

***Telonema* - an early chromalveolate lineage**

Shalchian-Tabrizi K.¹, Eikrem W.¹, Klaveness D.¹, Vaultot D.², Le Gall F.², Romari, K.², Throndsen, J.¹, Massana, R.³, Thomsen, H.A.⁴, Jakobsen K. S.¹

¹ Department of Biology, University of Oslo, N-0316 Oslo, Norway;

² Station Biologique, UMR 7127 CNRS et Université Pierre et Marie Curie, BP74, 29682 Roscoff, France

³ Institut de Ciències del Mar, CMIMA, Passeig Marítim de la Barceloneta, 37-49. 08003 Barcelona, Catalonia, Spain

⁴ Danish Institute for Fisheries Research, Dept. of Marine Ecology and Aquaculture, Kavalergården 6, 2920 Charlottenlund, Denmark

Keywords: *Telonema*, environmental samples, small subunit rRNA, phylogenetic analysis, Bayes factor, covariotide, chromalveolate, chromist, alveolate, tripartite hairs, cortical alveoli, cytoskeleton.

* Corresponding author: Kjetill S Jakobsen

Recent molecular investigations of marine samples taken from different environments, including tropical¹, temperate², and polar areas³ as well as deep thermal vents⁴, have revealed an unexpectedly high diversity of protists, some of them forming deep-branching clades within important lineages such as the alveolates and stramenopiles⁵. Using the same approach on coastal samples, we have identified a novel group of protist SSU rDNA sequences that do not correspond to any phylogenetic group previously identified. Phylogenetic inferences based on a covariotide evolution model, allowing variable substitution rates between sequences, place the new group at the base of alveolates and stramenopiles (a chromist lineage). Comparison with other SSU sequences obtained from cultures of heterotrophic protists showed that the environmental sequences grouped together with *Telonema*, a genus known since 1913⁶ but of uncertain taxonomic affinity. Morphological examination of *Telonema* revealed characteristic traits that define both the chromist and alveolate lineages. Altogether, the deep-branching phylogeny and the mixed chromist and alveolate traits indicate the *Telonema* clade is a very early chromoalveolate lineage, possibly ancestral to the chromoalveolates⁷, and thus a key clade for understanding the early evolution of chromist and alveolates. [186 words]

Phylogenetic inferences of genetic markers have recently resolved some of the most intriguing evolutionary relationships between protist lineages⁸. Two protist clades, the chromists⁹ and alveolates⁵, comprise large groups of ecologically important heterotrophic and photosynthetic species, such as diatoms, brown seaweeds, ciliates, apicomplexans (harbouring malaria causing agents) and dinoflagellates. Chromists and alveolates have usually been regarded as separate evolutionary lineages^{9,10}, but recent investigations suggest that all chromists, as well as both chromists and alveolates, have a monophyletic origin¹¹, in agreement with the chromalveolate hypothesis⁷. However, the close affiliation of chromists and alveolates is controversial and inconsistent with some contemporary phylogenetic studies¹²⁻¹⁴. Furthermore, phylogenetic reconstructions and examination of the cellular structures of these groups have so far failed to recover any likely common ancestor to the chromists and alveolates that could provide crucial knowledge about the early evolution and classification of these lineages.

In our search for new and hitherto undiscovered protist species we investigated coastal picoplankton samples from the English Channel (Roscoff) and Mediterranean Sea (Blanes), pre-filtered through 3 µm filters, to construct clone libraries of the small subunit rDNA gene (SSU) at different sampling dates, spread over the year¹⁵. Partial sequencing (500 bp starting from position 528) of clone libraries revealed the presence of many sequences that did not present any significant similarities to sequences available in gene databases, indicating a huge and unknown genetic diversity in marine pico- and nanoplankton. Screening of protist cultures from the English Channel by the same sequencing approach uncovered two cultures containing sequences with high similarity to five of the clone library

