Vol. 77, No. 1

Diversity of Phytoplankton Nitrate Transporter Sequences from Isolated Single Cells and Mixed Samples from the East China Sea and mRNA Quantification[▽]

Lee-Kuo Kang,[1] Hsuan-Fan Wang,[1] and Jeng Chang[1,2]*

Institute of Marine Biology, National Taiwan Ocean University, Keelung 20224, Taiwan, Republic of China,[1] and Institute of Marine Environmental Chemistry and Ecology, National Taiwan Ocean University, Keelung 20224, Taiwan, Republic of China[2]

The transcript abundances of nitrate transporter genes (*Nrt2*) were proposed as potential markers for nitrogen deficiency in marine diatoms. To correctly quantify diatom *Nrt2* mRNA in the East China Sea (ECS), we utilized both mixed-species sequencing and single-cell PCR to expand the sequence database for this region. Using the single-cell method of PCR, 9 new diatom *Nrt2* sequences belonging to 5 genera, the *Nrt2* sequences of which have never been reported before, were obtained. On the other hand, 291 sequences homologous to *Nrt2* were retrieved from mixed-species sequencing using degenerate primers, and these sequences were clustered into 12 major groups according to a phylogenetic analysis. Based on sequence alignments, 11 pairs of group-specific PCR primers were designed to detect *Nrt2* mRNA levels in the ECS, and 3 of these primer pairs showed high specificity to target species. In ECS phytoplankton samples, environmental RNA was amplified via antisense RNA amplification followed by cDNA production. Subsequently, *Nrt2* transcript levels were readily detected using quantitative PCR. Our results indicated that investigating sequence diversity followed by careful primer design and evaluation is a good strategy to quantify the expression of genes of ecologically important phytoplankton.

Marine phytoplankton are responsible for more than 45% of the photosynthetic primary production on earth (16), and a proper supply of inorganic nutrients is essential for their growth. In many locations, nutrients are believed to be the major factor controlling phytoplankton distribution and proliferation. As a result, the development of a simple, accurate method to determine the nutrient-related physiological status has always been an active research topic in the study of phytoplankton ecology.

Among various nutrients, nitrogen is the macronutrient frequently mentioned to limit phytoplankton growth in the ocean (24). Recently, a group of nitrate transporter genes (*Nrt2*) was proposed as a potential marker for evaluating nitrogen deficiencies in eukaryotic phytoplankton (23, 26, 27). *Nrt2* genes encode high-affinity nitrate transporters, which are considered the primary transporter system responsible for nitrate uptake (23, 26, 27, 43). In addition, genes homologous to *Nrt2* were widely identified in bacteria, fungi, algae, and higher plants (for recent reviews, see references 12, 18, and 42). In most marine eukaryotic algae studied, *Nrt2* transcript levels are repressed in the presence of ammonium, moderately expressed under a nitrate-sufficient condition, and highly expressed under a nitrogen-deprived condition (26, 43). Also, the maximum and minimum transcript levels can be achieved by artificial removal and addition of nitrogenous nutrients, which is helpful in interpreting the physiological state of the original cell population (27).

To detect *Nrt2* expression in natural assemblages of phytoplankton, it is first necessary to construct a database so that highly efficient probes or primers can be designed. One approach is the use of universal primers/probes that target conserved regions of the *Nrt2* gene. However, data interpretation can become complicated because nitrogen requirements vary among phytoplankton species (13, 35, 36), which implies substantial variations in *Nrt2* expression as well. Another approach is to design species- or genus-specific primers/probes using genetically divergent regions. Data interpretation with this approach is straightforward and allows one to focus on ecologically important species. For example, coastal upwelling systems containing high concentrations of nutrients usually lead to a phytoplankton community dominated by a limited number of diatom species (9, 32, 49). *Nrt2* primers and probes targeting these diatoms can be used to elucidate how nitrogen utilization influences bloom formation and its spatial distribution.

There are several strategies for constructing a sequence database of marker genes. Some studies mainly constructed sequence databases from cultivated strains such as the nitrate reductase (NR) database (2). This strategy generates sequences with well-defined taxonomical associations, but the number of cultivable strains is very limited compared to the number of species in the ocean. On the other hand, mixed-species sequencing from field samples with degenerate primers results in highly diverse sequences but often contains unidentified clades with no particular affinity to any known species (1, 2). Single-cell PCR was proposed to be a complementary tool to obtain sequences from uncultivable protists that can be

---

* Corresponding author. Mailing address: Institute of Marine Biology, National Taiwan Ocean University, 2 Pei-Ning Rd., Keelung 20224, Taiwan, Republic of China. Phone: 886-2-24622192, ext. 5308. Fax: 886-2-24633152. E-mail: jengchang@mail.ntou.edu.tw.
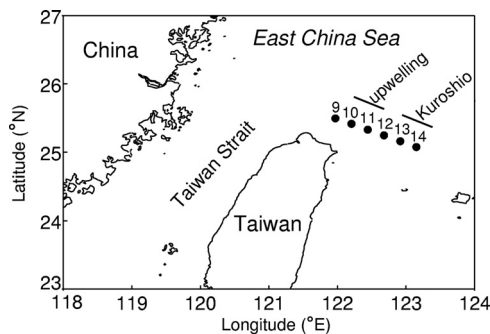
FIG. 1. Map of the southern East China Sea and locations of sampling stations. Regions occupied by shelf-break upwelling and an adjacent segment of the Kuroshio Current are marked above the corresponding stations. (Map adapted from reference 34 with permission.)

morphologically identified (14, 20, 37). Similar techniques are widely applied to genetic diagnosis and molecular differentiation in medical research (20, 21). Its applications to phytoplankton mainly focused on identifying marine dinoflagellates based on single-cell rRNA gene sequences (15, 22, 29, 30, 44).

In this study, we utilized both mixed-species sequencing and single-cell PCR methods to obtain *Nrt2* sequences from numerous diatom species in field samples. Subsequently, single-cell sequences were combined with known *Nrt2* sequences in GenBank, and the combined set was used to sort sequences of unknown taxonomical association from the southern East China Sea (ECS). On the basis of the alignment of these sequences, 11 sets of specific primers were designed, and their specificity and efficiency were evaluated. Our results indicated that such designed primers can reliably detect diatom *Nrt2* transcript levels in natural assemblages.

MATERIALS AND METHODS

Culture conditions. Two diatom cultures, *Chaetoceros affinis* CCMP 160 and *Thalassiosira oceanica* CCMP 1003, were obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (West Boothbay Harbor, ME). The unialgal culture of *Tetraselmis chui* clone TA was provided by H.-M. Su of the Tungkang Marine Laboratory, Pingtung, Taiwan (10). All cultures were grown at 20°C in f/2 enriched seawater medium (19a) under continuous illumination with an intensity of 145 μE m^{-2} s^{-1}. Aeration was not provided, but cultures were gently swirled daily to increase gas exchange.

