ANNUAL REVIEWS

Annual Review of Plant Biology

Comparative and Functional Algal Genomics

Crysten E. Blaby-Haas[1] and Sabeeha S. Merchant[2,3]

[1]Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA; email: cblaby@bnl.gov

[2]Departments of Plant and Microbial Biology and Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

[3]Institute for Genomics and Proteomics, University of California, Los Angeles, California 90095, USA

Annu. Rev. Plant Biol. 2019. 70:605–38

First published as a Review in Advance on March 1, 2019

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

https://doi.org/10.1146/annurev-arplant-050718-095841

Keywords

flavodoxin, sequence similarity network, cytochrome c_{6A}, chloroplast, endosymbiosis

Abstract

Over 100 whole-genome sequences from algae are published or soon to be published. The rapidly increasing availability of these fundamental resources is changing how we understand one of the most diverse, complex, and understudied groups of photosynthetic eukaryotes. Genome sequences provide a window into the functional potential of individual algae, with phylogenomics and functional genomics as tools for contextualizing and transferring knowledge from reference organisms into less well-characterized systems. Remarkably, over half of the proteins encoded by algal genomes are of unknown function, highlighting the volume of functional capabilities yet to be discovered. In this review, we provide an overview of publicly available algal genomes, their associated protein inventories, and their quality, with a summary of the statuses of protein function understanding and predictions.

Contents

1. INTRODUCTION

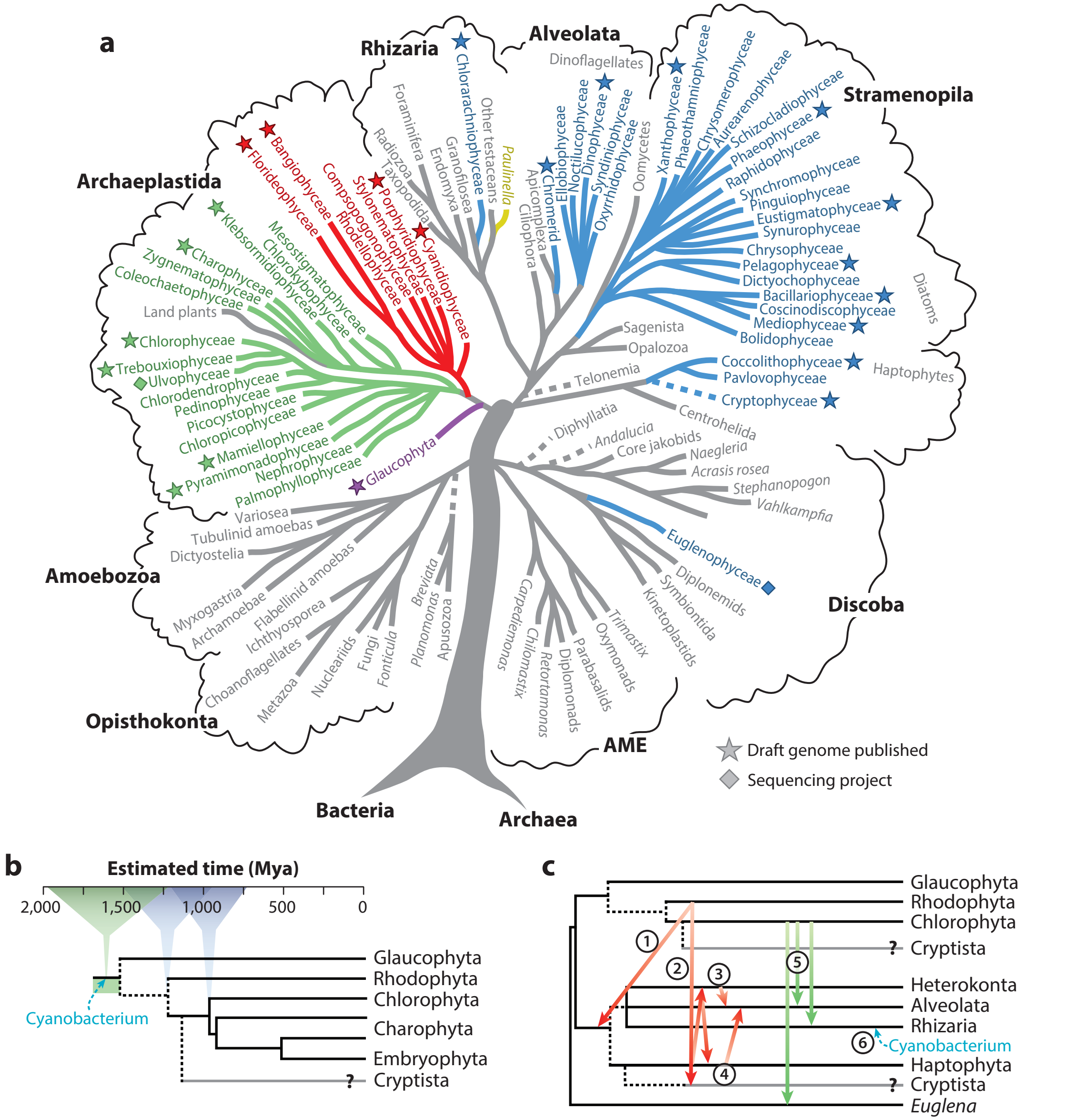
The evolution of oxygenic photosynthesis and the consequential rise in atmospheric oxygen levels drastically altered biology. The increase in global primary productivity and the availability of oxygen for new biochemical capabilities caused a shift in evolutionary trajectories resulting in the wealth of diversity we associate with life. The genes responsible for oxygenic photosynthesis and assimilation of carbon from CO_2 also had direct influence on the evolutionary landscape. Endosymbiosis accompanied by endosymbiotic gene transfer (the transfer of genes from an endosymbiont to the host nuclear genome) has spread the ability to photosynthesize between kingdoms, from bacteria to eukaryotes, and across Eukarya. Out of all of the resulting organisms capable of oxygenic photosynthesis, the algae represent the most diverse, complex, and understudied group.

1.1. What Is an Alga?

If we look at the diversity of photosynthetic eukaryotes that exists today, land plants (embryophytes), the only photosynthetic eukaryotes not referred to as algae, belong to a small corner of the tree of life that evolved roughly 500 million years ago, with the flowering plants (angiosperms) appearing around 200 million years ago (144, 173). In contrast, the algae are composed of organisms with deep branches in the eukaryotic lineage corresponding to an evolutionary time span of roughly 1.5 billion years (67, 92, 158) (**Figure 1**). Also unlike land plants, algae are distributed throughout the eukaryotic tree of life; they do not have a single common ancestor in the traditional sense. Instead, most algae are related through endosymbiosis that resulted in the transfer of plastids and genes to various eukaryotic hosts and created distinct lineages of algae outside of Archaeplastida (the phylogenetic clade to which photosynthetic eukaryotes with primary plastids reside) (**Figure 1**). The algal group is, therefore, polyphyletic because the most recent common ancestor of all eukaryotic algae was not an alga, and many algae are more closely related to nonphotosynthetic protists than they are to other algae (**Figure 1**). Until recently, the plastids of algae were thought to have a monophyletic origin: a single primary endosymbiotic event involving the engulfment of a now-presumed-extinct cyanobacterium by the common ancestor of Archaeplastida. Nevertheless, at least one exception is known. The photosynthetic bodies in *Paulinella* species, referred to as chromatophores, originated from an independent primary endosymbiotic event that occurred as recently as 60–90 million years ago (55, 153).

1.2. Tracing the Origins of Plastids

The unifying characteristic of the algal group is the presence of a photosynthetic plastid (97, 122, 154, 185), but tracing the evolutionary history of these organelles is not straightforward. Algae in Archaeplastida and *Paulinella* contain primary (1°) plastids that originated from an endosymbiotic



(*Caption appears on following page*)

Figure 1 (*Figure appears on preceding page*)

(*a*) A cartoon depiction of the eukaryotic tree of life based on consensus phylogeny of eukaryotes from Baldauf (15, 16) and incorporating phylogenetic relationships of algal classes from References 63, 123, and 235. Branches containing algae are colored green for green algae from Chlorophyta and Streptophyta, red for rhodophytes, purple for glaucophytes, blue for algae whose photosynthetic plastids are derived from endosymbiosis of a eukaryotic alga (i.e., secondary or tertiary endosymbiosis), and yellow for the genus *Paulinella*, which contains algae with primary plastids that are distinct from the primary plastids of Archaeplastida. Dashed lines represent uncertain relationships. AME is used to abbreviate amitochondriate excavates. For each lineage of algae shown in the tree, the presence of at least one publicly released whole-genome assembly or sequencing project is indicated with a star or diamond, respectively. (*b*) The major lineages of Archaeplastida. The ranges of estimated time in million years ago (Mya) for the endosymbiotic event (*green*) and last common ancestors (*blue*) are based on molecular clock estimates from References 67, 92, 101, 102, 104, 115, 158, 189, 239, and 243. Affinity between Cryptista and Archaeplastida observed in Reference 37 is represented with a dotted line. Among the algal lineages within Archaeplastida, Charophyta, which contains the streptophyte algae, is polyphyletic. (*c*) Various theories for the transfer of plastids and genes from Archaeplastida to protists. Events involving a red alga are indicated with red arrows, and events involving a green alga are indicated with green arrows: ① The chromalveolate hypothesis proposed by Cavalier-Smith (44) suggests that endosymbiosis of a red alga happened only once in a common ancestor of algae bearing chlorophyll *c*–containing plastids and that all nonplastid relatives are examples of former algae. ② In recent years, the chromalveolate hypothesis has been superseded by hypotheses involving multiple engulfment events. Shown here is the cryptophyte-first hypothesis proposed by Stiller et al. (207), which holds that a cryptophyte was the original host of a red alga–derived plastid that was subsequently spread to an ancestral ochrophyte within Heterokonta and then to haptophytes. ③ As an example of an extension of the cryptophyte-first hypothesis, the *Vitrella brassicaformis* plastid may be derived from an ochrophyte (195). ④ The modern-day plastid in dinoflagellates from Kareniaceae is derived from a haptophyte, which replaced the peridinin plastid that is found in other dinoflagellates. ⑤ The plastids in euglenophytes (*Euglena*), chlorarachniophytes (Rhizaria), and *Lepidodinium* (Alveolata) are hypothesized to have originated from different green algae. ⑥ *Paulinella* chromatophores are derived from endosymbiosis of a cyanobacterium.

relationship with a cyanobacterium (reviewed in 154). Algae outside these two groups contain plastids that originated from an endosymbiotic relationship with a eukaryotic alga. We refer to these plastids as secondary or tertiary (2°/3°) plastids because they are derived from the engulfment of an alga with a 1° plastid or 2° plastid, respectively. However, higher-order relationships involving engulfment of algae with 3° plastids have been proposed (discussed in 36).

The engulfment and retention of eukaryotic algae has created a complex array of plastids across the algal group. The plastids in euglenophytes, chlorarachniophytes, and the dinoflagellate genus *Lepidodinium* are derived from independent endosymbiotic relationships with a green alga from Chlorophyta (115). Dinoflagellates in Kareniaceae have a fucoxanthin-containing plastid derived from a haptophyte alga, whereas most other dinoflagellates have a peridinin-containing plastid (87, 114). In the case of other algae within the stramenopile, chromerid, dinoflagellate, cryptophyte, and haptophyte groups, the exact order and number of endosymbiotic events is a topic of contention (36, 227), but their plastids are most closely related to red algae (Rhodophyta) (**Figure 1**). In addition to the plastid genome of cyanobacterial origin found in all photosynthetic plastids, chlorarachniophytes and cryptophytes have periplastid-localized nucleomorphs, which are relatively small remnants of nuclear genomes from the engulfed eukaryotic green or red alga, respectively (141). There also exist a number of organisms that are sometimes referred to as algae but have impermanent plastids. These acquired phototrophs either have a symbiotic relationship with an alga, such as *Paramecium bursaria* with endosymbiotic *Chlorella* spp. (119), or are able to engulf and steal plastids from algae (kleptoplasty). An example of the latter is the ciliate *Myrionecta rubra*, which acquires plastids from cryptophyte algal prey (116). In these cases, plastid retention is temporary and feeding on algae is needed to replenish their supply of plastids.

1.3. Phenotypic Diversity of Algae

In addition to the evolutionary distances among the major algal groups, the phenotypic variety observed in algae is remarkable. As an example, the chlorophyte lineage contains both the smallest and the largest known free-living single-celled eukaryotes, *Ostreococcus tauri* (50) and *Caulerpa taxifolia* (137), respectively, and multicellular forms that range in size from the colonial alga *Tetrabaena socialis* (5) at 20 µM to the seaweed *Ulva lactuca* at 3 ft (206). The largest alga, *Macrocystis pyrifera*, a heterokont, grows in underwater beds commonly compared to redwood forests since this brown alga can reach a length of 200 ft (152). Algae occupy a wide range of ecological niches and are typically found in temperate and tropical soil, fresh water, and the oceans. Extremophilic algae have also been described. The halophilic green alga *Dunaliella salina* inhabits the Northern arm of the Great Salt Lake, Utah, where the NaCl concentration is oversaturated (35). The green alga *Dunaliella acidophila* survives in an environment of 1 M H^+ (pH 0), with a growth maximum at pH 1 (93), whereas red algae in the order Cyanidiales thrive at pH 0.5–3 and high temperature (50–55°C) (43). Psychrophilic green algae, such as *Chlamydomonas raudensis* (UWO 241), inhabit permanently ice-covered lakes in Antarctica (143, 164), whereas snow algae often blanket glaciers (107), and diatoms inhabit brine channels in sea ice (213). Other algae—such as the endolithic algae of the hyperarid, polyextreme Atacama Desert in Chile (233) or the green alga *Chlorella ohadii*, which was isolated from the Negev Desert in Israel (219)—have evolved to cope with extremes in temperature, desiccation, and light intensity.

2. GENOMES

The phenotypic and ecological niche diversity among algae hints at the breadth of functional capabilities encoded by their genomes. Algae contain at least three separate genomes, with the nuclear genome containing the vast majority of genetic material and coding potential. The most gene-rich organelle genome known belongs to the red macroalga *Grateloupia taiwanensis* (61) and contains 233 protein-coding genes. The least gene-rich nuclear genome known belongs to the red microalga *Cyanidioschyzon merolae* and encodes 4,775 proteins (155). Because of the larger size of nuclear genomes and their propensity for repetitive regions, which makes assembly of sequencing reads more difficult (217), we have access to fewer nuclear genomes compared with plastid and mitochondrial genomes. Nevertheless, with the use of hybrid strategies incorporating short- and long-read technologies [e.g., Illumina and PacBio (186)], and because of greater speed and quality combined with decreasing costs, we have access to ever more algal nuclear genomes each year. We are approaching the 100th published algal nuclear genome, and with increasing recognition of the biotechnological, nutraceutical, and environmental value of these organisms, this number is expected to double in the next two to three years (**Figure 2**). With ambitious projects, such as the 10KP Genome Sequencing Project (45), which proposes to sequence the genomes of at least 1,000 green algae and 3,000 photosynthetic and nonphotosynthetic protists in the next five years, access to algal genomes is expected to increase rapidly. This review is intended to provide a timely snapshot of an exponentially growing field with an emphasis on the role of genomics in generating new paradigms for the way we understand algal biology.

