Origin and evolution of the mitochondrial proteome

Kurland, Charles; Andersson, S G E

Published in:
Microbiology and Molecular Biology Reviews

2000

MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Dec. 2000, p. 786–820 Vol. 64, No. 4
1092-2172/00/$04.00+0

Origin and Evolution of the Mitochondrial Proteome

C. G. KURLAND[1,2]* AND S. G. E. ANDERSSON[1]

Department of Molecular Evolution, Evolutionary Biology Centre, University of Uppsala, Uppsala SE 752 36,[1] *and Department of Microbiology, Lund University, Lund SE 223 62,*[2] *Sweden*

INTRODUCTION

Mitochondria are the ATP-generating organelles of eukaryotes, and in most organisms they are oxygen respiring. Roughly 2 billion years ago, the ambient oxygen tension of Earth's atmosphere increased rapidly. Here, rapidly means that the oxygen tension went from roughly 1% to more than 15% of present levels within less than 200 million years (88). Many believe that the origins of mitochondria as organelles in primitive eukaryotes can be associated with this environmental trauma (121).

Nevertheless, the Earth's atmosphere during the billions of years prior to this global oxygen shock was probably not the heavy reducing atmosphere suggested by Oparin (142). Geochemical evidence suggests that the oxygen tension in the atmosphere may have been as much as 1% of present levels from the very beginning (88, 157). In other words, during the entire history of the biosphere, oxygen was accessible at low levels in the atmosphere and quite possibly at higher levels locally. The continuous presence of oxygen matches the ancient origins of the terminal oxidases characteristic of mitochondria. Thus, the monophyletic lineage of cytochrome oxidases is well represented in the archaea, bacteria, and eukaryotes (40, 41, 104, 161, 166).

Phylogenetic reconstructions and distance measurements based on the sequences of cytochrome *c* oxidase and cytochrome *b* are consistent with divergence of mitochondria from bacteria between 1.5 and 2.0 billion years ago (165). Accordingly, the oxidative respiratory system that was introduced into eukaryotes by way of the primitive mitochondrion was already an ancient enzymatic system. There is now overwhelming support for the idea that the vehicle that introduced the respiratory system into the eukaryotic lineages was an endosymbiotic α-proteobacterium (20, 77, 78, 79).

The endosymbiotic theory of plastid as well of mitochondrial origins arose in the nineteenth century and was given new life by Margulis (121) precisely when molecular methods could begin to test some of its predictions. The discovery of mitochondrial genomes and the results of phylogenetic reconstructions with sequences for rRNA as well as for a few proteins strengthened confidence in this theory (32, 79, 206). As a consequence, when we reviewed the literature on codon preferences in this journal 10 years ago, we found it convenient to treat the mitochondrial genome as though it was just another kind of bacterial genome (12).

Since then, detailed comparisons and phylogenetic reconstruction with relevant genome sequences have very much expanded our view of the mitochondrion. Most informative have been the mitochondrial genomes of protists (77, 80, 111), the nuclear genome of the yeast *Saccharomyces cerevisiae* (87; http://www.proteome.com), and the genome of the α-proteobacterium *Rickettsia prowazekii* (20, 76). The genomic comparisons show unambiguously that the coding sequences of the mitochondrial genomes are predominantly the descendants of α-proteobacterial homologues. Accordingly, some version of the endosymbiotic theory is in all probability relevant to the origins of mitochondria. However, to account for some of the new data, this theory needs to be modified significantly.

First, it turns out that only a small fraction of all proteins functioning in mitochondria are the descendants of the ances-

* Corresponding author. Mailing address: Munkarpsv. 21, SE 243 32 HÖÖR, Sweden. Phone: 46 413 558747. Fax: 46 413 558746. E-mail: chuck@alpha2.bmc.uu.se.

tral free-living α-proteobacterium. Most of the remaining proteins are descendants of nuclear genes with no bacterial antecedents (17, 95). That most of the genes of the ancestral α-proteobacterium have disappeared from the mitochondrial genome has been understood for some time (13, 14, 15, 16, 19, 20, 22, 77). The magnitude of the loss can be estimated as follows. To our knowledge, the smallest genome of a free-living α-proteobacterium is that of *Bartonella henselae*, with less than 2×10^6 base pairs, and the largest is *Bradyrhizobium japonicum*, with 8.7×10^6 base pairs (107, 155). Since the *Bartonella* genome encodes 1,600 or more proteins (Andersson et al., unpublished), we can take this figure as a conservative size estimate for the proteome encoded by the free-living α-proteobacterial ancestor of mitochondria. What then accounts for the enormous size discrepancy between the coding capacity of α-proteobacterial and mitochondrial genomes? Here, we need to compare 1,600 proteins with the 67 proteins encoded by the mitochondrion with the largest coding repertoire, that of *Reclinomonas americana* (76, 77).

There are at least two large-scale reductive tendencies that will account for fact that contemporary mitochondrial genomes have evolved into mere vestiges of the ancestral genome. One is the massive loss of genes that are not essential to life in the eukaryotic cytosol. Thus, genes in the nuclear genome can replace many gene products originally encoded by ones in the endosymbiont's genome (6, 13, 15, 16, 108). This means that the suspension of purifying selection allows redundant genes in the mitochondria to be inactivated and deleted by random mutation (101). In addition, unique essential genes can be transferred to the nucleus if their protein products can be recruited from the cytosol for function in the mitochondrion (17, 27, 75, 95). A recent model for the evolution of mitochondrial genomes predicts that eventually, when such transfer is not destructive, all coding sequences will be displaced from the mitochondria to the nucleus (27).

Genes transferred to the nucleus can encode proteins that will be transported to the mitochondria by a specific transport system (138, 162). The same transport system can also assist mitochondria to recruit nonbacterial proteins encoded in the nucleus. For example, the nucleus of *Saccharomyces cerevisiae* contributes more than 400 proteins to the mitochondrion (87). Phylogenetic analysis suggests that half of the nuclear proteins that augment the mitochondrial proteome have no bacterial affinities (19, 95). This half is likely to be purely eukaryotic in origin (see below).

In effect, mitochondria have evolved in two distinctive modes. One is the reductive mode that reflects an extreme adaptation to an intracellular existence. The other is an expansive mode in which the mitochondria are the beneficiaries of nuclear evolution. In the following we document these modes of genome evolution. We stress the importance of extending the analysis of mitochondria beyond the relatively small and highly variable contribution of contemporary mitochondrial genomes. A narrow focus on the genomes of the organelle tends to obscure most of its evolutionary history. Since there are descendants of both ancient α-proteobacterial genes and more recent eukaryotic genes cooperating in mitochondrial functions, it is most convenient to view the evolution of the organelle as the evolution of a proteome. Viewed from this vantage, mitochondria no longer seem to be just another sort of bacteria.

Space limitations have meant that we cannot do justice to the vast amount of information that is available on many other aspects of the mitochondrial genome. We recommend that interested readers supplement their background information with the aid of an excellent book, *Mitochondrial Genomes* (205), edited by Wolstenholme and Jeon. This book contains a chapter by Michael W. Gray (75) that is a must.

Ox-Tox MODEL

The endosymbiotic theory, as explicated by Margulis (121), was an eclectic formulation that concerned much about cellular evolution besides the origins of mitochondria and plastids, but convention has reduced the common use of the term. By 1998, the standard model to describe the origins of mitochondria was quite specific: the endosymbiont was identified as an α-proteobacterium, such as *Paracoccus*, and the host as an archaeon (50, 75, 79, 137). An important aspect of the standard model in all of its shifting forms is that it assumes, often tacitly, that the symbiosis leading to the mitochondrial lineages involved an exchange of ATP produced aerobically by the symbiont for organics provided by the anaerobic host.

This view was challenged recently on biochemical grounds. In particular, it was recognized that a free-living bacterium such as *Paracoccus* would probably not be able to actively transport ATP to a prospective host because bacteria do not in general have ATP exporters (10, 20, 124, 185). Two additional observations, to which we return below, are relevant. First, only two endocellular parasitic bacteria are known to have ATP transport proteins. These are importers of ATP that are clearly related to plastid homologues but unrelated to the ATP exporters of mitochondria (10, 17, 20, 201). Second, the ATP transporters of mitochondria seem to have evolved after the divergence of eukaryotes (17, 20, 95). From where do the ATP transport functions of the mitochondria come?

It turns out that this question of seemingly small detail opens into much larger issues. In particular, if the initial symbiotic relationship between the α-proteobacterium and its host did not depend on the sharing of ATP produced through the aerobic respiration of the symbiont, what was that relationship? Two current views of the initial symbiotic relationship between the ancestor of mitochondria and its host have emerged. One sort of model favors an evolutionary path that is initially supported by anaerobic syntrophy (115, 124). The other, which involves aerobic mutualism (17), we describe first.

As noted above, all coding sequences of characterized mitochondria are found in the mitochondrial genome of *R. americana* (77). This, along with the clear similarity of these to coding sequences of *R. prowazekii* and *B. henselae* (20; Andersson et al., unpublished data), suggests that the mitochondria arose only once and that they, together with the putative α-proteobacterial ancestor, make up a monophyletic lineage. This is consistent with the supposition that the initial endosymbiotic relationship to the ancestral host was an aerobic one.

Nevertheless, there are two details of this putative aerobic scenario that are challenging. First, it is not possible at present to identify the host of the ancestral symbiont with any confidence. Current preferences vacillate between an archaeon and a primitive eukaryote as the likely host (see, for example, references 76 and 131). Either way, we assume for simplicity that the host was a heterotroph that could provide the endosymbiont with substrates such as pyruvate. This is not a drastic assumption because of the near ubiquity of glycolytic pathways among archaea, bacteria, and eukaryotes (46). In addition, it is known that cytochrome oxidases may function as cytochrome *c* oxidases, quinol oxidases, or nitrogen oxide (NO) oxidases. Nevertheless, all members of this gene family belong to the same monophyletic lineage, and all three may have been present in the last common ancestor (40, 41). This, together with the particularly close monophyletic relationship between the cytochromes employed by bacteria and by mitochondria, is

strong evidence that the ancestral endosymbiont had already acquired an aerobic respiratory chain (20, 77, 165). We return to the origins of eukaryotic heterotrophy below.

Second, as mentioned, neither transporters such as the ATP/ADP translocases of *Rickettsia* and *Chlamydia* nor protein transport systems such as those that recruit proteins into the mitochondria are found among free-living bacteria. This means that there is no reason to suppose that the ancestral endosymbiont could export ATP or import proteins, as do modern mitochondria. Instead, we suggest that initially the aerobic symbiont interacted with its host in ways that are not characteristic of modern mitochondria.

One possibility is that the ancestral α-proteobacterium was an aerobic symbiont that consumed oxygen with the aid of a respiratory chain ending in cytochrome oxidase and that, in return, its heterotrophic anaerobic host made pyruvate accessible (17, 95). Here, the host may not have benefited initially by sequestering the ATP produced aerobically by the symbiont. Instead, it is suggested that the consumption of oxygen per se constituted the service provided by the ancestral symbiont in the initial phase of the evolution of mitochondria. In effect, the cytochrome oxidase activity of the symbiont detoxified the host cytosol by converting oxygen to water. The benefit here is that elements of the host's anaerobic metabolism that were sensitive to oxygen would be protected by the activities of the endosymbiont's cytochrome oxidase.

The credibility of this conjecture derives from numerous examples of modern symbiotic relationships with an oxygen-scavenging function assigned to one of the partners (60, 61, 62). Roughly two billion years ago, the oxygen tension increased from less than 1.5% to greater than 15% of present levels (88). At that time, the demands for an oxygen-consuming symbiont to support an essentially anaerobic host would have been, if anything, more pressing than they are today. Thus, in modern organisms, activities such as peroxidases, catalases, and superoxide dismutases protect cells against the toxic effects of oxygen respiration. These activities might not have been so widespread two billion years ago. Indeed, even some modern cells are killed or debilitated by exposure to less than ambient atmospheric oxygen tensions (60, 61, 62).

In the Ox-Tox model, the evolution of the mitochondrion from the endosymbiont required the evolution of characteristic mitochondrial control and export functions that were derived from nuclear genes. The evolution of novel nuclear gene products for recruitment by the mitochondria is typified by the integration of the ATP/ADP translocase into the workings of the primitive mitochondrion. Thus, this activity is found universally in mitochondria, which dates its debut to a time prior to the divergence of the major branches of eukaryotes. This novel recruit to the mitochondrial proteome made possible an efficient supply of ATP to the host cell from the evolving mitochondrion. For this reason, the integration of the ATP/ADP translocase into the workings of the endosymbiont may be taken as a marker for the transformation of the endosymbiont into an organelle.

FROM BACTERIAL GENOME TO VESTIGE

One group of α-proteobacteria, the rickettsiae, are of special interest both as models for the evolution of mitochondria and as possible descendants of the endosymbiotic ancestor to mitochondria (11, 14, 58, 81). These organisms, like the putative ancestor of the mitochondria, are thought to be the descendants of free-living α-proteobacteria (191, 197). Furthermore, phylogenetic reconstructions for diverse protein-coding sequences suggest that rickettsiae are the closest modern relatives of the mitochondria (11, 81, 141, 166, 186). Once the genome sequence of *Rickettsia prowazekii* (20) and that of its close relative, *Bartonella henselae* (Andersson et al., unpublished data), became available, their intimate phylogenetic relationships to mitochondrial genomes became incontrovertible. Nevertheless, there is a glaring gap between, on the one hand, discovering this phylogenetic relationship and, on the other, understanding precisely what that ancestor was.

Furthermore, there is the enormous discrepancy between the number of coding sequences in mitochondria and that in free-living α-proteobacteria. Much has been made about the limited coding capacity of animal mitochondria (22, 204) as well as of the stark contrast between these and plant mitochondria with their relatively large, complicated genomic architectures (83, 180). Compared to the little rickettsial genome with its 834 protein-coding sequences (20), the coding capacity of any mitochondrion is insignificant. The numerical range of protein-coding sequences in mitochondria extends from 2 in *Plasmodium falciparum* to 67 in *Reclinomonas americana* (77). The protein-coding capacities of all known plant, animal, and fungal mitochondria are nested between those of these two protists. In fact, most mitochondrial genomes can boast between 12 and 20 or so protein-coding genes along with rRNA and tRNA genes—not much with which to run an organelle.

If an α-proteobacterium such as *Bartonella* with its 1,600 genes was the ancestor of mitochondria, why did nearly all of the coding capacity of this genome disappear from the organelle? Alternatively, why does the mitochondrial genome have any protein-coding capacity at all? Again, why are most of the genes needed by mitochondria found in the nucleus and not in the mitochondrial genome? Finally, are the nuclear mitochondrial genes of bacterial origin or of eukaryotic origin?

Genome Degradation

A free-living bacterium that initiates a symbiotic relationship with another cell will be bathed in the metabolic intermediates of its host. These metabolites make some of the symbiont's genes redundant as long as it shares the host metabolism. Thus, we expect some genes in the symbiont to be neutralized by the host's biochemical activities. Neutralized genes are subject to mutational degradation (101). When genes required for the free-living mode are forfeited, the facultative symbiont has evolved into an obligate symbiont or an obligate parasite, with a coding capacity that can be extremely limited (70, 71, 72, 94, 147, 170, 172). For example, the obligate parasites of the genus *Rickettsia*, like mitochondria, have virtually no genes for amino acid or nucleoside biosynthesis, but their facultative parasitic relatives, the *Bartonella* spp., are fully able to produce these intermediates in their free-living mode (20; Andersson et al., unpublished data).

The *Rickettsia*, like their relatives the mitochondria, have a well-developed oxidative metabolism that exploits the Krebs cycle along with an ATP-generating electron transport chain that terminates with cytochrome oxidase. Both sorts of genomes are devoid of genes for anaerobic glycolysis, and this may be attributed to the fact that their respective hosts supply them with pyruvate as the precursor to the Krebs cycle. In contrast, the *Bartonella* genome has a complete glycolytic repertoire (B. Canbäck, U. C. M. Alsmark, S. G. E. Andersson, and C. G. Kurland, unpublished data). In effect, a good deal of the difference in the gene complements of these two bacteria, which amounts to circa 1 million base pairs, may simply be the difference between the needs of an obligate and a facultative parasite. Likewise, the difference between the 834 genes of the *Rickettsia* genome and the roughly 400 genes that specify mi-

tochondrial functions (see below) may reflect differences in the needs of an infective parasite and those of a captive organelle.

These streamlining effects are to some extent a reflection of the population structure common to the genomes of obligate symbionts, obligate parasites, and cellular organelles (15). These sorts of genomes tend to propagate as asexual lineages, which are characterized by small population sizes. Under these conditions, sublethal deleterious mutations accumulate, and these may include the inactivation or loss of nonessential genes (15, 107, 117, 118). In the case of mitochondria, such mutations will be subject to purifying selection at the cellular level (28). However, the efficiency of this selection will depend on the population size and the magnitude of the selective disadvantage of the mutations. As a consequence, asexual genomes in small populations or in populations subject to recurrent bottlenecks will tend to be degraded by the inroads of weakly deleterious mutations (59), i.e., by Muller's ratchet. The ratchet has been demonstrated experimentally in bacteriophages (43) and in free-living bacteria (6). The influence of Muller's ratchet has also been inferred in the genomes of endosymbionts (35, 110, 130, 192) as well as in mitochondria (108, 117, 118).

Muller's ratchet may also account for some of the reduction in the effective gene complement of the evolving mitochondrion (59, 137). The magnitude of the influence of the ratchet on a genome is related to the degree of mutational diversity in the genome population. This follows from the fact that the mutation frequency per genome is likely to be proportional to the size of the genome. This implies that the larger genome characteristic of an early stage of mitochondrial evolution should have been more vulnerable to the inroads of Muller's ratchet than that of a modern mitochondrion (27).

So far we have considered genetic mechanisms that influence the size of the mitochondrial genome. We may also consider the molecular mechanisms that mediate the degradation of genomes. There are at least two different ways that sequences may be extirpated from a genome. One would be a slippage mechanism in which short runs of nucleotides are removed. This is a slow but sure way to delete sequences. Indeed, traces of this mechanism are observed in the highly derived genome of *Rickettsia*. For example, nearly one quarter of the *Rickettsia prowazekii* genome is noncoding sequence (7, 20). It is possible to study the mutation spectrum of noncoding sequences from different species of *Rickettsia*. Such a comparison shows that short deletions provide the dominant evolutionary mode in these sequences (8, 9). Thus, noncoding sequences that are thought to be mutation-degraded versions of nonessential coding sequences can slowly depart the genome by virtue of small deletions (7, 8, 9, 11, 15).

A more dramatic deletion mechanism and one that has left a more obvious signature on highly derived genomes is that attending intrachromosomal recombination at repeat sequences. This sort of recombination event has been observed as the most common mechanism of large-scale deletions in bacteria under laboratory conditions (143). Such deletions leave at least two signatures. First, they lead to the loss of intervening sequences between two repeat sequences along with the deletion of one of the repeats. Second, they lead to rearrangements of the flanking sequences surrounding the original repeat sequences. Such rearrangements may be detected in descendants of the deleted genome as the loss of highly conserved sequence motifs, such as those of common operons.

In the reduced genomes of *Rickettsia*, gene duplications common to other bacteria such as multiple rRNA operons and duplicated elongation factor Tu genes are missing (21, 171). These, along with short repeat sequences that are common in free-living bacteria, seem to have been consumed by intrachromosomal recombination in the genome of *Rickettsia* (20). In addition, the correlate is observed. Thus, the highly conserved operons for rRNA, proteins of the translation apparatus, and some metabolic enzymes are either gone from the *Rickettsia* genome or are retained in scrambled form (7, 19, 21, 171). Such depredations are even more in evidence in the genomes of mitochondria. These commonly have their minimal coding sequences arranged with little rhyme or reason, except among some primitive protists and plants (75, 77).

As mentioned earlier, the largest number of coding sequences observed so far in mitochondria belongs to *Reclinomonas americana* (77, 80, 111). Although it has only 67 protein-coding genes, it is a giant among mitochondria. There is much to recommend interest in this genome, which is in some ways very unlike other mitochondrial genomes (77). Like some other protists and plant mitochondrial genomes, it has recognizable gene motifs, such as the rRNA operon and the giant ribosomal protein cluster seen in bacteria (20, 80, 111). The presence of such motifs is hard to explain other than by the conservation of ancient bacterial motifs.

That *R. americana*'s mitochondrial genome contains 18 protein-coding sequences not seen in other mitochondrial genomes is not as remarkable as the fact that all of the proteins found in all other mitochondrial genomes are among the remaining 49 protein-coding genes (77, 80, 111). This simple fact speaks forcefully for the monophyletic character of the mitochondrial lineages, particularly when it is recalled that there are hundreds of mitochondrial proteins coded by genes in the nuclei of eukaryotes (77).

Phylogenetic reconstructions with the coding sequences from the *R. americana* mitochondrion along with those from other mitochondria and from bacterial genomes are unambiguous: the α-proteobacteria *Rickettsia* and *Bartonella* have a common ancestor with the mitochondrial lineage. That common ancestor was a free-living bacterium with a genome that was probably larger than that of *Bartonella* and certainly much larger than those of *Rickettsia* as well as mitochondria. Nevertheless, there is a very important difference between the genomes of *Rickettsia* and those of mitochondria. While the proteome of *Rickettsia* is at most twice the size of the mitochondrial proteome, typically less than 10% of the mitochondrial proteome is encoded by the mitochondrial genome. Thus, there is a dimension to the reductive genome evolution of mitochondria that is not shared by the *Rickettsia*.

Transfer to the Nucleus

There is a decisive difference in the evolutionary coupling between the host cell genome and the genome of an endocellular parasite and between the genomes of an organelle and of an endosymbiont (15). The fitness of the endosymbiont, like that of the organelle, is coupled positively to that of the cell. In contrast, the parasite's fitness is negatively coupled to the cell's fitness. Consequently, mutations that adversely affect the parasite will benefit the host genome. Conversely, deleterious mutations in the mitochondrion can be compensated for by changes in the host genome that enhance the combined fitness of the two genomes. Much of the evolution of the mitochondrion can be understood with the help of this distinction.

The asexual character of mitochondrial lineages suggests that they might be particularly vulnerable to Muller's ratchet, especially compared to nuclear genomes, with their well-developed sexual mechanisms (107, 117, 119). Thus, Muller's ratchet (132) might account for the fact that the vast majority of genes constituting the mitochondrial proteome are found in the nuclear genome. Simply stated, it is conceivable that genes

in the mitochondria would have a much heavier mutational load than the same genes in the nucleus. If a transfer mechanism existed, it would be advantageous for the cell to move genes from the organelle to the nucleus. Indeed, the data for metazoan mitochondria tend to support this notion, because in these organisms, the mutation rates are much higher in their mitochondria than in their nuclei (204). Furthermore, it can be shown that as long as a transfer mechanism exists to shuttle genes from mitochondria to nucleus, the mutational load will inevitably drive genes to the nuclear genome (27). However, there is a serious limitation to this model: it is only applicable to a small fraction of this planet's eukaryotes, primarily the metazoans.

Thus, the genomes of plant mitochondria tend to be less mutation prone than plant nuclear genomes, and in fungi the mutation frequencies of the two genomes are more or less equivalent (117, 118, 146, 202). Nevertheless, the coding capacities of plant and yeast mitochondria are in general not very different from those of animal cells; in all cases, the overwhelming majority of mitochondrial genes are found in the nucleus (75, 77). From these observations, it follows that mutational load alone cannot drive the migration of organelle genes to the nuclei of organisms other than animals.

For the most realistic situation of eukaryotic organisms in finite populations, what is required in addition to random mutations is a biased transfer mechanism. When the transfer mechanism is adequately biased in the direction of the nucleus, it can overcome a mutational gradient in the opposite direction (27). For example, if a cellular transfer mechanism favors moving genes from mitochondria to the nucleus over transferring genes from the nucleus to mitochondria, it can do so as long as its bias is greater than the mutational bias of the nucleus compared to the mitochondrion.

There are data indicating that transfers between mitochondrial and nuclear genomes are an ongoing evolutionary process (1, 31, 45, 126, 140, 144–146, 176, 177). Furthermore, there is an experimental system to study and quantify the transfer of sequences between mitochondrial and nuclear genomes in the yeast *Saccharomyces cerevisiae* (66, 175, 176, 178). Here, plasmids have been introduced into mitochondria, and the transfer of coding sequences from these plasmids to nuclei has been studied quantitatively (38, 82, 178, 190). The genetic data suggest that the primary pathway for the uptake of mitochondrial coding sequences that are transferred and expressed in the nucleus is provided by autophagy of mitochondria by cellular vacuoles (phagolysosomes). Nucleic acid fragments liberated by disintegration of mitochondria may become intermediates in the transfer to nuclei.

Thorsness and Fox (175, 176) have estimated rates of transfer of coding sequences to and from the mitochondria. For wild-type *S. cerevisiae*, there is roughly 1 transfer event to the nucleus from the mitochondria per 10^5 generations. Their experiment failed to detect transfer in the opposite direction, suggesting that this rate is less than 1 transfer/10^{10} generations. Thus, in this experimental model, the transfer process is expressed at least 10^5 times more frequently from mitochondria to the nucleus than in the reverse direction. This means that we have found a highly polar process to transport genes from the mitochondria to the nuclei in *S. cerevisiae*. How general is this likely to be?

One indication of the generality of the transfer process is that autophagic vacuoles are ubiquitous in eukaryotes. In addition, it seems that exogenous fragments of nucleic acids do not normally get into mitochondria. Thus, the experimental transfer to mitochondria of exogenous coding sequences such as those on plasmids has been accomplished to our knowledge only in *S. cerevisiae*. This transfer requires what is referred to as high-velocity microprojectile bombardment (66). In effect, this unique experimental system requires that the transferred sequences be shot into the cells to effect penetration of some mitochondria. It would seem that effective physical barriers normally prevent transfer of coding sequences into mitochondrial genomes.

Another argument favoring the asymmetric transfer to nuclei by vacuoles is that these universal organelles consume the mitochondria as well as peroxisomes in the normal course of their function. It is this degradation process that apparently releases the fragments of nucleic acids that are taken up by the nuclei (38, 82, 178, 190). Obviously, the destruction of nuclei by vacuoles would be lethal to a cell, which would prevent vacuole-dependent gene transfer from nuclei to mitochondria by this route. For these reasons, we are inclined to believe that preferential transfer of coding sequences to nuclei is the rule rather than the exception. This does not preclude the transfer of sequences from the nucleus to mitochondria by other transfer processes (123, 177).

