Theoretical Population Biology **54**, 91–104 (1998)
Article No. TP981376

TPB

MINIREVIEW

Life's Third Domain (*Archaea*): An Established Fact or an Endangered Paradigm?

A New Proposal for Classification of Organisms Based on Protein Sequences and Cell Structure

Radhey S. Gupta

Department of Biochemistry, McMaster University, Hamilton, Ontario L8N 3Z5, Canada
E-mail: gupta@fhs.mcmaster.ca

Received April 13, 1998

The three-domain proposal of Woese *et al.* (*Proc. Natl. Acad. Sci. USA* 87, 4576 (1990)) divides all living organisms into three primary groups or domains named *Archaea* (or archaebacteria), *Bacteria* (or eubacteria), and *Eucarya* (or eukaryotes), with *Eucarya* being relatives (or descendants) of Archaea. Although this proposal is currently widely accepted, sequence features and phylogenies derived from many highly conserved proteins are inconsistent with it and point to a close and specific relationship between archaebacteria and gram-positive bacteria, whereas gram-negative bacteria are indicated to be phylogenetically distinct. A closer relationship of archaebacteria to gram-positive bacteria in comparison to gram-negative bacteria is generally seen for the majority of the available gene/protein sequences. To account for these results, and the fact that both archaebacteria and gram-positive bacteria are prokaryotes surrounded by a single cell membrane, I propose that the primary division within prokaryotes is between Monoderm prokaryotes (surrounded by a single membrane) and Diderm prokaryotes (i.e., all true gram-negative bacteria containing both an inner cytoplasmic membrane and an outer membrane). This proposal is consistent with both cell morphology and signature sequences in different proteins. Protein phylogenies and signature sequences also show that *all eukaryotic cells* have received significant gene contributions from both an archaebacterium and a gram-negative eubacterium. Thus, the hypothesis that archaebacteria and eukaryotes shared a common ancestor exclusive of eubacteria, or that the ancestral eukaryotic cell directly descended from an archaea, is erroneous. These results call into question the validity of the currently popular three-domain proposal and the assignment of a domain status to archaebacteria. A new classification of organisms consistent with phenotype and macromolecular sequence data is proposed.

Key Words: **Archaebacteria; gram-positive bacteria; gram-negative bacteria; prokaryote phylogeny; eukaryotic cell origin; cell structure.**

INTRODUCTION

How many fundamentally different forms of life exist on this planet and how they are evolutionarily related constitute some of the most challenging problems in biology. Based upon microscopic observations of cells, biologists have divided living organisms into two primary groups: prokaryotes and eukaryotes (Chatton, 1937; Murray,

0040-5809/98 $25.00

1968; Stanier *et al.* 1976; Margulis, 1993). Prokaryotes and eukaryotes differ from each other in many respects; however, the hallmark feature distinguishing the two types of cells is the membrane-bounded nucleus, and any organism lacking it is considered a prokaryote (Chatton, 1937; Murray, 1968). This view of the primary division of living organisms into prokaryotes and eukaryotes was challenged by Woese and co-workers based on the oligonucleotide sequence catalogs and phylogenies derived from 16S rRNA (see Woese, 1987, 1992). These studies revealed that one group of previously little studied prokaryotes, termed archaebacteria, were no more closely related to other bacteria (termed eubacteria) than to the eukaryotes. The distinctness of archaebacteria was also supported by phylogenies based on a number of proteins (viz., elongation factors EF-1/Tu and EF-2/G, RNA polymerase subunits II and III, and F- and V-type ATPases) (Iwabe *et al.*, 1989; Puhler *et al.*, 1989; Gogarten *et al.*, 1989), and their unique ether-linked membrane lipids (Kandler and Konig, 1993). This led to the proposal that archaebacteria (or *Archaea*) are totally distinct from other prokaryotes and constitute a fundamentally different form of life (i.e., life's third domain), in addition to eubacteria (*Bacteria*) and eukaryotes (*Eucarya*) (Woese *et al.*, 1990). The phylogenetic analysis of the above proteins also indicated that the eukaryotic homologs for these were more closely related to archaebacteria than to eubacteria and that the root of the universal tree lay between archaebacteria and eubacteria (Gogarten *et al.*, 1989; Iwabe *et al.*, 1989). This view of the relationship between the prokaryotic and eukaryotic organisms, commonly referred to as the archaebacterial or the three-domain tree of life, is now widely accepted and constitutes the current paradigm in the field (Woese *et al.*, 1990).

In the past 2-3 years, with the sequencing of complete genomes from a number of archaebacteria, eubacteria, and the eukaryote *Saccharomyces cerevisiae* (Bult *et al.*, 1996; Fleischmann *et al.*, 1995; Fraser *et al.*, 1995, see Tiger Microbial Database website for other references to completed genomes), many data have become available to examine critically the evolutionary relationships amongst various organisms. These data have thus far been interpreted to provide strong vindication of the view that the primary division within prokaryotes is between archaebacteria and eubacteria and that the eukaryotic cells have directly descended from an archaebacterial ancestor (Bult *et al.*, 1996; Belfort and Weiner, 1997; Dennis, 1997; Edgell and Doolittle, 1997; Olsen and Woese, 1997; Reeve *et al.*, 1997). As a consequence, the three-domain proposal has now assumed the aura of established fact. But does this view depict the true relationship among living organisms? Karl Popper (1968) in his classic treatise "The Logic of Scientific Discovery" has stressed that any valid scientific theory should make some predictions that could be empirically or experimentally tested. These predictions, if proven wrong or not supported by data, should lead to a rejection/reconsideration of the theory. The two main tenets of the three-domain proposal (Woese *et al.*, 1990) are: (i) archaebacteria (or Archaea) constitute a *monophyletic* domain and phylogenetically they are totally distinct from the rest of the prokaryotes (i.e., eubacteria); and (ii) archaebacteria and eukaryotic cells shared a common ancestor exclusive of any eubacteria. In other words, the eukaryotic *nuclear genome* (exclusive of organellar genes) has originated directly from an archaebacterial cell. Let us examine whether these tenets, which are the foundation of the three-domain hypothesis, are supported by the available data.

To understand the global relationship among living organisms, we must first understand the evolutionary relationship within the prokaryotes that predated the eukaryotes (Schopf, 1978). It should be emphasized that this question of evolutionary relationships between prokaryotes is completely independent of the question of how the eukaryotic species evolved from them. In the past, most phylogenies have attempted to explain the evolutionary relationship between various prokaryotic and eukaryotic species at the same time and have led to conflicting and confusing relationships (Woese, 1987; Gogarten *et al.*, 1989; Iwabe *et al.*, 1989; Puhler *et al.*, 1989; Rivera and Lake, 1992; Gupta and Singh, 1994; Golding and Gupta, 1995; Brown and Doolittle, 1995, 1997; Doolittle, 1995; Baldauf *et al.*, 1996; Forterre, 1997). However, once the evolutionary relationship between prokaryotic organisms is understood, the question of how the eukaryotic species are related to them can be meaningfully approached. To understand the evolutionary relationship between prokaryotic organisms, which spans >3.5 Ga, the sequences that are highly conserved (i.e., the best preserved molecular fossils) and that are present in most prokaryotes should be the molecules of choice. The molecules or sequences that are present in only particular groups of prokaryotes, or that show limited similarity between different groups of prokaryotes (i.e., where much of the evolutionary record has been eroded), will be of either very limited or no use in this regard.

EVOLUTIONARY RELATIONSHIPS AMONG PROKARYOTES

Since the discovery of the gram-staining reaction by Hans Christan Gram in 1884, prokaryotic organisms

have been divided into two primary groups, the gram-positive and gram-negative, based on the gram-stain retention characteristics of their cell walls (Stanier *et al*., 1976; Murray, 1986; Holt *et al*., 1994). Although gram-staining itself is not a reliable criterion, its basis lies in an important structural characteristic of the cells: the nature of the bounding layer (Stanier *et al*., 1976). All gram-positive bacteria are bounded by a single cell membrane and most of these contain a thick cell wall, which is responsible for retaining the gram stain. In contrast all "true" gram-negative bacteria contain only a thin peptidoglycan layer (i.e., cell wall) lying in between two different cell membranes (Stanier *et al*., 1976). I have used the latter criterion to define gram-positive and gram-negative bacteria in the present work.

