Eur. J. Biochem. **270**, 1599–1618 (2003) © FEBS 2003 doi:10.1046/j.1432-1033.2003.03499.x

REVIEW ARTICLE

Mitochondrial connection to the origin of the eukaryotic cell

Victor V. Emelyanov

Gamaleya Institute of Epidemiology and Microbiology, Moscow, Russia

Phylogenetic evidence is presented that primitively amitochondriate eukaryotes containing the nucleus, cytoskeleton, and endomembrane system may have never existed. Instead, the primary host for the mitochondrial progenitor may have been a chimeric prokaryote, created by fusion between an archaebacterium and a eubacterium, in which eubacterial energy metabolism (glycolysis and fermentation) was retained. A Rickettsia-like intracellular symbiont, suggested to be the last common ancestor of the family Rickettsiaceae and mitochondria, may have penetrated such a host (pro-eukaryote), surrounded by a single membrane, due to tightly membrane-associated phospholipase activity, as do present-day rickettsiae. The relatively rapid evolutionary conversion of the invader into an organelle may have occurred in a safe milieu via numerous, often dramatic, changes involving both partners, which resulted in successful coupling of the host glycolysis and the symbiont respiration. Establishment of a potent energy-generating organelle made it possible, through rapid dramatic changes, to develop genuine eukaryotic elements. Such sequential, or converging, global events could fill the gap between prokaryotes and eukaryotes known as major evolutionary discontinuity.

Keywords: endosymbiotic origin; energy metabolism; mitochondrial ancestor; respiration; rickettsiae; fusion hypothesis; eukaryogenesis; phylogenetic analysis; paralogous protein family.

From a genomics perspective, it is clear that both archaebacteria (domain Archaea) and eubacteria (domain Bacteria) contributed substantially to eukaryotic genomes [1–7]. It is also evident that eukaryotes (domain Eukarya) acquired eubacterial genes from a single mitochondrial ancestor during endosymbiosis [8–14], which probably occurred early in eukaryotic evolution [10,11,15–17]. This does not, however, necessarily mean that the mitochondrial ancestor was the only source of bacterial genes, although the number of transferred genes could be large enough given the fundamental difference in gene content between bacteria and organelles [10,11]. According to the archaeal hypothesis (Fig. 1A, left panel), a primitively amitochondriate eukaryote originated from an archaebacterium, and eubacterial genes were acquired from a mitochondrial symbiont [1, 18–20]. The alternative fusion, or chimera, theory (Fig. 1A, right panel) posits that an amitochondriate cell emerged as a fusion between an archaebacterium and a eubacterium, with their genomes having mixed in some way [1,3,6,21–24]. The so-called Archezoa concept (Fig. 1A) implies that the host for the mitochondrial symbiont has been yet a eukaryote, i.e. possessed at least some features distinguishing eukaryotes from prokaryotes [1,17,25–30]. The gene ratchet hypothesis, recently proposed by Doolittle [28], suggests that such an archezoon might have acquired eubacterial genes via endocytosis upon feeding on eubacteria. In effect, these firmly established facts and relevant ideas address two important, yet simple, questions about mitochondrial origin. (a) Were the genes of eubacterial provenance first derived from the mitochondrial ancestor or already present in the host genome before the advent of the organelle? (b) Did eukaryotic features such as the nucleus, endomembrane system, and cytoskeleton evolve before or after mitochondrial symbiosis?

There is little doubt that mitochondria monophyletically arose from within the α subdivision of proteobacteria, with their closest extant relatives being obligate intracellular symbionts of the order Rickettsiales [9–11,13,22,31–44]. This relationship was established by phylogenetic analyses of both small [34,37,39] and large [34] subunit rRNA, as well as Cob and Cox1 subunits of the respiratory chain using all α-proteobacterial sequences from finished and unfinished genomes known to date (V. V. Emelyanov, unpublished results). The four corresponding genes always reside in the organellar genomes and are therefore appropriate tracers for the origin of the organelle itself [10,45]. Thus, a sister-group relationship of eukaryotes and rickettsiae to the exclusion of free-living micro-organisms of the α subdivision revealed in phylogenetic analysis of a particular gene (protein), regardless of whether or not it serves an organelle, would confirm the acquisition of such a gene by Eukarya from a

Correspondence to V. V. Emelyanov, Department of General Microbiology, Gamaleya Institute of Epidemiology and Microbiology, Gamaleya Street 18, 123098 Moscow, Russia.
Fax: + 7095 1936183, Tel.: + 7095 7574644,
E-mail: vvemilio@jscc.ru

Abbreviations: ER, endoplasmic reticulum; LGT, lateral gene transfer; LBA, long-branch attraction; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; TPI, triose phosphate isomerase; PFO, pyruvate–ferredoxin oxidoreductase; Bya, billion years ago; ValRS, valyl-tRNA synthetase; MSH, MutS-like; IscS, iron–sulfur cluster assembly protein; AlaRS, alanyl-tRNA synthetase.

Dedication: This paper is dedicated to Matti Saraste, Managing Editor of *FEBS Letters*, who died on 21 May 2001.

(Received 30 October 2002, revised 20 December 2002, accepted 4 February 2003)

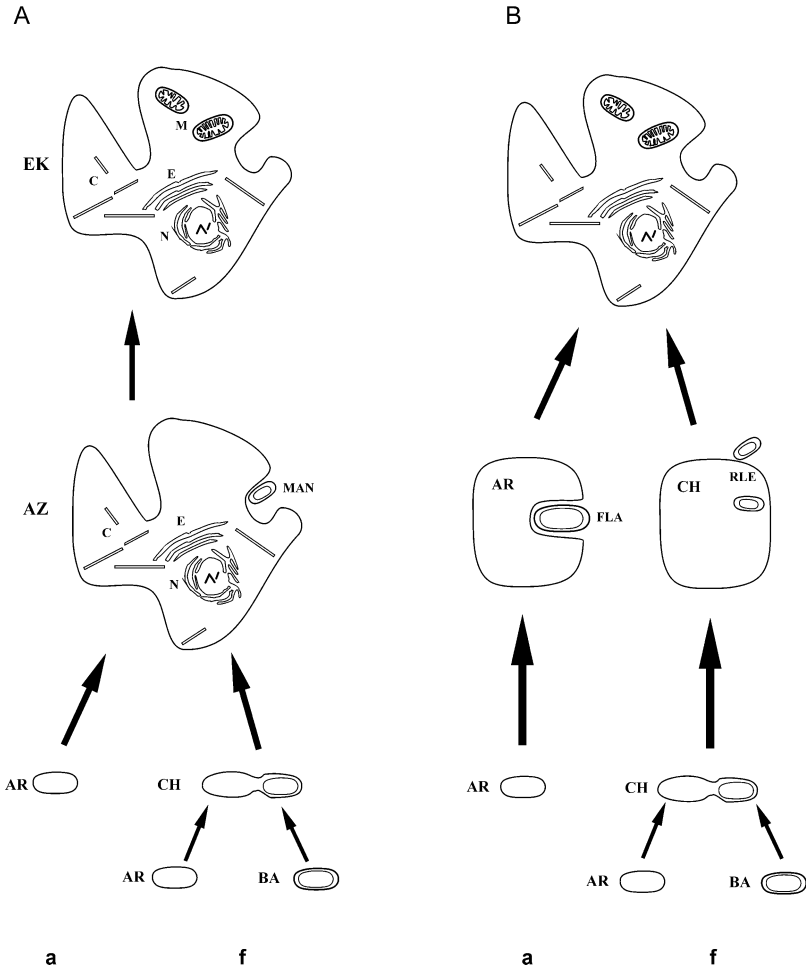


Fig. 1. The main competing theories of eukaryotic origin. Schematic diagrams describing the Archezoa (A) and anti-Archezoa (B) hypotheses, and their archaeal (a) and fusion (f) versions as envisioned from genomic and biochemical perspectives. Abbreviations: AR, archaeon; BA, bacterium; CH, chimeric prokaryote; AZ, archezoon; EK, eukaryote; MAN, mitochondrial ancestor; FLA, free-living α-proteobacterium; RLE, rickettsia-like endosymbiont; N, nucleus with multiple chromosomes; E, endomembrane system; C, cytoskeleton; M, mitochondria.

mitochondrial progenitor. This canonical pattern for the endosymbiotic origin may provide a reference framework in attempts to distinguish between the above hypotheses.

It should be realized that the archaeal hypothesis is much easier to reject than to confirm. Indeed, the latter may be accepted only if most eubacterial-like eukaryal genes turned out to be α-proteobacterial in origin, with the origin of the remainder being readily ascribed to lateral gene transfer (LGT). Of importance to this issue, several cases of a putative LGT from various eubacterial taxa to some protists have recently been reported [46–54] in good agreement with the above gene transfer ratchet. It is, however, an open question whether such acquisitions occurred early in eukaryotic evolution, e.g. before mitochondrial origin.

Whereas the sources of eubacterial genes may in principle be established in this way on the basis of multiple phylogenetic reconstructions, how and when the characteristically eukaryotic structures (and hence the eukaryote itself) appeared is difficult to assess. At first glance, there can be no appropriate molecular tracers for the origin of the nucleus, endomembrane, and cytoskeleton. Nonetheless, phylogenetic methods can still be applied to proteins, the appearance of which might have accompanied the origin of the respective eukaryotic compartments [21,23]. Unfortunately if one considers a specifically eukaryotic protein (which implies poor homology with bacterial orthologs), reliable alignment of the sequences needed for phylogenetic analysis are hardly possible. This is best exemplified by the cytoskeletal proteins actin and tubulin, the distant homologs of which have been suggested to be prokaryotic FtsA and FtsZ, respectively [55,56]. Curiously, actin was recently argued to derive from MreB [57]. On the other hand, when one considers a eukaryotic protein highly homologous to bacterial counterparts and show that it arose from the same lineage as the mitochondrion, the possibility remains that it first appeared in Eukarya even before the endosymbiotic event, but was subsequently displaced by an endosymbiont homolog. Furthermore, such a single ubiquitous protein would not be characteristic of a eukaryote.

One way to circumvent this problem was prompted by Gupta [23]. As convincingly argued in this work, the emergence of endoplasmic reticulum (ER) forms of conserved heat shock proteins via duplication of ancestral genes in a eukaryotic lineage may be indicative of the origin of ER *per se* [23]. Here I put forward an approach based on logical interpretation of phylogenetic data involving such eukaryotic paralogs (multigene families). If phylogenetic analysis reveals branching off of the sequences from free-living α-proteobacteria before a monophyletic cluster represented by rickettsial and paralogous eukaryotic sequences, i.e. a canonical pattern, this would mean that paralogous

duplication (multiplication) of protein, which must have accompanied the origin of the corresponding eukaryotic structure, occurred subsequent to mitochondrial origin. Otherwise it would be improbable that this protein was multiplied to meet the requirements of the emerging eukaryotic compartment prior to mitochondrial symbiosis, but subsequently, two or more copies were simultaneously replaced by a mitochondrial homolog that similarly multiplied to accomodate them.

In addition to *Rickettsia prowazekii* [9], complete genomes of free-living α-proteobacteria [58–62] and *Rickettsia conorii* [63], as well as sequences from unfinished genomes of *Wolbachia* sp., *Ehrlichia chaffeensis*, *Anaplasma phagocytophila* (http://www.tigr.org/tdb/mdb/mdbinprogress.html) and *Cowdria ruminantium* (http://www.sanger.ac.uk/projects/microbes) – species of a taxonomic assemblage closely related to or belonging within the family Rickettsiaceae [34] – have now become available, thus providing an opportunity to answer the above questions. I here present phylogenetic data, based on the broad use of α-proteobacterial protein sequences, which support the fusion hypothesis for a primitively amitochondriate cell (pro-eukaryote) and suggest that the host for the mitochondrial symbiont was a prokaryote.

Molecular phylogeny

Prokaryotes and eukaryotes (similarly bacteria and organelles) are so fundamentally different that complex characters, such as morphological traits, are of no use in discerning their relatedness [11,17,29]. It is the common belief that evolutionary relationships, including distant ones, can be deduced from multiple phylogenetic relationships of conserved genes and proteins using the methods of molecular phylogeny [1,13,23]. A simple rationale underlying the molecular approach is the following: the larger the number of replications (generations) separating related sequences from each other, the more different (i.e. less related) the sequences are, because of accumulation of mutational changes. There are three main phylogenetic methods: maximum likelihood (ML), the distance matrices-based methods (DM methods), and maximum parsimony (MP) [64–67]. The respective computer programs use alignment of the gene and protein sequences to produce phylogenetic trees. As the above methods interpret sequence alignments in different ways, the results are regarded as very reliable if they do not depend on the method used. The quality of alignment is strongly affected by the degree of sequence similarity. The regions that cannot be unambiguously aligned are normally removed, so as to obtain similar sequences of equal length. This procedure seems to be unbiased, given that highly variable regions usually contain mutationally saturated positions with little phylogenetic signal [68,69]. Generally, there are three types of homology. Proteins may be (partially) homologous due to convergence towards a common function (convergent similarity), in which case nothing can be ascertained about the evolutionary relationship. Two other types of homology are more evolutionarily meaningful. Homologous genes (proteins) of these types are called orthologous and paralogous genes (proteins). By definition, orthologous genes arose in different taxonomic groups by means of vertical gene transfer (i.e. from ancestor to progeny). Orthologous proteins usually have the same function and localize to the same or similar subcellular compartment. Paralogous genes emerged via duplication (multiplication) of a single gene followed by specialization of the resulting copies either recruited to different compartments/structures or adapted to serve different functions. As the different paralogs can be inherited separately and independently, their mixing up would be detrimental to phylogenetic inferences. On the contrary, recognized paralogy may be highly useful in this regard [1,70]. In particular, very ancient duplications have been widely used for unbiased rooting of the tree of life (reviewed in [1]). For instance, it has been argued that EF-Tu/EF-G paralogy originated in the universal ancestor via duplication of the primeval gene followed by assignment to each copy of a distinct role in translation [71]. Indeed, bipartite trees, with each subtree comprising one and only one sort of paralog, were always produced in phylogenetic analyses based on the combined alignments of such duplicated sequences. In most cases, reciprocal rooting of this kind (both subtrees serve the outgroups to one another) revealed a sister-group relationship of archaebacteria and eukaryotes [1,71–73], a notable exception being phylogenetic evidence based on valyl-tRNA synthetase/isoleucyl-tRNA synthetase paralogy (see below).

As for paralogy, apparent cases of LGT are not disturbing but instructive; however, the biological meaning of the gene transfer needs to be understood [46,52,74–76]. At face value, the events of an LGT look like a polyphyly of the expectedly monophyletic groups, the representatives of which served the recipients of the transferred genes. (Although monophyletic groups can be cut off the phylogenetic tree by splitting a single stem entering the group, two or more branches lead to polyphyletic assemblages [25].)

The reliability of phylogenetic relationships inferred from the above methods is commonly assessed by performing a bootstrap analysis. In particular, a nonparametric bootstrap analysis serves to test the robustness of the sequence relationships as if scanning along the alignment. To this end, the original alignment is modified in such a way that some randomly selected columns are removed, and others are repeated one or more times to obtain 100 or more different alignments, each containing the original number of shuffled columns. It is clear from this that the longer the aligned sequences, the more bootstrap replicates are to be used. Phylogenetic analysis is then performed on each of the resampled data to produce the corresponding number of phylogenetic trees. A consensus tree is inferred from these trees by placing bootstrap proportions at each node. The bootstrap proportions show how many times given branches emanate from a given node, and are thus interpreted as confidence levels. Normally, values above 50% are regarded as significant.

