Metatranscriptome analyses indicate resource partitioning between diatoms in the field

Harriet Alexander[a,b], Bethany D. Jenkins[c,d], Tatiana A. Rynearson[c], and Sonya T. Dyhrman[b,e,1]

[a]Massachusetts Institute of Technology–Woods Hole Oceanographic Institution Joint Program in Oceanography/Applied Ocean Science and Engineering, Cambridge, MA 02139; [b]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543; [c]Graduate School of Oceanography, University of Rhode Island, Narragansett, RI 02882; [d]Department of Cell and Molecular Biology, University of Rhode Island, Kingston, RI 02881; and [e]Department of Earth and Environmental Sciences, Lamont–Doherty Earth Observatory, Columbia University, Palisades, NY 10964

Edited by Sallie W. Chisholm, Massachusetts Institute of Technology, Cambridge, MA, and approved March 20, 2015 (received for review November 17, 2014)

Diverse communities of marine phytoplankton carry out half of global primary production. The vast diversity of the phytoplankton has long perplexed ecologists because these organisms coexist in an isotropic environment while competing for the same basic resources (e.g., inorganic nutrients). Differential niche partitioning of resources is one hypothesis to explain this "paradox of the plankton," but it is difficult to quantify and track variation in phytoplankton metabolism in situ. Here, we use quantitative metatranscriptome analyses to examine pathways of nitrogen (N) and phosphorus (P) metabolism in diatoms that cooccur regularly in an estuary on the east coast of the United States (Narragansett Bay). Expression of known N and P metabolic pathways varied between diatoms, indicating apparent differences in resource utilization capacity that may prevent direct competition. Nutrient amendment incubations skewed N/P ratios, elucidating nutrient-responsive patterns of expression and facilitating a quantitative comparison between diatoms. The resource-responsive (RR) gene sets deviated in composition from the metabolic profile of the organism, being enriched in genes associated with N and P metabolism. Expression of the RR gene set varied over time and differed significantly between diatoms, resulting in opposite transcriptional responses to the same environment. Apparent differences in metabolic capacity and the expression of that capacity in the environment suggest that diatom-specific resource partitioning was occurring in Narragansett Bay. This high-resolution approach highlights the molecular underpinnings of diatom resource utilization and how cooccurring diatoms adjust their cellular physiology to partition their niche space.

phytoplankton | diatom | nutrient physiology | niche partitioning | metatranscriptomics

The stability and primary productivity of ecosystems have long been linked to the diversity of primary producers (1, 2). This linkage is well documented in terrestrial systems (3–7) and is increasingly being established for marine systems (8–11). Marine phytoplankton generate roughly half of global primary production (12–14) and play a critical role in oceanic ecosystem structure and function. Within the phytoplankton, the diatoms generate an estimated 40% of primary production (15). Thus, diatoms alone exert a profound influence over marine primary production and global carbon (C) cycling, particularly in coastal margins and estuaries.

Phytoplankton are extremely diverse, with estimates of over 200,000 extant species (16, 17). This dramatic level of taxonomic diversity in the plankton is difficult to resolve with the apparently limited number of niches in the pelagic habitat because these organisms compete for the same two basic resources: light and nutrients. As was highlighted by Hutchinson (18), the phytoplankton violate Gause's law of competitive exclusion, which posits that two organisms competing for the same resources cannot coexist. Much thought has gone toward identifying the cause of the "paradox of the plankton," including explanations such as "contemporaneous disequilibrium" of patchy phytoplankton distributions (19), life history differences (20), species oscillations (21), environmental fluctuation (22), intraspecific variation (23), and differential niche partitioning (24). Of these potential factors, one of the most difficult to observe directly in the plankton is niche partitioning. Different species may have unique strategies that allow them to specialize on certain resources or nutrient forms, and species may have different responses to resource shifts that allow them to avoid competition. Such specialization in ecoevolutionary strategy may underlie the "winner-loser" dynamics observed in productive estuaries and coastal systems, yet resolving patterns of species-specific resource metabolism in the field remains a central challenge.

It is accepted that the macronutrients nitrogen (N) and phosphorus (P) are central to the structuring of phytoplankton communities across large spatial and temporal scales (25–27), and that phytoplankton compete for nutrients in the natural environment (28, 29). Studies focused on nutrient geochemistry and phytoplankton quotas or uptake have emphasized the importance of nutrients to community dynamics, but these studies do not generally examine resource partitioning between individual species (30, 31). Transcriptional studies provide species-specific resolution, but few studies have examined the global expression of nutrient metabolism pathways in the field (32) or in organisms lacking a fully sequenced genome (33, 34), and as a result, the mechanistic underpinnings of phytoplankton resource metabolism in situ are not well understood. In situ global gene expression analyses (metatranscriptome profiling) are a means for elucidating a species' metabolic capacity and examining

Significance

Nutrient availability plays a central role in driving the activities and large-scale distributions of phytoplankton, yet there are still fundamental gaps in understanding how phytoplankton metabolize nutrients, like nitrogen (N) and phosphorus (P), and how this metabolic potential is modulated in field populations. Here, we show that cooccurring diatoms in a dynamic coastal marine system have apparent differences in their metabolic capacity to use N and P. Further, bioinformatic approaches enabled the identification and species-specific comparison of resource-responsive (RR) genes. Variation of these RR gene sets highlights the disparate transcriptional responses these species have to the same environment, which likely reflects the role resource partitioning has in facilitating the vast diversity of the phytoplankton.

Author contributions: H.A., B.D.J., T.A.R., and S.T.D. designed research; H.A., B.D.J., T.A.R., and S.T.D. performed research; H.A. contributed new reagents/analytic tools; H.A. analyzed data; and H.A., B.D.J., T.A.R., and S.T.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The field sequence data reported in this paper have been deposited in the National Center for Biotechnology Information Sequence Read Archive, www.ncbi.nlm.nih.gov/sra (accession no. SRP055134).

[1]To whom correspondence should be addressed. Email: sdyhrman@ldeo.columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1421993112/-/DCSupplemental.

patterns in resource utilization potential through time by tracking the expression of species' resource-responsive (RR) genes. When simultaneously applied to multiple species in a sample, this approach can resolve differences in the expressed gene compliment and how it is modulated, which may reflect resource partitioning of phytoplankton niche space (35). For example, this approach has uncovered species-specific expression of genes for the transport of organic compounds in the bacterioplankton (36–38), highlighting potential differences in resource partitioning. Although increasingly critical for identifying resource utilization in the bacterioplankton, metatranscriptome profiling has only recently been used to examine resource utilization in coastal eukaryotic phytoplankton populations (39), largely due to challenges in quantifying a transcriptional response in a mixed population and, until recently, the lack of reference genomes and transcriptomes for determining the origin of the transcriptional response. Cooccurring phytoplankton may possess different metabolic capabilities and responses to resource availability, which may then enable resource partitioning and the segregation of the fundamental niche or the realized niche. Knowledge of if and how these organisms modulate their niche space would allow predictive models to better resolve species distribution and ecosystem structure and function in the future ocean (26).

Herein, we examined pathways of resource metabolism between two cooccurring diatoms from the genera *Thalassiosira* and *Skeletonema*, sampled from a time-series site in Narragansett Bay. Narragansett Bay is a highly productive and dynamic estuarine environment on the east coast of the United States with an estimated bay-wide average net production of 269 $gC{\cdot}m^2{\cdot}y^{-1}$ (40). Quantitative metatranscriptomic techniques were developed and used to (*i*) assign taxonomic designation, (*ii*) assess and track changes in known metabolic capacity by means of the quantitative molecular fingerprint (QMF), (*iii*) statistically identify the RR gene set, and (*iv*) proportionalize the expression of RR genes to track species-specific responses through time using standardized transcriptional differentiation (STD) scores. This multifaceted computational approach enabled the unprecedented resolution of the unique strategies these two diatoms use for resource acquisition.

Results and Discussion

Samples and Sequencing. Narragansett Bay has seasonal blooms of diatoms that have been monitored through weekly cell counts for over 50 y at a long-term time-series station (41, 42). Five eukaryotic surface metatranscriptome samples were taken from surface seawater collected during May and June of 2012 at the time-series site, yielding over 358 million 100-bp, paired-end cDNA reads from the field [sample 1 (S1)–S5] (*SI Appendix*, Table 1). In conjunction with these field-based surveys, a nutrient amendment incubation experiment was performed with natural communities on May 30, 2012 (S3) to drive the community toward opposite extremes in the N/P ratio (Redfield ratio) (*SI Appendix*, Table 2). Eukaryotic metatranscriptomes from the five incubation treatments produced over 264 million 100-bp, paired-end cDNA reads (*SI Appendix*, Table 1).

