Multigene phylogenies of clonal *Spumella*-like strains, a cryptic heterotrophic nanoflagellate, isolated from different geographical regions

Thorsten Stoeck,1 Steffen Jost1† and Jens Boenigk2

1School of Biology, University Kaiserslautern, Erwin-Schrödinger Str. 14, D-67663 Kaiserslautern, Germany
2Austrian Academy of Sciences, Institute for Limnology, Mondseestr. 9, A-5310 Mondsee, Austria

Many nanoflagellate morphospecies comprise an enormous variation of genotypes, probably indicating cryptic species. One of the best-investigated morphospecies with respect to molecular and ecophysiological variation are flagellates of the *Spumella* morphotype. Here, we have phylogenetically analysed three protein-coding genes (actin, alpha-tubulin, beta-tubulin), internal transcribed spacers (ITS1, ITS2) and the 5.8S rDNA of 17 *Spumella*-like strains isolated from soil, freshwater and marine samples in order to (i) test the validity of the current *Spumella*-like phylogenetic classification system based exclusively on small subunit (SSU) rDNA, (ii) elucidate the phylogenetic associations of SSU rDNA-unresolved strains and (iii) evaluate the validity of the assignment of ecophysiological adaptations to previously identified SSU rDNA sequence clades. All single-gene analyses show different patterns of support, are incongruent and identify a number of conflicting nodes. Likewise, a concatenation of all protein genes fails to recover specific SSU rDNA clades. However, a combined analysis of all genes confidently resolved the conflicts of the single genes and the protein-gene concatenations and resulted in a tree topology that is identical to the SSU rDNA analysis, but with enhanced phylogenetic resolution and decisively greater support. We conclude that, depending on the genes concatenated, a ‘supergene’ analysis minimizes artefactual effects of single genes and may be superior in its performance in phylogenetically analysing cryptic species. We confirm the validity of the SSU rDNA *Spumella*-like phyloclades and support the suggestion that these clades indeed seem to reflect certain ecophysiological adaptations.

INTRODUCTION

In protistology, species descriptions often pragmatically conform to the morphological species concept: microbial eukaryotes have traditionally been identified on the basis of their morphological diversity (Finlay et al., 1996). In general, such a concept seems to work well for a number of larger protists. However, numerous smaller species (nano- and picoeukaryotes) are morphologically indistinguishable (Coleman, 2002; Hackstein, 1997; Nanney, 2004), as evolution and speciation in microbial eukaryotes are not necessarily accompanied by a perceptible morphological change (Machelon et al., 1984). As a consequence, distinct microbial eukaryote species are often classified as a single species (=cryptic species). The prevalence of such cryptic species impairs biodiversity estimates and analyses of spatial distribution patterns and ecological functions (Bickford et al., 2006; Coleman, 2002; Hackstein, 1997). A group of organisms that falls into this category are colourless chrysophytes often referred to as *‘Spumella*-like' flagellates or *Spumella* spp. (Auer & Arndt, 2001; Cleven & Weisse, 2001; Weitere & Arndt, 2003). These organisms account for a substantial proportion of the heterotrophic nanoflagellate (HNF) biomass, especially in freshwater systems (Domaizon et al., 2003; Felip et al., 1999; Zhao et al., 2003), and play a major role in microbial ecology as...
they contribute significantly to carbon transfer through freshwater food webs (Arndt et al., 2000; Boenigk & Arndt, 2002; Sherr & Sherr, 1994). Field investigations and food web models have so far been focused primarily on HNFs as a functional group (Bojanic et al., 2006; Gasol et al., 1995; Weitere & Arndt, 2002). However, merging different taxa into ‘black boxes’ is insufficient to describe specific interactions and pathways within microbial food webs. Therefore, an increasing effort to focus on species- or taxon-specific investigations has arisen (Clever & Weisse, 2001).

The advent of DNA sequencing techniques has given biologists a new tool for the identification of such cryptic microbial eukaryotes (Darling et al., 2004; Hausmann et al., 2006; Scheekenbach et al., 2006; Slapeta et al., 2006). Recently, Boenigk et al. (2005) took advantage of this technique and analysed 28 Spumella-like strains from different geographical regions. The authors discovered sequence dissimilarities in the 18S (small subunit; SSU) rDNA phylogeny is characterized by some shortcomings: (i) four SSU rDNA phylogenetic clades. However, this SSU rDNA failed to resolve several Spumella-like phylogenetic clades are not supported, (ii) none of the subclades in the most prominent clade (C clade) are supported, (iii) the SSU rDNA was characterized by some shortcomings: (i) four out of six identified major Spumella-like phylogenetic clades are not supported, (ii) none of the subclades in the most prominent clade (C clade) are supported, (iii) the SSU rDNA failed to resolve several Spumella-like isolates from different geographical regions and (iv) several strains are characterized by unusually long branches, hampering phylogenetic reconstruction (Felsenstein, 1978). As a result, the discussion of SSU rDNA Spumella-like phylolocades in the context of ecological adaptation (ecotypes) and biogeography could lead to premature and erroneous pictures.

