Niche of harmful alga Aureococcus anophagefferens revealed through ecogenomics

Christopher J. Gobler1,2,* , Dianna L. Berry1,2, †, Sonya T. Dyhrman3, †, Steven W. Wilhelm4, †, Asaf Salamov5, Alexei V. Lobanov6, Yan Zhang6, Jackie L. Collier2, Louie L. Wurch3, Adam B. Kustka7, Brian D. Dill8, Manesh Shah9, Nathan C. VerBerkmoes8, Alan Kuo5, Astrid Terry5, Jasmyn Pangilinan5, Erika Lindquist5, Susan Lucas5, Ian Paulsen10, Theresa K. Hattenrath1,2, Stephanie C. Talmage1,2, Elyse A. Walker1,2, Florian Koch1,2, Amanda M. Burson1,2, Maria Alejandra Marcoval1,2, Ying-Zhong Tang1,2, Gary R. LeCleir3, Kathryn J. Coyne11, Gry Mine Berg12, Erin M. Bertrand13, Mak A. Saito13, 14, Vadim Gladyshev5, Igor V. Grigoriev4,* 1

1School of Marine and Atmospheric Sciences, Stony Brook University, Southampton, NY 11968, USA. 2School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY 11794-5000, USA. 3Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, 02543, USA. 4Department of Microbiology, The University of Tennessee, Knoxville, TN 37996, USA. 5US Department of Energy, Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA. 6Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston MA 02115, USA. 7Department of Earth and Environmental Sciences, Rutgers University, Newark, New Jersey 07102, USA. 8Chemical Sciences and 9Biosciences Divisions, Oak Ridge National Laboratory, Oak Ridge, TN, 37830, USA. 10Department of Chemistry and Biomolecular Sciences, Macquarie University Sydney, 2109, NSW, Australia. 11College of Earth, Ocean, and Environment, University of Delaware,
Lewes, DE, 19958 USA.\textsuperscript{12}Department of Environmental Earth System Science, Stanford University, 397 Panama Mall, Stanford, California 94305, USA. \textsuperscript{13}Massachusetts Institute of Technology and Woods Hole Oceanographic Institution Joint Program in Chemical Oceanography. \textsuperscript{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, \textit{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.
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 $5 \times 10^6$ 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 sulfatase 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.


Figure Legends

Figure 1. Field observations from Quantuck Bay, NY, USA. a, Macro- and microscopic images (inset) of an estuary (Quantuck Bay, NY, USA) under normal conditions from 9 June 2009, prior to a brown tide (note the diatom in the inset micrograph image). b, similar macro- and microscopic images (inset) taken 6 July 2009 during a harmful brown tide bloom caused by *A. anophagefferens* (note the dominance of *A. anophagefferens* in the inset micrograph). c, The dynamics of dissolved inorganic nitrogen (DIN), dissolved organic nitrogen (DON) and the extinction coefficient of light within seawater during the spring and summer of 2009 in Quantuck Bay. d, The dynamics of phytoplankton during the spring and summer of 2009, a year when *A. anophagefferens* bloomed almost to the exclusion of other phytoplankton, including picoeukaryotes that are often dominated by *Ostreococcus* sp. in estuaries which host brown tides (6, 7, 8), and *Thalassiosira* and *Phaeodactylum*, genera which are found in this system (6). The shaded region in panel C and D indicates the period when *A. anophagefferens* blooms, highlighting that *A. anophagefferens* blooms when levels of DIN and light levels are low, and DON levels are high and that *A. anophagefferens* blooms can persist for more than a month during summer when this species dominates phytoplankton biomass inventories. e, The dynamics of *A. anophagefferens* cell densities during 2007, 2008, and 2009 with the dates of samples collected for metaproteome analyses (6/26/07 and 7/9/07) indicated within the dashed circled. The inset metaproteome pie chart specifically depicts the mean relative abundance of unique spectral counts of peptides matching proteins from *A. anophagefferens*, *P. tricornutum* (9), *T. pseudonana* (10), *O. tauri* (11), *O. lucimarinus* (11), *Synechococcus* (CC9311) (12), *Synechococcus* clone CC9902, and heterotrophic bacteria.

Figure 2. Comparisons of gene compliment between *A. anophagefferens* and other co-occurring phytoplankton species. Aa, Pt, Tp, Ot, Ol, S1, and S2 are *Aureococcus anophagefferens*, *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*, *Ostreococcus tauri*, *Ostreococcus lucimarinus*, *Synechococcus* clone CC9311 and *Synechococcus* clone CC9902, respectively. a, The number of light harvesting complex (LHC) genes present in each phytoplankton genome (red bars; left axis) and *I*$_{\text{max}}$, the irradiance level required to achieve maximal growth rates in each phytoplankton (black squares; right axis). Among these species, *A. anophagefferens* possesses the greatest number of LHC genes, achieves a maximal growth rate at the lowest level of light, and blooms when light levels are low. b, The number of genes associated with the degradation of nitriles, asparagine, and urea in each phytoplankton genome. *A. anophagefferens* grows efficiently on organic nitrogen because it possesses more nitrilase, asparaginase, and urease genes than other phytoplankton. c, Inter-species comparison of the genes encoding proteins that contain the metals Se, Cu, Mo, Ni, and Co (left axis) and *Se*$_{\text{max}}$, the selenium level (added as selenite shown as log concentrations) required to achieve maximal growth rates in *A. anophagefferens*, *P. tricornutum*, *T. pseudonana*, and *Synechococcus* (white circles; right axis). The range of dissolved selenium concentrations found in estuaries is depicted as a yellow bar on the right y-axis. *A. anophagefferens* has the largest number of proteins containing Se, Cu, Mo, and Ni and blooms exclusively in shallow estuaries where inventories of these metals are high. See supplementary materials and methods for details of irradiance- and Se-dependent growth data and Se concentrations in estuaries.
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. Light extinction coefficient
b. 10µm
DIN (µM), light extinction m⁻¹

D. 2009
A. anophagefferens
Diatoms
Picoeukaryotes
Synechococcus

E. 2007
2008

Heterotrophic bacteria
Synechococcus CC9311
Synechococcus CC9902
O. lucimarinus
O. tauri
P. tricornutum
T. pseudonana
A. anophagefferens
a. Number of LHC genes

b. Number of gene copies

c. Number of metalloproteins
**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)

**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)

**Shared:**
- Alpha-glucosidase (4)
- Beta-xylosidase
- di-N-acetylchitobiose (4)
- Endo-1,3-beta-glucanase
- Glucan 1,4-beta-glucosidase
- Prunasin hydrolase
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*.

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