Copyright

by

Matt Peter Ashworth

2013

The Dissertation Committee for Matt Peter Ashworth Certifies that this is the approved version of the following dissertation:

Rock snot in the age of transcriptomes: Using a phylogenetic framework to identify genes involved in diatom extracellular polymeric substance-secretion pathways

Committee:

Edward C. Theriot, Supervisor

Robert K. Jansen, Co-supervisor

John W. La Claire

David Cannatella

Zengjian J. Chen

Rock snot in the age of transcriptomes: Using a phylogenetic framework to identify genes involved in diatom extracellular polymeric substance-secretion pathways

by

Matt Peter Ashworth, B.S., M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2013

Dedication

This dissertation is dedicated to my parents, Pat and Bob, and my wife Alison for their support through the process and for twice now not objecting to my desire to quit my job and go back to school to study molecular evolution in algae.

Acknowledgements

I would like to thank my fellow authors of the published versions of these dissertation chapters: Elizabeth Ruck, Teofil Nakov, Dwight Romanovicz. I would also like to acknowledge the help I received from my committee—in particular John La Claire for his assistance in developing the RNA extraction protocol for diatoms and interpretation of TEM images.

I would also like to thank the following people for collection of live material or facilitating the collection of live material: Haj Allali, Andrew Alverson, Craig Aumack, Mariska Brady, Tom Frankovich, Michael Gruenstaeudl, Paul Hargraves, Sheryl Kanno, Daryl Lam, Elizabeth Ruck, Eric Salomaki, Maria Schefter and Anna Yu. I am also extremely grateful to the Texas Advanced Computing Center (TACC) for access to the RANGER and LONESTAR supercomputers and to Dhivya Arasappan, Tracey Ruhlman and Zhang Jin for assistance building and interpreting my transcriptome data. I must also acknowledge the efforts of the Marine Microbial Eukaryote Transcriptome Project for sequencing *Cyclophora tenuis*.

I am deeply indebted to my advisors, Ed Theriot and Bob Jansen, for their support, friendship and demonstrated faith in my abilities. And a profound thanks to Chris Lobban—coauthor, unofficial advisor and friend—for his enthusiasm, efforts and advice in all things tropical and taxonomic where diatoms are concerned.

This research was supported by an NSF grant to Ed Theriot (EF 0629410), the Blumberg Fellowship to Ed Theriot and the C.P. "Pete" Oliver Memorial Endowment for Genetic Research Award to Matt Ashworth.

Rock snot in the age of transcriptomes: Using a phylogenetic framework to identify genes involved in diatom extracellular polymeric substance-secretion pathways

Matt Peter Ashworth, Ph.D.

The University of Texas at Austin, 2013

Supervisors: Edward C. Theriot & Robert K. Jansen

We are coming to understand that the ecological importance of diatoms is not limited to primary productivity, as many diatoms produce extracellular polymeric substances (EPS), which are vital components in algal and bacterial "biofilms." While great effort has been made to chemically identify the types of molecules and polymers used to create and modify diatom EPS there is still much about the process we do not know. Rather than studying this process chemically, we have elected to search for the genes involved in EPS production and secretion. We assembled transcriptomes from three EPS-producing diatoms (*Cyclophora tenuis, Lucanicum concatenatum, Thalassionema frauenfeldii*) and two diatoms which do not (*Astrosyne radiata, Thalassionema* sp. 'BlueH20'). In an attempt to limit the differences to EPS-related transcripts, the taxa were selected in a phylogenetic framework (which is also discussed in this dissertation), where EPS-producing taxa were closely-related to taxa which did not produce EPS (*A. radiata, C. tenuis, L. concatenatum* as one set, *T. frauenfeldii* and *T.* sp. 'BlueH20' as the other). The resulting pool of transcripts sorted for contigs which appeared in the EPS-producing taxa but not their closely-related non EPS-producing counterparts, and those contigs were then compared to two annotated diatom genomes and sorted by function, looking specifically for genes related to secretion, polysaccharide assembly or modification and carbohydrate metabolism. In the *Thalassionema* clade, 41 contigs with the aforementioned annotations were found, while 22 such contigs were found in the *Cyclophora/Lucanicum/Astrosyne* clade. These putative EPS-related markers are identified in this dissertation for further study on their function and evolution across diatoms.

Table of Contents

List of Tables

List of Figures

Chapter 1: A revision of the genus *Cyclophora* and description of *Astrosyne* gen. nov. (Bacillariophyta), two genera with the pyrenoids contained within pseudosepta.

INTRODUCTION

It is clear that there is much undescribed diversity amongst tropical marine periphyton, particularly diatoms. In particular, the coral reefs of Guam harbor an interesting assemblage of diatoms that seem to be maintained by damselfishes known colloquially as "farmer fish." While some of the larger diatoms have been characterized from this assemblage (Navarro & Lobban 2009, Lobban & Jordan 2010), recent work has been focused on describing new taxa amongst the smaller diatoms found in the farmer fish turfs, particularly the diversity within some genera that are rarely reported yet common in these habitats, including *Climaconeis* Grunow (Lobban *et al.* 2010) and *Bleakeleya* Round (Lobban *et al.* 2011). Another such genus is *Cyclophora* Castracane.

Castracane (1878) described a tabellarioid diatom, *Cyclophora tenuis,* noting in particular the alternating apex-to-apex pattern of colony formation and the distinct "annulus or central locule" found on only one valve of each diatom. This area lacks pores and is bordered internally by a pseudoseptum, which has either straight walls or forms a cup with an opening at the base [see Round et al. (1990), p. 439 figs i vs. k]. *Cyclophora* is equally striking when alive: Round *et al.* (1990) noted that it is unique among the araphid pennate diatoms in the radial arrangement of plastids arrayed from the center of the valve (first observed by Mereschkowsky 1901). Castracane (1866) further described and illustrated this species in his report on the *Challenger* diatoms.

The early history of *Cyclophora* contains some discussion which indicates that the genus was very poorly understood. Most remarkable is Van Heurck's (1896: 238) assertion that it was "not a true genus, but only … a craticular state of a Naviculoid

form." Van Heurck (1896) nevertheless illustrated an unnamed new species, differing from *C. tenuis* in being constricted in the middle and in lacking any central nodule (whereas he supposed *C. tenuis* to have a "central nodule indistinctly surrounded by an elliptic annulus"). However, Van Heurck (1896: 237) indicates [by (Castr.!) in the distribution list] that he had examined Castracane's specimen. He also relates that "Count Castracane has of late years withdrawn the name of *Cyclophora*, and given to … *C. tenuis* that of *Diatoma hyalina forma Cyclophora,*" an opinion with which he disagreed on the grounds of "the presence of a true raphe." Peragallo & Peragallo (1897–1908) had abundant material from tanks at the marine station at Banyuls (Mediterranean coast of France) and while setting aside any doubts about it being naviculoid nevertheless retained it in the raphid diatoms (Achnanthaceae, because of the heterovalvy). Hustedt (1931) recognized that it was araphid—as had Castracane—and put it back in Fragillarioideae.

To add to the confusion, several forms and a new species were added during these early years when the structure was seriously misunderstood. Grunow in Van Heurck (1881) and Mereschkowsky (1901) named varieties of *C. tenuis*. Castracane (1886) himself commented on *C. tenuis* var. *tropica* Grunow, noting that if there were terminal nodules, as described by Grunow, it would be a different genus. Castracane (1886) also added an unnamed variety distinguished apparently by some inflation, and Mereschkowsky (1901) added *C. tenuis* var. *abbreviata.* Østrup (1904) described a second species, *C. siamensis*, based on his observations of "terminal nodi" at the apices of the valve and the lack of "a perforation in the central part" which Peragallo & Peragallo (1897–1908) describe as a character of *C. tenuis*. This perforation apparently refers to the opening at the base of the pseudoseptum when it is cup-like. Østrup claimed his species was similar to the undescribed *Cyclophora* "n. sp." drawn by Van Heurck (1896) specifically in regard to the terminal nodi. Whether these nodi refer to the

rimoportulae (labiate processes) proximal to the apical pore fields or refer to some other character is unclear, though it should be noted that neither Østrup nor the Peragallos noted any other character at the apices that might be interpreted as rimoportulae.

In recent years *C. tenuis* has been well illustrated by Navarro (1982) and Round *et al.* (1990), and while the latter mention it as rarely collected, it has been reported from widespread localities and is now clearly described. Round & Crawford (in Round *et al.* 1990) reclassified *Cyclophora* to its own family, Cyclophoraceae and order, Cyclophorales.

There are several diatom genera that have internal crosswalls as part of their frustule ultrastructure. Crosswalls oriented parallel to the valve, usually emerging from girdle elements, are called "septa," such as in *Grammatophora* Ehrenberg or *Striatella unipunctata* (Lyngbye) Agardh (Muller 1886, von Stosch 1975). In contrast, thin crosswalls emerging from the valve perpendicularly to the valve face are termed "pseudosepta," such as those found in *Biddulphia biddulphiana* (J.E. Smith) Boyer, *Anaulus* Ehrenberg and *Eunotogramma* Weisse (Anonymous 1975). In *Cyclophora* the pseudoseptum is one more-or-less circular structure although it sometimes appears as a pair of curved walls.

While sampling in Guam, we found a diatom with a *Cyclophora*-like pseudoseptum, but that with regard to frustule morphology, appeared to be centric: circular valves with pores arrayed radially around what appeared to be a central annulus. In the course of studying this species and addressing the question of what is in the pseudoseptum, we also found additional *Cyclophora* species in Guam and Florida that led to emending the genus diagnosis, with description of three additional species. The circular species is described as a new genus and species, *Astrosyne radiata*, and its

relationship to *Cyclophora* assessed on the basis of valve morphology, gene sequences, and localization of the pyrenoids in the pseudosepta.

METHODS AND MATERIALS

Benthic samples were collected by hand using SCUBA or snorkel from subtidal environments or collected by plankton net with a 20 μm mesh size. Diatoms were isolated from this material for culture and the remaining material was preserved in a formaldehyde-seawater mix. Samples were coded according to the systems in use at the respective laboratories: in Lobban's collection the samples are identified by an island code (GU = Guam, TK = Chuuk), a site number, an alphabetic date code, and a sample number (e.g., GU6D-1). In Theriot's laboratory samples are assigned sequential numbers (e.g., ECT 3854). Slides and SEM stubs are retained in the authors' collections except for the deposited type materials.

Isolates from Guam were grown in marine f/2 medium (Guillard 1975) in a Percival model I-36LL incubation chamber at 27^{o} C, *Rhabdonema arcuatum* was grown at 14^{o} C in the incubator, and all other isolates were grown at room temperature. Fluorescent lighting was used in the incubation chambers in a 12:12 h light:dark photoperiod. Wild and cultured material was acid-cleaned with hydrogen peroxide and nitric acid and washed until pH neutral. Permanent mounts for light microscopy were made using Naphrax mounting medium.

Electron Microscopy Preparation

At UT, cleaned wild material was filtered and dried onto Millipore 1.2 μm Isopore membrane filters while cultured material was dried onto 12 mm round coverslips. Material was then coated with iridium using a Cressington 208 Bench Top Sputter Coater

and observed with a Zeiss SUPRA 40 VP scanning electron microscope. At UOG, specimens were dried onto Millipore filters, coated with Au-Pd and observed with a Phenom G2 Pro desktop SEM.

For TEM imaging, cultured cells were fixed first in a mixture of 2.5% glutaraldehyde and 2% formaldehyde prepared from EM grade fixatives (Electron Microscopy Sciences, www.emdiasum.com) in f/2 medium buffered with 0.1M cacodylate pH 7.4. This was followed by fixation with 2% reduced osmium (2% osmium tetroxide and 2% potassium ferrocyanide in 0.1M cacodylate buffer). After 1h in 2% uranyl acetate the cells were encased in agar, dehydrated with EtOH, and embedded in epoxy resin. Thin sections were cut with a diamond knife, picked up on Formvar-coated slot grids, and imaged at 80kV with an FEI Tecnai Spirit BioTwin.

DNA Extraction, Sequencing and Phylogenetic Analysis

Aliquots of cultured isolates were pelleted in a Sorvall RC-5B refrigerated superspeed centrifuge for 20 minutes at 8,000 rpm. DNA was extracted from the diatom pellets using a QIAGEN DNeasy Plant Mini Kit (QIAGEN Sciences, Maryland) after 45 seconds of cell disruption using 1.0 mm glass beads in a Mini-Beadbeater (Biospec Products, Inc). PCR amplification and sequencing of small-subunit nuclear rRNA, and the chloroplast-encoded *rbc*L and *psb*C markers, follows the primers and protocols of Theriot *et al.* (2010).

Phylogenetic analysis was conducted using the entire 136 taxon, 3 gene dataset in Theriot et al. (2010) using *Bolidomonas pacifica* as the outgroup; Genbank accession numbers for the new sequences presented in this paper are listed in Table 1. Sequence data were partitioned by gene and by codon position and run in RAxML ver. 7.0.4 (Stamatakis 2006) for Maximum Likelihood analysis using the 6-substitution GTR +

Gamma model. The Likelihood analysis was run for 25 generations with 500 bootstrap replicates each and the generation with the best log likelihood score is presented here. Bootstrap replicates were compiled from the tree with the best log-likelihood score and the next six-best trees, which differed by 0.0001 log-likelihood units. Two Bayesian Inference runs were also conducted using MRBAYES, using a 6-substitution model with partitioning by gene and by codon, for 20 million generations. All but the final million generations were discarded as "burn-in" and the remaining trees were compared in AWTY using the "compare" function (Wilgenbusch *et al.* 2004) for evidence the two runs had sampled the same distribution of posterior probabilities. The final 10,000 trees (sampled every 1,000 generations) were used to generate the final phylogenetic tree diagram.

Trees constraining *Cyclophora* and *Astrosyne* to a single clade (*C. tenuis, C. castracanei, C. tabellariformis* + *A. radiata*, excluding *Florella pascuensis*) and *Astrosyne* to the centric diatoms (*A. radiata* + radial centrics + bipolar centrics) were constructed with Mesquite ver. 2.74 (Maddison & Maddison 2010) and used to generate new likelihood analyses with RAxML. SH-Tests were conducted using RAxML between the best unconstrained tree and the best tree for each of the constraint analyses. The SH-Test was chosen for ease of analysis and because only pairwise-comparison of trees were conducted, as discussed in Theriot *et al.* (2011).

Taxon	Locality and Voucher Code	GenBank Accession (SSU, *rbc*L, *psb*C)
Astrosyne radiata Ashworth & Lobban	Gab Gab Reef, Guam (ECT3697)	JN975238, JN975252, JN975267
Biddulphia biddulphiana (J.E. Smith) Boyer	Kona, Hawaii, HI (ClayHI)	JN975239, JN975253, JN975268
Cyclophora tenuis Castracane	Kahana Beach Park, Oahu, HI (ECT3854)	JN975240, JN975254, JN975269
Cyclophora tenuis Castracane	Long Beach, CA (ECT3838)	JN975241, JN975255, JN975270
Cyclophora castracanei Ashworth & Lobban	Gab Gab Reef, Guam (GU44AB-6)	JN975242, JN975256, JN975271
Cyclophora tabellariformis Ashworth & Lobban	Carrabelle, FL (ECT3892)	JN975243, JN975257, JN975272
Dimeregramma sp. Ralfs	Port Aransas, TX (28II10-1A)	JN975244, JN975258, JN975273
Eunotogramma sp. Greville	Bald Head Island, NC (ECT3886)	JN975245, JN975259, JN975274
Florella pascuensis Navarro	Agat, Guam (ECT3756)	JN975246, JN975260, JN975275
Hyalosira interrupta (Ehrenberg) Navarro	Gab Gab Reef, Guam (ECT3700)	JN975247, JN975261, JN975276
Licmophora flucticulata Lobban, Schefter & Ruck	Cocos Wall, Guam (GU56-A)	HQ997923, JN975262, JN975277
Licmophora remulus Grunow	Outhouse Beach, Guam (GU52-O)	JN975248, JN975263, N/A
Opephora sp Petit	Ward Island, TX (ECT3831)	JN975249, JN975264, JN975278
Psammoneis japonica Sato, Kooistra & Medlin	Outhouse Beach, Guam (GU52-O)	JN975250, JN975265, JN975279
Rhabdonema arcuatum (Lyngbye; Agardh) Kützing	Pebble Beach, CA (ECT3898)	JN975251, JN975266, JN975280

Table 1. Taxa and genes added to the Theriot et al 2011 data set for this paper. Locality indicates source of isolate for DNA extraction. Locality and voucher codes in parentheses beginning with "GU" refer to Lobban lab collection numbers, while all others refer to Theriot lab collection numbers.

RESULTS

Taxonomic revisions

Based on the results of the phylogenetic analysis and microscopy detailed below, we amend the diagnosis of *Cyclophora* to reflect the previously-undescribed diversity of valve morphology within this genus. The following diagnosis is derived from Castracane's descriptions (1878, p. 10 and 1886, p. 58).

***Cyclophora* Castracane, emend. Ashworth & Lobban**

Frustules unequal or equal, one or both valves with more or less circular pseudoseptum below the center, containing pyrenoids. Frustules tabulate-rectangular and joined in chains by alternating gelatinous pads.

The various species are described below and the characters summarized in Table 2.

***Cyclophora tenuis* Castracane** Figs 1, 5–6, 14–18

REFERENCE ILLUSTRATIONS: Castracane 1886, pl. 25, fig. 3; Peragallo & Peragallo 1897–1908, pl. 1, figs 27–32; Navarro 1982, figs 9–17; Round *et al.* 1990, pp. 438–439; Navarro & Lobban 2009, figs 59, 60; Lobban & Jordan 2010, fig. 5b.

RECORDS: *Guam*: GU44X-2; GU44U-1B; GU6D-1; GU44AK-5; *Chuuk, Federated States of Micronesia*: (TK28); *Hawaii* (ECT 3854); *California* (ECT 3838).

DESCRIPTION: Cells attached in zigzag chains at the poles, heterovalvar, multiple ribbon-like plastids of various lengths arising from pseudoseptum (Figure 1). Valves linear, often tapering toward the apices and commonly constricted around the pseudoseptum (Figures 6, 14–15); bluntly rounded ends; length 44–76 µm, width 7–11 µm. Striae 30–32 in 10 µm, parallel, absent within pseudoseptum (usually including attached part of mantle). On the pseudoseptate valve, the mantle is elongated around the pseudoseptum (Figure 17). Sternum distinct, rimoportula at each end of the sternum

(indicated by arrows on Figure 14). Pseudoseptum elliptical to circular or sometimes appearing as two separate, curved crosswalls in wider valves (Figure 15). Large apical field of slits in a U-shaped pattern on the mantle, sometimes partially on valve face, slits at ends of U oblique/radial (Figure 18). Girdle bands perforated by two rows of slits (Figure 16, 17).

COMMENTS: There is a range of valve shape and size and the early literature is unreliable. Proper resolution of the true character of *C. tenuis* will require examination of the various type materials, which is beyond the scope of the present work. For now, we regard Navarro's (1982) LM and SEM images as representing the modern interpretation of *C. tenuis*. The description in Round *et al.* (1990) is in agreement, though they do not specifically identify their specimens as *C. tenuis*. The most useful characters, and the most stable in culture are (1) the apical pore field, (2) tapering apices. On this basis, a large number of specimens from Guam that would otherwise have been classified as *C. tenuis* must be included in *C. tabellariformis,* described below.

***Cyclophora castracanei* Ashworth & Lobban, sp. nov.** Figs 2, 7, 19-23

Frustules isovalvar, valves elongate with central sternum, both with internal pseudoseptum at valve center. Four plastids in pairs arising from one pseudoseptum (perhaps both), together appearing butterfly-like in girdle view. Valve length 35–47 μm, breadth 3–6 μm. Pseudoseptum circular to subcircular, with decreasing diameter from internal valve surface to opening. Striae 30–31 in 10 μm. Pores absent from pseudoseptum and valve within pseudoseptum. Rimoportulae proximal near pore fields, external tube absent. Straight apical fields of parallel slit-like pores on the mantle, shorter slits towards the distal ends of the field.

HOLOTYPE: GU44AA-5 (slide 445), deposited in the California Academy of Sciences, CAS accession #627406, slide 223020.

TYPE LOCALITY: Gab Gab Reef, Apra Harbor, Guam. 13.4427^{o}N, 144.6428^{o}W

PARATYPES: GU44AB-8 (slide "C.cast 9XII2009"), deposited in the California Academy of Sciences, CAS accession #627407, slide 223021; GU44AF-2 (material preserved in glutaraldehyde), CAS accession # 627411 deposited in the California Academy of Sciences.

ETYMOLOGY: Named in honor of Conte Abate Francesco Castracane Degli Antelminelli, who described the genus *Cyclophora*.

RECORDS: *Guam*: GU44AA-5, GU44AB-8, GU44AK-5; *Majuro, Marshall Is.*: M2.

DESCRIPTION: Living cells epiphytic on macroalgae, joined in zigzag chains and to the substratum by mucilage pads. Plastids in a "butterfly" configuration (Figure 2) arising from one of the pseudosepta (perhaps both). Valve morphology similar to *C. tenuis*, with the most obvious difference in the isovalvar nature of the frustule and the capitate apices, which are more consistently swollen in comparison to the rest of the valve than in *C. tenuis* which more often has parallel or tapered margins at the valve apex. Areolae in uniseriate rows except within the boundaries of the pseudosepta, where no pores are present (Figures 19, 21). As in *C. tenuis*, valve mantle is extended where the pseudosepta are located (Figure 21). Both valves bear pseudosepta (Figure 7), which vary in shape from simple, seemingly-discrete convex walls across the valve to a fused, funnel-shaped cup (Figure 20). Single rimoportulae are situated at each apex of the frustule, just beside the sternum on both valves (Figure 23). Apical pore field slits arranged in straight row across valve mantle (Figure 22), with length of slits decreasing at the distal ends of the pore field.

COMMENTS: Differing from *C. tenuis* in the presence of pseudosepta on both valves, the shape of the apical pore fields, and the number of plastids. This species can co-occur with *C. tenuis* and cannot be reliably separated from it in acid cleaned material in either LM or SEM unless both valves are present, since it is similar in size and shape to the narrower range of *C. tenuis*. The apical pore field of *C. castracanei* is unlike the U-shaped arrangement in *C. tenuis*.

***Cyclophora tabellariformis* Ashworth & Lobban, sp. nov.** Figs 3, 8-9, 24-30

Cells typically joined in short chains at apices. Elongate plastids (4–6) radiating from pseudoseptum on one valve. Valves more or less linear, slightly constricted below apices, 34–114 µm long, 3–5µm wide, striae 34–40 in 10 µm. Pseudoseptum circular with decreasing diameter from internal valve surface to opening. Apical pore field straight with numerous pervalvar slits. Distinguished from *C. tenuis* by the shape of the apical pore field and shape of apices, and from *C. castracanei* by having only one pseudoseptate valve.

HOLOTYPE: ECT3892 (slide "CyclophoraXL 20X2010"), deposited in the California Academy of Sciences, CAS accession #627408, slide 223022. CAS accession includes material preserved in glutaraldehyde.

PARATYPE: GU52J-3 (slide 254), deposited in the California Academy of Sciences, CAS accession #627414, slide 223026.

TYPE LOCALITY: Carrabelle River inlet, Carrabelle, FL, 29.86^{o}N, 84.681^{o}W

ETYMOLOGY: Named for the outline of the valve, which resembles valves of the genus *Tabellaria* Ehrenberg ex Kützing.

RECORDS: *Florida*: ECT3892; *Guam*: GU52J-3; GU52P-4; GU52P-8; GU52P-9; GU44I-4; *Palau*: PW2009-36.

DESCRIPTION: Another diatom typically found attached in zig-zag chains, though the cells from Florida typically form much shorter chains than *C. tenuis* or *C. castracanei*. The elongate plastids are arrayed in an "X"-shape in girdle view (Figure 3), much like *C. castracanei*. Frustules are elongate and heterovalvar—as *C. tenuis*, only one valve bears the internal pseudoseptum (Figures 8-9, 24-26, 30). Dimensions of the valves in Guam samples were smaller than the isolate from Florida (Table 2), with somewhat fewer striae. Valves are narrow, but widen at the capitate apices and at the center of the valve. Rimoportula number, location and morphology like *C. tenuis* and *C. castracanei* (Figure 27). Apical pore fields are like those of *C. castracanei*: a row of elongate slits down the mantle, decreasing in size at the distal ends of the field (Figure 28).

COMMENTS: Original collection of this taxon from Florida was in a plankton sample in a muddy coastal inlet, though the presence of apical pads, attached chain-forming habit in culture and the epiphytic habit of the other species in the genus suggest that the collected specimens were tychoplanktonic, having been dislodged from their benthic habit. This sample caused us to review Guam samples that appeared to be *C. tenuis* and to separate out those matching the new species from those matching illustrations of *C. tenuis* in Navarro (1982) (see above). The two species occurred together in several Guam samples. The valve apex illustrated by Navarro & Lobban (2009), fig. 61 appears to belong to *C. tabellariformis*.

***Cyclophora minor* Ashworth & Lobban, sp. nov.** Figs 4, 31-43

Valves elliptic-lanceolate, length 6.9–17.1 µm, width 2.0–3.0 µm, one or both valves with pseudoseptum. Striae 40–46 in 10 µm, parallel throughout, absent within the

area enclosed by the pseudoseptum; on pseudoseptate valves hyaline area reaches to valve border. Valve mantle deeper at the pseudoseptum. One rimoportulae near each pole, internally small. Pseudoseptum pyramidal, attached to mantle as well as valve face, truncated on abvalvar side with small aperture. Small apical pore field of 4–5 slits, without rim. Copulae several to many, with 1–2 rows of areolae.

HOLOTYPE: GU44I-4 (slide 235), deposited in the California Academy of Sciences, CAS accession #627405, slide 223019.

TYPE LOCALITY: GabGab reef, Apra Harbor Guam, 13° 26' 34" N, 144° 38' 34" E, associated with filamentous turf algae in farmer-fish territory.

ETYMOLOGY: Named for the diminutive size of the taxon.

RECORDS: *Guam*: GU44I-4, GU3D; *Palau*: PW2009-36.

DESCRIPTION: Frustules usually seen in girdle view (Figures 31–36), not observed alive but presumed to adhere in zigzag chains. Valves elliptic-lanceolate (Figures 4, 37, 40, 43), one or both with a pseudoseptum. Broken frustules with only one valve were inconclusive, but we observed complete valves with both one and two pseudosepta (Figures 31–33 vs. 34–36), in contrast to the consistent numbers (1 or 2) in other species. The fine striae and sternum are not resolvable in LM. In SEM we observed valves with continuous striation, presumably lacking a pseudoseptum (Figure 43), whereas those with a pseudoseptum had a broad hyaline saddle across the middle of the valve, reaching the valve borders (Figures 39, 42). One valve was observed with a small central circular hyaline area (Figure 44). Rimoportulae are small and positioned at the apices, as with the other *Cyclophora* taxa (Figures 37-49). Circular pseudoseptum (Figure 37) and copulae with two rows of pores (Figures 41-42) also similar to other *Cyclophora* taxa.

COMMENTS: Observations based on numerous acid cleaned valves in one Guam sample; a single frustule seen in the Palau sample. In the same Guam sample *C. tabellariformis* (Figure 30) was abundant and we observed some as short as 18 μm, but the apices in this species were always blunt with numerous apical slits, in contrast to those of *C. minor*. While in LM this species may appear close to *Hustedtiella*, it is distinguished in SEM by the presence of pore fields rather than ocelli at the apices, by the shape of the pseudoseptum, and by the presence of two rimoportulae.

***Astrosyne* Ashworth & Lobban, gen. nov.**

Valves circular with an internal, circular pseudoseptum at valve center. Pores radiate from pseudoseptum to valve margin, absent within pseudoseptum. Numerous rimoportulae. Numerous elongate plastids radiate from pseudosepta.

ETYMOLOGY: From the Greek "aster" + substantival suffix –osyne, indicating a special feature, for the star-like arrangement of chloroplasts radiating from the center of the frustule in valve view.

TYPE SPECIES: *Astrosyne radiata* Ashworth & Lobban

With only cursory observation, there is nothing that overtly suggests that this genus is not a radial centric. The valve outline is circular and the simple pores radiate from the central hyaline area. The chloroplasts are elongate, and in valve view radiate out from the central axis of the cell in a "starburst" arrangement for which the genus is named (Figure 10). This central hyaline area is bordered on the inside of the valve by a circular pseudoseptum which decreases in diameter at the opening (Figures 12-13, 45). The rimoportulae are similar in morphology to those in *Cyclophora* and are arrayed on the outer edge of the valve and on the valve mantle (Figures 44-47).

***Astrosyne radiata* Ashworth & Lobban sp. nov.** Figs 10-13, 44-48

Valve diameter 30–63 µm. Diameter of pseudoseptum 5–6.7 µm, decreasing from internal valve surface to opening. Striae 28-30 in 10 µm. Pores absent from pseudoseptum and valve within pseudoseptum. Rimoportulae scattered in the outer third of the valve.

HOLOTYPE: GU44Z-15 (slide 453), deposited in the California Academy of Sciences, CAS accession #627409, slide 223023.

TYPE LOCALITY: Gab Gab Reef, Apra Harbor, Guam. 13.4427^{o}N, 144.6428^{o}W

PARATYPES: ECT3697 (slide "starry 22VII2008"), deposited in the California Academy of Sciences, CAS accession #627410, slide 223024. Material preserved in glutaraldehyde deposited in the California Academy of Sciences, CAS accession # 627412.

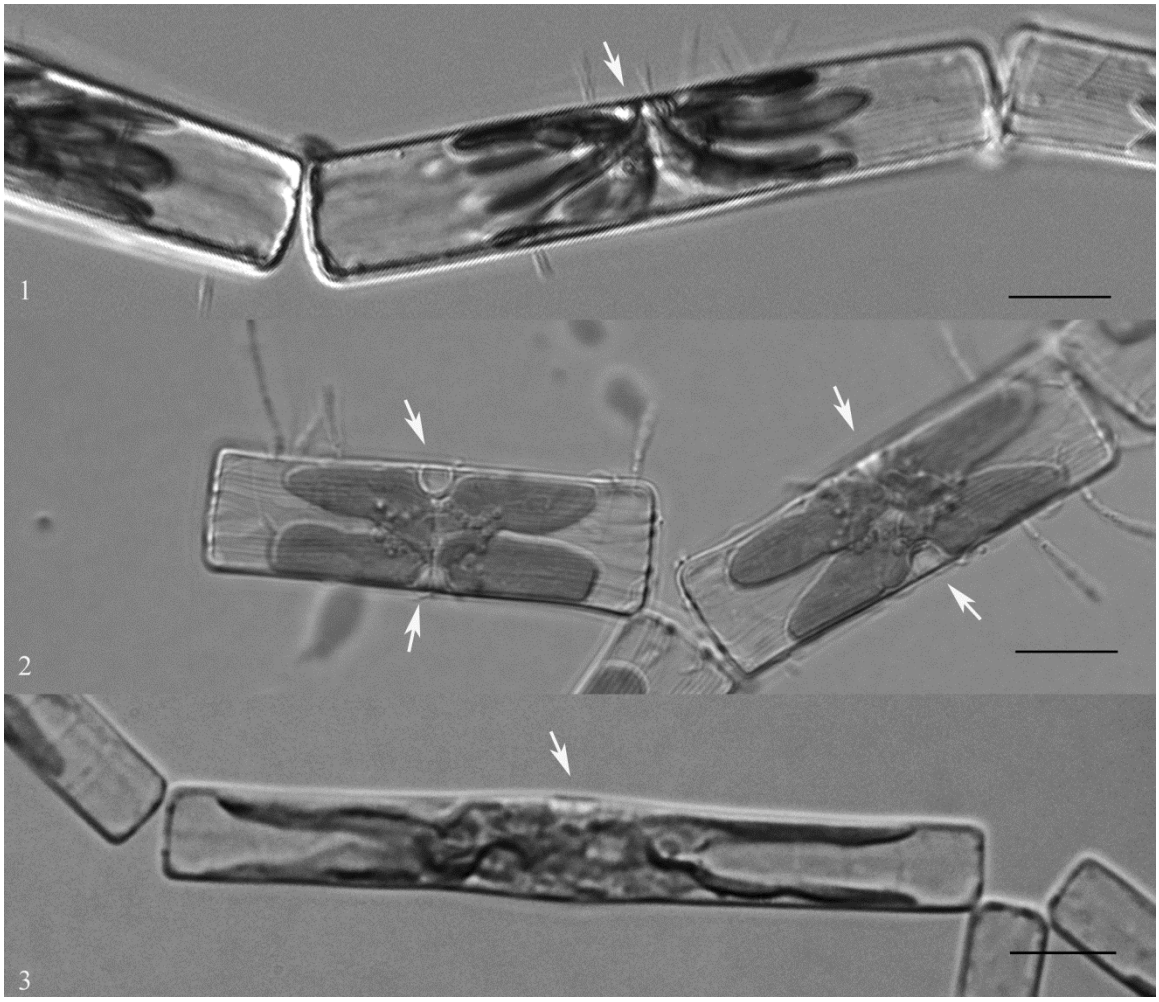
ETYMOLOGY: Named for the circular shape of the frustule, unusual in araphid pennate diatoms.

RECORDS: *Guam*: GU44Y-14, GU44Z-15, GU44AK-3; *Chuuk, Federated States of Micronesia*: TK4, TK28; *Majuro, Marshall Islands*: M2.

DESCRIPTION: Cells solitary in culture and field material. No mucilage pads have been observed holding cells together into chains or attaching cells to a substrate, though collections in Guam were associated with damselfish-tended macroalgal turfs. Rarely observed, even in these turf samples. Frustules are isovalvar, with the circular pseudosepta found on both valves of this species (Figure 12-13).

Character	*Cyclophora tenuis* [a]	*Cyclophora castracanei*	*Cyclophora tabellariformis*	*Cyclophora minor*
Valve shape	Linear	Linear	Linear	Lanceolate
Plastids	Numerous	4	4-6	Unknown
Pseudosepta	One valve	Both valves	One valve	One or both valves
Valve length (μm)	44–76 [66–70]	35–47	80–114 (Florida) 34–69 (Guam)	6.9–17.1
Max. valve width (μm)	7–11 [12]	3–6	4–5 (Florida) 3–5 (Guam)	2.0–3.0
Striae in 10 μm	30–32 [28–30]	30–31	40 (Florida) 34–37 (Guam)	40–46
Apical pore field	U-shaped with slit-like pores	Straight, tapering, with slit-like pores	Straight, with slit-like pores	Straight, with only 4-5 slit-like pores

Table 2. Comparison of characters of the *Cyclophora* species described in this paper.
[a] [Data from Navarro 1982 in brackets]

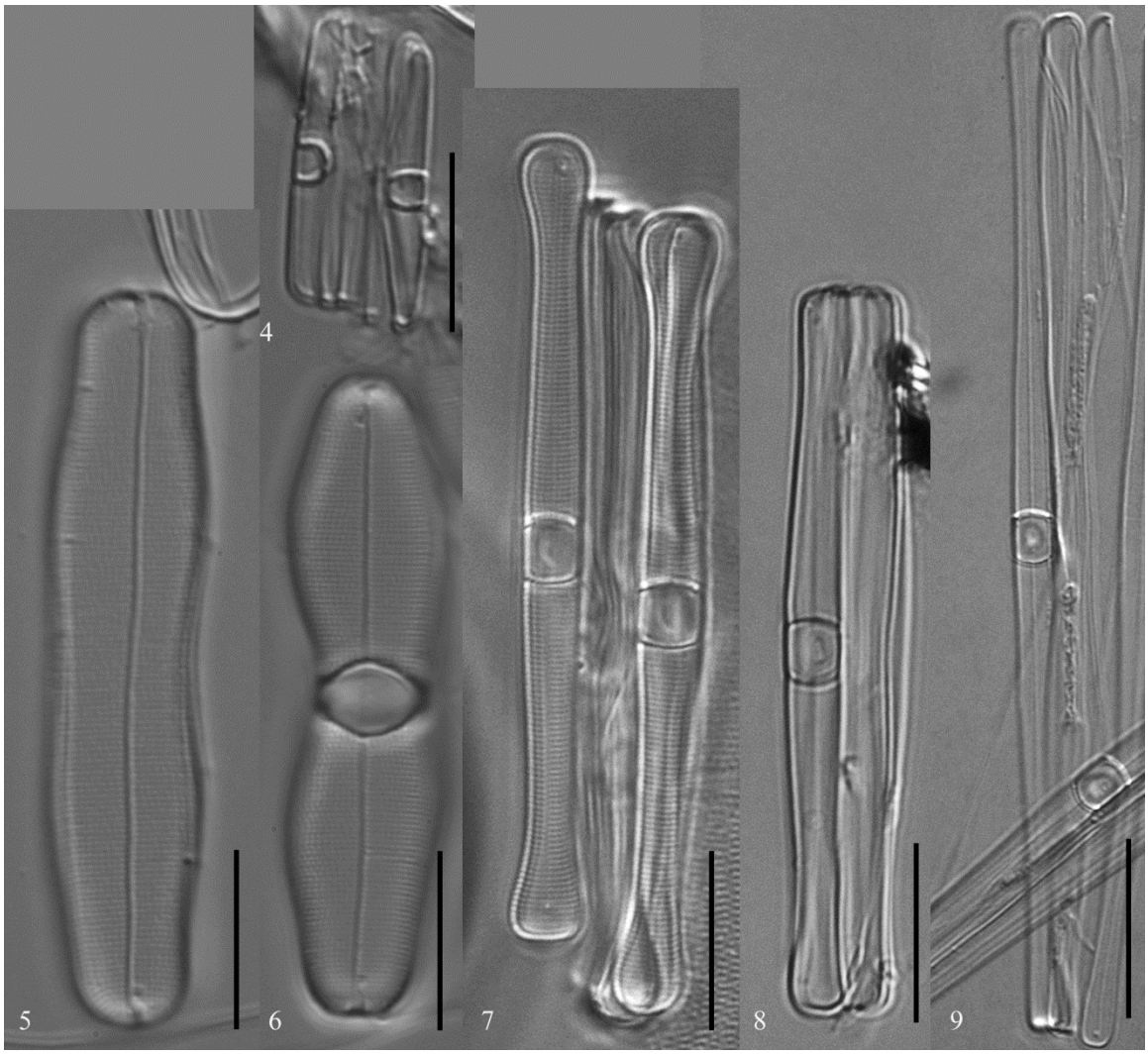


Figures 1-3: Light micrographs of *Cyclophora* spp. live chains. Scale bars = 10 µm. Arrows indicate pseudosepta

Figure 1: Micrograph of live *Cyclophora tenuis* (Guam, GU44Y) in girdle view.

