MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Dec. 1997, p. 456–502 Vol. 61, No. 4
1092-2172/97/$04.00+0


# *Archaea* and the Prokaryote-to-Eukaryote Transition

JAMES R. BROWN[1,2]* AND W. FORD DOOLITTLE[1]

*Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada,[1] and Department of Bioinformatics, SmithKline Beecham Pharmaceuticals, Collegeville, Pennsylvania[2]*



* Corresponding author. Mailing address: Department of Bioinformatics, SmithKline Beecham Pharmaceuticals, P.O. Box 5089, UP1345, Collegeville, PA 19426-0989. Phone: (610) 917-6374. Fax: (610) 917-7901. E-mail: James_R_Brown@sbphrd.com.

## INTRODUCTION

For over 50 years, scientists were confident with the notion that there were two basic kinds of living organisms, eubacteria and eukaryotes (306–308). In the late 1970s, this fundamental belief was shattered by the revelations of Woese and coworkers that life consisted not of two but three distinct groups of organisms (123, 337)—eukaryotes and two kinds of prokaryotes, the eubacteria and the archaebacteria. Since then, our knowledge of the latter has reached the point where the first complete genome sequence from an archaebacterium (49) is now known, with several more soon to follow (31, 58, 161a). Given the growth of DNA sequence databases, science is now poised to make broad and sweeping comparisons of living organisms. However, well before the dawning of this new age of entire genome sequences, the cumulative efforts of many highly determined and capable researchers had led to an impressive knowledge base about the archaebacteria. This review is a modest attempt to summarize some of that research in the context of what has been learned about the nature of the universal tree of living organisms. Hopefully, from viewing the archaebacteria in this way, a better appreciation can be gained about the evolutionary significance of these remarkable organisms.

Contemporary views on early cellular evolution have been strongly shaped by molecular phylogenetics. Ever since Woese and coworkers demonstrated the distinctiveness of archaebacteria on the basis of cluster dendrograms of data based on RNase $T_1$ oligonucleotide catalogs of rRNAs, phylogenetic analyses have played a pivotal role in the maturation, and often the upheaval, of macroevolutionary theory (336). Thus, a second goal of this review is to attempt a synthesis of universal trees based on different protein coding genes by reviewing and occasionally updating earlier phylogenetic studies and by adding analyses of new gene families. The dynamic growth of sequence databases makes it impossible to assemble a highly current yet comprehensive collection of species or genes prior to publication. Therefore, the multiple gene phylogenies presented here are best seen as a general overview of the universal tree of life from different biochemical perspectives.

## *ARCHAEA* ALONE

In 1990, Woese et al. (339) strongly advocated the replacement of the bipartate view of life with a new tripartite scheme based on three urkingdoms or domains; the *Bacteria* (eubacteria), *Archaea* (archaebacteria) and *Eucarya* (eukaryotes) (Fig. 1). The rationale behind this revision came from a growing body of evidence, in particular rRNA phylogenies, that the archaebacteria were worthy of the same taxonomic status as eukaryotes and eubacteria. Since then, the database of small-subunit rRNA sequences has grown to include over 2,000 prokaryotes and several hundred eukaryotes. Despite the addition of a large number of new species, the existence of three major groups or clades of organisms is consistent throughout rRNA phylogenies (59, 246, 247, 303, 336).

However, the three-domain classification has been contested, most notably by Mayr (232), Margulis and Guerrero (224), and Cavalier-Smith (57). As a result, there is an awkward coexistence between the terminology of the old and new

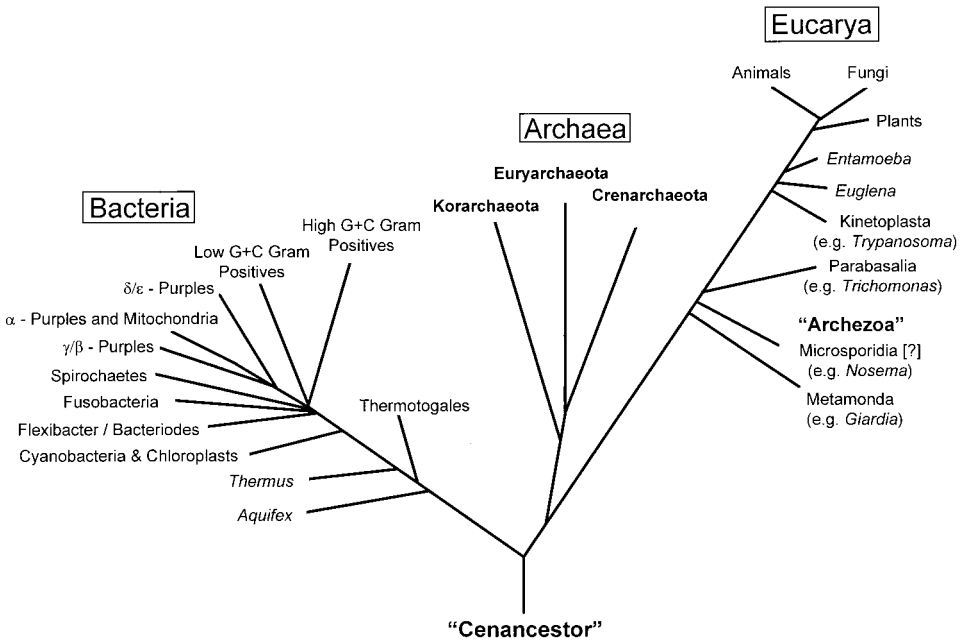


FIG. 1. Schematic drawing of a universal rRNA tree showing the relative positions of evolutionary pivotal groups in the domains *Bacteria*, *Archaea*, and *Eucarya*. The location of the root (the cenancestor) corresponds to that proposed by reciprocally rooted gene phylogenies (43, 133, 164). The question mark beside the Archezoa group Microsporidia denotes recent suggestions that it might branch higher in the eukaryotic portion of the tree. (Branch lengths have no meaning in this tree.)

schemes in the literature. Since the tripartate scheme basically suggests new subdivisions only in the prokaryotes, here we refer to the former archaebacteria and eubacteria as the *archaea* and *bacteria*, while eukaryotes will still be called as such.

According to rRNA trees, there are two groups within the *Archaea*: the kingdoms *Crenarchaeota* and *Euryarchaeota* (336). The kingdom *Crenarchaeota* generally consists of hyperthermophiles or thermoacidophiles (some genera are *Sulfolobus*, *Desulfurococcus*, *Pyrodictium*, *Thermoproteus* and *Thermofilum*). The kingdom *Euryarchaeota* spans a broader ecological range and includes hyperthermophiles (some genera are *Pyrococcus* and *Thermococcus*), methanogens (e.g., *Methanosarcina*), halophiles (some genera are *Halobacterium* and *Haloferax*), and even thermophilic methanogens (some genera are *Methanothermus*, *Methanobacterium*, and *Methanococcus*). Recent PCR amplifications of rRNA sequences from water and sediment samples have revealed a plethora of new archaeal species belonging to either kingdom living in mesophilic environments such as temperate marine coastal waters, the Antarctic Ocean, freshwater lakes, and even marine sponges (87, 88, 125, 261, 310). Recently, PCR surveys of hot spring microbiota detected new archaeal rRNA sequences that branch either deeply within the *Crenarchaeota* or just below the *Crenarchaeota-Euryarchaeota* divergence (14, 15). As such, these "organisms" have been tentatively assigned to a third kingdom, the *Korarchaeota* (Fig. 1).

It is also important to consider cellular and biochemical features of archaea with respect to the coherence of the domain *Archaea* itself and its evolutionary relationship to *Bacteria* and eukaryotes. The comparative biochemistry and the cellular biology of *Archaea*, *Bacteria*, and eukaryotes have been extensively reviewed elsewhere; thus, only the salient points will be discussed here (47, 175, 181, 342). Archaea have some unique characteristics as well as unique combinations of characteristics once thought to be exclusive to either the bacteria or eukaryotes. Some solely archaeal characteristics include isopranyl ether lipids, the absence of acyl ester lipids and fatty acid synthetase, modified tRNA molecules, a split in one of the RNA polymerase subunits, and a specific range of antibiotic sensitivities (reviewed in references 175, 181, and 187).

The structure of archaeal membranes has been thoroughly reviewed elsewhere (127, 174, 194). Briefly, archaeal lipids differ from those of bacteria and eukaryotes in four significant ways. First, Archaea have ether linkages established between glycerol and hydrocarbon chains while bacteria and eukaryotes have ester linkages. Second, archaea have highly methyl-branched isopranyl chains, while hydrocarbons in bacteria and eukaryotes are predominantly straight-chain fatty acyl chains. Third, archaeal glycerol ethers contain 2,3-*sn*-glycerol, which differs from the 1,2-*sn*-glycerols found in the two other domains. Fourth, some lipids in archaea are tetraethers for which ester lipids have no comparable structures (127). Some lipid biosynthetic enzymes in archaea are also different. The methanogen *Methanobacterium thermoautotrophicum* has a bifunctional prenyltransferase that provides precursors to both squalene and isoprenoid glyceryl lipids, whereas bacteria and eukaryotes use separate enzymes for the synthesis of these short-chain molecules (64, 65).

Among species of *Archaea*, there are a variety of metabolic regimes which differ greatly from the better-known metabolic pathways of the *Bacteria* and eukaryotes (reviewed in references 81 through 83 and 294). For example, both ATP-dependent and pyrophosphate-dependent phosphofructokinases can occur in bacteria and eukaryotes, while archaea use either ADP-dependent or pyrophosphate-linked kinases. Hexokinase is ATP dependent in bacteria, eukaryotes, and *Thermoproteus* but ADP-linked in *Pyrococcus* (182). The conversion of pyruvate to acetate, a reaction which bridges glycolytic and citric acid cycles, is catalyzed by a pyruvate dehydrogenase multienzyme complex in bacteria and eukaryotes, whereas archaea and some anaerobic eukaryotes (*Entamoeba*, *Giardia*, and *Trichomonas*) appear to employ pyruvate:ferredoxin oxidoreductase (81, 294). Dihydrolipoamide dehydrogenase, a component of the pyruvate dehydrogenase complex, has been detected in halophilic archaea, although its function is unclear (84).

## *ARCHAEA* AND *BACTERIA*

The members of the *Archaea* and *Bacteria* are united in the "realm of prokaryotes" by similar general cell sizes, the lack of a nuclear membrane and organelles, and the presence of a large circular chromosome occasionally accompanied by one or more smaller circular DNA plasmids. As an example, the chromosome complement of the halophilic archaeon *Haloferax volcanii* consists of one large circular genome roughly 2.92 Mbp in size and four smaller plasmid genomes with sizes of 690, 442, 86, and 6 kbp (62). Nearly all of the 60 to 70 identified *H. volcanii* genes map to the large circular genome. Although the origin of DNA replication of any archaeal large chromosome has yet to be confirmed experimentally, at present, there is little suggestion for any significant departure from the bacterial model of a single replication initiation site (122). The overall closer similarity between archaeal and bacterial topoisomerases and gyrases provides indirect evidence of comparable chromosome structure among the two groups.

Many archaeal genes appear to be organized into *Bacteria*-like operons. Furthermore, many archaeal operons and gene clusters are arranged in a similar fashion to those of the *Bacteria* (reviewed in references 181 and 268). As an example, ribosomal operons in bacteria and chloroplasts are arranged in the order 16S-23S-5S. Archaea have the same organization for these rRNAs with some variation, such as a $tRNA^{Ala}$ gene inserted between the 16S and 23S genes of methanogens and halophiles and a distal location of the 5S rRNA gene in some thermoacidophiles and methanogens (reviewed in reference 47).

In *Escherichia coli*, the ribosomal proteins RP L11 and RP L1 are clustered upstream of another group, containing RP L10 and RP L12. These genes occur in the same order in *Sulfolobus solfataricus* and *Halobacterium cutirubrum*, while *Methanococcus vannielii* has RP L11 translocated (12, 269, 296). Of the 11 genes in the *E. coli* spectinomycin (*spc*) operon, the same order occurs for 9 and 11 genes in *S. acidocaldarius* and *Methanococcus vannielii*, respectively (8, 268). Both species have three small additional open reading frames (ORFs) within their *spc* operons.

In comparison to the eight-gene S10 operon in *E. coli*, those of *Methanococcus vannielii* and *Halobacterium marismortui* are missing the first gene, RP S10, but have the remaining seven in similar orders (7). In *E. coli*, the S10 operon is distally located from the streptomycin (*str*) operon. The *str* operon has been sequenced from a wide phylogenetic range of bacteria, and in most cases, the genes for RP S12, RP S7, elongation factor G (EF-G) (*fus*), and EF-Tu (*tuf*) are found. However, the *str* operon of the hyperthermophilic bacterium *Aquifex pyrophilus*, which is the deepest-branching bacterial species in rRNA trees (50), lacks both ribosomal protein genes (31). Other variations include the translocation of the gene for RP S10 from the S10 operon to downstream of *tuf* in the cyanobacterial *Spirulina platenis* (287) and *Cyanophora paradoxa* cyanelles (242).

In the archaeon *M. vannielii*, the *str* operon is similarly or-

ganized, with the RP S10 gene downstream of the Ef-1α gene (bacterial EF-G and EF-Tu are homologous to EF-2 and EF-1α, respectively, from archaea and eukaryotes). The gene order is varied in *S. acidocaldarius*, where the EF-2 gene is missing and a $tRNA^{Ser}$ gene follows RP S10 (10). In *Pyrococcus furiosus*, the genes for Ef-1α, RP S10, and $tRNA^{Ser}$ are together but RP S7 is absent (76), while in *H. halobium*, RP S12, RP S7, and EF-2 are present but EF-1α and RP S10 are absent (163, 214).

The operons encoding the three largest subunits of DNA-dependent RNA polymerase (RNAP) in the *Archaea* and *Bacteria* can be potentially traced back to an ancestral organization. Here, the transcriptional polarity and gene order of the *S. acidocaldarius* RNAP (*rpo*) subunits B, A1, and A2 are reversed in comparison to *E. coli* (177). Furthermore, the *E. coli* genes *nusA*, encoding a transcription termination factor, and *infB*, encoding translation initiation factor 2, are located upstream of the *rpo* operon in the same transcriptional polarity. In *Sulfolobus*, the *rpo* operon splits the *nusA-infB* operon in two. The *infB* gene is now located downstream of the *rpo* operon, with the opposite transcription polarity, while the *nusA* gene is situated upstream between two ribosomal protein genes and is probably transcribed in the same direction. Keeling et al. (177) suggested that a single inversion event might be sufficient to explain this major rearrangement between *E. coli* and *Sulfolobus rpo* operons.

Like mRNAs in the *Bacteria*, archaeal mRNAs do not have 5′-end caps and often have Shine-Dalgarno ribosome binding sites (reviewed in reference 5). However, the locations of putative Shine-Dalgarno sequences relative to the translational initiation codon are more variable in the members of the *Archaea*, and the upstream sequences of several highly expressed genes bear little resemblance to Shine-Dalgarno motifs (see Fig. 6 in reference 268). Little is known about the actual process of translation initiation in the *Archaea*, although several sequences similar to eukaryotic translation initiation factors have been detected in *Sulfolobus* (17, 178) and *Methanococcus jannaschii* (49). However, even in the *Bacteria* and eukaryotes, the evolutionary relationships and exact functions of several initiation factors are unclear. Until protein synthesis has been better studied, one can only predict that translation initiation in the *Archaea* probably deviates somehow from the *E. coli* model.

The bacterial cell division protein FtsZ has been discovered in several species of *Archaea* (19, 327). FtsZ is thought to be a distant homolog to eukaryotic tubulins, since both proteins are GTPases that polymerize into filaments in the presence of GTP (107). *Arabidopsis* has a nucleus-encoded FtsZ homolog which is directed to the chloroplast (248). Phylogenetic analysis also supports the notion that FtsZ is a distant homolog of tubulin, with the FtsZ of the *Archaea* and *Bacteria* forming distinct yet closely related groups (19). Among these prokaryotic FtsZ homologs, those of the *Archaea* are marginally closer to the large eukaryotic family of tubulins. Some cell division proteins have been identified in the *M. jannaschii* genome but not the entire suite of genes known to function in cell septation or chromosome partitioning in the *Bacteria* (49). Either more cell division homologs remain to be found, or the archaea use a unique system based on particular *Bacteria*-like components.

Many members of the *Archaea* also have type II restriction enzyme systems, similar to those of the *Bacteria* (summarized in reference 47). Plasmid genes encoding for restriction-modification enzymes, endodeoxyribonuclease and DNA methyltransferase, in *Methanobacterium thermoformicium* show significant homology to those of the bacterium *Neisseria gonorrhoeae* (244).

Archaeal and bacterial protein-coding genes lack spliceosomal introns typically found in eukaryotic genes. A full discussion of the origin and distribution of introns is beyond the scope of this review. Noteworthy is that when comparing homologous genes, introns found in "higher eukaryotes," the metazoans, are frequently absent in "lower eukaryotes," i.e., the protists (250). Therefore, the absence of spliceosomal introns should not be considered a defining characteristic of prokaryotes. The genes encoding 16S and 23S rRNAs and tRNAs in several archaea have introns with ORFs that correspond to the homing endonucleases of group I introns found in mitochondria and bacteriophages (22, 47). Inteins are unusual introns spliced at the protein rather than the mRNA level, which were first found in the catalytic subunit of the yeast vacuolar type ATPase (158, 172). The first archaeal intein was found in the DNA polymerase gene of *Thermococcus litoralis* (254). Several new intein insertions were reported in the whole genome sequence of *M. jannaschii* (49).

## *ARCHAEA* AND EUKARYOTES

Although the domains *Archaea* and *Bacteria* appear very similar in terms of general genome organization, many archaeal genes show greater similarity to eukaryotic homologs. Hints of genetic homology among the *Archaea* and eukaryotes were found in early studies with antibiotics (summarized in reference 5). Bacteria are sensitive to streptomycin, an anti-70S ribosome-directed inhibitor. Archaea and eukaryotes are both refractory to streptomycin but are sensitive to certain anti-80S ribosome directed inhibitors (such as anisomycin). Archaea and eukaryotes also share sensitivity to aphidicolin, an inhibitor of DNA polymerase, to which bacteria are refractory (reviewed in reference 122).

Later studies showed significant similarities between archaeal and eukaryotic DNA replication, transcriptional, and translational components. Archaeal and eukaryotic DNA polymerases are homologous and not related to any bacterial DNA polymerase except that of *E. coli* (reviewed in reference 99). Several other DNA replication components are similarly shared only between *Archaea* and eukaryotes. Archaeal RNAPs are evolutionarily closer to those of eukaryotes (263, 345). In addition, bacterial RNAPs have a simpler structure composed of only four major subunits while eukaryotic and archaeal RNAPs have a minimum of seven subunits, of which several are homologous (139, 207, 345). Phylogenetic trees obtained with the sequences for large RNAP subunits strongly show that eukaryotic and archaeal genes are close relatives (188, 263, 344). The *Archaea* and eukaryotes also share other features of transcription that are apparently absent in the *Bacteria*. These include TATA box-like binding sites (275, 345) and some transcription factors (226, 249, 273, 283).

Pre-rRNA processing in eukaryotes involves many different small nucleolar rRNAs (snoRNAs) associating with the protein fibrillarin. The gene for fibrillarin has been cloned and sequenced from methanogenic members of the *Archaea* (6, 49), and antibodies against fibrillarin have precipitated sno RNAs in *Sulfolobus* (91). Although a report of U3-like RNA cloned from *Sulfolobus acidocaldarius* (258) was subsequently determined to be an error (284), it is still possible that related snoRNAs are involved in the processing of archaeal pre-rRNA (96). Introns in tRNA genes of the *Archaea* and eukaryotes are of similar size and occur in mostly the same positions, although some archaeal tRNA introns have shifted locations (24). The excision of tRNA introns in eukaryotes involves a site-specific endonuclease which is composed of two subunits (186, 321).

The subunits are similar to each other as well as being homologous to archaeal tRNA endonucleases.

Members of the *Archaea* and eukaryotes share a pathway of isoprenoid biosynthesis that involves the synthesis of mevalonate from 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) by the enzyme HMGCoA reductase. The first archaeal HMGCoA reductase gene was sequenced from *H. volcanii* (204). The gene and the enzyme activity have been recently characterized from *S. solfataricus* (32). Bacteria lack HMG-CoA, although *Pseudomonas mevalonii* has a highly divergent HMGCoA reductase, which might be a specific adaptation to the use of mevalonate as a carbon source (129).

In the *Bacteria*, including the deep-branching thermophile *Thermotoga maritima*, acetyl-CoA is converted to acetate through the coordinated activities of two enzymes, phosphate acetyltransferase and acetate kinase. However, hyperthermophilic archaea employ a single enzyme, an ADP-forming acetyl-CoA synthetase which is absent in bacteria (reviewed in reference 294). Interestingly, ADP-forming acetyl-CoA synthetases were initially discovered in two anaerobic eukaryotes, *Entamoeba histolytica* (274) and *Giardia lamblia* (220). However, the opposing activation of acetate back to acetyl-CoA in the *Archaea* is not catalyzed by acetyl-CoA synthetase (ADP-forming). For this reaction, methanogens use acetate kinase and phosphate acetyltransferase (both encoding genes in *Methanosarcina thermophila* have been sequenced and are homologous to those of *E. coli* [211]) while *Thermoproteus neutrophilus*, a member of the *Crenarchaeota*, uses an AMP-forming acetyl-CoA synthetase (294).

Protein degradation is an important regulatory step in many cellular processes. In eukaryotes, an abundant protein complex, the 26S proteasome, is responsible for ATP-dependent proteolysis. The 26S proteasome consists of a 20S core subunit flanked by two 19S caps (222). Archaea have a reduced proteasome complex equivalent to the 20S subunit which was reported first for *Thermoplasma acidophilum* (80, 265, 348, 349) and later for *Methanosarcina thermophila* (230), *Pyrococcus furiosus* (18), and *Methanococcus jannaschii* (49). The eukaryotic proteasome is a complex structure involving at least 34 polypeptides, while the functional *T. acidophilum* proteasome consists of only two subunits (330, 348). Sequence similarities between eukaryotic proteasomal subunits and the large ($\alpha$) or small ($\beta$) subunits of the *Thermoplasma* proteasome suggest that the simplified archaeal proteasome might be ancestral (348, 349). Many bacteria have sequences similar to 20S proteasome $\beta$-type subunits, further suggesting that proteasomes or proteasome-like precursors existed in the last common ancestor (223). However, bacterial multisubunit proteasomes have been reported only in actinomycetes, and it has been proposed that those were horizontally transferred from either a eukaryote or an archaeon (222).

Studies of archaeal DNA-binding proteins or HMf, initially from the methanogen *Methanothermus fervidus*, suggest a strong similarity to eukaryotic histones both at the primary sequence level (138, 290) and in three-dimensional structure (309). Histone-encoding genes have now been determined from different members of the *Euryarchaeota*, where their number can vary between species (273). Putative archaeal nucleosomes differ from eukaryotic nucleosomes in that DNA strands are constrained about HMf particles as positive supercoils rather than the conventional negative supercoil conformation imposed by histones (239, 273). While eukaryotic histones form only H2A-H2B and H3-H4 heterodimers, archaeal histones can assemble as both homodimers and heterodimers. Archaea also have proteins similar to bacterial DNA-binding proteins, known as HU, which are not evolutionary linked to histones, although they perform similar functions (30).

In summary, there are some features that distinguish the *Archaea* from the *Bacteria* and eukaryotes, most notably the structure and composition of their membranes. Primarily, the members of the *Archaea* are unique in having a combination of traits which, until now, were believed to be exclusive to either the *Bacteria* or eukaryotes. However, until we have genome sequences from more diverse groups of prokaryotes and eukaryotes, evolutionary scenarios based on the presence of a character in two domains and its presumed absence in the third must be considered to be highly provisional.

## ROOTING THE UNIVERSAL TREE

There are three possible scenarios for the evolution of the three domains of life: (i) *Bacteria* diverged first from a lineage producing *Archaea* and eukaryotes, (ii) a proto-eukaryotic lineage diverged from a fully prokaryotic (*Bacteria* and *Archaea*) lineage, or (iii) *Archaea* diverged from a lineage leading to eukaryotes and *Bacteria*. However, on the basis of a solitary gene, it is impossible to derive an objective rooting for the universal tree. Typically, the rooting for a particular organismal tree, for example, all mammalian species, would be determined by including sequence data from a known outgroup species, such as some cold-blooded vertebrates. However, outgroup species are not available for a gene tree consisting of all living organisms unless specific assumptions are made such as the progression of life from a prokaryotic to a eukaryotic cell. Therefore, the branching order of the three domains emerging from their last common ancestor—which Fitch and Upper called the cenancestor (116)—can only be established by some method unrelated to either outgroup organisms or theories about primitive and advanced states.

In 1989, a solution to this problem, using ancient duplicated genes, was simultaneously proposed in separate papers by Gogarten et al. (133) and Iwabe et al. (164). Their collective reasoning was as follows: although there can be no organism which is an outgroup for a tree relating all organisms, one could root a tree based on the sequences of outgroup genes produced by an early gene duplication (Fig. 2A). Iwabe et al. (164) applied this concept by reciprocal rooting of trees for paralogous elongation factor genes. Elongation factors are a family of GTP-binding proteins which facilitate the binding of aminoacylated tRNA molecules to the ribosome (EF-Tu in *Bacteria* and EF-1$\alpha$ in eukaryotes and *Archaea*) and the translocation of peptidyl-tRNA (EF-G in *Bacteria* and EF-2 in eukaryotes and *Archaea*). Iwabe et al. aligned amino acids from five conserved regions shared by the EF-Tu/1$\alpha$ and EF-G/2 genes of the archaeon *Methanococcus vannielii* and several species of the *Bacteria* and eukaryotes. According to protein sequence similarity and neighbor-joining trees, both the EF-1$\alpha$ and EF-2 genes of the *Archaea* were more similar to their respective eukaryotic rather than bacterial homologs.

Gogarten et al. (133) developed composite trees based on a second gene duplication, that of the V type (found in *Archaea* and eukaryotes) and F-type (found in *Bacteria*) ATPase subunits. The catalytic $\beta$ subunit of F-type ATPases is most similar to the A or 70-kDa subunit of V-type ATPases, while the $\alpha$ subunit of F-type ATPases is most similar to the B or 60-kDa subunit of V-type ATPases (reviewed in reference 286). In agreement with the elongation factor rooting, reciprocally rooted ATPase subunits trees showed the archaeon, in this case *Sulfolobus acidocaldarius*, to be closer to eukaryotes than to the *Bacteria.*

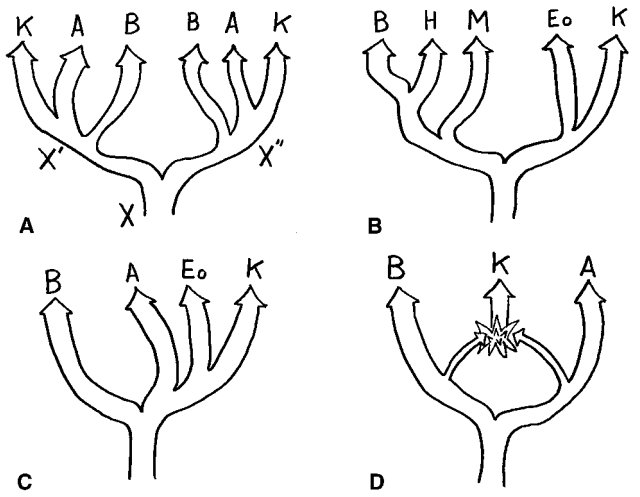


FIG. 2. Alternative scenarios about early cellular evolution. (A) Conceptual rooting of the universal tree by using paralogous genes. Suppose that some gene (X) was duplicated (X′ and X″) in the cenancestor such that all extant organisms have both genes. Provided that some sequence similarity still exists between genes X′ and X″, reciprocally rooted gene trees could be constructed. The positioning of *Archaea* (A) and eukaryotes (K) as sister groups with the *Bacteria* (B) as the outgroup, has been consistently supported by such rootings (43, 133, 164). (B) The 1988 eocyte tree as proposed by Lake and coworkers (199). The *Crenarchaeota* or eocytes (Eo) form a clade with eukaryotes, while the *Euryarchaeota*, namely, halophiles (H) and methanogens (M), cluster with the *Bacteria*. (C) The most recent eocyte tree advocated by Rivera and Lake in 1992 (278). The eocytes are the closest group of the *Archaea* to the eukaryotes, with the *Euryarchaeota* (A) being more distantly related although no longer branching with the *Bacteria*. (D) The "chimeric" or "fusion" hypothesis, which suggests that eukaryotes arose when a "gram-negative" bacterium engulfed an archaeon and components of their genomes fused (147, 202, 342, 343).

## ACCEPTANCE AND DOUBT

Archaebacteriologists (or "archaeologists") were already primed to accept the conclusions of these duplicated gene rootings of the universal tree, since early on there was a general feeling that the *Archaea* was somehow the "missing link" between the *Bacteria* and eukaryotes. At the time, eukaryote-like functional and structural characteristics of archaeal RNA and DNA polymerases and some ribosomal proteins were known. Woese et al. (339) incorporated the protein rooting in their formulation of the three domains, *Archaea*, *Bacteria*, and *Eucarya*. Although archaeal and bacterial rRNA sequences are slightly more similar, Woese et al. placed the root of the ribosomal tree such that *Archaea* and *Eucarya* were sister groups. Here, it is important to emphasize that their "archaeal" universal tree was formulated from three different data analyses. The separate monophyly of the *Archaea*, *Bacteria* and eukaryotes was suggested by rRNA gene trees (336), while the grouping of *Archaea* and eukaryotes together arose from the reciprocally rooted gene trees for elongation factors and ATPase subunits (133, 164).

However, some researchers have seriously challenged the topology of the archaeal universal tree. In an rRNA tree first proposed in 1988, Lake (199) broke up the *Archaea* (then archaebacteria) by placing the *Crenarchaeota*, which he called eocytes, in a clade with eukaryotes, now named karyotes (Fig. 2B). Methanogens, halophiles, and members of the *Bacteria* were in a separate group called the parkaryotes, which notably had the *Bacteria* in a clade with halobacteria (200). This early revisionist universal tree was based on differences in ribosome shapes (152, 201) and a novel phylogenetic analysis of rRNA sequences (199). However, as the rRNA data set grew and more ribosome structures were determined, Lake's 1988 eocyte tree became untenable. More recently, Rivera and Lake (278) found new support for the eocyte tree, this time in the analysis of a specific 11-amino-acid insertion shared in the EF-1α genes of eukaryotes and the *Crenarchaeota* but absent from the *Euryarchaeota* and *Bacteria*. This 1992 eocyte tree resembled the 1988 tree by still having a *Crenarchaeota*-eukaryote clade but differed in the reassignment of *Euryarchaeota* and *Bacteria* into individual clades (Fig. 2C).

Cammarano and coworkers (53, 75, 320) added several elongation factor sequences from new species of the *Archaea* and a deeply branching bacterium (*Thermotoga maritima*). Although they did not use a reciprocal rooting, their analyses of EF-G/2 sequences showed strong support for a monophyletic clade of the *Archaea*, subdivided into the kingdoms *Crenarchaeota* and *Euryarchaeota*. However, another analysis (13) of elongation factor genes, including many deep-branching species from all domains, found support, albeit statistically weak, for the divergence of eukaryotes within the *Archaea* as a sister group to the *Crenarchaeota*, which somewhat bolsters the eocyte tree of Rivera and Lake (278).

Forterre et al. (120) vigorously argued that neither the elongation factor nor the ATPase data set can settle the issue of rooting the universal tree. Their major criticisms concerned the paucity of taxa (which was largely addressed in the study by Balduaf et al. [13]) and the fact that only 120 amino acids could be aligned with confidence between EF-Tu/1α and EF-G/2, which are 390 to 460 and 700 to 860 amino acids long, respectively.

ATPase subunit gene phylogenies are also more problematic. Based on greater similarities between archaeal and eukaryotic V-type ATPases over bacterial $F_0F_1$-type ATPases, earlier analyses placed the root of the universal tree in the *Bacteria* (133, 164). At the time, known bacterial ATPases were of the $F_0F_1$ type while V-type ATPases were exclusive to the *Archaea* and eukaryotes. It was proposed that a gene duplication in the cenancestor resulted in the F-type β/V-type A (or 70-kDa) subunit, on the one hand, and the F-type α/V-type B (or 60-kDa) subunit, on the other.

Subsequently, archaeal V-type ATPases were reported for two bacterial species, *Thermus thermophilus* (322) and *Enterococcus hirae* (171), and an $F_1$-ATPase β-subunit gene was found in the archaeon *Methanosarcina barkeri* (313) (Fig. 3). Forterre et al. (120) suggested that the ATPase subunit gene family had not been fully determined and that other paralogous genes might exist. However, Hilario and Gogarten (156) suggest that the observed distribution of ATPase subunits is the result of a few lateral gene transfers between species of *Archaea* and *Bacteria*. In support of their view, broader surveys have failed to detect archaeal V-type ATPases in other bacterial species (132).

More recently, the rooting of the universal tree was attempted by using another duplicated gene family, the aminoacyl-tRNA synthetases (43). Novel isoleucyl-tRNA synthetases from species belonging to deep evolutionary lineages of the *Bacteria*, *Archaea*, and eukaryotes were used to construct a universal gene tree rooted by valyl- and leucyl-tRNA synthetases. The sisterhood of eukaryotes and *Archaea*, as well as the separate monophyly of all three domains, was strongly supported by this analysis. A similar conclusion was reached in the phylogenetic analysis of tryptophanyl- and tyrosyl-tRNA synthetases (46). Now that more archaeal aminoacyl-tRNA synthetases are known, further opportunities exist to derive multiple rooted universal trees.

