RESEARCH NOTE

A proposed protocol for nomenclaturally effective DNA barcoding of microalgae

KATHARINE M. EVANS AND DAVID G. MANN*
Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5LR, United Kingdom


A mechanism for giving DNA barcodes nomenclatural status in microalgae via culture-derived epitypes is demonstrated with reference to four species of the freshwater diatom Sellaphora. A fifth, Sellaphora capitata, is described as a new species to illustrate application of barcoding via a holotype. Using cox1 barcodes, it is shown that S. capitata has a worldwide distribution, consistent with the ubiquitous dispersal hypothesis.

KEY WORDS: Biogeography, cox1, Diatom, DNA barcoding, New species, Molecular typification, Nomenclature, Sellaphora, Ubiquitous dispersal hypothesis

INTRODUCTION

In this paper, we demonstrate how DNA barcodes can be made nomenclaturally effective for existing species of microalgae or for new species that cannot be preserved as living or metabolically inactive cultures. The basic problem, which has been identified previously (e.g. Hoef-Emden et al. 2007), is to reconcile barcoding with conventional nomenclature to provide a binding molecular typification. Our explanation and examples concern diatoms because barcoding is already feasible in this group and a nomenclatural mechanism is urgent, but the principles are general.

Those who use diatoms as bioindicators or for biological research need accurate taxonomy and therefore encounter three problems: (1) the diversity of diatoms is unknown but undoubtedly much greater than ever envisaged (Mann & Droop 1996), (2) identification becomes ever more difficult as existing taxonomic aids are rendered obsolete by new discoveries, and (3) contrary to the long-held belief that all taxa are cosmopolitan, species (Kooistra et al. 2008) and even genera (Vyverman et al. 1998, 2007) can possess biogeographies. The effects of inadequate taxonomy can be considerable (cf. Darling et al. 2004) but may not become obvious until after it has become impossible to rework the original material. Traditionally, diatom taxa have been characterized and identified by aspects of valve morphology, but exclusive reliance on these traits is an insufficient basis for future work because cryptic diversity is too great (Sarno et al. 2005; Amato et al. 2007; Vanormelingen et al. 2008), the extent of phenotypic plasticity is usually unknown, and describing and communicating slight morphological differences is very difficult, even among experienced taxonomists (cf. Kelly et al. 2002).

The freshwater morphospecies Sellaphora pupula (Kützing) Mereschkowsky sensu lato exemplifies current difficulties. Within the United Kingdom alone, there are at least 36 subtly different species within this single morphospecies (Mann et al. 2004, 2008), and molecular phylogenetic studies (18S rDNA, rbcL and cox1) have demonstrated the prevalence of morphological homoplasy and paraphyly or polyphly of previous taxa (Evans et al. 2007, 2008). Even after 25 years of studying Sellaphora, one of us (DGM) is not confident that he can identify many Sellaphora species without extra, nonmorphological information, such as provenance (which presupposes an existing, accurate catalogue of diversity) or the results of mating experiments (which are obviously impractical for routine identification). Such cases are not uncommon and are prime candidates for the development of DNA barcoding (Hebert et al. 2003). We recently demonstrated the power of barcoding in Sellaphora (Evans et al. 2007). We trialed four candidate genes and, in line with animals (Hajibabaei et al. 2006) and red algae (Saunders 2005; Robba et al. 2006), proposed part of the mitochondrion-encoded cytochrome oxidase I (cox1) as the gene of choice because it is short, variable, easy to align, and also a useful phylogenetic marker in combination with other genes. We now routinely use cox1 barcoding to identify Sellaphora clones and discover new species and, in collaboration with Caroline Souffreau (University of Ghent) and Dr Rosa Trobajo (IRTA, Catalunya), have extended barcoding to Pinnularia and Nitzschia, with promising results (unpublished).