Based on the idea that slowly evolving markers, such as SSU rDNA, may be insufficient to resolve the phylogenetic relations of closely related taxa (Avise, 2004; Roselló-Mora & Amann, 2001; Schlegel & Meisterfeld, 2003; Tsuchiya et al., 2003) and inadequate to detect ecophysiological differences between populations of the same morphospecies (Lowe et al., 2005; Rodriguez et al., 2005), we used multigene analyses to evaluate the validity of (i) the current Spumella-like phylogenetic classification system and (ii) the ecological and biogeographical conclusions that are based on this classification. To this end, we sequenced three different protein-coding genes (actin, alpha-tubulin and beta-tubulin) and fast-evolving ribosomal internal transcribed spacer (ITS) fragments of 17 Spumella-like flagellate strains isolated from freshwater and soil habitats. Additionally, we included a novel marine strain. Our phylogenetic analyses were based on single genes and concatenations of different genes.

**METHODS**

**Isolation and culturing of strain Got220.** We applied a modified filtration–acclimatization method (Hahn et al., 2004) for the isolation of strain Got220. Briefly, 20 ml sample water from the Gotland Deep in the Baltic Sea was filtered through 5 μm pore-size filters (Minisart syringe filters; Sartorius) and collected in sterile Erlenmeyer flasks. After 24 h acclimatization at room temperature, subsamples of 5 ml were stepwise diluted (three subsequent 1:1 dilutions) with sterile-filtered sample water. After 2–4 h, the flagellates were counted using a Sedingewick–Rafter chamber and a subsample was diluted to a final flagellate abundance of 0.5–1 flagellates ml$^{-1}$ and subsequently transferred to a 24-well cell culture plate. Wells were supplemented with food bacteria (Listonella pelagia) at a concentration of 3–5 × 10⁶ bacteria ml$^{-1}$. Wells were checked every second day for a period of at least 2 weeks for positive growth under the microscope using a total magnification of × 200. When flagellate growth was detected, the medium was transferred to a 50 ml Erlenmeyer flask containing Schmalz–Pratt medium (l⁻¹: 0.01 g K₂HPO₄, 3H₂O, 0.1 g KNO₃, 1.45 g CaCl₂, 2H₂O, 6.92 g MgSO₄, 7H₂O, 5.51 g MgCl₂, 6H₂O, 0.67 g KCl and 28.15 g NaCl) and fresh food bacteria. After 2–6 days, the subsamples were further diluted to final concentrations of 0.05, 0.1, 0.2 and 0.4 flagellates ml$^{-1}$ and supplemented with fresh food bacteria at a concentration of 15–25 × 10⁶ bacteria ml$^{-1}$. Each of these dilutions was transferred to wells of sterile 24-well cell-plate cultures (1 ml per well) and incubated at room temperature. Screening of the wells for the growth of flagellates was again performed by direct microscopic investigation every second day. Finally, flagellates were transferred to an Erlenmeyer flask containing fresh Schmalz–Pratt medium and food bacteria. This procedure was repeated until pure cultures were established, but at least four times. Pure cultures were acclimatized at 15 °C and transferred to permanent culture. During permanent culture, the flagellate strains were held on a mixed bacterial community on a wheat grain at 16 °C. Based on light and electron microscope investigations, strain Got220 was identified as a Spumella-like morphotype (data not shown).

Even though the Baltic Sea sample from which we isolated strain Got220 was characterized by a salinity of only 20 %, optimal growth conditions were only achieved in a medium with a salinity of 30 %. Growth at 20 % was only marginal. Under freshwater conditions, the strain did not survive. Thus, despite its origin from a brackish water environment, it is reasonable to assume that Got220 is indeed a marine taxon, which might have entered the Baltic Sea via saline deep currents from the North Sea.

