Org Divers Evol
DOI 10.1007/s13127-011-0050-6

ORIGINAL ARTICLE

ORGANISMS
DIVERSITY &
EVOLUTION

Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols

Jonas Zimmermann · Regine Jahn · Birgit Gemeinholzer

Received: 5 July 2010 / Accepted: 15 July 2011
© Gesellschaft für Biologische Systematik 2011

Abstract Diatoms are present in all types of water bodies and their species diversity is influenced greatly by environmental conditions. This means that diatom occurrence and abundances are suitable indicators of water quality. Furthermore, continuous screening of algal biodiversity can provide information about diversity changes in ecosystems. Thus, diatoms represent a desirable group for which to develop an easy to use, quick, efficient, and standardised organism identification tool to serve routine water quality assessments. Because conventional morphological identification of diatoms demands specialised in-depth knowledge, we have established standard laboratory procedures for DNA barcoding in diatoms. We (1) identified a short segment (about 400 bp) of the SSU (18S) rRNA gene which is applicable for the identification of diatom taxa, and (2) elaborated a routine protocol including standard primers for this group of microalgae. To test the universality of the primer binding sites and the discriminatory power of the proposed barcode region, 123 taxa, representing limnic diatom diversity, were included in the study and identified at species level. The effectiveness of the barcode was also scrutinised within a closely related species group, namely the *Sellaphora pupula* taxon complex and relatives.

J. Zimmermann (✉) · R. Jahn
Botanic Garden and Botanical Museum Berlin-Dahlem,
Freie Universität Berlin,
Königin-Luise-Str. 6-8,
14195 Berlin, Germany
e-mail: j.zimmermann@bgbm.org

J. Zimmermann · B. Gemeinholzer
Justus Liebig University,
Heinrich-Buff-Ring 38,
35392 Giessen, Germany

Keywords 18S (SSU) rRNA gene · Bacillariophyta · DNA barcoding · Diatoms · Standard laboratory procedure

Introduction

Diatoms are unicellular photoautotrophic eukaryotes which are responsible for at least 25% of the global carbon dioxide fixation (Falkowski et al. 1998; Field et al. 1998; Mann 1999; Smetacek 1999). They are an important part of benthic and planktonic biocoenoses and occur nearly ubiquitously in limnic, marine, and terrestrial ecosystems as well as in aerosols (Jahn et al. 2007). Therefore, diatoms are often used as biodindicators in water monitoring assessments and ecological studies (Stevenson and Pan 1999; Stoermer and Smol 1999). Even closely related taxa (excluding cryptic species) are often indicative of different ecological conditions (Poulíčková et al. 2008; Vanelslander et al. 2009). Hence, unambiguous identification of organisms down to species level is crucial for the quality of these studies. Archibald (1984) and Morales et al. (2001) have pointed out that many ecological and monitoring studies are misleading, because identifications have not been verified by experienced diatom taxonomists. To identify diatoms morphologically beyond the genus level is difficult and requires expert knowledge, especially because frustule morphology can vary considerably even within a population (Babanazarova et al. 1996; Bailey-Watts 1976; Jahn 1986; Medlin et al. 1991).

In cases of groups with poor morphological resolution, Hebert et al. (2003) promoted the concept of a DNA barcode to help with the identification of taxa. A DNA barcode is an instrument for the correlation of a taxonomically undetermined individual to a taxon with similar genetic sequence in a given reference database (Ratnasingham and Hebert 2007).

Published online: 26 July 2011

However, a suitable barcode marker has to meet three requirements. The ideal barcode marker (1) consists of a short sequence that can be easily amplified and sequenced in one read following a standardised laboratory protocol, (2) is flanked by a conserved region in which universal primers can be placed, and (3) has the power to resolve organisms at species level (e.g. Hebert et al. 2003; Moritz and Cicero 2004; Stoeckle 2003). Therefore, as in any environmental sampling approach, the quality of the method is not only related to the extent and quality of the reference database but also to the number of taxa that can be identified unambiguously (Erickson et al. 2008), and to the rate at which taxa are retrieved from environmental samples.

Applying the DNA barcoding concept to diatoms promises great potential to resolve the problem of inaccurate species identification and thus facilitate analyses of the biodiversity of environmental samples. In particular, the use of DNA barcodes in diatoms can serve various purposes, such as (1) DNA-based species characterisation and (2) surveying the genetic diversity in an environment of interest. Each of these goals implies different requirements with respect to sequence characteristics. Whereas species characterisation needs sequences with high discriminatory power for defining and identifying even cryptic species, it is not necessarily dependent on fast and universal laboratory protocols. A survey of genetic diversity in environmental samples, however, often relies on high-throughput techniques and therefore needs universal primers and standard protocols where taxa do not have be resolved on the finest scale (e.g. subspecies, cryptic species) (e.g. Hamsher et al. 2011).

Various gene regions have been proposed as barcode markers for diatoms. The mitochondrial cytochrome oxidase I gene (*cox*1) has been widely used for barcoding animals and other organism groups (e.g. Blaxter 2004; Blaxter et al. 2004; Hajibabaei et al. 2006a; Hebert et al. 2004; Robba et al 2006; Saunders 2005; Seifert et al. 2007; Ward et al. 2005). Evans et al. (2007, 2008) successfully tested *cox*1 as a barcoding marker in 22 *Sellaphora* species and three other raphid genera of diatoms. Their study also included a test of the chloroplast ribulose-1,5-bisphosphate carboxylase oxygenase gene (*rbcL*), which was less variable than *cox*1 within the species sampling. However, in other organism groups such as red algae (e.g. Robba et al. 2006; Saunders 2005, 2008), brown algae (Kucera and Saunders 2008) and some green algae (e.g. Lewis and Flechtner 2004; McManus and Lewis 2005), the *rbcL* gene proved to be a promising barcode marker. Moniz and Kaczmarska (2009, 2010) proposed a combination of the nuclear 5.8S rRNA gene and ITS2 upon screening the most species-rich classes of diatoms including mainly marine taxa of the Mediophyceae and Bacillariophyceae. Furthermore, binary characteristics, such as presence/absence of compensatory base changes (CBCs) in the secondary structure of ITS2 or the presence/absence of certain indels have been used to resolve species level diversity in all kind of organisms, including diatoms (Müller et al. 2007). This, however, includes the additional procedural step of calculating and analysing the secondary structure and, therefore, is too laborious for standard high-throughput analyses of environmental samples.

In existing sequence databases, the most extensive data record available for diatoms concerns the nuclear small ribosomal subunit (SSU-rRNA gene), as the latter has been used widely for phylogenetic and taxonomic purposes (e.g. Behnke et al. 2004; Beszteri et al. 2001; Friedl and O'Kelly 2002; Kooistra and Medlin 1996; Medlin et al. 1996; Medlin and Kaczmarska 2004; Sarno et al. 2005; Sorhannus 2007). This means that a substantial reference volume is already available (Hajibabaei et al. 2007), even though identification quality often is not verifiable and therefore does not meet DNA barcoding requirements. The 18S rRNA gene has been suggested as a potential barcoding marker for various organism groups, e.g. nematodes, tardigrades, and diatoms (Bhadury et al. 2006; Blaxter 2004; Blaxter et al. 2004; Floyd et al. 2002; Jahn et al. 2007; Powers 2004). The 18S region has been tested for diatoms in a pilot study by Jahn et al. (2007) and has been used as a marker in other protist groups (Scicluna et al. 2006; Utz and Eizirik 2007).

The present study proposes a 390–410 bp long fragment of the 1800 bp long 18S rRNA gene locus as a barcode marker for the analysis of environmental samples with high-throughput technologies such as 454 sequencing or microarrays, and discusses its use and limitations for diatom identification. The partial 18S region includes a section that is termed V4 in the nomenclature of Nelles et al. (1984) and represents the largest and most complex of the highly variable regions within the 18S locus (Nickrent and Sargent 1991).

Using newly designed universal primers for the V4 region that are introduced below, the region is identified as the most applicable one for barcoding on the 18S locus. Furthermore, an optimised standard laboratory protocol (including DNA extraction, PCR amplification and sequencing) is provided which was developed using diatoms from various limnic genera across many families to represent the freshwater diatom diversity. The study includes taxa from the three major divisions of diatoms: Coscinodiscophyceae (e.g. *Aulacoseira* spp.), Mediophyceae (e.g. *Cyclotella* spp., *Stephanodiscus* spp.) and Bacillariophyceae, with both raphid (e.g. *Nitzschia* spp.) and araphid representatives (e.g. *Fragilaria* spp.) (Table 1).

Table 1 List of all taxa, including strains, EMBL accession numbers (18 s rRNA), voucher identification codes in the Herbarium Berolinense (B), and sampling localities

Taxon	Strain	EMBL no.	B voucher	Locality
Achnanthidium cf. *longipes* Kisselev	–	AY485500	–	–
Achnanthidium minutissimum (Kützing) Czarnecki	D05_008	FR873231	B40 0040686	Germany, Iffeldorf
	–	AJ866992	–	–
	–	AY485500	–	–
Aulacoseira alpigena (Grunow) Krammer	–	AY569578	–	–
Aulacoseira ambigua (Grunow) Simonsen	–	AY569579	–	–
	–	AY569580	–	–
	–	AY569581	–	–
	–	AY569582	–	–
	–	AY569583	–	–
	–	X85404	–	–
Aulacoseira granulata (Ehrenberg) Simonsen	Aula2	FR873232	–	Germany, Berlin
	–	AY569584	–	–
	–	AY569585	–	–
Aulacoseira crenulata (Ehrenberg) Thwaites	–	AY569586	–	–
Aulacoseira granulata var. angustissima (O.Müller) Simonsen	–	AY485493	–	–
Aulacoseira nyassensis (O.Müller) Simonsen	–	AJ535187	–	–
	–	AY121819	–	–
Aulacoseira skvortzowii Edlund, Stoermer & C.M.Taylor	–	AJ535184	–	–
Aulacoseira subarctica (O.Müller) E.Y.Haworth	–	AY569573	–	–
	–	AY569574	–	–
	–	AY569576	–	–
	–	AY569577	–	–
Aulacoseira valida (Grunow) Krammer	–	AY569586	–	–
Cocconeis cf. molesta Kützing	–	AJ535148	–	–
Cocconeis pediculus Ehrenberg	D36_020[a]	FR873233	B40 0040644	Germany, Berlin
	Coco1	FR873234	B40 0040687	Germany, Berlin
	LuCoc03	FR873235	B40 0040734	Germany; Lubmin
	–	AM502010	–	–
Cocconeis placentula Ehrenberg	D36_012[a]	FR873236	B40 0040647	Germany, Berlin
	D17_011	FR873237	B40 0040735	Faroe Islands, Eysturo
	D26_016	FR873238	B40 0040736	South Korea, Tae An Sa
	WiCoc01	FR873239	B40 0040737	Germany, Wismar
	–	AM502013	–	–
Cyclotella atomus Hustedt	–	DQ514858	–	–
Cyclotella bodanica Grunow	–	DQ514901	–	–
Cyclotella choctawhatcheeana A.K.S.Prasad	–	AM712618	–	–
Cyclotella distinguenda Hustedt	–	DQ514859	–	–
Cyclotella gamma Sovereign	–	DQ514852	–	–
Cyclotella meneghiniana Kützing	D15_036	FR873240	B40 0040688	Faroe Islands, Eiði
	–	AM236073	–	–
	–	AJ535172	–	–
	–	AY496206	–	–
	–	AY496207	–	–
	–	AY496210	–	–
	–	AY496211	–	–
	–	AY496212	–	–