In order to be recruited by the mitochondrion, a protein-coding sequence transferred to the nucleus often requires an addressing signal to direct its product back to the organelle (24, 138, 162, 163). Splicing pathways such as those that support exon shuffling might accelerate the tagging of newly transferred genes with appropriate addressing sequences. Indeed, the discovery of an intron between an addressing sequence and a mitochondrial gene that had been transferred to the nucleus in some plants confirms this expectation (45, 140). Accordingly, it is conceivable that one reason that splicing systems spread through primitive eukaryotic nuclear genomes was to satisfy the need to tag newly transferred genes from mitochondria and chloroplasts with addressing signals.

The process that transfers a gene to the nucleus can be envisioned as a neutral process with several identifiable states (27). First, an inactive gene is transferred to the nucleus, while the active version is retained by the mitochondrion. Then, the tagging process provides an addressing sequence to the nuclear version so that both mitochondrion and nucleus have active versions of the gene. Finally, mutation inactivates and eventually deletes the mitochondrial gene, and the nuclear allele takes over its function.

Recent events in the evolution of legumes have produced intermediate stages of this sort for the transfer of *cox2* from mitochondria to the nucleus (1, 45, 140). The *cox2* sequence in the nucleus of legumes seems not to have been copied from mitochondrial DNA. Since this nuclear sequence does not require editing for expression, we may infer that it was copied from an edited RNA fragment (45, 140). Therefore, this transferred gene was probably copied from an RNA fragment that was released by the destruction of mitochondria.

The neutral transfer model implies that eventually all of the genes of the mitochondria that can be transferred without ill effect will be transferred to the nucleus (27). Why then are there any genes left in contemporary mitochondrial genomes? Obviously, some genes remaining in mitochondrial genomes may be ones for which the transfer process is not neutral or for which the transfer requires very rare mutational events. If the entire group of mitochondrial genes is destined for transfer to the nucleus, sequence technology may have discovered the last ones poised for transfer. We recall here that only two such proteins are left in the mitochondrial genome with the most limited coding capacity. It has been suggested that these two genes must remain fixed in the organelle's genome because their protein products can regulate their own expression in tune with the redox potential of the mitochondria (5). Consis-

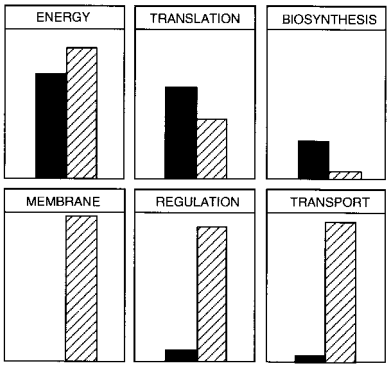

FIG. 1. Schematic illustration of the relative fraction of mitochondrial proteins with bacterial homologues (black bars) and without bacterial homologues (hatched bars). (Data taken from Karlberg et al. [95].)

tent with this interpretation is the identity of the two proteins encoded by all known mitochondrial genomes; both are cytochrome subunits (77).

In summary, we can identify evolutionary forces that tend to reduce the coding capacity of mitochondrial genomes. These evolutionary forces are either destructive, leading to the loss of coding sequences from the cell, or conservative, leading to the transfer of coding sequence to the nucleus. In the next section we also document the existence of an expansive evolutionary tendency that has supplemented the ancestral endosymbiont's proteome and transformed it into that of a cellular organelle.

THE YEAST MITOCHONDRIAL PROTEOME

The discovery of a major fraction of mitochondrial proteins that are not the descendants of a bacterial ancestor was entirely unexpected (95). This discovery required access to a definitive genome sequence and data defining the mitochondrial proteome of the same organism. For the moment, the yeast *S. cerevisiae* is the only eukaryote for which these two requirements have been met. In this section, we summarize and illustrate the phylogenetic reconstructions obtained for the mitochondrial proteome of *S. cerevisiae*.

The relative fraction of genes with homologues in bacteria and in eukaryotes for each of six functional categories is summarized in Fig. 1. Roughly one third of all the proteins are classified as ambiguous. These may cluster with bacterial taxa at the ends of long branches in the phylogenetic trees, or they may have homology to proteins from both bacterial and eukaryotic taxa (95). No attempt has been made to deduce their origins. In contrast, a cohort of circa 50 mitochondrial proteins that are clearly most closely related to α-proteobacteria have been identified. These support the identification of α-proteobacteria as ancestors of the mitochondria (95).

Surprisingly, half of the mitochondrial proteome of *S. cerevisiae*, circa 200 proteins, have no discernable alignments with any bacterial homologues ($P < e^{-10}$). They cluster exclusively as eukaryotic homologues (95). The presence of a sizable eukaryotic cohort contradicts an expectation of the endosymbiotic theory, which implies that the mitochondrial proteome is exclusively the descendant of an ancestral bacterial proteome. The data suggest that the endosymbiotic theory requires modification.

It is evident from Fig. 1 that the phylogenetic clustering of the mitochondrial proteins into α-proteobacterial and eukaryotic homologues goes hand in hand with the functional profiles of the clusters. Thus, the bacterial homologues seem to be mainly involved in translation and energy metabolism. In contrast, the eukaryotic proteins are typically associated with transport and regulatory functions.

A way of viewing the phylogenetic differences between the different functional categories of proteins is as follows (17, 95).

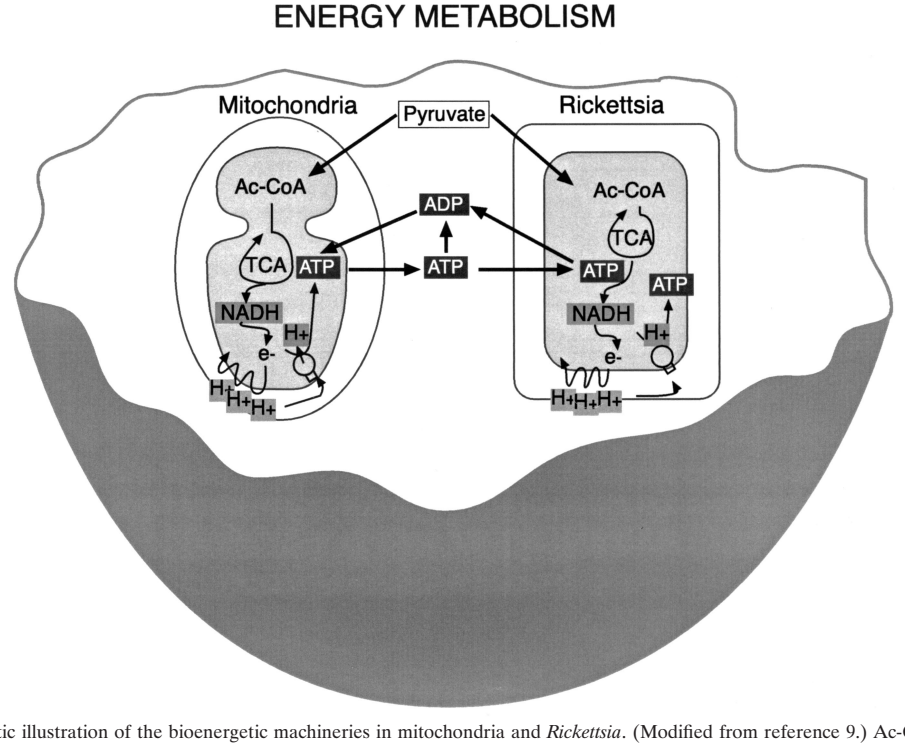

FIG. 2. Schematic illustration of the bioenergetic machineries in mitochondria and *Rickettsia*. (Modified from reference 9.) Ac-CoA, acetyl-CoA.

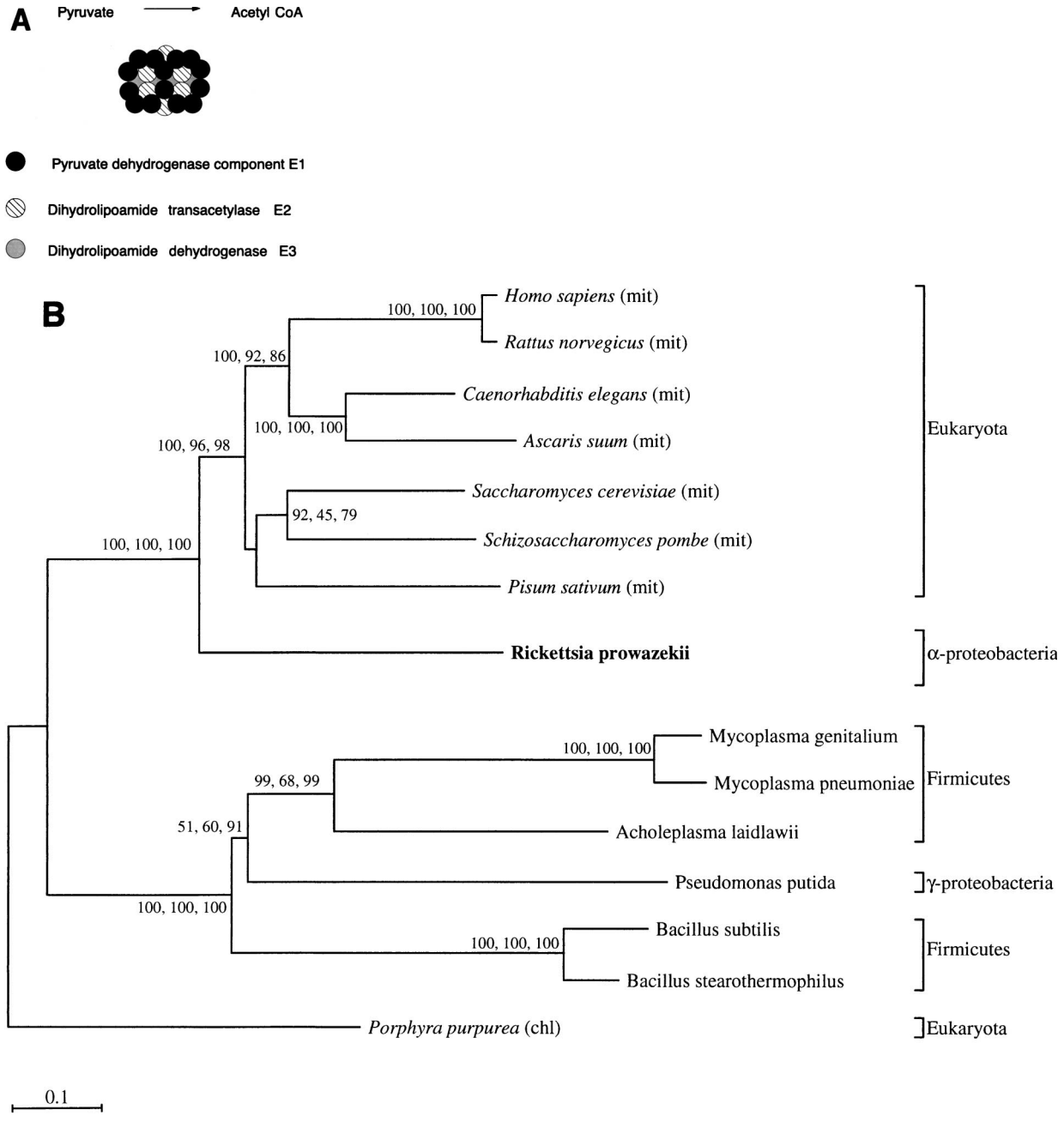

FIG. 3. (A) Schematic illustration of the pyruvate dehydrogenase complex. (B, C, and D) Phylogenetic reconstructions are based on the combined protein sequences of the α and β subunits of the pyruvate dehydrogenase E1 component (B), the dihydrolipoamide acetyltransferase E2 component (C), and the dihydrolipoamide dehydrogenase E3 component (D) from representative species. Names of species from the α-proteobacteria are shown in boldface and those from mitochondria (mit) are in italics in this and all subsequent figures where bootstrap numbers are indicated above the horizontal lines. The phylogenetic trees were constructed as described by Karlberg et al. (95).

First, we imagine that the ancestral bacterial symbiont introduced genes encoding proteins that have been perpetuated as an essential core of proteins functioning in aerobic respiration, the tricarboxylic acid (TCA) cycle, and gene expression. Furthermore, we suggest that the evolution of the mitochondria from the endosymbiont required that novel accessory proteins that arose in the eukaryotic genome complement such core functions. Some of these eukaryotic homologues augment the functions of the core proteins by participating in the assembly of complexes, while others function in regulation. In addition, there is a larger group of eukaryotic proteins with novel gene functions, such as ATP and protein transport. We suggest that the coevolution of the core α-proteobacterial components and the complementary eukaryotic nuclear components transformed the endosymbiont into an organelle.

We have also observed a number of bacterial protein homologues that may be examples of horizontal gene transfers from diverse bacteria. This means that there are proteins descended from three sorts of genomic ancestors in the yeast mitochondrial proteome: (i) homologues descended from an α-proteobacterium ancestor that encode core function, (ii) some ill-defined orthologues, some of which may have been recruited from a diverse group of bacteria through horizontal gene transfer, and (iii) the dominant group of eukaryotic pro-

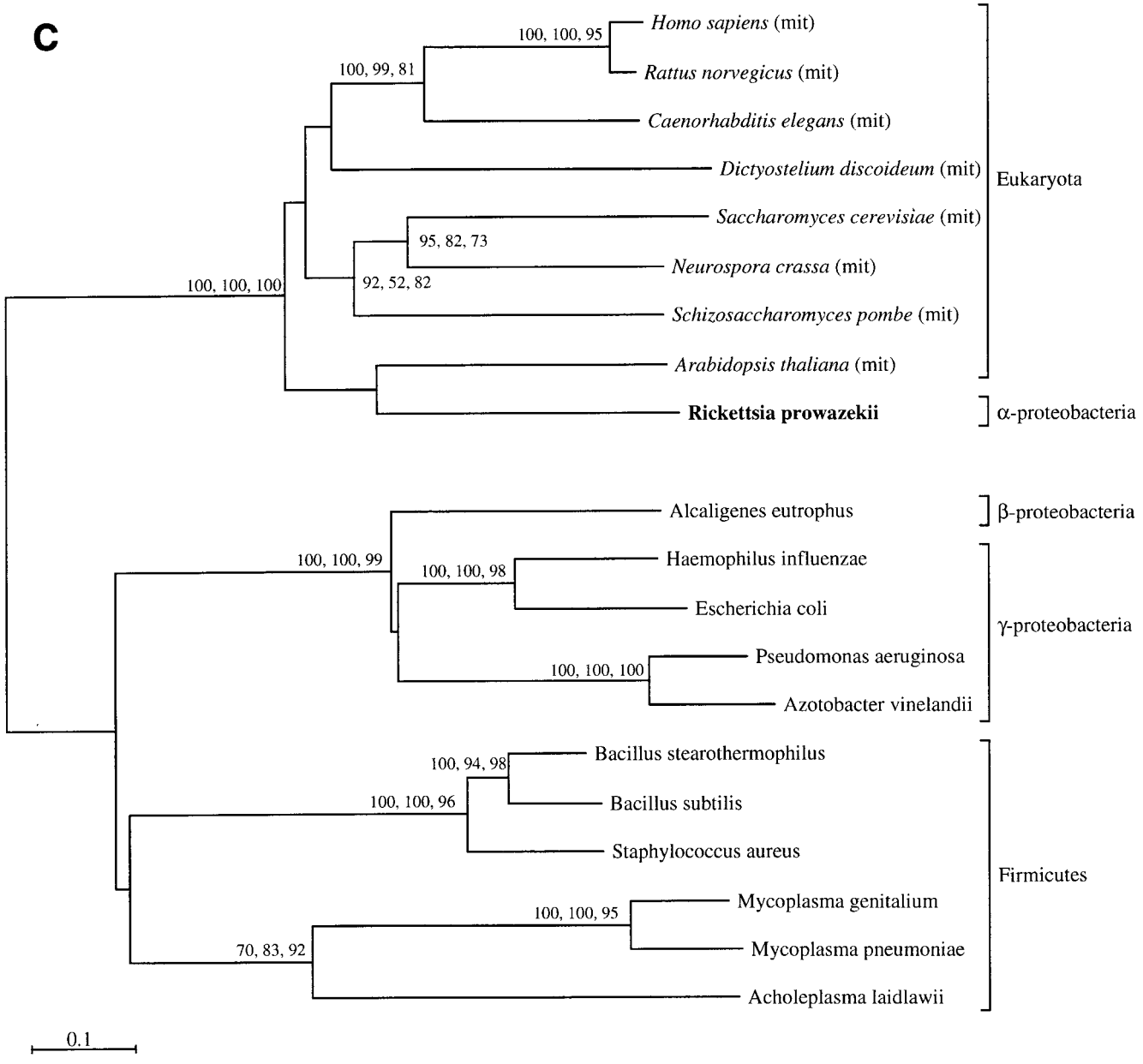

FIG. 3—*Continued.*

teins that have been recruited from the nuclear genome. In effect, the ancestral α-proteobacterial proteome consisting of 1,600 or more proteins has been reduced to roughly 50 proteins. This residual core has been complemented by circa 350 novel proteins, mostly recruited from the eukaryotic nuclear genome.

This interpretation rests on detailed phylogenetic analyses of the mitochondrial proteins of *S. cerevisiae* as well as of other eukaryotes that we present next. Here, all references to genes and their encoded proteins are for *S. cerevisiae* unless otherwise specified. Some relevant informational details may assist the reader in studying the phylogenetic reconstructions. All of the phylogenetic reconstructions are to be found at http://web1.ebc.uu.se/molev/publications/cfg2000 (95). The sequences of proteins of the yeast mitochondrial proteome are found in the Yeast Protein Database at http://www.proteome.com (87). The sequences of the mitochondrial genome of *Reclinomonas america* are at http://megasun.bch.umontreal.ca/ogmp/projects/other/mtcomp.html. Sequences for the nematode *Caenorhabditis elegans* are in the database WormPep17 at ftp://ftp.sanger.ac.uk/pub/databases/wormpep/. Other sequences from the SwissProt and NCBI databases were also used for the phylogenetic reconstructions. Alignments of homologous proteins ($P < e^{-10}$) were used to construct phylogenetic trees (95, 174). To structure the discussion, we have followed the somewhat arbitrary biochemical classification used by Karlberg et al. (95).

Energy Metabolism

ATP production in eukaryotes from glucose and oxygen normally consists of two catabolic processes: the conversion of glucose to pyruvate via a glycolytic pathway, and the oxidative

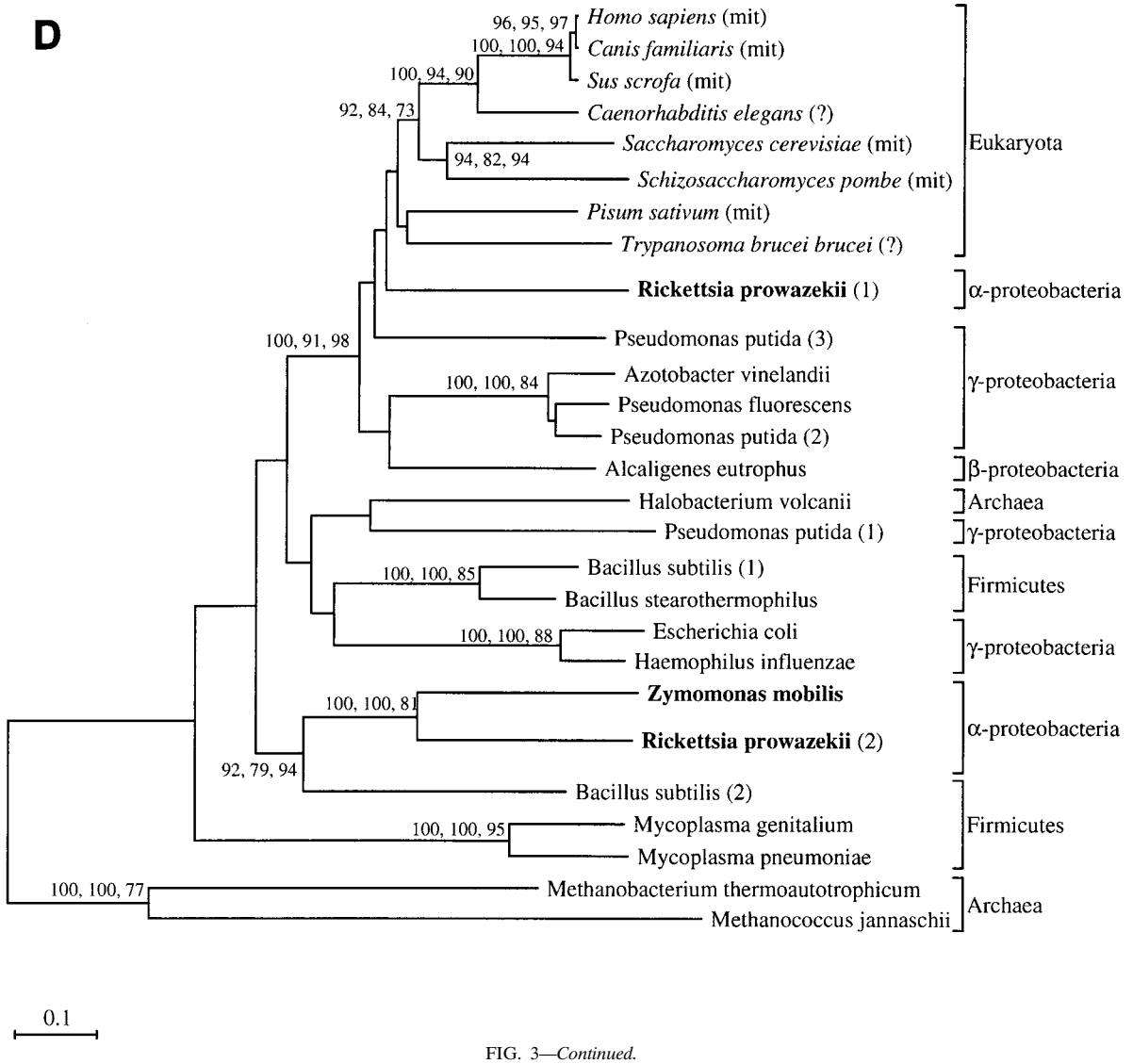

FIG. 3—*Continued.*

conversion of pyruvate to H_2O and CO_2. Both *Rickettsia* and mitochondria rely for pyruvate on the glycolytic systems of their host cells. More than 100 mitochondrial enzymes are involved in the oxidation of pyruvate. In *S. cerevisiae*, the relevant enzymes correspond to five proteins in the pyruvate dehydrogenase complex, 14 proteins in the ATP synthase complex, 16 proteins in the TCA cycle, and more than 70 proteins in the respiratory chain complexes. The ATP produced by this system is exported into the cytoplasm by the ATP/ADP translocases, which transport 1 molecule of ATP in exchange for 1 molecule of ADP.

The *Rickettsia* genome contains a similar cohort of proteins involved in the pyruvate dehydrogenase complex, the TCA cycle, the respiratory chain complex, and the ATP synthase complex. There are in addition five genes coding for ATP/ADP translocases, but these are required for the import of ATP from the host cell cytosol. Thus, *Rickettsia* and mitochondria have functionally related systems for ATP production (Fig. 2). However, ATP transport is somewhat different in these two systems (see below).

The pyruvate dehydrogenase complex that converts pyruvate to acetyl coenzymeA (acetyl-CoA) consists of multiple copies of each of three enzymatic components: pyruvate dehydrogenase (E1, two subunits), dihydrolipoamide acetyltransferase (E2), and lipoamide dehydrogenase (E3), as summarized in Fig. 3A. All three mitochondrial subunits are encoded by the nuclear genome and are more closely related to their homologs from α-proteobacteria than to those from other bacteria (Fig. 3B, C, and D). A notable exception is the *pdx*1 gene, which encodes a protein required anchoring the E3 to the E2 component in the complex. This protein has no bacterial homologues, so it may be a relatively recent addition to the pyruvate dehydrogenase complex.

Acetyl-CoA, produced by the pyruvate dehydrogenase complex, is fed into the TCA cycle (Fig. 4A). There are eight enzyme complexes in the TCA cycle, all of which are encoded by the nucleus in at least some eukaryotes. Three enzymes display particularly strong relationships to the α-proteobacteria, whereas the evolutionary history of the others is more complex. Genes encoding succinyl-CoA synthetase, succinate dehydrogenase, and fumarase seem to descend from an ancestral α-proteobacterium, and these have been transferred subsequently into the nucleus. For example, α-proteobacterial enzymes of the succinate dehydrogenase complex (Sdh) are

closely related to their mitochondrial homologs whether these are encoded in the nucleus or in the mitochondrion (Fig. 4B). However, the membrane-anchoring subunit of this complex, Sdh4, has so far been found only in eukaryotes.

Most if not all of the genes encoding enzymes in the TCA cycle have been duplicated. The resulting paralogues are often recruited to different subcellular compartments. For example, there are three genes encoding malate dehydrogenases in the nuclear genome of *S. cerevisiae*. These are targeted to the mitochondrion, the cytoplasm, and the peroxisome. All three enzymes form a phylogenetic cluster which is closely related to the mitochondrial malate dehydrogenases in other species (Fig. 4C). This suggests that the yeast cytosolic and peroxisomal forms are derived relatively recently from the mitochondrial malate dehydrogenase. On the other hand, the mouse cytosolic and mitochondrial malate dehydrogenases are highly divergent, which suggests that they arose in a more ancient gene duplication. The malate dehydrogenase in *R. prowazekii* seems not to be particularly closely related to either the mitochondrial or the cytoplasmic form of malate dehydrogenases in eukaryotes.

There are in *S. cerevisiae* at least two genes encoding the aconitase hydratase. The bacterial homologues cluster more closely to the cytoplasmic homologues and are more distant from the mitochondrial ones (Fig. 4D). This suggests that the homologue recruited for cytoplasmic functions and that recruited to the mitochondrion may have had different ancestors. The history of isocitrate dehydrogenases is even more complex. There are two forms of isocitrate dehydrogenase, the NAD^+- and the $NADP^+$-specific forms, both of which have the same function in the TCA cycle. The NAD^+-specific protein is a two-subunit protein in *S. cerevisiae* and a three-subunit protein in higher eukaryotes. The two yeast NAD^+ subunits cluster closely together, suggesting that they have been derived from the same ancestral gene. However, the $NADP^+$-specific isocitrate dehydrogenases are only remotely related to the NAD^+-specific proteins. The phylogenetic analysis suggests that the cytoplasmic, mitochondrial, and chloroplast $NADP^+$ homologues have common origin but have been recruited to different subcellular compartments in a pattern that is species specific. The $NADP^+$ homologues in mitochondria are most similar to those in *Sphingomonas yanoikuyae* (α-proteobacterium) and *Mycobacterium tuberculosis*, whereas more distantly related paralogues are found in all other bacteria. The isocitrate dehydrogenase found in *Rickettsia prowazekii* is highly divergent from all of these enzymes.