**Origin of freshwater and soil strains and culture conditions.** In addition to the marine strain Got220, 16 further flagellate strains affiliated with the Spumella morphotype (Boenigk et al., 2006) were selected for in-depth investigation (Table 1). Rather than use a virtually meaningless binomen resulting from the current confusion regarding the classification of ‘Spumella-like’ flagellates, we will refer to them as Spumella sp. throughout this document. The selection comprised flagellates affiliated with different SSU rDNA sequence clades as defined earlier (C1, C2, C3 and E; Boenigk et al., 2005, 2006). Further selection criteria were the place of origin (strains from distant sampling sites), different habitat types (soil, freshwater) and different temperature tolerances (Boenigk et al., 2006). Details of the strains used in this study along with their temperature adaptation traits (Boenigk et al., 2007) are shown in Table 1. Based on light and electron microscope investigations, these strains all belong to the same morphospecies (data not shown; Boenigk et al., 2004).

We used NSY inorganic basal medium for cultivation of freshwater and soil strains (Hahn et al., 2003). The gammaproteobacterium Listonella pelagia (= Vibrio pelagius) strain CBS 8 (Hahn & Höpff, 1998) was supplied as a food source. The bacteria were also grown in NSY medium (3 g organic matter per litre; Hahn et al., 2003). During permanent culture, the flagellate strains were held on a mixed bacterial community on a wheat grain at 16 °C.

All flagellate strains are available at no cost from the culture collection at the Institute of Limnology of the Austrian Academy of Sciences (contact authors J. B. and S. J.). The accession codes of the strains are identical to the names of the strains.

http://ijs.sgmjournals.org
Table 1. Origin of strains used in this study and SSU rDNA sequence GenBank accession numbers

Isolate names are identical to strain accession numbers in the Spumella culture collection of the Institute of Limnology (Austrian Academy of Sciences), Mondsee, Austria. Optimum growth temperatures (\(T_{\text{opt}}\)) were taken from Boenigk et al. (2007). \(\text{NA}\), Not available.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Accession number</th>
<th>Place of origin</th>
<th>Latitude/longitude</th>
<th>(T_{\text{opt.}}) ((^{\circ})C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JBAF32</td>
<td>AY651076</td>
<td>Uganda, freshwater</td>
<td>0° 37’ N 30° 16’ E</td>
<td>33.4</td>
</tr>
<tr>
<td>JBA45</td>
<td>DQ88541</td>
<td>USA, Hawaii, freshwater</td>
<td>19° 33’ N 154° 53’ W</td>
<td>26.7</td>
</tr>
<tr>
<td>JBC07</td>
<td>AY651097</td>
<td>China, Taihu, freshwater</td>
<td>31° 30’ N 120° 20’ E</td>
<td>33.0</td>
</tr>
<tr>
<td>JBCS24</td>
<td>AY651082</td>
<td>China, Shanghai, soil</td>
<td>31° 6’ N 121° 22’ E</td>
<td>27.3</td>
</tr>
<tr>
<td>JBCS23</td>
<td>AY651081</td>
<td>China, Badaling, soil</td>
<td>40° 20’ N 115° 58’ E</td>
<td>24.9</td>
</tr>
<tr>
<td>JBNZ41</td>
<td>AY651075</td>
<td>New Zealand, Aviemore, freshwater</td>
<td>44° 40’ S 170° 22’ E</td>
<td>32.3</td>
</tr>
<tr>
<td>JBNZ39</td>
<td>AY651078</td>
<td>New Zealand, Karangarua, freshwater</td>
<td>43° 37’ S 169° 46’ E</td>
<td>32.7</td>
</tr>
<tr>
<td>JBM10</td>
<td>AY651074</td>
<td>Austria, Mondsee, freshwater</td>
<td>47° 52’ N 13° 20’ E</td>
<td>34.3</td>
</tr>
<tr>
<td>JBM08</td>
<td>AY651098</td>
<td>Austria, Mondsee, freshwater</td>
<td>47° 52’ N 13° 20’ E</td>
<td>NA</td>
</tr>
<tr>
<td>JBM18</td>
<td>AY651092</td>
<td>Austria, St. Gilgen, freshwater</td>
<td>47° 47’ N 13° 23’ E</td>
<td>19.9</td>
</tr>
<tr>
<td>JBL14</td>
<td>AY651086</td>
<td>Austria, Lunz, freshwater</td>
<td>47° 51’ N 15° 3’ E</td>
<td>22.7</td>
</tr>
<tr>
<td>Got220</td>
<td>EF027354</td>
<td>Baltic Sea, Gotland Basin, marine</td>
<td>57° 18’ N 19° 0’ E</td>
<td>30.1</td>
</tr>
<tr>
<td>1031</td>
<td>DQ388563</td>
<td>Antarctica, Signy Island, freshwater</td>
<td>60° 42’ S 45° 36’ W</td>
<td>11.8</td>
</tr>
<tr>
<td>1006</td>
<td>DQ388558</td>
<td>Antarctica, Signy Island, soil</td>
<td>60° 42’ S 45° 36’ W</td>
<td>17.7</td>
</tr>
<tr>
<td>194f</td>
<td>DQ388551</td>
<td>Antarctica, Alexander Island, freshwater</td>
<td>71° 00’ S 68° 00’ W</td>
<td>16.4</td>
</tr>
<tr>
<td>39f</td>
<td>DQ388557</td>
<td>Antarctica, Davis Valley, freshwater</td>
<td>82° 27’ S 51° 21’ W</td>
<td>13.2</td>
</tr>
<tr>
<td>376hm</td>
<td>DQ388553</td>
<td>Antarctica, Davis Valley, freshwater</td>
<td>82° 28’ S 50° 36’ W</td>
<td>17.7</td>
</tr>
<tr>
<td>199hm</td>
<td>DQ388552</td>
<td>Antarctica, Davis Valley, freshwater</td>
<td>82° 28’ S 50° 36’ W</td>
<td>16.2</td>
</tr>
</tbody>
</table>

