Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability

Adrian Marchetti[a,1,2,3], David M. Schruth[a,1], Colleen A. Durkin[a], Micaela S. Parker[a], Robin B. Kodner[a], Chris T. Berthiaume[a], Rhonda Morales[a], Andrew E. Allen[b], and E. Virginia Armbrust[a,2]

[a]School of Oceanography, University of Washington, Seattle, WA 98105; and [b]J. Craig Venter Institute, San Diego, CA 92121

Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved December 20, 2011 (received for review November 9, 2011)

In vast expanses of the oceans, growth of large phytoplankton such as diatoms is limited by iron availability. Diatoms respond almost immediately to the delivery of iron and rapidly compose the majority of phytoplankton biomass. The molecular bases underlying the subsistence of diatoms in iron-poor waters and the plankton community dynamics that follow iron resupply remain largely unknown. Here we use comparative metatranscriptomics to identify changes in gene expression associated with iron-stimulated growth of diatoms and other eukaryotic plankton. A microcosm iron-enrichment experiment using mixed-layer waters from the northeastern Pacific Ocean resulted in increased proportions of diatom transcripts and reduced proportions of transcripts from most other taxa within 98 h after iron addition. Hundreds of diatom genes were differentially expressed in the iron-enriched community compared with the iron-limited community; transcripts of diatom genes required for synthesis of photosynthesis and chlorophyll components, nitrate assimilation and the urea cycle, and synthesis of carbohydrate storage compounds were significantly overrepresented. Transcripts of genes encoding rhodopsins in eukaryotic phytoplankton were significantly underrepresented following iron enrichment, suggesting rhodopsins help cells cope with low-iron conditions. Oceanic diatoms appear to display a distinctive transcriptional response to iron enrichment that allows chemical reduction of available nitrogen and carbon sources along with a continued dependence on iron-free photosynthetic proteins rather than substituting for iron-containing functional equivalents present within their gene repertoire. This ability of diatoms to divert their newly acquired iron toward nitrate assimilation may underlie why diatoms consistently dominate iron enrichments in high-nitrate, low-chlorophyll regions.

RNA-seq | bloom | geoengineering | climate mitigation

In large areas of contemporary oceans, phytoplankton growth and primary production are limited by iron availability, mainly due to minimal iron inputs into oceanic waters far from coastal margins and the short residence time of bioavailable iron in oxygenated seawater (1). Over a dozen large-scale iron enrichment experiments have been performed in oceanic high-nitrate, low-chlorophyll (HNLC) regions of the subarctic Pacific, the equatorial Pacific, and the Southern Ocean. All have shown that adding iron to surface waters results in dramatic increases in phytoplankton biomass (2). Before the addition of iron, small fast-growing phytoplankton such as cyanobacteria, chlorophytes, and haptophytes dominate the biomass of the phytoplankton communities and grow primarily on reduced forms of nitrogen (N), such as ammonium. The abundance of these small phytoplankton is suggested to be controlled largely by a combination of heterotrophic grazing and some degree of iron stress (3, 4). In the northeastern (NE) Pacific, large cells, such as diatoms, are usually rare members of the iron-limited plankton community. When iron is added to surface waters in this HNLC region, phytoplankton growth is stimulated and within days, diatoms dominate the phytoplankton biomass, fueled by the high concentrations of available nitrate (5). Newly available iron facilitates carbon (C) and N assimilation into components such as chlorophyll and energy storage compounds (6), the mechanisms of which are poorly understood. The net result of iron fertilization is enhanced organic matter accumulation leading to a temporary flux of atmospheric CO_2 into the upper ocean and subsequent export of C to the ocean interior (7–9).

A mechanistic understanding of ways that diatoms cope with exceedingly low iron availability and respond rapidly to iron enrichment has come from a combination of field studies and studies with model species under controlled laboratory conditions. Iron-limited diatoms have reduced chlorophyll contents, photosynthetic efficiencies (10, 11), and rates of N assimilation (12, 13), all of which contribute to reduced growth rates. When iron limited, cells may temporarily replace iron-requiring proteins with less efficient iron-free functional equivalents to minimize their iron demands, including the alternation between the iron-containing photosynthetic electron transfer protein ferredoxin and the iron-free version flavodoxin (14, 15). Most cultured isolates of diatoms preferentially use ferredoxin when grown under iron-replete conditions (16–18) although inconsistencies between laboratory and field studies have been reported (5, 19). In some instances, reliance on iron-free versions of proteins may be permanent. For example, coastal diatoms rely on the iron-requiring cytochrome c_6 for electron transport whereas certain open ocean species have replaced cytochrome c_6 with the copper-containing protein plastocyanin, presumably reducing their overall need for iron (20). The availability of whole-genome sequences for diatoms has uncovered additional examples of genetic variation among genera. For example, three sets of iron-responsive gene clusters identified in the pennate diatom *Phaeodactylum tricornutum* that encode putative components of iron capture and uptake are not found within the genome of the centric diatom *Thalassiosira pseudonana* (6).

Author contributions: A.M. and E.V.A. designed research; A.M., C.A.D., and R.M. performed research; A.M. and D.M.S. contributed new reagents/analytic tools; A.M., D.M.S., C.A.D., M.S.P., R.B.K., C.T.B., A.E.A., and E.V.A. analyzed data; and A.M., D.M.S., M.S.P., and E.V.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The 454 and SOLiD sequences reported in this paper have been deposited in the National Center for Biotechnology Information's Sequence Read Archive (accession no. SRP006906).

[1]A.M. and D.M.S. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: amarchetti@unc.edu or armbrust@u.washington.edu.

[3]Present address: Department of Marine Sciences, University of North Carolina, Chapel Hill, NC 27514.

See Author Summary on page 1828.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1118408109/-/DCSupplemental.

These results underscore the impressive functional diversity of diatoms and suggest specific adaptations have occurred over evolutionary timescales to allow diatoms to subsist and sometimes thrive in chronically low-iron HNLC waters.

Here we use comparative metatranscriptomics to characterize the molecular responses of an iron-limited eukaryotic plankton community to experimental iron enrichment at Ocean Station Papa (OSP; 50° N, 145° W) in the NE Pacific Ocean, a well-characterized iron-limited region (21). A combination of 454 GS FLX and SOLiD sequencing platforms was used to identify differentially expressed (DE) genes from plankton communities within ambient (iron-limited) waters and from incubations with and without added iron. Normalization techniques allowed us to distinguish between potential shifts in taxonomic cell abundance and gene expression patterns (22). We focus on the changes in gene transcripts in diatoms as they displayed the largest physiological response to the iron enrichment. Transcripts derived from an isolate of *Pseudo-nitzschia granii* (a dominant diatom in the iron-enriched community) obtained from the microcosm experiment were included in the analyses to complement existing eukaryotic organism reference databases.

Results and Discussion

Offshore waters along the Line P transect and at OSP had lower mean maximum photochemical yields (F_V:F_M) relative to the onshore stations and contained low phytoplankton biomass (<0.3 $\mu g \cdot L^{-1}$) dominated by small cells (*SI Appendix*, Fig. S1). At the time of sampling, chemical features in the upper mixed layer at OSP were characteristic of iron-limited, HNLC waters with relatively low dissolved iron concentrations (0.03 $nmol \cdot L^{-1}$) and high NO_3^-, $Si(OH)_4$, and PO_4^{3+} concentrations (Table 1). An iron enrichment was simulated by incubating 240 L of surface seawater with 4 $nmol \cdot L^{-1}$ $FeCl_3$ (plusFe) alongside control incubations with no added iron (Ctl). The mean F_V:F_M of the ambient waters was 0.51 and increased to 0.63 within 98 h following iron enrichment, a value typically associated with iron-replete, healthy phytoplankton cells (23, 24) and similar to values measured in surface waters of the coastal stations along the Line P transect (*SI Appendix*, Fig. S1). Within this same timeframe, total chlorophyll *a* concentrations reached 2.4 $\mu g \cdot L^{-1}$ in plusFe, which were 2.7-fold higher than the Ctl and 11-fold higher than ambient seawater concentrations. On the basis of flow cytometry measurements, the phytoplankton assemblage shifted from one dominated by the ~2-μm cyanobacteria *Synechococcus* and a population of ultraphytoplankton (likely chlorophytes) in the ambient community to one dominated by large cells with an 8-fold increase in numbers of the largest cell type (Table 1). Flow cytometry-sorted cells from this larger population contained pennate diatoms, primarily of the genera *Pseudo-nitzschia* and *Fragilariopsis* (*SI Appendix*, Fig. S2). Nitrate concentrations continued to decline in the plusFe sample after 98 h, indicating that at the time of sampling, the iron-stimulated phytoplankton community was in the early to midstages of bloom development. On the basis of continuous flow cytometry measurements in parallel iron-enriched incubation experiments, these large cells peaked in cell abundance on day 5 (~120 h), at which time NO_3^- concentrations were fully depleted (25). A modest, yet significant increase in phytoplankton biomass in the Ctl was observed compared with that in the ambient waters, possibly resulting from changes in light conditions, minor iron contamination, and/or reductions in grazing pressure associated with the incubations. In the Ctl, F_V:F_M declined to 0.40 within 98 h following incubation, likely due to shifts in the composition of the iron-limited phytoplankton community with a reduced dominance of cyanobacteria (Table 1).