Cruises and sampling. Two research cruises were conducted in the East China Sea (ECS) in the summer of 2007 on board the R/V *Ocean Researcher II*. The first cruise was conducted on July 9 and 10, and the second cruise was on September 1 to 3. Sampling was performed at 6 stations along a cross-shelf cruise track from the mid-shelf to the Kuroshio Current (Fig. 1). A detailed description of the hydrographic characteristics and nutrient concentrations can be found in reference 34. Phytoplankton samples were collected with a 20-μm-mesh plankton net with a mouth diameter of 0.5 m. At each station, an oblique tow was performed from a depth of 5 m for 10 min with the ship speed set at 1 knot. After the net tow, 2 subsamples were removed from the receiving bottle (cod end). One subsample was preserved with acidic Lugol's solution and stored at 4°C for microscopic examination. The other subsample was fixed in ethanol, replaced with fresh ethanol after 24 h, and then stored at −20°C until single-cell isolation. Phytoplankton remaining in the receiving bottle were collected with a 20-μm-mesh screen, transferred to cryotubes, and stored in liquid nitrogen for RNA and DNA isolation.

Total RNA and genomic DNA isolation. Samples frozen in liquid nitrogen were thawed by adding RLT buffer (Qiagen, Valencia, CA) with 1% β-mercaptoethanol (Sigma-Aldrich, St. Louis, MO), and then homogenized by supersonic disruption (sonicator ultrasonic processor XL; Heat System Ultrasonics, Farmingdale, NY). Total RNA was isolated using an RNeasy plant minikit (Qiagen) according to the manufacturer's instructions. The RNase-free DNase I set

(Qiagen) was applied for on-column digestion of residual DNA during RNA purification. The extracted RNA was eluted in diethyl pyrocarbonate-treated water. On the other hand, some frozen samples were used to extract genomic DNA using the phenol-chloroform method with the addition of cetyltrimethyl-ammonium bromide (11). The concentrations of RNA and DNA were determined with a spectrophotometer (ND-1000; NanoDrop Technologies, Wilmington, DE) at wavelengths of 260 and 280 nm.

Single-cell isolation and DNA extraction. Before isolation, phytoplankton in ethanol-fixed samples were transferred to Tris-EDTA (TE) buffer with 1 wash via centrifugation at 3,000 × g for 5 min. Diatom cells of interest were captured using an inverted microscope (IX51; Olympus, Tokyo, Japan) equipped with a microinjector attached to a micromanipulator (IM-9B; Narishige, Tokyo, Japan). Captured cells were washed several times with sterile TE buffer on a clean microscope slide to remove contaminants. Then, photographs were taken for species identification at a later time. Individual diatom cells were placed in 200-μl thin-walled PCR tubes containing 10 μl of TE buffer. DNA was extracted according to the enzymatic DNA extraction method with the addition of proteinase K (29, 31).

PCR amplification and DNA sequencing. Degenerate primers for amplifying *Nrt2*, 18S rRNA, and *rbcL* genes were designed from the conserved regions of homologous amino acid sequences in cyanobacteria, algae, and higher plants (Table l). Nested PCRs were performed with NrtF2 and NrtR7 primers in the first round and NrtF2 and NrtR5 primers in the second round. For sequencing *Nrt2* genes from ECS stations, 0.5 μg of total genomic DNA extracted from samples of water taken from individual stations was used as the template for the PCR. The PCR products were purified from a low-melting-point agarose gel and cloned into pGEM-T vectors (Promega, Madison, WI). DNA sequencing of the cloned fragments was performed using an ABI Prism 377A DNA sequencer with a PRISM Ready Reaction BigDye termination cycle sequencing kit (Applied Biosystems, Foster City, CA). In order to obtain gene fragments in a single cell, all crude lysates were used in the PCR. In addition, due to the minute amount of DNA extracted from a single cell, the product of the first-round PCR for *Nrt2* was used as the template to amplify 18S rRNA and *rbcL* gene fragments from the same cell. 18S rRNA gene fragments were amplified using primers 18S5N and 18S3C, and *rbcL* fragments were amplified by nested PCR using primers rbcLF1 and rbcLR1b in the first round and primers rbcLF2 and rbcLR1 in the second round (Table 1). The PCR products were purified, cloned, and sequenced as described above.

Phylogenetic tree construction. DNA sequences of *Nrt2* genes were translated to deduced amino acid sequences using DNAstar software (Lasergene, Madison, WI). The amino acid sequences were aligned using ClustalW (45), and highly fragmented regions were manually edited with the Bioedit sequence alignment editor version 5.0.9 (Department of Microbiology, North Carolina State University, Raleigh, NC) (28). The resultant alignment files were used to construct phylogenetic trees using the PHYLIP software (the PHYLogency Inference Package [http://evolution.genetics.washington.edu/phylip.html]). Pairwise distances were calculated using the Kimura formula in the PROTDIST program, and neighbor-joining trees were generated with the NEIGHBOR program. Bootstrap values were obtained with 1,000 bootstrap replicates to estimate the degree of confidence.

aRNA amplification and real-time Q-RT-PCR. Total RNA (1.0 μg) extracted from ECS samples was amplified using a MessageAmpII aRNA amplification kit (Ambion, Austin, TX) following the manufacturer's instructions. After amplification, 1.0 μg of antisense RNA (aRNA) was reverse transcribed into cDNA using random hexamers (Promega) and ImpromII reverse transcriptase (Promega) at 25°C for 10 min and 55°C for 1 h. Quantitative PCRs were initiated by adding the cDNA into a mixture containing 1× SYBR green PCR master mix (Applied Biosystems) and 300 nM concentrations of the forward and reverse primers (Table 1). In total, 11 sets of specific primers were designed for various diatom species in divergent regions of aligned *Nrt2* sequences, and each primer possessed at least 4 mismatched nucleotide differences from other *Nrt2* sequences. The reactions were then carried out in a GeneAmp 7000 sequence detection system (Applied Biosystems). PCR settings were 95°C for 10 min followed by 40 cycles, with 1 cycle consisting of 15 s at 95°C and 1 min at 60°C for 1 min. The fluorescence intensity from the complex formed by SYBR green and the double-stranded PCR product was continuously monitored from cycles 1 to 40. The threshold cycle (C_T) at which the fluorescence intensity became higher than a preset threshold was used to compare gene transcript levels. Since PCR quantifications using different primer sets were carried out with equal amounts of cDNA from the same aRNA sample, expression levels between *Nrt2* homologs were compared by the comparative C_T method according to the manufacturer's instructions (4). Each quantitative reverse transcription-PCR (Q-RT-PCR) was performed in triplicate to enable the calculation of a coefficient of

