Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic ''supergroups''

Vladimir Hampl[a,b,c], Laura Hug[a], Jessica W. Leigh[a], Joel B. Dacks[d,e], B. Franz Lang[f], Alastair G. B. Simpson[b], and Andrew J. Roger[a,1]

[a]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Canada B3H 1X5; [b]Department of Biology, Dalhousie University, Halifax, NS, Canada B3H 4J1; [c]Department of Parasitology, Faculty of Science, Charles University, 128 44 Prague, Czech Republic; [d]Department of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom; [e]Department of Cell Biology, University of Alberta, Edmonton, AB, Canada T6G 2H7; and [f]Departement de Biochimie, Université de Montréal, Montréal, QC, Canada H3T 1J4

Edited by Jeffrey D. Palmer, Indiana University, Bloomington, IN, and approved January 22, 2009 (received for review August 12, 2008)

Nearly all of eukaryotic diversity has been classified into 6 suprakingdom-level groups (supergroups) based on molecular and morphological/cell-biological evidence; these are Opisthokonta, Amoebozoa, Archaeplastida, Rhizaria, Chromalveolata, and Excavata. However, molecular phylogeny has not provided clear evidence that either Chromalveolata or Excavata is monophyletic, nor has it resolved the relationships among the supergroups. To establish the affinities of Excavata, which contains parasites of global importance and organisms regarded previously as primitive eukaryotes, we conducted a phylogenomic analysis of a dataset of 143 proteins and 48 taxa, including 19 excavates. Previous phylogenomic studies have not included all major subgroups of Excavata, and thus have not definitively addressed their interrelationships. The enigmatic flagellate *Andalucia* is sister to typical jakobids. Jakobids (including *Andalucia*), Euglenozoa and Heterolobosea form a major clade that we name Discoba. Analyses of the complete dataset group Discoba with the mitochondrion-lacking excavates or ''metamonads'' (diplomonads, parabasalids, and Preaxostyla), but not with the final excavate group, *Malawimonas*. This separation likely results from a long-branch attraction artifact. Gradual removal of rapidly-evolving taxa from the dataset leads to moderate bootstrap support (69%) for the monophyly of all Excavata, and 90% support once all metamonads are removed. Most importantly, Excavata robustly emerges between unikonts (Amoebozoa + Opisthokonta) and ''megagrouping'' of Archaeplastida, Rhizaria, and chromalveolates. Our analyses indicate that Excavata forms a monophyletic suprakingdom-level group that is one of the 3 primary divisions within eukaryotes, along with unikonts and a megagroup of Archaeplastida, Rhizaria, and the chromalveolate lineages.

Chromalveolata | Discoba | long-branch attraction

For decades, molecular phylogeneticists have attempted to infer the deepest relationships within the eukaryotic domain of the tree of life. The first such studies examined phylogenies of single ubiquitous genes such as ribosomal RNAs, elongation factors and tubulins (e.g., refs. 1–5). This approach suffered from lack of resolution (stochastic error) because of the low number of informative sites and systematic error in tree estimation caused by model violations such as compositional heterogeneity among sequences, and problems related to long-branch attraction (LBA) (6–11). Worse, in some cases single genes had been transferred laterally between the lineages under examination (12). More recently, the availability of vast quantities of data from genome-sequencing and expressed sequence tag (EST) projects over a wide range of eukaryotes has allowed phylogenetic estimation from "supermatrices" of moderate (13–15) to large (16–23) numbers of genes. Although such phylogenomic approaches are less sensitive to stochastic error, they can, when the phylogenetic model is misspecified, reinforce systematic errors such as LBA, yielding apparently strong support for an incorrect phylogeny (16, 19, 24). Some recent analyses employ objective data filtering approaches that isolate and remove the sites or taxa that contribute most to these systematic errors (19, 24).

The prevailing model of eukaryotic phylogeny posits 6 major supergroups (25–28): Opisthokonta, Amoebozoa, Archaeplastida, Rhizaria, Chromalveolata, and Excavata. With some caveats, solid molecular phylogenetic evidence supports the monophyly of each of Rhizaria, Archaeplastida, Opisthokonta, and Amoebozoa (16, 18, 29–34). However, the monophyly of both the Chromalveolata and the Excavata remains controversial (34), with recent evidence indicating that some, but not all, of the chromalveolate lineages are most closely related to Rhizaria (20, 22, 35). Of critical importance to both the placement of the root, and to the overall classification of eukaryotes, the branching order among the 6 supergroups remains obscure.

The supergroup Excavata was proposed on the basis of shared morphological characters—a ventral feeding groove and associated cytoskeletal structures (38–40), with some additional taxa (parabasalids, euglenids, and oxymonads) linked to the group primarily through molecular studies (41–45). There is no solid molecular phylogenetic evidence for the monophyly of Excavata as a whole, but 3 subgroups are often individually recovered as clades: (*i*) Preaxostyla (*Trimastix* and oxymonads), (*ii*) Fornicata (diplomonads, retortamonads, *Carpediemonas*) plus Parabasalia; and (*iii*) an unnamed clade consisting of Euglenozoa, Heterolobosea, and Jakobida (14, 20, 40, 42, 43, 45–55). The first 2 subclades comprise anaerobes/microaerophiles without classical mitochondria, and are now often classified together as Metamonada (55), but the relationship between the two is controversial (compare ref. 14 with ref. 15). Finally, 2 small but crucial groups are of uncertain placement. The first, *Malawimonas*, exhibits a typical excavate morphology, but is not robustly associated with any one of the 3 main subclades in published molecular phylogenetic studies (15, 53). The second, *Andalucia*, is morphologically similar to Jakobida, but an affiliation with jakobids is usually not recovered, nor statistically supported, in analyses of small subunit rRNA genes (53, 55) or small multiprotein datasets (56).