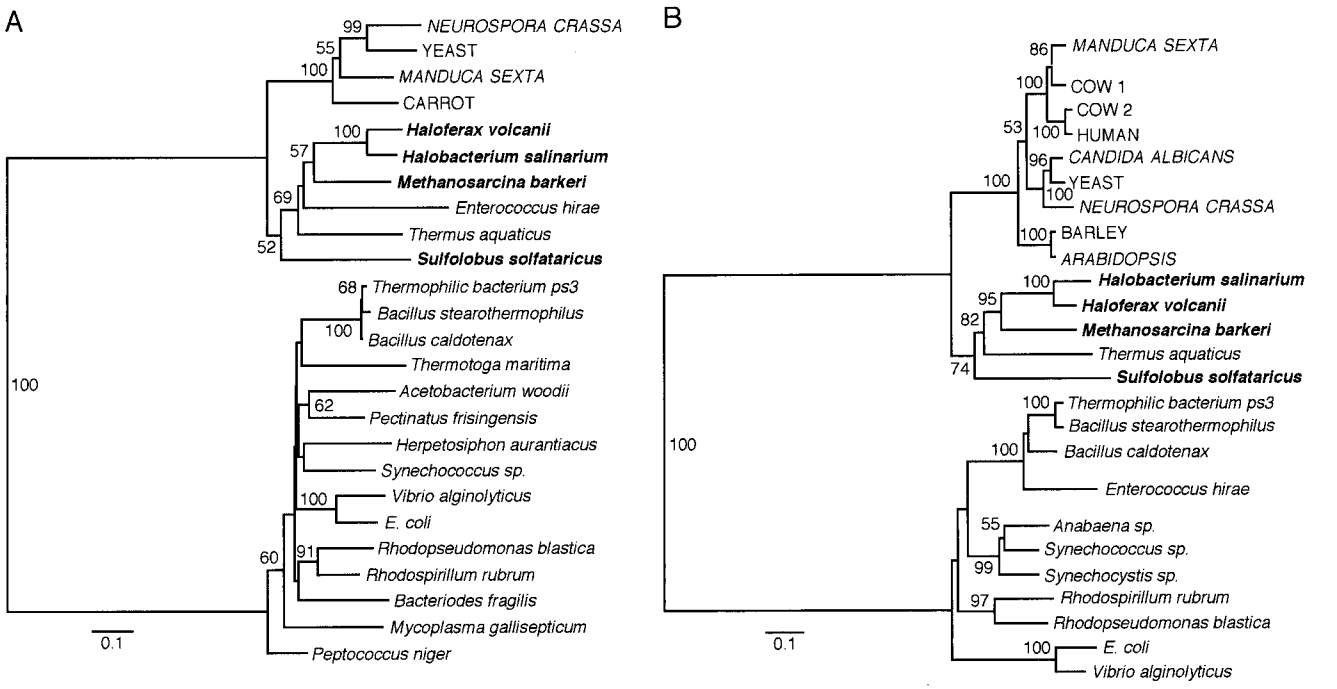


FIG. 3. Phylogenies for ATPase F-type β/V-type A subunit (A) and F-type α/V-type B subunit (B) genes from selected species. In this phylogenetic figure and all others, the different typefaces indicate whether the species belong to the eukaryotes (all capitals), *Archaea* (lowercase boldface), or Bacteria (lowercase lightface). Certain highly familiar species are referred to by their common genus or abbreviated species names to conserve space. Eukaryotic genes which are nucleus encoded yet targeted to the mitochondrion (mito.) or chloroplast (chl.) are indicated as such. Gene trees were constructed by the neighbor-joining method based on pairwise distance estimates of the expected number of amino acid replacements per site (0.1 in the scale bars). Numbers show the percent occurrence of nodes in 100 bootstrap replications, and only values greater than 50% are shown. The programs SEQBOOT, PROTDIST, NEIGHBOR, and CONSENSE of the PHYLIP 3.57 (114) package were used for the analysis.

## ORIGIN OF EUKARYOTES

Three different paralogous gene phylogenies (and a recently published fourth rooting obtained by using carbamoyl-phosphate synthetase [210]) provide a general consensus that the root of the universal tree lies somewhere in the *Bacteria*, thus positioning *Archaea* and eukaryotes as sister groups. However, there is still uncertainty about this rooting, since each duplicated gene data set has its own particular, and significant, shortcomings. Furthermore, three or four genes spanning a few thousand base pairs may not be representative of entire genomes with thousands of genes and, at least, several million base pairs.

There are many properties that make rRNA a suitable molecular marker for phylogenetic reconstruction: it occurs in all living organisms, its sequences are highly conserved, and there is no compelling evidence for interspecific transfers of rRNA genes. However, even the monophyly of *Archaea*, *Bacteria*, and eukaryotes, as strongly suggested by rRNA trees, is open to challenge. The higher G+C content of rRNA genes of certain organisms, such as thermophilic members of the *Archaea* and *Bacteria*, could be biasing phylogenetic reconstruction, and new environmental PCR-amplified rRNA sequences have in some instances reduced the overall statistical support for archaeal monophyly (15). More critically, there are growing numbers of protein gene phylogenies which challenge the notion of monophyletic domains and, indirectly, *Archaea*-eukaryote sisterhood.

The various universal protein gene phylogenies have been summarized previously (94, 134), and the further elaboration of this list is an important aspect of this review (Fig. 4). It seems that all combinations of domain branching orders are possible. Some gene trees, those like for glutamine synthetase (GS), glutamate dehydrogenase (GDH), and the 70-kDa heat shock protein (HSP70), position the *Archaea* as a paraphyletic group within the *Bacteria*. In other gene phylogenies, like those for glyceraldehyde-3-phosphate dehydrogenase (GAPDH), 3-phosphoglycerate kinase (PGK), and enolase, the *Bacteria* and eukaryotes cluster together. As with rRNA, any universal tree rooting based on a single gene can be inferred only by the midpoint method, which simply places the root somewhere in the center of the tree, as nearly equidistant from all organisms as possible. There is also a critical question of how well a particular gene tree reflects the actual evolution of the organisms. Horizontal gene transfer and lineage-specific differences in evolutionary rates could result in a gene tree radically different from the species phylogeny.

Just how widespread these phylogenetic distortion effects are and to what extent they have affected particular gene trees are important issues to molecular evolution. Developing evolutionary theories solely on the basis of any single gene phylogeny is, at best, highly speculative. However, if the majority of gene trees do correctly reflect the correct evolutionary origins for different bits of the genome, one might conclude that the descent of the *Bacteria*, *Archaea*, and eukaryotes from the cenancestor involved a more complex series of genetic events. In this vein, Sogin (304) and Zillig et al. (343, 344) theorized that the eukaryotic cell did not directly evolve from an archaea-like ancestor but, rather, that the eukaryotic nucleus arose from the cellular fusion between either a bacterium or "proto-eukaryote" and an archaeon (Fig. 2D).

Sogin's (304) version adheres to the initial conceptions of Woese and Fox of the cenancestor as a progenote (an organ-

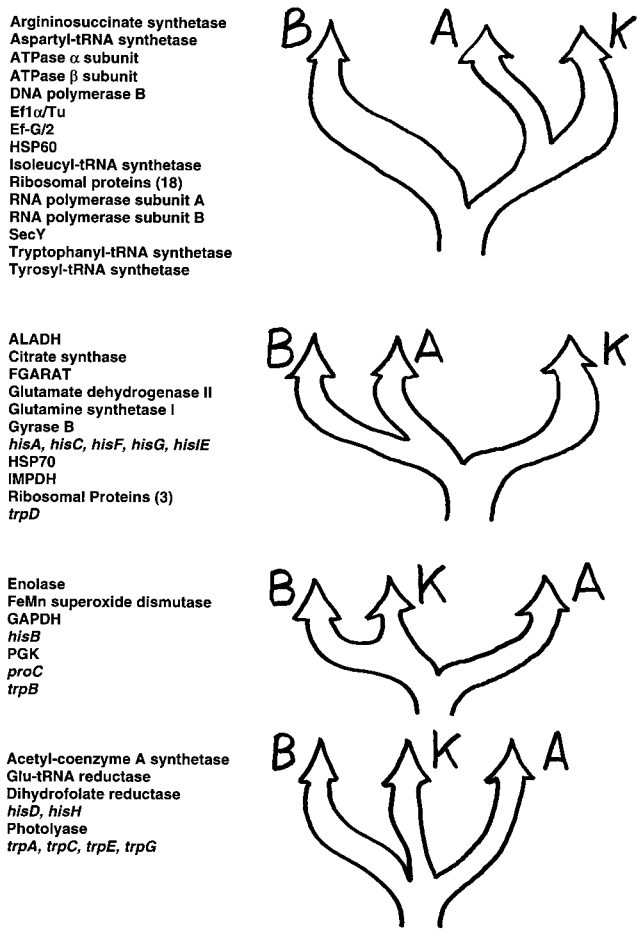


FIG. 4. Alternative rootings of the universal tree and the single-protein gene phylogenies that support them. Individual gene trees and abbreviations are given in the text.

ism where the genotype and phenotype were more loosely coupled than today [338]) from which two cellular lineages emerged: one composed of prokaryotic, DNA-based organisms (the *Bacteria* and *Archaea*) and the other composed of sophisticated RNA-based organisms (the putative proto-eukaryote). Subsequently, the *Archaea* and *Bacteria* diverged before the engulfment of an archaeon by the proto-eukaryote. The proto-eukaryote was proposed to have had a rudimentary cytoskeleton, since such an innovation would be necessary for phagocytosis. The engulfed archaeon formed the cell nucleus which led to the replacement of the host RNA genome by a DNA-based one. Sogin suggested that contradictions between rRNA phylogenies, which show contemporary the *Archaea* and *Bacteria* as most similar, and paralogous protein gene trees, which show *Archaea* and eukaryotes as sister groups, exist because eukaryotic rRNA is a remnant of the proto-eukaryote genome.

Zillig et al. (344) similarly proposed that separate lineages of the *Archaea* and *Bacteria* descended from the cenancestor but that eukaryotes did not exist until a cellular fusion occurred between species from the two prokaryotic groups. This model, unlike that of Sogin (304), has the cenancestor as a genote, a prokaryote with a fully functional genome, rather than a progenote. Gupta and Golding (144) have elaborated upon the hypothesis of Zillig et al. in proposing that it was a gram-negative bacterium that engulfed an archaeon; later, Gupta and Singh (147) suggested that it was an eocyte. Both theories suggest that any genome fusion event occurred prior to the widely accepted bacterial endosymbiosis leading to intracellular organelles, such as plastids and mitochondria.

Collectively, the theories of Sogin (304) and Zillig et al. (344) have been referred to as the chimeric or fusion hypothesis of the origin of the eukaryotic genome, although the version of Zillig et al. has been more widely considered (134). This terminology is somewhat confusing since the chimeric nature of the eukaryotic cell has been long recognized with respect to the endosymbiotic origin of organelles (reviewed in reference 137). In addition, it has been established that the eukaryotic genome is a chimera where genes of ancient eukaryotic ancestry coexist with genes more recently acquired from bacterial endosymbionts. In the context of this review, the term "chimera hypothesis" will be applied to suggestions that the eukaryotic genome originated from a fusion between two independent, noneukaryotic genomes, while the term "archaeal hypothesis" will refer to the more conventional view, i.e., that eukaryotes and the *Archaea* diverged recently from a common ancestor.

Unfortunately, there are few objective criteria for the rejection or acceptance of any of the chimera hypotheses. These hypotheses predict a mixture of phyletic relationships among different gene families which is self-evident. Other explanations for the observed mixing of domain relationships, such as unequal mutation rates, hidden gene paralogy, and horizontal gene transfers, cannot be strictly ruled out. Furthermore, only very broad speculation can be made about possible candidates for cell fusion participants. Although bacteria living intracellularly in a different bacterial species have been reported (209), phagocytosis by a bacterium has never been observed. Nor is there any evidence for the existence of the sophisticated RNA-based organisms with cytoskeletons integral to Sogin's proto-eukaryotic model. On the other hand, the endosymbiosis hypothesis, clearly establishes α-proteobacteria and cyanobacteria as the respective progenitors of mitochondria and plastids (chloroplasts).

Nonetheless, the version of the chimeric hypothesis proposed by Zillig et al. (344) has found some interest and support. Over the past 20 years following Woese's initial universal rRNA trees, many archaeal protein-coding genes have been sequenced that are homologous to counterparts in the *Bacteria* and eukaryotes. Golding and Gupta (134) constructed unrooted phylogenetic trees for 24 protein genes which were common to all three domains. They found that nine of the proteins gave statistically significant support for the monophyly of domains (sensu Woese's rRNA trees) and the closer association of the *Archaea* and eukaryotes (by midpoint rooting). However, seven of the protein gene trees supported an alternative topology with two clades, one of gram-positive bacteria and the *Archaea*, and the other of "*gram*-negative" bacteria and eukaryotes. The remaining eight protein trees were not statistically significant. Golding and Gupta stated that their results supported "the hypothesis of a chimeric origin for the eukaryotic cell nucleus formed from the fusion of an archaebacteria and a gram-negative Bacteria."

Such a conclusion is probably premature since new sequences can radically alter gene trees. If the full range of available sequences are considered, none of the seven protein gene trees support the monophyletic groups of gram-positive bacteria and the *Archaea* or gram-negative bacteria and eukaryotes (281). These gene trees, which include asparatate aminotransferase (AAT), GS, GDH, HSP70, histidinol phosphate synthetase, and pyrroline 5-carboxylate reductase, are described in more detail below. (Eukaryotic isoforms of the

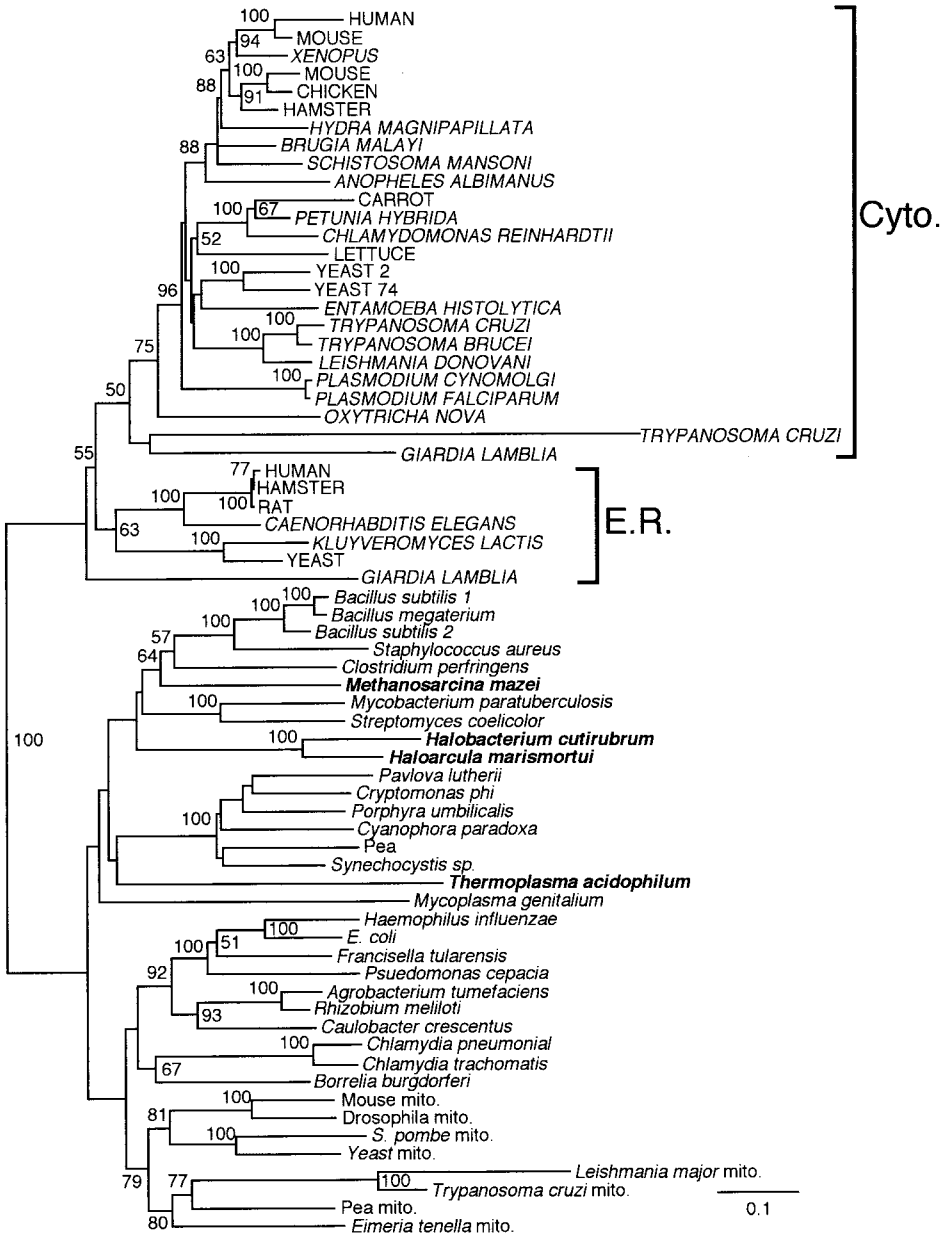


FIG. 5. The 70-kDa heat shock protein (HSP70) gene phylogeny. Cytostolic (Cyto.) and endoplasmic reticulum (E.R.) eukaryotic isoforms are indicated.

seventh gene, ferredoxin, occur only in plants and have probable plastid/cyanobacterial origins.) Recently, Gupta and Golding derived phylogenetic trees of phosphoribosylformylglycinamidine synthetase (FGAM synthetase) which clustered gram-positive bacteria and the *Archaea*, on one hand, and proteobacteria and eukaryotes on the other hand (145). However, subsequent analyses showed that the cyanobacterium *Synechococcous* also clusters with the group of gram-positive bacteria and *Archaea* (281).

Gupta et al. have extensively used HSP70 gene phylogenies as support for the chimeric origin of eukaryotes hypothesis (143–147). HSP70 genes from the *Archaea* and gram-positive bacteria are clearly more similar, with respect to both phylogeny (Fig. 5), and the absence of a specific 25-amino-acid sequence found in HSP70 genes of other bacteria and eukaryotes. However, a universal phylogeny based on HSP70 suffers the same drawbacks of any single-gene tree, in that it cannot be uniquely rooted. Since the rooting is subjective, the two clustering possibilities for the clade of gram-negative bacteria, either with eukaryotes or with gram-positive bacteria and *Archaea*, have equal validity. In addition, it is difficult to evaluate the proposition that eukaryotes arose from the specific fusion of an archaeon "eocyte" with a gram-negative bacterium since HSP70 gene sequences from *Crenarchaeota* are unknown.

According to Gupta et al., the earliest divergence was that separating gram-positive bacteria and the *Archaea*, yet all phylogenetic treatments of HSP70 genes show these groups as paraphyletic rather than monophyletic (whereas eukaryotes and all other bacteria can be resolved as separate monophy-

letic groups). They suggested that the intermixing of archaeal species with those of gram-positive bacteria and the *Archaea* reflected past patterns of alternatively evolving lineages. However, if such stochastic processes were involved in the early evolution of this HSP70 gene, the statistical reliability of later branching points is also open to debate. Furthermore, to reconcile the observed HSP70 gene phylogeny, one would have to postulate the unlikely case of a rapid acceleration in mutation rates in eukaryotes while gram-positive bacteria and the *Archaea* evolved so slowly as to appear similar to each other (281).

## UNIVERSAL PROTEIN GENE TREES

### Methods and Caveats

**Sequences.** Complete genomes are being sequenced from an ever-increasing number of bacterial, archaeal, and eukaryotic species. This flood of DNA sequence data brings new challenges and opportunities to the field of evolutionary biology (41). In this context, the comparative analysis of many individual universal gene trees, such as that initiated by Golding and Gupta (134), has value as a tentative, conceptual framework for the evolutionary analyses of genomes. Rather than relying on the selection and interpretation of a single gene phylogeny, a combined analysis of multiple gene trees might arrive at some broader consensus picture of the universal tree. This analysis could be extended to search for possible correlations between gene phylogenies, cellular functions, and suggested species relationships. Later, when more genomes have been sequenced, the spatial arrangement of genes might also be considered.

The remainder of this review will summarize the interdomain relationships implied by different universal protein gene trees. The phylogenies of some of these proteins have been well studied previously, while others are shown here for the first time (to the best of our knowledge). The discussion of these gene trees will be structured around the generalized metabolic pathways in which the encoded enzymes participate. Riley (277) provided a useful framework for the biochemical classification of gene products in *E. coli*, which will be very loosely adapted here. However, as Riley herself points out, the assignment of certain enzymes to particular functional categories can be arbitrary. For instance, carbamoyl phosphate synthase serves in both pyrimidine and arginine biosynthesis, and so this gene phylogeny is arbitrarily included with those of other amino acid biosynthetic enzymes largely out of convenience. Already discussed are two gene phylogenies, those of ATPase subunits, categorized as ATP-proton motive force interconversion by Riley (277), and HSP70, considered here as a chaperonin while Riley designated *dnaK* a component of DNA replication, restriction-modification, recombination, and repair.

Phylogenetic analyses were conducted only on gene sequences judged to be homologous rather than paralogous or analogous (217). The elongation factors, ATPase subunits, and aminoacyl-tRNA synthetases are all examples of paralogous gene families. Here, interdomain distances and gene trees were determined only among orthologous genes within these larger families of genes. An example of analogous genes are the DNA ligases, which are known for the *Archaea*, *Bacteria*, and eukaryotes and perform similar functions in each organism. However, bacterial ligases show no meaningful sequence similarities to archaeal or eukaryotic ligases (193), thus rendering it impossible to use these genes to derive meaningful evolutionary relationships among the three domains.

The GenBank, SwissProt, and PIR databases were initially searched (up to January 1996) for all archaeal entries (which were the most limiting); this was followed by downloading all possible bacterial and eukaryotic orthologs. Excluded were tRNA and rRNA genes, gene families for which eukaryotic versions only exist in organellar genomes, and partially sequenced genes. Of course, with the generation of new sequence entries proceeding at an ever-accelerating pace, such a survey can only aspire to be a static picture or "snapshot" of the global database at one particular moment. For example, shortly after most of these gene phylogenies were compiled, the complete genome sequence of the thermophilic methanogen *M. jannaschii* was released (49), and during the revision of this paper, the publication of the *Archaeoglobus fulgidus* genome (161a) was pending. For the most part, these new sequences were not included in the analysis, although the number of gene families available to derive universal trees has been dramatically increased. However, it is unlikely that these new archaeal sequence data would have greatly changed the presented gene tree topologies, since species of *Euryarchaeota*, in particular methanogens, are widely represented.

Considered here are the phylogenies of over 60 universal, orthologous proteins which are represented by over 1,200 different sequences. Of these proteins, 21 are complexed with the ribosome. Despite the large size of this data set, there were some serious limitations with respect to species diversity. Less than 15% of the sequences are from the *Archaea*. Few proteins are available from protist groups that branch deeply in eukaryotes, such as Metamonda, Microsporidia, and Parabasalia (Fig. 1). In addition, the bacterial sample is biased toward the best-known species, namely, *E. coli*, *Bacillus subtilis*, *Haemophilus influenzae* (117), and *Mycoplasma genitalium* (124). The degree of representation from evolutionarily diverse groups is an important consideration in the interpretation of any "global" phylogenetic analysis.

**Phylogenetic methods.** Prior to phylogenetic analyses, amino acid sequences from orthologous genes were aligned by the program MULTALIN (72) with the BLOSUM62 amino acid substitution matrix (153) and a gap penalty of 20. As necessary, sequence alignments were visually edited with previously published alignments as guides, and all gap positions were removed prior to phylogenetic analyses. To estimate the evolutionary distance, pairwise distances between all taxa were calculated with the program PROTDIST (from the PHYLIP 3.57 package [114]), which estimates the number of expected amino acid replacements per site by using a model based on the Dayhoff PAM substitution matrix (86). The resultant distance matrix was then used to draw a neighbor-joining tree (with the program NEIGHBOR). The statistical confidence for each node in a particular tree was estimated by bootstrapping, which involves the generation of 100 multiple random subsets of the alignments with the program SEQBOOT, and subsequently recalculating the distance matrices and neighbor-joining trees. In the figures, only node frequencies greater than 50% are reported.

The gene trees depicted here are only rough approximations of species relationships. More complete and precise phylogenetic analysis would entail a greater number of bootstrap replications and use a variety of different methods such as maximum parsimony and maximum likelihood. Furthermore, the rooting of any universal tree based on a single gene can only be implied—usually in the longest branch leading to the most distantly related domain. However, a midpoint rooting is not necessarily the correct one. For example, isoleucyl-tRNA synthetases of the *Archaea* and *Bacteria* have slightly higher sequence similarity, but with the inclusion of paralogous valyl-

and leucyl-tRNA synthetases as the outgroups, the rooting of the universal isoleucyl-tRNA synthetase tree is firmly in the *Bacteria* (43). For ATPase subunits and elongation factor gene phylogenies, both midpoint and paralogous rooting methods agree in the placement of the root in the *Bacteria*. Therefore, we stress that the distance trees shown here are highly provisional, and, where possible, we defer to studies involving more extensive analyses.

**Interdomain distance estimates.** Mean interdomain distances for each protein (*Archaea* to eukaryotes, *Archaea* to *Bacteria*, and eukaryotes to *Bacteria*) were determined by averaging the pairwise distances (calculated by PROTDIST) between all available sequences from species of different domains (see Table 1). Nucleus-encoded yet organelle-targeted isoforms which phylogenetically clustered with species of either proteobacteria or cyanobacteria (for example, many ribosomal proteins) were considered to be members of the *Bacteria* when the mean interdomain distances were calculated. For GS, prokaryotic GSI isoforms were compared to eukaryotic GSII isoforms while bacterial GSII versions were omitted. This decision was based on the best representation of major groups in the phylogeny, since bacterial GSII isoforms are fewer and generally coexist with GSI isoforms in the same organism. A somewhat more arbitrary choice was made for GDH, where eukaryotic type I and bacterial and archaeal type II isoforms were selected. However, both respective isoforms of GS and GDH were considered in evaluations of the domain coherence (monophyly). A simple nonparametric analysis of variance (ANOVA) test was used to judge whether any interdomain comparison was significantly ($P < 0.05$) smaller than the other two possibilities (herein, these trees will be called significant). Nearly always, comparisons of distances measured from either of the two most related domains to the outgroup were not significantly different.

## DNA Replication and Repair Enzymes

Prokaryotes and eukaryotes have fundamental differences in chromosome structure and organization which, in turn, affect the mode of DNA replication. Eukaryotes have multiple DNA replication initiation sites, while bacteria apparently have only a single origin, designated *oriC*. The origin of DNA replication has yet to be confidently identified in an archaeon, but overall similarities to bacterial chromosomes suggest that a single replication origin is likely (122).

Although the process of strand elongation during DNA replication is functionally similar among the *Bacteria*, *Archaea*, and eukaryotes, surprisingly few DNA replication proteins are homologous across all three domains (99). Several replication proteins are homologous among the *Archaea* and eukaryotes but have no counterpart in *Bacteria*. The *Methanococcus jannaschii* genome revealed several putative homologs to proteins involved with the replication factor complex (*rfc*) in eukaryotes (49). Two genes possibly encode proteins homologous to a protein associated with replication initiation in yeast. Another *M. jannaschii* protein, first identified in *Sulfolobus solfataricus*, corresponds to *pelota*—a protein known from *Drosophila* to be involved in the early stages of meiosis and mitosis (267). Flap endonuclease I (FEN-1) and RAD2 are DNA repair enzymes encoded by duplicate genes in eukaryotes to which there are single gene homologs in the *Archaea* but no bacterial versions (92).

In eukaryotes, the paralogs Dmc1 and Rad51 are similar to bacterial RecA proteins, which serve in homologous recombination, DNA repair, and the SOS response (37). Archaeal homologs to eukaryotic Dmc1 and Rad51, called RadA, have been found in the species *Haloferax volcanii*, *M. jannaschii*, *Methanobacterium thermoautotrophicum*, *Archaeoglobus fulgidus*, and *S. solfataricus* (92, 289). Recently determined was a second archaeal gene cluster separate from the Dmc1, Rad51, RadA, and RecA groups, which consisted of sequences from the species *M. jannaschii*, *M. thermoautotrophicum*, and *A. fulgidus*, as well as *Pyrococcus furiosus* and *Pyrococcus* sp. strain KOD1 (92, 270). Another highly divergent yet universal gene family encodes nucleoside triphosphate-binding proteins involved in chromosome condensation and DNA recombination and repair (104).

**DNA gyrases and topoisomerases.** The conformation of DNA has considerable importance in gene expression and genome compaction. In eukaryotes, DNA strands are negatively supercoiled while wrapped about histone-based nucleosome complexes but revert to a relaxed conformation once freed of DNA-binding proteins (reviewed in reference 122). The role of DNA-binding proteins in prokaryotic genomes is less well understood, but it appears that negative supercoiling is predominant. In both prokaryotes and eukaryotes, either positive or negative supercoils can be locally introduced into DNA strands during transcription.

The regulation of supercoiling in DNA strands is the role of specific enzymes called topoisomerases. Two major kinds of topoisomerase are found in the *Archaea*, *Bacteria*, and eukaryotes. Type I DNA topoisomerase is a monomer and is usually ATP-independent, whereas type II DNA topoisomerase is ATP dependent and multimeric. The two types differ by their mode of strand breakage prior to the crossing of two DNA strands: type I DNA topoisomerase catalyzes a transient single-strand break, while type II DNA topoisomerase performs a double-strand break.

Eukaryotic and prokaryotic type I DNA topoisomerases are functionally and structurally quite different. Eukaryotic type I DNA topoisomerases can relax either negative or positive supercoils by binding to the 3′ end of the DNA break. The two described type I topoisomerases from bacteria, *E. coli* protein ω (TOP2) and topoisomerase III, attach to the 5′ end of the DNA break and act only on negative supercoils. Eukaryotic 3′ DNA-binding and bacterial 5′ DNA-binding type I topoisomerases appear to be evolutionarily unrelated at the sequence level. However, recent studies have shown that bacterium-like 5′ DNA-binding type I topoisomerases exist in many archaea as part of the reverse gyrase (69) and in yeast as the enzyme TOP3 (326).

Reverse gyrases are unusual in that they generate positive supercoiling and appear restricted to thermophilic species of either the *Archaea* (such as sulfur-dependent thermophiles and thermophilic methanogens [35]) or *Bacteria* (*Thermotogales* strains [34]). Forterre et al. have suggested that reverse gyrase activity is necessary for the stabilization of the genome at high temperatures, although the exact mechanistic principles are not understood (121, 122). The reverse gyrase gene from *Sulfolobus acidocaldarius* appears to be a gene fusion involving a helicase N terminal followed by a type I DNA topoisomerase (69). The helicase-like motif might account for the reverse gyrase being ATP dependent while other type I DNA topoisomerases are ATP independent. The alignment and phylogenetic tree of various 5′ DNA-binding type I topoisomerases both suggest that the topoisomerase domain of reverse gyrase is most similar to bacterial protein ω, although all four type I topoisomerases are highly divergent (120, 121).

Although eukaryotic type II DNA topoisomerases function differently from bacterial and archaeal versions, the enzymes of all three domains show significant similarity in amino acid sequence (121). Prokaryotic type II DNA topoisomerases gen-

erally show gyrase activity which, in the presence of ATP, introduces negative supercoils into DNA. An exception is a type II DNA topoisomerase in *E. coli*, called topoisomerase IV, which lacks gyrase activity. The gyrase activity of type II DNA topoisomerases probably plays an important role in maintaining the negative supercoiling of bacterial genomes. Bacterial gene expression is known to be modulated by supercoiling, which is regulated by the opposing activities of topoisomerase I and DNA gyrase, with the fine control of the latter enzyme being affected by available pools of ATP (95).

Mutations in the DNA gyrase B gene convey novobiocin resistance in bacteria and at least one archaeon, a *Haloferax* sp. (159). Sequence comparisons and phylogenetic analysis show that the bacterial and archaeal type II topoisomerases are more closely related to each other than to their eukaryote homologs. However, *E. coli* type IV and bacteriophage T4 type II topoisomerases do not group with either bacterial/archaeal or eukaryotic isoforms (121) (Fig. 6A). In both *Haloferax* and gram-positive bacteria (*B. subtilis*), genes encoding gyrases A and B are cotranscribed, but in *E. coli*, they are unlinked.

However, archaea also have eukaryote-like DNA topoisomerases. Type II topoisomerase enzymes purified from *Sulfolobus acidocaldarius* (183) and *S. shibatae* (28) are reminiscent of eukaryotic type II topoisomerases in that they lack DNA gyrase activity, are refractory to novobiocin, and show sensitivity to a number of eukaryote-specific inhibitors. These archaeal genes have yet to be sequenced, which might prove interesting. A topoisomerase reported from the hyperthermophilic methanogen *Methanopyrus kandleri* appears biochemically similar to eukaryotic type I DNA topoisomerases (300).

**DNA polymerase B.** DNA polymerases are important, universally distributed enzymes involved in DNA replication and repair. There are five known eukaryotic (α, β, δ, ε, and γ) and three bacterial (I, II, and III) DNA polymerases (122). Replicases are the specific DNA polymerases involved with strand elongation from the DNA replication fork and are multimeric in structure. Bacteria have a single replicase, DNA polymerase III, while eukaryotes have several replicases specific to either nuclear (α, δ, and ε) or mitochondrial (γ) genome replication. The remaining eukaryotic and prokaryotic replicases are monomeric units involved with either DNA strand repair or the replication of short DNA strands, such as the extension of Okasaki fragments by *E. coli* DNA polymerase I.

The DNA polymerases have been classified into four families (A, B, C, and X) on the basis of similarities in sequence and drug sensitivities. Family A polymerases includes only the bacterial DNA polymerase I. Family B polymerases are sensitive to aphidicolin and have several highly conserved short amino acid motifs. Bacterial replicase DNA polymerase III form family C enzymes, while family X consists of DNA polymerase β, which is specific to eukaryotes. Polymerases of families A, B and C all share 3′ to 5′ exonuclease activity and three short consensus sequences, called exo boxes. However, little sequence similarity is found elsewhere among the three polymerase families.

Family B DNA polymerases are the only ones to have been found in species from all three domains. The eukaryotic family B polymerases are replicases, while the *E. coli* polymerase appears to be a repair enzyme. Surprisingly, DNA polymerase B has not been found in any other bacterium, including *H. influenzae* (117), and *M. gentialium* (124). Initial studies showed that DNA replication in species of the *Archaea* could be inhibited by aphidicolin, suggesting that archaeal and eukaryotic DNA replicases might be closely related (122). Due to commercial biotechnology interests, DNA polymerases have been cloned and sequenced from three different species of the hyperthermophilic genus *Pyrococcus*. These and other known archaeal DNA polymerases appear to be similar to eukaryotic family B types (36). In *Sulfolobus solfataricus*, there are two or perhaps three distinct but closely related family B DNA polymerases, which is suggestive of past gene duplication events (260).

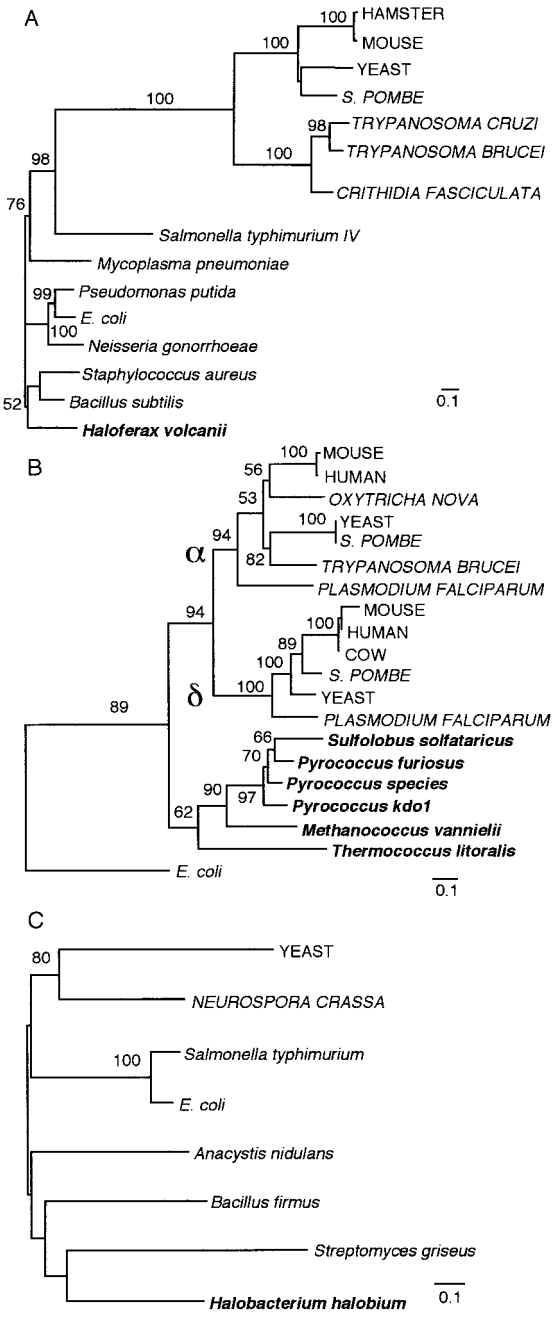


FIG. 6. Phylogenies for the gyrase B/topoisomerase II (A), family B DNA polymerase (the α and δ groups of eukaryotes are indicated) (B), and photolyase (C) genes.