TABLE 1. Primers used in this study

Primer	Orientation	Nucleotide sequence[a] (5′→3′)	Target gene
Degenerate primers			
NrtF2	Forward	GSI GGMTGG GGI AAY MTI GGI GGI GG	*Nrt2*
NrtR5	Reverse	GMI YIC WRA KRC GIC CRC GCA TIC C	*Nrt2*
Nrtr7	Reverse	GCW CCR RYR TTK CCI CCI GCI CC	*Nrt2*
rbcLF1	Forward	GSW ACW TGG ACN RYW GTW TGG AC	*rbcL*
rbcLF2	Forward	ATW AAY TCW CAA CCN TTY ATG CG	*rbcL*
rbcLR1b	Reverse	GGR TGN CCN AWW GTW CCA CCA CC	*rbcL*
rbcLR1	Reverse	TGN CCN AWW GTW CCA CCA CC	*rbcL*
Universal primers			
18S5N	Forward	GCG GTA ATT CCA GCT CCA ATA GC	18S rRNA
18S3C	Reverse	GTG TGT ACA AAG GGC AGG GAC G	18S rRNA
Primers for Q-RT-PCR			
Group04QF	Forward	TTC AAG CTT GGT ATG AGT GCT G	Group04 *Nrt2*
Group04QR	Reverse	ATC GGA GAT CTT GAG GAT AAG G	Group04 *Nrt2*
Group01QF	Forward	CGG AAT GAG CGA CGA GAA GG	Group01 *Nrt2*
Group01QR	Reverse	GCA GTC ATC GGA AAT CCT AAG A	Group01 *Nrt2*
Group07QF	Forward	TCA AGA CCG GCA TGT CCT CC	Group07 *Nrt2*
Group07QR	Reverse	GCG ACG CCG GAC ATG AAG C	Group07 *Nrt2*
Group10QF	Forward	TGA AAA TCC TGC CGA GAG AGC	Group10 *Nrt2*
Group10QR	Reverse	CAT CAC TAG TTG TAT AGA TCA TAA C	Group10 *Nrt2*

[a] Nucleotide single-letter code: A, adenosine; C, cytosine; G, guanine; T, thymine; W, A or T; Y, C or T; R, A or G; M, A or C; K, G or T; S, G or C; N, A or T or C or G; I, inosine.

variation. The specificity of the Q-RT-PCR products was confirmed by a melting temperature analysis performed on the GeneAMP 7000 machine (Applied Biosystems) from 65 to 95°C for 20 min, and it was also examined by electrophoresis on 3% agarose gels containing 0.5× Tris-boric acid-EDTA buffer.

Phytoplankton identification and enumeration. Samples preserved in Lugol's solution were placed in Sedgewick-Rafter counting slides (Hausser Scientific, Horsham, PA) and examined with a Nikon Optiphot-2 microscope at a magnification of ×100. All cells of the entire counting slide were examined and enumerated. If there were fewer than 200 cells on a counting slide, additional counting slides were prepared from the same sample and examined similarly to obtain statistically meaningful results.

Nucleotide sequence accession numbers. The sequences obtained in this study were deposited in GenBank under the accession numbers HM347518 to HM347564.

RESULTS

***Nrt2* sequences from the single-cell PCR.** With the help of a micromanipulator, 69 diatom colonies were isolated from samples from the East China Sea (ECS) and then 14 DNA fragments homologous to *Nrt2* genes (named ECS_Sp01 [Sp01] to ECS_Sp14 [Sp14]) were obtained by single-cell PCR. In addition to morphological identification, taxonomical affiliations of these colonies were further confirmed by 18S rRNA and *rbcL* gene sequences amplified from the same colony (Table 2). In agreement with the morphological identification, 18S rRNA gene sequences of Sp01 and Sp02 were 94% identical to that of *Lauderia annulate*, and the *rbcL* sequence of Sp01 was 97% identical to that of *L. annulate*. Sp03 and Sp04 were morphologically identified as 2 species of *Chaetoceros*, and their 18S rRNA gene sequences were highly similar to those of different *Chaetoceros* spp. (Table 2). Sp05, Sp06, and Sp07 were morphologically identified as *Thalassionema*, *Rhizosolenia*, and *Odontella*, respectively, and their *rbcL* sequences supported these results. Similar results were observed for Sp08 and Sp09, which were identified as *Coscinodiscus* by both methods. In contrast, the taxonomical associations of Sp10 to Sp14 were

less clear. Sp10 was morphologically identified as *Guinardia*, but its 18S rRNA gene sequence did not match the known sequence of *Guinardia delicatula* well (73% identity; GenBank accession no. AY485487). Sp11 and Sp12 were identified as *Pseudonitzschia* and *Asterolampra*, respectively, by microscopic examination, but efforts to obtain 18S rRNA and *rbcL* gene sequences were unsuccessful (Table 2). Finally, Sp13 and Sp14 were small, short cylindrical cells with no obvious morphological characteristics for identification, and their 18S rRNA and *rbcL* gene sequences did not match any known sequences.

In order to properly categorize ECS *Nrt2* sequences obtained from environmental DNA samples, a reference tree was first constructed based on single-cell sequences with affirmative identification (ECS_Sp01 to -Sp09) and other *Nrt2*-homologous sequences from cyanobacteria, eukaryotic algae, and higher plants (Fig. 2). All sequences were translated into amino acid sequences before alignment, and the results indicated that NRT2 sequences belonging to cyanobacteria, haptophytes, chlorophytes, and diatoms formed 4 distinctive clades at the phylum level. In the diatom clade, multiple copies of *Nrt2* genes within the same species formed clusters with significant bootstrap support except for 4 sequences found in *Chaetoceros affinis*, *Chaetoceros muellerii*, *Phaeodactylum tricornutum*, and *Thalassiosira pseudonana*. At the genus level, however, genes from certain genera, such as *Chaetoceros* and *Thalassiosira*, were grouped together but without strong bootstrap support (<60%). Only the clade containing *Thalassiosira pseudonana Nrt2* and *Thalassiosira weissflogii Nrt2* genes was supported by bootstrap analysis (79%). These results indicated that the cluster boundary between genera was not always clear in this *Nrt2* phylogenetic tree.

***Nrt2* sequence diversity in the ECS.** In total, 291 gene fragments homologous to *Nrt2* were obtained from genomic DNA extracted from samples from the ECS. These gene fragments

TABLE 2. Identifications and taxonomical affiliations of 14 single-diatom colonies isolated by micromanipulator from samples from the East China Sea

Colony	Taxonomic identification by morphology	Taxonomic identification using[a]:		Accession no. of gene:		
		18S rRNA gene	*rbcL*	*Nrt2*	18S rRNA	*rbcL*
ECS_Sp01	*Lauderia*	*Lauderia annulate* (DQ514849, 94%)	*Lauderia annulate* (DQ514769, 97%)	HM347524	HM347538	HM347544
ECS_Sp02	*Lauderia*	*Lauderia annulate* (DQ514849, 94%)		HM347525	HM347539	
ECS_Sp03	*Chaetoceros*	*Chaetoceros* sp. (GQ243434, 91%)		HM347526	HM347540	
ECS_Sp04	*Chaetoceros*	*Chaetoceros* sp. (AJ535167, 93%)		HM347527	HM347541	
ECS_Sp05	*Thalassionema*		*Thalassionema frauenfeldii* (AY604698, 97%)	HM347528		HM347545
ECS_Sp06	*Rhizosolenia*		*Rhizosolenia shrubsolei* (FJ002128, 97%)	HM347529		HM347546
ECS_Sp07	*Odontella*		*Odontella sinensis* (Z67753, 93%)	HM347530		HM347547
ECS_Sp08	*Coscinodiscus*		*Coscinodiscus* sp. (FJ002134, 95%)	HM347531		HM347548
ECS_Sp09	*Coscinodiscus*		*Coscinodiscus* sp. (FJ002134, 94%)	HM347532		HM347549
ECS_Sp10	*Guinardia*	No match	No match	HM347533	HM347542	HM347550
ECS_Sp11	*Pseudonitzschia*			HM347534		
ECS_Sp12	*Asterolampra*			HM347535		
ECS_Sp13	Unidentified	No match	No match	HM347536	HM347543	HM347551
ECS_Sp14	Unidentified		No match	HM347537		HM347552