Table 1 (continued)

Taxon	Strain	EMBL no.	B voucher	Locality
	–	AY496213	–	–
	–	DQ514853	–	–
	–	DQ514860	–	–
	–	DQ514861	–	–
	–	GQ148714	–	–
	–	GQ148716	–	–
Cyclotella ocellata Pant.	–	DQ514904	–	–
Diatoma hyemalis (Roth) Heib.	–	AB085829	–	–
Diatoma tenuis C. Agardh	D03_036	FR873241	B40 0040689	Germany, Berlin
	–	AJ535143	–	–
	–	AM497730	–	–
	–	AM497731	–	–
	–	EF423403	–	–
	–	EU260466	–	–
Diatoma vulgare var. *linearis* Grunow	–	EF465466	–	–
Encyonema caespitosum Kützing	–	AM502035	–	–
Encyonema minutum (Hilse) D.G.Mann	D36_007	FR873242	B40 0040690	Germany, Berlin
	–	AM501961	–	–
Encyonema silesiacum (Bleisch) D.G.Mann	D16_038	FR873243	B40 0040691	Faroe Islands, Viðareiði
Encyonema triangulum (Ehrenberg) Kützing	–	AJ535157		
Eolimna minima (Grunow) Lange–Bertalot[b]	D03_030	FR873244	B40 0040692	Germany, Berlin
	–	AJ243063	–	–
Eolimna subminuscula (Manguin) Gerd Moser, LangeBertalot & Metzeltin	–	AJ243064	–	–
Fragilaria bidens Heib.	–	AM497732	–	–
Fragilaria capucina Desmazières	–	EF465492	–	–
Fragilaria mesolepta Rabenhorst	D03_041	FR873245	B40 0040693	Germany, Berlin
Fragilaria nanana Lange–Bertalot	–	AM497738	–	–
	–	AM497739	–	–
Fragilaria pinnata Ehrenberg	–	EF192988	–	–
Fragilaria striatula Ehrenberg	–	AY485474	–	–
	–	EU090016	–	–
	–	EU090017	–	–
Fragilaria ulna (Nitzsch) Lange–Bertalot	–	AJ866993	–	–
Fragilaria vaucheriae (Kützing) J.B.Petersen	–	AM497733	–	–
	–	AM497735	–	–
	–	AM497741	–	–
Gomphonema acuminatum Ehrenberg	–	AM502019	–	–
Gomphonema micropus Kützing	–	AM501965	–	–
Gomphonema parvulum Kützing	D33_006	FR873246	B40 0040694	Mexico, Ixtlán de los Hervores
	–	AJ243062	–	–
Gomphonema productum (Grunow) Lange–Bertalot & E.Reichardt	–	AM501993	–	–
Gomphonema truncatum Ehrenberg	–	AM501956	–	–
Hantzschia amphioxys (Ehrenberg) Grunow	D28_006	FR873247	B40 0040695	Germany, Berlin
	D27_008	FR873248	B40 0040696	Germany, Berlin
Mayamaea atomus (Kützing) Lange–Bertalot var. *atomus*	–	AM501968	–	–
Mayamaea atomus var. *permitis* (Hustedt) Lange–Bertalot	D06_107	FR873249	B40 0040697	Germany, Berlin

Table 1 (continued)

Taxon	Strain	EMBL no.	B voucher	Locality
	–	AM501968	–	–
	–	AM501969	–	–
Melosira cf. *octogona* (Grunow) Hustedt	–	AY485518	–	–
Melosira varians C. Agardh	D34_009	FR873250	B40 0040698	Spain, Mallorca, Sa Calobra
	–	AJ243065	–	–
	–	AY569590	–	–
Navicula brockmanni Hustedt	–	AM50202	–	–
Navicula capitatoradiata H.Germ.	–	AM502012	–	–
Navicula cari Ehrenberg	–	GU295220	–	–
Navicula cryptocephala Kützing	D06_059	FR873251	B40 0040700	Germany, Berlin
	–	AM501973	–	–
	–	AM501996	–	–
Navicula cryptotenella Lange–Bertalot	–	AM502011	–	–
	–	AM502029	–	–
Navicula diserta Hustedt	–	AJ535159	–	–
Navicula glaciei van Heurck	–	EF106788	–	–
Navicula gregaria Donkin	D08_002	FR873252	B40 0040703	Sweden; Ängelholm
	–	AM501974	–	–
Navicula lanceolata (C.Agardh) Ehrenberg	–	AY485484	–	–
Navicula pelliculosa (Bréb. ex Kützing) Hilse	–	EU260468	–	–
Navicula phyllepta Kützing	–	FJ624253	–	–
Navicula radiosa Kützing	–	AM501972	–	–
	–	AM502027	–	–
	–	AM502034	–	–
Navicula ramosissima (C.Agardh) Cleve	–	AY485512	–	–
Navicula reinhardtii (Grunow) Grunow	–	AM501976	–	–
Navicula saprophila Lange–Bertalot & Bonik	–	AJ867025	–	–
Navicula slesvicensis Grunow	D06_038	FR873253	B40 0040705	Germany, Berlin
Navicula sp.	D08_051	FR873254	B40 0040699	Sweden; Ängelholm
Navicula subminuscula Manguin	–	AJ867026	–	–
Navicula tripunctata (O.F.Müller) Bory	D03_093	FR873255	B40 0040706	Germany, Berlin
	–	AM502028	–	–
Nitzschia acicularis (Kützing) W.Smith	D03_095	FR873256	B40 0040707	Germany, Berlin
	–	AJ867000	–	–
Nitzschia cf. *fonticola* Grunow	D15_023	FR873257	B40 0040712	Faroe Islands, Eiði
Nitzschia cf. *semirobusta* Lange–Bertalot	D32_012	FR873258	B40 0040708	Mexico, Ojo de Agua
	–	AJ867277	–	–
Nitzschia cf. *supralitorea* Lange–Bertalot	–	AJ867020	–	–
Nitzschia communis Rabenhorst	–	AJ867014	–	–
	–	AJ867278	–	–
*Nitzschia closterium (*Ehrenberg) W.Smith	–	AY485455	–	–
	–	HQ121419	–	–
Nitzschia dissipata (Kützing) Rabenhorst	D03_120	FR873259	B40 0040710	Germany,Berlin
	–	AJ867018	–	–
Nitzschia inconspicua Grunow	D10_009	FR873260	B40 0040709	Hungary, Sofiok
	–	AJ867021	–	–
Nitzschia filiformis (W.Sm.) van Heurck	–	AJ866999	–	–
Nitzschia linearis (C.Agardh) W. Smith	D05_031	FR873261	B40 0040711	Germany, Iffeldorf

Table 1 (continued)

Taxon	Strain	EMBL no.	B voucher	Locality
	–	AJ867011	–	–
	–	AJ867012	–	–
	–	AJ867013	–	–
Nitzschia palea (Kützing) W.Smith	–	AJ867006	–	–
	–	AJ867009	–	–
	–	AJ867010	–	–
Nitzschia palecea (Grunow) Grunow	–	AJ866996	–	–
Nitzschia paleaformis Hustedt	–	AJ866997	–	–
	–	AJ866998	–	–
Nitzschia pusilla Grunow	–	AJ867015	–	–
	–	AJ867016	–	–
Nitzschia sigma (Kützing) W.Sm.	–	AJ867279	–	–
Nitzschia sigmoidea (Nitzsch) W.Smith	Nitz1	FR873262	B40 0040713	Germany, Berlin
Nitzschia supralitorea Lange–Bertalot	–	AJ867019	–	–
Nitzschia thermalis (Kützing) Auerswald	–	AY485458	–	–
Nitzschia vitrea G.Norman	–	AJ867280	–	–
Pinnularia acrosphaeria W.Smith	–	AM502021	–	–
Pinnularia anglica Krammer	–	AM501980	–	–
Pinnularia appendiculata (C.Agardh) Cleve	–	AM743095	–	–
Pinnularia brauniana (Grunow) Studnicka	–	AM743097	–	–
Pinnularia cf. *gibba* (Ehrenberg) Ehrenberg	–	EF151977	–	–
Pinnularia divergens W.Smith	D31_023	FR873263	B40 0040714	Mexico, Barranca del Aguacate
Pinnularia gentilis (Donkin) Cleve	–	AM743099	–	–
Pinnularia mayeri Krammer	–	AM743102	–	–
Pinnularia mesolepta (Ehrenberg) W.Smith	–	AM501994	–	–
	–	AM502024	–	–
Pinnularia microstauron (Ehrenberg) Cleve	–	AM501981	–	–
	–	AM501982	–	–
	–	AM501983	–	–
	–	AM501984	–	–
Pinnularia obscura Krasske	–	AM743104	–	–
	–	AM743105	–	–
Pinnularia rupestris Hantzsch	–	AM501992	–	–
Pinnularia subcapitata W.Gregory	–	AM501979	–	–
Pinnularia substreptoraphe Krammer	–	AM502036	–	–
Pinnularia viridis (Nitzsch) Ehrenberg	–	AM502023	–	–
Stauroneis anceps Ehrenberg	–	AM502008	–	–
Stauroneis constricta Ehrenberg	–	AY485521	–	–
Stauroneis gracilior E.Reichardt	–	AM501988	–	–
Stauroneis kriegeri R.M.Patrick	–	AM501990	–	–
	–	AM502037	–	–
Stauroneis phoenicenteron (Nitzsch) Ehrenberg	Stau1	FR873264	B40 0040715	Germany, Berlin
	–	AM501987	–	–
	–	AM502031	–	–
Stephanodiscus agassizensis Håkansson. & H.J.Kling	–	DQ514895	–	–
Stephanodiscuc binderanus (Kützing) Willi Krieger	–	DQ514896	–	–
Stephanodiscus hantzschii Grunow	–	DQ514914	–	–
Stephanodiscus minutulus (Kützing) Cleve & Möller	D03_012	FR873265	B40 0040716	Germany, Berlin

Table 1 (continued)

Taxon	Strain	EMBL no.	B voucher	Locality
	–	DQ514900	–	–
	–	DQ514911	–	–
	–	DQ514915	–	–
	–	DQ514916	–	–
Stephanodiscus neoastraea Håkansson & B.Hickel	–	DQ514906	–	–
Stephanodiscus niagarae Ehrenberg	–	DQ514907	–	–
	–	DQ514908	–	–
Stephanodiscus reimeri Theriot	–	DQ514909	–	–
Stephanodiscus yellowstonensis Simonsen	–	DQ514910	–	–

[a] Epitype strain

[b] Treated as *Sellaphora minima* (Grunow) D.G. Mann by some authors

Methods

Taxon sampling

One hundred twenty three taxa from a wide range of genera throughout Bacillariophyta were used to test the universal applicability of different primer pairs of the 18S rRNA gene. The taxa sampled, the sample origins and/or corresponding EMBL numbers are listed in Tables 1 and 2. Vouchers of sequenced material are deposited in the Herbarium of the Botanic Garden and Botanical Museum Berlin-Dahlem (B), and described in more detail in AlgaTerra (Jahn and Kusber 2002+).

