Strategies among phytoplankton in response to alleviation of nutrient stress in a subtropical gyre

Supplementary Information

Robert H. Lampe[1,2], Seaver Wang[3], Nicolas Cassar[3,4], and Adrian Marchetti[2]*

Affiliations:

[1]Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA
[2]Department of Marine Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3]Division of Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, NC, USA
[4]Laboratoire des Sciences de l'Environnement Marin (LEMAR), UMR 6539 UBO/CNRS/IRD/IFREMER, Institut Universitaire Européen de la Mer (IUEM), Brest, FR

***For correspondence:**

Dr. Adrian Marchetti
Department of Marine Sciences
University of North Carolina at Chapel Hill
3202 Venable Murray Hall CB 3300
Chapel Hill, NC 27599-3300, USA
Email: amarchetti@unc.edu
Tel. (+1) 919 843 3473
Fax (+1) 919 962 1254

This file includes:

Supplementary Text

Phylogenetic analysis of nitric oxide dioxygenase (NOD)

NOD sequences from haptophytes, dinoflagellates, and diatoms (i.e., the Hacrobia+SAR supergroup)[1] form a distinct clade from yeast and heterotrophic bacteria that still maintain the characteristic globin and FAD-binding domains and conservation of the iron and FAD binding sites, although some sequences show a tyrosine to phenylalanine mutation at the FAD binding site (Fig. S12)[2]. Within this eukaryotic phytoplankton clade, *NOD* sequences do not necessarily correspond to taxonomic groupings. For example, *NOD* in the model diatom *Phaeodactylum tricornutum* is dissimilar from other diatoms and most closely related to *NOD* in the cryptophyte, *Cryptomonas curvata*. *NOD* in *Thalassiosira gravida* is similar to that found in most dinoflagellates, but some dinoflagellate *NOD sequences* are similar to other diatoms including *Thalassiosira gravida.* Haptophyte *NOD* sequences also are similar and mostly form a distinct clade apart from *E. huxleyi*, where there are two copies, one within this haptophyte clade and one similar to dinoflagellates. Although these results indicate the presence of *NOD* in haptophytes, expression was not detected at either site.

NOD is an ideal candidate for HGT as it exists as a single-gene metabolic module; the reaction is self-contained and catalyzed by a single protein. [3]. *NOD* in yeast and diplomonads is believed to have originated via horizontal gene transfer (HGT) from heterotrophic bacteria [4]. This conclusion is supported in our analysis by a distinct clade of heterotrophic bacteria and fungi with the diplomonad, *Giardia lamblia*, as well as the amoeba *Dictyostelium discoideum* that also likely acquired the gene via HGT. It is possible that multiple HGT events have occurred for eukaryotic phytoplankton to inherit the gene.

Supplementary Materials and Methods

Satellite data

Satellite-derived data were obtained from the NOAA CoastWatch Browser and plotted with matplotlib for Python v2.7 [5]. Chlorophyll *a* concentrations on a 0.025° grid and sea surface temperature on a 0.005° grid were obtained from MODIS on board the Aqua (EOS PM) and from the NOAA GOES system respectively.

Nutrient analyses

Seawater (15 mL) was syringe filtered through Whatman GF/F filters into acid-rinsed polypropylene tubes and immediately frozen at -20°C. Dissolved nitrate + nitrite (NO_2 + NO_3), phosphate (PO_4), and silicic acid (H_4SiO_4) were quantified with an Astoria Analyzer [6]. The detection limit for NO_2 + NO_3 was 0.007 µmol L^{-1}. Iron concentrations in the initial seawater are provided in Tang et al. [7].

Chlorophyll

Seawater (400 mL) was gravity-filtered through a 5 µm polycarbonate filter (47 mm) followed by a 0.45 µm mixed cellulose ester membrane filter (25 mm) under gentle vacuum pressure (<100 mm Hg). Funnels were rinsed with a sterile artificial saline solution, and the filters were immediately frozen at -80°C until analysis onshore. Chlorophyll *a* extraction was performed using 90% acetone at -20°C for 24 h and measured via *in vitro* fluorometry on a 10-AU fluorometer (Turner Designs, San Jose, CA, USA) using the acidification method [8].

Particulate nitrogen and ^{15}N-based uptake rates

Subsamples from the cubitainers (2 L) were spiked with 0.1 µmol L^{-1} $Na^{15}NO_3$ for the T0 and surface treatments or 0.5 µmol L^{-1} $Na^{15}NO_3$ for the DSW treatment and incubated in the same on-deck incubators in polycarbonate bottles. After 8 hours, samples were gravity-filtered through a 5 µm polycarbonate filter (47 mm) followed by a pre-combusted (450°C for 5 h) Whatman GF/F filter under gentle vacuum pressure (<100 mm Hg). Particles on the 5 µm filter were then rinsed onto a separate pre-combusted GF/F filter using a sterile artificial saline solution. Filters were stored at -20°C until preparation for analysis.

For quantification, filters were first dried at 60°C for 24 h then encapsulated in tin. Particulate nitrogen (PN) and percent ^{15}N atom were measured with an elemental analyzer paired with a continuous flow isotope ratio mass spectrometer (EA-IRMS) at the University of California, Davis Stable Isotope Facility. Absolute uptake rates (ρ) and biomass (PN)-specific uptake rates (*V*) were blank corrected and calculated according to the equations provided in Dugdale and Wilkerson (9).

18S rDNA sequencing and analysis

Samples were collected and analyzed following Wang et al. [10]. Briefly, 1-2 L of near-surface (5 m) seawater was collected from the same CTD casts, filtered onto a 0.22 µm filter (Millipore, Billerica, MA, USA) using a peristaltic pump and flash frozen in liquid nitrogen. DNA was later extracted using the Qiagen DNeasy Plant Mini Kit (Qiagen, Germantown, MD, USA) with the addition of internal DNA standards as described in Lin et al. [11]. Importantly, the standards (0.679 ng/5.78 x 10^6 rRNA gene copies of *Schizosaccharomyces pombe* gDNA) per sample were added before an initial bead beating lysis step prior to DNA extraction. The addition of the internal standards allows for quantification of the abundance of rDNA reads per L of seawater [12].