Since the first draft whole-genome sequence of an alga was released in 2004 (134), sequencing technology and the accompanying computational methods for assembly and structural annotations have improved and continue to do so. The *Chlamydomonas reinhardtii* genome, which was published in 2007 (139), serves as an example of how advances in both sequencing technologies and computational methods have contributed to continuous improvements over the intervening decade and emphasizes that genome projects for reference organisms are not end points at publication (20). Only the relatively small nuclear genomes of the red alga *C. merolae* (155) and the prasinophytes *Micromonas commoda* RCC299 (236) and *Ostreococcus lucimarinus* (157) are considered finished [i.e., telomere-to-telomere assembled chromosomes without gaps; however, finished does not necessarily equate to perfect. Errors could still be present, such as assembly artifacts

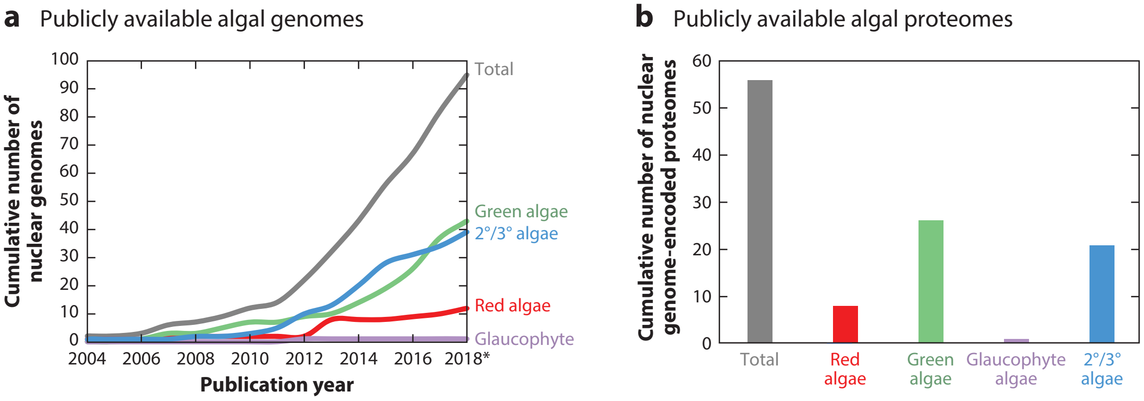


Figure 2

As second- and third-generation sequencing technologies have increased the speed and decreased the cost of whole-genome sequencing, access to algal nuclear genomes is growing. (*a*) The cumulative number of nuclear genome assemblies reported over time for each major algal lineage and for all lineages (total). The asterisk signifies that the data are current as of June 2018. (*b*) The quality of these assemblies varies significantly, and as such, only about half of these genome-sequencing projects are accompanied by a public release of gene models and predicted protein sequences.

caused by repeat regions, especially in centromeres and telomeres (196)]. The completeness of the other publicly available algal genomes ranges from chromosome-level assemblies with few gaps and unplaced continuous runs of sequence (contigs), such as for *Chromochloris zofingiensis* (187), to assemblies with tens of thousands of scaffolds containing ordered contigs separated by gaps, such as for the large, highly repetitive draft genome of *Cymbomonas tetramitiformis* (38) (**Table 1**).

In addition to the number of contigs and scaffolds, several metrics are used to assess quality, contiguity, and completeness of the assembly. The most popular technical metrics are the N50 length and the L50 count. These metrics can give a sense of the contiguity of the assembly and can be used to judge whether the assembled contigs are long enough to have captured most genes as full-length sequences. N50 length and L50 count can also be useful for assessing improvements in an assembly (assuming the overall sequenced length does not change); contiguity increases as the N50 length increases and L50 count decreases. Caution should be used when using the N50 and L50 statistics to compare genomes because these statistics are a measure of only assembly contiguity, not completeness. For instance, the most recent published genome assemblies for the dinoflagellate *Symbiodinium minutum* (199) and the kelp *Saccharina japonica* (238) are associated with similar N50 and L50 statistics, but only 50% of the *S. minutum* genome was assembled, whereas 98% of the *S. japonica* genome was assembled. The sizes of algal genomes that are sequenced or are being sequenced differ by nearly two orders of magnitude; the finished genome of the red alga *C. merolae* is 16.5 Mbp (155), whereas 616 Mbp of the estimated 1,500 Mbp of the *S. minutum* genome have been sequenced (199). The *S. minutum* genome is actually considered small among dinoflagellates, with the largest known dinoflagellate genome estimated to be 185 Gbp (42). Therefore, across the algal lineages, genome sizes vary by four orders of magnitude.

A separate method for assessing and comparing genome sequences is estimating completeness based on searches for universally conserved protein sets, such as with the Core Eukaryotic Genes Mapping Approach (CEGMA) (159) and Benchmarking Universal Single-Copy Ortholog (BUSCO) (200, 229). These ortholog searches can be used to assess the genome assembly and the set of gene models independently. The assumption is that the more full-length universal orthologs are found, the more complete is the genome assembly or set of gene model predictions.

Table 1 Published algal genomes with publicly available gene models

Taxonomy	Organism	Sequenced length (Mb)	Nucleotide sequence accession number[a]	Scaffolds[b]	Contigs	Protein count	Complete BUSCOs found	Reference(s)
Chlorophyta								
Mamiellophyceae	*Bathycoccus prasinos* RCC1105	15.07	GCA_002220235.1	21	41	7,847	255	142
	Micromonas sp. RCC299	21.109	GCA_000090985.2	19	19	10,137	272	236
	Micromonas sp. CCMP1545	21.958	GCA_000151265.1	21	497	10,242	265	236
	Micromonas sp. ASP10–01a	19.582	GCA_001430725.1	1,069	3,159	22,252	179	56
	Ostreococcus lucimarinus CCE9901	13.205	GCA_000092065.1	21	21	7,603	256	157
	Ostreococcus tauri RCC4221	13.033	GCF_000214015.3	22	499	7,766	265	28, 62
	Ostreococcus tauri RCC1115	14.763	GCA_002158475.1	70	200	8,099	255	27
Trebouxiophyceae	*Chlorella variabilis* NC64A	46.159	GCA_000147415.1	414	3,810	9,780	258	26
	Auxenochlorella protothecoides	22.924	GCA_000733215.1	374	1,386	7,014	258	89
	Chlorella sorokiniana UTEX 1602	59.568	GCA_002245835.2	NA	160	9,587	287	11
	Micractinium conductrix SAG 241.80	61.02	GCA_002245815.1	NA	301	10,070	290	11
	Chloroidium sp. UTEX 3007	52.5	**https://datadryad.org/resource/doi:10.5061/dryad.k83g4**	710	NA	48,392	9[c]	148
	Coccomyxa subellipsoidea C-169 NIES 2166	48.827	GCA_000258705.1	29	29	9,839	275	25
	Helicosporidium sp. ATCC 50920	12.374	GCA_000690575.1	5,666	5,666	6,033	170	166
	Picochlorum SENEW3 (SE3)	13.39	GCA_000876415.1	880	883	7,367	279	84
	Picochlorum soloecismus DOE101	15.252	GCA_002818215.1	38	56	7,844	257	96

(*Continued*)

Table 1 (*Continued*)

Taxonomy	Organism	Sequenced length (Mb)	Nucleotide sequence accession number[a]	Scaffolds[b]	Contigs	Protein count	Complete BUSCOs found	Reference(s)
Chlorophyceae	*Chlamydomonas reinhardtii* CC-503	111.1	v5.5 (Phytozome)	53	1,512	17,741	291	139
	Volvox carteri f. nagariensis, Eve	131.2	v2.1 (Phytozome)	434	4,100	14,247	287	169
	Chlamydomonas eustigma NIES-2499	66.63	GCA_002335675.1	520	2,507	14,161	289	106
	Dunaliella salina CCAP 19/18	343.704	GCA_002284615.1	5,512	55,168	18,801	235	165
	Gonium pectorale NIES-2863	148.806	GCA_001584585.1	2,373	13,664	17,984	247	99
	Monoraphidium neglectum SAG 48.87	69.712	GCA_000611645.1	6,720	12,077	16,755	177	30
	Tetrabaena socialis NIES-571	135.78	GCA_002891735.1	5,856	20,418	14,296	138	78
	Chromochloris zofingiensis SAG 211–14	60.13	v5.2.3.2 (Phytozome)	42	175	15,274	256	187
	Raphidocelis subcapitata NIES-35	51.163	GCA_003203535.1	300	1,620	13,383	278	209
Charophyta								
Klebsormidiophyceae	*Klebsormidium nitens* NIES-2285 (formerly *K. flaccidum*)	104.21	GCA_000708835.1	1,814	3,731	16,063	298	110
Rhodophyta								
Florideophyceae	*Chondrus crispus* Stackhouse (Gigartinales)	104.98	GCA_000350225.2	926	3,242	9,807	232	47
Cyanidiophyceae	*Cyanidioschyzon merolae* 10D	16.547	GCA_000091205.1	20	20	4,803	283	134, 155
	Galdieria phlegrea DBV009	11.4	**http://cyanophora.rutgers.edu/gphlegrea/**	NA	9,311	7,828	207	172
	Galdieria sulphuraria	13.712	GCA_000341285.1	433	518	7,174	279	193
Bangiophyceae	*Porphyra umbilicalis*	87.7	v1.5 (Phytozome)	2,125	2,183	13,360	189	33
	Pyropia yezoensis U-51	43.484	SRA061934	NA	46,634	10,327	135	146

(*Continued*)

Table 1 (*Continued*)

Taxonomy	Organism	Sequenced length (Mb)	Nucleotide sequence accession number[a]	Scaffolds[b]	Contigs	Protein count	Complete BUSCOs found	Reference(s)
Porphyridiophyceae	*Porphyridium purpureum* CCMP1328	19.452	GCA_000397085.1	3,014	3,014	8,355	273	18
Glaucophyta								
Glaucophyceae	*Cyanophora paradoxa*	70.2	SRP009206	NA	60,119	32,167	175	168
Miozoa								
Dinophyceae	*Symbiodinium minutum*	609.476	GCA_000507305.1	21,899	33,816	41,925	191	199
	Symbiodinium kawagutii	935	SRA148697	NA	NA	36,850	82	128
	Symbiodinium microadriaticum CCMP2467	808.227	GCA_001939145.1	9,688	44,596	43,403	210	6
Apicomonadea	*Chromera velia* CCAP 1602/1	193.6	ERP006228 (ENA)	5,953	13,987	26,112	241	235
	Vitrella brassicaformis	72.7	GCA_001179505.1	1,064	4,175	23,034	264	235
Cercozoa								
Chlorarachnea	*Bigelowiella natans* CCMP2755	91.406	GCA_000320545.1	3,736	3,736	21,708	263	52
Ochrophyta								
Pelagophyceae	*Aureococcus anophagefferens* CCMP1984	56.66	GCA_000186865.1	1,185	5,239	11,520	203	94
	Cladosiphon okamuranus	169.731	GCA_001742925.1	732	4,525	13,640	245	150
	Ectocarpus siliculosus (Dillwyn) Lyngbye	195.811	GCA_000310025.1	1,561	13,533	16,269	274	46
Mediophyceae	*Cyclotella cryptica*	161.7	**http://genomes.mcdb.ucla.edu/Cyclotella/download.html**	NA	116,817	21,121	257	216
	Thalassiosira oceanica CCMP1005	92.1856	GCA_000296195.2	NA	50,892	34,642	197	130
	Thalassiosira pseudonana CCMP1335	32.437	GCA_000149405.2	64	115	11,673	245	9

(*Continued*)

Table 1 (*Continued*)

Taxonomy	Organism	Sequenced length (Mb)	Nucleotide sequence accession number[a]	Scaffolds[b]	Contigs	Protein count	Complete BUSCOs found	Reference(s)
Bacillariophyceae	*Fragilariopsis cylindrus* CCMP1102	80.54	GCA_001750085.1	271	4,602	18,111	247	140
	Fistulifera solaris	49.74	GCA_002217885.1	295	1,305	20,429	273	210
	Phaeodactylum tricornutum CCAP 1055/1	27.451	GCA_000150955.2	88	179	10,408	257	31
Eustigmatophyceae	*Nannochloropsis gaditana* CCMP526	33.987	GCA_000240725.1	1,883	5,619	3,554	78	174
	Nannochloropsis gaditana B-31	27.589	GCA_000569095.1	684	3,880	10,929	241	49
	Nannochloropsis oceanica CCMP1779	28.7	SRP013753	NA	3,731	11,973	230	226
Haptophyta								
Prymnesiophyceae	*Chrysochromulina tobin* CCMP291	59.073	GCA_001275005.1	3,412	34,112	16,765	190	111
	Emiliania huxleyi CCMP1516	167.676	GCA_000372725.1	7,795	16,921	38,554	225	181
	Tisochrysis lutea	54.38	SRR3156597	NA	7,662	20,582	217	41
Cryptista								
Cryptophyceae	*Guillardia theta* CCMP2712	87.1453	GCA_000315625.1	669	5,126	24,840	254	52

Abbreviations: BUSCO, Benchmarking Universal Single-Copy Ortholog; ENA, European Nucleotide Archive; NA, not available.

[a] Accession numbers are from the US National Center for Biotechnology Information (NCBI) unless otherwise noted. Except for the number of BUSCOs found, information is given as reported by NCBI or in the indicated publication.

[b] NA signifies that the information was not found on NCBI nor in the referenced publication or that the assembly does not contain scaffolds.

[c] For the released protein set from *Chloroidium* sp. UTEX 3007, 125 fragmented BUSCOs were detected in addition to these complete BUSCOs.

2.1. Structural Annotations: Caveat Emptor

The raw genome sequence gives little insight into biological function by itself. The first step in decoding the genome sequence is structural annotation. Structural annotations specify where in the assembly genomic features, such as genes, coding sequences, transcription start and stop sites, and alternative splice sites, are located. The completeness and accuracy of structural annotations vary across algae. In part, gene model predictions are only as accurate as the underlying genome sequence. A gap in the assembly that falls within a gene can result in incorrect splitting of that gene into two or more models, and exons may be partially or entirely missing from the model. Sometimes sequence gaps in exons are represented as stretches of X's in predicted amino acid sequences, but other potential inaccuracies are not always evident from protein sequences deposited in databases. Depending on users' research objectives, manual assessment of the quality of individual gene models may be needed. In the worst-case scenario, genes may be missing from the assembly because of lack of sequence coverage or presence of highly repetitive regions that are recalcitrant to the assembly of sequencing reads. Even with finished genomes, structural annotations can be inaccurate, and typically algal research communities must invest significant resources to increase the number of evidence-based gene models both prior to and subsequent to publication of the genome sequence (20, 28, 48, 176, 225).

2.2. Defining the Parts List: Functional Annotations

In the postgenomic era, research efforts are focused on the development of tools to decode the functional significance of specific sequences in the context of biology. No single approach or battery of techniques is useful to generate an experimentally determined functional annotation for every protein encoded in every organism's genome. Even for *Arabidopsis thaliana*, arguably the most thoroughly investigated photosynthetic eukaryote, only 30% of functional annotations are associated with experimental evidence (12, 40). In algae, which as a group are relatively uncharacterized at the genetic level, the functional annotations of most proteins, like all nonreference (and many reference) organisms, are derived from sequence similarity searches against one or more databases. Compared with *A. thaliana*, which was the first photosynthetic eukaryote to have its genome sequenced, nearly half of predicted algal proteins are not associated with a Pfam domain, nor do they map onto any of the nearly 1.2 million orthologous groups defined by the EggNOG database (**Figure 3**), giving us an indication of the considerable potential for new discovery.