There are several other examples of mitochondrial and cytosolic isoforms that are highly divergent from bacterial homologues. For example, the mitochondrial citrate synthase seems to have originated from within the eukaryotic genome and displays no sequence identity with its bacterial analogue. The phylogenetic analysis of the enzymes of the TCA cycle reveals a complex evolutionary network. The complexities include (i) some genes that have been transferred from an α-proteobacterial ancestor into the nuclear genome and retargeted to the mitochondrion, (ii) others that have been transferred from an α-proteobacterial genome to the nucleus but targeted to other subcellular compartments, and (iii) still others that seem to have been recruited from different bacterial or eukaryotic ancestors.

The electron transport system consists of three energy-coupling sites: (i) the NADH dehydrogenase complex, (ii) the cytochrome *bc1* complex, and (iii) the cytochrome oxidase complex (Fig. 5A). Although some subunits of the NADH dehydrogenases from diverse eukaryotes are encoded by mitochondrial genomes, most are encoded by nuclear genomes. For example, the mitochondrial genome of *Reclinomonas americana* contains as many as 10 genes encoding subunits of the NADH dehydrogenase complex. This arrangement is reminiscent of the nine NADH dehydrogenase subunits located in immediate proximity to each other in *Escherichia coli*. A reduced version of this arrangement is also seen in *R. prowazekii*. Phylogenetic reconstructions based on a combined set of mitochondrial NADH dehydrogenase subunits suggest that these are derived from the α-proteobacteria (Fig. 5B). In contrast, only a single nuclear gene that encodes one of the subunits in the NADH dehydrogenase complex has been identified in *S. cerevisiae*. The corresponding protein is roughly 30% identical to the NADH dehydrogenase subunit of the γ-proteobacteria, *E. coli*, and *Haemophilus influenzae*. This low level of sequence identity suggests that the yeast homologue may not be a descendant of bacterial orthologues. Indeed, it seems likely that the bacterial NADH dehydrogenase genes may have been discarded from the yeast mitochondrial as well as nuclear genomes and replaced by eukaryotic analogues that are too dissimilar to be identified by current routines for identifying sequence similarity.

The second coupling site of the electron transport system is the cytochrome *bc1* complex, which contains a core of proteins such as cytochrome *b*, cytochrome *c1*, and the Rieske iron-sulfur protein. This complex has been isolated from many α-proteobacterial species, such as *Paracoccus denitrificans*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum*, and *Bradyrhizobium japonicum*. The gene encoding cytochrome *b* is located in the mitochondrial genome and is closely related to its α-proteobacterial relatives (Fig. 5C). The nuclear gene encoding cytochrome *c1* (*cyc1*) is also related to one found in γ-proteobacteria. Homologues for the Rieske iron-sulfur protein (*rip1*) have been found in α- and γ-proteobacteria as well as in cyanobacteria and green sulfur bacteria. All three mitochondrial proteins cluster closely with their α-proteobacterial relatives. In *S. cerevisiae*, the cytochrome *bc1* complex is composed of as many as 10 subunits. The seven other components of this complex in yeast cells are related to eukaryotic homologues, but they are unrelated to any bacterial analogues. These seven appear to be later additions to the central core derived from α-proteobacteria.

The cytochrome oxidase complex in *S. cerevisiae* provides another example of a complex where proteins of eukaryotic descent have been added to a core of α-proteobacterial proteins. The core genes are *cox*I, *cox*II, and *cox*III, encoding cytochrome oxidase *c* subunits I, II, and III, respectively. These three subunits are almost always mitochondrially encoded and cluster closely with their α-proteobacterial homologues (Fig. 5C). In contrast, none of the other yeast proteins in this complex can be aligned with bacterial homologues. These nonbacterial homologues are only occasionally found in other eukaryotes, which is consistent with the interpretation that they are recent nuclear contributions. Only 1 of the 13 yeast proteins responsible for the assembly of the cytochrome oxidase complex has a homologue in *R. prowazekii*. Four of these assembly proteins can be identified in the proteomes of other eukaryotes. The remaining assembly proteins for the cytochrome oxidase complex are specific for *S. cerevisiae*.

The gene order for the core components of the ATP synthetase complex (Fig. 6A) is highly conserved among bacterial genomes (Fig. 6B). There are 14 proteins in the ATP synthase complex in *S. cerevisiae*, half of which can be identified as bacterial descendants. Three of the latter are encoded in yeast mitochondria (*atp*6, *atp*8, and *atp*9), and each displays a close phylogenetic relationship to α-proteobacterial homologues. Similarly, phylogenetic reconstruction with the concatenated

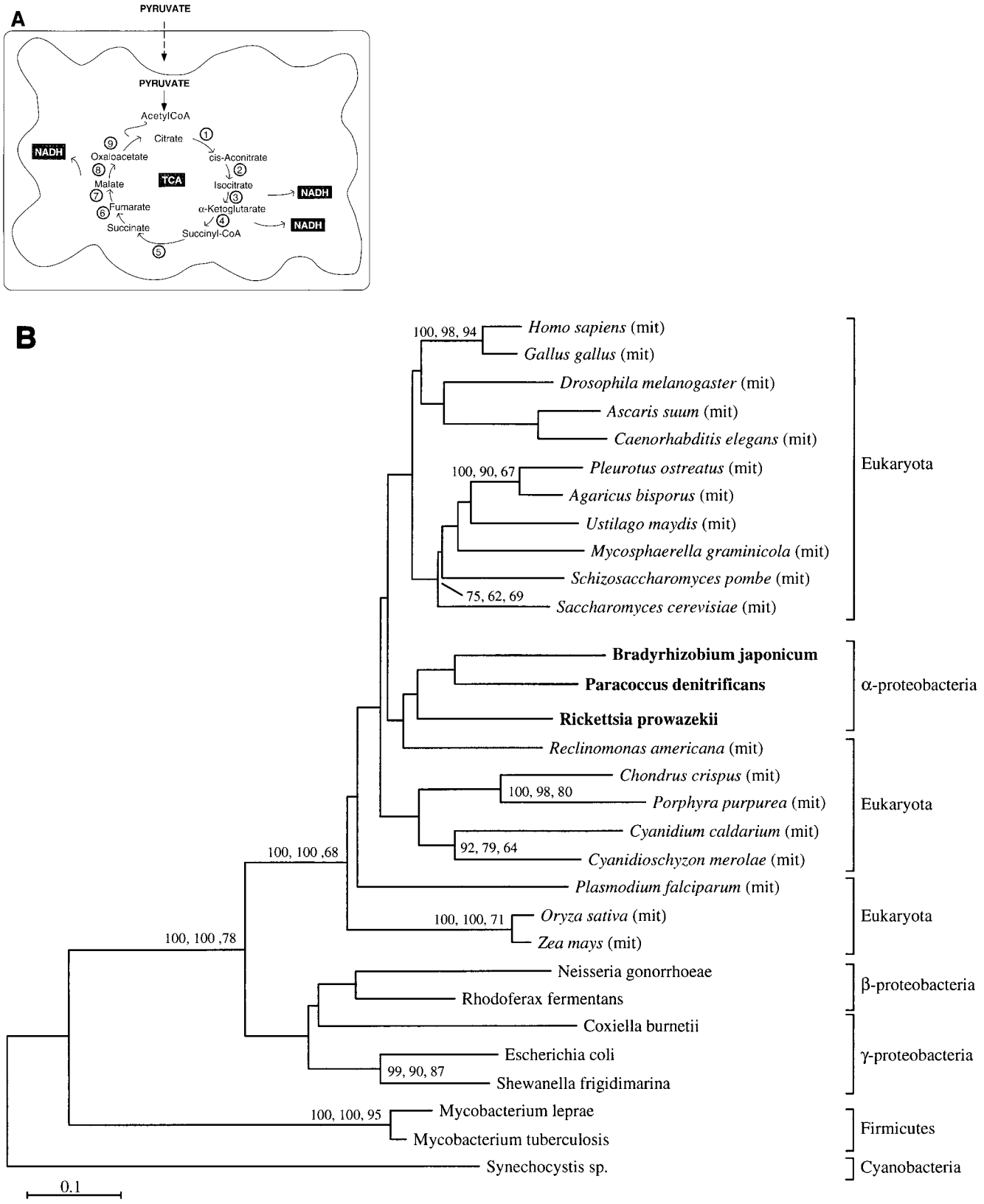

FIG. 4. (A) Schematic illustration of the TCA cycle. (B, C, and D) Phylogenetic reconstructions based on the succinate dehydrogenase iron sulfur protein (B), malate dehydrogenase (C), and aconitase (D) from representative species. The phylogenetic trees were constructed as described by Karlberg et al. (95). gly,glycosome; per, peroxisome; cyt, cytosol; chl1 and chl2, chloroplast 1 and 2, respectively; mit, mitochondrion.

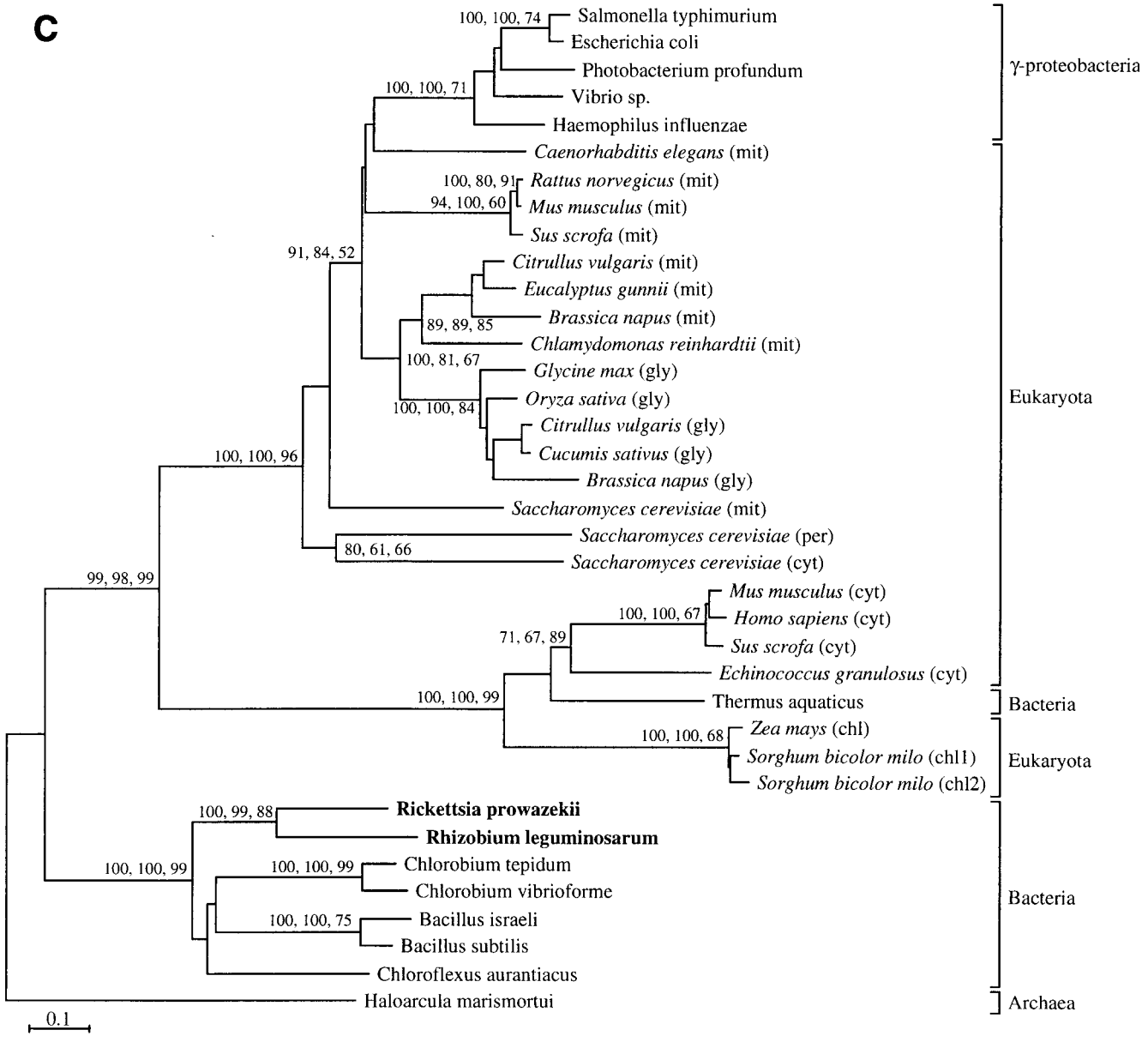

FIG. 4—*Continued.*

alignments of proteins encoded by the nuclear genes for the α and γ subunits of the ATP synthase (*atp*1 and *atp*3) reveals a cluster with strong bootstrap support for mitochondrial and α-proteobacterial homologues (Fig. 6C). A similar cluster is also observed for the genes *atp*2 and *atp*5. Of the remaining seven yeast proteins in this complex, four are found only in eukaryotes and three appear to be specific to *S. cerevisiae*.

In summary, phylogenetic reconstructions for the core components of the respiratory chain complexes identify these as α-proteobacterial descendants that are often encoded by the mitochondrial genomes of contemporary eukaryotes. Other components of the mitochondrial proteome descended from α-proteobacteria, such as *atp*1 and *atp*3, have been transferred to nuclei. Associated with these core components are accessory proteins such as those that assist in the assembly of complexes for which no homologues exist in bacteria. These seem to have evolved after the ancestral endosymbiotic event that contributed the core components. In total, 42% of the proteins participating in the aerobic ATP-generating system of yeast mitochondria have been found so far only in *S. cerevisiae* and other eukaryotes.

Information Processes

Processes such as replication, transcription, and translation are supported by more than one third of the mitochondrial proteome. Genes encoding components of these systems are typically located in the nuclear genome (Fig. 7), although the mitochondrial genomes of many protists encode different subsets of ribosomal proteins. In particular, almost of half of the mitochondrial genome of *Reclinomonas americana* consists of genes involved in sequence information processing.

rRNA sequences have served as the key reference sequences for phylogenetic reconstructions (197). Indeed, the very first

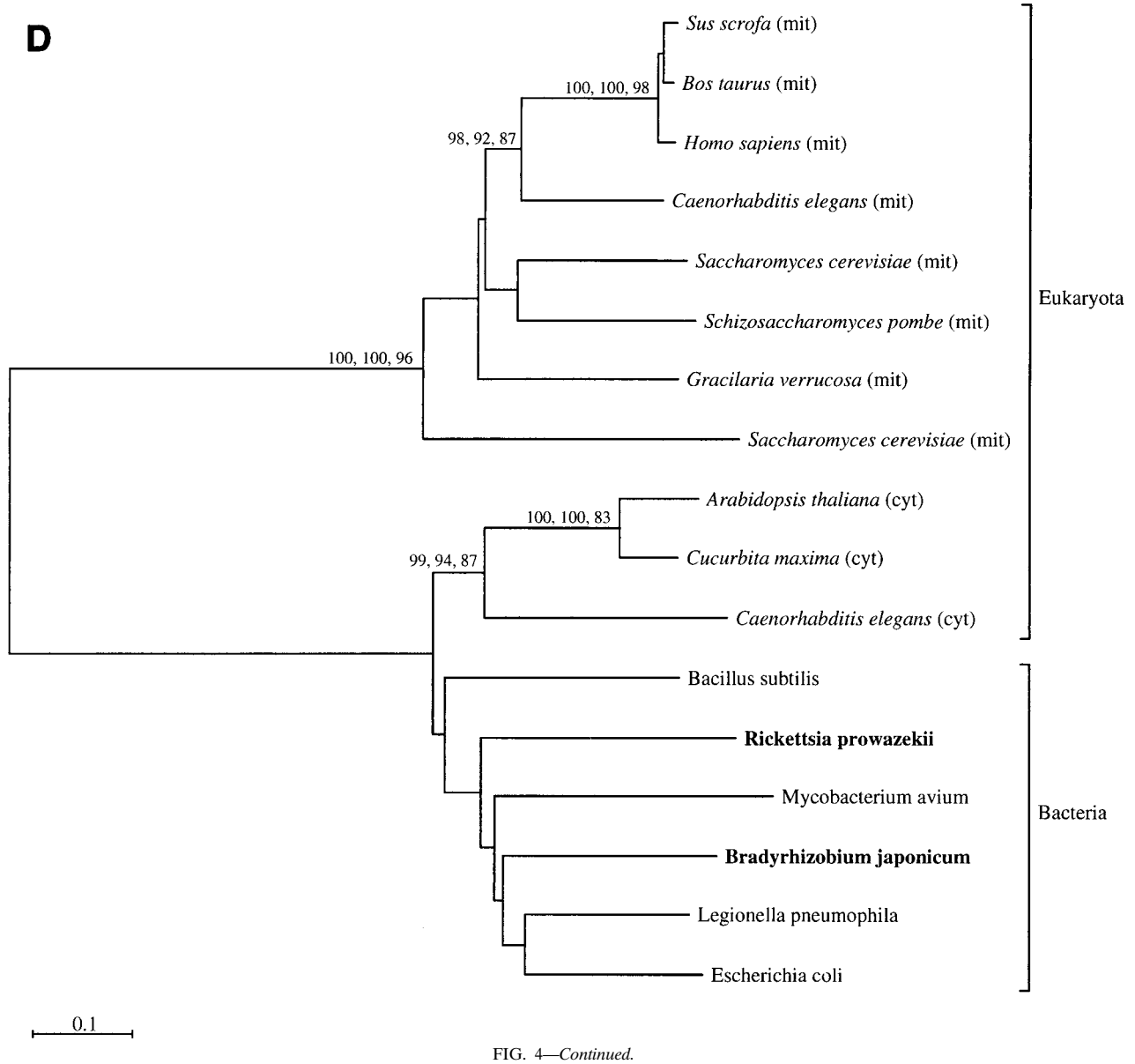

FIG. 4—*Continued.*

molecular data supporting the notion that mitochondria are derived from α-proteobacteria were obtained from phylogenetic reconstructions based on rRNA sequences (75, 197). More refined studies based on larger data sets suggested that there may be a particularly close relationship with the group of bacteria to which the *Rickettsia* belong, the *Rickettsiaceae* (Fig. 8) (75, 141, 197).

The interpretation of these phylogenetic reconstructions is not free of ambiguities. That many mitochondrial genomes are organized with extreme economy provides one complication. Thus, there is a direct correlation between the size of the mitochondrial genomes and their rRNA genes (13). In particular, in animal mitochondrial genomes smaller than 20 kb, the small- and large-subunit rRNAs are as short as ca. 850 and ca. 1,500 nucleotides, respectively. These lengths can be compared to those of the plant mitochondrial genomes longer than 300 kb that encode rRNAs with ca. 1,500 and 3,000 nucleotides. Fortunately, a conserved core of nucleotide sequences retained in all rRNA sequences facilitates alignments of rRNA sequences from genomes as diverse as those of bacteria and animal mitochondria despite their substantial size variation.

Another complication is that mutation biases and rates of nucleotide substitutions vary markedly among the mitochondrial genomes. The small animal mitochondrial genomes have a strong A+T mutation bias and evolve more than 50 times faster than the large plant mitochondrial genomes. This may account for the particularly close phylogenetic relationship observed between α-proteobacteria and plant mitochondria, with their long rRNA sequences and slow rates of nucleotide substitution. Although the rapid, biased sequence evolution of rRNA in animal mitochondria complicates phylogenetic reconstruction, it is likely that these mitochondrial rRNAs also descend from those of the α-proteobacterial subdivision (75).

Only a small fraction of the genes for the ubiquitous core of mitochondrial ribosomal proteins are found in the organelle's genome. The largest ribosomal protein assembly in a mito-

chondrial genome is found in *R. americana*, which encodes 11 large-subunit ribosomal proteins and 7 small-subunit ribosomal proteins. The organization of genes for these proteins in the protist's genome resembles the super-ribosomal protein gene operon found in bacteria (Fig. 9A) (111). Indeed, much detail of bacterial gene order with some departures is very well conserved in the mitochondrial genome of *R. americana* (77). A similar string of genes has been retained as a contiguous segment in the *Bacillus subtilis* genome. In the *E. coli* genome, this string has been divided into two segments, with the *rif* region located at 89 min and the ribosomal protein gene operons *str*-S10-spc and α located at 74 min (165). So much gene order seems to have been preserved in *R. americana* that it is tempting to infer that the entire string of genes associated with its ribosomal protein cluster are as they were in the genome of the ancestral endosymbiont.

Nevertheless, *R. americana* is exceptional. Most other mitochondrial genomes appear to have lost all traces of the gene order of the putative bacterial ancestor. In effect, the gene order of the *R. prowazekii* genome is intermediate between that of *R. americana* and the more representative genomes of mitochondria (14, 171). Thus, ancestral sequence motifs for genes encoding translation and transcription components are recognizable in *R. prowazekii*. However, these gene orders are slightly scrambled (Fig. 9A), presumably as a result of intrachromosomal recombination within a genome that originally was arranged as in some modern bacteria (19, 171).

Phylogenetic reconstructions based on the concatenated sequences of eight small ribosomal proteins and three large ribosomal proteins strongly support a clustering of the mitochondrial and α-proteobacterial sequences (Fig. 9B). A particularly interesting feature of this tree is that the nucleus-encoded yeast ribosomal proteins cluster with the mitochondrion-encoded ribosomal proteins. This supports the interpretation that ribosomal protein genes from the α-proteobacterial ancestor were transferred to the nucleus from a mitochondrial-endosymbiont intermediate.

The very small sizes of the rRNA sequences and the unusually large protein-RNA ratios in some mitochondria suggest that functions previously provided by the missing parts in these rRNA sequences may have been taken over by nucleus-encoded ribosomal proteins (T. O'Brien, personal communication). Indeed, the yeast mitochondrial ribosome contains a total of 23 small-subunit and 37 large-subunit proteins, many of which have no homologues in bacteria or other eukaryotes. Similarly, the human mitochondrial ribosome contains a large number of ribosomal proteins not found in bacteria. These examples of species-specific ribosomal proteins seem to represent novel solutions to the problems created by the tendency of mitochondria to delete short stretches of sequence from rRNA genes. Here, specific protein structures may replace the functions of deleted rRNA patches in the ribosomes of some mitochondria.

Most of the mitochondrial tRNAs are encoded by the organelle's genome. They are presumably the descendants of tRNAs from the ancestral endosymbiont (18, 98). As an apparent exception, several plant mitochondrial tRNA genes appear to have been replaced rather recently by chloroplast tRNA genes (75). It is conceivable that tRNA genes have crossed organelle boundaries several times in the evolutionary past.

The total number of mitochondrial tRNA genes varies in different species, with a minimum of only 22 tRNA genes in some animal mitochondria. It has been suggested that this reduction in the diversity of the tRNA population was associated with the codon reassignments characteristic of animal mitochondria (13, 37, 108). For example, AUA specifies methionine rather than isoleucine in animal and yeast mitochondria. AGA and AGG encode serine instead of arginine in the insect and echinoderm mitochondrial genes, while these same codons are termination codons in mammalian mitochondria. The standard $tRNA^{Arg}_{AGR}$ is absent in animal mitochondria. In insect and echinoderm systems, the translation of AGR codons is performed by the $tRNA^{Ser}_{AGY}$.

An absolute minimal tRNA set would theoretically consist of one tRNA per amino acid, i.e., 20 different tRNAs. However, since a minimum of two different tRNAs are required for the translation of arginine, isoleucine, serine, and leucine codons, a more realistic minimum of 24 different tRNAs might be required to translate the conventional genetic code. On the other hand, the diversity of the tRNA population could be further reduced if codon recognition patterns are altered so that a single tRNA can translate the amino acids encoded by three or six codons in the conventional genetic code (13, 108).

Thus, extreme pressure to minimize the set of tRNA species encoded in the mitochondrial genomes might lead to the following sorts of codon reassignments: AUA can be reassigned from isoleucine to methionine, UUG and UUA from leucine to phenylalanine, and AGA as well as AGG from arginine to serine or, alternatively, AGU and AGC from serine to arginine. Each of these reassignments could reduce the number of tRNA species required to translate all 64 triplets of the genetic code. Obviously, the reduction in the minimum number of tRNA isoacceptors requires that the remaining tRNA species expand their codon degeneracy. This may be done by reducing the contribution of the third codon position to function in what has been referred to as hyperwobble (108). A conversion of any of these codons into termination codons would serve the same purpose. Instances of each of these sense codon reassignments in animal mitochondria have been observed. Accordingly, it was suggested that the observed codon reassignments in animal mitochondria are a consequence of the minimization of the canonical set of tRNA genes inherited from the ancestral endosymbiont (13, 108).

Point mutations such as those affecting $tRNA^{Leu}_{UUR}$ are particularly common in human mitochondrial diseases. A spontaneous suppressor of such a mutant has been identified as a heteroplasmic alteration of the anticodon of $tRNA^{Leu}_{CUN}$ that enables the suppressor to translate leucine codons of the form UUR (129). In this case the suppressor mutation is carried by approximately 10% of the mitochondrial genomes (129). A similar scenario might account for the loss of tRNA species and the evolution of codon reassignments during the evolution of animal mitochondria. Here, a mutant with an ancestral tRNA gene that has accumulated point mutations and/or deletions might be rescued by other mutant tRNA variants that expand their codon recognition range to compensate for the first tRNA variant's defects.

In the yeast *Schizosaccharomyees pombe*, a viable mutant tRNA isoacceptor with an altered codon recognition pattern has been identified (160). The existence of such viable mutants contradicts the common assumption that mutations that alter codon recognition patterns are lethal. Such tRNA variants with or without accompanying suppressor tRNA species may be fixed in a lineage of mitochondrial genomes. Here, the transition from the canonical tRNA ensemble to one with novel codon recognition patterns might require minimum levels of heteroplasmy to support a transition from the standard translation pattern to an atypical one.

