Phylogeny, ecology, morphological evolution, and reclassification of the diatom orders Surirellales and Rhopalodiales

Elizabeth C. Ruck^{a,*1}, Teofil Nakov^{a1}, Andrew J. Alverson^a and Edward C. Theriot^b

* Corresponding author:

Elizabeth C. Ruck

^aDepartment of Biological Sciences

University of Arkansas

1 University of Arkansas, SCEN 601

Fayetteville, AR 72701

e-mail: ruck@uark.edu

Teofil Nakov

^aDepartment of Biological Sciences

University of Arkansas

1 University of Arkansas, SCEN 601

Fayetteville, AR 72701

Andrew J. Alverson

^aDepartment of Biological Sciences

University of Arkansas

1 University of Arkansas, SCEN 601

Fayetteville, AR 72701

Edward C. Theriot

^bDepartment of Integrative Biology

The University of Texas at Austin

1 University Station (A6700)

Austin, Texas 78712 USA

Keywords: canal raphe, diatom, reclassification, Rhopalodiales, salinity, Surirellales, trait evolution

¹ Co-first authors

Abstract

The Surirellales and Rhopalodiales are large, widespread, and morphologically diverse groups of raphid pennate diatoms (Bacillariophyta) whose raphe, a structure that facilitates active motility, opens internally into a siliceous canal. We collected 202 representatives of the lineage and sequenced genes from the nuclear, plastid, and mitochondrial genomes to infer phylogenetic relationships as a basis for comparative study of ecology and morphological evolution as well as reclassification. The lineage was ancestrally marine, and we report the first evidence for a 'stepping stone' model of marine–freshwater transitions in which freshwater invasions were preceded by adaptation to intermediate brackish habitats. Phylogenetic comparative analyses also showed that the shift from an apical (e.g., *Entomoneis*) to transapical major axis of development (e.g., *Surirella*) did not have to proceed through subcircular intermediate forms (i.e., *Campylodiscus*). Rather, subcircular forms evolved both within lineages with longer apical axis or longer transapical axis. We also used the inferred phylogeny as a basis for genus-level reclassification of the lineage. *Campylodiscus* now includes the fastuosoid members of *Surirella* and *Campylodiscus*, but excludes other marine *Campylodiscus* which are now classified as *Coronia*. *Surirella* includes the *Surirella striatula* clade, *Surirella* Pinnatae group, and species formerly classified as *Cymatopleura*. We resurrected the genus *Iconella* Jurilj to accommodate *Stenopterobia* and the robustoid members of *Surirella* and *Campylodiscus*. We broadened *Epithemia* to include members of the paraphyletic genus *Rhopalodia*. Finally, we discuss the challenges of constructing a classification that best leverages available phylogenetic data, while minimizing disruption to the research community and recognizing practical considerations stemming from the slow rate of progress on systematic studies of understudied organisms.

1. Introduction

Raphid pennate diatoms (Bacillariophyta: Bacillariophyceae) are characterized by the ability to actively move along a substrate using a structure called a raphe (Round et al., 1990). Raphe-enabled motility occurs via an actin–myosin motility system working with mucilagenous secretions extruded through slits in the silica cell wall (Poulsen et al. 1999; Harper 1977). Raphid pennates exhibit extraordinary morphological diversity, including a broad range of complexity in their raphe systems (Round et al., 1990). In some species, for example, the raphe slit opens internally into a siliceous canal. These canal raphe systems are often, but not always, raised onto keels and are most commonly positioned along the valve margins (Ruck and Kociolek, 2004; Ruck and Theriot, 2011). Communication with the cell interior occurs via tubes or openings of the canal, and the latter is structurally supported by siliceous braces (Krammer, 1989a; Round et al., 1990; Sims and Paddock, 1982). These kinds of raphe modifications are thought to improve structural stability of canal raphes and enhance cell mobility (Krammer, 1989b). Although the exact roles of these structures have not been established, the canal raphe likely confers some advantage to the cell, as it has evolved independently in two distantly related and highly diverse lineages, the Bacillariales (*Nitzschia* Hassall and allies) and a clade consisting of the orders Surirellales and Rhopalodiales (Ruck and Theriot, 2011).

The Surirellales and Rhopalodiales together comprise >2,000 nominal species (Fourtanier and Kociolek, 2011). The Surirellales includes ten genera arranged into three families (Guiry and Guiry, 2016; Round et al., 1990) and is noteworthy for the broad diversity of cell shapes, which range from circular (*Campylodiscus* Ehrenberg *ex* Kützing) to iso- or heteropolar ellipsoids (*Auricula* Castracane, *Cymatopleura* W. Smith, *Entomoneis* Ehrenberg, *Surirella* Turpin). Their raphe systems can be positioned along the apical axis of the cell, following a sigmoid or straight course (Entomoneidaceae; Fig. 1a), or along the valve margin in a semi- to fully circumferential arrangement (Auriculaceae and Surirellaceae; Fig. 1a; Round et al., 1990). In Rhopalodiales, cells are generally strongly dorsiventral along the apical plane and wedge-shaped in the transapical plane (Round et al., 1990). Their canal raphe systems can be positioned along the dorsal (e.g., *Rhopalodia* Müller) or ventral cell margin (e.g., *Epithemia* Kützing; Fig 1a) or medially (e.g., *Tetralunata* Hamsher, Graeff, Stepanek & Kociolek; Hamsher et al., 2014). Importantly, phylogenetic analyses based on both morphological and molecular data have found paraphyly of many named groups, including the two orders relative to each other (Ruck and Kociolek, 2004; Ruck and Theriot, 2011).

An outstanding question regarding the evolutionary history of the Surirellales and Rhopalodiales pertains to the evolution of cell wall architecture, specifically the evolutionary shift in major axis of development, from elongation about the apical axis to elongation along the transapical axis. Most pennate diatoms have an elongate shape. The long axis of the cells is referred to as the "apical" axis (AA), whereas the short or "transapical" axis (TA) lies perpendicular to the long axis. Cell development in bilaterally symmetrical diatoms, including the raphid pennate clade (Theriot et al., 2015), commonly includes expansion along the axis parallel to the raphe (the apical axis; e.g., Idei et al., 2013). This type of morphogenesis is an ancestral feature of the Surirellales + Rhopalodiales clade (SR-clade), present in all Rhopalodiales, some early diverging clades of Surirellales (Ruck and Theriot, 2011), and the sister lineage to SR-clade (e.g. Sabbe et al., 2004). In species of Surirellales with a circumferential raphe, the cell's apical and transapical axes can be nearly equally long, resulting in (sub)circular cells (*Campylodiscus*), or the transapical axis can be longer than the apical axis, forming ellipsoids (*Surirella*) or spindle-shaped cells (*Stenopterobia* Brébisson *ex* Van Heurck) whose long axis is perpendicular to the ancestrally major

transapical axis. Although it is tempting to suggest that elongation of the transapical axis (i.e., becoming wider) inevitably leads to more-or-less circular shapes before arriving at a state where the cell's transapical axis is longer, alternative scenarios, in which the *Campylodiscus*-type circular cell architecture evolved within lineages with an ancestrally transapical axis, have also been proposed (Jurilj, 1949). The relative support for these competing evolutionary hypotheses can be formally tested by comparing ordered models of character evolution, but these tests require phylogenetic hypotheses with dense taxonomic sampling.

Members of the Surirellales and Rhopalodiales generally live as solitary cells and are associated with fine sediments in the benthos of marine, brackish, and freshwater habitats (Lowe, 2003). Although strictly marine and strictly freshwater genera exist [e.g., *Epithemia* (Ehrenberg) Kützinger, *Stenopterobia*, *Petrodictyon* (Ehrenberg) D.G. Mann], the distributions of the largest genera (*Entomoneis*, *Rhopalodia*, *Surirella*, *Campylodiscus*) span the entire salinity gradient. Species distributed in brackish waters are common and include the type species of the two most species-rich genera, *Campylodiscus* and *Surirella*. However, the history of traversals across the salinity gradient—including ancestral states and the number and directions of transitions across environments—is unknown. It is unclear, for example, if the history of this group is consistent with a model of infrequent and irreversible transitions from marine to freshwaters (Mann, 1999) or with the two-way transitions common in the "centric" diatom order Thalassiosirales (Alverson et al., 2007). Moreover, the potential role of brackish habitats as a transitory stage toward adaptation to freshwater environments has not been investigated. Testing these hypotheses in a phylogenetically younger and almost exclusively benthic lineage, compared to the much older and planktonic Thalassiosirales, should provide valuable insights into the nature of the marine–freshwater barrier to diatom distribution (Alverson et al., 2007; Mann, 1999).

To address the phylogeny, classification, and evolutionary history of the Surirellales and Rhopalodiales, we produced a large phylogenetic dataset with genes from the nuclear, plastid, and mitochondrial genomes. The phylogenetic and ecological breadth of sampled taxa allowed us to formally test hypotheses regarding the evolution of cell architecture and the transitions among marine, brackish and freshwater habitats. We also used the phylogeny to propose a partial reclassification of taxa with both morphological and molecular synapomorphies.

2. Material and Methods

2.1 Taxon sampling and cell cultures

Previous phylogenetic studies recovered *Halamphora* (Cleve) Levkov as sister to the SR-clade (Ruck and Theriot, 2011; Stepanek and Kociolek, 2014), so we sampled three genera within the order Thalassiophysales as outgroups. Of the 243 culture strains included in this study, 181 represent Surirellales, 21 represent Rhopalodiales, and 41 were outgroup taxa from the Thalassiophysales (Table S1). Sequence data for 34 taxa, mostly outgroup Thalassiophysales, were obtained from Genbank, and we generated all of the remaining sequence data. We obtained 29 diatom strains from public culture collections, and the remaining strains were isolated from field collections. These either were isolated into monoclonal cultures or were used for single-cell PCR amplifications to generate the sequence data (Table S1). Cells were grown in batch culture at varying temperatures from 14–27°C on a 12:12 light-dark cycle. Marine species were grown in L1 medium (Guillard and Hargraves, 1993), and freshwater species were grown in WC, COMBO, or biphasic soil medium (Czarnecki, 1987; Guillard and Lorenzen, 1972; Kilham et al., 1998). Strains isolated from estuarine environments were grown in

diluted seawater with L1 or WC nutrients added. Images of voucher specimens are available in the online data package on Zenodo (<https://zenodo.org/record/57998>).

