Note

Algae 2010, 25(2): 65-70
DOI: 10.4490/algae.2010.25.2.065
Open Access

Nuclear rDNA characteristics for DNA taxonomy of the centric diatom *Chaetoceros* (Bacillariophyceae)

Hye-Young Oh[1], Ju-Yong Cheon[1], Jin Hwan Lee[1], Sung Bum Hur[2] and Jang-Seu Ki[1,*]

[1]*Department of Green Life Science, Sangmyung University, Seoul 110-743, Korea*
[2]*Korea Marine Microalgae Culture Center, Department of Aquaculture, Pukyong National University, Busan 608-737, Korea*

The genus *Chaetoceros* provides highly diversified diatoms in marine systems. Morphological descriptions of the genus are well-documented, yet the DNA taxonomy of *Chaetoceros* has not been satisfactorily established. Here, the molecular divergences of the 18S-28S rDNA of *Chaetoceros* were assessed. DNA similarities were relatively low in both 18S (93.1 ± 3.9%) and 28S rDNA (81.0 ± 4.6%). Phylogenies of the 18S, 28S rDNAs showed that *Chaetoceros* was divided according to individual species, clustering the same species into single clades. Statistical analysis with corrected genetic (*p*-) distance scores showed that nucleotide divergence of *Chaetoceros* 28S rDNA significantly differed from that of 18S rDNA (Student's *t*-test, $p < 0.05$). This finding suggests that the 28S rDNA may be treated as a more suitable marker for species-level taxonomic distinctions of *Chaetoceros*.

Key Words: *Chaetoceros*; DNA taxonomy; marine diatom; nucleotide divergence; ribosomal DNA

INTRODUCTION

Chaetoceros Ehrenberg, 1844, is the largest and most species-rich genus of marine planktonic diatoms (Rines and Hargraves 1988). To date, approximately 400 species of *Chaetoceros* have been morphologically described (Hasle and Syvertsen 1997). Some species are responsible for marine algal blooms. High concentrations of *Chaetoceros* cells may clog the gills of farmed fish and the spiny *Chaetoceros* setae can penetrate the gill tissue (Rensel 1993). These environmental and economically important effects have spurred many studies on *Chaetoceros*, which have improved the understanding of their biology, systematics, and ecology (Rensel 1993, Rines and Theriot 2003). In taxonomic and environmental monitoring purposes, discrimination of *Chaetoceros* species is generally achieved by microscopic observations, considering certain morphological characters such as forms of the chains, shapes of the aperture, and shapes of the valves. Particularly, the fine structures of the diatoms, including *Chaetoceros*, are observed with scanning electron microscopy. However, it is often very difficult to distinguish between *Chaetoceros* species (von Quillfeldt 2001) because of small size and their morphological similarity, and they can exhibit morphological changes under different culture conditions. In addition, morphological identification demands specialized in-depth knowledge.

DNA-based molecular tools are sometimes very effective for the species discriminations of microscopic-size organisms like diatoms (e.g., Jung et al. 2010). Recently, the concept of DNA barcoding was introduced to diatom taxonomy (Evans et al. 2007, Kaczmarska et al. 2007). The promise of DNA barcoding is based on a small DNA fragment divergence coinciding with biological species separation (Moniz and Kaczmarska 2009). Several pioneer studies on the diatom barcoding were performed with

Received 28 March 2010, **Accepted** 25 April 2010
***Corresponding Author**
E-mail: kijs@smu.ac.kr
Tel: +82-2-2287-5449 Fax: +82-2-2287-0700

several molecular markers such as nuclear ribosomal DNA (rDNA), chloroplast *rbcL*, and mitochondrial *cox1* gene (Evans et al. 2007, Moniz and Kaczmarska 2009, 2010). Also, many DNA-based studies have been done with regard to the evolutionary history and phylogenetic relationships of diatoms (Damsté et al. 2004, Alverson et al. 2007, Choi et al. 2008).