Figure 2: Micrograph of live *Cyclophora castracanei* (Guam, GU44Y-13) in girdle view.

Figure 3: Micrograph of live *Cyclophora tabellariformis* (FL, ECT3892) in girdle view.



Figures 4-9: Light micrographs of acid-cleaned *Cyclophora* spp. valves. Scale bars = 10 μm.

Figure 4: Micrographs of wild *Cyclophora minor* frustules in valve view (Guam, GU44I-4).

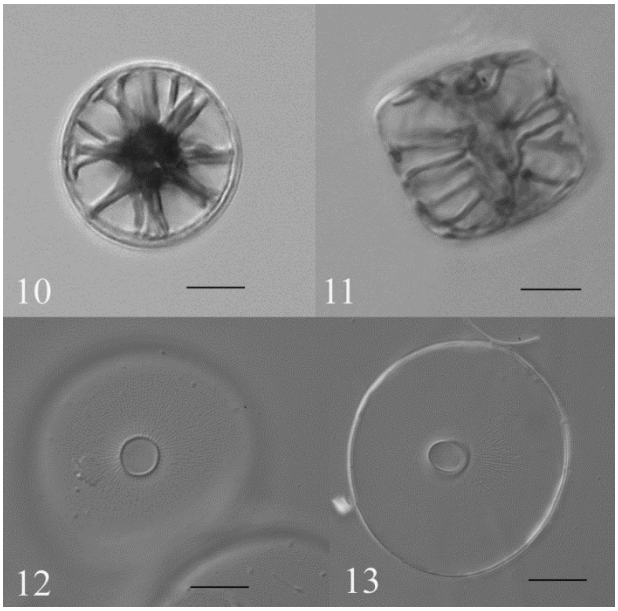
Figure 5: Micrograph of the non-pseudoseptate valve of cultured *C. tenuis* (Guam, ECT3723) cell.

Figure 6: Micrograph of the pseudoseptate valve of an acid-cleaned cultured *C. tenuis* (Guam, ECT3723) cell.

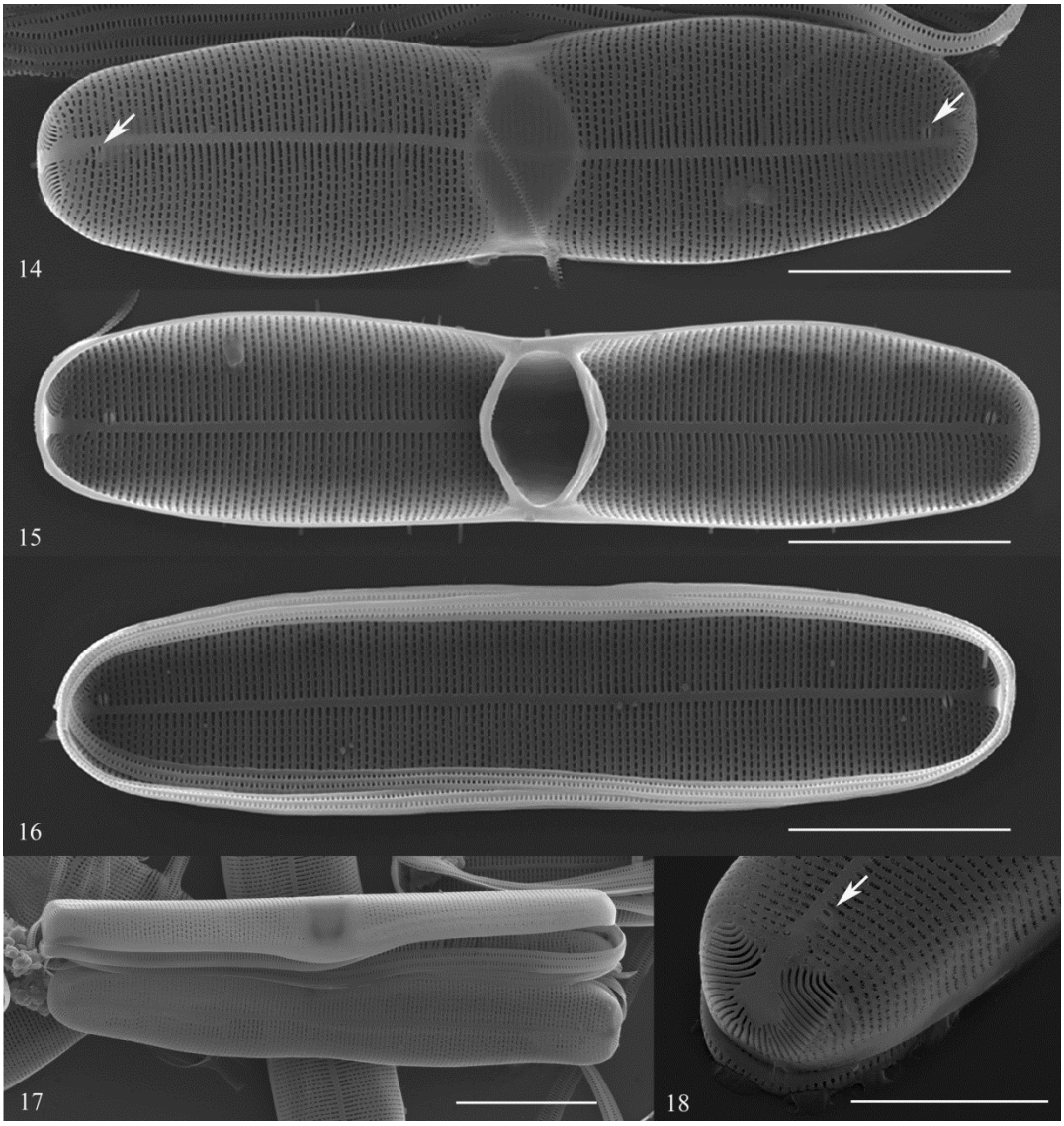
Figure 7: Micrograph of the frustule of wild *C. castracanei* (Guam, GU44AA-5) in valve view. Note pseudoseptae on both valves.

Figure 8: Micrograph of wild *C. tabellariformis* (Guam, GU52J-3) in valve view showing both pseudoseptate and non-pseudoseptate valves.

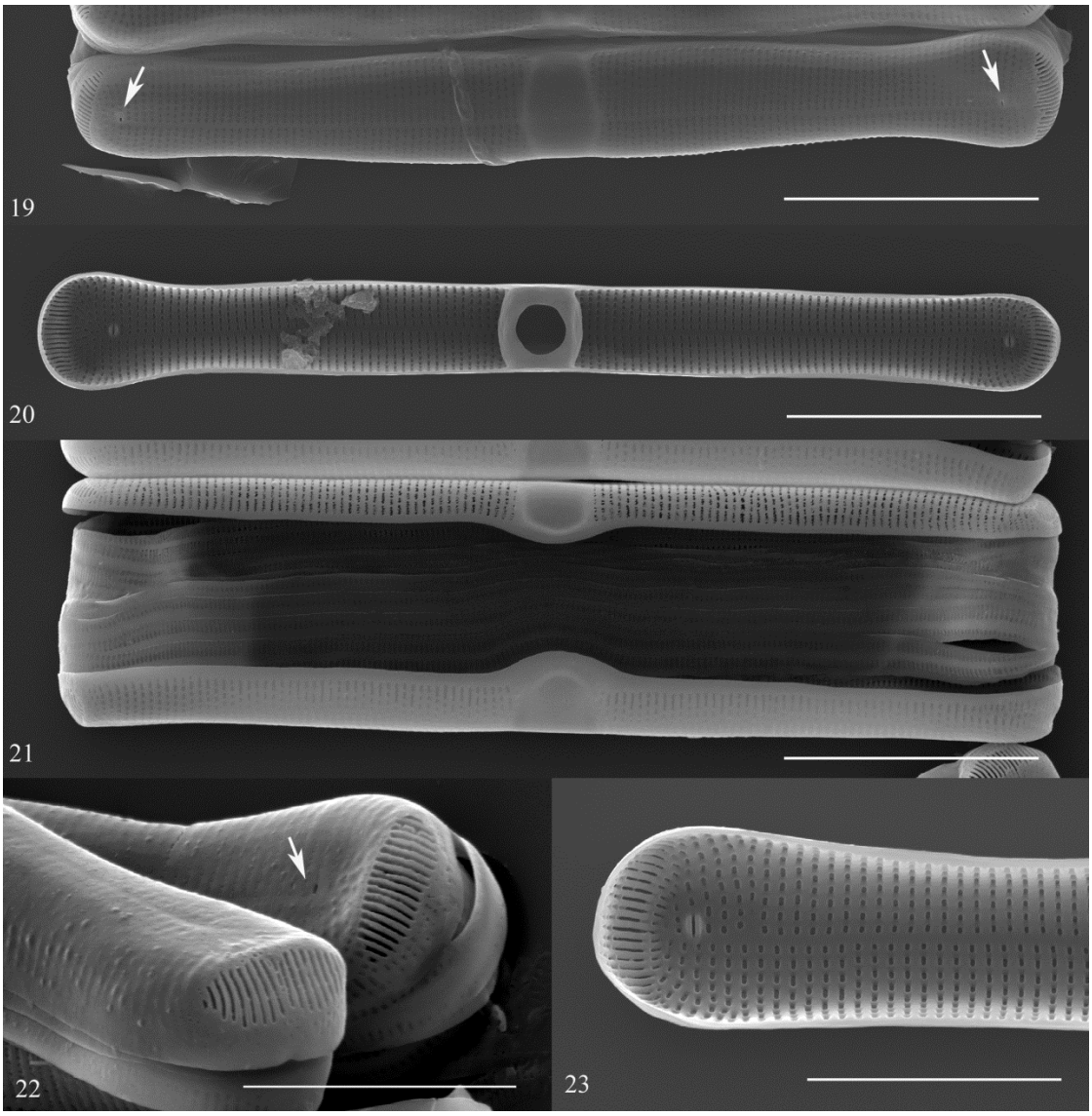
Figure 9: Micrograph of cultured *C. tabellariformis* (FL, ECT3892) in valve view showing both pseudoseptate and non-pseudoseptate valves.



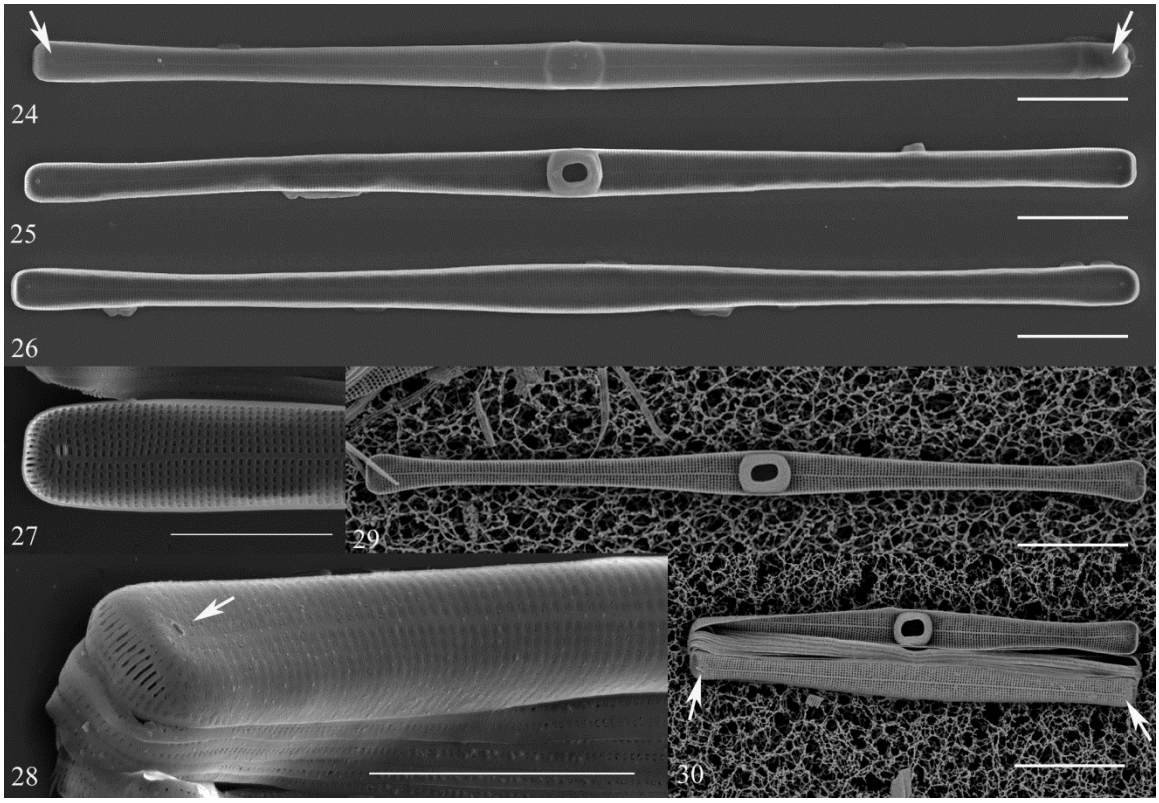
Figures 10-13: Light micrographs of acid-cleaned and live cultured *Astrosyne radiata* cells (Guam, ECT3697). Scale bars = 10 μm.
Figure 10: Micrograph of live cell in valve view. Note the distinctive pattern of plastids radiating from the center of the valve.
Figure 11: Micrograph of live cell in girdle view.
Figures 12-13: Micrograph of acid-cleaned cell in valve view at two focal planes to demonstrate the conical shape of the pseudoseptum.



Figures 14-18: Scanning electron micrographs of acid-cleaned, cultured *Cyclophora tenuis* (Guam, ECT3723) cells. External openings to rimoportulae are noted with arrows. Scale bars: Figs. 15-18 = 10 µm; Fig. 19 = 5 µm.
Figure 14-16: Micrograph of the exterior of the pseudoseptate valve, interior of the pseudoseptate valve and interior of the non-pseudoseptate valve, respectively. Note the associated girdle elements fallen into valve in the lower half of Figure 17.
Figure 17: Micrograph of the frustule in girdle view, showing the elongation of the mantle at the pseudoseptae and apices.
Figure 18: Detail of the apical pore field, showing the characteristic horseshoe-shaped arrangements of slits.



Figures 19-23: Scanning electron micrographs of acid-cleaned, cultured *Cyclophora castracanei* (Guam, GU44AB-6) cells. External openings to rimoportulae are noted with arrows. Scale bars: Figs. 20-22 = 10 µm; Figs. 23-24 = 5 µm.
Figure 19: Micrograph of the valve exterior.
Figure 20: Micrograph of valve interior.
Figure 21: Micrograph of the frustule in girdle view to show the mantle elongation at the pseudoseptae.
Figure 22: Detail of the valve apex, showing the linear arrangement of slits in the apical pore field.
Figure 23: Detail of the valve apex interior in valve view, showing the internal structure of the rimoportula.



Figures 24-30: Scanning electron micrographs of acid-cleaned, cultured and wild *Cyclophora tabellariformis* cells. External openings to rimoportulae are noted with arrows. Scale bars: Figs. 25-27, 30-31 = 10 µm; Figs. 28-29 = 5 µm.

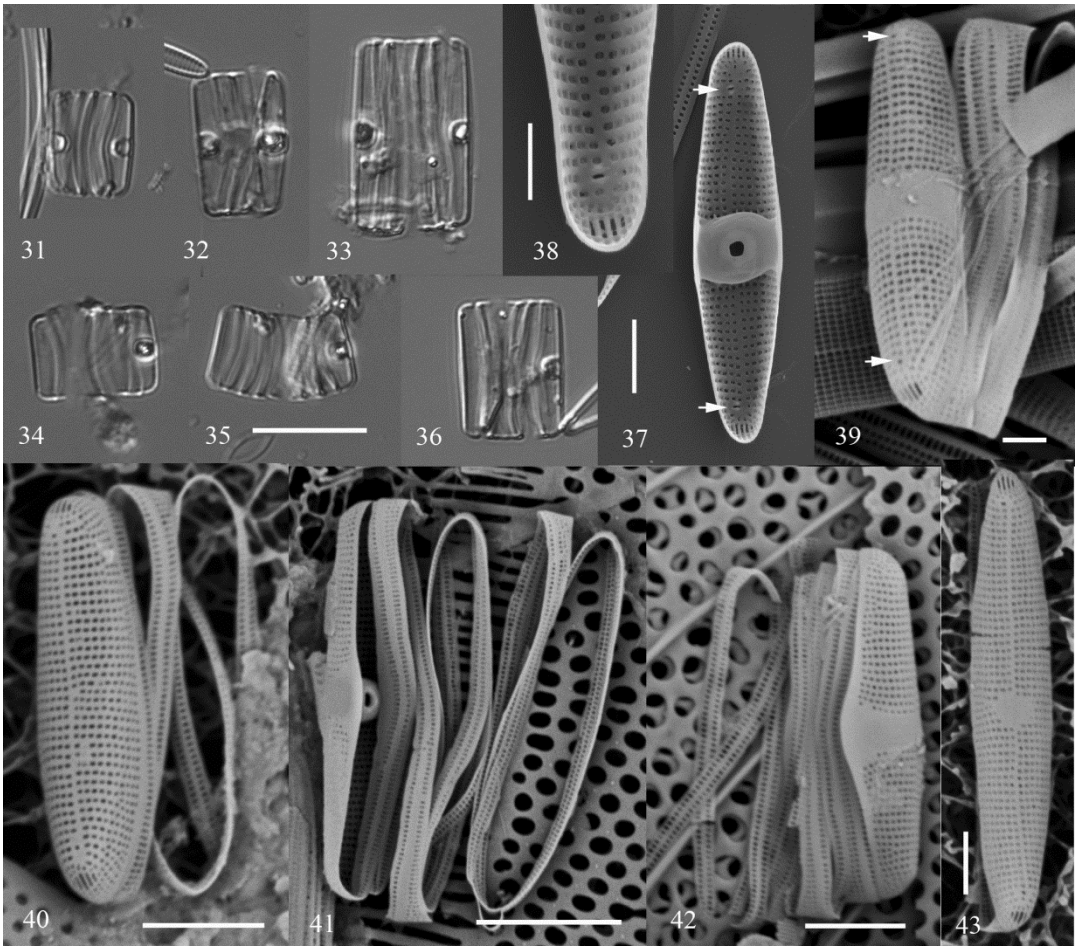
Figure 24-26: Micrographs of cultured cells (FL, ECT3892) in valve view showing the pseudoseptate valve exterior, interior and non-pseudoseptate valve interior, respectively.

Figure 27: Detail of the cultured (FL, ECT3892) valve apex interior in valve view, showing the internal structure of the rimoportula.

Figure 28: Detail of the cultured (FL, ECT3892) valve apex, showing the linear arrangements of slits in the apical pore field.

Figure 29: Micrograph of wild cell (Palau, PW2009-36) in valve view showing interior of the pseudoseptate valve.

Figure 30: Micrograph of wild frustule (Guam, GU44I-4) in valve view showing interior of the pseudoseptate valve and exterior of non-pseudoseptate valve.



Figures 31–43: Light and scanning electron micrographs of *Cyclophora minor* wild specimens (GU44I-4). Figs 31–36 LM images, Figs 37–43 SEM images. Scale bars: Figs 31–36 = 10 μm; Figs 37, 39 = 1 μm; Figs. 40, 42 = 3 μm; Fig. 41 = 5 μm; Figs 38, 43 = 2 μm.

Figures 31–36: Cells across size range, with one or two pseudosepta, showing valve view in Fig. 32.

Figure 37: Valve interior showing pseudoseptum and rimoportulae (arrows).

Figure 38: Detail of valve interior showing rimoportula and apical slits.

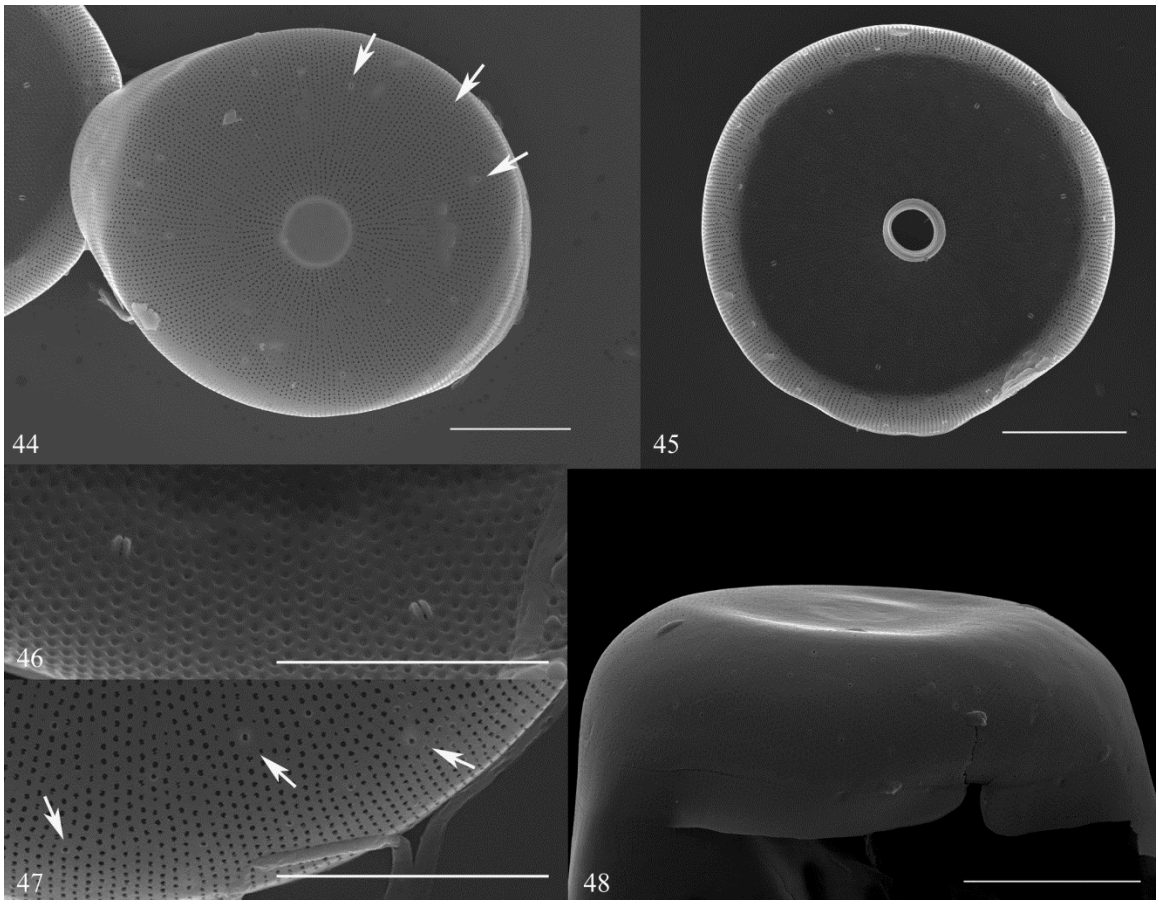
Figure 39: Exterior view of pseudoseptate valve showing rimportulae (arrows), apical slit field, and copulae.

Figure 40: Non-pseudoseptate valve showing apical slits and copulae.

Figure 41: Pseudoseptate valve, oblique view showing pseudoseptum and copulae.

Figure 42: External girdle view of pseudoseptate valve with copulae.

Figure 43: External valve view of valve with small hyaline area; presence/absence of pseudoseptum unknown.

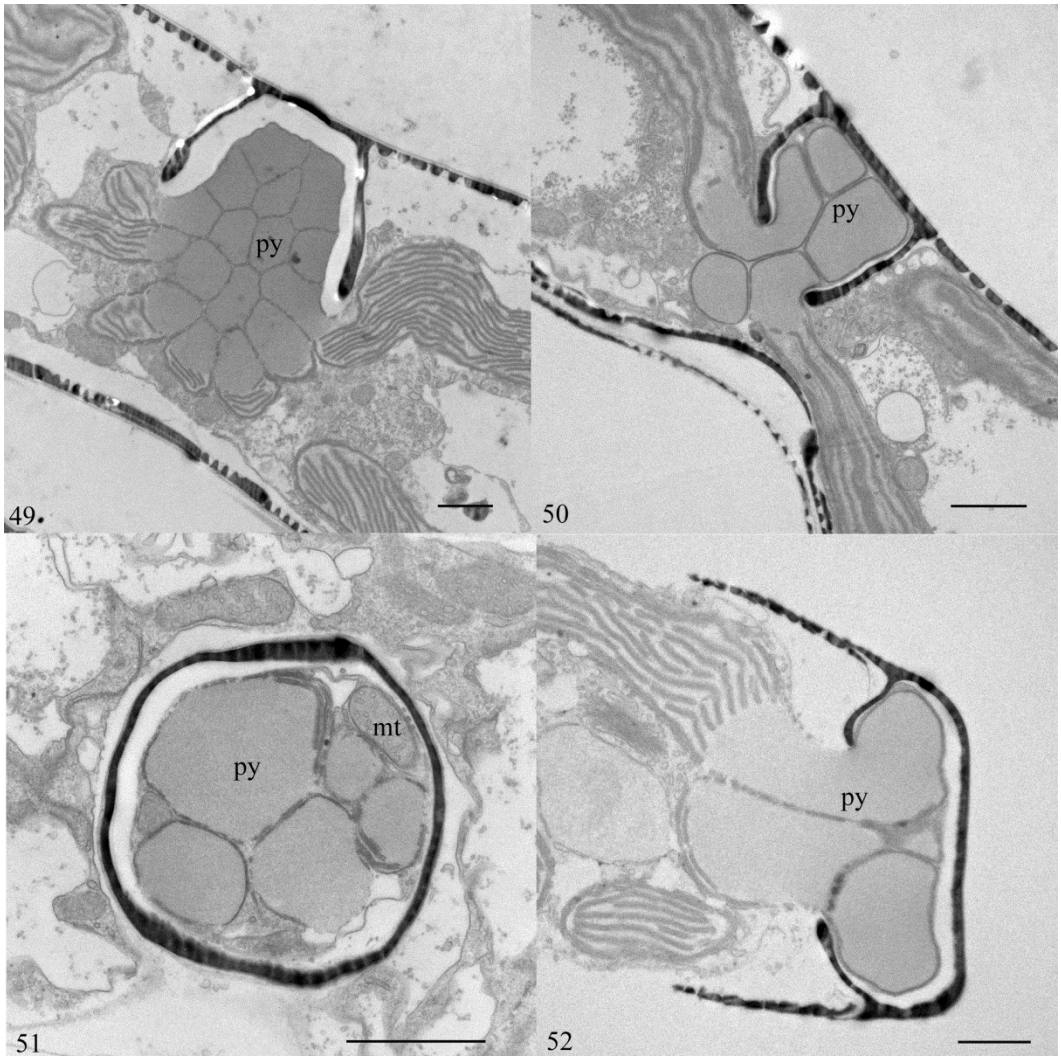


Figures 44-48: Scanning electron micrographs of acid-cleaned cultured *Astrosyne radiata* (Guam, ECT3697) cells. External openings to rimoportulae are noted with arrows. Scale bars: Figs. 44-45, 48 = 10 µm; Figs 46-47 = 5 µm.
Figure 44: Micrograph of exterior valve. Note the radial arrangement of pores around the pseudoseptum at center.
Figure 45: Micrograph of valve interior, demonstrating the narrowing of the pseudoseptum at the opening. Rimoportulae can be seen scattered about the outer third of the valve.
Figure 46: Micrograph of valve interior, showing the internal structure of the rimoportulae.
Figure 47: Micrograph of valve exterior. Pores are occluded by single rota and rimoportulae are surrounded by slightly-elevated rim.
Figure 48: Micrograph of frustule in girdle view showing valve and open valvocopula.

Cellular Ultrastructure

Sectioned cells were observed by TEM, focusing primarily on the chloroplasts and the pseudosepta, which appear strongly associated with the chloroplasts in both *Cyclophora* and *Astrosyne* (Figures 49–54). In both genera, the elongate chloroplasts appear to have a polarity, with one end devoid of thylakoids and with a pyrenoid, which is common in the chloroplasts of many diatom taxa (Schmid 2001, Bedoshvili et al. 2009). The pseudoseptum in *A. radiata* is crowded with the pyrenoids of these chloroplasts (Figures 53-54), with the occasional mitochondrion observed amongst the pyrenoids.

These same patterns of asymmetrical chloroplast ultrastructure and the localization of pyrenoids are found across *Cyclophora* taxa as well. Sectioned cells of *C. castracanei*, *C. tenuis*, and *C. tabellariformis* all showed pyrenoids within or just outside the pseudosepta (Figures 49–52).



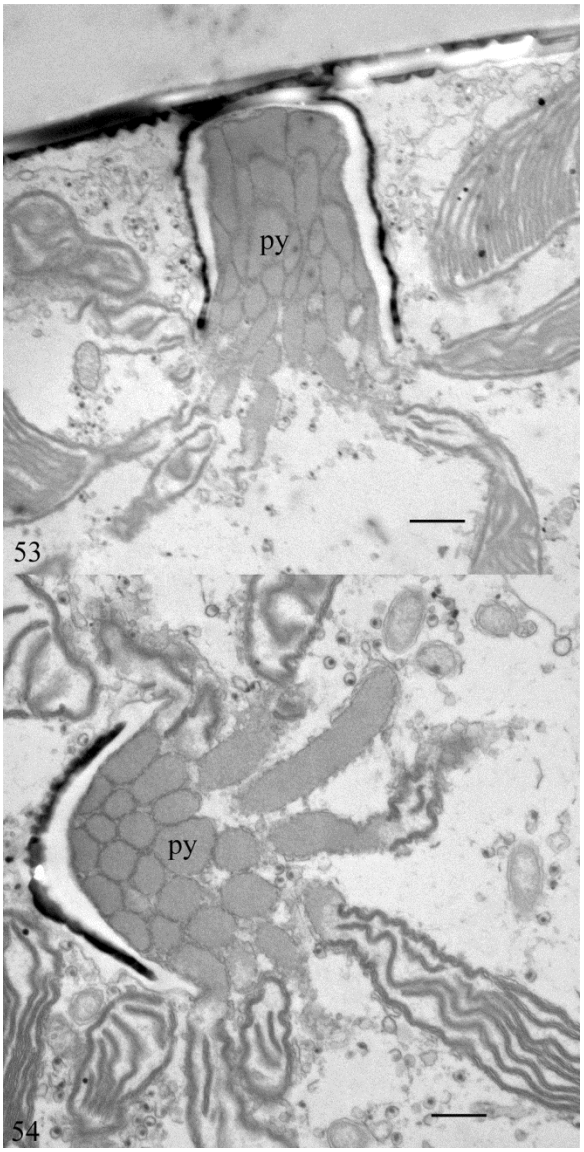
Figures 49-52: Transmission electron micrographs of cultured *Cyclophora* spp. Scale bars = 1 μm. "py" = pyrenoid, "mt" = mitochondrion.

Figure 49: Micrograph of *C. tenuis* (Guam, ECT3723) apical section. Pyrenoids can be seen clustered within and just outside of the pseudoseptum.

Figure 50: Micrograph of *C. castracanei* (Guam, GU44AB-6) apical section. Pyrenoids can be seen clustered within and just outside of the pseudoseptum.

Figure 51: Micrograph of *C. castracanei* (Guam, GU44AB-6), valvar section. Pyrenoids and mitochondria are observed within the pseudoseptum.

Figure 52: Micrograph of *C. tabellariformis* (FL, ECT3892) median transapical section. Pyrenoids can be seen clustered within the pseudoseptum.



Figures 53-54: Transmission electron micrographs of cultured *Astrosyne radiata* (Guam, ECT3697). Scale bars = 1 µm. "py" = pyrenoid. As seen in *Cyclophora spp*, the elongate plastids are oriented so the pyrenoids ("py") are clustered within the pseudoseptum.

Figure 53: Micrograph of cell in median valvar section.

Figure 54: Micrograph of cell in diagonal valvar section.

Phylogenetic analysis

Results of the combined data Maximum Likelihood and Bayesian Inference (Figure 55) analysis confirm the monophyly of the genus *Cyclophora,* as *C. tenuis*, C. *castracanei* and *C. tabellariformis* form a well-supported clade (bootstrap support=93%, posterior probability=1.0). The species with linear pore fields, *C. castracanei* and *C. tabellariformis*, formed a clade (bs=100%, pp=1.0), which was sister to the *C. tenuis* clade consisting of strains from California, Hawaii and Guam (bs=62%, pp=0.98). Sister to the *Cyclophora* clade is another clade consisting of *Astrosyne radiata* and *Florella pascuensis* Navarro (bs=95%, pp=1.0). In combined likelihood analyses run without *F. pascuensis*, *A. radiata* is sister to the *Cyclophora* clade (bs=77%), *F. pascuensis* is within the *C. tenuis* clade (bs=73%) without *A. radiata* included in the combined analysis. The idea that combined analyses can increase resolution in the diatom phylogeny was discussed in Theriot *et al.* (2011). The strong shift in bootstrap support values when the combined SSU/plastid DNA versus SSU and plastid DNA alone with this data set seems to support this idea. The *Cyclophora/Astrosyne/Florella* clade appears to be fairly well-derived amongst the araphid pennate diatoms, but there is little support via likelihood or Bayesian analyses for a sister lineage within the other araphid clades.

In Maximum Likelihood analyses run on ribosomal small subunit alone, *A. radiata* is sister to the *Cyclophora* clade (bs=68%), while *F. pascuensis* is sister to the *A. radiata*/*Cyclophora* clade (bs=34%). Using likelihood on the plastid-encoded data alone, *A. radiata* and *F. pascuensis* formed a clade (bs=56%) which was sister to the *C. tenuis* isolates from Hawaii and Guam (bs=34%). The California *C. tenuis* isolate was sister to this clade (bs=38%) and this assemblage was sister to the *C. tabellariformis*/*C. castracanei* clade (bs=50%).

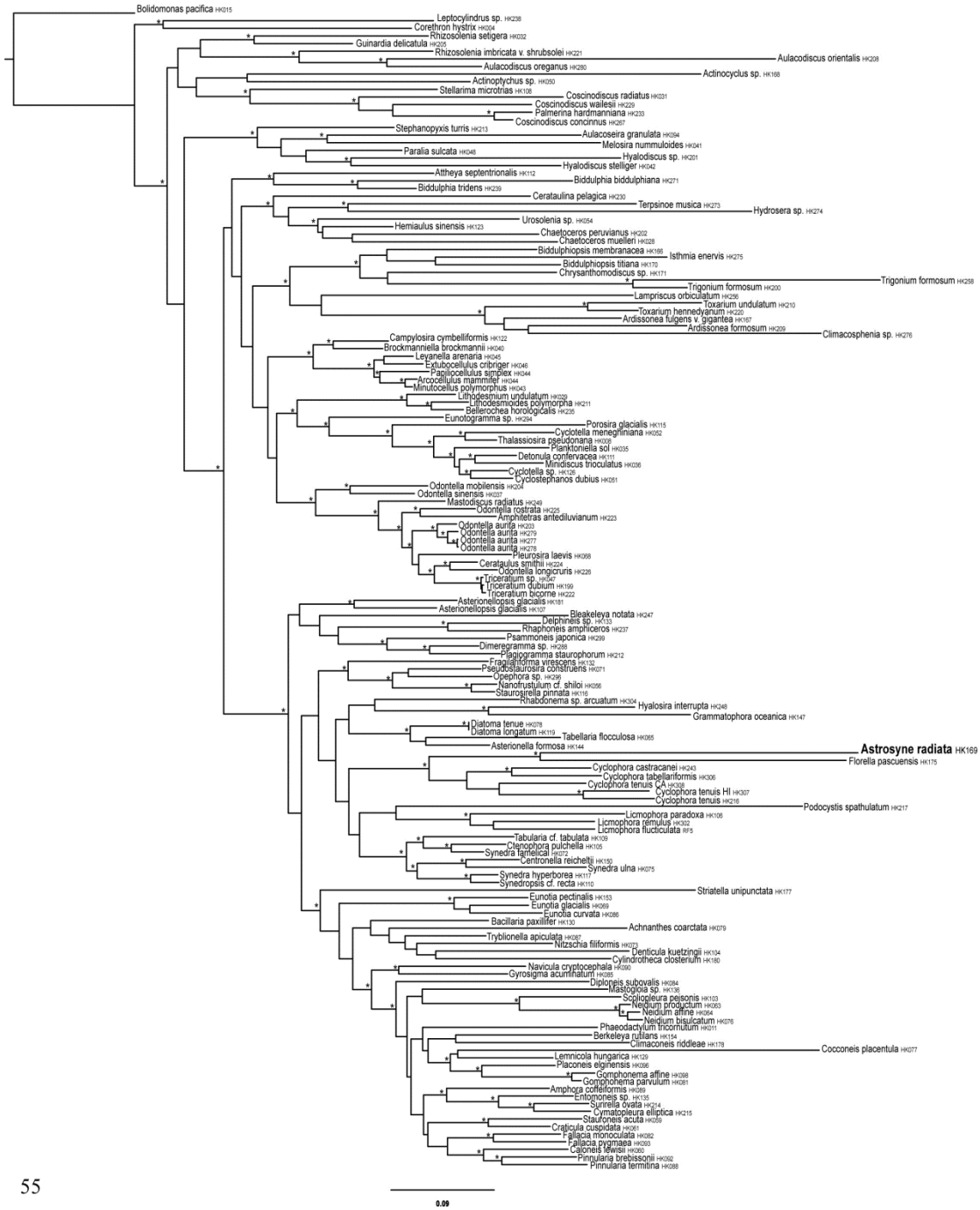


Figure 55: Phylogram depicting the best tree from the 152-taxon likelihood and Bayesian phylogenetic analyses. Likelihood tree was run with 500 bootstrap replicates. Bayesian results are from the final 10,000 trees after 20 million generations. Asterisks (*) denote clades with ≥75% bootstrap support and ≥0.95 posterior probability.

