A consensus secondary structure of ITS2 for the diatom Order Cymatosirales (Mediophyceae, Bacillariophyta) and reappraisal of the order based on DNA, morphology, and reproduction

Brajogopal Samanta a,⁎, James M. Ehrman b, Irena Kaczmarska a

a Department of Biology, Mount Allison University, Sackville, New Brunswick E4L 1G7, Canada
b Digital Microscopy Facility, Mount Allison University, Sackville, New Brunswick E4L 1G7, Canada

ARTICLE INFO

Keywords:
Cymatosirales
Cymatosiraceae
Diatoms
ITS2
Leyanellaceae
Phylogeny

ABSTRACT

In 1983, Hasle and colleagues removed cymatosiroid diatoms from the pennates, and erected a new centric diatom family, the Cymatosiraceae, mainly to accommodate for their newly discovered mode of sexual reproduction. The new family consisted of two subfamilies differing in frustule structure. The family was later elevated to the rank of Order Cymatosirales Round and Crawford. We revisited intra-ordinal relationships within Cymatosirales using combined genetic (DNA sequences), morphological (valve and frustule structure), and reproductive (auxospore type) characters. In total, 36 cymatosiroid strains from 19 species representing 13 genera (80% of the total number of extant genera; nine of them represented by their generitypes) were used in this study. Instead of only the commonly used loci (18S rRNA and plastidal genes) to infer diatom phylogeny, we developed a consensus secondary structure model of the Internal Transcribed Spacer 2 (ITS2) for this order and applied it to aid in sequence alignments for ITS2. This improved the alignment and thus the robustness of the phylogenetic framework. The compensatory base changes (CBCs) found in ITS2 secondary structures were mapped onto the multi-gene (18S rRNA + ITS2 + rbcL) phylogenetic tree topology. In all these trees, all species grouped into two morphologically and genetically distinct clades. Each clade was supported by multiple CBCs, as did all the clades representing genera. However, these clades did not correspond to the previously established subfamilies. Consequently, we amend the Order Cymatosirales and family Cymatosiraceae, and propose a new family, the Leyanellaceae. The structure of the auxospore was an additional synapomorphic character for Cymatosirales. Overall, we demonstrate a novel approach to study diatom phylogeny across a broader taxonomic range using ITS2 secondary structural information. Our results suggest that this approach might be useful in establishing higher taxonomic relationships in other groups of diatoms.

1. Introduction

Diatoms are one of the most diverse groups of unicellular eukaryotic protists. They are encased within architecturally complex siliceous walls, the morphological features of which have long been used as a basis for their identification and classification (Round et al., 1990). Technological advancements have revolutionized diatom systematics over the last 40 years; as a result, several taxonomic reappraisals have been proposed at all taxonomic levels (e.g., Medlin and Kaczmarska, 2004; Round et al., 1990; Simonsen, 1979). Importantly, the advent of scanning electron microscopy (SEM) in diatom systematics revealed taxonomically important ultrastructural features in the diatom cell wall that were not visible using light microscopy (Round et al., 1990). More recently, molecular and reproductive data have added more information to consider in diatom systematics (e.g. Ashworth et al., 2013; Li et al., 2015; Lim et al., 2018; Medlin, 2016; Medlin and Kaczmarska, 2004; Parks et al., 2018). In diatom phylogenetic trees, monophyletic clades are often poorly supported by synapomorphic valve morphological characters, with many characters thus far examined being plesiomorphic. For example, recent studies with the diatom families Plagiogrammaceae, Biddulphiaceae, and Eupodiscaceae have shown that phylogenetic lineages within these families were not always consistent with the morphological synapomorphies (Ashworth et al., 2013; Li et al., 2015). Moreover, the higher taxonomic level relationships among some diatom families and orders remain insufficiently understood, as their positions in tree topologies and statistical support for nodes differ depending on the analyses and taxa included (Ashworth et al., 2013; Medlin, 2016; Medlin and

⁎ Corresponding author.
E-mail address: brajomicro@gmail.com (B. Samanta).
https://doi.org/10.1016/j.ympev.2018.08.014
Received 18 May 2018; Received in revised form 18 August 2018; Accepted 23 August 2018
Available online 25 August 2018
1055-7903/ © 2018 Elsevier Inc. All rights reserved.
Kaczmarska, 2004; Sorbans, 2007, 2004; Theriot et al., 2015, 2010). These higher-level relationships might not be resolvable by genes thus far used and molecular approaches alone (Parks et al., 2018; Theriot et al., 2015).

The Internal Transcribed Spacer 2 (ITS2) is a fast-evolving part of the eukaryotic nuclear-encoded RNA operon located between the 5.8S and 28S rRNA genes. The inclusion of the ITS2 secondary structure improves the accuracy and robustness of phylogenetic tree reconstruction (Keller et al., 2010). ITS2 has been shown effective in distinguishing cryptic/pseudo-cryptic species via compensatory base changes (or CBCs, Coleman, 2009, 2000; Mülller et al., 2007; Wolf et al., 2013), thus resulting in its frequent use to resolve lower taxonomic affinities in many eukaryotic lineages including plants (Adedowale et al., 2016), animals (Coleman and Vacquier, 2002; Ruhl et al., 2010), fungi (Ahvenniemi et al., 2009; Schoch et al., 2012), and diatoms (Amato et al., 2007; Balzano et al., 2017; Casteleyn et al., 2008; Franco et al., 2016; Kaczmarska et al., 2014; Lim et al., 2018, 2013; MacGillivary and Kaczmarska, 2012; Percopo et al., 2016; Poulicková et al., 2010; Wang et al., 2012).