sequences. Both cultures contained *Telonema subtilis*, a species previously described by Griesmann⁶ from Roscoff and Naples in 1913. In parallel, we isolated from the Oslofjord a novel species, *Telonema antarcticum*¹⁶, containing an SSU sequence very similar to the environmental and *T. subtilis* sequences. To investigate the evolutionary origin of the *Telonema* species and the *Telonema*-like sequences, we obtained full SSU rDNA sequences from both environmental clones and cultured strains. These sequences were analysed by Bayesian inference (BI)¹⁷ and maximum likelihood (ML) using a general time reversible substitution model (GTR) together with gamma shaped distribution of site rate probability (Γ) and an invariable site parameter (I; together Γ +I). Since deep branching rRNA phylogenetic analyses can be seriously affected by covariotide substitution processes causing the proportion of invariable sites to change with time^{17,18}, we conducted BI with a Γ model and two alternative ways of accommodating covariotide patterns: 1) by allowing the invariable sites to change between invariable and variable state (Γ +cov), and, 2) by permitting a fraction of sites to stay constant while other sites could switch between invariable and invariable states (Γ +I+cov). In addition, Maximum parsimony (MP) and Minimum evolution (LogDet distances; ME-LD)¹⁹ criteria were applied. The inferences of an SSU alignment consisting of 1378 characters and 69 sequences consistently cluster the *Telonema* and *Telonema*-like sequences as a monophyletic group with high support (BI=1.0, MP=100%, ME-LogDet=100%), in which the environmental sequences from Roscoff grouped with *T. subtilis*, while the sequence obtained from Blanes grouped together with *T. antarcticum*¹⁶ (Fig. 1). The *Telonema* clade was excluded from stramenopiles with high support in all individual trees (BI=1.0, MP=99%, ME-LogDet=92%), as well as from the other chromist and alveolate phyla (BI=1.0, MP >99% ME-LogDet>92%). Importantly, the alveolates were consistently placed closer to the stramenopiles than to *Telonema* in all individual trees (BI=99, MP =71%, ME-LogDet=56%). The cercozoans clustered with the stramenopiles (BI=0.96 and ME-LogDet<50%) except in the parsimony tree where they constitute a sister group to the haptophytes (MP=77%). Haptophytes were either weakly supported as a sister group to green algae and plants in the BI-covariotide tree, or placed at the base of *Telonema* in all other inferred trees (MP=82%, ME-LogDet =73%; see supplementary BI and ML trees). The cryptophytes were tentatively placed together with the glaucophytes. Altogether, the results are largely in agreement with recent nuclear phylogenies of rDNA and protein-coding genes^{8,12} in dividing the constituent phyla into well-supported clades. Importantly, our trees including the *Telonema* group are also essentially congruent with the majority of the nuclear-based⁸ inferences in neither rigorously rejecting nor confirming the single origin of all chromist groups or the chromalveolate lineage. Evidence for such a single origin has previously only been deduced from plastid-related sequences^{11,20}. Statistical comparisons between the marginal likelihood scores of trees reconstructed with stationary (H_0) (Γ +I; ln=-20591.05) and covariotide models (H_1) (Γ +cov: ln=-20454.80 and Γ +I+cov: ln=-20435.75) show a likelihood difference of minimum 136.2 and a difference between the covariotide models of 19.05. The Bayes factor,

describing the posterior probability of the model given that prior probability of each model is equal, is defined as 2 times the ratio of the marginal likelihood scores of different models²¹, here calculated to be 272.40 and 38.10 in favour of the Γ +I+cov model. Since Bayes factors higher than 4.15 are regarded as decisive evidence for H_1 ²¹ it follows that covariotide models, in particular the Γ +I+cov model, do indeed exceed homogeneous models in describing the evolution of the SSU gene (see supplementary information for methods). Importantly, the node clustering *Telonema* to alveolates, stramenopiles and cercomonads received considerably higher posterior probability values in the Γ +I+cov tree (BI=0.90) as compared to the stationary Γ +I model (BI=0.68; see supplementary tree 1). This suggests that the covariotide patterns obscure tree-inferences when assuming homologous substitution processes. Thus, the optimal covariotide tree indicates that the *Telonema* lineage belongs to the base of the alveolates, the stramenopiles, as well as the cercomonads.

Members of the genus *Telonema* are biflagellated and have a proboscis-like structure located at the flagellar pole (Fig 2a, b). The ultrastructure of *T. subtilis* and *T. antarcticum* revealed characters that provide clues for the evolutionary position of the *Telonema* clade, including mitochondria with tubular cristae (tubulocristae) and a highly complex cytoskeleton composed of layers of microtubuli and microfilaments (Fig. 2c). Additional traits defining the phylogeny of the group were exclusively identified in *T. antarcticum*; these are characteristic cortical alveoli located just beneath the cell membrane (Fig. 2d, e) and tripartite tubular hairs on the long flagellum (Fig. 2f).

Cortical alveoli, as identified in *T. antarctica* are the main character defining the alveolates⁷, but have also been described in the stramenopile group (Raphidophyceae) as well as among the glaucophytes (Plantae)⁷. In addition, haptophytes appear to contain similar structures (as their subsurface cisterna act in the process of directed deposition of Golgi-derived scales to the cell surface²²). The subcortical vesicles in all these taxa may represent homologous structures already present in a primitive form before the radiation of the Alveolata, Chromista and Plantae as an adaptation for strengthening the cell and allow the evolution of increased cell size⁷. On the basis of this single character *Telonema* could belong to any of these phyla, but both the molecular phylogeny, and the presence of tubulocristate mitochondria suggest that *Telonema* is more closely related to alveolates and some of the chromists than to glaucophytes.

In contrast to many other groups of protists, the stramenopiles swim with a forward directed flagellum that pulls the cell owing to a number of tubular hairs attached to the surface of the flagellum²³. Typically, the stramenopiles are supposed to have tubular flagellar hairs composed of a hollow base and tubular shaft as well as non-tubular distal fibres, the so-called tripartite hairs²⁴. Since the tripartite hairs are synthesized in a composite process involving synthesis of distinct components in the endoplasmic reticulum (ER) and golgi apparatus, as well as precise assembly and transport to the

surface of the flagella , it seems unlikely that such hairs would have evolved several times independently. Tripartite hairs have until now been regarded as a synapomorphic trait defining the stramenopiles²⁵; Thus, on the basis of the complex biosynthesis and marginal distribution of tripartite hairs, *Telonema* should belong to the stramenopiles. However, phylogenetic analyses of the SSU sequences clearly reject this hypothesis, by placing alveolates as the sister group to the stramenopiles, consistent with phylogenies generated using other molecular markers⁸. Interestingly, one of the other chromist groups, the cryptophytes, also possesses similar tubular flagellar hairs (that are bipartite), produced in the ER^{24,26}. Given that the chromists are monophyletic, as suggested from recent interpretation of plastid sequence phylogeny²⁷, the most parsimonious explanation for the evolution of these hairs would be an inheritance from a common ancestor of chromists and *Telonema* lineages. However, the discrete cytoskeleton identified in *Telonema* has not been found in any of the chromist lineages, but resembles the cortical epiplasm structures in alveolates²⁸. In summary, on the basis of structural features (i.e. cortical alveoli, tubular tripartite hairs and the complex cytoskeleton) and the molecular phylogeny, we conclude that *Telonema* probably constitutes an early branch among the chromists (before the cytoskeleton was reduced in chromists), or possibly even preceding the separation of chromists and alveolates – and thus being a common ancestor of the chromalveolates⁷ (Fig 3)