To assign taxonomic designation, sequences from the time series were conservatively mapped (such that if a read mapped to more than one gene, it was discarded) to a sequence library containing all assembled sequences and annotations generated through the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) (43), which were made public as of March 17, 2014. The custom sequence library contained 401 transcriptomes across 209 species or cultured isolates. Between 62% and 71% of reads from the in situ samples mapped to the MMETSP database with diatoms dominating the libraries, representing 30–46% of the total mapped reads (Fig. 1). The peak in diatom representation coincided with a bloom of *Skeletonema* spp. detected in time-series cell counts (*SI Appendix*, Fig. 1) and

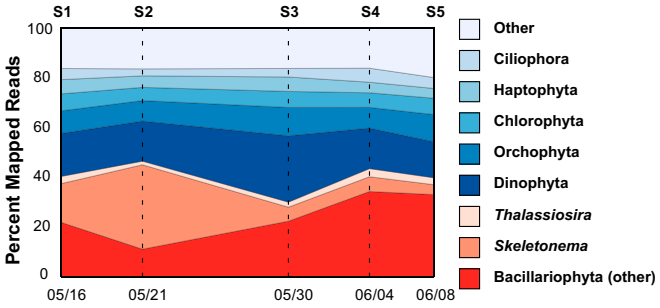


Fig. 1. Taxonomic classification of RNA-sequencing paired-end reads across the five field samples. Classification was determined by mapping to a database composed of all publicly available transcriptomes through the MMETSP as of March 17, 2014.

a period of historical overlap between the *Skeletonema* and *Thalassiosira* genera. *Skeletonema* and *Thalassiosira* were well represented during the time period studied in both mapped RNA (Fig. 1) and cell counts (*SI Appendix*, Fig. 1). *Thalassiosira rotula* was present, but at low abundance, during the time series, whereas *Skeletonema* spp. was abundant, with sampling spanning a bloom of *Skeletonema* (>10 million cells per liter) with peak cell densities in S2 (May 21, 2012) (*SI Appendix*, Fig. 1). As such, subsequent analyses were focused on these two groups by remapping the data to representative transcriptomes: *T. rotula* and *Skeletonema costatum* (*SI Appendix*, Table 1). *S. costatum* was chosen because it was the transcriptome from the genus *Skeletonema* that recruited the most hits in the MMETSP database. Because *Skeletonema* is known to include morphologically cryptic species that can only be identified by SEM (44–46), it is referred to here as *Skeletonema* spp. for clarity. Up to 17.5% and 54.9% of reads from a single sample mapped to *T. rotula* and *S. costatum*, respectively. As a point of comparison, reads were also mapped to the genome of a second thalassiosirid, *Thalassiosira pseudonana,* a diatom that is not known to be abundant in Narragansett Bay (*SI Appendix*, Table 1). Although displaying high identity with the 18S rDNA to *T. rotula* and *S. costatum* (96% and 93% identity, respectively), less than 1% of the metatranscriptome reads mapped *T. pseudonana* (*SI Appendix*, Table 1), highlighting the specificity of the approach.

Temporal Plasticity in Expressed Metabolic Capacity. Metatranscriptome short reads were mapped to transcriptomes that had been annotated with Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology (KO) (Dataset 1), allowing the expression of KO gene families within a KEGG module (higher level groupings of KO gene families into pathway or functional classifications) to be examined over time. Normalizing the expression of KEGG modules to the total KEGG annotated reads for each organism across time yielded the QMF, which highlighted differences between the two species and differences across time for each species (Fig. 2). A comparison of the total number of annotated genes falling into each of the KEGG modules revealed a close to one-to-one linear relationship (slope of 1.0948, $R^2 = 0.9123$) (*SI Appendix*, Fig. 2), indicating that the observed differences are not an artifact of gene distribution between organisms. The QMFs of the two organisms were distinct, and there were significant shifts in the QMF of each species over time reflecting considerable plasticity in the expressed metabolic capacity (*SI Appendix*, Fig. 3). Central carbohydrate metabolism, C fixation, and other carbohydrate metabolism were some of the most highly expressed KEGG modules in the field for both *Skeletonema* spp. and *T. rotula*, although higher for *Skeletonema* spp., where expression of these pathways peaked during S4, representing over 84% of mapped KEGG reads (Fig. 2). The

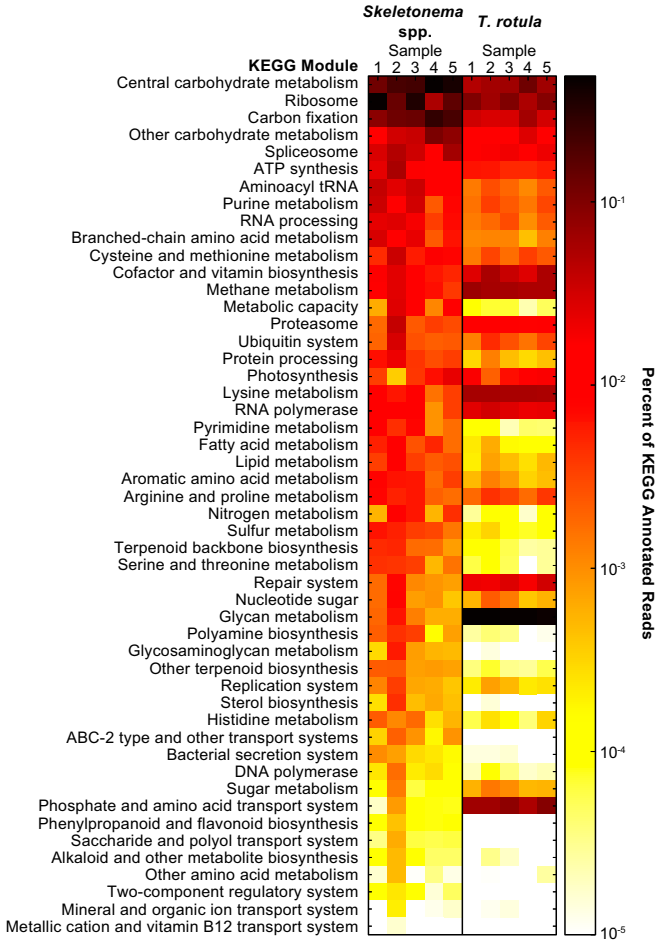


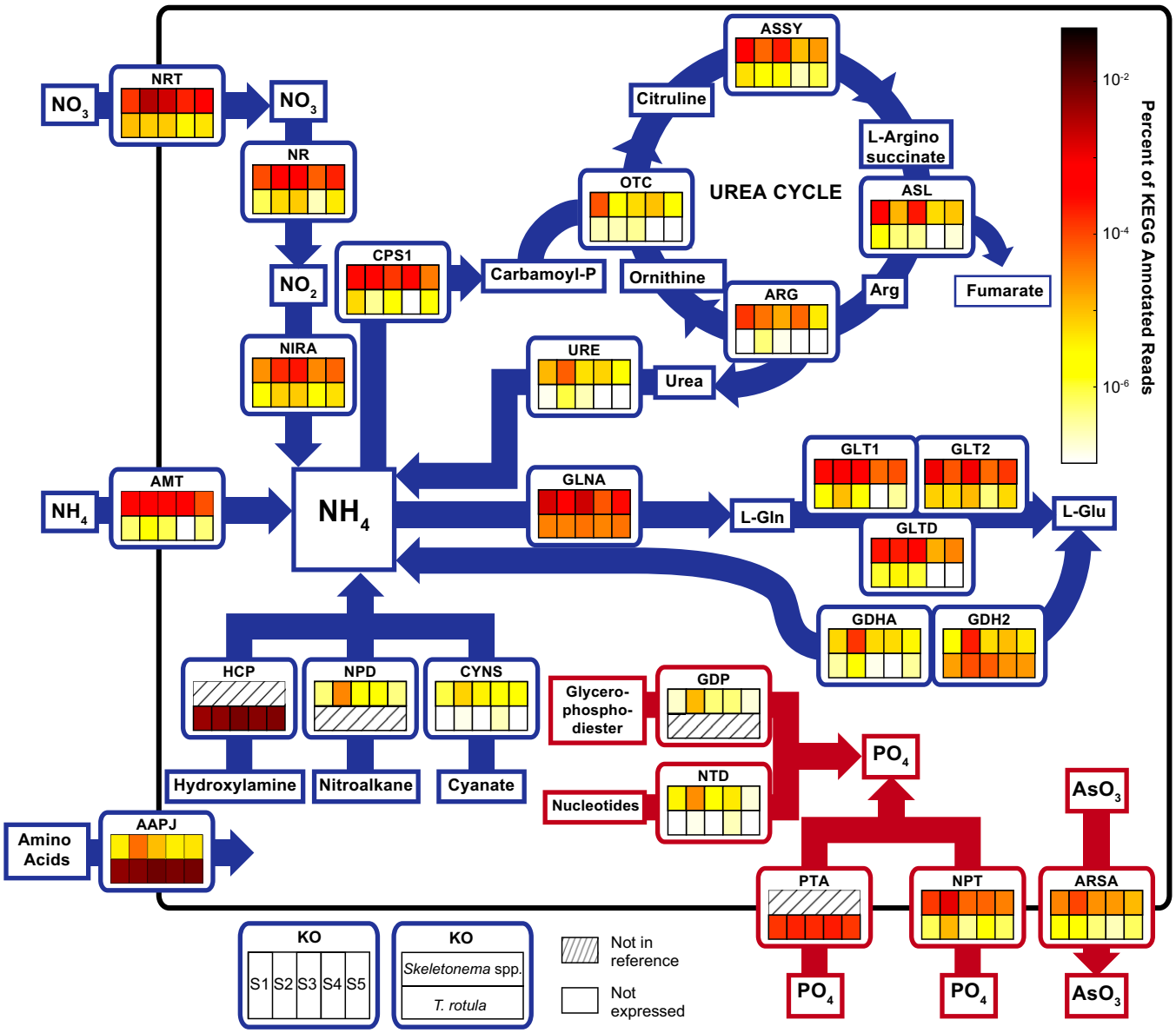
Fig. 2. Quantitative metabolic fingerprint depicting the relative expression of KEGG modules for *Skeletonema* spp. and *T. rotula* in Narragansett Bay across the five sampling time points (S1–S5). Color indicates the proportion of total reads mapping to each KEGG module relative to all KEGG annotated reads.