However, it is important to note that a typical eukaryotic ITS2 sequence folds into an open palm-like secondary structure with typically four helices and highly conserved motifs (Coleman, 2007, 2003; Mai and Coleman, 1997; Schultz et al., 2005). ITS2 research grows rapidly and currently several hundred thousand sequence-structure pairs are available in the ITS2 database I-V (Ankenbrand et al., 2015; Koetschan et al., 2012; Schultz et al., 2006, 2005; Selig et al., 2008; Wolf et al., 2005a). The formation of the secondary structure of ITS2 within the assemblage pre-ribosomal particle is a prerequisite for accurate and efficient pre-rRNA processing (Côté and Peculis, 2001; Fromm et al., 2017; Peculis and Greer, 1998). The common-core ITS2 secondary structures present in all eukaryotes might be necessary for cell function and thus will be evolutionarily conserved (Coleman, 2007; Côté et al., 2002). This secondary structural information was demonstrated to be useful for phylogenetic studies at taxonomic ranks higher than species in green algae and flowering plants (Mai and Coleman, 1997). Later, Coleman (2003) proposed that ITS2 sequences coupled with their secondary structures could be a marker suitable for evolutionary comparisons of taxa of even greater taxonomic range (subspecies to order level), and thus more informative than other gene markers from the same organism. In advancement of sequence-structure alignment, Wolf and colleagues introduced a new approach. The sequences and their individual secondary structures can be aligned using a specific scoring matrix, fitted to a 12-letter alphabet (pseudo-proteins) encoding the sequence structure information (Schultz and Wolf, 2009; Seibel et al., 2006, 2008; reviewed in Wolf et al., 2014). This approach increased robustness of the phylogenetic relationships recovered because they could be reconstructed based on the simultaneous consideration of primary sequence and secondary structure information (Keller et al., 2010). Several studies demonstrated that the secondary structure of ITS2 provided a well-supported background for phylogenetic application at high taxonomic and evolutionary comparisons in green algae (Buchheim et al., 2012, 2011; Caisová et al., 2013, 2011; Heeg and Wolf, 2015), terrestrial plants (Coleman, 2003; Merget and Wolf, 2010), some protists (Miao et al., 2008), and different animal lineages (Aguilar and Sánchez, 2007; Young and Coleman, 2004). Green algae in particular revealed novel and unanticipated insights into processes underlining ITS2 evolution and taxonomic significance of ITS2 characters (Caisová et al., 2013, 2011), which were surprising even to the authors. A similar approach is yet to be applied to higher taxa in diatoms.

In 1983, Hasle and colleagues proposed the Cymatosiraceae as a new centric diatom family. They removed this group from its earlier placement among the pennates (Hustedt, 1959 [1985]), and included it into the polar centrics based on their discovery that cymatosiraceans reproduced oogamously and produced uniflagellate sperm, as did all centrics studied to that date. They further subdivided the Cymatosiraceae into two subfamilies, Cymatosiroidae and Eutubocelluloideae, based on their frustule morphology (Hasle et al., 1983). The family was later elevated to the rank of Order Cymatosirales (Round et al., 1990). Recent molecular phylogenies based on three gene markers (18S rRNA, rbcL, and psbA) have supported the Order Cymatosirales as monophyletic within the Mediophyceae (Ashworth et al., 2013; Dąbek et al., 2017). This Order has been recovered as sister to a large, poorly supported clade containing a number of polar centric diatom orders and families, e.g., Thalassiosirales, Lutheosimusales, and Eupodiscaceae, in addition to some members of Chaetocerotales (Ashworth et al., 2013; Dąbek et al., 2017; Medlin, 2017) suggesting their phylogenetic relationship with other mediophycean orders and families remains unclear. Furthermore, Cymatosirales divided into two strongly supported subclades (Dąbek et al., 2017; Samanta et al., 2017) which did not agree with the delineation of subfamilies proposed by Hasle and colleagues (Hasle et al., 1983). Therefore, further investigation is required to clarify the evolutionary relationship among the Order members.

In this study, for the first time in diatoms, we apply ITS2 secondary structure to infer higher level taxonomic relationships in diatoms by selecting a widely distributed, morphologically diverse, but relatively small group of extant representatives of the Order Cymatosirales. We then compare ITS2 based phylogeny to those obtained from other, widely used markers (rbcL, 18S, and morphology). Our specific objectives were: (i) develop a consensus ITS2 secondary structure model for the Order Cymatosirales; (ii) search for molecular evolutionary changes (compensatory base changes or CBCs and hemi-CBCs or hCBCs) in the ITS2 helices; (iii) map all the molecular evolutionary changes and key morphological characters on the individual and concatenated multigene trees; and (iv) revise the current systematics of the Order, if appropriate.

2. Materials and methods

2.1. Cymatosiroid diatom strains and microscopy

In total, the morphology and molecular information from 36 cymatosiroid strains from 19 species representing 13 genera (80% of the total number of recently known extant genera) were used in all analyses (Table 1). From these, 18 strains from 13 species representing 10 genera were used for ITS2 secondary structure analysis. Twelve strains were obtained from the National Center for Marine Algae and Microbiota (NCMA, Bigelow, USA; https://ncma.bigelow.org/; names with CCMP numbers). Eight strains were obtained from the Roscoff Culture Collection (RCC, Station Biologique de Roscoff, France; http://roscoff-culture-collection.org/; names with RCC numbers). Finally, monoclonal cultures of Plagiogrammopsis (14 clones) and Cymatosira (3 clones) were established in our laboratory by the micropipetting method (Andersen, 2005) from the intertidal mud-flats of the Bay of Fundy, Canada. Additionally, cell pellet and genomic DNA of Pseudoleanyella lunata NG0001 was provided by the authority of this taxon (Nakamura et al., 2016). The isolates were grown in f/2 medium in a controlled 12:12 h light:dark photoperiod at 15 °C and 26 μmol photons m−2 s−1 irradiance. The CCMP and RCC strains were maintained in our laboratory under growth conditions and in media recommended by the culture providers. All strains were examined using scanning electron microscopy (SEM) to confirm their morphological identity at the beginning of the study. For SEM examination samples were prepared following Kaczmarska et al. (2005) and observed with a Hitachi SU3500 SEM (Hitachi High-Technologies Canada, Inc., Toronto, Canada), operating at 10 kV accelerating voltage and 5 mm working distance.

