

'Araphid' diatom classification and the 'absolute standard'

DAVID M. WILLIAMS

Department of Botany, The Natural History Museum, Cromwell Road,
London SW7 5BD, United Kingdom

'Araphid' diatom classification is discussed from the point of view of an 'absolute standard' for taxonomic rank. The 'absolute standard' is the phylogenetic tree, its nodes, the included monophyletic groups and sub-groups. To illustrate this point a few species from the genus *Licmophora* are re-analysed and the resulting phylogenetic tree is discussed in terms of a possible classification, the groups and sub-groups and their ranks.

Keywords: Diatom, araphid, classification, phylogeny

Introduction

At the 18th International Diatom Symposium I discussed diatom phylogenetic trees and their interpretation. Some of my presentation appeared in a recent book on the classification of 'algae' (WILLIAMS 2007) and in a shorter review paper (WILLIAMS and KOCIOLEK 2007). Since that time a few more papers have appeared on 'araphid' diatom phylogeny, particularly from a molecular perspective (SATO et al. 2008a, 2008b, MEDLIN et al. 2008a, 2008b). A contribution of some significance is MEDLIN et al. (2008b), as it presents a large quantity of data and proposes one of the first trees of relationships to include a good number of the diverse 'araphid' diatoms. In this short note I want to add a few more comments on the relationship between phylogenies (branching diagrams) and classification with respect to 'araphid' diatoms.

The purpose of this paper is not to examine in detail the phylogenetic trees presented in MEDLIN et al. (2008b) but to offer some commentary on a few general points. In particular I will focus on an idea previously articulated in MANN (1997) but recently stated more succinctly, concerning the apparent elusive 'absolute standard' of classification (SATO et al. 2008a: 386). I begin with a short discussion on the parameters of classification.

* Corresponding author: dmw@nhm.ac.uk

Genera, taxa, monophyly and the 'absolute standard'

Over a decade ago, the Japanese Journal of Diatomology (Diatom 13, 1997) presented a series of essays in honour of the late Professor Hiromu Kobayasi (1926–1996). The topic was the species concept in diatoms but, inevitably, discussion spilled over into concerns with other taxa at all ranks, partly because from the late 1980s onwards a veritable flood of new genera had been described (e.g. ROUND *et al.* 1990; see KOCIOLEK 1997, 1998; LANGE-BERTALOT 1997; ROUND 1997a, b). Reading the series of papers in *Diatom* today leaves one with the feeling that the core problem remains unresolved. A general question might be framed from those contributions: How might one find the right 'measure' for any particular taxon, the genus, for example? According to SATO *et al.* (2008a: 386), there is no absolute standard for the amount of sequence difference that justifies generic status.

Particular questions, such as 'What is a genus?' and 'What is a species?' might benefit from re-structuring. One might more profitably view the problem from the perspective of *artificial* and *natural classifications* rather than a quest for some measure. These two approaches – *artificial* and *natural classification* – might differ, inasmuch as the first is concerned with naming organisms, the second with naming schemes of relationships. Examples of artificial diatom classifications (SMITH 1853, HUSTEDT 1930) clearly considered identification to be of great importance and the primary purpose of a classification. Examples of natural classifications are harder to come across, as by their very definition they are temporary, subject to change. Some examples of persons clearly searching for such a system are AGARDH (1823–1828) and MEREZHKOWSKY (1902) (see WILLIAMS 2007 for a fuller review). That the two – a set of names for identification and a classification of relationships – might differ has been accepted for quite some time (WILLIAMS and KOCIOLEK 2007), although that may not be an altogether happy circumstance (WILLIAMS and EBACH 2007, 2009).

It seems fair to suggest that, over time, the goals of natural classification have been lost, or at least disappeared, even considered of little significance. This might possibly be because there are a vast number of general biologists who simply require names for their organisms, rather than a classification that reflects relationships.