largest global shift in KEGG module expression was seen in *Skeletonema* spp. on S2 (*SI Appendix*, Fig. 3), when its density peaked at 11,520,000 cells per liter. The S2 time point for *Skeletonema* spp. had increased QMF signals in ATP synthesis, proteasome, and ubiquitin systems and decreased QMF signals in photosynthesis and C metabolism relative to other time points. For example, 0.03% of annotated transcripts mapped in the photosynthesis KEGG module in S2, an order of magnitude lower than the other samples which ranged between 0.3 to 2.2% (Fig. 2). The temporal plasticity of transcript allocation to different aspects of metabolism for both species was striking and likely reflects the dynamic environment that they inhabit: an estuary where the geochemistry is highly variable (47).

Temporal plasticity in the KEGG module expression patterns, including a shift away from the expression of C fixation and photosynthesis, suggests that the elevated *Skeletonema* spp. cell numbers observed in S2 may have occurred after this diatom reached peak bloom biomass. A significant proportion of the KEGG modules expressed were classified as ribosomes (5–45% for *Skeletonema* spp. and 5–9% for *T. rotula*). Gifford et al. (35) suggested that ribosomal protein expression correlates with growth rate, and applying this principle to these eukaryotic data suggests growth rates for both *Skeletonema* spp. and *T. rotula* fluctuated, with peaks in growth rate occurring during S1 and S3 for *Skeletonema* spp. This pattern for *Skeletonema* spp. did not track with the relative abundance of the organism, which peaked in the S2 sample, again suggesting that this sample was taken during the bloom decline. These growth dynamics cannot be fully resolved without a more detailed sample set.

Skeletonema spp., the dominant diatom during the study period (Fig. 1), had a higher proportion of transcripts related to growth relative to *T. rotula*, such as those transcripts encoding aspects of C metabolism, N metabolism, sulfur metabolism, and lipid metabolism (Fig. 2). Conversely, several KEGG modules were more highly expressed in *T. rotula* compared with *Skeletonema* spp., particularly those KEGG modules for glycan metabolism, phosphate, and amino acid transport systems, as well as repair system modules (Fig. 2). The majority of highly expressed KO modules (e.g., N metabolism) were based on moderate to high expression across several KO gene families, but the differences in expression at the module level were due to differences in the expression of a single KO gene family within the KEGG module in some cases. For example, the driver of the difference in the expression of glycan metabolism, which represented upward of 41% of all KEGG annotated reads for *T. rotula* compared with less than 0.6% for *Skeletonema* spp., was primarily associated with the high expression of a putative UDP–*N*-acetylglucosamine–dolichyl-phosphate *N*-acetylglucosaminephosphotransferase (K01001). This gene was identified as a silaffin-like response gene associated with silica polymerization (48). Differences in silica metabolism may partially drive how the fundamental niche is segregated between these two diatoms. Taken together, the contrast in QMF between the two diatoms underscores the fundamental differences in expressed metabolic capacity that are present in these two cooccurring diatoms and highlights traits of a successful competitor (e.g., high expression of C metabolism).

Species-Specific Resource Utilization Underpins Physiological Ecology. KO gene families related to N and P metabolism were examined in the field samples to identify species-specific patterns in resource utilization. *Skeletonema* spp. and *T. rotula* both possess and express core pathways of N and P metabolism (e.g., the ornithine-urea cycle) (Fig. 3). Expression of these individual KO gene families was temporally variable, as was observed with the expression of KEGG modules, but related enzymes in a pathway exhibited a coordinated response (Fig. 3). For example, the nitrate transporter (K02575), nitrate reductase (K10534), and nitrite reductase (K00366) in *Skeletonema* spp. all had peak expression in S2 (Fig. 3). *Skeletonema* spp. and *T. rotula* share pathway homologs, including the same suite of N transporters (ammonium, nitrate, and amino acid), but these genes often had disparate patterns of expression between the two species (Fig. 3). *Skeletonema* spp., the more abundant diatom, had high expression of KO gene families associated with the acquisition of nitrate and ammonia that were particularly amplified during the S2 bloom event. *T. rotula* had low expression of both of those transporters but high expression of a general amino acid transporter (Fig. 3). Amino acid transport (49) and nitrate transport (50) have previously been found to correlate inversely with intracellular nitrate concentration in the cell or in the presence of ammonia in the media. However, here, two closely related diatoms existing in the same parcel of water and the same nutrient environment are expressing genes to access different pools of dissolved N. Similar to nitrate transport, there was high expression of nitrate/nitrite reductase KO gene families in *Skeletonema* spp., whereas *T. rotula* appears to possess a different N reduction metabolism. This difference is observed in a KO gene family that is absent from the reference transcriptome of *Skeletonema* spp.: hydroxylamine reductase (Fig. 3). This gene has been found in the genomes of both *T. pseudonana* and *Phaeodactylum tricornutum*, and it is thought to have been acquired via lateral transfer from bacteria (51). The enzyme may potentially aid redox balancing and electron cycling from nitrate reduction (52).



AAPJ	General L-amino acid transport system (K09969)	**GLT2**	Glutamate synthase (ferredoxin) (K00284)
AMT	Ammonium transporter (K03320)	**GLTD**	Glutamate synthase; small chain (K00266)
ARG	Arginase (K01476)	**HCP**	Hydroxylamine reductase (K05601)
ARSA	Arsenite-transporting ATPase (K01551)	**NIRA**	Nitrite reductase (ferredoxin) (K00366)
ASL	Argininosuccinate lysase (K01755)	**NPD**	Nitronate monooxygenase (K00459)
ASSY	Argininosuccinate synthase (K01940)	**NPT**	Sodium-phosphate cotransporter (K14683)
CPS1	Carbamoyl-phosphate synthase (K01948)	**NR**	Nitrate reductase (NADP/NADPH) (K10534)
CYNS	Cyanate lyase (K01725)	**NRT**	Nitrate transporter (K02575)
GDH2	Glutamate dehydrogenase (K15371)	**NTD**	5'-bisphosphate nucleotidase (K01082)
GDHA	Glutamate dehydrogenase (NADP+) (K00262)	**OTC**	Ornithine carbamoyltranferase (K00611)
GDP	Glycerophosphoryl diester phosphodiesterase (K01126)	**PTA**	Putative phosphate transporter (K02036)
GLNA	Glutamine synthetase (K01915)	**URE**	Urease (K01427)
GLT1	Glutamate synthase (NADP/NADH) (K00264)		

Fig. 3. Schematic cell model depicting the relative expression of KO gene families associated with N and P metabolic pathways for *Skeletonema* spp. and *T. rotula* in Narragansett Bay across the five sampling time points (S1–S5). Color indicates the proportion of total reads mapping to each KEGG module relative to all KEGG annotated reads.

Although the absence of this gene in *Skeletonema* spp. has not been definitively shown, the marked high expression of this gene in *T. rotula* suggests that this gene product represents a potential point of segregation in the metabolic capacity of these two species. Together, these data suggest that these species have disparate strategies for acquiring N and these differences

may partially drive the relative success of *Skeletonema* spp. over the sample period.

