REVIEWS

Eukaryotic evolution, changes and challenges

T. Martin Embley[1] & William Martin[2]

The idea that some eukaryotes primitively lacked mitochondria and were true intermediates in the prokaryote-to-eukaryote transition was an exciting prospect. It spawned major advances in understanding anaerobic and parasitic eukaryotes and those with previously overlooked mitochondria. But the evolutionary gap between prokaryotes and eukaryotes is now deeper, and the nature of the host that acquired the mitochondrion more obscure, than ever before.

New findings have profoundly changed the ways in which we view early eukaryotic evolution, the composition of major groups, and the relationships among them. The changes have been driven by a flood of sequence data combined with improved—but by no means consummate—computational methods of phylogenetic inference. Various lineages of oxygen-shunning or parasitic eukaryotes were once thought to lack mitochondria and to have diverged before the mitochondrial endosymbiotic event. Such key lineages, which are salient to traditional concepts about eukaryote evolution, include the diplomonads (for example, *Giardia*), trichomonads (for example, *Trichomonas*) and microsporidia (for example, *Vairimorpha*). From today's perspective, many key groups have been regrouped in unexpected ways, and aerobic and anaerobic eukaryotes intermingle throughout the unfolding tree. Mitochondria in previously unknown biochemical manifestations seem to be universal among eukaryotes, modifying our views about the nature of the earliest eukaryotic cells and testifying to the importance of endosymbiosis in eukaryotic evolution. These advances have freed the field to consider new hypotheses for eukaryogenesis and to weigh these, and earlier theories, against the molecular record preserved in genomes. Newer findings even call into question the very notion of a 'tree' as an adequate metaphor to describe the relationships among genomes. Placing eukaryotic evolution within a time frame and ancient ecological context is still problematic owing to the vagaries of the molecular clock and the paucity of Proterozoic fossil eukaryotes that can be clearly assigned to contemporary groups. Although the broader contours of the eukaryote phylogenetic tree are emerging from genomic studies, the details of its deepest branches, and its root, remain uncertain.

The universal tree and early-branching eukaryotic lineages

The universal tree based on small-subunit (SSU) ribosomal RNA[1] provided a first overarching view of the relationships between the different types of cellular life. The relationships among eukaryotes recovered from rRNA[2], backed up by trees of translation elongation factor (EF) proteins[3], provided what seemed to be a consistent, and hence compelling, picture (Fig. 1). The three protozoa at the base of these trees (*Giardia*, *Trichomonas* and *Vairimorpha*), along with *Entamoeba* and its relatives, were seen as members of an ultrastructurally simple, paraphyletic group of eukaryotes called the Archezoa[4]. Archezoa were thought to primitively lack mitochondria, having split from the main trunk of the eukaryotic tree before the mitochondrial endosymbiosis: all other eukaryotes contain mitochondria because they diverged after this singular symbiotic event[5]. Therefore, Archezoa were interpreted as contemporary descendants of a phagotrophic, nucleated, amitochondriate cell lineage that included the host for the mitochondrial endosymbiont[6]. The apparent agreement between molecules and morphology depicted the relative timing of the mitochondrial endosymbiosis (Fig. 1) as a crucial, but not ancestral, event in eukaryote phylogeny.

Chinks in the consensus

Mitochondrial genomes studied so far encode less than 70 of the proteins that mitochondria need to function[5]; most mitochondrial proteins are encoded by the nuclear genome and are targeted to

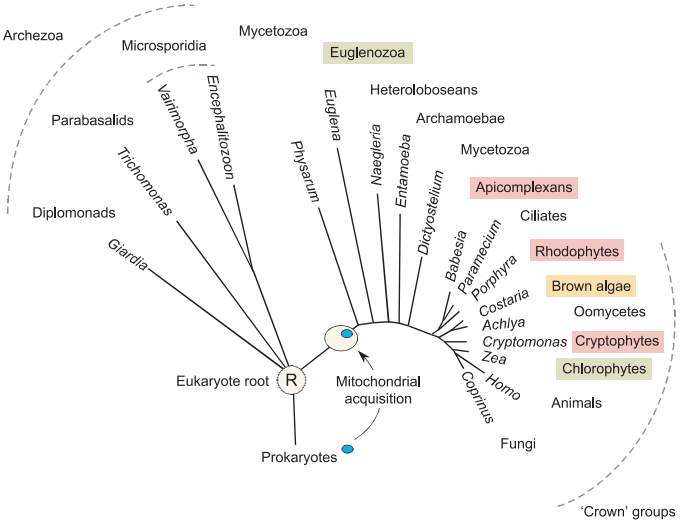

Figure 1 | The general outline of eukaryote evolution provided by rooted rRNA trees. The tree has been redrawn and modified from ref. 92. Until recently, lineages branching near the root were thought to primitively lack mitochondria and were termed Archezoa[4]. Exactly which archezoans branched first is not clearly resolved by rRNA data[2], hence the polytomy (more than two branches from the same node) involving diplomonads, parabasalids and microsporidia at the root. Plastid-bearing lineages are indicated in colours approximating their respective pigmentation. Lineages furthest away from the root, including those with multicellularity, were thought to be the latest-branching forms and were sometimes misleadingly (see ref. 60) called the 'crown' groups.

[1]School of Biology, The Devonshire Building, University of Newcastle upon Tyne, Newcastle NE1 7RU, UK. [2]Institute of Botany III, University of Düsseldorf, D-40225 Düsseldorf, Germany.

mitochondria using a protein import machinery that is specific to this organelle[7]. The mitochondrial endosymbiont is thought to have belonged to the α-proteobacteria, because some genes and proteins still encoded by the mitochondrial genome branch in molecular trees among homologues from this group[5,8]. Some mitochondrial proteins, such as the 60- and 70-kDa heat shock proteins (Hsp60, Hsp70), also branch among α-proteobacterial homologues, but the genes are encoded by the host nuclear genome. This is readily explained by a corollary to endosymbiotic theory called endosymbiotic gene transfer[9]: during the course of mitochondrial genome reduction, genes were transferred from the endosymbiont's genome to the host's chromosomes, but the encoded proteins were re-imported into the organelle where they originally functioned. With the caveat that gene origin and protein localization do not always correspond[9], any nuclear-encoded protein that functions in mitochondria and clusters with α-proteobacterial homologues is most simply explained as originating from the mitochondrion in this manner.

By that reasoning[10], the discovery of mitochondrial Hsp60 in *E. histolytica* was taken as evidence that its ancestors harboured mitochondria. A flood of similar reports on mitochondrial Hsp60 and Hsp70 from all key groups of Archezoa ensued[11], suggesting that their common ancestor also contained mitochondria. At face value, those findings falsified the central prediction of the archezoan concept. However, suggestions were offered that lateral gene transfer (LGT) in a context not involving mitochondria could also account for the data. But that explanation, apart from being convoluted, now seems unnecessary: the organisms once named Archezoa for lack of mitochondria not only have mitochondrial-derived proteins, they have the corresponding double-membrane-bounded organelles as well.