The plankton community response to the addition of iron was investigated using a comparative metatranscriptomics approach. Metatranscriptomics captures changes in both species composition and gene expression patterns in response to experimental conditions. To minimize the impacts of changes in species abundances on inferring gene expression, the trimmed mean of *M* values (TMM) normalization method (22) was applied to the metatranscriptomic read counts and used to calculate the normalization ratio (*Materials and Methods*). Inherent to this method is the assumption that expression of a large proportion of genes in a given organism remains unchanged regardless of the treatment or condition. Predicted genes that displayed a significant difference (adjusted *P* value <0.05) in the read count ratio between

Table 1. Comparison of chemical and biological properties in the ambient seawater (10 m) at Ocean Station Papa (50° N, 145° W), the unamended-control and the iron-enriched incubations after 98 h

Parameter	Ambient	Control	PlusFe
Dissolved and particulate nutrients			
NO_3^- ($\mu mol \cdot L^{-1}$)	15.43 ± 0.25	14.43 ± 0.15	12.90 ± 0.20
$Si(OH)_4$ ($\mu mol \cdot L^{-1}$)	23.35 ± 0.17	22.58 ± 0.21	21.57 ± 0.12
PO_4 ($\mu mol \cdot L^{-1}$)	1.32 ± 0.01	1.23 ± 0.02	1.09 ± 0.03
Particulate organic carbon ($\mu mol \cdot L^{-1}$)	12.43 ± 0.35	17.36 ± 1.00	23.21 ± 0.53
Particulate organic nitrogen ($\mu mol \cdot L^{-1}$)	1.75 ± 0.16	2.46 ± 0.22	3.36 ± 0.12
Biogenic silica ($\mu mol \cdot L^{-1}$)	0.60 ± 0.04	1.47 ± 0.38	2.43 ± 0.10
Chlorophyll *a* concentrations and photosynthetic efficiency			
Total Chl *a* ($\mu g \cdot L^{-1}$)	0.21 ± 0.04	0.89 ± 0.19	2.36 ± 0.20
<5 μm Chl *a* ($\mu g \cdot L^{-1}$)	0.10 ± 0.02	0.40 ± 0.14	0.69 ± 0.06
≥5 μm Chl *a* ($\mu g \cdot L^{-1}$)	0.11 ± 0.04	0.48 ± 0.12*	1.67 ± 0.19*,†
F_V:F_M	0.51 ± 0.02	0.40 ± 0.01	0.63 ± 0.01
Phytoplankton cell concentrations‡			
Synechococcus ($cells \cdot mL^{-1}$)	18,941 ± 3265	6,498 ± 9,719	1,617 ± 451*
Picophytoplankton ($cells \cdot mL^{-1}$)	195 ± 65	384 ± 196	318 ± 196
Ultraphytoplankton ($cells \cdot mL^{-1}$)	6,996 ± 1272	3,522 ± 4,299	888 ± 275
Nanophytoplankton 1 ($cells \cdot mL^{-1}$)	990 ± 21	1,818 ± 229	2,427 ± 769*
Nanophytoplankton 2 ($cells \cdot mL^{-1}$)§	214 ± 119	1,026 ± 413	1,730 ± 740*

Values are means ± SD (n = 3).
*Significantly different from the ambient ($P < 0.05$).
†Significantly different from the control ($P < 0.05$).
‡Size classes of phytoplankton populations are described in Ribalet et al. (25).
§Pennate diatoms such as *Pseudo-nitzschia* spp. are present in this population.

libraries relative to the TMM were considered either overrepresented (above TMM) or underrepresented (below TMM) following iron enrichment and we refer to them here as significantly differentially expressed.

A combination of 454 and SOLiD metatranscriptome sequencing was used to identify differentially expressed genes in the ambient seawater (ambient library), in the unamended control (control library), and under iron-enriched conditions (plusFe library), with the 454 sequences serving as reference for SOLiD read alignment. Sequence libraries using the 454 platform were created using RNA collected from the ambient waters and the iron-enriched conditions. Approximately 57% of the 224,211 reads in the ambient 454 library and 59% of the 173,198 reads in the plusFe 454 library were assigned to a taxonomic group (tBLASTx, *e*-value $<10^{-3}$) (*SI Appendix*, Table S1). About a quarter of the total reads in both 454 libraries could be assigned a putative function on the basis of sequence homology (BLASTx, *e*-value $<10^{-3}$). More than two-thirds of predicted genes from both 454 libraries were represented by a single read; <0.1% of the predicted genes were represented by >100 reads (*SI Appendix*, Table S1). A total of 121 predicted genes (46 overrepresented and 75 underrepresented) were significantly DE in the 454 libraries following iron enrichment (*SI Appendix*, Fig. S3).

Detection of DE genes was greatly enhanced by SOLiD sequencing using the same RNA samples, which increased the total number of sequence reads for analysis by ~60-fold and 100-fold in the ambient and plusFe SOLiD libraries, respectively (*SI Appendix*, Fig. S3). Taxonomic distributions of sequence reads were consistent between the 454 and SOLiD libraries with a few exceptions in some phyla (*SI Appendix*, Fig. S4). SOLiD sequencing of the unamended, control condition identified transcripts associated with potential bottle artifacts (e.g., changes in light climate and/or grazing pressure) resulting from the incubations. Genes with transcripts significantly DE in the plusFe library relative to both the ambient and control libraries were considered to be specifically responsive to iron enrichment. Diatom, haptophyte, and arthropod sequences dominated the ambient and control SOLiD libraries, averaging 25%, 21%, and 18%, respectively, of the taxonomically assigned reads with a mixture of other phyla making up the remainder (Fig. 1). After iron enrichment, a doubling in the proportion of reads assigned to diatoms (Fig. 1) was due largely to an increase in transcripts assigned to *P. granii* (*SI Appendix*, Fig. S5). Sequences assigned to all other phyla comprised either similar or lower proportions in the plusFe library compared with the ambient and control libraries.

Alignment of SOLiD reads to the 454 reference reads resulted in a significant increase in sensitivity, with a majority of predicted genes represented by more than a single read and ~5% of the predicted genes represented by >100 reads. Over an order of magnitude more DE genes were detected with this combined approach: Transcripts for 2,845 (1,442 overrepresented and 1,403 underrepresented) and 3,888 (1,537 overrepresented and 2,351 underrepresented) predicted genes were significantly DE in the plusFe vs. ambient and plusFe vs. control SOLiD libraries, respectively (*SI Appendix*, Fig. S3). Further analysis of these DE genes between SOLiD libraries was restricted to subsets of four taxonomic groups: the three dominant phytoplankton groups, diatoms, haptophytes, and chlorophytes (Fig. 2), and the combined heterotrophic phyla (*SI Appendix*, Fig. S6). Transcripts for genes discussed below displayed similar trends and were significantly DE in the plusFe library relative to both the ambient and control libraries and are thus considered iron responsive (*SI Appendix*, Fig. S7).

Almost half the iron-responsive genes in diatoms are annotated as encoding either hypothetical or predicted proteins, highlighting that much of the transcriptional response of diatoms to iron involves genes encoding proteins of unknown function. Interestingly, several of these unknown genes are highly expressed under iron-starved conditions within the diatom *P. tricornutum* and encode iron-starvation–induced protein 1 (ISIP1), ISIP2A, and ISIP3 (Fig. 2) (6). In our metatranscriptomic libraries, transcripts for these genes were identified in diatoms, haptophytes, and chlorophytes and most were significantly underrepresented following iron enrichment.

Each taxonomic group displayed distinct transcriptional responses to iron enrichment, with only three genes shared among the dominant phytoplankton groups (*SI Appendix*, Fig. S8). Diatoms displayed the broadest transcriptional response to iron enrichment: Transcripts for 98 predicted genes were underrepresented and transcripts for 285 predicted genes were overrepresented following iron enrichment. About 10% of the DE diatom genes were also DE by haptophytes. In chlorophytes, a roughly equal number of predicted genes were underrepresented and overrepresented following iron enrichment.

Iron-Rich Metabolic Pathways. Iron serves as an electron carrier in numerous metalloproteins due to its high reduction–oxidation potential (26). In phytoplankton, a high proportion of iron is contained within components of the photosynthetic apparatus and the mitochondrial electron transport chain (27, 28). Iron limitation appears to lead to a preference for oxidative pathways such as fatty acid metabolism, mitochondrial oxidative phosphorylation, and other components involved in cellular respiration (Fig. 3). Addition of iron led to significant increases, particularly in diatoms, for transcripts of genes associated with several photosynthesis-related pathways: light-dependent photosynthetic reaction centers photosystems I and II; the light-harvesting complexes; the protein complex cytochrome b_6f; and subunits of f-type ATP synthases, chlorophyll biosynthesis, and steps of carotenoid biosynthesis (Fig. 3 and *SI Appendix*, Table S2). Ribosomal and RNA polymerase activity appears to be unaffected by iron (*SI Appendix*, Fig. S9).

