VIE ET MILIEU - LIFE AND ENVIRONMENT, 2016, 66 (2): 129-154

PHYLOGENY OF 'ARAPHID' DIATOMS INFERRED FROM SSU AND LSU RDNA, *RBC*L AND *PSB*A SEQUENCES

L. K. MEDLIN[1*], *Y. DESDEVISES*[2]

[1] *Marine Biological Association of the UK, the Citadel, Plymouth, PL1 2PB UK Royal Botanic Gardens, Edinburgh, Scotland, UK*

[2] *Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des Organismes Marins (BIOM), Observatoire Océanologique, F-66650, Banyuls/Mer, France*

** Corresponding author: lkm@mba.ac.uk*

ARAPHID
DIATOMS
DIVERGENCE TIME ESTIMATION
LSU
PINNATE
PHYLOGENY
*PSB*A
*RBC*L
SSU
SUPERTREE

ABSTRACT. – Phylogenies of the diatoms have largely been inferred from SSU rDNA sequences. Because previously published SSU rDNA topologies of araphid pennate diatoms have varied, a supertree was constructed in order to summarize those trees and used to guide further analyses where problems arose. As previously seen with the SSU trees, araphid diatoms were divided into two clades: basal and core araphids. The basal clade is sister to a clade containing other araphids (core) and the raphid diatoms. Several subclades recovered in the supertree did not correspond to current taxonomy in the diatoms but were supported by ecological and/or morphological characters. A phylogeny of diatoms was then estimated using four gene markers, SSU and LSU rDNA, *rbc*L and *psb*A (total 4352 bp) with 42 diatom species chosen to resolve problems in the supertree. Two rooting strategies were explored: 1) one bolidomonad as the closest outgroup of the diatoms and, 2) one bolidomonad and more distantly related heterokontophyte outgroups. Two different strategies were employed to analyze the four gene tree with both Maximum Likelihood (ML) and Baysian Inference (BI) methods. In the first strategy, the variable regions of the LSU rDNA and the third codon position of *rbc*L were recoded into R (A+G) and Y (T+C) because of substitution saturation detected at these positions in these genes. In the second, these regions were not recoded. Tree topologies of pennates were nearly identical in all analyses. Pennates were divided into three major clades, basal araphid, core araphid and raphid diatoms, as shown in the supertree. The four gene trees displayed better resolution and had stronger bootstrap within the subclades than those of the SSU supertree. The divergence time of the pennates with a Bayesian estimation was estimated, allowing for simultaneous constraints from the fossil record and varying rates of molecular evolution of different branches in the phylogenetic tree. The radiation of pennates into three major clades took place in a short period of geological time before their first appearance in the fossil record and earlier than that proposed by other clocks using single genes.

INTRODUCTION

The diatoms have historically been divided into two groups: the centrics and the pennates, which can be distinguished by their pattern centres or symmetry of the valve, mode of sexual reproduction, and plastid number and structure (Round *et al*. 1990). The centrics are oogamous, show a radially symmetrical ornamentation of their valves and usually possess numerous discoid plastids. The pennates are isogamous, have bilaterally symmetrical pattern centres in their valves and in general possess fewer, plate-like plastids, although there are some exceptions with numerous discoid plastids, *e.g. Nitzschia longissima* group. The classification system, in which diatoms receives the rank of a division, has been recently revised by Medlin & Kaczmarska (2004) mainly based on rDNA phylogenies but also supported by selected morphological, reproductive and cytological features. This system based on the nuclear-encoded small subunit ribosomal DNA (*SSU*) phylogenies (*sensu* Medlin & Kaczmarska 2004) separates the division Bacillariophyta into two groups at the rank of subdivision: Coscinodiscophytina comprising radial centrics, and Bacillariophytina comprising the rest of the diatoms that exhibit polarity in the shape of their valves, except for the Thalassiosirales, which is assumed to have lost the ability to produce bands for the auxospores and have reverted to the ancestral state, *i.e*. radial symmetrical valves (Medlin 2016a). The Bacillariophytina are divided into two classes: Mediophyceae for the polar centrics plus the Thalassiosirales and the Bacillariophyceae. The pennates form a well-supported monophyly in all the published *SSU* phylogenies (Sinninghe-Damsté *et al*. 2003, Medlin & Kaczmarska 2004, Alverson *et al*. 2006, Sorhannus 2007, Ashworth *et al*. 2012). These species all share a midrib or sternum, and rows of poroids, called striae, perpendicular to the midrib or sterum. The Mediophyceae and the Coscinodiscophyceae can be recovered as monophyletic if two conditions are met: 1) alignment by secondary structure of the ribosomal genes, and 2) multiple outgroups. Other wise they are recovered as a grade of clades in most cases (see review in Medlin 2016b).

Morphologically, pennates can further be subdivided into two groups by the presence or absence of a slit, called a raphe, in the valve. The "raphid" pennates, *i.e.*, those that possess a raphe, are potentially actively motile. That group is monophyletic. In contrast, the "araphid" pennates, *i.e.*, those lacking a raphe, do not move actively, or at times, move very slowly (Sato & Medlin 2006), are paraphyletic in all studies to date (Medlin & Kaczmarska 2004, Alverson *et al*. 2006, Kooistra *et al*. 2007b, Sorhannus 2007, Sato *et al*. 2008a, Li *et al*. 2015, Theriot *et al*. 2015), and it would appear that the unique feature likely shared by all araphids is the absence of a raphe or slit in the valve, the hair like appendage on the ameboid male gamete used to attach to the female gamete and pull them together (Sato *et al*. 2011) and the release of one or both of the gametes from the gametangia. Labiate processes, a special tube through the valve found in centrics and in many araphids, except where secondarily lost, is only found in one raphid order, the Eunotiales, which considered to be primitive raphid diatoms because the labiate process is generally believed to be the valve structure that evolved into the raphe (Hasle 1974) and which is usually either outside the raphe lineage in some phylogenies (Theriot *et al*. 2015, SSU only), as the basal lineage inside the raphid diatoms (Medlin & Kaczmarska 2004, Rim *et al*. 2011), or as the basal lineage to the Naviculales (Theriot *et al*. 2015, plastid genes and combined trees).

In the doctoral works of Sato (2008), araphid diatoms were extensively sampled to evaluate phylogenetic relationships among various groups. Because most of the previous studies of higher rank diatoms phylogeny had been inferred solely from *SSU* sequences (Alverson & Theriot 2005, Mann & Evans 2007), Sato utilized additional genes outside the *SSU, viz.,* the D1/D2 region of the nuclear-encoded large subunit ribosomal DNA (*LSU*), the plastid encoded ribulose-1,5-bisphosphate carboxylase/oxygenase gene (*rbc*L), and the photosystem II reaction center D1 protein gene (*psb*A). In this paper, the terms *centric*, *araphid* and *raphid* are used, despite the paraphyly of two of these groups, because they refer to key morphological valve features or their absence. This does not imply that this corresponds to a mono- (holo-) phyletic group, or that it should be accorded any taxonomic status (Medlin & Sato 2009) except where the group is monophyletic (raphid).

Both the fossil record and molecular phylogenies indicate that the pennates evolved from the centrics as suggested by early workers (Fritsch 1935, Simonsen, 1979). However, the fossil record of the diatoms is not entirely free of problems. Although most extant taxonomic groups have fossil representatives, there is no substantial fossil record of the key evolutionary transitions from centrics to pennates, and from araphid to raphid diatoms. This is likely because the habitats where araphids are abundant (nearly shore coastal, benthic) are under-sampled or not available for stratigraphic studies. The earliest fossil date for a diatom at 180 My is taken from a sponge that was boiled in HCl acid, leaving only siliceous frustules (Medlin 2016a).

Molecular phylogenetic methods can address these issues by calibrating fossil dates with internodes in phylogenies inferred from sequence data, thus allowing the estimation of divergence times across the entire gene tree of a group. In the past, this has been accomplished assuming a molecular clock, that is, constancy of evolutionary rates across lineages (Kooistra & Medlin 1996). Under this assumption, the estimated branch lengths can be converted into absolute divergence times using fossil calibration. However, most datasets appear to violate the clock model (Graur & Martin 2004), which can cause serious biases in divergence time estimation (Rambaut & Broham 1998, Soltis *et al*. 2002). Consequently, a relaxed clock model has recently been introduced to overcome the inconstancy of the clock to study divergence time estimation of diatoms (Sorhannus 2007, see also Berney & Pawlowski 2006). However, previous divergence time estimations of diatoms utilised, as in most phylogenetic studies of these algae, solely *SSU* sequence data. The divergence time of the pennates was estimated in the present study using four genetic markers and Bayesian frameworks that account for rate variation of substitution when estimating divergence times and incorporate multiple genetic markers and multiple fossil calibration points (Thorne & Kishino 2002).

MATERIALS AND METHODS

Supertree construction: In order to summarise previously published *SSU* phylogenies, and to see consensus trends of these topologies, Sato (2008) constructed a supertree using Clann 3.0.0 (Creevey & McInerney 2005) with default settings under the Matrix representation using Parsimony (MRP) method. Topologies used to construct the supertree were taken from Medlin *et al*. (2000), Medlin & Kaczmarska (2004), Kooistra *et al*. (2003a, b, 2004), Sinninghe-Damsté *et al*. (2004), Alverson *et al*. (2006, Fig. 4), Sato *et al*. (2008a, c, e), Sorhannus (2004, 2007). All topologies were manually coded into newick format and entered into the program. In this procedure, all centrics were simply coded into one OTU 'centric' to root the supertree, and all raphid diatoms were also coded as one OTU 'raphid diatoms'. The araphid diatom *Striatella unipunctata* (Lyngbye) Agardh was pruned from the input trees when this species appeared in the clade of raphid diatoms (Medlin *et al*. 2000, Kooistra *et al*. 2003a, b, 2004, Sato *et al*. 2008a, c, e). The topology of Medlin *et al*. (2008a) was not included in the analysis because the phylogenetic tree was rooted with an araphid diatom *Rhabdonema* Kützing, making it impossible to compare the phylogenetic relationship of araphid diatoms using the same standard as the other studies, which were rooted by centrics and more distant outgroups. Also, other information (ecological/morphological features) were mapped onto Sato's supertree (Fig. 1) referring

to the original description of each species or some floristic and monographic works (Hustedt 1959, Round *et al*. 1990, Snoeijs 1992, Witkowski *et al*. 2000, Hasle 2001, Kobayasi *et al*. 2006) to see whether any other features correlated with any clade.

Taxon sampling of four gene analyses: The ideal taxon sampling should be broad and equal. Sato (2008) selected a few representative species from each supertree clade for four gene analyses to resolve the interclade relationships that were unresolvable in the supertree. In order to reduce the computational cost, some OTUs, which clearly belong to the same lineage with the representatives judging from morphology and/or the other information, were omitted from the dataset. For example, *Talaroneis* (Ricard) Kooistra & De Stefano, *Neofragilaria* Desikachary, Prasad & Prema, *Plagiogramma* Greville and *Psammoneis* Sato, Kooistra & Medlin (all belong to Family Plagiogrammaceae) were omitted because *Dimeregramma* Ralfs was selected as a representative of this morphologically well-defined family, whose monophyly was also supported by the *SSU* and *LSU* analyses (Sato *et al*. 2008a, f, respectively, Li *et al*. 2015). Two araphid species (*Pseudohimantidium* Hustedt & Krasske and *Pteroncola* Holmes & Croll), whose phylogenetic positions had never been reported so far, were newly added to his dataset.

The basis for taxon sampling in the centric and raphid diatoms was slightly different. The centric and raphid diatoms were rather regarded as close- and inner outgroup because the primary focus of this study was the araphid diatoms. They had been selected equally from broad range of each grade/clade referring to the previously published *SSU* trees of diatoms (Medlin & Kaczmarska 2004).

As a result, the dataset comprised 42 diatom species: 5 Coscinodiscophyceae and 9 Mediophyceae (close outgroups) and 28 Bacillariophyceae, of which the latter included 6 raphid (inner outgroup) and 22 araphid diatoms. The whole genome sequences of *Thalassiosira pseudonana* Hasle & Heimdale and *Phaeodactylum tricornutum* Bohlin are available so that four genes used in this study were retrieved from those databases. The plastid genome sequence of *Odontella sinensis* (Greville) Grunow is also available so that we sequenced the nuclear gene markers from the same CCMP strain. Culture strains used in this study (Table I) are currently available upon request to the first author, but may not survive long-term in culture (Chepurnov *et al*. 2004), instead, voucher specimens of cleaned material of the strains were, if available, mounted as permanent slides and have been deposited in the Hustedt Collection, Alfred Wegener Institute, Bremerhaven, Germany (Table I). Older material can be found in the slide collections of LKM.

A similar strategy was followed in Li *et al*. (2015) in which the primary taxa of interest were the araphid diatoms and out of 157 taxa only 10 radial and 19 bipolar centrics were sampled along with 62 raphids. Thus the araphid percentage of that study was 42 %, similar to that used in this study (52 %).

Because outgroup selection can alter the topology of the ingroup (Milinkovitch & Lyons-Weiler 1998, Tarrio *et al*. 2000), two strategies were deployed to compare the effect of outgroup selection. The first dataset used *Bolidomonas pacifica* Guillou & Chrétiennot-Dinet (Bolidophyceae, Guillou *et al*. 1999) as single closest outgroup to root the diatoms. The second dataset used multiple outgroups of heterokont algae, including *Bumilleriopsis filiformis* Vischer (Xanthophyceae), *Dictyota dichotoma* (Hudson) Lamouroux (Phaeophyceae) and *Heterosigma akashiwo* (Hada) Hada (Raphidophyceae) together with *Bolidomonas*. This rooting strategy was used because Medlin & Kaczmarska (2004) suggested that inclusion/exclusion of distant outgroups affects the monophyly of three major classes: Coscinodiscophyceae, Mediophyceae and Bacillariophyceae. Medlin (2014) has extensively tested multiple distant outgroups using the SSU gene and found that they had a profound affect on class monophyly. Hereafter, each dataset will be referred to as Bolido-root and Distant-root, respectively. In each diatom species and in *Bolidomonas*, four genes were sequenced from the identical strain, although the distant outgroups should be considered chimeric sequences because their sequence data were retrieved from GenBank from different strains of the same species.

DNA extraction, PCR, and sequencing: The following protocols and analysis results are reproduced from Sato (2008) with his permission because he is no longer interested in phylogenetic analysis of the diatoms, having moved more into sexual reproduction studies. "Samples of *c*. 500 mL of culture were harvested by filtration and DNA was extracted using a modified CTAB protocol (Doyle & Doyle 1990) or by the PAN Plant Kit (PAN Biotech, Aidenach, Germany). The quantity and quality of DNA were examined by agarose gel electrophoresis against known standards. Partial fragments of *SSU* (~1657 bp), *LSU* (~659 bp), *rbc*L (~1461 bp) and *psb*A (~933 bp) were amplified by PCR (see Table II for primers).

The markers were PCR-amplified in 25 μL volumes containing 10 ng DNA, 1 mM dNTPs, 0.5 μM of forward primer, 0.5 μM of reverse primer, 1 x Roche diagnostics PCR reaction buffer (Roche Diagnostics, GmbH, Mannheim, Germany), and 1 unit *Taq* DNA polymerase (Roche). The PCR cycling of nuclear genes comprised an initial 4-min heating step at 94 °C, followed by 35 cycles of 94 °C for 2 min, 56 °C for 4 min, and 72 °C for 2 min, and a final extension at 72 °C for 10 min, of plastid genes comprised an initial 5-min heating step at 94 °C, followed by 35 cycles of 95 °C for 1.5 min, 47 °C for 2 min, and 72 °C for 2 min, and a final extension at 72 °C for 6 min. The quantity and length of products were examined by agarose gel electrophoresis against known standards. Excess primers and dNTPs were removed from PCR product using the QIAQuick purification kit (QIAGEN, Hilden, Germany). Sequencing reactions took place in a PCR cycler using Big Dye Terminator v3.1 sequencing chemistry (Applied Biosystems, CA, USA) with sequencing primers described in Elwood *et al*. (1985). The PCR products were then electrophoresed on an ABI 3100 Avant sequencer (Applied Biosystems, CA, USA)."

For the Sato analysis, "rDNA sequences were aligned first using ClustalX (Thompson *et al*. 1997), and then refined by referring to some secondary structure models of the rRNA (Van de Peer *et al*. 1998, Alverson *et al*. 2007 for *SSU*, Sato *et al*. 2008a for *LSU*). There is extreme length variation in some

Table I. – List of species used in this analysis.