[a] The GenBank accession number and the percent similarity are shown in parentheses after the species. No match, no matching sequence found.

were easily identified as algal *Nrt2* without conspicuous intron regions. Only 4 fragments were found to contain a short span of intron sequences, and the intron boundaries matched the consensus intron sequences in diatoms (5). The taxonomical affiliations of ECS NRT2 sequences were determined by comparing them to sequences in the reference tree. According to the results of phylogenetic analysis, sequences located in the same clade and sharing more than 95% identities were collected as individual groups. The 291 ECS sequences were thus categorized into 12 groups called ECS_group01 to ECS_group12 (Fig. 3). Of the 12 groups, 10 groups were apparently composed of diatom NRT2 sequences. The other 2 groups, ECS_group11 and ECS_group12, occupied a branch between diatoms and haptophytes, and their taxonomical affiliations require further clarification. Within the diatom branch, some groups showed high similarities with known NRT2 sequences, and those relationships were supported by bootstrap values (>60%). For example, ECS_group04 very likely contained NRT2 sequences belonging to *Skeletonema* spp., and ECS_group05 very likely contained sequences belonging to *Cylindrotheca* spp. (Fig. 3). The same conclusion could not be extended to other groups, such as ECS_group02 or ECS_group03, because of weak support from bootstrapping. Therefore, these uncertain groups were temporarily named genus-like according to the most similar genus in the phylogenetic tree.

A systematic, cross-shelf change was observed in the relative dominances of individual *Nrt2* groups (Table 3). In September 2007, group04 (*Skeletonema*) and group07 (*Coscinodiscus*-like) *Nrt2* sequences dominated the genomic DNA pool in the upwelling region (stations 9 to 11) but were replaced by group08 (*Phaeodactylum*-like), group10 (*Pseudonitzschia*-like), and group12 *Nrt2* sequences in the Kuroshio region (stations 12 to 14). However, the distribution pattern was not consistent with the species composition

determined by microscopic examination. Diatom abundances were relatively high in the upwelling region, and the dominant species were *Chaetoceros*, *Rhizosolenia*, and *Pseudonitzschia*. The abundances of these species gradually decreased toward the Kuroshio region, and the dominant species were replaced by *Bacteriastrum* and *Hemiaulus*.

Detection of group-specific *Nrt2* transcript levels. According to the alignment of *Nrt2* sequences, 11 sets of group-specific primers were designed. *Nrt2* transcript levels were detected using Q-RT-PCR with amplified aRNA from different ECS stations as the templates. Electrophoretic analysis of the Q-RT-PCR products showed that amplicons generated by primers specific for group04 (*Skeletonema*), group10 (*Pseudonitzschia*-like), and group01 (*Chaetoceros*) formed single bands with the expected sizes (Fig. 4A). The single-peak melting curves also demonstrated the purity and specificity of these amplicons (Fig. 4B). In addition, these primers readily detected *Nrt2* transcripts in a separate cruise (station 11 in July 2007). In contrast, the other 8 primer pairs generated either multiple bands or diffuse bands (Fig. 4) and were apparently not suitable for Q-RT-PCR assays in the ECS.

Using the 3 primer pairs with good specificities, quantitative relationships of *Nrt2* mRNA between diatom groups were determined in the ECS (Fig. 5). In September 2007 at station 9, the C_T values of the *Skeletonema* group and the *Pseudonitzschia*-like group were 32.5 and 39.0, respectively. Although a C_T at 39.0 approaches the detection limit of Q-RT-PCR, the difference still indicated that the *Nrt2* transcript level of the *Skeletonema* group was substantially higher than that of the *Pseudonitzschia*-like group. Similarly, the *Chaetoceros* group possessed the highest *Nrt2* transcript level at station 13 in September, and both the *Pseudonitzschia*-like and *Chaetoceros* groups were major contributors to *Nrt2* transcripts at station 11 in July (Fig. 5).
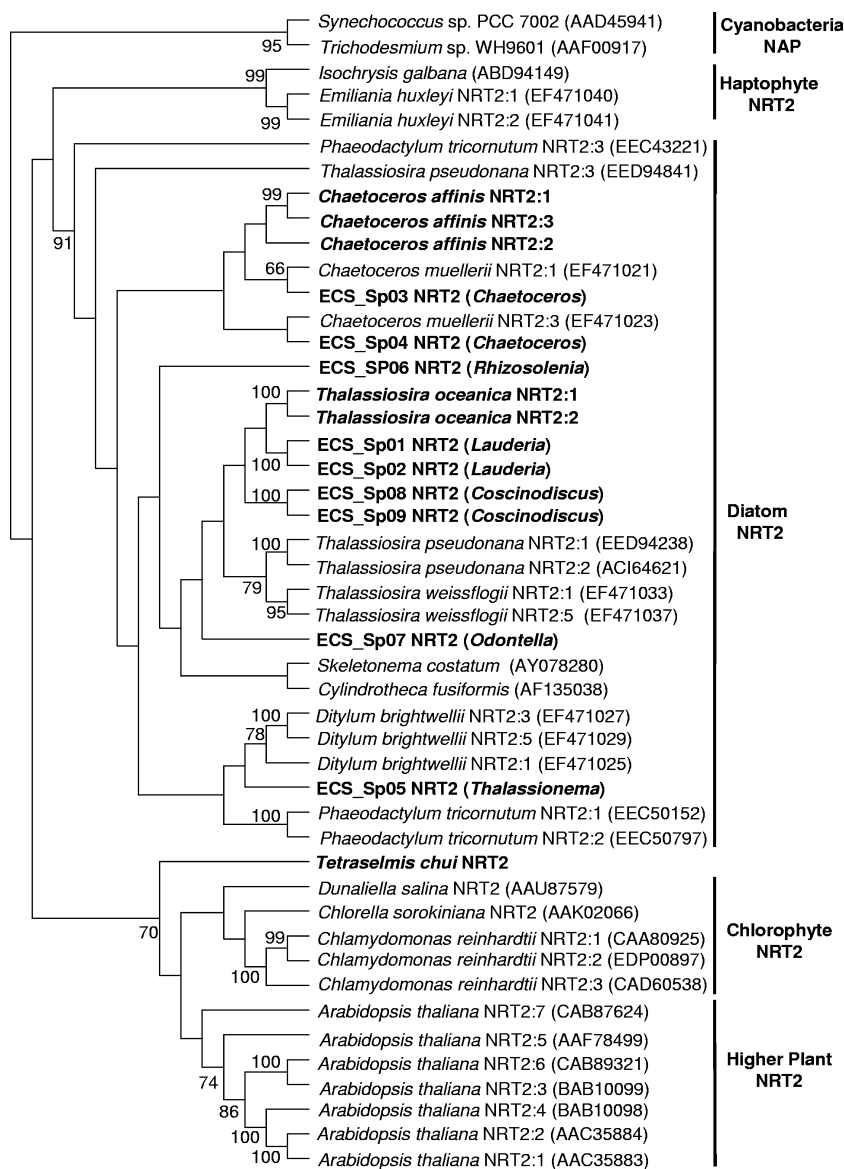
FIG. 2. Phylogenetic tree of nitrate transporter (NRT2) amino acid sequences in cyanobacteria, eukaryotic phytoplankton, and higher plants. The NRT2 sequences obtained in this study are shown in boldface type. Numbers at the nodes are the bootstrap values based on 1,000 resamplings, and only values that are >60% are shown. GenBank accession numbers are shown in parentheses.

