Classification: BIOLOGICAL SCIENCES (Ecology)

Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics

Christopher J. Gobler[1,2,*], Dianna L. Berry[1,2, †], Sonya T. Dyhrman[3, †], Steven W. Wilhelm[4, †], Asaf Salamov[5], Alexei V. Lobanov[6], Yan Zhang[6], Jackie L. Collier[2], Louie L. Wurch[3], Adam B. Kustka[7], Brian D. Dill[8], Manesh Shah[9], Nathan C. VerBerkmoes[8], Alan Kuo[5], Astrid Terry[5], Jasmyn Pangilinan[5], Erika Lindquist[5], Susan Lucas[5], Ian Paulsen[10], Theresa K. Hattenrath[1,2], Stephanie C. Talmage[1,2], Elyse A. Walker[1,2], Florian Koch[1,2], Amanda M. Burson[1,2], Maria Alejandra Marcoval[1,2], Ying-Zhong Tang[1,2], Gary R. LeCleir[3], Kathryn J. Coyne[11], Gry Mine Berg[12], Erin M. Bertrand[13], Mak A. Saito[13, 14], Vadim Gladyshev[5], Igor V. Grigoriev[4,*]

[1]School of Marine and Atmospheric Sciences, Stony Brook University, Southampton, NY 11968, USA. [2]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY 11794-5000, USA. [3]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA. [4]Department of Microbiology, The University of Tennessee, Knoxville, TN 37996, USA. [5]US Department of Energy, Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA. [6]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston MA 02115, USA. [7]Department of Earth and Environmental Sciences, Rutgers University, Newark, New Jersey 07102, USA, [8]Chemical Sciences and [9]Biosciences Divisions, Oak Ridge National Laboratory, Oak Ridge, TN, 37830, USA. [10]Department of Chemistry and Biomolecular Sciences, Macquarie University Sydney, 2109, NSW, Australia. [11]College of Earth, Ocean, and Environment, University of Delaware,

Lewes, DE, 19958 USA.[12]Department of Environmental Earth System Science, Stanford University, 397 Panama Mall, Stanford, California 94305, USA. [13]Massachusetts Institute of Technology and Woods Hole Oceanographic Institution Joint Program in Chemical Oceanography. [14]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA.

*To whom correspondence should be addressed. E-mail: christopher.gobler@stonybrook.edu or IVGrigoriev@lbl.gov

†These authors contributed equally to this work.

Keywords: Harmful algal blooms, HABs, genome sequence, ecogenomics, metaproteomics, eutrophication, *Aureococcus anophagefferens*,

Harmful algal blooms (HABs) cause significant economic and ecological damage worldwide. Despite considerable efforts, a comprehensive understanding of the factors that promote these blooms has been lacking because the biochemical pathways that facilitate their dominance relative to other phytoplankton within specific environments have not been identified. Here, biogeochemical measurements demonstrated that the harmful alga *Aureococcus anophagefferens* outcompeted co-occurring phytoplankton in estuaries with elevated levels of dissolved organic matter and turbidity and low levels of dissolved inorganic nitrogen. We subsequently sequenced the first HAB genome (*A. anophagefferens*) and compared its gene complement to those of six competing phytoplankton species identified via metaproteomics. Using an ecogenomic approach, we specifically focused on the gene sets that may facilitate dominance within the environmental conditions present during blooms. *A. anophagefferens* possesses a larger genome (56 mbp) and more genes involved in light harvesting, organic carbon and nitrogen utilization, and encoding selenium- and metal-requiring enzymes than competing phytoplankton. Genes for the synthesis of microbial deterrents likely permit the proliferation of this species with reduced mortality losses during blooms. Collectively, these findings suggest that anthropogenic activities resulting in elevated levels of turbidity, organic matter, and metals have opened a niche within coastal ecosystems that ideally suits the unique genetic capacity of *A. anophagefferens* and thus has facilitated the proliferation of this and potentially other HABs.

\body

Harmful algal blooms (HABs) are caused by phytoplankton that have a negative impact on ecosystems and coastal fisheries world-wide (1 – 4) and cost the US economy alone hundreds of millions of dollars annually (5). The frequency and impacts of HABs have intensified in recent decades and anthropogenic processes including eutrophication have been implicated in this expansion (1 - 3). While there is great interest in mitigating the occurrence of HABs, traditional approaches which have characterized biogeochemical conditions present during blooms do not identify the aspects of the environment which are favorable to an individual algal species. Predicting where, when, and under what environmental conditions HABs will occur has further been inhibited by a limited understanding of the cellular attributes that facilitate the proliferation of one phytoplankton species to the exclusion of others.

Aureococcus anophagefferens is a pelagophyte that causes harmful brown tide blooms with densities exceeding 10^6 cells mL^{-1} for extended periods in estuaries in the eastern US and in South Africa (6). Brown tides do not produce toxins that poison humans, but have decimated multiple fisheries and seagrass beds due to toxicity to bivalves and extreme light attenuation, respectively (6). Brown tides are a prime example of the global expansion of HABs as these blooms had never been documented prior to 1985, but have recurred in the US and South Africa annually since then (6). Like many other HABs, *A. anophagefferens* blooms in shallow, anthropogenically modified estuaries when levels of light and inorganic nutrients are low and organic carbon and nitrogen concentrations are elevated (1 - 3).

For this study, we utilized a novel ecogenomic approach to assess the extent to which the gene set of *A. anophagefferens* may permit its dominance under the environmental conditions present in estuaries during brown tides. We characterized the biogeochemical conditions present

in estuaries before, during, and after *A. anophagefferens* blooms. Sequencing the first HAB genome (*A. anophagefferens*), we compared its genome to those of six phytoplankton species identified via metaproteomics to co-occur with this alga during blooms events. Using this ecogenomic approach, we investigated how the gene sets of *A. anophagefferens* differ from the six comparative phytoplankton species, and how these differences may affect the ability of *A. anophagefferens* to compete in the physical (*e.g.* light harvesting), chemical (*e.g.* nutrients, organic matter, trace metals), and ecological (*e.g.* defense against predators and allelopathy) environment present during brown tides.

Results and Discussion

During an investigation of an US estuary, Quantuck Bay, NY, from 2007 through 2009, brown tides occurred annually from May through July, achieving abundances exceeding 10^6 cells mL^{-1} or 5 x 10^6 μm^3 mL^{-1} (Fig 1). *A. anophagefferens* was observed to bloom after spring diatom blooms and outcompeted small (< 2 µm) eukaryotic and prokaryotic phytoplankton (e.g. *Ostreococcus* and *Synechococcus*) during summer months (Fig 1D), a pattern consistent with prior observations (7, 8). Concurrently, dissolved inorganic nitrogen levels were reduced to < 1 µM during blooms while dissolved organic nitrogen levels and light extinction were elevated resulting in a system with decreased light availability and concentrations of dissolved organic nitrogen exceeding those of dissolved inorganic nitrogen (Fig. 1C). Metaproteomic analyses of planktonic communities were performed to identify phytoplankton that *A. anophagefferens* may compete with during blooms by quantifying organism-specific peptides among the microbial community. Performing such analyses on the plankton present in this estuary highlight the dominance of *A. anophagefferens* and co-existence of the six phytoplankton species for which

complete genome sequences have been generated (Fig 1E): two coastal diatom species, *Phaeodactylum tricornutum* (clone CCMP632) (9) and *Thalassiosira pseudonana* (clone CCMP 1335 (10) isolated from an embayment that now hosts brown tides (6)), and coastal zone isolates of *Ostreococcus* (*O. lucimarinus* and *O. tauri* (11)) and *Synechococcus* (clones CC9311 (12) and CC9902), small eukaryotic and prokaryotic phytoplankton, respectively, (Table 1, Fig 1). To assess the extent to which the gene set of *A. anophagefferens* may permit its dominance within the geochemical environment found in this estuary (Fig. 1C), the gene complement of *A. anophagefferens* was determined by genome sequencing and was compared to those of the six competing phytoplankton species (Table 1, Fig 1E).

