SCIENTIFIC REPORTS

natureresearch

OPEN

Deciphering phylogenetic relationships and delimiting species boundaries using a Bayesian coalescent approach in protists: A case study of the ciliate genus *Spirostomum* (Ciliophora, Heterotrichea)

Shahed Uddin Ahmed Shazib [1,6], Peter Vďačný [2,6], Marek Slovák[3,4], Eleni Gentekaki [5] & Mann Kyoon Shin [1]*

The ciliate genus *Spirostomum* comprises eight morphospecies, inhabiting diverse aquatic environments worldwide, where they can be used as water quality indicators. Although *Spirostomum* species are relatively easily identified using morphological methods, the previous nuclear rDNA-based phylogenies indicated several conflicts in morphospecies delineation. Moreover, the single locus phylogenies and previous analytical approaches could not unambiguously resolve phylogenetic relationships among *Spirostomum* morphospecies. Here, we attempt to investigate species boundaries and evolutionary history of *Spirostomum* taxa, using 166 new sequences from multiple populations employing one mitochondrial locus (CO1 gene) and two nuclear loci (rRNA operon and alpha-tubulin gene). In accordance with previous studies, relationships among the eight *Spirostomum* morphospecies were poorly supported statistically in individual gene trees. To overcome this problem, we utilised for the first time in ciliates the Bayesian coalescent approach, which accounts for ancestral polymorphisms, incomplete lineage sorting, and recombination. This strategy enabled us to robustly resolve deep relationships between *Spirostomum* species and to support the hypothesis that taxa with compact macronucleus and taxa with moniliform macronucleus each form a distinct lineage. Bayesian coalescent-based delimitation analyses strongly statistically supported the traditional morphospecies concept but also indicated that there are two *S. minus*-like cryptic species and *S. teres* is non-monophyletic. *Spirostomum teres* was very likely defined by a set of ancestral features of lineages that also gave rise to *S. yagiui* and *S. dharwarensis*. However, molecular data from type populations of the morphospecies *S. minus* and *S. teres* are required to unambiguously resolve the taxonomic problems.

Ciliates are common inhabitants of various environments worldwide. They are considered to be the top consumers of prokaryotic microbes in aquatic biofilms and they also play essential roles in different aspects of biological applications[1,2]. Since the discovery of the first ciliate *Paramecium* species by John Hill in 1752, more than 8,000 ciliate species have been described[3]. With the aid of advancements in molecular biology, the evolutionary history

[1]Department of Biological Science, University of Ulsan, Ulsan, 44610, South Korea. [2]Department of Zoology, Comenius University in Bratislava, 842 15, Bratislava, Slovakia. [3]Plant Science and Biodiversity Centre, Institute of Botany, Slovak Academy of Sciences, 845 23, Bratislava, Slovakia. [4]Department of Botany, Charles University, 128 01, Prague, Czech Republic. [5]School of Science, Mae Fah Luang University, Chiang Rai, 57100, Thailand. [6]These authors contributed equally: Shahed Uddin Ahmed Shazib and Peter Vďačný. *email: mkshin@ulsan.ac.kr

and relationships among ciliates have been studied at levels ranging from genera to classes. However, comparatively few studies have focused on integrated systematics at the species level and among closely related species[4–17].

The heterotrichean ciliate genus *Spirostomum* Ehrenberg, 1834 represents an appropriate model for studying ciliate evolution at the subgeneric level, due to the relatively easy morphological identification of its species. *Spirostomum* includes vermiform, large ciliates (150–4000 μm long) characterised by a long collecting canal of the contractile vacuole, extending from the posterior to the anterior body end along the dorsal cell side. The ciliary rows of *Spirostomum* become spiral during cell contraction. Another characteristic diagnostic trait of this genus includes a continuous paroral membrane distinctly thickened at its proximal end after silver staining[3,6,12,18–22]. The morphological taxonomy of *Spirostomum* is mostly based on body shape and size, macronuclear pattern (i.e., shape, size, and number of macronuclear nodules), and the number of ciliary rows and cortical granule rows between them[6,12,23]. *Spirostomum* ciliates are relatively common in fresh and brackish water environments[5,24,25], where they can be used as water quality indicators[23,26–29]. *Spirostomum* species are mostly found in microaerophilic or anaerobic conditions[23,30] and according to transcriptome analyses they are able to respire under anoxic conditions[31].

Recently, considerable research effort has focused on the morphology and molecular phylogeny of *Spirostomum* species[6,12,19,20,22,31–34]. Boscaro *et al.*[6] markedly improved *Spirostomum* systematics using both morphological and gene sequence data, but species boundaries were not unambiguously identified at the molecular level. Shazib *et al.*[12] used secondary structure information of the nuclear internal transcribed spacer 2 (ITS2) to delimit species boundaries, an approach commonly utilised for species identification and improvement of molecular phylogenetic reconstruction in various groups of organisms[35–43]. However, both primary and secondary sequence information merged several morphologically distinct species (i.e., *S. teres*, *S. yagiui* and *S. dharwarensis*) into a single group, and only a few relations were robustly resolved. All previous phylogenetic studies were limited to the nuclear rRNA gene sequences and suggested the presence of several cryptic species, especially, in *S. minus* and *S. teres*. Interrelationships among *Spirostomum* species were mostly left unclear.

In this study, 86 *Spirostomum* populations belonging to eight morphospecies were investigated, using three ribosomal markers and two protein-coding genes. This diverse dataset enabled us for the first time to explore the effectiveness of these molecular markers in species delimitation and to better understand evolutionary relationships among *Spirostomum* species. To complement the traditional tree building methods, the Bayesian coalescent approach and species network analysis were also utilised, which account for ancestral polymorphisms and incomplete lineage sorting. Finally, our aim was to compare the congruence of morphology and molecules in studying species boundaries.

Results

Sequence variation and genetic divergence analyses.
In total, 166 new *Spirostomum* gene sequences were obtained during the course of this study, including 32 new 18S rRNA gene sequences, 33 new ITS1-5.8S-ITS2 region sequences, 33 new D1D2 region sequences of the 28S rRNA gene, 31 new alpha-tubulin gene sequences, and 37 new mitochondrial cytochrome oxidase subunit 1 (CO1) gene sequences. In addition, one new alpha-tubulin and CO1 gene sequence from *Anigstenia* sp. were also obtained. All sequences were deposited in the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/nucleotide/). The corresponding GenBank accession numbers, length and GC content were summarised in Supplementary Table S1.

The intra- and inter-specific genetic distances are collated for all alignments in Supplementary Tables S2–20. To summarise, the 18S rRNA gene was the slowest evolving marker, while the mitochondrial CO1 gene was the fastest. The insert size of the CO1 barcode region was 285–288 nucleotides and 95–96 amino acids long. In the CO1 gene sequence alignment, one additional codon was commonly found in populations of *S. caudatum*, *S. minus*, *S. teres* and *S. yagiui* (SKS787), and was located in the insert region (Supplementary Fig. S1). The alpha-tubulin and CO1 gene sequences had greater mean genetic distances than the three rRNA gene sequences. *S. ambiguum* had low nucleotide variability in the rRNA locus and the alpha-tubulin gene, but markedly high diversity in the insert region of the CO1 gene. A similar pattern was also observed in *S. subtilis* and *S. yagiui*, but diversity in the CO1 gene was not so pronounced. The intra-specific genetic diversity in the morphospecies *S. minus* was comparatively higher in the ITS region than in the two other rRNA loci, but the mean intra-specific divergence was significantly lower in both protein coding genes, possibly due to the limited population sampling. The morphospecies *S. teres* showed markedly high mean intra-specific nucleotide diversities both in the rRNA locus and the CO1 gene.

Gene trees.
We used different datasets to investigate the effect of taxa sampling and masking on phylogenetic inferences. In almost all resultant phylogenetic trees (Figs 1–3), *Spirostomum* species were divided into two major clades. Clade 1 included *S. ambiguum*, *S. minus*, *S. semivirescens*, and *S. subtilis*, while clade 2 was comprised of *S. caudatum*, *S. dharwarensis*, *S. teres*, and *S. yagiui*. However, *S. semivirescens* and *S. dharwarensis* were present only in the 89 taxa dataset (Fig. 1). Single gene trees were, in general, congruent with concatenated trees, especially with regards to the well-supported monophyletic lineages. Although the branching pattern of phylogenetic trees inferred from the ribosomal locus (18S rRNA, ITS, 28S rRNA) and from the protein-coding gene sequences (alpha tubulin, CO1) was inconsistent, the conflicting topologies were not statistically supported. The most significant disagreement was the placement of the *S. ambiguum* + *S. subtilis* clade. However, its discordant phylogenetic position was also not statistically supported.