In this article we examine the monophyly and the phylogenetic position of Excavata within eukaryotes using rigorous analyses of a

Author contributions: V.H., L.H., J.W.L., A.G.B.S., and A.J.R. designed research; V.H. and L.H. performed research; B.F.L. contributed new reagents/analytic tools; V.H., L.H., J.W.L., J.B.D., and A.J.R. analyzed data; and V.H., A.G.B.S., and A.J.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence data reported in this paper have been deposited at www.natur.cuni.cz/~vlada/phylogenomicanalysis.

[1]To whom correspondence should be addressed. E-mail: andrew.roger@dal.ca.

This article contains supporting information online at www.pnas.org/cgi/content/full/0807880106/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0807880106

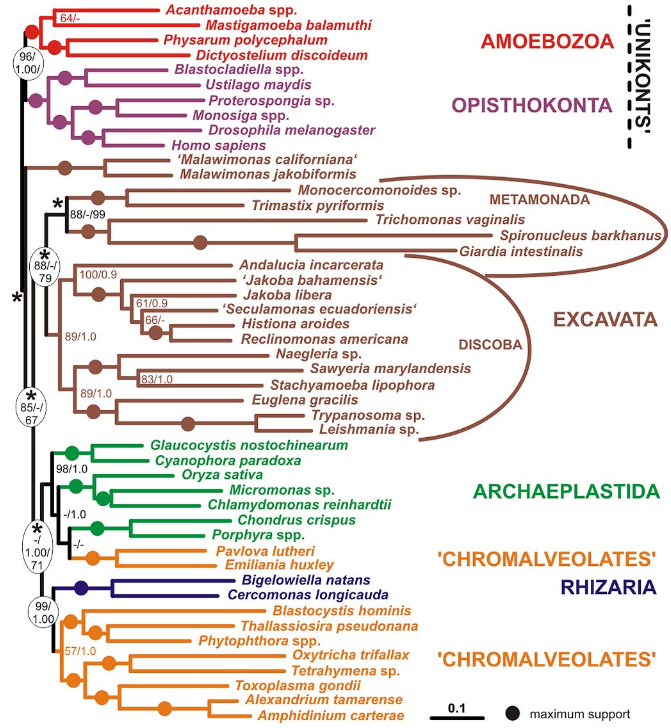

Fig. 1. The phylogenetic tree estimated from the main dataset. This topology received the highest likelihood in the exhaustive search of unconstrained nodes using the WAG+ Γ model; branch-lengths were calculated in RAxML using the WAG+ Γ model. The representatives of the 6 supergroups are color-coded. Asterisks indicate the nodes that were not constrained during the exhaustive search. The numbers at the nodes indicate bootstrap support calculated by RAxML bootstrapping/PhyloBayes posterior probability. At nodes that were not constrained during the exhaustive search in the separate analysis (asterisks), the third number indicates the RELL bootstrap value. Branches that received maximum possible support by all methods are indicated by full circles. Dashes indicate bootstrap values <50%, or posterior probabilities <0.5. Although the analyses did not assume a root, the tree is displayed with the basal split between ''unikonts'' and bikonts as suggested in ref. 37.

dataset of 143 genes and 48 taxa—the most complete phylogenomic representation of Excavata to date.

Results

We constructed a multigene dataset containing 143 genes (35,584 positions) for 48 taxa representing all 6 of the proposed eukaryotic supergroups (Tables S1 and S2). We chose to omit prokaryotic outgroup sequences, because previous analyses have shown that the long branch leading to these lineages causes long-branch attraction artifacts (LBA) that can compromise the accurate reconstruction of relationships among the major eukaryote taxa (24, 58). Phylogenetic congruence between all genes was not rejected by our analyses with Concaterpillar (59) at an α-level of 0.01, indicating that combined analyses of the genes were permissible. We performed a maximum likelihood (ML) analysis using RAxML (PROTCAT-WAG + Γ) with 100 bootstrap replicates and a Bayesian analysis using PhyloBayes using the CAT+ Γ model. The 2 methods produced slightly different trees. Based on these results, poorly supported deep nodes among the Excavata and eukaryotes were collapsed (nodes marked by asterisk in the Fig. 1) and the resulting 945 possible topologies were exhaustively searched and the branch support was estimated by resampling-estimated log-likelihood (RELL) bootstrap analysis (Fig. 1). The 3 established groupings of Excavata are all well supported. Notably, we recovered a strong clade of *Andalucia* with other jakobids, with this full jakobid clade branching as sister to Euglenozoa and Heterolobosea. We refer to this larger clade as Discoba (formally defined in the *SI Text*). Furthermore, we recovered bootstrap support 88% for a Metamonada grouping (i.e., Preaxostyla + diplomonads and parabasalids). This initial analysis did not support the monophyly of Excavata but instead recovered this taxon as 2 sequentially branching clades, (*i*) *Malawimonas* (bootstrap support 100%) and (*ii*) all other Excavata (bootstrap support 88%), emerging between unikonts and the rest of eukaryotes. The branch separating the 2 groups of Excavata (i.e., placing *Malawimonas* in a clan with unikonts and other Excavata in a clan with archaeplastids, chromalveolates and rhizarians) receives moderate to strong bootstrap support (85% RAxML boostrap support; 67% RELL bootstrap support in the constrained analysis). However, several topologies generated by rearranging the position of *Malawimonas* in the optimal tree to obtain a monophyletic Excavata could not be rejected by statistical tests (see *SI Text* and Table S3).