2.2 DNA extraction, PCR and sequencing

DNA was extracted using either a DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) or, if grown in biphasic media, a PowerSoil® kit (MO BIO Laboratories, Carlsbad, CA, USA). When culture establishment was unsuccessful, total genomic DNA was extracted by heating single cell isolations in a 10% Chelex 100® solution to 95 degrees for 20 minutes (Richlen and Barber, 2005). Nuclear *SSU* and partial *LSU* rDNA and plastid *rbcL* and *psbC* were PCR-amplified and sequenced using published PCR programs and primer sequences (Ruck and Theriot, 2011; Scholin et al., 1994). PCR and sequencing primers for mitochondrial cytochrome b (*cob*) are listed in Table S2. PCR reactions consisted of 1.0–5.0 µL of DNA extract; 6.5 µL of Failsafe Buffer E (Epicentre Technologies); 0.5 µL of each primer (20 µM stocks); 0.5 units Taq polymerase; and ddH₂O to a final volume of 25 µL. When necessary, a nested PCR strategy was used. The PCR cycles for amplification of *cob* (if only one reaction was required) were as follows: 95°C for 5 min, 36 cycles of (95°C for 60 s, 53°C for 60 s, 72°C for 60 s), and final extension of 72°C for 15 m. If nested PCR was required for amplification of *cob*, then the second PCR program was similar to the first but with the annealing temperature decreased to 45°C.

PCR products were purified with Exonuclease I (Exo) and Shrimp Alkaline Phosphatase (SAP) following Ruck and Theriot (2011). Forward and reverse strands were cycle sequenced with BigDye (Applied Biosystems, Foster City, CA, USA) and resolved with an ABI 3130 or 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Sequence chromatograms were edited and assembled into contigs with Sequencher ver. 4.5 (Gene Codes Corporation, Ann Arbor, MI, USA).

2.3 Sequence Alignments

The dataset included five genes: plastid *rbcL* and *psbC*, mitochondrial *cob*, and nuclear *SSU* rDNA and partial *LSU* rDNA. Protein-coding genes were unambiguously aligned with Mesquite ver. 3.04 (Maddison and Maddison, 2015) based on their conceptual protein translations. Ribosomal RNA genes were aligned using covariance models of secondary structure as implemented in the program SSU-Align (Cannone et al., 2002; Nawrocki, 2009). For the *SSU* gene, we used the covariance model of secondary structure for eukaryotes available in SSU-Align and derived from a manually curated full-length alignment available from the Comparative RNA Website (Cannone et al., 2002). For alignment of the partial *LSU* region, we used a custom-built covariance model of *LSU* secondary structure for heterokonts (Nakov et al., 2014), based again on a manually curated alignment of full-length *LSU* sequences available from the Comparative RNA website (Cannone et al., 2002). Both alignments were built in the context of broader sampling of diatoms and subsequently subsampled to include only those taxa relevant for aligning sequences from Surirellales and Rhopalodiales. Both alignments were also masked to remove any positions with low posterior probability of positional homology as calculated by SSU-mask (Nawrocki 2009). All newly generated DNA sequences were deposited in GenBank (Table S1), and alignments were deposited in Zenodo (<http://dx.doi.org/10.5281/zenodo.51332>).

2.4 Phylogenetic analyses

We partitioned the alignment based on codon position within protein-coding genes and secondary structure information (paired vs. unpaired sites) for the rRNA genes. This yielded a total of 13 partitions, three each for the *rbcL*, *psbC* and *cob* genes and two each for the *SSU* and *LSU* genes. We used the model selection routine implemented in the program IQ-TREE v.1.4.1 (option -m TEST) to identify the best substitution model for each partition (Nguyen et al., 2015). Among-site rate heterogeneity was accommodated through a discretized Gamma distribution (Γ), proportion of invariable sites (I), their combination ($\Gamma + I$), or by the FreeRates model (R; Yang, 1995; option -m TESTNEW). Thus, for each partition, this routine selected the substitution model from among the Generalized Time-Reversible (GTR) family and rate heterogeneity component that provided the best fit to the data based on the Bayesian Information Criterion (BIC).

After the model selection process, we performed a partition-merging procedure to identify partitions that could be combined without loss of model fit. The procedure implemented in IQ-TREE is similar to the one used in PartitionFinder (Lanfear et al., 2012). BIC was employed again to take into account model complexity (i.e., number of estimated parameters) when comparing partition models.

We inferred phylogenies using maximum likelihood (ML) and Bayesian inference (BI) methods. For ML inference in IQ-TREE v.1.4.1, we used the best partition scheme and substitution models identified as described above. We maintained the same set of branch lengths across partitions while allowing each partition to evolve under its own rate (IQ-TREE edge-proportional model, option -spp). IQ-TREE combines a hill climbing algorithm with a stochastic component to more efficiently search tree space. The stochastic component is controlled by a parameter for the strength of perturbation of the randomized nearest neighbor interchange (IQ-TREE option -pers; default=0.5), and the tree search is stopped after a certain number of unsuccessful iterations (IQ-TREE option -numstop, default=100). To ensure our ML optimizations were not getting stuck in local optima, we searched for the best phylogeny 31 times while varying the perturbation strength between 0.3 and 0.6 and requiring at least 250 unsuccessful iterations before the optimizations were stopped. Clade support was assessed with IQ-TREE's ultra fast bootstrap option (-bb; Minh et al., 2013) with 1000 pseudoreplicates.

We used MrBayes 3.2.3 for Bayesian tree inference (Ronquist et al., 2012), partitioning the data as for ML analyses, but instead of fixing the substitution model, we allowed the samplers to sample models from the GTR family in proportion to their posterior probability (MrBayes option lset nst=mix). Rate heterogeneity within each partition was accommodated with a combination of the discretized gamma distribution and proportion of invariable sites ($\Gamma + I$), and each partition was allowed to evolve under its own rate by setting a variable rate prior (MrBayes option prset ratepr=variable). All parameters except the rate multipliers, topologies, and branch lengths were unlinked across partitions. These settings allowed us to assess the topology and branch lengths while accommodating uncertainties in model selection; that is, parameters were averaged over the credible set of substitution models for each partition. We performed four MrBayes runs—each with one cold and three heated chains—for 30 million generations, thinning the posterior by sampling every 1000th iteration. We assessed convergence using the potential scale reduction factor statistic as output from MrBayes along with a suite of routines available in Are We There Yet? (AWTY; Nylander et al., 2008) as implemented in the R Package RWTY (Warren et al., 2016). We summarized the posterior distributions after discarding 2/3 of the samples as burn-in.

2.5 Evolution of cell architecture and habitat

Due to its importance in the classification of the group, the evolution of cell wall architecture in Surirellales and Rhopalodiales has been the topic of much debate but little formal analysis (Hustedt, 1929; Jurilj, 1949; Mereschkowsky,

1904). A major component of variation in Surirellales lies in the relationship between the cell's apical and transapical axes. Similar to most other pennate diatoms, cells of outgroup taxa (*Amphora* Ehrenberg ex Kützinger, *Halamphora*, *Thalassiophysa* Conger) and the families Entomoneidaceae and Rhopalodiaceae are elongated along their apical axis (Fig. 1a; [AA > TA]). *Campylodiscus* cells are, in contrast, generally circular in shape, with apical and transapical axes roughly equal in length (Fig. 1a; [AA = TA]). Most *Surirella* species are longer along their transapical axis (Fig. 1a; [AA < TA]). To model the evolution of cell wall architecture, we coded the ingroup + *Thalassiophysa* according to the three arrangements described above as a single three-state character ([AA > TA]: 1; [AA = TA]: 2; [AA < TA]: 3). All redundant accessions (i.e., conspecific strains) were removed from the dataset for these analyses.

To assess the evolution of this character we used phylogenetic Markov models as implemented in the R package corHMM (Beaulieu et al., 2013). We considered four pathway models: i) ordered-reversible, in which evolution proceeded directionally from [AA > TA] to [AA = TA] to [AA < TA], allowing for reversals at each step ($1 \leftrightarrow 2 \leftrightarrow 3$; $k=4$; Fig. 1b); ii) an ordered-reversible-2 model in which the first step was irreversible ($1 \rightarrow 2 \leftrightarrow 3$; $k=3$; Fig. 1c); and iii) an ordered-irreversible model, in which both the gain of circular cells and the gain of cells longer along the transapical axis were irreversible ($1 \rightarrow 2 \rightarrow 3$; $k=2$; Fig. 1d). In addition to these ordered models, we considered an unordered, all-rates-different model that allowed bidirectional transitions among all states ($k=6$; Fig. 1e). We fit these models to 10^3 phylogenies sampled at random from the post-burnin posterior distributions of the MrBayes runs, summarizing the results while taking phylogenetic uncertainty into account. Trees were scaled to relative time chronograms using the program TreePL with the best value for the branch-smoothing parameter estimated via cross-validation (Smith and O'Meara, 2012). We compared the four models using AIC scores corrected for sample size (AICc). To obtain a one-tree summary of these results, we used TreeAnnotator to find the maximum credibility tree from the combined post-burnin distributions of the four MrBayes runs. We then reconstructed ancestral states on the maximum credibility tree using the cross-tree average transition rate parameters as estimated by each of the credible character evolution models (AICc weight > 0.05). The topology of the best ML tree and Bayesian consensus and maximum credibility trees differed, so we also fit these models to the best ML topology.