Although N has been observed to be a primary nutritional driver in Narragansett Bay (47, 53, 54), P may also drive the dynamics of these two diatoms. *Skeletonema* spp. shows elevated expression of a sodium phosphate cotransporter (NPT), again with peak expression during S2 (bloom). *T. rotula* does not express the NPT as highly but, by comparison, has a much higher transcript count for a putative P transporter that is not detected in *Skeletonema* spp. (Fig. 3). These transporters may have different kinetic properties that allow the two species to diverge in their PO_4 uptake strategies. Genes associated with the scavenging of P from organic molecules, such as glycerophosphoryl diester phosphodiesterase (GDP), also suggest differences in expressed metabolic capacity between the two species. GDP may be associated with exogenous metabolism of dissolved organic P (DOP) or internally in the cleaving of P from lipids (55, 56). The expression of GDP by *Skeletonema* spp., with a peak around S2, and the apparent absence of this transcript in *T. rotula* suggest *Skeletonema* spp. may be accessing a pool of DOP that is not being used by *T. rotula*. In *T. pseudonana,* related transcripts are tightly linked to concomitant changes in the proteome and biochemical activities (56). If these transcriptional patterns are linked to similar changes in activities, then these insights suggest that there is a fundamental difference in the metabolic capacity being expressed in the same environment by the two diatoms. *Skeletonema spp.* is both actively taking up PO_4 and hydrolyzing organic sources, whereas *T. rotula* is not hydrolyzing DOP and is taking up inorganic PO_4 by a different mechanism. In summary, these data suggest that these two diatoms have a unique metabolic capacity for the utilization of specific forms of N and P. Such disparate resource utilization potential could be a niche-defining feature that underpins diatom diversity as well as the "winner-loser" dynamic observed here with the differences in cell abundance between the species.

Identification and Modulation of RR Genes in Situ Highlight Species-Specific Differences. To identify and quantitatively track RR genes in situ, incubation experiments were used to examine species-specific transcriptional responses to shifts in N/P ratios. Comparing the expression patterns between like nutrient treatments (+N vs. −N and +P vs. −P) for each of the organisms enabled the statistical identification of a suite of RR genes (57) and stable reference genes (58). RR gene counts were normalized to the stable reference genes (*SI Appendix*, Fig. 4), resulting in stable gene normalized counts (SGNCs). Calculation of an SGNC is similar in concept to the reference gene normalization done in quantitative RT-PCR (qRT-PCR) (59) or metatranscriptome studies of prokaryotes (60), with the added value of not having to rely on reference genes from model diatoms.

Of the transcripts expressed at greater than two tags per million (TPM) for at least one treatment, 24.5% and 17.9% were identified as RR by being significantly up- or down-regulated in the N or P treatments for *Skeletonema* spp. and *T. rotula*, respectively (*SI Appendix*, Table 3 and Dataset 2). As is common with phytoplankton studies (32), the majority of the RR genes do not have a KEGG annotation (Fig. 4*A* and Dataset 2). The portion of the RR gene set annotated with KEGG ontology for *Skeletonema* spp. and *T. rotula* revealed that, relative to the full KEGG profile, genes comprising genetic information processing associated with replication (encompassing ribosomes, nucleotide replication, and processing) were underrepresented for both organisms in the RR set (*SI Appendix*, Fig. 5). By contrast, the RR sets were enriched for energy metabolism, carbohydrate metabolism, and lipid metabolism, which encompass pathways known to be associated with the metabolism of N and P (Fig. 4*A* and *SI Appendix*, Fig. 5). Specific genes in this set included, but were not limited to, those genes associated with N

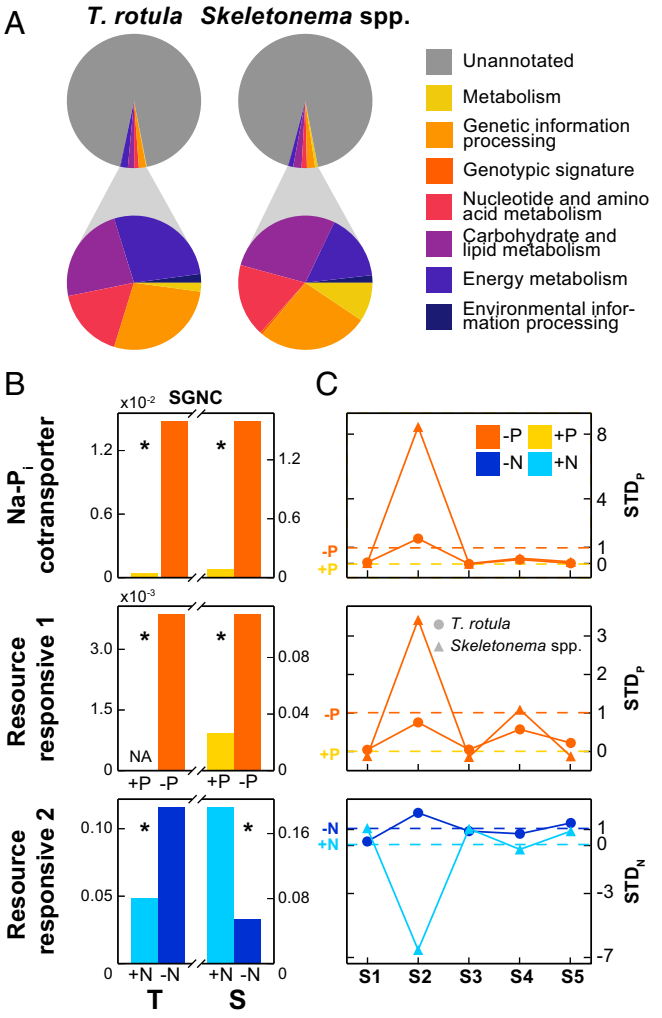


Fig. 4. Functional composition of the RR gene sets for *T. rotula* and *Skeletonema* spp. (*A*), relative expression in the incubation samples (*B*), and STD scores (*C*) for a known P-responsive gene, sodium-phosphate cotransporter, and two RR gene families. (*A*) RR gene sets were identified through cross-comparison of like-nutrient incubations (i.e., +N vs. −N and +P vs. −P), using ASC (fold change = 2, post-p > 0.95) (57). The relative functional categorization of the RR gene set for *T. rotula* and *Skeletonema* spp., based on KEGG ontology as assigned by the KEGG Automatic Annotation Server, is depicted at the module level relative to the portion unannotated with the KEGG. (*B*) Expression pattern in SGNCs of the genes from the associated gene cluster from *T. rotula* (T) and *Skeletonema* spp. (S) plotted in related incubations (i.e., RR1 shows expression from +P and −P incubations). The asterisk indicates significance between pair-wise comparisons (fold change = 2, post-p > 0.95) (57). (*C*) STD scores plotted across the five sample points showing STD_P for the P-significant genes and STD_N for the N-significant genes. Dashed horizontal lines at 0 and 1 indicate the +P or +N and −P or −N for corresponding significant genes.

assimilation (e.g., glutamate dehydrogenase, glutamine synthase, nitrate reductase), dissolved organic N utilization (e.g., urease, aminopeptidase, amino acid transport system), P scavenging (e.g., phosphate transporter, NPT), and DOP utilization (e.g., phosphatases) (Dataset 2). A number of these genes have been shown to be N- or P-responsive in transcriptional studies with cultures of the diatom *T. pseudonana* (56, 61), and transporters and enzymes for the processing of organic N or P, as observed here, are well known to be RR in many phytoplankton (56, 62–64). Overall, these genes demonstrated patterns of regulation in situ (Fig. 4*B* and *SI Appendix*, Fig. 6) similar to what has been observed in culture (56,

61). In the incubations, the NPT was significantly up-regulated in the −P treatment for both species (Fig. 4*B*), which is consistent with P regulation of a *T. pseudonana* NPT homolog (Thaps_24435) observed in culture experiments (56). Nitrate reductase, which has been shown to be regulated by N in *T. pseudonana* (Thaps_25299) (61), was up-regulated in −N for *T. rotula*, but not for *Skeletonema* spp. (*SI Appendix*, Fig. 6). In fact, members of this large gene family (*SI Appendix*, Fig. 7) showed disparate regulation in both species (*SI Appendix*, Fig. 6). These data demonstrate that the use of nutrient amendments is robust for normalizing and identifying N- and P-responsive genes in the field that are consistent with known signals, but they also point to the value of in situ analyses, because application of a priori knowledge about regulation from model diatoms could lead to misinterpretations.

