Morphological and genetic diversity of Beaufort Sea diatoms with high contributions from the Chaetoceros neogracilis species complex

Sergio Balzano, Isabella Percopo, Raffaele Siano, Priscillia Gourvil, Mélanie Chanoine, Dominique Marie, Daniel Vaulot, Diana Sarno

Morphological and genetic diversity of Beaufort Sea diatoms with high contributions from the *Chaetoceros neogracilis* species complex

SERGIO BALZANO[1*]

SORBONNE UNIVERSITÉS, UPMC UNIV PARIS 06, CNRS, UMR7144, STATION BIOLOGIQUE DE ROSCOFF, 29680 ROSCOFF, France

ISABELLA PERCOPO

INTEGRATIVE MARINE ECOLOGY DEPARTMENT, STAZIONE ZOOLOGICA ANTON DOHRN, VILLA COMUNALE, 80121 NAPLES, ITALY

RAFFAELE SIANO

IFREMER, DYNECO PELAGOS, BP 70 29280, PLOUZANE, FRANCE.

PRISCILLIA GOURVIL, MÉLANIE CHANOINE, DOMINIQUE MARIE, DANIEL VAULOT

SORBONNE UNIVERSITÉS, UPMC UNIV PARIS 06, CNRS, UMR7144, STATION BIOLOGIQUE DE ROSCOFF, 29680 ROSCOFF, FRANCE

DIANA SARNO[2]

INTEGRATIVE MARINE ECOLOGY DEPARTMENT, STAZIONE ZOOLOGICA ANTON DOHRN, VILLA COMUNALE, 80121 NAPLES, ITALY

[*] PRESENT ADDRESS: NIOZ ROYAL NETHERLANDS INSTITUTE FOR SEA RESEARCH, DEPARTMENT OF MARINE MICROBIOLOGY AND BIOGEOCHEMISTRY, P.O. BOX 59, 1790 AB DEN BURG, TEXEL, THE NETHERLANDS

[1] CORRESPONDING AUTHOR: SERGIO.BALZANO@NIOZ.NL

Editorial Responsibility: M. Wood (Associate Editor)

Abstract

Seventy-five diatoms strains isolated from the Beaufort Sea (Canadian Arctic) in the summer of 2009 were characterized by light and electron microscopy (SEM and TEM) as well as 18S and 28S rRNA gene sequencing. These strains group into 20 genotypes and 17 morphotypes and are affiliated with the genera *Arcocellulus*, *Attheya*, *Chaetoceros*, *Cylindrotheca*, *Eucampia*, *Nitzschia*, *Porosira*, *Pseudo-nitzschia*, *Shionodiscus*, *Thalassiosira*, *Synedropsis*. Most of the species have a distribution confined to the northern/polar area. *Chaetoceros neogracilis* and *Chaetoceros gelidus* were the most represented taxa. Strains of *C. neogracilis* were morphologically similar and shared identical 18S rRNA gene sequences, but belonged to four distinct genetic clades based on 28S rRNA, ITS-1 and ITS-2 phylogenies. Secondary structure prediction revealed that these four clades differ in hemi-compensatory base changes (HCBCs) in paired positions of the ITS-2, suggesting their inability to interbreed. Reproductively isolated *C. neogracilis* genotypes can thus co-occur in summer phytoplankton communities in the Beaufort Sea. *Chaetoceros neogracilis* generally occurred as single cells but can also form short colonies. It is phylogenetically distinct from an Antarctic species, erroneously identified in some previous studies as *C. neogracilis* but named here as *Chaetoceros* sp. This work provides taxonomically validated sequences for 20 Arctic diatom taxa, which will facilitate future metabarcoding studies on phytoplankton in this region.

Key index words: biogeography, ITS, ITS2 secondary structure, LSU, morphology, phylogeny, polar diatoms, SSU

Abbreviations: CCMP, National Centre for Marine Algae and Microbiota; DCM, Deep Chlorophyll Maximum; ITS, Internal Transcribed Spacer; ITS-1, first internal transcribed spacer; ITS-2, second internal transcribed spacer; RCC, Roscoff Culture Collection; T-RFLP, terminal-RFLP;

INTRODUCTION

Due to fluctuations in light, temperature, salinity and sea ice extent, Arctic phytoplankton undergo high seasonal variability in abundance and composition. Higher temperatures and longer daylight between March and September, lead to an increase in algal biomass and primary production (Sherr et al. 2003, Wang et al. 2005). Diatoms account for a high portion of Arctic phytoplankton, especially in coastal locations (Booth & Horner 1997, Lovejoy et al. 2002) and species belonging to the genera *Chaetoceros* Ehrenberg and *Thalassiosira* Cleve can dominate phytoplankton communities in different regions (Tuschling et al. 2000, Booth et al. 2002, Ratkova & Wassmann 2002).

The Beaufort Sea is a major component of the Arctic Ocean, and is highly influenced by the Mackenzie River, which plays a key role in disrupting the winter ice in early spring promoting primary production and phytoplankton blooms (Carmack & MacDonald 2002). In addition, periodic wind-driven upwelling events can bring nutrient rich waters up to the surface layer and promote phytoplankton growth (Pickart et al. 2013). Except during episodic upwelling events, the water column is highly stratified, the nutrient concentration in the upper layers is extremely low, leading to the prevalence of picoeukaryotes, mostly represented by the psychrophilic *Micromonas* Manton & Parke ecotype corresponding to the single genetic clade named "Arctic *Micromonas*" (Lovejoy et al. 2007, Balzano et al. 2012b), within the phytoplankton community. Diatoms tend to be more abundant near the coast (Hill et al. 2005), occasionally blooming in late spring (Hill et al. 2005, Sukhanova et al. 2009). The algal biomass and the contribution of diatoms to the phytoplankton community increase in summer (Hill et al. 2005) and diatoms bloom more frequently at the deep chlorophyll maximum (DCM; Sukhanova et al. 2009). Autumn communities include higher contributions of dinoflagellates, which can dominate the community along with diatoms (Brugel et al. 2009).

The MALINA oceanographic expedition sailed in July 2009 from the Pacific coast of Canada to the Beaufort Sea where an extensive multidisciplinary sampling effort was undertaken until mid-August. Pigment analyses (Coupel et al. 2015) and light microscopy techniques (http://malina.obs-vlfr.fr/data.html) confirmed previous findings on phytoplankton community composition and revealed that Prymnesiophyceae, Mamiellophyceae and Dinophyceae dominated offshore waters while diatoms accounted for most abundance and biomass on the Mackenzie Shelf (Coupel et al. 2015). Within diatoms the cold-water ecotype of *Chaetoceros socialis* described recently as *Chaetoceros gelidus* (Degerlund et al. 2012, Chamnansinp et al. 2013), several other *Chaetoceros* taxa, and with lower abundances, *Thalassiosira nordenskioeldii*, and *Pseudo-nitzschia* spp. prevailed (http://malina.obs-vlfr.fr/data.html). Molecular techniques [cloning/sequencing and terminal-RFLP (T-RFLP) on the 18S rRNA gene] on photosynthetic populations (Balzano et al. 2012b) partially agree with pigment analyses and phytoplankton microscopy counts indicating that Arctic *Micromonas* (Lovejoy et al. 2007) was the only photosynthetic picoplankter (< 2 μm) detected in most stations, whereas nanoplankton (2-20 μm) genetic libraries were dominated by the diatoms *C. gelidus* (referred therein as *C. socialis*) and *Chaetoceros neogracilis* in DCM and surface waters, respectively (Balzano et al. 2012b).

Seasonal succession and geographic distribution of phytoplankton species have thus been partially elucidated for the Beaufort Sea, but species level diversity has still not been fully assessed for diatoms, due to the limited resolution power of the morphological and molecular methods employed. Light microscopy, that has been applied in most the studies, does not allow the observation of the fine ultrastructural details often required to distinguish diatom species. Similarly, the 18S rRNA gene did not allow discrimination among some species of the genera *Chaetoceros* and *Pseudo-nitzschia* H. Peragallo, which were well represented in the area (Balzano et al. 2012b). Other ribosomal genes have a higher resolution power: the 28S rRNA gene can successfully discriminate most of the species within the genera *Chaetoceros* (Kooistra et al. 2010) and *Pseudo-nitzschia* (Lundholm et al. 2002) and is considered a good discriminatory molecular marker among

centric diatom species (Lee et al. 2013). A gene fragment extending from the 5' end of the 5.8S to the 3' end of the helix III of ITS-2 (5.8S + ITS-2) has been proved to separate the 99.5 % of diatom species (Moniz & Kaczmarska 2010).

Coupling culture isolation with morphological and genetic characterization allows detailed species identification. This approach has been applied to photosynthetic flagellates collected during the MALINA cruise. Photosynthetic pico- and nanoeukaryotic populations were dominated by cultured microorganisms (Balzano et al. 2012b) and 104 strains belonging to the Chlorophyta, Dinophyta, Haptophyta, Cryptophyta and Heterokontophyta divisions were isolated and characterized by both light microscopy (LM) and 18S rRNA gene sequencing (Balzano et al. 2012a).

A recent study investigated Arctic dinoflagellates coupling morphological and genetic approaches (Gu et al. 2013), but similar information on diatoms is missing. In the present paper, we focus on diatom strains isolated from the Beaufort Sea. We combined LM, TEM, and SEM with 18S and 28S rRNA gene sequencing to identify the isolated strains. We also sequenced the ITS operon of the rRNA gene from a number of *C. neogracilis* strains sharing highly similar 18S and 28S rRNA gene sequences to further investigate the occurrence of distinct genetic entities and we reconstructed the secondary structure of the ITS-2 of these strains in order to predict their reproductive isolation.

MATERIALS AND METHODS

Phytoplankton sampling, isolation and maintenance. Strains were isolated from seawater samples collected during the MALINA (http://www.obs-vlfr.fr/Malina) cruise which sailed the 06/07/09 from Victoria (British Columbia, Canada) to the Beaufort Sea where an extensive sampling effort was carried out in late summer from 1/08/09 to 24/08/09 (Table S1 in the Supporting Information). Samples were collected with a bucket from surface waters in the North Pacific and at different depths with Niskin bottles mounted on a CTD frame in the Beaufort Sea. Phytoplankton strains were isolated both onboard and back in the laboratory (Table 1) as described previously (Le Gall et al. 2008, Balzano et al. 2012a). Overall we isolated 75 diatom strains, 60 of which are currently (March 2016) available from the Roscoff Culture Collection (RCC: http://www.roscoff-culture-collection.org/). Most of the strains were isolated from the Beaufort Sea but we also included four strains from the North Pacific sampled during the first leg of the MALINA cruise for comparison purposes. The strains were maintained in K or K/2-medium (Keller et al. 2009) with addition of silicate, prepared from sterile seawater at a salinity of 35 and kept at 4°C at an irradiance of 50 µmol photons $\cdot m^{-2} \cdot s^{-1}$ in a 12:12 light dark regime. Some of the *C. neogracilis* strains were incubated at low light intensity (about 10 µmol photons $\cdot m^{-2} \cdot s^{-1}$) in f/2 medium (Guillard 1975) with nitrate supplied at a concentration 10-fold lower (88 µM) to induce resting spore formation, since spore morphology can help species identification in the genus *Chaetoceros* (Hasle & Syvertsen 1997).

DNA extraction and PCR. Genomic DNA was extracted from 75 MALINA strains as described previously (Balzano et al. 2012a) using the NucleoSpin Tissue kit (Mackerey Nagel, Hoerdt, France) and following the instructions provided by the manufacturer.

The 18S rRNA gene, the Internal Transcribed Spacer (ITS) of the rRNA operon and the 28S rRNA gene were then amplified by PCR on genomic DNA. For the 18S rRNA gene the primers 63f

(5'-ACGCTT-GTC-TCA-AAG-ATTA-3') and 1818r (5'-ACG-GAAACC-TTG-TTA-CGA-3') were used (Lepère et al. 2011) as described previously (Balzano et al. 2012a).

The ITS region of the rRNA operon was amplified from 35 MALINA strains of *C. neogracilis* (Table 1) and 3 Antarctic strains of *Chaetoceros* purchased from the National Centre for Marine Algae and Microbiota (Bigelow, USA) and previously thought to belong to *C. neogracilis* (CCMP187, CCMP189, and CCMP190; Table S2 in the Supporting Information). The ITS was amplified using primers 329f (5'-GTG-AAC-CTG-CRG-AAG-GAT-CA-3') and D1R-R (5'- TAT-GCT-TAA-ATT-CAG-CGG-GT-3') which correspond to the reverse complements of the reverse primer for 18S 329r (Guillou et al. 2004) and the 28S forward primer D1R (Lenaers et al. 1989), respectively. PCR condition included an initial incubation step at 95°C during 5 min, 35 amplification cycles (95°C for 1 min, 55°C for 45 s, and 72°C for 1 min 15 s) and a final elongation step at 72 °C for 7 min. From 72 diatom strains, the 28S rRNA gene was amplified using primers D1R (5'-ACC-CGC-TGA-ATT-TAA-GCA-TA-3') and D3Ca (5'-ACG-AAC-GAT-TTG-CAC-GTC-AG-3') targeting the D1–D3 region of the nuclear LSU rRNA (Lenaers et al. 1989, Orsini et al. 2002). PCR reactions were as follows: 30 amplification cycles of 94°C for 1 min, 55°C for 1 min 30 s, and 72°C for 1 min.

18S rRNA, ITS, and 28S rRNA amplicons were purified using Exosap (USB products, Santa Clara, USA) and partial sequences were determined by using Big Dye Terminator V3.1 (Applied Biosystems, Foster city, USA). The hypervariable V4 region (Dunthorn et al. 2012) of the 18S rRNA gene was sequenced from all the strains using the internal primer Euk528f (Zhu et al. 2005), whereas the primers 63f and 1818R were used to sequence the full 18S rRNA gene from selected strains. The ITS region was sequenced using both forward and reverse primers described above whereas the forward primer D1R was used to sequence the 28S rRNA gene. Sequencing was carried out on an ABI prism 3100 sequencer (Applied Biosystems).

Phylogenetic analysis. V4 sequences were compared to those available in Genbank using BLAST (blast.ncbi.nlm.nih.gov/Blast.cgi), aligned using ClustalW2

(http://www.ebi.ac.uk/Tools/msa/clustalw2) and then grouped into 17 different 18S genotypes based on 99.5% sequences similarity, using the Bioedit software (Hall 1999). The full 18S rRNA gene was sequenced from at least one strain per genotype (19 strains in total). For all the phylogenetic trees shown in this paper, relationships were analyzed using Maximum Likelihood (ML) and Neighbour Joining (NJ) methods (Nei & Kumar 2000) and bootstrap values were estimated using 1,000 replicates (Felsenstein 1985) for both methods. MEGA5 software (Tamura et al. 2011) was used to construct the phylogenetic trees based on the ML topology.

Full 18S rRNA sequences were aligned with reference sequences from Genbank (http://www.ncbi.nlm.nih.gov/nucleotide, Table S2) for a total of 84 sequences using clustalW2 as described above. Highly variable regions of the alignment were removed and the final dataset contained 1,465 nucleotide positions. A Tamura Nei model (Tamura & Nei 1993) was selected as the best model to infer both NJ and ML 18S phylogeny.

For the D1-D3 region of the 28S rRNA gene 64 sequences were aligned using clustalw2 and a subset, containing at least one sequence per genotype, was used to construct three phylogenetic trees (centric diatoms, pennate diatoms and *C. neogracilis* strains). Highly variable regions were removed from the alignments. For the centric diatoms, the alignment included, 65 sequences and 504 positions and the phylogeny was inferred using a Kimura-2 model (Kimura 1980). Phylogenetic relationships were then inferred as described above and 5 sequences from the genus *Attheya* West were used as an outgroup and were then removed from the tree for clarity. For the pennate diatoms, the alignment included 35 sequences and 490 nucleotide positions and the phylogeny was inferred using a Tamura-Nei model (Tamura & Nei 1993) and sequences from the genus *Attheya* were also used as an outgroup. A third phylogenetic tree was constructed for *C. neogracilis*, which included 36 MALINA strains from this species, 1 sequence of the strain CPH9 identified as *Chaetoceros fallax* Prosckina Lavrenko, 3 GenBank sequences from the Antarctic strains CCMP163, CCMP189 and CCMP190 (Table S2) and one sequence from *C. gelidus*

(RCC2271) which was used as an outgroup. The analysis was performed on 41 sequences for a total of 590 positions using a Kimura-2-parameter model.

We also sequenced the ITS operon of the rRNA gene from the MALINA strains affiliated to *C. neogracilis* as well as the Antarctic strains attributed by CCMP to *C. neogracilis*. Since the 5.8S is a region highly conserved at interspecific level, we identified the boundary between ITS-1 and 5.8S based on 5.8S sequences from other *Chaetoceros* species (Moniz & Kaczmarska 2010) available in GenBank. We then constructed a phylogenetic tree based on the ITS-1 and another phylogenetic tree consisting in a region starting at the 5' end of 5.8S and ending in the conserved motif of helix III of ITS-2. Some sequences did not cover the entire ITS length and were excluded from the alignment of either the ITS-1 or the 5.8S/ITS-2. The ITS-1 alignment included 30 sequences and 227 nucleotide positions and was analysed using a Jukes Cantor model (Jukes & Cantor 1969). For the 5.8/ITS-2 alignment the end of helix III was annotated based on the secondary structure of the ITS-2 from *T. weissflogii* (Grunow) Fryxell & Hasle (Sorhannus et al. 2010), which is the species most closely related to the genus *Chaetoceros* for which the secondary structure of the ITS-2 has been reconstructed. The final alignment included 30 sequences and 384 nucleotide positions and both ML and NJ phylogenies were inferred using a Kimura-2 model (Kimura 1980). The ITS could not be sequenced from the strain MALINA E43.N2, but it was attributed to Clade II based on its 28S sequence. Similarly since both the ITS-1 and the 5.8 + ITS-2 sequences from RCC2268, RCC2277 and RCC2318 were not sufficiently long to be included in the ITS-1 and the 5.8S/ITS-2 alignments, a neighbour joining phylogenetic tree for the entire ITS fragment which included 34 sequences for a total 483 positions (Fig. S1 in the Supporting Information) was constructed in order to identify the genetic clade of these strains.

ITS-2 structure prediction. To characterize our MALINA strains of *C. neogracilis* in deeper detail we reconstructed the secondary structure of the ITS-2 operon of the rRNA. The ITS-2 boundaries were then annotated using Hidden Markov Models of the flanking 5.8S and 28S regions (Keller et al. 2009). The secondary structure of the ITS-2 was first inferred for the strain RCC2014

using the RNA structure program (Mathews et al. 2004) and then transferred onto other *Chaetoceros* sequences through homology modelling (Wolf et al. 2005) using the ITS-2 database (Merget et al. 2012).

Microscopy. At least one strain per genotype, for a total of 61 strains (Table 1), was observed and photographed in light microscopy. Cells were harvested during the exponential phase of their growth and observed using an Olympus BX51 microscope (Olympus, Hamburg, Germany) with a 100X objective using differential interference contrast (DIC). Cells were imaged with a SPOT RT-slider digital camera (Diagnostics Instruments, Sterling Heights, MI, USA). Micrographs are available at http://www.roscoff-culture-collection.org for a large set of strains.

Selected strains, covering most genetic diversity based on both 18S and 28S rRNA, were observed using Light Microscopy (36 strains), TEM (25 strains) and/or SEM (28 strains) at Stazione Zoologica Anton Dohrn (Table 1). To remove organic matter, samples were treated with nitric and sulfuric acids (1:1:4, sample:HNO_3:H_2SO_4), boiled for a few seconds and washed with distilled water. LM observations were performed using a Zeiss Axiophot 200 equipped with a Axiocam Digital Camera (Carl Zeiss, Oberkochen, Germany). Acid-cleaned material was mounted on Formvar- coated grids and observed with a LEO 912AB transmission electron microscope (LEO, Oberkochen, Germany) and/or mounted on stubs, sputter-coated with gold-palladium and observed with a JEOL JSM-6500F SEM (JEOL-USA Inc., Peabody, MA, USA). Fixed samples not subjected to cleaning were placed on Nuclepore 3 µm pore size (Nuclepore, Pleasanton, CA, USA) polycarbonate filters, rinsed with distilled water, dehydrated in an ethanol series (25, 50, 75, 95, and 100%), and critical-point-dried. Dried filters were mounted on stubs, sputter-coated and observed with SEM.

249 RESULTS

250

251 In the present study, we characterized 75 diatom strains using a combination of morphological
252 and molecular techniques (Table 1). We sequenced the V4 region of the 18S rRNA from all the
253 strains and then we sequenced the full 18S rRNA from at least one strain from each unique
254 genotype. Moreover we sequenced the 28S rRNA from most of our strains and the ITS operon of
255 the rRNA from all the strains affiliated to *C. neogracilis.* The strains grouped into 17 genotypes
256 based on 18S and 28S rRNA phylogenies (Figs. 1 and 2). 28S rRNA and ITS analyses indicate that
257 36 strains of *C. neogracilis* sharing identical 18S rRNA gene sequence make up 4 distinct genetic
258 clades (Figs. 2 and 3). The most represented genera were *Chaetoceros* and *Thalassiosira*.
259 ***Bacillariaceae.*** We isolated 9 Bacillariaceae strains from the genera *Cylindrotheca, Nitzschia*
260 and *Pseudo-nitzschia.* The 18S rRNA gene (Fig. 1) discriminated the different *Cylindrotheca* and
261 *Nitzschia* representatives but was poorly resolutive for the different *Pseudo-nitzschia* species.
262 *Cylindrotheca closterium* (Ehrenberg) Lewin & Reimann.
263 Cells are 85 to 108 µm long, fusiform with rostrated ends and possess two chloroplasts (Hasle
264 1964, Jahn & Kusber 2005). The valve face is unperforated, transversed by transapical slightly
265 silicified ribs. The central raphe is interrupted by a central nodule. The fibulae (13-17 in 10 µm) are
266 narrow, irregularly spaced, and joined directly to the valve face (Fig. 4A; Hasle 1964, Jahn &
267 Kusber 2005).
268 *Cylindrotheca closterium* was considered as a cosmopolitan species but it was demonstrated to
269 constitute a species complex of similar morphotypes belonging to different genetic lineages (Haitao
270 et al. 2007). It has been repeatedly observed in the Arctic (Table 2).The 18S rRNA gene sequence
271 from *C. closterium* strain RCC1985 (Fig. 1) groups with the other *C. closterium* sequences forming
272 a moderately-supported clade (sequence similarity > 97.8 %), but does not cluster to any of the two
273 lineages described to date for the *C. closterium* species complex (Haitao et al. 2007). The 28S

rRNA gene sequence from *C. closterium* strain RCC1985 (Fig. 2A) branches with two other sequences from *C. closterium.*

Nitzschia pellucida Grunow.

Cells (apical axis: 35 µm; transapical axis: 3.0-3.5 µm) are solitary and possess two chloroplasts. Cells are lanceolate, tapering towards the poles, in valve view (Fig. 4B), and rectangular when observed in girdle view. The densities of fibulae and striae are 12-15 and 35-40 in 10 µm, respectively. Each stria contains one row of rounded poroids. A central larger interspace is present (Fig. 4, C and D).

Nitzschia pellucida has been previously reported in Arctic and Antarctic waters but also in European freshwater environments (Table 2). The 18S rRNA gene sequence from *N. pellucida* strain RCC2276 is highly related to that of *Nitzschia dubiiformis* (99.6 % sequence identity) and branches with other *Nitzschia* species (Fig. 1). The 28S rRNA gene sequence from *N. pellucida* strain RCC2276 groups with *Nitzschia laevis* and *N. pellucida* from GenBank (sequence identity 97.5 and 97.4, respectively). This clade branches with different *Nitzschia* and *Cylindrotheca* species (Fig. 2A), which supports the assertion of Lundholm et al. (2002) describing the genus *Nitzschia* as polyphyletic.

Pseudo-nitzschia granii (Hasle) Hasle.

Cells (apical axis: 17-25 µm; transapical axis: 1.4-1.8 µm) have two chloroplasts and colonies were not observed in culture conditions. Valves are lanceolate with a central swelling, one side of the valves is linear and the other convex (Fig. 4E). Apices are rounded. The striae (54-55 in 10 µm) are composed of a single row of poroids divided in 5-7 sectors. In the strain RCC2006, most of the valves have striae barely silicified that lack complete poroids (Fig. 4F) or have few poroids entirely formed (Fig. 4G). The fibulae (16-18 in 10 µm) are irregularly spaced and the central interspace is absent.

Pseudo-nitzschia granii has been reported in northern cold waters, including Arctic and subarctic regions (Table 2).

