Published in final edited form as:
Science. 2007 October 12; 318(5848): 245–250. doi:10.1126/science.1143609.

The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions

Sabeeha S. Merchant[1,*], **Simon E. Prochnik**[2,*], **Olivier Vallon**[3], **Elizabeth H. Harris**[4], **Steven J. Karpowicz**[1], **George B. Witman**[5], **Astrid Terry**[2], **Asaf Salamov**[2], **Lillian K. Fritz-Laylin**[6], **Laurence Maréchal-Drouard**[7], **Wallace F. Marshall**[8], **Liang-Hu Qu**[9], **David R. Nelson**[10], **Anton A. Sanderfoot**[11], **Martin H. Spalding**[12], **Vladimir V. Kapitonov**[13], **Qinghu Ren**[14], **Patrick Ferris**[15], **Erika Lindquist**[2], **Harris Shapiro**[2], **Susan M. Lucas**[2], **Jane Grimwood**[16], **Jeremy Schmutz**[16], **Pierre Cardol**[3,18], **Heriberto Cerutti**[19], **Guillaume Chanfreau**[1], **Chun-Long Chen**[9], **Valérie Cognat**[7], **Martin T. Croft**[20], **Rachel Dent**[21], **Susan Dutcher**[22], **Emilio Fernández**[23], **Patrick Ferris**[15], **Hideya Fukuzawa**[24], **David González-Ballester**[17], **Diego González-Halphen**[25], **Armin Hallmann**[26], **Marc Hanikenne**[18], **Michael Hippler**[27], **William Inwood**[21], **Kamel Jabbari**[28], **Ming Kalanon**[29], **Richard Kuras**[3], **Paul A. Lefebvre**[11], **Stéphane D. Lemaire**[30], **Alexey V. Lobanov**[31], **Martin Lohr**[32], **Andrea Manuell**[33], **Iris Meier**[34], **Laurens Mets**[35], **Maria Mittag**[36], **Telsa Mittelmeier**[37], **James V. Moroney**[38], **Jeffrey Moseley**[17], **Carolyn Napoli**[39], **Aurora M. Nedelcu**[40], **Krishna Niyogi**[21], **Sergey V. Novoselov**[31], **Ian T. Paulsen**[14], **Greg Pazour**[41], **Saul Purton**[42], **Jean-Philippe Ral**[43], **Diego Mauricio Riaño-Pachón**[44], **Wayne Riekhof**[45], **Linda Rymarquis**[46], **Michael Schroda**[47], **David Stern**[48], **James Umen**[15], **Robert Willows**[49], **Nedra Wilson**[50], **Sara Lana Zimmer**[48], **Jens Allmer**[51], **Janneke Balk**[20], **Katerina Bisova**[52], **Chong-Jian Chen**[9], **Marek Elias**[53], **Karla Gendler**[39], **Charles Hauser**[54], **Mary Rose Lamb**[55], **Heidi Ledford**[21], **Joanne C. Long**[1], **Jun Minagawa**[56], **M. Dudley Page**[1], **Junmin Pan**[57], **Wirulda Pootakham**[17], **Sanja Roje**[58], **Annkatrin Rose**[59], **Eric Stahlberg**[34], **Aimee M. Terauchi**[1], **Pinfen Yang**[60], **Steven Ball**[61], **Chris Bowler**[28,62], **Carol L. Dieckmann**[37], **Vadim N. Gladyshev**[31], **Pamela Green**[46], **Richard Jorgensen**[39], **Stephen Mayfield**[33], **Bernd Mueller-Roeber**[44], **Sathish Rajamani**[63], **Richard T. Sayre**[34], **Peter Brokstein**[2], **Inna Dubchak**[2], **David Goodstein**[2], **Leila Hornick**[2], **Y. Wayne Huang**[2], **Jinal Jhaveri**[2], **Yigong Luo**[2], **Diego Martínez**[2], **Wing Chi Abby Ngau**[2], **Bobby Otillar**[2], **Alexander Poliakov**[2], **Aaron Porter**[2], **Lukasz Szajkowski**[2], **Gregory Werner**[2], **Kemin Zhou**[2], **Igor V. Grigoriev**[2], **Daniel S. Rokhsar**[2,6,‡], and **Arthur R. Grossman**[17,‡]

[1]Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA 90095, USA

[2]U.S. Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA

[3]CNRS, UMR 7141, CNRS/Université Paris 6, Institut de Biologie Physico-Chimique, 75005 Paris, France

[4]Department of Biology, Duke University, Durham, North Carolina 27708, USA

[‡]To whom correspondence should be addressed. dsrokhsar@lbl.gov (D.S.R.); arthurg@stanford.edu (A.R.G.).
[*]These authors contributed equally to this work.

Supporting Online Material
www.sciencemag.org/cgi/content/full/318/5848/245/DC1
Materials and Methods
SOM Text
Figs. S1 to S25
Tables S1 to S14
References and Notes

[5]Department of Cell Biology, University of Massachusetts Medical School, Worcester, MA 01655, USA

[6]Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA94720, USA

[7]Institut de Biologie Moléculaire des Plantes, CNRS, 67084 Strasbourg Cedex, France

[8]Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143, USA

[9]Biotechnology Research Center, Zhongshan University, Guangzhou 510275, China

[10]Department of Molecular Sciences and Center of Excellence in Genomics and Bioinformatics, University of Tennessee, Memphis, TN 38163, USA

[11]Department of Plant Biology, University of Minnesota, St. Paul MN 55108, USA

[12]Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011, USA

[13]Genetic Information Research Institute, Mountain View, CA 94043, USA

[14]The Institute for Genomic Research, Rockville, MD 20850, USA

[15]Plant Biology Laboratory, Salk Institute, La Jolla, CA 92037, USA

[16]Stanford Human Genome Center, Stanford University School of Medicine, Palo Alto, CA 94304, USA

[17]Department of Plant Biology, Carnegie Institution, Stanford, CA 94306, USA

[18]Plant Biology Institute, Department of Life Sciences, University of Liège, B-4000 Liège, Belgium

[19]University of Nebraska-Lincoln, School of Biological Sciences–Plant Science Initiative, Lincoln, NE 68588, USA

[20]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK

[21]Department of Plant and Microbial Biology, University of California at Berkeley, Berkeley, CA 94720, USA

[22]Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA

[23]Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad de Córdoba, Campus de Rabanales, 14071 Córdoba, Spain

[24]Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan

[25]Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, México 04510 DF, Mexico

[26]Department of Cellular and Developmental Biology of Plants, University of Bielefeld, D-33615 Bielefeld, Germany

[27]Department of Biology, Institute of Plant Biochemistry and Biotechnology, University of Münster, 48143 Münster, Germany

[28]CNRS UMR 8186, Département de Biologie, Ecole Normale Supérieure, 75230 Paris, France

[29]Plant Cell Biology Research Centre, The School of Botany, The University of Melbourne, Parkville, Melbourne, VIC 3010, Australia

[30]Institut de Biotechnologie des Plantes, UMR 8618, CNRS/Université Paris-Sud, Orsay, France

[31]Department of Biochemistry, N151 Beadle Center, University of Nebraska, Lincoln, NE 68588–0664, USA

[32]Institut für Allgemeine Botanik, Johannes Gutenberg-Universität, 55099 Mainz, Germany

[33]Department of Cell Biology and Skaggs Institute for Chemical Biology, Scripps Research Institute, La Jolla, CA 92037, USA

[34]PCMB and Plant Biotechnology Center, Ohio State University, Columbus, OH 43210, USA

[35]Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637, USA

[36]Institut für Allgemeine Botanik und Pflanzenphysiologie, Friedrich-Schiller-Universität Jena, 07743 Jena, Germany

