Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group.

Matthew W. Brown[1,2], Aaron Heiss[3,4], Ryoma Kamikawa[5], Yuji Inagaki[6,7], Akinori Yabuki[8], Alexander K Tice[1,2], Takashi Shiratori[6], Ken-ichiro Ishida[6], Tetsuo Hashimoto[6,7], Alastair G. B. Simpson[3]*, and Andrew J. Roger[9]*

1. Department of Biological Sciences, Mississippi State University USA
2. Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, USA
3. Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biology Dalhousie University, Halifax, N.S., B3H 4R2, Canada.
4. Department of Invertebrate Zoology and Sackler Institute for Comparative Genomics, American Museum of Natural History, Central Park West at 79th St., New York, NY, 10024, USA
5. Graduate School of Human and Environmental Studies, Graduate School of Global Environmental Studies, Kyoto University, Japan
6. Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan
7. Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan
8. Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15, Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan.

9. Centre for Comparative Genomics and Evolutionary Bioinformatics, Dept. of Biochemistry and Molecular Biology, Dalhousie University, Halifax, N.S., B3H 4R2, Canada

* **Corresponding Authors:** E-mail: Andrew.Roger@dal.ca and Alastair.Simpson@dal.ca

Abstract

Recent phylogenetic analyses position certain 'orphan' protist lineages deep in the tree of eukaryotic life, but their exact placements are poorly resolved. We conducted phylogenomic analyses that incorporate deeply sequenced transcriptomes from representatives of collodictyonids (diphylleids), rigifilids, *Mantamonas* and ancyromonads (planomonads). Analyses of 351 genes, using site-heterogeneous mixture models, strongly support a novel supergroup-level clade that includes collodictyonids, rigifilids and *Mantamonas*, which we name 'CRuMs'. Further, they robustly place CRuMs as the closest branch to Amorphea (including animals and fungi). Ancyromonads are strongly inferred to be more distantly related to Amorphea than are CRuMs. They emerge either as sister to malawimonads, or as a separate deeper branch. CRuMs and ancyromonads represent two distinct major groups that branch deeply on the lineage that includes animals, near the most commonly inferred root of the eukaryote tree. This makes both groups crucial in examinations of the deepest-level history of extant eukaryotes.

Introduction

Our understanding of the eukaryote tree of life has been revolutionized by genomic and transcriptomic investigations of diverse protists, which constitute the overwhelming majority of

47 eukaryotic diversity (Burki 2014; Simpson and Eglt 2016). Phylogenetic analyses of super-
48 matrices of proteins typically show a eukaryote tree consisting of five-to-eight 'super-groups'
49 that fall within three even-higher-order assemblages: i) Amorphea (Amoebozoa plus Obazoa, the
50 latter including animals and fungi), ii) Diaphoretickes (primarily Sar, Archaeplastida, Cryptista
51 and Haptophyta), and iii) Excavata (Discoba and Metamonada) (Adl et al. 2012). Recent
52 analyses (Derelle et al. 2015) place the root of the eukaryote tree somewhere between Amorphea
53 and the other two listed lineages; Derelle *et al*. (2015) termed this the 'Opimoda-Diphoda' root.
54 There is considerable debate over the position of the root, however (e.g., (Cavalier-Smith 2010;
55 Katz et al. 2012; He et al. 2014))

56 Nonetheless, there remain several 'orphan' protist lineages that cannot be assigned to any
57 super-group by cellular anatomy or ribosomal RNA phylogenies (e.g., (Brugerolle et al. 2002;
58 Glücksman et al. 2011; Heiss et al. 2011; Cavalier-Smith 2013; Yabuki, Eikrem, et al. 2013;
59 Yabuki, Ishida, et al. 2013)). Recent phylogenomic analyses including *Collodictyon*,
60 *Mantamonas,* and ancyromonads indicate that these particular 'orphans' branch near the base of
61 Amorphea (Zhao et al. 2012; Cavalier-Smith et al. 2014), the same general position as the
62 purported Opimoda-Diphoda root. This implies, i) that these lineages are of special evolutionary
63 importance, but also, ii) that uncertainty over their phylogenetic positions will profoundly impact
64 our understanding of deep eukaryote history. Unfortunately their phylogenetic positions indeed
65 remain unclear, with different phylogenomic analyses supporting incompatible topologies, and
66 often showing low statistical support (Cavalier-Smith et al. 2014). This is likely due in part to the
67 modest numbers of sampled genes for some/most species examined to date (Cavalier-Smith et al.
68 2014; Torruella et al. 2015). Therefore, we undertook phylogenomic analyses that incorporated

deeply-sequenced transcriptome data from representatives of Collodictyonidae, *Mantamonas*, Ancyromonadida, and Rigifilida.