If one examines various available gene/protein sequences from prokaryotes, particularly from a number of prokaryotic genomes that have been completely sequenced (see Tiger Microbial Database website), one finds that a large proportion of the genes or potentially coding sequences either are unique to only the particular group of prokaryotes or have homologs in only closely related species (Fraser *et al*., 1995; Fleischmann *et al*., 1995; Bult *et al*., 1996; Koonin *et al*., 1997; Olsen and Woese, 1997; Gupta, unpublished results). Many of the genes involved in DNA replication and transcription in archaebacterial genomes, for which no eubacterial homologs are found, fall into this category (Belfort and Weiner, 1997; Edgell and Doolittle, 1997; Olsen and Woese, 1997; Reeve *et al*., 1997). These genes, because of their limited distribution, are not useful in understanding the evolutionary relationships within prokaryotes. If one studies various other gene/protein sequences, the homologs for which are found in all main groups of prokaryotes, two common patterns are observed. First, for many genes/proteins, the majority of which are related to transcription and translation processes, e.g., rRNAs, EF-1/Tu, EF-2/G, RNA polymerase II and III subunits, most ribosomal proteins, and aminoacyl-tRNA synthetases, the sequence signatures and phylogenies generally indicate a grouping of archaebacteria distinct from the rest of the prokaryotes (Langer *et al*., 1995; Golding and Gupta, 1995; Brown and Doolittle, 1995, 1997; Hashimoto and Hasegawa, 1996; Dennis, 1997; Olsen and Woese, 1997). The elongation factor EF-1/Tu (Fig. 1a) and ribosomal protein L5 (Fig. 1b) provide two examples of proteins for which conserved indels that distinguish archaebacteria from various eubacteria are present in the homologs. Similar signature sequences are present in a number of other proteins (Gupta, 1998, and unpublished results). The genes/proteins showing this pattern are referred to as Group I sequences in this work.

The second type of recurring pattern is observed for a number of proteins primarily involved in metabolic functions. Such proteins include the 70-kDa heat shock molecular chaperone protein (Hsp70), glutamine synthetase I (Gln I), asparginyl-tRNA synthetase, diaminopimelate epimerase, glutamate dehydrogenase (GDH), and inosine 5′-monophosphate dehydrogenase (IMPDH). In these cases, a close evolutionary relationship is observed between various archaebacteria and gram-positive bacteria, whereas the gram-negative bacteria are indicated to form a distinct clade (Gupta, 1998; Karlin *et al*., 1995; Tiboni *et al*., 1993; Benachenhou-Lahfa *et al*., 1993; Brown *et al*., 1994; Gupta and Singh, 1994; Golding and Gupta, 1995; Gupta *et al*., 1997, and unpublished results). A close and specific evolutionary relationship between archaebacteria and gram-positive bacteria and the distinctness of gram-negative bacteria are most readily seen by the signature sequences present in several of these proteins (Fig. 2). In contrast to phylogenetic trees, where the inferred relationship among species is dependent upon a number of variables, including sequence alignment, regions of sequences used in analysis, difference in evolutionary rates between species, number of species included, and phylogenetic methods employed, and often not resolved (Brown and Doolittle, 1997; Gupta, 1998), when a conserved indel (i.e., insertion or deletion) of defined length and sequence (referred to as a signature sequence) is present at precisely the same position in homologs from different species, the simplest and most parsimonious explanation for this observation is that the indel was introduced only once during the course of evolution and then passed on to all descendants (Rivera and Lake, 1992; Gupta, 1998). Thus, based on the presence or absence of a signature sequence, the species containing or lacking the signature can be divided into two unambiguous groups, which bear a specific evolutionary relationship to each other (Gupta, 1998).

Figure 2 shows a prominent signature sequence in the Hsp70 protein, where a conserved insert of 23–25 amino acids (a.a.) is found in all gram-negative bacteria but not in any archaebacteria or gram-positive bacteria. As seen, the insert is present in all of the gram-negative bacteria, including members of the α, β, γ, δ, and ε-subdivisions of the proteobacteria, chlamydia and spirochetes, cytophaga, flavobacteria, cyanobacteria, green nonsulfur bacteria, deinococcus, and thermus divisions. A conserved indel of 26 a.a. common to archaebacteria and gram-positive bacteria is also present in the Gln I sequences (Fig. 3A) (Brown *et al*., 1994). Likewise, in asparginyl-tRNA synthetase (Gupta, 1997) and diaminopimelate epimerase (Fig. 3B), conserved indels that distinguish archaebacteria and gram-positive bacteria from gram-negative bacteria

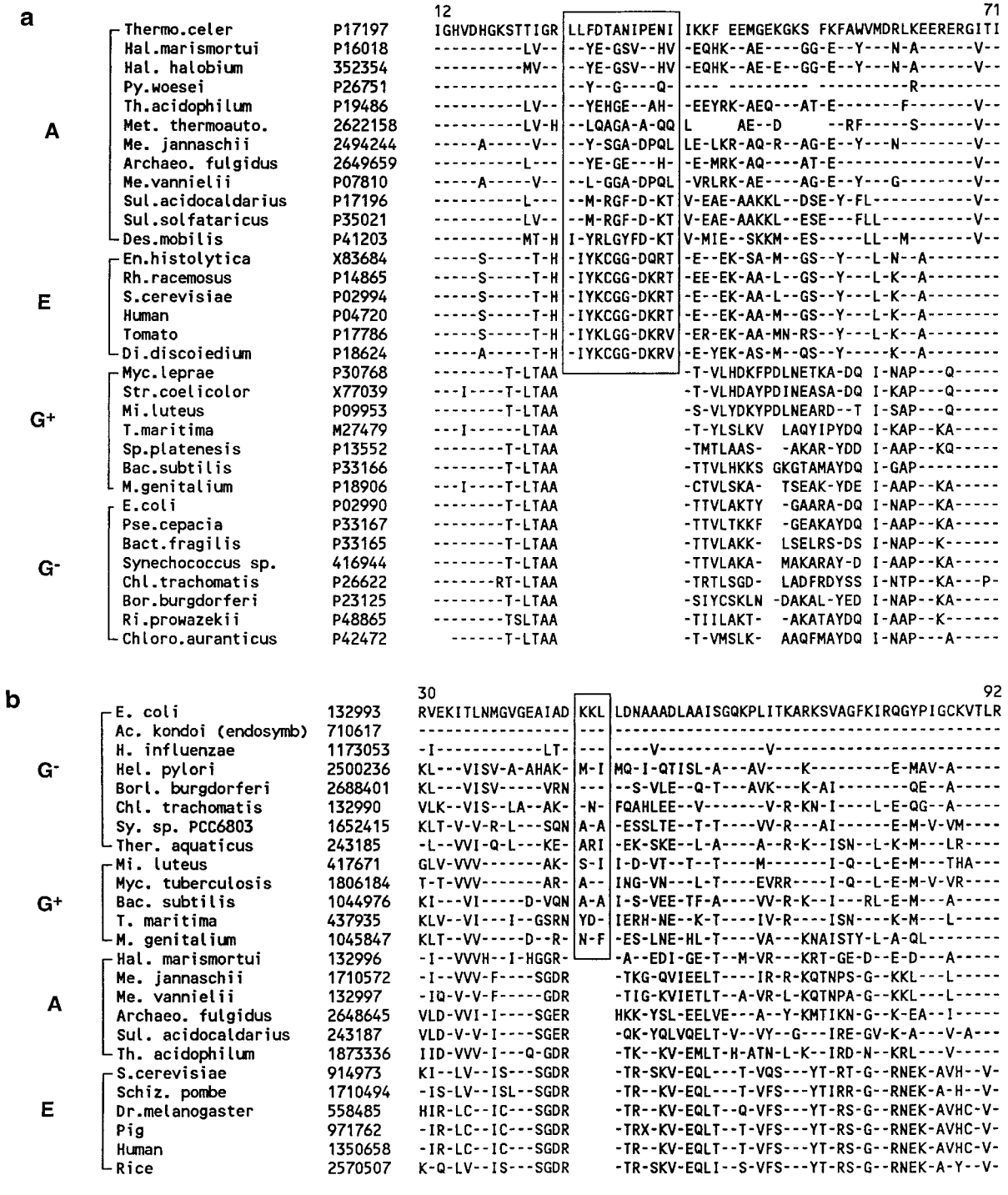

FIG. 1. Signature sequences in (a) EF-1/Tu and (b) ribosomal L5 proteins showing the distinctness of archaebacterial species (**A**) from eubacteria (**G**$^+$ and **G**$^-$). The abbreviations **A**, **G**$^+$, **G**$^-$, and **E** refer to archaebacteria, gram-positive bacteria, gram-negative bacteria, and eukaryotes, respectively. The shared indels that distinguish archaebacteria from eubacteria are boxed. The dashes (-) in all alignments denote identity with the amino acid in the top line. The accession numbers of sequences are in the second column. The abbreviations in the species names are Ac., *Acrythosiphon*; Archaeo., *Archaeoglobus*; Bac., *Bacillus*; Bact., *Bacteriodes*; Bor., *Borrelia*; Chl., *Chlamydia*; Chloro., *Chloroflexus*; Des., *Desulfurococcus*; E., *Escherichia*; H., *Haemophilus*; Hal., *Halobacterium*; Hel., *Helicobacteria*; M., *Mycoplasma*; Me., *Methanococcus*; Met. thermauto., *Methanobacterium thermoautotrophicum*; Mi., *Micrococcus*; Myc., *Mycobacterium*; Pse., *Pseudomonas*; Py., *Pyrococcus*; Ri., *Rickettsia*; Sp., *Spirulina*; Str., *Streptomyces*; Sul., *Sulfolobus*; Sy., *Synechocystis*; Syn., *Synechococcus*; T., *Thermotoga*; Th., *Thermoplasma*; Ther., *Thermus*; Thermo., *Thermococcus*.

