

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

ORIGINAL PAPER

Barcoding of Diatoms: Nuclear Encoded ITS Revisited

Mónica B.J. Moniz¹, and Irena Kaczmarska

Biology Department, Mount Allison University, 63B York Street, Sackville, NB E4L 1G7, Canada

Submitted September 29, 2008; Accepted July 4, 2009
Monitoring Editor: Marina Montresor

DNA-barcoding is based on the premise that the divergence of a small DNA fragment coincides with biological separation of species. If true, it offers an additional tool for worldwide consistent species recognition even in cases of semi-cryptic species. Our study includes 618 sequences representing 114 diatom species belonging to the two most species-rich classes of diatoms (Mediophyceae and Bacillariophyceae). A 99.5% success rate in separating biologically defined species and a 91% success rate in separating all species tested was obtained when using the proposed barcode starting at the 5' end of 5.8S and ending in the conserved motif of helix III of ITS2 (300 to 400 bp). Including the whole 5.8S+ITS2 region did not significantly improve species resolution. We tested our barcode on 17 unidentified, misidentified or contaminated strains derived mostly from a culture collection, and these were correctly flagged as erroneous by their ITS sequences. We conclude that the proposed barcode represents for the Mediophyceae and Bacillariophyceae a robust, economical, and rapid way to recognize and identify most species (when a reference sequence is available) that is as good as or better than other molecular markers thus far proposed.

Crown Copyright © 2009 Published by Elsevier GmbH. All rights reserved.

Key words: centrics; diatom; DNA-barcode; ITS; pennates; species identification.

Introduction

Recent reports of economically important species in danger of extinction and of geographic expansion of invasive and potentially harmful biota are expressions of the current changes in the oceans, necessitating research involving interactions between marine biodiversity and ecosystem functioning (Duffy and Stachowicz 2006). However, such studies are compromised by uncertain or incorrect identification of marine microbiota, which play key roles in many of these processes. Bortolus (2008) has recently illustrated how incorrect taxonomy leads to incorrect conclusions in ecological and other biological research, high-

lighting the need for accurate and globally consistent α -(species level)-taxonomy.

In eukaryotes, the morphological concept of species is most widely used. The relationship between diagnostic morphology and biological species boundaries is well understood in complex macro-biota, but it is largely unknown in many micro-eukaryotes (Chantangsi et al. 2007; Evans et al. 2007; Gómez et al. 2006; Kaczmarska et al. 2008; Mann 1999; Medlin 1997; Scicluna et al. 2006; Shao et al. 2004). Among micro-eukaryotes

Abbreviations: CBC, compensatory base change; *cox1*, cytochrome oxidase 1; ITS, internal transcribed spacers; LSU, large ribosomal subunit; SSU, small ribosomal subunit

¹Corresponding author; fax +1 506 364 2505
e-mail mbmoniz@mta.ca (M.B.J. Moniz).

such as unicellular green algae, ciliates, stramenopile parasites and fungi, morphology based taxonomy has either been replaced or at least supported by molecular taxonomy (Barth et al. 2006; Chantangsi et al. 2007; Lynn and Strüder-Kypke 2006; Scicluna et al. 2006). The central debatable issues with this methodology are which molecular markers are most suitable at the species level and what level of DNA-sequence divergence is most effective in discriminating species. This is of crucial importance since the level of intraspecific divergence varies between groups and further depends on which molecular marker is used and what parameters are used for alignments (Amato et al. 2007; Fazekas et al. 2008; Kaczmarska et al. 2007; Litaker et al. 2007).

Recently it was postulated that a small, standard portion of a single gene (or “barcode”) may be used to recognize and identify species from a wide range of animal taxa (Cywinska et al. 2006; Hajibabaei et al. 2006; Hebert et al. 2003, 2004; Janzen et al. 2009; Ratnasingham and Hebert 2007; Ward et al. 2005). The hope was that such a barcode is a species-specific tag that would bring global consistency to α -taxonomy. The gene sequence initially chosen to serve as a universal barcode was a 600 bp fragment located at the 5' end of the mitochondrial gene *cytochrome oxidase 1* (*cox1*, Hebert et al. 2003). Animal phyla by and large have been successfully, expeditiously and inexpensively barcoded using this mitochondrial gene segment (Cywinska et al. 2006; Hajibabaei et al. 2006; Hebert et al. 2003, 2004; Ratnasingham and Hebert 2007; Ward et al. 2005). *Cox1*-based barcodes, however, have proven inadequate or impractical to use in some other phyla and divisions of life (Chase et al. 2005; Cowan et al. 2006; Fazekas et al. 2008, 2009; Kress et al. 2005; Kress and Erickson 2007; Moniz and Kaczmarska 2009; Newmaster et al. 2006, 2008; Seifert 2009). Consequently, alternative barcodes are being considered in plants (Fazekas et al. 2008) and different protist groups, e.g. dinoflagellates (Litaker et al. 2007), parasitic stramenopiles (Scicluna et al. 2006) and peritrich ciliates (Utz and Eizirik 2007). It is becoming apparent that because the evolutionary history of genes is not uniform across all lineages, there may not be one gene that can be used as a universal barcode for all life.

Unlike many other microbial protists, diatom frustules are rich in micro-architecture, and these have been traditionally exploited as taxonomically informative characters. Unfortunately, morphology is a complex marker susceptible to subjective

application (Mann 1999) and sometimes presenting insufficient information in the separation of species (Amato et al. 2007; Evans et al. 2007; Kaczmarska et al. 2008; Lundholm et al. 2006; Mann 1999; Sarno et al. 2005). With the number of diatom species estimates ranging around 200,000 (Mann and Droop 1996) and possibly no more than a few percent formally described to date, there is an obvious need for more research in diatom biodiversity. Unfortunately, the relationship of diagnostic morphology to biological boundaries of diatom species are known for only about 15 species (Amato et al. 2007; Behnke et al. 2004; Casteleyn et al. 2008, 2009; Kaczmarska et al. 2008, in press; Vanormelingen et al. 2007, 2008). Such studies are difficult, time consuming and expensive to conduct, which means this approach will unlikely prove practical for resolving tens of thousands of species.

Because of such limitations, the last decade witnessed several research efforts to better understanding of the relationship between morphological, molecular and biological delineations of protistean species (Amato et al. 2007; Behnke et al. 2004; Beszteri et al. 2005; Casteleyn et al. 2008; Coleman 2000, 2003, 2007; Coleman and Mai 1997; Kaczmarska et al. 2008; Vanormelingen et al. 2007, 2008). The goal was to identify variations in DNA sequences which correlated well with morphologically defined species and that could be used to develop cost-effective, high throughput molecular methods for routine species identification. Among diatoms, one of the first studies showing that the nuclear encoded internal transcribed spacers (ITS) 1 and 2 and the 5.8S rDNA gene correlates with sexual compatibility was performed for the fresh-water diatom species complex *Sellaphora pupula* (Behnke et al. 2004) where mating experiments helped to distinguish four sexually incompatible phenodemes that later were elevated to separate species (Mann et al. 2004). Adding a mitochondrial marker and others to their study, Evans et al. (2007) showed that the fragment of *cox1* corresponding to the DNA-barcode (Hebert et al. 2003) to be successful but would require group-specific primers.

One of the first studies to suggest the nuclear encoded ITS region carries potential for species-specific markers in microalgae involved volvoccean green algae where interbreeding potential is very difficult to test directly (Coleman et al. 1994). However, alignments proved difficult for taxa higher than species. Nonetheless, Coleman and Mai (1997) demonstrated how the use of the secondary structure of the RNA transcript may

improve alignments. ITS2 in particular was proven to contain sections of significant conservation, including possibly pan-eukaryotic motifs and conserved positions in green algae, terrestrial plants, stramenopiles, chloromonads, cryptophytes, haptophytes, alveolates, red algae, fungi and animals ranging from sponges to mammals (Coleman 2007; Schultz et al. 2005; Seibel et al. 2006). These more conserved sections are presumed necessary for forming the proper secondary structure of the ITS2 transcript, which like its counterparts in ITS1, guides the enzymatic complex process that degrades ITS2 regions, liberating the small ribosomal subunit (SSU), 5.8S and the large ribosomal subunit (LSU) RNAs. The value of ITS2 has been proven at least as a guide in identification of unknown specimens, in the search for closely related organisms, in verification of identity of a culture strain (Coleman and Mai 1997) and recently as a possible barcode region for dinoflagellates (Litaker et al. 2007). In several species of diatoms, particularly the taxonomically difficult and potentially toxigenic diatom genus *Pseudo-nitzschia*, only ITS2 out of three ribosomal and one plastid markers coincided unequivocally with their biologically defined species, including several newly described and semi-cryptic species (Amato et al. 2007).

The goal of our study was to test the feasibility of using the nuclear encoded ITS-region as a marker for recognition and identification of diatoms on a wide range of taxa. We considered 5.8S alone and in combination with ITS2 in species belonging to the two most species-rich classes of diatoms, the Mediophyceae and Bacillariophyceae, the majority of which are simply available in the Culture Collection for Marine Phytoplankton (CCMP). We investigated intragenomic, intraspecific and interspecific distances for all DNA fragments used in this study. Intragenomic versus intraspecific variability was compared in order to ensure that putative variation arising from multiple copies of the nuclear rDNA cistron normally present in any individual does not confound the results, overestimating intraspecific variability. We identified a sufficient, species-specific level of divergence in a short fragment within the 5.8S+ITS2 region and propose it as a DNA-barcode for diatoms.

Results

Our evaluation of the barcoding potential of the ITS region was based on 618 sequences (exclud-

ing repeats); 238 sequences that we obtained in our lab and 380 from GenBank. This data set included 114 species: 56 species belonging to 19 genera, eight families and five orders from the class Mediophyceae and 58 species representing 21 genera, 14 families, and ten orders from the class Bacillariophyceae. Among these the most sequence-rich in GenBank were diatoms from seven species, including four from the genus *Pseudo-nitzschia* (16% of sequences), *Cyclotella meneghiniana* (19%), *Eunotia bilunaris* (6%) and *Navicula phyllepta* (10%). The majority of diversity of species was obtained from sequencing we performed for this study, all obtained through direct sequencing. Methods used to recover sequences available from GenBank are indicated in Table 1; the final tally of sequencing method was: 1) sequences of species obtained only by cloning: 9; only by direct PCR: 99; by both methods: 3, obtained by methods undisclosed by the authors: 8; and 2) sequences of strains obtained from cloning: 104; by direct PCR: 336; and by methods unspecified by the authors: 33.

In order to determine if there was an overlap in genetic variation within and between species we analysed intragenomic, intraspecific and interspecific variability in the 5.8S, 5.8S+ITS2 fragments combined, and a defined region beginning with the start codon of 5.8S and ending in the conserved region of helix III of the secondary structure of the ITS2 transcript. We focused on seven families from Mediophyceae and nine from Bacillariophyceae that were available from CCMP cultures, in order to determine what part of the ITS region offers sufficient resolution to meet the requirements of a DNA-barcode for diatoms.

Alignment Parameters

Alignment parameters are of critical importance when genetic distances are compared and may increase or decrease depending on restrictions and penalties applied. Biologically, gaps are assumed to be insertions or deletions and it should be ensured as much as possible that existing gaps represent true indels and not artificial ones. Therefore, we tested gap penalties within the range that is often used in alignments (open gap penalty of 10.0 and gap extent penalty of 0.1, Hall 2004) and similar to those that were applied in a study of an ITS-barcode in dinoflagellates (open gap penalty of 10.0 and gap extent penalty of 5.0, Litaker et al. 2007). For the test we set four matrices including six araphid species (nine sequences) and 33 raphid species

Table 1. List of strains used in this study; species/strain names listed as they appear in publications, culture collection, and GenBank⁵. Names/words in parentheses indicate the name the sequence was registered either in GenBank or CCMP and were corrected due to taxonomic and/or sequence analyses. Accession numbers in bold identify which sequences were obtained in our lab for this and parallel studies.

Species/Authority (corrected name)	Strain ID	Accession #	Species/Authority (corrected name)	Strain ID	Accession #
Mediophytes					
<i>Attheya longicornis</i> ⁶	ATTPB ³	FJ864273	<i>N. gregaria</i>	K8 ²	DQ235779
Crawford & Gardner			<i>N. gregaria</i>	K5 ²	DQ235780
<i>Arcocellulus mammifer</i>	CCMP0132 ³	GQ330310	<i>N. incerta</i> Grunow ex Van Heurck	CCMP 0542 ³	GQ330356
Hasle, von Stosch & Syvertsen					
<i>Bacteriastrium hyalinum</i> Lauder	CCMP0141 ³	GQ330314	<i>N. phyllepta</i> Kützing	BA3 cl.1 ²	DQ193566
<i>Brockmanniella brockmannii</i> (Hustedt)	CCMP0151 ³	GQ330315	<i>N. phyllepta</i>	BA3 cl.2 ²	DQ193569
Hasle, von Stosch & Syvertsen					
<i>Chaetoceros affinis</i> Lauder	CCMP0159 ³	GQ330316	<i>N. phyllepta</i>	BA3 cl.3 ²	DQ193568
<i>C. calcitrans</i> Takano	CCMP1315 ⁴	DQ358111	<i>N. phyllepta</i>	BA3 cl.4 ²	DQ193567
<i>C. calcitrans</i> f. <i>pumilus</i> Takano	CCAP1010/11 ⁴	DQ358117	<i>N. phyllepta</i>	BA4 cl.1 ²	DQ193565
<i>C. cf. jonquieri</i>	na ⁴	DQ858214	<i>N. phyllepta</i>	BA4 cl.3 ²	DQ193564
<i>C. contortus</i> Schütt	CCMP1578 ³	GQ330318	<i>N. phyllepta</i>	BA4 cl.4 ²	DQ193563
<i>C. curvisetus</i> Cleve	ch.5 ⁴	AY229895	<i>N. phyllepta</i>	BA12 cl.3 ²	DQ193561
<i>C. decipiens</i> Cleve	CCMP0173 ³	GQ330319	<i>N. phyllepta</i>	BA12 cl.2 ²	DQ193562
<i>C. gracilis</i> Pantocsek	ch.18 ⁴	AY229897	<i>N. phyllepta</i>	BA12 cl.4 ²	DQ193560
<i>C. gracilis</i> (<i>C. debilis</i>)	ch.4 ⁴	AY229896	<i>N. phyllepta</i>	K4 cl.12 ⁴	DQ193559
<i>C. gracilis</i>	UTEX LB 2375 ⁴	DQ358112	<i>N. phyllepta</i>	CCY0218 cl.12 ²	DQ193558
<i>C. gracilis</i>	UTEX LB 2658 ⁴	DQ358113	<i>N. phyllepta</i>	CCY0222 cl.11 ²	DQ193557
<i>C. gracilis</i> (<i>Chaetoceros</i> sp.)	NMBguh003-1 ⁴	DQ886408	<i>N. phyllepta</i>	K3 cl.1 ⁴	DQ193554
<i>C. gracilis</i> (<i>Chaetoceros</i> sp.)	NMBguh003-2 ⁴	DQ886409	<i>N. phyllepta</i>	K3 cl.2 ⁴	DQ193553
<i>C. muelleri</i> Lemmermann	CCAP1010/3 ⁴	DQ358116	<i>N. phyllepta</i>	CCY9804 cl.1 ²	DQ193552
<i>C. muelleri</i>	CCMP1316 ³	GQ330320	<i>N. phyllepta</i>	CCY9804 cl.2 ²	DQ193551
<i>C. muelleri</i> (<i>C. neogracile</i>)	CCMP1318 ³	GQ330321	<i>N. phyllepta</i>	CCY 0226 ²	DQ193550
<i>C. muelleri</i> (<i>C. simplex</i>)	CCAP1085/3 ³	DQ358115	<i>N. phyllepta</i>	CCY0213 cl.11 ²	DQ193549
<i>C. muelleri</i> (<i>C. simplex</i> var. <i>calcitrans</i>)	CSIROCS251 ⁴	DQ358114	<i>N. phyllepta</i>	CCY0213 cl.1 ²	DQ193548
<i>C. muelleri</i> (<i>C. calcitrans</i>)	arg.13 ⁴	DQ897644	<i>N. phyllepta</i>	CCY0213 cl.2 ²	DQ193547
<i>C. radicans</i> Schütt	CCMP0197 ³	GQ330322	<i>N. phyllepta</i>	CCY0201 cl.2 ²	DQ193546
<i>C. simplex</i> Ostenfeld	CCMP0200 ³	GQ330323	<i>N. phyllepta</i>	CCY0212 cl.3 ²	DQ193545
<i>C. socialis</i> Lauder	CCMP0203 ³	FJ864274	<i>N. phyllepta</i>	CCY0212 cl.2 ²	DQ193544
<i>C. socialis</i>	CCMP1579 ³	FJ864275	<i>N. phyllepta</i>	CCY0212 cl.1 ²	DQ193543
<i>Cyclotella cryptica</i>	CCMP0331 ³	GQ330325	<i>N. phyllepta</i>	CCY 0227 ²	DQ235783
Reimann, Lewin & Guillard					
<i>C. cryptica</i>	CCMP0332 ³	GQ330324	<i>N. salinicola</i> Hustedt	CCMP1730 ³	GQ330357
<i>C. meneghiniana</i> Kützing	G8W7 cl.1 ²	AY906809	<i>Nitzschia alexandrina</i> (Cholnoky) Lange-Bertalot & Simonsen	WDCM NCC35 ⁴	AY574379
			<i>N. compressa</i> (Bailey) Boyer	WDCM NCC38 ⁴	AY574377
<i>C. meneghiniana</i>	G8W7 cl.2 ²	AY906810			