Although phytoplankton genome size generally scales with cell size (15,16), *A. anophagefferens* (2 µm) has a larger genome (56 Mbp) and more genes (~11,500) than the six competing phytoplankton species (2.2 – 32 Mbp; 2,301 - 11,242 genes; Table 1 and Tables S1 to S4). Its small cell size and thus larger surface area to volume ratio allows it to kinetically outcompete larger phytoplankton for low levels of light and nutrients (17) while its large gene content and more complex genetic repertoire may provide a competitive advantage over other small phytoplankton with fewer genes. The *A. anophagefferens* genome contains the largest number of unique genes relative to the six competing phytoplankton examined here (209 v. 12 - 79 unique genes; Table 1). Many of these enriched or unique genes are associated with light harvesting, organic matter utilization, and metalloenzymes, as well as the synthesis of microbial predation and competition deterrents (Supplementary Tables S5-S17). These enriched and unique gene sets are involved in biochemical pathways related to the environmental conditions prevailing during brown tides (Fig. 1), and thus are likely to facilitate the dominance of this alga during chronic blooms that plague estuarine waters.

Light harvesting - Phytoplankton rely on light to photosynthetically fix carbon dioxide into organic carbon, but the turbid, low light environment characteristic of estuaries and intense shading during dense algal blooms (Fig. 1B,C) can strongly limit photosynthesis. *A. anophagefferens* is better adapted to low light than the comparative phytoplankton species that require at least three-fold higher light levels to achieve maximal growth rates (Fig. 2A). Its genome contains the full suite of genes involved in photosynthesis, including 62 genes encoding light harvesting complex (LHC) proteins (Fig. 2A). This is 1.5- to 3-times more than other eukaryotic phytoplankton sequenced thus far (Fig. 2A and Table S7) and a feature that likely enhances adaptation to low and/or dynamic light conditions found in turbid estuaries. LHC proteins bind antenna chlorophyll and carotenoid pigments that augment the light capturing capacity of the photosynthetic reaction centers (18,19). Twenty-six *A. anophagefferens* LHC genes belong to a group that has only six representatives in *T. pseudonana* and one in *P. tricornutum* (branch 'PHYMKG' in Fig. 3 and Fig. S1), but are similar to the multi-cellular brown macroalgae, *Ectocarpus siliculosus* (20). Similar LHC genes in the microalgae *Emiliania huxleyi* have recently been shown to be up-regulated under low light (21). We hypothesize that these LHC genes encode the major light harvesting proteins for *A. anophagefferens*, and that the enrichment of these proteins impart a competitive advantage in acquiring light under the low irradiance conditions that prevail during blooms (Fig 1C).

Organic matter utilization - In addition to being well adapted to low light, *A. anophagefferens* also outcompetes other phytoplankton in estuaries with elevated organic matter concentrations (6) (Fig 1C), and can survive extended periods with no light (22). Consistent with these observations, the genome of *A. anophagefferens* contains a large number of genes that may permit the degradation of organic compounds to support heterotrophic metabolism. For

example, its genome encodes proteins involved in the transport of oligosaccharides and sugars that are not found in competing phytoplankton, including genes for glycerol, glucose, and D-xylose uptake (Table S8). The *A. anophagefferens* genome also encodes more nucleoside sugar transporters and major facilitator family sugar transporters than other comparative phytoplankton species (Table S8). It is also highly enriched in genes associated with the degradation of mono-, di-, oligo- and polysaccharides, as well as sulfonated polysaccharides. *A. anophagefferens* possesses 47 sufatase genes including those targeting sulfonated polysaccharides such as glucosamine-(N-acetyl)-6-sulfatases, while the diatoms contain a total of 3 to 4 sulfatases, and the comparative picoplankton contain none (Table S9). *A. anophagefferens* also possesses many more genes involved in carbohydrate degradation than competing phytoplankton (85 v. 4 – 29 genes in comparative phytoplankton) including 29 such genes present only in *A. anophagefferens* (Fig. 4 and Tables S10 and S11). Collectively, these genes (Tables S9 to S12) provide this alga with unique metabolic capabilities regarding the degradation of an array of organic carbon compounds, many of which may not be accessible to other phytoplankton. In an ecosystem setting, such a supplement of organic carbon would be critical for population proliferation within the low light environments present in estuaries, particularly during dense algal blooms (Fig. 1C).

A. anophagefferens, like many HABs, blooms when inorganic nitrogen levels are low but organic nitrogen levels are elevated (Fig. 1C) (1 - 3), *A. anophagefferens* is known to efficiently metabolize organic compounds for nitrogenous nutrition (6, 23). Notably, this niche strategy is reflected within the *A. anophagefferens* genome which encodes transporters specific for a diverse set of organic nitrogen compounds including urea, amino acids, purines, nucleotide-sugars, nucleosides, peptides, and oligopeptides (Table S8) (24). Relative to competing phytoplankton, *A. anophagefferens* is enriched in genes encoding enzymes that degrade organic nitrogen

compounds such as nitriles, asparagine, and urea (Fig. 2B). *A. anophagefferens* is also the only species among the phytoplankton genomes examined that possesses a membrane-bound dipeptidase, several histidine ammonia-lyases, cysteine dioxygenase, tripeptidyl peptidase, and several other enzymes (Table S13) that could collectively play a role in metabolizing organic nitrogen compounds that are not bioavailable to other phytoplankton. Furthermore, the *A. anophagefferens* genome also contains enzymes that degrade amino acids, peptides, proteins, amides, amides, and nucleotides, often possessing more copies of these genes than competing phytoplankton (Supplementary Table 13). This characteristic, along with its unique gene set, may provide *A. anophagefferens* with a greater capacity to use organic compounds for nitrogenous nutrition compared to its competitors, a hypothesis supported by its dominance in systems with elevated ratios of dissolved organic nitrogen to dissolved inorganic nitrogen and the reduction in dissolved organic nitrogen concentrations typically observed during the initiation of brown tides (6, 25).