To specifically test the power of the proposed barcode region to distinguish between closely related species, the genus *Sellaphora* (incl. *Sellaphora pupula*-group) was chosen as a test case (Table 2). This is a diatom genus with well-defined biological species concepts (Evans et al. 2007, 2008) as well as vouchered sequences.

Cultivation

DNA was isolated from non-axenic unialgal cultures derived from single cells isolated from environmental samples. The cultures were raised on a modified WC medium (Guillard and Lorenzen 1972) with salt concentrations of 28 g/l of $CaCl_2$, 21 g/l of Na_2SiO_3 and 0.01 g/l of $CuSO_4$. The cultures were stored in petri dishes sealed with Parafilm® M (American National Can Group; Chicago, IL) at 15–17°C and a 12 hday/night rhythm, or at room temperature and the ambient day/night cycle.

DNA isolation

The harvested cultures were transferred to 1.5 ml tubes. DNA was isolated using either Dynal® DynaBeads (Invitrogen Corporation; Carlsbad, CA, USA) or Qiagen® Dneasy Plant Mini Kit (Qiagen Inc.; Valencia, CA) following the respective product instructions. DNA concentrations were checked using gel electrophoresis (1.5% agarose gel) and Nanodrop® (PeqLab Biotechnology LLC; Erlangen, Germany). DNA samples were stored at −20°C until further use.

Secondary structure analysis

The secondary structure of the V4 region was analysed using Mfold (Zuker 2003) running under standard RNA settings (default), and compared to the secondary structure of a consensus sequence (Alverson et al. 2006) to identify possible primer regions within the 18S locus. Primers were designed manually. To assess the variability of the fragment within any given primer pairing, the consensus sequence of Alverson et al. (2006) was used.

Primer testing

All primers given in Table 3 were also tested for amplification and sequencing success at annealing temperatures of 50–54°C under the PCR regime mentioned below. Melting temperature, dimerisation between primer pairs and within single primers, as well as GC content were determined using SeqState under default settings (Müller 2005).

PCR amplification

The V4 region of the 18S locus was amplified using different primer combinations (Table 3). The polymerase chain reaction (PCR) mix (25 µl) consisted of 14.65 µl HPLC H_2O, 2.5 µl 10× buffer S, 1.5 µl $MgCl_2$, 2.5 µl pecGOLD dNTPs, 0.5 µl BSA, 1 µl of each primer (20 pm/µl), 0.35 µl pecGOLD Pur Taq® (all products by PeqLab Biotechnology), and 1 µl DNA sample. The PCR regime included an initial denaturation at 94°C (2 min), then five

Table 2 List of the tested *Sellaphora* taxa with corresponding phenodemes, clone names (both after Evans et al. 2007, 2008) and EMBL accession numbers (18S rRNA)

Taxon	Corresponding phenodeme	Clone	EMBL no.
Sellaphora auldreekie D.G.Mann & S.M.McDonald	–	DUN1	EF151965
Sellaphora bacillum (Ehrenberg) D.G.Mann	–	BLA3	EF151980
Sellaphora blackfordensis D.G.Mann & S.Droop	–	BLA6	EF151969
Sellaphora capitata D.G.Mann & S.M.McDonald	–	BLA10	EF151971
Sellaphora cf. *minima* (Grunow) D.G.Mann	–	BM42	EF151966
Sellaphora cf. *seminulum* (Grunow) D.G.Mann	–	TM37	EF151967
Sellaphora lanceolata D.G.Mann & S.Droop	–	BLA13	EF151978
Sellaphora laevissima (Kützing) D.G.Mann	–	SCOT	AJ544655
	–	THR1	EF151979
	–	THR4	EF151981
Sellaphora pupula (Kützing) Mereschkovsky	Afro	AFR1	EF151975
	cf. capitate	THR7	EF151976
	cf. obese	BEL2	EF151973
	Elliptical	RGB1	EF151962
	Elliptical	THR9	EF151972
	Europa	GER1	EF151984
	Large	THR11	EF151970
	Little	RGB2	EF151964
	Small lanceolate	THR14	EF151963
	Spindle	BLA16	EF151974
	Southern capitate	AUS1	EF151982
	Southern pseudocapitate	AUS4	EF151983

Table 3 Tested primer sequences for 18S amplificates (M13 tails shown in italics; → =forward, ← =reverse)

Primer	Sequence 5′-3′	Direction	Source	GC content	MT in °C
D1NS1	CTA GTC ATA CGC TCG TCT C	→	Brinkmann et al. (pers. comm.)	52.0%	64.6
D2NS1	GTA GTC ATA CCC TCG TCT C	→	Brinkmann et al. (pers. comm.)	52.0%	64.6
D1800R	GCT TGA TCC TTC TGC AGG T	←	Brinkmann et al. (pers. comm.)	52.0%	64.6
Algen F	CTG GTT GAT CCT GCC AGT AG	→	This paper	53.8%	66.4
Primer IR3	GGT AAT TTA CGC GCC TGC T	←	This paper	52.0%	64.6
D1294for 18S	TTY CGT TAA CGA ACG AGA CC	→	This paper	48.1%	64.0
D514for 18S	TCC AGC TCC AAT AGC GTA	→	This paper	50.0%	61.8
D356rev 18S	GGT AAT YTA CGC GCS TGC	←	This paper	54.3%	63.4
D514rev 18S a	ATA CGC TAT TGG AGC TGG	←	This paper	50.0%	61.8
D1069rev 18S a	TCT TTA AGT TTC AGC CTT GC	←	This paper	42.3%	61.6
D512for 18S	ATT CCA GCT CCA ATA GCG	→	This paper	50.0%	61.8
D978rev 18S	GAC TAC GAT GGT ATC TAA TC	←	This paper	42.3%	61.6
M13F-D512for 18S	*TGT AAA ACG ACG GCC AGT* ATT CCA GCT CCA ATA GCG	→	This paper; M13 tail after Messing (1983)	–	–
M13R-D978rev 18S	*CAG GAA ACA GCT ATG AC* GAC TAC GAT GGT ATC TAA TC	←	This paper; M13 tail after Messing (1983)	–	–
M13F (-21)	TGT AAA ACG ACG GCC AGT	→	Messing (1983)	–	–
M13R (-27)	CAG GAA ACA GCT ATG AC	←	Messing (1983)	–	–

MT melting temperature

cycles consisting of denaturation at 94°C (45 s), annealing at 52/54°C (45 s), respectively, and elongation at 72°C (1 min), followed by 35 cycles in which the annealing temperature was lowered to 50/52°C, and a final elongation at 72°C (10 min). PCR products were visualised in a 1.5% agarose gel and cleaned with MSB Spin PCRapace® (Invitek LLC; Berlin, Germany) following standard procedure. DNA content was measured using Nanodrop (PeqLab Biotechnology).

A second PCR following the same protocol and primers (modified with 6 bp long 454 primertails for sample identification) was run to produce samples for the 454 sequencing. After PCR they were also cleaned with MSB Spin PCRapace® (Invitek LLC) following standard procedure. The samples were normalised to a total DNA content >200 ng using Nanodrop (PeqLab Biotechnology).

Sequencing

Sanger sequencing was used for the establishment of reference sequences, whereas 454 sequencing was conducted to establish intragenomic diversity. The Sanger sequencing was conducted by Starseq® (GENterprise LLC; Mainz, Germany). As sequencing primers the M13 tails were used (Table 3), following Ivanova et al. (2007). M13 tails consist of 17–18 bases that are attached at the 5′ end of the regular PCR primer during oligo synthesis. The M13 sequences become amplified at both ends of the PCR product and subsequently can be used as sequencing primers. This prevents loss of sequence information compared to the use of normal internal sequencing primers. As M13 tails can be attached to any primer, only one pair of sequencing primers are necessary regardless of the PCR primers used.

The sequences were edited in ChromasPro (Technelysium Pty. Ltd.; Tewantin, Australia), aligned using ClustalW (Larkin et al. 2007), and manually improved in BioEdit (Hall 1999).

Sequences for intragenomic comparisons were generated with a 454 sequencer (454 Life Sciences; Roche Company; Branford, CT) using GS FLX Titanium® chemistry, following the manufacturer's instructions. All sequences were compared against the reference sequence database created via Sanger sequencing. Only sequences with a complete primer sequence and longer than 250 bp were included.

Statistics

For analysis of the intraspecific and intrageneric variation, sequences from Sanger sequencing (35 sequences; Table 1, EMBL accession numbers FR873231 to FR873265) were used and complemented with sequences downloaded from EMBL (164 sequences; Table 1, all remaining EMBL accession numbers).

Uncorrected *p*-distances were computed using both DOINK (J. Ehrman, Digital Microscopy Facility, Mount Allison University, Sackville, NB, Canada) and PAUP 4.0b10 (Swofford 2002), as the former program cannot interpret ambiguity coding, whereas the latter does not distinguish between gaps and missing data. The significance of the divergence between intraspecific and intrageneric genetic distances was tested with the Wilcoxon rank-sum test using R (R Development Core Team 2005).

Results

DNA isolation

Non-destructive DNA isolation with the Dynal® DynaBeads generally yielded more DNA (up to 50%; details available from the authors upon request) than isolation with the Qiagen® Dneasy Plant Mini Kit for which the diatom frustules were crushed before the extraction procedure was started.

PCR protocol

First the entire 18S rRNA gene was screened for genetic variability between several diatom taxa for barcoding purposes. Then different fragments of high variability, short enough to be sequenced in one read (454 and Sanger), were tested for universal primer binding sites, PCR amplification and sequencing success. A summary of amplification and sequencing success, fragment lengths and variable positions within a fragment is given in Table 4. Among the tested primer pairs, D512for 18S and D978rev 18S as well as their M13 derivates were successful in 100% of the tested taxa in both amplification and sequencing, with the PCR regime given below gaining the most PCR products. All other primer pairings were less suitable as barcoding primers due to poorer amplification and sequencing success and/or to a worse fragment length/variability ratio (Table 4, Fig. 1). Furthermore, the fragment enclosed by the D512for/D978rev primer pair is short enough to be sequenced in one read and has at least 60 putatively variable basepair (bp) positions. The automated primer design software SeqState (Müller 2005) also favoured the application of this primer pair.

DNA sequencing

Sanger sequencing produced sequences of 35 taxa from unialgal cultures (Table 1, EMBL accession numbers FR873231 to FR873265).