Nucleic acid extraction, PCR amplification, cloning and sequencing. Samples (50 ml) of each clonal culture (\(5 \times 10^4\) cells ml\(^{-1}\)) were centrifuged, the supernatant was discarded and genomic DNA was extracted using the DNeasy Tissue kit (Qiagen) according to the protocol of the manufacturer. PCR primers for SSU rDNA, ITS1–5.8S–ITS2 fragment, alpha-tubulin, beta-tubulin and actin are given in Table 2. Each PCR contained 10–20 ng DNA template, 2.5 U HotStar Taq DNA polymerase (Qiagen) in the manufacturer-provided reaction buffer [Tris/HCl (pH 8.7), KCl and \((\text{NH}_4)_2\text{SO}_4\)], 1.5 mM MgCl\(_2\), 200 mM of each dNTP and 0.5 mM of each oligonucleotide primer. The final volume was adjusted to 50 µl with sterile water. The PCR protocol consisted of an initial hot-start incubation (15 min at 95 °C) followed by 30 identical amplification cycles of denaturing at 95 °C for 45 s, annealing for 1 min (temperatures for specific primer sets are outlined in Table 2) and extension at 72 °C for 2.5 min and a final extension at 72 °C for 7 min. The resulting PCR products were purified (PCR MinElute kit; Qiagen) and cloned into a vector (TA Cloning kit; Invitrogen). Plasmids were isolated (Qiaprep Spin Miniprep kit; Qiagen) from overnight cultures and reamplified by PCR using primers M13F and M13R to screen for inserts of the expected sizes. Positive clones for each strain and amplified gene were sequenced (M13F and M13R primers) using an Applied Biosystems ABI 3730 DNA Stretch Sequencer, with the XL Upgrade and the ABI Prism BigDye Terminator version 3.1 cycle sequencing ready reaction kit. By bidirectional sequencing, we obtained twofold coverage of each gene and clone. Prior to sequence assembly, we performed Phred20 processing of the sequences using the program CodonCode Aligner (CodonCode Corp.) to ensure high-quality sequences. SSU rDNA sequences for freshwater

Table 2. PCR primer sequences and corresponding annealing temperatures in PCRs used in this study