[37]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA

[38]Department of Biological Science, Louisiana State University, Baton Rouge, LA 70803, USA

[39]Department of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA

[40]Department of Biology, University of New Brunswick, Fredericton, NB, Canada E3B 6E1

[41]Department of Physiology, University of Massachusetts Medical School, Worcester, MA 01605, USA

[42]Department of Biology, University College London, London WC1E 6BT, UK

[43]Unité de Glycobiologie Structurale et Fonctionnelle, UMR8576 CNRS/USTL, IFR 118, Université des Sciences et Technologies de Lille, Cedex, France

[44]Universität Potsdam, Institut für Biochemie und Biologie, D-14476 Golm, Germany

[45]Department of Medicine, National Jewish Medical and Research Center, Denver, CO 80206, USA

[46]Delaware Biotechnology Institute, University of Delaware, Newark, DE 19711, USA

[47]Institute of Biology II/Plant Biochemistry, 79104 Freiburg, Germany

[48]Boyce Thompson Institute for Plant Research at Cornell University, Ithaca, NY 14853, USA

[49]Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney 2109, Australia

[50]Department of Anatomy and Cell Biology, Oklahoma State University, Center for Health Sciences, Tulsa, OK 74107, USA

[51]Izmir Ekonomi Universitesi, 35330 Balcova-Izmir Turkey

[52]Institute of Microbiology, Czech Academy of Sciences, Czech Republic

[53]Department of Plant Physiology, Faculty of Sciences, Charles University, 128 44 Prague 2, Czech Republic

[54]Bioinformatics Program, St. Edward's University, Austin, TX 78704, USA

[55]Department of Biology, University of Puget Sound, Tacoma, WA 98407, USA

[56]Institute of Low-Temperature Science, Hokkaido University, 060-0819, Japan

[57]Department of Biology, Tsinghua University, Beijing, China 100084

[58]Institute of Biological Chemistry, Washington State University, Pullman, WA 99164, USA

[59]Appalachian State University, Boone, NC 28608, USA

[60]Department of Biology, Marquette University, Milwaukee, WI 53233, USA

[61]UMR8576 CNRS, Laboratory of Biological Chemistry, 59655 Villeneuve d'Ascq, France

[62]Cell Signaling Laboratory, Stazione Zoologica, I 80121 Naples, Italy

[63]Graduate Program in Biophysics, Ohio State University, Columbus, OH 43210, USA

Abstract

Chlamydomonas reinhardtii is a unicellular green alga whose lineage diverged from land plants over 1 billion years ago. It is a model system for studying chloroplast-based photosynthesis, as well as the structure, assembly, and function of eukaryotic flagella (cilia), which were inherited from the common ancestor of plants and animals, but lost in land plants. We sequenced the ~120-megabase nuclear genome of *Chlamydomonas* and performed comparative phylogenomic analyses, identifying genes encoding uncharacterized proteins that are likely associated with the function and biogenesis of chloroplasts or eukaryotic flagella. Analyses of the *Chlamydomonas* genome advance our understanding of the ancestral eukaryotic cell, reveal previously unknown genes associated with photosynthetic and flagellar functions, and establish links between ciliopathy and the composition and function of flagella.

Chlamydomonas reinhardtii is a ~10-μm, unicellular, soil-dwelling green alga with multiple mitochondria, two anterior flagella for motility and mating, and a chloroplast that houses the photosynthetic apparatus and critical metabolic pathways (Fig. 1 and fig. S1) (1). *Chlamydomonas* is used to study eukaryotic photosynthesis because, unlike angiosperms (flowering plants), it grows in the dark on an organic carbon source while maintaining a functional photosynthetic apparatus (2). It also is a model for elucidating eukaryotic flagella and basal body functions and the pathological effects of their dysfunction (3,4). More recently, *Chlamydomonas* research has been developed for bioremediation purposes and the generation of biofuels (5,6).

The Chlorophytes (green algae, including *Chlamydomonas* and *Ostreococcus)* diverged from the Streptophytes (land plants and their close relatives) (Fig. 2) over a billion years ago. These lineages are part of the green plant lineage (Viridiplantae), which previously diverged from opisthokonts (animals, fungi, and Choanozoa) (7). Many *Chlamydomonas* genes can be traced to the green plant or plant-animal common ancestor by comparative genomic analyses. Specifically, many *Chlamydomonas* and angiosperm genes are derived from ancestral green plant genes, including those associated with photosynthesis and plastid function; these are also present in *Ostreococcus* spp. and the moss *Physcomitrella patens* (Fig. 2). Genes shared by *Chlamydomonas* and animals are derived from the last plant-animal common ancestor and many of these have been lost in angiosperms, notably those encoding proteins of the eukaryotic flagellum (or cilium) and the associated basal body (or centriole) (8). *Chlamydomonas* also displays extensive metabolic flexibility under the control of regulatory genes that allow it to inhabit distinct environmental niches and to survive fluctuations in nutrient availability (9).

Genome sequencing and assembly

The 121-megabase (Mb) draft sequence (10) of the *Chlamydomonas* nuclear genome was generated at 13× coverage by whole-genome, shotgun end-sequencing of plasmid and fosmid libraries, followed by assembly into ~1500 scaffolds (1). Half of the assembled genome is contained in 25 scaffolds, each longer than 1.63 Mb. The genome is unusually GC-rich (64%) (Table 1), which required modification of standard sequencing protocols. Alignments of expressed sequence tags (ESTs) to the genome suggest that the draft assembly is 95% complete (1).

The *Chlamydomonas* nuclear genome comprises 17 linkage groups (figs. S2 to S18) presumably corresponding to 17 chromosomes, consistent with electron microscopy of meiotic synaptonemal complexes (11). Seventy-four scaffolds, representing 78% of the draft genome, have been aligned with linkage groups (Fig. 3 and figs. S2 to S18). Sequenced ESTs from a field isolate (1) of *Chlamydomonas*, fertile with the standard laboratory strain, identified 8775 polymorphisms, resulting in a marker density of 1 per 13 kb (12, 13). By comparing physical marker locations on scaffolds with genetic recombination distances, we estimated 100 kb per centimorgan (cM) on average.

The *Chlamydomonas* genome has approximately uniform densities of genes, simple sequence repeats, and transposable elements. Several AT-rich islands coincide with gene- and transposable element–poor regions (figs. S2 to S18). As in most eukaryotes, the ribosomal RNA (rRNA) genes are arranged in tandem arrays. They are located on linkage groups I, VII, and XV, although assembly has only been completed on the outermost copies. We identified 259 transfer RNAs (tRNAs) (1) (table S1), 61 classes of simple repeats, ~100 families of transposable elements (1), and 64 tRNA-related short interspersed elements (SINEs) (tables S2 and S3), which is unusual for a microorganism. We also identified tRNAs clusters and a number of recent tRNA duplications (fig. S19), as well as clusters of genes associated with specific biological functions (fig. S20). Few chloroplast and mitochondrial genome fragments were detected in the nuclear genome ("cp" and "mito" in Fig. 3, and figs. S2 to S18).