Metalloenzymes - *A. anophagefferens* blooms in shallow, enclosed estuaries (6) where concentration of metals and elements like selenium are elevated (26 – 28), but never dominates deep estuaries or continental shelf regions (6) that are characterized by lower metal and trace element inventories (26 - 28). *A. anophagefferens* has a large and absolute requirement for some trace elements, such as selenium (Fig. 2C). In comparison, phytoplankton such as *Synechococcus* do not require this element while others, such as *T. pseudonana* and *P. tricornutum*, have lower selenium requirements for maximal growth (Fig. 2C). The *A. anophagefferens* genome is consistent with these observations as it is enriched in numerous classes of proteins that require metals and elements like selenium as cofactors (Fig. 2C). It possesses at least 56 genes encoding selenocysteine-containing proteins, twice the number

present in *O. lucimarinus* genome, which previously had the largest known eukaryotic selenoproteome (11, 29), and four-fold more than the diatom genomes (Fig. 2C). The *A. anophagefferens* selenoproteome includes nearly all known eukaryotic selenoproteins, as well as selenoproteins that were previously described only in bacteria (29) and several novel selenoproteins (Table S14). In addition, several selenoprotein families are represented by multiple isozymes (Table S14). Half of the selenoproteins are methionine sulfoxide reductases, thioredoxin reductases, glutathione peroxidases, glutaredoxins, and peroxiredoxins (Table S14). Together, these enzymes help protect cells against oxidative stress in the dynamic and ephemeral conditions present in estuaries through the removal of hydroperoxides and the repair of oxidatively damaged proteins. Moreover, selenocysteine residues are often superior catalytic groups compared to cysteine (30 - 32), and thus allow *A. anophagefferens* to more efficiently execute multiple metabolic processes and increase its competitiveness relative to other phytoplankton in the anthropogenically modified estuaries where it blooms.

The *A. anophagefferens* genome is also enriched in genes encoding for molybdenum-, copper-, and nickel-containing enzymes (Fig. 2C). For example, the *A. anophagefferens* genome includes twice the number of genes encoding molybdenum-containing oxidases found in competing species (6 v. 1 – 3 genes; Fig. 2C and Tables S15 and S16), and has the largest number of molybdenum-specific transporters (Table S8). Similarly, *A. anophagefferens* possesses four-times more genes that encode copper-containing proteins than its competitors (27 v. 1 – 6 genes; Fig. 2C), including five multi-copper oxidases and 20 tyrosinase-like proteins (Tables S15 and S16). Several of the *A. anophagefferens* tyrosinase and multi-copper oxidase family proteins are heavily glycosylated (>4 glycosylation sites; Table S16) and thus are likely secretory proteins, while the few present in the other comparative algal species are not. These

copper containing enzymes degrade lignin, catalyze the oxidation of phenolics, and can have anti-microbial properties (33, 34) and thus may provide nutrition or confer protection to *A. anophagefferens* cells. *A. anophagefferens* is also the only phytoplankton species with a homolog of the CutC copper homeostasis protein, which permits efficient cellular trafficking of this metal (Table S8). With three nickel-requiring ureases, *A. anophagefferens* has more nickel-containing enzymes than other comparative phytoplankton (Fig. 2B, C). Consistent with its ecogenomic profile, these ureases allow *A. anophagefferens* to meet its daily N demand from urea while other phytoplankton do not (35). Perhaps to support the synthesis and use of urease, *A. anophagefferens* is the only comparative phytoplankton species with a high-affinity nickel transporter (HoxN) (36). *A. anophagefferens* is not universally enriched in metalloenzymes, as other phytoplankton contain equal numbers of cobalt-containing enzymes (Fig. 2C). However, the formation of blooms exclusively in shallow estuaries ensures *A. anophagefferens* has access to a rich supply of the selenium, copper, and nickel required to synthesize these ecologically important and catalytically superior enzymes (30, 31, 37).

Microbial defense - While genes associated with the adaptation to low light, the use of organic matter, and metals permit *A. anophagefferens* to dominate a specific geochemical niche found within estuaries, genes involved in the production of compounds that inhibit predators and competitors may further promote blooms (2). Although specific toxins have yet to be identified in *A. anophagefferens*, it is grazed at a low rate during blooms (2, 6) and its genome contains two- to seven-times more genes involved in the synthesis of secondary metabolites than the comparative phytoplankton genomes (Fig. S2). *A. anophagefferens* also possesses a series of genes involved in the synthesis of putative anti-microbial compounds that are largely absent from the competing phytoplankton species (Table S17). For example, *A. anophagefferens* has

five berberine bridge enzymes involved in the synthesis of toxic isoquinoline alkaloids (38, 39) (Table S17). *A. anophagefferens* uniquely possesses a membrane attack complex gene and multiple phenazine biosynthetase genes (Table S17) that encode enzymes that may provide defense against microbes and/or protistan grazers (40, 41). There are two- to four-fold more ATP-binding cassette (ABC) transporters in *A. anophagefferens* compared to competing species (112 v. 30 – 54 ABC transporters; Table S8) and it is specifically enriched in ABC multidrug efflux pumps ((P-glycoprotein (ABCB1), MRP1 (ABCC1) and ABCG2 (BCRP)) that protect cells from toxic xenobiotics and endogenous metabolites (42, 43). Finally, the *A. anophagefferens* genome encodes 16-fold more Sel-1 genes (130 v. 0- 8 genes; Table S6), four-fold more ion channels (82 v. 1- 19 ion channels; Table S8), four-fold more protein kinases, and two-fold more WD40 domain genes than other phytoplankton (Table S6). These genes may collectively mediate elaborate cell signaling and sensing by dense bloom populations (44 - 46), processes which would be important for detecting competitors, predators, other *A. anophagefferens* cells, and the environment. Together, genes involved in the synthesis of microbial deterrents, export of toxic compounds, and cell signaling may contribute toward the proliferation of this species with reduced population losses and thus assist in promoting these HABs (2).

Conclusions - The global expansion of human populations along coastlines has led to a progressive enrichment in turbidity (47), organic matter including organic nitrogen (1, 47, 48), and metals (26, 28) in estuaries. Matching the expansion of HAB events around the world in recent decades, *A. anophagefferens* blooms were an unknown phenomenon prior to 1985, but have since become chronic, annual events in US and South African estuaries (6) with the potential for further expansion. The unique gene complement of *A. anophagefferens* encodes a

disproportionately greater number of proteins involved in light harvesting and organic matter utilization, as well as metal and selenium-requiring enzymes relative to competing phytoplankton. Collectively, these genes reveal a niche characterized by conditions (low light, high organic matter, and elevated metal levels) that have become increasingly prevalent in anthropogenically-modified estuaries, suggesting that human activities have enabled the proliferation of these HABs. In estuaries which host *A. anophagefferens* blooms, anthropogenically nutrient loading promotes algal growth and, as a result, elevated levels of organic matter and turbidity (6) whereas high concentrations of metals have been attributed to maritime paints and some fertilizers (27, 49). Collectively, these findings establish a context within which to prevent and control HABs specifically by ameliorating anthropogenically altered aspects of marine environments that harmful phytoplankton are genomically pre-disposed to exploit. Like *A. anophagefferens*, many HAB-forming dinoflagellates are known to exploit organic forms of carbon and nitrogen for growth (1 - 4), grow well under low light (45), and have elevated requirements of copper, molybdenum, and selenium (46, 47). Continued ecogenomic analyses of HABs will reveal the extent to which these events can be attributed to human activities that have transformed coastal ecosystems to suit the genetic capacity of these algae.

Materials and Methods

The environmental conditions and plankton community composition within a brown tide-prone estuary (Quantuck Bay, NY, USA) were monitored biweekly from spring through fall of 2009. Nutrient levels were assessed via wet chemical and combustion techniques, whereas the composition of the plankton community was assessed via immuno-fluorescent assays, flow

cytometry, and standard microscopy. Metaproteomes were generated using two-dimensional, nano-liquid chromatography – tandem mass spectrometry (LC-MS/MS) and spectra were analyzed using SEQUEST and DTASelect algorithms. The genome of *A. anophagefferens* was sequenced using whole-genome shotgun approach using Sanger platform, assembled with JAZZ assembler, and annotated using JGI Annotation tools. Complete information regarding all methods used for all analyses reported here is available in Supplementary Information.