The present work clearly highlights the importance of establishing cultures in parallel to the acquisition of sequences direct from the environment when aiming to address the evolution of eukaryotic cellular structures. We have demonstrated that it is possible to achieve an improved reconstruction of rRNA phylogeny by taking into account heterogeneous (covariotide) evolutionary processes. Finally, our results also emphasize the notion that the so-called "uncultivated" species²⁹ may very well correspond to existing taxonomic entities – for example established from field observations by classical protistologists - but for which no cultures have been available or, if available, were not examined in detail until now. The (re)discovery of the *Telonema*-clade is a glaring example.

[1576 words]

Methods [405 words]

- Cultures:

- Two cultures of *Telonema subtilis* were isolated from surface water sampled at the Astan station off Roscoff (English Channel, Brittany, France) on 12 April and 11 July 2000. Water pre-filtered through 3 mm filter was enriched with K medium (at 1/100 dilution) and grown at 15° C under 12h:12h light-dark cycles at 150 mmol quanta m⁻² s⁻¹. These cultures are deposited in the Roscoff Culture Collection (RCC: www.sb-roscoff.fr/phyto/collect.html). Cultures of *T. antarcticum* were isolated from surface sea water samples taken at the inner Oslofjord winter 1998 and grown in enriched seawater f/2 or IMR/2 transferring weekly to a vigorously growing culture of the Cryptophyceae *Rhinomonas* sp.. The cultures were held at 17 °C on a 14/10 h light/dark-cycles 230 yE m⁻² sec⁻¹

- Clone libraries and sequencing

- Samples were collected from surface waters of the English Channel off Roscoff (Astan station, 12 April and 10 December 2000) and of the Mediterranean Sea off Blanes (Catalan coast, 25 June 2001) and pre-filtered through 3 mm filter. DNA was collected on 0.2 µm filters and extracted by classical methods as described elsewhere³⁰. The SSU rRNA gene was amplified by PCR using primers described in Moon et al. (2000)³¹ and cloned using the TOPO-TA cloning kit (Invitrogen, Carlsbad, CA). Positive clones were partially sequenced (550 bp) by Qiagen Genomics Sequencing Services using the internal primer Euk528f (5'- GCG GTA ATT CCA GCT CCA A -3') and selected clones were fully sequenced by the same company. The *T. antarcticum* SSU sequence was generated as previously described¹⁶.

- Phylogenetic analyses

- The *Telonema* SSU sequences were aligned with a wide range of heterotrophic and algal species (see supplementary material for accession numbers) using ClustalX³² and modified by eye according to the secondary structure of the gene¹². Bayesian inferences were conducted with MrBayes v.2.01¹⁷. In addition, analyses with maximum likelihood, maximum parsimony and minimum evolution (LogDet distances) criteria were performed by using PAUP*¹⁹. Model tests were performed in a Bayesian framework calculating the Bayes factor, equivalent with the ratio of marginal likelihood values obtained under the various models. A detailed description of the phylogenetic methods is given as supplementary information at the Nature web site.

- Electron microscopy

- TEM of *Telonema subtilis* was performed by following standard procedures as described in Eikrem et al. (1998)³³. SEM of *T. antarcticum* and *T. subtilis* and TEM of *T. Antarctica* were conducted according to Shalchian-Tabrizi et al. (2003)¹⁶

Acknowledgments

We thank J. G. Ormerod for critically reading of the manuscript. This work was supported in part by grants from the Norwegian Research Council (to KSJ and DK) and the programs PICODIV (EVK3-CT-1999-00021), PicoManche, CNRS-Aventis fund, CNRS PROOF BIOSOPE, CPER Souchothèque de Bretagne and CRB (to DV).

Figure legends [306 words]

Figure 1. SSU phylogeny of *Telonema*. Tree inferred with Bayesian inference by applying a covarion model ($\Gamma+I+cov$) of evolution. The *Telonema* clade is clearly excluded from the stramenopiles (the only other phyla known to possess tripartite hairs) as well as the other chromist and alveolate phyla. The numbers at the nodes represent from top to bottom posterior probability values > 0.90 (BI) and bootstrap values $> 50\%$ received from analyses with MP and ME-LD methods, respectively.

Figure 2. Morphology and fine structure of *Telonema*. *Telonema subtilis* (a, b, c): **a**, Whole cell, scanning electron micrograph from natural sample (Gulf of Naples). **b**, Light micrograph of cultured cell (RCC404 from Roscoff); **c**, Longitudinal section of embedded material (RCC 404), showing the complex cytoskeleton (arrow), n = nucleus, fb = flagellar bases, fv = food vacuole. *Telonema antarcticum* (d, e, f): **d**, Whole cell, cultured material from the Oslofjord, showing cortical alveoli (arrow); **e**, Section through cortical alveoli (arrow) and cytoskeleton (arrowhead); **f**, Detail of flagellum with flagellar tubular tripartite hairs (arrow points to the terminal filament) as revealed by shadow cast whole-mount. **a, b, d** - scale bar 5 μm ; **c, e, f** - scale bar 1 μm .