After addition of iron, the short-term response by diatoms appears to be a continued utilization of what are generally considered inefficient, iron-free versions of certain photosynthetic proteins (14). Transcripts associated with flavodoxin were abundant in all libraries and overrepresented after iron enrichment (Fig. 2), suggesting that some or all diatoms continue relying on this protein for photosynthetic electron transfer rather than switching entirely to the use of the iron-containing version, ferredoxin. Gene homologs for both flavodoxin and ferredoxin are present in the four diatoms with available whole-genome sequences. Diatom transcripts for plastocyanin were also highly abundant under both conditions, but were underrepresented after iron enrichment. Few diatom transcripts were detected for the iron-requiring counterpart cytochrome c_6 in any of the sequence libraries. Laboratory studies suggest that coastal species use cytochrome c_6 whereas open ocean species use plastocyanin, with some diatoms containing both proteins (e.g., *P. granii* and *Fragi-*

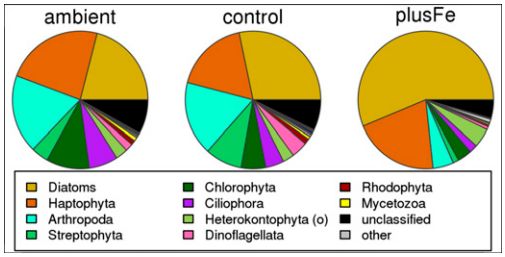


Fig. 1. Phylum-based taxonomic proportions of transcripts in the ambient, control, and plusFe SOLiD sequence libraries. Heterkontophyta have been separated into diatoms and other (o) heterokontophytes. Transcripts of chloroplast-encoded genes are omitted. Unclassified transcripts are SOLiD reads that align to 454 sequences with no taxonomic hit (tBLASTx, *e*-value $\leq 10^{-3}$). For additional phyla comprising "other", see *SI Appendix*, Fig. S5.

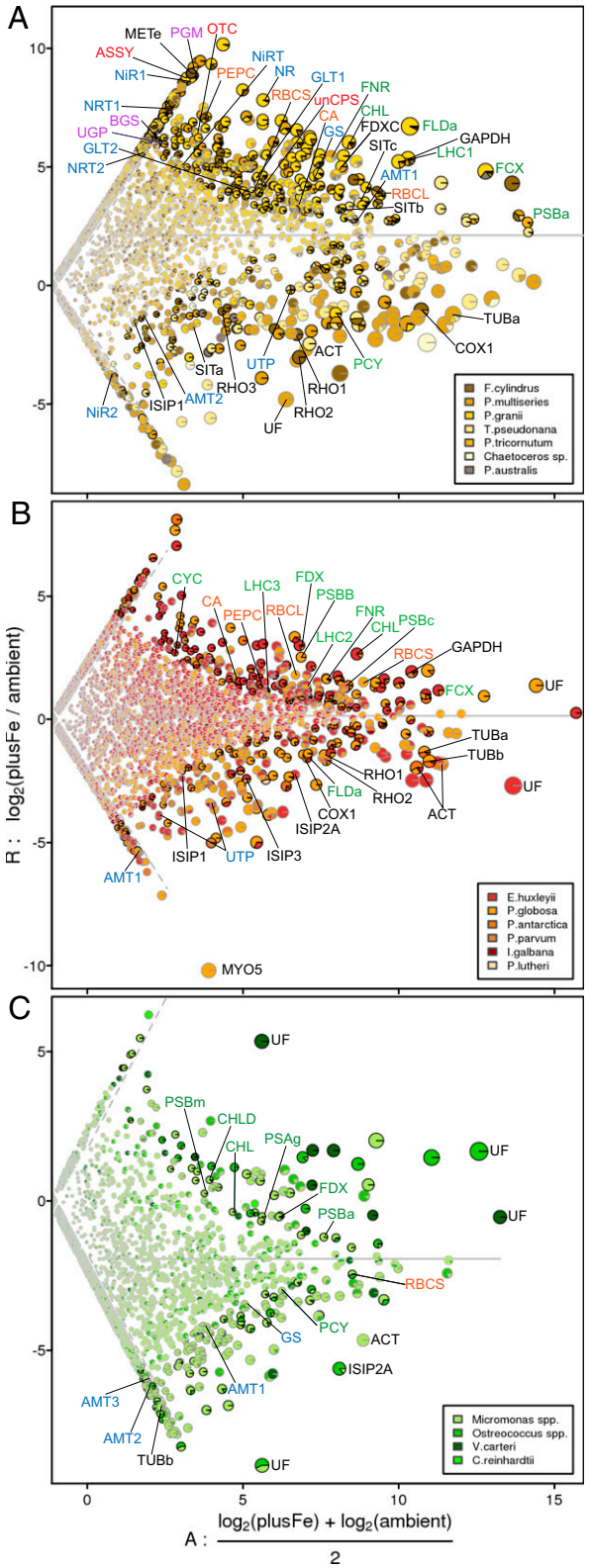


Fig. 2. Microbial Assemblage Normalized Transcript Analysis (MANTA) ratio-average (RA) plots of taxa-specific differential gene expression in response to iron enrichment for the major phytoplankton taxa: (*A*) diatoms, (*B*) haptophytes, and (*C*) chlorophytes. Each circle represents a collection of transcripts assigned to a predicted gene on the basis of homology (BLASTx, *e*-value $\leq 10^{-3}$) and annotated via KEGG orthology. Plotted are the fold-change ratio (*R*) and the average (*A*) of read counts in the plusFe and ambient SOLiD libraries for a given gene. Circles increase in size with absolute values of its coordinates to optimize visibility. Gene circles are colored according to taxonomic affiliation of the reference sequences. Gray horizontal line indicates the trimmed mean of fold-change values (TMM, *Materials and Methods*). Significantly DE genes above or below the TMM indicate over- or underrepresentation, respectively, after iron enrichment. Genes unique to one library are added into the plot left of the gray dashed lines. Gene circles with black borders indicate transcripts that are significantly DE (adjusted *P* value <0.05) in plusFe relative to both the ambient and control libraries; gene circles with dark gray borders indicates transcripts that are significantly DE in the plusFe vs. ambient libraries only. Gene circles with light gray borders are not DE. Full genus names are provided in *SI Appendix*, Fig. S5. Select gene names are colored and labeled according to predicted protein functions and grouped as follows: photosynthesis and related proteins (green): CHL, magnesium chelatase; CYC, cytochrome c_6; FCX, fucoxanthin; FDX, ferredoxin; FLD, flavodoxin I; FNR, ferredoxin-NADP$^+$ reductase; LHC, light-harvesting complex protein; PCY, plastocyanin; PSA, photosystem I protein; PSB, photosystem II protein. Carbon fixation proteins (orange): CA, carbonic anhydrase; PEPC, phosphoenolpyruvate carboxylase; RBC-L (large) and -S (small), ribulose-bisphosphate carboxylase. Nitrogen assimilation proteins (blue): AMT, ammonium transporter; GLT1, glutamate synthase (NADPH); GLT2, glutamate synthase (ferredoxin); GS, glutamine synthetase; NR, nitrate reductase; NiRc, nitrite reductase (ferredoxin); NiRb, nitrite reductase (NADPH); NiRT, nitrite transporter; NRT, nitrate transporter; UTP, urea transporter. Chrysolaminarin biosynthesis proteins (purple): BGS, chrysolaminarin synthase; PGM, phosphoglucomutase; UGP, UDP-glucose pyrophosphorylase. Urea cycle proteins (red): ASSY, argininosuccinate synthase; OTC, ornithine carbamoyltransferase; unCPS, carbamoyl-phosphate synthase. Other proteins (black): ACT, actin; METe, cobalamin-independent methionine synthase; COX1, cytochrome *c* oxidase 1; FDXC, ferredoxin component; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; ISIP, iron starvation-induced protein; MYO5, myosin V; RHO, rhodopsin; SIT, silicic acid transporter; TUB, tubulin; UF, unknown function. Where provided, lowercase letters indicate subunits unless defined otherwise. Note *y*-axis scales are different.

lariopsis cylindrus). Our results suggest that natural communities of oceanic diatoms preferentially rely on plastocyanin rather than cytochrome c_6, even when iron is available, substantially reducing their need for iron while increasing the need for copper, which is more abundant in oceanic waters (20).

Diatoms also continued to use iron-free versions of superoxide dismutases (SODs) to minimize oxidative stress. After iron enrichment, diatoms appeared to switch from a reliance on Mn-requiring SODs to a greater reliance on Cu/Zn SODs. Diatom transcripts associated with Fe/Mn SODs remained at low levels regardless of their iron status. In plants, Cu/Zn SODs provide less protection than iron-containing SODs (29), suggesting that oceanic diatoms preferentially maintain a proteome with less reliance on iron despite potential inefficiencies.