For the half of protein sequences that can be annotated by sequence similarity, the reliability of many annotations is unknown (21, 167), and automated functional annotation can be prone to both misannotation and overannotation (126, 192). While some reliability estimates of similarity-based approaches to functional annotation are available (214), this method of annotation is confounded by the observation that function may not be conserved between even highly similar sequences. Examples include RAF2, which based on sequence similarity alone would be predicted to be a pterin-4α-carbinolamine dehydratase, but phylogenomic and functional characterization suggests this protein lacks enzymatic activity and, surprisingly, is involved in assembly of Rubisco (79, 147, 230). Another example is offered by the algal protein LFO1, which was originally annotated based on its similarity to the antibiotic monooxygenases in the Pfam database. Subsequent phylogenomic analysis and experimental characterization instead supports a role for this protein as a heme-degrading enzyme involved in the response to Fe limitation (129). Conversely, proteins that do not share sequence similarity may have the same function. Classic examples are distinct families of enzymes, such as the three classes of carbonic anhydrases (39), the three superoxide dismutase families (2, 163), or plastocyanin and cytochrome c_6 (54), whose shared functions arose through convergent evolution. While genome-wide searches against databases are a quick way

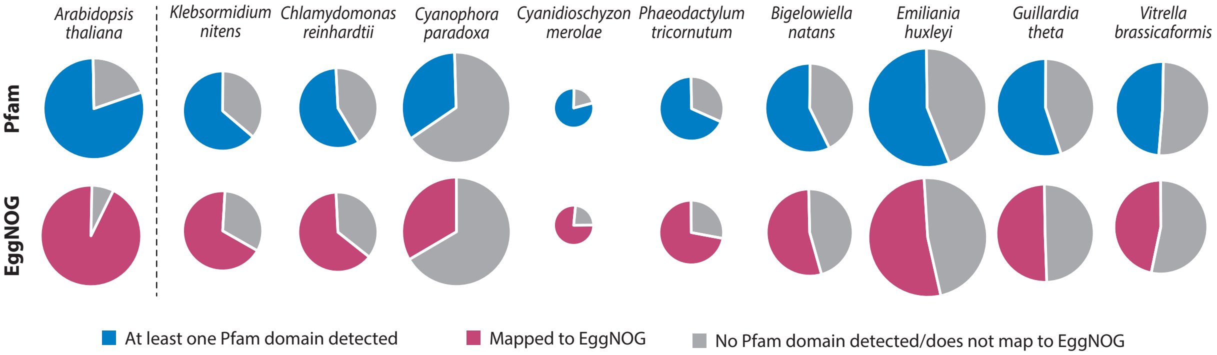


Figure 3

Algal genomes provide a reservoir of undiscovered functional capabilities. As an indicator of the amount of protein function yet to be discovered in algae, the number of predicted proteins from each whole-genome sequence is shown, where each protein either contains at least one Pfam-A domain (detected by using hmmsearch with Pfam-trusted score cutoffs) (82) or maps to an orthologous group from the EggNOG database using DIAMOND (113). For comparison, the same analyses for the choice reference land plant *Arabidopsis thaliana* are also shown. The pie charts are scaled relative to the total size of each predicted proteome. For each of the major lineages of algae (streptophyte algae, chlorophyte algae, glaucophytes, red algae, stramenopiles, rhizarians, haptophytes, cryptophytes, and alveolates), only one representative is shown and was chosen as having the highest number of complete Benchmarking Universal Single-Copy Orthologs (BUSCOs) compared with other predicted proteomes from closely related algae. Although the proportion of each proteome that maps to Pfam or EggNOG is roughly similar, the two are not perfectly overlapping. For instance, for *Cyanophora paradoxa*, about 2,300 proteins have at least one Pfam domain but do not map to EggNOG, whereas about 2,200 proteins map to EggNOG but do not have a Pfam domain. These analyses are conservative and serve to illustrate the distance between algal proteomes and sequences presently in databases. For instance, a more sensitive method, Domain Annotation by a Multi-objective Approach (DAMA) and CLADE, was recently used to detect conserved domains in 99% of the proteins in *Phaeodactylum tricornutum* (176). Assessing the number of proteins in these algae where the function is known or can be confidently predicted based on sequence similarity to a database, such as the Reference Sequence Database (RefSeq) of the US National Center for Biotechnology Information (NCBI) (170), UniProt (223), EggNOG, or a match to Pfam's domain models, is difficult or impossible to determine quantitatively. As part of the Gene Ontology (GO) Consortium's Reference Genome Project (182), a comprehensive set of GO annotations for the *A. thaliana* genome is available, and, as of June 2018, 30% of the genes in the *A. thaliana* genome are associated with an experimentally supported GO term for either a molecular function or a biological process (40).

to identify highly conserved proteins or domains, orthology predictions and community-led efforts involving manual curation provide more reliable functional annotations (86, 201). Indeed, in the postgenomic era, a goal for the foreseeable future is integrating and leveraging available genomic and postgenomic data to decipher protein function, prioritizing targets for experimental characterization, and incorporating that knowledge into a functional framework for each system.

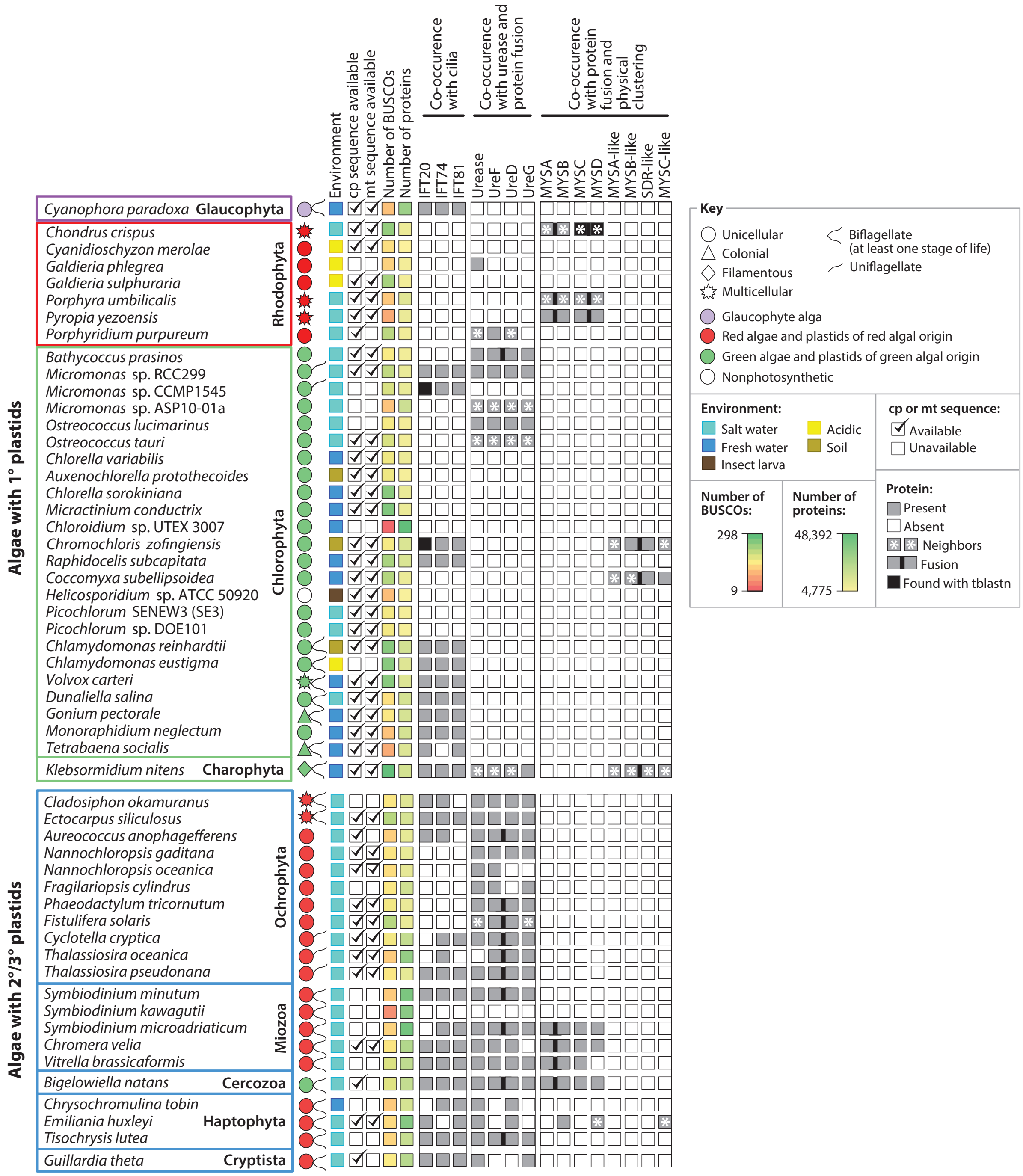
2.2.1. From sequence to function: integrating genomic and postgenomic data. Phylogenomics is a term put forth by Eisen and colleagues (68, 69) in the late 1990s to refer to a strategy for improving function predictions by considering the evolutionary history of proteins. The basic tenet of this approach is that orthologs share the same function, while paralogs may have diverged in function. Phylogeny provides a framework upon which different types of information from different family members are combined to inform the function of the family and/or subfamilies. In parallel with phylogeny, protein similarity networks are useful for visualizing large protein families and for defining subgroups of proteins based on pairwise amino acid sequence similarity (13, 91). Sequence similarity networks have been used to explore proteins involved in algal metal transport (22), plastid-targeted transporters (120), and algal transient receptor potential channels (8). Phylogeny is a tool for reconstructing protein families based on evolutionary models of amino acid substitution from a multiple sequence alignment. However, other than the evolutionary model used to calculate the E value of a pairwise BLASTp score, sequence similarity networks do not provide a rigorous hypothesis regarding the evolutionary distances or relationships between protein family members. Instead, these networks allow one to visualize which protein sequences are more similar to each other based on the E value threshold cutoff used to define similarity. When combined with phylogenetic trees, sequence similarity networks are powerful tools in visualizing functional information, such as the presence of protein domains, condition-specific gene regulation, and phenotypes of corresponding mutants.

2.2.2. Comparative genomics as a tool for protein function discovery. There are several types of comparative genomic analyses that can provide additional inferences for protein function prediction and for building confidence in an automated functional annotation. The basic principle is guilt by association (7). Associations between proteins of known function and unknown or uncertain function can provide evidence for the function of the latter. The two main types of guilt-by-association data that come directly from structurally annotated eukaryotic genomes are phylogenetic profiles and protein fusions (**Figure 4**).

Phylogenetic profiles are used to infer functional coupling between proteins by assuming that during evolution functionally related proteins are maintained or eliminated in a correlated manner (156, 162). Functional coupling can also be generated between a set of proteins and a phenotype, morphology, or metabolic capability. Large-scale analyses of this type have identified algal proteins involved in photosynthesis and chloroplast biology (121, 139), cilia biogenesis (127), carbohydrate uptake and metabolism (17), and biosynthesis of sulfated polysaccharides (80). Occurrence profiles can also be useful in identifying missing genes in pathways. Candidates for both locally missing genes (the gene for a particular step in a pathway is found in some genomes but missing from others), which are likely cases of nonorthologous gene displacement (231), and globally missing genes (the gene encoding a particular pathway step has not been identified in any genome) can be identified based on co-occurrence with other pathway genes (156).

Fusion proteins are valuable resources in that two proteins with separate activities are encoded by a single gene. These fusions, often termed Rosetta stone proteins, can provide evidence for a functional interaction, such as membership in the same pathway and/or components of a metabolon. As such, the known function of one domain can inform the function of the fused domain (132). It is estimated that up to 65% of eukaryotic proteins are multidomain proteins (72), and given the mosaicism created by horizontal gene transfer from endo- and exosymbionts to the nuclear genomes of photosynthetic eukaryotes, algal genomes are expected to be rich in unique gene fusions. Indeed, a recent comparative genomic analysis identified 67 protein families from various algae that contain fusion proteins where at least one domain is predicted to have originated from the cyanobacterial endosymbiont (136).

2.2.3. Functional genomics as a tool for protein function discovery. Functional genomics data sets, such as transcriptomics, proteomics, and genome-wide mutant screens, can provide additional layers of gene-specific functional data. While these experiments supply global snapshots of cellular behavior under different conditions, functional inference and associations can also be derived by identifying the specific conditions under which a protein is expressed (e.g., when and in what situation the cell requires that protein), by determining coexpressed proteins (e.g., proteins involved in the same process), and by localizing proteins to specific subcompartments within the cell (100, 188). Examples of high-throughput experimentally determined functional inferences associated with sequenced algae that can be used for protein function predictions include identification of proteins found in cilia (85, 160), the eyespot (71), oil bodies (149), the pyrenoid (241), the nucleus (234), the mitochondrion (14), or the chloroplast (19, 108, 212). Studies in which



(*Caption appears on following page*)

Figure 4 (*Figure appears on preceding page*)

A list of published algal proteomes predicted from whole-genome sequencing projects (as of June 2018). Citations for each genome are available in **Table 1**. Icons to the right of the name designate cellularity, whether the alga has cilia, and whether the alga [or the origin of the plastid from algae with secondary/tertiary (2°/3°) plastids] is from the glaucophyte, red, or green lineage. *Helicosporidium* sp. ATCC 50920 is from the Chlorophyta lineage but has recently lost photosynthetic capacity (211). The most common environment from which each alga has been isolated is also given. For most but not all algae listed, both the chloroplast-localized (cp) and the mitochondrion-localized (mt) genome sequences are also publicly available; availability is indicated with a checked box. The quality/completeness of the predicted proteome is indicated with a heatmap representing the number of complete Benchmarking Universal Single-Copy Orthologs (BUSCOs) detected (out of 303). The number of total proteins in each proteome is also shown as a heatmap. Both the number of BUSCOs and total number of proteins can be found in **Table 1**. Phylogenetic profiles require high-quality genomes and gene models for confidence that the absence of a gene is due to evolution rather than a gap in the sequence. For instance, the *Galdieria phlegrea* urease assembly factors were found by targeted sequencing (172), but these genes are not present in the public genome sequence. In some cases, a missing protein is due to an inaccurate gene model and can be recovered by searching the genome with tblastn. Whether a gene/protein was lost/never acquired or whether it is missing from the assembly can be better predicted if that gene/protein is also missing from closely related whole-genomes (for instance from the same genus). Additionally, the absence of functionally related proteins, such as cohorts in the same pathway, can provide support for gene loss. Comparative genomics–based analyses present hypotheses about the biology of organisms. As an example, protein components of intraflagellar transport (IFT) (three components are shown here) are encoded only by genomes belonging to algae that have cilia. Although cilia have not been described in *Chromochloris zofingiensis* and *Raphidocelis subcapitata*, the presence of genes related to cilia biogenesis and biology supports the presence of cilia in a yet-to-be-described stage of life (187). The co-occurrence of urease and assembly factors provide an example of how proteins that interact with each other in the cell co-occur. As seen previously, two or more proteins may be fused and proteins may be encoded by genes that are next door to one another in the genome. Both of these observations strengthen functional association between these enzymes. Although neighborhoods of functionally related genes are thought not to be as prevalent in eukaryotic genomes as in prokaryotes (because eukaryotes typically lack operons transcribed as polycistronic mRNA), physical proximity of some functionally related genes has been observed in algal genomes, as shown here for genes encoding urease and assembly factors and for genes predicted to be responsible for biosynthesis of UV-absorbing/screening mycosporine-like amino acids (MAAs) (33). Orthology between red algal MAA biosynthesis proteins with related proteins in green algae is not clear, but physical clustering of the corresponding genes supports a functional link between these green algal homologs, which leads to the prediction that these proteins may be responsible for MAA biosynthesis or, likely, a similar product since the MYSB homolog is uniquely fused to a protein from the short-chain dehydrogenase/reductase (SDR) family. While protein fusions can be an artifact of inaccurate gene models, the presence of at least two protein fusions from independent genomes is a good indication that the fusion is real.

collections of mutants are sequenced to identify affected loci causing a specific phenotype are another way to group genes involved in specific processes. Collections of temperature-sensitive lethal alleles (34, 221) and sequenced mutants with photosynthetic defects (59, 60) are available. Over 5,500 accessions of sequenced RNA from algae are deposited in the Sequence Read Archive (SRA) of the US National Center for Biotechnology Information (NCBI) (accessed July 2018). More than half of these data are from algae with published genome assemblies and gene models (**Table 1**). The other half are de novo assembled transcriptomes or from algae whose genomes are presently being sequenced or are publicly available but the predicted proteins are not public (**Supplemental Table 1**).

