

ORIGINAL PAPER

Polyubiquitin Insertions and the Phylogeny of Cercozoa and Rhizaria

David Bass^{a,1}, David Moreira^b, Purificación López-García^b, Stephane Polet^c, Ema E. Chao^a, Sophie von der Heyden^a, Jan Pawlowski^c, and Thomas Cavalier-Smith^a

^aDepartment of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

^bUnité d'Ecologie, Systématique et Evolution, UMR CNRS 8079, Université Paris-Sud, Bâtiment 360, 91405 Orsay CEDEX, France

^cDepartment of Zoology and Animal Biology, University of Geneva, CH-1211 Geneva 4, Switzerland

Submitted July 8, 2004; Accepted March 26, 2005
Monitoring Editor: Michael Melkonian

A single or double amino acid insertion at the monomer–monomer junction of the universal eukaryotic protein polyubiquitin is unique to Cercozoa and Foraminifera, closely related ‘core’ phyla in the protozoan infrakingdom Rhizaria. We screened 11 other candidate rhizarians for this insertion: Radiozoa (polycystine and acanthorean radiolaria), a ‘microheliozoan’, and Apusozoa; all lack it, supporting suggestions that Foraminifera are more closely related to Cercozoa than either is to other eukaryotes. The insertion’s size was ascertained for 12 additional Cercozoa to help resolve their basal branching order. The earliest branching Cercozoa generally have a single amino acid insertion, like all Foraminifera, but a large derived clade consisting of all Monadofilosa except *Metopion*, *Helkesimastix*, and *Cercobodo agilis* has two amino acids, suggesting one doubling event and no reversions to a single amino acid. *Metromonas* and *Sainouron*, cercozoans of uncertain position, have a double insertion, suggesting that they belong in Monadofilosa. An alternative interpretation, suggested by the higher positions for *Metopion* and *Cercobodo* on Bayesian trees compared with most distance trees, cannot be ruled out, i.e. that the second insertion took place earlier, in the ancestral filosan, and was followed by three independent reversions to a single amino acid in *Chlorarachnea*, *Metopion* and *Cercobodo*.

© 2005 Elsevier GmbH. All rights reserved.

Key words: Cercozoa; Rhizaria; polyubiquitin insertion; Radiozoa; Foraminifera.

Introduction

Recent advances in eukaryote phylogenetic reconstruction stem from progress in four areas: novel organisms being cultured, thereby providing raw materials for genetic and morphological

analysis; increased choice and sophistication of analytical techniques; basing phylogenetic analyses on several or many genes (e.g. Baldauf et al. 2000; Baptiste et al. 2002; Martin et al. 1998); and refining sequence-based trees with data from molecular cladistic characters, which can be highly resolving (Archibald et al. 2003; Stechmann and Cavalier Smith 2002, 2003), although such

¹Corresponding author;
fax +44 1865 281310
e-mail david.bass@zoo.ox.ac.uk (D. Bass)

data should be treated cautiously as indel characters may be homoplastic, for example by convergence or reversion (Baptiste and Philippe 2002). Increasingly many previously unclassified taxa are being placed into known groups, and misleading classifications and taxon designations arising from single-gene trees are being corrected (e.g. Cavalier-Smith, 2003a, 2004; Philippe and Adoutte 1998; Philippe et al. 2000). Consequently, the eukaryote tree is resolving into a topology with a rather small number of internally diverse taxonomic 'supergroups' (Berney et al. 2004; Cavalier-Smith 2003a, 2004). One of these, the biciliate infrakingdom Rhizaria (Cavalier-Smith 2002), has recently been the subject of several studies (Burki et al. 2002; Cavalier-Smith 2004; Cavalier-Smith and Chao 2003a–c; Nikolaev et al. 2003, 2004), which have sought to determine its member phyla, their interrelationships, and its closest relatives.

The two 'core rhizarian' phyla are the recently established Cercozoa (Bass and Cavalier-Smith 2004; Cavalier-Smith 1998; Cavalier-Smith and Chao 2003c) and the long-known Foraminifera (reviewed in Pawlowski et al. 2002). A common origin of Cercozoa and Foraminifera was initially suggested by actin trees (Keeling 2001) and by the marked propensity of both groups to form reticulopodia (Cavalier-Smith 2002). Support for this relationship was strengthened by the discovery that both phyla carry an insertion of one or two amino acid residues at the monomer–monomer junctions of the polyubiquitin gene (Archibald et al. 2003; Archibald and Keeling 2004), and by RNA polymerase II phylogenies (Longet et al. 2003).

Ubiquitin is a 76 aminoacid protein found in all eukaryotes, rivalling histone 4 in its degree of conservatism (Sharp and Li 1987). Ubiquitin genes occur in three main forms: single open reading frames, fused to ribosomal protein genes, or as head-to-tail multimers of ubiquitin coding regions (polyubiquitin). The protein is essential to many basic processes in eukaryotic cells, for example apoptosis, signal transduction, DNA repair, stress response, transcriptional regulation, endocytosis, and cell cycle regulation (Finley and Chau 1991; Hershko and Ciechanover 1998; Sun and Chen 2004).

Members of Haeckel's Radiolaria also belong in Rhizaria, even though Radiolaria itself is now shown to be polyphyletic (Polet et al. 2004). Phaeodarea (e.g. *Coelodendrum*, *Aulosphaera*, *Aulacantha*) are actually monadofilosan cercozoans (Polet et al. 2004), probably most closely related to the order Tectofilosida (Bass and

Cavalier-Smith 2004), while Acantharea and Polycystinea often group together (López-García et al. 2002). The latter grouping constitutes phylum Radiozoa (Cavalier-Smith 1993, despite the removal of Phaeodarea: Cavalier-Smith 2004). Recent molecular studies (Nikolaev et al. 2004) also support the grouping of Sticholonchea with Acantharea as the subphylum Spasmaria within the Radiozoa (Cavalier-Smith 1993), confirming that the previous placement of *Sticholonche* in Heliozoa (Febvre-Chevalier 1990) was erroneous. Radiozoa are considered to be rhizarians on the basis of a shared propensity to make reticulopodia and their usual grouping with Foraminifera and Cercozoa on rRNA trees (Cavalier-Smith and Chao 2003b; López-García et al. 2002; Nikolaev et al. 2004; Polet et al. 2004).

However, the branching order of Cercozoa, Foraminifera, and Radiozoa is not well resolved. On 18S rRNA trees that include all three groups, Foraminifera often branch within or as sister to Radiozoa (Cavalier-Smith and Chao 2003b,c), or sometimes as sister to the cercozoan haplosporidia (Nikolaev et al. 2004). When Radiozoa are not included, Foraminifera branch as sister to the cercozoan *Gromia oviformis* (Berney and Pawlowski 2003; Longet et al. 2004; Nikolaev et al. 2003). In the protein (actin, RPB1) trees, in the absence of Radiozoa, Foraminifera may branch within core Cercozoa (Filosa) (Keeling 2001), or as sister to plasmodiophorids (Archibald and Keeling 2004) or to *Gromia* (Longet et al. 2003; 2004). However, when radiolarian actin was included, one of the foraminiferan actin paralogues branched with Radiozoa (Nikolaev et al. 2004). Here we sequence polyubiquitin genes from several acantharea and polycystines and an additional foraminiferan in order to decide whether Foraminifera are closer to Radiozoa or Cercozoa.