In contrast with paralogy and LGT, the long-branch attraction (LBA) artefact and related phenomena are real drawbacks of phylogenetic methods associated with unequal rates of evolution [68,69,77]. In contradiction to the evolutionary model, long branches (which are highly deviant and fast evolving, but not closely related sequences) tend to group together on phylogenetic trees [42,77]. Obviously, certain cases of LBA may be erroneously interpreted as LGT. ML methods are known to be relatively robust to the LBA artefact [64]. Furthermore, modern

applications of ML and DM methods take account of among-site rate variation, invoking the so-called gamma shape parameter α, a discrete approximation to gamma distribution of the rates from site to site. This correction is known to minimize the impact of LBA on phylogeny [69,78].

Several statistical tests have been developed to assess evolutionary hypotheses [66,79,80]. Approximately unbiased and Shimodaira-Hasegawa tests are strongly recommended rather than Templeton and Kishino-Hasegawa tests, when *a posteriori* obtained trees are compared with the user-defined trees representing the competing hypotheses of evolutionary relationship [80]. Relative rate tests are commonly used to address the question of whether mutational changes occur in the sequences in a clock-like fashion [66,79]. Various four-cluster analyses can help to assess the validity of three possible topologies of the unrooted trees consisting of four monophyletic clusters [66,79].

A search for sequence signatures [particular characters and insertions/deletions (indels)] is another, cladistic, approach aimed to resolve phylogenetic relationships. It is argued that such signatures, uniquely present in otherwise highly conserved regions of certain sequences, but absent from the same regions of all others, may be shared traits derived from a common ancestor (reviewed in detail in [23]).

As briefly discussed here, molecular phylogenetics provides a powerful tool for evolutionary studies. However, it is becoming evident that phylogenetic data should be considered in conjunction with geological, ecological and biochemical data, when the issue of eukaryotic origin is concerned [13,19,23,24].

Chimeric nature of the pro-eukaryote

Origin of eukaryotic energy metabolism

The fundamentally chimeric nature of eukaryotic genomes is becoming apparent, with genes involved in metabolic pathways (operational genes) being mostly eubacterial and information transfer genes (informational genes) being more related to archaeal homologs [1,2,4,7]. In particular, eukaryotic enzymes of energy metabolism tend to group on phylogenetic trees with bacterial homologs [1,9,11,13,20, 46–48,50,51,53,81–87]. This fundamental distinction has received partial support from the study of archaeal signature genes. In this study, genes unique to the domain Archaea were shown to be primarily those of energy metabolism [88]. The aforementioned version of the Archezoa hypothesis implies that the primitively amitochondriate eukaryote, a direct descendent of the archaebacterium, might have acquired eubacterial genes by a process involving endocytosis. If, however, this archezoon possessed energy metabolism of a specifically archaeal type, it is unlikely that eubacterial genes for energy pathways were acquired one by one via gene transfer ratchet. These considerations suggest that energy metabolism as a whole might have been acquired by Eukarya in a single, i.e. endosymbiotic, event.

The most popular version of the archaeal hypothesis, the so-called hydrogen hypothesis (Fig. 1B, left panel), claims that all genes encoding enzymes of energy pathways were derived by an archaebacterial host from a mitochondrial symbiont. The latter is envisioned as a versatile free-living α-proteobacterium capable of glycolysis, fermentation, and oxidative phosphorylation [19,20,85,89]. Indeed, earlier phylogenetic analysis of triose phosphate isomerase (TPI) involving an incomplete sequence from *Rhizobium etli* revealed affiliation of this single α-proteobacterial sequence with those of eukaryotes. Keeling & Doolittle [90] pointed out, however, that an alternative tree topology placing γ-proteobacteria as a sister group to Eukarya was insignificantly worse. On the contrary, recent reanalysis of TPI showed a sisterhood of eukaryotes and γ-proteobacteria [85]. This result was corroborated by detailed phylogenetic analysis involving all α-proteobacterial sequences known to date (Fig. 2A). It should be noted that some data sets included *R. etli*. In agreement with published data [1,47,85], a close relationship between eukaryal and γ-proteobacterial sequences was also shown using glyceraldehyde-3-phosphate dehydrogenase (GAPDH), another glycolytic enzyme (Fig. 2B). The same relationship was observed when phylogenetic analysis was conducted on glucose-6-phosphate isomerase ([86] and data not shown). Collectively, these data revealed a complex evolutionary history of certain glycolytic enzymes [47,49,50,53,54,82,85,86,93,94]. In particular, an exceptional phyletic position of the amitochondriate protist *Trichomonas vaginalis* on the GAPDH tree (Fig. 2B) was assumed to be due to LGT [94]. Nonetheless, the present and published observations suggest that not the α but the γ subdivision of proteobacteria, or a group ancestral to β and γ proteobacteria (see below), might be a donor taxon of eukaryotic glycolysis. A recently published detailed phylogenetic analysis of glycolytic enzymes also revealed no α-proteobacterial contribution to eukaryotes [95]. Given an aberrant branching order of some eubacterial phyla on the above trees (Fig. 2 and [95]), compared with one based on small subunit rRNA [39] and exhaustive indel analyses [23], it might be suggested that the glycolytic enzymes are prone to orthologous replacement and that an initial endosymbiotic origin of eukaryotic glycolysis has subsequently been obscured by promiscuous LGT. It would be strange, however, if none of the glycolytic enzymes escaped such a replacement.

It is worth noting the presence of the genes for GAPDH, enolase and phosphoglycetrate kinase in the *Wolbachia* (endosymbiont of *Drosophila*) and *E. chaffeensis* genomes. Thus, ehrlichiae possess three of 10 key glycolytic enzymes, whereas *R. prowazekii* [9] and *R. conorii* [63] have none. It is particularly important, bearing in mind the divergence of the tribes Wolbachieae and Ehrlichieae after the tribe Rickettsieae (e.g [96]). This means that the last common ancestor of the family Rickettsiaceae and mitochondria still possessed the above three glycolytic enzymes, and their loss from *Rickettsia* may be an autapomorphy.

Curiously, the functional TPI–GAPDH fusion protein was recently shown to be imported into mitochondria of diatoms and oomycetes. Notwithstanding the sister relationship of γ proteobacteria and Eukarya, these data were interpreted as evidence for the mitochondrial origin of the eukaryotic glycolytic pathway [85]. Likewise, pyruvate–ferredoxin oxidoreductase (PFO), a key enzyme in fermentation, was suggested to have been acquired from a mitochondrial symbiont [19,89,97]. Observations that mitochondria of the Kinetoplastid *Euglena gracilis* and the Apicomplexan *Cryptosporidium parvum* lack pyruvate

dehydrogenase but instead possess pyruvate–$NADP^+$ oxidoreductase, an enzyme that shares a common origin with PFO, were assumed to support this idea [97,98]. However, the above data may be easily explained in another way. Some cytosolic proteins, the origin of which actually predated mitochondrial symbiosis, might be secondarily recruited to the organelle merely on acquisition of the targeting sequence and other rearrangements. Such a retargeting of fermentation enzymes was earlier suggested to have taken place during evolutionary conversion of mitochondria into hydrogenosomes [34,41].

Recent phylogenetic analysis of PFO failed to show a specific affiliation of eubacterial-like, monophyletic eukaryal proteins with those of proteobacterial phyla [83]. It is worth mentioning the rather scarce distribution of this enzyme among α-proteobacteria. In particular, none of the complete α-proteobacterial genomes harbor the gene encoding PFO. It is, however, quite a widespread protein in β and γ subdivisions (finished and unfinished genomes). Neither was hydrogenosomal hydrogenase, another fermentation enzyme, shown to be α-proteobacterial in origin [51,84,87].

As mentioned above, numerous molecular data point to the common origin of mitochondria and the order Rickettsiales. Detailed phylogenetic analyses of the best-characterized small subunit rRNA and chaperonin Cpn60 sequences have consistently shown a sister-group relationship between the family Rickettsiaceae and mitochondria to the exclusion

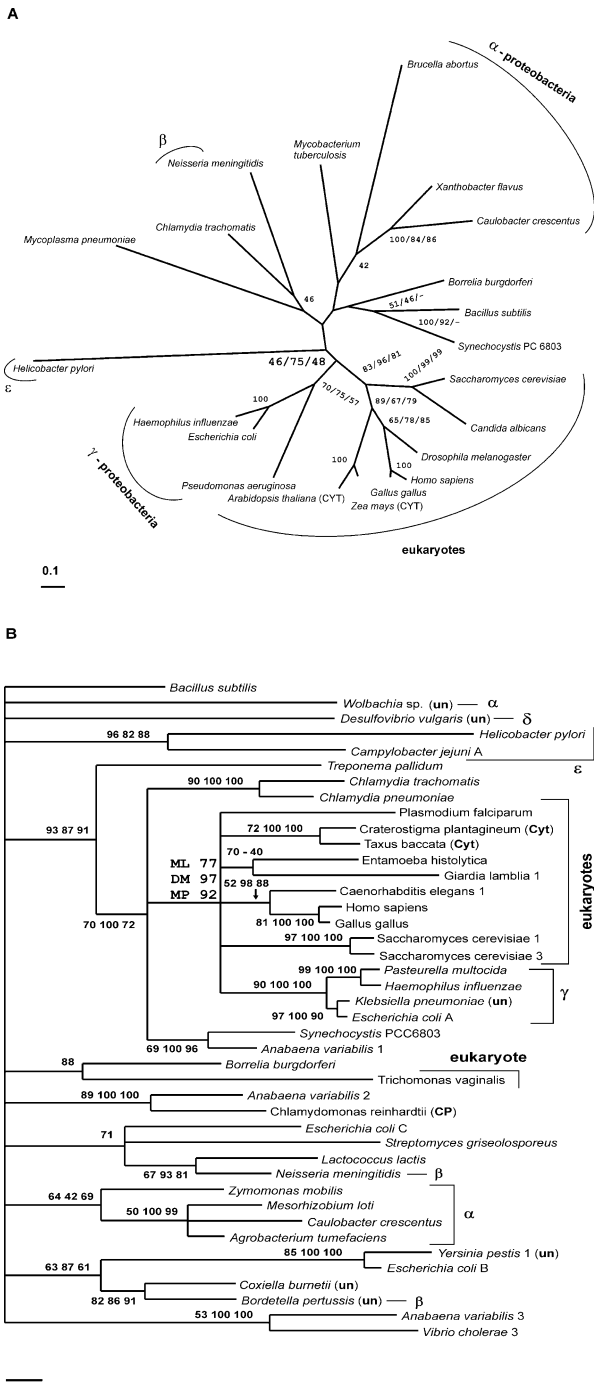


Fig. 2. Phylogenetic analysis of the glycolytic enzymes TPI (A) and GAPDH (B). Representative maximum likelihood (ML) trees are shown. Particular data sets included protists, other β and γ proteobacteria, and all α-proteobacteria for which the sequences are available in databases. Species sampling was proven to have no impact on the relationship of eukaryotic and proteobacterial sequences except for the cases of a putative LGT [85]. Bootstrap proportions (BPs) shown in percentages from left to right were obtained by ML, distance matrix (DM) and maximum parsimony (MP) methods, with those below 40% being indicated with hyphens. A single BP other than 100% pertains to the ML tree. Otherwise, support was 100% in all analyses. Scale bar denotes mean number of amino-acid substitutions per site for the ML tree. Dendrograms were drawn using the TREEVIEW program [91]. The sequences were obtained from GenBank unless otherwise specified. Abbreviations: Cyt, cytoplasm; CP, chloroplast; un, unfinished genomes. (A) ML majority rule consensus tree (ln likelihood = −7335.8) was inferred from 200 resampled data using SEQBOOT of the PHYLIP 3.6 package [65], PROTML of MOLPHY 2.3 [64], and PHYCON (http://www.binf.org/vibe/software/phycon/phycon.html) with the Jones, Taylor, and Thornton replacement model adjusted for amino-acid frequencies (JTT-f), as described elsewhere [83,92]. DM analysis was carried out by the neighbor-joining method using JTT matrix and Jin-Nei correction for among-site rate variation (PHYLIP) with the gamma shape parameter α estimated in PUZZLE. Unweighted MP analysis was performed by 50 rounds of random stepwise addition heuristic searches with tree bisection-reconnection branch swapping by using PAUP*, version 4.0 [67]. In DM and MP analysis, the data were bootstrapped 200 times. The MP trees were also inferred that constrained Eukarya to α-proteobacteria (PAUP), then evaluated by several statistical tests, as installed in the CONSEL 0.1d package [80]. The best constrained tree was not rejected at the 5% confidence level, with the *P* value of the most adequate approximately unbiased test [80] being 0.053. (B) The ML tree was constructed in PUZZLE with 10 000 puzzling steps using the JTT-f substitution model and one invariable plus eight variable rate categories (JTT-f + Γ + inv). The gamma shape parameter α (1.09) was estimated from the data set. DM analysis using ML distances was conducted on 200 resampled data by the FITCH program (PHYLIP) with global rearrangement and 15 permutations on sequence input order (G and J options). Distances were generated with PUZZLEBOOT (http://www.tree-puzzle.de/puzzleboot.sh) using the JTT-f + Γ + inv model. The MP consensus tree was inferred as above. Constrained trees were inferred as for TPI and evaluated as described above. The tree topology placing eukaryotic sequences with those from α-proteobacteria was strictly rejected by all tests of CONSEL.

of rickettsia-like endosymbionts classified in the order [34]. On the basis of these data, the mitochondrial origin was suggested to have been predisposed by the long-term mutualistic relationship of a rickettsia-like bacterium with a pro-eukaryote. In this way, the mitochondrial ancestor was regarded to be a highly reduced intracellular symbiont, which possessed both aerobic and anaerobic respiration, yet had lost many genes specifying redundant metabolic pathways such as glycolysis, fermentation and biosynthesis of small molecules [34]. In agreement with the fusion theory [21,23], these were assumed to have previously been inherited by the host mainly from a eubacterial fusion partner. Obviously, the above data are consistent with this contention.

Molecular dating

Timing of the appearance of eubacterial genes in eukaryotic genomes is another way to attempt to distinguish between different hypotheses about the origin of the pro-eukaryotic genome. Available data of this kind are rather controversial. On the one hand, Feng *et al.* [2] showed that archaeal genes appeared in Eukarya about 2.3 billion years ago (Bya) while eubacterial genes appeared 2.1 Bya. It was suggested that both estimates relate to the same event, fusion between an archaebacterium and a eubacterium, and the shift in the appearance time of bacterial genes to the present day was merely due to involvement in the analysis of mitochondrial and α-proteobacterial sequences. The above small difference would thus just reflect a more recent endosymbiotic event [96]. On the other hand, Rivera *et al.* [7] argued that archaeal (informational) genes were acquired by Eukarya in a single, very ancient event, whereas acquisitions of eubacterial (operational) genes were scattered along the timescale [7]. One may realize here that most eubacterial genes appeared in eukaryotes during both the fusion and subsequent endosymbiotic event, while others were derived from various bacterial groups more recently, when the true eukaryotes capable of endocytosis emerged (see below). Dating of the divergence of Rickettsiaceae and mitochondria, i.e. effectively the mitochondrial origin, was recently attempted by using the sequences of Cpn60, a ubiquitous, conserved protein with clock-like behavior. Rickettsiaceae and mitochondria were shown to have emerged 1.78 ± 0.17 Bya [96], i.e. significantly later than the appearance of eubacterial genes in eukaryotic genomes dated in the above-cited work [2] using a comparable approach.