Spirostomum minus, *S. semivirescens*, *S. ambiguum* and *S. subtilis* formed a monophyletic group with variable support in the majority of trees (Figs 1–3). Their interrelationships were, however, inconsistent across analyses. *Spirostomum ambiguum* and *S. subtilis* were grouped in a sister position to *S. minus* in the concatenated trees, i.e. CON1-38 BI, ML trees (Fig. 2A), CON2-38 ML tree (Fig. 2B), CON1-32 BI, ML trees (Fig. 3F), CON2-32 BI, ML trees (Fig. 3G), and CON4-32 BI, ML trees (Fig. 3I), while they were separated from *S. minus* and grouped with
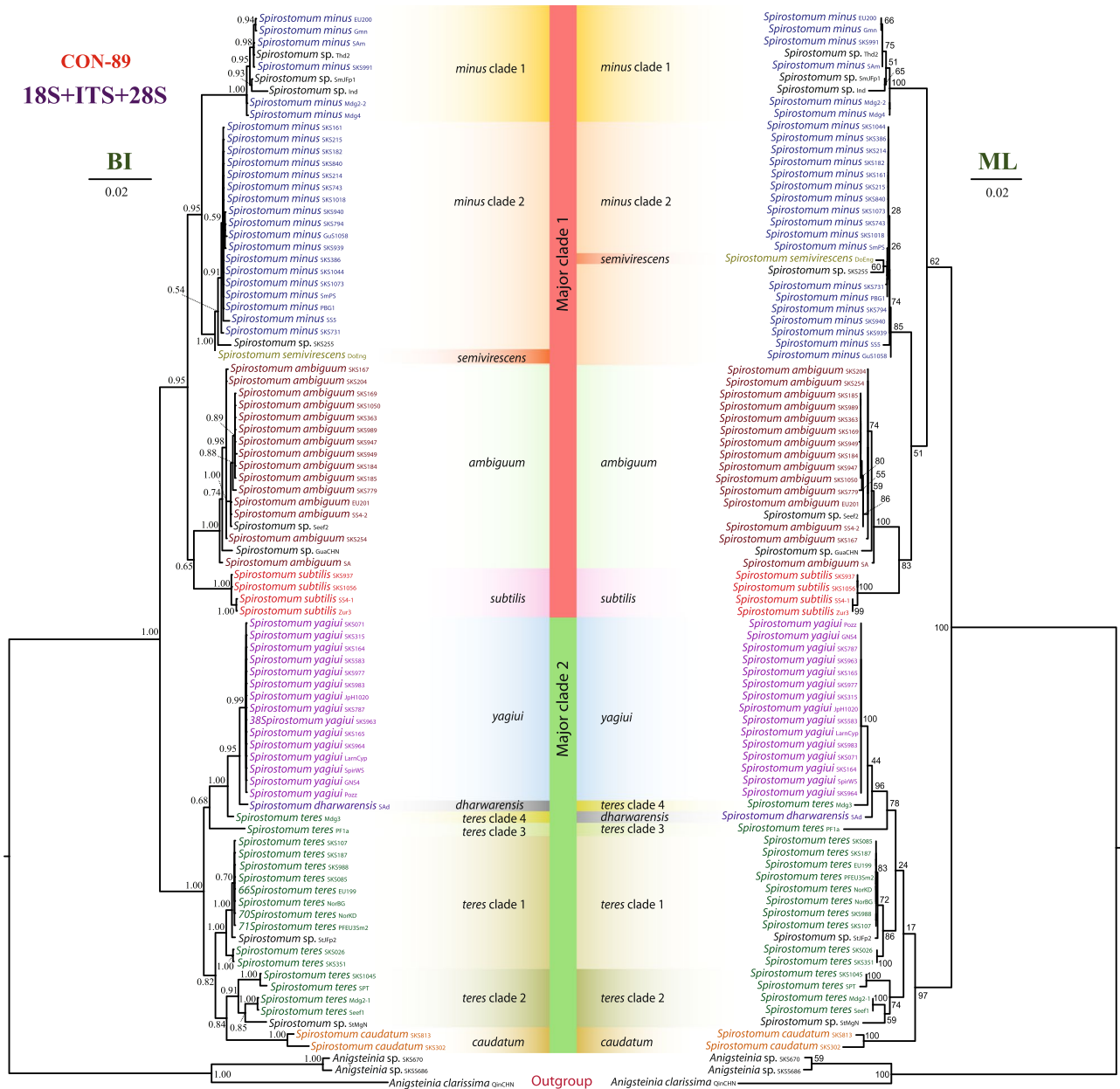
Figure 1. Phylogenetic relationships among 86 *Spirostomum* populations inferred from the 18S rRNA-ITS-28S rRNA concatenated dataset (CON-89). The Bayesian Inference (BI) tree is on the left and the Maximum Likelihood (ML) tree is on the right. Scale bars correspond to the number of nucleotide substitutions.

major clade 2 in the protein-coding gene trees (Fig. 3D,E,H). Nonetheless, this conflicting relationship received no statistical support. *Spirostomum minus* was not monophyletic and its populations formed two well supported clades in trees inferred from the 89 taxa dataset, whereby *S. semivirescens* was integrated into a clade along with *S. minus* clade 2 and *Spirostomum* sp. SKS255 in both BI and ML trees (Fig. 1). However, *S. minus* was recovered as a monophyletic lineage in trees based on the 32 and 38 taxa datasets (Figs 2 and 3), very likely due to the limited population sampling. *Spirostomum ambiguum* and *S. subtilis* were consistently grouped together in all concatenated analyses with low to strong support (Figs 1–3).

Spirostomum teres, *S. yagiui*, *S. caudatum*, and *S. dharwarensis* formed a strongly statistically supported clade in most concatenated analyses. However, their interrelationships varied depending on the molecular markers and phylogenetic technique used. This instability was also reflected in the weak statistical support for the majority of the clusters. The CON-89 dataset indicated that *S. caudatum* might have branched off first in major clade 2 (ML tree) but bootstrap support for this position was very weak (17%). In the BI tree, it was depicted as a sister taxon of the *S. teres* clade 2 (Fig. 1). Its position was also inconsistent in trees inferred from other datasets (Figs 2 and 3). *Spirostomum teres* remained non-monophyletic both in the single gene and the concatenated analyses based on four or more molecular markers (CON2-32 in Fig. 3G) but was recovered as a monophyletic lineage in the
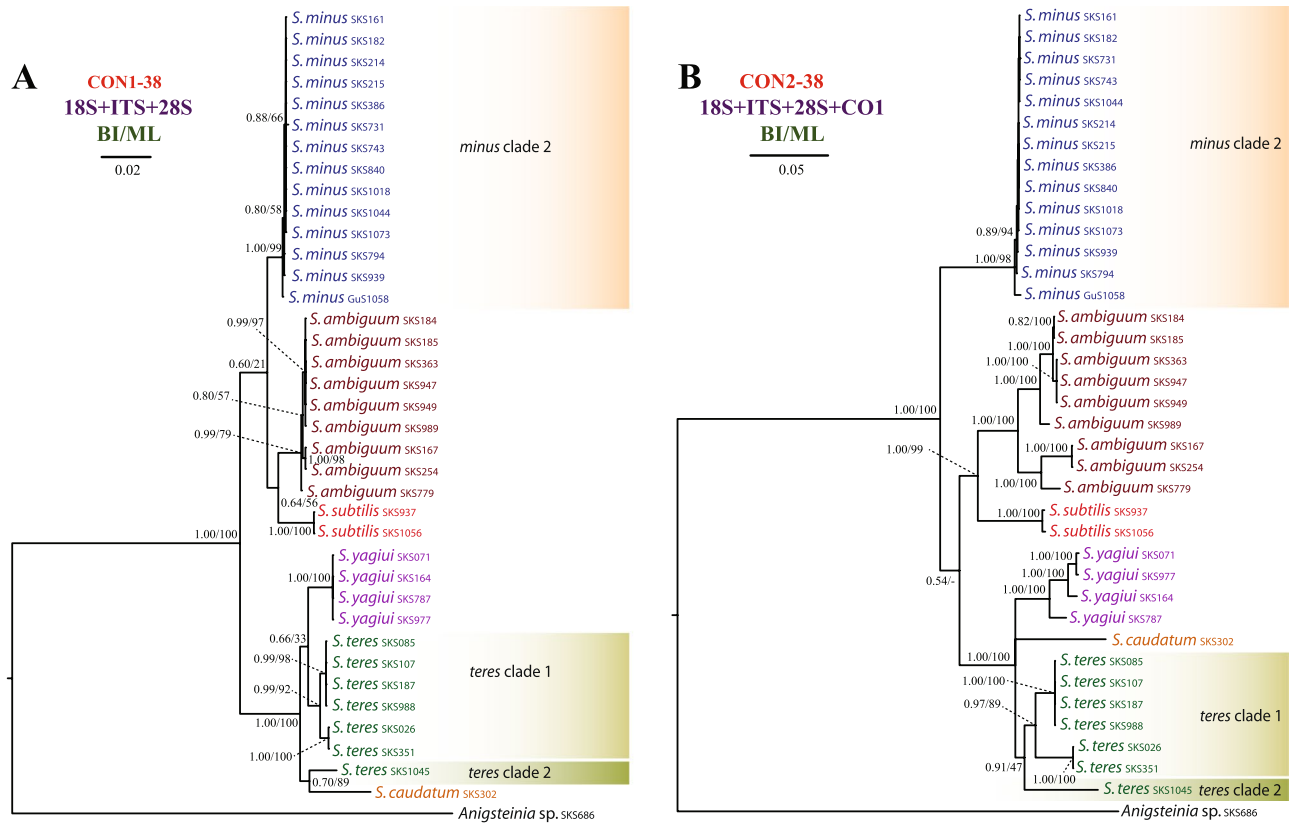
Figure 2. Phylogenetic relationships among 37 *Spirostomum* populations inferred from the CON1-38 and CON2-38 datasets. Results from the maximum likelihood (ML) bootstrap analyses were mapped onto the Bayesian Inference (BI) tree. Scale bars correspond to the number of nucleotide substitutions.

datasets CON2-38 and CON4-32 (Figs 2B and 3I) and the alpha-tubulin gene tree (Fig. 3D) with insignificant statistical support. This may be due to the limited population sampling.

All statistical tree topology tests did not reject the monophyletic origin of *Spirostomum* species having a moniliform macronucleus (i.e., clade 1) and monophyly of *Spirostomum* species with a compact macronucleus (i.e., clade 2). Regarding the CON-89 alignment, all statistical tests refuted monophyly of the four *S. teres* clades, but did not exclude monophyly of *S. teres* clades 1, 2, and 3. *Spirostomum teres* might be monophyletic when the single *S. teres* clade 4 population (Mdg3) is excluded. The monophyly of the two *S. minus* clades was firmly rejected by the AU test ($p = 0.036$) conducted on the concatenated 89 taxa alignment (Supplementary Table S21).

Species trees and multispecies coalescent analyses. Species trees were built from five datasets, as specified in Figs 4 and 5. All trees, except those inferred from the two protein-coding genes (Fig. 4D), consistently recognised two main clades within the genus *Spirostomum*. These clades matched those depicted in the majority of gene trees but they received high or full statistical support in multispecies coalescent analyses (cp. Figs 1–5). The branching pattern within the first main clade was robustly resolved in four and five marker trees: *S. minus* clade 2 branched off first and *S. ambiguum* and *S. subtilis* were sister taxa (Fig. 4). In species trees based on three markers, the two *S. minus* lineages did not group together but the *S. minus* clade 2 was placed in a sister position to the *S. semivirescens + Spirostomum* sp. SKS255 clade (Fig. 5), as also indicated in gene trees (Fig. 1). The grouping of *S. semivirescens + Spirostomum* sp. SKS255 was, however, left statistically unsupported in species trees. As concerns the branching pattern within the second main *Spirostomum* clade, results depended on the dataset analysed. Nevertheless, coalescent trees indicated that *S. caudatum* might have branched off first and *S. teres*, *S. yagiui*, and *S. dharwarensis* might cluster together. Relationships between *S. teres* clades 1–3 were left unresolved and, therefore, it cannot be excluded that they might belong to the same species (Fig. 5A). *Spirostomum teres* clade 4 was consistently classified with full statistical support in a clade along with *S. yagiui* and *S. dharwarensis* (Fig. 5A,B). This indicates that the *S. teres* clade 4 population (Mdg3) might represent a distinct species.