Protein coding genes and structure

Ab initio and homology-based gene prediction, integrated with EST evidence, was used to create a reference set of 15,143 protein-coding gene predictions (1) (tables S4, S5, and S6). More than 300,000 ESTs were generated from diverse environmental conditions; 8631 gene models (56%) are supported by mRNA or EST evidence (14), and 35% have been edited for gene structure and/or annotated by manual curation, as of June 2007. Protein-coding genes have, on average, 8.3 exons per gene and are intron-rich relative to other unicellular eukaryotes and land plants (15) (fig. S21); only 8% lack introns (Table 1) (1). The average *Chlamydomonas* intron is longer (373 bp) than that of many eukaryotes (16), and the average intron number and size are more similar to those of multicellular organisms than those of protists (fig. S21) (1,17). Only 1.5% of the introns are short (<100 bp), and we did not observe the bimodal intron size distribution typical of most eukaryotes (fig. S21A). Furthermore, 30% of the intron length is due to repeat sequences (1), which suggests that *Chlamydomonas* introns are subject to creation or invasion by transposable elements.

Gene families

We identified 1226 gene families in *Chlamydomonas* encoding two or more proteins (1); of these, 26 families have 10 or more members (table S7). The genes of 317 of the 798 two-gene families are arranged in tandem, which suggests extensive tandem gene duplications. Gene families contain similar proportions of the total gene complement of *Chlamydomonas*, human, and *Arabidopsis*. As in *Arabidopsis*, *Chlamydomonas* has large families of kinases and cytochrome P-450s, but the largest one is the class III guanylyl and adenylyl cyclase family. With 51 members, the *Chlamydomonas* family is larger than that in any other organism (18). Although these cyclases are not found in plants, in animals they catalyze the synthesis of cGMP and cAMP (18), which serve as second messengers in various signal transduction pathways. Cyclic nucleotides are critical for mating processes, as well as flagellar function and regulation in *Chlamydomonas* (19–21), and may be vital for acclimation to changing nutrient conditions (22,23). *Chlamydomonas* also encodes diverse families of proteins critical for nutrient acquisition (23,24).

Transporters

The transporter complement in *Chlamydomonas* suggests that it has retained the diversity present in the common plant-animal ancestor. *Chlamydomonas* is predicted to have 486 membrane transporters (figs. S22 and S23) (1) that fall into the broad classes of 61 ion channels, 124 primary (active) adenosine triphosphate (ATP)–dependent transporters and 293 secondary transporters; eight are unclassified. The 69-member ATP-binding cassette (ABC) and 26-member P-type adenosine triphosphatase (ATPase) families are large, as in *Arabidopsis*, and overall, the complement of transporters in *Chlamydomonas* resembles that of both *Ostreococcus* spp. and land plants (fig. S22). Furthermore, a number of plant transporters not found in animals are encoded on the *Chlamydomonas* genome (fig. S22 and table S8).

We also found copies of genes encoding animal-associated transporter classes, including some with activities related to flagellar function (e.g., the voltage-gated ion channel superfamily) (25) (fig. S22 and table S8). A number of these transporters redistribute intracellular Ca^{2+} in response to environmental signals such as light. Changing Ca^{2+} levels may modulate the activity of the flagella, which are structures found in animals but not in vascular plants (see below).

The *Chlamydomonas* genome also encodes a diversity of substrate-specific transporters that are important for acclimation of the organism to the fluctuating, often nutrient-poor, conditions of soil environments (24). Of the eight sulfate transporters, four are in the H^+/SO_4^{2-} family (characteristic of the plant lineage), three are in the Na^+/SO_4^{2-} family (not found in plants but present in opisthokonts), and one is a bacterial ABC-type SO_4^{2-} transporter (associated with the plastid envelope). The 12-member PiT phosphate transporter and 6-member KUP potassium channel families are larger than in other unicellular eukaryotes, and the former underwent a lineage-specific expansion. *Chlamydomonas* has 11 AMT ammonium transporters, which is only surpassed by the number in rice.

Phylogenomics and the origins of *Chlamydomonas* genes

To explore the evolutionary history of *Chlamydomonas*, we initially compared the *Chlamydomonas* proteome to a representative animal (human) and angiosperm (*Arabidopsis*) proteome (1). We plotted the best matches, calculated on the basis of BLASTP (Basic Local Alignment Search Tool for searching protein collections) scores, of every *Chlamydomonas* protein to the *Arabidopsis* and human proteomes (Fig. 4A). Most *Chlamydomonas* proteins exhibit slightly more similarity to *Arabidopsis* than to human proteins. Many *Chlamydomonas* proteins with greater similarity to animal homologs are present in the flagellar and basal body proteomes (Fig. 4A and below). This is consistent with the maintenance of flagella and basal bodies as cilia and centrioles, respectively, in animals (8), and their loss in angiosperms.

A mutual best-hit analysis of *Chlamydomonas* proteins against proteins from organisms across the tree of life (1) identified 6968 protein families of orthologs, co-orthologs (in the case of recent gene duplications), and paralogs (1). Of the *Chlamydomonas* proteins, 2489 were homologous to proteins from both *Arabidopsis* and humans (Fig. 4B). *Chlamydomonas* and humans shared 706 protein families (774 and 806 proteins, respectively), but these were not shared with *Arabidopsis*. These genes were either lost or diverged beyond recognition in green plants (table S9), and are enriched for sequences encoding cilia and centriole proteins (8,26). Conversely, 1879 protein families are found in both *Chlamydomonas* and *Arabidopsis* (1968 and 2396 proteins, respectively), but lack human homologs. *Chlamydomonas* proteins with homology to plant, but not animal, proteins were either (i) present in the common plant-animal ancestor and retained in *Chlamydomonas* and angiosperms, but lost or diverged in animals; (ii) horizontally transferred into *Chlamydomonas*; or (iii) arose in the plant lineage after divergence

of animals (but before the divergence of *Chlamydomonas*). This set is enriched for proteins that function in chloroplasts (table S9 and below).

The plastid and plant lineages

The plastids of green plants and red algae are primary plastids, i.e., direct descendants from the primary cyanobacterial endosymbiont (27). Diatoms, brown algae, and chlorophyll a– and c–containing algae are also photosynthetic, but their photosynthetic organelles were acquired via a secondary endosymbiosis (28,29). Because of shared ancestry, nucleus-encoded plastid-localized proteins derived from the cyanobacterial endosymbiont are closely related to each other and to cyanobacterial proteins.

We searched the 6968 families that contain *Chlamydomonas* proteins for those that also contained proteins from *Ostreococcus*, *Arabidopsis* and moss, but that did not contain proteins from nonphotosynthetic organisms. The search identified 349 families, which we named the GreenCut (Fig. 5A, table S10 and table SA); each of these families has a single *Chlamydomonas* protein. On the basis of manual curation of GreenCut proteins of known function (1) (table S11), we estimated ~5 to 8% false-positives and ~14% false-negatives (1). By comparing GreenCut proteins to those of the red alga *Cyanidioschyzon merolae*, which diverged before the split of green algae from land plants (Fig. 2), we identified the subset of proteins present across the plant kingdom; we named this subset the PlantCut (Fig. 5A,table S10 and table SA). GreenCut protein families that also included representatives from the diatoms *Thalassiosira pseudonana* (30) or *Phaeodactylum tricornutum* (31) were placed in the DiatomCut (Fig. 5A and table S10 and table SA). Given the phylogenetic position of diatoms and their secondary endosymbiosis-derived plastids, we hypothesize that protein families present in both the PlantCut and DiatomCut should contain only those GreenCut proteins associated with plastid function. This subset is referred to as the PlastidCut (Fig. 5A).