Acknowledgements: Genome sequencing, annotation, and analysis were conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Efforts were also supported by awards from New York Sea Grant to Stony Brook University, National Oceanic and Atmospheric Administration Center for Sponsored Coastal Ocean Research award #NA09NOS4780206 to Woods Hole Oceanographic Institution, NIH grant GM061603 to Harvard University, and NSF award IOS-0841918 to The University of Tennessee. Assembly and annotations of *Aureococcus anophagefferens* are available from JGI Genome Portal at http://www.jgi.doe.gov/Aureococcus and were deposited at DDBJ/EMBL/GenBank under the project accessions (ACJI00000000), respectively.

References

1. Heisler J, Glibert PM, Burkholder JM, Anderson DM, Cochlan W, Dennison WC, Dortch Q, Gobler CJ, Heil CA, Humphries E, Lewitus A, Magnien R, Marshall HG, Sellner K, Stockwell DA, Stoecker DK, Suddleson M (2008) Eutrophication and harmful algal blooms: A scientific consensus. *Harmful Algae* 8:3-13.
2. Sunda WG, Graneli E, Gobler CJ (2006) Positive feedback and the development and persistence of ecosystem disruptive algal blooms. *J Phycol* 42: 963-974.
3. Anderson DM, Burkholder JM, Cochlan WP, Glibert PM, Gobler CJ, Heil CA, Kudela R, Parsons ML, Renseli JEJ, Townsend DW, Trainerk VL, Vargo GA. (2008) Harmful algal blooms and eutrophication: Examples of linkages from selected coastal regions of the United States *Harmful Algae* 8: 39-53.
4. Smayda TJ (1997) Harmful algal blooms: Their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnol Oceanogr* 42: 1137-1153.
5. Hoagland P, Scatasta S (2006) in Ecology of Harmful Algae, eds Graneli E, Turner J (Springer-Verlag) pp 391-402.
6. Gobler CJ, Lonsdale DJ, Boyer GL (2005) A synthesis and review of causes and impact of harmful brown tide blooms caused by the alga, *Aureococcus anophagefferens*. *Estuaries* 28: 726-749.
7. O'Kelly CJ Sieracki ME Their EC, Hobson IC (2003) A transient bloom of *Ostreococcus* (Chlorophyta, Prasinophyceae) in West Neck Bay, Long Island, New York. *J Phycol* 39: 850-854.
8. Sieracki, ME, Gobler CJ, Cucci, T, Thier E, Hobson I (2004) Pico- and nanoplankton dynamics during bloom initiation of Aureococcus in a Long Island, NY bay. *Harmful Algae* 3:459-470.
9. Bowler C, et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom *genomes*. *Nature* 456: 239–244.
10. Armbrust EV, et al. (2004) The Genome of the Diatom *Thalassiosira pseudonana*: Ecology, Evolution, and Metabolism. *Science* 306: 79 – 86.
11. Palenik B, et al. (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci USA* 104: 7705-7710.
12. Palenik B, et al. (2006) Genome sequence of Synechococcus CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci USA* 103: 13555–13559.
13. Boeckmann B et al. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31:365-370.
14. Finn RD et al., (2010) The Pfam protein families database: *Nucleic Acids Res* 38:D211-222.
15. Connolly JA, Oliver MJ, Beaulieu JM, Knight CA, Tomanek L, Moline MA (2008) Correlated evolution of genome size and cell volume in diatoms (Bacillariophyceae). *J Phycol* 44: 124-131.
16. Hessen DO, Jeyasingh PD, Neiman M, Weider LJ (2010) Genome streamlining and the elemental costs of growth. *Trends Ecol Evol* 25:75-80.
17. Raven JA, Kubler JE (2002) New light on the scaling of metabolic rate with the size of algae. *J Phycol* 38: 11-16.
18. Green BR, Durnford DG (1996) The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol* 47: 685–714.
19. Durnford DG, Deane JA, Tan S, McFadden GI, Gantt E, Green, B.R. (199) A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *J Molecular Evol* 48: 59-68.
20. Cock JM, et al. (2010) The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* 465: 617-621
21. Lefebvre SC, Harris G, Webster R, Leonardos N, Geider RJ, Raines RA, Read BA, Garrido JL (2010) Characterization and expression analysis of the LHCf gene family in *Emiliania huxleyi* (Haptophyta) reveals differential responses to light and CO_2. *J Phycol* 46:123-134.
22. Popels LC, MacIntyre HL, Warner ME, Yaohong Z, Hutchins DA (2007) Physiological responses during dark survival and recovery in *Aureococcus anophagefferens* (Pelagophyceae). *J Phycol* 43: 32-42.
23. Mulholland MR, Gobler CJ, Lee C (2002) Peptide hydrolysis, amino acid oxidation, and nitrogen uptake in communities seasonally dominated by *Aureococcus anophagefferens*. *Limnol Oceanogr* 47: 1094-1108.
24. Wurch LL, Haley ST, Orchard ED Gobler CJ, Dyhrman ST. (2010) Nutrient-regulated transcriptional responses in the brown tide 1 forming alga *Aureococcus anophagefferens*. Environ Microbiol, in press.

