CHARACTERIZATION OF DIATOM (BACILLARIOPHYCEAE) NITRATE REDUCTASE GENES AND THEIR DETECTION IN MARINE PHYTOPLANKTON COMMUNITIES

Andrew E. Allen, Bess B. Ward
Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA
and
Bongkeun Song
Department of Biological Sciences, University of North Carolina at Wilmington, Wilmington, North Carolina 28401-1003, USA

The complete assimilatory nitrate reductase (NR) gene from the pennate diatom *Phaeodactylum tricornutum* Bohlin was sequenced from cDNA and compared with NR sequences from fungi, green algae, vascular plants, and the recently sequenced genome of the centric diatom *Thalassiosira pseudonana* Hasle and Heimdal CCMP1335. In all the major eukaryotic nitrate reductase (Euk-NR) functional domains, diatom NR gene sequences are generally 50%–60% identical to plant and alga sequences at the amino acid level. In the less conserved N-terminal, hinge 1, and hinge 2 regions, homology to other NR sequences is weak, generally <30%. Two PCR primer sets capable of amplifying Euk-NR from plants, algae, and diatoms were designed. One primer set was used to amplify a 750-base pair (bp) NR fragment from the cDNA of five additional diatom strains. The PCR amplicon spans part of the well-conserved dimer interface region, the more variable hinge 1 region, and part of the conserved cytochrome *b* heme binding region. The second primer set, targeted to the dimer region, was used to amplify an approximately 400-bp fragment of the NR gene from DNA samples collected in Monterey Bay, California and in central New Jersey inner continental shelf (LEO-15 site) waters. Only diatom-like NR sequences were recovered from Monterey Bay samples, whereas LEO-15 samples yielded NR sequences from a range of photosynthetic eukaryotes. The prospect of using DNA- and RNA-based methods to target the NR genes of diatoms specifically is a promising approach for future physiological and ecological experiments.

**Key index words:** diatoms; DNA sequence; nitrate; nitrate reductase; nitrogen; upwelling

**Abbreviations:** Euk-NR, eukaryotic nitrate reductase; NR, nitrate reductase; RACE, rapid amplification of cDNA ends

New production, primary production in the sea that results from inputs of nitrogen from outside the euphotic zone, determines the amount of carbon available for utilization by higher trophic levels. It is largely regulated by the rate and magnitude of nitrate (NO$_3^-$) uptake and assimilation by pelagic microbes (Dugdale and Goering 1967). Within marine phytoplankton, diatoms are among the best competitors for high levels of NO$_3^-$ (>2 μM) (Wilkerson et al. 2000). Therefore, in regions where high rates of nutrient supply are sustained, such as upwelling environments, and on continental margins, diatoms often constitute a large fraction of the photosynthetic biomass (Kudela and Dugdale 2000). The ability of diatoms to thrive under upwelling-induced nutrient-rich conditions makes them the basis for the world’s shortest and most efficient food webs (Steele 1974). Some of the world’s largest fisheries are driven and maintained primarily by diatom-based new production (Jickells 1998). Diatom photosynthesis produces approximately 25% of the 45–50 billion tons of organic carbon fixed annually in the sea (Nelson et al. 1995).

Nitrogen uptake and assimilation differ in that uptake is defined as NO$_3^-$ transport across the membrane and NO$_3^-$ assimilation is the reduction and subsequent incorporation of reduced N (nitrogen) into cellular biomass. Diatoms, like other photosynthetic eukaryotes, reduce NO$_3^-$ to NO$_2^-$ using an assimilatory nitrate reductase enzyme (NR) (E.C.1.6.6.1-3) before incorporating N into biomass. Nitrate reduction appears to be the rate-limiting process in nitrogen acquisition (Campbell 1999). Regulation of the eukaryotic NR protein by different environmental conditions has been studied in vascular plants, green algae, and diatoms. For example, in the marine green alga *Dunaliella tertiolecta*, NR mRNA transcripts have been shown to be nitrate inducible and ammonium repressible, and transcription is not induced under nitrogen depletion (Song and Ward 2004). In the freshwater green algae *Chlorella vulgaris*, *Chlamydomonas reinhardtii*, and *Volvox carteri*, NR transcription was induced, more rapidly than in *D. tertiolecta*, by the removal of repressors such as ammonia and metabolites of ammonia (Cannons and Pendleton 1994, Quesada and Fernández 1994). In diatoms, NR enzyme assays indicate that NR activity...
is nitrate inducible and ammonium repressible as well as light activated and temperature dependent (Lomas and Gilbert 1999, 2000, Gao et al. 2000). Nitrate reductase activity in diatoms is quantitatively related to NO$_3^-$ incorporation under steady-state growth conditions (Berges and Harrison 1995).