PCR amplification was performed to target the V4 region of the 18S rDNA gene using universal primers for marine eukaryotic taxa modified to improve coverage of haptophytes (Forward: 5'-CCAGCASCYGCGGTAATTCC-3', Reverse: 5'-ACTTTCGTTCTTGAT-3') [13, 14]. Primers were each dual-indexed with 6 bp long barcodes using a heterogeneity spacer approach [15]. The PCR reactions consisted of 2.5 µL 10x PCR buffer, 0.5 µL dNTPs (10 mM each), 1 µL MgSO4 (50 mM), 0.5 µL of each primer (10 µM), 0.2 µL Platinum Taq DNA Polymerase High Fidelity (Thermo Fisher Scientific, Waltham, MA, USA), and 19.3 µL of sterile water. Reactions were run with an initial 3 min denaturing step at 94°C, then 30 cycles consisting of 94°C for 30 s, 57°C for 1 min, and 72°C for 1 min, followed by a final step of 72°C for 10 min. PCR products were purified using the Qiagen QIAquick PCR Purification Kit and quantified with a Qubit 3.0 fluorometer (Life Technologies, Carlsbad, CA, USA).

Barcoded samples were pooled at equimolar concentrations and sequenced on an Illumina MiSeq at the Duke Center for Genomic and Computation Biology with V3 chemistry and 300 bp paired-end reads. Raw reads were trimmed, assembled and quality filtered with PANDAseq [16] following Fadrosh et al. [15]. Primer and other non-biological spacer sequences were removed

using TagCleaner [17]. OTUs were clustered based on 97% similarity using USEARCH [18] and annotated using the SILVA database (Release 123.1)[19]. Taxonomy assignment was conducted using the RDP classifier v2.2 [20]. Alignment was performed using PyNAST [21]. Absolute abundance of each OTU was calculated by dividing the number of OTU reads by the recovery ratio of the internal standards and the volume of seawater filtered [12]. After taxonomy assignment, internal standard and metazoan sequences were extracted and filtered from the taxonomy tables.

Phylogenetic analysis

Sequences were obtained from PhyloDB which includes sequences from the Marine Microbial Eukaryote Transcriptome Sequencing Project [22, 23] and from the JGI Genome Portal [24], KEGG [25], and UniProt [26] then were aligned with MUSCLE in Geneious Pro v10.2.2 [27]. A maximum likelihood phylogenetic tree of the reference sequences was created with RAxML v8.2.9 (PROTGAMMALG model) with 100 bootstraps [28]. Final trees were visualized and edited with Archaeopteryx v0.9916 [29].

Supplementary Figures

Fig. S1. Temperature (blue) and on-deck photosynthetically active radiation (PAR; green) during the incubations for each site. Plotted values are for every 15 minutes from a HOBO Data Logger (Onset, Cape Cod, MA, USA).

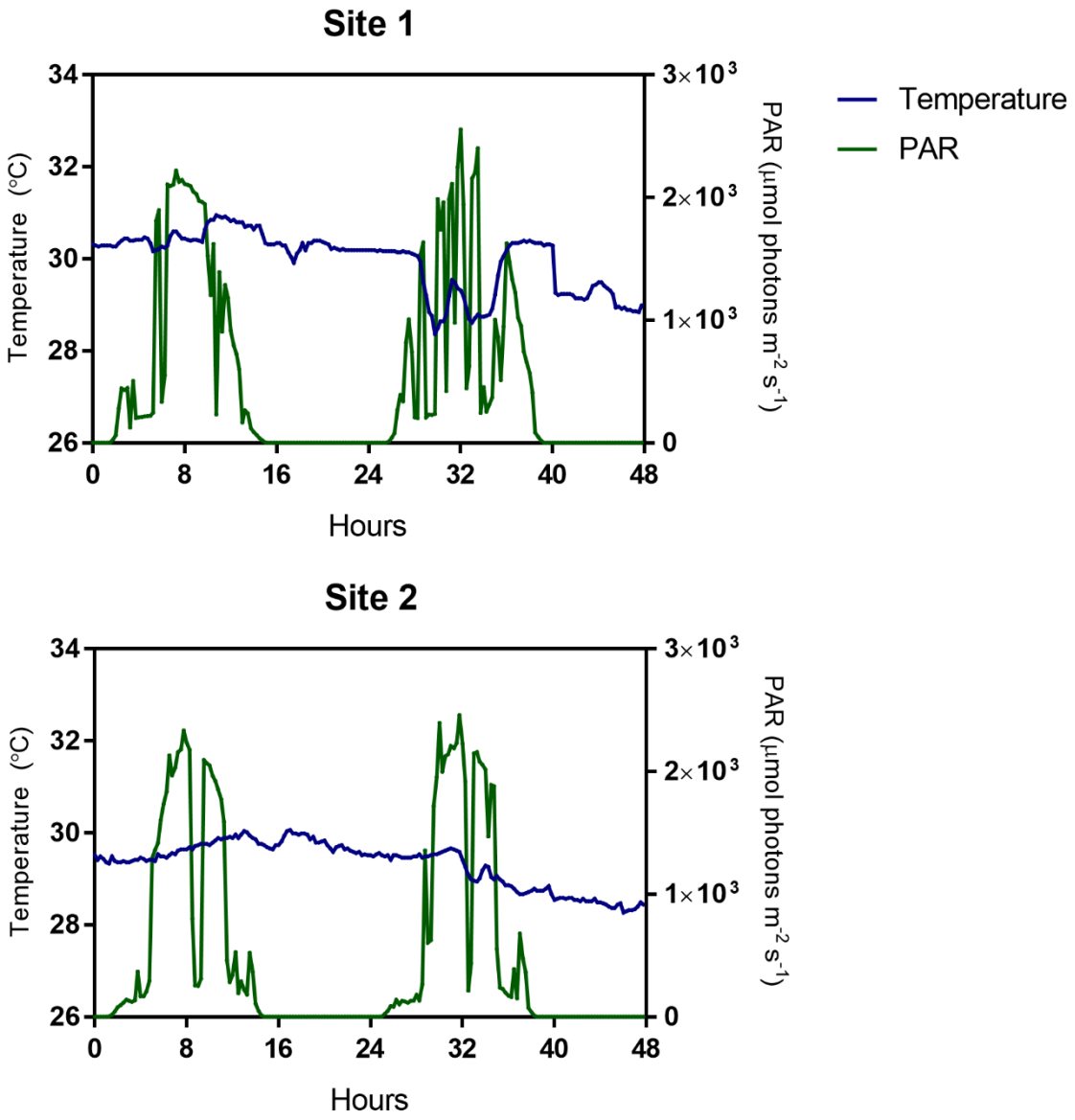


Fig S2. Relative abundances of the taxonomic groups from the initial seawater based on V4-18S ribosomal DNA and the initial seawater and the surface treatments based on RNA transcripts after 48 hours of incubation for both experimental sites. More detailed taxonomic information is provided in Figs. S3, S4, S5, and S7 as well as Dataset S2.