Of the RR gene sets for *Skeletonema* spp. and *T. rotula*, only 17.7% and 12.7% of the genes, respectively, could be annotated with KEGG ontology (Fig. 4*A*). Identifying differentially regulated genes in situ through experimental manipulations allowed the expression patterns of genes to be tracked even when their function was unknown. As an example, two RR gene families were identified with homologs in *Skeletonema* spp. and *T. rotula* (Fig. 4*B* and *SI Appendix*, Fig. 7). RR gene family 1 (RR1) was up-regulated in −P compared with +P for both species (Fig. 4*B*). Homologs from RR1 were also identified in other diatom genomes (Fracy_268075, Phatr_19661, Psemu_246578, Psemu_319824, and Thaps_32459) (*SI Appendix*, Fig. 7). Annotations for these genes were limited, although Fracy_268075 was identified as possibly involved in intracellular trafficking, secretion, or vesicular transport, suggesting these proteins may be involved in poly-P metabolism (65). RR2 demonstrated significantly different patterns of regulation in the two species: up-regulated in −N compared with +N for *T. rotula* but down-regulated in −N compared with +N for *Skeletonema* spp. (Fig. 4*B*). A homolog from RR2 was identified in *T. pseudonana* (Thaps_22648) (*SI Appendix*, Fig. 7) and was poorly characterized, with the best BLAST hit to a human dentin sialophosphoprotein. This finding suggests RR2 could be associated with biomineralization.

To enable cross-comparison of the RR genes between species, their expression was put into a greater metabolic context by proportionalizing the expression in the field to the transcriptional range observed in the incubations with extremes in the N/P ratio. This technique is similar in concept to targeted assays using qRT-PCR to compare expression patterns between species in culture (66). Briefly, the SGNC of a gene in the field was bounded by the SGNC from each of the nutrient treatments to yield the STD score for both N (STD_N) and P (STD_P) (Fig. 4*C*). The STD score was used to compare expression directly relative to its maximum and minimum capacity, where values of STD ≥ 1 indicate signals were similar to the deplete condition and values of STD ≤ 0 indicate similarity to the replete condition. The STD_N and STD_P were plotted for genes from the NPT and the two highlighted RR gene families over the time series (Fig. 4*C*). The NPT for both *Skeletonema* spp. and *T. rotula* showed elevated expression during S2. RR1, which was also identified as significantly expressed in −P compared with +P, also showed elevation during S2 (the bloom). The expression of RR1, however, was also elevated on S4 for both diatoms, which was not seen for the NPT. However, the STD_P was >1 for *Skeletonema* spp., indicating a far more P-deficient response in *Skeletonema* spp. compared with *T. rotula*, which never demonstrated P-sensitive expression in the field comparable to the P-sensitive expression observed in the −P incubations (Fig. 4*C*). RR2 showed different patterns of expression across time for both species. Most interesting perhaps was the low STD_N score for *Skeletonema* spp. during S2 (the bloom), indicating that RR2 expression was more similar to the +N treatment, whereas the STD_N for *T. rotula* was greater than 1, suggesting that RR2 expression was more similar to the −N treatment (Fig. 4*C*). These three targeted examples suggest that during the large bloom of *Skeletonema* spp., *Skeletonema* spp. was expressing genes in a pattern more similar to the −P and +N treatments, whereas *T. rotula* was expressing genes similar only to the −N treatment. Notably, these orthogonal patterns were associated with the same environment.

The STD_N and STD_P for all of the RR genes were calculated (Dataset 2) to expand upon the single-gene analyses above. The RR genes were plotted based on the STD_N/STD_P ratio (*SI Appendix*, Fig. 8) to examine how similar the pattern was to the incubation N/P ratio (Fig. 5*A* and *SI Appendix*, Fig. 9). Redfield regimes have historically been used to characterize different aquatic environments based on the ratio of nutrient resources required for growth. For example, a Redfield ratio of N/P = 16, here called "Redfield," would predict neither P nor N limitation. As expected, RR genes identified as N-regulated genes fall primarily into the N/P < Redfield quadrant and P-regulated genes fall primarily into the N/P > Redfield quadrant for both *Skeletonema* spp. and *T. rotula* (Fig. 5*A*). Observing patterns in the distribution of these genes across time, S2 stands out among the time points, where a significant (88%) proportion of the P-regulated genes from *Skeletonema* spp. move far into the N/P > Redfield quadrant (Fig. 5*A*). This N/P > Redfield physiology is consistent with the single-gene analyses (Fig. 4*C*) and suggests P availability may have constrained *Skeletonema* spp. populations during the bloom sample (S2). Conversely, a large proportion (59%) of the N-regulated genes in *T. rotula* move into the N/P < Redfield quadrant (Fig. 5*A*) consistent with the divergent responsiveness of RR2 observed for *T.* rotula compared with *Skeletonema* spp. (Fig. 4*C*). In fact, with the exception of S4 and S5, where *T. rotula* had even distribution between the N/P > Redfield and N/P < Redfield quadrants, the two species always showed statistically significant [Tukey's honest significant difference analysis ($p < 0.05$)] orthogonal responses in the distribution of the RR gene set across the two quadrants (Fig. 5*B* and *SI Appendix*, Fig. 10). These patterns, combined with the temporal variability in gene expression patterns, indicate a finely tuned response to the environment, which is distinctive for each diatom species. Although there are many potential controls on diatom dynamics in Narragansett Bay, including top-down processes like predation (67, 68), these patterns of RR gene expression suggest the presence of bottom-up nutrient control on diatom population dynamics in Narragansett Bay.

This work addresses fundamental knowledge gaps in how phytoplankton species are able to cooccur while they compete for the same basic resources. Cooccurring diatoms appear to have different functional capabilities in N and P metabolism, and this metabolic potential is modulated in field populations in a distinctive way for each diatom. These findings suggest that differential resource partitioning is occurring between these two diatoms in situ. Such resource partitioning could facilitate the vast diversity of the phytoplankton and the structure, function, and productivity of aquatic ecosystems. In culture studies, resource-related transcriptional changes have been shown to be tightly choreographed with changes in proteins, activities, and biochemical pools (56, 62, 69). If further work is similarly able to link the transcriptional patterns observed here with changes in enzymatic activities or uptake rates, then shifts in the RR gene sets might reflect aspects of the realized niche and how it differs between these two species. These detailed in situ transcriptional comparisons would not have been possible without proportionalization to metabolic capacity (STD), which provides a quantitative means to compare transcriptional patterns directly between species. This approach could be applied to other systems, organisms, or environmental parameters to identify responsive genes and proportionalize their expression, with the aim of answering similar questions about how cooccurring species adjust their cellular physiology to partition their niche space.