Information currently available concerning the regulation of NR in diatoms was derived from in vitro experiments using protein extracts or antibody probing for protein abundance (Smith et al. 1992, Berges and Harrison 1995, Vergara et al. 1998, Lomas 2004). Although these studies have undoubtedly increased our understanding of factors that regulate N uptake and metabolism in diatoms, there are limitations to these methods. Protein based enzyme activity assays depend on the extraction of intact protein in the presence of saturating substrates and cofactors, and it is difficult to relate in vitro measurements of enzyme activity quantitatively to in situ processes (Berges and Harrison 1995).

Likewise, $^{15}$N tracer techniques, which historically have been used to estimate rates of NO$_3^-$ incorporation and to investigate the effect of environmental factors on assimilation and uptake rates, have well-documented limitations. Additions of labeled nitrogen may alter the dynamics of nutrient-limited systems (Dugdale and Wilkerson 1986). The $^{15}$N based methods can also suffer from containment artifacts such as enhanced grazing and altered light (Collos et al. 1993). Finally, and most importantly, $^{15}$N techniques and enzyme based assays do not provide a high degree of phylogenetic resolution when applied to water samples. It is very difficult to estimate nitrogen incorporation or protein activity for particular taxa, species, or even broadly defined functional groups (i.e. photosynthetic organisms vs. heterotrophic bacteria).

One means to overcome some of these problems involves conducting assays to collect information at the level of DNA and RNA. To address ecological questions concerning variability in the diversity, abundance, and expression of eukaryotic NR genes, it is first necessary to construct a sequence database so that appropriate probes can be designed to detect groups and clades of interest. We present here the first published, complete and partial, NR sequences for diatoms. These data will enable the investigation of transcriptional variability within different groups of phytoplankton in response to availability of NO$_3^-$, NH$_4^+$, and dissolved organic nitrogen and will pave the way for in situ ecological investigations designed to evaluate the response of marine diatom populations to differing oceanic N regimes. A preliminary analysis of eukaryotic nitrate reductase (Euk-NR) genes retrieved from clone libraries indicates a dominance of diatoms in Monterey Bay samples and a higher proportion of diatoms relative to other types of phytoplankton at 1 m compared with 15 m in inner continental shelf waters off the central coast of New Jersey.

**MATERIALS AND METHODS**

**NR from diatom cultures.** Strains and growth conditions: Diatom strains used in this study were obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP, West Boothbay Harbor, ME, USA) and the Plymouth Culture Collection of Algae (PCC, Plymouth, UK). The following strains were used: Phaeodactylum tricornutum Bohlin (Bacillariophyceae, clone CCMP 630), Chaetoceros mulleri Lemmermann (Coscinodiscophyceae, clone CCMP 1316), Asterionellopsis glacialis (Castracane) Round (Fragilariophyceae, clone CCMP 139), Thalassiosira weissflogii (Grun.) Fryxell et Hasle (Coscinodiscophyceae, clone CCMP 1336), Coscinodiscus granii Gough (Coscinodiscophyceae, clone CCMP 1817), Entomoneis cf. alata Ehrenberg (Bacillariophyceae, clone CCMP 1522), Amorpha sp. (Bacillariophyceae, clone CCMP 1405), Thalassiosira oceanica (Hustedt) Hasle et Heimdal (Bacillariophyceae, clone CCMP 1005), and Skeletonema costatum (Bacillariophyceae, clone PCC 582).