Altogether, six Bayesian species delimitation analyses were conducted, as specified in Table 1. Analyses based on the datasets CON2-38, CON2-32, and CON4-32 were fully consistent and recognised all lineages depicted in coalescent trees as distinct species with a posterior probability of 1.00. Thus, multiple molecular markers statistically corroborated very well the morphospecies concept of the genus *Spirostomum* and supported the validity of *S. ambiguum*, *S. subtilis*, *S. minus* (clade 2), *S. caudatum*, *S. teres* (clades 1 + 2), and *S. yagiui*. A similar result was obtained from the dataset CON3-32. The single exceptions were *S. caudatum* and *S. yagiui*, which were not statistically significantly delimited as distinct species, indicating alpha-tubulin and CO1 genes do not harbour enough phylogenetic signal to delimit *Spirostomum* species within the second main clade. With regard to the CON-89 dataset, two scenarios were considered
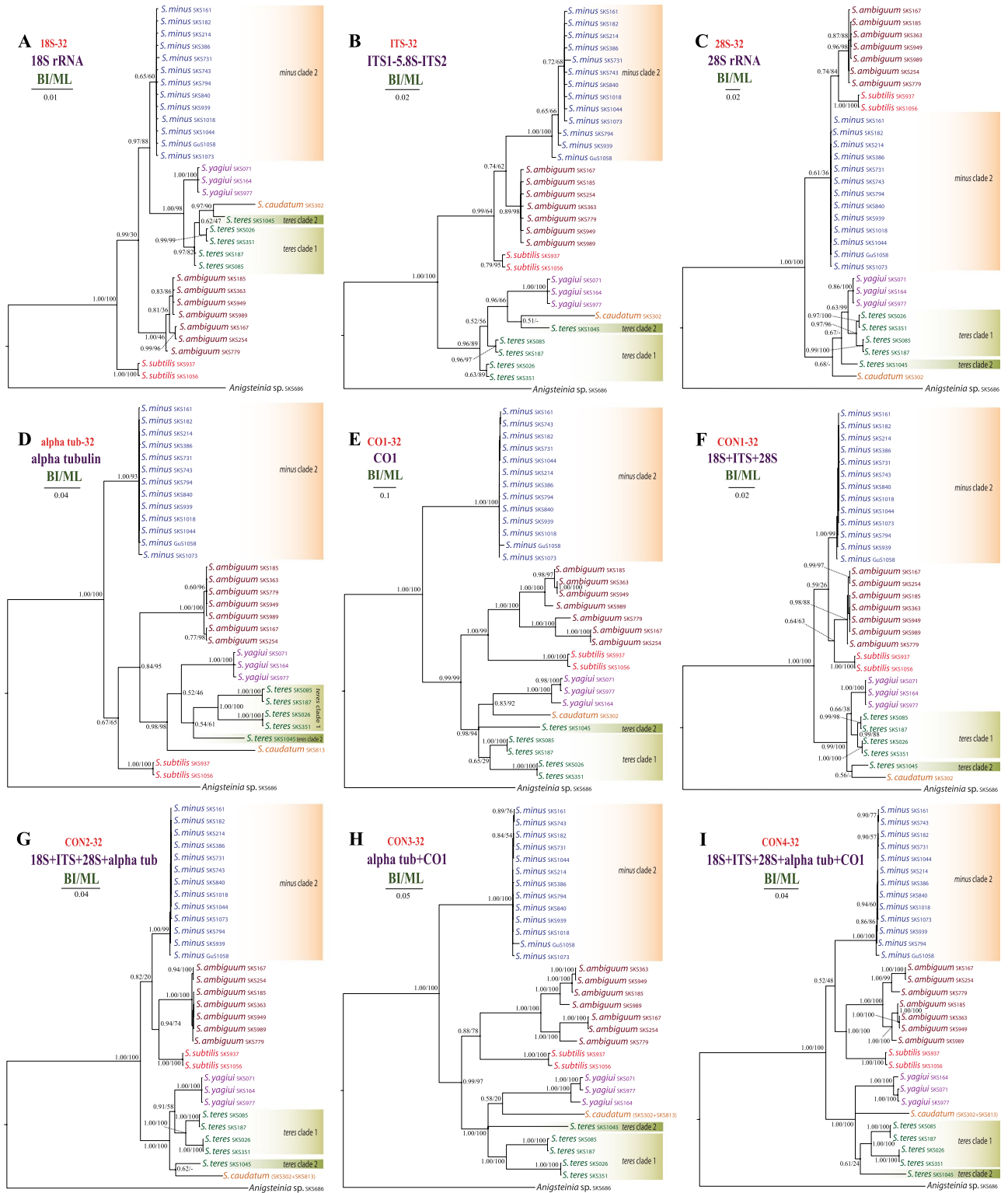
Figure 3. Phylogenetic relationships among 31 *Spirostomum* populations inferred from nine different datasets. Results from the maximum likelihood (ML) bootstrap analyses were mapped onto the Bayesian Inference (BI) tree. Scale bars correspond to the number of nucleotide substitutions.

given the results of statistical topology tests (Supplementary Table S21): altogether 15 lineages (i.e., four distinct clades were assumed in *S. teres*) were analysed in the first model and only 13 lineages (i.e., *S. teres* clades 1–3 were merged and only clade 4 was left separated) in the second model. Respective Bayesian delimitation analyses consistently recognised all assumed lineages, except for *Spirostomum* sp. SKS255 and *S. semivirescens* which were not statistically significantly
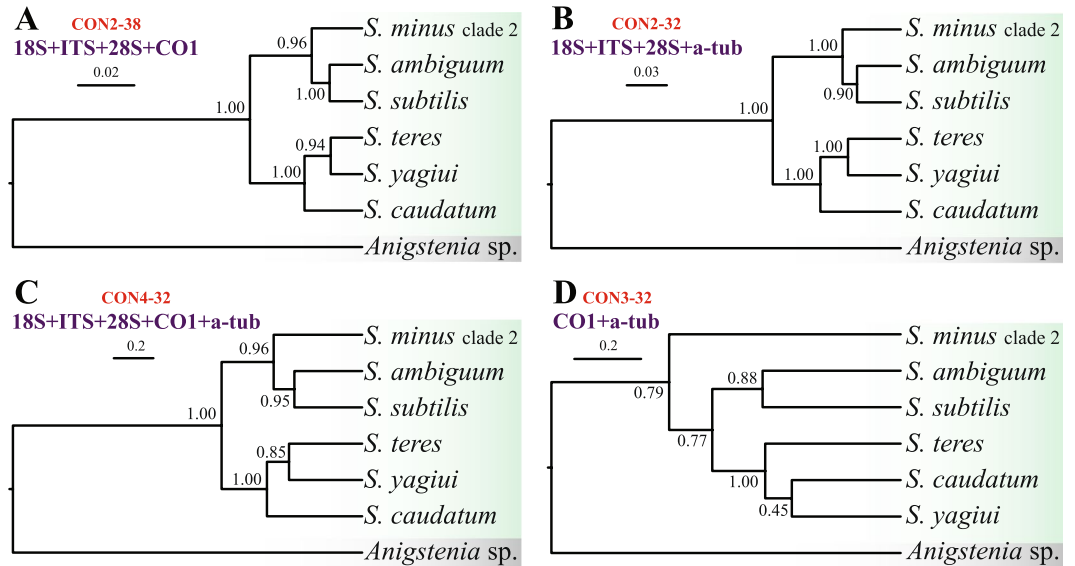
Figure 4. Phylogenetic relationships among six *Spirostomum* species inferred from four different datasets. Species trees were estimated using the Bayesian multispecies coalescent method. Values at nodes represent posterior probabilities. Scale bar corresponds to the number of nucleotide substitutions.
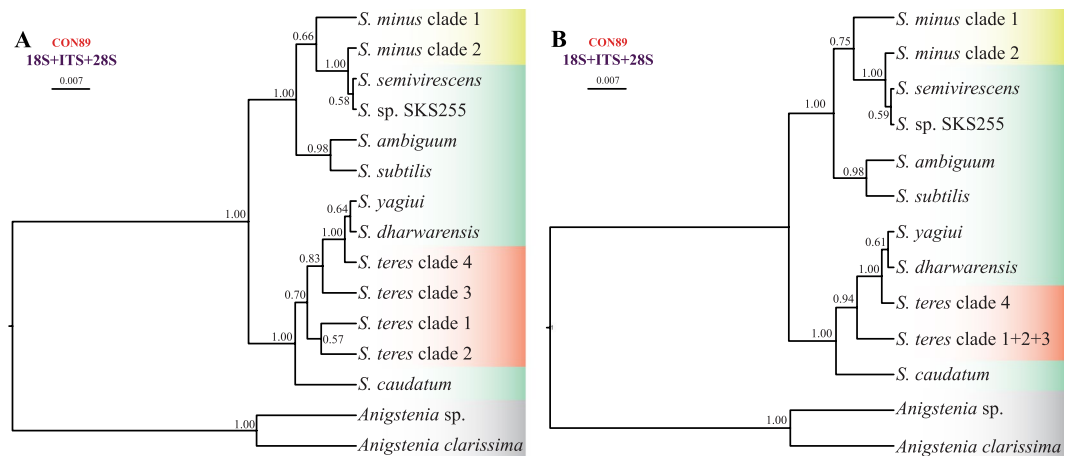

Figure 5. Phylogenetic relationships among *Spirostomum* clades inferred from the CON-89 dataset. Species trees were estimated using the Bayesian multispecies coalescent method. Two models were considered, with 15 (**A**) and 13 (**B**) separate lineages. Values at nodes represent posterior probabilities. Scale bar corresponds to the number of nucleotide substitutions.

delimited as distinct taxa, indicating they might be conspecific (Table 1). However, the distinctness of both *S. minus* clades was robustly supported and the morphospecies *S. teres* was depicted as non-monophyletic.

Species network analyses brought additional insights into conflicts between gene trees and poor statistical supports at some nodes. Networks computed with a maximum of zero reticulation nodes had topologies similar to the coalescent species trees (pseudo-likelihoods from −3188.94 to −3409.56), except for the position of *S. minus* (clade 2), which was depicted as the deepest branching species (Fig. 6A). The highest pseudo-likelihood (−2296.43) was obtained in an analysis limited to 10 reticulation nodes, whereby the resulting network had six reticulation nodes (Fig. 6B). The pattern of the species network most likely reflects a very deep incomplete lineage sorting, ancestral polymorphism, and/or population substructure along the branches of the phylogeny.

Selection tests. Branch-site Unrestricted Statistical Test for Episodic Diversification (BUSTED) was conducted to test for positive selection and to reveal the proportion of neutral, negative, and positively selected codon positions in the alpha-tubulin and CO1 genes of *Spirostomum* species. The null constrained model disallowing positive selection was not rejected either for the alpha-tubulin (likelihood ratio test LRT, *p*-value = 0.987) or the CO1 (LRT, *p*-value = 0.861) gene. Thus, both protein-coding genes very likely did not experience positive selection during the *Spirostomum* evolution.

