Trends in Plant Science
The Algal Revolution
--Manuscript Draft--

Manuscript Number:	PLANTS-D-17-00020R1
Article Type:	Review
Corresponding Author:	Debashish Bhattacharya, PhD Rutgers University New Brunswick, New Jersey UNITED STATES
First Author:	Debashish Bhattacharya, PhD
Order of Authors:	Debashish Bhattacharya, PhD
	Juliet Brodie
	Cheong Xin Chan
	Olivier De Clerck
	J. Mark Cock
	Susana M Coelho
	Claire Gachon
	Arthur R Grossman
	Thomas Mock
	John A Raven
	Alison G Smith
	Hwan Su Yoon
Abstract:	Algae are (mostly) photosynthetic eukaryotes that occupy multiple branches of the tree of life, and are vital for planet function and health. This review highlights a transformative period in studies of the evolution and functioning of this extraordinary group of organisms and their potential for novel applications, wrought by high-throughput 'omic' and reverse genetic methods. It covers the origin and diversification of algal groups, explores advances in understanding the link between phenotype and genotype, considers algal sex determination, and reviews progress in understanding the roots of algal multicellularity. Experimental evolution studies to determine how algae evolve in changing environments are highlighted, as is their potential as production platforms for compounds of commercial interest such as biofuel precursors, nutraceuticals, or therapeutics.

1   **Review**

2   # The Algal Revolution

3

4   Juliet Brodie,[1,†] Cheong Xin Chan,[2] Olivier De Clerck,[3] J. Mark Cock,[4] Susana M. Coelho,[4]

5   Claire Gachon,[5] Arthur R. Grossman,[6] Thomas Mock,[7] John A. Raven,[8,9] Alison G. Smith,[10]

6   Hwan Su Yoon,[11] Debashish Bhattacharya[12, †,*]

7

8   [1] Natural History Museum, Department of Life Sciences, London SW7 5BD, United Kingdom

9   [2] Institute for Molecular Bioscience, The University of Queensland, Brisbane QLD 4072,

10   Australia

11   [3] Research Group Phycology, Ghent University, Krijgslaan 281, S8, 9000 Gent, Belgium

12   [4] CNRS, Sorbonne Université, UPMC University Paris 06, Algal Genetics Group, UMR 8227,

13   Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, Roscoff F-

14   29688, France

15   [5] Scottish Association for Marine Science, Scottish Marine Institute, PA37 1QA Oban, United

16   Kingdom

17   [6] Department of Plant Biology, The Carnegie Institution, Stanford, CA 94305, USA

18   [7] School of Environmental Sciences, University of East Anglia, Norwich Research Park,

19   Norwich NR4 7TJ, United Kingdom

20   [8] Division of Plant Sciences, University of Dundee at the James Hutton Institute, Dundee DD2

21   5DA, United Kingdom (permanent address)

22   [9] School of Plant Biology, University of Western Australia, Crawley, WA 6009, Australia

23   [10] Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United

24   Kingdom

25   [11] Department of Biological Sciences, Sungkyunkwan University, Suwon 440-746, Korea

26   [12] Department of Ecology, Evolution and Natural Resources, Rutgers University, New

27   Brunswick, NJ 08901, USA

28   [†] These authors contributed equally to this work.

29   [*] For correspondence: d.bhattacharya@rutgers.edu (D. Bhattacharya)

30   Laboratory website: http://dblab.rutgers.edu/

34

35   **Abstract**

36   Algae are (mostly) photosynthetic eukaryotes that occupy multiple branches of the tree of life,

37   and are vital for planet function and health. This review highlights a transformative period in

38   studies of the evolution and functioning of this extraordinary group of organisms and their

39   potential for novel applications, wrought by high-throughput 'omic' and reverse genetic

40   methods. It covers the origin and diversification of algal groups, explores advances in

41   understanding the link between phenotype and genotype, considers algal sex determination, and

42   reviews progress in understanding the roots of algal multicellularity. Experimental evolution

43   studies to determine how algae evolve in changing environments are highlighted, as is their

44   potential as production platforms for compounds of commercial interest such as biofuel

45   precursors, nutraceuticals, or therapeutics.

46

The diversity and ecological importance of algae

Algae represent a vast array of photosynthetic and non-photosynthetic eukaryotes of ancient origin scattered throughout the tree of life (Fig. 1A; **Box 1**), but less than 10% have been formally described [1]. Algae are critical to the health of our planet because they are dominant in the oceans, which cover about 71% of the Earth's surface where they provide core ecosystem services and produce about one-half of the oxygen that we breathe [2]. Algae are also integral to human activities; they can serve as a food source or in some cases represent a significant nuisance or health hazard (**Box 2**). These organisms occupy a vast range of habitats from desert crusts to coastal and oligotrophic oceans where massive algal blooms can extend over thousands of km^2. Furthermore, the body mass of algae can differ by 20 orders of magnitude (Fig. 1B). *Ostreococcus*, a green microalga has a volume of <1 μm^3 making it the smallest known free-living eukaryote [3], the giant kelp (*Macrocystis pyrifera*) can grow up to 45 m in length [4], and the siphonous green alga *Caulerpa* has cells that grow to a meter in length and differentiate into distinct 'organs' [5]. The brown and red macroalgae are two of the few groups of organisms that have made the transition from unicellularity to complex multicellularity (**Box 1**; and see below). In addition to the photosynthetic lineages that exploit sunlight to reduce CO_2 to organic carbon, there exist a myriad of mixotrophs and facultative heterotrophs, some of which have lost the plastid and evolved into obligate heterotrophs such as *Plasmodium* and *Toxoplasma* that parasitize other organisms [6-8]. The study of algae is undergoing a revolution, largely due to the development of increasingly sophisticated 'omics' (e.g., genomics, proteomics, epigenomics, metagenomics), bioinformatics, systems biology, and novel reverse genetic approaches such as the use of CRISPR-Cas9 for editing their genomes [9-12]. In contrast to the routine application of genomic editing methods to land plant species, their development in different algal groups such as diatoms, haptophytes, and red algae is highly challenging because these taxa are ancient splits in the eukaryotic tree of life. For example, tools such as transformation protocols, transgene promoters, and nuclear localization signals usually have to be developed specifically for each lineage. Nonetheless, modern high-throughput methods have allowed researchers to generate testable hypotheses that set the stage for years of detailed biochemical, cell biological, and genetic experimentation. Here, we document the transformative impact of modern omics and reverse genetics research on our understanding of algae.

Exploring the origins and functions of algal genes

79 To understand the success of algae, it is important to identify the major genetic processes that

80 have shaped their genomes and how these processes impacted phenotypes. However,

81 disentangling the effects of genotype versus phenotype on algal evolution and fitness is

82 challenging because both are linked and highly plastic. In this section, we discuss the interplay

83 between genotype and phenotype, and review new tools for functional analysis in algae (**Box 3**).

84

85 *Endosymbiotic gene transfer*

86 The most significant introduction of foreign genes into the nuclear genome of all photosynthetic

87 organisms was the result of endosymbiotic gene transfer (**EGT**; see Glossary) [13], which

88 enriched the nuclear genome with cyanobacterial genes originating from the primary plastid

89 endosymbiont (Fig. 1C and **Box 1**). Most EGT-derived gene products are targeted back to the

90 organelle (via identifiable N-terminal targeting sequences called transit peptides) where they

91 express their original or related function, whereas others evolved novel roles in the host cell [14,

92 15]. The plastid-destined EGT proteins encode a range of functions, many of which are

93 associated with photosynthesis, metabolite biosynthesis, transcription and translation, plastid

94 biogenesis and macromolecular complex assembly, and the regulation of photosynthetic activity

95 [16, 17]. EGT-derived cyanobacterial genes may also have fused with various DNA sequences

96 from the host or from various prokaryotes to create the so-called symbiogenetic genes (**S-genes;**

97 see Glossary) [18]. S-genes play key roles in algal and plant metabolism, including responses to

98 oxidative stress, phototropism, and adaptation to N limitation [18]. In the following section, we

99 focus on the search for functions encoded by EGT-derived genes using the model green alga,

100 *Chlamydomonas reinhardtii* and the so-called **GreenCut** (see Glossary).