Batch cultures (1 L) were grown under continuous fluorescent light at 18°C in 0.2 µm filtered and autoclaved New Jersey coastal seawater amended with 1/2 nutrients with NO$_3^-$ as the sole nitrogen source (Guillard 1975). Cells were harvested in late log phase by filtration onto a 47-mm diameter 3.0-µm pore size Durapore filter (Millipore, Bedford, MA, USA) and transferred by scraping with a sterile razor blade into a sterile 2.0-mL microcentrifuge tube.

**RNA isolation:** Total RNA was extracted from cultures using an Ambion RNAqueous-4PCR kit (Ambion, Austin, TX, USA) following the manufacturer’s instructions. The extracted RNA was treated with DNase (2 U) (1 h, 37°C) to eliminate genomic DNA contamination. The RNA integrity was examined by visualizing the RNA on a denaturing agarose gel. Quantity and purity of the RNA was evaluated by measuring the A260/A280 ratio with a spectrophotometer. Extended total RNA was stored at −80°C until used for experiments.

**Degenerate rapid amplification of cDNA ends (RACE) PCR primer design and overall strategy for NR genes in cultured algae:** Degenerate oligonucleotide primers for the assimilatory NR gene were designed based on a DNA alignment of three plant and three algal genes that encode the NR protein. Initially, three primers, NRPt547F, NRPt1000F, and NRPt1389R, were designed to be used in nested RACE PCR reactions (Table 1, Fig. 1). Briefly, the forward primer NRPt547F and the 5' RACE universal primer mix generated a PCR product that was used as the template in a nested PCR reaction with the primers NRPt1000F and NRPt1389R. This reaction produced robust 390-bp PCR products with each of the diatom samples. The PCR products were cloned and sequenced, and gene specific PCR primers were designed to amplify the complete NR coding sequence with 5’ and 3’ RACE PCR. The 5’ and 3’ RACE PCR products from *P. tricornutum* were cloned and sequenced because this culture amplified most robustly.

**Complete 3’ and 5’ RACE PCR products were obtained by using the NRPt3’ RACE and NRPt5’ RACE primers, which were designed specifically for the *P. tricornutum* NR gene (Table 1).** Primers were designed from sequence information obtained by cloning and sequencing the 390-bp PCR product obtained with degenerate primers (above). To obtain the 5’ region of the *P. tricornutum* NR gene, the reverse transcription reaction to synthesize 5’ RACE cDNA was modified. Total RNA was used as a template with a SMART oligo primer (BD Science Clontech, Palo Alto, CA, USA) and an NR gene specific reverse primer (NRPt5’ RACE), which replaced the oligo(dT)-primer in the manufacturer’s instruction. The 5’ RACE PCR amplification was performed with the *P. tricornutum* NR gene specific reverse primer (NRPt5’ RACE) and the 5’ RACE universal primer mix (Table 2).

**RACE PCR amplification:** Total RNA was used as a template to synthesize oligo(dT)-primed cDNA using a RACE...
The primers are labeled according to position numbers that refer to the base pair position in the *P. tricornutum* NR DNA sequence.