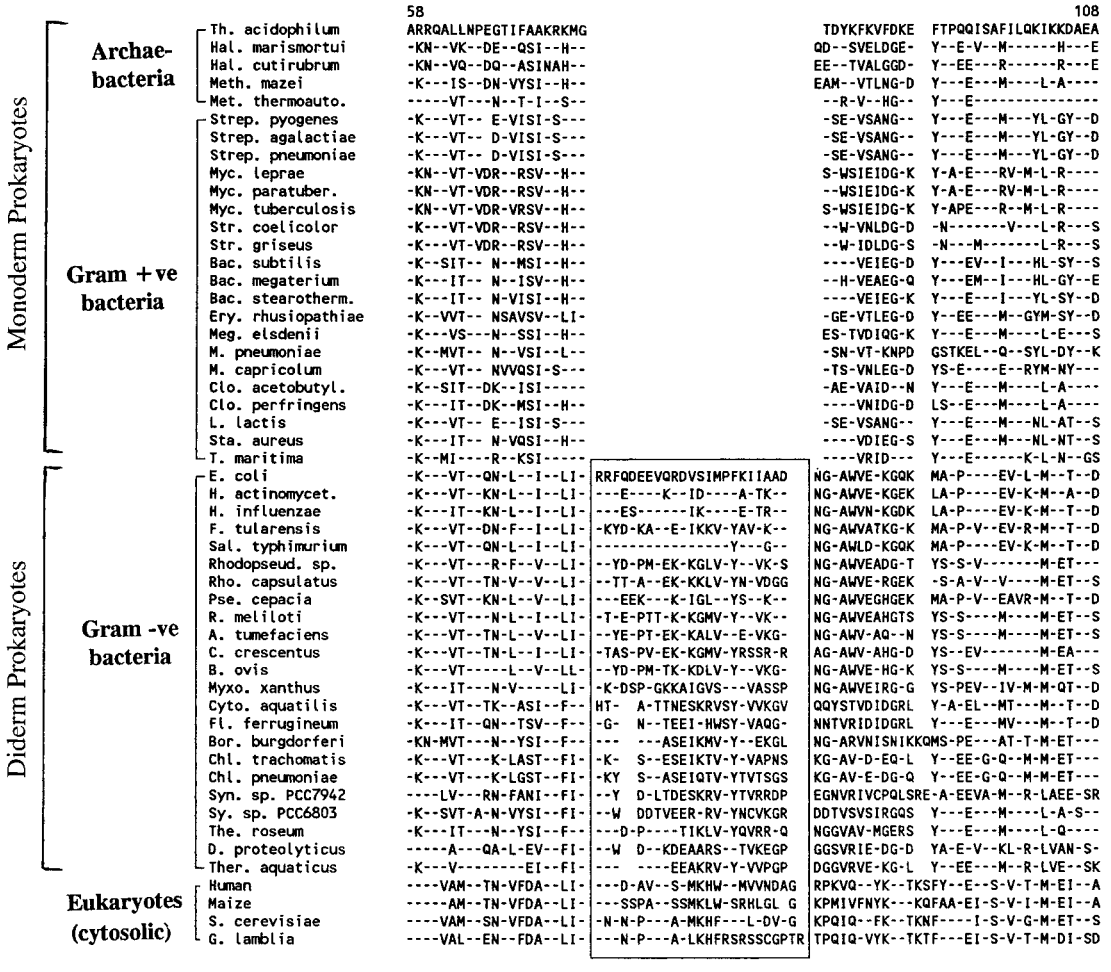

FIG. 2. Signature sequence in the Hsp70 proteins showing a specific relationship between archaebacteria (**A**) and gram-positive bacteria (**G**$^+$) (both monoderm prokaryotes) and pointing to the distinctness of gram-negative bacteria (**G**$^-$) (diderm prokaryotes). The boxed region shows the large insert present in all gram-negative bacteria but absent from all monoderm prokaryotes. For eukaryotic species, sequences for only representative cytosolic homologs are shown. The presence of this shared insert in the eukaryotic cytosolic homologs (**E**) provides evidence that they are derived from gram-negative bacteria rather than archaebacteria. A detailed description and analysis of eukaryotic organellar and cytosolic homologs and various signatures that distinguish between them can be found in earlier publications (Gupta and Singh, 1994; Gupta and Golding, 1996). New abbreviations in species names are A., *Agrobacterium*; B., *Brucella*; Bac. stearotherm., *Bac. stearothermophilus*; C., *Caulobater*; Clo. acetobutyl., *Clostridium acetobutylicum*; Cyto., *Cytophaga*; D., *Deinococcus*; Di., *Dictyostelium*; Dr., *Drosophila*; En., *Entameoba*; Ery., *Erysiopelothrix*; F., *Francisella*; Fl., *Flavobacteria*; H. actinomycet., *H. actinomycetemcomitans*; L., *Lactococcus*; Meg., *Megasphaera*; Meth., *Methanosarcina*; Myxo., *Myxococcus*; R., *Rhizobium*; Rh., *Rhizomucor*; Rho., *Rhodobacter*; Rhodopseud., *Rhodopseudomonas*; Sal., *Salmonella*; S., *Saccharomyces*; Schiz., *Schizosaccharomyces*; Sta., *Staphylococcus*; Strep., *Streptococcus*; The., *Thermomicrobium.*

are observed. The signature sequences in Hsp70 (Fig. 2) and Gln I (Fig. 3A) also provide evidence that the species *T. maritima*, which is currently grouped with gram-negative bacteria (Woese, 1987; Woese *et al.*, 1990; Holt *et al.*, 1994) is in fact a gram-positive bacteria. This inference is consistent with the observation that *T. maritima* contains only a single cell membrane (Cavalier-Smith, 1992), which is my defining criterion for gram-positive bacteria.

In the past, supporters of the three-domain proposal have ignored or dismissed the above results showing a close relationship of archaebacteria to gram-positive bacteria as possible anomalies (Baldauf *et al.*, 1996; Olsen and Woese, 1996; Brown and Doolittle, 1997). But is this so? In the reported phylogenies for some of the Group I genes/proteins, viz., 16S rRNA, EF-1/Tu, EF-2/G, RNA polymerase, aminoacyl-tRNA synthetases and various ribosomal proteins, which form the basis for defining archaebacteria as a unique domain, *T. maritima*, which is now known to be a gram-positive bacterium, shows the closest relationship to archaebacteria (Woese, 1987; Baldauf *et al.*, 1996; Hashimoto and Hasegawa, 1996; Brown and Doolittle, 1995, 1997). In a recent