Pseudo-nitzschia arctica Percopo & Sarno.

Four *Pseudo-nitzschia* strains isolated during the MALINA cruise have been recently described as a new species, *Pseudo-nitzschia arctica* (Percopo et al. 2016). Cells occur in colonies and each cell overlaps the next sibling cell for ca 1⁄8 of its length (Fig. 4H). Cells (apical axis: 26-60 µm; transapical axis: 1.6-2.5 µm) are lanceolate in valve view. The valve ends are broadly pointed. The fibulae are not always regularly spaced. The two central fibulae have a larger interspace and the raphe is here interrupted by a central nodule (Fig. 4I). The densities of fibulae and interstriae are 17-24 and 34-39 in 10 µm, respectively. The striae contain 1 row of rounded poroids, 5-6 poroids in 1 µm. Each poroid most often contains 1-6 sectors. Some striae are simply composed of more lightly silicified areas without any perforations.

Pseudo-nitzschia arctica seems to have a distribution confined to the northern polar area, possibly representing one of the endemic components of the Arctic diatom flora (Percopo et al. 2016).

Pseudo-nitzschia arctica and *P. granii* share highly similar 18S rRNA gene sequences (99.6 % sequence identity, Fig. 1) and the two species can be better separated using 28S rRNA phylogeny (Fig. 2A) where their sequences differ by 1.2 % sequence identity.

Fragilariaceae.

Synedropsis hyperborea (Grunow) Hasle, Medlin & Syvertsen.

Cells (apical axis: about 55 µm; transapical axis: 2.7-3.5 µm) are lanceolate in valve view (Fig. 4J). No colonies were observed. The uniseriate striae (22-23 in 10 µm) are parallel towards the apices and alternate in the some parts of the valve (Fig. 4K). The apical fields are composed of 5-7 slits (Fig. 4, L-N) slightly different from that reported in the original description (4-6 slits, Hasle et al. 1994). A single rimoportula is located two or three striae from one of the two valve apices (Fig. 4L). The rimoportula opens externally into a hole larger than the surrounding areolae (Fig. 4, M and N).

325 *Synedropsis hyperborea* is typical of the Northern cold region and it is commonly reported in
326 Arctic waters (Table 2).
327 Fragilariaceae taxonomy was not well resolved based on 18S rRNA gene since the sequence
328 from *S. hyperborea* strain RCC2043 shares very high similarities with a sequence from Genbank
329 affiliated to *S. hyperborea* (99.9 %) as well as *Synedra minuscula* (99.9 %), *Fragilaria* sp. (99.9 %)
330 and *Grammonema striatula* (99.5 %, Fig. 1). MALINA strains RCC2043 and RCC2520 belonging
331 to *S. hyperborea* share identical 28S rRNA gene sequences and group with a sequence from
332 *Synedropsis hyperboreoides* from GenBank (98.5 % sequence identity). The 28S rRNA gene
333 sequences from these strains are also related to *Thalassionema frauenfeldii* and 3 *Fragilaria* species
334 (Fig. 2A).
335 *Attheyaceae*. *Attheya septentrionalis* (Østrup) Crawford.
336 Cells (apical axis: 3.5-6.4 µm; pervalvar axis: 7-11.7) are solitary and bear four slightly wavy
337 horn-like projections (Fig. 5, A and B). One or two plate-like chloroplasts are present. Valves are
338 almost circular and lack the rimoportula (Fig. 5C). The length of the horns is variable (12-35 µm)
339 and the ratio between horn length and cell diameter ranges from 2.9 to 4.4. The number of
340 longitudinal strips can be 3 or 4 in both examined strains (Fig. 5, D and E).
341 *Attheya septentrionalis* is distributed in the northern cold region and it is common in Arctic
342 waters (Table 2).The 18S rRNA gene from the MALINA strains RCC1986 and RCC2042 branches
343 with that of sequences of *A. septentrionalis* (99.9 % sequence identity) and *Attheya longicornis*
344 (99.8 %) from GenBank and is related to sequences from three *Biddulphia* spp. (Fig. 1). The 28S
345 rRNA gene from the two MALINA strains is also highly related to that of sequences of *A.*
346 *septentrionalis* and *A. longicornis* (≈ 98 % sequence identity, Fig. 2A).
347 *Thalassiosiraceae.* We isolated 12 Thalassiosiraceae strains (Table 1) affiliated to the genera
348 *Thalassiosira, Porosira,* and *Shionodiscus*. Both 18S and 28S rRNA gene allowed the
349 discrimination of the different species found here (Figs. 1 and 2B).
350 *Thalassiosira gravida* Cleve.

Cylindrical cells (diameter: 28.5-30.5 µm) held in colonies by a single thick thread composed of several strands (Fig. 5F). A number of fultoportulae (or strutted processes, 11-15) are grouped in a central cluster and several fultoportulae are scattered on the valve face. The marginal fultoportulae are arranged to form 3-4 rings placed between the margin of the valve face and the mantle. A single rimoportula (or labiate process) is located within the inner ring of marginal fultoportulae (Fig. 5G). Different valves have a variable degree of silicification, but in general the areolae are well-formed on the margins of the valve (16-20 in 10 µm) and poorly developed in the central part, where siliceous radial ribs separate perforated areas.

Thalassiosira gravida is regarded as a bipolar, cold to temperate water species and it has been previously observed in Arctic and Antarctic waters (Table 2).

The 18S rRNA gene sequence from *T. gravida* strain RCC1984 clusters with sequences from both *T. gravida* and *T. rotula* (sequence identity > 99.5 %) and is highly related with a sequence from *Thalassiosira eccentrica* (99.3 % sequence identity, Fig. 1). The 28S rRNA gene sequences from both our strains of *T. gravida* group with two other sequences from *T. gravida* and are highly related to 2 sequences from *T. rotula* (99.2 % sequence identity, Fig. 2B).

Thalassiosira cf. *hispida* Syvertsen.

Cells (diameter: 6.5-13.5 µm) possess several chloroplasts, and form colonies of few cells (3-4 cells) connected by one central thread. The areolae (30 in 10 µm) have a similar size on both valve face and mantle. One ring of marginal fultoportulae (5 in 10 µm) and one central fultoportula are present on the valve face (Fig. 5H). The marginal fultoportulae have long external tubes (Fig. 5, H-K). All the fultoportulae have four satellite pores at their base (Fig. 5I). The rimoportula is positioned slightly inside the ring of marginal fultoportulae, between two of them. It can be either closer to one of them or in the middle. A broad hyaline margin is present. Short and minute spines and hairs emerge throughout the valve (Fig. 5I). The girdle is formed by a valvocopula, a copula and several open bands. The valvocopula has a broad abvalvar imperforated rim and one advalvar row of areolae (Fig. 5J). MALINA strain of *T.* cf. *hispida* is morphologically very similar to the

original description of *T. hispida* but possesses a higher number of areolae (18 and 24-26 in 10 µm on valve face and mantle, respectively, in Syvertsen 1986). Very similar is the dense covering of spinules on the valve surface, which however is not specific for *T. hispida*, but can be developed to a lesser extent in other *Thalassiosira* species, and the presence of a broad hyaline margin on the valve and a valvocopula with a wide non-pierced edge.

Thalassiosira hispida has only been reported in northern cold water regions (Table 2). 18S rRNA gene sequences from *T. hispida* are not available on the GenBank and the 18S rRNA gene sequence from our strain RCC2521 clusters with sequences of *Thalassiosira allenii* (98.5 % sequence identity) and *Thalassiosira angulata* (98.6 %, Fig. 1). The 28S rRNA gene sequence from RCC2521 (Fig. 2B) groups with *T. allenii* (97.5 %), *Thalassiosira aestivalis* (97.1 %) and *T. nordenskioeldii* (96.5 %) but the clade is poorly (< 50 % ML and NJ) supported.

Thalassiosira minima Gaarder.

Cells (diameter: 4.5-13µm) have two chloroplasts and do not form colonies under our culture conditions. In girdle view, cells are rectangular with a pervalvar axis generally shorter than the cell diameter and with a valve face slightly depressed in the centre (Fig. 5L). The areolae (30-35 in 10 µm) are hexagonal in shape (Fig. 5M). On the valve, a ring of marginal fultoportulae (4-6 in 10 µm) with short external tubes and one or two central fultoportulae are present (Fig. 5, M and N). Five fultoportulae have been occasionally observed in one single valve (Fig. 5O). A large rimoportula is placed between two marginal fultoportulae, slightly closer to one of them (Fig. 5, M and N). Each marginal fultoportula is accompanied with a small external labiate-shaped protrusion (Fig. 5P). The species has a worldwide distribution (Table 2) and it is reported for the first time in the Arctic Ocean.

The 18S rRNA gene sequence from our *T. minima* strain RCC2265 is highly similar to that of the *T. minima* sequence from the strain CCMP990 (99.7 %, Fig. 1). Our strains of *T. minima* from both the Beaufort Sea and the North Pacific Ocean (Table 1) share highly similar 28S rRNA gene sequences with the Antarctic strain RCC2707 (99.1 %) and group with the *T. minima* strain

CCMP990 forming a well-supported clade (Fig. 2B). Consistent with the 18S rRNA gene phylogeny, *Thalassiosira curviseriata* is the species most closely related to all the *T. minima* strains.

Thalassiosira nordenskioeldii Cleve.

Cells (diameter: 12-15 µm) possess several chloroplasts and form long colonies connected by a central thread (Fig. 6A). Areolae are 17-18 on valve face and 18-20 in 10 µm on mantle (Fig. 6B). Valves are characterized by a pronounced concavity in the centre, a high (4-6 areolae) and oblique mantle, a marginal ring of fultoportulae (3-4 in 10 µm) with long external tubes bearing a terminal collar, one central fultoportula and one rimoportula positioned within two marginal fultoportulae (Fig. 6, B and C).

Thalassiosira nordenskioeldii is a species typical of northern cold to temperate regions, common in Arctic waters (Table 2).

The 18S rRNA gene sequence from strain RCC2000 groups with sequences from *T. aestivalis* and *T. nordenskioeldii* forming a well-supported clade (Fig. 1). *Thalassiosira nordenskioeldii* RCC2000 shares identical 28S rRNA gene sequence with another *T. nordenskioeldii* strain from the GenBank and highly similar 28S sequence (99.8 %) with *T. nordenskioeldii* RCC2021. These strains form a clade with a sequence from *T. aestivalis* (Fig. 2B).

Porosira glacialis (Grunow) Jørgensen.

Cells (diameter: 30-40 µm) are cylindrical, possess several chloroplasts and can form short colonies (2-3 cells; Fig. 6, D and E). Numerous fultoportulae are scattered over the valve surface (3-4 in 10 µm). The striae (24-27 areolae in 10 µm) are wavy and radially arranged. A central annulus is present and a large rimoportula process is situated inside the margin of the valve (Fig. 6F).

Porosira glacialis is reported in Arctic and Antarctic waters (Table 2).

RCC2039 18S rRNA is identical with that from the Antarctic strain CCMP1099 (Fig. 1). The 28S rRNA gene sequence from the MALINA strain RCC2039 is highly related, but not identical (99.6 %), to that of the two Antarctic strains CCMP1099 and RCC2709 (Fig. 2B).

Shionodiscus bioculatus (Grunow) Alverson, Kang & Theriot.

Cells (diameter: 23-41 µm) are solitary and possess a large number of discoid chloroplasts (Fig. 6G). The pervalvar axis is generally longer than the diameter. The valve face is slightly convex and the mantle is rounded. The areolation is fasciculate (20-23 areolae in 10 µm) with a single fultoportula in the valve centre and a subcentral rimoportula (Figs. 6, H and I). The marginal fultoportulae (4-7 µm apart) have internal tube-like projections and no external extensions. Strain RCC1991 is the first representative of *S. bioculatus* sequenced to date, both 18S and 28S rRNA gene sequences from this strain group with sequences of *Shionodiscus oestrupii* and *Shionodiscus ritscheri* (Figs. 1 and 2B).

Cymatosiraceae.

Arcocellulus cornucervis. Hasle, von Stosch &Syvertsen.

Cells are solitary, very small (apical axis: 3.0–3.5 µm; pervalvar axis: 1.4-1.7 µm; transapical axis: 1.7-2.2 µm) and slightly curved in broad girdle view. Each cell possesses two different valves, a process valve and a pili valve, which are convex and concave, respectively, in larger cells (Fig. 6, J and K). Each valve has two ocelluli (Fig. 6, K and L). The pili cross each other and bear conspicuous branches (Fig. 6J). The process valve possesses a central process (Figs. 6K and 7A). A marginal row of poroids is always present along the margin of the valve and a variable number of poroids can be present on the valve face. The basal siliceous layer may be smooth or ornamented by costae which can be indistinct or more convoluted (Fig. 6K). Costae seem to be more pronounced in process valves. Patches of short spinules can be present near the pilus base (Fig. 7A).

Arcocellulus cornucervis has been reported in temperate and cold waters of both hemispheres, including Arctic Ocean (Table 2).

18S phylogeny could not discriminate *Arcocellulus* spp. from the closely related genus *Minutocellulus* (Fig. 1). The 18S rRNA gene sequence from *A. cornucervis* RCC2270 is indeed highly related to two sequences from *Minutocellus polymorphus* (99.5 % sequence identity) and both form a well-supported (96% ML, 100 % NJ) clade which branches with that of other

455 representatives from the family Cymatosiraceae, namely *Papiliocellulus elegans, Cymatosira*
456 *belgica*, and *Brockmanniella brockmanni* (Fig. 1).
457 The 28S rRNA gene of *A. cornucervis* strain RCC2270 is closely related to that of two
458 unidentified Cymatosiraceae (95 and 95.9 % sequence identity) isolated from temperate waters
459 (Fig. 2B).
460 *Hemiaulaceae.*
461 *Eucampia groenlandica* Cleve.
462 Cells (apical axis: 7-24 µm) are rectangular in girdle view, slightly silicified and possess several
463 chloroplasts. Cells form colonies which can be straight or slightly curved in broad girdle view with
464 square to hexagonal apertures (Fig. 7B). A rimoportula is present on the centre of the valve (Fig.
465 7C).
466 *Eucampia groenlandica* was first reported from Baffin Bay in Davis Strait and is considered
467 typical of the northern cold waters (Table 2).
468 The 18S rRNA gene sequence from *E. groenlandica* strain RCC1996 groups with sequences of
469 *Eucampia zodiacus* (99.2 %) and *Eucampia antarctica* (99.0 %) forming a well-supported clade
470 (Fig. 1). The 28S rRNA gene from our strains is related to a sequence from *E. zodiacus* (96.9 %
471 sequence identity, Fig. 2B).
472 *Chaetocerotaceae.* We isolated 45 strains of the genus *Chaetoceros* and using the 28S rRNA
473 (Fig. 2B, and 3A) and ITS phylogeny (Fig. 3, B and C) we grouped these strains into 6 genotypes, 2
474 of them corresponding to the species *Chaetoceros decipiens* and *C. gelidus* respectively, and 4 other
475 being closely related genotypes affiliated to *C. neogracilis*.
476 *Chaetoceros decipiens* Cleve.
477 Cells (apical axis: 11-22 µm) were generally solitary in culture conditions but a few colonies
478 have been observed (Fig. 7, D and E). Each cell possesses several chloroplasts.
479 Chains are straight and the apertures are elliptical. All setae lie in the apical plane. The intercalary
480 setae emerge from the valve margin without a basal part and may fuse for a shorter or longer

481 distance. Terminal setae are U or V shaped (Fig. 7, D and E). The valve, with a high mantle, is
482 almost flat in girdle view (Fig. 7, F and G). Valves have a central annulus from which irregular ribs
483 radiate and are perforated with small poroids. The mantle is high and a marginal ridge is present
484 between the valve face and mantle (Fig. 7F). Terminal valves possess a very small central process
485 with a short external projection (Fig. 7H). Girdle bands are ornamented with parallel transverse
486 costae interspaced by hyaline areas with scattered small poroids (Fig. 7I). The setae are polygonal,
487 mostly four-sided, in cross section, with spines on the edges and a single longitudinal row of large
488 pores on each side.

489 *Chaetoceros decipiens* is a cosmopolitan species, common in arctic waters (Table 2).

490 The 18S rRNA gene sequence from our strain of *C. decipiens* (RCC1997) groups with a
491 GenBank sequence from *Chaetoceros* cf. *lorenzianus* (97.1 % sequence similarity, Fig. 1) and,
492 similarly, the 28S rRNA gene is closely related to GenBank sequences from *Chaetoceros*
493 *lorenzianus* (99.2 %), and groups with *Chaetoceros affinis* and *Chaetoceros diadema* (Fig. 2B).

494 *Chaetoceros gelidus* Chamnansinp, Li, Lundholm & Moestrup.

495 Cells (apical axis. 4-12 µm) with a single lobed chloroplast are joined in curved chains (Fig. 7J).
496 Several chains group together forming a spherical colony (Fig. 7 K). The setae emerge inside the
497 valve margin and merge after a short basal part forming narrow hexagonal apertures (Fig. 7L). In
498 valve view, the valve is circular to oval, in girdle view it is slightly concave with a small central
499 inflexion (Figs. 7L and 8A). Generally the cells have three short curved setae and one long straight
500 seta. The short setae have densely spirally arranged spines occurring throughout its length. In
501 contrast the long straight seta does not exhibit spines on its basal part, whereas on its distal part it
502 possesses spines which are more distant between each other (Fig. 8B). Both valves from each
503 resting spore are convex and smooth.(Fig. 8C). The crest reported in the original description
504 (Chamnansinp et al. 2013) is absent here. Variability in spore morphology of *C. gelidus* was
505 already reported (Degerlund et al. 2012, therein as *C. socialis*, northern strains).

506 The species has been reported from northern cold waters, including Arctic Ocean (Table 2).

507 The 18S sequence of *C. gelidus* clusters with a sequence of *C. socialis* (97.2 % sequence
508 identity, Fig. 1) and 28S rRNA sequences are identical to that of the type strain of *C. gelidus* (Fig.
509 2B).
510 *Chaetoceros neogracilis* (Schütt) VanLandingham.
511 Twenty-eight of the 36 strains of *C. neogracilis* isolated here have been observed by LM and
512 photographs are available for most of them (http://www.roscoff-culture-collection.org). Seven
513 strains have been further examined using EM (Table 1). Cells are generally solitary (Fig. 8, D-F)
514 but short colonies (3-6 cells) have been occasionally observed (Fig. 8G) in 9 strains. Cells are
515 relatively small (apical axis: 4-12 µm) and possess a single lobed chloroplast (Fig. 8, D-G). No
516 significant morphological and ultrastructural difference has been observed among the different
517 strains, with the exception of a certain variability in the orientation of the setae. As single cells,
518 some strains have straight setae diverging at an angle of 45°, whereas others have setae
519 perpendicular to pervalvar axis, and others have more curved setae (Fig. 8, D-F), but this variability
520 might be associated to the different cell sizes of the strains. In the colonies, cells are joined to form
521 straight chains and they are separated by apertures varying from elliptically shaped (Fig. 8, G and
522 H) to narrow slits (not shown). Terminal setae are U or V shaped. Valves are ornamented with
523 irregular costae originating from a central annulus. In the terminal valves, a slit-like process is
524 located in the centre of the annulus and it bears an external flattened tube (Fig. 8, I and J). The
525 central process is absent in the intercalary valves of the colonies, confirming that the chains are real
526 colonies and not cells in division (Fig. 8K). Intercalary setae originate from the valve apices, cross
527 immediately at the chain margin and diverge running in different directions (Fig. 8, H and L). The
528 setae are circular in cross section. They are composed by long spiral costae ornamented with
529 arrowhead-shaped spines (about 2 spines per l µm) and interconnected by short transverse costae
530 (Fig. 8, M and N). Spores were not observed in any of the tested strains.
531 The name *C. neogracilis* (basionym: *C. gracile* Schütt) has been attributed almost
532 indiscriminately to many small, unicellular *Chaetoceros* taxa collected worldwide (see Rines &

533 Hargraves 1988 for a discussion). The specific epithet can be found in the literature spelled as *C.*
534 *gracile* or *C. neogracile*, because the genus *Chaetoceros* was considered to be neutral, rather than
535 masculine. However the genus is currently recognised as a masculine word and the correct name of
536 the species is *C. neogracilis*. In more recent years, the species has been consistently reported as a
537 significant component of microbial communities in Arctic and Baltic (Table 2) as well as Antarctic
538 regions.
539 All the *C. neogracilis* strains isolated during the MALINA cruise share 100% identity in the V4
540 region of the 18S rRNA gene (data not shown). The full 18S rRNA gene has been sequenced for
541 strains RCC2016 and RCC2318. These two MALINA strains share identical 18S rRNA gene
542 sequence with the two Arctic strains ArM0004 e ArM0005 and form a well-supported clade with
543 the sequence from the Antarctic strain AnM0002 (98.9 % sequence identity, Fig. 1). The 28S rRNA
544 gene sequences from the MALINA strains of *C. neogracilis* cluster together (Fig. 2B) as well as
545 with a GenBank sequence from the Baltic strain CPH9 attributed to *C. fallax* (Chamnansinp et al.
546 2013) and have a sister clade which includes the sequences from three Antarctic strains (CCMP163,
547 CCMP189 and CCMP190). All these sequences branch with *Chaetoceros tenuissimus* forming a
548 well-supported clade (Fig. 3A).
549 *Genetic diversity of* Chaetoceros neogracilis *strains.* The MALINA strains of *C. neogracilis*
550 shared highly similar although not identical 28S rRNA gene sequence. Sequences can diverge by up
551 to 0.5 %. Both ITS markers as well as 28S rRNA gene indicate significant differences between the
552 Arctic and the Antarctic strains (Fig. 3), since the two groups form two separate branches. For
553 example the Arctic *C. neogracilis* RCC2014 shares with the Antarctic strain *Chaetoceros* sp.
554 CCMP189 95 %, 86 %, and 85 % sequence similarity for the 28S, ITS-1, and 5.8S + ITS-2,
555 respectively. The MALINA strains of *C. neogracilis* form four different clades based on all the
556 three markers used. Overall, based on either or both 28S rRNA (Fig. 3A) and ITS phylogeny (Figs.
557 3, B and C, S1), 20 strains belong to Clade I, 8 to Clade II, 2 to Clade III and 6 to Clade IV (Table
558 1). The 28S rRNA gene phylogeny (Fig. 3A) separates the *C. neogracilis* strains in two groups,

559 both with high (> 75 % in both ML and NJ) bootstrap support. One group consists of *C. neogracilis*
560 Clade I, whereas the second group includes the other 3 clades. Specifically strains from Clade II are
561 at the base of the group from which Clade III and Clade IV emerge with moderate (> 50 %) support
562 in both ML and NJ (Fig. 3A). The strain CPH9 falls within Clade II and the Antarctic strains
563 CCMP163, CCMP189 and CCMP190 are fully separated from *C. neogracilis*. Both ITS-1 and 5.8S
564 + ITS-2 trees includes 27 Arctic sequences from *C. neogracilis*, with 15 of them forming Clade I, 4
565 strains belonging to Clade II, 2 strains to Clade III, and 6 strains to Clade IV. Strains from each
566 clade cluster between them with moderate support in ITS-1 phylogeny and Clade II, Clade III, and
567 Clade IV group together with high bootstrap support (Fig. 3B). In 5.8S+ITS-2 phylogeny Clade II
568 and Clade III are highly supported, whereas Clade I and Clade IV are moderately supported; Clade
569 III groups with Clade I and some differences occur between the different strains from Clade II (Fig.
570 3C).
571 *Secondary structure of ITS-2.* We predicted the secondary structure of ITS-2 rRNA for our
572 strains of *C. neogracilis* to further investigate their genetic differences. We determined
573 Compensatory Base Changes (CBC) and Hemi-CBC in positions paired in the helices of the
574 secondary structure according to Coleman (2009). The secondary structure of ITS-2 from our
575 strains exhibits four helices (I, IIa, III and IV) typical of all eukaryotes (Coleman 2009) as well as
576 an additional helix (IIb) located between helix IIa and helix III (Fig. 9). Differences in the ITS-2
577 sequences from our strains occur at 14 positions, 9 of them located in paired positions of the
578 helices. This variability in paired positions consists in Hemi-CBC for 6 nucleotides, and CBC for 2
579 nucleotides. Two hemi-CBC occur in helix I (GC ↔ AC, and CG ↔ UG), 3 in helix III (CG ↔
580 UG, GC ↔ GU, GU ↔ AU), and 1 in helix IV (GU ↔ GC). Moreover 1 CBC occurs on helix IIa
581 between clade I and II (AU) versus clade IV (GC), with clade III showing a Hemi-CBC (GU)
582 towards the other three clades (Fig. 9).
583

584 DISCUSSION

585

586 *Combining microscopy and genetic data.* The combination of morphological and molecular
587 approaches on phytoplankton strains isolated during the MALINA cruise allowed the
588 characterization of cultured diatoms from the Beaufort Sea. To date about 10^4 species have been
589 described based solely on their morphology (Guiry 2012) and the application of molecular
590 approaches during the last decade revealed a considerable genetic diversity within key planktonic
591 morphospecies such as *Asterionellopsis glacialis* (Castracane) Round (Kaczmarska et al. 2014),
592 *Leptocylindrus danicus* Cleve (Nanjappa et al. 2013), *Pseudo-nitzschia pseudodelicatissima*
593 (Lundholm et al. 2003, Lundholm et al. 2006, Amato & Montresor 2008, Lundholm et al. 2012,
594 Lim et al. 2013, Orive et al. 2013), and *Skeletonema costatum* (Sarno et al. 2005, Sarno et al. 2007,
595 Kooistra et al. 2008). It has been suggested that the number of extant diatom species exceeds by
596 one order of magnitude those described to date (Mann & Vanormelingen 2013).
597 Our work provides both 18S and 28S rRNA gene sequences validated with detailed
598 morphological and ultrastructural information for 17 morphotypes. Both genes have been sequenced
599 here for the first time for 6 diatom species (*A. cornucervis, C. decipiens, E. groenlandica, S.*
600 *bioculatus,* and *T.* cf. *hispida*). The 18S gene of *C. gelidus, N. pellucida,* and *P. arctica* has been
601 also sequenced for the first time. Moreover, most of the gene sequences obtained from the Arctic
602 strains were different from sequences from conspecific strains collected from different geographic
603 areas that are available in GenBank. Finally, we investigated the genetic rRNA diversity of 36
604 *Chaetoceros* strains sharing the same 18S gene sequence, and clarified the identity of *C.*
605 *neogracilis,* a taxon that dominated genetic libraries from the Beaufort Sea.
606 *Genetic markers and species delimitation.* The taxonomic resolution of the genetic markers used
607 here was different according to the genus investigated, but it also varied within a given genus,
608 depending on the phylogenetic distance existing between congeneric species.