Eukaryotic valyl-tRNA synthetase

With regard to the origin of the pro-eukaryotic genome, one important finding has been reported [77,96]. In eukaryotes, a single gene is known to encode cytosolic and mitochondrial valyl-tRNA synthetases (ValRSs), which are different in that a precursor of the organellar enzyme contains a mitochondrial-targeting sequence [99–101]. Hashimoto *et al.* [18] previously found that ValRS sequences of eukaryotes, including amitochondriate *T. vaginalis* and *Giardia lamblia*, and γ-proteobacteria contain a characteristic 37-amino-acid insertion which is absent from the sequences of all other known prokaryotes. Paralogous rooting of the ValRS tree with the most closely related isoleucyl-tRNA synthetases, which lack the insert, revealed the presence of the insert to be a derived state. The authors interpreted these data as evidence for acquisition of ValRS by eukaryotes from the mitochondrial symbiont, but pointed out a contemporary lack of relevant information from α-proteobacteria. These results were subsequently reanalyzed [96] involving archaeal-like ValRS from *R. prowazekii* [9] and a sequence from the unfinished genome of *Caulobacter crescentus* (a free-living α-proteobacterium). Figure 3A shows a comprehensive alignment of ValRS including all sequences from α, δ and ε subdivisions known to date, as well as the representatives from Eukarya and several prokaryotic taxa. It can be seen that only ValRS sequences of eukaryotes and β/γ-proteobacteria contain the characteristic 37-amino-acid insertion. Importantly, free-living α-proteobacteria possess insert-free enzyme of the eubacterial type, otherwise highly homologous to β/γ-proteobacterial counterparts, whereas Rickettsiaceae (*R. prowazekii*, *R. conorii*, *Wolbachia*, *E. chaffeensis* and *C. ruminantium*) also have the insert-free ValRS but of archaeal genre. Phylogenetic analysis of ValRS, performed at both the protein and DNA level, revealed monophyletic emergence of Rickettsiaceae from within Archaea (also supported by numerous sequence signatures) and a sister relationship of the free-living α-proteobacteria and β/γ-proteobacteria exclusive of Eukarya (data not shown). The latter means that the 37-amino-acid insert appeared in ValRS of β/γ-proteobacteria early during their diversification. The most parsimonious explanation of these data is that the pro-eukaryote inherited ValRS from β or γ proteobacteria, or their common ancestor before mitochondrial symbiosis (see also [77,96]). It is worth mentioning an apparent evolutionary (not convergent) origin of the insert itself (Fig. 3B). Apart from the origin of the pro-eukaryote, ValRS data shed light on the intriguing question of the extent and evolutionary significance of LGT [52,53,75,76]. The inference here is that acquisition of the archaeal enzyme by the family Rickettsiaceae or the order Rickettsiales shaped the evolutionary history of the rickettsial lineage.

Fig. 3. Signature sequence (37-amino-acid insertion) in ValRS that is uniquely shared by β-proteobacteria, γ-proteobacteria, and Eukarya (A) and phylogenetic analysis of insertion (B). The present alignment includes all known ValRSs from proteobacteria of α, δ and ε subdivisions, and several ValRSs from other phyla. All sequences of eukaryotes and β/γ-proteobacteria, which could be retrieved from finished and unfinished genomes using the BLAST server [102], contain a characteristic insert. It is lacking in ValRS of other prokaryotes and in isoleucyl-tRNA synthetase [18]. Identical amino-acid residues are shaded, and conserved ones are in bold. Two signatures showing the relatedness of rickettsial (R) homologs to Archaea (A) are printed in italics. Number and 's' on the top of the alignment indicate the sequence position of *R. prowazekii* ValRS and the above two signatures, respectively. Accession numbers of published entries follow the species names. The unrooted ML tree of the ValRS insert shown here was constructed using PUZZLE 4.0. DM analysis (FITCH) was based on ML distances obtained in PUZZLEBOOT. MP analysis was carried out using PROTPARS of PHYLIP with the J option. (A similar tree was obtained with PAUP parsimony.) For phylogenetic methods and other details, see legend to Fig. 2.

A

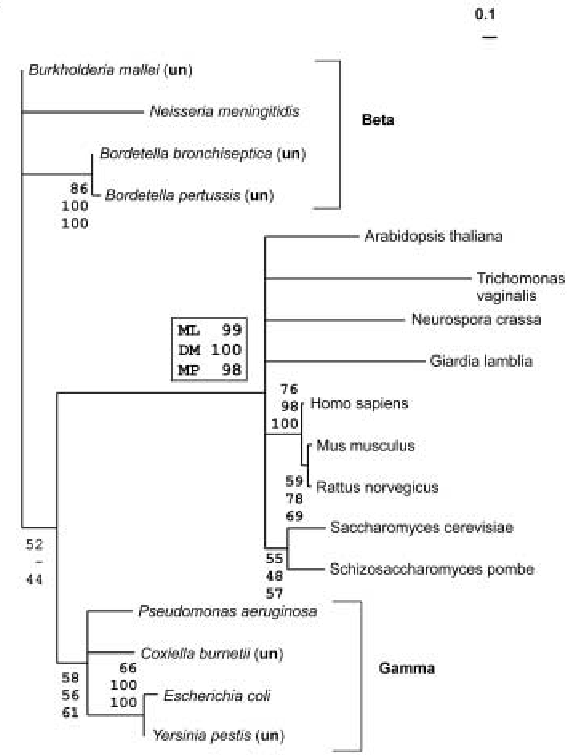
ValRS

```
                                             536                                                                            ss
Homo sapiens  M98326   [eukaryotes] KMSKSLGNVIDPLDVIYGISLQGLHNQLLNSNLDPSEVEKAKEGQKADFPAGIPECGTDALRFGL
Mus musculus  AAD26532              KMSKSLGNVIDPLDVIHGVSLQGLYDQLLNSNLDPSEVEKAKEGQKADFPAGIPECGTDALRFGL
Rattus norvegicus  Q04462           KMSKSLGNVIDPLDVIHGVSLQGLHDQLLNSNLDPSEVEKAKEGQRADFPAGIPECGTDALRFGL
Saccharomyces cerevisiae  J02719    KMSKSLGNVIDPLDVITGIKLDDLHAKLLQGNLDPREVEKAKIGQKESYPNGIPQCGTDAMRFAL
Schizosaccharomyces pombe CAA21241  KMSKSLGNVVDPIDVIEGISLQALHDKLLVGNLDSREVEKAKKGQRLSYPKGIPQCGTDALRFTL
Neurospora crassa  P28350           KMSKSLGNVIDPLDIIRGIELEDLHAKLLVGNLKEEEVARATKYQKTAFPGGIPECGADAMRFTL
Arabidopsis thaliana  P93736        KMSKSLGNVIDPLEVINGVTLGGLHKRLEEGNLDPKEVIVAKEGQVKDFPNGIPECGTDALRFAL
Trichomonas vaginalis  BAA28842     KMSKSLGNVIDPRHVINGIELEDLVAEIENSTFDDKEKKIAIDGRKADFPNGIPQCGTDAMRLAL
Giardia lamblia  AB008525           KMSKSKGNVVDPIDVIKGITLQEMGDKVRATNLPPKEIERALELQSKDFPIGIPECGTDALRFAL
Escherichia coli  X05891    [gamma] KMSKSKGNVIDPLDMVDGISLPELLEKRTGNMMQPQLADKIRKRTEKQFPNGIEPHGTDALRFTL
Pseudomonas aeruginosa  AAG07221    KMSKSKGNVLDPLDIVDGIDLDTLLQKRTSGMMQPKLAEKIAKQTRAEFPEGIASYGTDALRFTF
Coxiella burnetii                   KMSKSKGNVIDPIDIIDGISLDALIEKRTHALLQPKMAKTIEKMTRKEFPNGIASFGTDALRFTF
Burkholderia mallei          [beta] KMSKSKGNTLDPIDIVDGIGLDALVAKRTTGLMNPRQAATIEKKTRKEFPDGIPAFGTDALRFTM
Neisseria meningitidis  AAF40631    KMSKSEGNVIDPVDLIDGIGLEKLLVKRTTGLRKPETAPKVEEATKKLFPEGIPSMGADALRFTM
Bordetella pertussis                KMSKSKGNTLDPVDLIDGIDLKGLVRKRTFGLMHPKQAGAIEKATRRQYPDGIPAFGTDALRFTM
Rickettsia conorii  AAL03591    [R] KMSKSKGNVLVPEKLLE                                         QYGSDVIRYWS
Rickettsia prowazekii  CAA15124     KMSKSKGNVLVPEKLLE                                         RYGADVIRYWS
Wolbachia sp.                       KMSKSKGNIITPHIILE                                         TYGADVVRYWA
Ehrlichia chaffeensis               KMSKSKGNTLTPNKLLE                                         EYGADVVRYWA
Cowdria ruminantium                 KMSKSKGNALIPNQLLQ                                         EYGADVIRYWA
Brucella melitensis AAL52208[alpha] KMSKSKGNVIDPLELMD                                         EYGADALRFTL
Caulobacter crescentus  AAK23301    KMSKSKGNVMDPLILID                                         ELGCDAVRFTL
Agrobacterium tumefaciens AAK87491  KMSKSKGNVIDPLELID                                         EYGADALRFTL
Rhodobacter sphaeroides             KMSKSLGNVLDPLELID                                         EFGADAVRFTL
Sphingomonas aromaticivorans        KMSKSKGNVVDPLGLID                                         KYGADALRFFM
Sinorhizobium meliloti  CAC46095    KMSKSKGNVIDPLELID                                         EYGAGALRFTL
Mesorhizobium loti  BAB48552        KMSKSKGNVIDPLDLID                                         EYGADALRFTL
Rhodopseudomonas palustris          KMSKSKGNVIDPLNLID                                         EYGADALRFTL
Magnetospirillum magnetotacticum    KMSKSKGNIIDPLDLIE                                         KYGCDALRFTL
Silicibacter pomeroyi               KMSKSTGNVIDPLEIVD                                         EFGADALRFTM
Helicobacter pylori P56000[epsilon] KMSKSKGNVIDPLEMIE                                         KYGADSLRFTL
Campylobacter jejuni  U15295        KMSKSLGNVIDPNESIK                                         EYSADILRFTL
Geobacter sulfurreducens    [delta] KMSKSKGNVIDPLVVID                                         QYGTDAFRFTL
Desulfovibrio vulgaris              KMSKSTGNVIDPLAMID                                         KYGTDSLRFTL
Methanococcus jannaschii Q58413 [A] KMSKSRGNVVEPDEIIA                                         KYGADALRLWA
```

IleRS

```
Escherichia coli  X00776            KMSKSIGNTVSPQDVMN                                         KLGADILRLWV
Methanococcus jannaschii  Q58357    KMSKSLGNVVNPDDVVE                                         KYGADLLRFYL
```

B



Evolutionary ancestry of mitochondrial proteins

Ample data on the origin of mitochondrial proteins come from the study of the *Saccharomyces cerevisiae* mitochondrial proteome. It has been shown that as many as 160 of 210 bacterial-like mitochondrial proteins are not α-proteobacterial in origin [13,103]. Curiously, these values were far outnumbered in more recent work [14]. The simplest explanation of these data is that eubacterial genes related to the mitochondrion were present in the pro-eukaryotic genome before endosymbiosis, and easily recruited to serve the organelle during its origin. Indeed, it is very unlikely that the above 160 proteins were initially contributed by the mitochondrial ancestor and, hence, adapted to function in mitochondria, but subsequently replaced by their orthologs from other (bacterial) sources. Not to mention that recruitment of pre-existing genes would require one step less than acquisition by other ways that first require gene transfer to the host genome.

The data described in this section could be explained by pervasive LGT [20,76] mainly to the mitochondrial ancestor. However, it would be too strange a creature, an α-proteobacterial progenitor of mitochondria, with too many genes of non-α-proteobacterial origin. Of fundamental importance in this regard is the almost always observed monophyly of α-proteobacteria (e.g [95] and Fig. 2), with a striking exception being the above case for ValRS. Together, the present data reject the archaeal hypothesis and favor the fusion hypothesis for the primitively amitochondriate cell.

Taming of the mitochondrial symbiont: first step towards the eukaryote

It is evident that 'domestication' of the mitochondrial symbiont by the pro-eukaryotic host was accompanied by multiple changes in both the host and invader. These changes are particularly reflected in the protein sequences, ranging from smooth variations to dramatic ones. As shown in the above-cited studies [13,103], 47 mitochondrial proteins are α-proteobacterial in origin. They function mainly in energy metabolism (Krebs cycle and aerobic respiration) and translation. The authors were, however, surprised that as many as 208 proteins of the yeast mitoproteome have no apparent homologs among prokaryotes. They were referred to as specifically eukaryotic proteins [13]. It may well be, however, that some, or even many, of these proteins descended from a mitochondrial progenitor, but changed during coevolution of the host and endosymbiont to such an extent that they can no longer be recognized as α-proteobacterial in origin. A prime example may be accessory proteins of respiratory complexes and additional constituents of ribosomes. The proteins with transport functions deserve special attention, because this category comprises the smallest number of proteins with prokaryotic homologs [103]. The best example of a protein that has undergone minor changes is Atm1, a transporter of iron-sulfur clusters. True to expectations, Atm1-based phylogenetic reconstruction showed a sisterhood of mitochondria and *R. prowazekii* [13]. Another example, mitochondrial protein translocase Oxa1p, reflects an intermediate situation. There is little doubt that its ortholog is bacterial YidC [104], also present in Rickettsiaceae ([9,63] and unfinished genomes). There is even little doubt that a phylogeny of Oxa1p/YidC would have revealed an affiliation of mitochondria with rickettsiae. Unfortunately, poor homology of Oxa1p and YidC impedes phylogenetic analysis. Finally, an instance of not merely (dramatic) changes but of full replacement is the ATP/ADP carrier (AAC). It has been suggested [34] that the bacterial carrier protein, found only in obligate intracellular *Rickettsia* and *Chlamydia* [9,105], originated in rickettsia-like endosymbionts or was acquired by them from chlamydiae, and played a pivotal role in the establishment of mitochondrial symbiosis. Like mitochondrially encoded Cox1 [106], this bacterial inner membrane protein contains 12 transmembrane domains, and therefore might have been unimportable across the outer membrane subsequent to gene transfer from the rickettsia-like endosymbiont to the host genome in the course of mitochondrial origin. This rickettsial-type AAC was therefore suggested [34] to have been replaced by an unrelated mitochondrial carrier with six transmembrane domains in each of two subunits [107]. The latter is a member of the mitochondrial carrier family of tripartite proteins [107], the single repeat of which might in principle have derived from some of the rickettsial-like carriers. These have been suggested to have evolved during a long-term symbiotic relationship between the intracellular bacterium and the pro-eukaryote [34].

In summary, various changes in the course of mitochondrial origin are believed to represent the very first stage of a global evolutionary event, the conversion of an amitochondriate pro-eukaryote into a fully fledged mitochondriate eukaryote.