Taxon	Strain[b]	*SSU*	*LSU*	*rbc*L	*psb*A
Coscinodiscophyceae [Centrics]					
Aulacoseira granulata (Ehrenberg) Simmons	p778	AB430586	AB430619	AB430659	AB430699
Hyalodiscus scoticus (Kützing) Grunow	s0284[c]	AB430587	AB430620	AB430660	AB430700
Melosira dubia Kützing	s0076	AB430588	AB430621	AB430661	AB430701
Rhizosolenia setigera Brightwell	p1692	AB430589	AB430622	AB430662	AB430702
Stephanopyxis turris (Greville & Arnott) Ralfs	p121	AB430590	AB430623	AB430663	AB430703
Mediophyceae [centrics]					
Ardissonea baculus (Gregory) Grunow	wk76	AF525668	AB430624	AB430664	AB430704
Cyclotella meneghiniana Germain	p567	AB430591	AB430625	AB430665	AB430705
Chaetoceros radicans Schütt	CCMP197	AB430592	AB430626	AB430666	AB430706
Cymatosira belgica Grunow	p189	X85387	AB430627	AB430667	AB430707
Eunotogramma laevis Grunow *in* Van Heurck	s0382	AB430593	AB430628	AB430668	AB430708
Lampriscus kittonii Schmidt	p535	AF525667	AB430629	AB430669	AB430709
Odontella sinensis (Greville) Grunow	CCMP1815	Y10570	AB430630	Z67753[a]	
Thalassiosira pseudonana Hasle & Heimdal	CCMP1335	EF208793[a]		EF067921[a]	
Stephanodiscus sp.	p404	AB430594	AB430631	AB430670	AB430710
Bacillariophyceae [Pennate]					
Araphid diatoms					
Asterionella formosa Hassall	s0339	AB430595	AB430632	AB430671	AB430711
Asteroplanus karianus (Grunow) Gardner & Crawford	s0381	AB430596	AB430633	AB430672	AB430712
Cyclophora tenuis Castracane	p438	AJ535142	AB430634	AB430673	AB430713
Diatoma moniliforme Kützing	s0383	AB430597	AB430635	AB430674	AB430714
Dimeregramma minor var. *nana* (Gregory) Ralfs	s0355	AB430598	AB425083	AB430675	AB430715
Fragilaria bidens Heiberg	s0327	AB430599	AB430636	AB430676	AB430716
Grammatophora marina (Lyngbye) Kützing	s0190	AB430600	AB430637	AB430677	AB430717
Hyalosira delicatula Kützing	p439	AF525654	AB430638	AB430678	AB430718
*Licmophora parado (*Lyngbye) Agardh	s0213	AB430601	AB430639	AB430679	AB430719
Nanofrustulum shiloi (Lee, Reimer & McEnery) Round, Hallsteinsen & Paasche	p194	AM746971	AB430640	AB430680	AB430720
Rhaphoneis sp.	s0366	AB430602	AB430641	AB430681	AB430721
Rhabdonema minutum Kützing	s0351	AB430603	AB430642	AB430682	AB430722
Opephora sp.	s0357	AB430604	AB430643	AB430683	AB430723
Plagiostriata goreensis S Sato & Medlin	s0388 (Zu6/60)	AB430605	AB430644	AB430684	AB430724
Pseudohimantidium pacificum Hustedt & Krasske	mhk033	AB430606	AB430645	AB430685	AB430725
Pseudostriatella pacifica S Sato & Medlin	s0384 (Zu6/38)	AB379680	AB430646	AB430686	AB430726
Pteroncola inane (Giffen) Round	s0247	AB430607	AB430647	AB430687	AB430727
Pseudostaurosira brevistriata (Grunow in Van Heurck) Williams & Round	s0398	AB430608	AB430648	AB430688	AB430728
Striatella unipunctata Agardh	s0208	AB430609	AB430649	AB430689	AB430729
Tabularia laevis Kützing	s0021	AB430610	AB430650	AB430690	AB430730
Thalassiothrix longissima Cleve & Grunow	p441	AB430611	AB430651	AB430691	AB430731
Hyalosira tropicalis Navarro	s0252	AB430612	AB430652	AB430692	AB430732
Raphid diatoms					
Campylodiscus thuretii Brébisson	s0223	AB430613	AB430653	AB430693	AB430733
Cocconeis stauroneiformis (Rabenhorst) Okuno	s0230	AB430614	AB430654	AB430694	AB430734
Navicula sp.	s0020	AB430615	AB430655	AB430695	AB430735
Nitzschia dubiiformis Hustedt	s0311	AB430616	AB430656	AB430696	
Phaeodactylum tricornatum Bohlin	CCAP1055/1	EF553458[a]		EF067920[a]	
Psammodictyon constrictum (Gregory) Mann	s0309	AB430617	AB430657	AB430697	AB430737
Outgroup					
Bolidomonas pacifica Guillou & Chréteinnot-Dinet [Bolidophyceae]	p380	AB430618	AB430658	AB430698	AB430738
Bumilleriopsis filiformis Vischer [Xanthophyceae]	NA	AF083398	NA	X79223	

Table I. Continued.

Taxon	Strain[b]	*SSU*	*LSU*	*rbc*L	*psb*A
Dictyota dichotoma (Hudson. Lamouroux [Phaeophyceae]	NA	AF350227	AF331152	AY748321	
Heterosigma akashiwo (Hara. Hara [Raphidophyceae]	NA	DQ470662		AY119759	

[a] Whole genome accession number
[b] Strains with a p number should refer to Table 1 in Medlin & Kaczmarska (2004) and those with ZU65/xxx are deposited in the Hustedt Diatom Collection in Bremerhaven, Germany
[c] Samples from Sato without a voucher number in the Hustedt collection have been deposited but not yet curated.

rRNAs (Gillespie *et al*. 2005) and replication slippage often leads to convergence on similar primary and secondary structures (Hancock & Vogler 2000, Shull *et al*. 2001). Homology assessment in such regions was difficult or impossible, so that the highly variable regions (mostly peripheral regions of the rRNA secondary structure) were removed from the alignment using BioEdit 7.0.2 (Hall 1999), by referring to the variability map of *Saccharomyces cerevisiae* (Van de Peer *et al*. 1993 for *SSU*, Ben Ali *et al*. 1999 for *LSU*). Secondary structure models referred to align the *SSU* alignment were not the same especially in some variable regions, however, these differences had no substantial effect to the further analyses because these regions were highly variable and were removed in this procedure", but see the effect in the second analysis below. "Sequences for the protein coding genes *psb*A and *rbc*L lacked indels and were aligned manually. The final dataset comprised 4352 bp." The dataset examined in this study is available from Shinya Sato." In the second analysis set below, both the SSU and the LSU were aligned using the SILVA alignment in ARB.

Sequence analysis: "Substitutional noise in a gene marker makes it difficult to retrieve phylogenetic signal, especially when the gene is evolving fast. Saturation is caused by multiple substitutions at the same site. Because transitions occur more frequently than transversions (DeSalle 2005), transitions can likely suffer from saturation especially in a fast-evolving region of a gene, *i.e*., D1/D2 region of the nuclear *LSU* and third codon position of the protein coding genes in our dataset. Therefore, saturation of both transitions and transversions were tested by DAMBE 4.5.56 (Xia & Xie 2001)." These positions were recoded RY in one analysis and left uncoded in the second one.

Phylogenetic analyses: "All phylogenetic analyses were performed on both datasets, Bolido-root and Distant-root. For model based estimations, maximum likelihood (ML) and Bayesian inference (BI), each nuclear gene and each codon position of protein coding genes were treated as a separate partition (*i.e*., 8 partitions in the dataset) with independent model parameters, whereas a concatenated dataset was used for neighbour joining (NJ) and maximum parsimony (MP) analyses."

Two different analyses were done. The first analysis used RY recoding of the third codon position in the protein genes and in hypervariable regions of the LSU. Analyses were performed using ML, BI, MP, and NJ methods using both Bolido-root and the Distant-root datasets. The second analysis used uncoded original sequences only from the Distant-root dataset analysis based on results from the recoded one. Also full length SSU sequences were used.

In the first analysis, "the Message Passing Interface (MPI) version of MrBayes 3.1.2 (Huelsenbeck & Ronquist 2001, Ronquist & Huelsenbeck 2003, Altekar 2004) was used for Bayesian analyses with the GTR + I + G model for each partition, as suggested by MrModeltest 2.2 (Nylander 2004), to estimate the posterior probability distribution using Metropolis-Coupled Markov Chain Monte Carlo (MCMCMC) (Ronquist & Huelsenbeck 2003). Gamma correction values and a proportion of invariable sites of each partition were obtained automatically by the program. MCMCMC from a random starting tree were used in this analysis with two independent runs and 1 cold and 9 heated chains with temperature set 0.05. Bayesian analyses were run for 20 million generations each with trees sampled every 100th generation. To increase the probability of chain convergence, we sampled trees after the standard deviation values of the two runs dipped below 0.01 to calculate the posterior probabilities (*i.e*., after 810,000 and 500,000 generations for Bolido- and Distant-root, respectively). The remaining phylogenies were discarded as burn-in.

"RAxML-VI-HPC, v2.2.3 (Stamatakis *et al*. 2005) was used for ML analyses with the GTRMIX model. The Gamma correction value of each partition was obtained automatically by the program. The analyses were performed 1,000 times to find the best topology receiving the best likelihood using different random starting MP trees (one round of taxon addition) and the rapid hill-climbing algorithm (*i.e*., option -f d in RAxML). Bootstrap values were obtained by 1,000 replications with GTRCAT model."

"For NJ analyses, PAUP* was used with JC distances. Nodal support was estimated using NJ bootstrap analyses using the same settings (1,000 replicates). MP searches were done with the "new technology" search algorithm implemented in the Willi Hennig Society edition of TNT 1.1 (Goloboff *et al*. 2008). One hundred random addition sequence replicates were performed with default values. Nonparametric bootstrap analyses were done 1,000 times with the "traditional" search algorithm in TNT. MP analyses were not weighted."

In the second analysis, because BT support was low for clades of interest in the Distant-root dataset with RY recoded third positions, we performed an additional BI analysis with the third codon and the LSU not recoded and with full length SSU sequences. Phylogenetic reconstructions were carried out with the concatenated alignment of the 4 DNA markers, uncoded, using Bayesian inference with Mr Bayes 3.2 (Ronquist *et al*. 2012) with 2 runs of 4 chains of 1 million generations, trees sam-

Table II. – Primers used in this study

Marker	Name	Reference
SSU	A[a]	Medlin *et al.* (1988)[c]
	528F	Elwood (1985)
	1055F	Elwood (1985)
	536R	Elwood (1985)
	1055R	Elwood (1985)
	B[b]	Medlin *et al.* (1988)[c]
LSU	D1RF[a]	Scholin *et al*. (1994)
	D2CR[b]	Scholin *et al*. (1994)
*rbc*L	DPrbcL1[a]	Daugbjerg & Andersen (1997)
	AraphidF[a]	This study[d]
	16F	Jones *et al.* (2005)
	14R	Jones *et al*. (2005)
	DPrbcL7[b]	Daugbjerg & Andersen (1997)
*psb*A	psbA-F[a]	Yoon *et al*. (2002)
	psbA500F	Yoon *et al*. (2002)
	psbA-R2[b]	Yoon *et al.* (2002)
	psbA600R	Yoon *et al*. (2002)

[A] Forward PCR amplification primer.
[b] Reverse PCR amplification primer.
[c] Without polylinkers.
[d] 5'-GTCTCAATCTGTATCAGAAC-3

pled every 100 generations, and burnin value set to 20 % of the sampled trees. We checked that standard deviation of the split frequencies fell below 0.01 to ensure convergence in tree search. For each DNA marker, the best evolutionary model was selected via Akaike Information Criterion using jModelTest v2 (Darriba *et al*. 2012). In all cases, a general time reversible model was selected, accounting for rate heterogeneity across sites via a Gamma distribution and a proportion of invariant sites. Model parameters were independently estimated for each marker, and relaxed across codon position for coding genes (*psb*A and *rbc*L). *Heterosigma* was used as outgroup because in this analysis the missing LSU gene in *Bumilleriopsis* caused it to be placed on a very long branch (data not shown). The remaining topology was the same as that with *Bumilleriopsis* as the outgroup.

Testing alternative hypotheses: To compare some tree alternatives, Sato (2008) used "the Shimodaira-Hasegawa-test as implemented in PAUP* (Shimodaira & Hasegawa 1999, test option FullOpt, 1,000 bootstrap replicates). He also generated alternative topologies by introducing constraints, *i.e*., monophyly of centrics and araphid diatoms."

Divergence time estimation: "Divergence times of pennates were estimated following Bayesian methods (Thorne *et al*. 1998, Kishino *et al*. 2001, Thorne & Kishino 2002) using the program package of mulidistribute 9/25/03 (Thorne & Kishino 2002) and PAML 3.15 (Yang 1997). Original sequences were used in this process (*i.e*., all sequences were not RY re-coded), because the preliminary analyses using RY re-coded dataset showed that the likelihoods obtained from the baseml and estbranches analyses were very different, suggesting that one or both programs failed to optimize the likelihood. Analyses were performed either using BI and ML topologies. Because Sato focused on dating the evolution of the araphid pennates, there was no difference within pennate topology between the two datasets Bolido- and Distant-root. Therefore, the topologies of Bolido-root dataset were used in order to reduce calculation time. Although the entire topology was used for each analysis, Sato only presented the results of pennate lineage because the divergence pattern of pennates is nearly identical using BI and ML analyses making the results comparable. Branch lengths were estimated with the estbranches 8/5/03 of multidistribute in conjunction with the BI/ML topology. Sato used the F84 model incorporating among-site rate variation modelled by a gamma distribution (Yang 1994) of sequence evolution (F84+G model). This is the most complex model implemented in this program. This model is less parameterized than the best-fit models selected by Mr Modeltest 2.2 (above), however, previous studies (Yang & Yoder 2003, and references therein) have shown that it is actually the rate variation among site parameter that has the greatest effect on divergence time estimation. All the parameters within the model as well as the branch lengths were estimated separately for every gene. Markov chain Monte Carlo (MCMC) approximations were obtained with a burnin period of 100,000 proposal cycles. Thereafter, samples of the Markov chain were taken every 100 cycles until a total of 1,000,000 generations were obtained. The uncertainty of divergence time estimates was accounted for by using the 95 % credibility intervals of these 10,000 samples. To diagnose possible failure of the Markov chains to converge to their stationary distribution, Sato performed two replicate MCMC runs with different initial starting points for each analysis. Application of the multidivtime program requires a value for the mean of the prior distribution for the time separating the ingroup root from the present (rttm). Sato used a maximum (250 Ma) and minimum estimate (190 Ma) as rttm (see below about the calibration strategy), and the SD for the prior on the root rate (rttmsd) was set at half the prior on the root rate. The highest possible time between tips and root (bigtime) was also set at 250 and 190 My for each analysis. Other parameters for running multidivtime were set referring Rutschmann (2004) as follows: rtrate = X/rttm, where rtrate is the mean of prior distribution for the rate at the root node and X is the median amount of evolution from the ingroup root to the ingroup tips, which was obtained by TreeStat 1.1 (available at http://tree.bio.ed.ac.uk/software/treestat/), rtratesd = rtrate, where rtratesd is the standard deviation of rtrate, rttm*brownmean = 1, where brownmean is the mean of the prior distribution for the autocorrelation parameter (m), and brownsd = brownmean, where brownsd is the standard deviation of the prior distribution for m."

Reference fossils: For a minimum age constraint, the occurrence of a taxon in the fossil record was used. Based on a molecular clock, Medlin *et al*. (1997) and Medlin (2006) speculated that the pigmented heterokonts, to which the diatoms belong, diversified following the P/T boundary (*c*. 250 Ma). "Although using a constraint that was obtained solely by molecular clock method as a calibration point might be risky, the age after P/T boundary was also supported by paleoecological discussions

(Medlin *et al*. 1997, Medlin 2011), making this assumption reasonable. Tappan (1980) reported no diatoms from the well-preserved Palaeozoic cherts (659-*c*. 250 Ma), which contained radiolarians and sponges. Thus, Sato assumed the maximum age of the diatoms at 250 Ma. Recent dating study of diatoms by Sorhannus (2007) suggested that the first diatoms evolved some time between 250 and 183 Ma, validating our assumption of the maximum age of the diatoms at 250 My. The assumption of the minimum age of the origin is much more straightforward. 190 My was set for the minimum age of the origin. This assumption is based on the earliest generally accepted record of diatoms described by Rothpletz (1896, 1990), which has been found from the Toarcian stage of the Jurassic (*ca*. 190 My by Sims *et al*. 2006)." Medlin (2016a) has had the original German translated and it appears that Rothpletz boiled a bath sponge in HCl to obtain the diatoms he described, thus this 180 Ma record is a valid diatom one. It is not a deposit as traditionally assumed because the sponge was boiled in the acid. Thus, the time estimation was undertaken for four conditions using BI and ML topologies, assuming the origin of the diatoms 250 and 190 Ma. Please see Medlin (2016a) for further rationales for using 250 and 190 My as ages for the diatom root.

"The other calibration points were selected from the review of the fossil diatoms by Sims *et al*. (2006), in which reliable fossils were introduced, allowing them to be used in Sato's divergence time estimation. Calibrated ages and fossils used here are shown in Table III. Although the fossil records of *Thalassiosira* Cleve and its potential affinities, *e.g*., *Praethalassiosiropsis* Gersonde & Harwood and *Thalassiosiropsis* Hasle, are abundant in fossil records, they were not used for the calibration by way of caution because the evolutional relationship of these fossil and the modern taxa is not confidently established". In the cladistic analysis in Medlin (2016a) showed that these fossil diatoms can only be in the thalassiosiralean lineage. Similarly, fossil records of resting spores 'presumably' formed by *Chaetoceros* were not used in the present study.

RESULTS AND DISCUSSION

SSU supertree of araphid diatoms

12 *SSU* trees displaying the paraphyletic nature of araphid diatoms, which consistently comprised two clades: basal and core araphids (Fig. 1) were used by Sato (2008) to construct a supertree. These two clades are now given the rank of sub class (Medlin 2015). Several clades in the supertree did not correspond to current diatom taxonomy, but rather they were supported by ecological and/or morphological characters. Sato (2008) described the characters supporting each clade. Furthermore, expected/predicted members of extant genera to be included in each clade based on the morphology of that clade are also listed. This summary of each clade is designed to show monophyletic groups whose morphological, ecological and perhaps cytological and reproductive features should be investigated more closely to define the clades taxonomically.

Clade 1 – Family Fragilariaceae, Thalassionemataceae

Genera included in this clade in the *SSU* supertree: *Catacombas* Williams & Round, *Fragilaria* Lyngbye, *Fragilariforma* (Ralfs) Williams & Round, *Grammonema* Agardh, *Synedra* Ehrenberg, *Synedropsis* Hasle, Medlin & Syvertsen, *Tabularia* (Kützing) Williams & Round, *Thalassionema* Grunow *ex* Mereschkowsky.