101

102 *The GreenCut proteins*

103 Comparative genomics using the *C. reinhardtii* genome and genomes from other green lineage

104 organisms, including the terrestrial embryophytes *Arabidopsis thaliana*, *Physcomitrella patens*,

105 *Oryza sativa*, *Populus trichocarpa*, as well as *Ostreococcus* species (*O. tauri*, *O. lucimarinus*,

106 or *Ostreococcus* sp. RCC809) revealed a set of proteins designated the GreenCut. These genes

107 are absent from, or are highly diverged in heterotrophic (non-photosynthetic) organisms

108 [16]. The latest GreenCut (designated GreenCut2) includes a set of 597 *C. reinhardtii* proteins

109 [16, 17], of which almost one-half have unknown functions, although many have domains that

110 provide some information about function. Whereas a subset of these proteins are exclusively

111 conserved in green lineage organisms (Conserved in Green Lineage, CGL), others are also found

112 in at least one red alga (Conserved in PLantae, CPL), at least one diatom (Conserved

113 in Green Lineage and Diatoms, CGLD), or in the Plantae (in the Archaeplastida) and diatoms

114 (Conserved in the PLantae and Diatoms, CPLD) [19]. Orthologs of many GreenCut proteins are

115 also present in cyanobacteria. The use of the GreenCut suite of proteins assumes that the

116 functions of member proteins of any specific ortholog group are conserved among green lineage

117 organisms.

118 Approximately 70% of GreenCut proteins are predicted to be targeted to chloroplasts

119 where they function in photosynthesis and several other metabolic pathways, chloroplast

120 biogenesis/assembly, and the regulation of photosynthetic activity [16, 17]. A recently

121 characterized GreenCut protein, CGL71, resides in thylakoid membranes, has a tetratricopeptide

122 repeat domain, and has been shown to be required for accumulation/maintenance of photosystem

123 I (PSI) [20-22]. A *C. reinhardtii* mutant that is null for CGL71 exhibits very low levels of PSI,

124 cannot grow as a photoautotroph, and when grown heterotrophically, is high light sensitive.

125 Amazingly, this mutant can be largely rescued (ca. 70% recovery of PSI activity) when it is

126 maintained under hypoxic conditions. These results suggest that CGL71 functions under

127 atmospheric O_2 conditions to prevent oxidative disruption of the maturation/assembly of PSI,

128 which might reflect the need for PSI to associate with O_2 sensitive iron-sulfur (Fe-S)

129 clusters. These findings have implications for the evolution of the Earth's atmosphere. Oxygenic

130 photosynthesis evolved roughly 2.5 BYA when the atmosphere was largely anoxic. Therefore,

131 even an O_2-sensitive cofactor/complex that became integral to the ancestral photosynthetic

132 electron transport system would have likely remained stable in the early Earth's atmosphere.

133 However, as the atmosphere transitioned to oxic conditions as a consequence of photosynthetic

134 O_2 evolution, specific proteins, like CGL71, may have evolved to assist in assembling and

135 stabilizing complexes containing O_2-sensitive cofactors. The need for mechanisms to prevent

136 oxic disruption of both assembly intermediates and mature macromolecular complexes extends

137 beyond photosynthetic processes; i.e., it would be critical for any complex/activity in the cell that

138 requires an O_2 sensitive cofactor.

139

Algae in a changing environment

Microalgae have proven to be ideal models for experimental evolution approaches due to their ability to adapt and/or acclimate to changing environmental conditions. Their short generation times and asexual mode of reproduction makes it possible to cultivate them for hundreds to thousands of generations within a relatively short time. Despite these advantages, the large number of interacting variables associated with increasing atmospheric CO_2 and temperature makes it a challenge to design short-term studies to probe the acclimation of algae to these factors, and longer-term experimental evolution studies of genetic change. Accordingly, most experimental work has involved only a single variable: increased CO_2 [23, 24]. One of the longest experiments (4 years, 2100 generations [24]) examined the response of a high-CO_2-adapted population of the bloom-forming *Emiliania huxleyi* after it was returned to its original CO_2 condition and concluded that phytoplankton may evolve complex phenotypic plasticity. An instructive example for algal systems focused on N_2 fixation by the marine cyanobacterium *Trichodesmium* that was exposed to elevated CO_2 (750 p.p.m.) for 4.5 years (about 850 generations). These cultures showed significantly higher N_2 fixation rates and growth rates under P-limited conditions, as well as shifts in the diel occurrence of peak N_2 fixation [25]. The results were consistent with those for short-term (two week) incubations in elevated CO_2. These effects were maintained even when the cultures were returned to the ancestral CO_2 levels (380 p.p.m.) for 2 years. However, analysis of the proteome and enzyme activities did not reveal the basis of these changes [25]. These results might indicate a high level of phenotypic plasticity and surprisingly, genetic fixation of the adaptive phenotypes after a relatively short exposure to high CO_2. To test this hypothesis would require identifying genetic and epigenetic processes of adaptation, which are currently not well understood in cyanobacteria and algae [26].

Padfield et al. [27] investigated the changes in the metabolism of freshwater *Chlorella vulgaris* over 100 generations of exposure to a higher temperature, but without demonstrating that the effects were due to adaptation rather than acclimation. Schlüter et al. [28] examined the impact of increased temperature and increased CO_2, separately and together, on *E. huxleyi* and showed that there was no interaction of CO_2 and temperature. Bermúdez et al. [29] investigated the nutritional value of marine foods using a fully factorial design (three CO_2 concentrations, two temperatures) and reported adverse effects of high CO_2 on the synthesis of essential amino acids and polyunsaturated fatty acids in *Cylindrotheca fusiformis* over 250 generations. Innovative

171   work by Schaum and Collins [30], Schaum et al. [31], and Doblin and Sebille [32] provide hope

172   for predicting the adaptation of phytoplankton strains to changes in CO_2 and temperature levels.

173   Transcriptomic data are available for some acclimation and adaptation studies [33], but

174   there is very little genomic analysis coupled to adaptation experiments, except for the work done

175   by Perrineau et al. [34, 35] on the evolution of laboratory cultures of *C. reinhardtii*. As long as

176   we do not understand how the genetic repertoire of algae underpins the response to changing

177   environmental conditions, the evolutionary mechanisms remain elusive. For instance, the polar

178   diatom *Fragilariopsis cylindrus* has evolved to cope with what can be considered one of the

179   most extreme environmental changes in nature: the transformation from a liquid (sea water) into

180   a partly habitat (sea ice) [36]. The genome sequence of this diatom revealed that almost a quarter

181   of the genes had markedly divergent alleles that were differentially expressed under changing

182   environmental conditions. Metatranscriptomes from natural communities of *F. cylindrus* were

183   dominated by these divergent alleles, providing evidence that the alleles are important for

184   cellular function and were selected by the polar environment as an evolutionary mechanism

185   underpinning the success of *F. cylindrus* under highly variable conditions. Whether these

186   diverged alleles will provide an advantage for coping with global warming remains unknown,

187   but with an effective population size (N_e) $\approx$ 16.5 x 10^7, it might be said that there is an allele for

188   every occasion.

189

190   **Algal sex determination**

191   Sexual cycles are an ancestral feature of eukaryotes, and genomic analyses demonstrate the

192   presence of meiotic genes in all supergroups, including those containing algae [37]. The systems

193   that regulate mating (mating type loci and sex chromosomes) are more recent in origin,

194   remarkably diverse, and have emerged independently and repeatedly during evolution [38].

195   In genetically controlled sexual systems, mating types or sexes are determined by defined

196   non-recombining chromosomal regions that can be as small as a single locus or as large as an

197   entire chromosome. Analysis of green and brown (stramenopile) algal lineages has led to novel

198   and important contributions to our understanding of how sexes have emerged and how sex-

199   determining mechanisms evolve. One particularly interesting group is the volvocine algae, which

200   possess a broad range of sexual systems ranging from unicellular, isogamous species such as *C.*

201   *reinhardtii*, which have two equal-sized gametes of plus and minus mating types, to

202 multicellular, oogamous species such as *Volvox carteri*, with sperm-producing males and egg-

203 producing females. Comparative genomic analyses of volvocine species have shed new light on

204 the origin of male and female sexes, providing clear evidence that the sexes emerged from

205 mating types in this lineage [39-41]. Surprisingly, the emergence of a sexual system with male

206 sperm and female eggs in *V. carteri* does not appear to have involved the recruitment of

207 additional genes (influencing gamete size for example) into the sex-determining genomic region,

208 as was previously predicted [42]. Rather, sexual dimorphism can arise from isogamy largely via

209 adaptations of the master sex-determining gene (*MID*) itself. These changes appear to re-wire

210 regulatory networks, such that in an isogamous organism, the result is a simple system that

211 determines mating type. In contrast, oogamous organism such as *Volvox* require more complex

212 developmental programs that lead to the determination of spermatogenesis or oogenesis [39].

