3B2v8.06a/w (Dec 5 2003).51c XML:ver:5.0.1 | PROTIS : 25094 | Prod.Type:FTP pp.1–16(col.fig.:NIL) | ED:SumalathaP.N. PAGN:Anvar SCAN:

Protist, Vol. ▮, ▮▮▮–▮▮▮, ▮▮ ▮▮▮▮
http://www.elsevier.de/protis
Published online date • • •

Protist

ORIGINAL PAPER

# An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta)

**Katharine M. Evans[a,1], Alexandra H. Wortley, and David G. Mann**

[a]Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK

Submitted December 13, 2006; Accepted April 7, 2007
Monitoring Editor: Robert A. Andersen

**Due to limited morphological differentiation, diatoms can be very difficult to identify and cryptic speciation is widespread. There is a need for a narrower species concept if contentious issues such as diatom biodiversities and biogeographies are to be resolved. We assessed the effectiveness of several genes (*cox*1, *rbc*L, 18S and ITS rDNA) to distinguish cryptic species within the model 'morphospecies', *Sellaphora pupula* agg. This is the first time that the suitability of *cox*1 as an identification tool for diatoms has been assessed. A range of *cox*1 primers was tested on *Sellaphora* and various outgroup taxa. Sequences were obtained for 34 isolates belonging to 22 *Sellaphora* taxa and three others (*Pinnularia, Eunotia* and *Tabularia*). Intraspecific divergences ranged from 0 to 5 bp (= 0.8%) and interspecific levels were at least 18 bp (= c.3%). *Cox*1 divergence was usually much greater than rbcL divergence and always much more variable than 18S rDNA. ITS rDNA sequences were more variable than *cox*1, but well-known problems concerning intragenomic variability caution against its use in identification. More information and less sequencing effort mean that *cox*1 can be a very useful aid in diatom identification. The usefulness of *cox*1 for determining phylogenetic relationships among Sellaphora species was also assessed and compared to rbcL. Tree topologies were very similar, although support values were generally lower for *cox*1.**


**Keywords:** *cox*1; diatom; DNA barcoding; microbial species concept; *rbc*L; *Sellaphora pupula* agg.

## Introduction

Diatoms are single-celled eukaryotes of enormous ecological importance, accounting for at least 20% of all carbon fixed globally through photosynthesis each year (Mann 1999). They are major constituents of benthic and planktonic communities, occurring in terrestrial, marine and freshwater environments worldwide. Despite their key role in the functioning of many ecosystems and their use as bioindicators of water quality and past climates, major aspects of their biology are poorly understood and consequently contentious, for example their biodiversity and geographical distributions (e.g. Finlay et al. 2002; Mann and Droop 1996; see also Hughes Martiny et al. 2006). In common with all microbial groups, this is in large part due to the lack of a robust taxonomic system. Diatom species have previously been distin-

[1]Corresponding author. Fax +44 131 248 2901.
e-mail k.evans@rbge.ac.uk (K.M. Evans).


doi:10.1016/j.protis.2007.04.001

guished according to the shape, size, patterning and ultrastructure of their silica exoskeleton, with little or no understanding of the genetic basis of these features. Recent evidence from molecular data and mating experiments have shown that traditional classifications are too coarse: 'morphospecies' (the outcome of applying the 'morphological species concept' sensu Mayr 1942) can contain several to many pseudo- or semi-cryptic taxa, for example in *Pseudo-nitzschia* (e.g. Lundholm et al. 2006), *Skeletonema* (Sarno et al. 2005) and *Neidium* (Mann and Chepurnov 2005).

Perhaps the best studied example of cryptic speciation in diatoms is in *Sellaphora pupula* agg., a common and cosmopolitan freshwater benthic species complex, for which many morphological, mating and sequence data have been generated (Behnke et al. 2004; Evans et al. unpubl.; Mann 1984, 1989, 1999; Mann and Droop 1996; Mann et al. 1999, 2004). These studies have demonstrated the existence of a large number of pseudo- and semi-cryptic species. For example, within Blackford Pond in Edinburgh, nine *S. pupula* agg. taxa live sympatrically without intergradation, of which six have now been described as separate species (Mann et al. 2004). Although these species are morphologically similar, they are distinct and reproductively isolated, possess different mating systems, exhibit different degrees of genetic relationship to each other, and differ in their sensitivity to parasites, and molecular data and the fossil record suggest that they have been diversifying for at least 12 million years (Behnke et al. 2004; Mann et al. 2004).

Morphology is not always a reliable indicator of species boundaries and phylogenetic relationships in *S. pupula* agg. and even where morphological differences do exist, for example between the sister taxa *S. pupula* agg. "pseudocapitate" and *S. blackfordensis*, they are so subtle that they cannot be used for routine identification by eye: at first, despite our long experience with the genus, *S. pupula* agg. "pseudocapitate" clones were identified as *S. blackfordensis* (Behnke et al. 2004). Consequently, sequence data are being used increasingly to aid the recognition of species boundaries in *S. pupula* agg. and other diatom taxa (Amato et al. 2007; Behnke et al. 2004; Medlin et al. 1991; Sarno et al. 2005; Vanormelingen et al. 2007). We initially chose *rbc*L as the core identification region because (1) it is protein coding and so alignment is straightforward, (2) it is more variable than 18S rDNA and (3) being a plastid gene, the likelihood of amplifying contaminant DNA (e.g. fungal) is much reduced.

Sequence data have also been used to identify macroalgae (e.g. Brodie et al. 1996), fungi (e.g. Taylor et al. 2000) and prokaryotes (e.g. Pace 1997). In recent years, there has been a push to extend this approach to animals and higher plants under the banner 'DNA barcoding' (Hebert et al. 2003; Savolainen et al. 2005). For animals, an approximately 650 bp fragment of the mitochondrial gene, cytochrome *c* oxidase (*cox*1), has been chosen after promising results from tests on a wide range of organisms (e.g. Blaxter et al. 2004; Hajibabaei et al. 2006; Hebert et al. 2004; Witt et al. 2006). Saunders (2005) and Robba et al. (2006) have shown that *cox*1 is also suitable for discriminating red algal taxa. Recently, Seifert et al. (2007) have demonstrated its use in identifying fungal taxa. In contrast, in higher plants, *cox*1 is insufficiently variable and other genes are being investigated instead (Chase et al. 2005; Kress et al. 2005; Newmaster et al. 2006). The only study of *cox*1 in diatoms, by Ehara et al. (2000a), was not designed to test its use as a barcode for species, but showed that the gene is sufficiently variable to resolve some higher order relationships.

For organisms with robust taxonomies (for example most higher plants and many animal groups), the main role of DNA barcoding is to aid the identification of previously described species (e.g. Cowan et al. 2006; Hajibabaei et al. 2006). The ultimate goal is the same in diatoms and other microbes, but because the true diversities of these organisms are often unknown (Pedrós-Alió 2006; Šlapeta et al. 2005), it would be an advantage if a barcode region gave preliminary information on interspecific relationships.

The pros and cons surrounding a DNA barcoding approach to identification and discovery of new species have been discussed extensively elsewhere (e.g. Savolainen et al. 2005 and references therein). Two of the major concerns are whether rigid threshold approaches to intra versus interspecific variation are appropriate (e.g. Meyer and Paulay 2005; Witt et al. 2006) and whether reliance on a single, uniparentally inherited gene provides sufficiently accurate information (e.g. Hurst and Jiggins 2005; Rubinoff 2006). Whilst we acknowledge that DNA barcoding needs refinement (e.g. Pons et al. 2006) and further testing, such an approach is vital for identifying microbial species because of the prevalence of cryptic diversity (Fig. 1). In this paper we aim to:

(1) Examine the performance of existing algal *cox*1 primer sets in amplifying the DNA of



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

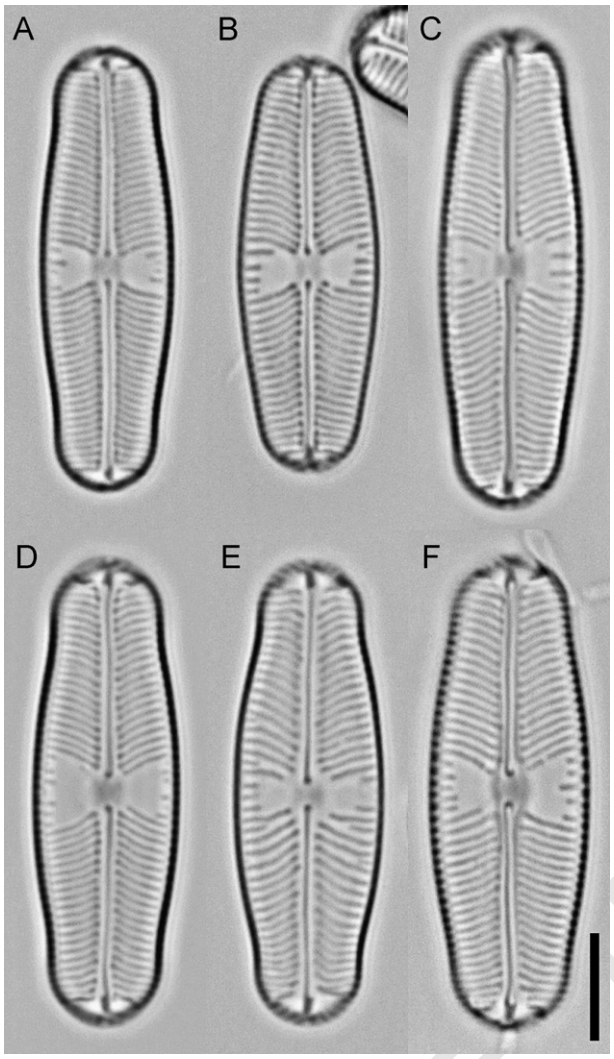


**Figure 1**. *Sellaphora pupula* agg.: valves of selected demes. **A.** "Small capitate" DUN7. **B.** "Europa" GER1. **C.** "Southern pseudocapitate" AUS4. **D.** "Southern capitate" AUS1. **E.** "Afro" AFR1. **F.** "Pseudocapitate" L845 (parental to F1 clone; Table 1). Bar = 5 μm. Diatoms become progressively smaller with each asexual division; size of valves is therefore of limited use in distinguishing taxa.

