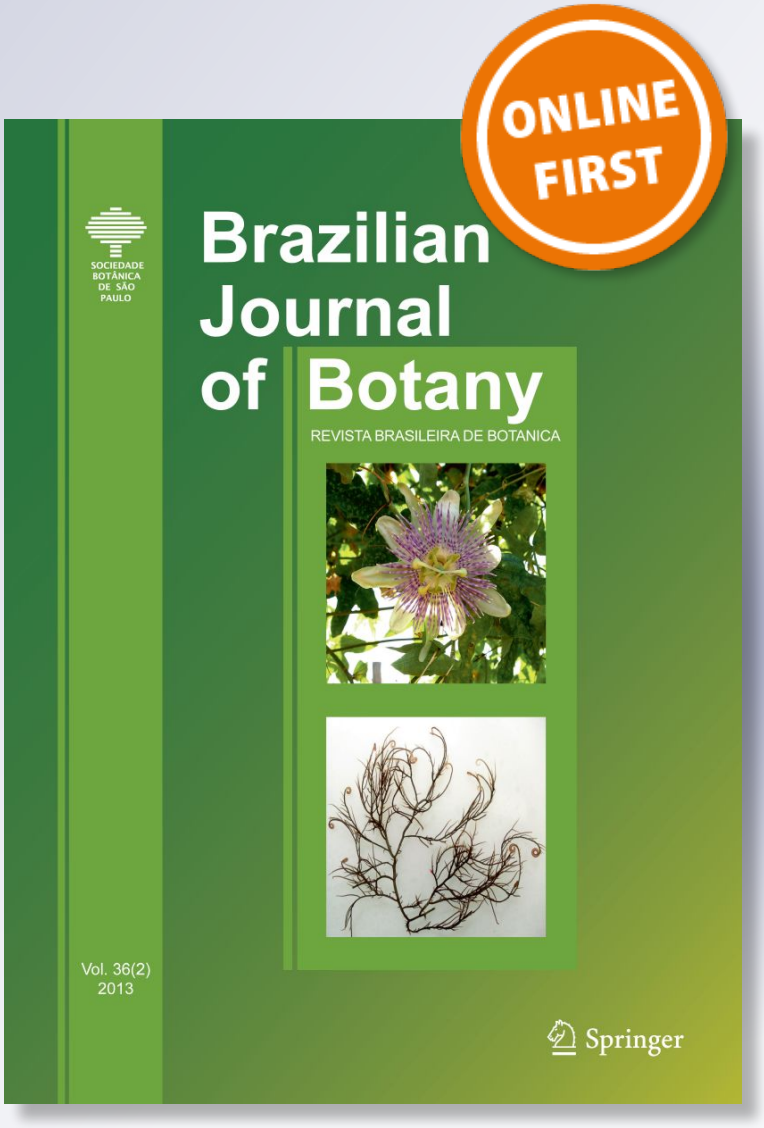
Mini review: Diatom species as seen through a molecular window

Linda K. Medlin

Brazilian Journal of Botany

ISSN 0100-8404

Braz. J. Bot
DOI 10.1007/s40415-018-0444-1



Springer

Brazilian Journal of Botany
https://doi.org/10.1007/s40415-018-0444-1

S.I. : USE OF MOLECULAR TOOLS IN THE SYSTEMATICS OF PLANTS AND FUNGI

Mini review: Diatom species as seen through a molecular window

Linda K. Medlin[1]

Received: 12 September 2017 / Accepted: 6 January 2018

Abstract
It has been accepted that we know less than 10% of the identified diversity in the marine microbial world and the diatoms are no exception. Even the species that we think we can easily recognize are often cryptic species, and even less is known of their life histories and spatial and temporal trends in their abundance and distribution. With new molecular and analytical techniques, we can advance our knowledge of a species to understand its morphological range, biogeographies and reproductive isolation. Moreover, some of molecular techniques are very sensitive. Depending on the species-level question(s) being asked, the molecular tools appropriate to answer them differ greatly.

Keywords Bacillariophyta · Barcoding · Molecular tools · Species concepts

1 Current status of the diatom species concept

One of the most controversial topics in biology is the recognition of species. Different definitions or concepts of species exist, such as the morphological, biological, phylogenetic and geological species concepts (Manhart and McCourt 1992; Gosling 1994; Mann 1999). Because of the differences in these definitions, species concepts can be very narrow or very broad. Several have been applied to the diatoms and even to the same group of diatoms. Because diatoms have an extensive fossil record and a well-defined morphology, and because extant taxa are relatively easily cultured, making it possible to study phenotypic plasticity and perform breeding experiments, it is sometimes easier to apply one concept over another to any particular group of diatoms.

The morphological species is the most widely used concept among all organisms except for prokaryotes. This is especially true for the diatoms whose intricate cell walls have been used for centuries to delimit its species (Williams 2007; Cox 2009). However, this concept as applied to diatoms is not foolproof because the size usually diminishes with each cell division and concomitantly as well as its shape and pattern (Medlin and Fryxell 1984; Goldman et al. 1990; Hasle et al. 1994), though there are examples of those diatoms that don't decrease in size (Rose and Cox 2013). Morphological changes are especially noticeable in the pennate diatoms, whereas in centrics they may be more subtle, e.g., as changes in the number of processes in 10 mm. Without knowing the full range of morphological changes, from the initial valves formed after auxosporulation to the smallest known cell, it can appear that more than one species is involved and more than once, at least two different species/genera have been described from opposite ends of the size range (Klee and Houk 1996; Rose and Cox 2014; Wetzel et al. 2015). It is also known that the same species can produce different diatom morphologies according to the seasons or habitat (see references in Cox 2014) and it has been suggested that the taxonomic level of *forma* should be used to reflect morphologies that change with particular environmental conditions (Cox 2014). The dynamics of cell size reduction and changes in shapes can be predicted and modeled (Woodard et al. 2016).

The biological species—separation of species according to their inability to interbreed—can be the most difficult to document because in many organisms, especially plants and unicellular algae, sexual reproduction is unknown. In the diatoms, we are only just beginning to understand life histories; the type of sexual reproduction has been most

✉ Linda K. Medlin
lkm@mba.ac.uk

[1] Marine Biological Association of the UK, The Citadel, Plymouth PL1 2PB, UK

Published online: 14 February 2018

commonly used in the diatoms to characterize higher taxa (e.g., Mizuno 2006, 2008; Kaczmarska and Ehrman 2015; Medlin 2016a) rather than examine species boundaries. Nevertheless, studies of 'model' genera and species, particularly *Pseudo-nitzschia* and *Sellaphora*, have investigated the degree to which morphological species are able to hybridize with other, closely related morphological species (Mann et al. 2004; Amato et al. 2007; Vanormelingen et al. 2013; Kaczmarska and Ehrman 2015).

The phylogenetic species concept uses monophyletic groupings to define species. These are clusters of individuals that are diagnosably distinct from other such clusters and should display a parental pattern of ancestry and descent (Cracraft 1989; Vanderlaan et al. 2013). An element of time is also introduced into the species concept by the phylogenetic species concept (Gosling 1994), because divergences of clades can be calibrated with the timing of certain events. In the diatoms, with their extensive fossil record, divergence times can be estimated with molecular clocks from other divergences that are known from the fossil record (Medlin 2008; Souffreau et al. 2011). The phylogenetic species in the diatoms can be assessed through cladistic analysis of their siliceous wall components (see Methods in Kitching et al. 1998 and examples in Kooistra et al. 2010; Edgar et al. 2015; Pennesi et al. 2016), through geometric morphometric analyses (Beszteri et al. 2005; Edgar et al. 2015; Urbánková et al. 2016) or through sequence analysis of one or more genes (Medlin 2016b; Theriot et al. 2015), or combinations of these methods (see each reference above). When known taxa arise within other known taxa making the parent taxa paraphyletic, the taxonomic decision to split the parent taxon into two separate entries can be problematic and difficult for many diatomists to reconcile (see Discussions in Medlin 2014) and this often results in the loss of well-known taxa (e.g., *Hemidiscus* Wallich, Gomez et al. 2017), when the acceptance of paraphyletic taxa as a natural course of evolution would eliminate such controversies (Medlin 2014). Pinseel et al. (2017a) described a new species within the *Pinnularia borealis* Ehrenberg complex. Within this complex, eight distinct lineages were recovered and only one of them was described as a new species, making *P. borealis*. Clearly, the taxonomy of the diatoms has the potential to become very confused if in one study, paraphyletic taxa are not accepted and well-known taxa are lost and in another, paraphyletic taxa are accepted and new taxa are described.