There is evidence for additional DNA polymerases in the *Archaea*. Aphidicolin-insensitive DNA polymerases (therefore not of family B) have been detected in various archaeal species (reviewed in reference 49), although the genome of *Methanococcus jannaschii* coded for only a single family B DNA poly-

merase (49). The activities of two different DNA polymerases, one of which is sensitive to aphidicolin while the other is not, have been documented in *Halobacterium halobium* (305). Thus, it is not clear whether family B polymerases are the principal replicase in all members of the *Archaea*.

Given the great sequence divergence among family B DNA polymerases and the presence of only a single bacterial representative, DNA polymerases do not make a robust universal tree (Fig. 6B). The highest sequence similarities occur over eight specific regions which include three exo motifs (256), but fewer than 200 amino acid positions can be aligned with confidence across domains. The DNA polymerase B gene tree shows monophyletic clusters for the *Archaea* and eukaryotes (36, 260). Certainly, the evolution of DNA polymerases requires further study, in particular with respect to determining the full range of types and their functions in all three domains.

**Photolyase.** Central to photoreactivation repair of double-stranded DNA is the enzyme photolyase, which catalyzes the monomerization of UV-induced pyrimidine dimers. Class I photolyases of eukaryotes (yeast and *Neurospora*) and proteobacteria (*E. coli* and *Salmonella typhimurium*) employ a folate derivative cofactor, 5,10-methylenyltetrahydrofolate, which optimizes enzyme activity at a wavelength of about 380 nm (340, 341). Photolyases of gram-positive bacteria, cyanobacteria, and archaea use an entirely different cofactor, 7,8-didemethyl-8-hydroxy-5-deazaflavin, which bestows optimal activity at a higher wavelength, 440 nm. In *H. halobium*, photolyase is a 481-amino-acid peptide and the gene is located immediately upstream of the gene for superoxide dismutase (SOD) (316). Eukaryotic and proteobacterial photolyases are closely related at the sequence level, which suggests a recent shared ancestry. Class II photolyases, which are distantly related to class I photolyases, have been recently confirmed to exist in all three domains as well (341).

In the phylogenetic tree shown in Fig. 6C, only the branching points separating eukaryotes and proteobacteria show strong statistical support, while the clustering of species on the basis of secondary cofactors is only weakly supported. However, overall sequence similarities and common cofactors suggest that lateral gene transfer and replacement between proteobacteria and eukaryotes might have occurred, perhaps as a consequence of some early endosymbiosis (as discussed below). The HDF-type photolyases known from two gram-positive bacteria, a cyanobacterium and a halophilic archaeon, group together, although the node is not statistically strong, leaving the branching order unresolved.

## Transcriptional Proteins

Differences between bacteria and eukaryotes in transcription are also profound. In *E. coli*, the RNAP holoenzyme consists of four main subunits, $\alpha_2$, $\beta$, $\beta'$, and $\omega$, with a single exogenous protein, factor $\sigma$, required for transcriptional activation. The promoter region is typically located $-10$ and $-35$ bp upstream from the transcriptional start site (reviewed in reference 102). Additional activators might be involved in enhancing promoter strength, but none of these proteins appear homologous to eukaryotic transcription factors (reviewed in reference 52).

Transcription in eukaryotes is considerably more complex. The typical eukaryotic promoter consists of a TATA box sequence located about $-30$ bp upstream of the transcriptional start nucleotide. The eukaryotic core RNAP does not contact the DNA template strand directly at this site; rather, the enzyme attaches to a specific transcription factor, TFIID, bound to the DNA strand. The fully assembled initiation complex of RNAP II consists of at least five transcription factors, TFIIA, TFIIB, TFIID, TFIIE, and TFIIF, with a sixth factor, TFIIS, binding to the RNA polymerase once strand elongation commences (RNAP I, II, and III have specific suites of transcription factors that are numbered accordingly). TFIID is a multimeric protein which includes the important TATA-binding protein (TBP). TBP is a general transcription factor insofar as it appears to be required for the initiation of transcription of all RNAP II-transcribed genes, including those without a recognizable TATA box, as well as genes transcribed by RNAP I and III.

Reiter et al. (275) demonstrated the importance of TATA box-like upstream sequences in the transcription of the *Sulfolobus* 16S/23S rRNA gene. Later, archaeal homologs of eukaryotic TFIIB, TFIIS, and TBP were identified. A TFIIB homolog in *Pyrococcus woesei* was discovered via a database search (249) and was subsequently cloned from *S. shibatae* (266). Later, the TBP major component of TFIID was found in *Thermococcus celer* (226) and *Pyrococcus woesei*, where it was elegantly demonstrated to be functional in transcription (283). More recently, it was demonstrated that yeast and human TBPs can substitute for native transcription factors in a cell-free archaeal transcription system (318, 331).

The ancestral eukaryotic and archaeal RNAP might well have been a type II homolog, since TBP, a general transcription factor for all eukaryotic RNAPs, also exists in the *Archaea* (reviewed in references 187, 207, and 345). However, 5′ capping, persistent poly(A) tailing, monocistronic mRNAs, and spliceosomal introns are still unique to eukaryotic mRNA; therefore, these features were probably derived after the eukaryote-*Archaea* divergence.

Eukaryotes have three types of RNAPs (called RNAP I, II, and III), while both bacteria and archaea have just one kind (reviewed in reference 207). Eukaryotic RNAPs are distinguishable from one another on the basis of genes transcribed and reactions to the compound $\alpha$-amanitin. Eukaryotic RNAP I transcribes 5.8S, 18S, and 28S rRNA and is refractory to $\alpha$-amanitin. RNAP II produces all its mRNA from protein-coding genes as well as certain small nuclear RNAs (snRNA), and it is highly sensitive to $\alpha$-amanitin inhibition. RNAP III transcribes the remaining snRNAs, tRNAs, and 5S rRNA and shows an intermediate blockage response to $\alpha$-amanitin. In addition, there are other types of RNAPs specific for organellar gene transcription.

Zillig and co-workers have extensively studied the structure, function and evolution of archaeal RNA polymerases. All three eukaryotic polymerases are more structurally similar to archaeal homologs than to bacterial ones (190, 264). The holoenzyme complex of the *Bacteria* consists of just four subunits, while eukaryotic and archaeal RNAPs consist of seven or more homologous subunits (139, 190, 207, 212). The two largest archaeal subunits, B and A, correspond to the RpoB ($\beta$) and RpoC ($\beta'$) subunits in *E. coli* and the RpoB and RpoA subunits of eukaryotic RNAP II, respectively. In all archaea, the second largest subunit, A, is split into two parts, called A′ and A″ (344, 345). In *Methanococcus*, *Halobacterium*, and *Archaeoglobus*, the B subunit is also split into two parts (designated B′ and B″), while *Sulfolobus*, *Thermoplasma*, and *Thermococcus* have a unified B subunit (264).

Puhler et al. (263) from Zillig's group first compared the archaeal A′A″ subunit (then referred to as the AC subunit) with the A subunits of eukaryotic RNAP I, II, and III and bacterial $\beta'$ subunits. They showed that archaeal and eukaryotic RNAPs were overwhelmingly more similar to each other than either were to bacterial RNAPs. Furthermore, RNAP II and III were slightly closer to the archaeal polymerase while

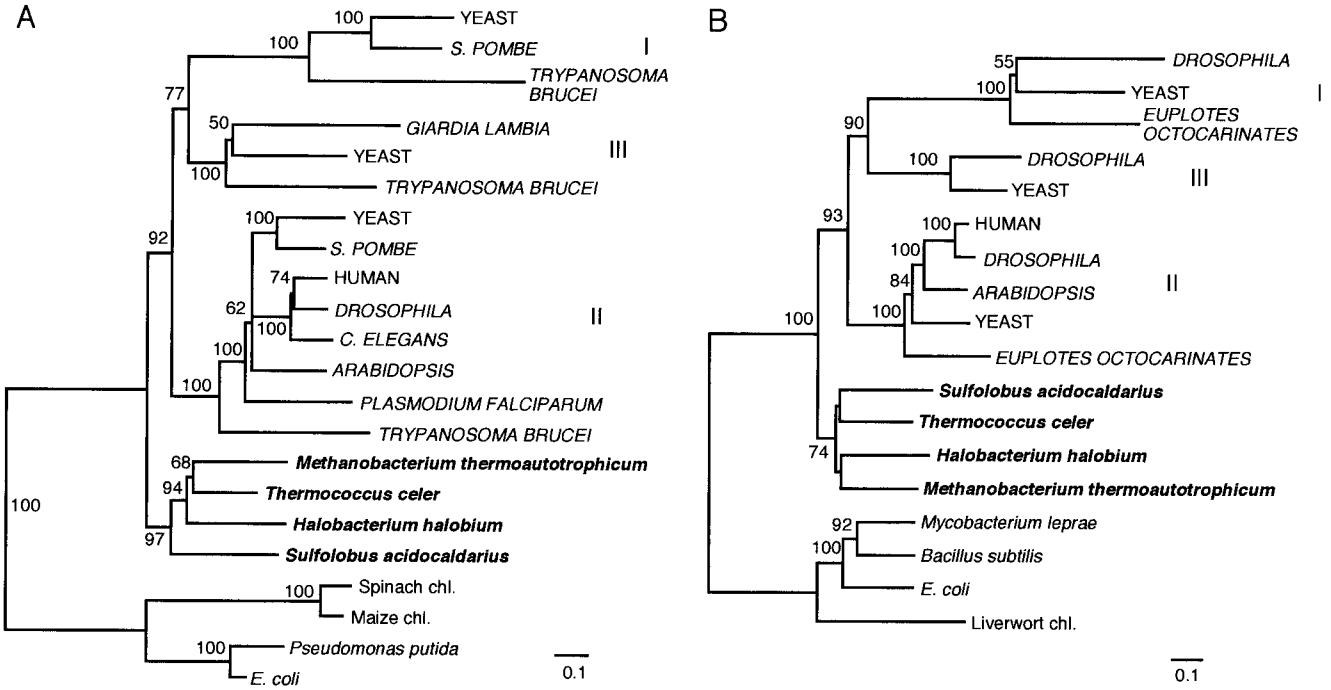


FIG. 7. Phylogenies for RNAP A subunit (A) and B subunit (B) genes from selected species. The three different eukaryotic types (I, II, and III) are indicated, as well as chloroplast-targeted (chl.) genes.

RNAP I "shared a bifurication" with the bacterial branch (although it was most similar at the sequence level to eukaryotic and archaeal RNAPs). A subsequent analysis from the same laboratory showed similar relationships in trees derived from the archaeal B′B″, bacterial β, and eukaryotic B subunits (188).

This observation of paraphyletic eukaryotic RNAP types formed the basis of Zillig's chimera hypothesis for the origin of eukaryotes. He suggested that the RNAP tree could be best explained if the eukaryotic cell was actually a chimera resulting from the fusion of a "prebacterium" (which contributed RNAP I) with a "prearchaeum" (the source of RNAP II and III [342]). However, Iwabe et al. (165) later reanalyzed the sequence data for both RNAP subunits and found that different phylogenetic methods produced different tree topologies. Maximum-likelihood and distance/neighbor-joining methods depicted the different RNAP types as a single monophyletic group in eukaryotes, while the maximum-parsimony method resulted in a paraphyletic spread of eukaryotic RNAPs. The latter result was attributed to more rapid sequence evolution among type I RNAP sequences. Iwabe et al. concluded that the strongest overall support was for eukaryotic RNAP I, II, and III forming a single clade. This would suggest a later gene duplication event, one that occurred well after the formation of the eukaryotic lineage, which led to the three types of polymerase.

Klenk et al. (188) further analyzed the two RNAP subunits, this time including a number of new archaeal and bacterial sequences, by using a "fuzzy-logic" approach which combined the results of several different tree-building methods. They concluded that there was sufficient statistical support for paraphyly among eukaryotic RNAP types. However, Klenk et al. (192) later analyzed several RNAP large subunits from lower eukaryotes and reported both paraphyletic and monophyletic eukaryote RNAP trees by the maximum-parsimony and distance methods, respectively. The RNAP phylogenies shown here concur with these early analyses by the neighbor-joining/distance method, where the eukaryotic types appear to be monophyletic (Fig. 7).

Regardless of controversies revolving around phylogenetic methodologies, a consistent result of RNAP trees is the early appearance of different polymerases in eukaryotes—a conclusion which was further supported by the discovery of a type III RNAP in *Giardia lambdia* (213). All phylogenetic analyses also support the monophyly of the *Archaea*, thus casting doubt on eocyte-like scenarios of eukaryotic origins. Phylogenetic analyses of RNAPs are particularly important since sequences are now available from archaeal species belonging to both *Euryarchaeota* and *Crenarchaeota*, as well as from deep evolutionary branches of the eukaryotes and *Bacteria*. RNAP sequences have been determined from the thermophilic bacteria *Thermotoga maritima* and *Aquifex pyrophilus*, whose rRNA sequences branch closest to the root of the *Bacteria* (188). Interestingly, the branching orders of species within the *Bacteria* and *Archaea* differ between RNAP and rRNA trees (188, 189). Klenk et al. (188, 189) propose that these differences might be the result of the high G+C nucleotide content found in the rRNA genes of thermophilic organisms, which might favor their placement at the base of the phylogenetic tree—a bias which they suggest is less important in phylogenetic reconstruction by using RNAP and other protein-coding genes.

### Translational Proteins

Both eukaryotes and the *Bacteria* use generally similar suites of catalytic enzymes, ribosomal proteins, and RNAs in protein synthesis. However, there are major differences with respect to the assembly and final structure of the ribosomal initiation complex (reviewed in reference 215). In the *Bacteria*, the small 30S subunit binds to purine-rich Shine-Dalgarno sequences located just upstream of the initiator AUG codon. The 30S

initiation complex is formed by the joining of three protein initiation factors, IF-1, IF-2, and IF-3, along with GTP and a formylated $tRNA^{Met}$. IF-3 is later released, allowing the 50S subunit to join the 30S initiation complex. The binding of the 50S subunit leads to the hydrolysis of GTP and the release of IF-1 and IF-2. The assembled 70S ribosome then proceeds with translation, while all the initiation factors are recycled for the next round of initiation. Of course, translation and transcription can be coupled in the *Bacteria*, something that is prohibited in eukaryotes by the additional processing required for mRNA maturation [addition of 5′ caps and poly(A) 3′ trailers and intron excision].

In the eukaryotes, the first step in ribosome assembly is the binding of the small 40S subunit to an initiator $tRNA^{Met}$, GTP, and a eukaryote-specific initiation factor, eIF-2. This 40S-$tRNA^{Met}$ complex then attaches to the capped 5′ end of the mRNA, just downstream of the translation start site. The 40S initiation complex then "scans" the mRNA for the first AUG codon. Once located, the 60S subunit joins the 40S complex, and the assembled 80S ribosome is then translationally active. Several initiation factors appear necessary for eukaryotic translational activation, although their exact number and respective functions are not clear.

The steps leading to the assembly of the archaeal ribosome are not well understood, although existing evidence points to a melding of bacterial and eukaryotic models (reviewed in reference 5). Like bacteria, archaeal mRNAs are not 5′-end capped, and some, but not all, genes have putative upstream Shine-Dalgarno initiation sites (48). The archaeal ribosome has rRNA components which are similar in number and sizes (23S, 16S and 5S) to those found in bacterial ribosomes. However, archaea lack N-formylated initiator Met-tRNA, and recognizable Shine-Dalgarno motifs are not found upstream of archaeal genes.

Similar to the situation in transcription, archaea appear to have several eukaryote-like translation initiation factors. In both *S. acidocaldarius* (177) and *M. jannaschii* (49), a putative IF-2 protein has been found that is more similar to yeast FUN12, a suggested eukaryotic initiation factor, than to bacterial IF-2 homologs. All IF-2 proteins appear to belong to a single gene family which, like EF-G/2, self-recycle GTP, and the two gene families appear to be paralogs (177, 178). These proteins belong to an even larger multigene family which includes EF-1α/Tu and eIF-2γ. However, the last two enzymes both require a guanine nucleotide exchange factor for GTP recycling. At least 11 different translation factor proteins have been identified from *M. jannaschii*, of which three match eukaryotic homologs (49). As in *Sulfolobus*, a hypusine-containing protein showing similarity to eukaryotic translation initiation factor eIF-5a (17) was found in *Methanococcus*, although recent experiments in yeast suggest that this protein is dispensable, thereby calling into question its role in translation (173).

Another archaeal translational component more similar to eukaryotic homologs is methionine aminopeptidase (MetAP), an enzyme which functions in all three domains to cleave the amino-terminal methionine from newly translated polypeptides (179). In phylogenies, MetAPs from *S. solfataricus* and *Methanothermus fervidus* (20) preferentially cluster with eukaryotic MetAP-2 types over those from the *Bacteria* (encoded by *MAP* genes) or eukaryotic MetAP-1 types. The latter eukaryotic version is suggested to have originated from the mitochondria, since MetAP-1 genes cluster strongly with homologs from the proteobacteria.

Comparative analyses of translational components have fueled several evolutionary controversies. Lake and coworkers argued for the eocyte universal-tree topology based on similarities in ribosome shapes among the *Crenarchaeota* and eukaryotes (152, 199–201). However, the different lobes and protubrences supposedly diagnostic of eocyte ribosomes were later found in species of halophiles and thermophilic methanogens (312, 313). The ribosomes of eukaryotes and the *Crenarchaeota* are also more protein rich than those of other members of the *Archaea* and *E. coli*. However, the bulking-up of ribosomes with proteins is more likely to be a general adaptation to high-temperature environments, since several thermophilic methanogenic archaea, as well as a hyperthermophilic bacterium, *Aquifex pyrophilus*, have ribosomes with high ratios of protein to rRNA (1).

**Elongation factors.** The orthologous genes EF-Tu (bacterial version) and EF-1α (archaeal and eukaryotic versions) act in translation to bring aminoacylated tRNA molecules into the A site of the ribosome. A second pair of orthologs, EF-G (in the *Bacteria*) and EF-2 (in the *Archaea* and eukaryotes), catalyze the subsequent release of deacylated tRNAs during ribosome translocation. EF-Tu/1α and Ef-G/2 are believed to be paralogous genes derived from an ancient duplication that occurred before the divergence of prokaryotes and eukaryotes. As discussed above, this gene paralogy was exploited by Iwabe et al. (164) to derive a rooted universal tree.

Since then, elongation factor genes have been sequenced from many other species. Cammarano et al. (53) constructed an EF-G/2 gene tree from new sequences from the archaea *Methanococcus vannielii*, *Halobacterium halobium*, *Sulfolobus acidocauldarius*, and *Thermoplasma acidophilum* and earlier data from a deep-branching bacterium, *Thermotoga maritima* (320). Their elongation factor trees concurred with rRNA phylogenies in finding strong support for the monophyly of domains with the bifurication of the kingdoms *Crenarchaeota* and *Euryarchaeota* internal to the *Archaea*. By adding EF-2 genes from *Pyrococcus woesei* and *Desulfurococcus mobilis* to their phylogenies, Creti et al. (75) showed that the *Archaea* was strongly monophyletic for EF-2/G genes but only weakly so for EF-Tu/1α genes. Hasegawa et al. (149, 150) originally reached similar conclusions after applying the maximum likelihood method for phylogenetic reconstruction. However, Hashimoto and Hasegawa (151) recently reported stronger support for paraphyletic archaea by using similar methods but a different alignment.

Rivera and Lake used an 11-amino-acid insertion, present in the *Crenarchaeota* and eukaryote EF-1α genes but not in *Euryarchaeota* EF-1α or *Bacteria* EF-Tu genes, to argue for the eocyte tree (278). Baldauf et al. (13) combined the EF-2/G and EF-Tu/1α data sets, derived a multiple-sequence alignment based on new crystallography data (2), and found weak statistical support in favor of the rooting described by Rivera and Lake (278). However, Creti et al. (77), in a reanalysis of their data set with a similar structure-based alignment, still found support for archaeal monophyly. Here, in a less rigorous treatment, gene trees for EF-Tu/1α and EF-G/2 (Fig. 8) show weak support for paraphyletic *Archaea* and fairly robust support for monophyletic *Archaea*, respectively. Thus, from the perspective of elongation factor genes, the question whether the origin of the eukaryotes can be directly linked to the *Crenarchaeota* remains open.

**Aminoacyl-tRNA synthetases.** In translation, the esterification or "charging" of a single amino acid to its cognate tRNA molecule is catalyzed by a specific aminoacyl-tRNA synthetase (abbreviated here as XxxRS, where Xxx is substituted for the three-letter code for a specific amino acid). The function and structure of aminoacyl-tRNA synthetases have been intensely studied, especially with respect to mechanisms of amino acid charging and tRNA specificity (reviewed in references 55 and

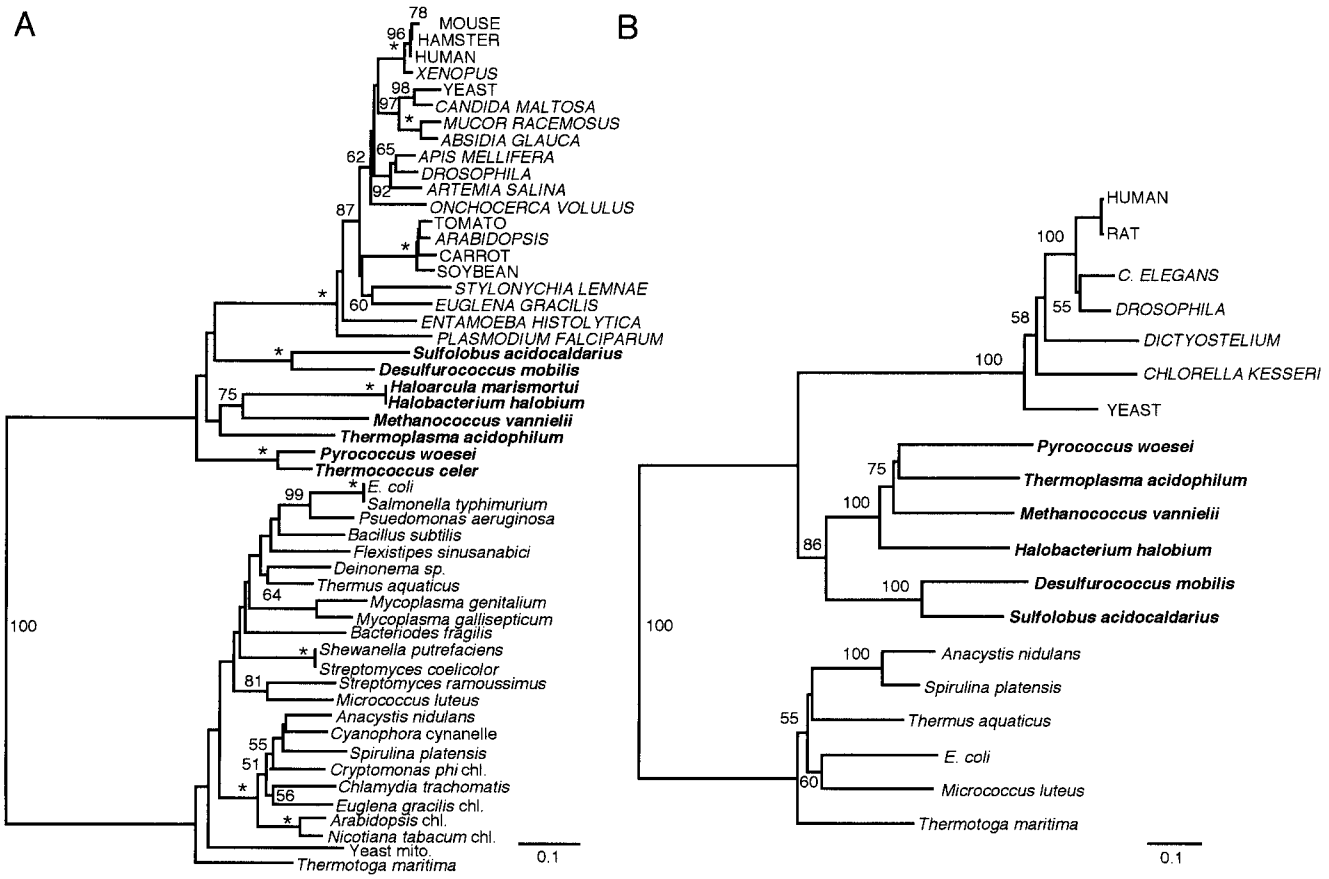


FIG. 8. Phylogenies for EF-Tu/1α (A) and EF-G/2 (B) genes from selected species. Some nodes found in 100% of bootstrap replications are indicated (*).

236). Aminoacyl-tRNA synthetases are variable in size, ranging from 334 (TrpRS) to 1,112 (PheRS) amino acids in *E. coli* (240). On the basis of structural and functional similarities, these enzymes are classified as belonging to either group I (specific for Glu, Gln, Trp, Tyr, Val, Leu, Ile, Met, Cys, and Arg) or group II (specific for Thr, Pro, Ser, Lys, Asp, Asn, His, Ala, Gly, and Phe) (106). All group I synthetases share two highly conserved amino acid motifs, HIGH and KMSKS, which are indicative of an ATP-binding structure. Group II synthetases lack this "Rossman fold" and the associated sequences. They have an entirely different structure and three other signature motifs, the most highly conserved of which has the general amino acid sequence, GLER. X-ray crystallographic studies have confirmed the presence of significant structural differences between the two groups (reviewed in references 55 and 105). The two groups of aminoacyl-tRNA synthetase are also functionally different; group I enzymes acylate the 2′ OH of the ribose of the final tRNA nucleotide, while group II enzymes charge the 3′ OH (106). Thus, despite having similar catalytic functions, group I and II aminoacyl-tRNA synthetases appear to be evolutionarily unrelated.

Nagel and Doolittle (240, 241) showed that all synthetases of either the group I or II type are related at the sequence level and that they therefore comprise two separate multigene families. Since bacterial and eukaryotic aminoacyl-tRNA synthetases with the same amino acid specificity nearly always clustered in phylogenetic trees, these investigators suggested that individual synthetases probably arose from a gene duplication predating the divergence of prokaryotes and eukaryotes. This apparent gene paralogy allowed the construction of universal trees based on IleRS sequences which were independently rooted by ValRS and LeuRS sequences (43).

Here, the unrooted universal trees for AspRS and IleuRS genes are shown (Fig. 9A and B). The branching topology of the IleRS gene tree differs slightly from that previously published, since other amino acids as well as those confidently aligned with ValRS and LeuRS genes could be included in the alignment. *Archaea*, *Bacteria*, and eukaryotes are still resolved as monophyletic clades. However, on the basis of comparative distances, archaeal species are, on average, slightly more similar to bacteria than to eukaryotes. A midpoint rooting would place the *Archaea* and *Bacteria* together, which is at odds with the rooting obtained with paralogous genes. This case illustrates that a midpoint rooting is not always the correct one and that any inferred rooting based on a single gene is highly provisional.

Given that there are several different aminoacyl-tRNA synthetase genes, multiple opportunities exist to derive rooted universal trees with either group I or II synthetases. Phylogenetic analysis of the group II synthetase AspRS suggests that *Pyrococcus* (161) and eukaryotic homologs are most similar. This relationship holds true even when the AspRS gene tree is independently rooted by using gene sequences of closely related paralogs, LysRS and AsnRS (42).

However, some studies have now called into question assumptions about the ubiquity and ancient gene paralogy of aminoacyl-tRNA synthetase. For example, the synthesis of Gln-tRNA$^{Gln}$ can occur either through the action of a specific

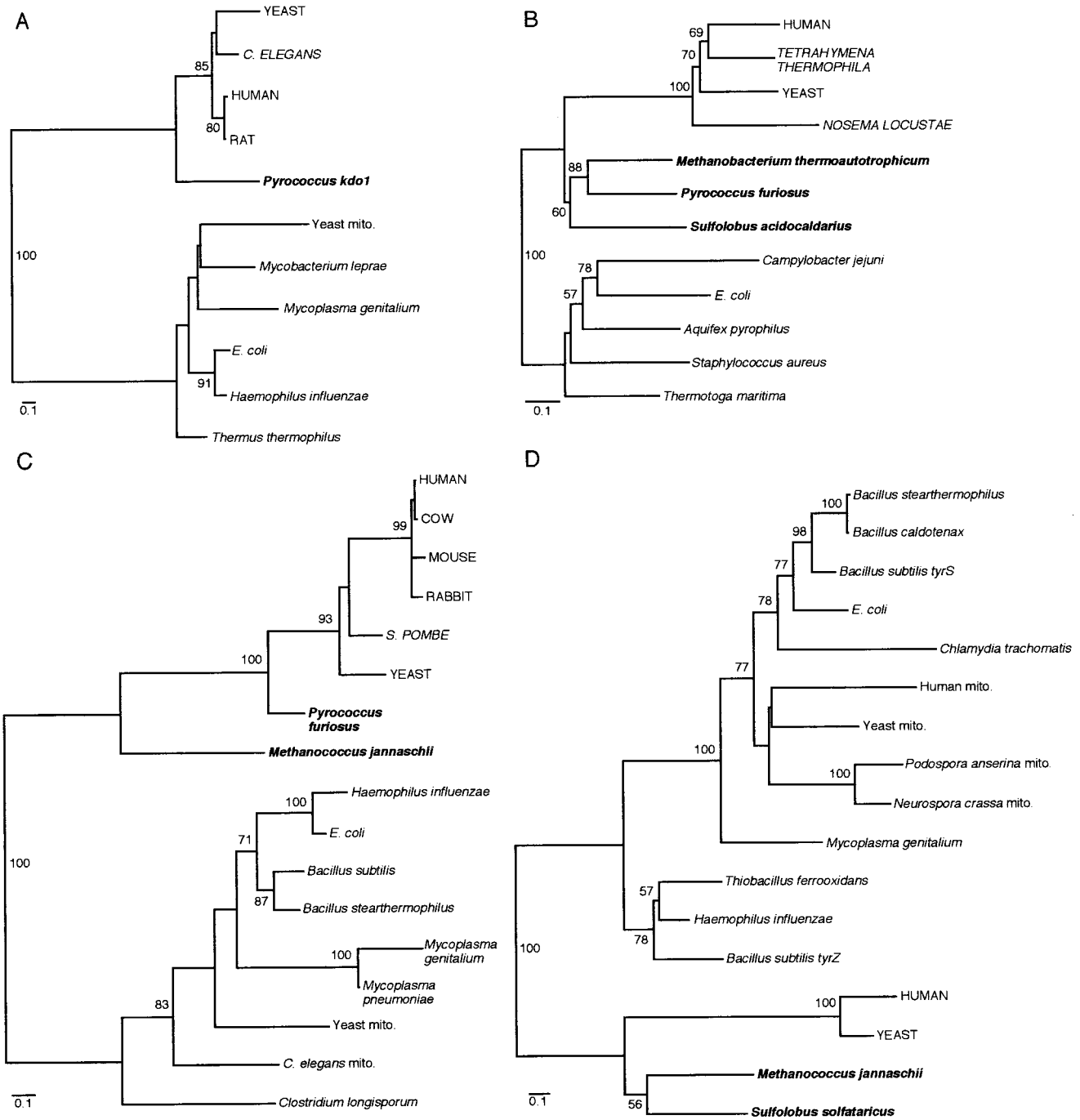


FIG. 9. Phylogenies for aspartyl-tRNA synthetase (A), isoleucyl-tRNA synthetase (B), tryptophanyl-tRNA synthetase (C), and tyrosyl-tRNA synthetase (D).

GlnRS or via the acylation of $tRNA^{Gln}$ by GluRS followed by transamidation of Glu to Gln. The latter pathway occurs in gram-positive bacteria, cyanobacteria, the proteobacterium *Rhizobium meliloti*, and *Thermus thermophilus*, as well as in mitochondria and chloroplasts, which apparently lack GlnRS (38, 126, 208). The GlnRS gene has been found in eukaryotes and two α-proteobacteria, *E. coli* and *H. influenzae.*

Phylogenies suggest that a duplication event in eukaryotes gave rise to their GluRS and GlnRS genes, and the latter copy was laterally transferred to α-proteobacteria (206). However, until recently, neither gene had been sequenced from an archaeon. Early biochemical studies suggested that transamidation of Glu to Gln also occurs in the *Archaea* (141, 332). Phylogenetic analyses of new archaeal GlxRS genes (since the precise charging function can not be assigned) show them to be closely related to, but branching outside of, eukaryotic GluRS and eukaryotic and proteobacterial GlnRS (44). Surprising, a similar transamidation pathway for the formation of $Asn\text{-}tRNA^{Asn}$ has been suggested to exist in *Haloferax volcanii* (79). Several aminoacyl-tRNA synthetase genes (the GlnRS, AsnRS, LysRS, and CysRS genes) could not be identified in the full genome sequence of *Methanococcus* (49).

A recent phylogenetic study of TrpRS and TyrRS genes by Ribas de Pouplana et al. (276) suggested that bacterial versions of either gene were more closely related to each other than to eukaryotic versions (TyrRS and TrpRS genes show significant sequence similarities; therefore, reciprocally rooted trees are possible). This would suggest some sort of independent evolution of either TyrRS, TrpRS or both proteins within the *Bacteria* and eukaryotes. However, archaeal versions were unavailable for their study. The inclusion of TrpRS sequences from the archaea, *Methanococcus jannaschii* and *Pyrococcus furiosus* and of TyrRS sequences from *M. jannaschii* and *Sulfolobus solfataricus* in the phylogenetic analysis produces quite different results (Fig. 9C and D). The two amino acid types of synthetases form separate monophyletic clades, which is in agreement with the conventional view of aminoacyl-tRNA synthetase gene evolution (46). Furthermore, for both the TyrRS and TrpRS portions of the tree, the *Archaea* and eukaryotes are sister groups, which confirms previous paralogous gene rootings of the universal tree.

**Ribosomal proteins.** The ribosome is a complex structure composed of three or four different RNAs and 50 to 90 different proteins. In *E. coli*, the ribosome consists of three major RNAs (5S, 16S, and 23S) and 55 ribosomal proteins (RP) (334). Names are given to new RPs based on their location, with either the large (L) or small (S) ribosomal subunits, and their homology to RPs in *E. coli* where the entire complement is known. However, this nomenclature is sometimes confusingly applied to eukaryotes and archaea, which have several additional RPs. The *Sulfolobus* 50S ribosomal subunit alone may have as many as 43 RPs (56), some which show little or no similarity to those in bacterial ribosomes (268). Furthermore, several archaeal RPs, such as RP L12, are shared with eukaryotes but not bacteria (reviewed in references 268 and 335). Conversely, a few RPs are found only in the *Bacteria* and *Archaea*, such as RP L1 located in the RP L10 gene cluster (12).

