Molecular Phylogenetics and Evolution 81 (2014) 71–85

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa

Thomas Cavalier-Smith *, Ema E. Chao, Elizabeth A. Snell, Cédric Berney [1], Anna Maria Fiore-Donno [2], Rhodri Lewis

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

ARTICLE INFO

Article history:
Received 24 January 2014
Revised 2 August 2014
Accepted 11 August 2014
Available online 23 August 2014

Keywords:
Cell evolution
Eukaryote phylogeny
Podiates
Protozoa
Transcriptome sequencing
Sulcozoa

ABSTRACT

Animals and fungi independently evolved from the protozoan phylum Choanozoa, these three groups constituting a major branch of the eukaryotic evolutionary tree known as opisthokonts. Opisthokonts and the protozoan phylum Amoebozoa (amoebae plus slime moulds) were previously argued to have evolved independently from the little-studied, largely flagellate, protozoan phylum, Sulcozoa. Sulcozoa are a likely evolutionary link between opisthokonts and the more primitive excavate flagellates that have ventral feeding grooves and the most primitive known mitochondria. To extend earlier sparse evidence for the ancestral (paraphyletic) nature of Sulcozoa, we sequenced transcriptomes from six gliding flagellates (two apusomonads; three planomonads; *Mantamonas*). Phylogenetic analyses of 173–192 genes and 73–122 eukaryote-wide taxa show Sulcozoa as deeply paraphyletic, confirming that opisthokonts and Amoebozoa independently evolved from sulcozoans by losing their ancestral ventral groove and dorsal pellicle: Apusozoa (apusomonads plus anaerobic breviate amoebae) are robustly sisters to opisthokonts and probably paraphyletic, breviates diverging before apusomonads; Varisulca (planomonads, *Mantamonas*, and non-gliding flagellate *Collodictyon*) are sisters to opisthokonts plus Apusozoa and Amoebozoa, and possibly holophyletic; Glissodiscea (planomonads, *Mantamonas*) may be holophyletic, but *Mantamonas* sometimes groups with *Collodictyon* instead. Taxon and gene sampling slightly affects tree topology; for the closest branches in Sulcozoa and opisthokonts, proportionally reducing missing data eliminates conflicts between homogeneous-model maximum-likelihood trees and evolutionarily more realistic site-heterogeneous trees. Sulcozoa, opisthokonts, and Amoebozoa constitute an often-pseudopodial 'podiate' clade, one of only three eukaryotic 'supergroups'. Our trees indicate that evolution of sulcozoan dorsal pellicle, ventral pseudopodia, and ciliary gliding (probably simultaneously) generated podiate eukaryotes from *Malawimonas*-like excavate flagellates.

1. Introduction

Phylogenetically, all eukaryotes have been assigned to just three supergroups: podiates, corticates (kingdoms Plantae and Chromista), and Eozoa (excavates and Euglenozoa) (Cavalier-Smith, 2013a). The entirely heterotrophic podiates, the focus of this paper, include Animalia, Fungi, and four protozoan phyla (Sulcozoa, Amoebozoa, Choanozoa Microsporidia); they are so called because of the general presence of pseudopodia except in the derived Fungi that lost them (Cavalier-Smith, 2013a). Originally 'excavates' excluded Euglenozoa (Simpson and Patterson, 1999), but later these distinctive flagellates were included despite not sharing excavate morphology (Cavalier-Smith, 2002, 2003; Simpson, 2003) under the influence of a probably erroneous assumption about the location of the eukaryotic root; here we follow Cavalier-Smith (2010a, 2013a) in excluding Euglenozoa from excavates. Multigene trees usually show corticates as a clade, but are contradictory concerning the boundary between the excavate Eozoa and the putatively basal podiate phylum Sulcozoa. The problem lies in the uncertain phylogenetic position of the excavate flagellate *Malawimonas* (Brown et al., 2013; Derelle and Lang, 2012; Hampl et al., 2009; Zhao et al., 2012, 2013). Some multigene trees show podiates as a clade with *Malawimonas* one node deeper (e.g. Brown et al., 2013); others place *Malawimonas* within podiates, typically as sister to the sulcozoan flagellate *Collodictyon* (e.g.

* Corresponding author.
E-mail address: tom.cavalier-smith@zoo.ox.ac.uk (T. Cavalier-Smith).
[1] Present address: Station Biologique de Roscoff, UMR7144/Groupe Evolution et10 PaléOcéans, CNRS/Université Paris VI, Roscoff, France.
[2] Present address: Zoology Institute, Terrestrial Ecology Group, Faculty of Mathematics and Natural Sciences, University of Cologne, Biozentrum Köln, Zülpicher Str. 47b, D-50674 Köln, Germany.

http://dx.doi.org/10.1016/j.ympev.2014.08.012
1055-7903/

Zhao et al., 2012, 2013). Either position is consistent with the postulate that the evolutionary transitions between the three supergroups (Fig. 1) all involved biciliate ventrally grooved cells morphologically similar to *Malawimonas* (Cavalier-Smith, 2013a). Podiates are proposed to have arisen from a *Malawimonas*-like ancestor by evolving ventral pseudopodia and novel dorsal semi-rigid pellicle and abandoning swimming in the plankton in favour of a benthic habitat, gliding on the posterior cilium over surfaces in search of prey (Cavalier-Smith, 2013a).

To clarify the deep branching of podiates one must determine the relationships of the various groups of Sulcozoa, a recently established protozoan phylum of mainly gliding flagellates. Sulcozoa are morphologically unified by a unique cell structure, combining ventral feeding groove, pseudopodia to catch prey (not for locomotion), and rigid dorsal pellicle (Cavalier-Smith, 2013a). Sulcozoa comprise subphyla Apusozoa (apusomonads and breviates) and Varisulca, proposed as the most basal podiate group (Cavalier-Smith, 2013a). Sulcozoa are phylogenetically important as likely ancestors of two major groups, whose cell structure is proposed to have been radically simplified from the complex cytoskeleton of the ventral feeding groove of Sulcozoa and excavates: opisthokonts, whose ancestor lost the anterior cilium and became uniciliate like human sperm, and Amoebozoa, which developed pseudopodia for efficient locomotion (Cavalier-Smith, 2013a). A 16-gene tree weakly grouped three sulcozoan orders as a clade, but was based on only three or four genes for apusomonads and planomonads (Katz et al., 2011), the best studied gliding sulcozoan flagellates; a sulcozoan clade was not seen in a 30-gene tree with four Sulcozoa, because *Malawimonas* was weakly within Sulcozoa (Grant et al., 2012). A study of 159 genes weakly excluded *Malawimonas* from Sulcozoa and podiates, but showed two or three distinct sulcozoan clades at the base of podiates, though sampling only five Sulcozoa from three orders, and only *Collodictyon* representing subphylum Varisulca (Brown et al., 2013). Brown et al. (2013) concluded that Apusozoa were paraphyletic; their trees also confirmed earlier evidence that *Collodictyon* branched deeper still, as the most divergent podiate lineage to date, making Sulcozoa the ancestral (paraphyletic) podiate group.

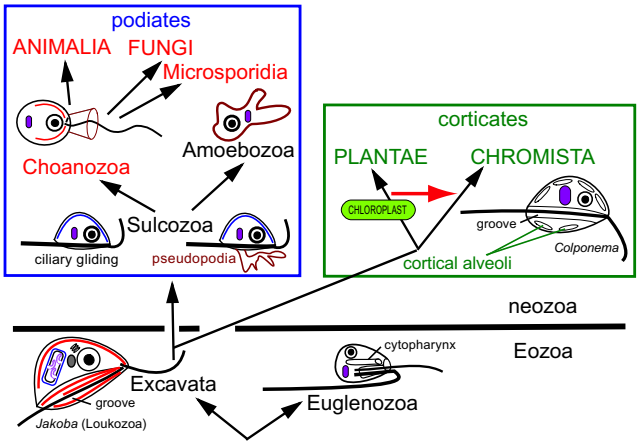

Fig. 1. Likely relationships amongst the eukaryote supergroups (podiates, corticates, Eozoa) highlighting major steps in eukaryote cell diversification. Group names in red are opisthokonts; those in green were ancestrally photosynthetic. The non-photosynthetic phagotrophic last common ancestor of all eukaryotes had two cilia, stemming from a pair of nucleus-attached centrioles anchored within the cell by three microtubular roots R1–R3 (red lines). The eukaryotic tree's root is shown within Eozoa, between excavates (with ventral feeding groove supported by a split right microtubular root R2 and left root R1, and divergent centrioles and cilia) and the grooveless Euglenozoa, whose unsplit, reflexed R2 supports their cytopharynx with complex mouth parts (centrioles and ciliary bases parallel), as explained elsewhere (Cavalier-Smith, 2010a,c, 2013a, 2014), though this precise position is not universally accepted. The groove-supporting microtubular skeleton of loukozoan excavates was arguably inherited by neozoa, but simplified when Choanozoa (ancestral opisthokonts) lost the anterior cilium when evolving a filopodial collar (Cavalier-Smith, 2013a). Our multigene trees showing that the excavate flagellate *Malawimonas* branches beside (or less likely within) Sulcozoa (Figs. 2 and 3, S1 and S3–S5) prove that the ancestors of podiates prior to the origin of their dorsal pellicle and posterior ciliary gliding had an excavate-like cytoskeleton, and vane-bearing posterior cilium, like the chromist *Colponema* and many loukozoan excavates. The sulcozoan dorsal rigidifying pellicular layer(s) (lost by opisthokonts and Amoebozoa) overlying the dorsal root R3 is blue, pseudopodia brown, mitochondria purple. Chloroplasts originated in a biciliate corticate with cortical alveoli by symbiogenetic enslavement of a cyanobacterium; after green plants and red algae diverged, chromists acquired chloroplasts by secondary transfer of a red algal cell (red arrow) into a host cell similar to *Colponema* (Alveolata), whose cortical alveoli and excavate-like cytoskeleton represent the ancestral state for corticates (Cavalier-Smith, 2013a,b). The fact that chromists are evolutionary chimaeras of two phylogenetically distinct corticate eukaryotes (Cavalier-Smith, 2013b), plus likely differential retention across chromist lineages of originally redundant genes of distinct ancestry, may explain some of the frequent contradictions within the corticate clade between multigene trees with different corticate taxon and gene sampling (Deschamps and Moreira, 2009).

Most Sulcozoa are gliding not swimming flagellates: apusomonads (Cavalier-Smith and Chao, 2010) and planomonads (Cavalier-Smith et al., 2008) and *Mantamonas* (Glücksman et al., 2011) (the latter two grouped as Glissodiscea) glide over surfaces on their posterior cilium held rigidly behind their cell like a ski. Sulcozoa also include Diphylleida (Brugerolle et al., 2002; Cavalier-Smith, 2003), alga-eating grooved swimming flagellates with two or four cilia, which emit pseudopodia from their groove and whose cytoskeleton is substantially modified compared with gliding Sulcozoa (Cavalier-Smith and Chao, 2010; Cavalier-Smith, 2013a). Diphylleids, sometimes included in Apusozoa or Excavata (Cavalier-Smith, 2003), were recently grouped with Glissodiscea and the non-flagellate Rigifilida (Yabuki et al., 2013) as Varisulca (Cavalier-Smith, 2013a). Previously Glissodiscea included the gliding flagellate *Discocelia* (Vørs, 1988; Cavalier-Smith, 2013a), but rRNA trees put it in Cercozoa (Cavalier-Smith et al., in preparation). Breviate amoebae lost the posterior cilium and glide instead by the remaining anterior cilium held rigidly ahead (Heiss et al., 2013), and have become secondarily anaerobic. Ciliary gliding, pseudopodia, and dorsal pellicle differentiate Sulcozoa from the excavate phylum Loukozoa, which includes *Malawimonas*, secondarily anaerobic metamonad flagellates like *Giardia*, and jakobid flagellates with the most primitive mitochondria (Lang et al., 1997; Burger et al., 2013).