Constraint trees testing the position of *A. radiata* on the phylogeny were built and compared to the tree in Figure 55. Both monophyly of *Cyclophora* and *Astrosyne* (*Cyclophora spp. + A. radiata* to the exclusion of *F. pacsuensis*) and inclusion of *A. radiata* in the centric diatoms were significantly worse than the "best" tree (Table 3).

Phylogenetic Tree	RAxML likelihood score	Significantly different? (5%)	Significantly different? (2%)	Significantly different? (1%)
Best tree (Figure 55)	-121429.921			
Astrosyne + *Cyclophora* spp = monophyletic	-122987.409	Yes	Yes	Yes
Astrosyne = centric diatom	-123032.489	Yes	Yes	Yes

Table 3. Results of SH-tests conducted with RAxML between the "best tree" in Figure 55 and two constraint trees.

DISCUSSION

Taxonomic implications

We have adopted Navarro's (1982) illustrations of *C. tenuis* as a working model to sort out specimens in our collections. However, the taxonomy of *C. tenuis* clearly requires additional work, both to establish the characters of Castracane's materials and other early taxa of *Cyclophora*, and to further explore the question of the molecular variation apparent in our widespread collections of *C. tenuis*. *C. tenuis* has been recognized in recent years on the basis of one dominating character, the presence of a pseudoseptum on one valve, but the new species, with pseudosepta on one or both valves, show that isolated pseudoseptate valves cannot be confidently assigned to *C. tenuis*, and

that species with only one pseudoseptum cannot be distinguished in girdle view. The ideal view has both valves adjacent and in valve view (e.g., Figures 4, 7-9) but such views are not common.

Ultrastructural implications

Citing studies of chloroplast ultrastructure and pyrenoid location being utilized to classify other lineages of algae, Schmid (2001) set out to test if this was viable in diatoms as well. Though the diversity of diatoms sampled was small, she did note that diatoms with the peripheral (or "extra-thylakoidal") pyrenoids, such as those in *Cyclophora* and *Astrosyne*, had so far only been observed amongst pennate lineages (both araphid and some biraphid pennates).

No diatom, however, has been observed to localize the pyrenoids within a silica chamber as seen in *Cyclophora* and *Astrosyne*. There is evidence to suggest the pyrenoids in diatoms are primarily composed of RuBisCO (McKay & Gibbs 1991, Okada 1992), so it might be logical to conclude that this clustering has something to do with the functionality of RuBisCO. While it is interesting to note that these pyrenoids are clustered in a pseudoseptum devoid of pores typically found in the diatom valve, experimentation would be required before making any suppositions regarding the functional aspects of this clustering.

Clustering of pyrenoids has been observed with TEM in other organisms as well. The euglenoid *Eutreptiella eupharyngea* sometimes arranges its plastids in stellate aggregations, with the pyrenoids clustered together and bordered by paramylon plates (Walne *et al*. 1986). No functional hypothesis was presented, however. The stellate arrangement of individual chloroplasts around an assemblage of terminal pyrenoids is

well-documented in the phaeophycean taxon *Bachelotia antillarum* (Magne 1976) and is one of the defining characters for the phaeophycean genera *Asteronema* and *Asterocladon* (summarized in Ouriques & Bouzon 2000). The pyrenoid bundles in these phaeophycean taxa generally do not appear to be bound by any cytological structure, such as the paramylon in *Eutreptiella* or silica in *Astrosyne* and *Cyclophora*.

Cyclophora and Astrosyne in the Diatom Phylogeny

While other pseudoseptate diatoms were included in the analysis—*Biddulphia tridens* (Ehrenberg) Ehrenberg, *B. biddulphiana*, and *Eunotogramma* sp.—there was no expectation that those taxa would form a clade with *Cyclophora* or *Astrosyne*. The bipolar centric *Biddulphia* has numerous discoid plastids that are not condensed around the pseudosepta, and there are no hyaline areas on the valves. While the plastids are not as well characterized in *Anaulus* or *Eunotogramma*, single, lobed plastids without distinctive inclusion into the area between pseudosepta were observed in *A. creticus* (Drebes & Schulz 1981) and the *Eunotogramma* included in this study (Figure 56). In addition, all of the aforementioned genera feature a central rimoportula, rather than the radially-arranged rimoportulae of *Astrosyne* or the apical rimoportulae of *Cyclophora*.

It should be noted that *A. radiata* in valve view strongly resembles Greville's sketches of *Porodiscus* (1863). However, the central area in *Porodiscus* spp. is not a pseudoseptum, but a deep depression in the valve face. With recent SEM studies confirming the presence of pores on the depressed central area, loculate areolae and a markedly different girdle structure than *Astrosyne* (Strelnikova *et al*. 2004), it is unlikely the two genera would be related.

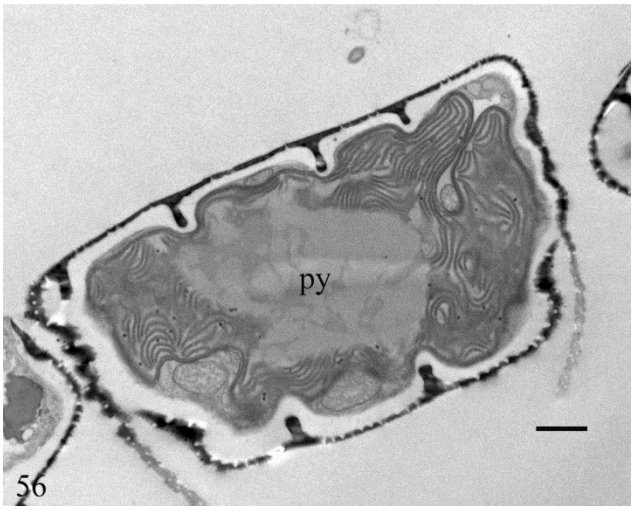


Figure 56: Transmission electron micrograph of *Eunotogramma* sp (NC, ECT3886), apical section. Note the multiple pseudoseptae and the large, central pyrenoid ("py") within the chloroplast. Scale bar = 1 μm

The presence of *Florella pascuensis* in the *Cyclophora/Astrosyne* clade, sister to *Astrosyne radiata* is a curious issue. This taxon is one of two species of *Florella* Navarro in the monotypic Florellaceae (Navarro 1996, 2002), and nothing about the morphology of this genus suggests a close relationship to *Astrosyne* or *Cyclophora*. Navarro (1996) placed the Florellaceae in the Striatellales (Round), but noted several key differences between *Florella portoricensis* Navarro and the other Striatellales, such as the lack of a strongly-delineated apical pore field or internally-distinct rimoportulae in *Florella*. However, both *Florella* species do possess septa, such as those found in *Grammatophora spp*., *Hyalosira spp*. and *Striatella spp*—all genera placed into Striatellales by Round *et al*. (1990). It should be noted that taxa from those genera were all included in the phylogenetic analysis presented in this paper, and while *Grammatophora oceanica* Ehrenberg and *Hyalosira interrupta* (Ehrenberg) Navarro formed a clade (bs=100%, pp = 1.0), this clade was not particularly close to *Florella pascuensis* or *Striatella unipunctata* (Figure 55).

In fact, *Florella* spp. lack pseudosepta found in *Cyclophora* and *Astrosyne* as well. Chloroplast arrangement in *F. portoricensis* and *F. pascuensis* is described as stellate chloroplasts scattered across the cell (Navarro 1996, 2002), and pyrenoid position is unknown. As both *A. radiata* and *F. pascuensis* occupy very long branches on the phylogenetic tree, and the taxon sampling of known araphid diatoms included in the analysis is sparse, it is entirely possible this association is a long-branch attraction artifact. The fact that *A. radiata* remains sister to the *Cyclophora* clade when *F. pascuensis* is removed, but *F. pascuensis* moves into the *C. tenuis* clade when *A. radiata* is removed from the analysis seems to support this.

The addition of taxa from the araphid pennate genus *Hustedtiella* Simonsen to future analyses is likely to be very informative with regard to the phylogenetic position of *Cyclophora* and *Astrosyne*. Both *Hustedtiella baltica* Simonsen and *H. sinuata* Crawford possess pairs of pseudosepta around the central area of both valves, with no pores on the valve or valve margin within the area of the pseudosepta. Unlike *Cyclophora* and *Astrosyne*, however, the pseudosepta have not been observed joining into a single, circular structure (Crawford *et al*. 1993). Although the plastids of *H. baltica* are more lobose than the elongated plastids of *Astrosyne* and *Cyclophora*, they do appear to extend into the area encompassed by the pseudosepta (Crawford *et al*. 1993). While *H. baltica* has a rimoprotula proximal to the pseudosepta, in *H. sinuata* the position of the rimoportulae is closer to the apices, as in *Cyclophora*.

Crawford *et al*. (1993) noted the similarities between *Hustedtiella* and *Cyclophora*, but refrained from combining the two into one genus. In support of this decision, they cited differences in habit, heterovalvy and plastid morphology; while they could not see evidence at the time from drawings that the plastids of *C. tenuis* entered the pseudosepta, our TEM images of *C. tenuis* show that this does occur. While *C.*

castracanei is also isovalvar, its strongly-supported relationship to the heterovalvar *C. tabellariformis* in the molecular phylogeny make its transfer to *Hustedtiella* difficult to resolve without also including *C. tabellariformis* in the transfer. The apical pore fields of both *Hustedtiella* spp. differ from *C. castracanei*, in that there are far more slits comprising the field in *C. castracanei* (17-20) than *H. baltica* (4-5) and the pore fields in *Hustedtiella spp*. are more sunken and sharply defined, to the point where Crawford *et al*. (1993) refer to them as "ocelli." *C. minor* is more similar to *Hustedtiella* spp. in size and number of slits in the apical pore field but here, too, the field is not sunken into an ocellus. Without DNA or ultrastructural data we are uncomfortable making any assessment of the taxonomy of *Hustedtiella* beyond acknowledging that it is likely to be closely related to *Cyclophora* and *Astrosyne*.

"Centric" Morphology in "Pennate" Diatoms

Gross morphology and symmetry have been important in diatom systematics for as long as diatoms have been described, especially with regard to higher-level taxonomy. However, as more taxa are discovered and previously-described taxa are critically re-examined, it has become apparent that the classical distinctions between the major groups of diatoms are not always so clear. Other traits and characters have been suggested, such as oogamy (Cherupanov *et al*. 2004), cellular ultrastructure and valve morphogenesis (Medlin & Kaczmarska 2004). While these characters may present a more realistic reflection of the true evolutionary history of the diatoms, relative ease of observing frustule morphology makes it unlikely to disappear from common usage.

Astrosyne radiata is not the first diatom taxon that does not fit cleanly into "centric" or "pennate" morphology. As more taxa from the bipolar centric and araphid

pennate lineages are added to molecular phylogenetic analyses, the distinction between the two seems to have blurred somewhat; much has already been written about the "extremely elongate" bipolar centric diatoms *Toxarium, Climacosphenia* and *Ardissonea* (Kooistra *et al*. 2003, Alverson *et al*. 2006, Medlin *et al*. 2008) and the Plagiogrammaceae, which have been suggested as bipolar centrics (Round *et al*. 1990) and araphid pennates (Simonsen 1979, Kooistra *et al*. 2004).

Conversely, at least one round diatom originally described as centric is now seen as belonging to an araphid pennate lineage—*Psammodiscus nitidus* (Gregory) Round & Mann. Originally described as a *Coscinodiscus* species, Round & Mann (1980) noted that *P. nitidus* lacked the typical coscinodiscoid loculate areolae and rim of rimoportulae around the valve margin, including the macrorimoportula. Instead, *P. nitidus* possesses a single, small rimoportula (if any) in the central area of the valve and simple, occluded pores. Based on those characters, the authors suggested a closer relationship to the araphid genus *Rhaphoneis*, which also has relatively simple occluded pores and tend to be epipsammic in habit, as is *P. nitidus*. This hypothesis has yet to be tested with molecular data, however.

While a lack of taxon sampling within the araphid pennates prevents any strongly-supported estimates to the closest ancestor to the *Cyclophora* clade, the position of *Astrosyne* would suggest that this is likely to be a secondarily-derived loss of lateral symmetry (or regaining of radial symmetry). The molecular data and the shared ultrastructural characters associated with the pseudoseptum between *Cyclophora* and *Astrosyne* make it extremely unlikely that *Astrosyne* could be viewed as some "transitional" taxon between centric and pennate diatoms. Despite the difference in gross valve morphology, we feel that between the DNA data and the ultrastructural similarity in

the pyrenoid-containing pseudoseptum there is enough evidence to suggest a close relationship between *Astrosyne* and *Cyclophora*.

Chapter 2: Revisiting Ross and Sims (1971): Towards a molecular phylogeny of the Biddulphiaceae and Eupodiscaceae (Bacillariophyceae)

INTRODUCTION

Schütt (1896) created the subfamily Biddulphioideae and defined it as those diatoms with two or more poles, generally elliptical in shape, with poles elevated as bumps or horns. He included in the Biddulphioideae a number of genera which broadly resemble *Biddulphia* Gray, such as *Odontella* C. Agardh, *Triceratium* Ehrenberg, *Eucampia* Ehrenberg, *Isthmia* C. Agardh, *Hemiaulus* Heiberg, *Anaulus* Ehrenberg, *Hemidiscus* G.C. Wallich, *Eupodiscus* J.W. Bailey, *Cerataulus* Ehrenberg and *Chaetoceros* Ehrenberg. Molecular data place these diatoms together with a number of other polar and even radially structured non-pennate diatoms (the Thalassiosirales) in a group formally designated the Mediophyceae (Medlin and Kaczmarska 2004). Support for Mediophyceae monophyly has been equivocal (Theriot et al. 2009, 2010, 2011) and some authors consider the group to be paraphyletic (Adl et al. 2005).

Hendey (1971) lamented the "feeling of apathy" by diatomists towards the diatoms fitting Schütt's definition (Hendey used the term "Biddulphiaceae"), and described the group as "a heterogenous collection of wholly dissimilar units, bound together only by our inability to see clearly taxonomic boundaries providing bases for useful subdivisions." That same year, Ross and Sims (1971) used the scanning electron microscope (SEM) in an attempt to find such boundaries. They argued that there were generally two types of pore fields present in the diatoms they observed: fine pore fields surrounded by a solid, unmarked rim ("ocelli"), or apices marked by a gradation of the valve markings to a field of fine pores ("pseudocelli"). Furthermore, Ross and Sims

grouped taxa observed into general categories based on valve perforation: porose valves with pore occlusions that ranged from simple bars (such as in *Biddulphiopsis* von Stosch and Simonsen or *Lampriscus* A.Schmidt) to flat or domed vela (such as those in *Biddulphia biddulphiana* [J.W. Smith] C.S. Boyer and *Odontella aurita* [Lyngbye] C. Agardh, respectively); or loculate, chambered valves (such as *Odontella mobiliensis* [J.W. Bailey] Grunow or *Odontella rhombus* [Ehrenberg] Kützing). They also identified a number of taxa which bore pores with domed cribra surrounded by incomplete and irregular chambers, dubbed "pseudoloculate" valves (*Triceratium favus* Ehrenberg and *Triceratum dictyotum* P.A. Sims and R. Ross).

Using the different types of apical pore fields and valve perforation characters, Ross and Sims created four informal groupings: taxa with porose valves and pseudocelli, taxa with loculate valves and pseudocelli, taxa with porose valves and ocelli and taxa with loculate valves and ocelli. They suggested the porose/pseudocellate taxa belong in *Biddulphia*, though they posited that those taxa without internal ribs might belong in a different genus. The loculate/pseudocellate taxa were assigned to *Trigonium* Cleve. The ocellate taxa they examined showed a greater degree of variation in valve morphology, however. They divided the loculate/ocellate taxa into two genera: *Cerataulus*, for taxa which featured a contorted valve (i.e. *Cerataulus turgidus* [Ehrenberg] Ehrenberg) and *Zygoceros* Ehrenberg, for taxa without torsion (i.e. *O. mobiliensis, O. rhombus*). The porose/ocellate taxa presented a particularly difficult situation, because of a great deal of morphological variability beyond these two characters. Thus, Ross and Sims (1971) proposed two options: either all porose/ocellate taxa should fall under *Odontella* (type = *O. aurita*), or these taxa should be divided into a number of genera, including *Odontella*, the pseudoloculate *Triceratium*, *Amphitetras* Ehrenberg and *Amphipentas* Ehrenberg.

Other classifications generally followed Ross and Sims (1971) in terms of suprageneric groupings, although there was still considerable disagreement about the relationship (and rank) of these suprageneric groups. Simonsen (1979) created the "Biddulphiineae" which was similar in composition to Schütt's Biddulphioideae. The Family Biddulphiaceae was characterized by pseudocelli while the Eupodiscaceae was characterized by ocelli. Simonsen's (1979) Biddulphiineae was not monophyletic. He placed the Biddulphiaceae in a clade with Chaetoceraceae and Lithodemiaceae and the Eupodiscaceae were placed as sister to pennates. Round et al. (1990) proposed a similar scheme, but elevated what were essentially Simonsen's Biddulphiaceae and Eupodiscaceae to the ordinal level: Biddulphiales and Triceratiales, respectively. These were placed in the subclass "Biddulphiophycidae." Nikolaev (1990) split the biddulphioids across three orders: the Triceratiaceae in the Pyxidiculales, the Trigoniaceae and Eupodiscaceae in the Coscinodiscales (resembling Schütt) and the Auliscaceae, Biddulphiaceae and Isthmiaceae in the Biddulphiales. However, Simonsen's two-family proposal seems to occur most commonly in the literature. This ambiguity is actually seen in the molecular data as well—while certain clades of diatoms are well-supported, the relationships between these large clades (roughly analogous to order and family-level) are poorly-resolved (Theriot et al. 2011).

Far less has been written regarding the phylogenetic relationships within the Eupodiscaceae and Biddulphiaceae, possibly due to the taxonomic uncertainty discussed previously. Glezer (1979) looked at the phylogeny of the Biddulphiaceae of extant and fossil forms, but focused primarily on the areolar structure and the overall frustule shape, asserting that cylindrical forms were ancestral to an ellipsoidal lineage (such as *Eunotogramma* Weisse, *Eucampia*, *Biddulphia*) and prismatic lineage (such as *Trinacria* Heiberg, *Amphitetras*, or the vaguely-described *Sheshukovia* Glezer). Glezer mentions

Ross and Sims' ocellus, but suggests that it has evolved several times across the diatoms: in *Odontella* (derived from *Biddulphia* in her tree diagram), in *Amphitetras* (derived from *Sheshukovia*) and in *Triceratium* (an altogether different lineage not included in the tree). In a later paper, Ross and Sims (1987) point to the variability in valve outline, even within a species (using *Odontella retiformis* [A, Mann] von Stosch as an example) as evidence against Glezer's scheme, however.

In a later paper on family and order-level taxonomy in the "centric" diatoms, Ross and Sims (1973) state that the defining character for the Eupodiscaceae is the ocellus, and suggest that this structure was derived from the "field of smaller areolae on the elevations that characterizes [the Biddulpiaceae]." Simonsen (1979) also used the ocellus as a key character of the Eupodiscaceae. They acknowledge that the loculate and pseudoloculate valve conditions occur outside of the Eupodiscaceae (they are present on *Coscinodiscus* Ehrenberg and *Stephanopyxis* [Ehrenberg] Ehrenberg, respectively). However, they suggest it is more likely for loculate valves to have evolved twice (once in the Eupodiscaceae and once in the Coscinodiscaceae) than for ocelli to have evolved twice.

Simonsen (1979) disagreed regarding the importance of the ocellus, stating that the "true ocellus" could be found in other taxa with no affinity to the Eupodiscaceae. As an example, he cites *Cymatosira lorenziana* Grunow. The Cymatosiraceae do feature elevated, rimmed apical pore fields. However, they typically possess fewer pores, feature short spikes between the pores and are concave in structure rather than convex like the ocelli of *Odontella* and *Triceratium*. To reflect these differences the pore fields of the Cymatosiraceae were given the name "ocellulus" (Hasle et al. 1983). Hasle et al. (1983) do not comment on the relationship between the ocellulus and the eupodiscoid ocellus except to say they are similarly structured. The term "ocellus" has been used to describe

several different structures, such as the apical pore field in the araphid pennate diatom *Bleakeleya* Round and is discussed in Lobban et al. (2011).

Part of the ambiguity in previous attempts at classifying these diatoms may lie in the reliance of plesiomorphic characters to delimit taxonomy. This is a well-documented problem in diatoms; classical structural groups such as "centric" and "araphid pennate", defined by lack of apomorphic characters, continue to be used in formal classifications despite their non-monophyletic nature (see Williams and Kociolek 2007). Some authors continue to argue against monophyly as the necessary and sufficient basis for classification (Medlin 2010). Within the biddulphioid/eupodiscoid diatoms, all authors who have written on the subject seem to agree that valve perforation characters are important, but the issue becomes how these characters are interpreted on a phylogeny. For example, groups of larger pores grading grading rapidly into smaller pores ("pseudocellus") may be homologous within some deeper level of the diatom tree, representing a plesiomorphic condition relative to a pore field surrounded by a hyaline ring ("ocellus"). In turn, this may reflect plesiomorphic resemblance to even more derived conditions such as the ocellulus, representing synapomorphy within ocellate diatoms.

Certainly a helpful and needed approach would be an analysis of a full matrix to resolve questions of synapomorphy versus plesiomorphy in the morphological data. We suggest that a molecular approach to the phylogeny is a needed "first step" to guide further studies with a full matrix; the matrix would have to include detailed analysis of these and other structures from numerous fossil taxa, (not necessarily just those in the Biddulphiaceae and Eupodiscaceae, since their relationship to the rest of the diatom tree is still poorly-understood) for which there is no DNA data for comparison anyway. Instead, we will treat the various characters argued as diagnostic for groups (particularly

the characters of Ross and Sims [1971]), as proposed synapomorphies and examine the extent to which there is agreement between hypotheses suggested by molecular data and these hypothesized morphological synapomorphies.

Our molecular dataset extends the three gene dataset (nuclear-encoded SSU rDNA, and chloroplast encoded *rbc*L and *psb*C) of Theriot et al. (2010, 2011), by nearly doubling the number of diatom sequences, with a particular increase in Mediophyceae. We assess the correspondence between molecular data and previously proposed classification schemes (Ross and Sims 1973, Simonsen 1979, Glezer 1979) or putative diagnostic characters (untested synapomorphies; Ross and Sims 1971) by comparing the likelihoods of trees constrained to reflect previous hypotheses to the unconstrained tree. This approach has been used to address questions of classification and character evolution in a wide array of taxa including diatoms (e.g. Theriot et al. 2011, Ruck & Theriot 2011), protists (e.g. Embley et al. 1995, Deschamps et al. 2011), and metazoans (e.g. Sperling et al. 2009). We recognize that this approach has its limitations and that a cladistic analysis of morphological characters would be a preferred way of testing morphological synapomorphies or correspondence/conflict between molecular and morphological data (cf. Losos et al. 2012). However, the hypothesis testing performed here may be our best course of action until an extensive, coded morphological dataset is compiled. We also argue that enough morphological and molecular evidence exists to transfer three species from Ross and Sims' concept of *Zygoceros—O. mobiliensis* , *Odontella regia* (M. Schultze) Simonsen and *Odontella sinensis* (Greville) Grunow—to a new genus: *Trieres*.

METHODS AND MATERIALS

Samples were taken from a variety of benthic and planktonic habitats (Appendix 1). Benthic taxa were isolated from hand-collected macrophytes or scraped directly from hard substrates by SCUBA or snorkel divers. Planktonic isolates were isolated from collections made by net with a 20 μm mesh size. The remaining sampled material was preserved in a formaldehyde-seawater mix. All marine isolates were grown in f/2 medium (Guillard 1975) at 27° C (Guam isolates) or at 14° C (northern CA, OR, WA isolates) in a Percival model I-36LL incubator under a 12:12 light:dark photoperiod under flourescent lights. All other isolates were grown at room temperature in a north-facing window in ambient light at around 21-22°C (TX, FL, HI and Canary Island isolates).

Microscopy Preparation

Wild and cultured material was acid-cleaned with hydrogen peroxide and nitric acid and washed until pH neutral. Permanent mounts for light microscopy were made using Naphrax mounting medium.

For SEM preparation, cleaned wild material was filtered and dried onto Millipore 1.2 μm Isopore membrane filters while cultured material was dried onto 12 mm round coverslips. Material was then coated with iridium or platinum/palladium using a Cressington 208 Bench Top Sputter Coater and observed with a Zeiss SUPRA 40 VP scanning electron microscope.

All strains used for DNA sequencing have LM and/or SEM photo vouchers, available from http://www.protistcentral.org/Project/get/project_id/79 or the senior author.

DNA Extraction, Sequencing and Phylogenetic Analysis

Aliquots of cultured isolates were pelleted in a Sorvall RC-5B refrigerated superspeed centrifuge for 20 minutes at 8,000 rpm. DNA was extracted from the diatom pellets using a QIAGEN DNeasy Plant Mini Kit (QIAGEN Sciences, Maryland) after 45 seconds of cell disruption using 1.0 mm glass beads in a Mini-Beadbeater (Biospec Products, Inc). PCR amplification and sequencing of small-subunit nuclear rRNA, and the chloroplast-encoded *rbc*L and *psb*C markers, follows the primers and protocols of Theriot et al. (2010).

Phylogenetic analyses were conducted using 245 diatom taxa plus *Bolidomonas pacifica* L.Guillou & M.-J.Chrétiennot-Dinet as the outgroup. Genbank accession numbers for the new sequences presented in this paper are listed in Appendix 1. All extracted DNA were given an accession number local to the Theriot lab which corresponds to a slide and SEM stub voucher ("HK###"). Sequence data were partitioned by gene and by codon position and run in RAxML ver. 7.2.8 (Stamatakis 2006) for Maximum Likelihood (ML) analysis using the 6-substitution GTR + G + I model. To obtain a most likely tree, we ran 1024 ML optimizations each starting from a parsimony tree. Clade support was obtained from 1000 nonparametric bootstrap replicates using the rapid bootstrap algorithm in RAxML. The tree with the highest log likelihood score from 1024 unconstrained runs with bootstrap proportions is presented in Figures 57 and 58. All RAxML runs were completed on the RANGER or LONESTAR supercomputers at the Texas Advanced Computing Center (TACC). Two Bayesian Inference runs were also conducted with MrBayes ver. 3.1.2, using the same substitution model and partitioning strategy as the ML analyses. Two Markov chain Monte Carlo (MCMC) runs with 4 chains each (1 cold and 3 heated) were iterated for 10^8 generations. The posterior distributions of tree topology, branch lengths and parameter estimates were samples

every 1000^{th} iteration producing a total of 10^5 samples. The posterior samples of the two runs were compared in AWTY (Wilgenbusch et al. 2004) for evidence of convergence. The final 10^3 trees were used to generate a majority rule consensus tree and obtain posterior probabilities for nodes (plotted to nodes of the ML phylogram in Figures 57 and 58).

For the constraint analysis, the eupodiscoid and biddulphioid taxa used in this analysis were assigned groupings based on Ross and Sims' character categories (Table 4). Examples of these characters can be found on Figure 59: ocelli (a-e), pseudocelli (f-j), porose valves (k-l), pseudoloculate porose valves (m-n) and loculate valves (o-p). We used Mesquite ver. 2.74 (Maddison and Maddison 2010) to create constraint topologies reflecting previous classification hypotheses or grouping of taxa based on shared morphological features (e.g. valve poration and apical process—see Table 4). Using the same procedure as for the unconstrained tree, we performed ML searches under six topological constraints (Table 5). We then found the best optimization for each of the constraint analyses and calculated the per-site log likelihoods of all alignment columns given the tree. Using these files, we performed the Shimodaira-Hasegawa test (SH; Shimodaira and Hasegawa 1999) between the best out of 1024 optimizations for each constraint and the best unconstrained tree (Figures 57, 58).

Taxon (Voucher Code HKXXX)	Valve Perforation	Apical Pore Field
Biddulphia biddulphiana (HK271)	**Porose**	**Pseudocellus**
Biddulphia biddulphiana (HK328)	**Porose**	**Pseudocellus**
Biddulphia tridens (HK239)	**Porose**	**Pseudocellus**
Biddulphia tridens (HK327)	**Porose**	**Pseudocellus**
Biddulphia reticulata (HK252)	**Porose**	**Pseudocellus**
Biddulphia cf *reticulata* (HK329)	**Porose**	**Pseudocellus**
Biddulphia alternans (HK292)	**Porose**	**Pseudocellus**
Biddulphiopsis titiana (HK170)	**Porose**	**Pseudocellus**
Isthmia minima (HK275)	**Porose**	**Pseudocellus**
Terpsinoë musica (HK273)	**Porose**	**Pseudocellus**
Hydrosera sp (HK274)	**Porose**	**Pseudocellus**
Trigonium formosum (HK200)	**Loculate**	**Pseudocellus**
Trigonium formosum (HK258)	**Loculate**	**Pseudocellus**
Trigonium formosum f. quadrangularis (CX17)	**Loculate**	**Pseudocellus**
Cerataulus smithii (HK224)	**Porose (pseudoloculate)**	**Ocellus**
Cerataulus smithii (HK332)	**Porose (pseudoloculate)**	**Ocellus**
Cerataulus smithii (HK331)	**Porose (pseudoloculate)**	**Ocellus**
Pleurosira laevis (HK068)	**Loculate**	**Ocellus**
Pleurosira laevis f. polymorpha (HK326)	**Loculate**	**Ocellus**
Mastodiscus radiatus (HK249)	**Porose**	**Ocellus**
Pseudauliscus peruvianus (HK330)	**Loculate**	**Ocellus**
Odontella aurita (HK203)	**Porose**	**Ocellus**
Odontella aurita (HK333)	**Porose**	**Ocellus**
Odontella aurita (HK334)	**Porose**	**Ocellus**
Odontella aurita v minima (HK336)	**Porose (pseudoloculate)**	**Ocellus**
Odontella aurita v. minima (HK337)	**Porose (pseudoloculate)**	**Ocellus**
Odontella rostrata (HK225)	**Porose**	**Ocellus**
Odontella longicruris (HK226)	**Porose**	**Ocellus**
Odontella longicruris v. hyalina (HK284)	**Porose**	**Ocellus**
Odontella longicruris v. hyalina (HK338)	**Porose**	**Ocellus**
Odontella sp. (HK335)	**Porose**	**Ocellus**
Odontella sp. "pseudoloc" (HK341)	**Porose (pseudoloculate)**	**Ocellus**
Odontella rhombus f. trigona (HK340)	**Loculate**	**Ocellus**
Odontella rhomboides (HK339)	**Loculate**	**Ocellus**
Odontella rhomboides (HK282)	**Loculate**	**Ocellus**
Amphitetras antediluviana (HK223)	**Porose**	**Ocellus**
Amphipentas pentacrinus (HK289)	**Porose**	**Ocellus**
Triceratium bicorne (HK222)	**Porose (pseudoloculate)**	**Ocellus**
Triceratium dictyotum (HK281)	**Porose (pseudoloculate)**	**Ocellus**
Triceratium dubium (HK254)	**Porose (pseudoloculate)**	**Ocellus**
Triceratium dubium (HK342)	**Porose (pseudoloculate)**	**Ocellus**
Triceratium dubium (HK199)	**Porose (pseudoloculate)**	**Ocellus**
Triceratium sp (HK047)	**Porose (pseudoloculate)**	**Ocellus**
Trieres mobilensis (HK204)	**Loculate**	**Ocellus**
Trieres mobilensis (HK227)	**Loculate**	**Ocellus**
Trieres mobilensis (HK251)	**Loculate**	**Ocellus**
Trieres regia (HK290)	**Loculate**	**Ocellus**
Trieres regia (HK322)	**Loculate**	**Ocellus**
Trieres sinensis (HK037)	**Loculate**	**Ocellus**
Trieres sinensis (HK323)	**Loculate**	**Ocellus**

Table 4. Biddulphioid taxa used in constraint analyses with assigned morphological character classes. For the "Valve Perforation" character, an extra notation is made for taxa with "pseudoloculate valves".

Constraint Tree	-lnL	ΔlnL	Significantly worse?
Unconstrained Best Tree	-164645.29	0.0	
Monophyletic Porose valve/Ocellus	-164673.56	-28.27	No
Monophyletic Loculate valve/Ocellus	-164855.86	-210.57	Yes
Monophyletic Pseudoloculate valve/Ocellus	-164659.43	-14.14	No
Monophyletic Dome-shaped Cribra	-166908.61	-2263.32	Yes
Monophyletic Biddulphiaceae + Hemiaulaceae	-165062.54	-417.25	Yes

Table 5. Shimodaira-Hasegawa test for the topological constraint trees. The "unconstrained best tree" is the tree shown in Figs. 57 & 58. Significance was assessed at $\alpha = 0.05$. The difference in likelihood (ΔlnL) was calculated by subtracting the log likelihood score of each constraint from the score of the best unconstrained tree.

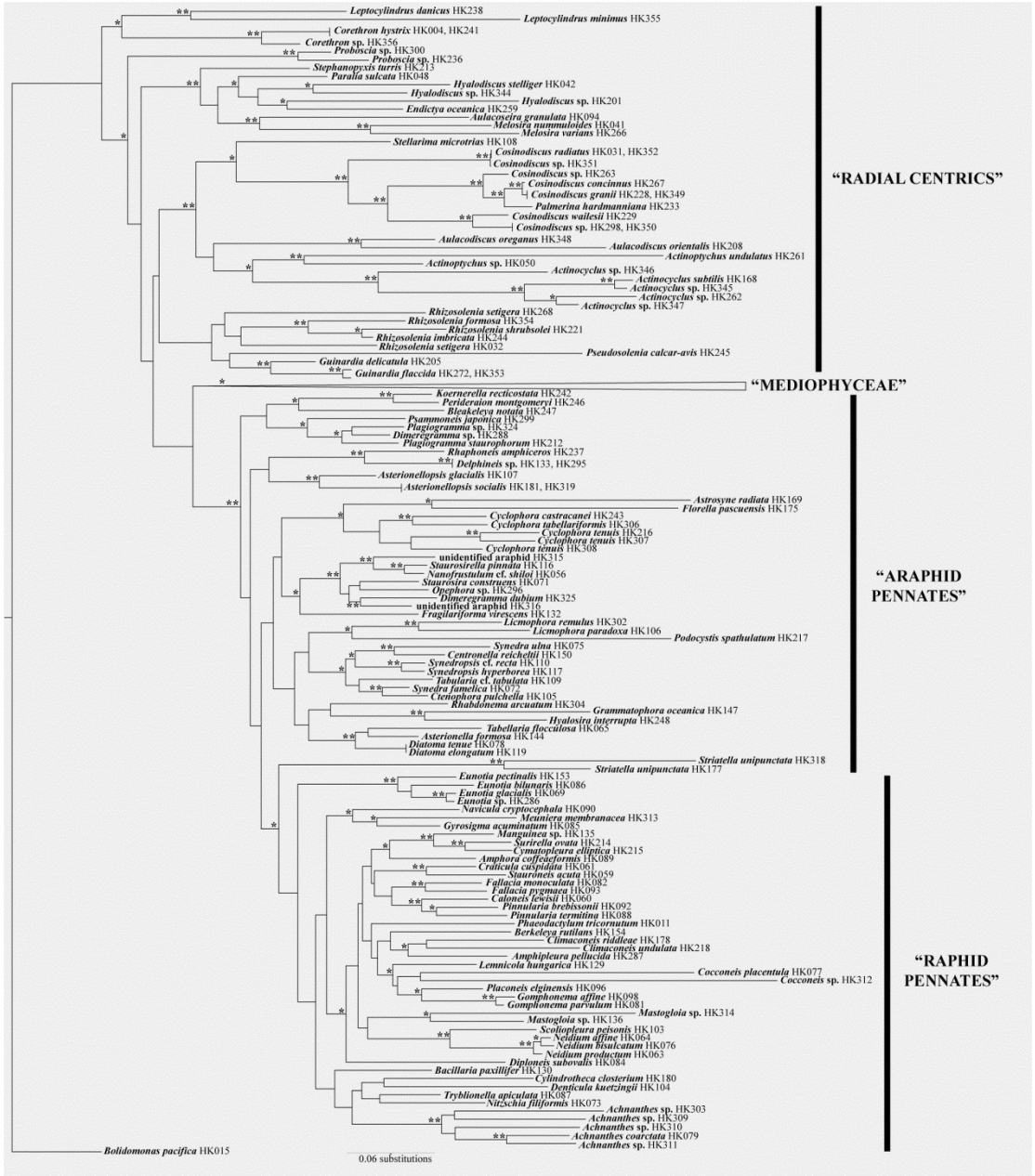


Figure 57. Phylogram depicting the best tree from the Maximum Likelihood and Bayesian phylogenetic analyses. Nodes with double asterisks (**) denote clades with ≥ 100% bootstrap support and ≥ 1.00 posterior probability, while nodes with single asterisks (*) denote clades with ≥ 75% bootstrap support and ≥ 0.95 posterior probability. The mediophycean clade is collapsed here for clarity and shown in Figure 58.