An additional way for cells to minimize iron requirements is to use reduced, rather than oxidized forms of N; reduction of nitrate requires iron-containing enzymes such as nitrate and nitrite reductases. As a result, the iron requirement for cultured phytoplankton grown on nitrate is 1.6 times that of growth on ammonium (26). As expected for iron-limiting conditions, diatoms appear to rely on reduced nitrogen compounds such as ammonium and urea despite the presence of high concentrations of available nitrate (Fig. 2). After iron enrichment, overrepresented transcripts include those associated with nitrate and nitrite transporters and nitrate and nitrite reductases as well as the enzymes, glutamine synthetase and both NADPH-dependent and ferredoxin-dependent glutamate synthetases, required for nitrate assimilation (Fig. 4). These results complement previous microcosm and large-scale iron fertilization experiments conducted in the NE Pacific that indicate increased nitrate uptake rates (30, 31) and nitrate reductase protein activity (32) within diatoms after iron addition.

Together, these results suggest that in the open ocean where iron inputs are ephemeral, diatoms continuously rely on non-iron–containing, less efficient proteins rather than carrying out a

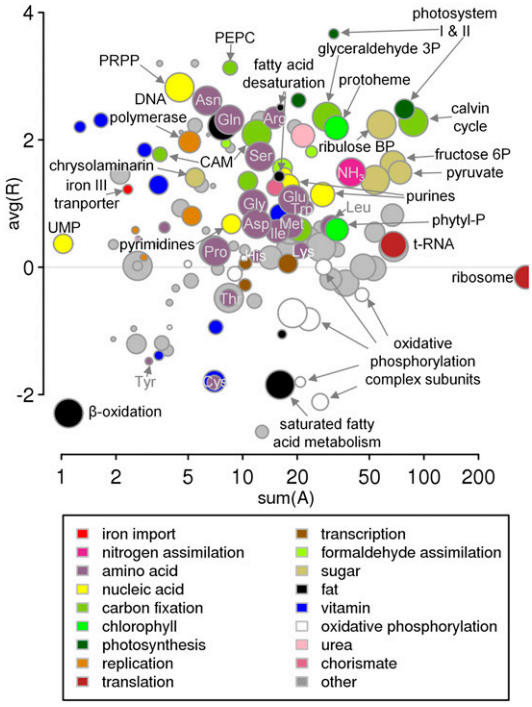


Fig. 3. Metabolic-pathway color-coded scatter plot of the average iron-induced fold change in transcript abundance in diatoms. Plotted are the average adjusted ratios [avg(*R*)] of transcripts in the plusFe SOLiD library over the ambient/control SOLiD libraries vs. the sum of transcripts for all genes in a module [sum (*A*)]. Each point represents the molecular product of a multienzyme/protein biosynthetic chain or cycle determined by counts of transcripts mapping to KEGG genes and modules. Circle size indicates the percentage of enzymes in each KEGG module with reads mapping to the underlying KEGG genes. The largest circles are 100% complete modules whereas the smallest have at least a single gene mapped. Gray horizontal line indicates no change in avg(*R*). Both axes labels are in log scale (base 2). Scatter plots for the other dominant taxonomic groups are provided in *SI Appendix*, Fig. S9. Select modules are labeled. Amino acids are identified with their three-letter abbreviations. Others: CAM, crassulacean acid metabolism; PRPP, phosphoribosyl pyrophosphate; UMP, uridine monophosphate biosynthesis.

widespread replacement of these proteins with their iron-containing, more efficient counterparts. Hedging their bets by maintaining a reduced iron-requiring protein repertoire would allow diatoms to rapidly reacclimate to iron-limiting conditions. This strategy would also allow them to take advantage of available nitrate by partitioning newly acquired iron toward nitrate reduction and subsequent ammonium assimilation for which there are no iron-free protein equivalents. Diatoms may compensate for a continued dependence on these less efficient, iron-free proteins by overproducing photosynthetic antennae, overcoming a short-term electron transport inefficiency with a longer-term investment in increased photon focusing infrastructure (Fig. 4).

The distinctiveness of the diatom response becomes even more apparent compared with the haptophytes, the other dominant member of the phytoplankton community after iron enrichment. Haptophytes display a more typical response to iron enrichment by switching to use of the iron-containing proteins ferredoxin and cytochrome c_6 (Fig. 2). Similarly, haptophytes appear to switch to use of Fe/Mn SODs. Of particular interest is that after iron enrichment, no haptophyte or chlorophyte transcripts were detected for any genes involved in nitrate transport or reduction (Fig. 4) despite the presence of all genes necessary to carry out these processes within their genomes (33, 34). This result suggests that both groups are unresponsive to the large concentrations of available nitrate following iron enrichment, perhaps because their newly acquired iron is partitioned for other uses such as a more efficient photosynthetic electron transfer.

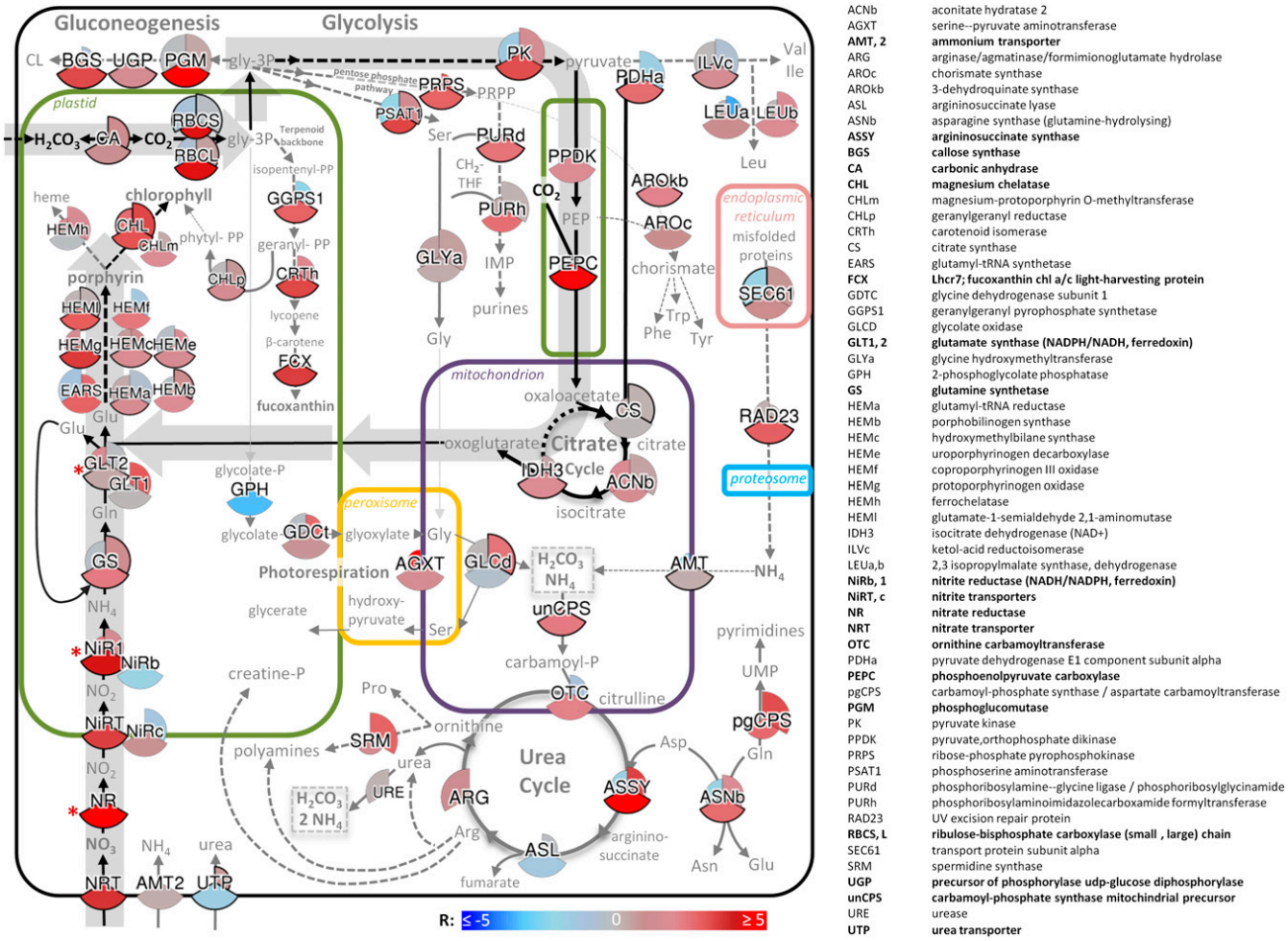
Diatom Bloom Formation in Response to Iron. The addition of iron to diatoms stimulates what appears to be a system-wide pattern of gene expression in clear alleviation of chlorosis and in preparation for rapid cell division (Fig. 4). Specifically, we observed significant increases in transcripts for genes encoding proteins that catalyze the biosynthesis of nucleic and amino acids, sugars, and chlorophyll (and its precursors, i.e., porphyrin) as well as the prerequisites for long-chain polyamines, which may play an important role in silica precipitation for diatom cell wall formation (Fig. 3). One of the strongest, iron-influenced pathways appears to be concentrated on C fixation: RuBisCO (RBC) and carbonic anhydrase (CA) were overrepresented following iron enrichment (35) (Fig. 4). Transcripts for phosphoenolpyruvate carboxylase (PEPC) were also overrepresented following iron enrichment. PEPC may play a role in C4–carbon fixation (33) or alternatively may provide anaplerotic C input for citric acid cycle intermediates that support more rapid N assimilation (36) (Fig. 4). Transcripts for genes that encode three of the known required enzymes for chrysolaminarin biosynthesis were overrepresented, suggesting accumulation and storage of reduced C as carbohydrates following iron enrichment (35) (Figs. 2 and 4).