Mixed-base sites (wobble positions) are denoted by standard IUB nomenclature: D, G/A/T; M, A/C; R, A/G; W, A/T; Y, C/T. Euk, Eukaryote-specific SSU rDNA primer; F, forward; R, reverse.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’–3’)</th>
<th>Annealing temperature (°C)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ActinF</td>
<td>GGCCTGGAARCCAYTTNCGRTGNA</td>
<td>56</td>
<td>Harper et al. (2005)</td>
</tr>
<tr>
<td>ActinR</td>
<td>TGGGAYGAYATGGARARATHTGG</td>
<td>62</td>
<td>Harper et al. (2005)</td>
</tr>
<tr>
<td>Alpha-Tubulin-F</td>
<td>CGCGGCCTCARTGNGNAAYGCTNTGYTGGA</td>
<td>59</td>
<td>Harper et al. (2005)</td>
</tr>
<tr>
<td>Alpha-Tubulin-R</td>
<td>CGCGGCCATNCCYTNCNNACRTACCA</td>
<td>53</td>
<td>White et al. (1990)</td>
</tr>
<tr>
<td>Beta-Tubulin-F</td>
<td>GCCATGCAGNCTRGGYGGNAAYCA</td>
<td>56</td>
<td>Medlin et al. (1988)</td>
</tr>
<tr>
<td>Beta-Tubulin-R</td>
<td>TCCCTGAGTRAAYTCTTCYTRTCCAT</td>
<td>1990hm</td>
<td></td>
</tr>
<tr>
<td>ITS1</td>
<td>TCCGTAAGGTCAGCTGCGG</td>
<td>53</td>
<td>White et al. (1990)</td>
</tr>
<tr>
<td>ITS4</td>
<td>TCCCTGCGTTATGATATGC</td>
<td>56</td>
<td>Medlin et al. (1988)</td>
</tr>
<tr>
<td>Euka (F)</td>
<td>AACCTGTGTTGACCTCGCCAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EukB (R)</td>
<td>TGATCCTTCTGCAGGTTCACCTAC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and soil strains were obtained from the GenBank database (accession numbers in Table 1; details of other accession numbers are given in Supplementary Table S1, available in IJSEM Online), while that for the marine strain Got220 was obtained in this study.

**Phylogenetic analyses.** A total of eight CLUSTAL_X alignments (Chenna et al., 2003) were subjected to phylogenetic analyses. Five of these were the respective alignments for the unique genes. The three remaining alignments were concatenations of different genes. One was a concatenation of all protein-coding genes, the second was a concatenation of the protein-coding genes and the SSU rDNA and a third was a concatenation of all genes. The alignments were refined manually in MacClade (Maddison & Maddison, 1992) using phylogenetically conserved regions. The alignments are available from the authors upon request. Subsequently, the program MODELTEST (Posada & Crandall, 2001) was applied to choose the model of DNA substitution that best fitted our dataset from among 56 possible models. We calculated both an evolutionary distance (ED) and a maximum-likelihood (ML) tree for each dataset using the PAUP software package 4.0b10 (Swofford, 2002) and also a Bayesian inference tree using MrBayes (Ronquist & Huelsenbeck, 2003). We assessed the relative stability of tree topologies using 1000 bootstrap replicates for the ED and ML trees. Heuristic searches for bootstrap analyses employed stepwise addition: starting trees with simple addition of sequences followed by tree bisection-reconnection (TBR) branch-swapping. Settings for bootstrap analyses were chosen according to the MODELTEST output. The parameters used to calculate the ‘best’ trees are given in the legends of the respective figures. For the concatenations, we only calculated ED and Bayesian trees.

For the Bayesian tree, we ran two simultaneous, completely independent analyses starting from different random trees. The analyses also employed GTR + I + G as the DNA substitution model with the gamma distribution shape parameter, the proportion of invariable sites, base frequencies and a rate matrix for the substitution model as assessed by MrBayes. Metropolis coupling with three heated chains and one ‘cold’ chain was employed to improve the Markov chain Monte Carlo sampling of the target distribution. We ran 1 000 000 generations and sampled every 1000th generation, resulting in 1001 samples from the posterior probability (PP) distribution. For further analysis, the first 250 trees generated were discarded (burnin).

**RESULTS**

**Single-gene phylogenies**

In accordance with previous analyses, the strains investigated in this study group into four different SSU rDNA