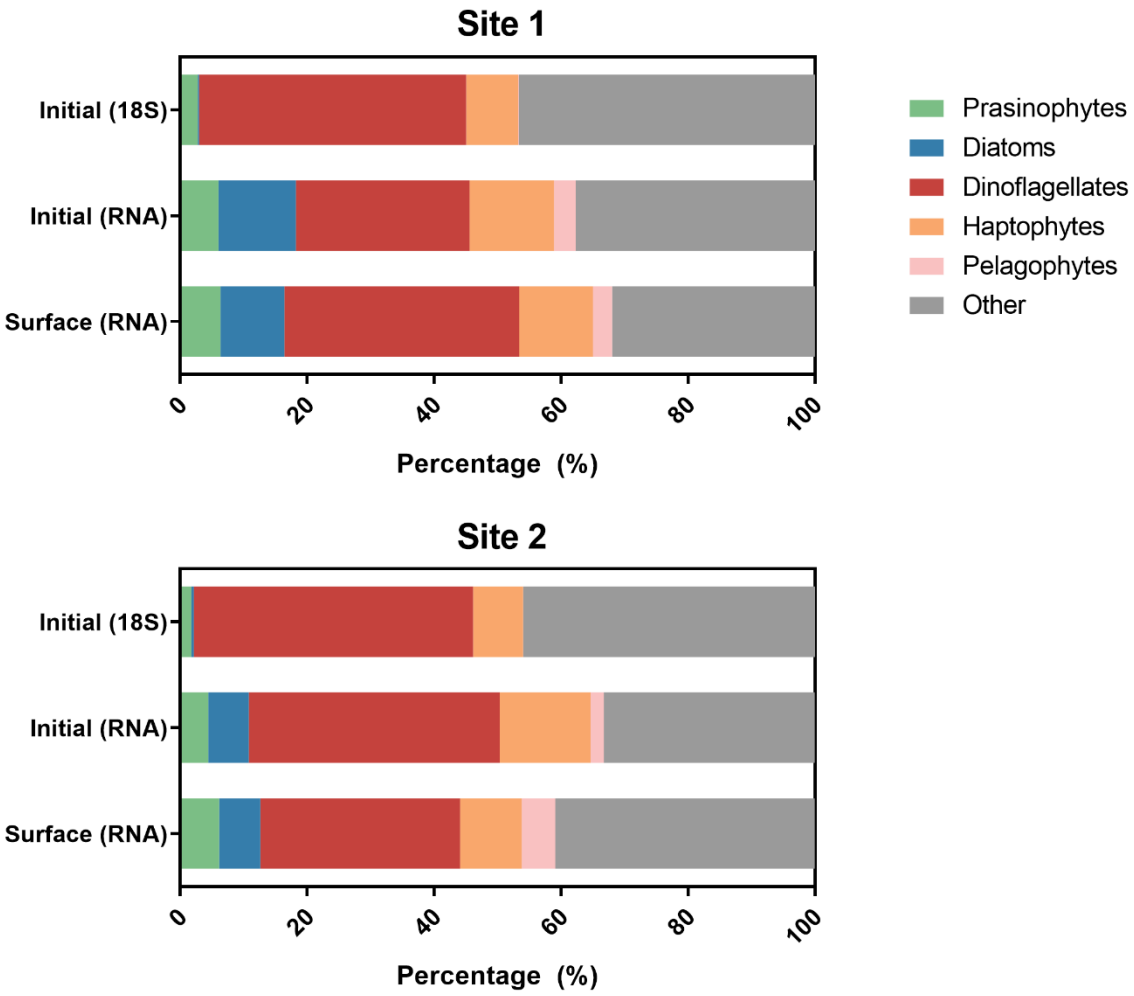


Fig. S3. 18S-based relative abundances of the dinoflagellates in the initial seawater at both experimental sites on the family and genus level (SILVA). Unassigned OTUs were unable to be resolved on the displayed taxonomic levels.