*Sellaphora* and other diatoms and, if necessary, design and test additional primers.

(2) Assess how well *cox*1 and *rbc*L perform as barcode regions, by reference to levels of divergence within and among *Sellaphora* taxa. We also discuss more limited data-sets available for 18S and ITS rDNA.

(3) Compare the effectiveness of *cox*1 and *rbc*L in determining phylogenetic relationships in *Sellaphora.*

## Results

*Sellaphora pupula* agg. demes that we recognize as separate biological entities, but have not yet described as distinct species, are referred to as *S. pupula* agg. followed by a morphological and/or geographical identifier, e.g. *S. pupula* agg. "small lanceolate" and *S. pupula* agg. "southern capitate" (Table 1); from here onwards, we will omit the *S. pupula* agg. prefix. *Sellaphora* isolates are coded depending on their geographic origin, for example THR isolates were collected from Threipmuir Reservoir (Table 1).

### Algal *cox*1 Primers

The success of the various algal *cox*1 primers is summarized in Table 2. Briefly, the diatom primers pC1 and pB1 (Ehara et al. 2000a) amplified the expected 1100 bp product in all 13 *Sellaphora, Pinnularia* cf. *gibba, Seminavis* cf. *robusta* and *Paralia sulcata* DNAs that were tested, but only the red algal primer GazR1 (Saunders 2005) was successful as a sequencing primer (in seven instances: BLA2, BLA16, F1, BEL3, RBG1, THR2 and *P.* cf. *gibba*; Table 1). Alignment of these and ten GenBank diatom *cox*1 sequences enabled the design of the internal primer KEintF, which extended the seven sequences in the 3′ direction, allowing the design of a diatom-specific reverse barcode primer approximately 700 bp into the gene, KEdtmR (Table 3).

The brown algal *cox*1 barcode forward primer, GazF2 (Table 3) and KEdtmR amplified a single band of the expected size (707 bp) for 28 of the 30 *Sellaphora* and outgroup taxa (exceptions were *S. auldreekie* and "small lanceolate"). GazF2 and KEdtmR were also successful sequencing primers for all *Sellaphora* taxa except *S. lanceolata*, *S. auldreekie* and "small lanceolate"; the failure of the latter two was not unexpected, because more than one product was amplified. In these instances, substituting KEdtmR with the brown algal *cox*1 reverse barcode primer GazR2 led to no improvement in either amplification or sequencing success (Table 2).

Sequences were also obtained for three of the seven outgroup taxa: the two raphid pennates *P.* cf. *gibba* and *Eunotia* sp., and the araphid pennate *Tabularia* sp. For outgroup and ingroup sequencing failures, internal primers, KEint2F and KEintR (Table 3) were trialled, and in the cases of *S. auldreekie* (BLA2), *S. lanceolata* and *Tabularia* sp., extra information was gained. When PCRs and sequencing were repeated on other representa-



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

**Table 1.** Identities and sources of isolates (UK National Grid References or latitudes and longitudes are given where possible), voucher slides (all in the diatom herbarium of the Royal Botanic Garden, Edinburgh), and GenBank accession numbers for *cox*1 and *rbc*L sequences.

| Taxon | Identifier | Original designation and synonyms | Voucher | Origin | Isolated by | Date | GenBank accessions | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | *Cox*1 | *Rbc*L |
| *Sellaphora auldreekie* | BLA2[a] | SEL504B | E3652 | Scotland: Blackford Pond (NT 253709) | D.G. Mann/A. Poulíčková | 2004 | EF164932 | EF143303 |
| *Sellaphora auldreekie* | DUN1 | 25D | E3018 | Scotland: Dunsapie Loch (NT 281731) | V.A. Chepurnov | 1996 | — | EF143276 |
| *Sellaphora bacillum* | BLA4 | SEL438B | E3640 | Scotland: Blackford Pond | V.A. Chepurnov | 2003 | EF164941 | EF143311 |
| *Sellaphora bacillum* | ST-MAR1 | 17SM | E3149 | Scotland: St Margaret's Loch (NT 276739) | V.A. Chepurnov | 1996 | EF164930 | EF143273 |
| *Sellaphora blackfordensis* | DUN5[a] | SEL223D | E3632 | Scotland: Dunsapie Loch | V.A. Chepurnov | 2003 | EF164949 | EF143319 |
| *Sellaphora blackfordensis* | BLA17[a] | BI.85 | — | Scotland: Blackford Pond | D.G. Mann | 2004 | EF164948 | — |
| *Sellaphora blackfordensis* | BAL1[a] | (Bal7)3 | E3266 | Scotland: Balgavies Loch (NO 531508) | D.G. Mann | 1997 | EF164935 | EF143310 |
| *Sellaphora capitata* | DUN6[a] | Dunsapie 138 | E3648 | Scotland: Dunsapie Loch | D.G. Mann/A. Poulíčková | 2004 | EF164946 | EF143286 |
| *Sellaphora capitata* | BEL3[a] | Mf21 | — | Belgium: De oude Scheldearm (51°01′N, 3°45′E) | V.A. Chepurnov | 2005 | EF164939 | — |
| *Sellaphora capitata* | BLA18[a] | BI.89 | E3608 | Scotland: Blackford Pond | D.G. Mann | 2004 | EF164947 | — |
| *Sellaphora laevissima* | THR2 | 24Tr | E3174 | Scotland: Threipmuir Reservoir (NT 168636; NT 163634) | V.A. Chepurnov | 1996 | EF164931 | EF143265 |
| *Sellaphora laevissima* | THR4 | SEL301TM | E3651 | Scotland: Threipmuir Reservoir | D.G. Mann | 2003 | EF164943 | EF143309 |
| *Sellaphora lanceolata* | BLA12 | SEL432B | E3635 | Scotland: Blackford Pond | V.A. Chepurnov | 2003 | EF164953 | EF143315 |
| *Sellaphora pupula* agg. "afro" | AFR1 | NgNg42 | E3622 | Tanzania: NgoroNgoro crater | V.A. Chepurnov | 2002 | EF164940 | EF143300 |
| *Sellaphora pupula* agg. "cf. capitata" | THR7[a] | Thr37 | E3365 | Scotland: Threipmuir Reservoir | D.G. Mann | 1998 | EF164937 | EF143302 |
| *Sellaphora pupula* agg. "cf. obese" | BEL2[a] | V8 | E3620 | Belgium: Kraenepoel (51°04′N, 3°29′E) | V.A. Chepurnov | 2001 | EF164933 | EF143266 |
| *Sellaphora pupula* agg. "cf. small blunt-capitate" | THR8 | SEL292TM | E3628 | Scotland: Threipmuir Reservoir | D.G. Mann | 2003 | EF164950 | EF143308 |
| *Sellaphora pupula* agg. "elliptical" | RBG1 | 13R | E3024 | Scotland: Royal Botanic Garden Edinburgh Pond (NT 248753) | V.A. Chepurnov | 1996 | EF164957 | EF143271 |



