The new phylogeny of eukaryotes

Hervé Philippe*, Agnès Germot† and David Moreira‡

Molecular phylogeny has been regarded as the ultimate tool for the reconstruction of relationships among eukaryotes – especially the different protist groups – given the difficulty in interpreting morphological data from an evolutionary point of view. In fact, the use of ribosomal RNA as a marker has provided the first well resolved eukaryotic phylogenies, leading to several important evolutionary hypotheses. The most significant is that several early-emerging, amitochondriate lineages, are living relics from the early times of eukaryotic evolution. The use of alternative protein markers and the recognition of several molecular phylogeny reconstruction artefacts, however, have strongly challenged these ideas. The putative early emerging lineages have been demonstrated as late-emerging ones, artefactually misplaced to the base of the tree. The present state of eukaryotic evolution is best described by a multifurcation, in agreement with the 'big bang' hypothesis that assumes a rapid diversification of the major eukaryotic phyla. For further resolution, the analysis of genomic data through improved phylogenetic methods will be required.

Addresses
*Equipe Phylogénie, Bioinformatique et Génome, UMR CNRS 7622, 9 quai Saint-Bernard, Case 24 75252, Paris Cedex 05, France
†Equipe Atipe UPRESA CNRS 8080, Développement & Evolution Université Paris, 11 91405 Orsay Cedex, France
‡División de Microbiología, Facultad de Medicina, Universidad Miguel Hernández, 03550 San Juan de Alicante, Spain

Current Opinion in Genetics & Development 2000, **10**:596–601

0959-437X/00/$ – see front matter

Abbreviations
LBA long-branch attraction
rRNA ribosomal RNA

Introduction

Detailed comparative studies of morphological data, made possible mainly through the use of electron microscopy, was very useful in delineating many eukaryotic groups but failed to resolve their relationships [1]. A great hope was thus placed in molecular phylogeny and, indeed, a eukaryote tree rapidly emerged from the analysis of the small subunit ribosomal RNA (rRNA) sequences [2]. This tree, which is still used as a reference framework, can be split into two parts: the 'crown', in which the major phyla — green plants, animals, fungi, stramenopiles, alveolates, and red algae — emerge almost simultaneously, and the base, which shows the paraphyletic emergence of numerous protist phyla: various amoeboid and amoeboflagellate organisms, euglenozoans, and the first-emerging lineages, diplomonads, microsporidia and trichomonads. Three important hypotheses have been developed on the basis of this tree: First, the three early-branching lineages, which lack mitochondria, emerged before the mitochondrial endosymbiosis and are living relics from an amitochondrial period of eukaryotic evolution (the Archezoa hypothesis) [3]; second, "molecular evolutionary distances between divergent eukaryotic taxa eclipse those observed in the entire prokaryotic world" [2]; and third, "the separation of the major groups of eukaryotes occurred nearly simultaneously" [2]. As stated by Sogin [2], these hypotheses can be tested with improved analytical techniques to recover phylogenetic information and with the sequencing of alternative phylogenetic markers unrelated to rRNA. As discussed here, the first two hypotheses have been challenged by recent results.

Secondary loss of mitochondria and the Archezoa hypothesis

The Archezoa hypothesis can be tested easily. As many mitochondrial genes have been transferred to the nucleus — with their products located in the mitochondria or in the cytosol — it is theoretically possible to recognize genes that remained in the nucleus after mitochondrial loss. In fact [4,5•], such genes have been found in all eukaryotes for which searches have been carried out. It has been suggested that transient endosymbioses, distinct from the mitochondrial one, however, could also have been the source of these genes [6]. Val-tRNA synthetase may represent such an example because the eukaryotic gene shares an insertion of ~40 amino acids with β- and γ-Proteobacteria [7] but not with α-Proteobacteria (*Caulobacter*, *Rhodobacter*, *Rhodopseudomonas* and *Sinorhizobium*; *Rickettsia* being a special case because its gene is of archaeal origin) (H Philippe, unpublished data). A non-mitochondrial origin seems unlikely for Cpn60 and Hsp70 because sequences from amitochondriate protists clearly emerge within the mitochondrial clade. Moreover, the discovery of an organelle containing the mitochondrial Cpn60 in *Entamoeba* [8,9] and of a hydrogenosome — H_2-producing organelle particularly present in trichomonads — containing its own genome [10•], strongly suggest that mitochondria still persist in many, if not all, amitochondriate protists as small cryptic organelles or as hydrogenosomes. The finding of mitochondrial-type Cpn60 and Hsp70 is fully compatible with this persistence of membrane-bound mitochondria-related organelles as these proteins are involved in the import of proteins across the mitochondrial membrane.

