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,5 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 Chlamydomonas Annotation Team,† JGI Annotation Team,† Igor V. Grigoriev,4 Daniel S. Rokhsar,6,16 Arthur R. Grossman17†

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 (cilium), 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) (2) over a billion years ago. These lineages are part of the green plant lineage (Viridiplantae), which previously diverged from opisthokonts (animals, fungi, and Chordozoa) (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 centrosome) (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, result-

---

*These authors contributed equally to this work.
†Full author list is included at the end of the manuscript.
‡To whom correspondence should be addressed. E-mail: drolksar@lbl.gov (D.S.R.); arthurg@stanford.edu (A.R.G.)

---

**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).
ing 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) (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 (I) (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) (I). 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) (I, 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 (I), 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 (I); 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) (I) 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 Ca2+ in response to environmental signals such as light. Changing Ca2+ 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+/SO42- family (characteristic of the plant lineage), three are in the Na+/SO42- family (not found in plants but present in opisthokonts), and one is a bacterial ABC-type SO42- 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 endo-

---

### Table 1. Comparison of *Chlamydomonas* genome statistics to those of selected sequenced genomes. nd, Not determined. [Source for all but *Chlamydomonas* (11)]

<table>
<thead>
<tr>
<th></th>
<th><em>Chlamydomonas</em></th>
<th><em>Ostreococcus tauri</em></th>
<th><em>Cyanidioschyzon</em></th>
<th><em>Arabidopsis</em></th>
<th><em>Human</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly length (Mb)</td>
<td>121</td>
<td>12.6</td>
<td>16.5</td>
<td>140.1</td>
<td>2,851</td>
</tr>
<tr>
<td>Coverage</td>
<td>13×</td>
<td>6.7×</td>
<td>11×</td>
<td>nd</td>
<td>~8×</td>
</tr>
<tr>
<td>Chromosomes</td>
<td>17</td>
<td>20</td>
<td>20</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>G+C (%)</td>
<td>64</td>
<td>58</td>
<td>55</td>
<td>36</td>
<td>41</td>
</tr>
<tr>
<td>G+C (%) coding sequence</td>
<td>68</td>
<td>59</td>
<td>57</td>
<td>44</td>
<td>52</td>
</tr>
<tr>
<td>Gene number</td>
<td>15,143</td>
<td>8,166</td>
<td>5,331</td>
<td>26,341</td>
<td>~23,000</td>
</tr>
<tr>
<td>Genes with EST support</td>
<td>63</td>
<td>36</td>
<td>86</td>
<td>60</td>
<td>nd</td>
</tr>
<tr>
<td>Gene density (per kb)</td>
<td>0.125</td>
<td>0.648</td>
<td>0.323</td>
<td>0.190</td>
<td>~0.0008</td>
</tr>
<tr>
<td>Average bp per gene</td>
<td>4312</td>
<td>nd</td>
<td>1553</td>
<td>2232</td>
<td>27,000</td>
</tr>
<tr>
<td>Average bp per transcript</td>
<td>1580</td>
<td>1257</td>
<td>1552</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>Average number of amino acids per polypeptide</td>
<td>444</td>
<td>387</td>
<td>518</td>
<td>413</td>
<td>491</td>
</tr>
<tr>
<td>Average number of exons per gene</td>
<td>8.33</td>
<td>1.57</td>
<td>1.005</td>
<td>5.2</td>
<td>8.8</td>
</tr>
<tr>
<td>Average exon length</td>
<td>190</td>
<td>750</td>
<td>1540</td>
<td>251</td>
<td>282*</td>
</tr>
<tr>
<td>Genes with introns (%)</td>
<td>92</td>
<td>39</td>
<td>0.5</td>
<td>79</td>
<td>85†</td>
</tr>
<tr>
<td>Mean length of intron</td>
<td>373</td>
<td>103</td>
<td>248</td>
<td>164</td>
<td>3,365</td>
</tr>
<tr>
<td>Coding sequence (%)</td>
<td>16.7</td>
<td>81.6</td>
<td>44.9</td>
<td>33.0</td>
<td>~1</td>
</tr>
<tr>
<td>Number of rDNA units (28S/18S/5.8S + 5S)</td>
<td>3 + 3</td>
<td>4 + 4</td>
<td>3 + 3†</td>
<td>12 + 700</td>
<td>5 + nd</td>
</tr>
<tr>
<td>Number tRNAs</td>
<td>2595</td>
<td>nd</td>
<td>589</td>
<td>497</td>
<td></td>
</tr>
<tr>
<td>Selenocysteine (Sec) tRNAs</td>
<td>1</td>
<td>nd</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

* National Center for Biotechnology Information (NIH) NCBI 36 from Ensembl build 38.
† [Source (56)]. § Three regions contain 55 rDNA exclusively, and three regions contain 28S/18S/5.8S rDNAs exclusively. ‡ 65 tRNAs that were included in SINE elements were removed from the tRNA-scanSE predictions.
symbiosis (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, antioxidative 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 enzymes 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 homologs 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 (SOUL5) 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-

---

**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*.
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-sensory cilia or plastoglobules and finally 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 eye spot 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 (cilial and eukaryotic flagellates are essentially identical organelles). In addition to motile cilia, animals contain nonmotile cilial flagella 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 acelulates, 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

---

**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 *Cyanidioschoen**x** 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).
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 proteins, 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 those encoded by the *Physcomitrella* genome indicates that *Physcomitrella* has lost five of the outer dynein arm proteins (Fig. 1, table S14). However, *Caenorhabditis* 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 bryophyte moss *Uglandium* has 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 (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 evolution of transportor gene families, as well as sensory flagellar and eyspot 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.

References and Notes
1. Materials and methods and supplemental online (SOM) text are available as supporting material on Science Online.
Dislocation Avalanches, Strain Bursts, and the Problem of Plastic Forming at the Micrometer Scale

Ferenc F. Csikor,1,2 Christian Motz,3 Daniel Wegyard,3 Michael Zaiser,5 Stefano Zapperi5,*

Under stress, many crystalline materials exhibit irreversible plastic deformation caused by the motion of lattice dislocations. In plastically deformed microcrystals, internal dislocation avalanches lead to jumps in the stress-strain curves (strain bursts), whereas in macroscopic samples plasticity appears as a smooth process. By combining three-dimensional simulations of the dynamics of interacting dislocations with statistical analysis of the corresponding deformation behavior, we determined the distribution of strain changes during dislocation avalanches and established its dependence on microcrystal size. Our results suggest that for sample dimensions on the micrometer and submicrometer scale, large strain fluctuations may make it difficult to control the resulting shape in a plastic-forming process.

In recent years, experimental evidence has accumulated that indicates that plastic flow is—at least on the micrometer scale—characterized by intermittent strain bursts with scale-free (i.e., power-law) size distributions (1–8). The phenomenology of these strain bursts closely resembles that of macroscopic plastic instabilities: Stress-strain curves are characterized by serrated yielding under displacement control and assume a staircase shape under conditions of stress control. Temporal intermittency is associated with spatial localization because each strain burst corresponds to the formation of a narrow slip line or slip band (9). On the macroscopic scale, spatiotemporal localization of plastic deformation associated with plastic instabilities is well known to have a detrimental effect on formability. A classic example is the strain bursts discovered by Portevin and le Chatelier (PLC effect), which arise from the interaction between dislocations and diffusing solutes (10). The PLC effect limits the applicability of many aluminum alloys in sheet metal–forming processes, but only arises under specific deformation conditions. Thus, the instability can be circumvented by appropriately choosing the process path, avoiding those temperature and strain rate conditions.