The GreenCut contains proteins of the photosynthetic apparatus, including those involved in plastid and thylakoid membrane biogenesis, photosynthetic electron transport, carbon fixation, antioxidant generation, and a range of other primary metabolic processes (table S11 and table SA). Although light-harvesting chlorophyll-binding proteins are poorly represented (1), we identified specialized chlorophyll-binding proteins, as well as a photosynthesis-specific kinase, involved in state transitions. Numerous GreenCut entries are enzymes of plastid-localized metabolic pathways (lipid, amino acid, starch, nucleotide, and pigment biosynthesis) or are unique to plants or highly divergent from animal counterparts. Although tRNA synthetases are conserved between kingdoms, those in the GreenCut represent organellar isoforms that are often targeted to both plastids and mitochondria in plants (32). GreenCut proteins that do not function in the plastids tend to be green lineage–specific or highly diverged from animal counterparts. For example, the *Chlamydomonas* GreenCut protein TOM20 (1), an outer mitochondrial membrane receptor involved in protein import, evolved convergently from a different ancestral protein in plants than in fungi and animals (33).

Of the 214 proteins in the GreenCut without known function, 101 have no motifs or homologies from which function can be inferred, and we can predict only a general function for the others (table S12). Given that 85% of the known proteins in the GreenCut are localized to chloroplasts (table S13), we predict that the set of unknowns contains many novel, conserved proteins that function in chloroplast metabolism and regulation.

The most reducing and oxidizing biological molecules are generated in chloroplasts via the activity of photosystem I and photosystem II, respectively. The flow of electrons through the photosystems causes damage to cellular constituents as a consequence of the accumulation of reactive oxygen species. Therefore, regulation of these molecules is important. Accordingly,

plastids house more redox regulators than do mitochondria. Thioredoxins are critical redox-state regulators, and we identified novel thioredoxins in the GreenCut (table S12). These novel thioredoxins have noncanonical active sites or are fused to domains of inferred function (e.g., a vitamin K–binding domain) in plastid metabolism (fig. S1). These findings reveal the potential for identifying unique redox signaling pathways with selectivity and midpoint potentials associated with specific thioredoxin redox sensors (1).

Chlamydomonas has a structure called the eyespot (Fig. 1) which can sense light and trigger phototactic responses. The eyespot is composed of several layers of pigment granules, similar to plastoglobules in plants, and thylakoid membrane, which are directly apposed to the chloroplast envelope and a region of the plasma membrane carrying rhodopsin-family photoreceptors. The pigment granules or plastoglobules contain many proteins with unknown function, many of which are present in the GreenCut, and are likely critical to plastid metabolism; these include SOUL domain, AKC (see below), and PLAP (plastid- and lipid-associated protein) protein families (34–36). SOUL domain proteins of the GreenCut (SOUL4 and SOUL5) have homologs in the *Arabidopsis* plastoglobule proteome (34,35), and at least one (SOUL3) is associated with the eyespot. The SOUL domain, originally identified in proteins encoded by highly expressed genes in the retina and pineal gland, can bind heme (37,38). This domain may be important as a heme carrier and/or in maintaining heme in a bound, non-phototoxic form until it associates with proteins or may function in signaling circadian cues.

We also identified plant-specific AKCs (ABC1 kinase in the chloroplast, AKC1 to 4 in the GreenCut), one of which (designated EYE3) is required for eyespot assembly (39). These AKCs are distinct from the mitochondrial ABC1 kinase that regulates ubiquinone production (40). Protein phosphatases present in the GreenCut and plastoglobules may turn off signaling initiated by the AKCs.

The PLAPs (PLAP1 to 4 in the GreenCut), also called plastoglobulins, are also associated with the eyespot or plastoglobule. These proteins were originally identified by their abundance in carotenoid-rich fibrils and chromoplast plastoglobules and may be structural or organizational components of this plastid subcompartment. Other GreenCut proteins associated with plastoglobules (34,36) include short-chain dehydrogenases, an aldo-keto isomerase, various methyltransferases with unspecified substrates, esterases and lipases, and a protein with a pantothenate kinase motif.

In sum, the eyespot or plastoglobules contain proteins that likely function in the synthesis, degradation, trafficking, and integration of pigments and lipophilic cofactors into the metabolic machinery of the cell and, most notably, into the photosynthetic apparatus, where they are in high demand. The numerous proteins in the GreenCut associated with the eyespot/plastoglobules may reflect the diverse repertoire of compounds, such as quinones, tocopherols, carotenoids, and tetrapyrroles (fig. S1B), required by photosynthetic organisms.

The 90 proteins in the PlastidCut (Fig. 5A) are likely to function in basic plastid processes because they are conserved in all plastid-containing eukaryotes. Sixty-one of these have unknown functions, with genes for most (except CPLD6 and CPLD29) expressed in chloroplast-containing cells, as assessed from EST representation in *Chlamydomonas* and *Physcomitrella*. For *Arabidopsis* homologs, expression (41) indicates that the genes represented in the PlastidCut tend to be expressed in leaves or all tissue, similar to genes that function in photosynthesis or primary chloroplast metabolism. Greater than 70% of previously unknown PlastidCut proteins have homologs in cyanobacteria, which suggests a critical, conserved, plastid-associated function.

Flagellar and basal body gene complement

Chlamydomonas uses a pair of anterior flagella to swim and sense environmental conditions (Fig. 1). Each flagellum is rooted in a basal body, which also functions as a centriole during cell division. The flagellar axoneme has the nine outer doublet microtubules plus a central pair (9+2) (Fig. 1) characteristic of motile cilia (cilia and eukaryotic flagella are essentially identical organelles). In addition to motile cilia, animals contain nonmotile cilia that function as a sensory organelle and typically lack outer and inner dynein arms, radial spokes, and central microtubules (Fig. 1), all of which are involved in the generation and regulation of motility. Both types of cilia have sensory functions and share conserved sensing and signaling components.

The loss of flagella in angiosperms, most fungi, and slime molds allowed us to identify cilia-specific genes through searches for proteins retained only in flagellate organisms (8,26). We searched the 6968 *Chlamydomonas* protein families (see above) for those that also contained proteins from human and a *Phytophthora* spp., but not from aciliates, and identified 186 protein families that we named the CiliaCut; these families contain 195 *Chlamydomonas* (Fig. 5B and table SB) and 194 human proteins. One hundred and sixteen of the *Chlamydomonas* proteins had been computationally identified (8,26), and 45 were identified in this study (1).

The *Chlamydomonas* CiliaCut proteins of unknown function that are missing from *Caenorhabditis*, which has only nonmotile sensory cilia (26), were designated MOT (motile flagella), whereas proteins of unknown function shared with *Caenorhabditis* were designated SSA (sensory, structural and assembly) (Fig. 5B). Thirty-five percent of CiliaCut proteins are in the *Chlamydomonas* flagellar proteome (42), double the number known from previous studies, and 27 of 101 previously identified flagellar proteins (42) are present in the CiliaCut. The CiliaCut contained δ-tubulin, which is required for basal body assembly (43), and a previously undescribed dynein light chain. Some flagellar proteins were not found by this analysis because they have orthologs in plants and fungi, whereas others are absent because they lack human orthologs. Most dynein heavy chains are missing, most likely due to the difficulty of identifying members of large gene families with a mutual best hit approach (1).

We manually curated 125 CiliaCut proteins (fig. S24) and identified large subsets as flagellar structural components (16%), mediating protein-protein interactions (26%), signaling (11%), GTP-binding (6%) and trafficking (6%). These results are consistent with proteomic analysis of the flagellum (42) and highlight the importance of signaling even in motile flagella.

The 62 CiliaCut proteins that *Chlamydomonas* shares with *Caenorhabditis* are predicted to have structural, sensory, or assembly roles in the cilium. As expected, the 133 CiliaCut proteins missing from *Caenorhabditis* (Fig. 5B) (1), designated the MotileCut, include a number of proteins associated with motility (42) (table S14). This data set also contains 31 proteins of unknown function found in the flagellar and basal body proteomes, 36 known but uncharacterized proteins, and 55 novel proteins (designated MOT1 to MOT55); these flagellar proteins are all predicted to be involved specifically in motility.