MEDLIN *et al.* (2008b) did not concern themselves with classification beyond the generic level, as their intention was to test existing genera and to test some relatively new proposals. Happily, there is less literature concerning the genus concept than there is concerning the species concept. That does not necessarily make it easier to digest. I will not attempt to do so (for those interested, these more general references provide an easily accessible sample from the last 60 years: BARTLETT 1940, ANDERSON 1940, GREENMAN 1940, SHERFF 1940, LEGENDRE 1971, CLAYTON, 1972, 1983, STEVENS 1983). Here I simply quote from the conclusion of Peter Stevens's paper: »If a decision is made that the classification is not to be at least congruent with a phylogeny, and no attempt is made to produce such a phylogeny, then the work we produce will be of limited relevance to biologists« (STEVENS 1983: 463). Phylogeny might be directly equated with natural classification. There are now many attempts at producing phylogenies, the data source usually being DNA sequences but that need not be the case, as morphology is equally as useful (WILLIAMS and REID 2006). So the issue might be seen as relating the one – phylogenetic trees – to the other – classification.

The issue of an 'absolute standard', regardless of whether data are molecular or morphological, is explored in the rest of this paper, with an example derived from a re-analysis of the 'araphid' phylogenetic tree.

Re-analysis of the 18S rDNA 'Araphid' Diatom Data

Results from the most comprehensive 18S rDNA sequence analyses of 'araphid' diatoms were summarised in a tree with a total of 45 nodes (MEDLIN et al. 2008b: Figs. 1, 2). The groups so formed were recognised, informally, as five clades, labelled 1–5; clade 5 is subdivided into four sub-clades, 5A–5D, clade 2 is further subdivided into two sub-clades, 2A and 2B (MEDLIN et al. 2008b: Figs. 1, 2). None of the clades or sub-clades was named. Clade 3 has 5 specimens related as in Figure 1 (redrawn from MEDLIN et al. 2008b: Fig. 2). Clade 3 is interesting for a number of reasons:

- (1) All the specimens belong to different genera, therefore the sample is insufficient to determine *any* generic relationships;
- (2) Clade 3 includes *Striatella*, a genus usually most closely related to all pennate diatoms but »...18S rDNA analyses undertaken so far have placed *S.[triatella] unipunctata* in various phylogenetic positions«, thus its position in this tree is also somewhat anomalous (SATO et al. 2008a: 386);
- (3) A larger number of species from the genus *Licmophora* are now available in GenBank, hence a more detailed analysis can be undertaken.

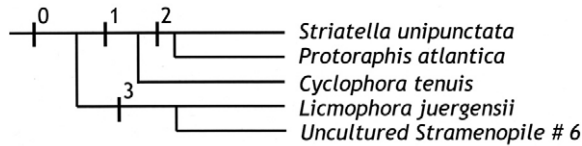


Fig. 1. Part of the phylogenetic tree from MEDLIN et al. 2008b, Figure 2, illustrating members and their relationships of their clade 3.

To reinvestigate 'araphid' relationships, 164 18s rDNA sequences were aligned using the program BioEdit (1997–2004) and ClustalW (LARKIN et al. 2007), the latter from within BioEdit's accessory applications (ClustalW was implemented using BioEdit's default options). Of the 164 sequences, 7 (*Arcocellulus mammifer*, *Cymatosira belgica*, *Extubocellulus spinifer*, *Minutocellus polymorphus*, *Minutocellus* sp. CCMP1701?, *Papiliocellulus elegans*, and *Talaroneis posidoniae*) are included as outgroup taxa, all from Cymatosiraceae, with the exception of *Talaroneis posidoniae*, which currently resides in Plagio-grammaceae (SATO et al. 2008b). A problem with both this re-analysis and that of MEDLIN et al. (2008b) is the lack of any raphid, pennate diatoms, which should be included as it is now accepted that 'araphid' diatoms are paraphyletic, not a natural group. The reason for their omission in this study was to make the analysis of 'araphid' data and the recognition of their sub-groups more tractable, given the size a matrix would need to be if relevant raphid diatoms were included. In any case, the focus of attention is the relationships of the genus *Licmophora*.

The complete alignment, including all introduced gaps, is 2830 bases long. Actual taxon sequence length ranges from 981 (*Synedropsis* sp. CCMP2747) to 2350 (*Hyalosira* sp. CCMP469); average sequence length ranges between 1750–1800 (a few relevant taxa with sequences in GenBank were omitted as they had too few bases to offer anything of significance).

Using the program Winclada (NIXON 1999–2002), uninformative characters (those that do not specify any relationships, those not shared by more than two taxa) can be removed, which discards a total of 2210, leaving 620 (c. 28%); if uninformative characters are removed using the option available in the parsimony program Nona (GOLOBOFF 1999), a total of 2245 are discarded, leaving 585 (c. 21%).