2.2. DNA extraction, amplification, and sequencing

Cultured cells were harvested from exponential growth phase and biomasses were concentrated by centrifugation. DNA was extracted
<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Taxon name</th>
<th>Strain</th>
<th>Accession numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td><em>B. brockmannii</em></td>
<td>CCMF151</td>
<td>MH129006 MF140305 MF140314 MF140296 BRR010-17</td>
</tr>
<tr>
<td>2.</td>
<td>B. brockmannii</td>
<td>HK040</td>
<td>– KB284711 KB284711 KB284707 –</td>
</tr>
<tr>
<td>3.</td>
<td><em>Plagiogrammosis vanheurckii</em></td>
<td>BS,Cymato_C11</td>
<td>MH129007 MH138039 MH138054 MH141537 PLV015-17</td>
</tr>
<tr>
<td>4.</td>
<td>P. vanheurckii</td>
<td>ECT3885</td>
<td>– KC390504 KC390504 KC390578 –</td>
</tr>
<tr>
<td>5.</td>
<td>P. vanheurckii</td>
<td>ECT3886</td>
<td>– KJ577870 KJ577870 KJ577907 –</td>
</tr>
<tr>
<td>6.</td>
<td><em>Cymatoa bellica</em></td>
<td>JM,Cymato_C14</td>
<td>MH129008 MH138040 MH138055 MH141538 CYMT001-18</td>
</tr>
<tr>
<td>7.</td>
<td>C. bellica</td>
<td>CCMF345</td>
<td>EF192986 JX477413 MH141539 CYMT002-18</td>
</tr>
<tr>
<td>8.</td>
<td>C. lorentziana</td>
<td>SZCZH114</td>
<td>KU556408 KU556408 KU556401 –</td>
</tr>
<tr>
<td>9.</td>
<td>C. lorentziana</td>
<td>ECT3874</td>
<td>– KC390490 KC390490 KC390562 –</td>
</tr>
<tr>
<td>10.</td>
<td>Campylostria cymbeliformis</td>
<td>–</td>
<td>CCI1 – HQ912623 HQ912623 HQ912487 –</td>
</tr>
<tr>
<td>11.</td>
<td><em>Lambertocellus africana</em></td>
<td>SZCZH74</td>
<td>– KU556409 KU556409 KU556402 –</td>
</tr>
<tr>
<td>12.</td>
<td><em>Leyanelia arenaaria</em></td>
<td>RCG1863</td>
<td>MH129009 MH138041 MH138056 MH141540 CYMT003-18</td>
</tr>
<tr>
<td>13.</td>
<td>L. pasciperlus</td>
<td>X14X15</td>
<td>– KU556412 KU556412 KU556405 –</td>
</tr>
<tr>
<td>14.</td>
<td>L. probus</td>
<td>SZCZH588</td>
<td>– KU556411 KU556411 KU556404 –</td>
</tr>
<tr>
<td>15.</td>
<td>L. probus</td>
<td>ECT3888</td>
<td>– KU556410 KU556410 KU556403 –</td>
</tr>
<tr>
<td>16.</td>
<td>L. probus</td>
<td>ECT3888</td>
<td>– KU556410 KU556410 KU556403 –</td>
</tr>
<tr>
<td>17.</td>
<td>Pseudoleyanella lunata</td>
<td>NG0001</td>
<td>MH129011 MH138043 MH138057 MH141541 CYMT005-18</td>
</tr>
<tr>
<td>18.</td>
<td>P. philippinensis</td>
<td>CCMF3125</td>
<td>MH129012 MH138044 MH138058 MH141542 CYMT006-18</td>
</tr>
<tr>
<td>19.</td>
<td>P. simplex</td>
<td>RCC2624</td>
<td>MH129013 MH138045 MH138059 MH141543 CYMT007-18</td>
</tr>
<tr>
<td>20.</td>
<td>P. simplex</td>
<td>CS431</td>
<td>– HQ912630 HQ912630 HQ912484 –</td>
</tr>
<tr>
<td>21.</td>
<td>Arcocellulus mammifer</td>
<td>CCMF132</td>
<td>MH129014 HQ912569 MH138057 MH141544 CYMT008-18</td>
</tr>
<tr>
<td>22.</td>
<td>A. coruscus</td>
<td>RCC2270</td>
<td>MH129016 MH138046 MH138061 MH141544 CYMT009-18</td>
</tr>
<tr>
<td>23.</td>
<td>A. coruscus</td>
<td>RCC3057</td>
<td>MH129015 MH138047 MH138062 MH141545 CYMT010-18</td>
</tr>
<tr>
<td>24.</td>
<td>Minutocellus polymorphus</td>
<td>CCMF499</td>
<td>MH129017 MH138048 MH138063 MH141546 CYMT011-18</td>
</tr>
<tr>
<td>25.</td>
<td>M. polymorphus</td>
<td>RCC741</td>
<td>MH129018 MH138049 MH138064 MH141547 CYMT012-18</td>
</tr>
<tr>
<td>27.</td>
<td>M. polymorphus</td>
<td>CCMF499</td>
<td>– AY485478 AY485478 AY4952432 CYMT014-18</td>
</tr>
<tr>
<td>28.</td>
<td>Minutocellus cf. sp.</td>
<td>CCMF1701</td>
<td>– AY485520 AY485520 FJ002118 CYMT015-18</td>
</tr>
<tr>
<td>29.</td>
<td>Estoxobocellus cribiger</td>
<td>CCMF391</td>
<td>MH129019 HQ912571 MH138058 MH141549 CYMT016-18</td>
</tr>
<tr>
<td>30.</td>
<td>E. spinifer</td>
<td>CCMF393</td>
<td>MH129020 AY485504 AY485504 MH141549 CYMT015-18</td>
</tr>
<tr>
<td>31.</td>
<td>E. spinifer</td>
<td>RCC795</td>
<td>MH129021 MH138050 MH138066 MH141550 CYMT017-18</td>
</tr>
<tr>
<td>32.</td>
<td>E. spinifer</td>
<td>RCC858</td>
<td>MH129022 MH138051 MH138066 MH141551 CYMT018-18</td>
</tr>
<tr>
<td>33.</td>
<td>E. spinifer</td>
<td>RCC8675</td>
<td>MH129023 MH138052 MH138068 MH141552 CYMT019-18</td>
</tr>
<tr>
<td>34.</td>
<td>E. spinifer</td>
<td>CCMF396</td>
<td>– MH138053 MH138069 MH141553 CYMT016-18</td>
</tr>
<tr>
<td>35.</td>
<td>Pierroccomprisia catenulae</td>
<td>R2</td>
<td>– HQ1413685 HQ1413685 HQ1413685 CYMT017-18</td>
</tr>
<tr>
<td>36.</td>
<td>P. catenulae</td>
<td>R10</td>
<td>– HQ1413684 HQ1413684 HQ1413686 –</td>
</tr>
<tr>
<td>37.</td>
<td>Lithodendrum undulatum</td>
<td>CCMF1797</td>
<td>GQ330350 HQ912559 HQ912559 HQ912483 –</td>
</tr>
</tbody>
</table>

using an UltraClean® Soil DNA Isolation Kit (Qiagen Sciences, Germantown, MD, USA [formerly Mo Bio Laboratories]) as per manufacturer’s instructions. ITS2 sequences were amplified using a suitable combination of primers SR12cF, 5.8SF, 25F1R, and ITS4R (Lang and Kaczmarska, 2011). A ~321 bp fragment of the conservative region (hereafter referred to as “conservative”) of the nuclear encoded 18S rRNA marker including partial V2 and V3 variable regions was amplified using primers 18F and 18R (Iwatani et al., 2005). Furthermore, the folding pattern was well supported by the Order Cymatosoriales which were also included into the phylogenetic analysis (Table 1). The correct start and end points of the ITS2 sequences were determined using Hidden Markov Models (HMMs) of the flanking 5.8S and 28S rRNA regions (Eddy, 1998; Keller et al., 2009).