DISCUSSION

The single-cell PCR is an effective method to expand the sequence database of functional genes in natural assemblages of phytoplankton. The ability to obtain rRNA and *rbcL* gene sequences from the same tube of cell lysates is also very powerful to confirm the taxonomical affiliation of target cells. This is especially true for species that lack clear morphological characteristics for identification. In this study, we obtained 9 new diatom *Nrt2* sequences with affirmative taxonomical affiliations (Table 2), including the *Nrt2* sequences of 5 genera that had never previously been reported. These single-cell sequences contributed more than 30% to diatom *Nrt2* sequences used in the reference phylogenetic tree (Fig. 3), which are

essential to an appropriate categorization of the sequences from the ECS environmental DNA samples.

In the reference phylogenetic tree of *Nrt2*, sequences were categorized into distinctive branches only at the phylum level (Fig. 2). Within diatoms, although sequences belonging to the same species often appeared on the same branch, the boundaries between genera were not always clear. This was in part due to the fact that a species might possess multiple forms of *Nrt2* (6). With more orthologs and paralogs continuously being created over time, the phylogenetic tree of *Nrt2* is not necessarily identical to that of a more-conserved gene such as that for 18S rRNA. A similar phenomenon was also observed in the phylogenetic analysis of diatom silicon transporters (3). On the bright side, such heterogeneity in *Nrt2* sequences within
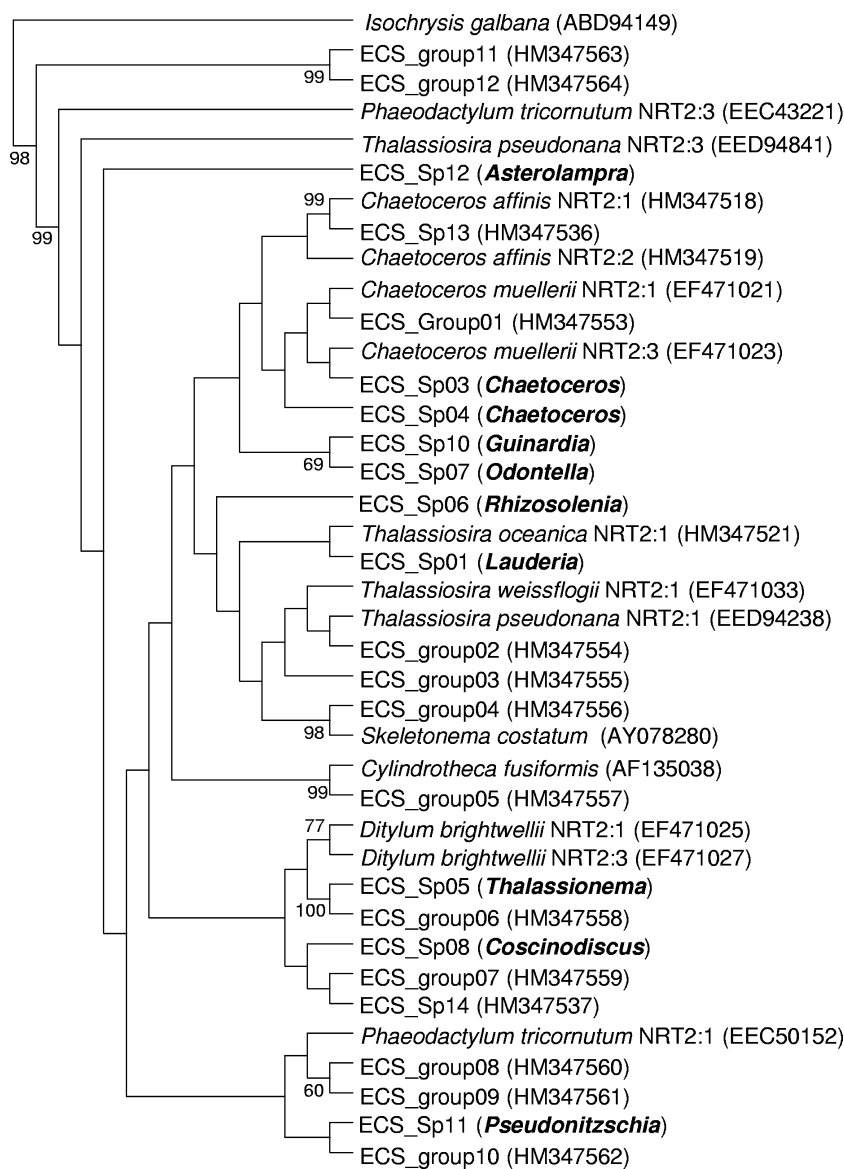
FIG. 3. Phylogenetic analysis of nitrate transporter (NRT2) sequences obtained from environmental DNA samples in the East China Sea. Selected diatom NRT2 sequences with known taxonomical associations were used as a basis for comparison, and the corresponding GenBank accession numbers are included in parentheses. This phylogenetic tree is a condensed version of a more-detailed analysis, and individual ECS_groups in the present alignment are represented by the most-conserved NRT2 sequence within that group. Numbers at the nodes are bootstrap values based on 1,000 resamplings, and only values that are >60% are shown.

the same genus provides ample opportunities for the design of group-specific primers.

The cross-shelf distributions of *Nrt2* sequences were not consistent with the diatom species composition based on microscopic examination (Table 3), which indicates preferential amplification of some species while others were ignored. For example, *Eucampia* and *Hemiaulus* were common in our samples, but single-cell PCR failed to generate any *Nrt2* sequence from these species. Similar biases were observed in an analysis of ribosomal genes (17). Even when "universal" primers are used, the PCR results might still be incomplete and not include all sequences in a community (7). In this study, *Nrt2* sequences obtained from the ECS might have disproportionately come

from a subset of species that are closely related to those used in the primer design, including *Cylindrotheca*, *Skeletonema*, and *Thalassiosira*. A more-thorough coverage of *Nrt2* sequences in marine environments will rely on certain primer-free methods such as metagenomic sequencing (40).