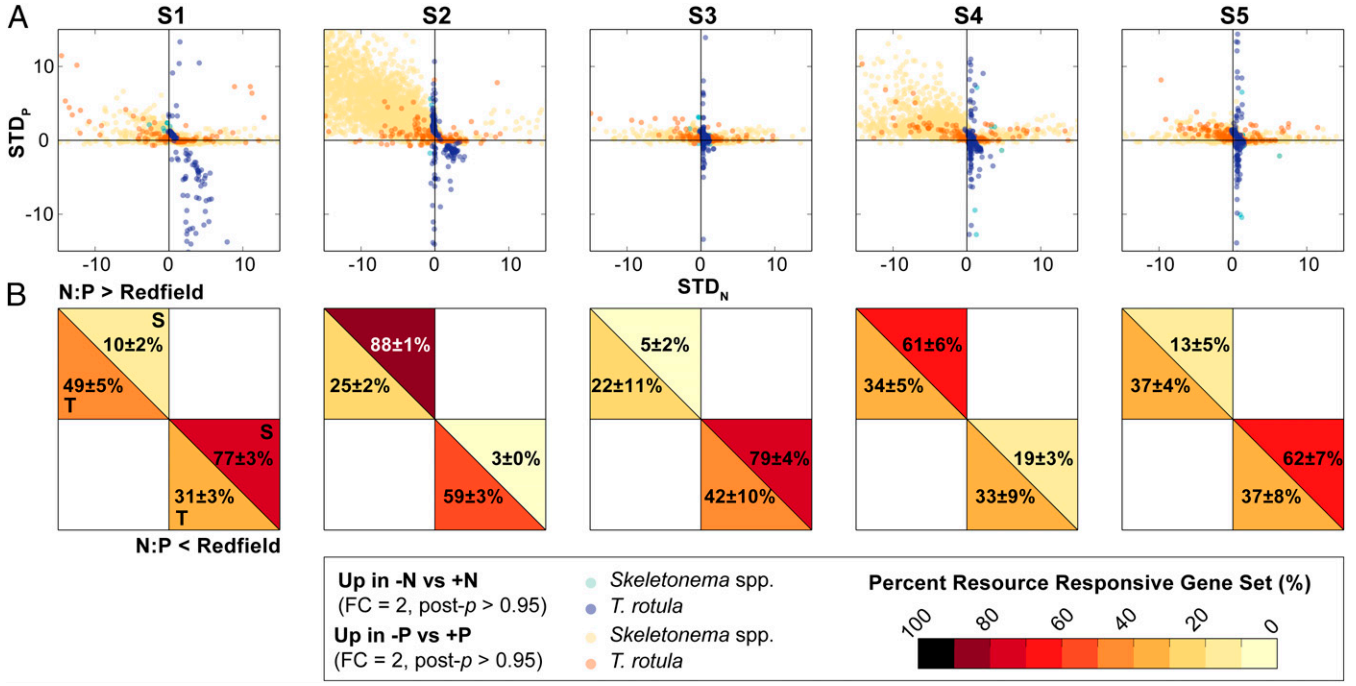


Fig. 5. Evolution of RR gene partitioning over time in Narragansett Bay for *T. rotula* and *Skeletonema* spp. (*A*) Stable gene normalized field signals for each gene identified as significantly (twofold change, post-$p > 0.95$) up-regulated in −P vs. +P for *Skeletonema* spp. (yellow) and *T. rotula* (orange) and in −N vs. +N for *Skeletonema* spp. (cyan) and *T. rotula* (dark blue) were proportionalized relative to the expression of those genes in nutrient incubations, yielding the STD_N and STD_P for each gene. These data are plotted for S1–S5. (*B*) Proportion of identified RR genes falling into the N/P > Redfield and N/P < Redfield quadrants for *T. rotula* (T) and *Skeletonema* spp. (S). FC, fold change.

Materials and Methods

Experimental Setup and Sample Collection. Surface seawater was collected and sampled for total community RNA at the long-term sampling site in Narragansett Bay (41°34′12″ N, 71°23′24″ W) during 2012 (May 16, May 21, May 30, June 4, and June 8, here called S1 through S5) in conjunction with the weekly time-series sampling effort. To diminish the influence of diel signals, samples were collected and processed between 0830 and 0900 local time. Near-surface water was collected in an acid-washed carboy and then filtered onto polycarbonate filters (5.0-μm pore size, 47 mm) using a peristaltic pump. Filters were then placed in cryovials and stored in liquid N until RNA extraction. In this manner, all samples were preserved within 15 min of collection. In addition to sampling for total community RNA, phytoplankton abundance was measured as part of the long-term weekly survey (70, 71).

A nutrient amendment incubation experiment was performed on May 30, 2012, with S3 representing the $t = 0$ of the experiment. Water collected in conjunction with S3 was prefiltered through 200-μm mesh to remove large zooplankton grazers and placed into acid-washed 2.5-L bottles. Triplicate bottles were then amended with nutrients to create five treatments: +N, +P, −N, −P, and ambient control. The +N and +P treatments were designed to eliminate the N and P stress signals, respectively, whereas the −N and −P treatments were supplemented with everything except the nutrient in question (e.g., the −N treatment was amended with P, Si, Fe, and vitamins) to force the drawdown of N and P, respectively (*SI Appendix*, Table 2). N and P amendment concentrations were selected to be ∼10-fold the seasonal average ambient N and P concentrations in the surface waters of Narragansett Bay measured at station II. The Si, Fe, and f/5 vitamin amendments were made in proportion to the f/5 media ratios (72). Bottles were placed in a flow-through incubator at ambient temperatures and photosynthetically active radiation (PAR) to mimic the collection depth. The incubation was run for 48 h, at which point all treatments were sampled for total community RNA as described above by filtering and snap-freezing 2 L of biomass from each replicate bottle.

RNA Extraction and Sequencing. Filters from triplicate bottles, representing ∼6 L of water, were pooled by treatment and extracted for each of the in situ and incubation experiment samples. RNA was extracted from individual filters with the RNeasy Mini Kit (Qiagen), following a modified version of the yeast protocol. Briefly, lysis buffer and RNA-clean zircon beads were added to the filter, and samples were vortexed for 1 min, placed on ice for 30 s, and then vortexed again for 1 min. Samples were processed following the yeast protocol. The resulting RNA was eluted in water and then treated for possible DNA contamination using a TURBO DNA-free Kit (Ambion) following the Rigorous Dnase protocol. RNA from each triplicate was then pooled by sample or treatment, using the RNA Cleanup Protocol from the RNeasy Mini Kit. The total RNA (>1,000 ng for each sample) was then enriched for eukaryotic mRNA through a poly-A pull-down onto oligo-dT beads. The resulting enriched RNA sample then went through library preparation with a TruSeq RNA Prep Kit (Illumina). Libraries were sequenced at the Columbia University Genome Center with an Illumina HiSeq2000. Each sample was sequenced to produce ∼60 million, 100-bp, paired-end reads (*SI Appendix*, Table 1). Raw sequence data quality was visualized using FastQC (73) and then cleaned and trimmed using Trimmomatic version 0.27 (paired-end mode; 6-bp-wide sliding window for quality below 20; minimum length of 25 bp) (74). All project sequence reads are available at the National Center for Biotechnology Information (NCBI) under accession number SRP055134.

Transcriptome and Genome Mapping. To assign a taxonomic identification to the reads, a database was created from transcriptomes made publicly available through the MMETSP as of March 17, 2014. In total, 401 transcriptomes from 209 species or cultured isolates were collected. Like-species transcriptomes were combined (regardless of strain or condition) using CD-HIT-EST (98% identity, word size of 9). The resulting clustered set of transcripts was considered to be the representative transcriptome for the species or cultured isolate. The 209 transcriptomes created in this manner were concatenated to form a comprehensive species-level transcriptome database from the MMETSP library. Due to the large size of the resulting MMETSP database, trimmed reads were mapped to the MMETSP using the Burrows–Wheeler aligner (75) and then counted using the HTSeq. 0.6.1 package (76).

Transcriptomes from two ecologically relevant diatom species in Narragansett Bay were selected: *S. costatum* RCC1716 (MMETSP0013, accessed from the publicly available transcriptome databases of the Moore Foundation Marine Microbiology Initiative-supported MMETSP, National Center for Genome Resources) and *T. rotula* CCMP3096 (a custom assembly available at NCBI under BioSample SAMN03349676). These transcriptomes were individually clustered using CD-HIT-EST (parameters: −c 0.98, −n 9) (77). The resulting clustered set of transcripts was then concatenated to form a reference transcriptome database. Trimmed reads from the field and incubation samples were mapped to this transcriptome database using

Bowtie2 version 2.2.1 (parameters: –a, –sensitive) (78). As a point of comparison, reads were also mapped using Bowtie2 version 2.2.1 under the same parameters to the genome of the model centric diatom species, *T. pseudonana* CCMP1335 (version 3.0), an organism not known to be abundant in Narragansett Bay. Mapped reads were then counted by transcript using the HTSeq. 0.6.1 python package (parameters: –m union, –s no) (76). Reads aligning to more than one full transcript were not counted. KEGG pathways were assigned to the assembled sequences with the online KEGG Automatic Annotation Server, using the bidirectional best-hit method to obtain KO annotations. In this study, only genes with a normalized count (NC) (raw count/total number of genes mapped to an organism) of at least two TPM in at least one of the field or incubation samples were included, thus limiting the sample set to 4,318 genes for *T. rotula* (19.3% of the transcriptome) and 20,921 genes for *Skeletonema* spp. (75.6% of the transcriptome). This difference in coverage is directly related to their relative abundance in the population.

Transcriptome Clustering. To assess relatedness of genes within *Skeletonema* spp. and *T. rotula*, the transcriptomes were translated using ORF predictor (proteomics.ysu.edu/tools/OrfPredictor.html) using a reference BLASTx alignment against the NCBI database with an 1e-5 cutoff (79). These translated peptide sequences were then combined with the translated proteins from the diatom genomes *Fragilariopsis cylindrus* CCMP1102 version 1.0, *P. tricornutum* CCMP632 version 2.0, *Pseudonitzschia multiseries* CLN-47 version 1.0, and *T. pseudonana* CCMP1335 version 3.0, which were collected from the Joint Genome Institute database (genome.jgi-psf.org/). A protein similarity network was then created using EGN, a software program that automates the reconstruction of gene networks from protein sequences through reciprocal BLASTp analysis (e-value <1e-5, 20% hit identity threshold, 5% best reciprocal threshold of best e-value, 90% minimal match coverage threshold) (80, 81). Networks were then visualized and manipulated using Cytoscape 3.0, where the layout of the network was produced using an edge-weighted, spring-embedded model based on e-value, meaning that genes that are closer together are more similar (82, 83). Known RR genes from previous transcriptome studies of the diatom species *T. pseudonana* were selected for analysis: (*i*) the P-responsive gene, Thaps_24435, which is an NPT (56) and (*ii*) the N-responsive gene, Thaps_25299, which is an assimilatory nitrate reductase (61).