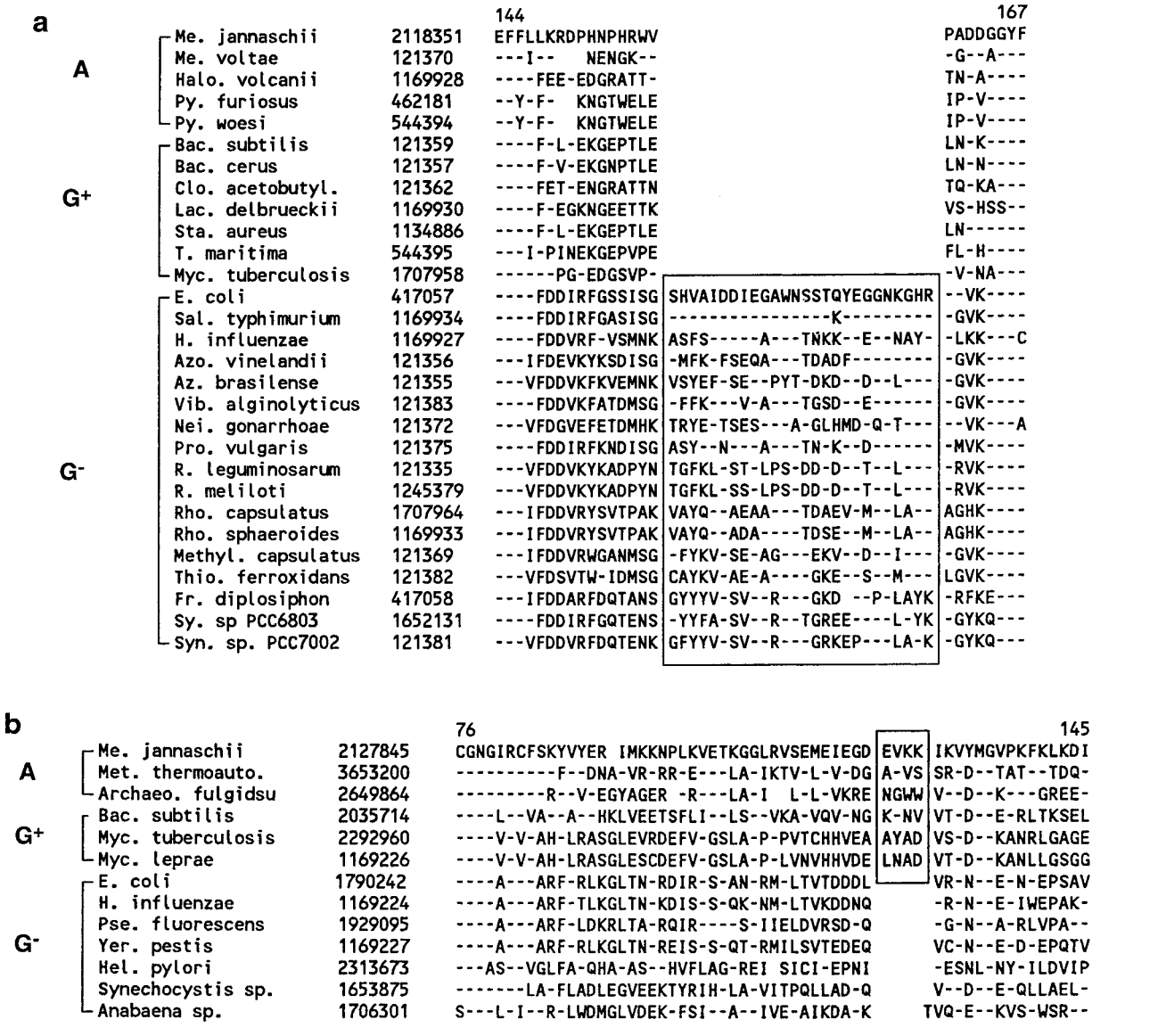

FIG. 3. Signature sequences in (a) glutamine synthetase I and (b) diaminopimelate epimerase showing the relatedness of archaebacteria (**A**) to gram-positive bacteria ($\mathbf{G}^+$) and the distinctness of gram-negative bacteria ($\mathbf{G}^-$). Additional abbreviations are Az., *Azospirillum*; Azo., *Azotobacter*; Fr., *Fremyella*; G., *Giardia*; Halo., *Haloferax*; Lac., *Lactobacillus*; Methyl., *Methylococcus*; Nei., *Neisseria*; Pro., *Proteus*; Thio., *Thiobacillus*; Vib., *Vibrio*; Yer., *Yersinia*.

review, Brown and Doolittle (1997) have reported phylogenies based on 66 protein sequences for which sequence information was available from archaebacteria, eubacteria, and eukaryotes. As pointed out earlier and as seen from the results of this study, it is confounding to consider the evolutionary relationships between prokaryotes and eukaryotes without a good understanding of the relationship within prokaryotes. However, if one examines the phylogenetic trees reported in this review and asks the question which group of prokaryotes are the closest relatives of archaebacteria, then for more than two-third of the genes studied, *T. maritima* or another gram-positive bacterium was found to be the closest relative of archaebacteria. These genes include Hsp70, gyrase B, photolyase, EF-1/Tu, EF-2/G, isoleucyl-tRNA synthetase, tryptophanyl-tRNA synthetase, tyrosyl-tRNA synthetase, ribosomal proteins L2, L11, L14, L15, L22, L23, L30, S5, S10, S15, S19, enolase, acetyl-CoA synthetase, citrate synthase, *trpA*, *trpB*, *trpC*, *trpD*, *hisC*, *hisD*, *hisF*, *hisH*, IMPDH, FGAM synthetase, glutamyl-tRNA reductase, 5-aminolevulinic acid dehydratase, SecY, FeMn-superoxide dismutase, Hsp60/Tcp-1, GDH, GlnI,

aspartate aminotransferase, histidinol-phosphate aminotransferase, and argininosuccinate synthetase (Brown and Doolittle, 1997). For a number of other genes (viz., RNA polymerase A, DNA polymerase B, *hisF*) no homologs from eubacteria or gram-positive bacteria were included. Thus a close relationship of archaebacteria to gram-positive bacteria is not restricted to a few genes but generally observed for most genes, including the various Group I genes/proteins sequences (Brown and Doolittle, 1997). Further, for several of the genes, for which sequence information was available from adequate number of sources (viz., Hsp70, glutamate dehydrogenase, Gln I, *hisC*, *hisF*, *hisH*, *trpB*, *trpD*), archaebacteria do not form a monophyletic group but exhibit paraphyletic branching with gram-positive bacteria (Benachenhou-Lahfa *et al.*, 1993; Tiboni *et al.*, 1993; Brown *et al.*, 1994; Gupta and Singh, 1994; Golding and Gupta, 1995; Brown and Doolittle, 1997; Gupta *et al.*, 1997).

What is the significance of the observed close relationship between archaebacteria and gram-positive bacteria? The answer to this question becomes strikingly clear upon consideration of cell structures of the prokaryotes (Stanier *et al.*, 1976; Tipper and Wright, 1979). Based upon their cell structures, the prokaryotic organisms can be divided into two major groups—those bounded by a single membrane (termed monoderms) and those containing an inner and outer membranes (termed diderms) that define the periplasmic compartment (Fig. 4). All archaebacteria and gram-positive bacteria belong to the first group. Some species which lack a cell wall (e.g., *Mycoplasma*, *Thermoplasma*) or show gram-negative staining due to other unusual characteristics (e.g., *Megasphaera*, *Thermotoga*) are also bounded by a single membrane. The signature sequences and phylogenies based on Hsp70 and other highly conserved proteins (viz., Figs. 2 and 3) cluster all prokaryotes with a single membrane (i.e., monoderm prokaryotes) in one taxon and all bacteria with two different membranes (i.e., diderm prokaryotes) in a separate taxon. Thus a close relationship between archaebacteria and gram-positive bacteria as seen for Hsp70 and a large number of other genes is not anomalous but is in accord with the most fundamental difference in cellular organization in prokaryotes (Fig. 4).

In addition to the presence of an outer membrane defining a periplasmic compartment, the diderm prokaryotes differ from monoderms in several other respects, including cell wall thickness and general response to environment. Tipper and Wright (1979) claim: "The Gram-negative bacteria has a fundamentally different strategy toward the external environment than the Gram-positive cell. In the Gram-negative cells a membrane is present, external to the peptidoglycan layer, that acts as

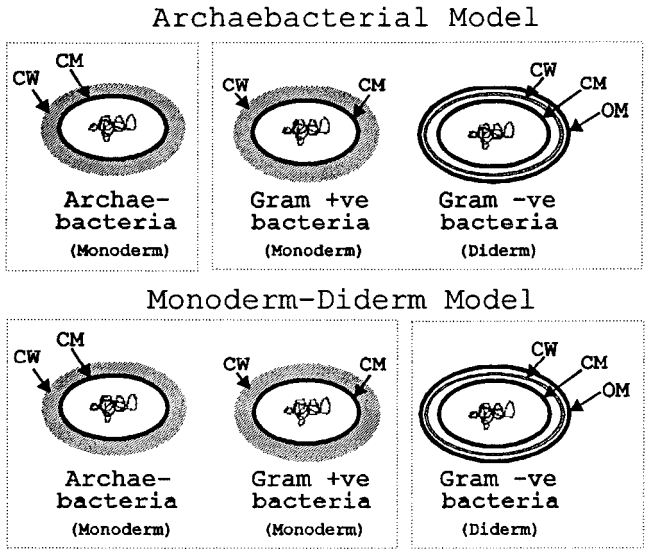

FIG. 4. The structural characteristics of the prokaryotes and a comparison of the two models for the primary division within prokaryotes. The top panel shows the primary division according to the currently popular archaebacterial or three-domain model (Woese *et al.*, 1990). The three-domain model does not recognize diderm prokaryotes (i.e., gram-negative bacteria) as a distinct taxon. According to this proposal one group of monoderm prokaryotes form one primary group, whereas in the other group different monoderm (gram-positive bacteria) and diderm prokaryotes show polyphyletic branching (Woese, 1992; Olsen and Woese, 1993; Woese, 1987). The lower panel shows the monoderm–diderm model proposed in this work. This model, which is supported by signature sequences in a number of highly conserved proteins, unites all monoderm prokaryotes (i.e., archaebacteria and gram-positive bacteria) into one taxon and all true gram-negative bacteria (i.e., diderm prokaryotes) in a separate taxon. Abbreviations: CM, cytoplasmic membrane; OM, outer membrane; CW, cell wall.

a permeability barrier between the external environment and the cytoplasmic membrane. It is an essential component of all Gram-negative cells and apparently cannot be dispensed with, even under laboratory conditions." It is unlikely that such fundamental and complex characteristics which involved formation of a new compartment evolved more than once independently or that they could be acquired/transferred nonspecifically. These observations are inconsistent with the three-domain proposal, which does not recognize diderm prokaryotes as a separate taxon and in which one group of monoderm prokaryotes (viz., archaebacteria) form one domain, whereas in the other domain a paraphyletic branching is observed for various monoderm and diderm phyla (Olsen and Woese, 1997; Woese, 1987; Woese *et al.*, 1990). Thus in contrast to the statements by Woese (1992) that "Morphology is not the guiding principle," and Olsen and Woese (1993) that "Bacterial morphologies, physiologies, and most other whole cell properties are too simple or too volatile to be reliable phylogenetic indicators," the

molecular sequence data are in fact in good agreement with, and strongly support, the most obvious and important structural distinction seen within prokaryotes.