Analysis of the informative character matrix (585) with Nona yields in excess of 10,000 trees (the maximum tree number to be retained was set to 10,000; the implication being that there are more trees to be found should the search be prolonged and the number of trees retained increased); trees are of length 3079, $ci=51$, $ri=83$; using the parsimony ratchet (NIXON 1999, GOLOBOFF 1999), a subset of 67 trees are found, also of length 3079, $ci=51$, $ri=83$ (the strict consensus tree has a length of 3148, $ci=50$, $ri=83$).

Naming the tree (Fig. 2)

The analysis undertaken here includes a group of 13 specimens, of which 11 are named as species in the genus *Licmophora* (distributed among 7 species); the remaining two specimens are named as *Cyclophora tenuis* and *Protoraphis atlantica*. The relevant part of the complete tree is fully resolved with 11 nodes (Fig. 2, taken from the consensus tree). The 11 specimens recognised as belonging to *Licmophora* include a node relating *Cyclophora tenuis* and *Protoraphis atlantica* most closely to *Licmophora flabellata* and *L. communis* (Fig. 2, node 4) rendering *Licmophora* non-monophyletic (paraphyletic), if the genus name is applied to the root of the tree (Fig. 2, node 0). That is, all species of *Licmophora* relate to more than one node (Fig. 2, nodes 6, 7, and 9). One option would be to transfer both *Cyclophora tenuis* and *Protoraphis atlantica* to *Licmophora* creating a highly variable, but monophyletic, genus at node 0 (Fig. 2). However, several branches in the tree have monophyletic groups that include some species of *Licmophora* and could be named if one so desired: nodes 3, 6, 7, and 9, of which only one can be *Licmophora* s.s., the group that includes its type, *Licmophora argentescens* C.A. Agardh. (VANLANDINGHAM understands *Licmophora argentescens* to be synonymous with *L. flabellata*. If this is so, species from node 9 will be the genus *Licmophora* s.s. VANLANDINGHAM 1971, see ROUND et al. 1990.)

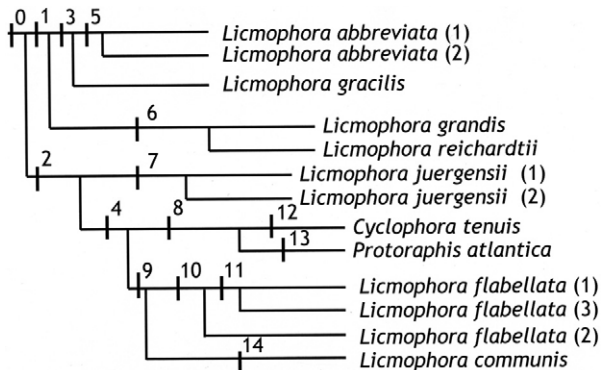


Fig. 2. Part of phylogenetic tree derived from a new analysis of 164 18s rDNA sequences, which includes a group of 13 specimens, 11 from the genus *Licmophora*, one each from *Cyclophora tenuis* and *Protoraphis atlantica*.

The branch leading to *Licmophora juergensii* (Figure 2, node 7) relates two specimens of the same species, basal to node 2. In this case, the only reasonable option is to place *L. juergensii* in a separate genus even though it is, at this time, monotypic.

Cyclophora and *Protoraphis* are represented by one species apiece; other species in each genus are said to exist but have yet to be investigated relative to rDNA data (possibly one more for *Cyclophora*; possibly two more for *Protoraphis*). Thus, at this time, it would be wise to retain their generic names representing (probable) monophyletic groups.

In total, then, for the species in figure 2, it is possible to name six genera: three monophyletic genera (nodes 3, 6, and 9) and one monospecific genus (node 7) for the species of *Licmophora*, plus *Cyclophora* and *Protoraphis* (Fig. 2, terminals 12 and 13). While the taxonomic level of genus is arbitrary, it is the fact that the groups recognised are monophyletic that is of significance.