Identification of Stable and Nutrient-Responsive Genes. Intercomparison of nutrient-incubation experiments enabled the identification of both nutrient-responsive genes and stably expressed reference genes for *T. rotula* and *Skeletonema* spp. For each organism, RR genes were identified by comparing the counts for that organism in +N to the –N incubation and the +P to the –P incubation, respectively, using analysis of sequence counts (ASC), an empirical Bayes method, which estimates the prior distribution from the data itself (57). ASC analyses were run using raw count data from each species separately. Genes were considered to be differentially regulated between treatments if for a fold change of 2.0, the posterior probability (post-*p*) was greater than 0.95 (56). After surveying the output of several different post-*p* cutoffs (*SI Appendix*, Fig. 11), stable genes were identified using ASC, as described by Alexander et al. (58), through pairwise comparisons of each of the incubation treatments (fold change of 1.25, post-$p < 0.1$).

Normalization of Metatranscriptome Data. Counts from the field were first normalized to the sequences belonging to the species in the library (Eq. **1**). For a particular species, *c*, the number of reads mapping to a gene *g*, $c_{i,g}$, was normalized to the sum of all of the counts across all genes for that organism, yielding the NC, similar to normalization techniques used for metatranscriptome data (32, 84):

NC (Eq. **1**):

$$NC_{i,g} = \frac{c_{i,g}}{\sum_{g \in G} c_{i,g}}. \quad [1]$$

Henceforth, only genes for which NC > 2 TPM in at least one sample (incubation or field) were considered. To facilitate interspecies comparisons, the NC was normalized to the geometric mean of the set of stable reference genes, *R*, yielding an SGNC. The calculation of an SGNC (Eq. **2**) for metatranscriptome data was designed to emulate the normalization used in qRT-PCR studies (85).

SGNC (Eq. **2**):

$$SGNC_{i,g} = \frac{NC_{i,g}}{\left(\prod^{R} NC_{i,g}\right)^{1/R}}. \quad [2]$$

The nutrient-responsive genes identified as differentially expressed in the nutrient incubations (*SI Appendix*, Table 2) were then selected for investigation in the field metatranscriptomes (S1–S5). The SGNCs from the field ($SGNC_{field}$) for these nutrient-related genes were bounded by the SGNCs from like nutrient incubations to calculate the STD_N and STD_P (Eqs. **3** and **4**).

STD_N (Eq. **3**):

$$STD_N = \frac{SGNC_{field} - SGNC_{+N}}{SGNC_{-N} - SGNC_{+N}}. \quad [3]$$

STD_P (Eq. **4**):

$$STD_P = \frac{SGNC_{field} - SGNC_{+P}}{SGNC_{-P} - SGNC_{+P}}. \quad [4]$$

For example, in calculating STD_N, the $SGNC_{field}$ is put in the range of the $SGNC_{+N}$ and $SGNC_{-N}$. In consequence, if the STD_N for a gene in the field equals 0, it is more similar in expression to the +N treatment, and if it equals 1, it is more similar in expression to the –N treatment. As such, a plot of STD_N against STD_P can divide the space into two main theoretical quadrants N/P > Redfield ($STD_P > 1$ and $STD_N < 0$) and N/P < Redfield ($STD_N > 1$ and $STD_P < 0$) (*SI Appendix*, Fig. 8). The total number of genes falling into each of the quadrants was counted by varying the bounds considered: the N/P > Redfield ratio quadrant ($STD_P > C$; $STD_N < C$ for $0.25 < C < 0.75$) and the N/P < Redfield ratio quadrant ($STD_P < C$; $STD_N > C$ for $0.25 < C < 0.75$). To approximate variation conservatively, the value of C was varied over 10 different values, and the average and SD for the percentages of genes falling into each of the quadrants were quantified. Similarity of data between species by quadrant was assessed using an ANOVA with a generalized linear model. The results from a post hoc Tukey test show the divergence of species across time ($P < 0.05$).

ACKNOWLEDGMENTS. We thank L. Whitney and M. Mercier for their efforts to grow *T. rotula* for the assembly. The MMETSP samples used in this study were sequenced, assembled, and annotated with the Assembly by Short Sequences (ABySS) pipeline at the National Center for Genome Resources. H.A. was supported by the Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program. This research was supported by funds from the National Science Foundation (NSF) Environmental Genomics and NSF Biological Oceanography Programs through Grant OCE-0723667 (to S.T.D., B.D.J., and T.A.R.) and Grant OCE-0962208 (to B.D.J.), and by the Joint Genome Institute/Department of Energy Community Sequencing Program through Grant CSP795793 (to B.D.J., S.T.D., and T.A.R.). Support was also provided by the Gordon and Betty Moore Foundation through Grant 2637 to the National Center for Genome Resources for the MMETSP.

1. Elton CS (1958) *The Ecology of Invasions by Animals and Plants* (Springer, Boston).
2. Cardinale BJ, et al. (2012) Biodiversity loss and its impact on humanity. *Nature* 486(7401):59–67.
3. Naeem S, Thompson LJ, Lawler SP, Lawton JH, Woodfin RM (1994) Declining biodiversity can alter the performance of ecosystems. *Nature* 368(6473):734–737.
4. Tilman D, et al. (2001) Diversity and productivity in a long-term grassland experiment. *Science* 294(5543):843–845.
5. Cadotte MW (2013) Experimental evidence that evolutionarily diverse assemblages result in higher productivity. *Proc Natl Acad Sci USA* 110(22):8996–9000.
6. Balvanera P, et al. (2006) Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol Lett* 9(10):1146–1156.
7. Tilman D, Wedin D, Knops J (1996) Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* 379(6567):718–720.
8. Behl S, Donval A, Stibor H (2011) The relative importance of species diversity and functional group diversity on carbon uptake in phytoplankton communities. *Limnol Oceanogr* 56(2):683–694.
9. Striebel M, Behl S, Stibor H (2009) The coupling of biodiversity and productivity in phytoplankton communities: Consequences for biomass stoichiometry. *Ecology* 90(8): 2025–2031.
10. Steiner CF, Long ZT, Krumins JA, Morin PJ (2005) Temporal stability of aquatic food webs: Partitioning the effects of species diversity, species composition and enrichment. *Ecol Lett* 8(8):819–828.
11. Ptacnik R, et al. (2008) Diversity predicts stability and resource use efficiency in natural phytoplankton communities. *Proc Natl Acad Sci USA* 105(13):5134–5138.
12. Nielsen ES (1960) Productivity of the oceans. *Annu Rev Plant Physiol* 11:341–362.
13. Strickland JD (1965) Phytoplankton and marine primary production. *Annu Rev Microbiol* 19:127–162.
14. Field CB, Behrenfeld MJ, Randerson JT, Falkowski PG (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281(5374): 237–240.
15. Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B (1995) Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with