A better sulcozoan phylogeny is crucial for understanding key steps in eukaryote cell evolution before the origins of animals and fungi. This is impeded by lack of multigene data for most sulcozoan lineages, whose relationships rely on morphological evidence and weakly resolved and conflicting rDNA phylogenies (Cavalier-Smith and Chao, 2010; Cavalier-Smith et al., 2008; Glücksman et al., 2011; Paps et al., 2013; Yabuki et al., 2013). Previously large-scale transcriptomic data existed for only six species, not all included together in one tree: the apusomonads *Thecamonas trahens* (the only sulcozoan with genome completely sequenced) and *Manchomonas bermudensis*, which grouped as sister to opisthokonts (Brown et al., 2013; Grant et al., 2012; Torruella et al., 2012; Zhao et al., 2013); three uniciliate breviate amoebae, *Breviata anathema*, *Subulatomonas tetraspora* and *Pygsuia biforma*, now grouped with apusomonads in Apusozoa (Cavalier-Smith, 2013a) but previously regarded as Amoebozoa (Cavalier-Smith, 2009; Minge et al., 2009); and the diphylleid *Collodictyon* that grouped weakly as sister to opisthokonts plus Apusozoa and Amoebozoa (Brown et al., 2013; Zhao et al., 2012, 2013). The first transcriptome-derived multigene tree with fewer proteins and no apusomonads grouped breviates with Amoebozoa (Minge et al., 2009). However breviates are now sisters either to apusomonads (forming a clade Apusozoa on maximum likelihood (ML) trees) or to opisthokonts plus apusomonads on evolutionarily more realistic site-heterogeneous (CAT-GTR-GAMMA) Bayesian trees (Brown

et al., 2013; Grant et al., 2012; Zhao et al., 2012, though their fast-site removal trees contradictorily placed them in Amoebozoa).

This paper studies basal podiate phylogeny and Sulcozoa in particular by sequencing partial transcriptomes from sulcozoan flagellates representing three major groups, two not previously represented on multigene trees. We did cDNA pyrosequencing for six sulcozoan species from three different orders: Apusomonadida (*Thecamonas oxoniensis*, *Multimonas media* (Cavalier-Smith and Chao, 2010)); Mantamonadida (*Mantamonas plastica* (Glücksman et al., 2011)); Planomonadida (*Ancyromonas sigmoides*, *Fabomonas tropica*, *Nutomonas howeae* (Glücksman et al., 2013)). For the evolutionarily key *Mantamonas* and planomonads we provide the first large-scale protein data. New apusomonad sequences supplement those for *Thecamonas trahens* and *Manchomonas bermudensis* (Brown et al., 2013; Torruella et al., 2012). For the first time we include both the breviate amoeboflagellates *Pygsuia* (Brown et al., 2013) and *Subulatomonas* (Grant et al., 2012) in the same tree, making our phylogenetic analyses the most comprehensive to date. Overall they indicate that Sulcozoa comprise at least three phylogenetically distinct lineages at the base of podiates, strongly support the separate phyletic position of the sulcozoan subphyla, Apusozoa and Varisulca (Cavalier-Smith, 2013a), and show that Apusozoa, Sulcozoa, Loukozoa, and Excavata are probably all ancestral paraphyletic groups of key importance for understanding early stepwise evolution of eukaryotic cells ancestral to higher kingdoms. Our analyses provide further evidence for excavate paraphyly by strongly suggesting that *Malawimonas* is sister to podiates whereas all other Eozoa are evolutionarily more distant; the evolutionarily most realistic CAT-model trees show podiates as a clade.

2. Materials and methods

RNA was extracted from uniprotist cultures of all six sulcozoans, cDNA libraries made, 454-pyrosequencing done, and 192-gene alignments (52,824 amino acid positions) constructed for >200 eukaryotes. Supplementary Table S3 lists the genes. Because of the complexity of data preparation, comprehensive technical details are in the electronic supplementary material. Briefly however, RNA was extracted using TRI-Reagent (Sigma) protocol (Chomczynski and Sacchi, 1987); cDNA libraries were constructed by Vertis AG using their full-length enriched protocol and PCR amplification – those for *Multimonas media* and *Ancyromonas sigmoides* were normalised. Sequencing was multiplex: tagging individual libraries with specific oligonucleotide markers allowed computational separation of taxa after pooling DNA for two 454-sequencing runs on two separate half plates that also included 18 other individually tagged protist libraries (reported in separate papers). Some cross-contamination between certain samples must have occurred at some stage, as some incorrectly labelled sequences were unambiguously identified in the ML trees run for quality control of all 192 single-gene alignments and excluded from the multigene analyses (as were similar contaminants identified in databases from other laboratories). Bootstrapped single-gene trees combined with careful inspection of all alignments were used to remove duplicate and incorrect paralogue sequences. Amino acid positions included in the analyses were similar to those of the seed alignments from published multigene analyses (see supplementary material) to which our and additional outgroup sequences were added. Sequences were concatenated for multigene trees by SCaFoS (Roure et al., 2007) for each gene and taxon sample analysed.

A eukaryote-wide subset of 75 taxa was used for phylogenetic analysis by the best available site-heterogeneous amino-acid substitution model (PhyloBayes-MPI v.1b GTR-CAT-Γ-4rates) (Lartillot and Philippe, 2008; Philippe et al., 2011) and for maximum likelihood (ML) by RAxML-MPI v.7.2.8 or 7.7.2 PROTGAMMALGF (a site-homogeneous model) as detailed in supplementary material. After *Pygsuia* sequences became available (Brown et al., 2013) we used both methods to study a second subset of 73 taxa that excluded all eozoan outgroup sequences but was substantially augmented by including additional podiate and corticate taxa (i.e. neozoa) as well as adding extra sequences for already included taxa from recently sequenced complete genomes represented only by incomplete transcriptomes in our first analyses. To reduce the proportion of missing data, analyses for these 73 neozoa excluded 19 less well sampled genes (marked in Table S3) and thus used only 173 genes. To study the effects of incomplete data for some taxa, we also ran 73-taxon trees by both methods for shorter alignment subsets that excluded successively more less-well represented genes; for that purpose we used SCaFoS (Roure et al., 2007) to construct shorter alignments restricted to genes missing in less than 50% of taxa, in less than 40% of taxa, and in less than 30% of taxa. ML trees were also done for taxonomically richer alignments (e.g. 86 or 122 taxa, not computationally practicable for PhyloBayes) for the 192 genes, as well as for alignments including only the 178 genes represented in our new sulcozoan data and for 98 eukaryote-wide taxa using the same 173 genes as for the 73-taxon trees (and for reduced alignments for these taxa for genes with <50%, <40%, and <30% missing data). For the 192-gene alignment we also used the program SlowFaster (Kostka et al., 2008) to make five other reduced alignments, which excluded the fastest evolving 5%, 10%, 19%, 30% and 40% of amino acid positions; see supplementary material) and calculated ML trees for each. New sequences obtained in this work have been deposited in GenBank under six BioProject numbers: PRJNA195917-9; PRJNA195922; PRJNA195923; PRJNA195925.

3. Results

3.1. Distinct phylogenetic positions of Varisulca and Apusozoa

The two varisulcan groups lacking any previous phylogenetic trees based on numerous genes (planomonads and *Mantamonas*) invariably branch in all our analyses at the very base of podiates close to *Collodictyon*, the only other varisulcan previously with extensive multigene sequences. All three planomonads form a maximally supported clade by both methods in all trees, with *Nutomonas* sister to *Ancyromonas* (invariably maximally supported), robustly confirming the weakly supported 18S rDNA tree topology (Glücksman et al., 2013). In Fig. 2, podiates are a clade by both methods (moderately supported by CAT, insignificantly by ML) and all five Varisulca clearly branch below the last common ancestor of opisthokonts and Amoebozoa. By contrast, the two lineages of the other sulcozoan subphylum Apusozoa branch above the opisthokont–Amoebozoan divergence as sisters to opisthokonts only – either as a single clade by ML or as two successive breviate and apusomonad clades by CAT, just as Brown et al. (2013) previously showed. For the first time this shows unambiguously that *Mantomonas* is not related to apusomonads, contrary to 28S rDNA trees (Glücksman et al., 2011), but is more closely related to *Collodictyon* (with which it is weakly sister by CAT, PP 0.51) or planomonads (weakly sister by ML: BS 57%). Though these distinct and contrasting phylogenetic positions of the two sulcozoan subphyla Varisulca and Apusozoa are consistent by both methods, ML and CAT differ for their internal phylogeny: with ML, Apusozoa (breviates, apusomonads) and Glissodiscea (planomonads, *Mantamonas*) are both holophyletic but with CAT both are paraphyletic. The two methods are also contradictory for *Malawimonas*: sister to podiates by CAT (PP 0.78) and insignificantly (39%) to metamonads by ML.

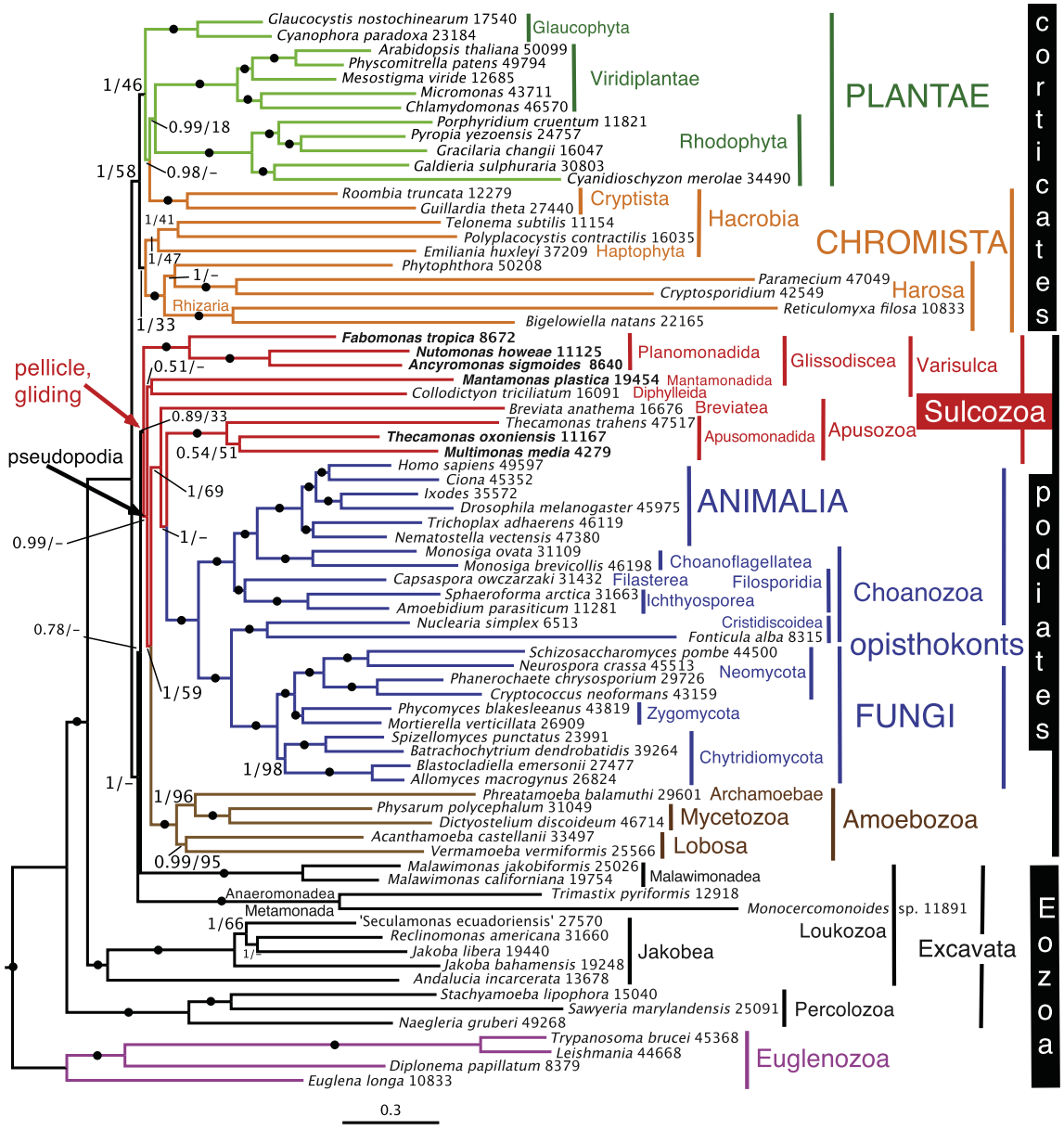

Fig. 2. Bayesian GTR-CAT tree for 75 eukaryotes based on 52,824 amino acid positions in 192 protein-coding genes. Two well-converged chains (maxdiff 0.080514) were summed. Support values for bipartitions are posterior probabilities (PP left) and bootstrap support (BS) percentages (100 pseudoreplicates) for a separate RAxML PROTGAMMALGF tree for the same alignment (right) are given; a black blob means both were maximal (1.0 or 100%); bipartitions not found in ML are shown by -. The number of amino acids included for each taxon follows its name; the six sulcozoan taxa sequenced here are in bold. In eight genera sequences for closely related species are combined as specified in Supplementary Table S1. Vertical labels denote the three eukaryotic supergroups and capitals the four derived kingdoms: all other branches belong to the basal kingdom Protozoa. The tree is rooted between Euglenozoa and excavates as shown by universal ribosomal protein trees (Lasek-Nesselquist and Gogarten, 2013), and a dozen other arguments (Cavalier-Smith, 2010a, 2013a). Arrows show likely points of origin of sulcozoan pellicle, ciliary gliding, and pseudopodia if this topology showing paraphyletic Varisulca were correct; however if Varisulca are holophyletic as in all corresponding ML trees and also the 73-taxon CAT tree with the lowest proportion of missing data (Fig. S2), arguably more likely (see text), then all three of these key sulcozoan characters probably evolved simultaneously in the ancestral podiate. The basal podiate and corticate near-simultaneous rapid radiations may be related to the origin of chloroplasts (perhaps corresponding to increased fossil cell complexity ~800 My ago (Cavalier-Smith, 2013a,b)), apparently long after the primary eukaryote bifurcation between excavates and Euglenozoa.

