VIE ET MILIEU - LIFE AND ENVIRONMENT, 2015, 65 (4): 219-238

A TIMESCALE FOR DIATOM EVOLUTION BASED ON FOUR MOLECULAR MARKERS: REASSESSMENT OF GHOST LINEAGES AND MAJOR STEPS DEFINING DIATOM EVOLUTION

LINDA K. MEDLIN

Marine Biological Association of the UK, The Citadel, Plymouth PL1 2PB UK
lkm@mba.ac.uk

DIATOMS
EVOLUTION
MOLECULAR CLOCK
PHYLOGENY
FOSSIL RECORDS
MULTI-GENE ANALYSIS

ABSTRACT. – Phylogenies and molecular clocks of the diatoms have largely been inferred from SSU rDNA sequences. A new phylogeny of diatoms was estimated using four gene markers SSU and LSU rDNA *rbc*L and *psb*A (total 4352 bp) with 42 diatom species. The four gene trees analysed with a maximum likelihood (ML) and Baysian (BI) analysis recovered a monophyletic origin of the new diatom classes with high bootstrap support, which has been controversial with single gene markers using single outgroups and alignments that do not take secondary structure of the SSU gene into account. The divergence time of the classes were calculated from a ML tree in the MultliDiv Time program using a Bayesian estimation allowing for simultaneous constraints from the fossil record and varying rates of molecular evolution of different branches in the phylogenetic tree. These divergence times are generally in agreement with those proposed by other clocks using single genes with the exception that the pennates appear much earlier and suggest a longer Cretaceous fossil record that has yet to be sampled. Ghost lineages (i.e. the discrepancy between first appearance (FA) and molecular clock age of origin from an extant taxon) were revealed in the pennate lineage, whereas those ghost lineages in the centric lineages previously reported by others are reviewed and referred to earlier literature.

INTRODUCTION

Molecular clocks are useful tools for unravelling the evolution of protistan taxa and have been used to reconstruct biogeographic histories and to estimate divergence times of many protists ranging from their origins to the divergence of cryptic species (see review in Medlin 2008). Microalgae, such as diatoms, dinoflagellates and coccolithophores, have mineralized/tough organic walls that may be preserved under appropriate conditions. These microalgal fossils have better preserved fossil records than their metaphyton and metazoan counterparts and molecular clocks made using calibrations from microalgae are better calibrated (Berney & Pawlowski 2006). Sorhannus (2007) has constructed a molecular clock for the origin of the diatoms, which is an improvement over that constructed by Kooistra & Medlin (1996) because the true sister group of the diatoms the Bolidophyceae (Guillou *et al*. 1999) is now known. This enables the root of the diatom lineage to be defined precisely. Taking too distant an outgroup only provides the origin of the entire lineage's last common ancestor of the ingroup and the outgroup rather than the true origin of the ingroup with its closest sister. For example, Philippe *et al*. (1994) used a relative rate test comparison among diatoms and other eukaryotic groups and concluded that the diatoms had a 300 m.y. gap in their fossil record. They used ciliates as the sister group and thus the branch length leading from the diatoms to ciliates corresponds to the entire branch length leading to the origin of the Heterokonta, which is the kingdom to which the diatoms belong plus the branch length to the ciliates. Thus, the relative rate of evolution in the diatoms as compared to the ciliates took into account far more base substitutions than are actually present in the branch length separating diatoms from bolidomonads. Their conclusion that there was a 300 m.y. gap in the diatom fossil record is compromised because they used a too distant outgroup for their relative rate test. Sorhannus (1997) calculated a molecular clock for the origin of the diatoms and used the chrysophytes as the nearest sister group to the diatoms, which are likely still too far away from the ingroup to give a reliable date for their origin. This date of origin was 330-400 Ma based on the SSU and LSU genes analysed separately. In his 2007 clock based on a single gene, the SSU rRNA gene, Sorhannus used a relaxed molecular clock (PATHd8), which was calibrated from single dating points sequentially and used the bolidomonads as the correct sister group of the diatoms. He found that each fossil date corroborated the overall results. In the clock for inferring the age and divergences of the classes within the pigmented heterokonts (Ochrophyta) (Brown & Sorhannus 2010), the Bolidophyceae were included as the sister group to the diatoms.

A new molecular clock for the diatoms based on four molecular markers is presented here to calibrate the divergence/diversification of the three diatom classes using the bolidomonads plus multiple distant heterokont groups as outgroups (Medlin & Kaczmarska 2004). The tree upon which this clock is based was taken from a study on the evolution of araphid diatoms by Sato (2008), which pro-

duced a tree with three monophyletic classes, albeit with low bootstrap support when the third codon position was RY coded (Supplementary Fig. 1). Reanalysis of his data set using different models for amino acid substitution with both ML and BI analyses recovered the classes as being monophyletic with high bootstrap support (Supplementary Figs 2 & 3). Other trees in Sato's dissertation, using a single bolidomonad as outgroup, did not recover monophyletic classes but those trees are not used here because they did not recover monophyletic classes. A similar study emphasizing basal araphid diatoms by Li *et al*. (2015) has also produced three monophyletic classes with 159 taxa and three genes (Supplementary Fig. 4). This tree has very high bootstrap support for the three classes adding support to the tree and the taxon selection from Sato (2008) that was selected to make this newest molecular clock. It would appear that many araphid taxa are critical to recovering the classes monophyletic. "Ghost lineages" where the molecular date precedes the known stratigraphic first appearance (FA) in both the present tree and that of Sorhannus (2007) are discussed.

Because recovery of the three monophyletic diatoms has now been achieved independently from those analyses by Medlin with separate multi-gene datasets, I hypothesize that first appearance dates of diatom taxa can be used to infer the divergence of the major evolutionary lineages in the diatoms and the appearance of morphological and reproductive features can be associated with these divergences. Two dates were used to constrain the root of the diatoms and many internal nodes were used to calibrate the clock.

MATERIAL AND METHODS

Taxon Sampling: To cover a wide range of diatoms taxon sampling was carried out using a previously published phylogenetic tree Medlin & Kaczmarska 2004). Taxa used in this study are shown in Table I and formed part of a larger study on the evolution of the araphid diatoms (Sato 2008, Sato *et al*. unpubl.). Multiple outgroups of heterokont algae including *Bumilleriopsis filiformis* (Xanthophyceae), *Dictyota dichotoma* (Phaeophyceae) and *Heterosigma akashiwo* (Rhaphidophyceae) as well as *Bolidomonas* sp. from the Bolidophyceae the closest relatives of the diatoms (Guillou *et al*. 1999) were used. Among the diatom taxa dataset five taxa are Coscinodiscophyceae, nine are Mediophyceae and 28 are Bacillariophyceae, which include six raphid and 22 araphid diatoms (Table I). The araphid gene sequences are part of the study done by Sato (2008), whereas other groups are taken from Genbank (http://www.ncbi.nlm.nih.gov/nucleotide/). Because Sato's dissertation was focused on evolution within the araphid diatoms the number of araphid taxa exceeded the other taxa (slightly over 50 %) as did the study by Li *et al*. (2015) whose goal was also to study one araphid family. However, in support of this limited taxon sampling, a much wider taxon sampling of 157 diatoms of which 61 were araphids (nearly 50 %) using three genes also recovered monophyletic classes with high bootstrap support (Supplementary Fig. 4).

Database assembly: Partial fragments of SSU (~1657 bp) LSU (~659 bp) rbcL (~1461 bp) and psbA (~933 bp) were amplified by PCR (see Supplementary Table I for primers). Details for the PCR reactions for each gene can be found in Sato (2008) who produced this data set for the study of the evolution of araphid diatoms.

This study represents a re-analysis of one tree from that data set using a molecular clock. The rDNA sequences were first aligned using ClustalX (Thompson *et al*. 1997) and this alignment was refined by referring to the secondary structure model of the rRNA in the rRNA structure database of Van de Peer *et al*. (1998) for SSU and Sato (2008) for LSU. There is extreme length variation in some rRNAs (Gillespie *et al*. 2005) but not in most diatoms and replication slippage can lead to convergence of similar primary and secondary structures (Hancock & Vogler 2000, Shull *et al*. 2001). Homology assessment in such regions can be difficult or impossible so highly variable regions (mostly peripheral regions of the rRNA secondary structure see supplementary file 1) were removed from the alignment using BioEdit 7.0.2 (Hall 1999) using the variability map of *Saccharomyces cerevisiae* (Van de Peer *et al*. 1993 for SSU and Ben Ali *et al*. 1999 for LSU). Sequences for the protein coding genes *psb*A and *rbc*L lacked indels and were aligned manually. The final dataset/alignment comprising 4352 bp can be obtained from Dr. Shinya Sato, ssato@fpu.ac.jp.

Phylogenetic Analyses: RAxML VIHPC v2.2.3 (Stamatakis *et al*. 2005) was used for ML analyses with GTR-MIX model. Gamma correction value of each gene or partition of the dataset was obtained automatically by the program. The analyses were performed 1000 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).

Divergence Time Estimation: With the dataset divergence times were estimated using Bayesian methods (Thorne *et al*. 1998 2000, Kishino *et al*. 2001, Thorne & Kishino 2002) in the program Multi-divtime 9/25/03 (Thorne & Kishino 2002) and PAML 3.15 (Yang 1997). Original sequences for the first and second codon were used in this process, whereas the third codon was RY recoded as a purine or pyrimidine base because preliminary analyses using a RY recoded dataset (Sato 2008) showed that the likelihoods obtained from the base-ml and est-branches analyses were very different suggesting that one or both programs failed to optimise the likelihood. Often if gene substitution is saturated so that all phylogenetic signal is lost then recoding the sequence data to indicate that the base is a pyrimidine (Y) or a purine (R) is the only way that a reliable phylogenetic signal can be obtained. Multi-divtime is the only molecular clock program that can cope with multi-locus data sets and multiple dating points. It is a relaxed molecular clock, termed rate smoothing that allows for rate variation across lineages and

Table I.– List of species used in this analysis. Accession numbers beginning with AB are new to this study.