213   The brown algae represent a key group for studying the evolution of sexes because they

214 too exhibit a broad range of different sexual characteristics (e.g., isogamy/ anisogamy, sex

215 determination in either the diploid or the haploid phase of the life cycle, and varying levels of

216 sexual dimorphism) [43]. Sexual systems that function during the haploid phase of the life cycle

217 (so-called UV sexual systems, where U and V refer to the female and male sex chromosomes,

218 respectively) are of particular interest because they exhibit novel evolutionary characteristics

219 compared to the better studied diploid phase systems (XY and ZW sex chromosomes) [38, 43,

220 44]. Analysis of the UV sex chromosomes of the brown alga *Ectocarpus* provided the first

221 detailed genetic description of a sex-determination system for a multicellular species outside the

222 opisthokont and green plant lineages [44]. These data supported theoretical predictions about the

223 evolutionary dynamics of sex chromsomes that could be tested empirically. For example, it has

224 been proposed that purifying selection during the haploid phase of the life cycles should prevent

225 degeneration of both the U and V sex-determining regions. Analysis of the *Ectocarpus* sex-

226 determining region did not suggest marked degeneration but there was, nonetheless, evidence for

227 some minor genic erosion. Theoretical models also predict that the presence of sexually

228 antagonistic genes in the recombining regions of sex chromosomes may drive expansion of the

229 non-recombining sex-determining region [45]. The small size of the *Ectocarpus* sex-determining

230 region is consistent with this prediction, given that this alga exhibits a low degree of sexual

231 dimorphism and has few sex-biased genes, indicative of a low level of sexual antagonism [46].

232 The recombining regions of the *Ectocarpus* sex chromosome, known as pseudoautosomal

8

233 regions, also have exceptional evolutionary and structural features. These regions are enriched in

234 orphan genes that may be selectively maintained because of their important roles in sporophytes

235 [47]. Ongoing analysis of the diverse algal sexual systems across the brown algal tree of life is

236 expected to bring new insights not only to the understanding of the mechanics of sex

237 chromosome function, but also to the interplay between the sexual system and sex chromosome

238 evolution.

239

240 **Advances in algal biotechnology**

241 With the availability of >30 microalgal genomes and >500 transcriptomes [48], there are

242 significant opportunities to exploit algal metabolism for biotechnological purposes such as

243 biofuels, pharmaceuticals, nutraceuticals and biomaterials. For the purpose of this review, we

244 will only highlight some of these areas because the field of algal biotechnology is rapidly

245 expanding with many studies being done on a variety of algal species. The reason why algae are

246 promising organisms for biotechnology is again rooted in their evolution. Their genetic diversity

247 results in biochemical diversity, which offers opportunities to discover novel metabolic pathways

248 and novel active molecules to serve many different biotechnological purposes [49]. Moreover,

249 being photosynthetic means they offer an advantage over the more traditional bacterial or yeast

250 hosts that require inputs of fixed carbon, and so in principle algae are more sustainable.

251 However, this is not guaranteed, and any industrial process requires careful life cycle assessment

252 to establish its level of sustainability [50]. The fact that many algae can grow mixotrophically or

253 heterotrophically would potentially provide an alternative if growth in photobioreactors is

254 unsustainable or too constrained by the footprint of any commercial venture.

255   Algae have been the subject of much investigation for biofuel production because many

256 of them accumulate triacylglycerides (TAGs), which can be used as a feedstock for biodiesel

257 [51], although in many cases the accumulation of TAG only occurs in response to nutrient

258 deprivation (e.g., nitrogen [N]). As a result, cell growth is inhibited leading to poor overall

259 productivity, which is one of the major challenges that need to be overcome to make algal

260 biofuels a commercial reality [52, 53]. Recent work shows that microalgae encode multiple

261 genes for enzymes that catalyze the last step in the metabolic pathway for the production of

262 TAG, the addition of a third acyl group onto diacylglycerol [54]. By analyzing RNAseq data it

263 was possible to find which of these genes were upregulated upon N-starvation, and therefore

264 potentially involved in the increase in TAG synthesis during starvation. In *C. reinhardtii* only

265 one of the five genes encoding diacylglycerol acyltransferase type 2 (DGTT1) had this

266 characteristic [55]. However, overexpression of this and two other DGAT2 genes in *C.*

267 *reinhardtii* under the control of the strong light-responsive *PSAD* promoter had no effect on total

268 lipid or TAG levels, likely indicating tight regulation of this pathway. A similar experiment in

269 the diatom *Phaeodactylum tricornutum* found that expression of Pt*DGAT2A* under control of the

270 light-responsive *FCPC* promoter somewhat increased neutral lipid levels, but also had an effect

271 on the proportion of unsaturated fatty acids in all cellular glycerolipids, not just TAGs [56]. This

272 illustrates a common observation, which is the generation of unexpected consequences from the

273 introduction of genes for metabolic enzymes.

274   More detailed systems-level expression has started to identify key factors important in the

275 cellular response to alterations in the C:N ratio. Transcriptomic analysis of N-starved *P.*

276 *tricornutum* cells provided evidence that N limitation led to a remodeling of intermediary

277 metabolism that shifted the flux of photosynthetically assimilated carbon from amino acid

278 biosynthesis towards lipids, and helped to conserve N further by recycling products of protein

279 degradation through the urea cycle [57]. Boyle et al. [58] analyzed *C. reinhardtii* RNAseq data

280 and found that a transcript for a SQUAMOSA promoter-binding protein domain transcription

281 factor increased just prior to the transcript encoding DGTT1. Mutants in this gene, named

282 *NRR1* for nitrogen response regulator, showed much lower TAG accumulation. A slightly

283 different approach was taken with *P. tricornutum*. Promoters of genes that were activated early

284 and strongly upregulated upon N deprivation contain certain overrepresented motifs. A RING-

285 domain protein (RGQ1) identified by yeast 1-hybrid analysis that binds these N-deprivation

286 motifs acts as a transcription regulator [59]. These insights suggest a number of strategies that

287 might provide effective targets for manipulation, such as the reduction of lipid catabolism to

288 engineer the synthesis of TAG without compromising growth. An example of this approach was

289 the knockdown of lipid catabolism, specifically lipases that catalyze the release of free fatty

290 acids (FAs) from lipids, and therefore increase lipid accumulation. Targeted knockdown of a

291 multifunctional lipase/phospholipase/ acyltransferase increased lipid yields without affecting

292 growth in the diatom *Thalassiosira pseudonana* [60]. As also observed in *C. reinhardtii*,

293 overexpressing genes encoding enzymes of TAG biosynthesis was less successful for stimulating

294 lipid accumulation [60].

295   Recently, the focus on algal bioproducts has turned away from low-value, high volume

296   biofuels, to high value products such as nutraceuticals (e.g., vitamins, pigments, antioxidants),

297   omega-3 fatty acids, or other novel chemicals [61, 62]. The diversity of algal species means that

298   many novel pathways remain to be identified. For example *Botryococcus braunii*, a colonial

299   green alga secretes copious amounts of various straight-chain and branched hydrocarbons

300   between cells in the colony but grows extremely slowly, reducing its potential as a production

301   strain. However, genes thought to encode specific enzymes involved in hydrocarbon synthesis

302   such as squalene synthase-like enzymes (SSL) and triterpenoid methyltransferase (TMT) for the

303   synthesis of botryococcenes, C30-C37 triterpenoids typical of *B. braunii* Race B, were identified

304   by sequence similarity, and their identities have been validated by expression in yeast [63].

305   Similarly, a recent analysis of *B. braunii* Race L identified a gene encoding an SSL involved in

306   the synthesis of the C40 tetraterpenoid lycopadiene [64], and again the activity of the encoded

307   protein was verified in yeast. These resources offer the means to reconstitute a novel microalgal

308   pathway in a heterologous host, and indeed up to 0.5 mg g^{-1} fresh weight of botryococcenes were

309   produced in tobacco plants into which SSL and TMT genes were introduced, although there were

310   adverse effects on plant growth and morphology [65].