### Table 1. Oligonucleotide primers used in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5' to 3')</th>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRP547F</td>
<td>CGYAARGARAVATG</td>
<td>RKEQNM</td>
</tr>
<tr>
<td>NRP907F</td>
<td>GGYGGNNNGATGGTBAAGTGGCT</td>
<td>GGRMIKWL</td>
</tr>
<tr>
<td>NTP1000F</td>
<td>GRRGGHTTGGTGTACACGCC</td>
<td>GGWWYKPC</td>
</tr>
<tr>
<td>NRP1389R</td>
<td>GTTGTYYMACATBCCCAT</td>
<td>MGGMNN</td>
</tr>
<tr>
<td>NRP1727F</td>
<td>ATVGAATCDGKCCACCCNGGRTG</td>
<td>HPGGIDSI</td>
</tr>
<tr>
<td>NRP2325R</td>
<td>GGGVGATRCCCHGTVCDCDCVGC</td>
<td>AGGTGTFTP</td>
</tr>
<tr>
<td>NRP3'RACE</td>
<td>AAGAGACAGGCAACGGACTAGAAGTGT</td>
<td></td>
</tr>
<tr>
<td>NRP5'RACE</td>
<td>ACATTCCGTAGTCCGGTGGCTTGT</td>
<td></td>
</tr>
</tbody>
</table>

Degenerate *PCR* primer design and reverse transcriptase (RT)-*PCR* amplification: The complete genome sequence for the centric diatom *Thalassiosira pseudonana* CCMP1335 (http://genome.jgi-psf.org/thaps1/) became available after we had completed sequencing the *P. tricornutum* NR gene. The DNA sequence alignments were therefore constructed with the complete *P. tricornutum* and *T. pseudonana* NR sequences, complete mRNA sequences from vascular plant and alga representatives, and the 390-bp fragment obtained with the degenerate primers NRP1000F and NRP1389R from eight other diatoms (as described above).

Degenerate primers capable of amplifying NR gene fragments from vascular plants, green algae, and diatoms were selected. The NRP1000F and NRP1389R primers did not require modification, and three additional primers were designed: RP907F, NRP1727R, and NRP2325R (Table 1, Fig. 1). First round RT:cDNA synthesis reactions were primed with the reverse primer NRP2325R. Reactions consisted of 2.0 µL of 10 × RT buffer, 2.0 µL of dNTP mix (5 mM each), 1.0 µL of Omniscript Reverse Transcriptase (Qiagen, Valencia, CA, USA), 1 µL Rnasin (10 U/µL in 1× RT buffer) (Promega, Madison, WI, USA), approximately 100 ng of total RNA, 1.0 µM of the reverse primer NRP2325R, and RNase free water to 20 µL. Reverse transcriptase reactions were followed by nested PCR reactions with NRP907F and NRP2325R primers in the first round of PCR and NRP1000F and NRP1727R primers in the second round. A total of 2.0 µL of the RT reaction was used as a template in first round PCR reactions and 1.5 µL first round PCR product was used as the template in second round reactions (Table 2). The second round PCR reaction with primers NRP1000F and NRP1727R produced an approximately 730-bp product on each of the diatom templates. The PCR products from *S. costatum*, *T. glacialis*, *Amorpha*, and *C. mulleri* were cloned for sequencing because they generated the strongest PCR products. Cloned PCR products were sequenced M13F and M13R primers (see below).

NR sequences from seawater samples. Collection of environmental samples: Water samples from Monterey Bay were collected from 10 m at a mid-bay station (36° 45′ N, 122° 01′ W) in April and October 1998. Seawater (4 L) was concentrated using tangential flow filter cassettes (Filtron Ultrasette 300 KD open channel, Filtron Technology Corp., Northborough, MA, USA). The retentate was collected on 47-mm diameter 0.2 µm pore size Supor filters (Gelman, Covina, CA, USA) and frozen immediately. The filters were archived at −80°C until DNA was extracted.

Other seawater samples were collected from 1 and 15 m at the Node A site of the Rutgers University Marine Field Station Long-Term Ecosystem Observatory (LEO-15) (http://marine.rutgers.edu/mts/LEO/LEO15.html), which is located in inner continental shelf waters off the central coast of New Jersey (39° 27′ N, 74° 15′ W). Approximately 1 L of water from each depth was collected on a 0.2-µm sterile filter capsule (Millipore) using a peristaltic pump. Samples were collected at 12:00 on 22 July 2002, flash frozen in liquid nitrogen, transported on dry ice, and stored until analysis in a −80°C freezer at the lab.