<i>C. meneghiniana</i>	G8W7 cl.3 ²	AY906811	<i>N. laevis</i> Hustedt	WDCM NCC39 ⁴	AY574378
<i>C. meneghiniana</i>	G8W7 cl.4 ²	AY906812	<i>Pauliella taeniata</i> (Grunow) Round & Basson	CCMP1115 ³	GQ330372
<i>C. meneghiniana</i>	G8W7 cl.5 ²	AY906813	<i>Phaeodactylum</i> <i>tricornutum</i> Bohlin	CCAP1052/1A ³	DQ085802
<i>C. meneghiniana</i>	G17W3 cl.1 ²	AY906805	<i>P. tricornutum</i>	CCAP1052/1B ³	DQ085803
<i>C. meneghiniana</i>	G17W3 cl.2 ²	AY906806	<i>P. tricornutum</i>	CCAP1052/6 ³	DQ085804
<i>C. meneghiniana</i>	G17W3 cl.3 ²	AY906807	<i>P. tricornutum</i>	CCMP0630 cl.1 ³	FJ864293
<i>C. meneghiniana</i>	G17W3 cl.4 ²	AY906808	<i>P. tricornutum</i>	CCMP0630 cl.2 ³	DQ085805
<i>C. meneghiniana</i>	G18W41 cl.1 ²	AY906823	<i>P. tricornutum</i>	CCMP0631 ³	DQ085806
<i>C. meneghiniana</i>	G18W41 cl.2 ²	AY906824	<i>P. tricornutum</i>	CCMP0632 ³	DQ085801
<i>C. meneghiniana</i>	G18W41 cl.3 ²	AY906825	<i>P. tricornutum</i>	CCMP0633 ³	DQ085809
<i>C. meneghiniana</i>	G18W41 cl.4 ²	AY906826	<i>P. tricornutum</i>	CCMP1327 cl.1 ³	GQ330373
<i>C. meneghiniana</i>	G18W41 cl.5 ²	AY906827	<i>P. tricornutum</i>	CCMP1327 cl.2 ³	DQ085807
<i>C. meneghiniana</i>	G18W41 cl.6 ²	AY906828	<i>P. tricornutum</i>	NEPCC640 ³	DQ085808
<i>C. meneghiniana</i>	G18W41 cl.7 ²	AY906829	<i>P. tricornutum</i>	Pt10 ⁴	DQ655656
<i>C. meneghiniana</i>	G183K cl.1 ²	AY906814	<i>P. tricornutum</i>	MACCb228 ³	AY574376
<i>C. meneghiniana</i>	G183K cl.2 ²	AY906815	<i>Pseudo-nitzschia</i> <i>americana</i> ⁶ (Hasle) Fryxell	WDCM NCC45 ⁴	GQ330380
<i>C. meneghiniana</i>	G183K cl.3 ²	AY906816	<i>P. americana</i> ⁶	WB224Aug ³	GQ330381
<i>C. meneghiniana</i>	G183K cl.4 ²	AY906817	<i>P. americana</i> ⁶	WB313July ³	GQ330378
<i>C. meneghiniana</i>	G183K cl.5 ²	AY906818	<i>P. americana</i> ⁶	WA424Aug ³	GQ330379
<i>C. meneghiniana</i>	G188D cl.1 ²	AY906819	<i>P. australis</i> Frenguelli	IO 93-01 ³	GQ330382
<i>C. meneghiniana</i>	G184 cl.1 ²	AY906798	<i>P. australis</i>	IO 84-01 ³	GQ330383
<i>C. meneghiniana</i>	G184 cl.2 ²	AY906799	<i>P. calliantha</i> Lundholm, Moestrup & Hasle	AL-56 ³	DQ813834
<i>C. meneghiniana</i>	G184 cl.3 ²	AY906800	<i>P. calliantha</i> Lundholm, Moestrup & Hasle	C-AL-1 ³	DQ813842
<i>C. meneghiniana</i>	G184 cl.4 ²	AY906801	<i>P. calliantha</i>	AL-101 ³	DQ813839
<i>C. meneghiniana</i>	G184 cl.5 ²	AY906802	<i>P. calliantha</i>	P5 ³	as DQ813839
<i>C. meneghiniana</i>	G184 cl.6 ²	AY906803	<i>P. calliantha</i>	P6 ³	as DQ813839
<i>C. meneghiniana</i>	G184 cl.7 ²	AY906804	<i>P. calliantha</i>	AL-112 ³	DQ813841
<i>C. meneghiniana</i>	G8W4 cl.1 ²	AY906794	<i>P. calliantha</i>	AL-117 ³	as DQ813841
<i>C. meneghiniana</i>	G8W4 cl.2 ²	AY906795	<i>Pseudo-nitzschia</i> sp. 1 (<i>P. cf. granii</i>)	CCMP2093 ³	GQ330384
<i>C. meneghiniana</i>	G8W4 cl.3 ²	AY906796	<i>Pseudo-nitzschia</i> sp. 1 (<i>P. seriata</i>)	CCMP1309 ³	GQ330428
<i>C. meneghiniana</i>	G8W4 cl.4 ²	AY906797	<i>P. cuspidata</i> (Hasle) Hasle <i>emend.</i> Lundholm, Moestrup & Hasle	AL-17 ³	DQ813827
<i>C. meneghiniana</i>	G8W6 cl.1 ²	AY906789	<i>P. cuspidata</i>	AL-28 ³	as DQ813827
<i>C. meneghiniana</i>	G8W6 cl.2 ²	AY906790	<i>P. cuspidata</i>	AL-57 ³	as DQ813827

Table 1. (continued)

Species/Authority (corrected name)	Strain ID	Accession #	Species/Authority (corrected name)	Strain ID	Accession #
<i>C. meneghiniana</i>	G8W6 cl.3 ²	AY906791	<i>P. cuspidata</i>	AL-61 ³	as DQ813827
<i>C. meneghiniana</i>	G8W6 cl.4 ²	AY906792	<i>P. delicatissima</i> (Cleve) Heiden	PLY1St.85Z ³	AM118052
<i>C. meneghiniana</i>	G8W6 cl.5 ²	AY906793	<i>P. delicatissima</i>	PLY1St.85B ³	AM118051
<i>C. meneghiniana</i>	G8W5 cl.1 ²	AY906786	<i>P. delicatissima</i>	PLY1St.85A ³	AM118050
<i>C. meneghiniana</i>	G8W5 cl.2 ²	AY906787	<i>P. delicatissima</i>	PLY1St.77C ³	AM118049
<i>C. meneghiniana</i>	G8W5 cl.3 ²	AY906788	<i>P. delicatissima</i>	PLY1St.48A ³	AM118048
<i>C. meneghiniana</i>	G183N cl.1 ²	AY906780	<i>P. delicatissima</i>	PLY1St.46A ³	AM118047
<i>C. meneghiniana</i>	G183N cl.2 ²	AY906781	<i>P. delicatissima</i>	PLY1St.45E ³	AM118046
<i>C. meneghiniana</i>	G183N cl.3 ²	AY906782	<i>P. delicatissima</i>	PLY1St.43C ³	AM118045
<i>C. meneghiniana</i>	G183N cl.4 ²	AY906783	<i>P. delicatissima</i>	PLY1St.42A ³	AM118044
<i>C. meneghiniana</i>	G183N cl.5 ²	AY906784	<i>P. delicatissima</i>	PLY1St.5 ³	AM118043
<i>C. meneghiniana</i>	G183N cl.6 ²	AY906785	<i>P. delicatissima</i> ⁶	W002 ³	EU599146
<i>C. meneghiniana</i>	G16W1 cl.1 ²	AY906776	<i>P. delicatissima</i>	LaesØ2 ³	DQ329206
<i>C. meneghiniana</i>	G16W1 cl.2 ²	AY906777	<i>P. delicatissima</i>	ØM2 ³	DQ329207
<i>Detonula confervacea</i> (Cleve) Gran	CCMP0352 ³	GQ330329	<i>P. delicatissima</i> ⁶	A3 A3 22F ³	EU599158
<i>D. confervacea</i>	CCMP0353 ³	GQ330330	<i>P. delicatissima</i> ⁶	A3 B2 22F ³	EU599160
<i>Ditylum brightwellii</i> (West) Grunow	CCMP0361 ³	GQ330332	<i>P. delicatissima</i> ⁶	A2 A2 22F ³	EU599151
<i>D. brightwellii</i> ⁶	SFPB44A1 ³	GQ330335	<i>P. delicatissima</i> ⁶	C4 B1 28F ³	GQ330385
<i>D. brightwellii</i> ⁶	SFPB44B4 ³	GQ330333	<i>P. delicatissima</i> ⁶	A2 A3 22F ³	EU599152
<i>D. brightwellii</i> ⁶	SFPB44C2 ³	GQ330334	<i>P. delicatissima</i> ⁶	W007 B2 ³	EU599148
<i>D. brightwellii</i>	CCMP0356 ³	FJ864280	<i>P. delicatissima</i> ⁶	W007 B3 ³	EU599149
<i>D. brightwellii</i>	CCMP0357 ³	FJ864279	<i>P. delicatissima</i> ⁶	W007 C2 ³	EU599150
<i>D. brightwellii</i>	CCMP0358 ³	FJ864283	<i>P. delicatissima</i> ⁶	W001 II B2 ³	GQ330417
<i>D. brightwellii</i>	CCMP0359 ³	FJ864281	<i>P. delicatissima</i> ⁶	C1 B4 24F ³	EU599171
<i>D. brightwellii</i>	CCMP1582 ³	GQ330331	<i>P. delicatissima</i> ⁶	W001 I B2 ³	EU599180
<i>D. brightwellii</i>	CCMP1810 ³	FJ864284	<i>P. delicatissima</i> ⁶	W001 II C2 ³	EU599145
<i>D. brightwellii</i> ⁶	DitMartin ³	FJ864282	<i>P. delicatissima</i> ⁶	W003 ³	EU599147
<i>D. brightwellii</i>	PSB3M_1 ³	DQ329268	<i>P. delicatissima</i> ⁶	A2 B2 24J ³	EU599154
<i>D. brightwellii</i> (<i>C. closterium</i>)	CCMP0339 ³	as FJ864284	<i>P. delicatissima</i> ⁶	B3 B3 24F ³	EU599181
<i>Extubocellulus cribriger</i>	CCMP0391 ³	GQ330336	<i>P. delicatissima</i> ⁶	B1 B2 23F ³	EU599162
Hasle, von Stosch & Syvertsen <i>E. spinifer</i>	CCMP0393 ³	GQ330338	<i>P. delicatissima</i> ⁶	A2 B3 24J ³	EU599155
Hasle, von Stosch & Syvertsen <i>Helicotheca tamesis</i>	CCMP0825 ³	GQ330347	<i>P. delicatissima</i> ⁶	B3 A2 24F ³	EU599165
(<i>Shrubsole</i>) Ricard <i>H. tamesis</i>	CCMP1687 ³	GQ330348	<i>P. delicatissima</i> ⁶	C2 B2 28F ³	EU599172
<i>Leyanella arenaria</i>	CCMP0471 ³	GQ330349	<i>P. delicatissima</i> ⁶	A2 B1 24J ³	EU599153
Hasle, von Stosch & Syvertsen <i>Lithodesmium undulatum</i> Ehrenberg	CCMP1797 ³	GQ330350	<i>P. delicatissima</i> ⁶	B3 B2 19J ³	EU599166
<i>Minidiscus trioculatus</i> (Taylor) Hasle	CCMP0496 ³	GQ330352	<i>P. delicatissima</i> ⁶	A2 B3 22F ³	EU599156
<i>M. trioculatus</i> var. <i>monoculatus</i> ⁶	Mini nova1 ³	GQ330351	<i>P. delicatissima</i> ⁶	A3 A3 31J ³	EU599157
Kaczmarek					