Many of the Excavata have very long branches in the tree, including the longest branches of all, diplomonads (*Giardia* and *Spironucleus*) and *Trichomonas*. By contrast, the lengths of the *Malawimonas* branches were extremely short. This great disparity led us to suspect that an LBA artifact might be responsible for the non-monophyly of the Excavata in the optimal tree. Specifically, the longer-branched Excavata could be attracted to the long branches of certain stramenopile, alveolate, and archaeplastid taxa, thereby separating them from *Malawimonas*, which clusters instead with the shorter unikont branches. To test this LBA hypothesis, we used 5 approaches intended to counter model misspecification and/or mutational saturation in the dataset that could potentially contribute to an LBA artifact: (*i*) amino acid recoding by functional categories, (*ii*) progressive fast-evolving site removal, (*iii*) use of an evolutionary model allowing gene-specific branch lengths (“separate analysis”), (*iv*) progressive long-branch taxon removal, and (*v*) progressive long-branch gene sequence removal. The first 2 approaches had virtually no effect on the results regarding the monophyly of Excavata (*SI Text* and Figs. S1–S4). However, it should be noted that maximum-likelihood based fast site removal resulted in strong support (peaking at 95%) for the position of the 2 Excavata lineages between unikonts and the rest of eukaryotes (Fig. S4), a result that will be discussed in greater depth below. Approach *iii* still resulted in an ML tree in which Excavata was polyphyletic. However, relative to the uniform model, the RELL bootstrap support for monophyletic Excavata increased, as did topology test *P* values (Table S3).

With regard to the monophyly of Excavata, the impact of approaches *iv* and *v* is described below.

Removal of Long-Branch (LB) Taxa. In this approach we gradually removed from the dataset the longest branching taxa as determined by measuring the distance of each taxon from a hypothetical root of the tree (root-to-tip distance). In this fashion, the robustness of a particular larger grouping can still be examined so long as at least 1 representative of the candidate component groups is still retained. To calculate these distances the root of the tree was placed in the center of the branch between the unikonts and the rest of the eukaryotes, consistent with the position proposed by Stechmann, Richards, and Cavalier-Smith (32, 36, 37), although other positions were also examined (see below). The bootstrap supports for branches were plotted against the number of taxa removed (Fig. 2*A*).

The support for the Preaxostyla (oxymonads + *Trimastix*) was always 100%, indicating that the shorter-branched *Trimastix* can be validly assumed to represent the whole clade in later analyses once the longer-branched oxymonad (*Monocercomonoides*) is removed. Before the removal of diplomonads and *Trichomonas*, Preaxostyla branched with these 2 groups (i.e., formed a Metamonada clade) with ≈80% bootstrap support (Fig. S5, points 0–2). With more

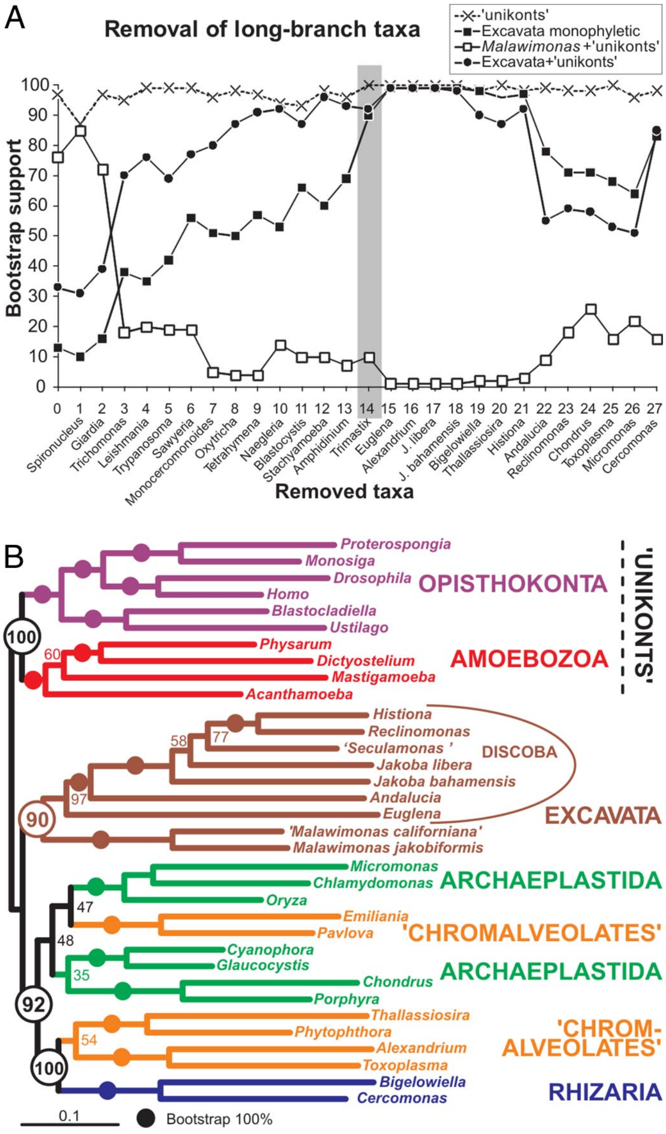

Fig. 2. LB taxon removal. (*A*) The support for the nodes of interest calculated by RAxML bootstrapping is plotted against the number of long-branch taxa that were removed from the concatenate. The support for the unikont bipartition (X) is used as a control. A root position at the midpoint of the branch connecting unikonts with the rest of eukaryotes was used to calculate root-to-tip distances of taxa. The order of taxon removal is given on the *x* axis. (*B*) A maximum likelihood tree after removal of 14 taxa (gray box in the part A). The tree was constructed in RAxML using WAG+ Γ model and colored as described in Fig. 1. The numbers at the nodes indicate bootstrap support calculated by RAxML bootstrapping. Branches that received maximum support by all methods are indicated by full circles.

caveats, Preaxostyla, and especially *Trimastix* may also be taken to represent the entire Metamonada grouping.

