Molecular Phylogenetics and Evolution 170 (2022) 107433

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Molecular signature characters complement taxonomic diagnoses: A bioinformatic approach exemplified by ciliated protists (Ciliophora, Oligotrichea)

Maximilian H. Ganser [a,*], Luciana F. Santoferrara [b], Sabine Agatha [a,*]

[a] Department of Environment & Biodiversity, Paris Lodron University of Salzburg, 5020 Salzburg, Austria
[b] Department of Biology, Hofstra University, 11549 New York, USA

ARTICLE INFO

Keywords:
DNA sequence alignment
Integrative taxonomy
rRNA genes
Diagnostic character
Taxon identification

ABSTRACT

Traditionally, taxa following the botanical or zoological codes of nomenclature are diagnosed mainly by morphological characters, although integrative taxonomy advocates including additional features. While many taxonomic studies include DNA sequence analyses, a systematic integration of diagnostic molecular characters (signature characters) is still rare. Here, we suggest a practical guideline for the detection and evaluation of signature characters that provides the means necessary to complement diagnoses and facilitates identifications. The guideline comprises generally applicable criteria exemplified by a case study on an ecologically important group of planktonic protists, the Oligotrichea. The detection of signature characters and their discrete states in multiple sequence alignments is facilitated by the recently developed tool DeSignate. Moreover, we introduce a novel bioinformatic approach to test the influence of different alignment programs on the consistency of signature characters. Our workflow enabled detection of consensus signature characters for most tested taxa and inclusion of such characters in the diagnoses of three orders, eight families, and two genera in the Oligotrichea. The suggested approach is a step towards an integrative taxonomy linking reliable molecular sequence data to organisms' traits.

1. Introduction

Systematics is based on the similarity of taxa in putative homologous features that are used to infer evolutionary relationships. Taxa are per definition monophyletic groupings that have been established in accordance with the codes of nomenclature. While taxonomy was traditionally based on morphological traits, integrative approaches were fostered by the ability of molecular phylogenies to discover non-monophylies. By compiling all available data from different bodies of evidence for taxon descriptions, the molecular data are linked with information about the organism's traits (morphology, lifestyle, etc.) that are key for understanding biological, evolutionary, and ecological processes. Achieving this goal requires practical means to incorporate formally molecular data in taxonomic diagnoses (for example, Recommendation 13A in the International Code of Zoological Nomenclature; ICZN, 1999). Methods relying on sequence (dis-)similarity thresholds are not suitable to delimit taxa because they merely represent arbitrary values not applicable to diagnoses (Goldstein and DeSalle, 2011). Instead, character-based distinguishing features are required for taxon definitions in accordance with the international codes of nomenclature for metazoans and heterotrophic protists (ICZN, 1999), algae, fungi, and plants (ICNafp; Turland et al. 2018), and prokaryotes (ICNP; Parker et al. 2019).

The columns of a gene sequence alignment represent putatively homologous characters (Dewey and Pachter, 2006) whose discrete states (i.e., the nucleotides A, C, G, or T) are suitable for differentiating one taxon from its relatives and thereby provide signature characters that can be included in taxon diagnoses. Signature characters represent the historical phylogenetic signal that is not the result of chance similarities, but support monophyletic groups in the corresponding phylogenetic tree (Wägele and Mayer, 2007). Early molecular work on prokaryotes compiled 16S rRNA gene signatures suitable for distinguishing the now well-established Bacteria and Archaea domains, as well as various lower-rank bacterial taxa (Woese, 1987). However, the application of discrete signature characters for distinguishing prokaryotes has not become popular and the current recommendations for improving

* Corresponding authors at: Department of Environment & Biodiversity, Paris Lodron University of Salzburg, Hellbrunnerstraße 34, 5020 Salzburg, Austria.
E-mail addresses: maximilian.ganser@plus.ac.at (M.H. Ganser), sabine.agatha@plus.ac.at (S. Agatha).

https://doi.org/10.1016/j.ympev.2022.107433
Received 30 November 2021; Received in revised form 13 January 2022; Accepted 20 January 2022
Available online 4 February 2022

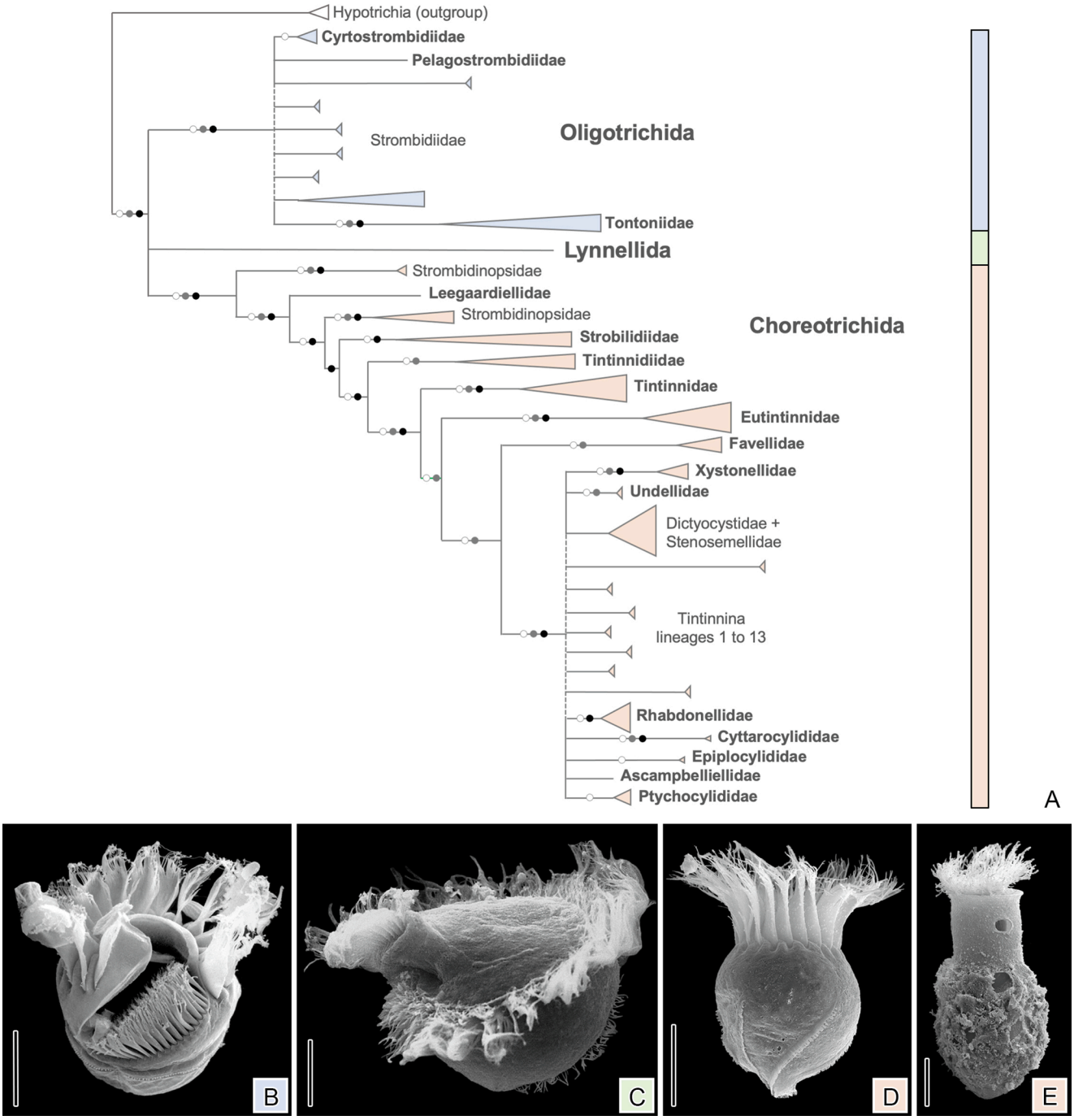

Fig. 1. Oligotrichea (Alveolata, Ciliophora). **(A)** Schematic representation of phylogenetic relationships as inferred from rRNA gene sequences (see the full trees in Supplementary Data S1). Only nodes consistently recovered with high support are shown (white, grey, and black dots indicate an IQTree ultrafast bootstrap support equal to or higher than 95% in the 18S rRNA gene, ITS region, and 28S rRNA gene trees, respectively). The monophyletic orders and families tested for signature characters are shown in bold. Simplified representations are included for the non-monophyletic Strombidiidae and the *incertae sedis* Tintinnina lineages (for details, see Santoferrara and McManus, 2021; Santoferrara et al., 2017). **(B-E)** Representatives of the main clades of oligotrichean ciliates in the scanning electron microscope. The taxa are characterised by an apical adoral zone of membranelles used for locomotion and filter feeding. (B) Oligotrichida: oblique top view of *Strombidium capitatum*. (C) Lynnellida: oblique top view of *Lynnella* sp. (D) Aloricate Choreotrichida: lateral view of *Strobilidium caudatum*. (E) Loricate Choreotrichida (Tintinnina): lateral view of *Codonellopsis schabi*. Its lorica is composed of a hyaline collar with a window and an agglutinated bowl. Scale bars: 20 μm.

prokaryotic taxonomy and nomenclature do not consider them (Hugenholtz et al., 2021). In eukaryotes, the identification of potential signature characters in taxonomic studies are rare and inconsistent. In some studies, signature characters were added only descriptively, while in others they were included in diagnoses. For instance, signature characters and their positions in the respective nuclear and/or plastid-encoded gene sequence alignments were visually identified in ciliates (Agatha and Strüder-Kypke, 2012; Hirt et al., 1995; Sun et al., 2012; Wang et al., 2020; Zhang et al., 2014). In other studies, the signature characters were included in diagnoses of, e.g., unicellular chlorophytes at different hierarchical ranks (Marin and Melkonian, 2010), euglenids (Marin et al., 2003), fungi (Rosling et al., 2011), and metazoans (Churchill et al., 2014; Delić et al., 2017; Grosse et al., 2021; Johnson et al., 2015; Jörger and Schrödl, 2013; Parapar et al., 2020; Satler et al.,

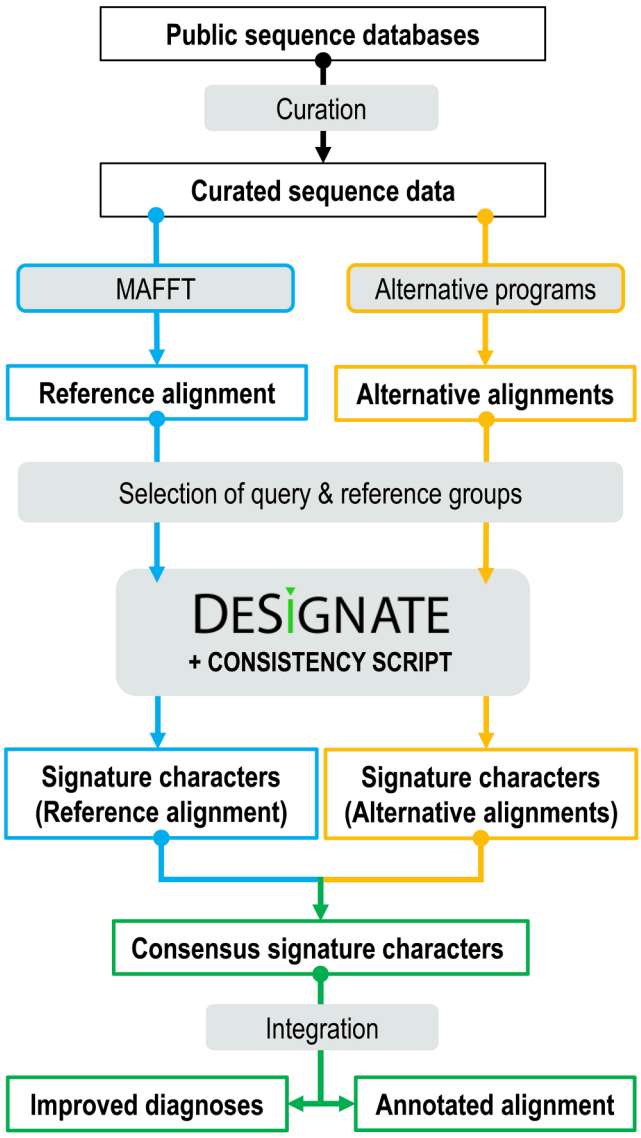

Fig. 2. Workflow displaying the main processes and tools (light grey) used to obtain consensus signature characters for improving taxonomic diagnoses (for details see text).

2013; Teixeira et al., 2020; Wang et al., 2016; Zielske and Haase, 2015). While a discussion of the plesiomorphic and apomorphic states is a common procedure for morphological characters, only few papers have contrasted the signature character states of gene sequences (e.g., Marin and Melkonian, 2010; Marin et al., 2003).

So far, signature characters included in taxonomic diagnoses were mostly detected by the program CAOS (Sarkar et al., 2008), by means of the maximum parsimony function with a labelled tree reconstruction showing the apomorphies (Marin and Melkonian, 2010; Marin et al., 2003), or by using the Diagnostic Character tool implemented in BOLD Systems V4 (http://www.barcodinglife.org; Ratnasingham and Hebert, 2007). However, the systematic addition of signature characters to formal, mostly morphology-based taxonomic diagnoses has been hampered by the lack of clear definitions, practical guidelines, as well as efficient and user-friendly tools. Owing to the pressing need, several tools (DeSignate, QUIDDICH, and FastaChar) have recently and almost simultaneously been developed (Hütter et al., 2020; Kühn and Haase, 2020; Merckelbach and Borges, 2020), but their broad utility remains to be tested.

The quality of the alignments is crucial. Each signature character (alignment column) should ideally comprise homologous nucleotides, i.e., the nucleotides may differ in their states but all of them initially derived from a single ancestral nucleotide (Dewey and Pachter, 2006). However, as it is impossible to retroactively observe the historical processes that led to the extant diversity of sequences, every alignment can only be an approximation containing information (historical signal) that might represent actual evolutionary relationships inferred from phylogenetic trees (Morrison, 2009).

Several programs exist that efficiently automate the process of aligning sequences by employing algorithms mainly based on the methodology of optimal path finding (Needleman and Wunsch 1970) to determine the similarities between two sequences. Therefore, most alignment algorithms are optimised for sequence similarity, which covers one important aspect of homology (Morrison, 2009). Although sequence similarity is a shared objective of these algorithms, various additional criteria and strategies implemented in the most popular programs lead to diverging results that also depend on the properties of each dataset (Chatzou et al., 2016; Pais et al., 2014). Accordingly, only those signature characters consistently detected in different alignments confirm various homology hypotheses.

In the present paper, we suggest a practical guideline for the integration of signature character data into taxonomic diagnoses. The critical steps are exemplified for an ecologically important clade of planktonic ciliates, the Oligotrichea (Alveolata, Ciliophora, Spirotricha), namely, (i) the curation of sequence data for the 18S and 28S rRNA genes and the ITS region, (ii) the identification of signature characters with the user-friendly DeSignate tool (Hütter et al., 2020), (iii) the evaluation of signature character consistency by means of a novel comparative approach, and (iv) the formal inclusion of signature characters into taxonomic diagnoses.

The Oligotrichea are among the most diverse and abundant ciliates in the sea. These heterotrophic or mixotrophic protists provide a comparatively high number of distinguishing features concerning morphology, ultrastructure, cell division patterns, and life cycle stages for their classification. Compared to other ciliates, for the Oligotrichea, a large body of morphology-based taxonomic literature and a considerable number of rRNA gene sequences in publicly accessible databases exist (reviewed by Agatha and Strüder-Kypke, 2014; Santoferrara et al., 2017). According to the current systematics, the Oligotrichea consist of the orders Oligotrichida (oligotrichids) and Choreotrichida (choreotrichids), the latter including the monophyletic shell-forming (loricate) tintinnids besides aloricate taxa; an additional proposed order, the Lynnellida, has an uncertain position in both gene trees and cladistic analyses (Fig. 1; Küppers et al., 2019; Liu et al., 2015; Santoferrara et al., 2017). A global comparison of 18S sequences from described morphospecies and environmental surveys has indicated a vast majority of molecularly uncharacterised or novel aloricate lineages (except for Lynnellida, which are monotypic and present consistently low diversity in environmental data), while most tintinnid sequences could be linked to described taxa (Santoferrara et al., 2017).

The oligotrichid families Tontoniidae, Cyrtostrombidiidae, and Pelagostrombidiidae are monophyletic in cladograms and phylogenies, while the fourth family, the Strombidiidae, is paraphyletic. The aloricate choreotrichids do not form a monophyletic group, and neither does the family Strombidinopsidae. The Leegaardiellidae are currently represented by a single sequence, and the Strobilidiidae appear as monophyletic in molecular phylogenies. The molecular phylogeny of the tintinnids displays a consecutive branching of the Tintinnidiidae, Tintinnidae, Eutintinnidae, and Favellidae. The remaining tintinnids form several statistically well-supported clades whose relationships are unresolved in gene trees. While many uncertainties remain in oligotrichean systematics, the proposed orders and 16 of their families are consistently supported by morphological characters and phylogenetic trees, and we thus use them (along with the oligotrichid genus *Novistrombidium* and the tintinnid genera *Antetintinnidium* and *Dartintinnus*) as models for the detailed evaluation of signature characters and their integration in taxonomic diagnoses.

2. Materials and methods

2.1. Curated sequence data

The workflow for detection of signature characters begins with the curation of the sequence data to ensure that only reliable information is considered (Fig. 2). For the Oligotrichea, the curated datasets established by Santoferrara et al. (2017) were revised and completed with 37 sequences from GenBank (last updated in January 2021). The datasets comprise the sequences of the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the D1-D2 regions of the 28S rRNA gene (in the following referred to as 18S, ITS region, and 28S, respectively), which fulfilled certain criteria. Gene sequences obtained from specimens that could reliably be identified to species or genus rank based on morphological data in the corresponding publications were assigned to categories 1 and 2, respectively. Gene sequences of low quality (ambiguous/missing nucleotides), insufficient length (<50% of average sequence length in the alignment), or missing published data to prove the identity of the sequenced specimen (e.g., micrographs and/or measurements) were considered unreliable and were excluded, as they hamper the detection of reliable signature characters (see 2.3. and 2.4.).

The updated and curated datasets comprise 398 oligotrichean sequences and four hypotrich outgroup sequences (Supplementary Table S1A). The datasets cover 158 out of about 1,200 described oligotrichean species, 53 out of 105 genera, 20 out of 22 families, and all three orders (Supplementary Table S1B). Each species is represented by a single sequence in at least one dataset resulting in 201, 103, and 94 sequences for the 18S, ITS region, and 28S, respectively.