The number of generated sequences (454 sequencing) for calculating the intragenomic variation varies between 16 and 112 per taxon (total 1010; Table 5). All sequences >250 bp

Table 4 Percentage of successful amplifications (annealing temperature regime 1 (52–50°C) / regime 2 (54–52°C)) and percentage of successful sequences of the amplificates from PCR regimes 1 and 2 in all 35 taxa; fragment length and number of variable positions on the given fragment following Alverson et al. (2006) for each primer pair

Primer pair	PCR success (regimes 1 / 2)	Sequencing success (regimes 1 / 2)	Fragment length in bp	Positions >20% variable
D514for 18S, D1069rev 18S	93% / 100%	100% / 100%	ca. 500	70
AlgenF, Primer IR3	66% / 90%	68% / 85%	ca. 400	43
D1294for 18S, D1800R	79% / 76%	96% /100%	ca. 700	69
(D1NS1, D2NS1), D514rev 18S	72% / 72%	71% / 71%	ca. 450	43
(D1NS1, D2NS1), D356rev 18S	66% / 62%	68% / 68%	ca. 400	42
D512for 18S, D978rev 18S	100% / 100%	100% / 100%	ca. 390–410	60
M13F D512for 18S, M13R D978rev 18S	100% / 100%	100% / 100%	ca. 390–410	60

from the 454 run could be assigned unambiguously to one of the reference sequences from the Sanger sequencing.

Genetic distances and statistics

To analyse genetic distances between and within strains (several sequences analysed for one unialgal culture), species and genera for the proposed 18S rRNA gene fragment (V4), uncorrected p-distances were calculated. The average, minimum and maximum p-distance values are given in Table 5. Average genetic distance within one strain varied between p=0.000 (*Nitzschia acicularis*, *N. linearis*) and p=0.005 (*Hantzschia amphioxys*). Intraspecific variation also ranged between p=0.000 (e.g. *Achnanthidium minutissimum*) and p=0.005 (*Nitzschia pusilla*, *Pinnularia mesolepta*, *Stauroneis kriegeri*). Intrageneric distance varied between p=0.011 (*Mayamaea* spp.) and p=0.174 (*Melosira* spp.), except for *Stephanodiscus* spp., in which the average intrageneric variation was only p=0.001 (Table 5). Except for *Stephanodiscus*, intrageneric (heterospecific) variation was always higher than both, intraspecific variation and the variation within each strain (for example, intraspecific variation in *Aulacoseira* varied between p=0.000 and p=0.001 while intrageneric distance was p=0.048; Table 5). The Wilcoxon rank-sum test showed that the genetic distances within the species of the 16 tested genera (Table 5) is significantly lower than between the single species in these genera ($p = 2.2 \times 10^{-16}$; Fig. 2).

Genetic distance among taxa in *Sellaphora* ranged between p=0.003 (*Sellaphora blackfordensis*/*Sellaphora pupula* phenodeme southern pseudocapitate) and p=0.087 (*Sellaphora* cf. *minima*/*Sellaphora pupula* phenodeme europa), with an average p=0.039 (Table 6). The average intraspecific genetic distance within *Sellaphora laevissima* is p=0.005 (min. p= 0.000, max. p=0.007; number of sequences=3; Table 6); within *Sellaphora pupula* phenodeme elliptical it is p=0.000 (number of sequences=2; Table 6).

Discussion

The analysis of environmental samples via DNA barcoding needs to facilitate the detection of—in this case diatom—diversity as well as the identification of species present in the respective sample. For the first part a standard laboratory protocol (including universal primers) is essential, for the second a critical assessment of intra- versus interspecific variation is needed.

Standard laboratory protocol

The development of a standard laboratory protocol considered DNA extraction as well as fragment amplification and sequencing including primer design. The DNA extraction using Dynal® DynaBeads is a non-destructive process that leaves the frustules intact and available for microscopic examination and taxonomic determination, e.g. if species have not yet been deposited in a reference database and morphological vouchers have to be cross-checked after sequencing or if mixed samples have to be analysed microscopically and valves have to be counted for quantification. Even if the Qiagen® Dneasy Plant Mini Kit is used non-destructively it includes more centrifuging steps that could damage especially the larger diatom frustules or fragile frustule characteristics that can be crucial for identification.

Concerning the Dynal® DynaBeads method it has to be noted that after the extraction the residue containing the frustules has to be centrifuged, the supernatant removed, and replaced by pH neutral storing buffer. Otherwise the frustules might be dissolved. The DNA yield is higher than with the Qiagen® Dneasy Plant Mini Kit. Because of the better performance and the conservation of the frustules, the non-destructive DNA isolation was chosen.

Of the six different primer pairs that were tested, D512for 18S and D978rev 18S, as well as their M13

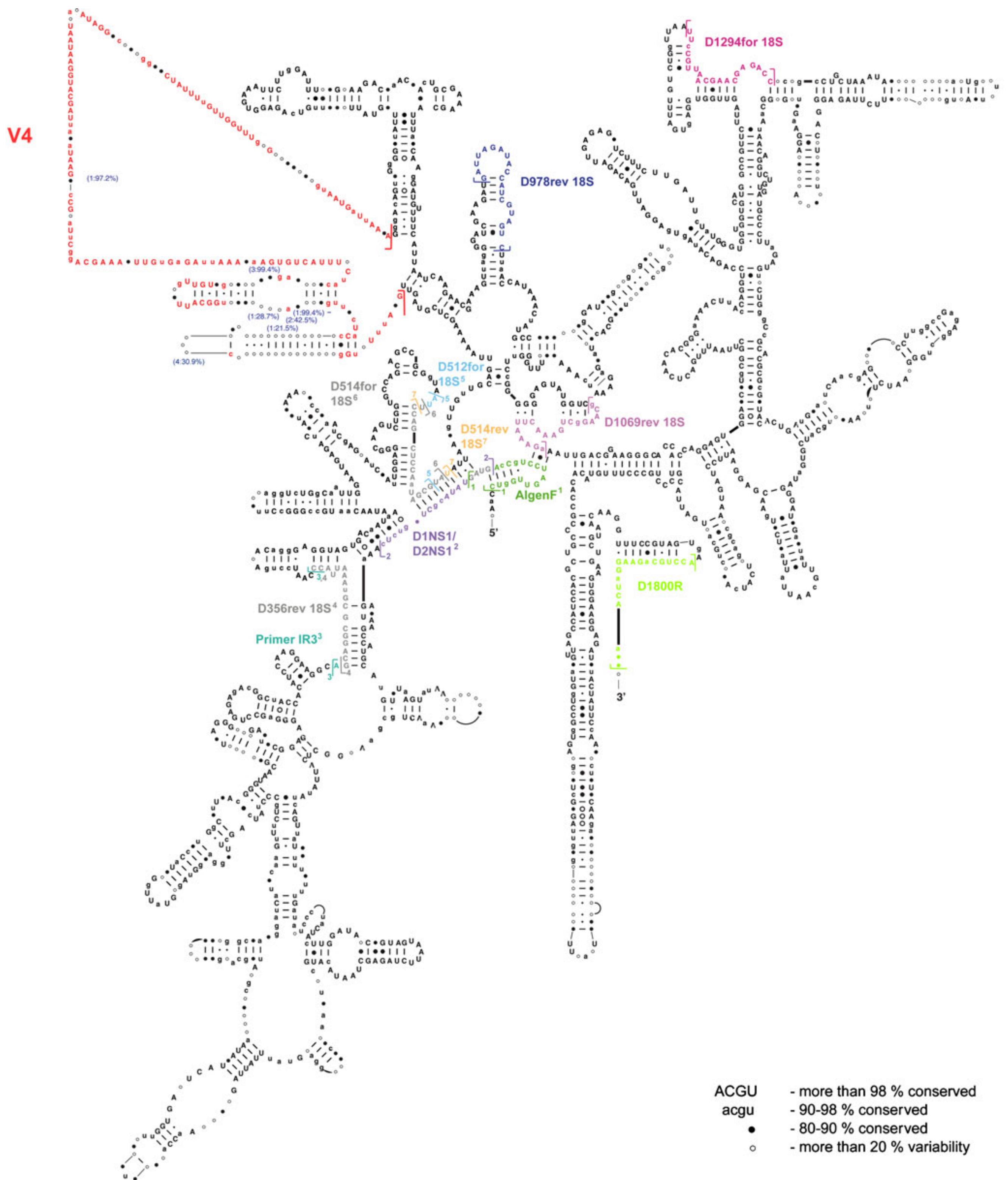

Fig. 1 Consensus secondary structure of the 18S locus (SSU rRNA gene) in diatoms (181 sequences), based on the *Toxarium undulatum* 18S secondary structure model as reference sequence. Upper-case letters indicate that nucleotides at corresponding positions are conserved in 98–100% of sequences, lower-case letters indicate 90–98% conservation, dots 80–90% conservation, circles indicate greater than 20% variability. V4 region and primer binding sites (see Table 3) shown highlighted and in brackets. Where primers overlap their names and brackets are numbered accordingly. Tags at V4 region indicate indels relative to *Toxarium undulatum* sequence; tag format is (maximum length of indel: percentage of sequences showing length polymorphisms). Figure modified after Alverson et al. (2006) and Gillespie et al. (2006)

Table 5 Uncorrected *p*-distances given as average, minimum and maximum values; *n* = number of sequences (per strain) or number of individuals (per species/genus), respectively

Taxon	*n*	*p*-distance		
		Average	Minimum	Maximum
Within strains				
Achnanthidium minutissimum	74	0.003	0.000	0.007
Aulacoseira granulata	24	0.001	0.000	0.002
Cyclotella meneghiniana	104	0.001	0.000	0.004
Diatoma tenue	20	0.003	0.000	0.007
Encyonema minutum	112	0.002	0.000	0.006
Fragilaria melosepta	29	0.003	0.000	0.006
Gomphonema parvulum	17	0.003	0.000	0.005
Hantzschia amphioxys	50	0.005	0.000	0.008
Mayamaea atomus var. *permitis*	67	0.002	0.000	0.008
Melosira varians	93	0.001	0.000	0.008
Navicula slesvicensis	48	0.002	0.000	0.007
Navicula tripunctata	16	0.001	0.000	0.003
Nitzschia acicularis	96	0.000	0.000	0.002
Nitzschia cf. *semirobusta*	54	0.001	0.000	0.006
Nitzschia linearis	95	0.000	0.000	0.000
Pinnularia divergens	24	0.002	0.000	0.008
Stephanodiscus minutulus	87	0.001	0.000	0.004
Within species				
Achnanthidium minutissimum	2	0.000	–	–
Aulacoseira ambigua	6	0.001	0.000	0.002
Aulacoseira baicalensis	2	0.000	–	–
Aulacoseira granulata	3	0.001	0.000	0.002
Aulacoseira nyassensis	2	0.001	0.000	0.002
Aulacoseira islandica	3	0.000	0.000	0.000
Aulacoseira subarctica	4	0.001	0.000	0.002
Cocconeis pediculus	4	0.000	0.000	0.000
Cocconeis placentula	5	0.000	0.000	0.000
Cyclotella meneghiniana	2	0.004	0.000	0.007
Diatoma tenue	6	0.001	0.002	0.000
Encyonema minutum	2	0.000	–	–
Eolimna minima	2	0.000	–	–
Fragilaria nanana	2	0.000	–	–
Fragilaria striatula	3	0.003	0.002	0.005
Fragilaria vaucheriae	3	0.000	0.000	0.000
Gomphonema parvulum	2	0.008	–	–
Hantzschia amphioxys	2	0.001	–	–
Mayamaea atomus var. *permitis*	3	0.000	0.000	0.000
Melosira varians	3	0.002	0.000	0.003
Navicula cryptocephala	3	0.000	0.000	0.000
Navicula cryptotenella	2	0.002	0.002	0.002
Navicula gregaria	2	0.000	–	–
Navicula radiosa	3	0.000	0.000	0.000
Navicula tripunctata	2	0.000	–	–
Nitzschia acicularis	2	0.000	–	–
Nitzschia cf. *semirobusta*	2	0.000	–	–