Taxa	**Strain (Voucher)**	***SSU***	***LSU***	***rbc*L**	***psb*A**
Coscinodiscophyceae [radial centrics]					
Aulacoseira granulata (Ehrenberg) Simmonsen	p778	AB430586	AB430619	AB430659	AB430699
Hyalodiscus scoticus (Kützing) Grunow	s0284	AB430587	AB430620	AB430660	AB430700
Melosira dubia Kützing	s0076	AB430588	AB430621	AB430661	AB430701
Rhizosolenia setigera Brightwell	p1692	AB430589	AB430622	AB430662	AB430702
Stephanopyxis turris (Greville et Arnott) Ralfs	p121	AB430590	AB430623	AB430663	AB430703
Mediophyceae [polar 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]	
Stephanodiscus sp.	p404	AB430594	AB430631	AB430670	AB430710
Thalassiosira pseudonana Hasle et Heimdal	CCMP1335	EF208793[a]		EF067921[a]	
Bacillariophyceae [pennates]					
Araphid diatoms (basal araphids)					
Asteroplanus karianus (Grunow) Gardner et Crawford	s0381	AB430596	AB430633	AB430672	AB430712
Dimeregramma minor var. nana (Gregory) Ralfs	s0355	AB430598	AB425083	AB430675	AB430715
Rhaphoneis sp.	s0366	AB430602	AB430641	AB430681	AB430721
Pseudostriatella pacifica S. Sato et Medlin	s0384 (Zu6/38)	AB379680	AB430646	AB430686	AB430726
Striatella unipunctata Agardh	s0208	AB430609	AB430649	AB430689	AB430729
Araphid diatoms (core araphids*)* *Asterionella formosa* Hassall	s0339	AB430595	AB430632	AB430671	AB430711
Cyclophora tenuis Castracane	p438	AJ535142	AB430634	AB430673	AB430713
Diatoma moniliforme Kützing	s0383	AB430597	AB430635	AB430674	AB430714
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
Hyalosira tropicalis Navarro	s0252	AB430612	AB430652	AB430692	AB430732
Licmophora paradoxa (Lyngbye) Agardh	s0213	AB430601	AB430639	AB430679	AB430719
Nanofrustulum shiloi (Lee, Reimer & McEnery) Round, Hallsteinsen & Paasche	p194	AM746971	AB430640	AB430680	AB430720
Opephora sp.	s0357	AB430604	AB430643	AB430683	AB430723
Plagiostriata goreensis S. Sato et Medlin	s0388	AB430605	AB430644	AB430684	AB430724
Pseudohimantidium pacificum Hustedt et Krasske	mhk033	AB430606	AB430645	AB430685	AB430725
Pteroncola inane (Giffen) Round	s0247	AB430607	AB430647	AB430687	AB430727
Pseudostaurosira brevistriata (Grun. in VH) Will. et Round	s0398	AB430608	AB430648	AB430688	AB430728
Rhabdonema minutum Kützing	s0351	AB430603	AB430642	AB430682	AB430722
Tabularia laevis Kützing	s0021	AB430610	AB430650	AB430690	AB430730
Thalassiothrix longissima Cleve et Grunow	p441	AB430611	AB430651	AB430691	AB430731
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	AB430736
Phaeodactylum tricornatum Bohlin	CCAP1055/1	EF553458[a]		EF067920[a]	

Table I. – Continued.

Taxa	Strain (Voucher)	*SSU*	*LSU*	*rbc*L	*psb*A
Psammodictyon constrictum (Gregory) Mann	s0309	AB430617	AB430657	AB430697	AB430737
Outgroup					
Bolidomonas pacifica Guillou et Chrétiennot Dinet [Bolidophyceae]	p380	AB430618	AB430658	AB430698	AB430738
Bumilleriopsis filiformis Vischer [Xanthophyceae]	NA	AF083398	NA	X79223	
Dictyota dichotoma (Hudson) Lamouroux [Phaeophyceae]	NA	AF350227	AF331152	AY748321	
Heterosigma akashiwo (Hara) Hara [Raphidophyceae]	NA	DQ470662		AY119759	

[a] Whole genome

across genes. Data sets without the RY coding recovered the same topology but with higher bootstrap support (Supplementary Figs 2 & 3).

The F84 model incorporating within-site rate variation modelled by a gamma distribution (Yang 1997) of sequence evolution (F84+G model) was used. This is the most complex model implemented in this program. This model is less parameterised than the best-fit models selected by MrModeltest 2.2 (in Sato 2008 the GTR+I+G was the best model selected for this data set for routine phylogenetic analysis). 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 and this parameter is accommodated in the F84+G model. All the parameters within the model as well as the branch lengths were estimated separately for each gene. Markov chain Monte Carlo (MCMC) approximations were obtained with a burn-in period of 100000 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. Two replicate MC-MC runs with different initial starting points for each analysis were performed and allowed to run until the standard deviation values of log likelihoods of the two runs dipped below 0.01. Application of the multi-divtime program requires a value for the mean of the prior distribution for the time separating the ingroup root from the present (rttm). A maximum (250 Ma) and minimum estimate (190 Ma) setting standard deviation of rttm (rttmsd) = 0 (see below about the calibration strategy) was used. Other parameters for running multi-divtime were set following Rutschmann (2004): 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.

Calibration Points: The FA of taxa in fossil record can be used as a minimum age constraint for the node in the tree where the two taxa diverge; however, in reality, the true first appearance can never be known because in most cases molecular divergence precedes morphological divergence. Nevertheless, I have consistently used the FA as the minimum age for that node in the tree. If both taxa have fossil record the older age was taken as the minimum constraint assuming that molecular divergence always precedes morphological divergence. Calibrated ages are shown in Table II. All are based on morphological FA dates, with the exception of the biochemical dating for *Rhizosolenia* (Sinninghe-Damsté *et al*. 2004). References that likely contain contaminations were avoided.

The origin of the diatoms was constrained at either 250 or 190 Ma. These two dates were used to constrain the date of the root or origin of the diatoms for the maximum dates and minimum dates respectively. Because of the relative uncertainty of these two dates the maximum and minimum constraints were not set in the same analysis but rather ran both as independent analyses. This was done upon advice from Thorne who wrote Multi-divtime.

Parsimony analysis: A parsimony analysis (run in PAUP) of selected valve features was performed to study the position of two key fossil thalassiosiroids in the phylogenetic tree and the results are presented in the supplementary file 2. This was done in order to ascertain the position of two genera, *Praethalassiosira* and *Thalassiosiropsis,* in either the thalassiosiroid or coscinodiscoid lineage to provide justification for earlier dates for the thalassiosiroid lineage, which has an FA at 40 Ma.

RESULTS

Justification for selection of dates to constrain the root of the tree

Medlin *et al*. (1997) and Medlin (2011) speculated that the pigmented heterokonts to, which the diatoms belong diversified after the mass extinction at the P/T boundary (ca. 250 Ma) and this was mainly based on molecular clock methods that took geological events into account. Clock calculations for the same node in the clock by Berney & Pawlowski (2006) are older at 310 Ma. Sorhannus (2007) estimated the origin of the diatoms at c. 250-183 Ma with a relaxed molecular clock from SSU

Table II. – Geologic 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]	Inferred Clock Date root at 250 MA	Literature[c]
Asterionellopsis[d]	Late Miocene (Messinian): 6.5-5.3	5.3	170	Schrader & Gersonde (1978)
Aulacoseira	Uppermost Late Cretaceous	65.5[e]	150	Ambwani *et al.* (2003), Wolfe & Edlund (2005)
Chaetoceros	Paleocene: *c.* 65-55	55	180	Fenner (1991)
Coscinodiscophyceae	Early Cretaceous (Aptian-Albian): 115-110	110	230	Gersonde & Harwood (1990), Harwood & Gersonde (1990)
Cyclotella	Early Eocene: 24	24	80	Bradbury & Krebs (1995)
Cymatosira	Early Eocene: *c.* 50-55	50	180	Homann (1991), Fenner (1994)
Diatoma	Late Eocene to Oligocene	33.9	88	Lupkina & Dolmatova (1975)
Dimeregramma	Miocene	5.33	162	Schrader & Fenner (1976), Reháková (1980)
Eunotogramma	Late Cretaceous	65.5	180	- (p. 377 & Sims *et al.* (2006)
Fragilaria	Late Eocene: *c.* 45-40	40	105	Lohman & Andrews (1968)
Grammatophora	Late Eocene: *c.* 45-40	40	115	Desikachary & Sreelatha (1989), Edwards (1991)
Hyalodiscus	Cretaceous (Albian-Campanian)	70.6	208	Tapia (1996), Tapia & Harwood (2002)
Mediophyceae	Early Cretaceous (Aptian-Albian): 115-110	110	218	Gersonde & Harwood (1990), Harwood & Gersonde (1990)
Odontella	Late Cretaceous	65.5	198	Hajòs & Stradner (1975), Harwood (1988)
Opephora	Early late Miocene	7.25	132	VanLandingham (1985)
Pennate	Late Cretaceous (Campanian): 75	75	218	- (p. 381 & Sims *et al.* (2006)
Rhabdonema	Late Eocene: *c.* 45-40	40	105	Desikachary & Sreelatha (1989), Edwards (1991)
Rhaphoneis	Late Eocene: *c.* 45-40	40	162	Andrews (1975)
Raphid diatoms	Late Cretaceous (Maastrichtian)	65.5	165	Singh *et al.* (2007)
Rhizosolenia[f]	Late Cretaceous (Upper Turonian): 91) 5±1.5	90-93[g]	225	Shinninghe-Damsté *et al.* (2004)
Staurosira[h]	Miocene	5.33	85	Hajòs (1968)
Stephanodiscus	Miocene	5.33	80	VanLandingham (1967)
Stephanopyxis	Late Cretaceous (Late Cenomanian-Santonian): 95-80	80[i]	218	Tapia (1996)
Surirellaceae	Middle Miocene	16.61	80	Reinhold (1937), Hajòs (1968 & 1986)
Thalassiothrix	Early Oligocene	28.4	85	- (p. 388 & Sims *et al.* (2006)

[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 and 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 Sims *et al.* (2006), then that is new original information.
[d] Used for the constraint of *Asteroplanus*.
[e] Early Cretaceous marine genus *Archepyrgus* can be an ancestor of *Aulacoseira* (Gersonde and Harwood (1990; Harwood and 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 my 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 staurosiroid diatom.
[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.)

sequences. Therefore it is reasonable to assume that the diatoms originated no earlier than 250 Ma because the younger date (250 vs. 310 Ma) is based on clocks generated by both Medlin and Sorhannus (*op. cit.*) with more diatom FA dates (calibration points) in their trees than that of Berny & Pawlowski (2006). To set this upper limit

(oldest age) strictly and to disallow this age any deviation standard deviation of rttm was set to zero and also the upper and lower constraints were set to 250 m.y. (note that the lower constraint was actually set to 249.99 because of the technical limits of the program). In addition Ogawa and Kawata (1996) reported no diatoms from a well-preserved Opal A oceanic radiolarian and sponge spicule-rich deposit from Upper Jurassic (Tithonian) deep sea deposit. Nevertheless this does not exclude the possibility that the diatoms originated before 250 Ma because an ancestral 'Ur' diatom likely had a weak/no silica cell wall (Round & Crawford 1981) unlike recent taxa making it difficult for them to be preserved before the silica cell wall was well developed.

The assumption of the minimum age of the origin may be also controversial. 190 m.y. was set for the minimum age of the diatom origin. This assumption is based on the earliest known but unconfirmed record of diatoms described by Rothpletz (1896 1900), which has been found from the Toarcian stage of the Jurassic (c. 190 Ma by Sims *et al*. 2006). A translation of the 1896 study by Rothpletz reveals he examined sponges and that the treatment of the sample confirms that they have to be siliceous microfossils and not calcareous e.g. Schizosphaerellids.

Rothpletz (1896): "Es kann keinem Zweifel unterliegen dass die kleinen Kieselpanzer welche in manchen dieser liasischen Schwämme in ungezählten Massen in anderen aber auch seltener vorkommen isolierte Schalen von Diatomeen sind." There can be no doubt that the small siliceous shells but, which occur in countless masses in some Liassic bath sponges but also more rarely in others, are isolated shells of diatoms.

"Um zunächst ihre äussere Form zu studieren löst man am besten ein Stückchen des fossilen Schwammes in verdünnter Salzsäure auf. Die Kieselpanzer bleiben dann zurück während die Coccolithen Spongiennadeln und Foraminiferengehäuse alle in Lösung gehen". To study first their outer form it is best to dissolve a piece of fossil sponge in diluted hydrochloric acid. The silica shells remain then while all the coccoliths, sponge spicules and Foraminifera dissolve.

"Besser eignen sich zu ihrer Beobachtung die Panzer in den Dünnschliffen die nicht selten aber erst mit starker Vergrösserung eine äusserst feine und regelmässige Gitterzeichung erkennen lassen". The shells of the diatoms are best suited to their observation in the thin sections through which an extremely fine and regular grid drawing can be quite often recognized but only with strong enlargement.