DNA isolation: DNA was purified from the sterile filters using a DNA extraction kit (Gentra, Minneapolis, MN, USA) following the directions of the manufacturer, except that the volume of the lysis buffer and of all other reagents was doubled. Plankton DNA samples purified from seawater were used as the template for nested PCR reactions with NRP907F and NRP2325R primers in the first round and NRP1000F and NRP1389R primers in the second round.

![Fig. 1. Deduced model of *Phaeodactylum tricornutum* CCMP 630 nitrate reductase based on sequence homology to *Arabidopsis thaliana* NIA2 (GenBank accession no. J03240) (Campbell 1999). All primer locations are shown. The numbers, based on the *P. tricornutum* sequence, signify the boundaries of the five structural domains and the hinge regions. Domain identities are explained in the text.](image-url)
RESULTS AND DISCUSSION

Full-length diatom NR sequences. The full-length diatom NR amino acid sequences from *P. tricornutum* CCMP 630 and *T. pseudonana* CCMP 1355 3H are 65% identical to one another and are each, on average, 40% identical to NR sequences from green algae and vascular plants. In plants, the NR monomer is composed of five structural domains, the boundaries of which have been defined: Mo-molybdopterin, dimer interface, cytochrome b, FAD, and NADH (Campbell 1999) (Fig. 1). Three sequence regions without similarity to other proteins and varying in sequence among different NRs are the N-terminal region, the hinge 1 region, and the hinge 2 region. Phylogenetically, diatoms cluster together with high bootstrap support, separate from green and red algae, vascular plants, fungi, and the stramenopile *Phytophthora* (Fig. 2).

By aligning the complete diatom NR proteins with those from plants, we were able to estimate domain delineations for the diatom NR subunits (Figs. 1 and 3). Based on the sequence homology and organization of the major functional domains, diatom NR genes appear to be structurally very similar to plant and algal NR genes. Twenty-one invariant residues have been identified in plant Euk-NR genes that are important for functions such as binding of nitrate, heme, and FAD. Diatom sequences are conserved in all these positions except in the case of a serine residue, in the hinge 1 region that is thought to be important for light dependent regulatory phosphorylation (Tischner 2000). One unique feature that is well conserved between the *P. tricornutum* and *T. pseudonana* NR gene sequences is an insertion that is identical at 10 of 11 amino acid sites in the Mo-molybdopterin domain between the nitrate and molybdopterin binding residues (in the 275–300 residue region, Fig. 3).

The N-terminal region of diatom NR genes, which has received attention recently because of its potential role in temperature sensitivity (Berges et al. 2002), has a length of 82 and 89 amino acids for *P. tricornutum* and *T. pseudonana* respectively. The sequences of the hinge 1 region that is thought to be important for light regulation via phosphorylation and the hinge 2 region are important for functions such as binding of nitrate, heme, and FAD. Diatom sequences are conserved in all these positions except in the case of a serine residue, in the hinge 1 region that is thought to be important for light dependent regulatory phosphorylation (Tischner 2000). One unique feature that is well conserved between the *P. tricornutum* and *T. pseudonana* NR gene sequences is an insertion that is identical at 10 of 11 amino acid sites in the Mo-molybdopterin domain between the nitrate and molybdopterin binding residues (in the 275–300 residue region, Fig. 3).

The N-terminal region of diatom NR genes, which has received attention recently because of its potential role in temperature sensitivity (Berges et al. 2002), has a length of 82 and 89 amino acids for *P. tricornutum* and *T. pseudonana*, respectively. This is more similar in length to the approximately 90 amino acid residue N-terminal region in vascular plants than to the green algal N-terminal regions of approximately 50 amino acids. Aspartic and glutamic acid-rich regions, which are important for light regulation via phosphorylation in plants (Pigaglio et al. 1999), are missing in the diatom N-terminal regions. Overall, the diatom NR genes appear to be structurally very similar to those of other NR genes, although there are two short segments of three to five residues that are important for functions such as binding of nitrate, heme, and FAD. Diatom sequences are conserved in all these positions except in the case of a serine residue, in the hinge 1 region that is thought to be important for light dependent regulatory phosphorylation (Tischner 2000). One unique feature that is well conserved between the *P. tricornutum* and *T. pseudonana* NR gene sequences is an insertion that is identical at 10 of 11 amino acid sites in the Mo-molybdopterin domain between the nitrate and molybdopterin binding residues (in the 275–300 residue region, Fig. 3).