Table 5 (continued)

Taxon	*n*	*p*-distance		
		Average	Minimum	Maximum
Nitzschia closterium	2	0.000	–	–
Nitzschia communis	2	0.000	–	–
Nitzschia dissipata	2	0.000	–	–
Nitzschia inconspicua	2	0.000	–	–
Nitzschia linearis	4	0.003	0.000	0.005
Nitzschia palea	3	0.002	0.000	0.003
Nitzschia paleaformis	2	0.000	–	–
Nitzschia pusilla	2	0.005	–	–
Pinnularia mesolepta	2	0.005	–	–
Pinnularia microstauron	4	0.000	0.000	0.000
Pinnularia obscura	2	0.000	–	–
Stauroneis kriegeri	2	0.005	–	–
Stauroneis phoenicenteron	3	0.000	0.000	0.000
Stephanodiscus minutulus	5	0.000	0.000	0.000
Within genera				
Achnanthidium	2	0.129	–	–
Aulacoseira	9	0.048	0.018	0.085
Cocconeis	3	0.134	0.044	0.181
Cyclotella	7	0.059	0.018	0.098
Diatoma	3	0.029	0.021	0.038
Encyonema	4	0.032	0.019	0.056
Eolimna	2	0.104	–	–
Fragilaria	8	0.074	0.018	0.124
Gomphonema	5	0.072	0.042	0.101
Mayamaea	2	0.011	–	–
Melosira	2	0.174	–	–
Navicula	20	0.069	0.0155	0.186
Nitzschia	19	0.093	0.013	0.212
Pinnularia	17	0.168	0.015	0.329
Stauroneis	5	0.064	0.015	0.129
Stephanodiscus	8	0.001	0.000	0.005

variants, were the most successful with respect to amplification and sequencing success, and exhibited the best fragment length/variability ratio (Table 4, Fig. 1). PCR amplification with primers D512for 18S and D978rev 18S was successful in all taxa in our study and in many other taxa (e.g. *Skeletonema* spp., *Phaeodactylum* spp., *Surirella* spp., *Campylodiscus* spp.; authors' unpublished data). This high amplification efficiency is due to the placement of the primers in highly conserved stemloop sections of the 18S rRNA gene (Fig. 1) that exhibit low mutation rates and are conserved across a wide range of diatom taxa, therefore make ideal binding sites for universal primers. The M13 tails were used as universal sequencing primers (Ivanova et al. 2007), which contributed to the high sequencing success.

Importantly, the primer combination D512for 18S and D978rev 18S includes the highly variable V4 region of the 18S rRNA gene (Fig. 1) which encloses many indel regions that contribute to the increased information level on this short fragment (Alverson et al. 2006). The other tested primer pairs also result in short variable segments, but with lower universality concerning the laboratory success. The fragments are also less variable, thus do not allow species-level identification within diatoms (Fig. 1).

Besides the primer universality, the V4 region has another promising feature for barcoding environmental samples: The association of the sequences produced by 454 sequencing to the reference data generated via Sanger sequencing was always unambiguously possible—due to the systematic selection procedure—without much computing

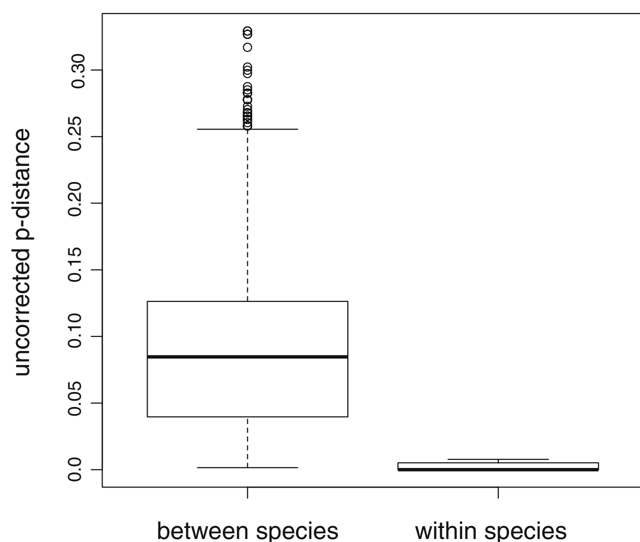

Fig. 2 Box-and-whisker plot of intraspecific and intrageneric (x-axis) genetic distances measured in uncorrected *p*-distances (y-axis). Thick black lines indicate median values, boxes represent upper and lower quartiles, whiskers indicate value ranges, circles represent outliers

and editing effort after sequencing. In addition, no problems emerged in the present study concerning homopolymer errors in the sequences as are often encountered when applying pyrosequencing (Huse et al. 2007).

For high-throughput studies it is also important that the barcode does not exceed a certain length, currently around 400 bp. This length keeps increasing along with the development of sequencing techniques and computation capacity (Schloss 2010), but the cost of sequencing increases accordingly. This is one reason why Hajibabaei et al. (2006b) proposed a 100 bp barcode, which would also work with high-throughput technologies that only produce shorter read length such as Illumina. The V4 region (Fig. 1) in itself is only about 60 bp long, so that it could qualify as such a short barcode without losing its resolving power. Some studies already use very short sequences to evaluate prokaryotic diversity in environmental samples (Huber et al. 2009; Huse et al. 2007; Schloss 2010).

For these reasons, standard laboratory protocols, primer universality, informational indels on a short fragment, the V4 region—maybe only a 60 bp part of it—show high potential for the use in fast, high-throughput approaches to environmental barcoding using next-generation sequencing.

Species identification

For the assessment of the 18S fragment's power to resolve taxa at species level, uncorrected *p*-distances were used. All species tested in this study feature uniform sequences allowing unambiguous resolution at species level, with the only exception concerning *Stephanodiscus*. This genus is well known as problematic in morphological discriminations due to small size of the individuals and to valve plasticity which is often overlapping between species (Håkansson and Kling 1989; 1990; Kobayasi et al. 1985; Spamer and Theriot 1997; Teubner 1997; Wolf et al. 2002). Molecular species identification in *Stephanodiscus* is also difficult (Moniz and Kaczmarska 2009, 2010), possibly because some taxa have diverged only very recently, e.g. *S. niagarae* and *S. yellowstonensis* about 12.000 to 8.000 years ago (Zechman et al. 1994).

Intraspecific variation was very low in general, not exceeding $p=0.005$ (*Hantzschia amphioxys*, Table 5). Intrageneric variation was significantly higher than intraspecific variation in all cases (Table 5). This leads to the assumption that, even though the *p*-distances are comparatively low compared to other markers (e.g. Huang et al. 2007; Wu et al. 2008; Xia et al. 2003), the 18S fragment (V4) used in the present study still has informative value as a barcoding marker to resolve taxa at the species level.

So far, the resolving ability of a given barcode marker has been assessed using either a fixed threshold or the concept of the "barcode gap" (Hollingsworth et al. 2009), meaning a well-defined difference between the levels of intra- and interspecific variation, often calculated by means of a ratio. Initially some studies used a 10-fold increase to gauge the applicability of a certain marker (Hebert et al. 2003). More recently, however, it has been shown that taxa differ considerably in their genetic variation, so that different studies now use very different ratios and thresholds depending on the respective organism group and marker (e.g. Cywinska et al. 2006; Hajibabaei et al. 2006a; b; Hebert et al. 2004; Hickerson et al. 2006; Meyer and Paulay 2005; Ward et al. 2005). For the *cox*1 gene a threshold of $p=0.04$ is considered sufficient in red algae (Saunders 2005), for the ciliate genus *Tetrahymena* $p=0.11$ (Chantangsi et al. 2007), and for *Paramecium* $p=0.20$ (Barth et al. 2006). Moniz and Kaczmarska (2009) give a minimum intrageneric distance of $p=0.07$ for a combination of the 5.8S rRNA gene and ITS2 within diatoms.

The variation in the 18S rRNA gene has been considered as too low for a barcoding marker in diatoms (Moniz and Kaczmarska 2009, 2010). This, however, refers to the complete 18S locus, which is much longer (1800 bp) than the one used in the present study (ca. 390–410 bp). As most of the 1800 bp fragment comprises extremely conserved regions, the genetic distance between species is reduced if the complete 18S rRNA gene locus is used. In the present study the region responsible for species identification is mainly the only ca. 60 bp long V4 region (Fig. 1). As mentioned above, the V4 region comprises not only many variable character sites but also many inversions, insertions and deletions, resulting in a highly concentrated information content on a very short fragment (Alverson et al. 2006).

Table 6 Uncorrected *p*-distances among tested taxa (clones and/or phenodemes) in the genus *Sellaphora*. Values in boxes labeled a–e are discussed in detail