Although molecular data can enhance any of the other species concepts, it is most easily incorporated into the phylogenetic species concept (Vanderlaan et al. 2013). The coupling of molecular data with more traditional means of delineating species provides a powerful tool to help in our assessment of a species. Molecular data can (1) identify multi-species complexes (cryptic species) and help better define a species' limits, (2) provide an objective framework upon which to interpret the taxonomic level to which physiological and morphological differences can be applied, (3) interpret gene flow and dispersal mechanisms and (4) depict the phylogenetic history of a group and interpret its biogeographic distribution. The resolution of the species concept in *Skeletonema costatum* (sensu lato) (Greville) Cleve provided one of the earliest applications of molecular data in providing answers to each of the areas listed above (Medlin et al. 1991; Sarno et al. 2005; Zingone et al. 2005; Kooistra et al. 2008). Indeed, *Skeletonema pseudocostatum* Medl. (Medlin et al. 1991) was the first diatom species to be detected first with molecular data, and then morphological data were associated with its molecular signal to describe it as a new species.

Although many species may be molecularly distinct, they may be reproductively isolated (syngens) (see *Sellaphora auldreekie* D.G. Mann & S.M. McDonald—like strains = Vanormelingen et al. 2013) or may not be (see the coccolithophorid *Emiliania huxleyi* (Lohm.) Hay & Mohl. in Bendif et al. 2014, 2015). Often, conflicts may occur between phylogenetic/molecular speciation and morphological speciation (see examples from the coccolithophores in Young et al. 2014). When molecular sequences and morphology change at the different rates, this complicates our interpretation of how much morphological variation constitutes a difference between species. Often a morphological feature used to separate older interpretations of species can be found across many molecular clades (John et al. 2014), making the molecular data the only feature delineating the species. These so-called cryptic (sibling) species can be identified with molecular data (Beszteri et al. 2007; Amato and Montresor 2008; Nanjappa et al. 2014); they may be morphologically identical; they may or may not be reproductively isolated but are molecularly divergent from one another (Fig. 1). However, in Rovira et al. (2015) there were morphological differences among clones of *Nitzschia inconspicua* Grun., and they found no consistent differences among genotypes belonging to different clades, which are definable only through sequence data. Nevertheless, separating the genotypes could be important for ecological purposes because two different ecophysiological responses to salinity were encountered.

This implies that a fitness of form has been achieved for survival in the environment. This fitness of form, which results in similar if not identical morphotypes, has analogies at all taxonomic levels (Knowlton 1993; Sournia 1988). Fitness of form also suggests that most groups may be under-classified at the species level. Molecular data are just beginning to unravel the extent of cryptic species not

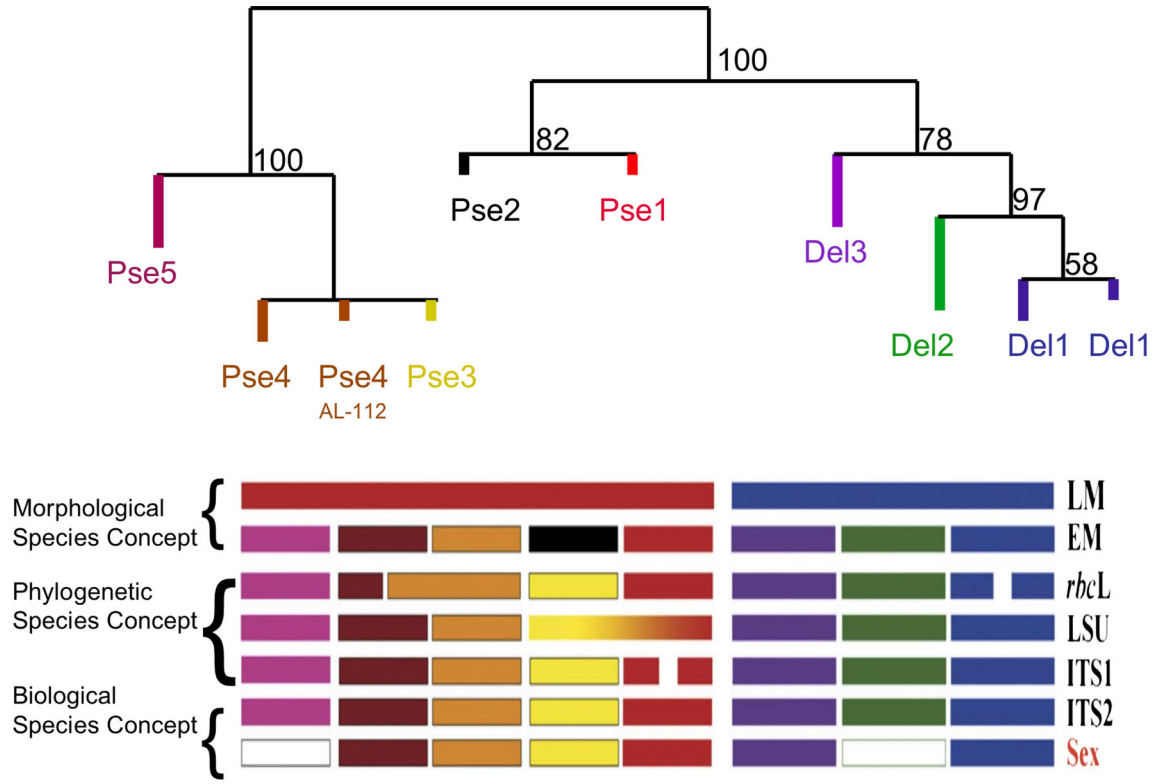


Fig. 1 Comparison of three species concepts in *Pseudo-nitzschia* illustrating the resolution of several species using a range of tools and genes available to diatomists. Pse 1 = *P. pseudodelicatissima*, Pse2 = *P. cuspidata*, Pse3 = *P. calliantha2*, Pse4 = *P. calliantha*, Pse5 = *P. caciantha,* Del1 = *P. delicatissima,* Del2 = *P. delicatissima2,* Del3 = *P. dolorosa*. Redrawn from Amato et al. (2007)

only in the diatoms but in many other groups, especially picoplankton, which have very few morphological markers (e.g., the prasinophyte *Ostreococcus* species, Subirana et al. 2013).

The region of the genome to be used in species-level questions must be carefully selected in order to provide answers to any questions being about species limitations. The interpretation of how much molecular divergence constitutes species-level divergences will depend on the (1) rate of evolution in the coding (gene) or non-coding (spacer) region selected for analysis (2) the geological age of taxa investigated and (3) the subjective perception we have as taxonomists as to how much morphological variation exists between species that are clearly separated molecularly, although the morphological variation can become more objective if cladistic analyses are applied (Kooistra et al. 2010; Edgar et al. 2015). At lower taxonomic levels, non-coding spacer regions may be more appropriate, because these are not under the same functional constraints as those of the coding regions and they are free to evolve at a faster rate to provide greater resolution among closely related species or even below the species at the population level.

Such conflicts between morphological and molecular data in interpreting species limits could be best resolved by breeding information, which, depending on the group, may or may not be available. Often the use of multiple genes can result in a consensus opinion as to how to interpret species limits from molecular data. Limited resolution between species with one genomic region may be amplified if faster evolving regions are used (Vanormelingen et al. 2013). By comparing variation in molecular markers with species-level markers from other concepts, one can interpret how much variation in molecular markers can be tolerated within a species.

DNA/DNA hybridization was a technique used in the latter part of the twentieth century to decide if two groups of organisms were the same species but it is rarely used today. However, values of less than 70% similarity were generally taken to indicate that separate eukaryotic species are involved (Huss et al. 1986, 1989; Amann et al. 1992). With such data to hand, one can compare differences in coding regions. Species (strains within a species) with 97% similarity in their DNA are usually separated by more than 4 nucleotides (99.8% similar) when coding regions (e.g., rRNA genes) have been compared (Amann et al. 1992; Huss personal communication). Once rapid sequencing methods and PCR became available, this technique became obsolete because for each new species added to the data set, all hybridizations had to be redone, each requiring 1–2

mgs of DNA. Today, with whole genomes being sequenced, this comparison can be done in silico to demonstrate the divergence among genomes (Basu et al. 2017).

The amount of nucleotide substitution in coding regions among clones known to belong to a species complex or a group of cryptic species will vary depending on the evolutionary age of the species and the rate of evolution in the gene/lineage selected. For example, in the coccolithophorids, *E. huxleyi* cannot be separated from *Gephyrocapsa* using 4 different genes, although the two are known to have separated ca 150,000 years ago and are morphologically distinct. Of commonly used species-level molecular markers, only the tufA gene can separate these taxa (Medlin et al. 1996) and interbreeding can still occur between the two (Bendif et al. 2014, 2015).