The aminoacyl-tRNA synthetases of the mitochondria are largely descendants of the α-proteobacterial ancestor that have been transferred to nuclear genomes. However, the transfer pattern is not without its complexities. In contrast to ribosomal

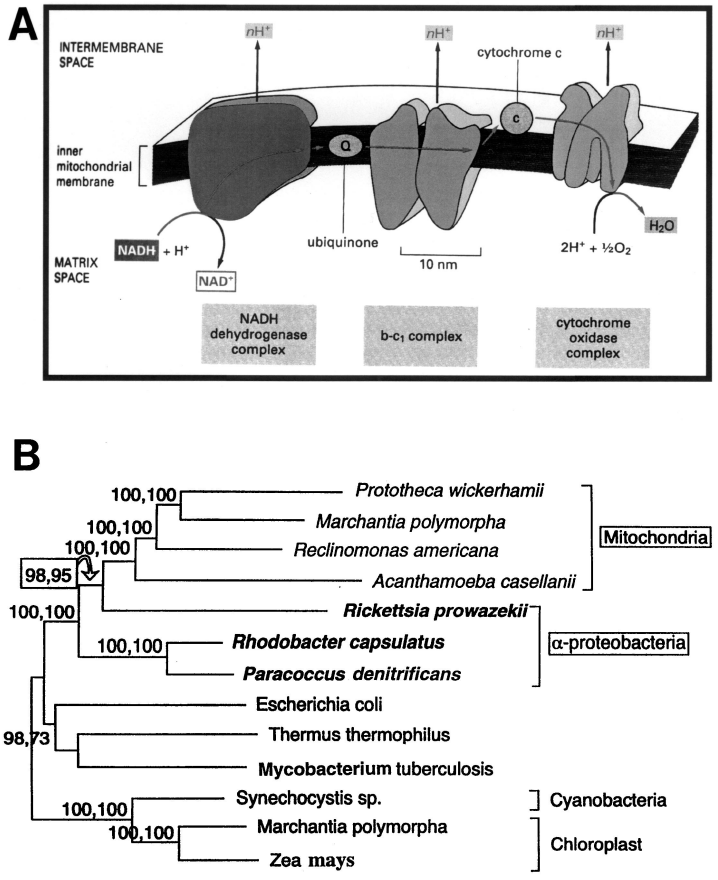

FIG. 5. (A) Schematic illustration of the respiratory chain complex. (B and C). Phylogenetic reconstructions based on the combined protein sequences of NADH dehydrogenase I chains A, J, K, L, M, and N (B) and the combined protein sequences of cytochrome *b* and cytochrome *c* oxidase subunit I (C) from representative species. The phylogenetic trees were constructed as described by Karlberg et al. (95). (B) The arrow indicates the bootstrap value at that node.

components, each of the aminoacyl-tRNA synthetases needs to interact with only a few molecules, such as the amino acids, ATP, and the tRNAs. This, along with the universality of their functions, may account for the fact that many aminoacyl-tRNA synthetases are associated with complex phylogenetic patterns that reflect the influence of horizontal gene transfers (199). The mitochondrial synthetases present opportunities for particularly complex phylogenetic patterns.

The initial association of the ancestral α-proteobacterium and its host brought together two complete translation systems with a total of 40 different aminoacyl-tRNA synthetases. The combination of the symbiont's synthetase complement in the presumptive organelle and the host's complement in the cytoplasm might have provided opportunities for extensive gene replacements and gene losses. In particular, after two billion years it is conceivable that the number of synthetases could have been reduced to a minimum of 20 that service both the mitochondrial and cytoplasmic compartments.

What is observed is more complex and only partially consistent with this conjecture. In *S. cerevisiae*, for example, three different aminoacyl-tRNA synthetases descended either from the symbiont or the host have been duplicated to serve the mitochondrion and the cytoplasm, while the complementary synthetase has been lost. Likewise, single genes that most likely function in the cytoplasm as well as in the mitochondrion encode a total of four synthetases. Surprisingly, a majority of the 20 nominal aminoacyl-tRNA synthetases in *S. cerevisiae* have retained the presumed ancestral pattern, represented by the presence of both a mitochondrial and a cytoplasmic protein of distinct phylogenetic origin. It is possible that the exceptional structures of mitochondrial tRNA species (see, for example, reference 108) and the requirement for these to be recognized by a coadapted synthetase have in some cases constrained the evolution of cell compartment-specific synthetase homologues.

We can imagine that the aminoacyl-tRNA synthetases have evolved in a three-stage process. (i) In gene transfer, the bacterial gene is transferred from the mitochondrion to the nuclear genome, which already contains an ancestral eukaryotic synthetase gene of the same specificity. Both genes are expressed, and the bacterially derived synthetase is targeted back to the mitochondrion, while the eukaryotic enzyme is targeted to the cytoplasm. (ii) In gene duplication and replacement, the nuclear and/or bacterial gene is duplicated, and one of the ancestral genes is replaced by the new paralogous gene copy. (iii) In functional duality and gene loss, signal sequences are added to the duplicated pair of genes so that their products can be recruited to the cytoplasm as well as to the mitochondrion. Finally, all of the unnecessary gene copies are purged by random mutation.

The first stage in this process seems to be represented by as many as 12 different aminoacyl-tRNA synthetases in *S. cerevisiae*, those for Glu, Phe, Leu, Met, Tyr, Asn, Asp, Trp, Ile, Lys, Ser, and Pro (Fig. 10A). Each synthetase is represented by a mitochondrial and a cytoplasmic form, each with a separate origin. The mitochondrial synthetases are normally similar to

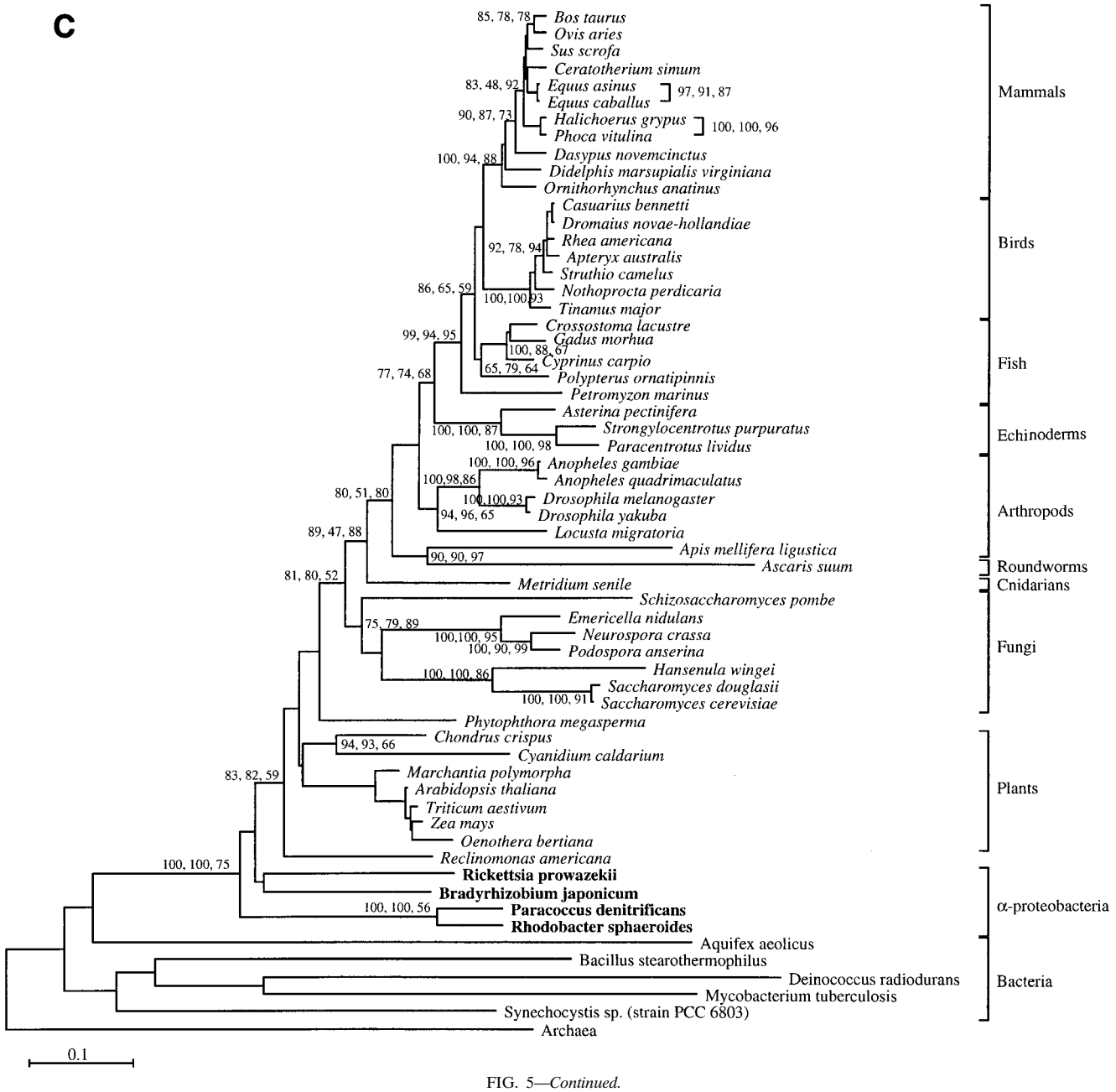

FIG. 5—*Continued.*

their bacterial homologues, but there is typically no specific clustering with the *rickettsiae*.

The disparity between rickettsial and mitochondrial synthetases has several sources. In the simplest case, the *Rickettsia* do not have an Asn tRNA synthetase. For other synthetases, the absence of this specific relationship is explained by the exceptional placement of *Rickettsia* in the phylogenetic tree of synthetases. Indeed, several examples of putative horizontal gene transfer events have been identified among the rickettsial aminoacyl-tRNA synthetases (20). For example, phylogenetic reconstructions suggest that the *Rickettsia* Met tRNA synthetase clusters more closely with a *Mycobacterium tuberculosis* homologue than with other α-proteobacterial versions of the enzyme. *Rickettsia* is also unusual in that it contains a class I Lys tRNA synthetase which is more similar to the class I Lys tRNA synthetase in the archaea than to the class II Lys tRNA synthetase of other proteobacteria. Finally, *Rickettsia* may have recruited the gene encoding the cytoplasmic Ile tRNA synthetase, since the rickettsial and the yeast cytoplasmic Ile tRNA synthetases are more similar to each other than either is to the yeast mitochondrial homologue.

The second stage in this process is characterized by the three yeast aminoacyl-tRNA synthetases specific for Arg, Thr, and Gly (Fig. 10B). There are for each of these synthetases two gene copies that cluster together in phylogenetic reconstructions with strong bootstrap support. For example, the cytoplasmic and the mitochondrial Thr tRNA synthetases form a cluster that appears to be of bacterial rather than of archaeal origin. In this instance, a mitochondrial Thr tRNA synthetase seems to have replaced the cytoplasmic one. Similarly, the

mitochondrial and cytoplasmic Arg tRNA synthetases cluster closely together with the bacterial Arg tRNA synthetase.

The third and final stage in this process is represented by the four yeast aminoacyl-tRNA synthetases for His, Val, Ala, and Cys (Fig. 10C). Each of these enzymes seems to be encoded by a single gene, suggesting that its gene product may serve both the mitochondrial and cytoplasmic compartments. It is sometimes difficult to infer the phylogenetic relationships for this class of synthetases because of putative horizontal gene transfer events. For example, the single yeast gene encoding ValRS is clearly of the bacterial type, whereas the Val tRNA synthetase of *Rickettsia* seems to be of the archaeal type.

Heat Shock Proteins

The best-sampled heat shock protein genes are the bacterial 10-kDa, 60-kDa, and 70-kDa heat shock proteins (HSP10, HSP60, and HSP70, respectively). Phylogenetic reconstructions based on HSP60, also known as GroEL, provide very strong support for a close phylogenetic relationship between mitochondria and α-proteobacteria (Fig. 11). A particularly close relationship between mitochondria and the group of bacteria to which *Rickettsia* and *Ehrlichia* belong is observed in this reconstruction (186). Homologues of HSP60 have also been identified in amitochondrial eukaryotes. We discuss these below.

HSP10, also known as GroES, is a short protein, which makes phylogenetic inferences from its sequences less reliable than those for larger proteins. We obtain phylogenetic reconstructions based on this protein that identify a mitochondrial as well as an α-proteobacterial cluster in which *Rickettsia* represents an early-diverging taxon. However, the bootstrap support values are not strong enough to support a specific relationship between mitochondria and α-proteobacteria for this protein.

HSP70 is encoded by the gene *dnaK* in bacteria. Phylogenetic reconstructions based on HSP70, like HSP10, suggest that *Rickettsia* represents an early-diverging clade within the α-proteobacteria. There are as many as three mitochondrial homologues of HSP70 in *S. cerevisiae*; each of these has roughly 50% sequence identity to the others. They all cluster with other mitochondrial proteins and seem to have arisen as ancient gene duplications. Finally, the yeast genome also encodes the mitochondrial heat shock protein HSP78, which is remotely related to bacterial heat shock proteins within the ClpB family of ATP-dependent proteases.

Biosynthesis, Regulation, and Transport

Other categories of genes encode proteins that complement both the respiratory system and the genetic sequence processing systems in mitochondria. Such proteins control the supply of substrates for protein and nucleic acid biosynthesis, the regulation of gene expression, and the transport of molecules in and out of the mitochondrion. With the exception of genes involved in biosynthetic processes, most of these accessory proteins seem to have originated within eukaryotic genomes.

More than 30 yeast genes are associated with biosynthetic pathways in mitochondria. A total of 16 genes encode proteins involved in amino acid metabolism, most of which have orthologues in all three domains. However, it has proven difficult to infer evolutionary relationships within this category of proteins due to the presence of large paralogous gene families. The multigene family of aminotransferases provides a good example of this complexity. The exceptional variability of substrate specificity among the aminotransferases may account for the phylogenetic complexity of this multigene family (26, 173).

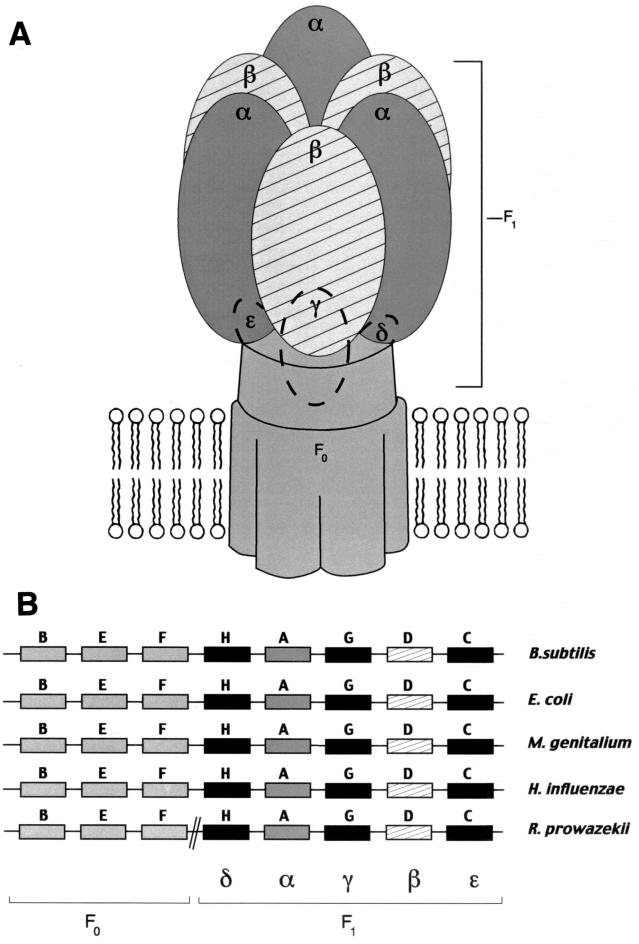

FIG. 6. (A) Schematic representation of the ATP synthase complex and (B) organization of the ATP synthase genes. (C) Phylogenetic reconstructions based on the combined protein sequences of the α and γ subunits of the ATP synthase complex from representative species. The phylogenetic trees were constructed as described by Karlberg et al. (95). Chl/Nuc refers to chloroplast/nucleus.

Since the *Rickettsia* lack most genes involved in amino acid biosynthesis, we are waiting for the annotation of the *Bartonella henselae* genome to be completed so that the α-proteobacterial genes in this family can be identified.

There is weak support for a clustering of mitochondria and *Rickettsia* homologues in phylogenetic reconstructions based on one gene involved in the isoleucine biosynthetic pathway (ILV1). Another two proteins from the same pathway (IL V5 and IL V6) support a clustering of mitochondria with proteobacteria. However, phylogenetic associations with other bacterial groups such as *Lactococcus* are also observed. Genes involved in biosynthetic functions may be particularly prone to changes in substrate specificity as well as horizontal gene transfer events (see below). Such tendencies might complicate the interpretation of phylogenetic reconstructions based on their sequences.

The overwhelming majority of proteins involved in the regulation of mitochondrial gene expression, mRNA stability, or splicing lack bacterial homologues. Accordingly, we suggest that they arose within eukaryotic nuclear genomes subsequent to the establishment of the endosymbiotic ancestor of mitochondria. A striking exception is provided by MSS1, which is involved in the expression of *cox*1 and *cox*3. Bacterial homo-

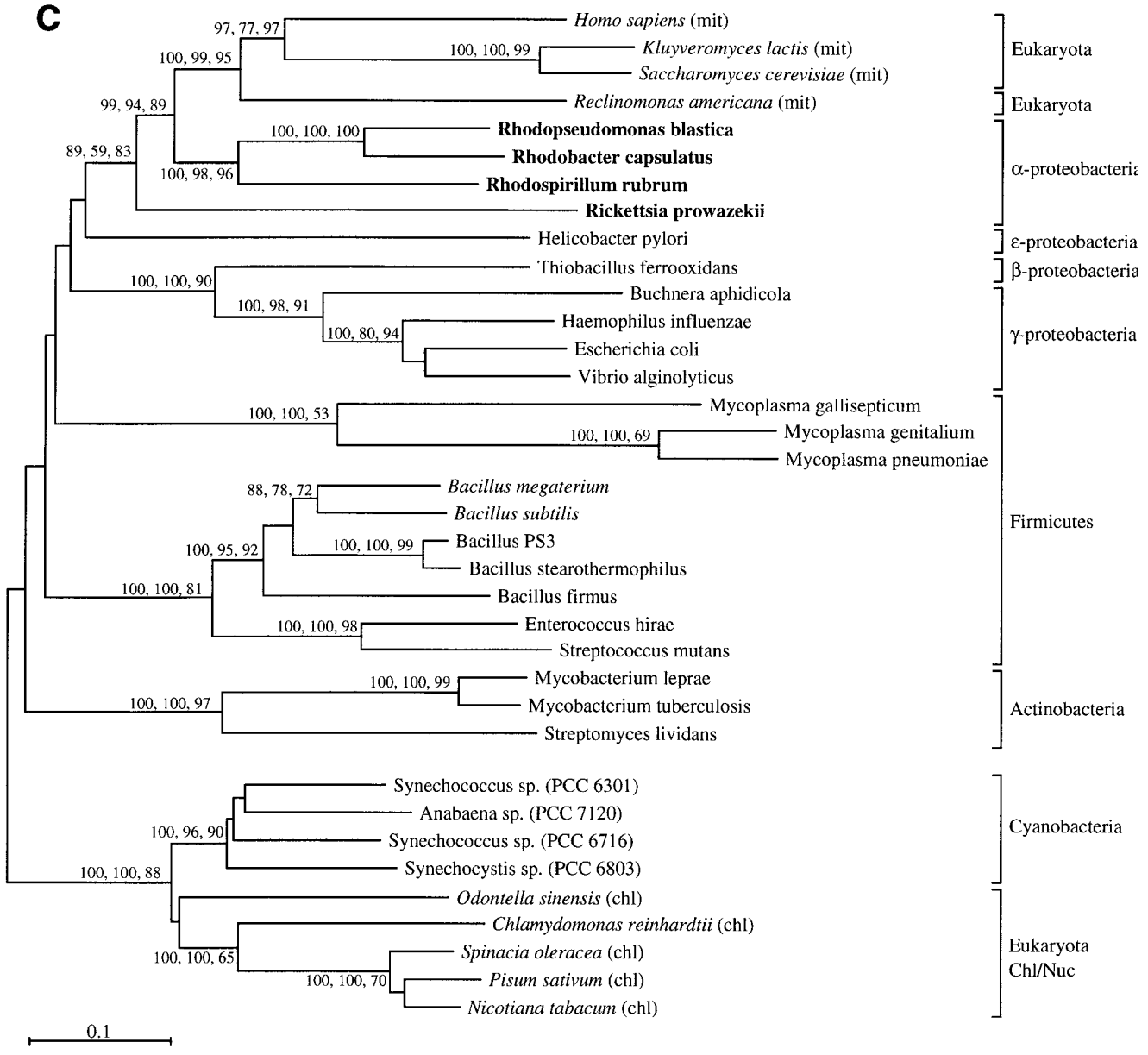

FIG. 6—*Continued.*

logues are available for this protein, and phylogenetic reconstructions of these indicate that homologues from *Rickettsia, S. cerevisiae*, and *C. elegans* form a tight cluster (Fig. 12A). Bacterial homologues are also available for MSS116, a mitochondrial RNA helicase required for splicing of group II introns of *cox*1 and *cob*. Unfortunately, the MSS116 protein is part of a large family of paralogous proteins, which complicates phylogenetic inferences. All other mitochondrial proteins involved in regulatory functions are unique to eukaryotic genomes.

The mitochondrial transport proteins also seem to have descended almost exclusively from eukaryotic genes. A provocative exception is the ATP-binding cassette (ABC) transporter protein ATM1, which is required for iron homeostasis in mitochondria. *R. prowazekii* has an orthologue of the ATM1 protein that clusters within the mitochondrial clade in phylogenetic reconstructions (Fig. 12B). These proteins are members of a large family of ABC transporters. Paralogous genes from this family have been found in a variety of organisms. A few other proteins active in sequestering iron, such as MMT1 and MMT2, have homologues in one or a few bacterial species.

More typical for *S. cerevisiae* are a relatively large number of inner and outer membrane proteins involved in protein import and referred to as TIM and TOM, respectively. These are exclusively eukaryotic proteins, with no similarities to bacterial genes. Homologues are known for some of these proteins in organisms such as humans, mice, and rats, while others appear to be unique to *S. cerevisiae*.

Another example of a protein family not found in prokaryotes is the ADP/ATP translocase gene family. This is a membrane-bound transport system that catalyzes the export of ATP in exchange for the import of ADP. The genes encoding the mitochondrial ADP/ATP translocators have been identified in the nuclear genomes of a large number of eukaryotes, such as fungi, plants, insects, and animals (Fig. 13A). Members

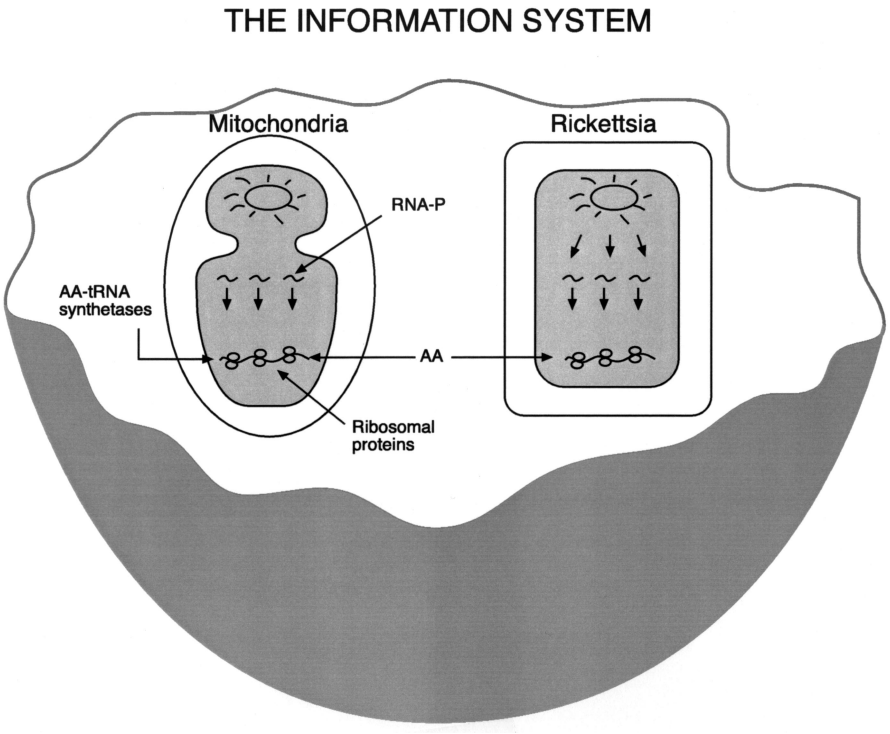

FIG. 7. Schematic illustration of the import of cytosolic components for the gene expression systems in mitochondria and *Rickettsia*. AA, amino acid; RNA-P, RNA polymerase.

of this superfamily have polypeptide chains that are approximately 300 amino acids long and that function as dimers. The yeast genome contains three genes that code for ADP/ATP translocators. Phylogenetic reconstructions based on the ADP/ATP translocators show that the mammalian and insect ADP/ATP translocases cluster separately from their fungal and plant counterparts (Fig. 13A). This pattern suggests that the mitochondrial ADP/ATP translocases are derived from a single ancestral gene that evolved vertically among the nuclear genomes. The phylogenetic reconstructions suggest that the mitochondrial ADP/ATP translocase genes originated early in the evolution of the eukaryotes, before the divergence of plants and animals (20, 84, 95, 103, 112, 188, 195, 196, 201).

Multiple gene duplications must be invoked to explain some parts of the tree describing the mitochondrial translocases. For example, the branching pattern suggests that the three human gene types originated prior to the divergence of mice, rats, and humans. In contrast, the gene duplication event that generated the two isoforms found in wheat and maize seem to have originated subsequent to the divergence of these species. Similarly, the three paralogues in *S. cerevisiae* cluster tightly, which suggests that they arose in recent gene duplications. As many as seven genes encoding ATP/ADP translocases have been identified in the *Caenorhabditis* genome. However, three of these are considerably more divergent than others; they are likely to be pseudogenes. The remaining four genes in *C. elegans* form a tight cluster.

In summary, the mitochondrial ADP/ATP transporters appear to belong to an ancient monophyletic lineage. The phylogenetic features of this gene family are consistent with the interpretation that mitochondria are monophyletic. These translocases belong to a superfamily of mitochondrial carrier

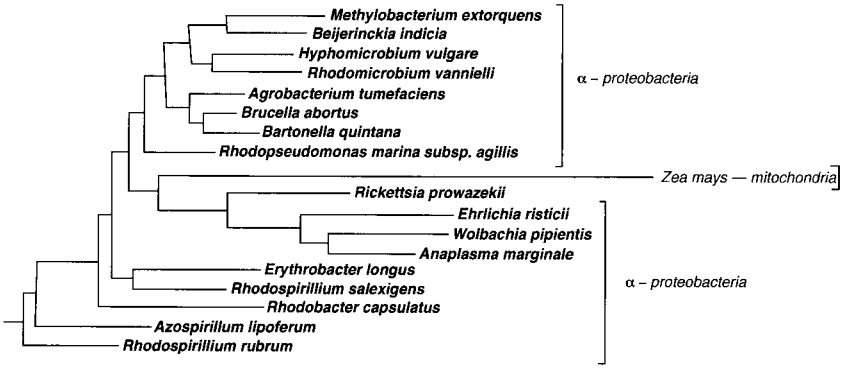

FIG. 8. Phylogenetic reconstructions based on rRNA sequences from *Zea mays* mitochondria and representative α-proteobacterial species. The tree is drawn according to Olsen et al. (141).