Figure 3. The evolutionary origin of *Telonema* as deduced within the framework of the chromalveolate hypothesis^{7,11,27} acknowledging the monophyly of chromists and alveolates. The stramenopile/alveolate affiliation of *Telonema* in the molecular phylogeny is verified by the presence of typical chromist and alveolate traits in *T. antarcticum*, such as tripartite tubular hairs on flagella and the highly elaborated cytoskeleton; altogether suggesting that *Telonema* belongs to a very early chromalveolate lineage, possibly ancestral to chromalveolates⁷. The phylogenetic tree represents the commonly held view of the relationships between chromist and alveolate groups based on plastid sequences and morphology⁷ and therefore does not exactly resemble the nuclear 18S rRNA topology in Fig 1.

References

1. Moon-van der Staay, S. Y., De Wachter, R. & Vaultot, D. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**, 607-610 (2001).
2. Díez, B., Pedrós Alió, C. & Massana, R. Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl. Environ. Microbiol.* **67**, 2932-2941 (2001).
3. López-García, P., Rodríguez-Valera, F., Pedrós-Alió C. & Moreira, D. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**, 603-607 (2001).
4. Edgcomb, V. P., Kysela, D. T., Teske, A., de Vera Gomez, A. & Sogin, M. L. Benthic eukaryotic diversity in the Guaymas Basin hydrothermal vent environment. *Proc. Natl. Acad. Sci. U S A* **99**, 7658-7662 (2002).
5. Moreira, D. & López-García, P. The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends. Microbiol.* **10**, 31-38 (2002).
6. Griessmann, K. Über marine Flagellaten. *Arch. Protistenk.* **32**, 1-78 (1913).
7. Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297-354 (2002).
8. Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972-977 (2000).
9. Medlin, L. K., Kooistra, W. H. C. F., Potter, D., Saunders, G. W. & Andersen, R. A. in *Origins of algae and their plastids* (ed. Bhattacharya, D.) 237-261 (Springer., Wien., 1997).
10. Palmer, J. D. & Delwiche, C. F. in *Molecular systematics of plants II: DNA sequencing* (eds. Soltis, D. E., Soltis, P. S. & Doyle, J. J.) (Chapman-Hall, New York, 1998).
11. Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. Nuclear-encoded, plastid-targeted genes suggest a single common origin for Apicomplexan and dinoflagellate plastids. *Mol. Biol. Evol.* **18**, 418-426 (2001).
12. Van de Peer, Y., Baldauf, S. L., Doolittle, W. F. & Meyer, A. An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. *J. Mol. Evol.* **51**, 565-576 (2000).
13. Martin, W. et al. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U S A* **99**, 12246-12251 (2002).
14. Funes, S. et al. A green algal apicoplast ancestor. *Science* **298**, 2155 (2002).
15. Vaultot, D., Romari, K. & Not, F. Are autotrophs less diverse than heterotrophs in marine picoplankton? *Trends Microbiol.* **10**, 266-267 (2002).
16. Shalchian-Tabrizi, K., Klaveness, D., Thomsen, H. A. & Jakobsen, K. S. *Telonema antarcticum* sp.nov., a common marine phagotrophic flagellate within a new phylum. *Int. J. Syst. Evol. Microbiol.* submitted (2003).

17. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001).
18. Huelsenbeck, J. P. Testing a covariotide model of DNA substitution. *Mol. Biol. Evol.* **19**, 698-707 (2002).
19. Swofford, D. L. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) (Sinauer Associates, Inc. Publishers, Sunderland, MA, 1998).
20. Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. The single, ancient origin of chromist plastids. *Proc. Natl. Acad. Sci. U S A* **99**, 15507-15512 (2002).
21. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773-795 (1995).
22. Brown, R. M. & Romanovicz, D. K. Biogenesis and structure of Golgi-derived cellulosic scales in Pleurochrysis. I. Role of the endomembrane system in scale assembly and exocytosis. *Appl. Polym. Symp.* **28**, 537-585 (1976).
23. Sleight, M. A. in *The Cytoskeleton of Flagellate and Ciliate Protists* (eds. Melkonian, M., Anderson, R. A. & Schnepf, E.) 45-53 (Springer, Wien, 1991).
24. Inouye, I. in *Ultrastructure of microalgae* (ed. Berner, T.) 99-133 (CRC press inc., London, Tokyo, 1993).
25. Patterson, D. J. The Diversity of Eukaryotes. *Am. Nat.* **154**, S96-S124 (1999).
26. Hibberd, D. J. Observations of the ultrastructure of the flagella and periplast in the Cryptophyceae. *Br. Phycol. J.* **6**, 61-72 (1971).
27. Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. The single, ancient origin of chromist plastids. *Proc. Natl. Acad. Sci. U S A* **99**, 15507-15512 (2002).
28. Huttenlauch, I. & Stick, R. Occurrence of articulins and epiplasmins in protists. *J. Eukaryot. Microbiol.* **50**, 15-28 (2003).
29. Pace, N. R. A Molecular View of Microbial Diversity and the Biosphere. *Science* **276**, 734-740 (1997).
30. Romari, K. & Vaultot, D. Composition and Temporal Variability of Picoeukaryote Communities at a Coastal Site of the English Channel Using Molecular Approaches. *In preparation*.
31. Moon-van der Staay, S. Y., van der Staay, G. W. M., Guillou, L., Vaultot, D., Claustre, H. & Medlin, L. K. Abundance and diversity of prymnesiophytes in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences. *Limnol. Oceanogr.* **45**, 98-109 (2000).
32. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids. Res.* **25**, 4876-4882 (1997).