The next confirmed fossil date of the diatoms is at 140-135 Ma in Korea (Harwood *et al*. 2007, Chang & Park 2008). There is a radical change in diatom morphology between the early Albian diatoms 115 Ma (Harwood & Gersonde 1990) and the late Albian amber diatoms 100 Ma (Girard *et al*. 2009). Adding 40 m.y. to the confirmed Korean deposit yields an age nearer the Rothpletz date. An age of 190 Ma is not unrealistic and the treatment of the Rothpletz deposit in hydrochloric acid con-

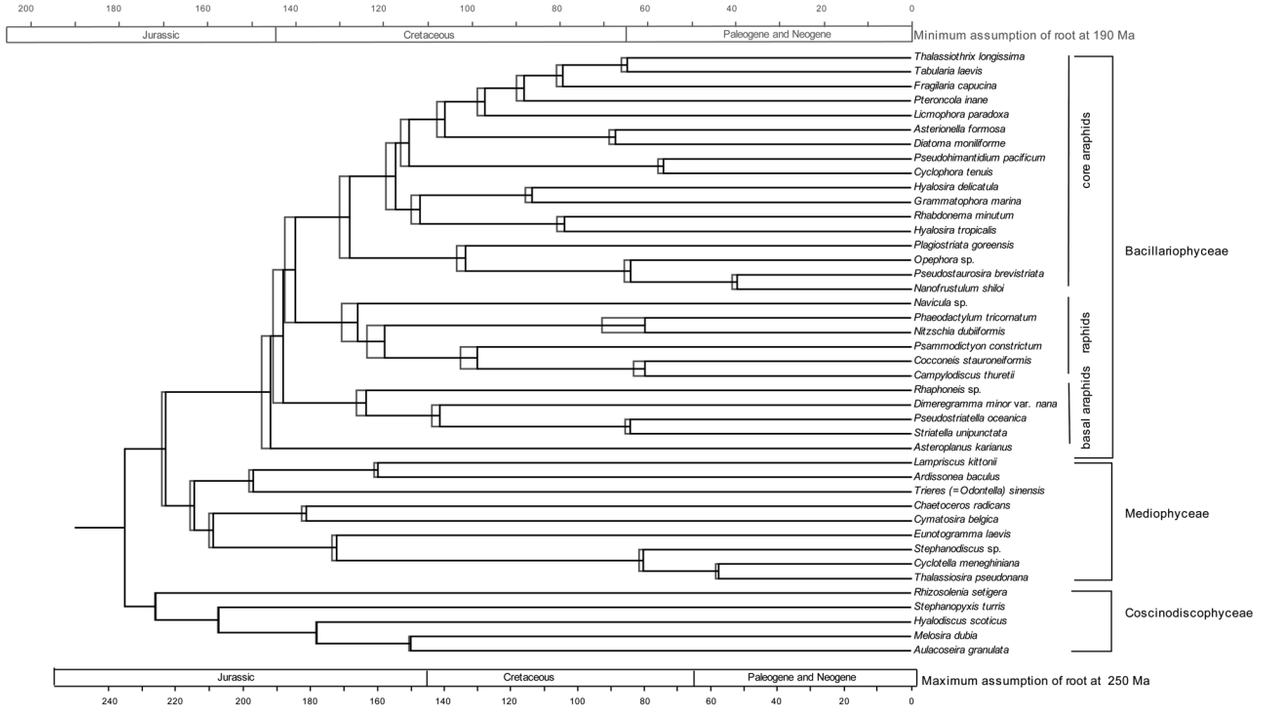

Fig. 1. – Molecular timescale for diatoms using a ML analysis of four genes with a maximum inference of the age of the root a 190 Ma (gray) and 250 Ma (black) tree based on four taxa outgroup rooting dataset. Priors of 250 and 190 Ma for the root of diatoms and respective time scales are indicated above and below the tree topology.

firms that it is siliceous. This leaves some time but probably not enough for the Ur diatom to evolve realistically from a naked flagellate to one covered with organic scales for the scales to begin to incorporate opaline silica for the cells to become non motile and change their life cycle for the resting stage to become the vegetative stage and finally for the unusual size restoration feature of their sexual reproduction to be developed. These are major physiological changes that would have had to occur in the lineage before any wall material would have been sufficiently developed to leave enough material to be preserved. Thus a lower limit at 190 Ma and an upper limit of 250 Ma seem reasonable.

Molecular dating of the diatom classes

In the ML tree (Fig. 1) using multiple outgroups plus bolidomonads as the closest relative to the diatoms, all three classes of the diatoms (Coscinodiscophyceae, Mediophyceae and Bacillariophyceae) are monophyletic: (1) the radial centrics (2) the bi(multi) polar centrics and (3) the pennates. This new clock based on an ML tree suggests that the radial centrics Subdivision Coscinodiscophytina, Class Coscinodiscophyceae diverged from the remaining diatoms belonging to the Subdivision Bacillariophytina at 230 (180) Ma [max (min)]. Within the Bacillariophytina the bi(multi) polar centrics Class Mediophyceae diverged from the Class Bacillariophyceae at 218 (172) Ma. The Bacillariophyceae radiated at 190 (143) Ma (Tables II, IV, Fig. 1). The tree in Fig. 1 shows the range of divergences using either the maximum age of the root at 250 Ma (black lines) or a minimum age of the root at 190 Ma (grey lines). The differences between the inferred ages of the lineages in the radial centric Coscinodiscophyceae are negligible for this clade. There are minor differences for the bi(multi) polar centrics between the maximum and minimum ages, whereas the greatest differences between the maximum and minimum ages inferred from the molecular clock occur in the pennate lineage with the divergence of *Phaeodactylum/Nitzschia* showing the greatest anomaly.

Phylogenetic analysis within the major clades and their dates of divergence

The topology of the coscinodiscophycean clade contained the following relationship: (*Rhizosolenia* (*Stephanopyxis* (*Hyalodiscus* (*Melosira Aulacoseira*)))). The major difference between this 4 gene tree and that of the SSU alone is that *Stephanopyxis* is not sister to *Melosira* and *Aulacoseira* making the Order Melosirales paraphyletic. Dates for the emergence of *Rhizosolenia* at 220 (170) pre-date the appearance of isoprenoid compounds in the fossil record at 93 Ma. The inferred divergence of *Melosira* from *Aulacoseira* at 150 Ma is not unreasonable given the presence of *Aulacoseira's* most likely fossil ancestor of *Archepyrgus* at 110 Ma.

All of the divergence times for taxa in the Mediophyceae greatly exceeded their first appearances in the fossil record (Table II). The topology of the mediophyean clade is very different from that found in other studies. A clade of *Eunotogramma* and Thalassiosirales (*Stephanodiscus* (*Thalassiosira Cyclotella*)) diverged at the root of the Mediophyceae suggesting that the radial valve of Thalassiosirales may not be a consequence of secondary loss of bi(multi) polar shape as suggested before based on SSU phylogenies (Sims *et al*. 2006). However this earlier divergence may be an artefact of taxon sampling because the Lithodesmiales are not included in this tree. A highly elongated diatom *Ardissonia* always diverged last among mediophycean diatoms as sister to *Lampriscus* reconfirming that this diatom is not a pennate diatom but a highly elongated bipolar centric diatom (Kooistra *et al*. 2003, Medlin *et al*. 2008). *Trieres* (= *Odontella* in part) was sister to this clade as recovered in the SSU tree. *Chaetoceros* and *Cymatosira* were sister taxa and this relationship is a new one because in a previous SSU tree with a single outgroup species, Cymatosirales are often the last clade to diverge before the pennate diatoms. This relationship is likely influenced by the absence of *Attheya* from this dataset.

The pennates have formed a well-supported clade in all analyses done to date including this one, which displayed essentially the same topology in the 4 gene tree as in SSU trees. There are two groups of araphids (basal and core) with the core araphids sister to the raphid diatoms. The only incongruence was the position of the araphid diatom *Asteroplanus*, which has appeared in the core araphid lineage (Medlin *et al*. 2000) but here appears at the root of the pennates to make a third araphid lineage. In the analysis that produced Supplementary Fig. 1, there were also three groups of araphid diatoms. To elucidate the true phylogenetic position of *Asteroplanus,* other affinities of the genus should be included in the further analysis, such as *Asterionellopsis* known as the sister genus in SSU tree and *Bleakeleya* a putative sister genus judged by its morphology (Round *et al*. 1990) although molecular data indicate that they are not (Lobban *et al*. 2011). Li *et al*.'s analysis (2015) did not include this taxon. Inferred dates for the emergence of the basal araphid taxa at 180-140 Ma were two – three times older than their FAs (Table I).

The basal araphid lineage consisted of a robust clade of marine araphid diatoms (*Rhaphoneis* (*Dimeregramma* (*Striatella Pseudostriatella*))). The inclusion of *Striatella* and *Pseudostriatella* into the basal araphids is new because in other phylogenies based only on SSU data the position of *Striatella* has been variable (Medlin *et al*. 2000, Medlin & Kaczmarska 2004, Sims *et al*. 2006) often at the base of the raphids (Li *et al*. 2015). Again inferred dates for diversification within the clade greatly exceeded their known first appearances (Table I).

The core araphid diatoms diverged from the raphid diatom clade at 165-140 Ma and inferred divergence times of these diatoms are in the same order of magnitude as in the basal araphid diatoms (Table I). The first divergence of the core araphid diatoms (155-125 Ma) was a clade comprising small-celled diatoms have greatly reduced or no labiate processes (*Plagiostriata* (*Opephora* (*Pseudostaurosira Nanofrustulum*))) from all remaining araphids with a labiate process unless secondarily lost. Subsequently a marine epiphytic clade ((*Hyalosira delicatula Grammatophora*) (*Hyalosira tropicalis Rhabdonema*)) diverged at 140 to 110 Ma. The recovery of this clade is supported by the fact that both *Grammatophora* and *Rhabdonema* share anisogamic sexual reproduction that has never been observed in other diatoms (Magne-Simon 1962). The clade has no exclusive characters except that they all share internal septa on the girdle bands except for one species. *Hyalosira delicatula* the generitype was sister to *Grammatophora*; whereas the other species *H. tropicalis*, which lacks septate girdle bands was sister to *Rhabdonema* and this relationship has recently been recovered in SSU and plastid trees (Lobban & Ashworth 2013) and a taxonomic revision is necessary and a new genus must be erected to contain *H. tropicalis* or it should be moved into Rhabodema with an emended description. Subsequently, a clade of *Cyclophora* and *Pseudohimantidium* diverged. Both genera are attached marine taxa (epiphytic or epizoic, respectively) and both have a slit structure located at their valve ends. The placement of *Pseudohimantidium* within the core araphid clade suggests that their raphe-like slit called a labiate groove is not a precursor of the raphe (Simonsen 1970). Following the separation of a fresh water clade comprising *Diatoma* and *Asterionella, Licmophora* diverged from clade of elongated araphid diatoms, which included *Pteroncola* at the root at 120-92 Ma. Although only one species of *Licmophora* was sampled here, Sims *et al*. (2006), Medlin *et al*. (2008) and Medlin (2015) found *Licmophora* separated into two clades depending on their means of attachment.

The raphid diatoms diverge into naviculoid vs. nitzschioid as in the SSU tree between 160 and 125 Ma. The canal raphe has also arisen twice (*Nitzschia* vs. *Campylodiscus* and *Psammodictyon*) as shown in earlier single gene phylogenies (Medlin *et al*. 2000, Medlin & Kaczmarska 2004). The marine monoraphid diatoms are also more closely related to canal raphe diatoms than they are to naviculoid diatoms as found in the SSU tree.