Dataset	Prior	Posterior probability for number of delimited lineages	Conclusion
CON2-38	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 2000)	$P_7 = 1.00$	Morphospecies concept statistically supported
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 200)	$P_7 = 1.00$	Morphospecies concept statistically supported
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 2000)	$P_7 = 1.00$	Morphospecies concept statistically supported
	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 200)	$P_7 = 1.00$	Morphospecies concept statistically supported
CON2-32	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 2000)	$P_7 = 1.00$	Morphospecies concept statistically supported
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 200)	$P_7 = 1.00$	Morphospecies concept statistically supported
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 2000)	$P_7 = 1.00$	Morphospecies concept statistically supported
	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 200)	$P_7 = 1.00$	Morphospecies concept statistically supported
CON4-32	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 2000)	$P_7 = 1.00$	Morphospecies concept statistically supported
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 200)	$P_7 = 1.00$	Morphospecies concept statistically supported
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 2000)	$P_7 = 1.00$	Morphospecies concept statistically supported
	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 200)	$P_7 = 1.00$	Morphospecies concept statistically supported
CON3-32	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 2000)	$P_7 = 0.84$, $P_6 = 0.13$	*S. caudatum* and *S. yagiui* not statistically significantly delimited as distinct species
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 200)	$P_7 = 0.89$, $P_6 = 0.10$	*S. caudatum* and *S. yagiui* not statistically significantly delimited as distinct species
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 2000)	$P_7 = 0.78$, $P_6 = 0.20$	*S. caudatum* and *S. yagiui* not statistically significantly delimited as distinct species
	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 200)	$P_7 = 0.94$, $P_6 = 0.05$	*S. caudatum* and *S. yagiui* not statistically significantly delimited as distinct species
CON-89	Assuming 15 lineages		
	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 2000)	$P_{15} = 0.73$, $P_{14} = 0.27$	*Spirostomum* sp. SKS255 and *S. semivirescens* not statistically significantly delimited as distinct species
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 200)	$P_{15} = 0.44$, $P_{14} = 0.56$	*Spirostomum* sp. SKS255 and *S. semivirescens* not statistically significantly delimited as distinct species
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 2000)	$P_{15} = 0.68$, $P_{14} = 0.32$	*Spirostomum* sp. SKS255 and *S. semivirescens* not statistically significantly delimited as distinct species
	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 200)	$P_{15} = 0.58$, $P_{14} = 0.42$	*Spirostomum* sp. SKS255 and *S. semivirescens* not statistically significantly delimited as distinct species
CON-89	Assuming 13 lineages		
	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 2000)	$P_{13} = 0.71$, $P_{12} = 0.29$	*Spirostomum* sp. SKS255 and *S. semivirescens* not statistically significantly delimited as distinct species
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 200)	$P_{13} = 0.46$, $P_{12} = 0.54$	*Spirostomum* sp. SKS255 and *S. semivirescens* not statistically significantly delimited as distinct species
	$\theta \sim$ G(2, 100), $\tau \sim$ G(2, 2000)	$P_{13} = 0.63$, $P_{12} = 0.36$	*Spirostomum* sp. SKS255 and *S. semivirescens* not statistically significantly delimited as distinct species
	$\theta \sim$ G(2, 1000), $\tau \sim$ G(2, 200)	$P_{13} = 0.55$, $P_{12} = 0.45$	*Spirostomum* sp. SKS255 and *S. semivirescens* not statistically significantly delimited as distinct species

Table 1. Posterior probabilities for the number of delimited lineages/species, using different priors for model parameters. For the CON-89 dataset, two scenarios were considered given the results of statistical topology tests (Supplementary Table S21): altogether 15 lineages (i.e., four distinct clades were assumed in *S. teres*) were analysed in the first model, while only 13 lineages (i.e., *S. teres* clades 1–3 were merged and only clade 4 was left separated) in the second model. For species trees and lineages, see Figs 4 and 5.

The rate ratio of non-synonymous/synonymous changes $\omega = d_N/d_S$ was smaller than 0.001 in 99.88% of the codon positions, documenting very strong negative selection acting on alpha-tubulin. Only 0.12% of the codon positions in alpha-tubulin evolved neutrally. In the case of CO1, the ω ratio was below 0.02 in 85.77% and below 0.20 in 8.57% of the codon positions. Only 5.66% of the codon positions in CO1 evolved neutrally. According to the BUSTED analyses, a higher proportion of codon positions were under strong negative selection in the pre- and post-insert parts of CO1 than in the CO1 insert part (Fig. 7). Thus, only 0.29% of the codon positions evolved neutrally in the pre- and post-insert parts of CO1, while 9.93% of the codon positions evolved neutrally in the CO1 insert part. Strong negative selection might have generated signal that is different from species ancestry in both protein-coding genes during the *Spirostomum* evolution.

Discussion

Utility of genetic markers in species delimitation and molecular phylogeny.
In the present paper, we analysed the utility of five genetic markers in tree inferences and species delimitation: the mitochondrial cytochrome oxidase 1 (CO1) gene, the nuclear alpha-tubulin gene as well as the nuclear 18S rRNA gene, the ITS1-5.8S-ITS2 region, and the D1D2 domains of the 28S rRNA gene. These markers are commonly employed in reconstruction of phylogenetic relationships among closely related ciliate species[5–12,14–17,33,44–51]. Among the five genetic markers analysed, CO1 gene sequences showed the highest genetic divergence between morphospecies of the genus *Spirostomum* (Supplementary Table S2). However, rRNA gene sequences including the fast evolving ITS region and the D1D2 domains of the 28S rRNA gene also showed some degree of variability inside the same

morphospecies. This variability was suitable for discriminating individual species, since only rRNA genes brought results consistent with coalescent trees based on three, four, and five gene datasets as well as with morphological evolutionary scenarios. This suggests that *Spirostomum* species with a moniliform macronucleus cluster together and species with a compact macronucleus form a separated sister clade[12]. In contrast, trees inferred solely from individual protein-coding genes or their combination displayed different topologies, which might reflect purifying selection rather than speciation processes. Indeed, the BUSTED analyses revealed that strong purifying negative selection might have acted on the alpha-tubulin and CO1 gene (Fig. 7). Selection pressure might have generated signal that is different from species ancestry[52,53]. The significance of protein-coding genes in discriminating species boundaries was also questioned by the present Bayesian species delimitation based only on the alpha-tubulin and CO1 genes. In the coalescent species tree inferred solely from these two protein-coding genes, *S. yagiui* and *S. caudatum* were depicted as sister taxa and the model where their ancestral node is collapsed was favoured in Bayesian delimitation analyses. This result indicated that both species might represent a single species in the light of alpha-tubulin and CO1, which is in contradiction with all three, four, and five gene delimitation analyses and also with morphological data[12].

Evolutionary relationships among *Spirostomum* species.   Phylogenetic relationships among *Spirostomum* species using single or multiple gene trees have not been clearly resolved in previous analyses. Although most gene trees indicated a split of *Spirostomum* into two major clades, statistical support for these clades was usually poor[6,12,31,32]. On the other hand, the species trees presented herein, which were based on the coalescent model taking into account phenomena such as incomplete lineage sorting, ancestral polymorphisms and/or recombination[54], recovered both major clades with strong statistical support. This finding supports previous morphological evolutionary scenarios, suggesting that *Spirostomum* species with a moniliform macronucleus form a distinct clade that is sister to the clade of *Spirostomum* species with a compact macronucleus[12]. Specifically, all members of major clade 1 (*S. ambiguum*, *S. minus*, *S. semivirescens*, *S. subtilis*) possess a moniliform macronucleus, while species from major clade 2 (*S. caudatum*, *S. dharwarensis*, *S. teres* and *S. yagiui*) exhibit a compact ellipsoidal or elongated curved macronucleus.

The coalescent approach also helped to resolve phylogenetic relationships better within major clade 1 and supported the existence of one morphologically cryptic species that was identified as *S. minus* in the previous studies[6,12]. *Spirostomum minus* clade 2 is closely related to *S. semivirescens* which could not be unambiguously separated from *Spirostomum* sp. SKS255 in Bayesian species delimitation analyses. On the other hand, *S. ambiguum* and *S. subtilis* are consistently depicted as sister species. The latter two species also have large bodies, being 1,000–4,000 μm long in *S. ambiguum* and 700–1,000 μm long in *S. subtilis*. *Spirostomum minus* is much smaller, typically being only 300–400 μm long,[6,19,23,34] while *S. semivirescens* is typically 600–2000 μm long with numerous symbiotic algae in the cytoplasm[6,31,55].

Phylogenetic relationships within the second major clade are much more intricate, which might be due to the paucity of phylogenetic signal and/or presence of non-historical signal[56]. Our coalescent network analyses suggest a very deep incomplete lineage sorting in the genus *Spirostomum* (Fig. 6). Hybridization, another source of reticulation in species networks, is highly unlikely since hybrids of even closely related ciliate species are usually not viable[57]. The present coalescent analyses indicate that *S. caudatum* branches off first, *S. teres* is non-monophyletic, and *S. yagiui* and *S. dharwarensis* are sister taxa. The two latter species share an elongated macronucleus, while the two former morphospecies have an ellipsoidal macronucleus, which was very likely an ancestral condition in the main *Spirostomum* clade 2[12]. The present phylogenetic trees support the hypothesis of Boscaro *et al.*[6] that the morphospecies *S. teres* was very likely defined by a set of ancestral features of lineages that gave also origin to *S. yagiui* and *S. dharwarensis*.

Taxonomic implications.   Boscaro *et al.*[6] proposed eight valid *Spirostomum* morphospecies. Gene trees have indicated that *S. minus* and *S. teres* very likely represent species complexes and/or include multiple cryptic species[6,12,19]. According to the present Bayesian delimitation analyses, *S. minus* should indeed be split into two species, which supports the previous crypticity hypothesis. The mean genetic divergences among *S. minus* populations were 0.37%, 2.57%, and 1.91% in the 18S rRNA gene, ITS region, and D1D2 domains of the 28S rRNA gene, respectively (89 taxa dataset) (Supplementary Table S2). Very similar genetic distances have also been inferred for the ITS region by Shazib *et al.*[12]. The mitochondrial CO1 sequences suggested a mean nucleotide diversity of 0.44% (from 14 clones), which is lower than the mean nucleotide diversity in the ITS region from the 89 taxa dataset. However, this result needs to be taken with caution because CO1 sequences are not available for members of the *S. minus* clade 1. Boscaro *et al.*[6] could not find any key morphological characters that would enable reliable discrimination between the two *S. minus* clades. So far, the classification of *S. minus*-like specimens into one of the two clades is based only on molecular information. Although it would be appropriate to create a new formal name for one of the *S. minus* clades, sequences from populations originally studied by Roux[58] are needed to determine which clade corresponds to the "true" *S. minus*.