Until the removal of diplomonads and parabasalids, *Malawimonas* continued to group with the unikonts, and not other Excavata sequences (Fig. 2*A*, points 0–2). With the removal of the diplomonad and parabasalid sequences, the support for the *Malawimonas* + unikonts clan quickly dropped, and there was a concomitant increase in support for: (*i*) *Malawimonas* with the Preaxostyla (grouping I), (*ii*) the monophyly of Excavata as a whole (grouping II) and, (*iii*) the clan of Excavata + unikonts (grouping III) (Fig. 2*A*, point 3, Fig. S5, points 3 and 9). The bootstrap support for each of these 3 groupings reached 84%, 69% and 93% respectively after removal of 13 taxa (Fig. 2*A* and Fig. S6), at which point the dataset still retained members of the Discoba, *Trimastix* and *Malawimonas*, that is, members of all recognized Excavata subgroups, assuming Metamonada is considered a subgroup. After removal of *Trimastix* in the 14th iteration, the relationship of *Malawimonas* plus the remaining candidate excavates (grouping II) increased to 90% (Fig. 2*B*) and after removal of *Euglena* in the 15th iteration to 99% (Fig. 2*A*, point 15). In the deletion of 19–26 taxa, the support decreased for both grouping II and grouping III. This decrease was probably caused by LBA between the long branches of *Cercomonas* and typical jakobids (*Reclinomonas* and *Seculamonas*) that emerged when their sister taxa, *Bigelowiella* and *Andalucia*, were removed at levels 19 and 22, respectively. With the removal of *Cercomonas* at level 27, the support for both groups rose again.

Calculating root-to-tip distances of lineages from a root situated between opisthokonts and all other eukaryotes as suggested by Arisue et al. (17) yields exactly the same deletion series and hence the same results. Furthermore, broadly similar results were seen when other proposed eukaryote root positions were used (see *SI Text* and Figs. S7 and S8), with support for an Excavata clade containing *Malawimonas*, *Trimastix*, and Discoba peaking at 92% in the diplomonad/parabasalid rooting and 66% in the midpoint rooting.

Early in all taxon removal series, we observed steadily increasing support for the split in the unrooted eukaryotic tree between unikonts and Excavata on one side and Archaeplastida, Rhizaria, and chromalveolate lineages on the other. The support reached 99% for a region midway through each taxon deletion series before falling again as the number of taxa remaining in the dataset became very small (Fig. 2*A* and Fig. S7 and S8).

Removal of Long-Branch (LB) Gene Sequences. In this approach we aimed to remove individual gene sequences that had accumulated large numbers of changes, rather than removing entire taxa. We measured root-to-tip distances of each sequence in each gene tree and removed the 250, 500, 750, 1,000, 1,250, 1,500, 1,750, 2,000, and 2,250 longest-branched sequences from their respective gene alignments. The bootstrap support values for the nodes of interest were plotted against the number of LB sequences removed (Fig. 3*A*). As in the LB taxon removal analyses, the support for *Malawimonas* branching exclusively with the unikonts dropped continually and concomitant with an increase of the support for the monophyly of the Excavata as a whole and for the grouping of Excavata + unikonts. The support for the Excavata reached a peak (54%) after removing 1,750 LB gene sequences (Fig. 3).

Notwithstanding the limited support for excavate monophyly obtained by this method, removal of 750+ LB sequences resulted in very strong bootstrap support (always >90%, maximum 100%) for the unrooted split between unikonts and Excavata on one side and Archaeplastida, Rhizaria, and chromalveolate lineages on the other (Fig. 3*A*).

Analyses Using the CAT+ Γ Model. The 14 LB taxa-removed and 1,750 LB gene-removed datasets were also analyzed with the CAT+ Γ model implemented in PhyloBayes (60). In both cases, MCMC runs converged on a tree in which Excavata were not monophyletic. We performed Bayes factor analyses on multiple independent chains to determine whether the monophyly of Excavata topology was significantly worse than the topology recovered by PhyloBayes (see *SI Text*). Curiously, our analyses indicated that under the CAT+ Γ model, the monophyletic Excavata topology recovered by ML was actually slightly preferred (for the 14 LB taxa-removed dataset) or strongly preferred (for the 1,750 LB gene-removed dataset) over the paraphyletic Excavata tree recovered by unconstrained PhyloBayes analysis. The reasons for this discrepancy are currently unclear, but call into question the results of the unconstrained PhyloBayes analyses.