A similar conclusion can be reached for spliceosomal introns, which have been proposed to be of recent origin (after the emergence of Euglenozoa) [11]. Introns have now been described in *Euglena* [12] and in microsporidia [13], however, and an essential spliceosomal component, PRP8, is found in trichomonads [14] and diplomonads (accession number AC040617). Therefore, current data imply that the common ancestor of extant eukaryotes possessed a mitochondrion and most likely had spliceosomal introns. The

simplest eukaryotes studied to date, thought to be living relics of premitochondrial eukaryotes, are, on the contrary, highly derived organisms which have lost complex features in contradiction to the Archezoa hypothesis.

Tree-reconstruction artefacts and eukaryotic phylogeny

Demonstration of the significant impact of tree-reconstruction artefacts on eukaryotic phylogeny, especially on the rRNA-based one, has been an important development of the past two years. When we first proposed at meetings in 1996 that all the lineages branching before the crown in the rRNA tree were misplaced because of tree-reconstruction artefacts [15], evolutionists at best were willing to accept this notion for only a few lineages, such as microsporidia [16,17]. In 1998, two papers published in this journal [4,18] suggested that the early emergence of several basal lineages could be caused by the long-branch attraction (LBA) artefact [19]. This artefact is especially troublesome in the case of eukaryotes because the very distantly related outgroup (generally Archaea), a long branch itself, attracts the long branches of fast-evolving groups.

Several recent papers have provided evidence supporting our 1996 claim. First, many genes used for inferring the eukaryotic phylogeny (e.g. rRNA, actin, tubulin and elongation factor 1α) are highly mutationally saturated [15,20]. Furthermore, saturated genes make believe in a molecular clock even if evolutionary rates are actually highly variable [18]. This explains why rRNA has incorrectly been considered a good molecular chronometer [2] and why LBA is so difficult to detect.

Second, the evolutionary rates within different eukaryotic phyla have been estimated for several genes and it has been shown that the faster a phylum evolves the earlier it emerges (for example, euglenozoans for rRNA and ciliates for actin) [15]. This result is in agreement with our hypothesis that the order of emergence is dictated by the evolutionary rate (through the LBA artefact) and not by the historical pattern.

Third, the addition of new sequences in phylogenetic analyses, which is known to reduce the impact of the LBA artefact [18], results in an upward movement of the early-branching species in the tree [21]. This is also congruent with the fact that these early-branching species are, in fact, fast-evolving ones.

Fourth, the use of more realistic models of sequence evolution — known to attenuate the impact of LBA [18] — leads, in rRNA trees, to a later emergence of euglenozoans [22,23], microsporidia [23,24], *Physarum* [23], trichomonads and heteroloboseans [25•]. This significant improvement is obtained by modelling the rate heterogeneity among sequence sites with a Γ law. A simpler approach is the removal of invariant sites, which is less efficient but computationally faster than the use of a Γ law [26•]. Although existing models assume that the evolutionary rate of each site is constant throughout time, this is often not true, in agreement with the covarion model [27]. A detailed comparison of a classic rRNA tree (i.e. inferred without taking into account rate heterogeneity) with a revised tree, in which all the lineages that branch before the crown have been relocated to positions supported by other markers, also revealed the importance of the covarion model violation [28••]. In fact, the revised tree is strongly rejected by a Kishino–Hasegawa test when rate heterogeneity is ignored ($\Delta\ln L = -5.16$ s.e.), still rejected when rate heterogeneity is handled through a Γ law ($\Delta\ln L = -2.27$ s.e.) but not rejected ($\Delta\ln L = -1.35$ s.e.) when rate heterogeneity is taken into account and when covarion-like sites (i.e. sites showing heterogeneous evolutionary rates throughout the tree) are removed [28••]. The importance of covarions is also demonstrated by the correlation between the number of variable sites within a lineage and its order of emergence [29••,30•]. The larger the number of variable sites is (i.e. the faster evolving), the earlier the emergence is, which is congruent with the idea that basal emergence is caused by LBA.

Fifth, several lines of evidence — highly heterogeneous rRNA length, large number of unique substitutions, attraction by artificial random sequences and high RASA (relative apparent synapomorphy analysis) taxon variance — suggested that the basal lineages of the rRNA tree are fast-evolving [31]. However, these four phenomena would also be expected even if the basal rRNA tree were correct. For instance, simulation studies demonstrate that RASA does not differentiate long branches caused by high evolutionary rates from long branches reflecting long history (P Lopez, H Philippe, unpublished data).