311   Biomineralizing coccolithophores and diatoms have been widely explored for

312   nanotechnology purposes such as drug delivery, nano-sensors, solar technology, microfluidics,

313   catalyst production and biosensing [66-69]. The structural and physical properties of the

314   biomineralized cell walls as in the frustule of diatoms underpin these applications. Purified

315   frustules are used as filter material and their replicas as biosensors. Genetically engineered

316   frustules of *T. pseudonana* that displayed an immunoglobulin G (IgG)-binding domain on their

317   surface had antibodies attached to selectively target and kill cancer cells [69]. Treatment with the

318   drug-loaded frustules led to tumor growth regression in mice. Despite the variety of uses offered

319   by algal cell walls, the main hindrance in being able to exploit fully their structural and physical

320   properties lies in a lack of knowledge about the genes and proteins required for their formation.

321   Furthermore, the functions of those genes and proteins that have been identified thus far are

322   primarily derived from biochemical *in vitro* studies with recombinant proteins. Thus, *in vivo*

323   studies utilizing reverse genetic approaches are needed to reveal their biological functions and

324   therefore to unravel molecular processes that are responsible for the formation of the

325   morphologically complex algal cell walls. The recent establishment of CRISPR/Cas9 to edit the

326   genome of the model diatom *T. pseudonana* [12] is the first step towards elucidating the

327   biogenesis and structural and physical properties of the diatom frustule.

328   Concurrently, the ability to introduce transgenes into microalgae means that it is possible

329   to consider the development of microalgal platforms for industrial production, not just of

330   endogenous molecules but also of a range of non-native compounds, ranging from therapeutics

331   (e.g., plant natural products, vaccines) to platform chemicals used for plastics [70, 71]. As

332   microbes, microalgae can be cultivated in enclosed photobioreactors rather than in open ponds

333   that are susceptible to dynamic environmental conditions, introduction of contaminating

334   eukaryotes and prokaryotes, and predation. Increasingly sophisticated molecular tools are being

335   developed particularly for *C. reinhardtii*, *P. tricornutum*, *T. pseudonana*, and *Nannochloropsis*

336   species. These include a wide range of vectors, promoters, and targeting sequences for transgene

337   expression [72, 73], as well as ways to edit their genomes using TALEN [74] and CRISPR-Cas9

338   [9-12]. Increasingly, genome sequence and transcriptomic data are being mined to identify new

339   regulatory sequences, and this approach also provides information that can be used to manipulate

340   other algae [75-77]. At the same time, synthetic biology approaches (Fig. 2) that exploit

341   engineering design principles are increasingly being applied to algae [78]. The application of the

342   Design-Build-Test cycle coupled with high-throughput methods and automation will speed up

343   identification of parts, verification of gene function, and analysis of mutant phenotypes, to make

344   manipulation much easier. Nevertheless, significant knowledge gaps need to be filled between

345   omics output and assigning gene functions and building metabolic and regulatory networks that

346   will ultimately lead to a systems-level understanding of algal biology.

347

348   **Concluding remarks**

349   Algae are extraordinary organisms that exhibit wide diversity in morphology, physiology, gene

350   content, and sexual systems. They have independently 'discovered' multicellularity (both simple

351   and complex) on several occasions and offer a wide range of biotechnological opportunities to

352   produce high value commercial products. From sustaining many ecosystems through primary

353   production and providing an array of human foods, algae are increasingly being targeted for

354   omics approaches to elucidate their diverse properties. Apart from commercial or academic

355   concerns, the impacts of a warming climate on algal health and the role of these taxa as

356   biomarkers of environmental change are also of paramount importance. The coming years will

357 prove pivotal for algal biologists and the broader public alike as the full array of transformative

358 scientific methods is brought to bear upon these organisms.

359 **Acknowledgements**

365

366 **References**

367 1.   de Vargas, C. et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. Science 348,
368   1261605
369 2.   Barsanti, L. and Gualtieri, P. (2014) *Algae: Anatomy, Biochemistry, and Biotechnology*.
370   CRC Press
371 3.   Courties, C. et al. (1994) Smallest eukaryotic organism. Nature 370, 255-255
372 4.   Steneck, R.S. et al. (2002) Kelp forest ecosystems: biodiversity, stability, resilience and
373   future. Environ. Conserv. 29, 436-459
374 5.   Jacobs, W.P. (1994) *Caulerpa*. Sci. Am. 271, 100-105
375 6.   Armstrong, E. et al. (2000) The first record of *Nitzschia alba* from UK coastal waters with
376   notes on its growth potential. J. Mar. Biol. Assoc. U. K. 80, 355-356
377 7.   Horiguchi, T. (2015) Diversity and phylogeny of marine parasitic dinoflagellates. In
378   *Marine Protists: Diversity and Dynamics* (Ohtsuka, S. et al., eds) pp. 397-419, Springer
379   Japan
380 8.   Stoecker, D.K. et al. (2017) Mixotrophy in the marine plankton. Annu. Rev. Mar. Sci. 9,
381   311-335
382 9.   Wang, Q. et al. (2016) Genome editing of model oleaginous microalgae *Nannochloropsis*
383   spp. by CRISPR/Cas9. Plant J. 88, 1071-1081
384 10.   Shin, S.E. et al. (2016) CRISPR/Cas9-induced knockout and knock-in mutations in
385   *Chlamydomonas reinhardtii*. Sci. Rep. 6, 27810
386 11.   Nymark, M. et al. (2016) A CRISPR/Cas9 system adapted for gene editing in marine algae.
387   Sci. Rep. 6, 24951
388 12.   Hopes, A. et al. (2016) Editing of the urease gene by CRISPR-Cas in the diatom
389   *Thalassiosira pseudonana*. Plant Methods 12, 49
390 13.   Martin, W. and Herrmann, R.G. (1998) Gene transfer from organelles to the nucleus: how
391   much, what happens, and why? Plant Physiol. 118, 9-17
392 14.   Deusch, O. et al. (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a
393   heterocyst-forming plastid ancestor. Mol. Biol. Evol. 25, 748-761
394 15.   Martin, W. et al. (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and
395   chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in
396   the nucleus. Proc. Natl. Acad. Sci. U. S. A. 99, 12246-12251

397   16.   Karpowicz, S.J. et al. (2011) The GreenCut2 resource, a phylogenomically derived
398   inventory of proteins specific to the plant lineage. J. Biol. Chem. 286, 21427-21439
399   17.   Wittkopp, T.M. et al. (2016) The GreenCut - functions and relationships of proteins
400   conserved in green lineage organisms. In *Chloroplasts: Current Research and Future*
401   *Trends* (Kirchhoff, H., ed) pp. 241-278, Caister Academic Press
402   18.   Méheust, R. et al. (2016) Protein networks identify novel symbiogenetic genes resulting
403   from plastid endosymbiosis. Proc. Natl. Acad. Sci. U. S. A. 113, 3579-3584
404   19.   Merchant, S.S. et al. (2007) The *Chlamydomonas* genome reveals the evolution of key
405   animal and plant functions. Science 318, 245-251
406   20.   Kim, J. et al. (2016) Flux balance analysis of primary metabolism in the diatom
407   *Phaeodactylum tricornutum*. Plant J. 85, 161-176
408   21.   Stöckel, J. et al. (2006) The evolutionarily conserved tetratrico peptide repeat protein pale
409   yellow green7 is required for photosystem I accumulation in *Arabidopsis* and copurifies
410   with the complex. Plant Physiol. 141, 870-878
411   22.   Wilde, A. et al. (2001) Characterization of the cyanobacterial ycf37: mutation decreases
412   the photosystem I content. Biochem. J. 357, 211-216
413   23.   Reusch, T.B. and Boyd, P.W. (2013) Experimental evolution meets marine phytoplankton.
414   Evolution 67, 1849-1859
415   24.   Schlüter, L. et al. (2016) Long-term dynamics of adaptive evolution in a globally important
416   phytoplankton species to ocean acidification. Sci. Adv. 2, e1501660
417   25.   Hutchins, D.A. et al. (2015) Irreversibly increased nitrogen fixation in *Trichodesmium*
418   experimentally adapted to elevated carbon dioxide. Nat. Commun. 6, 8155
419   26.   Mock, T. et al. (2016) Bridging the gap between omics and earth system science to better
420   understand how environmental change impacts marine microbes. Glob Chang Biol 22, 61-
421   75
422   27.   Padfield, D. et al. (2015) Rapid evolution of metabolic traits explains thermal adaptation in
423   phytoplankton. Ecol. Lett. 19, 133-142
424   28.   Schlüter, L. et al. (2014) Adaptation of a globally important coccolithophore to ocean
425   warming and acidification. Nat. Clim. Change 4, 1024-1030
426   29.   Bermúdez, R. et al. (2015) Long-term conditioning to elevated pCO_2 and warming
427   influences the fatty and amino acid composition of the diatom *Cylindrotheca fusiformis*.
428   PLoS ONE 10, e0123945
429   30.   Schaum, C.E. and Collins, S. (2014) Plasticity predicts evolution in a marine alga. Proc. R.
430   Soc. B 281, 20141486
431   31.   Schaum, C.E. et al. (2016) Environmental stability affects phenotypic evolution in a
432   globally distributed marine picoplankton. ISME J. 10, 75-84
433   32.   Doblin, M.A. and van Sebille, E. (2016) Drift in ocean currents impacts intergenerational
434   microbial exposure to temperature. Proc. Natl. Acad. Sci. U. S. A. 113, 5700-5705
435   33.   Lohbeck, K.T. et al. (2014) Gene expression changes in the coccolithophore *Emiliania*
436   *huxleyi* after 500 generations of selection to ocean acidification. Proc. R. Soc. B 281,
437   20140003
438   34.   Perrineau, M.M. et al. (2014) Using natural selection to explore the adaptive potential of
439   *Chlamydomonas reinhardtii*. PLoS ONE 9, e92533
440   35.   Perrineau, M.M. et al. (2014) Evolution of salt tolerance in a laboratory reared population
441   of *Chlamydomonas reinhardtii*. Environ. Microbiol. 16, 1755-1766