As concerns *S. teres*, both single and concatenated gene trees (except for trees inferred from the alpha tub-32 and the CON4-32 dataset; Fig. 3D,I) revealed this morphospecies to be non-monophyletic and its populations were consistently separated into two or four distinct clusters, depending on the dataset. Our previous results based on the primary and secondary structure of the ITS2 molecule could not support the existence of cryptic species in *S. teres*[12]. In this study, we analysed the genetic variability of the mitochondrial CO1 gene sequences at the population level. Our data suggested that the mean genetic diversity is significantly higher in the CO1 gene sequences (10.25% from 7 clones). Nevertheless, this is still below the 18% threshold that is considered as the mean intra-species variability in *Carchesium polypinum*[8]. The intra-specific genetic divergences in *S. teres* reach up to 20.81%, a value above that of *Carchesium* but distinctly below the 26% threshold in some arthropods[59]. Interestingly, about 19% sequence divergences were observed between two morphologically indistinguishable
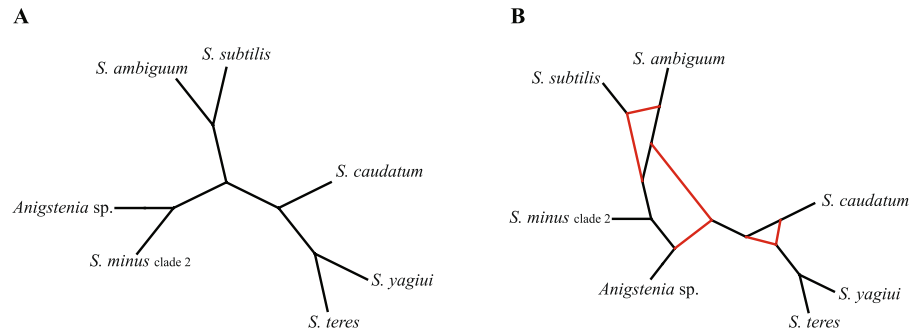
Figure 6. Species networks inferred from the 18S rRNA-ITS-28S rRNA, CO1 and alpha-tubulin Bayesian consensus trees based on the CON1-32, CO1-32, and alpha tub-32 datasets, respectively. Network retrieved with a maximum of zero (**A**) and ten (**B**) reticulation nodes allowed.
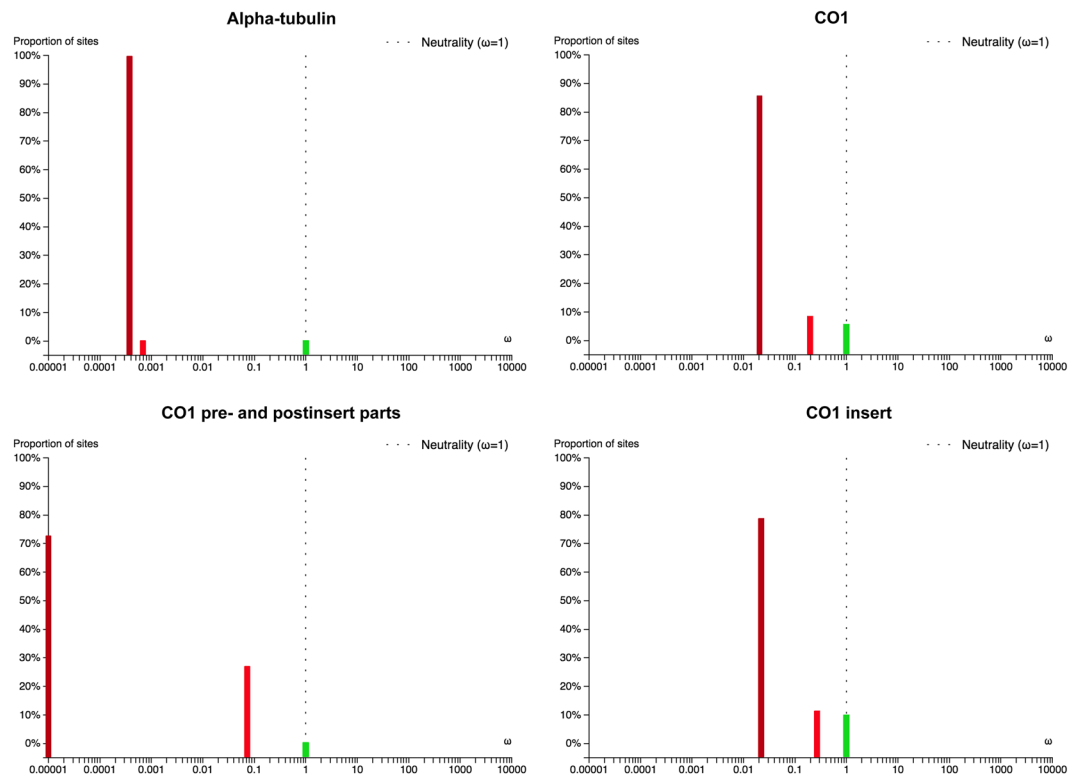

Figure 7. Proportion of sites belonging to three ω classes ($\omega_1 \leq \omega_2 \leq \omega_3$ and $\omega_3 = 1$) in alpha-tubulin and CO1 estimated with BUSTED.

populations of the millipede *Bicoxidens flavicollis*[60]. Nonetheless, the present statistical tree topology tests (Supplementary Table S21) and delimitation analyses (Table 1) suggest that at least the population Mdg3 of the *S. teres* morphospecies might represent a distinct taxon. The erection of a new species is, however, prevented by the lack of type material from the Mdg3 population. As in *S. minus*, molecular data from type population of *S. teres* are needed to determine which populations correspond to the "true" *S. teres*.

In all phylogenetic analyses, *S. ambiguum* populations were grouped together. Specifically, they formed an unstructured cluster in all gene trees, except for the CO1 gene tree where they were classified into two sister lineages (Fig. 3E). Furthermore, the mean genetic divergence among nine *S. ambiguum* isolates was 11.63%, with a maximum of 20.46%, and there were 167 polymorphic nucleotide sites (38 taxa) in the CO1 nucleotide sequences. Such a high intra-specific divergence in CO1 gene sequences is considered as a presence of species complexes or cryptic species[8,61]. However, the high divergence between *S. ambiguum* populations is concentrated mostly in the insert region of the CO1 gene. The quickly evolving ITS region sequences and D1D2 domains of the 28S rRNA gene do not indicate the presence of cryptic species within *S. ambiguum* (Figs 1, 2A, 3A–D,F,G). Therefore, we consider the nucleotide CO1 sequence variability to mostly reflect synonymous mutations in the insert region of

the CO1 gene and not speciation processes. Conjugation experiments and further neutral evolving markers are, however, needed to test if *S. ambiguum* might contain cryptic species or not.

Materials and Methods

Sampling, species identification and processing. Samples were collected from different habitats and localities, as summarised in Supplementary Table S1. Collected material was immediately stored in an icebox and brought to the laboratory, where individual ciliate morphospecies were isolated and used to set clonal cultures in Petri dishes. Each clonal culture started from a single individual that was washed in sterile distilled water. Cultures were maintained at 18–24 °C and contained filtrated original medium and/or sterile seawater for marine species and commercial mineral water (Evian, France) for freshwater species. Wheat grains were periodically added to stimulate the growth of prey bacteria.

After one week, specimens from each clonal culture were examined under an optical microscope Zeiss Axio Imager A1 at low (50–400X) and high (1,000X, oil immersion) magnifications, using bright field and differential interference contrast optics. The protargol staining method was used to reveal the ciliary pattern and nuclear apparatus[62]. Species identification was performed according to the following studies: Berger *et al.*[27], Boscaro *et al.*[6], Foissner *et al.*[23], Repak and Isquith[21] and Shazib *et al.*[12]. Main morphological characters of the studied *Spirostomum* species were summarised in Supplementary Fig. S2 and Table S22.

DNA extraction, PCR amplification and sequencing. After morphological identification, one or more cells from each identified population were isolated, washed several times in distilled water, and transferred into 10 μl extraction solution buffer (Sigma, St. Louis, MO, USA) in 1.5 ml microtubes. When more cells were collected from a population, they were kept separate and each cell represented a distinct sample. Subsequently, genomic DNA was extracted using the RED Extract-N-Amp Tissue PCR Kit (Sigma, St. Louis, MO), with modifications mentioned by Shazib *et al.*[12]. DNA amplifications were performed with polymerase chain reaction (PCR) using the TaKaRa Ex Taq polymerase kit, which has a higher fidelity than standard Taq polymerase with a mutation rate approximately 4.5 times lower (TaKaRa Bio-medicals, Otsu, Japan). Amplicons containing the 18S rRNA-ITS-28S rRNA region were obtained with the eukaryotic universal forward primer Euk A (5′-AAC CTG GTT GAT CCT GCC AG-3′)[63] and the reverse primer D1D2-R2 (5′-ACG ATC GAT TTG CAC GTC AG-3′)[64] under the PCR cycling conditions of Kim *et al.*[65]. The alpha tubulin gene sequences were amplified with the forward TUB-1 primer (5′-AAG GCT CTC TTG GCG TAC AT-3′) and the reverse TUB-2 primer (5′-TGA TGC CTT CAA CAC CTT CTT-3′)[66]. PCR conditions were as follows: 1 cycle for 5 min at 95 °C, 30 cycles for 1 min at 94 °C, 2 min at 60 °C and 2 min at 72 °C and 1 cycle for 10 min at 72 °C. The mitochondrial CO1 gene sequences were amplified using the forward F388dT primer (5′-TGT AAA ACG ACG GCC AGT GGW KCB AAA GAT GTW GC-3′) and the reverse R1184dT primer (5′-CAG GAA ACA GCT ATG ACT ADA CYT CAG GGT GAC CRA AAA TC A-3′)[61]. We also slightly modified the CO1 primers as follows: 5′-GGN KCN AAA GAT GTW GC-3′ for the forward CO1-F388dT17 primer and 5′-CAG GGT GAC CGA AAA ATC-3′ for the reverse CO1-R1184dT18 primer by aligning all available *Spirostomum* CO1 gene sequences and checking their suitability with polymerase chain reaction. The condition of the PCR cycles for both sets of primers were: 1 cycle for 4 min at 94 °C, 40 cycles for 45 s at 94 °C, 75 s at 47 °C and 90 s at 72 °C and 1 final extension cycle for 10 min at 72 °C. The size of the amplified DNA was confirmed by electrophoresing in 1.2% agarose gel and 1X TAE buffer at 80 V for 50–60 min. PCR products were visualised with the SYBR Green I nucleic acid gel stain (Sigma Aldrich) and UV transillumination. Finally, PCR products were purified and bi-directional sequenced with PCR primers on an ABI 3730 automatic sequencer (Macrogen Inc., Seoul, South Korea). For ribosomal gene sequences, five additional internal primers were used as specified in Shazib *et al.*[12].