The urea cycle of diatoms is hypothesized to play a key role in C and N assimilation by facilitating turnover and redistribution of metabolites involved in metabolic recovery from prolonged nitrogen limitation (37). Iron enrichment resulted in the overrepresentation of transcripts for genes encoding three of the five urea-cycle enzymes: a mitochondrial carbamoyl-phosphate synthetase (unCPS), ornithine carbamoyltransferase (OTC), and argininosuccinate synthase (ASSY) (Fig. 4). These enzymes are involved in the first three steps of the urea cycle. Transcripts for the gene encoding argininosuccinate lyase, involved in the next step, were also detected but were not significantly DE. Our data suggest that components of the diatom urea cycle likely play a significant role in facilitating the iron response leading to bloom formation following the alleviation of iron limitation and/or low-iron–induced N limitation. More specifically, unCPS likely plays an important role in mitochondrial recovery of inorganic C and N from both photorespiration (38) and UV damaged, misfolded, or otherwise degraded proteins (37).

Interestingly we detected few diatom transcripts for arginase and urease, suggesting either the disproportionate importance of unCPS, at the time of sampling, for increasing levels of fixed C and N in the form of urea cycle intermediates or an alternative fate for these intermediates. Newly recycled C and N appear to be directed toward the formation of more reduced, larger N/C-rich molecules such as arginino succinate and arginine that may themselves serve as important N storage compounds (39) rather than temporary intermediates for longer-term N storage molecules such as creatine phosphate (38). Metabolites derived from the urea cycle are also used to make polyamines such as putrescine and other long-chain polyamines that facilitate silica precipitation in diatoms (40).

Our pathway analysis of gene expression changes in response to iron enrichment suggests diatoms orchestrate a distinctive and efficient growth response leading to rapid bloom formation. In combination with an enhanced capacity to store iron (24), oceanic diatoms appear to prioritize newly available iron to reduce N that likely provides them with an ability to continue growth and cell division long after external iron concentrations have returned to low, background levels. Consequently, oceanic diatoms must also invoke strategies to subsist within the prolonged periods of iron limitation to enable them to maintain adequate seed populations to respond to ephemeral iron inputs.

Rhodopsins in Eukaryotic Phytoplankton. In the marine environment, proteorhodopsins are generally known as prokaryotic pro-



ACNb aconitate hydratase 2
AGXT serine--pyruvate aminotransferase
AMT, 2 **ammonium transporter**
ARG arginase/agmatinase/formimionoglutamate hydrolase
AROc chorismate synthase
AROkb 3-dehydroquinate synthase
ASL argininosuccinate lyase
ASNb asparagine synthase (glutamine-hydrolysing)
ASSY **argininosuccinate synthase**
BGS **callose synthase**
CA **carbonic anhydrase**
CHL **magnesium chelatase**
CHLm magnesium-protoporphyrin O-methyltransferase
CHLp geranylgeranyl reductase
CRTh carotenoid isomerase
CS citrate synthase
EARS glutamyl-tRNA synthetase
FCX **Lhcr7; fucoxanthin chl a/c light-harvesting protein**
GDTC glycine dehydrogenase subunit 1
GGPS1 geranylgeranyl pyrophosphate synthetase
GLCD glycolate oxidase
GLT1, 2 **glutamate synthase (NADPH/NADH, ferredoxin)**
GLYa glycine hydroxymethyltransferase
GPH 2-phosphoglycolate phosphatase
GS **glutamine synthetase**
HEMa glutamyl-tRNA reductase
HEMb porphobilinogen synthase
HEMc hydroxymethylbilane synthase
HEMe uroporphyrinogen decarboxylase
HEMf coproporphyrinogen III oxidase
HEMg protoporphyrinogen oxidase
HEMh ferrochelatase
HEMl glutamate-1-semialdehyde 2,1-aminomutase
IDH3 isocitrate dehydrogenase (NAD+)
ILVc ketol-acid reductoisomerase
LEUa,b 2,3 isopropylmalate synthase, dehydrogenase
NiRb, 1 **nitrite reductase (NADH/NADPH, ferredoxin)**
NiRT, c **nitrite transporters**
NR **nitrate reductase**
NRT **nitrate transporter**
OTC **ornithine carbamoyltransferase**
PDHa pyruvate dehydrogenase E1 component subunit alpha
PEPC **phosphoenolpyruvate carboxylase**
pgCPS carbamoyl-phosphate synthase / aspartate carbamoyltransferase
PGM **phosphoglucomutase**
PK pyruvate kinase
PPDK pyruvate,orthophosphate dikinase
PRPS ribose-phosphate pyrophosphokinase
PSAT1 phosphoserine aminotransferase
PURd phosphoribosylamine--glycine ligase / phosphoribosylglycinamide
PURh phosphoribosylaminoimidazolecarboxamide formyltransferase
RAD23 UV excision repair protein
RBCS, L **ribulose-bisphosphate carboxylase (small , large) chain**
SEC61 transport protein subunit alpha
SRM spermidine synthase
UGP **precursor of phosphorylase udp-glucose diphosphorylase**
unCPS **carbamoyl-phosphate synthase mitochindrial precursor**
URE urease
UTP **urea transporter**

Fig. 4. Schematic of the predicted iron-induced, metabolic flow of nitrogen (left) and carbon (top) into diatom biosynthetic components: glyceraldehyde and glycerate 3 phosphate (gly-3P), amino acids (three-letter abbreviations), chrysolaminarin (CL), nucleotides, chlorophyll, and polyamines. Metabolic reactions are illustrated as fixed-angle wedged pie charts. Relative transcript abundance is indicated by wedge radius (proportional to log SOLiD library averaged read counts) for the three major phytoplankton groups: haptophytes (right), chlorophytes (left), and diatoms (bottom). Wedge color gradients: blue indicates underrepresented in plusFe vs. ambient/control libraries; red, overrepresented in plusFe vs. ambient/control libraries; and gray, no change after iron addition. Thick wedge outline indicates significant differential expression. Genes with known iron dependencies are labeled with a red asterisk. Localization of gene products is based on predictions for diatoms from previous research. Dashed lines indicate multireaction biosynthesis pathways. Proposed biosynthesis of chlorophyll is indicated via the thick black reaction lines with gray background. Gene names in boldface type are also indicated in Fig. 2.

teins that act as light sensors for phototaxis or as light-driven proton pumps that generate a proton motive force to drive synthesis of ATP, cell motility, and/or import of small molecules (41, 42). Recently, proteorhodospin has been proposed to fuel light-driven ATP synthesis in eukaryotes on the basis of the identification of rhodopsin-type genes in several heterotrophic and mixotrophic dinoflagellates (43, 44). All other rhodopsins previously identified in eukaryotic phytoplankton (chlorophytes and cryptomonads) appear to be sensory rhodopsins (43, 45).

We identified putative rhodopsins in the oceanic diatoms *P. granii* and *F. cylindrus* and the haptophyte, *Phaeocystis globosa* (Fig. 5). Interestingly, we did not identify this gene in the coastal diatoms *Pseudo-nitzschia multiseries*, *T. pseudonana*, or *P. tricornutum*; the coastal pelagophyte *Aureococcus anophagefferens*; or two oomycetes *Phytophthora* spp., all of which belong to the heterokontophytes and have whole-genome sequences available. Diatom rhodopsins are more similar to the proton-pump variants rather than to sensory rhodopsins as they contain the retinal-binding pocket and an essential residue required for proton transport (*SI Appendix*, Fig. S10). Rhodopsin-like environmental transcripts encode for proteins that form a clade with haptophytes, dinoflagellates, and diatoms, all of which cluster with xanthorhodopsins from the bacteria *Gloeobacter* and *Salinibacter* (46) (Fig. 5). Thus, the erratic presence of rhodopsins among the photosynthetic marine eukaryotes, coupled with their clustering within a bacterial clade, suggests these eukaryotic plankton groups may have acquired rhodopsins through lateral gene transfer events (47, 48).