Sixth, if a basal emergence in the rRNA tree is correct, one expects that the slowly-evolving positions, which contain most of the ancient phylogenetic information, will provide strong support for the basal branching, and that the fast-evolving positions, which contain mainly noise for ancient events, will only weaken this support. When using the 'Slow-Fast' method (which allows the identification of the slowly evolving positions and reconstructs trees from them) [32], however, exactly the opposite phenomenon is observed [29••]. The basal taxa in the standard rRNA tree do not emerge early when only slow-evolving positions are used but display very long branches. They appear to emerge increasingly earlier as fast-evolving positions are added in the analysis. This study thus provides strong support for the idea that the basal emergence in the rRNA tree is only the result of an LBA artefact. In addition, the same phenomenon has been shown also for proteins (elongation factor-1α and α-tubulin) [29••]. As a result, the LBA artefact supplies a simple explanation to the contradictions observed between phylogenies on the basis of different markers, because the fast-evolving lineages which are artefactually early-branching ones are not the same for the various markers.

These analyses demonstrate that all basal lineages of the rRNA tree are actually part of the crown, such as microsporidia within fungi [24,26•,33••]. As the crown was already difficult to resolve because of its short internal branches, its resolution is now even more difficult because of the addition of new phyla (increasing the number of nodes), often fast evolving (increasing the quantity of noise). Therefore, phylogenies based on single genes will most likely provide resolution for only a few clades but not for the entire phylogeny. Moreover, specific biases (special evolutionary properties of one phylum, horizontal gene transfer etc.) lower the reliability of single-gene approaches. Resolution of the branching orders within the enlarged crown will most likely be achieved either first, through a combined analysis of multiple genes in order to have a sufficient number of informative positions or second, from the analysis of structural genomic data less prone to bias — such as gene fusion, insertion/deletion, and gene duplication.

Recent progress in the resolution of the eukaryotic phylogeny

Several protein markers are now available for a significant diversity of eukaryotes — for instance, actin [15,34•,35], elongation factor 1α [20,21,34•], elongation factor *2* [36•], enolase [37•,38], Hsp70 [30•], malate dehydrogenase [39,40], RNA polymerase B1 [26•,41,42], TBP (TATA box binding protein) [43], and tubulins [15,33••,34•]. As expected, in the trees derived from these markers, the relationships between major phyla are always poorly resolved, except for the grouping of fungi/microsporidia (supported by EF2, RPB1 and tubulins) and of red algae/green plants (supported by EF2). Single-gene analysis appears to be more useful for determining the phylogenetic status of morphologically enigmatic protists, such as *Reticulomyxa filosa*, a giant freshwater amoeba, within Foraminifera [35]. One notable exception could be the emergence of trichomonads at the base of eukaryotes derived from enolase sequences, supported by high bootstrap values and by two small deletions of only one amino acid each [38]. These deletions are also present in one archaeon and one bacterium [37•], however, suggesting that this is an unstable region. Moreover, the phylogeny based on the amino-terminal part of the enolase places the trichomonads with eukaryotes, while the phylogeny based on the carboxy-terminal part places them within bacteria. This incongruent branching pattern derived from the two parts of the protein sequence suggests that trichomonad enolase derives from a recombination event between their ancestral eukaryotic-type gene and a gene of bacterial origin (E Bapteste, H Philippe, unpublished data). The source of the bacterial-type enolase could be the same as the bacterial-type GAPDH of trichomonads [44]. This exemplifies the weakness of the single-gene approach to solve the eukaryotic phylogeny.

To increase the resolving power of molecular phylogeny, the use of larger amounts of information is necessary; however, the simultaneous analysis of combined sequences displaying various biases can be biased if the tree reconstruction method is inefficient [45]. Few combined analyses have been carried out and the results, albeit encouraging, are still poorly resolved [28••,30•,46•]. A fusion of four genes [34•] has provided several robust nodes, but the phylogeny was unrooted — making it impossible to differentiate between early- and late-emerging groups — and included few fast-evolving sequences, which are known to generate inference problems. The largest fusion (13 proteins, 5171 positions) yields a fully resolved tree [36•] but the three fastest-evolving lineages — diplomonads, euglenozoans and apicomplexans — emerge at the base, suggesting LBA artefacts, all the more so that few species are used. Only two well-supported groupings were consistently retrieved from these analyses: the animal/fungi and the red algae/green plants clades. Combined analysis especially requires that the tree reconstruction method be consistent, otherwise the longer the sequences, the stronger the bias.

What's next?