A comparison of CiliaCut proteins with proteins encoded by the *Physcomitrella* genome indicates that *Physcomitrella* has lost five of the outer dynein arm proteins (Fig. 1, table S14). However, *Physcomitrella* contains inner dynein arm subunits IDA4 and DHC2, as well as subunits of the central microtubules, the radial spokes, and the dynein regulatory complex (table S14). From this we conclude that *Physcomitrella* sperm flagella have a “9+2” axoneme containing inner dynein arms, central microtubules, and radial spokes, but lack the outer dynein arms. Although the structure of the *Physcomitrella* sperm flagellum is not known, sperm flagella of the bryalean moss *Aulacomnium palustre* have just such an axoneme (44).

In contrast, the motile flagella of centric diatoms lack the central pair of microtubules (45, 46). Orthologs of 69 of the 195 CiliaCut proteins (named CentricCut, Fig. 5B) were predicted to be present in the centric diatom *Thalassiosira*. As expected, *Thalassiosira* lacks all central pair proteins. However, it also lacks all radial spoke and inner dynein arm proteins, but has most of the outer dynein arm proteins. The contrasting patterns of loss of axonemal structures predicted for *Physcomitrella* and *Thalassiosira* suggest that the central pair and radial spokes function as a unit with the inner arms, but are dispensable for the generation of motility by the outer arms.

Intraflagellar transport (IFT), which is conserved in ciliated organisms except malaria parasites (47), is essential for flagellar growth (48). The IFT machinery consists of at least 16 proteins in two complexes (A and B) that are moved in anterograde and retrograde directions by the molecular motors kinesin-2 and cytoplasmic dynein 1b, respectively (Fig. 1). Our analysis of *Thalassiosira* reveals that it has components of the anterograde motor and complex B, but has lost the retrograde motor and complex A (table S14). This is intriguing, as retrograde IFT is essential for flagellar maintenance in *Chlamydomonas* (49) and is important for recycling IFT components (50). In addition, both *Physcomitrella* and *Thalassiosira* have lost the Bardet-Biedl syndrome (BBS) genes. BBS gene products are associated with the basal body in *Chlamydomonas* and mammals (8,51) and sensory cilia in *Caenorhabditis* (52), where they may be involved in IFT (53).

We searched the CiliaCut proteins for proteins shared with *Ostreococcus* spp., a green alga lacking a flagellate stage. The *Ostreococcus* spp. retain 46 (24%) of the 195 CiliaCut proteins but, consistent with loss of the flagellum, are missing genes encoding the IFT-particle proteins and motors, the inner and outer dynein arm proteins, the radial spoke and central pair proteins, and 32 out of 39 flagella-associated proteins (FAPs) (table S14). They have also lost many genes encoding basal body proteins, including all BBS proteins (table S14), which suggests that *Ostreococcus* also lack basal bodies. However, *Ostreococcus* spp. have retained many other CiliaCut proteins (table S14), which suggests either that they recently lost their flagella, or that they retained flagellar proteins for other cellular functions.

Conclusions

This analysis of the *Chlamydomonas* genome sheds light on the nature of the last common ancestor of plants and animals and identifies many cilia- and plastid-related genes. The gene complement also provides insights into life in the soil environment where extreme competition for nutrients likely drove expansion of transporter gene families, as well as sensory flagellar and eyespot functions (e.g., facilitating nutrient acquisition and optimization of the light environment). As more of the ecology and physiology of *Chlamydomonas* and other unicellular algae are explored, additional direct links between gene content and functions associated with the soil life-style will be unmasked with increased potential for biotechnological exploitation of these functions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank R. Howson for help with drawing figures, E. Begovic and S. Nicholls for comments on the manuscript. SM is supported by the grants NIH GM42143, DOE DE-FG02-04ER15529 USDA 2004-35318-1495. SP and DSR are funded by USDA and DOE, Joint Genome Institute. ARG is supported by USDA 2003-35100-13235, DOE DE-AC36-99GO10337 and the NSF-funded *Chlamydomonas* Genome Project, MCB 0235878. SJK was supported in part by a Ruth L. Kirschstein National Research Service Award GM07185. The authors declare they have no conflicts of

interest. Genome assembly together with predicted gene models and and annotations were deposited at DDBJ/EMBL/GenBank under the project accession ABCN00000000. Since manual curation continues, some models or anotations are changing and the latest set of gene models and annotations is available from www.jgi.doe.gov/chlamy. The most recent set, which includes a number of changes compared with the frozen set used for this analysis, was submitted as the first version, ABCN01000000.

References and Notes

1. Materials and methods and supplemental online (SOM) text are available as supporting material on *Science* Online.
2. Harris EH. Annu Rev Plant Physiol Plant Mol Biol 2001;52:363. [PubMed: 11337403]
3. Keller LC, Romijn EP, Zamora I, Yates JR 3rd, Marshall WF. Curr Biol 2005;15:1090. [PubMed: 15964273]
4. Pazour GJ, Agrin N, Walker BL, Witman GB. J Med Genet 2006;43:62. [PubMed: 15937072]
5. Vilchez C, Garbayo I, Markvicheva E, Galvan F, Leon R. Bioresour Technol 2001;78:55. [PubMed: 11265789]
6. Ghirardi ML, et al. Annu Rev Plant Biol 2007;58:71. [PubMed: 17150028]
7. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. Mol Biol Evol 2004;21:809. [PubMed: 14963099]
8. Li JB, et al. Cell 2004;117:541. [PubMed: 15137946]
9. Grossman AR, et al. Curr Opin Plant Biol 2007;10:190. [PubMed: 17291820]
10. Chlamydomonas reinhardtii v 3.0. DOE Joint Genome Institute; www.jgi.doe.gov/chlamy
11. Storms R, Hastings PJ. Exp Cell Res 1977;104:39. [PubMed: 836405]
12. Kathir P, et al. Eukaryot Cell 2003;2:362. [PubMed: 12684385]
13. Rymarquis LA, Handley JM, Thomas M, Stern DB. Plant Physiol 2005;137:557. [PubMed: 15665247]
14. Jain M, et al. Nucleic Acids Res 2007;35:2074. [PubMed: 17355987]
15. Yuan Q, et al. Plant Physiol 2005;138:18. [PubMed: 15888674]
16. Yandell M, et al. PLoS Comput Biol 2006;2:e15. [PubMed: 16518452]
17. Palenik B, et al. Proc Natl Acad Sci USA 2007;104:7705. [PubMed: 17460045]
18. Schaap P. Front Biosci 2005;10:1485. [PubMed: 15769639]
19. Hasegawa E, Hayashi H, Asakura S, Kamiya R. Cell Motil Cytoskeleton 1987;8:302. [PubMed: 2826019]
20. Pasquale SM, Goodenough UW. J Cell Biol 1987;105:2279. [PubMed: 2824527]
21. Gaillard AR, Fox LA, Rhea JM, Craige B, Sale WS. Mol Biol Cell 2006;17:2626. [PubMed: 16571668]
22. Gonzalez-Ballester D, de Montaigu A, Higuera JJ, Galvan A, Fernandez E. Plant Physiol 2005;137:522. [PubMed: 15665251]
23. Pollock SV, Pootakham W, Shibagaki N, Moseley JL, Grossman AR. Photosynth Res 2005;86:475. [PubMed: 16307308]
24. Grossman A, Takahashi H. Annu Rev Plant Physiol Plant Mol Biol 2001;52:163. [PubMed: 11337396]
25. Somlo S, Ehrlich B. Curr Biol 2001;11(9):R356. [PubMed: 11369247]
26. Avidor-Reiss T, et al. Cell 2004;117:527. [PubMed: 15137945]
27. Gray MW. Curr Opin Genet Dev 1999;9:678. [PubMed: 10607615]
28. Bhattacharya D, Yoon HS, Hackett JD. Bioessays 2004;26:50. [PubMed: 14696040]
29. Keeling P. Protist 2004;155:3. [PubMed: 15144050]
30. Armbrust EV, et al. Science 2004;306:79. [PubMed: 15459382]
31. Phaeodactylum tricornutum, v2.0. DOE Joint Genome Institute; www.jgi.doe.gov/phaeodactylum
32. Duchêne AM, et al. Proc Natl Acad Sci USA 2005;102:16484. [PubMed: 16251277]
33. Perry AJ, Hulett JM, Likic VA, Lithgow T, Gooley PR. Curr Biol 2006;16:221. [PubMed: 16461275]
34. Ytterberg AJ, Peltier JB, van Wijk KJ. Plant Physiol 2006;140:984. [PubMed: 16461379]