Results

Using a custom phylogenomic pipeline plus manual curation we generated a dataset of 351 orthologs. The dataset was filtered of paralogs and potential cross-contamination by visualizing each protein's phylogeny individually, then removing sequences whose positions conflicted with a conservative consensus phylogeny (as in (Tice et al. 2016; Kang et al. 2017)) (supplementary methods). We selected data-rich species to represent the phylogenetic diversity of eukaryotes. Our primary dataset retained 61 taxa, with metamonads represented by two short-branching taxa (*Trimastix* and *Paratrimastix*). We also analyzed a 64-taxon dataset containing three additional longer-branching metamonads. Maximum likelihood (ML) and Bayesian analyses were conducted using site-heterogeneous models; LG+C60+F+Γ and the associated PMSF model (LG+C60+F+Γ-PMSF) as implemented in IQ-TREE (Wang et al. 2017) and CAT-GTR+Γ in PHYLOBAYES-MPI, respectively. Such site-heterogeneous models are important for deep-level phylogenetic inference with numerous substitutions along branches (Lartillot et al. 2007; Le et al. 2008; Wang et al. 2008; Pisani et al. 2015; Wang et al. 2017).

Our analyses of both 61- and 64-taxon datasets robustly recover well-accepted major groups including Sar, Discoba, Metamonada, Obazoa, and Amoebozoa (Fig. 1, S1). Cryptista (e.g. cryptomonads and close relatives) branches with Haptophyta (Fig. 1) in the LG+C60+F+Γ-PSMF analyses as well as in one set of two converged PHYLOBAYES-MPI chains under the CAT-GTR model (Fig. S2). However another pair of converged chains places Haptophyta as sister to Sar while Cryptista nests within Archaeplastida (Fig. S3), which is largely consistent with some

other recent phylogenomic studies (e.g., (Burki et al. 2016)). Excavata was never monophyletic, with Discoba forming a clan with Diaphoretickes taxa (Sar, Haptophyta, Archaeplastida+Cryptista) and Metamonada grouping with Amorphea plus the four orphan lineages targeted in this study (see below). Malawimonads, which are morphologically similar to certain metamonads and discobids (Simpson 2003), also branch amongst the 'orphans' (see below).