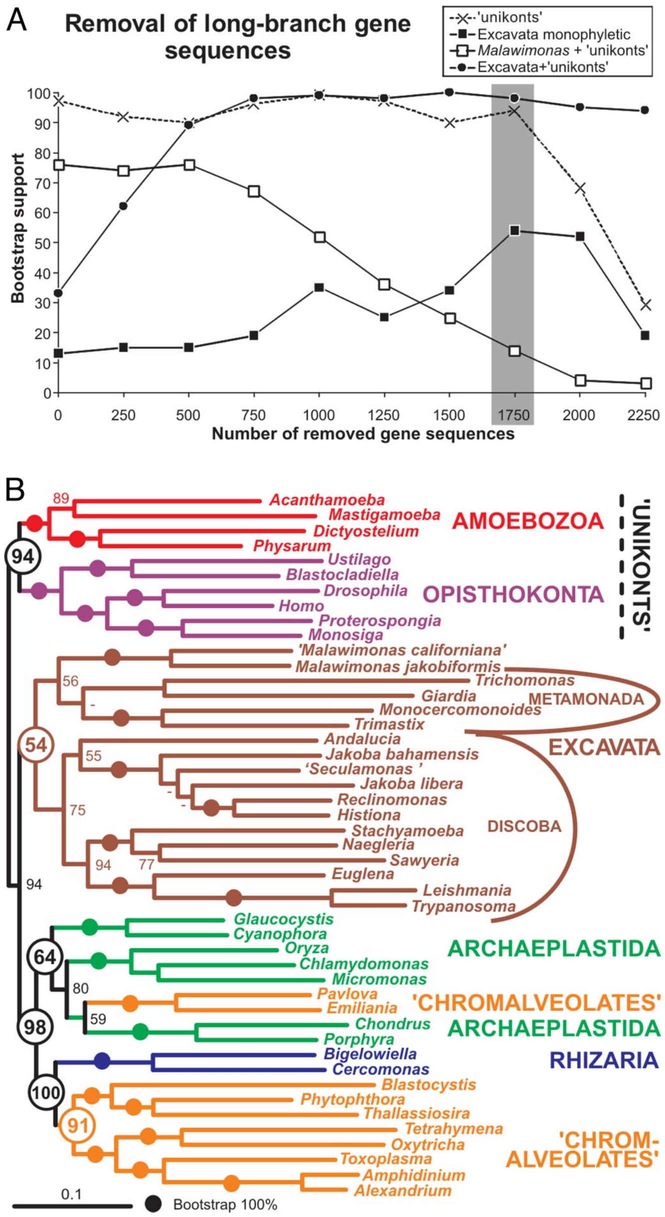

Fig. 3. LB gene sequence removal. (*A*) The support for the nodes of interest calculated by RAxML bootstrapping is plotted against the number of longest-branch gene sequences that were removed from the concatenate. The support for unikonts (X) is used as a control. (*B*) A maximum likelihood tree after the removal of 1,750 of the longest-branch gene sequences (gray box in the *A*). The tree was constructed in RAxML using the WAG+ Γ model and colored as described in Fig. 1. The numbers at the nodes indicate bootstrap support calculated by RAxML bootstrapping. Branches that received maximum support by all methods are indicated by full circles, dashes indicate bootstrap values <50%.

The data for these analyses were assembled at the end of 2006, at which time data from some important groups, specifically cryptophytes, were not available. To test whether inclusion of cryptophytes would significantly influence our results, we added cryptophytes to the dataset from Fig. 1 (see *SI Text* and Fig. S9). Cryptophyta branches weakly as sister to haptophytes and strongly within the Archaeplastida + Haptophyta clade. This suggests that inclusion of Cryptophyta has little effect on our results regarding Excavata monophyly and the deepest-level relationships among supergroups.

Discussion

Our analyses support the monophyly of a number of recognized higher eukaryotic taxa: Amoebozoa, Opisthokonta, Rhizaria, Alveolata, and Stramenopila. Because of the unexpected and strongly supported branching of haptophytes within Archaeplastida, our analyses do not support the monophyly of Archaeplastida (nor Chromalveolata). However, apart from the placement of haptophytes (and cryptophytes), the analyses recovered the unrooted branching order among eukaryotic supergroups with very good support.

Monophyly of Excavata. Although previous molecular analyses have not recovered an excavate clade, 3 higher-order groupings among excavates have been established: Metamonada, Discicristata + Jakobida, and *Malawimonas*. Our analyses that exclude the most rapidly evolving genes or species show that representatives of these 3 groupings are specifically related, suggestive of a complete Excavata clade. This was achieved largely due to the incorporation of *Trimastix*, which proved to be the shortest branched representative of Metamonada to be included in phylogenomic analysis to date. Although this result is not as straightforward as would be the recovery of a single excavate clade, given the extreme long-branch artifacts influencing the placement of the various taxa in question, this represents an important advance.

In the analyses of the full dataset, excavates formed 2 clades rather than one: the minor clade consisting of the 2 species of *Malawimonas* and the main clade consisting of the other 17 representatives. Because the 2 groups of Excavata differed markedly in their branch lengths, it is likely that LBA is responsible for their non-monophyly. A similar effect was observed by Rodriguez-Ezpeleta et al. in their analyses of a dataset including *Malawimonas*, jakobids, euglenozoans, and heteroloboseids (20). On theoretical grounds, it is expected that the use of more realistic models should mitigate the effect of the LBA that, in likelihood-based analyses of large datasets, is associated with model misspecification. Indeed, the use of a more realistic evolutionary model (independent model parameters and branch lengths for each gene-separate analysis) improved the support for Excavata monophyly, although this grouping was still not recovered as globally optimal.

Because saturation of substitutions at sites is expected to contribute to phylogenetic artifacts, we used various methods to detect and decrease the influence of saturation. Recoding and fast site removal analyses did not result in systematic changes in estimated topologies. However, the progressive removal of either LB taxa or LB gene sequences yielded Excavata monophyly. This is probably because the saturation occurs specifically in the sequences of the long branching lineages rather than in the sites whose average substitution rate across the whole alignment is increased. The LB taxa and LB gene sequence removal came at the cost of an extensive loss of data for long-branch Excavata, namely the diplomonads and *Trichomonas*. This, in turn, led to a poor representation of the Excavata for the LB taxon removal analysis and was probably responsible for the low support for Excavata monophyly recovered by the second gene-based LB sequence removal method.

In any case, a consistent trend toward the monophyly of the Excavata was observed in both LB removal analyses. In particular, a robust relationship between the Metamonada and the clade Discoba is seen in analyses of the initial dataset (Fig. 1), as is the affiliation of *Malawimonas* first with Preaxostyla and then with Discoba in the taxon removal studies (Fig. 2*B*). Most compellingly, a monophyletic Excavata clade is ultimately recovered midway through the long-branch taxon deletion series in datasets that still contain representatives of the established excavate higher order groupings, namely *Malawimonas*, *Trimastix* (representing metamonads), and several members of Discoba (bootstrap support 69%, 92%, and 66% in 3 different taxon removal series).

Internal Excavata Relationships—the Position of Malawimonas. As discussed earlier, the position of *Malawimonas* is of key importance to Excavata monophyly. After removal of LB taxa, *Malawimonas* branches within Excavata as a sister branch of Metamonada, robustly as a sister of Preaxostyla (*Trimastix* and *Monocercomonoides*) when diplomonads and parabasalids were excluded, and with the remaining excavates, with very high support, once all metamonads are removed (Fig. 2*B*). This strongly suggests that *Malawimonas* is a member of Excavata despite its separation from the others in analyses of the initial dataset.

The position of *Malawimonas* within Excavata is not clear. The long-branch taxon removal studies indicate a relationship with the various specific metamonad groups, but, because of low bootstrap support, cannot exclude the placement of *Malawimonas* cladistically within the metamonads. An affiliation of *Malawimonas* with Metamonada is consistent with the striking similarities in the flagellar apparatuses of *Malawimonas* and the diplomonad relative *Carpediemonas* (38, 51, 53).