35. Schmidt M, et al. Plant Cell 2006;18:1908. [PubMed: 16798888]
36. Vidi PA, et al. J Biol Chem 2006;281:11225. [PubMed: 16414959]
37. Zylka MJ, Reppert SM. Brain Res Mol Brain Res 1999;74:175. [PubMed: 10640688]
38. Sato E, et al. Biochemistry 2004;43:14189. [PubMed: 15518569]
39. Lamb MR, Dutcher SK, Worley CK, Dieckmann CL. Genetics 1999;153:721. [PubMed: 10511552]
40. Do TQ, Hsu AY, Jonassen T, Lee PT, Clarke CF. J Biol Chem 2001;276:18161. [PubMed: 11279158]
41. Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W. Plant Physiol 2004;136:2621. [PubMed: 15375207]
42. Pazour GJ, Agrin N, Leszyk J, Witman GB. J Cell Biol 2005;170:103. [PubMed: 15998802]
43. O'Toole ET, Giddings TH, McIntosh JR, Dutcher SK. Mol Biol Cell 2003;14:2999. [PubMed: 12857881]
44. Bernhard DL, Renzaglia KS. Bryologist 1995;98:52.
45. Manton I, Kowallik K, von Stosch HA. J Cell Sci 1970;6:131. [PubMed: 5417690]
46. Heath IB, Darley WM. J Phycol 1972;18:51.
47. Briggs LJ, Davidge JA, Wickstead B, Ginger ML, Gull K. Curr Biol 2004;14:R611. [PubMed: 15296774]
48. Rosenbaum JL, Witman GB. Nat Rev Mol Cell Biol 2002;3:813. [PubMed: 12415299]
49. Pazour GJ, Dickert BL, Witman GB. J Cell Biol 1999;144:473. [PubMed: 9971742]
50. Qin H, Diener DR, Geimer S, Cole DG, Rosenbaum JL. J Cell Biol 2004;164:255. [PubMed: 14718520]
51. Ansley SJ, et al. Nature 2003;425:628. [PubMed: 14520415]
52. Blacque OE, et al. Genes Dev 2004;18:1630. [PubMed: 15231740]
53. Ou G, et al. Mol Biol Cell 2007;18:1554. [PubMed: 17314406]
54. Ciccarelli FD, et al. Science 2006;311:1283. [PubMed: 16513982]
55. Keeling PJ, et al. Trends Ecol Evol 2005;20:670. [PubMed: 16701456]
56. Eichinger L, et al. Nature 2005;435:43. [PubMed: 15875012]

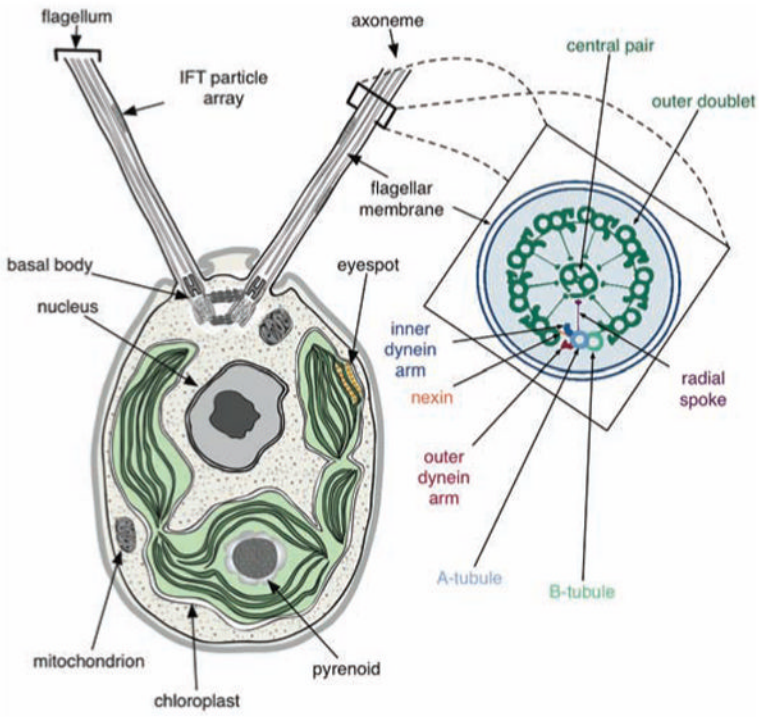


Fig. 1.

A schematic of a *Chlamydomonas* cell (from transmission electron micrographs) showing the anterior flagella rooted in basal bodies, with intraflagellar transport (IFT) particle arrays between the axoneme and flagellar membrane, the basal cup-shaped chloroplast, central nucleus and other organelles. An expanded cross section of the flagellar axoneme, as redrawn from (48), shows the nine outer doublets and the central pair (9+2) microtubules; axoneme substructures are color-coded and labeled (see inset).

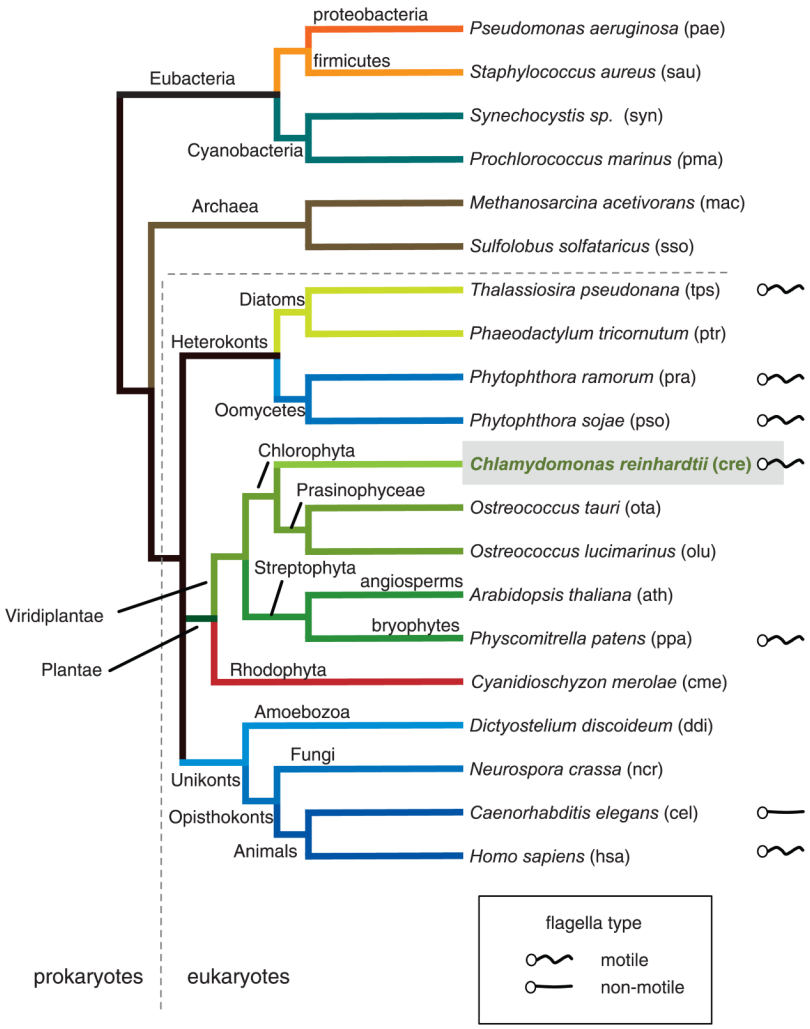


Fig. 2.