As we have seen, the existing consensus view of eukaryotic phylogeny is still far from settled for the relationships between the major phyla, except for the alveolates and the two clades already discussed (Figure 1). Interestingly, this means that molecular phylogenies, despite the use of many thousands of informative characters, do not yield a tree much more resolved than the one generated using a few hundred morphological characters [47]. The lack of molecular and morphological signatures defining supergroups is in agreement with our big-bang hypothesis [15], which proposes that all the eukaryotic phyla emerged in a relatively short period of time. This could explain the generalised lack of resolution of molecular phylogenies. Lack of data and inefficient tree reconstruction methods, however, are valid alternative explanations [4,29••] and much more work is thus needed to evaluate the actual timespan for eukaryotic diversification. Nevertheless, it now appears that all extant, known eukaryotic lineages diverged in a timespan considerably shorter than the time elapsed since — advancing the important conclusion that no contemporary eukaryotic group is significantly more primitive than the others. Therefore, the quest for 'living relic' eukaryotic lineages is exciting but, unfortunately, will most likely prove fruitless. In agreement with this idea, some late-emerging lineages have, for instance, retained mitochondrial genes, such as the eubacterial-type RNA polymerase in *Reclinomonas americana* [46•,48] or MutS in the coral *Sarcophyton glaucum* [49], which are not found in presumed earlier-emerging lineages. Therefore, the precise knowledge of the characteristics of the common ancestor of eukaryotes will require the in-depth analysis and comparison of as many groups as possible and not only of the few ones suspected to be primitive.

Progress towards the complete resolution of eukaryotic phylogeny is necessary. The huge amounts of information coming from complete genome and expressed sequence tag sequencing projects will soon provide the required raw

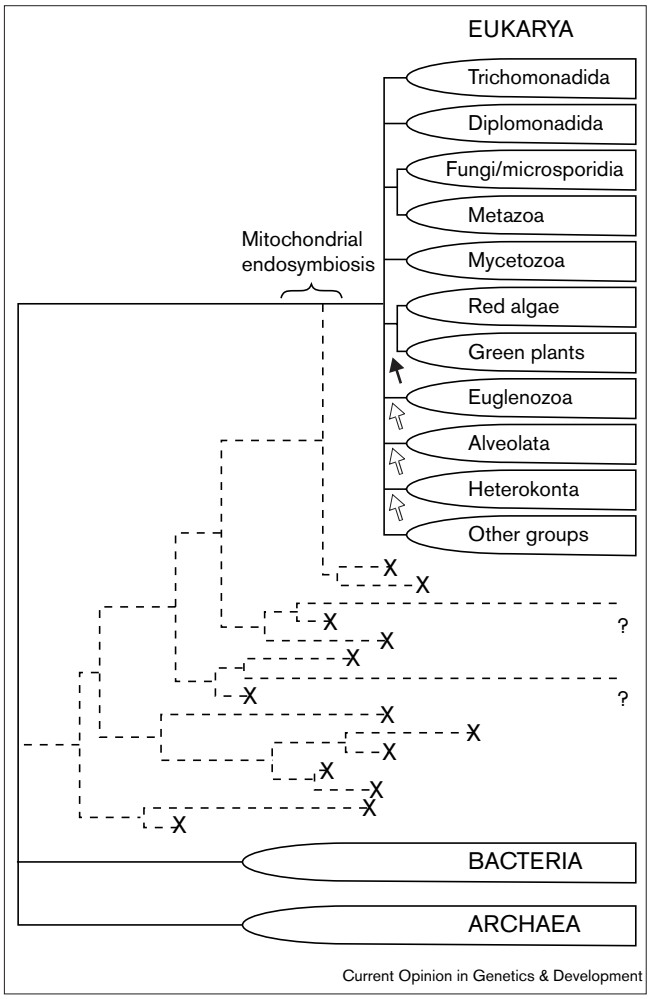
data. As discussed above, however, analyses made with existing phylogenetic reconstruction methods, even of such very large data sets, will be sensitive to artefacts. Improvement of these methods, in particular by taking into account the covarion model, is thus essential [28••,50]. On the other hand, the genomic data should be subjected to exploration by different means (see [18]): first, the analysis of all the unambiguous insertions/deletions; second, the study of specific gene fusions, such as that of *cox*1 and *cox*2, (supporting the monophyly of mycetozoa [5]) or of dihydrofolate reductase and thymidilate synthase genes (suggesting an early emergence of fungi/metazoa [29••]); third, shared gene duplications, such as those of release factors RFS and RF3 [51] and V-ATPase C-subunit (E Bapteste, H Philippe, unpublished data); and fourth, shared horizontally transferred genes [52]. For the latter, the available evidence suggests that horizontal transfers from bacteria to eukaryotes are more frequent than transfers between eukaryotes (with no example known yet, except for selfish elements). These known Bacteria→Eukarya transfers appear to specially affect metabolic genes, in particular those involved in the adaptation to energy production under anaerobic conditions, which explains the mosaic distribution of genes found in several anaerobic protists (such as *Giardia*, *Trichomonas*, and *Entamoeba*) [37•]. Obviously, these transfers may mislead phylogenetic reconstruction if they remain undetected but, on the other hand, they may help to find inter-phyla relationships if they are characterised appropriately. Anyway, on the basis of available data, even these Bacteria→Eukarya transfers seem to be much more infrequent than transfers between prokaryotes.