33. Eikrem, W. & Moestrup, O. Structural analysis of the flagellar apparatus and the scaly periplast in *Chrysochromulina scutellum* sp. nov. (Prymnesiophyceae, Haptophyta) from the Skagerrak and the Baltic. *Phycologia* **37**, 132-153 (1998).

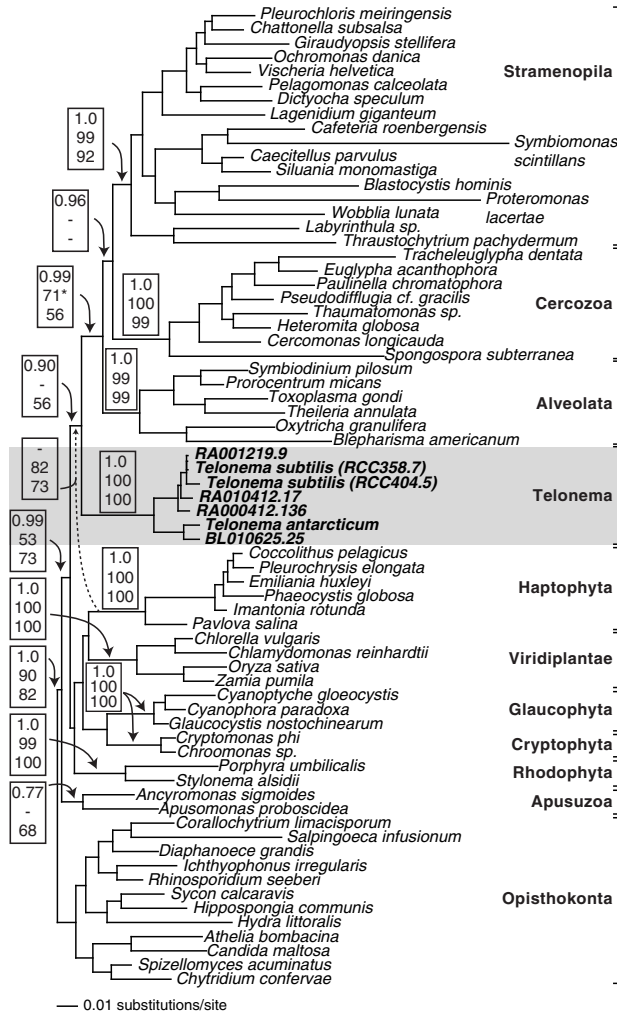


Fig 1

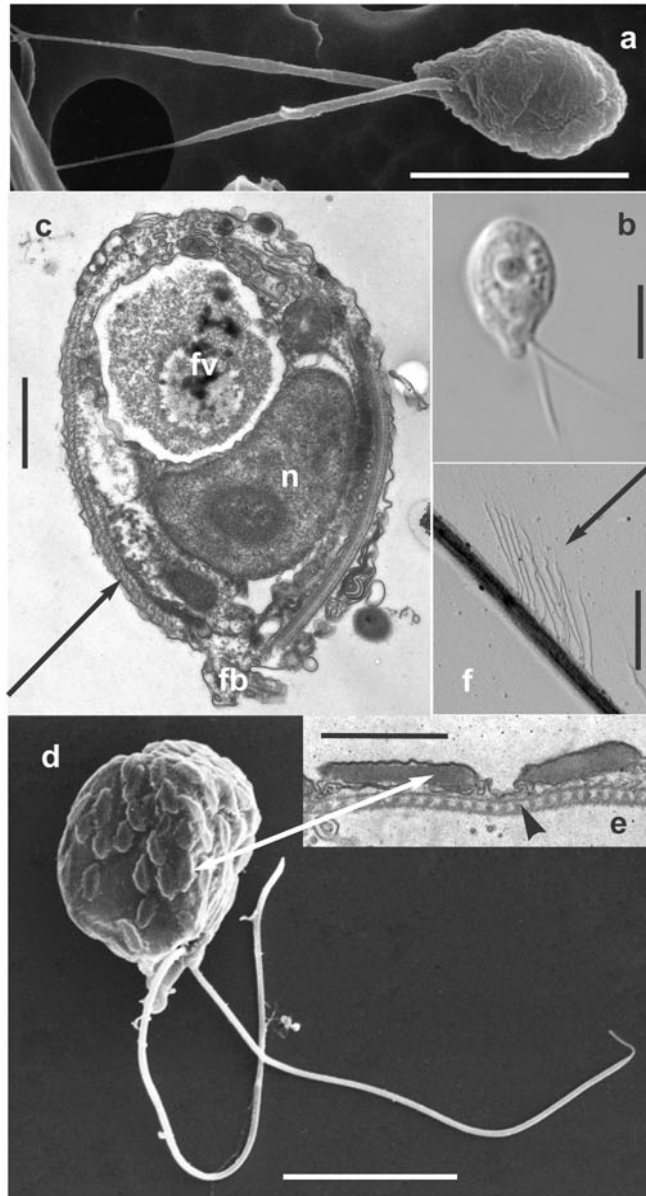


Fig 2

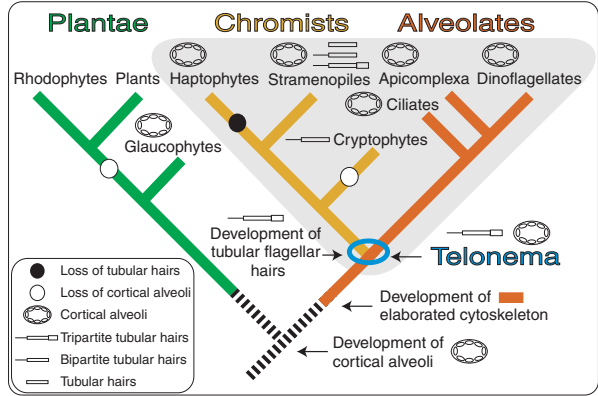


Fig 3

Supplementary information

Content:

1. Phylogenetic methods.
2. Model tests undertaken in a Bayesian framework.
3. References.
4. Table with all accession numbers and species names of sequences used in the molecular phylogeny.
5. Supplementary tree 1 (Bayesian inference done with $\Gamma+I$ model).
6. Supplementary tree 2 (Maximum likelihood tree inferred with $\Gamma+I$ model).

1. Phylogenetic analysis

Bayesian inferences were undertaken with the program MrBayes v2.01². Variable rates across sites was accommodated by allowing each site to evolve with a rate defined as a random variable drawn from a gamma distributed rate probability (Γ). Four discrete rate categories were used to approximate the gamma distribution of rates across sites. This parameter were used in three analyses in combination with either a proportion of invariable sites parameter ($\Gamma+I$) (see supplementary tree 1), a covariotide substitution model that allow sites to convert between invariable and variable modus; i.e off>on and on>off ($\Gamma+cov$), and finally together with both a constant proportion of invariable site parameter and the covariotide model ($\Gamma+I+cov$). In addition, the general time reversible model (GTR) was implemented. Priors for all other model parameters were set to default values. Metropolis coupling was used with three heated (temperature parameter 0.2) and one cold chain. Randomly generated trees were used as starting point for Marchov chain Monte Carlo chains that carried out for 1,000,000 generations. Sampling of trees was done every 100 generations for a total of 10,000 trees. Burn-in of trees (i.e. sampled before the MCMC chains reached convergence) was set to 3,000 trees based on assessment of the likelihood plots. Consensus of the remaining 7,000 trees was used to calculate the posterior probabilities of the clades. Two separate runs were performed to confirm the stationarity of the chains; the clade probabilities were very similar between the independent runs supporting our conclusion that the chains produced a reasonable sample from the posterior distribution within the chosen burn-in period.

In addition, phylogenetic analyses were performed with maximum likelihood (ML), maximum parsimony (MP) and minimum evolution (ME) criteria by using PAUP*¹. The ML tree was inferred by using an evolutionary model estimated from a tree generated with the Neighbor joining algorithm and Kimura 2-parameter (K2P) distances, and included GTR+ Γ +I

parameters (supplementary figure 2). The ME tree was calculated on the basis of LogDet distances with the use of proportion of invariable sites parameter estimated with ML criteria from a K2P tree, and the number of invariable characters was removed according to base frequencies estimated from the constant sites only. Inference with MP criteria was performed by reweighting characters according to the rescaled consistency index. All ML, MP and ME trees were reconstructed by 10 heuristic searches, random addition of sequences and tree tree-bisection-reconnection branch swapping. Bootstrap analyses were done on 100 pseudoreplicates with MP and ME criteria and tree-search methods as in analysis of the original data sets.

2. Model testing

The Bayes factor, defined as the posterior probability of a hypothesis given that the prior probabilities of the alternative evolutionary models are equal, was calculated as the ratio of marginal likelihood values (i.e. the harmonic mean value) obtained from stationary phases of MCMC runs, and calibrated to frequentist statistics by applying the Schwarz criterion³. Since the compared models are nested, the Schwarz criterion implies that $2 \log B_{10} \approx t^2 - (d_1 - d_2) \log(n)$, where B_{10} is the Bayes factor for H_1 against H_0 , t is the t statistics for frequentist testing of the models, $d_1 - d_2$ the differences of degree of freedom and n is the size of the sequence alignments. The approximate t value corresponds to different grades of evidence. The statistical tests have been performed hierarchically in two steps (i.e. $\Gamma + I$ vs $\Gamma + \text{cov}$ and $\Gamma + \text{cov}$ vs $\Gamma + I + \text{cov}$), where each comparison were carried out between models that only differ in one degree of freedom, resulting in $(d_1 - d_2) = 1$. It implies that for «decisive» evidence the minimum t value can be estimated with the formula $t^2 = \log(n) + 10^3$. In this work the alignment consists of 1387 characters, thus the approximate minimum t value that favour the H_1 model in each test in this work is $t = 4.15$.