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Sellaphora pupula* agg. "elliptical" | BLA14[a] | SEL9B | E3634 | Scotland: Blackford Pond | D.G. Mann | 2003 | EF164944 | EF143294 |
| *Sellaphora pupula* agg. "elliptical" | THR9 | SEL293TM | E3625 | Scotland: Threipmuir Reservoir | D.G. Mann | 2003 | EF164958 | EF143297 |
| *Sellaphora pupula* agg. "elliptical" | THR10[a] | SEL299TM | E3642 | Scotland: Threipmuir Reservoir | D.G. Mann | 2003 | EF164952 | EF143296 |
| *Sellaphora pupula* agg "europa" | GER1[a] | Fr.See.18 | E3325 | Germany: Friedheimer See (53°31′23″N, 8°39′54″E) | V.A. Chepurnov | 1997 | EF164955 | EF143318 |
| *Sellaphora pupula* agg. "large" | THR11 | 79TmNE | E3415 | Scotland: Threipmuir Reservoir | V.A. Chepurnov | 1997 | EF164945 | EF143285 |
| *Sellaphora pupula* agg. "little" | RBG2 | 20R | E3020 | Scotland: Royal Botanic Garden Pond | V.A. Chepurnov | 1996 | EF164928 | EF143275 |
| *Sellaphora pupula* agg. "pseudocapitate" | F1[a] | (L845.5 × Thr42) | E3384 | $F_1$ clone (see Methods) | V.A. Chepurnov | 1999 | EF164942 | EF143291 |
| *Sellaphora pupula* agg. "small blunt-capitate" | THR13 | 6Tm | E3026, E3161-2 | Scotland: Threipmuir Reservoir | D.G. Mann | 1996 | EF164954 | EF143272 |
| *Sellaphora pupula* agg. "small capitate" | DUN7 | SEL92D | E3638 | Scotland: Dunsapie Loch | D.G. Mann | 2003 | EF164936 | EF143269 |
| *Sellaphora pupula* agg. "small lanceolate" | THR14 | 5TM | E3160 | Scotland: Threipmuir Reservoir | V.A. Chepurnov | 1996 | — | EF143264 |
| *Sellaphora pupula* agg. "southern capitate" | AUS1[a] | HL5 | E3623 | Australia: Hacking River (34°04′22″S, 151°03′27″E) | V.A. Chepurnov | 2001 | EF164959 | EF143312 |
| *Sellaphora pupula* agg. "southern pseudocapitate" | AUS4 | HL18 | E3466 | Australia: Hacking River | V.A. Chepurnov | 2001 | EF164934 | EF143317 |
| *Sellaphora pupula* agg. "spindle" | BLA16[a] | KE68 | E3658 | Scotland: Blackford Pond | K.M. Evans | 2005 | EF164951 | EF143298 |
| *Sellaphora* cf. *minima* | — | BM 42 | E3211 | Scotland: Blackford Pond | D.G. Mann | 1997 | EF164929 | EF143279 |
| *Sellaphora* cf. *seminulum* | — | TM 37 | E3168 | Scotland: Threipmuir Reservoir | D.G. Mann | 1997 | EF164956 | EF143280 |
| *Eunotia* sp. | — | EUN 392T | — | Scotland: Loch Tulla (NN 294428) | D.G. Mann | 2003 | EF164960 | — |
| *Melosira nummuloides* | — | B4 | — | Scotland: North Berwick (NT 553858) | D.G. Mann | *c.* 2000 | — | — |
| *Odontella sinensis* | — | Odontella Can. Is. | — | Canary Isles: Gran Canaria | A. de la Jara/ D.G. Mann | 1999 | — | — |
| *Paralia sulcata* | —[a] | NBK-2 | — | Scotland: North Berwick | D.G. Mann | 1999 | — | — |
| *Pinnularia* cf. *gibba* | —[a] | Clone 12 | E3603 | Scotland: Loch of Craiglush (NO 041444) | A. Poulíčková | 2004 | EF164938 | EF143304 |
| *Seminavis* cf. *robusta* | —[a] | F1-10A | — | $F_1$ clone (see Methods) | T. Moens | 2000 | — | — |
| *Tabularia* sp. | — | Tabularia Naples | — | Italy: Naples | W.H.C.F. Kooistra | — | EF164927 | — |

[a]Clones are alive and available upon request; the remainder are dead, predominantly due to asexual reproduction which eventually results in cells that are too small to remain viable.



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

**Table 2.** Amplification and sequencing successes of a range of algal *cox*1 primers.

| Primer set/ primer | Amplification success | Sequencing success |
|---|---|---|
| GazF1 | 0/13 | 0/13 |
| GazR1 | | 7/13[a] |
| pC1 | 13/13 | 0/13 |
| pB1 | | 0/13 |
| GazF2 | 37/40 | 32/40 |
| KEdtmR | | 32/40 |
| GazF2 | 0/3[b] | 0/4[b] |
| GazR2 | | |
| KEintF | n/a | 7/7 |
| KEint2F | n/a | 3/8 |
| KEintR | n/a | 3/8 |

[a]As a sequencing primer of the pC1/pB1 PCR product.
[b]GazR2 only tested in instances where KEdtmR had failed.

**Table 3.** Sequences of brown algal and diatom *cox*1 primers used in this study.

| Name | Sequence (5′-3′) |
|---|---|
| *Brown algal* | |
| GazF2 | CAA CCA YAA AGA TAT WGG TAC |
| GazR2 | GGA TGA CCA AAR AAC CAA AA |
| *Diatom* | |
| KEintF | GAG AGC AAA AAG TTT ACC ATT TCA |
| KEdtmR | AAA CTT CWG GRT GAC CAA AAA |
| KEint2F | GAA GCW GGW GTW GGT ACW GGW TG |
| KEintR | CAA ATA AAA TTR ATW GCW CCT AA |

tives of the same outgroup taxa, in two cases (*S.* cf. *robusta* and *Melosira nummuloides*), non-diatom sequences were obtained; BLAST searches listed an invertebrate and oomycete, respectively, as the closest matches, both possible contaminants in the cultures.

In summary, *cox*1 barcodes were obtained for 22 of the 23 *Sellaphora* taxa ("small lanceolate" was the one exception) and for three of the seven outgroup taxa. In all instances sequences were unambiguous. Primer set GazF2 and KEdtmR will probably work well for most species within the raphid pennate lineage and for some araphids, but probably not for those belonging to the centric lineage. Although GazR2 was not tested extensively, due to sequence overlap with KEdtmR (Table 3), it is likely that it will perform comparably to KEdtmR when used together with GazF2. The 34 (31 ingroup and 3 outgroup) *cox*1 sequences were aligned with no gaps over 624 bp (the barcode region, excluding the primers was 665 bp). Seven isolates lacked some sequence information: up to 16 bp could not be determined for six isolates and 115 bp for *Tabularia* sp.

## Intra- and Interspecific *cox1* and rbc*L* Divergences

*Cox*1 data were obtained for five *Sellaphora* taxa represented by more than one isolate, viz. *S. capitata* ($n = 3$), *S. blackfordensis* ($n = 3$), "elliptical" ($n = 4$), *S. bacillum* ($n = 2$) and *S. laevissima* ($n = 2$; Table 1). Despite the geographic range over which the *S. capitata* isolates were sampled (Scotland and Belgium), all three had identical *cox*1 and *rbc*L sequences; the two *S. bacillum* isolates were also identical at both regions (Table 4). In contrast, although the three *S. blackfordensis* isolates had identical *rbc*L sequences, in *cox*1 there were 5 bp differences between the two Edinburgh isolates (BLA17 and DUN5; lakes separated by *c.* 5 km) and 1 bp difference between the isolate from Dunsapie Loch (DUN5) and an isolate from Balgavies Loch (BAL1; lakes separated by *c.* 80 km; Table 4). The *cox*1 data for *S. laevissima* and "elliptical" isolates strongly support our previous suspicions of additional cryptic diversity (Table 4). The *cox*1 sequences of the two

**Table 4.** Levels of *cox*1 (*c.* 624 bp) and *rbc*L (*c.* 1400 bp) divergence among *Sellaphora* isolates originally thought to belong to the same taxon: percentage divergence and number of bp differences. The degree of divergence among "elliptical" and *Sellaphora laevissima* isolates indicates additional cryptic diversity.

| Taxon | *cox*1 % divergence (no. bp) | *rbc*L % divergence (no. bp) |
|---|---|---|
| *Sellaphora capitata* | 0 | 0 |
| *Sellaphora bacillum* | 0 | 0 |
| *Sellaphora blackfordensis* | 0.8% (5) | 0 |
| "Elliptical" | 3.2% (20) | 0.4% (5) |
| *Sellaphora laevissima* | 7.9% (49) | 2.2% (30) |



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

*S. laevissima* isolates differed by 49 bps (= 7.9%) compared with 30 *rbc*L bp differences (= 2.2%; Fig. 3). There were up to 20 *cox*1 bp differences (= 3.2%) between "elliptical" isolates compared to 5 *rbc*L differences (= 0.4%).

In all cases, interspecific *cox*1 divergence was much greater than the corresponding *rbc*L divergence (Fig. 3). For example, between *S. blackfordensis* and its closest known relative, "pseudocapitate", with which it can form an $F_1$ (Behnke et al. 2004), there were up to 23 bp differences (= 3.7%) in *cox*1 compared to 21 *rbc*L differences (= 1.6%); and between *S. capitata* and the closely related "afro", there were 18 *cox*1 differences (= 2.9%) compared to 13 *rbc*L differences (= 0.9%; no data are available for reproductive compatibility of these two taxa).