Typically eukaryotic traits probably emerged subsequent to the origin of the mitochondrion

Characteristically eukaryotic proteins

Prokaryote to eukaryote transition first resulted in the appearance of such subcellular structures as the nucleus with multiple chromosomes, endomembrane system, and cytoskeleton [17,25–29]. The question was addressed of whether these features emerged before or after the advent of the mitochondrion. As stated above, a sister relationship of Rickettsiales and Eukarya exclusive of free-living α-proteobacteria, revealed in phylogenetic analysis of a particular protein, may be taken as evidence that the eukaryotic compartment, necessarily involving this protein, originated after an endosymbiotic event.

A study initially focused on specifically eukaryotic proteins, which have, nevertheless, highly homologous orthologs among the prokaryotes. In this regard, two proteins, which are also present in the *R. prowazekii* proteome, seemed attractive [9]. These are Sec7, an essential component of the Golgi apparatus [105], and adducin, a protein that plays a part in F-actin polymerization [108]. An exhaustive search for finished and unfinished prokaryotic genomes revealed that Sec7 is a feature of *R. prowazekii*. Interestingly, Sec7 is lacking in *R. conorii*, another species of the genus *Rickettsia* [63]. It may be therefore that this case represents reverse LGT, i.e. from Eukarya to rickettsia [105]. An alternative view that Sec7 was produced by a

rickettsia-like endosymbiont and transferred to eukaryotes via a mitochondrial progenitor cannot be ruled out, however. Adducin is a modular protein composed of an N-terminal globular (head) domain, and extended central and C-terminal domains [108]. Phylogenetic analysis after a careful search for databases revealed that the head domain, also known as class II aldolase, emerged via paralogous duplication of the quite widespread fuculose aldolase and transferred to eukaryotes and rickettsiae from free-living α-proteobacteria. However, adducin *per se* seems to be characteristic only of animals, including *Drosophila* and *Caenorhabditis elegans*. These data imply that this cytoskeletal protein may be dispensable in lower eukaryotes, albeit its presence in protists cannot be excluded. Of interest, *S. cerevisiae* lacks adducin, whereas *Schizosaccharomyces pombe* (unfinished genome) probably bears the head domain alone, i.e. class II aldolase, which is monophyletic with the head domain of eukaryotic adducins (V.V. Emelyanov, unpublished data).

Compartment-specific paralogous families of conserved proteins

According to Gupta and associates [21,23,109], duplication of the genes encoding eukaryotic (i.e. nucleocytoplasmic) heat shock proteins (Hsp40, Hsp70, and Hsp90) that gave rise to cytosolic and ER isoforms may have accompanied the origin of ER. While mitochondrial and mitochondrial-type Hsp70s are thought to have derived from a rickettsia-like progenitor of the organelle (see below), the origin of nucleocytoplasmic proteins remains obscure. As indicated by the presence of a characteristic insertion (indel) in the N-terminal quadrant of proteobacterial and eukaryotic homologs, which is lacking in Hsp70 of archaea and Gram-positive bacteria, as well as in its distant paralog MreB, eukaryal proteins derive from proteobacteria. This inference is also supported by other sequence signatures [21,23]. In contrast, phylogenetic analysis failed to establish with confidence the position of cytosolic and ER sister groups among eubacterial phyla. It is only clear from these data that paralogous duplication of Hsp70 occurred early in eukaryotic evolution, and that monophyletic eukaryotic clade may not be considered an outgroup given the presence of the above insert to be a derived state [23]. On the basis of a four-amino-acid insert that is uniquely present in β and γ proteobacteria, the latest diverging proteobacterial groups [110], Gupta [23] concluded that the donor taxon of eukaryotic Hsp70 must have been the α, δ, or ε subdivision. Thus, one may suggest (see also [111]) that paralogous ER and cytoplasmic Hsp70s are descended from an endosymbiont homolog. (No cases of δ and ε proteobacterial contributions to eukaryotes have been found: see, e.g., Figure 2.) If so, the ER itself might have originated subsequent to mitochondrial origin (see the Introduction). This might have occurred during quite rapid conversion of a pro-eukaryote into a fully developed eukaryote via tandem duplication of an endosymbiont gene followed by rapid speciation of two copies destined to the cytoplasm and ER. However, the possibility cannot be ruled out that nucleocytosolic Hsp70 appeared in Eukarya via a primary fusion event involving a lineage leading to β/γ-proteobacteria, in which the characteristic four-amino-acid insert originated after fusion but before diversification of β and γ proteobacteria. Consistent with this idea, thorough indel analysis showed that neither a β nor a γ proteobacterium could be a fusion partner [110].

Like the situation for Hsp70, the phyletic position of paralogous cytosolic and ER isoforms of Hsp40 and Hsp90, which also originated via ancient duplications [23,109], was proven to be uncertain ([112] and unpublished results). Only one indel was found within a moderately conserved region of Hsp90 sequences which may indicate the evolutionary origin of the above two eukaryotic heat shock proteins (Fig. 4). This observation still suggests that nucleocytosolic Hsp90 may have derived from an α-proteobacterial ancestor of mitochondria [112].

Recent phylogenetic analysis of eukaryotic protein disulfide isomerases discerned a complex evolutionary history of these enzymes catalyzing disulfide bond formation during protein trafficking across ER. The nearest relatives of eukaryotic proteins, including as many as five *G. lamblia* paralogs, were shown to be prokaryotic and eukaryotic thioredoxins [113]. These data encouraged the phylogenetic analysis of thioredoxins by using the sequences from a broad variety of prokaryotic taxa. Curiously, eukaryal thioredoxins were shown to group with chlamydial ones. Far-reaching conclusions are, however, difficult to reach because of the small protein size (82 alignable positions) and low bootstrap support for this relationship (V. V. Emelyanov, unpublished observations).

As pointed out above, the appearance of ER-specific proteins by means of paralogous multiplication may indicate the origin of ER *per se*. Similarly, multiplication of the enzymes of DNA metabolism may be tied to the origin of the nucleus with multiple chromosomes. A case in point is the multigene family of eukaryotic MutS-like (MSH) proteins. This group of DNA mismatch repair enzymes consists of at least six paralogous members. Among them, MSH1 is the mitochondrial form, and MSH4 and MSH5 are specific to meiosis ([114] and references therein). Curiously, the MutS (MSH1) gene was reported to persist in the mitochondrial genome of octocoral *Sarcophyton glaucum*, a possible relic linking a mitochondrial symbiont with a nucleocytosolic MSH family [115]. It was recently shown that nucleocytosolic MSHs constitute a monophyletic clade, with MSH1 of yeast and MutS of *R. prowazekii* being their closest relatives [114]. In this work, however, data sets included a limited number of eubacterial sequences. In particular, α-proteobacteria were represented by only *R. prowazekii*. Figure 5A shows the results of phylogenetic analysis of the MSH/MutS family involving all α-proteobacterial sequences known to date. Of the MSHs, only the least deviant MSH1 from *Sch. pombe* and *S. cerevisiae* was included. Given that an alignment of diverse MSHs is somewhat problematic [114], the use of only mitochondrial proteins allowed properly alignment of as many as 558 positions. A relationship of mitochondrial and α-proteobacterial enzymes was also supported by two sequence signatures (Fig. 5B). Bearing in mind the canonical pattern of endosymbiotic ancestry, it is clear from these and published data [114,116] that the origin of mitochondria predated the origin of the multigene MSH family. Importantly, a gene encoding MSH2 was recently characterized for the kinetoplastid *Trypanosoma cruzi* [116].

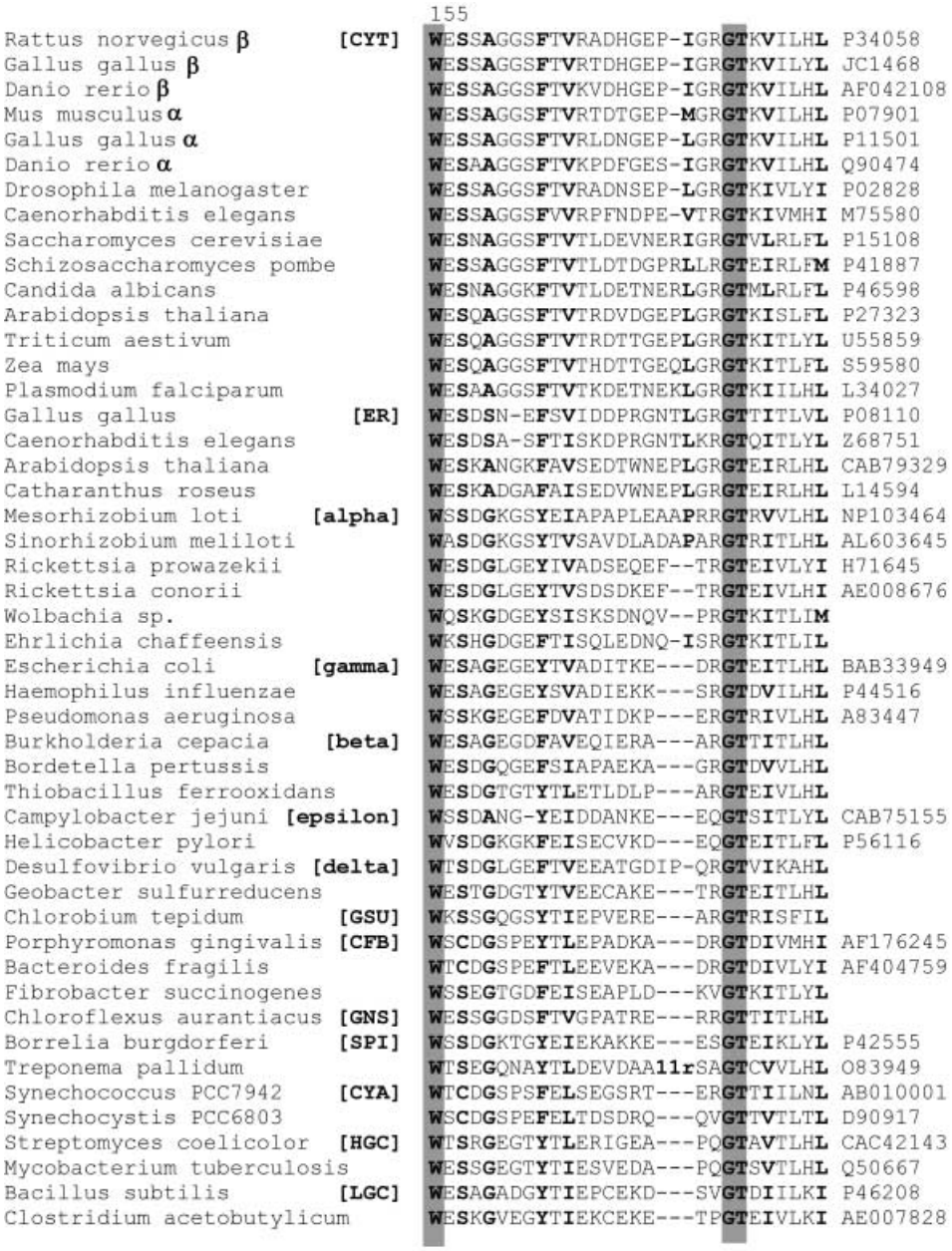


Fig. 4. Excerpt from the Hsp90 sequence alignment showing an insert that is present mostly in eukaryotic and α-proteobacterial homologs. It should be noted that Archaea and many eubacterial species including α-proteobacteria *Agrobacterium tumefaciens* and *C. crescentus* lack the *htpG* gene encoding Hsp90 [112]. It can be seen from alignment that rickettsial, animal cytoplasmic, and other eukaryotic plus α-proteobacterial homologs contain an insert one, two, and three residues in length, respectively. Only some representatives of β/γ-proteobacteria, cyanobacteria, and Gram-positive bacteria are shown. Of the two δ-proteobacterial sequences known to date, one contains a two-amino-acid insert. Like *T. pallidum*, *T. denticola* (unfinished genome, not shown) has an 11-residue insert whereas *Borrelia burgdorferi* does not. Essentially incomplete sequences from unfinished genomes of the free-living α-proteobacteria are not shown. Among them, *Magnetospirillum magnetotacticum* apparently lacks the insert, and *Rhodopseudomonas palustris* has a five-amino-acid insert. The number at the top refers to position in the *Mesorhizobium loti* sequence. Accession numbers are placed at the end of the alignment. If not present, the sequences were retrieved from unfinished genomes (TIGR). Other details are as in Fig. 3A. Abbreviations: CYT, cytoplasm; ER, endoplasmic reticulum; GSU, green sulfur bacteria; GNS, green nonsulfur bacteria; CFB, Cytophaga–Fibrobacter–Bacteroides group; SPI, spirochaetes; CYA, cyanobacteria; HGC and LGC, Gram-positive bacteria with high and low G + C content.

Kinetoplastids are known to be among the earliest emerging mitochondriate protists [25]. On the basis of these data, the following scenario for the origin of the nucleus can be proposed. A host for the mitochondrial symbiont was a chimeric prokaryote, and as such possessed a single *MutS* gene acquired from a eubacterial fusion partner (Archaea lack MutS [114]). During mitochondrial origin, the endosymbiont gene (occasionally) replaced this pre-existing gene,

giving rise to the paralogous MSH family, the diversification of which accompanied the origin of the nucleus. An alternative scenario would be the following. A host for the mitochondrion was a eukaryote with the true nucleus. Thus, like present-day eukaryotes, it possessed several *MutS*-related genes. Subsequently, an endosymbiont gene was introduced, giving rise to the (observed) MSH family. Thereafter, several pre-existing MutS-like proteins, which were still adapted to function in the (already existing) nucleus, were simultaneously lost. The absurdity of this scenario is apparent.

With respect to linear chromosome origin, telomere-like retroelements have to date been reported only in two linear mitochondrial plasmids of a primitive fungus *Fusarium oxysporum*. These data suggest that mitochondrial structures may be an evolutionary antecedent of eukaryotic telomeres [117].

Collectively, the present data argue that typically eukaryotic compartments, such as the nucleus with multiple linear chromosomes and the ER, probably originated after mitochondrial symbiosis.