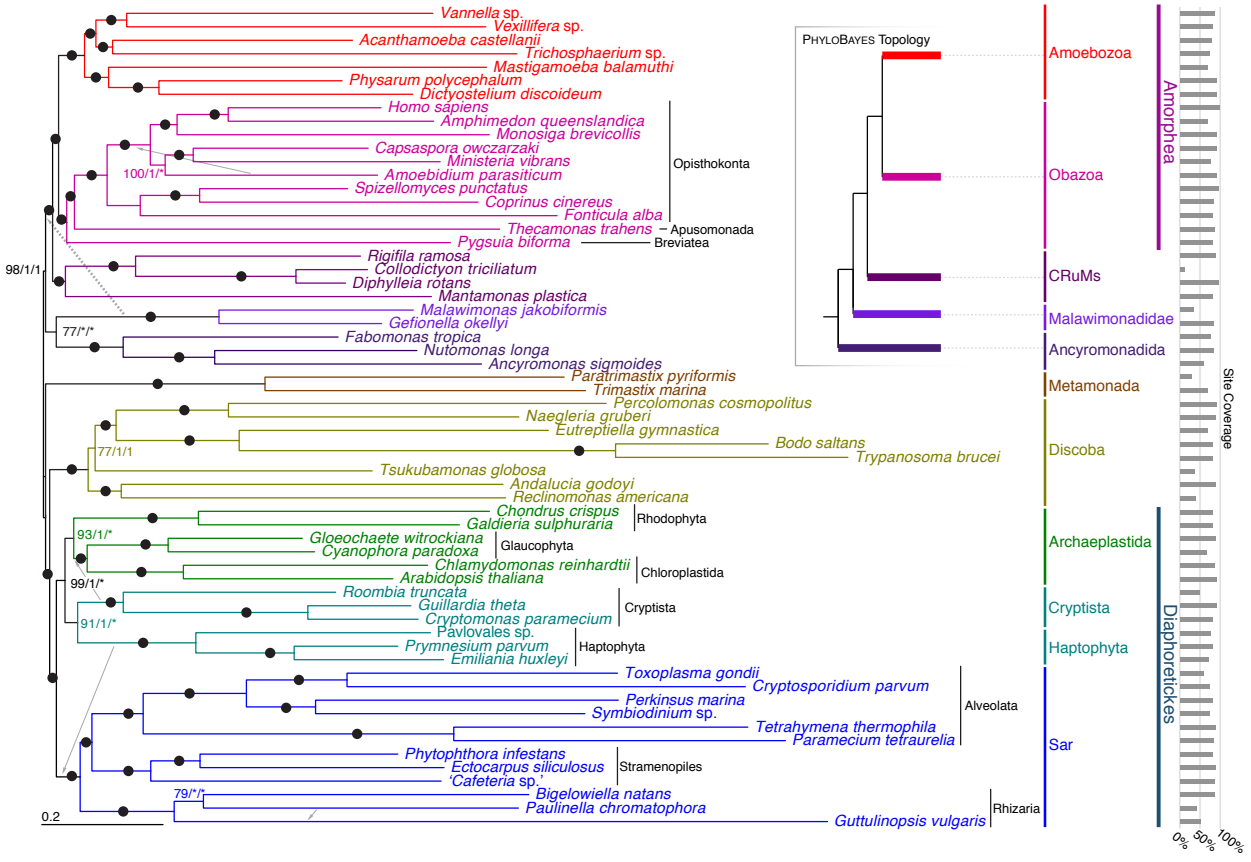


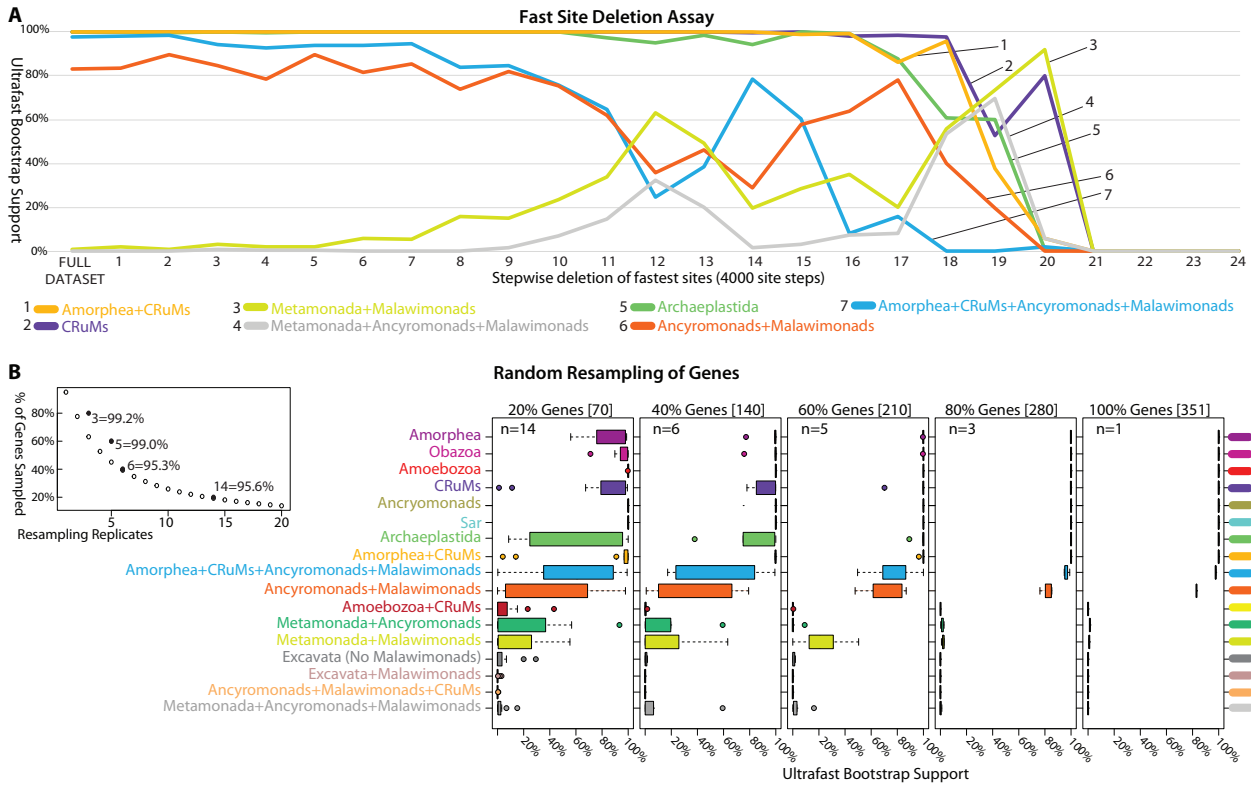
Figure 1. Phylogenetic tree for 61 eukaryotes, inferred from 351 proteins using Maximum Likelihood (LG+C60+F+Γ-PMSF model). The numbers on branches show (in order) support values from 100 real bootstrap replicates (LG+C60+F+Γ-PMSF model) and posterior probabilities both sets of converged chains in PHYLOBAYES-MPI under CAT-GTR+Γmodel (i.e., MLBS/PP/PP). Filled circles represent maximum support with all methods; asterisks indicate a clade not recovered in the PHYLOBAYES analysis. Further, the dashed arrow indicates the placement of malawimonads inferred with PHYLOBAYES-MPI (see also inset summary tree), and grey arrows indicate the placements of other lineages in the PHYLOBAYES-MPI analyses.

Phylogenies of both datasets place all four orphan taxa near the base of Amorphea (Fig. 1, Fig. S1). The uncertain position of the eukaryotic root (discussed above) therefore makes it unclear which bipartitions are truly clades, and which could be interrupted by the root. To allow efficient communication, we discuss the phylogenies as if the orphan taxa all lie on the Amorphea side of the root. We will also consider Amorphea as previously circumscribed (Adl et al. 2012): the least-inclusive clade or clan containing Amoebozoa and Obazoa.