Taxon (clone and/or phenodeme)	*p*-distance																					
	S. blackfordensis clone BLA6	*S. bacillum* clone BLA3	*S. lanceolata* clone BLA13	*S. capitata* clone BLA10	*S.* cf. *seminulum* clone TM37	*S.* cf. *minima* clone BM42	*S. auldreekie* clone DUN1	*S. laevissima* clone THR4	*S. laevissima*. strain SCOT	*S. laevissima* clone THR1	*S. pupula* clone GER1 europa	*S. pupula* clone THR9 elliptical	*S. pupula* clone BLA16 spindle	*S. pupula* clone RBG1 elliptical	*S. pupula* clone RBG2 little	*S. pupula* clone AUS4 southern pseudocapitate	*S. pupula* clone AUS1 southern capitate	*S. pupula* clone THR7 cf. capitata	*S. pupula* clone THR11 large	*S. pupula* clone THR14 small lanceolate	*S. pupula* clone AFR1 afro	*S. pupula* clone BEL2 cf. obese
S. blackfordensis clone BLA6	-																					
S. bacillum clone BLA3	0.026	-																				
S. lanceolata clone BLA13	0.033	0.008	-																			
S. capitata clone BLA10	0.020	0.038	0.046	-																		
S. cf. *seminulum* clone TM37	0.038	0.036	0.038	0.054	-																	
S. cf. *minima* clone BM42	0.074	0.069	0.069	0.084	0.056	-																
S. auldreekie clone DUN1	0.026	0.026	0.028	0.036	0.041	0.071	-															
S. laevissima clone THR4	0.031	0.031	0.038	0.038	0.048	0.071	0.041	-														
S. laevissima strain SCOT	0.031	0.031	0.038	0.038	0.048	0.071	0.041	0.000														
S. laevissima clone THR1	0.036	0.038	0.046	0.036	0.054	0.077	0.048	0.008	0.008	-												
S. pupula clone GER1 europa	0.043	0.043	0.046	0.051	0.056	0.087	0.038	0.046	0.046	0.051	-											
S. pupula clone THR9 elliptical	0.031	0.031	0.033	0.038	0.043	0.074	0.026	0.031	0.031	0.038	0.018	-										
S. pupula clone BLA16 spindle	0.031	0.033	0.036	0.038	0.043	0.079	0.028	0.036	0.036	0.043	0.026	0.010	-									
S. pupula clone RBG1 elliptical	0.031	0.031	0.033	0.038	0.043	0.074	0.026	0.031	0.031	0.038	0.018	0.000	0.010	-								
S. pupula clone RBG2 little	0.036	0.036	0.038	0.043	0.041	0.071	0.031	0.036	0.036	0.043	0.023	0.008	0.013	0.008	-							
S. pupula clone AUS4 southern pseudocapitate	0.003	0.028	0.036	0.020	0.041	0.077	0.028	0.028	0.028	0.033	0.043	0.031	0.031	0.031	0.036	-						
S. pupula clone AUS1 southern capitate	0.008	0.028	0.036	0.015	0.041	0.071	0.028	0.031	0.031	0.033	0.046	0.033	0.033	0.033	0.038	0.010	-					
S. pupula clone THR7 cf. capitata	0.015	0.033	0.041	0.015	0.048	0.079	0.033	0.038	0.038	0.038	0.051	0.038	0.038	0.038	0.043	0.018	0.010	-				
S. pupula clone THR11large	0.041	0.041	0.043	0.048	0.054	0.084	0.036	0.041	0.041	0.048	0.026	0.015	0.018	0.015	0.018	0.041	0.043	0.048	-			
S. pupula clone THR14 small lanceolate	0.048	0.036	0.043	0.059	0.064	0.079	0.048	0.054	0.054	0.059	0.054	0.041	0.048	0.041	0.046	0.048	0.054	0.059	0.054	-		
S. pupula clone AFR1 afro	0.020	0.038	0.046	0.000	0.054	0.084	0.036	0.038	0.038	0.036	0.051	0.038	0.038	0.038	0.043	0.020	0.015	0.015	0.048	0.059	-	
S. pupula clone BEL2 cf. obese	0.010	0.036	0.043	0.020	0.048	0.084	0.036	0.038	0.038	0.043	0.048	0.036	0.036	0.036	0.041	0.010	0.018	0.026	0.046	0.054	0.020	-

The V4 region appears to allow discrimination between species to a degree sufficient for environmental DNA barcoding. Therefore, to further test the power of this region for species identification in a closely related taxon complex, an exclusive in silico analysis within the *Sellaphora pupula*-group and sister taxa was performed. The genus *Sellaphora* is a genus with well-established species concepts and extensive data on mating behaviour, morphology, ecology and DNA sequence variation within the genus (Evans et al. 2007, 2008). The *Sellaphora pupula*-group consists of very closely related species, thus provides a strong test of the reliability of the proposed barcode region. The V4 region was able to discriminate between all the included taxa (following Evans et al. 2008).

There are some taxon pairs with very low genetic distances (Table 6, b–d), one of them comprising *Sellaphora blackfordensis* and *S. pupula* clone AUS4 phenodome southern pseudocapitate, (Table 6, b), the second *S. blackfordensis* and *S. pupula* clone AUS1 phenodome southern capitate (Table 6, c). These three taxa also form a well-supported clade in the *rbcL*-based phylogenetic tree provided by Evans et al. (2008). The third such pair contains *Sellaphora lanceolata* and *S. bacillum* (Table 6, d), showing a relationship which is consistent with the findings of Evans et al. (2008) as well. That the genomic variation between these pairs is lower than or similar to the variation within *Sellaphora laevissima* could indicate, for instance, that the V4 region is not powerful enough to distinguish

between all cryptic species, or that the species circumscriptions do not necessarily reflect the genetic diversity.

Within the former *Sellaphora pupula* taxon there are two identical sequences (Table 6, e), both designated as *S. pupula* phenodeme elliptical by Evans et al. (2008). Whether the genetic distances between these phenodemes represent population differences or variation between cryptic species needs further consideration (e.g. Evans et al. 2008). This shows that the V4 region also may have some potential for identifying closely related species, even though it might not be enough for defining them.

The V4 region of the 18S locus as a barcode marker

Various other barcodes have been proposed for various groups of organisms, among them the plastid regions *rbcL*, *matK*, *trnH-psbA*, the 23S rRNA gene, the mitochondrial gene *cox*1, and the nuclear markers ITS, entire 18S (SSU) rRNA gene and 28S (LSU) rRNA gene (e.g. Bhadury et al. 2006; Fazekas et al. 2008; Hebert et al. 2004; Hollingsworth et al. 2009; Kress and Erickson 2007; Kress et al. 2005; Newmaster et al. 2008; Summerbell et al. 2005). However, *cox*1, ITS, 18S and *rbcL* are the only ones which have been applied to diatoms, with mixed results, i.e. *cox*1 was very variable but no universal primers could be found, ITS was variable but is not universally amplifiable with standard laboratory protocols, *rbcL* was less variable, and 18S (whole gene) was not variable enough (e.g. Evans et al. 2007, 2008; Jahn et al. 2007; Moniz and Kaczmarska 2009, 2010).

That the *cox*1 gene is variable enough to discriminate between very similar taxa (e.g. cryptic species) has been stated for many groups throughout the tree of life (Barth et al. 2006; Chantangsi et al. 2007; Hebert et al. 2003; Kucera and Saunders 2008; Lynn and Strüder-Kypke 2006; Saunders 2005). However, a preliminary study using a dataset of over 60 diatom species from various groups to design universal primers for the *cox*1 gene (unpublished data) showed that it is virtually impossible to do so, because the locus lacks sufficiently conserved regions for primer binding. Universal primers constitute an essential condition for environmental analyses. Various publications have shown that this problem occurs not only within diatoms (e.g. Evans et al. 2007, 2008; Moniz and Kaczmarska 2009) but also in many other eukaryotic organism groups, e.g. in land plants (Cowan et al. 2006), dinoflagellates (Ferrell and Beaton 2007), gastropods (Kane et al. 2008), and fungi (Seifert et al. 2007). Most studies on the use of the *cox*1 gene as a barcoding marker for protists are limited to very confined groups, e.g. genera, and use group-specific primers (Chantangsi et al. 2007; Evans et al. 2007, 2008). In diatoms this high variability of the *cox*1 locus could be due to the occurrence of intron events and introgression of bacterial genes, both common in diatoms (Armbrust et al. 2004; Bowler et al. 2008; Ehara et al. 2000; Imanian et al. 2007; Ravin et al. 2010).

The combination of the 5.8S rRNA gene and ITS2 has been suggested as an alternative barcoding locus (Moniz and Kaczmarska 2009, 2010). Its potential to identify species is promising and has been demonstrated in many protists, fungi and plant groups (e.g. Gemeinholzer et al. 2006; Kelly et al. 2010; Litaker et al. 2007; Taylor et al. 2008). There are, however, some problems, the main one being that ITS is not easy to amplify and sequence with standard laboratory protocols (unpublished data; see also Hamsher et al. 2011). Furthermore, studies in fungi using ITS suggested that errors in amplification/sequencing—especially in high throughput—could easily lead to overestimation of diversity in environmental samples (Bellemain et al. 2010).

Plastid markers such as the *rbcL* gene could be problematic for DNA barcoding, as the plastid inheritance in diatoms is not uniform but can be either uniparental or biparental (Casteleyn et al. 2009; Jensen et al. 2003; Levialdi Ghiron et al. 2008; Round et al. 1990), and there are rare reports of natural hybrids (Casteleyn et al. 2009).

The 18S rRNA gene locus is often used to estimate the relative abundances and diversities of species in environmental samples (Liao et al. 2007), due to its low intraspecific but high interspecific variation. It also has been used to define operational taxonomic units (OTUs) in various eukaryots (Ciliophora, Dinophyceae, Cercozoa und Fungi; Lefèvre et al. 2007). The analysis of water samples via a 550 bp long fragment of the 18S rRNA gene locus was able to resolve organisms of the metazoans (e.g. nematodes), the algae Prasinophyceae, Cryptophyceae, Dinophyceae and Prymnesiophyceae, as well as heterotrophic Cercozoa, Choanoflagellates, Stramenopiles, and Cilitates (Romari and Vaulot 2004). It has been shown that the 18S rRNA gene can also discriminate diatoms in most cases of environmental samples, often to the species level (Jahn et al. 2007; Savin et al. 2004).

The main advantage of the V4 fragment of the 18S locus is that it is very easy to amplify with the proposed universal primers using our documented standard laboratory protocol, while it still has considerable power to resolve taxa on the species level. Both of these characteristics are crucial for its successful use in environmental studies. The potential of the V4 fragment to discriminate between (semi-)cryptic species has to be further evaluated. However, while this aspect is desirable it is not necessary for its use in environmental studies, as the members of cryptic-species complexes generally seem to have similar ecology (Beszteri et al. 2005a,b, 2007).

A further advantage of the 18S locus is its high representation in databases. A good retrieval rate for correct identifications strongly depends on the reference data. But

even though the reference database for the 18S rRNA gene is more extensive than for many other proposed barcode regions, it nevertheless has to be extended, especially with voucher-based sequences.

Conclusions

The crucial problem in selecting an applicable barcode is the balance between variability and primer-binding universality. For the analysis of environmental samples primer universality and reproducible laboratory protocols are of high importance, whereas for the detection and delimitation of cryptic species these aspects are often secondary.

For the detection of cryptic species other, more variable barcodes might be more feasible. But as discussed in many other studies, some problems, such as species delimitation and α-taxonomy, presumably cannot be solved with only one barcode (e.g. Chase et al. 2007; Cowan et al. 2006; Kress and Erickson 2007). A single barcode represents only a fraction of an organism's variation; therefore its power to define a taxon should not be overestimated. Consequently, a combination of the V4 region with other barcodes such as ITS should be discussed.

The 18S rRNA gene fragment proposed in the present study shows enough variation to unambiguously identify almost all tested taxa. Furthermore, the highly conserved primer binding sites allow amplification following a standard procedure. Due to its relatively short length it is also feasible for time- and cost-saving high-throughput analysis methods. The V4 region of the 18S locus therefore is a good candidate for barcoding diatoms in environmental samples.

Acknowledgements The authors wish to thank Martin Pfannkuchen and Daniela Maric, Monica Moniz and Irena Kaczmarska, James Ehrman, Neela Enke, Nelida Abarca, Daniel Lauterbach, Wolf-Henning Kusber and Weliton da Silva for fruitful discussions, Oliver Skibbe and Jana Bansemer for diatom cultivation. We also thank Michael Kube and Richard Reinhardt (MPI for Molecular Genetics, Berlin) for providing time and guidance at the 454 sequencer. The Association of the Friends of the Botanic Garden and Botanical Museum Berlin-Dahlem and the Academic Senate of the Freie Universität Berlin have provided financial support.

References

Alverson, A. J., Cannone, J. J., Gutell, R. R., & Theriot, E. C. (2006). The evolution of elongate shape in diatoms. *Journal of Phycology, 42*, 655–668.