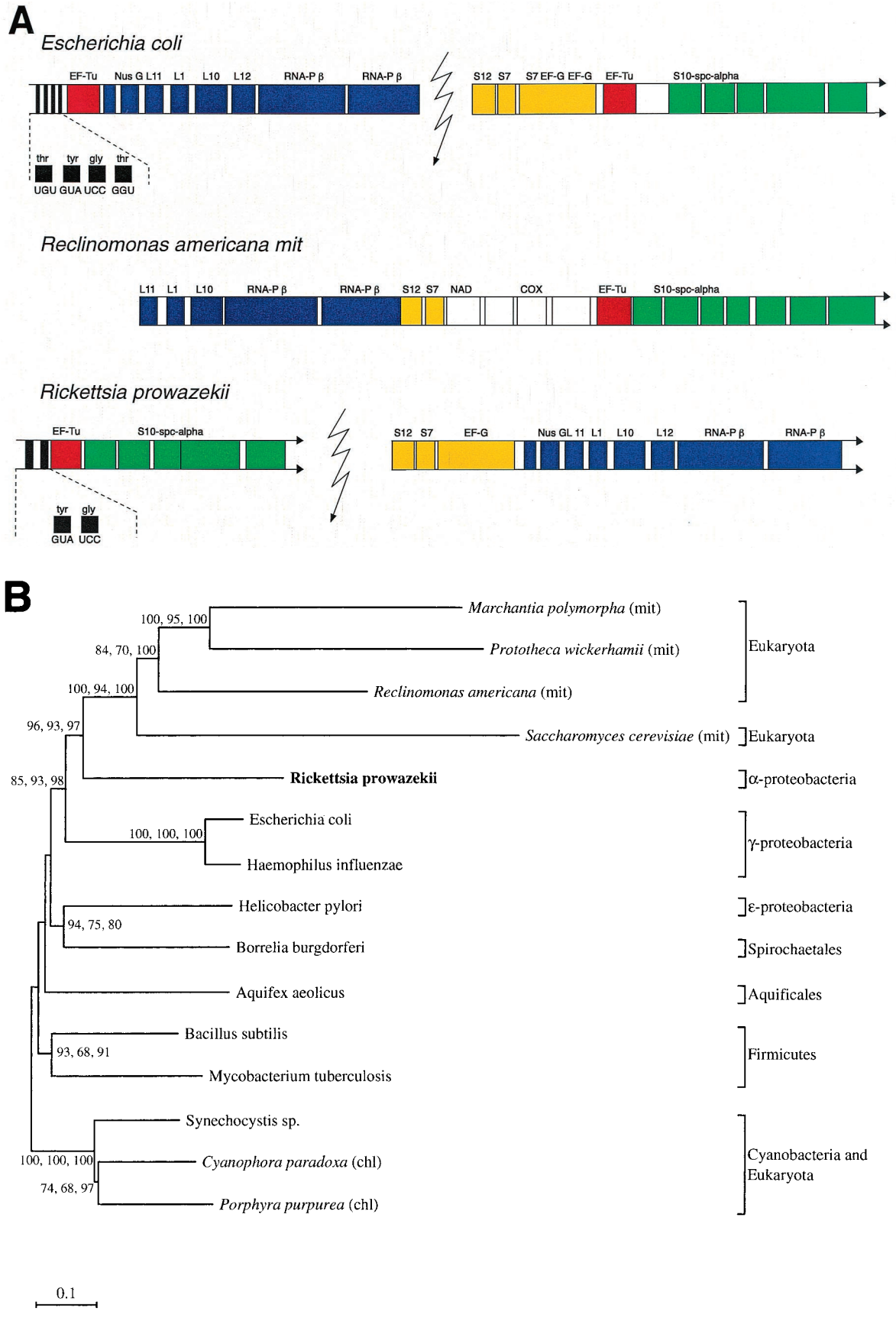

FIG. 9. (A) Schematic illustration of the organization of the ribosomal protein genes in *Escherichia coli, Rickettsia prowazekii*, and the mitochondrial (mit) genome of *Reclinomonas americana*. (B) Phylogenetic reconstructions based on the combined protein sequences of the ribosomal proteins S2, S7, S10, S12, S13, S14, S19, L5, L6, and L16 from representative species. The phylogenetic trees were constructed as described by Karlberg et al. (95).

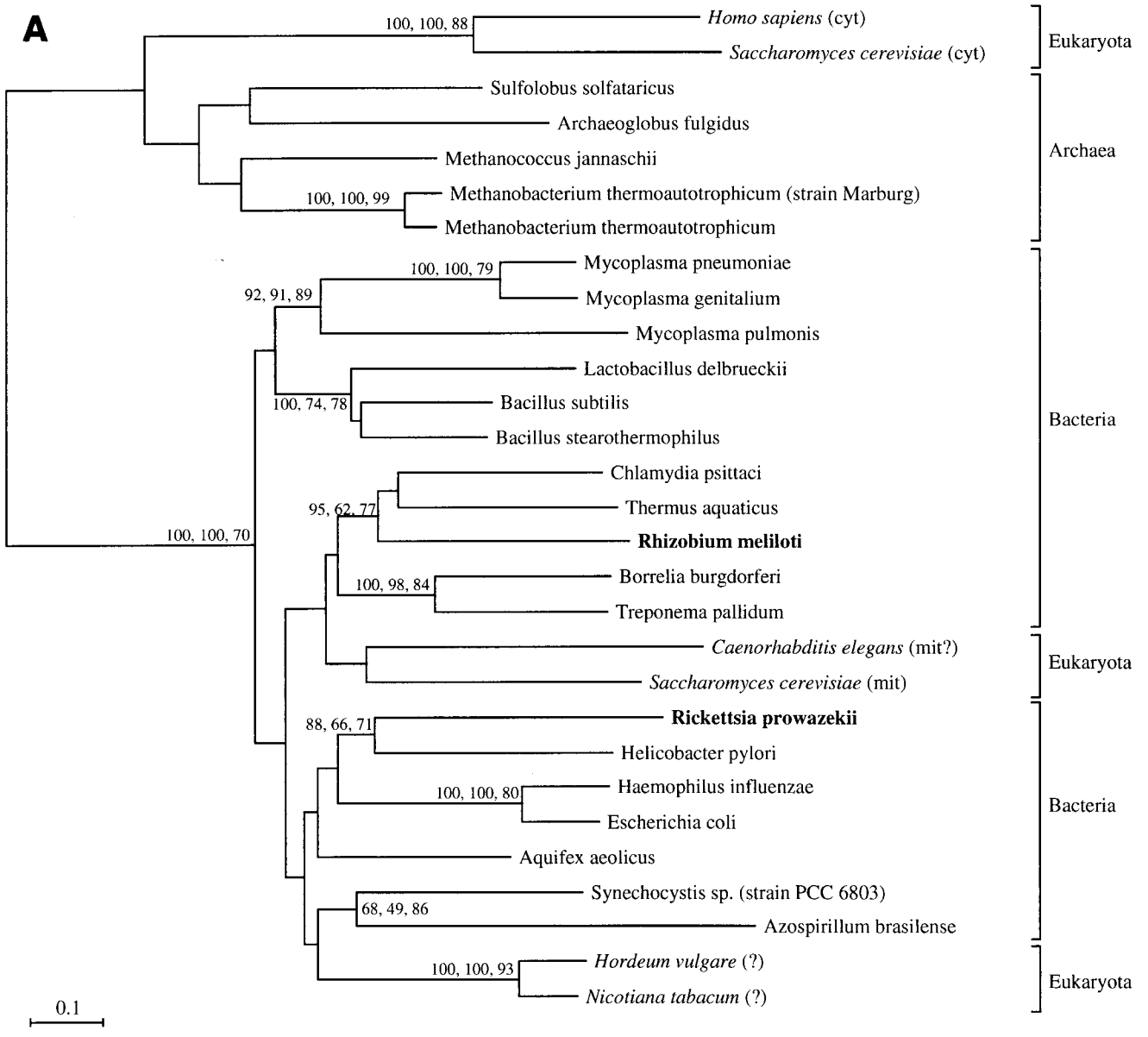

FIG. 10. Schematic illustration of the evolution of aminoacyl-tRNA synthetases. The phylogenetic trees are based on glutamyl-tRNA synthetases (A), arginyl-tRNA synthetases (B), and histidyl-tRNA synthetases (C) from representative species. The phylogenetic trees were constructed as described by Karlberg et al. (95).

proteins, including the phosphate carrier and the uncoupler protein (106, 158, 189). Sequence comparisons have shown that these three types of proteins contain a threefold repeat sequence of approximately 100 amino acids containing two membrane-spanning α-helices (159). The ATP/ADP translocases along with the other two classes of mitochondrial carrier proteins appear to be descendants of a primitive eukaryotic gene that evolved in multiple gene duplications.

Plant cells harbor two seemingly unrelated ATP/ADP transport systems. One is found in the plastid membrane, and the other is found in the mitochondrial membrane. Normally, these two types of transporters function with an opposite polarity with respect to the cytosol: the plastid system imports ATP, and the mitochondrial system exports it. The amino acid sequences of the bacterial ATP/ADP transports, found so far only in *Chlamydia* and *Rickettsia*, are unambiguously related to the plastid homologues, but they have no apparent sequence homology with the mitochondrial transporters (Fig. 13B) (17, 20, 95, 105, 200).

The ubiquity of the mitochondrial ATP/ADP translocases suggests that they arose prior to the divergence of plants, animals, and fungi but after the divergence of mitochondria from α-proteobacteria. The free-living proteobacterial ancestor of the mitochondria is unlikely to have encountered ATP in its environment. Hence, the mitochondrial ADP/ATP translocases could have arisen simultaneously with or shortly after the endosymbiotic event that led to modern mitochondria. It seems that both classes of ATP/ADP translocases may have originated within eukaryotic genomes after the endosymbiotic events that initiated the mitochondria and plastid lineages. The main point here is that that the genes coding for the mitochondrial ATP/ADP translocase as well as those for other characteristic mitochondrial functions do not descend from an α-proteobacterial ancestor.

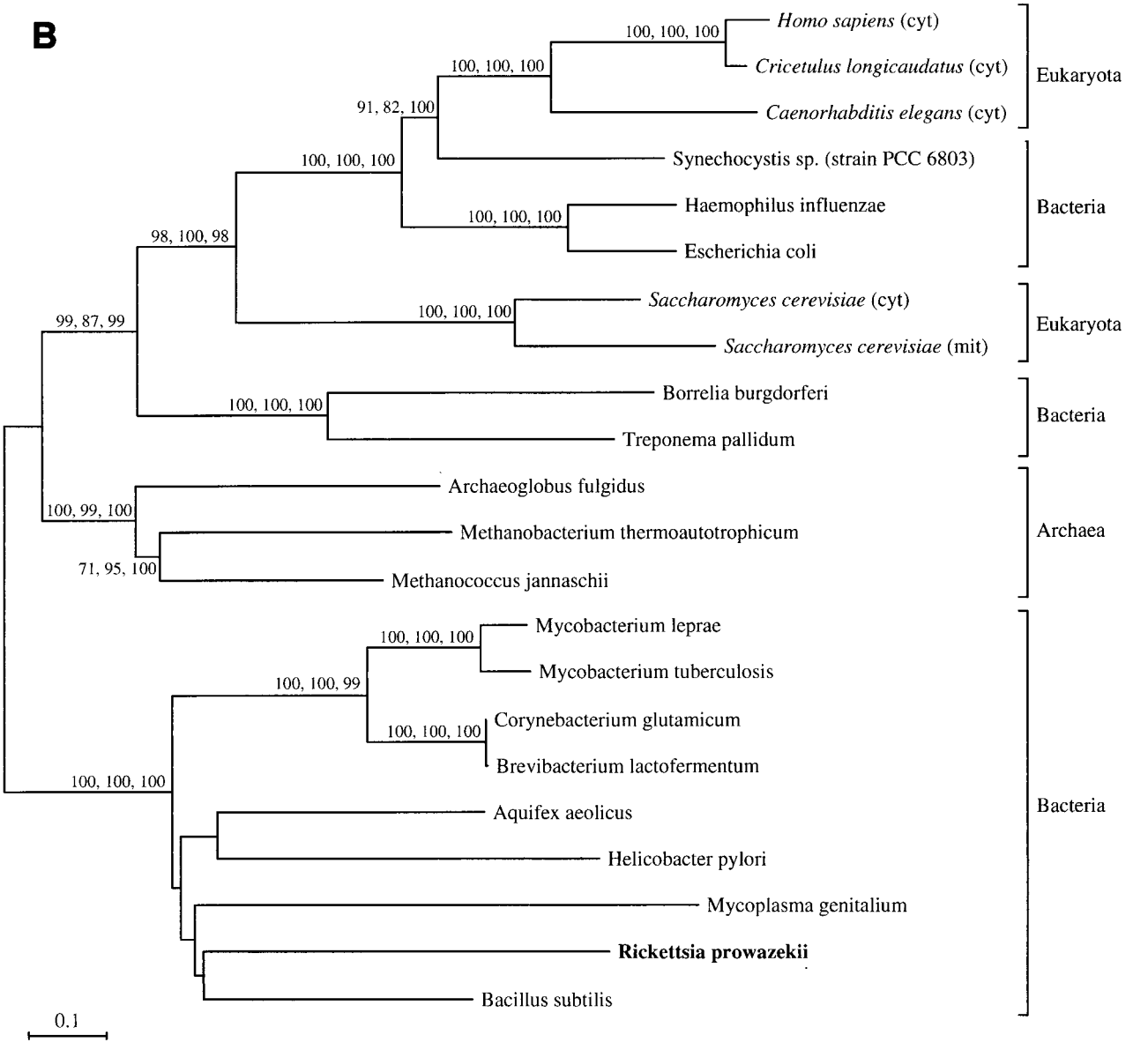

FIG. 10—*Continued.*

Summary

The mitochondrial proteome of the yeast *S. cerevisiae* is the first such cohort for which it has been possible to carry out a comprehensive phylogenetic analysis. Two unanticipated results have emerged (95). On the one hand, the number of α-proteobacterial descendents in this proteome, 47, is surprisingly small. On the other, the number of eukaryotic orthologues without discernable bacterial alignments, 197, is surprisingly large. The clustering of many of these bacterial and eukaryotic orthologues from *S. cerevisiae* with those from the mitochondria of other taxa suggests that *S. cerevisia* is not entirely exceptional. Accordingly, we conclude that the endosymbiotic hypothesis for the origin of mitochondria requires modification.

First, it is clear that the proteomes of mitochondria are descended primarily from two sorts of genomes. One of these is that the genome of the ancestral α-proteobacterium is retained in two forms: first, as the 3 to 67 coding sequences of vestigial mitochondrial genomes, and second, as a still larger set of coding sequences that have been transferred to nuclear genomes. The second genomic source is that of the eukaryotic nucleus, which in *S. cerevisiae* has contributed roughly 200 coding sequences to the mitochondrial proteome. That there are mitochondrial genes in the yeast nucleus not shared by nematodes suggests that new components of the mitochondrial proteome continued to evolve after the divergence of the major eukaryotic branches. Proteins typifying these later, more specialized additions to mitochondrial proteomes are some of the cytochrome oxidase subunits and ribosomal proteins encoded in nuclear genomes. We expect that numerous other proteins have been recruited from within the nuclear genome to fine tune the mitochondria in ever more taxon-specific ways. This expectation will be testable when more genomic data for eukaryotes are available.

The functional identities of the α-proteobacterial and eukaryotic descendants among mitochondrial proteomes are also

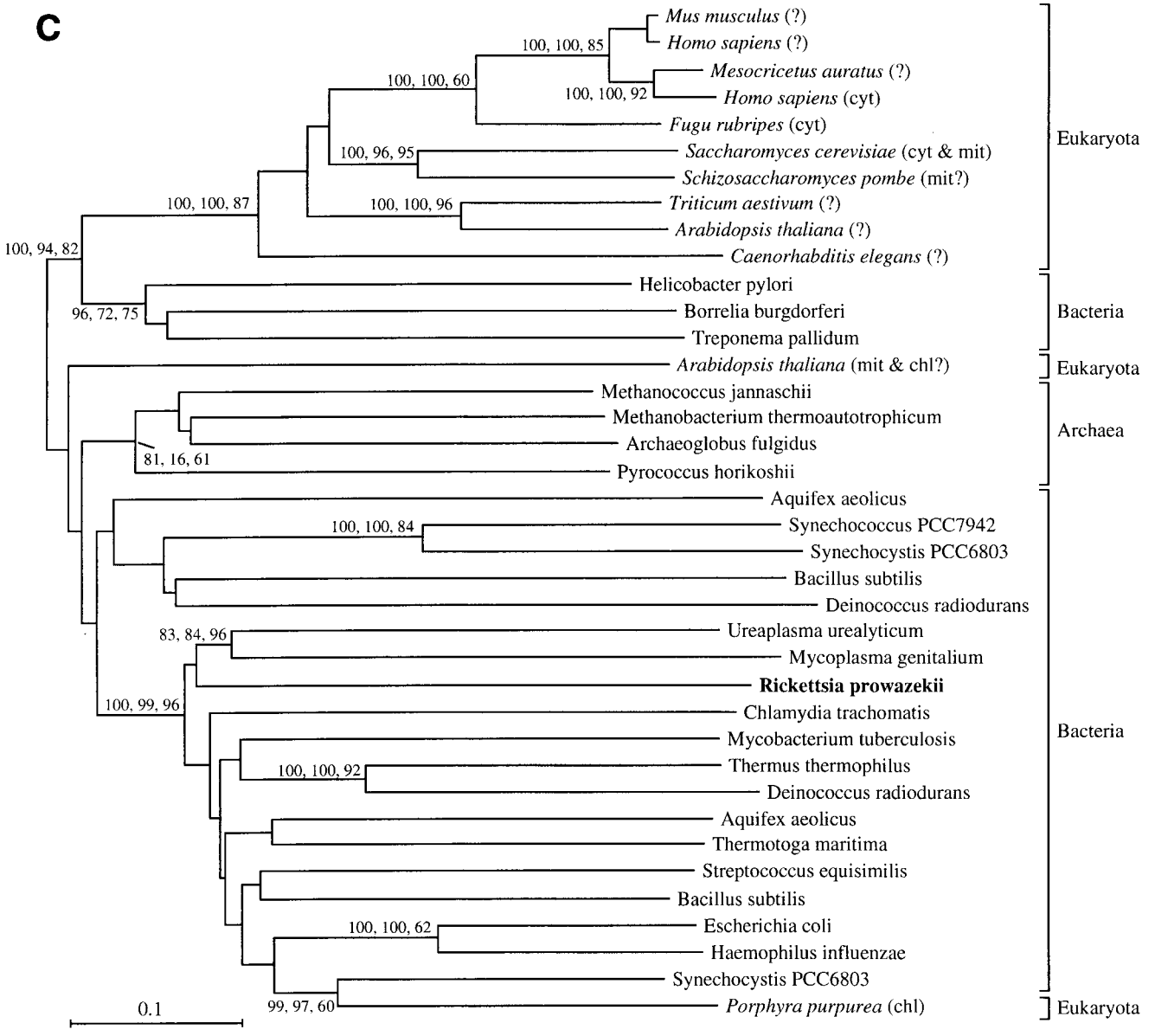

FIG. 10—*Continued.*

revealing. The core respiratory and translation proteins are primarily α-proteobacterial descendants, while accessory proteins in these functional complexes along with transport and regulatory functions tend to be predominantly eukaryotic descendants. Such a nonrandom distribution of functional attribution and phylogenetic descent is consistent with the two modes of genomic evolution that are postulated in the ox-tox model. Thus, the data are consistent with an initial symbiotic relationship by an aerobic symbiont that functions as an oxygen scavenger. Here the symbiont's genome evolves predominantly in a reductive mode, eventually retaining its core functions. In parallel, the host genome contributes new functions to the symbiont's proteome that transform it into an organelle with an ATP export function. Further vertical evolution of the nuclear genomic contribution involves progressively taxon-specific novelties.

THE HOST

A weakness of all current theories of mitochondrial origins is the uncertain identity of the host of the ancestral endosymbiont. Most authors tend to choose an archaeon or an amitochondrial eukaryote as a likely suspect (80, 124, 131, 150, 193). If anything, the ambiguities have been deepened by the genome sequence data acquired recently (49, 63, 151). In this section, we review the recent attempts to identify the host of the α-proteobacterial/mitochondrial lineage and to define the relationship of this lineage to the nuclear genomes of eukaryotes.

Archaeazoa

One recent view of the host, which is thought to have acquired its mitochondria approximately two billion years ago,

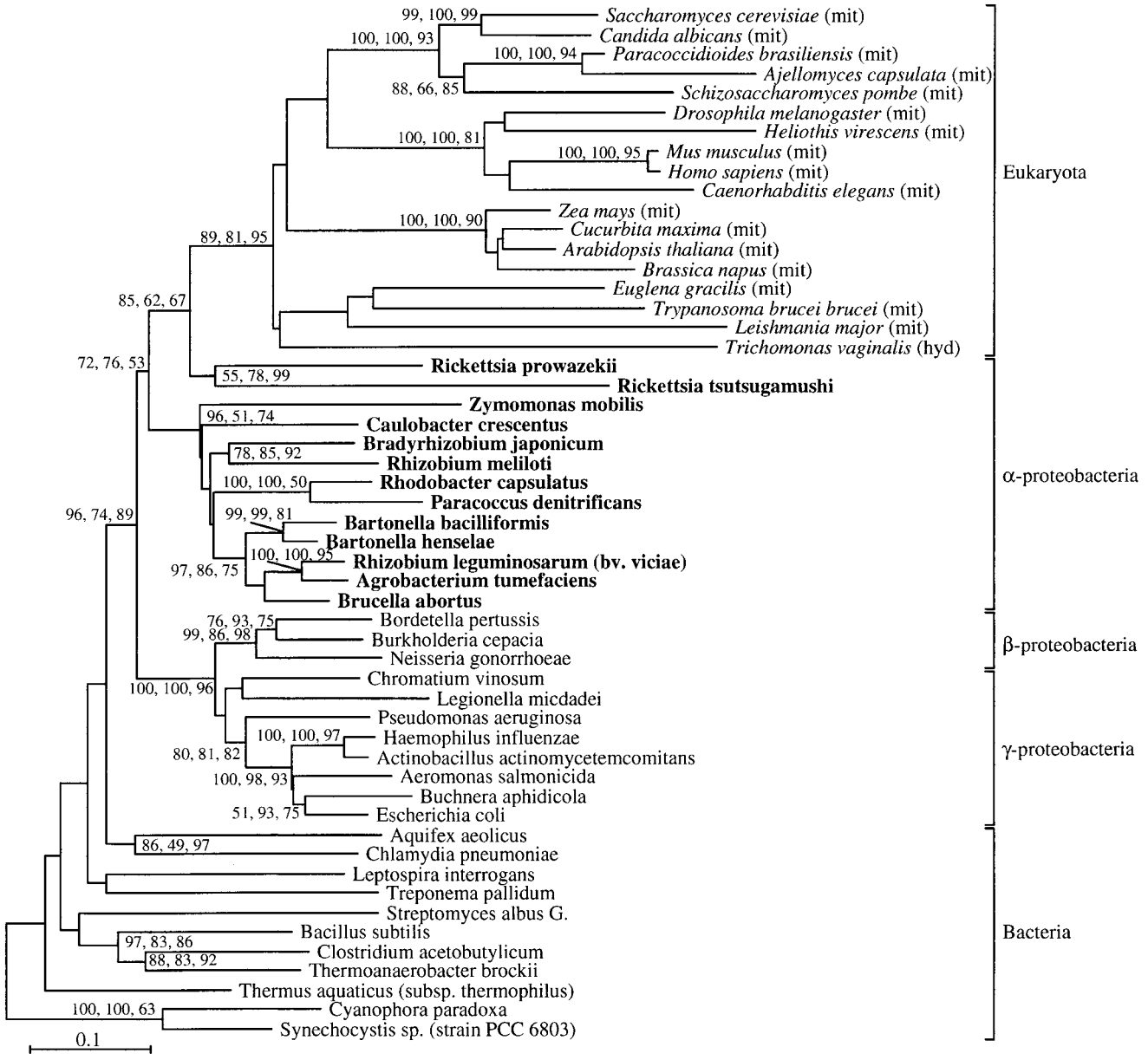

FIG. 11. Phylogenetic reconstructions based on heat shock protein HSP60 from representative species. The phylogenetic trees were constructed as described by Karlberg et al. (95). hyd, hydrogenosome clade; mit, mitochondrion.

was that it was an amitochondrial eukaryote that resembled modern amitochondrial protists, the so-called archaeazoans (42). The most remarkable feature of this attractive interpretation was the speed with which it was dropped following the discovery of what were purported to be mitochondrial homologues in modern archaeazoans (36, 44, 68, 69, 90, 135, 152, 153).

The deconstruction of the archaeazoa began with the demonstration by Viale and Arakaki (186) of close similarities between HSP60 from the *Ehrlichia/Rickettsia* subgroup of α-proteobacteria and homologues from mitochondria. This finding provided a probe with which to search for traces of the latter in amitochondrial protists. Although the sequence comparisons with homologues in *Entamoeba hystolytca* and *Trichomonas vaginalis* were not nearly as detailed or specific, there is little doubt that the data relate heat shock proteins from these amitochondrial protists to homologues in α-proteobacteria (36, 44, 68, 152). One inference from these studies might be that cells with mitochondria were the ancestors of the nominal archaeazoans.

However, similar alignments for a few heat shock proteins do not by themselves demonstrate that these organisms are secondarily amitochondrial. This caveat was particularly pertinent in view of the then mounting claims of substantial horizontal transfer of protein-encoding genes from bacteria to eukaryotic genomes (see below). For these reasons, sequence homologies that supplemented those for heat shock proteins were needed to confirm these far-reaching conclusions.