RPs themselves are short, highly conserved polypeptides (generally less than 300 amino acids) and thus are of limited use phylogenetically. However, as a large group of independently evolved yet physical-interacting proteins and RNAs, the ribosome is perhaps the best example of an ancient coadapted macromolecular complex. At the sequence level, the different RP genes appear evolutionarily distinct, with the possible exception of RPs L10 and L12, which share a conserved C-terminal motif (297). In eukaryotes, RP L12 apparently underwent a second gene duplication, since two distinct yet obviously related gene copies exist (218).

Although particular RPs are unique to only one or two domains, many proteins are homologous across the three domains. Previous phylogenetic analyses have been applied to RPs L2, L10, L11, L12, L15, L30, S8, S11, and S17 (9, 218, 335). A list of RP homologs was tabulated by Ramírez et al. (Table 2 in reference 268). When multiple alignments of RPs were assembled for the present review, it became apparent that certain proteins, although similarly labeled in the different domains, were clearly too distant to be considered homologs. For example, RP S3 sequences are similar between the *Archaea* and eukaryotes but not the *Bacteria*. Large interdomain distances were also estimated for "homologs" of RP L18. In total, 21 RPs were identified as reasonable choices for universal phylogenies. Several of these gene trees had very poor statistical resolution of internal nodes, while others provided fairly robust topologies. A total of 10 and 11 RPs were associated with the small and large ribosomal subunits, respectively.

As discussed above, RPs often occur in similarly organized operons in the *Bacteria* and *Archaea*. However, this was not reflected in either gene phylogenies or mean interdomain distance scores, where 18 of the 21 RP genes analyzed (for RPs L2, L3, L5, L6, L10, L14, L15, L22, L23, L30, S5, S7, S9, S10, S11, S12, S15, and S19) showed eukaryotes and archaea as sister groups (Fig. 10 through 13A to E). In some instances, archaeal and eukaryotic RPs shared N- or C-terminal extensions that were missing from bacterial homologs. The remaining three RPs (RPs L11, S8, and S17) showed archaea and bacteria as neighbors.

These findings generally agree with earlier phylogenetic studies with smaller data sets. Liao and Dennis (218) also found bacterial and archaeal RP L11 genes to be most similar whereas archaeal and eukaryotic RP L10 genes were most closely related. Wittmann-Liebold et al. (335) constructed dendrograms of RP L2 and RP S11 sequences which positioned eukaryotes and archaea as sister groups. Auer et al. (9) derived unrooted phylogenetic trees for RP L15, L30, S8 and S17 (as well as EF-Tu/1α), in which RP L15 showed eukaryotes and archaea together while major branching points were unresolved in the remaining RP gene trees. With few exceptions, the majority of RPs follow the evolutionary pattern of other protein synthesis gene which show the sisterhood of the *Archaea* and eukaryotes.

One might propose that the monophyly of the *Archaea*, *Bacteria*, and eukaryotes, the salient feature of rRNA trees, should be evident in the protein complement of the ribosome as well. However, with the exclusion of RPs where there was only a single species from a domain, 7 of 10 large-subunit and 3 of 7 small-subunit RP gene trees depicted the *Archaea* as paraphyletic with respect to eukaryotes. In these instances of paraphyly, where representative species of the *Crenarchaeota* and *Euryarchaeota* were both present, there was little to suggest that eukaryotes were consistently closer to a particular archaeal kingdom (sensu eocytes [199]). Phylogenetic analysis of RP L11 genes by Liao and Dennis (218) and here (Fig. 11A) suggests that *Sulfolobus* is the closest archaeon to the eukaryotes but with low bootstrap support. Conversely, the RP L10 gene tree has *Sulfolobus* as the farthest archaeon from eukaryotes. The RP S10 gene tree, which has five divergent archaea, shows the *Archaea* as monophyletic but, again, with low bootstraps (Fig. 12F).

## Enzymes in Central Metabolism

Hetrotrophic catabolism consists of two fundamental processes, the conversion of glucose and other sugars to pyruvate via the glycolytic pathway and the subsequent catabolism of pyruvate, either by anaerobic fermentation or by the aerobic citric acid cycle (CAC). The CAC also interacts with various amino acid synthetic and catabolic pathways as well as the chains of electron transport and oxidative phosphorylation. As such, the inherent processes of central metabolism are crucial to the viability of all living cells. The operation of central metabolic pathways in the *Bacteria* and eukaryotes have been well studied. Current knowledge of metabolic enyzmology in the *Archaea* has been reviewed by Danson (81–83) and, more recently, by Schönheit and Schäfer (294). Here, only metabolic enzyme genes characterized from at least one species each of the eukaryotes, *Bacteria*, and *Archaea* will be discussed. However, a brief review of archaeal metabolism might prove useful.

In eukaryotes and most bacteria, glycolysis proceeds along the Embden-Meyerhoff pathway, where the conversion of glucose to pyruvate results in the release of two molecules of ATP. Until recently, the activity of phosphofructokinase, a key Embden-Meyerhoff pathway enzyme, had not been detected in the

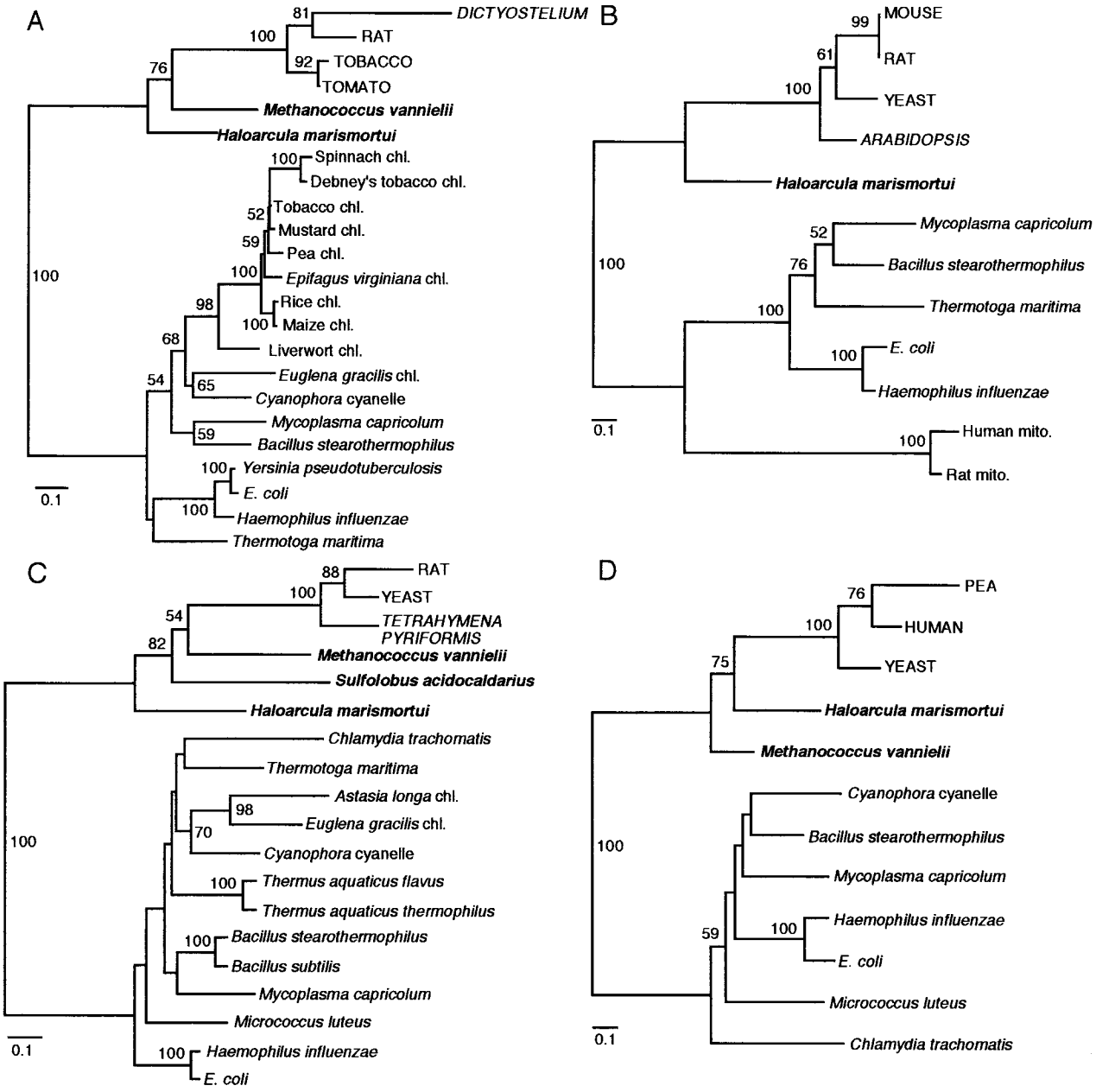


FIG. 10. Phylogenies for the RP L2 (A), RP L3 (B), RP L5 (C), and RP L6 (D) genes.

*Archaea*. This led to the assumption that the *Archaea* (and some bacteria) employed the Entner-Doudoroff pathway, an alternative mode of glycolysis which yields only a single ATP per glucose molecule (83). However, a novel hexokinase and phosphofructokinase dependent upon ADP, rather than ATP, has been found in *Pyrococcus furiosus* (182). Similarly, phosphofructokinase activity in *Thermoproteus neutrophilus* has been linked to pyrophosphate, while its hexokinase uses ATP as a phosphoryl donor (299). This modified Entner-Doudoroff pathway in thermophiles does not yield ATP molecules. Also in the pathway is a highly divergent form of triosephosphate isomerase recently found in *Methanothermus fervidus* and *Pyrococcus furiosus* (195). Thus, a modified Embden-Meyerhoff pathway might exist in these species alongside a nonphosphorylated Entner-Doudoroff pathway (292). There is some evidence for the presence of components of the Embden-Meyerhoff pathway in methanogens, mainly from the analyses of reverse reactions of the pathway leading to carbohydrate synthesis (81). However, methanogens are autotrophs, and rather than relying on glucose as a source of energy, they fix carbon directly as acetyl-CoA. Also, a third glucose metabolic route found in the *Bacteria* and eukaryotes, the pentose phosphate pathway, has been suggested to occur in several halophilic and thermophilic archaea (81).

The Embden-Meyerhoff, Entner-Doudoroff, and pentose phosphate pathways have several common catalyzed "trunk" reactions which synthesize pyruvate from glyceraldehyde 3-phosphate. Full-length gene sequences from the *Archaea* are known for a few of these enzymes, including glyceraldehyde-3-phosphate dehydrogenase (GAPDH), 3-phosphoglycerate

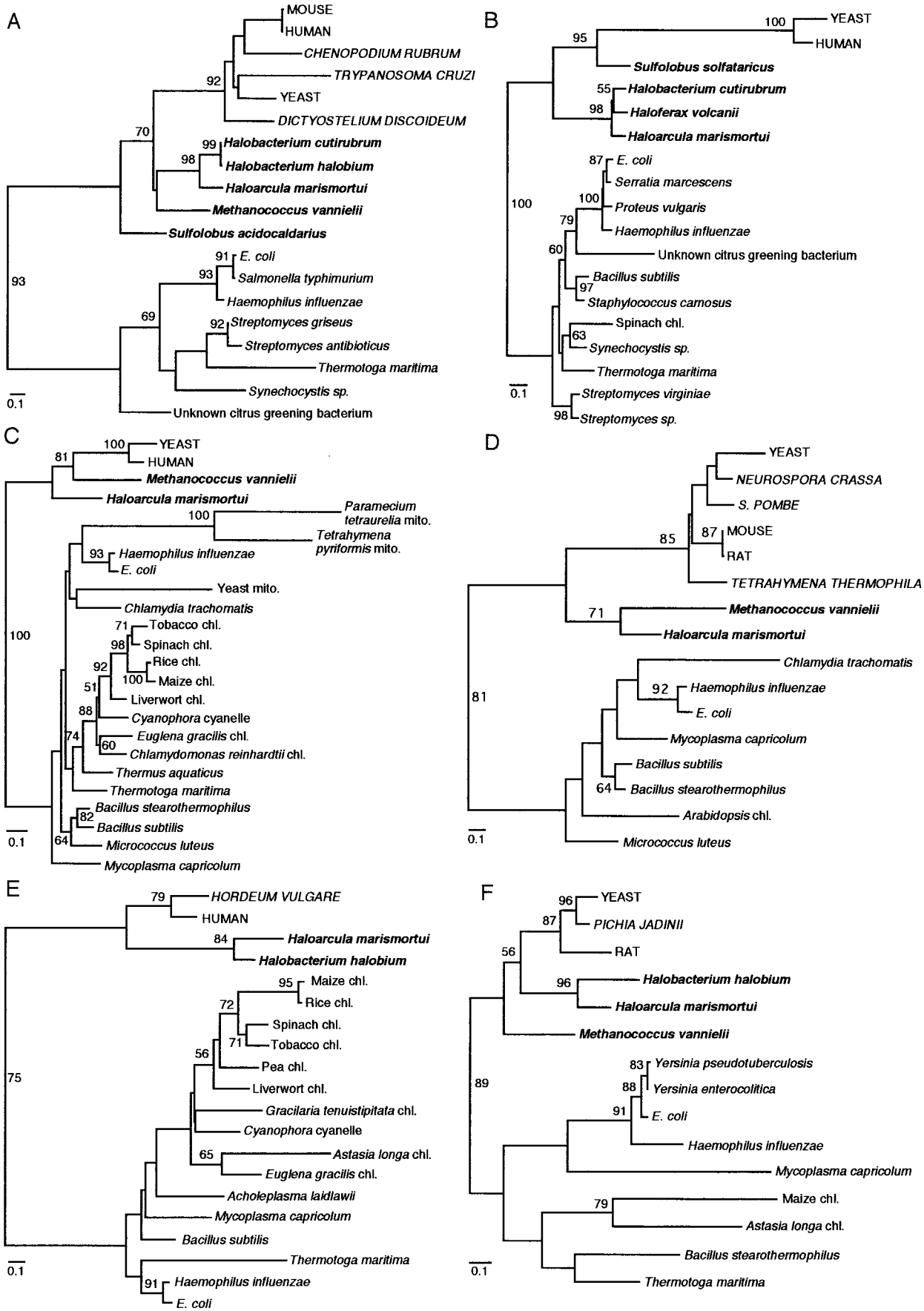


FIG. 11. Phylogenies for the RP L10 (A), RP L11 (B), RP L14 (C), RP L15 (D), RP L22 (E), and RP L23 (F) genes.

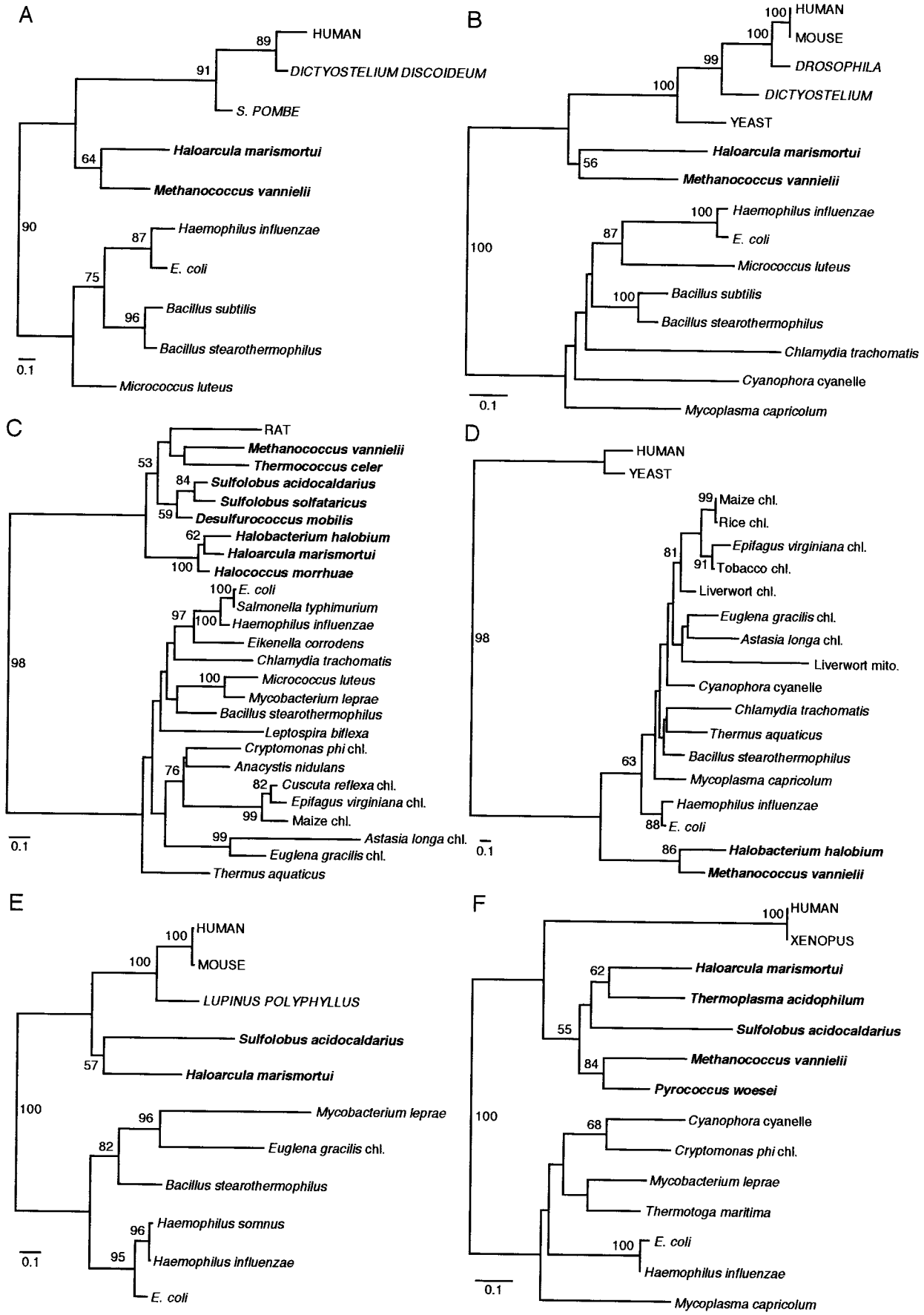


FIG. 12. Phylogenies for the RP L30 (A), RP S5 (B), RP S7 (C), RP S8 (D), RP S9 (E), and RP S10 (F) genes.

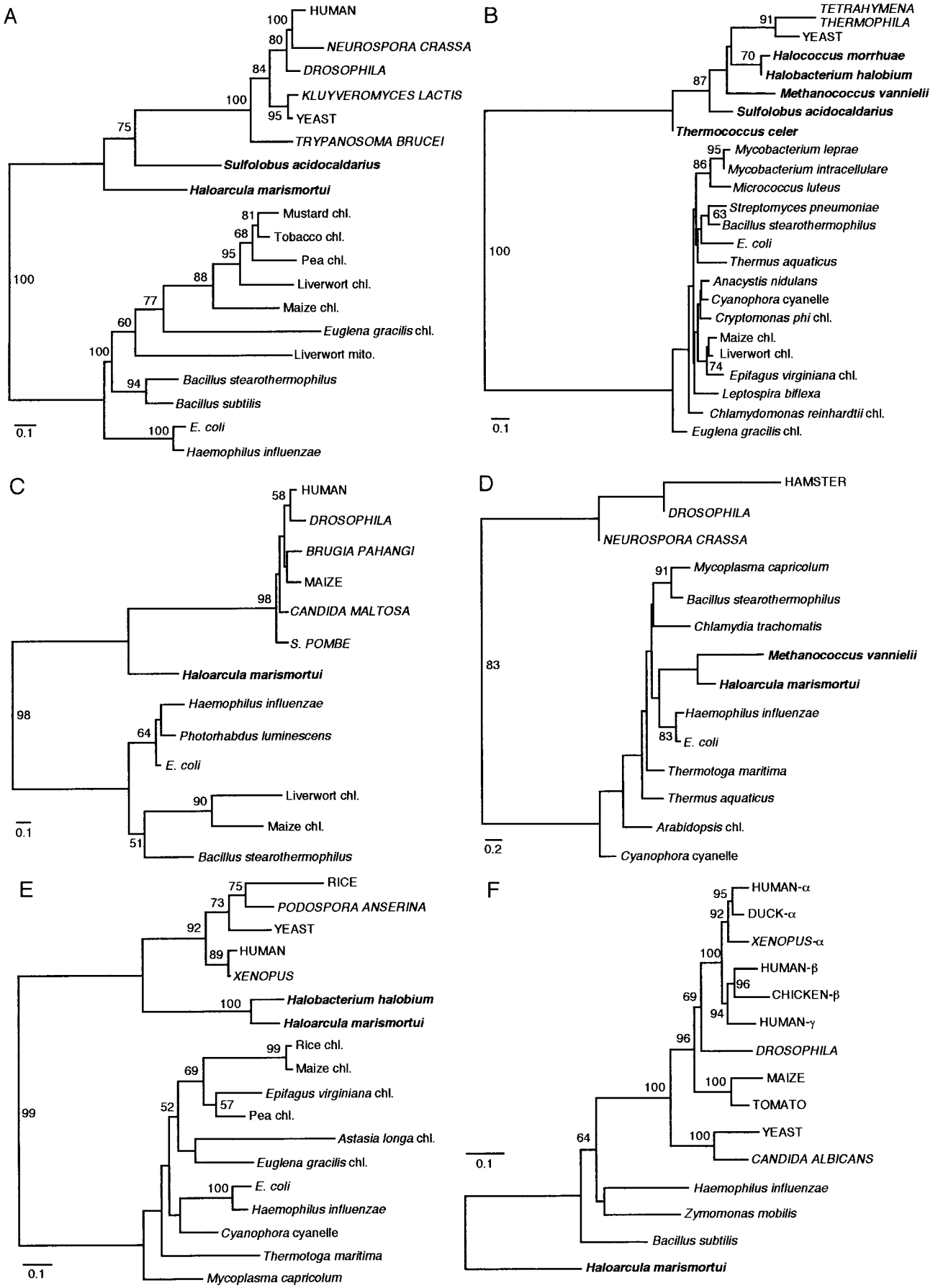


FIG. 13. Phylogenies for the RP S11 (A), RP S12 (B), RP S15 (C), RP S17 (D), RP S19 (E), and enolase (F) genes.

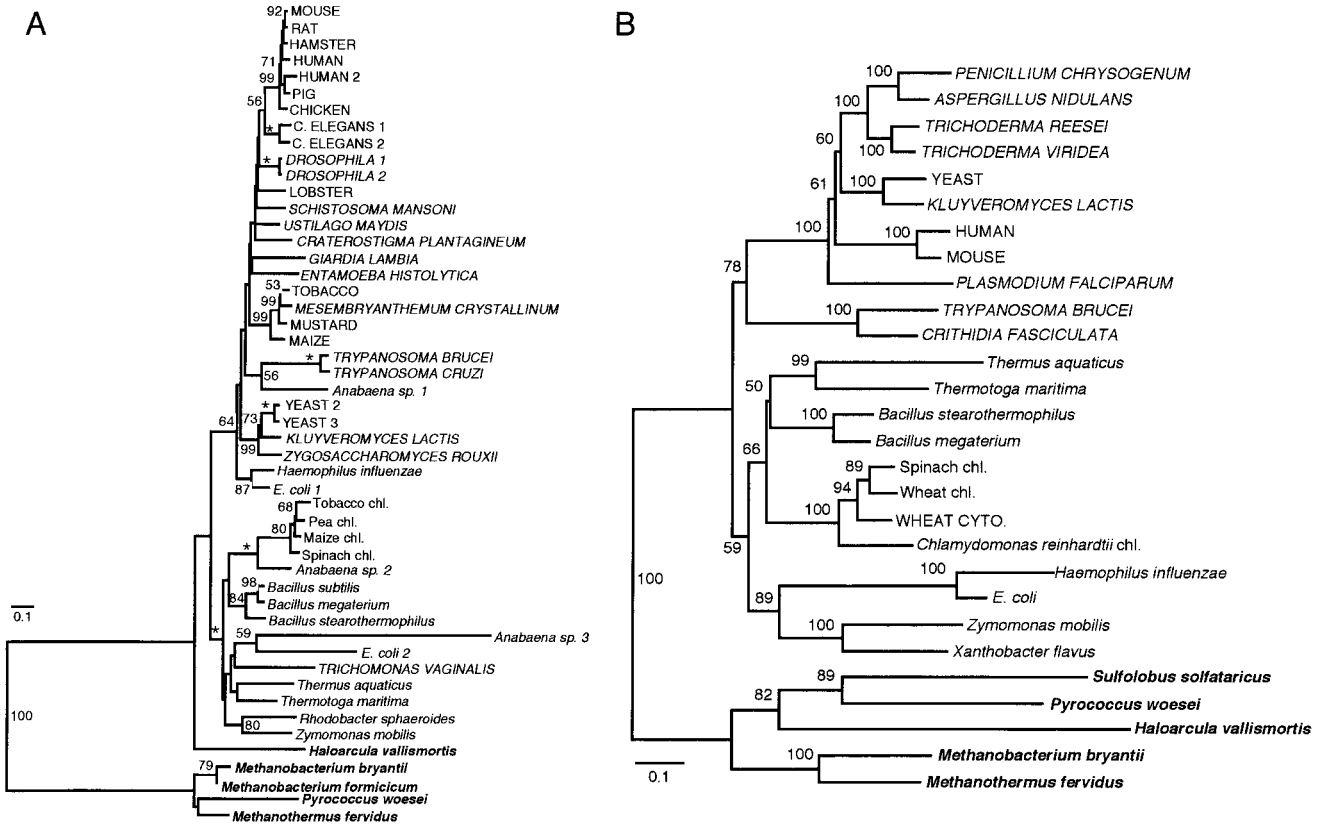


FIG. 14. Phylogenies for the GAPDH (A) and PGK (B) genes. For PGK, cytoplasmic (cyto.) and chloroplast-targeted (chl.) isoforms of wheat are indicated.

kinase (PGK), and enolase. Also, a small portion of the pyruvate kinase gene from *Thermoplasma acidophilum* has been sequenced (259). With respect to the involvement of nonphosphorylated intermediates, the Entner-Doudoroff pathway in *Sulfolobus*, *Thermoplasma*, and *Pyrococcus* differs from that in halophiles (81). In these thermophiles, equimolar amounts of pyruvate and glyceraldehyde are derived from the direct aldol cleavage of 2-keto-3-deoxygluconate. Glyceraldehyde is then oxidized to glycerate, which is subsequently converted into 2-phosphoglycerate by glycerate kinase. Thus, the primary roles of GAPDH and PGK might not be related to glucose metabolism.

Thus far, the oxidation of pyruvate to acetyl-coA is known to occur similarly in all members of the *Archaea*. Pyruvate oxidoreductases have also been found in the *Archaea* (and some anaerobic bacteria), although the electron receptors of enzymes from methanogens (which use a deazaflavin derivative) are different from those of halophiles and thermophiles (which use ferredoxin). In eukaryotes and the *Bacteria*, the analogous reaction is catalyzed by a multienzyme complex consisting of 2-oxoacid decarboxylase, dihydrolipoyl acyltransferase, and dihydrolipoamide dehydrogenase. Surprisingly, dihydrolipoamide dehydrogenase has been found in several archaeal species. Unlike in the *Bacteria* and eukaryotes, this enzyme is not part of a multimeric complex in the *Archaea*. The function of dihydrolipoamide dehydrogenase in the *Archaea* is not clear, although Danson (81) suggests a possible correlation between the presence of this enzyme and lipoic acid. Enzymatic assays have detected dihydrolipoamide dehydrogenase activity in members of the *Euryarchaeota* but not of the *Crenarchaeota.*

The CAC also exists in the *Archaea*, although it is specifically modified in different groups. Halophiles and *Thermoplasma acidophilum* have complete oxidative CAC, while *Sulfolobus* species exhibit only limited oxidative respiration under heterotrophic growth conditions. The reductive CAC has been observed in *Sulfolobus* spp. and *Thermoproteus* under autotrophic growing conditions, while partial oxidative and reductive CAC pathways are known to exist in methanogens (81).

In summary, central metabolism in the *Archaea* differs significantly from that of eukaryotes and most members of the *Bacteria*. Furthermore, among ecologically distinct *archaea*, there exist unusual embellishments in glucose-to-pyruvate conversion and the CAC. Our knowledge about archaeal metabolism is limited by the small number of metabolic genes sequenced from the *Archaea*, in particular the *Crenarchaeota*. Furthermore, new biochemical variants are often not easily identified through sequence comparisons alone.

**Glyceraldehyde 3-phosphate dehydrogenase.** The final stages of the Embden-Meyerhoff, Entner-Doudoroff, and pentose phosphate pathways share several similarly catalyzed reactions. Since the catalyzing enzymes are reversible, they can play a secondary role in gluconeogenesis. The first enzyme of these "trunk" sequences is GAPDH, which catalyzes the bidirectional conversion between glyceraldehyde 3-phosphate and 1,3-di (or bi)-phosphoglycerate in the presence of NADH.

Sequence analysis has shown that archaeal GAPDH genes are highly diverged from their bacterial and eukaryotic homologs (109, 111, 154, 347). Doolittle et al. (93) suggested that archaeal GAPDH genes might have been recruited from an entirely different gene family, possibly related to the $NAD^+$/NADH transhydrogenases. However, there may be multiple and divergent isoforms of GAPDH within the *Archaea*, as is

the case for several members of the *Bacteria* and eukaryotes (Fig. 14A). Indeed, a second class of GAPDH genes detected in the extreme halophile *Haloarcula vallismortis* is more similar to eukaryotic and bacterial homologs than to other archaeal GAPDH genes (228, 262).

Among bacterial and eukaryotic GAPDH genes, there are several instances of unexpected sequence similarities. *E. coli* is known to have two GAPDH genes. One copy, *GapB*, is similar to other bacterial GAPDH genes and occurs adjacent to genes encoding other glycolytic enzymes, 3-phosphoglycerate kinase and fructose 1,6-biphosphate aldolase (3). The second copy, *GapA*, is clearly similar to eukaryotic GAPDH genes (*GapC*), and was first believed to be eukaryotic in origin (93).

In 1993, Martin et al. (229) published three GAPDH gene sequences from the cyanobacterium *Anabaena variabilis*, one of which, *gap1*, was similar to eukaryotic *GapC* types. Of the two remaining isoforms, *gap2* was related to Calvin cycle enzymes from chloroplasts while *gap3* was similar to bacterial GAPDH genes. (The GAPDH nomenclature can be somewhat confusing. Martin et al. refer to *E. coli GapA* and *GapB* as *gap1* and *gap2*, respectively, while renaming the two subunits of the Calvin cycle GAPDH *GapA* and *GapB*. Thus, *Anabaena gap1* and eukaryotic *GapC* are directly orthologous to *E. coli gap1*, previously known as *GapA*.) In *Anabaena*, the product of *gap1* is probably utilized in glucose metabolism since it is located in the same operon as two other pentose phosphate cycle enzymes, pyruvate kinase and transaldolase. The divergence of these three cyanobacterial GAPDH genes is suggested to have occurred early in bacterial evolution. Martin et al. (229) postulated that multiple transfers of endosymbiotic GAPDH genes into the nucleus might have occurred, and perhaps *GapC* is an instance where the bacterial copy completely displaced the preexisting eukaryotic version—so-called gene replacements. These transfers must have occurred early in eukaryotic evolution as well, since GAPDH sequences of several deeply diverged protists, such as *Giardia* and *Entamoeba*, cluster with those from other eukaryotes. This also suggests that these amitochondrial protists might have witnessed lateral transfer of bacterial genes into the nucleus without the concomitant endosymbiosis leading to organelle formation (155). Such direct bacterial origins have also been suggested for two nonglycolytic genes characterized from *Entamoeba* (67).

However, it is not clear whether all eukaryotic GAPDH genes evolved from a single ancestor. Few branching points among eukaryotes have high bootstrap values, and the protists *Giardia* and *Entamoeba* branch much higher in the GAPDH tree than do other gene phylogenies. In our tree and that of Henze et al. (155), *Trypanosoma* GAPDH genes are highly divergent, being most similar to homologs from proteobacteria and *Anabaena*. In fact, *Trypanosoma brucei* contains two different GAPDH isoforms—one version is specifically active in the glycosome, an organelle-like microbody, while the other transits into the cytostol (235). In contrast, the GAPDH gene from the amitochondrial protist *Trichomonas vaginalis* convincingly branches with *Bacteria*-specific *GapB* (*gap2*) genes rather than with other eukaryotic *GapC* sequences (225). *Trichomonas GapC* might have had a bacterial origin separate from those that lead to GAPDH gene acquisitions in other eukaryotes. However, possible hidden gene paralogy among different GAPDH types cannot be excluded.

**3-Phosphoglycerate kinase.** In glycolysis, the enzyme PGK sequentially follows GAPDH by catalyzing the reaction 3-phosphoglyceroyl phosphate + Mg · ADP $\rightleftharpoons$ 3-phosphoglycerate + Mg · ATP. PGK-coding genes have been sequenced from several members of the *Bacteria*, eukaryotes, and *Archaea*, both *Crenarchaeota* and *Euryarchaeota*. As with GAPDH, archaeal PGK genes are highly diverged from those of the *Bacteria* and eukaryotes (39, 110). The archaeon *Haloarcula vallismortis* has a *gap-pgk* gene cluster, which also occurs in several bacteria, linked with the gene for the glycolytic enzyme triose-phosphate isomerase (*tpi*).

In PGK gene phylogenies, bacteria and eukaryotes largely form separate monophyletic yet closely related, groups (Fig. 14B). Overall sequence similarities suggest that eukaryotes obtained their nuclear PGK gene from the *Bacteria* (39). A separate gene transfer and replacement event might account for the unusual positions of nucleus-encoded PGK isoforms in wheat (*Triticum aestivum*) (221). Nucleus- and chloroplast-targeted wheat isoforms branch as a group within the *Bacteria*. Brinkmann and Martin (39) suggest a lateral gene transfer event, similar to that involving cyanobacterial *GapC*, was responsible for both plant PGK genes being highly similar to bacterial versions. However, additional sequences of PGK genes from cyanobacteria and plants are necessary to conclusively prove this hypothesis.

**Enolase.** The next step in glycolysis, the interconversion between 3-phosphoglycerate and 2-phosphoglycerate, is catalyzed by phosophoglycerate mutase. Although the activity of this enzyme has been detected in the *Archaea*, the gene has yet to be sequenced. The subsequent reaction, the conversion of 2-phosphoglycerate to phosphoenolpyruvate and $H_2O$, is facilitated by phosphoenolpyruvate synthetase or enolase. Enolase genes have been sequenced from only a few species of the *Bacteria* and only one species of the *Archaea*, *Haloarcula marismortui* (196). Phylogenetic analyses show eukaryotic enolase genes as being most similar to those of proteobacteria, although the only other bacterial enolase was that of *B. subtilis* (Fig. 13F).