In order to ensure specificity, the group-specific *Nrt2* primers were placed in divergent regions as revealed by sequence alignments. Each group-specific primer possessed at least 4 mismatched nucleotides compared to homologous sequences belonging to other groups, and a single mismatch on the 3′ end was added to further reduce the between-group PCR amplification (8, 19, 33). For example, in the 29 *Nrt2* sequences contained in ECS_group04, 22 sequences exactly matched the

TABLE 3. Distributions of *Nrt2* homologs and diatom species composition at various stations in the southern East China Sea during the September cruise in 2007

Phytoplankton group or parameter	% of *Nrt2* homologs[a]						Diatom relative abundance (%)[b]					
	St. 9	St. 10	St. 11	St. 12	St. 13	St. 14	St. 9	St. 10	St. 11	St. 12	St. 13	St. 14
Nrt2 obtained[c]												
Skeletonema (group04)	58	54	26	32	2			2	4	1		
Coscinodiscus-like (group07)	20	28	52	22	5							
Phaeodactylum-like (group09)	15			2								
Thalassiosira-like (group02)	4	12	9	7								
Chaetoceros (group01)		3					13	33	52	15	3	6
Thalassionema (group06)		2		2				6	4			
Thalassiosira-like (group03)		2	11	2								
Cylindrotheca (group05)					7		16	5	1	4	3	
Phaeodactylum-like (group08)	2			3	43							
Pseudonitzschia-like (group10)						32						
ECS_group11				5								
ECS_group12			2	20	41	68						
Other diatoms	2			7	2							
Nrt2 not obtained												
Rhizosolenia							28	3	1		6	
Pseudonitzschia							3	22	9		6	
Bacteriastrum							9	5	6	29	18	28
Hemiaulus							7	3	2	40	38	44
Other diatoms							24	21	21	11	26	22
n	55	65	46	60	44	22	2,600	5,400	3,463	900	425	225

[a] The percentage of homologs from samples taken from stations (St.) 9 to 14 belonging to a certain diatom group using total *Nrt2* homologs obtained from mixed-species sequencing as the basis of normalization. *n* is the total number of *Nrt2* sequences obtained at each station.

[b] Diatom relative abundance is expressed as the percentage of diatom cells belonging to a certain taxonomical group based on microscopic observation. *n* is the number of diatom cells ml^{-1} in net-collected samples at each station.

[c] Species or group from which the *Nrt2* sequences were obtained.

group04 forward primer in the annealing region, such as ST10N043 and ST10N123 (Fig. 6). The other 7 sequences possessed 1 or 2 internal mismatches in the annealing region, such as ST10N123 (Fig. 6). On average, sequence identity in the annealing region between group04 *Nrt2* sequences and

their corresponding primers reached 0.99. In contrast, the average identity between group04 primers and other nongroup04 sequences dramatically decreased to 0.44, which explained the good performance of this primer pair.

In this report, an SYBR green-based assay was used to quantify *Nrt2* mRNA in diatoms. Highly specific primers were designed to detect gene expression in a particular group of organisms. A somewhat different approach was adopted to quantify *rbcL* mRNA levels in diatoms and pelagophytes (25, 48). In this alternative approach, degenerate primers were used to amplify *rbcL* cDNA from a broad array of phytoplankton, including diatoms, pelagophytes, haptophytes, dinoflagellates, etc. Next, *rbcL* mRNA in target species was quantified by
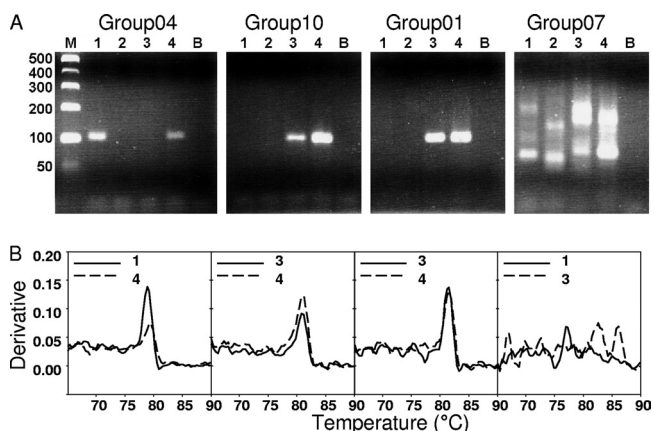

FIG. 4. Specificity of group-specific *Nrt2* primers in the quantitative reverse transcription-PCR (Q-RT-PCR) assay. (A) Electrophoretic analysis of Q-RT-PCR products. The templates were amplified antisense RNAs (aRNAs) from the East China Sea (ECS) in 2007. Lanes 1, 2, and 3, samples from the ECS at stations 9, 11, and 13 from the September cruise; lane 4, sample from the ECS at station 11 from the July cruise; lanes B, no-template control. Molecular weight markers are shown in lane M. (B) Melting curve analysis. The numbers with the lines in the symbol key indicate sample identities as explained above for panel A.
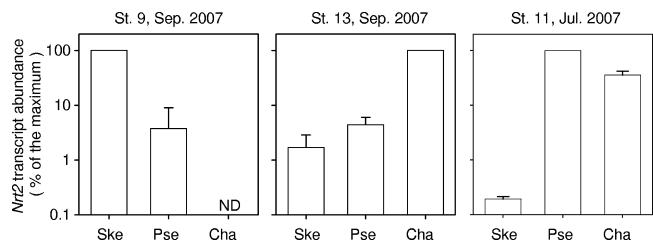

FIG. 5. Group-specific *Nrt2* transcript levels detected by Q-RT-PCR. *Nrt2* transcript levels are shown as relative percentages of diatom groups using the maximum mRNA level at individual stations (stations [St.] 9, 13, and 11) as 100%. The diatom groups were the *Skeletonema* group (group04) (Ske), *Pseudonitzschia*-like group (group10) (Pse), and *Chaetoceros* group (group01) (Cha). ND indicates that the transcript level was lower than the detection limit. The error bars represent 1 standard deviation.
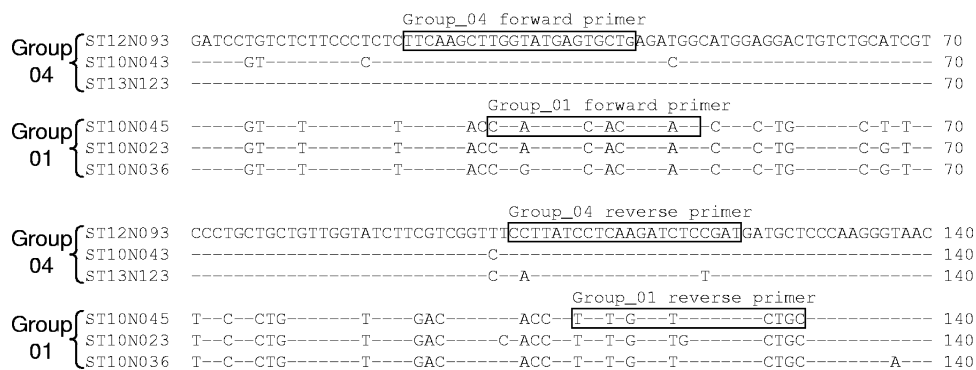
```
                                      Group_04 forward primer
Group ┌ST12N093  GATCCTGTCTCTTCCCTCTC┌TTCAAGCTTGGTATGAGTGCTG┐AGATGGCATGGAGGACTGTCTGCATCGT 70
  04  ┤ST10N043  -----GT---------C----------------------------------C-------------------------- 70
      └ST13N123  ----------------------------------------------------------------------------- 70
                                      Group_01 forward primer
Group ┌ST10N045  -----GT---T--------T------AC┌C--A-----C-AC----A-┐C---C-TG------C-T-T-- 70
  01  ┤ST10N023  -----GT---T--------T------ACC--A-----C-AC----A--C---C-TG------C-G-T-- 70
      └ST10N036  -----GT---T--------T------ACC--G-----C-AC----A--C---C-TG------C-G-T-- 70
                                      Group_04 reverse primer
Group ┌ST12N093  CCCTGCTGCTGTTGGTATCTTCGTCGGTTT┌CCTTATCCTCAAGATCTCCGAT┐GATGCTCCCAAGGGTAAC 140
  04  ┤ST10N043  -----------------------------------C-------------------------------------- 140
      └ST13N123  -----------------------------------C--A---------------T------------------- 140
                                      Group_01 reverse primer
Group ┌ST10N045  T--C--CTG-------T----GAC-------ACC-┌T--T-G---T--------CTGC┐----------- 140
  01  ┤ST10N023  T--C--CTG-------T----GAC-------C-ACC--T--T-G---TG-------CTGC----------- 140
      └ST10N036  T--C--CTG-------T----GAC-------ACC--T--T-G---T-------CTGC--------A-- 140
```