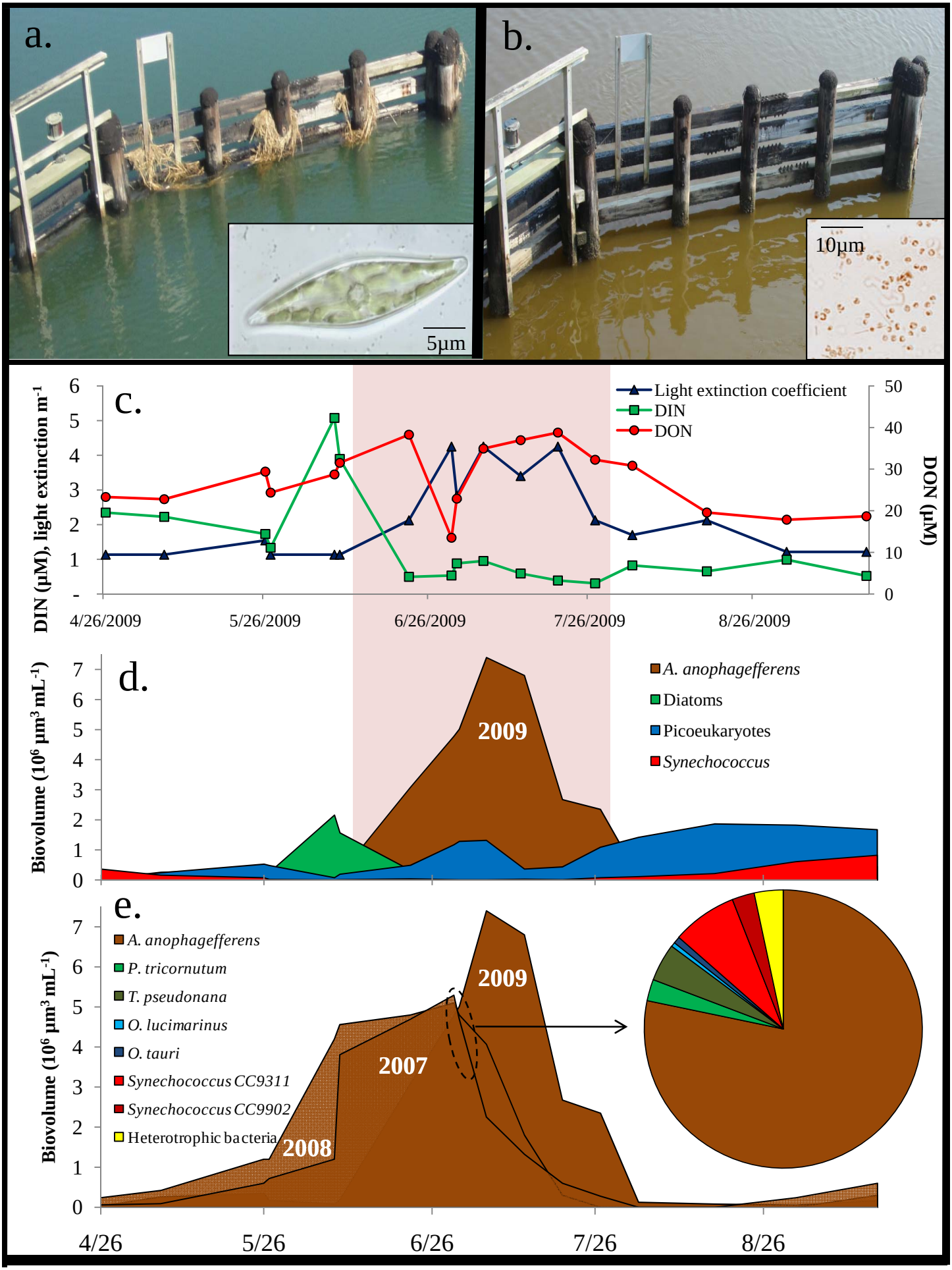
25. LaRoche J, Nuzzi R, Waters R, Wyman K, Falkowski PG, Wallace DWR (1997) Brown tide blooms in Long Island's coastal waters linked to variability in groundwater flow. *Global Change Biol* 3: 397-410.
26. Sañudo-Wilhelmy SA, Flegal AR (1993) Comparable levels of trace-metal contamination in two semi-enclosed embayments: San Diego Bay and South San Fransiscio Bay. *Environ Sci Technol* 27: 1934-1936.
27. Breuer E, Sañudo-Wilhelmy SA, Aller RC (1999). Distributions of trace metals and dissolved organic carbon in an estuary with restricted river flow and a brown tide. *Estuaries* 22: 603-615.
28. Cutter GA, Cutter LS (2004) Selenium biogeochemistry in the San Francisco Bay estuary: changes in water column behavior. *Estuarine Coastal Shelf Sci* 61:463–476.
29. Lobanov AV Fomenko DE, ZhangY, Sengupta A, Hatfield DL, Gladyshev VN (2007) Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic and small with terrestrial life. *Genome Biol.* 8: R198.
30. Stadtman TC (1996) Selenocysteine. *Annu Rev Biochem* 65, 83-100.
31. Hatfield DL, Gladyshev VN (2002) How Selenium Has Altered Our Understanding of the Genetic Code. *Mol Cell Biol* 22, 3565-3576.
32. Kim HY, Gladyshev VN (2005) Different Catalytic Mechanisms in Mammalian Selenocysteine- and Cysteine-Containing Methionine-R-Sulfoxide Reductases. *PLoS Biol.* 3: e375.
33. Score AJ, Palfreyman JW, White NA (1997) Extracellular phenoloxidase and peroxidase enzyme production during interspecific fungal interactions. *Int. Biodeterrioration Biodegradation* 39: 225-233.
34. Mayer AM (2006) Polyphenol oxidases in plants and fungi: Going places? A review. *Phytochem* 67: 2318-2331.
35. Fan C, Glibert PM, Alexander J, Lomas MW (2003) Characterization of urease activity in three marine phytoplankton species, *Aureococcus anophagefferens, Prorocentrum minimum*, and *Thalassiosira weissflogii*. *Mar Biol* 142: 949-958.
36. Wolfram L, Friedrich B, Eitinger T (1995) The Alcaligenes eutrophus protein HoxN mediates nickel transport in Escherichia coli. *J Bacteriol* 177: 1840-1843.
37. Messerschmidt A, Huber R, Wieghart K, Poulos T (2005) Handbook of Metalloproteins, Vol. 1–3. (Wiley).
38. Facchini PJ (2001) Alkaloid biosynthesis in plants: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. *Annu Rev Plant Physiol Plant Mol Biol* 52: 29–66.
39. Schmeller T, Latz-Bruing B, Wink M (1997) Biochemical activities of berberine, palamatinem and sanguinarine mediating chemical defence against microorganisms and herbivores. *Phytochemistry* 44: 257-266.
40. Rosado CJ, et al. (2007) A common fold mediates vertebrate defense and bacterial attack. *Science* 317:1548-1551.
41. Pierson LS, Gaffney T, Lam S, Gong F (1995) Molecular analysis of genes encoding phenazine biosynthesis in the biological control bacterium *Pseudomonas aureofaciens* 30–84. *FEMS Microbiol Lett* 134: 299 – 307.
42. Sharom FJ (2008)ABC multidrug transporters: structure, function and role in chemoresistance. *Pharmacogenomics* 9: 105-27.
43. van Veen HW, Konings WN (1998) The ABC family of multidrug transporters in microorganisms. Biochimica *et Biophysica Acta Bioenergetics* 1365:31-36.
44. Mittl PRE, Schneider-Brachert W (2007) Sel1-like repeat proteins in signal transduction. *Cellular Signaling* 19: 20-31.
45. Quarmy LM (1999) Signal transduction in the sexual life of *Chlamydomonas*. *Plant Mol Biol* 26:1271-87.
46. Neer EJ, Schmidt CJ, Nambudripad R, Smith TF (1994)The ancient regulatory-protein family of WD-repeat proteins. *Nature* 371:297-300.
47. Lotze HK Lenihan HS Bourque BJ, Bradbury RH, Cooke RG, Kay MC, Kidwell SM, Kirby MX, Peterson CH, Jackson, J.B.C. (2006) Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science* 312, 1806-1809.
48. Paerl HW, Pinckney JL, Fear JM, Peierls BL (1998) Ecosystem responses to internal and watershed organic matter loading: consequences for hypoxia in the eutrophying Neuse river estuary, North Carolina, USA. *Mar Ecol Prog Ser* 166, 17-25.
49. McBride MB Spiers G (2001) Trace element content of selected fertilizers and dairy manures as determined by ICP-MS. *Comm Soil Sci Plant Anal* 32:139–156.
50. MacIntyre HL, Lomas MW, Cornwell J, Suggett DJ, Gobler CJ, Koch EW, Kana TM (2004) Mediation of benthic-pelagic coupling by microphytobenthos: an energy- and material-based model for initiation of blooms o*f Aureococcus anophagefferens. Harmful Algae* 3: 403-437.