Secondarily amitochondriate nature of archezoa

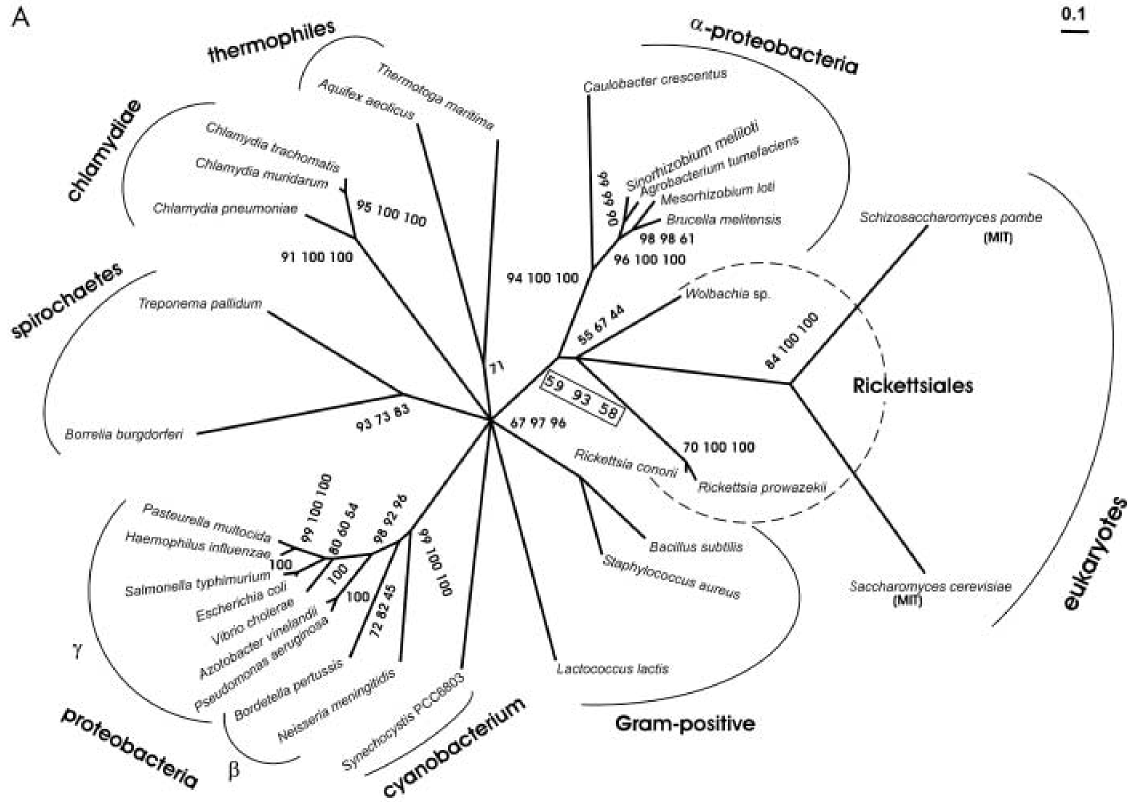
Mitochondrial-like proteins in amitochondriate protists

The archezoa hypothesis emerged several decades ago as the favored model of eukaryogenesis, and continues to have a major influence on the field [16,17,25–29]. As usually described, the concept of the Archezoa, primitively amitochondriate eukaryotes, implies that the very first eukaryotic cell (archezoon) possessed at least an endomembrane system and cytoskeleton. Thus, it had an advantage over other biota because it could engulf and digest surrounding microorganisms. In the framework of this concept (Fig. 1A), the eubacterial progenitor of mitochondrion escaped from being completely digested by some enigmatic mechanism, and subsequently gave rise to an energy-generating organelle (e.g. [29]). Three groups of amitochondriate protists, Microsporidia, Parabasalia, and Diplomonada (Metamonada), have long been considered as the candidates for Archezoa. Indeed, some trees have consistently shown these groups to have emerged before mitochondrion-bearing eukaryotes [15–17,25,118,119]. However, accumulating molecular data have challenged the Archezoa concept, instead raising the possibility that these provisional archezoons are secondarily without mitochondria [15–17,120]. The most compelling evidence is based on phylogenetic analysis of the mitochondrial-like heat shock proteins, Cpn60 and Hsp70, reported in several representatives of the above protist phyla [22,31,34–36,38,40–42]. In addition, a mitochondrial-like iron-sulfur cluster assembly protein (IscS) was recently characterized for the diplomonad *G. lamblia* and the parabasalid *T. vaginalis* [43]. As this study involved a single α-proteobacterium, *R. prowazekii*, extended phylogenetic analysis of the IscS gene and protein was performed which revealed a full canonical pattern of mitochondrial ancestry (V.V. Emelyanov, unpublished data). It is notable that both the eukaryotic IscS and Atm1 are mitochondrial in origin (see above). As to Microsporidia, these were argued to be, in effect, highly derived fungi [36,92,121–124]. Parabasalia, for example, the most well-studied *T. vaginalis*, lack mitochondria but instead possess hydrogenosomes [19,25,31,35, 119,125, 126]. It is now becoming evident that the latter are biochemically modified mitochondria [119,125,127–139]. Assuming an irreversible conversion of mitochondria or their ancestor into hydrogenosomes [34], Parabasalia are in no way primitively amitochondriate protists.

Transient symbioses?

However, after a period of enthusiasm, new data appeared that called into question the idea about the secondarily amitochondriate nature of Diplomonada. First, an archaeal-type alanyl-tRNA synthetase (AlaRS) was reported in *G. lamblia*, in contrast with other eukaryotes which possess closely related cytoplasmic and mitochondrial enzymes of bacterial genre [140–142]. These data suggested that diplomonads may have experienced abortive mitochondrial symbiosis, in which only some genes (e.g. *Cpn60*) may have been transferred to the nuclear genome. Diplomonads thus preserved archaeal AlaRS, while the full establishment of the organelle in the remaining eukaryotes resulted in replacement of the pre-existing archaeal gene with an endosymbiont ortholog, which gave rise to both cytoplasmic and mitochondrial forms [141]. Another explanation is, however, possible. One may suggest that ancient eukaryotes, such as Diplomonada, preserved both archaeal and eubacterial AlaRS for some time after the advent of the mitochondrion. The loss of this organelle in diplomonads was accompanied by the eventual loss of eubacterial-derived enzymes, whereas the stable presence of the mitochondrion in other eukaryotic lineages resulted in the loss of archaeal-derived enzymes. As shown in the same work [140,142], along with canonical CysRS, *G. lamblia* bears archaebacterial dual-specificity ProCysRS. These observations imply that archaeal proteins may persist in primitive protist lineages [93]. The situation would be reminiscent of the preservation of fermentation enzymes in mitochondriate eukaryotes [83,84].

Secondly, eubacterial-like Hsp70 (DnaK) of *G. lamblia* was recently sequenced as part of a genome sequencing project [143]. Phylogenetic analysis showed variable positions of giardial Hsp70, which sometimes diverged before α-proteobacterial orthologs. These data were interpreted as supporting the idea, proposed by Sogin [144,145], that some primitive eukaryotes may have undergone cryptic endosymbiosis or harbored an endosymbiont related to a mitochondrial progenitor. The present reanalysis (Fig. 6) used similar methods but slightly different input data. An alignment involved a smaller number of eukaryotic (both nucleocytoplasmic and mitochondrial/mitochondrial-like) entries. Also, the highly divergent *Ehrlichia* sp. (HGE agent) sequence was replaced by a less divergent sequence from *Wolbachia* (unfinished genome). Not unexpectedly, the eubacterial/mitochondrial-type sequence of *G. lamblia* always grouped with the mitochondrial clade (see legend to Fig. 6). Although in most analyses the *Giardia* affiliation to fast evolving lineages may be caused by an LBA artefact [77,83,92,146], distance matrix analysis with maximum likelihood distances revealed the deepest rooting within the mitochondrial clade with bootstrap support of 45% (Fig. 6). Thus, there is no compelling reason to suggest that



B

		776 S S	
Saccharomyces cerevisiae	[MT]	GKSTFLRQNAIIVILAQIGCFVPCSKARVGIVDKLFSRVGSADDL	P25846
Schizosaccharomyces pombe		GKSTFLRQNAIISILAQIGSFVPASNARIGIVDQIFSRIGSADNL	Z99091
Rickettsia prowazekii	[alpha]	GKSTYLRQNAIITIIAQIGSFVPAKSAKIGVVDKIFSRIGAADDL	E71685
Rickettsia conorii		GKSTFLRQNAIIAIIAQIGSFVPAKSAKIGVVDKIFSRIGAADDL	AAL02939
Wolbachia *sp.*		GKSTFLRQNALIAIILAHMGSFVPAESAHIGVIDKIFSRVGATDNI	
Ehrlichia chaffeensis		GKSTFLRQNALIGIILAHIGSFVPAQHAHIGVIDKVFSRVGASDNI	
Cowdria ruminantium		GKSTFLRQNALIGIILAHIGSFVPAEYAHIGVIDKVFSRVGASDNI	
Anaplasma phagocytophila		GKSTFLRQNALIAVLAHIGSFVPAEHAHIGVIDKIFSRVGASDNI	
Brucella melitensis		GKSTFLRQNALIAIILAQMGSFVPAGSAHIGVVDRLFSRVGASDDL	AAL52982
Sinorhizobium meliloti		GKSTFLRQNALIAIMAQTGSFVPAAAAHIGVVDRLFSRVGASDDL	P56883
Mesorhisobium loti		GKSTFLRQNALIAIILAQTGSFVPATSAHIGVVDRLFSRVGASDDL	BAB51800
Caulobacter crescentus		GKSTFLRQNALLAIILAQSGCYVPAASFRLGVVDRLFSRVGAGDDL	AAK22000
Agrobacterium tumefaciens		GKSTFLRQNALIAIILAQIGSFVPAEAAHIGVVDRLFSRVGASDDL	AAK86162
Rhodopseudomonas palustris		GKSTFLRQNALIALLAQVGSFVPAEAAHIGIVDRLFSRVGAADDL	
Magnetospirillum magnetotacticum		GKSTFLRQNAVIAIILAQMGSFVPAESVHMGVVDRLFSRVGAADDL	
Sphingomonas aromaticivoras		GKSTFLRQNALIVLLAQAGGFVPARSATVGLVDRLFSRVGASDNL	
Rhodospirillum rubrum		GKSTFLRQNALIAVLAQMGSFVPAESAEIGVIDRLFSRVGAADDL	
Escherichia coli	[gamma]	GKSTYMRQTALIALMAYIGSYVPAQKVEIGPIDRIFTRVGAADDL	P23909
Salmonella typhimurium		GKSTYMRQTALIALLAYIGSYVPAQNVEIGPIDRIFTRVGAADDL	U16303
Haemophilus influenzae		MRGKSTYQTALITLLAYIGSFVPADSARIGPIDRIFTRIGASDDL	P44834
Pasteurella multocida		GKSTYMRQTALITLMAYMGSFVPAESAVIGPIDRIFTRIGASDDL	P57972
Vibrio cholerae		GKSTYMRQTALIALMAHIGSYVPAESASIGPLDRIFTRIGASDDL	B82312
Azotobacter vinelandii		GKSTYMRQTALIVLLAHIGSFVPAQSCELSLVDRIFTRIGSSDDL	M63007
Pseudomonas aeruginosa		GKSTYMRQTALIVLLAHIGSFVPAARCELSLVDRIFTRIGSSDDL	B83193
Bordetella pertussis	[beta]	GKSTYMRQVALIALLARTGSFVPATRARVGRLDRIFTRIGAADDL	
Neisseria meningitidis		GKSTYMRQVALIVLLAHTGCFVPADAATIGPIDQIFTRIGASDDL	CAB83555
Borrelia burgdorferi	[spirochaetes]	GKSTYLRQVALITLMAHIGSFVPASKALIGITDKIFCRIGASDNI	O51737
Treponema pallidum		GKSTFLRQTALICLIAQVGSFVPAEKAELTPVDRIFCRVGAADNL	AAC65315
Chlamydia muridarum	[chlamydiae]	GKSTYIRQIALLVIMAQMGSFIPARSAHIGIIDKIFTRIGAGDNL	G81733
Chlamydia pneumoniae		GKSTYIRQIALLVIMAQMGSYIPAKSAHIGVIDKIFTRIGAGDNL	D81552
Chlamydia trachomatis		GKSTYIRQIALLVIMAQMGSFIPARSAHIGIVDKIFTRIGAGDNL	O84797
Aquifex aeolicus	[Aquificaceae]	GKSSYIRQVGVLTLLSHIGSFIPARRAKIPVVDALFTRIGSGDVL	O66652
Thermotoga maritima	[Thermotogales]	GKSTFIRQVGLISLMAQIGSFVPAQKAILPVFDRIFTRMGARDDL	U71155
Synechocystis	[cyanobacterium]	GKSCYLRQVGLIQLMAQTGSFIPAKTATLSICDRIFTRVGAVDDL	P73769
Bacillus subtilis	[gram-positive]	GKSTYMRQIALISIMAQIGCFVPAKKAVLPIFDQIFTRIGAADDL	C69663
Lactococcus lactis		GKSTYMRQFALTVIMAQIGSFVPAETANLPIFDAIFTRIGASDNL	AAK06308
Staphylococcus aureus		GKSTYMRQVAIISIMAQMGAYVPCKEAVLPIFDQIFTRIGAADDL	AF378369

Fig. 5. Phylogenetic analysis (A) and alignment (B) of MutS/MSH1 proteins. (A) ML tree (lnL = −21319.6). The tree was inferred using PUZZLE (see Fig. 2B) with parameter $\alpha = 1.18$ (weak rate heterogeneity) estimated from the data set. The DM method with ML distances was used as described in the legend to Fig. 2B (400 resamplings). MP analysis was performed on 400 resampled data using PROTPARS with the J option. Similar trees were obtained when employing the same three methods from other packages, and when using the input data that involved *E. chaffeensis* and different free-living α-proteobacteria (unfinished genomes). For other details see the legend to Fig. 2. Statistical tests were applied to a tree [(a,b) (c,d)] composed of monophyletic clusters: a, mitochondria + Rickettsiales; b, free-living α-proteobacteria; c, β/γ-proteobacteria; d, other phyla. Using PHYLTEST 2.0 with combined Poisson and gamma correction [79], an interior branch was shown to be significantly of nonzero length. Four-cluster likelihood mapping (PUZZLE) revealed that the occupancy of the area, representing above the tree, among the main three areas [66] was 91.3%. (B) The number above the alignment indicates the sequence position of *S. cerevisiae* MSH1. Two signatures distinguish mitochondria and α-proteobacteria from other species. Other details are as in Fig. 4.

Diplomonada acquired *dnaK* via a separate LGT, but not in the course of mitochondrial origination. Affiliation of *G. lamblia* mitochondrial-like Hsp70 with a mitochondrial cluster has also been reported [147].

Diplomonads and parabasalids

Notably, Hsp70-based phylogenetic analysis typically shows poor resolution of intergroup relationships [146]. In particular, a sisterhood of Rickettsiales and mitochondria was only rarely observed, being sensitive to species sampling (V.V. Emelyanov, unpublished data). Analysis of Cpn60, on the contrary, robustly showed not only branching of *G. lamblia* with a mitochondrial cluster, but also a sister-group relationship of rickettsiae and mitochondria [34,42,148]. Horner & Embley [148] recently reported that Cpn60 of *Spironucleus barkhanus*, another diplomonad, groups with the *G. lamblia* homolog deep in the mitochondrial clade. Unlike *Giardia*, its chaperonin contains an N-terminal extension similar to the mitochondrial-targeting sequence. This observation suggests that *S. barkhanus* may harbor a sort of remnant organelle resembling the crypton/mitosome described in secondarily amitochondriate *Entamoeba histolytica* [149,150].

The secondary absence of mitochondria in diplomonads is also strongly supported by the often observed sister relationship of *G. lamblia* and *T. vaginalis* [18,42,43,54,123,148,151]. Parabasalia may appear to be an even more ancient group than Diplomonada, as indicated by the presence of an indel in an enolase uniquely shared by *T. vaginalis* and prokaryotes to the exclusion of *G. lamblia* and other eukaryotes [152]. Taken together, these data argue for the secondary absence of mitochondria in diplomonads.