For ancestral reconstructions of habitat, we coded terminal taxa based on the salinity of the habitat where the corresponding culture strain was isolated from. If two strains of a single species were collected from habitats with different salinities, we included both strains in the analysis and coded them according to their origin. We coded habitat as a three-state character with marine (32–12‰), brackish (12–1‰) and freshwater (1–0‰) states. As for cell architecture, we used a similar framework with ordered and unordered pathway models to test if invasions of freshwater were stepwise and irreversible (i.e., marine $\rightarrow$ brackish $\rightarrow$ freshwater; ordered-irreversible; $k=2$), stepwise and reversible (i.e., marine $\leftrightarrow$ brackish $\leftrightarrow$ freshwater; ordered-reversible; $k=4$) or all transitions possible (i.e., marine $\leftrightarrow$ freshwater, marine $\leftrightarrow$ brackish, brackish $\leftrightarrow$ freshwater; all-rates-different; $k=6$). As before, we fit these models across 10^3 phylogenies and constructed a one-tree summary as described above.

3. Results

3.1 Taxon sampling

Our multigene matrix included 243 OTUs. Taxon sampling at the generic level was nearly complete, with only three genera from the Surirellales—*Hydrosilicon* Brun, *Plagiodiscus* Grunow & Eulenstein and the newly described

Platichthys Lange-Bertalot & Kulikovskiy —not represented. Each of these have only two described species. For Rhopalodiales, only the newly described genus, *Tetralunata* Hamsher et al. (2014), was not represented in the analysis.

3.2 Alignment

The per-gene final taxon count and alignment lengths were as follows: *rbcL*: sequences= 243, final alignment length (FAL)= 1434 nucleotides (nt); *psbC*: sequences= 154, FAL= 1212 nt; *cob*: sequences= 140, FAL= 714 nt; *SSU* rDNA: sequences= 173, FAL= 1591 nt; *LSU* rDNA: sequences= 105, FAL= 380 nt. The five genes were concatenated into a single alignment of 5,331 nt in length, with 10% missing data.

3.3 Substitution models and partition schemes

The starting partitioning scheme treated each codon position of each protein-coding gene ($k=9$ subsets) as well as the paired vs. unpaired sites in each rDNA gene as separate partitions ($k=4$ subsets), for a total of 13 character partitions. The model selection routine in IQ-TREE identified a diverse set of substitution rate matrices (Table S3). The most complex model in the reversible family, GTR, was selected for third codon positions of *rbcL* and *psbC* and the first codon position of *cob*. The FreeRates model (Yang, 1995) accounted for among-site rate heterogeneity better than the discretized Gamma distribution (Γ), the proportion of invariable sites (I), and their combination ($\Gamma + I$); the latter was selected only twice (Table S3). We found that this partitioning scheme could be greatly simplified. IQ-TREE's partition-merging procedure combined these 13 partitions into three: one for the four rDNA partitions, another for all codon positions of protein-coding genes except the third codon positions in *cob*, which was assigned its own partition (Table S4). The rDNA partition was best accommodated with a GTR substitution matrix with five FreeRates categories, the majority of protein-coding data was best modeled by a Tanimura-Nei (TN) model with six FreeRates categories, and the third codons of *cob* by a TIM + Γ + I (Table S4). All IQ-TREE analyses used this partitioning scheme and these models.

Bayesian analyses used the same partitioning scheme and, in addition, sampled substitution rate matrices in proportion to their posterior probability (PP). For the rDNA partition, GTR and three other models with five or six parameters were in the credible set ($PP > 0.05$; Table S5). For the majority of protein-coding DNA, the TN model, plus another three models each with 3–4 parameters, were credible (Table S5). For the third-codon positions of *cob*, there were seven credible models with 4–6 rate parameters, including the TIM2 model identified before (Table S5). Rate heterogeneity was accounted for with $\Gamma + I$.

3.4 Phylogeny of Surirellales and Rhopalodiales

The “best” tree from the IQ-TREE (IQ) ML analyses and the MrBayes (MB) majority rule consensus topology reconstructed the SR-clade as monophyletic [IQ-TREE UFboot=100%; MrBayes PP=1.00 (heretofore written as IQ/MB=xxx/xxx); Figs 2 and 3; Figs. S1 and S2]. Previous studies with broader taxon sampling among raphid pennate diatoms found the genus *Halamphora* as sister to the SR clade (Ruck and Theriot, 2011; Stepanek and Kociolek, 2014). Our analyses, however, strongly support the monotypic genus *Thalassiosiphysa* as sister to the SR-clade, with *Halamphora* as sister to *Thalassiosiphysa* + SR-clade (Figs 2-3). Relationships among the outgroup amphoroid taxa are consistent with the most recent phylogenetic treatment of this group (Stepanek and Kociolek, 2014).

Although IQ-TREE and MrBayes used different tree searching strategies and employed different substitution, rate heterogeneity, and branch length models, results of the two analyses generally agreed with one another. *Entomoneis* was sister to all other members of the SR clade, which themselves were a strongly supported monophyletic group (IQ/MB = 100/1.00; Figs 2 and 3). Relationships among Surirellaceae, Auriculaceae and Rhopalodiales were uncertain, however, and varied between analyses. These discrepancies were due to uncertainty about the placements of *Auricula* and a clade of marine *Campylodiscus* (Fig. 2).

In the Bayesian majority-rule consensus tree, *Auricula* was monophyletic but with low support (IQ/MB = NA/0.63; Fig 2a and Fig 3, clade 2; Fig. S1), and Rhopalodiales was monophyletic but also with low support (IQ/MB = NA/0.60; Fig. 2a). A clade of marine *Campylodiscus* was recovered as sister to all other Surirellaceae with PP=0.77 (Figs 2a and 3). *Auricula*, Rhopalodiales and Surirellaceae were a polytomy (Fig. 2a). In the maximum credibility tree from the Bayesian analysis, *Protokeelia* Reimer & Lee was reconstructed as sister to a clade composed of all other Rhopalodiales, Auriculaceae, and Surirellaceae (Fig. 2b).

In the IQ-TREE analysis, the clade of marine *Campylodiscus* was reconstructed, with low support, as sister to a clade that included Rhopalodiales + *Auricula flabelliformis* Voigt (IQ/MB = 79/NA; Fig. 2c). Relationships among *Protokeelia*, *A. flabelliformis*, and the remaining Rhopalodiales were also poorly supported (Fig. 2c). The remaining two species of *Auricula* were sister to this entire clade, again with poor support (IQ/MB = 83/NA; Fig. 2c). This clade of *Auricula*, Rhopalodiales and part of *Campylodiscus* was sister to a strongly supported clade of the remaining Surirellaceae (IQ/MB = 99/NA; Figs 2c and 3; Figs S1 and S2).

Monophyly of the remaining Surirellaceae—a clade that included the genera *Campylodiscus pro parte*, *Cymatopleura*, *Petrodictyon*, *Surirella*, and *Stenopterobia*—was strongly supported in both analyses (IQ/MB = 99/0.94; Fig. 3; Figs S1 and S2). *Petrodictyon*+*Surirella febigeri* Lewis was sister to either all remaining Surirellaceae (IQ-TREE UFboot = 59) or sister to a clade of Fastuosae groups of *Surirella* and *Campylodiscus*, the Pinnatae group of *Surirella*, the type of *Surirella* (*S. striatula*), and *Cymatopleura* (MrBayes, with support 0.55; Fig. 3, clades 5–10; Figs S1 and S2). The uncertain placement of *Petrodictyon* + *S. febigeri* aside, Surirellaceae can be divided into two large clades. The first of these clades included the mixed marine–freshwater clade of *Surirella* Fastuosae, *Campylodiscus* Fastuosi, *S. striatula*, *Surirella* Pinnatae and *Cymatopleura*, the monophyly of which was strongly supported in the Bayesian analysis but less so in the ML tree (IQ/MB = 89/1.00; Fig. 3, clades 5-10; Figs S1 and S2). The second Surirellaceae clade was comprised of *Surirella* Robustae, *Campylodiscus* Robusti, and *Stenopterobia* (IQ/MB = 100/1.00; Fig. 3, clade 11).

3.5. Evolution of cell architecture

Cell symmetry in Surirellales and Rhopalodiales can be divided in three more-or-less discrete states: cells longer along their apical axis (AA) than their transapical axis (TA) [AA > TA], cells circular [AA = TA], and cells longer along their transapical axis than their apical axis [AA < TA] (Fig 1a). If we assume the ancestral symmetry was [AA > TA]—as is the case for many other raphid pennates including the immediate outgroup (Ruck and Theriot, 2011)—then it is plausible that expansion in the transapical direction would inevitably lead to circular symmetry before arriving at the state in which the "short" transapical axis is actually longer than the "long" apical axis (Fig 1a). This stepwise evolutionary history in cell symmetry could be irreversible at one or both steps (Fig 1b–d). We tested three ordered pathway models that reflect the above scenarios and found that disallowing bidirectional changes provides a poor fit to the observed data (Table 1; Fig 1b–

d). Specifically, the estimates of the rate of transition from $[AA < TA]$ to $[AA = TA]$ are comparable to the rate of transition from $[AA > TA]$ to $[AA = TA]$ and from $[AA = TA]$ to $[AA < TA]$ (Table 1; Fig 1b-d). Therefore, disallowing $[AA = TA]$ to $[AA > TA]$ transition rate (ordered-reversible-2 model; Fig 1c) comes with no penalty, but completely removing all reversibility from the model (the ordered-irreversible model) offers a significantly worse fit (Table 1; Fig 1d). Optimizations on the best ML tree, whose topology differed in parts of the tree that potentially affected these reconstructions, qualitatively agreed with results obtained from the Bayesian population of trees and highlighted the robustness of our results with respect to the underlying phylogenetic hypothesis. These results strongly support the hypothesis that transitions between the circular and longer transapical axis states occurred in both directions, as observed, for example, in the derived position of *Campylodiscus Robusti* within the *Surirella Robustae* clade (Figs 3 and 4, clade 11).