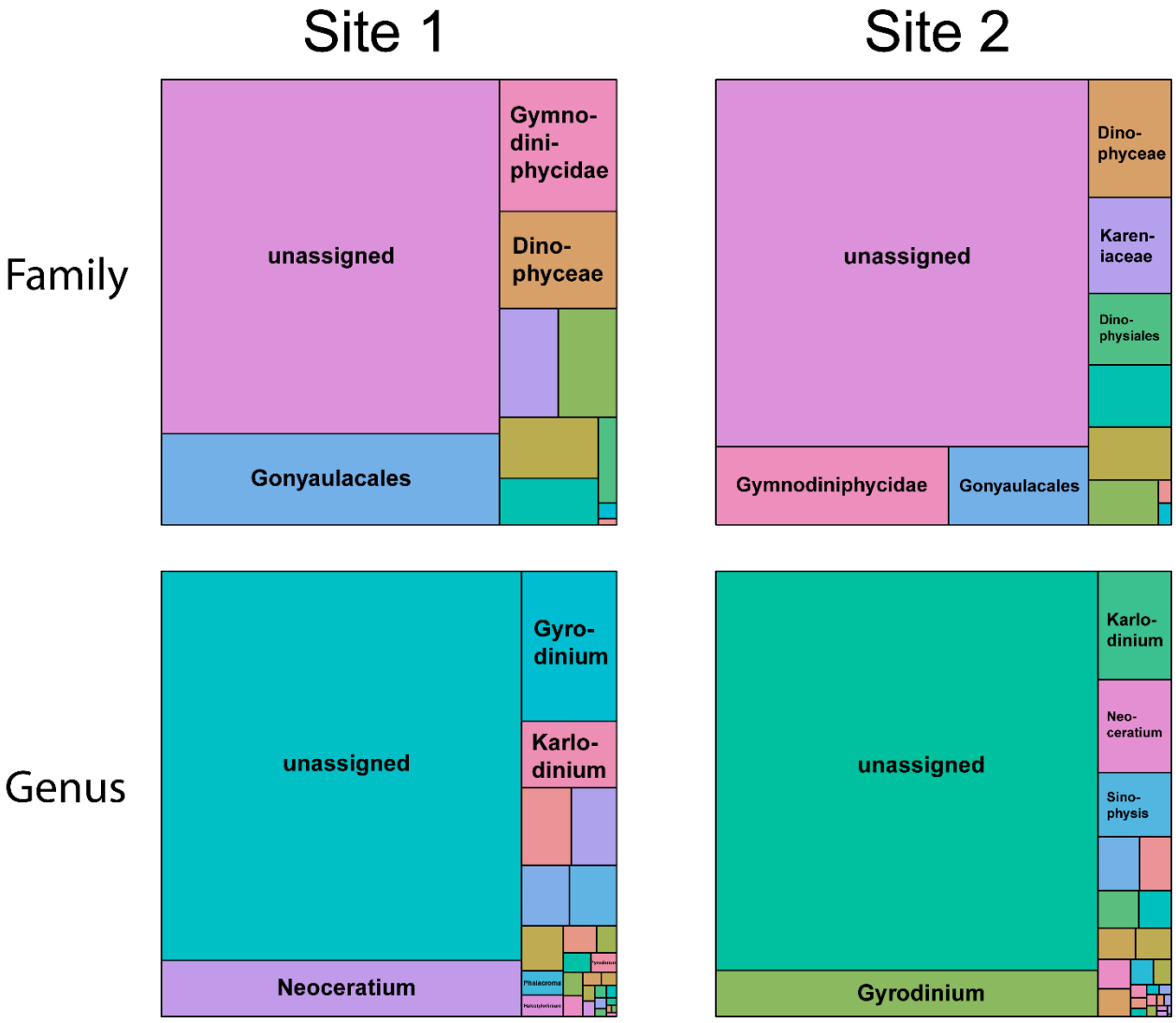


Fig. S4. 18S-based relative abundances of the haptophytes in the initial seawater at both experimental sites on the order level (SILVA). Family and genus levels are not displayed as most OTUs did not have family or genus level assignments. Unassigned OTUs were unable to be resolved on the order level.

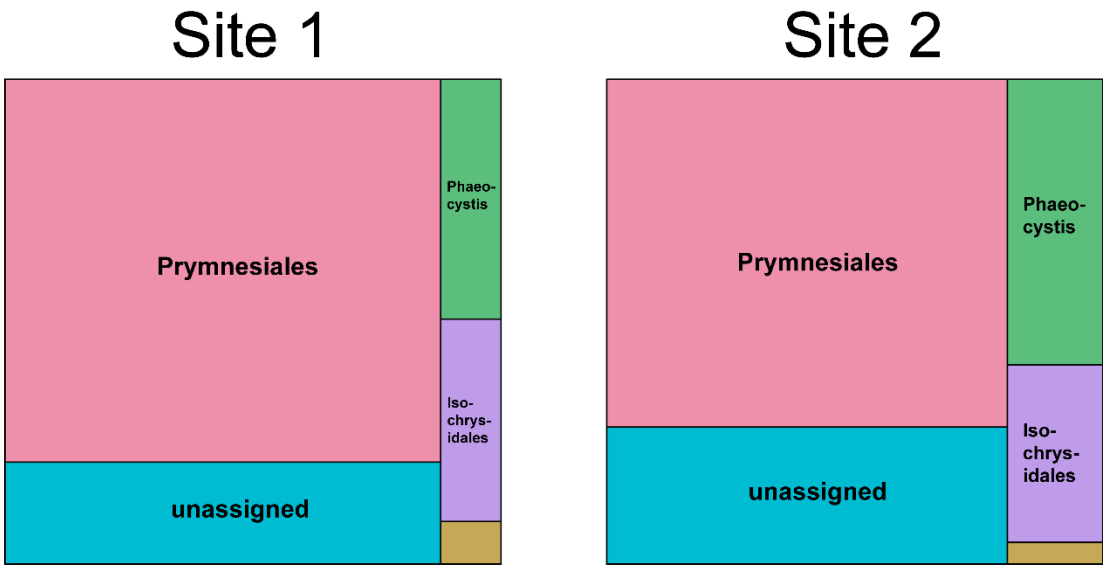


Fig. S5. 18S-based relative abundances of the prasinophytes in the initial seawater at both experimental sites on the family level (SILVA). Unassigned OTUs were unable to be resolved on the family level.

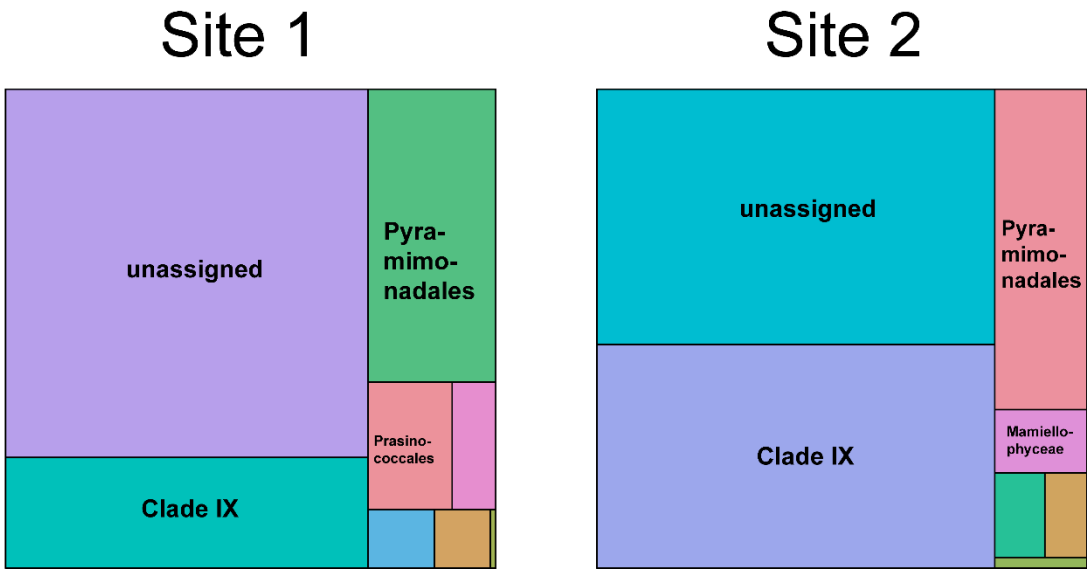


Fig. S6. Abundance of diatoms (rDNA copies L^{-1}) during the cruise within the Sargasso Sea. Seawater for the incubations was collected as stations 28 and 54 (blue). The 10% trimmed mean of the dataset is shown with a horizontal gray dashed line. Values on the y-axis are plotted on a log scale.