Conclusions

Only a part of the eukaryotic rRNA-based phylogeny, albeit inferred from a single gene with simple methods, has been challenged by the huge amount of new data and the use of improved methods: all the basal lineages in this tree now emerge in the crown. This rules out the Archezoa hypothesis and the hypothesis of eukaryotic diversity, being greater than the prokaryotic one in terms of evolutionary distance. In fact, because all extant eukaryotic lineages are derived from a mitochondrion-bearing ancestor, their diversity is smaller than that of a single eubacterial phylum (α-proteobacteria) and is probably not much larger than that of plants, animals and fungi. On the contrary, the third hypothesis derived from the rRNA tree — a rapid diversification of the crown taxa — has been entirely confirmed.

The new phylogenetic framework for eukaryotes derived from the big-bang hypothesis is characterised by a long, unbroken basal branch and a rapid diversification of extant lineages (Figure 1). This implies that no actual primitive eukaryotes are known. Although this surprising lack may merely be caused by a bias of sampling in the study of protist diversity, an alternative explanation may be that all true primitive forms have disappeared. Such a long, unbroken basal branch is not a unique phenomenon because mammals, birds or angiosperms also display one. An important multidisciplinary research field will be the identification of the biological or non-biological cause for the massive extinction and massive diversification of eukaryotes. Such a momentous extinction might have been provoked by a key evolutionary innovation that would have increased the adaptive fitness of the common ancestor of modern eukaryotes, which would have displaced all the less competitive lineages. This key innovation could have been the same as the one that triggered the massive diversification shown by extant lineages. An attractive candidate could be the acquisition of mitochondria although several authors prefer to put this event at the very beginning of eukaryotic origins.

Figure 1



The 'big bang' hypothesis for eukaryotic evolution. Extant eukaryotic phyla have emerged from a massive multifurcation, within which only a few supergroups have been unambiguously recognised (alveolates, fungi plus microsporidians plus metazoans, and red algae plus green plants). A very long, unbroken branch separates eukaryotes from prokaryotes. This long branch does not mean that no eukaryotic lineages have arisen during this vast evolutionary time but they most probably became extinct (represented by dotted lines). Putative undetected, surviving ancient lineages are indicated by question marks. The solid arrowhead indicates the unique primary plastid acquisition, whereas open arrowheads indicate secondary photosynthetic plastid acquisitions.

At any rate, the long evolutionary distance separating the last common ancestor of extant eukaryotes from the first eukaryote (Figure 1) suggests that the latter organism might have been a very different organism from the former, itself being different from contemporary eukaryotes. In addition, as the problem of the rooting of the universal tree of life is still unsettled (as are the relationships between the three domains of life) [32], the intriguing question of the origin of eukaryotes remains widely open. This is an active and strongly debated field, an example being that the authors of this article advocate for entirely different hypotheses. Recent proposals range from claims for a eukaryotic ancestor for all life to symbiosis between archaea and bacteria forming the primeval eukaryotes. The first are based both on the apparent antiquity of some eukaryotic RNA-related molecular mechanisms [53] and on the reinterpretation of molecular phylogeny data in the light of the artefacts that have most likely biased the reconstruction of universal trees [54]. The second are hypotheses that try to account for the apparently composite nature of eukaryotes genomes — informational genes more similar to archaeal homologues and operational genes more similar to bacterial homologues — by looking for the origin of eukaryotes in metabolic symbioses between methanogenic archaea and various kinds of bacteria [55,56].

Update

Two recent works have provided good evidence in favor of a clade grouping Alveolata and Heterokonta. The first [57•] is based on the combined analysis of four genes (encoding α- and β-tubulins, actin and EF-1α) for a large number of species (61). The second is based on a specific gene duplication of the cytosolic GAPDH, with the corresponding proteins being exported into the chloroplasts (NM Fast *et al*., personal communication). This new super-group is especially interesting as it most likely reduces the number of secondary chloroplastic endosymbioses, in agreement with a recent hypothesis of Cavalier-Smith [58].

Acknowledgements

We are grateful to Philippe Lopez and Miklós Müller for their critical reading of the manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Patterson DJ: **The diversity of eukaryotes.** *Am Nat* 1999, **154**:S96-S124.

2. Sogin ML: **Early evolution and the origin of eukaryotes.** *Curr Opin Genet Dev* 1991, **1**:457-463.

3. Cavalier-Smith T: **Eukaryotes with no mitochondria.** *Nature* 1987, **326**:332-333.

4. Embley TM, Hirt RP: **Early branching eukaryotes?** *Curr Opin Genet Dev* 1998, **8**:624-629.

5. • Lang BF, Gray MW, Burger G: **Mitochondrial genome evolution and the origin of eukaryotes.** *Annu Rev Genet* 1999, **33**:351-397.
A clear review of mitochondrial genome evolution in light of recent sequencings, especially of the remarkably eubacteria-like mtDNA in the jakobid *Reclinomonas*.

