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 Salamony,2 Lillian K. Fritz-Laylin,6 Laurence Maréchal-Drouard,7 Wallace F. Marshall,8 Liang-Hu Ou,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 Susan M. Lucas,2 Jane Grimwood,16 Anton A. Sanderfoot,11 Martin H. Spalding,12 Vladimir V. Kapitonov,13 Qinghu Ren,14 Jeremy Schmutz,16 Chlamydomonas Annotation Team,† JGI Annotation Team,† Igor V. Grigoriev,4 Daniel S. Rokhsar,2,6 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 (cilias), 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 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...

---

1Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA 90095, USA. 2U.S. Department of Energy (DOE) Joint Genome Institute (JGI), Walnut Creek, CA 94598, USA. 3CNRS, Université Paris 6, Institut de Biologie Physico-Chimique, 75005 Paris, France. 4Department of Biology, Duke University, Durham, NC 27708, USA. 5Department of Cell Biology, University of Massachusetts Medical School, Worcester, MA 01655, USA. 6Department of Moleculer and Cell Biology, University of California at Berkeley, Berkeley, CA 94720, USA. 7Institut de Biologie Moléculaire des Plantes, CNRS, 67084 Strasbourg Cedex, France. 8Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143, USA. 9Biotechnology Research Center, Zhongshan University, Guangzhou 510275, China. 10Department of Molecular Sciences and Center of Excellence in Genomics and Bioinformatics, University of Tennessee, Memphis, TN 38163, USA. 11Department of Plant Biology, University of Minnesota, St. Paul, MN 55108, USA. 12Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011, USA. 13Genetic Information Research Institute, Mountain View, CA 94043, USA. 14The Institute for Genomic Research, Rockville, MD 20850, USA. 15Plant Biology Laboratory, Salk Institute, La Jolla, CA 92037, USA. 16Stanford Human Genome Center, Stanford University School of Medicine, Palo Alto, CA 94304, USA. 17Department of Plant Biology, Carnegie Institution, Stanford, CA 94306, USA.

*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: drosrokhar@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 most complete. 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 tRNA elements (SINEs) (tables S2 and S3), which is

**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 guanayl 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 triphosphate (ATPase) families are large, as in Arabidopsis, and overall, the complement of transporters in *Chlamydomonas* resembles that of both Osteoecoccus 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-
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* (1).]

<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>13x</td>
<td>6.7x</td>
<td>11x</td>
<td>nd</td>
<td>~8x</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>259§</td>
<td>nd</td>
<td>30</td>
<td>589</td>
<td>497</td>
</tr>
<tr>
<td>Selenocysteine (Sec) tRNAs</td>
<td>1</td>
<td>nd</td>
<td>nd</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*National Center for Biotechnology Information (NIH) NCBI 36 from Ensembl build 38. †Source (56). ‡Three regions contain 5S 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 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 (31), 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-

**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.
The numerous proteins in the GreenCut associated synthetic apparatus, where they are in high demand. and lipophilic cofactors into the metabolic machin- radation, trafficking, and integration of pigments proteins that likely function in the synthesis, deg-

protein with a pantothenate kinase motif. (conserved only in the green lineage. The GreenCut of their occurrence in various cuts. CGL refers to genes encoding proteins of unknown function that were not but not in nonphotosynthetic organisms. Genes en-

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, Physcomi-
trella, and Ostreococcus tauri and O. lucimaranus), but not in nonphotosynthetic organisms. Genes en-
coding 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).
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), described in 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 components, 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 just have 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 Ostrovoccus spp., a green alga lacking a flagellate stage. The Ostrovoccus 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 Ostrovoccus also lack basal bodies. However, Ostrovoccus spp. have retained many other CiliaCut proteins (table S14), which suggests either that they recently lost their flagellum, or that they retained flagellar proteins for other cellular functions.

Conclusions. The 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.