Relatively recent emergence of mitochondriate protists

In an attempt to determine the divergence time of Protozoa, the apparently paraphyletic nature of the lineage aside [25], Cpn60-based dating (see above) was extended by involvement of protist sequences. It appeared that the sequences from Kinetoplastida (*Trypanosoma brucei* and *Trypanosoma cruzi*) and Apicomplexa (*Plasmodium falciparum* and *Plasmodium yoelii*) passed a relative rate test involving animals, fungi, and several rickettsiae [96]. Using PHYLTEST 2.0 with both Poisson and gamma correction [79], these mitochondriate protists were shown to have emerged 1.45 ± 0.12 Bya, in contrast with 1.78 ± 0.17 Bya for the origin of mitochondria. One may be sceptical about both the molecular clock and the above estimates. Nevertheless, the present data are thought to be quite reliable for the following reasons. First, the Cpn60 sequences used throughout this study have been shown to exhibit clock-like (quasi-linear) behavior. Given a smooth difference between sequences, a choice of outgroup for the relative rate test was always straightforward. Secondly, the same event, i.e. divergence of animals and fungi [96], was taken as a reference timepoint for both measurements. Thirdly, time is nothing more than a measure and a convenient representation of linear or periodic processes. In other words, the above estimates may essentially be interpreted as relative genetic distances and definitely not as absolute time estimates. In view of these considerations, the first eukaryotes are thought to have emerged at some time within the above time frame. The big-bang hypothesis, based on phylogenetic data, assumes that all major eukaryotic lineages emerged during a short period of time [69,77,146]. One may suggest that this rapid diversification immediately followed the origin of mitochondria. In all probability, truly amitochondriate eukaryotes have not so far been found because they do not exist and have never even existed.

Conclusions

At least two hypotheses have been advanced that describe the host for the mitochondrial symbiont as a prokaryote. Both imply that the primitively amitochondriate host was a sort of archaebacterium [19,153]. According to Vellai *et al.* [153] only the establishment of an efficient energy-producing organelle made it possible for truly eukaryotic elements such as the nucleus with multiple chromosomes to develop. The main idea of the hydrogen hypothesis of Martin and Müller (Fig. 1B, left panel) is a syntrophy-based strict dependence of the host (obligately anaerobic methanogen) upon waste fermentation products (H_2 and CO_2) of the facultatively anaerobic symbiont (free-living α-proteobacterium capable of glycolysis, fermentation, and respiration). Briefly, the hydrogen hypothesis states that a methanogenic archaeon embraced a symbiont to a greater and greater extent to gain H_2 and CO_2 with maximum efficiency. This entailed transfer of genes for carrier proteins, which supply the symbiont with reduced organic compounds, to the host genome and incorporation of the carriers into the host envelope. Finally, glycolysis was relocated to the host cytoplasm, and methanogenesis vanished to prevent futile cycling of carbohydrates [19]. Although attractive in its premise, i.e. syntrophy, the hypothesis suffers several shortcomings. First of all, a time-course of this process, in which the host engulfs the symbiont, remains obscure. It is also difficult to understand how such a chimera might propagate. More generally, the hydrogen hypothesis requires that several unique evolutionary events (in effect, a symbiont completely engulfed by a

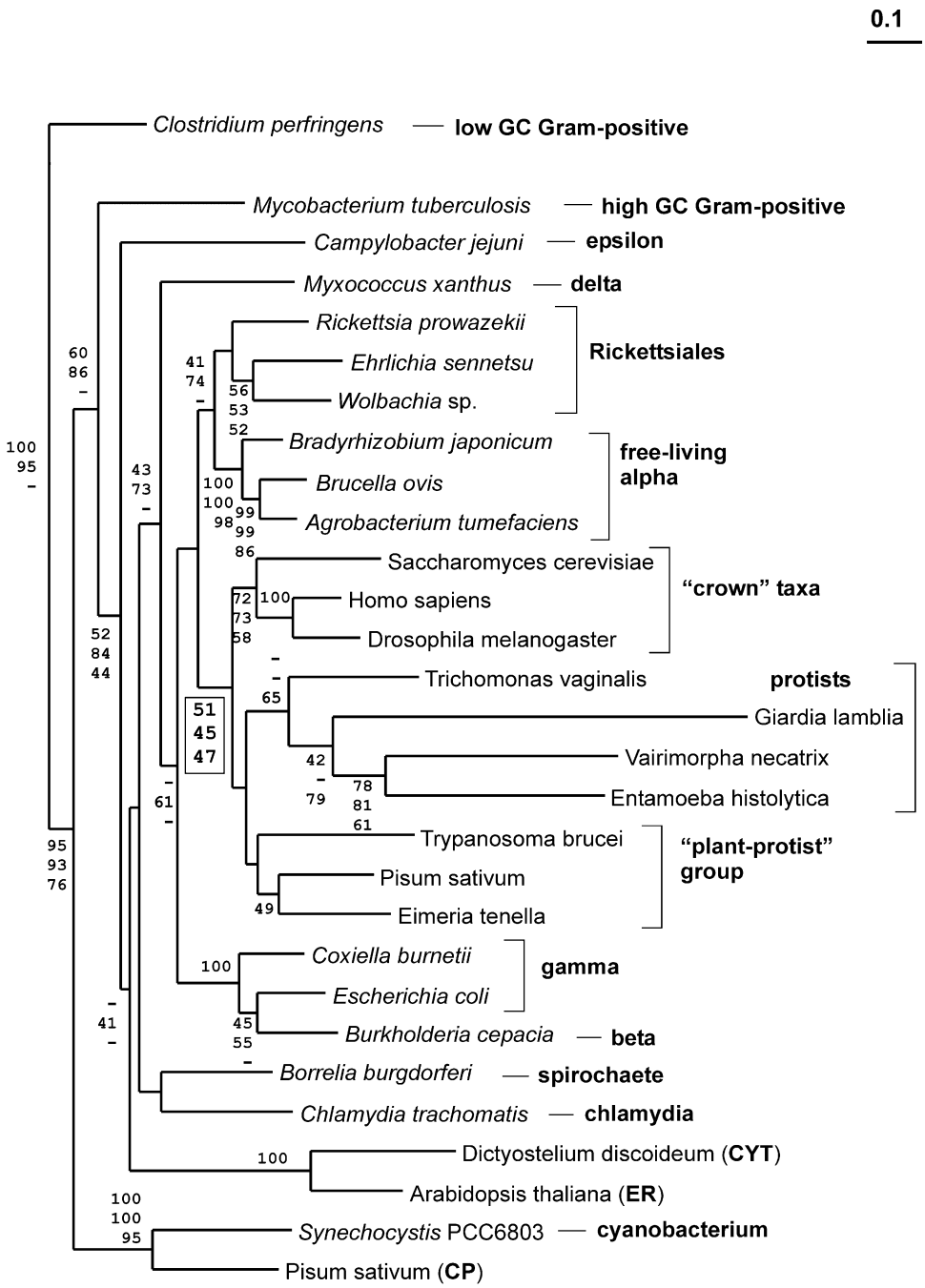


Fig. 6. ML majority consensus tree (lnL = −17895.2) based on global alignment of Hsp70 sequences. An archaebacterium *Methanosarcina mazei* (not shown) served as an outgroup. Phylogenetic analyses were conducted on 508 alignable positions. ML analysis was carried out on 400 bootstrap replicates as described in the legend to Fig. 2A. DM and MP methods were used as above (Fig. 2B) also with 400 resampled data. The gamma shape parameter used in distance analysis was $\alpha = 0.78$. BPs shown from top to bottom apply to ML, DM and MP trees, respectively. The MP tree (lnL = −17933.7) constrained for monophyly of mitochondrial/mitochondrial-like sequences excluding *G. lamblia* was not rejected by statistical tests. It is noteworthy that the sister relationship of a mitochondrial clade and α-proteobacteria exclusive of β/γ-proteobacteria on the MP trees constrained for monophyly of mitochondrial (including *Giardia*) sequences. Nonetheless, only weak support for a tree topology grouping mitochondrial cluster with α-proteobacteria to the exclusion of γ-proteobacteria and a fourth group represented by all other taxa was obtained in four-cluster analyses (see legend to Fig. 5A). It should be noted that trees similar to those depicted here were obtained using other programs. Furthermore, the relative branching order of γ-proteobacteria, α-proteobacteria, and a mitochondrial cluster including *G. lamblia* proved to be robust to exclusion from alignment of a proportion of constant sites, the category of fastest evolving sites [83,92,143], and idiosyncratic sites [96].

host, with its genes for carriers transferred to the host genome, carrier proteins retargeted to the host membrane, and glycolysis relocated to the host cytoplasm) have occurred simultaneously. Otherwise, any such commitment alone would be fatal to the emerging creature. Finally, the hydrogen hypothesis is not supported by the molecular data described in this review.

In contrast, eukaryogenesis is hypothesised here (Fig. 1B, right panel) which is thought to be consistent with most phylogenetic data. Successive steps towards the construction of the eukaryotic cell are briefly summarized in Fig. 7. A primarily amitochondriate organism (pro-eukaryote) emerged as a true chimera [23], with genetic apparatus acquired from an archaebacterium and core metabolism from a eubacterium. Bearing in mind the early origin of respiratory chains [106,154], the bacterial fusion partner must have been a facultative anaerobe, e.g. similar to the γ-proteobacterium *Escherichia coli*, capable of oxidative

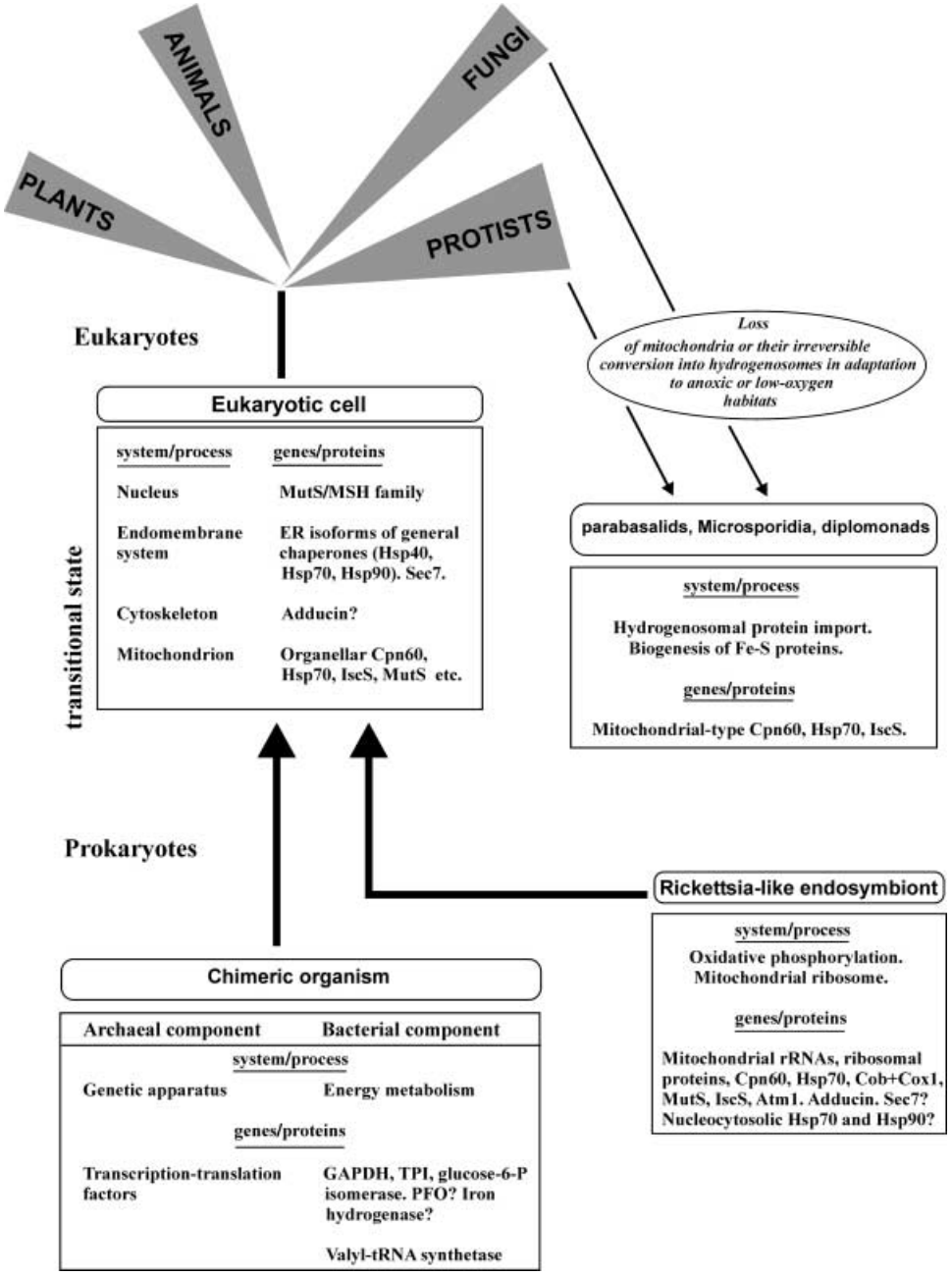


Fig. 7. Prokaryote–eukaryote transition and diversification of eukaryotes. As some parts of the eukaryotic tree (in particular, the basal position of the archezoan protists) cannot be resolved clearly [17,77], the branching order of the major eukaryotic groups is represented as a multifurcation. Shown in the boxes are some key systems and processes related to hypothetical evolutionary events, and protein/gene data supporting these events. Question marks indicate cases in which available molecular data are either equivocal or permit an equally parsimonious alternative interpretation (see text for details). Notably, cysteine desulfurase IscS (but not eubacterial-type Cpn60) was reported in a microsporidian *Encephalitozoon cuniculi* [122]. Phylogenetic analysis placed this organism inside the fungal clade (V. V. Emelyanov, unpublished data).

phosphorylation. In effect, it could resemble a mitochondrial progenitor in many respects. Such a similarity may, in particular, account for the otherwise enigmatic fact that as many as 75% of bacterial-like mitochondrial proteins are not endosymbiotic in origin [13,103]. Interestingly, this even pertains to some enzymes of the tricarboxylic acid cycle [13,20,81]. At that time (≈ 2.3 Bya), respiration would be inefficient, because of low oxygen concentration [155], and disappear, with oxygen-scavenging mechanisms being preserved. Thus, an advantage of the chimeric cell over Archaea could be more advanced energy metabolism (glycolysis and fermentation) and oxygen insensitivity, and a selective advantage over bacteria could be resistance to antibiotics produced by Gram-positive bacteria [24]. A sort of syntrophy [19,156] may also be invoked as a driving force that forged an amitochondriate host. Such a double advantage could have allowed the chimeric cell to rapidly propagate, having once become a target for a rickettsia-like α-proteobacterium. From the start, the latter possessed both anaerobic and aerobic energy pathways. Like modern rickettsiae [157], the symbiont entered the host because of membrane-bound phospholipase activity, with a single plasma membrane being subsequently repaired to prevent cytoplasm leakage. Curiously, an endosymbiotic relationship involving two different proteobacteria was recently reported [158]. Also of interest, rickettsiae are 3–5 times shorter than free-living bacteria such as *E. coli* [159]. The common evolutionary history of Rickettsiales and mitochondria (the first part of mitochondrial history) proceeded by loss of many genes specifying redundant metabolic pathways including glycolysis. The growing tension of atmospheric oxygen provided unique conditions under which the host glycolysis and symbiont oxidative phosphorylation were successfully combined. Domestication of the endosymbiotic bacterium by a former pro-eukaryote involved dramatic processes which resulted in the emergence of an ATP-producing organelle with most of its genes, directly or indirectly supporting respiration, being fixed in the host genome. This event initiated or merged with another global evolutionary event, creation of the true eukaryote with nucleus, endomembrane system, and cytoskeleton.