Of the molecular markers used in taxonomic studies, nuclear rDNA in eukaryotes is typically composed of tandem arrays of a basic unit that contain the transcription unit (e.g., 18S, 5.8S, 28S) and an intervening intergenic spacer region. The different subunits and regions of the rDNA locus have different degrees of sequence variability and varying suitability for comparison at the intergeneric or inter-species level. Recent data indicated that nuclear rDNA is a suitable molecular marker for DNA-based taxonomy or DNA barcoding of diatoms (Alverson et al. 2007, Evans et al. 2007, Kaczmarska et al. 2007, Moniz and Kaczmarska 2009, 2010, Jung et al. 2010). However, DNA-based discriminations should be carefully applied to the strongly diversified diatoms considering their molecular divergences of the rDNA, because they are variable according to different rDNA molecules and taxonomic categories (e.g., Jung et al. 2010). For example, the centric diatoms *Cyclotella* and *Discostella* show high divergences of both 18S and 28S rDNA (Jung et al. 2010), while their close relative diatom *Stephanodiscus* shows highly conserved 18S rDNA sequences within this genus, indicating the non-suitability of the 18S rDNA for their DNA taxonomy (Ki 2009). Taking this into account, it is necessary to evaluate genetic divergences of individual rDNA locus according to taxonomic categories, particularly at the generic level. In the case of *Chaetoceros*, although their morphological phylogenetic relationships have been studied (Rines and Theriot 2003), few studies on the molecular phylogeny have been attempted to date. Most studies have been carried out through broader diatom phylogenetic analyses (Damsté et al. 2004, Choi et al. 2008). Also, little is known of the genetic divergences of *Chaetoceros* rDNA for DNA taxonomy.

In the present study, we characterized molecular characteristics including genetic divergences and DNA similarity of the 18S-28S rDNA sequences from several selected *Chaetoceros*. In addition, phylogenetic and statistical analyses were performed to evaluate the usefulness of the 18S and 28S rDNA for the DNA taxonomy of *Chaetoceros*.

MATERIALS AND METHODS

Taxon samplings

In this study, a total of 32 rDNA sequences from *Chaetoceros* were used for the extensive analyses. The 18S rDNA sequences were determined from eight *Chaetoceros* species: *C. calcitrans* (GenBank accession numbers AY485449, AY625894, EU240879, EU240880), *C. curvisetus* (AY229895), *C. debilis* (AY229896), *C. gracilis* (AY625895), *C. muellerii* (AY485453, AY625896), *C. neogracile* (EU090012), *C. rostratus* (X85391), and *C. socialis* (AY485446). The partial 28S rDNA were from twelve *Chaetoceros* species: *C. atlanticus* (EF423454), *C. brevis* (EF423469), *C. compressus* (EF423429), *C. costatus* (EF423471-4), *C. curvisetus* (EF423476-7), *C. danicus* (EF423447), *C. debilis* (EF423466), *C. diadema* (EF423433), *C. lorenzianus* (EF423435-6), *C. peruvianus* (EF423449), *C. pseudo-curvisetus* (EF423478-9), and *C. socialis* (EF423467-8).

DNA sequence characteristics

Intra-specific genetic variations of *Chaetoceros* were investigated by comparing DNA similarities and genetic distances of both 18S and partial 28S rDNA sequences. For the extensive analyses, we constructed two data matrixes of the selected 18S and 28S rDNA sequences. These contained eight sequences for 18S and twelve sequences for 28S. Multiple alignments were performed with each dataset using the Clustal W 1.8 (Thompson et al. 1994). The aligned sequences were trimmed at each end to the same length and obvious base errors that were only found in single strands were manually removed. Finally, we used identical positions (e.g., 1,706 out of 1,815 alignment positions for 18S; 757 out of 800 for 28S) of the aligned sequences. DNA similarities of the 18S-28S rDNA were measured separately in BioEdit version 5.0.6 (North Carolina State University). The corrected pairwise (*p*-) genetic distances were calculated with Kimura 2-parameter model in MEGA 4.0 (Tamura et al. 2007). Sequence characteristics, including parsimony informative (PI) site, were analyzed using MEGA version 4.0. Statistical analyses of the nucleotide comparisons were performed using SPSS version 10.0.7 (SPSS Inc., Chicago, IL, USA).