Keeling and Doolittle (97), in a much-cited paper, analyzed the phylogenetic descent of triosephosphate isomerase (TPI). They presented two phylogenetic reconstructions, one an unweighted parsimony tree and the other a neighbor-joining dis-

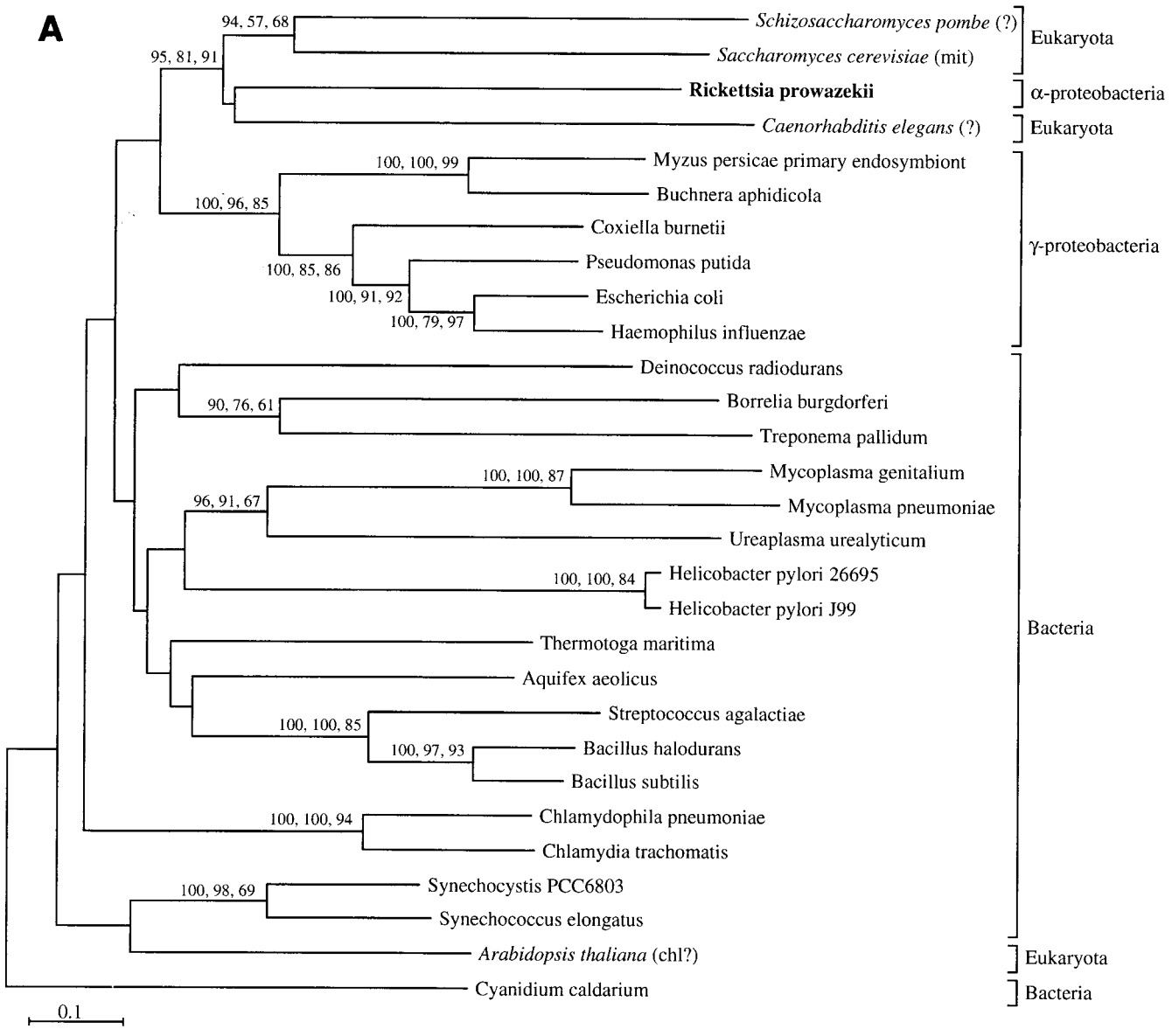

FIG. 12. Illustration of mitochondrial regulatory and transport proteins of bacterial origin. Phylogenetic reconstructions are based on the regulatory protein MSS1 (A) and the ABC transporter protein ATM1 (B) from representative species. The phylogenetic trees were constructed as described by Karlberg et al. (95). mit, mitochondrion; chl, chloroplast; vac, vacuole.

tance tree. The bootstraps for the critical nodes at the divergence of TPI from *Rhizobium etli*, the only α-proteobacterium in the study, from the eukaryotic cluster of TPIs were less than 30 and 50%, respectively. They also inferred that the eukaryotic TPI homologues cluster closely with other proteobacteria and distantly from archaea. The absence of strong support for a specific α-proteobacterial descent of eukaryotic TPI led them to subject the data to further statistical analyses. Here, they could find topologies in which the eukaryotic cluster of TPI was closest to α-proteobacterial TPI, but "the difference between *R. etli* specifically and all the proteobacteria was often insignificant" (97). Nevertheless, they summarize their data by suggesting that "TPI genes present in eukaryotic genomes were derived from an alpha-proteobacterial genome (possibly that of the protoendosymbiont)" (97). They also note that among these putative descendants of α-proteobacterial TPI is that of the amitochondrial *Giardia* spp.

The phylogenetic affinities of the TPI gene family have been reinvestigated using data not available in 1997. An analysis of three α-proteobacterial homologues of TPI together with 50 other homologues from eukaryotes and bacteria extend the findings of Keeling and Doolittle (97). The data confirm the observation that there is no specific relationship between the TPIs of α-proteobacteria and those of eukaryotes (Canbäck et al., unpublished).

Also relevant are the claims for valyl-tRNA synthetase from *T. vaginalis*, which was analyzed by Brown and Doolittle (34). They found that this synthetase clustered with homologues from other eukaryotes along with the homologue from *E. coli*. No homologues from α-proteobacteria were included in the

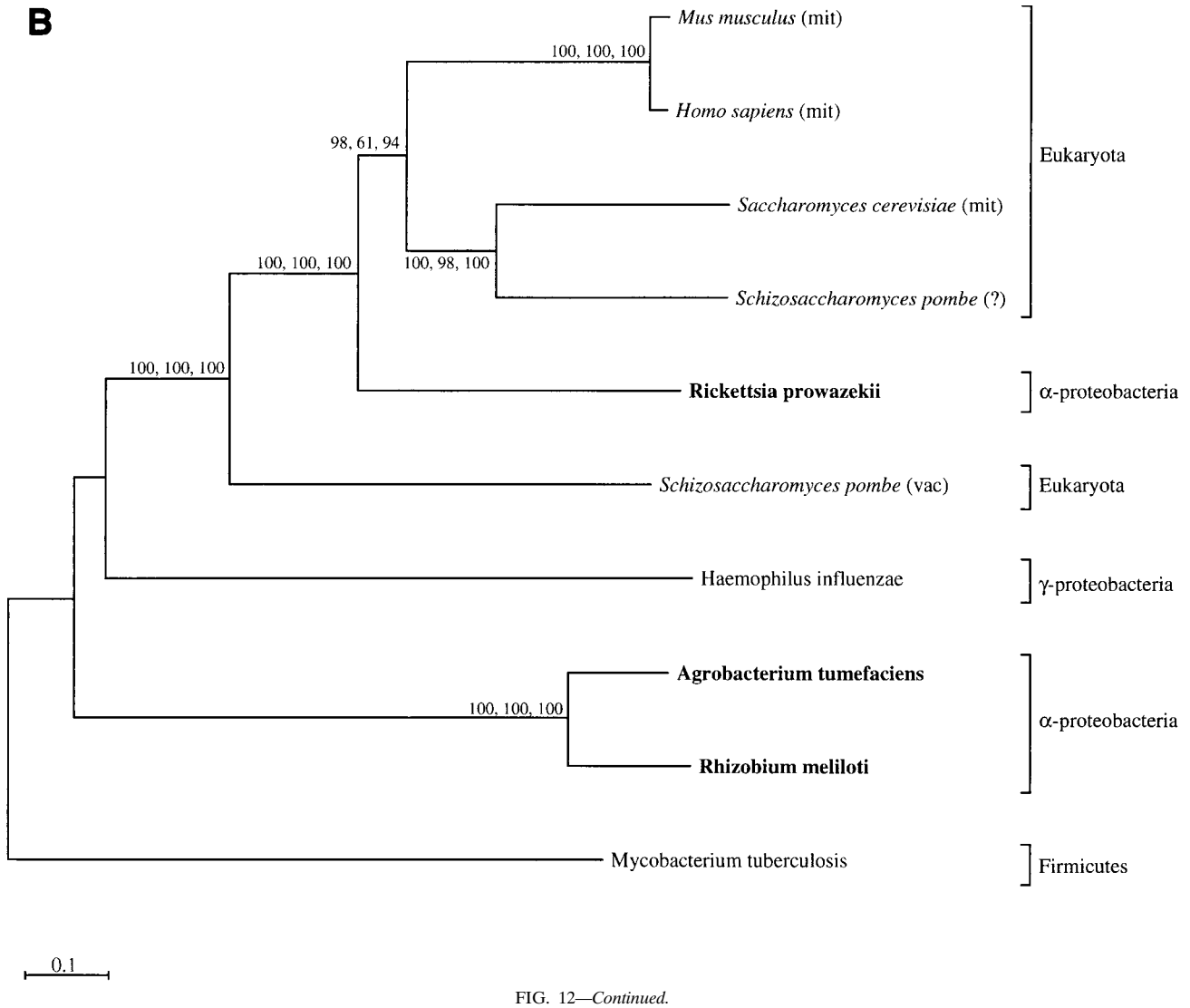

FIG. 12—*Continued.*

analysis, but it was recalled that the synthetase is one that serves both the cytoplasmic and the mitochondrial translation systems in eukaryotes. The authors summarize the analysis with the comment that the eukaryotic Val-tRNA synthetases "may have been transferred from the mitochondrial genome to the nuclear genome, suggesting that this amitochondrial trichomonad once harbored an endosymbiotic bacterium" (34). In fact, there are no data to support this speculation. Instead, as we have observed in the yeast genome (see above), bacterial homologues other than those from α-proteobacteria may be transferred to the nucleus. These in turn may duplicate, and different cellular compartments, including the mitochondria, may recruit their products.

Earlier, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) had been studied in *T. vaginalis*. At that time, Markos et al. noted a clustering of the *T. vaginalis* GAPDH with bacterial homologues, in particular, an *E. coli* homologue (122). This GAPDH was distinguishable from a second *E. coli* homologue that was more closely related to those of other eukaryotes, including *Entamoeba histolytica* and *Giardia lamblia*. No additional claims were made for GAPDH in this study, but in a later paper the widespread occurrence of this gene in bacteria was noted by Henze et al. (85). They reasoned that GAPDH might have transferred to eukaryotes by a process that they refer to as cryptic endosymbiotic transfer. By this, Henze et al. (85) mean transfer of genes from bacteria during abortive or transient endosymbiotic associations that do not result in a new cellular organelle. It was not made clear why endosymbiosis is a requirement for the transfer or how this speculation later became an argument for the mitochondrial origin of GAPDH (64, 139).

Finally, there are the characteristic proteins of the hydrogenosome, pyruvate:ferredoxin oxidoreductase (PFO) and hydrogenase, which are found in some amitochondrial cells. These too have been found to cluster with bacterial homologues, but again with no specific relationship to the α-proteobacteria (89, 91, 93).

Now, it is evident that, aside from the exemplary studies of heat shock proteins, there is no direct support for the interpretation that proteins of the amitochondrial eukaryotes are descendants of the α-proteobacterial/mitochondrial lineages. There is, however, some reason to believe that a number of nominally eukaryotic proteins, including those of the amitochondrial lineages, may be the descendants of diverse bacterial

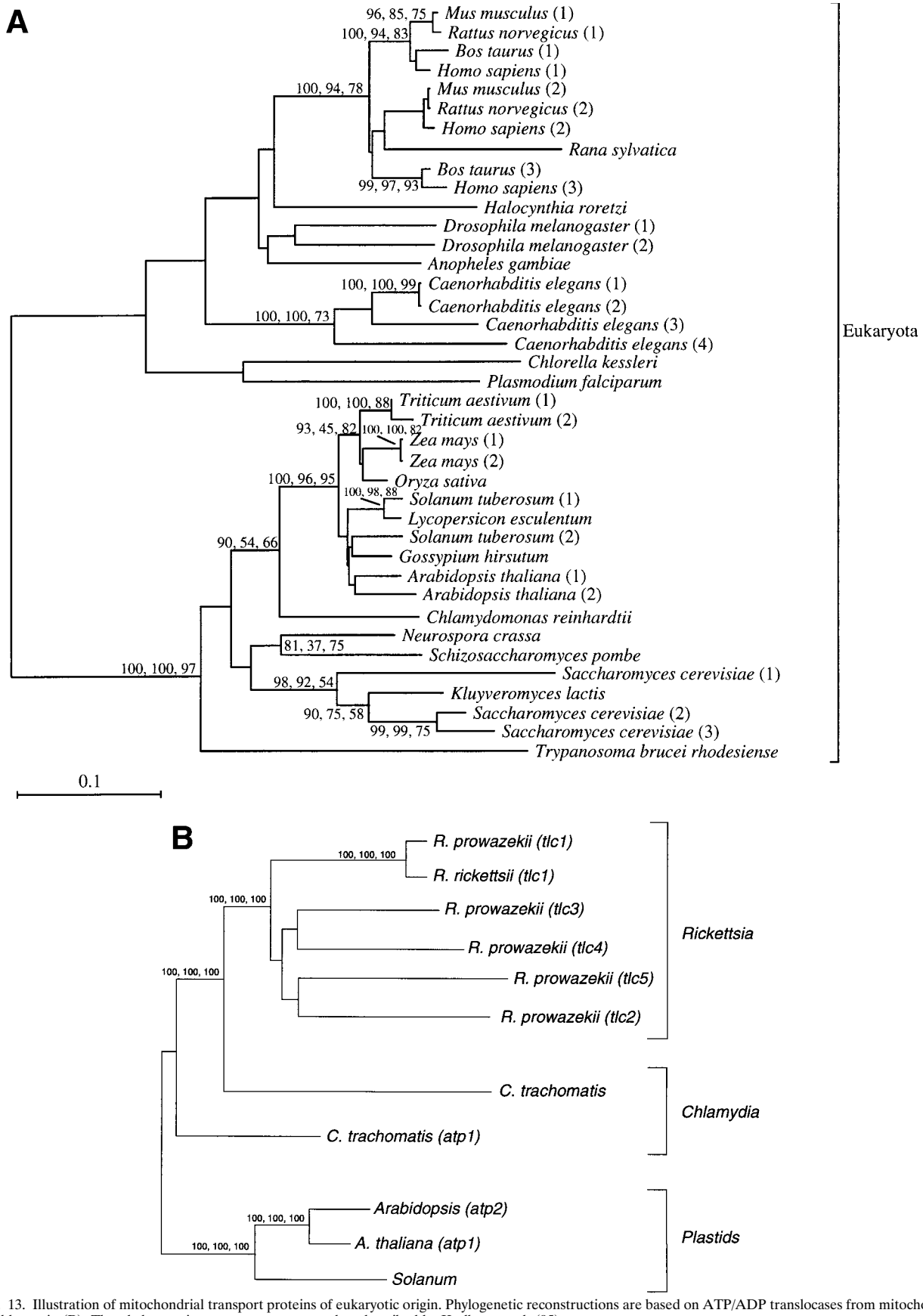

FIG. 13. Illustration of mitochondrial transport proteins of eukaryotic origin. Phylogenetic reconstructions are based on ATP/ADP translocases from mitochondria (A) and bacteria (B). The phylogenetic trees were constructed as described by Karlberg et al. (95).

ancestors. We postpone a resolution of this issue until we consider other data suggesting that the nominal archaeazoans are not taxa from deeply diverging branches of the eukaryote tree (50, 53, 55, 86, 96, 154, 156).

Hydrogenosomes

Hydrogenosomes, like mitochondria, are organelles that process pyruvate, but they do so anaerobically, typically with the aid of PFO and hydrogenase to produce ATP along with H_2 and CO_2. These anaerobic organelles were first described in trichomonads and subsequently found in ciliates as well as in chytrid fungi (57, 114, 133, 134, 136, 179). Chytrid fungi are somewhat atypical in that pyruvate formate-lyase seems to have replaced PFO in their hydrogenosomes (3). Here, a so-called malic enzyme provides the alternative source of H_2 by oxidative decarboxylation of malate to pyruvate, which can then be converted to acetyl-CoA and formate by pyruvate formate-lyase. At least one class of fungal hydrogenosomes seems to have a DNA genome (4).

The production of H_2 by the hydrogenosome is the inspiration for an alternative way to describe the initial endosymbiotic relationship that led to the evolution of mitochondria. Martin and Mueller (124) begin their scenario by identifying the host as a strict anaerobe, such as a hydrogen-dependent autotrophic archaeon. They suggest that this archaeon established a syntrophic relationship with a hydrogen-secreting α-proteobacterium. Here, methane itself or some other organic substrate could be metabolized anaerobically by the symbiont to produce hydrogen, which is consumed by the methanogenic archaeon host. In this scenario, the symbiont is the descendant of a facultative anaerobe that in the initial stage functions like a modern hydrogenosome by producing the molecular hydrogen required by the methanogenic host. H_2-based syntrophy of this sort is seen in some modern anaerobic ciliates (62).

Martin and Mueller (124) postulate that their autotrophic host is bereft of genes to support heterotrophy. They speculate that at a later evolutionary stage of the syntrophic relationship, genes for heterotrophy would be transferred from the genome of the symbiont to that of the host. Accordingly, a prediction of this hypothesis is that genes for heterotrophic metabolic pathways in eukaryotes are the descendants of orthologues from the α-proteobacterial endosymbiont.

Finally, in the hydrogen hypothesis, the symbiont is identified as the common ancestor of both hydrogenosomes and mitochondria (124). However, Martin and Mueller (124) also realized that in order for the dual evolutionary role of the symbiont to be realized, it is necessary that it retain functional copies of the genes for Krebs cycle and respiratory chain functions during the initial anaerobic phase dominated by the anaerobic syntrophy. To solve this problem, they suggest that the anaerobic host could survive intermittent periods of exposure to oxygen assisted by the respiratory system of the symbiont (124). Thus, by functioning occasionally as a scavenger system for toxic oxygen, the respiratory system would be subjected periodically to selective purification, which would preserve its functions until they were recruited for service in mitochondria. Thus, both the hydrogen hypothesis (124) and the ox-tox model (17, 95) exploit in their different ways a similar oxygen-scavenging function to support an aerobic mode for the ancestral endosymbiont. The significant difference between these two aerobic scavenging modes is the timing of their contributions to the evolution of hydrogenosomes.

It had been suggested prior to the publication of the hydrogen hypothesis that hydrogenosomes are degenerate, metabolically transformed mitochondria that support an amitochondrial group of eukaryotic microbes with ATP under conditions of anoxia (30, 42, 54, 65). In this interpretation, the oxygen-scavenging function of the ox-tox model would precede both the evolution of both the mitochondria and their transformation in some lineages into hydrogenosomes. The new interpretation of the hydrogen hypothesis was that hydrogenosome functions were primitive in the sense that the facultative aerobe, which is identified as the ancestor of mitochondria, initially functioned very like a hydrogenosome and exported H_2 its host (124). In this case, the aerobic scavenging function intermittently attends the initial hydrogen-exporting phase. These two principal views of the origins of hydrogenosomes are conveniently referred to as hydrogen exporter-late and hydrogen exporter-early, respectively.

Hydrogen Exporter—Late

Diplomonads and microsporidia together with trichomonads were previously identified in phylogenetic reconstructions as the earliest-diverging eukaryotes; accordingly, they were included in the archaeazoa (42). Neither mitochondria nor hydrogenosomes are found in diplomonads or microsporidia. It seemed for a while that some archaeazoa might be latterday versions of the primordial eukaryotic cell prior to its acquisition of organelles. That view was enjoyed until the heat shock proteins of the α-proteobacterial/mitochondrial lineage were detected among the archaeazoa. In fact, the phylogenetic data for heat shock proteins do not distinguish mitochondrial, endosymbiotic, or more casual donors of the transferred heat shock protein genes to cells with hydrogenosomes. However, other observations help to establish a relationship between mitochondria and hydrogenosomes.

For example, double membranes reminiscent of those of mitochondria contain the hydrogenosomes of trichomonads as well as of ciliates and fungi (42, 65, 181, 182). Likewise, targeting sequences that resemble the mitochondrial targeting sequences in related organisms have been observed in coding sequences for some hydrogenosomal enzymes (33, 181, 182). Finally, enzymes found typically in both fungal mitochondria and the chytrid fungi lack the mitochondrion-targeting sequences in the chytrids. The revealing point is that these exceptional chytrid proteins are found in the cytosol of the fungus but not in hydrogenosomes (2).

Accordingly, it can be argued that one adaptation to anoxia by cells with mitochondria is to retarget some enzymes of this organelle to the cytosol and to replace these enzymes with those that support the activities of hydrogenosomes (2, 3, 54, 56, 128). In such scenarios, the mitochondrial "bag" can be retained along with the transport system for recruiting proteins to the organelle. Reassignment of targeting sequences specific to this transport system leads to a functional refitting of the organelle that is more appropriate to its evolving anaerobic function. It may be relevant that cells with both hydrogenosomes and mitochondria are not observed even though cells with comparable metabolic capacities are viable (65). Presumably, the evolutionary transition between the functional modes of the bag is rapid.

More generally, biophysical studies have revealed striking similarities between the two organelles with respect to their transmembrane pH and electrochemical gradients (30, 127). This makes it easier to accept the idea that the same sort of proton motive force-dependent transport system that recruits proteins to the mitochondria from the cytosol could recruit the alternative proteins of the hydrogenosome. Nevertheless, there is a missing link in the hydrogen exporters-late scenario. This

concerns the origins of the hydrogenosome-specific functions, to which we return below.

Hydrogen Exporter—Early

The hydrogenosomes of ciliates have been viewed as problematic because they are found among clades of anaerobic ciliates that are near relatives of aerobic clades with mitochondria (54, 62, 183, 184). Thus, the phylogenetic reconstructions suggest that the hydrogenosomes in these clades of ciliates must have arisen independently several times. We are therefore confronted with a choice between two alternatives. Either the hydrogenosome as a whole was reinvented independently in each of the relevant ciliate clades, or the mitochondria in these clades have devolved some of their proteins, which have been replaced by those that are characteristic of hydrogenosomes. Because it involves many fewer events, the transformation of mitochondria into hydrogenosomes in ciliates seems to be the most probable of the two scenarios (54).

Also relevant are recent revisions of the phylogenetic status of members of the nominal archaeazoa. These revised phylogenies have eroded the view that the nominal archaeazoa represent a coherent ancestral group of amitochondrial eukaryotes. Among the first of the archaeazoa to be promoted to the crown of the eukaryote tree were the microsporidia, for which phylogenetic reconstructions with tubulin and heat shock protein sequences favor a late divergence close to the fungi (53, 86, 96). The next to be displaced was *Entamoeba* (*Archamoeba*), for which rRNA sequences were found to diverge later than originally thought (169). Most recently, the placement of trichomonads and *Giardia* early in the eukaryotic tree has been challenged on the basis of phylogenetic reconstructions with RNA polymerase orthologues (86). It would seem that the phylogenetic interpretations favoring the creation of the archaeazoa rested in part on reconstructions that were pressed beyond the limits of reliability (86, 148, 149).

If the earliest-diverging clades of eukaryotes contained mitochondria, while all those containing hydrogenosomes are higher in the tree, the hydrogen exporter-early hypothesis becomes rather complicated. Thus, to preserve this scenario, it is necessary to assert the following. First, the hydrogenosome-like functions arose in and supported the ancestral α-proteobacterial endosymbiont under anaerobic conditions. Second, under subsequent aerobic conditions, the genes for hydrogenosome-specific functions were purged from the evolving mitochondrion and replaced by aerobic enzymes. Finally, during a return to anaerobic conditions, the mitochondria were purged of aerobic enzymes and refitted with proteins to support hydrogenosome functions. This is not an impossible scenario but is extremely improbable because it takes so much undoing and redoing. Furthermore, evidence that the ancestral symbiont engaged in H_2-dependent syntrophy is notable by its absence.

For example, we might expect from the hydrogen exporter-early hypothesis that characteristic hydrogenosome proteins such as PFO are the descendants of α-proteobacterial homologues. While it seems from the limited data available that this protein may be a descendant of bacterial homologues, its roots are found in bacteria distant from α-proteobacteria (86, 91, 100, 102). Conceivably, this and other hydrogenosomal proteins may have been recruited to the ancestor by horizontal transfer. However, this notion does little to strengthen the hydrogen exporter-early interpretation, since it is equally applicable to the hydrogen exporter-late interpretation.

Eukaryotic Heterotrophy

Finally, we come to the suggestion that the heterotrophic metabolism of eukaryotes originates with the α-proteobacterial endosymbiont (124). It is an undisputed fact that glycolysis and glucogenesis are almost as widespread among archaea, bacteria, and eukaryotes as are ribosomes (46, 99, 161, 164, 165, 194). Furthermore, even the facultatively autotrophic archaeon *Methanococcus maripalusis* has its versions of these enzymes (207). Accordingly, it is not clear what problem is solved by speculating that a methanogenic archaeon was the beneficiary of a transfer of genes for heterotrophy from the ancestral α-proteobacterial endosymbiont (124). Nevertheless, this speculation has its testable prediction, which is that a significant number of genes for heterotrophic metabolism in eukaryotes should be related to those of α-proteobacteria.

Initially, the conclusion of Keeling and Doolittle (97) that TPI of eukaryotes is specifically related to the TPI from *R. etli* had been forwarded as evidence for the α-proteobacterial origins of eukaryotic heterotrophy (124). However, as we have seen, the data of Keeling and Doolittle (97) do not in fact specifically relate the TPI from *R. etli* to eukaryotic taxa (Canbäck et al., unpublished). Similarly, the phylogenetic data summarized by Martin and Schnarrenberger (125) for other glycolytic enzymes in eukaryotes suggest that none of these is specifically related to homologues in α-proteobacteria. The sequencing of *Bartonella henselae's* genome (Andersson et al., unpublished data) has provided data that can be used for complementary phylogenetic reconstructions. The briefest description of these data is that there is no example of a eukaryotic glycolytic enzyme in this cohort (circa 50 eukaryotic taxa) that is descended from an α-proteobacterial orthologue (Canbäck et al., unpublished), which confirms the conclusion drawn from previous studies of glycolytic enzyme families.

In order to preserve the hydrogen hypothesis, it has been argued that the failure to detect α-proteobacterial descendants among the eukaryotic glycolytic enzymes is a reflection of extensive gene replacements by horizontal gene transfer from other, possibly endosymbiotic bacteria (125). Nevertheless, the α-proteobacterial descent of 50 different mitochondrial proteins has been documented (95). Likewise, another cohort of roughly 200 eukaryotic descendants of mitochondrial proteins has been established (95). One may wonder how the phylogenetic coherence of these cohorts escaped the inroads of the putative gene transfer from bacteria. In other words, why should the transfer from other bacteria specifically obscure the α-proteobacterial descent of glycolytic enzymes? It seems probable that part of the answer to these questions is that horizontal gene transfer is not as frequent as assumed by Martin and Schnarrenberger (125). Another part of the answer may be that none of the eukaryotic glycolytic enzymes are specifically descended from α-proteobacterial homologues.