If slow-evolving coding regions are used, such as the 18S rRNA gene, and more than 4 bp differences are found, then the possibility of a species complex (cryptic or sibling species) should be investigated (Medlin et al. 1991, 1994). If faster evolving coding or non-coding regions are used (see examples from *E. huxleyi* below), then it is important to separate species-level variation from population-level variation among individuals (Coleman and Mai 1997; Beszteri et al. 2007; Ruggerior et al. 2015; Balzano et al. 2017). If breeding information is known for a group of cryptic species, then firmer conclusions can be drawn between the molecular differences and species limits (Mann et al. 1999, 2010; Amato et al. 2007; Vanormelingen et al. 2008).

If the reproductive isolation of the cryptic species is too recent in geological time, then the cryptic species will not have coexisted long enough to accumulate nucleotide substitutions in either coding or non-coding regions of the genome and therefore will have identical molecular sequence data (Medlin et al. 1996). Some recently evolved species may have arisen as a result of selection for functional morphological traits but there is no divergence in the molecular signal. Such recent speciations can probably only be detected using population-level methods, such as microsatellites (Rynearson and Armbrust 2005; Evans et al. 2007; Vanormelingen et al. 2007; Rynearson et al. 2009; Whitaker and Rynearson 2017) and perhaps measurements of genome size (Vaulot et al. 1994). Please refer to Medlin and Kooistra (2010) for detailed summaries of population-level studies in microalgae.

In the absence of breeding data, construction of molecular secondary structures in rRNA has been used to help interpret the significance of base substitutions found in coding or non-coding regions of potential cryptic species (Fig. 1). Compensatory base changes across an rRNA helix cannot be interpreted as microheterogeneity within a gene or even random sequencing or amplification errors and must be given more weight in interpreting species limits than base changes in regions of the molecule that do not form a double helix (see Rappé et al. 1995; Coleman and Mai 1997; Coleman 2000; Engesmo et al. 2016). With breeding data, the interpretation of compensatory base changes as a means of delimiting species using molecular data is greatly strengthened. In the green microalgae *Chlamydomonas* and *Gonium*, compensatory base changes across helices in the non-coding regions within the ribosomal cistron correlate with boundaries between syngens that cannot hybridize (Coleman and Mai 1997; Coleman 2000). New species of the raphidophyte *Heterosigma* were delimited in part by compensatory changes across a helix (Engesmo et al. 2016).

These molecular data can be used to predict mating success and define a species' limits. This method has been applied in several cases to determine if new species are involved. *Chaetoceros neogracilis* S.L.VanLandingham has identical 18S rRNA genes but can be separated into 4 clades with two other faster markers. Secondary structure analyses of the ITS regions showed with hemi-compensatory base changes across a helix, which suggests that the four clades cannot interbreed and should be regarded as separate species, which have recently speciated (Balzano et al. 2017). Secondary structure analyses to predict species limits have been used in other diatom studies (Medlin 1997; Behnke et al. 2004).

2 Selected case studies

Most diatom work utilizing type material is based on light and electron microscopy but in a few instances the type material or isotype material has been examined molecularly. In the Abarca et al. (2014) study of the highly morphologically variable *Gomphonema parvulum* (Kütz.) Kütz., the type species was defined with an epitype culture, varieties were raised to species level, and new species were added, all based first on molecular data, which were then supplemented by morphological data. Isolates identified as belonging to the *P. lanceolatum* (Bréb. ex Kütz.) Lange-Bert./*P. frequentissimum* (Lange-Bert.) Lange-Bert., species complex, were divided into 5 clades, with new species being described, varieties raised to species, rarely encountered species being recognized as more widespread. (Jahn et al. 2017). The value of a sinus versus a cavum was substantiated and divided the group naturally into two clades.

Skeletonema costatum (sensu lato) is a classic example of how one cosmopolitan polymorphic taxon was divided into five taxa following molecular analyses, each with a different morphology once the molecular clades were defined (Sarno et al. 2005). Type material was

reinvestigated and found to contain two taxa but based on the original drawings it was possible to establish which taxon was the type species (Zingone et al. 2005).

Vanormelingen et al. (2008) found *Eunotia bilunaris* (Her.) Schaars. populations, belonging to three clades, had different valve morphologies. The predominant mode of auxosporulation was heterothallic sexual reproduction; however, apomixis was also found. Interbreeding between the clades was possible but most F1 offspring were sterile. Hybrid sterility had evolved in two sympatric species with an ITS sequence divergence of 11.5–12.3%.

Pinseel et al. (2017b) investigated *Achnanthidium minutissimum* (Kütz.) Czarn. and found 12 lineages using a three-gene phylogeny. One new species was described, and they suggested that more molecular data were needed to solve the taxonomic problems associated with this group, which would eventually result in a better understanding of the biogeography and niche differentiation of different species belonging to the *A. minutissimum* complex.

More recently, automated species detection based on coalescence theory using two genes revealed that over 70 lineages (species??) could be recovered in the *P. borealis* complex and it was suggested that the true diversity of that species could be several hundreds of lineages, some cosmopolitan, others with a more restricted distribution (Pinseel et al. 2017c).

Cyclotella comensis Grun., *C. pseudocomensis* W. Schef. and *C. costei* Druart & Straub were investigated using 4 genes (Kistenich et al. 2014). They are indistinguishable using morphological and DNA sequence data, which suggest that they are one and the same taxon, although they had originally been delineate by differences in their central area. Certainly in *Cyclotella meneghiniana* Kütz., nearly identical populations were separated by geometric morphometrics (Beszteri et al. 2005). Population-level markers may shed more light on any obviously very young species complex.

Evans et al. (2008, 2009), using microsatellites and ITS sequence data in global isolates of *Sellaphora capitata* (Kütz.) Meres., found that the former could recover population structure but the latter could not. Other studies in diatoms [*Eunotia bilunaris* (Vanormelingen et al. 2007, 2015), the marine *Pseudo-nitzschia pungens* (Grun. ex Cleve) Hasle (Casteleyn et al. 2008) and *Ditylum brightwellii* (T.West) Grun. (Rynearson et al. 2009)] using these same two molecular tools, could recover population structure. Mating studies indicated that gene flow across large geographic areas was still possible.

Trobajo et al. (2009, 2010) analyzed *Nitzschia palea* (Kütz.) W. Smith with cox 1, LSU, *rbc*L and recovered at least four lineages. Intra-specific differences between these lineages were 0–0.8%, and inter-specific differences were 2.9–3.7%. Cox 1 was the most variable but could not be amplified from all species, undermining its usefulness as a gene to separate diatom taxa at the species level. Rimet et al. (2014) expanded these studies to include more global isolates and found slight increases in genetic distance with increasing geographical distance but could not find an objective criterion to choose a precise molecular threshold for marking species boundaries. If the cryptic species groups differed physiologically, then their recognition as a different species could improve monitoring. Certainly, Whitaker and Rynearson (2017) found that environmental and ecological selection likely exerted a stronger influence than dispersal on the divergence and connectivity of planktonic populations of *Thalassiosira rotula* (Meun.).

3 Barcoding

DNA barcoding is a technique that uses a short standardized DNA region to identify species (Hebert et al. 2003). For this approach to work, the "barcoding community" needs to agree on the gene fragment to use so that barcodes from different species are comparable. A fragment of the mitochondrial COI gene (cytochrome oxidase I) is most often used for DNA barcoding, especially in animals, but it is not sure yet that this is the best choice for a range of organisms, including phytoplankton (see review of potential barcodes in diatoms in Mann et al. 2010 and Moniz and Kaczmarska 2009, 2010). There is no general agreement on which region to use for protists. Pawlowski et al. (2012) have recommended a two-step barcoding, a first screen with the V4 region of the 18S rDNA and then a second screen with a variety of other genes (Fig. 2). Other genes, i.e., *rbc*L, might give a better resolution and identification in certain groups but *rbc*L is only present in photoautotroph organisms. It might well be in the end, that DNA barcodes have to be developed using several different genes. Another possible disadvantage is the short length of the barcoding sequence. Whereas this is deliberate to make development and analysis easier, it has the drawback that the information content of the sequence is limited and it might not be possible to distinguish between species based on these short sequences. Sequence reads are becoming longer as the technology develops (Table 1). Moniz and Kaczmarska (2009, 2010) found that 5.8S + ITS-2 provided the best resolution coupled with PCR amplification success rate and recommended this region as a diatom barcoding region.