The chromosomal location of the enolase gene (*eno*) is variable in prokaryotes. In *E. coli*, *eno* is part of an operon with *pyrG*, which also encodes CTP synthetase (329), while in *Zymomonas mobilis*, *eno* is a solitary transcriptional unit (51). In *B. subtilis*, a single operon encodes four glycolytic enzymes, PGK (*pgk*), triose-phosphate isomerase (*tpi*), phosphoglycerate mutase (*pgm*), and enolase (*eno*) (in 5′-to-3′ order) (216). In the archaeon *H. marismortui*, an entirely different operon exists, with the $tRNA^{Leu}$ gene; the genes encoding RPs HL29, L13, and S9; two unidentified ORFs; *eno*; and a third unknown ORF (which shows some distant similarity to the ORF encoding the vertebrate laminin receptor protein [196]). The RP L13 and S9 genes also occur as a two-gene cluster in *E. coli* but are located more than 0.5 Mbp away from the *pyrG-eno* gene operon.

**Acetyl-CoA synthetase.** The activation of free acetate to acetyl-CoA in the presence of ATP is essential for several biochemical pathways including lipid biosynthesis and the oxidative CAC. In bacteria, eukaryotes and a few archaea (*Thermoproteus neutrophilus* and some methanogens [81]), this is accomplished with the assistance of acetyl-CoA synthetase. This ATP-dependent (AMP-forming) acetyl-CoA synthetase is not to be confused with ADP-forming acetyl-CoA synthetase. The latter enzyme has been found in some archaeal halophiles and thermophiles, as well as in anaerobic protists, where it produces ATP from the hydrolysis of acetyl-CoA to acetate.

Gene sequence data for ATP-dependent acetyl-CoA synthetase are limited and are known for only one archaeal species, the methanogen *Methanothrix soehngenii* (101). Overall, the bacterial and archaeal sequences show the greatest similarity to *M. soehngenii* and *B. subtilis*, a gram-positive bacterium (Fig. 15A).

**Citrate synthase.** Citrate synthase facilitates the entry of carbon as citrate into the CAC through the condensation of

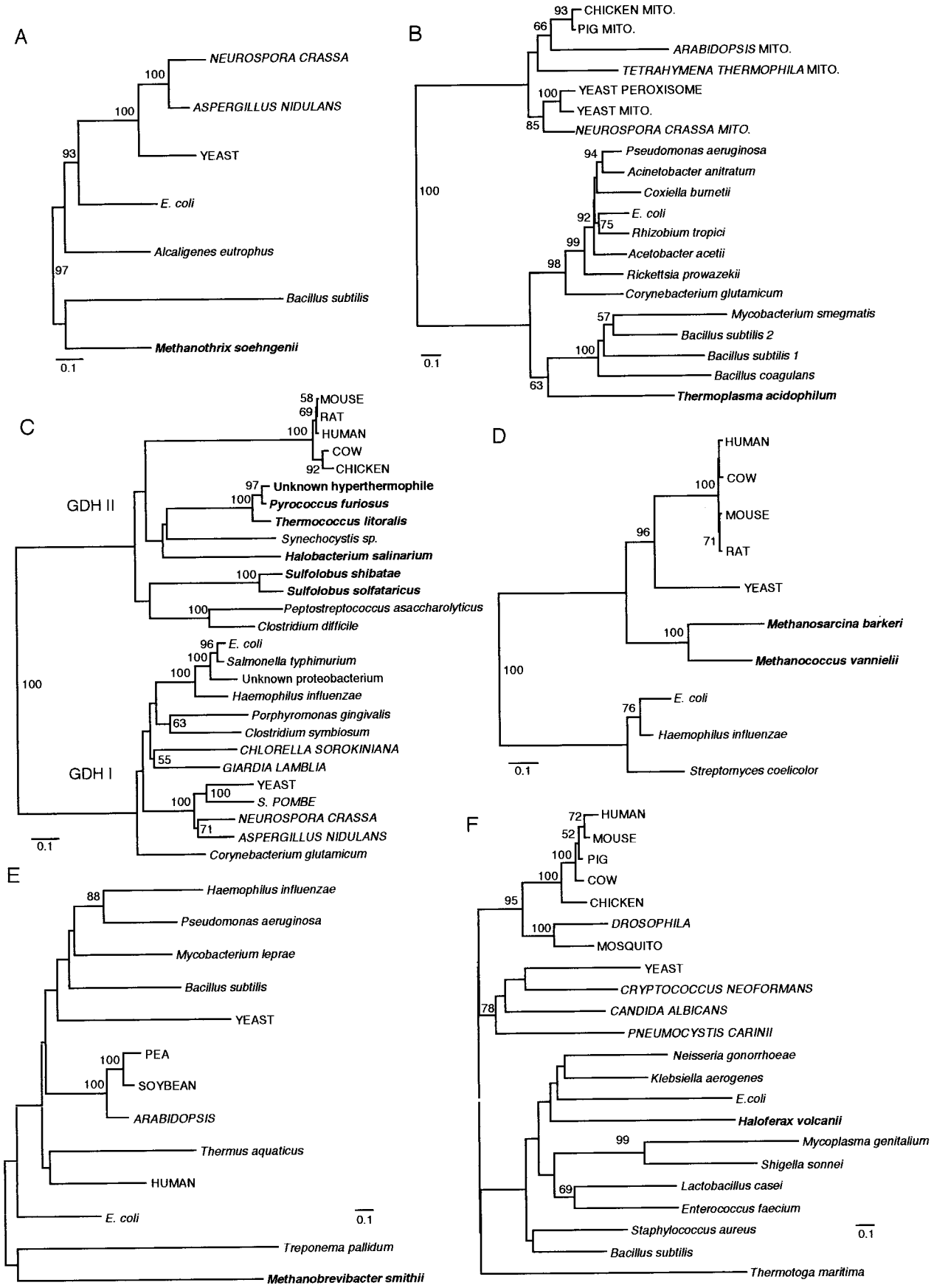


FIG. 15. Phylogenies for the acetyl-CoA synthetase (A), citrate synthase (B), GDH (C), argininosuccinate synthetase (*argG*) (D), pyrroline-5-carboxylate reductase (*proC*) (E), and DHFR (F) genes.

acetyl-CoA with oxaloacetate and water. Citrate synthase activity has been documented in species from all three domains, including different archaeal groups (with the possible exception of some methanogens [81]). Partial N-terminal sequences have been determined for citrate synthases from *Pyrococcus furiosus*, *Sulfolobus solfataricus*, and *Haloferax volcanii* (70, 219, 238).

The complete citrate synthase genes have been determined for *Thermoplasma acidophilum* (314) and, more recently, *P. furiosus* (238). In the former species, this protein is 384 amino acids in length. Although only about 20% similar at the amino acid level, the crystal structures of *T. acidophilum* and pig heart citrate synthases show a high degree of congruence (285). In phylogenetic analyses, archaeal citrate synthases cluster closest with those from gram-positive bacteria (Fig. 15B). A similar arrangement occurs in gene trees of GS, where a lateral transfer between species of the *Archaea* and gram-positive bacteria has been proposed (45, 319).

Eukaryotic citrate synthase genes encoding separate mitochondrial and cytostolic isoforms are highly divergent from those of the *Bacteria* or *Archaea*. Unlike the situation for eukaryotic GAPDH genes, mitochondrion- and peroxisome-targeted citrate synthase isoforms do not seem to have originated from a bacterial endosymbiont.

**Malate dehydrogenase.** In the final step of the CAC, malate dehydrogenase (MDH) catalyzes the bidirectional reaction L-malate + $NAD^+ \rightleftharpoons$ oxaloacetate + NADH + $H^+$. In eukaryotes, there are two MDH isoforms, a mitochondrial type, which functions in the CAC, and a cytosolic version, which is part of the malate-aspartate shuttle cycle along with aspartate transaminase.

In their 1989 paper, Iwabe et al. (164) suggested that MDH and L-lactate dehydrogenase (LDH) held promise as yet another ancient duplicated gene family, although the archaeal sequences were unknown. Subsequently, full-length MDH gene sequences were determined for *Methanothermus fervidus* (160) and *Haloarcula marismortui* (60) and partial amino-terminal sequences were determined for *Sulfolobus acidocaldarius* (136). Contrary to the predictions of Iwabe et al. (164), these archaeal MDH genes are highly divergent not only from bacterial and eukaryotic genes but also from each other (MDH phylogeny is not shown). The MDH gene of *H. marismortui* shows greater affinity, in both sequence and structure, to eukaryotic and bacterial LDH genes, while the *M. fervidus* LDH gene is highly diverged from all forms (60). On the basis of X-ray crystallography of *H. marismortui* MDH, Dym et al. suggested that particular amino acid substitutions might be related to specific adaptations to halophilic environments (97).

Cytosolic and mitochondrial MDH genes of eukaryotes are also highly divergent. Mitochondrial isoforms tend to be closely related to *E. coli* MDH, which might reflect an origin from a proteobacterial endosymbiont. However, the full extent of the MDH and LDH gene family may not be fully realized, since an MDH sequence from the bacterium *Thermus flavus* is highly similar to eukaryotic cytosolic isoforms.

### Amino Acid Biosynthesis and Degradation

Degradation products from excess amino acids comprise a small but significant portion of the total energy budget of the cell. Utilization of these by-products in the CAC provides an indirect linkage between amino acid metabolic and central metabolic pathways. Despite the abundance of amino acid biosynthetic enzymes within the cell (277), relatively few examples of gene sequences are known for the *Archaea* (reviewed in reference 272). However, two gene families, the GS and GDH genes, have been fairly intensely studied from a phylogenetic perspective.

**Aminotransferases.** The aminotransferases are a large group of pyridoxal-5′-phosphate-dependent enzymes which catalyze the reversible transformation of amino acids to oxo acids. The aminotransferases fall into four distantly related groups (168, 233). The closest relationships are among subgroup I enzymes, which include aspartate, alanine, tyrosine, phenylalanine, and histinol-phosphate aminotransferases. These enzymes commonly catalyze the transfer of an amino group to 2-oxoglutamate, which results in L-glutamate and a reaction-specific oxo acid.

The existence of aminotransferases in the *Archaea*, *Bacteria*, and eukaryotes suggests that this multigene family diverged very early in cellular evolution. However, evolutionary relationships among subgroup I aminotransferases are complex, and the switching of substrates among orthologous genes might have occurred several times. The best-represented aminotransferase in the database is aspartate aminotransferase (AAT). In eukaryotes, AAT isoforms are targeted to specific subcellular compartments and are encoded by nuclear genes. In phylogenetic trees, eukaryotic isoforms are organized into four distinct clades according to their cellular compartments, which are the cytosol (where animals and plants form monophyletic clades), mitochondria (where animals and plants are polyphyletic), and plastid (333) (Fig. 16A). The exceptions are mitochondrial and cytosolic AAT isoforms of yeast, which are highly divergent and branch outside other eukaryotic sequences. Thus, it may be that the various isoforms arose from an early eukaryotic gene duplication. To resolve the evolution of eukaryotic aminotransferases, sequences from lower eukaryotes, such as amitochondrial protists, might be especially important.

Prokaryotic AAT genes are very diverse. *E. coli* and *H. influenzae* AAT are similar to their eukaryotic homologs. Initially, the only example of an archaeal AAT, that of *Sulfolobus solfataricus*, was found to be highly similar to a homolog characterized from a *Bacillus* thermophile (78). Mehta et al. (233) suggested that the closer similarity between these two AAT genes reflected structural constraints imposed by thermotolerance. Golding and Gupta (134) provided these data as support for an evolutionary affinity between gram-positive bacteria and the *Archaea*. However, new sequence data have dissolved both hypotheses. Two AAT genes characterized from separate strains of *Rhizobium meliloti*, a soil-dwelling α-proteobacterium, (328), also cluster with *Sulfolobus* and *Bacillus* homologs in phylogenies (333). In addition, newly available AAT genes from cyanobacteria and *Thermus* cluster with these bacterial and archaeal genes.

Interestingly, AAT genes are not monophyletic with respect to the remaining subgroup I aminotransferases. Eukaryotic AAT isoforms are closely related not only to AAT genes from *E. coli* and *Haemophilus* but also to tyrosine-reactive aromatic amino acid aminotransferases (AroAT) from *E. coli* and *Rhizobium* (Fig. 16A) (233). Furthermore, *Bacteria-Archaea*-type AATs are, in fact, more similar to eukaryotic tyrosine and alanine aminotransferases than to eukaryote-proteobacteria-type AATs.

The subgroup I aminotransferase gene family is even more heterogeneous if histidinol-phosphate aminotransferases (HAPT) are considered. This enzyme functions in the histidine biosynthetic pathway, and in the *Bacteria* it is encoded by the *hisC* gene (see below). HAPT is known from an archaeon, the extreme halophile *Haloferax volcanii* (71). All known HAPT genes seem to form a cohesive group, with the *H. volcanii* gene being more closely related to those from gram-positive bacte-

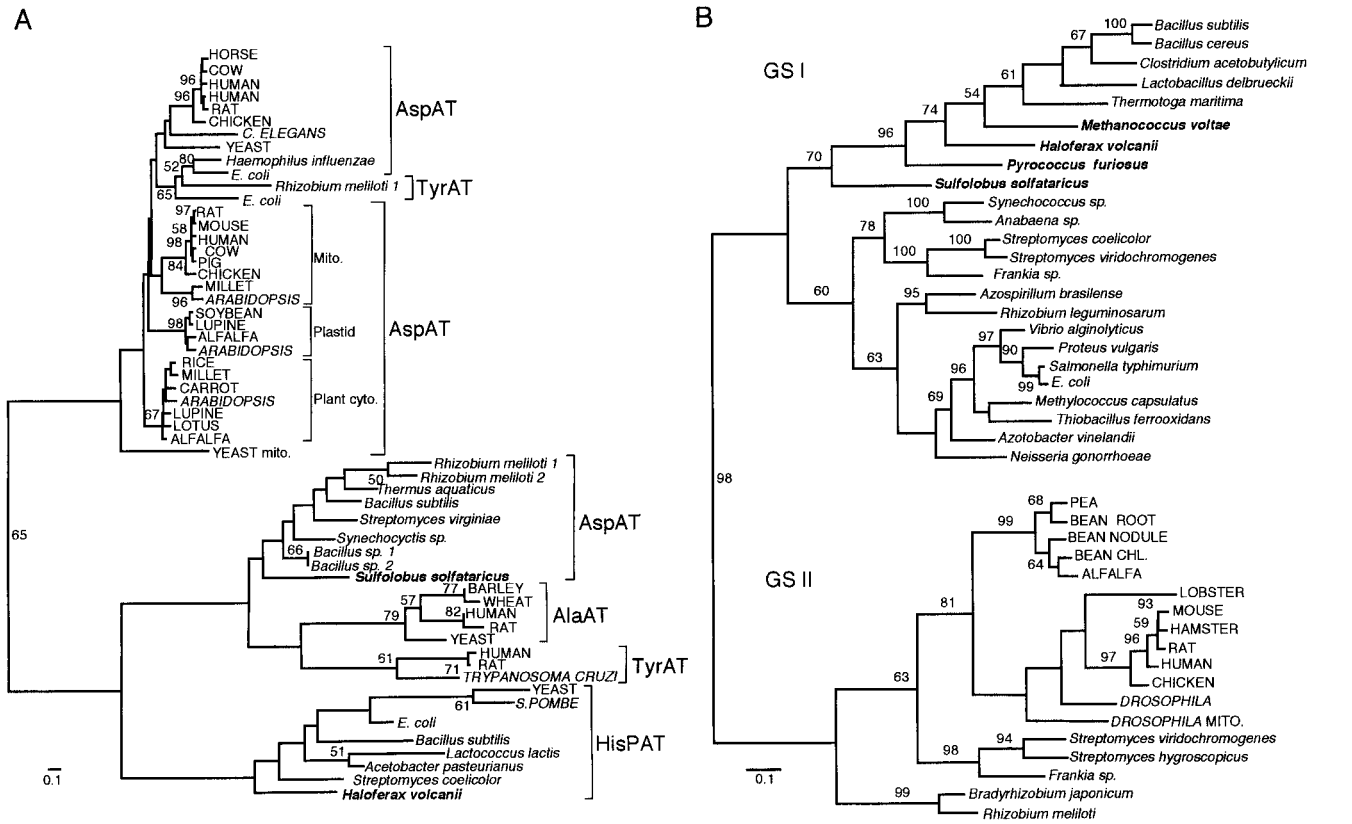


FIG. 16. Phylogenies for the subgroup I aminotransferases (AT) (A) and GS (B) genes. The different amino acid classes of AT are indicated, as well as the subcellular targeted isoforms of eukaryotic AspAT. For GS, the two major gene families are also labelled.

ria. The HAPT group falls between the two types of AAT from eukaryote-proteobacteria and *Bacteria-Archaea.*

The phylogenetic pattern of aminotransferases cannot be reconciled with notions of either direct organismal evolution or fixed enzyme specifities since the cenancestor. Benner et al. (27) suggested three other possible evolutionary scenarios. First, extensive amino acid switching might have occurred among aminotransferases, a theory somewhat supported by mutagenesis experiments which show that relatively few changes in the protein can redirect substrate specificity. The second scenario involves different patterns of replacement of some ribozyme precursor, relics of the "RNA world," by protein-based aminotransferases in the cenancester. The third theory postulates a late origin of aminotransferases, after the divergence of the *Archaea*, the *Bacteria*, and perhaps the eukaryotes and subsequent dispersion among the three domains by multiple lateral transfer events. Jensen and Gu (168) suggest that family I aminotransferases evolved by progressive specialization from an ancestral enzyme which had a broad specificity but lower efficiency.

Alternatively, we may have only a fragmented picture of the aminotransferase gene family and many orthologs may remain to be uncovered, such as an archaeal AAT gene similar to eukaryotic homologs. To this end, further sequence data from evolutionarily pivotal taxa of the *Bacteria*, *Archaea*, and eukaryotes will be important.

**Glutamine synthetase.** GS is an essential enzyme in ammonia assimilation and glutamine biosynthesis. Complete GS sequences (on average, 450 amino acids in length) have been determined for a number of species of the *Bacteria*, *Archaea*, plants, and animals. Archaeal species include members of both the *Euryarchaeota* (*Haloferax volcanii* [45], *Methanococcus voltae* [257], *Pyrococcus furiosus* [45], and *P. woesei* [319]) and the *Crenarchaeota* (*Sulfolobus solfataricus* [288]).

There are several isoforms of GS which are highly divergent from one another. GS type I (GSI) genes occur in the *Bacteria* and *Archaea* but not eukaryotes, while GS type II (GSII) genes exist in eukaryotes and a few members of the *Bacteria* including *Rhizobium* (85), *Frankia* (100, 280), and *Streptomyces* (198). Most bacteria with the GSII isoform also have the typically prokaryotic GSI copy. In the bacteria *Bacteriodes fragilis* (157), *Rhizobium leguminosarum* (66), and *Butyrivibrio fibrisolvens* (135), genes for a third variant, GSIII, have been found. Although distantly related, GSI and GSII genes are readily alignable while the exact evolutionary relationship of GSIII genes remain indeterminate, since these genes show little sequence similarity to either GSI or GSII isoforms.

Gene phylogenies show that GSI type genes can be further split into two distinct subdivisions, α and β (45, 319). GSI-β genes uniquely have a specific 25-amino-acid insertion and undergo posttranslational regulation by reversible adenylylation, whereas GSI-α and GSII genes do not. Among prokaryotes, GSI-α genes occur in the *Archaea* (with the possible exception of *Sulfolobus*), low-G+C gram-positive bacteria, and *Thermotoga maritima*, while all remaining bacteria have GSI-β genes. Although cyanobacterial sequences are characteristic of GSI-β, adenylylation control of enzyme activity is not evident (115), which suggests a secondary loss of this regulatory mechanism (45).

In GS gene phylogenies, members of the *Euryarchaeota*, the

low-G+C gram-positive bacteria, and *Thermotoga* form a single clade, although the internal branching order is poorly resolved (45, 302, 319). The sole *Crenarchaeota* representative, *S. solfataricus*, falls either at the base of GSI-α genes or all GSI genes depending on the phylogenetic method used. However, *Sulfolobus* is not an eocyte—its GS gene is like bacterial GSI, not eukaryotic GSII. While the GS gene is not congruent with the archaeal gene tree of eukaryotic origins, it does roughly mirror the closer relationship between archaea and gram-positive bacteria seen in HSP70 (146) and GDH (25) gene trees.

One possibility is that all bacteria once had the GSI-β gene, which was later replaced by an archaea specific GSI-α gene in some lineages (45, 319). However, to be reconciled with the divergent positions of *Thermotoga* and low-G+C gram-positive bacteria in the rRNA gene trees, there would have to been at least two independent GSI-α gene transfers (319). Alternatively, taxonomic repositioning of *Thermotoga* among the gram-positive bacteria must be considered (45)—a suggestion made earlier by Cavalier-Smith (58) based on similarities in cell envelope structure.

A second evolutionary paradox is the simultaneous occurrence of GSII genes in certain members of the *Bacteria* and eukaryotes. When bacterial GSII genes were initially discovered in nitrogen-fixing plant-symbiotic rhizobia and agrobacteria, they were suggested to have originated from plants (54). However, later phylogenetic analyses showed that bacterial GSII genes evolved before the divergence of animals and plants (197, 255, 295, 301). Kumada et al. (197) speculated that GSI and GSII genes were the result of an early gene duplication in the cenancestor. If GS genes did arise this way, alternative homologs of GSI and GSII should be found in eukaryotes and the *Archaea*, respectively. However, coexisting GSI and GSII genes have been found in only a few species of the *Bacteria*. Therefore, any theory of ancient GS gene paralogy would also need to explain the complex loss/gain pattern of GSI and GSII homologs seen in different lineages.

**Glutamate dehydrogenase.** There are two distinct functional kinds of GDHs. Tetrameric GDHs are NAD dependent and are instrumental in glutamate catabolism, while hexameric GDHs can be either NADP or NAD dependent and function in ammonia assimilation.

Of the two, only hexameric GDHs gene sequences are known for a number of species of eukaryotes and the *Bacteria* and *Archaea*. Forterre and coworkers have examined the phylogenetic relationships among hexameric GDH genes and proposed a further subdivision into two gene families. Family I GDH genes (GDH-I) occur in proteobacteria, the gram-positive *Clostridium symbiosum*, several fungi, and *Giardia lamblia* (25, 26) (Fig. 15C). Family II GDH genes (GDH-II) have been found in other gram-positive bacteria (*Peptostreptococcus asaccharolyticus* and *Clostridium difficile*), cyanobacteria, homeothermic animals, six archaeal species, and, recently, *Thermotoga maritima* (119). The animal genes are highly diverged from the bacterial and archaeal GDH-II genes, and the branching orders among these are poorly resolved except for the terminal clusters. While some phylogenetic reconstructions place *Halobacterium salinarium* GDH-II in a very earlier branch of animals (25), no single topology of archaeal GDH persists (156).

The deep split separating these two GDH families might suggest yet another instance of "cryptic" paralogy, yet the simultaneous existence of GDH-I and GDH-II in the same organism has not been proven. Although GDH-I and GDH-II genes are found only in different species of *Clostridium*, this gram-positive genus also appears as paraphyletic in rRNA trees, suggesting problems with species classification. Particularly problematic is the occurrence of GDH-I and GDH-II family genes in different groups of eukaryotes. If GDH-I and GDH-II genes are ancient paralogs, then, as for GSI and GSII, awkward scenarios of multiple-gene loss or gain in widely divergent taxa still must be addressed.

**Argininosuccinate synthetase.** In arginine biosynthesis and the cytosolic portion of the urea cycle, the enzyme argininosuccinate synthetase catalyzes the reversible condensation of citrulline and aspartate to argininosuccinate in the presence of ATP. Nearly a decade ago, argininosuccinate synthetase (*argG*) genes were found in *Methanosarcina barkerii* and *Methanococcus vannielii*, through the complementation of *argG* locus mutations in *E. coli* (237). In *M. barkerii*, there is an ORF just downstream of *argG* which corresponds to the large subunit of carbamoyl phosphate synthetase (CPS), another enzyme functioning in arginine metabolism (see below).

Although known for only a few taxa, argininosuccinate synthetase might have potential as another phylogenetic reporter molecule for deep evolutionary events. The gene is fairly long (about 397 amino acids long in *M. vannielii*) and encodes several well-conserved amino acid motifs which facilitates multiple sequence alignments. Phylogenetic analyses of known *argG* genes provide strong support for domain monophyly, with eukaryotic and archaeal sequences being the most similar (Fig. 15D).

**Carbamoyl phosphate synthetase.** Carbamoyl phosphate, a precursor compound in the biosynthesis of arginine and pyrimidine, is formed by the action of the enzyme CPS. CPS is a dimeric enzyme consisting of two asymmetrical subunits. The small subunit (42 kDa), through the hydrolysis of glutamine, provides a free amino group to the large subunit (120 kDa), which catalyzes the formation of carbamoyl phosphate in a complex reaction involving ATP, carbon dioxide, ammonia, and water (293).

In eukaryotes, mitochondrial and cytosolic versions of CPS are encoded by separate nuclear genes. Organelle-targeted CPS I participates in arginine biosynthesis, while cytosol-specific CPS II functions in pyrimidine biosynthesis. *Bacillus subtilis* has independently regulated CPS enzymes for arginine and pyrimidine biosynthesis (253), while proteobacteria make do with a single CPS enzyme. In *E. coli*, the *carA* and *carB* genes encode the small and large subunits, respectively. In *Methanosarcina barkeri*, the CPS large-subunit gene is located immediately upstream of the *argG* gene (237).

The N- and C-terminal halves of the *E. coli carB* gene show a high degree of sequence similarity, suggested to be the result of gene evolution by duplication (245). Schofield (293) proposed that eukaryotic and archaeal *carB* genes have similar internal, tandem arrangements, implying that this duplication might have occurred before the divergence of domains. If so, ammonia-metabolizing CPS genes might prove useful in reciprocally rooting the universal tree.

Lawson et al. (210) recently described CPS genes from the archaeon *Sulfolobus solfataricus*. In this organism, the amidotransferase and synthetase domains, encoded by *carA* and *carB* genes, respectively, are located just downstream from the *argG* and *argH* genes which encode the final two enzymes in the arginine biosynthetic pathway. Cotranscription of all four genes might occur. Lawson and co-workers constructed a reciprocally rooted CPS phylogeny by using the internal duplication of *carB* gene. In agreement with other paralogous universal trees, their phylogeny convincingly positioned the *Archaea* and eukaryotes as sister groups.

Another recent study suggests that methanogen *carB* subunits are more similar to those of the *Bacteria* while the *Sulfolobus* sequences are more like eukaryotic homologs (131).

Thus, the *Archaea* is paraphyletic, and the universal tree topology resembles that proposed by the eocyte hypothesis. However, it is also possible that lateral gene transfers between methanogens and members of the *Bacteria* have occurred.

Other, more recent CPS gene duplications might have occurred as well. Eukaryotic CPS I and II isoforms (and a newly discovered CPS III in fish) appear to be the result of gene duplications specific to eukaryotes—that is, the mitochondrially targeted isoform is unlikely to have originated from the endosymbiont (210). This gene duplication might have occurred recently, since some Apicomplexan protists apparently have a single CPS enzyme (118). Dual CPS enzymes in *B. subtilis* appear to be the result of another, independent gene duplication within the gram-positive bacteria.

**Tryptophan biosynthesis.** Nearly 40 years of study of the *trp* and *his* operons has been essential to our understanding of bacterial gene structure and regulation. The *trp* operon encodes five enzymes responsible for the synthesis of tryptophan from chorismate. The organization and sequences of *trp* operon genes have been determined for many bacteria and several archaea (73, 74, 243). In the *Bacteria*, the consensus gene order is anthranilate synthase α subunit (*trpE*), anthranilate synthase β subunit (*trpG*), anthranilate phosphoribosyl transferase (*trpD*), indole-3-glycerol phosphate synthetase (*trpC*), phosphoribosyl anthranilate isomerase (*trpF*), tryptophan synthase β subunit (*trpB*), and tryptophan synthase α subunit (*trpA*), or *trpEGDCFBA*. The gene fusions *trpGD*, *trpEG*, and *trpCF* and the breaking up of some clusters are seen in various bacterial species (184).

In contrast, gene orders of archaeal *trp* operons are highly variable with respect to both the bacterial model and each other. Among the *Euryarchaeota*, *Methanobacterium thermoautotrophicum* has the gene order *trpEGCFBAD* (234), which agrees with an earlier partial sequence of the *trp* operon (*trpFBA* only) from *Methanococcus voltae* (298). In *Haloferax volcanii*, the *trp* operon is broken into two isolated clusters, *trpCBA* and *trpDFEG*, which are nearly 1,200 kbp apart on the chromosome (203, 205). In the *Crenarchaeota S. solfataricus*, there is an intact operon, *trpBADFEGC*, which resembles the methanogen operon in gene order except that the *trpBAD* cluster is translocated to the 5′ end of the cluster (323). In fungi, *trp* genes are not part of a unified transcriptional unit, although some novel gene fusions do occur such as *trpGC* or *trpGCF*, and *trpAB* (74).

The fact that certain linkage groups, such as *trpBA* and *trpEG*, are retained throughout the *Bacteria* and *Archaea* suggests that some transcriptional units might be ancestral. If this is true, contemporary gene orders would be the result of specific arrangements that occurred after the *Archaea-Bacteria* divergence in different lineages. However, an alternative scenario of genes being distally scattered in the genome of the cenancestor, with the subsequent formation of operons occurring separately in the *Archaea* and *Bacteria*, cannot be ruled out.

Any reconstruction of the tentative ancestral *trp* operon might be enhanced if robust phylogenies could be derived from the individual *trp* genes. Unfortunately, *trp* genes make relatively poor phylogenetic reporter molecules, since their gene trees seldom provide statistically significant support for major branching points (Fig. 17). In the *trpA* gene tree, the archaea are closer to the bacteria, in particular the gram-positive bacteria. However, *Thermus aquaticus* also clusters with *Bacillus* species, and the proteobacteria are not a single clade (*Pseudomonas* sp. and *Caulobacter crescentus* in one group, and γ-proteobacteria in the other). Other *trp* gene trees are also inconsistent with bacterial evolution scenarios portrayed in either rRNA or *recA* gene phylogenies (103). In the *trpB* phylogeny, all domains are either para- or polyphyletic. Nucleus-encoded *trpB* genes in plants might have originated from plastids since they cluster with that of *Synechocystis*, a cyanobacterium. In all *trp* gene trees, the members of the *Archaea* either were paraphyletic with respect to the *Bacteria* (*trpA*, *trpB*, *trpE*, and *trpG*) or formed a clade which, in a single instance (*trpC*), was statistically nonsignificant.

Possible anthranilate synthase α subunit gene homologs are *phnA* (phenazine pathway) and *pabB* (ρ-aminobenzoate pathway) genes, while the anthranilate synthase β subunit gene is probably homologous to the *phnB* and *pabA* genes (108). Inclusion of these homologs in the respective *trpE* and *trpG* gene phylogenies did not improve the resolution of the eukaryote, archaeal, or bacterial clades (phylogenies not shown). However, the *phnA* and *phnB* genes of *Pseudomonas aeruginosa* were more similar to the respective *trpE* and *trpG* genes of enteric bacteria, as suggested previously (74, 108). Inconsistent rates of evolution, specific environmental adaptations, and, in the case of anthranilate-synthesizing enzymes, multiple events of gene duplication and convergence are all potential causes of the confusing species branching orders seen in *trp* gene trees.

**Histidine biosynthesis.** Recent reviews by Alifano et al. (4) and Fani et al. (113) have covered, in great depth, the histidine biosynthetic pathway and its related genes and their evolution. Briefly, in *E. coli* and *S. typhimurium*, nine intermediates and eight catalytic enzymes are known to participate in the biosynthesis of histidine from 5-phosphoribosyl-α-1-pyrophosphate and ATP. These γ-proteobacteria (and *H. influenzae*) have a single operon with the genes *hisGDCHBAF(IE)*, which encode, in order, ATP-phosphoribosyl transferase (*hisG*), histidinol dehydrogenase (*hisD*), HAT (*hisC*), glutamine amidotransferase (*hisH*), imidazoleglycerol-phosphate dehydratase/histidinol-phosphate-phosphatase (*hisB*), phosphoribosyl-formimino-5-amino-1-phosphoribosyl-4-imidazole carboxamide isomerase (*hisA*), cyclase (*hisF*), and pyrophosphohydrolase: phosphoribosyl-AMP cyclohydrolase (*hisIE*). Enzymes encoded by *hisB*, *hisD* and *hisIE* are bifunctional in histidine biosynthesis.

However, unlike the *trp* operon, the different prokaryotic examples of the *his* operon show radically different organizations. *Azospirillum brasilense*, an α-proteobacterium, has only a few *his* genes encoded in the operon *hisBdHorf168AFEorf122*, where *hisBd* is the distal dehydratase domain of the bifunctional enzyme. The gram-positive bacterium *Lactococcus lactis* has an operon consisting of 12 ORFs, of which eight, *hisCGDBdHAF(IE)*, correspond to those of *E. coli* (89, 90). *Staphylococcus aureus* has a six-gene cluster, *hisEABCDG*, while the same *his* genes seem to be dispersed in *B. subtilis* (113, 252).

In methanogenic and halophilic species of the *Archaea*, the few known *his* genes are not found in cotranscriptional units. The *Haloferax volcanii hisC* gene is not flanked by any recognizable ORFs (71), while the *Methanococcus vannielii hisA* and *hisI* genes are estimated to be over 10 kb apart (22). However, two genome-sequencing projects have recently revealed complete *his* operons for thermophilic species from both *Crenarchaeota* and *Euryarchaeota*. In *Sulfolobus solfataricus*, the operon gene order, *hisCGABFDEHI*, represents several novel rearrangements (63), while the *Pyrococcus furiosus* operon, with the operon structure *hisGDBHAFIEC*, is only one gene translocation, that of the distal placement of *hisC*, removed from the *E. coli* model (279). In *Saccharomyces cerevisiae*, the seven *his* genes exist on six different chromosomes (4). However, yeast *HIS4* does have similarities to *hisIE* and *hisD* and yeast *HIS7* has similarities to *hisH* and *hisF*, which suggests similar gene fusions.