Supplemental Material >

3. WHAT HAVE ALGAL GENOMES REVEALED SO FAR?

The value of whole-genome sequencing cannot be overstated. To understand the genetic underpinnings of algal biology and achieve systems and synthetic biology objectives for algae, genome sequences are essential. De novo transcriptomics is a powerful tool for providing a snapshot of expressed genes/proteins under the conditions sampled, but high-quality genomes are needed for access to promoters and regulatory elements, intron/exon structure, centromeres, epigenomics, and a complete repertoire of genes. Whole-genome sequences are also invaluable resources for designing and building the genetic tools needed for both bioengineering applications and protein function discovery. As an example, a prevailing roadblock in algal-based industrial biofuel production is the observation that the storage lipid triacylglycerol accumulates typically during stress conditions, which increases triacylglycerol content per cell but inhibits growth. To address this issue, genetic or metabolic engineering efforts have targeted neutral lipid production, and successful strategies are reported largely for algae where genomic resources are available (3, 58, 77, 95, 131, 151, 218, 237).

Genomes provide unprecedented insight into the evolution of algae and their nonalgal relatives and are providing mechanistic insight into the genomic foundation of adaptation. Whole-genome sequences have strengthened support for a single event of plastid endosymbiosis at the base of Archaeplastida (134, 168) and weakened support for a single event of red plastid acquisition outside of Archaeplastida (207). Phylogenomic analyses exploring the origins and conservation of genes have revealed extensive mosaicism, such as the retention of animal-like genes (9, 139), the presence of genes from green and red algae in the nuclear genomes of algae with 2°/3° plastids (52, 66, 145), and the presence of bacterial genes, such as those from the cyanobacterial progenitor of the chloroplast (53, 64, 133, 185) and others from a relative of *Chlamydia* (76, 171). Life forms in the oceans have acquired entire pathways and processes from marine bacteria through horizontal gene transfer, generating a melting pot of protein repertoires (4). Evidence is growing that points to the foreign genes retained by algae as drivers in colonizing new niches (4, 180, 194). In addition to helping build our understanding of organisms and their ecosystems, these algal adaptations offer bioengineers a reservoir of unique functional capabilities that operate or cooperate in a photosynthetic cell.

At the same time, whole-genome sequencing has confirmed how different algae are from one another. The supergroups to which algae belong, based on the evolutionary origins of the heterotrophic hosts, are estimated to have diverged within 300 million years of the last eukaryotic universal ancestor at least 1–1.9 billion years ago (73). The phylogenetic affinity between algae with primary plastids and algae with 2°/3° plastids pertains to only a subset of genes. Although estimates vary among algal genomes, horizontal and endosymbiotic gene transfer are estimated to have contributed roughly 2,000 green and red algal proteins to *Phaeodactylum tricornutum* (176). At the same time, nearly 6,000 proteins are unique to *P. tricornutum* and other stramenopiles [based on reciprocal BLAST best hits with an E value of 1×10^{-10}; however, a sensitive homology search detects conserved protein domains in 99% of proteins (176)]. The gene count from individual unique isolates of marine green algae from the genus *Micromonas* can vary as much as 10% (236). Similar diversity was found for isolates of the marine coccolithophorid *Emiliania huxleyi*, where over 5,000 genes in the reference genome were not found in one or more of three isolates (181).

3.1. The Role of Algal Genomics in Opening Doors to New and Novel Approaches in Biotechnology

Both the cultivation and the engineering of algal strains for industrial-scale bioproduct harvesting and bioprospecting for functional capabilities have benefited immensely from genomics. Whole-genome sequencing has enabled the transfer of knowledge about proteins and pathways from bacteria, fungi, plants, and animals to algae. Reference organisms, such as *C. reinhardtii* and *P. tricornutum*, have been particularly useful for experimentally characterizing algal-specific adaptations at the genetic and molecular levels. In this way, research into the use of newly isolated or newly sequenced algae as factories for bioproducts does not have to start at ground zero. Studies illuminating organism-specific traits and research with potential commercial strains build upon a core of shared knowledge derived through the common ancestry of metabolism revealed by genomics.

Genomes are also informative about what they do not contain. Alkanes and alkenes are high-value chemicals that can be derived from fatty acids. These hydrocarbons are used as liquid transportation fuel and to make plastics, but production by engineered microbes is still more expensive than extracting them from crude oil or natural gas (118). Like some cyanobacteria (190), some algae have the ability to convert C_{16} and C_{18} fatty acids into alka(e)nes, but genes encoding known hydrocarbon-forming enzymes are not found in their genomes (203). This absence motivated the recent discovery of a new fatty acid photodecarboxylase, which is an alga-specific enzyme that is catalytically activated by light. This is an extremely rare but biotechnologically desired property for the development of industrial catalysts, which has the potential to impact biotechnology far beyond fuels (202).

3.2. Adapting to Feast and Famine

In nature, access to nutrients can be variable. Acclimation to seasonal or daily changes in the environment and competition with neighbors has left its mark on the genomes of algae. There are large repertoires of proteins involved in responding to the changing availability of each essential macro- and micronutrient for terrestrial and aquatic life. Genomes have also shed light on the unique biogeochemistry of each alga's environment and the adaptations that have been selected for during competition and cooperation within these ecological niches.

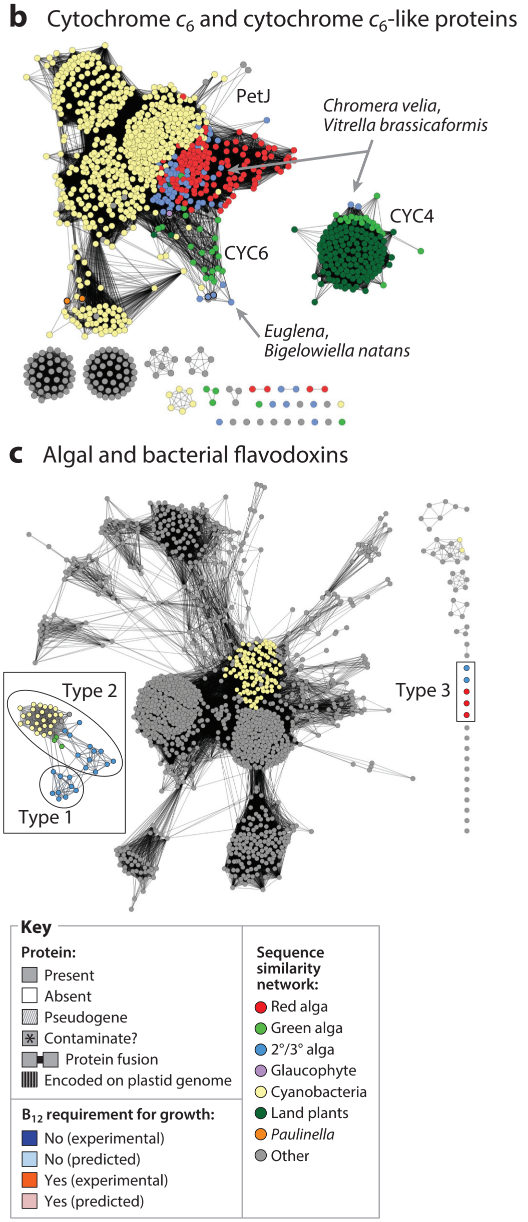
The presence of genes encoding orthologs of well-characterized transporters and assimilatory proteins can provide an initial survey of the nutrient sources that an organism can use. For instance, transporters for four nitrogen sources (nitrate, ammonium, urea, and amino acids) were found during initial analysis of the genome from the diatom *Thalassiosira pseudonana* (9). Nitrate, ammonium, and urea transporters were also found in the genome of the prasinophyte *O. tauri*. Based on a copy number comparison between the two marine algae, *O. tauri* may be more competitive for ammonium, whereas *T. pseudonana* may be more competitive for nitrate and urea (62). Competition for Fe is also evident in algal genomes, with many algae containing multiple types of Fe transporters and auxiliary components that are either unique to algae (FEA and ISIP2a) or shared with yeast (ferroxidase dependent), animals (transferrin dependent), or land plants (divalent cation transporters) (22).

All algae have an absolute requirement for metal cofactors to catalyze many of the reactions essential to life, including electron transfer during both respiration and photosynthesis. It therefore should come as no surprise that Fe, Cu, and Zn have been demonstrated to be limiting nutrients for algal growth in the environment. The abundance and bioavailability of metal ions are fundamental characteristics of each environment, and to be successful, an alga must adapt to the geochemistry and competition within each niche. Unlike the macroelements, which cannot be fully replaced, there is some flexibility in the use of specific transition metal ions. This plasticity is due to convergent evolution where two proteins with the same function have evolved independently to use different metal ions, or in some cases, no metal at all. When both isoforms are encoded in a genome, the corresponding genes can be differentially regulated depending on cofactor availability. Classic examples include the Cu-regulated switch between plastocyanin (Cu dependent) and cytochrome c_6 (Fe dependent) (138) and the Fe-regulated switch between ferredoxin (Fe dependent) and flavodoxin (flavin dependent) (74).

Comparative genomics can be used to determine if algae have the potential for these mechanisms (**Figure 5**). However, except in rare cases, we lack an understanding of the regulatory sequences that determine condition-specific gene expression in algae, and transcriptomics and/or proteomics are also required to inform the biological role of these proteins in acclimation to nutrient availability. For instance, the *Thalassiosira oceanica* genome encodes both plastocyanin and an ortholog of cytochrome c_6 (130), which suggests this diatom could be capable of a Cu-dependent switch between the two proteins. However, plastocyanin is constitutively expressed (161). The inability to dispense plastocyanin during Cu limitation may explain the growth defect of *T. oceanica* during Cu limitation (105) because *C. reinhardtii*, where the Cu-dependent switch occurs, does not display a growth defect during Cu limitation (124). Ferredoxin and flavodoxin represent slightly different examples that emphasize the importance of establishing orthological relationships for accurate functional annotation propagation. A combination of expression analysis and phylogenetics revealed that pelagophytes contain two flavodoxin genes; one is regulated by Fe but the other is not (232). Although the *T. pseudonana* genome encodes a ferredoxin and a flavodoxin, the Fe-regulated paralog has been lost (232) (**Figure 5**). A similar scenario has occurred involving

a

		CYC6	CYC4	PC	PAA1	PAA2	FD	FLD Type 1	FLD Type 2	FLD Type 3	B_{12}	METE	METH	Urease	UreF	UreD	UreG	CuSOD	Fe/MnSOD	NiSOD
	Cyanophora paradoxa	■	□	□	□	□	▥	□	□	□	🟧	□	■	□	□	□	□	■	■	□
	Chondrus crispus	▥	□	□	□	□	▥	□	□	■	🩵	■	■	□	□	□	□	■	■	□
	Cyanidioschyzon merolae	▥	□	□	□	□	▥	□	□	□	🟦	■	□	□	□	□	□	□	■	□
	Galdieria phlegrea	□	□	□	□	□	□	□	□	□	🩵	■	□	■	□	□	□	□	■	□
	Galdieria sulphuraria	▥	□	□	□	□	▥	□	□	□	🩵	■	□	□	□	□	□	□	■	□
	Porphyra umbilicalis	▥	□	□	□	□	▥	□	□	■	🩵	■	■	□	□	□	□	■	■	□
	Pyropia yezoensis	▥	□	□	□	□	▥	□	□	■	🩵	■	■	□	□	□	□	■	■	□
	Porphyridium purpureum	▥	□	□	□	□	▥	□	□	□	🟧	⊛	■	■	■	■	□	□	■	□
Algae with 1° plastids	*Bathycoccus prasinos*	□	■	■	■	■	■	□	□	□	🩷	□	■	■	■▪	■	■	■	■	■
	Micromonas sp. RCC299	□	■	■	■	■	■	□	□	□	🩷	□	■	■	■	■	■	■	■	■
	Micromonas sp. CCMP1545	□	■	■	■	■	■	□	□	□	🟧	□	■	□	□	□	□	■	■	■
	Micromonas sp. ASP10-01a	□	□	■	■	■	□	□	□	□	🩷	□	□	■	■	■	■	■	■	■
	Ostreococcus lucimarinus	□	■	■	■	■	■	□	■	□	🩷	□	■	■	■	■	■	■	■	■
	Ostreococcus tauri	□	■	■	■	■	■	□	□	□	🟧	□	■	■	■	■	■	■	■	■
	Chlorella variabilis	■	■	■	■	■	■	□	□	□	🟦	■	■	□	□	□	□	□	■	□
	Auxenochlorella protothecoides	□	□	■	□	□	■	□	□	□	🟦	■	■	□	□	□	□	□	■	□
	Chlorella sorokiniana	■	□	■	■	■	■	□	□	□	🟦	■	■	□	□	□	□	□	■	□
	Micractinium conductrix	■	■	■	■	■	■	□	□	□	🩵	■	■	□	□	□	□	□	■	□
	Chloroidium sp. UTEX 3007	□	■	■	□	□	□	□	□	□	□	□	■	□	□	□	□	□	□	□
	Chromochloris zofingiensis	■	■	■	■	■	■	□	■	□	🩵	■	■	□	□	□	□	□	■	□
	Raphidocelis subcapitata	■	■	■	□	□	■	□	□	□	🩵	■	■	□	□	□	□	□	■	□
	Coccomyxa subellipsoidea	■	■	■	■	■	■	□	■	□	🟦	■	□	□	□	□	□	□	■	□
	Helicosporidium sp. ATCC 50920	□	□	□	□	□	□	□	□	□	🩵	■	■	□	□	□	□	□	■	□
	Picochlorum SENEW3 (SE3)	■	■	■	■	■	■	□	□	□	🩵	■	■	□	□	□	□	□	■	□
	Picochlorum sp. DOE101	■	■	■	■	■	□	□	□	□	🩵	■	■	□	□	□	□	□	■	□
	Chlamydomonas reinhardtii	■	■	■	■	■	■	□	□	□	🟦	■	■	□	□	□	□	□	■	□
	Chlamydomonas eustigma	□	■	■	■	■	■	□	□	□	🩵	■	■	□	□	□	□	□	■	□
	Volvox carteri	□	□	■	■	■	■	□	□	□	🟧	□	■	□	□	□	□	□	■	□
	Dunaliella salina	■	□	■	■	■	■	□	□	□	🟦	■	■	□	□	□	□	□	■	□
	Gonium pectorale	■	■	■	■	■	■	□	□	□	🟧	▨	■	□	□	□	□	□	■	□
	Monoraphidium neglectum	■	■	■	■	■	■	□	□	□	🩵	■	■	□	□	□	□	□	■	□
	Tetrabaena socialis	□	■	■	■	■	■	□	□	□	🩵	■	■	□	□	□	□	□	■	□
	Klebsormidium nitens	□	■	■	■	■	■	□	□	□	🩵	■	■	■	■	■	■	■	■	□
	Cladosiphon okamuranus	□	□	□	□	□	□	□	□	□	🩵	■	■	■	■	■	■	■	■	□
	Ectocarpus siliculosus	▥	□	□	□	□	▥	□	□	□	🟦	■	■	■	■	■	■	■	■	□
	Aureococcus anophagefferens	■	□	□	□	□	■	□	■	■	🟧	□	■	■	■▪	■	■	■	■	■
	Nannochloropsis gaditana	▥	□	□	□	□	▥	□	□	□	🩵	■	■	■	■	■	■	□	■	□
	Nannochloropsis oceanica	▥	□	□	□	□	▥	□	□	□	🩵	■	■	■	■	□	□	□	■	□
	Fragilariopsis cylindrus	■	□	■	□	□	□	■	■	□	🟦	■	■	■	■	□	■	□	■	■
	Phaeodactylum tricornutum	■	□	□	□	□	▥	■	■	□	🟦	■	■	■	■▪	■	■	■	■	■
	Fistulifera solaris	■	□	□	□	□	▥	■	□	□	□	□	□	■	■▪	■	■	■	■	■
Algae with 2°/3° plastids	*Cyclotella cryptica*	■	□	□	□	□	□	■	□	□	🩵	■	■	■	■▪	■	■	□	■	■
	Thalassiosira oceanica	■	□	■	□	□	■	■	■	■	🩷	□	■	□	■▪	■	■	□	■	■
	Thalassiosira pseudonana	■	□	□	□	□	▥	■	□	□	🟧	□	■	■	■▪	■	■	□	■	□
	Symbiodinium minutum	■	□	□	□	□	■	□	□	□	🩷	□	■	■	■▪	■	■	□	■	■
	Symbiodinium kawagutii	□	□	□	□	□	■	□	■	□	🩷	□	■	□	□	□	□	■	■	■
	Symbiodinium microadriaticum	■	□	□	□	□	■	□	■	□	🩷	□	■	■	■▪	■	■	□	■	■
	Chromera velia	■	■	□	□	□	■	□	■	□	🩵	■	■	■	■	■	□	□	■	□
	Vitrella brassicaformis	■	■	□	□	□	■	□	■	□	🩵	■	□	■	■	■	■	□	■	□
	Bigelowiella natans	■	□	□	□	□	■	□	■	□	🩷	□	■	■	■▪	■	■	■	■	■
	Chrysochromulina tobin	□	□	□	□	□	■	□	□	□	🩷	□	■	■	□	■	□	■	□	■
	Emiliania huxleyi	■	□	■	□	□	■	□	■	□	🟦	□	■	■	□	■	□	□	■	■
	Tisochrysis lutea	■	□	□	□	□	■	□	□	□	🩷	□	■	■	■▪	■	■	■	■	■
	Guillardia theta	■	□	□	□	□	▥	□	□	□	🩷	□	■	■	□	□	■	□	■	■