DISCUSSION

Monophyly of the diatom classes and morphological support

In this study, the three classes of diatoms were recovered as monophyletic clades with a ML analysis of four molecular markers with multiple distant outgroups within the stramenopiles. The monophyly of the two centric classes has been controversial (Sorhannus 2004, Theriot *et al*. 2009) but recovering a grade of clades is useful for sequences of evolution (Medlin & Kaczmarska 2004).

Two factors affect the recovery of the monophyly of the three classes: secondary structure alignment and multiple distant outgroups: Monophyly is not recovered unless a secondary structure alignment following the van de Peer model (van de Peer 1993, 1998) and not the Gutell model (Theriot *et al*. 2009), which excludes any secondary structure for the highly variable and phylogenetic informative V4 region of the SSU gene. The tree used for this clock and that also found in the Supplementary Figs 1-4 were made from the secondary structure analysis of the SSU and LSU genes using the van de Peer model with only a few of the most variable positions in the V4 regions excluded (Supplementary file 2) plus 3 or 2 protein genes respectively and have recovered monophyletic classes. Medlin (2009, 2014) and Medlin *et al*. (1993, 2009) have shown empirically that using an SSU alignment not based on secondary structure will also cause the centric classes to become a grade of clades rather than two monophyletic groups.

Even though secondary structure analyses have been used in more recent other analyses from other workers (Theriot *et al*. 2009, Theriot *et al*. 2010, Ashworth *et al*. 2013, Theriot *et al*. 2015) the failure to recover the three monophyletic classes is likely the result of using a single outgroup species. A single bolidomonad outgroup is often used in analyses using single genes (SSU) and now in multi-gene analyses (Theriot *et al*. 2009, Theriot *et al*. 2010, Ashworth *et al*. 2013, Theriot *et al*. 2015). In only two cases have the three classes all found to be monophyletic when a single outgroup has been used (Table III). In 2004, Medlin and Kaczmarska first suggested that having more distant outgroups would influence the monophyly of the classes. Sato (2008) and Medlin (2014) have shown empirically that the choice of the number of outgroups will affect the monophyly of the classes in some analyses but not in others with more distant outgroups influencing the monophyly of the classes. Unless multiple representatives of each distant outgroup are selected, long branch attraction may occur and affect the topology of the ingroup of interest (Phillippe 2000). Certainly the BI analysis did not recover monophyletic classes in the Sato data set (Sato 2008) but the ML analysis did and it is this tree that was subjected to a molecular clock analysis in this study. Reanalysis of his data using different models for amino acid substitution have recovered three monophyletic classes with high bootstrap support (Supplementary Figs 2 & 3).

Because not all of the analytical methods with multiple genes and outgroups and a secondary structure alignment for the SSU gene render the three classes monophyletic, Sato (2008) among others, considers the monophyly of

Table III. – Summary of the results from major studies on the evolution of the diatoms and whether or not the classes were recovered as monophyletic (**M**) and if not how many clades could be assigned to each class for the centric diatoms. For the araphid diatoms, the recovery of basal and core araphids is indicated. Numbers in brackets refer to the highest bootstrap or posterior probability support for the clade, even if multiple types of analyses were conducted. The number of outgroups used in each study is also indicated because this will have an effect on the monophyly of the clades (Medlin 2014).

Study	**Genes Used**	**Coscinodiscophyceae**	**Mediophyceae**	**Bacillariophyceae**	**Basal +Core Monophyletic clades**	**No. of outgroups**
Study	SSU	**M**[1]	4	**M**[1]	yes[1]	2 bolidomonads
Alverson *et al.* 2006, fig. 3	SSU	**M**[1]	3	**M**[1]	yes[1]	2 bolidomonads
Alverson *et al.* 2006, fig. 4	SSU	2 plus *Ellerbeckia*	2	**M** (69)	Basal (97) + 4 core	Multiple bolidomonads + Chrysophyte
Alverson *et al.* 2006, fig. 5	SSU	3 plus *Ellerbeckia*	1 + 2 clade polytomy	**M** (96)	Basal (59) + 5 core polytomy	Multiple bolidomonads + Chrysophyte
Alverson *et al.* 2006, fig. 6	SSU, rbcL psaC	3	**M** > 95	**M** (100)	Basal (<50) + 2 core	1 bolidomonad
Ashworth *et al.* 2012, fig. 48		5	**M** (>75)	**M** (100)	2 basal + 1 core	1 bolidomonad
Ashworth *et al*. 2013 Supp. fig. 1	SSU	2	**M** (100)	**M** (100)	Only basal	Mulitple crown group
Brown & Sorhannus 2010[2]	SSU	2	**M** (70)	**M** (98)	yes (92 +58)	Multiple crown groups
Cavalier-Smith & Chao 2006	SSU	**M** (80)	3	**M** (100)	yes (99+100)	One bolidomonad
Choi *et al.* 2008, fig. 7[2]	*rbc*L	3 clades	**M (79)**	3 clades with one centric	Only basal	One bolidomonad
Choi *et al.* 2008, fig. 8[2]	COX 1	**M** (79)	**M** (93)	**M** (-)	Only basal	Haptophytes and Heterokonts
Ehara *et al.* 2000 NJ analysis	LSU	2	2	n/a	One Pennate	
Lee *et al.* 2013	SSU, LSU, *rbc*L, *psb*C	**M** (66)	**M** (66)	**M** (100)	yes (76 + <50)	Single bolidomonad
Li *et al.* 2015.	SSU, *rbc*L *psb*C	**-**	**M** (100)	**M** (100)	Basal (< 50) + 3 core	Mediophytes
Lobban & Ashworth 2014	SSU	**M**	2-3 clades	**M**	yes (100+97)	Multiple heterokonts
Kooistra & Medlin 1996	SSU	**M** (88)	3 + *Biddulphiopsis*	**M** (99)	yes (96+ -)	4 radial centrics
Kooistra *et al.* 2003[2]	SSU	**M** (63)	**M** (98)	**M** (100)	Not shown	Multiple
Medlin & Kaczmarska 2004 fig. 1	SSU	4	4	**M** (100)	Basal (100) + 5 core	Multiple bolidomonads
Medlin & Kaczmarska 2004 fig. 3	SSU	**M** (68)	**M** (95)	**M** (100)	yes (53+93)	Multiple
Medlin *et al.* 1993	SSU	**M** (91)	3	**M** (100)	yes (93+100)	Single
Medlin *et al.* 1996a	SSU	**M** (77)	3	**M**	yes (91+100)	Multiple heterokonts
Medlin *et al.* 1996b	SSU	**M** + *Ellerbeckia*	4	**M** (91)	yes (90+ <50)	Single
Medlin *et al.*2000	SSU	**M** (100)	**M** (100)	**M** (99)	yes (100+97)	4 radial centrics
Medlin *et al.* 2008	SSU, LSU, *rbc*L, *psb*A	**M**(< 50)	**M** (51)	**M** (91)	yes (76+<50)	Multiple heterokonts

Table III. – Continued.

Study	Genes Used	Coscinodiscophyceae	Mediophyceae	Bacillariophyceae	Basal +Core Monophyletic clades	No. of outgroups
Sato 2008, fig 4, Publication 10	SSU	**M** (100)	**M** (100)	**M** (100)	yes (100+100)	Multiple heterokonts
Sato *et al.* unpubl.	SSU LSU	**M**[1]	**M** (only 1 taxon)	**M**[1]	yes[1]	Multiple heterokonts
Sims *et al.* 2006	SSU	3	3	**M**[1]	yes[1]	Single
Sinninghe-Damsté *et al.*2004	SSU	**M**	**M**	**M**		Single
Sorhannus 1997, Fig. 1	SSU	3, polytomy	**M**[1]	**M**[1]	yes[1]	Single
Sorhannus 2004, Fig. 1	SSU	**M**[1]	4	**M**[1]	yes[1]	Single
Sorhannus 2004, Fig. 2	SSU	3 + *Ellerbeckia*	8	**M**	yes[1]	Single
Sorhannus 2004, Fig. 3	SSU	2	3	**M**[1]	Basal[1] +multiple core	Multiple
Sorhannus, 2007	SSU	3	**M** (90)	**M** (99)	Basal (77) + multiple core	Single
Theriot *et al.* (2009, Fig. 1) Unweighted MP	SSU	3	**M** (52) + *Attheya*	**M** (100)	yes (95+95)	Rooted midway radial
Theriot *et al.* (2009, Fig. 2)	SSU, *rbc*L, *psb*C	2	4	**M** (100)	2 basal+2 core	Single
Theriot *et al.* (2009, Fig. 8)	SSU, *rbc*L, *psb*C	2 + *Corethron*	**M** (56)	**M** (100)	Not shown	Single
Theriot *et al.* (2010 fig 3)	SSU, *rbc*L, *psb*C *psb*A, *psa*A, *psa*B, *atp*B	3	3	**M** (83)	Yes (53+79)	Single
Theriot *et al.* (2011, fig. 1)	SSU	**M** (100)	**M** (88)	**M** (95)	Not included	multiple

[1] no bootstrap values given
[2] did not use secondary structure alignment

the three classes a subject of continued study and analysis. However trees are now being recovered that consistently find the Mediophyceae to be monophyletic (see Table III and now in Supplementary Figs 1-4, the Coscinodiscophyceae are monophyletic). Taking multiple outgroups outside of the heterokonts has eliminated this problem of some analytical methods with SSU genes alone (Kaczmarska & Medlin 2004, Medlin 2014), whereas with multiple genes grades of clades are still recovered with single outgroups, viz. bolidomonads. Using three genes Theriot *et al*. (2010) have recovered a monophyletic Mediophyceae and multiple clades for the Coscinodiscophyceae using a single outgroup the Bolidophyceae. Ashworth *et al*. (2012, 2013) recovered a monophyletic Mediophyceae with high bootstrap support and multiple clades for the Coscinodiscophyceae again using a single outgroup the Bolidophyceae. Li *et al*. (Supplementary Fig. 4 using the van de Peer secondary structure model) have recovered all three classes monophyletic with high bootstrap support using two bolidomonads but with 157 diatoms and three genes with a high proportion of araphid taxa. Clearly the number of outgroups needed to recover monophyly of the classes is a factor that should be continuously rigorously tested in the diatoms. Medlin (2014) has begun a first attempt to analyze the effect of various outgroups on the class monophyly using only the SSU gene. Here ML analyses with multiple genes have recovered monophyletic classes, although there is no bootstrap support for this (Supplementary Fig. 1). However there is high bootstrap support for the most recent analysis of the diatom classes using 3 genes and 159 taxa (Supplementary Figs 2-3); thus the tree used in this clock calculation, despite the low bootstrap support originally obtained by Sato (2008), is reasonable for estimating the divergence of the three classes. Medlin in Sims *et al*. (2006) and in Bowler *et al*. (2008) and Medlin (2014) have obtained 100 % posterior probabilities (PP) for the three classes in a Bayesian analysis with a much larger data set based on the 18S alone and multiple outgroups.