<i>M. trioculatus</i> var. <i>monoculatus</i> ⁶	Mini nova 2 ³	as GQ330351	<i>P. delicatissima</i> ⁶	A3 B2 31J ³	EU599159
<i>M. variabilis</i> Kaczmarska	CCMP0495 1 ³	GQ330353	<i>P. delicatissima</i> ⁶	B3 B2 24F ³	EU599167
<i>M. variabilis</i>	CCMP0495 2 ³	as GQ330353	<i>P. delicatissima</i> ⁶	B2 B2 23F ³	EU599163
<i>M. variabilis</i> f. <i>inornata</i> Kaczmarska ⁶	IKEMs011706 ³	GQ330354	<i>P. delicatissima</i> ⁶	B2 B3 23F ³	EU599164
<i>Minutocellus polymorphus</i> (Hargraves & Guillard) Hasle, von Stosch, & Syvertsen	CCMP0499 ³	FJ864292	<i>P. delicatissima</i> ⁶	BC002 ³	EU599174
<i>Odontella aurita</i> ⁶ (Lyngbye) Agardh	IKEOa011706 ³	GQ330369	<i>P. delicatissima</i> ⁶	DM002 ³	EU599176
<i>O. aurita</i>	CCMP0145 ³	GQ330365	<i>P. delicatissima</i> ⁶	DM004 ³	EU599177
<i>O. aurita</i>	CCMP1108 ³	GQ330364	<i>P. delicatissima</i> ⁶	LK001 ³	EU599175
<i>O. aurita</i>	CCMP0816 ³	GQ330366	<i>P. delicatissima</i> ⁶	BBA4 ³	EU599178
<i>O. longicruris</i> (Greville) Hoban	CCMP1808 ³	GQ330367	<i>P. delicatissima</i> ⁶	CT B3 ³	EU599179
<i>O. mobiliensis</i> (Bailey) Grunow	CCMP0598 ³	GQ330368	<i>P. delicatissima</i>	21-02 ³	AY519350
<i>O. mobiliensis</i>	CCMP0146 ³	GQ330370	<i>P. delicatissima</i>	AL-22 ³	DQ813829
<i>O. cf. sinensis</i>	CCMP1815 ³	GQ330371	<i>P. delicatissima</i>	AL-38 ³	DQ813832
<i>Porosira glacialis</i> (Grunow) Jorgensen	CCMP1274 ³	GQ330377	<i>P. delicatissima</i>	S-AL-1 ³	DQ813843
<i>P. glacialis</i>	CCMP0670 ³	GQ330376	<i>P. delicatissima</i> (del1)	18-02 ³	AY519334
			(Cleve) Heiden, in part ¹		
<i>P. glacialis</i>	CCMP0980 ³	GQ330375	<i>P. delicatissima</i> (del1)	24-02 ³	AY519335
<i>P. glacialis</i>	CCMP1099 ³	GQ330374	<i>P. delicatissima</i> (del1)	26-02 ³	AY519337
<i>Skeletonema dohrni</i> Sarno & Kooistra	CCMP 2479 ⁴	EF433513	<i>P. delicatissima</i> (del1)	27-02 ³	AY519336
<i>S. dohrnii</i> (<i>Skeletonema</i> sp.)	CCMP1715 ⁴	EF433512	<i>P. delicatissima</i> (del1)	AL-24 ³	DQ813830
<i>S. dohrnii</i>	CCMP0789 ³	GQ330440	<i>P. delicatissima</i> (del1)	AL-110 ³	DQ813840
<i>S. dohrnii</i>	CCMP0785 ³	GQ330432	<i>P. dolorosa</i> Lundholm & Moestrup	20-02 ³	AY519275
<i>S. dohrnii</i>	CCMP0779 ³	GQ330434	<i>P. dolorosa</i>	AL-59 ³	DQ813835
<i>S. dohrnii</i>	CCMP0782 ³	GQ330435	<i>P. dolorosa</i>	AL-67 ³	DQ813837
<i>S. grethae</i> Zingone & Sarno	CCMP0775 ³	GQ330437	<i>P. dolorosa</i>	AL-74 ³	DQ813838
<i>S. grethae</i>	CCMP0776 ³	GQ330436	<i>P. dolorosa</i>	C-AL 2 ³	as DQ813835
<i>S. marinoi</i> Sarno & Zingone	CCMP0781 ³	GQ330433	<i>P. fraudulenta</i> (Cleve)	IO 83-03 ³	GQ330422
			Hasle		
<i>S. marinoi</i>	CCMP1009 ⁴	EF433511	<i>P. multiseriata</i> (Hasle)	CCMP2708 ³	GQ330424
			Hasle		
<i>S. marinoi</i>	CCMP2507 ⁴	EF433517	<i>P. multistriata</i> (Takano)	IO 92-01 ³	GQ330426
			Takano		
<i>S. marinoi</i>	CCMP2505 ⁴	EF433516	<i>P. multistriata</i>	cl. A/B4 ²	EF636681
<i>S. marinoi</i>	KAL N3 ³	AY954684	<i>P. multistriata</i>	cl. A/B3 ²	EF636680
<i>S. marinoi</i>	KAL M2 ³	AY954683	<i>P. multistriata</i>	cl. A/B2 ²	EF636679
<i>S. marinoi</i>	KAL H1 ³	AY954682	<i>P. multistriata</i>	cl. A/B1 ²	EF636678
<i>S. marinoi</i>	CCMP2497 ⁴	EF433515	<i>P. multistriata</i>	cl. B5 ²	EF636677
<i>S. marinoi</i> (<i>S. costatum</i>)	Ifremer-Bouin ⁴	DQ897642	<i>P. multistriata</i>	cl. B4 ²	EF636676
<i>S. marinoi</i> (<i>S. costatum</i>)	KFU1 ³	AY748228	<i>P. multistriata</i>	cl. B3 ²	EF636675
<i>S. marinoi</i> (<i>S. costatum</i>)	KFP2 ³	AY748227	<i>P. multistriata</i>	cl. B2 ²	EF636674
<i>S. marinoi</i> (<i>S. costatum</i>)	KFN2 ³	AY748226	<i>P. multistriata</i>	cl. B1 ²	EF636673
<i>S. marinoi</i> (<i>S. costatum</i>)	KFC36 ³	AY748225	<i>P. multistriata</i>	cl. A ²	EF636672
<i>S. marinoi</i> (<i>S. costatum</i>)	KFC23 ³	AY748222	<i>P. pseudodelicatissima</i> (Hasle) Hasle emend.	AL-60 ³	DQ813836

Table 1. (continued)

Species/Authority (corrected name)	Strain ID	Accession #	Species/Authority (corrected name)	Strain ID	Accession #
<i>S. marinoi</i> (<i>S. costatum</i>)	KFC33 ³	AY748224	Lundholm, Hasle & Moestrup		
<i>S. marinoi</i> (<i>S. costatum</i>)	KFC27 ³	AY748223	<i>P. pseudodelicatissima</i>	AL-41 ³	DQ813833
<i>S. marinoi</i> (<i>S. costatum</i>)	PORTF1 ³	AY748237	<i>P. pseudodelicatissima</i>	AL-29 ³	DQ813831
<i>S. marinoi</i> (<i>S. costatum</i>)	PORTJ1 ³	AY748234	<i>P. pseudodelicatissima</i>	AL-19 ³	DQ813828
<i>S. marinoi</i> (<i>S. costatum</i>)	PORTG3 ³	AY748233	<i>P. pseudodelicatissima</i>	AL-15 ³	DQ813826
<i>S. marinoi</i> (<i>S. costatum</i>)	PORTD2 ³	AY748232	<i>P. pseudodelicatissima</i>	SZN-B109 ²	AY550126
<i>S. marinoi</i> (<i>S. costatum</i>)	PORTD1 ³	AY748231	<i>P. pseudodelicatissima</i>	SZN-B111 ²	AY550128
<i>S. marinoi</i> (<i>S. costatum</i>)	PORTC3 ³	AY748230	<i>P. pseudodelicatissima</i>	AL-21 ³	as DQ813833
<i>S. marinoi</i> (<i>S. costatum</i>)	PORTA1 ³	AY748229	<i>P. pseudodelicatissima</i>	AL-20 ³	as DQ813828
<i>S. marinoi</i> (<i>S. costatum</i>)	A1 ³	AY748226	<i>P. pseudodelicatissima</i>	AL-27 ³	as DQ813828
<i>S. marinoi</i> (<i>S. costatum</i>)	CANK2 ³	AY748221	<i>P. pseudodelicatissima</i>	AL-40 ³	as DQ813828
			<i>P. pungens</i> (Grunow ex Cleve) Hasle	CT B1 ³	EU599141
<i>S. marinoi</i> (<i>S. costatum</i>)	CANJ1 ³	AY748220	<i>P. pungens</i>	NA233 cl.20 ²	FM207625
<i>S. marinoi</i> (<i>S. costatum</i>)	CANG1 ³	AY748219	<i>P. pungens</i>	NA233 cl.15 ²	FM207624
<i>S. marinoi</i> (<i>S. costatum</i>)	CAND2 ³	AY748218	<i>P. pungens</i>	NA233 cl.14 ²	FM207623
<i>S. marinoi</i> (<i>S. costatum</i>)	CANB2 ³	AY748217	<i>P. pungens</i>	NA233 cl.10 ²	FM207622
<i>S. marinoi</i> (<i>S. pseudocostatum</i>)	CCAP1077/7 ⁴	DQ897643	<i>P. pungens</i>	NA231 cl.9 ²	FM207621
<i>S. menzellii</i>	CCMP0786 ³	GQ330439	<i>P. pungens</i>	NA231 cl.8 ²	FM207620
Guillard, Carpenter & Reim					
<i>S. menzellii</i>	CCMP0787 ³	GQ330438	<i>P. pungens</i>	NA204 cl.12 ²	FM207619
<i>S. menzellii</i>	CCMP0792 ⁴	DQ897643	<i>P. pungens</i>	NA179 cl.3 ²	FM207618
<i>Skeletonema</i> sp.	M2 ⁴	DQ280324	<i>P. pungens</i>	NA179 cl.7 ²	FM207617
<i>Skeletonema</i> sp.	GFC-2005.5 ⁴	DQ280328	<i>P. pungens</i>	NA179 cl.9 ²	FM207616
<i>Skeletonema</i> sp.	GFC-2005.4 ⁴	DQ280327	<i>P. pungens</i>	NA179 cl.10 ²	FM207615
<i>Skeletonema</i> sp.	GFC-2005.3 ⁴	DQ280326	<i>P. pungens</i>	NA178 cl.28 ²	FM207614
<i>Skeletonema</i> sp.	GFC-2005.1 ⁴	DQ280324	<i>P. pungens</i>	NA178 cl.27 ²	FM207613
<i>Stephanodiscus hantzschii</i> Grunow	na ³	U03078	<i>P. pungens</i>	NA178 cl.17 ²	FM207612
<i>S. niagarae</i> Ehrenberg	I132 ³	U03074	<i>P. pungens</i>	NA241 cl.21 ²	FM207611
<i>S. niagarae</i>	N1P1 ³	U03076	<i>P. pungens</i>	NA241 cl.19 ²	FM207610
<i>S. niagarae</i>	LL16 ³	U03075	<i>P. pungens</i>	NA192 cl.37 ²	FM207609
<i>S. yellowstonensis</i> Simonsen	na ³	U03077	<i>P. pungens</i>	NA192 cl.30 ²	FM207608
<i>Thalassiosira aestivalis</i> Gran	C1 ³	EF208797	<i>P. pungens</i>	NA178 cl.20 ²	FM207607
<i>T. angulata</i> ⁶ (Gregory) Hasle	IIIB2 ³	GQ330450	<i>P. pungens</i>	NA178 cl.18 ²	FM207606
<i>T. angulata</i> ⁶	IIIB3 ³	GQ330448	<i>P. pungens</i>	NA178 cl.16 ²	FM207605
<i>T. angulata</i> ⁶	IIC3 ³	GQ330446	<i>P. pungens</i>	NA116 cl.45 ²	FM207604
<i>T. angulata</i> ⁶	PBC2 ³	GQ330445	<i>P. pungens</i>	NA116 cl.9 ²	FM207603
<i>T. angulata</i> ⁶	PBA1 ³	GQ330449	<i>P. pungens</i>	NA116 cl.48 ²	FM207602
<i>T. angulata</i> ⁶	PBC3 ³	GQ330447	<i>P. pungens</i>	NA233 cl.11 ²	FM207601
<i>T. anguste-lineata</i> (Schmidt) Fryxell & Hasle	C3 ³	EF208800	<i>P. pungens</i>	NA233 cl.7 ²	FM207600
<i>T. antarctica</i> Comber	CCMP0982 ³	GQ330452	<i>P. pungens</i>	NA200 cl.1 ²	FM207599

<i>T. floridana</i> (Cooper) Hasle	CCMP0985 ³	GQ330453	<i>P. pungens</i>	NA200 cl.6 ²	FM207598
<i>T. gravida</i> Cleve	CCMP0986 ³	GQ330454	<i>P. pungens</i>	NA200 cl.17 ²	FM207597
<i>T. guillardii</i> Hasle	CCMP0988 ³	GQ330455	<i>P. pungens</i>	NA200 cl.10 ²	FM207596
<i>T. minima</i> Gaarder	CCMP0991 ³	GQ330468	<i>P. pungens</i>	NA179 cl.1 ²	FM207595
<i>T. minuscula</i> Krasske	CCMP1093 ³	GQ330456	<i>P. pungens</i>	NA177 cl.7 ²	FM207594
<i>T. nordenskiöldii</i> Cleve	1B3 ³	GQ330458	<i>P. pungens</i>	NA177 cl.16 ²	FM207593
<i>T. nordenskiöldii</i>	1C4 ³	GQ330457	<i>P. pungens</i>	NA177 cl.25 ²	FM207592
<i>T. nordenskiöldii</i>	2C4 ³	GQ330459	<i>P. pungens</i>	NA2 ³	FM207591
<i>T. nordenskiöldii</i>	CCMP0997 ³	FJ864301	<i>P. pungens</i>	NA1 ³	FM207590
<i>T. oceanica</i> Hasle	CCMP1006 cl.1 ²	EF362631	<i>P. pungens</i>	NA233 cl.23 ²	FM207626
<i>T. oceanica</i>	CCMP1006 cl.2 ²	EF362632	<i>P. pungens</i>	NA242 cl.24 ²	FM207628
<i>T. oceanica</i>	CCMP1006 cl.3 ²	EF362633	<i>P. pungens</i>	NA242 cl.21 ²	FM207627
<i>T. oceanica</i>	CCMP1001 ³	GQ330461	<i>P. pungens</i>	Jp-14 ³	AM778814
<i>T. oceanica</i> ⁶	IKET062105 ³	GQ330460	<i>P. pungens</i>	Jp-1 ³	AM778812
<i>T. oceanica</i>	CCMP1004 ²	EF208794	<i>P. pungens</i>	US-96 ³	AM778803
<i>T. proschkiniae</i> ⁶ Makarova	IKETp091305 ³	GQ330462	<i>P. pungens</i>	US-94/f ²	AM778802
<i>T. pseudonana</i> Hasle & Heimdal	CCMP1007 ³	as GQ330463	<i>P. pungens</i>	US-94/e ²	AM778801
<i>T. pseudonana</i>	CCMP1011 ³	EF208789	<i>P. pungens</i>	US-94/d ²	AM778800
<i>T. pseudonana</i>	CCMP1012 ³	EF208790	<i>P. pungens</i>	US-94/c ²	AM778799
<i>T. pseudonana</i>	CCMP1014 ³	EF208791	<i>P. pungens</i>	US-94/b ²	AM778798
<i>T. pseudonana</i>	CCMP1015 ³	EF208792	<i>P. pungens</i>	US-94/a ²	AM778797
<i>T. pseudonana</i>	CCMP1335 ³	EF208793	<i>P. pungens</i>	Cn-218 ³	AM778794
<i>T. pseudonana</i>	CCMP1335 ³	GQ330463	<i>P. pungens</i>	Cn-213 ³	AM778791
<i>T. punctigera</i> ⁶ (Castracane) Hasle	IKETp121404 ³	FJ864303	<i>P. pungens</i>	Cn-193 ³	AM778788
<i>T. rotula</i> Meunier	CCMP1647 ³	GQ330466	<i>P. pungens</i>	Cn-172 ³	AM778786
<i>T. rotula</i>	CCMP1018 ³	GQ330465	<i>P. pungens</i>	Vigo-4 ³	AM778785
<i>T. rotula</i>	CCMP1812 ³	GQ330467	<i>P. pungens</i>	Vigo-3/g ²	AM778784
<i>T. weissflogii</i>	CCMP1010 ³	EF208782	<i>P. pungens</i>	Vigo-3/f ²	AM778783
(Grunow) Fryxell & Hasle					
<i>T. weissflogii</i>	CCMP1047 ³	EF208779	<i>P. pungens</i>	Vigo-3/e ²	AM778782
<i>T. weissflogii</i>	CCMP1048 ³	EF208780	<i>P. pungens</i>	Vigo-3/d ²	AM778781
<i>T. weissflogii</i>	CCMP1049 ³	GQ330470	<i>P. pungens</i>	Vigo-3/c ²	AM778780
<i>T. weissflogii</i>	CCMP1051 ³	EF208787	<i>P. pungens</i>	Vigo-3/b ²	AM778779
<i>T. weissflogii</i>	CCMP1052 ³	EF208784	<i>P. pungens</i>	Vigo-3/a ²	AM778778
<i>T. weissflogii</i>	CCMP1053 ³	EF208785	<i>P. pungens</i>	Vigo-2 ³	AM778777
<i>T. weissflogii</i>	CCMP1336 cl.1 ³	EU029232	<i>P. pungens</i>	Vigo-1 ³	AM778776
<i>T. weissflogii</i>	CCMP1336 cl.2 ³	GQ330471	<i>P. pungens</i>	S(4)6/g ²	AM778767
<i>T. weissflogii</i>	CCMP1587 cl.1 ³	FJ864303	<i>P. pungens</i>	S(4)6/f ²	AM778766
<i>T. weissflogii</i>	CCMP1587 cl.2 ³	as FJ864303	<i>P. pungens</i>	S(4)6/e ²	AM778765
<i>T. weissflogii</i>	1587C4 ³	EU022313	<i>P. pungens</i>	S(4)6/d ²	AM778764
<i>T. weissflogii</i>	1587C2 ³	EU022312	<i>P. pungens</i>	S(4)6/c ²	AM778763
<i>T. weissflogii</i>	BILB2001 cl.C ³	EU018147	<i>P. pungens</i>	S(4)6/b ²	AM778762
<i>T. weissflogii</i>	BILB2001 cl.A ³	EU018146	<i>P. pungens</i>	6 ³	AM778735
<i>T. weissflogii</i>	BILIB2001 ³	EF208786	<i>P. pungens</i>	NZ-74 ³	AM778817
unid. sp., centric	CCMP2018 ³	GQ330337	<i>P. pungens</i>	NZ-67 ³	AM778816
unid. sp., centric	CCMP2125 ³	GQ330339	<i>P. pungens</i>	NZ-49 ³	AM778815
unid. sp., centric	CCMP2063 ³	GQ330464	<i>P. pungens</i>	Jp-11 ³	AM778813
unid. sp., centric	CCMP1277 ³	GQ330469	<i>P. pungens</i>	US-135 ³	AM778811