Evolutionary relationships of 20 species with sequenced genomes (54,55) used for the comparative analyses in this study include cyanobacteria and nonphotosynthetic eubacteria, Archaea and eukaryotes from the oomycetes, diatoms, rhodophytes, plants, amoebae and opisthokonts. Endosymbiosis of a cyanobacterium by a eukaryotic protist gave rise to the green (green branches) and red (red branches) plant lineages, respectively. The presence of motile or nonmotile flagella is indicated at the right of the cladogram.

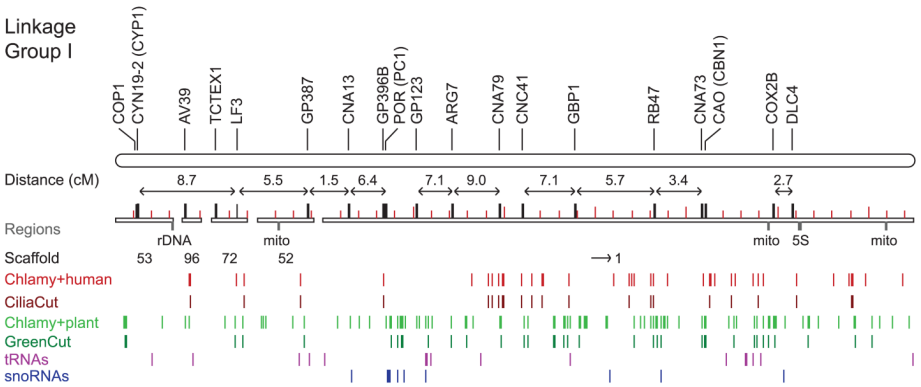


Fig. 3.

Linkage group I depicted as a long horizontal rod, with genetically mapped scaffolds shown as open rectangles below (the scaffold number is under each scaffold, and arrows indicate the orientation of the scaffold where it is known; other scaffolds were placed in their most likely orientation on the basis of genetic map distances. The scale of each map is determined by molecular lengths of the mapped scaffolds. Short and long red ticks are drawn on scaffolds every 0.2 Mb and 1.0 Mb, respectively. We assumed small 50 kb gaps between scaffolds. Genetic distances between markers (centimorgans), where they are known, are shown by two-headed arrows above the scaffold, with the gene symbol and any synonyms in parentheses shown at the top. Genomic regions are labeled below the scaffolds: 5*S*, rDNA, mito (insertion of mitochondrial DNA). *Chlamydomonas* genes with homologs in other organisms/lineages ("Cuts" as defined in the text and Fig. 5) are shown as tracks of vertical bars: light red, genes shared between *Chlamydomonas* and humans, but not occurring in nonciliated organisms; dark red, genes in CiliaCut; light green, genes shared between *Chlamydomonas* and *Arabidopsis*, but not in nonphotosynthetic organisms; dark green, genes in GreenCut; magenta, predicted tRNAs, including those that represent SINE sequences; dark blue, small nucleolar RNAs (snoRNAs).

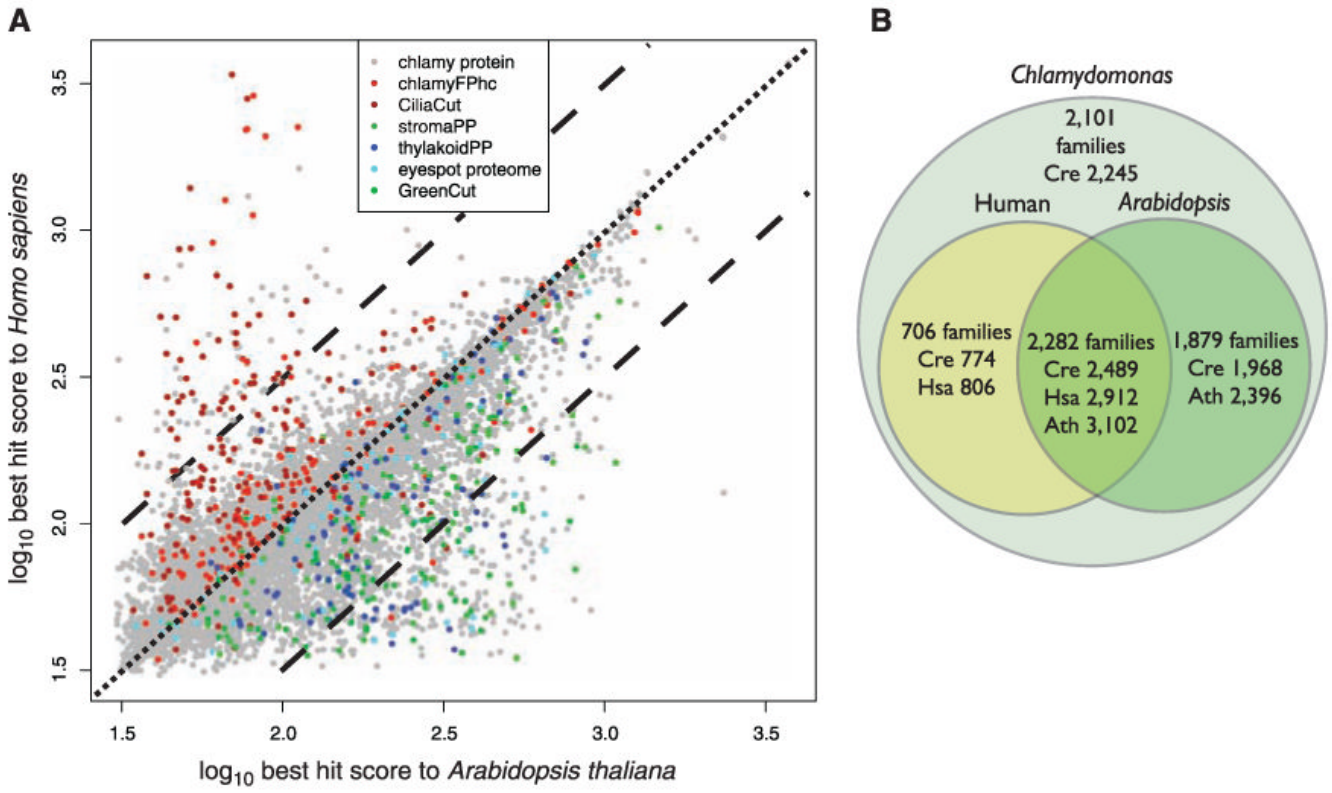
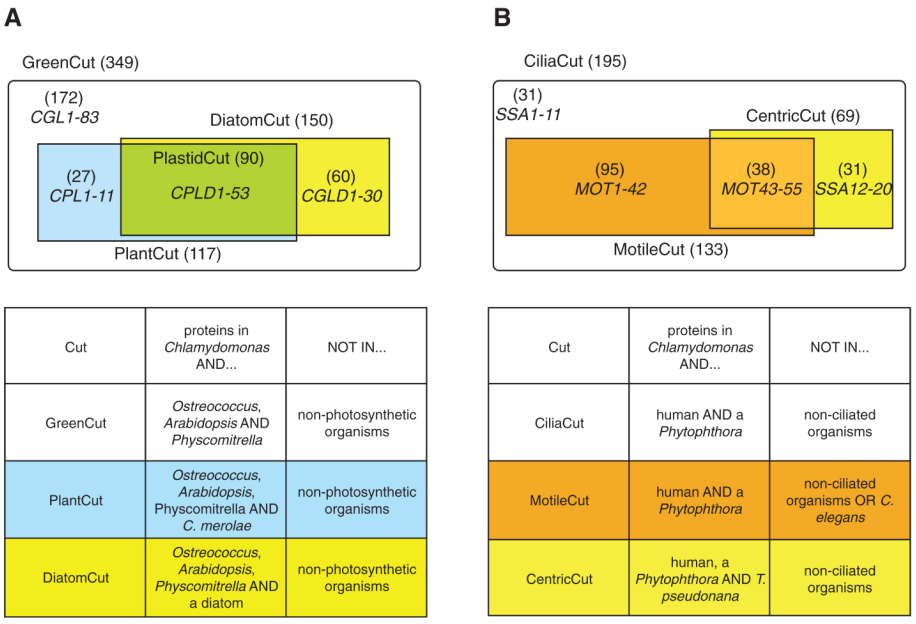