Three of the orphan lineages are specifically related in our trees (Fig. 1, S1). In both 61-taxon and 64-taxon analyses, *Rigifila ramosa* (representing Rigifilida) forms a maximally-supported clade with the collodictyonids *Collodictyon triciliatum* and *Diphylleia rotans*. *Mantamonas plastica* then branches as their closest relative, with maximal support. This Collodictyonid+Rigifilida+*Mantamonas* clade ('CRuMs') forms the sister group to Amorphea, again with maximal support.

ML analyses and the converged PHYLOBAYES chains grouped ancyromonads, malawimonads, and CRuMs with Amorphea, with strong bootstrap support and Bayesian posterior probability (Fig. 1, 61 taxa; PMSF BS=98%, PP=1). Ancyromonads and malawimonads formed a clade in the ML analyses, but with equivocal support (Fig. 1, 61 taxa; BS=77%). Both sets of converged chains of the Bayesian analyses instead grouped malawimonads with CRuMs+Amorphea to the exclusion of ancyromonads (Fig. S2, S3, PP=1 for both); however some unconverged chains support an ancyromonad+malawimonad clade (data not shown). Lack of convergence amongst multiple chains using the CAT-GTR+Γ model is unfortunately common for large datasets, and often cannot be resolved by increasing the number of generations of Markov chain Monte Carlo within a reasonable time frame (Pisani et al. 2015;