Two other groups originally but more doubtfully included in Rhizaria are centrohelid heliozoa and Apusozoa (Cavalier-Smith 2002). Recent 18S rDNA phylogenetic analyses, in combination with ultrastructural data, have led to these being omitted from the current circumscription of Rhizaria (Cavalier-Smith 2003a, 2004; Cavalier-Smith and Chao 2003a,b). However, in view of their historical association with Rhizaria, it was important to check whether polyubiquitin structure is consistent with this exclusion. We have therefore also screened two members of Apusozoa and a marine microheliozoan (a probable sister to centrohelids: Cavalier-Smith and Chao 2003a) in order to see if they have polyubiquitin insertions or not.

As some Cercozoa have one and others two amino acids in the polyubiquitin insertion, this character could in principle be used to partition Cercozoa cleanly into two groups if the number of amino acids was doubled only once and has never undergone reversion. This would be particularly valuable phylogenetically and taxonomically as the basal branches of existing cercozoan rDNA trees are already known to be poorly resolved, especially within the subphylum Filosa (Cavalier-Smith and Chao 2003c). Present evidence suggested that the cercozoan subphylum Endomyxa has an insertion of a single amino acid like Foraminifera (which may branch within it on sparsely sampled trees) and that all studied Filosa except for chlorarachnean algae have an insertion of two amino acids (Archibald et al. 2003; Archibald and Keeling 2004). Therefore, we also sequenced polyubiquitins from a much larger variety of filosan Cercozoa, to see which lineages have insertions of one or two amino acids and whether the evolutionary transition between the two types is indeed sufficiently rare to be a valuable phylogenetic marker within Filosa.

Results

Figure 1 shows the polyubiquitin sequences of 59 taxa, including numerous cercozoans. It can be seen that all Cercozoa have an insertion of either one or two amino acids compared with most eukaryotes. By contrast all Foraminifera have a single amino acid insertion, except for one species, where two different genes were found, two copies of one with a single insertion, which we consider the authentic foraminiferan sequence, and a single copy with no insertion that is likely to be a contaminant. None of the new sequences from Radiozoa, Apusozoa, and Heliozoa contained inserts.

To enable us to interpret the insertion pattern within Cercozoa, especially of *Cercobodo agillis* for which no previous sequences were available, we constructed more comprehensive 18S rDNA trees for Cercozoa (Figs 2,3).

Discussion

We have now shown that the unnamed marine microheliozoan, two Apusozoa, and five Radiozoa (two acanthareans and three polycystineans) all lack amino acid insertions between the ubiquitin monomers of the polyubiquitin gene. On the simplest interpretation, which assumes that this

insertion once gained has never been lost, the absence of the insertion in these two groups suggests that Foraminifera and Cercozoa are more closely related to each other than they are to any taxa lacking an insertion (Fig. 4). Because we cannot rule out the possibility that the insertion might sometimes be lost, other evidence is needed to confirm this probable closer relationship of Foraminifera and Cercozoa.

The Evolutionary Position of Radiozoa

If it is true that the insertion has never been lost in any group, then the frequent grouping of Foraminifera with Polycystinea on 18S rDNA trees (Cavalier-Smith and Chao 2003b,c) is most likely a long-branch attraction artifact (Embley and Hirt 1998; Philippe and Adoutte 1998; Philippe et al. 2000). Initially, that grouping of Radiozoa and Foraminifera seemed to support the classification of both groups as the supergroup Retaria (Cavalier-Smith 1999). The discovery that Phaeodarea are cercozoans (Polet et al. 2004) has shown that the taxon Retaria as originally conceived (Cavalier-Smith 1999) was polyphyletic. Therefore, Foraminifera and Radiozoa (modified by excluding Phaeodarea: Cavalier-Smith 2003a) are again treated as distinct phyla and Retaria not used as a formal taxon (Cavalier-Smith 1993, 2004). Unless the polyubiquitin insertion was lost in Radiozoa, our results indicate that even omitting Phaeodarea, Retaria would not be holophyletic.

The term 'core rhizaria' was coined for Cercozoa, Foraminifera, and Radiolaria, all characterized by a propensity to form reticulopodia and/or filopodia and all grouping together on rDNA trees (Cavalier-Smith and Chao 2003b). However, with the exclusion of Apusozoa from Rhizaria this term would become synonymous with Rhizaria, which currently comprises three phyla: Cercozoa, Foraminifera, and Radiozoa (Cavalier-Smith 2004), and is very probably a clade (Fig. 4). We suggest that it is now useful to restrict the term 'core Rhizaria' to Cercozoa plus Foraminifera, the two groups with the shared synapomorphy of the polyubiquitin insertion. Thus Rhizaria comprise core Rhizaria plus Radiozoa.

Exclusion of Apusozoa from Rhizaria

When Apusozoa were originally tentatively placed in Rhizaria it was recognised that their inclusion made the group paraphyletic (Cavalier-Smith