442   36.   Mock, T. et al. (2017) Evolutionary genomics of the cold-adapted diatom *Fragilariopsis*
443   *cylindrus*. Nature, DOI:10.1038/nature20803 (ahead of print)

444   37.   Speijer, D. et al. (2015) Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic
445   life. Proc. Natl. Acad. Sci. U. S. A. 112, 8827-8834

446   38.   Bachtrog, D. et al. (2014) Sex determination: why so many ways of doing it? PLoS Biol.
447   12, e1001899

448   39.   Geng, S. et al. (2014) Evolution of sexes from an ancestral mating-type specification
449   pathway. PLoS Biol. 12, e1001904

450   40.   Hamaji, T. et al. (2016) Sequence of the *Gonium pectorale* mating locus reveals a complex
451   and dynamic history of changes in volvocine algal mating haplotypes. G3 (Bethesda) 6,
452   1179-1189

453   41.   Umen, J.G. (2011) Evolution of sex and mating loci: an expanded view from volvocine
454   algae. Curr. Opin. Microbiol. 14, 634-641

455   42.   Charlesworth, B. (1978) The population genetics of anisogamy. J. Theor. Biol. 73, 347-357

456   43.   Luthringer, R. et al. (2014) Sexual dimorphism in the brown algae. Perspect. Phycol. 1, 11-
457   25

458   44.   Ahmed, S. et al. (2014) A haploid system of sex determination in the brown alga
459   *Ectocarpus* sp. Curr. Biol. 24, 1945-1957

460   45.   Jordan, C.Y. and Charlesworth, D. (2012) The potential for sexually antagonistic
461   polymorphism in different genome regions. Evolution 66, 505-516

462   46.   Lipinska, A.P. et al. (2015) Development of PCR-based markers to determine the sex of
463   kelps. PLoS ONE 10, e0140535

464   47.   Luthringer, R. et al. (2015) The pseudoautosomal regions of the U/V sex chromosomes of
465   the brown alga *Ectocarpus* exhibit unusual features. Mol. Biol. Evol. 32, 2973-2985

466   48.   Keeling, P.J. et al. (2014) The Marine Microbial Eukaryote Transcriptome Sequencing
467   Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans
468   through transcriptome sequencing. PLoS Biol. 12, e1001889

469   49.   Kim, S.K. and Chojnacka, K., eds (2015) *Marine Algae Extracts : Processes, Products,*
470   *and Applications, Volumes 1 and 2*. Wiley-VCH Verlag GmbH & Co.

471   50.   Kazamia, E. and Smith, A.G. (2014) Assessing the environmental sustainability of
472   biofuels. Trends Plant Sci. 19, 615-618

473   51.   Rodolfi, L. et al. (2009) Microalgae for oil: strain selection, induction of lipid synthesis and
474   outdoor mass cultivation in a low-cost photobioreactor. Biotechnol. Bioeng. 102, 100-112

475   52.   Georgianna, D.R. and Mayfield, S.P. (2012) Exploiting diversity and synthetic biology for
476   the production of algal biofuels. Nature 488, 329-335

477   53.   Scott, S.A. et al. (2010) Biodiesel from algae: challenges and prospects. Curr. Opin.
478   Biotechnol. 21, 277-286

479   54.   Chen, J.E. and Smith, A.G. (2012) A look at diacylglycerol acyltransferases (DGATs) in
480   algae. J. Biotechnol. 162, 28-39

481   55.   Miller, R. et al. (2010) Changes in transcript abundance in *Chlamydomonas reinhardtii*
482   following nitrogen deprivation predict diversion of metabolism. Plant Physiol. 154, 1737-
483   1752

484   56.   Niu, Y.F. et al. (2013) Improvement of neutral lipid and polyunsaturated fatty acid
485   biosynthesis by overexpressing a type 2 diacylglycerol acyltransferase in marine diatom
486   *Phaeodactylum tricornutum*. Mar. Drugs 11, 4558-4569

487 57. Levitan, O. et al. (2015) Remodeling of intermediate metabolism in the diatom
488 *Phaeodactylum tricornutum* under nitrogen stress. Proc. Natl. Acad. Sci. U. S. A. 112, 412-
489 417
490 58. Boyle, N.R. et al. (2012) Three acyltransferases and nitrogen-responsive regulator are
491 implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*.
492 J. Biol. Chem. 287, 15811-15825
493 59. Matthijs, M. et al. (2016) Profiling of the early nitrogen stress response in the diatom
494 *Phaeodactylum tricornutum* reveals a novel family of RING-domain transcription factors.
495 Plant Physiol. 170, 489-498
496 60. Trentacoste, E.M. et al. (2013) Metabolic engineering of lipid catabolism increases
497 microalgal lipid accumulation without compromising growth. Proc. Natl. Acad. Sci. U. S.
498 A. 110, 19748-19753
499 61. Adarme-Vega, T.C. et al. (2014) Towards sustainable sources for omega-3 fatty acids
500 production. Curr. Opin. Biotechnol. 26, 14-18
501 62. Scranton, M.A. et al. (2015) *Chlamydomonas* as a model for biofuels and bio-products
502 production. Plant J. 82, 523-531
503 63. Niehaus, T.D. et al. (2011) Identification of unique mechanisms for triterpene biosynthesis
504 in *Botryococcus braunii*. Proc. Natl. Acad. Sci. U. S. A. 108, 12260-12265
505 64. Thapa, H.R. et al. (2016) A squalene synthase-like enzyme initiates production of
506 tetraterpenoid hydrocarbons in *Botryococcus braunii* Race L. Nat. Commun. 7, 11198
507 65. Jiang, Y. and Yu, D. (2016) The WRKY57 transcription factor affects the expression of
508 jasmonate ZIM-domain genes transcriptionally to compromise *Botrytis cinerea* resistance.
509 Plant Physiol. 171, 2771-2782
510 66. Viji, S. et al. (2014) Diatom-based label-free optical biosensor for biomolecules. Appl.
511 Biochem. Biotechnol. 174, 1166-1173
512 67. Ren, F. et al. (2013) Enhancing surface plasmon resonances of metallic nanoparticles by
513 diatom biosilica. Opt. Express 21, 15308-15313
514 68. Li, A. et al. (2016) Towards uniformly oriented diatom frustule monolayers: experimental
515 and theoretical analyses. Microsyst. Nanoeng. 2, 16064
516 69. Delalat, B. et al. (2015) Targeted drug delivery using genetically engineered diatom
517 biosilica. Nat. Commun. 6, 8791
518 70. Gimpel, J.A. et al. (2015) In metabolic engineering of eukaryotic microalgae: potential and
519 challenges come with great diversity. Front. Microbiol. 6, 1376
520 71. Guarnieri, M.T. and Pienkos, P.T. (2015) Algal omics: unlocking bioproduct diversity in
521 algae cell factories. Photosynth. Res. 123, 255-263
522 72. Rasala, B.A. et al. (2014) Enhanced genetic tools for engineering multigene traits into
523 green algae. PLoS ONE 9, e94028
524 73. Scaife, M.A. et al. (2015) Establishing *Chlamydomonas reinhardtii* as an industrial
525 biotechnology host. Plant J. 82, 532-546
526 74. Daboussi, F. et al. (2014) Genome engineering empowers the diatom *Phaeodactylum*
527 *tricornutum* for biotechnology. Nat. Commun. 5, 3831
528 75. Thiriet-Rupert, S. et al. (2016) Transcription factors in microalgae: genome-wide
529 prediction and comparative analysis. BMC Genomics 17, 282
530 76. Vieler, A. et al. (2012) Genome, functional gene annotation, and nuclear transformation of
531 the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. PLoS Genet. 8,
532 e1003064