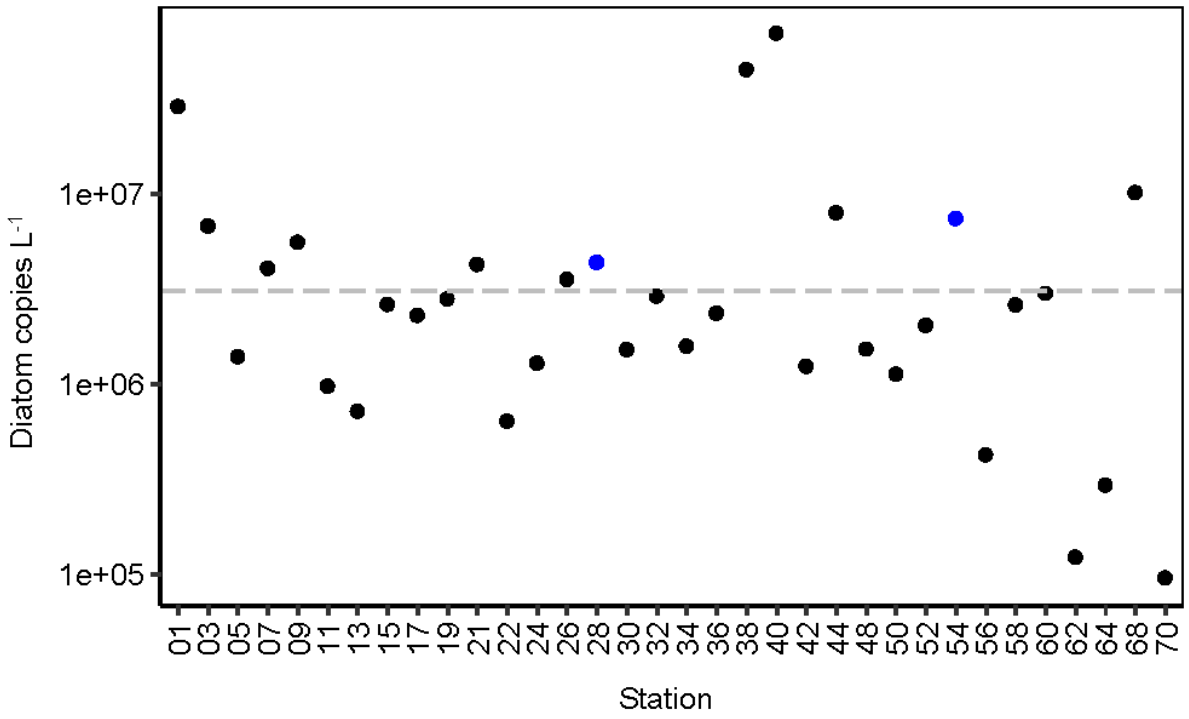


Fig. S7. 18S-based relative abundances of the diatoms in the initial seawater at both experimental sites on the genus level (SILVA).

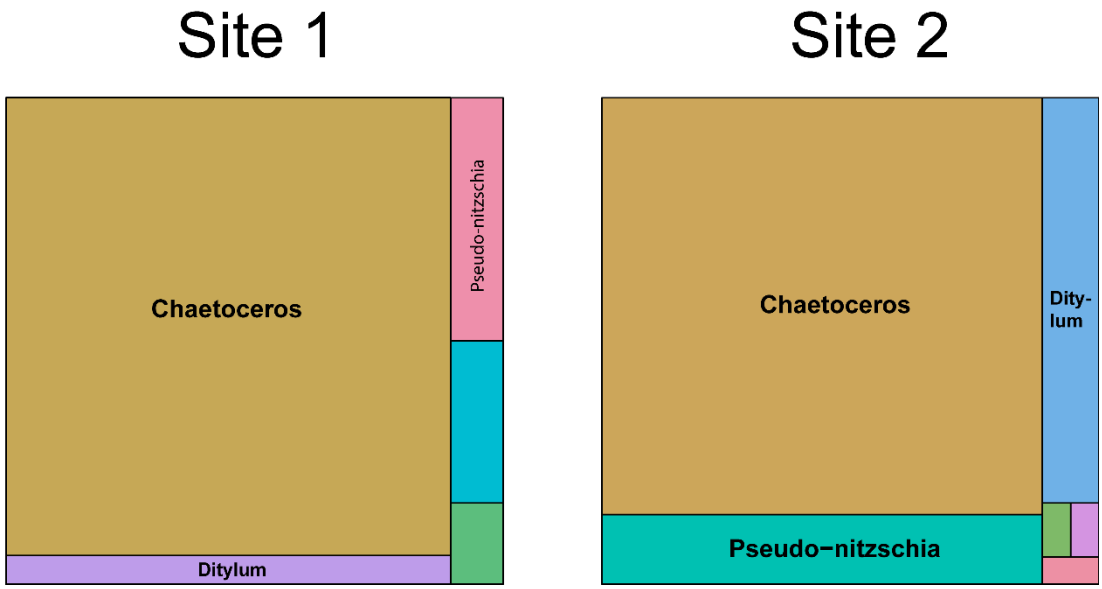


Fig. S8. Heatmaps for (A) the 927 expressed KOs where the gene was not detected in one group other than diatoms, and (B) the 1,358 expressed KOs where the gene was detected in diatoms and only one other group. Gray bars indicate that the gene was not detected.

