Supplementary Material

Cultures

T. subtilis was grown with *Imantonia rotunda* as food source in Erdschreiber medium (Foyn 1934) at 16ºC for 14 days under light/dark cycles of 14h/10h, respectively, and finally grown in dark for periods of 5-10 days to maximize the *Telonema* cells from the culture. About 2 L of *T. subtilis* culture were harvested by centrifugation at 4000 rpm for 10 min and flash frozen in liquid nitrogen. *R. contractilis* was monoxenically grown with the green alga *Chlorogonium elongatum* as food source as described earlier (Sakaguchi and Suzaki 1999). The cells were cultured at 20ºC for 14 days under light/dark periods of 14h/10h, respectively. After confirming that most of the *C. elongatum* cells were consumed by *R. contractilis*, ~0.9 liters of culture was collected by centrifugation at 500g for 3 min and the cells were transferred into RNAlater solution (Ambion, Austin, USA). *Plagioselmis nannoplanctica* culturing will be described elsewhere as part of an independent project involving *Collodictyon*.

cDNA library construction and 454 pyrosequencing

Normalized cDNA libraries were constructed by Vertis Biotechnology AG (Germany) according to their Random-Primed (RPD) cDNA protocol. Frozen cells were ground under liquid nitrogen and total RNA isolated from the cell powder using the mirVana RNA isolation kit (Ambion). Poly(A)+ RNA was prepared from total RNA. First-strand cDNA synthesis was primed with a N6 randomized primer and second-strand cDNA was synthesized according to the classical Gubler-Hoffman protocol (Gubler and Hoffman 1983). Double stranded DNA (dsDNA) was blunted and 454 adapters A and B ligated to the 5' and 3' ends. dsDNA carrying both adapter A and adapter B attached to its ends was selected and amplified with PCR using a proof reading enzyme (24 cycles). To ensure a reduction in highly expressed genes, an equalization of the gene representation was performed with a method developed by Vertis Biotechnology. For 454 sequencing the cDNA in the size range of 250 – 600 bp was eluted from a preparative agarose gel. Half a plate of a GSFLX instrument (Standard chemistry) was sequenced for *T. subtilis* by the Norwegian ultra-high throughput sequencing service unit at the University of Oslo, yielding about 210,000 reads. For *R. contractilis*, half a plate was sequenced by Macrogen Inc (South-Korea) generating about 360,000 reads.

Bootstrapping of genes analyses

Bootstrapping analyses in which the genes instead of the sites are randomly sampled from the total number of genes were performed. We constructed 200 replicates, each containing 127 concatenated genes, which resulted in supermatrices of variable length (ranging from 25,045 aa to 33,017aa). This procedure was repeated on 2 sets of species, that is with *T. subtilis* and *R. contractilis* included (**Supp. Figure 1**) and excluded (**Supp. Figure 2**). Overall, the relationships we obtained were very similar to those based on the original alignment. In particular, this approach confirmed the evolutionary affinities of telonemids, centrohelids et haptophytes. In **Supp. Figure 1**, cryptomonads weakly branched together with some excavates (Discoba, Hampl et al. 2009). After removing *T. subtilis* and *R. contractilis*, cryptomonads were placed again as sister to haptophytes (**Supp. Figure 2**). Importantly, we could not detect any clear correlation between the inferred evolutionary relationships and the length of the alignments or the combination of genes that were sampled in each replicate. Hence, the bootstrapping of genes approach did not reveal any obvious conflict in the phylogenetic signal in the data.

Phylogenetic analyses after removing both, or one of T. subtilis or R. contractilis

In order to see the impact on the topology and supports of *T. subtilis* and *R. contractilis*, analyses were performed using concatenated datasets that did not contain these species (**Supp. Figure 3**), as well as with one or the other removed (**Supp. Figure 4 and 5**). In all cases the major groups of eukaryotes were recovered as in Figure 1 and the relationships among them were very consistent. Interestingly, we observed more robust support for the association between cryptomonads and haptophytes (corresponding to node 1 in Figure 1, see the main text) and the sister position of this grouping to SAR (node 2 in Figure 1) when both species were not included (**Supp. Figure 3**). This is consistent with our interpretation of the ancient origin of telonemids and centrohelids. Indeed, if relatively few sequence synapomorphies accumulated during a brief period of shared common ancestry with cryptomonads and haptophytes, and even fewer now remain following hundreds of millions of years of divergence, one expects that *T. subtilis* and *R. contractilis* will randomly branch elsewhere in the tree, thus lowering the statistical support for the whole CCTH/CCTH-SAR groups.