Table 1. (continued)

Species/Authority (corrected name)	Strain ID	Accession #	Species/Authority (corrected name)	Strain ID	Accession #
Pennates					
<i>Achnanthes brevipes</i> Agardh	CCMP0100 ³	GQ330304	<i>P. pungens</i>	US-132 ³	AM778810
<i>A. longipes</i> Agardh	CCMP0101 ³	GQ330305	<i>P. pungens</i>	US-123/e ²	AM778809
<i>Amphora coffeaeformis</i> ⁶	IKEAc ³	GQ330306	<i>P. pungens</i>	US-123/d ²	AM778808
(Agardh) Kützing			<i>P. pungens</i>	US-123/c ²	AM778807
<i>A. coffeaeformis</i>	CCMP0127 ³	GQ330307	<i>P. pungens</i>	US-123/b ²	AM778806
<i>A. coffeaeformis</i>	CCMP1405 ³	GQ330309	<i>P. subpacifica</i> (Hasle)	IO 81-01 ³	GQ330430
<i>A. salina</i> Smith	CCMP1119 ³	GQ330308	Hasle		
<i>Asterionellopsis glacialis</i>	WDCM NCC5 ⁴	AY574380	<i>Rhabdonema</i> sp. ⁶	RhCTIIIA4 ³	GQ330431
			<i>Sellaphora auldreekie</i>	SMALL 8B c ²	AJ544679
			Mann & McDonald		
			(<i>S. pupula</i>)		
(Castracane) Hasle (<i>Asterionella</i> sp.)					
<i>A. glacialis</i>	CCMP0139 ³	FJ864272	<i>S. auldreekie</i> (<i>S. pupula</i>)	SMALL 8B b ²	AJ544678
<i>A. glacialis</i>	CCMP1581 ³	GQ330312	<i>S. auldreekie</i> (<i>S. pupula</i>)	SMALL 8B a ²	AJ544677
<i>A. glacialis</i>	CCMP0135 ³	GQ330311	<i>S. blackfordensis</i>	RECT-6	AJ544666
			Mann & Droop	(BS14xBM44) ⁴	
			(<i>S. pupula</i>)		
<i>A. glacialis</i>	CCMP1717 ³	GQ330313	<i>S. blackfordensis</i>	RECT-5 (BAL7) ⁴	AJ544665
			(<i>S. pupula</i>)		
<i>Cylindrotheca closterium</i>	CCMP1725 ³	FJ864277	<i>S. blackfordensis</i>	RECT-4(BS7) ³	AJ544664
			(<i>S. pupula</i>)		
(Ehrenberg) Lewin & Reimann					
<i>C. closterium</i>	CCMP2086 ³	GQ330326	<i>S. blackfordensis</i>	RECT-4(BS7)3a ²	AJ544663
			(<i>S. pupula</i>)		
<i>C. fusiformis</i> Reimann & Lewin	CCMP0344 ³	FJ864278	<i>S. blackfordensis</i>	RECT-3 BLA17 ²	AJ544662
			(<i>S. pupula</i>)		
<i>C. fusiformis</i>	CCMP0343 ³	GQ330327	<i>S. blackfordensis</i>	RECT-2 ²	AJ544661
			(<i>S. pupula</i>)		
<i>Delphineis</i> sp.	CCMP1095 ³	GQ330328	<i>S. blackfordensis</i>	RECT-1 BS13 ²	AJ544660
			(<i>S. pupula</i>)		
<i>Eunotia bilunaris</i> (Ehrenberg) Mills	DM22-5 ²	AM747216	<i>S. capitata</i> Mann & McDonald	CAP ²	AJ544651
			(<i>S. pupula</i>)		
<i>E. bilunaris</i>	DM22-4 ²	AM747215	<i>S. pupula</i> (Kützing)	BLUNTTHR15 ²	AJ544676
			Mereschkovskiy deme blunt		
<i>E. bilunaris</i>	DM33-2 cl.a ²	AM747214	<i>S. pupula</i> deme small blunt capitata	SM-BLCAP b ²	AJ544675
<i>E. bilunaris</i>	DM33-2 cl.c ²	AM747213	<i>S. pupula</i> deme small blunt capitata	SM-BLCAP a ²	AJ544674
<i>E. bilunaris</i>	DM33-2 cl.b ²	AM747212	<i>S. pupula</i> deme pseudocapitata	PSEUDOCAP-4 ²	AJ544673

<i>E. bilunaris</i>	DM33-17 ²	AM747211	<i>S. pupula</i> deme pseudocapitate	PSEUDOCAP-3 ²	AJ544672
<i>E. bilunaris</i>	DM33-16 ²	AM747210	<i>S. pupula</i> deme pseudocapitate	PSEUDOCAP-2 ²	AJ544671
<i>E. bilunaris</i>	DM33-1 cl.e ²	AM747209	<i>S. pupula</i> deme pseudocapitate	PSEUDOCAP-1 18TmNE d ²	AJ544670
<i>E. bilunaris</i>	DM33-1 cl.d ²	AM747208	<i>S. pupula</i> deme pseudocapitate	PSEUDOCAP-1 18TmNE c ²	AJ544668
<i>E. bilunaris</i>	DM33-1 cl.c ²	AM747207	<i>S. pupula</i> deme pseudocapitate	PSEUDOCAP-1 18TmNE b ²	AJ544667
<i>E. bilunaris</i>	DM33-1 cl.b ²	AM747206	<i>S. pupula</i> deme pseudocapitate	PSEUDOCAP-1 18TmNE a ²	AJ544667
<i>E. bilunaris</i>	DM33-1 cl.a ²	AM747205	<i>Stauroneis amphorooides</i> Grunow	CCMP797 ³	GQ330441
<i>E. bilunaris</i>	DM33-8 ²	AM747204	<i>S. constricta</i> Ehrenberg	CCMP1120 ³	GQ330442
<i>E. bilunaris</i>	DM33-9 cl.d ²	AM747203	<i>Synedra fragilarioides</i> Fritsch & Rich	CCMP0844 ³	GQ330443
<i>E. bilunaris</i>	DM33-5 cl.c ²	AM747202	<i>Synedropsis hyperborea</i> (Grunow) Hasle, Medlin & Syvertsen	CCMP1422 ³	FJ878725
<i>E. bilunaris</i>	DM33-5 cl.e ²	AM747201	<i>Tabularia affinis</i> ⁶ (Kützing) Snoeijs	206A uk ³	EU883366
<i>E. bilunaris</i>	DM33-5 cl.d ²	AM747200	<i>T. fasciculata</i> ⁶ (Agardh) Williams & Round	T0206A ³	FJ878740
<i>E. bilunaris</i>	DM33-5 cl.b ²	AM747199	<i>T. fasciculata</i> ⁶	T0206B ³	FJ878739
<i>E. bilunaris</i>	DM33-5 cl.a ²	AM747198	<i>T. fasciculata</i> ⁶	T0206E ³	FJ878764
<i>E. bilunaris</i>	DM33-9 cl.e ²	AM747197	<i>T. fasciculata</i> ⁶	T0206F ³	FJ878763
<i>E. bilunaris</i>	DM33-9 cl.b ²	AM747196	<i>T. fasciculata</i> ⁶	T0207U ³	FJ878757
<i>E. bilunaris</i>	DM33-9 cl.c ²	AM747195	<i>T. fasciculata</i> ⁶	T0208A ³	FJ878728
<i>E. bilunaris</i>	DM33-9 cl.a ²	AM747194	<i>T. fasciculata</i> ⁶	T0208AA ³	FJ878727
<i>E. bilunaris</i>	DM33-18 cl.a ²	AM747193	<i>T. fasciculata</i> ⁶	T0208AB ³	EU883356
<i>E. bilunaris</i>	DM33-18 cl.c ²	AM747192	<i>T. fasciculata</i> ⁶	T0208ABS ³	FJ878741
<i>E. bilunaris</i>	DM33-18 cl.d ²	AM747191	<i>T. fasciculata</i> ⁶	T0208AD ³	FJ878743
<i>E. bilunaris</i>	DM33-18 cl.b ²	AM747190	<i>T. fasciculata</i> ⁶	T0208AE ³	EU883360
<i>E. bilunaris</i>	BST1 ²	AM234666	<i>T. fasciculata</i> ⁶	T0208AF ³	EU883361
<i>E. bilunaris</i>	BST10 ²	AM234668	<i>T. fasciculata</i> ⁶	T0208AL ³	as EU883361
<i>E. bilunaris</i>	BST19 ²	AM234669	<i>T. fasciculata</i> ⁶	T0208W ³	FJ878750
<i>E. bilunaris</i>	BST6 ²	AM234667	<i>T. fasciculata</i> ⁶	T0208 × ³	FJ878749
<i>E. bilunaris</i>	JP11 a ²	AM234654	<i>T. fasciculata</i> ⁶	T0214AA ³	FJ878756
<i>E. bilunaris</i>	JP11 b ²	AM234655	<i>T. fasciculata</i> ⁶	T0214AB ³	FJ878747
<i>E. bilunaris</i>	LB2 a ²	AM234656	<i>T. fasciculata</i> ⁶	T0214AC ³	EU883362
<i>E. bilunaris</i>	LB2 b ²	AM234659	<i>T. fasciculata</i> ⁶	T0214AH ³	EU883357
<i>E. bilunaris</i>	LB2 c ²	AM234657	<i>T. fasciculata</i> ⁶	T0214AI ³	FJ878771
<i>E. bilunaris</i>	LB2 d ²	AM234658	<i>T. fasciculata</i> ⁶	T0214AJ ³	FJ878770
<i>E. bilunaris</i>	LB5 a ²	AM234660	<i>T. fasciculata</i> ⁶	T0214AL ³	FJ878746
<i>E. bilunaris</i>	LB5 b ²	AM234661	<i>T. fasciculata</i> ⁶	T0214AM ³	FJ878769

Table 1. (continued)

Species/Authority (corrected name)	Strain ID	Accession #	Species/Authority (corrected name)	Strain ID	Accession #
<i>E. bilunaris</i>	LB5 c ²	AM234662	<i>T. fasciculata</i> ⁶	T0214AN ³	FJ878752
<i>E. bilunaris</i>	LB5 d ²	AM234663	<i>T. fasciculata</i> ⁶	T0214AO ³	FJ878768
<i>E. bilunaris</i>	LB5 e ²	AM234664	<i>T. fasciculata</i> ⁶	T0214AP ³	FJ878753
<i>E. bilunaris</i>	LB5 f ²	AM234665	<i>T. fasciculata</i> ⁶	T0214AQ ³	FJ878745
<i>Fragilaria pinnata</i> Ehrenberg	CCMP0398 ³	GQ330340	<i>T. fasciculata</i> ⁶	T0214AS ³	FJ878754
<i>F. cylindrus</i> (Grunow) Krieger	CCMP1102 ³	GQ330342	<i>T. fasciculata</i> ⁶	T0214L ³	as FJ878754
<i>F. cylindrus</i> (<i>Nitzschia curta</i>)	CCMP1430 cl.1 ³	GQ330361	<i>T. fasciculata</i> ⁶	T0214Q ³	FJ878755
<i>F. cylindrus</i> (<i>Nitzschia curta</i>)	CCMP1430 cl.2 ³	as GQ330361	<i>T. fasciculata</i> ⁶	T0214S ³	FJ878772
<i>F. cylindrus</i> (<i>Nitzschia curta</i>)	CCMP0552 ³	GQ330360	<i>T. fasciculata</i> ⁶	T0214T ³	FJ878767
<i>F. cylindrus</i> (<i>Nitzschia curta</i>)	CCMP1273 ³	GQ330362	<i>T. fasciculata</i> ⁶	T0214U ³	FJ878726
<i>F. cylindrus</i> (<i>Thalassiosira antarctica</i>)	CCMP0981 ³	GQ330451	<i>T. fasciculata</i> ⁶	T0214W ³	FJ878744
<i>F. cylindrus</i>	F cyl new ³	EF660055	<i>T. fasciculata</i> ⁶	T0214 x ³	FJ878766
<i>F. cylindrus</i>	Real 9 ³	EF660056	<i>T. fasciculata</i> ⁶	T0217C ³	FJ878765
<i>F. cylindrus</i>	Real 1 ³	EF660057	<i>T. fasciculata</i> ⁶	T0217D ³	as FJ878765
<i>F. kerguelensis</i> (O'Meara) Hustedt	4-20 ³	EF660061	<i>T. fasciculata</i> ⁶	T0217I ³	FJ878773
<i>F. nana</i> (Nielsen) Paasche	2-E-F ³	EF660058	<i>T. fasciculata</i> ⁶	206E uk ³	FJ878774
<i>F. nana</i>	ErikaFcyl ³	EF660059	<i>T. fasciculata</i> ⁶	206H uk ³	EU883363
<i>F. nana</i>	PA plate 12A4 ³	EF660060	<i>T. tabulata</i> ⁶ (Agardh) Snoeijs	421D ³	EU883365
<i>F. oceanica</i> ⁶ (Cleve) Hasle	SV1B1 ³	GQ330345	<i>T. tabulata</i> ⁶	421I ³	as EU883365
<i>F. oceanica</i> ⁶	SV2C3 ³	GQ330346	<i>T. cf. tabulata</i>	CCMP0846 ³	EU883364
<i>F. oceanica</i> ⁶	SV1A3 ³	GQ330344	<i>Thalassionema nitzschioides</i> ⁶ (Grunow) Mereschkowsky	CT25 ³	GQ330444
<i>Grammatophora oceanica</i> Ehrenberg	CCMP0410 ³	FJ864287	unid. sp., pennate	CCMP2024 ³	GQ330359
<i>Grammonema striatula</i> ⁶ Lyngbye	IIA3 ³	FJ864286	unid. sp., pennate	CCMP2025 ³	GQ330358
<i>G. striatula</i> ⁶	IIIA2 ³	FJ878724	unid. sp., pennate	CCMP2122 ³	GQ330355
<i>G. striatula</i> ⁶	CCMP1094 ³	GQ330341	unid. sp., pennate	CCMP2035 ³	GQ330429
<i>Navicula arenaria</i> Donkin	CCY0228 ²	DQ235781	unid. sp., pennate	CCMP2038 ³	GQ330343
<i>N. arenaria</i>	K6 ²	DQ235782	unid. sp., pennate	CCMP2041 ³	GQ330363
<i>N. gregaria</i> Donkin	K7 ²	DQ235778			

¹see explanation in Kaczmarska et al. (2008) and Amato et al. (2007) regarding taxonomic status of this clade; ²sequences obtained through bacterial cloning, details in publication associated with GenBank accession number; ³sequences obtained from multiple, separate direct-PCR, following rationale of Thornhill et al. (2007); ⁴no information associated published on how sequence was obtained; na=no strain indicated; ⁵s trains obtained from known culture collections are identified as follows: CCMP-Provasoli-Guillard national center for culture of marine phytoplankton, CCAP – Culture collection of algae and protozoa, CSIRO – Commonwealth scientific and industrial research organization collection of living microalgae, IO – Instituto de oceanografia culture collection of Lisbon University, UTEX – University of Texas, Austin culture collection of algae; ⁶Our strains.

(39 sequences), with four different gap penalties (0.1, 1.2, 2.5 and 5.0) to compare how alignment changed with changing parameters. Intraspecific nucleotide difference changed only slightly over the various gap penalties tested, while interspecific and intrafamilial distance increased between gap penalty 0.1 and 1.2 and reached a plateau thereafter. The plateau coincided with alignments that improved low-scoring fragments without disrupting the known anchor-regions as recommended by Hall (2004). Consequently, we chose a gap penalty of 1.2. Using these parameters (open gap penalty of 10 and gap extent penalty of 1.2) we performed a multiple alignment on all regions studied in order to test intraspecific and interspecific genetic distances.

We readily found the conserved motif of helix III of ITS2 by comparing the species for which this region has already been described with new sequences. We do not exclude the possibility that future studies of secondary structure will improve these alignments. But these improvements will not significantly change the boundaries of intraspecific vs heterospecific divergence, since closely related species have very similar sequences which align very easily.

Intragenomic Variation

To calculate intragenomic nucleotide divergence 39 strains from three species belonging to the class Mediophyceae and eight species from the class Bacillariophyceae were used. Intragenomic uncorrected genetic distances (p) observed in the 5.8S sequences of examined mediophytes and pennates (Table 2) ranged from $p=0$ to 0.01 differences (substitutions or indels) per site (diff./site; Table 2).