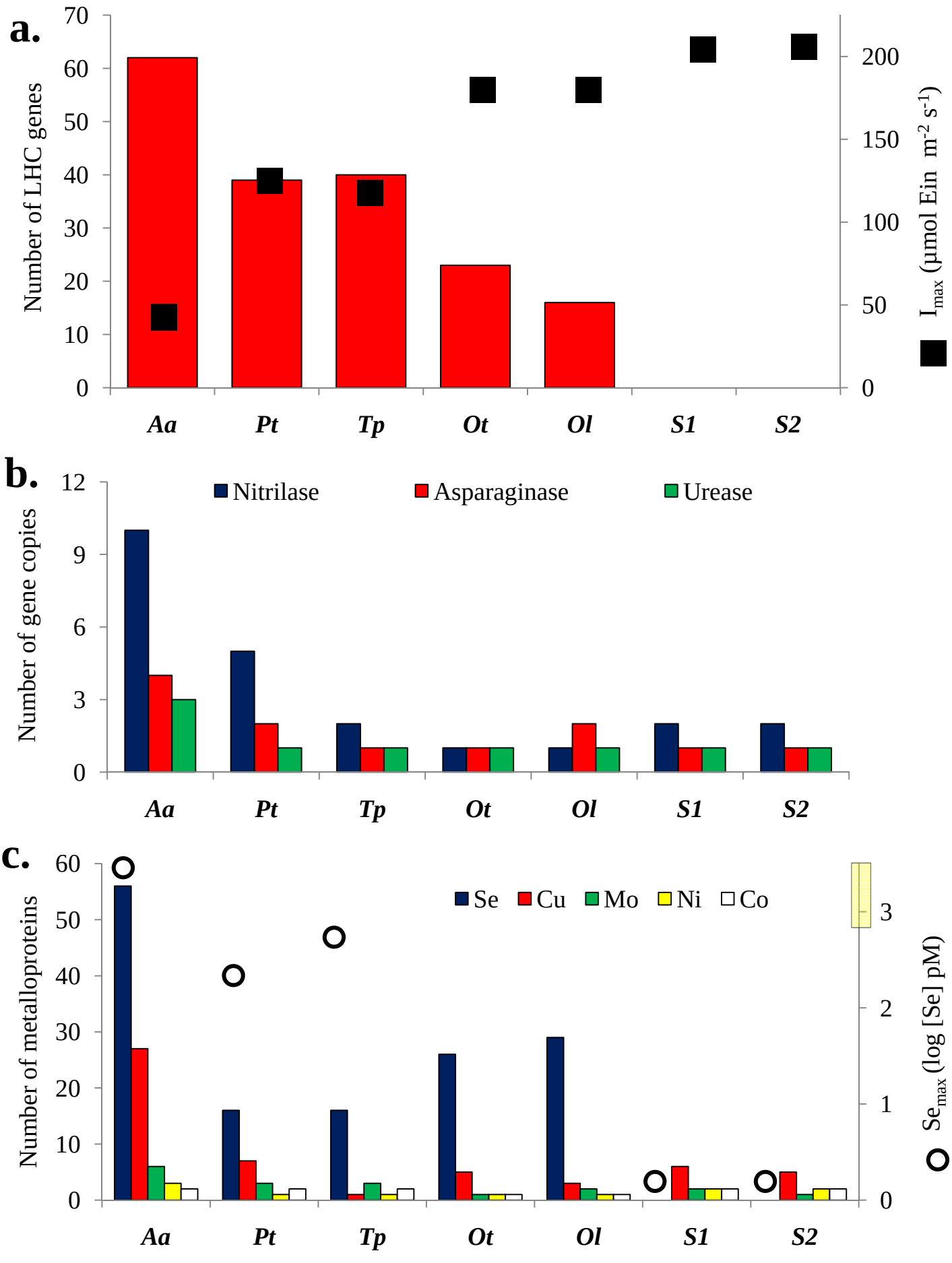
51. Quigg A, Finkel ZV, Irwin AJ, Reinfelder JR, Rosenthal Y, HoT-Y, Schofield O, Morel FMM, Falkowski PG (2003) The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. *Nature* 425: 291-294.
52. Doblin MA, Blackburn SI, Hallegraeff GM (1999) Comparative study of selenium requirements of three phytoplankton species: *Gymnodinium catenatum, Alexandrium minutum* (Dinophyta) and *Chaetoceros cf. tenuissimus* (Bacillariophyta). *J Plankton Res* 21: 1153-1169.