**Fig. 1.** (a) Minimum-evolution SSU rDNA phylogenetic tree of Spumella-like flagellate strains. The tree was constructed under ML criteria by using a GTR + I + G DNA substitution model with the variable-site gamma distribution shape parameter (G) at 0.5340, the proportion of invariable sites (I) at 0.7032 and base frequencies and a rate matrix for the substitution model as suggested by MODELTEST (Posada & Crandall, 2001), based on 1675 unambiguously aligned positions. (b) Minimum-evolution phylogenetic tree of Spumella-like flagellate strains based on a concatenation of sequences for three protein-coding genes (alpha-tubulin, beta-tubulin and actin) and rDNA fragments (SSU rDNA, ITS1, 5.8S rDNA, ITS2). The tree was constructed under ML criteria by using a GTR + I + G DNA substitution model with the variable-site gamma distribution shape parameter (G) at 0.8404, the proportion of invariable sites (I) at 0.5401 and base frequencies and a rate matrix for the substitution model as suggested by MODELTEST (Posada & Crandall, 2001), based on 4356 unambiguously aligned positions. In (a), support values are shown for nodes that received support over 50 % and are (left to right) evolutionary-distance bootstraps (1000 replicates)/maximum-likelihood bootstraps (1000 replicates)/Bayesian posterior probabilities (1001 sampled trees). In (b), support values are shown for nodes that received support over 50 % and are evolutionary distance bootstraps (1000 replicates)/Bayesian posterior probabilities (1001 samples trees). Because the position of the root is unclear, both trees are drawn with a basal trichotomy.
clades (C1, C2, C3 and E; Fig. 1a). The two Austrian freshwater strains, JBM08 and JBM18, fall into clade E and branch as sister to the clade C3, consisting of the freshwater strains JBC07, JBM10, JBAF32 and JBNZ41. These four strains are from different geographical locations (China, Austria, Uganda and New Zealand, respectively), but their SSU rDNA genes are identical, resulting in unresolved relations between these strains. Clades E and C3 are characterized by unusually long branches. Subclade C1 includes two strains isolated from soil in China (JBCS23 and JBCS24) and one from Antarctica (1006) as well as the marine strain (Got220) isolated from an oxygen-depleted basin in the Baltic Sea. Strains JBCS24 and 1006, which comprise one of the two clusters in the clade C1, cannot be distinguished from each other based on their SSU rDNA sequences. Interestingly, in this analysis, the Chinese strains S23 and S24 are more similar to the marine Got220 and the polar strain 1006, respectively, than to each other. However, the support for the cluster S23 and Got220 is low and cluster JBCS24 and 1006 is unsupported (ED/ML/PP <50%). Also, subclade C2, consisting entirely of freshwater strains, is split into three clusters: (i) JBL14 from Austria and 391f from Antarctica, with a very low support (64/79/−), (ii) strains 199hm and 376hm from Antarctica, which cannot be discriminated based on SSU rDNA, as well as the Antarctic strains 1031 and 194f, and (iii) two freshwater strains, JBNZ39 from New Zealand and JBNAA45 from Hawaii, which are highly supported (100/100/100).

The ITS and the actin, alpha- and beta-tubulin genes resolved larger numbers of strains than SSU rDNA (Supplementary Figs S1–S4, available in IJSEM Online). Interestingly, different genes failed to discriminate different strains: JBC07 and JBM10 are unresolved in the ITS phylogeny (Supplementary Fig. S1), JBL14 and JBNAA45 in the actin phylogeny (Supplementary Fig. S2) and JBM08 and JBM18 in the beta-tubulin phylogeny (Supplementary Fig. S4). All strains can be discriminated using alpha-tubulin as a marker gene (Supplementary Fig. S3). Additionally, these single-gene trees have higher statistical support compared with the SSU rDNA tree. However, they are incongruent and identify a number of conflicting nodes. For example, the branching orders within the ITS subclades C1 and C2 (Supplementary Fig. S1) are distinctly different from the respective SSU rDNA subclades (Fig. 1a). In the actin tree (Supplementary Fig. S2), strains JBNZ39 and JBNAA45 branch within the actin clades C1 and C2, respectively, while in the SSU rDNA phylogeny both strains branch in C2 with high support. Furthermore, strain JBAF32, branching in the SSU rDNA clade C3, is unassigned and characterized by an extremely long branch in the actin phylogeny (Supplementary Fig. S2). In the alpha-tubulin tree (Supplementary Fig. S3), the Antarctic freshwater strain 194f and the New Zealand freshwater strain JBNZ39 do not cluster with the other freshwater strains in subclade C2 as in the SSU rDNA phylogeny, but rather in the soil/marine subclade C1. Also, the beta-tubulin phylogeny (Supplementary Fig. S4) fails to recover the SSU rDNA subclades C1 and C2.

### Concatenated gene phylogenies

All multiple gene concatenations succeeded in resolving all strains (Fig. 1b and Supplementary Figs S5 and S6). It is conspicuous that, in the concatenated proteome-gene tree, strains JBNZ39 and JBNAA45 are assigned to subclades C1 and C2, respectively (Supplementary Fig. S5), whereas they seem genetically related to one another in the SSU rDNA tree and clearly assigned to subclade C2 (Fig. 1a).

Adding data for SSU rDNA, 5.8S rDNA, ITS1 and ITS2 to the concatenated proteome dataset yielded a tree (Fig. 1b) with topology that was different from that of the proteome-gene concatenation (Supplementary Fig. S5) but essentially identical to that of the SSU rDNA tree (Fig. 1a) with some qualitative improvements: a decisive overall increase for the support of clade C2 and a slight increase for the support of C1, no exceptionally long branches stand out in the concatenated supergene tree, compared with the SSU rDNA gene tree and all strains can be easily distinguished.