Taxon-rich trees are commonly better than sparse ones because more internal branches are broken allowing more accurate phylogenetic reconstruction, so we also ran trees with substantially more neozoan and podiate taxa. Fig. 3 omits the more distant excavate outgroups to allow a denser taxonomic sampling for neozoa (podiates and corticates). It also omitted most genes not represented in our six new transcriptomes, so as to reduce possible artefacts from too high a proportion of missing data (Roure et al., 2013), and includes three more apusozoan sequences unavailable when Fig. 2 was run (Brown et al., 2013); moreover, for some taxa, extra genes were added from full genomes previously absent from partial transcriptomes. Theoretically all four differences should make the Fig. 3 trees more accurate. Reassuringly, despite considerable differences in taxon and gene sampling, podiate and sulcozoan topology is identical to Fig. 2 for CAT except for a different and now convincingly and robustly supported internal phylogeny of the better sampled apusomonads. Moreover, ML now also weakly shows the *Mantamonas*/*Collodictyon* clade (59%) and thus now agrees in this respect with CAT; also support for obazoa (Apusozoa, opisthokonts) increased from 69% to 89% as it does for the branching of Varisulca below obazoa from 59% to 81%. These improvements to the ML tree show the greater importance

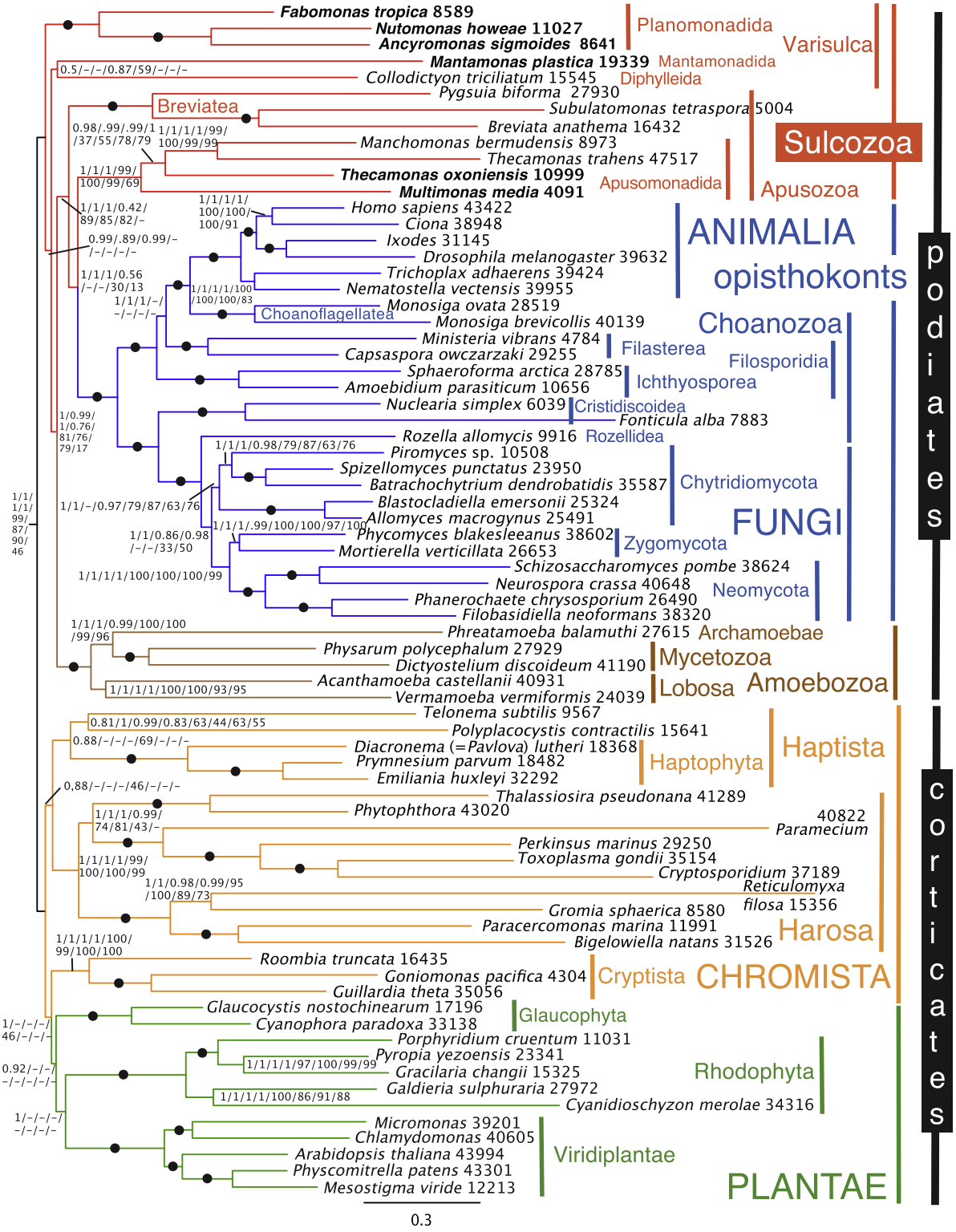

Fig. 3. Bayesian GTR-CAT tree for 73 neozoan eukaryotes based on 45,194 amino acid positions in 173 protein-coding genes. Support values for bipartitions are posterior probabilities (leftmost for the complete alignment, next for the <50% missing genes subalignment – 30,538 amino acids, then the <40% missing genes subalignment – 17,470 amino acids, and last the <30% missing genes alignment - 8680 amino acids) and bootstrap percentages (100 pseudoreplicates, first for the complete alignment, next for the smaller <50%, <40%, and <30% missing gene subalignments) for separate RAxML PROTGAMMALGF trees are given; black blobs indicate maximal (1.0 or 100%) support in all eight trees; bipartitions not found in relevant trees are shown by -. The number of amino acids included for each taxon in the complete alignment follows its name (augmented for some taxa by adding extra genes compared with Figs. 2 and 5 and S1); the six sulcozoan taxa we sequenced are in bold. This tree is for one of four independent chains that did not converge because each had a different topology within corticates; as all four had identical topology for all Sulcozoa the one most consistent with other evidence for chromists is used for this figure; as discussed in the text, two chains also conflicted with others for the position of *Rozella* within opisthokonts. The two chains run for the CAT tree for the <50% and <40% missing gene alignments also did not converge because of the same conflict over *Rozella*; bootstrap values for them are for the chain with the topology shown; both these pairs of chains agreed for all Sulcozoa and chromists. However, the two chains for the CAT tree for the <30% missing gene alignments did converge (maxdiff 0.228) and Varisulca had a different topology (see Fig. S2) that is identical to that in ML for the <40% and <30% missing gene alignments (see text).

Table 1
Summary of the numbers of genes and taxa for the various trees.

Figure	Taxa	Composition	Algorithm	Genes	Amino acids	Purpose
Fig. 2	75	Eukaryote-wide	CAT/ML	192	52,824	Simple summary
Fig. 3	73	Neozoa (no Eozoa)	CAT/ML	173 or less[a]	45,194[a]	Missing data test
Figs. 4 + S3–S5	98	Eukaryote-wide	ML	173 or less[a]	45,194[a]	Missing data test for more taxa
Fig. 5	122	Eukaryote-wide	ML	178	47,510	Faster site removal
Fig. S1	86	Non-chromists extra excavates	CAT	178	47,510	Taxon sampling test
Fig. S1A	86	Non-chromists extra excavates	ML	178	47,510	Taxon sampling test
Fig. S2	73	Neozoa (no Eozoa)	CAT	[b]	8680	Missing data test

[a] These tests calculated trees for both the full 173-gene alignment and also after omitting successively more genes for which data were absent for >50%, >40%, and 30% of taxa, thus also analysed fewer genes and amino acids (the indicated number of amino acids is for 173 genes; those for the smaller alignments with proportionally less missing data are in Fig. 3 legend).
[b] This tree is for the same taxa and gene set as Fig. 3 after excluding all genes missing in >30% of taxa.

of good taxon- and gene-sampling for ML than for CAT, which maximally supports these clades in both trees. Corticate basal phylogeny is also improved, with Plantae monophyletic in Fig. 3, not polyphyletic as in Fig. 2; however chromists still appear paraphyletic with cryptists closer to Plantae than to other Chromista. Thus podiate basal tree topology is more stable than it is for corticates.

Nonetheless the closeness of the varisulcan and *Malawimonas*/metamonad divergences and the conflicts between ML and CAT near this interface between podiates and excavates on both trees made further tests of tree stability desirable, for which numerous additional trees were run. Overall these trees drew on 152 eukaryotic taxa and represent the most thorough and extensive multigene analysis to date of basal eukaryotic branching. To orient the reader in their more detailed description below, Table 1 summarises the different figures and alignments analysed and their purposes.

A key issue is the potential distorting effect of disproportionately high levels of missing sequence in some taxa, an artefact to which ML seems more prone than CAT, though neither is immune (Roure et al., 2013). This is especially important for Varisulca, *Malawimonas*, and the metamonad anaeromonads as no genomes or near-complete transcriptomes are available for any of them and they therefore inevitably have proportionally more missing data than most other clades on the trees. We therefore investigated this for the Fig. 3 alignment by systematically excluding genes missing in higher proportions of the taxa (Fig. 3). To study whether increasing taxon sampling substantially above the number computationally feasible for running the CAT trees of the 75 and 73 taxon alignments used for Figs. 2 and 3 would make ML trees more compatible with the probably more accurate CAT trees, we carried out similar sparsely-represented-gene removal tests on a eukaryote-wide 98-taxon sample including more excavates than before (Fig. 4). A different test, used by others concerned by the inconstant position of *Malawimonas* on trees (Derelle and Lang, 2012; Hampl et al., 2009; Zhao et al., 2012), is faster-evolving-site removal to examine tree reproducibility when slower evolving sites are given greater weight. As this sometimes makes ML trees agree better with CAT trees, Fig. 5 shows an ML tree for the largest eukaryote-wide alignment to date for protist multigene analyses (122 taxa), and summarises the results of removing faster-evolving sites (corresponding individual trees after faster-site removal are in Figs. S3–S5).

Finally, to investigate effects of including more excavate outgroups on the basal branching order of podiates and related excavates, a sample of 86 eukaryotes and 178 genes (excluding some of the most sparsely represented genes in our larger alignment) was analysed by CAT and LG ML (Fig. S1), primarily to see the effect of including the long-branch diplomonad excavates, excluded in other trees because of concerns about their possibly distorting the rest of the tree. This had precisely the same sulcozoan topology as Fig. 2 for ML and Fig. 3 for CAT, showing that for Sulcozoa it does not matter whether long-branch diplomonads and trichomonads are both included (Fig. S1), both excluded (Fig. 2) or only Parabasalia included (Fig. 3) – though excavate topology differed in Fig. S1. Fig. S1 excluded Chromista, partly to reduce computational time and partly because Deschamps and Moreira (2009) showed that including chromist outgroups can distort basal branching order of Plantae, and in Fig. 2 Cryptista even intruded into Plantae, and we wanted to check the internal branching order of Plantae in the absence of possibly distorting chromists, using a much larger alignment and many more outgroups and genes than did Deschamps and Moreira (2009). The CAT topology of Fig. S1 confirms that found by Deschamps and Moreira in the absence of chromists (red algae, Viridiplantae sisters), but the corresponding ML topology (Fig. S1A) was contradictory with glaucophytes strongly (92%) sisters of Viridiplantae (as in some other ML trees Figs. 4 and 5, not Fig. 2). However, in contrast to Fig. S1 where metamonads were seemingly polyphyletic (Trichozoa sister to obazoa not to anaeromonads) Metamonada are a clade with maximal support and branching as sister to podiates/*Malawimonas* plus corticates in ML trees (Figs. 5 and S1A). More importantly for this paper, excluding all Chromista, like all Eozoa, did not alter the sulcozoan branching order, which is therefore robust to immense changes in sampling distant taxa.