Genetic distances calculated for 5.8S+ITS2 and the proposed barcode in mediophytes and pennates did not differ and ranged from 0 to $p=0.06$ diff./site. The strains showing greatest variation were two from the species-complex *Eunotia bilunaris* (Vanormelingen et al. 2007, 2008) and the strain *Sellaphora pupula* 18tmNe, which were obtained by cloning and not direct PCR. Except for these two species, divergence between sequences obtained by direct PCR and by cloning was equal.

Intraspecific Variation

To estimate intraspecific nucleotide divergence, 80 mediophyte clones from 25 species (Table 3) and 124 pennate clones from 34 species were

used (Table 4). Within these species, 15 species included strains that were sexually compatible (Table 4) and were considered biological species in the classical sense of having the capacity to interbreed. Some of these interbreeding isolates came from geographically dispersed areas. An additional 19 morpho-species included clones from different biogeographic regions.

Intraspecific uncorrected genetic distance using the 5.8S gene was equal to or slightly higher than the intragenomic range. For mediophytes, it ranged from $p=0$ to 0.06 diff./site (Table 3) and for pennates from $p=0$ to 0.03 diff./site (Table 4). The species showing the maximum variation were *Odontella mobiliensis*, *Pseudo-nitzschia delicatissima* clade 1 (sensu Lundholm et al. 2006; Amato et al. 2007) and *Sellaphora blackfordensis*.

The region 5.8S+ITS2 presented higher intraspecific than intragenomic distances. For mediophytes, it ranged from $p=0$ to 0.11 diff./site (Table 3) and for pennates from $p=0$ to 0.18 diff./site (Table 4). The higher distances in this case were found in the species *Odontella aurita* and *Asterionellopsis glacialis*.

The proposed barcode showed similar results to those observed with the whole 5.8S+ITS2 region. In mediophytes the range recovered was 0 to 0.09 diff./site (Table 3) and in pennates it was 0 to 0.17 diff./site (Table 4), with *Chaetoceros gracilis* and *Sellaphora* deme "blunt" presenting the highest distances.

Interspecific Variation

To calculate the number of interspecific nucleotide differences for congener species, eight mediophyte genera (47 species) and 10 pennate genera (45 species) were used. In this dataset, for 5.8S alone, the interspecific distances greatly overlapped intraspecific genetic distances for both classes. In mediophytes, the distances ranged from $p=0$ to 0.19 diff./site (Table 3) with identical sequences for different species of the same genus found in all genera except *Extubocellulus* and *Odontella*. In 75% of the cases interspecific differences fell under the highest value found for intraspecific variability for this class and 51% of the cases fell under the highest intraspecific variability value found for each family. In bacillariophytes, it ranged from $p=0$ to 0.17 diff./site (Table 4) with identical sequences found in every genus except *Achnanthes*, *Sellaphora* and *Stauroneis*. In 95% of the cases interspecific differences fell under the highest value found for intraspecific variability for pennates and 62% of

Table 2. Ranges of intra-strain and intragenomic distances in the 5.8S, 5.8S+ITS2 and proposed DNA-barcode sequences.

Strain/clone	N	5.8S	5.8S+ITS2	Proposed barcode	ITS ¹
Mediophyceae					
<i>Thalassiosira oceanica</i> CCMP1006 ²	3	0	0	0	
<i>T. weissflogii</i> BILB2001 ³	3	0	0-0.01	0-0.01	
<i>T. weissflogii</i> CCMP1587 ³	2,2	0-0.01	0-0.01	0-0.01	
<i>Minidiscus trioculatus</i> CCMP0495 ³	2	0	0	0	
<i>M. trioculatus</i> var. <i>monoculatus</i> ³	2	0	0	0	
Bacillariophyceae					
<i>Eunotia bilunaris</i> LB2 ²	4	0	0-0.06	0-0.06	
<i>E. bilunaris</i> LB5 ²	6	0	0-0.06	0-0.06	
<i>E. bilunaris</i> DM33-2 ²	3	0	0-0.01	0-0.01	
<i>E. bilunaris</i> DM33-1 ²	5	0	0-0.02	0-0.01	
<i>E. bilunaris</i> DM33-9 ²	5	0-0.01	0-0.04	0-0.05	
<i>E. bilunaris</i> DM33-5 ²	5	0-0.01	0-0.06	0-0.06	
<i>E. bilunaris</i> DM33-18 ²	4	0	0-0.03	0-0.01	
<i>Phaeodactylum tricornutum</i> CCMP0630 ³	2	0	0	0	
<i>P. tricornutum</i> CCMP1327 ³	2	0	0	0	
<i>Sellaphora pupula</i> 18tmNe ²	4	0-0.01	0.01-0.06	0.01-0.06	
<i>S. auldreekie</i> 8B ²	3	0.01	0-0.01	0-0.01	
<i>Navicula phyllepta</i> CCY 0212 ²	3	0	0-0.01	0-0.01	
<i>N. phyllepta</i> CCY 0213 ²	3	0.01	0-0.01	0-0.01	
<i>N. phyllepta</i> BA3 ²	4	0	0	0	
<i>N. phyllepta</i> BA4 ²	3	0	0-0.01	0-0.01	
<i>Pseudo-nitzschia delicatissima</i> ²		nd	nd	nd	4
<i>P. multistriata</i> AA5 ²		nd	nd	nd	0-0.06
<i>P. multistriata</i> DD19 ²	12	nd	nd	nd	0-0.06
<i>P. multistriata</i> DD11 ²	13	nd	nd	nd	0-0.06
<i>P. pungens</i> NA233 ²	6	0-0.01	0-0.01	0-0.01	
<i>P. pungens</i> NA231 ²	2	0.01	0	0	
<i>P. pungens</i> NA179 ²	4	0-0.01	0-0.01	0-0.01	
<i>P. pungens</i> NA178 ²	5	0	0	0	
<i>P. pungens</i> NA241 ²	2	0	0	0	
<i>P. pungens</i> NA192 ²	2	0	0	0	
<i>P. pungens</i> NA116 ²	3	0-0.01	0	0	
<i>P. pungens</i> NA200 ²	4	0	0	0	
<i>P. pungens</i> NA177 ²	3	0	0	0	
<i>P. pungens</i> NA242 ²	2	0	0	0	
<i>P. pungens</i> US-94 ²	5	0	0	0	
<i>P. pungens</i> Vigo-3 ²	7	0	0	0	
<i>P. pungens</i> S(4)6 ²	5	0	0-0.01	0-0.01	
<i>P. pungens</i> US-123 ²	3	0	0	0	
<i>Fragilariopsis cylindrus</i> CCMP1430 ³	2	0	0	0	

¹According to authors cited; ²sequences obtained through bacterial cloning, details in publication associated with GenBank accession number (Table 1); ³sequences obtained from multiple, separate direct-PCR, following rationale of Thornhill et al. (2007); ⁴lower than variability of multiple strains within a clade ([Orsini et al. 2004])

the cases were not separated using the threshold for each family, using 5.8S alone.

In the mediophytes, the combined 5.8S+ITS2 interspecific variability ranged from $p=0$ to

0.48 diff./site calculated from the 245 comparisons, species by species (47 total species, Table 3). Only 17% of the species comparisons fell into the 0 to 0.11 range observed for within

Table 3. Intraspecific, intrageneric and intrafamilial distances for the markers 5.8S, 5.8S+ITS2 and the proposed barcode calculated for seven families of the Class Mediophyceae; N_p indicates number of pairwise comparisons and N_s shows the number of species represented by the sequences. For each marker and each family the average uncorrected distance is shown in each cell above with the range below. Intrageneric and intrafamilial overlap (in parentheses) indicates percent of pairs of sequences falling into the ranges calculated for intraspecific distances (within intrageneric category) or intrageneric distances (within intrafamilial category). Values presented in the row named Class indicates number of species, average and ranges of pairwise sequence comparisons calculated for the Class in the same way as for each family. Overlap percentages were calculated using the Class threshold (first value) and the threshold calculated for each family (second value in bold).

Family	Intraspecific				Intrageneric				Intrafamilial			
	N_p N_s	5.8S	5.8+ITS2	barcode	N_p N_s	5.8S (overlap)	5.8+ITS2 (overlap)	barcode (overlap)	N_p N_s	5.8S (overlap)	5.8+ITS2 (overlap)	barcode (overlap)
Thalassiosiraceae	49 8	0 0-0.01	0.01 0-0.10	0.01 0-0.08	143 18	0.03 (24%)	0.21 (5%)	0.18 (5%)	89 19	0.03 (100%)	0.26 (79%)	0.23 (76%)
Skeletonemataceae	561 7	0.01 0-0.03	0.02 0-0.07	0.01 0-0.05	21 7	0.01 (95%)	0.11 (43%)	0.05 (43%)	6 6	0.08 (0%)	0.26 (0%)	0.22 (0%)
Stephanodiscaceae	9 2	0.01 0-0.02	0.01 0-0.03	0.01 0-0.02	7 4	0.01 (100%)	0.09 (25%)	0.08 (25%)	8 13	0.07 (0%)	0.28 (0%)	0.23 (0%)
Chaetocerotaceae	32 4	0.01 0-0.04	0.03 0-0.09	0.03 0-0.09	68 12	0.09 (16%)	0.34 (4%)	0.30 (3%)	12 2	0.12 (92%)	0.44 (92%)	0.29 (100%)
Lithodesmiaceae	56 2	0 0-0.05	0.01 0-0.05	0 0-0.01	nd nd	nd (nd)	nd (nd)	nd (nd)	5 3	0.04 (nd)	0.25 (nd)	0.22 (nd)
Cymatosiraceae	Nd Nd	nd nd	nd nd	nd nd	1 2	0.02 (nd)	0.18 (nd)	0.21 (nd)	14 6	0.08 (21%)	0.38 (0%)	0.33 (0%)
Tricerataceae	7 2	0.01 0-0.06	0.05 0-0.11	0.04 0-0.08	5 4	0.13 (20%)	0.38 (0%)	0.32 (0%)	nd nd	nd (nd)	nd (nd)	nd (nd)
Class	714 25	0.01 0-0.06	0.02 0-0.11	0.02 0-0.09	245 47	0.05 (75%/ 51%)	0.22 (17%/ 15%)	0.19 (16%/ 15%)	134 49	0.07 (79%/ 43%)	0.31 (99%/ 34%)	0.25 (100%/ 35%)

Table 4. Intraspecific, intrageneric and intrafamilial distances for the markers 5.8S, 5.8S+ITS2 and the proposed barcode calculated for nine families of the Class Bacillariophyceae; N_p indicates number of pairwise comparisons and N_s shows the number of species represented by the sequences. For each marker and each family the average uncorrected distance is shown in each cell above with the range below. Intrageneric and intrafamilial overlap (in parentheses) indicates percent of pairs of sequences falling into the ranges calculated for intraspecific distances (within intrageneric category) or intrageneric distances (within intrafamilial category). Values calculated only with biological species are indicated separately. Values presented in the row named Class indicates number of species, average and ranges of pairwise sequence comparisons calculated for the Class in the same way as for each family. Overlap percentages were calculated using the Class threshold (first value) and the threshold calculated for each family (second value in bold).

Family	Intraspecific				Intrageneric				Intrafamilial			
	N_p N_s	5.8S	5.8+ITS2	barcode	N_p N_s	5.8S (overlap)	5.8+ITS2 (overlap)	barcode (overlap)	N_p N_s	5.8S (overlap)	5.8+ITS2 (overlap)	barcode (overlap)
Fragilariaceae	832	0	0.02	0.02	9	0.01	0.22	0.24	25	0.03	0.31	0.36
	3	0-0.02	0-0.06	0-0.06	5	0-0.01 (100%)	0.16-0.30 (0%)	0.19-0.31 (0%)	6	0.01-0.05 (0%)	0.16-0.39 (20%)	0.23-0.44 (0%)
Fragilariaceae (biological species)	821	0	0.03	0.02	3	0.01	0.18	0.23	nd	nd	nd	nd
	2	0-0.02	0-0.05	0-0.06	3	0-0.01 (100%)	0.16-0.21 (0%)	0.19-0.31 (0%)		(nd)	(nd)	(nd)
Eunotiaceae ¹ (biological species)	33	0	0.02	0.02	103	0	0.17	0.17	nd	nd	nd	nd
	4	0-0.01	0-0.05	0-0.05	4	0-0.01 (100%)	0.13-0.21 (0%)	0.13-0.21 (0%)		(nd)	(nd)	(nd)
Achnantheaceae	nd	nd	nd	nd	1	0.10	0.49	0.40	2	0.10	0.40	0.36
					2	(nd)	(nd)	(nd)	2	0.09-0.10 (100%)	0.38-0.41 (100%)	0.33-0.39 (100%)
Sellaphoraceae ² (biological species)	46	0.01	0.02	0.02	9	0.03	0.30	0.27	nd	nd	nd	nd
	2	0-0.03	0-0.13	0-0.17	5	0.01-0.06 (38%)	0.09-0.41 (21%)	0.08-0.39 (21%)		(nd)	(nd)	(nd)
Phaeodactylaceae	55	0	0.01	0.01	nd	nd	nd	nd	nd	nd	nd	nd
	1		0-0.02	0-0.03		(nd)	(nd)	(nd)		(nd)	(nd)	(nd)

Naviculaceae	28	nd	nd	nd	21	0.03	0.32	0.25	nd	nd	nd	nd
	8				8	0-0-05 (nd)	0.13-0.40 (nd)	0.13-0.34 (nd)		(nd)	(nd)	(nd)
Stauroneidaceae	nd	nd	nd	nd	2	0.17	0.51	0.45	nd	nd	nd	nd
					1	(nd)	(nd)	(nd)		(nd)	(nd)	(nd)
Bacillariaceae	1973	0	0	0	168	0.02	0.23	0.17	213	0.02	0.38	0.25
	15	0-0.03	0-0.02	0-0.03	28	0-0.06 (71%)	0.01-0.48 (1%)	0-0.43 (1%)	28	0-0.08 (98%)	0.30-0.40 (100%)	0.08-0.43 (100%)
Bacillariaceae (biological species)	1424	0	0	0	45	0.02	0.17	0.15	nd	nd	nd	nd
	7	0-0.02	0-0.02	0-0.03	10	0-0.04 (62%)	0.03-0.24 (0%)	0.02-0.24 (2%)		(nd)	(nd)	(nd)
Catenulaceae	3	0.01	0.03	0.03	3	0.04	0.36	0.36	nd	nd	nd	nd
	1		0-0.05	0-0.05	2	0.03-0.04 (0%)	0.35-0.36 (0%)	0.35-0.36 (0%)		(nd)	(nd)	(nd)
Class	2970	0	0.02	0.02	316	0.05	0.33	0.29	240	0.05	0.32	0.33
	34	0-0.03	0-0.18	0-0.17	45	0-0.17 (95%/ 62%)	0.01-0.41 (41%/ 4%)	0.02-0.45 (41%/ 4%)	36	0-0.10 (90%/ 66%)	0.16-0.41 (95%/ 73%)	0.10-0.44 (98%/ 67%)

¹following criteria of Vanormelingen et al. (2008); ²following criteria of Mann et al. 2008.

species variation. This includes the case of the two closely related morpho-species *Stephanodiscus niagarae* and *S. yellowstonensis*, thought to have diverged only 12,000-8,000 years ago (Zechman et al. 1994) and which present an identical sequence. Also, only 15% of comparisons fell into the range observed within each family. In the bacillariophytes, the 5.8S+ITS2 interspecific distances ranged from 0.01 to 0.41 calculated for the 316 species by species comparisons (45 total species, Table 4). Although 41% of the species comparisons fell into the 0 to 0.18 range observed for within species variation, only 4% of the comparisons fell into the range observed within each family; which means 85% separation in mediophyte species and 96% separation in bacillariophyte species using thresholds specific for each family. For both classes, these results indicate an excellent level of species resolution considering that not only biologically defined but also morpho-species were included.