Acknowledgements

I would like to thank the head and staff of the joint supercomputer centre (JSCC, Moscow), who provided the opportunity to make computations.

References

1. Brown, J.R. & Doolittle, W.F. (1997) Archaea and prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**, 456–502.
2. Feng, D.-F., Cho, G. & Doolittle, R.F. (1997) Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl Acad. Sci. USA* **94**, 13028–13033.
3. Gupta, R.S. & Golding, G.B. (1996) The origin of the eukaryotic cell. *Trends Biochem. Sci.* **21**, 166–171.
4. Horiike, T., Hamada, K., Kanaya, S. & Shinozawa, T. (2001) Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. *Nat. Cell Biol.* **3**, 210–214.
5. Lake, J.A. & Rivera, M.C. (1994) Was the nucleus the first endosymbiont? *Proc. Natl Acad. Sci. USA* **91**, 2880–2881.
6. Ribeiro, S. & Golding, G.B. (1998) The mosaic nature of eukaryotic nucleus. *Mol. Biol. Evol.* **15**, 779–788.
7. Rivera, M.C., Jain, R., Moore, J.E. & Lake, J.A. (1998) Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244.
8. Adams, K.L., Daley, D.O., Qiu, Y.L., Whelan, J. & Palmer, J.D. (2000) Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature (London)* **408**, 354–357.
9. Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C.M., Podowski, R.M., Näslund, A.K., Eriksson, A.-S., Winkler, H.H. & Kurland, C.G. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature (London)* **396**, 133–140.
10. Gray, M.W., Burger, G. & Lang, B.F. (1999) Mitochondrial evolution. *Science* **283**, 1476–1481.
11. Lang, B.F., Gray, M.W. & Burger, G. (1999) Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397.
12. Gray, M.W. (2000) Mitochondrial genes on the move. *Nature (London)* **408**, 302–305.
13. Kurland, C.G. & Andersson, S.G.E. (2000) Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* **64**, 786–820.
14. Marcotte, E.M., Xenarios, I., van der Bliek, A.M. & Eisenberg, D. (2000) Localizing proteins in the cell from their phylogenetic profiles. *Proc. Natl Acad. Sci. USA* **97**, 12115–12120.
15. Embley, T.M. & Hirt, R.P. (1998) Early branching eukaryotes? *Curr. Opin. Genet. Dev.* **8**, 624–629.
16. Keeling, P.J. (1998) Kingdom's progress: Archezoa and the origin of eukaryotes. *Bioessays* **20**, 87–95.
17. Roger, A.J. (1999) Reconstructing early events in eukaryotic evolution. *Am. Nat.* **152**, S146–S163.
18. Hashimoto, T., Sanchez, L.B., Shirakura, T., Müller, M. & Hasegawa, M. (1998) Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. *Proc. Natl Acad. Sci. USA* **95**, 6860–6865.
19. Martin, W. & Müller, M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature (London)* **392**, 37–41.
20. Schnarrenberger, C. & Martin, W. (2002) 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.
21. Gupta, R.S., Aitken, K., Falah, M. & Singh, B. (1994) Cloning of *Giardia lamblia* heat shock protein HSP70 homologs: implications regarding origin of eukaryotic cells and of endoplasmic reticulum. *Proc. Natl Acad. Sci. USA* **91**, 2895–2899.
22. Gupta, R.S. (1995) Evolution of the chaperonin families (Hsp60, Hsp10 and Tcp-1) of proteins and the origin of eukaryotic cells. *Mol. Microbiol.* **15**, 1–11.
23. Gupta, R.S. (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* **62**, 1435–1491.
24. Gupta, R.S. (1999) Origin of eukaryotic cells: was metabolic symbiosis based on hydrogen driving force? *Trends Biochem. Sci.* **24**, 423.
25. Cavalier-Smith, T. (1998) A revised six-kingdom system of life. *Biol. Rev.* **73**, 203–266.
26. Cavalier-Smith, T. (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297–354.
27. Doolittle, R.F. (2000) Searching for the common ancestor. *Res. Microbiol.* **151**, 85–89.

28. Doolittle, W.F. (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet*. **14**, 307–311.
29. Doolittle, W.F. (1998) A paradigm gets shifty. *Nature (London)* **392**, 15–16.
30. Margulis, L., Dolan, M.F. & Guerrero, R. (2000) The chimeric eukaryote: origin of the nucleus from karyomastigont in amitochondriate protists. *Proc. Natl Acad. Sci. USA* **97**, 6954–6959.
31. Bui, E.T., Bradley, P.J. & Johnson, P.J. (1996) A common evolutionary origin for mitochondria and hydrogenosomes. *Proc. Natl Acad. Sci. USA* **93**, 9651–9656.
32. Clark, C.G. & Roger, A.J. (1995) Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc. Natl Acad. Sci. USA* **92**, 6518–6521.
33. Duncan, R., Faggart, M.A., Roger, A.J. & Cornell, N.W. (1999) Phylogenetic analysis of the 5-aminolevulinate synthase. *Mol. Biol. Evol*. **16**, 383–396.
34. Emelyanov, V.V. (2001) Evolutionary relationship of rickettsiae and mitochondria. *FEBS Lett*. **501**, 11–18.
35. Germot, A., Philippe, H. & LeGuyader, H. (1996) Presence of a mitochondrial-type 70-kDa heat shock protein in *Trichomonas vaginalis* suggests a very early mitochondrial endosymbiosis in eukaryotes. *Proc. Natl Acad. Sci. USA* **93**, 14614–14617.
36. Germot, A., Philippe, H. & LeGuyader, H. (1997) Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol. Biochem. Parasitol*. **87**, 159–168.
37. Gray, M.W. (1998) Rickettsia, typhus and the mitochondrial connection. *Nature (London)* **396**, 109–110.
38. Horner, D.S., Hirt, R.P., Kilvington, S., Lloyd, D. & Embley, T.M. (1996) Molecular data suggest an early acquisition of the mitochondrion endosymbiont. *Proc. R. Soc. London Ser. B* **263**, 1053–1059.
39. Olsen, G.J., Woese, C.R. & Overbeek, R. (1994) The winds of (evolutionary) change: breathing new life into microbiology. *J. Bacteriol*. **176**, 1–6.
40. Peyretaillade, E., Broussolle, V., Peyret, P., Metenier, G., Gouy, M. & Vivares, C.P. (1998) Microsporidia, amitochondrial protists, possess a 70-kDa heat shock protein gene of mitochondrial origin. *Mol. Biol. Evol*. **15**, 683–689.
41. Roger, A.J., Clark, C.G. & Doolittle, W.F. (1996) A possible mitochondrial gene in the early branching amitochondriate protist *Trichomonas vaginalis*. *Proc. Natl Acad. Sci. USA* **93**, 14618–14622.
42 Roger, A.J., Svärd, S.G., Tovar, J., Clark, C.G., Smith, M.W., Gillin, F.D. & Sogin, M.L. (1998) A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to progenitor of mitochondria. *Proc. Natl Acad. Sci. USA* **95**, 229–234.
43. Tachezy, J., Sanchez, L.B. & Müller, M. (2001) Mitochondrial type iron-sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as indicated by the phylogeny of IscS. *Mol. Biol. Evol*. **18**, 1919–1928.
44. Viale, A.M. & Arakaki, A.K. (1994) The chaperone connection to the origins of the eukaryotic organelles. *FEBS Lett*. **341**, 146–151.
45. Palmer, J.D. (1997) Organelle genomes: going, going, gone! *Science* **275**, 790–791.
46. Field, J., Rosenthal, B. & Samuelson, J. (2000) Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenase from anaerobic prokaryotes to *Entamoeba histolytica*. *Mol. Microbiol*. **38**, 446–455.
47. Figge, R.M., Schubert, M., Brinkmann, H. & Cerff, R. (1999) Glyceraldehyde-3-phosphate dehydrogenase gene diversity in eubacteria and eukaryotes: evidence for intra- and interkingdom gene transfer. *Mol. Biol. Evol*. **16**, 429–440.
48. Gerbod, D., Edgcomb, V.P., Noel, C., Vanacova, S., Wintjens, R., Tachezy, J., Sogin, M.L. & Viscogliosi, E. (2001) Phylogenetic relationships of class II fumarase genes from trichomonad species. *Mol. Biol. Evol*. **18**, 1574–1584.
49. Hannaert, V., Brinkmann, H., Nowitzki, U., Lee, J.A., Albert, M.-A., Sensen, C.W., Gaasterland, T., Müller, M., Michels, P. & Martin, W. (2000) 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*. **17**, 989–1000.
50. Henze, K., Horner, D.S., Suguri, S., Moore, D.V., Sanchez, L.B., Müller, M. & Embley, T.M. (2001) Unique phylogenetic relationships of glucokinase and glucose phosphate isomerase of the amitochondriate eukaryotes *Giardia intestinalis*, *Spironucleus barkhanus* and *Trichomonas vaginalis*. *Gene* **281**, 123–131.
51. Horner, D.S., Foster, P.G. & Embley, T.M. (2000) Iron hydrogenases and the evolution of anaerobic eukaryotes. *Mol. Biol. Evol*. **17**, 1695–1709.
52. de Koning, A.P., Brinkman, F.J., Jones, S.J. & Keeling, P.J. (2000) Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol. Biol. Evol*. **17**, 1769–1773.
53. Qian, Q. & Keeling, P.J. (2001) Diplonemid glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and prokaryote-to-eukaryote lateral gene transfer. *Protist* **152**, 193–201.
54. Wu, G., Henze, K. & Müller, M. (2001) Evolutionary relationships of the glucokinase from the amitochondriate protist, *Trichomonas vaginalis*. *Gene* **264**, 265–271.
55. Doolittle, R.F. (1998) Microbial genomes opened up. *Nature (London)* **392**, 339–342.
56. Faguy, D.M. & Doolittle, W.F. (1998) Cytoskeletal proteins: the evolution of cell division. *Curr. Biol*. **8**, R338–R341.
57. van den Ent, F., Amos, L.A. & Löwe, J. (2001) Prokaryotic origin of the actin cytoskeleton. *Nature (London)* **413**, 39–44.
58. Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J., Boistard, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S., Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Masuy, D., Pohl, T., Portetelle, D., Puhler, A., Purnelle, B., Ramsperger, U., Renard, C., Thebault, P., Vandenbol, M., Weidner, S. & Galibert, F. (2001) Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc. Natl Acad. Sci. USA* **98**, 9877–9882.
59. DelVecchio, V.G., Kapatral, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltsman, E., Selkov, E., Elzer, P.H., Hagius, S., O'Callaghan, D., Letesson, J.J., Haselkorn, R., Kyrpides, N. & Overbeek, R. (2002) The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. *Proc. Natl Acad. Sci. USA* **99**, 443–448.
60. Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B.S., Cao, Y., Askenazi, M., Halling, C., Mullin, L., Houmiel, K., Gordon, J., Vaudin, M., Iartchouk, O., Epp, A., Liu, F., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Flanagan, C., Crowell, C., Gurson, J., Lomo, C., Sear, C., Strub, G., Cielo, C. & Slater, S. (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* **294**, 2323–2331.
61. Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. & Tabata, S. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res*. **7**, 331–338.
62. Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Alley, M.R., Ohta, N.,

Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M.B., Khouri, H., Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L., Fraser, C.M. & Eisen, J. (2001) Complete genome sequence of *Caulobacter crescentus*. *Proc. Natl Acad. Sci. USA* **98**, 4136–4141.

63. Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.-E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.-M. & Raoult, D. (2001) Mechanisms of evolution in *Rickettsia conorii* and *Rickettsia prowazekii*. *Science* **293**, 2093–2098.
64. Adachi, J. & Hasegawa, M. (1996) MOLPHY, Version 2.3. *Comput. Sci. Monogr.* **28**.
65. Felsenstein, J. (1999) *PHYLIP*, *Phylogeny Inference Package*, Version 3.6. University of Washington, Seattle, WA.
66. Strimmer, K. & von Haeseler, A. (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl Acad. Sci. USA* **94**, 6815–6819.
67. Swofford, D.L. (1998) *PAUP*, Phylogenetic Analysis Using Parsimony (*and other methods)*, Version 4.0. Sinauer Associates, Sunderland, MA.
68. Brinkmann, H. & Philippe, H. (1999) Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. *Mol. Biol. Evol.* **16**, 817–825.
69. Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Müller, M. & Le Guyader, H. (2000) Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc. R. Soc. London Ser. B* **267**, 1213–1221.
70. Archibald, J.M., Logsdon, J.M. jr. & Doolittle, W.F. (2000) Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. *Mol. Biol. Evol.* **17**, 1456–1466.
71. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. (1989) Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. USA* **86**, 9355–9359.
72. Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T., Konishi, J., Denda, K. & Yoshida, M. (1989) Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **86**, 6661–6665.
73. Gribaldo, S. & Cammarano, P. (1998) The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. *J. Mol. Evol.* **47**, 508–516.
74. Boucher, Y. & Doolittle, W.F. (2000) The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol. Microbiol.* **37**, 703–716.
75. Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* **284**, 2124–2129.
76. Martin, W. (1999) Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *Bioessays* **21**, 99–104.
77. Philippe, H., Germot, A. & Moreira, D. (2000) The new phylogeny of eukaryotes. *Curr. Opin. Genet. Dev.* **10**, 596–601.
78. Reyes, A., Pesole, G. & Saccone, C. (2000) Long-branch attraction phenomenon and the impact of among-site rate variation on rodent phylogeny. *Gene* **259**, 177–187.
79. Kumar, S. (1996) *PHYLTEST: a program for testing phylogenetic hypotheses*. Pennsylvania State University.
80. Shimodaira, H. & Hasegawa, M. (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247.
81. Baughn, A.D. & Malamy, M.H. (2002) A mitochondrial-like aconitase in the bacterium *Bacteroides fragilis*: implications for the evolution of the mitochondrial Krebs cycle. *Proc. Natl Acad. Sci. USA* **99**, 4662–4667.
82. Henze, K., Morrison, H.G., Sogin, M.L. & Müller, M. (1998) Sequence and phylogenetic position of a class II aldolase gene in the amitochondriate protist, *Giardia lamblia*. *Gene* **222**, 163–168.
83. Horner, D.S., Hirt, R.P. & Embley, T.M. (1999) A single eubacterial origin of eukaryotic pyruvate:ferredoxin oxidoreductase genes: implications for the evolution of anaerobic eukaryotes. *Mol. Biol. Evol.* **16**, 1280–1291.
84. Horner, D.S., Heil, B., Happe, T. & Embley, T.M. (2002) Iron hydrogenases: ancient enzymes in modern eukaryotes. *Trends Biochem. Sci.* **27**, 148–153.
85. Liaud, M.-F., Lichtle, C., Apt, K., Martin, W. & Cerff, R. (2000) Compartment-specific isoforms of TPI and GAPDH are imported into diatom mitochondria as a fusion protein: evidence in favor of a mitochondrial origin of the eukaryotic glycolytic pathway. *Mol. Biol. Evol.* **17**, 213–223.
86. Nowitzki, U., Flechner, A., Kellermann, J., Hasegawa, M., Schnarrenberger, C. & Martin, W. (1998) Eubacterial origin of nuclear genes for chloroplast and cytosolic glucose-6-phosphate isomerase from spinach: sampling eubacterial gene diversity in eukaryotic chromosomes through symbiosis. *Gene* **214**, 205–213.
87. Voncken, F.G., Boxma, B., van Hoek, A.H., Akhmanova, A.S., Vogels, G.D., Huynen, M., Veenhuis, M. & Hackstein, J.H. (2002) A hydrogenosomal [Fe]-hydrogenase from the anaerobic chytrid *Neocallimastix* sp. L2. *Gene* **284**, 103–112.
88. Graham, D.E., Overbeek, R., Olsen, G.J. & Woese, C.R. (2000) An archaeal genomic signature. *Proc. Natl Acad. Sci. USA* **97**, 3304–3308.
89. Müller, M. & Martin, W. (1999) The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. *Bioessays* **21**, 377–381.
90. Keeling, P.J. & Doolittle, W.F. (1997) Evidence that eukaryotic triosephosphate isomerase is of alpha-proteobacterial origin. *Proc. Natl Acad. Sci. USA* **94**, 1270–1275.
91. Page, R.D.M. (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357–358.
92. Hirt, R.P., Logsdon, J.M. jr., Healy, B., Dorey, M.W., Doolittle, W.F. & Embley, T.M. (1999) Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96**, 580–585.
93. Suguri, S., Henze, K., Sanchez, L.B., Moore, D.V. & Müller, M. (2001) Archaebacterial relationships of the phosphoenolpyruvate carboxykinase gene reveal mosaicism of *Giardia intestinalis* core metabolism. *J. Euk. Microbiol.* **48**, 493–497.
94. Viscogliosi, E. & Müller, M. (1998) Phylogenetic relationships of the glycolytic enzyme, glyceraldehyde-3-phosphate dehydrogenase, from parabasalid flagellates. *J. Mol. Evol.* **47**, 190–199.
95. Canbäck, B., Andersson, S.G.E. & Kurland, C.G. (2002) The global phylogeny of glycolytic enzymes. *Proc. Natl Acad. Sci. USA* **99**, 6097–6102.
96. Emelyanov, V.V. (2001) Rickettsiaceae, rickettsia-like endosymbionts, and the origin of mitochondria. *Biosci. Rep.* **21**, 1–17.
97. Rotte, C., Stejskal, F., Zhu, G., Keithly, J.S. & Martin, W. (2001) Pyruvate:$NADP^+$ oxidoreductase from the mitochondrion of *Euglena gracilis* and from the apicomplexan *Cryptosporidium parvum*: a biochemical relic linking pyruvate metabolism in mitochondriate and amitochondriate protists. *Mol. Biol. Evol.* **18**, 710–720.
98. Nakazawa, M., Inui, H., Yamaji, R., Yamamoto, T., Takenaka, S., Ueda, M., Nakano, Y. & Miyatake, K. (2000) The origin of