3.2. The global eukaryotic tree

Corticates are a maximally supported clade in the CAT trees of Figs. 2 and 3, and S2; they are also a clade with low to strong support in all ML trees (Figs. 2–5). The Fig. 2/3 CAT trees have a maximally supported bipartition between podiates and Eozoa. Fig. 2 also weakly shows this bipartition weakly by ML. All CAT trees show *Malawimonas* and metamonads as more closely related to podiates (usually their immediate outgroups) than to other excavates or to corticates. Three ML trees did not show the podiate/eozoan bipartition because *Malawimonas* intrudes into podiates and into Varisulca, becoming sister to *Collodictyon* and displacing *Mantamonas* (Figs. 4, 5 and S1) as Zhao et al. (2012) observed with a CAT model. Almost all branches within opisthokonts are consistently maximally supported. Neither method showed Plantae and Chromista as clades, because Cryptista was within Plantae – sister to glaucophytes in ML (BS 62%), contradictorily sister to the red algal/green plant clade in PhyloBayes CAT (PP 0.98). With both phylogenetic methods and taxon samples (Figs. 2 and 3), Sulcozoa appear as two distinct groups corresponding to subphyla Apusozoa and Varisulca, each weakly holophyletic on ML trees.

3.3. Detailed sulcozoan phylogeny

3.3.1. Apusozoa

The anaerobic breviate amoeboflagellates are sisters to the aerobic pseudopodial apusomonads in ML trees for 192 or 173 genes: support for an apusozoan clade, is 71% for Fig. 2 and 75% for Fig. 3).

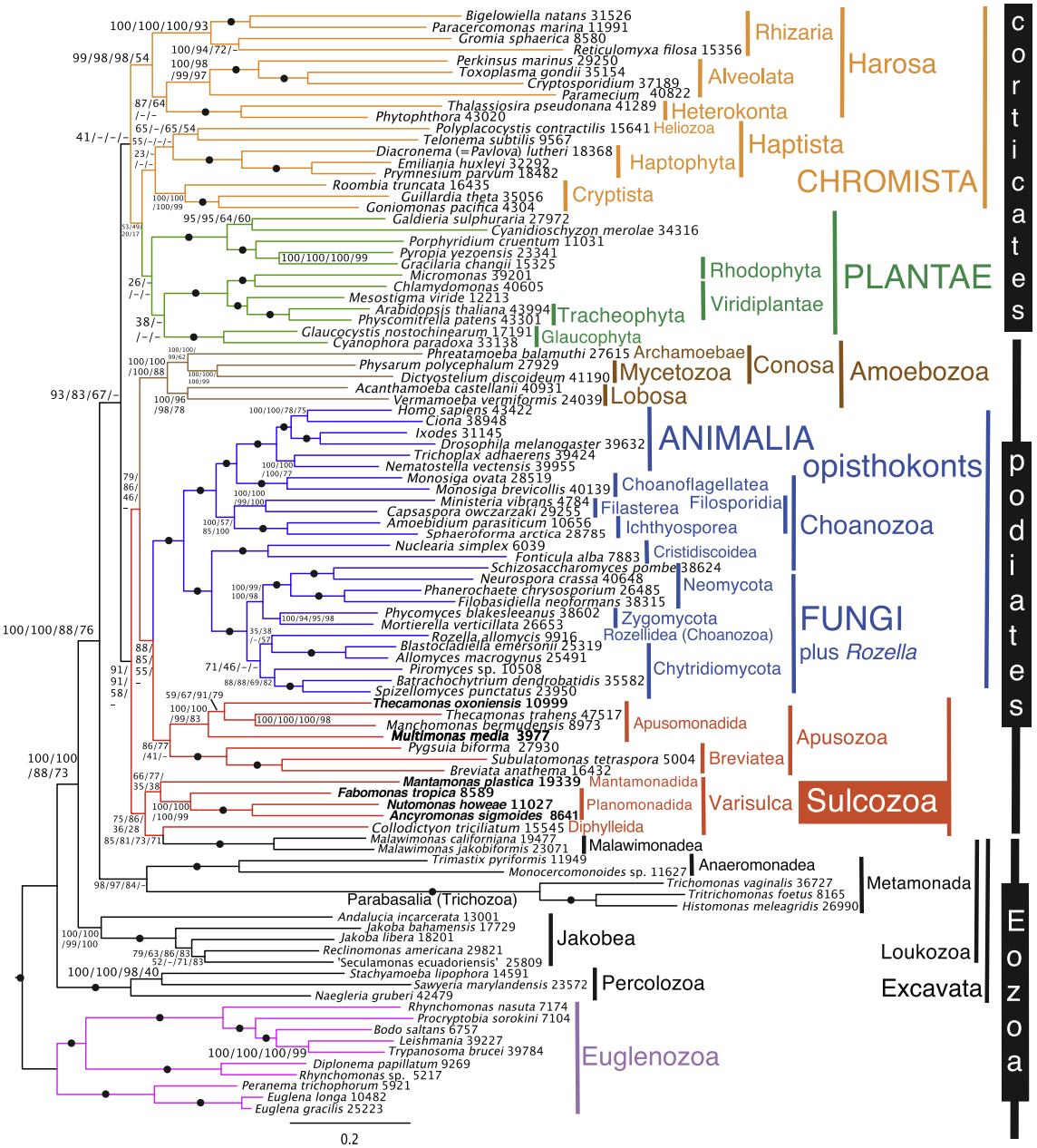

Fig. 4. Maximum likelihood tree for 98 eukaryotes based on 45,194 amino acid positions in 173 protein-coding genes. Support values for bipartitions are bootstrap percentages (100 pseudoreplicates, first for the complete alignment, next for the smaller <50%, <40%, <30% missing gene subalignments, as in Fig. 3) for separate RAxML PROTGAMMALGF trees are given; black blobs indicate maximal (100%) support in all four trees; bipartitions not found in relevant trees are shown by -. The number of amino acids included for each taxon in the complete alignment follows its name (augmented for some taxa by adding extra genes compared with Figs. 2 and 5, and S1); the complete trees for the subalignment showing the number of amino acids retained for each taxon are in supplementary material (Figs. S3–S5); the six sulcozoan taxa we sequenced are in bold.

However all CAT trees for the present taxon samples instead show apusomonads as sister to opisthokonts with breviates branching one node more deeply as their sisters (CAT: PP 1 in Figs. 2 and 3 and Supplementary Fig. S1 for 86 taxa). Fig. 3 shows that when genes with proportionally lower taxonomic representation are successively excluded from the analyses, this contradiction between LG and CAT trees disappears and eventually Apusozoa become paraphyletic by both methods with identical topology, albeit with lower support by the less realistic LG than with CAT. The ML tree for the full alignment and that using only genes missing in <50% of taxa (30,538 amino acid positions, 35.21% gaps) showed Apusozoa as a clade, those using only genes missing in <40% (17,470 amino acids, 27.39% gaps) or <30% (8680 amino acids, 21.22% gaps) of taxa showed Apusozoa as paraphyletic and with identical topology to CAT trees. Thus Apusozoa are probably paraphyletic, as Brown et al. (2013) also argued from the markedly better fit of the data to CAT than to LG. The contradiction by ML using genes with >40% of taxa missing is probably because the site-homogeneous LG model is less able to model ancestral states for the apusozoan clade than is the evolutionarily more realistic CAT model when genes are missing for too high a fraction of the taxa (Roure et al., 2013). This multigene tree has the most comprehensive sampling for Apusozoa to date, and shows for the first time that *Subulatomonas* is robustly sister to *Breviata* not *Pygsuia* and shows that

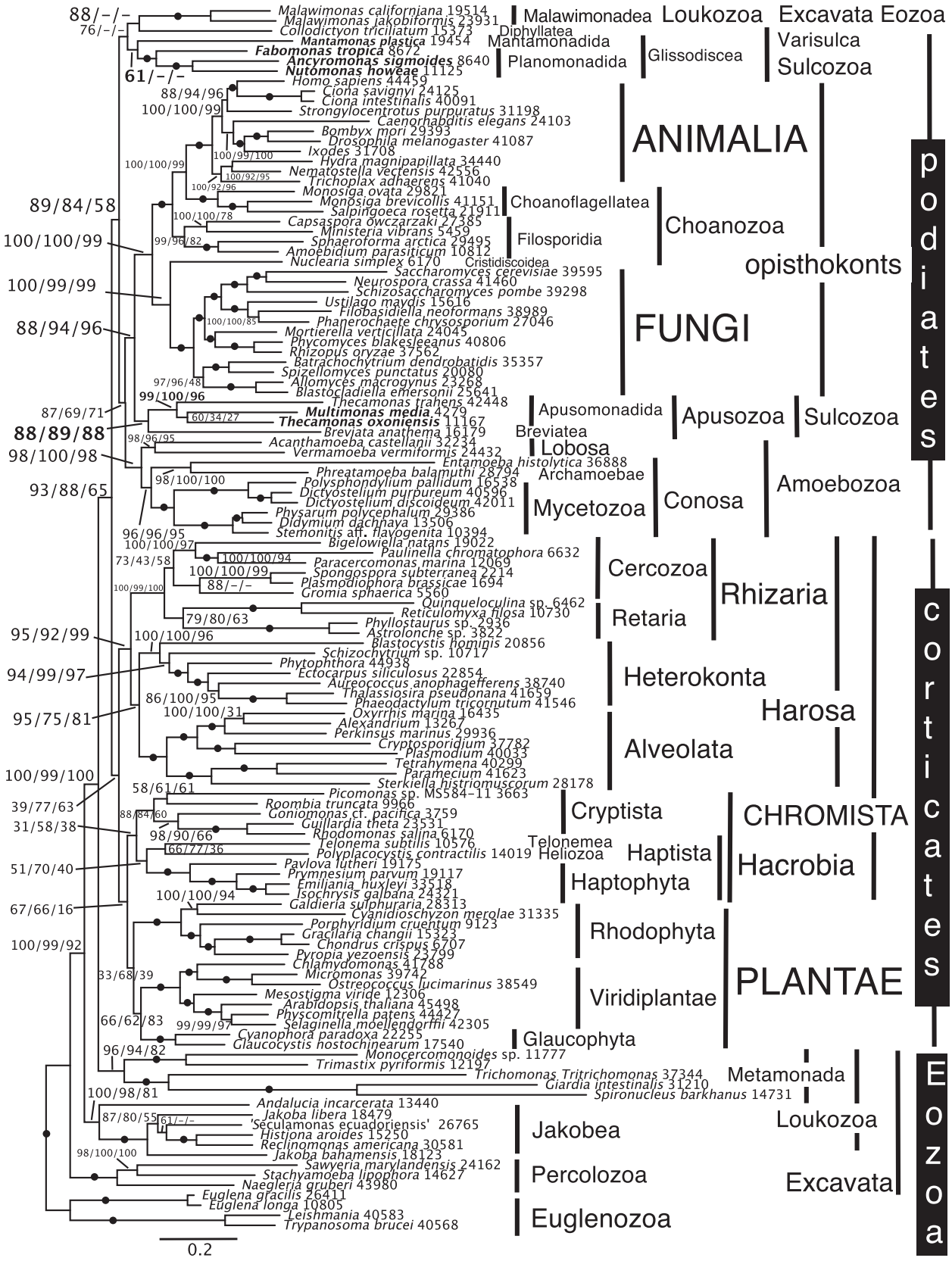

Fig. 5. Maximum likelihood tree for 122 eukaryotes based on 47,510 amino acid positions in 178 genes (14 genes unrepresented in our new sulcozoan data excluded). This RAxML PROTGAMMALGF (4 gamma rates) tree is rooted within Eozoa between Euglenozoa and Excavata according to Cavalier-Smith (2010a). The number of amino acids included for each taxon follows its name; the six sulcozoan taxa we sequenced are in bold. Vertical labels denote the three eukaryotic supergroups and capitals the four derived kingdoms: all other branches belong to the basal kingdom Protozoa. Bootstrap support (100 pseudoreplicates) for bipartitions are shown from left to right for (1) the complete data; (2) after removing the fastest 19%, or (3) 40%, of amino acid positions; a black blob means 100% support in all 6 trees with varying site removal. In six cases sequences for closely related species are combined as specified in Supplementary Table S1.

Thecamonas trahens is much more closely related to *Manchomonas* than to *Thecamonas oxoniensis*, proving that *Thecamonas* as presently constituted is paraphyletic (suspected from 18S rDNA trees: Cavalier-Smith and Chao, 2010).