(*Caption appears on following page*)

Figure 5 (*Figure appears on preceding page*)

Capitalizing on the wealth of knowledge for cofactor homeostasis and usage in algae, comparative genomics provides a method for exploring the presence of these processes across the algal lineages. (*a*) A co-occurrence plot for the presence of orthologs for selected metal-dependent enzymes and either their functional isoform or their assembly factors. These types of analyses provide a window into cofactor dependencies. For instance, axenic culture of algae revealed that some algae require the vitamin B_{12} but others do not. No eukaryote is known to be able to synthesize B_{12} de novo, and the reason for this difference remained a mystery until whole-genome sequencing revealed the presence of B_{12}-dependent and B_{12}-independent methionine synthases in the genomes of algae (51, 103). Organisms that have B_{12}-dependent methionine synthase require exogenously supplied B_{12}, whereas organisms that use a B_{12}-independent isoform do not require B_{12} in their diet. This knowledge enables the prediction of B_{12} dependency based on genome sequencing. (*b*) Homolog searches are typically not sufficient for accurately predicting protein function. CYC6 transfers electrons between cytochrome *f* and photosystem I but CYC4 does not. Accurately differentiating between these two proteins requires phylogenetics or sequence similarity networks, as shown here. (*c*) Flavodoxin can functionally substitute for ferredoxin during Fe limitation as a strategy to reduce dependency on Fe. Diatoms contain two flavodoxin proteins. The expression of only one paralog has been shown to respond to Fe nutrition, whereas the other paralog is regulated by the diel cycle. Analysis of similar proteins reveals that some green algae also have a flavodoxin that clusters with the Fe-regulated form but that flavodoxins from red algae and two additional diatom sequences (one of which is a fragment) do not share high enough sequence similarity with proteins in the network to inform function. Abbreviations: CuSOD, Cu-dependent superoxide dismutase; CYC4: cytochrome *c*-like (Fe); CYC6, cytochrome c_6 (Fe); FD, ferredoxin (Fe); Fe/MnSOD, Fe- or Mn-dependent superoxide dismutase; FLD, flavodoxin (flavin); METE, B_{12}-independent methionine synthase; METH, B_{12}-dependent methionine synthase; NiSOD, Ni-dependent superoxide dismutase; PAA, chloroplast-localized Cu transporter; PC, plastocyanin (Cu); PetJ, an ortholog of cytochrome c_6 (Fe); UreF/D/G, subunits of the urease molecular chaperone that inserts Ni.

plastocyanin and cytochrome c_6 homologs. *C. reinhardtii* contains plastocyanin, a Cu-regulated cytochrome c_6 (CYC6), and a homolog of cytochrome c_6 termed CYC4 that is not regulated by Cu. Although the function of CYC4 is still unknown (112), most green algal genomes encode orthologs of both CYC6 and CYC4; red algal genomes encode only an ortholog of CYC6 (termed PetJ and encoded on the chloroplast genome), whereas land plant genomes encode only an ortholog of CYC4 termed cytochrome c_{6A} (**Figure 5**).

3.3. Carbon

Algae are significant contributors to global and local carbon cycling and storage. As fast-growing, primary producers, algae typically form the foundation of ecosystems, although their importance in some habitats was overlooked until recently (224). On the global scale, phytoplankton, composed of algae and cyanobacteria, is estimated to contribute 46.2% of the annual global net primary production (81). However, the placement of algae within food webs is often complicated by their metabolic flexibility. In addition to phototrophy, some algae are capable of heterotrophy and mixotrophy and can assimilate reduced carbon sources, such as sugars, and ingest bacteria and eukaryotes. Duality as producer and consumer, a common strategy for acquiring nutrients in the oceans, was only recently incorporated into a global simulation of the marine food web (228). Through genome sequencing, the genetic adaptations that enable this lifestyle are starting to be explored. Comparative genomic analysis of the phago-mixotrophic green alga *C. tetramitiformis* with other phagotrophic and nonphagotrophic organisms has produced a list of nearly 400 putative proteins predicted to be specific to the phagotrophic lifestyle (38). Phagocytosis is a capability that *C. tetramitiformis* shares with the last common ancestor of the Archaeplastida and is a feeding strategy that is thought to be responsible for the capture of the cyanobacterial progenitor of the chloroplast.

Comparative genomic analyses have also been performed to gain insight into the use of organic carbon sources in nonphagotrophic algae. Based on comparative genomic analysis of the red alga *Galdieria sulphuraria*, which can use over 50 organic carbon sources, and *C. merolae*, which is an obligate photoautotroph, the presence of genes encoding proteins involved in carbohydrate metabolism was not a reliable indicator of potential carbon usage (17). Instead, the ability to use exogenous sources of organic carbon was attributed to the relatively large number of carbohydrate transporters encoded specifically in the *G. sulphuraria* genome (17). Prediction of carbohydrate usage based on analysis of transporter inventory is supported by the observation that introduction of a nonnative plasma-localized glucose transporter gene into the genome of either the diatom *P. tricornutum* or the green alga *C. reinhardtii* confers the ability to grow heterotrophically with glucose (65, 240). However, when predicting carbon usage based on the presence of genes encoding putative carbohydrate transporters, the potential localization of the proteins has to be considered. Chloroplast membranes contain a suite of transporters that function in shuttling carbohydrates (83). Without experimental evidence or robust localization predictions, some of these transporters could be mistakenly predicted to function in assimilation of carbon sources from the environment.

For assimilation of inorganic carbon, many algal genomes contain genes for carbon concentrating mechanisms (CCMs) (197). While Rubisco is responsible for irreversible carbon fixation and is the first enzyme in the Calvin-Benson pathway for generating reduced carbon from CO_2, algal genomes typically encode a suite of protein components of the CCM. These proteins function in assimilating and concentrating CO_2 at the site of carbon fixation, thus effectively increasing photosynthetic efficiency. CCMs are not unique to algae; cyanobacteria and some land plants with C_4 assimilation employ mechanisms to saturate Rubisco with CO_2, but the diversity of CCMs in algae is greater (177). Even Rubisco has been subject to evolutionary tinkering, with four phylogenetically distinct forms found in different algae resulting from combinations of endosymbiotic and horizontal gene transfer (179). In aquatic environments, the diffusion of CO_2 in still water can be up to 10,000 times slower than through air (10), and pH can have a significant effect on the ratio of CO_2 and HCO_3^- (because lipid bilayers are more permeable to CO_2 than to HCO_3^-). Algae with 2°/3° plastids also have to contend with additional membranes that act as barriers for getting inorganic carbon to Rubisco, although the membranes of diatoms appear to be more permeable to CO_2 than the membranes of some green algae (109, 208, 220). Overcoming these challenges and acclimating to changes in the environment that affect inorganic carbon concentration and speciation have resulted in the evolution of different CCMs. Experimentally characterized components of algal CCMs include active transport of bicarbonate and/or CO_2 transporters, CO_2 channels (204, 205), carbonic anhydrases (90), and proteins involved in pyrenoid biogenesis. Some evidence for C_4-like metabolism in individual algae has been presented (62, 117, 183, 184), but the prevalence or contribution of these mechanisms to CCM in algae remains controversial (75, 178).

3.4. Understanding Postendosymbiotic Innovation Through Phylogenomics and Experimentation

In addition to the genetic contribution from the proteobacterial progenitor of the mitochondrion shared by all eukaryotes, evolution of the plastid was accompanied by the transfer of genes from the cyanobacterial endosymbiont to the host. Whole-genome sequencing has revealed that due to endosymbiotic gene transfer, up to 20% of the genes in the nuclear genomes from the green lineage is estimated to have originated from the cyanobacterial endosymbiont (53, 64, 133, 185). Often this transfer was accompanied by genetic adaptions that can be traced through genomics. Eventual domestication of the endosymbiont and its transformation into an organelle involved both gene loss (125) and gene fusion (135, 136) as well as adaptations that were required after transfer of genes to the host nucleus. Examples include acquisition of localization signals and integration of host transcription and translation signals, regulatory sequences, and introns (see 29, 154). In addition to these adaptions that had to take place for expression and proper targeting of cyanobacterial proteins, some of these genes were duplicated, resulting in neofunctionalization or subfunctionalization.

3.4.1. Evolution of the plastid from the perspectives of the host and endosymbiont. Carbon metabolism is a defining aspect of algal biology. Indeed, the ability of photosynthesis to fix and reduce CO_2 was the main selective advantage behind endosymbiosis, evolution of the chloroplast, and its transfer across Eukarya by secondary and tertiary endosymbiosis. Phylogenetic analysis of envelope-localized transporters suggests that a majority of transporters, particularly carbohydrate transporters, are of host origin (222). However, the relationship between the host and the endosymbiont was not one sided, and the requirement to sustain the endosymbiont within the host cytosol would have served as a driving force for adaptation and fixation of genes acquired by endosymbiotic gene transfer.

Metal ions, in particular, would have been a challenge. The reactivity of Fe, Cu, Mn, and Zn has made these metals useful in biology, but their very reactivities render them toxic in excess, especially in the presence of oxygen (generated by the newly-acquired symbiont) where Fe and Cu can generate reactive oxygen species that are deleterious to biological macromolecules. Photosynthesis has an absolute requirement for Fe and Cu (within plastocyanin-containing algae) in electron transfer and for Mn in the water-splitting reaction of photosystem II. Endosymbiosis must have presented a challenge to both the host and the endosymbiont. If a nutrient was limiting, induction of high-affinity uptake by the endosymbiont, as occurs in extant cyanobacteria, could have starved the host. At the same time, because metal transport and trafficking are highly controlled processes in eukaryotes, without regulated provision, the endosymbiont could itself be starved of metal ions.

3.4.2. Transition from a free-living organism to an organelle: adaptations involving transport capabilities. Distributive transporters critical for metalloprotein biogenesis have been retained during evolution of the chloroplast, but many of the high-affinity metal transporters found in extant cyanobacteria are not present in the genomes of land plants and algae. The transport of Cu and Mn serves to illustrate this point and provides an example of the synergy between genomics and experimentation in understanding functional implications involving chloroplast evolution.

In the cyanobacterium *Synechocystis* sp. PCC 6803, two P_{1B}-type Cu^+-ATPases function collaboratively to provide plastocyanin with the Cu required for its activity (215): PacS, which is a typical Cu-detoxification exporter with a high efflux rate, and CtaA, which has a lower efflux rate typical of other Cu^+-ATPases involved in metalloprotein biogenesis (175). Orthologs of CtaA but not of PacS are found in the genomes of green algae and land plants (98). In addition to green algae and land plants, several diatom genomes and a haptophyte genome encode a plastocyanin homolog, but like red algae, which are the modern relatives of the engulfed alga that became their plastid, these algae are also missing orthologs of CtaA and PacS. This suggests that a different pathway exists to metallate plastocyanin (23).

Photosystem II is dependent on Mn for the water-splitting reaction during photosynthesis. The metal transporter Mnx/SynPAM71, which is a member of the UPF0016 family, functions in transporting Mn for biogenesis and possible reassembly of the Mn-cluster in *Synechocystis* sp. PCC 6803 (32, 88). A functional homolog of Mnx is present in land plants and green algae, but its evolutionary origin is not clear, with the chloroplast transporters branching before homologs from cyanobacteria, fungi, and metazoans (57, 191). This phylogeny was previously interpreted as a host origin of the chloroplast transporter (222). Given the propensity within this family for domain duplication and fusion and conservation in the genomes of algae outside of Archaeplastida, further analysis is needed. However, as observed for Cu transport, the ABC-type high-affinity Mn transporter present in extant cyanobacteria is not found in algae or land plants.

3.4.3. Duplication and neofunctionalization. Another adaptation necessary for chloroplast Cu and Mn transport relates to the localization of the target metalloproteins. Transporters in cyanobacteria are made on cytoplasmic ribosomes and inserted into the plasma membrane or thylakoid membrane from the cytoplasmic side. In algae and land plants, chloroplast transporters are synthesized outside the organelle on cytoplasmic ribosomes, then transported as unfolded proteins either to the envelope or through the envelope to the thylakoid membrane. These structural differences appear to have necessitated duplication of both the Mnx functional homolog and the CtaA endo-ortholog followed by neofunctionalization. Localization studies and phenotypes of the corresponding mutants in *A. thaliana* suggest that each pair of transporters acts in tandem. One paralog [CMT1 (Mn) (70, 242) and PAA1 (Cu) (198)] is targeted to the envelope membrane for transport of metal ions from the host cytosol into the stroma, and the other paralog [PAM71 (Mn) (191) and PAA2 (Cu) (1)] is targeted to the thylakoid membrane for transport of metal ions from the stroma into the thylakoid lumen.