	Monomer N	Monomer N+1			
	Euglypha rotunda	TLHLVLRRLRGGSGMQIFVKTLTGK			
	* Thaumatomonas sp.	TLHLVLRRLRGGSGMQIFVKTLTGK			
	* 'Spongomonas minima UT1'	TLHLVLRRLRGGSGMQIFVKTLTGK			
Clade A	Cercomonas sp. E	TLHLVLRRLRGGSGMQIFVKTLTGK			
	Cercomonas sp. 22 (ATCC 50318)	TLHLVLRRLRGGSGMQIFVKTLTGK			
	* Cercomonas 'longicauda'-1	TLHLVLRRLRGGSGMQIFVKTLTGK			
Clade B	* Cercomonas 'longicauda'-2	TLHLVLRRLRGG S AMQIFVKTLTGK	Monadofilosa		
	Cercomonas sp. 18-1 (ATCC 50316)	TLHLVLRRLRGGSGMQIFVKTLTGK			
	Cercomonas sp. 18-2 (ATCC 50316)	TLHLVLRRLRGG S AMQIFVKTLTGK			
	* Heteromita sp.	TLHLVLRRLRGGSGMQIFVKTLTGK			
	* Sainouron mikroteron	TLHLVLRRLRGG AG MQIFVKTLTGK		Filosa	
	* Metromonas simplex	TLHLVLRRLRGGSGMQIFVKTLTGK			
	* Massisteria marina	TLHLVLRRLRGG NG MQIFVKTLTGK	Proteomyxidea		
	* Dimorpha-like	TLHLVLRRLRGGSGMQIFVKTLTGK			
	* Gymnophrys sp.	TLHLVLRRLRGGSGMQIFVKTLTGK			
	* Metopion fluens	TLHLVLRRLRGG S -MQIFVKTLTGK	basal filosan flagellates		
	* Helkesimastix sp.	TLHLVLRRLRGG A -MQIFVKTLTGK			
	* Cercobodo agilis	TLHLVLRRLRGG S -MQIFVKTLTGK			
	Bigelowiella natans	TLHLVLRRLRGG S -MQIFVKTLTGK			
	Lotharella amoebiformis 1	TLHLVLRRLRGG A -MQIFVKTLTGK	Chlorarachnea		
	Lotharella amoebiformis 2	TLHLVLRRLRGG S -MQIFVKTLTGK			
	Lotharella globosa	TLHLVLRRLRGG S -MQIFVKTLTGK			
	Spongospora subterranea	TLHLVLRRLRGG T -MQIFVKTLTGK			
	Plasmodiophora brassicae	TLHLVLRRLRGG T -MQIFVKTLTGK	Plasmodiophorida	Endomyxa	
	* Gromia oviformis	TLHLVLRRLRGG S -MQIFVKTLTGK			
	Reticulomyxa filosa	TLHLVLRRLRGG A -MQIFVKTLTGK			
	Haynesina germanica	TLHLVLRRLRGG A -MQIFVKTLTGK			
	* Bathysiphon spp. 1&2	TLHLVLRRLRGG T -MQIFVKTLTGK	FORAMINIFERA		
	* Bathysiphon sp. 3	TLHLVLRRLRGG -- MQIFVKTLTGK			
	* Collozoum sp.	TLHLVLRRLRGG -- MQIFVKTLTGK			
	* Collozoum inerme 1	TLHLVLRRLRGG -- MQIFVKTLTGK	Polycystinea		
	* Sphaerozoum italicum	TLHLVLRRLRGG -- MQIFVKTLTGK		RADIOZOA	
	* Stauracon pallidus	TLHLVLRRLRGG -- MQIFVKTLTGK			
	* Xiphacantha alata	TLHLVLRRLRGG -- MQIFVKTLTGK	Acantharea		
	* marine microheliozoan	TLHLVLRRLRGG -- MQIFVKTLTGK	HELIOZOA		
	* Ancyromonas sp.	TLHLVLRRLRGG -- MQIFVKTLTGK	APUSOZOA		
	* Amastigomonas sp.	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Trichomonas vaginalis	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Naegleria fowleri	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Tetrahymena pyriformis	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Euplotes eurystomus	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Plasmodium falciparum	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Trypanosoma cruzi	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Phytophthora infestans	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Acanthamoeba castellanii	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Dictyostelium discoideum	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Physarum polycephalum	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Volvox carteri	TLHLVLRRLRGG -- MQIFVKTLTGK	OTHER EUKARYOTES		
	Gracilaria gracilis	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Arabidopsis thaliana	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Oryza sativa	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Pinus sylvestris	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Saccharomyces cerevisiae	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Neurospora crassa	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Candida albicans	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Homo sapiens	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Drosophila melanogaster	TLHLVLRRLRGG -- MQIFVKTLTGK			
	Caenorhabditis elegans	TLHLVLRRLRGG -- MQIFVKTLTGK			

Figure 1. Junction between two ubiquitin monomers in the polyubiquitin genes of rhizarians, apusozoans, a marine microheliozoan (possibly sister to the centrohelid heliozoa: Cavalier-Smith and Chao 2003a), and a range of other eukaryotes. The insertions of one and two amino acids between the monomers are marked in bold. The 23 results generated by this study are marked with an asterisk.

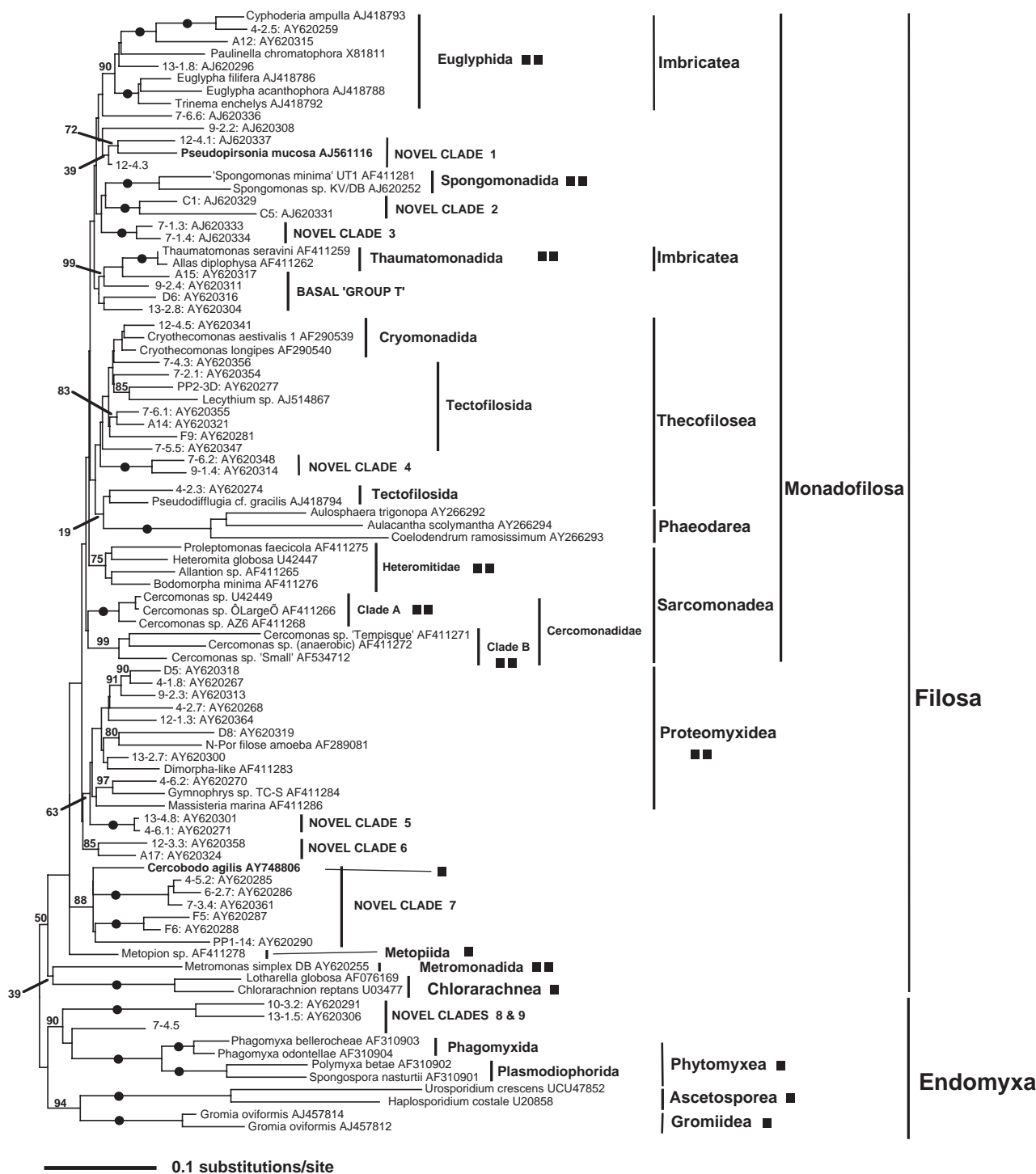


Figure 2. Distance tree of 91 cercozoan 18S rRNA sequences (including some from environmental libraries and novel clades first described in Bass and Cavalier-Smith 2004) using 1163 positions (GTR+ Γ + i model). All bootstrap percentages $\geq 75\%$ are shown; others lower than this are shown for groups of particular interest to this study. Black blobs indicate 100% support. Black squares to the right of the tree represent the number of amino acid residues in the polyubiquitin insertion in the representative member(s) of the taxa indicated.