In a recent survey of British Sellaphora, Mann et al. (2008) specified GenBank cox1 sequences that might act as barcodes for five formally recognized Sellaphora species and eight informally named genodemes (putative species). The aim was to provide unambiguous reference data for identification. However, in order for barcodes to have real value, it is essential to find a way to give them formal nomenclatural status, as ‘molecular types’, and this was not achieved in our earlier paper (Mann et al. 2008). There are two obstacles: (1) there is no provision in the International Code of Botanical Nomenclature (McNeill et al. 2006) for designating gene sequences as types, and (2) existing
diatom types are almost invariably cleaned preparations of the silica cell walls. By contrast, type specimens of macroalgae sometimes contain DNA that is sufficiently intact to allow use in phylogeny (e.g. Hughey et al. 2002) and hence potentially also for barcoding. Cultures of algae are acceptable as types if kept metabolically inactive, such as by cryopreservation (McNeill et al. 2006, Article 8.4), but this is not yet possible for most diatoms, including Sellaphora (Dr V.A. Chepurnov and O. Chepurnova, personal communication), and it may never be. Furthermore, because of the obligate link between size restitution and sexual reproduction in most diatoms, living cultures are particularly liable to genetic alteration with time, and the high frequency of heterothallism in the pennates (which are by far the most speciose group) means that many species, including those dealt with here, cannot be maintained as monoclonal cultures beyond a few months or years.

We will employ five species of Sellaphora as worked examples. The mechanism we use to give a barcode nomenclatural effect is via the holotype, if the species is described for the first time, or via an epitype, that is, ‘a specimen or illustration selected to serve as an interpretative type when the holotype, lectotype, or previously designated neotype, or all original material associated with a validly published name, is demonstrably ambiguous and cannot be critically identified for purposes of the precise application of the name of a taxon’ (McNeill et al. 2006, Article 9.7). Four Sellaphora clones from Blackford Pond, Edinburgh, and one from Loch Leven, Perthshire, were isolated, vouched, analysed, and photographed as described in Evans et al. (2007, 2008) to provide types. All slide preparations are kept in the diatom herbarium of the Royal Botanic Garden Edinburgh (E).

For specimen records to gain formal BARCODE status in global sequence databases, including GenBank, seven stringent requirements are in place: (1) a species name (although it can be interim), (2) voucher data (catalogue number and institution at which the voucher is stored), (3) collection record (collector, collection date, and location with GPS coordinates), (4) identifier of the specimen, (5) cox1 sequence of at least 500 base pairs (bp), (6) PCR primers used to generate the amplicon, and (7) deposition of accompanying, overlapping electropherograms (Ratnasingham & Hebert 2007). To adhere to the seventh of these requirements, cox1 regions of the five type specimens were sequenced using superior chemistry to that employed previously (Evans et al. 2007). The 665-bp diatom cox1 region was amplified using primers GazF2 and KEdtmR as detailed in Evans et al. (2007), and PCR products were purified using ExoSAP-IT (USB Corporation). Sequencing was conducted in 10-µl volumes using 0.32 µM of PCR primer, 1 µl of BigDye v3.1, and 2 µl of sequencing reaction buffer (Applied Biosystems). Sequencing PCR conditions were 25 cycles of 95°C for 30 s, 50°C for 20 s, and 60°C for 4 min. Excess dye-labelled nucleotides were removed using the Performa DTR V3 cleanup system (EdgeBio), and sequence products were run on an ABI 3730 DNA sequencer (Applied Biosystems). Forward and reverse reads were edited and aligned using Sequencher 4.5 (GeneCodes Corporation). Full 665-bp reads were obtained for all type specimens; GenBank BARCODE accessions are listed with each species. To document the distributions of the five species as an example of barcode use, all available cox1 data for these five species were uploaded to GenBank as standard submissions.

Specimens from four clones are illustrated in Figs 1–9 and agree morphologically with previous descriptions of S. pupula (Mann 2001; Mann et al. 2004), S. blackfordensis D.G. Mann & S. Droop and S. capitata D.G. Mann & S.M. McDonald (Mann et al. 2004), and S. caput K.M. Evans & D.G. Mann sp. nov. (Mann et al. 2008, as the ‘caput’ deme). Accordingly, these clones are designated as epitypes or (for S. caput) the holotype, the type specimens taking the form of preserved cell walls and DNA, with associated information (cox1 sequences); the fifth species, S. bacillum (Ehrenberg) D.G. Mann, was illustrated by Jahn et al. (2008).