533 77. Yao, L. et al. (2015) RNA-Seq transcriptomic analysis with Bag2D software identifies key
534   pathways enhancing lipid yield in a high lipid-producing mutant of the non-model green
535   alga *Dunaliella tertiolecta*. Biotechnol. Biofuels 8, 191
536 78. Scaife, M.A. and Smith, A.G. (2016) Towards developing algal synthetic biology.
537   Biochem. Soc. Trans. 44, 716-722
538 79. Adl, S.M. et al. (2005) The new higher level classification of eukaryotes with emphasis on
539   the taxonomy of protists. J. Eukaryot. Microbiol. 52, 399-451
540 80. Ball, S.G. et al. (2013) Metabolic effectors secreted by bacterial pathogens: essential
541   facilitators of plastid endosymbiosis? Plant Cell 25, 7-21
542 81. Price, D.C. et al. (2012) *Cyanophora paradoxa* genome elucidates origin of photosynthesis
543   in algae and plants. Science 335, 843-847
544 82. Yoon, H.S. et al. (2004) A molecular timeline for the origin of photosynthetic eukaryotes.
545   Mol. Biol. Evol. 21, 809-818
546 83. Marin, B. et al. (2007) The ancestor of the *Paulinella* chromatophore obtained a
547   carboxysomal operon by horizontal gene transfer from a *Nitrococcus*-like gamma-
548   proteobacterium. BMC Evol. Biol. 7, 85
549 84. Nowack, E.C.M. et al. (2008) Chromatophore genome sequence of *Paulinella* sheds light
550   on acquisition of photosynthesis by eukaryotes. Curr. Biol. 18, 410-418
551 85. Reyes-Prieto, A. et al. (2010) Differential gene retention in plastids of common recent
552   origin. Mol. Biol. Evol. 27, 1530-1537
553 86. Delaye, L. et al. (2016) How really ancient is *Paulinella chromatophora*? PLoS Currents
554   Tree of Life, March 15
555 87. Nowack, E.C.M. et al. (2016) Gene transfers from diverse bacteria compensate for
556   reductive genome evolution in the chromatophore of *Paulinella chromatophora*. Proc.
557   Natl. Acad. Sci. U. S. A. 113, 12214-12219
558 88. Bhattacharya, D. et al. (2012) Single cell genome analysis supports a link between
559   phagotrophy and primary plastid endosymbiosis. Sci. Rep. 2, 356
560 89. Matt, G. and Umen, J. (2016) *Volvox*: A simple algal model for embryogenesis,
561   morphogenesis and cellular differentiation. Dev. Biol. 419, 99-113
562 90. Cock, J.M. et al. (2010) The *Ectocarpus* genome and the independent evolution of
563   multicellularity in brown algae. Nature 465, 617-621
564 91. Cock, M.J. and Collén, J. (2015) Independent emergence of complex multicellularity in the
565   brown and red Algae. In *Evolutionary Transitions to Multicellular Life: Principles and
566   mechanisms* (Ruiz-Trillo, I. and Nedelcu, M.A., eds) pp. 335-361, Springer Netherlands
567 92. Knoll, A.H. (2011) The multiple origins of complex multicellularity. Annu. Rev. Earth
568   Planet. Sci. 39, 217-239
569 93. Niklas, K.J. and Newman, S.A. (2013) The origins of multicellular organisms. Evol. Dev.
570   15, 41-52
571 94. Ye, N. et al. (2015) *Saccharina* genomes provide novel insight into kelp biology. Nat.
572   Commun. 6, 6986
573 95. Collén, J. et al. (2013) Genome structure and metabolic features in the red seaweed
574   *Chondrus crispus* shed light on evolution of the Archaeplastida. Proc. Natl. Acad. Sci. U.
575   S. A. 110, 5247-5252
576 96. Macaisne, N. et al. (2017) The *Ectocarpus IMMEDIATE UPRIGHT* gene encodes a
577   member of a novel family of cysteine-rich proteins that have an unusual distribution across
578   the eukaryotes. Development, DOI:10.1242/dev.141523 (ahead of print)

579 97. Dillehay, T.D. et al. (2008) Monte Verde: seaweed, food, medicine, and the peopling of
580 South America. Science 320, 784-786
581 98. Tseng, C.K. and Chang, C.F. (1984) Chinese seaweeds in herbal medicine. Hydrobiologia
582 116, 152-154
583 99. FAO (2014) *The State of World Fisheries and Aquaculture*. Food and Agriculture
584 Organization of the United Nations
585 100. Duarte, C.M. et al. (2007) Rapid domestication of marine species. Science 316, 382-383
586 101. Cottier-Cook, E.J. et al. (2016) *Safeguarding the future of the global seaweed aquaculture*
587 *industry*. United Nations University (INWEH) and Scottish Association for Marine Science
588 Policy Brief
589 102. Smetacek, V. and Zingone, A. (2013) Green and golden seaweed tides on the rise. Nature
590 504, 84-88
591 103. Varela-Álvarez, E. et al. (2012) Mediterranean species of *Caulerpa* are polyploid with
592 smaller genomes in the invasive ones. PLoS ONE 7, e47728
593 104. Loureiro, R. et al. (2015) Seaweed cultivation: potential and challenges of crop
594 domestication at an unprecedented pace. New Phytol. 206, 489-492
595 105. Botebol, H. et al. (2015) Central role for ferritin in the day/night regulation of iron
596 homeostasis in marine phytoplankton. Proc. Natl. Acad. Sci. U. S. A. 112, 14652-14657
597 106. Kilian, O. et al. (2011) High-efficiency homologous recombination in the oil-producing
598 alga *Nannochloropsis* sp. Proc. Natl. Acad. Sci. U. S. A. 108, 21265-21269
599 107. Lozano, J.C. et al. (2014) Efficient gene targeting and removal of foreign DNA by
600 homologous recombination in the picoeukaryote *Ostreococcus*. Plant J. 78, 1073-1083
601 108. Minoda, A. et al. (2004) Improvement of culture conditions and evidence for nuclear
602 transformation by homologous recombination in a red alga, *Cyanidioschyzon merolae* 10D.
603 Plant Cell Physiol. 45, 667-671

604

Figure legends

606 **Figure 1**. A) The major eukaryotic supergroups (kingdoms) showing the polyphyletic origins of

607 algae in the tree of life. *Incertae sedis* (unknown affiliations) are shown with the dotted lines.

608 **Archaeplastida** (see Glossary) are shown in violet text, whereas algae with (solely, or primarily)

609 red algal-derived plastids are shown with the red text and those with green algal derived plastids

610 in green text. The plastid in photosynthetic *Paulinella* species (indicated in blue) is an example

611 of an independent cyanobacterial primary endosymbiosis. Image adapted from Kim et al. [20].

612 B) Moving clockwise from top left: a *Sargassum fusiforme* (stramenopile) farm in South Korea;

613 the invasive green alga (Viridiplantae) *Caulerpa taxifolia*; *Saccharina japonica* (stramenopile);

614 an underwater kelp 'forest' of *Undaria pinnatifida* (stramenopile); (top) dividing cells of

615 *Paulinella microporus* KR01 (Rhizaria; image prepared by Jong Im Kim); (bottom) light

616 microscope image showing the single sheet of cells that comprises the *Porphyra umbilicalis*

617   (Rhodophyta) blade; an underwater tropical forest of the stramenopile *Sargassum* (credit: R.

618   Terada); (below) examples of harmful algal blooms and their biotic interactions. Shown are two

619   unreported protistan parasites infecting the toxic diatom (stramenopile) *Pseudo-Nitszchia* sp.

620   (scale bars = 10 µm; images prepared by Andrea Garvetto). Images of seaweed farms were

621   provided by the National Institute of Fisheries Science in South Korea. C) Major sources of

622   foreign genes in the nuclear genome of the Archaeplastida ancestor that underwent primary

623   plastid endosymbiosis. EGT refers to genes derived via intracellular gene transfer from the

624   cyanobacterial endosymbiont. HGT refers to a variety of genes derived from multiple non-plastid

625   sources, including symbionts, prey, viruses (represented as the blue and magenta ovals), or other

626   sources of DNA that were present in the cell or entered it from the environment.