Reference libraries of DNA barcodes are being developed for animals, plants and fungi. The Protist Working Group (ProWG), initiated by the Consortium for the Barcode of Life (CBOL, http://www.barcodeoflife.org/) has the tasks of identifying the standard barcode regions for protists and assembling a reference library. Reference libraries for diatoms have been established (Zimmermann

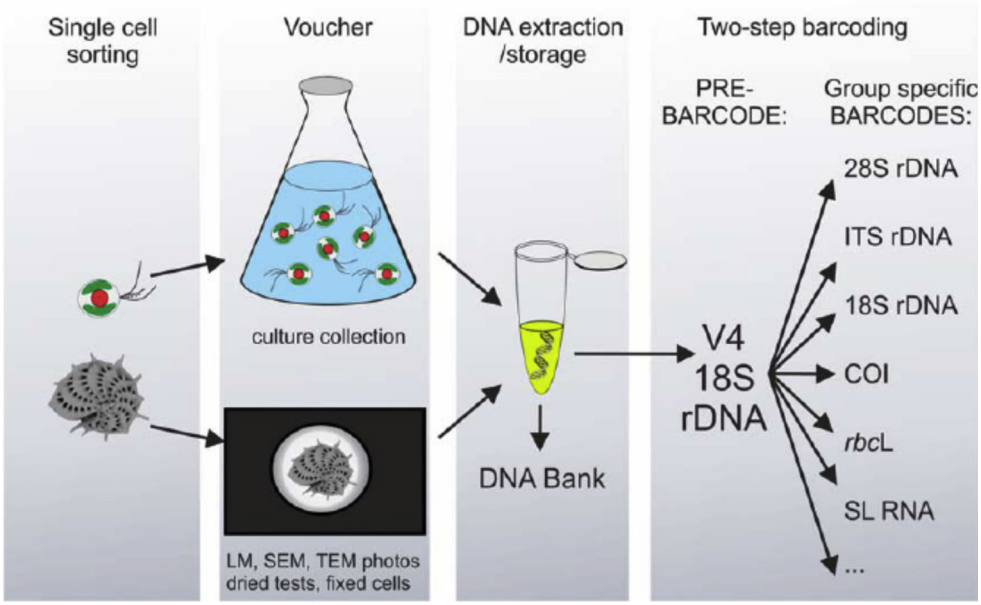


Fig. 2 Two-step protist barcoding pipeline. Protistan species, spanning four orders of cell size magnitude from < 1 to 10 mm are individually sorted from the environment, phenotyped directly or after culturing, DNA extracted and barcodes using a two-step, nested strategy. Reproduced from Pawlowski et al. (2012), open access

Table 1 Comparison of different sequencing platforms capabilties based on optimal conditions. Taken from Garrido-Cardenas et al. 2017)

Platform	Read length (bp)	Accuracy (%)	Run time	Bases per run (Gb)	Cost/Gb
454 Roche	1000	99	24 h	0.54	$10,000
SOLiD	75	99.9	7 d	520	$10
Illumina	300	99.9	3 d	1800	$10
Ion torrent	400	99	2 h	15	$100
Pacific bioscience	20,000	90	3 h	12,000	$600
Oxford nanopore	10,000	90	2 d	42	$1000

bp base pairs, *Gb* gigabase pairs, *h* hours, *d* days

et al. 2014b). According to Hamsher et al. (2011), there is currently no perfect barcode marker for both pennate and centric diatom species but see comments by Moniz and Kaczmarska. Zimmermann et al. (2011) suggested that the DNA barcoding of diatoms promises great potential to resolve the problem of inaccurate species identification and thus facilitate analyses of the biodiversity of environmental samples and have reviewed the markers and their resolution, citing advantages and disadvantages of each.

With the barcode markers used in Kistenich et al. (2014), there was very little sequence variation in 18S V4, LSU D2/D3, *rbc*L and cox 1 fragments, indicating that all strains of *Cyclotella comensis* (Grun.) studied could belong to the same species. Piredda et al. (2017) performed HTS metabarcoding of the V4 region of the 18S rDNA extracted from a surface sediment sample and simultaneously performed serial dilution cultures and were able to recover nearly all NGS sequences in their cultures, but as would be expected more ribotypes (82) were recovered in the molecular analysis than number of taxa (30) recovered by culturing. This pilot study also illustrates the resolution capability of the V4 barcode region in different genera. In some cases, the V4 region discriminates between species that are indistinguishable in light microscopy, i.e., *Leptocylindrus danicus* Cl. and L. *hargravesii* Nanj. & Zing. (Nanjappa et al. 2013). In other cases, this marker region cannot distinguish between morphologically and phylogenetically distinct species, e.g., *Skeletonema pseudocostatum* Medl. and *S. tropicum* Cl. In particular, these two species, among others in the genus, differ in other regions of the 18S or 28S rDNA region, which is used in phylogenetic analyses (Sarno et al. 2005), but share the same V4 barcode region (Luddington et al. 2012).

4 Metabarcoding

DNA barcoding can be a powerful tool in taxonomy and in analyzing biodiversity in environmental samples (metabarcoding). The high-throughput approach and the comparability of data will help to address many questions about cryptic and invasive species and will help to identify quickly microbial diversity in any given water sample. As with all methods, the limits of this technique must be

	no. taxa/site+method															
	RL1.1		RL2.1		RL3.1		RL4.1		RL5.1		RL6.1		RL7.1		RL1.1-RL7.1	
Genus	LM	NGS	LM	NGS	LM	NGS	LM	NGS	LM	NGS	LM	NGS	LM	NGS	LM	NGS
Achnanthidium	2	8	1	9	1	1	1	1	2	3	1	1	1	2	4	14
Amphora	-	-	1	3	1	1	-	-	1	2	2	6	2	3	2	8
Asterionella	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1
Caloneis	-	-	-	-	-	-	-	-	-	-	1	1	-	-	1	1
Cocconeis	-	-	1	2	1	1	1	3	1	3	1	7	2	2	2	15
Craticula	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	2
Cymatopleura	-	-	-	-	-	-	-	-	-	1	-	-	1	5	1	5
Cymbella	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	2
Diadesmis	1	1	-	-	-	-	-	-	-	-	-	-	-	-	1	1
Diatoma	1	1	1	1	1	1	1	1	4	3	2	3	3	4	6	11
Encyonema	1	-	3	8	3	3	2	3	2	4	-	1	1	2	3	7
Eolimna	1	-	2	7	1	1	1	1	1	1	1	2	1	1	2	7
Eunotia	1	6	-	-	-	-	-	-	-	-	-	-	-	-	1	5
Fragilaria	5	7	3	6	5	3	5	4	3	4	4	9	5	5	18	29
Frustulia	-	-	1	1	1	-	-	-	-	-	-	-	-	-	2	1
Geissleria	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
Gomphonema	2	1	4	27	2	4	2	6	6	14	2	12	3	6	10	28
Hantzschia	-	-	1	1	-	1	-	-	-	-	-	-	-	-	1	2
Hippodonta	-	-	1	1	1	1	1	2	1	1	-	-	1	-	2	2
Karayevia	-	-	-	-	1	1	-	-	-	-	1	-	1	-	2	1
Luticola	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-
Mayamaea	-	-	-	-	-	-	-	-	1	1	1	1	-	-	1	2
Meridion	2	-	1	-	-	-	1	-	1	-	-	-	2	-	2	-
Navicula	6	19	4	11	5	12	6	17	6	24	5	14	5	8	9	40
Nitzschia	3	5	6	8	7	24	6	18	5	13	5	17	5	14	17	49
Parlibellus	-	-	-	-	-	-	-	-	1	-	-	-	1	-	1	-
Pinnularia	1	1	-	-	-	-	2	4	-	-	-	-	-	-	3	5
Planothidium	1	2	2	3	2	2	1	1	1	2	1	1	1	2	2	9
Reimeria	-	-	1	2	1	1	1	1	1	1	1	1	1	1	1	7
Rhoicosphenia	-	-	1	-	1	-	1	-	1	-	1	-	1	-	1	-
Sellaphora	1	1	-	-	1	-	1	1	1	1	1	1	1	2	2	6
Stauroneis	1	-	-	-	-	-	1	-	-	-	-	-	-	-	1	-
Surirella	2	2	1	3	1	2	2	2	2	3	-	-	2	3	3	8
Tryblionella	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	2

Fig. 3 Heat map and numbers of taxa detected in the specific genera for NGS and LM methods at the sampling sites long the German/Polish border. Color code ranges from light to darker shades for the number of taxa at each site. Reproduced from Zimmermann et al. (2014b), with permission. (Color figure online)

assessed not to overestimate what DNA barcodes can really tell us.

To obtain a semi-quantitative overview of the diversity of an environmental sample, theoretically the barcodes of all possible organisms in the biosphere should first be known. We are unfortunately still very far removed from achieving that and it is questionable if it will ever be known, although the Tara ocean program data sets appear to be recovering a good estimate of the total OTUs at the genus level in the marine environment (Malviya et al. 2017). It cannot be assumed that the barcode of a single individual of a species is representative of the species because different individuals in a population and individuals in different geographic populations may possess slightly different barcode sequences. The extent of intraspecific variation is critical and should remain far less than differences among species.