The ordered models did not test specifically whether the evolution of cell wall architecture in Surirellales and Rhopalodiales followed a strictly stepwise model of the kind $[AA > TA] \rightarrow [AA = TA] \leftrightarrow [AA < TA]$ (Fig 1b-d). To test whether direct transitions from a longer apical to a longer transapical axis were possible, we compared the ordered models to an unordered, all-rates-different model (Fig. 1e). The unordered model had a lower relative likelihood due to the increased number of parameters but still represented a plausible scenario (Table 1). The ancestral state reconstructions under the ordered-reversible and the all-rates-different models were almost identical (Fig. 4). The one important difference was the ancestral state at the node representing the most-recent common ancestor of Surirellaceae. The ancestor of this clade was reconstructed as circular in shape under the ordered models, but under the all-rates-different model, the most-recent common ancestor of Surirellaceae was reconstructed as ambiguous between a form with a longer apical axis and a form with a longer transapical axis (Fig. 4). That is, there was nearly equal probability that the most-recent common ancestor of Surirellaceae had the symmetry typical of *Auricula* or of *Surirella* (Fig. 1a). Taken together, these results suggest that transition from $[AA > TA]$ to $[TA > AA]$ did not necessarily have to proceed through an intermediate circular shape. Alternatively, if such an intermediary circular shape was a required stepping stone, our results could be due the underrepresentation of intermediary forms in our phylogeny, due to extinction, taxon sampling, or both.

3.6 Salinity of ancestral habitats

As with many other groups of raphid pennate diatoms, Surirellales and Rhopalodiales have large cohorts of marine and freshwater species. Many taxa are most commonly encountered in brackish waters with intermediate salinity. The taxonomic and ecological breadth of species represented in our molecular phylogeny provided the first opportunity to investigate the evolutionary history of transitions between habitats of different salinity in this group. The SR-clade is ancestrally marine, so we hypothesized that the brackish state represents a stepping stone towards adaptation to strictly freshwater environments. To test this hypothesis, we compared a model that forced transitions in the direction of marine $\rightarrow$ brackish $\rightarrow$ freshwater (ordered-irreversible model, $k=2$) to a model that allowed all types of changes, including bidirectional transitions between fully marine and fully freshwater environments with no transitional brackish habitat (all-rates-different model, $k=6$). Ancestral state reconstructions across models revealed that the backbone of the phylogeny was entirely marine with incursions to brackish and, eventually, fresh waters later on in the evolutionary history of the lineage (Fig. 5). We found that the ordered model was significantly better (average $\Delta AICc = 8.02$; Table 2), supporting the hypothesis that the brackish environment was a stepping stone on the route to colonization of fully freshwater habitats (Table 2). Moreover, parameters for direct transitions between marine and freshwater habitats in the all-rates-different

model had zero estimates across phylogenies. To further investigate the possibility of reversible transitions between habitats of different salinities over evolutionary timescales, we compared the ordered-irreversible model ($k=2$) to an ordered-reversible model ($k=4$). The latter allowed freshwater diatoms to migrate back to brackish waters and brackish diatoms to recolonize fully marine environments. The irreversible model—which forbids recolonization of the ancestral habitat—was again favored, but the model allowing bidirectional transitions between brackish and freshwaters could not be ruled out ($\Delta\text{AICc} = 4.07$; Table 2). It is important to note, however, that given the current taxon sampling, not all freshwater lineages have a brackish representative (e.g. Fig. 5, clade 3, 11). In these cases, a hypothetical brackish transition state was inferred as a result of the directional constraint imposed by the best-fitting model. Across models, transitions from brackish to freshwater habitats were the fastest, which is expected given the small number of brackish taxa represented on the tree. In all, these results highlight the marine environment as the ancestral habitat for the SR-clade and the most important one, given that the main types of cell architecture evolved in marine clades. In summary, colonization of habitats with lower salinity (brackish or freshwater) by marine ancestors occurred at least seven times and proceeded in a stepwise manner, with the brackish state as intermediary prior to invasions of freshwater (Fig. 5; orange bars). In many cases, the brackish lineages went on to colonize and diversify in freshwaters (Fig. 5; red bars).

4. Discussion

4.1 Morphological evolution in the *Surirellales* + *Rhopalodiales*

The SR-clade lineage is defined by the presence of a canal raphe gained independently from the similar, but homoplasious, canal raphe system of Bacillariales (Ruck and Theriot, 2011). Diversification within SR-clade has resulted in an assemblage of species that is highly heterogeneous in many aspects of morphology and development. The evolution of a circumferential canal raphe system along with a fundamental switch in cell development, in which the major axis of expansion is perpendicular to the corresponding axis in most other bilaterally symmetrical diatoms (Fig 1a), are unique among raphid pennate diatoms. The evolution of these derived morphologies in *Surirellales* has, consequently, long attracted interest from diatom systematists (Mereschkowsky, 1904; Hustedt, 1929; reviewed in Ruck and Kociolek, 2004). Mereschkowsky (1904) proposed a hypothesis for the evolution of *Surirella*-type morphologies (raphe circumferential, cells elongated along transapical axis) through a series of "translational stages" beginning with *Auricula*-like forms (raphe marginal, semi- to fully-circumferential, cells elongated along apical axis). His model was based on the observation that, in *Auricula*, the marginal raphe can cover different fractions of the circumference of the cell, so he proposed that progressive increase in this fraction eventually led to the fully circumferential raphe typical of *Surirella*. Hustedt (1929) proposed a different scenario in which the *Surirella* body form was derived from *Nitzschia*, which we now know is only distantly related to *Surirella*. Neither author dealt explicitly with the observation that the major plane of development in *Surirella* is along the transapical axis and perpendicular to that found in *Auricula* and *Nitzschia*. Notably, these early hypotheses of the evolutionary history of the canal-raphe-bearing diatoms viewed the evolution of *Surirella*-type cell architecture without reference to *Campylodiscus*. Similarities between groups of *Surirella* and *Campylodiscus* were noted (Peragallo and Peragallo, 1897–1908), but *Campylodiscus* never featured in these hypotheses of translational stages despite its more-or-less circular shape, which represents a plausible intermediate between taxa with long apical axes and forms with long transapical axes. Instead, most authors have considered *Campylodiscus* as derived from *Surirella*-type morphology (e.g., Jurilj 1949).

The phylogeny of canal-raphe diatoms was first investigated with morphological cladistic analysis under the assumption of a single origin of the canal raphe system in diatoms (Ruck and Kociolek, 2004). Using Bacillariales as the outgroup, this analysis suggested that taxa with a circumferential raphe (Surirellaceae) were paraphyletic, with equivocal reconstruction as to the number of times a circumferential raphe evolved (Ruck and Kociolek, 2004). Since *Nitzschia* was used as an outgroup, this study could not contrast the evolutionary scenarios proposed by Mereschkowsky (1904) and Hustedt (1929). However, the analysis of Ruck and Kociolek (2004) recovered *Auricula* as paraphyletic, grading into a clade of taxa with circumferential raphe systems, supporting a single origin of this character in the SR-clade. The maximum likelihood and Bayesian trees reconstructed here differed with respect the relationships in this part of the tree, and neither method of tree inference provided strong support for alternative arrangements (Figs 2 and 3). Thus, our results were more equivocal with respect to Mereschkowsky's (1904) hypothesis of translational stages from *Auricula* to *Surirella* (Figs 2 and 3). This apparent discrepancy is likely due to the absence of Rhopalodiales in the morphology-derived tree of Ruck and Kociolek (2004), but, as our results show, Rhopalodiales are an important component for the resolution of this part of the phylogeny, and consequently the evolution of the circumferential raphe (Fig 2).

The evolution of circumferential raphe systems appears to be closely tied to the evolutionary shift in the major axis of development within the SR-clade, from cell elongation along the "long" apical axis to elongation along the ancestrally "short" transapical axis (Fig 1a). Our findings that ordered models were only marginally better than a model that allowed the circular state to be bypassed (Table 1; Fig. 4), imply that the developmental shift did not have to transition through a (sub)circular state. Support for alternative models was to a large degree dependent on the phylogenetic position of a clade of marine *Campylodiscus*, whose placement was variable and/or poorly supported across phylogenetic analyses (Figs 2 and 3, clade 4). Optimizations under different, but similarly plausible, models resulted in different ancestral state reconstructions for the most recent common ancestor of Surirellaceae (Table 1; Fig. 4). Consequently the cell wall architecture of this ancestor could not be reliably determined with this dataset (Fig. 4). However, if the clade of marine *Campylodiscus* remains sister to all other Surirellaceae after more complete taxon sampling, then the a model of ordered transitions with a (sub)circular intermediary state would be more plausible.

Our modeling results also supported bidirectional transitions between circular forms and "wide" forms that are longer about the transapical axis (Table 1; Fig. 4). This result suggested that cell architecture in Surirellaceae, where these two states are most common, is more labile compared to other diatom lineages. For comparison, transitions from circular forms to "narrow" forms longer along the apical axis were not inferred in either of our models (Fig. 4). The evolution of *Campylodiscus*-type cell architecture supports a longstanding hypothesis that at least some *Campylodiscus* were derived from *Surirella*-like ancestors (Figs 3-4, clade 11), perhaps evolving by incremental increase in torsion of the cell and rotation of the two valves relative each other (Jurilj, 1949). Cell torsion and valve rotation represent important architectural changes within Surirellaceae. However, apart from a very few *Surirella* species (e.g., *S. spiralis*, *S. iconella*, and *S. helissella*), cell torsion is restricted to various lineages of the polyphyletic *Campylodiscus*. As such, these characters have a nearly 1:1 correlation with the character state (sub)circular cell architecture. So although we did not explicitly consider these characters, the transition to twisted cells and 90° valve rotation are implicitly accounted for in the models.