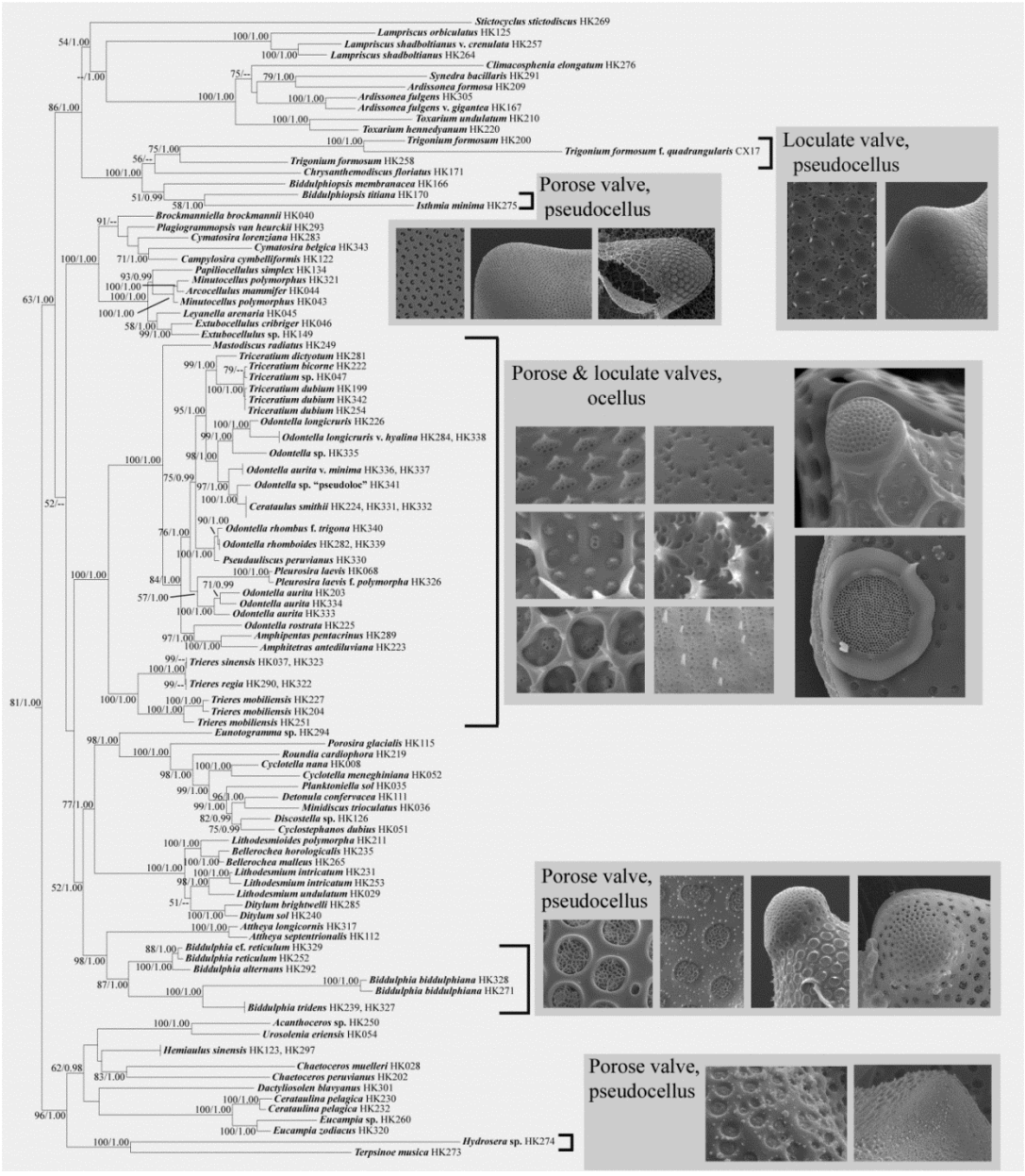


Figure 58. Phylogram depicting the mediophycean clade of the best tree from the Maximum Likelihood and Bayesian phylogenetic analyses. Values at nodes correspond to ML bootstrap values and BI posterior probability values, respectively. Only bootstrap values over 50% and posterior probabilities over 0.95 are shown on the tree. Biddulphiacean and eupodiscacean clades are identified by valve perforation and apical pore field morphology, with SEM exemplars in shaded boxes adjacent to each clade.

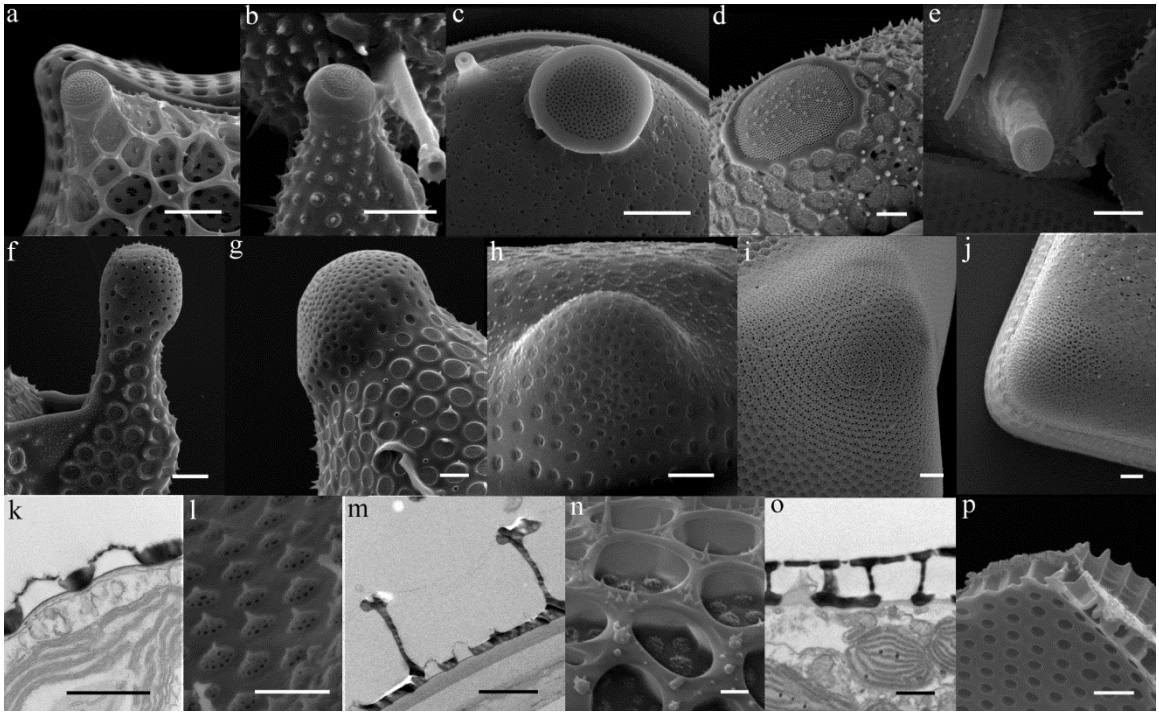


Figure 59. Scanning electron micrographs showing the different morphologies of ocellus (a-e), pseudocellus (f-j) and valve perforation (l, n, p) exhibited by the biddulphioid diatoms in this study. Transmission electron micrographs are also provided to illustrate valve perforation (k, m, o). Porose valves shown in (k, l), pseudoloculate valves shown in (m, n) and loculate valves shown in (o, p). (a) *Triceratium dubium* (HK342). (b) *Odontella aurita* (HK333). (c) *Pseudauliscus peruvianus* (HK330). (d) *Amphitetras antediluviana* (HK223). (e, p) *Trieres mobilensis* (HK227). (f) *Biddulphia tridens* (HK327). (g) *Biddulphia biddulphiana* (GU44AK4, Guam). (h) *Biddulphia reticulum* (HK252). (i) *Trigonium diaphanum* (GU44AN7, Guam). (j) *Trigonium formosum* (HK258). (k) *Odontella aurita* (HK334). (l) *Odontella longicruris* (HK226). (m, n) *Triceratium dictyotum* (HK281). (o) *Trigonium formosum* (GU44AK-6). Scale bars: 1 µm in (k, m, o); 2 µm in (a-j, l, n, p).

RESULTS AND DISCUSSION

Phylogenetic Analysis

In our combined three-gene analysis, the Mediophyceae formed a clade (Figure 57, ML bootstrap = 81%, posterior probability = 1.00) sister to the pennate diatoms (ML bs = 100%, pp = 1.00). The remaining radially-symmetrical non-pennate diatoms formed

a basal "grade of clades," with clades of the genera *Leptocylindrus* Cleve/*Corethron* Castracane (ML bs = 85%, pp = 1.00), *Proboscia* Sundström (ML bs = 100%, pp = 1.00), *Stephanopyxis*/*Paralia* Heiberg/*Hyalodiscus* Ehrenberg/*Endictya* Ehrenberg/*Aulacoseira* Thwaites/*Melosira* C. Agardh (ML bs = 100%, pp = 1.00), *Stellarima* Hasle & P.A. Sims/*Coscinodiscus* Ehrenberg/*Palmerina* G.R. Hasle plus *Aulacodiscus* Ehrenberg/*Actinoptychus* Ehrenberg/*Actinocyclus* Ehrenberg (ML bs = 100%, pp = 1.00) and *Rhizosolenia* Brightwell/*Pseudosolenia* B.G. Sundström/*Guinardia* H. Peragallo (ML bs = 100%, pp = 1.00).

Within the Mediophyceae, there are a number of clades which received strong statistical support in both Bayesian and ML analyses (Figure 58). The Thalassiosirales, Lithodesmidales and Cymatosiraceae are all strongly-supported (ML bs = 100%, pp = 1.00). Relationships between these clades however, are not as clear. For example, the internode between Thalassiosirales plus *Eunotogramma* sp. and Lithodesmiales has high support in the Bayesian inference (pp=1.00) but only modest ML support (bs=77%). The internodes between the Cymatosiraceae and the "Lithodesmiales/*Eunotogramma* + Thalassiosirales" clade are marked by very low ML support (bs =52%).

There are also a number of other mediophycean clades that are strongly-supported but for which diagnostic morphological features are not immediately obvious (such as the bilabiate process of the Lithodesmiales or the strutted process of the Thalassiosirales). The two *Attheya* T. West species, for example, are sister to the *Biddulphia* clade with fairly high support values (ML bs = 98%, pp = 1.00). The other taxa with setae-like apical projections, *Chaetoceros muelleri* Lemmermann and *C. peruvianus* Brightwell, can be found in a fairly well-supported (ML bs = 98%, pp = 1.00) clade with *Acanthoceras* H. Honigmann, *Urosolenia eriensis* (H.L. Smith) F.E. Round and R.M. Crawford, *Terpsinoë musica* Ehrenberg, *Hydrosera* G.C. Wallich, *Dactyliosolen*

blavyanus (H. Peragallo) Hasle, and the polyphyletic Hemiaulales (*Hemiaulus*, *Cerataulina* Peragallo ex Schütt and *Eucampia* in this analysis), though none of the relationships between these taxa are well-supported. The remaining mediophycean taxa in the analysis can be found in a fairly well-supported clade (ML bs = 86%, pp = 1.00), though relationships within the clade are also not well-supported. The "bifacial annulus" group (Medlin et al. 2008) consisting of *Ardissonea* spp. De Notaris, *Climacosphenia elongatum* J.W. Bailey and *Toxarium* spp. J.W. Bailey is resolved within this clade (ML bs = 100%, pp = 1.00), as are the monophyletic *Lampriscus* spp (ML bs = 100%, pp = 1.00). The last clade in this group includes *Trigonium formosum* (Brightwell) Cleve, *Chrysanthemodiscus floriatus* A. Mann and *Biddulphiopsis* spp., which are paraphyletic with respect to *Isthmia minima* Harvey & Bailey.

The ocellus-bearing taxa form a strongly-supported clade (ML bs = 100%, pp = 1.00), with the planktonic, loculate taxa *Trieres sinensis comb. nov*, *T. regia comb. nov.* and *T. mobiliensis comb. nov.* sister to the rest of the clade (ML bs = 100%, pp = 1.00). The remaining loculate, ocellate taxa (*Odontella rhomboides* R. Jahn & Kusber, *Pseudauliscus peruvianus* [Kitton ex Pritchard] A. Schmidt, *Pleurosira laevis* [Ehrenberg] Compère) are more derived in the tree, with *O. rhomboides* and *P. peruvianus* forming a clade (ML bs = 100%, pp = 1.00) but little support for *Pleurosira*'s position.

The porose, ocellate taxa are also not monophyletic. The disc-shaped *Mastodiscus radiatus* A.K.S.K. Prasad & J.A. Nienow is sister to the rest of the ocellates, except for the *Trieres* clade (ML bs = 100%, pp = 1.00). *Odontella rostrata* (Hustedt) Simonsen, *Amphipentas pentacrinus* Ehrenberg and *Amphitetras antediluviana* Ehrenberg form a clade (ML bs = 97%, pp = 1.00) and the *O. aurita* strains form a clade (ML bs = 100%, pp = 1.00) sister to *P. laevis*, though with low support (ML bs = 57%, pp = 1.00). The

pseudoloculate taxa are not monophyletic, as the porose taxa *Odontella longicruris* (Greville) M.A. Hoban, *O. longicruris v. hyalina* (J.L.B. Schröder) M.A. Hoban and *Odontella sp*. "ECT3854" are sister to the *Cerataulus smithii* Ralfs ex Pritchard /*Odontella aurita v. minima* (Grunow in Van Heurck) De Toni/*Odontella sp*. "pseudoloc" clade (ML bs = 98%, pp = 1.00), This group is sister to the clade containing *Triceratium dictyotum , T. bicorne* Cleve*, T. dubium* Brightwell and *Triceratium sp*. "CCMP147" (ML bs = 95%, pp = 1.00).

Hypothesis Testing

The results of the hypothesis testing of trees for which valve perforation and apical pore field type were constrained to be monophyletic are summarized in Table 5. Constraining the pseudocellate taxa with porose valves lead to a tree which was significantly worse than the unconstrained tree, as the taxa with these characters (ex: *Biddulphia*, *Isthmia*, *Terpsinoë*) were split across three distinct clades. Constraining the pseudocellate/loculate valve condition to monophyly was not necessary as the unconstrained search recovered the monophyly of taxa (*Trigonium* spp.) from this category.

Among the constraint trees for the ocellate taxa, only the tree constraining ocellate taxa with loculate valves was significantly worse than the unconstrained tree. This is not particularly surprising; as Ross and Sims (1971) point out when they created these categories, the loculate valve is a likely plesiomorphic character state. The fact that all ocellate taxa can be found in a single clade is in agreement with the importance placed on that character by the authors in defining the Eupodiscaceae.

Constraining the tree to a sister relationship between the Hemiaulaceae and Biddulphaceae, as suggested by Ross and Sims (1973) and Simonsen (1979) also produced a tree significantly worse than the unconstrained tree. In fact, the Hemiaulaceae was not monophyletic in our analysis—taxa from the Chaetoceraceae are sister to several Hemiaulacean taxa. Interestingly, the biddulphiacean taxa closest to the Hemiaulaceae in our analyses were *Terpsinoë* and *Hydrosera*, two biddulphiacean taxa which do not feature elevated apices.

Ocellate taxa (Eupodiscaceae)

The DNA data in our analysis suggests that the Eupodiscaceae are monophyletic and the convex-topped, undecorated ocellus is likely a synapomorphy of that group, in agreement with the suggestion made by Ross and Sims (1973). The relationship of this eupodiscacean ocellus to the cymatosiracean ocellulus is unclear at this time on the basis of the characteristics of this feature alone. Each could be interpreted as independently derived from pseudocelli or the ocellulus could be derived from the eupodiscacean ocellus. In combination with our molecular results, the former interpretation is preferred. While some authors have referred to the apical pore field of *Lampriscus* spp. as both a pseudocellus or ocellus, we would suggest that it does not cleanly fit the definition of either. There is indeed a hyaline rim surrounding the pores (unlike a pseudocellus), but pore size and density changes drastically before this rim (unlike the eupodiscacean ocellus). Additionally, large linking spines are found attached to the pore-field rim (Figures 60a, b, Navarro 1981, Navarro and Lobban 2009), whereas eupodiscacean ocelli have structureless rims bordering the pore field. This molecular analysis also suggests that other characters proposed by Ross and Sims (1971) are not entirely characteristic to

the Eupodiscaceae; the deflected valve mantle is absent in the *Trieres* clade (Figures 61a-c) and present in the pseudocellate *Biddulphia* clade (*B. biddulphiana* and *B. tridens* [Ehrenberg] Ehrenberg, Figures 62a, b), and the orientation of rimoportula (a character also discussed in detail by Simonsen [1979]) across the eupodiscacean clade is highly variable (Figure 63).

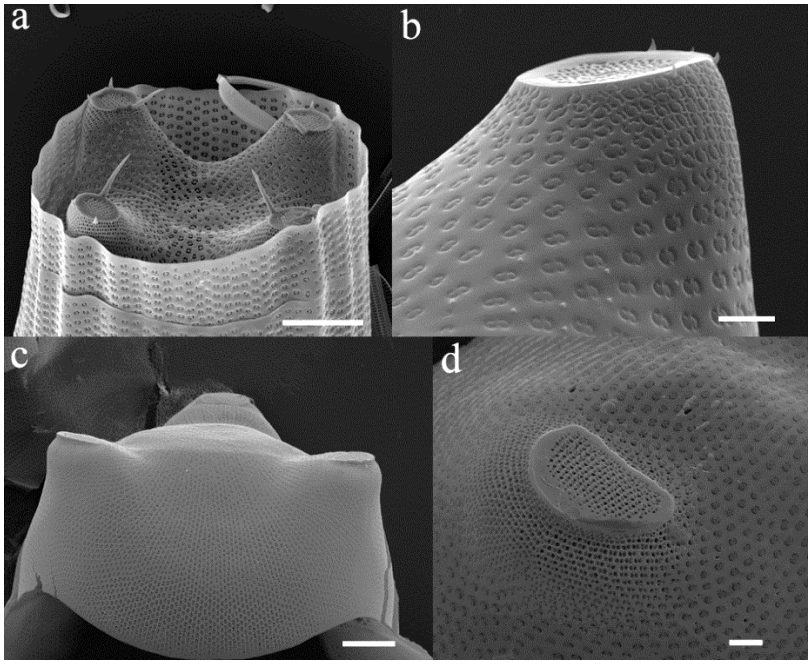


Figure 60. Scanning electron micrographs illustrating the *Lampriscus* taxa used in this study, with details of the apical pore fields (b, d). (a) *Lampriscus shadboltianus v. crentulata* (GU44I-4, Guam). (b) *Lampriscus shadboltianus* (HK264). (c,d) *Lampriscus orbiculatus* (HK125). Scale bars: 10 µm (a, c); 2 µm (b, d).

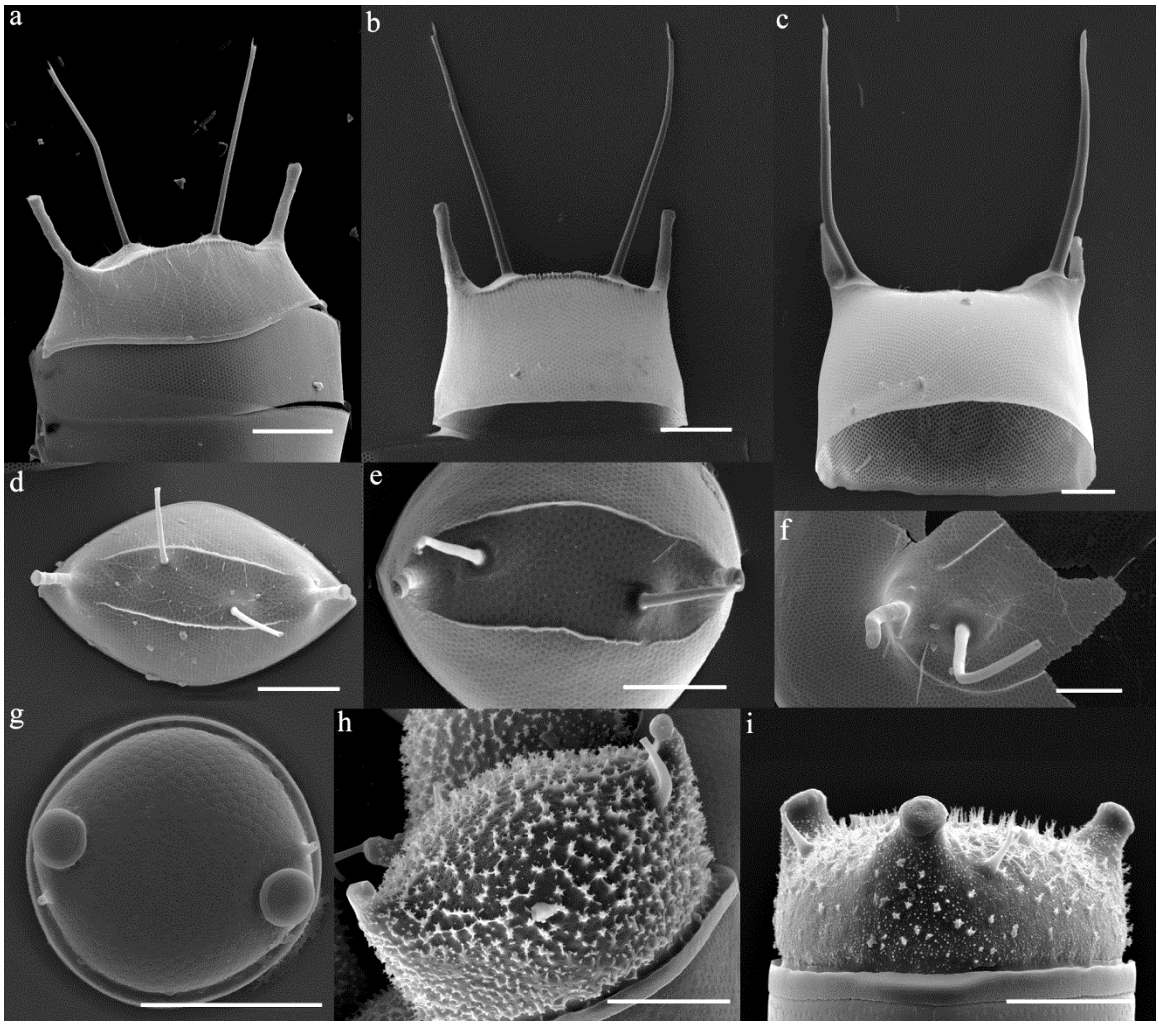


Figure 61. Scanning electron micrographs showing some of the ocellus-bearing, loculate-valved eupodiscacean taxa used in this study. Cells are shown in gridle view (a-c, h, i) and valve view (d-g). (a, d) *Trieres mobiliensis* (HK204). (b, e) *Trieres regia* (HK322). (c, f) *Trieres sinensis* (HK323). (g) *Pseudauliscus peruvianus* (HK330). (h) *Odontella rhomboides* (HK282) (i) *Odontella rhombus f. trigona* (HK340). Scale bars: 10 µm.

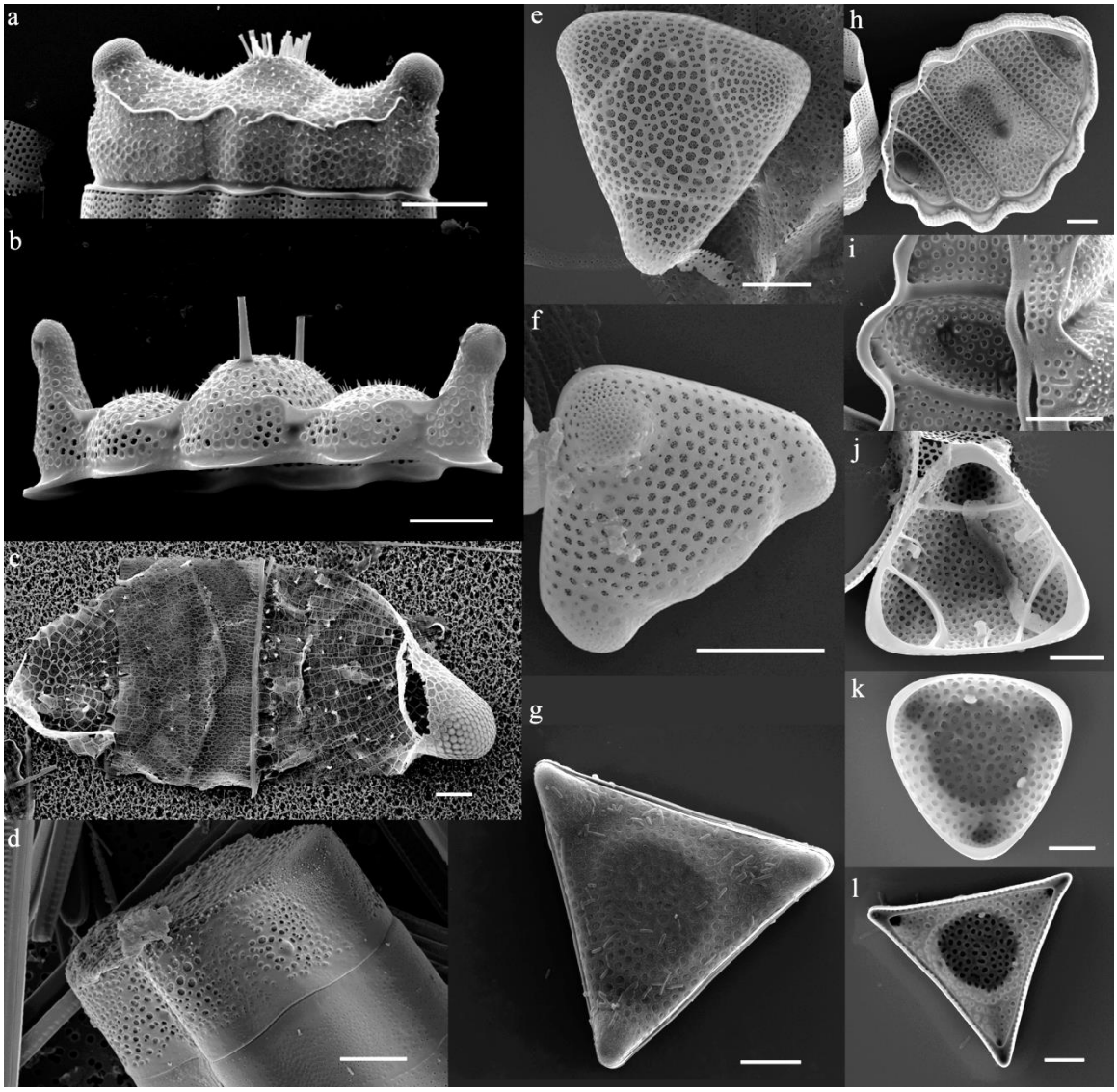


Figure 62. Scanning electron micrographs showing some of the pseudocellus-bearing, porose-valved biddulphiacean taxa used in this study. External views of the various valves are shown in (a-g), while the internal views can be seen in (h-l). (a, h) *Biddulphia biddulphiana* (HK271). (b, i) *Biddulphia tridens* (HK239). (c) *Isthmia minima* (GU44Y-8, Guam). (d) *Terpsinoë musica* (HK273). (e, j) *Biddulphia alternans* (HK292). (f, k) *Biddulphia reticulum* (HK252). (g, l) *Biddulphia* cf *reticulum* (HK329). Scale bars: 20 µm (a) ; 10 µm (b, c, d, f, h, g) ; 5 µm (e, i, j, k, l).

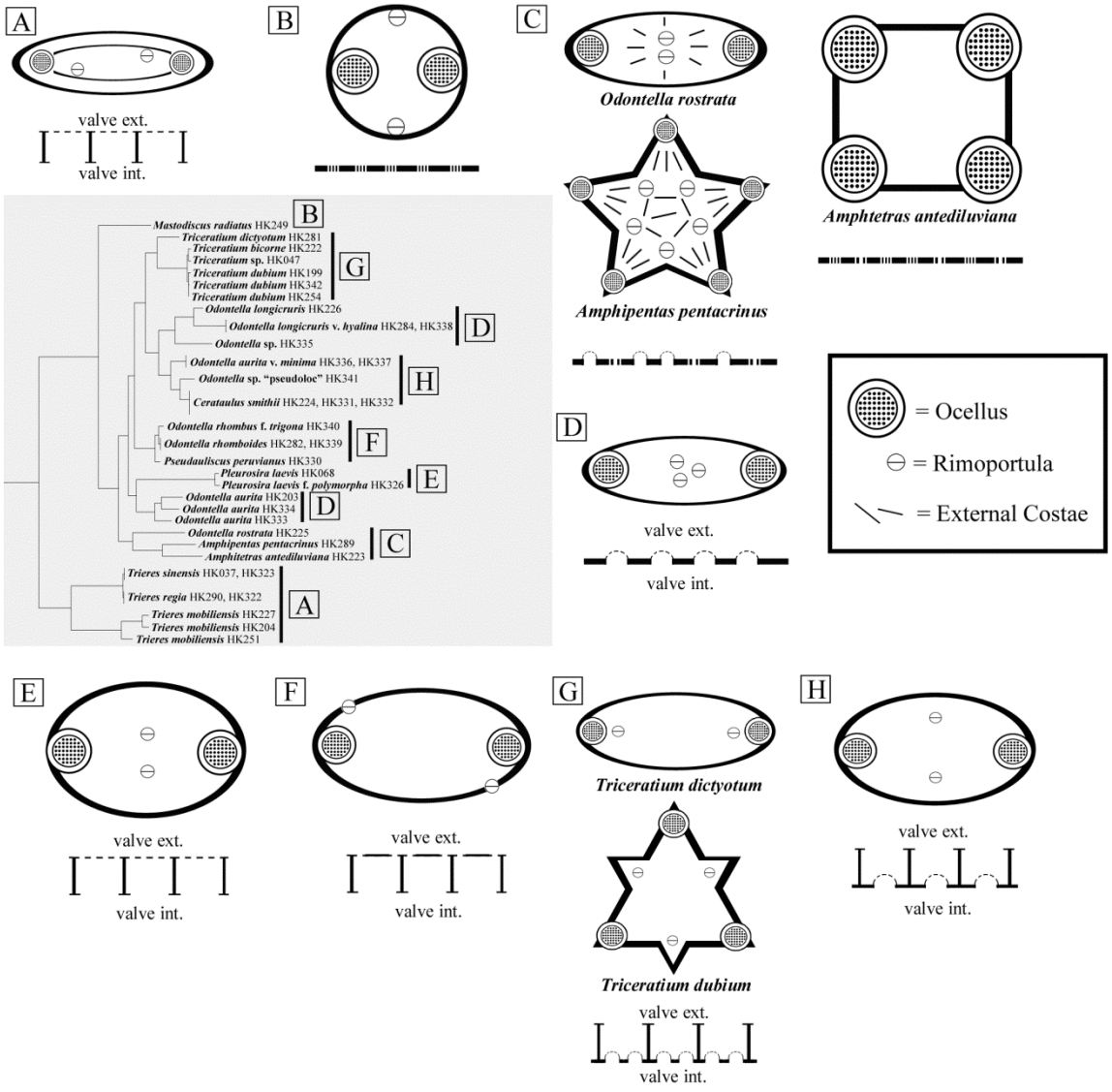


Figure 63. Diagrammatic representation of the valve morphology and perforation of the eupodiscacean clades. Each clade of the "best tree" from Figure 58 has a diagram showing the relative position of the ocellus and rimoportula on the valve face, with a cross-section of the valve diagramed below. Note the two different pore types on the valve cross-section of clade C.

Our results seem to support Ross and Sims' (1971) contention that almost every ocellate mediophycean diatom with a porose valve is an *Odontella*, or none are. The taxa with a convex, undecorated ocellus formed a well-supported clade, which includes the type species of *Odontella*, *O. aurita*, which is sister to *P. laevis*. Other genera making *Odontella* paraphyletic include *Triceratium*, *Amphitetras*, *Amphipentas*, *Cerataulus* and *Mastodiscus* J.W. Bailey. Transferring all these diatoms to *Odontella* would result in a monophyletic, but large and highly diverse genus *Odontella*, by both morphology (Ross and Sims 1971) and DNA data.

The alternative is to break *Odontella*, as presently conceived, into smaller genera and to also retain existing genera which are marked by synapomorphies. For example, the pseudoloculate taxa form two clades—one with multiple pores with dome-shaped cribra per pseudolocular chamber (Figure 64a) and one with a single pore with domed cribra per chamber (Figure 64d). Both clades fits Sims and Ross' (1990) definition of the genus *Triceratium*—bearing pseudoloculate valves and elevated ocelli—but while the former clade (multiple pores) contains strains that might be considered "classical" *Triceratium*, the later clade contains taxa that have never been considered as *Triceratium* in their taxonomic record. This is likely an artifact of the historical idea that *Triceratium* spp. were tri- or multi-angular in shape (Roper 1859, De Toni 1894, Schütt 1896), whereas the single-pore clade strains are bipolar (Figure 64e, f) or round (Figure 64g). At this point we should note that the name attached to one of these strains, *Odontella* sp. HK341 (Figure 64e), is merely provisional, and reflects the currently ambiguous nature of the taxonomy of the genus *Odontella*.

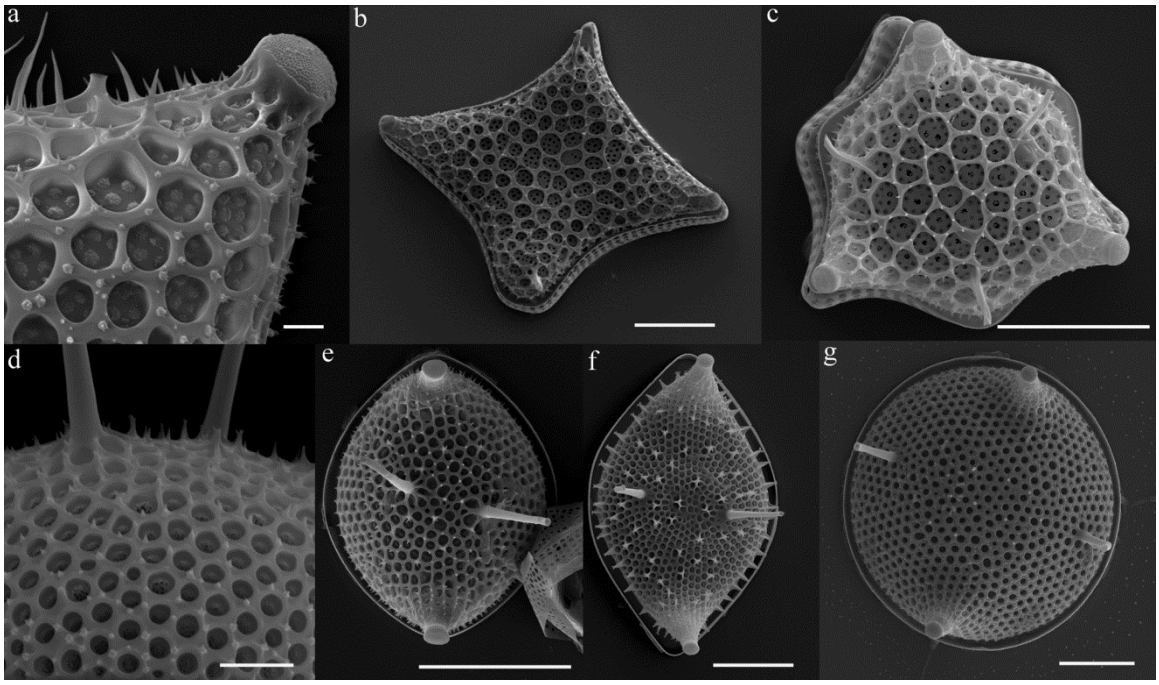


Figure 64. Scanning electron micrographs showing some of the ocellus-bearing, pseudoloculate eupodiscacean taxa used in this study. Exemplar taxa illustrating the two types of pseudoloculate valves—many pores per pseudolocule or one pore per pseudolocule—are shown in (a-c, d-g) respectively. (a) *Triceratium dictyotum* (HK281). (b) *Triceratium bicorne* (HK222). *Triceratium dubium* (HK342). (d, e) *Odontella sp*. "pseudoloc" (HK341). (f) *Odontella aurita v. minima* (HK336). (g) *Cerataulus smithii* (HK331). Scale bars: 10 µm (b, c, e-g); 2 µm (a, d).

At this point it may be premature to simply subsume all pseudoloculate taxa into *Triceratium*. The non-pseudoloculate *O. longicruris* (Figure 65d) strains form a clade between the two pseudoloculate clades (though constraining the pseudoloculate taxa to a single clade results in a tree that is not significantly different than the best tree—see Table 5). Morphologically, there is little to link *O. longicruris* to the pseudoloculate clades beyond possessing ocelli, external rimoportula tubes and domed cribra. Even domed cribra are present in several ocellate clades, such as the *O. aurita* clade (Figure 65a, Figure 63d) and the *O. rostrata* clade (Figure 65c, f , Figure 63c). Constraining the

domed-cribra taxa to a single clade results in a tree that is significantly different from the "best," unconstrained tree (Table 5).

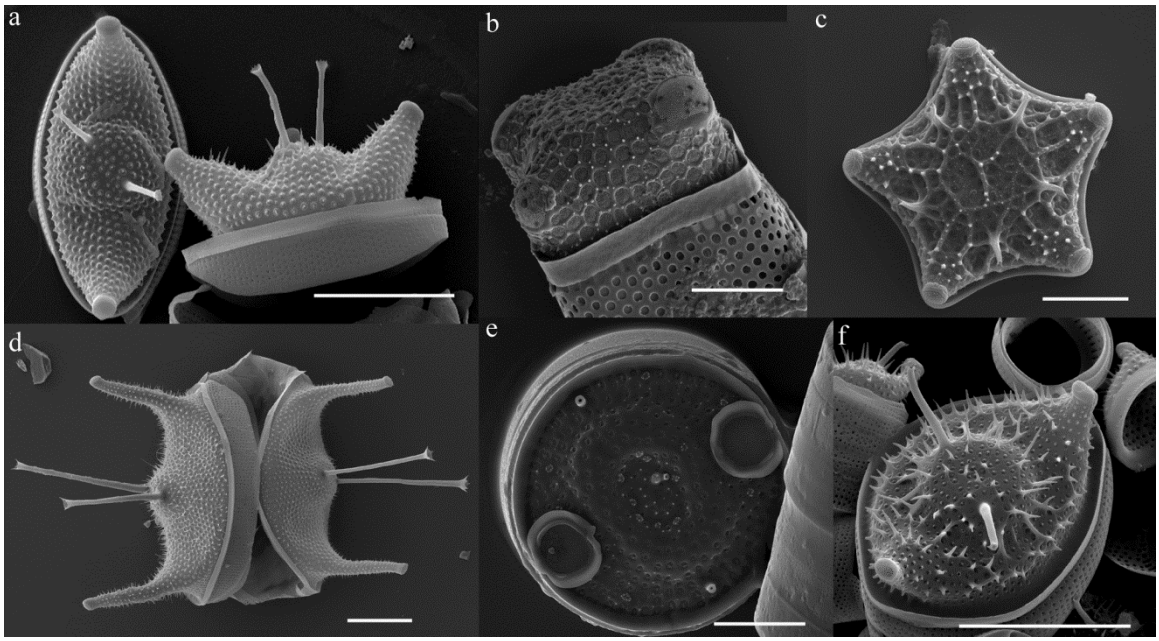


Figure 65. Scanning electron micrographs showing some of the ocellus-bearing, porose-valved eupodiscacean taxa used in this study. (a) *Odontella aurita* (HK333) shown in both valve (left) and girdle views (right). (b) *Amphitetras antediluviana* (HK223). (c) *Amphipentas pentacrinus* (HK289). (d) *Odontella longicruris* (Alt. Plank #8, FL). (e) *Mastodiscus radiatus* (HK249). (f) *Odontella rostrata* (HK225). Scale bars: 10 µm.