3.3.2. Varisulca

Our trees are the first using numerous genes to include glissodisceans (planomonads and *Mantamonas*). Planomonads are always a clade with maximal support. *Nutomonas* and *Ancyromonas* are maximally supported as sisters, which decisively supports their grouping as Ancyromonadidae in contrast to the earlier diverging *Fabomonas* in agreement with ciliary structure and much less robust 18S rDNA trees (Glücksman et al., 2013). Several CAT trees weakly group *Mantamonas* with *Collodictyon*; support for this clade varies with taxon and gene sampling from PP 0.5 to 0.87 (Figs. 2, 3 and S1). With LG this clade is seen only if *Malawimonas* is excluded (Fig. 3), but it is found only with the complete gene set (59% support); whenever less well represented genes are excluded *Mantamonas* groups instead with planomonads with weak support (<50% BS 39%; <40% 54%; <30% 26%) forming a clade corresponding to the class Glissodiscea – the ML support value peaking at the intermediate degree of gene removal suggests this may be the correct topology. Removing genes missing in >50% of taxa moved *Mantamonas* away from *Collodictyon*, breaking Varisulca into three distinct lineages: planomonads sister to Amoebozoa/Apusozoa/opisthokonts (PP 0.88), *Collodictyon* sister to all of them (PP 0.52) and *Mantamonas* the deeply branch sister to all other podiates; however, the CAT tree for <40% missing retained a *Mantamonas*/*Collodictyon* clade; most strikingly, when only genes missing in less than 30% of taxa were used CAT topology agreed with that of all three gene-removal ML trees (PP 0.64 for Glissodiscea Fig. S2; based on both chains, which converged even for *Rozella* – unlike the other 73-taxon CAT trees). The fact that varisulcan topology agrees by both methods when the proportion of missing data is sufficiently reduced and that it is also the only one found that agrees with the morphological reasons for establishing class Glissodiscea (Cavalier-Smith, 2013a) makes it likely that Glissodiscea is a clade and that CAT and ML trees may be distorted in contradictory ways when the proportion of missing data is too high. Had we not conducted the sparse-gene removal study for both methods we might have concluded (probably wrongly) that the 192- and 173-gene CAT gene trees were correct and that Glissodiscea are paraphyletic. *Collodictyon* was sister to Glissodiscea (making a varisulcan clade) in all Fig. 3 trees showing clade Glissodiscea (<50% BS 65%; <40% 63%; <30% 12%, PP 0.45), weakly indicating that Varisulca are probably holophyletic. Whenever *Malawimonas* is included in ML trees it groups with *Collodictyon* (BS 71–85%), and *Mantamonas* groups instead (with weak to moderate support) with planomonads forming a glissodiscean clade (BS 57% for Fig. 2 and 61% Fig. 4).

These trees provide the first strong evidence that both planomonads and *Mantamonas* branch more deeply than the common ancestor of Amoebozoa and opisthokonts in the same tree region as the diphylleid *Collodictyon*, and that Varisulca collectively are the most deeply divergent of all podiates, i.e. *Mantamonas* does not group with apusomonads as it did on 28S rDNA trees (Glücksman et al., 2013). Less strongly we conclude that Varisulca are probably a clade. The trees for subalignments with proportionally less missing data suggest that incomplete data are a problem in several other areas in the tree, most seriously at the base of corticates as indicated by changing support values that this produced (Fig. 3).

To see if such conflicts are reduced by better taxon sampling within Eozoa (but excluding the longest diplomonad branch present in Fig. S1) we also analysed by RaxML-MPI v. 7.7.2 PROTGAMMALGF an enlarged eukaryote-wide alignment for 98 taxa using all 173 genes (45,190 amino acid positions, 47.15% gaps) plus three reduced alignments including only genes with <50% (24, 657 amino acid positions, 34.87% gaps), <40%, (17,470 amino acid positions, 27.39% gaps) and <30% (8680 amino acids, 21.22% gaps) missing taxa. As Fig. 4 shows, in all four trees *Malawimonas* was still within podiates (BS 91%, 91%, 58%, 33%) and Varisulca (75%, 86%, 36%, 28%), and sister to *Collodictyon* (85%, 81%, 73%, 71%). The steady drop in support for this position and thus reduced conflict with the CAT trees supports the interpretation that there are not enough data for a site-homogeneous model to reconstruct accurately ancestral sequences for *Malawimonas* and Varisulca. *Mantamonas* was always sister to planomonads (clade Glissodiscea: BS 66%, 77%, 35%, 38%). As in the 73-taxon neozoan trees (Fig. 3), support for Apusozoa being a clade declined markedly as genes with more missing data were removed (Fig. 4: BS 86%, 77%, 41%, -); and in the <30% missing-gene tree breviates moved to group weakly with anaeromonad excavates and Amoebozoa, probably an artefact of insufficient data. Within podiates most established groups (opisthokonts, Amoebozoa, breviates, and planomonads) were invariably maximally supported as clades, as were apusomonads except in the smallest <40% and <30% alignments where they remained strongly holophyletic and their internal topology robustly identical to the 73-taxon tree.

To see whether the intrusion of *Malawimonas* into podiates on ML trees is just a consequence of using the LG model we also ran a RAxML GTRGAMMA tree for the complete alignment from Fig. 4. Its only topological difference was that Harosa were misplaced below the divergence of podiates/*Malawimonas* and other corticates (not shown). This suggests that for the long-branch Harosa, the possibly overgeneralised GTR model reconstructs ancestral sequences less well than LG, whose parameters are derived from a much larger empirical data set than the present alignment and may therefore be more accurate. Most bootstrap values were identical, though a few non-maximal ones were higher and a similar number lower.

As Supplementary Fig. S1 has PhyloBayes tree topology identical to Figs. 2 and 3 except for the weakly supported internal branching of Jakobea and Apusomonadida, the main features of CAT trees are insensitive to removing all 10 chromists and adding 21 taxa to the remaining groups, including trichozoan metamonads omitted from Fig. 2 because of their long branches and because Hampl et al. (2009) found their presence or absence altered the position of *Malawimonas* on their 48-taxon 143-gene trees. Reassuringly, using many more taxa and genes stabilised the position of *Malawimonas* on our Bayesian trees; neither its position nor that of any sulcozoan groups altered with big changes in metamonad sampling. Unexpectedly however, Fig. S1 shows trichozoan metamonads within podiates, separately from anaeromonad metamonads; the corresponding ML tree put *Malawimonas* sister to *Collodictyon* (78% support). The position of *Malawimonas* may therefore be affected by taxon sampling more on ML than CAT trees. ML trees for an even larger taxon sample of 122 taxa also placed *Malawimonas* within Varisulca and robustly monophyletic Metamonada significantly lower (Fig. 5).

3.4. Removing faster sites less useful than proportionally reducing missing data

In marked contrast to earlier studies with no sulcozoans or only one, successive removal of up to 40% of faster evolving sites did not cause *Malawimonas* to move downwards and join other excavates (Fig. 3). The composition and separate positions of Varisulca and Apusozoa remained identical and strongly supported, though within Varisulca *Malawimonas* and *Mantamonas* interchanged positions with 10% or more removal; the resulting planomonad/*Malawimonas* 'clade' moved immediately below the new

Mantamonas/*Collodictyon* clade (which remains sister to opisthokonts/Apusozoa/Amoebozoa); the position of these two clades remained stable when 19–40% of faster sites were removed. Thus after 10% faster-site removal varisulcan ML tree topology was the same as with CAT, except for *Malawimonas* also being within Varisulca as sister to planomonads only rather than sister to all podiates. Fast-site removal did not reduce the stability of or reverse the likely artefactual holophyly of Apusozoa (Fig. 5), as did reducing the proportion of missing data (Fig. 3), suggesting that for some clades the latter is a better way of reducing site-homogeneous-model tree reconstruction artefacts. Removing faster sites did move *Mantamonas* to be sister to *Collodictyon* and thus made this tree contradict Fig. 4 and the CAT < 30% tree that showed a glissodiscean clade, arguably making the tree worse. Faster site removal is probably not a good way of eliminating long-branch problems, and could reduce some real phylogenetic signal. Previous interpretations of the movement downwards of *Malawimonas* following fast-site removal as correct history (Derelle and Lang, 2012; Hampl et al., 2009; Zhao et al., 2012) are not supported by our analyses using more genes and taxa. In one case, site-removal clearly wrongly placed *Breviata* apparently robustly within Amoebozoa (Zhao et al., 2012). Adding here many more Sulcozoa, by anchoring *Breviata* in a probably more correct position, also supported by ultrastructural characters (Cavalier-Smith, 2013a; Heiss et al., 2013), eliminated that signal-loss artefact. Breaking up branches by adding more taxa, especially shorter-branch ones, is a sounder way than faster site removal of reducing long-branch artefacts.

In all analyses both site-heterogeneous and homogeneous trees robustly place all varisulcan lineages (planomonads, *Collodictyon*, *Mantamonas*) below the common ancestor of Amoebozoa and opisthokonts, and put Apusozoa (apusomonads and breviates) above that common ancestor, specifically with opisthokonts, confirming previous arguments that Sulcozoa are a paraphyletic phylum ancestral independently to Amoebozoa and opisthokonts (Cavalier-Smith, 2013a). Figs. 2, 3, and S1 with very different taxon sampling and somewhat different gene sampling confirm that paraphyly of Apusozoa on site-heterogeneous trees is robust to taxon sampling, and Figs. 2–5 and S1–S5 with extremely different taxon and gene sampling and two contrasting algorithms all show Varisulca as the deepest branching podiate clade.

3.5. Paraphyly of excavates

All our trees show that the excavate *Malawimonas* is more closely related to podiates than to corticates or to jakobid or percolozoan excavates. Our best CAT trees suggest that metamonad excavates are sister to podiates plus *Malawimonas* (Fig. 2 confined to the short-branch anaeromonad metamonads) or even branch within podiates (Fig. S1, including also the long-branch Trichozoa, but we trust this tree less as its large taxon number allowed running only one chain, which plateaued well and have found using larger taxon samples including chromists and an additional parabasalian (not shown) that all metamonads more often than not form a robust clade on CAT trees that is sister to podiates plus *Malawimonas*). All but two ML trees contradictorily placed Metamonada as sister to podiates/*Malawimonas* plus corticates; however, it is possible that this deeper position compared with CAT trees is a long-branch artefact; support for this deeper position declined progressively as the proportion of missing data was lowered (93%, 83%, 67%, -), and the <30% tree (Fig. S4) placed anaeromonads but not Trichozoa weakly within podiates. Thus all our trees show that both *Malawimonas* and Metamonada are more closely related to podiates than they are to jakobids. The 159-protein (43,319 amino acid) 68-taxon trees of Brown et al. (2013) weakly grouped *Malawimonas* and the sole included metamonad (*Trimastix*) together, but we never saw that on our trees more broadly sampled for excavates. However, the position of their *Malawimonas*/*Trimastix* clade by CAT (and of *Malawimonas* alone by ML) was the same as for these genera separately on our CAT trees, though made less obvious by their arbitrary rooting of the tree between podiates and other eukaryotes – which is contradicted by prokaryote-rooted ribosomal protein and mitochondrial protein multigene trees (Lasek-Nesselquist and Gogarten, 2013; Zhao et al., 2013) and the dozen characters discussed by Cavalier-Smith (2010a, 2013a), which all place the root within Eozoa, either between excavates and Euglenozoa or (mitochondrial protein tree only) within excavates between *Malawimonas* and jakobids. Support for the deeper branching of metamonads than excavates in Fig. 2 is not high (PP 0.61, 79%) so our trees cannot exclude the possibility that *Malawimonas* and metamonads are a clade that is sister to podiates (Brown et al., 2013), rather than two successively branching independent sister clades to podiates, as our taxonomically better sampled Fig. 2 suggests (supported also by several unshown CAT trees with still greater taxon sampling).

Irrespective of where one places the eukaryotic root, excavates are clearly a paraphyletic or ancestral group from which at least two of the groups podiates, corticates, and euglenoids evidently evolved by independently losing the ultrastructural characters that define excavates. If the root is either between jakobids and *Malawimonas* or between jakobids and all other eukaryotes, then all three groups evolved from excavates; if the root were between podiates and all other eukaryotes, then corticates and Euglenozoa evolved from excavates; if it were between Euglenozoa and Excavata, as the largest mass of evidence indicates, then only corticates and podiates evolved from excavates, and Euglenozoa diverged earlier.

3.6. Corticate outgroup phylogeny with contrasting algorithms and alignments

As proteins used for multigene trees consistently fit the site-heterogeneous CAT model better than the homogeneous LG model (Philippe et al., 2011), it is unsurprising that the Fig. S1 CAT tree correctly shows the monophyly of vascular plants, whereas its ML tree wrongly groups the tracheophyte *Selaginella* with the moss *Physcomitrella*. Practical superiority of CAT is also shown by another improvement in Plantae: Viridiplantae are sister to Rhodophyta with maximal support on all three PhyloBayes trees (and very weakly so on the 75-taxa ML tree), as is probably correct from other evidence (wrongly grouped with Glaucophyta in 86-, 98-, and 122-taxa ML: 92%, 32%, and 66% support). Intrusion of Cryptista into Plantae on one PhyloBayes trees (Fig. 2), sometimes seen by others (Brown et al., 2013; Burki et al., 2012), is probably incorrect, and did not occur in the better-sampled Fig. 3. Some ML trees (Figs. 2 and 4) show the same, but Figs. 3 and 5 show holophyletic Plantae. Instability in ML basal branching order within Plantae and inclusion or not of Cryptista may simply result from basal corticated stems being shorter than for Sulcozoa, making their order even harder to reconstruct reliably.