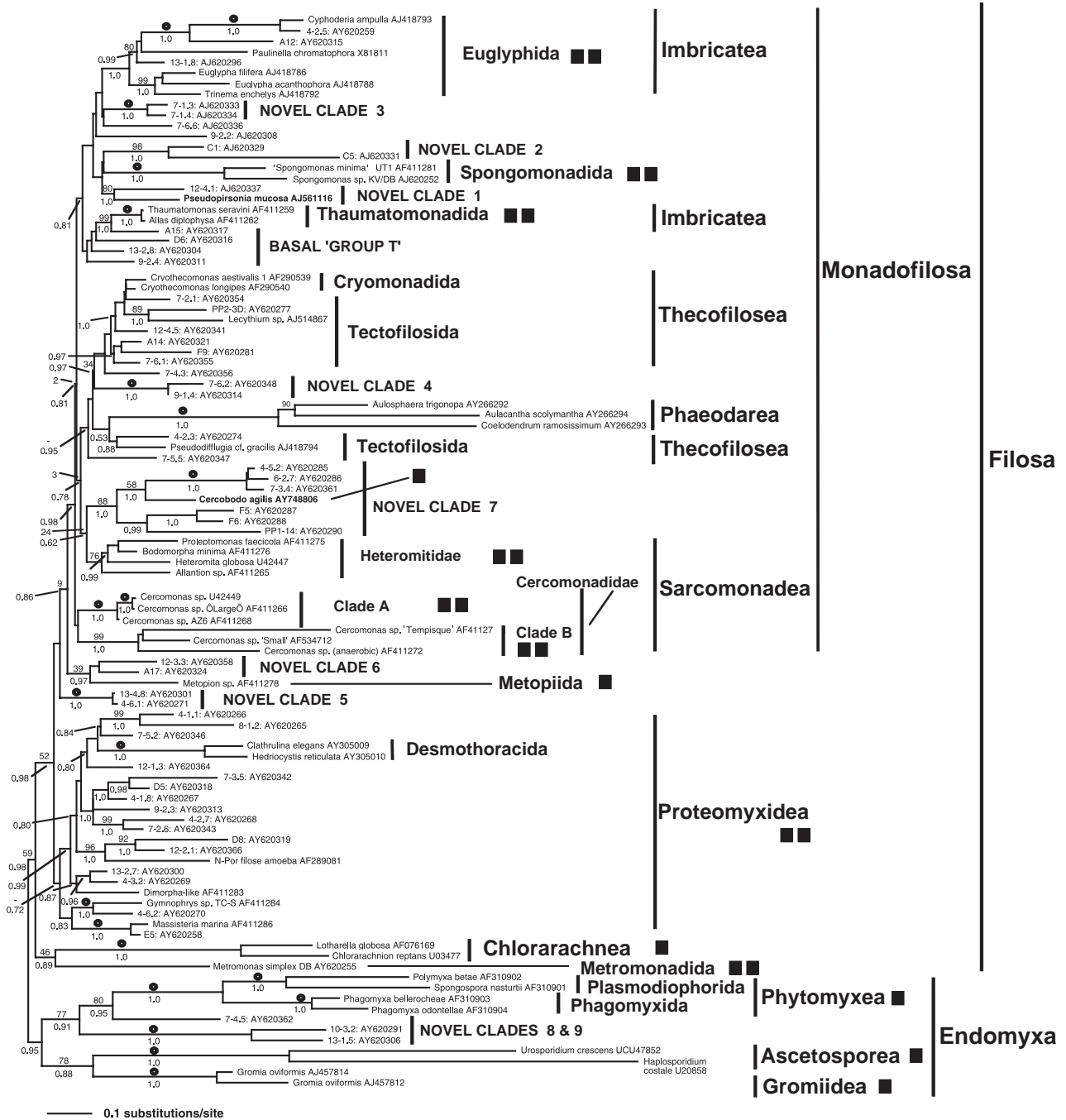


Figure 3. Bayesian tree of 100 cercozoan sequences using 1118 positions. The figures and blobs at the nodes are bootstrap percentages (upper; blobs = 100% support) and Bayesian posterior probabilities (lower) for this dataset. See Methods for details. The criteria for inclusion of bootstrap support values are the same as for Figure 2 and also apply to the PP values.

2002). Subsequently a possible relationship of Apusozoa to Excavata was suggested, enabling Apusozoa to be excluded from Rhizaria (Cavalier-

Smith 2003a, b). Neither *Amastigomonas* nor *Ancyromonas* polyubiquitin genes carry the insertion, from which we infer that Apusozoa do not

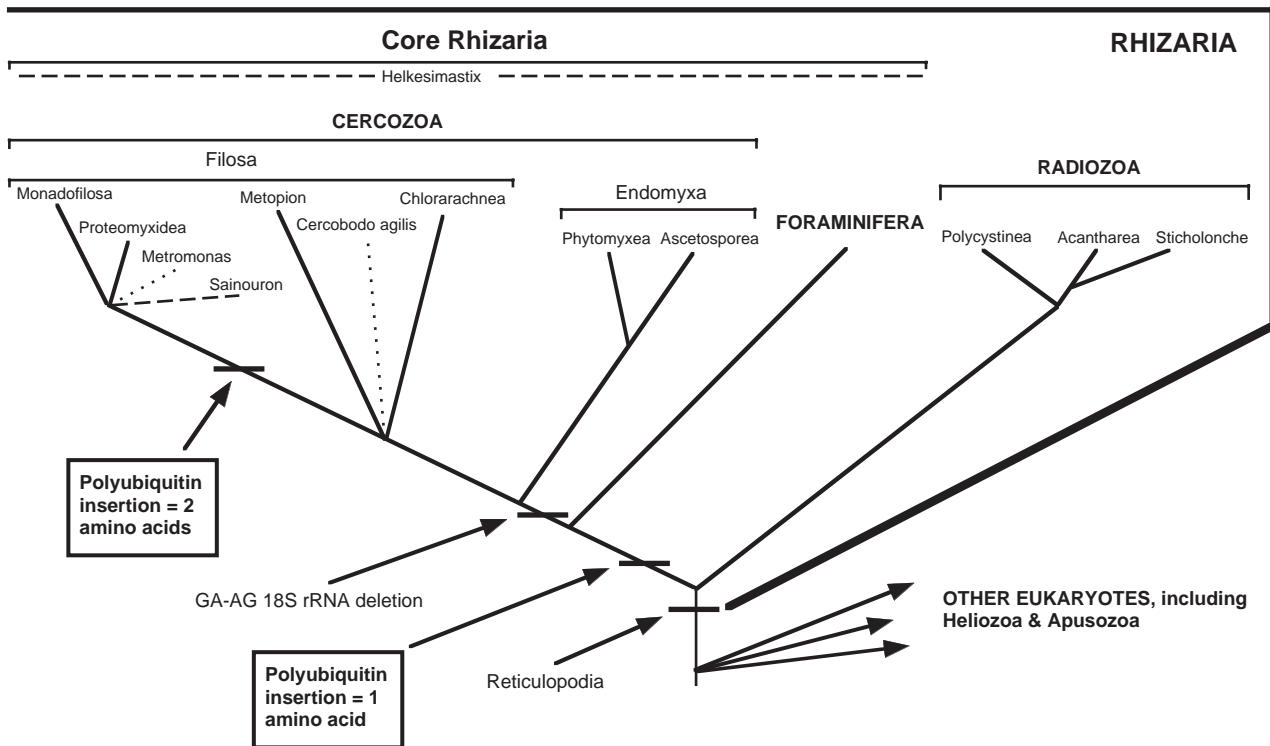


Figure 4. Schematic tree showing relative phylogenetic positions of the taxa under study as jointly informed by 18S rRNA gene phylogenies and the partitioning suggested by the distribution of presence/absence, and the size, of the polyubiquitin insertion. Solid branches summarize information from 18S rDNA analyses, morphological, and polyubiquitin data; dashed lines indicate those taxa (*Sainouron* and *Helkesimastix*) for which 18S rDNA sequences are unavailable. The dotted line to *Metromonas* signifies concordance of morphological and polyubiquitin insertion data, but discordance with 18S rDNA sequence data; that to *Cercobodo* indicates its position according to some analyses and as suggested by the nature of the insertion in its polyubiquitin gene, but acknowledges the instability of its branching position (cf. Figs 2, 3).