The immediate sister clade to the "domed cribra" clade is also diagnosable by morphological characters. *Psuedauliscus peruvianus* (Figure 61g), *O. rhomboides* (Figure 61h) and *Odontella rhombus* f. *trigona* (Figure 61i), have rimoportulae positioned along the margin of the valve face in a counter-clockwise rotation (rather than equidistant along the margin between ocelli), a condition unique to this group among the ocellate taxa sampled here. Ross and Sims (1971) suggested that some taxa in this loculate/ocellate

group could be split from the genus *Odontella*, into *Zygoceros* (type = *Z. rhombus* Ehrenberg) and *Cerataulus* , primarily distinguished by valve torsion (type = *C. turgidus*). Hoban (1979) disagreed, suggesting this would create a genus that included very different morphologies, such as *Z. rhombus* and *Trieres mobiliensis*, and instead proposed "Zygocerae" as a section of *Odontella*. This morphological diversity, however, is mostly in valve outline which is variable even within species of the Eupodiscaceae; therefore, we agree with the Ross and Sims (1971) interpretation where *O. rhomboides* and *O. rhombus* would be returned to *Zygoceros*, with the planktonic loculate/ocellate taxa transferred to *Trieres*, which is discussed separately below. Defining *Zygoceros* here, however, is potentially complicated by *Odontella atlantica* (Frenguelli) Sar, which has the same loculate valve construction as *O. rhombus* and *O. rhomboides*, but has rimoportula immediately adjacent to the ocelli (Sar et al. 2007), rather than submarginally like the latter taxa (*O. rhomboides* = Figure 61h). Commenting on the genus *Cerataulus* with our current data is also problematic: the genus is only represented by *C. smithii* in our data set, which is so radically different from the type *C. turgidus* in the pseudoloculate valve (Figure 64g) and lack of valve torsion that it should likely be transferred to a different genus altogether.

Odontella rostrata (Figure 65f) shares valve characteristics of those taxa with which it is grouped by molecular data: external ribs on the valve like the multipolar *A. pentacrinus* and two types of pore occlusions like both *A. pentacrinus* and the multipolar *A. antediluviana*. (Figures 65c and b, respectively). Ross and Sims (1971) previously made a case for separating *A. pentacrinus* and *A. antediluviana* from *Triceratium* on the basis of the pseudoloculate valve, which the former taxa lack. To this we add the molecular evidence and the second pore occlusion type in the *Amphitetras*/*Amphipentas* clade.

Our proposed new genus, *Trieres* has clockwise orientation of the rimoportulae, which are located on or near the marginal ridge. All three are planktonic in habit, a relative rarity among the ocellate diatoms. All three taxa possess tall, thin ocellus-bearing processes and external rimoportula tubes and a distinctive marginal ridge (Figures 61a-f). Komura (1999) described four ocellate genera from the Miocene which also have tall, thin ocellus-bearing processes and external rimoportula tubes as well as pronounced marginal ridges. *Trieres* is distinguished from the most similar form, *Parodontella* Komura, by the lack of a sternum. It is interesting to note the disparate level of genetic diversity in our *Trieres* strains, with so much genetic diversity in the *T. mobiliensis* strains. We used the descriptions in Hoban (1979), as this work contains extensive LM and SEM documentation of all three species we found, and all our *T. mobiliensis* isolates fit into Hoban's description of that species. If there are several species represented in our *T. mobiliensis* isolates, the difference is not immediately obvious and will require further study. The apparent lack of genetic variation between *T. sinensis* and *T. regia* is not particularly surprising; the two taxa share a very similar morphology, with the only difference being that the rimoportula of *T. sinensis* are raised on the same elevations as the ocellus-bearing processes while the rimoportula of *T. regia* are adjacent but not co-occurring on the ocellate elevations (Hoban 1979).

Thus most smaller clades recovered with molecular data within the ocellate group seem to share morphological characters (or are at least united by overall morphology). We suggest this supports the notion that the ocellate taxa should be reorganized into several genera with a much restricted diagnosis of *Odontella*.

There is one significant exception, however. The strains we identified as *O. aurita*, the generitype, share the externally deflected cribra morphology (Figure 65a) of *Triceratium* and its allies, but the *O. aurita* strains are sister to *P. laevis* which has

external cribra and internal foramina. There is also significant genetic variation among our *O. aurita* strains, whose identification should be considered provisional at this time. This is not a huge surprise, as this diatom was described and illustrated in such a way that many different diatoms fit into the description. Lyngbye's (1819) original sketch showed the frustules in girdle view only, and the description was limited to a description of the arrangements of chloroplasts and the "eared" state of the cell wall, no doubt a reference to the ocelli. As a result, many morphologies have been presented as "*Odontella aurita*" throughout the literature (e.g Peragallo and Peragallo 1898-1908, pl. 98, fig. 3-6, Navarro 1981, fig. 48, Witkowski et al. 2000, pl. 8, figs 12, 13 and pl. 9, figs 1-3) and several varieties of the species have been described. For this study, we have used Ross and Sims' characterization of *O. aurita* (1971): porose, undecorated valves with 2-3 external tubes on the central elevation. For the most part, we have avoided the use of varieties (except *O. aurita v. minima* --see below), but the level of genetic divergence demonstrated in our *O. aurita* isolates suggests further study may well reveal this clade to be a complex of lineages. Thus, while our identification is consistent with Lyngbye's original sketch and the definition of Ross and Sims, it must be admitted that a review of type material is necessary to have confidence that the type species is represented in what we call the *O. aurita* clade.

The uncertainty is highlighted by the fact that the strains we identified as *O. aurita v. minima* are in a clade separate from *O. aurita* var. *aurita*. De Toni (1894) describes two small varieties, *v. minor* and *v. minima*; our isolates do not match De Toni's striae count for *v. minor*, but do match the description of acuminate apices in *v. minima*. To our knowledge, these varieties have never been illustrated using SEM, and all other illustrations and light micrographs of this diminutive diatom are presented in girdle view. Our strains so identified are morphologically distinct from the *O. aurita* isolates—

the *O. aurita v. minima* strains have pseudoloculate valves (Figure 64f).Without looking at type material in the SEM, this identification should also be considered provisional.

These issues, along with the uncertain placement of *P. laevis* highlight the main reason our results need to be cautiously interpreted. Analysis of the diatom phylogeny has been sensitive to taxon sampling, perhaps even more than sampling of genes or method of optimization (Theriot et al. 2009, 2010, 2011). This explains why, for the most part, we are hesitant to begin a full-scale reclassification of the ocellate taxa. Only for *Trieres* and allies do we believe that there has been the necessary monographic work for apparently closely-related diatoms, including revision of fossil taxa (Komura 1999) which we can also confidently tie to our molecular exemplars through the similarity of morphological characters.

Pseudocellate taxa (Biddulphiaceae)

In contrast to the situation for the Eupodiscaceae, the molecular results do not support a monophyletic Biddulphiaceae. *Biddulphia* spp. form a clade which is sister to *Attheya* and several nodes away from the clades bearing other taxa often associated with the Biddulphiaceae such as *Trigonium*, *Eunotogramma*, *Isthmia* and *Terpsinoë*, which themselves are scattered across the Mediophyceae. Ross and Sims (1971) concept of a pseudocellate Biddulphiaceae was defined in part by possession of a straight valve mantle, but some *Biddulphia* species themselves have a deflected valve mantle characteristic of most Eupodiscaceae (Figures 62a, b). Other characters suggested by Ross and Sims (1971) were already highly variable within their Biddulphiaceae, such as internal ribs or ridges, which are present in *B. biddulphiana* (Figures 62a, h), *B. alternans* (J.W. Bailey) Van Heurck (Figures 62e, j) and *Terpsinoë musica* but absent in *Biddulphia*

reticulum (Ehrenberg) C.S. Boyer (Figures 62f, g, k, l) and *I. minima* (Figures 62c). The internal ribs also appear to be missing in *B. tridens* (Figures 62b, i); while there is a hyaline area on the interior of the valve where the valve face is depressed, this depression does not appear to extend into the interior space as it does in the ribs of the aforementioned taxa.

With so many other mediophycidean lineages represented solely by fossil taxa and unavailable for this study, it is difficult to draw conclusions about the relationships of the Biddulphiaceae and Eupodiscaeae to the rest of the diatom tree at this time. However, as mentioned previously, several authors have proposed a close relationship of the Biddulphiaceae and the Hemiaulaceae (Ross, Sims and Hasle 1977, Glezer 1979, Simonsen 1979). This proposal was based on the fact that Biddulphiaceae and Hemiaulaceae share the feature of strongly-elevated apices on their valves; *Hemiaulus* has hollow tubes at the apices while *Eucampia* and *Cerataulina* feature costate pore fields at the end of the elevations. Not only do we recover the Biddulphiaceae as polyphyletic, our analyses recover the Chaetoceraceae embedded in a paraphyletic Hemiaulaceae, so a sister group relationship between Biddulphiaceae and Hemiaulaceae to the exclusion of Chaetoceraceae and other taxa is rejected by the topological constraint test using per-site log-likelihoods (Table 5).

Smaller groups of former Biddulphiaceae, however, seem well supported by both molecular and morphological data. In particular, Ross and Sims' generic concept of *Biddulphia* holds up very well with molecular data. Isolates identified as *Biddulphia* in this data set (*B. alternans*, *B. biddulphiana*, *B. reticulum* and *B. tridens*) all have pseudocelli and similarly-occluded simple pores across the valve. Internal ribs, another character Ross and Sims associated with this genus, were only found in two of the taxa: the aforementioned *B. alternans* and *B. biddulphiana*. The authors mention *Triceratium*

reticulum Ehrenberg as likely belonging to this genus despite the lack of internal ribs, which is supported by the molecular data; this taxon bears the more appropriate name *B. reticulum* in our analysis and should not be confused with the ocellate/pseudoloculate *Biddulphia reticulata* Roper (see Sims and Ross 1990 for more detail on this taxonomic issue). However, at the moment we cannot suggest a synapomorphy for this taxon based on the presented data.

The genus *Biddulphiopsis* has been associated with the Biddulphiaceae (Round et al. 1990) based mostly on its taxonomic history, but *Biddulphiopsis titiana* (Grunow) von Stosch & Simonsen and *B. membranacea* (Cleve) von Stosch & Simonsen differ by the presence (*B. titiana*) or absence (*B. membranacea*) of pseudocelli (von Stosch and Simonsen 1984). In our analysis, *Isthmia minima* is embedded in *Biddulphiopsis*, sharing marginal rimoportulae and pseudocelli with its sister taxon, *B. titiana*. *I. minima* is on a long branch in our molecular tree, and also seems to be highly derived morphologically, having autapomorphic characters such as the presence of only a single pseudocellus per cell and deep, sac-like depressions on the valve mantle. The latter character, however, is unique to *I. minima* and is not found in other *Isthmia* species (Andresen 1995). Both genera require further molecular and morphological analysis.

That the freshwater taxa *Terpsinoë musica* and *Hydrosera sp*. form a clade is hardly surprising, considering the undulate rimoportula both taxa share as an apparent synapomorphy. Both taxa do feature apical pore fields which may not be considered "pseudocelli," as they are not elevated above the valve surface (Figure 62d).

The generic concept for *Trigonium* also held up strongly with molecular data, although that could well be an artifact for poor taxon sampling with only three strains on the analysis—all of them nominally agreeing with the admittedly vague species diagnosis of *T. formosum* (Figure 66) originally provided by Brightwell (described under the name

Triceratium striolatum Ehrenberg in 1853, then renamed *Triceratium formosum* Brightwell in 1856). Hendey (1971) and Ross and Sims (1971) both note that only slight difference in the valve perforation (density at the center of the valve and degree of velum convexity) exists between the few *Trigonium* species each examined and compared (*T. arcticum* Brightwell, *T. formosum* and *T. graeffeanum* [Witt] Hendey in Hendey; *T. arcticum*, *T. formosum* and *T. montereyi* Brightwell in Ross and Sims). Close examination by SEM of the type material for *Trigonium* species should be a priority moving forward. Additionally, the placement of *Trigonium* in the diatom phylogeny as a whole remains uncertain. We recover it as sister to *Chrysanthemodiscus* (with low statistical support). The latter lacks the loculate areolae and apices of *Trigonium*, and has a central pore field instead of pore fields on elevated apices.

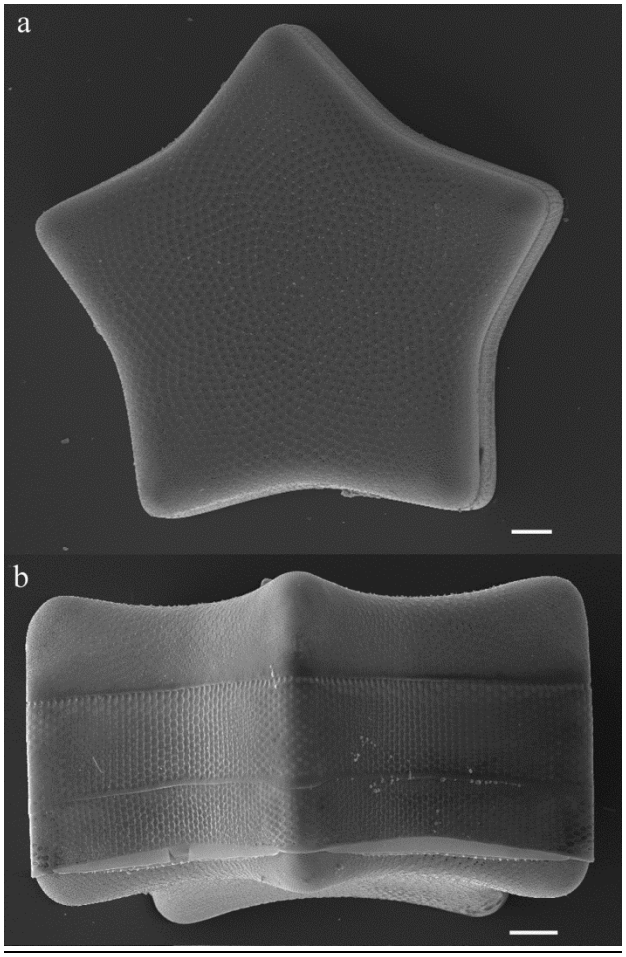


Figure 66. Scanning electron micrographs showing the pseudocellus-bearing, loculate-valved biddulphiacean taxa used in this study: *Trigonium formosum.* (a) *T. formosum* (HK258) shown in valve view. (b) *T. formosum* (HK258) shown in girdle view. Scale bars: 10 µm.

CONCLUSIONS

In a plea for improved illustration included in scientific texts, G. C. Wallich (1858) used *Triceratium* as an exemplar of a genus potentially undone by improper documentation of characters. We would argue that his prophecy is at least partly fulfilled. The phylogeny of the Biddulphiaceae and Eupodiscaceae appears to be a case where the

use of molecular characters does not so much "disagree" with morphological characters as they indicate uncertainty present in both types of data. On one hand, difficulties in interpretation of morphology, may lead to reliance on plesiomorphy to delineate taxa or to discount "minutiae" as synapomorphies in their own right. On the other hand, although much easier to interpret, molecular data suffers from higher propensity for convergent change and therefore perhaps higher risk to mislead phylogenetic inference. It is evident that more work is needed to resolve the uncertainties uncovered to date or identify conflicting inferences that withstand statistical scrutiny (as in lizards: Losos et al. 2012).

Ross and Sims (1971) focused primarily on the apical pore fields (ocellus vs. pseudocellus) as the primary character to distinguish the Biddulphiaceae and Eupodiscaceae, and suggested several other observed valvar characters to support this distinction. They also suggested (1973) that the pseudocellus gave rise to the eupodiscoid ocellus, making the pseudocellus plesiomorphic. Our molecular results, however, does not seem to support this interpretation in its entirety, although it supports the hypothesis that the eupodiscoid ocellus is synapomorphic. Thus, our analysis does not agree with Glezer's assertion that the eupodiscoid ocellus evolved several times from the pseudocellus (1979).

This work is an important step in laying the groundwork for future systematic studies of the Mediophyceae. With the most complete taxon sampling to date, the monophyly of the ocellate Eupodiscaceae seems well supported. Some groups within the Eupodiscaceae appear to be well supported (e.g., *Triceratium*). *Odontella*, however, is non-monophyletic. It is scattered among several molecular lineages each of which, in our preliminary analyses, are at least phenetically coherent if not supported by synapomorphy. Likewise the Biddulphiaceae, formerly diagnosed only by presumed plesiomorphy, appears to be non-monophyletic with groups scattered across the

Mediophyceae. Some clades of former Biddulphiaceae again appear to be phenetically coherent if not diagnosable by synapomorphy.

Despite our addition of many new mediophytes to molecular analysis, there are still many taxa which need to be sampled. The variety of phylogenetic placement of the pseudocellate lineages in this study underscores the massive undertaking that will be the cladistic, morphological analysis on the fossil and extant Mediophyceae. It will be difficult to select exemplars, as the "biddulphioid" state actually is present on many different branches of the diatom tree. Nevertheless, the agreement of the molecular data with a synapomorphic, eupodiscacean ocellus suggests that the situation is not as chaotic as it might seem.

NOMENCLATURAL AMENDMENTS

Based on the conservation of morphological and molecular characters discussed above, we recommend the following taxa be transferred to the genus *Trieres* gen. nov.:

Trieres M.P. Ashworth & E.C Theriot, gen. nov.

DESCRIPTION: Frustules with large pervalvar distance, 1-2 times the transapical distance. Valves alveolate, with finely-perforate vela arrayed in box-like sections and single, circular foramen on the valve interior. Valves elongate with tall ocelli at apices and offset, opposed rimoportula with conspicuous external tubes on the valve face. Valve margin edged with a wall-like projection running between the ocelli. Cells planktonic in habit.

ETYMOLOGY: From the Greek "*triērēs*," a type of ancient Greek sailing vessel which resembles the valves of this genus seen in girdle view.

TYPE SPECIES: *Trieres sinensis* (Greville) M.P. Ashworth & E.C. Theriot, illustrated in Williams 1988; pl. 25, fig. 3.

Trieres mobiliensis (J.W. Bailey) Ashworth and Theriot comb. nov.

Basionym: *Zygoceros mobiliensis* J.W. Bailey 1851, p. 40, pl. 2: figs 34-35

Homotypic Synonyms: *Biddulphia mobiliensis* (J.W. Bailey) Grunow 1882, (in Van Heurck), pl. 101: figs. 4-6

Odontella mobiliensis (J.W. Bailey) Grunow 1884, p. 58

Trieres regia (M. Schultze) M.P. Ashworth & E.C. Theriot comb. nov.

Basionym: *Denticella regia* M. Schultze 1859, p. 21, pl. 2: figs. 11, 12

Homotypic Synonyms: *Biddulphia regia* (M. Schultze) Ostenfeld 1908, p. 7, fig. 3

Odontella regia (M. Schultze) Simonsen 1974, p. 27

Trieres sinensis (Greville) M.P. Ashworth & E.C. Theriot comb. nov.

Basionym: *Biddulphia sinensis* Greville 1866, p. 81, pl. 9: fig. 16

Homotypic Synonyms: *Odontella sinensis* (Greville) Grunow 1884, p. 58

Chapter 3: Rock snot in the age of transcriptomes: Using a phylogenetic framework to identify genes involved in diatom extracellular polymeric substance-secretion pathways

INTRODUCTION

The diatoms are exceptional among the heterokont algae for their intricate silica-based shells and for their amazing species diversity, both in extant and fossil forms. Some estimates indicate that benthic microalgae (predominantly diatoms) may contribute just as much to the primary productivity of certain habitats as planktonic species (Wetzel 1964). We are coming to understand that the ecological importance of diatoms is not limited to primary productivity. Many diatoms produce extracellular polymeric substances (EPS), sometimes in copious quantities. These are vital components in algal and bacterial "biofilms" (an aggregation of microorganisms growing on a solid substrate) as potential carbon sources and help stabilize sediments for biological colonization (Smith & Underwood 2000, Thornton 2002). There are commercial reasons to study this issue as well, since diatom biofilms are major contributors to ship biofouling, which can significantly increase the friction on a ship's hull and therefore increase fuel consumption (Alberte et al. 1992). Diatoms are resistant to conventional toxic anti-biofouling hull coatings (Johnson et al. 1995). In addition to biofouling, new concerns over the invasive diatom species *Didymosphenia geminata* have brought diatom EPS production to the fore as this diatom can produce tough, macroscopic stalks of mucilage which have been reported to smother all other benthic life in the rivers of New Zealand in which it has been found (Kilroy et al 2008).

Diatoms produce and utilize EPS in a number of different forms, most of which are related to motility and adhesion. One clade of diatoms, the raphid pennates, use EPS

secreted through a slit (the raphe) in the siliceous shell to pull themselves along the substrate, though the exact method employed is still disputed (Rosowski 1980, Webster et al. 1985). Some planktonic centric diatoms secrete long chains of β-chitin, presumably to modify their position in the water column by increasing the cell's buoyancy (Durkin et al 2009).

Both centric and pennate diatoms use EPS-derived mucilage to form films, pads (both for cell-cell adhesion in colony formation and cell-surface adhesion), stalks and even macroscopic, complex tubes often mistaken for the thalli of brown algae (Drum 1969). Pad, film and stalk formation is seen in many lineages of both centric and pennate diatoms, and several lineages of pennate diatoms form tubes, demonstrated in the phylogenetic tree diagram in Figure 67. This begs the question—did these diatoms evolve these EPS morphotypes independently, or is there some ancestral pathway in the genomes of all diatoms, which is simply not expressed or even lost in many diatom lineages?

The first step is to identify some of the genes involved in the assembly and secretion of these EPS morphotypes. Efforts have been made to chemically identify the types of polysaccharides, proteins and other molecules used to create and modify diatom EPS, but beyond general classification of compounds (i.e. polysaccharides and proteins) there is significant variation in the compounds between species (Hoagland et al. 1993, Wustman et al. 1997, Wustman et al. 1998, Wang et al. 2000). Far fewer attempts have been made to unlock the genetic pathways involved in these processes, though a recent examination of the interaction between diatoms and bacterial components in biofilms has proposed several proteins involved in extracellular signaling for EPS production in diatoms (Bruckner et al 2011).

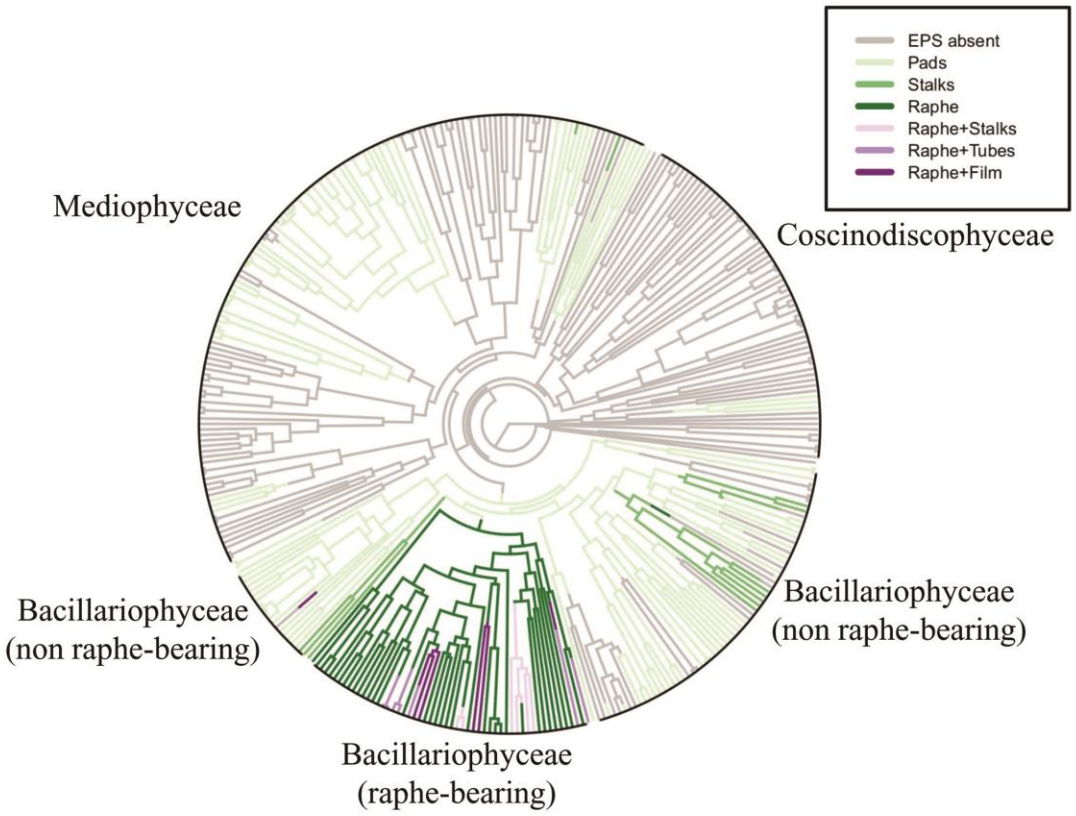


Figure 67. Cladogram generated from three genes of DNA sequence data (nuclear-encoded ribosomal small subunit and chloroplast encoded *rbc*L and *psb*C) from 350 diatom taxa. DNA sequences analyzed under maximum likelihood with RAxML 7.0.4. EPS morphotype characters exhibited by the taxa are mapped on the tree by color.

This study took a phylogenetic approach to the problem of EPS assembly and secretion. Assuming closely-related diatoms grown under identical conditions have similar transcriptomes, comparison of the transcriptome of an EPS stalk- or pad-producing diatom to a closely-related diatom that does not produce one of these EPS morphotypes, would suggest that the proteins involved with EPS assembly or secretion are differently-expressed. We do, however, have to make several assumptions about the

EPS process: the genes involved are orthologous, the process is not largely epigenetic and that the diatom is largely responsible for the process.

Several clades (Figure 67) are candidate taxa for this comparison. Two clades were chosen: the *Cyclophora* clade and *Thalassionema* clade (Figure 68), both within the araphid 'pennate' diatoms. In the *Thalassionema* clade, *Thalassionema frauenfeldii* secretes EPS pads at one apex of each cell to hold together planktonic colonies, while *Thalassionema* sp. 'BlueH2O' is solitary in culture, with no obvious EPS pads or stalks. In the *Cyclophora* clade, *Cyclophora tenuis* and *Lucanicum concatenatum* produce EPS pads between the apices of adjacent cells, while the closely-related diatom *Astrosyne radiata* does not produce any obvious EPS morphotype.

In comparing the expressed transcripts from these diatoms, we hope to elucidate some of the molecular and genetic machinery behind EPS assembly (carbohydrate modification, extracellular transport or packaging). This would be an important step in determining how and why this process has apparently evolved in so many different lineages of diatoms.

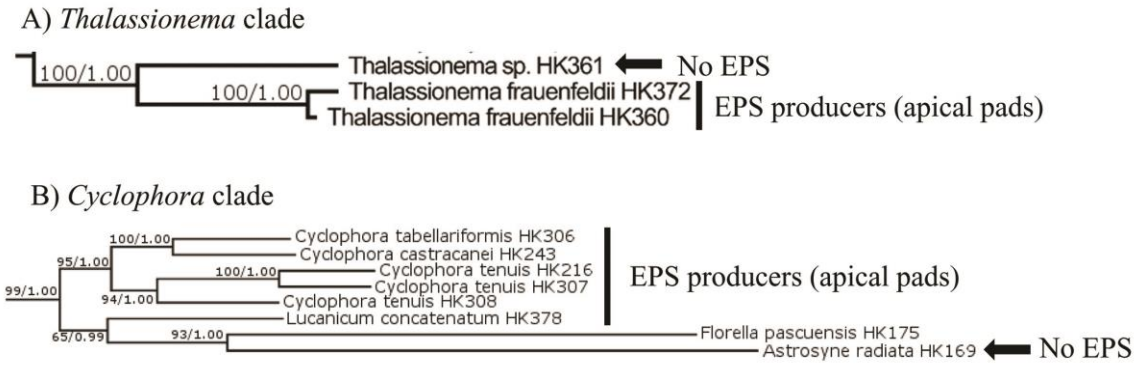


Figure 68. Portion of the tree in Figure 67 generated from 3 genes of DNA sequence data (nuclear-encoded ribosomal small subunit and chloroplast encoded *rbc*L and *psb*C) from 350 diatom taxa. DNA sequences analyzed under maximum likelihood with RAxML 7.0.4. These two clades show the taxa from which transcriptomes were generated for this study.

METHODS AND MATERIALS

RNA extraction and Sequencing

Cultures were grown in f/2 media (Guillard 1975) at either room temperature (20-22^{o}C) near a north-facing window (*Cyclophora tenuis*, *Thalassionema* spp.) or in a Percival model I-36LL incubation chamber at 27^{o}C (*Astrosyne radiata*, *Lucanicum concatenatum*), depending on the growth requirements of the culture. Culture details are outlined in Table 6. Cells were harvested in the exponential growth phase and disrupted with 1.0 mm glass beads in a Mini-Beadbeater (Biospec Products Inc., Bartlesville, OK, USA) for 45 seconds. RNA was extracted with a QIAGEN RNeasy Plant Minikit (QIAGEN Sciences, MD, USA) and treated with DNase I (Thermo Scientific, Wilmington, DE, USA) to degrade any co-extracted DNA. RNA concentration and quality was measured on a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA).

Illumina Hi-Seq 2000 paired-end sequencing was done by the Marine Microbial Eukaryote Transcriptome Project for *Astrosyne radiata* (MMETSP0418) and *Cyclophora tenuis* (MMETSP0397) and at the University of Texas Genomic Sequencing and Analysis Facility for *Astrosyne radiata*, *Lucanicum concatenatum* and both *Thalassionema* spp.

Taxon	Voucher ID (HK###)	Strain ID #	Collection Locality	Growth Temp.	EPS morphotype
Astrosyne radiata	HK169	ECT3697Astro; MMETSP0418	GabGab Beach, Guam	27°C	None
Cyclophora tenuis	HK307	ECT3854Cycten; MMETSP0397	Kahana State Park, HI	20-22°C	Apical pads
Lucanicum concatenatum	HK378	GU44AI-3Lucan	GabGab Beach, Guam	27°C	Apical pads
Thalassionema frauenfeldii	HK360	ECT3929ThalXL	Gulf of Mexico, TX	20-22°C	Apical pads
Thalassionema sp. 'BlueH2O'	HK361	ECT3929ThalBluH20	Gulf of Mexico, TX	20-22°C	None

Table 6. Taxonomic and culture information for the diatom strains used in this study. Voucher ID (HK###) refers to the Theriot Lab. Voucher images can be found on http://www.protistcentral.org/Project/get/project_id/79.

Transcriptome assembly and analysis

Paired-end reads were assembled with Trinity ver. r20130225 (Grabherr et al 2011) on the LONESTAR supercomputer at the Texas Advanced Computing Center (TACC). The assembled contigs were translated to amino acid sequences with Trinity's 'transdecoder' utility. BLASTP databases were made from each taxon's assembled contigs.

Serial BLASTP ver. 2.2.25 searches were conducted as outlined in Figure 69. In the case of the *Thalassionema* clade, contigs from EPS-producing taxon *Thalassionema*

frauenfeldii HK360 were searched against the contigs from the non EPS-producing taxon *Thalassionema* sp. 'BlueH2O' HK361, and the non-matching contigs were annotated using Blast2GO (Conesa et al 2005). For the *Cyclophora* clade, first the EPS-producing taxa *Cyclophora tenuis* and *Lucanicum concatenatum*) were BLASTed to each other and the contigs in common were sorted and annotated with Blast2GO. These contigs were then BLASTed against the contigs from the non EPS-producing *Astrosyne radiata*. The contigs that were present in the EPS-producing taxa but were not present in the non EPS-producing *Astrosyne* were then sorted out and annotated with Blast2GO.

The annotated contigs were then searched for several terms thought to be related to EPS synthesis, modification and secretion. In addition to the broad terms of 'polysaccharide,' 'oligosaccharide' and 'secretory,' terms were searched related to the chemical classification studies previously conducted on diatom EPS. Underwood et al. (2004) suggested the glucan pathway was involved in EPS secretion, while Wustman et al. (1997) found sulfated polysaccharides bound together by cation-mediated cross-links (such as Ca^{+2}) to be major constituents of EPS stalks.

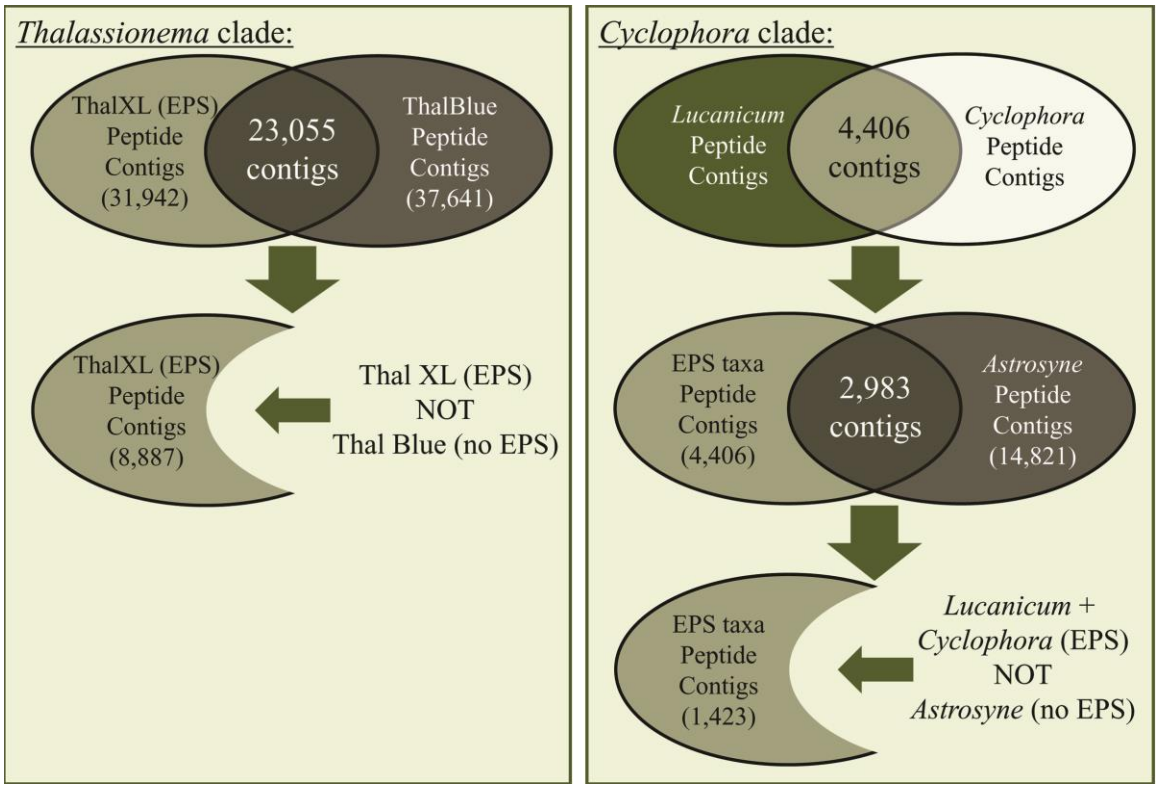


Figure 69. Graphical representation of the steps undertaken for the serial BLASTP searches conducted on the EPS-producing and non EPS-producing transcriptomes.

RESULTS

Assembly statistics for the transcriptomes are presented in Table 7. The number of assembled contigs varied among taxa, but especially among clades, with the *Thalassionema* clade having 10,000 more assembled contigs than any taxon in the *Cyclophora* clade. The *Cyclophora tenuis* assembly was also noteworthy in that it had the highest N50 value (821 bp) but the lowest maximum contig length (996 bp). The maximum contig length for *Lucanicum concatenatum* (1,072 bp) was comparable to *C. tenuis*, but the other taxa had much higher maxima.

Taxon	Number of reads (Million)	Contigs Assembled (**translated*)**	Max. Contig Length (bp) (**translated*)**	Contig N50 (bp)
Astrosyne radiata	66	60,831 **14,821***	7,804 **1,653***	503
Cyclophora tenuis	26	33,429 **21,819***	4,428 **996***	821
Lucanicum concatenatum	78	40,009 **12,875***	7,279 **1,072***	443
Thalassionema frauenfeldii	81	74,100 **31,942***	9,299 **1,859***	598
Thalassionema sp. 'BlueH2O'	80	83,391 **37,641***	10,086 **1,504***	692

Table 7. Statistics for the data assembled using Trinity. Minimum contig length was set at 200 bp for all taxa and therefore not reported here. The bold and starred numbers are amino acid counts while the normal text above is nucleotide counts.

Thalassionema Clade

Comparison of the peptide contigs of the *Thalassionema* spp. yielded 8,887 contigs that were present in the EPS-producing *Thalassionema frauenfeldii* that did not match contigs from the non EPS-producing *Thalassionema sp*. 'BlueH20.' Of those contigs, roughly a quarter resulted in a diatom sequence as the first BLASTP hit (Figure 70A). The largest functional groups (Figure 71A) seem to be ion binding (3.6%), heterocyclic compound binding (3.2%), organic cyclic compound binding (3.2%), transferase activity (2.7%), hydrolase activity (2.6%) and protein binding (2.4%). Of the 8,887 contigs distinct to *T. frauenfeldii*, 6,781 had no matches in the annotation databases beyond 'predicted protein' or 'hypothetical protein.'