Fig. 4.

(**A**) Scatter plot of best BLASTP hit score of *Chlamydomonas* proteins to *Arabidopsis* proteins versus best BLASTP hit score of *Chlamydomonas* proteins to human proteins. Functional or genomic groupings are colored [see inset key in (A)]: *Chlamydomonas* flagellar proteome (42) high confidence set (chlamyFPhc); CiliaCut; *Arabidopsis* stroma plastid proteome (stromaPP); *Arabidopsis* thylakoid plastid proteome (thylakoidPP); eyespot proteome; GreenCut; remaining proteins are gray. (**B**) *Chlamydomonas* protein paralogs were grouped into families together with their homologs from human and *Arabidopsis*. The outer circle represents the proteins in *Chlamydomonas*, 7476 (out of 15,143 total), that fall into 6968 families. Another 7937 proteins cannot be placed in families. Counts of families (and the numbers of proteins from each species in them) with proteins from *Chlamydomonas* and human only, *Chlamydomonas* and *Arabidopsis* only, and *Chlamydomonas* and human and *Arabidopsis*, are shown in the inner circles and the overlap between the two inner circles, respectively. Cre, *Chlamydomonas*; Hsa, human; Ath, *Arabidopsis*.



Cut	proteins in *Chlamydomonas* AND...	NOT IN...
GreenCut	*Ostreococcus*, *Arabidopsis* AND *Physcomitrella*	non-photosynthetic organisms
PlantCut	*Ostreococcus*, *Arabidopsis*, Physcomitrella AND *C. merolae*	non-photosynthetic organisms
DiatomCut	*Ostreococcus*, *Arabidopsis*, *Physcomitrella* AND a diatom	non-photosynthetic organisms

Cut	proteins in *Chlamydomonas* AND...	NOT IN...
CiliaCut	human AND a *Phytophthora*	non-ciliated organisms
MotileCut	human AND a *Phytophthora*	non-ciliated organisms OR *C. elegans*
CentricCut	human, a *Phytophthora* AND *T. pseudonana*	non-ciliated organisms

Fig. 5.

Summary of genomic comparisons to photosynthetic and ciliated organisms. (**A**) GreenCut: The GreenCut comprises 349 *Chlamydomonas* proteins with homologs in representatives of the green lineage of the Plantae (*Chlamydomonas*, *Physcomitrella*, and *Ostreococcus tauri* and *O. lucimarinus*), but not in nonphotosynthetic organisms. Genes encoding proteins of unknown function that were not previously annotated were given names on the basis of their occurrence in various cuts. CGL refers to conserved only in the green lineage. The GreenCut protein families, which also include members from the red alga *Cyanidioschyzon* within the Plantae, were assigned to the PlantCut (blue plus green rectangles). CPL refers to conserved in the Plantae. GreenCut proteins also present in at least one diatom (*Thalassiosira* and *Phaeodactylum*) were assigned to the DiatomCut (yellow plus green rectangle). CGLD refers to conserved in the green lineage and diatoms. Proteins present in all of the eukaryotic plastid-containing organisms in this analysis were assigned to the PlastidCut (green rectangle). CPLD refers to conserved in the Plantae and diatoms. The criteria used for the groupings associated with the GreenCut are given in the lower table. (**B**) CiliaCut: The CiliaCut contains 195 *Chlamydomonas* proteins with homologs in human and species of *Phytophthora*, but not in nonciliated organisms. This group was subdivided on the basis of whether or not a homolog was present in *Caenorhabditis*, which has only nonmotile sensory cilia. The 133 CiliaCut proteins without homologs in *Caenorhabditis* were designated the MotileCut (orange rectangle). Unnamed proteins in this group were named MOT (motility). Proteins with homologs in *Caenorhabditis* are associated with nonmotile cilia (white and yellow areas). Proteins in this group that were not already named were named SSA. The CentricCut (yellow plus light orange box) is made up of 69 CiliaCut homologs present in the centric diatom *Thalassiosira*. These proteins can be divided into those also in the MotileCut (38 proteins; light orange box) or those not present in the MotileCut (31 proteins; yellow box).

Table 1

Comparison of *Chlamydomonas* genome statistics to those of selected sequenced genomes. nd, Not determined. [Source for all but *Chlamydomonas* (1)]

	Chlamydomonas	***Ostreococcus tauri***	***Cyanidioschyzon***	***Arabidopsis***	**Human**
Assembly length (Mb)	121	12.6	16.5	140.1	2,851
Coverage	13×	6.7×	11×	nd	~8×
Chromosomes	17	20	20	5	23
G+C (%)	64	58	55	36	41
G+C (%) coding sequence	68	59	57	44	52
Gene number	15,143	8,166	5,331	26,341	~23,000
Genes with EST support (%)	63	36	86	60	nd
Gene density (per kb)	0.125	0.648	0.323	0.190	~0.0008
Average bp per gene	4312	nd	1553	2232	27,000
Average bp per transcript	1580	1257	1552	nd	nd
Average number of amino acids per polypeptide	444	387	518	413	491
Average number of exons per gene	8.33	1.57	1.005	5.2	8.8
Average exon length	190	750	1540	251	282*
Genes with introns (%)	92	39	0.5	79	85†
Mean length of intron	373	103	248	164	3,365
Coding sequence (%)	16.7	81.6	44.9	33.0	~1
Number of rDNA units (28*S*/18*S*/5.8*S* + 5*S*)	3 + 3	4 + 4	3 + 3‡	12 + 700	5 + nd
Number tRNAs	259§	nd	30	589	497
Selenocysteine (Sec) tRNAs	1	nd	nd	0	1

*National Center for Biotechnology Information (NIH) NCBI 36 from Ensembl build 38.

†[Source (56)].

‡Three regions contain 5*S* rDNA exclusively, and three regions contain 28*S*-18*S*-5.8*S* rDNAs exclusively.

§65 tRNAs that were included in SINE elements were removed from the tRNA-scanSE predictions.