regional data and relationship to biogenic sedimentation. *Global Biogeochem Cycles* 9(3):359–372.
16. Sournia A, Chrdtiennot-Dinet M-J, Ricard M (1991) Marine phytoplankton: How many species in the world ocean? *J Plankton Res* 13(5):1093–1099.
17. Tett P, Barton E (1995) Why are there about 5000 species of phytoplankton in the sea? *J Plankton Res* 17(8):1693–1704.
18. Hutchinson GE (1961) The paradox of the plankton. *Am Nat* 95:137–145.
19. Richerson P, Armstrong R, Goldman CR (1970) Contemporaneous disequilibrium, a new hypothesis to explain the "paradox of the plankton". *Proc Natl Acad Sci USA* 67(4):1710–1714.
20. Huisman J, Johansson AM, Folmer EO, Weissing FJ (2001) Towards a solution of the plankton paradox: The importance of physiology and life history. *Ecol Lett* 4(5): 408–411.
21. Huisman J, Weissing F (1999) Biodiversity of plankton by species oscillations and chaos. *Nature* 402(6760):407–410.
22. Roy S, Chattopadhyay J (2007) Towards a resolution of "the paradox of the plankton": A brief overview of the proposed mechanisms. *Ecol Complex* 4(1-2):26–33.
23. Menden-Deuer S, Rowlett J (2014) Many ways to stay in the game: Individual variability maintains high biodiversity in planktonic microorganisms. *J R Soc Interface* 11(95):20140031.
24. Connel JH (1980) Diversity and the coevolution of competitors, or the ghost of competition past. *Oikos* 35(2):131–138.
25. Margalef R (1963) On certain unifying principles in ecology. *Am Nat* 97:357–374.
26. Follows MJ, Dutkiewicz S, Grant S, Chisholm SW (2007) Emergent biogeography of microbial communities in a model ocean. *Science* 315(5820):1843–1846.
27. Johnson ZI, et al. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311(5768):1737–1740.
28. Sommer U (1983) Nutrient competition between phytoplankton species in multispecies chemostat experiments. *Arch Hydrobiol* 96(4):399–416.
29. Sommer U (1985) Comparison between steady state and non-steady state competition: Experiments with natural phytoplankton. *Limnol Oceanogr* 30(2):335–346.
30. Hutchins DA, Witter AE, Butler A, Luther GW (1999) Competition among marine phytoplankton for different chelated iron species. *Nature* 400(6747):858–861.
31. Zubkov MV, Fuchs BM, Tarran GA, Burkill PH, Amann R (2003) High rate of uptake of organic nitrogen compounds by Prochlorococcus cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl Environ Microbiol* 69(2):1299–1304.
32. Marchetti A, et al. (2012) Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc Natl Acad Sci USA* 109(6):E317–E325.
33. Frischkorn KR, Harke MJ, Gobler CJ, Dyhrman ST (2014) De novo assembly of *Aureococcus anophagefferens* transcriptomes reveals diverse responses to the low nutrient and low light conditions present during blooms. *Front Microbiol* 5:375.
34. Moustafa A, et al. (2010) Transcriptome profiling of a toxic dinoflagellate reveals a gene-rich protist and a potential impact on gene expression due to bacterial presence. *PLoS ONE* 5(3):e9688.
35. Gifford SM, Sharma S, Booth M, Moran MA (2013) Expression patterns reveal niche diversification in a marine microbial assemblage. *ISME J* 7(2):281–298.
36. Poretsky RS, Sun S, Mou X, Moran MA (2010) Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environ Microbiol* 12(3):616–627.
37. Rinta-Kanto JM, Sun S, Sharma S, Kiene RP, Moran MA (2012) Bacterial community transcription patterns during a marine phytoplankton bloom. *Environ Microbiol* 14(1):228–239.
38. Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA (2011) Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME J* 5(3):461–472.
39. Dupont CL, et al. (October 21, 2014) Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J*, 10.1038/ismej.2014.198.
40. Oviatt C, Buckley B, Nixon S (1981) Annual phytoplankton metabolism in Narragansett Bay calculated from survey field measurements and microcosm observations. *Estuaries* 4(3):167–175.
41. Borkman DG, Smayda T (2009) Multidecadal (1959–1997) changes in *Skeletonema* abundance and seasonal bloom patterns in Narragansett Bay, Rhode Island, USA. *Journal of Sea Research* 61(1-2):84–94.
42. Li Y, Smayda TJ (1998) Temporal variability of chlorophyll in Narragansett Bay, 1973–1990. *ICES J Mar Sci* 55(4):661–667.
43. Keeling PJ, et al. (2014) The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* 12(6):e1001889.
44. Sarno D, Kooistra WHCF, Medlin LK, Percopo I, Zingone A (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of *S. costatum*-like species with the description of four new species. *J Phycol* 41(1): 151–176.
45. Zingone A, Percopo I, Sims PA, Sarno D (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). I. A reexamination of the type material of *S. costatum* with the description of *S. grevillei* sp. nov. *J Phycol* 41(1):140–150.
46. Smayda TJ (2011) Cryptic planktonic diatom challenges phytoplankton ecologists. *Proc Natl Acad Sci USA* 108(11):4269–4270.
47. Nixon SW, Granger SL, Nowicki BL (1995) An assessment of the annual mass balance of carbon, nitrogen, and phosphorus in Narragansett Bay. *Biogeochemistry* 31(1): 15–61.
48. Shrestha RP, et al. (2012) Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics* 13(1):499.
49. North BB, Stephens GC (1972) Amino acid transport in *Nitzschia ovalis arnott*. *J Phycol* 8(1):64–68.
50. Serra JL, Llama MJ, Cadenas E (1978) Nitrate utilization by the diatom *Skeletonema costatum*: II. Regulation of nitrate uptake. *Plant Physiol* 62(6):991–994.
51. Bowler C, et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456(7219):239–244.
52. Allen AE, et al. (2008) Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc Natl Acad Sci USA* 105(30):10438–10443.
53. Smayda TJ (1974) Bioassay of the growth potential of the surface water of lower Narragansett Bay over an annual cycle using the diatom *Thalassiosira pseudonana*. *Limnol Oceanogr* 19(6):889–901.
54. Sakshaug E (1977) Limiting nutrients and maximum growth rates for diatoms in Narragansett Bay. *J Exp Mar Biol Ecol* 28(2):109–123.
55. Van Mooy BAS, et al. (2009) Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* 458(7234):69–72.
56. Dyhrman ST, et al. (2012) The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS ONE* 7(3):e33768.
57. Wu Z, et al. (2010) Empirical bayes analysis of sequencing-based transcriptional profiling without replicates. *BMC Bioinformatics* 11(1):564.
58. Alexander H, et al. (2012) Identifying reference genes with stable expression from high throughput sequence data. *Front Microbiol* 3:385.
59. Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 25(2):169–193.
60. McCarren J, et al. (2010) Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc Natl Acad Sci USA* 107(38):16420–16427.
61. Bender SJ, Durkin CA, Berthiaume CT, Morales RL, Armbrust EV (2014) Transcriptional responses of three model diatoms to nitrate limitation of growth. *Front Mar Sci* 1:3.
62. Wurch LL, Bertrand EM, Saito MA, Van Mooy BAS, Dyhrman ST (2011) Proteome changes driven by phosphorus deficiency and recovery in the brown tide-forming alga *Aureococcus anophagefferens*. *PLoS ONE* 6(12):e28949.
63. Dyhrman ST, et al. (2006) Long serial analysis of gene expression for gene discovery and transcriptome profiling in the widespread marine coccolithophore Emiliania huxleyi. *Appl Environ Microbiol* 72(1):252–260.
64. Bruhn A, LaRoche J, Richardson K (2010) *Emiliania huxleyi* (prymnesiophyceae): Nitrogen-metabolism genes and their expression in response to external nitrogen sources. *J Phycol* 46(2):266–277.
65. Ogawa N, DeRisi J, Brown PO (2000) New components of a system for phosphate accumulation and polyphosphate metabolism in *Saccharomyces cerevisiae* revealed by genomic expression analysis. *Mol Biol Cell* 11(12):4309–4321.
66. Kang L-K, Hwang S-PL, Lin H-J, Chen P-C, Chang J (2009) Establishment of minimal and maximal transcript levels for nitrate transporter genes for detecting nitrogen deficiency in the marine phytoplankton *Isochrysis glabana* (Prymnesiophycae) and *Thalassiosira pseudonana* (Bacillariophycae). *J Phycol* 45(4):864–872.
67. Martin JH (1970) Phytoplankton-zooplankton relationships in Narragansett Bay. IV. The seasonal importance of grazing. *Limnol Oceanogr* 15(3):413–418.
68. Lawerence C, Menden-Deuer S (2012) Drivers of protistan grazing pressure: Seasonal signals of plankton community composition and environmental conditions. *Mar Ecol Prog Ser* 459:39–52.
69. Bertrand EM, et al. (2012) Influence of cobalamin scarcity on diatom molecular physiology and identification of a cobalamin acquisition protein. *Proc Natl Acad Sci USA* 109(26):E1762–E1771.
70. Furnas M (1983) Community structure, biomass and productivity of size-fractionated summer phytoplankton populations in lower Narragansett Bay, Rhode Island. *J Plankton Res* 5(5):637–655.
71. Furnas MJ (1982) Growth rates of summer nanoplankton (<10 micrometer) populations in lower Narragansett Bay, Rhode Island, USA. *Mar Biol* 70:105–115.
72. Guillard RRL (1975) *Culture of Marine Invertebrate Animals*, eds Smith WL, Chanley MH (Springer, Boston).
73. Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. Available at: www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed March 31, 2014.
74. Lohse M, et al. (2012) RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res* 40(Web Server issue):W622–W627.
75. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.
76. Anders S, Pyl PT, Huber W (2015) HTSeq: A python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.
77. Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.
78. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
79. Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: Predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res* 33(Web Server issue):W677–W680.
80. Halary S, McInerney JO, Lopez P, Bapteste E (2013) EGN: A wizard for construction of gene and genome similarity networks. *BMC Evol Biol* 13(1):146.
81. Halary S, Leigh JW, Cheaib B, Lopez P, Bapteste E (2010) Network analyses structure genetic diversity in independent genetic worlds. *Proc Natl Acad Sci USA* 107(1): 127–132.
82. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T (2011) Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics* 27(3):431–432.
83. Saito R, et al. (2012) A travel guide to Cytoscape plugins. *Nat Methods* 9(11):1069–1076.
84. Ottesen EA, et al. (2011) Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *ISME J* 5(12):1881–1895.
85. Vandesompele J, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3(7):RESEARCH0034.