2.4. ITS2 consensus secondary structure prediction and sequence alignments

ITS2 secondary structures were predicted by comparing RNA folding patterns of complete ITS2 sequences and, if necessary, of single helices, using the Mfold web server (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form2.3; Zuker, 2003) with default parameters except for temperature, which was set according to the growth conditions for the strains. Usually, Mfold generated multiple alternative folding patterns for each ITS2 sequence with similar minimal free energy values. The correct folding pattern corresponded to the original secondary structure model of Mai and Coleman (1997) that later on was corroborated by hundreds of thousands of secondary structures (Schultz et al., 2005). Furthermore, the folding pattern was well supported by CBCs and hCBCs, which were revealed by comparisons of related strains. The ITS2 secondary structures were visually inspected with VARNA (Darty et al., 2009).
The ITS2 primary sequences and their individual secondary structures were aligned in 4SALE v1.7.1 (http://4sale.bioapps.biozentrum.uni-wuerzburg.de/; Seibel et al., 2008, 2006). To obtain a consensus secondary structure for the Cymatosirales, the required outputs were corrected in the alignment editor mode of 4SALE through careful comparative analysis of each position (nucleotide) in the sequence at the taxonomic level defined by the phylogenetic reconstruction. First, the consensus secondary structures were identified at the lowest taxonomic level (e.g. comparison of species within one genus → consensus of the genus) and subsequently at progressively higher-level taxa following the phylogenetic reconstruction. To specify the conservation level of individual ITS2 positions in the alignment, a 70% majority rule consensus secondary structure was generated following Caisová et al. (2011). Subsequently, a variability category was specified for each position in the consensus secondary structure of ITS2. The compensatory base substitutions were counted based on the sequence structure alignment using CBCAnalyzer as implemented in 4SALE (Wolf et al., 2005b). All substitutions were then mapped onto the multi-gene phylogenetic tree topology. The hCBCs were counted manually based on the sequence-structure alignment (Appendix A – Supplementary Data 1).

2.5. Phylogenetic analyses

2.5.1. Phylogeny of multi-gene dataset

Two separate multi-gene phylogenetic analyses were undertaken. The first analysis was based on the 18 strains from 13 species representing 10 genera with sequences available for all three molecular markers (conservative and V4 region of 18S rRNA + rbcL + ITS2). The multiple sequence-structure alignment of ITS2 was generated in 4SALE v1.7.1 using ClustalW (Thompson et al., 1994) and saved in one letter encoded/pseudo-proteins (12 × 12 scoring matrix) format (Wolf et al., 2014). The conservative and V4 region of 18S rRNA and rbcL sequences were aligned using MUSCLE (http://www.ebi.ac.uk/Tools/muscle/; Edgar, 2004). Multi-gene datasets were constructed by head to tail concatenation of the aligned sequences. The alignment file was manually inspected in BioEdit v7.0 before phylogenetic analyses. Bayesian Inference (BI) was performed in MrBayes v3.2 (Ronquist et al., 2012). The multi-gene dataset was partitioned by genes and the substitution model was estimated from the data. A four-chain run for 10 million generations was undertaken with trees sampled every 5000 generations. Posterior probabilities were estimated with 50% burn-in, and a majority rule consensus tree was constructed.

A second analysis used only the conservative and V4 regions of the 18S rRNA and the rbcL genes available for 36 strains (18 from the first analysis plus an additional 18 for which only those two genes sequences were available). The conservative and V4 region of 18S rRNA and rbcL sequences were aligned using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/; Edgar, 2004). Multi-gene datasets were constructed by head to tail concatenation of the aligned sequences. Phylogenetic analyses were performed using three different methods: Bayesian Inference (BI), Maximum Likelihood (ML), and Maximum Parsimony (MP). BI was performed in MrBayes v3.2 (Ronquist et al., 2012). The multi-gene dataset was partitioned by genes and the substitution model was estimated from the data. A four-chain run for 10 million generations was undertaken with trees sampled every 5000 generations. Posterior probabilities were estimated with 50% burn-in, and a majority rule consensus tree was constructed. An online version of PHYML (Phylogenetic inferences using Maximum Likelihood; www.atgc-montpellier.fr/phyml/; Guindon et al., 2010) was used to construct the ML tree by selecting AIC (Akaike Information Criterion; Posada and Buckley, 2004) for the appropriate substitution model (Lefort et al., 2017) and NNI (Nearest-Neighbor Interchanges; Felsenstein, 2004) for branch swapping. The MP analysis was performed using the Tree-Bi-section-Regrafting (TBR; Felsenstein, 2004) algorithm in MEGA v7.0.18 (Kumar et al., 2016). Bootstrap supports (Felsenstein, 1985) were obtained based on 1000 replicates for ML and MP analyses. Lithodesmium undulatum CCMP1797 was the outgroup for all the analyses.

2.5.2. Phylogeny of multi-gene dataset and morphological characters

A separate phylogenetic analysis (BI) was undertaken based on the matrix generated from the multi-gene dataset (conservative and V4 region of 18S rRNA + rbcL + ITS2 [12 × 12 scoring matrix]) and key morphological characters of 18 representative strains of 13 species belonging to 10 genera. Seven morphological characters (i.e., frustule form, process type, spine type and location, marginal ridges, specialized end valve, fascia, and pili) were used in morphological character matrix generation (Appendix A – Supplementary Data 2). The mixed dataset was partitioned into morphology and molecular subsets. The molecular data subset was further partitioned by genes. The last setting was nst = 1, rates = gamma for morphology and nst = 6, rates = invgamma (BIC criterion) for the molecular subset. A four-chain run for 10 million generations was used and trees were sampled every 5000 generations. Posterior probabilities were estimated with 50% burn-in, and a majority rule consensus tree was constructed.