branch within core rhizaria. Apusozoa was a candidate for inclusion in Rhizaria on the basis of its mode of locomotion (active anterior and trailing posterior flagellum used for gliding on surfaces characterise Cercozoa) and sharing the character of kinetocyst extrusomes with some Cercozoa. The phylogenetic position (or even holophyly) of Apusozoa cannot be resolved by 18S rDNA analyses (Atkins et al. 2000; Cavalier-Smith 2000; Cavalier-Smith and Chao 2003b; Cavalier-Smith 2004). It is now considered incertae sedis in subkingdom Biciliata; whether it is the most divergent bikont group or is related to excavates remains unclear (Cavalier-Smith 2004; Cavalier-Smith and Chao 2003b).

Heliozoa

Centrohelid heliozoa, like Apusozoa, fail to group consistently with any other phylum on rDNA trees

(Cavalier-Smith and Chao 2003b) and are now also provisionally excluded from Rhizaria (Cavalier-Smith 2004). As all our stocks of genomic DNA from cultures of centrohelid Heliozoa are probably contaminated with other (prey) eukaryotes, we did not consider it worthwhile to sequence polyubiquitin from them, as its provenance would have been uncertain. Heliozoa are therefore represented in our polyubiquitin data only by the unnamed marine microheliozoan (which feeds on bacteria) that usually groups weakly as sister to centrohelids in 18S rDNA trees (Cavalier-Smith and Chao 2003a). However, it is ultrastructurally uncharacterised and probably not a centrohelid. The absence of the insertion from this microheliozoan suggests that it is not part of the cercozoan-foraminiferan clade, but this should not be generalised to centrohelids until there is direct evidence from them or firmer evidence that the two are related, though it is likely that centrohelids will turn out also to lack the insert.

Centrohelids have also been considered putative Rhizaria: their hexagonally arranged axopodial axonemes superficially resemble those of radiolarians, and they also have kinetocyst extrusomes, but both characters may be convergent. They do not group within or as sister to Rhizaria on 18S rDNA and actin trees (Cavalier-Smith and Chao 2003a; Nikolaev et al. 2004). As desmothoracids (e.g. *Clathrulina*, *Hedriocystis*) are not related to centrohelids but group within Cercozoa, as suggested by Cavalier-Smith and Chao (2003a,c) and shown in Nikolaev et al. (2004), they should carry the insert. This is currently being investigated.

The Relationship between Foraminifera and Cercozoa

There is now evidence of single amino acid insertions in the polyubiquitin gene of three phylogenetically divergent foraminiferan genera—*Haynesina*, *Reticulomyxa* (both have an alanine insert) (Archibald et al. 2003), and *Bathysiphon* (threonine) (this study). This information, in combination with several other lines of evidence (actin, RPB1, and 18S rDNA phylogenies; see references above), confirms that Foraminifera and Cercozoa are very closely related, but whether they are sisters or whether Foraminifera occupy a derived position within Cercozoa remains uncertain.

The facts that Foraminifera do not share the unique single nucleotide deletion in the 18S rRNA gene shown by all Cercozoa (Cavalier-Smith and Chao 2003b), and that more highly sampled 18S rDNA distance and ML trees do not show Foraminifera nested within Cercozoa, suggest that Foraminifera and Cercozoa may be sisters. However, it is possible, given the rapid evolution of the foraminiferan rRNA genes (Pawlowski and Berney 2003) that the thymine-containing nucleotide in Foraminifera at the site of the cercozoan deletion might be secondarily inserted, and that Foraminifera are actually derived from Cercozoa. This would be congruent with some actin trees, which show Foraminifera branching within Cercozoa, as sister to *Cercomonas* (Keeling 2001) or as sister to Plasmodiophorida (Archibald and Keeling 2004). However, establishing the phylogenetic position of Foraminifera using actin sequences is made difficult by the presence in Foraminifera of two paralogous actin genes, which often do not group together. For example, in some recent actin phylogenies, one paralogue branches with *Gromia oviformis*, while the other appears as sister to Polycystinea (Longet et al. 2004; Nikolaev et al.

2004); another recent very well sampled actin tree did not even group Foraminifera and Cercozoa (Stechmann and Cavalier-Smith unpublished). This problem was resolved recently by increasing taxon sampling of foraminiferan actin and by removing fast-evolving actin sequences from the analyses (Flakowski et al. 2005). However, actin is too conserved to resolve the exact relationship between Cercozoa and Foraminifera and other proteins should be investigated.

Phylogeny of Cercozoa

As Figure 1 indicated, the following cercozoans have a single amino acid insertion: *Bigelowiella* and *Lotharella* (both chlorarachneans), *Cercobodo agilis*, *Metopion*, and *Gromia*. *Helkesimastix* also has a single amino acid insertion and is therefore a core rhizarian, but we cannot confirm that it is a cercozoan as data from other genes are inconclusive, even though its gliding zooflagellate morphology strongly suggests that it is; the presence of only one amino insertion, in contrast to the two in *Cercomonas*, provides additional evidence that it is not a cercomonad. The taxa with a single amino acid insertion either belong to the subphylum Endomyxa (*Gromia*), or are basal lineages in subphylum Filosa in many 18S rDNA trees (e.g. Fig. 2). The possible exceptions to this are *Metopion* and *Cercobodo*, which in some analyses group above proteomyxids and as sister to heteromitids respectively, but with negligible support (Fig. 3). At other times they branch below proteomyxids, but equally unreliably.

In many cases the evidence from 18S rDNA trees and the polyubiquitin insertion data are concordant. The few cases where there are weakly supported discrepancies between trees and the distribution of the polyubiquitin insertion are priorities for further analysis. If there are no proven cases of loss of either amino acid residues in the insertion, our data would provide strong evidence that taxa with a single amino acid insertion are indeed the most basal filosaurs, and those with insertions of two amino acids are derived. However, as Figure 3 suggests that this condition may have been violated at least three times, albeit with weak support (*Cercobodo agilis*, *Metopion*, and *Metromonas*), evidence from additional sources is required to resolve the branching order within Cercozoa, and thus verify that some reversions from a double to a single amino acid insertion have occurred.