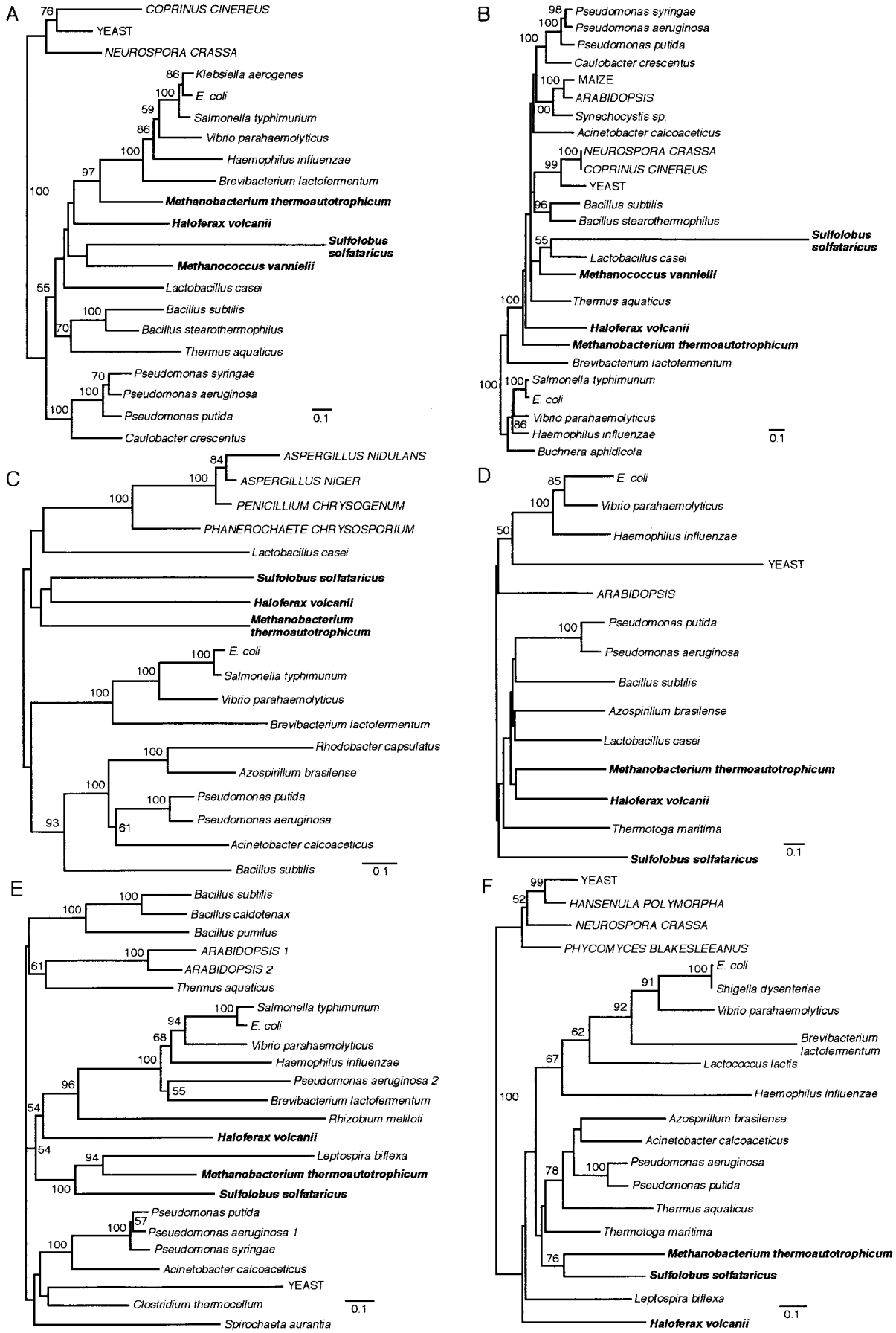


FIG. 17. Phylogenies for the *trpA* (A), *trpB* (B), *trpC* (C), *trpD* (D), *trpE* (E), and *trpG* (F) genes.

Gene fusions and duplications were probably an integral part of *his* gene evolution. Fani et al. (112) proposed, on the basis of internal sequence similarities, that the evolution of *hisA* and *hisF* genes occurred via two rounds of gene duplication. First, the *hisA* gene arose from a duplication and fusion of a primordial gene which was half as long as the current *hisA* gene. Then *hisA* duplicated to give rise to *hisF*. This scenario is supported by sequence similarities among the corresponding segments in *hisA* and *hisF* genes from species of the *Bacteria* and *Archaea*, which also suggests that the duplication events were completed before the divergence of the three domains. Later gene fusion events might have involved *hisB* and *hisIE*, which encode bifunctional enzymes in some but not all organisms (113). However, *hisD*, whose product catalyzes the final two steps in the histidine pathway, is a bifunctional complex in all known species, which is suggestive of an ancient gene fusion.

Available sequence data permit the construction of three domain species trees for *hisA*, *hisB*, *hisC*, *hisD*, *hisF*, *hisG*, *hisH*, and *hisIE* (Fig. 18). As with the genes of the *trp* operon, the resultant phylogenies have little resolution of critical branch points and cannot reach any consensus universal tree (63). The *Archaea*, with representives from both kingdoms, appears to group with the *Bacteria*, which is in general agreement with multiple *his* gene phylogenies generated by Fani et al. (113).

The phylogenies of the *his* and *trp* genes would suggest that these sequences are not highly informative markers for deep evolutionary events. The different gene orders, even within operons, indicate that *his* genes have been frequently rearranged in the past, perhaps erasing any recognizable gene patterns leading back to the cenancestor. However, the linkage groups *trpBA* and *trpEG* are widespread enough to suggest some long-term conservation of these spatial gene arrangements. Further, the existence of tryptophan and histidine biosynthetic pathway components in the *Archaea*, *Bacteria*, and some eukaryotes suggests that these pathways functioned in early cells, although interdomain transfer of some amino acid biosynthetic genes is still a possibility.

**Pyrroline-5-carboxylate reductase.** Pyrroline-5-carboxylate reductase (encoded by the *proC* gene) promotes the final step of proline biosynthesis, which is the reduction of pyrroline 5-carboxylic acid by either NADPH or NADH. The *proC* gene has been characterized for only a few species of bacteria and eukaryotes and one archaeon, *Methanobrevibacter smithii* (148). The *proC* gene tree (Fig. 15E) does not resolve any particular branching order, although Golding and Gupta (134) claimed that the *proC* phylogeny supported a eukaryote–gram-negative bacterium clade. Although *M. smithii* clusters with the "gram-negative" spirochete *Treponema pallidum*, this grouping is probably a methodological artifact of long branch lengths.

## Cofactors

The nucleoside thymine is formed by methylation of uracil. The specific methyl group comes from the oxidation of a tetrahydrofolate derivative to dihydrofolate. The enzyme dihydrofolate reductase (DHFR) plays an important role in the regeneration of tetrahydrofolate through the reduction of dihydrofolate by NADH. Eukaryotic DHFR genes are distinct from those of the *Bacteria* and *Archaea*; the archaeal gene is known from *Haloferax volcanii* (346). However, the relative branching orders among the different groups of the *Bacteria* and *Archaea* are unresolved (Fig. 15F).

## Purine Ribonucleotide Biosynthesis

**IMP dehydrogenase.** The enzymes and genes involved in purine ribonucleotide biosynthesis in the *Bacteria* are well known (98, 277). However, relatively few genes for this important pathway have been characterized in the *Archaea* (reviewed in reference 251). The universal phylogenies of two purine pathway proteins are considered here.

The first enzyme, inosine monophosphate dehydrogenase (IMPDH, encoded by the *guaB* gene), catalyzes the reaction $IMP + NAD^+ + H_2O \rightarrow$ xanthosine-5-phosphate + NADH. IMPDH gene sequences are known from the archaeon *Pyrococcus furiosus* (68), several gram-positive and γ-proteobacteria, and some eukaryotes, including the diplomonad *Tritrichomonas foetus* (21). The IMPDH gene tree suggests that *Pyrococcus* is an ally of the gram-positive bacteria as seen in GS and HSP70 gene phylogenies (68) (Fig. 19A).

**Phosphoribosylformylglycinamidine synthetase.** Phosphoribosylformylglycinamidine synthetase (FGAM synthetase, encoded by the *purL* gene) functions in the early steps of purine biosynthesis (21). In *B. subtilis*, two separate genes, *purL* and *purQ*, together encode a polypeptide that is equivalent in length and function to the single *purL* gene product in *E. coli*. A partial *purL* sequence has been found immediately downstream of the isoleucyl-tRNA synthetase gene in the archaeon *Methanobacterium thermoautotrophicum*, and the two genes might be cotranscribed (167).

The evolution of FGAM synthetase has figured in debates about the evolution of the eukaryotic cell. Gupta and Golding (145) recently suggested that a phylogeny of FGAM synthetase (which they called FGARAT) provided support for the closer evolutionary relationship between gram-positive bacteria and archaea on the one hand, and gram-negative bacteria and eukaryotes on the other—the cornerstone of the chimera hypothesis for the origin of eukaryotes. Supporting evidence consisted of the FGAM synthetase gene tree and two separate deletions shared among gram-positive bacteria and archaea but not gram-negative bacteria and eukaryotes. The inclusion of a cyanobacterial FGAM synthetase challenges these arguments (Gupta and Golding considered cyanobacteria to be gram-negative bacteria [134, 145]). Not only does *Synechococcus* FGAM synthetase branch strongly with *Methanobacterium* and gram-positive bacteria in the phylogenetic tree, but also the sequence has the same two deletions (281) (Fig. 19B). Thus, the exclusive grouping of the *Archaea* and gram-positive bacteria with respect to FGAM synthetase is untenable.

## Respiration

Superoxides ($O_2^-$) are toxic partial products of oxygen formed in respiration and various hydroxylation and oxygenation reactions. The cell neutralizes these compounds through the action of superoxide dismutase (SOD), which converts superoxides into hydrogen peroxide ($H_2O_2$) and dioxygen ($O_2$) (reviewed in reference 29). Given the rapid accumulation of superoxides in respiring cells, these enzymes are maintained at high cellular concentrations. There are three types of SODs, which are distinguishable on the basis of metal cofactors; CuZnSOD, FeSOD, and MnSOD. CuZnSOD activity has been found throughout the eukaryotes and in some proteobacteria but not in the *Archaea*. However, either FeSODs or MnSODs, which are closely related evolutionarily, are found throughout all living organisms, and the two enzymes coexist in some species of bacteria and plants (301). In eukaryotes, Fe/MnSOD proteins are active in mitochondria while the unrelated CuZnSOD protein functions in the cytosol.

MnSOD genes have been sequenced from archaeal halo-

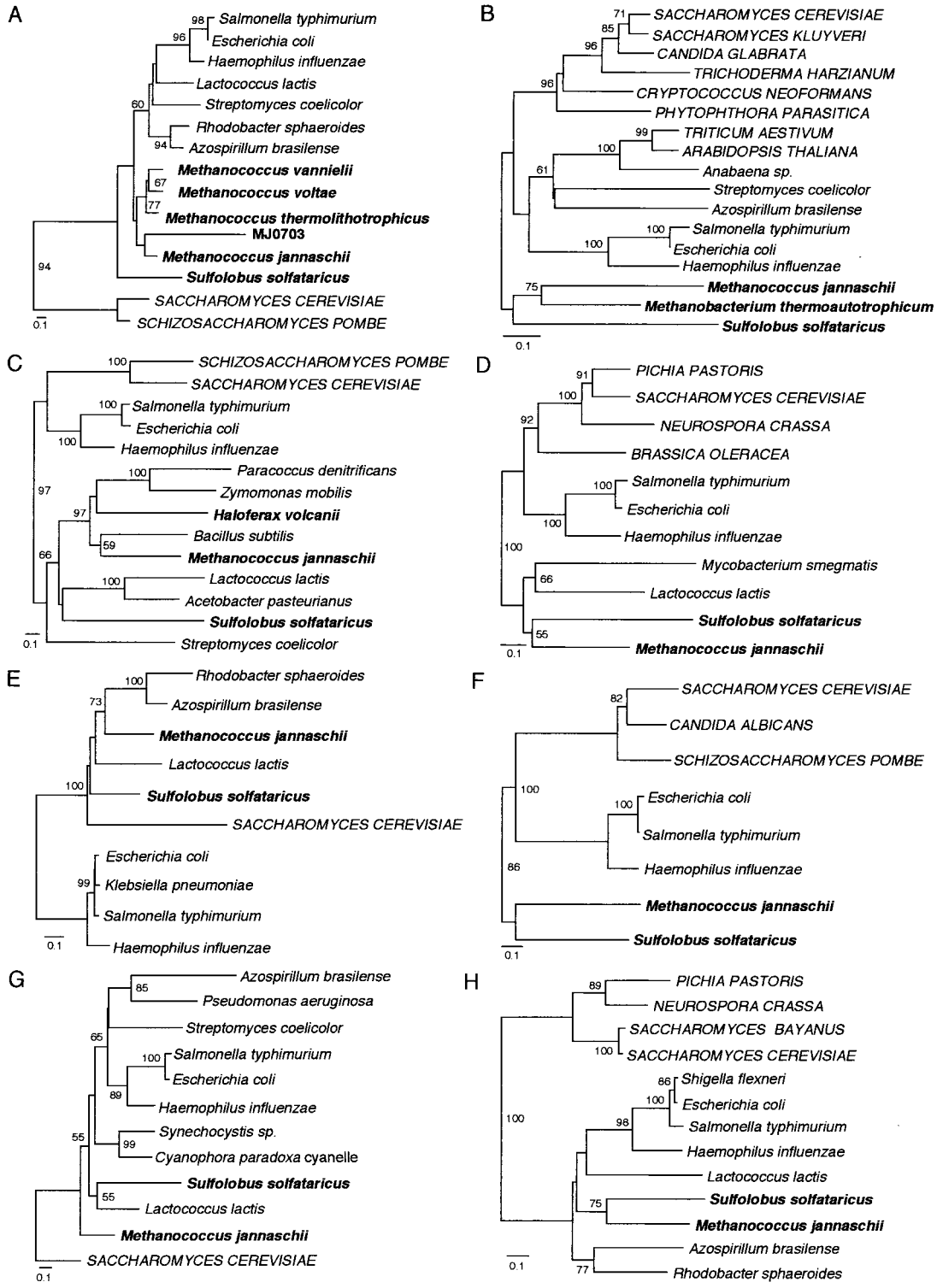


FIG. 18. Phylogenies for the *hisA* (A), *hisB* (B), *hisC* (C), *hisD* (D), *hisF* (E), *hisG* (F), *hisH* (G), and *hisIE* (H) genes.

philes (169, 231), and *Sulfolobus acidocaldarius* (191), while two FeSOD genes have been found in *Methanobacterium thermoautotrophicum* (315). In phylogenies of Fe/MnSOD genes, the archaeal isoforms cluster with eukaryotic FeSODs (Fig. 20A). *Sulfolobus* MnSOD and *Methanobacterium* FeSOD form a subcluster, which suggests that switching from Mn to Fe metal cofactors might have occurred recently in methanogens. The inclusion of *Mycobacterium* SODs among eukaryotic and archaeal SODs, as initially observed by Smith et al. (302), is still supported by the present neighbor-joining analysis, which includes more archaeal sequences.

Within the halobacteria, there is a pattern of recent dupli-

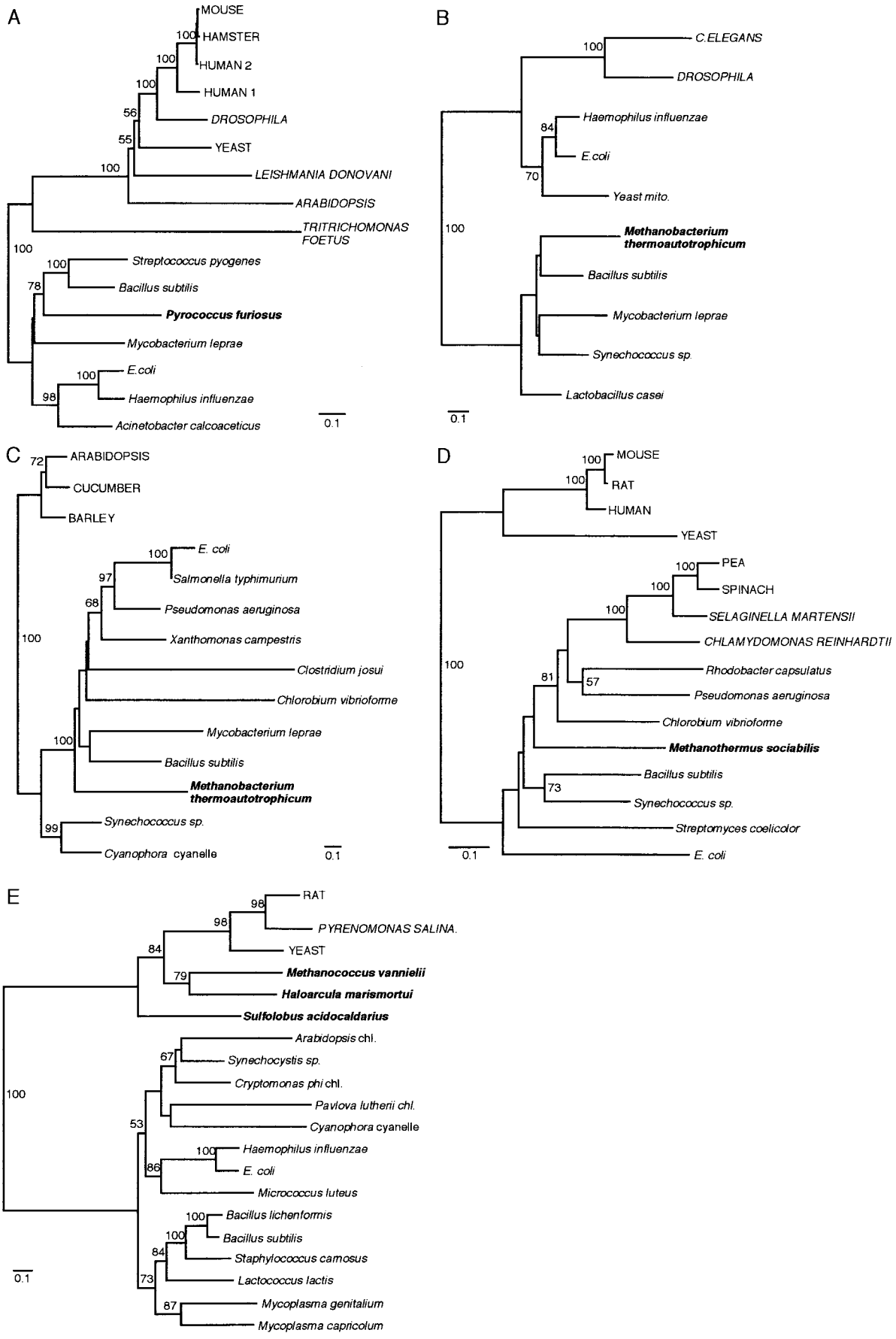


FIG. 19. Phylogenies for the IMPDH (A), FGAM synthetase (B), glutamyl-tRNA reductase (C), ALADH (D), and SecY (E) genes.

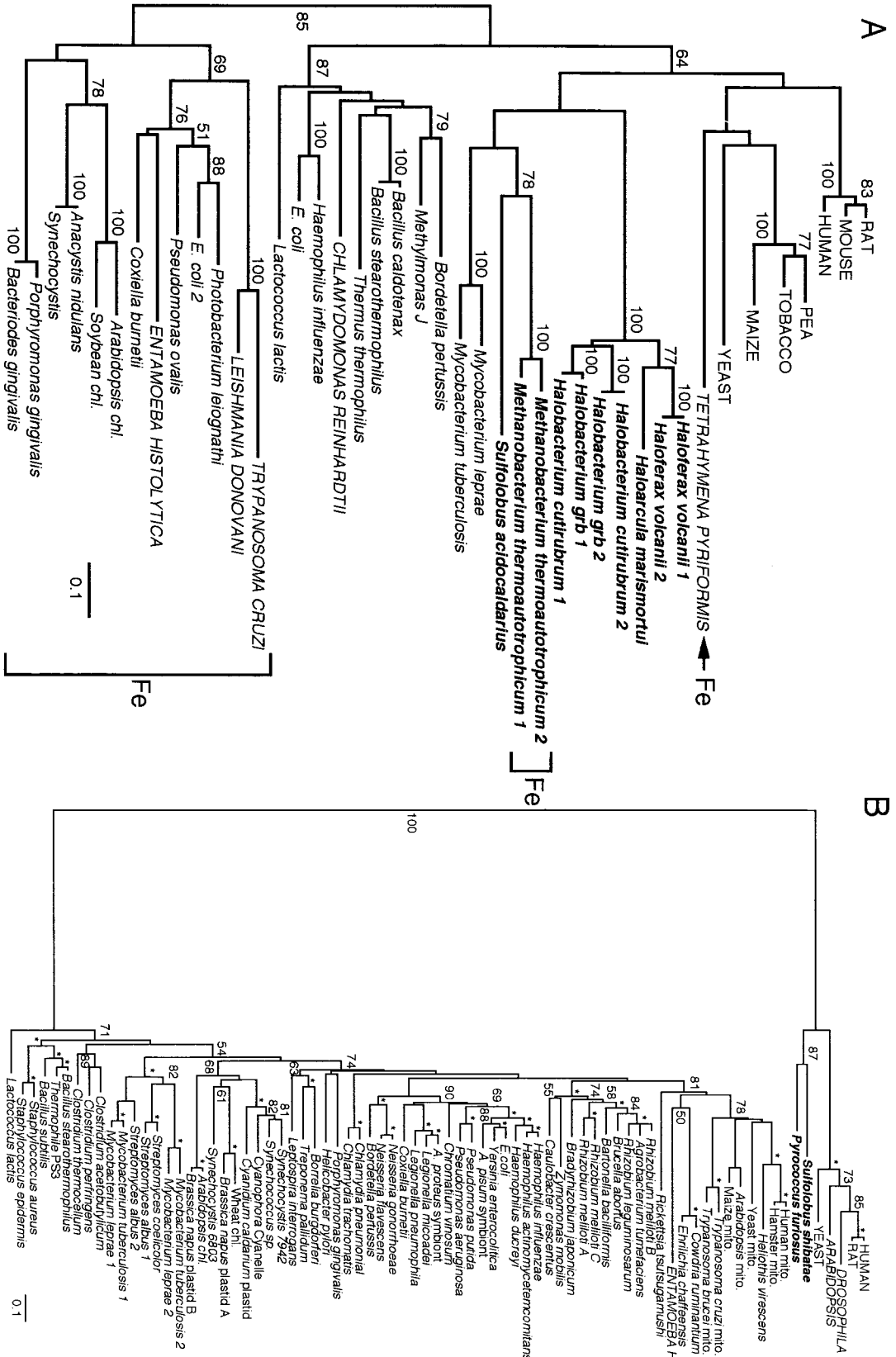


FIG. 20. Phylogenies for the FeMnSOD (A) and HSP60/TCP-1/TF-55 (B) genes. For the SOD genes, those utilizing Fe cofactors are indicated.

cation and divergence among MnSOD genes. Three species of halobacteria, *Halobacterium cutirubrum*, *Halobacterium* strain GRB, and *Haloferax volcanii*, each have duplicated MnSOD genes, named *sod* and *slg* (for *sod*-like gene), while a fourth species, *Haloarcula marismortui*, apparently has only a single gene (170). The genes *sod* and *slg* have very similar nucleotide and amino acid sequences but are differentially expressed by the cell. In the presence of paraquat, a generator of superoxide anions, both mRNA synthesis and protein expression of the *sod* gene are elevated while *slg* mRNA levels are unaffected. However, the *slg* gene is actively transcribed under other conditions; therefore, it is unlikely to be a pseudogene. Joshi and Dennis (170) suggested that the species with a single *sod* gene, *Haloarcula marismortui*, is ancestral to the other three halophilic species with multiple *sod* and *slg* genes. However, the SOD gene tree shown here does not position *Haloarcula* as an outgroup to other halophiles. The fact that duplicated SOD genes have recently been found in *Methanobacterium thermoautotrophicum* as well suggests that SOD gene duplications might have occurred independently in other archaeal lineages (315).

Further complexities in FeSOD and MnSOD gene evolution are seen when relationships between eukaryotic and bacterial homologs are considered. Chloroplast-targeted FeSOD genes are similar to those of cyanobacteria, which points to a likely endosymbiotic origin (302). Most eukaryotes have SOD genes of the Mn type, with the notable exceptions of several protists, which have FeSOD genes. Although the SOD gene of *Tetrahymena pyriformis* gene employs an iron cofactor, it is apparently more similar to other eukaryotic MnSOD types, which suggests a substitution of metal cofactors (16). On the other hand, the FeSOD gene of the amitochondrial protist *Entamoeba histolytica* is highly similar to FeSOD genes of the *Bacteria*, in particular the proteobacteria (317). In this instance, a lateral transfer of an FeSOD gene from a bacterium to a eukaryote has been suggested (302). Interestingly, FeSOD genes have also been found in other parasites, *Trypanosoma cruzi* and *Leishmania donovani* (162).

## Porphyrin Biosynthesis

**Glutamyl-tRNA reductase.** Many important respiratory and photosynthetic molecules, such as hemes, cytochromes, corrins, and chlorophylls, are based on porphyrin ring-like structures. Porphyrin itself is derived from eight molecules of 5-aminolevulinic acid (ALA), whose synthesis is known to occur via two different pathways. Some bacteria and mitochondria utilize the Shemin ($C_4$) pathway where ALA synthase catalyzes a single-step condensation of succinyl-CoA and glycine (see reference 324 and references therein). ALA synthesis in the *Archaea*, chloroplasts, cyanobacteria, *E. coli*, and *B. subtilis* proceeds via the $C_5$ pathway, which begins with glutamate. The biochemistry of the $C_5$ pathway is particularly unusual in that the initial substrate is a Glu-tRNA$^{Glu}$ molecule (166). In the presence of NADPH, glutamyl-tRNA reductase (GluTR) reduces Glu-tRNA$^{Glu}$ to glutamate-1-semialdehyde with the release of the tRNA. The conversion of glutamate-1-semialdehyde to ALA is then catalyzed by the enzyme glutamate-1-semialdehyde-2,1-aminomutase. The same tRNA$^{Glu}$ is used in both protein and ALA synthesis. The occurrence of eukaryotic GluTRs, known only from plant chloroplasts, suggests a derivation from cyanobacteria. Such a scenario is supported by the GluTRs phylogeny, which shows plant and cyanobacterial sequences as highly divergent neighbors relative to homologs from other bacteria and the archaeon *Methanobacterium thermoautotrophicum* (Fig. 19C). The closer association of the archaeal GluTR with the gram-positive bacteria than with the proteobacteria is tenuously suggested by the gene tree.

**5-Aminolevulinic acid dehydratase.** The enzyme 5-aminolevulinic acid dehydratase (ALADH) (encoded by *hemB* in the *Bacteria*) catalyzes the condensation of two molecules of ALA to form porphobilinogen, the monopyrrole precursor to large ringed structures. Unlike GluTR, ALADH is found in animals, plants, and fungi as well as in different bacteria and at least one archaeon, *Methanothermus sociabilis* (40). The gene phylogeny of ALADH is confusing, since plant genes appear to be more closely related to prokaryotic *hemB* than to other eukaryotic homologs (Fig. 19D). This might suggest that plant ALADH genes originated from a plastid endosymbiont were it not for the more distal placement of the cyanobacterium *Synechococcus*. As indicated by low bootstrap replication frequencies, the branching order among plants, bacteria, and archaea is not well resolved. Regardless, the general presence of porphyrin biosynthetic pathways in the *Archaea*, *Bacteria*, and eukaryotes suggests that the capacity to manufacture complex ring-like macromolecules could have existed in the cenancestor.

## Chaperonins

The term "chaperonins" defines a diverse group of enzymes whose general function is to stabilize the tertiary or folded configurations of other proteins. Although evolutionarily unrelated, bacterial 10-, 60-, and 70-kDa heat shock proteins (HSP10, HSP60, and HSP70, respectively) show similar chaperonin-like activities (142). The evolutionary analyses of HSP70 were discussed earlier. There are no eukaryotic or archaeal homologs to bacterial HSP10, also known as Cpn10 and GroES. However, bacterial and organellar HSP60 (also known as Cpn60, GroEL, and bacterial common antigen) appear to be distantly related to the eukaryotic TCP-1 (also called CCT) complex and the archaeal Tf-55 protein (185). Archaeal Tf-55 and eukaryotic TCP-1 are closely related, yet both domains are mono- or holophyletic (Fig. 20B), unlike the situation for HSP70.

Since HSP60 is a major bacterial antigenic protein implicated in autoimmune response, its sequences have been determined for a wide range of bacteria. Phylogenetic analyses of HSP60 sequences show that cyanobacteria and α-proteobacteria are nearest to chloroplasts and mitochondria, respectively, thus lending support to the endosymbiotic origins of organelles (142, 325). In particular the genera *Rickettsia* and *Ehrlichia*, α-proteobacteria living intracellularly in eukaryotic cells, were most closely related to the mitochondria. In this respect, the phyletic placement of the HSP60 gene from the amitochondrial protist *Entamoeba histolytica* with mitochondrion-like sequences postulates a secondary loss of organelles in this organism (67).

## Membranes

The transmembrane movements of proteins in the *Bacteria* are mediated by a group of interacting proteins known as the *sec* system. SecA and SecB are cytoplasmic proteins which interact with the cell membrane, while SecE and SecY are integral membrane components. The gene encoding SecY (also known as *prlA*) has been characterized from several members of the *Bacteria* and plastids, whereas the gene encoding the eukaryotic endoplasmic reticulum homolog, Sec61, has been sequenced from yeast and mammals. *secY* genes have been determined from the Archaea *Methanococcus vanielii* (11), *Haloarcula marismortui* (7), and *Sulfolobus acidocaldarius* (140, 176). As in *E. coli*, the archaeal *secY* gene is the last

component of the *spc* operon, being located just downstream of the gene for RP L15.

Phylogenetic analyses suggest that archaeal SecY and eukaryotic Sec61 are close relatives (Fig. 19E) (271). These proteins are highly conserved, with 10 membrane-spanning domains (totaling about 569 amino acids) which can be easily aligned among homologs. The present phylogeny suggests that the archaea are not monophyletic, with the halophiles and methanogens, rather than *Sulfolobus*, being the closest group to eukaryotes. Rensing and Maier (271) interpreted their SecY phylogeny as lending credence to Sogin's (304) chimeric origins of the eukaryotic nucleus. However, the present analysis suggests that the SecY tree supports the more conventional view of a bacterial root to the universal tree with a more recent divergence between the *Archaea* and eukaryotes.

## EVOLUTIONARY TRENDS IN MULTIPLE PROTEIN PHYLOGENIES

### Nearest Domains

At risk of oversimplifying some highly complex evolutionary scenarios, Table 1 summarizes the estimates of interdomain distances and domain coherence derived from 66 protein phylogenies described above. Some deeply paralogous genes, such as AAT, CPS, and MDH, which probably represent multigene families, were excluded from these comparisons. *Archaea* and eukaryotes (AK) are the closest related domains in 34 comparisons, while *Archaea* and *Bacteria* (AB) and *Bacteria* and eukaryotes (BK) are nearest domains in 21 and 11 comparisons, respectively. If instances of nonsignificant $P > 0.05$ [ANOVA]) differences among all three groups are excluded (leaving 56 gene trees), the clusterings AK, AB, and BK occur in 31, 18, and 7 protein comparisons, respectively.

On the basis of biochemical functions, most of the proteins involved in either DNA replication, transcription, or translation, so-called information roles, seem to favor the AK grouping. The specific proteins are elongation factors, RNAP, DNA polymerase, aminoacyl-tRNA synthetases, and most ribosomal proteins. Notable exceptions are gyrase B and photolyase, although for the latter, a second class of photolyase has been found where the *Archaea* and eukaryotes appear highly similar (341). That *Bacteria* and *Archaea* might have similar gyrase B/topoisomerase II molecules is not surprising given that the chromosome configurations for the two groups appear comparable. Furthermore, there are possibly other topoisomerases in the *Archaea* that are more akin to their eukaryotic counterparts. The total evidence would suggest that the *Archaea* and eukaryotes have a considerable shared ancestry with respect to core information pathways.

However, interdomain relationships become much less clear when metabolic enzymes are considered. Three proteins of the glycolytic pathway provide the strongest support for a BK grouping. More questionable support for the BK grouping comes from results for FeMnSOD, where many protist sequences are clearly bacterium-like while those of higher eukaryotes are near to the *Archaea*. Argininosuccinate synthetase, SecY, HSP60, and ATPase subunits (provisional that any similarities among archaeal and bacterial ATPase sequences have limited species distribution) generally support an AK grouping. Other biosynthetic and chaperone genes, like those encoding HSP70, GDH, and GS, suggest an AB grouping. Gene trees derived from the enzymes of the tryptophan and histidine biosynthetic pathway show little favor toward any particular rooting. In summary, single-gene trees derived from metabolic and biosynthetic enzymes provide no clear support for any particular grouping of domains whereas gene trees based on informational enzymes, as well as the few rootings with paralogous genes, tend to more solidly support the sisterhood of the *Archaea* and eukaryotes.

### Integrity of Domains

Considering only the gene trees with two or more species from each domain, universal phylogenies showing two or more paraphyletic domains are at least twice as frequent as phylogenies depicting all three domains as monophyletic sensu rRNA gene trees (33 versus 17 gene trees). Interestingly, all gene trees which showed monophyletic domains also supported the AK grouping. In the BK or AB grouping trees, species from two or more domains are always intermixed. The occurrence of monophyletic domain gene trees is not restricted by protein function, since examples could be found among informational pathway molecules (such as RNAPs and RPs) as well as metabolic/biosynthetic enzymes (such as argininosuccinate synthetase).

A closer examination of the instances of paraphyly reveals some interesting, although tentative, trends in species relationships. According to the most recent rendering of the eocyte hypothesis, the *Archaea* should follow a specific paraphyletic structuring where the members of the kingdom *Crenarchaeota* are the direct ancestors of eukaryotes while the members of the *Euryarchaeota* (as well as the *Bacteria*) form a separate clade (278). As discussed above, support for this version of the eocyte hypothesis now rests largely on the structural and phylogenetic analyses of elongation factor genes.

From the present analysis, there are now 21 different gene trees which might provide a suitable test of the eocyte hypothesis. Suitability is based on two criteria: first, that a level of significant support for an AK grouping exists, and second, that species from both archaeal kingdoms are present (Table 1). (The phylogenies of both ATPase subunits were included in this total since *Thermus* and *Enterococcus* probably suffered a gene transfer from the *Archaea* rather than the other way round.) Of the 21 phylogenies, 13 depicted the *Archaea* as monophyletic with various levels (55 to 97%) of bootstrap support. Included among these trees was that of EF-G/2, which, according to the present analysis, supported the monophyly of the *Archaea* in 86% of bootstrap replications.

Of the eight AK group phylogenies that were paraphyletic, only three depicted a branching order consistent with the eocyte hypothesis. Although the EF-Tu/1α gene tree showed the *Crenarchaeota* as the closest archaeal group to eukaryotes, bootstrap support for that clustering was less than 50%. However, the present analysis differs from earlier studies (13, 278) in that only a single tree-building method was used and the EF-Tu/1α and EF-G/2 data sets were not combined. The two remaining gene trees, those of the RP L11 and RP S11 genes, had high bootstrap values of 95 and 75%, respectively, for the eocyte grouping. The other five paraphyletic Archaea gene trees, had either methanogens or halophiles with equal frequency as the closest relative to eukaryotes.