Phylogenetic analysis of *Chaetoceros*

For the phylogenetic analysis of *Chaetoceros*, DNA sequences were aligned in the same way used in the sequence comparisons, and unambiguously aligned sequences for the phylogenetic analyses: 1,706 out of 1,806 alignment positions for 18S, and 652 out of 806 for 28S,

respectively. As best-fit models for the present 18S, 28S datasets, the General Time Reversible plus Gamma distributed model (GTR+G) was selected for 18S (- lnL = 5046.1) and for 28S (- lnL = 3933.2) from the Akaike Information Criterion in MrModeltest2 (Nylander 2004). Bayesian analysis of the 18S rDNA was implemented in MrBayes version 3.1.2 (Huelsenbeck and Ronquist 2001) using the selected GTR+G model with among-site rate variation, while the rates for variable sites were drawn from a gamma distribution. The Markov chain Monte Carlo process was set at two chains, and a million generations were conducted. Sampling frequency was assigned as every 100 generations. After analysis, the first 2,000 trees were deleted as burn-in and the consensus tree was constructed. For 28S rDNA tree, Bayesian analysis was performed in the same way using the 18S sequences.

RESULTS AND DISCUSSION

In this study, we characterized nuclear 18S and 28S rDNA sequences of *Chaetoceros* using available DNA sequences (12 sequences of 18S and 20 sequences of 28S) obtained in the public databases. These included nearly complete 18S rDNA sequences and partial 28S rDNA. Particularly, the 28S rDNA, the largest rDNA coding region, contains relatively conserved core segments and 12 hypervariable, divergent (D) domains (Hassouna et al. 1984). The present 28S data contained 28S rDNA D1 to D3 and their adjacent partial core regions.

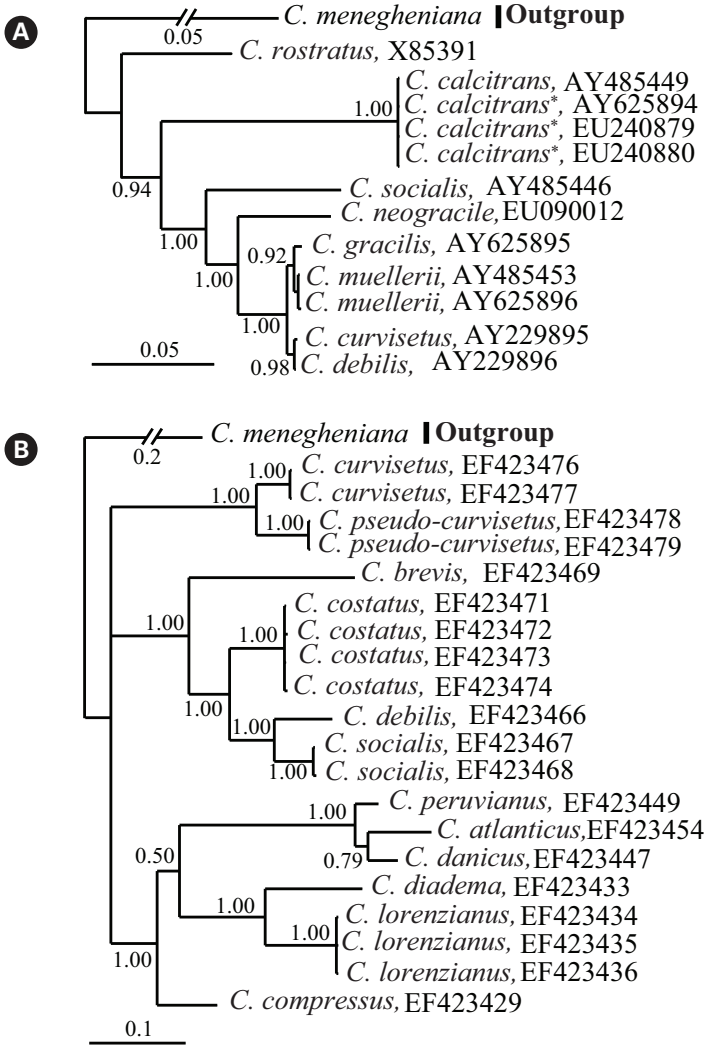

Fig. 1. Phylogenetic relationships of *Chaetoceros* inferred from (A) nearly complete 18S rDNA and (B) partial 28S rDNA sequences with Bayesian algorithms. Bayesian likelihood scores were recorded at –lnL = 5067.4 in 18S tree and at –lnL = 3970.5 in 28S tree, respectively. The numbers at each node represent posterior probability (> 0.50). **Chaetoceros. calcitrans* f. *pumilus*.