## *Cox*1 and rbc*L* Phylogenetic Analyses

For *cox*1 (alignment 624 bp), 298 characters were variable and 206 were parsimony-informative. Within *Sellaphora* itself, 225 characters were variable and 158 were parsimony-informative. Maximum parsimony (MP) analysis found four trees of length 899 steps. No new islands of trees were found using the second search strategy. For maximum likelihood (ML) analyses, the hierarchical likelihood ratio (HLR) test selected the model GTR+I+G, for which $-\ln L = 4755.8711$, $K = 10$, and base frequencies were A = 0.2910, C = 0.1123, G = 0.1375 and T = 0.4592. The rate matrix was A−C = 1.0447, A−G = 3.6703, A−T = 0.7520, C−G = 0.0005, C−T = 6.7541 and G−T = 1.0000. The gamma distribution shape parameter was 0.6504. ML analysis produced a tree with likelihood $-\ln L = 4751.08707$ (Fig. 2A), which is very similar in topology to the MP strict consensus tree (not shown).

For *rbc*L (alignment 1399 bp), 252 characters were variable and 150 were parsimony-informative. Within *Sellaphora* itself, 188 characters were variable and 111 were parsimony-informative. MP analysis found 369 trees of length 577 steps. No new islands of trees were found using the second search strategy. The HLR test selected the model TrN+I+G, for which $-\ln L = 5021.2383$, $K = 7$, and base frequencies were A = 0.2997, C = 0.1689, G = 0.1984 and T = 0.3330, with rates of change A−G = 3.0015 and

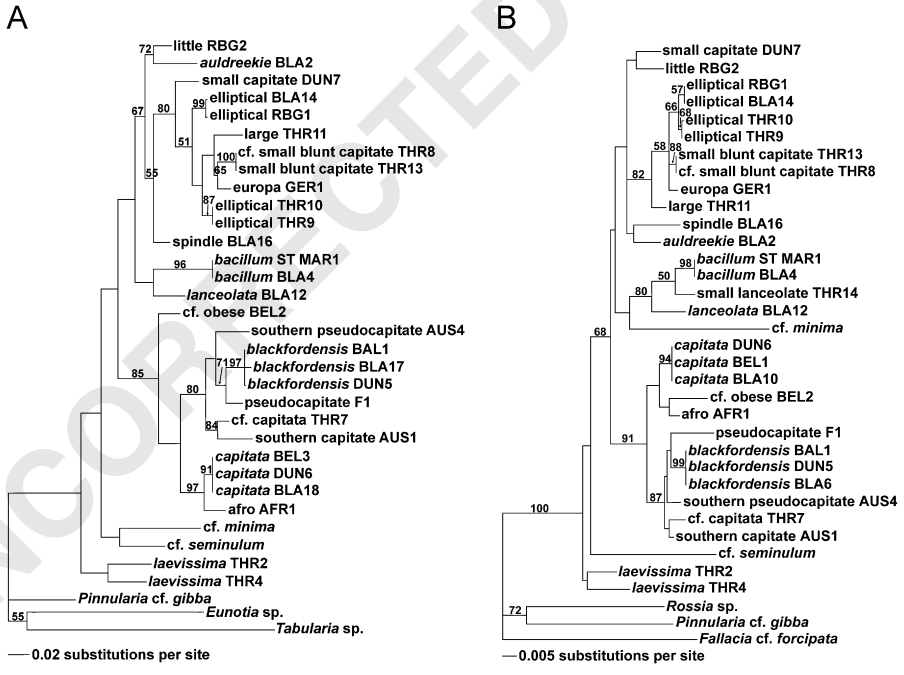


**Figure 2**. Phylogenies obtained from ML analysis of *cox*1 (A) and *rbc*L (B) sequences. *Sellaphora* isolates are identified by species or deme name plus a suffix indicating their provenance (see Table 1). In the *rbc*L analysis (B), three *Sellaphora* isolates are different to those used in the *cox*1 analysis (A), but are considered sufficiently closely related (i.e. closer than to any other taxon in the tree) to act as placeholders for the same taxa (see Methods). Branch lengths are proportional to the number of substitutions inferred to have occurred upon them. Numbers above branches represent JK support values where greater than 50%.



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

C—T = 5.3232. The gamma distribution shape parameter was 0.6793. ML analysis produced a tree with likelihood $-\ln L = 4998.8777$ (Fig. 2B), which is very similar in topology to the MP strict consensus tree (not shown).

As Fig. 2A shows, in *cox*1 analysis, all *Sellaphora* taxa formed a monophyletic group, with 84% bootstrap (BS) in MP analysis (not shown), although jack-knife (JK) support in ML analysis was lower. Whereas *S. pupula* agg. was paraphyletic, *S. bacillum* was strongly supported as a monophyletic group in both types of analysis (100% MP BS, 96% ML JK) and *S. laevissima* was well supported in MP (97% BS); strong support was also found for the monophyly of two recently described species (Mann et al. 2004) for which three clones of different provenance were included, viz. *S. blackfordensis* (100% MP BS, 97% ML JK) and *S. capitata* (98% MP BS, 91% ML JK). The two "elliptical" groupings (Table 4) were strongly supported (over 98% MP BS, over 85% ML JK), as were two deme pairs (100% MP BS, over 95% ML JK): "small blunt-capitate"+"cf. small blunt-capitate" and *S. capitata*+"afro". Within *Sellaphora,*one clade, the "*blackfordensis*" group was particularly well supported (98% MP BS, 85% ML JK). This group comprises *S. blackfordensis, S. capitata,* "afro," "cf. obese," "southern capitate," "cf. capitata," "pseudocapitate" and "southern pseudocapitate".

A comparison of Figs. 2A and 2B shows that the *cox*1 and *rbc*L data-sets give very similar results in terms of topology, although support values are generally higher for *rbc*L. As for *cox*1, all *Sellaphora* formed a monophyletic group, although with stronger support (94% MP BS, 100% ML JK). *Sellaphora bacillum, S. blackfordensis* and *S. capitata* were also all strongly supported as monophyletic groups (at least 95% MP BS and ML JK). Both "elliptical" clades were supported although less strongly than in the *cox*1 analysis (over 80% MP BS, over 55% ML JK). As for *cox*1, the deme pair "small blunt-capitate"+"cf. small blunt-capitate" was well supported (97% MP BS, 88% ML JK). The same was true for the "*blackfordensis*" group (93% MP BS, 91% ML JK). The major difference between the two phylogenies is the position of *S.* cf. *minima*, which is resolved outside *S. pupula* agg. using *cox*1 and within *S. pupula* agg. using *rbc*L.

## Discussion

### Is *cox*1 a Suitable Gene for Diatom Identification?

**Level of Variation Within and Among *Sellaphora* Taxa:** Despite the much shorter length of the *cox*1 sequences, the absolute amount of intraspecific divergence was mostly either equal or greater than that for *rbc*L. For example, for *S. capitata* and *S. bacillum* isolates there was neither *cox*1 nor *rbc*L divergence, whereas for the three *S. blackfordensis* isolates from Dunsapie Loch, Blackford Pond and Balgavies Loch, although there was no *rbc*L divergence, there were up to five *cox*1 bp differences (Table 4). Therefore, for some species, *cox*1 may provide useful information on population structure or racial differences: breeding experiments have shown no intrinsic reproductive barriers between Dunsapie, Blackford and Balgavies clones of *S. blackfordensis* (Behnke et al. 2004; Chepurnov and Mann, unpubl.), though these data do not refer to the same clones as those used in the present study. "Elliptical" and *S. laevissima* have been shown by *rbc*L (and in the case of *S. laevissima,* also by rDNA) analysis to contain two molecular clades (Behnke et al. 2004; Evans et al. unpubl.) and in both cases the *cox*1 sequences were more divergent than the *rbc*L data (Table 4 and Fig. 2). Two groupings of "elliptical" isolates were well supported (Fig. 2), viz. the 'upland' isolates, THR9 and THR10 (identical *cox*1 sequences, but one *rbc*L bp difference), both from the mesotrophic, upland Threipmuir Reservoir and the 'urban' isolates, RBG1 and BLA14 (two *cox*1 bp differences, but *rbc*L sequences identical), both from highly eutrophic ponds in Edinburgh. The degree of divergence between the two *S. laevissima* isolates and the two "elliptical" subclades supports their delineation as separate taxa (Table 4: both sets of taxa also differ subtly in their morphologies; Evans et al. unpubl.), but before formalizing this, it will be important to obtain additional data from nuclear markers (e.g. rDNA). Levels of interspecific *cox*1 divergence between previously recognized taxa or closely related demes were at least 18 bp or 2.9% (between *S. capitata* and "afro"; Fig. 3). Interspecific *cox*1 divergences were always much greater than corresponding *rbc*L data (Fig. 3).