There were another 11 gene trees which had a statistically significant AE grouping but lacked a *Crenarchaeota* representative. Five of these phylogenies also depicted a paraphyletic *Archaea* domain with no trend in which species (methanogen, halophile or thermophile) was nearest to eukaryotes. Thus, the paraphyletic *Archaea* can occur even if the *Euryarchaeota* is considered alone, which suggests that the *Archaea* paraphyly might be artifactual. Given that the paraphyletic *Archaea* can occur in various forms in different phylogenies and that the eocyte branching order is an infrequent tree topology, it re-

TABLE 1. Summary of protein gene phylogenies

| Function | Gene product | Interdomain distance[a] | | | Significance[b] | Domain integrity[c] | | |
|---|---|---|---|---|---|---|---|---|
| | | A-K | A-B | K-B | | K | A | B |
| DNA repair and replication | DNA polymerase II | **1.231** | 1.910 | 2.047 | * | | | |
| | Gyrase B | 2.406 | **0.706** | 2.327 | * | | | |
| | Photolyase class I | 1.318 | **1.188** | 1.310 | NS | | | |
| Transcription | RNAP subunit A | **1.138** | 1.803 | 2.126 | * | | | |
| | RNAP subunit B | **1.187** | 1.612 | 2.010 | * | | | |
| Translation | EF-G/2 | **1.153** | 1.337 | 1.475 | * | | | |
| | EF-Tu/1α | **0.638** | 1.152 | 1.289 | * | | | |
| | Isoleucyl-tRNA synthetase | 1.272 | **1.188** | 1.613 | * | | | |
| | Aspartyl-tRNA synthetase | **1.330** | 4.469 | 4.276 | * | | | |
| | Tryptophanyl-tRNA synthetase | **1.849** | 2.665 | 3.297 | * | | | |
| | Tyrosyl-tRNA synthetase | **1.283** | 2.564 | 3.048 | * | | | |
| | RP L2 | **1.031** | 1.558 | 1.989 | * | | | |
| | RP L3 | **1.209** | 2.176 | 2.700 | * | | | |
| | RP L5 | **1.066** | 1.686 | 1.990 | * | | | |
| | RP L6 | **1.267** | 2.271 | 2.969 | * | | | |
| | RP L10 | **1.752** | 3.599 | 4.600 | * | | | |
| | RP L11 | 1.941 | **1.367** | 2.591 | * | | | |
| | RP L14 | **0.775** | 1.141 | 1.279 | * | | | |
| | RP L15 | **2.012** | 2.910 | 3.194 | * | | | |
| | RP L22 | **1.374** | 3.186 | 2.769 | * | | | |
| | RP L23 | **0.852** | 1.685 | 1.656 | * | | | |
| | RP L30 | **1.799** | 1.811 | 2.582 | * | | | |
| | RP S5 | **0.950** | 1.411 | 1.650 | * | | | |
| | RP S7 | **0.788** | 2.215 | 2.400 | * | | | |
| | RP S8 | 4.049 | **2.257** | 3.978 | * | | | |
| | RP S9 | **1.238** | 2.097 | 2.022 | * | | | |
| | RP S10 | **1.188** | 1.230 | 1.564 | * | | | |
| | RP S11 | **0.712** | 1.213 | 1.492 | * | | | |
| | RP S12 | **0.684** | 2.995 | 3.425 | * | | | |
| | RP S15 | **1.629** | 2.728 | 3.594 | * | | | |
| | RP S17 | 4.505 | **1.435** | 4.135 | * | | | |
| | RP S19 | **0.976** | 1.916 | 1.864 | * | | | |
| Central metabolism | GAPDH | 2.760 | 2.699 | **0.684** | * | | | |
| | PGK | 1.481 | 1.490 | **0.808** | * | | | |
| | Enolase | 1.095 | 0.900 | **0.665** | * | | | |
| | Acetyl-CoA synthetase | 1.065 | **0.992** | 1.119 | NS | | | |
| | Citrate synthase | 1.910 | **1.231** | 2.047 | * | | | |
| Amino acid biosynthesis | Argininosuccinate synthetase | **1.141** | 2.115 | 2.003 | * | | | |
| | GDH-II | 1.200 | **0.866** | 1.209 | * | | | |
| | GSI | 2.899 | **0.977** | 2.826 | * | | | |
| | *hisA* product | 3.274 | **1.477** | 3.980 | * | | | |
| | *hisB* product | 0.931 | 1.025 | **0.872** | * | | | |
| | *hisC* product | 2.175 | **1.815** | 2.008 | * | | | |
| | *hisD* product | 1.121 | 1.082 | **0.945** | NS | | | |
| | *hisF* product | 1.072 | **0.859** | 1.335 | * | | | |
| | *hisG* product | 1.607 | **1.431** | 1.498 | * | | | |
| | *hisH* product | 1.326 | **1.282** | 1.345 | NS | | | |
| | *hisIE* product | 1.696 | **1.191** | 1.597 | * | | | |
| | *proC* product | 2.235 | 2.408 | **1.418** | * | | | |
| | *trpA* product | 1.454 | 1.406 | **1.332** | NS | | | |
| | *trpB* product | 1.087 | 1.093 | **0.638** | * | | | |
| | *trpC* product | **1.296** | 1.373 | 1.338 | NS | | | |
| | *trpD* product | 1.629 | **1.180** | 1.579 | * | | | |
| | *trpE* product | **1.196** | 1.271 | 1.358 | NS | | | |
| | *trpG* product | **0.858** | 1.013 | 0.870 | NS | | | |
| Cofactors | DHFR | 2.060 | 1.939 | **1.810** | NS | | | |
| Purine biosynthesis | IMPDH | 1.377 | **0.767** | 1.231 | * | | | |
| | FGAM synthetase | 2.282 | **1.084** | 1.827 | * | | | |
| Respiration | FeMnSOD | 1.138 | 1.032 | **0.810** | * | | | |
| Porphyrins | ALADH | 1.128 | **0.851** | 1.255 | * | | | |
| | Glu-tRNA reductase | 1.539 | 1.591 | **1.500** | NS | | | |
| Chaperones | 70-kDa HSP | 0.867 | **0.658** | 0.816 | * | | | |
| | 60-kDa HSP | **1.113** | 2.342 | 2.687 | * | | | |
| Membrane | SecY protein secretion | **1.434** | 2.942 | 3.119 | * | | | |
| ATP-proton | ATP synthase F1 α subunit | **0.704** | 1.888 | 1.909 | * | | | |
| | ATP synthase F1 β subunit | **0.636** | 2.114 | 2.061 | * | | | |

[a] Interdomain distance was the expected number of amino acid replacements per site, averaged other all pairwise comparisons between species of the *Archaea* (A), eukaryotes (K), and *Bacteria* (B). The lowest mean value (closest pair of domains) for a particular protein are in boldface type.

[b] *, significantly different ($P < 0.05$ by ANOVA); ns, not significantly different.

[c] Domain integrity was determined through inspection of neighbor-joining, bootstrapped gene trees for either monophyletic (solid boxes) or para/polyphyletic (shaded boxes) domains. Where only a single species from a particular domain was known, the box for that domain was left blank.

mains questionable whether the origin of eukaryotes can be specifically linked to the kingdom *Crenarchaeota.*

A lesser yet still considerable number of protein gene phylogenies suggest universal trees with either an AB or BK clustering. In trees with more than one archaeal, bacterial, and eukaryotic species, the two most closely related domains, whether they be *Archaea* and *Bacteria* or *Bacteria* and eukaryotes, were never monophyletic. The sole exception was the RP S8 gene tree, which had *Archaea* and *Bacteria* as closely related, monophyletic clades.

Are there any consistent trends in species associations in either AB or BK grouping gene trees? The answer is not clear since the domains in question were not monophyletic and major branch points were often ill resolved. In addition, species diversity was often low, especially with respect to different groups of bacteria and archaea. However, there are some species relationships in these nonarchaeal gene trees that are worthy of further study. In phylogenies supporting an AB grouping, the archaeal branches are often among those of the gram-positive bacteria. GS and IMPDH gene trees clearly cluster the *Archaea* with low-G+C gram-positive bacteria. The HSP70 gene phylogeny also positions archaeal species among the gram-positive bacteria, although the exact branching order among these groups and the cyanobacteria are poorly resolved. Notably, an extensive RecA protein phylogeny suggested that cyanobacteria are closely related to the high-G+C gram-positive bacteria (103). In other phylogenies, an association between *Archaea* and Gram-positive bacteria is less clear, although citrate synthase, gyrase B, and GDH I potentially provide other examples of the *Archaea* clustering with gram-positive bacteria.

### Endosymbiosis—More Than Just Organelles

In several gene trees, eukaryotes branch near those contemporary bacterial species now believed to be the closest living relatives to plastid (cyanobacteria) and mitochondrial α-proteobacteria) endosymbionts (137). Phylogenies for the proteins GAPDH, PGK, and enolase, as well as triosephosphate isomerase (TPI), suggest that these eukaryotic nuclear genes originated from proteobacteria and/or cyanobacteria (39, 155, 180).

Recently, there have been suggestions that bacterial endosymbiosis might have occurred very early in eukaryotic evolution, perhaps even before the divergence of those lower, amitochondrial protist lineages collectively known as the archezoa (227). Plausibly, the archezoa either secondarily lost their organelles or underwent some kind of endosymbiosis which resulted in the successful fixation of several bacterial genes in the nuclear genome but not an intracellular organelle. *Trichmononas vaginalis*, a parabaslia, has both mitochondrion-specific HSP60 (282) and mitochondrion-targeted HSP70 genes (128). Clark and Roger (67) describe HSP60 and another mitochondrion-specific gene, the pyridine nucleotide transhydrogenase gene from the protist *Entamoeba histolytica* which lacks mitochondria but is not considered to be an archezoan.

The TPI gene of *Giardia lamblia* clusters with other eukaryotic versions, which, in turn, are near an α-proteobacteria TPI gene (180). In addition, the GAPDH genes of *E. histolytica* and *G. lamblia* fall within the same cluster of eukaryotic homologs suggested to have emanated from a proteobacterium (155). A less species-rich phylogeny suggested that valyl-tRNA synthetases of eukaryotes, including *T. vaginalis*, also have proteobacterial origins (43). The general process where certain genes from the endosymbiont may have actually replaced the original nuclear copy has been called endosymbiotic gene replacement by Martin (227). The more specialized instances where contemporary species have the organellar gene but not the organelle are termed cryptic endosymbiosis, which invokes the notion of a temporal state of endosymbiosis followed by loss of the bacterial endosymbiont (155).

The timing of these various lateral gene transfers is an open question. While not embracing the chimera hypothesis, Gogarten and coworkers have suggested that extensive lateral gene transfers in early cellular evolution are primarily responsible for the confusing network of organismal relationships depicted by multiple gene trees (130, 156). They also noticed the tendency in some gene trees for the *Archaea* and the gram-positive bacteria to be in one cluster and for eukaryotes and the *Bacteria* to be in another. Gogarten et al. suggested that any transfer of genes from the *Bacteria* to eukaryotes had occurred earlier than organellar endosymbiosis (132).

Aside from the possibility of variable rates of evolution or gene convergence, two basic evolutionary scenarios might be behind the existence of so many conflicting protein gene trees. First, the eukaryotic nuclear genome is a chimera which resulted from some past cellular and genome fusion event involving a bacterium and an archaeon. As such, one subset of eukaryotic genes would be similar to archaeal homologs while a different subset would look most like bacterial counterparts. Furthermore, the proteins that are most similar among the *Archaea* and *Bacteria* should be reflective of a very ancient divergence event. According to such reasoning, it follows that the two prokaryotic domains should be resolved as monophyletic groups; however this is never true.

The second scenario depicts the archaeal tree as being essentially correct—that is, from the cenancestor, two lineages emerged, one leading to the *Bacteria* with the other later splitting into the *Archaea* and eukaryotes. Accordingly, the majority of gene trees should position eukaryotes and the *Archaea* as sister domains and should depict each as separate monophyletic groups, which is often the case. Protein gene trees incongruent with the archaeal tree must be explained in terms of either lateral gene transfers or specific gene losses which occurred after the emergence of the three domains. Collectively, phylogenetic evidence from both duplicated (hence rooted) and unrooted gene trees would seem generally, but not absolutely, consistent with the latter scenario.

### Rates Ain't Misbehavin'

Extensive heterogeneity among different lineages in the rates of evolution can confound phylogenetic inferences. Are rate differences between homologous proteins from the *Bacteria*, *Archaea*, and eukaryotes extensive? As an example, a very high rate of amino acid substitutions in eukaryotes might result in the AB grouping in a phylogeny. To assess the rate differentials among the three domains, a version of the relative rate test (291) was applied to each of the three gene tree subsets, i.e., the AK, AB, and BK grouping trees. Both significant and nonsignificant protein distances were included. For each tree subset, the mean distances from the two ingroups to the outgroup were plotted for each protein (Fig. 21A to C). Highly significant regressions were obtained for tree subsets corresponding to groupings of AK ($r^2 = 0.86$; $P < 0.001$), AB ($r^2 = 0.89$; $P < 0.001$), and BK ($r^2 = 0.98$; $P < 0.001$). The regression coefficients of all three equations were not significantly different with respect to slope and intercept values, which were close to 0.0 and 1.0, respectively.

What do these regressions tell us? First, the lower the amino acid substitution rate, the lower the protein will be on the

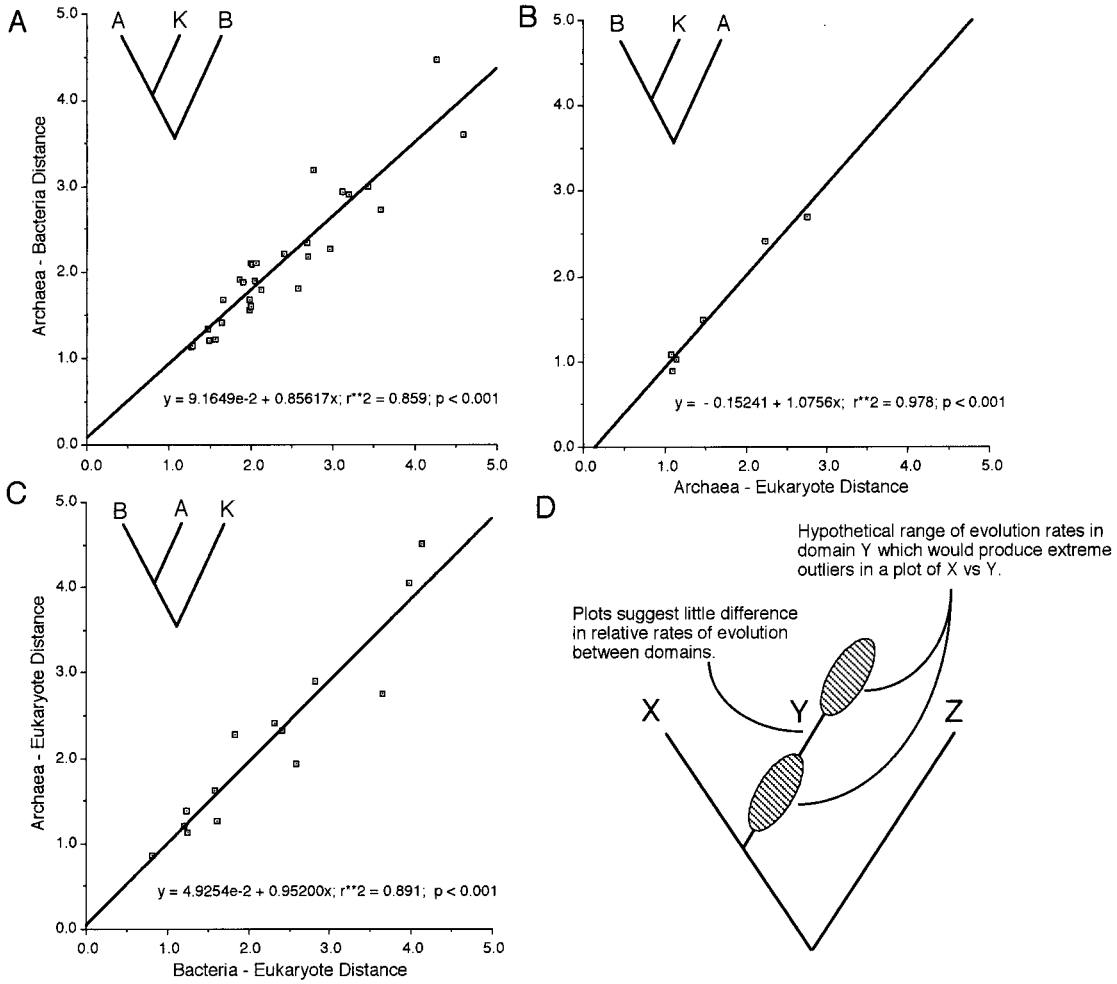


FIG. 21. Regression plots of the mean interdomain distances among different proteins. (A to C) As indicated by the universal tree insets, separate plots were done for proteins which support either an *Archaea*-eukaryote (A), *Bacteria*-Eukaryote (B), *Bacteria-Archaea* (C) clustering. Plotted coordinates are the separate distances from the two closest domains to the outgroup domain. In each case, the regression lines were highly significant and have an approximate slope of 1.0 with a near-zero intercept, which suggests few outlier points. (D) An interpretation of these plots is that for a given protein, the closest domains (X and Y) have evolved away from the outgroup domain (Z) at similar rates. Of course, the possibility of highly extreme differences in rates, such that rooting of the universal tree would be changed (thus X and Z appear as sister groups with Y as the outgroup), cannot be ruled out.

regression line. There is an upper limit to the distance values since highly rapidly evolving proteins would cease to have recognizable homology. The lower limit would suggest a minimal evolutionary distance between domains, which, from Table 1, is around 0.638 amino acid substitution per site for EF-Tu/1α. Only slightly higher minimal distance values were found in the AB and BE subsets. Second, since these regression lines have a slope value near 1.0 and are highly significant, the closest two domains have been changing at more or less a constant rate away from the outgroup in all three protein subsets. If a particular protein had a higher amino acid substitution rate in one domain relative to the other, it would appear as an outlier relative to the estimated regression line (Fig. 21D). Widespread rate differentials would shift the slope of the line away from 1.0. However, there are no significant outliers (points far off the regression line) in any of the three plots. While the occurrence of extreme rate differences that result in the wrong tree cannot strictly be ruled out, intermediate differences in rate are not apparent among the three domains. Individual molecules from certain taxa are known to vary widely in amino acid and/or nucleotide substitution rates; however, in the broader overview, rate differences might not be responsible for different perceptions of macroevolutionary relationships among the domains *Bacteria*, *Archaea*, and eukaryotes. Speculatively, problems in reaching a consensus rooting of the universal tree are not due to differences in evolutionary rates among genes but, rather, are the result of alternative modes of gene acquisition and differential retention of genes by the host genome.

## WHAT NEXT?

"We need to have more data . . .," is the incessant refrain for most summaries of this sort. However, archaeal, bacterial, and eukaryotic genome sequencing projects are generating tremendous volumes of relevant data. Thus, the major challenge is the synthesis of a grander view of the prokaryote-eukaryote transition.

In some respects, additional genomic sequence data has complicated our perceptions about cellular biology—it was

easier to pose plausible hypotheses when, in our ignorance, we could speculate about the contents of a genome. Perhaps, the most notable features about the first bacterial genomes sequenced, those of *Haemophilus* and *Mycoplasma*, were the genes that were missing rather than those that were present. Decades of genetic experiments and partial genome sequences of the mainstays of microbiology, *E. coli* and *B. subtilis*, perhaps raised our expectations about the contents of these first entire bacterial genomes. Human errors in sequencing and/or annotation aside, if we are surprised by the absence of even a few types of genes from different organisms, the ubiquity of any gene family is open to question. In this sense, the development of more robust universal trees is even more urgent to determine where genes appear or disappear—in the twigs or the load-bearing branches of the tree of life?

It is also evident that the *Bacteria*, *Archaea*, and eukaryotes are much too diverse to be characterized by just one complete genome apiece. Fortunately, genomes will be available from many examples of each. As such, comparative genomics will be important in discerning pandomain characteristics from limited, intradomain variation. If some consensus about the basic gene composition of bacterial, archaeal, and eukaryotic genomes could be reached, the evolutionary origins of specific cellular characters might become more clear.

In this respect, the abstract reconstruction of the genome of the cenancestor or the last common ancestor might prove useful. Initially, the cenancestor should be viewed as a list of genes and genome structures which are known to be shared across all three domains. Some of these genes have been discussed above. Moving beyond a simple gene list, the structure of early biochemical pathways might then be suggested. The spatial organization of bits of the genome could be inferred from comparisons of operon structures in the *Bacteria* and *Archaea*, although extensive rearranging within either group would render this task difficult. Finally, ancestral gene sequences could be inferred from amino acid sequence motifs conserved across archaeal, bacterial, and eukaryotic genes.

While any reconstruction of the cenancestor genome will reflect commonality among contemporary species of the *Archaea*, *Bacteria* and eukaryotes, specific features cannot be assumed to have necessarily existed before the divergence of the three domains. Any innovation in one domain, say the *Bacteria*, might have been transferred later to early members of the *Archaea* and eukaryotes. Thus, it will be impossible to discern, with absolute certainty, whether some specific characters existed in the cenancestor.

One point made apparent by this review is the number of gene trees which project incongruent views of the universal tree. In some instances, we are clearly pushing the phylogenetic methods and assumptions too far by trying to resolve the very deepest evolutionary branches. At other times, the gene trees seem to be highly robust but "wrong" from the perspective of organismal evolution. We may have many correct phylogenetic reconstructions of gene evolution but relatively fewer reconstructions of organismal evolution. This would imply that the evolution of genomes, or parts thereof, is sometimes decoupled from that of the host organism. Lateral gene transfer is one example of a decoupled process. What is needed now is the development of new evolutionary paradigms where genomes, biochemistry, and organisms are all considered in concert.

## ACKNOWLEDGMENTS

We thank the following individuals for providing either comments on the manuscript, preliminary sequence alignments, or early access to unpublished or submitted manuscripts: S. Balduaf, P. Cammarano, M. J. Danson, P. P. Dennis, D. Edgell, P. Gogarten, P. Forterre, P. Keeling, H.-P. Klenk, J. Logsdon, W. Martin, A. Roger, and A. Stoltzfus. We also thank G. Olsen, an anonymous individual, and the editor for reviewing the manuscript.

W.F.D. is a Fellow of the Canadian Institute for Advanced Research, and J.R.B. was a Canadian Medical Research Council (MRC) Postdoctoral Fellow. This work was supported mainly by MRC operating grant MT4467 to W.F.D. J.R.B. also thanks SmithKline Beecham for support during the final stages of this review.

## REFERENCES

1. **Acca, M., M. Bocchetta, E. Ceccarelli, R. Creti, K. O. Stetter and P. Cammarano.** 1993. Updating mass and composition of archaeal and bacterial ribosomes, archaeal-like features of ribosomes from the deep-branching bacterium *Aquifex pyrophilus*. Syst. Appl. Microbiol. **16:**129–137.
2. **Ævarsson, A.** 1995. Structure based sequence alignment of elongation factors Tu and G with related GTPases involved in translation. J. Mol. Evol. **41:**1096–1104.
3. **Alefounder, P. R., and R. N. Perham.** 1989. Identification, molecular cloning and sequence analysis of a gene cluster encoding the Class II fructose 1,6-bisphosphate aldolase, 3-phosphoglycerate kinase and a putative second glyceraldehyde 3-phosphate dehydrogenase of *Escherichia coli*. Mol. Microbiol. **3:**723–732.
4. **Alifano, P., R. Fani, P. Lio, A. Lazcano, M. Bazzicalupo, M. S. Carlomagno, and C. B. Bruni.** 1996. Histidine biosynthetic pathway and genes: structure, regulation, and evolution. Microbiol. Rev. **60:**44–69.
5. **Amils, R., P. Cammarano, and P. Londei.** 1993. Translation in Archaea, p. 393–438. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The biochemistry of Archaea (Archaebacteria). Elsevier Biomedical Press, Amsterdam, The Netherlands.
6. **Amiri, K. A.** 1994. Fibrillarin-like proteins occur in the domain *Archaea*. J. Bacteriol. **176:**2124–2127.
7. **Arndt, E.** 1992. The genes for ribosomal protein L15 and the protein equivalent to secY in the archaebacterium *Haloarcula marismortui*. Biochim. Biophys. Acta **1130:**113–116.
8. **Auer, J., G. Spicker, and A. Böck.** 1989. Organization and structure of the *Methanococcus* transcriptional unit homologous to the *Escherichia coli* "spectinomycin operon": implications for the evolutionary relationship of 70S and 80S ribosomes. J. Mol. Biol. **209:**21–36.
9. **Auer, J., G. Spicker, and A. Böck.** 1990. Phylogenetic positioning of archaebacteria on the basis of ribosomal protein sequences. Syst. Appl. Microbiol. **13:**354–360.
10. **Auer, J., G. Spicker, and A. Böck.** 1991. Organization and nucleotide sequence of a gene cluster comprising the translation elongation factor Ef-1α from *Sulfolobus acidocaldarius*. Syst. Appl. Microbiol. **14:**14–22.
11. **Auer, J., G. Spicker, and A. Böck.** 1991. Presence of a gene in the archaebacterium *Methanococcus vannielii* homologous to *secY* of eubacteria. Biochimie **73:**683–688.
12. **Baier, G., B. Redl, and G. Stoffler.** 1990. Structure, organization, and evolution of the L1 equivalent ribosomal protein gene of the archaebacterium *Methanococcus vannielii*. Nucleic Acids Res. **18:**719–724.
13. **Baldauf, S. L., J. D. Palmer, and W. F. Doolittle.** 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.
14. **Barnes, S. M., R. E. Fundyga, M. W. Jefferies, and N. R. Pace.** 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc. Natl. Acad. Sci. USA **91:**1609–1613.
15. **Barnes, S. M., C. F. Delwiche, J. R. Palmer, and N. R. Pace.** 1996. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. USA **93:**9188–9193.
16. **Barra, D., M. E. Schinina, F. Bossa, K. Puget, P. Durosay, A. Guissani, and A. M. Michelson.** 1990. A tetrameric iron superoxide dismutase from the eucaryote *Tetrahymena pyriformis*. J. Biol. Chem. **265:**17680–17687.
17. **Bartig, D., K. Lemkemeier, J. Frank, F. Lottspeich, and F. Klink.** 1992. The archaebacterial hypusine-containing protein: structural features suggest common ancestry with eukaryotic translation initiation factor 5A. Eur. J. Biochem. **204:**751–758.
18. **Bauer, M. W., S. H. Bauer, and R. M. Kelly.** 1997. Purification and characterization of a proteasome from the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl. Environ. Microbiol. **63:**1160–1164.
19. **Baumann, P., and S. P. Jackson.** 1996. An archaebacterial homologue of the essential eubacterial cell division protein FtsZ. Proc. Natl. Acad. Sci. USA **93:**6726–6730.
20. **Bazan, J. F., L. H. Weaver, S. L. Roderick, R. Huber, and B. W. Matthews.** 1994. Sequence and structure comparison suggest that methionine aminopeptidase, prolidase, aminopeptidase P, and creatinase share a common fold. Proc. Natl. Acad. Sci. USA **91:**2473–2477.
21. **Beck, J. T., S. Zhao, and C. C. Wang.** 1994. Cloning, sequencing, and structural analysis of the DNA encoding inosine monophosphate dehydrogenase (EC 1.1.1.205) from *Tritrichomonas foetus*. Exp. Parasitol. **78:**101–112.

22. **Beckler, G. S., and J. N. Reeve.** 1986. Conservation of primary structure in the *hisI* gene of the archaebacterium, *Methanococcus vannielii*, the eubacterium *Escherichia coli*, and the eukaryote *Saccharomyces cerevisiae*. Mol. Gen. Genet. **204:**133–140.
23. **Belfort, M., M. E. Reaban, T. Coetzee, and J. Z. Dalgaard.** 1995. Prokaryotic introns and inteins: a panoply of form and function. J. Bacteriol. **177:**3897–3903.
24. **Belfort, M., and A. Weiner.** 1997. Another bridge between kingdoms: tRNA splicing in Archaea and eukaryotes. Cell **89:**1003–1006.
25. **Benachenhou-Lahfa, N., P. Forterre, and B. Labedan.** 1993. Evolution of glutamate dehydrogenase genes: evidence for paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life. J. Mol. Evol. **36:**335–346.
26. **Benachenhou-Lahfa, N., P. Forterre, and B. Labedan.** 1994. PCR-mediated cloning and sequencing of the gene encoding glutamate dehydrogenase from the Archaeon *Sulfolobus shibatae*: identification of putative amino-acid signatures for extremophilic adaptation. Gene **140:**17–24.
27. **Benner, S. A., M. A. Cohen, G. H. Gonnet, D. B. Berkowitz, and K. P. Johnsson.** 1993. Reading the palimpsest: contemporary biochemical data and the RNA world, p. 27–70. *In* R. F. Gesteland and J. F. Atkins (ed.), The RNA world. The nature of modern RNA suggests a prebiotic RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
28. **Bergerat, A., D. Gadelle, and P. Forterre.** 1994. Purification of a DNA topisomerase II from the hyperthermophilic Archaeon *Sulfolobus shibatae*. J. Biol. Chem. **269:**27663–27669.
29. **Beyer, W., J. Imlay, and I. Fridovich.** 1991. Superoxide dismutases. Prog. Nucleic Acid Res. Mol. Biol. **40:**221–253.
30. **Bianchi, M. E.** 1994. Prokaryotic HU and eukaryotic HMG1: a kinked relationship. Mol. Microbiol. **14:**1–5.
31. **Bocchetta, M., E. Ceccarelli, R. Creti, A. M. Sanangelantoni, O. Tiboni, and P. Cammarano.** 1995. Arrangement and nucleotide sequence of the gene (*fus*) encoding elongation factor G (EF-G) from the hyperthermophilic bacterium *Aquifex pyrophilus*: phylogenetic depth of hyperthermophilic Bacteria inferred from analysis of the EF-G/*fus* sequences. J. Mol. Evol. **41:**803–812.
32. **Bochar, D. A., J. R. Brown, W. F. Doolittle, H.-P. Klenk, W. Lam, M. E. Schenk, C. V. Stauffacher, and V. W. Rodwell.** 1997. 3-Hydroxy-3-methylglutaryl coenzyme reductase of *Sulfolobus solfataricus*: DNA sequence, phylogeny, expression in *Escherichia coli* of the *hmgA* gene, and purification and kinetic characterization of the gene product. J. Bacteriol. **179:**3632–3638.
33. **Borges, K. M., S. R. Brummet, A. Bogert, M. C. Davis, K. M. Hujer, S. T. Domke, J. Szasz, J. Ravel, J. DiRuggiero, C. Fuller, J. W. Chase, and F. T. Robb.** 1996. A survey of the genome of the hyperthermophilic Archaeon, *Pyrococcus furiosus*. Genome Sci. Technol. **1:**37–46.
34. **Bouthier de la Tour, C., C. Portemer, R. Huber, P. Forterre, and M. Duguet.** 1991. Reverse gyrase in thermophilic eubacteria. J. Bacteriol. **173:**3921–3923.
35. **Bouthier de la Tour, C., C. Portemer, M. Nadal, K. O. Stetter, P. Forterre, and M. Duguet.** 1990. Reverse gyrase, a hallmark of the hyperthermophilic archaebacteria. J. Bacteriol. **172:**6803–6808.
36. **Braithwaite, D. K., and J. Ito.** 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. **21:**787–802.
37. **Brendel, V., L. Brocchieri, S. J. Sandler, A. J. Clark, and S. Karlin.** 1997. Evolutionary comparisons of RecA-like proteins across all major kingdoms of living organisms. J. Mol. Evol. **44:**528–541.
38. **Breton, R., D. Watson, M. Yaguchi, and J. Lapointe.** 1990. Glutamyl-tRNA synthetases of *Bacillus subtilis* 168T and of *Bacillus stearothermophilus*. J. Biol. Chem. **265:**18248–18255.
39. **Brinkmann, H., and W. Martin.** 1996. Higher-plant chloroplast and cytosolic 3-phosphoglycerate kinases: a case of endosymbiotic gene replacement. Plant Mol. Biol. **30:**65–75.
40. **Bröckl, G., M. Berchtold, M. Behr, and H. König.** 1992. Sequence of the 5-aminolevulinic acid dehydratase-encoding gene from the hyperthermophilic methanogen, *Methanothermus sociabilis*. Gene **119:**151–152.
41. **Brown, J. R.** 1996. Preparing for the flood: evolutionary biology in the age of genomics. Trends Ecol. Evol. **11:**510–513.
42. **Brown, J. R.** Unpublished data.
43. **Brown, J. R., and W. F. Doolittle.** 1995. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc. Natl. Acad. Sci. USA **92:**2441–2445.
44. **Brown, J. R., and W. F. Doolittle.** Gene descent, duplication and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. J. Mol. Evol., in press.
45. **Brown, J. R., Y. Masuchi, F. T. Robb, and W. F. Doolittle.** 1994. Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. J. Mol. Evol. **38:**566–576.
46. **Brown, J. R., F. T. Robb, R. Weiss, and W. F. Doolittle.** 1997. Evidence for the early divergence of tyrosyl- and tryptophanyl-tRNA synthetases. J. Mol. Evol. **45:**9–16.
47. **Brown, J. W., C. J. Daniels, and J. N. Reeve.** 1989. Gene structure, organization and expression in archaebacteria. Crit. Rev. Microbiol. **16:**287–338.
48. **Brown, J. W., and J. N. Reeve.** 1985. Polyadenylated, noncapped RNA from the archaebacterium *Methanococcus vannielii*. J. Bacteriol. **162:**909–917.
49. **Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleishmann, G. G. Sutton, J. A. Blake, L. M. Fitzgerald, R. A. Clayton, J. D. Gocayne, A. R. Kertavage, B. A. Dougherty, J.-F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. M. Geoghagen, J. F. Weidman, J. L. Fuhrmann, D. Nguyen, T. R. Utterback, J. M. Kelley, J. D. Peterson, P. W. Sadow, M. C. Hanna, M. D. Cotton, K. M. Roberts, M. A. Hurst, B. P. Kaine, M. Borodovsky, H.-P. Klenk, C. M. Fraser, H. O. Smith, C. R. Woese, J. C. Ventor.** 1996. Complete genome sequence of the methanogenic Archaeon, *Methanococcus jannaschii*. Science **273:**1058–1073.
50. **Burggraf, S., G. J. Olsen, K. O. Stetter, and C. R. Woese.** 1992. A phylogenetic analysis of *Aquifex pyrophilus*. Syst. Appl. Microbiol. **15:**352–356.
51. **Burnett, M. E., J. Liu, and T. Conway.** 1992. Molecular characterization of the *Zymomonas mobilis* enolase (*eno*) gene. J. Bacteriol. **174:**6548–6553.
52. **Busby, S., and R. H. Ebright.** 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell **79:**743–746.
53. **Cammarano, P., P. Palm, R. Creti, E. Ceccarelli, A. M. Sanangelantoni, and O. Tiboni.** 1992. Early evolutionary relationships among known life forms inferred from elongation factor EF-2/EF-G sequences: phylogenetic coherence and structure of the archaeal domain. J. Mol. Evol. **34:**396–405.
54. **Carlson, T. A., and B. K. Chelm.** 1986. Apparent eucaryotic origin of glutamine synthetase II from the bacterium *Bradyrhizobium japonicum*. Nature **322:**568–570.
55. **Carter, C. W.** 1993. Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. Annu. Rev. Biochem. **62:**715–748.
56. **Casiano, C., A. T. Matheson, and R. R. Taut.** 1990. Occurrence in the archaebacterium *Sulfolobus solfataricus* of a ribosomal protein complex corresponding to *Escherichia coli* $(L7/L12)_4 \cdot L10$ and eukaryotic $(P1)_2/(P2)_2 \cdot P0$. J. Biol. Chem. **265:**18757–18761.
57. **Cavalier-Smith, T.** 1992. Bacteria and eukaryotes. Nature **356:**570.
58. **Cavalier-Smith, T.** 1992. Origins of secondary metabolism. Ciba Found. Symp. **171:**64–87.
59. **Cavalier-Smith, T.** 1993. Kingdom protozoa and its 18 phyla. Microbiol. Rev. **57:**953–994.
60. **Cendrin, F., J. Chroboczek, G. Zaccai, H. Eisenberg, and M. Mevarech.** 1993. Cloning, sequencing, and expression in *Escherichia coli* of the gene coding for malate dehydrogenase of the extremely halophilic archaebacterium *Haloarcula marismortui*. Biochemistry **32:**4308–4313.
61. **Charlebois, R. L., T. Gaasterland, M. A. Ragan, W. F. Doolittle, and C. W. Sensen.** 1996. The *Sulfolobus solfataricus* P2 genome project. FEBS Lett. **389:**88–91.
62. **Charlebois, R. L., L. C. Schalkwyk, J. D. Hofman, and W. F. Doolittle.** 1991. Detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2. J. Mol. Biol. **222:**509–524.
63. **Charlebois, R. L., C. W. Sensen, W. F. Doolittle, and J. R. Brown.** 1997. Evolutionary analysis of the *his*$CGAB_4$FDEHI operon from the archaeon *Sulfolobus solfataricus* P2. J. Bacteriol. **179:**4429–4432.
64. **Chen, A., and C. D. Poulter.** 1993. Purification and characterization of farnesyl diphosphate/geranylgeranyl diphosphate synthase. A thermostable bifunctional enzyme from *Methanobacterium thermoautotrophicum*. J. Biol. Chem. **268:**1002–11007.
65. **Chen, A., P. A. Kroon, and C. D. Poulter.** 1994. Isoprenyl diphosphate synthases: Protein sequence comparisons, a phylogenetic tree, and predictions of secondary structure. Protein Sci. **3:**600–607.
66. **Chiurazzi, M., R. Meza, M. Lara, A. Lahm, R. Defez, M. Iaccarino, and G. Espin.** 1992. The *Rhizobium leguminosarum* bivar *phaseoli glnT* gene, encoding glutamine synthetase III. Gene **119:**1–8.
67. **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.
68. **Collart, F. R., J. Osipiuk, J. Trent, G. J. Olsen, and E. Huberman.** 1996. Cloning, characterization and sequence comparison of the gene coding for IMP dehydrogenase from *Pyrococcus furiosus*. Gene **174:**209–216.
69. **Confalonieri, F., C. Elie, M. Nadal, C. Bouthier de La Tour, P. Forterre, and M. Duguet.** 1993. Reverse gyrase: a helicase-like domains and a type I topoisomerase in the same polypeptide. Proc. Natl. Acad. Sci. USA **90:**4753–4757.
70. **Connaris, H., D. W. Hough, and M. J. Danson.** Unpublished data.
71. **Conover, R. K., and W. F. Doolittle.** 1990. Characterization of a gene involved in histidine biosynthesis in *Halobacterium* (*Haloferax*) *volcanii*: isolation and rapid mapping by transformation of an auxotroph with cosmid DNA. J. Bacteriol. **172:**3244–3249.
72. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16:**10881–10890.
73. **Crawford, I. P.** 1989. Evolution of a biosynthetic pathway: the tryptophan paradigm. Annu. Rev. Microbiol. **43:**567–600.
74. **Crawford, I. P., and R. Milkman.** 1990. Orthologous and paralogous divergence, reticulate evolution, and lateral gene transfer in bacterial *trp* genes, p. 77–95. *In* R. K. Selander, A. G. Clark, and T. S. Whittam (ed.), Evolution