EVOLUTIONARY RELATIONSHIP BETWEEN PROKARYOTES AND EUKARYOTES

The second main premise of the three-domain model, that archaebacteria and eukaryotes share a common ancestor exclusive of eubacteria, was proposed to account for the fact that for a number of protein sequences, EF-1/Tu, EF-2/G, RNA polymerase II and III subunits, and F- and V-type ATPases, originally studied, the eukaryotic homologs exhibited greater similarity to archaebacteria than to eubacteria (Gogarten *et al*., 1989; Iwabe *et al*., 1989; Puhler *et al*., 1989). Recent sequencing and analysis of archaebacterial genomes indeed provide strong evidence that for the vast majority of genes involved in information transfer processes such as replication, transcription, and translation, the eukaryotic homologs are closely related to archaebacteria (Langer *et al*., 1995; Belfort and Weiner, 1997; Dennis, 1997; Edgell and Doolittle, 1997; Klenk *et al*., 1997; Koonin *et al*., 1997; Olsen and Woese, 1997; Reeve *et al*., 1997; Smith *et al*., 1997). The proteins EF-1/Tu and the ribosomal proteins L5 provide examples of proteins where all archaebacterial and eukaryotic homologs share common sequence signatures not found in any eubacteria (Fig. 1). For several genes involved in DNA replication and transcription no eubacterial homologs have been found (Edged and Doolittle, 1997; Olsen and Woese, 1997; Reeve *et al*., 1997). Thus, it is indisputable that archaebacteria contributed to the eukaryotic nuclear genome, particularly to the information transfer machinery. However, the key question in this area is whether *all* of the eukaryotic nuclear–cyotosolic genome (i.e., exclusive of organelles) derived from archaebacteria or whether other groups of prokaryotes also made significant contributions to it? The origin of the eukaryotic cell hinges critically on the resolution of this question.

The question of establishment of any eubacterial contribution to the eukaryotic nuclear–cytosolic genome is far more difficult than connecting archaebacteria and eukaryotes. The main difficulty lies in the fact that in contrast to archaebacteria, which have contributed only to the nuclear genome, two classes of eukaryotic cell organelle genomes, mitochondria and plastids, derive from eubacteria in later endosymbiotic acquisitions (Gray, 1992; Margulis, 1993). Most organellar genes were later transposed to the nucleus. Thus, eukaryotes often have multiple homologs of proteins with sequence similarity to eubacteria. For most sequences in the databases, information that clearly distinguishes nuclear–cytosolic from organellar homologs is lacking. The presence of multiple genes inside eukaryotes also raises the possibility that the genes for the nuclear–cytosolic proteins may be derived from the organellar genes by horizontal transfer followed by divergence. Thus establishment that a nuclear–cytosolic gene is of eubacterial origin and was acquired prior to and independent of the organellar genomes has proven difficult. However, in recent years, the enlarged sequence database and extensive characterization of many eukaryotic protein families have helped resolve this problem. In the well-studied Hsp70 family of proteins, different eukaryotic nuclear–cytosolic homologs are clearly distinguished from the organellar by numerous signature sequences (Gupta and Singh, 1994; Gupta and Golding, 1996). The hydrogenosomal Hsp70 from *Trichomonas vaginalis* also contains various distinctive sequence characteristics unique to mitochondrial homologs and branches with this group, providing evidence that it has originated from the same source as mitochondria (Bui *et al*., 1996; Germot *et al*., 1996; Martin and Muller, 1998). However, the nuclear–cytosolic homologs of Hsp70 from all species (including *Giardia lamblia*) are distinct from the mitochondrial and hydrogenosomal homologs based on various sequence signature and their branching in the phylogenetic trees (Gupta and Singh, 1994; Gupta *et al*., 1994; Gupta and Golding, 1996; Bui *et al*., 1996; Germot *et al*., 1996; Gupta, 1998). Yet, based upon the presence of the large insert in their N-terminal quadrants (see Fig. 2) and their branching within gram-negative bacteria (Gupta and Singh, 1994), these homologs are also derived from gram-negative eubacteria. These results provide evidence that the nuclear cytosolic homologs of Hsp70 have originated from gram-negative bacteria independent of mitochondria and hydrogenosomes and they are inconsistent with the recently proposed hydrogenosome hypothesis for the origin of eukaryotic cell (Martin and Muller, 1998).

Two additional examples of proteins for which the eukaryotic nuclear–cytosolic homologs are derived from eubacteria and not archaebacteria include Hsp90 and adenylosuccinate synthetase (ASS) (Fig. 5). Hsp90 homologs from gram-negative bacteria and eukaryotes, including *G. lamblia* (unpublished), lack an indel present in both low G + C and high G + C gram-positive bacteria (Fig. 5A). No homolog for Hsp90 has yet been found in any archaebacterium, including the three completed archaebacterial genomes (Bult *et al*., 1996; Klenk *et al*., 1997; Smith *et al*., 1997). The eukaryotic nuclear–cytosolic genes for this protein thus evolved from gram-negative eubacteria rather than from archaebacteria. The

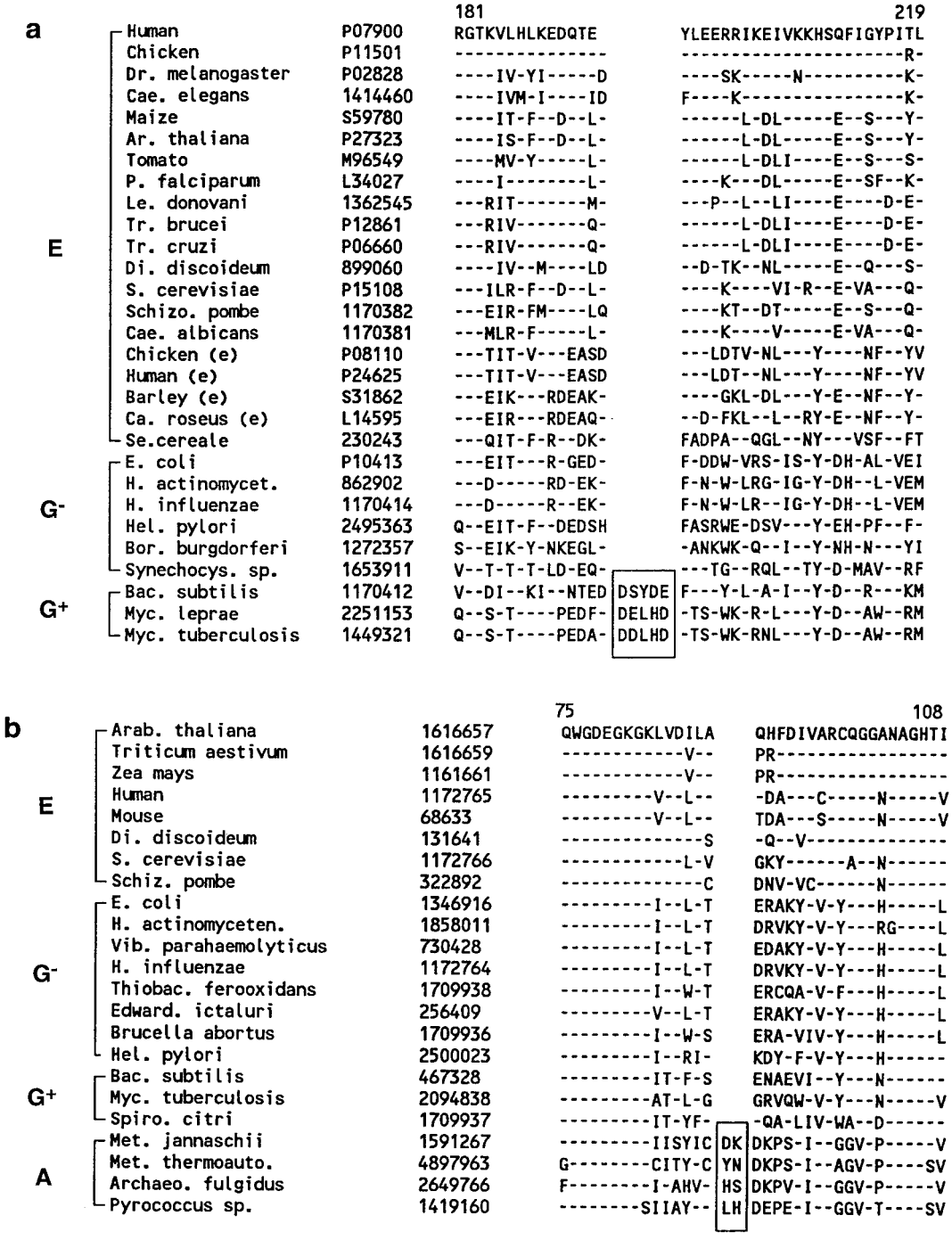

FIG. 5. Signature sequence in the (a) Hsp90 and (b) adenylosuccinate synthetase showing the relatedness of eukaryotic cytosolic homologs (**E**) to eubacteria ($\mathbf{G}^+$ and $\mathbf{G}^-$) rather than archaebacteria (**A**). For Hsp90, no archaebacterial homolog has been identified in the three completed archaebacterial genomes (Bult *et al.*, 1996; Klenk *et al.*, 1997; Smith *et al.*, 1997). Additional abbreviations: Arab., *Arabidopsis*; Ca., *Catharanthus*; Cae., *Caenorhabditis*; Di., *Dictyostelium*; Le., *Leischmania*, P., *Plasmodium*; Se., *Secale*; Tr., *Trypanosome*.