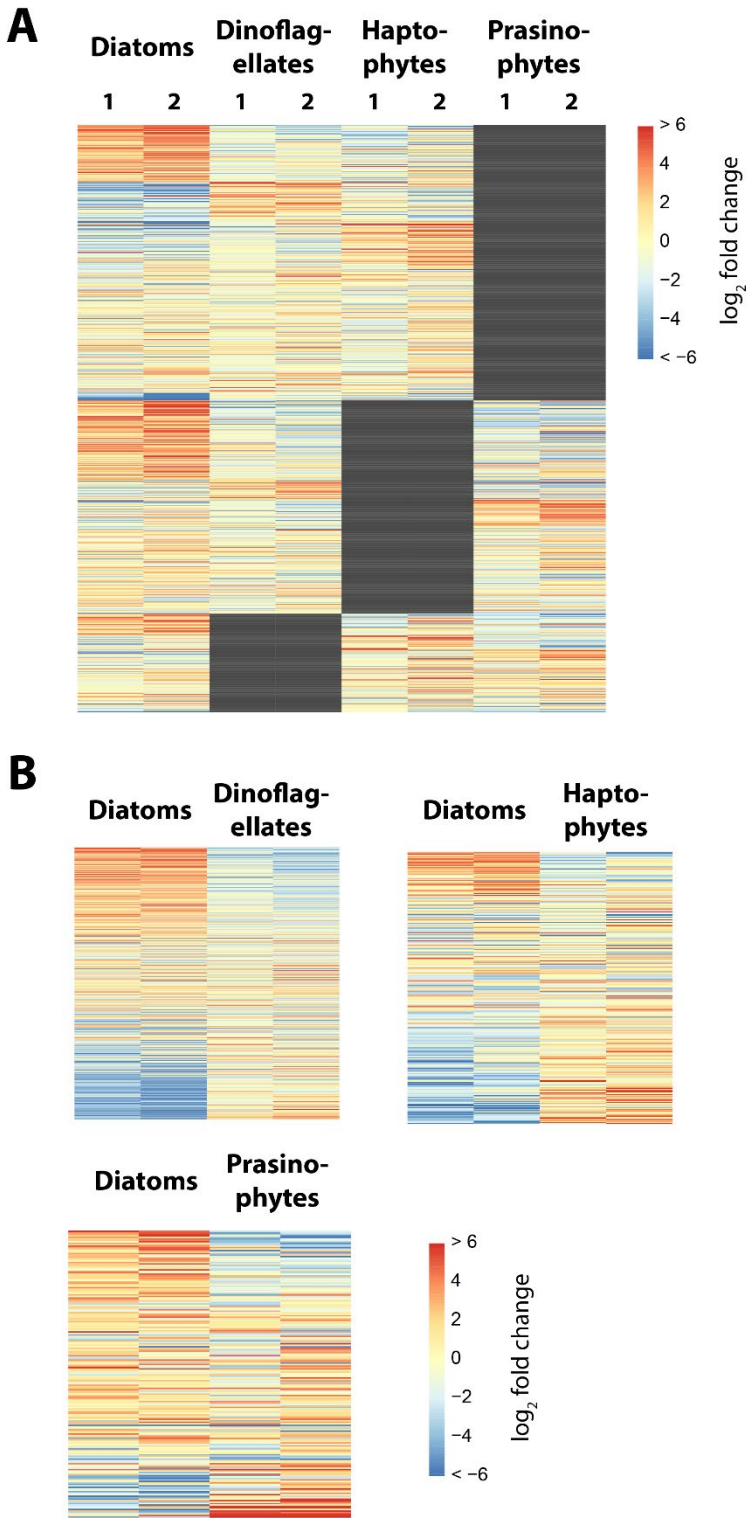


Fig. S9. Heatmap of KEGG ortholog (KO) abundances grouped by each KEGG module category (BRITE Level B) for the genes found to be uniquely overrepresented for each group shown in Fig. 5B.

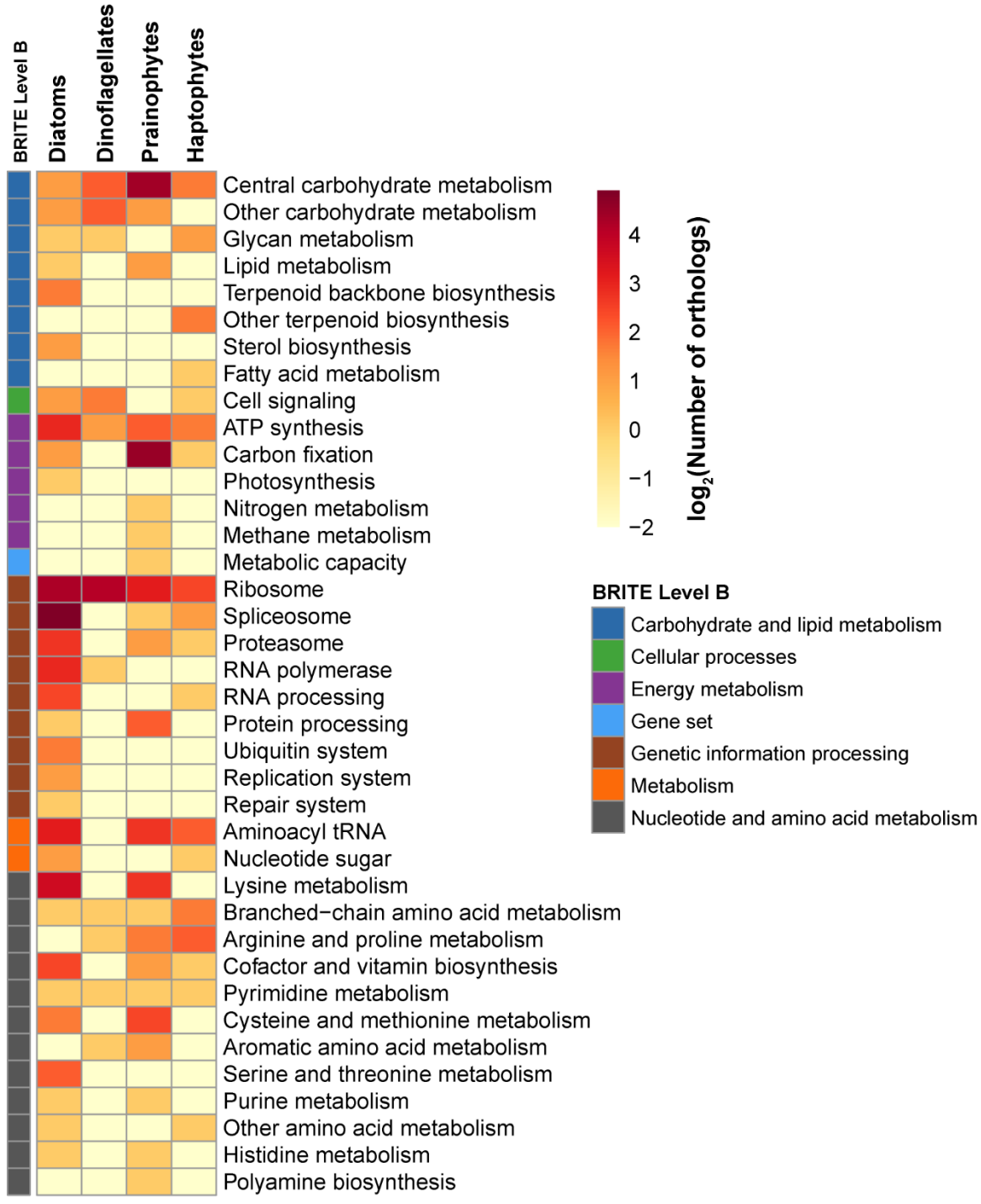


Fig. S10. Heatmap of variance stabilized transformed counts [30] for nitrogen metabolism genes displayed in Fig. 6.