Higher-level classification is more problematic, with a variety of ways of solving the problem. Here I offer one. The tree divides into two groups at nodes 1 and 2. Node 1 contains genus 1, from node 3 (node 5 is here ignored) and genus 2, from node 6 (Fig. 2). Node 2 subdivides into node 4 and genus 3, from node 7 (Fig. 2). Node 4 further subdivides into nodes 8 and 9, where node 8 relates *Cyclophora* and *Protoraphis* and node 9 relates most closely the two species in genus 4 (Fig. 2).

Tables 1–3 illustrate all possible nodes and their inclusive taxa (Tab. 1), related to groups and sub-groups (Tab. 2) and to possible ranks (Tab. 3). Not all nodes need naming; alternatives schemes are possible – but once again the 'absolute standard' is monophyly.

Node 7, the monotypic genus for *Licmophora juergensii*, remains a curiosity. If we assume an appropriate rank for the entire group (node 0) is an order, the next sub-division is into two families (nodes 1 and 2) (Tab. 3). Family 1 is composed of two genera. Family 2 is composed of further sub-divisions, the first being *Licmophora juergensii* plus all taxa from node 4. As node 4 has itself two further sub-divisions, *Licmophora juergensii* and all species included in node 4 could be sub-families. Thus, *Licmophora juergensii* will be in its own sub-family as well as its own genus.

The meaning of monotypic taxa is that the relationships of the included single species is unknown, relative to other closely related species, beyond their shared more basal node. In the case of *Licmophora juergensii*, its relationships are unknown other than it is being basal

Tab. 1. Nodes and inclusive taxa, derived from the tree in figure 2.

			Nodes	Taxa
0	1	3		<i>Licmophora abbreviata</i>
0	1	3		<i>Licmophora gracilis</i>
0	1		6	<i>Licmophora grandis</i>
0	1		6	<i>Licmophora reichardtii</i>
0	2		7	<i>Licmophora juergensii</i>
0	2	4	8	<i>Cyclophora tenuis</i>
0	2	4	8	<i>Protoraphis atlantica</i>
0	2	4	9	<i>Licmophora flabellata</i>
0	2	4	9	<i>Licmophora communis</i>

Tab. 2. Groups, sub-groups and genera, derived from the tree in figure 2.

Groups and sub-groups			Taxon	
Group 1	Sub-group 1 (1)		Genus 1 (3)	<i>Licmophora abbreviata</i>
			Genus 1 (3)	<i>Licmophora gracilis</i>
			Genus 2 (6)	<i>Licmophora grandis</i>
			Genus 2 (6)	<i>Licmophora reichardtii</i>
	Sub-group 2 (2)	Sub-sub-group 1 (7)	Genus 3 (7)	<i>Licmophora juergensii</i>
		Sub-sub-sub-group 1 (8)	Genus 4 (12)	<i>Cyclophora tenuis</i>
			Genus 5 (13)	<i>Protoraphis atlantica</i>
		Sub-sub-sub-group 2 (4)	Genus 6 (10)	<i>Licmophora flabellata</i>
		Sub-sub-sub-group 2 (9)	Genus 6 (14)	<i>Licmophora communis</i>

Tab. 3. Possible ranks for groups and sub-groups, derived from the tree in figure 2.

Ranks			Taxon	
Order 1 (0)	Family 1 (1)		Genus 1 (3)	<i>Licmophora abbreviata</i>
			Genus 1 (3)	<i>Licmophora gracilis</i>
			Genus 2 (6)	<i>Licmophora grandis</i>
			Genus 2 (6)	<i>Licmophora reichardtii</i>
	Family 2 (2)	Sub-family 1 (7)	Genus 3 (7)	<i>Licmophora juergensii</i>
		Sub-family 2 (4)	Tribe 1 (8)	Genus 4 (12)
			Genus 5 (13)	<i>Protoraphis atlantica</i>
			Tribe 2 (9)	Genus 6 (10)
			Genus 6 (14)	<i>Licmophora flabellata</i>
				<i>Licmophora communis</i>

to all species from node 2. A monotypic taxon – a single species – is neither mono-, para- nor polyphyletic; it is, strictly speaking, only most closely related to itself. Monotypy in this sense, serves a proper function, indicating the extent of our knowledge (or, rather, ignorance) (Fig. 2). This use is in contrast to that of Cyclophorales, for example, which was placed in its own family, Cyclophoraceae, with the intention of indicating its uniqueness in some unspecified way (ROUND et al. 1990). Thus, Cyclophorales and Cyclophoraceae are uninformative expressions of a certain amount of ignorance rather than acknowledgement of some unquantifiable uniqueness (PATTERSON 1982: 32).