121 Kang et al. 2017). Instead we treat the two topologies recovered in these analyses as candidate
122 hypotheses requiring further investigation.
123
124 We conducted approximately unbiased (AU) topology tests on the 61-taxon data set
125 under the LG+C60+F+Γ mixture model (Table S1). These tests rejected the Phylobayes trees, as
126 well as all trees optimized by enforcing constraints representing plausible alternative relative
127 placements of ancyromonads, malawimonads, and metamonads.
128 The fastest evolving sites are expected to be the most prone to saturation and systematic
129 error arising from model misspecification in phylogenomic analyses (Philippe et al. 2011). We
130 conducted a 'fast-site removal' analysis with the 61-taxon data set and generated ultrafast
131 bootstrap support (UFBOOT) values (Minh et al. 2013) for relevant groups as sites were
132 progressively removed from fastest to slowest (Fig. 2A). All groups of interest receive
133 reasonably strong support until ~44,000-48,000 sites were removed, when support fell markedly
134 for the ancryomonad+malawimonad clade and the
135 Amorphea+CRuMs+ancryomonad+*Malawimonas* clan. At this point, a notable proportion of the
136 bootstrap trees show malawimonads and/or ancyromonads grouping with metamonads. This
137 decline in support for the ancryomonad+malawimonad group reverses somewhat with further
138 site removal, before support falls again as overall phylogenetic structure is lost when ~76,000
139 sites are removed (Fig. 2A).



140

Figure 2. Effects of fast evolving sites and random subsampling of genes on our phylogenomic analyses. (A) Sites were sorted based on their rates of evolution estimated under the LG+F+Γ model and removed from the data set from highest to lowest rate. Each step has 4,000 of the fastest evolving sites removed progressively. The bootstrap values (UFBOOT; LG+C60+F+Γ-PSMF model) for each bipartition of interest are plotted. (B, C) Effects of random subsampling of genes within the 351-gene dataset. The following bipartitions were examined but received nearly 100% support accross the fast site deletion assay (data not shown); Amorphea, Obazoa, Amoebozoa, Ancryomonads, and Sar. The following bipartitions were examined but received nearly 0% support across the fast site deletion assay (data not shown); Amoebozoa+CRuMs, Metamonada+Ancyromonads, Excavata (No Malawimonads), Excavata+Malawimonads, and Ancyromonads+Malawimonads+CRuMs. (B) Effects of random resampling of genes on the bipartitions of interest. Inset panel is the calculation of the number of replicates (n) necessary for a 95% probability of sampling every gene when subsampling 20, 40, 60 and 80% of genes using the formula: $0.95 = 1 - (1 - x/100)^n$ where x is the percentage of genes subsampled. UFBOOT support values for all nodes of interest with the variability of support values illustrated by box-and-whisker plots.

141

142 To evaluate heterogeneity in phylogenetic signals amongst genes (Inagaki et al. 2009),
143 we also inferred phylogenies from subsamples of the 351 examined genes (61-taxon dataset; Fig.
144 2B,C). For each subsample 20-80% of the genes were randomly selected, without replacement,
145 with replication as per Fig. 2B (giving a >95% probability that a particular gene would be
146 sampled at each level), and UFBOOT support for major clades was inferred (Fig. 2C). The '80%
147 retained' replicates gave nearly identical results to the full dataset, indicating that there was little
148 stochastic error associated with gene sampling at this level. Support for the CRuMs clade is
149 almost always high when 40%+ of genes are retained, while subsamples containing 60% of
150 genes still showed differing support for a ancyromonad-malawimonad clade (as opposed to, for
151 example, malawimonads branching with metamonads).