On the basis of limited morphological evidence (light microscope and ultrastructural) *Metopion*, *Helkesimastix*, and *Cercobodo agilis* might be thought to belong to Sarcomonadea. However, weak indications from 18S rDNA trees are now strengthened by our new polyubiquitin data, which suggest that they represent at least two different, divergent filosan lineages (Figs 2,3). As *Cercobodo agilis* was originally described as *Cercomonas agilis* Morott, but had a single amino acid insert unlike all other *Cercomonas* strains, we sequenced the 18S rRNA gene to see whether or not it has been misclassified. As Figure 3 shows, *Cercobodo agilis* does not group with *Cercomonas* Clade A or B, but instead with Novel Clade 7 (Bass and Cavalier-Smith 2004), in some analyses one of the most deeply diverging filosan clades. The suggestion, on the basis of morphological observations, that *Sainouron* is a sarcomonad, and not a deep-branching filosan (Karpov 1990) is supported by the fact it has a two-amino acid insertion. Both *Sainouron* and *Helkesimastix* are very difficult to work with in terms of rDNA sequencing. As we have not been able to obtain complete 18S rDNA sequences from either our new data provide the first molecular evidence that these genera are cercozoan and core rhizarian respectively.

Almost all other cercozoans screened are more derived filosans (including Monadofilosa), according to 18S rDNA phylogenies, and have an insertion comprising two amino acids. The only exception to this is *Metromonas*, which has the same insertion of two amino acids as several monadofilosans but is the only taxon with such an insertion to group basally—as sister to, or near to *Chlorarachnea* in ML, distance, and Bayesian analyses. However, this relationship has very little statistical support (see also Bass and Cavalier-Smith 2004) and is not corroborated by morphological evidence. We suggest that *Metromonas* may really be related to monadofilosans, but more data are required to confirm this, and we do not rule out the possibility that *Metromonas* is a basal filosan and that either the insertion of two amino acids is a character convergent with that in the true monadofilosans or there have been secondary reversions to a single amino acid. Indeed, if the higher positions shown for *Metopion* and *Cercobodo* on the Bayesian tree (Fig. 3) turn out to be correct, rather than their deeper branching seen on distance trees such as Figure 2, a reasonably parsimonious alternative interpretation would be that the second insertion took place somewhat earlier than indicated in Figure 4, in the ancestral

filosan, and was later followed by three independent reversions to a single amino acid in *Chlorarachnea*, *Metopion* and *Cercobodo*.

Identification of Novel Cercozoan Clades

Of the nine novel clades revealed by screening environmental 18S rRNA gene libraries in Bass and Cavalier-Smith (2004) (our knowledge of them was previously limited to these sequence data), two have now been partially characterised by the knowledge that they contain *Pseudopirsonia mucosa* (Kühn et al. 2004) (a modified Novel Clade 1; 80% bootstrap support) and *Cercobodo agilis* (Novel Clade 7; 88% bootstrap support) (Fig. 3). Figures 2 and 3 are the first published trees showing convincingly that *Pseudopirsonia mucosa* is a monadofilosan cercozoan, not a heterokont as stated in GenBank (Bass and Cavalier-Smith 2004); although Kühn et al. (2004) also show *Pseudopirsonia* to be a cercozoan, parts of their tree in Figure 3 are discordant with previous studies of both heterokonts and Cercozoa; for example alveolates appear polyphyletic; within heterokonts Placidiales, Pseudofungi and hypogyristes are not monophyletic; and within Cercozoa Filosa are not holophyletic and the genus *Cercomonas* is polyphyletic, being split into three clades, one of which misleadingly appears to be sister to *Pseudopirsonia*. Furthermore, the Bayesian tree (Fig. 3) strongly suggests for the first time that Novel Clade 6 is closely related to *Metopion* and that Novel Clade 3 belongs in Thecofilosea.

DNA from Uncultivated Samples

It might be argued that as neither the polycystines nor the acantharea were cultivated, the DNA samples from them could have been contaminated by non-rhizarian protists and that it may be their DNA that was amplified by the polyubiquitin primers, and that Radiozoa themselves might actually have the insert. To ensure that this is not the case, our sequences are based on several independent samples processed in separate laboratories. Each sample was carefully washed in an attempt to exclude such contamination and all of them yielded only authentic radiozoan rRNA genes when tested using universal 18S rDNA primers. Although the possibility of contamination of all of them seems remote, because we cannot altogether exclude it (and also because there is a possibility that the ancestral radiozoa might have

lost the insertion), our conclusion that Foraminifera are more probably more closely related to Cercozoa than to the Radiozoa requires corroboration from other molecules.

In the foraminiferan *Bathysiphon*, even though two clones carried a single amino acid insertion as in all other studied foraminifera, one clone showed no insertion. This latter sequence is very unlikely to originate from a foraminiferan as it lacks three signature sequences from elsewhere in the polyubiquitin gene that are shared by all other sequenced foraminiferans, shown here in bold italics with the non-foraminiferan state as subscripts: GGA-MQIFVKTLTGKTITLDVEPN_SDTIQ_E/D/SNVKAKIQDKEGIPPE_D.

Conclusions

The possible partitions in the rhizarian tree in the light of the taxonomic distribution of the two types of polyubiquitin insertion in combination with information from 18S rDNA trees were summarised in Figure 4. Because they are the only two groups to share the insert, Cercozoa and Foraminifera appear to be more closely related to each other than any of the other groups screened, but the distribution of the polyubiquitin insertion does not allow us to decide whether Cercozoa are sisters of or ancestral to Foraminifera. If there turn out to be no cases where the insertion has been deleted, then its absence from Apusozoa, Radiozoa and the microheliozoan is strong evidence that these groups are excluded from core Rhizaria, although Radiozoa probably belong to Rhizaria, branching just outside the Cercozoa–Foraminifera clade.

For cercozoan phylogeny, Figure 3 is largely congruent with that in Bass and Cavalier-Smith (2004). The major novelties are that *Cercobodo agilis* is strongly supported as a member of Novel Clade 7 (Bass and Cavalier-Smith 2004) and desmothoracids branch well within the proteomyxids. Assuming that insertions have not arisen convergently, we propose that *Sainouron* is a monadofilosan and *Helkesimastix* is a more divergent cercozoan (though it might group outside Cercozoa, it is nonetheless definitely a core rhizarian). Because of the conflict between the distance and Bayesian trees in the positions of *Metopion* and *Cercobodo*, we cannot determine whether they are basal filosaurs that never had the second amino acid insertion (Fig. 4) or are part of the later sarcomonad radiation and have secondarily lost it. Thus the second amino acid insertion

occurred either in the ancestral filosan or somewhat later (Fig. 4), after the divergence of *Helkesimastix*, Novel Clade 7, Chlorarachnea, and the rest of Filosa. It also remains unclear whether *Metromonas* is sister to Chlorarachnea, as weakly indicated by the rDNA trees, or closer to Monadofilosa as morphology and the doubleness of its polyubiquitin insertion suggest.

Post Scriptum

Since this paper was written, Vickerman and Moreira (pers. comm.) have shown that a newly discovered cercozoan, '*Aurigamonas solis*', is related to *Cercobodo agilis*, belonging to Novel Clade 7 according to 18S rDNA phylogenies, and like *C. agilis*, has a single amino acid polyubiquitin insertion. Therefore it is possible that there was a reversion to a single amino acid insertion in the common ancestor of all NC7 lineages.