Though corticates normally form a reasonably well supported clade on multigene trees, as they do here, their basal branching order is notoriously unstable and different in almost all recently published trees (Burki et al., 2009, 2012, 2013; Brown et al., 2013), which the inconsistencies amongst our trees for the branching order of Hacrobia confirm. The simplest explanation of that instability is that only extremely short stems separate basal corticate lineages on the tree, so very few subsequently stable amino acid substitutions occurred in these short intervals. Coupled with the large number of very early diverging corticate clades, this makes basal tree reconstruction intrinsically harder than for podiates or Eozoa and indicates that Plantae and Chromista both

underwent an extremely rapid near simultaneous radiation shortly after the symbiotic origin of chloroplasts and the secondary transfer of a red algal plastid to chromists (Cavalier-Smith, 2009, 2013b). Addition here of several Varisulca to the tree shows that the basal podiate radiation also took place essentially contemporaneously with that of corticates and was similarly rapid.

3.7. Opisthokont phylogeny

In marked contrast to podiates and corticates, basal opisthokont radiations were later and well spread out over time, making them inherently easier to resolve and robustly supported on all trees. However, Figs. 3 and 4 include the flagellate parasite *Rozella*, sometimes classified as a chytridiomycete fungus, but which was recently excluded from Fungi and placed within the protozoan phylum Choanozoa (Cavalier-Smith, 2013a). In Fig. 3 the four chains did not converge, partly because in one *Rozella* was sister to fungi whereas in the other it was within Chytridiomycota as sister to *Piromyces* (and also because of branching differences within chromists). The same lack of convergence with contrasting positions for *Rozella* was evident for the <50% and <40% missing taxa alignments; strikingly when the proportion of missing data was further reduced below 30%, the CAT chains converged (maxdiff 0.233) and *Rozella* was sister to all fungi with maximal support, not within Chytridiomycota, and excluded from fungi with strong support (PP 0.98 Fig. S2), in agreement with the ML tree of James et al. (2013). This suggests that the difficulty of resolving its position is exacerbated by proportionally higher amounts of missing data. James et al. (2013) showed that on ML 200-gene trees *Rozella* is sister to microsporidia, excluded from our trees because of their very long branch, so neither *Rozella* nor microsporidia branch within fungi, but are jointly their sisters. Allomycetes and Chytridiomycetes are robustly sisters here (arguing against separate fungal phyla for them, and for including both in Chytridiomycota), not the alternatives on some previous trees (e.g. Torruella et al. (2012); see Cavalier-Smith (2013a) for discussion).

Previously the position of Filasterea within opisthokonts was uncertain: some multigene trees put them as sister to choanoflagellates plus animals forming a group called filozoa (Shalchian-Tabrizi et al., 2008), others with different taxon/gene sampling (e.g. Brown et al., 2013) put them as sister to Ichthyosporea. Three of our site-heterogeneous trees maximally support the latter grouping (Figs. 2 and 3 for the fewest genes, S1), but the three most gene-rich trees for one taxon sample (the smallest) aberrantly placed Filasterea with animals/choanoflagellates (Fig. 3), but the tree for this taxon sample with proportionally least missing data has Filasterea and Ichthyosporea as sisters (PP 0.99 Fig. S2), as in all 13 ML trees and Brown et al. (2013). As 16 of 19 trees (including all with the richest taxon sampling) show Filasterea/Ichthyosporea as a clade, mostly with strong (usually maximal) support, we name it here:

3.7.1. New opisthokont superclass Filosporidia and probable clade

New choanozoan superclass Filosporidia Cavalier-Smith (under ICZN). *Diagnosis*: trophic cells either naked with long thin filodigits (Cavalier-Smith, 2013a) not organised as a collar or walled cells that usually reproduce by naked amoebae or uniciliate zoospores. Phylogenetically includes all Choanofila except choanoflagellates, i.e. Filasterea, Corallochytrea, and Ichthyosporea.

4. Discussion

4.1. Overall eukaryote phylogeny

A recent synthesis suggested that podiates first evolved pellicle, ciliary gliding, and pseudopodia in their last common ancestor and constitute one of only three eukaryotic supergroups (Cavalier-Smith, 2013a). The other two supergroups (corticates and Eozoa) were ancestrally non-pseudopodial biciliates, though evolved pseudopodia independently in a minority of lineages (Cavalier-Smith, 1997): notably filose and reticulose pseudopodia in the chromist Rhizaria and eruptive lobose pseudopodia in the excavate phylum Percolozoa (Heterolobosea and others: Cavalier-Smith and Nikolaev, 2008), all generally morphologically distinct from podiate pseudopodia. Our Bayesian trees clearly separate the three supergroups and strongly support the monophyly of both podiates and corticates, but ML support is often weak (strongest (80%) for podiates on the 10% fast-site-removal tree).

Adding six Sulcozoa to the tree improves understanding of eukaryote large-scale phylogeny by showing that Varisulca all consistently branch more deeply within podiates than do Apusozoa or Amoebozoa. Irrespective of where one places the eukaryotic root, on all our trees both Sulcozoa and excavates in the classical sense (Simpson and Patterson, 1999) are highly paraphyletic ancestral groups (as O'Kelly who initiated the excavate concept (O'Kelly, 1993) but not the name (Simpson and Patterson, 1999) originally assumed for excavates). Multigene trees previously put apusomonads as sister to opisthokonts (Brown et al., 2013; Derelle and Lang, 2012; Torruella et al., 2012) in accord with the earliest rDNA trees for an apusomonad (Cavalier-Smith and Chao, 1995). Adding more apusomonads here strongly supports this relationship between Apusozoa and opisthokonts and robustly confirms that apusomonads are an ancient genetically diverse clade that probably first radiated even earlier than animals or fungi (Cavalier-Smith and Chao, 2003, 2010). Our CAT trees clearly indicate for the first time that *Malawimonas* and metamonads are probably successively sisters to podiates and that corticates are probably sisters to podiates/*Malawimonas*/metamonads, not to podiates alone.

4.2. Sulcozoan phylogeny and podiate evolution

Our multigene trees, together with those of Brown et al. (2013), provide strong sequence evidence for the deduction from cytoskeletal morphology that the uniciliate breviate amoebae are more closely related to apusomonads than to Amoebozoa (Cavalier-Smith, 2013a). Our Bayesian trees strongly suggest that Apusozoa are paraphyletic and that apusomonads are sisters to opisthokonts, but breviates branch slightly more deeply as sister to opisthokonts plus apusomonads. That is more likely to be correct than the alternative grouping of *Breviata* with apusomonads by ML, as the homogeneous LG model used by RAxML does not fit the evolutionary behaviour of the genes included in our analyses as well as does the CAT heterogeneous model, as Brown et al. (2013) also argued. Our systematic removal of taxonomically less well represented genes reduces the conflict with ML, eventually making it agree with the CAT trees; this strongly supports the assumption that apusozoan holophyly on ML trees is an artefact and suggests that missing data make inference of ancestral states for Apusozoa more difficult for homogeneous models than it does for site-heterogeneous ones (see the excellent discussion of these problems in Roure et al. (2013)), thereby distorting tree topology. Breviates are clearly unrelated to Amoebozoa, contrary to original assumptions (Cavalier-Smith et al., 2004) and early multigene trees using homogeneous models only that included no Sulcozoa (Minge et al., 2009); though more closely related to apusomonads than to Amoebozoa, breviates diverged from apusomonads before the primary radiation of extant apusomonads (our trees represent three of the five apusomonad lineages in the poorly-resolved basal radiation of five deep-branching clades (Cavalier-Smith and Chao, 2010)). The apusomonad/opisthokont affinity of breviates shown here and independently by Brown et al. (2013) agrees with the taxonomic composition of the sulcozoan subphylum Apusozoa as

recently revised by adding Breviatea and excluding Glissodiscea and Rigifilida (both now grouped instead with *Collodictyon* and other diphylleids as the new sulcozoan subphylum Varisulca: Cavalier-Smith, 2013a).

The strong grouping of *Breviata* and *Subulatomonas* found here for the first time is consistent with both sharing a tenuous relict pellicular layer (Cavalier-Smith, 2013a), seemingly absent from *Pygsuia* (Brown et al., 2013), implying that the *Pygsuia* lineage lost the typical sulcozoan pellicular layer after *Pygsuia* diverged from the common ancestor of *Breviata* and *Subulatomonas*; similar loss of a tenuous pellicular layer was postulated for *Sulcomonas* within diphylleids (Cavalier-Smith, 2013a). A thorough study of *Breviata* cytoskeletal ultrastructure (Heiss et al., 2013) strengthens the view that its single cilium is a derived state resulting from loss of the posterior cilium (Cavalier-Smith, 2013a), convergently with such loss in the uniciliate phalansteriids and Archamoebae (both Amoebozoa). In marked contrast, opisthokonts lost the anterior cilium and simplified the ancestral sulcozoan skeleton in a different way through convergently evolving more symmetrical cone-like microtubular arrays. Our trees are in harmony with these ultrastructural interpretations, this congruence reinforcing the idea that podiate ancestors were grooved biciliate cells with an asymmetric excavate-like cytoskeleton (Cavalier-Smith, 2013a).

That means that the 'unikont' condition (having a single centriole) of phalansteriids and Archamoebae is not ancestral for podiates, as was once postulated (Cavalier-Smith, 2000, 2002), and not even ancestral for Amoebozoa, which must now be considered originally biciliate (Cavalier-Smith, 2013a; Heiss et al., 2013). Our clear demonstration that breviates are derived from biciliate Sulcozoa, that on trees both opisthokonts and Amoebozoa are nested within Sulcozoa, and that Sulcozoa comprise three distinct clades, means that the former grouping of opisthokonts plus Amoebozoa alone, originally called unikonts (Cavalier-Smith, 2002) is polyphyletic. Though usage of 'unikonts' has sometimes been extended to include Apusozoa, e.g. Torruella et al. (2012), that name is now inappropriate for such a broader group, being phylogenetically and descriptively misleading. The clade name 'podiates' proposed by Cavalier-Smith (2013a) to embrace the original unikonts (opisthokonts, Amoebozoa; often uniciliate but seldom unicentriolar) plus all Sulcozoa (generally biciliate, sometimes uni-, tetra- or non-ciliate) is a more appropriate substitute for that broadened usage of unikont.

If planomonads were the deepest branching Sulcozoa, as Bayesian trees with proportionally more missing data strongly suggest (Figs. 2, 3 and S1), their ancestor would probably have lacked pseudopodia, and only the dorsal pellicle and ciliary gliding evolved (causing ventral vane loss) when Sulcozoa evolved from a *Malawimonas*-like ancestor, pseudopodia arising slightly later in the common ancestor of all Sulcozoa other than planomonads (Fig. 2). As myosin II (used for podiate pseudopodial motility) has not been found in planomonads (Richards and Cavalier-Smith, 2005) or non-podiates other than the heterolobosean *Naegleria* (which might have got it by lateral gene transfer), myosin II might have originated immediately after other podiates diverged from planomonads at the same time as pseudopodia, and been instrumental in their origin. However, if instead the ML topology of Figs. 2–5 and the CAT tree with proportionally least missing data (Fig. S2) is correct, varisulcan paraphyly on trees with proportionally more missing data is a tree reconstruction artefact, which is not unlikely, and Varisulca are holophyletic. If so, pseudopodia probably evolved in the ancestral podiate, and planomonads (the only Sulcozoa without pseudopods) lost them – as previously postulated (Cavalier-Smith, 2013a); genome sequences are needed for all varisulcan lineages to see if myosin II also originated then. If one judges lateral gene transfer for myosin II to be unlikely (Sebé-Pedrós et al., 2014), then it would probably have arisen in an early neokaryote and have been lost independently by *Giardia* and corticates. In either case, its extensive use by early podiates was probably a significant feature of their pseudopodial function. It would be valuable to have complete genome sequences for all the deepest branching sulcozoan lineages, not only to firm up their phylogeny but also to clarify the origins of the unique podiate pseudopodial cytoskeleton that subsequently played a central role in the development, physiology, and evolution of animals and Amoebozoa.

Within planomonads, the sister grouping of *Nutomonas* and *Ancyromonas* is much stronger on our trees than it was on rDNA trees, where the extremely long branch of *Nutomonas* 18S rDNA reduced bootstrap support for this relationship; we conclude that Ancyromonadidae sensu Glücksman et al. (2013) is certainly holophyletic, as also are Planomonadida and Apusomonadida.