152

153 **Discussion**

154 Our 351 protein (97,002 AA site) super-matrix places several orphan lineages in two separate
155 clades emerging between Amorphea and all other major eukaryote groups. All methods recover a
156 strongly supported clade comprising the free-swimming collodictyonid flagellates, the
157 idiosyncratic filose protist *Rigifila* (Rigifilida) and the gliding flagellate *Mantamonas*. This clade
158 is resilient to exclusion both of fast-evolving sites and of randomly selected genes. It is also
159 consistently placed as the immediate sister taxon to Amorphea. This represents the first robust
160 estimate of the positions of these three taxonomically poor but phylogenetically deep clades.
161 Previous phylogenomic analyses placed collodictyonids in various positions, such as sister to
162 either malawimonads or Amoebozoa, but often with low statistical support (Zhao et al. 2012;
163 Cavalier-Smith et al. 2014). Placements of *Mantamonas* have varied dramatically. A recent

phylogenomic study recovered a weak *Mantamonas*+collodictyonid clade in some analyses, but other analyses in the same study instead recovered a weak *Mantamonas*+ancyromonad relationship (Cavalier-Smith et al. 2014), and SSU+LSU rRNA gene phylogenies strongly grouped *Mantamonas* with apusomonads (Glücksman et al. 2011; Yabuki, Ishida, et al. 2013). Our study decisively supports the first of these possibilities. This is the first phylogenomic analysis incorporating Rigifilida: Previous SSU+LSU rRNA gene analyses recovered a negligibly supported collodictyonid+rigifilid clade, but not a relationship with *Mantamonas* (Yabuki, Ishida, et al. 2013).

Overall, the hypotheses that i) collodictyonids, rigifilids and *Mantamonas* form a major eukaryote clade, and ii) this clade is sister to Amorphea, are novel, plausible, and evolutionarily important. No name exists for this putative super-group, and it is obviously premature to propose a formal taxon. We suggest the place-holding moniker 'CRuMs' (Collodictyonidae, Rigifilida, *Mantamonas*), which is euphonic and evokes the species-poor nature of these taxa.

Whether ancyromonads branch outside Amorphea or within it has been disputed (Paps et al. 2013; Cavalier-Smith et al. 2014). Our study strongly places ancyromonads outside Amorphea, more distantly related to it than are the CRuMs. Ancyromonads instead fall 'amongst' the excavate lineages (Discoba, Metamonada, and Malawimonadidae). Resolving the relationships amongst 'excavates' is extremely challenging (Hampl et al. 2009; Derelle et al. 2015), and this likely contributed to our difficulty in resolving the exact position of ancyromonads vis-à-vis malawimonads. A close relationship between ancyromonads and some/all excavates would be broadly consonant with the marked cytoskeletal similarity between *Ancyromonas* and 'typical excavates' (Heiss et al. 2011). Certainly, our study flags ancyromonads as highly relevant to resolving relationships amongst excavates.

187 Both candidate positions for ancyromonads place them at the centre of a crucial open
188 question: locating the root of the eukaryote tree. As discussed above, the latest analyses (Derelle
189 et al. 2015) locate the root between Discoba+Diaphoretickes ('Diphoda') and a clade including
190 Amorphea, collodictyonids and malawimonads ('Opimoda'). Our phylogenies show the
191 ancyromonad lineage emerging close to this split. One of the two positions we recovered would
192 actually place ancyromonads either as the deepest branch within 'Diphoda', or the deepest
193 branch within 'Opimoda', or even as sister to all other extant eukaryotes. This demonstrates the
194 profound importance of including ancyromonads in future rooted phylogenies of eukaryotes,
195 using datasets optimized for this purpose.
196
197 **Materials and Methods**
198 Details of experimental methods for culturing, nucleic acid extraction and Illumina sequencing
199 are described in the supplemental text.
200
201 *Phylogenomic data set construction*.
202 A reference data set of 351 aligned proteins described in (Kang et al. 2017) was used as the
203 starting point for the current analysis, from which 61 or 64 taxa representing diverse eukaryotes
204 were selected (see Table S2). Extensive efforts were made to exclude contamination and
205 paralogs, as described in the supplemental text. Poorly aligned sites were excluded using BMGE
206 (Criscuolo and Gribaldo 2010), resulting in an alignment of 97,002 amino acid (AA) sites with
207 less than 25% missing data for both 61- and 64-taxon datasets (Table S2).
208
209 *Phylogenomic tree inference*