2.2. Reference alignments and phylogenetic trees

The curated datasets for the 18S, ITS region, and 28S sequences were aligned on the MAFFT v. 7 web server (Katoh and Standley, 2013), using the "*globalpair*" (G-INS-i alignment method) and "*maxiterate 1000*" commands, which resulted in three reference sequence alignments (Supplementary Data S1; Fig. 2). For signature character detection (see below), untrimmed alignments were used, except for the 28S dataset, which was trimmed after site 796 (pre-alignment position; subsequent nucleotides available for only 30% of the sequences were disregarded). For phylogenetic tree inference, the flanking regions of each alignment were trimmed based on the results of Gblocks v. 0.91b under default settings (Castresana, 2000). The IQTree webserver (Trifinopoulos et al., 2016) was used to infer phylogenetic trees for each of the three alignments (Supplementary Data S1). The gene trees were inferred with the substitution model determined by Modelfinder under the Akaike information criterion (Kalyaanamoorthy et al., 2017): GTR + F + I + G4 for 18S and TIM2e + I + G4 for ITS and 28S. Branch support was calculated with 1,000 iterations of the ultrafast bootstrap approximation (Minh et al., 2013) and the SH-like approximate likelihood-ratio test with 1,000 replicates (Guindon et al., 2010).

2.3. Terminology

Signature character (diagnostic molecular character) as defined by Hütter et al. (2020): A single column in a DNA sequence alignment comprising a diagnostic character state (in the following diagnostic state) that is uniform in one taxon (query group) but differs from the character state of one or several other taxa (reference group). Query and reference groups should represent taxa of the same taxonomic rank, as done for morphological characters. The numbering (position in the respective alignment) of congruent signature characters may differ between alignments. While the diagnostic state in the query group is defined to be uniform, several character states can occur in the reference group. Therefore, two types of signature characters are distinguished:

(1) *Binary signatures* comprise only two different character states, i.e., one in the query group (e.g., nucleotide base A) and a different one in the reference group (e.g., nucleotide base T);
(2) *Asymmetric signatures* comprise at least two character states in the reference group (e.g., nucleotide bases C, G, or T) that are different from the uniform character state in the query group (e.g., nucleotide base A).

2.4. Signature character analysis

Taxonomic groups of different systematic ranks (orders, families, and genera) within the Oligotrichea that are monophyletic in both cladistic analyses of morphological characters and 18S-based phylogenies (Fig. 1) were selected as query groups to ensure that there is a clear link between sequence data and the morphological description/diagnosis. Thereby, multiple lines of evidence are used to assess the stability of each taxon and prevent premature diagnoses based solely on putative signature characters. Known inconsistencies in ITS- and 28S-based phylogenies, likely due to the lower phylogenetic resolution of these markers, were disregarded (Santoferrara et al., 2017).

The signature characters and their diagnostic states for each query group were identified by means of the DeSignate web server (Hütter et al., 2020) and the untrimmed 18S, ITS, and 28S reference alignments (Fig. 2). DeSignate can be freely accessed through an intuitive web application that guides the user through the analysis process and displays all results at a glance. For the analyses, deletions (gaps) were considered and the query and reference groups were selected via the phylogenetic tree option or comma-separated lists (file input and group selection). The results for each group comparison and reference alignment, i.e., the entropy and diagnostic relevance ranking tables, were saved for downstream analyses.

The Shannon entropy is calculated by DeSignate either for discrete alignment positions or as moving average (k-window = 21, i.e., the ten alignment positions preceding the signature character position and the ten subsequent ones), considering the relative frequency of the different character states in each alignment column. High entropy values indicate positions with a high variability of character states (e.g., in the V4 or V9 regions) and low values indicate positions with conserved character states (Hadziavdic et al., 2014).

The consistency of signature characters was assessed by testing the influence of different alignment algorithms, i.e., which binary and asymmetric signatures were congruently detected by DeSignate in the reference and two alternative alignments of the same gene or region (=consensus signature characters). To this end, reference sequences were additionally aligned with MUSCLE v.3.8.31 (Edgar, 2004), T-COFFEE (EBI web server; Notredame et al., 2000), and CLUSTALO (Sievers et al., 2011), using default settings. Subsequent analyses were done with the MAFFT and MUSCLE alignments for all the markers, T-COFFEE for the 18S, and CLUSTALO for the 28S and ITS region (T-COFFEE generated better 18S alignments than CLUSTALO, while the opposite was true for the 28S and ITS region). Consensus signature characters were automatically determined with a custom python script (https://github.com/maxganser/consistency-script) which compares the signature characters detected by DeSignate for a selected pair of query and reference groups between the reference and alternative alignments. Such a comparison necessitates tracking each individual nucleotide and its position in each alignment, which is achieved by annotating each nucleotide of the aligned sequences with its position from the unaligned (gapless) sequences. In the generated nucleotide position matrix, the nucleotide bases are replaced by their positional numbers derived from the unaligned sequences and gaps are replaced by the number zero.

To test whether discrete and average entropy values might enable assessing the consistency of signature characters directly in the reference alignments, statistically significant differences between discrete and moving average entropy values for consensus and non-consensus

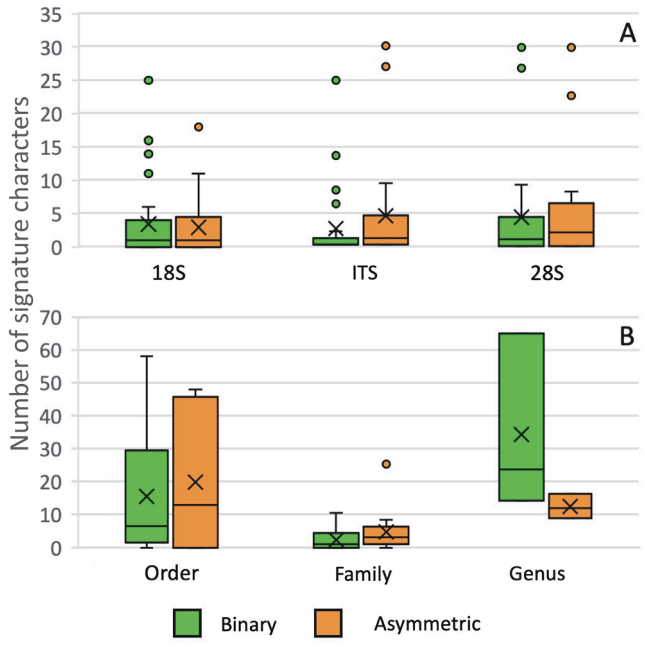

Fig. 3. Variability in the numbers of consensus signature characters. (A) Variation within each alignment (N = 25, 23, and 22 taxa tested for 18S, ITS, and 28S, respectively). (B) Variation within each systematic rank (N = 6, 16, and 3 taxa tested at the order/suborder, family, and genus ranks, respectively).

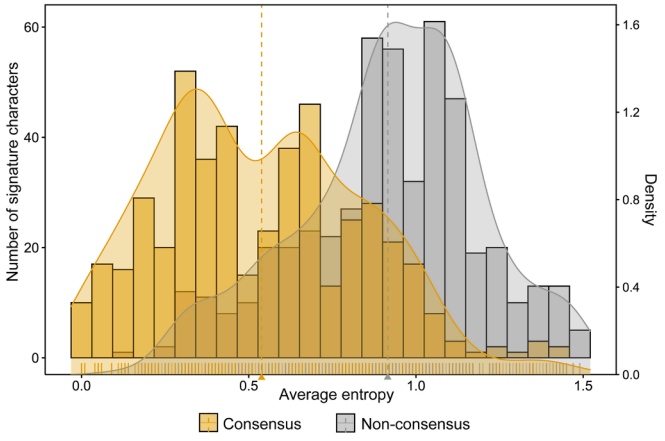

Fig. 4. Histograms and density plots of average entropy values for consensus (ochre) and non-consensus (grey) signature characters. Mean values (dashed lines and arrowheads on the x-axis) show a significant difference ($p < 2e\text{-}16$) indicating that average entropy values tend to be lower for consensus compared to non-consensus signature characters.

signature characters were determined with the unpaired Wilcoxon rank sum test, using R v. 3.6.2 (R Core Team, 2019).

2.5. Integration of consensus signature characters in diagnoses

The diagnoses of the query groups were complemented with consensus signature characters according to the following criteria: (i) binary consensus signatures were detected, (ii) the query group comprised at least two curated sequences (except for the monotypic order Lynnellida), and (iii) the query group was shown to be monophyletic by various phylogenetic analyses. Within the formal diagnoses, we mentioned the gene sequence or region analysed plus the character position in the reference alignment followed by the binary or asymmetric character state of the query group in brackets.

To facilitate retrieving the signature characters, the diagnostic states were annotated by means of the freely available Jalview v. 2.11.1.4 software in each reference alignment with the MAFFT-aligned 18S, ITS, and 28S sequences (Fig. 2; http://www.jalview.org; Waterhouse et al., 2009). The reference alignments provide the GenBank accession numbers followed by the curated species names for each sequence. The signature character type (binary or asymmetric) and the query and reference groups were specified. The annotation files (Supplementary Data S1) were saved in Jalview format that can be loaded for the respective alignment files to display every consensus signature character and its states.

2.6. Secondary structure visualisation

Secondary structures of ribosomal RNA contain valuable information for phylogenetic and taxonomic analyses (e.g., Abraham et al., 2021; Nickrent and Sargent, 1991; Tasneem et al., 2020; Wang et al., 2015). Therefore, the secondary structures of the 18S rRNA from five representative taxa of different taxonomic ranks (the lynnellid *Lynnella semiglobulosa*, the oligotrichid *Spirotontonia*, the aloricate choreotrichid *Leegaardiella*, and the tintinnids *Antetintinnidium mucicola* and *Dartintinnus alderae*) were visualised with the R2DT web server (https://rnacentral.org/r2dt). R2DT predicts and visualises the secondary structures in a standard layout by using a template of known structure to facilitate comparability (Sweeney et al., 2021). Here, we chose the template of *Saccharomyces cerevisiae* (SC_SSU_3D) based on recent models incorporating 3D structural information (Petrov et al., 2014). The template was visualised with RiboVision v. 1.15 (http://apollo.chemistry.gatech.edu/RiboVision/) to depict the secondary structure of the complete 18S and its hyper-variable V4 region, the latter being a suitable marker for metabarcoding eukaryotes and, especially, ciliates (Dunthorn et al., 2012). The naming of the segments, i.e., of the conserved and expansion segments in the secondary structures, follows Gerbi (1996). Binary and asymmetric signatures were tagged in the secondary structure diagrams of the V4 region for the abovementioned five taxa and in the complete 18S diagram of *L. semiglobulosa*.

3. Results

The critical steps in the completion of taxon diagnoses by adding signature characters (Fig. 2) are exemplified for an ecologically important clade of planktonic ciliates, the Oligotrichea. First, the consensus signature characters are identified for three datasets (18S, ITS region, and 28S). Whether entropy values are good indicators for the consistent discovery of signature characters in each alignment is subsequently analysed. Furthermore, potential distribution patterns in the alignments are scrutinised, and the positions of signature characters in the secondary structure of the 18S are particularly examined for representative taxa in more detail. Finally, the consensus signature characters emerging in all alignments composed of curated sequences are included into the taxon diagnoses.

3.1. Consensus signature characters

The comparative analyses of 18S, ITS region, and 28S alignments yielded a total of 468 consensus signature characters distributed among almost every selected query group (three orders, 13 out of 16 families, and three genera; Supplementary Table S2). The highest number of binary and asymmetric signatures was found in the 28S alignment (194 signatures among 821 positions) followed by the 18S alignment (160 signatures among 1,830 positions), while the ITS alignment (114 signatures among 584 positions) had the lowest number of total binary and asymmetric signatures (Fig. 3A).

Regarding systematic ranks, the numbers of binary and asymmetric consensus signatures show a higher variability at the order and genus ranks versus family rank (Fig. 3B). At order rank, comparisons of the

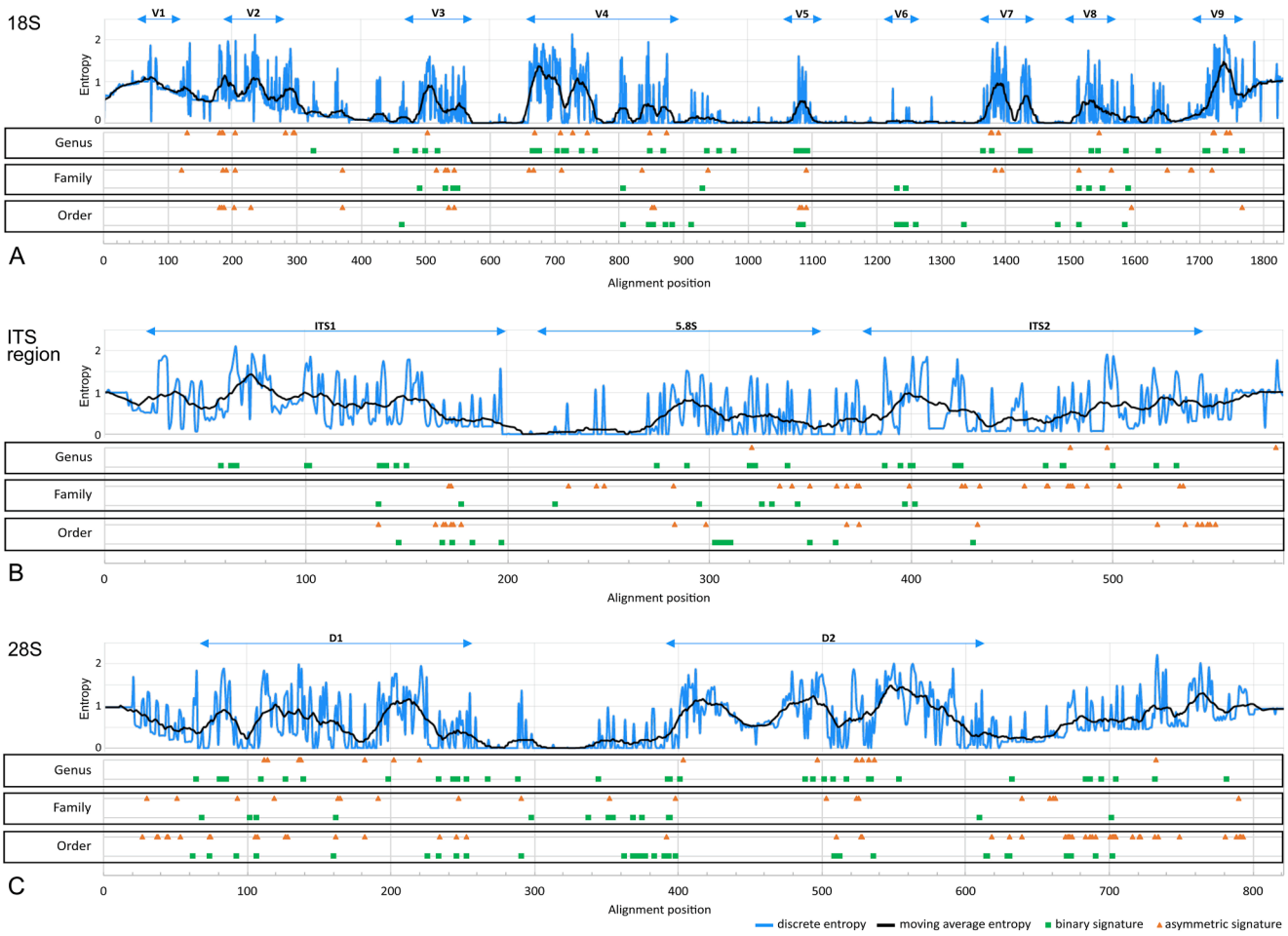

Fig. 5. Distribution of consensus signature characters in the 18S, ITS region, and 28S alignments. Binary (green squares) and asymmetric (orange triangles) signatures for each rank (order, family, and genus) are shown in separate lines. The entropy plot (top) indicates variable positions by peaks; the discrete entropy is denoted by a blue line and the moving average (each positional value is calculated from the average of 21 positions, including the 10 preceding and the 10 subsequent positions) is denoted by a black line. Sequence-specific variable regions in the alignments are indicated by double line arrows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Oligotrichida and Choreotrichida yielded up to eight signatures, whereas the Lynnellida (only comprising the monotypic genus *Lynnella)* yielded up to 29 binary and 29 asymmetric signatures when compared to either or both other orders (Lynnellida vs. Oligotrichida or Choreotrichida, Lynnellida vs. Oligotrichida and Choreotrichida). The oligotrichean families had from 1 to 7 binary and from 1 to 8 asymmetric signatures, except for the family Leegaardiellidae that is characterised by 10 binary and 24 asymmetric signatures. At genus rank, more binary (up to 39) than asymmetric (up to 24) signatures were found in total for the three investigated taxa.

In addition to the 468 consensus signature characters (see above), 490 non-consensus characters were detected. The ratio of consensus to non-consensus signature characters was higher for binary (233 vs. 125) than for asymmetric (234 vs. 371) signatures. Gaps were more frequent in non-consensus signature characters than in consensus signature characters for the query group (86 vs. 3; Supplementary Table S4).

The entropy values of consensus and non-consensus signature characters were compared for each systematic rank in the reference alignments (Supplementary Fig. S1, Supplementary Table S3). The discrete entropy values are highly variable for consensus (CV = 25%) and non-consensus (CV = 22%) signature characters and their ranges distinctly overlap, whereas average entropy values are less variable (CV = 9% and 7%, respectively). Overall, the average entropy values (Fig. 4) of consensus signature characters are significantly lower ($\bar{x}$ = 0.54 ± 0.30) compared to the values of non-consensus signature characters ($\bar{x}$ = 0.91 ± 0.27) for binary ($p < 2e\text{-}16$) and asymmetric ($p < 2e\text{-}16$) signatures.

Most significantly different average entropy values between consensus and non-consensus signature characters were detected in the 18S and 28S datasets (Supplementary Table S5). The ITS average entropy values are only significantly different at order rank for binary signatures and at order and family ranks for asymmetric signatures. In summary, the consistency of signature characters (consensus = consistent; non-consensus = not consistent) is best reflected by the average entropy values in the 18S and 28S and to a lesser extend in the ITS. This pattern seems to be related to their phylogenetic information content and resolution: the 18S is useful for relationships at higher ranks, while the 28S better resolves relationships at genus rank in Oligotrichea; the ITS is highly variable and does not reliably resolve relationships above genus rank.