Ultrastructure is unavailable for *Mantamonas*, but should clearly tell us whether mantamonads are more closely related to planomonads (ML trees and the <30% missing gene sample CAT tree) or to diphylleids (most other CAT trees in Figs. 2 and 3 and S1) or a third varisulcan lineage and the most divergent podiate of all (Fig. 3 CAT < 50% missing gene sample). Even without that, our trees show that the strong grouping of *Mantamonas* with apusomonads on 28S rDNA trees (Glücksman et al., 2011) was phylogenetically incorrect; by contrast 18S rDNA trees, weakly grouping *Mantamonas* with planomonads or having no significant resolution for their position relative to other Varisulca (Glücksman et al., 2011, 2013) closely fit our multigene trees, showing that a single-gene tree with low bootstrap support may sometimes be truer than a contradictory one with high support. This may occasionally also be true of multigene trees (Philippe et al., 2011), so the best test of their validity is not levels of statistical 'support' but congruence with independent data and stability to addition of many more taxa, stability to reduction in the proportion of missing data, and use of better phylogenetic algorithms.

The closeness of the branches at the base of podiates revealed by our trees implies a rapid early radiation, which is even more striking at the base of corticates, especially compared with the better-spaced and well-resolved deepest branches following the putatively primary eukaryotic divergence of Euglenozoa and excavates. This explosive bush-like radiation near the base of neozoa makes it difficult to resolve trees even with scores of genes, as does incompleteness in the data; the observed weak resolution in these circumstances is to be expected (Philippe et al., 2011). Though Apusozoa are probably paraphyletic and Varisulca probably holophyletic, further work is needed to confirm this. Better resolution than achieved here will require full genome sequences or near-complete transcriptomes for several representatives of each sulcozoan lineage, including Rigifilida (varisulcan non-flagellate filose amoebae: Yabuki et al., 2013), currently lacking transcriptomic data. Unlike their unsurprising conflict over the likely paraphyly/holophyly of Apusozoa and Varisulca, which we have resolved and partially explained, all our trees unambiguously indicate that Sulcozoa as a whole are paraphyletic: all Apusozoa are more closely related to opisthokonts than they are to Amoebozoa, whereas all Varisulca branch below the bifurcation between Amoebozoa and Apusozoa/opisthokonts. Support for that conclusion varied with taxon sampling, gene sampling, and method, being up to 1.0 with PhyloBayes (Figs. 2 and 3) and up to 88% with ML (Fig. 5).

We conclude that the sulcozoan body plan was most likely ancestral to all other podiate eukaryotes. In other words, dorsal pellicle, ventral pseudopodia, and posterior ciliary gliding all evolved in the podiate common ancestor of Varisulca, Amoebozoa, Apusozoa, and opisthokonts, which already had a ventral groove with a supporting cytoskeleton of two asymmetric posterior microtubular centriolar roots (R2 being split) inherited from a common ancestor of *Malawimonas* and podiates, i.e. a loukozoan excavate. If so, opisthokonts and Amoebozoa must have lost the

ancestral sulcozoan ciliary gliding, ventral groove, and dorsal pellicle independently, as earlier argued (Cavalier-Smith, 2013a). This gives strong evidence for descendants of Sulcozoa having undergone four independent ciliary losses and radical cytoskeletal simplifications to give rise to the more symmetric cone-like skeletons found in opisthokonts and breviate Sulcozoa, both of which retained two centrioles, and in the amoebozoan phalansteriids and Archamoebae that lost the posterior centriole altogether (possibly in a common ancestor, more likely independently if archamoebae are more closely related to the basically biciliate myxogastrid slime moulds than are phalansteriids as some 18S rDNA trees suggest).

A theoretically possible alternative is that Sulcozoa are polyphyletic. But that would mean that dorsal pellicles, posterior ciliary gliding, and ventral pseudopodia each evolved three times independently, this threefold combination all in this part of the eukaryotic tree only and in a relatively short time interval, nine unlikely coincidences and a most unparsimonious hypothesis. The simplest interpretation is that posterior ciliary gliding evolved once only in podiates, in their last common ancestor, and played a key role in their early evolution, as it probably did independently in that of Euglenozoa and the chromist phylum Cercozoa, and perhaps even in the origin of cilia (Cavalier-Smith, 2014); as in those two phyla, ancestral podiate ciliary gliding was probably multiply replaced by swimming as ancestrally surface-associated sulcozoa generated swimming planktonic descendant lineages. Note that Amorphea (Adl et al., 2012) is not a synonym for podiates, as it refers only to the podiate subclade that excludes Varisulca.

4.3. Paraphyly of excavates and the position of Malawimonas

Even the first 143-gene (35,584 positions) ML tree for 48 eukaryotes using the WAG substitution matrix showed Excavata as paraphyletic with *Malawimonas* sister to podiates (no Sulcozoa included) (Hampl et al., 2009); that was argued to be an artefactual result of including long-branch Trichozoa, but as the same topology can be found by both CAT and ML-LG when they are excluded (Fig. 2 and Brown et al., 2013) that argument is no longer defensible. Two other previous multigene papers also provided suggestive evidence that *Malawimonas* may be more closely related to opisthokonts/Amoebozoa (as seen on all three of our most gene-rich Bayesian trees) than to other excavates (Derelle and Lang, 2012; Zhao et al., 2012); one showing it as sister to *Collodictyon* (Zhao et al., 2012), generally assumed to be a tree-reconstruction artefact. That work showed that the position of *Malawimonas* was exceptionally sensitive to taxon and molecular site sampling. Our CAT trees are more taxon-rich and gene-rich and provide further evidence that *Malawimonas* is more closely related to podiates than to other excavates or corticates and suggest that the same may also be true for metamonads.

That is consistent with the argument that supergroup Eozoa is ancestral to neozoa and thus paraphyletic (Cavalier-Smith, 2010a,c, 2013a, 2014). The idea that a *Malawimonas*-like cell was ancestral to Sulcozoa (Cavalier-Smith and Chao, 2010) and other podiates (Cavalier-Smith, 2013a), is now very strongly supported, whether the excavate *Malawimonas* is truly sister to podiates as our Bayesian and a few ML trees indicate or instead really sister to neozoa (or even within podiates) as on other ML trees. Adding six Sulcozoa and many more genes to the tree somewhat stabilised the position of *Malawimonas* compared with earlier taxonomically sparser studies (Hampl et al., 2009; Zhao et al., 2012). Its closeness to the base of the podiate/corticate bifurcation and to Varisulca, plus the incompleteness of the data for both *Malawimonas* and Varisulca in the absence of genome sequences, make branching order in this region still somewhat problematic. Nonetheless, the likelihood that *Malawimonas* is sister to podiates, whereas corticates include *Colponema* [which similarly has a ventral groove with almost indistinguishable microtubular cytoskeleton and a posterior cilium with ventral vane (Cavalier-Smith, 2013a; Mignot and Brugerolle, 1975)], fits the idea that podiates and corticates both evolved from a *Malawimonas*-like loukozoan excavate (Cavalier-Smith, 2013a), but suggests they did so independently and are not directly sisters. If *Malawimonas* really is sister to podiates, and Metamonada sister to this clade as the Fig. 2 CAT tree indicates, even 'core excavates' are paraphyletic.

Contrary to what is sometimes erroneously thought, demonstration of the paraphyly of ancestral groups like Excavata, Sulcozoa, Apusozoa, and Choanozoa does not make them less interesting for evolution or unacceptable as taxa. Both paraphyletic and holophyletic groups arise as a result of the monophyletic origin of a particular character set, in contrast to polyphyletic ones that involve character convergence. The distinction between paraphyly and holophyly is more artificial than that between monophyly (i.e. paraphyly plus holophyly: Ashlock, 1974) and polyphyly, and evolutionarily less important. Clear demonstration of paraphyly reveals ancestral groups, so one can better understand early cell evolution, because it allows one to specify with confidence successive evolutionary steps in morphology, as in the progression in vertebrates through jawless fish, jawed fish, amphibian, reptile to (independently) birds and mammals; thus paraphyletic groups per se are not evolutionarily misleading, but highly informative – what is confusing is to treat paraphyletic groups as if they were holophyletic clades (see Cavalier-Smith, 2010b).

4.4. Importance of increasing gene and taxon representation in multigene trees

We found two instances (paraphyly of Apusozoa and exclusion of *Rozella* from fungi) where removing genes with more missing data increased the accuracy of ML, making it agree more with CAT trees. In a third case, Varisulca, removing genes with more missing data changed both CAT and ML trees in Fig. 3, harmonising their previously conflicting topology and providing evidence for the holophyly of both Glissodiscea and Varisulca and for both CAT and ML trees having been distorted in different ways by excessive proportions of missing data. A fourth example of likely distortion was the apparent paraphyly of the Ichthyosporea/Filasterea clade (Filosporidia) in three out of four CAT trees for one taxon sample only (Fig. 3) indicating that with such distortion a globally better fitting model can sometimes give a less accurate topology for an individual clade than one with globally better fit, possibly because as Roure et al. (2013) suggest errors of an inaccurate model may occasionally reinforce rather than contradict the true topology depending on the historical evolutionary accidents in a particular clade. That particular tree had a smaller taxon sample than the others (Figs. 2, 4 and S3–S5) that showed holophyletic Filosporidia, consistent with richer taxon sampling generally improving trees.

We showed that analyses of alignments with differing proportions of partially missing genes can be helpful in identifying such local exceptions to the reasonable expectation that models with globally better fit to the data should generally yield more accurate trees. Thus our extensive analyses of effects of proportionally reducing missing data, not previously done for protist deep branching, confirm studies with animal data that potentially misleading tree distortions can affect both CAT and ML tree-reconstruction algorithms and support the argument of Roure et al. (2013) that trees with fewer but taxonomically more completely represented genes can sometimes be more accurate than those with many more genes if the latter suffer from unequal taxonomic representation of too high a fraction of the included genes.

By contrast to these examples, the stem at the base of the clade comprising opisthokonts, Apusozoa, and Amoebozoa is consistently well supported by both ML and CAT trees, showing the robustness of all Varisulca being the deepest branching of all podiates. This difference is perhaps attributable to whole genome sequences being available for many opisthokonts and Amoebozoa and one apusozoan, whereas none is for Varisulca. The basal branching of Sulcozoa and therefore podiates as a whole will probably only be more robustly resolved when at least one complete genome or near-complete transcriptome is available for each of the three varisulcan lineages as well as for a breviate. The same problem applies to the still more uncertain position of *Mantamonas*. The stem at the base of the clade containing it and *Collodictyon* on Fig. 4 is just as short. Indeed this grouping is supported (weakly) only in the full alignment and for CAT in the <40% alignment (Fig. 3) and disappears for both ML and CAT with proportionally less missing data (Fig. 3); though ML always had a glissodiscean clade, the CAT tree with <50% missing data placed *Mantamonas* on its own as sister neither to planomonads nor to *Collodictyon*, but as sister to *Collodictyon* plus Amoebozoa, Apusozoa, and opisthokonts (low 0.51 support). The topology of the holophyletic Varisulca in the <30% missing gene tree (Fig. 2) ought technically to be the most reliable, and is in harmony with ML trees that exclude *Malawimonas*, and with all the cell evolutionary arguments underlying the establishment of that subphylum (Cavalier-Smith, 2013a). Though that makes it overall the most likely topology, more nearly complete transcriptomes for taxa branching in this region are essential to test it more thoroughly. That diphylleids and *Mantamonas* are represented by only single species, unlike all other sulcozoan clades, makes accurate tree construction even more difficult, so it will be desirable to break up these long branches by adding other members of both groups (and a rigifilid, which rDNA suggests is related to diphylleids: Yabuki et al., 2013).

5. Conclusions

We conclude that Varisulca (diphylleids plus Glissodiscea) are the most divergent podiates, and probably holophyletic, with important implications for the origin and earliest evolution of podiates. In harmony with 18S rDNA trees, we conclusively show, contrary to 28S rDNA trees (Glücksman et al., 2011), that *Mantamonas* is more closely related to planomonads and diphylleids than to apusomonads. We show for the first time that within breviates *Subulatomonas* is more closely related to *Breviata* than to *Pygsuia* and that the apusomonad genus *Thecamonas* as currently constituted is paraphyletic or polyphyletic. Our analyses clarify the reasons for and largely resolve previous contradictions between site-homogeneous ML and site-heterogeneous Bayesian trees with respect to the holophyly or paraphyly of Apusozoa (and similar contradictions seen here for Varisulca as well as two previously inconsistent branching patterns within opisthokonts) and confirm the conclusion that Apusozoa are paraphyletic (Brown et al., 2013); as breviates and apusomonads are probably not sisters, their shared apusozoan phenotype is ancestral to that of opisthokonts.