Maximum likelihood (ML) trees were inferred using IQ-TREE v. 1.5.5 (Nguyen et al. 2015). The best-fitting available model based on the Akaike Information Criterion (AIC) was the LG+C60+F+Γ mixture model with class weights optimized from the dataset and four discrete gamma (Γ) categories. ML trees were estimated under this model for both 61- and 64-taxon data sets. We then used this model and best ML tree under the LG+C60+F+Γ model to estimate the 'posterior mean site frequencies' (PMSF) model (Wang et al. 2017) for both 61 (Fig. 1) and 64 (Fig. S1) taxon data sets. This LG+C60+F+Γ-PMSF model was used to re-estimate ML trees, and for a bootstrap analysis of the 61-taxon dataset, with 100 pseudoreplicates (Fig. 1). AU topology tests under the LG+C60+F+Γ were conducted with IQ-TREE to evaluate whether trees recovered by the Bayesian analyses or alternative placements of the orphan taxa could be rejected statistically.

Bayesian inferences were performed using PHYLOBAYES-MPI v1.6j (Rodrigue and Lartillot 2014), under the CAT-GTR+Γ model, with four discrete Γ categories. For the 61-taxon analysis, 6 independent Markov chain Monte Carlo chains were run for ~4,000 generations, sampling every second generation. Two sets of two chains converged (at 800 and 2,000 generations, which were respectively used as the burnin), with the largest discrepancy in posterior probabilities (PPs) (maxdiff) < 0.05. The topologies of the converged chains are presented in Fig. S3 and S4 and are mapped upon Fig. 1. For the 64-taxon analysis, four chains were run for ~3,000 generations. Two chains converged at ~200 generations, which was used as the burnin, (maxdiff = 0) and the posterior probabilities are mapped upon the ML tree in Fig. S1.

Fast-site removal and gene subsampling analyses

For fast site removal, rates of evolution at each site of the 61-taxon dataset were estimated with DIST_EST (Susko et al. 2003) under the LG model using discrete gamma probability estimation. A custom PYTHON script was then used to remove fastest evolving sites in 4,000-site steps. Random subsampling of 20, 40, 60, or 80% of the genes in the 61-taxon dataset was conducted using a custom PYTHON script, with the number of replicates as given in Fig 2B. In both cases each step or subsample was analyzed using 1,000 UFBOOT replicates in IQ-TREE under the LG+C60+F+Γ-PMSF model.

Data availability

All new transcriptomic data have been deposited at the National Center for Biotechnology Information under BioProjects PRJNA401035, as detailed in Table S1. All single gene alignments, masked and unmasked, and phylogenomic matrices are available in supplemental file Brown_etal.2017.CRuMs.tgz

Acknowledgements

The authors thank Tom Cavalier-Smith and Ed Glücksman (Oxford University) for supplying cultures strains B-70 (*Ancyromonas sigmoides*), NYK3C (*Fabomonas tropica*), and Bass1 (*Mantamonas plastica*). The part of this work conducted at Dalhousie University was supported by NSERC Discovery grants awarded to AGBS (298366-2014) and AJR (2016-06792) respectively. AJR also acknowledges the Canada Research Chairs program for support. This project was supported in part by the National Science Foundation (NSF) Division of Environmental Biology (DEB) grant 1456054 (http://www.nsf.gov), awarded to MWB. Mississippi State University's High Performance Computing Collaboratory provided some

computational resources. The part of this work conducted at the University of Tsukuba was supported by grants from the Japan Society for the Promotion of Science (JSPS; 15H05606 and 15K14591 awarded to RK, 23117006 and 16H04826 awarded to YI, 15H04411 awarded to KI, and 15H05231 to TH) and by the "Tree of Life" research project (Univ. Tsukuba).

References

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, et al. 2012. The revised classification of eukaryotes. J. Eukaryot. Microbiol. 59:429–493.

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.

Burki F. 2014. The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb. Perspect. Biol. 6:a016147.

Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, Keeling PJ. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc. Biol. Sci. 283.

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

Cavalier-Smith T. 2013. 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, Chao EE, Snell EA, Berney C, Fiore-Donno AM, Lewis R. 2014. Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. Mol. Phylogenet. Evol. 81:71–85.

Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. 10:210.

Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, Vlček Č, Lang BF, Eliáš M. 2015. Bacterial proteins pinpoint a single eukaryotic root. Proc. Natl. Acad. Sci. U. S. A. 112:E693-699.