What is the possible role of light-driven proton pumps within the physiology of open ocean photosynthetic diatoms? Rhodopsin transcripts assigned to *P. granii* and *F. cylindrus* (as well as to *P. globosa*) were highly abundant in the ambient and control libraries, suggesting a putative role for rhodopsins in coping with iron deficiency (Fig. 2). In comparing the energy and N costs between oxygenic photosynthesis and rhodopsins for photon absorption and the energized products formed, Raven (49) suggested there are few situations where photochemical generation of ion gradients and ATP by rhodopsins would be selectively favored. Due to the high iron requirements in oxygenic photosynthesis and lack of trace metals required in rhodopsins, one of these situations was suggested to be in low-iron environments, which is consistent with our findings (49).

Previous studies have also shown that open ocean phytoplankton can use alternate light- and energy-management strategies when faced with nutrient-poor waters. For example, in cyano-

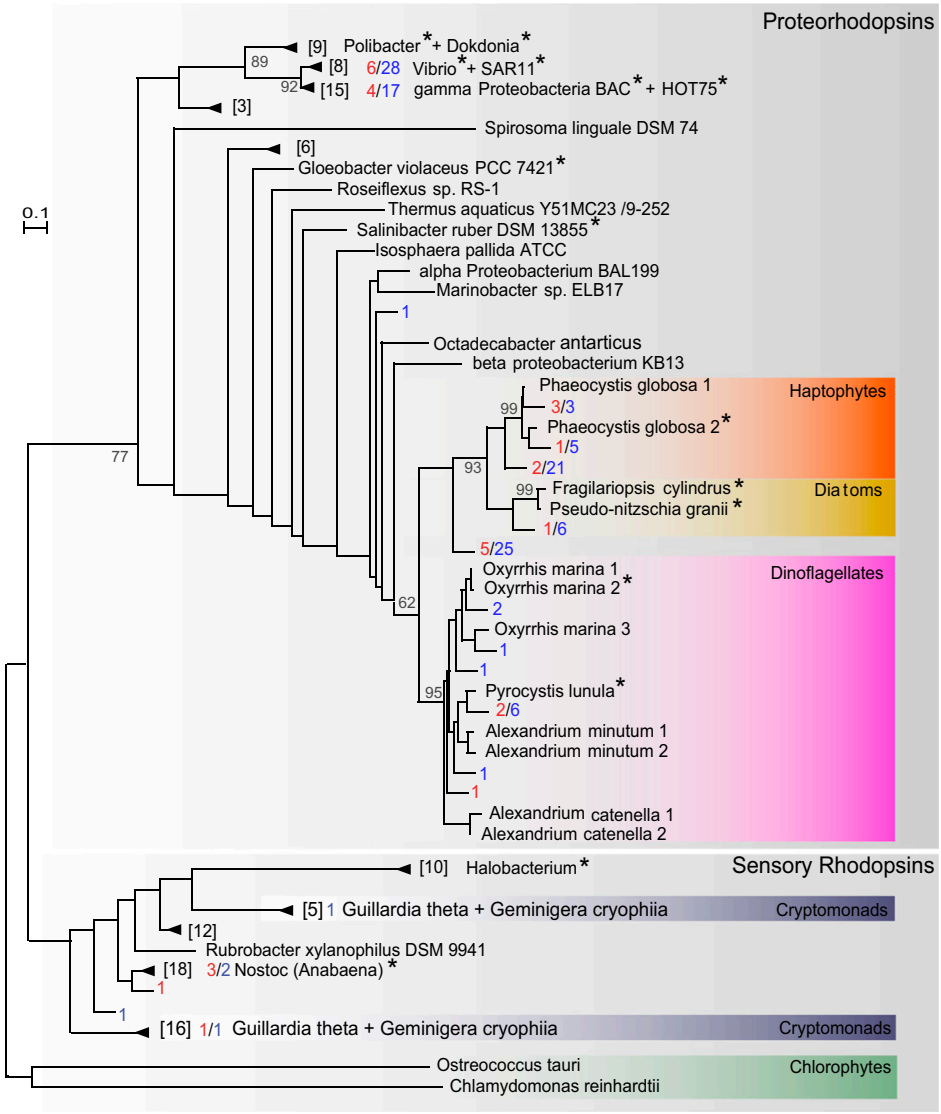


Fig. 5. Phylogenetic tree of rhodopsins. Reference tree is composed of proteorhodopsins and sensory rhodopsins from 114 taxa. Numbers at branch tips represent placement of environmental transcripts from the 454 sequence libraries: plusFe (red) and ambient (blue). Some nodes have been collapsed to simplify the tree with the numbers in brackets indicating the number of reference taxa at collapsed nodes (see Dataset S6 for a list of species at collapsed nodes). Colored boxes highlight eukaryotic plankton groups. Taxa with an asterisk are found in the rhodopsin alignment in *SI Appendix*, Fig. S10. Bootstrap values >50 are indicated at the branch points.

bacteria from low-iron environments, electrons derived from light energy harvested by the photosystem II complex (PSII) may be diverted to a plastoquinol oxidase (PTOX), away from the iron-rich photosystem I complex (reviewed in Zehr and Kudela) (50). It is possible that rhodopsin in open ocean diatoms plays a similar role, allowing for the production of ATP from light energy without the involvement of the iron-demanding PSI. Unlike PTOX, rhodopsin would not be dependent on PSII for light harvesting, further reducing the iron requirement for ATP synthesis. However, production of ATP from PTOX or rhodopsin presents the same challenge: a source of reductants, such as NADPH, which cells require for both C and N assimilation. Alternatively, eukaryotic rhodopsins may replace or supplement the proton motive force generated via V-ATPase hydrolysis of ATP (43), particularly when photosynthesis, and the generation of ATP, is compromised by low-iron conditions. V-ATPases hydrolyze ATP to generate a proton motive force. In diatoms, V-ATPases may help maintain the low pH within the silica deposition vesicle needed to create the silica cell wall (51). Many V-ATPase subunit transcripts in diatoms were overrepresented following iron enrichment.

Summary. Our study provides unique insights into the molecular underpinnings for the large physiological response of diatoms to iron enrichment. By combining next-generation sequencing technologies, we distinguished the differential expression of thousands of genes within a natural, diverse plankton assemblage. Our findings suggest that the immediate response of oceanic diatoms to iron enrichment is to continue to use iron-free proteins rather than switching over en masse to the iron-containing functional equivalents. This strategy may provide for a more rapid acclimation to a return to low-iron conditions. The selective partitioning of newly acquired iron into iron-requiring proteins involved in nitrate reduction and assimilation could explain how diatoms dominate the consumption of nitrate following iron enrichment. Many of the genes in diatoms that were most significantly overrepresented following iron enrichment belonged to pathways associated with the reduction of organic molecules to facilitate chlorophyll bio-

synthesis and storage of fixed N and C. The discovery of certain diatoms containing rhodopsin-related genes that may play a role in helping them cope with low-iron conditions provides further evidence of how diatoms have evolved exceptional capabilities that contribute to their success in marine environments.

Materials and Methods

Study Area, Experimental Design, and Water Sampling Collection. The microcosm iron enrichment experiment was performed from June 8 to June 13, 2008 during the Line P cruise in the NE Pacific Ocean from May 28 to June 17, 2008 onboard the Canadian Coast Guard Ship J. P. Tully. Seawater was collected at OSP (50° N, 145° W) from a depth of 10 m (corresponding to 30% of incident irradiance), using a trace-metal clean sampling system that consisted of a Teflon air bellows pump and Teflon-lined Kevlar tubing (0.75 inches in diameter) attached to Kevlar line. For the ambient condition, seawater was immediately filtered for RNA and prepared for all ancillary measurements as described below. For the iron-enriched treatment, seawater was placed into 10-L flexible acid-cleaned polyethylene cubitainers within an on-deck trace-metal clean positive pressure flowhood. Cleaning protocols for the cubitainers included soaking the inside walls in 1.2 mol·L^{-1} hydrochloric acid (reagent grade) for 3 d followed by three rinses with Milli-Q H_2O, soaking in 1.2 mol·L^{-1} hydrochloric acid (trace metal grade) for 1 wk followed by three rinses with Milli-Q H_2O, and soaking in 0.1 mol·L^{-1} acetic acid (trace-metal grade). Before filling the cubitainers with seawater, the dilute acetic acid was removed and the cubitainers were rinsed thoroughly three times with ambient, low-iron seawater. In total, 30 cubitainers were prepared of which 27 were inoculated with 4 nmol·L^{-1} $FeCl_3$ and the remaining 3 served as unamended controls. Cubitainers were placed in on-deck plexiglass incubators with flow-through seawater to maintain near-ambient surface temperatures and were covered with neutral density screening to reduce irradiance to ∼30% of the incident. Seawater from 24 cubitainers (240 L) was reserved for the iron-enriched RNA sequence libraries (described below) whereas seawater from the remaining 6 cubitainers (3 iron-enriched replicates and 3 unamended control replicates) was used for all ancillary measurements: size-fractionated chlorophyll *a*, F_V:F_M, dissolved and particulate nutrient concentrations, and flow cytometry (*SI Appendix, SI Methods*).