Looking at the combination of major axis of development and presence of a circumferential raphe reveals that taxa with circumferential raphe systems but long apical axes—like *Auricula flabelliformis*, *Plagiodiscus* (two species) and *Hydrosilicon* (two species)—are either uncommon over evolutionary scales or possibly extinct (e.g. *Surirella baldjickii*

Norman; Ognjanova-Rumenova and Buczkó 2016; Schmidt et al., 1874). The diversity of taxa with this combination of characters is dwarfed by the number of SR-clade taxa with a longer apical axis and *no* circumferential raphe, or those with a longer transapical axis *and* a circumferential raphe (Fourtanier and Kociolek, 2011). It might be that the combination of longer apical axis and circumferential raphe was not a stable strategy over evolutionary timescales.

4.2 Ecological transitions across the salinity gradient

The SR-clade is an ecologically diverse lineage and one of many raphid diatom lineages with representatives in both marine and freshwaters. Diatom lineages are generally thought to have originated in marine environments and only rarely colonized freshwaters (Mann, 1999)—a presumably difficult transition that requires overcoming strong gradients in ionic concentrations and osmotic pressure (Kirst, 1996). It has also been hypothesized that established freshwater lineages rarely if ever recolonize marine habitats (Mann, 1999). The sole test of these hypotheses to date was undertaken within a largely planktonic clade of "centric" diatoms, with results showing that freshwater lineages can, and often do, transition back to their ancestral marine environment (Alverson et al., 2007). Similar investigations in clades of different ages or inhabiting the benthos, like the SR-clade, have not been carried out but are essential for understanding the general nature of salinity as a historical barrier to diatom distribution. Several of the taxa studied here inhabit waters with intermediate salinity, allowing us to study the role of brackish habitats as a possible transition along the pathway from marine to freshwaters and vice versa. Our analyses reconstructed the ancestral habitat of the backbone of the SR-clade phylogeny as unequivocally marine with up to seven independent colonizations of brackish or freshwater habitats (Fig. 5, orange and red bars). A stepping stone model of habitat transition, in which the brackish state served as an intermediary for marine colonists on their route to adaptation to fully freshwater habitats, was strongly supported (Table 2). Under this stepping stone model, all major clades had at least one incursion into habitats of low salinity (Fig. 5, orange bars). In most cases these brackish ancestors gave rise to freshwater lineages, two in Rhopalodiaceae, and one each in *Entomoneis*, the *Surirella striatula* + *Surirella* Pinnatae + *Cymatopleura* clade, and the *Surirella* Robustae + *Campylodiscus* Robusti + *Stenopterobia* clade (Fig. 5, red bars).

Since stepwise colonization of freshwater could be irreversible at either or both transition steps, we contrasted ordered models that disallowed bidirectional transitions to ordered models that allowed bidirectional transitions. We found that, although weakly supported, a model permitting bidirectional transitions was credible, suggesting that transitions from freshwaters back to brackish habitats could not be ruled out (Fig. 5b, clade 10, white bar). It is important to note, however, that freshwater → brackish transitions were inferred only within the *Surirella* Pinnatae lineage, and here, only in a clade of *Surirella brebissonii* + *Surirella brightwellii* composed of pairs of conspecific accessions collected from environments with different salinities (i.e., freshwater vs. brackish; Fig. 5b, clade 10, white bar). Moreover, under the reversible model, the ancestral states within this clade were ambiguous between the freshwater and brackish states complicating inferences about the directions of habitat transitions (Fig. 5b). Thus, in SR-clade, recolonization of habitats with higher salinity from a freshwater background occurred recently, and it is unclear if such events were followed by diversification within these secondarily brackish or marine lineages.

The ecological distribution of the paraphyletic genus *Surirella* spans the entirety of the salinity gradient, with considerable diversity in brackish habitats and freshwater habitats with high electrolyte content (e.g., mineral waters) (Krammer and Lange-Bertalot, 1988). Using models that forbid freshwater-to-brackish transitions (ordered-irreversible

model), the ancestors along the backbone of the lineage of *Surirella striatula*, the *Surirella* Pinnatae group, and *Cymatopleura* were all found to be brackish (Fig. 5a, clades 8–10). Under this constraint of irreversibility of freshwater colonization, as many as eight invasions of freshwater were inferred within this clade alone (Fig. 5a, clades 8–10, red bars). A simpler, though not necessarily more likely, scenario was provided by models that permitted freshwater diatoms to recolonize brackish waters (Table 2; Fig. 5b). Here, after the split between between *S. striatula* + *S. sella* and *Surirella* Pinnatae + *Cymatopleura*, the backbone was inferred to be more likely freshwater, with one recolonization of brackish water in the *Surirella brebissonii* + *Surirella brightwellii* clade (Fig. 5a, clade 10, white bar). Although these results expose some conflict between the competing evolutionary models, we would argue, instead, that the models are likely oversimplifying a more complex evolutionary process that our limited data on the ecological tolerances of these species have not fully captured. Some of these species are known to have relatively broad salinity tolerances (e.g., *S. brightwellii*, 0–10 ppt; Ruck personal observations), and it is possible that the distributions of many other taxa in this group span a substantial, though currently unknown, salinity range. Therefore, a more conservative interpretation is that the high number of invasions of freshwaters, or recolonizations of brackish waters, are a by-product of the stochasticity of field collections from a lineage whose members are euryhaline to varying degrees. Detailed investigation of this question would likely require broader geographic sampling or controlled experiments to determine the ranges of salinity tolerance for members of this clade.

4.3 A revised classification of the Surirellales + Rhopalodiales

Indications that the classification of the SR-clade lineage might not reflect evolutionary history for at least some genera have been present in the diatom literature since the turn of the 20th century. Peragallo and Peragallo (1897–1908) created informal groups within *Surirella* and *Campylodiscus* and noted morphological similarities between some of their informal groups that suggested that some *Surirella* (e.g., Robustae) might be more similar to some *Campylodiscus* (e.g., Robusti) than to other congeners. Morphological studies using scanning electron microscopy provided further evidence for parallels in fine-scale morphology, especially morphology of the canal raphe features, between informal groups of *Surirella* and *Campylodiscus* (Lange-Bertalot and Metzeltin, 1996; Paddock, 1985; Reichardt, 1996). Formal phylogenetic treatment of morphology within the Surirellales eventually showed that the classification scheme did not correspond to phylogenetic relationships (Ruck and Kociolek, 2004). Specifically, some informal groups of *Surirella* and *Campylodiscus*, e.g., Robustae + Robusti and Fastuosae + Fastuosi, were found to be sister to each other, and the two largest genera, *Surirella* and *Campylodiscus*, were each found to be polyphyletic (Ruck and Kociolek, 2004). Although initially with modest taxon sampling, molecular phylogenies supported phylogenetic reconstructions based on morphological characters, and by revealing that Rhopalodiales are embedded within Surirellales, also showed that the classification of these two groups could be not treated separately (Ruck and Theriot, 2011).

Our findings expose further evidence for nonmonophyly at the family and genus levels (Fig. 3). The monogeneric family Auriculaceae is still poorly sampled and its monophyly and relationships to the Rhopalodiaceae and Surirellaceae are uncertain (Fig. 3; Figs S1 and S2). Monophyly of the Rhopalodiaceae is dependent on the placement of a single representative of *Protokeelia* (Figs 2–4). The genus *Rhopalodia* is paraphyletic with respect to *Epithemia* (Fig. 3) and—by extension—*Tetralunata*, due to the sister relationship between *Epithemia* and *Tetralunata* (Hamsher et al., 2014). *Surirella* and *Campylodiscus* are polyphyletic, and *Stenopterobia* is embedded within a clade of *Surirella* Robustae and

Campylodiscus Robusti. The type of *Surirella*, *S. striatula*, is sister to a clade of *Cymatopleura* + *Surirella* Pinnatae, and *Petrodictyon* is sister to *S. febigeri* (Fig. 3; Figs S1 and S2). In short, it is clear that the current classification scheme grossly misrepresents the phylogenetic relationships as inferred by both morphological and molecular data, and that substantial revision of generic boundaries is needed before we arrive a set of names that represent monophyletic groups.

To this end, we propose a partial reclassification of the SR-clade that will resolve some of the major issues outlined above. Since the Surirellales are paraphyletic with respect to the Rhopalodiales, these two names cannot coexist, and given that they were proposed in the same publication (Round et al. 1990), we retain the name Surirellales simply because it currently encompasses a broader range of taxa. The alternative, to divide the Surirellales into two orders—one for Entomoneidaceae and another for Surirellaceae + Auriculaceae—would overlook the uncertain phylogenetic position of *Auricula* (Figs 2 and 3). At the family level, only the monophyly of Entomoneidaceae appears highly supported and robust across analyses. Therefore, the most conservative approach is to avoid naming family-level taxa until datasets with more power to resolve these high-level relationships have been assembled. Nonmonophyly at the genus level requires decisions and tradeoffs that balance sinking some generic names into others, or creating (sometimes many) new names, to reflect monophyletic groups. To avoid a proliferation of new names, whenever possible we followed the former approach of establishing larger, more inclusive genera with the goal of retaining as many old names as possible, which we hope will facilitate the transition to a new phylogenetic classification of SR-clade. We detail our new proposed classification below and provide a list of required nomenclatural transfers in the supplementary material (List S1), focused only on taxa included in our phylogeny.

ORDER SURIRELLALES D.G. MANN 1990

Comments: The names Surirellales and Rhopalodiales appeared simultaneously (D. G. Mann in Round et al., 1990). We elected to retain Surirellales because as described it encompasses a larger fraction of the diversity of the lineage.

Synapomorphy: The presence of a canal raphe, together with a single variably lobed plastid positioned against valve face, distinguishes Surirellales from canal-raphe-bearing Bacillariales.