### DISCUSSION

#### Single-gene analyses

Our 18S (SSU) rDNA phylogeny is largely congruent with data published previously (Boenigk et al., 2005, 2006). Our data support the separate position of the tropical freshwater strains JBNZ39 (New Zealand) and JBNAA45 (Hawaii) within the subclade C2 (Boenigk et al., 2006). These strains, together with Spumella obliqua (see Boenigk et al., 2006), seem to be a basal sister group within or to the subclade C2.

As a rule, protists in marine and non-marine environments are largely distinct (Richards et al., 2005; von der Heyden & Cavalier-Smith, 2005; von der Heyden et al., 2004). Numerous major taxa are restricted to one or other habitat type. For example, radiolarians, pelagophytes, phaeodarians and xenophyophores are exclusively marine. In contrast, other taxa like myzoozoa are never marine. Even in groups where habitat shifts have occurred, several substantial subclades seem to be exclusive to one or other (e.g. bodonids and Gongiomanas; Koch & Ekelund, 2005; von der Heyden & Cavalier-Smith, 2005; von der Heyden et al., 2004). Interestingly, the novel marine strain isolated from the Baltic Sea clusters in the subcluster C1 that contained, so far, mainly isolates from soil (Boenigk et al., 2007). This result is supported by all other gene phylogenies conducted in this study, leaving no doubt about its accuracy.

On present evidence, the fraction of protist lineages able to live in both marine and non-marine habitats is very small. Yet the potential capacity of some Spumella-like flagellates to switch rapidly between marine and non-marine
environments might greatly facilitate a global dispersal of at least some clades. Unfortunately, to date, no further marine isolates are available to test this hypothesis. Thus, concerted efforts are in order to collect marine Spumella-like flagellates and shed light on their phylogenies and distribution.

Both the ITS1–5.8S rDNA–ITS2 fragment and the protein-coding genes analysed in this study resolve all investigated Spumella-like strains that are unresolved in the SSU rDNA tree. The SSU rDNA is highly conserved, with a relatively low evolutionary rate (Hillis & Dixon, 1991). Thus, this gene is generally well suited to the resolution of taxa that separated relatively early in eukaryote evolution. However, evidence is accumulating to show that the SSU rDNA often has insufficient resolving power to guarantee the delimitation of protist species. For example, as pointed out earlier (Massana et al., 2004), within eukaryotic microorganisms, cultures that share more than 99.5 % of the SSU rDNA sequence have been classified as belonging both to different species (Medlin et al., 1994; Schmidt et al., 2006) and the same species (Montresor et al., 2003).

A marker with a higher evolutionary rate is the ITS locus (Coleman & Mai, 1997), with mutation rates in protists of about 0.5–1 % per million years (e.g. Lajeunesse, 2005). Thus, this highly variable locus has attracted attention as a potential marker for the resolution of closely related taxa in different protists (Diggles & Adlard, 1997; Katz et al., 2004; Pelandakis et al., 2000; van Hoek et al., 1998). Also, the protein-coding genes we evaluated in this study (actin, alpha- and beta-tubulin) are known as fast-evolving in a range of different protist taxa (Kim et al., 2004; Nishi et al., 2005). Considering the wide geographical distribution of strains that cluster in a single SSU rDNA clade (e.g. in clade C3), a high evolutionary rate in functional genes may be ecologically advantageous while conquering new habitats.

Yet, our analyses revealed an unsatisfactory resolution of all strains under study regardless of the marker gene used. Unexpectedly, this also includes the ITS locus. This finding may question the widely held belief that the ITS locus belongs to the most variable regions in protist genomes and, thus, is most useful for high-resolution phylogenetic studies at a low taxonomic level. As speculated earlier by Coleman (2007), protists seem to present the most difficult choice of locus to sequence. The reasoning of the author is that the broad and ancient variety of cell types, which is greater than in fungi, animals or plants. As a consequence, the quest for a single common nuclear locus that is most useful in protist phylogeny remains a challenge.