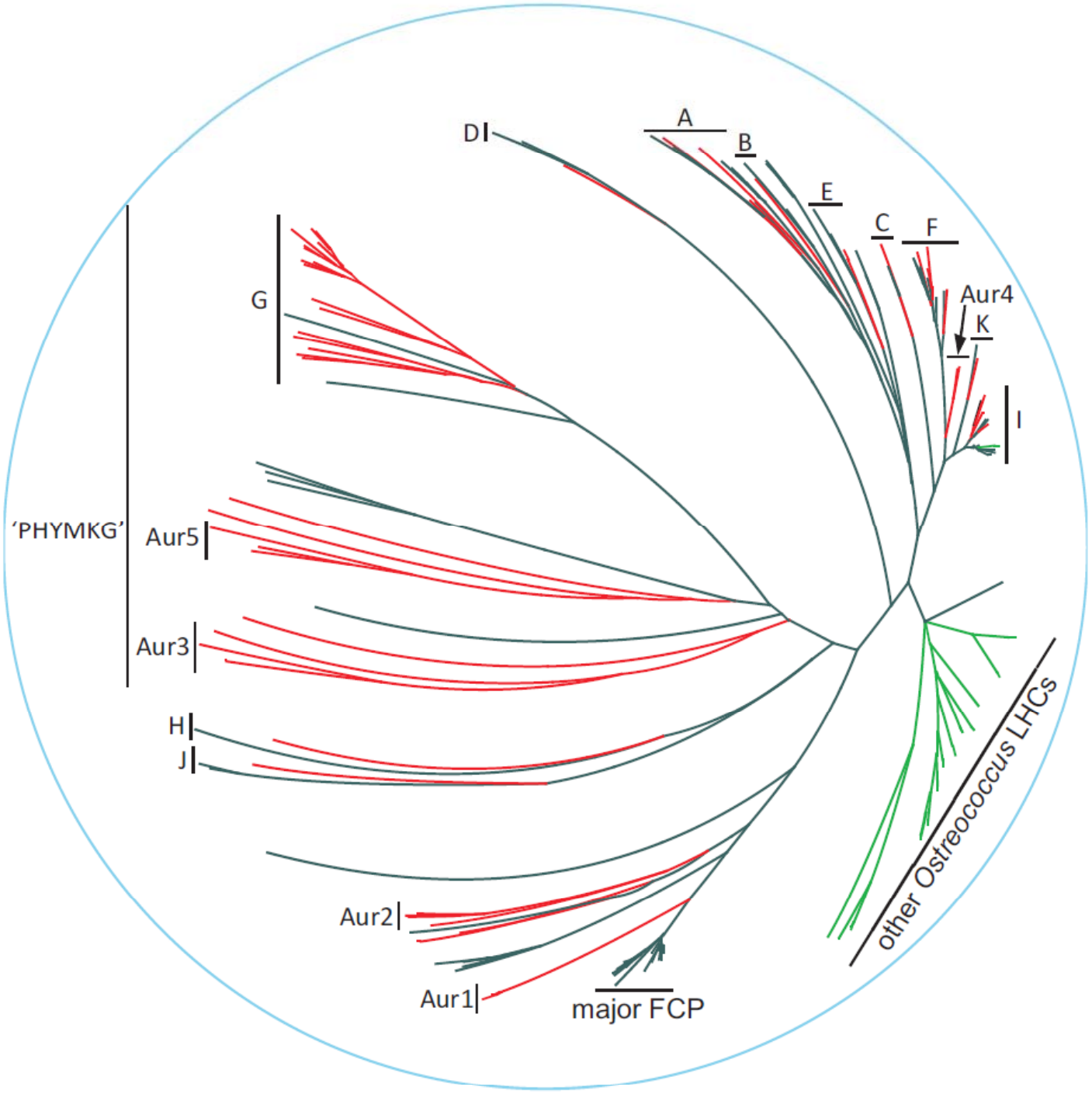
425 **Figure Legends**
426
427 **Figure 1.** Field observations from Quantuck Bay, NY, USA. a, Macro- and microscopic images
428 (inset) of an estuary (Quantuck Bay, NY, USA) under normal conditions from 9 June 2009, prior
429 to a brown tide (note the diatom in the inset micrograph image). b, similar macro- and
430 microscopic images (inset) taken 6 July 2009 during a harmful brown tide bloom caused by *A.*
431 *anophagefferens* (note the dominance of *A. anophagefferens* in the inset micrograph). c, The
432 dynamics of dissolved inorganic nitrogen (DIN), dissolved organic nitrogen (DON) and the
433 extinction coefficient of light within seawater during the spring and summer of 2009 in Quantuck
434 Bay. d, The dynamics of phytoplankton during the spring and summer of 2009, a year when *A.*
435 *anophagefferens* bloomed almost to the exclusion of other phytoplankton, including
436 picoeukaryotes that are often dominated by *Ostreococcus* sp. in estuaries which host brown tides
437 (6, 7, 8), and *Thalassiosira* and *Phaeodactylum*, genera which are found in this system (6). The
438 shaded region in panel C and D indicates the period when *A. anophagefferens* blooms,
439 highlighting that *A. anophagefferens* blooms when levels of DIN and light levels are low, and
440 DON levels are high and that *A. anophagefferens* blooms can persist for more than a month
441 during summer when this species dominates phytoplankton biomass inventories. e, The
442 dynamics of *A. anophagefferens* cell densities during 2007, 2008, and 2009 with the dates of
443 samples collected for metaproteome analyses (6/26/07 and 7/9/07) indicated within the dashed
444 circled. The inset metaproteome pie chart specifically depicts the mean relative abundance of
445 unique spectral counts of peptides matching proteins from *A. anophagefferens*, *P. tricornutum* (9)
446 *T. pseudonana* (10), *O. tauri* (11), *O. lucimarinus* (11), *Synechococcus* (CC9311) (12),
447 *Synechococcus* (CC9902), and heterotrophic bacteria.
448
449 **Figure 2.** Comparisons of gene compliment between *A. anophagefferens* and other co-occurring
450 phytoplankton species. Aa, Pt, Tp, Ot, Ol, S1, and S2 are *Aureococcus anophagefferens*,
451 *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*, *Ostreococcus tauri*, *Ostreococcus*
452 *lucimarinus*, *Synechococcus* clone CC9311 and *Synechococcus* clone CC9902, respectively. a,
453 The number of light harvesting complex (LHC) genes present in each phytoplankton genome
454 (red bars; left axis) and I_{max}, the irradiance level required to achieve maximal growth rates in
455 each phytoplankton (black squares; right axis). Among these species, *A. anophagefferens*
456 possesses the greatest number of LHC genes, achieves a maximal growth rate at the lowest level
457 of light, and blooms when light levels are low. b, The number of genes associated with the
458 degradation of nitriles, asparagine, and urea in each phytoplankton genome. *A. anophagefferens*
459 grows efficiently on organic nitrogen because it possesses more nitrilase, asparaginase, and
460 urease genes than other phytoplankton. c, Inter-species comparison of the genes encoding
461 proteins that contain the metals Se, Cu, Mo, Ni, and Co (left axis) and Se_{max}, the selenium level
462 (added as selenite shown as log concentrations) required to achieve maximal growth rates in *A.*
463 *anophagefferens*, *P. tricornutum*, *T. pseudonana*, *and Synechococcus* (white circles; right axis).
464 The range of dissolved selenium concentrations found in estuaries is depicted as a yellow bar on
465 the right y-axis. *A. anophagefferens* has the largest number of proteins containing Se, Cu, Mo,
466 and Ni and blooms exclusively in shallow estuaries where inventories of these metals are high.
467 See supplementary materials and methods for details of irradiance- and Se-dependent growth
468 data and Se concentrations in estuaries.
469
470

Figure 3. Phylogenetic tree constructed from amino acid sequences of predicted LHC proteins from two diatoms (*Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, black branches), two *Ostreococcus* species (*O. tauri* and *O. lucimarinus*, green branches), and *Aureococcus anophagefferens* (red branches). The tree constructed in MEGA4 (see Fig. S1) is displayed here after manipulation of the original branch lengths in Hypertree (http://kinase.com/tools/HyperTree.html) to aid visualization of major features of the tree. None of the *Aureococcus* LHCs were closely related to green plastid lineage LHCs, although four belonged to a group found in both the green and red plastid lineages (group I). None of the *Aureococcus* LHCs clustered with the 'major' fucoxanthin-chlorophyll binding proteins (FCP) of diatoms and other heterokonts (major FCP group). However, many *Aureococcus* LHCs did group with similar sequences from *P. tricornutum* and *T. pseudonana* (as well as LHCs from other red-lineage algae not included in this tree; groups A to K). There were also five groups of *A. anophagefferens* LHCs that were not closely related to any other LHCs (Aur1 to Aur5). Group G includes 16 LHCs from *A. anophagefferens* and two from *T. pseudonana*, and shares a unique PHYMKG motif near the end of helix two with 10 additional *A. anophagefferens* LHCs plus 5 more from the diatoms.

Figure 4. Genes encoding for enzymes involved in degrading organic carbon compounds in *A. anophagefferens*. The graph displays the portion and names of the genes encoding for functions which are unique to *A. anophagefferens* (red; 53%), enriched in *A. anophagefferens* relative to the six comparative phytoplankton (34%; green), and present at equal or lower numbers in *A. anophagefferens* relative to the six comparative phytoplankton (13%; blue). The number of genes present in multiple copies in *A. anophagefferens* is shown in parentheses. Further detail regarding these genes is presented in tables S10 and S11.


a.
5µm
b.
10µm
c.
DIN (µM), light extinction m-1
DON (µM)
Light extinction coefficient
DIN
DON
4/26/2009
5/26/2009
6/26/2009
7/26/2009
8/26/2009
d.
Biovolume (10^6 µm^3 mL^-1)
A. anophagefferens
Diatoms
Picoeukaryotes
Synechococcus
2009
e.
Biovolume (10^6 µm^3 mL^-1)
A. anophagefferens
P. tricornutum
T. pseudonana
O. lucimarinus
O. tauri
Synechococcus CC9311
Synechococcus CC9902
Heterotrophic bacteria
2009
2007
2008
4/26
5/26
6/26
7/26
8/26


a.
Number of LHC genes
70
60
50
40
30
20
10
0
200
150
100
50
0
I_{max} (µmol Ein m^{-2} s^{-1})
Aa
Pt
Tp
Ot
Ol
S1
S2
b.
Nitrilase
Asparaginase
Urease
Number of gene copies
12
9
6
3
0
Aa
Pt
Tp
Ot
Ol
S1
S2
c.
Se
Cu
Mo
Ni
Co
Number of metalloproteins
60
50
40
30
20
10
0
3
2
1
0
Se_{max} (log [Se] pM)
Aa
Pt
Tp
Ot
Ol
S1
S2