627   **Figure 2.** A synthetic biology approach to engineering microalgae. Using defined parts (e.g.,

628   promoters, terminators) with predictable behavior, and a standardized way to combine them,

629   many permutations can be tested for optimal expression of the gene of interest through the

630   Design-Build-Test-Learn cycle [78].

631

632   **BOX 1.** *Algae in the tree of life and the evolution of multicellularity*

633   **Primary endosymbiosis**

634   The primary plastid characteristic of the Archaeplastida (land plants, glaucophyte, green and red

635   algae) can be traced back to a single primary endosymbiosis of a cyanobacterium that occurred

636   in the Archaeplastida common ancestor [79-81] about 1.6 billion years ago [82]. Thereafter, the

637   plastid in red and green algae spread to other taxa through secondary or additional rounds of

638   eukaryote-eukaryote endosymbioses (Fig. 1A).

639

640   *Paulinella* **primary endosymbiosis**

641   There is only one other known example of a primary plastid endosymbiosis: in the

642   photosynthetic *Paulinella* species [83-85]. This plastid (known as a chromatophore) originated

643   90-140 million years ago [86] and is derived from an α-cyanobacterium with Form IA Rubisco,

644   unlike the Form IB Rubisco derived by EGT from the β-cyanobacterial ancestor of the

645   archaeplastidial plastid [13]. The chromatophore genome has undergone genome reduction [87]

646   but is ca. 1 Mbp in size, which is about 5-10 times larger than other plastid genomes. Genomic

647 and transcriptomic data demonstrate the existence of extensive horizontal gene transfer (HGT)

648 from non-endosymbiont bacterial sources (Fig. 1C). Many of these nuclear HGTs complement

649 functions lost in the chromatophore due to genome reduction [87]. This result suggests that

650 phagotrophy, which typified the heterotrophic ancestor of photosynthetic *Paulinella* species [88]

651 was crucial during the early phases of primary endosymbiosis to allow gene acquisition.

652

653 **Evolution of multicellularity**

654 Algae play an important role in improving our understanding of multicellularity. In particular,

655 the volvocine algae represent one of the best model systems for understanding the key initial

656 transition from unicellularity to multicellularity [89]. These organisms have provided several

657 compelling examples of how genes that were already present in the unicellular ancestor were co-

658 opted for functions related to multicellularity.

659   Transitions from unicellularity to multicellularity have occurred many times but only a

660 limited number of groups have evolved complex multicellularity. Animals, plants, fungi and

661 brown and red seaweeds, which all have organized macroscopic body plans consisting of

662 multiple cell types are generally considered to have made this transition [90-93]. Complex

663 multicellularity evolved independently in each of these groups providing an ideal situation to

664 apply comparative approaches to understand this important evolutionary process. Developmental

665 processes are well understood in land plants, however the two macroalgal groups remain poorly

666 studied at the molecular level. The availability of genome sequence information from both brown

667 [90, 94] and red macroalgae [95] has allowed some comparative analyses to be done, but

668 experimental investigation of the molecular basis of developmental processes in these algae is

669 essential. The recent demonstration that key developmental genes can be identified in the brown

670 algal model *Ectocarpus* using a forward genetic approach [96] represents a first step towards the

671 emergence of experimental macroalgal developmental biology.

672

673 **BOX 2.** *Algae in human affairs*

674 Human consumption of seaweed dates back at least 14,000 years in South America [97], with

675 medicinal uses recorded in Chinese and Japanese literature from 1500 to 2000 years ago [98].

676 Algae are increasingly used in an array of manufactured foods and feed, as a source for

677 hydrocolloids of commercial importance and in food processing, pharmaceutical, cosmetic,

678   biofuels and biomaterials industries. With a value of $5.6 billion USD in 2014, the seaweed

679   market is currently the fastest growing aquaculture sector (8% year^{-1}) and represents 25% of

680   global aquaculture production (27 million tonnes [Mt] in 2014) [99]. Five major groups of algae

681   contribute to 97% of production: eucheumatoids (*Eucheuma* spp., *Kappaphycus*) accounted for

682   about 11 Mt, the kelp *Saccharina japonica* for 7.6 Mt, the red alga *Gracilaria* spp. for 3.7 Mt,

683   wakame *Undaria pinnatifida* for 2.4 Mt, and laver (*Porphyra* spp. and *Pyropia*) for 1.8 Mts.

684   Major seaweed production areas include China (49%), Indonesia (37%), the Philippines (6%),

685   and South Korea (4%), followed by Malaysia, North Korea, and Japan. However, cultivation of

686   macro- and microalgae is developing quickly on all continents and consequently the number of

687   farmed species is rapidly rising. This domestication follows a similar trend for animal marine

688   species [100], and poses significant challenges to the sustainable exploitation and conservation

689   of marine genetic resources [101]. Some microalgae form blooms that are toxic for other

690   organisms, in particular fish and humans, and uncontrolled growth of seaweeds (e.g., *Ulva* spp.)

691   as a response to eutrophication along coastlines worldwide is a major cause of reduced

692   ecosystem functioning [102]. In addition, non-native seaweeds can cause dramatic ecological

693   shifts in marine ecosystems, for example the green seaweed *Caulerpa* spp. in the Mediterranean

694   Sea [103].

695

696   **Future domestication of algae**

697   Sustainable exploitation and domestication of algae will require greatly improved knowledge of

698   the ecological and molecular factors controlling growth and reproduction [104]. For general

699   applications, such as lipids for biofuel, or higher-value compounds for a wider range of

700   nutraceuticals, therapeutics or even bulk chemicals, microalgal production needs to be

701   reproducible at scale under ambient conditions [52]. Similarly, for seaweeds, the challenge is to

702   balance growth and reproduction with culture loss through diseases and pests. Identification of

703   ecotypic genetic variation and a comprehensive understanding of molecular mechanisms

704   controlling such intraspecific variation (e.g., quantitative trait loci), and the introduction of

705   agrigenomics tools, will be instrumental to facilitate successful algal domestication.

706

707   **BOX 3.** *New tools for functional analysis in algae*

708   Genome-enabled studies have revealed the significance of specific transcription factors, small

709   RNAs, epigenetics, and transposable elements in algal plasticity. Linking hypotheses generated

710   from genomic analyses to specific activities/phenotypes is facilitating the development of a new

711   range of tools.

712

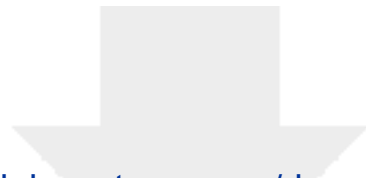713   **CRISPR/Cas9: RNA-guided gene-editing tool**

714   CRISPR/Cas9 gene editing is currently available for the algae *C. reinhardtii*, *Phaeodactylum*

715   *tricornutum*, *Nannochloropsis* sp. and *Thalassiosira pseudonana* [9-12] and high frequency gene

716   targeting by homologous recombination has been reported for the haploid genomes of

717   *Cyanidioschizon merolae*, *Ostreococcus tauri*, and *Nannochloropsis* sp. [105-108]. These tools,

718   because they can target genes or loci precisely, will enable insight into how (epi) genes impact

719   phenotypic expression in algae. Furthermore, the role of gene redundancy (e.g., isoforms from

720   different endosymbionts) can be dissected either by knocking out specific family members, or

721   increasing dosage by adding various synthetic isoforms. This approach would shed light on how

722   different endosymbionts and their gene repertoires contribute to the adaptation and evolution of

723   extant algae in addition to the functionality associated with genome plasticity. Genome editing

724   with CRISPR/Cas9 can target more than a single gene/locus at a time. Thus by simultaneously

725   editing a large number of nuclear genes encoding plastid proteins the contributions of "green" vs.

726   "red" genes for key metabolic processes such as carbon fixation, lipid and polysaccharide

727   synthesis/storage, and macronutrient uptake/assimilation could be determined.

Trends

- Application of modern omic and genetic methods has significantly advanced our understanding of the origin, evolution, and metabolic potential of unicellular and multicellular algae as well as their diverse modes of sexual reproduction.

- The GreenCut proteins, a conserved gene set in the Viridiplantae are primarily plastid targeted and play key roles in the function and regulation of photosynthesis, including the maintenance of photosynthetic reaction complexes.