Nevertheless, all trees show a high degree of congruence regarding the recovery of the same taxa in clade E and subclade C3. However, the analyses also reveal some conflicting nodes and topologies with different statistical support. These incongruences are restricted to clades C2 and C1. Inconsistencies in phylogenetic reconstruction using different marker genes are not uncommon in protists (Edgcomb et al., 2002; Nishi et al., 2005; Philippe & Adoutte, 1996; Shalchian-Tabrizi et al., 2006a, b) and highlight fundamental difficulties in phylogenetic reconstruction (Philippe & Adoutte, 1996). Firstly, evolutionary rates of different genes may be highly variable within the genes as well as between different organisms, resulting in different tree topologies (Philippe et al., 2000). Secondly, in contrast to ribosomal genes, which usually evolve in concert (Hillis & Dixon, 1991), ancient duplication and rearrangements of protein-coding genes may result in two or more gene loci (Henikoff et al., 1997; Kondrashov et al., 2002). The protein-gene copies in these loci might have evolved independently from each other over time, resulting in paralogue genes (Kondrashov et al., 2002; Maddison, 1997). As a matter of fact, the tubulin gene family (Baroin-Tourancheau et al., 1998; Edlund et al., 1996) as well as the actin gene (Carlini et al., 2000) may have several paralogue copies within an individual organism. However, in yeast and some alveolates, actin is encoded by a single gene (Cupple & Pearlman, 1986; Hightower & Meagher, 1986; Reece et al., 1992). While at higher taxonomic levels paralogues do not hamper phylogenetic reconstruction severely, this does not hold true at a low taxonomic level (Baroin-Tourancheau et al., 1998). Thus, even though we succeeded in resolving the taxa that were identical in the SSU rDNA phylogeny, none of the single-gene trees clearly identified the most likely tree topology, resolved conflicting nodes or identified the relations of all Spumella-like strains and phyloclades.

Combined supergene analyses

Recently, combined phylogenetic analyses using two or more genes have become increasingly popular when single genes do not provide sufficient resolution or offer conflicting results (Gontcharov et al., 2004). The reasoning for such multigene phylogenies is that a larger number of characters improves phylogenetic accuracy and resolution when the number of alignable and informative nucleotides in single-gene analyses is too limited or the rates of sequence evolution are too different (Poe & Swofford, 1999). This approach has been successful in minimizing difficulties based on single protein-coding gene evolution in plants (Bowe et al., 2000; Karol et al., 2001; Mallatt & Winchell, 2002), animals (Mallatt & Winchell, 2002), various groups of algae (Gontcharov et al., 2004; Hoef-Emden et al., 2002; Nozaki et al., 2000) and the radiation of eukaryotes (Baldauf et al., 2000; Parfrey et al., 2006). The combination of seven different genes seems to minimize any potential bias in phylogenetic reconstruction of Spumella-like strains that is caused by single genes in single taxa. Including 4356 aligned nucleotides in phylogenetic analyses of the ‘supergene’ increased the relative fraction of evolutionary informative positions distinctively (n=1045, compared with 175 in the SSU rDNA alignment).

A concatenation is no panacea in phylogenetic reconstruction, because a strong bias in evolutionary rates (leading to long-branch attraction) may still persist and even increase
when more and more characters are added (Sanderson & Shaffer, 2002). However, this was not the case with the data presented here. The resulting phylogeny was superior to all single-gene analyses when the statistical support of internal branches is considered. Specifically, the concatenations resolved the major conflict between the SSU rDNA tree and the protein-gene trees in subclades C2 and C1 in favour of the SSU rDNA analysis. Our approach ensures that conflicts between single-gene topologies are probably derived from different patterns of sequence evolution between the genes, as also observed in other protist taxa like green algae (Gontcharov et al., 2004). Several branches that, in single-gene analyses, were unresolved or weakly supported obtained greater significance in the combined analysis.

One could argue that the consistency in the topology between the concatenated supergene tree and the SSU rDNA tree is due to a significant influence of the relatively long SSU rDNA fragment itself. However, as the combined protein-gene data contribute to a larger extent to the fraction of informative positions in the ‘supergene’ than the SSU rDNA, we do not ascribe this result to a major influence of the SSU rDNA. Thus, it is the concatenation of as many genes as possible that leads to the ‘most optimal’ tree (as defined by statistical support, number of resolved branches and branch lengths).

Because phylogenetic supergene analyses confirmed the validity of the SSU rDNA *Spumella*-like phyloloclades (Boenigk et al., 2005, 2006), we support the suggestion that these clades indeed seem to reflect certain ecophysiological adaptations which may, however, be overlaid by a geographical pattern.

**ACKNOWLEDGEMENTS**

We would like to thank two anonymous reviewers and the editor for helpful comments on this manuscript. We thank the Deutsche Forschungsgemeinschaft (DFG), which supported this study with a grant to T.S. (STO414/2-2), and the German Society for the Promotion of Young Scientists in Protozoology (DGP-F), which contributed to this study with a grant to S.J.

**REFERENCES**