609 The 18S rRNA gene can successfully discriminate species within the genus *Nitzschia* (Rimet et
610 al. 2011) and the *C. closterium* species complex (Haitao et al. 2007). Both 18S and 28S rRNA
611 genes are commonly used for the taxonomic identification of *Thalassiosira* species (Kaczmarska et
612 al. 2006, Alverson et al. 2007, Hoppenrath et al. 2007) and here they provided a good taxonomic
613 resolution for all the Thalassiosiraceae representatives except *T. gravida*, which shares identical 18S
614 rRNA gene with *T. rotula* (Fig. 1). These two species show low phylogenetic distances also on 28S
615 rRNA gene phylogeny (Fig. 2B) and can be correctly separated only after ITS sequencing
616 (Whittaker et al. 2012).
617 The 28S rRNA gene is a relatively good molecular marker to discriminate most of *Pseudo-*
618 *nitzschia* species although a better resolution of phylogenetic relationships can be generally
619 achieved with the ITS rRNA possibly supplemented by the analysis of the secondary structure of
620 the ITS2 (Lundholm et al. 2003, Amato et al. 2007, Lundholm et al. 2012, Lim et al. 2013, Orive et
621 al. 2013, Percopo et al. 2016). *Pseudo-nitzschia arctica* and *P. granii* share highly similar 18S
622 rRNA gene sequences (Fig. 1) but can be better discriminated based on 28S rRNA (Fig. 2A), ITS
623 and *rbc*L phylogenies (Percopo et al. 2016).
624 Similarly, the MALINA strains of *C. neogracilis* and two Arctic strains isolated from the
625 Greenland Sea (EU090013 and EU090014; Choi et al. 2008) share identical 18S rRNA sequences
626 (Fig. 1), but they are genetically different at both 28S and ITS levels (Figs. 2B, and 3). 28S and ITS
627 rRNA phylogenies consistently grouped sequences from the Arctic strains of *C. neogracilis* into
628 four phylogenetically discrete clades (Fig. 3). The differences in the ITS secondary structure
629 confirm this grouping and would indicate reproductive isolation between the four clades of *C.*
630 *neogracilis* which may correspond to closely related but distinct cryptic species. Specifically a CBC
631 in helix IIa (Fig. 9) suggests reproductive isolation between clade I and clade II vs. clade IV, and
632 similarly the presence of at least a Hemi-CBC in the Helix III between Clade I and Clade II, as well
633 as between Clade III and all the other clades, suggests that the different clades are unable to
634 interbreed (Coleman 2009). The secondary structures of both ITS-1 and ITS-2 are involved in

635 ribosome assembly (Tschochner & Hurt 2003) and changes in paired positions likely affects gamete
636 compatibility preventing cells differing by CBC or Hemi-CBC from mating (Coleman 2001). For
637 diatoms, inability to interbreed has been demonstrated between strains differing by CBC or Hemi-
638 CBC in the ITS-2 within the *P. pseudodelicatissima* species complex (Amato et al. 2007).
639 The sympatric occurrence of distinct genetic clades of *C. neogracilis* in the Beaufort Sea gives
640 further support to the hypothesis that they should be considered separate species unable to
641 interbreed rather than different genotypes of a single species. Closely-related species or genotypes
642 can co-occur in the same environment and similar results were found previously in dinoflagellates.
643 Several ITS genotypes from the Atama complex, which consisted of *Alexandrium tamarense*
644 (Lebour) Balech, *Alexandrium fundyense* Balech, and *Alexandrium catenella* (Whedon & Kofoid)
645 Balech, co-occurred in the Chukchi Sea (Gu et al. 2013). In contrast, the Arctic *Micromonas*
646 (Lovejoy et al. 2007, Balzano et al. 2012b) consisted in a single ITS genotype (Balzano et al.
647 2012a), which dominated both surface and DCM, waters throughout the Beaufort Sea during the
648 MALINA cruise (Balzano et al. 2012b).
649 Notably, clone libraries based on 18S rRNA gene sequences, and high throughput amplicon
650 sequencing of the V4 or V9 regions of the 18S rRNA, which are widely used in environmental
651 studies (Stoeck et al. 2010, Comeau et al. 2011, Logares et al. 2012, Logares et al. 2014, Balzano et
652 al. 2015), failed to discriminate among the four clades of *C. neogracilis* and recovered them as a
653 unique genotype (Pawlowski et al. 2008, Lovejoy & Potvin 2011).
654 Both 18S and 28S rRNA genes are too conserved for some genera failing to discriminate the
655 different species. For example, *A. septentrionalis* shared identical 18S rRNA and 28S rRNA gene
656 sequences with *A. longicornis* (Figs 1, and 2A). These two species can be distinguished only using a
657 combination of several nuclear and plastidial encoded genes (Sorhannus & Fox 2012).
658 The 18S rRNA gene is highly conserved also within the family Cymatosiraceae, where *A.*
659 *cornucervis* strain RCC2270 shares almost identical 18S rRNA with two GenBank sequences from
660 *M. polymorphus* (Fig. 1), and the 2 species share 100% identity in the V4 region (Luddington et al.

2012). The extent of the variability of the 28S rRNA gene within the Cymatosiraceae is not clear since no other sequence from this family is available on GenBank and *A. cornucervis* RCC2270 shares highly similar 28S rRNA gene with two unidentified Cymatosiraceae strains (Fig. 2B).

Overall, ITS-2 provides a higher taxonomic resolution than 28S, but although it was proposed as a universal barcode for diatoms (Moniz & Kaczmarska 2010, Guo et al. 2015), very few ITS sequences are available to date in GenBank compared to 18S and 28S and its high variability makes the alignment between different genera difficult or even impossible. Similarly the 28S rRNA gene is less conserved than the 18S rRNA allowing a better discrimination between congeneric species but 28S sequences are available for a larger number of diatom species. Ideally sequencing the entire rRNA operon from the same specimen would allow the best taxonomic resolution and provide taxonomic annotation from most species in environmental studies. Single molecule sequencing technologies such as PacBio could allow the sequencing of reads as long as 5,000 bp (Mikheyev & Tin 2014, Schloss et al. 2016). For current sequencing technologies the 28S rRNA seems the best compromise between resolutive power and easiness of alignment, for environmental studies focused on diatoms, whereas 18S rRNA gene sequencing can be used for general studies on microbial eukaryotes.

Diatoms in the Beaufort Sea. Diatoms represented an important fraction of the nano- and microphytoplankton identified during the MALINA cruise (Balzano et al. 2012b, Coupel et al. 2015) with *Chaetoceros* and *Thalassiosira* being the most represented genera. Different species from these two genera are frequently observed in Arctic waters where they typically dominate phytoplankton assemblages (Booth & Horner 1997, Lovejoy et al. 2002, Ratkova & Wassmann 2002), eventually forming spring blooms (Booth et al. 2002, Sukhanova et al. 2009).

In spite of the high diversity reported in previous studies (Sukhanova et al. 2009), only few environmental ribotypes associated with *T. nordenskioeldii* were detected by T-RFLP among sorted photosynthetic eukaryotes during the MALINA cruise (Balzano et al. 2012b) and only *T. nordenskioeldii, T. gravida, Thalassiosira pacifica* and few undetermined species were observed by

687 microscopy counts (http://malina.obs-vlfr.fr), accounting for a low proportion of the phytoplankton
688 community. Clearly, *Thalassiosira* species did not bloom in the Beaufort Sea during late summer
689 2009 and *T. gravida, T.* cf. *hispida,* and *T. minima* were possibly only present in low abundance.
690 The high number of *Chaetoceros* strains (45), mostly represented by *C. gelidus* and *C.*
691 *neogracilis*, reflected the dominance of these two species in the summer phytoplankton
692 assemblages, already shown by the genetic libraries (Balzano et al. 2012b). Notably, phytoplankton
693 counts confirmed the high abundance of *C. gelidus* and other unidentified morphotypes, but barely
694 reported the occurrence of *C. neogracilis*. This discrepancy indicates that cells of *C. neogracilis*
695 might have been erroneously attributed to several different solitary species, such as *C. tenuissimus*
696 or *Chaetoceros simplex* Ostenfeld, or other undetermined *Chaetoceros.* We also suggest that cell
697 chains of *C. neogracilis,* which were described for the first time in this study, might have been
698 wrongly identified as the freshwater species *Chaetoceros wighamii* Brightwell (http://malina.obs-
699 vlfr.fr; see Bosak et al. 2015 for a discussion on *C. wighamii*). Similarly, the doubtful reports of *C.*
700 *wighamii* from the Baltic Sea and Danish waters could indeed refer to *C. neogracilis*, as suggested
701 by the morphological and ultrastructural similarity between Arctic strains of *C. neogracilis*
702 described in this study and culture material from Danish waters attributed to *C. wighamii* (see fig.
703 224 in Jensen & Moestrup 1998).
704 Other colonial *Chaetoceros* species found in the phytoplankton counts were not isolated in this
705 study because they might be more difficult to bring into culture compared to *C. gelidus* and *C.*
706 *neogracilis,* or because they are rare, as suggested by their absence in the 18S rRNA libraries and in
707 T-RFLP analyses (Balzano et al. 2012b).
708 Interestingly most of the *C. neogracilis* strains from Clade I and Clade II as well as all the strains
709 of Clade IV were isolated from surface waters (Table 1), whereas 5 out of 8 strains of *C. gelidus*
710 and both *C. neogracilis* Clade III strains were isolated from DCM waters. During the MALINA
711 cruise surface waters were warmer, less saline (Table S1), and poorer in nutrients
712 (http://malina.obs-vlfr.fr/data.html) compared to DCM waters. We do not know whether these

patterns are indicative of ecological preferences for these genotypes. However surface genotypes might be adapted to lower salinities, higher irradiation, higher temperatures and lower nutrient concentrations. Unfortunately, the different clades of *C. neogracilis* have identical T-RFLP ribotypes and therefore their relative contribution to the environmental samples from the MALINA cruise (Balzano et al. 2012b) cannot be discerned.

Notably, some of the strains isolated here show similarities with specimens from other environments affected by seasonal salinity shifts similar to those characterizing the Beaufort Sea. One of the *C. neogracilis* strains belonging to Clade II, CPH9 (Fig. 3A), was isolated in the Baltic Sea, and *C. closterium* RCC1985 forms a clade, in the 28S rRNA tree, with a strain (K-520, Fig. 2A) which has been isolated from Kattegat (Lundholm et al. 2002). Interestingly, a number of environmental sequences as well as photosynthetic flagellates isolated from the surface waters of the Beaufort Sea during the MALINA cruise are genetically related to strains or environmental sequences from the Baltic Sea (Balzano et al. 2012a, Balzano et al. 2012b). Despite the significant differences in temperature and salinity between the Beaufort Sea and both the Baltic Sea and the Kattegat, the genetic similarities found in samples from these areas might be associated with the seasonal ice and the shifts in salinity occurring in these environments.

The Chaetoceros neogracilis species complex. *Chaetoceros neogracilis* was originally described as *Chaetoceros gracile* Schütt from the Baltic Sea as solitary, small *Chaetoceros* species (Schütt 1895). Due to the scanty original description and to the lack of distinctive features in such small single cell-taxa, the name has most probably been attributed to different and not related taxa collected worldwide (Rines & Hargraves 1988). All the Arctic strains isolated during the MALINA cruise share a similar cell morphology with *C. neogracilis*, together with a prevalent absence of colony formation. Indeed, *C. neogracilis* was originally described as a solitary species whereas some of the MALINA strains have been observed forming short colonies. Notably, the ability to occasionally form colonies is common to other *Chaetoceros* species considered solitary, as it has also been observed in the related species *C. tenuissimus* (D. Sarno, pers. observation). The original

739 description of the species (Schütt 1895) includes a spiny spore that unfortunately has not been
740 observed in our study.
741 Based on the available information, it is not possible to provide the authoritative taxonomic
742 revision required by the International Code of Nomenclature for algae, fungi, and plants (McNeill et
743 al. 2012) to establish each of the four clades as valid species and to assess if one of them
744 corresponds to *C. neogracilis* sensu stricto. Further analyses are required to provide additional
745 ultrastructural information on a larger number of strains from the four clades to be compared with
746 the type material of *C. neogracilis* and eventually designate an epitype. In the meantime, we
747 propose that the Arctic *Chaetoceros* strains sharing very similar morphology and molecular
748 signatures described here are considered as affiliated to *C. neogracilis* species complex. The
749 provisional ascription of the name *C. neogracilis* to the Arctic *Chaetoceros* complex is supported
750 by the fact that one of the strains (i.e., CPH9, syn K-1665, http://www.sccap.dk/) belonging to
751 Clade II of the species complex, was isolated from Danish waters in the Baltic Sea, which is the
752 type locality of *C. neogracilis*. The morphologically similar Antarctic species, which has been
753 frequently identified as *C. neogracilis* and is represented in this study by the strains AnM0002,
754 CCMP187, CCMP189 and CCMP190 (Choi et al. 2008), corresponds to a related but genetically
755 distinct (Figs. 1, 2B, and 7) and probably undescribed species, here named as *Chaetoceros* sp.
756 *Biogeography of Arctic diatoms.* Most of the diatom species (10 out of 17) characterized in this
757 study have a distribution confined to the northern/polar area, including *Pseudo-nitzschia arctica*
758 (Percopo et al. 2016), and the *C. neogracilis* species complex, which was one of the few Arctic
759 phylotypes identified by their 18S rRNA gene (Lovejoy & Potvin 2011) (Table 2). In addition, the
760 MALINA strain of *C. closterium* (RCC1985) is phylogenetically distant from any lineage described
761 for this species complex (Haitao et al. 2007) and might correspond to an Arctic genotype.
762 Endemism has been recently suggested for a number of Arctic protists from the Baffin Bay and the
763 Beaufort Sea (Terrado et al. 2013). Endemic polar species include in particular the green alga Arctic
764 *Micromonas* (Lovejoy et al. 2007), several foraminiferan species (Darling et al. 2007, Pawlowski et

765 al. 2008), and the Antarctic terrestrial diatoms *Pinnularia borealis* Ehrenberg and *Hantzschia*
766 *amphioxys* (Ehrenberg) Grunow (Souffreau et al. 2013).

767 Two species found here, *Porosira glacialis* and *T. gravida*, are considered to have bipolar
768 distribution (McMinn et al. 2005, Whittaker et al. 2012, Goes et al. 2014). The presence of the same
769 species in ecologically related but geographically distant environments, such as the Arctic and the
770 Antarctic, has been suggested for two *Fragilariopsis* Hustedt species (Lundholm & Hasle 2008) as
771 well as the dinoflagellate *Polarella glacialis* Montresor, Procaccini & Stoecker (Montresor et al.
772 2003) and the ciliate *Euplotes nobilii* Valbonesi & Luporini (Di Giuseppe et al. 2014). Polar species
773 can hardly survive in temperate and tropical waters and the evolution of polar species is thus
774 unlikely to arise from transport of living cells between Arctic and Antarctic waters. The presence of
775 bipolar species could be associated with a migration occurred during the last glacial period, where
776 colder seawater at low latitudes would have permitted the survival of cells during their transport
777 across the globe or due to more recent transport of resting forms (Montresor et al. 2003). Such
778 resting forms could survive tropical waters or in alternative they might have been transported across
779 the globe via the global ocean conveyor belt or other deep cold currents.

780 Few (5) of the strains characterized in this study belong to species that are supposed to have a
781 wide geographical distribution (Table 2). Molecular methods have demonstrated conspecificity in
782 widely distributed morphospecies, as for example some *Pseudo-nitzschia* (Lelong et al. 2012) or
783 *Skeletonema* (Kooistra et al. 2008) species. Other studies on plankton biogeography indicate that
784 populations previously thought to make up unique cosmopolitan species are often genetically
785 distinct and reproductively isolated (Kooistra et al. 2008, Casteleyn et al. 2010). Indeed, the
786 northern/polar ecotype of the worldwide-considered species, *C. socialis*, has been recently
787 described as a distinct species, i.e., *C. gelidus*, based on physiological, morphological and molecular
788 evidence (Degerlund et al. 2012, Huseby et al. 2012, Chamnansinp et al. 2013). Subsequently all
789 the previous reports of *C. socialis* in Arctic waters (Booth et al. 2002, Ratkova & Wassmann 2002,

790 Sukhanova et al. 2009), including those reported for the MALINA cruise (Balzano et al. 2012b), are
791 likely to correspond to *C. gelidus*.
792 Similarly, the degree of interspecific divergence between the cosmopolitan *T. rotula* and the
793 bipolar *T. gravida* advocates they should be treated as separate species (Whittaker et al. 2012),
794 despite previous studies suggesting that the two morphotypes are likely to be a single species
795 (Syvertsen 1977, Sar et al. 2011). We cannot exclude that the use of more sensitive molecular
796 markers would allow to identify differences among geographical populations of bipolar or
797 cosmopolitan species, as demonstrated for the cosmopolitan species *Pseudo-nitzschia pungens*
798 (Casteleyn et al. 2010). Further analyses will be required to evaluate the slight difference here found
799 among the 28S rRNA gene sequences of the Arctic and Antarctic strains of *Porosira glacialis*.
800 Therefore while some species distribution patterns seem to support the hypothesis of ubiquity
801 (Finlay & Fenchel 2004), other species are far more restricted. The availability of validated
802 reference sequences for arctic diatoms will facilitate the interpretation of metabarcoding data and
803 will allow to test theories on dispersal and biogeographic patterns in protists using large scale
804 screening of environmental samples.

ACKNOWLEDGMENTS

We thank all participants to MALINA cruise for their help. This work was mainly supported by the MALINA project, in particular ANR (ANR-08-BLAN-0308), which funded SB post-doctoral work, the ASSEMBLE EU FP7 research infrastructure initiative (EU-RI-227799) which funded IP and RS fellowships, the ANR Project PhytoPol (ANR-15-CE02-0007-02) and the EU project MaCuMBA (FP7-KBBE-2012-6-311975). We are grateful to F. Iamunno and R. Graziano (Electron Microscopy Service, SZN) for EM support.

REFERENCES

Aizawa, C., Tanimoto, M. & Jordan, R. W. 2005. Living diatom assemblages from North Pacific and Bering Sea surface waters during summer 1999. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.* 52:2186-205.

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.

Amato, A., Kooistra, W. H. C. F., Ghiron, J. H. L., Mann, D. G., Proschold, T. & Montresor, M. 2007. Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* 158:193-207.

Amato, A. & Montresor, M. 2008. Morphology, phylogeny, and sexual cycle of *Pseudo-nitzschia mannii* sp. nov. (Bacillariophyceae): a pseudo-cryptic species within the *P. pseudodelicatissima* complex. *Phycologia* 47:487-97.

Balzano, S., Gourvil, P., Siano, R., Chanoine, M., Marie, D., Lessard, S., Sarno, D. & Vaulot, D. 2012a. Diversity of cultured photosynthetic flagellates in the northeast Pacific and Arctic Oceans in summer. *Biogeosciences* 9:4553-71.

Balzano, S., Marie, D., Gourvil, P. & Vaulot, D. 2012b. Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. *ISME J* 6:1480-98.

Balzano, S., Abs, E. & Leterme, S. C. 2015. Protist diversity along a salinity gradient in a coastal lagoon. *Aquat. Microb. Ecol.* 74:263-77.

Bérard-Therriault, L., Poulin, M. & Bossé, L. 1999. *Guide d'identification du phytoplancton marin de l'estuaire et du golfe du Saint-Laurent incluant également certains protozoaires Canadian* NRC Research Press, Ottawa, 387 pp.

Booth, B. C. & Horner, R. A. 1997. Microalgae on the Arctic Ocean Section, 1994: species abundance and biomass. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.* 44:1607-22.

Booth, B. C., Larouche, P., Belanger, S., Klein, B., Amiel, D. & Mei, Z. P. 2002. Dynamics of *Chaetoceros socialis* blooms in the North Water. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.* 49:5003-25.

Bosak, S., Gligora Udovic, M. & Sarno, D. 2015. Morphological study of *Chaetoceros wighamii* Brightwell (Chaetocerotaceae, Bacillariophyta) from Lake Vrana, Croatia. *Acta Bot. Croat.* 74: 233–244.

Brugel, S., Nozais, C., Poulin, M., Tremblay, J.-E., Miller, L. A., Simpson, K. G., Gratton, Y. & Demers, S. 2009. Phytoplankton biomass and production in the southeastern Beaufort Sea in autumn 2002 and 2003. *Mar. Ecol.-Prog. Ser.* 377:63-77.

Cărăus, I. 2012. *Algae of Romania. A distributional checklist of actual algae. Version 2.3 third revision.* Univ. Bacau, 809 pp.

Carmack, E. C. & MacDonald, R. W. 2002. Oceanography of the Canadian shelf of the Beaufort Sea: a setting for marine life. *Arctic* 55:29-45.

Caron, G., Michel, C. & Gosselin, M. 2004. Seasonal contributions of phytoplankton and fecal pellets to the organic carbon sinking flux in the North Water (northern Baffin Bay). *Mar. Ecol.-Prog. Ser.* 283:1-13.

Casteleyn, G., Leliaert, F., Backeljau, T., Debeer, A.-E., Kotaki, Y., Rhodes, L., Lundholm, N., Sabbe, K. & Vyverman, W. 2010. Limits to gene flow in a cosmopolitan marine planktonic diatom. *P. Natl. Acad. Sci. USA* 107:12952-57.

Chamnansinp, A., Li, Y., Lundholm, N. & Moestrup, O. 2013. Global diversity of two widespread, colony forming diatoms of the marine plankton, *Chaetoceros socialis* (syn. *C. radians*) and *Chaetoceros gelidus* sp. nov. *J. Phycol.* 49:1128-41.