Discussion

Above I presented part of a phylogenetic tree for species of *Licmophora* derived from a parsimony analysis of 18s rDNA sequence data. I also presented a classification derived from that tree. The tree allows precision in establishing the systematic position of each

taxon. Regardless of rank, each taxon relates to a particular node. Other classifications are possible but the tree – the nodes, the monophyletic groups – is the 'absolute standard'.

Of further significance, the tree allows precision when conducting further studies. For example, should another species of *Licmophora* be studied it will relate to some part of the tree. Should it relate directly to either node 3, 6 or 9, for example, it is part of that particular genus (Fig. 2). Should it relate to node 7 instead then support is gained for the genus based on *Licmophora juergensii*, providing it with a second species and demonstrating its monophyly (Fig. 2). In short, when further species are investigated their position and taxonomic status is determined in an absolute sense by the tree.

This can be contrasted with the viewpoint recently expressed and noted above. Defending the naming of the new monotypic genus *Pseudostriatella*, Sato, Mann and Medlin SATO et al. offered the following argument: »There are many morphological and ecological similarities between *P. oceanica* and *S. [Striatella] unipunctata*... [its most closest relative] [but] there are also many differences, which we regard as sufficient to differentiate these taxa at the rank of genus« (SATO et al. 2008a: 383). Of course, the notion of 'sufficient' remains unexplored and unquantified. Further, they state, that genetic distance may also be considered: »18S rDNA phylogeny places *P. oceanica* among the pennate diatoms and supports a close relationship between *P. oceanica* and *S. unipunctata*, but the genetic distance between them, coupled with the morphological differences, justifies separation at genus level« (SATO et al. 2008a: 383). Thus, 'many' differences linked to an unspecified genetic distance become 'sufficient' for 'justifying' a particular rank. But oddly, »...there is no absolute standard for the amount of sequence difference that justifies generic status« (SATO et al. 2008a: 386). One must assume SATO et al. guessed.

I have tried to show above that there *is* an 'absolute standard', which resides in the nodes of the phylogenetic tree, with the congruence of classification and phylogeny; in short, monophyly (WILLIAMS and KOCIOLEK 2007). I have also tried to show that classifications derived in this way are predictive, in the sense that any further information will relate – or meaningfully test – the relationships proposed; this is the essence of congruence – of which morphology provides data in abundance – and the essence of scientific investigation in general.

Acknowledgements

I wish to thank Mark Carine, Malte Ebach, Tony Gill and Pat Kociolek for their comments on various drafts of this paper and two anonymous referees for their comments.

References

- AGARDH, C. A. 1823–1828: Systema Algarum. Litteris Berlingianis, Lundae.
- ANDERSON, E. 1940: The concept of the genus: II. A survey of modern opinion. Bulletin of the Torrey Botanical Club 67, 363–369.
- BARTLETT, H. H. 1940: The concept of the genus: I. History of the generic concept in botany. Bulletin of the Torrey Botanical Club 67, 349–362.
- CLAYTON, W. D. 1972: Some aspects of the genus concept. Kew Bulletin 27, 281–287.