**Sellaphora pupula**

**EPITYPE DESIGNATED HERE:** Material of clone BLA21 (= SEL 721B), as preserved on slide E4215 (E), illustrated in Fig. 1, barcoded in GenBank accession FJ147204, and with preserved DNA at (E) as EDNA 08-01132. The epitype is selected to clarify the nature of the lectotype specimen at England Finder reference M45/2 on slide BM 17918 (Natural History Museum, London), which was designated and illustrated by Mann (2001, figs 2–6); this epitype supplements and does not replace the previous epitype illustrations and specimen designated by Mann et al. (2004) and available online at http://rbg-web2.rbge.org.uk/algae/research/types/Sellaphora_pupula_type.html.

Clone BLA21 (isolated 15 January 2008) and the previous epitypes were all derived from epipelon of soft mud at the SW end of Blackford Pond, Edinburgh (55°55′29″N, 3°11′49″W; UK National Grid Reference NT 253709).

**Sellaphora blackfordensis**

**EPITYPE DESIGNATED HERE:** Material of clone BLA17 (= BI 85), as preserved on slide E4701 (E), illustrated in Fig. 2, barcoded in GenBank accession EF164948, and with preserved DNA at (E) as EDNA 06-04760. The epitype is selected to clarify the nature of the holotype on slide E16/4 at England Finder reference S39/4, which was designated and illustrated by Mann et al. (2004, fig. 19) and is shown online at http://rbg-web2.rbge.org.uk/algae/research/types/Sellaphora_blackfordensis_type.html.

Clone BLA17 (isolated 5 October 2004) and the holotype were both derived from epipelon of soft mud at the SW end of Blackford Pond, Edinburgh (location as described previously).

**Sellaphora capitata**

**EPITYPE DESIGNATED HERE:** Material of clone BLA18 (= BI 89), as preserved on slide E3608 (E), illustrated in Fig. 3, barcoded in GenBank accession EF164947, and with preserved DNA at (E) as EDNA 06-04765. The epitype is selected to clarify the nature of the holotype on slide E16/4 at England Finder reference R41/0, which was designated and illustrated by Mann et al. (2004, fig. 20) and is shown
Sellaphora bacillum

In a previous paper (Jahn et al. 2008), we applied the same approach as adopted here to specify a ‘type’ cox1 sequence for S. bacillum (Ehrenberg) D.G. Mann, but we did not fulfill the requirements for it to be designated an official BARCODE. We have now upgraded the sequence and record (EF164941) to official quality and status.

Sellaphora capitata K.M. Evans & D.G. Mann, sp. nov.


Valves linear-elliptical with broad capitulate poles, becoming subcapitate or rostrate in small specimens, 18–38 × 6.25–7.3 μm (decreasing beyond 18 to c. 11 μm in old cultures). Striae radiate, slightly curved, with shorter ones intercalated at the centre, 22.5–26 (usually 23–25) in 10 μm. Areolae invisible in LM. Axial area very narrow. Central area neither ornamented nor ridged, transversely rectangular to ± bow-tie-shaped. Polar bars parallel to slightly radiate.

ETYMOLOGY OF THE EPITHET: Though originally inspired by the Latin word for ‘head’, the epithet capitata is to be regarded as an indeclinable noun in apposition, to maintain maximum compatibility with the catalogue of Sellaphora diversity by Mann et al. (2008), who provide further illustrations and information about the species.