Other genes commonly used to identify diatoms and to construct phylogenies are the nuclear rDNA regions, 18S and ITS, which are approximately 1800 bp and 650—850 bp in length, respectively



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

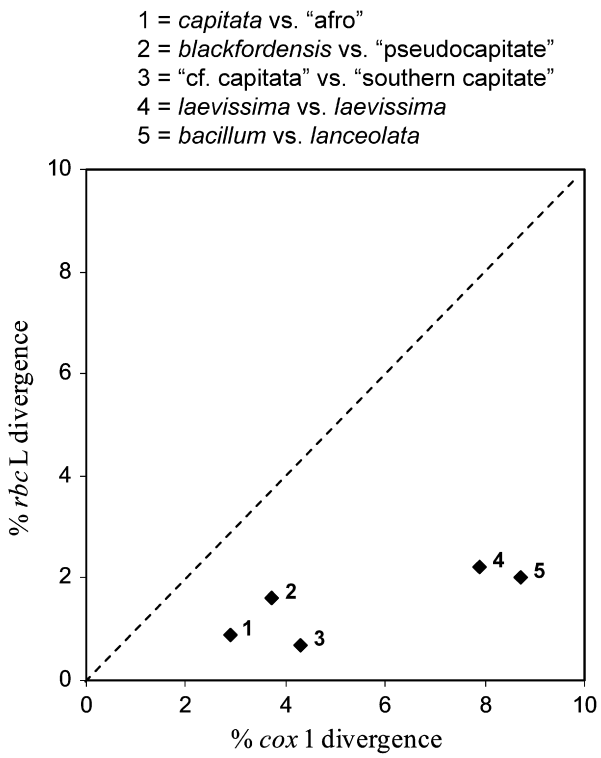


**Figure 3**. Divergences of *cox*1 (*c.* 624 bp) and *rbc*L (*c.* 1400 bp) sequences between selected taxa. The dashed line indicates identical levels of divergence.

(e.g. Beszteri et al. 2005; Orsini et al. 2004). As yet, there are few 18S and ITS data for *Sellaphora* taxa (Behnke et al. 2004, Evans et al. unpubl.) and so it is possible only to make a few comparisons with *cox*1 and *rbc*L and these are not direct (i.e. the sequences were not all obtained from the same isolates). Nevertheless, it is apparent that the 18S region is less variable than *rbc*L and much less variable than *cox*1; for example between two *S. laevissima* isolates that belong to different *rbc*L clades (THR1 and THR4: GenBank accessions EF151979 and EF151981, respectively), there are just 8 bp differences in 18S rDNA ( = *c.* 0.4%, compared with 7.9% for *cox*1 and 2.2% for *rbc*L); and between *S. blackfordensis* and "pseudocapitate" isolates (BLA6 and US1; EF151969 and AJ544649, respectively) there are just 3 bp differences ( = *c.* 0.2%, compared with 3.7% for *cox*1 and 1.6% for *rbc*L). In contrast to 18S, the ITS region (ITS-1—5.8S—ITS-2) is highly variable and is often chosen to investigate within-population genetic variation, for example in *Pseudo-nitzschia* (e.g. Orsini et al. 2004). Behnke et al. (2004) reported 23 differences ( = *c.* 2.8%) among six *S. blackfordensis* isolates, five of which originated from the same pond; and 92 differences ( = *c.* 11%) between *S. laevissima* isolates obtained from Scotland and the US (as for the *S. laevissima* isolates included in this study, these probably represent distinct taxa).

To sum up, we have shown that *cox*1 is sufficiently variable to (a) distinguish previously recognized taxa, (b) draw attention to potential cryptic taxa and (c) act in some cases as a geographical marker. This accords with the diagnostic potential of *cox*1 in metazoa (e.g. Hajibabaei et al. 2006) and red algae (Robba et al. 2006; Saunders 2005).

## Other Considerations

Besides being suitably variable, there are a number of other factors to consider before deciding on a gene(s) for barcode identification and preliminary phylogenetic analyses. The first is the mode of inheritance. Uniparental inheritance of organelles carries the risk that sequence data will not accurately reflect species boundaries if hybridization and subsequent introgression have occurred (e.g. Gompert et al. 2006). This scenario is probably less common in diatoms than in many land plants, because most diatom clones, like individual animals, have a finite and short life (of a few years only) because of cell size reduction, after which regeneration via a sexual phase is usually obligatory. Hence reduced fertility of hybrids could not be circumvented indefinitely by vegetative propagation, leading to the likelihood of strong selection for effective mechanisms of reproductive isolation, or restoration of full fertility, wherever incipient diatom species occur in sympatry. Nevertheless, it is desirable to also sequence a nuclear gene (Rubinoff 2006), particularly when examining closely related species.

In animals and land plants (excluding most conifers), mitochondria and chloroplasts are maternally inherited. In algae, maternal inheritance of organelles also appears to be dominant (e.g. Coyer et al. 2002), and may even occur, surprisingly, in organisms that reproduce isogamously. Thus, the isogamous brown alga *Scytosiphon lomentaria* exhibits uniparental inheritance of mitochondria, but biparental inheritance of the chloroplasts (Kato et al. 2006). The isogamous green alga *Chlamydomonas reinhardtii* transmits both mitochondria and chloroplasts uniparentally, but from different mating types (Aoyama et al. 2006). How diatoms inherit organelles during the sexual phase is poorly known. The sperm of some oogamous centric diatoms certainly lose any



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

visible sign of chloroplasts (during hologenous gametogenesis: Jensen et al. 2003), implying inheritance via the egg alone; however many centric diatoms are monoecious, producing eggs and sperm during different phases of the same life cycle (Chepurnov et al. 2004), so that each clone *will* transmit its chloroplasts to progeny. In pennate diatoms the parental chloroplasts appear to retain their integrity throughout sexual reproduction and are segregated at the first division of the $F_1$ initial cell (Mann 1996). Nothing is known about mitochondrial inheritance in diatoms; we are currently researching this issue, using the *cox*1 variation we detected among *S. blackfordensis* (Fig. 2A).

Second, it is important to know to which genome the gene belongs. For example in higher plants, gene transfer from the mitochondria to the nucleus is a frequent, ongoing process, leading to paralogous gene sequences (see references in Gray et al. 2004). So far, only one diatom mitochondrial genome has been sequenced in its entirety, that of *Thalassiosira pseudonana*. At least in this diatom, there is no evidence for recent large-scale transfers of mitochondrial DNA to the nucleus and *cox*1 is part of the mitochondrial genome (Armbrust et al. 2004).

Third, evolutionarily frequent horizontal transfer of mitochondrial genes between distantly related angiosperm taxa has been documented; in contrast, horizontal transfer of chloroplast DNA in land plants is likely to be much lower (Bergthorsson et al. 2003). Whether horizontal transfer of mitochondrial genes also occurs between algal taxa is unknown since much fewer mitochondrion-based phylogenetic analyses have been conducted.

Fourth, if there are introns within the barcoding gene, or if the gene is split into two or more segments, there could be amplification and sequencing problems. Unlike animals (Beagley et al. 1996), introns are abundant in land-plant mitochondria (Knoop 2004). They are also found in some algal mitochondrial genes, for example the *cox*1 genes of the diatoms *T. pseudonana* (2338 bp in length; Armbrust et al. 2004) and *T. nordenskioeldii* (2311—2468 bp in length; Ehara et al. 2000b) and several green algae (Watanabe et al. 1998). An isolate of the brown alga *Pylaiella littoralis* possessed three introns in a *cox*1 gene that was approximately 9000 bp in length (Fontaine et al. 1997); however, no introns were found in the *cox*1 genes of four other brown algae (Oudot-Le Secq et al. 2002, 2006). Lastly, in the haptophyte *Emiliania huxleyi*, *cox*1 is spliced from two mitochondrial segments, 10 kbp apart (Sánchez Puerta et al. 2004). So far, introns found in diatom *cox*1 genes occur outside the 'barcode region', i.e. beyond the first 700 bp. *Rbc*L can also contain introns (e.g. in green algae: Nozaki et al. 2002), but we have detected none in over 150 species of diatoms (unpubl.).

Fifth, although not an absolute requirement for DNA barcoding, ease of alignment is an advantage if the sequence data are to contribute to phylogenetic analyses. In contrast to 18S and ITS rDNA, both *cox*1 and *rbc*L are coding genes and consequently contain few indels, making sequences easy to align. For ITS in particular, this is not the case. For example, Behnke et al. (2004) found that the length of the ITS region (excluding 5.8S) in selected *Sellaphora* taxa ranged from 519 to 734 bp. Furthermore, ITS and to a lesser degree 18S rDNA have the additional problem of intra-individual variation (Álvarez and Wendel 2003; Alverson and Kolnick 2005; Vanormelingen et al. 2007). Within one *S. auldreekie* isolate, Behnke et al. (2004) recorded three types of ITS sequences that differed at 48 positions and two indels of 50 and 4 bp. Cloning of PCR fragments is therefore necessary to obtain accurate measurements of intra-individual ITS variation. Consequently, we do not consider ITS to be a suitable barcode region for diatoms. Surprisingly, ITS is one of the genes currently being considered for barcoding higher plants (Kress et al. 2005); in large part, this is because of the huge number of ITS sequences already available, which is not the case for diatoms.

Sixth, it is desirable to reduce as much as possible the likelihood of amplifying contaminant DNA. Although it is relatively easy to detect carry-over of other photosynthetic organisms during isolation, small heterotrophic flagellates are less easy to exclude. In addition, under sub-optimal culture conditions, mite-mediated fungal transfer between cultures can occur: two of the *cox*1 sequences in this study, obtained from preserved DNA collected originally for *rbc*L analysis, turned out to belong to non-diatoms. Using a chloroplast marker, it is unlikely that any non-specific products will be amplified as long as the culture is unialgal, which is easy to verify by low-resolution microscopy.