3. Results

3.1. Molecular analyses

3.1.1. Consensus secondary structure model of ITS2

Eighteen strains from 13 species representing 10 extant genera belonging to Cymatosirales were used for ITS2 secondary structure analysis. Nine of the ten genera were represented by their generic types. The primary sequences of ITS2 showed a notable length variation across the members of Cymatosirales, ranging from 273 nucleotides in P. vanheurckii to 390 in E. spinifer. Helix 3 alone differed in length by 83 nucleotides in these two species. E. spinifer compensated for this extra length by incorporating insertions in the variable part of Helix 3 (i.e., between base pair 77/165 and 97/146), but conserved the first three base pairings at the basal and a dozen so at the distal part of Helix 3 (Fig. 1; Appendix A – Supplementary Data 3). It was this high degree of secondary structure conservation that allowed the alignment of the most nucleotide positions and construction of a consensus secondary structure model for this diatom order (Fig. 1). In all but one genus analyzed here, the secondary structure retained the four helices (1–4) which were interconnected by highly conserved unpaired nucleotide sequences. The exception was the genus Arcocellulus, which had an additional helix (Helix 5) following Helix 4. The hallmark signatures for Helix 1 (GG at the base), Helix 2 (U*U mismatch), and Helix 3 (UGGU) on the 5′ side near apex of the helix) were found in all strains. The helices were unbranched in all, except the strain RCC1863, which had a 32-nucleotide long branch between positions 77 and 78 on Helix 3. The side branch in Helix 3 is also known from other diatoms (Coleman, 2007; Wolf, 2004).

A total of 186 homologous nucleotide positions were recovered in the consensus ITS2 secondary structure that were present in all strains investigated. They served as the backbone of the ITS2 consensus model for this order and included the basal part of all four helices and the spacer regions in addition to the near apical part of Helix 3 (shaded green in Fig. 1). Out of these 186, 144 positions were > 70% conserved (red circled nucleotides in Fig. 1) and 42 were < 70% conserved (white circled nucleotides in Fig. 1). Specifically, (1) the first seven nucleotides at the 5′ end of the consensus structure were 100% conserved; (2) out of the first 12 base pairs of Helix 1, seven were > 70% conserved; (3) three out of four nucleotides in the spacer region between Helix 1 and 2 were 100% conserved; (4) out of 13 base pairs in the basal part of Helix 2, six were > 70% conserved; (5) seven out of nine nucleotides in the spacer region between Helix 2 and 3 were 100% conserved; (6) out of 16 base pairs in the apical part of Helix 3, 15 were > 70% conserved; (7) out of six base pairs in the basal part of Helix 4, four were > 70% conserved; and finally (8) the last three nucleotides at the 3′ end of the...
The remaining ITS2 motifs, including the apical part of the helices and the near basal part of Helix 3, were highly variable (black circles and backbone of structure not shaded green in Fig. 1). We introduced a numbering system in the consensus secondary structure for unambiguous positional descriptions of base pairs, CBCs, hCBCs, and indels (Fig. 1). The highly variable positions with specific insertions, deletions, and length variation in each taxon are listed in Appendix A (Supplementary Data 3) and indicated with subscript numbers (1, 2, 3….) combined with the 5′-preceding nucleotide numbers (Fig. 1).

3.1.2 Multi-gene phylogenies

The combined multi-gene dataset resolved all the nodes in the tree topology (Fig. 2). All 16 nodes in the tree had > 0.95 posterior probability in BI analysis. The members of the Cymatosirales grouped into two distinct clades. C. belgica, B. brockmannii, and P. vanheurckii joined in Clade 1, whereas the remaining 10 species in Clade 2 (Fig. 2). The morphologically very unusual strain RCC1863 formed a separate basal branch within Clade 2.

We also performed a phylogenetic analysis using the conservative and V4 regions of 18S rRNA + rbcL, with a total number of 36 strains from 19 species to determine the overall general phylogenetic relationship of those taxa whose ITS2 sequences were not available in GenBank (Fig. 3). In this phylogenetic tree, two major clades were also recovered, each with similar topology for all lighter silicified, smaller species (Clade 2, Fig. 2). Heavier silicified Campylosira celliformis was placed as a sister taxon to Cymatosira belgica within Clade 1. However, heavier silicified Cymatosira loeziana (ECT3874 and SZCZCh114) and Lambertocellus africana SZCZP74 were positioned outside of Clades 1 and 2 (Fig. 3).

3.1.3 Compensatory base changes (CBCs) and hemi-Compensatory base changes (hCBCs)

In the consensus ITS2 secondary structure model, 17 base pair positions were 100% conserved across all taxa. Among these 17 positions, three were found in Helix 1 (8/32: U-G, 9/31: C-G, 12/28: C-G), two in Helix 2 (38/63: G-U, 44/57: G-C), and 12 in Helix 3 (75/167: C-G, 76/166: C-G, 80/164: G-C, 81/163: G-C, 82/161: C-G, 90/153: C-G, 104/142: A-U, 105/141: C-G, 106/140: C-G, 109/136: A-U, 111/134: C-G, 118/127: G-C; Fig. 1). Five base pair positions were 100% conserved across all taxa in Clade 1. Among these five positions, three were found in Helix 1 (10/30: G-C, 11/29: G-C, 14/25: G-C), one in Helix 2 (46/55: C-G), and one in Helix 4 (173/182: G-C) (Fig. 1). Two base pair positions (10/30: A-U in Helix 1 and 77/165: U-A in Helix 3) were 100% conserved across all taxa in Clade 2 (Fig. 1).

All CBCs and hCBCs in the conserved region (green in Fig. 1) were superimposed on the BI tree inferred from the combined dataset of 18 strains from 13 species and were linked to the nodes on which they evolved (Fig. 2). The node separating Clade 1 from Clade 2 was supported by several CBCs (Table 2). Moreover, all the nodes separating one genus from another were also supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1). The node separating A. mammifer from two strains of A. cornucervis was supported by both CBCs and hCBCs, as was the case for all the species (Table 2, Appendix A – Supplementary Data 1).
3.2. Morphological analyses

3.2.1. Morphological and molecular characterization of unnamed strains and taxa identity

The valve characters of our two newly established clonal isolates BS Cymato C11 and JM Cymato C14 confirmed that they represent Plagiogrammopsis vanheurckii and Cymatosira belgica, respectively (Fig. 4a–d). Taxonomic identity of several strains obtained from culture collections also had to be established as a part of this study. Specifically, out of eight cymatosiroid strains received from the RCC collection, four were identified only to genus level while a further two only as members of Cymatosiraceae. Our morphometric and molecular examination recognized seven of the eight RCC strains as already known species. Strains RCC741, RCC795, RCC2624, and RCC3057 were morphologically and molecularly consistent with Minutocellus polymorphus, Extubocellulus spinifer, Papillocellulus simplex, and Arcocellulus cornucervis, respectively (Fig. 4g–j, n–o). Two strains designated as Cymatosiracea RCC1858 and RCC2675 were found to be E. spinifer (Fig. 4e–f). Finally, the strain RCC1863 assigned to the genus Minutocellus sp. at the time of strain establishment, was found to be very different genetically (Figs. 2 and 3) and morphologically from any known Minutocellus species (Fig. 4h–l; Hasle et al., 1983). The siliceous valves of this strain were extremely delicate, consisted of an open mesh-like valve with one tubular process on the valve face (Fig. 4k–l). We consider this strain an unranked member of the Order Cymatosirales and in need of more in depth investigation and a separate study.