863 Choi, H. G., Joo, H. M., Jung, W., Hong, S. S., Kang, J. S. & Kang, S. H. 2008. Morphology and
864 phylogenetic relationships of some psychrophilic polar diatoms (Bacillariophyta). *Nova*
865 *Hedwigia*:7-30.
866 Cleve, P. 1896. Diatoms from Baffins Bay and Davis Strait. *Kongliga Svenska Vetenskaps-*
867 *Akademien* 22:1-22.
868 Coleman, A. W. 2001. Biogeography and speciation in the *Pandorina/Volvulina* (Chlorophyta)
869 superclade. *J. Phycol.* 37:836-51.
870 Coleman, A. W. 2009. Is there a molecular key to the level of "biological species" in eukaryotes? A
871 DNA guide. *Mol. Phylogenet. Evol.* 50:197-203.
872 Comeau, A. M., Li, W. K. W., Tremblay, J.-E., Carmack, E. C. & Lovejoy, C. 2011. Arctic Ocean
873 microbial community structure before and after the 2007 record sea ice minimum. *Plos One*
874 6:e27492.
875 Coupel, P., Matsuoka, A., Ruiz-Pino, D., Gosselin, M., Marie, D., Tremblay, J.-E. & Babin, M.
876 2015. Pigment signatures of phytoplankton communities in the Beaufort Sea.
877 *Biogeosciences* 12:991-1006.
878 Crawford, R. M., Gardner, C. & Medlin, L. K. 1994. The genus *Attheya*. I. A description of four
879 new taxa, and the transfer of *Gonioceros septentrionalis* and *G. armatus*. *Diatom Res.* 9:27-
880 51.
881 Darling, K. F., Kucera, M. & Wade, C. M. 2007. Global molecular phylogeography reveals
882 persistent Arctic circumpolar isolation in a marine planktonic protist. *P. Natl. Acad. Sci.*
883 *USA*104:5002-07.
884 Degerlund, M. & Eilertsen, H. C. 2010. Main species characteristics of phytoplankton spring
885 blooms in NE Atlantic and Arctic waters (68-80°N). *Estuar. Coast.* 33:242-69.
886 Degerlund, M., Huseby, S., Zingone, A., Sarno, D. & Landfald, B. 2012. Functional diversity in
887 cryptic species of *Chaetoceros socialis* Lauder (Bacillariophyceae). *J. Plankton Res.*
888 34:416-31.
889 Di Giuseppe, G., Erra, F., Frontini, F. P., Dini, F., Vallesi, A. & Luporini, P. 2014. Improved
890 description of the bipolar ciliate, *Euplotes petzi*, and definition of its basal position in the
891 *Euplotes* phylogenetic tree. *Eur. J. Protistol.* 50:402-11.
892 Dunthorn, M., Klier, J., Bunge, J. & Stoeck, T. 2012. Comparing the hyper-variable V4 and V9
893 regions of the small subunit rDNA for assessment of ciliate environmental diversity. *J.*
894 *Eukaryot. Microbiol.* 59:185-87.
895 Felsenstein, J. 1985. Confidence limits on phylogenies. An approach using the bootstrap. *Evolution*
896 39:783-91.
897 Finlay, B. J. & Fenchel, T. 2004. Cosmopolitan metapopulations of free-living microbial
898 eukaryotes. *Protist* 155:237-44.
899 Goes, J. I., Gothes, H. D. R., Haugen, E. M., McKee, K. T., D'Sa, E. J., Chekalyuk, A. M.,
900 Stoecker, D. K., Stabeno, P. J., Saitoh, S. & Sambrotto, R. N. 2014. Fluorescence, pigment
901 and microscopic characterization of Bering Sea phytoplankton community structure and
902 photosynthetic competency in the presence of a Cold Pool during summer. *Deep-Sea Res.*
903 *Part II-Top. Stud. Oceanogr.* 109:84-99.
904 Gosselin, M., Levasseur, M., Wheeler, P. A., Horner, R. A. & Booth, B. C. 1997. New
905 measurements of phytoplankton and ice algal production in the Arctic Ocean. *Deep-Sea Res.*
906 *Part II-Top. Stud. Oceanogr.* 44:1623-44.
907 Gu, H., Zeng, N., Xie, Z., Wang, D., Wang, W. & Yang, W. 2013. Morphology, phylogeny, and
908 toxicity of Atama complex (Dinophyceae) from the Chukchi Sea. *Polar Biol* 36:427-36.
909 Guillard, R. R. L. 1975. Culture of phytoplankton for feeding marine invertebrates. *In*: Smith, W. L.
910 & Chanley, M. H. [Eds.] *Culture of marine invertebrate animals.* Plenum Book Publication
911 Corporation, New York, pp. 29-60.
912 Guillou, L., Eikrem, W., Chretiennot-Dinet, M. J., Le Gall, F., Massana, R., Romari, K., Pedros-
913 Alio, C. & Vaulot, D. 2004. Diversity of picoplanktonic prasinophytes assessed by direct

914 nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from
915 oceanic and coastal marine ecosystems. *Protist* 155:193-214.
916 Guinder, V. A., Molinero, J. C., Popovich, C. A., Marcovecchio, J. E. & Sommer, U. 2012.
917 Dominance of the planktonic diatom *Thalassiosira minima* in recent summers in the Bahía
918 Blanca Estuary, Argentina. *J. Plankton Res.* 34:995-1000.
919 Guiry, M. D. 2012. How many species of algae are there? *J. Phycol.* 48:1057-63.
920 Guo, L., Sui, Z., Zhang, S., Ren, Y. & Liu, Y. 2015. Comparison of potential diatom 'barcode'
921 genes (the 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and
922 determining species taxonomy in the Bacillariophyta. *Int. J. Syst. Evol. Micr.* 65:1369-80.
923 Haitao, L., Guanpin, Y., Ying, S., Suihan, W. & Xiufang, Z. 2007. *Cylindrotheca closterium* is a
924 species complex as was evidenced by the variations of rbcL gene and SSU rDNA. *Journal*
925 *of Ocean University of China* 6:167-74.
926 Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignement editor and analysis
927 program for Windows 95/98/NT. *Nucleic Acid symposium Series* 41:95-98.
928 Hällfors, G. 2004. Checklist of Baltic Sea phytoplankton species (including some heterotrophic
929 protistan groups). *Baltic Sea Environment Proceedings* 95:1-208.
930 Hasle, G. R. 1964. *Nitzschia* and *Fragilariopsis* species studied in the light and electron
931 microscopes. I. Some marine species of the group Nitzschiaella and Lanceolate. *Skrifter*
932 *utgitt av Det Norske Videnskaps-Akademi i Oslo. I Matematisk-Naturvidenskapelig Klasse.*
933 *Ny serie* 16:1-48.
934 Hasle, G. R., Medlin, L. K. & Syvertsen, E. E. 1994. *Synedropsis* gen. nov., a genus of araphid
935 diatoms associated with sea ice. *Phycologia* 33:248-70.
936 Hasle, G. R. & Syvertsen, E. E. 1997. Marine diatoms. *In*: Tomas, C. R. [Ed.] *Identifying marine*
937 *phytoplankton.* Academic Press, San Diego, pp. 5-385.
938 Hasle, G. R., Von Stosch, H. A. & Syvertsen, E. E. 1983. Cymatosiraceae, a new diatom family.
939 *Bacillaria* 6:9-156.
940 Hill, V., Cota, G. & Stockwell, D. 2005. Spring and summer phytoplankton communities in the
941 Chukchi and Eastern Beaufort Seas. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.* 52:3369-
942 85.
943 Hoppenrath, M., Beszteri, B., Drebes, G., Halliger, H., Van Beusekom, J. E. E., Janisch, S. &
944 Wiltshire, K. H. 2007. *Thalassiosira* species (Bacillariophyceae, Thalassiosirales) in the
945 North Sea at Helgoland (German bight) and Sylt (North Frisian Wadden Sea). A first
946 approach to assessing diversity. *Eur. J. Phycol.* 42:271-88.
947 Horner, R. & Schrader, G. C. 1982. Relative contributions of ice algae, phytoplankton, and benthic
948 microalgae to primary production in nearshore regions of the Beaufort Sea. *Arctic* 35:485-
949 503.
950 Huseby, S., Degerlund, M., Zingone, A. & Hansen, E. 2012. Metabolic fingerprinting reveals
951 differences between northern and southern strains of the cryptic diatom *Chaetoceros*
952 *socialis*. *Eur. J. Phycol.* 47:480-89.
953 Jahn, R. & Kusber, W. H. 2005. Reinstatement of the genus *Ceratoneis* Ehrenberg and
954 lectotypification of its type specimen: *C. Closterium* Ehrenberg. *Diatom Res.* 20:295-304.
955 Jensen, K. G. & Moestrup, Ø. 1998. The genus *Chaetoceros* (Bacillariophyceae) in inner Danish
956 coastal waters. *Nord. J. Bot.* 18 :88
957 Jukes, T. H. & Cantor, C. R. 1969. Evolution of protein molecules. *In*: Munro, H. N. [Ed.]
958 *Mammalian protein metabolism.* Academic Press, New York, pp. 21-123.
959 Kaczmarska, I., Beaton, M., Benoit, A. C. & Medlin, L. K. 2006. Molecular phylogeny of selected
960 members of the order Thalassiosirales (Bacillariophyta) and evolution of the fultoportula. *J.*
961 *Phycol.* 42:121-38.
962 Kaczmarska, I., Mather, L., Luddington, I. A., Muise, F. & Ehrman, J. 2014. Cryptic diversity in a
963 cosmopolitan diatom known as *Asterionellopsis glacialis* (Fragilariaceae): Implications for
964 ecology, biogeography, and taxonomy. *Am. J. Bot.* 101:267-86.

965 Katsuki, K., Takahashi, K., Onodera, J., Jordan, R. W. & Suto, I. 2009. Living diatoms in the
966 vicinity of the North Pole, summer 2004. *Micropaleontology* 55:137-70.
967 Keller, A., Schleicher, T., Schultz, J., Muller, T., Dandekar, T. & Wolf, M. 2009. 5.8S-28S rRNA
968 interaction and HMM-based ITS2 annotation. *Gene* 430:50-57.
969 Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through
970 comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111-20.
971 Kooistra, W. H. C. F., Sarno, D., Balzano, S., Gu, H., Andersen, R. A. & Zingone, A. 2008. Global
972 diversity and biogeography of *Skeletonema* species (Bacillariophyta). *Protist* 159:177-93.
973 Kooistra, W., Sarno, D., Hernandez-Becerril, D. U., Assmy, P., Di Prisco, C. & Montresor, M.
974 2010. Comparative molecular and morphological phylogenetic analyses of taxa in the
975 Chaetocerotaceae (Bacillariophyta). *Phycologia* 49:471-500.
976 Le Gall, F., Rigaut-Jalabert, F., Marie, D., Garczarek, L., Viprey, M., Gobet, A. & Vaulot, D. 2008.
977 Picoplankton diversity in the South-East Pacific Ocean from cultures. *Biogeosciences* 5:203-
978 14.
979 Lee, M.-A., Faria, D. G., Han, M.-S., Lee, J. & Ki, J.-S. 2013. Evaluation of nuclear ribosomal
980 RNA and chloroplast gene markers for the DNA taxonomy of centric diatoms. *Biochem.*
981 *Syst. Ecol.* 50:163-74.
982 Lelong, A., Hégaret, H., Soudant, P. & Bates, S., S. 2012. *Pseudo-nitzschia* (Bacillariophyceae)
983 species, domoic acid and amnesic shellfish poisoning: revisiting previous paradigms.
984 *Phycologia* 51:168-216.
985 Lenaers, G., Maroteaux, L., Michot, B. & Herzog, M. 1989. Dinoflagellates in evolution. A
986 molecular phylogenetic analysis of large subunit ribosomal RNA. *J. Mol. Evol.* 29:40-51.
987 Lepère, C., Demura, M., Kawachi, M., Romac, S., Probert, I. & Vaulot, D. 2011. Whole Genome
988 Amplification (WGA) of marine photosynthetic eukaryote populations. *FEMS Microbiol.*
989 *Ecol.* 76:513-23.
990 Lim, H. C., Teng, S. T., Leaw, C. P. & Lim, P. T. 2013. Three novel species in the *Pseudo-nitzschia*
991 *pseudodelicatissima* complex: *P. batesiana* sp. nov., *P. lundholmiae* sp. nov., and *P. fukuyoi*
992 sp. nov. (Bacillariophyceae) from the strait of Malacca, Malaysia. *J. Phycol.* 49:902-16.
993 Logares, R., Audic, S., Santini, S., Pernice, M. C., de Vargas, C. & Massana, R. 2012. Diversity
994 patterns and activity of uncultured marine heterotrophic flagellates unveiled with
995 pyrosequencing. *ISME J* 6:1823-33.
996 Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J.,
997 Dolan, J. R., Dunthorn, M., Edvardsen, B., Gobet, A., Kooistra, W. H. C. F., Mahe, F., Not,
998 F., Ogata, H., Pawlowski, J., Pernice, M. C., Romac, S., Shalchian-Tabrizi, K., Simon, N.,
999 Stoeck, T., Santini, S., Siano, R., Wincker, P., Zingone, A., Richards, T. A., de Vargas, C. &
1000 Massana, R. 2014. Patterns of Rare and Abundant Marine Microbial Eukaryotes. *Curr. Biol.*
1001 24:813-21.
1002 Lovejoy, C., Legendre, L., Martineau, M. J., Bacle, J. & von Quillfeldt, C. H. 2002. Distribution of
1003 phytoplankton and other protists in the North Water. *Deep-Sea Res. Part II-Top. Stud.*
1004 *Oceanogr.* 49:5027-47.
1005 Lovejoy, C., Vincent, W. F., Bonilla, S., Roy, S., Martineau, M.-J., Terrado, R., Potvin, M.,
1006 Massana, R. & Pedros-Alio, C. 2007. Distribution, phylogeny, and growth of cold-adapted
1007 picoprasinophytes in arctic seas. *J. Phycol.* 43:78-89.
1008 Lovejoy, C. & Potvin, M. 2011. Microbial eukaryotic distribution in a dynamic Beaufort Sea and
1009 the Arctic Ocean. *J. Plankton Res.* 33:431-44.
1010 Luddington, I. A., Kaczmarska, I. & Lovejoy, C. 2012. Distance and character-based evaluation of
1011 the V4 region of the 18S rRNA gene for the identification of Diatoms (Bacillariophyceae).
1012 *Plos One* 7:e45664.
1013 Luddington, I. A., Lovejoy, C. & Kaczmarska, I. 2016. Species-rich meta-communities of the
1014 diatom order Thalassiosirales in the Arctic and northern Atlantic Ocean. *J. Plankton Res.*
1015 doi:10.1093/plankt/fbw030

1016 Lundholm, N., Daugbjerg, N. & Moestrup, O. 2002. Phylogeny of the Bacillariaceae with emphasis
1017 on the genus *Pseudo-nitzschia* (Bacillariophyceae) based on partial LSU rDNA. *Eur. J.*
1018 *Phycol.* 37:115-34.
1019 Lundholm, N., Moestrup, O., Hasle, G. R. & Hoef-Emden, K. 2003. A study of the *Pseudo-*
1020 *nitzschia pseudodelicatissima/cuspidata* complex (Bacillariophyceae): What is *P.*
1021 *pseudodelicatissima*? *J. Phycol.* 39:797-813.
1022 Lundholm, N., Moestrup, O., Kotaki, Y., Hoef-Emden, K., Scholin, C. & Miller, P. 2006. Inter- and
1023 intraspecific variation of the *Pseudo-nitzschia delicatissima* complex (Bacillariophyceae)
1024 illustrated by rRNA probes, morphological data and phylogenetic analyses. *J. Phycol.*
1025 42:464-81.
1026 Lundholm, N. & Hasle, G. R. 2008. Are *Fragilariopsis cylindrus* and *Fragilariopsis nana* bipolar
1027 diatoms? Morphological and molecular analyses of two sympatric species. *Nova Hedwigia*:
1028 231-50.
1029 Lundholm, N., Bates, S. S., Baugh, K. A., Bill, B. D., Connell, L. B., Leger, C. & Trainer, V. L.
1030 2012. Cryptic and pseudo-cryptic diversity in diatoms with description of *Pseudo-nitzschia*
1031 *hasleana* sp. nov. and *P. fryxelliana* sp. nov. *J. Phycol.* 48:436-54.
1032 Majaneva, M., Rintala, J. M., Piisila, M., Fewer, D. P. & Blomster, J. 2012. Comparison of
1033 wintertime eukaryotic community from sea ice and open water in the Baltic Sea, based on
1034 sequencing of the 18S rRNA gene. *Polar Biol.* 35:875-89.
1035 Mann, D. G. & Vanormelingen, P. 2013. An Inordinate Fondness? The number, distributions, and
1036 origins of diatom species. *J. Eukaryot. Microbiol.* 60:414-20.
1037 Marchetti, A., Lundholm, N., Kotaki, Y., Hubbard, K., Harrison, P. J. & Armbrust, E. V. 2008.
1038 Identification and assessment of domoic acid production in oceanic *Pseudo-nitzschia*
1039 (Bacillariophyceae) from iron-limited waters in the northeast subarctic pacific. *J. Phycol.*
1040 44:650-661.
1041 Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M. & Turner, D. H. 2004.
1042 Incorporating chemical modification constraints into a dynamic programming algorithm for
1043 prediction of RNA secondary structure. *P. Natl. Sci. USA*101:7287-92.
1044 McMinn, A., Pankowski, A. & Delfatti, T. 2005. Effect of hyperoxia on the growth and
1045 photosynthesis of polar sea ice microalgae. *J. Phycol.* 41:732-41.
1046 McNeill, J., Barrie, F. R., Buck, W. R., Demoulin, V., Greuter, W., Hawksworth, D. L., Herendeen,
1047 P. S., Knapp, S., Marhold, K., Prado, J., Prud'homme van Reine, W. F., Smith, G. F.,
1048 Wiersema, J. H. & Turland, N. J. 2012. *International Code of Nomenclature for Algae,*
1049 *Fungi, and Plants (Melbourne Code) adopted by the Eighteenth International Botanical*
1050 *Congress Melbourne, Australia, July 2011.* Koeltz Scientific Books, Königstein. 208 pp.
1051 Merget, B., Koetschan, C., Hackl, T., Forster, F., Dandekar, T., Muller, T., Schultz, J. & Wolf, M.
1052 2012. The ITS2 Database.*: JoVE* 61:e3806.
1053 Mikheyev, A. S. & Tin, M. M. Y. 2014. A first look at the Oxford Nanopore MinION sequencer.
1054 *Mol. Ecol. Resour.* 14:1097-102.
1055 Moniz, M. B. J. & Kaczmarska, I. 2010. Barcoding of Diatoms: Nuclear encoded ITS revisited.
1056 *Protist* 161:7-34.
1057 Montresor, M., Lovejoy, C., Orsini, L., Procaccini, G. & Roy, S. 2003. Bipolar distribution of the
1058 cyst-forming dinoflagellate *Polarella glacialis*. *Polar Biol.* 26:186-94.
1059 Nanjappa, D., Kooistra, W. H. C. F. & Zingone, A. 2013. A reappraisal of the genus *Leptocylindrus*
1060 (Bacillariophyta), with the addition of three species and the erection of *Tenuicylindrus* gen.
1061 nov. *J. Phycol.* 49:917-36.
1062 Nei, M. & Kumar, S. 2000. *Molecular Evolution and Phylogenetics.* Oxford University Press, New
1063 York, 352 pp.
1064 Olli, K., Riser, C. W., Wassmann, P., Ratkova, T., Arashkevich, E. & Pasternak, A. 2002. Seasonal
1065 variation in vertical flux of biogenic matter in the marginal ice zone and the central Barents
1066 Sea. *J. Mar. Syst.* 38:189-204.

1067 Orive, E., Perez-Aicua, L., David, H., Garcia-Etxebarria, K., Laza-Martinez, A., Seoane, S. &
1068 Miguel, I. 2013. The genus *Pseudo-nitzschia* (Bacillariophyceae) in a temperate estuary with
1069 description of two new species: *Pseudo-nitzschia plurisecta* sp. nov. and *Pseudo-nitzschia*
1070 *abrensis* sp. nov. *J. Phycol.* 49:1192-206.
1071 Orsini, L., Sarno, D., Procaccini, G., Poletti, R., Dahlmann, J. & Montresor, M. 2002. Toxic
1072 *Pseudo-nitzschia multistriata* (Bacillariophyceae) from the Gulf of Naples: morphology,
1073 toxin analysis and phylogenetic relationships with other *Pseudo-nitzschia* species. *Eur. J.*
1074 *Phycol.* 37:247-57.
1075 Pawlowski, J., Majewski, W., Longet, D., Guiard, J., Cedhagen, T., Gooday, A. J., Korsun, S.,
1076 Habura, A. A. & Bowser, S. S. 2008. Genetic differentiation between Arctic and Antarctic
1077 monothalamous foraminiferans. *Polar Biol.* 31:1205-16.
1078 Percopo, I., Siano, R., Cerino, F., Sarno, D. & Zingone, A. 2011. Phytoplankton diversity during the
1079 spring bloom in the northwestern Mediterranean Sea. *Bot. Mar.* 54:243-67.
1080 Percopo, I., Ruggiero, M. V., Balzano, S., Gourvil, P., Lundholm, N., Siano, R., Vaulot, D. &
1081 Sarno, D. 2016. *P. arctica* sp. nov., a new cold-water cryptic *Pseudo-nitzschia* species
1082 within the *P. pseudodelicatissima* complex. *J. Phycol.* 52:184-99.
1083 Pickart, R. S., Schulze, L. M., Moore, G. W. K., Charette, M. A., Arrigo, K. R., van Dijken, G. &
1084 Danielson, S. L. 2013. Long-term trends of upwelling and impacts on primary productivity
1085 in the Alaskan Beaufort Sea. *Deep-Sea Res. Pt I* 79:106-21.
1086 Ratkova, T. N. & Wassmann, P. 2002. Seasonal variation and spatial distribution of phyto- and
1087 protozooplankton in the central Barents Sea. *J. Mar. Syst.* 38:47-75.
1088 Rimet, F., Kermarrec, L., Bouchez, A., Hoffmann, L., Ector, L. & Medlin, L. K. 2011. Molecular
1089 phylogeny of the family Bacillariaceae based on 18S rDNA sequences: focus on freshwater
1090 *Nitzschia* of the section Lanceolatae. *Diatom Res.* 26:273-91.
1091 Rines, J. E. B. & Hargraves, P. E. 1988. *The Chaetoceros Ehrenberg (Bacillariophyceae). Flora of*
1092 *Narraganset Bay, Rhode Island, U.S.A.* Lubrecht & Cramer Ltd, Berlin. 196 pp.
1093 Sar, E. A., Sunesen, I. s., Lavigne, A. S. & Lofeudo, S. 2011. *Thalassiosira rotula*, a heterotypic
1094 synonym of *Thalassiosira gravida*: morphological evidence. *Diatom Res.* 26:109-19.
1095 Sarno, D., Kooistra, W., Medlin, L. K., Percopo, I. & Zingone, A. 2005. Diversity in the genus
1096 *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of *S. costatum*-like
1097 species with the description of four new species. *J. Phycol.* 41:151-76.
1098 Sarno, D., Kooistra, W., Balzano, S., Hargraves, P. E. & Zingone, A. 2007. Diversity in the genus
1099 *Skeletonema* (Bacillariophyceae): III. Phylogenetic position and morphological variability of
1100 *Skeletonema costatum* and *Skeletonema grevillei*, with the description of *Skeletonema*
1101 *ardens* sp. nov. *J. Phycol.* 43:156-70.
1102 Schloss, P. D., Jenior, M. L., Koumpouras, C. C., Westcott, S. L. & Highlander, S. K. 2016.
1103 Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system.
1104 *Peer J.* 4:e1869.
1105 Schütt, F. 1895. Arten von *Chaetoceras* und *Peragallia*. *Ein Beitrag zur Hochseeflora. Berichte der*
1106 *Deutsche Botanisch Gesellschaft* 13:35-50.
1107 Sherr, E. B., Sherr, B. F., Wheeler, P. A. & Thompson, K. 2003. Temporal and spatial variation in
1108 stocks of autotrophic and heterotrophic microbes in the upper water column of the central
1109 Arctic Ocean. *Deep-Sea Res. Pt I* 50:557-71.
1110 Sorhannus, U. & Fox, M. G. 2012. Phylogenetic analyses of a combined data set suggest that the
1111 *Attheya* lineage is the closest living relative of the pennate diatoms (Bacillariophyceae).
1112 *Protist* 163:252-62.
1113 Sorhannus, U., Ortiz, J. D., Wolf, M. & Fox, M. G. 2010. Microevolution and speciation in
1114 *Thalassiosira weissflogii* (Bacillariophyta). *Protist* 161:237-49.
1115 Souffreau, C., Vanormelingen, P., Van de Vijver, B., Isheva, T., Verleyen, E., Sabbe, K. &
1116 Vyverman, W. 2013. Molecular evidence for distinct Antarctic lineages in the cosmopolitan
1117 terrestrial diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. *Protist* 164:101-15.