Archibald, R. E. M. (1984). Diatom illustrations—an appeal. *Bacillaria, 7*, 173–178.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., et al. (2004). The genome of *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science, 306*, 79–86.

Babanazarova, O. V., Likhoshway, Y. V., & Sherbakov, D. Y. (1996). On the morphological variability of *Aulacoseira baicalensis* and *Aulacoseira islandica* (Bacillariophyto) of Lake Baikal, Russia. *Phycologia, 35*, 113–123.

Bailey-Watts, A. E. (1976). Planktonic diatoms and some diatom-silica relations in a shallow eutrophic Scottish loch. *Freshwater Biology, 6*, 69–80.

Barth, D., Krenek, S., Fokin, S. I., & Berendonk, T. (2006). Intraspecific genetic variation in *Paramecium* revealed by mitochondrial cytochrome *c* oxidase I sequences. *Journal of Eukaryotic Microbiology, 53*, 20–25.

Behnke, A., Friedl, T., Chepurnov, V. A., & Mann, D. G. (2004). Reproductive compatibility and rDNA sequence analyses in the *Sellaphora pupula* species complex (Bacillariophyta). *Journal of Phycology, 40*, 193–208.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., & Kauserud, H. (2010). ITS as an environmental DNA barcode for fungi: An in silico approach reveals potential PCR biases. *BMC Microbiology, 10*, 189. doi:10.1186/1471-2180-10-189.

Beszteri, B., Acs, E., Makk, J., Kovács, G., Márialigeti, K., & Kiss, K. T. (2001). Phylogeny of six naviculoid diatoms based on 18S rDNA sequences. *International Journal of Systematic and Evolutionary Microbiology, 51*, 1581–1586.

Beszteri, B., Ács, É., & Medlin, L. K. (2005a). Conventional and geometric morphometric studies of valve ultrastructural variation in two closely related *Cyclotella* species (Bacillariophyta). *European Journal of Phycology, 40*, 89–103.

Beszteri, B., Ács, É., & Medlin, L. K. (2005b). Ribosomal DNA sequence variation among sympatric strains of the *Cyclotella meneghiniana* complex (Bacillariophyceae) reveals cryptic diversity. *Protist, 156*, 317–333.

Beszteri, B., John, U., & Medlin, L. K. (2007). An assessment of cryptic genetic diversity within the *Cyclotella meneghiniana* species complex (Bacillariophyta) based on nuclear and plastid genes, and amplified fragment length polymorphism. *European Journal of Phycology, 42*, 47–60.

Bhadury, P., Austen, M. C., Bilton, D. T., Lambshead, P. J. D., Rogers, A. D., & Smerdon, G. R. (2006). Development and evaluation of a DNA-barcoding approach for the rapid identification of nematodes. *Marine Ecology Progress Series, 320*, 1–9.

Blaxter, M. L. (2004). The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London, Biological Sciences, 359*, 669–679.

Blaxter, M., Elsworth, B., & Daub, J. (2004). DNA taxonomy of a neglected animal phylum: An unexpected diversity of tardigrades. *Proceedings of the Royal Society of London, Biological Sciences, 271*, 189–192.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature, 456*, 239–244.

Casteleyn, G., Adams, N. G., Vanormelingen, P., Debeer, A. E., Sabbe, K., & Vyverman, W. (2009). Natural hybrids in the marine diatom *Pseudo-nitzschia pungens* (Bacillariophyceae): Genetic and morphological evidence. *Protist, 160*, 343–354.

Chantangsi, C., Lynn, D. H., Brandl, M. T., Cole, J. C., Hetrick, N., & Ikonomi, P. (2007). Barcoding ciliates: A comprehensive study of 75 isolates of the genus *Tetrahynema*. *International Journal of Systematic and Evolutionary Microbiology, 57*, 2412–2425.

Chase, M. W., Cowan, R. S., Hollingsworth, P. M., van den Berg, C., Madriñán, S., Petersen, G., et al. (2007). A proposal for a standardised protocol to barcode all land plants. *Taxon, 56*, 295–299.

Cowan, R. S., Chase, M. W., Kress, W. J., & 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, F. F., & Hebert, P. D. N. (2006). Identifying Canadian mosquito species through DNA barcodes. *Medical and Veterinary Entomology, 20*, 413–424.

Ehara, M., Watanabe, K. I., & Ohama, T. (2000). Distribution of cognates of group II introns detected in mitochondrial *cox*1 genes of a diatom and haptophyte. *Gene, 256*, 157–167.

Erickson, D. L., Spouge, J., Resch, A., Weigt, L. A., & Kress, W. J. (2008). DNA barcoding in landplants: Developing standards to quantify and maximise success. *Taxon, 57*, 1304–1316.

Evans, K. M., Wortley, A. H., & Mann, D. G. (2007). An assessment of potential diatom "barcode" genes (*cox*1, *rbcL*, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist, 158*, 349–364.

Evans, K. M., Wortley, A. H., Simpson, G. E., Chepurnov, V. A., & Mann, D. G. (2008). A molecular systematic approach to explore diversity within the *Sellaphora pupula* species complex (Bacillariophyta). *Journal of Phycology, 44*, 215–231.

Falkowski, P. G., Barber, R. T., & Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science, 281*, 200–207.

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., et al. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PloS One, 3*(7), e2802. doi:10.1371/journal.pone.0002802.

Ferrell, J., & Beaton, M. (2007). The evaluation of DNA barcoding for identification of dinoflagellates: A test using *Prorocentrum*. In: *Canadian barcode of life network 2007 science symposium* (pp. 37). Guelph: Blackwell.

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

Floyd, R., Abebe, E., Papert, A., & Blaxter, M. L. (2002). Molecular barcodes for soil nematode identification. *Molecular Ecology, 11*, 839–850.

Friedl, T., & O'Kelly, C. J. (2002). Phylogenetic relationships of green algae assigned to the genus *Planophila* (Chlorophyta): Evidence from 18S rDNA sequence data and ultrastructure. *European Journal of Phycology, 37*, 373–384.

Gemeinholzer, B., Oberprieler, C., & Bachman, K. (2006). Using GenBank data for plant identification: Possibilities and limitations using the ITS1 of Asteraceae species belonging to the tribes Lactuceae and Anthemideae. *Taxon, 55*, 173–187.

Gillespie, J. J., Johnston, J. F., Cannone, J., & Gutell, R. R. (2006). Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): Structure, organization, and retrotransposable elements. *Insect Molecular Biology, 15*, 657–686.

Guillard, R. R. L., & Lorenzen, C. J. (1972). Yellow green algae with chlorophyllide. *Journal of Phycology, 8*, 10–14.

Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., & Hebert, P. D. N. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the USA, 103*, 968–971.

Hajibabaei, M., Smith, A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., & Hebert, P. D. N. (2006). A minimalist barcode can identify specimens whose DNA is degraded. *Molecular Ecology, 6*, 959–964.

Hajibabaei, M., Singer, G. A. C., Hebert, P. D. N., & Hickey, D. A. (2007). DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics, 23*, 167–172.

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 Research, 4*, 269–288.

Håkansson, H., & Kling, H. (1990). The current status of some very small freshwater diatoms of the genera *Stephanodiscus* and *Cyclostephanos*. *Diatom Research, 5*, 273–287.

Hall, T. A. (1999). BioEdit: A user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleid Acids Symposium Series, 41*, 95–98.

Hamsher, S. E., Evans, K. M., Mann, D. G., Poulíčková, A., & Saunders, G. W. (2011). Barcoding diatoms: Exploring alternatives to COI-5P. *Protist, 162*, 405–422.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & de Waard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Biological Sciences, 270*, 313–321.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. *PLoS Biology, 2*, 1657–1663.

Hickerson, M. J., Meyer, C. P., & Moritz, C. (2006). DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology, 55*, 729–739.

Hollingsworth, M. L., Clark, A. A., Forrest, L. L., Richardson, J., Pennington, R. T., Long, D. G., et al. (2009). Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources, 9*, 439–457.

Huang, J., Xu, Q., Sun, Z. J., Tang, G. L., & Su, Z. Y. (2007). Identifying earthworms through DNA barcodes. *Pedobiologia, 51*, 301–309.

Huber, J. A., Morrison, H. G., Huse, S. M., Neal, P. R., Sogin, M. L., & Welch, D. M. (2009). Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environmental Microbiology, 11*, 1292–1302.

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., & Welch, D. M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology, 8*, R143.

Imanian, B., Carpenter, K. J., & Keeling, P. J. (2007). Mitochondrial genome of a tertiary endosymbiont retains genes for electron transport proteins. *Journal of Eukaryotic Microbiology, 54*, 146–153.

Ivanova, N. V., Zemlak, T. S., Hanner, R. H., & Hebert, P. D. N. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes, 7*, 544–548.

Jahn, R. (1986). A study of *Gomphonema augur* Ehrenberg: The structure of the frustule and its variability in clones and populations. In M. Ricard (Ed.), *Proceedings of the 8th International Diatom Symposium 1984* (pp. 191–204). Paris: Koeltz Scientific Books.

Jahn, R., & Kusber, W. H. (2002+). *AlgaTerra Information System (online)*. Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin. http://www.algaterra.org. Accessed 30 December 2010.

Jahn, R., Zetzsche, H., Reinhardt, R., & Gemeinholzer, B. (2007). Diatoms and DNA barcoding: A pilot study on an environmental sample. In W. H. Kusber & R. Jahn (Eds.), *Proceedings of the 1st Central European Diatom Meeting 2007* (pp. 63–68). Berlin: Botanic Garden and Botanical Museum Berlin-Dahlem.

Jensen, K. G., Moestrup, Ø., & Schmid, A. M. M. (2003). Ultrastructure of the male gametes from two centric diatoms, *Chaetoceros laciniosus* and *Coscinodiscus wailesii* (Bacillariophyceae). *Phycologia, 42*, 98–105.

Kane, R. A., Stothard, J. R., Emery, A. M., & Rollinson, D. (2008). Molecular characterization of freshwater snails in the genus *Bulinus*: a role for barcodes? *Parasites & Vectors*, *1*(15). doi:10.1186/1756-3305-1-15.

Kelly, L. J., Ameka, G. K., & Chase, M. W. (2010). DNA barcoding of African Podostemaceae (river-weeds): A test of proposed barcode regions. *Taxon, 10*, 251–260.

Kobayasi, H., Kobayashi, H., & Idei, M. (1985). Fine structure and taxonomy of the small and tiny *Stephanodiscus* (Bacillariophy-

ceae) species in Japan. 3. Co-occurrence of *Stephanodiscus minutulus* (Kütz.) Round and *S. parvus* Stoerm. & Håk. *Japanese Journal of Phycology, 33*, 293–300.

Kooistra, W. H. C. F., & Medlin, L. K. (1996). Evolution of the diatoms (Bacillariophyta): IV. A reconstruction of their age from small subunit rRNA coding regions and the fossil record. *Molecular Phylogenetics and Evolution, 6*, 391–407.

Kress, W. J., & Erickson, D. L. (2007). A two-locus global DNA barcode for landplants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS Biology, 2*, e508. doi:10.1371/journal.pone.0000508.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., & Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the USA, 102*, 8369–8374.