The N-terminal region of diatom NR genes, which has received attention recently because of its potential role in temperature sensitivity (Berges et al. 2002), has a length of 82 and 89 amino acids for *P. tricornutum* and *T. pseudonana*, respectively. This is more similar in length to the approximately 90 amino acid residue N-terminal region in vascular plants than to the green algal N-terminal regions of approximately 50 amino acids. Aspartic and glutamic acid-rich regions, which are important for light regulation via phosphorylation in plants (Pigaglio et al. 1999), are missing in the diatom N-terminal regions. Overall, the diatom NR N-terminal regions have very little homology to one another or to those of other NR genes, although there are two short segments of three to five residues that are identical between the diatoms and one segment that shares similarity with the green algal N-terminal regions (Fig. 3). The diatom N-terminal regions contain...
a conserved cysteine residue that is absent in the other lineages. The function of the N-terminal-region remains to be discovered. Berges et al. (2002) and Gao et al. (2000) demonstrated the relatively low temperature optima of *T. pseudonana* and other diatoms and suggested that diatoms and other chromophytic algal NR enzymes have a common feature that enables high activity at relatively low temperatures. Interestingly, as noted by Berges et al. (2002), a deletion of the NR N-terminal region in the plant *Nicotiana* caused a drop in the thermal optimum of the NR enzyme from 30°C to 15°C (Nussaume et al. 1995).

Partial diatom NR sequences from conserved and variable domains. The approximately 730-bp NR gene fragments that were amplified from additional diatom strains with the primers NRPt1000F and NRPt1727R were aligned with full-length genes from plants, green, and red algae. The aligned region is part of the well-conserved dimer interface region, the more variable hinge 1 region, and part of the conserved cytochrome *b* heme binding region (Fig. 3). As a group, the 7 diatom amino acid NR sequences are an average of 77% and 75% identical to one another and 53% and 50% identical to plant and algal NR sequences in the dimer interface and cytochrome *b* heme binding region of the NRpt1000F and NRpt1727R amplicon, respectively. In the less conserved hinge 1 region, however, diatom sequences are an average 58% identical to one another and only 25% identical to plant and algal sequences.

The approximately 390-bp gene fragments that were amplified with the NRpt1000F and NRpt1389R primers from additional diatom strains and DNA extracted from filtered seawater were translated into

---

**Fig. 2.** Inferred phylogenetic relationships of complete eukaryotic assimilatory nitrate reductase amino acid sequences. Bootstrap values greater than 50 (of 100) are shown at the nodes. The sulfite oxidase gene, a putative evolutionary ancestor of eukaryotic assimilatory nitrate reductase from *Arabidopsis thaliana*, was used to root the tree. GenBank accession numbers are given next to their designated sequence. Sequences contain approximately 900 amino acids.
Fig. 3. continued on next page
amino acids and aligned. Phylogenetic analysis of the alignment, which includes 10 partial NR sequences from cultured diatoms, indicated that the diatom NR sequences again form a distinct clade apart from algal and plant sequences, with significant bootstrap support (Fig. 4). Therefore, although the NRPt1000F and NRPt1389R primers target a short fragment of the highly conserved dimer interface region where sequences from diverse organisms share relatively high levels of homology (generally >50%), diatoms retain their group properties with even higher levels of identity (70%–80%).