A “separate” analysis was also conducted on the dataset that lacked *T. subtilis* and *R.*

contractilis. Here we specifically examined the relationships among 5 major groups — (1) cryptomonads plus haptophytes, (2) SAR group, (3) Plantae, (4) excavates, and (5) unikonts (opisthokonts + Amoebozoa). 123 genes (amounting to a total of 28'166 aa), which contained at least one representative taxon for each group of interest, were selected from the 127 genes used in the concatenation. The best tree was identical to the Bayesian and ML analyses of the supermatrix and RELL values were consistently higher than those on Figure 1 (**Supp. Figure 3)**.

When *R. contractilis* was removed from the alignment in isolation (**Supp. Figure 4**), *T. subtilis* branched within a clade also including cryptomonads and haptophytes (CTH group), and this group was sister to SAR. The Bayesian and ML approaches gave two different unsupported topologies for the position of *T. subtilis* within the CTH group, a poor resolution that was also observed in Figure 1. This means that adding *R. contractilis*, another enigmatic lineage that likely diverged soon after the origin of the CCTH-SAR grouping, did not help in recovering a good support for placing the telonemids. When *T. subtilis* was removed to see how *R. contractilis* alone influenced the results (**Supp. Figure 5**), we again recovered the same major eukaryotic groups and relationships, notably an assemblage enclosing cryptomonads, haptophytes, and centrohelids (CCH group) and its sister position to SAR. Consistent with our previous observations, the CCH group and the CCH-SAR relationship received in this analysis the lowest ML support of all analyses (60% BP), indicating once more that only weak phylogenetic signal remains in centrohelids sequences (probably less than in telonemids). It is worth noting that this lack of sequence synapomorphies resulting in poor phylogenetic signal is likely the main reason for the many unsuccessful attempts to place centrohelids in the tree of eukaryotes until this study (Cavalier-Smith and Chao 2003; Sakaguchi et al. 2005; Cavalier-Smith and von der Heyden 2007; Sakaguchi et al. 2007).

Topology comparisons based on the supermatrices

To better assess the phylogenetic position of *T. subtilis* and *R. contractilis*, we conducted topology comparisons using the approximately unbiased (AU) test. For each tested tree, per-site log likelihoods for the supermatrices were calculated using RAxML (Stamatakis 2006) and the AU tests were performed using CONSEL (Shimodaira and Hasegawa 2001) with default scaling and replicate values. The test trees were constructed by using the Bayesian and ML topologies shown in Figure 1, Supp. Figure 2 and Supp. Figure 3 and plac-

ing *T. subtilis* and *R. contractilis* on different branches (we did not test positions within monophyletic groups that received maximal supports) (**Supp. Figure 6, A-E**). These analyses generally confirmed the trends observed in the tree reconstructions, that is: (1) alternative positions for *T.subtilis* cannot be rejected, but only if placed within or sister to the CCTH or CTH groups; an exception was a non-rejected position on the branch leading to the red algae when *R. contractilis* was absent from the alignment (**Supp. Figure 6D**), but this branching was discarded when *R. contractilis* was present, underlying the importance of taxon-sampling (**Supp. Figure 6A and B**). (2) Alternative positions for *R. contractilis* even outside that of the CCTH-SAR groupings were kept in the set of plausible trees, precisely as sister to or within the excavates (**Supp. Figure 6B and E**); a sister relationship to the red algae was also accepted when *T. subtilis* was absent (**Supp. Figure 6E**), but it was similarly rejected when both species were analyzed together (**Supp. Figure 6 A and B**). Altogether these analyses suggest once more the early origin of telonemids and centrohelids; because several deep branchings could not be rejected for centrohelids, it is possible that this group diverged even earlier.

The AU tests retained in the pool of candidate trees a relationship between *T. subtilis* (or *R. contractilis*) and red algae when only one of these two species was considered. Although no obvious relationship with red algae was found in our single-gene tree reconstructions, this signal could be explained by genes of red origin that were transferred from a red algal endosymbiont to the nucleus in the ancestor of CCTH-SAR if the chromalveolate hypothesis is correct (Lane and Archibald 2008). In this context, it is interesting to note that a relationship between *R. contractilis* and red algae was previously observed on 18S rRNA (Cavalier-Smith and Chao 2003) and α- and β-tubulin phylogenies (Sakaguchi et al. 2005) and was not rejected in an analysis of six housekeeping genes (Sakaguchi et al. 2007). Yet neither of these genes has been shown to have a red algal ancestry in chromalveolate species.

References

Cavalier-Smith T, Chao EE. 2003. Molecular phylogeny of centrohelid heliozoa, a novel lineage of bikont eukaryotes that arose by ciliary loss. J Mol Evol. 56:387-396.