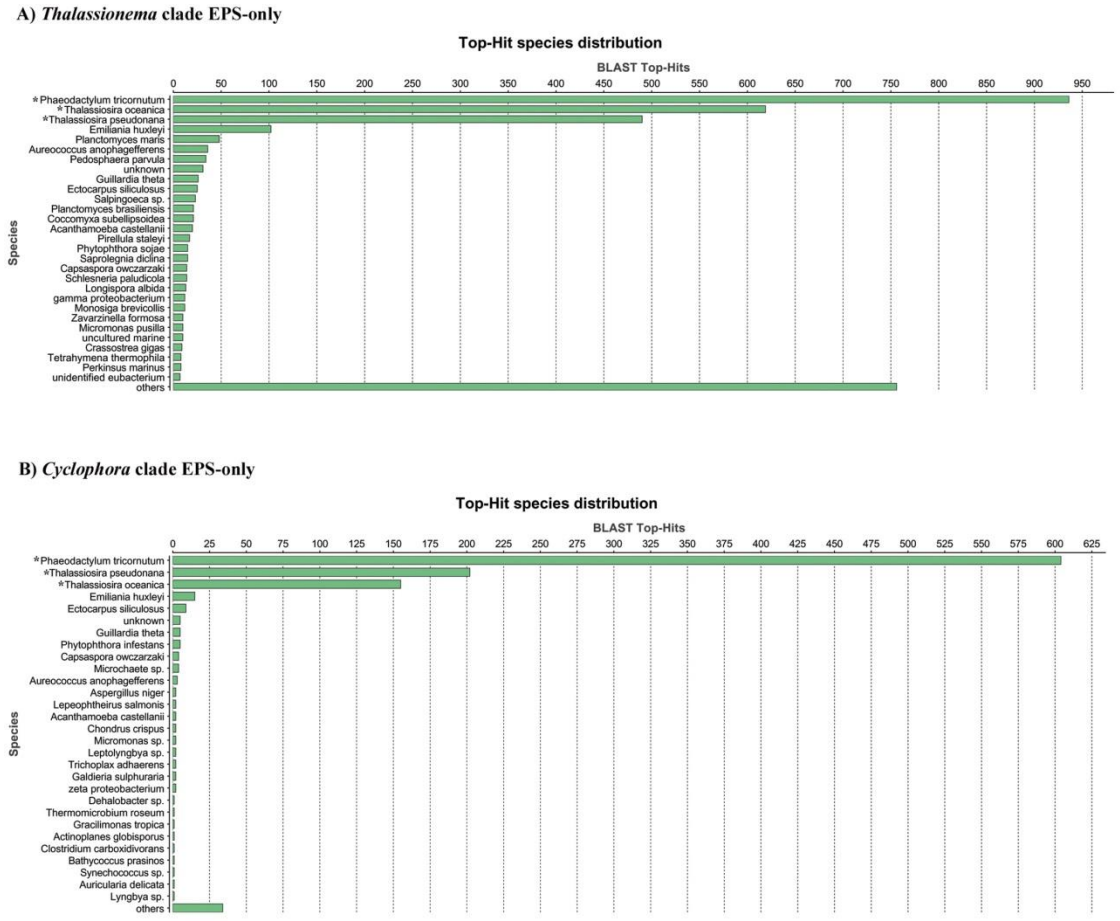


Figure 70. Graphs showing species from which the top BLASTP hit for each contig is found. A. Top hits from the EPS-producing *Thalassionema frauenfeldii* transcriptome contigs that did not match contigs from non EPS-producing *Thalassionema* sp. "BlueH20." B. Top hits from the contigs that match in the transcriptomes of the EPS-producing *Cyclophora tenuis* and *Lucanicum concatenatum*, which did not match contigs from the non EPS-producing *Astrosyne radiata*. Diatom taxa are preceded by an asterisk.

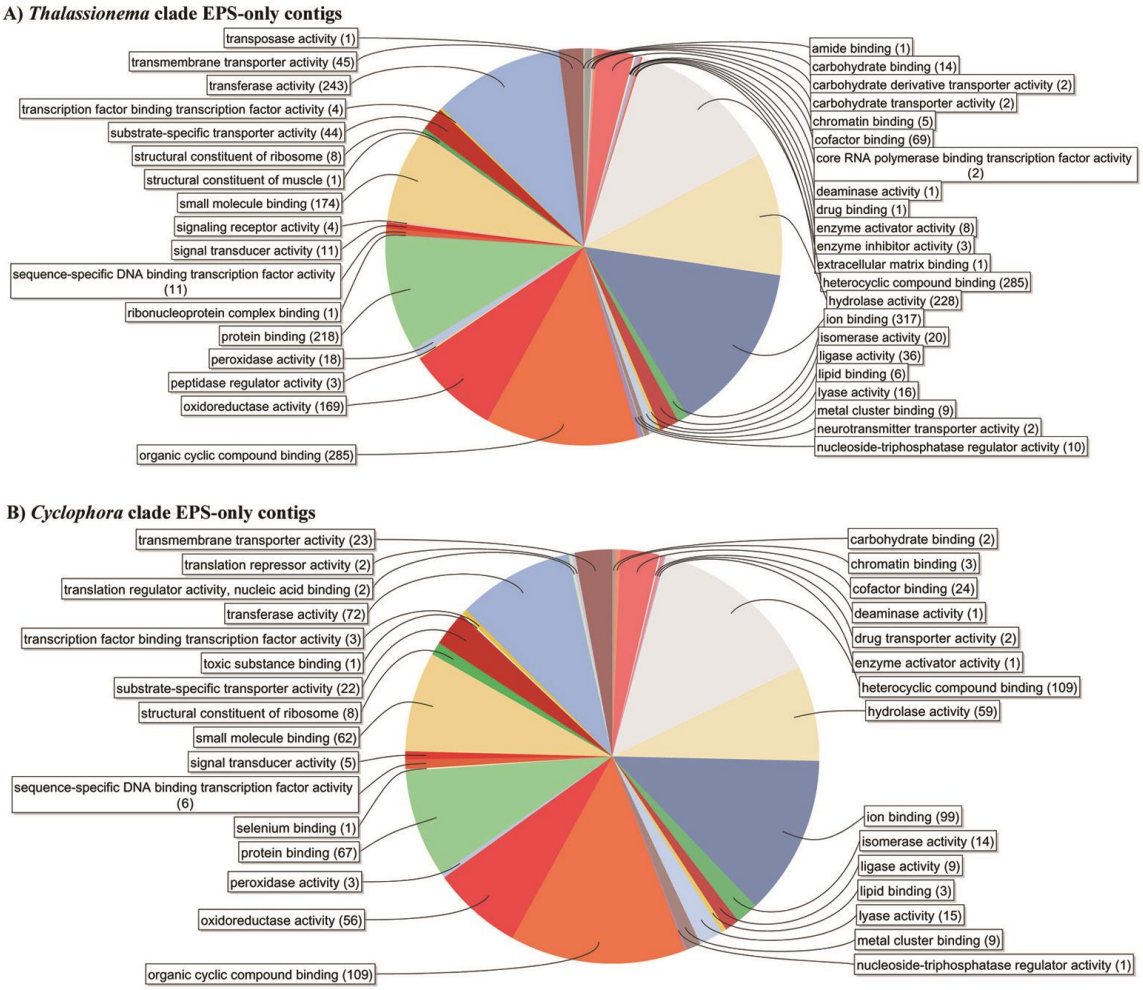


Figure 71. Graphs showing the breakdown of gene ontology (GO) function annotation for the transcriptome data indicated. A. Contigs from the EPS-producing *Thalassionema frauenfeldii* transcriptome that did not match contigs from non EPS-producing *Thalassionema* sp. "BlueH20." B. Contigs that match in the transcriptomes of the EPS-producing *Cyclophora tenuis* and *Lucanicum concatenatum,* which did not match contigs from the non EPS-producing *Astrosyne radiata*.

When the annotations were sorted for 'polysaccharide,' 'oligosaccharide,' 'carbohydrate' or 'glucan,' fifteen hits resulted (Table 8). Of those fifteen, eight contigs involved a putative function regarding glucan modification. Two were polysaccharide

deacetylases, four were polysaccharide/oligosaccharide biosynthesis or modification transferases. One was a dehydrorhamnose reductase, which has been linked to extracellular polysaccharide biosynthetic processes. In total, six of the 'polysaccharide'-related contigs had BLASTP top hits that were not diatoms: two of the endo-glucanase contigs were more similar to the haptophyte *Emiliana huxleyi*, the dehydrorhamnose annotation came from *Ostreococcus* (a prasinophyte), the deacetylases were annotated from bacterial sequences and the galactosyl transferase was from an louse sequence.

Screening of the non-matching *Thalassionema* contigs for 'secretory' or 'secretion,' yielded thirteen hits (Table 8). Five of the thirteen contigs were annotated to bacterial transmembrane facilitator proteins, and three were bacterial AMP-dependent synthetases involved in transmembrane transport. Three were annotated as being involved in extracellular ligand-gated ion channels, one was a HlyD-family secretion protein and one was a protein involved in type II protein secretion. There was also tartrate-resistant acid phosphatase type 5-like protein, which has been linked to the organization of extracellular matrices. Only the contigs annotated as ligand-gated ion channels had diatom sequences as the top BLASTP hits.

When screened for 'sulfo-' or 'sulfa-,' nine contigs were annotated to a form of sulfate modification. Five contigs were annotated as transferases, including two that were annotated as carbohydrate sulfotransferases, though the sequence identity was very low (<30%) and the annotated sequences were from vertebrates. Two contigs were identified as being involved in transmembrane sulfur transport, and one of the two contigs

identified as arylsulfatases had a gene ontology term of "proteinaneous extracellular matrix."

The contigs were also searched for calcium ion transport, as these ions have been suggested as being involved in the cross-linking of the polysaccharide polymers. Four contigs were annotated as having similarity to known calcium ion binding domains (Table 8). Two of those contigs had gene ontology terms suggesting they might be involved in carbohydrate binding, however, the source of those annotations was from a cyanobacterium rather than a diatom.

Cyclophora Clade

In the *Cyclophora* clade, there were 1,423 contigs shared with the EPS-producing *C. tenuis* and *L. concatenatum,* to which BLAST did not match any non EPS-producing *Astrosyne* contigs. Approximately two-thirds of those contigs yielded a diatom sequence as the first BLASTP hit (Figure 70B). The largest functional groups (Figure 71B) seem to be heterocyclic compound binding (7.7%), organic cyclic compound binding (7.7%), ion binding (7%), transferase activity (5.1%), protein binding (4.7%), hydrolase activity (4.1%) and small molecule binding (4.4%). None of the contigs showed any annotation related to 'secretory' or 'secretion' terms, and 337 contigs had no annotation at beyond 'predicted protein' or 'hypothetical protein.'

Screening the contigs for the terms 'polysaccharide,' 'oligosaccharide,' 'carbohydrate' or 'glucan,' recovered 19 hits (Table 9). Six of these were annotated to some form of glucan modification. Two contigs were identified as being involved in oligonucleotide/oligopolysaccharide binding, one was a polysaccharide deacetylase, and

one was a polysaccharide biosynthesis protein—gdp-l-fucose synthetase. Interestingly, the polysaccharide deacetylase was similar in amino acid identity (71% identity to *C. tenuis* and 68% identity to *L. concatenatum*) to a chitooligosaccharide deacetylase identified in *Cyclotella nana*; *C. nana* is known to secrete threads of β-chitin. Also of note was a contig that was annotated as a galactosyl transferase-like protein, as sulfated galactosyl has been found to be a component in some diatom EPS stalks (Wustman et al 1997). Another contig was annotated as being involved with carbohydrate binding and ER membrane involvement, perhaps part of intracellular transport of carbohydrates. Fourteen of these annotated contigs had diatom genomes as their top BLASTP hits, and four were from other stramenopiles (*Emiliana huxleyi* and *Phytophthora* spp.).

Only three contig annotations contained 'sulfa-' or 'sulfo-' terms. All three were annotated as sulfotransferase proteins (one from the slime mold *Dictyostelium* and two bacterial) with diatom 'hypothetical' or 'predicted' proteins as their top BLASTP hits. None of the annotated sequence identities were above 50%, and the sequence identity between the similar *C. tenuis* and *L. concatenatum* sequences was also comparatively low (below 60% identity on two of the three annotated contigs).

Thalassionema Clade/Cyclophora Clade Overlap

Few genes with EPS-related annotations can be found in both sets of contigs from EPS- producing taxa between the two clades (Table 10). Among the 'polysaccharide/carbohydrate' contigs, only endo-1,3-β-glucanase (ThXL-carb05, CyLu-carb04) annotations occurred independently in both clades. Comparing both of those annotations against the other clade in BLAST did yield hits, but with low sequence

identity (≤50%). Sulfotransferase annotations could also be found in the contigs from both clades (ThXL-sulf03, CyLu-sulf01, CyLu-sulf02, CyLu-sulf03). These annotations had low sequence identity hits when compared against the contigs of the other clade with BLAST (<43%). There were no secretory annotations found among the *Cyclophora* clade contigs, compared to the thirteen found in the *Thalassionema* clade, and while a putative intracellular carbohydrate transport protein was annotated from the *Cyclophora* clade contigs (CyLu-carb09), none were annotated from the *Thalassionema* clade.

Sequence identity values between the annotations from each clade and their top BLAST hit from the other clade tended to be low. Simply by the best sequence identity scores (close to or above 50%), the most promising candidates of conserved genes involved in EPS across the two clades are mostly from the 'polysaccharide/carbohydrate' annotations: CyLu-carb01, CyLu-carb02, CyLu-carb08, mannose-6-phosphate isomerase, sedoheptulose-bisphosphotase, ThXL-carb05, ThXL-carb06 and ThXL-carb07. There was one candidate each from the 'sulfa-/sulfo-' (ThXL-sulf04) and 'secretion' (ThXL-AMPsyn2) annotation groups.

DISCUSSION

There is still a level of uncertainty regarding inferred functionality of proteins based on sequence similarity. Though there is no agreed-upon threshold for the minimum percentage of sequence identity needed to infer protein function from an annotation, several studies cite a 40-60% range (Tian and Skolnick 2003, Fontana et al 2009), while others suggest functionality declines sharply below 70% similarity (Rost 2002). Using that 40% threshold to the sequence identity between the two clades limits the putative

EPS-related annotations to eight: CyLu-carb02, CyLu-carb08, mannose-6-phosphate isomerase, sedoheptulose-bisphosphotase, ThXL-carb06, ThXL-carb07, ThXL-sulf04, ThXL-AMPsyn2. Of those, only four have annotations to known proteins above the 40% identity threshold: mannose-6-phosphate isomerase, sedoheptulose-bisphosphotase, ThXL-sulf04, ThXL-AMPsyn2. Based on the entry in the INTERPRO protein database, mannose-6-phosphate isomerases might be involved in secretory targeting, so this could be a valid candidate as an EPS-related marker. Sedoheptulose bisphosphotase is a curious annotation—this enzyme is strongly associated with the Calvin Cycle (Raines et al 1999), so we would expect it to be conserved even in the non EPS-producing taxa. It is possible this particular contig codes for a similar protein, or the RNA sequencing just missed this transcript in *Astrosyne* and *Thalassionema* sp. 'BlueH20.' Due to the large number of reads sequenced (Table 7), the latter seems unlikely.

The ThXL-sulf04 annotation to arylsulfatase 'e' (ARSE) is intriguing. This enzyme is involved in sulfuric ester hydrolysis, and research on human ARSE has suggested that this particular form is associated with the Golgi apparatus (Daniele et al 1998), which suggests a role in the secretory pathways. While most arylsulfatase characterization has been conducted on mammals, at least one arylsulfatase has been identified in the green alga *Chlamydomonas reinhardtii* (de Hostos et al 1988). While sulfated esters have been detected in the EPS stalks of the diatoms *Achnanthes longipes* and *Cymbella mexicana* (Wustman et al. 1997), it is unclear whether those esters are an end-product of EPS secretion or some form of transitional product. The presence of so

many sulfated carbohydrates in the diatom EPS from the aforementioned Wustman survey suggests they have some importance.

ThXL-AMPsyn2 was annotated as a bacterial AMP-dependent synthetase, which is a broad characterization. There are almost certainly many transmembrane proteins involved in the secretion of EPS, but only this one matched contigs in all three EPS-producing taxa at an identity above the 40% threshold. We cannot completely rule out bacterial contamination in the transcriptome; while the library preparation for Illumina sequencing uses the poly-A tail to enrich the final RNA for eukaryotic transcripts, these cultures were not axenic. In fact, there is evidence that bacterial interactions do play some role in diatom EPS secretion (Bruckner et al. 2008, 2011), but the nature of that interaction is still unknown. At this time, we cannot determine the provenance of ThXL-AMPsyn2 as being from the diatom genome or an associated bacterial genome.

The molecular pathways involved in polysaccharide assembly, modification and secretion in diatoms is complex and largely unexplored. Though this survey has identified several genetic markers for EPS biosynthesis and secretion, almost 70% of the contigs found in the EPS-producing taxa could not be annotated. Still, finding the first few molecular markers putatively involved in the EPS process is an important step to guide further investigations. Once we understand some of the molecular machinery involved in the construction and maintenance of diatom EPS, we can begin to investigate how this machinery might be manipulated to control biofouling.

Polysaccharide/ Carbohydrate Metabolism						
Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Hypothetical protein 'ThXL-carb01'	ThalXL m.13016	ThalXL comp109288_c0_seq2	Hypothetical protein (*Emiliana* 38% identity)	Cycten m.25513 (22% identity)	Hydrolase activity	Endo-1,3:1,4-β-glucanase (*Phytophthora* 32% identity) Match in *Astrosyne*
Hypothetical protein 'ThXL-carb02'	ThalXL m.13017	ThalXL comp109288_c0_seq3	Hypothetical protein (*Emiliana* 40% identity)	Cycten m.25513 (22% identity)	Hydrolase activity	Endo-1,3:1,4-β-glucanase (*Phytophthora* 33% identity) Predicted protein (*Phaeodactylum* 37% identity) Match in *Astrosyne*
β-glucan elicitor receptor	ThalXL m. 21	ThalXL comp12782_c0_seq1	β-glucan elicitor receptor (*Phaeodactylum* 44% identity)	Cycten m.21378 (29% identity)	Hydrolase activity, acting on glycosyl bonds; Cell wall macromolecule catabolic process	Hypothetical protein (*Cyclotella nana* 44% identity) Match in *Astrosyne*
Predicted protein 'ThXL-carb03'	ThalXL m. 20133	ThalXL comp112691_c0_seq1	Predicted protein (*Cyclotella nana* 50% identity)	Cycten m.21378 (31% identity)	Hydrolase activity, acting on glycosyl bonds; Cell wall macromolecule catabolic process	β -glucan elicitor receptor (*Phaeodactylum* 39% identity) Match in *Astrosyne*

Table 8. Results of the BLASTP search of the contigs in common between the EPS-producing taxon *Thalassionema frauenfeldii* ("ThalXL") that were not a match to contigs from non EPS-producing taxon *Thalassionema* sp. "BlueH20."

Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Predicted protein 'ThXL-carb04'	ThalXL m. 20136	ThalXL comp112691_c0_seq2	Predicted protein (*Cyclotella nana* 60% identity)	Cycten m.21378 (31% identity)	Hydrolase activity, hydrolyzing O-glycosyl compounds	β -glucan elicitor receptor (*Phaeodactylum* 48% identity) Match in *Astrosyne*
Hypothetical protein 'ThXL-carb05'	ThalXL m.20780	ThalXL comp112914_c0_seq2	Hypothetical protein (*Thalassiosira oceanica* 38% identity)	Cycten m.11854 (50% identity) Lucan m.15729 (33% identity)	Hydrolase activity	Endo-1,3:1,4-β-glucanase (*Phytophthora* 30% identity)
Hypothetical protein 'ThXL-carb06'	ThalXL m.23105	ThalXL comp113646_c0_seq3	Hypothetical protein (*Thalassiosira oceanica* 59% identity)	Cycten m.15960 (45% identity) Lucan m.18574 (43% identity)	Hydrolase activity, hydrolyzing O-glycosyl compounds	Hypothetical protein (*Cyclotella nana* 57% identity) GPI-anchored β-glucanase (*Acanthamoeba* 32% identity)
Hypothetical protein 'ThXL-carb07'	ThalXL m.23106	ThalXL comp113646_c0_seq4	Hypothetical protein (*Thalassiosira oceanica* 59% identity)	Cycten m.15960 (45% identity) Lucan m.18574 (43% identity)	Hydrolase activity, hydrolyzing O-glycosyl compounds	Hypothetical protein (*Cyclotella nana* 57% identity) GPI-anchored β-glucanase (*Acanthamoeba* 32% identity)
3-β-galactosyl transferase-1	ThalXL m.7862	ThalXL comp104177_c0_seq1	3-β-galactosyl transferase-1 (insect 36% identity)	Cycten m.5964 (33% identity)	Receptor activity; Glucan endo-1,4-β-glucanase activity; Cell wall macromolecule catabolic process	Predicted protein (*Phaeodactylum* 30% identity) Match in *Astrosyne*

Table 8 continued.

Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Ddtp-4-deydrorhamnse reductase	ThalXL m.21175	ThalXL comp113057_c0_seq1	Ddtp-4-deydrorhamnse reductase (*Ostreococcus* 40% identity)	Cycten m.5729 (25% identity)	Ddtp-4-deydrorhamnse reductase; Extracellular polysaccharide biosynthetic process	Hypothetical protein (*Thalassiosira oceanica* 34% identity) Match in *Astrosyne*
Predicted protein 'ThXL-carb08'	ThalXL m.6748	ThalXL comp102370_c0_seq1	Predicted protein (*Cyclotella nana* 40% identity)	Cycten m.6885 (31% identity)	Transferase activity; Transferring glycosyl groups	Mannosyltransferose (bacteria; 29% identity) Match in *Astrosyne*
Polysaccharide deacetylase 'ThXL-deac1'	ThalXL m.8892	ThalXL comp105533_c0_seq1	Dehydrogenase (bacteria 40% identity)	Cycten m.16181 (38% identity)	Hydrolase activity; Carbohydrate metabolic process	Match in *Astrosyne*
Polysaccharide deacetylase 'ThXL-deac2'	ThalXL m.8893	ThalXL comp105533_c0_seq3	Dehydrogenase (bacteria 40% identity)		Hydrolase activity; Carbohydrate metabolic process	
Glycosyl transferase	ThalXL m.10195	ThalXL comp107005_c0_seq2	Glycosyl transferase (*Phaeodactylum* 72% identity)	Cycten m.3512 (45% identity) Lucan m.10195 (37% identity)	1,3-β-D-glucan synthase complex	*Thalassiosira oceanica* hypothetical protein (64% identity)
N-acetylglucosaminyl transferase	ThalXL m.11275	ThalXL comp108035_c0_seq3	N-acetylglucosaminyl transferase (*Phaeodactylum* 52% identity)	Cycten m.7642 (53% identity)	Transferase activity; Transfering glycosyl groups; Oligosaccharide biosynthesis process	Match in *Astrosyne*

Table 8 continued.

Sulfate Modification						
Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Predicted protein 'ThXL-sulf01'	ThalXL m.6981	ThalXL comp102713_c0_seq1	Predicted protein (*Phaeodactylum* 65% identity)	Cycten m.20099 (61% identity) Lucan m.17775 (61% identity)	Sulfate transmembrane transporter activity; Sulfate transport	Sulfate transporter (Vascular plant 57% identity) Match in *Astrosyne*
Predicted protein 'ThXL-sulf02'	ThalXL m.6984	ThalXL comp102713_c0_seq4	Predicted protein (*Phaeodactylum* 64% identity)	Cycten m.20099 (60% identity) Lucan m.17775 (60% identity)	Sulfate transmembrane transporter activity; Sulfate transport	Sulfate transporter (Vascular plant 58% identity) Match in *Astrosyne*
Hypothetical protein 'ThXL-sulf03'	ThalXL m.10698	ThalXL comp107515_c0_seq1	Hypothetical protein (*Emiliana* 36% identity)	Cycten m.19595 (40% identity) Lucan m.5139 (31% identity)	Sulfotransferase activity	Vascular plant sulfotransferase domain (32% identity)
Sulfatotransferase domain	ThalXL m.24674	ThalXL comp114068_c2_seq4	Sulfatotransferase domain (bacteria 35% identity)	Cycten m.12307 (31% identity) Lucan m.7442 (34% identity)	Transferase activity	
Hypothetical protein 'ThXL-sulf04'	ThalXL m.15660	ThalXL comp110825_c0_seq2	Hypothetical protein (*Thalassiosira oceanica* 53% identity)	Cycten m.14285 (43% identity) Lucan m.8191 (54% identity)	Catalytic activity; Arylsulfatase activity; Sulfuric ester hydrolase activity; Hydrolase activity; Proteinaceous extracellular matrix	Arylsulfatase e (Orca 45% identity)
Hypothetical protein 'ThXL-sulf05'	ThalXL m.15662	ThalXL comp110825_c0_seq4	Hypothetical protein (*Thalassiosira oceanica* 54% identity)	Cycten m.14285 (40% identity)	Catalytic activity	Arylsulfatase e (urchin 45% identity) Match in *Astrosyne*

Table 8 continued.

Putative Gene Name	Protein Contig ID	Nucleotide Contig ID	Top BLASTP Hit	Similar Contigs	GO terms	Additional BLASTP Hits
Hypothetical protein 'ThXL-sulf06'	ThalXL m.16261	ThalXL comp111141_c0_seq1	Hypothetical protein (*Thalassiosira oceanica* 59% identity)	Cycten m.13335 (28% identity)	Transferase activity	Rhodanase-related sulfur transferase (bacterial 28% identity) Found in *Astrosyne*
Hypothetical protein 'ThXL-sulf07'	ThalXL m.20096	ThalXL comp112674_c0_seq1	Hypothetical protein (*Thalassiosira oceanica* 30% identity)	Cycten m.22115 (28% identity)	Sulfotransferase activity	Carbohydrate sulfotransferase 15 (fish 23% identity) Match in *Astrosyne*
Hypothetical protein 'ThXL-sulf08'	ThalXL m.39570	ThalXL comp116425_c2_seq9	Hypothetical protein (*Thalassiosira oceanica* 32% identity)	Cycten m.22115 (28% identity)	Sulfotransferase activity	Carbohydrate sulfotransferase 15 (mammal 25% identity) Match in *Astrosyne*
Secretion/ Transmembrane Transport						
Hypothetical protein 'ThXL-sec01'	ThalXL m.31931	ThalXL comp115489_c1_seq1	Hypothetical protein (*Thalassiosira oceanica* 57% identity)	Cycten m.5912 (24% identity) Lucan m.5253 (40% identity)	Extracellular ligand-gated ion channel activity	Acetylcholine receptor (tunicate 26% identity)
Hypothetical protein 'ThXL-sec02'	ThalXL m.31933	ThalXL comp115489_c1_seq3	Hypothetical protein (*Thalassiosira oceanica* 60% identity)	Cycten m.5912 (24% identity) Lucan m.5253 (40% identity)	Extracellular ligand-gated ion channel activity	Chloride channel (roundworm 26% identity)
Hypothetical protein 'ThXL-sec03'	ThalXL m.31938	ThalXL comp115489_c1_seq9	Hypothetical protein (*Thalassiosira oceanica* 49% identity)	Cycten m.5912 (29% identity) Lucan m.5254 (40% identity)	Extracellular ligand-gated ion channel activity; ion transport	Amniobutyric acid receptor (flatworm 29% identity)

Table 8 continued.

Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Transmambrane facilitator protein	ThalXL m.24457	ThalXL comp114009_c0_seq1	Transmembrane facilitator protein (bacteria 50% identity)	Cycten m.7014 (21% identity) Lucan m.10903 (23% identity)	Transmembrane transport; Integral to membrane	Hypothetical protein (bacteria 48% identity)
Hypothetical protein 'ThXL-sec04'	ThalXL m.24460	ThalXL comp114009_c0_seq3	Hypothetical protein (bacteria 47% identity)	Cycten m.24248 (27% identity)	Transmembrane transport; Integral to membrane	Transmembrane facilitator protein (bacteria 44% identity) Match in *Astrosyne*
Hypothetical protein 'ThXL-sec05'	ThalXL m.24464	ThalXL comp114009_c0_seq6	Hypothetical protein (bacteria 56% identity)	Cycten m.24248 (29% identity)	Transmembrane transport; Integral to membrane	Transmembrane facilitator protein (bacteria 47% identity) Match in *Astrosyne*
Hypothetical protein 'ThXL-sec06'	ThalXL m.24466	ThalXL comp114009_c0_seq8	Hypothetical protein (bacteria 52% identity)	Cycten m.24248 (29% identity)	Transmembrane transport; Integral to membrane	Transmembrane facilitator protein (bacteria 49% identity) Match in *Astrosyne*
Hypothetical protein 'ThXL-sec07'	ThalXL m.24468	ThalXL comp114009_c0_seq11	Hypothetical protein (bacteria 54% identity)	Cycten m.26449 (33% identity) Lucan m.6128 (28% identity)	Transmembrane transport; Integral to membrane	Transmembrane facilitator protein (bacteria 51% identity)
AMP-dependent synthetase 'ThXL-AMPsyn1'	ThalXL m.30922	ThalXL comp115302_c1_seq1	AMP-dependent synthetase (bacteria 52% identity)	Cycten m.4614	Transferase activity; Transmembrane transport; Integral to membrane	Match in *Astrosyne*

Table 8 continued.

Putative Gene Name	Protein Contig ID	Nucleotide Contig ID	Top BLASTP Hit	Similar Contigs	GO terms	Additional BLASTP Hits
AMP-dependent synthetase 'ThXL-AMPsyn2'	ThalXL m.30924	ThalXL comp115302_c1_seq2	AMP-dependent synthetase (bacteria 56% identity)	Cycten m.10738 (45% identity) Lucan m.2877 (46% identity)	Transferase activity; Transfering acyl groups; Transmembrane transport; Integral to membrane	
AMP-dependent synthetase 'ThXL-AMPsyn3'	ThalXL m.30926	ThalXL comp115302_c1_seq3	AMP-dependent synthetase (bacteria 56% identity)	Cycten m.4614	Transferase activity; Transfering acyl groups; Transmembrane transport; Integral to membrane	Match in *Astrosyne*
Type II secretion system protein	ThalXL m.1842	ThalXL comp82756_c0_seq1	Type II secretion system protein (bacteria 71% identity)	Cycten m.19516 (27% identity)	Type II protein secretion system complex; Protein transporter activity	Match in *Astrosyne*
Hypothetical protein 'ThXL-sec08'	ThalXL m.4029	ThalXL comp95652_c0_seq1	Hypothetical protein (bacteria 51% identity)	Cycten m.6437 (34% identity) Lucan m. 17834 (29% identity)	Transport; HlyD family secretion protein	RND family efflux transporter (bacteria 50% identity)
Ca^{+2} ion binding						
Predicted protein 'ThXL-Ca01'	ThalXL m.23469	ThalXL comp113735_c0_seq1	Predicted protein (*Cyclotella nana* 50% identity)	Cycten m.14879 (29% identity)	Secreted protein rich in cysteine Ca binding region; Calcium ion binding	Hypothetical protein (Thalassiosira oceanica 40% identity) Match in *Astrosyne*
Hypothetical protein 'ThXL-Ca02'	ThalXL m.28267	ThalXL comp114818_c0_seq4	Allergen V5/Tpx-1-like protein (cyanobacteria 51%)		Calcium ion binding Carboxypeptidase activity; Carbohydrate binding	Hypothetical protein (Cyanobacteria 48% identity) Match in *Astrosyne*

Table 8 continued.

Putative Gene Name	Protein Contig ID	Nucleotide Contig ID	Top BLASTP Hit	Similar Contigs	GO terms	Additional BLASTP Hits
Hypothetical protein 'ThXL-Ca03'	ThalXL m.28272	ThalXL comp114818_c0_seq9	Allergen V5/Tpx-1-like protein (cyanobacteria 55%)		Calcium ion binding; Carboxypeptidase activity; Carbohydrate binding	Hypothetical protein (Cyanobacteria 49% identity) Match in *Astrosyne*
Predicted protein 'ThXL-Ca04'	ThalXL m.23472	ThalXL comp113735_c0_seq4	Predicted protein (*Cyclotella nana* 50% identity)	Cycten m.14879 (29% identity)	Secreted protein rich in cysteine Ca binding region; Calcium ion binding	Hypothetical protein (Thalassiosira oceanica 40% identity) Match in *Astrosyne*
Extracellular matrix organization						
Tartarate-resistant acid phosphatase type 5-like	ThalXL m.40117	ThalXL comp116470_c0_seq10	Tartarate-resistant acid phosphatase type 5-like (urchin 46% identity)	Cycten m.10481 (36% identity)	Extracellular matrix organization; Interspecies interaction between organisms; Response to oxidative stress	Hypothetical protein (*Cyclotella nana* 43% identity) Match in *Astrosyne*

Table 8 continued.

Polysaccharide/ Carbohydrate Metabolism						
Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Mannose-6-phosphate isomerase	Cycten m.25185 Lucan m.15460 (74% identity)	Cycten comp56709_c0_seq1 Lucan comp128351_c0_seq1	Mannose-6-phosphate isomerase (*Phaeodactylum* 69% identity)	ThalXL m.11425 (59% identity)	Mannose-6-phosphate isomerase activity; GDP-mannose biosynthetic process	Hypothetical protein (*Cyclotella nana* 66% identity; *Thalassiosira oceanica* 64% identity)
Hypothetical protein 'CyLu-carb00'	Cycten m.23226 Lucan m.6944 (89% identity)	Cycten comp53449_c0_seq1 Lucan comp122349_c1_seq2	Hypothetical protein (*Phaeodactylum* 86% identity)	ThalXL m.13587 (38% identity)	Polysaccharide biosynthesis process; GDP-1-fucose synthetase; Oxidoreductase activity	GDP-fucose synthetase (*Cyclotella nana* 81% identity)
Chitooligosaccharide deacetylase	Cycten m.3083 Lucan m.13276 (80% identity)	Cycten comp26204_c0_seq1 Lucan comp127515_c0_seq3	Chitooligosaccharide deacetylase (*Cyctlotella nana* 71% identity)	ThalXL m.26845 (39% identity)	Urate catabolism protein; Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds; Carbohydrate metabolic process	Hypothetical protein (*Thalassiosira oceanica* 71% identity) Predicted protein (*Phaeodactylum* 59% identity)

Table 9. Results of the BLASTP search of the contigs in common between the EPS-producing taxa *Cyclophora tenuis* ("Cycten") and *Lucanicum concatenatum* ("Lucan") that were not a match to contigs from non EPS-producing taxon *Astrosyne radiata*.

Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Predicted protein 'CyLu-carb01'	Cycten m.3403 Lucan m.1913 (39% identity)	Cycten comp26479_c0_seq1 Lucan comp105543_c0_seq1	Predicted protein (*Phaeodactylum* 45% identity)	ThalXL m.3314 (69% identity)	Hydrolase activity	β-1,6-glucan active enzyme (*Ectocarpus* 36% identity) Hypothetical protein (*Thalassiosira oceanica* 41% identity)
Predicted protein 'CyLu-carb02'	Cycten m.11702 Lucan m.545 (70% identity)	Cycten comp45905_c0_seq1 Lucan comp75890_c0_seq1	Predicted protein (*Phaeodactylum* 55% identity)	ThalXL m.3795 (55% identity)	Catalytic activity; Carbohydrate metabolic proces	Exo-1,3-β glucanase (*Phytophthora* 37% identity)
Predicted protein 'CyLu-carb03'	Cycten m.22796 Lucan m.545 (70% identity)	Cycten comp52687_c0_seq1 Lucan comp75890_c0_seq1	Predicted protein (*Phaeodactylum* 55% identity)	ThalXL m.3795 (25% identity)	Catalytic activity; Carbohydrate metabolic proces	Exo-1,3-β glucanase (*Phytophthora* 37% identity)
Predicted protein 'CyLu-carb04'	Cycten m.15960 Lucan m.18574 (65% identity)	Cycten comp49882_c0_seq1 Lucan comp129195_c0_seq5	Predicted protein (*Phaeodactylum* 46% identity)	ThalXL m.5696 (25% identity)	Hydrolase activity; Hydrolyzing O-glycosyl compounds Carbohydrate metabolic process	Endo-1,3-β glucanase (amoeba 37% identity)
Hypothetical protein 'CyLu-carb05'	Cycten m.18649 Lucan m.702 (59% identity)	Cycten comp50940_c2_seq1 Lucan comp80753_c0_seq2	Hypothetical protein (*Emiliana* 37% identity)	ThalXL m.201700 (33% identity)	Catalytic activity; Carbohydrate metabolic process	Endoglucanase (bacteria 27% identity)
Hypothetical protein 'CyLu-carb06'	Cycten m.18650 Lucan m.6704 (57% identity)	Cycten comp50940_c3_seq1 Lucan comp121980_c0_seq1	Hypothetical protein (*Emiliana* 34% identity)		Catalytic activity; Carbohydrate metabolic process	Endoglucanase (bacteria 32% identity)

Table 9 continued.