The application of DNA barcodes for real-time monitoring of environmental samples has been achieved only in a few cases. In the Tara Ocean samples, the barcode used was the V9 region of the rDNA (Malviya et al. 2017), which appears to discriminate diatoms best at the genus level. This region has been used for primarily in bacteria, which is why it was used for these samples, which covered the entire marine community. For the diatoms, NGS sequencing and linking of the OTUs to water quality have been achieved by Zimmermann et al. (2014a), Kermarrec et al. (2013, 2014) and Apothéloz-Perret-Gentil et al. (2017) using the V4 region of the rDNA. Zimmermann et al. made a direct comparison numbers of taxa derived by NGS (263) and light microscopy (102) and showed that the NGS eDNA barcoding approach is qualitatively at least on the same level as light microscopy, but often provides a greater taxonomic depth than the classical morphology-based approach (Fig. 3). Problem taxa, such as *Nitzschia*, were easily identified, whereas NGS revealed more taxa than LM. Ruggerior et al. (2015) used DNA barcoding from clone libraries to trace the distribution of cryptic and toxigenic *Pseudo-nitzschia* species and the diversity of this key diatom genus in the natural environment.

Apothéloz-Perret-Gentil et al. (2017) attempted a taxonomic free approach to calculate a molecular index from

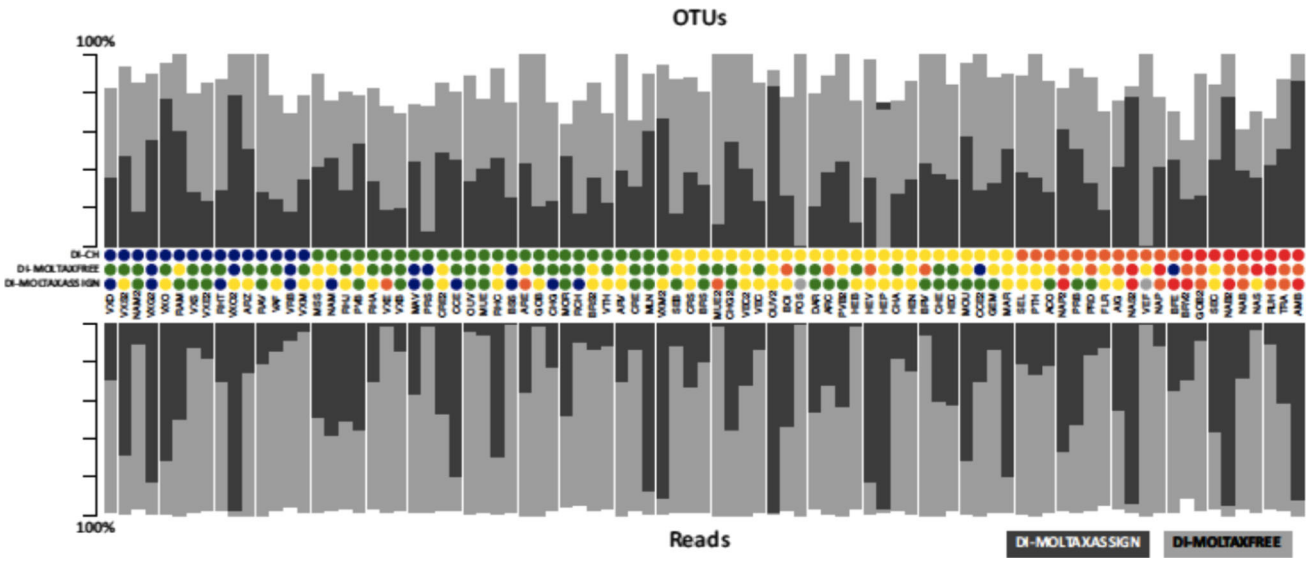


Fig. 4 Percentage of the NGS data used by the taxonomic assignment by known OTUs (dark gray) and the molecular index from all OTUs (light gray) method for each site. The OTUs bars are positioned at the top and the bottom of the graph, whereas the colored dots represent the ecological status as determined by light microscopy and cell counts (DI-CH), as inferred from the total OTUs (DI-MOLTAXFREE) or the identifiable OTUs (DI-MOLASSIGN). Blue, green, yellow, orange and red dots represent increasing poor water quality, with blue being very good and red very bad. Taken from Apothéloz-Perret-Gentil et al. (2017) with permission. (Color figure online)

the NGS data without any reference to species identification. The index provided an accurate assessment of 72% of the sites examined because all of the OTUs could used to calculate the index. Microscopic counts were made at the same time and from the species present, the ecological status of the sampling site was established to provide the calibration of the molecular index. Because of the limited sequence data in public databases, only, 35% of the OTUs received a taxonomic assignment. The calibration step is the limiting factor, but once achieved, then the taxon-free index could be used in routine monitoring (Fig. 4). However, obtaining cultures of each OUT is a daunting task to achieve calibration. One of the limiting factors that these workers noted was the lack of resolution in the V4 region of the 18S rDNA for critical species, such as those in *Navicula* that are needed for correct water quality assessment. They recommended using additional markers to improve resolution in these problem areas. Thus their initial work should likely be regarded cautiously.

A very different approach was used in the EU µAQUA project, in which a microarray was made for the detection of freshwater pathogens and diatoms as indicators of water quality. For this application, the barcodes are immobilized on a glass slides (the microarray or so-called phylochip). To tal rRNA was extracted from the environmental samples, fluorescently labeled and hybridized to the microarray. This is a PCR-free method that is semi-quantitative in that the intensity of the hybridization signal is proportional to the abundance of the target organisms for that particular sample, as based on calibration curves for selected taxa. In this project, cultures of indicator diatom species critical for determining water quality were assembled from known literature based primarily on an index used in Italy for diatoms. A barcode was designed for each species on the list from their 18S rRNA gene and spotted onto the slides. The microarray was field-tested in 5 countries and in Table 2 are the results for the diatoms from two locations: the Mondony River, Station Amelie and Canet Lagoon in SE France. Canet Lagoon experienced a poorer water quality than Mondony River over the two-year sampling period, as estimated from the species detected by the phylochip.

Despite the fact that multiple data sets are emerging (V4 and V9 of the rDNA gene, *rbc*L), barcoding provides a powerful tool for obtaining semi-quantitative data on the species composition of large numbers of environmental samples in a rapid and cost-effective way. DNA can be extracted, the barcode marker of interest PCR-amplified and the obtained plethora of gene copies sequenced by means of modern massive sequencing technologies. The resulting sequence reads are characterized taxonomically by automated means: samples in, semi-quantitative list of species out, which can take up to 3 months to produce the results because of all of the controls. The phylochip, which uses total rRNA as its target, is enzyme free and less prone to natural inhibition and gives immediate results, however, only of known sequence/species. One attractive alternative to this limitation of the phylochip is to use NGS sequences, once they are annotated and potentially identified, respotted in a microarray format, so that information on these

Table 2 Summary of diatom signals from the µAQUA phylochip and an interpretation of the water quality at the sampling date, using an Italian scale and a French one associated with the taxon on the phylochip

Date	Diatoms in order of decreasing intensity of signal I is worst	Water quality Italy	Water quality France 7 is worst
Mondony River			
Jun 2011	Order Surirellales	?III	C2–C4
Jul 2011	Order Surirellales	?III	C2–C4
Sept 2011	Order Surirellales, *Surirella angusta* Kütz.	?III, III	C2–C4, C3–C4
Oct 2011	Order Surirellales, *Surirella angusta*	?III	C2–C4, C3–C4
Dec 2011	Order Surirellales (very low)	?III	C2–C4
Mar 2012	Order Surirellales, *Navicula veneta* Kütz	?III	C2–C4, C2
Apr 2012	Order Surirellales	?III	C2–C4
May 2012	Order Surirellales	?III	C2–C4
Canet Lagoon			
July 2011	*Achnanthidium minutissima, Surirella angusta, Nitzschia palea, Amphora coffeaeformis* (C. Ag.) Kütz.**	I–II, III, IV, **Toxin producer	C5, C3–C4, C1, C2
Aug 2011	*Achnanthidium minutissima, Navicula veneta*	I–II, IV–V	C5, C2
Sept 2011	*Achnanthidium minutissima, Surirella angusta, Navicula veneta*	I–II, III, IV–V	C5, C3–C4, C2
Oct 2011	*Achanthidium minutissima, Surirella angusta*	I–II, III	C5, C3–C4,
Nov 2011	Genus *Surirella* Turp.	?III	C2–C4
Dec 2011	*Surirella angusta*	III	C3–C4
Mar 2012	Genus *Surirella, Amphora coffeaeformis***	?III, ** toxin producer	C2–C4, C2
Apr 2012	*Achnanthidium minutissima, Surirella angusta, Navicula veneta*	I–II, III, IV–V	C5, C3–C4, C2
May 2012	Order Surirellales, *Achnanthidium minutissima*	?III, I–II	C2–C4, C5

? is added for probes above the species level

Species/taxa are listed in order of decreasing signal intensity

OTUs is more or less immediate, thus eliminating long turn around time for analysis of NGS sequence data.