3.2.2. Character-state evolution and reappraisal of the families

The Bayesian analysis of the combined dataset of three molecular markers and key morphological characters (details in Appendix A – Supplementary Data 2) recovered the same two clades seen in molecular analyses alone (Fig. 5). All the species in Clade 1 are relatively heavily silicified, have a labiate process and marginal spines, whereas the species in Clade 2 show lighter silification, possess a tubular...
process and have either marginal ridges, face spines or are devoid of any such projections. Leyanella, Pseudoleyanella, Papiliocellulus, and Arcocellulus have marginal ridges instead of marginal spines, whereas spines scattered about the face are present in Extubocellulus. Finally, Minutocellus lacks any kind of spines or ridges.

On the basis of the conservation of molecular and morphological characters presented above, and a compilation of reproductive characters outlined below, we propose the following changes to the delimitation of the Family Cymatosiraceae and erect a new Family Leyanellaceae.

3.2.2.1. Order Cymatosirales Round and Crawford, emend. Samanta & Kaczmarska. Cells are chain forming or solitary. The frustules are either heterovalvate or isovalvate. Two apical ocelluli are present on valves, albeit very rudimentary in very small specimens. Process on the valve face is either labiate, tubular, or absent. Valves may carry pilli, marginal ridges, or spines on the valve margins or throughout valve face. Auxospores are globular or sub-globular with spinescent scales and crescent-shaped initial valves at the early stage of initial frustule development. The families falling into this order are: Cymatosiraceae and Leyanellaceae.

Molecular signatures: Seventeen base pair positions are 100% conserved in the ITS2 consensus secondary structure model for this order:

- H1: 8/32 (U-G)
- H1: 9/31 (C-G)
- H1: 12/28 (C-G)
- H2: 38/63 (G-U)
- H2: 44/57 (G-C)
- H3: 75/167 (C-G)
- H3: 76/166 (C-G)
- H3: 80/164 (G-C)
- H3: 81/163 (G-C)
- H3: 82/161 (C-G)
- H3: 90/153 (C-G)
- H3: 104/142 (A-U)
- H3: 105/141 (C-G)
- H3: 106/140 (C-G)
- H3: 109/136 (A-U)
- H3: 111/134 (C-G)
- H3: 114/121 (C-G)

Fig. 3. Phylogenetic tree from Bayesian inference (BI) based on concatenated conservative + V4 region of 18S rRNA + rbcL nucleotide sequences. Support values at nodes from left to right are posterior probabilities (BI) and bootstrap percentages (ML/MP). Only posterior probabilities > 0.95 and bootstrap values > 50% are shown at nodes. Lithodesmium undulatum CCMP1797 was the outgroup. Scale bar indicates substitutions per site. Strains with hash mark (#) were included in ITS2 secondary structure analysis (Fig. 1) and structure guided multi-gene phylogenies (Figs. 2 and 5).
3.2.2.2. Family Cymatosiraceae Hasle, von Stosch & Syvertsen, emend. Samanta & Kaczmarska. Cells are chain forming or solitary. Frustules are heterovalvate, one valve with a labiate process. Ocelluli and marginal ridges, pilli, or face spines are present in some of the species. Marginal ridges, pilli, or face spines are present; in some taxa marginal spines are linked. Marginal spines are present; in some taxa marginal spines are linked. Five base pair positions are 100% conserved in the ITS2 sequences of the polar diatoms, Cymatosirales. We recovered four helices in our consensus model and found the first consensus model of ITS2 secondary structure to examine diatoms from species to order level in a relatively small group of polar diatoms, Cymatosirales.

We recovered four helices in our consensus model and found the basal pairings of Helices 1, 2, and 4 are conserved at order level. As expected, the most conserved region in our structure was the distal part of Helix 3, as was found for ITS2 secondary structures in a wide range of organisms, but other hallmarks were also found (pan-eukaryotic sensu Coleman, 2009, 2007; Schultz et al., 2005; ITS2 database I-V in Ankenbrand et al., 2015; Koetschan et al., 2012; Schultz et al., 2006; Selig et al., 2008; Wolf et al., 2005a). Furthermore, we found that the unpaired nucleotides in the spacer regions were also highly conserved within the order. In higher level taxa (family and order) ITS2 secondary structure comparisons, the primary sequences of each helix sometimes needed to fold individually. In such cases, the highly conserved spacer regions were an additional useful tool in unambiguously defining helix boundaries. Friedman et al. (2010) demonstrated the general impact of topological optimization of secondary structure prediction based on ITS2 sequences of the flowering plant family Asteraceae. Caisová et al. (2013, 2011) later showed that the conserved spacer regions can be used for reproducible prediction of ITS2 helices for green algae. Our results support this notion and suggest that this may apply to a still wider range of organisms. Similar to others (MacGillivray and Kaczmarska, 2012; Miao et al., 2008; Yao et al., 2010), we also found significant ITS2 primary sequence length variation among Cymatosirales. The strains with long ITS2 sequences nonetheless preserved the conserved regions by accommodating the extra length at the variable part of the consensus structure. This facilitates the congruent alignment of the conserved regions first, and then of all sequences for phylogenetic analyses. Markert et al. (2012) showed that phylogenetic studies based on ITS2 sequence-structure alignments seem fairly robust against different structure conformations, as long as structures consistent in one conformation are used. In this study, all the predicted ITS2 secondary structures belong to one structural conformation and all of them were obtained from direct folding using Mfold.

4. Discussion

4.1. Molecular analyses

4.1.1. ITS2 primary sequences and secondary structure model

In spite of the highly divergent primary sequences of the fast-evolving ITS2, the use of its sequence-structure alignment and resulting structure comparisons, the primary sequences of each helix sometimes needed to fold individually. In such cases, the highly conserved spacer regions were an additional useful tool in unambiguously defining helix boundaries. Friedman et al. (2010) demonstrated the general impact of topological optimization of secondary structure prediction based on ITS2 sequences of the flowering plant family Asteraceae. Caisová et al. (2013, 2011) later showed that the conserved spacer regions can be used for reproducible prediction of ITS2 helices for green algae. Our results support this notion and suggest that this may apply to a still wider range of organisms. Similar to others (MacGillivray and Kaczmarska, 2012; Miao et al., 2008; Yao et al., 2010), we also found significant ITS2 primary sequence length variation among Cymatosirales. The strains with long ITS2 sequences nonetheless preserved the conserved regions by accommodating the extra length at the variable part of the consensus structure. This facilitates the congruent alignment of the conserved regions first, and then of all sequences for phylogenetic analyses. Markert et al. (2012) showed that phylogenetic studies based on ITS2 sequence-structure alignments seem fairly robust against different structure conformations, as long as structures consistent in one conformation are used. In this study, all the predicted ITS2 secondary structures belong to one structural conformation and all of them were obtained from direct folding using Mfold.