The highly elongated araphid diatoms (termed by Medlin *et al*. 2008b, length/width ratio grater than 15:1 in their maximum valve size) can only be seen in this clade. The end of the valve bears one or two labiate process(es) that a structure located inside the valve, which is believed to be involved with movement and the precursor of the raphe (see discussion in Medlin & Kazsmarska 2004). Chambered valve structure has exclusively been found in this clade in araphid diatoms. The ocellulimbus, a discrete area depressed below the surface of the valve and responsible for mucilage secretion (Williams 1986) at the end of valve is exclusively observed in this clade, except in the genus *Synedropsis* where the pores are replaced by slits and the entire apical pore field is not recessed below the valve surface. Exceptionally, however, the ocellulimbus can also be found in *Striatella unipunctata* (Hasle 1974), whose allocation in the *SSU* supertree remained unresolvable (Fig. 1) but has never been sistered or included in clade 1 diatoms in previous *SSU* phylogenies (*e.g*., Medlin & Kaczmarska 2004, Alverson *et al*. 2006, Li *et al*. 2015, Theriot *et al*. 2015, plastid trees). An araphid diatom *Pseudostriatella oceanica* Sato, Mann & Medlin, the closest relative of *Striatella* in *SSU* (Sato *et al*. 2008b) and four gene trees (Sato 2008, Fig. 3 here), bears no ocellulimbus, strongly suggesting the homoplastic nature of the ocellulimbus of *Striatella*.

Expected/Predicted genera to be included in this clade: *Centronella* Voigt (p. 348, Round *et al*. 1990), *Ctenophora* (Grunow) Williams & Round (Williams & Round 1986, p. 372, Round *et al*. 1990), *Entopyla* Ehrenberg (Prasad & Fryxell 1991), *Gephyria* Arnott (John, 1984, p. 440, Round *et al*. 1990), *Hannaea* Patrick (Patrick & Reimer 1966, p. 366, Round *et al*. 1990), *Hyalosynedra* Williams & Round (Williams & Round 1986, p. 380, Round *et al*. 1990), *Lioloma* Hasle (Hasle & Syvertsen 1996, Hasle 2001), *Neosynedra* Williams & Round (Williams & Round 1986, p. 372, Round *et al*. 1990), *Psammosynedra* Round (Round 1993), *Pteroncola* Holmes & Croll (Sullivan 1979, p. 390, Holmes & Croll 1984, p. 267, Round *et al*. 1990), *Reimerothrix* Prasad (Prasad *et al*. 2001), *Thalassiothrix* Cleve & Grunow (p. 426, Round *et al*. 1990, Hasle 2001), *Trichotoxon* Reid & Round (p. 428, Round *et al*., 1990, Hasle 2001), *Ulnaria* (Kützing) Compère (Compère 2001).

Table III. – Geographic records (in Ma. of diatoms introduced in Sims *et al*. (2006) and Singh *et al*. (2007) used as minimum age constraints in divergence time estimation except for the calibration of *Rhizosolenia* (see footnote f).

Genus/Clade	Appearance Period (Stage if available: Age [a])	Calibration [b]	Literature [c]
Asterionellopsis[d]	Late Miocene (Messinian): 6.5-5.3	5.3	Schrader & Gersonde (1978)
Aulacoseira	Uppermost Late Cretaceous	65.5[e]	Ambwani *et al*. (2003), Wolfe & Edlund (2005)
Chaetoceros	Paleocene: *c.* 65-55	55	Fenner (1991)
Coscinodiscophyceae	Early Cretaceous (Aptian-Albian): 115-110	110	Gersonde & Harwood (1990), Harwood & Gersonde (1990)
Cyclotella	Early Eocene: 24	24	Bradbury & Krebs (1995)
Cymatosira	Early Eocene: *c.* 50-55	50	Homann (1991), Fenner (1994)
Diatoma	Late Eocene to Oligocene	33.9	Lupkina & Dolmatova (1975)
Dimeregramma	Miocene	5.33	Schrader & Fenner (1976), Reháková (1980)
Eunotogramma	Late Cretaceous	65.5	- (p) 377 Sims *et al.* 2006)
Fragilaria	Late Eocene: *c.* 45-40	40	Lohman & Andrews (1968)
Grammatophora	Late Eocene: *c.* 45-40	40	Desikachary & Sreelatha (1989), Edwards (1991)
Hyalodiscus	Cretaceous (Albian-Campanian)	70.6	Tapia (1996), Tapia & Harwood (2002)
Mediophyceae	Early Cretaceous (Aptian-Albian): 115-110	110	Gersonde & Harwood (1990), Harwood & Gersonde (1990)
Odontella	Late Cretaceous	65.5	Hajòs & Stradner (1975), Harwood (1988)
Opephora	Early late Miocene	7.25	Van Landingham (1985)
Pennates	Late Cretaceous (Campanian): 75	75	- (p) 381 Sims *et al*. 2006)
Rhabdonema	Late Eocene: *c.* 45-40	40	Desikachary & Sreelatha (1989), Edwards (1991)
Rhaphoneis	Late Eocene: *c.* 45-40	40	Andrews (1975)
Raphid diatoms	Late Cretaceous (Maastrichtian)	65.5	Singh *et al*. (2007)
Rhizosolenia[f]	Late Cretaceous (Upper Turonian): 91.5 ± 1.5	90-93[g]	Shinninghe-Damsté *et al*. (2004)
Staurosira[h]	Miocene	5.33	Hajòs (1968)
Stephanodiscus	Miocene	5.33	VanLandingham (1967)
Stephanopyxis	Late Cretaceous (Late Cenomanian-Santonian): 95-80	80[i]	Tapia (1996)
Surirellaceae	Middle Miocene	16.61	Reinhold (1937), Hajòs (1968 1986)
Thalassiothrix	Early Oligocene	28.4	Sims *et al.* 2006, p. 388

[a] Ages are shown if indicated in Sims *et al.* (2006).
[b] Minimum ages were taken if range is shown in Sims *et al*. (2006). Otherwise dates are obtained by assigning them to top of reported chronostratigraphic unit in the Geologic Time Scale of Gradstein & Ogg (2004).
[c] Source literatures cited by Sims *et al*. 2006 are indicated except for Singh *et al.* (2007) which was published later. If the page number is shown in of Sims *et a*l. (2006) then that is new original information.
[d] Used for the constraint of *Asteroplanus* being regarded as a first appearance of the clade 7 in the *SSU* supertree.
[e] Early Cretaceous marine genus *Archepyrgus* can be an ancestor of *Aulacoseira* (Gersonde & Harwood 1990, Harwood & Nikolaev 1995, Sims *et al*. 2006, however the first appearance of *Archepyrgus* at 110 Ma was not used for a calibration point of *Aulacoseira* in this analysis because this assumption violates the maximum node constraint of *Rhizosolenia at* 93 Ma which diverged earlier than *Aulacoseira* in our cladogram.
[f] Not fossil but abrupt increase of C_{25} HBI alkene.
[g] Calibrated as minimum and maximum constraints.
[h] Used for the constraint of *Pseudostaurosira* being regarded as a first appearance of staurosioid form.
[i] Although Sims *et al.* (2006) found *Stephanopyxis* in the slide of Lower Cretaceous sediments of ODP Site 693, Harwood *et al.* (2007) p. 35. regards this as a contamination of Oligocene specimens but Sims does not (Sims pers comm).

Clade 2 – Family Licmophoraceae, Family Cyclophoraceae and Family Protoraphidaceae

GENERA INCLUDED IN THIS CLADE IN THE *SSU* SUPERTREE: *Cyclophora* Castracane, *Licmophora* Agardh, *Protoraphis* Simonsen.

Instead of an apical pore field that consists of small round pores, these genera have a row of openings at the valve end, usually in a strict tetragonal or hexagonal array. The elongation of the opening varies from elliptic pores (*Licmophora*, Honeywill, 1998) to long slits (*Cyclophora*, p. 439, Round *et al*. 1990, *Protoraphis*, Gibson, 1979a). These genera are marine epiphytic taxa (most *Licmophora* species, Honeywill 1998, Terasaka *et al*. 2005) or epizoic (some *Licmophora* species attach to copepods or whales, p. 404, Round *et al*. 1990, *Protoraphis* attaches to copepods, Hiromi *et al*. 1985). Labiate process are generally located at the end of valve, but their number varies greatly. The new genus *Astrosyne* is radial and thus appears to be another genus/clade whose members have lost the ability to make the bands that squeeze the auxospore to give it its polar shape (Ashworth *et al*. 2012).

EXPECTED/PREDICTED GENERA TO BE INCLUDED IN THIS CLADE: *Falcula* Voigt (Takano 1983, p. 388, Round *et al*.

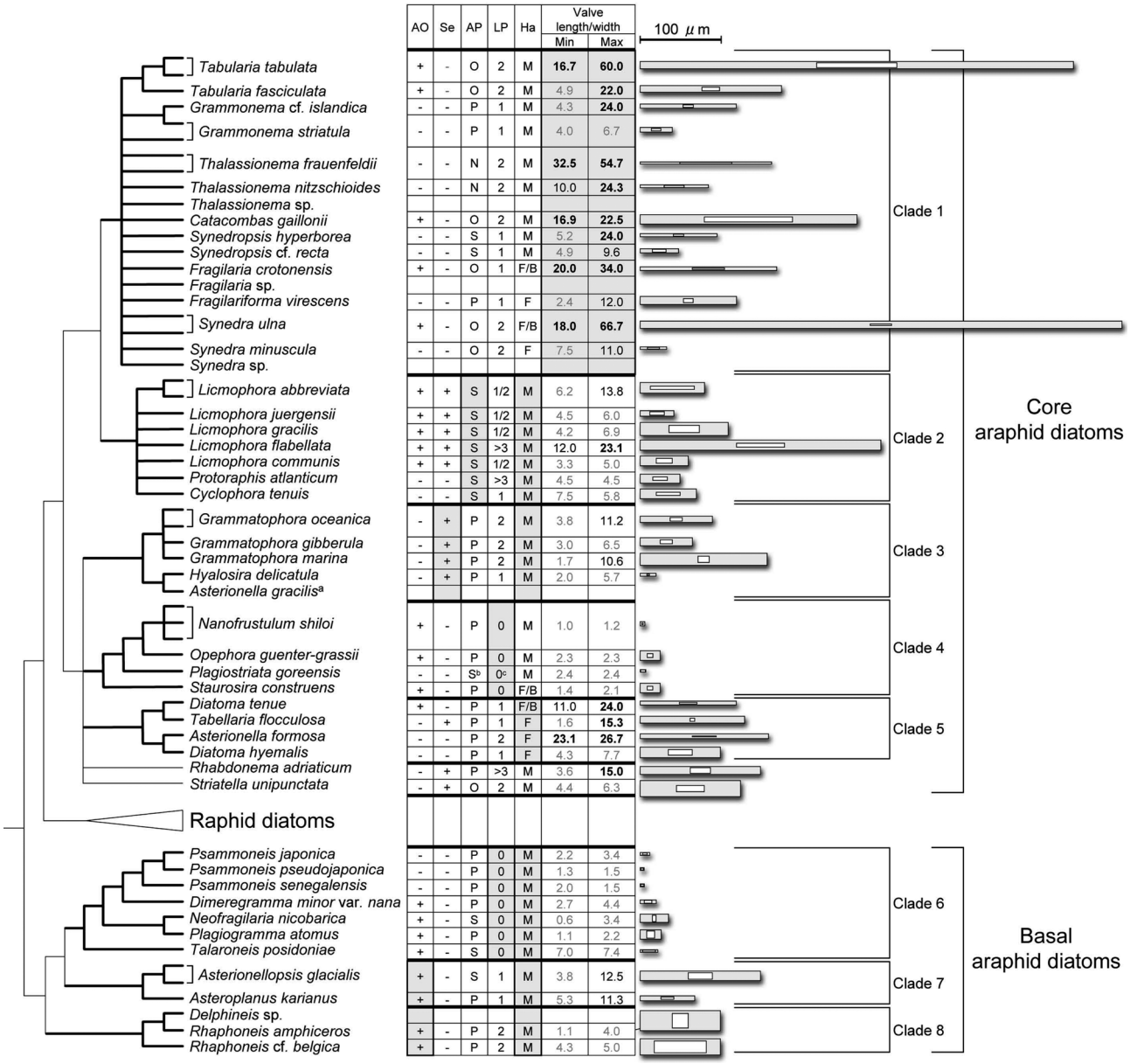


Fig. 1. – Supertree of pennates (but only araphid diatoms are emphasized) constructed from various *SSU* trees published so far (see Materials and methods for source topologies). Each clade was highlighted with bold line in the tree). Explanations of characters plotted here are as follows: Areolar Occlusion (AO) is a membrane-like structure occluding the pore of the valve, presence or absence is plotted); Septum (Se) is formed at girdle band extending inward the cell, presence or absence is plotted); Apical Pore area (AP) locates at the end(s) of valve usually possessing discrete area for mucilage secretion, such as Ocellulimbus (O) - depressed below the surface of the valve; apical Pore field (P) - aggregated small pores; apical Slit area (S)- one row of slits arranged in parallel; or No such area (N). Labiate Process (LP) is a structure situated inside of the valve; the number of LP is plotted. Habitat (Ha) ranges from Marine (M), Brackish (B), to Fresh water (F). The degree of valve elongation is shown as the length/width ratio, which is classified into three categories *sensu* Medlin *et al*. (2008b): l/w ratio greater than 15:1 (in bold black), 8:1 to 15:1 (black), and less than 8:1 (gray) The ratio was measured on maximum and minimum valve size of each species). Shadowed boxes are discriminative characters of each clade). Shared diagnostic character in each clade is emphasized by shadow. Also maximum and minimum size ranges of valve are illustrated based on the measurements of the original descriptions and some floristic works. It should be noted that the outer and inner squares do not correspond to the valve outline of each species but to the maximum and minimum size. Thus, all valve outlines fit within the gray area between outer and inner squares. Kooistra *et al*. (2003) misidentified *Protoraphis atlanticum* as *Pseudohimantidium pacificum*. Also the clade comprising three OTUs of *Microtabella* sp. in their paper was coded as *Hyalosira delicatula*. In Medlin *et al*. (2000) *Grammatophora* was coded as *Grammatophora oceanica*; *Diatoma* was coded as *Diatoma tenuis*; *Microtabella* was coded as *Hyalosira delicatula*.All environmental and endosymbiont sequences were removed from the analyses. [a] *Asterionella gracilis* (AY485447) in Shinninghe-Damsté *et al*. (2004) and Sorhannus (2007) is a *Hyalosira* species, see CCMP 469 which is the source of this sequence. [b] The apical slit of *Plagiostriata* is unique in shape (Sato *et al*. 2008e). [c] The LP of *Plagiostriata* is highly reduced (Sato *et al*. 2008e).

1990), *Hustedtiella* Simonsen (Simonsen 1960, Crawford *et al*. 1993), *Pseudohimantidium* (Gibson 1979b, p. 446, Round *et al*. 1990), *Sceptronema* Takano (Takano 1983), *Astrosyne* (Ashworth *et al*. 2012).

Clade 3 – Family Grammatophoraceae

GENERA INCLUDED IN THIS CLADE IN THE *SSU* SUPERTREE: *Grammatophora* Ehrenberg, *Hyalosira* Kützing, in part.

Septate girdle bands, marine epiphytic life forms, zigzag/straight chains attached by means of mucilaginous pads secreted from apical pore fields, labiate processes located at the one or both end(s) of valve (Navarro & Williams 1991, Sato *et al*. 2004a) are the features that characterize this clade. Septate girdle bands are observed in many lineages of diatoms (see Round *et al*. 1990 for *Licmophora*, p. 404, *Tabellaria* Ehrenberg, p. 398, a raphid diatom *Diatomella* Greville, p. 558) so that it may have been acquired independently in each lineage. Some *Hyalosira* species are more closely related to *Rhabdonema*, whereas the species in Sato's tree and in that of Lobban & Ashworth (2014) is most closely related to *Grammatophora* and others have been moved back into *Microtabella*.

EXPECTED/PREDICTED GENERA TO BE INCLUDED IN THIS CLADE: *Fossula* Hasle Syvertsen & Quillfeldt (Hasle *et al*. 1996).

Clade 4 – Family Fragilariaceae, in part (= Staurosiraceae fam. nov.)

GENERA INCLUDED IN THIS CLADE IN THE *SSU* SUPERTREE: *Nanofrustulum* Round, Hallsteinsen & Paasche, *Opephora* Petit, *Plagiostriata* Sato & Medlin, *Staurosira* (Ehrenberg) Williams & Round.

These taxa are relatively small in cell size with a highly reduced labiate process (*Plagiostriata*, Sato *et al*. 2008e) or absent (*Nanofrustulum*, Round *et al*. 1999, *Opephora*, Sabbe *et al*. 1995, *Staurosira*, Williams & Round 1987). Apical pore fields are narrow. No epiphytic species are reported from this clade, they are either planktonic or bottom dwelling (epipelic/epipsammic/epilithic) life form. This will be described as a new family below because this clade is very distant genetically to true fragilariacean species, all of whom have labiate processes.

EXPECTED/PREDICTED GENERA TO BE INCLUDED IN THIS CLADE: *Belonastrum* Round & Maidana (Round & Maidana 2001), *Martyana* Round (p 362, Round *et al*. 1990), *Pseudostaurosira* Williams & Round (Williams & Round 1987, p. 356, Round *et al*. 1990), *Pseudostaurosiropsis* Morales (Morales 2001), *Punctastriata* Williams & Round (Williams & Round 1987, p. 358, Round *et al*. 1990), *Sarcophagodes* Morales (Morales 2002), *Staurosirella* Williams & Round (Williams & Round 1987, p. 352, Round *et al*. 1990), *Synedrella* Round & Maidana (Round & Maidana 2001), *Trachysphenia* Petit (p. 384, Round *et al*.1990),

Clade 5 – Family Fragilariaceae in part and Family Tabellariaceae

GENERA INCLUDED IN THIS CLADE IN THE *SSU* SUPERTREE: *Asterionella* Hassall, *Diatoma* Bory, *Tabellaria* Ehrenberg ex Kützing.

These taxa inhabit non-marine environments (brackish/limnic) except for few species of *Diatoma,* attached by making zig-zag or stellate chains, a colony shape often observed in many lineages of araphid diatoms. Valves have slightly sunken apical pore fields with pores not really well differentiated from the striae areolae. The position of labiate process is not restricted to the end of valve.