Sequence processing, datasets and alignment procedures. Sequencing chromatogram files were checked, trimmed, and assembled into contigs using the software Geneious ver. 8.1.7[67] (http://www.geneious.com). The protein-coding genes were translated into amino acid sequences in Geneious with the ciliate nuclear genetic code for the alpha-tubulin gene and with the protozoan mitochondrial code for the CO1 gene to check for stop codons and frame shifts. Several datasets were constructed to examine the impact of taxonomic sampling on phylogenetic analyses (Table 2). Ribosomal gene sequences were aligned using the MAFFT algorithm and 100 bootstrap repeats on the online server GUIDANCE2[68] (http://guidance.tau.ac.il/ver2/). The protein-coding genes were aligned based on the predicted amino acid sequences with MEGA ver. 6.06[69]. Unreliable and poorly aligned columns were removed from the final rRNA gene sequence alignments according to the calculated confidence scores suggested by the GUIDANCE2 algorithm. No masking strategy was employed for the alpha-tubulin and CO1 gene sequences, as all columns were aligned unambiguously. The GC content of each sequence was calculated in Geneious. Numbers of parsimony informative (Pi) sites were estimated from each alignment using the software PAUP* ver. 4.0b10[70].

Distance analyses. Intraspecific as well as interspecific pairwise uncorrected *p*-distances and numbers of nucleotide differences were calculated separately in MEGA ver. 6.06[69]. All alignment positions with gaps were excluded from the distance analysis, using the complete and/or partial deletion option.

Construction of gene trees. We conducted Bayesian Inference (BI) and Maximum Likelihood (ML) analyses on all alignments. The Akaike information criterion (AIC) calculated in jModelTest ver. 2.0.1[71,72] was used to evaluate nucleotide substitution models of evolution for each dataset. The best fitting evolutionary models for all alignments are summarised in Table 2. Bayesian analyses were performed in MrBayes ver. 3.2.6[73] using the best evolutionary model, whereby MCMC chains were one million steps long and every 100th generation was sampled.

Dataset	No. of taxa	Molecular marker(s)	No. of char.	No. of PI char.	Evolutionary substitution model used in Bayesian analyses
18S-32	32	18S rRNA gene (=18S)	1580	63	GTR + I (=0.7750) + G (=0.6340)
ITS-32	32	ITS1-5.8S-ITS2 region (=ITS)	256	43	GTR + G (=0.1940)
28S-32	32	D1D2 domain of 28S rRNA gene (=28S)	495	50	GTR + I (=0.4570) + G (=0.4760)
alpha tub-32	32	Alpha-tubulin	936	203	GTR + I (=0.2560) + G (=0.2110)
CO1-32	32	Cytochrome oxidase subunit 1 (=CO1)	644	306	GTR + I (=0.3430) + G (=0.7290)
CON1-32	32	18S + ITS + 28S	2331	156	GTR + I (=0.6840) + G (=0.4610)
CON2-32	32	18S + ITS + 28S + alpha-tubulin	3267	359	GTR + I (=0.6180) + G (=0.3880)
CON3-32	32	Alpha-tubulin + CO1	1580	509	GTR + I (=0.5600) + G (2.1390)
CON4-32	32	18S + ITS + 28S + alpha-tubulin + CO1	3911	665	GTR + I (=0.5950) + G (=0.5640)
CON1-38	38	18S + ITS + 28S	2317	154	GTR + I (=0.6900) + G (=0.4635)
CON2-38	38	18S + ITS + 28S + CO1	2887	439	GTR + I (=0.6130) + G (=0.4930)
CON-89	89	18S + ITS + 28S	2428	333	GTR + I (=0.6850) + G (=0.5670)

Table 2. Characterisation of the datasets analysed. The number of parsimony informative characters (PI char.) was calculated in PAUP* and the best evolutionary substitution model was selected in jModelTest under the Akaike information criterion.

The first 2,500 sampled trees were discarded as burn-in, leaving 7,500 trees for calculating majority rule consensus trees and posterior probabilities of their branching patterns. ML analyses with 1,000 bootstrap replicates were carried out using RAxML-HPC2 ver. 8.2.10 on the CIPRES Science Gateway ver. 3.3 (http://www.phylo.org/index.php/portal/v33) with the GTRCAT evolutionary model to account for heterogeneity rate[74–77]. Phylogenetic trees were visualised and edited using FigTree ver. 1.4 (http://tree.bio.ed.ac.uk/software/figtree/) and MEGA. In all datasets, *Anigstenia* species were considered as the outgroup taxa for rooting the trees.

Construction of species trees, networks and Bayesian species delimitation. Species trees were calculated under the Bayesian multispecies coalescent model, using STACEY ver. 1.2.2[78] implemented in the computer package BEAST ver. 2.4.5[79]. Input files were prepared in BEAUti with the following settings: (i) best evolutionary substitution models as selected by jModelTest for each partition; (ii) four categories for substitution rate heterogeneity; (iii) uncorrelated lognormal clock; (iv) ploidy scalars at 1.0 for the mitochondrial partition and 2.0 for the nuclear partitions; (v) the Yule process model for the species tree prior; and (vi) 200 million generations and a sampling frequency of 20,000 in Markov Chain Monte Carlo analyses. In total, six different datasets were analysed, as detailed in the Results section. The convergence to stationary distribution (effective sample size >200 for all parameters) was checked in Tracer ver. 1.6 for all analyses. The maximum clade credibility trees were summarised in TreeAnnotator ver. 1.8.1[80] after discarding the first 10% of sampled trees.

A species network was constructed in PhyloNet ver. 3.6.1[81,82], using the maximum pseudo-likelihood framework. The network was computed with 0, 5, and 10 maximum numbers of reticulation nodes from the 18S + ITS + 28S, CO1 and alpha-tubulin Bayesian consensus trees based on the CON1-32, CO1-32, and alpha tub-32 datasets, respectively. The maximum number of reticulation nodes was determined iteratively. Zero reticulations represented a null model corresponding to a species tree, while a maximum of either 5 or 10 reticulations served to test how many reticulations could be present in the resulting phylogenetic networks. If there were exactly 10 reticulation nodes, another round of analyses with an increased number of reticulations would be needed. However, phylogenetic networks with a maximum of six reticulations were recovered (see the Results section); hence, no further analyses were needed. Each analysis was performed with 10 runs and default settings, generating five optimal networks. The species networks were visualised with Dendroscope ver. 2.7.4[83].

Bayesian species delimitation was conducted in BP&P ver. 2.2[84], with the same datasets as used in the construction of species trees. Coalescent species trees obtained with STACEY served as guide trees for species delimitation. Each species delimitation model was assigned equal prior probability. Four different combinations of prior settings for the ancestral population size (θ) and root age (τ) were tested to examine the robustness of the results: relatively small ancestral population size and shallow divergences ($\theta = G[2, 1000]$, $\tau = G[2, 2000]$), relatively large ancestral population size and deep divergences ($\theta = G[2, 100]$, $\tau = G[2, 200]$), relatively large ancestral population size and shallow divergences ($\theta = G[2, 100]$, $\tau = G[2, 2000]$), and relatively small ancestral population size and deep divergences ($\theta = G[2, 1000]$, $\tau = G[2, 200]$)[85]. The rjMCMC analyses were run for 100,000 generations with a sampling frequency of 2 and a burn-in of 10,000. A large fine-tuning parameter ($\varepsilon = 15$) was used to guarantee a good mixing in the reversible jump algorithm[84]. All analyses were conducted twice to confirm consistency between runs.

Statistical tree topology tests. Topology tests were carried out to assess the monophyletic origins of (1) *Spirostomum* species with moniliform macronucleus, (2) *Spirostomum* species with compact macronucleus, (3) two *S. minus* clades, and (4) four *S. teres* clades. The approximately unbiased (AU), the weighted Kishino-Hasegawa (WKH) and the weighted Shimodaira-Hasegawa (WSH) test were conducted, as implemented in CONSEL ver. 0.1[86–88]. The unconstrained and constrained ML trees, and their site-wise likelihoods were calculated in PAUP* ver. 4.0b10[70] under the best evolutionary models using the ML criterion, heuristic search, TBR branch swapping and 10 random sequence addition replications.

Selection tests. Gene-wide tests for positive selection acting on the alpha-tubulin and CO1 genes were performed with BUSTED (Branch-site Unrestricted Statistical Test for Episodic Diversification) on the Datamonkey Adaptive Evolution Server[89–91]. BUSTED simultaneously estimates the proportion of sites belonging to each of three ω classes. It holds $\omega_1 \leq \omega_2 \leq 1 \leq \omega_3$ in the unconstrained model, while $\omega_3 = 1$ in the constrained null model disallowing positive selection. If the null hypothesis is rejected, then there is evidence that at least one site has, at least some of the time, experienced positive selection[92].

Conclusions

Our analyses strongly statistically supported the following previous hypotheses about the genus *Spirostomum*: (1) taxa with compact macronucleus and taxa with moniliform macronucleus each form a distinct lineage; (2) the morphospecies *S. minus* contains two morphologically cryptic taxa; and (3) the morphospecies *S. teres* is non-monophyletic and defined by a set of ancestral features of lineages that also gave origin to *S. yagiui* and *S. dharwarensis*. Our analyses further revealed that ribosomal RNA genes and their spacers bear phylogenetic signal – which is consistent with species trees – and therefore have the highest phylogenetic informativeness and delimitation power in *Spirostomum*. On the other hand, the protein-coding CO1 and alpha-tubulin genes are useful in population structure analyses but might not have power to resolve phylogenetic relationships among *Spirostomum* species, possibly due to the purifying selection. The problem of purifying selection is especially pronounced in the gene coding for alpha-tubulin whose usage is therefore not recommended in species delimitation analyses of the genus *Spirostomum*, although this gene is highly variable.