1118 Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W. & Richards, T. A.
1119 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly
1120 complex eukaryotic community in marine anoxic water. *Mol. Ecol.* 19:21-31.
1121 Stonik, I. V., Orlova, T. Y. & Crawford, R. M. 2006. *Attheya ussurensis* sp. nov. (Bacillariophyta) -
1122 a new marine diatom from the coastal waters of the Sea of Japan and a reappraisal of the
1123 genus. *Phycologia* 45:141-147.
1124 Sukhanova, I. N., Flint, M. V., Pautova, L. A., Stockwell, D. A., Grebmeier, J. M. & Sergeeva, V.
1125 M. 2009. Phytoplankton of the western Arctic in the spring and summer of 2002: Structure
1126 and seasonal changes. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.* 56:1223-36.
1127 Syvertsen, E. E. 1977. *Thalassiosira rotula* and *T. gravida*: ecology and morphology. *Beiheft zur*
1128 *Nova Hedwigia* 54:99-112.
1129 Syvertsen, E. E. 1986. *Thalassiosira hispida* sp. nov., a marine planktonic diatom. In: M. Ricard
1130 [Ed.] *Proceedings of the Eighth International Diatom Symposium*, Koeltz, Koenigstein, pp.
1131 33-42.
1132 Syvertsen, E. E. & Hasle, G. R. 1983. The diatom genus *Eucampia*: morphology and taxonomy.
1133 *Bacillaria* 6:169-210.Tamura, K. & Nei, M. 1993. Estimation of the number of nucleotide
1134 substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol.*
1135 *Biol. Evol.* 10:512-26.
1136 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: molecular
1137 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
1138 maximum parsimony methods. *Mol. Biol. Evol.* 28:2731-39.
1139 Terrado, R., Scarcella, K., Thaler, M., Vincent, W. F. & Lovejoy, C. 2013. Small phytoplankton in
1140 Arctic seas: vulnerability to climate change. *Biodiversity* 14:2-18.
1141 Tschochner, H. & Hurt, E. 2003. Pre-ribosomes on the road from the nucleolus to the cytoplasm.
1142 *Trends Cell Biol.* 13:255-63.
1143 Tuschling, K., von Juterzenka, K., Okolodkov, Y. B. & Anoshkin, A. 2000. Composition and
1144 distribution of the pelagic and sympagic algal assemblages in the Laptev Sea during
1145 autumnal freeze-up. *J. Plankton Res.* 22:843-64.
1146 von Quillfeldt, C. H. 2000. Common diatom species in arctic spring blooms: Their distribution and
1147 abundance. *Bot. Mar.* 43:499-516.
1148 von Quillfeldt, C. H., Ambrose, W. G. & Clough, L. M. 2003. High number of diatom species in
1149 first-year ice from the Chukchi Sea. *Polar Biol.* 26:806-18.
1150 Wang, J., Cota, G. F. & Comiso, J. C. 2005. Phytoplankton in the Beaufort and Chukchi Seas:
1151 Distribution, dynamics, and environmental forcing. *Deep-Sea Res. Part II-Top. Stud.*
1152 *Oceanogr.* 52:3355-68.
1153 Whittaker, K. A., Rignanese, D. R., Olson, R. J. & Rynearson, T. A. 2012. Molecular subdivision of
1154 the marine diatom *Thalassiosira rotula* in relation to geographic distribution, genome size,
1155 and physiology. *Bmc Evolutionary Biology* 12. doi**:** 10.1186/1471-2148-12-209
1156 Wolf, M., Achtziger, M., Schultz, J., Dandekar, T. & Muller, T. 2005. Homology modeling
1157 revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures.
1158 *Rna-a Publication of the Rna Society* 11:1616-23.
1159 Zhu, F., Massana, R., Not, F., Marie, D. & Vaulot, D. 2005. Mapping of picoeucaryotes in marine
1160 ecosystems with quantitative PCR of the 18S rRNA gene. *Fems Microbiol. Ecol.* 52:79-92.

Table 1. List of the strain isolated during the MALINA cruise and used in the present study.

Family	Species	Strain code[1]	Isolation site[2]		Morphology[3]	Genbank accession numbers[4]		
			Station	Depth (m)		18S	28S	ITS
Bacillariaceae	*Cylindrotheca closterium*	RCC1985	280	30	LM, TEM	JF794039	JQ995403	
	Nitzschia pellucida	RCC2276	BEA130709A	0	LM,TEM	JF794052	JQ995450	
	Pseudo-nitzschia granii	RCC2006	PAC080709A	5	LM, TEM		JQ995420	
	Pseudo-nitzschia granii	RCC2008	PAC080709A	5	LM	JN934671	JQ995421	
	Pseudo-nitzschia granii	RCC2273	PAC060709A	0			JQ995391	
	Pseudo-nitzschia arctica	RCC2002	690	29	LM, TEM, SEM		JQ995416	
	Pseudo-nitzschia arctica	RCC2004	690	29	LM, TEM, SEM	JF794046	JQ995418	
	Pseudo-nitzschia arctica	RCC2005	690	29	LM, TEM		JQ995419	
	Pseudo-nitzschia arctica	RCC2517	690	29	LM, TEM, SEM		JQ995461	
Fragilariaceae	*Synedropsis hyperborea*	RCC2043	280	30	LM, TEM, SEM	JF794051	JQ995434	
	Synedropsis hyperborea	RCC2520	280	30	LM		JQ995463	
Attheyaceae	*Attheya septentrionalis*	RCC1986	280	30	LM, TEM, SEM	JF794040	JQ995404	
	Attheya septentrionalis	RCC1988	280	30	LM		JQ995405	
	Attheya septentrionalis	RCC2042	680	3	LM	JN934675	JQ995433	
Thalassiosiraceae	*Thalassiosira gravida*	RCC1984	280	30	LM	JN934669	JQ995402	
	Thalassiosira gravida	RCC1999	280	30	LM, TEM, SEM		JQ995414	
	Thalassiosira cf. *hispida*	RCC2521	680	40	TEM, SEM	JN934691	JQ995464	
	Thalassiosira minima	RCC2265	394	3	LM, TEM, SEM	JN934676	JQ995440	
	Thalassiosira minima	RCC2266	394	3	LM, TEM, SEM		JQ995441	
	Thalassiosira minima	RCC2269	PAC050709A	0	LM, TEM, SEM		JQ995444	
	Thalassiosira nordenskioeldii	RCC2000	690	29	LM, TEM, SEM	JF794045	JQ995415	
	Thalassiosira nordenskioeldii	RCC2021	680	3	LM, TEM, SEM		JQ995428	
	Thalassiosira nordenskioeldii	RCC2522	620	3	LM			
	Porosira glacialis	RCC1995	690	29	LM, SEM, TEM			
	Porosira glacialis	RCC2039	690	29	LM	JN934673	JQ995432	
	Shionodiscus bioculatus	RCC1991	620	65	LM, TEM, SEM	JF794041	JQ995408	
Cymatosiraceae	*Arcocellulus cornucervis*	RCC2270	ARC120709A	0	LM, SEM,TEM	JN934677	JQ995445	

Hemiaulaceae	*Eucampia groenlandica*	RCC1996	690	29	LM, TEM, SEM	JF794043	JQ995412	
	Eucampia groenlandica	RCC2037	690	29	LM, SEM		JQ995430	
	Eucampia groenlandica	RCC2038	690	29	LM, SEM		JQ995431	
Chaetoceraceae	*Chaetoceros decipiens*	RCC1997	690	29	LM, TEM, SEM	JF794044	JQ995413	
	Chaetoceros gelidus	RCC1990	620	65	LM		JQ995407	
	Chaetoceros gelidus	RCC1992	620	65	LM, TEM, SEM	JF794042	JQ995409	
	Chaetoceros gelidus	RCC1994	690	29	LM, SEM		JQ995411	
	Chaetoceros gelidus	RCC2046	280	30	LM		JQ995435	
	Chaetoceros gelidus	RCC2271	690	3	LM,SEM		JQ995446	
	Chaetoceros gelidus	MALINA E65 PG4	690	29			JQ995393	
	Chaetoceros gelidus	MALINA E65 PG18	690	29				
	Chaetoceros gelidus	MALINA S135	BEA140709A	0			JQ995396	
	Chaetoceros neogracilis clade I	RCC2003	690	29	LM		JQ995417	KT860511
	Chaetoceros neogracilis clade I	RCC2011	620	3	LM		JQ995423	KT860513
	Chaetoceros neogracilis clade I	RCC2017	760	3	LM, TEM		JQ995427	KT860517
	Chaetoceros neogracilis clade I	RCC2262	460	3	LM		JQ995437	KT860520
	Chaetoceros neogracilis clade I	RCC2263	235	3	LM		JQ995438	KT860521
	Chaetoceros neogracilis clade I	RCC2264	235	3	LM		JQ995439	KT860522
	Chaetoceros neogracilis clade I	RCC2267	394	3			JQ995442	KT860523
	Chaetoceros neogracilis clade I	RCC2274	620	3	LM		JQ995448	KT860526
	Chaetoceros neogracilis clade I	RCC2275	620	3	LM		JQ995449	KT860527
	Chaetoceros neogracilis clade I	RCC2278	320	3	LM		JQ995452	KT860529
	Chaetoceros neogracilis clade I	RCC2279	320	3	LM		JQ995453	KT860530
	Chaetoceros neogracilis clade I	RCC2280	760	3	LM		JQ995454	KT860531
	Chaetoceros neogracilis clade I	RCC2281	760	3	LM		JQ995455	KT860532
	Chaetoceros neogracilis clade I	RCC2507	235	25	LM		JQ995459	KT860536
	Chaetoceros neogracilis clade I	MALINA S441 P21-E6	320	3			JQ995397	KT860541
	Chaetoceros neogracilis clade I	MALINA S502 P27.B3	760	3			JQ995399	KT860540
	Chaetoceros neogracilis clade I	MALINA S509	760	3			KT884482	KT884482
	Chaetoceros neogracilis clade I	MALINA S510	760	3			KT884483	KT884483

Chaetoceros neogracilis clade I	MALINA S511	760	3			KT884484	KT884484
Chaetoceros neogracilis clade I	MALINA S512	760	3			KT884485	KT884485
Chaetoceros neogracilis clade II	RCC2261	460	3	LM		JQ995436	KT860519
Chaetoceros neogracilis clade II	RCC2268	BEA130709A	0	LM		JQ995443	KT860524
Chaetoceros neogracilis clade II	RCC2272	BEA130709A	0	LM,SEM		JQ995447	KT860525
Chaetoceros neogracilis clade II	RCC2277	BEA130709A	0	LM		JQ995451	KT860528
Chaetoceros neogracilis clade II	RCC2282	760	3	LM		JQ995456	KT860533
Chaetoceros neogracilis clade II	RCC2318	620	65	LM	JN934684	JQ995457	KT860534
Chaetoceros neogracilis clade II	RCC2506	235	3	LM		JQ995458	KT860535
Chaetoceros neogracilis clade II	MALINA E43.N2	BEA140709A	0			JQ995392	
Chaetoceros neogracilis clade III	RCC1989	620	65	LM		JQ995406	KT860509
Chaetoceros neogracilis clade III	RCC1993	620	65	LM		JQ995410	KT860510
Chaetoceros neogracilis clade IV	RCC2010	620	3	LM, SEM		JQ995422	KT860512
Chaetoceros neogracilis clade IV	RCC2012	110	3	LM, SEM,TEM		JQ995424	KT860514
Chaetoceros neogracilis clade IV	RCC2014	110	3	LM, TEM		JQ995425	KT860515
Chaetoceros neogracilis clade IV	RCC2016	760	3	LM, SEM	JF794049	JQ995426	KT860516
Chaetoceros neogracilis clade IV	RCC2022	680	3	LM, SEM		JQ995429	KT860518
Chaetoceros neogracilis clade IV	MALINA FT56.6 PG6	110	3			JQ995395	KT860542

[1] RCC: Roscoff culture collection. More information on the strains is available at http://roscoff-culture-collection.org/. Strains without an RCC code are no longer available.

[2] Sampling location of the MALINA cruise. See Table S1 for more details

[3] Technique used for the morphological identification: LM, Light Microscopy; TEM, Transmission Electron Microscopy; SEM, Scanning Electron Microscopy.

[4] Please note that the V4 region of the 18S rRNA gene has been sequenced from all the strains.

Table 2. Geographic distribution and morphological references of the species identified in the present study.

Species	Morphological references	Global distribution	Distribution in Arctic waters
Cylindrotheca closterium (Ehrenberg) Lewin & Reimann	Hasle & Syvertsen 1997, and references therein Jahn & Kusber 2005	Cosmopolitan (Hasle & Syvertsen 1997) Common in Arctic waters	Beaufort Sea (Horner & Schrader 1982) Chukchi Sea (von Quillfeldt 2000) White and Barents Sea (Luddington et al. 2016) Laptev Sea (Tuschling et al. 2000) Central Arctic Ocean (Katsuki et al. 2009)
Nitzschia pellucida Grunow	Bérard-Therriault et al. 1999, and references therein	Northern cold water region (Bérard-Therriault et al. 1999) Antarctica (Hällfors 2004) European freshwater environments (Cărăus 2012) Ishigaki Island, Japan (Lundholm et al. 2002)	Chukchi Sea (von Quillfeldt 2000)
Pseudo-nitzschia granii (Hasle) Hasle	Hasle & Syvertsen 1997, and references therein Marchetti et al. 2008	Northern cold water region to temperate? (Hasle & Syvertsen 1997) Northern Atlantic (Hasle 1964) Subarctic Pacific (Marchetti et al. 2008, this study)	Norwegian Sea (Hasle 1964) Chukchi Sea (von Quillfeldt et al. 2003) White and Barents Seas (Luddington et al. 2016)
Pseudo-nitzschia arctica Percopo & Sarno	Percopo et al. 2016	Recently described from Arctic waters (Percopo et al. 2016)	Beaufort Sea, Barrow Strait, Baffin Bay (Percopo et al. 2016)
Synedropsis hyperborea (Grunow) Hasle, Medlin & Syvertsen	Hasle et al. 1994	Northern cold water region (Hasle & Syvertsen 1997) Common in Arctic waters (Hasle et al. 1994)	Frobisher Bay, Greenland, Barents Sea (Hasle et al. 1994) Chukchi Sea (von Quillfeldt et al. 2003)
Attheya septentrionalis (Østrup) Crawford	Crawford et al. 1994 Stonik et al. 2006	Northern cold water region to temperate (Hasle & Syvertsen 1997) Common in Arctic waters	Nansen Basin (Gosselin et al. 1997) Chukchi Sea (von Quillfeldt et al. 2003)

			Baffin Bay (Caron et al. 2004) White and Barents Sea (Luddington et al. 2016) Laptev Sea (Tuschling et al. 2000)
Thalassiosira gravida Cleve	Syvertsen 1977	Northern and southern cold water regions (Whittaker et al. 2012) Common in Arctic waters	Nansen Basin, Chukchi Sea (Gosselin et al. 1997) Baffin Bay (Lovejoy et al. 2002) Laptev Sea (Tuschling et al. 2000) Central Arctic Ocean (Katsuki et al. 2009)
Thalassiosira cf. *hispida* Syvertsen	Syvertsen 1986	Northern cold water region to temperate (Hasle & Syvertsen 1997)	Amundsen Gulf (Luddington et al. 2016) Central Arctic Ocean (Katsuki et al. 2009) Svalbard and the Barents Sea (von Quillfeldt 2000) Chukchi Sea (von Quillfeldt et al. 2003)
Thalassiosira minima Gaarder	Hasle & Syvertsen 1997, and references therein Hoppenrath et al. 2007	Cosmopolitan excluding polar regions (Hasle & Syvertsen 1997) North Sea (Hoppenrath et al. 2007) North Atlantic Ocean (Luddington et al. 2016, as *T.* aff. *minima*) Warm waters in coastal and estuarine systems (Guinder et al. 2012)	First report in this study
Thalassiosira nordenskioeldii Cleve	Hasle & Syvertsen 1997, and references therein	Northern cold water region to temperate (Hasle & Syvertsen 1997) Common in Arctic waters	Amundsen Gulf (Luddington et al. 2016) Canadian Arctic (Aizawa et al. 2005) Baffin Bay (Caron et al. 2004) Barents Sea (Degerlund & Eilertsen 2010) Laptev Sea (Tuschling et al. 2000) Chukchi Sea (von Quillfeldt et al.

			2003) Central Arctic Ocean (Katsuki et al. 2009)
Porosira glacialis (Grunow) Jørgensen	Hasle & Syvertsen 1997, and references therein	Northern cold water region to temperate, southern cold water region (Hasle & Syvertsen 1997) Common in Arctic waters	Amundsen Gulf (Luddington et al. 2016) Chukchi Sea (Gosselin et al. 1997) Beaufort Sea (Sukhanova et al. 2009) White and Barents Seas (Olli et al. 2002) Central Arctic Ocean (Katsuki et al. 2009)
Shionodiscus bioculatus (Grunow) Alverson, Kang & Theriot	as *Thalassiosira bioculata*: Hasle & Syvertsen 1997, and references therein Bérard-Therriault et al. 1999, and references therein	Northern cold water region (Hasle & Syvertsen 1997) Common in Arctic waters	Amundsen Gulf (Luddington et al. 2016) Chukchi Sea (von Quillfeldt et al. 2003) White and Barents Sea (Luddington et al. 2016) Norwegian coastal waters (Degerlund & Eilertsen 2010) Baffin Bay (Booth et al. 2002) Central Arctic Ocean (Katsuki et al. 2009)
Arcocellulus cornucervis Hasle, von Stosch & Syvertsen	Hasle et al. 1983	Northern cold and temperate waters, New Zealand (Hasle & Syvertsen 1997) Mediterranean Sea (Percopo et al. 2011)	Baffin Bay (Lovejoy et al. 2002)
Eucampia groenlandica Cleve	Syvertsen & Hasle 1983	Northern cold water region (Hasle & Syvertsen 1997) Common in Arctic waters	Baffin Bay (Cleve 1896) Laptev Sea (Tuschling et al. 2000) Barents Sea (Luddington et al. 2016)
Chaetoceros decipiens Cleve	Hasle & Syvertsen 1997, and references therein Jensen & Moestrup 1998	Cosmopolitan (Hasle & Syvertsen 1997) Common in Arctic waters	North Pacific and Bering Sea (Aizawa et al. 2005) Baffin Bay (Caron et al. 2004) Barents Sea (Ratkova & Wassmann 2002)

			Norwegian coastal waters (Degerlund & Eilertsen 2010)
Chaetoceros gelidus Chamnansinp, Li, Lundholm & Moestrup	Chamnansinp et al. 2013	Northern cold water region (Chamnansinp et al. 2013)	Barents Sea, Norwegian Sea, Greenland (Chamnansinp et al. 2013)
Chaetoceros neogracilis (Schütt) VanLandingham	Schütt 1895 (as *C. gracile*) See discussion	Baltic Sea (Hällfors 2004, Majaneva et al. 2012)	Svalbard (Choi et al. 2008) Beaufort Sea (Lovejoy & Potvin 2011)

FIGURE LEGEND

Figure 1. Full 18S rRNA phylogenetic tree derived from Maximum Likelihood (ML) analysis. The tree includes at least one sequence from each genotype found within the diatom strains isolated during the MALINA cruise. Four sequences from radial centrics (*Corethron hystrix*, *Corethron pennatum*, *Rhizosolenia setigera*, and *Rhizosolenia similoides*) have been used as outgroup. The MALINA strains sequenced here are labelled in bold whereas other strains isolated from Arctic waters are underlined. The Genbank accession number is indicated next to the strain code. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches from left (ML) to right (Neighbour-joining). Missing percentage values and "_" indicate that bootstrap values < 50 % were obtained for the corresponding node. Asterisks indicate strains isolated from the North Pacific Ocean.

Figure 2. 28S rRNA phylogenetic tree inferred by maximum likelihood (ML) analysis for the (**A)** pennate and (**B**) centric diatoms isolated during the MALINA cruise. The MALINA strains sequenced here are labelled in bold whereas other strains isolated from Arctic waters are underlined. The evolutionary histories were inferred using maximum likelihood. The percentage of trees in which the associated taxa cluster together is shown next to the branches based on Maximum Likelihood (left) and Neighbour joining (right). ML and NJ values are indicated next to the branch nodes as described in Fig. 1. Asterisks indicate strains isolated from the North Pacific Ocean.

Figure 3. 28S rRNA (**A**), ITS-1 (**B**), and 5.8S+ITS-2 (**C**) phylogenetic trees for the strains of *Chaetoceros neogracilis* strains isolated from the Beaufort Sea. For the 28S, *C. gelidus* was used to root the phylogenetic tree whereas for the ITS-1 and 5.8S + ITS-2 trees, the Antarctic strains of *Chaetoceros* sp. (CCMP187, CCMP189, CCMP190) were used as outgroup. The bootstrap values are indicated next to the branches as for Figure 1.

Figure 4. (A) *Cylindrotheca closterium*: TEM micrograph, RCC1985. Detail of the valve in which is visible the raphe interruption. Scale bar, 2 µm. (B-D) *Nitzschia pellucida*: (B) TEM micrograph, RCC2276. Whole valve. Scale bar, 5 µm. (C) TEM micrograph, RCC2276. Detail of the valve. Note the central larger interspace. Scale bar, 2 µm. (D) TEM micrograph, RCC2276. Detail of cell apex. Scale bar, 2 µm. (E-G) *Pseudo-nitzschia granii*: (E) TEM micrograph, RCC2006. Whole valve. Scale bar, 5 µm. (F) TEM micrograph, RCC2006. Detail of the valve. Note the few incomplete poroids (arrows). Scale bar, 0.5 µm. (G) TEM micrograph, RCC2006. Detail of the valve with scattered complete poroids. Scale bar, 1 µm. (H-I) *Pseudo-nitzschia arctica*: (H) LM micrograph, RCC2002. A colony of two cells in girdle view. Scale bar, 20 µm. (I) TEM micrograph, RCC2004. Detail of the valve. Note the central larger interspace. Scale bar, 1µm. (J-N) *Synedropsis hyperborea*: (J) LM micrograph, RCC2043. Cell in valve view. Scale bar, 2µm. (K) SEM micrograph, RCC2043. External view of the central part of the valve. Scale bar, 1 µm. (L) SEM micrograph, RCC2043. Internal view of the apex. Note apical slit field and rimoportula. Scale bar, 0.1 µm. (M) SEM micrograph, RCC2043. External view of the apex. Note apical slit field and absence of rimoportula. Scale bar, 0.5 µm. (N) SEM micrograph, RCC2043. External view of the apex. Note apical slit field and rimoportula. Scale bar, 0.2 µm.

Figure 5. (A-E) *Attheya septentrionalis*: (A) LM micrograph, RCC1986. Cells in girdle view. Scale bar, 10 µm. (B) LM micrograph, RCC2042. A cell in girdle view. Scale bar, 20 µm. (C) TEM micrograph, RCC1986. A circular valve. Scale bar, 1 µm. (D) TEM micrograph, RCC2042. A horn with three longitudinal strips. Scale bar, 0.5 µm. (E) TEM micrograph, RCC1986. A horn with four longitudinal strips. Scale bar, 0.5 µm. (F-G) *Thalassiosira gravida*: (F) LM micrograph, RCC1999. Three cells joined in colony. Scale bar, 20 µm. (G) TEM micrograph, RCC1999. A valve with the central cluster of fultoportulae and several fultoportulae scattered on the valve face. Scale bar, 5 µm. (H-K) *Thalassiosira* cf. *hispida*: (H) SEM micrograph, RCC2521. A cell in valve view with a ring of marginal fultoportulae and one central fultoportula. Note the rimoportula between two marginal fultoportulae. Scale bar, 1 µm. (I) TEM micrograph, RCC2521. Detail of a valve; short and minute spines are present on the hyaline margin and in the areolae foramina. Scale bar, 1 µm. (J) SEM micrograph, RCC2521. The girdle composed by the valvocopula, a copula and open bands. Scale bar, 1 µm. (K) TEM micrograph, RCC2521. Detail of the fultoportula. Scale bar, 0.2 µm. (L-P) *Thalassiosira minima*: (L) LM micrograph, RCC2269. Cell in girdle view with two chloroplasts. Scale bar, 2 µm. (M) SEM micrograph, RCC2266. External view of the valve with a ring of marginal fultoportulae and two central fultoportulae. Note the rimoportula between two marginal fultoportulae. Scale bar, 1 µm. (N) SEM micrograph, RCC2269. Internal view of a valve. Scale bar, 1 µm. (O) TEM micrograph, RCC2266. Five central fultoportulae. Scale bar, 1 µm. (P) SEM micrograph,

RCC2266. Detail of two marginal fultoportulae with the small external labiate-shaped protrusions on the external face of the valve. Scale bar, 1 µm.