So should *cox*1 be chosen as a barcode gene for diatoms? From our preliminary data, it is certainly a serious contender, because it has a higher degree of variability than any other gene for which we possess information, apart from ITS (so that sequencing effort and expenditure are lower: two sequencing reactions are required compared



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

to four for *rbc*L and up to six for 18S rDNA), and as far as we are aware there are no examples of intragenomic *cox*1 variation. On the downside, relatively little is known about the evolution of algal mitochondrial genomes. In addition, mitochondrial mode of inheritance has not yet been established (although the same is true for chloroplasts), nor have diatom-wide primers been designed. Taking these factors into account, we now use *cox*1 as a relatively quick and cheap screening method to determine whether diatom clones are of particular interest to us; if so, we obtain *rbc*L, 18S rDNA or other data to place the clone phylogenetically.

## How Widely Can the Diatom *cox*1 Barcode Primers be Used?

The *cox*1 barcode primers, GazF2 and KEdtmR yielded a product of the anticipated size for all taxa except two belonging to the *S. pupula* species complex (Table 2). The same primers were also suitable for sequencing the majority of *Sellaphora* isolates as well as the other raphids *Pinnularia* cf. *gibba* (which is fairly closely related to *Sellaphora*: Behnke et al. 2004, Evans et al. unpubl.) and *Eunotia* sp. (which is widely separated from *Sellaphora*: Sims et al. 2006). Surprisingly, *cox*1 sequences could not be obtained for *Seminavis* cf. *robusta* despite its closer relationship to *Sellaphora* than *Eunotia* sp. Outside the raphid pennates their use was limited, working partially only for the araphid pennate *Tabularia* sp. No *cox*1 sequences were obtained for centric diatoms. Additional primers will therefore need to be developed for a universal *cox*1 diatom barcoding system to be viable.

## *Cox*1 as a Phylogenetic Marker

The *cox*1 *Sellaphora* phylogeny suggests essentially the same relationships as those that resulted from analysis of *rbc*L (Fig. 2), although support values are generally lower. In both cases, *Sellaphora* clearly emerged as a monophyletic group. Whereas *S. bacillum* and *S. laevissima* are natural groups, *S. pupula* agg. is paraphyletic, because the clade that contains the demes and segregate species of *S. pupula* agg. also includes *S. bacillum.* In the *cox*1 tree (Fig. 2A), the closest relative to *S. bacillum* is the recently described *S. lanceolata* (Mann et al. 2004). In the *rbc*L phylogeny (Fig. 2B), "small lanceolate" was even more closely related to *S. bacillum*, but we were unable to obtain a *cox*1 sequence for this taxon. *Sellaphora blackfordensis* and *S. capitata* formed strongly supported monophyletic groups in both data-sets, supporting the recent recognition of these as species (Mann et al. 2004). Two taxon pairs were very well supported in the *cox*1 analysis: "small blunt-capitate"+"cf. small blunt-capitate" and *S. capitata*+"afro". *Cox*1 sequences for "small blunt-capitate"+"cf. small blunt-capitate" were identical (as were *rbc*L sequences) suggesting that despite slight morphological differences (Evans et al. unpubl.), both isolates belong to the same taxon.

The *cox*1 and *rbc*L analyses resolved a number of identical clades within *Sellaphora*, one of which was particularly well supported in both: the "*blackfordensis*" group (*S. blackfordensis, S. capitata,* "cf. obese", "afro", "pseudocapitate", "southern pseudocapitate", "cf. capitata", and "southern capitate"). The strong support for this group correlated with several *cox*1 amino-acid changes in sites 55—65 (numbering according to the *T. pseudonana* exon, GenBank NC007405), producing the unique motif RPDSDFLNYNH. Three of these (55Arg, 57Asp and 65His) represent changes in amino-acid type (acidic, basic, polar, non-polar) relative to all other *Sellaphora*. Most other non-synonymous changes among *Sellaphora cox*1 sequences involved replacement of amino-acids by others of the same charge group. In contrast, there were no obvious correlations with specific synapomorphies in the *rbc*L data-set for the *blackfordensis* group.

The major difference between the two phylogenies is in the position of *S.* cf. *minima*, which is resolved outside *S. pupula* agg. using *cox*1 and within *S. pupula* agg. in *rbc*L (Fig. 2). All *S. pupula* agg. taxa and *S. bacillum* possess transverse bars at the poles of each cell; these bars are absent in the other *Sellaphora* taxa included in this study. *Sellaphora* cf. *minima* lacks polar bars and its position in the *rbc*L phylogeny, which has low support, is likely to be an anomaly.

Overall, our analyses show that the *cox*1 barcode region, despite its short length, can be a valuable phylogenetic marker. Though obtained primarily to characterize diatom taxa for identification, *cox*1 barcode sequences can provide enough phylogenetic information to indicate when more extensive studies, using multiple longer sequences, are likely to be rewarding.



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

## Methods

**Choice of isolates:** We used an extensive *rbc*L data-set (unpubl.) as a basis for selecting *Sellaphora* isolates for *cox*1 analysis. *Rbc*L data were available for isolates belonging to 23 *Sellaphora* taxa (19 *S. pupula* (Kütz.) Mereschk. agg. demes or species and four other *Sellaphora* species; see Table 1 for GenBank accessions); 18S and ITS rDNA data-sets are currently much more limited (Behnke et al. 2004; unpubl.). One isolate from each of the 23 *Sellaphora* taxa was chosen unless (a) isolates originated from more than one water body (*S. bacillum* (Ehrenb.) D. G. Mann, *S. blackfordensis* D. G. Mann et S. Droop, and *S. capitata* D. G. Mann et S. M. McDonald) and/or (b) molecular data (Behnke et al. 2004, unpubl.) suggested they harbour cryptic diversity (*S. laevissima* (Kütz.) D. G. Mann and "elliptical"). Although *rbc*L data suggested that isolates of "small blunt capitate" (THR13) and "cf. small blunt capitate" (THR8; Table 1) may belong to the same species (despite displaying morphological differences: unpubl.), both were included in this study to test for a similar lack of *cox*1 divergence.

To allow a diatom-wide assessment of the effectiveness of different primer sets in amplifying the *cox*1 region, seven non-*Sellaphora* taxa were included that represented other major evolutionary lineages of diatoms (Round et al. 1990). Non-*Sellaphora* raphid pennates were represented by *Eunotia* sp., *Seminavis* cf. *robusta* Danielidis et D. G. Mann and *Pinnularia* cf. *gibba* Krammer; araphid pennates by *Tabularia* sp.; and centrics by *Paralia sulcata* (Ehrenb.) Cleve, *Melosira nummuloides* (Dillw.) Ag. and *Odontella sinensis* (Grev.) Grun.

**Isolation, culture and identification of isolates:** *Sellaphora* and freshwater out-group taxa were obtained from epipelic communities growing on mud at 0.5—1 m depth in a range of freshwaters; marine outgroup taxa were obtained from various net plankton samples (Table 1; clones available on request). Epipelic communities were harvested as described by Mann and Chepurnov (2005). Clones were isolated either by pipette or by streaking on agar (freshwater taxa only). Cells were then transferred into small volumes of medium in the wells of Repli dishes and grown as monoclonal cultures in 50 mm diameter Petri dishes. Freshwater species were kept in WC medium with silicate (Guillard and Lorenzen 1972), marine species in Roshchin medium (Mann and Chepurnov 2004). Stock cultures were kept at 15 °C on a 12:12 h L—D cycle at low irradiance (*c.* 5—20 μmol photons $m^{-2}$ $s^{-1}$). For molecular studies, two 90 mm Petri dishes were inoculated, one for DNA extraction, the other to provide cells for a voucher slide. Two clones were isolated from among the $F_1$ progenies of two sexual crosses: (1) between a North American clone (L845: supplied by the late Dr. D. B. Czarnecki) and a Scottish clone (Thr 42) of "pseudocapitate"; and (2) between two clones of *S.* cf. *robusta* (75 and 88) isolated from Veerse Meer, Zeeland, The Netherlands (Chepurnov et al. 2002). Mating experiments were performed as described by Mann et al. (1999).

The preparation of voucher slides and compilation of digital images for identification purposes are described in more detail in the Supplementary Material. Identities were assigned to clones of the *S. pupula* complex (Table 1) on the basis of their morphological similarity to clones and demes studied previously (e.g. Behnke et al. 2004; Mann et al. 1999, 2004; Taylor 1996).

**DNA extraction, PCR, sequencing and *cox*1 primer design:** DNA was extracted using one of three methods, two of which are described in Sluiman and Guihal (1999) and Jones et al. (2005); for a few samples, a Qiagen Plant DNeasy kit (Qiagen, Crawley, UK) was used. *Rbc*L, ITS and 18S rDNA regions were amplified and sequenced as described in Jones et al. (2005) and Behnke et al (2004). The respective lengths of these regions are approximately 1400 bp, 650—850 bp and 1800 bp.