Data availability

Accession numbers for the newly obtained DNA sequences are 18S rRNA (MK688522 – MK688553); ITS1-5.8S-ITS2 (MK721433 – MK721465); 28S rRNA™(MK713375 – MK713407); alpha-tubulin (MK721466 – MK721497); CO1 (MK721498 – MK721535) listed in Supplementary Table S1 and are available at https://www.ncbi.nlm.nih.gov/nucleotide/. Results of all analyses are included in this published article and its Supplementary Information files. The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Dopheide, A., Lear, G., Stott, R. & Lewis, G. Molecular characterization of ciliate diversity in stream biofilms. *Appl. Environ. Microbiol.* **74**, 1740–1747, https://doi.org/10.1128/AEM.01438-07 (2008).
2. Hakenkamp, C. C. & Morin, A. The importance of meiofauna to lotic ecosystem functioning. *Freshw. Biol.* **44**, 165–175, https://doi.org/10.1046/j.1365-2427.2000.00589.x (2000).
3. Lynn, D. H. *The Ciliated Protozoa. Characterization, Classification, and Guide to the Literature.* 3rd edn (Springer Verlag, 2008).
4. Finlay, B. J., Esteban, G. F., Brown, S., Fenchel, T. & Hoef-Emden, K. Multiple cosmopolitan ecotypes within a microbial eukaryote morphospecies. *Protist* **157**, 377–390, https://doi.org/10.1016/j.protis.2006.05.012 (2006).
5. Boscaro, V., Fokin, S. I., Verni, F. & Petroni, G. Survey of *Paramecium duboscqui* using three markers and assessment of the molecular variability in the genus *Paramecium*. *Mol. Phylogenet. Evol.* **65**, 1004–1013, https://doi.org/10.1016/j.ympev.2012.09.001 (2012).
6. Boscaro, V. *et al.* Focusing on genera to improve species identification: Revised systematics of the ciliate *Spirostomum*. *Protist* **165**, 527–541, https://doi.org/10.1016/j.protis.2014.05.004 (2014).
7. Chantangsi, C. & Lynn, D. H. Phylogenetic relationships within the genus *Tetrahymena* inferred from the cytochrome c oxidase subunit 1 and the small subunit ribosomal RNA genes. *Mol. Phylogenet. Evol.* **49**, 979–987, https://doi.org/10.1016/j.ympev.2008.09.017 (2008).
8. Gentekaki, E. & Lynn, D. H. High-level genetic diversity but no population structure inferred from nuclear and mitochondrial markers of the peritrichous ciliate *Carchesium polypinum* in the Grand River basin (North America). *Appl. Environ. Microbiol.* **75**, 3187–3195, https://doi.org/10.1128/AEM.00178-09 (2009).
9. Gentekaki, E. & Lynn, D. H. Evidence for cryptic speciation in *Carchesium polypinum* Linnaeus, 1758 (Ciliophora: Peritrichia) inferred from mitochondrial, nuclear, and morphological markers. *J. Eukaryot. Microbiol.* **57**, 508–519, https://doi.org/10.1111/j.1550-7408.2010.00505.x (2010).
10. Kher, C. P. *et al.* Barcoding *Tetrahymena*: discriminating species and identifying unknowns using the cytochrome c oxidase subunit I (cox-1) barcode. *Protist* **162**, 2–13, https://doi.org/10.1016/j.protis.2010.03.004 (2011).
11. Przyboś, E., Tarcz, S., Rautian, M. & Sawka, N. Delimiting species boundaries within a paraphyletic species complex: insights from morphological, genetic, and molecular data on *Paramecium sonneborni* (*Paramecium aurelia* species complex, Ciliophora, Protozoa). *Protist* **166**, 438–456, https://doi.org/10.1016/j.protis.2015.07.001 (2015).
12. Shazib, S. U. A., Vďačný, P., Kim, J. H., Jang, S. W. & Shin, M. K. Molecular phylogeny and species delimitation within the ciliate genus *Spirostomum* (Ciliophora, Postciliodesmatophora, Heterotrichea), using the internal transcribed spacer region. *Mol. Phylogenet. Evol.* **102**, 128–144, https://doi.org/10.1016/j.ympev.2016.05.041 (2016).
13. Strüder-Kypke, M. C., Wright, A. D. G., Fokin, S. I. & Lynn, D. H. Phylogenetic relationships of the genus *Paramecium* inferred from small subunit rRNA gene sequences. *Mol. Phylogenet. Evol.* **14**, 122–130, https://doi.org/10.1006/mpev.1999.0686 (2000).
14. Syberg-Olsen, M. J. *et al.* Biogeography and character evolution of the ciliate genus *Euplotes* (Spirotrichea, Euplotia), with description of *Euplotes curdsi* sp. nov. *PLoS One* **11**, e0165442, https://doi.org/10.1371/journal.pone.0165442 (2016).
15. Tarcz, S. *et al. Paramecium putrinum* (Ciliophora, Protozoa): The first insight into the variation of two DNA fragments – Molecular support for the existence of cryptic species. *Mol. Phylogenet. Evol.* **73**, 140–145, https://doi.org/10.1016/j.ympev.2014.01.019 (2014).
16. Zhao, Y. *et al.* Utility of combining morphological characters, nuclear and mitochondrial genes: an attempt to resolve the conflicts of species identification for ciliated protists. *Mol. Phylogenet. Evol.* **94**, 718–729, https://doi.org/10.1016/j.ympev.2015.10.017 (2016).
17. Zhao, Y., Yi, Z., Warren, A. & Song, W. Species delimitation for the molecular taxonomy and ecology of the widely distributed microbial eukaryote genus *Euplotes* (Alveolata, Ciliophora). *Proc. R. Soc. B* **285**, 20172159, https://doi.org/10.1098/rspb.2017.2159 (2018).
18. Curds, C. R., Gates, M. A. & Roberts, D. M. L. *British and Other Freshwater Ciliated Protozoa. Part II. Ciliophora: Kinetofragminophora.* The Linnean Society of London and the estuarine and brackish-water sciences association. (Cambridge University Press, 1983).

19. Fernandes, N. M. & da Silva Neto, I. D. Morphology and 18S rDNA gene sequence of *Spirostomum minus* and *Spirostomum teres* (Ciliophora: Heterotrichea) from Rio de Janeiro, Brazil. *Zool. (Curitiba)* **30**, 72–79, https://doi.org/10.1590/S1984-46702013000100009 (2013).
20. Jang, S. W., Kwon, C. B. & Shin, M. K. First records of two *Spirostomum* ciliates (Heterotrichea: Heterotrichida: Spirostomidae) from Korea. *Animal Syst. Evol. Divers.* **28**, 29–35, https://doi.org/10.5635/ASED.2012.28.1.029 (2012).
21. Repak, A. J. & Isquith, I. R. The systematics of the genus *Spirostomum* Ehrenberg, 1838. *Acta Protozool.* **12**, 325–333 (1974).
22. Shazib, S. U. A., Vďačný, P., Kim, J. H., Jang, S. W. & Shin, M. K. Phylogenetic relationships of the ciliate class Heterotrichea (Protista, Ciliophora, Postciliodesmatophora) inferred from multiple molecular markers and multifaceted analysis strategy. *Mol. Phylogenet. Evol.* **78**, 118–135, https://doi.org/10.1016/j.ympev.2014.05.012 (2014).
23. Foissner, W., Berger, H. & Kohmann, F. Taxonomische und ökologische Revision der Ciliaten des Saprobiensystems — Band II: Peritrichia, Heterotrichida, Odontostomatida. *Informationsberichte Bayer. Landesamtes für Wasserwirtschaft* **5/92**, 1–502 (1992).
24. Bradley, M. W., Esteban, G. F. & Finlay, B. J. Ciliates in chalk-stream habitats congregate in biodiversity hot spots. *Res. Microbiol.* **161**, 619–625, https://doi.org/10.1016/j.resmic.2010.05.016 (2010).
25. Finlay, B. J. & Esteban, G. F. Freshwater protozoa: Biodiversity and ecological function. *Biodivers. Conserv.* **7**, 1163–1186, https://doi.org/10.1023/A:1008879616066 (1998).
26. Berger, H. & Foissner, W. Illustrated guide and ecological notes to ciliate indicator species (Protozoa, Ciliophora) in running waters, lakes, and sewage plants in *Handbuch* Angewandte Limnologie (eds Steinberg, C., Calmano, W., Klapper, H. & Wilken, R.-D.) 1–160. (Ecomed Verlag, 2003).
27. Berger, H., Foissner, W. & Kohmann, F. *Bestimmung und Ökologie der Mikrosaprobien nach DIN 38410*. (Gustav Fischer, 1997).
28. Nałecz-Jawecki, G. & Sawicki, J. Spirotox – a new tool for testing the toxicity of volatile compounds. *Chemosphere* **38**, 3211–3218, https://doi.org/10.1016/S0045-6535(98)00482-2 (1999).
29. Twagilimana, L., Bohatier, J., Groliere, C. A., Bonnemoy, F. & Sargos, D. A new low-cost microbiotest with the protozoan *Spirostomum teres*: culture conditions and assessment of sensitivity of the ciliate to 14 pure chemicals. *Ecotoxicol. Environ. Saf.* **41**, 231–244, https://doi.org/10.1006/eesa.1998.1698 (1998).
30. Tirjaková, E., Krajčovičová, K., Illyová, M. & Vďačný, P. Interaction of ciliate communities with cyanobacterial water bloom in a shallow, hypertrophic reservoir. *Acta Protozool.* **55**, 173–188, https://doi.org/10.4467/16890027AP.16.017.5749 (2016).
31. Hines, H. N., Onsbring, H., Ettema, T. J. & Esteban, G. F. Molecular investigation of the ciliate *Spirostomum semivirescens*, with first transcriptome and new geographical records. *Protist* **169**, 875–886, https://doi.org/10.1016/j.protis.2018.08.001 (2018).
32. Chen, X., Kim, J. H., Shazib, S. U. A., Kwon, C. B. & Shin, M. K. Morphology and molecular phylogeny of three heterotrichid species (Ciliophora, Heterotrichea), including a new species of *Anigsteinia*. *Eur. J. Protistol.* **61**, 278–293, https://doi.org/10.1016/j.ejop.2017.06.005 (2017).
33. Fernandes, N. M., Paiva, T. da S., da Silva-Neto, I. D., Schlegel, M. & Schrago, C. G. Expanded phylogenetic analyses of the class Heterotrichea (Ciliophora, Postciliodesmatophora) using five molecular markers and morphological data. *Mol. Phylogenet. Evol.* **95**, 229–246, https://doi.org/10.1016/j.ympev.2015.10.030 (2016).
34. Kim, M. S., Kim, J. H. & Shin, M. K. First record of two *Spirostomum* species (Spirostomatidae, Heterotrichida, Heterotrichea) of ciliates from Jindo Island in Korea. *Animal Syst. Evol. Divers.* **9**, 1–9, https://doi.org/10.5635/ASED.2016.SIN9.041 (2016).
35. Caisová, L., Marin, B. & Melkonian, M. 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 (2011).
36. Coleman, A. W. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.* **19**, 370–375, https://doi.org/10.1016/S0168-9525(03)00118-5 (2003).
37. Goertzen, L. R., Cannone, J. J., Gutell, R. R. & Jansen, R. K. ITS secondary structure derived from comparative analysis: Implications for sequence alignment and phylogeny of the Asteraceae. *Mol. Phylogenet. Evol.* **29**, 216–234, https://doi.org/10.1016/S1055-7903(03)00094-0 (2003).
38. Keller, A. *et al.* Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. *Biol. Direct.* **5**, 1–12, https://doi.org/10.1186/1745-6150-5-4 (2010).
39. Mai, J. C. & Coleman, A. W. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J. Mol. Evol.* **44**, 258–271, https://doi.org/10.1007/PL00006143 (1997).
40. Marin, B. & Melkonian, M. 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 (2010).
41. Müller, T., Philippi, N., Dandekar, T., Schultz, J. & Wolf, M. Distinguishing species. *RNA* **13**, 1469–1472, https://doi.org/10.1261/rna.617107 (2007).
42. Wolf, M., Chen, S., Song, J., Ankenbrand, M. & Müller, T. Compensatory base changes in ITS2 secondary structures correlate with the biological species concept despite intragenomic variability in ITS2 sequences – A proof of concept. *PLoS One* **8**, e66726, https://doi.org/10.1371/journal.pone.0066726 (2013).
43. Wolf, M. ITS so much more. *Trends Genet.* **31**, 175–176, https://doi.org/10.1016/j.tig.2015.02.005 (2015).
44. Barth, D., Krenek, S., Fokin, S. I. & Berendonk, T. U. Intraspecific genetic variation in *Paramecium* revealed by mitochondrial cytochrome c oxidase I sequences. *J. Eukaryot. Microbiol.* **53**, 20–25, https://doi.org/10.1111/j.1550-7408.2005.00068.x (2006).
45. Gao, F. *et al.* The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). *Sci. Rep.* **6**, 24874, https://doi.org/10.1038/srep24874 (2016).
46. Lynn, D. H. & Strüder-Kypke, M. C. Species of *Tetrahymena* identical by small subunit rRNA gene sequences are discriminated by mitochondrial cytochrome c oxidase I gene sequences. *J. Eukaryot. Microbiol.* **53**, 385–387, https://doi.org/10.1111/j.1550-7408.2006.00116.x (2006).
47. Santoferrara, L. F., McManus, G. B. & Alder, V. A. 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 (2013).
48. Stoeck, T., Przybos, E. & Dunthorn, M. The D1-D2 region of the large subunit ribosomal DNA as barcode for ciliates. *Mol. Ecol. Resour.* **14**, 458–468, https://doi.org/10.1111/1755-0998.12195 (2014).
49. Yi, Z. *et al.* Molecular phylogeny of *Pseudokeronopsis* (Protozoa, Ciliophora, Urostylida), with reconsideration of three closely related species at inter- and intra-specific levels inferred from the small subunit ribosomal RNA gene and the ITS1-5.8S-ITS2 region sequences. *J. Zool.* **275**, 268–275, https://doi.org/10.1111/j.1469-7998.2008.00438.x (2008).
50. Yi, Z., Katz, L. A. & Song, W. Assessing whether alpha-tubulin sequences are suitable for phylogenetic reconstruction of Ciliophora with insights into its evolution in euplotids. *PLoS One* **7**, e40635, https://doi.org/10.1371/journal.pone.0040635 (2012).
51. Zhang, Q., Simpson, A. & Song, W. Insights into the phylogeny of systematically controversial haptorian ciliates (Ciliophora, Litostomatea) based on multigene analyses. *Proc. R. Soc. B* **279**, 2625–2635, https://doi.org/10.1098/rspb.2011.2688 (2012).
52. Townsend, T. M. *et al.* Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation. *Mol. Phylogenet. Evol.* **61**, 363–380, https://doi.org/10.1016/j.ympev.2011.07.008 (2011).
53. Rajter, Ľ. & Vďačný, P. Selection and paucity of phylogenetic signal challenge the utility of alpha-tubulin in reconstruction of evolutionary history of free-living litostomateans (Protista, Ciliophora). *Mol. Phylogenet. Evol.* **127**, 534–544, https://doi.org/10.1016/j.ympev.2018.05.011 (2018).
54. Drummond, A. J. & Bouckaert, R. R. *Bayesian Evolutionary Analysis with BEAST* (Cambridge University Press, 2015).