Using the same analyses for the proposed DNA-barcode region within the class Mediophyceae, the interspecific distance ranged from $p=0$ to 0.41 diff./site calculated for the 245 species by species comparisons (47 total species, Table 3). Similar to the longer fragment, only 16% of the species comparisons fell into the $p=0$ to 0.09 diff./site range observed for within species variation and only 15% fell into the range observed within each family. In the bacillariophytes, the barcode interspecific distances ranged from $p=0.02$ to 0.45 diff./site calculated for the 316 species by species comparisons (45 total species, Table 4). Although 41% percent of the species comparisons fell into the 0 to 0.17 range observed for within species variation for the whole class, 4% of species comparisons fell into the range observed within each family, indicating that our proposed barcode performs as well as the longer fragment discussed above. Again, this translates to 85% separation in mediophyte species and 96% separation in bacillariophyte species, using thresholds specific for each family.

Both 5.8S/ITS2 and the proposed barcode are good at discriminating the vast majority of species examined in our species set assembled by chance. For mediophytes, the proposed barcode and the whole 5.8S+ITS2 are roughly equivalent in discriminating species. For bacillariophytes, the proposed barcode is slightly less sensitive than 5.8S/ITS2 in discriminating species. No two species ever presented the same sequence either for the whole 5.8S/ITS2 and only a few did for the proposed barcode.

To expand the phylogenetic range of data examined, intrageneric but intrafamilial divergence was also calculated for 17 mediophyte genera (49 species, Table 3) and 9 pennate genera (36 species, Table 4). Families which showed the highest overlap for intrageneric/intrafamilial variability coincide with families that showed the lowest overlap for intraspecific/interspecific variability, these being Thalassiosiraceae and Chaetocerotaceae using 5.8S and all families using the other two markers.

Ultimate Test – BLAST

The effectiveness of the barcode was further supported using the tool BLAST on all sequences, including the ones used in intra- versus interspecific variability, on ten sequences of unknown taxa, seven ill-fitting sequences, and six sequences belonging to families represented by single strains. The sequences obtained were BLASTed against our database using the software provided by Genome Quebec (McGill). For seven strains their sequences were flagged as suspect by this marker. The misidentification was confirmed by subsequent morphological analysis. These strains were “*Nitzschia curta*” (CCMP 0552), “*N. cylindrus*” (CCMP 1273), “*N. curta*” (CCMP 1430) and “*Thalassiosira antarctica*” (CCMP 0981) all of which proved to be the same as CCMP 1102 (*Fragilariopsis cylindrus*). Another strain, “*Pseudo-nitzschia seriata*” (CCMP 1309) was found to be the same as *Pseudo-nitzschia* cf. *granii* (CCMP 2093) while “*Cylindrotheca closterium*” (CCMP 0339) was in fact *Ditylum brightwellii*. Using the same procedure, we also used our proposed barcode to identify ten strains from CCMP that were listed as unidentified centrics or pennates (“unid. sp.”). BLASTed to our database, these strains were identified as indicated: CCMP 2018 correctly as *Extubocellulus spinifer*, CCMP 2122 as *Navicula phyllepta*, CCMP 2024 (is *Navicula tripunctata*) and CCMP 2025 as species of *Navicula* correctly, CCMP 2035 as a species of *Pseudo-nitzschia* (is a *Nitzschia* sp.), CCMP 2038 as a species of *Fragilariopsis* (is *Nitzschia* cf. *spatulata*), CCMP 2125 as *Fragilaria pinnata* (is *Nanofrustulum shiloi*), CCMP 1277 and CCMP 2063 as *Thalassiosira* sp. correctly (is *T.* cf. *mediolana*), CCMP 2041 as a species of *Nitzschia* (is *Psammodictyon* cf. *panduriformis*). These identities were confirmed by a follow up SEM-based examination; identities of the closest morpho-species are indicated in parentheses. For example, we identified CCMP 2024 as

Navicula tripunctata, a species for which we did not yet have a sequence. The strains identified only to the genus level were confirmed to be species different from those for which reference sequences were available in our dataset. When a reference barcode for the species existed in our database, the identification of “unid. sp.” was correct.

Discussion

What Makes a Good Barcode and How to Test It?

A practical and useful barcode must fulfill at least three prerequisites: possess genetic interspecific distances exceeding intraspecific distances (monophyletic clusters of representative barcodes for a species), consist of a sequence length obtainable in a single amplification and contain conserved flanking fragments so as to facilitate the design of universal primers.

The concept of DNA-barcoding has been well accepted and successfully practiced, particularly in the higher animal kingdom. In these organisms, biological delineation of species coincides fairly well with phenotypical criteria normally used to identify species, and intraspecific *cox1* sequence variability is low, in contrast to diatoms (Evans et al. 2007; Ehara et al. 1995; Moniz and Kaczmarek 2009). Since *cox1* (the chosen barcode for the animal kingdom) separates the majority of animal species in well defined genera, it is considered a good marker (Cywinska et al. 2006; Hajibabaei et al. 2006; Hebert et al. 2003, 2004), although there is less consensus regarding the threshold of divergence coinciding with separation of species (Hebert et al. 2003; Lefebvre et al. 2006; Spies et al. 2006; Ward et al. 2005). In microprotists however, *cox1*-based barcode testing including sexual compatibility has only been conducted for the diatom genus *Sellaphora* (Evans et al. 2007; Evans and Mann 2009; Mann et al. 2008), and the ciliate genus *Tetrahymena* (Lynn and Strüder-Kypke 2006).

Here, we discuss the efficacy of an alternative marker (5.8S+ITS2) applied to a wider still phylogenetic range and larger number of diatom species obtained from our research and available from the literature (Amato et al. 2007; Behnke et al. 2004; Evans et al. 2007; Kaczmarek et al. 2008; Moniz and Kaczmarek 2009; Vanormelingen et al. 2007, 2008) for species delineated

morphologically (Tables 3 and 4) and biologically (Table 4, in part).

Using our proposed DNA-barcoding, ca. 75% of all species examined here were resolved using the highest value of intraspecific variability found for the whole class as a threshold and 91% using the highest value for intraspecific variability found for each family as a threshold. This success rate is lower than recently reported in animal phyla (Janzen et al. 2009) but higher than, for example, the 70% success achieved using the three-gene barcode *matk+rpoB+rpoC1* for land plants, which represented one of the highest success rates found for plants (Fazekas et al. 2008). It appears that diatoms harbor comparatively more intraspecific sequence variability in our barcode region even among sexually compatible clones than some other protists (Litaker et al. 2007). We nonetheless anticipate the success rate of the proposed barcode for diatoms to improve further with expansion of our reference set of sequences and better taxonomic resolution of “difficult” taxa.

In this study we included (a) 15 species from four genera containing species tested for sexual compatibility, (b) morpho-species whose identity was confirmed solely by light microscopy and (c) taxonomically challenging genera comprised of sequences available in GenBank. In the second group, we include several pairs of sister morpho-species in genera such as: *Cylindrotheca*, *Extubocellulus*, *Fragilariopsis*, *Odontella*, *Pseudonitzschia*, *Skeletonema*, *Tabularia*, *Thalassiosira* and *Minidiscus*. In the last group (c), we accepted species identity given by associated sources, some of which remain unpublished. These include ca. 15 species from genera which are either undergoing taxonomic re-appraisal or have been identified as requiring re-appraisal. These include: *Chaetoceros* (Castillo et al. 1992; Ostenfeld 1912; Rines and Hargraves 1988), *Cyclotella* (Carvalho et al. 1995; Flower et al. 1990; Håkansson 1989; Håkansson and Khursevich 1997; Sabater and Klee 1990), *Skeletonema* (Kooistra et al. 2008; Sarno et al. 2005, 2007; Zingone et al. 2005) and *Stephanodiscus* (Håkansson and Kling 1989, 1990; Håkansson and Meyer 1994). Therefore, some of the sequences from group (c) may represent strains where species assignment is uncertain or incorrect. These taxonomic uncertainties would affect the intraspecific genetic distance range recovered in our study but we included them here to illustrate cases of taxa representing a range of variance in species taxonomic resolution, such as will likely be encountered in initial stages of future barcoding

efforts of such taxa. It is encouraging to observe that even in these cases the barcode places the diatom in the correct genus and often in the correct section of the genus when the reference sequence of a related species is available in the database.

5.8S Gene – an Anchor Point

In the chance-selected assemblage of diatoms we examined here, the 5.8S fragment shows enough variability to separate some species and most genera, which can be seen by the fact that the families included either have low intraspecific/intrageneric overlap or low intrageneric/intrafamilial overlap. Kress et al. (2005) reported that in flowering plants this fragment acts as an anchor point for alignments in both phylogenetic and barcoding studies. In our study and those from green and brown algae, terrestrial plants, and most animal lineages (Coleman 2003), sequences were often spotted as misidentified or mislabeled because the 5.8S fragment differed significantly from what was expected. When a species sequence is unavailable in the database being used, this fragment promptly eliminates many candidates and often correctly indicates to which higher taxon this unknown organism belongs. For these reasons, we decided to keep the 5.8S gene when testing our barcode.

ITS2 Fragment – Adding Resolution

As a non-coding fragment, ITS2 does not have the triplet code to help with alignment and therefore is a difficult marker to use in global alignments with all taxa. However, recent reports have shown that ITS2 contains conserved cores throughout the eukaryotes (Coleman 2000, 2003, 2007; Coleman and Mai 1997; Müller et al. 2007; Schultz et al. 2005). These core structures can now be readily aligned automatically for very large sets of sequences (Müller et al. 2007; Seibel et al. 2006; Wolf et al. 2005) making ITS2 more suitable as a marker over a broader taxonomic range (Coleman 2003; Mai and Coleman 1997; Schultz et al. 2005).

The 5.8S+ITS2 fragment reliably separated intraspecific and interspecific divergence in 91% of the cases we studied using the family-specific threshold. These successful cases included species from the genera *Thalassiosira*, *Cyclotella*, *Odontella*, *Chaetoceros*, *Fragilaria*, *Tabularia*, *Eunotia*, *Sellaphora* and most nitzschoids.

In a similar study, Litaker et al. (2007) also evaluated the ITS fragment and its different

constituents as a barcode for dinoflagellates. Although the 5.8S fragment did not show enough resolution to separate dinoflagellate species, both ITS1 and ITS2 alone did, suggesting that these fragments alone could serve as a tool for identifying dinoflagellate species. Using the entire ITS sequence, they found differences higher than $p=0.04$ as indicative of interspecific divergence level. For diatoms and the 5.8S+ITS2 fragment, this value would be a very low threshold, but dinoflagellates seem to show less divergence than other taxa even when using *cox1* (Ferrell 2008). Using just the ITS2 fragment, most of the dinoflagellate species studied by Litaker et al. (2007) were separated with uncorrected genetic distances greater than $p=0.0706$, which is more similar to distances we found in the combined 5.8S+ITS2 fragment in diatoms. Shao et al. (2004), using the ITS fragment for species identification in Gymnodiniaceae, found much higher interspecific genetic distances ($p=0.19$ to 0.20), more in line with our findings for diatoms using uncorrected distances in a general time reverse model.

Proposed Barcode

The proposed barcode shows enough divergence to separate 91% of the diatom morpho-species and 99.5% of all biologically defined species using the family-specific threshold. Not only did the proposed barcode confirm separation of well-defined species with a class-specific threshold of $p=0.11$ diff./site for mediophytes and $p=0.18$ diff./site for pennates, it also improved the alignment and thus resolution of taxa such as the genera from the family Chaetocerotaceae showing more overlap when using the entire 5.8S+ITS2 fragment aligned with no aid of secondary structure.

Since a barcode should be precisely defined, we chose to test a fragment beginning with the start codon at the 5' end of 5.8S and ending in the widely conserved motif on the 5' end of helix III of the ITS2 secondary structure (Coleman 2008). This fragment fulfills the requirements of a reliable and practical barcode because it contains a short sequence length compatible with single amplifications: ca. 300 bp in mediophytes and 400 bp in pennates. This fragment also has conserved flanking fragments (SSU and LSU) that increase the probability of developing universal primers. We used the primers ITS1 and ITS4 for a large majority of our sequences, but we believe that with the diversity of species that have in the meantime been sequenced, either by us or by others, improvement on these primers might be obtained.

The ITS region has already been used as a specific tag in a number of protist phyla, some animal groups (Wang and Guo 2008; Moritz et al. 2001) and fungi (Seifert 2009). Recently, the Internal Sub-commission on Fungal Barcoding selected ITS as a barcode for fungi (Seifert 2009), and now prepares the formal proposal to the Consortium for the Barcode of Life (CBOL, Seifert 2009) to designate it as an official barcode for fungi. Litaker et al. (2007) suggested using this marker to identify dinoflagellates. Even among flowering plant species, Gemeinholzer et al. (2006), explored advantages and disadvantages of using ITS1 for species identification of Asteraceae from the tribes Lactuceae and Anthemiodeae. The limitations for use of the ITS-region in phylogenetic studies are well documented in flowering plants (Álvarez and Wendel 2003) and more challenging due to polyploidization being a key mechanism of speciation; 50-80+% of flowering plant species are thought to have arisen via polyploidization (Soltis et al. 2004, 2009).

In polyploids and hybrids, particularly in taxa where concerted evolution is slow, intraindividual sequence variation may be high (Álvarez and Wendel 2003). The presence of different multiple, equally frequent copies of the DNA barcode sequence would render identification of such species more difficult, but not impossible to detect (D'Álelio et al. 2009). Although polyploid species have not yet been documented in diatoms (Armbrust et al. 2004; Bowler et al. 2008; Kociolek and Stoermer 1989) species chromosome numbers are remarkably similar in closely related taxa (Kociolek and Stoermer 1989) suggesting their non-polyploid origin. If polyploid diatom species are discovered and/or their DNA proves inaccessible to direct PCR amplification, an alternative barcode may be needed for them. Polyploidy however, is not without advantages because it may identify species ancestry (Feliner and Rosselló 2007).

Using ITS region sequence data, hybrids have recently been shown to exist among morphologically indistinguishable sympatric populations of *Pseudonitzschia multistriata* (D'Álelio et al. 2009), two varieties of *P. pungens* (Casteleyn et al. 2009) and among sterile hybrids of two reproductively isolated heterogeneous clones of *Eunotia bilunaris* (Vanormelingen et al. 2008). In all these cases the hybrid nature of the sequences was apparent and did not prevent their identification. Nonetheless, routine evaluation of sequences irrespective of suspect polyploid or hybrid nature is advisable in all biota (D'Álelio et al. 2009; Feliner and Rosselló 2007).

Paralogy, better documented in nuclear multi-gene, multicopy gene families, is not absent in organelles, as recently documented for *cox1* in the fungal genus *Fusarium* (Gilmore et al. 2009), and should be anticipated in the diatom organelle genome as well, particularly in species with hologamic syngamy. All these data necessitate re-emphasis of our message, here and in previous publications (Kaczmarska et al. 2007; Moniz and Kaczmarska 2009) that "unusual" DNA-barcodes for a given species should be viewed as "flags" to initiate further in-depth taxonomic analysis.

Comparison with Cox1

In animals the success rate in species identification using *cox1* barcodes has been reported to be as high as 98-100% in North American birds (Hebert et al. 2004), Australian fish (Ward et al. 2005) and tropical Lepidoptera (Hajibabaei et al. 2006). The success rate declined when a comprehensive search for inter- and intraspecific variation also including "difficult" species was carried out as we did in our study. Meyer and Paulay (2005) found that in a comprehensive dataset of cowry species (small marine gastropods) the majority of individuals (>96%) were successfully identified by the *cox1*-barcode. However, they noted that a certain percentage of evolutionarily recent and/or closely related species (0 to 8% in their case) will still show 100% identity with their sister species, similar to our case of two morpho-species of *Stephanodiscus* and *Minidiscus*. In certain genera with currently unresolved taxonomy, as was the case in some of the taxa included in our study (e.g. some species of *Chaetoceros*) some species will appear non-monophyletic, increasing the barcoding identification error rate for taxa defined only by morphology. Meyer and Paulay (2005) also caution against the strict use of thresholds in screening for new species, because this would overlook at least one-fifth of less divergent species, e.g. cowries, turbinids and limpets. Therefore, thresholds may be set for various taxa more specifically, in tune with their evolutionary history and the time of separation from their last common ancestor.