For *cox*1, red algal (GazF1 and GazR1; Saunders 2005) and diatom (pC1 and pB1; Ehara et al. 2000a) *cox*1 primers were tested, predominantly using *Sellaphora* DNAs (AUS1, BAL1, BLA2, BLA16, DUN6, F1, BEL3, RBG1, THR2 and THR7), but also *P.* cf. *gibba, P. sulcata* and *S.* cf. *robusta* (Table 1). GazF1 and GazR1 were designed by Saunders to amplify a 710 bp *cox*1 barcode in red algae, whereas pC1 and pB1 were used to amplify 1100 bp of diatom *cox*1 (also from near the 5′ end of the gene) to construct a limited phylogeny. PCR conditions were as detailed in Saunders (2005) and Ehara et al. (2000a). Amplification products were purified using a GFX PCR purification kit (Amersham Biosciences, Little Chalfont, UK) and subsequent concentrations estimated by electrophoresis on a 1% agarose gel. PCR fragments were sequenced directly in 10 μl volumes using 1 μM of PCR primer, 2 μl of dye terminator cycle sequencing reagents and 1.5 μl of sequencing reaction buffer (Beckman Coulter, High Wycombe, UK). Sequencing PCR conditions were 35 cycles of 96 °C for 20 s, 50 °C



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

for 20 s and 60 °C for 4 min. Excess dye-labelled nucleotides were removed by standard ethanol—-sodium acetate precipitation. Sequence products were re-suspended in 35 μl of sample loading solution (Beckman Coulter) and run on a CEQ8000 DNA sequencer (Beckman Coulter).

Sequences were edited using CEQ software (Beckman Coulter) and aligned in Sequencher 4.5 (Gene Codes Corporation, Ann Arbor, MI, USA). Also included in the alignment were the ten publicly available (GenBank) diatom *cox*1 sequences: *Thalassiosira pseudonana* Hasle & Heimdal (NC007405, the focus of the diatom genome sequencing project; Armbrust et al. 2004), *T. nordenskioeldii* Cleve (AB020229), *Aulacoseira ambigua* (Grun. in Van Heurck) Simonsen (AB009418), *Skeletonema costatum* (Grev.) Cleve (AB020227), *Ditylum brightwellii* (T. West) Grun. (AB020223), *Rhizosolenia setigera* Brightwell (AB020226), *Fragilaria striatula* Lyngbye (AB020224), *Thalassionema nitzschioides* (Grun.) Grun. ex Hustedt (AB020228), *Cylindrotheca closterium* (Ehrenb.) Reimann & Lewin (AB020222) and *Nitzschia frustulum* (Kütz.) Grun. in Cleve & Grun. (AB020225).

PRIMER3 (Rozen and Skaletsky 2000) was used to design additional primers, primarily focusing on *Sellaphora*. These were tested in combination with brown algal *cox*1 barcode primers, GazF2 and GazR2 (Table 3; these unpublished primer sequences were kindly provided by Prof. Gary Saunders subsequent to the initial tests). Reaction volumes were 50 μl and contained 2 μl of DNA template, 1 × $NH_4$ buffer, 2.5 mM $MgCl_2$, 200 μM of dNTPs, 0.3 μM of each primer and 1.25 units of *Taq* (Bioline, London, UK). PCR conditions were 95 °C for 3 min; 35 cycles of 95 °C for 30 s, 50 °C for 1 min and 72 °C for 1.5 min; and, finally, 72 °C for 5 min. PCR purifications, sequencing, editing and aligning were performed as detailed above. GenBank accessions are provided in Table 1.

**Assessing levels of intra- and interspecific *cox*1 and *rbc*L divergence:** Intra- and interspecific thresholds of *cox*1 and *rbc*L divergences were estimated by calculating the number of bp differences between *Sellaphora* isolates belonging to the same taxa and those belonging to selected closely related taxa (at least 60% MP BS in either the *cox*1 or *rbc*L phylogenetic analyses).

***Cox*1 and *rbc*L phylogenetic analyses:** *Cox*1 and *rbc*L phylogenies were constructed using exactly the same *Sellaphora* isolates except for two *S. capitata* clones (*rbc*L data were from BEL1 and BLA10: GenBank accessions EF143301 and EF143295, respectively) and one *S. blackfordensis* clone (*rbc*L data were from BLA6: EF143282); these clones were nevertheless from the same locations. Two of the three outgroup taxa used in the *rbc*L analysis differed from those used in the *cox*1 analysis because they were chosen with different criteria in mind (*Fallacia* cf. *forcipata* (Grev.) Stickle et Mann and *Rossia* sp.: EF143289 and EF143281, respectively).

MP and ML analyses were carried out using PAUP* 4.0b10 (Swofford 2002). MP analyses implemented two heuristic search strategies. Search strategy I involved 1000 replicates, saving all trees and using tree-bisection-reconnection (TBR) branch-swapping, with all characters equally weighted and gaps treated as missing data; starting trees were obtained via random stepwise addition. Strategy II involved 10,000 replicates, saving only two trees at each step, to check for the presence of islands of trees not found in the first search. Parsimony BS values were calculated with 100 replicates of a 1000-replicate heuristic search, saving all trees, with TBR branch-swapping. ML analyses were conducted using models selected by the HLR test in ModelTest (Posada and Crandall 1998). Full heuristic ML searches of 100 replicates were carried out, with the molecular clock not enforced, using TBR branch-swapping, saving two trees per replicate. JK values were calculated from a 100-replicate ML analysis, resampling 50% of characters, with starting trees obtained by neighbour-joining.

## Acknowledgements

We thank Prof. Gary Saunders for very kindly providing us with the brown algal *cox*1 barcode primer sequences, Dr. Michelle Hollingsworth for advice on molecular work, Dr. Pete Hollingsworth for fruitful discussions, and Prof. Wim Vyverman, Jeroen Van Wichelen, Dr. Wiebe Kooistra, Prof. A. Poulíčková, and especially Dr. Victor Chepurnov, for providing material and for isolating clones. Lastly, we thank two anonymous reviewers for their constructive comments. This work was supported by a Natural Environment Research Council Fellowship (NE/C518373/1) to KME.

## Appendix A. Supplementary Material

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.protis.2007.04.001.

## References

**Álvarez I, Wendel JF** (2003) Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenet Evol **29**: 417—434

**Alverson AJ, Kolnick L** (2005) Intragenomic nucleotide polymorphism among small subunit (18S) rDNA paralogs in the diatom genus *Skeletonema* (Bacillariophyta). J Phycol **41**: 1248—1257

**Amato A, Kooistra WHCF, Levialdi Ghiron JH, Mann DG, Pröschold T, Montresor M** (2007) Reproductive isolation among sympatric cryptic species in marine diatoms. Protist **158**: 193—207

**Aoyama H, Hagiwara Y, Misumi O, Kuroiwa T, Nakamura S** (2006) Complete elimination of maternal mitochondrial DNA during meiosis resulting in the paternal inheritance of the mitochondrial genome in *Chlamydomonas* species. Protoplasma **228**: 231—242

**Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kröger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS** (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution and metabolism. Science **306**: 79—86

**Beagley CT, Okada NA, Wolstenholme DR** (1996) Two mitochondrial group I introns in a metazoan, the sea anemone *Metridium senile*: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. Proc Natl Acad Sci USA **93**: 5619—5623

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

**Bergthorsson U, Adams KL, Thomason B, Palmer JD** (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature **424**: 197—201

**Beszteri B, Ács E, Medlin LK** (2005) Ribosomal DNA sequence variation among sympatric strains of the *Cyclotella meneghiniana* complex (Bacillariophyceae) reveals cryptic diversity. Protist **156**: 317—333

**Blaxter M, Elsworth B, Daub J** (2004) DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. Proc R Soc Lond B **271**: S189—S192

**Brodie J, Hayes PK, Barker GL, Irvine LM** (1996) Molecular and morphological characters distinguishing two *Porphyra* species (Rhodophyta: Bangiophycidae). Eur J Phycol **31**: 303—308

**Chase MW, Salamin N, Wilkinson M, Dunwell JM, Prasad Kesanakurthi R, Haidar N, Savolainen V** (2005) Land plants and DNA barcodes: short-term and long-term goals. Phil Trans R Soc Lond B **360**: 1889—1895

**Chepurnov VA, Mann DG, Sabbe K, Vyverman W** (2004) Experimental studies on sexual reproduction in diatoms. Int Rev Cytol **237**: 91—154

**Chepurnov VA, Mann DG, Vyverman W, Sabbe K, Danielidis DB** (2002) Sexual reproduction, mating system and protoplast dynamics of *Seminavis* (Bacillariophyceae). J Phycol **38**: 1004—1019

**Cowan RS, Chase MW, Kress WJ, Savolainen V** (2006) 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. Taxon **55**: 611—616

**Coyer JA, Peters AF, Hoarau G, Stam WT, Olsen JL** (2002) Inheritance patterns of ITS1, chloroplasts and mitochondria in artificial hybrids of the seaweeds *Fucus serratus* and *F. evanescens* (Phaeophyceae). Eur J Phycol **37**: 173—178