In the case of chloroplast Cu transport, additional adaptations have been suggested. The chloroplast Cu transporters, like CtaA, are unidirectional ATPases. Therefore, maintaining the topology of the cyanobacterial ancestor would result in transport of Cu from the stroma into the intermembrane space by the envelope-targeted Cu-ATPase PAA1. This topology is at odds with genetic evidence supporting the function of PAA1 as a chloroplast importer (198). Based on topology experiments with purified envelope vesicles, PAA1 does appear to be situated in the envelope membrane with the amino terminus facing the intermembrane space (24), which would enable ATP-driven transport of Cu from the intermembrane space into the stroma. While this transporter is flipped relative to the orientation of the ortholog in extant cyanobacteria, the experimentally determined topology would be consistent with the direction of transport of homologous P_{1B}-type Cu^{+}-ATPases in the Golgi and vacuole of eukaryotes. This result has yet to be confirmed in vivo, but by surveying sequenced genomes, it becomes apparent that *PAA1* orthologs uniquely encode a conserved glycine-stretch next to the transit peptide that may function in the localization and topology of PAA1 in the inner envelope membrane.

Eukaryotic Cu homeostasis involves routing pathways composed of Cu chaperones and Cu^{+}-ATPases. The evolution of a Cu chaperone is a third adaptation, which involves the endo-ortholog of *ctaA*. In *A. thaliana* and *C. reinhardtii*, the Cu chaperone and PAA1 are expressed from the same gene through an alternative splicing event. Comparative genomic analysis revealed that duplication followed by subfunctionalization occurred independently in different land plant lineages and resulted in the Cu chaperone and transporter being encoded by separate genes (24). This snapshot suggests that alternative splicing can serve as an intermediary state prior to gene duplication in the evolution of new functions involving genes derived from endosymbiotic gene transfer. As more high-quality algal genomes and associated transcriptome resources become available, it will be exciting to see the extent to which such evolutionary mechanisms have had a functional impact on algae across the various lineages. For instance, transcript sequencing in *P. tricornutum* (176) and the chlorarachniophyte *Bigelowiella natans* (52) suggests abundant alternative splice forms in these algae.

SUMMARY POINTS

1. To understand the genetic underpinnings of algal biology and achieve systems and synthetic biology objectives, genome sequences from this polyphyletic group are essential.
2. The large number of proteins of unknown function encoded on algal genomes indicates that there is much to be discovered.
3. Comparing genomes and analyzing functional genomics data are needed to contextualize and predict protein function, but researchers should be aware of the quality of published

genome assemblies and associated gene model predictions. With some exceptions, most genome assemblies are incomplete and, although improving, many structural annotations are inaccurate.

4. Laboratory reference organisms, such as *Chlamydomonas reinhardtii* and *Phaeodactylum tricornutum*, are essential for providing answers to individual investigator-initiated questions, which can be propagated with due diligence to other algae through genome-based evolutionary relationships.
5. Algal genomes are a melting pot of unique functional capabilities encoded by genes of disparate evolutionary origin.
6. We have a limited understanding of how inorganic nutrients are supplied and transported across the three or four chloroplast membranes in algae outside of Archaeplastida, but largely due to advanced genomic and genetic resources, we are beginning to piece together the evolutionary history and functional implications of chloroplast metal transport within the green lineage.

FUTURE ISSUES

1. Algal diversity, with respect to both evolutionary history and ecological niche, is expansive and provides fertile ground for discovery.
2. In addition to enabling a genome-based understanding of algal biology, algal genome sequences offer a reservoir of unique functional capabilities that can be employed for the design of new capabilities in crops and beyond.
3. In addition to sequencing more algal genomes that better represent the diversity of algal biology, high-quality genome sequences and high-quality structural annotations are needed to facilitate protein function prediction and contextualization of functional annotations.
4. Whole-genome sequencing and comparative genomics, together with the application of CRISPR-Cas systems and other genome-engineering technologies, will enable a broader range of organisms to ascend to the level of a reference and ultimately expand our knowledge of diverse algal biology.
5. The collection and collation of large functional genomics data sets, such as from proteomics, transcriptomics, and mutants and their phenotypes, will give rise to functional inferences and ultimately generate evidence-based annotations in reference algae to serve as resources for effective genome curation.
6. While sequencing-based genome-wide experiments, such as transcriptomics, are providing valuable insight into the adaptation and acclimation of algae to their environment, improved methods for metabolite profiling (and metabolite discovery) are needed to elucidate the metabolic capabilities of diverse groups of algae and link genes to function.

DISCLOSURE STATEMENT

S.S.M. is a consultant for ExxonMobil Research and Engineering.

ACKNOWLEDGMENTS

This work was supported by the US Department of Energy Office of Science Biological and Environmental Research program awards DE-SC0018301 and DE-FC02-02ER63421. We are especially grateful to Dr. Ian K. Blaby for critical feedback and Dr. Fatima Foflonker for calculating Benchmarking Universal Single-Copy Ortholog (BUSCO) conservation.

LITERATURE CITED

1. Abdel-Ghany SE, Müller-Moulé P, Niyogi KK, Pilon M, Shikanai T. 2005. Two P-type ATPases are required for copper delivery in *Arabidopsis thaliana* chloroplasts. *Plant Cell* 17:1233–51
2. Abreu IA, Cabelli DE. 2010. Superoxide dismutases—a review of the metal-associated mechanistic variations. *Biochim. Biophys. Acta Proteins Proteom.* 1804:263–74
3. Ajjawi I, Verruto J, Aqui M, Soriaga LB, Coppersmith J, et al. 2017. Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. *Nat. Biotechnol.* 35:647–52
4. Allen AE, Dupont CL, Obornik M, Horak A, Nunes-Nesi A, et al. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473:203–7
5. Arakaki Y, Kawai-Toyooka H, Hamamura Y, Higashiyama T, Noga A, et al. 2013. The simplest integrated multicellular organism unveiled. *PLOS ONE* 8:e81641
6. Aranda M, Li Y, Liew YJ, Baumgarten S, Simakov O, et al. 2016. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Sci. Rep.* 6:39734
7. Aravind L. 2000. Guilt by association: contextual information in genome analysis. *Genome Res.* 10:1074–77
8. Arias-Darraz L, Cabezas D, Colenso CK, Alegría-Arcos M, Bravo-Moraga F, et al. 2015. A transient receptor potential ion channel in *Chlamydomonas* shares key features with sensory transduction-associated TRP channels in mammals. *Plant Cell* 27:177–88
9. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306:79–86
10. Armstrong W. 1980. Aeration in higher plants. In *Advances in Botanical Research*, Vol. 7, ed. HW Woolhouse, pp. 225–332. New York: Academic
11. Arriola MB, Velmurugan N, Zhang Y, Plunkett MH, Hondzo H, Barney BM. 2018. Genome sequences of *Chlorella sorokiniana* UTEX 1602 and *Micractinium conductrix* SAG 241.80: implications to maltose excretion by a green alga. *Plant J.* 93:566–86
12. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25:25–29
13. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. 2009. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLOS ONE* 4:e4345
14. Atteia A, Adrait A, Brugière S, Tardif M, van Lis R, et al. 2009. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the α-proteobacterial mitochondrial ancestor. *Mol. Biol. Evol.* 26:1533–48
15. Baldauf SL. 2003. The deep roots of eukaryotes. *Science* 300:1703–6
16. Baldauf SL. 2008. An overview of the phylogeny and diversity of eukaryotes. *J. Syst. Evol.* 46:263–73
17. Barbier G, Oesterhelt C, Larson MD, Halgren RG, Wilkerson C, et al. 2005. Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. *Plant Physiol.* 137:460–74
18. Bhattacharya D, Price DC, Chan CX, Qiu H, Rose N, et al. 2013. Genome of the red alga *Porphyridium purpureum*. *Nat. Commun.* 4:1941
19. Bienvenut WV, Espagne C, Martinez A, Majeran W, Valot B, et al. 2011. Dynamics of post-translational modifications and protein stability in the stroma of *Chlamydomonas reinhardtii* chloroplasts. *Proteomics* 11:1734–50

20. Blaby IK, Blaby-Haas CE, Tourasse N, Hom EF, Lopez D, et al. 2014. The *Chlamydomonas* genome project: a decade on. *Trends Plant Sci*. 19:672–80

21. Blaby-Haas CE, de Crécy-Lagard V. 2011. Mining high-throughput experimental data to link gene and function. *Trends Biotechnol*. 29:174–82

22. Blaby-Haas CE, Merchant SS. 2012. The ins and outs of algal metal transport. *Biochim. Biophys. Acta Mol. Cell Res.* 1823:1531–52

23. Blaby-Haas CE, Merchant SS. 2017. Regulating cellular trace metal economy in algae. *Curr. Opin. Plant Biol.* 39:88–96

24. Blaby-Haas CE, Padilla-Benavides T, Stübe R, Argüello JM, Merchant SS. 2014. Evolution of a plant-specific copper chaperone family for chloroplast copper homeostasis. *PNAS* 111:E5480–87

25. Blanc G, Agarkova I, Grimwood J, Kuo A, Brueggeman A, et al. 2012. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol*. 13:R39

26. Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, et al. 2010. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* 22:2943–55

27. Blanc-Mathieu R, Krasovec M, Hebrard M, Yau S, Desgranges E, et al. 2017. Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Sci. Adv.* 3:e1700239

28. Blanc-Mathieu R, Verhelst B, Derelle E, Rombauts S, Bouget FY, et al. 2014. An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina *de novo* assemblies. *BMC Genom*. 15:1103

29. Bock R. 2017. Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. *Annu. Rev. Genet.* 51:1–22

30. Bogen C, Al-Dilaimi A, Albersmeier A, Wichmann J, Grundmann M, et al. 2013. Reconstruction of the lipid metabolism for the microalga *Monoraphidium neglectum* from its genome sequence reveals characteristics suitable for biofuel production. *BMC Genom*. 14:926

31. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239–44

32. Brandenburg F, Schoffman H, Kurz S, Krämer U, Keren N, et al. 2017. The *Synechocystis* MANGANESE EXPORTER Mnx is essential for manganese homeostasis in cyanobacteria. *Plant Physiol*. 173:1798–810

33. Brawley SH, Blouin NA, Ficko-Blean E, Wheeler GL, Lohr M, et al. 2017. Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *PNAS* 114:E6361–70

34. Breker M, Lieberman K, Cross FR. 2018. Comprehensive discovery of cell-cycle-essential pathways in *Chlamydomonas reinhardtii. Plant Cell* 30:1178–98

35. Brock TD. 1975. Salinity and the ecology of *Dunaliella* from Great Salt Lake. *Microbiology* 89:285–92

36. Burki F. 2017. The convoluted evolution of eukaryotes with complex plastids. In *Advances in Botanical Research*, Vol. 84, ed. Y Hirakawa, pp. 1–30. New York: Academic

37. Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, et al. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B* 283:20152802

38. Burns JA, Paasch A, Narechania A, Kim E. 2015. Comparative genomics of a bacterivorous green alga reveals evolutionary causalities and consequences of phago-mixotrophic mode of nutrition. *Genome Biol. Evol.* 7:3047–61

39. Capasso C, Supuran CT. 2015. An overview of the alpha-, beta- and gamma-carbonic anhydrases from *Bacteria*: Can bacterial carbonic anhydrases shed new light on evolution of bacteria? *J. Enzym. Inhib. Med. Chem.* 30:325–32

40. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, et al. 2009. AmiGO: online access to ontology and annotation data. *Bioinformatics* 25:288–89

41. Carrier G, Baroukh C, Rouxel C, Duboscq-Bidot L, Schreiber N, Bougaran G. 2018. Draft genomes and phenotypic characterization of *Tisochrysis lutea* strains. Toward the production of domesticated strains with high added value. *Algal Res*. 29:1–11

42. Casabianca S, Cornetti L, Capellacci S, Vernesi C, Penna A. 2017. Genome complexity of harmful microalgae. *Harmful Algae* 63:7–12
43. Castenholz RW, McDermott TR. 2010. The Cyanidiales: ecology, biodiversity, and biogeography. In *Red Algae in the Genomic Age*, ed. J Seckbach, DJ Chapman, pp. 357–71. Dordrecht: Springer
44. Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* 46:347–66
45. Cheng S, Melkonian M, Smith SA, Brockington S, Archibald JM, et al. 2018. 10KP: a phylodiverse genome sequencing plan. *GigaScience* 7:giy013
46. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* 465:617–21
47. Collén J, Porcel B, Carré W, Ball SG, Chaparro C, et al. 2013. Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *PNAS* 110:5247–52
48. Cormier A, Avia K, Sterck L, Derrien T, Wucher V, et al. 2017. Re-annotation, improved large-scale assembly and establishment of a catalogue of noncoding loci for the genome of the model brown alga *Ectocarpus*. *New Phytol.* 214:219–32
49. Corteggiani Carpinelli E, Telatin A, Vitulo N, Forcato C, D'Angelo M, et al. 2014. Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. *Mol. Plant* 7:323–35
50. Courties C, Vaquer A, Troussellier M, Lautier J, Chrétiennot-Dinet MJ, et al. 1994. Smallest eukaryotic organism. *Nature* 370:255
51. Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. 2005. Algae acquire vitamin B_{12} through a symbiotic relationship with bacteria. *Nature* 438:90–93
52. Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492:59–65
53. Dagan T, Roettger M, Stucken K, Landan G, Koch R, et al. 2012. Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biol. Evol.* 5:31–44
54. De la Rosa MA, Molina-Heredia FP, Hervás M, Navarro JA. 2006. Convergent evolution of cytochrome c_6 and plastocyanin. In *Photosystem I: The Light-Driven Plastocyanin: Ferredoxin Oxidoreductase*, ed. JH Golbeck, pp. 683–96. Advances in Photosynthesis and Respiration Ser. 24. Dordrecht, Neth.: Springer
55. Delaye L, Valadez-Cano C, Pérez-Zamorano B. 2016. How really ancient is *Paulinella* chromatophora? *PLOS Curr. Tree Life* 8:e68a099364bb1a1e129a17b4e06b0c6b
56. Delmont TO, Eren AM, Vineis JH, Post AF. 2015. Genome reconstructions indicate the partitioning of ecological functions inside a phytoplankton bloom in the Amundsen Sea, Antarctica. *Front. Microbiol.* 6:1090
57. Demaegd D, Colinet A-S, Deschamps A, Morsomme P. 2014. Molecular evolution of a novel family of putative calcium transporters. *PLOS ONE* 9:e100851
58. Deng XD, Gu B, Li YJ, Hu XW, Guo JC, Fei XW. 2012. The roles of acyl-CoA: diacylglycerol acyltransferase 2 genes in the biosynthesis of triacylglycerols by the green algae *Chlamydomonas reinhardtii*. *Mol. Plant* 5:945–47
59. Dent RM, Haglund CM, Chin BL, Kobayashi MC, Niyogi KK. 2005. Functional genomics of eukaryotic photosynthesis using insertional mutagenesis of *Chlamydomonas reinhardtii*. *Plant Physiol.* 137:545–56
60. Dent RM, Sharifi MN, Malnoë A, Haglund C, Calderon RH, et al. 2015. Large-scale insertional mutagenesis of *Chlamydomonas* supports phylogenomic functional prediction of photosynthetic genes and analysis of classical acetate-requiring mutants. *Plant J.* 82:337–51
61. DePriest MS, Bhattacharya D, Lopez-Bautista JM. 2013. The plastid genome of the red macroalga *Grateloupia taiwanensis* (Halymeniaceae). *PLOS ONE* 8:e68246
62. Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, et al. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *PNAS* 103:11647–52
63. Derelle R, López-García P, Timpano H, Moreira D. 2016. A phylogenomic framework to study the diversity and evolution of stramenopiles (=heterokonts). *Mol. Biol. Evol.* 33:2890–98

64. Deusch O, Landan G, Roettger M, Gruenheit N, Kowallik KV, et al. 2008. Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol. Biol. Evol.* 25:748–61