D
A
B
E
C
F
Aur4
K
I
G
'PHYMKG'
Aur5
Aur3
H
J
Aur2
Aur1
major FCP
other Ostreococcus LHCs

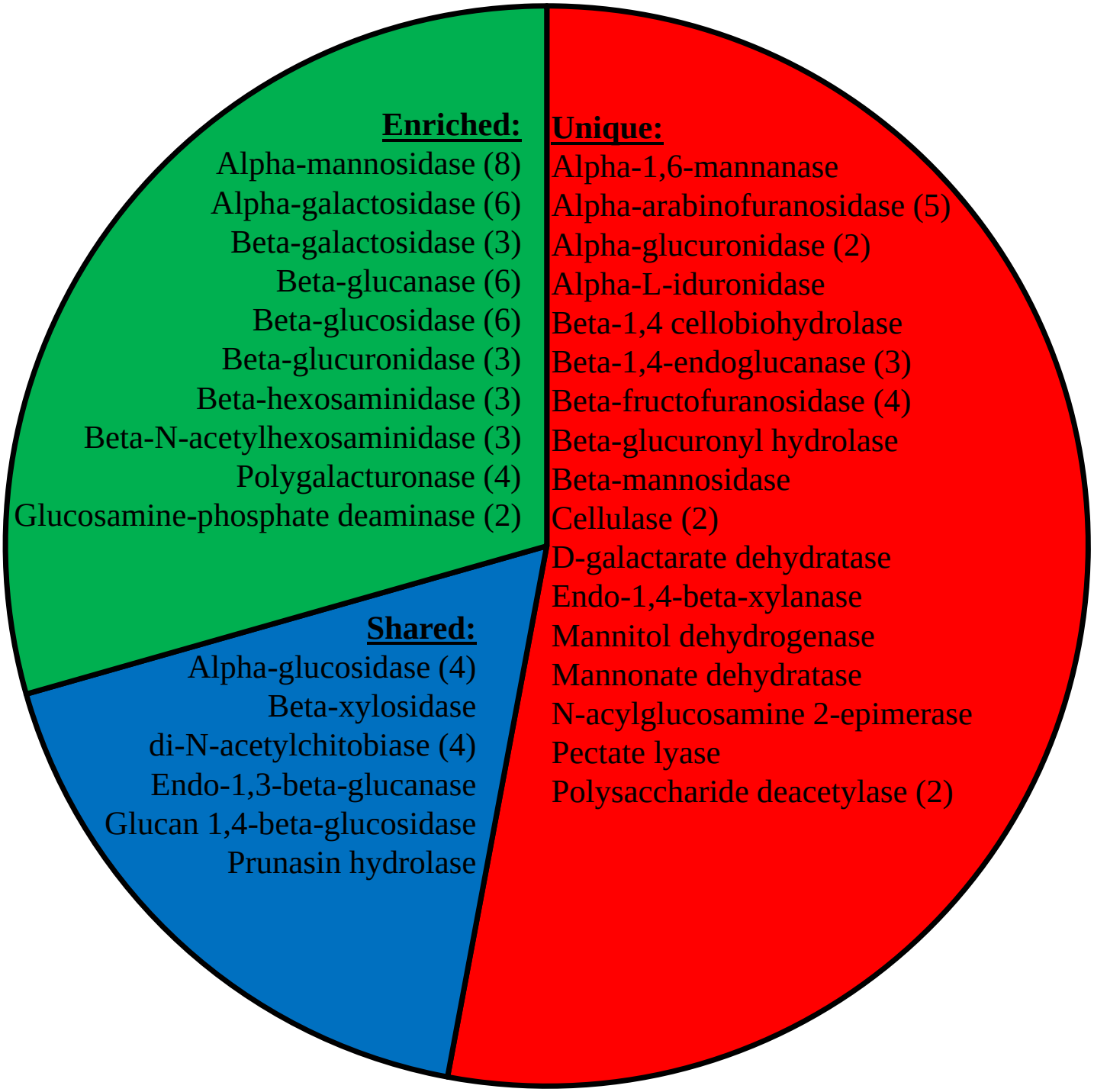

Enriched:
Alpha-mannosidase (8)
Alpha-galactosidase (6)
Beta-galactosidase (3)
Beta-glucanase (6)
Beta-glucosidase (6)
Beta-glucuronidase (3)
Beta-hexosaminidase (3)
Beta-N-acetylhexosaminidase (3)
Polygalacturonase (4)
Glucosamine-phosphate deaminase (2)
Shared:
Alpha-glucosidase (4)
Beta-xylosidase
di-N-acetylchitobiase (4)
Endo-1,3-beta-glucanase
Glucan 1,4-beta-glucosidase
Prunasin hydrolase
Unique:
Alpha-1,6-mannanase
Alpha-arabinofuranosidase (5)
Alpha-glucuronidase (2)
Alpha-L-iduronidase
Beta-1,4 cellobiohydrolase
Beta-1,4-endoglucanase (3)
Beta-fructofuranosidase (4)
Beta-glucuronyl hydrolase
Beta-mannosidase
Cellulase (2)
D-galactarate dehydratase
Endo-1,4-beta-xylanase
Mannitol dehydrogenase
Mannonate dehydratase
N-acylglucosamine 2-epimerase
Pectate lyase
Polysaccharide deacetylase (2)

Table 1. Major features of the genomes of *A. anophagefferens*, and six competing algal species *P. tricornutum* (9), *T. pseudonana* (10), *O. tauri* (11), *O. lucimarinus* (11), *Synechococcus* (CC9311) (12), *Synechococcus* (CC9902). Genes with known functions were identified using Swiss-Prot, a curated protein sequence database, with an e-value cut-off of $< 10^{-5}$ (13). Pfam domains are sequences identified from a database of protein families represented by multiple sequence alignments and hidden Markov models (13). The compressed nature of *P. tricornutum* cells (11 x 2.5 µm) makes its biovolume smaller than *T. pseudonana*.

	A. anophagefferens	***P. tricornutum***	***T. pseudonana***	***O. tauri***	***O. lucimarinus***	***Synechococcus* (CC9311)**	***Synechococcus* (CC9902)**
Cell diameter (µm)	2.0	11.0	5.0	1.2	1.3	1.0	1.0
Cell volume (µm3)	6	61	88	1.8	2.0	1.2	1.2
Genome size (Mbp)	57	27	32	13	13	2.6	2.2
Predicted gene number	11,501	10,402	11,242	7,892	7,651	2,892	2,301
Genes with known functions	8,560	6,239	6,797	5,090	5,322	1,607	1,469
Genes with Pfam domains	6,908	5,398	5,791	4,763	4,214	1,636	1,488
Genes with unique Pfam domains	209	79	75	23	51	55	12