Cavalier-Smith T, von der Heyden S. 2007. Molecular phylogeny, scale evolution and taxonomy of centrohelid heliozoa. Mol Phylogenet Evol. 44:1186-1203.

Foyn B. 1934. Lebenszyklus, cytology und sexualitat der Chlorophyceae Cladopora subriana Kützing. Arch Protistenkd. 83:1-56.

Gubler U, Hoffman BJ. 1983. A simple and very efficient method for generating cDNA libraries. Gene. 25:263-269.

Hampl V, Hug LA, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ. 2009. Phylogenomic Analyses Support the Monophyly of Excavata and Resolve Relationships among Eukaryotic "Supergroups". Proc Natl Acad Sci U S A. 106:3859-3864.

Lane CE, Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. Trends Ecol Evol. 23:268-275.

Sakaguchi M, Inagaki Y, Hashimoto T. 2007. Centrohelida is still searching for a phylogenetic home: analyses of seven Raphidiophrys contractilis genes. Gene. 405:47-54.

Sakaguchi M, Nakayama T, Hashimoto T, Inouye I. 2005. Phylogeny of the Centrohelida inferred from SSU rRNA, tubulins, and actin genes. J Mol Evol. 61:765-775.

Sakaguchi M, Suzaki T. 1999. Monoxenic culture of the heliozoon Actinophrys sol. Eur J Protistol. 35:411-415.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246-1247.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688-2690.

Supp. Figure 1 and Figure 2

Cladograms representing the majority rule consensus trees resulting from the bootstrapping of genes analyses. In Supp. Figure 1 *T. subtilis* and *R. contractilis* were included; these 2 species were excluded from the analysis resulting in Supp. Figure 2. Black dots are on nodes defining groups that were recovered in all 200 replicates (100%); when not maximal the percentages of trees recovering the groups are indicated. The white thick bars are the groups that were originally included in the chromalveolates. Assemblages indicated by capitalized names correspond to the hypothetical supergroups of eukaryotes.

Supp. Figure 3

Phylogeny summarizing the relationships among the major groups of eukaryotes when *T.subtilis* and *R.contractilis* are not included in the analysis. This tree was obtained with phylobayes ran under the CAT model (consensus between two independent Markov chains), and subsequently schematized in FigTree (http://tree.bio.ed.ac.uk/software/figtree/) with the "Cartoon" option. Black dots correspond to 1.0 posterior probability (PP) and 100% ML bootstrap (BP), otherwise values at node represent PP (above) and BP (below) when not maximal. Black squares show the constrained bifurcations used in the separate analysis and RELL bootstraps (RBP) are indicated.

Supp. Figure 4 and Figure 5

These trees represent a Bayesian phylogeny of eukaryotes, obtained from the consensus between two independent Markov chains, run under the CAT model implemented in phylobayes. The curved dashed lines indicate the alternative branchings recovered in the ML analysis of the same dataset. Black dots correspond to 1.0 posterior probability (PP) and 100% ML bootstrap (BP), otherwise values at node represent PP (above) and BP (below) when not maximal. The white thick bars are the groups that were originally included in the chromalveolates. Assemblages indicated by capitalized names correspond to the hypothetical supergroups of eukaryotes. The scale bar represents the estimated number of amino acid substitutions per site.

Supp. Figure 6

Summary of the AU tests based on the concatenated alignments, showing the alternative branching points that were tested (numbers on branches) and the *P*-values higher than 0.05. The values in circles correspond to the positions that were not rejected by the AU tests. When both *T. subtilis* and *R. contractilis* were present, only one species was moved at a time leaving the other in its inferred position. (A) Bayesian tree as in Figure 1, *T. subtilis* or *R. contractilis* were successively placed on alternative branches; (B) ML tree as in Figure 1, *T. subtilis* or *R. contractilis* were successively placed on alternative branches; (C) Bayesian tree as in Figure 1, both *T. subtilis* and *R. contractilis* were successively placed on alternative branches; (D) Bayesian tree as in Supp. Figure 2, *T. subtilis* was successively placed on alternative branches; (E) Bayesian tree as in Supp. Figure 2, *R. contractilis* was successively placed on alternative branches. Ma: *Malawimonas*; Tr: *Trimastix*; Di: *Discoba*; Re: *Red algae*; Gr: *Green algae*; Gl: *Glaucophytes*; Cr: *Cryptomonads*; Ha: *Haptophytes*; Te: *T.subtilis*; Ra: *R.contractilis*; Un: *Unikonts*.