Figure 6. (A-C) *Thalassiosira nordenskioeldii*: (A) LM micrograph, RCC2000. A colony in girdle view. Scale bar, 5 µm. (B) TEM micrograph, RCC2000. A valve with a marginal ring of fultoportulae, one central fultoportula and one rimoportula positioned within two marginal fultoportulae. Scale bar, 2 µm. (C) SEM micrograph, RCC2000. A cell with ring of fultoportulae with long external tubes bearing a terminal collar. Scale bar, 1 µm. (D-F) *Porosira glacialis*: (D) LM micrograph, RCC1995. Cell in girdle view. Scale bar, 10 µm. (E) LM micrograph, RCC1995. Cell in valve view. Scale bar, 10 µm. (F) TEM micrograph, RCC1995. A valve with numerous fultoportulae scattered over the valve surface. Note the central annulus and the marginal rimoportula (arrow). Scale bar, 10 µm. (G-I) *Shionodiscus bioculatus*: (G) LM micrograph, RCC1991. A cell in girdle view. Scale bar, 20 µm. (H) SEM micrograph, RCC1991. External view of a cell; note the marginal ring of fultoportulae, the single fultoportula in the valve centre and a subcentral rimoportula. Scale bar, 10 µm. (I) TEM micrograph, RCC1991. Whole valve. Scale bar, 10 µm. (J-L) *Arcocellulus cornucervis*: (J) SEM micrograph, RCC2270. A slightly curved cell in girdle view. Note the conspicuous branches of the pili (arrow). Scale bar, 1 µm. (K) SEM micrograph, RCC2270. A pili valve (left) and a process valve (right). Note the ocelluli (arrows). Scale bar, 1 µm. (L) A pili valve in which the short spinules are visible near the pilus base (arrows). Scale bar, 2 µm.

Figure 7. (A) *Arcocellulus cornucervis*: TEM micrograph, RCC2270. A process valve in which the two ocelluli are visible (arrows). Scale bar, 1 µm. (B-C) *Eucampia groenlandica*: (B) LM micrograph, RCC2037. Part of a colony. Scale bar, 5 µm (C) SEM micrograph, RCC2037. A valve with the central rimoportula (arrow). Scale bar, 1 µm. (D-I) *Chaetoceros decipiens*: (D) LM micrograph, RCC1997. Part of a colony. Scale bar, 20µm. (E) LM micrograph, RCC1997. A solitary cell. Scale bar, 20 µm. (F) TEM micrograph, RCC1997. A terminal valve. Scale bar, 5 µm. (G) SEM micrograph, RCC1997. Two intercalary valves. Scale bar, 1 µm. (H) TEM micrograph, RCC1997. A terminal valve. Note the central process. Scale bar, 5 µm. (I) TEM micrograph, RCC1997. A girdle band with parallel costae and small poroids. Scale bar, 5 µm. (J-L) *Chaetoceros gelidus*: (J) LM micrograph, RCC2271. A curved chain. Note the two straight setae on the upper part of the picture. Scale bar, 20 µm. (K) LM micrograph, RCC2271. A spherical colony. Scale bar, 50 µm. (L) SEM micrograph, RCC2271. Two intercalary valves with the narrow aperture. Scale bar, 1 µm.

Figure 8. (A-C) *Chaetoceros gelidus*: (A) TEM micrograph, RCC2271. Intercalary valve. Scale bar, 1 µm. (B) SEM micrograph, RCC2271. Detail of the two types of setae, the short (on the upper part of the picture) and the straight long seta (crossing the picture). Note the absence of spines in a large part of the long seta. Scale bar, 5 µm. (C) SEM micrograph, RCC2271. A spore. Scale bar, 1 µm. (D-N) *Chaetoceros neogracilis*: (D) LM micrograph, RCC2272. A solitary cell. Scale bar, 10 µm. (E) LM micrograph, RCC2017. A solitary cell. Scale bar, 10 µm. (F) LM micrograph, RCC2016. A solitary cell. Scale bar, 10 µm. (G) LM micrograph, RCC1989. A colony of four cells. Scale bar, 5 µm. (H) SEM micrograph, RCC2012. Detail of a colony. Note quite narrow apertures. Scale bar, 1 µm. (I) SEM micrograph, RCC2012. Terminal valve with

the external tube. Scale bar, 2 µm. (J) TEM micrograph, RCC2012. Terminal valve with the central slit-like process. Scale bar, 2 µm. (K) TEM micrograph, RCC2012. Intercalary valve. Scale bar, 2 µm. (L) TEM micrograph, RCC2271, Intercalary valve. Scale bar, 2 µm. (M) SEM micrograph, RCC2271. Setae with arrowhead-shaped spines. Scale bar, 1 µm. (N) SEM micrograph, RCC2271. Detail of a seta with spines and long spiral costae interconnected by short transverse costae. Scale bar, 0.5 µm.

Figure 9. FIG. 9. Diagrams of the secondary structure of the ITS-2 transcripts of Chaetoceros neogracilis Clade I RCC2279. The boxes indicate the structural variations found in C. neogracilis Clade I with respect to the other clades. Nucleotides which differ between C. neogracilis Clade I and the other three clades are marked with black background.

Supporting information

Table S1. Details of the strain isolated during the MALINA cruise and used in the present study. Most strains are available at Roscoff Culture Collection (RCC)

Table S2. List of the strains and species from which the sequences were used in the present study for the phylogenetic trees. Most strains are currently available at different institutions or culture collections. CCMP: National Centre for Marine Algae and Microbiota (ncma.bigelow.org), UNC: Culture Collection at University of North Carolina (www.unc.edu/), NIOZ: Culture Collection at Netherland Institute for Sea Research (www.nioz.nl), UTEX: Culture Collection of Algae at University of Texas Austin (utex.org/), CCAP: Culture Collection of Algae and Protozoa (www.ccap.ac.uk), TCC: Thonon Culture Collection (www6.inra.fr/carrtel-collection_eng), CS: Australia National Algae Culture Collection (www.csiro.au/en/Research/Collections/ANACC/About-our-collection), SZN: Stazione Zoologica Anton Dohrn, Naples (www.szn.it), RCC: Roscoff Culture Collection (http://roscoff-culture-collection.org).

Figure S1. Phylogenetic tree of the ITS operon of the *Chaetoceros* sp. strains isolated in the present study. The Antarctic strains of *Chaetoceros* sp. (CCMP187, CCMP189, CCMP190) were used as outgroup. The bootstrap values are indicated next to the branches as for Figure 6.

Figure 1

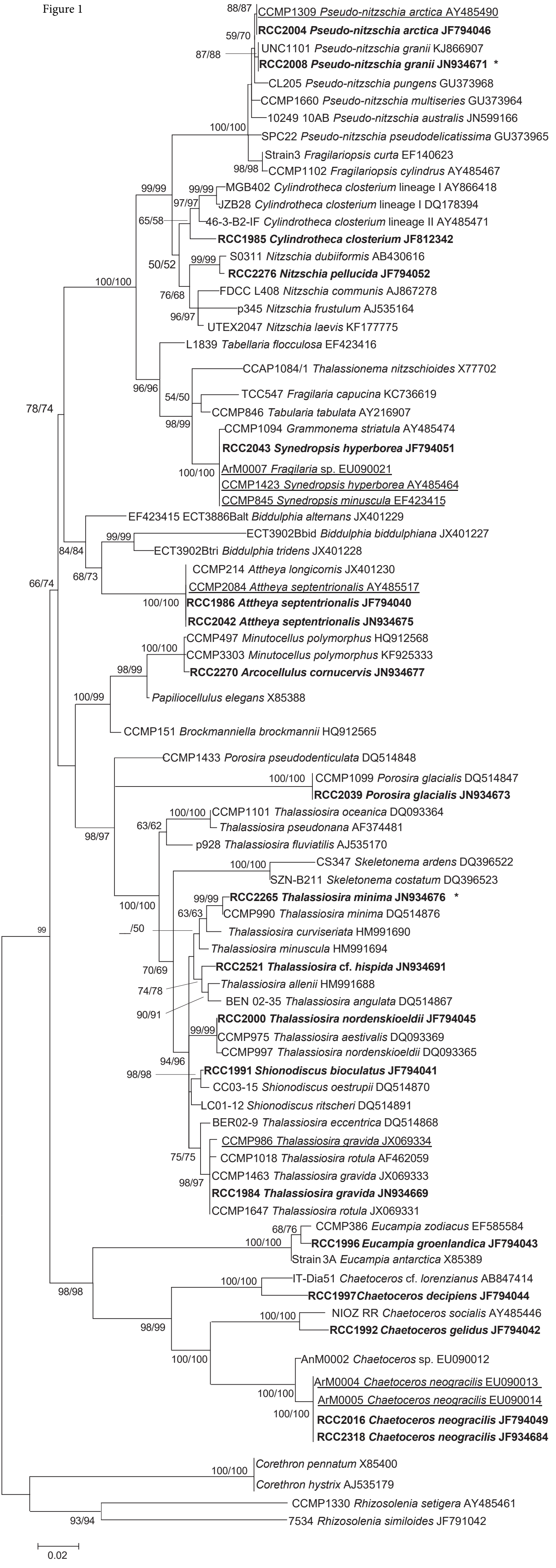

88/87
CCMP1309 Pseudo-nitzschia arctica AY485490
59/70
RCC2004 Pseudo-nitzschia arctica JF794046
87/88
UNC1101 Pseudo-nitzschia granii KJ866907
RCC2008 Pseudo-nitzschia granii JN934671 *
CL205 Pseudo-nitzschia pungens GU373968
CCMP1660 Pseudo-nitzschia multiseries GU373964
10249 10AB Pseudo-nitzschia australis JN599166
100/100
SPC22 Pseudo-nitzschia pseudodelicatissima GU373965
Strain3 Fragilariopsis curta EF140623
98/98
CCMP1102 Fragilariopsis cylindrus AY485467
99/99
99/99
MGB402 Cylindrotheca closterium lineage I AY866418
97/97
JZB28 Cylindrotheca closterium lineage I DQ178394
65/58
46-3-B2-IF Cylindrotheca closterium lineage II AY485471
RCC1985 Cylindrotheca closterium JF812342
50/52
99/99
S0311 Nitzschia dubiiformis AB430616
RCC2276 Nitzschia pellucida JF794052
100/100
76/68
FDCC L408 Nitzschia communis AJ867278
p345 Nitzschia frustulum AJ535164
96/97
UTEX2047 Nitzschia laevis KF177775
L1839 Tabellaria flocculosa EF423416
CCAP1084/1 Thalassionema nitzschioides X77702
96/96
54/50
TCC547 Fragilaria capucina KC736619
78/74
CCMP846 Tabularia tabulata AY216907
98/99
CCMP1094 Grammonema striatula AY485474
RCC2043 Synedropsis hyperborea JF794051
100/100
ArM0007 Fragilaria sp. EU090021
CCMP1423 Synedropsis hyperborea AY485464
CCMP845 Synedropsis minuscula EF423415
EF423415 ECT3886Balt Biddulphia alternans JX401229
99/99
ECT3902Bbid Biddulphia biddulphiana JX401227
84/84
ECT3902Btri Biddulphia tridens JX401228
68/73
CCMP214 Attheya longicornis JX401230
66/74
CCMP2084 Attheya septentrionalis AY485517
100/100
RCC1986 Attheya septentrionalis JF794040
RCC2042 Attheya septentrionalis JN934675
CCMP497 Minutocellus polymorphus HQ912568
100/100
CCMP3303 Minutocellus polymorphus KF925333
98/99
RCC2270 Arcocellulus cornucervis JN934677
100/99
Papiliocellulus elegans X85388
CCMP151 Brockmanniella brockmannii HQ912565
CCMP1433 Porosira pseudodenticulata DQ514848
100/100
CCMP1099 Porosira glacialis DQ514847
RCC2039 Porosira glacialis JN934673
100/100
CCMP1101 Thalassiosira oceanica DQ093364
63/62
Thalassiosira pseudonana AF374481
98/97
p928 Thalassiosira fluviatilis AJ535170
100/100
CS347 Skeletonema ardens DQ396522
SZN-B211 Skeletonema costatum DQ396523
99/99
RCC2265 Thalassiosira minima JN934676 *
100/100
63/63
CCMP990 Thalassiosira minima DQ514876
99
/50
Thalassiosira curviseriata HM991690
Thalassiosira minuscula HM991694
70/69
RCC2521 Thalassiosira cf. hispida JN934691
74/78
Thalassiosira allenii HM991688
90/91
BEN 02-35 Thalassiosira angulata DQ514867
RCC2000 Thalassiosira nordenskioeldii JF794045
99/99
CCMP975 Thalassiosira aestivalis DQ093369
CCMP997 Thalassiosira nordenskioeldii DQ093365
94/96
98/98
RCC1991 Shionodiscus bioculatus JF794041
CC03-15 Shionodiscus oestrupii DQ514870
LC01-12 Shionodiscus ritscheri DQ514891
BER02-9 Thalassiosira eccentrica DQ514868
CCMP986 Thalassiosira gravida JX069334
75/75
CCMP1018 Thalassiosira rotula AF462059
98/97
CCMP1463 Thalassiosira gravida JX069333
RCC1984 Thalassiosira gravida JN934669
CCMP1647 Thalassiosira rotula JX069331
68/76
CCMP386 Eucampia zodiacus EF585584
100/100
RCC1996 Eucampia groenlandica JF794043
Strain 3A Eucampia antarctica X85389
100/100
IT-Dia51 Chaetoceros cf. lorenzianus AB847414
98/98
RCC1997 Chaetoceros decipiens JF794044
100/100
NIOZ RR Chaetoceros socialis AY485446
98/99
RCC1992 Chaetoceros gelidus JF794042
100/100
AnM0002 Chaetoceros sp. EU090012
ArM0004 Chaetoceros neogracilis EU090013
100/100
ArM0005 Chaetoceros neogracilis EU090014
100/100
RCC2016 Chaetoceros neogracilis JF794049
RCC2318 Chaetoceros neogracilis JF934684
100/100
Corethron pennatum X85400
Corethron hystrix AJ535179
CCMP1330 Rhizosolenia setigera AY485461
93/94
7534 Rhizosolenia similoides JF791042
0.02

Figure 2

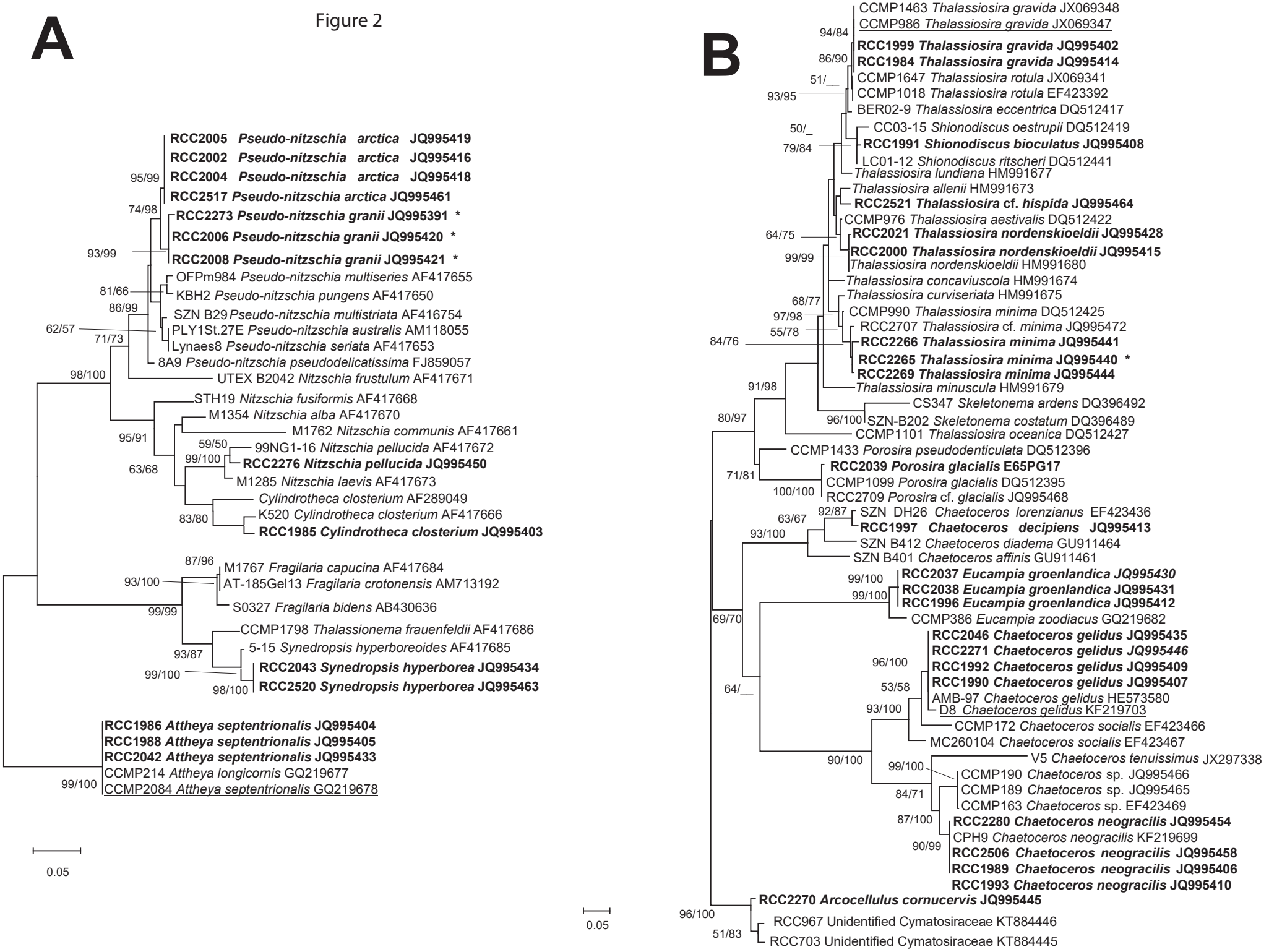
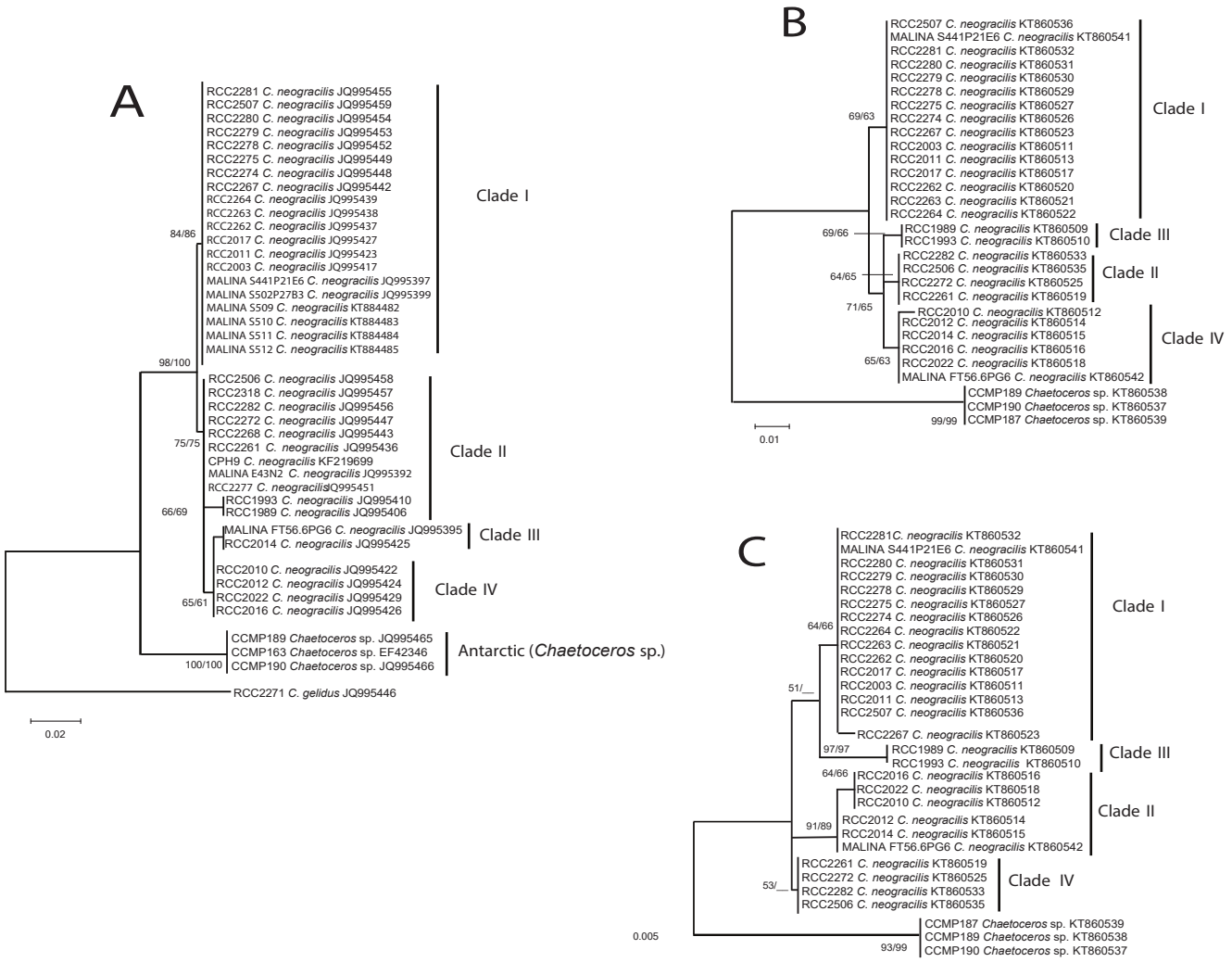
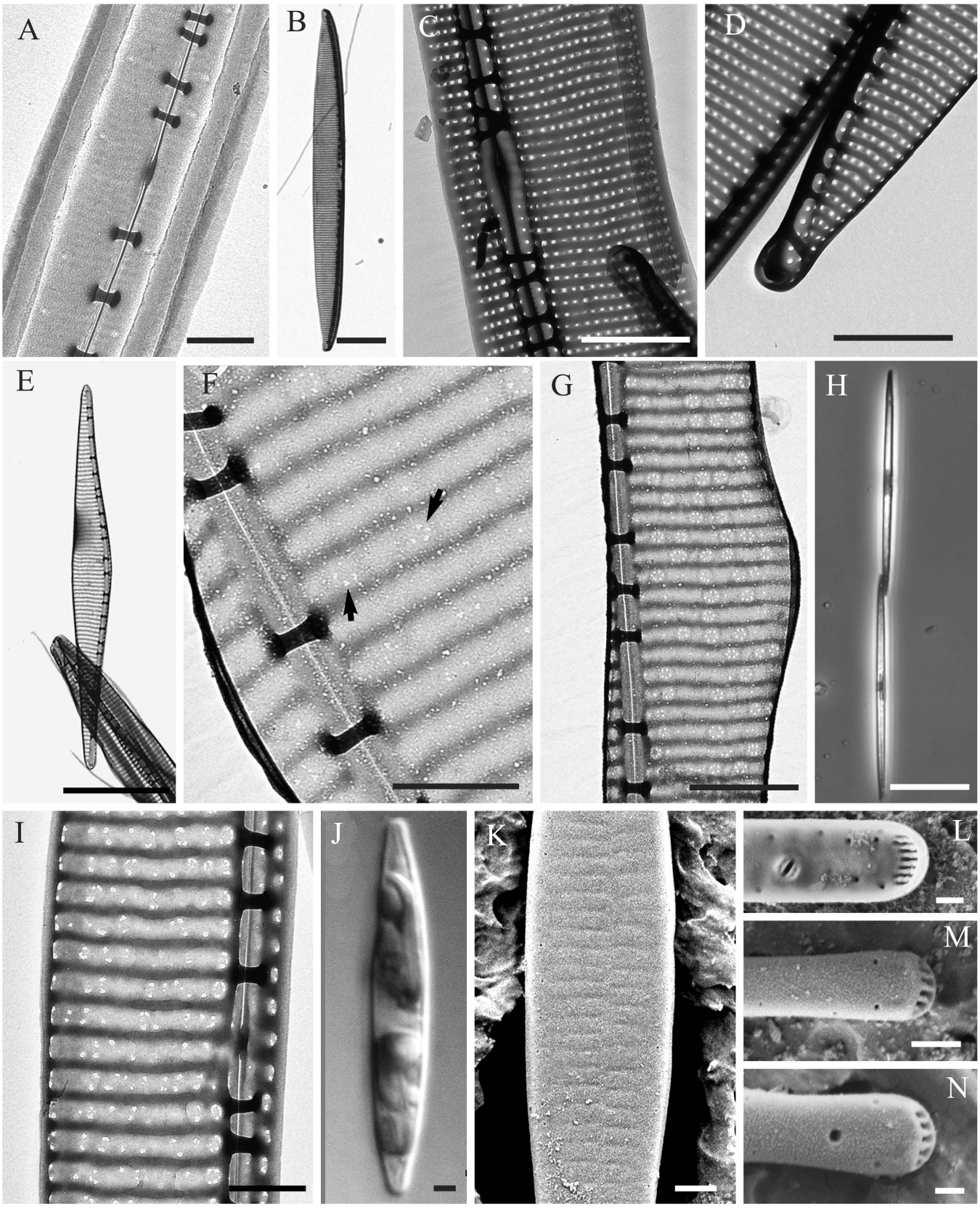
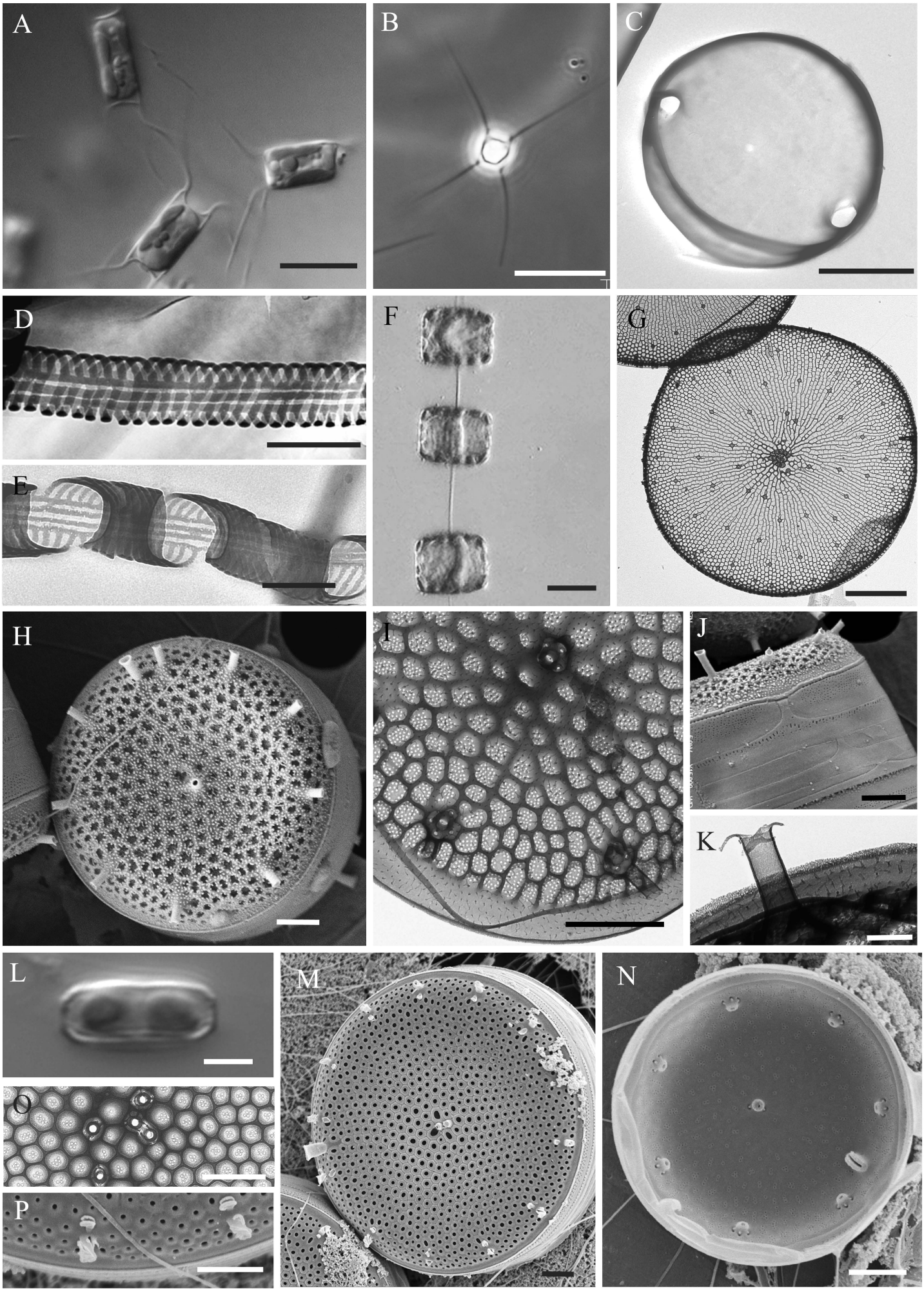
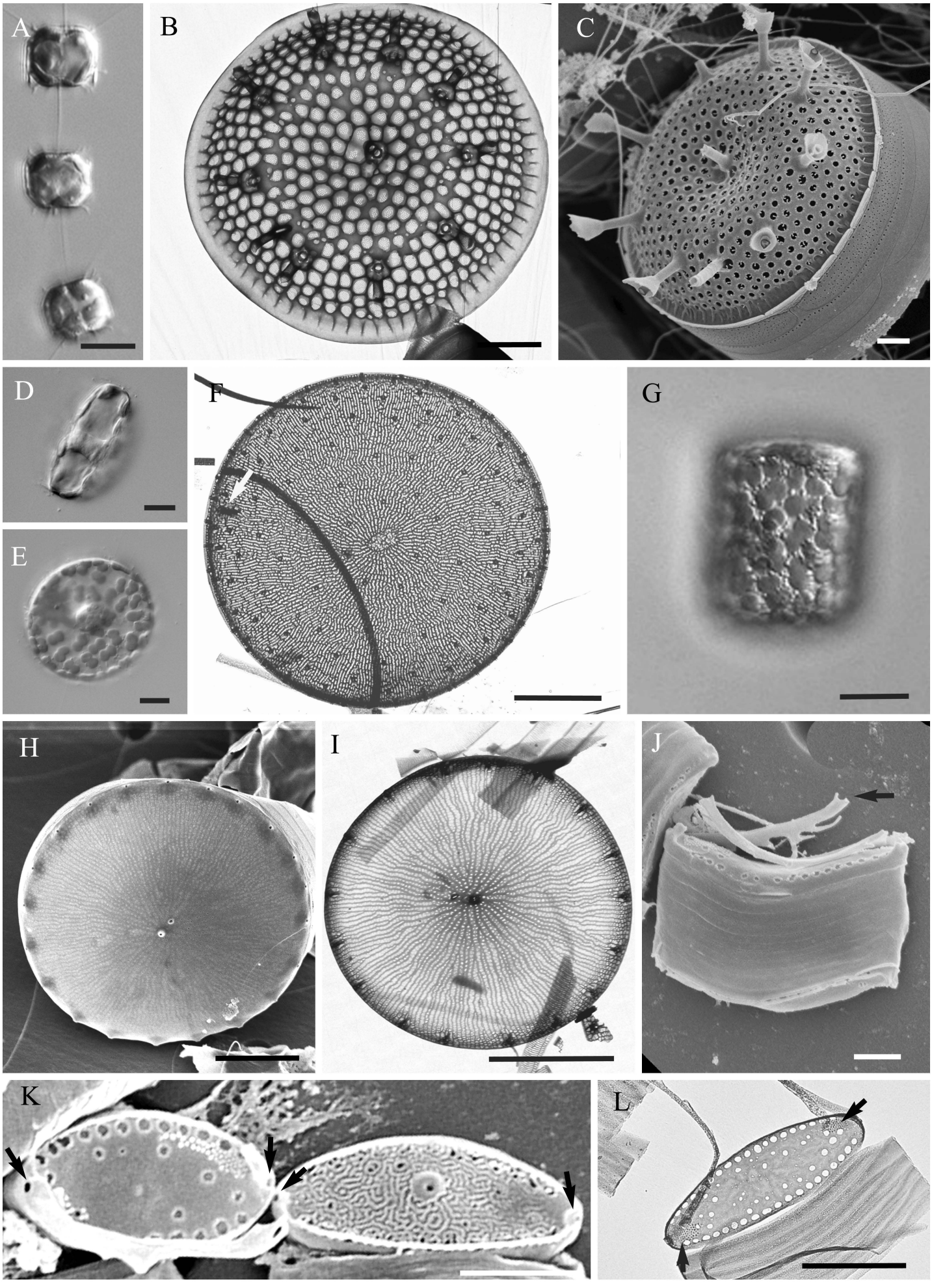