Detection of NR sequences from diatoms and other eukaryotes in seawater: Eight of 10 clones screened from the LEO-15 15 m and Monterey Bay April libraries each contained Euk-NR gene fragments. Most of the clones screened from the LEO-15 15 m (5 of 5) and Monterey Bay October (4 of 5) libraries contained Euk-NR gene fragments. The cloned fragments that were not Euk-NR gene fragments were not similar to one another and are probably the result of PCR artifacts associated with highly degenerate primers. Many (18 of 25) of the Euk-NR sequences retrieved from Monterey Bay and LEO-15 cluster with a high level of confidence with known diatom NR sequences. Seven of the 25 NR sequences fall outside of the diatom clade and are associated with eukaryotic organisms without particular affinity to any sequenced representatives.

The four LEO-15 15-m sequences that form a tight group with bootstrap support of 100 are approximately 95% identical to one another and only 45% identical to diatom and other known NR sequences. Three other sequences retrieved from the LEO-15 samples fall outside the diatom cluster; the Leo212(15m)2 and Leo212(1m)1 sequences form a distinct clade that is 5% more similar to vascular plants than to any other group, and the Leo212(1m)2 sequence is deeply branching and does not cluster with any of the existing NR sequences. Each of these unknown marine environmental representatives share conserved amino acid motifs with the known sequences, but each of them also contains unique and, in the case of the well-supported cluster of four LEO-15 sequences, well-conserved motifs. The challenge still remains to uncover the biochemical and phenotypic significance of such unique motifs in otherwise very conserved regions of the NR gene.

**Phylogenetic comparisons and ecological applications.** In the past decade, research on Euk-NR has
received substantial attention from plant scientists, biochemists, and ecologists working across many taxa (Berges 1997, Campbell 2001). Exciting new findings and techniques include the observation in plants that NR transcription is controlled in part by photosynthetic electron flow and mechanistically linked to CO₂ reduction (Sherameti et al. 2002) and the development of an immunolabeling cytometric method to quantify NR protein abundance in diatoms (Jochem et al. 2000). Also, in diatoms, fundamental experiments on the effect of different environmental regulators such as nitrogen source, light, and temperature on protein activity and abundance have shown that NR activity is highly influenced by N source, has a diel cycle, is correlated with internal carbon pools, and is likely the rate-limiting step of N incorporation (Berges and Harrison 1995, Vergara et al. 1998, Lomas and Gilbert 1999, Lomas 2004). Vergara et al. (1998) developed a model to account for their measurements of NR protein activity and abundance and concluded that diel variations in NR activity are controlled primarily by transcriptional regulation.

Despite these significant increases in understanding NR activity dynamics, basic studies at the transcriptional level for biogeochemically important marine phytoplankton such as diatoms have not been possible because of the absence of sequence data. Although the T. pseudonana genome has been completely sequenced, it is important to examine a range of sequences from many diatoms to conduct ecologically meaningful experiments. New techniques based on microarray

**FIG. 4.** Inferred phylogenetic relationships of the approximately 130 amino acid eukaryotic assimilatory nitrate reductase gene fragments amplified with the primers NRP1000F and NRP1389R. Bootstrap values greater than 500 (of 1000) are shown at the nodes. The tree is rooted through the homologous region of the Arabidopsis thaliana sulfite oxidase gene.
technology and real-time PCR rely on sequence data from a diversity of environmentally important taxa.

Eukaryotic NR is present in vascular plants; red, green, and brown algae; all the lineages of eukaryotic phytoplankton; and in some heterotrophic organisms such as fungi and nonphotosynthetic stramenopiles. Eukaryotic NR does not have a prokaryotic homolog, and cyanobacteria have a different form of NR. Eukaryotic assimilatory NR therefore is thought to have originated from a single evolutionary event early in the history of eukaryotes, before the symbiotic origin of chloroplasts (Stolz and Basu 2002). Diatoms were derived by secondary endosymbiosis whereby a nonphotosynthetic eukaryote acquired a chloroplast by engulfing a photosynthetic eukaryote, probably a red algal endosymbiont (Armbrust et al. 2004). Phylogenetically, diatom NR does not have any particular affinity to the red algae *Cyanidioschyzon merolae*, relative to green algae and vascular plants (Figs. 2–4). It appears therefore that in secondary endosymbiotic lineages, such as the diatom, NR is also derived from the heterotrophic host. Eukaryotic NR sequences tend to be relatively well conserved and consistent with rRNA-defined taxonomic lines. As a group, for example, diatoms form a deep coherent branch distinct from algae and plants (Fig. 2). At this point it is unclear whether or not this clade includes chromophytes other than diatoms.