Mitochondria in multiple guises

The former archezoans are mostly anaerobes, avoiding all but a trace of oxygen, and like many anaerobes, including various ciliates and fungi that were never grouped within the Archezoa, they are now known to harbour derived mitochondrial organelles—hydrogenosomes and mitosomes. These organelles all share one or more traits in common with mitochondria (Fig. 2), but no traits common to them all, apart from the double membrane and conserved mechanisms of protein import, have been identified so far. Mitochondria typically—but not always (the *Cryptosporidium* mitochondrion lacks DNA[12])—possess a genome that encodes components involved in oxidative phosphorylation[5]. With one notable exception[13], all hydrogenosomes

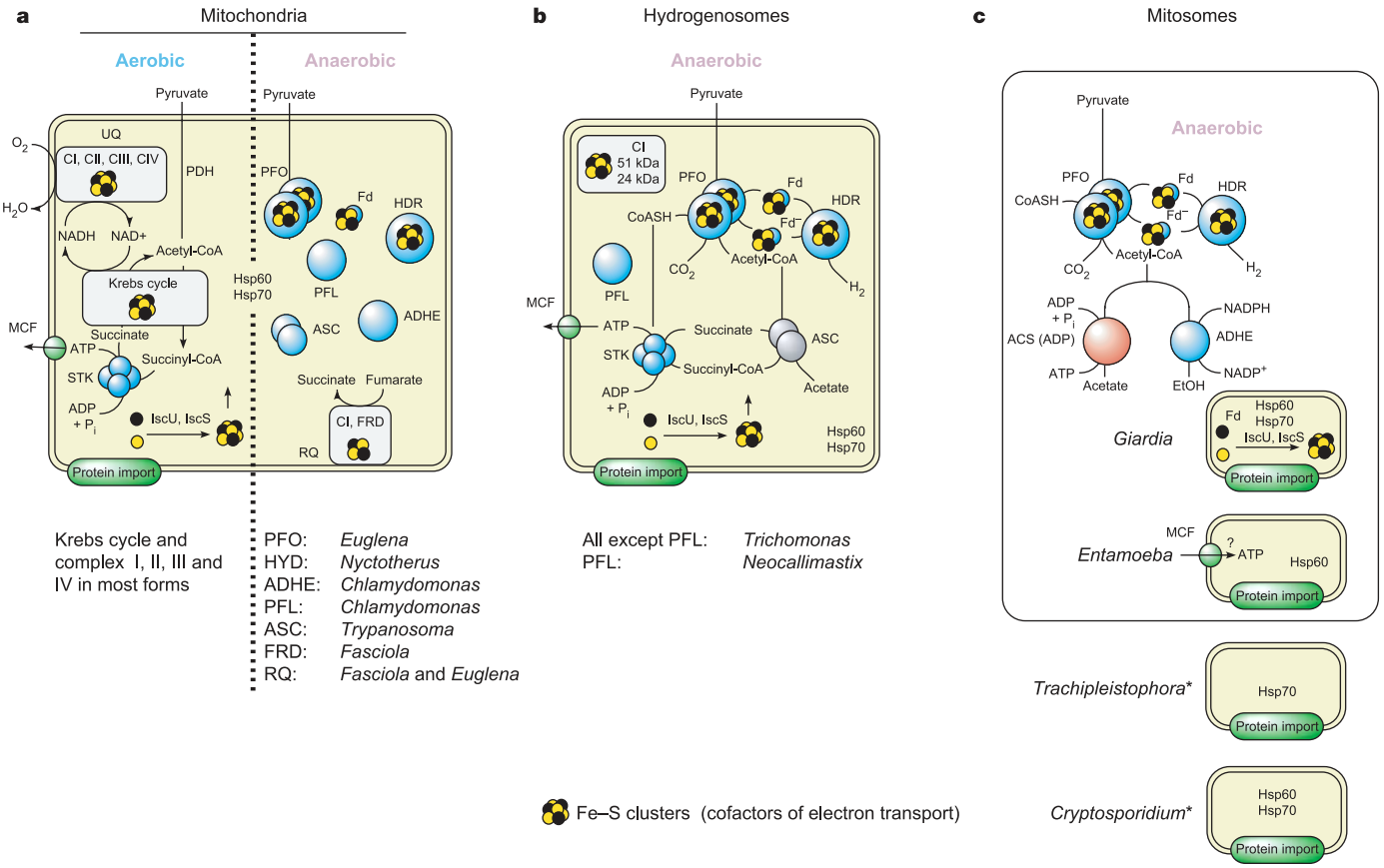

Figure 2 | Enzymes and pathways found in various manifestations of mitochondria. Proteins sharing more sequence similarity to eubacterial than to archaebacterial homologues are shaded blue; those with converse similarity pattern are shaded red; those whose presence is based only on biochemical evidence are shaded grey; those lacking clearly homologous counterparts in prokaryotes are shaded green. **a**, Schematic summary of salient biochemical functions in mitochondria[5,88], including some anaerobic forms[16,17]. **b**, Schematic summary of salient biochemical functions in hydrogenosomes[14,19]. **c**, Schematic summary of available findings for mitosomes and 'remnant' mitochondria[32–34,93]. The asterisk next to the *Trachipleistophora* and *Cryptosporidium* mitosomes denotes that these organisms are not anaerobes in the sense that they do not inhabit O_2-poor niches, but that their ATP supply is apparently O_2-independent. UQ, ubiquinone; CI, mitochondrial complex I (and II, III and IV, respectively); NAD, nicotinamide adenine dinucleotide; MCF, mitochondrial carrier family protein transporting ADP and ATP; STK, succinate thiokinase; PFO, pyruvate:ferredoxin oxidoreductase; PDH, pyruvate dehydrogenase; CoA, coenzyme A; Fd, ferredoxin; HDR, iron-only hydrogenase; PFL, pyruvate:formate lyase; ASC, acetate-succinate CoA transferase; ADHE, bi-functional alcohol acetaldehyde dehydrogenase; FRD, fumarate reductase; RQ, rhodoquinone; Hsp, heat shock protein; IscU, iron–sulphur cluster assembly scaffold protein; IscS; cysteine desulphurase; ACS (ADP), acetyl-CoA synthase (ADP-forming).

and mitosomes studied so far lack a genome. The organisms in which they have been studied generate ATP by fermentations involving substrate-level phosphorylations, rather than through chemiosmosis involving an F_1/F_0-type ATPase[12,14,15]. *Entamoeba*, *Giardia* and *Trichomonas* live in habitats too oxygen-poor to support aerobic respiration[14], while others, like *Cryptosporidium* and microsporidia have drastically reduced their metabolic capacities during adaptation to their lifestyles as intracellular parasites[12,15].

Between aerobic mitochondria, which use oxygen as the terminal electron acceptor of ATP-producing oxidations, and *Nyctotherus* hydrogenosomes, which (while retaining a mitochondrial genome) use protons instead of oxygen[13], there are a variety of other anaerobically functioning mitochondria. They occur in protists such as *Euglena*, but also in multicellular animals such as *Fasciola* and *Ascaris*, which typically excrete acetate, propionate or succinate, instead of H_2O or H_2, as their major metabolic end-products[16,17]. Hence, mitochondria, hydrogenosomes and mitosomes are viewed most simply as variations on a single theme, one that fits neatly within the framework provided by classical evolutionary theory[18]. They are evolutionary homologues that share similarities because of common ancestry, but—like forelimbs in vertebrates—differ substantially in form and function across lineages owing to descent with modification.