ASS homologs from various archaebacteria contain a 2-a.a. insert absent in eubacterial and eukaryotic sequences (Fig. 5B), indicating that the eukaryotic homologs of this protein are also of eubacterial and not archaebacterial origin. Another striking characteristic of eukaryotic cells not explained by archaebacterial origin is their membrane lipid composition (Zillig *et al.*, 1989; Lake and Rivera, 1994; Gupta and Golding, 1996). All eukaryotic cell membranes contain ester-linked fatty acid lipids like those in eubacteria rather than the ether-linked lipids that define archaebacteria (Woese *et al.*, 1990; Kandler and Konig, 1993). Thus the eukaryotic cell membranes are of eubacterial rather than archaebacterial origin. Therefore, the premise that archaebacteria and the ancestral eukaryotic cell shared a common ancestor exclusive of all eubacteria is incorrect.

The global phylogenies and signature sequences of various gene/protein sequences that have been examined show that the eukaryotic nuclear homologs for different genes exhibit greater similarity to either archaebacteria or gram-negative bacteria (Golding and Gupta, 1995; Gupta and Golding, 1996; Gupta, 1997, 1998; Karlin *et al.*, 1997). In both cases, the results are strongly supported statistically (Golding and Gupta, 1995). To explain these mutually discordant histories of eukaryotic nuclear genes, we have proposed that the ancestral eukaryotic cell arose by a unique fusion event involving an archaebacterium (most likely an eocyte, based on Rivera and Lake's (1992) data on EF-1 and EF-2 sequences) and a gram-negative bacterium (Fig. 6) (Gupta and Singh, 1994; Golding and Gupta, 1995; Gupta and Golding, 1996; Gupta, 1997). The chimeric origin of eukaryotic cells by fusion between an archaebacterium and a eubacterium was originally proposed by Zillig *et al.* (1989) but this view has not been favored in their subsequent publications (Klenk and Zillig, 1994; Langer *et al.*, 1995). At an early stage following fusion, an assortment or selection of genes from the two fusion partners occurred, during which most of the genes for information transfer such as replication, transcription, and translation (which probably function as a unit) were retained from the archaebacterial (monoderm) partner, whereas genes for other metabolic

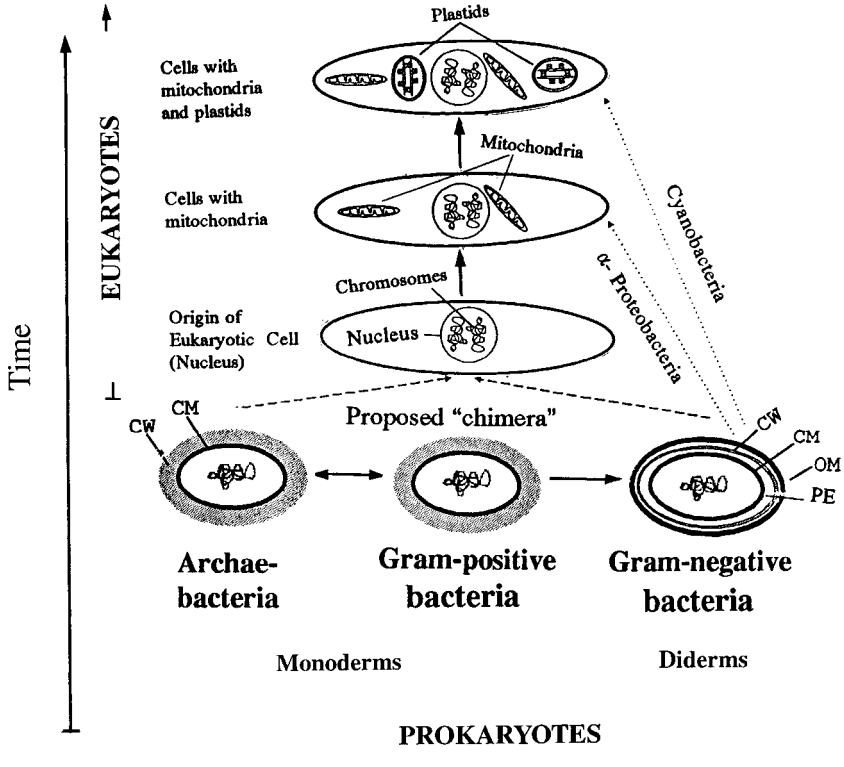

FIG. 6. Evolutionary relationships among living organisms as inferred from protein sequence data and morphology. The solid arrows identify taxa that evolved from each other in the directions shown by accumulation of mutations. The double-headed arrow between archaebacteria and gram-positive bacteria points to a paraphyletic relationship between these groups for a number of genes. The dashed lines indicate the first fusion between an archaebacterium and a gram-negative bacterium that gave rise to the ancestral eukaryotic cell (Gupta *et al.*, 1994; Gupta and Golding, 1996). The dotted lines denote subsequent symbiotic events that led to the acquisition of mitochondria and plastids (Margulis, 1970; Gray, 1992). Abbreviations: CM, cell membrane; CW, cell wall; OM, outer membrane, PE, periplasm.

functions such as membrane lipids, Hsp70, Hsp90, ASS, were kept from the gram-negative (diderm) bacterium. The ancestral eukaryotic cell is thus a chimera that contains characteristics from each of the prokaryotic parents (Zillig *et al*., 1989; Gupta and Singh, 1994; Golding and Gupta, 1995; Gupta and Golding, 1996; Gupta, 1997, 1998; Karlin *et al*., 1997). In contrast to the three-domain model, this chimeric model predicts that no eukaryotic cell, including amitochondriate and aplastidic cells, will ever be found that did not receive and retain contributions such as Hsp70 and Hsp90 genes, and membrane lipids, from a gram-negative eubacterium.

In addition to being consistent with most of the gene/protein phylogenies, the present chimeric model also provides a plausible explanation for the origin of the eukaryotic endomembrane system: the nucleus and endoplasmic reticulum (ER) (Gupta *et al*., 1994; Gupta, 1995; Gupta and Golding, 1996). Our earlier work indicates that these structures coevolved as a direct result of the primary fusion. The formation of these compartments was preceded or accompanied by duplication of genes for chaperone proteins (e.g., Hsp70, Hsp90) which are necessary for communication between the newly formed ER compartments and the rest of the cell (Gupta *et al*., 1994; Gupta, 1995; Gupta and Golding, 1996). The chimeric origin of the eukaryotes by a fusion and integration of the genomes of two very different prokaryotic organisms also explains the enormous structural differences seen between the prokaryotes and the eukaryotes (Mayr, 1990; Margulis, 1993), the absence of any intermediates in this transition, and the inferred similar times of divergence of eukaryotes from either archaebacteria or eubacteria (about 2 Ga ago) based on genetic distances in different proteins (Doolittle *et al*., 1996). These observations cast serious doubts concerning the origin of ancestral eukaryotic cell from an archaebacterium by normal evolutionary mechanisms (e.g., mutations and recombination) (Gupta, 1998).

THE CLASSIFICATION OF ORGANISMS: A NEW PROPOSAL

The questions should now be asked whether the three-domain proposal for the classification of organisms is justified and whether it is appropriate to ascribe the highest taxonomic status (domain) to archaebacteria? As indicated, neither of the two main premises of the three-domain model—(i) that archaebacteria are totally distinct from other prokaryotes and the primary division within prokaryotes is between archaebacteria and eubacteria and (ii) that the ancestral eukaryotic cell was a direct descendant of archaebacteria, which were cited as the main reasons for ascribing a domain status to archaebacteria (Woese *et al*., 1990)—is universally supported by the molecular data. Further, while the distinction between "Monoderm" and "Diderm" prokaryotes is supported by both morphological and molecular observations, the taxon "Archaea" is defined only by biochemical and sequence characteristics. Archaebacteria show *no unique morphological features* by which they could be distinguished from other monoderm prokaryotes, i.e., gram-positive eubacteria (Woese *et al*., 1990; Gupta, 1997). As pointed out by Murray (1986b), "It is not appropriate to separate kingdoms on any basis but a major, reasonably easily determined difference in organization." Thus, on the basis of the above I conclude that the assignment of domain status to *Archaea* is not justified.

To integrate the various structural and macromolecular sequence characteristics, I propose a new classification of organisms at the highest taxonomic levels which is consistent with all of the observations (Table 1). Since eukaryotic cells are structurally distinct from prokaryotes in the presence of nucleus, endomembrane system, etc. (Chatton, 1937; Murray, 1968; Cavalier-Smith, 1987; Margulis, 1993), and appear to have evolved by fusion of two vastly different prokaryotes, my proposal recognizes only two highest taxa: Domain *Procaryotae* and Domain *Eucaryotae* (Murray, 1968; Margulis, 1996). The domain *Procaryotae* is divided into two naturally defined, non-overlapping and unambiguous subdomains: Subdomain *Monodermata* [Greek monos–single, derma–skin] and subdomain *Didermata* [Greek dis–double]. The subdomain *Monodermata* is further divided into two subsubdomains: *Archaebacteria* and gram-positive bacteria. The subdomain *Didermata* contains all gram-negative bacteria containing in addition to the cytoplasmic membrane an outer membrane which defines the periplasmic compartment. The structural and molecular features which define and distinguish the various proposed taxa are summarized in Table 1.