The three classes are supported by clear morphological differences, viz. the structure of the specialised diatom zygote, the auxospore, the position of the Golgi bodies in the cell, and the position of the processes or tubes through the silica wall inside the annulus, no matter how much of the entire valve face the annulus occupies. All of these are discussed in detail in Medlin & Kaczmarska (2004) and in Medlin (2014). Kaczmarska & Ehrman (2015) have provided new data from the auxospore initial valve that

supports the three classes of the diatoms. Their evidence suggests that the auxospore initial valve in the Coscinodiscophyceae more spore-like in appearance than that in the Mediophyceae and the initial valves of the bacillariophycean or pennate initial valve are like vegetative valves. Thus, the initial valve had undergone three stages of evolution from being spore like and very different to vegetative valves in the basal centrics to nearly identical to vegetative valves in the pennates. This evidence supports the hypothesis by Mann & Marchant (1989) that the diatom valve has evolved from a spore like ancestral cell. In addition to the above points, an external cribrum covering the loculate areolae is present in the Class Coscinodiscophyceae. One exception to the external cribrum in the Coscinodiscophyceae is that of *Endictya* from the Upper Cretaceous by Harwood (1988), which has an internal cribrum and a marginal ring of processes that appear to be simple tubes (Round *et al*. 1990) but, which are in fact small labiates. This genus has modern representatives and would appear to be misplaced in the Coscinodiscophyceae (Round *et al*. 1990) because it exhibits eccentric areolation typical of the Thalassiosirales rather than radial areolation typical of the Coscinodiscophyceae and therefore it could be misclassified. However recent molecular analyses have placed it in the *Stephanopyxis* clade (Ashworth *et al*. 2013), which confirms its place in the Melosirales of the Coscinodiscophyceae. The position of the cribrum in the pseudo-loculate valve of *Stephanopyxis* is internal and it would appear that *Endictya* also has pseudo-loculate areolae (Anonymous 1975, see references in Medlin 2016a). Nearly all of the diatoms in the early Cretaceous deposit of Gersonde & Harwood (1990) have pseudo-loculate valves with internal cribra and *Stephanopyxis* is likely an early descendant of these diatoms with an early evolution of a labiate process (Sims *et al*. 2006), although Gersonde & Harwood (1990) regard the presence of this diatom as a contaminant in this deposit. Following the argument that the walls of the pseudo-loculate valve might not be analogous to the walls of the loculate valve because one opening are not restricted into a foramen (Anonymous 1975), then the ancestral state is a cribrum flush with the valve face as described in Gersonde & Harwood (1990). In the coscinodiscophycean diatoms, the loculate valves develop proximally to the cribrum and in mediophycean diatoms, valve morphogenesis is in the opposite direction. This has been substantiated with morphogenetic studies from both groups (see Round *et al*. 1990 and in the supplementary cladistic analysis, Supplementary File 2). An internal cribrum is found in most mediophycean diatoms with loculate areolae. Some exceptions, among others, are *Eupodiscus, Pleurosira* and a small group of *Odontella* spp. some of which have now been moved to a new genus (Ashworth *et al*. 2013). Some mediophycean diatoms with simple poroid areolae have the cribrum positioned nearer the internal valve surface e.g. *Isthmia, Eucampia* and *Hemiaulus,* whereas in others the cribrum is nearer the outer surface (see illustrations in Ashworth *et al*. 2013). Several genera in the Thalassiosirales, e.g. *Cyclotella,* have alveolate areolae, which because of their elongated sausage-shaped opening, requires an external cribrum to the areolae. Alveolate areolae and pseudo-loculate areolae are not believed to be homologous structures to loculate areolae (Anonymous 1975).

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. The only anomaly in the current tree is the early divergence of *Asteroplanus,* which should be reinvestigated because it has appeared at the base of the raphids making a third araphid clade and three clades, albeit different ones, were found in the data set shown in supplementary Fig. 4. Each of the major clades deserves to have equal rank in the diatom taxonomy; new subclasses have been described for each of these pennate clades (Medlin 2016b). The four-gene analysis points the way to which groups need further work in order to find defining morphological features, to erect new taxa and has provided better resolution at the base of the tree.

The core and basal araphids have different kinds of auxospores (Medlin & Sato 2009, Kaczmarska *et al*. 2013) with the basal araphids having a combination of both traverse bands similar to the properizonium of the mediophycean bi(multi) polar centrics and the longitudinal perizonial bands of the raphid pennate diatoms. The core araphids have transverse and longitudinal bands in their auxospores, which they share with the raphid pennates. The male sex cell in araphids appears to be different from the female sex cell. Where studied, a filament with microtubules has been seen attached to the amoeboid male gamete, which draws the sex cells together during sexual reproduction. The male sex cells are released from their gametangia and the female sex cell may or may not be released from its gametangium. The newly discovered appendage has been found in the araphids: *Pseudostaurosia trainorii* (Sato *et al*. 2011), *Tabularia fasciculate, T. tabulata* (Davidovich *et al*. 2012) and *Ulnaria ulna* (Davidovich 2012). Amoeboid gametes are known also from *Grammatophora* and *Rhabdonema* (see references in Sato *et al*. 2011) and it is likely this appendage may be the defining feature for all araphids. A summary of the major evolutionary steps in the sexual reproduction of the diatoms can be seen in Fig. 2, which assumes monophyly of the three diatom classes.

Divergence times

This analysis based on four molecular markers recovers the three classes of the diatom as being monophyletic regardless of the time restraint placed on the origin of the diatoms at 250 or 190 Ma. The younger date is not really realistic because it does not give enough time for the dia-

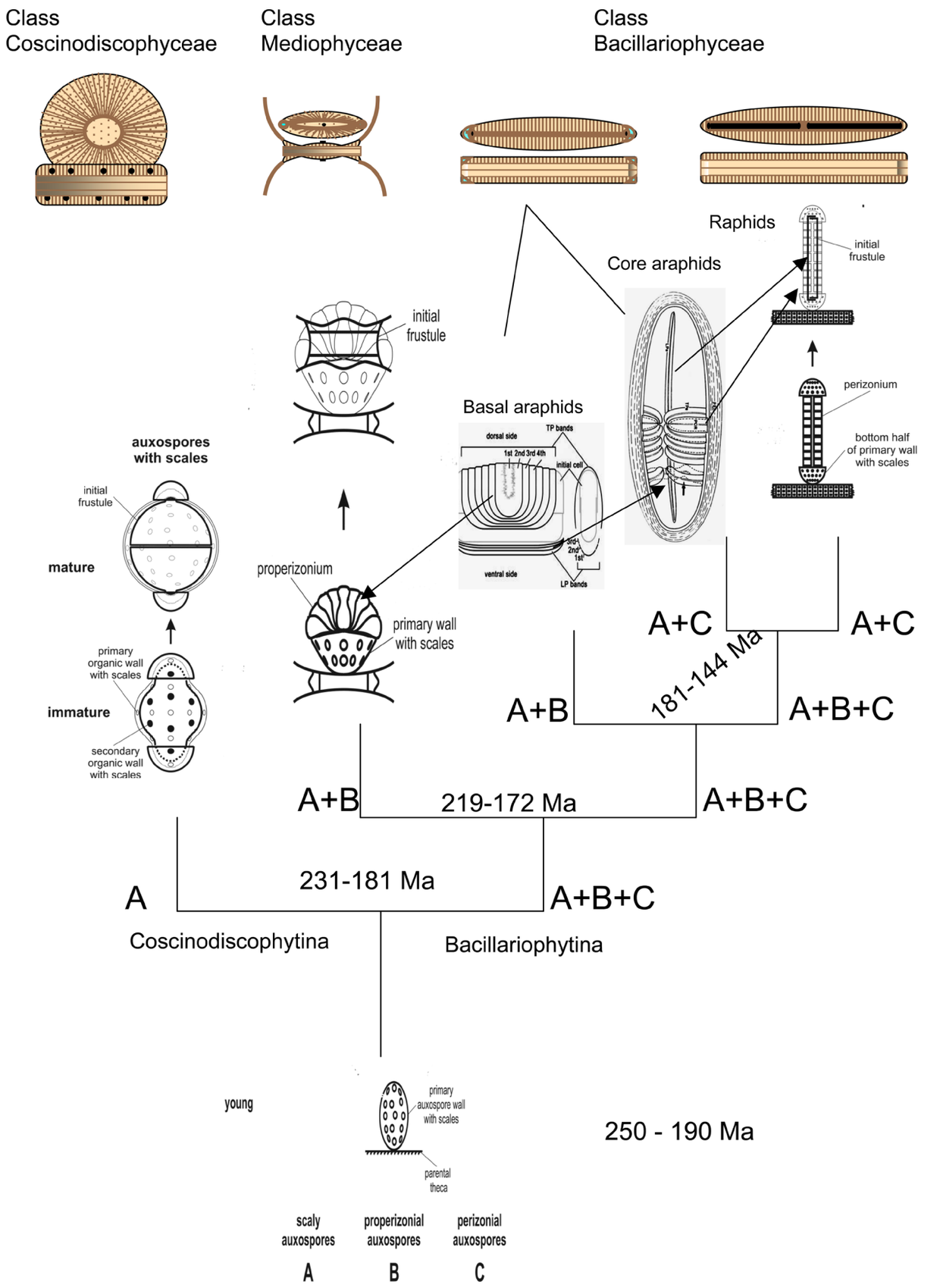

Fig. 2. – Summary of the major features of sexual reproduction in the diatoms and the respective clades in which this occurred (redrawn from Medlin 2014). Divergence times for the major clades are placed on the nodes.

tom wall to have sufficiently evolved from a scaly ancestor to a fully formed wall as those found in the sponges studied by Rothpletz. This clock places the origin of the diatoms [240-190] well within the time frame shown by Sorhannus (2007) Kooistra & Medlin (1996) and Medlin *et al*. (1997) with the exception of the age of the Class Bacillariophyceae, which is shown here to be very much older. Brown & Sorhannus (2010) inferred the origin of the Bacillariophyta to be ca. 370 Ma with a diversification at 320 Ma using a variety of calibration points from across the pigmented heterokonts as well as within the diatoms. Their tree suggests a long time interval from their origin before the diatoms diversify (ca. 65 m.y.) and this is not seen in the diatom tree constructed here from 4 genes. The length of the branch before the first divergence in this tree is ca. 10-15 m. y. from the origin to the first diversification. Likely taking calibration points outside of the diatoms in the Brown & Sorhannus tree instead of only within the diatoms as in the tree shown here has likely caused this large difference in the dating points in the two studies. Sorhannus (1997) estimated a divergence time of 330 Ma for the divergence of the Coscinodiscophytina from the Bacillariophytina, which is considerably older than predicted with this molecular clock.

The pennate diversification predates the FA of a pennate diatom in the Campanian at 75 Ma (Sims *et al*. 2006). With the assumption of a 250 Ma origin, the emergence of the pennates was inferred to be at 219 Ma and the raphid origin at 181 Ma. Therefore a possible range of the divergence time of origin of the pennates (max to min) is from 241-153 Ma and that of raphid diatoms is 205-144 Ma (Table II). These new ages greatly predate previous estimations based on SSU with the origin of pennates at 80 Ma (Sorhannus 1997), 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).