EXPECTED/PREDICTED GENERA TO BE INCLUDED IN THIS CLADE: *Distrionella* Williams (Williams 1990), *Meridion* Agardh (Williams 1985, p 368., Round *et al*. 1990), *Oxyneis* Round (p. 402, Round *et al*. 1990), *Tetracyclus* Ralfs (Williams, 1987, 1989, p. 400, Round *et al*. 1990), *Thalassioneis* Round (p. 386, Round *et al*. 1990).

Clade 6 – Family Plagiogrammaceae

GENERA INCLUDED IN THIS CLADE IN THE *SSU* SUPERTREE: *Dimeregramma*, *Neofragilaria*, *Plagiogramma*, *Psammoneis*, *Talaroneis*.

This clade corresponds to the Family Plagiogrammaceae, which appears in marine coastal region, being largely psammic, although *Talaroneis* is epiphytic (Kooistra *et al*. 2003b). The areolae on their valve bear perforated rotae occlusions, with no labiate processes.

EXPECTED/PREDICTED GENERA TO BE INCLUDED IN THIS CLADE: *Desikaneis* Prasad & Livingston (Prasad & Livingston 1993), *Glyphodesmis* Greville (Sullivan 1988, p. 240, Round *et al*. 1990), *Hyaloneis* Amspoker (Amspoker 2008), *Psammogramma* Sato & Medlin (Sato *et al*. 2008f).

Clade 7 – Family Fragilariaceae in part (= Asterionellopaceae, fam. nov.)

GENERA INCLUDED IN THIS CLADE IN THE *SSU* SUPERTREE: *Asterionellopsis* Round, *Asteroplanus* Gardner & Crawford.

These genera are marine planktonic. Valves are heteropolar, and the narrower end has a labiate process. Cells attach each other by broader end of valves to make stellate, undulate chains (Crawford & Gardner 1997). This clade is described as a new family here because it is genetically very distant from true fragilariacean diatoms and from *Asterionella*.

EXPECTED/PREDICTED GENERA TO BE INCLUDED IN THIS CLADE: *Bleakeleya* Round (p. 394, Round *et al*. 1990).

Clade 8 – Family Rhaphoneidaceae

GENERA INCLUDED IN THIS CLADE IN THE *SSU* SUPERTREE: *Rhaphoneis* Ehrenberg, *Delphineis* Andrews.

This clade corresponds to the Family Rhaphoneidaceae. They all appear in marine coastal regions, being largely psammic but *Delphineis* and *Neodelphineis* are planktonic (e.g., Sar *et al*. 2007 and Hernández-Becerril 1990,

respectively). The areolae on their valve bear perforated rotate occlusions. Labiate processes are small.

EXPECTED/PREDICTED GENERA TO BE INCLUDED IN THIS CLADE: *Adoneis* Andrews & Rivera (Andrews & Rivera 1987), *Diplomenora* Blazé (Blazé 1984, p. 408, Round *et al*. 1990), *Neodelphineis* Takano (Takano 1982, p. 412, Round *et al*. 1990), *Perissonoë* Andrews & Stoelzel (p. 412, Round *et al*. 1990, Watanabe *et al*. 2007), *Psammodiscus* Round & Mann (Round and Mann, 1980, p. 418, Round *et al*., 1990), *Sceptroneis* Ehrenberg (p. 416, Round *et al*., 1990).

Phylogeny of the diatoms

Estimation of saturation

For further analysis of the araphid diatoms, Sato (2008) selected four additional gene markers. Because two of these were protein coding genes, it was necessary to determine if these were saturated at the third codon position. The saturation plot showed a serious saturation in *LSU* and the third codon of *rbc*L (Fig. 2B, C). Therefore, Sato used only transversions in his phylogenetic analyses,

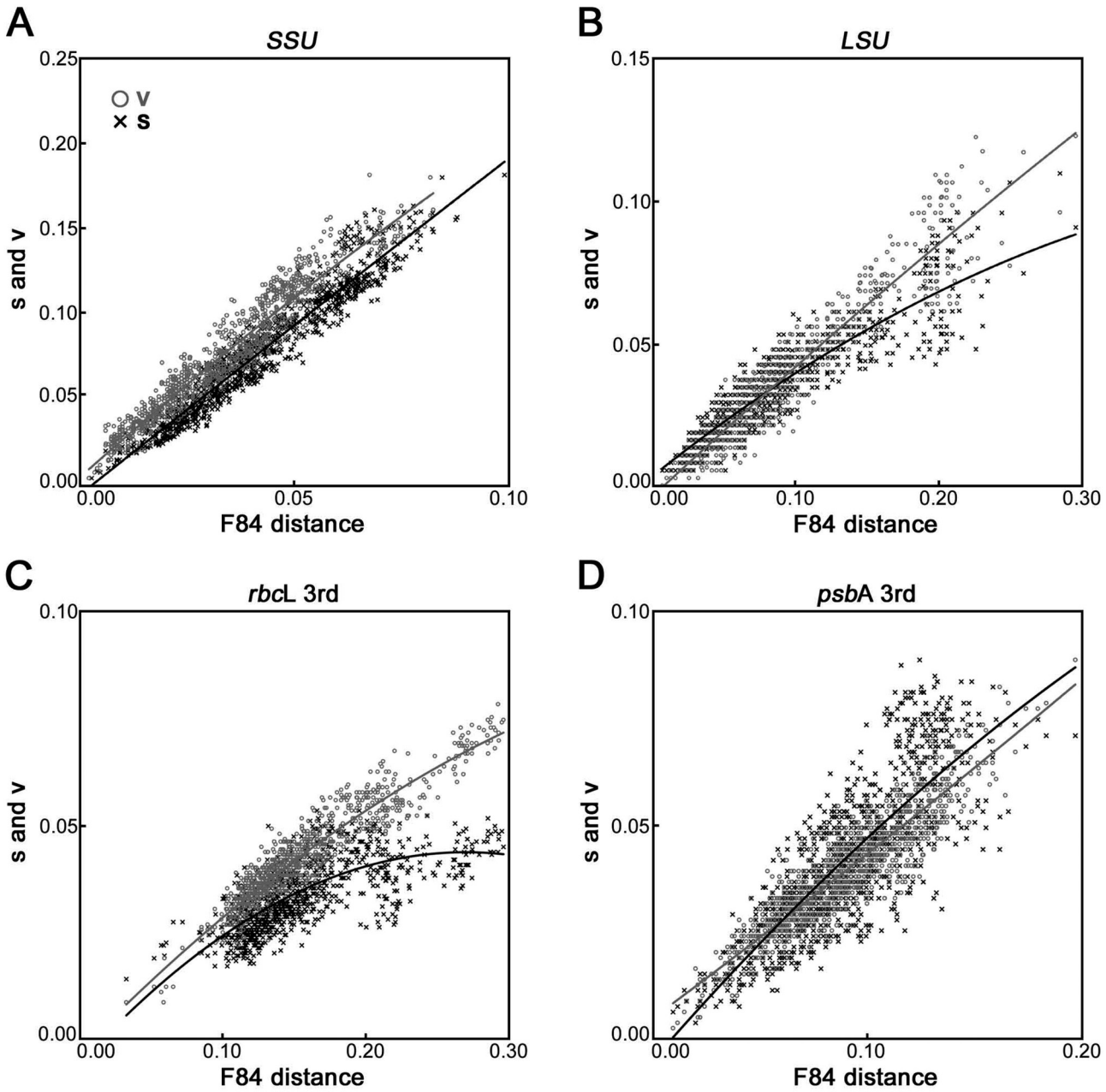


Fig. 2. – Transitions (s) and transversions (v) of *SSU* (**A**), *LSU* (**B**). Third codon position of *rbc*L (**C**) and *psb*A (**D**) sequences plotted versus F84 genetic distance. Each black cross or gray circle in the field represents a transition or transversion, respectively. Saturation can be seen in transition of B and C but not A and D.

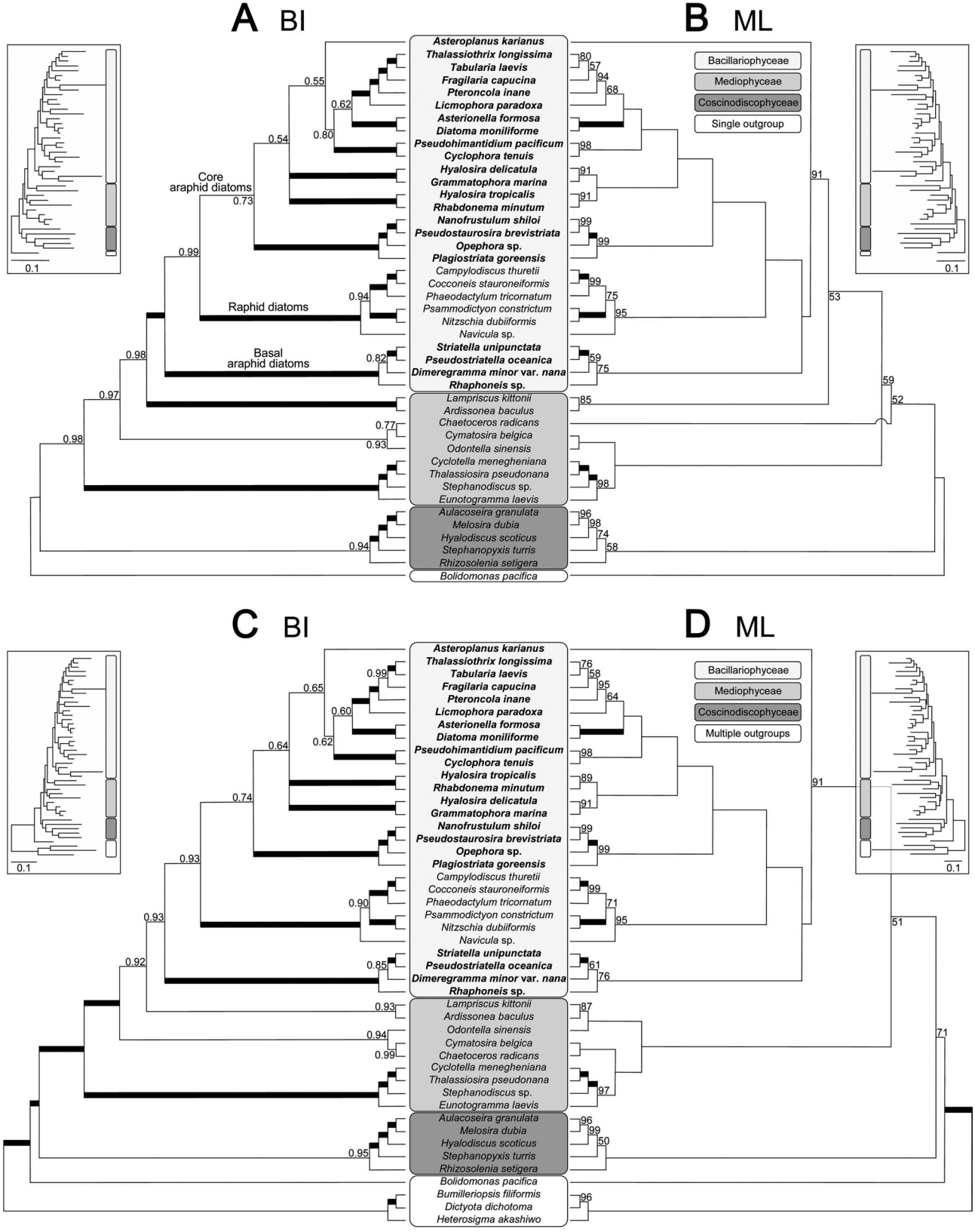


Fig. 3. – Evolutionary relationships of diatoms inferred from DNA sequences of *SSU* and *LSU rbc*L and *psb*A. Note *LSU* and third codon position of *rbc*L were re-coded into RY to ignore saturated transitions. Phylogenies inferred from Bayesian inferences (A C and Maximum Likelihood analyses (B D Trees are rooted with single outgroup *Bolidomonas* (A B or multiple outgroups of Heterokonta (C D Branch length of each topology can be seen in the small boxes. Each class in the current systematics is indicated by color. Three clades of pennates i.e. basal araphid, core araphid and raphid diatoms are shown in A, which are now recognised at the subclass level (Medlin 2014a). Taxon labels in bold cases are araphid diatoms. The thick nodes represent 1.00 Bayesian posterior probability in BI (**A**, **C**) or 100 % bootstrap value in ML analyses (**B**, **D**).

by coding purine (G and A) into R, and pyrimidine (C and T) into Y.

Single gene phylogenies

Single gene trees from Sato (2008) are provided in the supplemental material. None of the single gene phylogenies were congruent with one another with regards to the monophyly of the Coscinodiscaceae or the Mediophyceae. Only the pennate diatoms, as usual, were monophyletic and in the *SSU* analysis, the araphids were monophyletic, but with the basal araphids still basal in the pennate clade. In the *LSU* tree, araphid diatoms were divided into basal and core group, except for *Asteroplanus*, which appeared in centric lineage. The ancestral position of raphid diatoms in the tree, that has never been supported in the other phylogenetic studies nor fossil records, indicated that the sole use of *LSU* is insufficient for the phylogenetic reconstruction of diatoms (see similar discussions in Bruder & Medlin 2008). The *rbc*L tree shows that one of the clade 2 araphid diatoms (*Cyclophora* and *Pseudohimantidium*) were sisters to the curious clade comprising centric and raphid diatoms, and the rest of araphid diatoms were monophyletic. The araphid diatoms were polyphyletic in the *psb*A phylogeny, in which they were scattered in the tree, nevertheless, most of the supertree clades recovered by Sato (2008) were monophylytic.

The combined analyses of nuclear rDNA (*SSU+LSU*) vs. plastid (*psa*A +*rbc*L) genes were also presented in the supplementary figures, again giving different results. In Sato's rDNA topology, a clade of *Striatella* and *Pseudostriatella* were sisters to the monophyletic raphid diatoms (also recovered by Li *et al*. 2015), and the rest of the araphid diatoms formed a clade. The plastid tree displayed paraphyletic divergence of araphid diatoms, in that the clade of raphid diatoms emerged relatively diverged position of pennates being rooted by *Asteroplanus*. In the combined SSU and chloroplast trees by Theriot *et al*. (2015), they were recovered with the basal araphids.

Four gene phylogeny

Following RY recoding of the protein and LSU genes, Sato (2008) performed BI, ML, NJ, and MP analyses with two outgroups: Bolido-root and Distant-root to evaluate the effect of distant vs. close outgroups. The monophyly of the diatoms had been recovered in the most analyses (Fig. 3), only one exception was found in MP Distant-root analysis in which a Coscinodiscophycean diatom, *Stephanopyxis turris,* diverged before *Bolidomonas* and all other of diatoms (Topology not shown). MP analyses were not weighted analyses and this is likely the cause of these positional problems (Medlin 2014). The constrained topologies of the monophyly of the 'centrics' and 'araphid diatoms' performed by Sato (2008) were significantly rejected (Table IV).

In the recoded four gene phylogeny by Sato (2008), the class Coscinodiscophyceae was monophyletic in all analyses, but BS support in ML analyses were low (Fig. 3) with an identical topology in all analyses: (*Rhizosolenia*, (*Stephanopyxis*, (*Hyalodiscus*, (*Melosira*, *Aulacoseira*)))). Class Mediophyceae was monophyletic only in the ML Distant-root analysis with BS support < 50, whereas the class was resolved as a grade in the trees in the first BI analysis but as monophyletic in the second, non recoded Distant-root BI analysis (Fig. 4). In trees from the ML Bolido-root analysis, a clade of *Eunotogramma* and Thalassiosirales (*Stephanodiscus*, (*Thalassiosira*, *Cyclotella*)) always diverged at the root of the subdivision Bacillariophytina, suggesting that the radial valve of Thalassiosirales may not a consequence of secondary loss of bipolar shape, as suggested before based on *SSU* phylogenies, where it usually appears as a diverged lineage (Sims *et al*. 2006, Alverson *et al*. 2011). In the new uncoded BI analysis per-

Table IV. – Tests of Bayesian topologies versus alternative topologies derived from other optimization methods (maximum likelihood parsimony and neighbor joining trees as well as maximum likelihood searches with constraint assuming the Bayesian topology as best tree.

Topology	-ln L	Difference	KH-test	SH-test
Bayesian tree[a] (Figs 3A C)	36823.34274			
Maximum likelihood tree				
Bolido-root (Fig. 3B).	36832.10415	8.76141	0.685	0.892
Distant-root (Fig. 3D).	36835.36331	12.02057	0.599	0.824
Maximum parsimony tree				
Bolido-root	37039.32655	215.98381	< 0.001[b]	0.140
Distant-root	36848.35294	25.01019	0.353	0.733
Neighbor joining tree				
Bolido-root	36940.31975	116.97701	< 0.001[b]	0.329
Distant-root	36981.26151	157.91876	< 0.001[b]	0.233
Constrained trees				
Centrics as monophyletic	39693.97789	2870.63515	< 0.001[b]	< 0.001[b]
Araphid diatoms as monophyletic	39967.73758	3144.39483	< 0.001[b]	< 0.001[b]

The presented analyses test *a posteriori* hypotheses a situation in which the Kishino-Hasegawa-test (KH-test, Kishino & Hasegawa 1989 is not appropriate. Therefore we also used the Shimodaira-Hasegawa-test (SH-test, one-tailed error probability, Shimodaira & Hasegawa 1999. Nevertheless for comparative seasons we give the two-tailed error probabilities for the KH-test.

[a] Bayesian topologies were identical with either outgroup selection, a single (*Bolidomonas pacifica*. or multiple (*Heterosigma akashiwo Dictyota dichotoma Bumilleriopsis filiformis* and *Bolidomonas pacifica*. outgroup.

[b] Hypotheses that were rejected at $P = 0.05$.