A
RCC2005 Pseudo-nitzschia arctica JQ995419
RCC2002 Pseudo-nitzschia arctica JQ995416
RCC2004 Pseudo-nitzschia arctica JQ995418
RCC2517 Pseudo-nitzschia arctica JQ995461
RCC2273 Pseudo-nitzschia granii JQ995391 *
RCC2006 Pseudo-nitzschia granii JQ995420 *
RCC2008 Pseudo-nitzschia granii JQ995421 *
OFPm984 Pseudo-nitzschia multiseries AF417655
KBH2 Pseudo-nitzschia pungens AF417650
SZN B29 Pseudo-nitzschia multistriata AF416754
PLY1St.27E Pseudo-nitzschia australis AM118055
Lynaes8 Pseudo-nitzschia seriata AF417653
8A9 Pseudo-nitzschia pseudodelicatissima FJ859057
UTEX B2042 Nitzschia frustulum AF417671
STH19 Nitzschia fusiformis AF417668
M1354 Nitzschia alba AF417670
M1762 Nitzschia communis AF417661
99NG1-16 Nitzschia pellucida AF417672
RCC2276 Nitzschia pellucida JQ995450
M1285 Nitzschia laevis AF417673
Cylindrotheca closterium AF289049
K520 Cylindrotheca closterium AF417666
RCC1985 Cylindrotheca closterium JQ995403
M1767 Fragilaria capucina AF417684
AT-185Gel13 Fragilaria crotonensis AM713192
S0327 Fragilaria bidens AB430636
CCMP1798 Thalassionema frauenfeldii AF417686
5-15 Synedropsis hyperboreoides AF417685
RCC2043 Synedropsis hyperborea JQ995434
RCC2520 Synedropsis hyperborea JQ995463
RCC1986 Attheya septentrionalis JQ995404
RCC1988 Attheya septentrionalis JQ995405
RCC2042 Attheya septentrionalis JQ995433
CCMP214 Attheya longicornis GQ219677
CCMP2084 Attheya septentrionalis GQ219678
0.05
B
CCMP1463 Thalassiosira gravida JX069348
CCMP986 Thalassiosira gravida JX069347
RCC1999 Thalassiosira gravida JQ995402
RCC1984 Thalassiosira gravida JQ995414
CCMP1647 Thalassiosira rotula JX069341
CCMP1018 Thalassiosira rotula EF423392
BER02-9 Thalassiosira eccentrica DQ512417
CC03-15 Shionodiscus oestrupii DQ512419
RCC1991 Shionodiscus bioculatus JQ995408
LC01-12 Shionodiscus ritscheri DQ512441
Thalassiosira lundiana HM991677
Thalassiosira allenii HM991673
RCC2521 Thalassiosira cf. hispida JQ995464
CCMP976 Thalassiosira aestivalis DQ512422
RCC2021 Thalassiosira nordenskioeldii JQ995428
RCC2000 Thalassiosira nordenskioeldii JQ995415
Thalassiosira nordenskioeldii HM991680
Thalassiosira concaviuscola HM991674
Thalassiosira curviseriata HM991675
CCMP990 Thalassiosira minima DQ512425
RCC2707 Thalassiosira cf. minima JQ995472
RCC2266 Thalassiosira minima JQ995441
RCC2265 Thalassiosira minima JQ995440 *
RCC2269 Thalassiosira minima JQ995444
Thalassiosira minuscula HM991679
CS347 Skeletonema ardens DQ396492
SZN-B202 Skeletonema costatum DQ396489
CCMP1101 Thalassiosira oceanica DQ512427
CCMP1433 Porosira pseudodenticulata DQ512396
RCC2039 Porosira glacialis E65PG17
CCMP1099 Porosira glacialis DQ512395
RCC2709 Porosira cf. glacialis JQ995468
SZN DH26 Chaetoceros lorenzianus EF423436
RCC1997 Chaetoceros decipiens JQ995413
SZN B412 Chaetoceros diadema GU911464
SZN B401 Chaetoceros affinis GU911461
RCC2037 Eucampia groenlandica JQ995430
RCC2038 Eucampia groenlandica JQ995431
RCC1996 Eucampia groenlandica JQ995412
CCMP386 Eucampia zoodiacus GQ219682
RCC2046 Chaetoceros gelidus JQ995435
RCC2271 Chaetoceros gelidus JQ995446
RCC1992 Chaetoceros gelidus JQ995409
RCC1990 Chaetoceros gelidus JQ995407
AMB-97 Chaetoceros gelidus HE573580
D8 Chaetoceros gelidus KF219703
CCMP172 Chaetoceros socialis EF423466
MC260104 Chaetoceros socialis EF423467
V5 Chaetoceros tenuissimus JX297338
CCMP190 Chaetoceros sp. JQ995466
CCMP189 Chaetoceros sp. JQ995465
CCMP163 Chaetoceros sp. EF423469
RCC2280 Chaetoceros neogracilis JQ995454
CPH9 Chaetoceros neogracilis KF219699
RCC2506 Chaetoceros neogracilis JQ995458
RCC1989 Chaetoceros neogracilis JQ995406
RCC1993 Chaetoceros neogracilis JQ995410
RCC2270 Arcocellulus cornucervis JQ995445
RCC967 Unidentified Cymatosiraceae KT884446
RCC703 Unidentified Cymatosiraceae KT884445
0.05

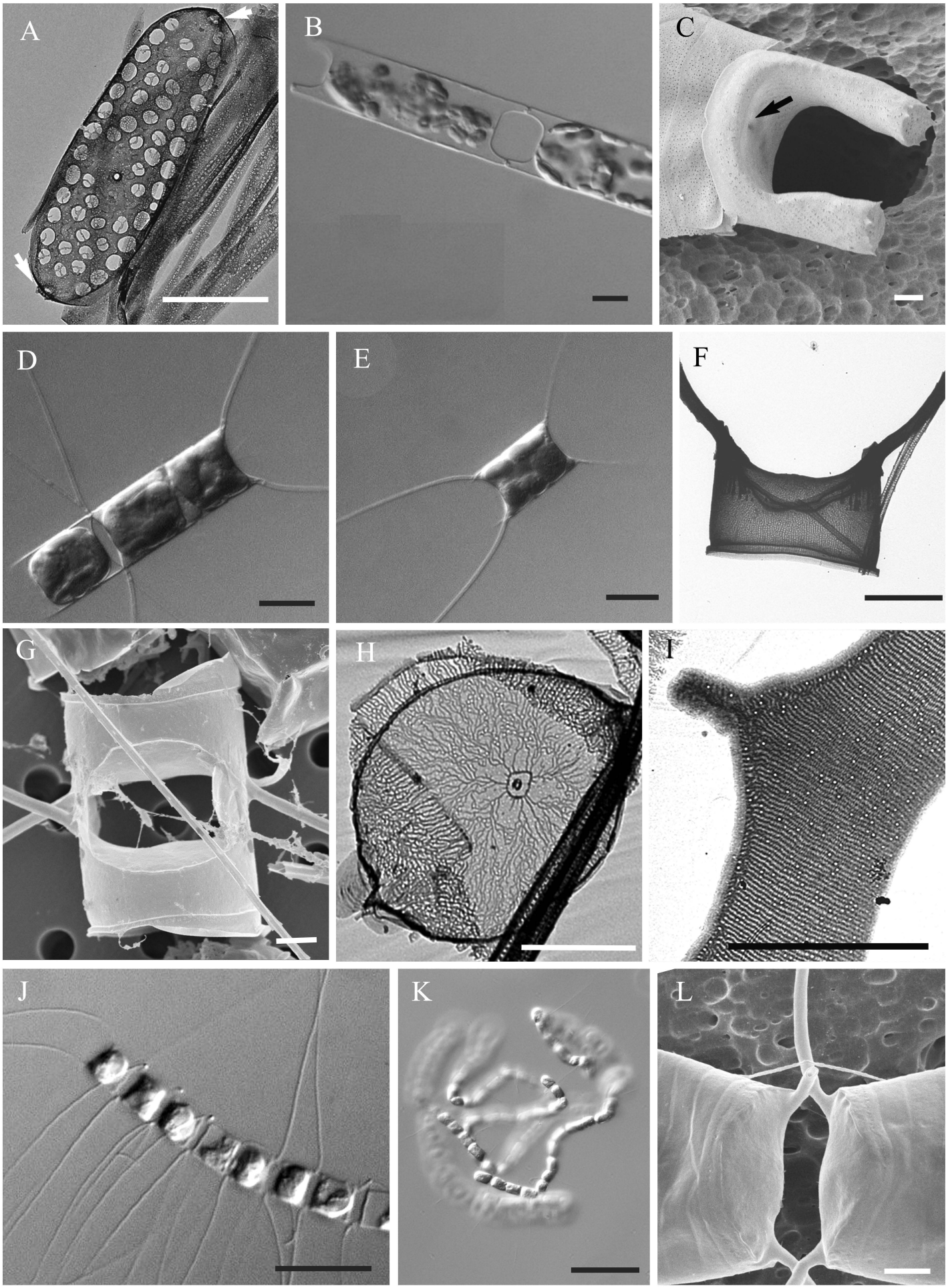
Figure 3


A
RCC2281 C. neogracilis JQ995455
RCC2507 C. neogracilis JQ995459
RCC2280 C. neogracilis JQ995454
RCC2279 C. neogracilis JQ995453
RCC2278 C. neogracilis JQ995452
RCC2275 C. neogracilis JQ995449
RCC2274 C. neogracilis JQ995448
RCC2267 C. neogracilis JQ995442
RCC2264 C. neogracilis JQ995439
RCC2263 C. neogracilis JQ995438
RCC2262 C. neogracilis JQ995437
RCC2017 C. neogracilis JQ995427
RCC2011 C. neogracilis JQ995423
RCC2003 C. neogracilis JQ995417
MALINA S441P21E6 C. neogracilis JQ995397
MALINA S502P27B3 C. neogracilis JQ995399
MALINA S509 C. neogracilis KT884482
MALINA S510 C. neogracilis KT884483
MALINA S511 C. neogracilis KT884484
MALINA S512 C. neogracilis KT884485
Clade I
84/86
98/100
RCC2506 C. neogracilis JQ995458
RCC2318 C. neogracilis JQ995457
RCC2282 C. neogracilis JQ995456
RCC2272 C. neogracilis JQ995447
RCC2268 C. neogracilis JQ995443
RCC2261 C. neogracilis JQ995436
CPH9 C. neogracilis KF219699
MALINA E43N2 C. neogracilis JQ995392
RCC2277 C. neogracilis JQ995451
Clade II
75/75
RCC1993 C. neogracilis JQ995410
RCC1989 C. neogracilis JQ995406
66/69
MALINA FT56.6PG6 C. neogracilis JQ995395
RCC2014 C. neogracilis JQ995425
Clade III
RCC2010 C. neogracilis JQ995422
RCC2012 C. neogracilis JQ995424
RCC2022 C. neogracilis JQ995429
RCC2016 C. neogracilis JQ995426
Clade IV
65/61
CCMP189 Chaetoceros sp. JQ995465
CCMP163 Chaetoceros sp. EF42346
CCMP190 Chaetoceros sp. JQ995466
Antarctic (Chaetoceros sp.)
100/100
RCC2271 C. gelidus JQ995446
0.02
B
RCC2507 C. neogracilis KT860536
MALINA S441P21E6 C. neogracilis KT860541
RCC2281 C. neogracilis KT860532
RCC2280 C. neogracilis KT860531
RCC2279 C. neogracilis KT860530
RCC2278 C. neogracilis KT860529
RCC2275 C. neogracilis KT860527
RCC2274 C. neogracilis KT860526
RCC2267 C. neogracilis KT860523
RCC2003 C. neogracilis KT860511
RCC2011 C. neogracilis KT860513
RCC2017 C. neogracilis KT860517
RCC2262 C. neogracilis KT860520
RCC2263 C. neogracilis KT860521
RCC2264 C. neogracilis KT860522
Clade I
69/63
RCC1989 C. neogracilis KT860509
RCC1993 C. neogracilis KT860510
Clade III
69/66
RCC2282 C. neogracilis KT860533
RCC2506 C. neogracilis KT860535
RCC2272 C. neogracilis KT860525
RCC2261 C. neogracilis KT860519
Clade II
64/65
71/65
RCC2010 C. neogracilis KT860512
RCC2012 C. neogracilis KT860514
RCC2014 C. neogracilis KT860515
RCC2016 C. neogracilis KT860516
RCC2022 C. neogracilis KT860518
MALINA FT56.6PG6 C. neogracilis KT860542
Clade IV
65/63
CCMP189 Chaetoceros sp. KT860538
CCMP190 Chaetoceros sp. KT860537
CCMP187 Chaetoceros sp. KT860539
99/99
0.01
C
RCC2281C. neogracilis KT860532
MALINA S441P21E6 C. neogracilis KT860541
RCC2280 C. neogracilis KT860531
RCC2279 C. neogracilis KT860530
RCC2278 C. neogracilis KT860529
RCC2275 C. neogracilis KT860527
RCC2274 C. neogracilis KT860526
RCC2264 C. neogracilis KT860522
RCC2263 C. neogracilis KT860521
RCC2262 C. neogracilis KT860520
RCC2017 C. neogracilis KT860517
RCC2003 C. neogracilis KT860511
RCC2011 C. neogracilis KT860513
RCC2507 C. neogracilis KT860536
Clade I
64/66
51/__
RCC2267 C. neogracilis KT860523
RCC1989 C. neogracilis KT860509
RCC1993 C. neogracilis KT860510
Clade III
97/97
RCC2016 C. neogracilis KT860516
RCC2022 C. neogracilis KT860518
RCC2010 C. neogracilis KT860512
Clade II
64/66
RCC2012 C. neogracilis KT860514
RCC2014 C. neogracilis KT860515
MALINA FT56.6PG6 C. neogracilis KT860542
91/89
RCC2261 C. neogracilis KT860519
RCC2272 C. neogracilis KT860525
RCC2282 C. neogracilis KT860533
RCC2506 C. neogracilis KT860535
Clade IV
53/__
CCMP187 Chaetoceros sp. KT860539
CCMP189 Chaetoceros sp. KT860538
CCMP190 Chaetoceros sp. KT860537
93/99
0.005