pyruvate:$NADP^+$ oxidoreductase in mitochondria of *Euglena gracilis*. *FEBS Lett*. **479**, 155–157.
99. Chatton, B., Walter, P., Ebel, J.-P., Lacroute, F. & Fasiolo, F. (1988) The yeast VAS1 gene encodes both mitochondrial and cytoplasmic valyl-tRNA synthetases. *J. Biol. Chem.* **263**, 52–57.
100. Hsieh, S.-L. & Campbell, R.D. (1991) Evidence that gene G7a in the human major histocompatibility complex encodes valyl-tRNA synthetase. *Biochem. J.* **278**, 809–816.
101. Souciet, G., Menand, B., Ovesna, J., Cosset, A., Dietrich, A. & Wintz, H. (1999) Characterization of two bifunctional *Arabidopsis thaliana* genes coding for mitochondrial and cytosolic forms of valyl-tRNA synthetase and threonyl-tRNA synthetase by alternative use of two in-frame AUGs. *Eur. J. Biochem.* **266**, 848–854.
102. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
103. Karlberg, O., Canbäck, B., Kurland, C.G. & Andersson, S.G.E. (2000) The dual origin of the yeast mitochondrial proteome. *Yeast* **17**, 170–187.
104. Scotti, P.A., Urbanus, M.L., Brunner, J., de Gier, J.-W., von Heijne, G., van der Does, Driessen, A.J.M., Oudega, B. & Luirink, J. (2000) YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J.* **19**, 542–549.
105. Wolf, Y.I., Aravind, L. & Koonin, E.V. (1999) Rickettsiae and chlamydiae: evidence of horizontal gene transfer and gene exchange. *Trends Genet.* **15**, 173–175.
106. Castresana, J., Lübben, M., Saraste, M. & Higgins, D.G. (1994) Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. *EMBO J.* **13**, 2516–2525.
107. Saraste, M. & Walker, J.E. (1982) Internal sequence repeats and the path of polypeptide in mitochondrial ADP/ATP translocase. *FEBS Lett.* **144**, 250–254.
108. Matsuoka, Y., Li, X. & Bennett, V. (2000) Adducin: structure, function and regulation. *Cell. Mol. Life Sci.* **57**, 884–895.
109. Gupta, R.S. (1995) Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species. *J. Mol. Evol.* **12**, 1063–1073.
110. Gupta, R.S. (2000) The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol. Rev.* **24**, 367–402.
111. Gribaldo, S., Lumia, V., Creti, R., de Macario, E.C., Sanangelantoni, A. & Cammarano, P. (1999) Discontinuous occurrence of the hsp70 (dnaK) gene among Archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J. Bacteriol.* **181**, 434–443.
112. Emelyanov, V.V. (2002) Phylogenetic relationships of organellar Hsp90 homologs reveal fundamental difference to organellar Hsp70 and Hsp60 evolution. *Gene* **299**, 125–133.
113. McArthur, A.G., Knodler, L.A., Silberman, J.D., Davids, B.J., Gillin, F.D. & Sogin, M.L. (2001) The evolutionary origins of eukaryotic protein disulfide isomerase domains: new evidence from the amitochondriate protist *Giardia lamblia*. *Mol. Biol. Evol.* **18**, 1455–1463.
114. Culligan, K.M., Meyer–Gauen, G., Lyons-Weiler, J. & Hays, J.B. (2000) Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. *Nucleic Acids Res.* **28**, 463–471.
115. Pont–Kingdon, G., Okada, N.A., Macfarlane, J.L., Beagley, C.T., Watkins-Sims, C.D., Cavalier-Smith, T., Clark-Walker, G.D. & Wolstenholme, D.R. (1998) Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. *J. Mol. Evol.* **46**, 419–431.
116. Augusto-Pinto, L., Bartholomeu, D.C., Teixeira, S.M., Pena, S.D. & Machado, C.R. (2001) Molecular cloning and characterization of the DNA mismatch repair gene class 2 from the *Trypanosoma cruzi*. *Gene* **272**, 323–333.
117. Walter, T.C. & Kennell, J.C. (1999) Linear mitochondrial plasmids of *Fusarium oxysporum* are novel, telomere-like retroelements. *Mol. Cell* **4**, 229–238.
118. Cavalier-Smith, T. (1987) Eukaryotes with no mitochondria. *Nature (London)* **326**, 332–333.
119. Embley, T.M., Horner, D.S. & Hirt, R.P. (1997) Anaerobic eukaryote evolution: hydrogenosomes as biochemically modified mitochondria. *Trends Ecol. Evol.* **12**, 437–441.
120. Fast, N.M. & Keeling, P.J. (2001) Alpha and beta subunits of pyruvate dehydrogenase E1 from the microsporidian *Nosema locustae*: mitochondrion-derived carbon metabolism in microsporidia. *Mol. Biochem. Parasitol.* **117**, 201–209.
121. Baldauf, S.L., Roger, A.J., Wenk-Siefert, I. & Doolittle, W.F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977.
122. Katinka, M.D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J. & Vivares, C.P. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature (London)* **414**, 450–453.
123. Keeling, P.J., Luker, M.A. & Palmer, J.D. (2000) Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol. Biol. Evol.* **17**, 23–31.
124. Van de Peer, Y., Ben Ali, A. & Meyer, A. (2000) Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene* **246**, 1–8.
125. Hackstein, J.H., Akhmanova, A., Boxma, B., Harhangi, H.R. & Voncken, F.G. (1999) Hydrogenosomes: eukaryotic adaptations to anaerobic environments. *Trends Microbiol.* **7**, 441–447.
126. Müller, M. (1993) The hydrogenosome. *J. Gen. Microbiol.* **139**, 2879–2889.
127. Akhmanova, A., Voncken, F., van Alen, T., van Hoek, A., Boxma, B., Vogels, G., Veenhuis, M. & Hackstein, J.H. (1998) A hydrogenosome with a genome. *Nature (London)* **396**, 527–528.
128. Akhmanova, A., Voncken, F.G., Harhangi, H., Hosea, K.M., Vogels, G.D. & Hackstein, J.H. (1998) Cytosolic enzymes with a mitochondrial ancestry from the anaerobic chytrid *Piromyces* sp. E2. *Mol. Microbiol.* **30**, 1017–1027.
129. Bradley, P.J., Lahti, C.J., Plumper, E. & Johnson, P.J. (1997) Targeting and translocation of proteins into the hydrogenosome of the protist *Trichomonas*: similarities with mitochondrial protein import. *EMBO J.* **16**, 3484–3493.
130. Dyall, S.D. & Johnson, P.J. (2000) Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis. *Curr. Opin. Microbiol.* **3**, 404–411.
131. Embley, T.M., Finlay, B.J., Dyal, P.L., Hirt, R.P., Wilkinson, M. & Williams, A.G. (1995) Multiple origins of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. *Proc. R. Soc. London Ser. B* **23**, 87–93.
132. Embley, T.M. & Martin, W. (1998) A hydrogen-producing mitochondrion. *Nature (London)* **396**, 517–519.
133. van der Giezen, M., Sjollema, K.A., Artz, R.R., Alkema, W. & Prins, R.A. (1997) Hydrogenosomes in the anaerobic fungus *Neocallimastix frontalis* have a double membrane but lack an associated organelle genome. *FEBS Lett.* **408**, 147–150.
134. van der Giezen, M., Slotboom, D.J., Horner, D.S., Dyal, P.L., Harding, M., Xue, G.P., Embley, T.M. & Kunji, E.R. (2002)

Conserved properties of hydrogenosomal and mitochondrial ADP/ATP carriers: a common origin for both organelles. *EMBO J.* **21**, 572–579.
135. Hausler, T., Stierhof, Y.D., Blattner, J. & Clayton, C. (1997) Conservation of mitochondrial targeting sequence function in mitochondrial and hydrogenosomal proteins from the early-branching eukaryotes *Crithidia*, *Trypanosoma* and *Trichomonas*. *Eur. J. Cell Biol.* **73**, 240–251.
136. van Hoek, A.H., Akhmanova, A.S., Huynen, M.A. & Hackstein, J.H. (2000) A mitochondrial ancestry of the hydrogenosomes of *Nyctotherus ovalis*. *Mol. Biol. Evol.* **17**, 202–206.
137. Plumper, E., Bradley, P.J. & Johnson, P.J. (2000) Competition and protease sensitivity assays provide evidence for the existence of a hydrogenosomal protein import machinery in *Trichomonas vaginalis*. *Mol. Biochem. Parasitol.* **106**, 11–20.
138. Rotte, C., Henze, K., Müller, M. & Martin, W. (2000) Origins of hydrogenosomes and mitochondria. *Curr. Opin. Microbiol.* **3**, 481–486.
139. Van Hellemond, J.J., Opperdoes, F.R. & Tielens, A.G.M. (1998) Trypanosomatidae produce acetate via a mitochondrial acetate:succinate CoA transferase. *Proc Natl Acad. Sci. USA* **95**, 3036–3041.
140. Bunjun, S., Stathopoulos, C., Graham, D., Min, B., Kitabatake, M., Wang, A.L., Wang, C.C., Vivares, C.P., Weiss, L.M. & Söll, D. (2000) A dual-specificity aminoacyl-tRNA synthetase in the deep-rooted eukaryote *Giardia lamblia*. *Proc. Natl Acad. Sci. USA* **97**, 12997–13002.
141. Chihade, J.W., Brown, J.R., Schimmel, P.R. & Ribas de Pouplana, L. (2000) Origin of mitochondria in relation to evolutionary history of eukaryotic alanyl-tRNA synthetase. *Proc. Natl Acad. Sci. USA* **97**, 12153–12157.
142. Woese, C.R., Olsen, G.J., Ibba, M. & Söll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**, 202–236.
143. Morrison, H.G., Roger, A.J., Nystul, T.G., Gillin, F.D. & Sogin, M.L. (2001) *Giardia lamblia* expresses a proteobacterial-like DnaK homolog. *Mol. Biol. Evol.* **18**, 530–541.
144. Sogin, M.L. (1997) History assignment: when was the mitochondrion founded? *Curr. Opin. Genet. Dev.* **7**, 792–799.
145. Sogin, M.L. (1997) Organelle origins: energy-producing symbionts in early eukaryotes? *Curr. Biol.* **7**, R315–R317.
146. Germot, A. & Philippe, H. (1999) Critical analysis of eukaryotic phylogeny: a case study based on the HSP70 family. *J. Euk. Microbiol.* **46**, 116–124.
147. Arisue, N., Sanchez, L.B., Weiss, L.M., Müller, M. & Hashimoto, T. (2002) Mitochondrial-type hsp70 genes of the amitochondriate protists, *Giardia intestinalis*, *Entamoeba histolytica* and two microsporidians. *Parasitol. Int.* **51**, 9–16.
148. Horner, D.S. & Embley, T.M. (2001) Chaperonin 60 phylogeny provides further evidence for secondary loss of mitochondria among putative early-branching eukaryotes. *Mol. Biol. Evol.* **18**, 1970–1975.
149. Ghosh, S., Field, J., Rogers, R., Hickman, M. & Samuelson, J. (2000) The *Entamoeba histolytica* mitochondrion-derived organelle (crypton) contains double-stranded DNA and appears to be bound by a double membrane. *Infect. Immun.* **68**, 4319–4322.
150. Tovar, J., Fischer, A. & Clark, C.G. (1999) The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol. Microbiol.* **32**, 1013–1021.
151. Roger, A.J., Morrison, H.G. & Sogin, M.L. (1999) Primary structure and phylogenetic relationships of a malate dehydrogenase gene from *Giardia lamblia*. *J. Mol. Evol.* **48**, 750–755.
152. Keeling, P.J. & Palmer, J.D. (2000) Parabasalian flagellates are ancient eukaryotes. *Nature (London)* **405**, 635–637.
153. Vellai, T., Takacs, K. & Vida, G. (1998) A new aspect to the origin and evolution of eukaryotes. *J. Mol. Evol.* **46**, 499–507.
154. Castresana, J. & Saraste, M. (1995) Evolution of energetic metabolism: the respiration-early hypothesis. *Trends Biochem. Sci.* **20**, 443–448.
155. Knoll, A.H. (1992) The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622–627.
156. Lopez-Garcia, P. & Moreira, D. (1999) Metabolic symbiosis at the origin of eukaryotes. *Trends Biochem. Sci.* **24**, 88–93.
157. Winkler, H.H. (1990) *Rickettsia* species (as organisms). *Annu. Rev. Microbiol.* **44**, 131–153.
158. von Dohlen, C.D., Kohler, S., Alsop, S.T. & McManus, W.R. (2001) Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. *Nature (London)* **412**, 433–435.
159. Emelyanov, V.V. (1990) A simple method for counting rickettsiae. *Mikrobiol. Z. (USSR)* **52**, 91–95.