Horizontal Transfer to and from Eukaryotes

In the section on archaeazoa, we reviewed earlier studies that inferred a mitochondrial descent for eukaryotic genes even when no sequence comparisons specifically related these genes to the α-proteobacterial/mitochondrial lineage. The most global application of this sort of phylogenetic inference was provided by Gogarten et al. (73).

The stimulus for Gogarten et al. (73) was an admittedly challenging paper by R. F. Doolittle et al. (49). Here, evidence of archaeal origins for eukaryotic sequences encoding the functions of gene expression was reinforced and supplemented by the discovery of sequence similarities with diverse bacterial taxa for the eukaryotic genes of intermediary metabolism (49).

The reliability and scope of these inferences were confirmed and extended subsequently (63, 151). In addition, extrapolations from extant sequences suggested that at least some of these same eukaryotic genes of bacterial descent diverged from their prokaryotic homologues roughly two billion years ago (49). The sensational implication of the paper by R. F. Doolittle et al. (49) is that the last common ancestor was extant two billion years ago.

Two billion years is very roughly the midline for life on this planet. It also corresponds to what is thought to be the time when the endosymbiotic events leading to the acquisition of organelles by eukaryotes took place. This prompted Gogarten et al. (73) to suggest an alternative interpretation in which bacterial homologues were transferred to the genomes of eukaryotes from the endosymbiotic bacteria that initiated organelle evolution. This interpretation neatly displaced the then uncomfortable inference that the last common ancestor of all organisms was extant a mere two billion years ago. A fresh analysis with additional data led Feng et al. (63) to endorse this interpretation. Nevertheless, it is worth emphasizing that what was observed in these simple sequence comparisons was that a significant number of genes encoding enzymes of intermediary metabolism in eukaryotes seem to be more similar to bacterial than to archaeal homologues. Evidence for a specific descent of these transferred genes from α-proteobacteria or cyanobacteria was conspicuously missing (49, 63, 151).

Rivera et al. (151) have referred to these metabolic genes as those encoding operational functions. In addition, a significant fraction of the genes encoding functions in replication, translation, and transcription also seem to be of ancient dignity but more closely related to archaeal homologues (49, 63, 151). These are referred to as informational genes (151). The most recent analysis of these genomic sequences suggests that the bacterial operational genes have not been introduced into eukaryotic genomes in a single burst, as might be expected from the hypothesis of Gogarten et al. (73). According to Jain et al., these transfers must have occurred gradually (92, 109).

What Jain et al. (92) observed is that the diversity of rootings for paralogous genes and the branching topologies for the informational and operational lineages are systematically different. They observed significantly fewer independent roots for phylogenetic trees within the archaeal lineage compared to the bacterial lineage. These, along with the branching patterns, are not consistent with a single burst of transfer for bacterial genes. In other words, models in which bacterial transfers occurred exclusively prior to or simultaneous with the endosymbiotic events leading to the mitochondria are not supported. Likewise, models based on cellular or genomic fusion between archaea and bacteria are not supported. Finally, exclusive transfer from an endosymbiont that already contained horizontally recruited genes from other bacteria is not supported. The patterns observed by Jain et al. (92) are consistent with there having been intermittent recruitment of bacterial operational genes by eukaryotes during the past two billion years.

We recall here our discussion of the transfer of genes from mitochondria to nuclei that has been studied experimentally by Thorsness, Fox, and their colleagues. They have suggested that such transfers result from the destruction of mitochondria by autophagous lysosomes, as is observed for peroxisomes (175). Obviously, this cellular mechanism may also account for the transfer of diverse bacterial genes to eukaryotic genomes. Thus, bacteria apparently never enter eukaryotic cells without being engulfed in one membranous structure or another. Such membranous structures normally mature into or fuse with a lytic organelle (29) that may release nucleic acids from the lysed bacteria and thereby provide a pathway for horizontal gene transfer to the nucleus.

Some pathogenic bacteria as well as protists have well-developed molecular strategies for facilitating their interaction with membranous structures to stimulate cellular engorgement by stepwise (zipper) or abrupt (trigger) mechanisms (reviewed in reference 67). The modes of intrusion are varied (reviewed in reference 167). Some pathogens modify the normal incorporation of host signal proteins into the so-called pathogen vacuoles, thereby retarding their maturation into functional phagolysosomes. Others inhibit fusion of their vacuoles with the phagolysosomal vacuoles. Still others actively lyse the membranous compartments and escape into the cytosol. Thus, cells that survive such intrusions may have the opportunity to recruit bacterial nucleic acid sequences into their nuclear genomes. Likewise, pinocytosis of bacteria (29) will provide an alternative route for lateral transfer from bacteria to eukaryotes. For example, protists that graze on nonpathogenic bacteria may also gather these in vacuoles that eventually release nucleic acid fragments and affect gene transfer from fodder to feeder (51). The point then is that intracellular associations of diverse bacteria with eukaryotes are common and recurrent. These almost certainly provide opportunities for a continuous procession of horizontal gene transfers from bacteria to eukaryotes.

A model of vacuole-dependent transfer of bacterial genes to eukaryotic nuclear genomes is consistent with the apparent diversity of bacterial ancestors to genes in eukaryotic genomes (49, 63). It is also consistent with the intermittent accumulation of such transfers during the past two billion years (92). Thus, we may imagine that bacterial genes are transferred to the nuclear genomes of eukaryotes by random mechanisms and that they are fixed in the eukaryotic populations either by drift or by selection (27). For example, a eukaryotic cell might acquire by horizontal transfer a second gene for an indigenous enzyme. If the two forms of the enzyme are functionally equivalent, the services of one or the other may be lost as a result of random mutation. The survivor of this neutral process might be the bacterial form of the enzyme.

Alternatively, a novel or redundant bacterial gene may be acquired through the random transfer process and then selected. Here, the recruited gene might encode a protein that either is superior in function to the eukaryotic analogue or functions as a stimulant or suppressor of some defective function of the eukaryote. In these cases, selection may eventually fix the bacterial variant in the eukaryote's genome. Indeed, the transfer may involve large coding sequences, as in the case of an *Arabidopsis* strain that has been found to have a nearly complete mitochondrial genome copy transferred to its nuclear genome (113).

There is an important detail that, to our knowledge, has been omitted from discussions of horizontal gene transfer. In order for horizontal transfer events to be propagated in multicellular eukaryotes, the transfers must be established in the germ line. Bacterial infections of the gonads in humans are rare, as indicated by the anecdotal observation that no medical practitioner that we have interviewed has ever seen such an infection. This may explain another anecdotal observation: reports of horizontal gene transfer to single-celled eukaryotes seem to be more prevalent in the literature than reports of such transfer to multicellular organisms. Furthermore, inspection of the relevant phylogenetic reconstructions suggests that a number of the horizontal transfers that have been recorded among metazoans may in fact be from animal to pathogenic bacterium.

Phylogenetic Inference and Gene Transfer

Although the putative transfer process from bacteria or archaea to eukaryotes is likely to be a random one, the fixation of transferred coding sequences in populations is likely to be highly selective. For example, it is not given that the transferred gene product is neutral or advantageous in its new cellular surrounding. It might be that the transferred gene is toxic or inhibitory (125). Thus, the cooperative interactions of the transferred gene product within the new cellular environment are likely to determine its fate. Woese and his colleagues (74, 198, 199) have developed this notion in an evolutionary context.

Woese (198) argues that the universal ancestor to all modern lineages was a population of relatively simple organisms with highly error-prone replication and expression systems. A telling characteristic of this life form was that it could not raise extensive barriers against horizontal gene transfer that make this sort of exchange a relatively infrequent event among modern organisms. On the contrary, Woese infers that horizontal gene transfer was so prevalent that it was the dominant mode of genome evolution for the universal ancestor. The transition to the contemporary situation, in which vertical transmission is the dominant mode of genome evolution, was accompanied by progressive enhancement of the barriers to horizontal transmission. Among these barriers are more refined molecular systems that operate with higher precision and depend on increasingly complex cooperative interactions between gene products. The increased cooperativity between components of particular molecular systems may then express itself as a barrier to the incorporation of foreign gene products at different times in evolution. So, for example, Woese (198) suggests that gene expression systems may have become more resistant to lateral transfers early on. At the other extreme, Woese (198) points to solitary enzymes of metabolism as proteins that can more freely wander by gene transfer between modern organisms.

Thus, enzymes such as the aminoacyl-tRNA synthetases seem to be subject to lateral transfer relatively frequently (199, 200). Indeed, such transfer from archaea to α-proteobacteria has been observed in the *Rickettsia* genome (20). Likewise, several transfer events between bacteria and eukaryotic genomes have been observed for the enzymes of glycolysis and glucogenesis (125; Canbäck et al., unpublished). In contrast, several hundred genes have been identified among genome sequences as being unique to archaea (74). These genes make up the genomic signature of archaea and include, for those that can be functionally classified, genes for nucleic acid processing, flagellum biosynthesis, and single-carbon metabolism (74). Similarly, the division of genes into informational and operational lineages that are, respectively, rarely or often subject to lateral gene transfer is explained by Jain et al. (92) in their complexity hypothesis. Here, members of the informational lineages are active in complex functional systems, while the operational ones are less so. As a consequence, members of the informational lineage transfer less frequently than do members of the operational lineage.

W. F. Doolittle (52, 168) has for some time been pondering the extent and diversity of horizontal gene transfer. He asks (52) whether horizontal gene transfer "can be treated as just a nuisance in phylogenetic classification, or is it the essence of the phylogenetic process (at least for prokaryotes and the earliest eukaryotes) and thus a threat to the whole enterprise of classification?" It seems to us that these alternatives are too extreme.

Horizontal gene transfer events, like other common genomic events such as deletions and other sorts of insertions, create discontinuities in the phylogenies of genes or genomes, but these discontinuities are not "lethal" to phylogeny. Rather, such singularities slightly complicate phylogenetic reconstructions of genes and genomes. These events can even provide useful clues to the phylogeny of lineages that they have marked. Furthermore, we do not believe that the solution to the conceptual problems created by horizontal gene transfer will be to "imagine organismic lineages to have a sort of emergent reality," as suggested by W. F. Doolittle (52). Rather, we would recommend focusing attention on the history of genomes as the basis of phylogeny. From this point of view, horizontal gene transfer is very obviously important but is not the "essence of the phylogenetic process" (52). Indeed, the analysis of the mitochondrial proteome that we have described here and elsewhere (95) suggests that even such a complex evolutionary entity can be analyzed by augmented conventional methods.

The discovery of hundreds of gene sequences that make up the genomic signature of archaea (74) confirms a fundamental phylogenetic division that was derived from simple phylogenetic reconstructions with rRNA sequences. Thus, horizontal gene transfer seems not to have eroded the reality of the archaea as a distinct phylogenetic superkingdom. Likewise, nuclear genes found in all eukaryotic taxa that encode mitochondrial proteins and that have only eukaryotic homology are likely to be members of the genomic signature of the eukaryotes. Such a characteristic gene family is the mitochondrial ATP/ADP translocase (Fig. 13). It does not seem unduly optimistic to suggest that the immediate future will witness the description of the characteristic gene families that make up the genomic signatures of the third domain defined by Woese and his colleagues on the basis of rRNA phylogeny (197). The irony here is that horizontal gene transfer events are likely to provide important data with which to trace genomic phylogeny. Furthermore, it remains to be seen whether or not phylogeny based on single gene families such as rRNA and some selected proteins (47, 116, 187, 197) is fatally complicated by the intrusions of horizontal gene transfer (52).

FUTURE DIRECTIONS

Phylogenetic reconstructions for more than 400 homologues are the basis of our analysis of the origins of mitochondrial proteomes (95). Although this number of phylogenetic reconstructions may seem large by current standards, it is only a modest portent of what we may expect from future genomic and proteomic studies. Proper analysis of even such small genomes as those of archaea and bacteria will require thousands of phylogenetic reconstructions. Comprehensive phylogenetic studies are necessary if we are to avoid the distortions and ambiguities inherent in studies based on the pairwise sequence comparisons favored by many recent authors. Routine procedures to carry out phylogenetic reconstructions for the gene cohorts of whole genomes are certain to be available in the immediate future.

Comparisons of sequence identity or similarity between pairs of homologues might give the impression that there has been intensive horizontal gene transfer from bacteria to eukaryotic genomes, particularly for enzymes of intermediary metabolism (52, 73, 124). In fact, such comparisons do not distinguish sequence similarities arising by gene transfer from those arising from common descent. Routine phylogenetic algorithms applied to a cohort of circa 400 proteins identified nearly 50 α-proteobacterial and roughly 200 eukaryotic homologues (95). The phylogenetic coherence of these gene families suggests to us that the global influence of horizontal gene transfer from bacteria to eukaryotes has been vastly overesti-

mated (52, 73, 124). More comprehensive phylogenetic comparisons should settle this issue in the near future.

The members of the yeast mitochondrial proteome that have been identified as eukaryotic descendants are not found uniformly in all eukaryotic genomes (95). Accordingly, once the mitochondrial proteomes of other organisms have been more extensively characterized, it will be possible to study the evolutionary processes that introduce as well as expunge the proteins recruited to the organelle. In particular, we expect the patterns of loss and gain within mitochondrial proteomes to provide new insights into cell evolution and eukaryote phylogeny.

ACKNOWLEDGMENTS

We thank Carl Woese for stimulating exchanges. We thank him along with Dieter Söll for sharing unpublished data. We also thank Russell Doolittle for helpful comments. Björn Canbäck and Olof Karlberg have helped us in many ways, not the least by keeping us informed about their work. We are grateful to Irmgard Winkler for help with the references and Olof Karlberg for help with the figures.

Our work is supported by the Swedish Natural Research Council, by the Strategic Research Fund, and by the Knut and Alice Wallenberg Foundation.

REFERENCES

1. **Adams, K. L., K. Song, P. G. Roessler, J. M. Nugent, J. L. Doyle, J. J. Doyle, and J. D. Palmer.** 1999. Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes. Proc. Natl. Acad. Sci. USA **96:**13863–13868.
2. **Akhmonova, A., F. G. J. Voncken, H. Harhangi, K. M. Hosea, G. D. Vogels, and J. H. P. Hackstein.** 1998. Cytosolic enzymes with a mitochondrial ancestry from the anaerobic chytrid Piromyces sp. E2. Mol. Microbiol. **30:** 1017–1027.
3. **Akhmanova, A., F. G. J. Voncken, K. M. Hosea, H. Harhangi, J. T. Keltjens, H. op den Camp, G. D. Vogels, and J. H. P. Hackstein.** 1999. A hydrogenosome with pyruvate formate-lyase: anaerobic chytrid fungi use an alternative route for pyruvate catabolism. Mol. Microbiol. **32:**103–114.
4. **Akhmanova, A., F. Voncken, T. van Alen, A. van Hoek, B. Boxma, G. Vogels, M. Veenhuis, and J. H. P. Hackstein.** 1998. A hydrogenosome with a genome. Nature (London) **396:**527–528.
5. **Allen, J.** 1993. Control of gene expression by redox potential and the requirement for chloroplast and mitochondria genomes. J. Theor. Biol. **165:** 609–631.
6. **Andersson, D. I., and D. Hughes.** 1996. Muller's ratchet decreases fitness of a DNA based microbe. Proc. Natl. Acad. Sci. USA **93:**906–907.
7. **Andersson, J. O., and S. G. E. Andersson.** 1997. Genomic rearrangements during evolution of the obligate intracellular parasite *Rickettsia prowazekii* as inferred from an analysis of 52015 bp nucleotide sequence. Microbiology **143:**2783–2795.
8. **Andersson, J. O., and S. G. E. Andersson.** 1999. Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev. **9:**664–671.
9. **Andersson, J. O., and S. G. E. Andersson.** 1999. Genome degradation is an ongoing process in Rickettsia. Mol. Biol. Evol. **16:**1178–1191.
10. **Andersson, S. G. E.** 1998. Bioenergetics of the obligate intracellular parasite *Rickettsia prowazekii*. Biochim. Biophys. Acta **1365:**105–111.
11. **Andersson, S. G. E., A.-S. Eriksson, A. K. Naeslund, M. S. Andersen, and C. G. Kurland.** 1996. The *Rickettsia prowazekii* genome: a random sequence analysis. Microb. Comp. Genomics **4:**293–315.
12. **Andersson, S. G. E., and C. G. Kurland.** 1990. Codon preferences in free-living microorganisms. Microbiol. Rev. **54:**198–210.
13. **Andersson, S. G. E., and C. G. Kurland.** 1991. An extreme codon preference strategy: codon reassignment. Mol. Biol. Evol. **8:**530–544.
14. **Andersson, S. G. E., and C. G. Kurland.** 1995. Genomic evolution drives the evolution of the translation system. Biochem. Cell Biol. **73:**775–787.
15. **Andersson, S. G. E., and C. G. Kurland.** 1998. Reductive evolution of resident genomes. Trends Microbiol. **6:**263–268.
16. **Andersson, S. G. E., and C. G. Kurland.** 1998. Ancient and recent horizontal transfer events: the origins of mitochondria. APMIS Suppl. **84:**5–14.
17. **Andersson, S. G. E., and C. G. Kurland.** 1999. Origin of mitochondria and hydrogenosomes. Curr. Opin. Microbiol. **2:**535–541.
18. **Andersson, S. G. E., and P. M. Sharp.** 1996. Codon usage and base composition in *Rickettsia prowazekii*. J. Mol. Evol. **42:**525–536.
19. **Andersson, S. G. E., D. R. Stothard, P. Fuerst, and C. G. Kurland.** 1999. Molecular phylogeny and rearrangement of rRNA genes in Rickettsia species. Mol. Biol. Evol. **16:**987–995.
20. **Andersson, S. G. E., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. M. Alsmark, R. M. Podowski, A. K. Näslund, A.-S. Eriksson, H. H. Winkler, and C. G. Kurland.** 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature (London) **396:**133–140.
21. **Andersson, S. G. E., A. Zomorodipour, H. H. Winkler, and C. G. Kurland.** 1995. Unusual organization of the rRNA genes in *Rickettsia prowazekii*. J. Bacteriol. **177:**4171–4175.
22. **Attardi, C.** 1985. Animal mitochondrial DNA: an extreme example of genetic economy. Int. Rev. Cytol. **93:**93–145.
23. **Baker, A., and C. J. Leaver.** 1985. Isolation and sequence analysis of a cDNA encoding the ADP/ATP translocator of *Zea mays* L. Nucleic Acids Res. **13:**5857–5867.
24. **Baker, K. P., and G. Schatz.** 1991. Mitochondrial proteins essential for viability mediate protein transport into yeast mitochondria. Nature (London) **349:**205–208.
25. **Battini, R., S. Ferrari, L. Kaczmarek, B. Calabretta, S-T. Chen, and R. Baserga.** 1987. Molecular cloning of a cDNA for a human ADP/ATP carrier which is growth regulated. J. Biol. Chem. **262:**4355–4359.
26. **Benner, S. A., M. A. Cohen, G. H. Gonne, D. B. Berkowitz, and K. P. Johnsson.** 1993. Reading the palimpset: contemporary biochemical data and the RNA world, p. 27–70. *In* R. F. Gesteland and J. F. Atkins (ed.), The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. **Berg, O. G., and C. G. Kurland.** 2000. Why mitochondrial genes are most often found in nuclei. Mol. Biol. Evol. **17:**951–961.
28. **Bergstrom, C. T., and J. Pritchard.** 1998. Germline bottlenecks and evolutionary maintenance of mitochondrial genomes. Genetics **149:**2135–2146.
29. **Beron, W., C. Alvarez-Dominguez, L. Mayorga, and P. D. Stahl.** 1995. Membrane trafficking along the phagocytic pathway. Trends Cell Biol. **5:** 100–104.
30. **Biagini, G. A., A. J. Hayes, M. T. E. Suller, C. Winters, B. J. Finlay, and D. Lloyd.** 1997. Hydrogenosomes of *Metopus contortus* physiologically resemble mitocondria. Microbiology **143:**1623–1629.
31. **Bittner-Eddy, P., A. F. Monroy, and R. Bramble.** 1994. Expression of mitochondrial genes in the germinating conidea of *Neurospora crassa*. J. Mol. Biol. **235:**881–897.
32. **Bonen, L., R. S. Cunningham, M. W. Gray, and W. F. Doolittle.** 1977. Wheat embryo mitochondrial 18S ribosomal RNA: evidence for its prokaryotic nature. Nucleic Acids Res. **4:**663–671.
33. **Bronjik, T. H. C., R. Durand, M. van der Giezen, J. C. Gottschal, R. A. Prins, and M. Fevre.** 1996. scsB, a cDNA encoding the hydrogenosomal protein b-succinyl-coA synthetase from the anaerobic fungus *Neocallimastix frontalis*. Mol. Gen. Genet. **253:**315–323.
34. **Brown, J. R., and W. F. Doolittle.** 1997. Archaea and the prokaryote-to-eukaryote transition. Microbiol. Mol. Biol. Rev. **61:**456–502.
35. **Brynnel, E. U., C. G. Kurland, N. A. Moran, and S. G. E. Andersson.** 1998. Evolutionary rates for tuf genes in endosymbionts of aphids. Mol. Biol. Evol. **15:**574–582.
36. **Bui, E. T. N., P. J. Bradley, and P. J. Johnson.** 1996. A common evolutionary origin for mitochondria and hydrogenosomes. Proc. Natl. Acad. Sci. USA **93:**9651–9656.
37. **Bulmer, M.** 1987. Coevolution of the genetic code and transfer RNA abundance. Nature (London) **352:**728–730.
38. **Campbell, C. L., and P. E. Thorsness.** 1998. Escape of mitochondrial DNA to the nucleus in yme1 yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments. J. Cell Sci. **111:**2455–2464.
39. **Reference deleted.**
40. **Castresana, J., M. Luebben, M. Sarraste and D. G. Higgins.** 1994. Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen. EMBO J. **13:**2516–2525.
41. **Castresana, J., and M. Saraste.** 1995. Evolution of energetic metabolism: the respiration-early hypothesis. Trends Biochem. **20:**443–448.
42. **Cavalier-Smith, T.** 1987. The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies, p. 55–71. *In* J. L. Lee and J. F. Frederick (ed), Endocytobiology III. New York Academy of Sciences, New York, N.Y.
43. **Chao, L.** 1990. Fitness of RNA virus decreased by Muller's ratchet. Nature (London) **348:**454–455.
44. **Clark, C. G., and A. J. Roger.** 1995. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. Proc. Natl. Acad. Sci. USA **92:**6518–6521.
45. **Covello, P. S., and M. W. Gray.** 1992. Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (cox2) in soybean: Evidence for RNA-mediated gene transfer. EMBO J. **11:**3815–3820.
46. **Dandekar, T., S. Schuster, B. Snel, M. Huynen, and P. Bork.** 1999. Pathway alignment: application to the comparative analysis of glycolytic enzymes. Biochem. J. **343:**115–124.
47. **De Peer, Y. V., J.-M. Neefs, P. De Rijk, P. De Vos, and R. De Wachter.** 1994. About the order of divergence of the major bacterial taxa during evolution. Syst. Appl. Bacteriol. **17:**32–38.
48. **Doolittle, R. F.** 1997. Microbial genomes opened up. Nature (London) **392:** 339–342.