Putative Gene Name	Protein Contig ID	Nucleotide Contig ID	Top BLASTP Hit	Similar Contigs	GO terms	Additional BLASTP Hits
Hypothetical protein 'CyLu-carb07'	Cycten m.3059 Lucan m.14774 (72% identity)	Cycten comp26167_c0_seq1 Lucan comp128106_c0_seq5	Hypothetical protein (*Emiliana* 44% identity)	ThalXL m.40103 (23% identity)	Polygalacturonase activity; Carbohydrate metabolic process	
Hypothetical protein 'CyLu-carb08'	Cycten m.7275 Lucan m.6653 (67% identity)	Cycten comp38151_c0_seq1 Lucan comp121907_c0_seq1	Hypothetical protein (*Thalassiosira oceanica*/*Cyclotella nana* 42% identity)	ThalXL m.3636 (48% identity)	Isomerase activity; Carbohydrate binding; Hexose metabolic process	Aldose-1-epimerase (bacteria 34% identity)
Galactosyl transferase-like protein	Cycten m.7620 Lucan m.2052 (35% identity)	Cycten comp39101_c0_seq1 Lucan comp107506_c0_seq2	Galactosyl transferase-like protein (vascular plant 30% identity)	ThalXL m.2290 (35% identity)	Response to chemical stimulus; Catalytic activity; Metabolic process	
O-glycosyl hydrolase superfamily protein	Cycten m.12359 Lucan m.10040 (48% identity)	Cycten comp46995_c0_seq1 Lucan comp125467_c0_seq1	O-glycosyl hydrolase superfamily protein (*Phytophthora* 34% identity)	ThalXL m.8583 (28% identity)	Carbohydrate metabolic process	
Phosphoketolase	Cycten m.14865 Lucan m.3822 (87% identity)	Cycten comp49014_c0_seq1 Lucan comp116334_c0_seq1	Phosphoketalose (*Phaeodactylum* 65% identity)	ThalXL m.42449 (39% identity)	Aldehyde-lysase activity; Carbohydrate metabolic process	Phosphoketalose (bacteria 61% identity)
β-glucosidase	Cycten m.17656 Lucan m.13625 (30% identity)	Cycten comp50586_c0_seq1 Lucan comp127661_c0_seq1	β-glucosidase (*Phaeodactylum* 42% identity)	ThalXL m.14448 (45% identity)	Hydrolase activity; Carbohydrate metabolic process	Glycoside (*Nannochloropsis* 36% identity)
Sedoheptulose-bidsphosphotase	Cycten m.10094 Lucan m.3490 (94% identity)	Cycten comp40704_c0_seq1 Lucan comp40704_c0_seq1	Sedoheptulose-bidsphosphotase (*Phaeodactylum* 82% identity)	ThalXL m.4791 (53% identity)	RNA methyltransferase; Carbohydrate metabolic process	

Table 9 continued.

Sulfate Modification						
Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Predicted protein 'CyLu-sulf01'	Cycten m.10816 Lucan m.6858 (71% identity)	Cycten comp41076_c0_seq1 Lucan comp122230_c0_seq2	Predicted protein (*Phaeodactylum* 50% identity)	ThalXL m.14029 (33% identity)	Sulfotransferase activity	Sulfotransferase domain protein (*Dictyostelium* 25% identity) Predicted protein (*Cyclotella nana* 42% identity)
Hypothetical protein 'CyLu-sulf02'	Cycten m.31234 Lucan m.5582 (53% identity)	Cycten comp71374_c0_seq1 Lucan comp120332_c0_seq3	Hypothetical protein (*Thalassiosira oceanica* 35% identity)	ThalXL m.17277 (42% identity)	Transferase activity; Integral to membrane; Sulfotransferase activity	Sulfotransferase-like protein (bacteria 32% identity)
Hypothetical protein 'CyLu-sulf03'	Cycten m.32434 Lucan m.9118 (56% identity)	Cycten comp76337_c0_seq1 Lucan comp124751_c0_seq3	Hypothetical protein (*Thalassiosira oceanica* 32% identity)	ThalXL m.17277 (42% identity)	Transferase activity; Integral to membrane; Sulfotransferase activity	Sulfotransferase-like protein (bacteria 32% identity)
Polysaccharide/ Carbohydrate Transport						
Predicted protein 'CyLu-carb09'	Cycten m.24443 Lucan m.7143 (79% identity)	Cycten comp55457_c0_seq1 Lucan comp122679_c0_seq2	Predicted protein (*Phaeodactylum* 63% identity)	ThalXL m.12323 (40% identity)	Carbohydrate binding	ER membrane protein complex subunit 7 homologue (vascular plant 37% identity)

Table 9 continued.

Oligosaccharide/ Modification						
Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Hypothetical protein 'CyLu-carb10'	Cycten m.11584 Lucan m.5042 (58% identity)	Cycten comp45748_c0_seq1 Lucan comp119391_c0_seq1	Hypothetical protein (*Cyclotella nana* 49% identity)		Oligonucleotide/ oligosaccharide binding; Helicase activity; Nucleotide binding	Helicase 2 (*Phaeodactylum* 43% identity)
Hypothetical protein 'CyLu-carb11'	Cycten m.27238 Lucan m.3494 (89% identity)	Cycten comp60504_c0_seq1 Lucan comp115309_c0_seq1	Hypothetical protein (*Cyclotella nana* 70% identity)		Oligonucleotide/ oligosaccharide binding; ATP-dependent helicase activity; Nucleotide binding	Hypothetical protein (Thalassiosira oceanica 61% identity) ATP-dependent RNA helicase (*Phytophthora* 40% identity)

Table 9 continued.

Polysaccharide/ Carbohydrate Metabolism						
Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Mannose-6-phosphate isomerase	Cycten m.25185 Lucan m.15460 (74% identity)	Cycten comp56709_c0_seq1 Lucan comp128351_c0_seq1	Mannose-6-phosphate isomerase (*Phaeodactylum* 69% identity)	ThalXL m.11425 (59% identity)	Mannose-6-phosphate isomerase activity; GDP-mannose biosynthetic process	Hypothetical protein (*Cyclotella nana* 66% identity; *Thalassiosira oceanica* 64% identity)
Hypothetical protein 'CyLu-carb00'	Cycten m.23226 Lucan m.6944 (89% identity)	Cycten comp53449_c0_seq1 Lucan comp122349_c1_seq2	Hypothetical protein (*Phaeodactylum* 86% identity)	ThalXL m.13587 (38% identity)	Polysaccharide biosynthesis process; GDP-1-fucose synthetase; Oxidoreductase activity	GDP-fucose synthetase (*Cyclotella nana* 81% identity)
Chitooligosaccharide deacetylase	Cycten m.3083 Lucan m.13276 (80% identity)	Cycten comp26204_c0_seq1 Lucan comp127515_c0_seq3	Chitooligosaccharide deacetylase (71% identity)	ThalXL m.26845 (39% identity)	Urate catabolism protein; Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds; Carbohydrate metabolic process	Hypothetical protein (*Thalassiosira oceanica* 71% identity) Predicted protein (*Phaeodactylum* 59% identity)
Predicted protein 'CyLu-carb02'	Cycten m.11702 Lucan m.545 (70% identity)	Cycten comp45905_c0_seq1 Lucan comp75890_c0_seq1	Predicted protein (*Phaeodactylum* 55% identity)	ThalXL m.3795 (55% identity)	Catalytic activity; Carbohydrate metabolic process	Exo-1,3-β glucanase (*Phytophthora* 37% identity)

Table 10. Results of the BLASTP search of the contigs in common between the EPS-producing taxa *Cyclophora tenuis* ("Cycten") and *Lucanicum concatenatum* ("Lucan") and *Thalassionema frauenfeldii* ("ThalXL") that were not a match to contigs from non EPS-producing taxa *Astrosyne radiata* and *Thalassionema* sp. "BlueH20."

Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Predicted protein 'CyLu-carb03'	Cycten m.22796 Lucan m.545 (70% identity)	Cycten comp52687_c0_seq1 Lucan comp75890_c0_seq1	Predicted protein (*Phaeodactylum* 55% identity)	ThalXL m.3795 (25% identity)	Catalytic activity; Carbohydrate metabolic proces	Exo-1,3-β glucanase (*Phytophthora* 37% identity)
Hypothetical protein 'ThXL-carb05'	ThalXL m.20780	ThalXL comp112914_c0_seq2	Hypothetical protein (*Thalassiosira oceanica* 38% identity)	Cycten m.11854 (50% identity) Lucan m.15729 (33% identity)	Hydrolase activity	Endo-1,3:1,4-β-glucanase (*Phytophthora* 30% identity)
Predicted protein 'CyLu-carb04'	Cycten m.15960 Lucan m.18574 (65% identity)	Cycten comp49882_c0_seq1 Lucan comp129195_c0_seq5	Predicted protein (*Phaeodactylum* 46% identity)	ThalXL m.5696 (25% identity)	Hydrolase activity; Hydrolyzing O-glycosyl compounds Carbohydrate metabolic process	Endo-1,3-β glucanase (amoeba 37% identity)
O-glycosyl hydrolase superfamily protein	Cycten m.12359 Lucan m.10040 (48% identity)	Cycten comp46995_c0_seq1 Lucan comp125467_c0_seq1	O-glycosyl hydrolase superfamily protein (*Phytophthora* 34% identity)	ThalXL m.8583 (28% identity)	Carbohydrate metabolic process	
Predicted protein 'CyLu-carb01'	Cycten m.3403 Lucan m.1913 (39% identity)	Cycten comp26479_c0_seq1 Lucan comp105543_c0_seq1	Predicted protein (*Phaeodactylum* 45% identity)	ThalXL m.3314 (69% identity)	Hydrolase activity	β-1,6-glucan active enzyme (*Ectocarpus* 36% identity) Hypothetical protein (*Thalassiosira oceanica* 41% identity)
Hypothetical protein 'CyLu-carb05'	Cycten m.18649 Lucan m.702 (59% identity)	Cycten comp50940_c2_seq1 Lucan comp80753_c0_seq2	Hypothetical protein (*Emiliana* 37% identity)	ThalXL m.201700 (33% identity)	Catalytic activity; Carbohydrate metabolic process	Endoglucanase (bacteria 27% identity)

Table 10 continued.

Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Hypothetical protein 'CyLu-carb07'	Cycten m.3059 Lucan m.14774 (72% identity)	Cycten comp26167_c0_seq1 Lucan comp128106_c0_seq5	Hypothetical protein (*Emiliana* 44% identity)	ThalXL m.40103 (23% identity)	Polygalacturonase activity; Carbohydrate metabolic process	
Hypothetical protein 'CyLu-carb08'	Cycten m.7275 Lucan m.6653 (67% identity)	Cycten comp38151_c0_seq1 Lucan comp121907_c0_seq1	Hypothetical protein (*Thalassiosira oceanica*/*Cyclotella nana* 42% identity)	ThalXL m.3636 (48% identity)	Isomerase activity; Carbohydrate binding; Hexose metabolic process	Aldose-1-epimerase (bacteria 34% identity)
Phosphoketolase	Cycten m.14865 Lucan m.3822 (87% identity)	Cycten comp49014_c0_seq1 Lucan comp116334_c0_seq1	Phosphoketalose (*Phaeodactylum* 65% identity)	ThalXL m.42449 (39% identity)	Aldehyde-lysase activity; Carbohydrate metabolic process	Phosphoketalose (bacteria 61% identity)
β-glucosidase	Cycten m.17656 Lucan m.13625 (30% identity)	Cycten comp50586_c0_seq1 Lucan comp127661_c0_seq1	β-glucosidase (*Phaeodactylum* 42% identity)	ThalXL m.14448 (45% identity)	Hydrolase activity; Carbohydrate metabolic process	Glycoside (*Nannochloropsis* 36% identity)
Hypothetical protein 'ThXL-carb06'	ThalXL m.23105	ThalXL comp113646_c0_seq3	Hypothetical protein (*Thalassiosira oceanica* 59% identity)	Cycten m.15960 (45% identity) Lucan m.18574 (43% identity)	Hydrolase activity, hydrolyzing O-glycosyl compounds	Hypothetical protein (*Cyclotella nana* 57% identity) GPI-anchored β-glucanase (*Acanthamoeba* 32% identity)
Hypothetical protein 'ThXL-carb07'	ThalXL m.23106	ThalXL comp113646_c0_seq4	Hypothetical protein (*Thalassiosira oceanica* 59% identity)	Cycten m.15960 (45% identity) Lucan m.18574 (43% identity)	Hydrolase activity, hydrolyzing O-glycosyl compounds	Hypothetical protein (*Cyclotella nana* 57% identity) GPI-anchored β-glucanase (*Acanthamoeba* 32% identity)

Table 10 continued.

Putative Gene Name	Protein Contig ID	Nucleotide Contig ID	Top BLASTP Hit	Similar Contigs	GO terms	Additional BLASTP Hits
Glycosyl transferase	ThalXL m.10195	ThalXL comp107005_c0_seq2	Glycosyl transferase (*Phaeodactylum* 72% identity)	Cycten m.3512 (45% identity) Lucan m.10195 (37% identity)	1,3-β-D-glucan synthase complex	*Thalassiosira oceanica* hypothetical protein (64% identity)
Galactosyl transferase-like protein	Cycten m.7620 Lucan m.2052 (35% identity)	Cycten comp39101_c0_seq1 Lucan comp107506_c0_seq2	Galactosyl transferase-like protein (vascular plant 30% identity)	ThalXL m.2290 (35% identity)	Response to chemical stimulus; Catalytic activity; Metabolic process	
Sedoheptulose-bidsphosphotase	Cycten m.10094 Lucan m.3490 (94% identity)	Cycten comp40704_c0_seq1 Lucan comp40704_c0_seq1	Sedoheptulose-bidsphosphotase (*Phaeodactylum* 82% identity)	ThalXL m.4791 (53% identity)	RNA methyltransferase; Carbohydrate metabolic process	
Sulfate Modification						
Hypothetical protein 'ThXL-sulf03'	ThalXL m.10698	ThalXL comp107515_c0_seq1	Hypothetical protein (*Emiliana* 36% identity)	Cycten m.19595 (40% identity) Lucan m.5139 (31% identity)	Sulfotransferase activity	Vascular plant sulfotransferase domain (32% identity)
Sulfatotransferase domain	ThalXL m.24674	ThalXL comp114068_c2_seq4	Sulfatotransferase domain (bacteria 35% identity)	Cycten m.12307 (31% identity) Lucan m.7442 (34% identity)	Transferase activity	
Predicted protein 'CyLu-sulf01'	Cycten m.10816 Lucan m.6858 (71% identity)	Cycten comp41076_c0_seq1 Lucan comp122230_c0_seq2	Predicted protein (*Phaeodactylum* 50% identity)	ThalXL m.14029 (33% identity)	Sulfostransferase activity	Sulfotransferase domain protein (*Dictyostelium* 25% identity) Predicted protein (*Cyclotella nana* 42% identity)

Table 10 continued.

Putative Gene Name	Protein Contig ID	Nucleotide Contig ID	Top BLASTP Hit	Similar Contigs	GO terms	Additional BLASTP Hits
Hypothetical protein 'CyLu-sulf02'	Cycten m.31234 Lucan m.5582 (53% identity)	Cycten comp71374_c0_seq1 Lucan comp120332_c0_seq3	Hypothetical protein (*Thalassiosira oceanica* 35% identity)	ThalXL m.17277 (42% identity)	Transferase activity; Integral to membrane; Sulfotransferase activity	Sulfotransferase-like protein (bacteria 32% identity)
Hypothetical protein 'CyLu-sulf03'	Cycten m.32434 Lucan m.9118 (56% identity)	Cycten comp76337_c0_seq1 Lucan comp124751_c0_seq3	Hypothetical protein (*Thalassiosira oceanica* 32% identity)	ThalXL m.17277 (42% identity)	Transferase activity; Integral to membrane; Sulfotransferase activity	Sulfotransferase-like protein (bacteria 32% identity)
Hypothetical protein 'ThXL-sulf04'	ThalXL m.15660	ThalXL comp110825_c0_seq2	Hypothetical protein (*Thalassiosira oceanica* 53% identity)	Cycten m.14285 (43% identity) Lucan m.8191 (54% identity)	Catalytic activity; Arylsulfatase activity; Sulfuric ester hydrolase activity; Hydrolase activity; Proteinaceous extracellular matrix	Arylsulfatase e (Orca 45% identity)
Secretion/ Transmembrane Transport						
Hypothetical protein 'ThXL-sec01'	ThalXL m.31931	ThalXL comp115489_c1_seq1	Hypothetical protein (*Thalassiosira oceanica* 57% identity)	Cycten m.5912 (24% identity) Lucan m.5253 (40% identity)	Extracellular ligand-gated ion channel activity	Acetylcholine receptor (tunicate 26% identity)
Hypothetical protein 'ThXL-sec02'	ThalXL m.31933	ThalXL comp115489_c1_seq3	Hypothetical protein (*Thalassiosira oceanica* 60% identity)	Cycten m.5912 (24% identity) Lucan m.5253 (40% identity)	Extracellular ligand-gated ion channel activity	Chloride channel (roundworm 26% identity)
Hypothetical protein 'ThXL-sec03'	ThalXL m.31938	ThalXL comp115489_c1_seq9	Hypothetical protein (*Thalassiosira oceanica* 49% identity)	Cycten m.5912 (29% identity) Lucan m.5254 (40% identity)	Extracellular ligand-gated ion channel activity; ion transport	Amniobutyric acid receptor (flatworm 29% identity)

Table 10 continued.

Putative Gene Name	**Protein Contig ID**	**Nucleotide Contig ID**	**Top BLASTP Hit**	**Similar Contigs**	**GO terms**	**Additional BLASTP Hits**
Transmambrane facilitator protein	ThalXL m.24457	ThalXL comp114009_c0_seq1	Transmembrane facilitator protein (bacteria 50% identity)	Cycten m.7014 (21% identity) Lucan m.10903 (23% identity)	Transmembrane transport; Integral to membrane	Hypothetical protein (bacteria 48% identity)
Hypothetical protein 'ThXL-sec07'	ThalXL m.24468	ThalXL comp114009_c0_seq1 1	Hypothetical protein (bacteria 54% identity)	Cycten m.26449 (33% identity) Lucan m.6128 (28% identity)	Transmembrane transport; Integral to membrane	Transmembrane facilitator protein (bacteria 51% identity)
AMP-dependent synthetase 'ThXL-AMPsyn2'	ThalXL m.30924	ThalXL comp115302_c1_seq2	AMP-dependent synthetase (bacteria 56% identity)	Cycten m.10738 (45% identity) Lucan m.2877 (46% identity)	Transferase activity; Transfering acyl groups; Transmembrane transport; Integral to membrane	
Hypothetical protein 'ThXL-sec08'	ThalXL m.4029	ThalXL comp95652_c0_seq1	Hypothetical protein (bacteria 51% identity)	Cycten m.6437 (34% identity) Lucan m. 17834 (29% identity)	Transport; HlyD family secretion protein	RND family efflux transporter (bacteria 50% identity)
Predicted protein 'CyLu-carb09'	Cycten m.24443 Lucan m.7143 (79% identity)	Cycten comp55457_c0_seq1 Lucan comp122679_c0_seq2	Predicted protein (*Phaeodactylum* 63% identity)	ThalXL m.12323 (40% identity)	Carbohydrate binding	ER membrane protein complex subunit 7 homologue (vascular plant 37% identity)

Table 10 continued.

Appendix

TAXA USED IN PHYLOGENETIC ANALYSIS, WITH LOCALITY, THERIOT LAB SPECIMEN VOUCHER ID (HK###) AND GENBANK ACCESSION NUMBERS. TAXA MARKED WITH ASTERISKS WERE USED IN THERIOT ET AL (2010) BUT WERE MISIDENTIFIED IN THAT MANUSCRIPT (* = "BIDDULPHIA ALTERANS;" ** = "ISTHMIA ENERVIS;" * = "COSCINODISCUS CONCINNIFORMIS").**

Taxon	Voucher (HK###)	Locality	GenBank Accession (SSU, *rbc*L, *psb*C)
Achnanthes coarctata (Brébisson ex W. Smith) Grunow in Cleve & Grunow	HK079	FD185 (UTEX)	HQ912594, HQ912458, HQ912287
Achnanthes sp Bory	HK303	SanNicholas1 (San Nicholas, Canary Islands)	KC309473, KC309545, KC309617
Achnanthes sp Bory	HK309	ECT3883 (Rainbow Harbor, Long Beach, California)	KC309474, KC309546, KC309618
Achnanthes sp Bory	HK310	ECT3911 (Long Beach, California)	KC309475, KC309547, KC309619
Achnanthes sp Bory	HK311	ECT3684 (Achang Reef, Guam)	KC309476, KC309548, KC309620
Amphipleura pellucida Kützing	HK287	ECT3568 (Lake Travis, Texas)	KC309477, KC309549, KC309621
Amphora coffeaeformis (Agardh) Kützing	HK089	FD75 (UTEX)	HQ912602, HQ912466, HQ912295
Bacillaria paxillifer (O. F. Müller) T. Marsson	HK130	FD468 (UTEX)	HQ912627, HQ912491, HQ912320
Berkeleya rutilans (Trentepohl ex Roth) Grunow	HK154	ECT3616 (Laguna Beach, California)	HQ912637, HQ912501, HQ912330
Caloneis lewisii Patrick	HK060	FD54 (UTEX)	HQ912580, HQ912444, HQ912273
Climaconeis riddleae Prasad	HK178	ECT3724 (Umatac Bay, Guam)	HQ912644, HQ912508, HQ912337
Climaconeis undulata (Meister) Lobban et al	HK218	ECT3743 (Talofofo Bay, Guam)	KC309478, KC309550, N/A
Cocconeis placentula Ehrenberg	HK077	FD23 (UTEX)	HQ912592, HQ912456, HQ912285
Cocconeis sp Ehrenberg	HK312	ECT3901 (Channel #5, US-1, Florida)	KC309479, KC309551, KC309622
Craticula cuspidata (Kützing) Mann	HK061	FD35 (UTEX)	HQ912581, HQ912445, HQ912274
Cylindrotheca closterium (Ehrenberg) Reimann & Lewin	HK180	CCMP1855	HQ912645, HQ912509, HQ912338
Cymatopleura elliptica (Brebisson ex Kutzing) W. Smith	HK215	L1333 (UTEX)	HQ912659, HQ912523, HQ912352
Denticula kuetzingii Grunow	HK104	FD135 (UTEX)	HQ912610, HQ912474, HQ912303
Diploneis subovalis Cleve	HK084	FD282 (UTEX)	HQ912597, HQ912461, HQ912290

Eunotia curvata Lagerstedt	HK086	FD412 (UTEX)	HQ912599, HQ912463, HQ912292
Eunotia glacialis Meister	HK069	FD46 (UTEX)	HQ912586, HQ912450, HQ912279
Eunotia pectinalis (Kützing) Rabenhorst	HK153	NIES461	HQ912636, HQ912500, HQ912329
Eunotia sp Ehrenberg	HK286	ECT3676 (Tinago River, Guam)	KC309480, KC309552, KC309623
Fallacia monoculata (Hustedt) Mann	HK082	FD254 (UTEX)	HQ912596, HQ912460, HQ912289
Fallacia pygmaea (Kützing) Stickle & Mann	HK093	FD294 (UTEX)	HQ912605, HQ912469, HQ912298
Gomphonema affine Kützing	HK098	FD173 (UTEX)	HQ912608, HQ912472, HQ912301
Gomphonema parvulum (Kützing) Kützing	HK081	FD241 (UTEX)	HQ912595, HQ912459, HQ912288
Gyrosigma acuminatum (Kützing) Rabenhorst	HK085	FD317 (UTEX)	HQ912598, HQ912462, HQ912291
Lemnicola hungarica (Grunow) Round	HK129	FD456 (UTEX)	HQ912626, HQ912490, HQ912319
Manguinea sp Paddock	HK135	CS782 (CSIRO)	HQ912631, HQ912495, HQ912324
Mastogloia sp Thwaites in W. Smith	HK136	29X07-6B (Mustang Island, Texas)	HQ912632, HQ912496, HQ912325
Mastogloia sp Thwaites in W. Smith	HK314	ECT3762 (Taeleyag Beach, Guam)	KC309481, KC309553, N/A
Meuniera membranacea (Cleve) P. C. Silva	HK313	ECT3896 (Port Aransas Jetty, Texas)	KC309482, KC309554, KC309624
Navicula cryptocephala Kützing	HK090	FD109 (UTEX)	HQ912603, HQ912467, HQ912296
Neidium affine (Ehrenberg) Pfitzer	HK064	FD127 (UTEX)	HQ912583, HQ912447, HQ912276
Neidium bisulcatum (Lagerstedt) Cleve	HK076	FD417 (UTEX)	HQ912591, HQ912455, HQ912284
Neidium productum (W. Smith) Cleve	HK063	FD116 (UTEX)	HQ912582, HQ912446, HQ912275
Nitzschia filiformis (W. Smith) Van Heurck	HK073	FD267 (UTEX)	HQ912589, HQ912453, HQ912282
Phaeodactylum tricornutum Bohlin	HK011	CCMP2561	HQ912556, HQ912420, HQ912250
Pinnularia brebissonii (Kützing) Rabenhorst	HK092	FD274 (UTEX)	HQ912604, HQ912468, HQ912297
Pinnularia termitina (Ehrenberg) Patrick	HK088	FD484 (UTEX)	HQ912601, HQ912465, HQ912294
Placoneis elginensis (Gregory) Cox	HK096	FD416 (UTEX)	HQ912607, HQ912471, HQ912300
Scoliopleura peisonis Grunow	HK103	FD13 (UTEX)	HQ912609, HQ912473, HQ912302
Stauroneis acuta W. Smith	HK059	FD51 (UTEX)	HQ912579, HQ912443, HQ912272
Surirella ovata Kützing	HK214	L1241 (UTEX)	HQ912658, HQ912522,

			HQ912351
Tryblionella apiculata Gregory	HK087	FD465 (UTEX)	HQ912600, HQ912464, HQ912293
Asterionella formosa Hassall	HK144	UTCC605	HQ912633, HQ912497, HQ912326
Asterionellopsis glacialis (Castracane) Round	HK107	CCMP134	HQ912613, HQ912477, HQ912306
Asterionellopsis socialis (Lewin & Norris) Crawford & Gardner	HK181	CCMP1717	HQ912646, HQ912510, HQ912339
Asterionellopsis socialis (Lewin & Norris) Crawford & Gardner	HK319	ECT3920 (Ft. Stevens State Park, Oregon)	JX413545, JX413562, JX413579
Astrosyne radiata Ashworth et Lobban	HK169	ECT3697 (Gab Gab Beach, Guam)	JN975238, JN975252, JN975267
Bleakeleya notata (Grunow in Van Heurck) F.E. Round	HK247	ECT3733 (Pago Bay, Guam)	HM627330, HM627327, HM627324
Centronella reicheltii Voigt	HK150	CCAP1011/1	HQ912635, HQ912499, HQ912328
Ctenophora pulchella (Ralfs ex Kuetzing) Williams & Round	HK105	FD150 (UTEX)	HQ912611, HQ912475, HQ912304
Cyclophora castracanei Ashworth et Lobban	HK243	GU44AB-6 (Gab Gab Beach, Guam)	JN975242, JN975256, JN975271
Cyclophora tabellariformis Ashworth et Lobban	HK306	ECT3892 (Carrabelle, Florida)	JN975243, JN975257, JN975272
Cyclophora tenuis Castracane	HK216	ECT3723 (Umatac Bay, Guam)	HQ912660, HQ912524, HQ912353
Cyclophora tenuis Castracane	HK308	ECT3838 (Long Beach, California)	JN975241, JN975255, JN975270
Cyclophora tenuis Castracane	HK307	ECT3854 (Kahana Beach Park, Oahu, Hawaii)	JN975240, JN975254, JN975269
Delphineis sp G.W. Andrews	HK133	CCMP1095	HQ912629, HQ912493, HQ912322
Delphineis surirella (Ehrenberg) Andrews	HK295	ECT3886 (Bald Head Island, North Carolina)	JX413544, JX413561, JX413578
Diatoma elongatum (Lyngbye) Agardh	HK119	UTCC62	HQ912622, HQ912486, HQ912315
Diatoma tenue Agardh	HK078	FD106 (UTEX)	HQ912593, HQ912457, HQ912286
Dimeregramma dubium (Grunow) H. Peragallo & M. Peragallo	HK325	GU44AI-3 (Gab Gab Beach, Guam)	JX413547, JX413564, JX413581
Dimeregramma sp J. Ralfs in A. Pritchard	HK288	ECT3864 (MSI, Port Aransas, Texas)	JN975244, JN975258, JN975273
Florella pascuensis Navarro	HK175	ECT3756 (Guam)	JN975246, JN975260, JN975275
Fragilariforma virescens (Ralfs) Williams & Round	HK132	FD291 (UTEX)	HQ912628, HQ912492, HQ912321
Grammatophora oceanica Ehrenberg	HK147	CCMP410	HQ912634, HQ912498, HQ912327
Hyalosira interrupta (Ehrenberg) Navarro	HK248	ECT3700 (Gab Gab Beach, Guam)	JN975247, JN975261, JN975276

Koernerella recticostata (Körner) Ashworth, Lobban et Theriot	HK242	GU44AB-8 (Gab Gab Beach, Guam)	HM627331, HM627328, HM627325
Licmophora paradoxa (Lyngbye) Agardh	HK106	CCMP2313	HQ912612, HQ912476, HQ912305
Licmophora remulus Grunow	HK302	GU52-O (Outhouse Beach, Guam)	JN975248, JN975263, N/A
Nanofrustulum cf *shiloi* (J.J. Lee, C.W. Reimer, & M.E. McEnery) F.E. Round, H. Hallsteinsen, & E. Paasche	HK056	CCMP2649	HQ912578, HQ912442, HQ912271
Opephora sp Petit	HK296	ECT3831 (Texas)	JN975249, JN975264, JN975278
Perideraion montgomeryii Lobban, Jordan et Ashworth	HK246	GU7 (UOG Marine Lab, Guam)	HM627332, HM627329, HM627326
Plagiogramma staurophorum (Gregory) Heiberg	HK212	ECT3776 (Taeleyag Beach, Guam)	HQ912656, HQ912520, HQ912349
Plagiogramma sp R.K. Greville	HK324	ECT3924 (Potlatch State Park, Washington)	JX413546, JX413563, JX413580
Podocystis spathulatum (Shadbolt) Van Heurck	HK217	ECT3733 (Anae Island, Guam)	HQ912661, HQ912525, HQ912354
Psammoneis japonica S. Sato, W. Kooistra & L. Medlin	HK299	GU52-O (Outhouse Beach, Guam)	JN975250, JN975265, JN975279
Rhabdonema arcuatum (Lyngbye; Agardh) Kützing	HK304	ECT3898 (Stillwater Cove, California)	JN975251, JN975266, JN975280
Rhaphoneis amphiceros (Ehrenberg) Ehrenberg	HK237	ECT3828 (Redfish Bay, Texas)	HQ912673, HQ912537, KC309625
Staurosira construens Ehrenberg	HK071	FD232 (UTEX)	HQ912587, HQ912451, HQ912280
Staurosirella pinnata (Ehrenberg) Williams & Round	HK116	CCMP330	HQ912620, HQ912484, HQ912313
Striatella unipunctata (Lyngbye) Agardh	HK177	ECT3648 (Asan Beach, Guam)	HQ912643, HQ912507, HQ912336
Striatella unipunctata (Lyngbye) Agardh	HK318	ECT3874 (Channel #5, US-1, Florida)	JX419383, JX419384, JX419385
Synedra famelica Kützing	HK072	FD255 (UTEX)	HQ912588, HQ912452, HQ912281
Synedra hyperborea Grunow	HK117	CCMP1423	HQ912621, HQ912485, HQ912314
Synedra ulna (Nitzsch) Ehrenberg	HK075	FD404 (UTEX)	HQ912590, HQ912454, HQ912283
Synedropsis cf *recta* Hasle, Medlin et Syvertsen	HK110	CCMP1620	HQ912616, HQ912480, HQ912309
Tabellaria flocculosa (Roth) Kützing	HK065	FD133 (UTEX)	HQ912584, HQ912448, HQ912277
Tabularia cf *tabulata* (C. A. Agardh) Snoeijs	HK109	CCMP846	HQ912615, HQ912479, HQ912308
unidentified araphid "Araphidribbon FL"	HK315	9X10-2 (FSU Marine Lab, Florida)	JX413542, JX413559, JX413576
unidentified araphid "Epipsammicchain FL"	HK316	10X10-2 (St. George Island, Florida)	JX413543, JX413560, JX413577
Acanthoceras sp H. Honigmann	HK250	6VII09-1A (Lake	HQ912676, HQ912540,

		Okoboji, Iowa)	HQ912367
Amphipentas pentacrinus Ehrenberg	HK289	ECT3874 (Channel #5, US-1, Florida)	KC309483, KC309555, KC309626
Amphitetras antediluviana Ehrenberg	HK223	ECT3627 (Montana de Oro State Park, California)	HQ912665, HQ912529, HQ912358
Arcocellulus mammifer Hasle, von Stosch & Syvertsen	HK044	CCMP132	HQ912569, HQ912433, HQ912263
Ardissonea formosa (Hantzsch in Rabenhorst) Grunow in Cleve & Grunow	HK209	ECT3655 (Gab Gab Beach, Guam)	HQ912653, HQ912517, HQ912346
Ardissonea fulgens v gigantea (Lobarzewsky) De Toni	HK167	ECT3655 (Gab Gab Beach, Guam)	HQ912639, HQ912503, HQ912332
Ardissonea fulgens (Greville) Grunow in Cleve & Grunow	HK305	26VI10-2A (Bald Head Island, North Carolina)	KC309484, KC309556, KC309627
Attheya longicornis Crawford & Gardner in Crawford, Gardner & Medlin	HK317	CCMP214	JX401230, JX401247, JX401265
Attheya septentrionalis (Østrup) Crawford in Crawford, Gardner & Medlin	HK112	CCMP2084	HQ912618, HQ912482, HQ912311
Bellerochea horologicalis von Stosch	HK235	ECT3829 (Redfish Bay, Texas)	HQ912672, HQ912536, HQ912364
Bellerochea malleus (Brightwell) Van Heurck	HK265	Har-1 (HBOI Boat Dock, Florida)	KC309485, KC309557, KC309628
Biddulphia alternans (J.W. Bailey) Van Heurck	HK292	ECT3886 (Bald Head Island, North Carolina)	JX401229, JX401246, JX401264
Biddulphia biddulphiana (J.E. Smith) Boyer	HK271	ClayHI (Kona, Hawaii)	JN975239, JN975253, JN975268
Biddulphia biddulphiana (J.E. Smith) Boyer	HK328	ECT3902 (Channel #5, US-1, Florida)	JX401227, JX401244, JX401262
Biddulphia tridens (Ehrenberg) Ehrenberg	HK239	ECT3838 (Long Beach, California)	HQ912674, HQ912538, HQ912365
Biddulphia tridens (Ehrenberg) Ehrenberg	HK327	ECT3902 (Channel #5, US-1, Florida)	JX401228, JX401245, JX401263
* *Biddulphia reticulum* (Ehrenberg) Boyer	HK252	ECT3856 (Kahana Bay, Oahu, Hawaii)	HQ912677, HQ912541, KC309629
Biddulphia cf *reticulum* (Ehrenberg) Boyer	HK329	ECT3891 (St. George Island, Florida)	KC309486, KC309558, KC309630
Biddulphiopsis membranacea (Cleve) von Stosch & Simonsen	HK166	ECT3655 (Gab Gab Beach, Guam)	HQ912638, HQ912502, HQ912331
Biddulphiopsis titiana (Grunow) von Stosch & Simonsen	HK170	ECT3697 (Gab Gab Beach, Guam)	HQ912641, HQ912505, HQ912334
Brockmanniella brockmannii (Hustedt) Hasle, von Stosch & Syvertsen	HK040	CCMP151	HQ912565, HQ912429, HQ912259
Campylosira cymbelliformis (Schmidt) Grunow in Van Heurck	HK122	CCC-1 (Corpus Christi Bay, Texas)	HQ912623, HQ912487, HQ912316
Cerataulina pelagica (Cleve) Hendey	HK230	ECT3845 (Ward Island, Texas)	HQ912669, HQ912533, HQ912361
Cerataulina pelagica (Cleve) Hendey	HK232	ECT3836 (Rainbow Harbor, Long Beach, California)	KC309487, KC309559, KC309631

Cerataulus smithii Ralfs in Pritchard	HK331	ECT3886 (Bald Head Island, North Carolina)	KC309488, KC309560, KC309632
Cerataulus smithii Ralfs in Pritchard	HK332	ECT3891 (St. George Island, Florida)	KC309489, KC309561, KC309633
Cerataulus smithii Ralfs in Pritchard	HK224	ECT3829 (Redfish Bay, Texas)	HQ912666, HQ912530, N/A
Chaetoceros muelleri Lemmermann	HK028	CCMP1316	HQ912558, HQ912422, HQ912252
Chaetoceros peruvianus Brightwell	HK202	ECT3821 (Ward Island, Texas)	HQ912650, HQ912514, HQ912343
Chrysanthomodiscus floriatus Mann	HK171	ECT3710 (Haputo Point, Guam)	HQ912642, HQ912506, HQ912335
Climacosphenia elongatum J.W. Bailey	HK276	ECT3758 (Scuba Beach, Guam)	HQ912685, HQ912549, HQ912371
Cyclostephanos dubius (Fricke) Round	HK051	Waco5 (Lake Waco, Texas)	HQ912575, HQ912439, HQ912268
Cyclotella meneghiniana Kützing	HK052	Waco1 (Lake Waco, Texas)	HQ912576, HQ912440, HQ912269
Cyclotella sp (F.T. Kützing) A. de Brébisson	HK126	LO-4-2 (Lake Ohrid, Macedonia)	HQ912625, HQ912489, HQ912318
Cymatosira lorenziana Grunow	HK283	ECT3874 (Channel #5, US-1, Florida)	KC309490, KC309562, KC309634
Cymatosira belgica Grunow in Van Heurck	HK343	ECT3892 (Carrabelle, Florida)	N/A, KC309563, KC309635
Dactyliosolen blavyanus (H. Peragallo) Hasle	HK301	ECT3891 (St. George Island, Florida)	KC309491, KC309564, KC309636
Detonula confervacea (Cleve) Gran	HK111	CCMP353	HQ912617, HQ912481, HQ912310
Ditylum sol (A. Schmidt) Cleve	HK240	Har-1 (HBOI Boat Dock, Florida)	KC309492, KC309565, KC309637
Ditylum brightwelli Peragallo & Peragallo	HK285	ECT3884 (Pacific Grove, California)	KC309493, KC309566, KC309638
Eucampia sp C.G. Ehrenberg	HK260	ECT3836 (Rainbow Harbor, Long Beach, California)	KC309494, KC309567, KC309639
Eucampia zodiacus Ehrenberg	HK320	ECT3896 (Port Aransas Jetty, Texas)	KC309495, KC309568, KC309640
Eunotogramma sp J.F. Weisse	HK294	ECT3886 (Bald Head Island, North Carolina)	JN975245, JN975259, JN975274
Extubocellulus cribriger Hasle, von Stosch & Syvertsen	HK046	CCMP391	HQ912571, HQ912435, HQ912265
Extubocellulus sp G.R. Hasle, H.A. von Stosch & E.E. Syvertsen	HK149	CCAP1018/1	N/A, KC309570, KC309642
Hemiaulus sinensis Greville	HK123	CCH-1 (Corpus Christi Bay, Texas)	HQ912624, HQ912488, HQ912317
Hemiaulus sinensis Greville	HK297	24I10-1A (Clear Lake, Texas)	KC309496, KC309569, KC309641
Hydrosera sp. G.C. Wallich	HK274	TN-CYTX025 (Austin, Texas)	HQ912683, HQ912547, N/A
** *Isthmia minima* Harvey & Bailey	HK275	CXCL (Guam)	HQ912684, HQ912548, N/A
Lampriscus shadboltianus v. crenulata	HK257	GU44AB-8 (Gab Gab	HQ912680, HQ912544, N/A