3.2. *Distribution of consensus signature characters: Sequences and secondary structures*

No distinct patterns emerged regarding the distribution of consensus signature characters for each systematic rank (order, family, and genus), viz., they are scattered over almost all variable regions in the reference alignments (Fig. 5). In the 18S reference alignment, binary signatures were primarily located between the variable V3 and V9 regions, whereas asymmetric signatures were also present in the first portion of the

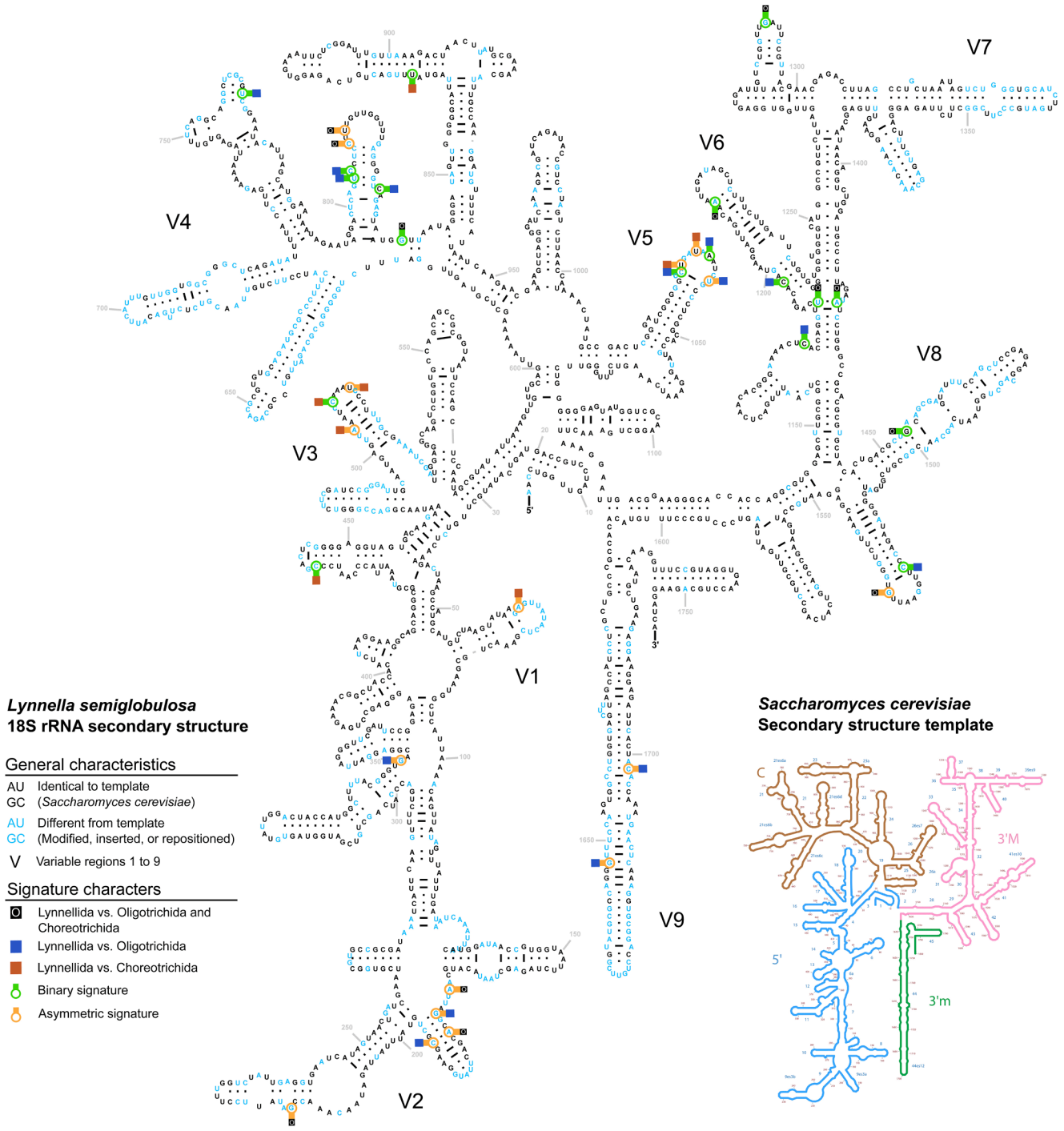

Fig. 6. An 18S rRNA secondary structure diagram of *Lynnella semiglobulosa* displaying the consensus signature characters detected for each comparison at order rank (Lynnellida vs. the Oligotrichida and Choreotrichida, vs. the Oligotrichida, and vs. the Choreotrichida). For further explanations, see legend. The schematic representation of the secondary structure template (lower right corner) is coloured according to the four domains (Central, 5 Prime, 3′minor, and 3′Major). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

alignment including the V2 region. Both types of signature characters mainly accumulated in the variable regions and only a few were identified in highly conserved portions. The first segment of the hypervariable V4 region contained exclusively binary signatures for the genus rank, while the second segment comprised signature characters for each investigated rank. In the ITS region, binary signatures were detected mainly at genus rank in the initial portion of the ITS1 and most of the ITS2. In the 28S, a comparatively high number of binary and asymmetric signatures for all taxonomic ranks was distributed quite homogeneously.

Alignment positions that yielded consensus signature characters for up to three different query groups (Supplementary Table S6) were identified in the reference alignments of the 18S (13 positions), the ITS (15 positions), and the 28S (23 positions). The query groups mainly contained taxa for which a high amount of binary and asymmetric signatures was detected (e.g., the family Leegaardiellidae, the order Lynnellida, as well as several tintinnid families and two genera).

To assess how consensus signature characters coincide with structural features of the 18S rRNA, their positions in the secondary structure diagrams were investigated (Fig. 6, Supplementary Fig. S2). Binary and

asymmetric signatures in the complete 18S secondary structure of *Lynnella semiglobulosa* representing the Lynnellida were compared with the Oligotrichida and the Choreotrichida separately and with both taxa simultaneously (Fig. 6). Both signature types are generally scattered across all domains (Central, 5 Prime, 3′minor, and 3′Major), but show accumulations of three to five signatures in the terminal segments of the V2, V3, V4, and V5 regions. Nearly twice the number of signature characters differentiate *Lynnella* from the Oligotrichida compared to the Choreotrichida (15 vs. 8). The V2 and V4 regions exclusively comprise signature characters that distinguish *Lynnella* from both taxa simultaneously or only the Oligotrichida, whereas signature characters in the V1 and V3 regions only concern the Choreotrichida. In addition, the terminal loop of the V5 region contains five adjacent signature characters discriminating *Lynnella* from either the Oligotrichida or the Choreotrichida. Finally, most signature characters (25 out of 35) were detected for nucleotides that differ from the ones in the secondary structure of the template (*Saccharomyces cerevisiae*) because of substitution events, insertions, or slight structural deviations.

In the V4 region, most order-, family-, and genus-specific consensus signature characters are found in expansion segment 6d (Supplementary Fig. S2), which is about 40 nucleotide bases long and forms a short stem structure with an internal and a hairpin loop. Only genus-specific consensus signature characters are present in expansion segments 6b and 6c. The two binary and two asymmetric signatures detected for the monotypic tintinnid genus *Dartintinnus* in these segments are compensatory base changes. Furthermore, one order-, two family-, and two genus-specific consensus signature characters are located in the conserved segments 22 and 23.

3.3. Improved diagnoses

Exhaustive evaluation of molecular signature characters has allowed us to generate a conservative approach (see criteria in 2.5.), by which such signatures are incorporated into the diagnoses of taxa that are supported by additional lines of evidence, especially morphology in the case of Oligotrichea. Owing to the unsettled oligotrichean systematics, many taxa could not be tested as they are not monophyletic or lack reliable sequences. In the majority of taxa we tested, however, consensus signature characters were detected (Supplementary Table S2). Thus, we could add consensus signature characters to the diagnoses of three orders, eight families, and two genera belonging to the Oligotrichea (Appendix A). Within the formal diagnoses, we did not differentiate between binary and asymmetric signatures of the query group.

4. Discussion

In the present study, we propose a formal and practical approach to integrate routinely molecular signature characters into diagnoses of taxa that are consistently inferred as monophyletic based on multiple data sources (e.g., morphology, ultrastructure, gene sequences). We provide recommendations on the choice of curated DNA sequence data, alignment methods, and query groups, as well as on the selection of the most consistent signature characters based on a novel comparative approach employing an automated script. Our strategy represents the first reliability check for signature characters that can be efficiently applied on the results of the recently developed bioinformatic tool DeSignate (Hütter et al., 2020). Considering our suggested guidelines, the utility of molecular signature characters should not replace traditional morphological characters, but highlight the benefits for integrative taxonomy (Dayrat, 2005; Santoferrara et al., 2016; Schlick-Steiner et al., 2010; Warren et al., 2017; Will et al., 2005).

4.1. Nomenclature in the era of integrative taxonomy

According to the International Code for Zoological Nomenclature (ICZN, 1999), every new name published must currently be accompanied by a description or definition (diagnosis) that states in words those characters that are purported to differentiate the taxon from related or similar taxa, by a bibliographic reference, or by an indication of a replacement name. Taxonomists are responsible for judging which of the observed features of the organisms qualify or not as taxonomic characters. Diagnoses have to include discrete homologous characters with different states to make taxon names available (Dubois 2017). Characters may be morphological, anatomical, genetic, biochemical, physiological, and behavioural, for instance. Traditionally, comparisons of similar or related taxa yielded distinguishing features, which are included together with their character states in taxon diagnoses. Diagnoses are usually generalisations deduced from the observation of several organisms and are monothetic, viz., they represent a unique combination of character states that are both necessary and sufficient for membership. Thereby, a diagnosis allows an unequivocal identification irrespective of any cladistic hypothesis, namely, assumptions about apomorphic and plesiomorphic character states (Dubois, 2017). In this work, we promote the inclusion of non-morphological features (signature characters of gene sequences) in diagnoses. We disagree with arguments that such an approach contradicts the very purpose of a diagnosis in providing distinguishing features allowing the taxon's identification (Hassemer et al. 2020). Actually, our approach facilitates a partial automation of the taxonomic workflow, while the comparability with historical data and knowledge is maintained (Gemeinholzer et al., 2020).

4.2. Sequence datasets and alignments

The prerequisite for a reliable detection of molecular signature characters is a trustworthy sequence dataset. DNA sequences should originate from well-documented specimens (e.g., by drawings, photos, videos, morphometric data) for confirming taxonomic affiliations. Existing publicly accessible databases (e.g., NCBI GenBank) do not provide the sufficient level of curation for obtaining reliable signature characters by simply using deposited sequences without prior revision, and thus additional taxon-specific expertise is required. Furthermore, if multiple markers are analysed, the sequences should ideally derive from a single individual to avoid mixing data from cryptic species. Sequences that include ambiguous nucleotides or are shorter than 50% of average sequence length yield potentially incorrect signature characters or hamper the detection of signatures.

To compute the reference alignments with our curated sequences, we chose the program MAFFT (Katoh and Standley, 2013). In comparison to other programs, MAFFT performs better in regions that are difficult to align and potentially contain phylogenetic information useful for tree inference (Morrison et al., 2015). Furthermore, MAFFT is among the most efficient and accurate alignment programs for various datasets (Pais et al., 2014). However, the benchmarks used to determine the best performance and accuracy for phylogenetic analyses mainly rely on simulated datasets created with evolutionary models that may not adequately reflect true evolutionary processes (Chatzou et al., 2016). Thus, these benchmarks may help to assess the overall accuracy in the frame of the applied models, but still do not provide the means to decide on the reliability of each alignment column (character). Instead, we suggest comparing alignments representing alternative hypotheses of homology computed with several different programs (e.g., MUSCLE, T-COFFEE, or CLUSTALO). Ultimately, which alignments are retained for molecular signature character analyses must be evaluated for each sequence dataset.

4.3. Detection and evaluation of signature characters

By comparing a query and a reference group in a multiple sequence alignment, DeSignate can detect signature characters and thus provide discrete distinguishing features as required by nomenclatural codes (Hütter et al., 2020).

A-posteriori-procedures to find signature characters that represent molecular apomorphies for a clade/taxon are not much different from the generally accepted cladistic strategy of interpreting morphological characters (Marin et al., 2003). In our novel approach, a script automatically compares the signature characters identified by DeSignate in alternative homology hypotheses expressed by different alignments. Consensus signature characters congruently detected in all selected alignments indicate that the optimal solution for aligning the respective nucleotides into columns is identical amongst the different programs. Thus, consensus signature characters constitute hypotheses about homologous nucleotides that are repeatedly supported. For non-consensus signature characters, putative homology cannot be inferred by the algorithms only considering compositional similarity. Other criteria that further improve alignments still need to be integrated into current programs, for example, pattern matching and modelling of known molecular mechanisms causing sequence variation (Morrison et al., 2015).

Average entropy values most reliably indicated the consistency of signature characters as they incorporate the variability of adjacent alignment columns. For example, signature characters with the lowest average entropy values are usually located between or in proximity to rather conserved characters, i.e., alignment columns with one or only a few differing nucleotides (character states). Therefore, only a single most parsimonious solution exists to align the nucleotides into the respective column. However, a clear distinction between consensus and non-consensus characters merely based on average entropy values is not always possible because of outliers that especially concern signature characters in alignments of sequences that are typically more divergent and problematic to align (e.g., the ITS region). Other means for quality assessment of alignments are based on reliability scoring metrics (e.g., Chang et al., 2014; Landan and Graur, 2007; Penn et al., 2010), masking ambiguously aligned positions (e.g., Capella-Gutiérrez et al., 2009; Castresana, 2000), or probabilistic methods (e.g., Ali et al., 2019; Ezawa, 2016). However, these metrics are usually intended to improve subsequent phylogenetic tree inferences and are not suitable to incorporate results from several alignment programs. Thus, they are insufficient for identifying reliable alignment positions that represent consensus signature characters.

The secondary structure of ribosomal RNA and associated internal spacer regions can be highly informative for alignments and the reliability of signature characters, especially as models have become more accurate by including 3D structural information (Petrov et al., 2014). Conserved and variable alignment segments of the 18S, 5.8S, and 28S rRNA are not randomly distributed because of structural constraints that retain ribosomal functions. For example, mutation events that affect nucleotides in paired regions (stems, helices) are compensated to restore structural stability (Gutell et al., 1994). These so-called compensatory base changes have a diagnostic value if they are based on robust secondary structure models to assess and compare homologous nucleotide base pairs between different taxa (e.g., Caisová et al., 2011; Ruhl et al., 2010). We detected only a few (about 7%) consensus signature characters that coincided with complementary base pairs in the inspected 18S/V4 diagrams of selected taxa (Fig. 6, Supplementary Figure S2).

Most signature characters involve nucleotides that vary compared to the template. The variation present in each evolutionary lineage might be the result of, for example, functional adaptations (Doris et al., 2015) or might not have any influence at all if no interaction with other components of the ribosome is affected (Taylor et al., 2009). However, a general pattern is not evident regarding the distribution of consensus signature characters in respect to the rather conserved secondary structure of the 18S rRNA. Therefore, further datasets need to be explored to verify a potential linkage between signature characters and secondary structures.

4.4. Complementing diagnoses with signature characters

Improving diagnoses is a regular and necessary process due to the continuous incorporation of new knowledge obtained, for example, by the application of new methods. Likewise, the splitting of taxa or the addition of new taxon members (further populations, new species, genera, etc.) may require narrowing or expanding diagnoses, respectively, and might further lead to new hypotheses on evolutionary relationships.

Since molecular signature characters are as inclusive as morphological characters, each member of a higher taxon (species of a genus, genus of a family, etc.) shows specific signatures besides those of the higher taxon. Hence, taxon sampling may influence the number of detected signature characters. For example, when only a single sequence representing a higher taxon is analysed (e.g., in the monotypic genus *Lynnella*), the number of signature characters is probably overestimated and may need revision when further members of the lineage are discovered.

Like in morphology-based diagnoses, intraspecific variability might affect the detection of molecular signature characters. However, the most commonly obtained sequences are those that represent the dominating haplotype for a given taxon (Wang et al., 2019; Zhao et al., 2019). Hence, less frequent haplotypes are usually not represented in databases, and it is thus unlikely that they will affect taxon diagnoses at or above genus rank as in our case study. The prerequisite for a query group is its monophyly in cladistic and molecular analyses, representing a valid taxon that by definition comprises its type. Since all further taxon members share the morphological and signature characters of a taxon, it is not required to analyse sequences of the type prior to a complementation of the diagnosis.

The consensus signature characters included in a diagnosis should be accompanied by information allowing its reliable retrieval and future usage, namely, the gene or region analysed, their positions in the reference alignment, and their character states in the query group; similar to morphological characters, a discussion of the signature characters should describe the character state/s for the reference group. Therefore, we suggest standards for a routinely implementation.

- Consensus signature characters are added to taxon diagnoses as depicted in the following example:

  18S rRNA gene: 854 (C) [Gene: Character position in the reference alignment (binary or asymmetric character state of the query group)].

- Diagnostic character data availability: Consensus signature characters are annotated in the reference alignments using Jalview (the annotation file comprises the signature character type, query and reference group names, and the position of the nucleotide in the reference sequence).

4.5. Limits of molecular signature characters

Our approach promotes integrative taxonomy, while we oppose delineating taxa solely by molecular signature characters. Taxon diagnoses based on gene sequences or signature characters alone are not useful for many taxonomists and ecologists and are unsuitable for a stand-alone reference system (Ahrens et al., 2021). Since phylogenetic relationships are not facts, but hypotheses, the establishment of a new taxon exclusively inferred from its placement in a molecular tree, i.e., without the provision of discrete diagnostic characters, results in a *nomen nudum* (a name not fulfilling the requirements of the ICZN, 1999). Exceptions to the integrative approach apply mainly to higher ranks in the systematic hierarchy. Here, it becomes more and more difficult to ascertain synapomorphies in morphological or ultrastructural features. This is particularly true in protists, a paraphyletic assembly of biologically distinct taxa (i.e., all eukaryotes except for the animals, land plants,

and true fungi), many of which show a relatively simple, but sometimes difficult to study structure because of their microscopic sizes. For instance, the eukaryotic SAR-group comprising the Stramenopiles (e.g., diatoms, brown algae), Alveolates (e.g., ciliates, dinoflagellates), and Rhizaria (e.g., foraminiferans, radiolarians) so far lacks unifying morphological or ultrastructural features, and it is exclusively defined phylogenetically, thus having no taxon-status (Adl et al., 2019). Since molecular signature characters seem to be as inclusive as morphological characters, none or very few signatures might be detected at such high systematic ranks in the uniformly used ribosomal RNA gene sequences. Hence, new markers might be required, as shown by an approach in metazoans using nearly universal single-copy nuclear protein-coding genes (Eberle et al. 2020). Another application of purely genetic taxon discrimination are cryptic species, i.e., species that cannot be delineated by morphological features at the current state of knowledge. In our study case, we restrict to systematic ranks of genera or above, owing to the insufficient number of sequences for detecting signature characters in species.

4.6. Remarks on Oligotrichea

Following the recommendations compiled above, we used the Oligotrichea as a case study. We analysed only taxa characterised by morphological features that form a monophylum in 18S gene trees generated by different algorithms. Due to unsettled and incomplete gene trees and several non-monophylies, these criteria ensure reliability but also restricted our testable taxa drastically (Table A.1). Yet, our limited taxonomic sampling yielded a considerable number of consensus signature characters.