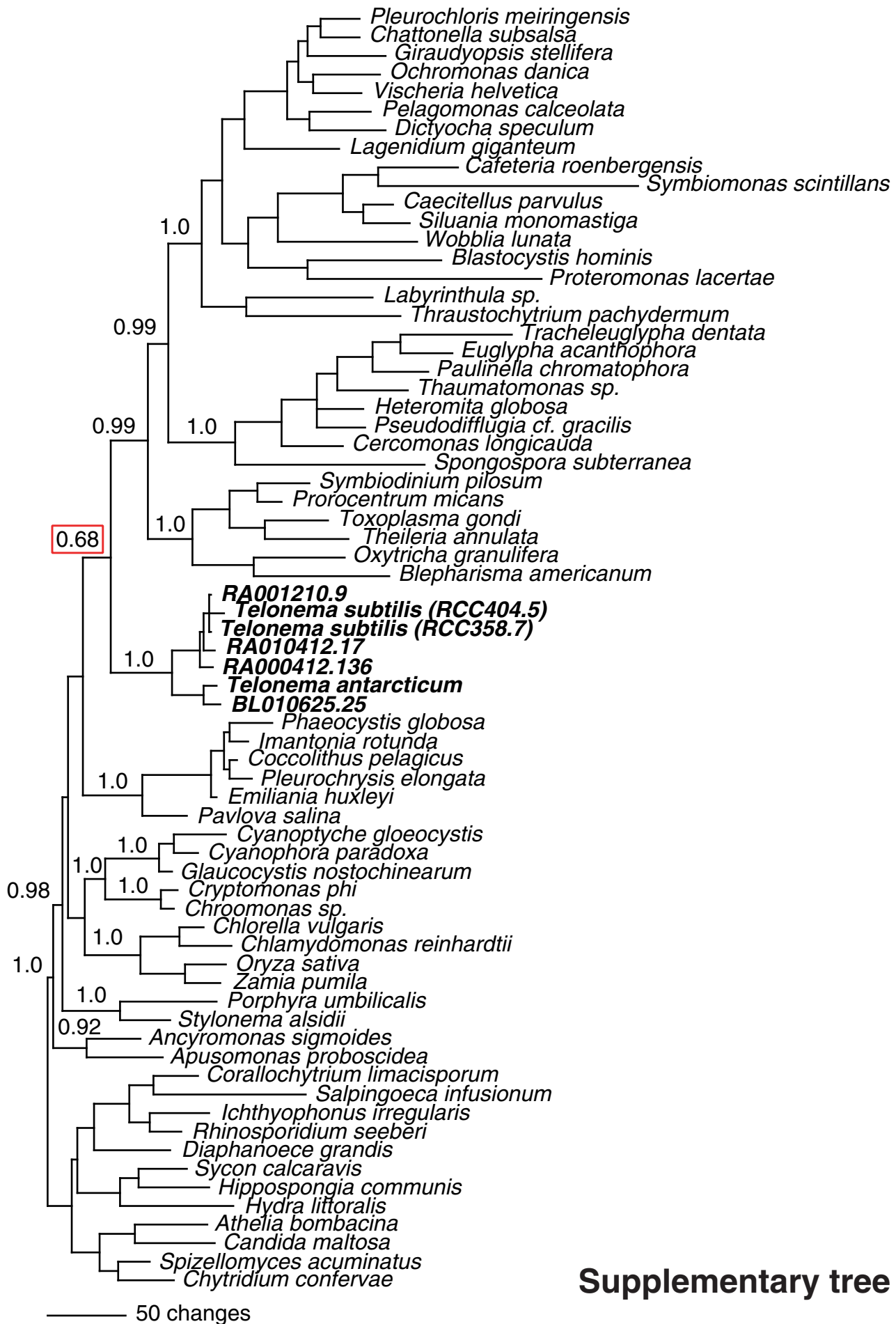
3. References

1. Swofford, D. L. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) (Sinauer Associates, Inc. Publishers, Sunderland, MA, 1998).
2. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001).
3. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773-795 (1995).

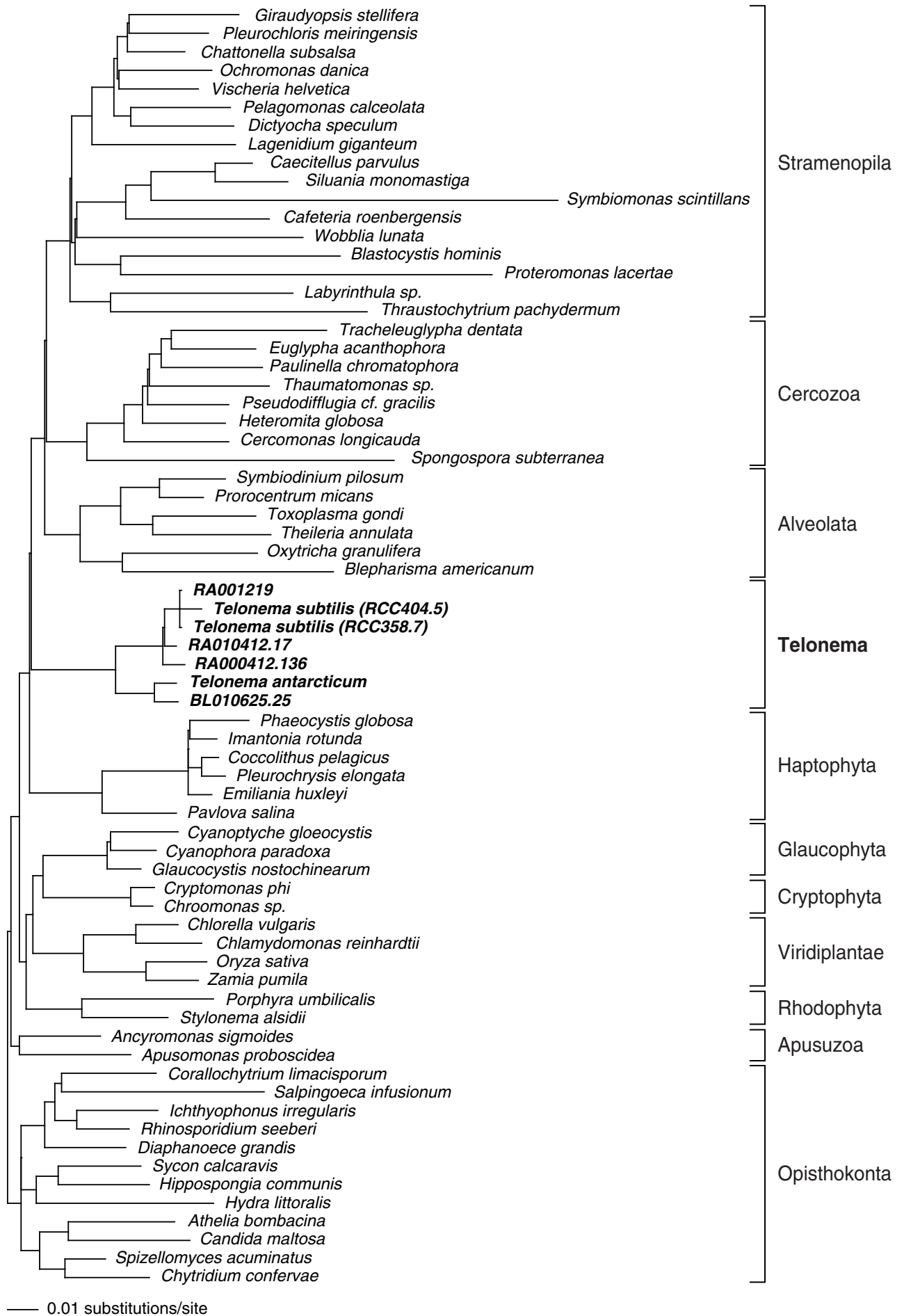
4. Table with accession numbers and species names included in the phylogenetic analyses:

<i>Ancyromonas sigmoides</i>	AF174363.1
<i>Apusomonas proboscidea</i>	L37037.1
<i>Athelia bombacina</i>	M55638.1
<i>Blastocystis hominis</i>	AB023499.1
<i>Blepharisma americanum</i>	M97909.1
<i>Caecitellus parvulus</i>	AF174368.1
<i>Cafeteria roenbergensis</i>	AF174364.1
<i>Candida maltosa</i>	D14593.1
<i>Cercomonas longicauda</i>	AF101052.1
<i>Chattonella subsalsa</i>	U41649.1
<i>Chlamydomonas reinhardtii</i>	M32703.1
<i>Chlorella vulgaris</i>	X13688.1
<i>Chroomonas sp.</i>	X81328.1
<i>Chytridium confervae</i>	M59758.1
<i>Coccolithus pelagicus</i>	AJ246261.1
<i>Corallochytrium limacisporum</i>	L42528.1
<i>Cryptomonas phi</i>	U53126.1
<i>Cyanophora paradoxa</i>	X68483.1
<i>Cyanoptyche gloeocystis</i>	AJ007275.1
<i>Diaphanoeca grandis</i>	L10824.1
<i>Dictyocha speculum</i>	U14385.1
<i>Emiliana huxleyi</i>	AF184167.1
<i>Euglypha acanthophora</i>	AJ418788.1
<i>Giraudyopsis stellifera</i>	U78034.1
<i>Glaucozystis nostochinearum</i>	X70803.1
<i>Guillardia theta</i>	X57162.1
<i>Heteromita globosa</i>	U42447.1
<i>Hippospongia communis</i>	AF246616.1
<i>Hydra littoralis</i>	U32392.1
<i>Ichthyophonus irregularis</i>	AF232303.1
<i>Imantonia rotunda</i>	AJ246267.1
<i>Labyrinthula sp.</i>	AB022105.1
<i>Lagenidium giganteum</i>	X54266.1
<i>Ochromonas danica</i>	M32704.1
<i>Oryza sativa</i>	AF069218.1
<i>Oxytricha granulifera</i>	X53486.1
<i>Paulinella chromatophora</i>	X81811.1
<i>Pavlova salina</i>	L34669.1
<i>Pelagomonas calceolata</i>	U14389.1
<i>Phaeocystis globosa</i>	X77476.1
<i>Pleurochloris meiringensis</i>	AF109728.1
<i>Pleurochrysis elongata</i>	AJ246264.1
<i>Porphyra umbilicalis</i>	L36049.1
<i>Prorocentrum micans</i>	M14649.1
<i>Proteromonas lacertae</i>	U37108.1
<i>Pseudodifflugia cf. gracilis</i>	AJ418794.1
<i>Rhinosporidium seeberi</i>	AF399715.1
<i>Salpingoeca infusionum</i>	AF100941.1
<i>Siluania monomastiga</i>	AF072883.1
<i>Spizellomyces acuminatus</i>	M59759.1
<i>Spongospora subterranea</i>	AF245217.1
<i>Stylonema alsidii</i>	L26204.1
<i>Sycon calcaravis</i>	D15066.1
<i>Symbiodinium pilosum</i>	X62650.1

<i>Symbiomonas scintillans</i>	AF185052.1
<i>Thaumatomonas sp.</i>	U42446.1
<i>Theileria annulata</i>	M64243.1
<i>Thraustochytrium pachydermum</i>	AB022113.1
<i>Toxoplasma gondii</i>	L37415.1
<i>Tracheleuglypha dentata</i>	AJ418790.1
<i>Vischeria helvetica</i>	AF045051.1
<i>Wobblia lunata</i>	AB032606.1
<i>Zamia pumila</i>	M20017.1



Supplementary tree 1



Supplementary tree 2