6. Sogin ML: **Organelle origins: energy-producing symbionts in early eukaryotes?** *Curr Biol* 1997, **7**:R315-R317.

7. Hashimoto T, Sanchez LB, Shirakura T, Muller M, Hasegawa M: **Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny.** *Proc Natl Acad Sci USA* 1998, **95**:6860-6865.

8. Tovar J, Fischer A, Clark CG: **The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*.** *Mol Microbiol* 1999, **32**:1013-1021.

9. Mai Z, Ghosh S, Frisardi M, Rosenthal B, Rogers R, Samuelson J: **Hsp60 is targeted to a cryptic mitochondrion-derived organelle ('crypton') in the microaerophilic protozoan parasite *Entamoeba histolytica*.** *Mol Cell Biol* 1999, **19**:2198-2205.

10. • Akhmanova A, Voncken F, van Alen T, van Hoek A, Boxma B, Vogels G, Veenhuis M, Hackstein JH: **A hydrogenosome with a genome.** *Nature* 1998, **396**:527-528.
Evidence for the presence of a hydrogenosome-associated genome in the anaerobic ciliate *Nyctotherus ovalis*. SSU rRNA clusters with mitochondrial SSU rRNA and hydrogenosomal-type hydrogenase displays similarities with subunits of the respiratory chain complex I of mitochondria, showing that the hydrogenosome may have evolved from a mitochondrion during adaptation to anaerobic environments.

11. Logsdon J: **The recent origins of spliceosomal introns revisited.** *Curr Opin Genet Dev* 1998, **8**:637-648.

12. Breckenridge DG, Watanabe Y, Greenwood SJ, Gray MW, Schnare MN: **U1 small nuclear RNA and spliceosomal introns in *Euglena gracilis*.** *Proc Natl Acad Sci USA* 1999, **96**:852-856.

13. Biderre C, Metenier G, Vivares CP: **A small spliceosomal-type intron occurs in a ribosomal protein gene of the microsporidian *Encephalitozoon cuniculi*.** *Mol Biochem Parasitol* 1998, **94**:283-286.

14. Fast NM, Doolittle WF: ***Trichomonas vaginalis* possesses a gene encoding the essential spliceosomal component, PRP8.** *Mol Biochem Parasitol* 1999, **99**:275-278.

15. Philippe H, Adoutte A: **The molecular phylogeny of Eukaryota: solid facts and uncertainties.** In *Evolutionary Relationships Among Protozoa*. Edited by Coombs G, Vickerman K, Sleigh M, Warren A. London: Chapman & Hall; 1998:25-56.

16. Keeling PJ, Doolittle WF: **Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family.** *Mol Biol Evol* 1996, **13**:1297-1305.

17. Edlind TD, Li J, Visvesvara GS, Vodkin MH, McLaughlin GL, Katiyar SK: **Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa.** *Mol Phylogenet Evol* 1996, **5**:359-367.

18. Philippe H, Laurent J: **How good are deep phylogenetic trees?** *Curr Opin Genet Dev* 1998, **8**:616-623.

19. Felsenstein J: **Cases in which parsimony or compatibility methods will be positively misleading.** *Syst Zool* 1978, **27**:401-410.

20. Roger AJ, Sandblom O, Doolittle WF, Philippe H: **An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes.** *Mol Biol Evol* 1999, **16**:218-233.

21. Moreira D, Le Guyader H, Philippe H: **Unusually high evolutionary rate of the elongation factor 1 alpha genes from the Ciliophora and its impact on the phylogeny of eukaryotes.** *Mol Biol Evol* 1999, **16**:234-245.

22. Tourasse NJ, Gouy M: **Evolutionary relationships between protist phyla constructed from LSU rRNAs accounting for unequal rates of substitution among sites.** In *Evolutionary Relationships Among Protozoa*. Edited by Coombs G, Vickerman K, Sleigh M, Warren A. London: Chapman & Hall; 1998:57-75.

23. Peyretaillade E, Biderre C, Peyret P, Duffieux F, Metenier G, Gouy M, Michot B, Vivares CP: **Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core.** *Nucleic Acids Res* 1998, **26**:3513-3520.

24. Van de Peer Y, Ben Ali A, Meyer A: **Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi.** *Gene* 2000, **246**:1-8.

25. • Silberman JD, Clark CG, Diamond LS, Sogin ML: **Phylogeny of the genera *Entamoeba* and *Endolimax* as deduced from small-subunit ribosomal RNA sequences.** *Mol Biol Evol* 1999, **16**:1740-1751.
The monophyly and phylogenetic position of Entamoebidae in the rRNA tree is examined using different models of sequence evolution. The fast-evolving lineages

corresponding to the early branching eukaryotes when rate across site is not taken into account are rather late-emerging species when correction is applied.