No pennate diatoms have been recovered in any of the 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 & Harwood 1997, Nikolaev *et al*. 2001) and Germany (Forti & Schulz 1932). These deposits are likely nearshore marine and would not be expected necessarily to contain any pennate species. Most extant pennates thrive in benthic areas, either freshwater or marine. Thus, if the early dates inferred here for the origins of pennate diatoms are correct, then that means that early deposits representing benthic habitats suitable for pennate diatoms have yet to be found. Future work should be concentrated to find early benthic deposits to uncover FA of pennate diatoms that would be compatible with the dates inferred by this clock. Rothpletz (1896 1900) suggests that more investigations of fossil sponges should be undertaken from these early times because some specimens contain abundant concentrations of diatoms. Although pyritized diatoms occur throughout fossil record the Late Cretaceous is known for its abundance of pyritized diatoms (Sims *et al*. 2006). Thus it is very possible that the earliest pennate has simply not been discovered. This clock estimation of the origin of the pennates between 219 and 173 Ma greatly predates their first occurrence of fossils from the deposit of the Campanian (Sims *et al*. 2006, Hajòs & Schrader 1975). The origin of the raphid diatoms was also estimated at 181 and 143 Ma predating their fossils from the deposit of the Maastrichtian c.70 Ma (Singh *et al*. 2006). Notably one of the earliest fossils of the raphid diatoms is now classified in the genus *Lyrella* (Witt 1886), 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) so early raphid pennates as well as araphid pennates are clearly missing from the fossil record. It is clear that more work is needed for the early deposits. Therefore if the molecular data are correct, then existing data from stratigraphic studies must underestimate the age of the pennates and this new origin is not so unrealistic (Sims *et al*. 2006).

My choice of molecular clocks used here is a relaxed clock so variable rates of evolution in the centrics vs. the pennates (Kooistra & Medlin 1996) are likely not a factor in pushing the dates of origins back in geological time. In general, most of the inferred fossil dates exceed the FA of the extant taxa in this tree (compare inferred tree dates with FA dates listed in Table I). These differences are most noticeable among the pennates. Because the differences between the max and the min ages for the centrics are not significantly different, it is unlikely that the dates for the rooting of the diatom origin are causing the FA discrepancies in the pennate lineage. There is no adequate explanation why this clock has pushed back the molecular origins of the pennates but it is likely a combination of a relaxed clock, the four genes and multiple calibration points (Welsh & Bromham 2005). In the most recent molecular clock for the haptophytes from four genes (Liu *et al*. 2009), discrepancies to other clocks were also found in not only the divergence of the entire group but also in divergences of some of the major clades. In each case the four-gene clock found younger times of origins. The use of a relaxed molecular clock was considered to be one of the major reasons for the time differences.

Ghost lineages

In Sorhannus (2007), some discrepancies in the molecular clock inferred age and the FA for two genera and gaps in the fossil record for one other lineage were also recovered. He referred to these differences as ‘ghost ranges’. These ghost lineages are also seen in this tree for the same taxa included in both trees. These new clock calculations also indicate that these lineages are older than their first

Table IV. – Estimated time (Ma) of the origin of major clades using ML topology (with DistIn dataset). 95 % credibility range is in square brackets.

Origin of diatoms assumed at	Separation of Coscinodiscophytina from Bacillariophytina	Separation of Mediophyceae from Bacillariophyceae	Diversification of pennate diatoms
190 Ma	180.6 [189.7-160.4]	172.1 [184.9-152.7]	143.9 [158.8 -136.5]
250 Ma	230.8 [249.3-193.2]	218.8 [241.0-182.8]	180.7 [204- 150.0]

appearances. However these anomalies in the phylogenies of the diatom clock and tree were already addressed in earlier clocks and phylogenetic studies and are not the result of misdiagnoses or as yet undiscovered taxa in the fossil record as believed by Sorhannus (2007). These earlier analyses are herein revisited to offer explanations why the molecular age of these ghost lineages precedes their FA. Alverson (2014) has also found ghost lineages in the recent molecular clock for the Thalassiosirales; discrepancies in FA and molecular inferred origins were found for *Stephanodiscus Shinodiscus* and *Cyclotella*.

The long branch leading to the Thalassiosirales before its modern radiation was first seen and discussed by Medlin *et al*. (1996) and extensively analysed in Kooistra and Medlin (1996). They reported that the origin of the Thalassiosirales lineage did not correspond to the fossil record of the first appearance of the group in the Eocene when we have the first documentation of a strutted process the defining feature of the order (Wolfe & Silver 2009). Thus, this FA is likely the first appearance of the strutted process and not the molecular origin of the lineage and thus corresponds to the modern radiation of the order, which is defined by the presence of a strutted process. This means that the ancestors of modern Thalassiosirales did not possess strutted processes and we should look for members in the fossil record that do not have this feature. Hasle and Syvertsen (1985), Medlin *et al*. (2000) and Gersonde & Harwood (1990) all suggested that *Thalassiosiropsis* could be the likely ancestor of modern Thalassiosirales. Medlin *et al*. (2000), Medlin & Kaczmarska (2004), Kaczmarska *et al*. (2006) and Sims *et al*. (2006) discussed possible changes that would have occurred in the morphology of the uvular tube process of *Archaegladiopsis* and the annular tube process seen in *Thalassiosiropsis* and *Gladiopsis* to evolve into the central strutted process of the Thalassiosirales and then into the central labiate of the bi (multi) polar centrics. A formal cladistic analysis is presented in the supplementary data (Supplementary File 2) and the steps outlined above in the transition from *Thalassiosiropsis* to *Thalassiosira* amount to gains of characters. *Praethalassiosiropsis* is considered to be directly linked to *Thalassiosiropsis* thus placing them either as direct ancestors of or a sister lineage that diverged from the lineage leading to modern Thalassiosirales as far back as 107-113 Ma (Harwood & Nikolaev 1995 now redated by McCartney *et al*. (2014). *Praethalasiosiropsis* and *Thalassiosiropsis* have to be in the Thalassiosirales lineage because there are no other radial centrics with an internal cribrum to their loculate areolae and central processes other than those of the Thalassiosirales and other mediophycean diatoms unless the areolae are pseudoloculate, which they are not because a foramen is present. Thus, it can be safely assumed that *Praethalasiosiropsis* and *Thalassiosiropsis* are early ancestors in the thalassiosiralean lineage and the cladistic analysis shown in Supplementary File 2 places them in this lineage and not with the coscinodiscophycean lineage.

All other radial centrics have an external cribrum and belong to the Coscinodiscophyceae (Round *et al*. 1990). *Endictya* and *Stephanopyxis* in the Coscinodisophyceae have pseudo-loculate areolae with an internal cribrum with eccentric areolation. Molecular analysis has shown *Endictya* to be related to *Stephanopyxis* (Ashworth *et al*. 2013) and the modern species sequenced has marginal labiate processes but the specimens in Round *et al*. (2002) appear to have only marginal tubes but these may be eroded specimens. Thus, the small labiate processes have likely evolved from simple tubes and these simple tubes through the valves in the fossil *Endictya* spp. (Round *et al*. 1990) could be considered forerunners of the small labiate processes.

Other molecular investigations have shown that the true sister group of the Thalassiosirales is the Lithodesmiales with high bootstrap support (Medlin & Kaczmasrska 2004, Kaczmarska *et al*., 2006 Alverson 2014). Kaczmarska *et al*. (2006) discussed possible evolutionary scenarios in which a structure similar to the marginal fringe found in modern Lithodesmiales found in the last common ancestor with the Thalassiosirales could have evolved into the strutted process of the Thalassiosirales. They also discussed the possibilities that the central strutted process is not homologous with the marginal ones and presented four possible scenarios for their separate or homologous evolution. The discovery of *Mediopyxis* in the Lithodesmiales with occluded strutted processes (Kuhn *et al*. 2006) the purported ancestor of the strutted process (Syvertsen & Hasle 1982) gives further support to these scenarios. The variety of shapes of central tube structures in *Mediopyxis* also support the hypothesis that a central tube structure could have evolved into the central labiate of the bi (multi) polar centrics or into a strutted process in the Thalassiosirales. All of these investigations have pointed the way to the type of diatom valves that should be considered as possible ancestors bridging the gap between the FA of *Praethalassiosiropsis* at 107-113 Ma through to *Thalassiosiropsis*/*Gladiopsis* at the late Cretaceous to modern Thalassiosirales with strutted processes with a FA at 40 Ma.

Kooistra & Medlin (1996) first introduced the possibility that the various labiate processes in the diatoms were not homologous and suggested that the central strutted process of the Thalassiosirales, the central labiate process of the bi(multi) polar centrics and the raphe of the pennates were all derived from the same ancestral structure a simple tube like that found in *Thalassiosiropsis*. Kaczmarska *et al*. (2006) and Sims *et al*. (2006) have also suggested this because the labiate process in both of the new classes is very different in both lineages. The macrolabiate process of the Coscinodiscophyceae is more similar morphologically to the unusual single and normally marginally positioned labiate process in the Thalassiosirales, whereas the microlabiate process of the Coscinodiscophyceae is more similar to the labiates of the bi (multi) polar centrics. Sims *et al*. (2006) also suggested that the central strutted process of the Thalassiosirales and the central labiate of bi (multi) polar centrics were both derived from the central structure consisting of a simple tube as seen in *Gladiopsis, Prethalassiosiropsis* and *Rhynchopyxis* in Gersonde and Harwood's Group 2 fossil diatoms (1990). The presence of both tubes and labiates process in *Endictya* spp. support the hypothesis that the microlabiate processes have evolved from simple tubes as have the strutted processes. Certainly the presence of multiple types of tubes in *Mediopyxis* also supports this hypothesis (Kuehn *et al*. 2006).

Most recently modern radial centric diatoms (*Stephanopyxis* and *Aulacoseira*-like cells) have been found in amber dated from the latest early Cretaceous (Girard *et al*. 2009). Many fossil diatoms in strata of this age are pyritized making them unidentifiable. The presence of these coscinodiscophycean diatoms at this time whose extant members have true labiate processes lends further support to the hypothesis that the labiate process has evolved more than once.

The second ghost lineage cited by Sorhannus (2007) is that of *Aulacoseira*, which has a first appearance in the fossil record in the middle Eocene ca. 48.6 Ma (Wolfe *et al*. 2006) but its molecular age of the lineage to which it belongs is older extending back to ca. 99 Ma in the clock of Sorhannus and to ca. 150 Ma in this new clock. Sorhannus claims that this discrepancy could be from diatoms misidentified or as yet undiscovered in the fossil record that could push back the FA of *Aulacoseira*. However Gersonde & Harwood inferred *Archepygrus* to be direct ancestors of *Aulacoseira* from a fossil deposit at 107-113 Ma in 1990 as did Medlin *et al*. in 2000 and Sims *et al*. in 2006. A direct comparison of the valve morphology of *Archepyrgrus* at 107-113 Ma and *Aulacoseira* (modern valve) was illustrated in Sims *et al*. (2006) and in Medlin *et al*. (2000). *Archepyrgrus* lacks the labiate processes of *Aulacoseira* as do all of the diatoms found by Gersonde & Harwood (1990) in the fossil deposit at 107-113 Ma, which had to be placed into new families and orders (Nikolaev & Harwood 2001, 2002 and Nikolaev *et al*. 2001) because they did not possess the labiate process whose presence location structure or loss defines all modern centric diatom orders and families. However, the two genera do share a marginal flange called a ringleist, which is only found in extinct and extant Aulacoseirales and the interlocking mechanism between sibling valves and the poroid areolae between these two genera are almost identical (see illustrations of modern *Aulacoseira* in Crawford & Likhoshway (1998). The ringleist is mentioned as a valve feature in the *Pyxidicula* illustrated by Rothpletz (1900). From this high similarity of valve morphology and the unique shared feature of the ringleist it has been inferred that the lineage containing *Melosira* and *Aulacoseira* is very old with origins in the early Cretaceous as documented from the appearance of *Archepyrgrus* in the 107-113 Ma deposit from Australia and Antarctica. The discrepancy between the first appearance age and the fossil date of *Aulacoseira* was first reported by Kooistra & Medlin (1996) because they used the early date of the Cretaceous material to date this node in their clock and the morphological links between the two taxa. The existence of taxa closely related to *Aulacoseira* was first noted by Gersonde and Harwood in 1990 and that *Aulacoseira* is specifically derived from *Archepyrgrus* is stated in Harwood *et al*. (2007).