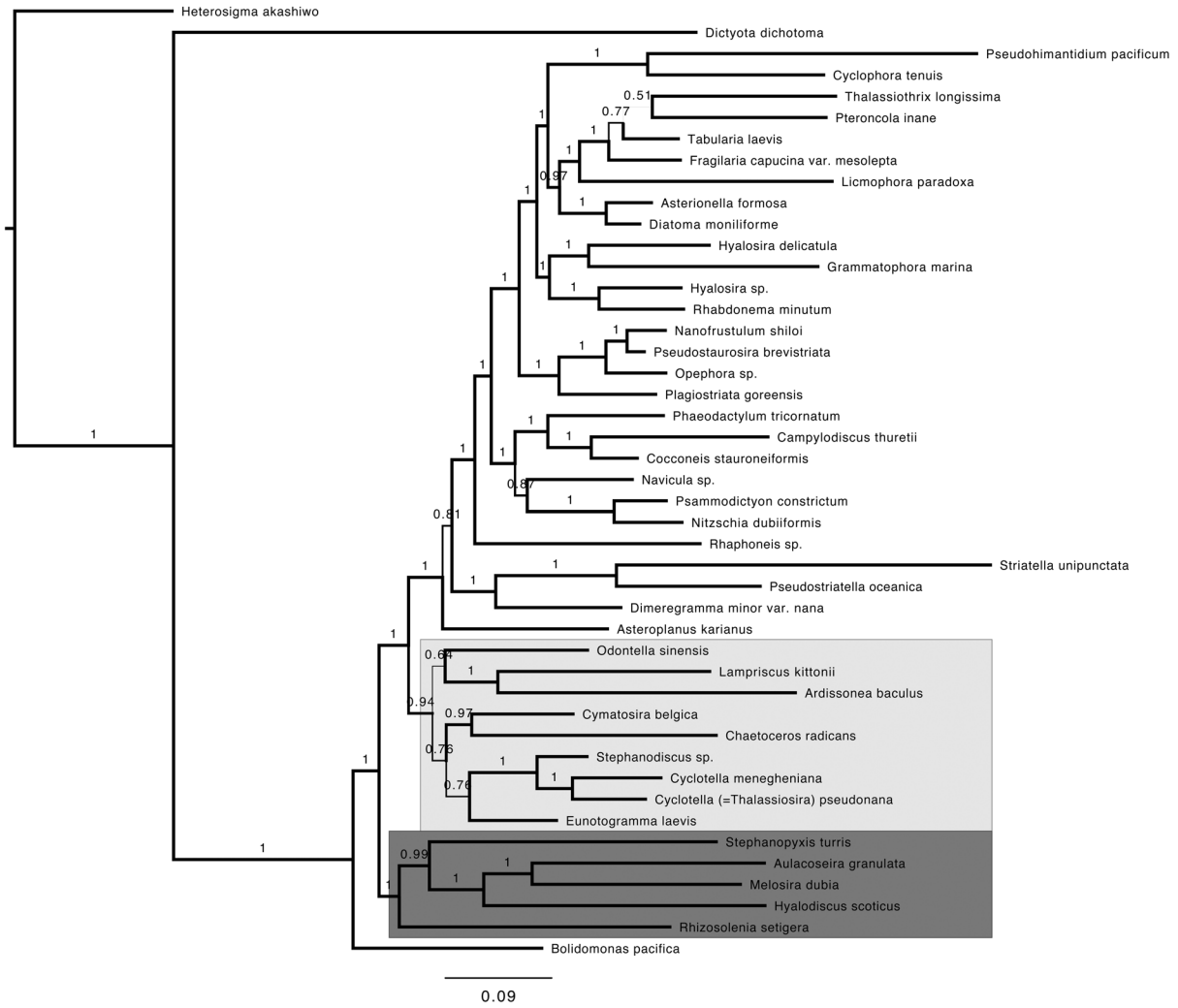


Fig. 4. – Evolutionary relationships of diatoms inferred from DNA sequences of *SSU* and *LSU rbc*L and *psb*A with RY recoding. Phylogenies inferred from Bayesian inferences multiple outgroups of Heterokonta but excluding *Bumilleriopsis* because the models used in this analysis were sensitive to missing data and the LSU gene was missing from this taxon.

formed here (Fig. 4), it appeared as the last divergence, supporting that they have likely lost the ability to make bands and have reverted to the ancestral state of radial valves. Thus its position is highly dependent on the type of analysis performed. A highly elongated diatom *Ardissonea* (Medlin *et al*. 2008b) always diverged last among mediophycean diatoms as sister to *Lampriscus* at the root of the pennate lineage with high BT support, in the ML Distant-root analysis they were part of the monophyletic mediophycean clade (Fig. 3). The relationship of the rest of mediophycean genera, *Chaetoceros*, *Cymatosira* and *Odontella*, remained unresolved in the first BI and ML analyses with bolido-root dataset but were well supported in the both BI analyses with multiple outgroups. Unsupported relationships for these genera were recovered in Theriot *et al*. 2015 with 6 plastid genes with a single outgroup.

In other analyses of multiple genes (3 genes in Ashworth *et al*. 2012), the mediophycean genera have been recovered as monophyletic but the coscinoodiscophycean diatoms were still a grade with only one outgroup, *Bolidomonas*. In Li *et al*. (2015), Coscinoodiscophyceae were monophyletic with high bootstrap support with two bolidomonads strains as outgroups. This study also had a high proportion of araphid taxa relative to the remainder of the diatoms (52 %) and it is likely that a higher proportion of araphid taxa has influenced the monophyly of the classes with these two sequences as outgroups. In the six gene analysis by Theriot *et al*. (2015) both classes are recovered as grades of clades with only one bolidomonad outgroup.

The pennates formed a robust clade with essentially the same topology in all analyses (Fig. 3), although the monophyly of *Grammatophora, Rhabdonema,* and two species of *Hyalosira* was not recovered in first coded BI analyses with both data sets done by Sato (2008) but in the non-RY coded dataset with multiple outgroups performed here and in the 6 gene phylogeny by Theriot *et al*. (2105) they

were highly supported (Fig. 4). The only incongruence in the coded dataset analysis was the position of the araphid diatom, *Asteroplanus*, which appeared in the core araphid lineage in the first BI analysis but at the root of the pennates in ML. The genera normally recovered as basal araphids were separated into three sequential clades prior to the core araphids diverging as sister to the raphid diatoms in the uncoded BI analysis done here. To elucidate the phylogenetic position of *Asteroplanus,* affinities of the genus should be included in the further analysis, such as *Asterionellopsis*, shown as the sister genus in *SSU* tree (see Fig. 1) and *Bleakeleya*, a putative sister genus judged by its morphology (p. 394-395, Round *et al*.1990). Theriot *et al*. (2015) place *Bleakeleya* in a clade with *Koeneriella* and *Perideraion,* separate from *Asterionellopsis*.

Some of the clades in the super tree were separated in the four gene analysis. The interclade relationships, which were unresolvable in the *SSU* supertree, were fully resolved in the four gene analyses done by Sato and done here in this study. The following divergence patterns of pennates were commonly observed in all analyses (Figs 3, 4). A robust clade of marine araphid diatoms, the basal araphids, (*Rhaphoneis*, (*Dimeregramma*, (*Striatella*, *Pseudostriatella*)) diverged first from the remaining pennate lineage. Then, the raphid diatoms' clade (*Navicula*, ((*Nitzschia*, *Psammodictyon*), (*Phaeodactylum*, (*Cocconeis*, *Campylodiscus*)))) diverged from the rest of the core araphid diatoms. The first divergence of the core araphid diatoms was a clade comprising small-celled diatoms (*Plagiostriata*, (*Opephora*, (*Pseudostaurosira*, *Nanofrustulum*)), which is equivalent to clade 4 in the *SSU* supertree. Subsequently in ML, a marine epiphytic clade diverged ((*Hyalosira delicatula*, *Grammatophora*), (*Hyalosira tropicalis*, *Rhabdonema*)). The recovery of this clade, even if only in ML analyses, is remarkable because *Grammatophora* and *Rhabdonema* share anisogamic sexual reproduction, which has never been observed in other diatoms. Ameboid gamete movement is also seen in other araphid diatoms where the hair like appendage attached to the male gamete has been documented (Sato *et al*. 2012). The clade was, however, not recovered in BI analyses and weakly supported by the other characters: it has no exclusive character but can only be characterized by the combination of some morphological and ecological features (see above for the characters of the clade 3 for more details). The topologies of four gene analyses strongly indicated a paraphyletic nature of the genus *Hyalosira*, whose generitype *H. delicatula* formed a clade with *Grammatophora*, whereas the other species, *H. tropicalis*, which lacks septate girdle bands unlike the other member members of *Hyalosira*, formed a clade with *Rhabdonema* (indicated as New Clade in Fig. 4), and this link has never been suggested in the previous *SSU* phylogenies. *H. tropicalis* should be described new genus, whereas *H. interrupta* has been restored to *Microtabella* (Lobban & Ashworth 2014). Subsequently, a clade of *Cyclophora* (epiphytic) and *Pseudohimantidium* (epizoic) diverged. The only shared character in this clade was their attached marine habitat and a slit structure located at the ends of their valve. The placement of *Pseudohimantidium* (and also *Protoraphis* as indicated in the *SSU* supertree) within the core araphid clade rejects the hypothesis that a raphe-like slit in these diatoms, called a labiate groove, is a precursor of the raphe (Simonsen 1970). Following the separation of a non-marine clade comprised comprising *Diatoma* and *Asterionella*, *Licmophora* diverged from clade of elongated araphid diatoms, which included *Pteroncola* at the root (*Pteroncola*, (*Fragilaria*, (*Tabularia*, *Thalassiothrix*)). In this four gene analyses, *Licmophora* separated from the remaining clade 2 species, although, in the *SSU* supertree the clade is supported by habitat and apical slit structure. Medlin *et al*. (2008a) and Sims *et al*. (2006) found *Licmophora* separated into two clades depending on their means of attachment.

This study reconfirms the paraphyly of araphid diatoms. The three clades recovered within the pennate (= Class Bacillariophyceae) are 1) basal araphids, 2) core araphids sister to 3) raphid diatoms. Each deserves to have equal rank in diatom taxonomy. Medlin (2015) has described two new subclasses to accommodate these two clades that are consistently recovered in most molecular analyses. The only unstable feature is that, in some analyses, the position of some genera are not stable, appearing sometimes in the basal araphids and sometime as an independent clade or in the core araphids. Our current results are insufficient to establish the precise nature of araphid pennate phylogeny and lacking some morphologically interesting members whose affinities are yet unknown in our studies and in those of others, such as *Diprora* (Main, 2003), *Licmosoma* (Round & Alexander 2002), *Nephroneis* (Amspoker 1989), *Omphalopsis* Greville (Greville 1863, Williams 1988), and *Porannulus* Hamilton & Poulin (Hamilton *et al*. 1997). However, our groupings in both the supertree and the four gene analysis point the way to which groups need further work in order to find defining morphological features to erect new taxa.

This study represents the only study in which the effect of single vs multiple outgroups has been tested with multiple genes outside of the SSU genes (Medlin 2014). Because our analyses stopped when the standard deviation between the duplicate runs dropped below 0.01, the number of generations that each analysis used to resolve a congruent tree differed between the two data sets. With only bolidomonads as the root 810,000 generations were used to resolve the tree, whereas with the multiple outgroups were used only 500,000 generations were needed. Thus, the multiple outgroups shortened the time to resolve the tree and monophyletic clades were recovered in the ML analyses and in the BI analysis with no RY coding. Medlin (2014) has extensively tested multiple outgroups with the SSU gene alone and found that monophyly of the

centric classes was only obtained with distant (outside the heterokonts) outgroups.

Assessment of basal araphid position of Striatella/ Pseudostriatella *based on auxospore structure*

An early divergence of the clade of *Striatella* and *Pseudostriatella* in four gene phylogeny among the pennates has never been shown by previous *SSU* studies, in these analyses, the position of *Striatella* has always been unstable and has even appeared within the raphid lineage (see Medlin & Kaczmarska 2004, Sato *et al*. 2008c) but has never appeared within in the basal araphid clade. Li *et al*. (2015) placed them as sister to the raphid clade with weak bootstrap support (50 %). Theriot *et al*. (2015) in their combined *SSU* and plastid analysis placed *Striatella* in the basal araphid clade, weakly supported to a *Bleakeleya* clade. Nevertheless, its placement in this clade can be supported by the morphological features of its auxospore, which is a special cell of diatoms generally produced only after the sexual reproduction, and known to have evolutionary significance (Medlin & Kaczmarska 2004, Kaczmarska *et al*. 2013). The auxospore of *Pseudostriatella* (Sato *et al*. 2008c) is largely covered with a structure called 'properizonium', consisted of saddle-shaped closed bands, and this is observed in some bipolar Mediophycean diatoms, *e.g*., *Chaetoceros* Ehrenberg (von Stosch 1982), *Lampriscus* Schmidt (Idei & Nagumo 2002), see table II in Medlin & Kaczmarska (2004) for a summary of auxospore morphologies of "centric" diatoms. Furthermore, interestingly, the ventral side of auxospore of *Pseudostriatella* has 'longitudinal perizonial bands' that is typically seen in the auxospore of the most of the raphid diatoms (*e.g*., Mann 1982, Nagumo 2003, Amato *et al*. 2005, Kaczmarska *et al*. 2007) and some core araphid diatoms (*e.g*., von Stosch 1982, Sato *et al*. 2008b). The other core araphid diatoms also have a structure that is likely homologous to the longitudinal perizonial bands lying at the ventral side of the auxospore (Sato *et al*. 2004b, 2008d). Therefore, *Pseudostriatella* possesses characters of both groups, implying that it has an intermediate state between Mediophyceae and the clade of core araphid/raphid diatoms, and supporting the result of four gene analyses. The perizonial structure of the family Rhaphoneidaceae and Plagiogrammaceae, closest relatives of *Striatella*/*Pseudostriatella*, yet remains unknown preventing further speculation. Nevertheless similar auxospore structures have now also been found in *Tabularia fasciculata, T. tabulata* (Davidovich *et al*. 2012) and in *Ulnaria ulna* (Davidovich 2012) and ameboid gametes are known also from *Grammatophora* and *Rhabdonema* (see references in Sato *et al*. 2011).

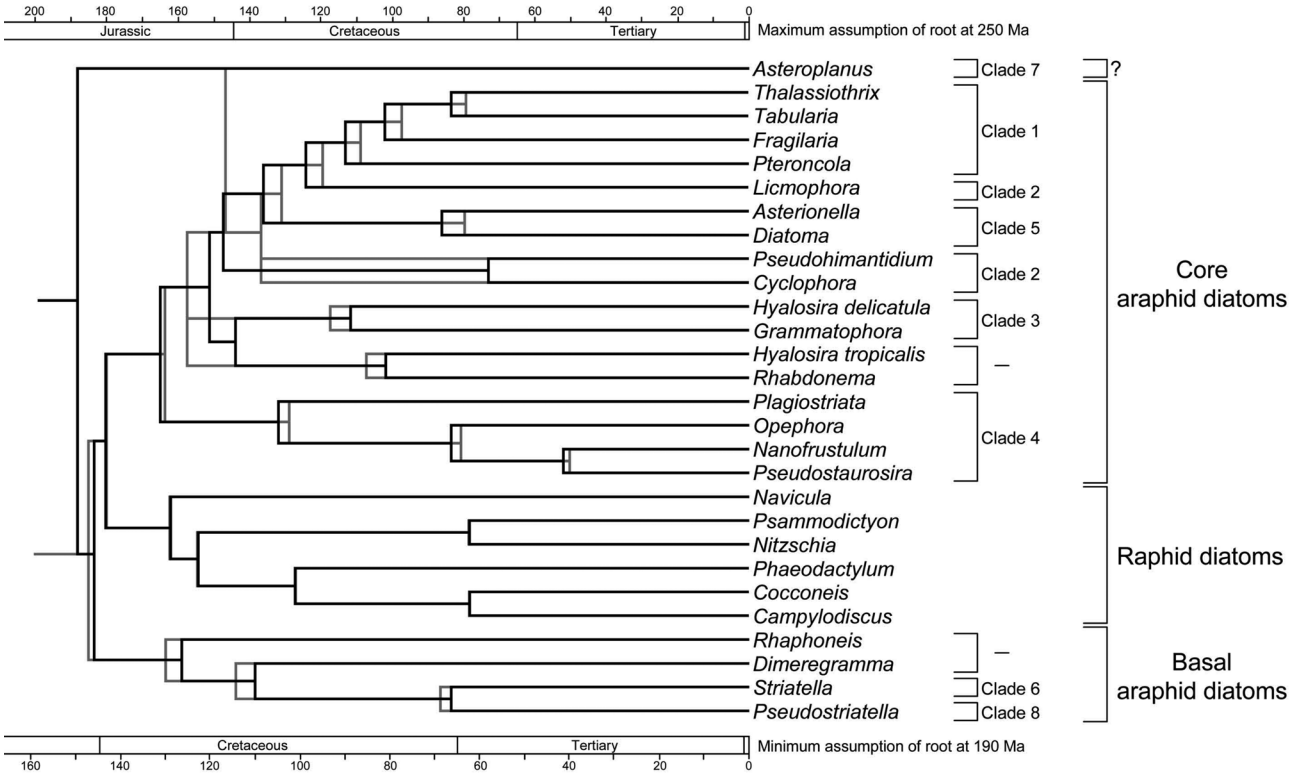


Fig. 5.- Molecular timescale for pennates of BI (gray) and ML (black. tree based on Bolido-root dataset in Fig. 3A and B, respectively). Priors of 250 and 190 Ma for the root of diatoms and respective time scales are indicated above and below topology. Clade in the *SSU* supertree in Fig. 1 is indicated. Basal and core araphid diatoms are also marked. Note the unresolved araphid, *Asteroplanus,* that belongs to basal araphid in BI but core araphid in a ML topology.

Divergence time estimation of pennate diatoms

Sato's results indicate that the early divergence of the pennates into three major clades, basal araphid, core araphid, and raphid diatoms, took place in a very short period (Fig. 5). All major clades of araphid diatoms (*i.e.*, clades 1 to 8) appeared by the end of the Cretaceous in all analyses.