AURICULA CASTRACANE 1873 (FIG. 3, CLADE 2) – MONOPHYLY UNCERTAIN

ENTOMONEIS (EHRENBERG) EHRENBERG 1845 (FIG. 3, CLADE 1)

Synapomorphy: Fibulate canal raphe raised on a sigmoid keel (Fig. 6a, b).

(See below for comments on an addition to this family proposed by Lange-Bertalot et al. 2015)

PROTOKEELIA REIMER & LEE 1984 - MONOPHYLY UNCERTAIN

EPITHEMIA KÜTZING 1844 (RHOPALODIA O. MÜLLER 1895 AND TETRALUNATA HAMSHER, GRAEFF, STEPANEK & KOCIOLEK 2014; FIG. 3, CLADE 3)

Comments: A recent phylogenetic study involving genera related to *Rhopalodia* found that lunate areolar openings created by flap-like occlusions (volae) and fibulae that span the entire width of valve ('entire' fibulae) were synapomorphies for the group of *Rhopalodia*, *Epithemia*, and *Tetralunata* (Hamsher et al., 2014). That study, however, included only one

Rhopalodia species (unnamed), that could not have represented the full range of morphological variability present in the grade of marine and freshwater *Rhopalodia*. This variability also extends to the characters used by Hamsher et al. (2014) for their phylogenetic analyses. For example, the 'entire' fibulae synapomorphy for the *Rhopalodia*, *Epithemia*, and *Tetralunata* identified by Hamsher et al. (2014) breaks down once marine *Rhopalodia* with shorter fibulae are considered (Fig. 7b). The character 'entire' fibulae used by Hamsher et al. (2014) is also not a synapomorphy for the clade that includes only the freshwater *Rhopalodia*, *Epithemia* and *Tetralunata*, as 'entire' fibulae are present in some of marine *Rhopalodia* species outside this freshwater clade. The same applies to the second synapomorphy, flap-like occlusions, that are also present in at least some marine *Rhopalodia* (Fig. 7b). Thus, by including only a single representative from *Rhopalodia* and *Tetralunata*, Hamsher et al. (2014) were not able to test putative synapomorphies or the monophyly of either genus.

Our results showed that the marine *Rhopalodia* grade into a clade of freshwater *Rhopalodia* + *Epithemia* (Fig. 7b). Thus, retaining *Epithemia* and *Tetralunata* as separate genera also retains a paraphyletic *Rhopalodia*. One way to rectify this problem is to subsume *Tetralunata* and *Rhopalodia* into *Epithemia*, creating a large monophyletic group defined by fibulae that extend beyond the canal raphe and onto valve face (Fig. 7 b,c). Another option is to restrict *Rhopalodia* to the mostly freshwater lineage sister to *Epithemia* and relegate the grade of marine *Rhopalodia* to *incertae sedis*. Following the second option preserves the current generic names, and under this scheme, the name *Rhopalodia* would correspond to the freshwater clade because it includes the type species, *R. gibba* (Ehrenberg) O. Müller (Fig. 3, clade 3). This scheme requires diagnosis of *Rhopalodia sensu stricto*, the clade of freshwater *Rhopalodia* species sister to *Epithemia*, to the exclusion of marine *Rhopalodia*. However, given the problems with synapomorphies outlined above, diagnosing *Rhopalodia sensu stricto*—to the exclusion of the marine *Rhopalodia* grade—is problematic. For example, the apomorphy for *Rhopalodia* proposed by Hamsher et al. (2014), a thickened raised structure holding the raphe (the 'keel'), is found in species across the *Rhopalodia* grade and so cannot be a synapomorphy for *Rhopalodia sensu stricto*. Finally, preserving the current generic names by restricting the definition of *Rhopalodia* to include only the freshwater clade sister to *Epithemia*, requires diagnosis of two additional genera to accommodate the grade leading to the crown clade of freshwater *Rhopalodia* and *Epithemia*. It is therefore most productive to subsume *Rhopalodia* and *Tetralunata* into *Epithemia*, creating a monophyletic genus defined by fibulae that extend beyond the canal raphe and into valve face.

This decision tries to balance several competing interests, many of which represent longstanding challenges in taxonomy and classification. These include: 1) minimizing overall disruption to the taxonomy of this group, 2) weighing the practical costs and benefits of establishing a single large genus *versus* three smaller genera, 3) minimizing the proliferation of new genus names, in this case to small clades that are difficult to diagnose, 4) recognizing the slow rate of historical progress in establishing a natural classification of this group, and 5) eliminating non-monophyletic groups that lack biological cohesion and meaning. In the end, we decided that the only tangible "harmful" outcome would be the preservation of non-monophyletic groups. Although in many ways we preferred to retain the name *Rhopalodia* for the freshwater clade sister to *Epithemia* and to relegate the remaining *Rhopalodia* to *incertae sedis*, allowing these to be resolved in future studies, decades may pass before such a solution is made. To our knowledge, Hamsher et al. (2014) was the sole paper in the last decade dealing with the systematics of the Rhopalodiales. Ruck & Kociolek (2004) and Ruck & Theriot (2011) were the only papers in the last decade focused on the systematics of Surirellales. The rate at which studies on the systematics of these groups are published suggests that these problems would not be resolved soon. We believe, therefore, that our proposed taxonomy is the most productive, practical, and least disruptive solution.

Synapomorphy: Fibulae that extend into the valve face (Fig. 7 b,c).

CAMPYLODISCUS (EHRENBERG) EHRENBERG EX KÜTZING 1844 (*C. CLYPEUS* (EHRENBERG) KÜTZING + *C. CF. BICOSTATUS* GRUNOW CLADE + *SURIRELLA* FASTUOSAE + *C. FASTUOSUS* EHRENBERG; FIG. 3, CLADE 6, 7)

Comments: Both the Bayesian maximum clade credibility tree and the best maximum likelihood phylogeny reconstructed a weakly supported monophyletic group composed of *Campylodiscus clypeus* + allies and the *Surirella* Fastuosae group. This clade was dissolved into a polytomy in the Bayesian majority rule consensus topology, however, filtering of the posterior distribution of trees found that such grouping cannot be rejected (posterior probability=0.37). Communication between the raphe canal and cell interior within this clade is through funnel- or chalice-shaped structure (infundibulum) that widens towards the canal raphe, which in turn is supported by a fibular plate (*Surirella fastuosa*; Fig. 9a,b). *C. fastuosus* did not fall within this clade, however it does have an infundibulum with simple strut-like fibulae (Fig. 9c,d) and its inclusion within the *C. clypeus* + allies and the *Surirella* Fastuosae clade could not be rejected (posterior probability=0.35). Moreover, the infundibulum was identified as a synapomorphy for this group of species in the morphological cladogram of Ruck and Kociolek (2004). Since *C. clypeus* is the type of the genus *Campylodiscus*, this clade retains that name and a number of *Surirella* Fastuosae group species are transferred to it (see List S1).

Synapomorphy: Communication between the raphe canal and cell interior through a funnel- or chalice-shaped structure (infundibulum; Fig. 9; see also Ruck and Kociolek 2004).

CORONIA C.G. EHRENBERG 1840 (*CAMPYLODISCUS* ECHENEIDII, *CAMPYLODISCUS* EXIMII, *CAMPYLODISCUS* DECORATI, *CAMPYLODISCUS* LIMBATI; FIG. 3, CLADE 4)

Comments: The diagnosis of *Campylodiscus* as the lineage that includes taxa with infundibula orphaned the taxa formerly placed in *Campylodiscus* but lacking the synapomorphy of *Campylodiscus sensu stricto*. Among these orphaned *Campylodiscus*, the Robusti group was transferred to *Iconella*, leaving only a clade of marine *Campylodiscus* without clear generic placement (Fig. 3, clade 4). This is a strongly supported clade and is morphologically diverse, containing members of the taxonomic sections Echeneidii, Eximii, Decorati, and Limbati. Despite strong support for monophyly, we were unable to identify a morphological synapomorphy. Nevertheless, we believe that naming this lineage, even without a synapomorphy, is preferable compared to placing these taxa under *incertae sedis*, which in practice would mean preserving a polyphyletic *Campylodiscus*.

The name *Coronia* was introduced by Ehrenberg (1840) as subgenus of *Campylodiscus* (Fourtanier and Kociolek 2011) and included the species *Campylodiscus (Coronia) echeneis* Ehrenberg. This species is present in our phylogeny (Fig. 3, clade 4) and because it is the type of the subgenus, it becomes the type of the genus *Coronia*.

Type species: *Coronia echeneis* Ehrenberg 1840, p. 206.

ICONELLA JURILJ 1949 (*SURIRELLA* ROBUSTAE, *CAMPYLODISCUS* ROBUSTI, *STENOPTEROBIA* BRÉBISSEON EX VAN HEURCK 1896; FIG. 3, CLADE 11)

Comments: The genus *Iconella* was created by Jurilj (1949) to accommodate a robustoid species, *Iconella variabilis* Jurilj (1949), from Lake Ohrid. Later, Jurilj reversed his opinion and placed *I. variabilis*, as well as several other species which he originally assigned to small or monotypic genera, into *Surirella* (Jurilj 1954). The robustoid *Surirella* and

Campylodiscus, together with *Stenopterobia*, form a strongly supported clade, and neither *Campylodiscus* nor *Surirella* are monophyletic, so there is no basis to keep members of this clade in their current genera. Since *Iconella* is a validly published name, it can be applied to the robustoid clade (Fig. 3, clade 11) with *I. variabilis* as the type.

The alar canal was thought to be the synapomorphy for this group (Ruck and Kociolek, 2004), but alar canals have since been documented in marine species of *Campylodiscus* in the Decorati and Eximii groups (Fig. 3, clade 4; Fig. 8e-f). Consequently, presence of an alar canal is not a synapomorphy for the clade comprised of robustoid *Surirella* species, robustoid *Campylodiscus*, and *Stenopterobia* (here joined into *Iconella*). However, the alar canals of *Iconella* are uniformly porous, whereas perforations of alar canals in marine *Campylodiscus sensu lato* are usually patterned (Fig. 8e). The simple, rimmed pores found in *Iconella* are absent in marine *Campylodiscus sensu lato* that have alar canals (Fig. 8e).