DISTRIBUTIONS: Using the barcode sequences as references, we have confirmed the presence of S. capitata in Blackford Pond and Dunsapie Loch (55°56′43.2″N, 3°09′12.5″W; GenBank accession FJ042923), Scotland; S. blackfordensis in Blackford Pond, Dunsapie Loch (EF164949), the Royal Botanic Garden Edinburgh pond (55°57′55.1″N, 3°12′20.8″W; FJ042934), Inverleith Pond (55°57′40.9″N, 3°13′4.6″W; FJ042936), Figgate Loch (55°57′02.3″N, 3°7′29.9″W; FJ042935), Balerno Millgate Pond (55°52′33.4″N, 3°19′58.0″W; FJ042937), and Balgavies Loch (56°38′49.4″N, 2°45′53.6″W; EF164935), Scotland, and Malham Tarn (54°5′53.9″N, 2°9′28.6″W; FJ042927), England; S. capitata in Blackford Pond, Dunsapie Loch (FJ042902), and Loch Leven (FJ042904), Scotland; Malham Tarn (FJ042905), England; Merelbeke
Pond (51°0′45.6″N, 3°44′49.8″W; FJ042907), Belgium; and Lake Purrumbete (38°17′1″S, 143°12′55″E; FJ042906), Australia; S. bacillum in Blackford Pond and St Margaret’s Loch (55°57′10.6″N, 3°09′37.0″W; EF164930), Scotland, and Malham Tarn (FJ042924) and a pond near Ashford-in-the-Water (53°13′50.5″N, 1°42′7.7″W; FJ042925), England; and S. caput in Loch Tulla (56°33′12.9″N, 4°45′2.6″W; FJ042926) and Loch Leven, Scotland. In no case was coax1 divergence greater than 1% between isolates regarded as the same species; in most cases it was much less. All these records are backed by voucher slides in E.

Deposit of vouchered sequences in GenBank provides a secure basis for biogeographical and ecological analysis. Thus, clear demonstration of S. caput in Loch Leven and Loch Tulla goes against the morphology-based survey of Mann et al. (2008) by showing that the species can occur in eutrophic as well as dystrophic habitats. The S. capitata records represent clear evidence, independent of morphology-based identification, that a freshwater diatom species can have a worldwide distribution, as required for the ubiquitous dispersal hypothesis (e.g. Finlay 2002); although, it remains to be established if this is natural or the result of recent accidental introduction by humans.

**DISCUSSION**

These examples illustrate how barcodes can be used to stabilize taxonomy by giving them nomenclatural status via holotypes or epitypes. An alternative would have been to simply mention a gene sequence in a new or emended description. We rejected this because the function of a description is to circumscribe a taxon, that is, to specify the range of variation among individuals that a particular taxonomist decides should be included within a particular species, variety, and so on. Not surprisingly, the circumscriptiopn, and hence the description, usually changes as knowledge increases. Like any fast-evolving part of the genome, the coax1 barcode region can also be expected to vary within species, and indeed, it does vary within several of the species we treat here. A description of these species would therefore have to list several coax1 sequences in order to circumscribe them, and the list would grow or shrink according to subsequent taxonomic research. Typification, on the other hand, has little to say about circumscription, serving instead to determine what name should be applied to each group of individuals; the barcode’s value is to improve typification, as an unambiguous molecular reference point.

Specification of a barcode as part of the holotype (as for S. caput) seems straightforward. Our proposal to add a barcode to an existing taxon via an epitype clone and its preserved remnants, and associated information requires extra steps beyond those required for a new species but no more than would be needed if the epitype were a photograph. For example, it must be true that the epitype and the holotype or lectotype it supports are the same taxonomically; if not, the types must be redefined (McNeill et al. 2006, Article 9.18). Hence, morphological agreement between the epitype material and the lectotype or holotype must be as exact as the lectotype or holotype will allow. In the case of two or more species that differ scarcely or not at all in their morphology and that have previously been confused with each other, the choice of barcoded epitype will often be arbitrary. The provenance of the original material may provide guidance, but we consider that it would be inadvisable to insist on geographical or ecological congruence between epitype and lecto- or holotype in every case because pre-barcode identifications will be inherently ambiguous. For extinct species, of course, our proposals offer no aid.

**ACKNOWLEDGEMENTS**

This work was supported by a Natural Environment Research Council Fellowship (NE/C518373/1) to KME.

**REFERENCES**


*Received 25 August 2008; accepted 23 September 2008*  
Associate editor: Alan Millar