- CLAYTON, W. D. 1983: The genus concept in practice. *Kew Bulletin* 38, 149–153.
- GREENMAN, J. M. 1940: The concept of the genus: III. Genera from the standpoint of morphology. *Bulletin of the Torrey Botanical Club* 67, 371–374.
- GOLOBOFF, P. 1999: NONA (NO NAME) ver. 2. Published by the author, Tucumán, Argentina.
- HUSTEDT, F. 1930. Bacillariophyta (Diatomeae). In: PASCHER, A. (ed), *Die Susswasserflora Mitteleuropas*, 10 Verlag von Gustav Fischer, Germany.
- KOCIOLEK, J. P. 1997: Historical constraints, species concepts and the search for a natural classification of diatoms. *Diatom* 13, 3–8.
- KOCIOLEK, J. P. 1998: Does each genus of diatoms have at least one unique feature? – A reply to Round. *Diatom Research* 13, 177–179.
- LANGE-BERTALOT, H. 1997: As a practical diatomist, how does one deal with the flood of new names? *Diatom* 13, 9–12.
- LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J., HIGGINS, D. G. 2007: Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948; doi:10.1093/bioinformatics/btm404
- LEGENDRE, P. 1971: Circumscribing the concept of the genus. *Taxon* 20, 137–139.
- MANN, D. G. 1997: Shifting sands: The use of lower taxonomic ranks in diatoms. *Diatom* 13, 13–17.
- MEDLIN L. K., SATO S., MANN D. G., KOOISTRA W. H. C. F. 2008a: Molecular evidence confirms sister relationship of *Ardissonaea*, *Climacosphenia* and *Toxarium* within the bipolar centric diatoms. *Journal of Phycology* 44, 1340–1348.
- MEDLIN L. K., JUNG, I., BAHULIKAR, R., MENDGEN, K., KROTH, P., KOOISTRA, W. H. C. F. 2008b: Evolution of the diatoms. VI. Assessment of the new genera in the araphids using molecular data. *Nova Hedwigia* 133, 81–100.
- MEREZHKOWSKY, C. 1902: On the classification of diatoms. *Annals and Magazine of Natural History*, 9, 65–68.
- NIXON, K. C. 1999: The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15, 407–414.
- NIXON, K. C. 1999–2002: WinClada ver. 1.0000. Published by the author, Ithaca, NY, USA.
- PATTERSON, C. 1982: Morphological characters and homology. In: JOYSEY, K. A., FRIDAY, A. E. (eds), *Problems of Phylogenetic Reconstruction*, 21–74. Academic Press, London.
- ROUND, F. E. 1997a: Genera, species and varieties – are problems real or imagined? *Diatom* 13, 25–29.
- ROUND, F. E. 1997b: Does each diatom genus have a unique feature? *Diatom Research* 12, 341–345.
- ROUND, F. E., CRAWFORD, R. M., MANN, D. G. 1990: *The diatoms – biology and morphology of the genera*. Cambridge University Press, Cambridge.
- SATO, S., MANN, D. G., MATSUMOTO, S., MEDLIN, L. K. 2008a: *Pseudostriatella* (Bacillariophyta): a description of a new araphid diatom genus based on observations of frustule and auxospore structure and 18S rDNA phylogeny. *Phycologia* 47, 371–391.

- SATO, S., KOOISTRA, W. H. C. F., WATANABE, T., MATSUMOTO, S., MEDLIN, L. K. 2008b: A new araphid diatom genus *Psammoneis* gen. nov. (Plagiogrammaceae, Bacillariophyta) with three new species based on SSU and LSU rDNA sequence data and morphology. *Phycologia* 47, 510–528.
- SHERFF, E. E. 1940: The concept of the genus: IV. The delimitations of genera from the conservative point of view. *Bulletin of the Torrey Botanical Club* 67, 375–380.
- SMITH, W. 1853: A Synopsis of the British Diatomaceae, 1. Smith and Beck, London.
- STEVENS, P. F. 1983: The genus concept in practice – but for what practice? *Kew Bulletin* 40, 457–465.
- VANLANDINGHAM, S. L. 1971: Catalogue of the fossil and recent genera and species of diatoms and their synonyms. Part IV. *Fragilaria* through *Naunema*. J. Cramer, Lehre.
- WILLIAMS, D. M. 2007: Classification and diatom systematics: The past, the present and the future. In: BRODIE, J., LEWIS, J. (eds), *Unravelling the algae*, 57–91. CRC Press, Florida.
- WILLIAMS, D. M., EBACH, M. C. 2007: *The Foundations of systematics and biogeography*. Springer, New York.
- WILLIAMS, D. M., EBACH, M. C. 2009: What, exactly, is cladistics? Re-writing the history of systematics and biogeography. *Acta Biotheoretica* 57, 249–268.
- WILLIAMS, D. M., KOCIOLEK, J. P. 2007: Pursuit of a natural classification of diatoms: History, monophyly and the rejection of paraphyletic taxa. *European Journal of Phycology* 42, 313–319.
- WILLIAMS, D. M., REID, G. 2006: *Amphorotia* nov. gen., a new genus in the family Eunotiaceae (Bacillariophyceae), based on *Eunotia clevei* Grunow in Cleve et Grunow. *Diatom Monographs* 6, 1–153.