Hydrogen-producing mitochondria

Hydrogenosomes oxidize pyruvate to H_2, CO_2 and acetate, making ATP by substrate-level phosphorylation[19] that they export to the cytosol using a mitochondrial-type ADP/ATP carrier[20,21]. They have been identified in trichomonads, chytridiomycetes and ciliates[13,22]; their hydrogen excretion helps to maintain redox balance[14] in these organisms. Important similarities between *Trichomonas* hydrogenosomes and mitochondria include the use of common protein import pathways[23], conserved mechanisms of iron–sulphur-cluster assembly[24], conserved mechanisms of NAD^+ regeneration[25], and conservation of a canonical ATP-producing enzyme of the mitochondrial Krebs cycle—succinate thiokinase[26]. On the basis of electron microscopy and ecology, additional, and diverse, eukaryotic lineages are currently suspected to contain hydrogenosomes[27,28], but hydrogen production—the defining characteristic of hydrogenosomes[19]—by those organelles has not yet been shown.

In contrast to most mitochondria, hydrogenosomes typically contain pyruvate:ferredoxin oxidoreductase (PFO) and iron [Fe] hydrogenase. Common among anaerobic bacteria, these enzymes prompted the early suggestion that trichomonad hydrogenosomes arose from a *Clostridium*-like endosymbiont[29]. In a recent rekindling of that idea[30,31], trichomonad hydrogenosomes were suggested to be hybrid organelles, derived from an endosymbiotic anaerobic bacterium (the source of PFO and hydrogenase genes), a failed mitochondrial endosymbiosis (the source of nuclear genes for mitochondrial Hsp60 and Hsp70), plus LGT from a mitochondrially related (but non-mitochondrial) donor (the source of NADH dehydrogenase). However, independent work suggested a mitochondrial, rather than hybrid, origin of the *Trichomonas* NADH dehydrogenase[25]. Furthermore, the hybrid hypothesis fails to account for the presence of [Fe]-hydrogenase homologues in algal chloroplasts, PFO homologues in *Euglena* mitochondria, or the presence of either enzyme and hydrogenosomes in other eukaryotic lineages[25]; hence, a single common ancestry of mitochondria and hydrogenosomes sufficiently accounts for current observations.

Mitochondria reduced to bare bones

Mitosomes were discovered in *Entamoeba*[32] as mitochondrion-derived organelles that have undergone more evolutionary reduction than hydrogenosomes. They are also found in *Giardia*[33] and microsporidia[34]. Mitosomes seem to have no direct role in ATP synthesis because, so far, they have been found only among eukaryotes whose core ATP synthesis occurs in the cytosol[14] or among energy parasites[15]. Mitosomes import proteins in a mitochondrial-like manner[35–37], and *Giardia* mitosomes contain two mitochondrial proteins of Fe–S cluster assembly—cysteine desulphurase (IscS) and iron-binding protein (IscU)[33]. Fe–S clusters are essential for life: they are cofactors of electron transfer, catalysis, redox sensing and ribosome biogenesis in eukaryotes[38]. Fe–S cluster assembly is an essential function of yeast mitochondria[38] and it has been widely touted as a potential common function for mitochondrial homologues[15,22]. It is the only known function of *Giardia* mitosomes, which, like *Trichomonas* hydrogenosomes[24,37], promote assembly of [2Fe–2S] clusters into apoferredoxin *in vitro*[33]. By contrast, and (so far) uniquely among eukaryotes, *Entamoeba* uses two proteins of non-mitochondrial ancestry for Fe–S cluster assembly[39]; the location of this pathway in *Entamoeba* is currently unknown.

Branch migrations and evolutionary models

The discovery of mitochondrial homologues in *Giardia*, *Trichomonas* and microsporidians, which had been the best candidates for eukaryotes that primitively lacked mitochondria, has pinned the timing of the mitochondrial origin to the ancestor of all eukaryotes studied so far. But that does not mean that the basal position of these groups in the SSU rRNA tree (Fig. 1) and EF trees[3] is necessarily incorrect. That issue hinges on efforts to construct reliable rooted phylogenetic trees depicting ancient eukaryotic relationships: a developing area of research that is fraught with difficulties. The tempo and mode of sequence evolution is far more complicated than is assumed by current mathematical models that are used to make phylogenetic trees[40]. In computer simulations, where the true tree is known, model mis-specification can produce the wrong tree with strong support[41].

Different sites in molecular sequences evolve at different rates, and failure to accommodate this rate variation, something early methods failed to do, can lead to strongly supported but incorrect trees owing to a common problem called 'long-branch-attraction'[42]. This occurs when branches that are long or 'fast evolving', relative to others in the tree, cluster together irrespective of evolutionary relationships. The molecular sequences of *Giardia*, *Trichomonas* and microsporidia often form long branches in trees and thus are particularly prone to this problem[25,43,44]. The traditional models that placed microsporidia deep within trees[2,3] assumed that all sequence sites evolved at the same rate, even though they clearly do not. In these trees, the long-branch microsporidia are next to the long branches of the prokaryotic outgroups. More data and better models have produced trees that agree in placing microsporidia with fungi[45,46], suggesting that the deep position of microsporidia in early trees was indeed an artefact.

The position of *Giardia* and *Trichomonas* sequences at the base of eukaryotic molecular trees is also suspect, given that they also form long branches in the trees that place them in this way, and because other trees and models place them together as an internal branch of a rooted eukaryotic tree[47]. Resolving which position is correct is particularly important, because *Giardia* and *Trichomonas* are still commonly referred to as 'early-branching' eukaryotes. Given the evident uncertainties of such phylogenies, and the importance of the problem, the onus is on those who would persist in calling these species 'early branching' to show that trees placing them deep explain the data significantly better than trees that do not.

The root of the eukaryotic tree

The usual way to root a phylogenetic tree is by reference to an outgroup; the rRNA and EF trees used prokaryotic sequences to root eukaryotes on either the *Giardia*, *Trichomonas* or microsporidia branch (Fig. 1), but these rootings have not proved robust[43–45]. The sequences of outgroups are often highly divergent compared to those of the ingroup, making it difficult to avoid model mis-specification and long-branch-attraction[44,48].

An alternative method of rooting an existing tree is to look for rare

changes in a complex molecular character where the ancestral state can be inferred. This method was used[49] to infer that the root of the eukaryotic tree lies between the animals, fungi and amoebozoa (together called unikonts) on the one side, and plants, algae and most protozoa (bikonts) on the other. In fungi and animals, the genes for dihydrofolate reductase (DHFR) and thymidylate synthase (TS) are separate[44], as they are in prokaryote outgroups; but they are fused in the bikonts sampled so far. Assuming that the fusion occurred only once and that its subsequent fission did not occur at all, the DHFR–TS fusion would be a derived feature uniting bikonts, suggesting that the eukaryote root lies outside this group[49]. The coherence of animals, fungi and various unicellular eukaryotes (together called opisthokonts) is supported by phylogenetic trees and other characters[50]. The presence of a type II myosin in opisthokonts and amoebozoa unites them to form the unikonts[51]. If both unikonts and bikonts are monophyletic groups, and together they encompass extant eukaryotic diversity, then the root of eukaryotes would lie between them.