Our trees are consistent with the subdivision of Sulcozoa into two subphyla, confirm that breviates belong in Apusozoa (not Amoebozoa) and that Apusozoa are sister to opisthokonts, and show that Sulcozoa are paraphyletic and thus ancestral to all other podiate eukaryotes, with subphylum Varisulca being sister to opisthokonts plus Apusozoa and Amoebozoa. That means that the simple microtubular skeletons of animal and other opisthokont cells arose by radically simplifying a much more complex cell body plan that first evolved in association with a ventral feeding groove in the ancestor of biciliate excavates, from which all eukaryotes other than Euglenozoa are argued to have descended (Cavalier-Smith, 2010a, 2013a). Overall, when initial conflicts between CAT and ML trees are resolved by reducing the proportion of missing data, our trees provide substantial support for the thesis that the sulcozoan dorsal pellicle, ventral pseudopodia, and posterior ciliary gliding all evolved simultaneously and coadaptively during the origin of ancestral podiates from a swimming, non-gliding, non-pseudopodial *Malawimonas*-like excavate (Cavalier-Smith, 2010a, 2013a). Moreover, CAT trees place *Malawimonas* and anaeromonad metamonads as sisters to podiates as in Brown et al. (2013), further strengthening that idea. Our analyses clarify hitherto contradictory interpretations of the phylogenetic position of malawimonads and make it likely that Loukozoa and excavates as a whole are paraphyletic.

Acknowledgments

This work was supported by the Natural Environment Research Council (Grant No. NE/E004156/1) and by the Leverhulme Trust (Grant No. R1008101). We thank Research Councils UK (15842) for open access funding.

RL thanks NERC for a research studentship. We thank Oxford University Advanced Research Computing (ARC) for access to their clusters and efficient advice, Matthew Brown, Fabien Burki, Jessica Grant, Vladimir Hampl, Timothy James, Naira Rodríguez-Espeleta, Ken-Ichiro Ishida (for *Paulinella*), and Kamran Shalchian-Tabrizi for protein alignments or sequences.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.08.012.

References

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., et al., 2012. The revised classification of eukaryotes. J. Eukaryot. Microbiol. 59, 429–493.

Ashlock, P., 1974. The use of cladistics. Annu. Rev. Ecol. Syst. 5, 81–99.

Brown, M.W., Sharpe, S.C., Silberman, J.D., Heiss, A.A., Lang, F.B., Simpson, A.G.B., Roger, A.J., 2013. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. Proc. R. Soc. B 280, 20131755.

Brugerolle, G., Bricheux, G., Philippe, H., Coffea, G., 2002. *Collodictyon triciliatum* and *Diphylleia rotans* (=*Aulacomonas submarina*) form a new family of flagellates (Collodictyonidae) with tubular mitochondrial cristae that is phylogenetically distant from other flagellate groups. Protist 153, 59–70.

Burger, G., Gray, M.W., Forget, L., Lang, B.F., 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol. Evol. 5, 418–438.

Burki, F., Inagaki, Y., Bråte, J., Archibald, J.M., Keeling, P.J., Cavalier-Smith, T., Sakaguchi, M., Hashimoto, T., Horak, A., Kumar, S., Klaveness, D., Jakobsen, K.S., Pawlowski, J., Shalchian-Tabrizi, K., 2009. Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, Telonemia and Centroheliozoa, are related to photosynthetic chromalveolates. Genome Biol. Evol. 1, 231–238.

Burki, F., Okamoto, N., Pombert, J.-F., Keeling, P.J., 2012. The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc. R. Soc. B, 2246–2254.

Burki, F., Corradi, N., Sierra, R., Pawlowski, J., Meyer, G.R., Abbott, C.L., Keeling, P.J., 2013. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in Rhizaria. Curr. Biol. 23, 1541–1547.

Cavalier-Smith, T., 1997. Amoeboflagellates and mitochondrial cristae in eukaryote evolution: megasystematics of the new protozoan subkingdoms Eozoa and Neozoa. Arch. Protistenkd. 147, 237–258.

Cavalier-Smith, T., 2000. Flagellate megaevolution: the basis for eukaryote diversification. In: Green, J., Leadbeater, B. (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., 2003. The excavate protozoan phyla Metamonada Grassé emend. (Anaeromonadea, Parabasalia, *Carpediemonas*, Eopharyngia) and Loukozoa emend. (Jakobea, *Malawimonas*): their evolutionary affinities and new higher taxa. Int. J. Syst. Evol. Microbiol. 53, 1741–1758.

Cavalier-Smith, T., 2009. Megaphylogeny, cell body plans, adaptive zones: causes and timing of eukaryote basal radiations. J. Eukaryot. Microbiol. 56, 26–33.

Cavalier-Smith, T., 2010a. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. Biol. Lett. 6, 342–345.

Cavalier-Smith, T., 2010b. Deep phylogeny, ancestral groups, and the four ages of life. Philos. Trans. Roy. Soc. B 365, 111–132.
Cavalier-Smith, T., 2010c. The origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. Biol. Direct 5, 7.
Cavalier-Smith, T., 2013a. Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa. Eur. J. Protistol. 49, 115–178.
Cavalier-Smith, T., 2013b. Symbiogenesis: mechanisms, evolutionary consequences, and systematic implications. Annu. Rev. Syst. Ecol. Evol. 44, 145–172.
Cavalier-Smith, T., 2014. The neomuran revolution and phagotrophic origin of eukaryotes and cilia in the light of intracellular coevolution and a revised tree of life. In: Keeling, P.J., Koonin, E.V. (Eds.), The Origin and Evolution of Eukaryotes. Cold Spring Harbor Perspectives Biology, pp. 41–77. doi: http://dx.doi.org/10.1101/cshperspect.a016006.
Cavalier-Smith, T., Chao, E.E., 1995. The opalozoan *Apusomonas* is related to the common ancestor of animals, fungi, and choanoflagellates. Proc. R. Soc. B 261, 1–6.
Cavalier-Smith, T., Chao, E.E., 2003. Phylogeny of Choanozoa, Apusozoa, and other Protozoa and early eukaryote megaevolution. J. Mol. Evol. 56, 540–563.
Cavalier-Smith, T., Chao, E.E., 2010. Phylogeny and evolution of Apusomonadida (Protozoa: Apusozoa): new genera and species. Protist 161, 549–576.
Cavalier-Smith, T., Nikolaev, S., 2008. The zooflagellates *Stephanopogon* and *Percolomonas* are a clade (Percolatea: Percolozoa). J. Eukaryot. Microbiol. 55, 501–509.
Cavalier-Smith, T., Chao, E.E., Oates, B., 2004. Molecular phylogeny of Amoebozoa and the evolutionary significance of the unikont *Phalansterium*. Eur. J. Protistol. 40, 21–48.
Cavalier-Smith, T., Chao, E.E., Stechmann, A., Oates, B., Nikolaev, S.I., 2008. Planomonadida ord. nov. (Apusozoa): ultrastructural affinity with *Micronuclearia podoventralis* and deep divergences within *Planomonas* gen. nov. Protist 159, 535–562.
Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.
Derelle, R., Lang, F.B., 2012. Rooting the eukaryotic tree with mitochondrial and bacterial proteins. Mol. Biol. Evol. 29, 1277–1289.
Deschamps, P., Moreira, D., 2009. Signal conflicts in the phylogeny of the primary photosynthetic eukaryotes. Mol. Biol. Evol. 261, 2745–2753.
Glücksman, E., Snell, E.A., Berney, C., Chao, E.E., Bass, D., Cavalier-Smith, T., 2011. The novel marine gliding zooflagellate genus *Mantamonas* (Mantamonadida ord. n.: Apusozoa). Protist 162, 207–221.
Glücksman, E., Snell, E.A., Cavalier-Smith, T., 2013. Phylogeny and evolution of Planomonadida (Sulcozoa): eight new species and new genera *Fabomonas* and *Nutomonas*. Eur. J. Protistol. 49, 179–200.
Grant, J., Lahr, D.J.G., Rey, F.E., Gordon Burleigh, J., Gordon, J.I., Knight, R., Molestina, R.E., Katz, L.A., 2012. Gene discovery from a pilot study of the transcriptomes from three diverse microbial eukaryotes: *Corallomyxa tenera*, *Chilodonella uncinata*, and *Subulatomonas tetraspora*. Prot. Genomics, 3–18.
Hampl, V., Hug, L., Leigh, J.W., Dacks, J.B., Lang, F.B., Simpson, A.G.B., Roger, A.J., 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". Proc. Natl. Acad. Sci. USA 106, 3859–3864.
Heiss, A.A., Walker, G., Simpson, A.G.B., 2013. The flagellar apparatus of *Breviata anathema*, a eukaryote without a clear supergroup affinity. Eur. J. Protistol. 49, 354–372.
James, T.Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., Stajich, J.E., 2013. Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. Curr. Biol. 23, 1548–1553.
Katz, L.A., Grant, J., Parfrey, L.W., Gant, A., O'Kelly, C.J., Anderson, R.O., Molestina, R.E., Nerad, T.A., 2011. *Subulatomonas tetraspora* nov. gen. nov. sp. is a member of a previously unrecognized major clade of eukaryotes. Protist 162, 762–773.
Kostka, M., Uzlikova, M., Cepicka, I., Flegr, J., 2008. SlowFaster, a user-friendly program for slow-fast analysis and its application on phylogeny of *Blastocystis*. BMC Bioinformatics 9, 341.
Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., Gray, M.W., 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387, 493–497.
Lartillot, N., Philippe, H., 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. Philos. Trans. R. Soc. B 363, 1463–1472.
Lasek-Nesselquist, E., Gogarten, J.P., 2013. The effects of model choice and mitigating bias on the ribosomal tree of life. Mol. Phylogenet. Evol. 69, 17–38.
Mignot, J.-P., Brugerolle, G., 1975. Étude ultrastructurale du flagellé phagotrophe *Colponema loxodes* Stein. Protistologica 11, 429–444.
Minge, M.A., Silberman, J.D., Orr, R.J., Cavalier-Smith, T., Shalchian-Tabrizi, K., Burki, F., Skjæveland, Å., Jakobsen, K.S., 2009. Evolutionary position of breviate amoebae and the primary eukaryote divergence. Proc. R. Soc. B 276, 597–604.
O'Kelly, C., 1993. The jakobid flagellates: structural features of *Jakoba*, *Reclinomonas* and *Histiona* and implications for the early diversification of eukaryotes. J. Eukaryot. Microbiol. 40, 627–636.
Paps, J., Medina-Chacón, L.A., Marshall, W., Suga, H., Ruiz-Trillo, I., 2013. Molecular phylogeny of unikonts: new insights into the position of apusomonads and ancyromonads and the internal relationships of opisthokonts. Protist 164, 2–12.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602.
Richards, T.A., Cavalier-Smith, T., 2005. Myosin domain evolution and the primary divergence of eukaryotes. Nature 436, 1113–1118.
Roure, B., Rodríguez-Ezpeleta, N., Philippe, H., 2007. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol. Biol. 7, S2.
Roure, B., Baurain, D., Philippe, H., 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol. Biol. Evol., 197–214.
Sebé-Pedrós, A., Grau-Bové, X., Richards, T.A., Ruiz-Trillo, I., 2014. Evolution and classification of myosins, a paneukaryotic whole genome approach. Genome Biol. Evol. 6, 290–305.
Shalchian-Tabrizi, K., Minge, M.A., Espelund, M., Orr, R.J., Ruden, T., Jakobsen, K.S., Cavalier-Smith, T., 2008. Multigene phylogeny of Choanozoa and the origin of animals. PLoS One 3, e2098.
Simpson, A.G.B., 2003. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int. J. Syst. Evol. Microbiol. 53, 1759–1777.
Simpson, A.G.B., Patterson, D.J., 1999. The ultrastructure of *Carpediemonas membranifera* (Eukaryota) with reference to the "excavate hypothesis". Eur. J. Protistol. 35, 353–370.
Torruella, G., Derelle, R., Paps, J., Lang, F.B., Roger, A.J., Shalchian-Tabrizi, K., Ruiz-Trillo, I., 2012. Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. Mol. Biol. Evol. 29, 531–544.
Yabuki, A., Ishida, K., Cavalier-Smith, T., 2013. *Rigifila ramosa* n. gen., n. sp., a filose apusozoan with a distinctive pellicle, is related to *Micronuclearia*. Protist 164, 75–88.
Zhao, S., Burki, F., Bråte, J., Keeling, P.J., Klaveness, D., Shalchian-Tabrizi, K., 2012. *Collodictyon* – an ancient lineage in the tree of eukaryotes. Mol. Biol. Evol. 29, 1557–1568.
Zhao, S., Shalchian-Tabrizi, K., Klaveness, D., 2013. Sulcozoa revealed as a paraphyletic group in mitochondrial phylogenomics. Mol. Phylogenet. Evol. 69, 462–468.