In these cases, our approach provided consensus signature characters that further corroborate the tree topology and can be implemented into the taxon diagnoses. The exceptions were the aloricate choreotrichid family Leegaardiellidae and the tintinnid genus *Dartintinnus*, for which only one identified individual per taxon has been sequenced so far, but environmental sequences indicate a much higher diversity of undescribed taxa (Santoferrara et al., 2017, 2018).

The enigmatic *Lynnella semiglobulosa* is the sole known species of its family and probably represents a new order that markedly differs from the Oligotrichida and Choreotrichida. This taxon somehow appears as a chimera between the Oligotrichida and Choreotrichida in respect to morphological and molecular features (see Appendix A). Depending on the alignment and tree-building algorithm, the species groups with the Oligotrichida or the Choreotrichida (Santoferrara et al., 2017). However, most morphological characters suggest a closer relationship to the Choreotrichida than the Oligotrichida (Adl et al., 2019; Agatha and Strüder-Kypke, 2014), which is supported by fewer signature characters distinguishing the Lynnellida from the former order (11 vs. 16 binary signatures in the 18S, none vs. 13 in the ITS region, and 10 vs. 29 in the 28S; Supplementary Table S2).

5. Outlook

In addition to taxonomic and nomenclatural utility, complementing taxon diagnoses with signature characters promises further potential applications linked with diverse aspects of biodiversity studies. The prerequisite for easy accessibility is the provision of the signature characters as metadata of curated reference sequences (e.g., see our reference alignment annotated with Jalview; Supplementary Data S1) in commonly employed databases. Biodiversity assessment of complex communities, for example, those of aquatic or soil protists, is now commonly done by metabarcoding, which has been based mostly on the short V4 or V9 regions of the 18S (Dunthorn et al., 2012; Santoferrara et al., 2020) and more recently on long-reads that comprise the 18S, ITS region, and 28S (Jamy et al., 2020). In addition to the multiple methods used for taxonomic annotation, signature characters might allow for a formal identification of a taxon based on its DNA sequence (order, family, or genus in our Oligotrichea example). As a result, annotations of environmental sequences can become more accurate even without expert knowledge on the particular taxonomic group. Conversely, sequences that do not match signature characters of closely related sequences represent potentially new taxa. Signature characters may also be employed in targeted retrieval approaches by employing FISH probes that enable subsequent morphological studies (Gimmler and Stoeck, 2015; Piwosz et al., 2021). For more in-depth analyses of communities, specific primers can be designed, using signature characters (Jung et al., 2012; Kohara et al., 2002). The search for signature characters may expand beyond rRNA markers and consider other nuclear genes, such as universal single-copy nuclear protein-coding genes (USCOs; Eberle et al., 2020). Our approach might be promising, especially in regard to the rapid identification of new sequences, given that the amount of processed signature character data is much smaller compared to that of hundreds of genes.

CRediT authorship contribution statement

Maximilian H. Ganser: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Luciana F. Santoferrara:** Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Sabine Agatha:** Conceptualization, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Thomas Hütter for improving the consistency script and Birgit Weißenbacher for annotating the reference alignments.

Funding

This research was supported by the Austrian Science Fund (FWF; grant number I 3268-B29) given to SA. LFS is supported by the U.S.A. National Science Foundation (Grant OCE-1924527).

Appendix A. Systematic overview and improved diagnoses of the Oligotrichea

Detailed systematic overview

The Oligotrichea comprise three orders: the Oligotrichida, the Choreotrichida, and the Lynnellida. These orders are monophyletic in both cladistic and genetic analyses (Fig. 1; Table A.1). However, the relationships among the orders are unresolved as the order Lynnellida with its single species shares morphologic and genetic features of the other two orders.

The Oligotrichida consist of four families: the Cyrtostrombidiidae, the Pelagostrombidiidae, the Strombidiidae, and the Tontoniidae. The families, except for the Strombidiidae, are characterised by apomorphies and are monophyletic in molecular genealogies. The Strombidiidae are, however, merely characterised by plesiomorphies and interspersed throughout the oligotrichids. Currently, the resolution of the

phylogenetic trees is insufficient regarding most genera; yet, it is highly unlikely that the genera *Strombidium, Spirostrombidium*, *Parallelostrombidium*, and *Propecingulum* will become monophyletic with increased taxon sampling or by the analyses of further gene sequences. Agatha and Strüder-Kypke (2014) established two subgenera within *Parallelostrombidium*. The type species *Parallelostrombidium paralatum* Xu et al., 2006 of the subgenus *Asymptokinetum* falls in a genetic clade together with congeners characterised by a broadly obovoidal to broadly ellipsoidal cell shape, while the other *Parallelostrombidium* clade comprises species with obconical to elongate obconical species like the type species of the subgenus *Parallelostrombidium*, *P. rhyticollare* (Corliss and Snyder, 1986) Agatha, 2004, for which genetic data are unfortunately not available. Apparently, the somatic ciliary pattern is not the optimal distinguishing feature in this group or at least should be combined with other reliable characters in the future.

The Choreotrichida consist of five morphologically well-characterised suborders: Leegaardiellina Laval-Peuto, Grain and Deroux, 1994; Lohmanniellina Laval-Peuto, Grain and Deroux, 1994; Strombidinopsina Small and Lynn, 1985; Strobilidiina Jankowski, 1980; and Tintinnina Kofoid and Campbell, 1929. In cladistic and molecular analyses, the aloricate suborders form a paraphyly, while the particular choreotrichid taxa are monophyletic, except for the Strombidinopsina. The genus *Strombidinopsis* is split in two genetically distinct clades, and the two *Parastrombidinopsis* species are again a separate grouping (Kim et al., 2020). Despite the lack of gene sequences for the Lohmanniellina and the strombidinopsid genus *Parastrombidium*, the probability is extremely low that the suborder Strombidinopsina becomes monophyletic.

The vase- or tube-shaped lorica is the main apomorphy of the suborder Tintinnina (tintinnids) with its currently 14 families, 77 genera, and more than 1,000 species. In contrast to the previously mentioned taxa, cytological features are barely known in tintinnids; hence, their taxonomy and classification are widely lorica-based. The lorica structure (entirely hyaline, entirely with adhered foreign particles = agglutinated, or with hyaline anterior and agglutinated posterior portion), the shape, size, structure (continuous wall or with spirals/rings), presence of windows, details of the opening rim (e.g., the presence of a gutter or a denticulation), and the texture of the wall (e.g., compact, with alveoli) were used for distinguishing species (e.g., Kofoid and Campbell, 1929, 1939). However, phylogenetic analyses indicated that the lorica features have a limited value for inferring relationships owing to homoplasies, phenotypic plasticity, and crypticity (Santoferrara and McManus, 2021). Likewise, transmission electron microscopic studies of the wall textures found only partly an ultrastructural basis of the formerly applied features (Agatha and Bartel, 2021; Laval-Peuto, 1994). Actually, the gene trees display a successive branching of the statistically well-supported monophyletic families Tintinnidiidae, Tintinnidae, Eutintinnidae, and Favellidae. The remaining families form an unresolved polytomy. The families Ascampbelliellidae, Cyttarocylididae, Epiplocylididae, Ptychocylididae, Rhabdonellidae, Undellidae, and Xystonellidae are monophyletic; yet, their contents partially changed due to the addition or withdrawal of genera based on their phylogenetic placement. Because of these surprising genetic relationships, some circumscriptions, using lorica features, are currently highly unsatisfying. Particularly, the genus *Tintinnopsis* with its about 160 species is split in 13 clades/lineages often together with taxa possessing hyaline loricae, indicating that the hard and entirely agglutinated loricae represent homoplasies or the plesiomorphic character state. The Dictyocystidae and Stenosemellidae are intertwined in gene trees, and the genus *Codonellopsis* (Fig. 1E) does not form a monophylum. In contrast to the other Dictyocystidae, several *Codonellopsis* species might not possess a lorica sac and closing apparatus. Further, several genera lost their home and are currently considered *incertae sedis*. Molecular data are missing for the Nolaclusiliidae Sniezek et al., 1991 (Table A.1).

Table A.1
Classification of the Oligotrichea (Alveolata, Ciliophora; cp. Fig. 1). Bold names mark taxa whose diagnoses are improved by the addition of consensus signature characters in the present study; underlining denotes tested taxa for which the diagnoses were not improved due to limitations: non-monophyly (a), insufficient number of gene sequences (b), absence of gene sequences (c), considerable cryptic diversity (Santoferrara et al., 2017, 2018) (d), absence of binary consensus signatures (e), lack of consensus signature characters (f), or identification to species or genus rank not verifiable (g).

	Class Oligotrichea Bütschli, 1887
	Order Oligotrichida Bütschli, 1887
	Family Cyrtostrombidiidae Agatha, 2004
	Genus *Cyrtostrombidium* Lynn and Gilron, 1993
b	Family Pelagostrombidiidae Agatha, 2004
g	Genus *Limnostrombidium* Krainer, 1995
b	Genus *Pelagostrombidium* Krainer, 1991
a	Family Strombidiidae Fauré-Fremiet, 1970 (Fig. 1B)
b	Genus *Antestrombidium* Liu et al., 2015
a	Genus *Apostrombidium* Xu et al., 2013
c	Genus *Foissneridium* Agatha, 2011
	Genus *Novistrombidium* Song and Bradbury, 1998
g	Genus *Omegastrombidium* Agatha, 2004
c	Genus *Opisthostrombidium* Agatha, 2011
a	Genus *Parallelostrombidium* Agatha, 2004
a	Genus *Propecingulum* Agatha and Strüder-Kypke, 2014
b	Genus *Sinistrostrombidium* Liu et al., 2015
a	Genus *Spirostrombidium* Jankowski, 1978
a	Genus *Strombidium* Claparède and Lachmann, 1859
b	Genus *Varistrombidium* Xu et al., 2009
b	Genus *Williophrya* Liu et al., 2011
	Family Tontoniidae Agatha, 2004
	Genus *Laboea* Lohmann, 1908
c	Genus *Paratontonia* Jankowski, 1978
	Genus *Pseudotontonia* Agatha, 2004
	Genus *Spirotontonia* Agatha, 2004
c	Genus *Tontonia* Fauré-Fremiet, 1914
	Order Choreotrichida Small and Lynn, 1985
b, d	Suborder Leegaardiellina Laval-Peuto et al., 1994
b, d	Family Leegaardiellidae Lynn and Montagnes, 1988
b, d	Genus *Leegaardiella* Lynn and Montagnes, 1988
c	Suborder Lohmanniellina Laval-Peuto et al., 1994
c	Family Lohmanniellidae Montagnes and Lynn, 1991
c	Genus *Lohmanniella* Leegaard, 1915
a	Suborder Strombidinopsina Small and Lynn, 1985
a	Family Strombidinopsidae Small and Lynn, 1985
	Genus *Parastrombidinopsis* Kim et al., 2005
c	Genus *Parastrombidium* Fauré-Fremiet, 1924
a	Genus *Strombidinopsis* Saville-Kent, 1881
	Suborder Strobilidiina Jankowski, 1980
	Family Strobilidiidae Kahl in Doflein and Reichenow, 1929 (Fig. 1D)
	Genus Pelagostrobilidium Petz et al., 1995
a	Genus *Rimostrombidium* Jankowski, 1978
b	Genus *Strobilidium* Schewiakoff, 1892
f	Suborder Tintinnina Kofoid and Campbell, 1929 (Fig. 1E)
b, e	Family Ascampbelliellidae Corliss, 1960
e	Family Cyttarocylididae Kofoid and Campbell, 1929
	Genus *Cyttarocylis* Fol, 1881
	Genus *Petalotricha* Saville-Kent, 1881
a	Family Dictyocystidae Haeckel, 1873
	Family Epiplocylididae Kofoid and Campbell, 1939
	Family Eutintinnidae Bachy et al., 2012
	Genus *Eutintinnus* Kofoid and Campbell, 1939
b, d	Genus *Dartintinnus* Smith and Santoferrara, 2017
	Family Favellidae Kofoid and Campbell, 1929
c	Family Nolaclusiliidae Sniezek et al., 1991
f	Family Ptychocylididae Kofoid and Campbell, 1929
f	Family Rhabdonellidae Kofoid and Campbell, 1929
e	Family Tintinnidae Claparède and Lachmann, 1858
	Family Tintinnidiidae Kofoid and Campbell, 1929
	Genus *Antetintinnidium* Ganser and Agatha, 2019
c	Genus *Membranicola* Foissner et al., 1999
	Genus *Tintinnidium* Saville-Kent, 1881
a	Family Stenosemellidae Campbell, 1954
f	Family Undellidae Kofoid and Campbell, 1929
	Family Xystonellidae Kofoid and Campbell, 1929
	Order Lynnellida Liu et al., 2015 (Fig. 1C)
	Family Lynnellidae Liu et al., 2011

Incertae sedis:

b	*Acanthostomella* Jörgensen, 1927
	Climacocylis Jörgensen, 1924
c	*Codonopsis* Kofoid and Campbell, 1939
	Helicostomella Jörgensen, 1924
b	*Leprotintinnus* Jörgensen, 1900
c	*Poroecus* Cleve, 1902
b	*Rhizodomus* Strelkow and Wirketis, 1950
c	*Rotundocylis* Kufferath, 1950
b	*Stylicauda* Balech, 1951
a	*Tintinnopsis* Stein, 1867

Nomen inquirendum: *Coxliella* Brandt, 1906.

Antetintinnopsis Wang et al., 2021: The genus is supposed to differ from the genus *Tintinnopsis* Stein, 1867 in the possession of a conspicuous ciliary tuft and densely arranged monokinetids in the ventral kinety's middle portion (Wang et al., 2021a,b). However, the absence of these features in *Tintinnopsis* cannot be inferred from the original description of its type species, *T. beroidea*, whose redescription by Bai et al. (2020) is based on misidentified specimens. Thus, the genus *Antetintinnopsis* is rejected by Agatha et al. (2021).

Improved diagnoses

Class Oligotrichea Bütschli, 1887

Remarks. The Oligotrichea comprise the Oligotrichida, Choreotrichida, and Lynnellida.

Improved diagnosis. Perilemmaphora with usually globular to obconical cell shape. Adoral zone of membranelles occupies apical cell portion, composed of three-rowed membranelles; occasionally, proximalmost ones two-rowed. Somatic kinetids as monokinetids, dikinetids with two cilia, dikinetids with cilia at the posterior or (initially) anterior basal body, or rarely unciliated dikinetids. Secondary absence of cirri. Reorganisation of ciliature indistinct or absent. Endoral membrane stichomonad, originates probably de novo, might secondarily be absent (Cyrtostrombidiidae). Stomatogenesis hypoapokinetal, cell division enantiotropic. Marine or brackish habitats, rarely freshwater.

Discussion: Molecular phylogenies display a sister group relationship of Oligotrichea and hypotrich ciliates, which are usually dorsoventrally flattened cells with the C-shaped adoral zone of membranelles extending around the anterior cell end and across the anterior left portion of ventral side. In hypotrichs, the individual membranelles are usually four-rowed, and a second membrane (paroral) is present. In most hypotrichs and the related euplotids, the somatic ciliature consists of ventral cirri and longitudinal dorsal kineties composed of dikinetids, each having associated a cilium only with the anterior basal body. Further, proter and opisthe have usually the same orientation during cell division (homothetogenic fission). The somatic and oral ciliature are partially or entirely reorganised during cell division in most hypotrichs, whose stomatogenesis is epiapokinetal (vs. hypoapokinetal in euplotids).

Order Oligotrichida Bütschli, 1887

Remarks: The order contains four families: the Cyrtostrombidiidae Agatha, 2004; Pelagostrombidiidae Agatha, 2004; Strombidiidae Fauré-Fremiet, 1970; and Tontoniidae Agatha, 2004.

Improved diagnosis. Oligotrichea with C-shaped, ventrally distinctly open adoral zone of membranelles; membranelles usually decrease in width towards cytostome, broad membranelles almost longitudinally arranged on collar. Endoral membrane extends longitudinally on inner wall of buccal lip, might secondarily be absent (Cyrtostrombidiidae). Peristome usually spoon-shaped to funnel-shaped, parallel to main cell axis. Mostly with acicular or rod-shaped extrusomes of trichite-type. Frequently with polysaccharidic cortical platelets in posterior cell portion. Somatic ciliature generally comprises two kineties: a curved and occasionally fragmented girdle kinety and a usually longitudinal ventral kinety. Somatic kinetids invariably as dikinetids, usually with clearly recognisable cilium only at (initially) anterior basal bodies; kinetodesmal fibril and postciliary microtubular ribbon at (initially) posterior basal bodies; presence of transverse microtubular ribbon uncertain. Stomatogenesis in subsurface tube, axes of proter and opisthe inverse in middle dividers. Marine or brackish habitats, rarely freshwater. 18S rRNA gene: 807 (C), 847 (G), 848 (T), 854 (-), 873 (T), 883 (A), 1079 (G), 1086 (T), 1231 (T), 1239 (C), 1246 (T), 1261 (G), 1336 (A), 1482 (G), 1514 (T), 1586 (T); ITS1-5.8S rRNA gene-ITS2 region: 146 (A), 168 (T), 169 (T), 173 (A), 183 (G), 197 (A), 303 (C), 306 (C), 307 (T), 309 (G), 311 (C), 350 (G), 363 (C), 431 (T); 28S rRNA gene: 27 (G), 54 (C), 63 (A), 74 (T), 93 (C), 107 (G), 161 (A), 226 (A), 234 (G), 246 (T), 253 (C), 291 (T), 363 (C), 373 (T), 384 (C), 391 (G), 392 (A), 394 (G), 399 (C), 512 (A), 513 (T), 536 (G), 615 (A), 616 (T), 630 (T), 631 (C), 671 (C), 672 (G), 673 (T), 674 (A), 691 (T), 703 (C).

Discussion. The Oligotrichida differ from the Choreotrichida and Lynnellida in several morphological features: (i) the shape of the adoral zone (C-shaped in oligotrichids and *Lynnella*, usually circular in choreotrichids); (ii) width of proximal adoral membranelles (smaller than distal membranelles in oligotrichids, elongated and extending into buccal cavity in choreotrichids and *Lynnella*); (iii) orientation of distal membranelles (longitudinal on collar in oligotrichids and *Lynnella*, across the peristomial rim forming a contorted pattern in choreotrichids); (iv) shape and orientation of peristomial field (spoon- or funnel-shaped and parallel to main axis in oligotrichids, rather flat and perpendicular to main cell axis in choreotrichids and *Lynnella*); (v) the position of the endoral membrane (extends on inner wall of buccal lip in oligotrichids and *Lynnella*, across peristomial field in choreotrichids); (vi) the structure of the somatic kinetids [oligotrichids with plesiomorphic state, choreotrichids and *Lynnella* with apomorphic states according to Agatha and Strüder-Kypke (2014)]; (vii) course of somatic kineties (one longitudinal and one distinctly curved and occasionally fragmented row in oligotrichids, several to many mostly longitudinal rows in choreotrichids, and two exclusively longitudinal rows in *Lynnella*); (viii) extrusomes of trichite-type (usually present in oligotrichids, absent in choreotrichids and *Lynnella*); (ix) polysaccharidic platelets (usually present in oligotrichids, absent in choreotrichids and *Lynnella*); (x) stomatogenesis (in subsurface tube in oligotrichids, in pouch in choreotrichids and *Lynnella*); and (xi) orientation of oral primordium (inverse in oligotrichids, perpendicular to proter's main axis in choreotrichids and *Lynnella*).

Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised 47, 16, and 14 sequences, respectively, while the reference group Choreotrichida comprised 153, 86, and 79 sequences, respectively; a single sequence each was available for the Lynnellida. The Oligotrichida differed from the Choreotrichida in one binary consensus signature in the 18S rRNA gene, one asymmetric consensus signatures in the ITS1-5.8S rRNA gene-ITS2 region, as well as in one binary and eight asymmetric consensus signatures in the 28S rRNA gene. The Oligotrichida are distinguished from the Lynnellida by several binary consensus signatures: 16 in the 18S rRNA gene, 13 in the ITS1-5.8S rRNA gene-ITS2 region, and 29 in the 28S rRNA gene (for character states in the reference groups, see Supplementary Tables S2, S6 and Supplementary Data S1).

Family Cyrtostrombidiidae Agatha, 2004

Remarks: The family consists of a single genus only, for which the diagnosis is thus identical.

Improved diagnosis: Oligotrichida having buccal membranelles and endoral membrane secondarily reduced. Pharyngeal fibres thick and cyrtos-like after protargol impregnation. Usually, one macronucleus. Stomatogenesis in transient tube. Marine or brackish habitats. ITS1-5.8S rRNA gene-ITS2 region: 136 (A), 533 (A); 28S rRNA gene: 162 (T).

Discussion: For the morphological comparison with the other oligotrichid families, see Strombidiidae. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised four, one, and one sequences, respectively, while the reference group "other Oligotrichida" comprised 43, 15, and 13 sequences, respectively. One binary and one asymmetric consensus signature were detected in the ITS1-5.8S rRNA gene-ITS2 region, while

only one binary consensus signature in the 28S rRNA gene (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1); no consensus signatures were found in the 18S rRNA gene.

Family Pelagostrombidiidae Agatha, 2004

Remarks: Since the family is represented by only a single 18S rRNA gene sequence, we refrained from improving the diagnosis. The permanent subsurface tube (neoformation organelle), in which stomatogenesis takes place, represents an apomorphy of the family. Hence, we provide the diagnosis given by Bardele et al. (2018). The family comprises two genera: *Limnostrombidium* Krainer, 1995 and *Pelagostrombidium* Krainer, 1991.

Diagnosis: Oligotrichida with collar and buccal membranelles and endoral membrane. Somatic cilia and ventral kinety might secondarily be absent (genus *Pelagostrombidium*). One macronucleus. Stomatogenesis in neoformation organelle. Freshwater habitats.

Discussion: For the morphological comparison with the other oligotrichid families, see Strombidiidae. Only one 18S rRNA gene sequence is available, in which three asymmetric consensus signatures were detected, using "other Oligotrichida" as reference group; data on the ITS1-5.8S rRNA gene-ITS2 region and the 28S rRNA gene are lacking (for character states in the reference group see Supplementary Tables S2, S6 and Supplementary Data S1).

Family Strombidiidae Fauré-Fremiet, 1970

Remarks: Since the oligotrichid family Strombidiidae is not monophyletic in molecular analyses and merely characterised by plesiomorphies in cladistic analyses (see below), we did not check for signature characters. The family comprises 13 genera: *Antestrombidium* Liu et al., 2015; *Apostrombidium* Xu et al., 2013; *Foissneridium* Agatha, 2011; *Novistrombidium* Song and Bradbury, 1998; *Omegastrombidium* Agatha, 2004; *Opisthostrombidium* Agatha, 2011; *Parallelostrombidium* Agatha, 2004; *Propecingulum* Agatha and Strüder-Kypke, 2014; *Sinistrostrombidium* Liu et al., 2015; *Spirostrombidium* Jankowski, 1978; *Strombidium* Claparède and Lachmann, 1859 (Fig. 1B); *Varistrombidium* Xu et al., 2009; and *Williophrya* Liu et al., 2011. Only the diagnosis of *Novistrombidium* is complemented here by signature characters owing to unclear phylogenetic relationships, non-monophylies, or unavailable genetic data in the remaining genera.

Diagnosis: Oligotrichida with collar and buccal membranelles and endoral membrane. Usually one macronucleus. Stomatogenesis in transient tube. Marine or brackish habitats, rarely freshwater.

Discussion: The Tontoniidae differ by a highly contractile tail and numerous macronuclear nodules, the Cyrtostrombidiidae are characterised by thick pharyngeal fibres besides a secondary absence of an endoral membrane and buccal membranelles, and the Pelagostrombidiidae are distinguished by a stomatogenesis in a permanent subsurface tube.

Genus *Novistrombidium* Song and Bradbury, 1998

Remarks: Agatha and Strüder-Kypke (2014) subdivided the genus *Novistrombidium* based on morphologic and ontogenetic differences and a considerable genetic dissimilarity in the 18S rRNA gene sequences, establishing the subgenera *Novistrombidium* and *Propecingulum*. Additionally, Li et al. (2013) spotted discrepancies in the ITS2 secondary structures between the subgenera. Since the subgenus *Propecingulum* had been raised to genus rank by Küppers et al. (2019; see below), the genus *Novistrombidium* is characterised by the features of its nominotypical subgenus only.

Improved diagnosis: Strombidiidae with longitudinal ventral kinety and girdle kinety commencing in right cell half and performing one dextral spiral. Extrusome attachment sites in question mark-shaped pattern directly posterior to adoral membranelles and in an arc on posterior dorsal side. Oral primordium originates between question mark-shaped pattern of extrusome attachment sites and girdle kinety in left half of ventral side. Marine and brackish habitats. ITS1-5.8S rRNA gene-ITS2 region: 140 (T), 274 (C), 320 (A), 321 (C), 323 (A), 467 (G), 479 (T), 497 (T), 500 (G), 522 (A), 532 (T), 581 (G); 28S rRNA gene: 136 (T), 137 (C), 199 (T), 202 (G), 244 (G), 247 (A), 508 (T), 518 (C), 524 (A), 528 (T), 533 (A), 535 (A), 537 (T), 633 (C), 687 (A).

Discussion: Propecingulum differs from *Novistrombidium* in several features. Its girdle kinety commences in the left cell half and performs slightly more than one dextral spiral. The extrusome attachment sites are directly anteriorly to the girdle kinety. The oral primordium originates anteriorly to the stripe of extrusome attachment sites. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised two, one, and one sequences, respectively, while the reference group "other Oligotrichida" 45, 15, and 13 sequences, respectively. Eight binary and four asymmetric consensus signatures were detected in the ITS1-5.8S rRNA gene-ITS2 region and eight binary and seven asymmetric consensus signatures in the 28S rRNA gene (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1); no consensus signatures were found in the 18S rRNA gene.

Genus *Propecingulum* Agatha and Strüder-Kypke, 2014

Remarks: We discuss *Propecingulum* here due to nomenclatural issues. Küppers et al. (2019) raised the subgenus to genus rank, although it is not monophyletic in 18S rRNA phylogenies and thus requires a revision. Agatha and Strüder-Kypke (2014) designated *Novistrombidium sinicum* Liu et al., 2009 as type of the subgenus and affiliated further species: *Novistrombidium ioanum* (Lynn and Gilron, 1993) Agatha and Strüder-Kypke, 2014; *Novistrombidium orientale* (Liu et al., 2009) Agatha and Strüder-Kypke, 2014; and *Novistrombidium platum* (Song and Packroff, 1997) Agatha and Strüder-Kypke, 2014. Later, *Novistrombidium* (*Propecingulum*) *rufinoi* da Silva Paiva et al., 2016 had been added. Since Agatha and Strüder-Kypke (2014) established *Propecingulum* simultaneously at genus and subgenus rank (Art. 43.1. of the ICZN, 1999) and made the new combinations at either rank, those combinations performed by Küppers et al. (2019) are incorrect. The same is true for the species established by da Silva Paiva et al. (2016). According to Article 50.3.1 of the ICZN (1999), the authorship is not affected by the rank used; thus, the authorship is *Propecingulum rufinoi* da Silva Paiva et al., 2016.

Family Tontoniidae Agatha, 2004

Remarks: The family comprises five genera: *Laboea* Lohmann, 1908 (monotypic); *Paratontonia* Jankowski, 1978; *Pseudotontonia* Agatha, 2004; *Spirotontonia* Agatha, 2004; and *Tontonia* Fauré-Fremiet, 1914.

Improved diagnosis: Oligotrichida with contractile, usually ciliated tail, which might secondarily be absent (genus *Laboea*). Oral ciliature with collar and buccal membranelles and endoral membrane. Usually, numerous macronuclear nodules. Stomatogenesis in transient tube. Marine and brackish habitats. 18S rRNA gene: 930 (A), 1383 (C), 1394 (G); ITS1-5.8S rRNA gene-ITS2 region: 177 (A), 331 (G), 341 (T), 479 (C); 28S rRNA gene: 165 (T), 369 (T), 395 (T), 398 (A), 524 (C), 702 (T).

Discussion: For the morphological comparison with the other oligotrichid families, see Strombidiidae. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised six, three, and two sequences, respectively, while the reference group "other Oligotrichida" comprised 41, 13, and 12 sequences, respectively. One binary and two asymmetric consensus signatures were detected in the 18S rRNA gene, two binary and two asymmetric consensus signatures in the ITS1-5.8S rRNA gene-ITS2 region, and three binary and three asymmetric consensus signatures in the 28S rRNA gene (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1).

Order Choreotrichida Small and Lynn, 1985

Remarks: The order comprises four aloricate suborders (Leegaardiellina Lava-Peuto, Grain and Deroux, 1994; Lohmanniellina Laval-Peuto, Grain and Deroux, 1994; Strobilidiina Jankowski, 1980; and Strombidinopsina Small and Lynn, 1985) and one loricate suborder (Tintinnina Kofoid and Campbell, 1929).

Improved diagnosis: Oligotrichea with circular adoral zone of membranelles, rarely with minute secondary ventral gap (genera *Parastrombidinopsis* and *Parastrombidium*). Collar membranelles extend across peristomial rim, forming a contorted pattern, proximalmost ones

usually elongated, extending into buccal cavity. Endoral membrane extends across peristomial field, plunging into buccal cavity. Peristomial field perpendicular to main cell axis, rather flat. Occasionally with potentially extrusive organelles. Somatic ciliature usually comprises several to many longitudinal kineties. Somatic kinetids as monokinetids, dikinetids with two cilia, and/or dikinetids with cilia at the posterior or (initially) anterior basal bodies. Stomatogenesis in subsurface pouch, oral primordium perpendicular to proter's main cell axis. Marine or brackish habitats, rarely freshwater. 18S rRNA gene: 463 (T), 541 (T), 852 (T), 854 (C), 883 (A), 913 (C), 1239 (C), 1261 (G), 1336 (A), 1482 (G), 1514 (T); 28S rRNA gene: 226 (A), 369 (T), 373 (T), 374 (G), 377 (G), 378 (C), 384 (C), 391 (A), 509 (A), 513 (T), 615 (A).

Discussion. For comparison with the other orders, see Oligotrichida. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised 153, 86, and 79 sequences, respectively, while the reference group Oligotrichida comprised 47, 16, and 14 sequences, respectively; the Lynnellida consist of a single sequence each. The Choreotrichida differed from the Oligotrichida in one binary consensus signature each in the 18S rRNA gene and 28S rRNA gene. The Choreotrichida are distinguished from the Lynnellida by several binary consensus signatures: 11 in the 18S rRNA gene and ten in the 28S rRNA gene (for character states in the reference groups, see Supplementary Tables S2, S6 and Supplementary Data S1).

Family Leegaardiellidae Lynn and Montagnes, 1988

Remarks: The suborder Leegaardiellina contains a single family and a single genus. Since only a single sequence is available and a high diversity is indicated by environmental samples (Santoferrara et al., 2017, 2018), we refrain from complementing the diagnosis by signature characters.

Diagnosis: Choreotrichida with bases (polykinetids) of collar membranelles divided into inner and outer portions. One or two globular macronuclear nodules. Somatic kineties anteriorly shortened, composed of dikinetids, each having two cilia or one cilium associated only with the anterior basal body. Marine and brackish habitats.

Discussion: The bipartited collar membranelles are the autapomorphy of the Leegaardiellidae. The Strobilidiidae differ mainly by the C-shaped macronucleus and the somatic kineties composed of condensed monokinetids with cilia directed leftwards due to kinetal lips covering their proximal portions (Fig. 1D). In the non-monophyletic Strombidinopsidae, the somatic ciliature is composed of dikinetids, each having associated invariably two cilia, and the adoral zone of membranelles might show a secondary gap on ventral side (genera *Parastrombidinopsis* and *Parastrombidium*). The Lohmanniellidae have somatic dikinetids, each with a cilium only at the posterior basal body (Agatha and Strüder-Kypke, 2007); genetic data are not available. The Tintinnina are characterised by the possession of a lorica. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised one sequence each, while the reference group "other Choreotrichida" comprised 152, 85, and 78 sequences, respectively. Six binary and nine asymmetric consensus signatures were detected in the 18S rRNA gene, nine asymmetric consensus signatures in the ITS1-5.8S rRNA gene-ITS2 region, and four binary and six asymmetric consensus signatures in the 28S rRNA gene (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1).

Family Strobilidiidae Kahl in Doflein and Reichenow, 1929

Remarks: The family is the sole member of the suborder Strobilidiina and comprises three genera (*Pelagostrobilidium* Petz et al., 1995; *Rimostrombidium* Jankowski, 1978; and the monotypic genus *Strobilidium* Schewiakoff, 1892; Fig. 1D). Hence, the diagnosis of the family is also applicable to the suborder.

Improved diagnosis: Choreotrichida with continuous collar polykinetids. Macronucleus usually C-shaped with ventral gap, transversely orientated underneath collar membranelles. Somatic kineties composed of condensed monokinetids with cilia directed leftwards due to kinetal lips covering their proximal portions, usually anteriorly and/or posteriorly shortened. Marine or brackish habitats, rarely freshwater. 28S rRNA gene: 69 (A), 247 (G).

Discussion: Kinetal lips and condensed monokinetids are autapomorphies of the Strobilidiidae; the rare reports of somatic dikinetids with two cilia each need verification. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised seven, two, and two sequences, respectively, while the reference group "other Choreotrichida" comprised 146, 84, and 77 sequences, respectively. One binary and one asymmetric consensus signature were detected in the 28S rRNA gene; no signatures were found in the 18S rRNA gene and the ITS1-5.8S rRNA gene-ITS2 region (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1).

Suborder Tintinnina Kofoid and Campbell, 1929

Remarks: The suborder consists of 14 families and 77 genera (Fig. 1E). Gruber et al. (2018) and Agatha et al. (2022) revealed by transmission electron microscopy unique kinetid types, which highly likely developed in at least the last common ancestor of tintinnids with hard loricae; possibly, Tintinnidiidae with soft loricae share this feature, but ultrastructural data are missing. The posterior dikinetidal basal bodies and monokinetids have overlapping postciliary ribbons, kinetodesmal fibrils, one microtubular ribbon extending horizontally leftwards, and one further microtubular ribbon extending obliquely to the anterior left. The anterior dikinetidal basal bodies are equipped with transverse ribbons and one microtubular ribbon extending obliquely to the anterior left.

Diagnosis: Choreotrichida with continuous collar polykinetids. Cell attached to bottom of vase- or tube-shaped lorica by contractile peduncle. Two, occasionally more macronuclear nodules; rarely, only one. Capsules (potentially extrusive, flask-shaped organelles) in striae (cytoplasmic strands adhered to adoral membranelles) and/or tentaculoids (pin-shaped cytoplasmic extensions between membranelles). Somatic kineties numerous, mostly longitudinal. Somatic kinetids as monokinetids, dikinetids with two cilia, and/or dikinetids with cilia only at posterior basal bodies. Marine or brackish habitats, rarely freshwater.

Discussion: Despite the characteristic apomorphic lorica of the tintinnid ciliates, no signatures were detected with the aloricate choreotrichids as reference group (18S rRNA gene: 136 vs. 17 sequences; ITS1-5.8S rRNA gene-ITS2 region: 78 vs. 8 sequences; 28S rRNA gene: 71 vs. 8 sequences).

Family Ascampbelliellidae Corliss, 1960

Remarks: The family contains besides the type genus *Ascampbelliella* Corliss, 1960, the *incertae sedis* genera *Luxiella* Lecal, 1953 and *Niemarshallia* Corliss, 1960. Based on molecular data, the genus *Acanthostomella* Jörgensen, 1927 is excluded from that family and considered *incertae sedis* in Tintinnina (Santoferrara and McManus, 2021; Santoferrara et al., 2018). Owing to the uncertain affiliation of *Luxiella* and *Niemarshallia*, the following (preliminary) diagnosis is identical to that of the type genus *Ascampbelliella*. We refrain from adding signatures to the diagnosis as the query group comprised only one sequence per marker, and merely three asymmetric consensus signatures could be identified.

Diagnosis: Tintinnina with hard and hyaline lorica very small and broadly urceolate in shape, with bipartite collar comprising an erected inner rim and an outer flaring one or a horizontally spreading rim, both separated by a gutter; rarely, outer rim repeated. Collar rims smooth. Posterior end closed, broadly rounded, bluntly pointed, or obconical. Lorica wall smooth, potentially trilaminar. Marine habitats.

Discussion: According to Kofoid and Campbell (1929), *Ascampbelliella* loricae differ from those of *Acanthostomella* in a smooth outer rim (vs. distinctly denticulate) and the general absence of agglutinated particles (vs. adhered coccoliths in some species; Burns, 1983). Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised one sequence each, while the reference group "other Tintinnina" comprised 135, 77, and 70 sequences, respectively. Three asymmetric consensus signatures were detected in the 18S rRNA gene (for character states in the reference group, see

Supplementary Tables S2, S6 and Supplementary Data S1), while no consensus signatures were found in the ITS1-5.8S rRNA gene-ITS2 region and the 28S rRNA gene.