Glücksman E, Snell EA, Berney C, Chao EE, Bass D, Cavalier-Smith T. 2011. The novel marine gliding zooflagellate genus *Mantamonas* (Mantamonadida ord. n.: Apusozoa). Protist 162:207–221.

Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AGB, Roger AJ. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." Proc. Natl. Acad. Sci. U. S. A. 106:3859–3864.

He D, Fiz-Palacios O, Fu C-J, Fehling J, Tsai C-C, Baldauf SL. 2014. An alternative root for the eukaryote tree of life. Curr. Biol. CB 24:465–470.

Heiss AA, Walker G, Simpson AGB. 2011. The ultrastructure of *Ancyromonas*, a eukaryote without supergroup affinities. Protist 162:373–393.

Inagaki Y, Nakajima Y, Sato M, Sakaguchi M, Hashimoto T. 2009. Gene sampling can bias multi-gene phylogenetic inferences: the relationship between red algae and green plants as a case study. Mol. Biol. Evol. 26:1171–1178.

Kang S, Tice AK, Spiegel FW, Silberman JD, Pánek T, Cepicka I, Kostka M, Kosakyan A, Alcântara DM, Roger AJ, et al. 2017. Between a pod and a hard test: the deep evolution of amoebae. Mol. Biol. Evol.

Katz LA, Grant JR, Parfrey LW, Burleigh JG. 2012. Turning the crown upside down: gene tree parsimony roots the eukaryotic tree of life. Syst. Biol. 61:653–660.

Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol. Biol. 7 Suppl 1:S4.

Le SQ, Lartillot N, Gascuel O. 2008. Phylogenetic mixture models for proteins. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 363:3965–3976.

Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30:1188–1195.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Paps J, Medina-Chacón LA, 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 DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9:e1000602.

Pisani D, Pett W, Dohrmann M, Feuda R, Rota-Stabelli O, Philippe H, Lartillot N, Wörheide G. 2015. Genomic data do not support comb jellies as the sister group to all other animals. Proc. Natl. Acad. Sci. U. S. A. 112:15402–15407.

Rodrigue N, Lartillot N. 2014. Site-heterogeneous mutation-selection models within the PhyloBayes-MPI package. Bioinformatics 30:1020–1021.

Simpson AGB. 2003. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int. J. Syst. Evol. Microbiol. 53:1759–1777.

Simpson AGB, Eglt Y. 2016. Protist Diversification. In: Encyclopedia of Evolutionary Biology. Vol. 3. Elsevier. p. 344–360.

Susko E, Field C, Blouin C, Roger AJ. 2003. Estimation of rates-across-sites distributions in phylogenetic substitution models. Syst. Biol. 52:594–603.

Tice AK, Shadwick LL, Fiore-Donno AM, Geisen S, Kang S, Schuler GA, Spiegel FW, Wilkinson KA, Bonkowski M, Dumack K, et al. 2016. Expansion of the molecular and morphological diversity of Acanthamoebidae (Centramoebida, Amoebozoa) and identification of a novel life cycle type within the group. Biol. Direct 11:69.

Torruella G, de Mendoza A, Grau-Bové X, Antó M, Chaplin MA, del Campo J, Eme L, Pérez-Cordón G, Whipps CM, Nichols KM, et al. 2015. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. Curr. Biol. 25:2404–2410.

Wang H, Minh B, Susko E, Roger AJ. 2017. Modeling Site Heterogeneity with Posterior Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. Syst. Biol. in press.

Wang H-C, Li K, Susko E, Roger AJ. 2008. A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. BMC Evol. Biol. 8:331.

Yabuki A, Eikrem W, Takishita K, Patterson DJ. 2013. Fine structure of *Telonema subtilis* Griessmann, 1913: a flagellate with a unique cytoskeletal structure among eukaryotes. Protist 164:556–569.

Yabuki A, Ishida K-I, 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 PJ, Klaveness D, Shalchian-Tabrizi K. 2012. *Collodictyon*--an ancient lineage in the tree of eukaryotes. Mol. Biol. Evol. 29:1557–1568.