26. Hirt RP, Logsdon JM Jr, Healy B, Dorey MW, Doolittle WF, Embley TM:
• **Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins.** *Proc Natl Acad Sci USA* 1999, **96**:580-585.
A report that provides for the second time, after tubulin analyses, strong support for the Microsporidia/Fungi clade, using the largest subunit of the RNA polymerase II and disregarding fast-evolving or invariant sites. This result prompts the authors to re-analyse the elongation factor phylogeny, which shows early branching of amitochondriate eukaryotes, by reducing site-rate heterogeneity and choosing a pertinent outgroup. In that case, none of these groups displays an early divergence among eukaryotes.

27. Fitch WM, Markowitz E: **An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution.** *Biochem Genet* 1970, **4**:579-593.

28. Philippe H, Germot A: **Phylogeny of eukaryotes based on**
•• **ribosomal RNA: long-branch attraction and models of sequence evolution.** *Mol Biol Evol* 2000, **17**:830-834.
We investigate the robustness of the eukaryotic phylogeny based on rRNA with consideration of different models of sequence evolution. We show that a revised phylogeny – on the basis of combination of results given by different protein markers – could not be rejected whan a more realistic model of sequence evolution is taken into account.

29. Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J,
•• Moreira D, Müller M, Le Guyader H: **Early branching or fast evolving eukaryotes? An answer based on slowly evolving positions.** *Philos Trans R Soc Lond B Biol Sci* 2000, **267**:1213-1221.
Use of a recent method, 'Slow-Fast' [32], to point out the fast-evolving species. In the case of eukaryotic phylogeny, re-examination of rRNA and protein data by this method argue in favour of the 'big bang' hypothesis.

30. Germot A, Philippe H: **Critical analysis of eukaryotic phylogeny:**
• **a case study based on the HSP70 family.** *J Eukaryot Microbiol* 1999, **46**:116-124.
An analysis of the eukaryotic phylogeny and the position of trichomonads on the basis of different Hsp70s and their combination. The rapid diversification of eukaryotes is proposed to explain the lack of resolution of the phylogenetic reconstruction. This is the first time that the number of variable positions peculiar to a lineage (trichomonads) is shown to be the cause of its artefactual placement at the base of eukaryotes or as a sister-group of fast-evolving species.

31. Stiller J, Hall B: **Long-branch attraction and the rDNA model of early eukaryotic evolution.** *Mol Biol Evol* 1999, **16**:1270-1279.

32. Brinkmann H, Philippe H: **Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies.** *Mol Biol Evol* 1999, **16**:817-825.

33. Keeling PJ, Luker MA, Palmer JD: **Evidence from beta-tubulin**
•• **phylogeny that microsporidia evolved from within the fungi.** *Mol Biol Evol* 2000, **17**:23-31.
The increased sampling of fungal and microsporidian β tubulin sequences reveals the membership of Microsporidia to the Fungi. The first strong evidence that Microsporidia evolved from a fungus after the divergence of chytrids.

34. Baldauf SL: **A search for the origins of animals and fungi: comparing**
• **and combining molecular data.** *Am Nat* 1999, **154**:S178-S188.
A detailed review about the origin of the animal–fungi clade on the basis of molecular data. The author adds original work founded on combined analysis of protein data (EF1α, actin, α and β tubulins) that supports the close relationship of Mycetozoa (*Dictyostelium* and *Physarum*) and Opisthokonta (animals and fungi).

35. Pawlowski J, Bolivar I, Fahrni JF, De Vargas C, Bowser SS: **Molecular evidence that *Reticulomyxa filosa* is a freshwater naked foraminifer.** *J Eukaryot Microbiol* 1999, **46**:612-617.

36. Moreira D, Le Guyader H, Philippe H: **The origin of red algae:**
• **implications for the evolution of chloroplasts.** *Nature* 2000, **405**:69-72.
Analyses of a single-gene (*Elongation factor 2*) and a multi-gene (13 proteins) provide the first significant evidence from nuclear genes for the sisterhood of red algae and green plants. Note how the fusion of sequences that individually do not give strong support can be used to analyse specific phylogeny problems.

37. Hannaert V, Brinkmann H, Nowitzki U, Lee JA, Albert M-A, Sensen CW,
• Gaasterland T, Müller M, Michels P, Martin W: **Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi*, and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway.** *Mol Biol Evol* 2000, **17**:989-1000.
Molecular and phylogenetic analyses of newly sequenced cytosolic enolase from three protists. In the case of *Euglena*, another enolase located in the chloroplast was identified. The corresponding gene arose after gene duplication to the cytosolic form. The authors also emphasise the use of indels in phylogeny, especially when plylogenetic reconstructions are not conclusive, but they also warn against their misinterpretation in the absence of sufficient data.