A
B
C
D
E
F
G
H
I
J
K
L
M
N


A
B
C
D
F
G
E
H
I
J
K
L
M
N
O
P


A
B
C
D
F
G
E
H
I
J
K
L


A
B
C
D
E
F
G
H
I
J
K
L

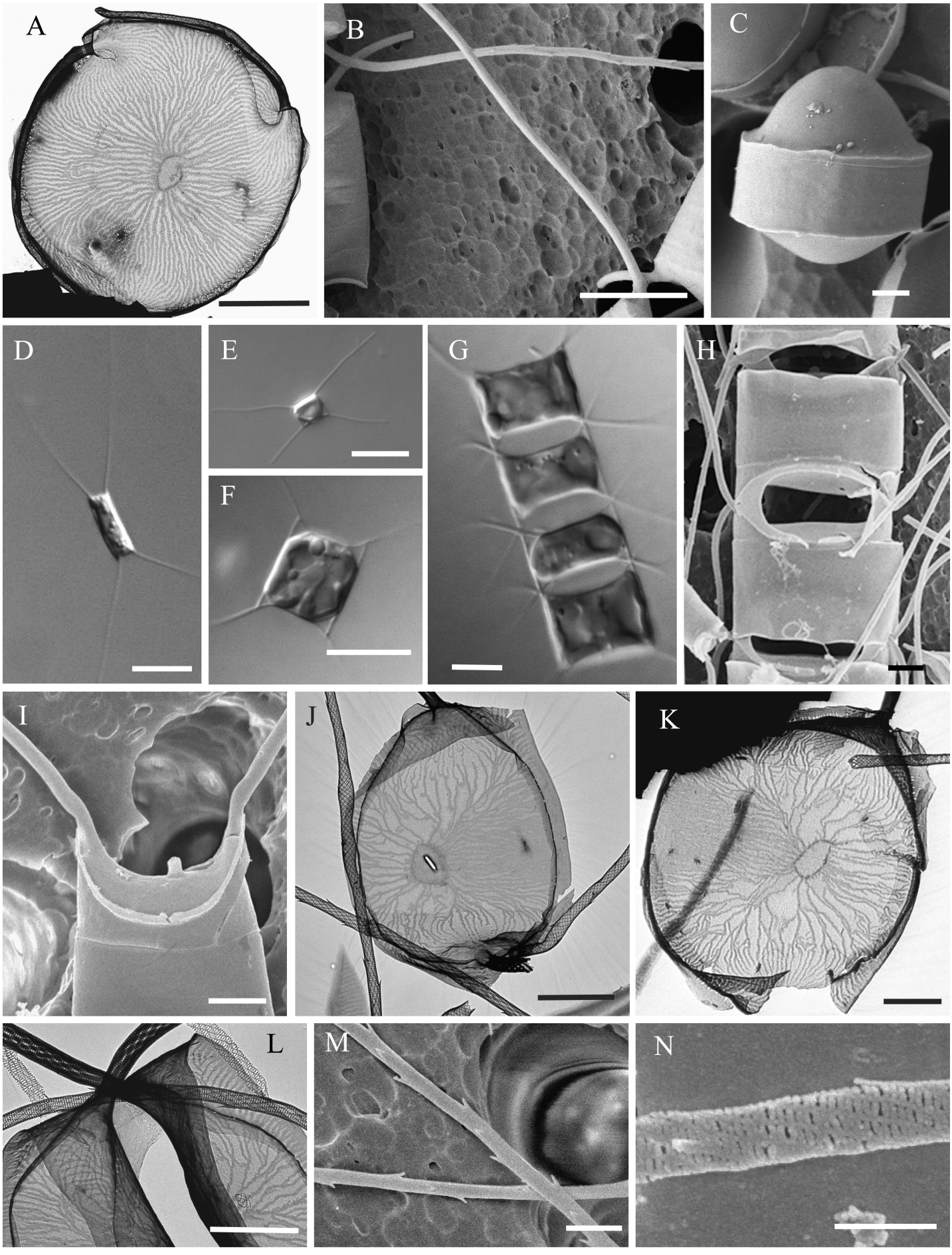
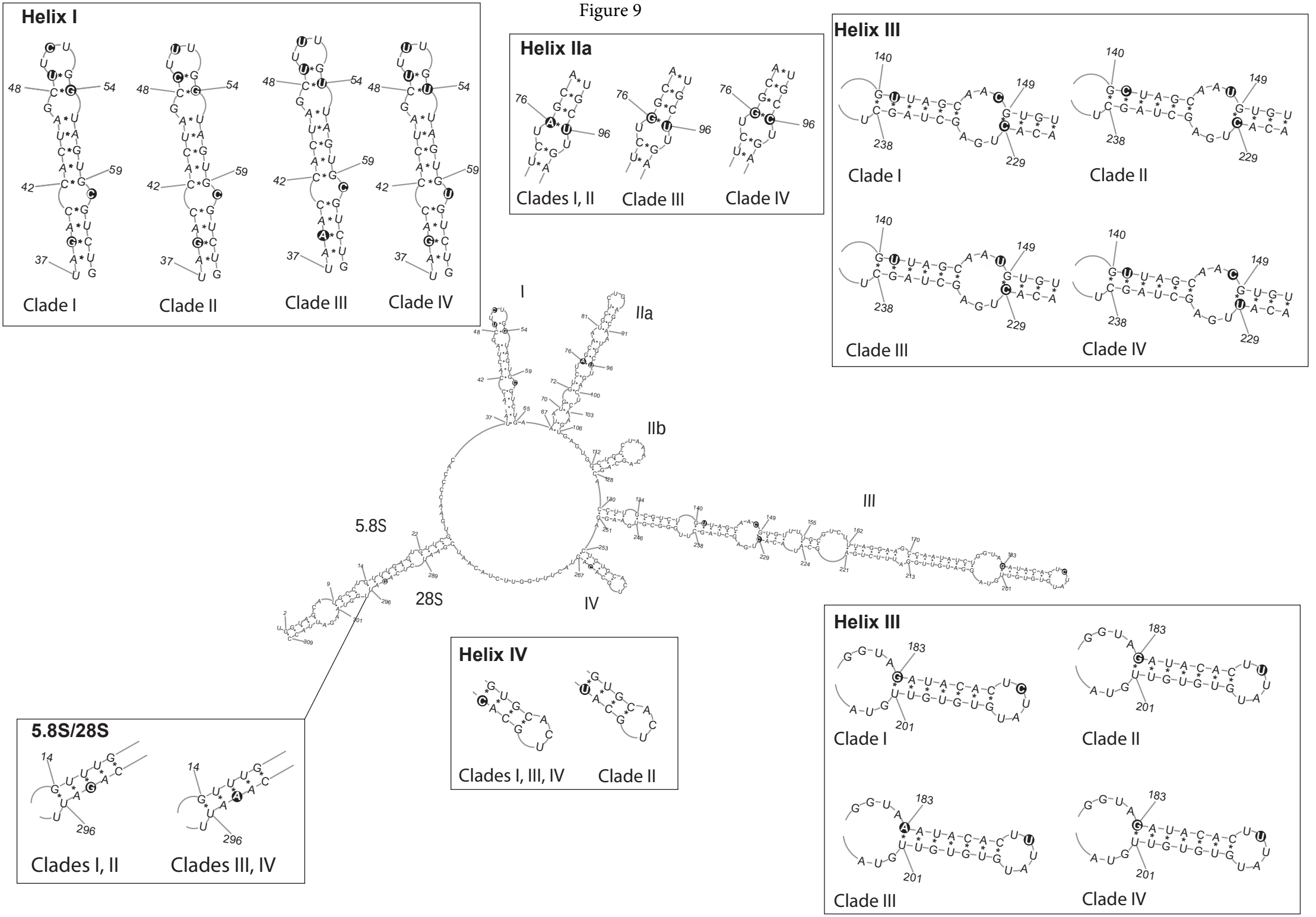
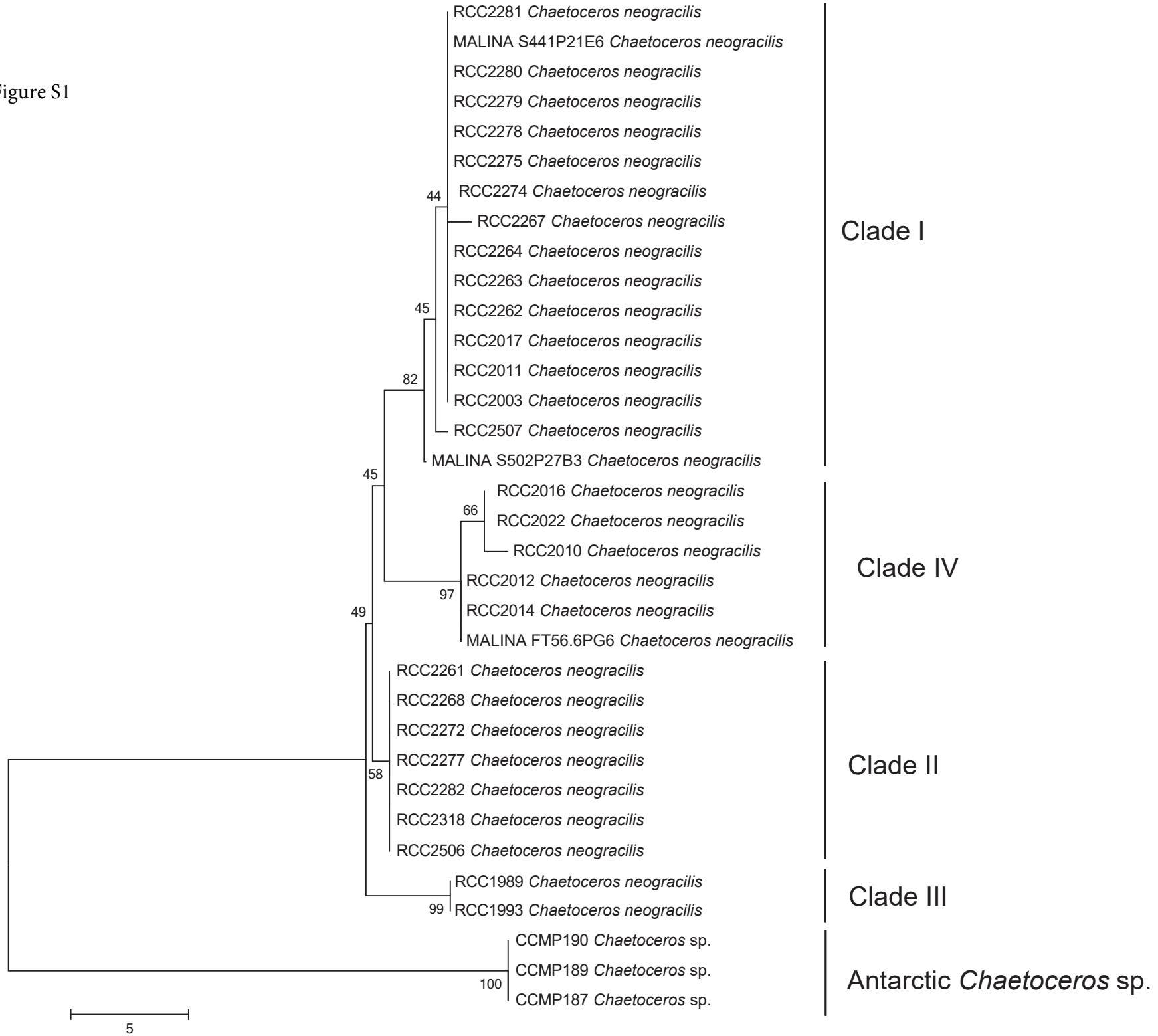

A
B
C
D
E
F
G
H
I
J
K
L
M
N

Figure 9


Helix I
Clade I
Clade II
Clade III
Clade IV
Helix IIa
Clades I, II
Clade III
Clade IV
Helix III
Clade I
Clade II
Clade III
Clade IV
I
IIa
IIb
III
IV
5.8S
28S
Helix IV
Clades I, III, IV
Clade II
5.8S/28S
Clades I, II
Clades III, IV
Helix III
Clade I
Clade II
Clade III
Clade IV

Supplementary Figure S1



Supplementary Table S1

Station	CTD	Depth	Latitude (°N)	Longitude (°W)	Temperature (°C)	Salinity (psu)	Cultures direct FCS[a]	Cultures TFF[b]		Culture Enrichments	
								FCS	Pipette isolation	FCS	Pipette isolation
PAC05		3								1	
PAC06		3	50.06	-139.53	12.1	32.5				1	
PAC08		3	53.36	159.29	11.8	32.7					2
ARC12		3	71.19	159.42	2	30.5	1				
BEA13		3	70.56	145.4	8.8	17.6	2			2	
BEA14		3	70.5	135.5	3.3	25.6	1			1	
110	56	3	71.7	-126.48	4.4	28.7			3		
235	191	3	71.76	-130.83	0	27.3		3			
235	191	25	71.76	-130.83	1.6	29.9		1			
280	42	30	70.87	-130.51	-0.7	32.2			8		
320	82	3	71.57	-133.94	-0.8	27		3			
394	38	3	69.85	-133.49	7	25.1	3				
460	145	3	70.68	-136.05			2				
620	99	3	70.68	-139.63	1.6	22.1		2	3		
620	99	65	70.68	-139.63	-1.1	30.7		1	5		
680	35	3	69.6	-138.23	8.3	14.7			3		
680	35	40	69.6	-138.23	-1.2	31.3		1			
690	31	3	69.49	-137.94	7.4	19	1				
690	31	29	69.49	-137.94	-1.3	31.2					16
760	106	3	70.55	-140.8	0.6	22.3		8	2		
Total							10	19	24	5	18

Supplementary Table S2

Strain ID	Strain name	Authors	Geographical origin	Genbank 18S	Genbank 28S	Genbank ITS
CCMP214	*Attheya longicornis*	Crawford & Gardner	Gulf of Maine, North Atlantic Ocean	JX401230	GQ219677	
CCMP2084	*Attheya septentrionalis*	(Østrup) Crawford	Baffin Bay, Arctic	AY485517	GQ219678	
ECT3886Balt	*Biddulphia alternans*	(Bailey) Van Heurck	Kahana Bay, North Pacific Ocean	JX401229		
ECT3902Bbid	*Biddulphia biddulphiana*	(J.E.Smith) Boyer	Unknown	JX401227		
ECT3902 Btri	*Biddulphia tridens*	(Ehrenberg) Ehrenberg	Long Beach, North Pacific Ocean	JX401228		
CCMP151	*Brockmanniella brockmanni*	(Hustedt) Hasle, Stosch & Syvertsen	Unknown	HQ912565		
SZN B401	*Chaetoceros affinis*	Lauder	Unknown		GU911461	
SZN-B412	*Chaetoceros diadema*	(Ehrenberg) Gran	Gulf of Naples, Mediterranean Sea		GU911464	
AMB-97	*Chaetoceros gelidus*	Chamnansinp, Li, Lundholm & Moestrup	Tromso, North Atlantic Ocean		HE573580	
D8	*Chaetoceros gelidus*	Chamnansinp, Li, Lundholm & Moestrup	Kattegat Bay, North Sea		KF219703	
SZN-DH26	*Chaetoceros lorenzianus*	Grunow	Gulf of Naples, Mediterranean Sea		EF423436	
IT-Dia51	*Chaetoceros* cf. *lorenzianus*	Grunow	Aki nada Sea, Indian Ocean	AB847414		
ArM0004	*Chaetoceros neogracilis*	(Schütt) VanLandingham	Svalbard, Arctic Ocean	EU090013		
ArM0005	*Chaetoceros neogracilis*	(Schütt) VanLandingham	Svalbard, Arctic Ocean	EU090014		
CPH9	*Chaetoceros neogracilis*	(Schütt) VanLandingham	Hellerup Harbour, Baltic Sea		KF219699	
CCMP172	*Chaetoceros socialis*	Lauder	Friday Harbour, North Pacific Ocean		EF423466	
MC260104	*Chaetoceros socialis*	Lauder	Gulf of Naples, Mediterranean Sea		EF423467	
NIOZ RR	*Chaetoceros socialis*	Lauder	Unknown	AY485446		
AnM0002	*Chaetoceros* sp.		King George Island, Antarctica	EU090012		
CCMP163	*Chaetoceros* sp.		Southern Ocean		EF426369	
CCMP187	*Chaetoceros* sp.		Weddell Sea, Antarctica			KT860539
CCMP189	*Chaetoceros* sp.		Weddell Sea, Antarctica		JQ995466	KT860538
CCMP190	*Chaetoceros* sp.		Weddell Sea, Antarctica		JQ995465	KT860537
V5	*Chaetoceros tenuissimus*	Meunier	Black Sea		JX297338	
Unavailable	*Corethron pennatum*	(Grunow) Ostenfeld	Unknown	X85400		
Unavailable	*Corethron hystrix*	Hensen	Unknown	AJ535179		
MGB402	*Cylindrotheca closterium*	(Ehrenberg) Lewin & Reimann	Qindao Bay, Indian Ocean	AY866418		
JZB28	*Cylindrotheca closterium*	(Ehrenberg) Lewin & Reimann	Qindao Bay, Indian Ocean	DQ178394		
NIOZ (46-3-B2-IF)	*Cylindrotheca closterium*	(Ehrenberg) Lewin & Reimann	Unknown	AY485471		
K520	*Cylindrotheca closterium*	(Ehrenberg) Lewin & Reimann	Kattegat Bay, North Sea		AF417666	
Unavailable	*Cylindrotheca closterium*	(Ehrenberg) Lewin & Reimann	Unknown		AF289049	
Strain 3A	*Eucampia antarctica*	(Castracane) Mangin	Unknown	X85389		
CCMP386	*Eucampia zodiacus*	Ehrenberg	Gulf of Maine, North Atlantic Ocean	EF585584	GQ219682	
S0327	*Fragilaria bidens*	Heiberg	Unknown		AB430636	
TCC547	*Fragilaria capucina*	Desmazières	Unknown	KC736619		
M1767	*Fragilaria capucina*	Desmazières	Cologne pond, Germany		AF417684	
AT-185Gel3	*Fragilaria crotonensis*	Kitton	Bremen pond, Germany		AM713192	
AnM0007	*Fragilaria* sp.		Svalbard, Arctic Ocean	EU090021		
Strain 3	*Fragilariopsis curta*	(Van Heurck) Hustedt	Mertz glacier, Antarctica	EF140623		
CCMP1102	*Fragilariopsis cylindrus*	(Grunow) Helmcke & Krieger	Islas Orcadas, Antarctica	AY485467		
CCMP1094	*Grammonema striatula*	C.Agardh	Gulf of Alaska, North Pacific Ocean	AY485474		
CCMP497	*Minutocellus polymorphus*	(Hargraves & Guillard) Hasle, Stosch, & Syv	Bermuda, North Atlantic Ocean	HQ912568		
CCMP3303	*Minutocellus polymorphus*	(Hargraves & Guillard) Hasle, Stosch, & Syv	Angola coast, South Atlantic Ocean	KF925333		
M1354	*Nitzschia alba*	J.C.Lewin & R.A.Lewin	Roscoff, English Channel		AF417670	
M1762	*Nitzschia communis*	Rabenhorst	Cologne Botanical Garden, Germany		AF417661	
FDCC L408	*Nitzschia communis*	Rabenhorst	Unknown	AJ867278		
S0311	*Nitzschia dubiformis*	Hustedt	Unknown	AB430616		
STH19	*Nitzschia fusiformis*	Grunow	Isefjord, Kattegat Bay, North Sea		AF417668	
p345	*Nitzschia frustulum*	(Kützing) Grunow	Unknown	AJ535164		
UTEXB2042	*Nitzschia frustulum*	(Kützing) Grunow	la Jolla, California, North Atlantic Ocean		AF417671	
UTEX2047	*Nitzschia laevis*	Hustedt	Woods Hole, North Atlantic Ocean	KF177775		
M1285	*Nitzschia laevis*	Hustedt	Dusseldorf pond, Germany		AF417673	
99NG1-16	*Nitzschia pellucida*	Grunow	Ishigaki Island, Japan		AF417672	
Unavailable	*Papiliocellulus elegans*	Hasle, Stosch & Syvertsen	Unknown	X85388		
CCMP1099	*Porosira glacialis*	(Grunow) Jørgensen	Islas Orcadas, Antarctica	DQ514847	DQ512395	
RCC2709	*Porosira* cf. *glacialis*	(Grunow) Jørgensen	Fildes Bay, Antarctica		JQ995468	
CCMP1433	*Porosira pseudodenticulata*	(Hustedt) Jousé	Ross Sea, Antarctica	DQ514848	DQ512396	
CCMP1309	*Pseudo-nitzschia arctica*	Percopo & Sarno	Barrow Strait, Canada	AY485490		
10249 10AB	*Pseudo-nitzschia australis*	Frenguelli	Monterey Bay, North Pacific Ocean	JN599166		
Ply1St.27E	*Pseudo-nitzschia australis*	Frenguelli	Loch Lihne, North Atlantic Ocean		AM118055	
UNC1101	*Pseudo-nitzschia granii*	(Hasle) Hasle	Gulf of Alaska, North Pacific Ocean	KJ866907		
CCMP1660	*Pseudo-nitzschia multiseries*	(Hasle) Hasle	Gulf St Lawrence, North Atlantic Ocean	GU373964		
OFPm984	*Pseudo-nitzschia multiseries*	(Hasle) Hasle	Ofunato Bay, Indian Ocean		AF417655	
SZNB29	*Pseudo-nitzschia multistriata*	(Takano) Takano	Gulf of Naples, Mediterranean Sea		AF416754	
SPC22	*Pseudo-nitzschia pseudodelicatissima*	(Hasle) Hasle	San Pedro Channel, North Pacific Ocean	GU373965		
8A9	*Pseudo-nitzschia pseudodelicatissima*	(Hasle) Hasle	Thermaikos Gulf, Mediterranean Sea		FJ859057	
CL205	*Pseudo-nitzschia pungens*	(Grunow ex Cleve) Hasle	Lennoz Channel, North Atlantic Ocean	GU373968		
KBH2	*Pseudo-nitzschia pungens*	(Grunow ex Cleve) Hasle	Khan Hoa Bay, Indian Ocean		AF417650	
Linaes8	*Pseudo-nitzschia seriata*	(Cleve) H.Peragallo	Isefjord, Kattegat Bay, North Sea		AF417653	
CCMP1440	*Pseudo-nitzschia* sp.		McMurdo Sound, Antarctica	GU373969		
CCMP1330	*Rhizosolenia setigera*	Brightwell	Massacchussets, USA	AY485461		
7534	*Rhizosolenia similoides*	Cleve	Gulf of Mexico, North Atlantic Ocean	JF791042		
CC03-15	*Shionodiscus oestrupii*	(Ostenfeld) Alverson, Kang & Theriot	Sargassum Sea, North Atlantic Ocean	DQ514870	DQ512419	
LC01-12	*Shionodiscus ritscheri*	(Hustedt) Alverson, Kang & Theriot	Drake Passage, Southern Ocean	DQ514891	DQ512441	
CS347	*Skeletonema ardens*	Sarno & Zingone	Gulf of Carpentaria, Indian Ocean	DQ396522	DQ396492	
SZN-B202	*Skeletonema costatum*	(Greville) Cleve	Indian River Lagoon, FL, USA		DQ396489	
SZN B211	*Skeletonema costatum*	(Greville) Cleve	Montevideo coast, South Atlantic Ocean	DQ396523		
CCMP1423	*Synedropsis hyperborea*	(Grunow) Hasle, Medlin & Syvertsen	Baffin Bay, Arctic	AY485464		
5-15	*Synedropsis hyperboreoides*	Hasle, Medlin & Syvertsen	Ross Sea, Antarctica		AF417685	
CCMP845	*Synedropsis minuscula*	(Grunow) Kooistra	Baffin Bay, Arctic	EF423415		
L1839	*Tabellaria flocculosa*	(Roth) Kützing	Unknown	EF423416		
CCMP846	*Tabularia tabulata*	(Agardh) Snoeijs	Gulf of Alaska, North Pacific Ocean	AY216907		
CCMP1798	*Thalassionema frauenfeldii*	(Grunow) Tempère & Peragallo	North Atlantic Ocean		AF417686	
CCAP1084/1	*Thalassionema nitzschioides*	(Grunow) Mereschkowsky	Unknown	X77702		
CCMP975	*Thalassiosira aestivalis*	Gran	Gulf of Maine, North Atlantic Ocean	DQ093369		
CCMP975	*Thalassiosira aestivalis*	Gran	Vancouver coast, North Pacific Ocean		DQ512422	
Unknown	*Thalassiosira allenii*	Takano	Unknown	HM991688	HM991673	
BEN 02-35	*Thalassiosira angulata*	(Gregory) Hasle	San Joaquin River, California, USA	DQ514867		
Unavailable	*Thalassiosira concaviuscola*	Makarova	Unknown		HM991674	
Unavailable	*Thalassiosira curviseriata*	Takano	Unknown	HM991690	HM991675	
BER02-9	*Thalassiosira eccentrica*	(Ehrenberg) Cleve	San Francisco Bay, North Pacific Ocean	DQ514868	DQ512417	
p928	*Thalassiosira fluviatilis*	Hustedt	Unknown	AJ535170		
CCMP986	*Thalassiosira gravida*	Cleve	Tromso, North Atlantic Ocean	JX069334	JX069347	
CCMP1463	*Thalassiosira gravida*	Cleve	McMurdo Sound, Antarctica	JX069333	JX069348	
Unknown	*Thalassiosira lundiana*	Fryxell	Unknown		HM991677	
CCMP990	*Thalassiosira minima*	Gaarder	Unknown	DQ514876	DQ512425	
RCC2707	*Thalassiosira* cf. *minima*	Gaarder	Fildes Bay, Antarctica		JQ995472	
Unavailable	*Thalassiosira minuscula*	Krasske	Unknown	HM991694	HM991679	
CCMP997	*Thalassiosira nordenskioeldii*	Cleve	Norwegian Sea, North Atlantic Ocean	DQ093365		
Unavailable	*Thalassiosira nordenskioeldii*	Cleve	Unknown		HM991680	
CCMP1101	*Thalassiosira oceanica*	Hasle	Islas Orcadas, Antarctica	DQ093364	DQ512427	
Unavailable	*Thalassiosira pseudonana*	Hasle & Heimdal	Unknown	AF374481		
CCMP1018	*Thalassiosira rotula*	Meunier	California Bight, North Pacific Ocean	AF462059	EF423392	
CCMP1647	*Thalassiosira rotula*	Meunier	Gulf of Naples, Mediterranean Sea	JX069331	JX069341	
RCC967	Unidentified Cymatosiraceae		Chile upwelling, South Pacific Ocean		KT884446	
RCC703	Unidentified Cymatosiraceae		South Africa Coast		KT884445	