Family Cyttarocylididae Kofoid and Campbell, 1929

Synonym: Petalotrichidae Kofoid and Campbell (1929)

Remarks: The genus *Cyttarocylis* was established by Fol (1881) with *Dictyocysta cassis* Haeckel, 1873 as type species (by monotypy; Aescht, 2001). In the same year, but later, Saville-Kent (1881) provisionally established the genus *Petalotricha* (p. 627), for which *Tintinnus ampulla* Fol, 1881 was subsequently designated as type species by Brandt (1907). For each genus, Kofoid and Campbell (1929) established a family. Corliss (1979) followed the suggestion of Laval-Peuto in synonymising the families with priority of the Cyttarocylididae. Gene sequence data (18S rRNA, 5.8S also demonstrated a high similarity of the type species of both) also demonstrated a high similarity of the type species of both genera (Bachy et al., 2012; Santoferrara et al., 2017). While Bachy et al. (2012) synonymised *Petalotricha* with *Cyttarocylis*, Santoferrara et al. (2017) kept the two type species separate and retained the genera based on the differences in their 28S rDNA gene sequences and lorica morphologies. We agree with the synonymisation of the Petalotrichidae with the Cyttarocylididae, which both merely contain their type genera (Laval-Peuto, 1994) and add the nomenclatorial framework: Article 24.2.2 of the ICZN (1999) states that if there are two names based on different types published on the same date in the same work [i.e., in Kofoid and Campbell (1929)], the precedence of the names is fixed by the first reviser [i.e., Corliss (1979)].

Laval (1972) studied the lorica of *Petalotricha ampulla* by transmission electron microscopy and later characterised the family Cyttarocylididae by a trilaminar and tubular wall texture (Laval-Peuto, 1994). Recently, Agatha and Bartel (2021) confirmed this texture in *Cyttarocylis*.

Improved diagnosis: Tintinnina with hard and hyaline lorica campanulate, kettle-shaped, or elongate obconical in shape, with flaring collar set off by nuchal constriction or slightly obconical portion, posteriorly closed. Lorica wall trilaminar, composed of electron-dense inner and outer layers and a tubular middle layer, distinctly reticulate or with smooth surface and some windows in anterior portion. Two or more macronuclear nodules. Marine habitats.

Discussion: The trilaminar and tubular wall texture of the Cyttarocylididae is unique. In gene trees, the family frequently groups with the Ascampbelliellidae, Epiplocylididae, Ptychocylididae, and Rhabdonellidae, which often have surface ridges and windows in their alveolar lorica walls (Agatha and Bartel, 2021). Regarding the 18S rRNA, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA, the Cyttarocylididae comprised two sequences each, while the reference group "other Tintinnina" comprised 134, 76, and 69 sequences, respectively. The family diagnosis was not complemented with molecular signature characters, as merely asymmetric consensus signatures were detected: two in the 18S rRNA gene, one in the ITS1-5.8S rRNA gene-ITS2 region, and three in the 28S rRNA gene (see Supplementary Tables S2, S6 and Supplementary Data S1).

Family Epiplocylididae Kofoid and Campbell, 1939

Remarks: The family comprises three genera: *Epicancella* Kofoid and Campbell, 1929; *Epiplocylis* Jörgensen, 1924; and *Epiplocyloides* Hada, 1938. Agatha and Bartel (2021) analysed own and published scanning electron micrographs for the latter two genera and light microscopic data on *Epicancella* (Brandt, 1906, 1907; Fernandes, 2004) and characterised the family by the possession of reticulate surface ridges and partially windows (apparently absent in *Epicancella*).

Improved diagnosis: Tintinnina with hard, hyaline, and roughly acorn-shaped lorica; anterior portion cylindroidal, posterior portion closed, obconical and bluntly pointed or with short posterior process. Opening rim smooth, occasionally with bipartite collar composed of erected inner and flaring outer rim. Lorica wall alveolar, with comparatively high reticulate surface ridges on posterior lorica portion, frequently with windows in centre of meshes. Two macronuclear nodules. Marine habitats. 18S rRNA gene: 1515 (G), 1687 (T); ITS1-5.8S rRNA gene-ITS2 region: 368 (C), 399 (T), 425 (G).

Discussion: Potentially, all Rhabdonellidae are characterised by surface ridges similar to those of the Epiplocylididae; however, their windows - if present - are much smaller. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised two, one, and one sequences, respectively, while the reference group "other Tintinnina" comprised 134, 77, and 70 sequences, respectively. One binary and one asymmetric consensus signature were detected in the 18S rRNA gene and three asymmetric consensus signatures in the ITS1-5.8S rRNA gene-ITS2 region (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1); no consensus signatures were found in the 28S rRNA gene.

Family Eutintinnidae Bachy, Gómez, López-García, Dolan and Moreira, 2012

Remarks: The family comprises the genus *Dartintinnus* Smith and Santoferrara, 2017 that forms a sister group to the genus *Eutintinnus* Kofoid and Campbell, 1939 in gene trees; the latter displays a deep bipartition. Information on the texture of the lorica wall comes from transmission electron microscopy on *Eutintinnus* species (Agatha and Bartel, 2021).

Improved diagnosis: Tintinnina with hard and hyaline lorica having very rarely adhered diatoms, with posterior opening, roughly cylindroidal to obconical, with truncate, rarely collapsing ends. Lorica wall trilaminar and compact, middle layer with highly ordered electron-dense rods perpendicular to lorica surface. Two or four macronuclear nodules. Monokinetidal ventral kinety occasionally fragmented and/or shifted posteriorly. Right and left ciliary fields, usually a lateral ciliary field. One or two, very rarely three dorsal kineties. Adoral zone of membranelles frequently somewhat oblique and elliptical in contracted specimens. Marine and brackish habitats. 18S rRNA gene: 371 (G), 491 (G), 1246 (C); 28S rRNA gene: 163 (C), 291 (C), 298 (G), 352 (C), 355 (C).

Discussion: Only the Tintinnidae Claparède and Lachmann, 1858 share compact and hard lorica walls. In *Amphorides* Strand, 1928, the texture of the middle wall layer is similar to that in *Eutintinnus* (Agatha and Bartel, 2021); however, the inner layer is much thicker than the outer layer and even thicker than the middle layer. The loricae of Tintinnidae are posteriorly closed, except for *Epicranella*, *Daturella*, *Salpingacantha*, and *Salpingella*, and have occasionally more or less longitudinal ridges. Further, they are seemingly characterised by a more oblique and elliptical adoral zone of membranelles in contracted specimens (Agatha and Strüder-Kypke, 2014). According to Bai et al. (2020), *Amphorellopsis* (Tintinnidae) has exclusively dikinetidal somatic kineties, i.e., one anterior dikinetid with two cilia is followed by dikinetids, each having a cilium associated only with the posterior basal body. The family Nolaclusiliidae differs from *Amphorellopsis* in the presence of a ventral kinety and a derived structure of the somatic kineties (composed of an anterior diciliated dikinetid and many monokinetids, rarely entirely monokinetidal). In contrast to the Eutintinnidae, however, the somatic ciliary pattern is less complex in the Nolaclusiliidae owing to the lack of a monokinetidal lateral ciliary field and dorsal kineties (Sniezek et al., 1991; Snyder and Brownlee, 1991). Accordingly, the Nolaclusiliidae branch off between the Tintinnidae and Eutintinnidae in cladistic analyses (Agatha and Strüder-Kypke, 2014). Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised 14, five, and six sequences, respectively, while the reference group "other Tintinnina" comprised 122, 73, and 65 sequences, respectively. Two binary and one asymmetric consensus signatures were detected in the 18S rRNA gene and three binary and two asymmetric consensus signatures were identified in the 28S rRNA gene (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1); no consensus signatures were found in the ITS1-5.8S rRNA gene-ITS2 region.

Genus *Dartintinnus* Smith and Santoferrara, 2017

Remarks: Currently, the genus is monotypic and only one reference sequence is available. However, a high diversity is expected within the Eutintinnidae based on environmental sequences (Santoferrara et al., 2018). Hence, we refrain from adding signatures, but slightly improve the diagnosis given by Smith et al. (2018).

Improved diagnosis: Eutintinnidae with collapsible lorica openings. Two macronuclear nodules. One dorsal kinety. Adoral zone of membranelles perpendicular to main cell axis and probably circular in contracted cells.

Discussion: Eutintinnus loricae are not capable to collapse, and the cells have usually four macronuclear nodules and two, very rarely three dorsal kineties. The loricae of Nolaclusiliidae Sniezek et al., 1991 are collapsible but posteriorly closed. Only one 18S rRNA gene sequence is available for *Dartintinnus*; data on the ITS1-5.8S rRNA gene-ITS2 region and the 28S rRNA gene are missing. DeSignate detected 25 binary and 18 asymmetric consensus signatures, using 'other Eutintinnidae' (13 sequences) as reference group (see Supplementary Tables S2, S6 and Supplementary Data S1).

Family Favellidae Kofoid and Campbell, 1929

Remarks: Data on signature characters are added to the improved diagnosis published by Agatha and Strüder-Kypke (2012). The family comprises a single genus, for which the given diagnosis is thus also applicable.

Improved diagnosis: Tintinnina with hard and hyaline lorica. Protolorica campanulate or cylindroidal, posterior portion broadly rounded to obconical, usually with hollow, posteriorly closed process bearing ribs, often with spiralled or annulated epilorica. Paralorica coxlielliform, with or without posterior process. Opening rims smooth or irregular, often with small, erected collar. Lorica wall potentially monolaminar with alveoli, surface smooth. Adoral zone of membranelles horizontally orientated in contracted specimens. Somatic ciliature comprises two dorsal kineties plus occasionally some kinety fragments, a monokinetidal ventral kinety as well as a lateral, right, and left ciliary field. Marine and brackish habitats. 18S rRNA gene: 531 (T), 532 (T), 1090 (A), 1651 (C); ITS1-5.8S rRNA gene-ITS2 region: 487 (T), 503 (A); 28S rRNA gene: 93 (G), 503 (A), 790 (C).

Discussion: At first glance, *Favella* resembles *Schmidingerella* in lorica shape; yet, they differ in the subapical bulge (absent vs. present), the wall (smooth and without windows vs. with reticulate surface ridges and windows), and the somatic ciliary pattern (e.g., ventral kinety monokinetidal vs. bipartite in monokinetidal anterior and dikinetidal posterior portion; broad unciliated stripe separating ventral kinety and right ciliary field absent vs. present; Agatha and Strüder-Kypke, 2012). Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised six, three, and one sequences, respectively, while the reference group "other Tintinnina" comprised 130, 75, and 70 sequences, respectively. One binary and three asymmetric consensus signatures were detected in the 18S rRNA gene, two asymmetric consensus signatures in the ITS1-5.8S rRNA gene-ITS2 region, and three asymmetric consensus signatures in the 28S rRNA gene (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1).

Families Ptychocylididae, Rhabdonellidae, Tintinnidae, and Undellidae

Remarks: In none of these families signature characters were detected in the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, except for the Tintinnidae that are distinguished from "other Tintinnina" by one asymmetric consensus signature in the ITS1-5.8S rRNA gene-ITS2 region. Hence, the diagnoses of the families could not be complemented by signature characters (see Supplementary Tables S2, S6 and Supplementary Data S1).

Family Tintinnidiidae Kofoid and Campbell, 1929

Remarks: The family comprises three genera: *Antetintinnidium* Ganser and Agatha, 2019; *Membranicola* Foissner et al., 1999; and *Tintinnidium* Saville-Kent, 1881.

Improved diagnosis: Tintinnina with usually cylindroidal lorica posteriorly closed by lorica wall or subterminal membrane. Lorica wall soft, gelatinous, with agglutinated particles. One or two macronuclear nodules. Somatic ciliature interrupted by distinct ventral stripe unciliated or with de-novo-originating ventral organelles. Somatic kineties exclusively dikinetidal, with dikinetids in anterior and monokinetids in posterior portion, or entirely monokinetidal. Buccal membranelle indistinct or absent. Freshwater habitats, occasionally in marine and brackish habitats; mostly planktonic, rarely sessile. 18S rRNA gene: 807 (T); ITS1-5.8S rRNA gene-ITS2 region: 224 (A), 244 (T), 248 (T), 282 (C), 295 (T), 326 (T), 344 (G), 397 (T), 402 (C), 427 (A), 480 (C).

Discussion: In contrast to the Tintinnidiidae, the remaining tintinnids possess hard loricae. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised seven, three, and five sequences, respectively, while the reference group "other Tintinnina" comprised 129, 75, and 66 sequences, respectively. One binary consensus signature was detected in the 18S rRNA gene, and six binary and five asymmetric consensus signatures were identified in the ITS1-5.8S rRNA gene-ITS2 region; no consensus signatures were found in the 28S rRNA gene (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1).

Genus *Antetintinnidium* Ganser and Agatha, 2019

Remarks: Antetintinnidium differs from *Tintinnidium* and *Membranicola* in the somatic ciliary pattern, namely, by the absence of ventral organelles in an otherwise unciliated longitudinal stripe separating the left and right ciliary fields (Ganser and Agatha, 2019). In the gene trees, the sequence of *Antetintinnidium* is an adelphotaxon to a specimen identified as *Tintinnidium* sp. by Santoferrara et al. (2013) based on the possession of a soft lorica, while its cell features are unknown. Since the remaining *Tintinnidium* sequences form a distinct, statistically well-supported clade, it seems very likely that the former *Tintinnidium* sequence rather represents an *Antetintinnidium* species. Actually, this alternative results in more signature characters for *Antetintinnidium* than the assumption that the sister sequence belongs to a *Tintinnidium* species (e.g., 14 vs. 6 binary consensus signatures and 6 vs. 4 asymmetric consensus signatures in the 18S rRNA gene).

Improved diagnosis: Tintinnidiidae with cylindroidal, posteriorly broadly rounded lorica. Two macronuclear nodules. Somatic kineties interrupted by unciliated ventral stripe; kineties exclusively composed of dikinetids, each having associated a cilium only with the posterior basal body, all originate by intrakinetal proliferation of basal bodies. With buccal membranelle. Planktonic in marine and brackish habitats. 18S rRNA gene: 183 (A), 503 (A), 666 (G), 670 (C), 705 (C), 709 (G), 957 (T), 1076 (A), 1093 (T), 1366 (A), 1377 (C), 1431 (G), 1587 (C), 1637 (A), 1711 (C), 1713 (G), 1741 (T), 1742 (G), 1743 (A), 1767 (C); ITS1-5.8S rRNA gene-ITS2 region: 58 (G), 63 (T), 64 (T), 65 (G), 66 (A), 101 (G), 102 (A), 137 (A), 138 (G), 139 (C), 140 (T), 145 (C), 150 (C), 289 (T), 321 (C), 339 (G), 387 (T), 395 (G), 400 (T), 401 (G), 422 (G), 425 (T), 475 (T), 476 (A); 28S rRNA gene: 65 (T), 81 (G), 82 (A), 86 (C), 110 (A), 112 (A), 114 (C), 127 (T), 138 (A), 140 (T), 182 (G), 199 (A), 220 (T), 234 (C), 253 (G), 268 (A), 289 (C), 345 (G), 393 (T), 395 (A), 402 (C), 404 (G), 489 (A), 494 (G), 497 (C), 502 (G), 533 (G), 554 (-), 684 (A), 695 (G), 705 (T), 732 (C), 733 (T), 782 (A).

Discussion: Tintinnidiidae belonging to the genera *Tintinnidium* and *Membranicola* possess two ventral organelles, which possibly originate de novo. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised two sequences for each marker, while the reference group "other Tintinnidiidae" comprised five, one, and three sequences, respectively. Fourteen binary and six asymmetric consensus signatures were detected in the 18S rRNA gene, 24 binary consensus signatures in the ITS1-5.8S rRNA gene-ITS2 region, and 26 binary and eight asymmetric consensus signatures in the 28S rRNA gene (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1).

Family Xystonellidae Kofoid and Campbell, 1929

Remarks: The family comprises four genera: *Parafavella* Kofoid and Campbell, 1929; *Spiroxystonella* Kofoid and Campbell, 1939; *Xystonella*

Brandt, 1906; and *Xystonellopsis* Jörgensen, 1924. In genetic analyses, *Parafavella* and *Xystonella* group with the genus *Dadayiella* Kofoid and Campbell, 1929 (Santoferrara and McManus, 2021), which is thus placed *incertae sedis* in the Xystonellidae but not included in the diagnosis below. Own and literature data from transmission electron microscopy were analysed by Agatha and Bartel (2021) who inferred an alveolar wall texture in the Xystonellidae. In the sister taxa *Xystonella* and *Dadayiella*, the wall is bilaminar with a compact outer layer and simple septae in the alveolar inner layer, while the *Parafavella* species are characterised by septae with alveolar prisms at their inner and outer ends.

Improved diagnosis: Tintinnina with hard, hyaline, and usually elongate obconical lorica with closed posterior end. Opening rim simple or with erected inner and flaring, occasionally denticulate outer rim. Lorica wall alveolar and smooth, rarely, with protruding rings/spirals. Marine. ITS1-5.8S rRNA gene-ITS2 region: 434 (C), 467 (G), 468 (G), 478 (G); 28S rRNA gene: 107 (G), 119 (G), 191 (T), 394 (A).

Discussion: The monographs of Kofoid and Campbell (1929, 1939), Laval-Peuto (1994), and Lynn (2008) provide descriptions of the family Xystonellidae, whose content, however, changed owing to the consideration of phylogenetic relationships inferred from molecular data. Furthermore, differences in the wall texture used by Kofoid and Campbell (1929, 1939) have no ultrastructural base (Agatha and Bartel, 2021; Laval-Peuto, 1994) and can thus not be applied to diagnose the family. Accordingly, the characterisation of the Xystonellidae is unsatisfying without the signature characters. Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised eight, six, and five sequences, respectively, while the reference group "other Tintinnina" comprised 128, 72, and 66 sequences, respectively. Four asymmetric consensus signatures were detected in the ITS1-5.8S rRNA gene-ITS2 region and two binary and two asymmetric consensus signatures in the 28S rRNA gene (for character states in the reference group, see Supplementary Tables S2, S6 and Supplementary Data S1); no consensus signatures were found in the 18S rRNA gene.

Order Lynnellida Liu, Yi, Xu, Clamp, Li, Lin and Song, 2015

Remarks: Depending on the tree-building algorithm, the monotypic genus *Lynnella* (Fig. 1C) groups with the Oligotrichida or the Choreotrichida. Liu et al. (2015) displayed part of the alignment with those nucleotide positions marked that distinguish *Lynnella* from either the Oligotrichida, the Choreotrichida, or both simultaneously. Together with the description of the enigmatic species *Lynnella semiglobulosa*, a new genus and the family Lynnellidae Liu et al., 2011 had been established (Liu et al., 2011). Later, Liu et al. (2015) proposed also a new order.

