoon, J. D. Hackett, G. Pinto, D. Bhattacharya, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15507 (2002).
8. F. Burki *et al.*, *PLoS ONE* **2**, e790 (2007).
9. D. Baurain *et al.*, *Mol. Biol. Evol.* **27**, 1698 (2010).
10. L. W. Parfrey *et al.*, *Syst. Biol.* **59**, 518 (2010).
11. J. W. Stiller, *Trends Plant Sci.* **12**, 391 (2007).
12. W. Löffelhardt, H. J. Bohnert, D. A. Bryant, *Plant Syst. Evol.* **11**, 149 (1997).
13. B. Pfanzagl *et al.*, *J. Bacteriol.* **178**, 332 (1996).
14. G. I. McFadden, G. G. van Dooren, *Curr. Biol.* **14**, R514 (2004).
15. A. P. M. Weber, M. Linka, D. Bhattacharya, *Eukaryot. Cell* **5**, 609 (2006).
16. A. Moustafa *et al.*, *Science* **324**, 1724 (2009).
17. W. Martin, R. G. Herrmann, *Plant Physiol.* **118**, 9 (1998).
18. J. T. Harper, P. J. Keeling, *Mol. Biol. Evol.* **20**, 1730 (2003).
19. W. Martin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12246 (2002).
20. N. Sato, M. Ishikawa, M. Fujiwara, K. Sonoike, *Genome Inform.* **16**, 56 (2005).
21. A. Moustafa, D. Bhattacharya, *BMC Evol. Biol.* **8**, 6 (2008).
22. The *Cyanophora paradoxa* Genome Project (http://dblab.rutgers.edu/cyanophora/home.php).
23. T. Wunder, R. Martin, W. Löffelhardt, E. Schleiff, J. M. Steiner, *BMC Evol. Biol.* **7**, 236 (2007).
24. J. M. Steiner, W. Löffelhardt, *Mol. Membr. Biol.* **22**, 123 (2005) (review).
25. C. Colleoni *et al.*, *Mol. Biol. Evol.* **27**, 2691 (2010).
26. J. Huang, J. P. Gogarten, *Genome Biol.* **8**, R99 (2007).
27. A. Moustafa, A. Reyes-Prieto, D. Bhattacharya, *PLoS ONE* **3**, e2205 (2008).
28. F. Yusa, J. M. Steiner, W. Löffelhardt, *BMC Evol. Biol.* **8**, 304 (2008).
29. J. Gross, D. Bhattacharya, *Trends Plant Sci.* **14**, 13 (2009).
30. S. Reumann, K. Inoue, K. Keegstra, *Mol. Membr. Biol.* **22**, 73 (2005) (review).
31. G. R. Wolfe, F. X. Cunningham, D. G. Durnford, B. R. Green, E. Gantt, *Nature* **367**, 566 (1994).
32. M. C. Posewitz, A. Dubini, J. Meuser, M. Seibert, M. Ghirardi, in *The Chlamydomonas Sourcebook: Organellar and Metabolic Processes*, D. Stern, Ed. (Academic Press, Oxford, ed. 2, 2009), vol. 2, pp. 217–255.
33. J. E. Meuser *et al.*, *J. Biotechnol.* **142**, 21 (2009).
34. B. L. Cantarel *et al.*, *Nucleic Acids Res.* **37** (Database issue), D233 (2009).
35. C. Plancke *et al.*, *Eukaryot. Cell* **7**, 247 (2008).
36. P. Deschamps, H. Moreau, A. Z. Worden, D. Dauvillée, S. G. Ball, *Genetics* **178**, 2373 (2008).
37. A. Reyes-Prieto *et al.*, *Mol. Biol. Evol.* **27**, 1530 (2010).
38. S. S. Merchant *et al.*, *Science* **318**, 245 (2007).

Acknowledgments: This work was made possible by grants from the National Science Foundation (MGSP 0625440 and MCB 0946528) awarded to D.B. and J.B. Partial support came from NSF grants EF 0827023 and DEB 0936884 awarded to D.B. and H.S.Y. and by BioGreen 21 (PJ008177) Rural Development Administration of South Korea awarded to H.S.Y. Air Force Office of Scientific Research grant FA9550-11-1-0211 supported work by J.E.M. and M.C.P. A.P.M.W. appreciates support from Deutsche Forschungsgemeinschaft CRC TR1. W.L. acknowledges support from the Austrian Science Foundation (P19683). G.B. acknowledges funding from Canadian Institute for Health Research operating grant MSP-14226. S.R. and A.S. acknowledge funding from the German Federal Ministry of Education and Research. G.B. acknowledges the Natural Sciences and Engineering Research Council of Canada (NSERC) for general laboratory funding. We are grateful to B. Andersen and the staff at the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP) for preparing the axenic culture of *C. paradoxa*. We thank I. Plante and L. Forget for technical assistance in mtDNA cloning and sequencing and N. Beck for development of the MFannot computer program. The sequence data used to assemble the draft *C. paradoxa* genome and the Illumina mRNA-seq reads from this alga are archived at the NCBI Sequence Read Archive (SRA) under accession number SRP00920. The assembled genome and cDNA contigs, gene models, gene annotations, supporting material that is not in the SOM, and the phylogenomic results are available at (*22*). The complete *C. paradoxa* and *G. nostochinearum* mitochondrial genomes are available at NCBI under accession numbers HQ849544 and NC_015117, respectively. The partial mitochondrial genome data from *G. nostochinearum* are available at the genome Web site as part of the aligned protein data set used to infer the plastid multigene tree. We are grateful for the helpful comments of two anonymous reviewers of this manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/335/6070/843/DC1
Materials and Methods
SOM Text
Figs. S1 to S9
Tables S1 to S4
References

5 September 2011; accepted 5 January 2012
10.1126/science.1213561

Downloaded from www.sciencemag.org on February 21, 2012