Methods

Detection of polyubiquitin insertion: Genomic DNA from the following taxa was amplified with primers UBIQ1: 5'-GGCCATGCARATHTTYGT-NAARAC-3' and IUB2: 5'-GATGCCYTCYTRT-CYTGDAYTT-3': '*Spongomonas minima* UT1' (ATCC 50405), '*Cercomonas longicauda*' (ATCC50344), *Heteromita* sp. (culture now dead), *Sainouron mikroteron* (ATCC 50340), *Metromonas simplex* (culture now dead), *Massisteria marina* (ATCC 50266), *Dimorpha*-like sp. (ATCC 50522), *Gymnophrys* sp. (ATCC 50638), *Metopion fluens* (AP Mylnikov, Borok, Russia), *Helkesimastix faecicola* (ATCC 50328), *Cercobodo agilis* (CCAP 1910/1), *Gromia oviformis**, *Bathysiphon* sp.*, *Collozoum* sp.**, *Collozoum inerme***, *Sphaerozoum italicum***, *Stauracon pallidus***, *Xiphacantha alata***, *Ancyromonas sigmoides* (ATCC 50267), *Amastigomonas* sp. (ATCC 50062), and marine microheliozoan (TC-S lab). The species marked with single asterisks are uncultivable. Those with two asterisks were collected directly from plankton samples taken from Villefranche sur Mer (Mediterranean coast of France) at 25 m depth using a plankton net. The different species were identified using a stereomicroscope, isolated and washed many times with sterile sea water before DNA extraction. *Bathysiphon* sp. was collected from the coast of Scotland. *G. oviformis* originated from the Mediterranean coast. As for the planktonic species, the foraminifera were thoroughly cleaned prior to extraction, which was performed

with 10–50 specimens. *Gromia*'s extraction was done using single specimens undergoing gametogenesis, as described in Arnold (1972). To ascertain the authenticity of ubiquitin sequences, amplifications were obtained from at least two DNA extracts for each uncultured species and 1–3 clones were sequenced for each PCR product.

PCR reactions were performed as follows: 3 min denaturation at 94 °C, followed by 35 or 45 cycles of 45 s at 92 °C, 1 min at 50 °C, and 1.5 min at 72 °C, followed by a final extension of 5 min at 72 °C. The products were run on a gel. This primer pair produces a ladder, the shortest fragment is less than a complete ubiquitin monomer; the bands increase in size by one whole monomer length. Bands corresponding to 1.5 and 2.5 monomer lengths were excised from the gel and purified. rDNA clone libraries were constructed using the TOPO TA Cloning kit (Invitrogen). White or light blue colonies were picked and grown in 2 ml LB medium:1 µl ampicillin overnight. The plasmid DNA was extracted and cleaned using the QIAquick PCR Purification kit (Qiagen). The cleaned products were run on a gel with known negative transformants; positive transformants were sequenced using dye terminators and separated on an automated ABI-377 sequencer. The sequencing primers used were M13 forward and M13 reverse. The sequences were edited and translated in the trace editor TED, then aligned by eye using the Genetic Data Environment software.

18S rDNA distance and Bayesian trees: a representative sample of 89 partial (about 1.26 kb only to allow inclusion of the incomplete novel clade sequences) cercozoan sequences from our database was aligned by eye using the Genetic Data Environment software. New 18S rDNA sequences for *Cercobodo agilis* (this study: culture from K. Vickerman, Glasgow) and *Pseudopirsonia mucosa* (GenBank) were aligned to this dataset. Gaps and ambiguously aligned positions were also removed, leaving an alignment of 1163 positions. ModelTest v.3.06 (Posada and Crandall 1998) selected the GTR model with gamma correction for intersite rate variation (Γ) and allowance for invariant sites (i) for the dataset ($\alpha = 0.631$; $i = 0.258$). These parameters were used for distance trees constructed using PAUP 4.0b10 (Swofford, 2003). A heuristic search (500 replicates; random addition) using minimum evolution was made (with TBR branch swapping). The tree was rooted between Filosa and Endomyxa as shown by maximum likelihood trees with close outgroups in Cavalier-Smith and Chao (2003c).

Bootstrap support values were estimated using 500 bootstrap replicates within one addition sequence for a distance analysis using a ML model of substitution with the same values of α and i .

The Bayesian tree (Fig. 3) was calculated using MrBayes v3.0b4 (Huelsenbeck and Ronquist 2001) based on 100 taxa, including all those in Figure 2 with the addition of desmothoracids (*Clathrulina* & *Hedriocystis*) and some more environmental cercozoan sequences; 1118 positions were used for the analysis. Two separate MCMCMC runs with randomly generated starting trees were carried out for 4 million generations each with one cold and three heated chains. The evolutionary model applied included GTR substitution matrix, gamma correction with shape parameter estimated from the data, the covarion model, and autocorrelation. Trees were sampled every 100 generations. Six thousand and five hundred trees were discarded as burn-in (trees sampled before the likelihood plots reached a plateau). A consensus of the remaining trees was generated to reveal the posterior probabilities (PPs) of the branching pattern. Both independent runs resulted in basically identical tree topologies and posterior probabilities, indicating that the runs lasted long enough to converge. The tree shown is not the consensus tree produced by the Bayesian analysis but the tree with the highest likelihood selected from the posterior distribution, annotated with the consensus PPs. The bootstrap support values on Figure 3 were estimated in PAUP 4.0b10 using 500 bootstrap replicates for each of 50 addition sequence replicates for a distance analysis using a GTR+ Γ + i model ($\alpha = 0.611$; $i = 0.237$).

Sequences have been deposited in GenBank under accession numbers AY748806–AY748813, AJ937550–AJ937554, DQ098270–DQ098278, and DQ100436–DQ100438. Sequence alignments are available on request from the corresponding author.

Acknowledgements

The work was supported by NERC Research Studentships to DB and SvdH. TCS thanks NERC UK for research grants and NERC and the Canadian Institute for Advanced Research Evolutionary Biology Program for fellowship support. DM, PL-G, SP, and JP would like to thank Jean and Colette Febvre (Station Biologique de Villefranche sur Mer) for hospitality and help during the

isolation and identification of Polycystinea and Acantharea. Their part of the work was supported by an ATIP grant of the CNRS to PL-G. Research of JP and SP was supported by Swiss NSF grant 3100A0-100415. We thank John M Archibald for the polyubiquitin primers.