Internal Excavata Relationships—the Jakobids Include *Andalucia* and Are Members of the Group Discoba. *Andalucia incarcerata* branched in all analyses as the sister to other Jakobida, with very strong statistical support. This corresponds well with the morphological similarities between *Andalucia* and other members of the Jakobida (51). *Andalucia* is a strikingly deep branch, however (the basal branch for Jakobida other than *Andalucia* is much longer than the basal branch for all jakobids), perhaps explaining why previous analyses based on 1–7 genes were unable to robustly resolve the position of this organism (15, 55, 56).

The Jakobida clade, including *Andalucia*, branched robustly as the sister group to a clade comprising Heterolobosea + Euglenozoa. A relationship between Jakobida, Heterolobosea, and Euglenozoa was recovered by several studies based on 1–6 genes, sometimes with strong statistical support (15, 20, 22), but the interrelationships among the 3 groups differed between datasets. Our results are consistent with those obtained recently by Rodriguez-Ezpeleta et al. without *Andalucia* (20).

For some time it has been common to consider Euglenozoa + Heterolobosea as one of the major clades of eukaryotes—Discicristata or "discicristates" (25, 41). This analysis resolves previous doubts about the monophyly of Discicristata with respect to all Jakobida based on some smaller datasets (15). The grouping of Discicristata with Jakobida represents a still more significant clade than Discicristata, and we think it useful to have a taxon name for this clade. We propose the name "Discoba" (defined in the *SI Text*).

Non-Monophyly of Chromalveolata and Archaeplastida. The Excavata are not the only supergroup to be examined in this study. Our analyses failed to support the monophyly of Archaeplastida and Chromalveolata because one putative chromalveolate group, the haptophytes, robustly branches within the Archaeplastida, as a sister to either the green plants or the rhodophytes. A similar position was recovered in other recent phylogenomic analyses (21, 23). Our data filtering experiments intended to minimize LBA showed that this artifact is unlikely to be responsible for this positioning of haptophytes. The recovered phylogenetic position of haptophytes, which are secondary algae, may result from the phylogenetic signal carried by genes that they acquired through endosymbiotic gene replacement associated with the origin of their plastids via secondary endosymbiosis of a red alga (61). However, the strong affiliation of haptophytes with the Archaeplastida yet weak affiliation to the red algae specifically suggests that either; (*i*) they may have acquired numerous genes from other primary algal lineages, or (*ii*) an unknown systematic bias in tree reconstruction obscures the phylogenetic signal from the red lineage. Another intriguing possibility is that the position of the haptophytes reflects the true phylogeny, with Chromalveolata being polyphyletic and the lineage of haptophytes + cryptophytes having arisen from within Archaeplastida, and having acquired their secondary plastid via an independent endosymbiosis from the one(s) that occurred in ancestors of stramenopiles and alveolates. Although not supported by our analyses, another formal possibility is that the position of Haptophyta within Archaeplastida reflects the true phylogenetic position of a monophyletic Chromalveolata (or a Chromoalveolata + Rhizaria clade).

All of our analyses strongly supported (86–100% bootstrap support) a relationship between Stramenopila, Alveolata, and Rhizaria, a recently reported assemblage of higher eukaryotic taxa (22, 35). Previous studies did not agree about the internal relationships within this "SAR" clade (20, 22, 23, 35). All of our analyses recovered the basal position of the Rhizaria within the SAR grouping. The Alveolata + Stramenopila clade was only weakly supported in analyses of the full dataset, but the bootstrap support increased to 91% after the removal of 1,750 LB gene sequences.

Relationships Among Supergroups: A Eukaryote Megagroup. Outside of the placement of the haptophytes, analyses with or without data filtering recovered a particular unrooted branching pattern for the major eukaryotic groups. All taxa classified as Excavata were placed between unikonts on one hand and Archaeplastida, Rhizaria and the chromalveolate lineages on the other hand. Although not well supported in the unfiltered analysis, the bootstrap support for this position rose considerably when either fast evolving sites, LB taxa or LB sequences were removed: 95% after exclusion of 18,584 fastest sites, 99% after exclusion of 15 LB taxa and 100% after exclusion of 1,500 LB gene sequences from gene trees.

No matter where the root of eukaryotes lies, our results indicate that Excavates are not uniquely related to any one of the other 5 iconic supergroups, and, if monophyletic, stemmed from a very deep branching event within the history of Eukaryotes. Probably the best-supported position for the root of the eukaryotic tree is between unikonts and all other Eukaryotes, in this context known as "bikonts" (32, 36, 37). If this rooting position is correct, the placement of Excavata that we have recovered implies that Excavata is the deepest branch among the well-studied bikonts, and consequently that Archaeplastida, Rhizaria and the chromalveolate lineages form a single massive clade to the exclusion of other well-known eukaryotes. This latter clade might be referred to as a megagroup, being composed of several supergroups. Evidence for this megagroup was also recently presented by Burki et al. (23). Although the position of the root of eukaryotes remains controversial, almost all contemporary hypotheses propose positions outside this megagroup, making it a strong working hypothesis for the deep-level phylogenetic structure within the eukaryotic domain.

Materials and Methods

The phylogenetic analyses included >100 individual bootstrapped ML analyses, and took several processor-years to complete. The sequence data were assembled and the alignments generated as described in *SI Text*. To assess phylogenetic congruence of the genes in the 143 gene supermatrix, we used Concaterpillar (59). The WAG+Γ model was used, and congruence was evaluated with an α-level cutoff of 0.01.

Tree Construction. From the concatenated alignment, alignments generated by LB removal (see below), and alignment of only Excavata (see *SI Text* and Fig. S10), trees were generated by the maximum likelihood method, using RAxML (62) under WAG+ Γ model (using the PROTGAMMACAT setting) with 4 categories of rate variation (100 bootstrap replicates were undertaken for estimation of node support). Bayesian analyses, implemented in PhyloBayes version 2.1c employing the CAT+ Γ model (60), were also performed on selected datasets.