The third example that Sorhannus cites as a 'ghost lineage' is that of *Paralia sol*, which is now classified as *Ellerbeckia sol* (Crawford & Sims 2006). This species is not in this 4-gene tree but is discussed here because the interpretation of its ghost lineage by Sorhannus is compromised because the combination of *Paralia sol* was used incorrectly. Sorhannus (2007) reported that the ghost lineage of *Paralia* amounted to 84 m.y. This species was transferred to *Ellerbeckia* and appeared under that name in the trees of Sims *et al*. (2006). *Ellerbeckia* is predominately a freshwater/terrestrial genus whereas *Paralia* is a neritic marine taxon.

The recent discovery of diatoms in the earliest Cretaceous (Valanginian) Dabokni Formation ~135 to 140 Ma of the Jasong synthem Chang & Park (2008) have revealed that these lightly metamorphosed sediments were deposited in a terrestrial setting (Harwood *et al*. 2004, 2007). Harwood *et al*. (2004) have put forth a hypothesis that the diatoms likely had a terrestrial origin. Medlin (2007) has drawn attention to the fact that *Ellerbeckia* as a terrestrial genus is the first diatom lineage to diverge using a single outgroup analysis placing potential ancestors of this extant genus in the terrestrial fossil deposit and lending support to the hypothesis by Harwood *et al*. (2004, 2007) that the diatoms had a terrestrial origin. An evolutionary scenario as to how unicellular non-silicified cells could have evolved into diatoms with a terrestrial origin with these terrestrial cells being flushed into marine waters as continents were re-flooded to evolve into near-shore marine waters was detailed in Medlin (2007) see also Nikolaev & Harwood (2001). Thus, there is no discrep-

ancy between the fossil first appearance and the molecular date inferred from the phylogenetic tree for this genus using a correct classification of this species in the terrestrial genus of *Ellerbeckia* and the report of what appears to be resting spore stages of very early diatoms with some similarities to *Ellerbeckia* now from the oldest confirmed diatom deposit a terrestrial deposit (Chang & Park 2008). The diatoms illustrated by Chang & Park (2008: figs 2 c,d,e) from the very early deposits have a valve structure (fenestrations on the valve face) similar to *Truania*, which is related to *Ellerbeckia* (Sims & Crawford 2007).

Thus, the three ghost lineages discussed by Sorhannus (2007) were already known & accounted for in earlier literature and the inferred divergences were also used in the molecular clock of Kooistra & Medlin (1996). Each of his three ghost lineage can be assumed to have existed at the time of its molecular origin of the lineage, which is earlier than that assumed by its FA. Sorhannus (2007) suggested the ghost lineages are the result of misdiagnoses or as yet undiscovered taxa in the fossil record; however the evidence presented here suggests that links in the fossil record were already noted in the literature and his claim is unfounded.

In contrast, the differences in the FA and the molecular dates for the pennate lineages in this current clock are almost certainly a result of undiscovered pennate taxa in the fossil record. This present clock provides support for protracted ghost lineages among all the major clades of pennate diatoms and suggests that the limited studies of shallow water habitats in the fossil record have not yet captured the evolutionary history of pennate diatoms and hence the biogeochemical services that may be attributable to them. Clearly these habitats should be the focus of more stratigraphic studies. This is best substantiated by the first appearance of *Lyrella* as the first raphid pennate diatom in the fossil record (Witt 1886), which is a genus that is not recently diverged in any molecular tree. Perhaps other biochemical data may help with the documentation of early raphid pennate diatoms. Some raphid pennates are known to produce isoprenoids and the report of these compounds in the Cretaceous attributed to *Rhizosolenia* could just have easily been from pennate diatoms. There are reports of non-marine and marine araphids from middle Late Cretaceous (see references in Harwood *et al*. 2007).

Sorhannus (2007) comments that the error bars in Medlin *et al*. (2000) are too great to give a reliable dating to their dates of origin. Whereas it is true that the upper 95 % confidence interval in Medlin *et al*. (2000) pushed the origin of the diatoms back to 266 Ma, the lower 95 % confidence interval was 55 m.y. younger than the first presumed and unconfirmed diatom fossil at 190 Ma and can be ignored. Much tighter confidence intervals are reported for the divergence of the diatom classes in this current clock based on four molecular markers (Table IV). Sorhannus (2007) does not provide any confidence intervals for his dates because these are not available in the PATHd8 program so only a comparison average ages or maximum/minimum ages in the clocks present by Sorhannus (2007) and the present clock can be made. In this paper and an earlier one (Medlin *et al*. 2000), the average age of the origin of the diatoms ranges from 135 to 266 Ma based on four genes and multiple calibration points and Sorhannus' work concludes that the average age of the diatom is from 183 to 250 Ma based on a single gene with one calibration point at a time, which are not too different from one another.

The major difference in Sorhannus' 2007 clock and the one presented here is that the origin of the pennate lineage is pushed much further back in time and the emergence of the three diatom classes can be dated because they are monophyletic with an analysis that uses a secondary structure alignment (van de Peer model) for the SSU genes with multiple distant outgroups. The Brown & Sorhannus clock (2010) did not use a secondary structure alignment and they did not recover monophyletic classes and therefore did not attempt to date any internal lineages in the diatom branch of their tree other than the origin and its diversification. The earlier age of the pennates can be supported because most fossil deposits within this time frame are not benthic in origin, so it is unlikely that their habitat has been sampled. The monophyly of the two centric classes has been a major point of contention in the diatom community and as demonstrated here empirical evidence of the effect of multiple markers and multiple outgroups recovers the monophyly of two classes of centric diatoms and suggest that future workers follow this approach to extend the available data base by adding species to this four-gene phylogeny and using multiple distant outgroups to root their tree. Unless trees that recover monophyletic classes are used for clock calculations, then the age of divergence of the three classes cannot be inferred and morphological and reproductive advances cannot be assigned to each major divergence. The clock presented here allows for such a time frame for diatom evolution to be estimated.

ACKNOWLEDGMENTS. – Dr S Sato performed the molecular clock analyses. Dr H Thüs Natural History Museum London UK translated the German text.

REFERENCES INCLUDING SUPPLEMENTARY ONES

Alverson AJ 2014. Timing marine-freshwater transitions in the diatom order Thalassiosirales. *Paleo* 40: 91-101.

Ambwani K, Sahni A, Kar RK, 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.

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

Anonymous 1975. Proposal for a standardization of diatom terminology and diagnoses. *Nov Hedw Beih* 53: 323-354.

Ashworth MP, Nakov T, Theriot EC 2013. Revisiting Ross and Sims 1971: toward a molecular phylogeny of the Biddulphiaceae and Eupodiscaceae Bacillariophyceae. *J Phycol* 49: 1207-1222.

Behnke A, Friedl T, ChepurnovVA, 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. *Nuc Acid Res* 27: 2825-2831.

Berney C, Pawlowski J 2006. A molecular timescale for eukaryotic evolution recalibrated with the continuous microfossil record. *Proc R Soc B* 273: 1867-1872.

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov EA, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Vérret F, Berges J, Brownlee A, Cadoret C, Chiovitti JP, Choi A, Chang JC, De Martino S , Detter A, Durkin J, Colleen C, Falciatore A, Fournet J, Haruta M, Huysman M, Jenkins B D, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroeger N, Kroth P, La Roche J, Lindquist E, Lommer M, Martin-Jézéquel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Petit Jean-L, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Schnitzler PM, Shapiro H, Siaut M, Stanley M, Sussman MJ, Taylor A, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV 2008. The *Phaeodactylum* genome reveals the dynamic nature and multi-lineage evolutionary history of diatom genomes. *Nature* 456: 239-244.

Bradbury JP, Krebs WN 1995. Fossil continental diatoms paleolimnology evolution and biochronology. *In* Blome CD *et al*. convenors ed, Siliceous microfossils Short Courses in Paleontology, n° 8. The Paleontological Society, Knoxville Tennessee: 119-138.

Brown JW, Sorhannus U 2010. A molecular genetic timescale for the diversification of autotrophic stramenopiles Ochrophyta: substantive underestimation of putative fossil ages. *PLoS ONE* 59: e12759. doi: 10 1371/journal Pone0012759.

Chacon-Baca E, Bernaldi-Campesi H, Cevalla-ferris SRS, Knoll AH, Golubic S 2002. 70 Ma non-marine diatoms from northern Mexico. *Geology* 30: 279-281.

Chang KH, Park SO 2008. Early Cretaceous tectonism and diatoms in Korea. *Acta Geol Sin* 82: 1179-1184.

Crawford RM, Likhoshway YV 1998. The velum of the species of the diatom genus *Aulacoseira* Thwaites. *In* John J ed, Proc 15th Int Diatom Symp. ARG Gantner Verlag, Liechtenstein: 275-287.

Crawford RM, Sims PA 2006. The diatoms *Radialiplicata sol* Ehrenb. Gleser and *R. clavigera* Grun. Gleser and their transfer to *Ellerbeckia* thus a genus with freshwater and marine representatives. *Nov Hed Beih* 130: 137-162.

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 SA, 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.

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

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

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

Dun WS, Rand 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. *N Z Geol Survey 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 1994. Diatoms of the Fur Formation their taxonomy and biostratigraphic interpretation: Results from the Harre borehole Denmark Aarhus. *Geoscience* 1: 99-163.

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

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

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.

Girard V, Saint Martin S, Saint Martin JP, Schmidt ARE, Struwe S, Perrichot V, Breton G, Néraudeau D 2009. Exceptional preservation of marine diatoms in upper Albian amber. *Geology* 37: 83-86.

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

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

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

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

Hajós M 1968. Die Diatomeen der Miozänen Ablagerungen des Matravorlandes Geologica Hungarica Series. *Palae* 37: 1-40.

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

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

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 HE 1977. A Lower Cretaceous Aptian diatom flora from Australia. *Nov Hed 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 Leg113 Site 693 Weddell Sea Part 2 Resting spores chrysophycean cysts an endoskeletal dinoflagellate and notes on the origin of diatoms. *Proc Ocn Drill Prog, Sci Res* 113: 403-425.

Harwood DM, Nikolaev VA 1995. Cretaceous diatoms: morphology taxonomy biostratigraphy. *In* Blome CD *et al*. convenors ed, Siliceous microfossils, Short Courses in Paleontology n° 8. The Paleontological Society Knoxville Tennessee: 81-106.

Harwood DM, Chang KH, Nikolaev VA 2004. Late Jurassic to earliest Cretaceous diatoms from Jasong Synthem Southern Korea: Evidence for a terrestrial origin. *In* Witkowski A, Radziejewska T, Wawrzyniak-Wydrowska B, Daniszewska-Kowalczyk G, Bok M eds, 18th Int Diatom Symp Abstracts. Gdansk: 81.

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 n° 13. The Paleontological Society, Knoxville Tennessee: 33-59.

Hasle GR, Syvertsen EE 1985. *Thalassiosiropsis* a new diatom genus from the fossil records. *Micropaleontology* 31: 82-91.

Homann M 1991. Die Diatomeen der Formation Alttertiä Limfjord/Dänemark. *Geol Jahr Ser* A 123: 1-285.

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

Kaczmarska I, Beaton M, Benoit AC, Medlin LK 2006. Molecular phylogeny of selected members of the Order Thalassiosirales (Bacillariophyta) and evolution of the fultoportula. *J Phycol* 42: 121-138.

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.