**Ehara M, Inagaki Y, Watanabe KI, Ohama T** (2000a) Phylogenetic analysis of diatom *cox*1 genes and implications of a fluctuating GC content on mitochondrial genetic code evolution. Curr Genet **37**: 29—33

**Ehara M, Watanabe KI, Ohama T** (2000b) Distribution of cognates of group II introns detected in mitochondrial *cox*1 genes of a diatom and a haptophyte. Gene **256**: 157—167

**Finlay BJ, Monaghan EB, Maberly SC** (2002) Hypothesis: the rate and scale of dispersal of freshwater diatom species is a function of their global abundance. Protist **153**: 261—273

**Fontaine J-M, Goux D, Kloareg B, Loiseaux-de Goër S** (1997) The reverse-transcriptase-like proteins encoded by group II introns in the mitochondrial genome of the brown alga *Pylaiella littoralis* belong to two different lineages which apparently coevolved with the group II ribosyme lineages. J Mol Evol **44**: 33—42

**Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM** (2006) Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. Mol Ecol **15**: 1759—1768

**Gray MW, Lang BF, Burger G** (2004) Mitochondria of protists. Annu Rev Genet **38**: 477—524

**Guillard RRL, Lorenzen CL** (1972) Yellow-green algae with chlorophyllide c. J Phycol **8**: 10—14

**Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN** (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci USA **103**: 968—971

**Hebert PDN, Cywinska A, Ball SL, deWaard JR** (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B **270**: 313—321

**Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM** (2004) Identification of birds through DNA barcodes. PLoS Biology **2**: 1657—1663

**Hughes Martiny JB, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Adams Krumins J, Kuske CR, Morin PJ, Naeem S, Øvreås L, Reysenbach A-L, Smith VH, Staley JT** (2006) Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol **4**: 102—112

**Hurst GDD, Jiggins FM** (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc R Soc Lond B **272**: 1525—1534



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

**Jensen KJ, Moestrup Ø, Schmid A-MM** (2003) Ultrastructure of the male gametes from two centric diatoms, *Chaetoceros laciniosus* and *Coscinodiscus wailesii* (Bacillariophyceae). Phycologia **42**: 98—105

**Jones HM, Simpson GE, Stickle AJ, Mann DG** (2005) Life history and systematics of *Petroneis* (Bacillariophyta), with special reference to British waters. Eur J Phycol **40**: 61—87

**Kato Y, Kogame K, Nagasato C, Motomura T** (2006) Inheritance of mitochondrial and chloroplast genomes in the isogamous brown alga *Scytosiphon lomentaria* (Phaeophyceae). Phycol Res **54**: 65—71

**Knoop V** (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet **46**: 123—139

**Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH** (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA **102**: 8369—8374

**Lundholm N, Moestrup Ø, Kotaki Y, Hoef-Emden K, Scholin C, Miller P** (2006) Inter- and intraspecific variation of the *Pseudo-nitzschia delicatissima* complex (Bacillariophyceae) illustrated by rRNA probes, morphological data and phylogenetic analyses. J Phycol **42**: 464—481

**Mann DG** (1984) Observations on copulation in *Navicula pupula* and *Amphora ovalis* in relation to the nature of diatom species. Ann Bot **54**: 429—438

**Mann DG** (1989) The species concept in diatoms: evidence for morphologically distinct, sympatric gamodemes in four epipelic species. Plant Syst Evol **164**: 215—237

**Mann DG** (1996) Chloroplast Morphology, Movements and Inheritance in Diatoms. In Chaudhary BR, Agrawal SB (eds) Cytology, Genetics and Molecular Biology of Algae. SPB Academic Publishing, Amsterdam, pp 249—274

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

**Mann DG, Chepurnov VA** (2004) What have the Romans ever done for us? The past and future contribution of culture studies to diatom systematics. Nova Hedwigia **79**: 237—291

**Mann DG, Chepurnov VA** (2005) Auxosporulation, mating system, and reproductive isolation in *Neidium* (Bacillariophyta). Phycologia **44**: 335—350

**Mann DG, Droop SJM** (1996) Biodiversity, biogeography and conservation of diatoms. Hydrobiologia **336**: 19—32

**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**: 152—170

**Mann DG, McDonald SM, Bayer MM, Droop SJM, Chepurnov VA, Loke RE, Ciobanu A, Hans du Buf JM** (2004) The *Sellaphora pupula* species complex (Bacillariophyceae): morphometric analysis, ultrastructure and mating data provide evidence for five new species. Phycologia **43**: 459—482

**Mayr E** (1942) Systematics and the Origin of Species. Columbia University Press, New York

**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

**Meyer CP, Paulay G** (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biology **3**: 2229—2238

**Newmaster SG, Fazekas AJ, Ragupathy S** (2006) DNA barcoding in land plants: evaluation of *rbc*L in a multigene tiered approach. Can J Bot **84**: 335—341

**Nozaki H, Takahara M, Nakazawa A, Kita Y, Yamada T, Takano H, Kawano S, Kato M** (2002) Evolution of *rbc*L group IA introns and intron open reading frames within the colonial Volvocales (Chlorophyceae). Mol Phylogenet Evol **23**: 326—338

**Orsini L, Procaccini G, Sarno D, Montresor M** (2004) Multiple rDNA ITS-types within the diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae) and their relative abundances across a spring bloom in the Gulf of Naples. Mar Ecol Progr Ser **271**: 87—98

**Oudot-le Secq MP, Kloareg B, Loiseaux-de Goër S** (2002) The mitochondrial genome of the brown alga *Laminaria digitata*: a comparative analysis. Eur J Phycol **37**: 163—172

**Oudot-Le Secq M-P, Loiseaux-de Goër S, Stam WT, Olsen JL** (2006) Complete mitochondrial genomes of the three brown algae (Heterokonta: Phaeophyceae) *Dictyota dichotoma, Fucus vesiculosus* and *Desmarestia viridis*. Curr Genet **49**: 47—58

**Pace NR** (1997) A molecular view of microbial diversity and the biosphere. Science **276**: 734—740

**Pedrós-Alió C** (2006) Marine microbial diversity: can it be determined? Trends Microbiol **14**: 257—263

**Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP** (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst Biol **55**: 595—609

**Posada D, Crandall KA** (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics **14**: 817—818

**Robba L, Russell SJ, Barker GL, Brodie J** (2006) Assessing the use of the mitochondrial *cox*1 marker for use in DNA barcoding of red algae (Rhodophyta). Am J Bot **93**: 1101—1108

**Round FE, Crawford RM, Mann DG** (1990) The Diatoms. Cambridge University Press, Cambridge

**Rozen S, Skaletsky HJ** (2000) PRIMER3 on the WWW for General Users and for Biologist Programmers. In Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, pp 365—386

**Rubinoff D** (2006) Utility of mitochondrial DNA barcodes in species conservation. Conservat Biol **20**: 1026—1033

**Sánchez Puerta MV, Bachvaroff R, Delwiche CF** (2004) The complete mitochondrial genome sequence of the haptophyte *Emiliania huxleyi* and its relation to Heterokonts. DNA Res **11**: 1—10

**Sarno D, Kooistra WHCF, Medlin LK, Percopo I, Zingone A** (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of *S. costatum*-like species



Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001

with the description of four new species. J Phycol **41**: 151—176

**Saunders GW** (2005) Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future application. Phil Trans R Soc Lond B **360**: 1879—1888

**Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R** (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Phil Trans R Soc Lond B **360**: 1805—1811

**Seifert KA, Samson RA, deWaard JR, Houbraken J, Lévesque CA, Moncalvo J-M, Louis-Seize G, Hebert PDN** (2007) Prospects for fungus identification using *CO1* DNA barcodes, with *Penicillium* as a test case. Proc Natl Acad Sci USA **104**: 3901—3906

**Sims PA, Mann DG, Medlin LK** (2006) Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia **45**: 361—402

**Šlapeta J, Moreira D, López-García P** (2005) The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. Proc R Soc Lond B **272**: 2073—2081

**Sluiman HJ, Guihal C** (1999) Phylogenetic position of *Chaetosphaeridium* (Chlorophyta), a basal lineage in the Charophyceae, inferred from 18S rDNA sequences. J Phycol **35**: 395—402

**Swofford DL** (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). 4.0b10. Sinauer, Sunderland

**Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC** (2000) Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol **31**: 21—32

**Taylor NG** (1996) Variation Within *Sellaphora pupula* (Bacillariophyta) in Relation to the Species Concept. PhD Thesis, University of Edinburgh

**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

**Watanabe KI, Ehara M, Inagaki Y, Ohama T** (1998) Distinctive origins of group I introns found in the *cox*1 genes of three green algae. Gene **213**: 1—7

**Witt JDS, Threloff DL, Hebert PDN** (2006) DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Mol Ecol **15**: 3073—3082





Please cite this article as: Evans K M, et al. An Assessment of Potential Diatom "Barcode" Genes (*cox*1, *rbc*L, 18S and ITS rDNA) and their Effectiveness in Determining Relationships in *Sellaphora* (Bacillariophyta), Protist (2007), doi:10.1016/j.protis.2007.04.001