Genetic variations in the rDNA of *Chaetoceros* intra-species were investigated with the DNA similarity scores. Mostly, high DNA similarities were measured from individual 18S and partial 28S comparisons within the same species (more than 99% similarity). For example, *C. calcitrans*, including *C. calcitrans* f. *pumilus*, had nearly identical genotypes of the 18S rDNA (99.9 ± 0.1% similarity) among four different isolates, and *C. muellerii* showed 99.8% similarity between CCMP 1316 (GenBank accession number AY485453) and CCAP 1010/3 (AY625896), respectively. Also, we detected high DNA similarity in comparisons of the intra-species 28S rDNA. At present, we detected few genetic variations in the rDNA of intra-species; however, the present data are quite limited and so generalization should not be done. Further studies are needed to determine the nucleotide sequences of the rDNA of increased number of samples, collected worldwide from different geographical regions.

Bayesian trees with the 18S-28S rDNAs showed *Chaetoceros* spp. studied here were divided according to their taxonomic positions (Fig. 1). In cases of the same species, they formed single clusters (e.g., *C. calcitrans, C. costatus, C. muellerii, C. lorenzianus*, and *C. socialis*), which were separate from other species. These were in accordance with the intra-species rDNA comparisons, in which *Chaetoceros* has the different genotypes of the 18S-

Table 1. Similarity scores (above diagonal) and genetic distances (below diagonal) between nine pairs of the aligned sequence data (1,734 sites) of the nearly complete 18S rDNA of *Chaetoceros*

	Species	GenBank accession no.	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
			[Similarity]							
[1]	*C. calcitrans*	AY485449		89.1	89.0	89.0	89.1	88.7	88.4	87.9
[2]	*C. curvisetus*	AY229895	10.7		99.8	99.3	99.3	95.1	91.6	93.0
[3]	*C. debilis*	AY229896	10.7	0.1		99.3	99.3	95.0	91.6	93.0
[4]	*C. gracilis*	AY625895	10.7	0.6	0.6		99.5	94.8	91.6	92.8
[5]	*C. muellerii*	AY625896	10.6	0.6	0.6	0.4		95.0	91.8	92.8
[6]	*C. neogracile*	EU090012	11.2	4.7	4.8	5.0	4.8		90.8	92.1
[7]	*C. rostratus*	X85391	11.2	8.1	8.1	8.1	7.9	8.8		90.5
[8]	*C. socialis*	AY485446	12.0	7.0	7.0	7.1	7.1	7.7	9.5	
			[% *p*-distance]							

Table 2. Similarity scores (above diagonal) and genetic distances (below diagonal) between 12 pairs of the aligned sequence data (758 sites) of partial 28S rDNA of *Chaetoceros*

	Species	GenBank accession no.	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
			[Similarity]											
[1]	*C. atlanticus*	EF423454		76.6	83.4	77.5	77.8	92.2	79.6	79.5	76.7	93.2	76.6	79.8
[2]	*C. brevis*	EF423469	22.1		81.1	85.8	77.7	76.8	82.7	77.5	73.0	78.5	77.3	84.2
[3]	*C. compressus*	EF423429	15.8	15.9		83.7	83.7	84.5	83.8	84.0	81.1	85.3	82.0	83.6
[4]	*C. costatus*	EF423471	21.0	11.7	12.6		80.9	79.0	88.6	80.4	76.7	80.7	79.6	89.5
[5]	*C. curvisetus*	EF423476	19.6	20.2	13.3	16.3		78.7	79.3	79.1	76.3	78.6	92.6	80.6
[6]	*C. danicus*	EF423447	5.3	20.1	13.7	17.6	19.6		78.7	81.7	78.4	94.4	77.0	79.6
[7]	*C. debilis*	EF423466	18.7	15.9	14.7	9.5	19.2	18.8		77.8	74.9	80.6	77.2	93.1
[8]	*C. diadema*	EF423433	17.8	19.7	12.8	16.6	18.6	15.9	20.5		84.0	81.3	78.5	79.1
[9]	*C. lorenzianus*	EF423434	18.2	21.2	12.3	16.5	18.4	16.2	19.4	9.1		78.8	76.7	77.0
[10]	*C. peruvianus*	EF423449	6.0	19.3	13.3	17.0	18.4	3.5	17.5	15.3	15.1		77.0	80.7
[11]	*C. pseudo-curvisetus*	EF423478	21.6	20.4	15.7	18.3	7.8	21.8	22.2	19.2	17.9	21.0		79.3
[12]	*C. socialis*	EF423467	18.4	14.3	13.7	8.1	17.0	17.2	6.1	18.3	16.3	17.4	19.3	
			[% *p*-distance]											