Placing the eukaryote root between unikonts and bikonts would help to bring order to chaos, if it is correct. However, it assumes that the underlying tree—over which the rooting character is mapped—is known, when in fact the relationships—especially for bikonts and many enigmatic protistan lineages[52]—remain uncertain. The rooting also depends upon a single character of unknown stability sampled from only a few species. An additional caveat is that *Giardia* and *Trichomonas* lack both DHFR and TS—parasites relinquish genes of various biosynthetic pathways, stealing the pathway products from their hosts instead. Hence, the missing fusion character does not address their position in the tree.

New hypotheses of eukaryotic relationships

New data and analyses from many laboratories have been used to formulate a number of hypotheses of eukaryotic relationships (Fig. 3) that fundamentally differ from those in the SSU rRNA tree. It is apparent that hydrogenosomes and mitosomes appear on different branches; the absence of traditional mitochondria and presence of a specialized anaerobic phenotype are neither rare nor 'primitive', as once thought. Mitochondria with a genome encoding elements of the respiratory pathway also appear on both sides of the tree (Fig. 3), suggesting that this pathway has been retained since earliest times; although, as modern examples attest[16,17], it need not have always used oxygen as the sole terminal electron acceptor. On the basis of the unfolding tree, it would seem entirely possible—if not likely—that aerobic and anaerobic eukaryotes, harbouring mitochondrial homologues of various sorts, have co-existed throughout eukaryote history.

The relationships between major groups of eukaryotes are uncertain because of the lack of agreement between different proteins and different analyses; this uncertainty is depicted as a series of polytomies in Fig. 3. Most groups are still poorly sampled for species and molecular sequences—factors that impede robust resolution[53]. It has been suggested[54] that the lack of resolution in deeper parts of the eukaryotic tree stems from an evolutionary 'big bang' or rapid radiation for eukaryotes, perhaps driven by the mitochondrial endosymbiosis[54]. However, both theory and computer simulations[40,41] suggest that a lack of resolution at deeper levels is to be expected given sparse data, our assumptions about sequence evolution, and the limitations of current phylogenetic methods. Thus, loss of historical signal provides a simple null hypothesis for the observed lack of resolution in deeper parts of the eukaryotic tree.

More good theories for eukaryotic origins than good data

Eukaryotic cell organization is more complex than prokaryotic, boasting, *inter alia*, a nucleus with its contiguous endoplasmic reticulum, Golgi, flagella with a '9+2' pattern of microtubule arrangement, and organelles surrounded by double membranes. There are no obvious precursor structures known among prokaryotes from which such attributes could be derived, and no intermediate cell types known that would guide a gradual evolutionary inference between the prokaryotic and eukaryotic state. Accordingly, thoughts on the topic are diverse, and new suggestions appear faster than old ones can be tested.

Biologists have traditionally derived the complex eukaryotic state from the simpler prokaryotic one. In recent years, even that has been

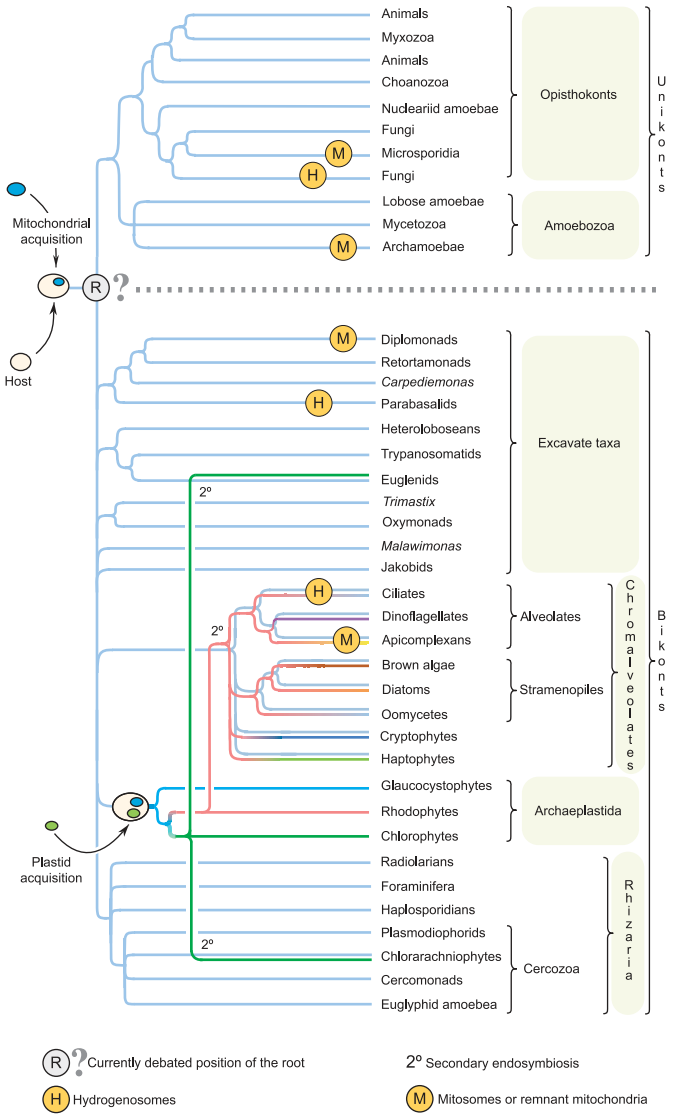

Figure 3 | Schematic tree of newer hypotheses for phylogenetic relationships among major groups of eukaryotes. The composite tree is based on work from many different laboratories and is summarized elswhere[52]; no single data set supports all branches. Polytomies indicate uncertainty in the branching order between major groups. The naming of groups follows current popular usage[52,60]. The current debate that the root of the tree may split eukaryotes into bikonts and unikonts is discussed in the text. Lineages containing species with comparatively well-studied hydrogenosomes (H) or mitosomes (M) are labelled. The depicted distribution of hydrogenosomes and mitosomes is almost certainly conservative, as relatively few anaerobic or parasitic microbial eukaryotes have been studied in sufficient detail to characterize their organelles. The strict coevolution of host nuclear and algal nuclear plus plastid genomes within the confines of a single cell in the wake of secondary endosymbiosis (2°), irrespective of whether or not the secondary nucleus or plastid has persisted as a separate compartment, is indicated by doubled branches. Diversity of pigmentation among photosynthetic eukaryote lineages is symbolized by different coloured branches.

called into question, as some phylogenies have suggested that prokaryotes might be derived from eukaryotes[55]. However, the ubiquity of mitochondrial homologues represents a strong argument that clearly polarizes the prokaryote-to-eukaryote transition: because the common ancestor of contemporary eukaryotes contained a mitochondrial endosymbiont that originated from within the proteobacterial lineage, we can confidently infer that prokaryotes arose and diversified before contemporary eukaryotes—the only ones whose origin requires explanation—did. This view is consistent with microfossil and biogeochemical evidence[56].