Improved diagnosis: Oligotrichea with C-shaped adoral zone of membranelles. Adoral membranelles almost longitudinally arranged on collar, proximalmost ones elongated, extending into buccal cavity. Endoral membrane extends longitudinally on inner wall of buccal lip. Peristomial field perpendicular to main cell axis, rather flat. Somatic kineties extend longitudinally between adoral zone and posterior cell end. Somatic kinetids as ciliated monokinetids or dikinetids, each having a cilium associated only with the posterior basal body. Stomatogenesis in subsurface pouch, oral primordium perpendicular to the proter's main cell axis. 18S rRNA gene: 72 (A), 180 (A), 184 (G), 187 (A), 203 (C), 228 (G), 371 (G), 463 (C), 535 (A), 541 (C), 545 (T), 807 (T), 847 (T), 848 (C), 852 (C), 854 (T), 873 (C), 883 (G), 913 (T), 1079 (C), 1080 (T), 1084 (T), 1086 (A), 1090 (T), 1231 (C), 1239 (T), 1246 (C), 1261 (A), 1336 (G), 1482 (A), 1514 (G), 1586 (C), 1595 (G), 1716 (G), 1767 (C); ITS1-5.8S rRNA gene-ITS2 region: 136 (C), 146 (G), 164 (A), 168 (G), 169 (T), 172 (A), 173 (G), 177 (G), 183 (A), 197 (G), 283 (T), 298 (A), 303 (T), 306 (T), 307 (C), 309 (T), 311 (T), 350 (C), 363 (G), 368 (A), 374 (G), 431 (C), 433 (G), 522 (T), 536 (T), 542 (A), 544 (A), 547 (G), 548 (A), 551 (A); 28S rRNA gene: 27 (G), 37 (A), 38 (T), 44 (A), 45 (T), 54 (C), 63 (G), 74 (A), 75 (A), 93 (T), 106 (G), 107 (T), 127 (A), 128 (A), 161 (G), 162 (A), 182 (C), 226 (G), 234 (C), 246 (A), 253 (G), 291 (C), 363 (T), 369 (C), 373 (C), 374 (T), 377 (A), 378 (A), 384 (T), 391 (A), 392 (T), 394 (T), 399 (A), 509 (C), 510 (T), 512 (G), 513 (A), 527 (C), 528 (A), 536 (A), 615 (T), 616 (C), 618 (A), 630 (A), 631 (T), 639 (C), 670 (G), 671 (T), 672 (A), 673 (T), 674 (G), 684 (C), 687 (A), 688 (G), 691 (A), 701 (C), 702 (T), 703 (T), 704 (G), 705 (G), 716 (G), 721 (G), 722 (C), 732 (A), 734 (T), 749 (G), 781 (C), 789 (C), 791 (C), 792 (C), 793 (T).

Discussion: Lynnella semiglobulosa is somehow a chimera between the Oligotrichida and Choreotrichida in respect to morphological and molecular features (see Oligotrichida). The majority of morphological characteristics, however, suggest a closer relationship with the Choreotrichida, which is also reflected by the lower number of signature characters detected by comparing the two taxa (see 3.2.). Regarding the 18S rRNA gene, the ITS1-5.8S rRNA gene-ITS2 region, and the 28S rRNA gene, the query group comprised one sequence each, while the reference group Oligotrichida comprised 47, 16, and 14 sequences, respectively, and the reference group Choreotrichida comprised 153, 86, and 79 sequences, respectively. The Lynnellida differed from the Oligotrichida by 16 binary and 11 asymmetric consensus signatures in the 18S rRNA gene, 13 binary and 12 asymmetric consensus signatures in the ITS1-5.8S rRNA gene-ITS2 region, and 29 binary and 22 asymmetric consensus signatures in the 28S rRNA gene. The Lynnellida are distinguished from the Choreotrichida by 11 binary and nine asymmetric consensus signatures in the 18S rRNA gene, ten asymmetric consensus signatures in the ITS1-5.8S rRNA gene-ITS2 region, and ten binary and 29 asymmetric consensus signatures in the 28S rRNA gene. The comparison of the Lynnellida with the Oligotrichida plus Choreotrichida as reference group yielded six binary and six asymmetric consensus signatures in the 18S rRNA gene, five asymmetric consensus signatures in the ITS1-5.8S rRNA gene-ITS2 region, and five binary and six asymmetric consensus signatures in the 28S rRNA gene (for character states in the reference groups, see Supplementary Tables S2, S6 and Supplementary Data S1).

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107433.

References

Abraham, J.S., Somasundaram, S., Maurya, S., Gupta, R., Makhija, S., Toteja, R., 2021. Characterization of *Euplotes lynni* nov. spec., *E. indica* nov. spec. and description of *E. aediculatus* and *E. woodruffi* (Ciliophora, Euplotidae) using an integrative approach. Eur. J. Protistol. 79, 125779. https://doi.org/10.1016/j.ejop.2021.125779.

Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., Cárdenas, P., Čepička, I., Chistyakova, L., Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A.A., Hoppenrath, M., James, T.Y., Karnkowska, A., Karpov, S., Kim, E., Kolisko, M., Kudryavtsev, A., Lahr, D.J.G., Lara, E., Le Gall, L., Lynn, D.H., Mann, D.G., Massana, R., Mitchell, E.A.D., Morrow, C., Park, J.S., Pawlowski, J.W., Powell, M.J., Richter, D.J., Rueckert, S., Shadwick, L., Shimano, S., Spiegel, F.W., Torruella, G., Youssef, N., Zlatogursky, V., Zhang, Q., 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. J. Eukaryot. Microbiol. 66, 4–119. https://doi.org/10.1111/jeu.12691.

Aescht, E., 2001. Catalogue of the generic names of ciliates (Protozoa, Ciliophora). Denisia 1, 1–350.

Agatha, S., Strüder-Kypke, M.C., 2007. Phylogeny of the order Choreotrichida (Ciliophora, Spirotricha, Oligotrichea) as inferred from morphology, ultrastructure, ontogenesis, and SSrRNA gene sequences. Eur. J. Protistol. 43, 37–63. https://doi.org/10.1016/j.ejop.2006.10.001.

Agatha, S., Strüder-Kypke, M.C., 2012. Reconciling cladistic and genetic analyses in choreotrichid ciliates (Ciliophora, Spirotricha, Oligotrichea). J. Eukaryot. Microbiol. 59, 325–350. https://doi.org/10.1111/j.1550-7408.2012.00623.x.

Agatha, S., Strüder-Kypke, M.C., 2014. What morphology and molecules tell us about the evolution of Oligotrichea (Alveolata, Ciliophora). Acta Protozool. 53, 77–90. https://doi.org/10.4467/16890027ap.14.008.1445.

Agatha, S., Bartel, H., 2021. A comparative ultrastructural study of tintinnid loricae (Alveolata, Ciliophora, Spirotricha) and a hypothesis on their evolution. J. Eukaryot. Microbiol. 00, e12877. https://doi.org/10.1111/jeu.12877.

Agatha, S., Ganser, M.H., Santoferrara, L.F., 2021. The importance of type species and their correct identification: A key example from tintinnid ciliates (Alveolata,

Ciliophora, Spirotricha). J. Eukaryot. Microbiol. 68, e12865. https://doi.org/10.1111/jeu.12865.
Agatha, S., Gruber, M.S., Bartel, H., Weißenbacher, B., 2022. Somatic infraciliature in tintinnid ciliates (Alveolata, Ciliophora, Spirotricha): an ultrastructural comparison. J. Eukaryot. Microbiol. 00, e12885 https://doi.org/10.1111/jeu.12885.
Ahrens, D., Ahyong, S.T., Ballerio, A., Barclay, M.V.L., Eberle, J., Espeland, M., Huber, B. A., Mengual, X., Pacheco, T.L., Peters, R.S., Rulik, B., Vaz-de-Mello, F., Wesener, T., Krell, F.-T., 2021. Is it time to describe new species without diagnoses? - A comment on Sharkey et al. (2021). Zootaxa 5027, 151–159. https://doi.org/10.11646/zootaxa.5027.2.1.
Ali, R.H., Bogusz, M., Whelan, S., 2019. Identifying clusters of high confidence homologies in multiple sequence alignments. Mol. Biol. Evol. 36, 2340–2351. https://doi.org/10.1093/molbev/msz142.
Bachy, C., Gómez, F., López-García, P., Dolan, J.R., Moreira, D., 2012. Molecular phylogeny of tintinnid ciliates (Tintinnida, Ciliophora). Protist 163, 873–887. https://doi.org/10.1016/j.protis.2012.01.001.
Bai, Y., Wang, R., Al-Rasheid, K.A.S., Miao, M., Hu, X., 2020. The type species of *Amphorellopsis* and *Tintinnopsis* (Protozoa: Ciliophora): a new ciliary pattern and some comments in Tintinnina. J. King Saud Univ. Sci. 32, 3454–3462. https://doi.org/10.1016/j.jksus.2020.10.006.
Bardele, C.F., Stockmann, N., Agatha, S., 2018. Some ultrastructural features of the planktonic freshwater ciliate *Limnostrombidium viride* (Alveolata, Ciliophora, Oligotrichida) and improved diagnoses of oligotrich taxa. Acta Protozool. 57, 169–193. https://doi.org/10.4467/16890027AP.18.014.10090.
Brandt, K., 1906. Die Tintinnodeen der Plankton-Expedition. Tafelerklärungen nebst kurzer Diagnose der neuen Arten. Ergebn. Plankton-Exped. Humboldt-Stiftung 3 La, 1-33 + Plates I-LXX.
Brandt, K., 1907. Die Tintinnodeen der Plankton-Expedition. Systematischer Teil. Ergebn. Plankton-Exped. Humboldt-Stiftung 3 La, 1-488.
Burns, D.A., 1983. The distribution and morphology of tintinnids (ciliate protozoans) from the coastal waters around New Zealand. N. Z. J. Mar. Freshw. Res. 17, 387–406. https://doi.org/10.1080/00288330.1983.9516015.
Caisová, L., Marin, B., Melkonian, M., 2011. A close-up view on ITS2 evolution and speciation - a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). BMC Evol. Biol. 11, 262. https://doi.org/10.1186/1471-2148-11-262.
Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.
Chang, J.-M., Di Tommaso, P., Notredame, C., 2014. TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. Mol. Biol. Evol. 31, 1625–1637. https://doi.org/10.1093/molbev/msu117.
Chatzou, M., Magis, C., Chang, J.-M., Kemena, C., Bussotti, G., Erb, I., Notredame, C., 2016. Multiple sequence alignment modeling: methods and applications. Brief. Bioinformatics 17, 1009–1023. https://doi.org/10.1093/bib/bbv099.
Churchill, C.K.C., Valdés, A., Foighil, D.Ó., 2014. Molecular and morphological systematics of neustonic nudibranchs (Mollusca: Gastropoda: Glaucidae: *Glaucus*), with descriptions of three new cryptic species. Invertebr. Syst. 28, 174–195. https://doi.org/10.1071/Is13038.
Corliss, J.O., 1979. The Ciliated Protozoa. Characterization, Classification and Guide to the Literature. Pergamon Press, Oxford, New York.
da Silva Paiva, T., Küppers, G.C., da Silva-Neto, I.D., 2016. Morphology and divisional morphogenesis of the brackish water ciliate *Novistrombidium rufinoi* sp. nov. (Ciliophora: Oligotrichia) from Brazil. Rev. Brasil. Zoociê. 17, 20–32.
Dayrat, B., 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85, 407–415. https://doi.org/10.1111/j.1095-8312.2005.00503.x.
Delić, T., Trontelj, P., Rendoš, M., Fišer, C., 2017. The importance of naming cryptic species and the conservation of endemic subterranean amphipods. Sci. Rep. 7, 3391. https://doi.org/10.1038/s41598-017-02938-z.
Dewey, C.N., Pachter, L., 2006. Evolution at the nucleotide level: the problem of multiple whole-genome alignment. Hum. Mol. Genet. 15, R51–R56. https://doi.org/10.1093/hmg/ddl056.
Doris, S.M., Smith, D.R., Beamesderfer, J.N., Raphael, B.J., Nathanson, J.A., Gerbi, S.A., 2015. Universal and domain-specific sequences in 23S–28S ribosomal RNA identified by computational phylogenetics. RNA 21, 1719–1730. https://doi.org/10.1261/rna.051144.115.
Dubois, A., 2017. Diagnoses in zoological taxonomy and nomenclature. Bionomina 12, 63–85. https://doi.org/10.11646/bionomina.12.1.8.
Dunthorn, M., Klier, J., Bunge, J., Stoeck, T., 2012. Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity. J. Eukaryot. Microbiol. 59, 185–187. https://doi.org/10.1111/j.1550-7408.2011.00602.x.
Eberle, J., Ahrens, D., Mayer, C., Niehuis, O., Misof, B., 2020. A plea for standardized nuclear markers in metazoan DNA taxonomy. Trends Ecol. Evol. 35, 336–345. https://doi.org/10.1016/j.tree.2019.12.003.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.
Ezawa, K., 2016. Characterization of multiple sequence alignment errors using complete-likelihood score and position-shift map. BMC Bioinf. 17, 133. https://doi.org/10.1186/s12859-016-0945-5.
Fernandes, L.F., 2004. Tintininos (Ciliophora, Tintinnina) de águas subtropicais na região Sueste-Sul do Brasil. I. Famílias Codonellidae, Codonellopsidae, Coxliellidae, Cyttarocylidae, Epiplocylidae, Petalotrichidae, Ptychocylidae. Tintinnididae e Undellidae. Rev. Bras. Zool. 21, 551–576. https://doi.org/10.1590/S0101-81752004000300019.
Fol, H., 1881. Contribution a la connaissance de la famille des Tintinnodea. Archs Sci. phys. nat. 5, 5-24 + Plate I.
Ganser, M.H., Agatha, S., 2019. Redescription of *Antetintinnidium mucicola* (Claparède and Lachmann, 1858) nov. gen., nov. comb. (Alveolata, Ciliophora, Tintinnina). J. Eukaryot. Microbiol. 66, 802–820. https://doi.org/10.1111/jeu.12728.
Gemeinholzer, B., Vences, M., Beszteri, B., Bruy, T., Felden, J., Kostadinov, I., Miralles, A., Nattkemper, T.W., Printzen, C., Renz, J., Rybalka, N., Schuster, T., Weibulat, T., Wilke, T., Renner, S.S., 2020. Data storage and data re-use in taxonomy - the need for improved storage and accessibility of heterogeneous data. Org. Divers. Evol. 20, 1–8. https://doi.org/10.1007/s13127-019-00428-w.
Gerbi, S.A., 1996. Expansion segments: regions of variable size that interrupt the universal core secondary structure of ribosomal RNA. In: Zimmermann, R.A., Dahlberg, A.E. (Eds.), Ribosomal RNA: Structure, Evolution, Processing and Function in Protein Synthesis. Telford- CRC Press, Boca Raton, FL, pp. 71–87.
Gimmler, A., Stoeck, T., 2015. Mining environmental high-throughput sequence data sets to identify divergent amplicon clusters for phylogenetic reconstruction and morphotype visualization. Environ. Microbiol. Rep. 7, 679–686. https://doi.org/10.1111/1758-2229.12307.
Goldstein, P.Z., DeSalle, R., 2011. Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. BioEssays 33, 135–147. https://doi.org/10.1002/bies.201000036.
Grosse, M., Capa, M., Bakken, T., 2021. Describing the hidden species diversity of *Chaetozone* (Annelida, Cirratulidae) in the Norwegian Sea using morphological and molecular diagnostics. Zookeys 1039, 139–176. https://doi.org/10.3897/zookeys.1039.61098.
Gruber, M.S., Mühlthaler, A., Agatha, S., 2018. Ultrastructural studies on a model tintinnid - *Schmidingerella meunieri* (Kofoid and Campbell, 1929) Agatha and Strüder-Kypke, 2012 (Ciliophora). I. Somatic kinetids with unique ultrastructure. Acta Protozool. 57, 195–214. https://doi.org/10.4467/16890027AP.18.015.10091.
Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. https://doi.org/10.1093/sysbio/syq010.
Gutell, R.R., Larsen, N., Woese, C.R., 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol. Rev. 58, 10–26. https://doi.org/10.1128/mr.58.1.10-26.1994.
Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., Troedsson, C., Voolstra, C.R., 2014. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PLoS ONE 9, e87624. https://doi.org/10.1371/journal.pone.0087624.
Hassemer, G., Prado, J., Baldini, R.M., 2020. Diagnoses and descriptions in plant taxonomy: are we making proper use of them? Taxon 69, 1–4. https://doi.org/10.1002/tax.12200.
Hirt, R.P., Dyal, P.L., Wilkinson, M., Finlay, B.J., Roberts, D.M., Embley, T.M., 1995. Phylogenetic relationships among karyorelictids and heterotrichs inferred from small subunit rRNA sequences: resolution at the base of the ciliate tree. Mol. Phylogenet. Evol. 4, 77–87. https://doi.org/10.1006/mpev.1995.1008.
Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D.H., Soo, R.M., 2021. Prokaryotic taxonomy and nomenclature in the age of big sequence data. ISME J. 15, 1879–1892. https://doi.org/10.1038/s41396-021-00941-x.
Hütter, T., Ganser, M.H., Kocher, M., Halkic, M., Agatha, S., Augsten, N., 2020. DeSignate: detecting signature characters in gene sequence alignments for taxon diagnoses. BMC Bioinf. 21, 151. https://doi.org/10.1186/s12859-020-3498-6.
ICZN (International Commission of Zoological Nomenclature), 1999. International Code of Zoological Nomenclature. Fourth edition adopted by the International Union of Biological Sciences, International Trust for Zoological Nomenclature, London, UK.
Jamy, M., Foster, R., Barbera, P., Czech, L., Kozlov, A., Stamatakis, A., Bending, G., Hilton, S., Bass, D., Burki, F., 2020. Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. Mol. Ecol. Resour. 20, 429–443. https://doi.org/10.1111/1755-0998.13117.
Johnson, S.B., Warén, A., Tunnicliffe, V., Dover, C.V., Wheat, C.G., Schultz, T.F., Vrijenhoek, R.C., 2015. Molecular taxonomy and naming of five cryptic species of *Alviniconcha* snails (Gastropoda: Abyssochrysoidea) from hydrothermal vents. Syst. Biodivers. 13, 278–295. https://doi.org/10.1080/14772000.2014.970673.
Jörger, K.M., Schrödl, M., 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. Front. Zool. 10, 59. https://doi.org/10.1186/1742-9994-10-59.
Jung, J.-H., Kim, S., Ryu, S., Kim, M.-S., Baek, Y.-S., Kim, S.-J., Choi, J.-K., Park, J.-K., Min, G.-S., 2012. Development of single-nucleotide polymorphism-based phylum-specific PCR amplification technique: application to the community analysis using ciliates as a reference organism. Mol. Cells 34, 383–391. https://doi.org/10.1007/s10059-012-0169-0.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.
Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.
Kim, S.Y., Xu, D., Jung, J.-H., Choi, J.K., 2020. Phylogeny and genetic/morphological variation of *Strombidinopsis minima*-like species (Ciliophora: Choreotrichia). J. Eukaryot. Microbiol. 67, 115–124. https://doi.org/10.1111/jeu.12761.