55. Esteban, G. F., Bradley, M. W. & Finlay, B. J. A case-building *Spirostomum* (Ciliophora, Heterotrichida) with zoochlorellae. *Eur. J. Protistol.* **45**, 156–158, https://doi.org/10.1016/j.ejop.2009.01.002 (2009).

56. Vďačný, P. Integrative taxonomy of ciliates: Assessment of molecular phylogenetic content and morphological homology testing. *Eur. J. Protistol.* **61**, 388–398, https://doi.org/10.1016/j.ejop.2017.02.001 (2017).

57. Valbonesi, A., Ortenzi, C. & Luporini, P. An integrated study of the species problem in the *Euplotes crassus–minuta–vannus* group. *J. Protozool.* **35**, 38–45, https://doi.org/10.1111/j.1550-7408.1988.tb04073.x (1988).

58. Roux, J. Faune infusorienne des eaux stagnantes des environs de Genève. *Mém. Inst. Natn. Génev.* **19**, 1–148 (1901).

59. Cognato, A. I. Standard percent DNA sequence difference for insects does not predict species boundaries. *J. Econ. Entomol.* **99**, 1037–1045, https://doi.org/10.1093/jee/99.4.1037 (2006).

60. Mwabvu, T., Lamb, J., Slotow, R., Hamer, M. & Barraclough, D. Is millipede taxonomy based on gonopod morphology too inclusive? Observations on genetic variation and cryptic speciation in *Bicoxidens flavicollis* (Diplopoda: Spirostreptida: Spirostreptidae). *Afr. Invertebr.* **54**, 349–356, https://doi.org/10.5733/afin.054.0203 (2013).

61. Strüder-Kypke, M. C. & Lynn, D. H. Comparative analysis of the mitochondrial cytochrome c oxidase subunit I (COI) gene in ciliates (Alveolata, Ciliophora) and evaluation of its suitability as a biodiversity marker. *Syst. Biodivers.* **8**, 131–148, https://doi.org/10.1080/14772000903507744 (2010).

62. Wilbert, N. Eine verbesserte Technik der Protargolimprägnation für Ciliaten. *Mikrokosmos* **64**, 171–179 (1975).

63. Medlin, L., Elwood, H. J., Stickel, S. & Sogin, M. L. The characterization of enzymatically amplified eukaryotic l6S-like rRNA-coding regions. *Gene* **71**, 491–499, https://doi.org/10.1016/0378-1119(88)90066-2 (1988).

64. Sonnenberg, R., Nolte, A. W. & Tautz, D. An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Front. Zool.* **4**, 6, https://doi.org/10.1186/1742-9994-4-6 (2007).

65. Kim, S. Y. *et al*. Morphological and ribosomal DNA-based characterization of six Antarctic ciliate morphospecies from the Amundsen Sea with phylogenetic analyses. *J. Eukaryot. Microbiol.* **60**, 497–513, https://doi.org/10.1111/jeu.12057 (2013).

66. Li, L. *et al*. Does *Kiitricha* (Protista, Ciliophora, Spirotrichea) belong to Euplotida or represent a primordial spirotrichous taxon? With suggestion to establish a new subclass Protohypotrichia. *Int. J. Syst. Evol. Microbiol.* **59**, 439–446, https://doi.org/10.1099/ijs.0.65801-0 (2009).

67. Kearse, M. *et al*. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649, https://doi.org/10.1093/bioinformatics/bts199 (2012).

68. Sela, I., Ashkenazy, H., Katoh, K. & Pupko, T. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* **43**, 7–14, https://doi.org/10.1093/nar/gkv318 (2015).

69. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729, https://doi.org/10.1093/molbev/mst197 (2013).

70. Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Ver. 4 (Sinauer Associates, 2003).

71. Guindon, S. & Gascuel, O. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Syst. Biol.* **52**, 696–704, https://doi.org/10.1080/10635150390235520 (2003).

72. Posada, D. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256, https://doi.org/10.1093/molbev/msn083 (2008).

73. Ronquist, F. *et al*. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542, https://doi.org/10.1093/sysbio/sys029 (2012).

74. Izquierdo-Carrasco, F., Smith, S. A. & Stamatakis, A. Algorithms, data structures, and numerics for likelihood-based phylogenetic inference of huge trees. *BMC Bioinform.* **12**, 470, https://doi.org/10.1186/1471-2105-12-470 (2011).

75. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490, https://doi.org/10.1371/journal.pone.0009490 (2010).

76. Stamatakis, A. Phylogenetic models of rate heterogeneity: a high performance computing perspective in *Proceedings 20th IEEE International Parallel & Distributed Processing Symposium* (2006).

77. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, https://doi.org/10.1093/bioinformatics/btu033 (2014).

78. Jones, G. Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *J. Math. Biol.* **74**, 447–467, https://doi.org/10.1007/s00285-016-1034-0 (2017).

79. Bouckaert, R. *et al*. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537, https://doi.org/10.1371/journal.pcbi.1003537 (2014).

80. Rambaut, A. & Drummond, A. Tracer v1.4, http://beast.bio.ed.ac.uk/Tracer (2007).

81. Than, C., Ruths, D. & Nakhleh, L. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinform.* **9**, 322, https://doi.org/10.1186/1471-2105-9-322 (2008).

82. Wen, D., Yu, Y., Zhu, J. & Nakhleh, L. Inferring phylogenetic networks using PhyloNet. *Syst. Biol.* **67**, 735–740, https://doi.org/10.1093/sysbio/syy015 (2018).

83. Huson, D. H. & Scornavacca, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61**, 1061–1067, https://doi.org/10.1093/sysbio/sys062 (2012).

84. Yang, Z. & Rannala, B. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. USA* **107**, 9264–9269, https://doi.org/10.1073/pnas.0913022107 (2010).

85. Yang, Z. The BPP program for species tree estimation and species delimitation. *Curr. Zool.* **61**, 854–865, https://doi.org/10.1093/czoolo/61.5.854 (2015).

86. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508, https://doi.org/10.1080/10635150290069913 (2002).

87. Shimodaira, H. Testing regions with nonsmooth boundaries via multiscale bootstrap. *J. Stat. Plan. Inference* **138**, 1227–1241, https://doi.org/10.1016/j.jspi.2007.04.001 (2008).

88. Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247, https://doi.org/10.1093/bioinformatics/17.12.1246 (2001).

89. Delport, W., Poon, A. F. Y., Frost, S. D. W. & Kosakovsky Pond, S. L. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455–2457, https://doi.org/10.1093/bioinformatics/btq429 (2010).

90. Pond, S. L. & Frost, S. D. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531–2533, https://doi.org/10.1093/bioinformatics/bti320 (2005).

91. Pond, S. L. K., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679, https://doi.org/10.1093/bioinformatics/bti079 (2004).

92. Murrell, B. *et al*. Gene-wide identification of episodic selection. *Mol. Biol. Evol.* **32**, 1365–1371, https://doi.org/10.1093/molbev/msv035 (2015).

Author contributions

S.U.A.S., P.V and M.K.S. designed the research. S.U.A.S. and M.K.S. conducted fieldwork. S.U.A.S. performed laboratory work. S.U.A.S., P.V., M.S. and E.G. prepared the data sets and conducted analyses. S.U.A.S. and P.V. wrote the first draft of the article. All authors contributed to revisions of the draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-52722-4.

Correspondence and requests for materials should be addressed to M.K.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.