Garrido-Cardenas et al. (2017) has reviewed the different platforms for NGS and the advantages and disadvantages of each. It can predicted that new technologies will continue to reduce the cost of each, while extending the amount of data generated with increasing confidence (Table 1).

Metabarcoding data will be best used to demonstrate changes in the total community over time. With climate change, more pronounced changes in temperature, paleoecological data sets allow conclusions regarding temperature effects on ecosystem function and biodiversity over evolutionary time scales (Hillebrand et al. 2010). It is predicted that species compositions will change more dramatically with increased temperatures over ecologically relevant time scales. Lohbeck et al. (2012) showed that *E. huxleyi* could adapt and mutate in response to increasing carbon dioxide concentrations in fewer than 2 years, or over ecologically relevant time scales. In contrast, in the Tara Ocean analysis of the total community, it appears that species associations, i.e., who you are associated with, were more important for the distribution of the diatoms than association by abiotic factors, which had a greater affect on the association of other protists (Malviya et al. 2017).

5 Concluding remarks and perspectives

From the above examples, it can be concluded that molecular sequences can be used at the species level in the diatoms and above. Polymorphisms in coding or non-coding regions usually indicate either a species complex or population-level variation that can be associated with biogeographic distribution of the species. Biogeographic clusters may be termed varieties or even ecotypes of a species. More information regarding species delineation and population-level variation can be gathered when data from other species-level measurements are also included. Given that sufficient care, such as sampling of inter-population variation, direct sequencing of PCR products, pooling of multiple clones for sequencing, construction of secondary structures, is used to produce sequence data free of methodological error for species-level enquiries, then these data can be reliably used to delineate species. The

combination of several species-level methods (both molecular and traditional) is recommended to resolve the issue of a species limit. The V4 region is often omitted from phylogenetic analyses even those where taxonomic decisions are made, e.g., Gomez et al. (2017), so it seems counterintuitive to recommend this as a barcoding region but eliminate from phylogenetic analyses using full-length 18S sequences. With molecular data finding its way into many traditional diatomist's bag of tools, it is ever more important to keep accurate records and voucher material. For the molecular age, this would include the DNA, a slide made of the culture harvested for molecular analysis as well as dried material. My DNA collection has been donated to the Botanischer Garten und Botanisches Museum Berlin-Dahlem, Freie Universität Berlin, Berlin, Germany, and aliquots can be obtained from them. When DNA becomes low, it is possible to perform a whole genome amplification to boost the limited supply (Medlin et al. 2008).

Acknowledgements Dr. D. G. Mann critically read my manuscript.

References

Abarca N, Jahn R, Zimmermann J, Enke N (2014) Does the cosmopolitan diatom *Gomphonema parvulum* (Kützing) Kützing have a biogeography? PLoS ONE 9:e86885. https://doi.org/10.1371/journal.pone.0086885

Amann R, Lin C, Key R, Montgomery L, Stahl DA (1992) Diversity among *Fibrobacter* isolates: towards a phylogenetic classification. Syst Appl Microbiol 15:23–31

Amato A, Montresor M (2008) Morphology, phylogeny, and sexual cycle of *Pseudo-nitzschia mannii* sp. nov. (Bacillariophyceae): a pseudo-cryptic species within the *P. pseudodelicatissima* complex. Phycologia 47:487–497

Amato A, Kooistra WHCF, Ghiron JHL, Mann DG, Proschold T, Montresor M (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. Protist 158:193–207

Apothéloz-Perret-Gentil L, Cordonier A, Straud F, Iseli J, Esling P, Pawlowski J (2017) Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Mol Ecol Res. https://doi.org/10.1111/1755-0998.12668

Balzano S, Percopo I, Sianao R, Gourvil P, Chanoine M, Marie D, Vaulot D, Sarno D (2017) Morphological and genetic diversity of Beaufort Sea diatoms with high contributions from the *Chaetoceros neogracilis* species complex. J Phycol 53:161–187

Basu S, Pati S, Mapleson D, Russo MT, Vitale L, Fevola C, Maumus F, Casotti R, Mock T, Caccamo M, Montresor M, Sanges R, Ferrante MI (2017) Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. New Phytol. https://doi.org/10.1111/nph.14557

Behnke A, Friedl T, Chepurnov V, Mann DG (2004) Reproductive compatibility and rDNA sequence analyses in the *Sellaphora pupula* species complex (Bacillariophyta). J Phycol 40:193–208

Bendif EM, Probert I, Diaz-Rosas F, van den Engh G, Young JR, von Dassow P (2014) Recent reticulate evolution in the ecologically dominant lineage of coccolithophores. Front Microbiol. https://doi.org/10.3389/fmicb.2016.00784

Bendif EM, Probert I, Young JR, von Dassow P (2015) Morphological and phylogenetic characterization of new *Gephyrocapsa* isolates suggests introgressive hybridization in the *Emiliania/Gephyrocapsa* complex (Haptophyta). Protist. https://doi.org/10.1016/j.protis.2015.05.003

Beszteri B, Acs E, Medlin LK (2005) Conventional and geometric morphometric studies of valve ultrastructural variation in two closely related *Cyclotella* species (Bacillariophyceae). Eur J Phycol 40:73–88

Beszteri B, John U, Medlin LK (2007) An assessment of cryptic genetic diversity within the *Cyclotella meneghiniana* species complex (Bacillariophyta) based on nuclear and plastid genes, and amplified fragment length polymorphisms. Eur J Phycol 42:47–60

Bowler C (2017) Tara oceans: eco-systems biology at planetary scale. Presentation at 7th applied phycology conference Nantes, June 2017

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 species? Harmful Algae 7:241–257

Coleman AW (2000) The significant of a coincidence between evolutionary landmarks in mating affinity and a DNA sequence. Protist 151:1–9

Coleman AW, Mai JC (1997) Ribosomal DNA ITS-1 and ITS-2 sequence comparisons as a tool for predicting genetic relatedness. J Mol Evol 44:258–271

Cox EJ (2009) What's in a name? Diatom classification should reflect systematic relationships. Acta Bot Croat 68:443–454

Cox EJ (2014) Diatom identification in the face of changing species concepts and evidence of phenotypic plasticity. J Micropalaeontol. https://doi.org/10.1144/jmpaleo2014-014

Cracraft J (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and process of differentiation. In: Otte D, Endler JA (eds) Speciation and its consequences. Sinauer Assoc, Sunderland, pp 28–59

Edgar RK, Saleh AI, Edgar SM (2015) A morphometric diagnosis using continuous characters of *Pinnunavis edkuensis,* sp. nov. (Bacillariophyta: Bacillariophyceae), a brackish-marine species from Egypt. Phytotaxa 212:1–56

Engesmo A, Eikrem W, Seoane S, Smith K, Edvardsen B, Hofgaard A, Tomas CR (2016) New insights into the morphology and phylogeny of *Heterosigma akashiwo* (Raphidophyceae), with the description of *Heterosigma minor* sp. nov. Phycologia 55:279–294

Evans KM, Wortley AH, Mann DG (2007) An assessment of potential diatom barcode genes (cox1 rbcL 18S, ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). Protist 158:349–361

Evans KM, Wortley AH, Simpson GE, Chepurnov VA, Mann DG (2008) A molecular systematic approach to explore diversity within the *Sellaphora pupula* species complex. J Phycol 44:215–231

Evans KM, Chepurnov VA, Sluiman HJ, Thomas SJ, Spears BM, Mann DG (2009) Highly differentiated populations of the freshwater diatom *Sellaphora capitata* suggest limited dispersal and opportunities for allopatric speciation. Protist 160:386–396

Garrido-Cardenas JA, Garcia-Maroto F, Alvares-Bermej JA, Manzanao-Agugliaro F (2017) DNA sequencing sensors: an overview. Sensors 17:588. https://doi.org/10.3390/s17030588

Goldman N, Paddock TBB, Shaw KM (1990) Quantitative analysis of shape variation in populations of *Surirella fastuosa*. Diatom Res 5:25–42

Gomez F, Wang L, Hernandez-Becerril DU, Lisunova Lopes RM, Lin S (2017) Molecular phylogeny suggests transfer of *Hemidiscus*

into *Actinocyclus* (Coscinodiscales, Coscinodiscophyceae). Diatom Res 32:21–28

Gosling EM (1994) Speciation and species concepts in the marine environment. In: Beaumont AR (ed) Genetics and evolution of aquatic organisms. Chapmann Hall, London, pp 1–15