Kofoid, C.A., Campbell, A.S., 1929. A conspectus of the marine and fresh-water Ciliata belonging to the suborder Tintinnoinea, with descriptions of new species principally from the Agassiz Expedition to the eastern tropical Pacific 1904–1905. Univ. Calif. Publs Zool. 34, 1–403.

Kofoid, C.A., Campbell, A.S., 1939. Reports on the scientific results of the expedition to the eastern tropical Pacific, in charge of Alexander Agassiz, by the U. S. Fish Commission Steamer "Albatross," from October, 1904, to March, 1905, Lieut.-Commander L. M. Garrett, U. S. N. Commanding. XXXVII. The Ciliata: The Tintinnoinea. Bull. Mus. comp. Zool. Harv. 84, 1-473 + Plates I-XXXVI.

Kohara, Y., Noda, H., Okano, K., Kambara, H., 2002. DNA probes on beads arrayed in a capillary, 'Bead-array', exhibited high hybridization performance. Nucleic Acids Res. 30, e87 https://doi.org/10.1093/nar/gnf086.

Kühn, A.L., Haase, M., 2020. QUIDDICH: QUick IDentification of DIagnostic CHaracters. J. Zool. Syst. Evol. Res. 58, 22–26. https://doi.org/10.1111/jzs.12347.

Küppers, G.C., da Silva Paiva, T., do Nascimento Borges, B., Alfaro, E.R., Claps, M.C., 2019. A new oligotrich (Ciliophora, Oligotrichia) from Argentina, with redefinition of *Novistrombidium* Song and Bradbury. Eur. J. Protistol. 69, 20–36. https://doi.org/10.1016/j.ejop.2019.02.006.

Landan, G., Graur, D., 2007. Heads or tails: a simple reliability check for multiple sequence alignments. Mol. Biol. Evol. 24, 1380–1383. https://doi.org/10.1093/molbev/msm060.

Laval, M., 1972. Ultrastructure de *Petalotricha ampulla* (Fol). Comparaison avec d'autres tintinnides et avec les autres ordres de ciliés. Protistologica 8, 369–386.

Laval-Peuto, M., 1994. Classe des Oligotrichea Bütschli, 1887. Ordre des Tintinnida Kofoid et Campbell, 1929. In: Puytorac, P. de (Ed.), Traité de Zoologie. Anatomie, systématique, biologie. 2. Infusoires ciliés. 2. Systématique. Masson, Paris, Milano, Barcelona, pp. 181-219.

Li, J., Liu, W., Gao, S., Warren, A., Song, W., 2013. Multigene-based analyses of the phylogenetic evolution of oligotrich ciliates, with consideration of the internal transcribed spacer 2 secondary structure of three systematically ambiguous genera. Eukaryot. Cell 12, 430–437. https://doi.org/10.1128/EC.00270-12.

Liu, W., Yi, Z., Lin, X., Al-Rasheid, K.A.S., 2011. Morphologic and molecular data suggest that *Lynnella semiglobulosa* n. g., n. sp. represents a new family within the subclass Choreotrichia (Ciliophora, Spirotrichea). J. Eukaryot. Microbiol. 58, 43–49. https://doi.org/10.1111/j.1550-7408.2010.00519.x.

Liu, W., Yi, Z., Xu, D., Clamp, J.C., Li, J., Lin, X., Song, W., Prentis, P., 2015. Two new genera of planktonic ciliates and insights into the evolution of the family Strombidiidae (Protista, Ciliophora, Oligotrichia). PLoS ONE 10, e0131726. https://doi.org/10.1371/journal.pone.0131726.

Lynn, D.H., 2008. The Ciliated Protozoa. Characterization, Classification, and Guide to the Literature. Springer, Dordrecht.

Marin, B., Melkonian, M., 2010. Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. Protist 161, 304–336. https://doi.org/10.1016/j.protis.2009.10.002.

Marin, B., Palm, A., Klingberg, M., Melkonian, M., 2003. Phylogeny and taxonomic revision of plastid-containg euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. Protist 154, 99–145. https://doi.org/10.1016/j.protis.2009.10.002.

Merckelbach, L.M., Borges, L.M.S., 2020. Make every species count: FASTACHAR software for rapid determination of molecular diagnostic characters to describe species. Mol. Ecol. Res. 20, 1761–1768. https://doi.org/10.1111/1755-0998.13222.

Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195. https://doi.org/10.1093/molbev/mst024.

Morrison, D.A., 2009. A framework for phylogenetic sequence alignment. Plant Syst. Evol. 282, 127–149. https://doi.org/10.1007/s00606-008-0072-5.

Morrison, D.A., Morgan, M.J., Kelchner, S.A., 2015. Molecular homology and multiple-sequence alignment: an analysis of concepts and practice. Aust. Syst. Bot. 28, 46–62. https://doi.org/10.1071/SB15001.

Needleman, S.B., Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453. https://doi.org/10.1016/0022-2836(70)90057-4.

Nickrent, D.L., Sargent, M.L., 1991. An overview of the secondary structure of the V4 region of eukaryotic small-subunit ribosomal RNA. Nucleic Acids Res. 19, 227–235. https://doi.org/10.1093/nar/19.2.227.

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217. https://doi.org/10.1006/jmbi.2000.4042.

Pais, F.-S.-M., de Cássia Ruy, P., Oliveira, G., Coimbra, R.S., 2014. Assessing the efficiency of multiple sequence alignment programs. Algorithms Mol. Biol. 9, 4. https://doi.org/10.1186/1748-7188-9-4.

Parapar, J., Capa, M., Nygren, A., Moreira, J., 2020. To name but a few: descriptions of five new species of *Terebellides* (Annelida, Trichobranchidae) from the North East Atlantic. Zookeys 992, 1–58. https://doi.org/10.3897/zookeys.992.55977.

Parker, C.T., Tindall, B.J., Garrity, G.M., 2019. International code of nomenclature of prokaryotes - Prokaryotic Code (2008 Revision). Int. J. Syst. Evol. Microbiol. 69, S1–S111. https://doi.org/10.1099/ijsem.0.000778.

Penn, O., Privman, E., Landan, G., Graur, D., Pupko, T., 2010. An alignment confidence score capturing robustness to guide tree uncertainty. Mol. Biol. Evol. 27, 1759–1767. https://doi.org/10.1093/molbev/msq066.

Petrov, A.S., Bernier, C.R., Gulen, B., Waterbury, C.C., Hershkovits, E., Hsiao, C., Harvey, S.C., Hud, N.V., Fox, G.E., Wartell, R.M., Williams, L.D., Barash, D., 2014. Secondary structures of rRNAs from all three domains of life. PLoS ONE 9, e88222. https://doi.org/10.1371/journal.pone.0088222.

Piwosz, K., Mukherjee, I., Salcher, M.M., Grujčić, V., Šimek, K., 2021. CARD-FISH in the sequencing era: opening a new universe of protistan ecology. Front. Microbiol. 12, 640066. https://doi.org/10.3389/fmicb.2021.640066.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Mol. Ecol. Notes 7, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x.

Rosling, A., Cox, F., Cruz-Martinez, K., Ihrmark, K., Grelet, G.-A., Lindahl, B.D., Menkis, A., James, T.Y., 2011. Archaeorhizomycetes: unearthing an ancient class of ubiquitous soil fungi. Science 333, 876-879. https://doi.org/doi:10.1126/science.1206958.

Ruhl, M.W., Wolf, M., Jenkins, T.M., 2010. Compensatory base changes illuminate morphologically difficult taxonomy. Mol. Phylogenet. Evol. 54, 664–669. https://doi.org/10.1016/j.ympev.2009.07.036.

Santoferrara, L.F., McManus, G.B., 2021. Diversity and biogeography as revealed by morphologies and DNA sequences - tintinnid ciliates as an example. In: Teodósio, M. A., Barbosa, A.B. (Eds.), Zooplankton Ecology. CRC Press, Boca Raton, pp. 85–118.

Santoferrara, L.F., McManus, G.B., Alder, V.A., 2013. Utility of genetic markers and morphology for species discrimination within the order Tintinnida (Ciliophora, Spirotrichea). Protist 164, 24–36. https://doi.org/10.1016/j.protis.2011.12.002.

Santoferrara, L.F., Bachy, C., Alder, V.A., Gong, J., Kim, Y.-O., Saccà, A., da Silva Neto, I. D., Strüder-Kypke, M.C., Warren, A., Xu, D., Yi, Z., Agatha, S., 2016. Updating biodiversity studies in loricate protists: the case of the tintinnids (Alveolata, Ciliophora, Spirotrichea). J. Eukaryot. Microbiol. 63, 651–656. https://doi.org/10.1111/jeu.12303.

Santoferrara, L.F., Alder, V.V., McManus, G.B., 2017. Phylogeny, classification and diversity of Choreotrichia and Oligotrichia (Ciliophora, Spirotrichea). Mol. Phylogenet. Evol. 112, 12–22. https://doi.org/10.1016/j.ympev.2017.03.010.

Santoferrara, L.F., Rubin, E., McManus, G.B., 2018. Global and local DNA (meta) barcoding reveal new biogeography patterns in tintinnid ciliates. J. Plankton Res. 40, 209–221. https://doi.org/10.1093/plankt/fby011.

Santoferrara, L., Burki, F., Filker, S., Logares, R., Dunthorn, M., McManus, G.B., 2020. Perspectives from ten years of protist studies by high-throughput metabarcoding. J. Eukaryot. Microbiol. 67, 612–622. https://doi.org/10.1111/jeu.12813.

Sarkar, I.N., Planet, P.J., DeSalle, R., 2008. CAOS software for use in character-based DNA barcoding. Mol. Ecol. Resour. 8, 1256–1259. https://doi.org/10.1111/j.1755-0998.2008.02235.x.

Satler, J.D., Carstens, B.C., Hedin, M., 2013. Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). Syst. Biol. 62, 805–823. https://doi.org/10.1093/sysbio/syt041.

Saville-Kent, W., 1881. A Manual of the Infusoria: Including a Description of all known Flagellate, Ciliate, and Tentaculiferous Protozoa, British and Foreign, and an Account of the Organization and Affinities of the Sponges. D. Bogue, London, pp. 473-720.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., Crozier, R.H., 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annu. Rev. Entomol. 55, 421–438. https://doi.org/10.1146/annurev-ento-112408-085432.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539. https://doi.org/10.1038/msb.2011.75.

Smith, S.A., Song, W., Gavrilova, N.A., Kurilov, A.V., Liu, W., McManus, G.B., Santoferrara, L.F., 2018. *Dartintinnus alderae* n. g., n. sp., a brackish water tintinnid (Ciliophora, Spirotrichea) with dual-ended lorica collapsibility. J. Eukaryot. Microbiol. 65, 400–411. https://doi.org/10.1111/jeu.12485.

Sniezek, J.H., Capriulo, G.M., Small, E.B., Russo, A., 1991. *Nolaclusilis hudsonicus* n. sp. (Nolaclusiliidae n. fam.) a bilaterally symmetrical tintinnine ciliate from the lower Hudson River estuary. J. Protozool. 38, 589–594.

Snyder, R.A., Brownlee, D.C., 1991. *Nolaclusilis bicornis* n. g., n.sp. (Tintinnina: Tintinnidiidae): a tintinnine ciliate with novel lorica and cell morphology from the Chesapeake Bay estuary. J. Protozool. 38, 583–589.

Sun, P., Xu, D., Clamp, J.C., Shin, M.K., 2012. Molecular and morphological characterization of a poorly known marine ciliate, *Myoschiston duplicatum* Precht, 1935: implications for phylogenetic relationships between three morphologically similar genera - *Zoothamnium*, *Myoschiston*, and *Zoothamnopsis* (Ciliophora, Peritrichia, Zoothamniidae). J. Eukaryot. Microbiol. 59, 163–170. https://doi.org/10.1111/j.1550-7408.2011.00609.x.

Sweeney, B.A., Hoksza, D., Nawrocki, E.P., Ribas, C.E., Madeira, F., Cannone, J.J., Gutell, R., Maddala, A., Meade, C.D., Williams, L.D., Petrov, A.S., Chan, P.P., Lowe, T.M., Finn, R.D., Petrov, A.I., 2021. R2DT is a framework for predicting and visualising RNA secondary structure using templates. Nat. Commun. 12, 3494. https://doi.org/10.1038/s41467-021-23555-5.

Tasneem, F., Shakoori, F.R., Ilyas, M., Shahzad, N., Potekhin, A., Shakoori, A.R., 2020. Genetic diversity of *Paramecium* species on the basis of multiple loci analysis and ITS secondary structure models. J. Cell. Biochem. 121, 3837–3853. https://doi.org/10.1002/jcb.29546.

Taylor, D.J., Devkota, B., Huang, A.D., Topf, M., Narayanan, E., Sali, A., Harvey, S.C., Frank, J., 2009. Comprehensive molecular structure of the eukaryotic ribosome. Structure 17, 1591–1604. https://doi.org/10.1016/j.str.2009.09.015.

Teixeira, M.A.L., Vieira, P.E., Pleijel, F., Sampieri, B.R., Ravara, A., Costa, F.O., Nygren, A., 2020. Molecular and morphometric analyses identify new lineages within a large *Eumida* (Annelida) species complex. Zool. Scr. 49, 222–235. https://doi.org/10.1111/zsc.12397.

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., Minh, B.Q., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44 (W1), W232–W235. https://doi.org/10.1093/nar/gkw256.

Turland, N.J., Wiersema, J.H., Barrie, F.R., Greuter, W., Hawksworth, D.L., Herendeen, P.S., Knapp, S., Kusber, W.-H., Li, D.-Z., Marhold, K., May, T.W., McNeill, J., Monro, A.M., Prado, J., Price, M.J., Smith, G.F., 2018. International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Regnum Vegetabile 159. Glashütten: Koeltz Botanical Books.

Wägele, J.W., Mayer, C., 2007. Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. BMC Evol. Biol. 7, 147. https://doi.org/10.1186/1471-2148-7-147.

Wang, P., Gao, F., Huang, J., Strüder-Kypke, M., Yi, Z., 2015. A case study to estimate the applicability of secondary structures of SSU-rRNA gene in taxonomy and phylogenetic analyses of ciliates. Zool. Scr. 44, 574–585. https://doi.org/10.1111/zsc.12122.

Wang, R., Song, W., Bai, Y., Warren, A., Li, L., Hu, X., 2020. Morphological redescriptions and neotypification of two poorly known tintinnine ciliates (Alveolata, Ciliophora, Tintinnina), with a phylogenetic investigation based on SSU rRNA gene sequences. Int. J. Syst. Evol. Microbiol. 70, 2515–2530. https://doi.org/10.1099/ijsem.0.004065.

Wang, R., Bai, Y., Hu, T., Xu, D., Suzuki, T., Hu, X., 2021a. Integrative taxonomy and molecular phylogeny of three poorly known tintinnine ciliates, with the establishment of a new genus (Protista; Ciliophora; Oligotrichea). BMC Ecol. Evol. 21, 115. https://doi.org/10.1186/s12862-021-01831-8.

Wang, R., Bai, Y., Hu, T., Xu, D., Suzuki, T., Hu, X., 2021b. Correction to: Integrative taxonomy and molecular phylogeny of three poorly known tintinnine ciliates, with the establishment of a new genus (Protista; Ciliophora; Oligotrichea). BMC Ecol. Evol. 21, 158. https://doi.org/10.1186/s12862-021-01875-w.

Wang, Y., Zhou, Q.-S., Qiao, H.-J., Zhang, A.-B., Yu, F., Wang, X.-B., Zhu, C.-D., Zhang, Y.-Z., 2016. Formal nomenclature and description of cryptic species of the *Encyrtus sasakii* complex (Hymenoptera: Encyrtidae). Sci. Rep. 6, 34372. https://doi.org/10.1038/srep34372.

Wang, Y., Wang, C., Jiang, Y., Katz, L.A., Gao, F., Yan, Y., 2019. Further analyses of variation of ribosome DNA copy number and polymorphism in ciliates provide insights relevant to studies of both molecular ecology and phylogeny. Sci. China Life Sci. 62, 203–214. https://doi.org/10.1007/s11427-018-9422-5.

Warren, A., Patterson, D.J., Dunthorn, M., Clamp, J.C., Achilles-Day, U.E.M., Aescht, E., Al-Farraj, S.A., Al-Quraishy, S., Al-Rasheid, K., Carr, M., Day, J.G., Dellinger, M., El-Serehy, H.A., Fan, Y., Gao, F., Gao, S., Gong, J., Gupta, R., Hu, X., Kamra, K., Langlois, G., Lin, X., Lipscomb, D., Lobban, C.S., Luporini, P., Lynn, D.H., Ma, H., Macek, M., Mackenzie-Dodds, J., Makhija, S., Mansergh, R.I., Martín-Cereceda, M., McMiller, N., Montagnes, D.J.S., Nikolaeva, S., Ong'ondo, G.O., Pérez-Uz, B., Purushothaman, J., Quintela-Alonso, P., Rotterová, J., Santoferrara, L., Shao, C., Shen, Z., Shi, X., Song, W., Stoeck, T., La Terza, A., Vallesi, A., Wang, M., Weisse, T., Wiackowski, K., Wu, L., Xu, K., Yi, Z., Zufall, R., Agatha, S., 2017. Beyond the "Code": a guide to the description and documentation of biodiversity in ciliated protists (Alveolata, Ciliophora). J. Eukaryot. Microbiol. 64, 539–554. https://doi.org/10.1111/jeu.12391.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54, 844–851. https://doi.org/10.1080/10635150500354878.

Woese, C.R., 1987. Bacterial evolution. Microbiol. Rev. 51, 221–271. https://doi.org/10.1128/mr.51.2.221-271.1987.

Zhang, Q., Yi, Z., Fan, X., Warren, A., Gong, J., Song, W., 2014. Further insights into the phylogeny of two ciliate classes Nassophorea and Prostomatea (Protista, Ciliophora). Mol. Phylogenet. Evol. 70, 162–170. https://doi.org/10.1016/j.ympev.2013.09.015.

Zhao, F., Filker, S., Xu, K., Li, J.u., Zhou, T., Huang, P., 2019. Effects of intragenomic polymorphism in the SSU rRNA gene on estimating marine microeukaryotic diversity: a test for ciliates using single-cell high-throughput DNA sequencing. Limnol. Oceanogr. Meth. 17, 533–543. https://doi.org/10.1002/lom3.10330.

Zielske, S., Haase, M., 2015. Molecular phylogeny and a modified approach of character-based barcoding refining the taxonomy of New Caledonian freshwater gastropods (Caenogastropoda, Truncatelloidea, Tateidae). Mol. Phylogenet. Evol. 89, 171–181. https://doi.org/10.1016/j.ympev.2015.04.020.