Current ideas on the origin of eukaryotes fall into two general classes: those that derive a nucleus-bearing but amitochondriate cell first, followed by the origin of mitochondria in a eukaryotic host[57–61] (Fig. 4a–d), and those that derive the origin of mitochondria in a prokaryotic host, followed by the origin of eukaryotic-specific features[62–64] (Fig. 4e–g). Models that derive a nucleated but amitochondriate cell as an intermediate (Fig. 4a–d) have suffered a substantial blow with the demise of Archezoa. Models that do not entail amitochondriate intermediates have in common that the host assumed to have acquired the mitochondrion was an archaebacterium not a eukaryote; hence, the steep organizational grade between prokaryotes and eukaryotes follows in the wake of radical chimaerism involving mitochondrial origins (Fig. 4e–g). A criticism facing all 'archaebacterial host' models is that phagotrophy (the ability to engulf bacteria as food particles) was once seen as an absolute prerequisite for mitochondrial origins[60]. This argument has lost some of its strength with the discovery of symbioses where one prokaryote lives inside another, non-phagocytotic prokaryote[65].

The elusive informational ancestor

With the exception of the neomuran hypothesis, which views both eukaryotes and archaebacteria as descendants of Gram-positive eubacteria[60,61] (Fig. 4d), most current theories for eukaryotic origins (Fig. 4) posit the involvement of an archaebacterium in that process. The archaebacterial link to eukaryote origins was first inferred from shared immunological and biochemical similarities of their DNA-dependent RNA polymerases[66]. Tree-based studies of entire genomes[67,68] extended this observation: most eukaryotic genes for replication, transcription and translation (informational genes) are related to archaebacterial homologues, while those encoding biosynthetic and metabolism functions (operational genes) are usually related to eubacterial homologues[8,67,68].

The rooted SSU rRNA tree[1] depicts eukaryotes and archaebacteria as sister groups, as in the neomuran (Fig. 4d) hypothesis[60,61]. By

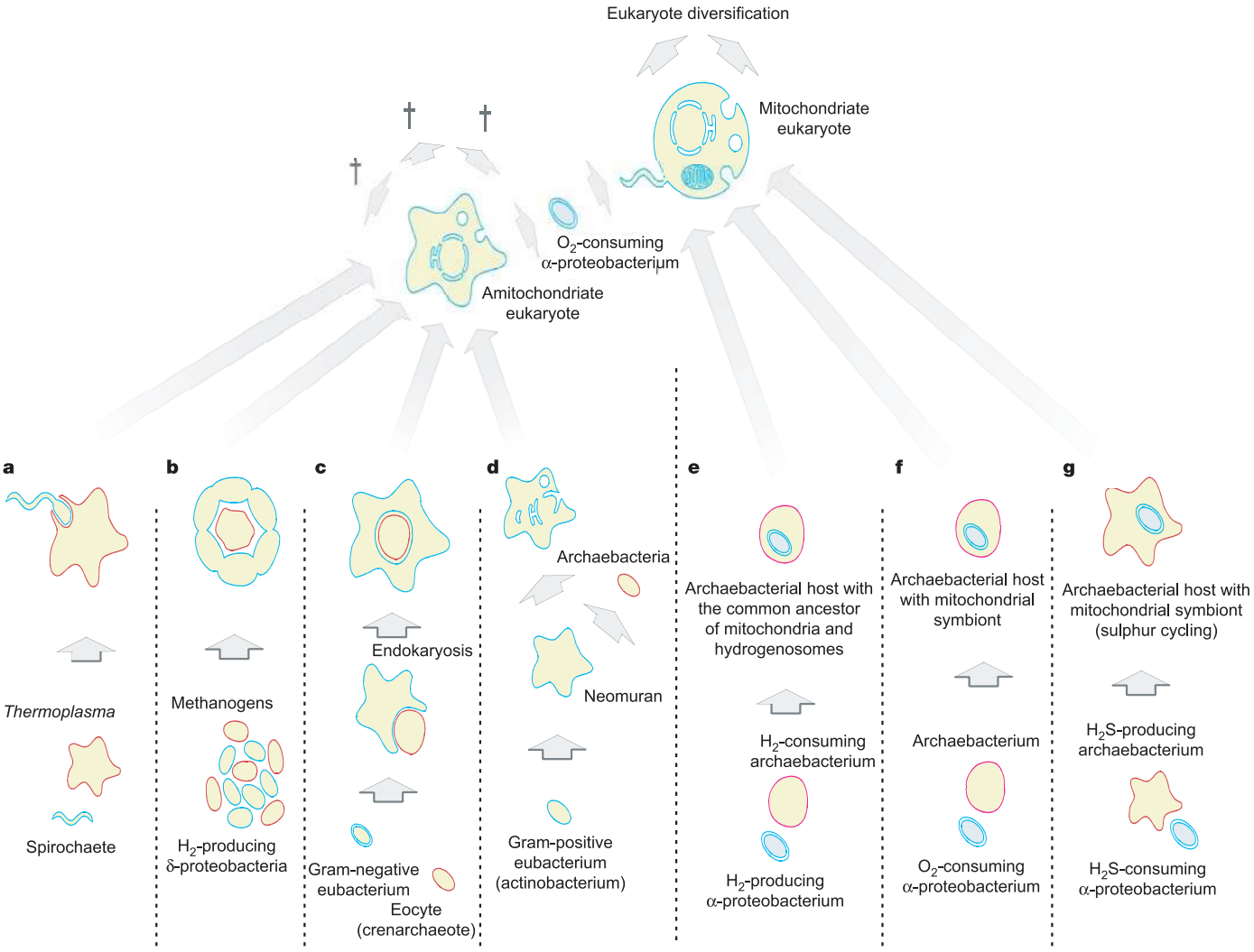

Figure 4 | Models for eukaryote origins that are, in principle, testable with genome data. a–d, Models that propose the origin of a nucleus-bearing but amitochondriate cell first, followed by the acquisition of mitochondria in a eukaryotic host. **e–g**, Models that propose the origin of mitochondria in a prokaryotic host, followed by the acquisition of eukaryotic-specific features. Panels **a–g** are redrawn from refs 57 (**a**), 58 (**b**), 59 (**c**), 60 and 61 (**d**), 62 (**e**), 63 (**f**) and 64 (**g**). The relevant microbial players in each model are labelled. Archaebacterial and eubacterial lipid membranes are indicated in red and blue, respectively.

contrast, the eocyte (Fig. 4c) hypothesis[69,70] proposes that eukaryotic informational genes originate from a specific lineage of archaebacteria called the eocytes, a group synonymous with the Crenarchaeota[1]. In the eocyte tree, the eukaryotic genetic machinery is descended from within the archaebacteria. Although the rooted rRNA tree is vastly more visible to non-specialists, published data are equivocal: for every analysis of a eukaryotic informational gene that recovers the neomuran topology, a different analysis of the same molecule(s) has recovered the eocyte tree[70–74], with the latter being favoured by more sophisticated phylogenetic analyses[69,73,74] and by a shared amino-acid insertion in eocyte and eukaryotic elongation factor 1-α[70].