28S rDNA among inter-species, but have nearly identical genotypes of the rDNA among intra-species (e.g., *C. calcitrans, C. muellerii, C. costatus*). The 18S tree (Fig. 1A) showed that *C. rostratus* formed the early divergent species (1.00 posterior probability [PP]). *C. gracilis* and *C. muellerii* as sister species formed a clade with *C. curvisetus* and *C. debilis* (1.00 PP). The two later species were not separated by our 18S phylogenetic analysis. On the other hand, the 28S Bayesian tree (Fig. 1B) separated individual species more clearly with long-branches compared with the present 18S rDNA phylogeny. The 28S Bayesian tree showed that *Chaetoceros* formed a polytomy (1.00 PP), in which species were separated into three clades: one cluster contained with *C. curvisetus* and *C. pseudo-curvisetus,* another included *C. costatus, C. debilis*, and *C. socialis*, and the other included *C. atlanticus, C, diadema, C. brevis,* and *C. lorenzianus*.

Molecular comparisons and phylogenies showed that sequence variations in the 18S and 28S rDNA within intra-species were not significantly different (Student's t-test, $p > 0.05$). Thus, we selected different *Chaetoceros* (e.g., eight for 18S rDNA; twelve for 28S rDNA) to extensively compare one another. Table 1 summarizes the DNA similarity and corrected *p*-distance scores between the eight pairs of aligned 18S rDNA sequences. DNA pairs of *C. curvisetus, C. debilis, C. gracilis*, and *C. muellerii* were recorded at high DNA similarities (> 99%, or < 0.6% *p*-distance), indicating that they could not be separated by the 18S rDNA divergences; however the other pairs showed relatively low similarities (< 95%, or > 4.8% *p*-distance). On the other hand, the *Chaetoceros* 28S rDNA showed high genetic divergences in the present analysis. Table 2 displays the DNA similarity and *p*-distance scores among the 12 compared species. In most cases, DNA divergences were considerably high in the 28S rDNA (81.0 ± 4.6% similarity). The highest similarity (94.4%) was recorded between *C. danicus* and *C. peruvianus*, and the lowest (77.2%) was recorded between *C. devilis* and *C. pseudo-curvisetus*.

In addition, comparative analysis showed that corrected *p*-distances of the 18S and 28S rDNAs were 7.3% and 16.3%, respectively (Fig. 2), based on pairwise genetic distance scores (Tables 1 & 2). Statistical testing revealed that divergences of the 28S rDNA were significantly different compared to the 18S rDNA (Student's t-test, $p < 0.05$). In further analysis, we found that the 28S rDNA contained more PI sites (28.6%) than 18S rDNA (11.0%). The 28S variation was approximately 2.60-times higher than that of the 18S as judged from the % PI values, and it was also 2.23-times by *p*-distance in the present data sets. These statistical, parsimonious results showed that the 28S rDNA D1-D3 (> 3.5% *p*-distance, > 5.4% dissimilarity) had a much greater genetic divergence than the 18S rDNA (> 0.4% *p*-distance, > 0.5% dissimilarity). These results were generally in accordance with other centric diatoms, *Cyclotella, Discostella*, and *Stephanodiscus* (Ki 2009, Jung et al. 2010). These results suggest that the 28S rDNA may be treated as a more suitable marker for species-level taxonomic distinctions of *Chaetoceros*.