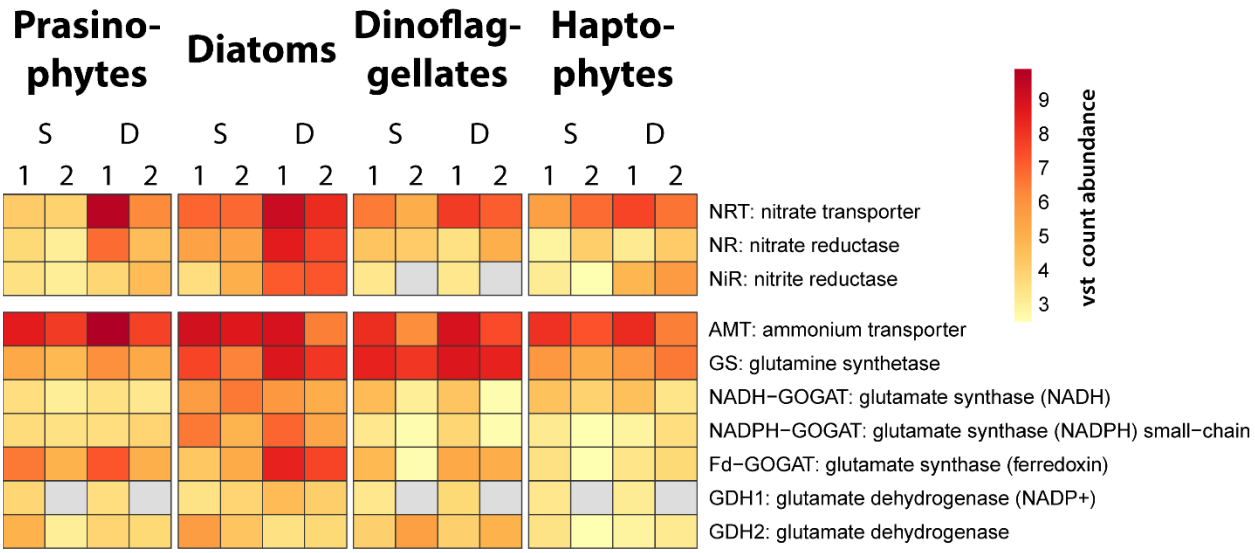


Fig. S11. Midpoint-rooted maximum likelihood phylogenetic tree of nitric oxide dioxygenase (*NOD*) sequences. Bootstrap values ≥ 50 are shown, and the scale bar represents 0.1 amino acid substitutions per position. Labels are colored corresponding to their taxonomic groupings. Nodes consisting of sequences of species within the same genera are collapsed with the genus name and the number of nodes/species denoted in brackets.

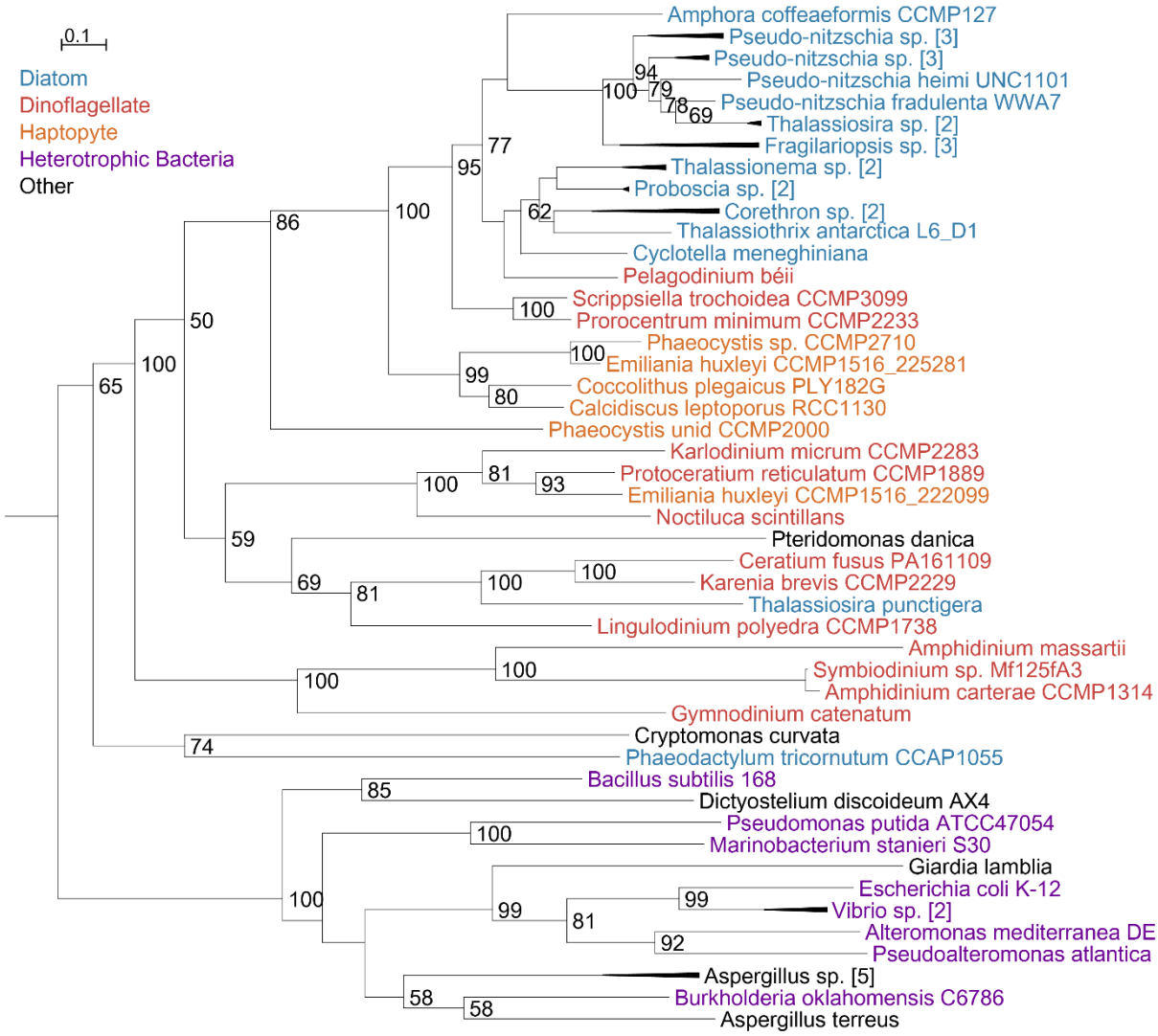
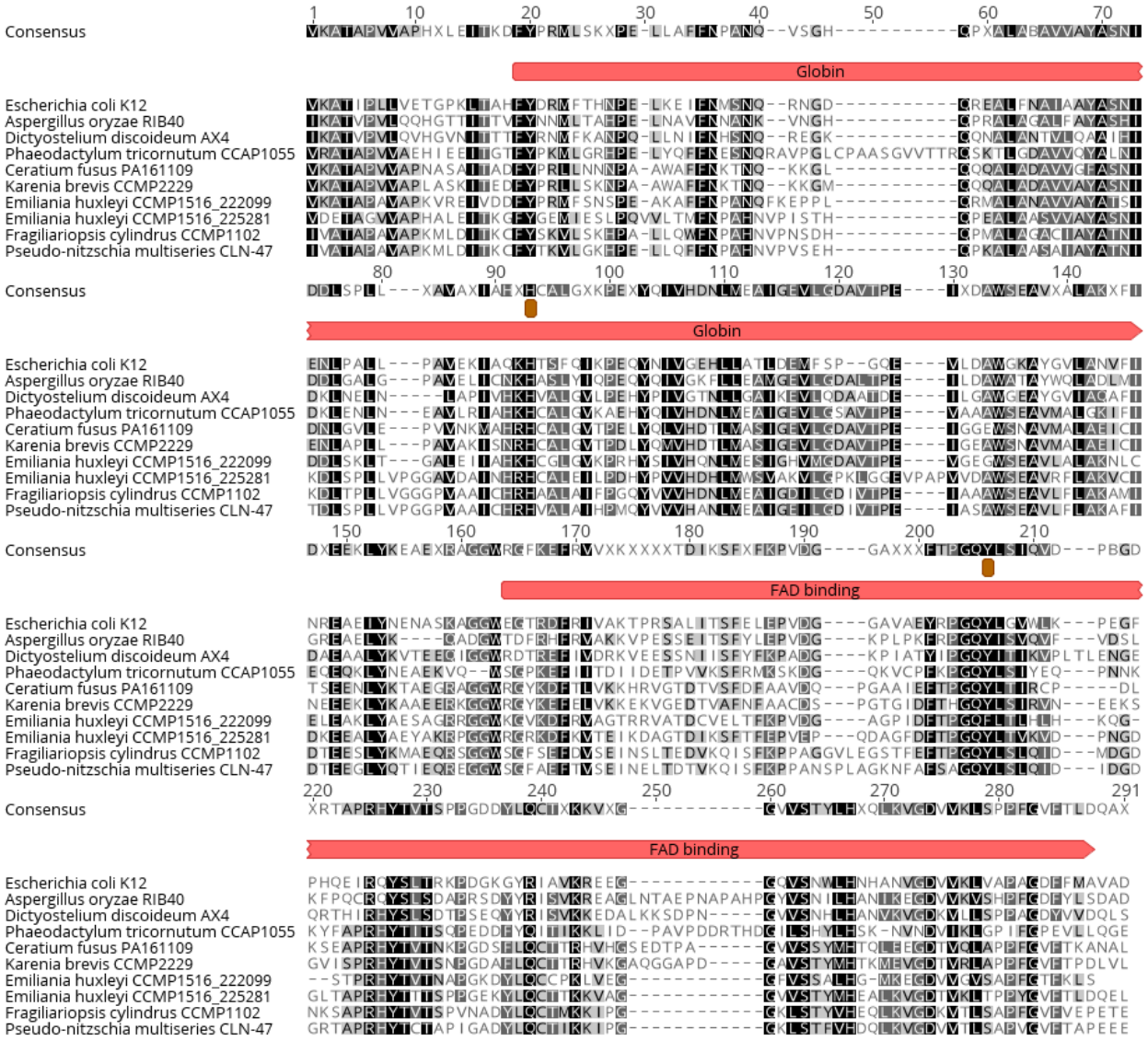