Although two topologies (BI and ML) of pennates used in the divergence time estimation were not identical as regarding the position of *Asteroplanus*, the timing of the divergence was very similar (Fig. 5, compare gray and black line, respectively). Estimated divergence times of the pennate and raphid diatoms are summarized in Table V. Under the assumption of a diatom origin at 190 Ma, pennate emerged from 159.4 (BI) to 156.9 (ML) Ma and raphid diatoms emerged from 144.6 (BI) to 143.1 (ML) Ma. On the other hand, with the assumption of a 250 Ma origin, the emergence of the pennates was push back to 200.0 (BI) to 199.0 (ML) and the raphid origin from 180.2 (ML) to 179.9 (BI) Ma. Therefore, a possible range of the divergence time of the pennate is from 200.0-156.9 Ma, and that of raphid diatoms is 180.2-143.1 Ma. These ages greatly predate previous estimations based on *SSU*, where the origin of pennates at 125 Ma (Sorhannus 2007), 98 Ma (Berney & Pawlowski 2006) or 86 Ma (Kooistra & Medlin 1996) and the origin of raphid diatoms at 93.8 Ma (Sorhannus 2007). This data set was also used to date the divergence times of the three classes of diatoms (Medlin 2016a).

No pennate diatoms have been recovered in well-preserved Early Cretaceous floras studied to date (Harwood *et al*. 2007), *i.e.*, Early Albian flora from near the Antarctic margin (Gersonde & Harwood 1990, Harwood & Gersonde 1990), Australia (Dun *et al*. 1901, Harper 1977, Haig & Barnbaum 1978, Nikolaev *et al*. 2001), and Germany (Forti & Schulz 1932). Although these floras have been intensively examined, none of these deposits are situated in a benthic environment, which is known as the main habitat for the most of the extant members of pennates. Moreover, the Late Cretaceous is known for its abundance of pyritized diatoms (Sims *et al*. 2006). Thus, it is still possible that the earliest pennate has simply not been discovered. In fact, our estimation showed that the origin of the pennates between 200.0 and 156.9 Ma, greatly predating their first occurrence of fossils from the deposit of the Campanian (Sims *et al*. 2006). The origin of the raphid diatoms was also estimated 180.2 and 143.1 Ma predating their fossils from the deposit of the Maastrichtian (Singh *et al*. 2006). Furthermore, one of the earliest fossils of the raphid diatoms is now classified in the genus *Lyrella*, which is not a basal lineage within the raphid diatoms, either in 18S rDNA or *rbc*L analyses (Behnke *et al*. 2004, Jones *et al*. 2005). Therefore, if the molecular data are correct, the fossil record must significantly underestimate the age of the pennates, pushing back the origin of the entire pennate group (Sims *et al*. 2006).

Sato (2008) noted that he could not exclude any possibility that the results suffered from an analytical artefact. The reconstruction may have been influenced by the unequal taxon sampling in the present study, *i.e.*, known as node-density artefact (Fitch & Bruschi 1987, Fitch & Beintema 1990). As Sato's primary emphasis in the present study was 1) to reveal the phylogeny of the 'araphid diatoms', and 2) to resolve unsupported nodes in the supertree from the SSU gene, it can be expected that taxon sampling could be uneven but the taxa of interest were not undersampled at the time of his dissertation study because nearly half (22 out of 46 OTUs) of the taxa sample were these diatoms. It might be suspected the densely sampled area of a tree can cause overestimated branch lengths, pushing the node back in time using a divergence time estimation. Although a statistical method (the delta test) that detects the node-density artifact in trees (Webster *et al*. 2003, available at http://www.evolution.reading.ac.uk/pe/index.htm) detected no evidence for punctuated evolution or the node density artefact in his ML topologies (BI topologies were not subjected to the test because the program rejects trees containing polytomies), the influence of the unequal taxon sampling in this estimation could not completely rejected. However it should be noted that in the Li *et al*. (2015) study, which also concentrated on araphid diatoms and nearly 50 % proportion of the diatoms in their data set were araphid. Notably they recovered the two monophyletic centric classes with only a single bolidomonad outgroup, suggesting that a high percentage of araphids in the dataset could affect centric class monophyly.

Table V. – Divergence time of pennates and its credibility intervals for the nodes (in Ma. With the assumption of root at 250 - 190 Ma. 95 % credibility range in square brackets.

	Origin of araphid diatoms (= origin of pennates)	Origin of raphid diatoms
This study		
BI topology	200.0 [226.6-163.0] - 159.4 [174.6-138.9]	179.9 [205.8-146.2] - 144.6 [159.7-125.8]
ML topology	199.0 [224.5-166.3] - 156.9 [171.4-138.6]	180.2 [203.1-150.0] - 143.1 [157.6-125.7]
Previous studies		
Kooistra & Medlin (1996)	86 [-159]	-
Berney & Pawlovski (2006)	98 [110-77]	-
Sorhannus (2007)	125	93.8

Because a perfect coverage of fossil deposits of Mesozoic, and probably even Cenozoic age, will never be achieved, the time estimation using the fossils for calibration will always involve some potential errors. The large confidence intervals of these estimations (Table V) rule out further discussions as to whether any geological event drove this pennate radiation. Nevertheless near-shore habitats have not been adequately sampled to uncover missing fossils to help substantiate or refute Sato's dates.

Influence of different outgroups

In the ML analysis with Distant-Out dataset and the BI analysis of uncoded positions, three major monophyletic lineages were recovered as clades, supporting the idea of Medlin & Kaczmarska (2004) that the higher level relationships in the diatoms can best be reconstructed using multiple distantly related outgroups. It is notable that our *SSU* sequences were aligned by Sato separately from the ARB dataset, which was used in Medlin & Kaczmarska (2004), so that the possible bias that the outcome of this study was drawn mainly by phylogenetic signals involved in the *SSU* of Medlin & Kaczmarska (2004) can be excluded. In Sato's dataset, the more variable regions of the SSU were excluded and in that of Medlin and Kaczmarska (2004) and the SSU alignment here none of the bases were excluded because of the ARB alignment and the application of a weighting mask. In the analysis by Theriot *et al*. (2015), they did not use the full length SSU. To compare the results of using the entire SSU molecule vs. elimination of some variable sites, see the main tree in Li *et al*. (2015) vs. the Supplementary Fig. 4 in Medlin (2016a) and compare differences in BT support for the same data set but different **nt** numbers from their SSU alignment without the V4 region and hers with it and also trees in Medin (2014). Furthermore, saturation of phylogenetic information resulted in adding distant outgroups (Wheeler 1990, Maddison *et al*. 1992) is also unlikely because the saturation test run by Sato (2008) displayed linear distribution of plots in our genetic markers used in this study (Fig. 2). However, because grades of clades were recovered in three of the four Mediophycean topologies, further analyses with even larger numbers of outgroups is needed to resolve fully this issue. However, with the limited dataset analysed here, it is clear that the number and distance that the number of outgroups have to the ingroup influence the outcome of the ingroup monophyly. Medlin (2014) has explored the effect of more distant outgroups with the SSU gene alone and found considerable effect on the monophyly of the three classes.

SUMMARY

The combined analyses of the pennate diatoms continue to support the non-monophyly of the araphid diatoms with two groups being consistently recovered: basal araphids and core araphids. A formal description of these two groups as subclasses has been published (Medlin 2015). With the exception of *Striatella* and *Pseudostriatella* whose position in the *SSU* trees has been unstable, the diatom genera in basal araphid group has remained the same but now with *Striatella* and *Pseudostriatella* included. The unique auxospore structure in *Pseudostriatella* suggests that this may be the feature that unites this basal group of pennate diatoms who possess very different valve morphologies, and further investigations to support this hypothesis should be aimed at studying auxospore formation in the basal araphid diatoms. The presence of the hair like appendage on the male gamete has now been reported in both basal and core araphids (Sato *et al*. 2011) and likely can be used as a feature to define an araphid because auxospore formation is relatively conservative in the diatoms. Furthermore the two araphid clades can be distinguished by the type and arrangement of their bands (properizonium and perizonium). We do note that in the recent revision of terminology associated with auxospores (Kaczmarska *et al*. 2013), it has been recommended that the two terms be synonymized. We have kept them separate because they are arranged differently on the auxospore and as such define the differences between the basal and core araphids.

We continue to provide preliminary support that the number of distantly related outgroups affects the monophyly of the new centric classes. Medlin *et al*. (2008b) and Medlin (2014) also provided more evidence that the alignment by secondary structure rather than by primary sequence similarity will also affect the monophyly of the centric classes. Both of these attributes should be investigated with larger datasets to resolve this issue about the monophyly of the two centric classes. We would recommend in future work that for each class used as an outgroup, *e.g*. Phaeophyceae, that at least two – three species representing that class be used to avoid any criticisms about long branch attraction. Outgroups outside of the heterokonts should be taken and haptophyte and chlorophytes were among the outgroups that recovered monophyletic clades in the SSU gene analyses (Medlin 2014).

The four gene analysis and the super tree constructed from many *SSU* trees provide more structure to the araphid diatoms and the clades recovered, some of which have defining morphological and/or ecological features, should be revised at higher taxon levels, *e.g*., clade 7 with *Asterionellopsis* and *Asteroplanus*. The super tree alone recovered many clades that were unsupported, which with the addition of more genes to the analysis greatly improved the support for the clades. Other clades with not so obvious morphological features should be targeted for further investigations to determine if there are defining morphological features for these clades. The Family Fragilariaceae is scattered over 4 clades, and with now well defined clusters of taxa, we can begin to search for

characters to break up this large morphologically, seemingly homogenous family. We provide the description of two new families to resolve this issue. In much the same way, the cosmopolitan polymorphic species, *Skeletonema costatum* (Greville) Cleve, was divided into six new species with distinct biogeographies after preliminary phylogenetic analyses were performed (Zingone *et al*. 2005, Sarno *et al*. 2005, 2007, Kooistra *et al*. 2007a), although *S. tropicum* and *S. menzelii* appear to be truly cryptic species morphologically but clearly divided into multiple well supported molecular clades.

We recognise that our data set is heavily weighted in terms of araphids (22 out of 42 taxa) but those were our taxa of interest and they were selected to analysis further those clades that were unsupported in the supertree. One sees the heavily weighted data sets in favour of the taxa of interest in other works: Alverson *et al*. 2012, 60 out of 82 taxa are the Thalassiosirales vs other taxa, and Ashworth *et al*. 2013 taxa of interest are the araphids, 39 out of 136 taxa and seven species from the genus of interest, *Cyclophora*. Li *et al*. (2015) had 62 araphids out of 157 diatoms, with a single outgroup with the family Plagiogrammaceae as their target of interest. However, as more and more genes and taxa are amassed, more balanced data sets will become available for more general analyses. Nevertheless, our work has indicated that multiple outgroups with multiple genes recover monophyletic classes with high bootstrap support, depending on the analysis, whereas single outgroups (Boldiomonads) (this study, Li *et al*. 2015, Theriot *et al*. 2015) usually do not. To date this the only study with multiple genes where this factor has been empirically tested.

New taxa

Subclass Urneidophycidae

Family Asterionellopaceae Medlin fam. nov.

Frustules heteropolar in both valve and girdle views. Two to numerous small discoid plastids. Cells attach to each other at broad or head pole of the valve faces to make flat, stellate to twisting, undulating chains. The genera in this family are marine planktonic, possibly epiphytic. Valves with alternating striae with simple poroid areolae covered with cribrate vela at the head pole to rotate vela at the foot pole, where studied. The valves markedly heteropolar, enlarged head pole narrows to an extension, which is very narrow in *Asterionellopsis* and broader in *Blekeleya* Round. Valves are without a raphe and belong to the subclass of araphids termed basal araphids, which are sister to another clade with other araphid and raphid diatoms. Valves have a narrow sternum, which can be almost indistinguishable, and spines may be present along the valve margin. The narrower end or foot pole of the valve has a single labiate process that lies either perpendicular or parallel to the striae. Both poles have apical pore fields that range from simple pores resembling valve poroids to elongated slits. Pore field at the head pole is separated from the valve face by a thickened, raised flange of silica that is also seen internally. Multiple narrow girdle bands, all perforated with multiple rows of pores, sometimes elongated.

Type genus *Asterionellopsis* Round

Subclass Fragilariophycidae

Family Staurosiraceae Medlin fam. nov.

Frustules rectangular to square in girdle view, often in ribbon chains, usually with two plastids. No epiphytic species are reported from this clade, they are either planktonic or bottom dwelling (epipelic/epipsammic/epilithic) life forms. Valves range from linear to elliptical in shape. Valves are without a raphe and the sternum can be narrow to very broad. Striae are alternate with poroid areolae that can either elongated into slits, uni or multi seriate. Molecularly, they are in a clade sister to all remaining core araphids with labiate processes. Labiate processes are highly reduced (*Plagiostriata*) or absent (*Nanofrustulum*, *Opephora*, *Staurosira*, *Staurosirella*, *Stauroforma*, *Pseudostaurosira, Pseudostaurosiropsis, Punctastriata*), which is the defining feature of this family. Apical pore fields range from being absent to one or two pores to small area of pores to a large area of pores, not bordered by a rim. Most genera have spines on or between the striae. Taxa are predominately freshwater, although *Nanofrustulum* is marine and can be endozooic. Girdle bands range from plain, at least the broad valvocopulae is, to perforated by a single row of pores and can have plaques.

Type genus *Staurosira* Ehrenberg

ACKNOWLEDGEMENTS. – The authors are grateful to S Sato for allowing us to use and publish data from his dissertation. It is regrettable that he did not wish to be part of this publication because of fears of criticism from the diatom community that the trees were only preliminary and because he could not defend that his tree was more correct than others. We have tried to acknowledge his contribution to the paper at every opportunity. We thank R Gersonde for invaluable comments and discussion about fossil records of diatoms, S Matsumoto, T Nagumo, B Petkus, Z Sun, K Valentin, W Wardle and T Watanabe, who provided us the living specimen, from which clonal cultures were established, S Ohtani for providing a *Pseudohimantidium pacificum* clone, A Witkowski for help sampling in Poland, S Frickenhaus (AWI, Germany) for establishing parallel processing for Bayesian analyses, F Hinz (AWI, Germany) for technical help with LM and SEM. This study was supported by DAAD for doctoral research fellowship to S Sato while at the A Wegener Institute for Polar and Marine Research, Bremerhaven, Germany.

REFERENCES

Altekar G, Dwarkadas S, Huelsenbeck J.P, Ronquist F 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20: 407-415.

Alverson AJ, Theriot EC 2005. Comments on recent progress toward reconstructing the diatom phylogeny. *J Nanosci Nanotech* 5: 57-62.

Alverson, AJ, Beszteri, B, Julius ML, Theriot EC 2011. The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus *Cyclotella*. *BMC Evol Biol* 11 125-133.

Amato A, Orsini L, D'Alelio D, Montresor M 2005. Life cycle size reduction patterns and ultrastructure of the pennate planktonic diatom *Pseudo-nitzschia delicatissima* (Bacillariophyta). *J Phycol* 41: 542-556.

Ambwani K, Sahni A, Kar R.K, Dutta D 2003. Oldest known non-marine diatoms (*Aulacoseira* from the uppermost Cretaceous Deccan Intertrappean Beds and Lameta formation of India. *Rev Micropal* 46: 67-71.

Amspoker MC 1989. *Nephroneis macintirei* gen. & sp. nov, a marine araphid diatom from California USA. *Diatom Res* 4: 171-177.

Amspoker MC 2008. Transfer of the marine diatom *Dimerogramma hyalinum* Hustedt to the new araphid genus *Hyaloneis*. *Diatom Res* 23: 11-18.

Andrews GW 1975. Taxonomy and stratigraphic occurrence of the marine diatom genus *Rhaphoneis*. *Nov Hedw Beih* 53: 193-227.

Andrews GW, Rivera P 1987. Morphology and evolutionary significance of *Adoneis pacifica* gen. & sp. nov. (Fragilariaceae Bacillariophyta) a marine araphid diatom from Chile. *Diatom Res* 2: 1-14.

Ashworth A, Ruck E, Lobban C, Romanovicz R, Theriot E 2012. A revision of the genus *Cyclophora* and description of *Astrosyne* gen. nov. (Bacillariophyta) two genera with the pyrenoids contained within pseudosepta. *Phycologia* 51: 684-699

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

Ben Ali A, Wuyts J, De Wachter R, Meyer A, Van de Peer Y 1999. Construction of a variability map for eukaryotic large subunit ribosomal RNA. *Nucleic Acids Res* 27: 2825-2831.

Berney C, Pawlowski JA 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Phil Trans R Soc Lond* Ser. B 273: 1867-1872.

Blazé KL 1984. Morphology and taxonomy of *Diplomenora* gen. nov. (Bacillariophyta). *Br J Phycol* 19 217-225.

Bradbury JP, Krebs W.N 1995. Fossil continental diatoms: paleolimnology evolution and biochronology. *In* Babcock LE, Ausich WI eds, Siliceous Microfossils, Short Courses in Paleontology no. 8) The Paleontological Society Knoxville Tennessee: 119-138.

Bruder K, Medlin LK 2008. Molecular assessment of phylogenetic relationships in selected species/genera in the naviculoid diatoms (Bacillariophyta). II. Selected genera and families. *Diatom Res* 23:331-347.

Chepurnov VA, Mann D.G, Sabbe K, Vyverman W 2004. Experimental studies on sexual reproduction in diatoms. *Int Rev Cytol* 237: 91-154.

Compère P 2001. *Ulnaria* (Kützing) Compère, a new genus name for *Fragilaria* subgen. *Alterasynedra* Lange-Bertalot with comments on the typification of *Synedra* Ehrenberg. *In* Jahn R, Kociolek JP, Witkowski A, Compère P eds, Lange-Bertalot-Festschrift: Studies on Diatoms, dedicated to Prof Dr Horst Lange-Bertalot on the occasion of his 65th Birthday. ARG Gantner Verlag. KG, Ruggell: 97-102.

Crawford RM, Gardner C 1997. The transfer of *Asterionellopsis kariana* to the new genus *Asteroplanus* (Bacillariophyceae) with reference to the fine structure. *Nov Hedw* 65 47-57.