Metatranscriptome Sample Preparation and Sequencing. For the ambient sequence libraries, seawater was collected using trace-metal clean techniques as described above and directly filtered onto Millipore Isopore membrane filters (0.4 μm pore size, 142 mm) by way of a peristaltic pump and tygon tubing. Filters were changed every 15 min or sooner if flow rate decreased. Filtering commenced at 12:00 AM [Pacific standard time (PST)] and ended at 3:00 AM (PST) on June 8, 2008. Filters were placed into Ziploc plastic bags, wrapped in aluminum foil, and immediately stored in liquid nitrogen. With a flow rate of 5–6 L·min^{-1}, an estimated volume of ∼1,000 L of seawater was filtered. For the plusFe and control sequence libraries, iron-amended seawater was incubated on deck as described above for 98 h. Immediately before dawn, seawater was removed from the incubators and stored in a dark, cool room until filtration. Incubated seawater was filtered using the same filtration setup and storage protocols as described for the ambient sequence libraries. Onshore, filters were stored at −80 °C until RNA extractions were performed.

For RNA extractions, filters were briefly thawed on ice. RNA was extracted from individual thawed filters, using the ToTALLY RNA Kit (Ambion) according to the manufacturer's protocol with the additional step that filter pieces were first vortexed in 7 mL of denaturation solution containing 0.5 mL of glass beads and the resulting lysate was centrifuged at 8,801 × *g* and 4 °C for 3 min. Trace DNA contamination was eliminated by DNase 1 (Ambion) digestion at 37 °C for 45 min. Polyadenosine [poly(A)$^+$] RNA was isolated with the Poly(A) Purist Mag Kit (Ambion) according to the manufacturer's protocol. For 454 pyrosequencing 5.9 μg (ambient library) and 9.3 μg (plusFe library) of poly(A)$^+$ selected mRNA were reverse transcribed into double-stranded cDNA (ds cDNA), using the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen). Sequence libraries were prepared according to protocols provided in Poinar et al. (52) and sequenced using a 454-GS-FLX as described in Miller et al. (53). Each library was sequenced on one full 454-GS-FLX plate. For the control SOLiD sequence library, due to insufficient yields of RNA, mRNA was linearly amplified using the MessageAmp II aRNA Amplification kit (Ambion). For SOLiD sequencing, 0.8 μg (ambient library) and 1.0 μg (plusFe and control libraries) of poly(A)+ selected mRNA were used for library preparations (see *SI Appendix, SI Methods* for SOLiD library preparations). For the ambient and plusFe SOLiD libraries, 41 million beads of each sample were loaded onto two spots of an eight-spot slide (one spot per library) and run on an Applied Biosystems (AB) SOLiD sequencer version 3 Plus. For the control SOLiD library, sequences were barcoded and run on an AB SOLiD sequencer version 4. Sequence library filtering and cleanup for both 454 and SOLiD sequencing are provided in *SI Appendix, SI Methods*.

Taxonomic Identification and Gene Annotation Pipeline. On the basis of lowest *e*-value (tBLASTx, *e*-value ≤10^{-3}), 454 sequence reads were assigned a taxonomic affiliation. The datasets examined were available marine eukaryotic genomes and ESTs, National Center for Biotechnology Information's EST database (excluded Fungi and Metazoa except Copepoda), and National Center for Biotechnology Information's nonredundant (nr) database (Dataset S1). An isolate of the diatom *P. granii* (UWOSP1E) was obtained from the iron-enriched treatment, and RNA isolated from cells grown under low-iron conditions in the laboratory was used to generate 454 sequence reads. ESTs were assembled using CABOG (v. 5.3) and used as a BLAST database. Sequence reads were mapped across reference sequences, with a slight bias for reads to align to the 3′ ends of genes (*SI Appendix*, Fig. S11). Positional distributions of sequence reads did not exhibit any apparent *e*-value biases. To provide consistent taxonomic classifications for all reference organisms with best hits to sequence reads, taxonomic classification of ranks was assigned using custom lookup tables (Dataset S2).

Potential function was assigned on the basis of best homology (BLASTx, *e*-value ≤ 10^{-3}) to proteins within the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the nr databases. For those sequences that displayed homology to "predicted", "hypothetical", or "unknown" proteins in KEGG, additional manual in silico annotation was performed with online search tools. For predicted genes in which a gene function was subsequently identified, a KEGG orthology (KO) was assigned (if possible) through a search in the KEGG database or by searching for best homology to annotated proteins in the nr database (Dataset S3). The top functional gene hits (if any) were assigned to each 454 sequence read on the basis of the lowest *e*-value, with priority given to hits to KEGG to standardize on KEGG's module and pathway hierarchy. A new composite gene identifier was created to track protein definitions where predicted genes with unknown functions were replaced with a suitable manual redefinition. Additionally, in cases where a gene did not have a KO number or corresponding definition, the raw sequence identifier or a KEGG sequence description line was used in this hybrid identifier. Taxonomic identifications and gene annotation best hits were then joined (Dataset S4).

Cell Abundance Normalization and Differential Expression Analysis. To facilitate analyses of the comparative metatranscriptome, we designed an approach called "Microbial Assemblage Normalized Transcript Analysis" (MANTA, v. 0.8.6), which performs RNA-seq normalization and DE on taxonomic subsets of a microbial community. To help correct for potential changes in taxonomic composition between conditions, the "TMM" normalization procedure available in the edgeR Bioconductor package was used for data normalization. Inherent to the TMM normalization procedure is the assumption that a large proportion of genes are not DE under two conditions (22). We extended this procedure to assume that a majority of genes within a mixed community are also not DE under two conditions. This TMM estimate is assumed here to closely track the asymptotic ratio of read counts for non-DE genes. This normalization point was subtracted from the log fold change ratio (*R*) values to calculate the "TMM-adjusted *R*" values for the genes of each taxonomic subset. After normalization, significance of DE genes was assessed using edgeR's "exactTest" function. We restricted our definition of DE genes to only those that had an adjusted *P* value <0.05 (determined by the Benjamini and Hochberg multiple-testing correction implemented in the "p.adjust" method of R) (54). Methods for KEGG modules, pathways, and gene set enrichment analyses are provided in *SI Appendix*, *SI Methods*. Some products of interest were not currently available in KEGG and therefore were manually created (Dataset S5).

Data Deposition. All environmental sequences and ESTs from *P. granii* obtained by 454 sequencing are publicly available on the CAMERA portal (http://camera.calit2.net/). Environmental sequences obtained by 454 and SOLiD sequencing are deposited in the National Center for Biotechnology Information's Sequence Read Archive under study accession no. SRP006906. BWA alignments, BLAST results, custom lookup tables, and analyses code are available at http://manta.ocean.washington.edu/. The MANTA package is also available at Bioconductor (http://bioconductor.org/packages/devel/bioc/html/manta.html). All rhodopsin sequences are deposited in the National Center for Biotechnology Information GenBank database (see *SI Appendix*, Fig. S10 legend for accession numbers).

Phylogenetic Trees. Reference rhodopsin sequences were recruited for the alignment using the HMMER hmmsearch function (*e*-value ≤10^{-5}) (55). A hidden Markov model profile was created using an alignment of experimentally verified rhodopsin sequences (*SI Appendix*, Fig. S10) and was searched against a reference database encompassing a collection of National Center for

Biotechnology Information refseq libraries, all available marine eukaryotic genomes and ESTs, and the ESTother sequence databases (see Dataset S1 for reference database composition). The resulting reference alignment was manually filtered to remove duplicate sequences and included at least one representative of each organism defined by National Center for Biotechnology Information taxa ID for each leaf of the tree. The bootstrapped tree was made using RAxML with the PROTGAMMAWAG model and 100 iterations (56). Metatranscriptomic reads were recruited using a second HMMER hmmsearch, using the reference alignment as the profile HMM. Reads were then placed on the reference tree, using pplacer in posterior probability mode (57), and were visualized and edited using Archeopteryx (58).

ACKNOWLEDGMENTS. We thank the scientists of Fisheries and Oceans Canada at the Institute of Ocean Sciences, Sidney, BC, Canada, and the officers and crew aboard the Canadian Coast Guard ship *John P. Tully* for assistance during the cruise. We are grateful to J. Barwell-Clarke, K. Johnson, and N. Sutherland (Institute of Ocean Sciences) for sharing nutrients and iron data; F. Ribalet for analyzing flow cytometry data; C. Moreno for assistance with gene annotations; and R. Morris and S. Amin for helpful discussions on rhodopsins. J. Hackett and J. Archibald generously provided EST databases for *Geminigera cryophila* (Hackett) and *Guillardia theta* (Archibald) used in our analysis. This study was supported by a Gordon and Betty Moore Foundation Marine Microbiology Investigator Award, National Science Foundation grants, and a National Institute of Environmental Health Sciences grant (to E.V.A.) as well as National Science Foundation Grant MCB-0732448 (to A.E.A.).