FIG. 6. Nucleotide sequence alignments of representative *Nrt2* sequences from ECS_group04 and ECS_group01. The primer sequences are boxed, and nucleotides identical to those in the top row are indicated by dashes.

applying group-specific TaqMan probes. An advantage of the TaqMan probe-based assay is that *rbcL* sequences from a spectrum of phytoplankton species can be simultaneously amplified from environmental mRNA. Difficulties associated with this approach include the possibility that an overdominant template in mixed assemblages may repress the amplification of minor templates (48). Cross-specificity between unrelated taxonomical groups was also observed such as between diatoms and pelagophytes (25). In comparison, using highly specific primers, our SYBR green-based assay quantified mRNA abundances at a more-focused genus-to-family level. However, sensitivity can be a problem because each pair of primers targets only a very small subset of total mRNA in environmental samples.

In order to analyze the transcript abundance from a small number of templates, amplification through the production of aRNA (aRNA amplification) is a commonly used technique (for reviews, see references 38 and 46). Since the aRNA method can maintain the relative levels of constituent mRNAs faithfully (39, 47), transcript levels quantified by various group-specific primer sets should be comparable within the same sample, since equal amounts of cDNA from the same batch of aRNA were used as the template (Fig. 5). However, comparisons of the levels of expression between different stations or cruises require the aid of reference genes, such as the histone H4 gene (41). If one of these genes is selected for the ECS study, its sequence diversity, target specificity, and amplification efficiency should be tested in a manner similar to the way *Nrt2* was tested. Normalization of *Nrt2* mRNA against the transcript level of a reference gene is necessary to clarify whether an increased transcript level is a result of elevated abundance of a target species or an actively expressed gene under nitrogen stress.

In conclusion, our results indicated that a survey of diatom *Nrt2* homologs combined with a single-cell PCR can be used to design group-specific primers to detect mRNA levels at the genus-to-family level. After proper RNA amplification, the *Nrt2* transcript levels of a group of closely related species can be detected in environmental samples by Q-RT-PCR. To make transcript levels from different samples comparable, a reference gene should be selected in the future. In addition, the concerns about multiple *Nrt2* homologs in each species and species-unique ranges of gene expression can be solved by

parallel incubations with ammonium addition and N deprivation (26, 27). Through this strategy, transcript levels can be used as indicators for the physiological state of eukaryotic phytoplankton in marine environments.

REFERENCES

1. **Adhitya, A., F. I. M. Thomas, and B. B. Ward.** 2007. Diversity of assimilatory nitrate reductase genes from plankton and epiphytes associated with a seagrass bed. Microb. Ecol. **54:**587–597.
2. **Allen, A. E., B. B. Ward, and B. K. Song.** 2005. Characterization of diatom (Bacillariophyceae) nitrate reductase genes and their detection in marine phytoplankton communities. J. Phycol. **41:**95–104.
3. **Alverson, A. J.** 2007. Strong purifying selection in the silicon transporters of marine and freshwater diatoms. Limnol. Oceanogr. **52:**1420–1429.
4. **Applied Biosystems.** 1997. Relative quantitation of gene expression: ABI PRISM 7700 sequence detection system. PerkinElmer, Foster City, CA.
5. **Armbrust, E. V.** 2000. Structural features of nuclear genes in the centric diatom *Thalassiosira weissflogii* (Bacillariophyceae). J. Phycol. **36:**942–946.
6. **Armbrust, E. V., J. A. Berges, C. Bowler, B. R. Green, et al.** 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science **306:**79–86.
7. **Baker, G., J. J. Smith, and D. A. Cowan.** 2003. Review and re-analysis of domain-specific 16S primers. J. Microbiol. Methods **55:**541–555.
8. **Bru, D., F. Martin-Laurent, and L. Philippot.** 2008. Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. Appl. Environ. Microbiol. **74:**1660–1663.
9. **Chen, Y. L. L.** 2000. Comparisons of primary productivity and phytoplankton size structure in the marginal regions of southern East China Sea. Cont. Shelf Res. **20:**437–458.
10. **Chung, C. C., S. P. L. Hwang, and J. Chang.** 2003. Identification of a high-affinity phosphate transporter gene in a pransiophyte alga, *Tetraselmis chui*, and its expression under nutrient limitation. Appl. Environ. Microbiol. **69:**754–759.
11. **Clark, C. G.** 1992. DNA purification from polysaccharide-rich cells, p. D-3.1–D-3.2. *In* J. J. Lee and A. T. Soldo (ed.), Protocols in protozoology. Society of Protozoology, Lawrence, KS.
12. **Crawford, N. M., and A. D. M. Glass.** 1998. Molecular and physiological aspects of nitrate uptake in plants. Trends Plant Sci. **3:**389–395.
13. **Dortch, Q.** 1990. The interaction between ammonium and nitrate uptake in phytoplankton. Mar. Ecol. Prog. Ser. **61:**183–201.
14. **Duff, R. J., H. Ball, and P. J. Lavrentyev.** 2008. Application of combined morphological-molecular approaches to the identification of planktonic protists from environmental samples. J. Eukaryot. Microbiol. **55:**306–312.
15. **Edvardsen, B., K. Shalchian-Tabrizi, K. S. Jakobsen, L. K. Medlin, E. Dahl,**

S. Brubak, and E. Paasche. 2003. Genetic variability and molecular phylogeny of *Dinophysis* species (Dinophyceae) from Norwegian waters inferred from single cell analyses of rDNA. J. Phycol. **39:**395–408.

16. Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281:**237–240.

17. Forney, L. J., X. Zhou, and C. J. Brown. 2004. Molecular microbial ecology: land of the one-eyed king. Curr. Opin. Microbiol. **7:**210–220.

18. Galvan, A., and E. Fernandez. 2001. Eukaryotic nitrate and nitrite transporters. Cell. Mol. Life Sci. **58:**225–233.