Crawford R, Simonsen R, Hinz F, Gardner C 1993. The diatoms *Hustedtiella baltica* and *H. sinuata* sp. nov. and the systematic position of the genus. *Nov Hedw Beih* 106: 151-160.

Creevey CJ, McInerney JO 2005. Clann: investigating phylogenetic information through supertree analyses. *Bioinformatics* 21: 390-392.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Meth* 9: 772.

Daugbjerg N, Andersen RA 1997. A molecular phylogeny of the heterokont algae based on analyses of chloroplast encoded *rbc*L sequence data. *J Phycol* 33: 1031-1041.

Davidovich NA 2012. Pseudopodial activity of the gamete surface in araphid pennate diatoms. (Poster). Abstracts XXII Int Diatom Symp*:* 150.

Davidovich NA, Kaczmarska I, Karpov S.A, Davidovich OI, Michael L, MacGillivary ML, Mather L 2012. Mechanism of male gamete motility in araphid pennate diatoms from the genus. *Tabularia* (Bacillariophyta). *Protist* 163: 480-494.

DeSalle R 2005. What's in a character? *J Biomed Inform* 39: 6-17.

Desikachary TV, Sreelatha PM 1989. Oamaru diatoms. *Bib Diatom* 19: 1-330.

Doyle JJ, Doyle JL 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Dun WS, Rands WH, David TWE 1901. Note on the occurrence of diatoms, radiolaria and infusoria in the Rolling Downs Formation (Lower Cretaceous). Queensland. *Proc Linn Soc NSW* 26: 299-309.

Edwards AR 1991. The Oamaru diatomite. *NZ Geol Surv Paleontol Bull* 64: 1-260.

Elwood HJ, Olsen GJ, Sogin ML 1985. The small subunit ribosomal DNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonichia pustulata*. *Mol Biol Evol* 2: 399-410.

Fenner JM 1991. Taxonomy stratigraphy and paleoceanographic implications of Paleocene diatoms. *Proc Ocean Drill Prog Sci Res* 114:123-154.

Fenner JM 1994. Diatoms of the Fur Formation their taxonomy and biostratigraphic interpretation – results from the Harre borehole Denmark. *Aarh Geosci* 1: 99-163.

Fitch WM, Beintema JJ 1990. Correcting parsimonious trees for unseen nucleotide substitutions: the effect of dense branching as exemplified by ribonuclease. *Mol Biol Evol* 7: 438-443.

Fitch WM, Bruschi M 1987. The evolution of prokaryotic ferredoxins - with a general method correcting for unobserved substitutions in less branched lineages. *Mol Biol Evol* 4: 381-394.

Flower RJ, Jones VJ, Round FE 1996. The distribution and classification of problematic *Fragilaria virescens* v. *exigua* Grun./*Fragilaria exiguiformis* (Grun) Lange-Bertalot: a new species or a new genus? *Diatom Res* 11: 41-57.

Forti A, Schulz P 1932. Erste Mitteilungen über Diatomeen aus dem Hannoverschen Gault. *Beih Bot Zen* 50: 241-246.

Fritsch E 1935. Structure and reproduction of the algae. Vol. 1. Cambridge University Press, Cambridge.

Gersonde R, Harwood DM 1990. Lower Cretaceous diatoms from ODP Leg 113 site 693 (Weddell Sea. Part 1: vegetative cells. *Proc Ocean Drill Prog Sci Res* 113: 365-402.

Gibson RA 1979a. *Protoraphis atlantica* sp. nov., a new marine epizoic diatom. *Bacillaria* 2: 109-126.

Gibson RA 1979b. An ultrastructure study of *Pseudohimantidium pacificum* Hust. & Krasske (Bacillariophyceae: Protoraphidaceae) with special reference to the labiate processes. *Nov Hedw Beih* 64: 147-156.

Gillespie JJ, McKenna CH, Yoder MJ, Gutell RR, Johnston JS, Kathirithamby J, Cognato AI 2005. Assessing the odd secondary structural properties of nuclear small subunit ribosomal RNA sequences (SSU of the twisted-wing parasites (Insecta: Strepsiptera). *Insect Mol Biol* 14: 625-643.

Goloboff PA, Farris JS, Nixon KC 2008. TNT a free program for phylogenetic analysis. *Cladistics* DOI 10.1111/j.1096-0031.2008.00217.x

Gradstein FM, Ogg JG 2004. Geological Time Scale 2004 – Why how and where next! *Lethaia* 37: 175-181.

Graur D, Martin W 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet* 20: 80-86.

Greville RK 1863. Descriptions of new genera and species of diatoms from the South Pacific. Part 1. *Edinburgh New Phil J* New Ser 18: 34-43.

Guillou L, Chrétiennot-Dinet MJ, Medlin LK, Claustre H, Loiseaux-de Goër S, Vaulot D 1999. *Bolidomonas*: a new genus with two species belonging to a new algal class the Bolidophyceae class. nov. (Heterokonta). *J Phycol* 35: 368-381.

Haig DW, Barnbaum D 1978. Early Cretaceous microfossils from the type Wallumbilla Formation Surat Basin Queensland. *Alcheringa: Australas J Palaeontol* 2: 159-178.

Hajós M 1968. Die Diatomeen der Miozänen Ablagerungen des Matravorlandes. *Geol Hung Ser Palaeontol* 37: 1-40.

Hajós M 1986. Stratigraphy of Hungary's Miocene diatomaceous earth deposits. *Geol Hung Ser Palaeontol* 49: 1-339.

Hajós M, Strander H 1975. Late Cretaceous Archaeomonadaceae Diatomaceae and Silicoflagellatae from the South Pacific Ocean. Deep Sea Drilling Project Leg 29 site 275. *Init Rep. Deep Sea Drill Proj* 29: 913-1109.

Hall TA 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nuc Acids Symp Ser* 41: 95-98.

Hamilton PB, Poulin M, Yang JR, Klöser H 1997. A new diatom genus *Porannulus* (Bacillariophyta. associated with marine sponges around King George Island South Shetland Islands Antarctica. *Diatom Res* 12 229-242.

Hancock JM, Vogler AP 2000. How slippage-derived sequences are incorporated into rRNA variable-region secondary structure: implications for phylogeny reconstruction. *Mol Phylogenet Evol* 14: 366-374.

Harper ME 1977. A Lower Cretaceous (Aptian) diatom flora from Australia. *Nov Hedw Beih* 54: 411.

Harwood DM 1988. Upper Cretaceous and lower Paleocene diatom and silicoflagellate biostratigraphy of Seymour Island eastern Antarctic Peninsula. *Geol Soc Am Mem* 169: 55-129.

Harwood DM, Gersonde R 1990. Lower Cretaceous diatoms from ODP Leg 113 Site 693 (Weddell Sea). Part 2: resting spores chrysophycean cysts an endoskeletal dinoflagellate and notes on the origin of diatoms. *Proc Ocean Drill Prog Sci Res* 113: 403-425.

Harwood DM, Nikolaev VA 1995. Cretaceous diatoms: morphology taxonomy biostratigraphy. *In* Babcock LE, Ausich WI eds, *Siliceous microfossils* (Short Courses in Paleontology no. 8. The Paleontological Society Knoxville Tennessee: 81-106.

Harwood DM, Nikolaev VA, Winter DM 2007. Cretaceous records of diatom evolution radiation and expansion. *In* Starratt S ed, Pond Scum to Carbon Sink: Geological and Environmental Applications of the Diatoms (Short Courses in Paleontology no. 13). The Paleontological Society Knoxville Tennessee: 33-59.

Hasle GR 1974. The 'mucilage pore' of pennate diatoms. *Nov Hedw Beih* 45: 167-194.

Hasle GR 2001. The marine planktonic diatom family Thalassionemataceae: morphology taxonomy and distribution. *Diatom Res* 16: 1-82.

Hasle GR, Syvertsen EE 1996. Marine diatoms *In* Tomas C.R ed, Identifying marine diatoms and dinoflagellates. Academic Press Inc, New York: 5-385.

Hasle GR, Syvertsen EE, von Quillfeldt CH 1996. *Fossula arctica* gen. nov, spec. nov, a marine Arctic araphid diatoms. *Diatom Res* 11: 261-272.

Hernández-Becerril DU 1990. Observations on the morphology and distribution of the planktonic diatom *Neodelphineis pelagica*. *Br J Phycol* 25: 315-319

Hiromi J, Kadota S, Takano H 1985. Diatom infestation of marine copopods (review). *Bull Tok Reg Fish Res Lab* 117: 37-46.

Holmes RW, Croll DA 1984. Initial observations on the composition of dense diatom growths on the body feathers of three species of diving seabirds. *In* Mann DG ed, Proceedings of the Seventh International Diatom Symposium, Philadelphia, August 22-27, 1982. Koeltz Science Publishers, Koenigstein: 265-278.

Homann M 1991. Die Diatomeen der Fur-Formation (Alttertiär Limfjord/ Dänemark). *Geol Jb*. ser A 123: 1-285.

Honeywill C 1998. A study of British *Licmophora* species and a discussion of its morphological features. *Diatom Res* 13: 221-271.

Huelsenbeck JP, Ronquist F 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*17: 754-755.

Hustedt F 1959. Die Kieselalgenn Deutschlands Österreichs und der Schweiz. *In* Rabenhorst L ed,. Kryptogamen-Flora von Deutschland Osterreich und der Schweiz Band VII Teil 2. Otto Koeltz Science Publishers Koenigstein Germany.

Idei M, Nagumo T 2002. Auxospore structure of the marine diatom genus *Lampriscus* with triangular/quadrangular forms. *In* Poulin M ed, Abstracts book of 17th Int Diatom Symp. Ottawa, Canada: 166.

John J 1984. Observation on the ultrastructure of *Gephyria media* W. Arnott. *In* Richard M ed, Proc 8th Int Diat Symp. Koeltz Scientific Books Koenigstein: 155-162.

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: 43-71.

Kaczmarska I, Davidovich NA, Ehrman JM 2007. Sex cells and reproduction in the diatom *Nitzschia longissima* (Bacillariophyta): discovery of siliceous scales in gamete cell walls and novel elements of the perizonium. *Phycologia* 46: 726-737.

Kaczmarska I, Poulíčková A, Sato S, Edlund MB, Idei M, Watanabe T, Mann DG 2013. Proposals for a terminology for diatom sexual reproduction auxospores and resting stages. *Diatom Res* 28: 1-32. DOI:10.1080/0269249X.2013.791344

Kishino H, Hasegawa M 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequences and the branching order of Hominoidea. *J Mol Evol* 29: 170-179.

Kishino H, Thorne JL, Bruno WJ 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol Biol Evol* 18: 352-361.

Kobayasi H, Idei M, Mayama S, Nagumo T, Osada K 2006. H. Kobayasi's atlas of Japanese diatoms based on electron microscopy. Vol 1, Uchida Rokakuho Publishing Company Tokyo.

Kooistra WHCF, De Stefano M, Mann DG, Medlin LK 2003a. The phylogeny of the diatoms. *Prog Mol Subcell Biol* 33: 59-97.

Kooistra WHCF, De Stefano M, Mann DG, Salma N, Medlin LK 2003b. Phylogenetic position of *Torium* a pennate-like lineage within centric diatoms (Bacillariophyceae). *J Phycol* 39: 185-197.

Kooistra WHCF, Forlani G, Sterrenburg FAS, De Stefano M 2004. Molecular phylogeny and morphology of the marine diatom *Talaroneis posidoniae* gen. & sp. nov. (Bacillariophyta). advocate the return of the Plagiogrammaceae to the pennate diatoms. *Phycologia* 43: 58-67.

Kooistra WHCF, Sarno D, Balzano S, Gu H, Andersen RA, Zingone A 2007. Global Diversity and Biogeography of *Skeletonema* Species (Bacillariophyta). *Protist* 159: 177-193.

Kooistra WHCF Gersonde R, Medlin LK, Mann DG 2007b. The origin and evolution of the diatoms: their adaptation to a planktonic existence. *In* Falkowski PG, Knoll AH eds, Evolution of primary producers in the sea. Elsevier Amsterdam: 207-249.

Kooistra WHCF, Medlin LK 1996. Evolution of the diatoms (Bacillariophyta). IV. A reconstruction of their age from small subunit rRNA coding regions and the fossil record. *Mol Phylogenet Evol* 6: 391-407.

Li C, Ashworth MP, Witkowski A, Dąbek P, Medlin LK, Kooistra WHCF, Sato S, Zgłobicka I, Kurzydłowski KJ, Theriot EC, Sabir JSM, Khiyami MA, Mutwakil MHZ, Sabir MH, Alharbi NS, Hajarah NH, Qing S, Jansen RK 2015. New insights into Plagiogrammaceae (Bacillariophyta) based on multigene phylogenies and morphological characteristics with the description of a new genus and three new species. *PLoS ONE*. dx.doi.org/10.1371/journal.pone.0139300

Lobban C, Ashworth, MP 2014. *Hanicella moenia*, gen. & sp. nov., a ribbon-forming diatom (Bacillariophyta) with complex girdle bands, compared to *Microtabella interrupta* and *Rhabdonema* cf. *adriaticum*: implications for Striatellales, Rhabdonematales, and Grammatophoraceae, fam. nov. *J Phycol* 50: 860-884.

Lohman KE, Andrews GW 1968. Late Eocene nonmarine diatoms from the Beaver Divide area Fremont County Wyoming. Contributions to Paleontology. *US Geol Surv Profl Pap* 593-E 1-26.

Lupkina YG, Dolmatova LM 1975. The Paleogene lagoonal diatom flora of Kamchatka. *Paleontol ZJ* 1: 120-128.

Main SP 2003. *Diprora haenaensis* gen. & sp. nov, a filamentous pseudoaerial araphid diatom from Kaua'i (Hawaiian Islands. *Diatom Res* 18: 259-272.

Mann DG 1982. Structure life history and systematics of *Rhoicosphenia* (Bacillariophyta). II. Auxospore formation and perizonium structure of *Rh. curvata*. *J Phycol* 18: 264-274.

Mann DG, Evans KM 2007. Molecular genetics and the neglected art of diatomics. *In* Brodie J, Lewis J eds, Unravelling the Algae – The Past Present and Future of Algal Systematics. CRC Press Boca Raton Florida: 231-265.

Medlin, Linda K. 2011. The Permian Triassic Extinction forces the radiation of the modern phytoplankton. *Phycologia* 50: 684-693.

Medlin LK 2014. Evolution of the Diatoms: VIII. Re-Examination of the SSU-rRNA gene using multiple outgroups and a cladistic analysis of valve features. *J Biodivers Biopros Dev* 1: 129. doi:10.4172/2376-0214.1000129.

Medlin LK 2015. Opinion: Can coalescent models explain deep divergences in the diatoms and argue for the acceptance of paraphyletic taxa at all taxonomic hierarchies? *Nova Hedw* DOI:10.1127/nova_hedwigia/2015/0295.

Medlin LK 2016a. A timescale for diatom evolution based on four molecular markers and clarification of ghost lineages. *Vie Milieu* 65(4): 219-238.

Medlin L.K. 2016b. Evolution of the diatoms: major steps in their evolution and a review of the supporting molecular and morphological evidence. *Phycologia* 55: 79-103.

Medlin LK, Sato S 2009. The biological reality of the core and basal groups of araphid diatoms. *Diatom Res* 24: 503-508.

Medlin L, Elwood HJ, Stickel S, Sogin ML 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA coding regions. *Gene* 71: 491-499.

Medlin LK, Kaczmarska I 2004. Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision. *Phycologia* 43: 245-270.

Medlin LK, Kooistra WHCF, Gersonde R, Sims PA, Wellbrock U 1997. Is the origin of the diatoms related to the end-Permian mass extinction? *Nov Hedw* 65: 1-11.

Medlin LK, Kooistra WHCF, Schmid AM 2000. A review of the evolution of the diatoms – a total approach using molecules morphology and geology. *In* Witkowski A, Sieminska J eds, The Origin and Early Evolution of the Diatoms*:* Fossil Molecular and Biogeographical Approaches. Szafer Institute of Botany Polish Academy of Sciences Cracow: 13-35.

Medlin LK, Jung I, Bahulikar R, Mendgen K, Kroth P, Kooistra WHCF 2008a. Evolution of the diatoms. VI. Assessment of the new genera in the araphids using molecular data. *Nova Hedw Beih* 133: 81-100.

Medlin LK, Sato S, Kooistra WHCF, Mann DG 2008b. Molecular evidence confirms sister relationship of *Ardissonea Climacosphenia* and *Toxarium* within the bipolar centric diatoms and cladistic analyses confirm that extremely elongated shape has arisen twice in the diatoms (Bacillariophyta Mediophyceae). *J Phycol* 44: 1340-1348.

Milinkovitch MC, Lyons-Weiler J 1998. Finding optimal ingroup topologies and convexities when the choice of outgroups is not obvious. *Mol Phygenet Evol* 9: 348-357.

Morales EA 2001. Morphological studies in selected fragilarioid diatoms (Bacillariophyceae. from Connecticut waters (USA). *Proc Acad Nat Sci Phil* 151: 105-120.

Morales EA 2002. Studies in selected fragilarioid diatoms of potential indicator value from Florida (USA) with notes on the genus *Opephora* Petit (Bacillariophyceae. *Limnologica* 32: 102-113.

Nagumo T 2003. Taxonomic studies of the subgenus *Amphora* Cleve of the genus *Amphora* (Bacillariophyceae) in Japan. *Bib Diatom* 49: 1-265.

Navarro NJ 1982. A survey of the marine diatoms of Puerto Rico IV. Suborder Araphidineae: Families Diatomaceae and Protoraphidaceae. *Bot Mar* 25: 247-263.

Navarro JN, Williams DM 1991. Description of *Hyalosira tropicalis* sp. nov. (Bacillariophyta) with notes on the status of *Hyalosira* Kützing and *Microtabella* Round. *Diatom Res* 6: 327-336.

Nikolaev VA, Harwood DM, Samsonov NI 2001. Lower Cretaceous Diatoms. Komarov Botanical Institute, Saint Petersburg, NAUKA.