65. Doebbe A, Rupprecht J, Beckmann J, Mussgnug JH, Hallmann A, et al. 2007. Functional integration of the *HUP1* hexose symporter gene into the genome of *C. reinhardtii*: impacts on biological H_2 production. *J. Biotechnol.* 131:27–33

66. Dorrell RG, Gile G, McCallum G, Meheust R, Bapteste EP, et al. 2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife* 6:e23717

67. Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *PNAS* 101:15386–91

68. Eisen JA. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* 8:163–67

69. Eisen JA, Kaiser D, Myers RM. 1997. Gastrogenomic delights: a movable feast. *Nat. Med.* 3:1076–78

70. Eisenhut M, Hoecker N, Schmidt SB, Basgaran RM, Flachbart S, et al. 2018. The plastid envelope CHLOROPLAST MANGANESE TRANSPORTER1 is essential for manganese homeostasis in *Arabidopsis*. *Mol. Plant* 11:955–69

71. Eitzinger N, Wagner V, Weisheit W, Geimer S, Boness D, et al. 2015. Proteomic analysis of a fraction with intact eyespots of *Chlamydomonas reinhardtii* and assignment of protein methylation. *Front. Plant Sci.* 6:1085

72. Ekman D, Björklund ÅK, Frey-Skött J, Elofsson A. 2005. Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *J. Mol. Biol.* 348:231–43

73. Eme L, Sharpe SC, Brown MW, Roger AJ. 2014. On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harb. Perspect. Biol.* 6:a016139

74. Erdner DL, Price NM, Doucette GJ, Peleato ML, Anderson DM. 1999. Characterization of ferredoxin and flavodoxin as markers of iron limitation in marine phytoplankton. *Mar. Ecol. Prog. Ser.* 184:43–53

75. Ewe D, Tachibana M, Kikutani S, Gruber A, Río Bártulos C, et al. 2018. The intracellular distribution of inorganic carbon fixing enzymes does not support the presence of a C_4 pathway in the diatom *Phaeodactylum tricornutum*. *Photosynth. Res.* 137:263–80

76. Facchinelli F, Colleoni C, Ball SG, Weber APM. 2013. Chlamydia, cyanobiont, or host: Who was on top in the ménage à trois? *Trends Plant Sci.* 18:673–79

77. Fan JH, Ning K, Zeng XW, Luo YC, Wang DM, et al. 2015. Genomic foundation of starch-to-lipid switch in oleaginous *Chlorella* spp. *Plant Physiol.* 169:2444–61

78. Featherston J, Arakaki Y, Hanschen ER, Ferris PJ, Michod RE, et al. 2017. The 4-celled *Tetrabaena socialis* nuclear genome reveals the essential components for genetic control of cell number at the origin of multicellularity in the volvocine lineage. *Mol. Biol. Evol.* 35:855–70

79. Feiz L, Williams-Carrier R, Belcher S, Montano M, Barkan A, Stern DB. 2014. A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *Plant J.* 80:862–69

80. Ficko-Blean E, Hervé C, Michel G. 2015. Sweet and sour sugars from the sea: the biosynthesis and remodeling of sulfated cell wall polysaccharides from marine macroalgae. *Perspect. Phycol.* 2:51–64

81. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281:237–40

82. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, et al. 2015. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44:D279–85

83. Fischer K, Weber APM, Kunz H-H. 2016. The transporters of plastids—new insights into an old field. In *Chloroplasts: Current Research and Future Trends*, ed. H Kirchhoff, pp. 209–40. Norfolk, UK: Caister Academic

84. Foflonker F, Price DC, Qiu H, Palenik B, Wang S, Bhattacharya D. 2015. Genome of the halotolerant green alga *Picochlorum* sp. reveals strategies for thriving under fluctuating environmental conditions. *Environ. Microbiol.* 17:412–26

85. Fu G, Nagasato C, Oka S, Cock JM, Motomura T. 2014. Proteomics analysis of heterogeneous flagella in brown algae (stramenopiles). *Protist* 165:662–75

86. Gabaldón T, Koonin EV. 2013. Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* 14:360–66

87. Gagat P, Bodył A, Mackiewicz P, Stiller JW. 2014. Tertiary plastid endosymbioses in dinoflagellates. In *Endosymbiosis*, ed. W Löffelhardt, pp. 233–90. Vienna: Springer

88. Gandini C, Schmidt SB, Husted S, Schneider A, Leister D. 2017. The transporter SynPAM71 is located in the plasma membrane and thylakoids, and mediates manganese tolerance in *Synechocystis* PCC 6803. *New Phytol.* 215:256–68

89. Gao C, Wang Y, Shen Y, Yan D, He X, et al. 2014. Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. *BMC Genom.* 15:582

90. Gee CW, Niyogi KK. 2017. The carbonic anhydrase CAH1 is an essential component of the carbon-concentrating mechanism in *Nannochloropsis oceanica*. *PNAS* 114:4537–42

91. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, et al. 2015. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta Proteins Proteom.* 1854:1019–37

92. Gibson TM, Shih PM, Cumming VM, Fischer WW, Crockford PW, et al. 2017. Precise age of *Bangiomorpha pubescens* dates the origin of eukaryotic photosynthesis. *Geology* 46:135–38

93. Gimmler H, Weis U. 1992. Dunaliella acidophila—life at pH 1.0. In *Dunaliella: Physiology, Biochemistry, and Biotechnology*, ed. M Avron, A Ben-Amotz, pp. 99–133. Boca Raton, FL: CRC

94. Gobler CJ, Berry DL, Dyhrman ST, Wilhelm SW, Salamov A, et al. 2011. Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *PNAS* 108:4352–57

95. Gong YM, Guo XJ, Wan X, Liang Z, Jiang M. 2011. Characterization of a novel thioesterase (PtTE) from *Phaeodactylum tricornutum*. *J. Basic Microbiol.* 51:666–72

96. Gonzalez-Esquer CR, Twary SN, Hovde BT, Starkenburg SR. 2018. Nuclear, chloroplast, and mitochondrial genome sequences of the prospective microalgal biofuel strain *Picochlorum soloecismus*. *Genome Announc.* 6:e01498–17

97. Gould SB, Waller RF, McFadden GI. 2008. Plastid evolution. *Annu. Rev. Plant Biol.* 59:491–517

98. Hanikenne M, Baurain D. 2013. Origin and evolution of metal P-type ATPases in Plantae (Archaeplastida). *Front. Plant Sci.* 4:544

99. Hanschen ER, Marriage TN, Ferris PJ, Hamaji T, Toyoda A, et al. 2016. The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat. Commun.* 7:11370

100. Hansen BO, Vaid N, Musialak-Lange M, Janowski M, Mutwil M. 2014. Elucidating gene function and function evolution through comparison of co-expression networks of plants. *Front. Plant Sci.* 5:394

101. Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB. 2001. Molecular evidence for the early colonization of land by fungi and plants. *Science* 293:1129–33

102. Hedges SB, Blair JE, Venturi ML, Shoe JL. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* 4:2

103. Helliwell KE, Wheeler GL, Leptos KC, Goldstein RE, Smith AG. 2011. Insights into the evolution of vitamin B_{12} auxotrophy from sequenced algal genomes. *Mol. Biol. Evol.* 28:2921–33

104. Herron MD, Hackett JD, Aylward FO, Michod RE. 2009. Triassic origin and early radiation of multicellular volvocine algae. *PNAS* 106:3254–58

105. Hippmann AA, Schuback N, Moon K-M, McCrow JP, Allen AE, et al. 2017. Contrasting effects of copper limitation on the photosynthetic apparatus in two strains of the open ocean diatom *Thalassiosira oceanica*. *PLOS ONE* 12:e0181753

106. Hirooka S, Hirose Y, Kanesaki Y, Higuchi S, Fujiwara T, et al. 2017. Acidophilic green algal genome provides insights into adaptation to an acidic environment. *PNAS* 114:E8304–13

107. Hoham RW, Ling H. 2000. Snow algae: the effects of chemical and physical factors on their life cycles and populations. In *Journey to Diverse Microbial Worlds: Adaptation to Exotic Environments*, ed. J Seckbach, pp. 131–45. Dordrecht, Neth.: Springer

108. Hopkins JF, Spencer DF, Laboissiere S, Neilson JAD, Eveleigh RJM, et al. 2012. Proteomics reveals plastid- and periplastid-targeted proteins in the chlorarachniophyte alga *Bigelowiella natans*. *Genome Biol. Evol.* 4:1391–406

109. Hopkinson BM, Dupont CL, Allen AE, Morel FMM. 2011. Efficiency of the CO_2-concentrating mechanism of diatoms. *PNAS* 108:3830–37

110. Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, et al. 2014. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5:3978

111. Hovde BT, Deodato CR, Hunsperger HM, Ryken SA, Yost W, et al. 2015. Genome sequence and transcriptome analyses of *Chrysochromulina tobin*: metabolic tools for enhanced algal fitness in the prominent order Prymnesiales (Haptophyceae). *PLOS Genet.* 11:e1005469

112. Howe CJ, Schlarb-Ridley BG, Wastl J, Purton S, Bendall DS. 2006. The novel cytochrome c_6 of chloroplasts: a case of evolutionary *bricolage*? *J. Exp. Bot.* 57:13–22

113. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34:2115–22

114. Ishida K, Green BR. 2002. Second- and third-hand chloroplasts in dinoflagellates: Phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *PNAS* 99:9294–99

115. Jackson C, Knoll AH, Chan CX, Verbruggen H. 2018. Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Sci. Rep.* 8:1523

116. Johnson MD. 2011. The acquisition of phototrophy: adaptive strategies of hosting endosymbionts and organelles. *Photosynth. Res.* 107:117–32

117. Johnston AM. 1991. The acquisition of inorganic carbon by marine macroalgae. *Can. J. Bot.* 69:1123–32

118. Kang MK, Nielsen J. 2017. Biobased production of alkanes and alkenes through metabolic engineering of microorganisms. *J. Ind. Microbiol. Biotechnol.* 44:613–22

119. Karakashian SJ. 1963. Growth of *Paramecium bursaria* as influenced by the presence of algal symbionts. *Physiol. Zool.* 36:52–68

120. Karkar S, Facchinelli F, Price DC, Weber APM, Bhattacharya D. 2015. Metabolic connectivity as a driver of host and endosymbiont integration. *PNAS* 112:10208–15

121. Karpowicz SJ, Prochnik SE, Grossman AR, Merchant SS. 2011. The GreenCut2 resource, a phylogenomically derived inventory of proteins specific to the plant lineage. *J. Biol. Chem.* 286:21427–39

122. Keeling PJ. 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant. Biol.* 64:583–607

123. Krabberød AK, Orr RJS, Bråte J, Kristensen T, Bjørklund KR, Shalchian-Tabrizi K. 2017. Single cell transcriptomics, mega-phylogeny, and the genetic basis of morphological innovations in Rhizaria. *Mol. Biol. Evol.* 34:1557–73

124. Kropat J, Gallaher SD, Urzica EI, Nakamoto SS, Strenkert D, et al. 2015. Copper economy in *Chlamydomonas*: prioritized allocation and reallocation of copper to respiration versus photosynthesis. *PNAS* 112:2644–51

125. Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, et al. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524:427–32

126. Kyrpides NC, Ouzounis CA. 1999. Whole-genome sequence annotation: 'Going wrong with confidence.' *Mol. Microbiol.* 32:886–87

127. Li JB, Gerdes JM, Haycraft CJ, Fan Y, Teslovich TM, et al. 2004. Comparative genomics identifies a flagellar and basal body proteome that includes the *BBS5* human disease gene. *Cell* 117:541–52

128. Lin S, Cheng S, Song B, Zhong X, Lin X, et al. 2015. The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* 350:691–94

129. Lojek LJ, Farrand AJ, Wisecaver JH, Blaby-Haas CE, Michel BW, et al. 2017. *Chlamydomonas reinhardtii* LFO1 is an IsdG family heme oxygenase. *mSphere* 2:e00176-17

130. Lommer M, Specht M, Roy AS, Kraemer L, Andreson R, et al. 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol.* 13:R66

131. Ma YH, Wang X, Niu YF, Yang ZK, Zhang MH, et al. 2014. Antisense knockdown of pyruvate dehydrogenase kinase promotes the neutral lipid accumulation in the diatom *Phaeodactylum tricornutum*. *Microb. Cell Fact.* 13:100

132. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. 1999. Detecting protein function and protein-protein interactions from genome sequences. *Science* 285:751–53

133. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, et al. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *PNAS* 99:12246–51

134. Matsuzaki M, Misumi O, Shin-i T, Maruyama S, Takahara M, et al. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428:653–57

135. Méheust R, Bhattacharya D, Pathmanathan JS, McInerney JO, Lopez P, Bapteste E. 2018. Formation of chimeric genes with essential functions at the origin of eukaryotes. *BMC Biol.* 16:30

136. Méheust R, Zelzion E, Bhattacharya D, Lopez P, Bapteste E. 2016. Protein networks identify novel symbiogenetic genes resulting from plastid endosymbiosis. *PNAS* 113:3579–84

137. Meinesz A. 1999. *Killer Algae*, transl. D Simberloff. Chicago: Univ. Chicago Press

138. Merchant SS, Bogorad L. 1986. Regulation by copper of the expression of plastocyanin and cytochrome c_{552} in *Chlamydomonas reinhardi*. *Mol. Cell. Biol.* 6:462–69

139. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318:245–50

140. Mock T, Otillar RP, Strauss J, McMullan M, Paajanen P, et al. 2017. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* 541:536–40

141. Moore CE, Archibald JM. 2009. Nucleomorph genomes. *Annu. Rev. Genet.* 43:251–64

142. Moreau H, Verhelst B, Couloux A, Derelle E, Rombauts S, et al. 2012. Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol.* 13:R74

143. Morgan RM, Ivanov AG, Priscu JC, Maxwell DP, Huner NPA. 1998. Structure and composition of the photochemical apparatus of the Antarctic green alga, *Chlamydomonas subcaudata*. *Photosynth Res.* 56:303–14

144. Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, et al. 2018. The timescale of early land plant evolution. *PNAS* 115:E2274–83

145. Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324:1724–26

146. Nakamura Y, Sasaki N, Kobayashi M, Ojima N, Yasuike M, et al. 2013. The first symbiont-free genome sequence of marine red alga, Susabi-nori (*Pyropia yezoensis*). *PLOS ONE* 8:e57122

147. Naponelli V, Noiriel A, Ziemak MJ, Beverley SM, Lye LF, et al. 2008. Phylogenomic and functional analysis of pterin-4a-carbinolamine dehydratase family (COG2154) proteins in plants and microorganisms. *Plant Physiol.* 146:1515–27

148. Nelson DR, Khraiwesh B, Fu W, Alseekh S, Jaiswal A, et al. 2017. The genome and phenome of the green alga *Chloroidium sp.* UTEX 3007 reveal adaptive traits for desert acclimatization. *eLife* 6:e25783

149. Nguyen HM, Baudet M, Cuiné S, Adriano JM, Barthe D, et al. 2011. Proteomic profiling of oil bodies isolated from the unicellular green microalga *Chlamydomonas reinhardtii*: with focus on proteins involved in lipid metabolism. *Proteomics* 11:4266–73

150. Nishitsuji K, Arimoto A, Iwai K, Sudo Y, Hisata K, et al. 2016. A draft genome of the brown alga, *Cladosiphon okamuranus*, S-strain: a platform for future studies of 'mozuku' biology. *DNA Res.* 23:561–70