19. Ghedira, R., N. Papazova, M. Vuylsteke, T. Ruttink, I. Taverniers, and M. De Loose. 2009. Assessment of primer/template mismatch effects on real-time PCR amplification of target taxa for GMO quantification. J. Agric. Food Chem. **57:**9370–9377.

19a. Guillard, R. R. L. 1983. Culture of phytoplankton for feeding marine invertebrates, p. 108–132. *In* C. J. Berg (ed.), Culture of marine invertebrates. Hutchinson Ross Publishing Co., Stroudsburg, PA.

20. Hahn, S., X. Y. Zhong, and W. Holzgreve. 2002. Single cell PCR in laser capture microscopy. Laser capture microscopy and microdissection. Methods Enzymol. **356:**295–301.

21. Hahn, S., X. Y. Zhong, C. Troeger, R. Burgemeister, K. Gloning, and W. Holzgreve. 2000. Current applications of single-cell PCR. Cell. Mol. Life Sci. **57:**96–105.

22. Handy, S. M., T. R. Bachvaroff, R. E. Timme, D. W. Coats, S. Kim, and C. F. Delwiche. 2009. Phylogeny of four Dinophysiacean genera (Dinophyceae, Dinophysiales) based on rDNA sequences from single cells and environmental samples. J. Phycol. **45:**1163–1174.

23. Hildebrand, M., and K. Dahlin. 2000. Nitrate transporter genes from the diatom *Cylindrotheca fusiformis* (Bacillariophyceae): mRNA levels controlled by nitrogen source and by the cell cycle. J. Phycol. **36:**702–713.

24. Howarth, R. W., and R. Marino. 2006. Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems: evolving views over three decades. Limnol. Oceanogr. **51:**364–376.

25. John, D. E., S. S. Patterson, and J. H. Paul. 2007. Phytoplankton-group specific quantitative polymerase chain reaction assays for RuBisCO mRNA transcripts in seawater. Mar. Biotechnol. **9:**747–759.

26. Kang, L. K., S. P. L. Hwang, G. C. Gong, H. J. Lin, P. C. Chen, and J. Chang. 2007. Influences of nitrogen deficiency on the transcript levels of ammonium transporter, nitrate transporter and glutamine synthetase genes in *Isochrysis galbana* (Isochrysidales, Haptophyta). Phycologia **46:**521–533.

27. Kang, L. K., S. P. L. Hwang, H. J. Lin, P. C. Chen, and J. Chang. 2009. Establishment of minimal and maximal transcript levels for nitrate transporter genes for detecting nitrogen deficiency in the marine phytoplankton *Isochrysis galbana* (Prymnesiophyceae) and *Thalassiosira pseudonana* (Bacillariophyceae). J. Phycol. **45:**864–872.

28. Keane, T., C. Creevey, M. Pentony, T. Naughton, and J. McInerney. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6:**29.

29. Ki, J. S., and M. S. Han. 2005. Sequence-based diagnostics and phylogenetic approach of uncultured freshwater dinoflagellate *Peridinium* (Dinophyceae) species, based on single-cell sequencing of rDNA. J. Appl. Phycol. **17:**147–153.

30. Ki, J. S., and M. S. Han. 2007. Rapid molecular identification of the harmful freshwater dinoflagellate *Peridinium* in various life stages using genus-specific single-cell PCR. J. Appl. Phycol. **19:**467–470.

31. Ki, J. S., G. Y. Jang, and M. S. Han. 2004. Integrated method for single-cell DNA extraction, PCR amplification, and sequencing of ribosomal DNA from harmful dinoflagellates *Cochlodinium polykrikoides* and *Alexandrium catenella*. Mar. Biotechnol. **6:**587–593.

32. Kudela, R. M., W. P. Cochlan, and R. C. Dugdale. 1997. Carbon and nitrogen uptake response to light by phytoplankton during an upwelling event. J. Plankton Res. **19:**609–630.

33. Kwok, S., D. E. Kellogg, N. Mckinney, D. Spasic, L. Goda, C. Levenson, and J. J. Sninsky. 1990. Effects of primer template mismatches on the polymerase chain reaction-human immunodeficiency virus type-1 model studies. Nucleic Acids Res. **18:**999–1005.

34. Liu, H. C., G. C. Gong, and J. Chang. 2010. Lateral water exchange between shelf-margin upwelling and Kuroshio waters influences phosphorus stress in microphytoplankton. Mar. Ecol. Prog. Ser. **409:**121–130.

35. Lomas, M. W., and P. M. Glibert. 1999. Interactions between NH_4^+ and NO_3^- uptake and assimilation: comparison of diatoms and dinoflagellates at several growth temperatures. Mar. Biol. **133:**541–551.

36. Lomas, M. W., and P. M. Glibert. 2000. Comparisons of nitrate uptake, storage, and reduction in marine diatoms and flagellates. J. Phycol. **36:**903–913.

37. Lynn, D. H., and M. Pinheiro. 2009. A survey of polymerase chain reaction (PCR) amplification of unicellular protists using single-cell PCR. J. Eukaryot. Microbiol. **56:**406–412.

38. Nygaard, V., and E. Hovig. 2006. Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. Nucleic Acids Res. **34:**996–1014.

39. Poirier, G. M. C., and M. G. Erlander. 1998. Postdifferential display: parallel processing of candidates using small amounts of RNA. Methods **16:**444–452.

40. Raes, J., and P. Bork. 2008. Molecular eco-systems biology: towards an understanding of community function. Nat. Rev. Microbiol. **6:**693–699.

41. Siaut, M., M. Heijde, M. Mangogna, A. Montsant, et al. 2007. Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. Gene **406:**23–35.

42. Slot, J. C., K. N. Hallstrom, P. B. Matheny, and D. S. Hibbett. 2007. Diversification of NRT2 and the origin of its fungal homolog. Mol. Biol. Evol. **24:**1731–1743.

43. Song, B., and B. B. Ward. 2007. Molecular cloning and characterization of high-affinity nitrate transporters in marine phytoplankton. J. Phycol. **43:**542–552.

44. Takano, Y., and T. Horiguchi. 2006. Acquiring scanning electron microscopical, light microscopical and multiple gene sequence data from a single dinoflagellate cell. J. Phycol. **42:**251–256.

45. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.

46. Wadenback, J., D. Clapham, D. Craig, R. Sederoff, G. Peter, S. Von Arnold, and U. Egertsdotter. 2005. Comparison of standard exponential and linear techniques to amplify small cDNA samples for microarrays. BMC Genomics **6:**61.

47. Wang, E., L. D. Miller, G. A. Ohnmacht, E. T. Liu, and F. M. Marincola. 2000. High-fidelity mRNA amplification for gene profiling. Nature Biotechnol. **18:**457–459.

48. Wawrik, B., J. H. Paul, and F. R. Tabita. 2002. Real-time PCR quantification of *rbcL* (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes. Appl. Environ. Microbiol. **68:**3771–3779.

49. Wilkerson, F. P., R. C. Dugdale, R. M. Kudela, and F. P. Chavez. 2000. Biomass and productivity in Monterey Bay, California: contribution of the large phytoplankton. Deep Sea Res. Part II Top. Stud. Oceanogr. **47:**1003–1022.