Navarro		Beach Guam)	
Lampriscus shadboltianus (Greville) Peragallo & Peragallo	HK264	Pan-1 (STRI Station, Panama)	KC309497, KC309571, KC309644
Lampriscus orbiculatus (Shadbolt) Peragallo & Peragallo	HK125	GU7 (UOG Marine Lab, Guam)	HQ912679, HQ912543, KC309643
Leyanella arenaria Hasle, von Stosch & Syvertsen	HK045	CCMP471	HQ912570, HQ912434, HQ912264
Lithodesmioides polymorpha von Stosch	HK211	ECT3772 (Taeleyag Beach, Guam)	HQ912655, HQ912519, HQ912348
Lithodesmium intricatum (T. West) Peragallo & Peragallo	HK253	ECT3850 (Kahalu'u, Hawaii)	HQ912678, HQ912542, HQ912368
Lithodesmium intricatum (T. West) Peragallo & Peragallo	HK231	ECT3836 (Rainbow Harbor, Long Beach, California)	HQ912670, HQ912534, HQ912362
Lithodesmium undulatum Ehrenberg	HK029	CCMP1797	HQ912559, HQ912423, HQ912253
Mastodiscus radiatus Prasad & Nienow	HK249	ECT3822 (Ward Island, Texas)	HQ912675, HQ912539, HQ912366
Minidiscus trioculatus (F. J. R. Taylor) Hasle	HK036	CCMP495	HQ912563, HQ912427, HQ912257
Minutocellus polymorphus (Hargraves & Guillard) Hasle, von Stosch & Syvertsen	HK043	CCMP497	HQ912568, HQ912432, HQ912262
Minutocellus polymorphus (Hargraves & Guillard) Hasle, von Stosch & Syvertsen	HK321	ECT3920 (Ft. Stevens State Park, Oregon)	KC309498, KC309572, KC309645
Odontella aurita Agardh	HK203	ECT3619 (Elkhorn Slough, California)	JX413551, JX413568, JX413585
Odontella aurita Agardh	HK333	ECT3888 (San Sebastian, Canary Islands)	KC309508, KC309582, KC309655
Odontella aurita Agardh	HK334	ECT3788 (Talofofo Bay, Guam)	KC309509, KC309583, KC309656
Odontella aurita v minima (Grunow in Van Heurck) De Toni	HK336	ECT3886 (Bald Head Island, North Carolina)	KC309510, KC309584, KC309657
Odontella aurita v minima (Grunow in Van Heurck) De Toni	HK337	ECT3892 (Carrabelle, Florida)	KC309511, KC309585, N/A
Odontella longicruris (Greville) Hoban	HK226	ECT3828 (Redfish Bay, Texas)	JX413553, JX413570, JX413587
Odontella longicruris v hyalina (Schröder) Hoban	HK284	ECT3884 (Pacific Grove, California)	KC309512, KC309586, KC309658
Odontella longicruris v hyalina (Schröder) Hoban	HK338	ECT3920 (Ft. Stevens State Park, Oregon)	KC309513, KC309587, KC309659
Odontella rhomboides Jahn & Kusber	HK282	ECT3845 (Ward Island, Texas)	KC309514, KC309588, KC309660
Odontella rhomboides Jahn & Kusber	HK339	ECT3905 (Florida Bay, Florida)	KC309515, KC309589, KC309661
Odontella rhombus f. trigona (Cleve ex Van Heurck) R. Ross	HK340	ECT3891 (St. George Island, Florida)	KC309516, KC309590, KC309662
Odontella rostrata (Hustedt) Simonsen	HK225	ECT3828 (Redfish Bay, Texas)	JX413552, JX413569, JX413586
Odontella sp Agardh	HK335	ECT3854 (Kahana State Park, Oahu, Hawaii)	KC309517, KC309591, KC309663

Odontella sp "pseudoloc" Agardh	HK341	ECT3883 (Rainbow Harbor, Long Beach, California)	JX413555, JX413572, JX413589
Papiliocellulus simplex Gardner & Crawford	HK134	CS431 (CSIRO)	HQ912630, HQ912494, HQ912323
Plagiogrammopsis van heurckii (Grunow) Hasle, von Stosch & Syvertsen	HK293	ECT3885 (Elkhorn Slough, California)	KC309504, KC309578, KC309651
Planktoniella sol (Wallich) Schütt	HK035	CCMP1608	HQ912562, HQ912426, HQ912256
Pleurosira laevis (Ehrenberg) Compère	HK068	FD482 (UTEX)	HQ912585, HQ912449, HQ912278
Pleurosira laevis f. polymorpha Compère	HK326	ECT3833 (Corona Del Mar, California)	KC309505, KC309579, KC309652
Porosira glacialis (Grunow) E. Jorgensen	HK115	CCMP668	HQ912619, HQ912483, HQ912312
Pseudauliscus peruvianus (Kitton ex Pritchard) Schmidt	HK330	ECT3896 (Port Aransas Jetty, Texas)	KC309506, KC309580, KC309653
Roundia cardiophora (Round) Makarova	HK219	ECT3681 (Achang Reef, Guam)	KC284712, KC284708, KC284700
Stictocyclus stictodiscus (Grunow) R. Ross	HK269	GU44AB-8 (Gab Gab Beach, Guam)	KC309507, KC309581, KC309654
Synerdra bacillaris (Grunow) Hustedt	HK291	ECT3874 (Channel #5, US-1, Florida)	JX413548, JX413565, JX413582
Terpsinoë musica Ehrenberg	HK273	NHOP43 (Brackenridge Field Lab, Texas)	HQ912682, HQ912546, HQ912370
Thalassiosira pseudonana Hasle & Heimdal	HK008	CCMP1335	HQ912555, HQ912419, HQ912249
Trieres mobilensis (Bailey) Ashworth & Theriot	HK204	ECT3828 (Redfish Bay, Texas)	JX413549, JX413566, JX413583
Trieres mobilensis (Bailey) Ashworth & Theriot	HK227	ECT3834 (Alamitos Bay, California)	KC309499, KC309573, KC309646
Trieres mobilensis (Bailey) Ashworth & Theriot	HK251	Har-1 (HBOI Boat Dock, Florida)	KC309500, KC309574, KC309647
Trieres regia (Schultze) Ashworth & Theriot	HK322	Har-1 (HBOI Boat Dock, Florida)	KC309501, KC309575, KC309648
Trieres regia (Schultze) Ashworth & Theriot	HK290	27II10-3C (Port Aransas, Texas)	KC309502, KC309576, KC309649
Trieres sinensis (Greville) Ashworth & Theriot	HK323	ECT3886 (Bald Head Island, North Carolina)	KC309503, KC309577, KC309650
Trieres sinensis (Greville) Ashworth & Theriot	HK037	CCMP1815	HQ912564, HQ912428, HQ912258
Trigonium formosum (Brightwell) Frenguelli	HK200	ECT3689 (Saluglula Pools, Guam)	HQ912648, HQ912512, HQ912341
Trigonium formosum f. quadrangularis (Greville) Desikachary & Sreelatha	CX17	ECT3671 (Haputo Point, Guam)	KC309518, N/A, KC309664
Trigonium formosum (Brightwell) Frenguelli	HK258	University of Hawaii, Manoa culture	JX413550, JX413567, JX413584
Toxarium undulatum J.W. Bailey	HK210	ECT3802 (Gab Gab Beach, Guam)	HQ912654, HQ912518, HQ912347
Toxarium hennedyanum (Gregory)	HK220	ECT3648 (Asan Beach,	HQ912662, HQ912526,

Pelletan		Guam)	HQ912355
Triceratium bicorne Cleve	HK222	ECT3821 (Ward Island, Texas)	HQ912664, HQ912528, HQ912357
Triceratium dictyotum Ross & Sims	HK281	Pan-1 (STRI Station, Panama)	JX413554, JX413571, JX413588
Triceratium dubium Brightwell	HK199	ECT3767 (Taeleyag Beach, Guam)	HQ912647, HQ912511, HQ912340
Triceratium dubium Brightwell	HK254	ECT3838 (Long Beach, California)	KC309519, KC309592, KC309665
Triceratium dubium Brightwell	HK342	ECT3888 (San Sebastian, Canary Islands)	KC309520, KC309593, KC309666
Triceratium sp C.G. Ehrenberg	HK047	CCMP147	HQ912572, HQ912436, N/A
Urosolenia eriensis (H.L. Smith) Round & Crawford	HK054	Y98-8 (Yellowstone Lake, Wyoming)	HQ912577, HQ912441, HQ912270
Actinocyclus sp C.G. Ehrenberg	HK345	ECT3910 (Key Largo, Florida)	KC309521, KC309594, KC309667
Actinocyclus sp, "tinydrum" C.G. Ehrenberg	HK346	ECT3899 (Pacific Grove, California)	KC309522, KC309595, KC309668
Actinocyclus sp C.G. Ehrenberg	HK347	GU52-O (Outhouse Beach, Guam)	KC309523, KC309596, KC309669
Actinocyclus sp C.G. Ehrenberg	HK262	CVPan-4 (STRI Station, Panama)	KC309524, KC309597, KC309670
Actinocyclus subtilis (Gregory) Ralfs	HK168	ECT3672 (Haputo Point, Guam)	HQ912640, HQ912504, HQ912333
Actinoptychus sp C.G. Ehrenberg	HK050	SA19Ai4 (South Africa)	HQ912574, HQ912438, HQ912267
Actinoptychus undulatus (Kützing) Ralfs	HK261	17iii09-2F (Stillwater Cove, California)	KC309525, KC309598, KC309671
Aulacodiscus orientalis Greville	HK208	ECT3746 (Talofofo Bay, Guam)	HQ912652, HQ912516, HQ912345
Aulacodiscus oreganus Harvey & Bailey	HK348	ECT3897 (Stillwater Cove, California)	JX413556, JX413573, JX413590
Aulacoseira granulata (Ehrenberg) Simonsen	HK094	FD301 (UTEX)	HQ912606, HQ912470, HQ912299
Corethron hystrix Hensen	HK004	CCMP307	HQ912554, HQ912418, HQ912248
Corethron hystrix Hensen	HK241	Har-1 (HBOI Boat Dock, Florida)	KC309526, KC309599, KC309672
Corethron sp A.F. Castracane	HK356	ECT3920 (Ft. Stevens State Park, Oregon)	KC309527, KC309600, KC309673
*** *Coscinodiscus concinnus* W. Smith	HK267	ECT3839 (Port O'Connor, Texas)	HQ912681, HQ912545, HQ912369
Coscinodiscus granii Gough	HK228	ECT3836 (Rainbow Harbor, Long Beach, California)	HQ912667, HQ912531, HQ912359
Coscinodiscus granii Gough	HK349	ECT3923 (Olympia, Washington)	KC309528, KC309601, N/A
Coscinodiscus radiatus Ehrenberg	HK031	CCMP310	HQ912560, HQ912424, HQ912254

Coscinodiscus radiatus Ehrenberg	HK352	ECT3923 (Olympia, Washington)	KC309529, KC309602, KC309674
Coscinodiscus sp C.G. Ehrenberg	HK298	ECT3874 (Channel #5, US-1, Florida)	KC309530, KC309603, KC309675
Coscinodiscus sp C.G. Ehrenberg	HK263	CVPan-4 (STRI Station, Panama)	KC309531, KC309604, KC309676
Coscinodiscus sp C.G. Ehrenberg	HK350	ECT3891 (St. George Island, Florida)	KC309532, KC309605, KC309677
Coscinodiscus sp C.G. Ehrenberg	HK351	ECT3900 (Moss Landing Harbor, California)	KC309533, KC309606, N/A
Coscinodiscus wailesii Gran & Angst	HK229	ECT3847 (Port Aransas ferry crossing, Texas)	HQ912668, HQ912532, HQ912360
Endictya oceanica Ehrenberg	HK259	Pan-1 (STRI Station, Panama)	KC309534, KC309607, KC309678
Guinardia delicatula (Cleve) Hasle	HK205	ECT3821 (Ward Island, Texas)	HQ912651, HQ912515, HQ912344
Guinardia flaccida (Castracane) H. Peragallo	HK272	ECT3870 (Boca Chica Channel, Florida)	KC309535, KC309608, KC309679
Guinardia flaccida (Castracane) H. Peragallo	HK353	ECT3891 (St. George Island, Florida)	KC309536, KC309609, KC309680
Hyalodiscus stelliger J.W. Bailey	HK042	CCMP454	HQ912567, HQ912431, HQ912261
Hyalodiscus sp C.G. Ehrenberg	HK344	ECT3898 (Stillwater Cove, California)	KC309537, KC309610, KC309681
Leptocylindrus danicus Schutt	HK238	ECT3845 (Ward Island, Texas)	JX413558, JX413575, JX413592
Leptocylindrus minimus Gran	HK355	ECT3920 (Ft. Stevens State Park, Oregon)	KC309538, N/A, N/A
Melosira nummuloides (Agardh) Greville	HK041	CCMP482	HQ912566, HQ912430, HQ912260
Melosira varians Agardh	HK266	ECT3833 (Corona Del Mar, California)	KC309539, KC309611, KC309682
Paralia sulcata (Ehrenberg) Cleve	HK048	CCAP1059/1	HQ912573, HQ912437, HQ912266
Palmerina hardmanniana (Greville) Hasle	HK233	ECT3847 (Port Aransas ferry crossing, Texas)	HQ912671, HQ912535, HQ912363
Podosira sp C.G. Ehrenberg	HK201	ECT3681 (Achang Reef, Guam)	HQ912649, HQ912513, HQ912342
Proboscia sp. B.G. Sundstrom	HK300	ECT3891 (St. George Island, Florida)	KC309540, KC309612, KC309683
Proboscia sp. B.G. Sundstrom	HK236	ECT3845 (Ward Island, Texas)	KC309541, KC309613, KC309684
Pseudosolenia calcar-avis (Schultze) Sundstrom	HK245	Har-1 (HBOI Boat Dock, Florida)	KC309542, KC309614, KC309685
Rhizosolenia imbricata Brightwell	HK244	Har-1 (HBOI Boat Dock, Florida)	KC309543, KC309615, KC309686
Rhizosolenia formosa H. Peragallo	HK354	ECT3896 (Port Aransas Jetty, Texas)	JX413557, JX413574, JX413591
Rhizosolenia shrubsolei Cleve	HK221	ECT3821 (Ward Island, Texas)	HQ912663, HQ912527, HQ912356

Rhizosolenia setigera Brightwell	HK032	CCMP1820	HQ912561, HQ912425, HQ912255
Rhizosolenia setigera Brightwell	HK268	ECT3845 (Ward Island, Texas)	KC309544, KC309616, KC309687
Stellarima microtrias (Ehrenberg) Hasle & Sims	HK108	CCMP806	HQ912614, HQ912478, HQ912307
Stephanopyxis turris (Greville & Arnott) Ralfs	HK213	ECT3828 (Redfish Bay, Texas)	HQ912657, HQ912521, HQ912350
Bolidomonas pacifica L.Guillou & M.-J.Chrétiennot-Dinet	HK015	CCMP1866	HQ912557, HQ912421, HQ912251

References

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S., Brugerolle, G.U.Y., Fensome, R.A., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., McCourt, R.M., Mendoza, L., Moestrup, O., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A.V., Spiegel, F.W. and Taylor, M.F.J.R. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology* 52:399-451.

Alberte, R.S., Snyder, S., Zahuranec, B.J. and Whetstone, M. 1992. Biofouling research needs for the United States Navy: program history and goals. *Biofouling* 6: 91-95.

Alverson, A.J., Cannone J.J., Gutell R.R. and Theriot E.C. 2006. The evolution of elongate shape in diatoms. *Journal of Phycology*, 42: 655–668.

Andresen, N. A. 1995. A taxonomic investigation of a little known species of *Isthmia*. In: J. P. Kociolek and M. J. Sullivan (eds.), A centrury of diatom research in North America: A tribute to the distinguished careers of Charles W. Reimer and Ruth Patrick. Koeltz Scientific Books USA, Champaign, IL. pp. 103-110.

Anonymous. 1975. Proposal for a standardization of diatom terminology and diagnoses. *Nova Hedwigia Beihefte* 53: 323–354.

Bailey, J.W. 1851. Microscopical observations made in South Carolina, Georgia and Florida. *Smithsonian Contributions to Knowledge* 2:1-48.

Bedoshvili, Y.D., Popkova, T.P. and Likhoshway, Y.V. 2009. Chloroplast structure of diatoms of different classes. *Cell and Tissue Biology* 3: 297-310.

Brightwell, T. 1853. On the genus *Triceratium*, with descriptions and figures of the species. *Quarterly Journal of Microscopical Science, London* 1: 245-252.

Brightwell, T. 1856. Further observations on the genus *Triceratium*, with descriptions and figures of new species. *Quarterly Journal of Microscopical Science, London* 4:272-276.

Bruckner, C.G., Bahulikar, R., Rahalhar, M., Schink, B. and Kroth, P.G. 2008. Bacteria associated with benthic diatoms from Lake Constance: Phylogeny and Influences on diatom growth and secretion of extracellular polymeric substances. *Applied and Environmental Microbiology* 74: 7740-7749.

Bruckner, C.G., Rehm, C., Grossart, H., and Kroth, P.G. 2011. Growth and release of extracellular organic compounds by benthic diatoms depend on interactions with bacteria. *Environmental Microbiology* 13: 1052-1063.

Castracane, F. 1878. Nuovo genere e specie di Diatomea. Atti dell'Accademia Pontificia de'Nuovi Lincei, 31, pp. 1-11.

Castracane, F. 1886. Report on the Scientific Results of the Voyage of H. M. S. Challenger During the Years 1873–76: Botany, vol. 2, 178 pp., 30 pl. Her Majesty's Stationery Office, London.

Chepurnov, V. A., Mann, D. G., Sabbe, K. and Vyverman, W. 2004. Experimental studies on sexual reproduction in diatoms. *International Review of Cytology* 237:91–154.

Conesa, A., Götz, S., Garcia-Gomes, J.M., Terol, J., Talon, M., & Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

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

Daniele, A., Parenti, G., d'Addio, M., Andria, G., Ballabio, A. and Meroni, G. 1998. Biochemical characterization of arylsulfatase e and functional analysis of mutations found in patients with X-linked chondrodysplasia punctate. *American Journal of Human Genetics* 62: 562-572.

de Hostos, E.L., Togasaki, R.K. and Grossman, A. 1988. Purification and biosynthesis of a depressible periplasmic arylsulfatase from Chlamydomonas reinhardtii. *Journal of Cell Biology* 106: 29-37.

De Toni, G.B. 1894. Sylloge algarum omnium hucusque cognitarum. Vol. II. Bacillarieae; sectio III. Cryptoraphideae. Typis Seminarrii, Patavii. pp. 819-1556.

Deschamps, P., Lara, E., Marande, W., López-García, P., Ekelund, F., and Moreira, D. 2011. Phylogenomic Analysis of Kinetoplastids Supports That Trypanosomatids Arose from within Bodonids. *Molecular Biology and Evolution* 28: 53-58.

Drebes, G. and Schultz, D. 1981. Anaulus creticus sp. nov., a new centric diatom from the Mediterranean Sea. *Bacillaria* 4: 161-176.

Drum, R.W. 1969. Light and electron microscope observations on the tube-dwelling diatom *Amphipleura rutilans* (Trentepohl) Cleve. *Journal of Phycology*. 5: 21-26.

Durkin, C.A., Mock, T. and Armbrust, E.V. 2009. Chitin in diatoms and its association with the cell wall. *Eukaryotic Cell* 8: 1038-1050.

Embley, T.M., Finlay, B.J, Dyal. P.L., Hirt, R.P., Wilkinson, M., and Williams, A.G.. 1995. Multiple Origins of Anaerobic Ciliates with Hydrogenosomes within the Radiation of Aerobic Ciliates. *Proceeding of the Royal Society B* 262: 87-93.

Fontana, P., Cestaro, A., Velasco, R., Formentin, E. and Toppo, S. 2009. Rapid annotation of anonymous sequences from genome projects using semantic similarities and a weighting scheme in gene ontology. PLOS ONE 4(2): e4619.

Glezer, Z.I. 1979. Evolution and systematics of the order Biddulphiales (Bacillariophyta). *Paleontological Journal* 13:103-111 (translated from Russian).

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* 29: 644-652.

Greville, R.K. 1863. Descriptions of new and rare diatoms. Series IX. *Transactions of the Microscopical Society of London* 11: 63-76, plates IV-V.

Guillard, R.R.L. 1975. Culture of phytoplankton for feeding marine invertebrates. In: Smith, W.L. and Chanley, M.H. (Eds) Culture of marine invertebrate animals. Plenum Press, New York, pp. 26–60.

Hasle, G.R., von Stosch H.A. and Syvertsen, E.E. 1983. Cymatosiraceae, a new diatom family. *Bacillaria* 6: 9-156.

Hein, M.K., Winsborough, B.M. and Sullivan, M.J. 2008. Bacillariophyta (diatoms) of the Bahamas. Iconogr. Diatom. 19. Gantner Verlag, Ruggell, Germany.

Hendey, N.I. 1971. Some marine diatoms from the Galapagos Islands. *Nova Hedwigia* 22:371-422.

Hoagland, K.D., Rosowski, J.R., Gretz, M.R., and Roemer, S.C. 1993. Diatom extracellular polymeric substances: function, fine structure, chemistry and physiology. *Journal of Phycology* 29: 537-566.

Hoban, M.A. 1979. Morphology, life histories and systematics of selected genera of biddulphioid diatoms. Ph. D. dissertation, University of Texas at Austin, USA.

Johnson, L.M., Hoagland, K.D. and Gretz, M.R. 1995. Effects of bromide and iodide on stalk secretion in the biofouling diatom *Achnanthes longipes* (Bacillariophyceae). *Journal of Phycology* 31: 401-412.

Kilroy, C., Snelder, T.H., Floerl, O., Vieglais, C.C., and Dey, K.L. 2008. A rapid technique for assessing the suitability of areas for invasive species applied to New Zealand's rivers. *Diversity and Distributions* 14: 262-272.

Komura, S. 1999. Barrel-shaped diatoms from the Miocene Nabuto Formation, central Japan. *Diatom* 15:51-78.

Kooistra, W.H.C.F., De Stefano, M., Mann, D.G., Salma, N. and Medlin, L.K. 2003. Phylogenetic position of *Toxarium*, a pennate-like lineage within centric diatoms (Bacillariophyceae). *Journal of Phycology*, 39: 185–197.

Kooistra, W.H.C.F, Forlani, G., Sterrenburg, F.A.S. and De Stefano, M. 2004. Molecular phylogeny and morphology of the marine diatom *Talaroneis posidoniae* gen. et sp. nov. (Bacillariophyta) advocate the return of the Plagiogrammaceae to the pennate diatoms. *Phycologia* 43: 58-67.

Lobban, C.S. and Jordan, R.W. 2010. Diatoms on coral reefs and in tropical marine lakes. In: J.P. Smol and E.F. Stoermer (Eds.) The diatoms: applications for the environmental and earth sciences (2nd ed.). Cambridge University Press, Cambridge, U.K. pp. 346-356.

Lobban, C.S., Ashworth, M.P. and Theriot, E.C. 2010. *Climaconeis* species (Bacillariophyceae: Berkeleyaceae) from western Pacific islands, including *C. petersonii* sp. nov. and *C. guamensis* sp. nov., with emphasis on the plastids. *European Journal of Phycology* 45: 293-307.

Lobban, C.S., Ashworth, M.P., Arai, Y., Jordan, R.W. and Theriot, E.C. 2011. Marine necklace-chain Fragilariaceae (Bacillariophyceae) from Guam, including descriptions of *Koernerella* and *Perideraion*, genera nova. *Phycological Research* 59: 174-193.

Losos, J.B., Hillis, D.M. and Greene, H.W. 2012. Who speaks with a forked tongue? *Science* 338: 1428-1429.

Lyngbye, H.C. 1819. Tentamen Hydrophytologiae Danicae Continens omnia Hydrophyta Cryptogama Daniae, Holsatiae, Faeroae, Islandiae, Groenlandiae hucusque cognita, Systematice Disposita, Descripta et iconibus illustrata, Adjectis Simul Speciebus Norvegicis. Hafniae. 248 pp., 70 pls.

Maddison, W. P. and Maddison, D.R. 2010. Mesquite: a modular system for evolutionary analysis. Version 2.74 http://mesquiteproject.org

Magne, F. 1976. Quelques caractères cytologiques particuliers du *Bachelotia antillarum* (Phéophycées, Ectocarpales). *Phycologia* 15: 309-319.

Mckay, R.M.L. and Gibbs, S.P. 1991. Composition and function of pyrenoids: cytochemical and immunochemical approaches. *Canadian Journal of Botany* 69: 1040-1052.

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

Medlin, L.K., Sato, S., Mann, D.G. and Kooistra, W.H.C.F. 2008. Molecular evidence confirms sister relationship of *Ardissonea*, *Climacosphenia* and *Toxarium* within the bipolar centric diatoms (Bacillariophyta, Mediophyceae), and cladistic analyses confirm that extremely elongate shape has arisen twice in the diatoms. *Journal of Phycology* 45: 1340–1348.

Medlin, L.K. 2010. Pursuit of a natural classification of diatoms: an incorrect comparison of published data. *European Journal of Phycology* 45: 155-166.

Mereschkowsky, C. 1901. Etjudy nad endokhromom diatomovykh vodoroslei (po frants). (Étude sur l'endochrome des Diatomées.) I. Zapiski Imperatorskoj Akademii Nauk po Fiziko-matematiceskomee Otdeleniju. *Memoires de l'academie imperiale des sciences de Saint Petersbourg, Classe physico-mathematique Series* 8, 11(6):1–40, 7 pl.

Müller O. 1886. Die Zwischenbänder und Septen der Bacillariaceen. *Berichte der Deutsche Botanisches Gesellschaft*. 6: 306-316.

Navarro, J.N. 1981. A survey of the marine diatoms of Puerto Rico II. Suborder Biddulphiineae: Families Biddulphiaceae, Lithodesmiaceae and Eupodiscaceae. *Botanica Marina* 24:615-630.

Navarro, J.N. 1982. A survey of the marine diatoms of Puerto Rico. IV. Suborder Araphidinae: Families Diatomaceae and Protoraphidaceae. *Botanica Marina* 25: 247–263.

Navarro, J.N. 1996. New observations on the araphid marine diatom *Florella portoricensis* (Florellaceae fam. nov.). *Diatom Research* 11: 207-304.

Navarro, J.N. 2002. *Florella pascuensis* sp. nov., a new marine diatom species from Easter Island (Isla de Pascua), Chile. *Diatom Research* 17: 283-289.

Navarro, J.N. and Lobban, C.S. 2009. Freshwater and marine diatoms from the western Pacific islands of Yap and Guam, with notes on some diatoms in damselfish territories. *Diatom Research* 24: 123–157.

Nikolaev, V.A. 1990. The system of centric diatoms. In: H. Simola (ed.), Proceedings of the Tenth International Diatom Symposium, Joensuu, Finland, August 28 - September 2, 1988. Koeltz Scientific Books, Koenigstein. pp. 103-110.

Okada, M. 1992. Recent studies on the composition and the activity of algal pyrenoids. In: Round, F.E. & Chapman, D.J. [eds.] *Progress in Phycological Research* 8: 117-138. Biopress, Bristol, UK.

Ouriques, L.C. and Bouzon, Z.L. 2000. Stellate chloroplast organization in *Asteronema breviarticulatum* comb. nov. (Ectocarpales, Phaeophyta). *Phycologia* 39: 267-271.

Ostenfeld, C.H. 1908. On the immigration of *Biddulphia sinensis* Grev. and its occurrence in the North Sea during 1903-1907 and on its use for the study of the direction and rate of flow of the currents. *Meddelelser fra Kommissionen for Danmarks Fiskeri-Og Havundersøgelser. Serie: Plankton (København)* 1:1-44.

Østrup, E. 1904. Marine Diatoms. In: Schmidt, J (Ed.). Flora of Koh Chang. Contributions to the knowledge of the vegetation in the Gulf of Siam. Part VIII. *Botanisk Tidsskrift* 26(1):115-161, pls. 1 and 2.

Peragallo, H. and Peragallo, M. 1897–1908. Diatomées marines de France et des districts maritimes voisins. M.J. Tempère, Grez-sur-Loing, 492 pp.

Raines, C.A., Lloyd, J.C. and Dyer, T.A. 1999. New insights into the structure and function of sedoheptulose-1,7-bisphosphatase; and important but neglected Calvin cycle enzyme. *Journal of Experimental Botany* 50: 1-8.

Roper, F.C.S. 1859. On the genus *Biddulphia* and its affinities. *Transcriptions of the Microscopical Society of London* 7:1-24.

Rosowski, J.R. 1980. Band and valve morphology of some freshwater diatoms. II. Integration of valves and bands in Navicula confervacea. *Journal of Phycology*. 16: 88-101.

Ross, R. and Sims, P.A. 1971. Generic limits in the Biddulphiaceae as indicated in the scanning electron microscope. In: V.H. Heywood (Ed.) Scanning electron microscopy, systematic and evolutionary application. Academic Press, London. Pp. 155-177.

Ross, R. and Sims, P.A. 1973. Observations on family and generic limits in the Centrales. *Nova Hedwigia Beiheft* 45:97-132.

Ross, R. and Sims, P.A. 1987. Further genera of the Biddulphiaceae (diatoms) with interlocking linking spines. *Bulletin of the British Museum (Natural History), Botany series* 16 (4):269-311.

Ross, R., Sims, P.A. and Hasle, G.R. 1977. Observations on some species of Hemiauloideae. *Nova Hedwigia Beiheft* 54:179-203.

Rost, B. 2002. Enzyme function less conserved than anticipated. *Journal of Molecular Biology* 318:595-608.

Round, F.E. and Mann, D.G. 1980. *Psammodiscus* nov. gen. based on *Coscinodiscus nitidus*. *Annals of Botany* 46: 368-373.

Round, F.E., Crawford, R.M. and Mann, D.G. 1990. The diatoms: biology and morphology of the genera. Cambridge University Press, Cambridge, UK. 747 pp.

Ruck, E.C., and Theriot E.C. 2011. Origin and evolution of the canal raphe system in diatoms. *Protist* 162: 723–737.

Sar, E.A., Sunesen, I., Jahn, R., Kusber, W-H. and Lavigne, A.S. 2007. Revision of *Odontella atlantica* (Frenguelli) Sar *comb. et stat. nov*. with comparison to two related species, *O. rhombus* (Ehrenb.) Kütz. and *O. rhomboides* R. Jahn et Kusber. *Diatom Research* 22:341-353.

Schmid, A-M.M. 2001. Value of pyrenoids in the systematics of the diatoms: their morphology and ultrastructure. In A. Economou-Amilli [Ed.] Proceedings of the 16th International Diatom Symposium. Amvrosiou Press, Athens, Greece, pp 1-32.

Schultze, M. 1859. Phenomena of internal movements in Diatomaceae of the North Sea, belonging to the genera *Coscinodiscus*, *Denticella*, *Rhizosolenia*. *Quarterly Journal of Microscopical Science* 7:13-21.

Schütt, F. 1896. Bacillariales (Diatomeae). In: Engler, A. and Prantl, K. (Eds.) Die natürlichen Pflanzenfamilien. Verlag von Wilhelm Engelmann. Leipzig. I(1b):31-150.

Shimodaira, H. and Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16:1114–1116

Simonsen R. 1974. The diatom plankton of the Indian Ocean Expedition of R/V Meteor 1964-5. "Meteor" Forschungsergebnisse. *Reihe D: Biologie* 19:1-107.

Simonsen, R. 1979. The diatom system: ideas on phylogeny. *Bacillaria* 2: 9–71.

Sims, P.A. and Ross, R. 1990. *Triceratium pulvinar* and *T. unguiculatum*, two confused species. *Diatom Research* 5:155-169.

Smith, D.J and Underwood, G.J.C. 2000. The production of extracellular carbohydrates by estuarine benthic diatoms: the effects of growth phase and light and dark treatment. *Journal of Phycology* 36: 321-333.

Sperling, E.A., Peterson, K.J. and Pisani, D. 2009. Phylogenetic-signal dissection of nuclear housekeeping genes supports the paraphyly of sponges and the monophyly of Eumetazoa. *Molecular Biology and Evolution* 26: 2261–2274.

Stamatakis, A. 2006. RaxML-VI-HPC; Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22 (21): 2688-2690.

von Stosch, H.A. 1975. An amended terminology of the diatom girdle. *Nova Hedwigia Beihefte* 53: 1-36.

von Stosch, H.A. and Simonsen, R. 1984. *Biddulphiopsis*, a new genus of the Biddulphiaceae. *Bacillaria* 7:9-36.

Strelnikova, N.I., Fourtanier, E., Kociolek, J.P. and Barron, J.A. 2004. A SEM study of the diatom genus *Porodiscus* Greville: morphology of the species and comparison with related genera. *Proceedings of the California Academy of Sciences* 55: 300-337.

Theriot, E.C., Cannone, J.J., Gutell, R.R., and Alverson, A.J. 2009. The limits of nuclear-encoded SSU rDNA for resolving the diatom phylogeny. *European Journal of Phycology* 44:277-290.

Theriot, E.C., Ashworth, M., Ruck, E., Nakov, T. and Jansen, R.K. 2010. A preliminary multigene phylogeny of the diatoms (Bacillariophyta): challenges for future research. *Plant Ecology and Evolution* 143: 1-18.

Theriot, E.C., Ruck, E., Ashworth, M., Nakov, T. and Jansen, R.K. 2011. Status of the pursuit of the diatom phylogeny. In: J. Seckbach and J.P. Kociolek [Eds] The Diatom World. Springer Science+Business Media. pp 119-142.

Tian, W. and Skolnick, J. 2003. How well is enzyme function conserved as a function of pairwise sequence identity? *Journal of Molecular Biology* 333: 863-882.

Thornton, D.C.O. 2002. Diatom aggregation in the sea: mechanisms and ecological implications. *European Journal of Phycology* 37: 149-161.

Van Heurck, H. 1880–1885. Synopsis des Diatomées de Belgique. Jucaju & Cie., Anvers.

Van Heurck, H. 1896. A treatise on the Diatomaceae. London, UK: Wesley & Son.

Underwood, G.J.C., Boulcott, M., Raines, C.A. and Waldron, K. 2004. Environmental effects on exopolymer production by marine benthic diatoms: dynamics, changes in composition, and pathways of production. *Journal of Phycology* 40: 293-304.

Wallich, G.C. 1858. On *Triceratium* and some new allied forms. *Quarterly Journal of Microscopical Science, London* 6:242-253.

Walne, P.A., Moestrup, Ø., Norris, R. and Ettl, H. 1986. Light and electron microscopical studies of *Eutreptiella eupharynea* sp. nov. (Euglenophyceae) from Danish and American waters. *Phycologia* 25: 109-126.

Wang, Y., Chen,Y., Lavin, C., and Gretz, M.R. 2000. Extracellular matrix assembly in diatoms (Bacillariophyceae). IV. Ultrastructure of *Achnathes longipes* and *Cymbella cistula* as revealed by high-pressure freezing/freeze substitution and cryo-field emission scanning electron microscopy. *Journal of Phycology* 36: 367-378.

Webster, D.R., Cooksey, K.E., and Rubin, R.W. 1985. An investigation of the involvement of cytoskeletal structures and secretion in gliding motility of the marine diatom *Amphora coffeaeformis*. *Cell Motility* 5: 103-122.

Wetzel, R.G. 1964. A comparative study of the primary productivity of higher aquatic plants, periphyton and phytoplankton in a large, shallow lake. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* 49: 1-64.

Wilgenbusch, J.C., Warren, D.L. and Swofford, D.L. (2004). AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. Available at: http://ceb.csit.fsu.edu/awty.

Williams, D.M. 1988. An illustrated catalogue of the type specimens in the Greville diatom herbarium. *Bulletin of the British Museum (Natural History), Botany Series*18:1-148.

Williams, D.M. and Kociolek, J.P. 2007. Pursuit of a natural classification of diatoms: History, monophyly and the rejection of paraphyletic taxa. *European Journal of Phycology* 42: 313-319.

Witkowski, A., Lange-Bertalot, H. and Metzelin, D. 2000. *Diatom flora of marine coasts* I. A.R.G. Gantner Verlag K.G., Ruggell, Germany. 925 pp.

Wustman, B.A., Gretz. M.R., and Hoagland, K.D. 1997. Extracellular matrix assembly in diatoms (Bacillariophyceae). I. A model of adhesives based on chemical characterization and localization of polysaccharides from the marine diatom *Achnanthes longipes* and other diatoms. *Plant Physiology* 113: 1059-1069.

Wustman, B.A., Lind, J., Wetherbee, R., and Gretz, M.R. 1998. Extracellular matrix assembly in diatoms (Bacillariophyceae). III. Organization of fucoglucuronogalactans within the adhesive stalks of *Achnanthes longipes*. *Plant Physiology* 116: 1431-1441.