- Lab evolution experiments demonstrate strong adaptability of microalgae to environmental changes that are associated with climate change, although it is unclear if these results will hold in natural ecosystems.

- Development of algae as 'cell factories' promises to allow the production not only of endogenous molecules but also non-native compounds useful in therapeutics and in the production of plastics.
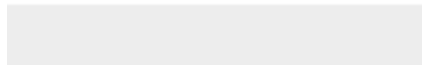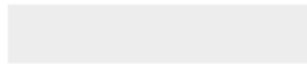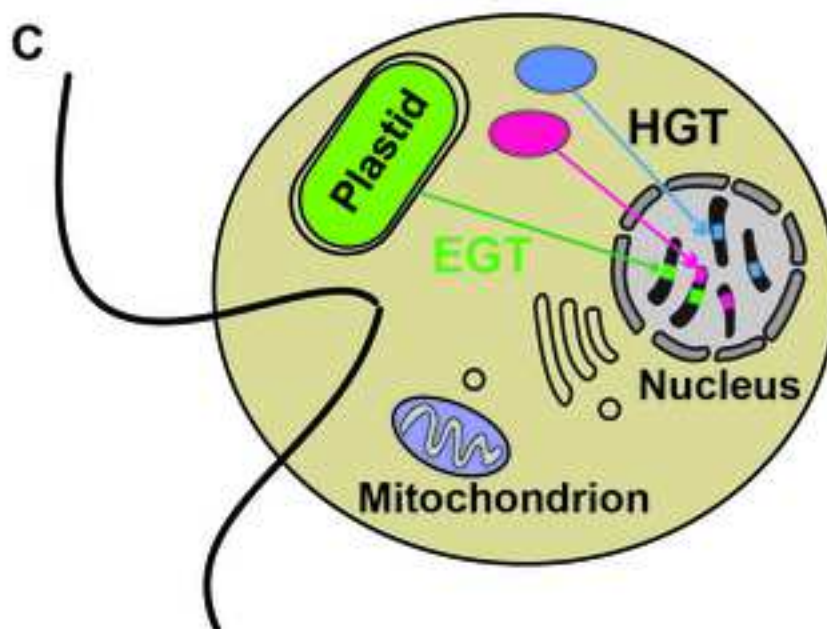
Outstanding Questions

- What role did EGT play in expanding the genetic inventory of algae and plants and how can we determine the roles of these foreign genes in extant cells?

- What do the GreenCut proteins teach us about how ancient Viridiplantae dealt with the rise in atmospheric oxygen precipitated by the great oxygenation event, the subsequent spread of eukaryotic photosynthetic lineages, and the stresses posed by hosting an oxygen-evolving organelle?

- Will the fluctuating environmental conditions forecasted for the coming century impact the health and distribution of aquatic microalgae, and which types of acclimatory and adaptive mechanisms do these taxa possess to deal with increasing temperatures and lower pH levels?

- What is the basis of sex determination in algae and seaweeds and do these mechanisms, such as the existence of sex-determining regions or sex chromosomes, follow the same patterns of evolution as described in well-studied, classical eukaryotic model systems?

- What are the unique and most promising aspects of, and the constraints on using algae as 'cell factories' and what types of advances in genetic manipulation techniques and genomic methods are needed to make these taxa more useful to the biotechnology industry?
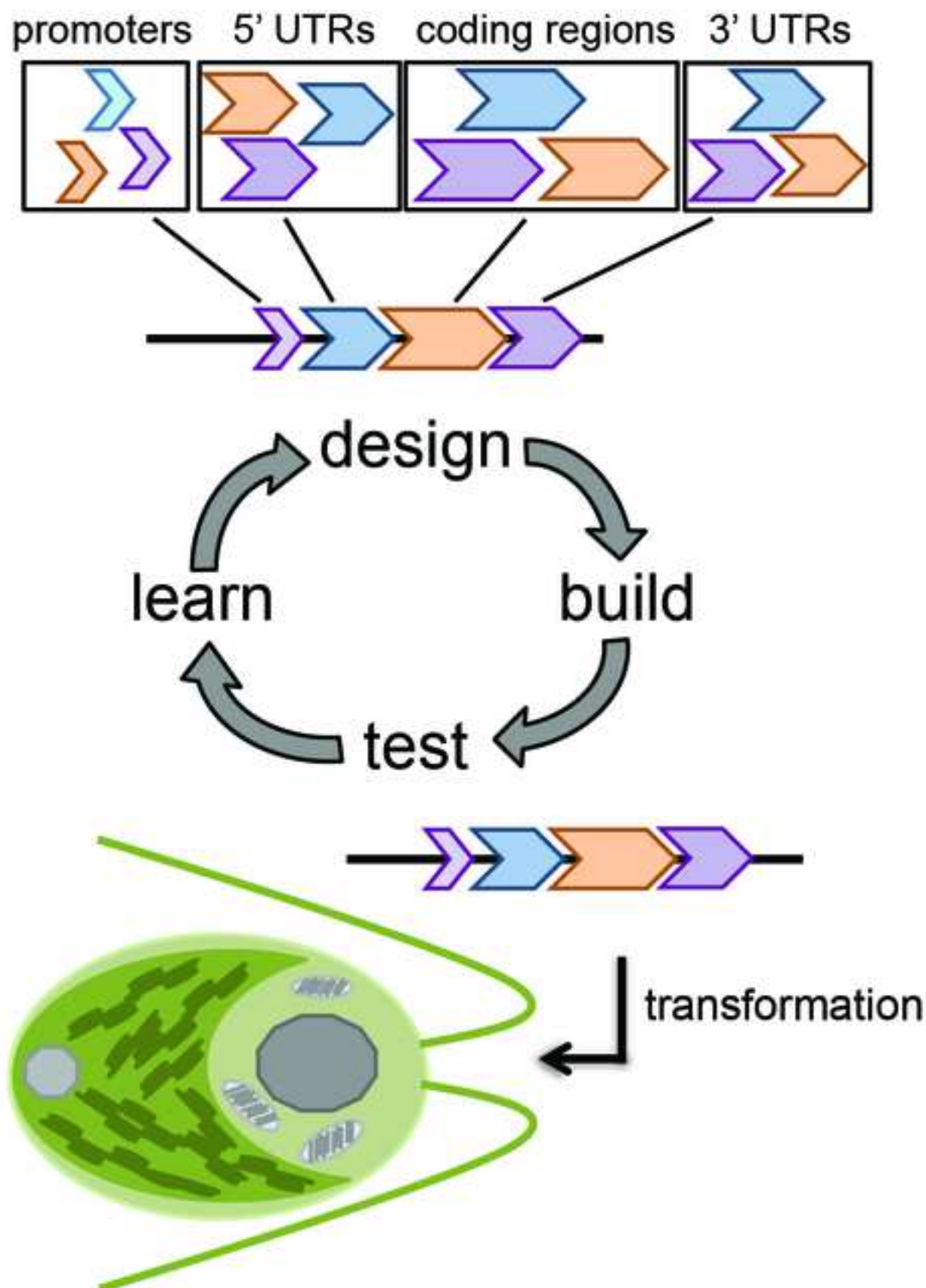
Click here to access/download
Author Supplementary Material
Glossary_TIPS.docx

A

Excavata
Metamonada
-Fornicata
-Parabasalia
-Preaxostyla
Malawimonas
Discoba
-Jakobida
-Discicristata
-Euglenozoa
(Euglenids)
Kinetoplastids)

Archaeplastida
Glaucophyta
Rhodophyta/
Viridiplantae

**Stramenopiles/
Alveolates/
Rhizaria (SAR)**
Heterokonta (stramenopiles)
Alveolata
(Dinoflagellates, Chromerida,
Apicomplexa, Ciliates)
Cercozoa
Chlorarachniophyta

Foraminifera, Radiolaria

Opisthokonta
Animals
Choanoflagellata
Ichthyosporea
Fungi
Nucleariidae

Amoebozoa
Dictyostelia
Tubulinea
Discosea
Archamoebae

Incertae sedis
Cryptophyta, Katablepharids
Haptophyta, Telonemia
Rappemonads
Picozoa, Centrohelida
Ancyromonadida
Apusomonadidae
Breviatea

B

C

Plastid

HGT

EGT

Nucleus

Mitochondrion

Click here to access/download

Interactive Questions

iq.pdf

Figure360

Click here to access/download
Figure360
TIPS_figure360_final_reduced.mov