4.1.2. Importance of CBCs and hCBCs in diatoms

Robust prediction of secondary structures in ITS2 is required for the detection of all CBCs. Coleman (2000) first proposed that even a single CBC in conserved regions of Helices 2 and 3 coincided with sexual incompatibility and could be used to recognize sexual incompatibility and thus biological species, initially in the green algae Volvocales. Unfortunately, the absence of a CBC did not always coincide with sexual compatibility. Later, Müller et al. (2007) also showed that the presence of a CBC between related species may indeed be sufficient to distinguish
them, and that they were equally distributed over all four helices. More recently Caisová et al. (2013, 2011) found that in some Chlorophytes, CBCs most often coincide with supraspecific divergence, e.g., species complexes, genera or even higher, and called them CBC grades. These authors concluded that genes controlling sexual compatibility and the genes involved in morpho-species separation apparently evolved faster than the CBCs in the ITS2 of the Ulvales. In the Cymatosirales ITS2 consensus model, CBCs that correspond to species separation phylogenetically as well as morphologically are not restricted to conserved regions of Helices 2 and 3 but extend to conserved regions of Helices 1 and 4 and support taxa at all levels, including the majority of species. Previous studies with *Pseudo-nitzschia* have shown that even CBCs in Helices 1 and 4 coincide with species divergence and reproductive isolation (Amato et al., 2007; Lim et al., 2013). The present study suggests that detection of CBCs over the conserved region of all four helices offers valuable insights into higher taxonomic relationships in cymatosiorids, and possibly in other diatoms.

We found that the nodes separating all but two morpho-species within a genus were also supported by CBCs in the conserved regions of ITS2 helices (green in Fig. 1). The two exceptions currently make up the genus *Papiocellulus* (*P. elegans* and *P. simplex*, Gardner and Crawford, 1992; Hasle et al., 1983). These two species are quite distinct morphologically (Fig. 4m vs. 4n-o, respectively), but the only CBC was found in the non-conserved region of Helix 3 and three hCBCs were found in the conserved region of the ITS2 model. This suggests that CBCs in non-conserved regions might be useful in distinguishing morpho-species in diatoms. Müller et al. (2007) showed that the probability of two ITS2 sequences belonging to different species, given at least one CBC, is 0.93 and the probability of two ITS2 sequences belonging to different species, given no CBC, is 0.76; albeit diatoms were not included in that study. In case of group specific CBC probabilities, Heeg and Wolf (2015, for Chlorellaceae) showed these probabilities are 0.74 and 0.93 respectively. In case of Cymatosirales, both the probabilities are 1.00. However, for Cymatosirales, using only the conserved region of consensus ITS2 secondary structure, the probabilities are 0.92 and 1.00 respectively.

Previous studies have also shown that the presence of hCBCs, even in non-conserved regions, separated morpho-species in some diatoms.
(Leptocylindrus, Nanjappa et al., 2013; Pseudo-nitzschia, Amato et al., 2007; Amato and Montresor, 2008; Percopo et al., 2016). Post-zygotic incompatibility has been observed within Pseudo-nitzschia between two closely related morpho-species (P. calliantha and P. mannii) differing only by hCBCs (two in Helix 1 and two in Helix 3) which suggests that hCBCs may also be sufficient to recognize them as separate biological species (Amato et al., 2007; Amato and Montresor, 2008), at least in this genus. Therefore, the absence of CBCs but presence of hCBCs in the ITS2 secondary structure could also be informative in recognition of some diatom morpho-species. In contrast, among members of the genus Extubocellulus, strain RCC2675 is morphologically similar to E. spinifer, but the node separating this strain from the other three E. spinifer strains was supported by two hCBCs. All the cases discussed above together suggest that hCBCs alone may also coincide with the emergence of morpho-species (Papiliocellulus, Leptocylindrus, Pseudo-nitzschia), and possibly the existence of morphological cryptic diversity (Extubocellulus). Nonetheless, hCBCs should be considered with additional evidences (e.g. mating experiments) until more data is available.

4.1.3. Phylogenies

In the absence of ITS2 sequences available for some species, we performed a more “general” phylogenetic analysis (combining 18S rRNA and rbcL sequences). This permitted gauging phylogenetic relationships among a greater number of species. Among these, the position of Cymatosira lorenziana and the recently erected Lambertocellus africana remains unresolved in our combined tree. Despite possessing all the key morphological characteristics seen in other members of the family Cymatosiraceae (Hasle et al., 1983), C. lorenziana did not join Cymatosiraceae. Similarly, even though Lambertocellus africana possesses the familial valve characters of Leyanellaceae (pilli and a tubular process, see Dąbek et al., 2017) it also did not join the clade with members to which it is most similar morphologically. In contrast,
Camyplosira cymbelliformis, possessing all the key morphological characters of the family Cymatosiraceae, did join the appropriate clade as a sister taxon to C. helgica. We observe nonetheless that neither of the unresolved species joined an “inappropriate” clade.

4.2. Morphological analyses and character evolution

Our results support the ordinal rank for Cymatosirales proposed by Round and Crawford (Round et al., 1990), and its monophyly. We identified morphological and molecular synapomorphies for this order and for its two families. The morphological synapomorphies for this order are in their valve morphology and the type of auxospores. First, cymatosiroid species have ocelluli at both apices of valves. Second, during auxospore development detailed in B. brockmannii (which belongs to Cymatosiraceae), the initial epivalves curl inside the globular auxospore walls and thus are crescent-shaped during early stages of valve genesis (Samanta et al., 2017). This was also observed in P. vanheurckii (Hoppenrath et al., 2009), and possibly in some members of Leyanellaceae (M. polymorphus and P. elegans; Samanta and Kaczmarska, unpublished). Our observations and those published by von Stosch (1982), Hasle et al. (1983) for three additional genera indicate that this unique mode of auxospore development evolved early in the cymatosiracean lineage and was likely present in the last common ancestor of both families.