151. Niu YF, Zhang MH, Li DW, Yang WD, Liu JS, et al. 2013. Improvement of neutral lipid and polyunsaturated fatty acid biosynthesis by overexpressing a type 2 diacylglycerol acyltransferase in marine diatom *Phaeodactylum tricornutum*. *Mar. Drugs* 11:4558–69

152. North WJ. 1971. Growth of individual fronds. In *The Biology of Giant Kelp Beds (Macrocystis) in California*, ed. WJ North, pp. 123–68. Beihefte zur Nova Hedwigia, Heft 32. Lehre, Ger.: J. Cramer

153. Nowack ECM, Melkonian M, Glöckner G. 2008. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr. Biol.* 18:410–18

154. Nowack ECM, Weber APM. 2018. Genomics-informed insights into endosymbiotic organelle evolution in photosynthetic eukaryotes. *Annu. Rev. Plant. Biol.* 69:51–84

155. Nozaki H, Takano H, Misumi O, Terasawa K, Matsuzaki M, et al. 2007. A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. *BMC Biol.* 5:28

156. Osterman A, Overbeek R. 2003. Missing genes in metabolic pathways: a comparative genomics approach. *Curr. Opin. Chem. Biol.* 7:238–51

157. Palenik B, Grimwood J, Aerts A, Rouzé P, Salamov A, et al. 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *PNAS* 104:7705–10

158. Parfrey LW, Lahr DJ, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *PNAS* 108:13624–29

159. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–67

160. Pazour GJ, Agrin N, Leszyk J, Witman GB. 2005. Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* 170:103–13

161. Peers G, Price NM. 2006. Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* 441:341–44

162. Pellegrini M, Marcotte E, Thompson M, Eisenberg D, Yeates T. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *PNAS* 96:4285–88

163. Perry JJP, Shin DS, Getzoff ED, Tainer JA. 2010. The structural biochemistry of the superoxide dismutases. *Biochim. Biophys. ActaProteins Proteom.* 1804:245–62

164. Pocock T, Lachance MA, Pröschold T, Priscu JC, Kim SS, Huner N. 2004. Identification of a psychrophilic green alga from Lake Bonney Antarctica: *Chlamydomonas raudensis* Ettl. (UWO 241) Chlorophyceae. *J. Phycol.* 40:1138–48

165. Polle JEW, Barry K, Cushman J, Schmutz J, Tran D, et al. 2017. Draft nuclear genome sequence of the halophilic and beta-carotene-accumulating green alga *Dunaliella salina* strain CCAP19/18. *Genome Announc.* 5:e01105–17

166. Pombert J-F, Blouin NA, Lane C, Boucias D, Keeling PJ. 2014. A lack of parasitic reduction in the obligate parasitic green alga *Helicosporidium*. *PLOS Genet.* 10:e1004355

167. Poptsova MS, Gogarten JP. 2010. Using comparative genome analysis to identify problems in annotated microbial genomes. *Microbiology* 156:1909–17

168. Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, et al. 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* 335:843–47

169. Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, et al. 2010. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carter*i. *Science* 329:223–26

170. Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 33:D501–4

171. Qiu H, Price DC, Weber APM, Facchinelli F, Yoon HS, Bhattacharya D. 2013. Assessing the bacterial contribution to the plastid proteome. *Trends Plant Sci.* 18:680–87

172. Qiu H, Price DC, Weber APM, Reeb V, Yang EC, et al. 2013. Adaptation through horizontal gene transfer in the cryptoendolithic red alga *Galdieria phlegrea*. *Curr. Biol.* 23:R865–66

173. Qiu H, Yoon HS, Bhattacharya D. 2016. Red algal phylogenomics provides a robust framework for inferring evolution of key metabolic pathways. *PLOS Curr. Tree Life* 8:7b037376e6d84a1be34af756a4d90846

174. Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, et al. 2012. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*. *Nat. Commun.* 3:686

175. Raimunda D, González-Guerrero M, Leeber BW, Argüello JM. 2011. The transport mechanism of bacterial Cu^{+}-ATPases: distinct efflux rates adapted to different function. *Biometals* 24:467–75

176. Rastogi A, Maheswari U, Dorrell RG, Vieira FRJ, Maumus F, et al. 2018. Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci. Rep.* 8:4834

177. Raven JA, Cockell CS, De La Rocha CL. 2008. The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Philos. Trans. R. Soc. B* 363:2641–50

178. Raven JA, Giordano M. 2017. Acquisition and metabolism of carbon in the Ochrophyta other than diatoms. *Philos. Trans. R. Soc. B* 372:20160400

179. Raven JA, Giordano M, Beardall J, Maberly SC. 2012. Algal evolution in relation to atmospheric CO_2: carboxylases, carbon-concentrating mechanisms and carbon oxidation cycles. *Philos. Trans. R. Soc. B* 367:493–507

180. Raymond JA, Kim HJ. 2012. Possible role of horizontal gene transfer in the colonization of sea ice by algae. *PLOS ONE* 7:e35968

181. Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, et al. 2013. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* 499:209–13

182. Ref. Genome Group Gene Ontol. Consort. 2009. The Gene Ontology's Reference Genome Project: a unified framework for functional annotation across species. *PLOS Comput. Biol.* 5:e1000431

183. Reinfelder JR, Kraepiel AM, Morel FM. 2000. Unicellular C_4 photosynthesis in a marine diatom. *Nature* 407:996–99

184. Reiskind JB, Bowes G. 1991. The role of phosphoenolpyruvate carboxykinase in a marine macroalga with C_4-like photosynthetic characteristics. *PNAS* 88:2883–87

185. Reyes-Prieto A, Weber APM, Bhattacharya D. 2007. The origin and establishment of the plastid in algae and plants. *Annu. Rev. Genet.* 41:147–68

186. Rhoads A, Au KF. 2015. PacBio sequencing and its applications. *Genom. Proteom. Bioinform.* 13:278–89

187. Roth MS, Cokus SJ, Gallaher SD, Walter A, Lopez D, et al. 2017. Chromosome-level genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates astaxanthin production. *PNAS* 114:E4296–305

188. Ruprecht C, Vaid N, Proost S, Persson S, Mutwil M. 2017. Beyond genomics: studying evolution with gene coexpression networks. *Trends Plant Sci*. 22:298–307

189. Sánchez-Baracaldo P, Raven JA, Pisani D, Knoll AH. 2017. Early photosynthetic eukaryotes inhabited low-salinity habitats. *PNAS* 114:E7737–45

190. Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB. 2010. Microbial biosynthesis of alkanes. *Science* 329:559–62

191. Schneider A, Steinberger I, Herdean A, Gandini C, Eisenhut M, et al. 2016. The evolutionarily conserved protein PHOTOSYNTHESIS AFFECTED MUTANT71 is required for efficient manganese uptake at the thylakoid membrane in Arabidopsis. *Plant Cell* 28:892–910

192. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLOS Comput. Biol.* 5:e1000605

193. Schönknecht G, Chen W-H, Ternes CM, Barbier GG, Shrestha RP, et al. 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* 339:1207–10

194. Schönknecht G, Weber APM, Lercher MJ. 2014. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *BioEssays* 36:9–20

195. Sevcikova T, Horak A, Klimes V, Zbrankova V, Demir-Hilton E, et al. 2015. Updating algal evolutionary relationships through plastid genome sequencing: Did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci. Rep.* 5:10134

196. Sevim V, Bashir A, Chin CS, Miga KH. 2016. Alpha-CENTAURI: assessing novel centromeric repeat sequence variation with long read sequencing. *Bioinformatics* 32:1921–24

197. Shen C, Dupont CL, Hopkinson BM. 2017. The diversity of CO_2-concentrating mechanisms in marine diatoms as inferred from their genetic content. *J. Exp. Bot.* 68:3937–48

198. Shikanai T, Müller-Moulé P, Munekage Y, Niyogi KK, Pilon M. 2003. PAA1, a P-type ATPase of Arabidopsis, functions in copper transport in chloroplasts. *Plant Cell* 15:1333–46

199. Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, et al. 2013. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr. Biol.* 23:1399–408

200. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–12

201. Sonnhammer ELL, Gabaldón T, Sousa da Silva AW, Martin M, Robinson-Rechavi M, et al. 2014. Big data and other challenges in the quest for orthologs. *Bioinformatics* 30:2993–98

202. Sorigué D, Légeret B, Cuiné S, Blangy S, Moulin S, et al. 2017. An algal photoenzyme converts fatty acids to hydrocarbons. *Science* 357:903–7

203. Sorigué D, Légeret B, Cuiné S, Morales P, Mirabella B, et al. 2016. Microalgae synthesize hydrocarbons from long-chain fatty acids via a light-dependent pathway. *Plant Physiol.* 171:2393–405

204. Soupene E, Inwood W, Kustu S. 2004. Lack of the Rhesus protein Rh1 impairs growth of the green alga *Chlamydomonas reinhardtii* at high CO_2. *PNAS* 101:7787–92

205. Soupene E, King N, Feild E, Liu P, Niyogi KK, et al. 2002. Rhesus expression in a green alga is regulated by CO_2. *PNAS* 99:7769–73

206. Steffensen DA. 1976. Morphological variation of *Ulva* in the Avon-Heathcote Estuary, Christchurch. *N. Z. J. Mar. Freshw. Res.* 10:329–41

207. Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J. 2014. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat. Commun.* 5:5764

208. Sultemeyer D, Rinast KA. 1996. The CO_2 permeability of the plasma membrane of *Chlamydomonas reinhardtii*: Mass-spectrometric ^{18}O-exchange measurements from $^{13}C^{18}O_2$ in suspensions of carbonic anhydrase-loaded plasma-membrane vesicles. *Planta* 200:358–68

209. Suzuki S, Yamaguchi H, Nakajima N, Kawachi M. 2018. *Raphidocelis subcapitata* (=*Pseudokirchneriella subcapitata*) provides an insight into genome evolution and environmental adaptations in the Sphaeropleales. *Sci. Rep.* 8:8058

210. Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, et al. 2015. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *Plant Cell* 27:162–76

211. Tartar A, Boucias DG, Becnel JJ, Adams BJ. 2003. Comparison of plastid 16S rRNA (*rrn16*) genes from *Helicosporidium* spp.: evidence supporting the reclassification of Helicosporidia as green algae (Chlorophyta). *Int. J. Syst. Evol. Microbiol.* 53:1719–23

212. Terashima M, Specht M, Naumann B, Hippler M. 2010. Characterizing the anaerobic response of *Chlamydomonas reinhardtii* by quantitative proteomics. *Mol. Cell Proteom.* 9:1514–32

213. Thomas DN, Dieckmann GS. 2002. Antarctic sea ice—a habitat for extremophiles. *Science* 295:641–44

214. Tian W, Skolnick J. 2003. How well is enzyme function conserved as a function of pairwise sequence identity? *J. Mol. Biol.* 333:863–82

215. Tottey S, Rich PR, Rondet SA, Robinson NJ. 2001. Two Menkes-type ATPases supply copper for photosynthesis in *Synechocystis* PCC 6803. *J. Biol. Chem.* 276:19999–20004

216. Traller JC, Cokus SJ, Lopez DA, Gaidarenko O, Smith SR, et al. 2016. Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnol. Biofuels* 9:258

217. Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13:36–46

218. Trentacoste EM, Shrestha RP, Smith SR, Gle C, Hartmann AC, et al. 2013. Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *PNAS* 110:19748–53

219. Treves H, Raanan H, Finkel OM, Berkowicz SM, Keren N, et al. 2013. A newly isolated *Chlorella* sp. from desert sand crusts exhibits a unique resistance to excess light intensity. *FEMS Microbiol. Ecol.* 86:373–80

220. Tu CK, Acevedo-Duncan M, Wynns GC, Silverman DN. 1986. Oyxgen-18 exchange as a measure of accessibility of CO_2 and $HCO_3{}^-$ to carbonic anhydrase in *Chlorella vulgaris* (UTEX 263). *Plant Physiol.* 80:997–1001

221. Tulin F, Cross FR. 2014. A microbial avenue to cell cycle control in the plant superkingdom. *Plant Cell* 26:4019–38

222. Tyra HM, Linka M, Weber APM, Bhattacharya D. 2007. Host origin of plastid solute transporters in the first photosynthetic eukaryotes. *Genome Biol.* 8:R212

223. UniProt Consort. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43:D204–12

224. Vadeboncoeur Y, Power ME. 2017. Attached algae: the cryptic base of inverted trophic pyramids in freshwaters. *Annu. Rev. Ecol. Evol. Syst.* 48:255–79

225. van Baren MJ, Bachy C, Reistetter EN, Purvine SO, Grimwood J, et al. 2016. Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. *BMC Genom.* 17:267

226. Vieler A, Wu G, Tsai CH, Bullard B, Cornish AJ, et al. 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. *PLOS Genet*. 8:e1003064

227. Wang Q, Sun H, Huang J. 2017. Re-analyses of "algal" genes suggest a complex evolutionary history of oomycetes. *Front. Plant Sci.* 8:1540

228. Ward BA, Follows MJ. 2016. Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *PNAS* 113:2958–63

229. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, et al. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35:543–48

230. Wheatley NM, Sundberg CD, Gidaniyan SD, Cascio D, Yeates TO. 2014. Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the α-carboxysome. *J. Biol. Chem.* 289:7973–81

231. Wheeler G, Ishikawa T, Pornsaksit V, Smirnoff N. 2015. Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife* 4:e06369

232. Whitney LP, Lins JJ, Hughes MP, Wells ML, Chappell PD, Jenkins BD. 2011. Characterization of putative iron responsive genes as species-specific indicators of iron stress in *Thalassiosiroid* diatoms. *Front. Microbiol.* 2:234

233. Wierzchos J, DiRuggiero J, Vítek P, Artieda O, Souza-Egipsy V, et al. 2015. Adaptation strategies of endolithic chlorophototrophs to survive the hyperarid and extreme solar radiation environment of the Atacama Desert. *Front. Microbiol.* 6:934

234. Winck FV, Riaño-Pachón DM, Sommer F, Rupprecht J, Mueller-Roeber B. 2012. The nuclear proteome of the green alga *Chlamydomonas reinhardtii*. *Proteomics* 12:95–100

235. Woo YH, Ansari H, Otto TD, Klinger CM, Kolisko M, et al. 2015. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *eLife* 4:e06974

236. Worden AZ, Lee JH, Mock T, Rouzé P, Simmons MP, et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324:268–72

237. Xue J, Niu YF, Huang T, Yang WD, Liu JS, Li HY. 2015. Genetic improvement of the microalga *Phaeodactylum tricornutum* for boosting neutral lipid accumulation. *Metab. Eng.* 27:1–9

238. Ye N, Zhang X, Miao M, Fan X, Zheng Y, et al. 2015. *Saccharina* genomes provide novel insight into kelp biology. *Nat. Commun.* 6:6986

239. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21:809–18

240. Zaslavskaia LA, Lippmeier JC, Shih C, Ehrhardt D, Grossman AR, Apt KE. 2001. Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. *Science* 292:2073–75

241. Zhan Y, Marchand CH, Maes A, Mauries A, Sun Y, et al. 2018. Pyrenoid functions revealed by proteomics in *Chlamydomonas reinhardtii*. *PLOS ONE* 13:e0185039

242. Zhang B, Zhang C, Liu C, Jing Y, Wang Y, et al. 2018. Inner envelope CHLOROPLAST MANGANESE TRANSPORTER 1 supports manganese homeostasis and phototrophic growth in *Arabidopsis*. *Mol. Plant* 11:943–54

243. Zimmer A, Lang D, Richardt S, Frank W, Reski R, Rensing SA. 2007. Dating the early evolution of plants: detection and molecular clock analyses of orthologs. *Mol. Genet. Genom.* 278:393–402