More recently, genome trees based on shared gene content have been reported. These methods are still new, and—just like gene trees—give different answers from the same data, recovering for informational genes either eukaryote–archaebacterial sisterhood[75], the eocyte tree[76] or a euryarchaeote ancestry[77]. The dichotomy of archaebacteria into euryarchaeotes and eocytes/crenarchaeotes[1] remains unchallenged. The issue, so far unresolved, is the relationship of eukaryotic informational genes to archaebacterial homologues: inheritance from a common progenitor (as in the neomuran hypothesis) or a direct descendant; and if by direct descent, from eocytes/crenarchaeotes like *Sulfolobus*[76], or euryarchaeotes such as *Thermoplasma*[64,78], *Pyrococcus*[77] or methanogens[58,62]. The problems associated with the phylogenetic relationships discussed above are exacerbated at such deep levels, and there is currently neither consensus on this issue nor unambiguous evidence that would clarify it.

The vexing operational majority

Of those eukaryotic genes that have detectable prokaryotic homologues, the majority[67], perhaps as much as 75%[8], are eubacterial and correspond to the operational class. Here arises an interesting point. Although individual analyses of informational genes arrive at fundamentally different interpretations[76,77], no one has yet suggested that more than one archaebacterium participated in eukaryote origins. The situation is quite different with operational genes, where differing phylogenies for individual genes are freely interpreted as evidence for the participation of more than one eubacterial partner. The contribution of gene transfers from the ancestral mitochondrion to nuclear chromosomes has been estimated as anywhere from 136–157 (ref. 77) to ~630 genes[79], depending on the method of analysis. An issue that still requires clarification concerns the origin of thousands of eukaryotic operational genes that are clearly eubacterial, but not specifically α-proteobacterial, in origin[8] (disregarding here the cyanobacterial genes in plants[80]).

There are currently four main theories that attempt to account for those genes. (1) In the neomuran hypothesis (Fig. 4d), they are explained through a direct inheritance from the Gram-positive ancestor[60,61]; however, few eukaryote genes branch with Gram-positive homologues. (2) In hypotheses entailing more than one eubacterial partner at eukaryote origins (Fig. 4a–c), they are explained as descending from the non-mitochondrial eubacterium; however, these genes branch all over the eubacterial tree, not with any particular lineage. (3) In models favouring widespread LGT from prokaryotes to eukaryotes, they are explained as separate acquisitions from individual donors[81]; although some LGT clearly has occurred[82], the jury is still out on its extent because of a lack of detailed large-scale analyses of individual genes using reliable methods. (4) In single-eubacterium models (Fig. 4e–g), they are either not addressed, or explained as acquisitions from the mitochondrial symbiont, with a twofold corollary[8] of LGT among free-living prokaryotes since the origin of mitochondria, and phylogenetic artefact.

LGT among prokaryotes[83] figures into the origin of eukaryotic operational genes in a fundamental manner that is often overlooked. Most claims of outright LGT to ancestral eukaryotes (that is, from donors distinct from the mitochondrion) implicitly assume a static chromosome model in which prokaryotes do not exchange genes among themselves; finding a eukaryotic gene that branches with a group other than α-proteobacteria is taken as evidence for an origin from that group (the vagaries of deep branches notwithstanding). But if we embrace a fluid chromosome model for prokaryotes, as some interpretations of the data suggest we should[84], then the expected phylogeny for a gene acquired from the mitochondrion would be common ancestry for all eukaryotes, but not necessarily tracing to α-proteobacteria, because the ancestor of mitochondria possessed an as yet unknown collection of genes.

The timing and ecological context of eukaryote origins

Diversified unicellular microfossils of uncertain phylogenetic affinity (acritarchs), but widely accepted as eukaryotes, appear in strata of ~1.45 billion years (Gyr) of age[85], providing a minimum age for the group. *Bangiomorpha*, a fossilized multicellular organism virtually indistinguishable in morphology from modern bangiophyte red algae, has been found in strata of ~1.2 Gyr of age[86], placing a lower bound on the age of the plant kingdom. A wide range of molecular clock estimates of eukaryote age have been reported, but these are still uncertain, being contingent both on the use of younger calibration points and on the phylogenetic model and assumed tree[87]. At present, a minimum age of eukaryotes at ~1.45 Gyr and a minimum age of the plant kingdom at ~1.2 Gyr seem to be criteria that the molecular clock must meet.

The classical view of early eukaryote evolution posits two main ecological stages: (1) the early emergence and diversification of anaerobic, amitochondriate lineages, followed by (2) the acquisition of an oxygen-respiring mitochondrial ancestor in one lineage thereof and the subsequent diversification of aerobic eukaryotic lineages[78]. Concordant with that view, mitochondrial origins have traditionally been causally linked to the global rise in atmospheric oxygen levels at ~2 Gyr ago and an assumed 'environmental disaster' for cells lacking the mitochondrial endosymbiont[63,88], providing a selective force (oxygen detoxification) for the acquisition of the mitochondrion[63,88]. Two observations challenge this model.

First, it is now clear that the contemporary anaerobic eukaryotes did not branch off before the origin of mitochondria. Second, new isotope studies indicate that anaerobic environments persisted locally and globally over the past 2 Gyr. That oxygen first appeared in the atmosphere at ~2 Gyr ago is still generally accepted, but it is now thought that, up until about 600 Myr ago, the oceans existed in an intermediate oxidation state, with oxygenated surface water (where photosynthesis was occurring), and sulphide-rich (sulphidic) and oxygen-lacking (anoxic) subsurface water[89,90]. Hence, the 'oxygen event' in the atmosphere should be logically decoupled from anoxic marine environments, where anaerobic eukaryotes living on the margins of an oxic world could have flourished, as they still do today[27].

Outlook

In the past, phylogenetic trees have produced a particular view of early eukaryote history that was appealing, but turned out to be wrong in salient aspects. Simply testing whether a model used to make a tree actually fits the data[40] would do much to restore confidence in the merits of deep phylogenetic analyses. The fact that monophyly of plants can be recovered using molecular sequences[91], an event that should predate 1.2 Gyr, suggests that ancient signal can be extracted, but how far back we might expect to be able to go is uncertain. The persistence of mitochondrially derived organelles in all eukaryotes, and plastids in some lineages, provides phylogeny-independent evidence for the occurrence of those symbiotic events. But independent evidence for the participation of other prokaryotic endosymbionts is lacking. Analysis of mitochondria in their various guises has revealed that their unifying trait is neither respiration nor ATP synthesis; the common essential function, if any, for contemporary eukaryotes remains to be pinpointed by

comparative study. It may still be that a eukaryote is lurking out there that never possessed a mitochondrion—a *bona fide* archezoan—in which case prokaryote-host models (Fig. 4e–g) for eukaryogenesis can be abandoned. However, morphological studies and environmental sequencing efforts performed so far from the best candidate habitats to harbour such relics—anaerobic marine sediments—have not uncovered new, unknown and more-deeply branching lineages; rather, they have uncovered a greater diversity of lineages with affinities to known mitochondriate groups[28,61]. The available phylogenetic findings from genomes are not fully consistent with any current hypothesis for eukaryote origins, the underlying reasons for which—biological, methodological or both—are as yet unclear. Genomes must surely bear some testimony to eukaryotic origins, but new approaches and more rigorous attention to the details of phylogenetic inference will be required to decipher the message.

1. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87,** 4576–4579 (1990).
2. Leipe, D. D., Gunderson, J. H., Nerad, T. A. & Sogin, M. L. Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Mol. Biochem. Parasitol.* **59,** 41–48 (1993).
3. Hashimoto, T., Nakamura, Y., Kamaishi, T. & Hasegawa, M. Early evolution of eukaryotes inferred from the amino acid sequences of elongation factors 1α and 2. *Arch. Protistenkd.* **148,** 287–295 (1997).
4. Cavalier-Smith, T. in *Endocytobiology II* (eds Schwemmler, W. & Schenk, H. E. A.) 1027–1034 (De Gruyter, Berlin, 1983).
5. Gray, M. W., Lang, B. F. & Burger, G. Mitochondria of protists. *Annu. Rev. Genet.* **38,** 477–524 (2004).
6. Cavalier-Smith, T. in *Endocytobiology II* (eds Schwemmler, W. & Schenk, H. E. A.) 265–279 (De Gruyter, Berlin, 1983).
7. Pfanner, N. & Geissler, A. Versatility of the mitochondrial protein import machinery. *Nature Rev. Mol. Cell Biol.* **2,** 339–349 (2001).
8. Esser, C. *et al.* A genome phylogeny for mitochondria among α-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21,** 1643–1660 (2004).
9. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* **5,** 123–135 (2004).
10. Clark, C. G. & Roger, A. J. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc. Natl Acad. Sci. USA* **92,** 6518–6521 (1995).
11. Roger, A. J. Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154,** S146–S163 (1999).
12. Abrahamsen, M. S. *et al.* Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304,** 441–445 (2004).
13. Boxma, B. *et al.* An anaerobic mitochondrion that produces hydrogen. *Nature* **434,** 74–79 (2005).
14. Müller, M. in *Molecular Medical Parasitology* (eds Marr, J. J., Nilsen, T. W. & Komuniecki, R. W.) 125–139 (Academic, Amsterdam, 2003).
15. Katinka, M. D. *et al.* Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414,** 450–453 (2001).
16. Tielens, A. G., Rotte, C., van Hellemond, J. J. & Martin, W. Mitochondria as we don't know them. *Trends Biochem. Sci.* **27,** 564–572 (2002).
17. Komuniecki, R. W. & Tielens, A. G. M. in *Molecular Medical Parasitology* (eds Marr, J. J., Nilsen, T. W. & Komuniecki, R.) 339–358 (Academic, Amsterdam, 2003).
18. Darwin, C. *The Origin of Species* Reprint edn (Penguin Books, London, 1968).
19. Müller, M. The hydrogenosome. *J. Gen. Microbiol.* **139,** 2879–2889 (1993).
20. van der Giezen, M. *et al.* Conserved properties of hydrogenosomal and mitochondrial ADP/ATP carriers: A common origin for both organelles. *EMBO J.* **21,** 572–579 (2002).
21. Tjaden, J. *et al.* A divergent ADP/ATP carrier in the hydrogenosomes of *Trichomonas gallinae* argues for an independent origin of these organelles. *Mol. Microbiol.* **51,** 1439–1446 (2004).
22. Embley, T. M. *et al.* Hydrogenosomes, mitochondria and early eukaryotic evolution. *IUBMB Life* **55,** 387–395 (2003).
23. Dyall, S. D. *et al.* Presence of a member of the mitochondrial carrier family in hydrogenosomes: Conservation of membrane-targeting pathways between hydrogenosomes and mitochondria. *Mol. Cell. Biol.* **20,** 2488–2497 (2000).
24. Sutak, R. *et al.* Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote *Trichomonas vaginalis*. *Proc. Natl Acad. Sci. USA* **101,** 10368–10373 (2004).
25. Hrdy, I. *et al. Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* **432,** 618–622 (2004).
26. Schnarrenberger, C. & Martin, W. Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. A case study of endosymbiotic gene transfer. *Eur. J. Biochem.* **269,** 868–883 (2002).
27. Fenchel, T. & Finlay, B. J. *Ecology and Evolution in Anoxic Worlds* (eds May, R. M. & Harvey, P. H.) (Oxford Univ. Press, Oxford, 1995).
28. Roger, A. J. & Silberman, J. D. Cell evolution: Mitochondria in hiding. *Nature* **418,** 827–829 (2002).
29. Whatley, J. M., John, P. & Whatley, F. R. From extracellular to intracellular: The establishment of mitochondria and chloroplasts. *Proc. R. Soc. Lond. B* **204,** 165–187 (1979).
30. Dyall, S. D., Brown, M. T. & Johnson, P. J. Ancient invasions: From endosymbionts to organelles. *Science* **304,** 253–257 (2004).
31. Dyall, S. D. *et al.* Non-mitochondrial complex I proteins in a hydrogenosomal oxidoreductase complex. *Nature* **431,** 1103–1107 (2004).
32. Tovar, J., Fischer, A. & Clark, C. G. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol. Microbiol.* **32,** 1013–1021 (1999).
33. Tovar, J. *et al.* Mitochondrial remnant organelles of *Giardia* function in iron–sulphur protein maturation. *Nature* **426,** 172–176 (2003).
34. Williams, B. A., Hirt, R. P., Lucocq, J. M. & Embley, T. M. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature* **418,** 865–869 (2002).
35. Regoes, A. *et al.* Protein import, replication and inheritance of a vestigial mitochondrion. *J. Biol. Chem.* **280,** 30557–30563 (2005).
36. Chan, K. W. *et al.* A Novel ADP/ATP transporter in the mitosome of the microaerophilic human parasite *Entamoeba histolytica*. *Curr. Biol.* **15,** 737–742 (2005).
37. Dolezal, P. *et al. Giardia* mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting. *Proc. Natl Acad. Sci. USA* **102,** 10924–10929 (2005).
38. Lill, R. & Muhlenhoff, U. Iron–sulfur-protein biogenesis in eukaryotes. *Trends Biochem. Sci.* **30,** 133–141 (2005).
39. Ali, V., Shigeta, Y., Tokumoto, U., Takahashi, Y. & Nozaki, T. An intestinal parasitic protist, *Entamoeba histolytica*, possesses a non-redundant nitrogen fixation-like system for iron–sulfur cluster assembly under anaerobic conditions. *J. Biol. Chem.* **279,** 16863–16874 (2004).
40. Penny, D., McComish, B. J., Charleston, M. A. & Hendy, M. D. Mathematical elegance with biochemical realism: The covarion model of molecular evolution. *J. Mol. Evol.* **53,** 711–723 (2001).
41. Ho, S. Y. W. & Jermiin, L. S. Tracing the decay of the historical signal in biological sequence data. *Syst. Biol.* **53,** 623–637 (2004).
42. Felsenstein, J. Cases in which parsimony or incompatibility methods will be positively misleading. *Syst. Zool.* **25,** 401–410 (1978).
43. Stiller, J. W. & Hall, B. D. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol. Biol. Evol.* **16,** 1270–1279 (1999).
44. Philippe, H. *et al.* Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc. R. Soc. Lond. B* **267,** 1213–1221 (2000).
45. Hirt, R. P. *et al.* 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 (1999).
46. Keeling, P. J., Luker, M. A. & Palmer, J. D. Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol. Biol. Evol.* **17,** 23–31 (2000).
47. Arisue, N., Hasegawa, M. & Hashimoto, T. Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. *Mol. Biol. Evol.* **22,** 409–420 (2005).
48. Penny, D. Criteria for optimising phylogenetic trees and the problem of determining the root of a tree. *J. Mol. Evol.* **8,** 95–116 (1976).
49. Stechmann, A. & Cavalier-Smith, T. The root of the eukaryote tree pinpointed. *Curr. Biol.* **13,** R665–R666 (2003).
50. Steenkamp, E. T., Wright, J. & Baldauf, S. L. The protistan origins of animals and fungi. *Mol. Biol. Evol.* **23,** 93–106 (2006); published online 8 September 2005 (doi:10.1093/molbev/msj011).
51. Richards, T. A. & Cavalier-Smith, T. Myosin domain evolution and the primary divergence of eukaryotes. *Nature* **436,** 1113–1118 (2005).
52. Adl, S. M. *et al.* The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* **52,** 399–451 (2005).
53. Graybeal, A. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47,** 9–17 (1998).
54. Philippe, H. & Adoutte, A. in *Evolutionary Relationships Among Protozoa* (eds Coombs, G. H., Vickerman, K., Sleigh, M. A. & Warren, A.) 25–56 (Kluwer Academic, Dordrecht, 1998).
55. Forterre, P. & Philippe, H. Where is the root of the universal tree of life? *Bioessays* **21,** 871–879 (1999).
56. Knoll, A. H. *Life on a Young Planet: The First Three Billion Years of Evolution on Earth* (Princeton Univ. Press, 2003).
57. Margulis, L., Dolan, M. F. & Whiteside, J. H. "Imperfections and oddities" in the origin of the nucleus. *Paleobiology* **31,** 175–191 (2005).
58. Moreira, D. & Lopez Garcia, P. Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *J. Mol. Evol.* **47,** 517–530 (1998).
59. Lake, J., Moore, J., Simonson, A. & Rivera, M. in *Microbial Phylogeny and Evolution Concepts and Controversies* (ed. Sapp, J.) 184–206 (Oxford Univ. Press, Oxford, 2005).

60. Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52,** 297–354 (2002).
61. Cavalier-Smith, T. Only six kingdoms of life. *Proc. R. Soc. Lond. B* **271,** 1251–1262 (2004).
62. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37–41 (1998).
63. Vellai, T., Takacs, K. & Vida, G. A new aspect to the origin and evolution of eukaryotes. *J. Mol. Evol.* **46,** 499–507 (1998).
64. Searcy, D. G. in *The Origin and Evolution of the Cell* (eds Matsuno, H. H. & Matsuno, K.) 47–78 (World Scientific, Singapore, 1992).
65. von Dohlen, C. D., Kohler, S., Alsop, S. T. & McManus, W. R. Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. *Nature* **412,** 433–436 (2001).
66. Zillig, W., Schnabel, R. & Stetter, K. O. Archaeabacteria and the origin of the eukaryotic cytoplasm. *Curr. Top. Microbiol. Immunol.* **114,** 1–18 (1985).
67. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95,** 6239–6244 (1998).
68. Ribeiro, S. & Golding, G. B. The mosaic nature of the eukaryotic nucleus. *Mol. Biol. Evol.* **15,** 779–788 (1998).
69. Lake, J. A. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331,** 184–186 (1988).
70. Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257,** 74–76 (1992).
71. Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. The root of the universal tree and the origin of eukaryotes based upon elongation factor phylogeny. *Proc. Natl Acad. Sci. USA* **93,** 7749–7754 (1996).
72. Brown, J. R. & Doolittle, W. F. Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61,** 456–502 (1997).
73. Tourasse, N. J. & Gouy, M. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol. Phylogenet. Evol.* **13,** 159–168 (1999).
74. Brown, J. R. *et al.* Universal trees based on large combined protein sequence data sets. *Nature Genet.* **28,** 281–285 (2001).
75. Daubin, V., Gouy, M. & Perriere, G. A phylogenomic approach to bacterial phylogeny: Evidence of a core of genes sharing a common history. *Genome Res.* **12,** 1080–1090 (2002).
76. Rivera, M. C. & Lake, J. A. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431,** 152–155 (2004).
77. Horiike, T., Hamada, K., Miyata, D. & Shinozawa, T. The origin of eukaryotes is suggested as the symbiosis of *Pyrococcus* into γ-proteobacteria by phylogenetic tree based on gene content. *J. Mol. Evol.* **59,** 606–619 (2004).
78. Margulis, L., Dolan, M. F. & Guerrero, R. The chimeric eukaryote: Origin of the nucleus from the karyomastigont in amitochondriate protists. *Proc. Natl Acad. Sci. USA* **97,** 6954–6959 (2000).
79. Gabaldon, T. & Huynen, M. A. Reconstruction of the proto-mitochondrial metabolism. *Science* **301,** 609 (2003).
80. Martin, W. *et al.* Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* **99,** 12246–12251 (2002).
81. Doolittle, W. F. You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* **14,** 307–311 (1998).
82. Loftus, B. *et al.* The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433,** 865–868 (2005).
83. Doolittle, W. F. Lateral genomics. *Trends Cell Biol.* **9,** M5–M8 (1999).
84. Kunin, V., Goldovsky, L., Darzentas, N. & Ouzounis, C. A. The net of life—reconstruction of the microbial phylogenetic network. *Genome Res.* **15,** 954–959 (2005).
85. Javaux, E. J., Knoll, A. H. & Walter, M. R. Morphological and ecological complexity in early eukaryotic ecosystems. *Nature* **412,** 66–69 (2001).
86. Butterfield, N. J. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26,** 386–404 (2000).
87. Benton, M. J. & Ayala, F. J. Dating the tree of life. *Science* **300,** 1698–1700 (2003).
88. Kurland, C. G. & Andersson, S. G. Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* **64,** 786–820 (2000).
89. Shen, Y., Knoll, A. H. & Walter, M. R. Evidence for low sulphate and anoxia in a mid-Proterozoic marine basin. *Nature* **423,** 632–635 (2003).
90. Poulton, S. W., Fralick, P. W. & Canfield, D. E. The transition to a sulphidic ocean ~1.84 billion years ago. *Nature* **431,** 173–177 (2004).
91. Rodriguez-Ezpeleta, N. *et al.* Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* **15,** 1325–1330 (2005).
92. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276,** 734–740 (1997).
93. Keithly, J. S., Langreth, S. G., Buttle, K. F. & Mannella, C. A. Electron tomographic and ultrastructural analysis of the *Cryptosporidium parvum* relict mitochondrion, its associated membranes, and organelles. *J. Eukaryot. Microbiol.* **52,** 132–140 (2005).

Acknowledgements We thank M. Müller, J. Archibald, R. Hirt, K. Henze and L. Tielens, and members of our laboratories, for discussions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to T.M.E. (martin.embley@ncl.ac.uk) or W.M. (w.martin@uni-duesseldorf.de).