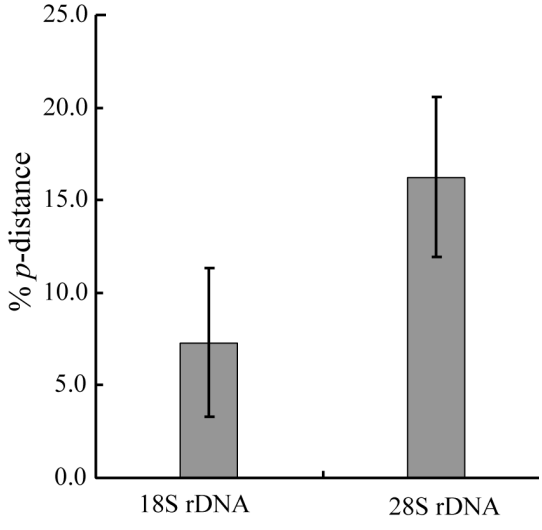

Fig. 2. Nucleotide divergences of *Chaetoceros* 18S and 28S rDNAs based on corrected *p*-distances. Values of the *p*-distances were measured at 7.3 ± 4.01 (n = 36) for 18S and at 16.2 ± 4.34 (n = 66), respectively.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2009-0084603).

REFERENCES

Alverson, A. J., Jansen, R. K. & Theriot E. C. 2007. Bridging the Rubicon: phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. Mol. Phylogenet. Evol. 45:193-210.

Choi, H. G., Joo, H. M., Jung, W., Hong, S. S., Kang, J. S. & Kang, S. H. 2008. Morphology and phylogenetic relationships of some psychrophilic polar diatoms (Bacillariophyta). Nova Hedwigia Beih. 133:7-30.

Damsté, J. S., Muyzer, G., Abbas, B., Rampen, S. W., Massé,

G., Allard, W. G., Belt, S. T., Robert, J. M., Rowland, S. J., Moldowan, J. M., Barbanti, S. M., Fago, F. J., Denisevich, P., Dahl, J., Trindade, L. A. & Schouten, S. 2004. The rise of the rhizosolenid diatoms. Science 304:584-587.

Evans, K. M., Wortley, A. H. & Mann, D. G. 2007. An assessment of potential diatom "barcode" genes (*cox1, rbcL*, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). Protist 158:349-364.

Hasle, G. R. & Syvertsen, E. E. 1997. Marine diatoms. *In* Tomas, C. (Ed.) *Identifying Marine Phytoplankton*. Academic Press, San Diego, pp. 5-385.

Hassouna, N., Michot, B. & Bachellerie, J. P. 1984. The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Res. 12:3563-3583.

Huelsenbeck, J. P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Jung, S. W., Han, M. S. & Ki, J. S. 2010. Molecular genetic divergence of the centric diatom *Cyclotella* and *Discostella* (Bacillariophyceae) revealed by nuclear ribosomal DNA comparisons. J. Appl. Phycol. 22:319-329.

Kaczmarska, I., Reid, C. & Móoniz, M. 2007. Diatom taxonomy: morphology, molecules and barcodes. *In* Kusber, W. H. & Jahn, R. (Eds.) Proc. 1st Central-European Diatom Meeting, Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Berlin, pp. 69-72.

Ki, J. S. 2009. Comparative molecular analysis of freshwater centric diatoms with particular emphasis on the nuclear ribosomal DNA of *Stephanodiscus* (Bacillariophyceae). Algae 24:129-138.

Moniz, M. B. J. & Kaczmarska, I. 2009. Barcoding diatoms: is there a good marker? Mol. Ecol. Resour. 9:65-74.

Moniz, M. B. J. & Kaczmarska, I. 2010. Barcoding of diatoms: nuclear encoded ITS revisited. Protist 161:7-34.

Nylander, J. A. A. 2004. MrModeltest v2. Evolutionary Biology Centre, Uppsala University, Uppsala.

Rensel, J. E. 1993. Severe blood hypoxia of Atlantic salmon (*Salmo salar*) exposed to the marine diatom *Chaetoceros concavicornis*. *In* Smayda, T. J. & Shimizu, Y. (Eds.) *Toxic Phytoplankton Blooms in the Sea*. Elsevier, New York, pp. 625-630.

Rines, J. E. B. & Hargraves, P. E. 1988. The *Chaetoceros* Ehrenberg (Bacillariophyceae) fiora of Narragansett Bay, Rhode island, U.S.A. Bibl. Phycol. 79:1-196.

Rines, J. E. B. & Theriot, E. C. 2003. Systematics of Chaetoceraotaceae (Bacillariophyceae). I. A phylogenetic analysis of the family. Phycol. Res. 51:83-98.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24:1596-1599.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 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.

von Quillfeldt, C. H. 2001. Identification of some easily confused common diatom species in Arctic spring blooms. Bot. Mar. 44:375-389.