Synapomorphy: Uniformly porous alar canals (Fig. 8 a,c).

TYPE SPECIES: *Iconella variabilis* Jurilj 1949, p. 185; figs 10 a-b, 11 a-b.

PETRODICTYON (EHRENBERG) D. G. MANN 1990 (FIG. 3, CLADE 5)

SURIRELLA TURPIN 1828 (*S. STRIATULA* TURPIN + *S. SELLA* HUSTEDT IN SCHMIDT ET AL. 1925, *SURIRELLA* PINNATAE, *CYMATOPLEURA* W. SMITH; FIG. 3, CLADE 8-10)

Comments: *Cymatopleura* differ from the *Surirella* Pinnatae group by their discontinuous internal raphe fissure at one pole (Ruck and Kociolek, 2004) and, in some species, the presence of longitudinal undulations. However, undulations are not universally present among species assigned to *Cymatopleura*, and the morphology of the raphe fissure in various groups of Surirellaceae is variable as well (see comments on *S. splendida* in Ruck and Kociolek, 2004). Otherwise, cell morphology, especially the communication between the raphe canal and the cell interior are consistent within this clade, which includes *S. striatula* (the type of *Surirella*), the *Surirella* Pinnatae group, and *Cymatopleura* (Fig. 10a-c). Thus, retaining *Cymatopleura* as a separate genus would reduce *Surirella* to a few nominal species and require the establishment of another genus to accommodate the *Surirella* Pinnatae group. Moreover, we know of no synapomorphies for *Cymatopleura* or the *Surirella* Pinnatae group, so both would be difficult to diagnose morphologically. We therefore conclude that subsuming *Cymatopleura* into *Surirella* is the best available solution given the data.

Synapomorphy: Communication between the raphe canal and cell interior is through simple portulae created by fibulae of various thickness (Fig. 10a-c; type 1 and type 2 fibulae in Ruck and Kociolek, 2004).

INCERTAE SEDIS

HYDROSILICON BRUN 1891

PLAGIODISCUS GRUNOW 1867

PLATICHTHYS LANGE-BERTALOT, KULIKOVSKIY, WITKOWSKI, SEDDON AND KOCIOLEK 2015

Comments: *Platichthys* Lange-Bertalot et al. was recently described and tentatively placed within the family Entomoneidaceae (Lange-Bertalot et al., 2015). This new genus is morphologically distinct from *Entomoneis*, lacking a sigmoid raphe, being linear-elliptical in girdle view (not bilobate) and having very fine, dense, uniseriate striae made up of simple areolae. The common features include the solid valve mantle and multiple (usually >4) components of the girdle.

Although similar in some respects to *Entomoneis*, the phylogenetic position of *Platichthys* is unknown and is included here for completeness.

SURIRELLA FEBIGERI LEWIS 1861

Acknowledgements

We thank the reviewers for helpful comments, especially the ones that prompted us to more carefully consider our proposed reclassification. ECR would like to extend special thanks to David Czarnecki, Mark Edlund, Becky Bixby, Joshua McDill, Matt Ashworth, Chris Lobban, Andrzej Witkowski, Malgorzata Bak, Thomas Hübener, and Kristina Buczkó for sharing cultures or help with field collections. Computational analyses used resources available through the University of Arkansas High Performance Computing Center, which is supported by numerous NSF awards. This research was supported by NSF EF 0629410 (ECT), Jane and Roland Blumberg Centennial Professorship in Molecular Evolution (ECT), NSF DEB-1353131 (AJA), and a Lewis & Clark award from the American Philosophical Society (ECR).

References

- Alverson, A.J., Jansen, R.K., Theriot, E.C., 2007. Bridging the Rubicon: Phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Mol. Phylogenet. Evol.* 45, 193–210. doi:10.1016/j.ympev.2007.03.024
- Beaulieu, J.M., O'Meara, B.C., Donoghue, M.J., 2013. Identifying hidden rate changes in the evolution of a binary morphological character: The evolution of plant habit in campanulid angiosperms. *Syst. Biol.* 62, 725–737. doi:10.1093/sysbio/syt034
- Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M., Pande, N., Shang, Z., Yu, N., Gutell, R.R., 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3, 2.
- Czarnecki, D.B., 1987. Freshwater diatom culture collection at Loras College, Dubuque, Iowa: *Notulae Naturae of The Academy of Natural Sciences of Philadelphia*, No. 465. *Not. Naturae* 465, 1–16.
- Ehrenberg, C.G., 1840. Charakteristik von 274 neuen Arten von Infusorien. *Ber. Akad. Wiss. Berlin* 1840:197–219.
- Fourtanier, E., Kociolek, J.P., 2011. Catalogue of Diatom Names, California Academy of Sciences, On-line Version [WWW Document]. URL <http://research.calacademy.org/research/diatoms/names/index.asp> (accessed 9.15.14).
- Guillard, R., Hargraves, P., 1993. *Stichochrysis immobilis* is a diatom, not a chrysophyte. *Phycologia* 32, 234–236.
- Guillard, R.R., Lorenzen, C.J., 1972. Yellow-green algae with Chlorophyllide C1, 2. *J. Phycol.* 8, 10–14.
- Guiry, M.D., Guiry, G.M., 2016. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. [WWW Document]. URL <http://www.algaebase.org> (accessed 5.2.16).
- Hamsher, S.E., Graeff, C.L., Stepanek, J.G., Kociolek, J.P., 2014. Frustular morphology and polyphyly in freshwater *Denticula* (Bacillariophyceae) species, and the description of *Tetralunata* gen. nov. (Epithemiaceae, Rhopalodiales). *Plant Ecol. Evol.* 147, 346–365.
- Harper, M.A. 1977. Movements, in: Werner, D. (Eds.), *The biology of diatoms*, pp. 224–249.

- Hustedt, F., 1929. Untersuchungen über den Bau der Diatomeen. VII-VIII. VII. Weitere Untersuchungen über die Kanalarphe der Nitzschioideae. VIII. Untersuchungen über die Kanalarphe der Gattung *Surirella*. Ber Dtsch. Bot Ges 47, 101–111.
- Idei, M., Sato, S., Watanabe, T., Nagumo, T., Mann, D.G., 2013. Sexual reproduction and auxospore structure in *Diploneis papula* (Bacillariophyta). Phycologia 52, 295–308.
- Jurilj, A., 1949. New diatoms, Surirellaceae, of Ochrida Lake in Yugoslavia and their phylogenetic significance. Jugosl. Acad. Sci. Arts Zagreb 24, 1–94.
- Jurilj, A., 1954. Flora and vegetation of Diatoms from Ochrida Lake in Yugoslavia. Jugosl. Acad. Sci. Arts Zagreb 26:99–190.
- Kilham, S., Kreeger, D., Lynn, S., 1998. COMBO: a defined freshwater culture medium for algae and zooplankton. Hydrobiologia 377, 147–159.
- Kirst, G.O., 1996. Osmotic adjustment in phytoplankton and macroalgae, in: Kiene, R.P., Visscher, P.T., Keller, M.D., Kirst, G.O. (Eds.), Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds. Springer US, pp. 121–129.
- Krammer, K., 1989a. Valve morphology and taxonomy in the genus *Stenopterobia* (Bacillariophyceae). Br. Phycol. J. 24, 237–243.
- Krammer, K., 1989b. Functional valve morphology in some *Surirella* species (Bacillariophyceae) and a comparison with the Naviculaceae. J. Phycol. 25, 159–167.
- Krammer, K., Lange-Bertalot, H., 1988. Bacillariophyceae. 2. Teil: Bacillariaceae, Epithemiaceae, Surirellaceae., Süßwasserflora von Mitteleuropa. VEB Gustav Fischer Verlag, Jena.
- Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.
- Lange-Bertalot, H., Metzeltin, D., 1996. Ultrastructure of *Surirella desikacharyi* sp. nov. and *Campylodiscus indianorum* sp. nov. (Bacillariophyta) in comparison with some other taxa of the “robustoid” type of construction. Nova Hedwig. Beih. 112, 321–328.
- Lange-Bertalot, H., Witkowski, A., Kulikovskiy, M., Seddon, A.W., Kociolek, J.P., 2015. Taxonomy, frustular morphology and systematics of *Platichthys*, a new genus of canal raphe bearing diatoms within the Entomoneidaceae. Phytotaxa 236, 135–149.
- Lowe, R.L., 2003. Keeled and cannaled raphid diatoms, in: Sheath, R.G., Wehr, J.D. (Eds.), Freshwater Algae of North America. Academic Press, Burlington, pp. 669–684.
- Maddison, W.P., Maddison, D.R., 2015. Mesquite: a modular system for evolutionary analysis.
- Mann, D.G., 1999. Crossing the Rubicon: The effectiveness of the marine/freshwater interface as a barrier to the migration of diatom germplasm. in: Ama, S., Idei, M., Koizumi, I. (Eds.), Proceedings Of The 14th International Diatom Symposium. Koeltz Scientific Books, Koenigstein, pp. 1–21.
- Mereschkowsky, C., 1904. Loi de translation des stades chez les diatomées. J. Bot. 18, 17–29.
- Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195.