Hamsher SE, Evans KM, Mann DG, Poulickova A, Saunders GW (2011) Barcoding diatoms: exploring alternatives to COI-5P. Protist 162:405–422

Hasle GR, Medlin LK, Syvertsen EE (1994) *Synedropsis* gen. nov. a genus of araphid diatoms associated with sea ice. Phycologia 33:248–270

Hebert PDN, Cywinska A, Ball LS, DeWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond Ser B 270:313–321

Hillebrand H, Soininen J, Snoeijs P (2010) Warming leads to higher species turnover in a coastal ecosystem. Glob Change Biol. https://doi.org/10.1111/j.1365-2486.2009.02045.x

Huss VAR, Don R, Grossman U, Kessler E (1986) Deoxyribonucleic acid reassociation in the taxonomy of the genus *Chlorella*. Arch Microbiol 145:329–333

Huss VAR, Huss G, Kessler E (1989) Deoxyribonucleic acid reassociation and interspecies relationships of the genus *Chlorella* Chlorophyceae. Plant Syst Evol 168:1–82

Jahn R, Abarca N, Gemeinholzer B, Mora D, Skibbe O, Kulikovskiy M, Gusev E, Kusber W-H, Zimmermann J (2017) *Planothidium lanceolatum* and *Planothidium frequentissimum* reinvestigated with molecular methods and morphology: four new species and the taxonomic importance of the sinus and cavum. Diatom Res 32:75–107. https://doi.org/10.1080/0269249X.2017.1312548

John U, Litaker RW, Montresor M, Murray S, Brosnahan B, Anderson DM (2014) Formal revision of the *Alexandrium tamarense* species complex (Dinophyceae) taxonomy: the introduction of five species with emphasis on molecular-based (rDNA) classification. Protist 165:779–804

Kaczmarska I, Ehrman JM (2015) Auxosporulation in *Paralia guyana* MacGillivary (Bacillariophyta) and possible new insights into the habit of the earliest diatoms. PLoS ONE 10:e0141150. https://doi.org/10.1371/journal.pone.0141150

Kermarrec L, Franc A, Rimet F et al (2013) Next generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms. Mol Ecol Res 13:607–619

Kermarrec L, Franc A, Rimet F et al (2014) A Next generation sequencing approach to river monitoring using benthic diatoms. Fresh Sci 33:349–363

Kistenich S, Dreßler M, Zimmermann J, Hübener T, Bastrop HR, Jahn R (2014) An investigation into the morphology and genetics of *Cyclotella comensis* and closely related taxa. Diatom Res. https://doi.org/10.1080/0269249X.2014.922125

Kitching IJ, Forey PL, Humphries CJ, Williams DM (1998) Cladistics: the theory and practice of parsimony analysis, 2nd edn. Oxford University Press, Oxford, 229 pp

Klee R, Houk V (1996) Morphology and ultrastructure of *Cyclotella woltereckii* Hustedt (Bacillariophyceae). Arch Protist 147:19–27

Knowlton N (1993) Sibling species in the sea. Ann Rev Ecol Syst 24:189–216

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

Kooistra WHCF, Sarno D, Hernandez-Becerril DU, Assmy P, Di Prisco C, Montresor M (2010) Comparative molecular and morphological phylogenetic analyses of taxa in the Chaetocerotaceae (Bacillariophyta). Phycologia 49:471–500

Lohbeck KT, Riebesell U, Thorsten BH, Reusch TBH (2012) Adaptive evolution of a key phytoplankton species to ocean acidification. Nat Geosci 5:346–351

Luddington IA, Kaczmarska I, Lovejoy C (2012) Distance and character-based evaluation of the v4 region of the 18 s rRNA gene for the identification of diatoms (Bacillariophyceae). PLoS ONE 7:e45664. https://doi.org/10.1371/journal.pone.0045664

Malviya S, Scalcob E, Audicc S, Vincenta F, Veluchamya A, Poulaind J, Winckerd P, Iudiconeb D, de Vargas C, Bittnera L, Zingone A, Bowler C (2017) Insights into global diatom distribution and diversity in the world's ocean. PNAS. https://doi.org/10.1073/pnas.1509523113

Manhart JR, McCourt RM (1992) Molecular data and species concepts in the algae. J Phycol 28:730–737

Mann DG (1999) The species concept in diatoms. Phycologia 38:437–495

Mann DG, Chepurnov VA, Droop SJM (1999) sexuality, incompatibility, size variation, and preferential polyandry in natural populations and clones of *Sellaphora pupula* (Bacillariophyceae). J Phycol 35:151–170

Mann DG, McDonald SM, Bayer MM, Droop SJM, Chepurnov VA, Loke RE, Ciobanu A, Du Buf JMH (2004) The *Sellaphora pupula* species complex (Bacillariophyceae): morphometric analysis, ultrastructure and mating data. Phycologia 43:459–482

Mann DG, Sato S, Trobajo R, Vanomelingen P, Souffreau C (2010) DNA barcoding for species identification and discovery in diatoms. Cryptogam Algol 31:557–577

Medlin LK (1997) Can molecular techniques help define species limits? Diatom 13:19–23

Medlin LK (2008) Molecular clocks and inferring evolutionary milestones and biogeography in the microalgae. In: Okada H, Mawatari SF, Suzuki N, Gautam P (eds) Proceedings of International Symposium "The Origin and Evolution of Natural diversity", 1–5 October 2007, Sapporo, pp 31–42

Medlin LK (2014) Evolution of the diatoms: VIII. Re-examination of the SSU-rRNA gene using multiple outgroups and a cladistic analysis of valve features. J Biodivers Biopros Dev 1:129. https://doi.org/10.4172/2376-0214.1000129

Medlin LK (2016a) Opinion: can coalescent models explain deep divergences in the diatoms and argue for the acceptance of paraphyletic taxa at all taxonomic hierarchies? Nova Hedwig. https://doi.org/10.1127/nova_hedwigia/2015/0295

Medlin LK (2016b) A timescale for diatom evolution based on four molecular markers: reassessment of ghost lineages and major steps defining diatom evolution. Vie et Milleu 65:219–238

Medlin LK, Fryxell GA (1984) Structure, life history and systematics of *Rhoicosphenia* (Bacillariophyta). IV. Changes in frustule morphology concomitant with size reduction in *Rhoicosphenia genuflexa* (Kütz.) Medlin. J Phycol 20:101–108

Medlin LK, Kooistra WHCF (2010) Methods to estimate the diversity in the marine photosynthetic protist community with illustrations from case studies: a review. Biol Divers Assess Mol Methods 2:973–1014

Medlin LK, Elwood HJ, Stickel S, Sogin ML (1991) Morphological and genetic variation within the diatom *Skeletonema costatum* Bacillariophyta: evidence for a new species *Skeletonema pseudocostatum*. J Phycol 27:514–524

Medlin LK, Lange M, Baumann MEM (1994) Genetic differentiation among three colony-forming species of *Phaeocystis*: further evidence for the phylogeny of the Prymnesiophyta. Phycologia 33:199–212

Medlin LK, Barker GLA, Campbell L, Green JC, Hayes PK, Marie D, Wrieden S, Vaulot D (1996) Genetic characterisation of *Emiliania huxleyi* (Haptophyta). J Mar Syst 9:13–31