Some protist groups have shown success when using *cox1* barcodes, e.g. macroalgae (Kucera and Saunders 2008; Saunders 2005), and some micro-protists, albeit in some groups on selected or phylogenetically narrow taxa such as isolates from the genus *Tetrahymena* (Chantangsi et al. 2007) and strains of *Paramecium tetraurelia* and *P. caudatum* (Barth et al. 2006). As well, using 22

Sellaphora species including sexually compatible and incompatible strains and three other raphid genera, [Evans et al. \(2007\)](#) and [Mann et al. \(2008\)](#) assessed *cox1* as a possible barcode for diatoms. Our own data supports the authors' conclusions that this marker affords clear separation of species and that diatom-universal primer design for this DNA fragment will be challenging ([Moniz and Kaczmarska 2009](#)). These difficulties led us to focus on the ITS-fragment as, based on our data to date, it appears to carry stronger potential to be used in barcoding all diatoms, using a few pairs of primers already available. A more systematic comparison of this and other markers is discussed in [Moniz and Kaczmarska \(2009\)](#).

Working Toward a Globally Consistent Diatom Identification Tool

DNA-barcoding is a kind of molecular taxonomy that utilizes a very short standardized DNA sequence to identify species ([Ratnasingham and Hebert 2007](#)). In practice, a recovered barcode sequence of an unknown specimen is compared to barcodes in an existing database assembled for specimens, identified or validated by expert taxonomists, available in museum and culture collections, archived from biodiversity inventories, etc. Each of these barcodes is tied to a voucher specimen, its biogeography and other auto-ecological data that is publicly available at the Barcoding of Life Data Systems (BOLD) linked via the "LinkOut" function to the DNA Barcoding in the GenBank TaxBrowser of the respective taxon. These databases provide molecular and classical taxonomic references for the animal species, inclusive of the range of representative barcodes attributed to a species and reflecting its genetic and morphological variability, a *de facto* barcode-description of the species. Such a database includes (whenever possible or after taxonomic re-appraisal), but is not limited to, barcodes for a species type (holotype, neotype, etc.). DNA barcode sequence variability (comprised in representative barcodes) has been a recognized phenomenon since the onset of barcoding and is evident in the attention given to detection of ranges in sequence divergence within and between species ([Hebert et al. 2003](#); [Saunders 2005](#)), although intraspecific barcode variability in animal groups is much lower than in diatoms, even when only biological species of diatoms are compared.

Application of DNA barcodes to identification of diatoms conducted as summarized above will

confer the same benefits. We consider barcoding of diatoms as a modern, electronically accessible and searchable extension and addition to the earlier floristic efforts (e.g., Diatoms of the United States, etc.) where voucher images, biogeography, etc. may now be accompanied by a standardized short fragment of DNA; the barcode. It is our contention that there is an obvious need to barcode a great number of specimens of a great number of species (imperfect as they currently may appear to be delineated) in order to assemble a reference barcode-database and begin a large scale systematic evaluation of the relationship between valve micro-architecture (morpho-species), DNA sequence variability (phylogenetically defined species) and when necessary (and possible), their relationship to breeding compatibility (biologically defined species). Widely cast barcoding will separate known species from yet undescribed taxa. If current predictions of diatom species richness are correct ([Mann 1994, 1999](#)) and species number is on the order of 200,000-1,000,000, compared to 10,000-20,000 species now formally described, the majority of extant diatom species will be described in the future and these ought to be typified by both morphology and type-(bar)code.

Among barcoded animal species, "unusual" barcodes initiate scrutiny of the specimens and when necessary involve interim names and codes ([Janzen et al. 2009, p. 10](#)). New species are then described, old ones amended and affiliated with the type or neo-typified (type-code) etc., following the rules of the Nomenclatural Code and, perhaps consideration of [Hoef-Emden et al. \(2007\)](#) and [Surek \(2008\)](#). This approach also applies to diatoms. The history of redefining old species and discovery of new species in *Pseudo-nitzschia* is an excellent example where barcoding could now be consistently applied to individuals collected worldwide even if few yet have their taxonomic types barcoded.

In a different approach, [Evans and Mann \(2009\)](#) illustrate how barcodes can be used to stabilize taxonomy of described species by giving them nomenclatural status in order to establish an "unambiguous molecular reference point" ([Evans and Mann 2009, p. 73](#)), with the implication that one species can only be represented by one barcode (holotype, epitype barcode, etc.). In that approach, "the barcode value is to improve species typification" ([Evans and Mann 2009, p. 73](#)). We concur that such a type-code reference point would indeed be an unambiguous reference

and improved typification ought to be the goal of follow-up taxonomic revisions when specimens believed to belong to one species are found to contain groups of different barcodes. We observe however that the type specimen is already an unambiguous reference point. Barcoding that is practiced as an identification tool for animal and macro-algae, described above (Hebert et al. 2003; Janzen et al. 2009; Saunders 2005) and proposed here has the potential to achieve more than improving typification. Ambiguities arise when the morphology (or sequence) of an unknown individual departs from 100% identity with the type. It is at this point that representative barcodes are most useful because they provide an empirical and quantitative framework for intraspecifically acceptable divergence in the standardized sequence. This is based on several hundred characters (nucleotides) and their positions, on which to evaluate the goodness of fit of the unknown individual to a closest barcoded species. Diatom cultures available in culture collections are a natural first choice to begin such DNA-barcoding. It is already clear that large scale taxonomic revisions in diatom taxonomy and systematics will result from using molecular tools for species identification (including DNA-barcoding), as it did following the introduction of electron microscopy to diatom research a few decades ago.

Summary

So far and in the species tested, the barcode we propose has succeeded in separating 91% of cases using thresholds determined at the family level. BLASTing of sequences correctly identified all sequences tested, detected misidentified sequences and afforded identification of unknown strains, including those which could be identified to species level when reference sequences were available in our dataset. Larger reference databases, when available, will be especially useful to all scientists requiring a globally consistent, correct and inexpensive identification of unknown organisms. Recently a comparison study of several methods of species identification came to the conclusion that BLASTing against a good reference database is one of the most reliable tools, even more so than strict tree-based phylogenetic methods (Ross et al. 2008) attesting to significance of a valid reference database.

Databases compatible with the proposed barcode fragment are growing very rapidly (Müller

et al. 2007; Seibel et al. 2006; Wolf et al. 2005). Furthermore, improved understanding of the secondary structure of ITS2 brings additional serendipitous advantages to our choice. An intriguing consistent coincidence between sexual incompatibility and nucleotide differences in these conserved fragments has been documented in a wide range of biota (Coleman 2008), including a few species in the genus *Pseudo-nitzschia* (Amato et al. 2007; Casteleyn et al. 2008; D'Alelio et al. 2009) and possibly in the genus *Sellaphora* (Behnke et al. 2004). If proven widely applicable to diatoms, this coincidence would be a very useful proxy for the detection of sexual separation in microbiota such as diatoms, for which mating compatibility is not easily determined on a routine basis and morphological characteristics of congeners may be misleading (Evans et al. 2008). Behnke et al. (2004) and Amato et al. (2007) have already suggested the presence of CBC (compensatory base change) in sexually incompatible clones of two *Sellaphora* genotypes and seven *Pseudo-nitzschia* genotypes. For all of these encouraging reasons, we propose the 5.8S+ITS2 fragment as a DNA-barcode for diatoms.

Methods

The identity and origin of culture strains examined in this study are listed in Table 1 and included 116 clones established in our laboratory, 117 strains received from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP), 5 from Instituto de Oceanografia da Faculdade de Ciências, Lisbon University (IO) and 380 sequences from GenBank. Available strains were selected to represent a wide range of geographic and phylogenetic diversity, preferably among multispecies genera. The procedure for clone isolation by micropipetting followed Andersen (2005). Non-axenic cultures were maintained in 20 mL tubes or 125 mL flasks in f/2 medium (Guillard 1983), a dilution thereof (f/4 to f/20), or in L1⁺ medium (Andersen 2005) at a temperature of approximately either 6 °C, 12 °C or 20 °C depending on strain requirements, with a 12:12 h light:dark (L:D) photocycle and a photon fluence rate of about 20–50 $\mu\text{E m}^{-2} \text{s}^{-1}$. Frozen pellets of all the strains used, their DNA and material prepared for SEM observations are available upon request. Live material of nearly all of the CCMP strains used in this study is available at the collection. Live material of some of our clones (or their genetic equivalents) is still available and will be provided upon request.

When necessary to verify the identity of a strain, culture samples were prepared for SEM as described in Kaczmarska et al. (2005) and observed with a JEOL-5600 SEM (JEOL USA, Peabody, MA, USA) operating at 10 kV and 8 mm working distance, at the Digital Microscopy Facility, Mount Allison University.

Cells in mid exponential growth phase were concentrated by centrifugation and pellets were frozen until needed. The UltraClean Soil DNA Kit (MoBio Laboratories, Carlsbad, CA,

USA) as well as the Power Plant DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA) were used to obtain DNA, following manufacturer specifications.

Seven hundred bp of the nuclear internal transcribed spacers (ITS1 and ITS2) and the 5.8S rDNA gene (hereinafter referred collectively as the ITS fragment) were amplified in most cases using primers ITS1 and ITS4 (White et al. 1990). In some cases, primers ITS5 (White et al. 1990) or the pair NS7m and LR1850 (Bhattacharya 1996; White et al. 1990) were used with better results. The 25 μ l PCR reaction mixes were carried out either in PuReTaqTM Ready-To-Go PCR Beads (GE Healthcare Biosciences, Piscataway, NJ, USA) following Kaczmarska et al. (2008) or using JumpStart (Sigma, Saint Louis, MO, USA) and included the same amounts of DNA extraction solution and primer mixes, 12.5 μ l of Taq Mix and 9 μ l of ultrapure water (DEPC treated water, Invitrogen Canada, Inc., Burlington, ON, Canada). The amplification regime followed Amato et al. (2007). PCR products were visualized in a 1.3% agarose gel. Cleaning of PCR products and sequencing were conducted at Nanuq (McGill University and Genome Québec) following their standard procedure.

We routinely applied BLASTing to all sequences obtained against the sequences present in GenBank and also our database using the tool provided by Nanuq (McGill University and Genome Quebec).

Electropherograms for the 5.8S and ITS2 fragments were edited with BioEdit version 7.0.5.3 (Hall 1999). All our sequences were based on high quality bidirectional readings. The different fragments (5.8S alone, 5.8S+ITS2 fragment and proposed barcode) of our sequences and sequences from GenBank (Table 1) were aligned with either BioEdit or ClustalW version 1.4 (Thompson et al. 1994). Thereafter alignments were verified and/or improved manually. A gap open of 10 and gap extent of 1.2 were used for analyses. Specifics of evaluation of alignment parameters and justification for those chosen in this study are presented in the second section of the results. Distance rates were estimated using simple uncorrected pair-wise distances (p) with the Visual Basic/Excel program DOINK (J.M. Ehrman, Digital Microscopy Facility, Mount Allison University, 2007). The program produces a nucleotide difference matrix identical to that produced by BioEdit but within an Excel spreadsheet. Genetic distances between sequences in each analysis were expressed as the number of substitutions, gaps or indels (differences) per site, following Litaker et al. (2007) for ease of comparison. When uncorrected $p \leq 0.1$, uncorrected distances are very similar to corrected distances such as the Kimura 2-Parameter model as in the current Barcode of Life Data System (BOLD, www.boldsystems.org, Ratnasingham and Hebert 2007). In fact, it was suggested by Nei and Kumar (2000) that when studying closely related species, it is better to use simpler distances, because they have smaller variance. Our use of uncorrected distances also makes our study comparable to other studies that also used uncorrected distances.

All fragments studied begin with the start codon of 5.8S. The 5.8S gene is very easily aligned. The whole 5.8S+ITS2 ends in the binding site of the reverse primer. The end of the proposed barcode was found by searching for the conserved motif on helix III, thus close to the end of ITS2 and consists of TGGTA inclusive in *Stephanodiscus* (Coleman 2007). In other species we found some variability in this sequence fragment. The common characteristic YGGTA was found in mediophytes and the sequence TGGTA was the most commonly found in pennates. The 5.8S region was comprised of ca. 155 sites. The 5.8S+ITS2 region included ca. 500 sites. The proposed

barcode region consisted of ca. 300 bp in mediophytes and 400 bp in pennates.

Acknowledgements

We thank R. Andersen and all CCMP staff for freely providing their cultures and J.M. Ehrman for all the SEM work involved in this study and for bioinformatics support; N. Davidovich, J. Martin, M. LeGresley and K. Pauley for providing phytoplankton samples from which our clones were isolated, and A. Cockshutt for molecular troubleshooting early in this study. Four reviewers provided valuable input that needs acknowledgement and we thank them for it. We also thank all colleagues in our laboratory for providing any help needed. We acknowledge funding through the Canadian Barcode of Life Network from Genome Canada through the Ontario Genomics Institute, NSERC, Mount Allison University and other sponsors listed at www.BOLNET.ca (to IK).

References

- Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Mol Phylogenet Evol* **29**: 417–434
- Amato A, Kooistra WHCF, Ghiron JHL, Mann DG, Pröschold T, Montresor M (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* **158**:193–207
- Andersen RA ed (2005) *Algal Culturing Techniques*. Elsevier Academic Press, Amsterdam
- Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski M, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**:79–86
- Barth D, Krenek S, Fokin SI, Berendonk T (2006) Intraspecific genetic variation in *Paramecium* revealed by mitochondrial cytochrome *c* oxidase I sequences. *J Eukaryot Microbiol* **53**:20–25
- Behnke A, Friedl T, Chepurinov VA, Mann DG (2004) Reproductive compatibility and rDNA sequence analyses in the *Sellaphora pupula* species complex (Bacillariophyta). *J Phycol* **40**:193–208
- Beszteri B, Ács É, Medlin LK (2005) Ribosomal DNA sequence variation among sympatric strains of the *Cyclotella*

- meneghiniana* complex (Bacillariophyceae) reveals cryptic diversity. *Protist* **156**:317–333
- Bhattacharya D** (1996) Analysis of the distribution of bootstrap tree lengths using the maximum parsimony method. *Mol Phylogenet Evol* **6**:339–350
- Bortolus A** (2008) Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *Ambio* **37**:114–118
- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otilar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret J, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Secq MO, Napoli C, Obornik M, Parker MS, Petit J, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV** (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**:239–244
- Carvalho LR, Cox EJ, Fritz SC, Juggins S, Sims PA, Gasse F, Battarbee RW** (1995) Standardizing the taxonomy of saline lake *Cyclotella* spp. *Diatom Res* **10**:229–240
- Casteleyn G, Adams NG, Vanormelingen P, Debeer A, Sabbe K, Vyverman W** (2009) Natural hybrids in the marine diatom *Pseudo-nitzschia pungens* (Bacillariophyceae): genetic and morphological evidence. *Protist* **160**:343–354
- Casteleyn G, Chepurnov VA, Leliaert F, Mann DG, Bates SS, Lundholm N, Rhodes L, Sabbe K, Vyverman W** (2008) *Pseudo-nitzschia pungens* (Bacillariophyceae): A cosmopolitan diatom species. *Harmful Algae* **7**:241–257
- Castillo PMS, León MAU, Round FE** (1992) Estudio de *Chaetoceros wighamii* Brightwell: un taxon mal interpretado. *Diatom Res* **7**:127–136
- Chantangsi C, Lynn DH, Brandl MT, Cole JC, Hetrick N, Ikonomi P** (2007) Barcoding ciliates: a comprehensive study of 75 isolates of the genus *Tetrahymena*. *Int J Syst Evol Microbiol* **57**:2412–2425
- Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RO, Haidar N, Savolainen V** (2005) Land plants and DNA barcodes: short-term and long-term goals. *Philos Trans R Soc Lond B Biol Sci* **360**:1889–1895
- Coleman AW** (2000) The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. *Protist* **151**:1–9
- Coleman AW** (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet* **19**:370–375
- Coleman AW** (2007) Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Res* **35**:3322–3329
- Coleman AW** (2008) Is there a molecular key to the level of “biological species” in eukaryotes? A DNA guide. *Mol Phylogenet Evol* doi:10.1016/j.ympev.2008.10.008.
- Coleman AW, Mai JC** (1997) Ribosomal DNA ITS-1 and ITS-2 sequence comparisons as a tool for predicting genetic relatedness. *J Mol Evol* **45**:168–177
- Coleman AW, Suarez A, Goff L** (1994) Molecular delineation of species and syngens in volvocacean green algae (Chlorophyta). *J Phycol* **30**:80–90
- Cowan RS, Chase MW, Kress WJ, Savolainen V** (2006) 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon* **55**:611–616
- Cywinska A, Hunter FF, Hebert PDN** (2006) Identifying Canadian mosquito species through DNA barcodes. *Med Vet Entomol* **20**:413–424
- D’Alelio D, Amato A, Kooistra WHCF, Procaccini G, Casotti R, Montresor M** (2009) Internal transcribed spacer polymorphism in *Pseudo-nitzschia multistriata* (Bacillariophyceae) in the Gulf of Naples: recent divergence or intraspecific hybridization?. *Protist* **160**:9–20
- Duffy JE, Stachowicz JJ** (2006) Why biodiversity is important to oceanography: potential roles of genetic, species, and trophic diversity in pelagic ecosystem processes. *Mar Ecol Prog Ser* **311**:179–189
- Ehara M, Inagaki Y, Watanabe KI, Ohama T** (1995) Phylogenetic analysis of diatom *cox1* genes and implications of a fluctuating GC content on mitochondrial genetic code evolution. *Curr Genet* **37**:29–33
- Evans KM, Mann DG** (2009) A proposed protocol for nomenclaturally effective DNA barcoding of microalgae. *Phycologia* **48**:70–74
- Evans KM, Wortley AH, Mann DG** (2007) An assessment of potential diatom “barcode” genes (*cox1*, *rbcL*, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist* **158**:349–364
- Evans KM, Wortley AH, Simpson GE, Chepurnov VA, Mann DG** (2008) A molecular systematic approach to explore diversity within the *Sellaphora pupula* species complex (Bacillariophyta). *J Phycol* **44**:215–231
- Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH** (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* **3**:e2802
- Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SCH, Newmaster SG, Hajibabaei M, Husband BC** (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers?. *Mol Ecol Resour* **9**:130–139
- Feliner GN, Rosselló JA** (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol Phylogenet Evol* **44**:911–919
- Ferrell J** (2008) The evaluation of DNA barcoding for species identification of dinoflagellates. Master of Science thesis, Mount Allison University, 134pp.
- Flower RJ, Gasse F, Håkansson H** (1990) A new species of *Cyclotella*, *C. azizgensis* sp. nov., described from modern material collected from upland lakes in Morocco. *Diatom Res* **5**:253–260