Removal of Long Branches. ***Removal of LB taxa using root-to-tip distances.*** The distance from each taxon to the proposed root of the maximum likelihood tree of the main dataset was calculated using TreeStat (http://tree.bio.ed.ac.uk/software/treestat). The longest branched taxa were progressively removed. Trees were constructed and bootstrap support was calculated from each reduced dataset as described above.

Removal of LB gene sequences. First, gene-specific branch lengths were calculated by constraining the topology of the gene tree as in Fig. 1 and they were rooted between unikonts and the rest of eukaryotes. For every gene tree, the distance from each taxon to the root of the tree was calculated using TreeStat. The longest branched sequences were progressively removed from the gene alignments, gene alignments were concatenated and trees and bootstrap supports were calculated as above. If the representation of a taxon in the concatenated alignment dropped <5% of positions the taxon was removed from the concatenated alignment. If the number of taxa in the gene alignment dropped below 4, the gene was removed from the concatenation.

ACKNOWLEDGMENTS. We thank Dr. Hervé Phillipe (Université de Montréal) for providing the seed alignments. *Trimastix pyriformis* EST data were generated as part of A.J.R.'s involvement in the Genome Atlantic/Genome Canada-sponsored Protist EST Program Project. Some of these analyses were performed on the CamGrid computational resource. This work was supported by Canadian Institutes for Health Research Grant MOP-62809 (to A.J.R.), Natural Sciences and Engineering Research Council of Canada Grants 298366-2004 (to A.G.B.S.) and 227085-05 (to A.J.R.), Grant Agency of the Czech Republic Project 204/08/P280, The Grant Agency of the Academy of Sciences of the Czech Republic Project KJB501110801, The Czech Ministry of Education Project 0021620828 (V.H.), a Natural Sciences and Engineering Research Council E.W.R. Steacie Memorial Fellowship (to A.J.R.), the Canadian Institute for Advanced Research Program in Integrated Microbial Diversity (A.J.R. and A.G.B.S.), a Parke Davis foundation fellowship (to J.B.D.), a Nova Scotia Health Research Foundation scholarship (to J.W.L.), and Natural Sciences and Engineering Research Council and Killam Pre-Doctoral fellowships (to L.H.).