Medlin LK, Sato S, Mann DG, Kooistra WCHF (2008) Molecular evidence confirms sister relationship of *Ardissonea, Climacosphenia*, and *Toxarium* within the bipolar centric diatoms (Bacillariophyta, Mediophyceae), and cladistic analyses confirm

that extremely elongated shape has arisen twice in the diatoms. J Phycol 44:1340–1348
Mizuno M (2006) Evolution of meiotic patterns of oogenesis and spermatogenesis in centric diatoms. Phycol Res 54:57–64
Mizuno M (2008) Evolution of centric diatoms inferred from patterns of oogenesis and spermatogenesis. Phycol Res 56:156–165
Moniz MBJ, Kaczmarska I (2009) Barcoding diatoms: is there a good marker? Mol Ecol Res 9:65–74
Moniz MBJ, Kaczmarska I (2010) Barcoding of diatoms: nuclear encoded ITS revisited. Protist 161:7–34
Nanjappa D, Kooistra WHCF, Zingone A (2013) A reappraisal of the genus *Leptocylindrus* (Bacillariophyta), with the addition of three species and the erection of *Tenuicylindrus* gen. nov. J Phycol 49:917–936
Nanjappa D, Audic S, Romac S, Kooistra WCHF, Zingone A (2014) Assessment of species diversity and distribution of an ancient diatom lineage using a DNA metabarcoding approach. PLoS ONE 9:e103810
Pawlowski J, Audic S, Adl S, Bass D, Belbahri L et al (2012) CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol 10:e1001419. https://doi.org/10.1371/journal.pbio.1001419
Pennesi C, Poulin M, Totti C (2016) Phylogenetic relationships and biogeography of the diatom genus *Mastogloia* (Bacillariophyceae): revision of the Section *Ellipticae* including the description of new taxa. Protist 167:148–173
Pinseel E, Hejdukova E, Vanormelingen P, Kopalova K, Vyverman W, Van de Vijver B (2017a) *Pinnularia catenaborealis* sp. nov. (Bacillariophyceae), a unique chain-forming diatom species from James Ross Island and Vega Island (Maritime Antarctica). Phycologia 56:94–107
Pinseel E, Vanormelingen P, Hamilton P, Vyverman W, Van de Vijver B, Kopalova K (2017b) Molecular and morphological characterization of the *Achnanthidium minutissimum* complex (Bacillariophyta) in Petuniabukta (Spitsbergen, High Arctic) including the description of *A. digitatum* sp. nov. Eur J Phycol. https://doi.org/10.1080/09670262.2017.1283540
Pinseel E, Vanormelingen P, Janssens S et al (2017c) Opening Pandora's box: a diatom species complex as a case study for the diversity and biogeography of terrestrial micro-eukaryotes. Abstracts from the 11th international phycological congress. Phycologia 56:146 (supplement 4)
Piredda R, Sarno D, Lange CB, Tomasino MP, Zingonne A, Montresor M (2017) Diatom resting stages in surface sediments: a pilot study comparing next generation sequencing and serial dilution cultures. Cryptogam Algol 38:31–46
Rappé MS, Kemp PF, Giovannoni SJ (1995) Chromophyte plastid 16S ribosomal RNA genes found in a clone library from Atlantic Ocean Seawater. J Phycol 31:979–988
Rimet F, Trobajo R, Mann DG, Kermarrec L, Franc A, Domaizon I, Bouchez A (2014) When is sampling complete? The effects of geographical range and marker choice on perceived diversity in *Nitzschia palea* (Bacillariophyta). Protist 165:245–259
Rose DL, Cox EJ (2013) Some diatom species do not show a gradual decrease in cell size as they reproduce. Fundam Appl Limnol 182:117–122
Rose DL, Cox EJ (2014) What constitutes *Gomphonema parvulum*? Long-term culture studies show that some varieties of *G. parvulum* belong with other *Gomphonema*. Plant Ecol Evol 147:366–373
Rovira L, Trobajo R, Sato S, Ibanez C, Mann DG (2015) Genetic and physiological diversity in the diatom *Nitzschia inconspicua*. J Euk Microbiol 62:815–832
Ruggerior MV, Barra L, Kooistra WHCF, Zingone A (2015) Diversity and temporal pattern of *Pseudo-nitzschia* species (Bacillariophyceae) through the molecular lens. Harmful Algae. https://doi.org/10.1016/j.hal.2014.12.001
Rynearson TA, Armbrust EV (2005) Maintenance of clonal diversity during a spring bloom of the centric diatom *Ditylum brightwellii*. Mol Ecol 14:1631–1640
Rynearson TA, Lin EO, Armbrust EV (2009) Metapopulation structure in the planktonic diatom *Ditylum brightwellii* (Bacillariophyceae). Protist 160:111–121
Sarno D, Kooistra WHCF, Medlin LK, Percopo I, Zingone A (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae): *Skeletonema costatum* (Bacillariophyceae) consists of several genetically and morphologically distinct species with the description of four new species. J Phycol 41:151–176
Souffreau C, Verbruggen H, Wolfe AP, Vanormelingen P, Siver PA, Cox EJ, Mann DG, Van de Vijver B, Sabbe S, Vyverman W (2011) A time-calibrated multi-gene phylogeny of the diatom genus *Pinnularia*. Mol Syst Evol 61:866–879
Sournia A (1988) *Phaeocystis* Prymnesiophyceae: how many species? Nova Hedwig 47:211–217
Subirana L, Péquin B, Michely S, Escand M-L, Meillanda J, Derelle E, Marine B, Piganeau G, Desdevises Y, Moreau H, Nigel H, Grimsley NH (2013) Morphology, genome plasticity, and phylogeny in the genus *Ostreococcus* reveal a cryptic species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). Protist 164:643–659
Theriot EC, Ashworth M, Nakov T, Ruck E, Jansen RK (2015) Dissecting signal and noise in diatom chloroplast protein encoding genes with phylogenetic information profiling. Mol Phylo Evol 89:28–36
Trobajo R, Clavero E, Chepurnov VA, Koen Kabbe, Mann DG, Ishihara S, Cox EJ (2009) Morphological, genetic and mating diversity within the widespread bioindicator *Nitzschia palea* (Bacillariophyceae). Phycologia 48:443–459
Trobajo R, Mann DG, Clavero E, Evans KM, Vanormelingen P, McGregor RC (2010) The use of partial cox1, rbcL and LSU rDNA sequences for phylogenetics and species identification within the *Nitzschia palea* species complex (Bacillariophyceae). Eur J Phycol 45:413–425
Urbánková P, Scharfen V, Kulichová J (2016) Molecular and automated identification of the diatom genus *Frustulia* in northern Europe. Diatom Res 31:217–229. https://doi.org/10.1080/0269249X.2016.1224780
Vanderlaan TA, Ebach MC, Williams DM, Wilkins JS (2013) Defining and redefining monophyly: Haeckel, Hennig, Ashlock, Nelson and the proliferation of definitions. Aust Syst Bot 26:347–355
Vanormelingen P, Chepurnov 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, Chepurnov VA, Mann DG, Sabbe K, Vyverman W (2008) Genetic divergence and reproductive barriers among morphologically heterogeneous sympatric isolates of *Eunotia bilunaris* sensu lato (Bacillariophyta). Protist 159:73–90
Vanormelingen P, Evans KM, Chepurnov VA, Vyverman W, Mann DG (2013) Molecular species discovery in the diatom *Sellaphora* and its congruence with mating trials. Fottea 13:133–148
Vanormelingen P, Evans KM, Mann DG, Lance SL, Debeer AE, D'Hondt S, Verstraete T, de Meester L, Vyverman W (2015) Genotypic diversity and differentiation among populations of two benthic freshwater diatoms as revealed by microsatellites. Mol Ecol. https://doi.org/10.1111/mec.13336
Vaulot D, Birrien J-L, Marie D, Casotti R, Veldhuis MJW, Kraay GW, Chretiennot-Dinet M-J (1994) Morphology, ploidy, pigment composition and genome size of cultured strains of *Phaeocysts*. J Phycol 30:1022–1035

Wetzel CE, Ector L, Van de Vijver B, Compère P, Mann DG (2015) Morphology, typification and critical analysis of some ecologically important small naviculoid species (Bacillariophyta). Fottea Olomouc 15:203–234

Whitaker K, Rynearson TA (2017) Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. PNAS. https://doi.org/10.1073/pnas.1612346114

Williams DM (2007) Classification and diatom systematics: the past, the present and the future. In: Brodie J, Lewis J (eds) Unravelling the algae—the past, present and future of algal systematics. CRC Press, Baton Rouge, pp 57–91

Woodard K, Kulichová J, Poláčková T, Neustupa J (2016) Morphometric allometry of representatives of three naviculoid genera throughout their life cycle. Diatom Res 31:231–242. https://doi.org/10.1080/0269249X.2016.1227375

Young JR, Liu H, Probert I, Aris-Brosou S, De Vargas C (2014) Morphospecies versus phylospecies concepts for evaluating phytoplankton diversity: the case of the coccolithophores. Cryptogam Algol 35:353–377

Zimmermann J, Jahn R, Gemeinholzer B (2011) Barcoding diatoms: evaluation of the V4 sub-region on the 18S rRNA gene, including new primers and protocols. Org Divers Evol. https://doi.org/10.1007/s13127-011-0050-6

Zimmermann J, Glöckner G, Jahn R, Enke N, Gemeinholzer B (2014a) Metabarcoding vs morphological identication to assess diatom diversity in environmental studies. Mol Ecol Res. https://doi.org/10.1111/1755-0998.12336

Zimmermann J, Abarca N, Enk N, Skibbe O, Kusber W-H et al (2014b) Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research. PLoS ONE 9:e108793. https://doi.org/10.1371/journal.pone.0108793

Zingone A, Percopo I, Sims PA, Sarno D (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). I. A re-examination of the type material of *Skeletonema costatum*, with the description of *S. grevillei* sp. nov. J Phycol 41:140–150