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

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

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

Kühn SF, Klein G, Halliger H, Hargraves P, Medlin LK 2006. A new diatom *Mediopyxis helysia* gen. nov. and sp. nov. Mediophyceae from the North Sea and the Gulf of Maine as determined from morphological and phylogenetic characteristics. *Nov Hed Beih* 130: 307-324.

Li C, Ashworth MP, Witkowski A, Dąbek P, Medlin LK, KooistraWHCF, 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*. doi:10.1371/journal.pone.0139300.

Liu H, Aris-Brosou S, Probert I, de Vargas C 2010. A time line of the environmental genetics of the haptophytes. *Mol Biol Evol* 27: 161-176.

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

Lobban CS, Ashworth MP, Arai Y, Jordan RW, Theriot EC 2011. Marine necklace-chain Fragilariaceae (Bacillariophyceae) from Guam including descriptions of *Koernerella* gen. nov. and *Perideraion* gen. nov. *Phycol Res* 59: 175-193.

Lohman KE, Andrews GW 1968. Late Eocene non-marine diatoms from the Beaver Divide area Fremont County Wyoming. *Cont Palae US Geo Sur Prof Pap* 593E: 1-26.

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

Mann DG, Marchant HJ 1989. The origin of the diatom and its life cycle. *In* Green JC, Leadbeater BSC, Diver WL, eds, The chromophyte algae: problems and perspectives. Clarendon Press, Oxford, UK: 307-323.

Magne-Simon MF 1962. L'auxosporulation chez une Tabellariacée marine, *Grammatophora marina* Lyngb. Kütz. (Diatomée). *Cah Biol Mar* 3: 79-89.

McCartney K, Witkowski J, Harwood DM 2014. New insights into skeletal morphology of the oldest known silicoflagellates: *Variramus Cornua* and *Gleserocha* gen. nov. *Rev Micropal*eontology 53: 75-91.

Medlin LK 2007. Continued ideas on the evolution of silica. *Diatom Res* 22: 217-226.

Medlin LK 2008. Molecular clocks and inferring evolutionary milestones and biogeography *In* Okada H, Mawatari SF, Suzuki N, Gautam P eds, The Microalgae Origin and Evolution of Natural Diversity, 21st. Century COE for Neo-Science of Natural History, Hokkaido: 31-42.

Medlin LK 2009. Pursuit of a natural classification of diatoms: An incorrect comparison of published data. *Eur J Phycol* 45: 155-166.

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 Bio Biopros Dev* 1: 129 doi:104172/2376-02141000129.

Medlin LK 2016a. 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 2016b. Opinion: Can coalescent models explain deep divergences in the diatoms and argue for the acceptance of paraphyletic taxa at all taxonomic hierarchies? *Nova Hedw* 102(1). DOI:10.1127/nova_hedwigia/2015/0295.

Medlin LK, Kaczmarska I 2004. Evolution of the Diatoms: Morphological and Cytological Support for the Major Clades and a Taxonomic Revision. *Phycologia* 43: 245-270.

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

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

Medlin LK, Williams DM, Sims PA 1993. The evolution of the diatoms (Bacillariophyta). I. Origin of the group and assessment of the monophyly of its major divisions. *Eur J Phycol* 28: 261-275.

Medlin LK, Gersonde R, Kooistra WHCF, Wellbrock U 1996. Evolution of the diatoms. (Bacillariophyta). III. Molecular evidence for the origin of the Thalassiosirales *Desikachary Fes Nova Hedw* 112: 221-234.

Medlin LK, Kooistra WCHF, Potter D. Saunders GW, Andersen RA 1997. Phylogenetic relationships of the 'golden algae ': Haptophytes, heterokont chromophytes and their plastids. *Plant Syst Evol* Suppl 11: 187-219.

Medlin LK, Kooistra WCHF, Schmid AM 2000. A review of the evolution of the diatoms a total approach using molecules morphology and geology. *In* The origin and early evolution of the diatoms, Wittkowski A, Sieminska J eds, *Frag Flor et Geobot* Sp vol: 13-35.

Medlin LK, Sato S, Mann DG, Kooistra WCHF 2008. Molecular evidence confirms sister relationship of *Ardissonea, Climacosphenia* and *Toxarium* within the bipolar centric diatoms (Mediophyceae Bacillariophyta). *J Phycol* 44: 1340-1348.

Moscovitz S, Ehrlich A, Soudry D 1983. Siliceous microfossils of the Upper Cretaceous Mishash Formation Central Negev Israel. *Cret Res* 4: 173-194.

Nikolaev VA, Harwood DM 1997. New process genus and family of Lower Cretaceous diatoms from Australia. *Diatom Res* 12: 24-33.

Nikolaev VA, Harwood DM 2001. Diversity and classification of centric diatoms. *In* Amilli A ed, Proc16th Int Diatom Symp. Economou-University of Athens, Athens: 127-152.

Nikolaev VA, Harwood DM 2002. Morphology Taxonomy and System of Classification of Centric Diatoms. Saint Petersburg NAUKA: 117 p, 36 plates.

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

Ogawa Y, Kawata T 1998. Preservation of Biogenic Opal-A in Earliest Cretaceous Radiolarian Claystone from the Western Pacific. *J Sed Res* 68: 435-439.

Phillippe H, Sorhanuus U, Baroin A, Perasso R, Gasse F, Adoutte A 1994. Comparison of molecular and palaeontological record in diatoms suggests a major gap in the fossil record. *J Evol Biol* 7: 247-265.

Phillippe H, 2000. Opinion: Long Branch Attraction and Protist Phylogeny. *Protist* 151:307-316.

Reháková Z 1980. Süsswasserdiatomeenflora des oberen Miozäns in der Tschechlowakei Sbornik Geologickyck ved Praha. *J Geo Sci Paleo* 23: 83-184.

Reinhold T 1937. Fossil diatoms of the Neogene of Java and their zonal distribution. *Ov Ver van Geol Mijn Gen Ned Kol Geol Ser* 12: 43-132.

Rothpletz A 1896. Ueber die Flysch Fucoiden und einige andere fossile Algen sowie überliasische Diatomeen fuhrende Hornschwamme. *Zeit Deutsch Geol Ges* 48: 854-915.

Rothpletz A 1900. Ueber einen neuen jurassichen Hornschwamme und die darin eingeschlossenen Diatomeen. *Zeit Deutsch Geol Ges* 52: 154-160.

Round FE, Crawford R M 1981. The lines of evolution of the Bacillariophyta. I. Origin. *Proc R Soc Lond B* 211: 237-260.

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/ multi-divtime. A step-by-step manual Version 1 4.

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

Scholin CA, Villac MC, Buck KR, Powers JM, Fryxell GA, Chavez F P 1994. Ribosomal DNA sequences discriminate among toxic and non-toxic *Pseudo-nitzschia* species. *Nat Toxins* 2: 152-165.

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

Schrader HJ, Gersonde R 1978. The Late Messian Mediterranean brackish to freshwater environment diatom floral evidence. *In 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. *Sys Biol* 50: 945-969.

Sims PA, Crawford RM 2007. The fossil diatom genus *Truania* Pantocsek and its relationship with *Ellerbeckia* Crawford. *Diatom Res* 22: 197-207.

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

Simonsen R 1970. Protoraphidaceae eine neue Familie der Diatomeen. *Beih Nov Hedw* 31: 377-394.

Singh RS, Stoermer EF, Kar R 2007. Earliest freshwater diatoms from the Deccan intertrappen Maastrichtian sediments of India. *Micropaleontology* 52: 545-551.

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 rhizosolenoid diatoms. *Science* 304: 584-87.

Sorhannus U 1997. The origination time of diatoms an analysis based on ribosomal RNA data. *Micropaleontology* 43: 215-218.

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

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

Stamatakis A, Ludwig T, Meier H 2005. RAxMLIII: A fast program for maximum likelihoodbased inference of large phylogenetic trees. *Bioinformation* 21: 456-463.

Syvertsen EE, Hasle GR 1982. The marine planktonic diatom *Lauderia annulata* Cleve with particular reference to the processes. *Bacillaria* 5: 243-56.

Tapia PM 1996. Campanian diatom biostratigraphy and paleoecology of Arctic Canada. MScThesis, University of Nebraska, USA.

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

Theriot E, Alverson AJ, Gutell R 2009. The limits of nuclear-encoded SSU rDNA for resolving the diatom phylogeny. *Eur J Phycol* 44: 277-290.

Theriot EC, Ashworth M, Ruck E, Nakov T, Jansen RK 2010. A preliminary multi-gene phylogeny of the diatoms. *Pl Ecol Evol* 143: 278-296.

Theriot EC, Ashworth MP, Nakov T, Ruck E, Jansen RK 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. *Nuc Acid Res* 24: 4876-4882.

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

Thorne BL, Grimaldi DA, Krishna K 2000. Early fossil history of termites. *In* Abe TD Bignell E, Higashi M eds, Termites: Evolution Sociality Symbioses Ecology. Kluwer Academic Publishers Dordrecht: 77-93.

Thorne JL, Kishino H 2002. Divergence time and evolutionary rate estimation with multilocus data. *Sys Biol* 51: 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, Caers A, De Rijk P, De Wachter R 1998. Database on the structure of small ribosomal subunit RNA. *Nuc Acid Res* 26: 179-182.

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

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

Welch J, Bromham L 2005. Molecular dating when rates vary. *Trends Ecol Evol* 20: 320-327.

Witkowski J, Harwood DM, Chin K 2011. Taxonomic composition paleoecology and biostratigraphy of Late Cretaceous diatoms from Devon Island Nunavut Canadian High Arctic. *Cret Res* 32: 2773000 doi: 10 1016/jcretres 2010 12 009.

Witt ON 1886. Über den Polierschiefer von Archangelsk Kurojedowo im Gouv Simbirsk. *Vern der Rus Kais Min Gesell St Petersburg Ser 2,* 22: 137-177.

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

Wolfe A P, Siver PA 2009. Three genera of extant thalassiosiroid diatoms from Middle Eocene lake sediments in northern Canada. *Am J Bot* 96: 487-497.

Wolfe AP, Edlund MB, Sweet AR, Creighton S 2006. A first account of organelle preservation in Eocene non-marine diatoms: observations and paleobiological implications. *Palaios* 21: 298-304.

Yang Z 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Com Appl Bio* 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. *PNAS USA* 99: 11724-1172

Received on July 13, 2015
Accepted on December 12, 2015
Associate editor: Y Desdevises

These supplementary tables are available on our website:

Supplementary Table I. – PCR primers used in Sato 2008 to produce the data set analyzed here.

Supplementary File I. – Cladistic analysis of morphological features of selected genera of the diatoms to test the position of *Praethallasiosira* and *Thalassiosiropsis* among the centric diatoms.

Supplementary File II. – Positions deleted from the alignment of the 18S SSU.

Supplementary Fig. I. – Bootstrap Tree for the ML analysis of the data set with multiple distant heterokont outgroups from Sato 2008. Third position of amino acids coded RY.

Supplementary Fig. 2. – Bootstrap Tree for the ML analysis of the data set with multiple distant heterokont outgroups from Sato 2008 without RY coding of the amino acid third position.

Supplementary Fig. 3. – Bootstrap Tree for the BI analysis of the data set with multiple distant heterokont outgroups from Sato 2008 without RY coding of the amino acids.

Supplementary Fig. 4. – Bootstrap Tree for the ML analysis of a data set with multiple bolidomonads outgroups and 157 diatoms using 3 genes. Data set from Li et al. 2015, except in this analysis the ARB alignment for the SSU and the LSU was used and this included the V4 region, which made the BT support higher than in Li et al. 2015.