1. Hashimoto T, et al. (1994) Protein phylogeny gives a robust estimation for early divergences of eukaryotes—phylogenetic place of a mitochondria-lacking protozoan *Giardia lamblia. Mol Biol Evol* 11:65–71.
2. Hashimoto T, et al. (1995) Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia* inferred from amino acid sequences of elongation factor 2 *Mol Biol Evol* 12:782–793.
3. Kumar S, Rzhetsky A (1996) Evolutionary relationships of eukaryotic kingdoms. *J Mol Evol* 42:183–193.
4. Sogin ML, Silberman JD (1998) Evolution of the protists and protistan parasites from the perspective of molecular systematics. *Int J Parasitol* 28:11–20.
5. Edgcomb VP, Roger AJ, Simpson AGB, Kysela DT, Sogin ML (2001) Evolutionary relationships among ''jakobid'' flagellates as indicated by alpha- and beta-tubulin phylogenies. *Mol Biol Evol* 18:514–522.
6. Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* 27:401–410.
7. Hirt RP, et al. (1999) Microsporidia are related to fungi: Evidence from the largest subunit of RNA polymerase II and other proteins. *Proc Natl Acad Sci USA* 96:580–585.
8. Stiller JW, Hall BD (1999) Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol Biol Evol* 16:1270–1279.
9. Philippe H (2000) Opinion: Long branch attraction and protist phylogeny. *Protist* 151:307–316.
10. Dacks JB, Marinets A, Doolittle WF, Cavalier-Smith T, Logsdon JM (2002) Analyses of RNA polymerase II genes from free-living protists: Phylogeny, long branch attraction, and the eukaryotic big bang. *Mol Biol Evol* 19:830–840.
11. Gribaldo S, Philippe H (2004) in *Organelles, Genomes and Eukaryote phylogeny*, eds Horner DS, Hirt RP (CRC, London), pp 27–53.
12. Andersson JO (2005) Lateral gene transfer in eukaryotes. *Cell Mol Life Sci* 62:1182–1197.
13. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290:972–977.
14. Hampl V, et al. (2005) Inference of the phylogenetic position of oxymonads based on nine genes: Support for Metamonada and Excavata. *Mol Biol Evol* 22:2508–2518.
15. Simpson AGB, Inagaki Y, Roger AJ (2006) Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of ''primitive'' eukaryotes. *Mol Biol Evol* 23:615–625.
16. Bapteste E, et al. (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium, Entamoeba,* and *Mastigamoeba*. *Proc Natl Acad Sci USA* 99:1414–1419.
17. Arisue N, Hasegawa M, Hashimoto T (2005) Root of the eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. *Mol Biol Evol* 22:409–420.
18. Burki F, Pawlowski J (2006) Monophyly of Rhizaria and multigene phylogeny of unicellular bikonts. *Mol Biol Evol* 23:1922–1930.
19. Rodriguez-Ezpeleta N, et al. (2007) Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol* 56:389–399.
20. Rodriguez-Ezpeleta N, et al. (2007) Toward resolving the eukaryotic tree: The phylogenetic positions of jakobids and cercozoans. *Curr Biol* 17:1420–1425.
21. Patron NJ, Inagaki Y, Keeling PJ (2007) Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol* 17:887–891.
22. Burki F, et al. (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* 2:e790.
23. Burki F, Shalchian-Tabrizi K, Pawlowski J (2008) Phylogenomics reveals a new ''megagroup'' including most photosynthetic eukaryotes. *Biol Lett* 4:366–369.
24. Brinkmann H, van der Giezen M, Zhou Y, Poncelin de RG, Philippe H (2005) An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst Biol* 54:743–757.
25. Baldauf SL (2003) The deep roots of eukaryotes. *Science* 300:1703–1706.
26. Simpson AGB, Roger AJ (2004) The real ''kingdoms'' of eukaryotes. *Curr Biol* 14:R693–R696.
27. Keeling PJ, et al. (2005) The tree of eukaryotes. *Trends Ecol Evol* 20:670–676.
28. Adl SM, et al. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukc Microbiol* 52:399–451.
29. Archibald JM, Longet D, Pawlowski J, Keeling PJ (2003) A novel polyubiquitin structure in Cercozoa and Foraminifera: Evidence for a new eukaryotic supergroup. *Mol Biol Evol* 20:62–66.
30. Nikolaev SI, et al. (2004) The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. *Proc Natl Acad Sci USA* 101:8066–8071.
31. Bass D, et al. (2005) Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. *Protist* 156:149–161.
32. Richards TA, Cavalier-Smith T (2005) Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436:1113–1118.
33. Rodriguez-Ezpeleta N, et al. (2005) Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Curr Biol* 15:1325–1330.
34. Parfrey LW, et al. (2006) Evaluating support for the current classification of eukaryotic diversity. *PLoS Genet* 2:e220.
35. Hackett JD, et al. (2007) Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of ''Rhizaria'' with chromalveolates. *Mol Biol Evol* 24:1702–1713.
36. Stechmann A, Cavalier-Smith T (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science* 297:89–91.
37. Stechmann A, Cavalier-Smith T (2003) The root of the eukaryote tree pinpointed. *Curr Biol* 13:R665–R666.
38. Simpson AGB, Patterson DJ (1999) The ultrastructure of *Carpediemonas membranifera* (Eukaryota) with reference to the ''Excavate hypothesis.'' *Eur J Protistol* 35:353–370.
39. Simpson AGB (2003) Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). *Int J Syst Evol Microbiol* 53:1759–1777.
40. Simpson AGB, Roger AJ (2004) in *Organelles, Genomes and Eukaryote phylogeny*, eds Horner DS, Hirt RP (CRC, London), pp 27–53.
41. Cavalier-Smith T (1998) A revised six-kingdom system of life. *Biol Rev Camb Philos Soc* 73:203–266.
42. Henze K, et al. (2001) Unique phylogenetic relationships of glucokinase and glucosephosphate isomerase of the amitochondriate eukaryotes *Giardia intestinalis, Spironucleus barkhanus* and *Trichomonas vaginalis*. *Gene* 281:123–131.
43. Dacks JB, et al. (2001) Oxymonads are closely related to the excavate taxon *Trimastix*. *Mol Biol Evol* 18:1034–1044.
44. Cavalier-Smith T (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol* 52:297–354.
45. Andersson JO, Sarchfield SW, Roger AJ (2005) Gene transfers from Nanoarchaeota to an ancestor of diplomonads and parabasalids. *Mol Biol Evol* 22:85–90.
46. Page FC, Blanton RL (1985) The Heterolobosea (Sarcodina, Rhizopoda), A new class uniting the Schizopyrenida and the Acrasidae (Acrasida). *Protistologica* 21:121–132.
47. Fenchel T, Patterson DJ (1986) *Percolomonas cosmopolitus* (Ruinen) n.gen., a new type of filter feeding flagellate from marine plankton. *J Mar Biol Ass UK* 66:465–482.
48. Flavin M, Nerad TA (1993) *Reclinomonas americana* n. g., n. sp., a new freshwater heterotrophic flagellate. *J Eukaryot Microbiol* 40:172–179.
49. O'Kelly CJ (1993) The jakobid flagellates—structural features of *Jakoba, Reclinomonas* and *Histiona* and implications for the early diversification of eukaryotes *J Euk Microbiol* 40:627–636.
50. Simpson AGB (1997) The identity and composition of the Euglenozoa. *Arch Protistenk* 148:318–328.
51. Simpson AGB, Patterson DJ (2001) On core jakobids and excavate taxa: The ultrastructure of *Jakoba incarcerata*. *J Euk Microbiol* 48:480–492.
52. Silberman JD, et al. (2002) Retortamonad flagellates are closely related to diplomonads—Implications for the history of mitochondrial function in eukaryote evolution. *Mol Biol Evol* 19:777–786.
53. Simpson AGB, et al. (2002) Evolutionary history of ''early-diverging'' eukaryotes: The excavate taxon *Carpediemonas* is a close relative of Giardia. *Mol Biol Evol* 19:1782–1791.
54. Cavalier-Smith T (2003) The excavate protozoan phyla Metamonada Grasse 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.
55. Lara E, Chatzinotas A, Simpson AGB (2006) *Andalucia* (n. gen.)—the deepest branch within jakobids (Jakobida; Excavata), based on morphological and molecular study of a new flagellate from soil. *J Euk Microbiol* 53:112–120.
56. Simpson AGB, Perley TA, Lara E (2007) Lateral transfer of the gene for a widely used marker, alpha-tubulin, indicated by a multi-protein study of the phylogenetic position of *Andalucia* (Excavata). *Mol Phylogenet Evol* 47:366–377.
57. Wilkinson MF, McInerney JO, Hirt RP, Foster PG, Embley TM (2007) (2007) Of clades and clans: Terms for phylogenetic relationships in unrooted trees. *Trends Ecol Evol* 22:114–115.
58. Brinkmann H, Philippe H (2007) The diversity of eukaryotes and the root of the eukaryotic tree. *Adv Exp Med Biol* 607:20–37.
59. Leigh JW, Susko E, Baumgartner M, Roger AJ (2008) Testing congruence in phylogenomic analysis. *Syst Biol* 57:104–115.
60. Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21:1095–1109.
61. Bhattacharya D, Yoon HS, Hackett JD (2004) Photosynthetic eukaryotes unite: Endosymbiosis connects the dots. *Bioessays* 26:50–60.
62. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.