38. Keeling PJ, Palmer JD: **Parabasalian flagellates are ancient eukaryotes.** *Nature* 2000, **405**:635-637.

39. Roger AJ, Morrison HG, Sogin ML: **Primary structure and phylogenetic relationships of a malate dehydrogenase gene from *Giardia lamblia*.** *J Mol Evol* 1999, **48**:750-755.

40. Wu G, Fiser A, ter Kuile B, Sali A, Muller M: **Convergent evolution of *Trichomonas vaginalis* lactate dehydrogenase from malate dehydrogenase.** *Proc Natl Acad Sci USA* 1999, **96**:6285-6290.

41. Stiller JW, Hall BD: **The origin of red algae: implications for plastid evolution.** *Proc Natl Acad Sci USA* 1997, **94**:4520-4525.

42. Stiller JW, Duffield EC, Hall BD: **Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II.** *Proc Natl Acad Sci USA* 1998, **95**:11769-11774.

43. Fast NM, Logsdon JM Jr, Doolittle WF: **Phylogenetic analysis of the TATA box binding protein (TBP) gene from *Nosema locustae*: evidence for a microsporidia-fungi relationship and spliceosomal intron loss.** *Mol Biol Evol* 1999, **16**:1415-1419.

44. Viscogliosi E, Muller M: **Phylogenetic relationships of the glycolytic enzyme, glyceraldehyde-3-phosphate dehydrogenase, from parabasalid flagellates.** *J Mol Evol* 1998, **47**:190-199.

45. Naylor GJ, Brown WM: **Structural biology and phylogenetic estimation.** *Nature* 1997, **388**:527-528.

46. Burger G, Saint-Louis D, Gray MW, Lang BF: **Complete sequence of**
• **the mitochondrial DNA of the red alga *Porphyra purpurea*. Cyanobacterial introns and shared ancestry of red and green algae.** *Plant Cell* 1999, **11**:1675-1694.
The study of the 36,753 base pair long mitochondrial genome of the red algae *Porphyra purpurea* reveals that red algal mitochondrial genomes are unusually uniform and provides strong evidence that red algae share a common ancestry with green plants, both from the analysis of individual or combined genes.

47. Taylor FJR: **Problem in the development of an explicit hypothetical phylogeny of the Lower Eukaryotes.** *Biosystems* 1978, **10**:67-89.

48. Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW: **An ancestral mitochondrial DNA resemb ling a eubacterial genome in miniature.** *Nature* 1997, **387**:493-497.

49. Culligan KM, Meyer-Gauen G, Lyons-Weiler J, Hays JB: **Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins.** *Nucleic Acids Res* 2000, **28**:463-471.

50. Lockhart PJ, Huson D, Maier U, Fraunholz MJ, Van De Peer Y, Barbrook AC, Howe CJ, Steel MA: **How molecules evolve in Eubacteria.** *Mol Biol Evol* 2000, **17**:835-838.

51. Inagaki Y, Doolittle WF: **Evolution of the eukaryotic translation termination system: origins of release factors.** *Mol Biol Evol* 2000, **17**:882-889.

52. Gogarten JP, Hilario E, Olendzenski L: **Gene duplications and horizontal gene transfer during early evolution.** In *Evolution of Microbial Life*. Edited by Roberts DM, Sharp P, Alderson G, Collins M. Cambridge University Press; 1996:109-126.

53. Jeffares DC, Poole AM, Penny D: **Relics from the RNA world.** *J Mol Evol* 1998, **46**:18-36.

54. Forterre P, Philippe H: **Where is the root of the universal tree of life?** *Bioessays* 1999, **21**:871-879.

55. Martin W, Müller M: **The hydrogen hypothesis for the first eukaryote.** *Nature* 1998, **392**:37-41.

56. Moreira D, López-García P: **Symbiosis between methanogenic archaea and delta-Proteobacteria as the origin of eukaryotes: the syntrophic hypothesis.** *J Mol Evol* 1998, **47**:517-530.

57. Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF: **Kingdom-level**
• **phylogeny of Eukaryotes based on combined protein data.** *Science* 2000, in press.
This is the first gene fusion for which there is a rich species sampling, obtained to the detriment of number of positions (only 1528). Several interesting groupings are found (such as Fungi+Microsporidia, Alveolata+Heterokonta or Euglenozoa+Percolozoa), albeit often with low statistical support. Because of the high number of paralogs, the detection of orthologs is, however, difficult - which can weaken the analysis.

58. Cavalier-Smith T: **Membrane heredity and early chloroplast evolution.** *Trends Plant Sci* 2000, **5**:174-182.