49. **Doolittle, R. F., D.-F. Feng, S. Tsang, G. Cho, and E. Little.** 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271:**470–477.
50. **Doolittle, W. F.** 1998. A paradigm gets shifty. Nature (London) **392:**15–16.
51. **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.
52. **Doolittle, W. F.** 1999. Phylogenetic classification and the universal tree. Science **284:**2124–2129.
53. **Edlind, T. D., J. Li, G. S. Visversvara, M. H. Vodkin, G. L. McLaughlin, and R. Katiyar.** 1996. Phylogenetic analysis of β-tubulin sequences from amitochondrial protozoa. Mol. Phylogenet. Evol. **5:**359–367.
54. **Embley, T. M., B. J. Finlay, P. L. Dyal, R. P. Hirt, M. Wilkinson, and A. G. Williams.** 1995. Multiple origin of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. Proc. R. Soc. Lond. B **262:**87–93.
55. **Embley, T. M., and R. P. Hirt.** 1998. Early branching eukaryotes? Curr. Opin. Genet. Dev. **8:**624–629.
56. **Embley, T. M., D. A. Horner, and R. P. Hirt.** 1997. Anaerobic eukaryote evolution: hydrogenosomes as biochemically modified mitochondria. Trends Ecol. Evol. **12:**437–441.
57. **Embley, T. M., and W. Martin.** 1998. A hydrogen-producing mitochondrion. Nature (London) **396:**517–519.
58. **Eremeeva, M. E., V. Roux, and D. Raoult.** 1993. Determination of genome size and restriction pattern polymorphism of *Rickettsia prowazekii* and *Rickettsia typhi* by pulsed field gel electrophoresis. FEMS Microbiol. Lett. **112:** 105–112.
59. **Felsenstein, J.** 1974. The evolutionary advantage of recombination. Genetics **78:**737–756.
60. **Fenchel, T., and B. J. Finlay.** 1990. Oxygen toxicity, respiration and behavioural responses to oxygen in free-living anaerobic ciliates. J. Gen. Microbiol. **136:**1953–1959.
61. **Fenchel, T., and B. J. Finlay.** 1991. The biology of free-living anaerobic ciliates. Europ. J. Protistol. **26:**201–215.
62. **Fenchel, T., and B. J. Finlay.** 1995. Ecology and evolution in anoxic worlds. Oxford University Press, Oxford, U.K.
63. **Feng, D.-F., G. Cho, and R. F. Doolittle.** 1997. Determining divergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. USA **94:**13028–13033.
64. **Figge, R. M., M. Schubert, H. Brinkmann, and R. Cerff.** 1999. Glyceraldehyde-3-phosphate dehydrogenase gene diversity in eubacteria and eukaryotes: evidence for intra-and inter-kingdom gene transfer. Mol. Biol. Evol. **16:**429–440.
65. **Finlay, B. J., and T. Fenchel.** 1989. Hydrogenosomes in anaerobic protozoa resemble mitochondria. FEMS Microbiol. Lett. **65:**311–314.
66. **Fox, T. D., L. S. Folley, J. J. Mulero, T. W. McMullin, P. E. Thorsness, L. O. Hedin, and M. C. Costanzo.** 1990. Analysis of yeast mitochondrial genes. Methods Enzymol. **194:**149–165.
67. **Garcia-del Portillo, F., and B. B. Finlay.** 1995. The varied lifestyles of intracellular pathogens within eukaryotic vacuolar compartments. Trends Microbiol. **3:**373–380.
68. **Germot, A., H. Phillipe, and H. Le Guyader.** 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:**14614–14618.
69. **Germot, A., H. Philippe, and H. Le Guyader.** 1997. Evidence for loss of mitochondria in microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. Mol. Biochem. Parasitol. **87:**159–168.
70. **Gilson, P. R., U. G. Maier, and G. I. McFadden.** 1997. Size isn't everything: lessons in genetic miniaturisation from nucleomorphs. Curr. Opin. Genet. Dev. **7:**800–806.
71. **Gilson, P. R., and G. I. McFadden.** 1996. The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns. Proc. Natl. Acad. Sci. USA **93:**7737–7742.
72. **Gilson, P. R., and G. I. McFadden.** 1997. Good things in small packages: the tiny genomes of chlorarachniophyte endosymbionts. Bioessays **19:**167–173.
73. **Gogarten, J. P., L. Olendzenski, E. Hilario, C. Simon, and K. E. Holsinger.** 1996. Dating the cenancester of organisms. Science **274:**1750–1751.
74. **Graham, D. E., R. Overbeek, G. J. Olsen, and C. R. Woese.** 2000. An archaeal genomic signature. Proc. Natl. Acad. Sci. USA **97:**3304–3308.
75. **Gray, M. W.** 1992. The endosymbiont hypothesis revisited. Int. Rev. Cytol. **141:**233–357.
76. **Gray, M. W.** 1998. Rickettsia, typhus and the mitochondrial connection. Nature (London) **396:**109–110.
77. **Gray, M. W., G. Burger, and B. F. Lang.** 1999. Mitochondrial evolution. Nature (London) **283:**1476–1481.
78. **Gray, M. W., R. Cedergren, Y. Abel, and D. Sankoff.** 1989. On the evolutionary origin of plant mitochondrion and its genome. Proc. Natl. Acad. Sci. USA **86:**2267–2271.
79. **Gray, M. W., and W. F. Doolittle.** 1982. Has the endosymbiont hypothesis been proven? Microbiol. Rev. **46:**1–42.
80. **Gray, M. W., B. F. Lang, R. Cedergren, G. B. Golding, C. Lemieux, D. Sankoff, M. Turmel, N. Brossard, E. Delege, T. G. Littlejohn, I. Plante, P. Rioux, D. Saint-Louis, Y. Zhu, and G. Burger.** 1998. Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res. **26:**865–878.
81. **Gray, M. W., and D. F. Spencer.** 1996. Organellar evolution, p. 109–126. *In* D. M. Roberts, P. M. Sharp, G. Alderson, and M. Collins (ed.), Evolution of microbial life. Cambridge University Press, Cambridge, U.K.
82. **Hanekamp, T., and P. E. Thorsness.** 1996. Inactivation of YME2, an integral inner mitochondrial membrane protein, causes increased escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. Mol. Biol. Cell **16:**2764–2771.
83. **Hanson, M. R., and O. Folkerts.** 1992. Structure and function of the higher plant mitochondrial genome, p. 129–173. *In* D. R. Wolstenholme and K. W. Jeon (ed.), Mitochondrial genomes. Academic Press, Inc., San Diego, Calif.
84. **Hatch, T. P., E. Al-Houssaine, and J. A. Silvermann.** 1982. Adenine nucleotide and lysine transport in *Chlamydia psittaci*. J. Bacteriol. **150:**662–670.
85. **Henze, K., A. Badr, M. Wettern, R. Cerff, and W. Martin.** 1995. A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc. Natl. Acad. Sci. USA **92:**9122–9126.
86. **Hirt, R. P., J. M. Logsdon, B. Healy, M. W. Dorby, W. F. Doolittle, and T. M. Embley.** 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.
87. **Hodges, P. E., A. H. Z. McKee, B. P. Davis, W. E. B. Pasyne, and J. I. Garrels.** 1999. Yeast Protein Database (YPD): a model for the organization and presentation of genome-wide functional data. Nucleic Acids Res. **27:** 69–73.
88. **Holland, H. D.** 1994. Early proterozoic atmospheric change, p. 237–244. *In* S. Bengtson (ed.), Early life on earth. Columbia University Press, New York, N.Y.
89. **Horner, D. S., R. P. Hirt, and T. M. Embley.** 1999. A single eubacterial origin of eukaryotic pyruvate:ferredoxin oxidoreductase genes: implications for the evolution of anaerobic eukaryotes. Mol. Biol. Evol. **16:**1280–1291.
90. **Horner, D. S., R. P. Hirt, S. Kilvington, D. Lloyd, and T. M. Embley.** 1996. Molecular data suggest an early acquisition of the mitochondrion endosymbiont. Proc. R. Soc. London Ser. B **263:**1053–1059.
91. **Hrdy, I., and M. Mueller.** 1995. Primary structure and eubacterial relationships of the pyruvate:ferredoxin oxidoreductase of the amitochondriate eukaryote *Trichomonas vaginalis*. J. Mol. Evol. **41:**388–396.
92. **Jain, R., M. C. Rivera, and J. A. Lake.** 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. USA **96:** 3801–3806.
93. **Johnson, P. J., C. J. Lahti, and P. J. Bradley.** 1993. Biogenesis of the hydrogenosome in the anaerobic protist *Trichomonas vaginalis*. J. Parasitol. **79:**664–670.
94. **Kalman, S., W. Mitchell, R. Marathe, C. Lammel, L. Fan, R. W. Hyman, L. Olinger, L. Grimwood, R. W. Davis, and R. S. Stephens.** 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nat. Genet. **21:**385–389.
95. **Karlberg, O., B. Canbäck, C. G. Kurland, and S. G. E. Andersson.** 2000. The dual origin of the yeast mitochondrial proteome. Yeast Comp. Functional Genomics. **17:**170–187.
96. **Keeling, P. J., and W. F. Doolittle.** 1996. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. Mol. Biol. Evol. **13:**1297–1305.
97. **Keeling, P. J., and W. F. Doolittle.** 1997. Evidence that eukaryotic triosephosphate isomerase is of alpha-proteobacterial origin. Proc. Natl. Acad. Sci. USA **94:**1270–1275.
98. **Keeling, P. J., and W. F. Doolittle.** 1997. Widespread and ancient distribution of a noncanonical genetic code in diplomonads. Mol. Biol. Evol. **14:** 895–901.
99. **Kengen, S. W. M., F. A. M. de Bok, N.-D. van Loo, C. Dijkema, A. J. M. Stams, and W. M. de Vos.** 1998. Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. J. Biol. Chem. **269:**17537–17541.
100. **Kerscher, L., and D. Oesterhelt.** 1982. Pyruvate:ferredoxin oxidoreductase—new findings on an ancient enzyme. Trends Biochem. Sci. **7:**371–374.
101. **Kimura, M.** 1983. The neutral theory of molecular evolution. Cambridge University Press, New York, N.Y.
102. **Kletzin, A., and M. W. W. Adams.** 1996. Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*. J. Bacteriol. **178:**248–257.
103. **Klingenberg, M.** 1989. Molecular aspects of adenine nucleotide carrier from mitochondria. Arch. Biochem. Biophys. **270:**1–14.
104. **Kobayashi, K., and A. Yoshimoto.** 1982. Studies on yeast sulfite reductase IV. Structure and steady state kinetics. Biochim. Biophys. Acta **705:**348–356.
105. **Kolarov, J., N. Kolarova, and N. Nelson.** 1990. A third ADP/ATP translocator gene in yeast. J. Biol. Chem. **265:**12711–12716.
106. **Kuan, J., and M. H. Saier.** 1993. The mitochondrial carrier family of transport proteins: structural, functional, and evolutionary relationships.

Crit. Rev. Biochem. Mol. Biol. **28:**209–233.
107. **Kündig, C., H. Hennecke, and M. Göttfert.** 1993. Correlated physical and genetic map of the *Bradyrhizobium japonicum* 110 genome. J. Bacteriol. **175:**613–622.
108. **Kurland, C. G.** 1992. Evolution of mitochondrial genomes and the genetic code. Bioessays **14:**709–714.
109. **Lake, J. A.** 1994. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. Proc. Natl. Acad. Sci. USA **91:**1455–1459.
110. **Lambert, J. D., and N. A. Moran.** 1998. Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **95:** 4458–4462.
111. **Lang, B. F., G. Burger, C. J. O'Kelly, R. Cedergren, G. B. Golding, C. Lemieux, D. Sankoff, M. Turmel, and M. W. Gray.** 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature (London) **387:**493–497.
112. **Lawson, J. E., and M. G. Douglas.** 1988. Separate genes encode functionally equivalent ADP/ATP carrier proteins in *Saccharomyces cerevisiae*: isolation and analysis of AAC2. J. Biol. Chem. **263:**14812–14818.
113. **Li, X., S. Kaul, S. Rounsley, T. P. Shea, M.-I. Benito, et al.** 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature **402:** 761–768.
114. **Lindmark, D. G., and M. Mueller.** 1973. Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Trichomonas foetus* and its role in pyruvate metabolism. J. Biol. Chem. **248:**7724–7728.
115. **Lopez-Garcia, P., and D. Moreira.** 1999. Metabolic symbiosis at the origin of eukaryotes. Trends Biotechnol. **24:**88–93.
116. **Ludwig, W., J. Neumaier, N. Klugbauer, E. Brockmann, C. Roller, S. Jilg, K. Reetz, I. Schachtner, A. Ludvigsen, M. Bachleitner, U. Fischer, and K. H. Schleifer.** 1993. Phylogenetic relationships of Bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes. Antonie van Leeuwenhoek J. Microbiol. Serol, **64:**285–305.
117. **Lynch, M.** 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol. Biol. Evol. **13:** 209–220.
118. **Lynch, M.** 1997. Mutation accumulation in nuclear, organelle and prokaryotic transfer RNA genes. Mol. Biol. Evol. **14:**914–925.
119. **Lynch, M., R. Burger, D. Butcher, and W. Gabriel.** 1993. The mutational meltdown in asexual populations. J. Hered. **84:**339–344.
120. **Maidak, B. L., G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and C. R. Woese.** 1996. The Ribosomal Database Project (RDP). Nucleic Acids Res. **24:**82–85.
121. **Margulis, L.** 1970. Origin of eukaryotic cells. Yale University Press, New Haven, Conn.
122. **Markos, A., A. Miretsky, and M. Mueller.** 1993. A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote *Trichomonas vaginalis*. J. Mol. Evol. **37:**631–643.
123. **Martin, W., and R. Herrmann.** 1998. Gene transfer from organelles to the nucleus: How much, what happens, and why? Plant Physiol. **118:**9–17.
124. **Martin, W., and M. Mueller.** 1998. The hydrogen hypothesis for the first eukaryote. Nature (London) **392:**37–41.
125. **Martin, W., and C. Schnarrenberger.** 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr. Genet. **32:**1–18.
126. **Martin, W., B. Stoebe, V. Goremykin, S. Hansmann, M. Hasegawa, and K. V. Kowallik.** 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature (London) **393:**162–165.
127. **Marvin-Sikkema, F. D., A. J. M. Driessen, J. C. Gottschal, and R. A. Prins.** 1994. Metabolic energy generation in the hydrogenosomes of the anaerobic fungus *Neocallimastix*: evidence for a functional relationship with mitochondria. Mycol. Res. **98:**205–212.
128. **Marvin-Sikkema, F. D., M. N. Kraak, M. Veenhuis, J. C. Gottschal, and R. A. Prins.** 1993. The hydrogenosomal enzyme hydrogenase from the anaerobic fungus *Neocallimastiz* sp. L2 is recognized by antibodies, directed against the C-terminal microbody protein targeting signal SKL. Eur. J. Cell Biol. **61:**86–91.
129. **Meziane, A. E., S. K. Lehtinen, N. Hance, L. G. Nijtmans, D. Dunbar, I. J. Holt, and H. T. Jacobs.** 1998. A tRNA suppressor mutation in human mitochondria. Nat. Genet. **18:**350–353.
130. **Moran, N. A.** 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93:**2873–2878.
131. **Moreira, D., and P. Lopez-Garcia.** 1998. Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. J. Mol. Evol. **47:**517–530.
132. **Mueller, M.** 1993. The hydrogenosome. J. Gen. Microbiol. **139:**2879–2889.
133. **Mueller, M.** 1997. Evolutionary origins of trichomonad hydrogenosomes. Parasitol. Today **13:**166–167.
134. **Mueller, M.** 1997. What are the microsporidia? Parasitol. Today **13:**455–456.
135. **Mueller, M.** 1998. Enzymes and compartmentation of core energy metabolism of anaerobic protists: a special case in eukaryotic evolution, p. 109–127. *In* G. H. Coombs, K. Vickerman, M. A. Sleigh, and A. Warren (ed.), Evolutionary relationships among protozoa. Kluwer Academic Publishers, London, U.K.
136. **Mueller, M., and W. Martin.** 1999. The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. Bioessays **21:**377–381.
137. **Muller, J. J.** 1964. The relation of recombination to mutational advance. Mutat. Res. **1:**2–9.
138. **Neupert, W.** 1997. Protein import into mitochondria. Annu. Rev. Biochem. **66:**683–717.
139. **Nowitzki, U., A. Flechner, J. Kellerman, M. Hasegawa, C. Schnarrenberger, and W. Martin.** 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.
140. **Nugent, J. M., and J. D. Palmer.** 1991. RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution. Cell **66:**473–481.
141. **Olsen, G. J., C. R. Woese, and R. Overbeek.** 1994. The winds of evolutionary change: breathing new life into microbiology. J. Bacteriol. **176:**1–6.
142. **Oparin, A. I.** 1957. The origin of life on the earth, 3rd ed., trans. by A. Synge. Oliver and Boyd, Edinburgh, U.K.
143. **Orston, L. N., E. L. Neidle, and J. E. Houghton.** 1990. Gene rearrangements, a force for evolutionary change; DNA sequence arrangements, a source of genetic constancy, p. 325–335. *In* K. Drlica and M. Riley (ed.), The bacterial chromosome. American Society for Microbiology, Washington, D.C.
144. **Palmer, J.** 1997. Organelle genomes: going, going, gone! Science **275:**790–791.
145. **Palmer, J. D.** 1997. The mitochondrion that time forgot. Nature (London) **387:**454–455.
146. **Palmer, J. D., and L. A. Herbon.** 1988. Plant mitochondrial DNA evolves rapidly in structure but slowly in sequence. J. Mol. Evol. **28:**87–94.
147. **Pang, H., and H. H. Winkler.** 1993. Copy numbers of the 16S rRNA gene in *Rickettsia prowazekii*. J. Bacteriol. **175:**3893–3896.
148. **Peyretaillade, E., C. Biderre, P. Peyret, F. Duffieuxx, G. Metenier, M. Gouy, B. Michot, and C. Vivares.** 1998. Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. Nucleic Acids Res. **26:**3513–3520.
149. **Peyretaillade, E., V. Broussolle, P. Peyret, G. Metenier, M. Gouy, and C. P. Vivares.** 1998. Microsporidia, amitochondrial protists, possess a 70-kDa heat shock protein gene of mitochondrial evolutionary origin. Mol. Biol. Evol. **15:**683–689.
150. **Ribeiro, S., and G. B. Golding.** 1998. The mosaic nature of the eukaryotic nucleus. Mol. Biol. Evol. **15:**779–788.
151. **Rivera, M. C., R. Jain, J. E. Moore, and J. A. Lake.** 1998. Genomic evidence for two functionally distinct gene classes. Proc. Natl. Acad. Sci. USA **95:** 6239–6244.
152. **Roger, A. J., C. G. Clark, and W. F. Doolittle.** 1996. A possible mitochondrial gene in the early-branching amitochondrial protist *Trichomonas vaginalis*. Proc. Natl. Acad. Sci. USA **93:**14618–14622.
153. **Roger, A. J., S. G. Svard, J. Tovar, C. G. Clark, M. W. Smith, F. D. Gillin, and M. L. Sogin.** 1998. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc. Natl. Acad. Sci. USA **95:**229–234.
154. **Rosenthal, B., M. Zhiming, D. Caplivski, S. Ghosh, H. Deka Vega, T. Graf, and J. Samuelsson.** 1997. Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite *Entamoeba histolytica*. J. Bacteriol. **179:**3736–3745.
155. **Roux, V., and D. Raoult.** 1993. Genotypic identification and phylogenetic analysis of the spotted fever group rickettsiae by pulsed-field gel electrophoresis. J. Bacteriol. **175:**4895–4904.
156. **Rozario, C., L. Morin, A. J. Roger, M. W. Smith, and M. Mueller.** 1996. Primary structure and phylogenetic relationships of glyceraldehyde-3-phosphate dehydrogenase gene of free-living and parasitic diplomonad flagellates. J. Eukaryot. Microbiol. **43:**330–340.
157. **Rubey, W. W.** 1951. Geological history of seawater: an attempt to state the problem. Geol. Soc. Am. Bull. **62:**1111–1148.
158. **Runswick, M. J., A. Philippdes, G. Laura, and J. R. Walker.** 1994. Extension of the mitochondrial transporter super-family: sequences of five members from the nematode worm, *Caenorhabditis elegans*. DNA Seq. **4:**281–291.
159. **Runswick, M. J., S. J. Powell, P. Nyren, and J. R. Walker.** 1987. Sequence of the bovine mitochondrial carrier protein: structural relationship to ADP/ATP translocase and the brown fat mitochondria uncoupling protein. EMBO J. **6:**1367–1373.
160. **Santos, M. A. S., V. M. Perreau, and M. F. Tuite.** 1996. Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*. EMBO J. **18:**5060–5066.
161. **Schaefer, G., M. Engelhard, and V. Mueller.** 1999. Bioenergetics of the

archaea. Microbiol. Mol. Biol. Rev. **63:**570–620.
162. **Schatz, G.** 1996. The protein import system of mitochondria. J. Biol. Chem. **271:**31763–31766.
163. **Schatz, G., and B. Dobberstein.** 1996. Common principles of protein translation across membranes. Science **271:**1519–1526.
164. **Selig, M., K. B. Xavier, H. Santos, and P. Schonheit.** 1997. Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium Thermotoga. Arch. Microbiol. **167:**217–232.
165. **Sichertz-Ponten, T., C. G. Kurland, and S. G. E. Andersson.** 1998. A phylogenetic analysis of the cytochrome b and cytochrome c oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. Biochim. Biophys. Acta **1365:**545–551.
166. **Siebers, B., and R. Hensel.** 1993. Glucose catabolism of the hyperthermophilic archaeum *Thermoproteus tenax*. FEMS Microbiol. Lett. **111:**1–8.
167. **Sinai, A. P., and K. A. Joiner.** 1997. Safe haven: the cell biology of nonfusogenic pathogen vacuoles. Annu. Rev. Microbiol. **51:**415–462.
168. **Smith, M. W., D.-F. Feng, and R. F. Doolittle.** 1992. Evolution by acquisition: the case for horizontal gene transfers. Trends Ecol. Evol. **17:**489–493.
169. **Sogin, M. L.** 1997. History assignment: when was the mitochondrion founded? Curr. Opin. Genet. Dev. **7:**792–799.
170. **Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. L. Marathe, W. P. Mithcell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis.** 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science **282:**754–759.
171. **Syvänen, A.-C., H. Amiri, A. Jamal, S. G. E. Andersson, and C. G. Kurland.** 1996. A chimeric disposition of the elongation factor genes in *Rickettsia prowazekii*. J. Bacteriol. **178:**6192–6199.
172. **Tetley, L., S. M. Brown, V. McDonald, and G.H. Coombs.** 1998. Ultrastructural analysis of the sporozoite of *Cryptosporidium parvum*. Microbiology **144:**3249–3255.
173. **Thomas, D., H. Cherest, and Y. Surdin-Kerjan.** 1989. Elements involved in *S*-adenosylmethionine-mediated regulation of the *Saccharomyces cerevisiae* MET25 gene. Mol. Cell. Biol. **9:**3292–3298.
174. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
175. **Thorsness, P. E., and T. D. Fox.** 1990. Escape of DNA from mitochondria to nucleus in *Saccharomyces cerevisiae*. Nature (London) **346:**376–379.
176. **Thorsness, P. E., and T. D. Fox.** 1993. Nuclear mutations in *Saccharomyces cerevisiae* that affect the escape of DNA from mitochondria to the nucleus. Genetics **134:**21–28.
177. **Thorsness, P. E., and E. Weber.** 1996. Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. Int. Rev. Cytol. **165:**207–234.
178. **Thorsness, P. E., K. H. White, and T. D. Fox.** 1993. Inactivation of *YME1*, a gene coding a member of the SEC18, CDC48 family of putative ATPases, causes increased escape of DNA from mitochondria in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **13:**5418–5426.
179. **Townson, S. M., J. A. Upcroft, and P. Upcroft.** 1996. Characterisation and purification of pyruvate:ferredoxin oxidoreductase from *Giardia duodenalis*. Mol. Biochem. Parasitol. **79:**183–193.
180. **Unseld, M., J. R. Marienfeld, P. Brandt, and A. Brennicke.** 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat. Genet. **15:**57–61.
181. **van der Giezen, M., K. A. Sjollema, R. R. Artz, W. Alkema, and R. A. Prins.** 1997. Hydrogenosomes in the anaerobic fungus *Neocallimastix frontalis* have a double membrane but lack an associated organelle genome. FEBS Lett. **408:**147–150.
182. **van der Geizen, M., K. B. Rechinger, I. Svendsen, R. Durand, R. P. Hirt, T. M. Embley, and R. A. Prins.** 1997. A mitochondrial-like targeting signal on the hydrogenosomal malic enzyme from the anaerobic fungus *Neocallimastix frontalis*: support for the hypothesis that hydrogenosomes are modified mitochondria. Mol. Microbiol. **23:**11–21.
183. **van Hoek, A. H. A. M., T. A. van Allen, V. S. I. Sprakel, J. H. P. Hackstein, and G. D. Vogels.** 1998. Evolution of anaerobic ciliates from the gastrointestinal tract: phylogenetic analysis of the ribosomal repeat from *Nyctotherus ovalis* and its relatives. Mol. Biol. Evol. **15:**1195–1206.
184. **van Hoek, A. H. A. M., T. A. van Allen, V. S. I. Sprakel, J. A. M. Leunissen, T. Brigge, G. D. Vogels, and J. H. P. Hackstein.** 2000. Multiple acquisition of methanogenic archaeal symbionts by anaerobic ciliates. Mol. Biol. Evol. **17:**251–258.
185. **Vellai, T., K. Takács, and G. Vida.** 1998. A new aspect to the origin and evolution of eukaryotes. J. Mol. Evol. **46:**499–507.
186. **Viale, A., and A. K. Arakaki.** 1994. The chaperone connection to the origins of the eukaryotic organelles. FEBS Lett. **341:**146–151.
187. **Viale, A. M., A. K. Arakaki, F. C. Soncini, and R. G. Ferreyra.** 1994. Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. Int. J. Syst. Bacteriol. **44:**527–533.
188. **Vignais, P. V.** 1976. Molecular and physiological aspects of adenine nucleotide transport in mitochondria. Biochim. Biophys. Acta **456:**1–38.
189. **Walker, J. E., and M. J. Runswick.** 1993. The mitochondrial transport protein superfamily. Bioenerg. Biomem. **25:**435–446.
190. **Weber, E. R., T. Hanekamp, and P. E. Thorsness.** 1996. Biochemical and functional analysis of the *YME1* gene product, an ATP- and zinc-dependent mitochondrial protease from *S. cerevisiae*. Mol. Biol. Cell **7:**307–317.
191. **Weisburg, W. G., M. E. Dobson, J. E. Samuel, G. A. Dasch, L. P. Mallavia, O. A. L. Mandelco, J. E. Sechrestr, E. Weiss, and C. R. Woese.** 1989. Phylogenetic diversity of the rickettsiae. J. Bacteriol. **171:**4202–4206.
192. **Wernegreen, J. J., and N. A. Moran.** 1999. Evidence for genetic drift in endosymbionts (Buchnera): analyses of protein-coding genes. Mol. Biol. Evol. **16:**83–97.
193. **Whatley, J. M., P. John, and F. R. Whatley.** 1979. From extracellular to intracellular: the establishment of mitochondria and chloroplasts. Proc. R. Soc. Lond. B **204:**165–187.
194. **Whitehouse, D. B., J. Tompkins, J. U. Lovegrove, D. A. Hopkinson, and W. O. McMillan.** 1998. A phylogenetic approach to the identification of phosphoglucomutase genes. Mol. Biol. Evol. **15:**456–462.
195. **Williamson, L. R., G. V. Plano, H. H. Winkler, D. C. Krause, and D. O. Wood.** 1989. Nucleotide sequence of the *Rickettsia prowazekii* ATP/ADP translocase-encoding gene. Gene **80:**269–278.
196. **Winkler, H. H.** 1976. Rickettsial permeability: an ADP-ATP transport system. J. Biol. Chem. **251:**389–396.
197. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
198. **Woese, C. R.** 1998. The universal ancestor. Proc. Natl. Acad. Sci. USA **95:**6854–6859.
199. **Woese, C. R., G. J. Olsen, M. Ibba, and D. Soell.** 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. **64:**202–236.
200. **Wolf, Y. I., L. Aravind, N. V. Grishin, and E. V. Koonin.** 1999. Evolution of aminoacyl-tRNA synthetases-analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Res. **9:**689–710.
201. **Wolf, Y. I., L. Aravind, and E. V. Koonin.** 1999. Rickettsiae and chlamydiae: evidence of horizontal gene transfer and gene exchange. Trends Genet. **15:**173–175.
202. **Wolfe, K. H., W-H. Li, and P. M. Sharp.** 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84:**9054–9058.
203. **Wolfe, K. H., and D. C. Shields.** 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature (London) **387:**708–713.
204. **Wolstenholme, D. R.** 1992. Animal mitochondrial-DNA: structure and evolution. Int. Rev. Cytol. **141:**173–216.
205. **Wolstenholme, D. R., and K. W. Jeon (ed.).** 1992. Mitochondrial genomes. Academic Press, Inc., San Diego, Calif.
206. **Yang, D., Y. Oyaizu, H. Oyaizu, G. J. Olsen, and C. R. Woese.** 1985. Mitochondrial origins. Proc. Natl. Acad. Sci. USA **82:**4443–4447.
207. **Yu, J.-P., J. Lapado, and W. B. Whitman.** 1994. Pathway of glycogen metabolism in *Methanococcus maripaludis*. J. Bacteriol. **176:**325–332.