References and Notes

1. Materials and methods and supplemental online (SOM) text are available as supporting material on Science Online.
39. 12 OCTOBER 2007 VOL 318 SCIENCE www.sciencemag.org 250
Dislocation Avalanches, Strain Bursts, and the Problem of Plastic Forming at the Micrometer Scale

Ferenc F. Csikor,1,2 Christian Motz,3 Daniel Weygand,4 Michael Zaiser,5 Stefano Zapperi4,5*

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—characterized by intermittent strain bursts with scale-free (i.e., power-law) size distributions (1–8)—is—at least on the micrometer scale—ly 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

Received 4 April 2007; accepted 5 September 2007

References and Notes
9 April 2007; accepted 5 September 2007. 10.1126/science.1143609

www.sciencemag.org SCIENCE VOL 318 12 OCTOBER 2007 251

1Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA 90095, USA. 2U.S. Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA. 3CNRS, UMR 7141, CNRS/Université Paris 6, Institut de Biologie Physico-Chimique, 75005 Paris, France. 4Institute of Biotechnology, University of Delaware, Newark, DE 19716, USA. 5Department of Molecular and Cellular Biology, University of California at Berkeley, Berkeley, CA 94720, USA. 6Institut de Biologie Moléculaire des Plantes, CNRS, UMR 8618, Centre National de la Recherche Scientifique, Villeneuve d’Ascq, France. 7Laboratory of Macromolecular and Molecular Sciences Center and Excellence in Genomics and Bioinformatics, University of Tennessee, Memphis, TN 38163, USA. 8Department of Plant Biology, University of Minnesota, St. Paul MN 55108, USA. 9Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011, USA. 10Institute of Plant Sciences and Biotechnology, University of Arizona, Tucson, AZ 85721, USA. 11Department of Physics, University of Massachusetts Medical School, Worcester, MA 01605, USA. 12Department of Biology, University College London, London WC1E 6BT, UK. 13Unité de Glycobiologie Structurale et Fonctionnelle, UMR8576 CNRS/ UFR JFR 118, Université des Sciences et Technologies de Lille, Cite DENI, France. 14Université Potsdam, Institut für Biochemie und Biologie, D-14476 Golm, Germany. 15Department of Chemistry, National Jewish Medical and Research Center, Denver, CO 80206, USA. 16Delaware Biotechnology Institute, University of Delaware, Newark, DE 19711, USA. 17Institute of Biology II, University of Freiburg, 79104 Freiburg, Germany. 18Box Thompson Institute for Plant research at Cornell University, Ithaca, NY 14853, USA. 19Department of Chemistry and BioMolecular Sciences, Maquette University, Sydney 2109, Australia. 20Department of Anatomy and Cell Biology, Oklahoma State University, Center for Health Sciences, Tulsa, OK 74107, USA. 21Institute for Economiski Universitet, 35330 Balcova-Izmir Turkey. 22Institute of Microbiology, Czech Academy of Sciences, Czech Republic. 23Department of Plant Physiology, Faculty of Sciences, Charles University, 128 44 Prague 2, Czech Republic. 24Bioinformatics Program, St. Edward’s University, Austin, TX 78704, USA. 25Department of Biology, University of Puget Sound, Tacoma, WA 98407, USA. 26Institute of Low-Temperature Science, Hokkaido University, 060-0819, Japan. 27Department of Biology, Tsinghua University, Beijing, China 100084. 28Institute of Biological Chemistry, Washington State University, Pullman, WA 99164, USA. 29Appalachian State University, Boone, NC 28608, USA. 30Institute of Biology, Marquette University, Milwaukee, WI 53233, USA. 31UMR8567 CNRS, Laboratory of Biochemical, 59655 Villeneuve d’Ascq, France. 32Cell Signaling Laboratory, Stazione Zoologica, I 80121 Naples, Italy. 33Graduate Program in Biophysics, Ohio State University, Columbus, OH 43210, USA.

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
9 April 2007; accepted 5 September 2007.

1Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA 90095, USA. 2U.S. Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA. 3CNRS, UMR 7141, CNRS/Université Paris 6, Institut de Biologie Physico-Chimique, 75005 Paris, France. 4Institute of Biotechnology, University of Delaware, Newark, DE 19716, USA. 5Department of Molecular and Cellular Biology, University of California at Berkeley, Berkeley, CA 94720, USA. 6Institut de Biologie Moléculaire des Plantes, CNRS, 67084 Strasbourg Cedex, France. 7Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, CA 94143, USA. 8Biotechnology Research Center, Zhanjiang University, Guangzhou 520275, China. 9Department of Molecular Sciences and Center of Excellence in Genomics and Bioinformatics, University of Tennessee, Memphis, TN 38163, USA. 10Department of Plant Biology, University of Minnesota,
The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions

Editor's Summary