Kucera, H., & Saunders, G. W. (2008). Assigning morphological variance of *Fucus* (Fucales, Phaeophyceae) in Canadian waters to recognized species using DNA barcoding. *Botany, 86*, 1065–1079.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*, 2947–2948.

Lefèvre, E., Bardot, C., Noël, C., Carrias, J., Viscogliosi, C., Amblard, C., et al. (2007). Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: Evidence from a molecular diversity study in a deep meromictic lake. *Environmental Microbiology, 9*, 61–71.

Levialdi Ghiron, J. H., Amato, A., Montresor, M., & Kooistra, W. H. C. F. (2008). Plastid inheritance in the planctonic raphid pennate *Pseudo-nitzschia delicatissima* (Bacillariophyceae). *Protist, 159*, 91–98.

Lewis, L. A., & Flechtner, V. R. (2004). Cryptic species of *Scenedesmus* (Chlorophyta) from desert soil communities of western North America. *Journal of Phycology, 40*, 1127–1137.

Liao, P. C., Huang, B. H., & Huang, S. (2007). Microbial community composition of the Danshui River estuary of northern Taiwan and the practicality of the phylogenetic method in microbial barcoding. *Microbial Ecology, 54*, 497–507.

Litaker, R. W., Vandersea, M. W., Kibler, S. R., Reece, K. S., Stokes, N. A., Lutzoni, F. M., et al. (2007). Recognizing dinoflagellate species using ITS rDNA sequences. *Journal of Phycology, 43*, 344–355.

Lynn, D. H., & Strüder-Kypke, M. C. (2006). Species of *Tetrahymena* identical by small subunit rRNA gene sequences are discriminated by mitochondrial cytochrome c oxidase I gene sequences. *Journal of Eukaryotic Microbiology, 53*, 385–387.

Mann, D. G. (1999). The species concept in diatoms. *Phycologia, 38*, 437–495.

McManus, H. A., & Lewis, L. A. (2005). Molecular phylogenetics, morphological variation, and colony-form evolution in the family Hydrodictyaceae (Sphaeropleales, Chlorophyta). *Phycologia, 44*, 582–595.

Medlin, L. K., & Kaczmarska, I. (2004). Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision. *Phycologia, 43*, 245–270.

Medlin, L. K., Elwood, H. J., Stickel, S., & Sogin, M. L. (1991). Morphological and genetic variation within the diatom *Skeletonema costatum* (Bacillariophyta): Evidence for a new species, *Skeletonema pseudocostatum*. *Journal of Phycology, 27*, 514–524.

Medlin, L. K., Kooistra, W. H., Gersonde, R., & Wellbrock, U. (1996). Evolution of the diatoms (Bacillariophyta). II. Nuclear-encoded small subunit rRNA sequence comparisons confirm a paraphyletic origin for the centric diatoms. *Molecular Biology and Evolution, 13*, 67–75.

Messing, J. (1983). New M13 vectors for cloning. *Methods in Enzymology, 101*, 20–78.

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology, 3*, e422. doi:10.1371/journal.pbio.0030422.

Moniz, M. B. J., & Kaczmarska, I. (2009). Barcoding diatoms: Is there a good marker? *Molecular Ecology Resources, 9*, 65–74.

Moniz, M. B. J., & Kaczmarska, I. (2010). Barcoding of diatoms: Nuclear encoded ITS revisited. *Protist, 161*, 7–34.

Morales, E. A., Siver, P. A., & Trainor, F. R. (2001). Identification of diatoms (Bacillariophyceae) during ecological assessments: Comparison between light microscopy and scanning electron microscopy techniques. *Proceedings of the Academy of Natural Sciences of Philadelphia, 151*, 95–103.

Moritz, C., & Cicero, C. (2004). DNA barcoding: Promise and pitfalls. *PLoS Biology, 2*, 1529–1531.

Müller, K. (2005). SeqState – primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics, 4*, 65–69.

Müller, T., Philippi, N., Dandekar, T., Schultz, J., & Wolf, M. (2007). Distinguishing species. *RNA, 13*, 1469–1472.

Nelles, L., Fang, B. L., Volckaert, G., Vandenberghe, A., & De Wachter, R. (1984). Nucleotide sequence of a crustacean 18S ribosomal RNA gene and secondary structure of eukaryotic small subunit ribosomal RNAs. *Nucleic Acid Research, 12*, 8749–8768.

Newmaster, S. G., Fazekas, A., Steeves, R., & Janovec, J. (2008). Testing candidate plant barcode regions in the Myristicaceae. *Molecular Ecology Resources, 8*, 480–490.

Nickrent, D. L., & Sargent, M. L. (1991). An overview of the secondary structure of the V4 region of eukaryotic small-subunit ribosomal RNA. *Nucleic Acid Research, 19*, 227–235.

Poulíčková, A., Špačková, J., Kelly, M. G., Duchoslav, M., & Mann, D. G. (2008). Ecological variation within *Sellaphora* species complexes (Bacillariophyceae): Specialists or generalists? *Hydrobiologia, 614*, 373–386.

Powers, T. (2004). Nematode molecular diagnostics: From bands to barcodes. *Annual Review of Phytopathology, 42*, 367–38.

R Development Core Team. (2005). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The barcode of life data system. *Molecular Ecology Notes, 7*, 355–364.

Ravin, N. V., Galachyants, Y. P., Merdanov, A. V., Beletsky, A. V., Petrova, D. P., Sherbakova, T. A., et al. (2010). Complete sequence of the mitochondrial genome of a diatom alga *Synedra acus* and comparative analysis of diatom mitochondrial genomes. *Current Genetics, 56*, 215–223.

Robba, L., Russell, S. J., Barker, G. L., & Brodie, J. (2006). Assessing the use of the mitochondrial *cox*1 marker for use in DNA barcoding of red algae (Rhodophyta). *American Journal of Botany, 93*, 1101–1108.

Romari, K., & Vaulot, D. (2004). Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnology and Oceanography, 49*, 784–798.

Round, F. E., Crawford, R. M., & Mann, D. G. (1990). *The diatoms – biology and morphology of the genera*. Cambridge: Cambridge University Press.

Sarno, D., Kooistra, W. H. C. F., Medlin, L. K., 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. *Journal of Phycology, 41*, 151–176.

Saunders, G. W. (2005). Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future application. *Philosophical Transactions of the Royal Society of London, Biological Sciences, 360*, 1879–1888.

Saunders, G. W. (2008). A DNA barcode examination of the red algal family Dumontiaceae in Canadian waters reveals substantial

cryptic species diversity. 1. The foliose *Dilsea-Neodilsea* complex and *Weeksia*. *Botany, 86*, 773–789.

Savin, M. C., Martin, J. L., Giewat, M., & Rooney-Varga, J. (2004). Plankton diversity in the Bay of Fundy as measured by morphological and molecular methods. *Microbial Ecology, 48*, 51–65.

Schloss, P. D. (2010). The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Computational Biology, 6*, e1000844. doi:10.1371/journal.pcbi.1000844.

Scicluna, S. M., Tawari, B., & Clark, C. G. (2006). DNA barcoding of *Blastocystis*. *Protist, 157*, 77–85.

Seifert, K. A., Samson, R. A., de Waard, J. R., Houbraken, J., Lévesque, C. A., Moncalvo, J. M., et al. (2007). Prospects for fungus identification using COI DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Academy of Sciences of the USA, 104*, 3901–3906.

Smetacek, V. (1999). Diatoms and the carbon ocean cycle. *Protist, 150*, 25–32.

Sorhannus, U. (2007). A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. *Marine Micropaleontology, 65*, 1–12.

Spamer, E. E., & Theriot, E. C. (1997). “*Stephanodiscus minutulus*”, “*S. minutus*”, and similar epithets in taxonomic, ecological, and evolutionary studies of modern and fossil diatoms (Bacillariophyceae: Thalassiosiraceae)—A century and a half of uncertain taxonomy and nomenclatural hearsay. *Proceedings of the Academy of Natural Sciences of Philadelphia, 148*, 231–272.

Stevenson, R. J., & Pan, Y. (1999). Assessing ecological conditions in rivers and streams with diatoms. In E. P. Stoermer & J. P. Smol (Eds.), *The diatoms: Applications to the environmental and earth sciences* (pp. 11–40). Cambridge: Cambridge University Press.

Stoeckle, M. (2003). Taxonomy, DNA and the barcode of life. *Bioscience, 53*, 2–3.

Stoermer, E. P., & Smol, J. P. (1999). *The diatoms: Applications to the environmental and earth sciences*. Cambridge: Cambridge University Press.

Summerbell, R. C., Lévesque, C. A., Seifert, K. A., Bovers, M., Fell, J. W., Diaz, M. R., et al. (2005). Microcoding: The second step in DNA barcoding. *Philosophical Transactions of the Royal Society of London, Biological Sciences, 360*, 1897–1903.

Swofford, D. L. (2002). *PAUP*: Phylogenetic Analyses Using Parsimony (* and other methods). 4.0 beta*. Sunderland: Sinauer Associates.

Taylor, J., Bruns, T., & Lutzoni, F. (2008). *ITS as the fungal barcode*. http://www.allfungi.com/its-barcode.php. Accessed 30 December 2010.

Teubner, K. (1997). Merkmalsvariabilität bei planktischen Diatomeen in Berlin-Brandenburger Gewässern. *Nova Hedwigia, 65*, 233–250.

Utz, L. R., & Eizirik, E. (2007). Molecular phylogenies of subclass Peritrichia (Ciliophora: Oligohymenophorea) based on expanded analyses of 18S rRNA sequences. *Journal of Eukaryotic Microbiology, 54*, 303–305.

Vanelslander, B., Créach, V., Vanormelingen, P., Ernst, A., Chepurnov, V. A., Sahan, E., et al. (2009). Ecological differentiation between sympatric pseudocryptic species in the estuarine benthic diatom *Navicula phyllepta* (Bacillariophyceae). *Journal of Phycology, 45*, 1278–1289.

Ward, R., Zemlack, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. N. (2005). DNA barcoding Australia’s fish species. *Philosophical Transactions of the Royal Society of London, Biological Sciences, 360*, 1847–1857.

Wolf, M., Scheffler, W., & Nicklisch, A. (2002). *Stephanodiscus neoastraea* and *Stephanodiscus heterostylus* (Bacillariophyceae) are one and the same species. *Diatom Research, 17*, 445–451.

Wu, S. G., Wang, G. T., Xi, B. W., Gao, D., & Nie, P. (2008). Molecular characteristics of *Camallanus* spp. (Spirurida: Camallanidae) in fishes from China based on its rDNA sequences. *Journal of Parasitology, 94*, 731–736.

Xia, X. H., Xie, Z., & Kjer, K. M. (2003). 18S ribosomal RNA and tetrapod phylogeny. *Systematic Biology, 52*, 283–295.

Zechman, F. W., Zimmer, E. A., & Theriot, E. C. (1994). Use of ribosomal DNA internal transcribed spacers for phylogenetic studies in diatoms. *Journal of Phycology, 30*, 507–512.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research, 31*, 3406–3415.