Fig. S12. Alignment of nitric oxide dioxygenase (NOD) sequences. Red bars indicate the globin and FAD binding domains. Binding sites are marked (brown). The amino acids at site 93 represent the iron binding domain and the amino acids at site 206 represent the FAD binding site.



Supplementary Datasets

Dataset S1. Sequencing information for each sample. The number of paired-end (PE) reads for both the high output (HO) and rapid run (RR) were quantified after quality trimming.

Dataset S2. 18Sv4 OTUs for both sites with SILVA taxonomy.

Dataset S3. Log_2 fold change values for KEGG Orthologs displayed in Fig. 5B.

Dataset S4. Differential gene expression results in diatoms from DESeq2 for the DSW versus surface condition.

Supplementary Information References

1. Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, et al. Phylogenomics reshuffles the eukaryotic supergroups. PLoS one. 2007;2:e790-e.
2. Ilari A, Bonamore A, Farina A, Johnson KA, Boffi A. The X-ray Structure of Ferric *Escherichia coli* Flavohemoglobin Reveals an Unexpected Geometry of the Distal Heme Pocket. J Biol Chem. 2002;277:23725-32.
3. Wisecaver JH, Alexander WG, King SB, Todd Hittinger C, Rokas A. Dynamic Evolution of Nitric Oxide Detoxifying Flavohemoglobins, a Family of Single-Protein Metabolic Modules in Bacteria and Eukaryotes. Mol Biol Evol. 2016;33:1979-87.
4. Andersson JO, Sjögren ÅM, Davis LAM, Embley TM, Roger AJ. Phylogenetic Analyses of Diplomonad Genes Reveal Frequent Lateral Gene Transfers Affecting Eukaryotes. Curr Biol. 2003;13:94-104.
5. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng. 2007;9:90-5.
6. Barwell-Clarke J, Whitney F. IOS nutrient methods and analysis. Can Tech Rep Hydrog Ocean Sci. 1996;182:1-43.
7. Tang W, Wang S, Fonseca-Batista D, Dehairs F, Gifford S, Gonzalez AG, et al. Revisiting the distribution of oceanic N_2 fixation and estimating diazotrophic contribution to marine production. Nat Commun. 2019;10:831.
8. Parsons TR, Maita Y, Lalli CM. A Manual of Chemical & Biological Methods for Seawater Analysis. Amsterdam: Pergamon; 1984.
9. Dugdale RC, Wilkerson FP. The use of ^{15}N to measure nitrogen uptake in eutrophic oceans; experimental considerations. Limnol Oceanogr. 1986;31:673-89.
10. Wang S, Lin Y, Gifford S, Eveleth R, Cassar N. Linking patterns of net community production and marine microbial community structure in the western North Atlantic. ISME J. 2018;12:2582-95.
11. Lin Y, Gifford S, Ducklow H, Schofield O, Cassar N. Towards quantitative marine microbiome community profiling using internal standards. Appl Environ Microb. 2018; :AEM.02634-18.
12. Satinsky BM, Gifford SM, Crump BC, Moran MA. Use of internal standards for quantitative metatranscriptome and metagenome analysis. Method Enzymol. 2013;531:237-50.
13. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microb. 2013;79:5112-20.
14. Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol. 2010;19:21-31.
15. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome. 2014;2:6.
16. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics. 2012;13:31.
17. Schmieder R, Lim YW, Rohwer F, Edwards R. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. BMC Bioinformatics. 2010;11:341.

18. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460-1.
19. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35:7188-96.
20. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Enviro Microbiol. 2007;73:5261-7.
21. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics (Oxford, England). 2010;26:266-7.
22. Bertrand EM, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, et al. Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. P Natl Acad Sci USA. 2015;112:9938-43.
23. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. PLoS Biol. 2014;12:e1001889.
24. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res. 2014;42:D26-D31.
25. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40:D109-D14.
26. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45:D158-D69.
27. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792-7.
28. Stamatakis A. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics. 2014;30:1312-1313.
29. Han MV, Zmasek CM. phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics. 2009;10:1-6.
30. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.