Nylander JAA 2004. MrModeltest. Program distributed by the author. Evol Biol Centre Uppsala Univ.

Patrick R, Reimer CW 1966. The diatoms of the United States. Vol 1. Monographs *Proc Acad Nat Sci Phil* 13: 1-688.

Prasad AKSK, Fryxell GA 1991. Habit, frustule morphology and distribution of the Antarctic marine benthic diatom *Entopyla australis* var. *gigantea* (Grev) Fricke (Entopylaceae). *Br J Phycol* 26: 101-122.

Prasad AKSK, Livingston RJ 1993. Frustule morphology of the plantonic pennate diatom *Fragilaria gessneri* Hustedt (Bacillariophyceae) from Florida coast of the Gulf of Mexico with a description of *Desikaneis* gen. nov. *Phycologia* 32: 434-443.

Prasad AKSK, Nienow JA, Riddle KA 2001. Fine structure taxonomy and systematics of *Reimerothrix* (Fragilariaceae Bacillariophyta) a new genus of synedroid diatoms from Florida Bay USA. *Phycologia* 40 35-46.

Reháková Z 1980. Süsswasserdiatomeenflora des oberen Miozäns in der Tschechlowakei. *J Geol Sci* 23: 83-184.

Reinhold T 1937. Fossil Diatoms of the Neogene of Java and their zonal distribution. *Verh Geol Mijn Genootsch Neder Kolon Geol Ser* 12 43-132.

Ronquist F, Huelsenbeck JP 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.

Ronquist FF, Teslenko MM, van der Mark PP, Ayres, DLD, Darling AA, Hahna SS, Larget B, Liu L, Suchard MA, Huelsenbeck JP 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61: 539-542.

Rothpletz A 1896. Über die Flysch-Fucoiden und einige andere fossile Algen sowie uberliasische Diatomeen fuhrende Hornschwamme. *Z Dtsch Geol Ges Wiss* 48: 854-915.

Rothpletz A 1900. Über einen neuen jurassischen Hornschwamm und die darin eingeschlossenen Diatomeen. *Z Dtsch Geol Ges Wiss* 52: 154-160.

Round FE 1993. The identity of *Synedra closterioides* Grun. and its transfer to a new genus *Psammosynedra*. *Diatom Res* 8: 209-213.

Round FE, Alexander CG 2002. Licmosoma - a new diatom genus growing on barnacle cirri. *Diatom Res* 17: 319-326.

Round FE, Mann DG 1980. *Psammodiscus* nov. gen. based on *Coscinodiscus nitidus*. *An Bot* 46: 367-373

Round FE, Maidana NI 2001. Two problematic freshwater araphid taxa, re-classified in new genera. *Diatom* 17 21-28.

Round FE, Hallsteinsen H, Paasche E 1999. On a previously controversial "fragilarioid" diatom now placed in a new genus *Nanofrustulum*. *Diatom Res* 14 343-356.

Round FE, Crawford RM, Mann DG 1990. The Diatoms: Biology and Morphology of the Genera. Cambridge University Press Cambridge.

Rutschmann F 2004. Bayesian molecular dating using PAML/multidivtime. A step-by-step manual. University of Zurich Switzerland. http://www.plant.ch.

Sabbe K, Vyverman W 1995. Taxonomy morphology and ecology of some widespread representatives of the diatom genus *Opephora*. *Eur J Phycol* 30: 235-249

Sar EA, Sunesen I, Fernández PV 2007. Marine diatoms from Buenos Aires coastal waters (Argentina). II. *Thalassionemataceae* and *Rhaphoneidaceae*. *Rev Chil Hist Nat* 80: 63-79.

Sarno D, Kooistra WHCF, Hargraves PE, Zingone A 2007. Diversity in the genus *Skeletonema* (Bacillariophyceae. III. Phylogenetic position and morphology of *Skeletonema costatum* and *Skeletonema grevillei* with the description of *Skeletonema ardens* sp. nov. *J Phycol* 43: 156-170.

Sarno D, Kooistra WCHF, Medlin LK, Percopo I, Zingone A 2005. Diversity in the genus *Skeletonema* (Bacillariophyceae. II. An assessment of the taxonomy *S. costatum*-like species with the description of four new species. *J Phycol* 41: 151-176.

Sato S 2008. Phylogeny of araphid diatoms inferred from morphological and molecular data. PhD Dissertation, Univ Bremen. http: //elibsuubunibremende/diss/docs/00011057.pdf.

Sato S, Medlin LK 2006. Note: Motility of non-araphid diatoms. *Diatom Res* 21: 473-477.

Sato S, Nagumo T, Tanaka J 2004a. Morphology and taxonomy of marine attached diatoms in genus *Grammatophora* Ehrenberg (Bacillariophyceae) in Japan. *Jpn J Phycol* 52 (suppl): 183-187.

Sato S, Nagumo T, Tanaka J 2004b. Auxospore formation and the morphology of the initial cell of the marine araphid diatom *Gephyria media* (Bacillariophyceae). *J Phycol* 40: 684-691.

Sato S, Kooistra WHCF, Watanabe T, Matsumoto S, Medlin LK 2008a. Phylogenetics and morphometrics reveal pseudocryptic diversity in a new araphid diatom genus *Psammoneis* (Plagiogrammaceae Bacillariophyta). gen. nov. *Phycologia* 47: 510-528.

Sato S, Kuriyama K, Tadano T, Medlin LK 2008b. Auxospore fine structure in a marine araphid diatom *Tabularia parva* (Bacillariophyta). *Diatom Res* 28: 23:423-433.

Sato S, Matsumoto S, Medlin LK 2008c. *Pseudostriatella pacifica* gen. & sp. nov. (Bacillariophyta): a new araphid diatom genus and its fine-structure auxospore and phylogeny. *Phycologia* 47: 371-391.

Sato S, Mann DG, Nagumo T, Tanaka J, Medlin LK 2008d. Life cycle of *Grammatophora marina* (Bacillariophyta) with special reference to its auxospore fine structure. *Phycologia* 47: 12-27.

Sato S, Matsumoto S, Medlin LK 2008e. Fine structure and 18S rDNA phylogeny of a marine araphid pennate diatom *Plagiostriata goreensis* gen & sp nov (Bacillariophyta). *Phycol Res* 57: 25-35.

Sato S, Watanabe T, Crawford RM, Kooistra WHCF, Medlin LK 2008f. Morphology of four plagiogrammacean diatoms: *Dimeregramma minor* var. *nana Neofragilaria nicobarica Plagiogramma atomus* and *Psammogramma vigoensis* gen. & sp. nov, and their phylogenetic relationship inferred from partial 28S rDNA. *Phycol Res* 56: 56: 255-268.

Sato S, Beakes G, Idei M, Nagumo T, Mann DG 2011. Novel sex cells and evidence for sex pheromones in diatoms. *PLoS ONE* 6: e26923.

Scholin MA, Villac MC, Buck KR, Powers JM, Fryxell GA, Chavez FP 1994. Ribosomal DNA sequences discriminate among toxic and non-toxic *Pseudonitzschia* species. *Nat Toxins* 2: 152-165.

Schrader HJ, Fenner J 1976. Norwegian Sea Cenozoic diatom biostratigraphy and taxonomy. Part 1. *Rep Deep Sea Drill Proj* 38: 921-1099.

Schrader HJ, Gersonde R 1978. The Late Messian Mediterranean brackish to freshwater environment diatom floral evidence. *Rep Deep Sea Drill Proj* 42: 761-775.

Shull VL, Vogler AP, Baker MD, Maddison DR, Hammond PM 2001. Sequence alignment of 18S ribosomal RNA and the basal relationships of adephagan beetles: evidence for monophyly of aquatic families and the placement of Trachypachidae. *Syst Biol* 50: 945-969.

Simonsen R 1960. Neue Diatomeen aus der Ostsee. II. *Kiel Meeres* 16: 126-160.

Simonsen R 1970. Protoraphidaceae eine neue Familie der Diatomeen. *Nov Hedw Beih* 31: 383-394.

Simonsen R 1979. The diatom system: ideas on phylogeny. *Bac* 2: 9-71.

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

Singh RS, Stoermer EF, Kar R 2006. Earliest freshwater diatoms from the deccan intertrappen (Maastrichtian) sediments of India. *Micropal* 52: 545-551.

Shimodaira H, Hasegawa M 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16: 1114-1116.

Simonsen R 1979.The diatom system: ideas on phylogeny. *Bac* 2: 9-71.

Sinninghe-Damsté JS, Muyzer G, Abbas B, Rampen SW, Masse G, Allard WG, Belt ST, Robert JM, Rowland SJ, Moldowan JM, Barbanti SM, Fago FJ, Denisevich P, Dahl J, Trindade LAF, Schouten S 2004. The rise of the rhizosolenid diatoms. *Science* 304: 584-587.

Snoeijs P 1992. Studies in the *Tabularia fasciculata* complex. *Diatom Res* 7: 313-344.

Sorhannus U 2004. Diatom phylogenetics inferred based on direct optimization of nuclear-encoded SSU rRNA sequences. *Clad* 20: 487-497.

Sorhannus U 2007. A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. *Mar Micropal* 65: 1-12.

Soltis PS, Soltis DE, Savolainen V, Crane PR, Barraclough TG 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proc Natl Acad Sci USA* 99 4430-4435.

Stamatakis A, Ludwig T, Meier H 2005. RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21: 456-463.

Stosch HA von 1982. On auxospore envelopes in diatoms. *Bac* 5: 127-156.

Sullivan MJ 1979. Taxonomic notes on epiphytic diatoms of Mississippi Sound USA. *Nov Hedw Beih* 64: 241-249.

Sullivan MJ 1988. A morphological study of the marine diatom *Glyphodesmis eximia*: the type of the genus. *In* Round FE ed, Proceedings of the 9th Int Diatom Symp. Bristol Biopress & Koenigstein Koeltz: 361-370.

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

Takano H 1982. New and rare diatoms from Japanese marine waters-VIII. *Neodelphineis pelagica* gen. & sp. nov. *Bull Tokai Reg Fish Res Lab* 106: 45-53

Takano H 1983. New and rare diatoms from Japanese marine waters. XI. Three new species epizoic on copepods. *Bull Tokai Reg Fish Res Lab* 111: 23-35.

Tapia P.M 1996. Campanian diatom biostratigraphy and paleoecology of Arctic Canada. MSc Thesis. University of Nebraska.

Tapia PM, Harwood M 2002. Upper Cretaceous diatom biostratigraphy of the Artic Archipelago and northern continental margin Canada. *Micropaleontology* 48: 303-342.

Tarrio R, Rodriguez-Trelles F, Ayala FJ 2000. Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: the *Drosophila saltans* and *willistoni*groups a case study. *Mol Phygenet Evol* 16: 344-349

Tappan H 1980. The paleobiology of plant protists. W.H. Freeman and Company, San Francisco.

Terasaka T, Nagumo T, Tanaka J 2005. Distribution and morphology of four common species of the marine pennate diatom genus *Licmophora* C. Agardh in Japan. *Diatom* 21: 91-106 (In Japanese with English abstract).

Theriot, EC, Ashworth, MP, Nakov T, Ruck E, Jansen RJ 2015. Dissecting signal and noise in diatom chloroplast protein encoding genes with phylogenetic information profiling. *Mol Phylogenet Evol* 89: 28-36.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG 1997 The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24: 4876-4882.

Thorne JL, Kishino H 2002. Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol* 51(5): 689-702.

Thorne JL, Kishino H, Painter IS 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15: 1647-1657.

Van de Peer Y, Neefs JM De Rijk P, De Wachter R 1993. Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. *J Mol Evol* 37: 221-232.

Van de Peer Y, Caers A, De Rijk P, De Wachter R 1998. Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res* 26: 179-182.

VanLandingham SL 1967. Paleoecology and microfloristics of Miocene diatomites from the Otis Basin-Juntura Region of Harney and Malheur Counties Oregon. *Nov Hedw Beih* 26: 1-77.

VanLandingham SL 1985. Potential Neogene diagnostic diatoms from the western Snake River basin Idaho and Oregon. *Micropaleontology* 31: 167-174.

Webster AJ, Payne RJH, Pagel M 2003. Molecular phylogenies link rates of evolution and speciation. *Science* 301**:** 478.

Watanabe T, Suzuki H, Nagumo T, Tanaka J 2007. Morphological investigation of *Perissonoë crucifera* Kitton Desikachary *et al*. collected from Japan. *J Jpn Bot* 82: 296-304.

Williams DM 1985. Morphology taxonomy and inter-relationships of the ribbed araphid diatoms from the genera *Diatoma* and *Meridion* (Diatomaceae: Bacillariophyta). *Bib Diatom* 8: 1-228.

Williams DM, Round FE 1986. Revision of the genus *Synedra*. *Diatom Res* 1: 319-339.

Williams DM, Round FE 1987. Revision of the genus *Fragilaria*. *Diatom Res* 2: 267-288.

Williams DM 1987. Observations on the genus *Tetracyclus* Ralfs (Bacillariophyta). I. Valve and girdle structure of the extant species. *Br J Phycol* 22: 383-399.

Williams DM 1989. Observations on the genus *Tetracyclus* Ralfs (Bacillariophyta). II Morphology and taxonomy of some species from the genus *Stylobiblium*. *Br J Phycol* 24: 317-327.

Williams DM 1988. An illustrated catalogue of the type specimens in the Greville diatom herbarium. *Bull Br Mus (Nat Hist Bot* Ser) 18: 1-148.

Williams DM 1990. *Distrionella* D.M. Williams nov. gen, a new araphid diatom (Bacillariophyta) genus closely related to *Diatoma* Bory. *Arch Prot* 138: 171-177.

Witkowski A, Lange-Bertalot H, Metzeltin D 2000. Diatom Flora of Marine Coasts I. Iconographia Diatomologica Vol. 7. A.R.G. Gantner Verlag K.G., Ruggell Liechtenstein.

Wolfe AP, Edlund MB 2005. Taxonomy phylogeny and paleoecology of *Eoseira wilsonii* gen. & sp. nov a Middle Eocene diatom (Bacillariophydceae: Aulcoseiraceae). from lake sediments at Horse Fly British Columbia Canada. *Can J Earth Sci* 42: 243-257.

Xia X, Xie Z 2001. DAMBE: Data analysis in Molecular Biology and Evolution. *J Hered* 92: 371-373.

Yang Z 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comp Appl Biosci* 13: 555-556.

Yang Z, Yoder AD 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points with application to a radiation of cute-looking mouse Lemur species. *Syst Biol* 52: 705-716.

Yoon HS, Hackett JD, Bhattacharya D 2002. A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc Natl Acad Sci USA* 99: 11724-11729.

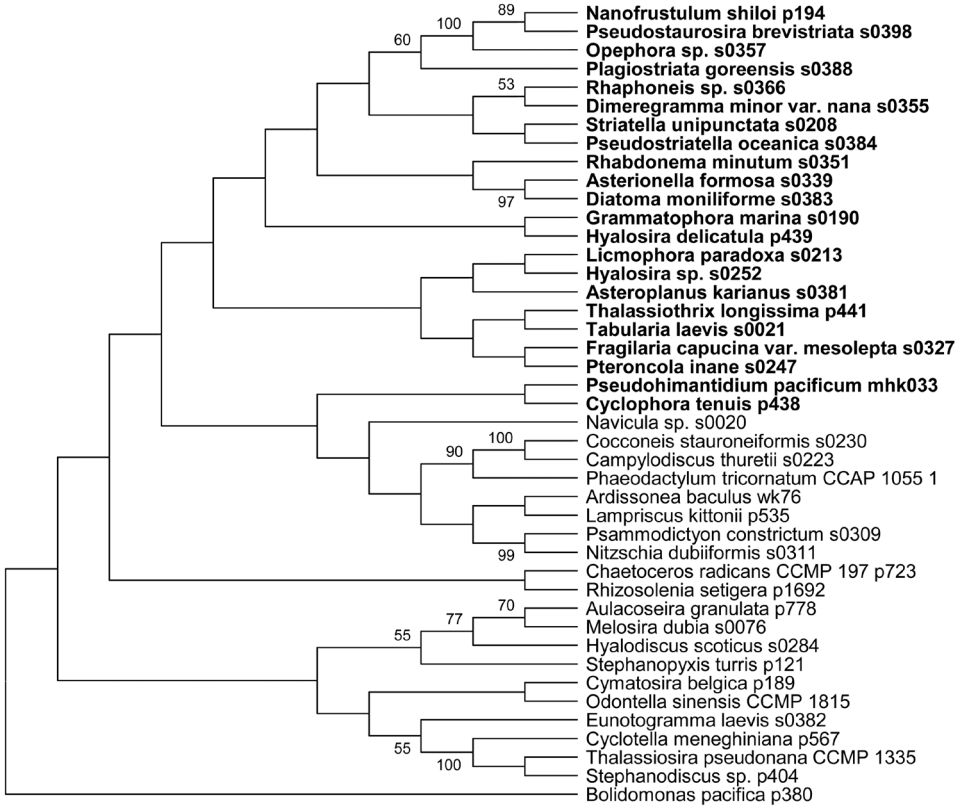
Zingone A, Percopo I, Sims PA, Sarno D 2005. Diversity in the genus *Skeletonema* (Bacillariophyceae). I. A re-examination of the type material of *Skeletonema costatum* with the description of *S. grevillei* sp. nov. *J Phycol* 41: 140-150.

Received on June 26, 2015
Accepted on April 27, 2016
Associate editor: E Magnanou

SUPPLEMENTAL TREES

RAxML trees constructed by each single gene nuclear (*SSU+LSU*. and plastid (*rbc*L+*psb*A. genes. Taxon labels of araphid diatoms are indicated by bold. The bootstrap value exceeding 50 % is shown. **S1**. *SSU*; **S2**. *LSU*; **S3**. *rbc*L; **S4**. *psb*A; **S5**. rDNA (*SSU+LSU*.; **S6**. cpDNA (*rbc*L+*psb*A).

*rbc*L



*psb*A