Reference

- Arnold ZM** (1972) Observations on the biology of the protozoan *Gromia oviformis*. Dujardin. Univ Calif Publ Zool **100**: 76–107
- Atkins MS, McArthur AG, Teske AP** (2000) Ancyromonadida: a new phylogenetic lineage among the protozoa closely related to the common ancestor of metazoans, fungi, and choanoflagellates (Opisthokonta). J Mol Evol **51**: 278–285
- Archibald JM, Keeling PJ** (2004) Actin and ubiquitin protein sequences support a Cercozoa/Foraminifera ancestry for the plasmodiophorid plant pathogens. J Eukaryot Microbiol **51**: 113–118
- Archibald JM, Longet D, Pawlowski J, Keeling PJ** (2003) A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. Mol Biol Evol **20**: 62–66
- Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF** (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science **290**: 972–977
- Baptiste E, Philippe H** (2002) The potential value of indels as phylogenetic markers: position of trichomonads as a case study. Mol Biol Evol **19**: 972–977
- Baptiste E, Brinkmann H, Lee JA, et al** (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. Proc Natl Acad Sci USA **99**: 1414–1419
- Bass D, Cavalier-Smith T** (2004) Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). Int J Syst Evol Microbiol **54**: 2393–2404
- Berney C, Pawlowski J** (2003) Revised small subunit rRNA analysis provides further evidence that Foraminifera are related to Cercozoa. J Mol Evol **57**: S120–S127
- Berney C, Fahrni J, Pawlowski J** (2004) How many novel eukaryotic “kingdoms” Pitfalls and limitations of environmental DNA surveys. BMC Biol **2**: 13
- Burki F, Berney C, Pawlowski J** (2002) Phylogenetic position of *Gromia oviformis* Dujardin inferred from nuclear-encoded small subunit ribosomal DNA. Protist **153**: 251–260
- Cavalier-Smith T** (1993) Kingdom Protozoa and its 18 phyla. Microbiol Rev **57**: 953–994
- Cavalier-Smith T** (1998) A revised six-kingdom system of life. Biol Rev Camb Philos Soc **73**: 203–266
- Cavalier-Smith T** (1999) Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryotic family tree. J Eukaryot Microbiol **46**: 347–366
- Cavalier-Smith T** (2000) Flagellate Megaevolution: the Basis for Eukaryote Diversification. In Green JR, Leadbeater BSC (eds) The Flagellates. Taylor and Francis, London, pp 361–390
- Cavalier-Smith T** (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. Int J Syst Evol Microbiol **52**: 297–354
- Cavalier-Smith T** (2003a) Protist phylogeny and the high-level classification of Protozoa. Europ J Protistol **39**: 338–348
- Cavalier-Smith T** (2003b) The excavate protozoan phyla Metamonada Grassé emend. (*Anaeromonadea*, *Parabasalia*, *Carpodimonas*, *Eopharyngia*) and Louksozoa emend. (*Jakobea*, *Malawimonas*): their evolutionary affinities and new higher taxa. Int J Syst Evol Microbiol **53**: 1741–1758
- Cavalier-Smith T** (2004) Only six kingdoms of life. Proc R Soc Ser B **271**: 1251–1262
- Cavalier-Smith T, Chao E-Y** (2003a) 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, Chao E-Y** (2003b) Phylogeny of Choanozoa, Apusozoa, and other protozoa and early eukaryote megaevolution. J Mol Evol **56**: 540–563
- Cavalier-Smith T, Chao E-Y** (2003c) Phylogeny and classification of phylum Cercozoa (Protozoa). Protist **154**: 341–358
- Embley TM, Hirt RP** (1998) Early branching eukaryotes? Curr Opin Genet Dev **8**: 624–629
- Febvre-Chevalier C** (1990) Phylum Actinopoda: Class Heliozoa. In Margulis L, Corliss JO, Melkonian M, Chapman D (eds) Handbook of Protozoa. Jones and Bartlett, Boston, MA, pp 399–416
- Finley D, Chau V** (1991) Ubiquitination. Annu Rev Cell Biol **7**: 25–69

- Flakowski J, Bolivar I, Fahrni J, Pawlowski J** (2005) Actin phylogeny of Foraminifera. *J Foraminifera Res* **35**: 93–102
- Hershko A, Ciechanover A** (1998) The ubiquitin system. *Annu Rev Biochem* **67**: 425–479
- Huelsenbeck JP, Ronquist F** (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755
- Karpov SA** (1990) System of Protists. Omsk, Mezhdunarodnaya
- Keeling PJ** (2001) Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? *Mol Biol Evol* **18**: 1551–1557
- Kühn S, Medlin LK, Eller G** (2004) Phylogenetic position of the parasitoid nanoflagellate *Pirsonia* inferred from nuclear-encoded small subunit ribosomal DNA and a description of *Pseudopirsonia* n. gen. and *Pseudopirsonia mucosa* (Drebes) comb. nov. *Protist* **155**: 143–156
- Longet D, Archibald JM, Keeling PJ, Pawlowski J** (2003) Foraminifera and Cercozoa share a common origin according to RNA polymerase II phylogenies. *Int J Syst Evol Microbiol* **53**: 1735–1739
- Longet D, Burki F, Flakowski J, Berney C, Polet S, Fahrni J, Pawlowski J** (2004) Multigene evidence for close evolutionary relationships between *Gromia* and Foraminifera. *Acta Protozoologica* **43**: 303–311
- López-García P, Rodríguez-Valera, Moreira D** (2002) Toward the monophyly of Haeckel's Radiolaria: 18S rRNA environmental data support the sisterhood of Polycystinea and Acantharea. *Mol Biol Evol* **19**: 118–121
- Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV** (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**: 162–165
- Nikolaev SI, Berney C, Fahrni J, Mylnikov AP, Aleshin VV, Petrov N, Pawlowski J** (2003) *Gymnophrys cometa* and *Lecythium* sp. are core Cercozoa: evolutionary implications. *Acta Protozool* **42**: 183–190
- Nikolaev SI, Berney C, Fahrni J, Bolivar I, Polet S, Mylnikov AP, Aleshin VV, Petrov NB, Pawlowski J** (2004) The twilight of Heliozoa and rise of Rhizaria: a molecular revolution in the phylogeny of amoeboid protists. *Proc Natl Acad Sci USA* **101**: 8066–8071
- Pawlowski J, Berney C** (2003) Episodic Evolution of Nuclear Small Subunit Ribosomal RNA Gene in the Stem Lineage of Foraminifera. In Donoghue PC, Smith MP (eds) *Telling the Evolutionary Time: Molecular Clocks and the Fossil Record*. Systematics Association Special Volume.No.66.. Taylor and Francis, London, pp 107–118
- Pawlowski J, Holzmann M** (2002) Molecular phylogeny of Foraminifera—a review. *Europ J Protistol* **38**: 1–10
- Philippe H, Adoutte A** (1998) The Molecular Phylogeny of Eukaryota: Solid Facts and Uncertainties. In Coombs G, Vickerman K, Sleigh M, Warren A (eds) *Evolutionary Relationships among Protozoa*. Chapman and Hall, London, pp 25–56
- Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Müller M, Le Guyader H** (2000) Early branching or fast evolving eukaryotes? An answer based on slowly evolving positions. *Proc R Soc Ser, B* **267**: 1213–1222
- Polet S, Berney C, Fahrni J, Pawlowski J** (2004) Small-subunit ribosomal RNA gene sequences of Phaeodarea challenge the monophyly of Haeckel's Radiolaria. *Protist* **155**: 53–63
- Posada D, Crandall KA** (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818
- Sharp PM, Li W-H** (1987) Molecular evolution of ubiquitin genes. *Trends Ecol Evol* **2**: 328–332
- Stechmann A, Cavalier-Smith T** (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science* **297**: 89–91
- Stechmann A, Cavalier Smith T** (2003) The root of the eukaryote tree pinpointed. *Curr Biol* **13**: R665–R666
- Sun L, Chen ZJ** (2004) The novel functions of ubiquitination in signalling. *Curr Opin Cell Biol* **16**: 119–126
- Swofford DL** (2003) PAUP*: Phylogenetic Analysis using Parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, MS