The four nested PCR primers that generated these 389-bp PCR products were intentionally designed to amplify a diverse range of phytoplankton Euk-NR gene fragments. At the time of primer design, the only photosynthetic taxa for which Euk-NR sequences were available were green algae, plants, and diatoms. Therefore, it is likely that our PCR primers are not capable of amplifying NR gene fragments from some groups of eukaryotic phytoplankton. Nevertheless, phylogenetic analysis of the sequences obtained from the seawater samples indicates that we retrieved some sequences of unknown identity that are in fact NR and most likely do not belong to any of the major taxa used for primer selection (Fig. 4). Degenerate-nested PCR facilitated the amplification of sequences from highly divergent taxa, which would not have been possible with other methods.

Although we did not sequence enough clones from any one library to draw any robust conclusions regarding the distribution of various type of NR genes, it is interesting to note that all the Monterey Bay sequences retrieved belong to diatoms. This is consistent with the ecology of Monterey Bay, where diatoms dominate production and NO$_3^-$ assimilation (Kudela and Dugdale 2000). At the LEO-15 site, six of eight sequences retrieved from 1 m belonged to the diatom group, whereas all five of the sequences retrieved from 15 m grouped outside of the diatoms, suggesting the presence of distinct communities at different depths. This pattern probably reflects differences in diversity and community composition between the depths sampled at the LEO site and Monterey Bay site, which could be described more quantitatively with a larger study of community composition at the NR gene level. Assays designed to detect Euk-NR mRNA in various phytoplankton lineages, therefore, will be useful for sampling the nutrient status and activity of specific groups within phytoplankton communities. Delineating particular conserved and variable regions of the Euk-NR gene in important marine phytoplankton such as diatoms is the first step toward introducing such powerful methods to aquatic microbial ecology.

The alignment in Figure 3 illustrates the different levels of conservation that occur between regions of the NR gene that are thought to have evolved independently and to be under differing levels of selection (Zhou and Kleinhofs 1996, Campbell 1999). In the poorly conserved hinge 1 region, which does not contain detectable similarity between diatoms and other taxa, a diatom signature is retained. This region could be useful for designing diatom-specific PCR primers or probes targeted to particular groups within the diatoms.

Based on data collected in this study, it is clear that diatom NR genes can be targeted as a coherent group at the DNA and RNA level. It should be noted that the primers designed in this study for the amplification of diatom NR genes do not amplify heterotrophic bacterial or eukaryotic dissimilatory NR genes. It is recognized that some groups of heterotrophic marine bacteria do assimilate nitrate aerobically using an NR enzyme with molybdopterin and heme cofactors (Allen et al. 2001), but such bacterial NR genes (*nasA*) are not homologous to eukaryotic NR genes. Eukaryotic dissimilatory genes (*Zvyagil’skaya et al. 1996*) are also not homologous to eukaryotic assimilatory NR genes. This is fortunate because it facilitates the investigation of the response of these ecologically distinct groups at the transcriptional level to environmental variables and will provide insight into community successional patterns in changing nutrient regimes, such as upwelling events. An important priority for the future is developing NR sequence databases that include other groups of ecologically important marine phytoplankton, such as dinoflagellates, haptophytes, and picoeukaryotes.

We thank Alan Milligan for assistance in maintaining the diatom cultures and Lee Kerkhoff for help in collecting the LEO-15 samples. Comments by John Berges helped to improve the quality of the manuscript. This research was supported by grants to B. B. W. from the U.S. Department of Energy and the National Science Foundation.