The members of the mediophycean family Eupodiscaceae also have ocelluli, but their structure is different from Cymatosirales (Hasle et al., 1983; Simonsen, 1979). Ashworth et al. (2013) showed that all ocelluli bearing Eupodiscaceae clustered into a monophyletic clade apart from the other morphological taxa with biddulphioid frustules. More recent phylogenetic studies recovered Eupodiscaceae as sister to Cymatosirales (Djabek et al., 2017; Nakov et al., 2018). The structure and development of the cymatosiracean auxospore wall supports this notion. Although spinescent scales in the auxospore walls are only known in these two groups, the structure of the spines is different in each (Hasle et al., 1983; Samanta et al., 2017; von Stosch, 1982). Phylogenetic and unusual shared morphological characters suggest that a new diatom Order Eupodiscidales might have to be considered as a sister to Cymatosirales by reappraising the Family Eupodiscaceae.

Hasle et al. (1983) placed all the cymatosiroids in the single Family Cymatosiraceae, subdivided into two subfamilies based on valve morphology: Cymatosiroidae included heterovalvate species while Exubocelluloideae were isovalvate. However, our combined molecular morphology; Cymatosiroideae included heterovalvate species while Exubocelluloideae were isovalvate. However, our combined molecular phylogenetic studies recovered Eupodiscaceae as sister to Cymatosirales (Hasle et al., 1983; Samanta et al., 2017; von Stosch, 1982). Phylogenetic and univalve genesis (Samanta et al., 2017). This was also observed in P. vanheurckii (Hoppenrath et al., 2009), and possibly in some members of Leyanellaceae (M. polymorphus and P. elegans; Samanta and Kaczmarska, unpublished). Our observations and those published by von Stosch (1982), Hasle et al. (1983) for three additional genera indicate that this unique mode of auxospore development evolved early in the cymatosiracean lineage and was likely present in the last common ancestor of both families.

The members of the mediophycean family Eupodiscaceae also have ocelluli, but their structure is different from Cymatosirales (Hasle et al., 1983; Simonsen, 1979). Ashworth et al. (2013) showed that all ocelluli bearing Eupodiscaceae clustered into a monophyletic clade apart from the other morphological taxa with biddulphioid frustules. More recent phylogenetic studies recovered Eupodiscaceae as sister to Cymatosirales (Hasle et al., 1983; Samanta et al., 2017; von Stosch, 1982). Phylogenetic and unusual shared morphological characters suggest that a new diatom Order Eupodiscidales might have to be considered as a sister to Cymatosirales by reappraising the Family Eupodiscaceae.

Hasle et al. (1983) placed all the cymatosiroids in the single Family Cymatosiraceae, subdivided into two subfamilies based on valve morphology: Cymatosiroidae included heterovalvate species while Exubocelluloideae were isovalvate. However, our combined molecular and morphological cladistics do not support this classification. Based on molecular and morphological synapomorphies, we amend the Order Cymatosirales and propose it to be subdivided into two families, Cymatosiraceae and Leyanellaceae, the first being amended while the second erected. The common ancestor shared with eupodiscoid taxa likely had spinescent scales in their young auxospore walls because spinescent auxospore walls are present in all the members of these groups thus far examined. While multiple labiate processes may be present in Eupodiscaceae, one labiate process may have become the stable synapomorphic character in Cymatosiraceae in the course of earlier evolution of the lineage, but eventually even this one was lost or modified into a tubular process in some Leyanellaceae. The tubular process is not a stable synapomorphic character among Leyanellaceae either. It is absent in Pseudoleyanella but its presence in the sister genus Leyanella suggests the loss of this character during the evolution of the former. It is important to note that a tubular process is also absent in Pierrecomperia (Sabbe et al., 2010) but it is at least occasionally present in Extubocellulus, and these two genera are closely related genetically within the Leyanellaceae (Fig. 3). The evolutionary pattern suggests that some species in the genus Extubocellulus may be in a state of transition toward the loss of the process, having been already lost in Pierrecomperia.

In addition to the labiate process, the presence of marginal spines is also a synapomorphic character for Cymatosiraceae. Leyanellaceae on the other hand possess either marginal ridges or face spines. It remains uncertain how the marginal spines of Cymatosiraceae evolved, and/or whether marginal ridges in Leyanellaceae represent a transformation of the marginal spines of the former family because valve morphology of the basal taxon of Leyanellaceae (Leyanellaceae X sp. RCC1863) is strongly dissimilar to any of the Cymatosiraceae or Leyanellaceae species examined. Our phylogenies suggest that the marginal ridges were lost twice, first in Extubocellulus, then again in Minutocellulas, when Leyanellaceae diversified. It is interesting to note that face spines evolved only in Extubocellulus and Cymatosirella. Furthermore, isovalvate in Extubocellulus and Cymatosirella is shared with Pierrecomperia, but the latter genus does not have face spines (Djabek et al., 2013; Sabbe et al., 2010). Morphologically, these three genera are closely related to each other. Unfortunately, the phylogenetic relationship of Cymatosirella remains unknown because there are no sequence data of any kind available for this genus. In our combined ITS2 rRNA and rcl phylogenetic tree, Pierrecomperia was recovered as a sister taxon to E. spinifer (Fig. 3), but we have no ITS2 sequences of Pierrecomperia and so could not include this genus in our secondary structure analyses.

In summary, using extant polar centric cymatosiroid diatoms as a case study, we demonstrate for the first time the application of ITS2 secondary structures and its sequence-structure alignment for the recovery of diatom phylogenetic relationships not only at the species but also at the genus, family and order level. In our ITS2 based phylogenetic trees, all branches are supported by a number of CBCs, correspond to new morphological synapomorphies and are consistent with multigene phylogenies. In all these trees, all our species group into two genetically and morphologically distinct clades. These clades do not correspond to the current classification based on different morphological synapomorphies, however. Consequently, we amend the Order Cymatosirales and family Cymatosiraceae, and propose a new family, the Leyanellaceae. We also showed the importance of CBCs and hCBCs in diatom morpho-species recognition and their possible implications in recognizing biological species separation. In Cymatosirales, both the probabilities of two ITS2 sequences belonging to different species, given at least one CBC, and the two ITS2 sequences belonging to different species, given no CBC are 1.00. Our results suggest that the approach used in our study might be useful in recovering both terminal and deep divergences in diatoms, at least up to the ordinal level.

Competing interests

Authors have no competing interests to declare.

Acknowledgements

We thank to L. Caisová for valuable suggestions and comments on ITS2 secondary structure prediction, and N. Nakamura and S. Mayama for sharing the cell pellet and DNA of P. lunata. We are also thankful to two anonymous reviewers for their helpful comments. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant to IK.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.08.014.

References


D42.


Wolf, M., Achtziger, M., Schultz, J., Dandekar, T., Muller, T., 2005a. Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. RNA 11, 1616–1623.