- Gemeinholzer B, Oberprieler CH, Bachmann K** (2006) Screening the applicability of molecular markers for plant identification using the ITS 1 sequences of Asteraceae species belonging to the tribes Lactuceae and Anthemideae. *Taxon* **55**:173–187
- Gilmore SR, Gräfenhan T, Louis-Seize G, Seifert KA** (2009) Multiple copies of cytochrome oxidase 1 in species of the fungal genus *Fusarium*. *Mol Ecol Resour* **9**:90–98
- Gómez A, Wright PJ, Lunt DH, Cancino JM, Carvalho GR, Hughes RN** (2006) Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan taxon. *Proc R Soc Lond B Biol Sci* **274**:199–207
- Guillard RRL** (1983) Culture of Phytoplankton for Feeding Marine Invertebrates. In Berg Jr. CJ (ed) *Culture of Marine Invertebrates. Selected Readings*. Hutchinson Ross Publ Co, Stroudsburg, PA, pp 108–132
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN** (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci USA* **103**:968–971
- Håkansson H** (1989) A light and electron microscopical investigation of the type species of *Cyclotella* (Bacillariophyceae) and related forms, using original material. *Diatom Res* **4**:255–267
- Håkansson H, Kling H** (1989) A light and electron microscope study of previously described and new *Stephanodiscus* species (Bacillariophyceae) from central and northern Canadian lakes, with ecological notes on the species. *Diatom Res* **4**:269–288
- Håkansson H, Kling H** (1990) The current status of some very small freshwater diatoms of the genera *Stephanodiscus* and *Cyclostephanos*. *Diatom Res* **5**:273–287
- Håkansson H, Khursevich G** (1997) *Tertiarius* gen. nov., a new genus in the Bacillariophyceae, the transfer of some cyclotelloid species and a comparison to closely related genera. *Diatom Res* **12**:19–33
- Håkansson H, Meyer B** (1994) A comparative study of species in the *Stephanodiscus niagarae*-complex and a description of *S. heterostylus* sp. nov. *Diatom Res* **9**:65–85
- Hall BG, ed.** (2004). *Phylogenetic Trees Made Easy: A How-To Manual*. 2nd ed, Sinauer Associates, Sunderland, MA.
- Hall TA** (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*, **41**, 95–98.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR** (2003) Biological identifications through DNA barcodes. *Proc R Soc Lond B Biol Sci* **270**:313–321
- Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM** (2004) Identification of birds through DNA barcodes. *PLoS Biol* **2**:e312
- Hoef-Emden K, Küpper FC, Andersen RA** (2007) Meeting report: Sloan Foundation workshop to resolve problems relating to the taxonomy of microorganisms and to culture collections arising from the barcoding initiatives; Portland ME, November 6–7, 2006. *Protist* **158**:135–137
- Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou J, Chacon I, Dapkey T, Deans AR, Epstein ME, Espinoza B, Franclemont JG, Haber WA, Hajibabaei M, Hall JPW, Hebert D, Gauld ID, Harvey DJ, Hausmann A, Kitching IJ, Lafontaine D, Landry J, Lemaire C, Miller JY, Miller JS, Miller L, Miller SE, Montero J, Munroe E, Green SR, Ratnasingham S, Rawlins JE, Robbins RK, Rodriguez JJ, Rougerie R, Sharkey MJ, Smith MA, Solis MA, Sullivan JB, Thiaucourt P, Wahl DB, Weller SJ, Whitfield JB, Willmott KR, Wood DM, Woodley NE, Wilson JJ** (2009) Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol Ecol Resour* **9**:1–26
- Kaczmarska I, Reid C, Moniz M** (2007). Diatom taxonomy: morphology, molecules and barcodes. In: Kusber WH, Jahn R (eds.) *Proceedings of the 1st Central-European Diatom meeting 2007, Berlin-Dahlem, 23 March–25 March 2007*. Berlin-Dahlem, Germany, pp 69–72.
- Kaczmarska I, LeGresley MM, Martin JL, Ehrman JM** (2005) Diversity of the diatom genus *Pseudo-nitzschia* Peragallo in the Quoddy Region of the Bay of Fundy, Canada. *Harmful Algae* **4**:1–19
- Kaczmarska I, Reid C, Martin JL, Moniz MBJ** (2008) Morphological, biological, and molecular characteristics of the diatom *Pseudo-nitzschia delicatissima* from the Canadian Maritimes. *Can J Bot* **86**:763–772
- Kaczmarska I, Ehrman JM, Moniz MBJ, Davidovich N** (in press). Phenotypic and genetic structure of interbreeding populations of the diatom *Tabularia fasciculata* (Bacillariophyta). *Phycologia*.
- Kocielek JP, Stoermer EF** (1989) Chromosome numbers in diatoms: a review. *Diatom Res* **4**:47–54
- Kooistra WHCF, Sarno D, Balzano S, Gu H, Andersen RA, Zingone A** (2008) Global diversity and biogeography of *Skeletonema* species (Bacillariophyta). *Protist* **159**:177–193
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH** (2005) Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA* **102**:8369–8374
- Kress WJ, Erickson DL** (2007) A two-locus global DNA barcode for land plants: the coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE* **2**:e508
- Kucera H, Saunders GW** (2008) Assigning morphological variants of *Fucus* (Fucales, Phaeophyceae) in Canadian waters to recognized species using DNA barcoding. *Botany* **86**:1065–1079
- Lefébure T, Douady CJ, Gouy M, Gibert J** (2006) Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Mol Phylogenet Evol* **40**:435–447
- Litaker RW, Vandersea MW, Kibler SR, Reece KS, Stokes NA, Lutzoni FM, Yonish FM, West MA, Black MND, Tester PA** (2007) Recognizing dinoflagellate species using ITS rDNA sequences. *J Phycol* **43**:344–355
- Lundholm N, Moestrup Ø, Kotaki Y, Hoef-Emden K, Scholin C, Miller P** (2006) Inter- and intraspecific variation of the *Pseudo-nitzschia delicatissima* complex (Bacillariophyceae) illustrated by rRNA probes, morphological data and phylogenetic analyses. *J Phycol* **42**:464–481
- Lynn DH, Strüder-Kypke MC** (2006) Species of *Tetrahymena* identical by small subunit rRNA gene sequences are discriminated by mitochondrial cytochrome c oxidase I gene sequences. *J Eukaryot Microbiol* **53**:385–387

- Mai JC, Coleman AW** (1997) The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J Mol Evol* **44**:258–271
- Mann DG** (1994) The Origins of Shape and Form in Diatoms: the Interplay Between Morphogenetic Studies and Systematics. In Ingram DS, Hudson A (eds) *Shape and Form in Plants and Fungi*. The Linnean Society of London, London, pp 17–38
- Mann DG** (1999) The species concept in diatoms. *Phycologia* **38**:437–495
- Mann DG, Droop SJM** (1996) Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* **336**:19–32
- Mann DG, Thomas SJ, Evans KM** (2008) Revision of the diatom genus *Sellaphora*: a first account of the larger species in the British Isles. *Fottea* **8**:15–78
- Mann DG, McDonald SM, Bayer MM, Droop SJM, Cherpurnov VA, Loke RE, Ciobanu A, du Buf JMH** (2004) The *Sellaphora pupula* species complex (Bacillariophyceae): morphometric analysis, ultrastructure and mating data provide evidence for five new species. *Phycologia* **43**:459–482
- Medlin LK** (1997) Can molecular techniques help define species limits?. *Diatom* **13**:19–23
- Meyer CP, Paulay G** (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol* **3**:e422
- Moniz MJB, Kaczmarek I** (2009) Barcoding diatoms: is there a good marker?. *Mol Ecol Resour* **9**:65–74
- Moritz G, Paulsen M, Delker C, Picl S, Kumm S** (2001). Identification of Thrips using ITS-RFLP Analysis. In Marullo R, Mound L (eds.) *Proceedings of the 7th International Symposium on Thysanoptera 2001*, Calabria, Italy, 2 July–7 July 2001. Regio Calabria, Italy, pp 365–367.
- Müller T, Philippi N, Dandekar T, Schultz J, Wolf M** (2007) Distinguishing species. *RNA* **13**:1469–1472
- Nei M, Kumar S eds** (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford
- Newmaster SG, Fazekas AJ, Ragupathy S** (2006) DNA barcoding in the land plants: evaluation of *rbcl* in a multigene tiered approach. *Can J Bot* **84**:335–341
- Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J** (2008) Testing candidate plant barcode regions with species of recent origin in the Myristicaceae. *Mol Ecol Notes* **8**:480–490
- Orsini L, Procaccini G, Sarno D, Montresor M** (2004) Multiple rDNA ITS-types within the diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae) and their relative abundances across a spring bloom in the Gulf of Naples. *Mar Ecol Prog Ser* **271**:87–98
- Ostenfeld CH** (1912) A revision of the marine species of *Chaetoceros* Ehb. sect. *Simplicia* Ostf. *Meddel. Komm. Havunders. Series Plankton* **1**:1–11
- Ratnasingham S, Hebert PDN** (2007) BOLD: the Barcode of Life Data system. *Mol Ecol Notes* **7**:355–364
- Rines JEB, Hargraves PE** (1988) The *Chaetoceros* Ehrenberg (Bacillariophyceae) flora of Narragansett Bay, Rhode Island, USA. *Bibl Phycol* **79**:1–196
- Ross HA, Murugan S, Li WLS** (2008) Testing the reliability of genetic methods of species identification via simulation. *Syst Biol* **57**:216–230
- Sabater S, Klee R** (1990) Observations on centric diatoms of the rivers Ebro phytoplankton, with special interest on small *Cyclotella* spp. *Diatom Res* **5**:141–154
- Sarno D, Kooistra WHCF, Medlin LK, Percopo I, Zingone A** (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). II: an assessment of the taxonomy of *S. costatum*-like species with the description of four new species. *J Phycol* **41**:151–176
- Sarno D, Kooistra WHCF, Balzano S, Hargraves PE, Zingone A** (2007) Diversity in the genus *Skeletonema* (Bacillariophyceae). III. Phylogenetic position and morphological variability of *Skeletonema grevillei*, with the description of *Skeletonema ardens* sp. nov. *J Phycol* **43**:156–170
- Saunders GW** (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Proc R Soc Lond B Biol Sci* **360**:1879–1888
- Schultz J, Maisel S, Gerlach D, Müller T, Wolf M** (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* **11**:361–364
- Scicluna SM, Tawari B, Clark CG** (2006) DNA barcoding of *Blastocystis*. *Protist* **157**:77–85
- Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M** (2006) 4SALE – a tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics* **7**:498
- Seifert KA** (2009) Progress towards DNA barcoding of fungi. *Mol Ecol Resour* **9**:83–89
- Shao P, Chen Y, Zhou H, Yuan J, Qu L, Zhao D, Lin Y** (2004) Genetic variability in Gymnodiniaceae ITS regions: implications for species identification and phylogenetic analyses. *Mar Biol* **144**:215–224
- Soltis DE, Soltis PS, Tate JA** (2004) Advances in the study of polyploidy since plant speciation. *New Phytol* **161**:173–191
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS** (2009) Polyploidy and angiosperm diversification. *Am J Bot* **96**:336–348
- Spies IB, Gaichas S, Stevenson DE, Orr JW, Canino MF** (2006) DNA-based identification of Alaska skates (*Amblyraja*, *Bathyraja* and *Raja*: Rajidae) using cytochrome *c* oxidase subunit I (col) variation. *J Fish Biol* **69**:283–292
- Surek B** (2008) Meeting report: algal culture collections 2008: an international meeting at the Culture Collection of Algae and Protozoa (CCAP), Dunstaffnage Marine Laboratory, Dunbeg, Oban, United Kingdom; June 8–11, 2008. *Protist* **159**:509–517
- Thompson JD, Higgins DG, Gibson TJ** (1994) CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighing, position specific gap penalties and weigh matrix choice. *Nucleic Acids Res* **22**:4673–4680
- Thornhill DJ, Lajeunesse TC, Santos SR** (2007) Measuring rDNA diversity in eukaryotic microbial systems: how intragenomic variation, pseudogenes, and PCR artifacts confound biodiversity estimates. *Mol Ecol* **16**:5326–5340

- Utz LRP, Eizirik E** (2007) Molecular phylogenetics of subclass Peritrichia (Ciliophora: Oligohymenophorea) based on expanded analyses of 18S rRNA sequences. *J Eukaryot Microbiol* **54**:303–305
- Vanormelingen P, Chepurinov VA, Mann DG, Cousin S, Vyverman W** (2007) Congruence of morphological, reproductive and ITS rDNA sequence data in some Australasian *Eunotia bilunaris* (Bacillariophyta). *Eur J Phycol* **42**:61–79
- Vanormelingen P, Chepurinov VA, Mann DG, Sabbe K, Vyverman W** (2008) Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric clones of *Eunotia bilunaris sensu lato* (Bacillariophyta). *Protist* **159**:73–90
- Wang Y, Guo X** (2008) ITS length polymorphism in oysters and ITS use in species identification. *J Shellfish Res* **27**:489–493
- Ward R, Zemlack TS, Innes BH, Last PR, Hebert PDN** (2005) DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci* **360**:1847–1857
- White TJ, Bruns TD, Lee SB, Taylor JW** (1990) Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) *PCR Protocols: a Guide to Methods and Applications*. Academic Press, London, pp 315–322
- Wolf M, Friedrich J, Dandekar T, Müller T** (2005) CBC analyzer: inferring phylogenies based on compensatory base changes in RNA secondary structures. *In Silico Biol* **5**:291–294
- Zechman FW, Zimmer EA, Theriot EC** (1994) Use of ribosomal DNA internal transcribed spacers for phylogenetic studies in diatoms. *J Phycol* **30**:507–512
- Zingone A, Percopo I, Sims PA, Sarno D** (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae): I. A re-examination of the type material of *S. costatum* with the description of the *S. grevillei* sp. nov. *J Phycol* **41**: 140–150

Available online at www.sciencedirect.com