- Nakov, T., Ruck, E.C., Galachyants, Y., Spaulding, S.A., Theriot, E.C., 2014. Molecular phylogeny of the Cymbellales (Bacillariophyceae, Heterokontophyta) with a comparison of models for accommodating rate variation across sites. *Phycologia* 53, 359–373. doi:10.2216/14-002.1
- Nawrocki, E.P., 2009. Structural RNA homology search and alignment using covariance models. Electron. Theses Diss.
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.
- Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2008. AWTY (Are we there yet?): A system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinforma. Oxf. Engl.* 24, 581–583. doi:10.1093/bioinformatics/btm388
- Ognjanova-Rumenova, N., Buczkó, K. 2016. Taxonomical re-investigation of Sarmatian diatoms from NE Bulgaria, hosted in the Pantocsek Collection (Budapest). *Geol. Balc.* 44, 39-50.
- Paddock, T., 1985. Observations and comments on the diatoms *Surirella fastuosa* and *Campylodiscus fastuosus* and on other species of similar appearance. *Nova Hedwig.* 41, 417–444.
- Peragallo, H., Peragallo, M., 1897. Diatomées marines de France et des districts maritimes voisins. Grez-sur-Loing.
- Poulsen, N.C., Spector, I., Spurck, T.P., Schultz, T.F., Wetherbee, R. 1999. Diatom gliding is the result of an actin-myosin motility system. *Cell Motil. Cytoskeleton* 44, 23–33.
- Reichardt, E., 1996. Die identität von *Campylodiscus levanderi* Hustedt. *Diatom Res.* 11, 81–87.
- Richlen, M.L., Barber, P.H., 2005. A technique for the rapid extraction of microalgal DNA from single live and preserved cells. *Mol. Ecol. Notes* 5, 688–691. doi:10.1111/j.1471-8286.2005.01032.x
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi:10.1093/sysbio/sys029
- Round, F.E., Crawford, R.M., Mann, D.G., 1990. *Diatoms: Biology and Morphology of the Genera*. Cambridge University Press.
- Ruck, E.C., Kociolek, P., 2004. Preliminary Phylogeny of the Family Surirellaceae-(Bacillariophyta), *Bibliotheca Diatomologica*. J. Cramer, Berlin, Stuttgart.
- Ruck, E.C., Theriot, E.C., 2011. Origin and evolution of the canal raphe system in diatoms. *Protist* 162, 723–737. doi:10.1016/j.protis.2011.02.003
- Sabbe, K., Chepurinov, V.A., Mann, D.G., Vyverman, W., 2004. Sexual behaviour, auxosporulation and chloroplast dynamics in a marine *Amphora* (Bacillariophyceae) studied in culture. *Bot. Mar.* 47, 53–63.
- Schmidt, A., Schmidt, M., Fricke, F., Heiden, H., Muller, O., Hustedt, F., 1874. *Atlas der Diatomaceen-kunde*. Koeltz, Leipzig.
- Scholin, C.A., Herzog, M., Sogin, M., Anderson, D.M., 1994. Identification of group- and strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae). II Sequence analysis of a fragment of the LSU rRNA gene. *J. Phycol.* 30, 999–1011.
- Sims, P.A., Paddock, T.B.B., 1982. The fenestral fibula: A new structure in the diatoms. *Bacillaria* 5, 7–42.
- Smith, S.A., O’Meara, B.C., 2012. TreePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinforma. Oxf. Engl.* 28, 2689–2690. Doi:10.1093/bioinformatics/bts492.

- Stepanek, J.G., Kociolek, J.P., 2014. Molecular phylogeny of *Amphora sensu lato* (Bacillariophyta): An investigation into the monophyly and classification of the amphoroid diatoms. *Protist* 165, 177–195.
doi:10.1016/j.protis.2014.02.002
- Theriot, E.C., Ashworth, M.P., Nakov, T., Ruck, E., Jansen, R.K., 2015. Dissecting signal and noise in diatom chloroplast protein encoding genes with phylogenetic information profiling. *Mol. Phylogenet. Evol.* 89, 28–36.
doi:10.1016/j.ympev.2015.03.012
- Warren, D.L., Geneva, A., Swofford, D.L., Lanfear, R., 2016. rwtY: R We There Yet? A package for visualizing MCMC convergence in phylogenetics.
- Yang, Z., 1995. A space-time process model for the evolution of DNA sequences. *Genetics* 139, 993–1005.

Figure 1. Contrasting cell architectures in Surirellales and Rhopalodiales. a) Light microscope photographs of exemplar genera and simplified diagrams of the relationship between the raphe and major developmental axis with description of the three typical arrangements: i) apical axis longer, raphe semi- to nearly fully-circumferential (*Entomoneis*, *Epithemia*, *Auricula*), ii) cell circular, raphe circumferential (*Campylodiscus*), and iii) transapical axis longer, raphe circumferential (*Surirella*). Micrographs are not to scale. b–e) pathway models for the evolution of these three character states. See text for details.

Figure 2. Simplified diagrams of the alternative reconstructions of the relationships between Auriculaceae, Rhopalodiaceae and Surirellaceae based on the Bayesian and maximum likelihood phylogenetic analyses. Majority rule consensus topology (a) and maximum clade credibility tree (b) from the combined post-burnin posterior distributions of the four MrBayes runs. (c) The “best” maximum likelihood tree as inferred by IQ-TREE. Numbers at nodes represent Bayesian posterior probabilities or ultra-fast ML bootstrap values. See Figs S1 and S2 for complete phylograms.

Figure 3. Majority rule consensus phylogram of the post-burnin distributions of the four MrBayes runs. The size and shading of symbols at nodes are proportional to the posterior probabilities. Branch background colors represent the current classification and clade annotations show the reclassification proposed here. Conspecific taxa were removed for clarity. See Figs S1 and S2 for complete phylograms inferred by ML and BI analyses.

Figure 4. Distribution of observed character states and ancestral state reconstructions for the evolution of cell architecture in Surirellales and Rhopalodiales. The phylogeny is the maximum credibility tree obtained by summarizing the post-burnin distributions from the four MrBayes runs. Pie charts represent the probabilities of ancestral states as reconstructed by the ordered-reversible-2 model (see text for details). Apical axis longer than transapical ([AA > TA]; blue), apical axis approximately equal to transapical axis ([AA = TA]; green) and apical axis shorter than transapical ([AA < TA]; grey). Reconstructions from ordered and unordered models are shown for the key node representing the most-recent common ancestor of Surirellaceae.

Figure 5. Distribution of observed character states and ancestral state reconstructions for marine, brackish and freshwater habitats in Surirellales and Rhopalodiales. Shown are reconstructions under the ordered-irreversible model (a) and ordered-reversible model for *Surirella sensu stricto* (*S. striatula* + *Surirella* Pinnatae + *Cymatopleura* clade) (b). The phylogeny is the maximum credibility tree obtained by summarizing the post-burnin distributions from the four MrBayes runs. Pie charts represent the probabilities of ancestral states: marine (grey), brackish (orange), and freshwater (red).

Figure 6. Scanning electron micrographs of *Entomoneis*. External valve view showing sigmoid raphe positioned on a keel (a) and an internal valve view showing the fibulae (b). Scale bars (a) 1 μm , (b) 5 μm .

Figure 7. Scanning electron micrographs. Internal valve view of an *Auricula* with small fine fibulae restricted to the raphe canal (a), *Rhopalodia iriomotensis* with marginal fibulae, fibulae that extend partway across internal valve face, and flap-

like occlusions of the pores (b), and *Epithemia* where the fibulae extend across entire internal valve face (c). Scale bars (a) and (b) 1 μm , (c) 2 μm .

Figure 8. Scanning electron micrographs. External and internal valve views of alar canal morphologies in *Iconella* compared to alar canals of an exemplar from the *Campylodiscus sensu lato* clade. *Iconella tenera* (a, b) and *I. hibernicus* (c, d) with moderate to high keels and alar canals with fenestral openings occluded by fenestral bars (a, c), internally the valve alar canal openings are round to elliptical, uniformly porous and pores are rimmed (b, d). *Campylodiscus samoensis* with a moderate keel height and alar canals alternating with fenestral openings that are often unoccluded (e), the internal opening of the alar canal is rhomboid, with patterned poration and simple (unrimmed) pores (f). Scale bars (a-f) 1 μm .

Figure 9. Scanning electron micrographs. External whole valve views of *Campylodiscus neofastuosus* (a) and *Campylodiscus fastuosus* (c). Internal valve view of infundibula of *Campylodiscus* with type II fibulae and fibular plate (b) compared to *C. fastuosus* infundibula with simple type I (strut-like) fibulae (d). Scale bars (a, b, d) 2 μm , (c) 5 μm .

Figure 10. Scanning electron micrographs. Internal valve view of the communication between the canal raphe and the cell interior through simple portulae formed by fibulae of various thickness (see also Ruck and Kocilek, 2004). (a) *Surirella striatula*, the type of *Surirella*, b) *Surirella brebissonii*, from the *Surirella* Pinnatae group, and c) *Surirella solea*, formerly *Cymatopleura solea*. Scale bars (a) 5 μm , (b, c) 1 μm .

Table 1. Model selection for the evolution of cell wall architecture in Surirellales + Rhopalodiales. Shown are the number of parameters (k), average log-likelihood from 10^3 optimizations, average Akaike score corrected for sample size, the difference in Akaike scores and the Akaike weight of each model (ω).

Model	k	Mean lnL	Mean AICc	Δ AICc	ω
Ordered-irreversible	2	-47.818	99.726	20.248	0.0
Ordered-reversible-2	3	-36.648	79.478	0	0.455
Ordered-reversible	4	-36.231	80.767	1.288	0.234
Unordered-ARD	6	-33.740	80.132	0.653	0.321

Table 2. Model selection for the evolutionary history of transitions between marine, brackish, and freshwater habitats in Surirellales + Rhopalodiales. Shown are the number of parameters (k), average log-likelihood from 10^3 optimizations, average Akaike score corrected for sample size, the difference in Akaike scores and the Akaike weight of each model (ω).

Model	k	Mean lnL	Mean AICc	Δ AICc	ω
Ordered-irreversible	2	-50.93	105.95	0.0	0.87
Ordered-reversible	4	-50.86	110.03	4.08	0.11
Unordered-ARD	6	-50.66	113.97	8.02	0.02

Figure 1

Fig 2

Figure 3

Figure 4

Figure 6

Figure 7

Figure 8

Figure 9

Figure 10

*Graphical Abstract