Of the two proposed primary taxa within prokaryotes, *Monodermata* is indicated to be the ancient lineage. Evidence in support of this comes from several lines of studies: First, one could use the sequence data for the duplicated EF-Tu/1 and EF-G/2 proteins to root the prokaryotic tree. These studies indicate that the root lies in between archaebacteria and *T. maritima* (Baldauf *et al*., 1996; Hashimoto and Hasegawa, 1996; Gupta, 1998), both of which are monoderm prokaryotes. Second, a similar inference has been reached independently based on the homologous isoleucine–, leucine–, and valine–tRNA synthetase sequences (Brown and Doolittle, 1995). Third, in earlier studies based on comparison of sequences of Hsp70 and the prokaryotic MreB protein, which shows

TABLE 1

Taxonomic Summary

Domain[a] *Procaryotae* (*Prokarya*) (Gk. pref. *pro* before; Gk. n. *karyon* nut, kernel (nucleus); Single and multicellular organisms that existed before the evolution of the membrane-bounded nucleus; DNA (nucleoplasm) in cells not separated from cytoplasm by membrane; contain smaller (70S) ribosomes with 16S and 23S types of rRNA; distinctive signature sequences found in the rRNA (Woese, 1987), Hsp70 protein (Gupta and Singh, 1994), and glucose–fructose 6-phosphate transaminase (Gupta, 1998).

Subdomain Monodermata (Gk. adj. *monos* single (or one); Gk. n. *derma* skin (layer): Prokaryotic cells surrounded by only a single layer of the unit lipoprotein membrane; most members contain a thick cell wall; distinguished from other prokaryotes by the absence of the large insert in their Hsp70 proteins (Fig. 2).

Subsubdomain Archaebacteria:[b] Monoderm prokaryotic cells; membrane lipids predominantly isoprenoid glycerol diethers or diglycerol tetraethers; distinguished from other monoderm prokaryotes (gram-positive bacteria) by unique signature sequences present in the rRNA (Woese, 1987) and EF-1/Tu and ribosomal L5 proteins (Fig. 1). Phenotypically methanogens, thermoacidophiles, and halophiles.

Subsubdomain Gram-positive bacteria:[b] Monoderm prokaryotic cells; membrane lipids predominantly diacylglycerol diesters; most show positive gram-staining reaction correlated with lack of outer membrane and presence of a thick peptidoglycan layer containing muramic acid; distinguished from Archaebacteria by signature sequences in the rRNA (Woese, 1987) and EF-1/Tu and ribosomal L5 (Fig. 1) proteins. Vast range of phenotypes.

Subdomain Didermata (*gram-negative bacteria*[b]) (Gk. adj. *dis* double): Prokaryotic cells with an inner and an outer unit cell membrane enclosing a periplasmic compartment; distinguished from other prokaryotes by the presence of a 23- to 25-amino-acid insert in their Hsp70 proteins (Fig. 2).

Domain Eucaryotae (*Eukarya* = *Eucarya*) (Gk. adj. *eu* good or true): Organisms composed of cells containing one to many membrane-bounded nuclei, chromatin organized into 2–10^3 chromosomes (DNA + protein) per cell. Heterogenomic cells formed by the fusion and integration of the genomes of an archaebacterium and a gram-negative bacterium. Later fusions with gram-negative prokaryotes gave rise to mitochondria and plastids (Margulis, 1970; Gray, 1992). Nucleoplasm separated from cytoplasm by a well-defined pore-studded membrane; most contain larger (80S) ribosomes with 18S and 28S type of rRNA and a cytoskeleton made up of actin and tubulin. Distinctive sequence signatures present in the rRNA (Woese, 1987), Hsp70 protein (Gupta, 1997), and glucose–fructose 6-phosphate transaminase (Gupta, 1998).

[a] I use "domain" and "subdomain" terminology to indicate the relative status of these groups.

[b] The names archaebacteria, gram-positive bacteria, and gram-negative bacteria are retained here for historical and practical considerarations only. These terms do not imply the defining characteristics of these groups.

significant sequence and structural similarity to the N-terminal half of Hsp70, we have provided evidence that the large indel in the N-terminal quadrant of Hsp70 (Fig. 2) is an insert in the diderm prokaryotes rather than a deletion in the monoderms (Gupta and Singh, 1992; Gupta and Golding, 1993). This observation indicates that the monoderm prokaryotes lacking this insertion are ancestral and diderms are derived from them. Fourth, based on the earth's geological history, the earliest organisms are postulated to be anaerobic and thermophilic (Schopf, 1978; Margulis, 1993). The wide spread occurrence of these characteristics in the *Monodermata* is in accordance with its ancestral nature (Pace, 1991). Last, the signature sequences in various proteins also provide evidence that the diderm prokaryotes are specifically related to, and have evolved from, the gram-positive bacteria (Gupta, 1997, 1998). The proposed relationships between prokaryotic and eukaryotic organisms are depicted in Fig. 6.

While the present proposal unifies both the phylogenetic and morphological characteristics of the organisms, one problem that remains concerns the exact evolutionary relationships between archaebacteria and the gram-positive bacteria, both of which are monoderm prokaryotes. As pointed out in this study, while archaebacteria form a monophyletic lineage by many gene phylogenies, the phylogenies based on several highly conserved proteins favor a polyphyletic distribution of these within the gram-positive bacteria. To explain this, some gene transfer events between these groups must be postulated. The nature of such events will be examined elsewhere (Gupta, 1998). However, any gene transfer between these two groups of monoderm prokaryotes will not affect their placement in the Monodermata taxon.

ACKNOWLEDGMENTS

I thank B. Singh for assistance with sequencing of Hsp70 genes and V. Johari for database searches. Thanks are also due to Dr. L. Margulis for helpful comments on an earlier draft of the paper. The work was supported by a research grant from the Medical Research Council of Canada.

REFERENCES

Baldauf, S. L., Palmer, J. D., and Doolittle, W. F. 1996. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny, *Proc. Natl. Acad. Sci. USA* **93**, 7749–7754.

Belfort, M., and Weiner, A. 1997. Another bridge between kingdoms: tRNA splicing in archaea and eukaryotes, *Cell* **89**, 1003–1006.

Benachenhou-Lahfa, N., Forterre, P., and Labedan, B. 1993. Evolution of glutamate dehydrogenase genes: Evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life, *J. Mol. Evol.* **36**, 335–346.

Brown, J. R., Masuchi, Y., Robb, F. T., and Doolittle, W. F. 1994. Evolutionary relationships of bacterial and archaeal glutamine synthetase genes, *J. Mol. Evol.* **38**, 566–576.

Brown, J. R., and Doolittle, W. F. 1995. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications, *Proc. Natl. Acad. Sci. USA* **92**, 2441–2445.

Brown, J. R., and Doolittle, W. F. 1997. *Archaea* and the prokaryote-to-eukaryote transition, *Microbiol. Rev.* **61**, 456–502.

Bui, E. T., Bradley, P. J., and Johnson, P. J. 1996. A common evolutionary origin for mitochondria and hydrogenosomes, *Proc. Natl. Acad. Sci. USA* **93**, 9651–9656.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., and Venter, J. C. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*, *Science* **273**, 1058–1073.

Cavalier-Smith, T. 1987. The origin of eukaryotic and archaebacterial cells, *Ann. N. Y. Acad. Sci.* **503**, 17–54.

Cavalier-Smith, T. 1992. Origins of secondary metabolism, *Ciba Found. Symp.* **171**, 64–80.

Chatton, E. 1937. "Titres et Travaux Scientifiques (1906–1937) de Edouard Chatton," E. Sottano, Sete, France.

Dennis, P. P. 1997. Ancient ciphers: Translation in Archaea, *Cell* **89**, 1007–1010.

Doolittle, R. F. 1995. Of archae and eo: What's in a name? *Proc. Natl. Acad. Sci. USA* **92**, 2421–2423.

Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G., and Little, E. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock, *Science* **271**, 470–477.

Edgell, D. R., and Doolittle, W. F. 1997. Archaea and the origin(s) of DNA replication proteins, *Cell* **89**, 995–998.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., and Merrick, J. M. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, *Science* **269**, 496–512.

Forterre, P. 1997. Protein versus rRNA: Problems in rooting the universal tree of life, *ASM News* **63**, 89–95.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., and Kelley, J. M. 1995. The minimal gene complement of *Mycoplasma genitalium*, *Science* **270**, 397–403.

Germot, A., Philippe, H., and Le Guyader, H. 1996. Presence of a mitochondrial-type 70-kDa heat shock protein in *Trichomonas vaginalis* suggest a very early mitochondrial endosymbiosis in eukaryotes, *Proc. Natl. Acad. Sci. USA* **93**, 14614–14617.

Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E. J., Bowman, B. J., Manolson, M. F., Poole, R. J., Date, T., and Oshima, T. 1989. Evolution of the vacuolar H+-ATPase: Implications for the origin of eukaryotes, *Proc. Natl. Acad. Sci. USA* **86**, 6661–6665.

Golding, G. B., and Gupta, R. S. 1995. Protein-based phylogenies support a chimeric origin for the eukaryotic genome, *Mol. Biol. Evol.* **12**, 1–6.

Gram, C. 1884. Ueber die isolierte farbung der Schizomyceten in Schnitt und Trockenpraparaten, *Fortschr. Med.* **2**, 185–189.

Gray, M. W. 1992. The endosymbiont hypothesis revisited, *Int. Rev. Cytol.* **141**, 233–357.

Gupta, R. S., Aitken, K., Falah, M., and Singh, B. 1994. Cloning of *Giardia lamblia* heat shock protein HSP70 homologs: Implications regarding origin of eukaryotic cells and of endoplasmic reticulum, *Proc. Natl. Acad. Sci. USA* **91**, 2895–2899.

Gupta, R. S. 1995. Phylogenetic analysis of the 90 kD heat shock family of protein sequences and an examination of the relationship among animals, plants, and fungi species, *Mol. Biol. Evol.* **12**, 1063–1073.

Gupta, R. S. 1997. Protein phylogenies and signature sequences: Evolutionary relationships within prokaryotes and between prokaryotes and eukaryotes, *Antonie van Leeuwenhoek* **72**, 49–61.

Gupta, R. S., Bustard, K., Falah, M., and Singh, D. 1997. Sequencing of heat shock protein 70 (DnaK) homologs from *Deinococcus proteolyticus* and *Thermomicrobium roseum* and their integration in a protein-based phylogeny of prokaryotes, *J. Bacteriol.* **179**, 345–357.

Gupta, R. S. 1998. Protein phylogenies and signature sequences: A reappraisal of the evolutionary relationships between archaebacteria, eubacteria, and eukaryotes, *Microbiol. Rev.*, submitted.

Gupta, R. S., and Golding, G. B. 1993. Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes, *J. Mol. Evol.* **37**, 573–582.

Gupta, R. S., and Golding, G. B. 1996. The origin of the eukaryotic cell, *Trends Biochem. Sci.* **21**, 166–171.

Gupta, R. S., and Singh, B. 1992. Cloning of the HSP70 gene from *Halobacterium marismortui*: Relatedness of archaebacterial HSP70 to its eubacterial homologs and a model for the evolution of the HSP70 gene, *J. Bacteriol.* **174**, 4594–4605.

Gupta, R. S., and Singh, B. 1994. Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus, *Curr. Biol.* **4**, 1104–1114.

Hashimoto, T., and Hasegawa, M. 1996. Origin and early evolution of eukaryotes inferred from the amino acid sequences of translation elongation factors 1alpha/Tu and 2/G, *Adv. Biophys.* **32**, 73–120.

Holt, J. G., Krieg, N. R., Sneath, P. H. A., Staley, J. T., and Stanley, T. W. 1994. "Bergey's Manual of Determinative Bacteriology," Williams &Wilkins, Baltimore.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes, *Proc. Natl. Acad. Sci. USA* **86**, 9355–9359.

Kandler, O., and Konig, H. 1993. Cell envelopes of archaea: Structure and chemistry, *in* "The Biochemistry of Archaea (Archaebacteria)" (M. Kates, D. J. Kushner, and A. T. Matheson, Eds.), p. 223, Elsevier, Amsterdam/New York.

Karlin, S., Weinstock, G. M., and Brendel, V. 1995. Bacterial classifications derived from recA protein sequence comparisons, *J. Bacteriol.* **177**, 6881–6893.

Karlin, S., Mrazek, J., and Campbell, A. M. 1997. Compositional biases of bacterial genomes and evolutionary implications, *J. Bacteriol.* **179**, 3899–3913.

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides,

N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., and Loftus, B. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*, *Nature* **390**, 364–NIL7.

Klenk, H. P., and Zillig, W. 1994. DNA-dependent RNA polymerase subunit B as a tool for phylogenetic reconstructions: Branching topology of the archaeal domain, *J. Molec. Evol.* **38**, 402–432.

Koonin, E. V., Mushegian, A. R., Galperin, M. Y., and Walker, D. R. 1997. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea, *Mol. Microbiol.* **25**, 619–637.

Lake, J. A., and Rivera, M. C. 1994. Was the nucleus the first endosymbiont? *Proc. Natl. Acad. Sci. USA* **91**, 2880–2881.

Langer, D., Hain, J., Thuriaux, P., and Zillig, W. 1995. Transcription in archaea: similarity to that in eucarya, *Proc. Natl. Acad. Sci. USA* **92**, 5768–5772.

Margulis, L. 1970. "Origin of Eukaryotic Cells," Yale Univ. Press, New Haven, CT.

Margulis, L. 1993. "Symbiosis in Cell Evolution," Freeman, New York.

Margulis, L. 1996. Archaeal–eubacterial mergers in the origin of Eukarya: Phylogenetic classification of life, *Proc. Natl. Acad. Sci. USA* **93**, 1071–1076.

Martin, W., and Muller, M. 1998. The hydrogenosome hypothesis for the first eukaryote, *Nature* **392**, 37–41.

Mayr, E. 1990. A natural system of organisms, *Nature* **348**, 491.

Murray, R. G. E. 1968. Microbial structure as an aid to microbial classification and taxonomy, *Spisy Prirodoved Fak. Univ. J. E. Purkyne Brne* **43**, 249–252.

Murray, R. G. E. 1986a. Kingdom Procaryote, *in* "Bergey's Manual of Systematic Bacteriology" (P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt, Eds.), p. 34, Williams & Willkins, Baltimore.

Murray, R. G. E. 1986b. The higher taxa, or, a place for everything...? *in* "Bergey's Manual of Systematic Bacteriology" (P. H. A. Sneath, N. S. Mair, M. E. Sharpe, and J. G. Holt, Eds.), p. 31, Williams & Wilkins, Baltimore.

Olsen, G. J., and Woese, C. R. 1993. Ribosomal RNA: a key to phylogeny, *FASEB J.* **7**, 113–123.

Olsen, G. J., and Woese, C. R. 1996. Lessons from an Archaeal genome: what are we learning from Methanococcus jannaschii? *Trend. Genet.* **12**, 377–379.

Olsen, G. J., and Woese, C. R. 1997. Archaeal genomics: An overview, *Cell* **89**, 991–994.

Pace, N. R. 1991. Origin of life—Facing up to the physical setting, *Cell* **65**, 531–533.

Popper, K. 1968. "The Logic of Scientific Discovery," Hutchison, London.

Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H. P., Lottspeich, F., Garrett, R. A., and Zillig, W. 1989. Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome, *Proc. Natl. Acad. Sci. USA* **86**, 4569–4573.

Reeve, J. N., Sandman, K., and Daniels, C. J. 1997. Archaeal histones, nucleosomes, and transcription initiation, *Cell* **89**, 999–1002.

Rivera, M. C., and Lake, J. A. 1992. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives, *Science* **257**, 74–76.

Schopf, J. W. 1978. The evolution of the earliest cells, *Sci. Am.* **239**, 110–120.

Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H. M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D. Y., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., and Bush, D. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* DeltaH: Functional analysis and comparative genomics, *J. Bacteriol.* **179**, 7135–7155.

Stanier, R. Y., Adelberg, E. A., and Ingraham, J. L. 1976. "The Microbial World," Prentice–Hall, New York.

Tiboni, O., Cammarano, P., and Sanangelantoni, A. M. 1993. Cloning and sequencing of the gene coding glutamine synthetase I from the archaeum *Pyrococcus woesi*: Anomalous phylogenies inferred from analysis of archael and bacterial glutamine synthetase I sequence, *J. Bacteriol.* **175**, 2961–2969.

TIGR Microbial Database Website, 1998. http://www.tigr.org/tdb/mdb/mdb.html.

Tipper, D. J., and Wright, A. 1979. The structure and biosynthesis of bacterial cell walls, *in* "The Bacteria" (J. R. Sokatch, and L. N. Ornston, Eds.), Vol. VII, p. 291, Academic Press, New York.

Woese, C. R. 1987. Bacterial evolution, *Microbiol. Rev.* **51**, 221–271.

Woese, C. R., Kandler, O., and Wheelis, M. L. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya, *Proc. Natl. Acad. Sci. USA* **87**, 4576–4579.

Woese, C. R. 1992. Prokaryote systematics: the evolution of a science, *in* "The Prokaryotes" (A. Balows, H. G. Truper, M. Dworkin, W. Harder, and K. H. Schleifer, Eds.), p. 3, Springer-Verlag, New York.

Zillig, W., Klenk, H. P., Palm, P., Leffers, H., Puhler, G., Gropp, F., and Garrett, R. A. 1989. Did eukaryotes originate by a fusion event? *Endocytobiosis Cell Res.* **6**, 1–25.