1. Martin JH, Fitzwater S (1988) Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic. *Nature* 331:341–343.
2. Boyd PW, et al. (2007) Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions. *Science* 315:612–617.
3. Morel FMM, Hudson RJM, Price NM (1991) Iron nutrition of phytoplankton and its possible importance in the ecology of ocean regions with high nutrient and low biomass. *Oceanography (Wash DC)* 4(2):56–61.
4. Cullen JJ (1995) Status of the iron hypothesis after the open-ocean enrichment experiment. *Limnol Oceanogr* 40:1336–1343.
5. Boyd PW, et al. (2005) The evolution and termination of an iron-induced mesoscale bloom in the northeast subarctic Pacific. *Limnol Oceanogr* 50:1872–1886.
6. Allen AE, et al. (2008) Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proc Natl Acad Sci USA* 105:10438–10443.
7. Buesseler KO, Andrews JE, Pike SM, Charette MA (2004) The effects of iron fertilization on carbon sequestration in the Southern Ocean. *Science* 304:414–417.
8. Denman KL, Voelker C, Pena MA, Rivkin RB (2006) Modelling the ecosystem response to iron fertilization in the subarctic NE Pacific: The influence of grazing, and Si and N cycling on CO2 drawdown. *Deep Sea Res Part II Top Stud Oceanogr* 53:2327–2352.
9. Pollard RT, et al. (2009) Southern Ocean deep-water carbon export enhanced by natural iron fertilization. *Nature* 457:577–580.
10. Greene RM, Geider RJ, Falkowski PG (1991) Effect of iron limitation on photosynthesis in a marine diatom. *Limnol Oceanogr* 36:1772–1782.
11. Marchetti A, Harrison PJ (2007) Coupled changes in the cell morphology and the elemental (C, N and Si) composition of the pennate diatom *Pseudo-nitzschia* due to iron deficiency. *Limnol Oceanogr* 52:2270–2284.
12. Milligan AJ, Harrison PJ (2000) Effects of non-steady-state iron limitation on nitrogen assimilatory enzymes in the marine diatom Thalassiosira weissflogii (Bacillariophyceae). *J Phycol* 36:78–86.
13. Timmermans KR, Stolte W, Debaar HJW (1994) Iron-mediated effects on nitrate reductase in marine phytoplankton. *Mar Biol* 121:389–396.
14. La Roche J, Geider RJ, Graziano LM, Murray H, Lewis K (1993) Induction of specific proteins in eukaryotic algae grown under iron-deficient, phosphorus-deficient, or nitrogen-deficient conditions. *J Phycol* 29:767–777.
15. McKay RML, Geider RJ, LaRoche J (1997) Physiological and biochemical response of the photosynthetic apparatus of two marine diatoms to Fe stress. *Plant Physiol* 114: 615–622.
16. Pankowski A, McMinn A (2009) Iron availability regulates growth, photosynthesis, and production of ferredoxin and flavodoxin in Antarctic sea ice diatoms. *Aquat Biol* 4:273–288.
17. McKay RML, La Roche J, Yakunin AF, Durnford DG, Geider RJ (1999) Accumulation of ferredoxin and flavodoxin in a marine diatom in response to iron. *J Phycol* 35: 510–519.
18. Lommer M, et al. (2010) Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. *BMC Genomics*, 10.1186/1471-2164-11-718.
19. Erdner DL, Price NM, Doucette GJ, Peleato ML, Anderson DM (1999) Characterization of ferredoxin and flavodoxin as markers of iron limitation in marine phytoplankton. *Mar Ecol Prog Ser* 184:43–53.
20. Peers G, Price NM (2006) Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* 441:341–344.
21. Harrison PJ (2002) Station Papa time series: Insights into ecosystem dynamics. *J Oceanogr* 58:259–264.
22. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, 10.1186/gb-2010-11-3-r25.
23. Kolber ZS, et al. (1994) Iron limitation of phytoplankton photosynthesis in the equatorial Pacific Ocean. *Nature* 371:145–149.
24. Marchetti A, et al. (2009) Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature* 457:467–470.
25. Ribalet F, et al. (2010) Unveiling a phytoplankton hotspot at a narrow boundary between coastal and offshore waters. *Proc Natl Acad Sci USA* 107:16571–16576.
26. Raven JA (1988) The iron and molybdenum use efficiencies of plant growth with different energy, carbon and nitrogen sources. *New Phytol* 1988:279–287.
27. Raven JA, Evans MCW, Korb RE (1999) The role of trace metals in photosynthetic electron transport in O_2-evolving organisms. *Photosynth Res* 60:111–149.
28. Raven JA (1990) Predictions of Mn and Fe use efficiencies of phototrophic growth as a function of light availability for growth and of C assimilation pathway. *New Phytol* 116:1–18.
29. Alscher RG, Erturk N, Heath LS (2002) Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *J Exp Bot* 53:1331–1341.
30. Marchetti A, Juneau P, Whitney FA, Wong CS, Harrison PJ (2006) Phytoplankton processes during a mesoscale iron enrichment in the NE subarctic Pacific: Part II - nutrient utilization. *Deep Sea Res Part II Top Stud Oceanogr* 53:2114–2130.
31. Boyd PW, et al. (1996) In vitro iron enrichment experiments in the NE subarctic Pacific. *Mar Ecol Prog Ser* 136:179–193.
32. Boyd PW, Berges JA, Harrison PJ (1998) In vitro enrichment experiments at iron-rich and poor sites in the NE subarctic Pacific. *J Exp Mar Biol Ecol* 227:133–151.
33. Bruhn A, LaRoche J, Richardson K (2010) *Emiliana huxleyi* (prymnesiophyceae): Nitrogen-metabolism genes and their expression in response to external nitrogen sources. *J Phycol* 46:266–277.
34. Peers G, Niyogi KK (2008) Pond scum genomics: The genomes of *Chlamydomonas* and *Ostreococcus*. *Plant Cell* 20:502–507.
35. Kroth PG, et al. (2008) A model for carbohydrate metabolism in the diatom Phaeodactylum tricornutum deduced from comparative whole genome analysis. *PLoS ONE* 3:e1426.
36. Doubnerová V, Ryšlavá H (2011) What can enzymes of C_4 photosynthesis do for C_3 plants under stress? *Plant Sci* 180:575–583.
37. Allen AE, et al. (2011) Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473:203–207.
38. Bender SJ, Parker MS, Armbrust EV (2011) Coupled effects of light and nitrogen source on the urea cycle and nitrogen metabolism over a diel cycle in the marine diatom Thalassiosira pseudonana. *Protist,* 10.1016/j.protis.2011.07.008.
39. Llácer JL, Fita I, Rubio V (2008) Arginine and nitrogen storage. *Curr Opin Struct Biol* 18:673–681.
40. Kröger N, Deutzmann R, Sumper M (2001) Silica-precipitating peptides from diatoms. The chemical structure of silaffin-A from Cylindrotheca fusiformis. *J Biol Chem* 276: 26066–26070.
41. Béjà O, et al. (2000) Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* 289:1902–1906.
42. Fuhrman JA, Schwalbach MS, Stingl U (2008) Proteorhodopsins: An array of physiological roles? *Nat Rev Microbiol* 6:488–494.
43. Slamovits CH, Okamoto N, Burri L, James ER, Keeling PJ (2011) A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nat Commun*, 10.1038/ncomms1188.
44. Lin S, Zhang H, Zhuang Y, Tran B, Gill J (2010) Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. *Proc Natl Acad Sci USA* 107:20033–20038.
45. Sineshchekov OA, Jung K-H, Spudich JL (2002) Two rhodopsins mediate phototaxis to low- and high-intensity light in Chlamydomonas reinhardtii. *Proc Natl Acad Sci USA* 99:8689–8694.
46. Morris RM, et al. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J* 4:673–685.
47. Andersson JO (2005) Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* 62: 1182–1197.
48. Keeling PJ (2009) Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr Opin Genet Dev* 19:613–619.
49. Raven JA (2009) Functional evolution of photochemical energy transformations in oxygen-producing organisms. *Funct Plant Biol* 36:505–515.
50. Zehr JP, Kudela RM (2009) Ocean science. Photosynthesis in the open ocean. *Science* 326:945–946.
51. Vartanian M, Desclés J, Quinet M, Douady S, Lopez PJ (2009) Plasticity and robustness of pattern formation in the model diatom Phaeodactylum tricornutum. *New Phytol* 182:429–442.
52. Poinar HN, et al. (2006) Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science* 311:392–394.
53. Miller W, et al. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:387–390.
54. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300.
55. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763.
56. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
57. Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 10.1186/1471-2105-11-538.
58. Han MV, Zmasek CM (2009) phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics*, 10.1186/1471-2105-10-356.