at the molecular level. Sinauer Associates, Sunderland, Mass.
75. **Creti, R., E. Ceccaralli, M. Bocchetta, A. M. Sanangelantoni, O. Tiboni, P. Palm, and P. Cammarano.** 1994. Evolution of translational elongation factor (EF) sequences: reliability of global phylogenies inferred from EF-1α(Tu) and EF-2(G) proteins. Proc. Natl. Acad. Sci. USA **91:**3255–3259.
76. **Creti, R., F. Citarella, O. Tiboni, A. Sanangelantoni, P. Palm, and P. Cammarano.** 1991. Nucleotide sequence of a DNA region comprising the gene for elongation factor 1α (Ef-1α) from the ultrathermophile archaeote *Pyrococcus woesei*: phylogenetic implications. J. Mol. Evol. **33:**332–342.
77. **Creti, R., A. M. Sanangelantoni, P. Palm, and P. Cammarano.** Evolution of translational elongation factors G and 2 [EF-G(2)] inferred from an optimized selection of alignment positions. J. Mol. Evol., in press.
78. **Cubellis, M. V., C. Rozzo, G. Nitti, M. I. Arnone, G. Marino, and G. Sannia.** 1989. Cloning and sequencing of the gene coding for aspartate aminotransferase from the thermoacidophilic archaebacterium *Sulfolobus solfataricus*. Eur. J. Biochem. **186:**375–381.
79. **Curnow, A. W., M. Ibba, and D. Söll.** 1996. tRNA-dependent aspargine formation. Nature **382:**589–590.
80. **Dahlmann, B., F. Kopp, L. Kuehn, B. Niedel, G. Pfeifer, R. Hegerl, and W. Baumeister.** 1989. The multicatalytic proteinase (proteasome) is ubiquitous from eukaryotes to archaebacteria. FEBS Lett. **251:**125–131.
81. **Danson, M. J.** 1993. Central metabolism of the Archaea, p. 1–24. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The biochemistry of Archaea (Archaebacteria). Elsevier Biomedical Press, Amsterdam, The Netherlands.
82. **Danson, M. J.** 1989. Central metabolism of the archaebacteria: An overview. Can. J. Microbiol. **35:**58–64.
83. **Danson, M. J.** 1988. Archaebacteria: the comparative enzymology of their central metabolic pathways. Adv. Microb. Physiol. **29:**165–231.
84. **Danson, M. J., R. Eisenthal, S. Hall, S. R. Kessell, and D. L. William.** 1984. Dihydrolipoamide dehydrogenase from halophilic archaebacteria. Biochem. J. **218:**811–818.
85. **Darrow, R. A., and R. R. Knotts.** 1977. Two forms of glutamine synthetase in free-living root-nodule bacteria. Biochem. Biophys. Res. Commun. **78:**554–559.
86. **Dayhoff, M. O., R. V. Eck, and C. M. Park.** 1972. A model of evolutionary change in proteins, p. 89–99. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5. National Biomedical Research Foundation, Washington, D.C.
87. **Delong, E. F.** 1992. Archaea in coastal marine environments. Proc. Natl. Acad. Sci. USA **89:**5685–5689.
88. **Delong, E. F., K. Y. Wu, B. B. Prézelin, and R. V. M. Jovine.** 1994. High abundance of Archaea in Antarctic marine picoplankton. Nature **371:**695–897.
89. **Delorme, C., S. D. Ehrlich, and P. Renault.** 1992. Histidine biosynthesis genes in *Lactococcus lactis* subsp. *lactis*. J. Bacteriol. **174:**6571–6579.
90. **Delorme, C., J. J. Gordon, S. D. Ehrlich, and P. Renault.** 1993. Gene inactivation in *Lactococcus lactis* histidine biosynthesis. J. Bacteriol. **175:**4391–4399.
91. **Dennis, P. P.** 1997. Ancient ciphers: translation in Archaea. Cell **89:**1007–1010.
92. **DiRuggiero, J. R. Brown, A. M. Bogert, and F. T. Robb.** DNA repair systems in Archaea: mementos from the last universal common ancestor? J. Mol. Evol., in press.
93. **Doolittle, R. F., D. F. Feng, K. L. Anderson, and M. R. Alberro.** 1990. A naturally occurring horizontal gene transfer from a eukaryote to a prokaryote. J. Mol. Evol. **31:**383–388.
94. **Doolittle, W. F. and J. R. Brown.** 1994. Tempo, mode, the progenote and the universal root. Proc. Natl. Acad. Sci. USA **91:**6721–6728.
95. **Drlica, K.** 1990. Bacterial topisomerase and the control of DNA supercoiling. Trends Genet. **6:**433–437.
96. **Durovic, P., and P. P. Dennis.** 1994. Separate pathways for excision and processing of 16S and 23S rRNA from the primary rRNA operon transcript from the hyperthermophilic archaebacterium *Sulfolobus acidocaldarius*: similarities to eukaryotic rRNA processing. Mol. Microbiol. **13:**229–242.
97. **Dym, O., M. Mevarech, and J. L. Sussman.** 1995. Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. Science **267:**1344–1346.
98. **Ebbole, D. J., and H. Zalkin.** 1987. Cloning and characterization of a 12-gene cluster from *Bacillus subtilis* encoding nine enzymes for *de novo* purine nucleotide synthesis. J. Biol. Chem. **262:**8274–8287.
99. **Edgell, D. R., and W. F. Doolittle.** 1997. Archaea and the origin(s) of DNA replication proteins. Cell **89:**995–998.
100. **Edmands, J., N. A. Noridge, and D. R. Benson.** 1987. The actinorhizal root-nodule symbiont *Frankia* sp. strain CpI1 has two glutamine synthetases. Proc. Natl. Acad. Sci. USA **84:**6126–6130.
101. **Eggen, R. I. L., A. C. S. Geerling, A. B. P. Boshoven, and W. M. de Vos.** 1991. Cloning, sequence analysis, and functional expression of the acetyl coenzyme A synthetase gene from *Methanothrix soehngenii* in *Escherichia coli*. J. Bacteriol. **173:**6383–6389.
102. **Eick, D., A. Wedel, and H. Heumann.** 1994. From initiation to elongation: comparison of transcription by prokaryotic and eukaryotic RNA polymerases. Trends Biochem. Sci. **10:**292–292.
103. **Eisen, J. A.** 1995. The recA protein as a model molecule for molecular systematic studies of Bacteria: comparison of trees of recA and 16S rRNAs from the same species. J. Mol. Evol. **41:**1105–1123.
104. **Elie, C., M.-F. Baucher, C. Fondrat, and P. Forterre.** 1997. A protein related to eucaryal and bacterial DNA-motor proteins in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*. J. Mol. Evol. **45:**107–114.
105. **Eriani, G., J. Cavarelli, F. Martin, L. Ador, B. Rees, J.-C. Thierry, J. Gangloff, and D. Moras.** 1995. The class II aminoacyl-tRNA synthetases and their active site: evolutionary conservation of an ATP binding site. J. Mol. Evol. **40:**499–508.
106. **Eriani, G., M. Delarue, O. Poch, J. Gangloff, and D. Moras.** 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature **347:**203–206.
107. **Erickson, H. P.** 1995. FtsZ, a prokaryotic homolog of tubulin? Cell **80:**367–370.
108. **Essar, D. W., L. Eberly, A. Hadero, and I. P. Crawford.** 1990. Identification and characterization of genes for a second anthranilate synthase in *Pseudomonas aeruginosa*: interchangeability of the two anthranilate synthases and evolutionary implications. J. Bacteriol. **172:**884–900.
109. **Fabry, S., and R. Hensel.** 1988. Primary structure of glyceraldehyde-3-phosphate dehydrogenase deduced from the nucleotide sequence of the thermophilic archaebacterium *Methanothermus fervidus*. Gene **64:**189–197.
110. **Fabry, S., P. Heppner, W. Dietmaier, and R. Hensel.** 1990. Cloning and sequencing the gene encoding 3-phosphoglycerate kinase from mesophilic *Methanobacterium bryantii* and thermophilic *Methanothermus fervidus*. Gene **91:**19–25.
111. **Fabry, S., J. Lang, T. Niermann, M. Vingron, and R. Hensel.** 1989. Nucleotide sequence of the glyceraldehyde-3-phosphate dehydrogenase gene from the mesophilic archaebacteria *Methanobacterium bryantii* and *Methanobacterium formicicum*. Eur. J. Biochem. **179:**405–413.
112. **Fani, R., P. Liò, I. Chiarelli, and M. Bazzicalupo.** 1994. The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for *hisA* and *hisF* genes. J. Mol. Evol. **38:**489–495.
113. **Fani, R., P. Liò, and A. Lazcano.** 1995. Molecular evolution of the histidine biosynthetic pathway. J. Mol. Evol. **41:**760–774.
114. **Felsenstein, J.** 1993. PHYLIP (Phylogeny Inference Package) version 3.57c. Department of Genetics, University of Washington, Seattle. Distribution: http://evolution.genetics.washington.edu/phylip.html.
115. **Fisher, R., R. Tuli, and R. Haselkorn.** 1991. A cloned cyanobacterial gene for glutamine synthetase functions in *Escherichia coli*, but the enzyme is not adenylylated. Proc. Natl. Acad. Sci. USA **87:**3393–3397.
116. **Fitch, W. M., and K. Upper.** 1987. The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. Cold Spring Harbor Symp. Quant. Biol. **52:**759–767.
117. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. Fitzhugh, C. A. Fields, J. D. Gocayne, J. D. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
118. **Flores, M. V., W. J. O'Sullivan, and T. S. Stewart.** 1994. Characterization of the carbamoyl phosphate synthetase gene from *Plasmodium falciparum*. Mol. Biochem. Parasitol. **68:**315–318.
119. **Forterre, P.** Personal communication.
120. **Forterre, P., N. Benachenhou-Lahfa, F. Confalonieri, M. Duguet, C. Elie, and B. Labedan.** 1993. The nature of the last universal ancestor and the root of the tree of life, still open questions. BioSystems **28:**15–32.
121. **Forterre, P., A. Bergerat, D. Gadelle, C. Elie, L. Lottspeich, F. Confalonieri, M. Duguet, M. Holmes, and M. Dyall-Smith.** 1994. Evolution of DNA topisomerases and DNA polymerases: a perspective from Archaea. Syst. Appl. Microbiol. **16:**746–758.
122. **Forterre, P. and C. Elie.** 1993. Chromosome structure, DNA topoisomerases and DNA polymerases in archaebacteria (Archaea), p. 325–365. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The Biochemistry of Archaea (Archaebacteria). Elsevier Biomedical Press, Amsterdam, The Netherlands.
123. **Fox, G. E., L. J. Magrum, W. E. Balch, R. S. Wolfe, and C. R. Woese.** 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. Proc. Natl. Acad. Sci. USA **74:**4537–4541.
124. **Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, E. F. Kirkness, G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. L. Fuhrmann, D. T. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J.-F. Tomb, B. A. Dougherty, K. F. Bott, P.-C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison III, and J. C. Venter.** 1995. The minimal gene complement of *Mycoplasma genitalium*. Science **270:**397–403.

125. **Fuhrman, J. A., K. McAllum and A. A. Davis.** 1992. Novel major archaebacterial group from marine plankton. Nature **356:**148–149.
126. **Gagnon, Y., L. Lacoste, N. Champagne, and J. Lapointe.** 1996. Widespread use of the Glu-tRNA$^{Gln}$ transamidation pathway among bacteria. A member of the alpha purple Bacteria lacks glutaminyl-tRNA synthetase. J. Biol. Chem. **271:**4856–4863.
127. **Gambacorta, A., A. Trincone, B. Nicolaus, L. Lama, and M. de Rosa.** 1993. Unique features of lipids of Archaea. Syst. Appl. Microbiol. **16:**18–27.
128. **Germot, A., H. Philippe, 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:**14614–14617.
129. **Gill, J. F., M. J. Beach, and V. W. Rodwell.** 1985. Mevalonate utilization in *Pseudomonas* species M. Purification and characterization of an inducible 3-hydroxy-3-methylglutaryl-CoA reductase. J. Biol. Chem. **260:**9393–9398.
130. **Gogarten, J. P.** 1995. The early evolution of cellular life. Trends Ecol. Evol. **10:**147–151.
131. **Gogarten, J. P., L. Olendzenski, V. Puente, J. Leguina, and A. Lazcano.** Personal communication.
132. **Gogarten, J. P., E. Hilario, and L. Olendzenski.** 1996. Gene duplications and horizontal transfer during early evolution, p. 267–292. *In* D. M. Roberts, G. Alderson, P. Sharp, and M. Collins (ed.), Evolution of microbial life. Society for General Microbiology Symposia 54. Cambridge University Press, Cambridge, United Kingdom.
133. **Gogarten, J. P., H. Kibak, P. Dittrich, L. Taiz, E. J. Bowman, B. J. Bowman, N. F. Manolson, R. J. Poole, T. Date, T. Oshima, J. Konishi, K. Denda, and M. Yoshida.** 1989. Evolution of the vacuolar $H^+$-ATPase: implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA **86:**6661–6665.
134. **Golding, G. B., and R. S. Gupta.** 1995. Protein-based phylogenies support a chimeric origin for the eukaryotic genome. Mol. Biol. Evol. **12:**1–6.
135. **Goodman, H. J. K., and D. R. Woods.** 1993. Cloning and nucleotide sequence of the *Butyrivibrio fibrisolvens* gene encoding a type III glutamine synthetase. J. Gen. Microbiol. **139:**1487–93.
136. **Gorisch, H., and K.-D. Jany.** 1989. Archaebacterial malate dehydrogenase: the amino-terminal sequence of the enzyme from *Sulfolobus acidocaldarius* is homologous to the eubacterial and eukaryotic malate dehydrogenases. FEBS Lett. **247:**259–262.
137. **Gray, M.** 1992. The endosymbiont hypothesis revisited. Int. Rev. Cytol. **141:**233–357.
138. **Grayling, R. A., K. Sandman, and J. N. Reeve.** 1994. Archaeal DNA binding proteins and chromosome structure. Syst. Appl. Microbiol. **16:**582–590.
139. **Gropp, F., W.-D. Reiter, A. Sentenac, W. Zillig, R. Schnabel, M. Thomm, and K. O. Stetter.** 1986. Homologies of components of DNA-dependent RNA polymerases of archaebacteria, eukaryotes and eubacteria. Syst. Appl. Microbiol. **7:**95–101.
140. **Guenther, I., D.-C. Yang, J. Auer, and A. T. Matheson.** 1994. The nucleotide sequence of the secY gene in the Archaeon *Sulfolobus acidocaldarius*. Syst. Appl. Microbiol. **16:**692–696.
141. **Gupta, R. S.** 1984. *Halobacterium volcanii* tRNAs. J. Biol. Chem. **259:**9461–9471.
142. **Gupta, R. S.** 1995. Evolution of the chaperonin families (Hsp60, Hsp10 and Tcp-1) of proteins and the origin of eukaryotic cells. Mol. Microbiol. **15:**1–11.
143. **Gupta, R. S., K. Aitken, M. Falah, and B. Singh.** 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.
144. **Gupta, R. S., and G. B. Golding.** 1993. Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria and eukaryotes. J. Mol. Evol. **37:**573–582.
145. **Gupta, R. S., and G. B. Golding.** 1996. The origin of the eukaryotic cell. Trends Biochem. Sci. **21:**166–171.
146. **Gupta, R. S., and B. Singh.** 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.
147. **Gupta, R. S., and B. Singh.** 1994. Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. Curr. Biol. **4:**1104–1114.
148. **Hamilton, P. T., and J. N. Reeve.** 1985. Structure of genes and an insertion element in the methane producing archaebacterium *Methanobrevibacter smithii*. Mol. Gen. Genet. **200:**47–59.
149. **Hasegawa, M., N. Iwabe, Y. Mukohata, and T. Miyata.** 1990. Close evolutionary relatedness of archaebacteria, *Methanococcus* and *Halobacterium*, to eukaryotes demonstrated by composite phylogenetic trees of elongation factors EF-Tu and EF-G: eocyte tree is unlikely. Jpn. J. Genet. **65:**109–114.
150. **Hasegawa, M., T. Hashimoto, and J. Adachi.** 1992. Origin and evolution of eukaryotes as inferred from protein sequence data, p. 107–130. *In* H. Hartman, and K. Matsuno (ed.), The origin and evolution of the cell. World Scientific, Singapore.
151. **Hashimoto, T., and M. Hasegawa.** 1996. Origin and early evolution of eukaryotes inferred from the amino acid sequences of translation elongation factors 1 alpha/Tu and 2/G. Adv. Biophys. **32:**73–120.
152. **Henderson, E., M. Oakes, M. W. Clark, J. A. Lake, A. T. Matheson, and W. Zillig.** 1984. A new ribosome structure. Science **225:**510–512.
153. **Henikoff, S., and J. G. Henikoff.** 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA **89:**10915–10919.
154. **Hensel, R., P. Zwickl, S. Fabry, J. Lang, and P. Palm.** 1989. Sequence comparison of glyceraldehyde-3-phosphate dehydrogenases from the three urkingdoms: Evolutionary implication. Can. J. Microbiol. **35:**81–85.
155. **Henze, K. A., 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.
156. **Hilario, E., and J. P. Gogarten.** 1993. Horizontal transfer of ATPase genes—the tree of life becomes the net of life. BioSystems **31:**111–119.
157. **Hill, R. T., J. R. Parker, H. J. K. Goodman, D. T. Jones, and D. R. Woods.** 1988. Molecular analysis of a novel glutamine synthetase of the anaerobe, *Bacteroides fragilis*. J. Gen. Microbiol. **135:**3271–3279.
158. **Hirata, R., Y. Ohsumk, A. Nakano, H. Kawasaki, K. Suzuki, and Y. Anraku.** 1990. Molecular structure of a gene, VMA-1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. J. Biol. Chem. **265:**6726–6733.
159. **Holmes, M. L., and M. L. Dyall-Smith.** 1991. Mutations in DNA gyrase result in novobiocin resistance in halophilic archaebacteria. J. Bacteriol. **173:**642–648.
160. **Honka, E., S. Fabry, T. Niermann, P. Palm, and R. Hensel.** 1990. Properties and primary structure of the L-malate dehydrogenase from the extremely thermophilic archaebacterium *Methanothermus fervidus*. Eur. J. Biochem. **188:**623–632.
161. **Imanaka, T., S. Lee, M. Takagi, and S. Fujiwara.** 1995. Aspartyl-tRNA synthetase of the hyperthermophilic Archaeon *Pyrococcus* sp. KOD1 has a chimerical structure of eukaryotic and bacterial enzymes. Gene **164:**153–156.

161a. **Institute of Genome Research.** 1997. *Archaeglobus fulgidus* DNA sequence. Available from: ftp://ftp.tir.org/pub/data/.

162. **Ismail, S. O., Y. A. Skeiky, A. Bhatia, L. A. Omara-Opyene, and L. Gedamu.** 1994. Molecular cloning, characterization, and expression in *Escherichia coli* of iron superoxide dismutase cDNA from *Leishmania donovani chagasi*. Infect. Immun. **62:**657–664.
163. **Itoh, T.** 1989. Sequence analysis of the peptide-elongation factor EF-2 gene, downstream from those of ribosomal proteins H-S12 and H-S7, from the archaebacterial extreme halophile, *Halobacterium halobium*. Eur. J. Biochem. **186:**213–219.
164. **Iwabe, N., K.-I. Kuma, M. Hasegawa, S. Osawa, and T. Miyata.** 1989. Evolutionary relationship of Archaea, Bacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA **86:**9355–9359.
165. **Iwabe, N., K.-I. Kuma, H. Kishino, M. Hasegawa, S. Osawa, and T. Miyata.** 1991. Evolution of RNA polymerases and branching patterns of the three major groups of archaebacteria. J. Mol. Evol. **32:**70–78.
166. **Jahn, D., E. Verkamp, and D. Söll.** 1992. Glutamyl-transfer RNA: a precursor of heme and chorophyll biosynthesis. Trends Biochem. Sci. **17:**215–218.
167. **Jenal, U., T. Rechsteiner, P.-Y. Tan, E. Bühlmann, L. Meile, and T. Leisinger.** 1991. Isoleucyl-tRNA synthetase of *Methanobacterium thermoautotrophicum* Marburg: cloning of the gene, nucleotide sequence, and localization of a base change conferring resistance to pseudomonic acid. J. Biol. Chem. **266:**10570–10577.
168. **Jensen, R. A., and W. Gu.** 1996. Evolutionary recruitment of biochemically specialized subdivisions of family I within the protein superfamily of aminotransferases. J. Bacteriol. **178:**2161–2171.
169. **Joshi, P., and P. P. Dennis.** 1993. Characterization of paralogous and orthologous members of the superoxide dismutase gene family from genera of the halophilic archaebacteria. J. Bacteriol. **175:**1561–1571.
170. **Joshi, P., and P. P. Dennis.** 1993. Structure, function, and evolution of the family of superoxide dismutase proteins from halophilic archaebacteria. J. Bacteriol. **175:**1572–1579.
171. **Kakinuma, Y., K. Igarishi, K. Konishi, and I. Yamato.** 1991. Primary structure of the alpha-subunit of vacuolar-type $Na^+$-ATPase in *Enterococcus hirae*, amplification of a 1000 bp fragment by polymerase chain reaction. FEBS Lett. **292:**64–68.
172. **Kane, P. M., C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebi, and T. H. Stevens.** 1990. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. Science **250:**651–657.
173. **Kang, H. A., and J. W. B. Hershey.** 1994. Effect of initiation factor eIF-5A depletion on protein synthesis and proliferation of *Saccharomyces cerevisiae*. J. Biol. Chem. **269:**3934–3940.
174. **Kates, M.** 1993. Membrane lipids of Archaea, P. 261–295. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The biochemistry of Archaea (Archaebacteria). Elsevier Biomedical Press, Amsterdam, The Netherlands.
175. **Kates, M., D. J. Kushner, and A. T. Matheson.** 1993. The biochemistry of Archaea (Archaebacteria). Elsevier Biomedical Press, Amsterdam, The Netherlands.

176. **Kath, T., and G. Schafer.** 1995. A secY homologous gene in the crenarchaeon *Sulfolobus acidocaldarius*. Biochim. Biophys. Acta **1264:**55–158.
177. **Keeling, P. J., S. L. Baldauf, W. F. Doolittle, W. Zillig, and H.-P. Klenk.** 1996. An *infB*-homolog in *Sulfolobus acidocaldarius*. Syst. Appl. Microbiol. **19:**312–321.
178. **Keeling, P. J., and W. F. Doolittle.** 1995. Archaea: narrowing the gap between prokaryotes and eukaryotes. Proc. Natl. Acad. Sci. USA **92:**5761–5764.
179. **Keeling, P. J., and W. F. Doolittle.** 1996. Methionine aminopeptidase-1: the *MAP* of the mitochondrion. Trends Biochem. Sci. **21:**285–286.
180. **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.
181. **Keeling, P. J., R. L. Charlebois, and W. F. Doolittle.** 1994. Archaebacterial genomes: eubacterial form and eukaryotic content. Curr. Opin. Genet. Dev. **4:**816–822.
182. **Kengen, S. W. M., F. A. M. de Bok, N.-D. van Loo, C. Dijkema, A. J. M. Stams, and W. M. de Vos.** 1994. Evidence of 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.
183. **Kikuchi, A., T. Shibata, and S. Nakasu.** 1986. Reverse gyrase and DNA supercoiling in *Sulfolobus*. Syst. Appl. Microbiol. **7:**72–78.
184. **Kim, C. W., P. Markiewicz, J. J. Lee, C. F. Schierle, and J. H. Miller.** 1993. Studies of the hyperthermophile *Thermotoga maritima* by random sequencing of cDNA and genomic libraries: identification and sequencing of the *trpEG(D)* operon. J. Mol. Biol. **231:**960–981.
185. **Kim, S., K. R. Willison, and A. L. Horwich.** 1994. Cytosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. Trends Biochem. Sci. **19:**543–547.
186. **Kleman-Leyer, K., D. W. Armbruster, and C. J. Daniels.** 1997. Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. Cell **89:**839–847.
187. **Klenk, H.-P., and W. F. Doolittle.** 1994. Archaea and eukaryotes versus Bacteria? Curr. Biol. **4:**920–922.
188. **Klenk, H.-P., P. Palm, and W. Zillig.** 1993. DNA-dependent RNA polymerases as phylogenetic marker molecules. Syst. Appl. Microbiol. **16:**138–147.
189. **Klenk, H.-P., and W. Zillig.** 1994. DNA-dependent RNA polymerase subunit B as a tool for phylogenetic reconstructions: Branching topology of the archaeal domain. J. Mol. Biol. **38:**420–432.
190. **Klenk, H.-P., P. Palm, F. Lottspeich, and W. Zillig.** 1992. Component H of the DNA-dependent RNA polymerases of Archaea is homologous to a subunit shared by the three eucaryal nuclear RNA polymerases. Proc. Natl. Acad. Sci. USA **89:**407–410.
191. **Klenk, H.-P., C. Schleper, V. Schwass, and R. Brudler.** 1993. Nucleotide sequence, transcription and phylogeny of the gene encoding the superoxide dismutase of *Sulfolobus acidocaldarius*. Biochim. Biophys. Acta **1174:**95–98.
192. **Klenk, H.-P., W. Zillig, M. Lanzendörfer, B. Grampp, and P. Palm.** 1995. Location of protist lineages in a phylogenetic tree inferred from sequences of DNA-dependent RNA polymerases. Arch. Protistenkd. **145:**221–230.
193. **Kletzin, A.** 1992. Molecular characterization of a DNA ligase gene of the extreme thermophilic Archaeon *Desulfurolobus ambivalens* show close phylogenetic relationship to eukaryotic ligases. Nucleic Acids Res. **20:**5389–5396.
194. **Koga, Y., M. Nishihara, H. Morii, and M. Akagawa-Matsushita.** 1993. Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. Microbiol. Rev. **57:**164–182.
195. **Kohlhoff, M., A. Dahm, and R. Hensel.** 1996. Tetrameric triosephosphate isomerase from hyperthermophilic Archaea. FEBS Lett. **383:**245–250.
196. **Krömer, K. J., and E. Arndt.** 1991. Halobacterial S9 operon: three ribosomal protein genes are cotranscribed with genes encoding a $tRNA^{Leu}$, the enolase, and a putative membrane protein in the archaebacterium *Haloarcula* (*Halobacterium*) *marismortui*. J. Biol. Chem. **266:**24573–24579.
197. **Kumada, Y., D. R. Benson, D. Hillemann, T. J. Hosted, D. A. Rochefort, C. J. Thompson, W. Wohlleben, and Y. Tateno.** 1993. Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. Proc. Natl. Acad. Sci. USA **90:**3009–3013.
198. **Kumada, Y., E. Takano, K. Nagaoka, and C. J. Thompson.** 1990. *Streptomyces hygroscopicus* has two glutamine synthetase genes. J. Bacteriol. **172:**5343–5351.
199. **Lake, J. A.** 1988. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. Nature **331:**184–186.
200. **Lake, J. A., M. W. Clark, E. Henderson, S. P. Fay, M. Oakes, A. Scheinman, J. P. Thornber, and R. A. Mah.** 1985. Eubacteria, halobacteria, and the origin of photosynthesis: the photocytes. Proc. Natl. Acad. Sci. USA **82:**3716–3720.
201. **Lake, J. A., E. Henderson, M. Oakes, and M. W. Clark.** 1984. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc. Natl. Acad. Sci. USA **81:**3786–3790.
202. **Lake, J. A., and M. C. Rivera.** 1994. Was the nucleus the first endosymbiont? Proc. Natl. Acad. Sci. USA **91:**2880–2881.
203. **Lam, W. L., A. Cohen, D. Tsouluhas, and W. F. Doolittle.** 1990. Genes for tryptophan biosynthesis in the archaebacterium *Haloferax volcanii*. Proc. Natl. Acad. Sci. USA **87:**6614–6618.
204. **Lam, W. L., and W. F. Doolittle.** 1992. Mevinolin-resistant mutations identify a promoter and the gene for a eukaryote-like 3-hydroxy-3-methylglutaryl-coenzyme A reductase in the archaebacterium *Haloferax volcanii*. J. Biol. Chem. **267:**5829–5834.
205. **Lam, W. L., S. M. Logan, and W. F. Doolittle.** 1992. Genes for tryptophan biosynthesis in the halophilic archaebacterium *Haloferax volcanii*: the *trpDFEG* cluster. J. Bacteriol. **174:**1694–1697.
206. **Lamour, V., S. Quevillon, S. Diriong, V. C. N'Guyen, M. Lipinski, and M. Mirande.** 1994. Evolution of the Glx-tRNA synthetase family: the glutaminyl enzyme as a case of horizontal gene transfer. Proc. Natl. Acad. Sci. USA **91:**8670–8674.
207. **Langer, D., J. Hain, P. Thuriaux, and W. Zillig.** 1995. Transcription in Archaea: similarity to that in Eucarya. Proc. Natl. Acad. Sci. USA **92:**568–5772.
208. **Lapointe, J., L. Duplain, and M. Proulx.** 1986. A single glutamyl-tRNA synthetase aminoacylates $tRNA^{Glu}$ and $tRNA^{Gln}$ in *Bacillus subtilis* and efficiently misacylates *Escherichia coli* $tRNA^{Gln}$ in vitro. J. Bacteriol. **165:**88–93.
209. **Larkin, J. M., and M. C. Henk.** 1996. Filamentous sulfide-oxidizing bacteria at hydrocarbon seeps of the Gulf of Mexico. Microsc. Res. Tech. **33:**23–31.
210. **Lawson, F. S., R. L. Charlebois, and J.-A. R. Dillon.** 1996. Phylogenetic analysis of carbamoylphosphate synthetase genes: evolution involving multiple gene duplications, gene fusions, and insertions and deletions of surrounding sequences. Mol. Biol. Evol. **13:**970–977.
211. **Latimer, M. T., and J. G. Ferry.** 1993. Cloning, sequence analysis, and hyperexpression of the genes encoding phosphotransacetylase and acetate kinase from *Methanosarcina thermophila*. J. Bacteriol. **175:**6822–6829.
212. **Lazendörfer, M., D. Langer, J. Hain, H.-P. Klenk, I. Holz, I. Arnold-Ammer, and W. Zillig.** 1993. Structure and function of the DNA-dependent RNA polymerase of *Sulfolobus*. Syst. Appl. Microbiol. **16:**156–164.
213. **Lazendörfer, M., P. Palm, B. Grampp, D. A. Peattie, and W. Zillig.** 1992. Nucleotide sequence of the gene encoding the largest subunit of the DNA-dependent RNA polymerase III of *Giardia lamblia*. Nucleic Acids Res. **20:**1145.
214. **Lefeers, H., F. Gropp, F. Lottspeich, W. Zillig, and R. A. Garrett.** 1989. Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. Mol. Biol. **206:**1–17.
215. **Lewin, B.** 1994. Genes V. Oxford University Press, Oxford, United Kingdom.
216. **Leyva-Vazquez, M. A., and P. Setlow.** 1994. Cloning and nucleotide sequences of the genes encoding triose phosphate isomerase, phosphoglycerate mutase, and enolase from *Bacillus subtilis*. J. Bacteriol. **176:**3903–3910.
217. **Li, W.-H. and Graur, D.** 1991. Fundamentals of molecular evolution. Sinauer Associates. Sunderland, Mass.
218. **Liao, D., and P. P. Dennis.** 1994. Molecular phylogenies based on ribosomal protein L11, L1, L10, and L12 sequences. J. Mol. Evol. **38:**405–419.
219. **Lill, U., S. Lefrank, A. Henschen, and H. Eggerer.** 1992. Conversion, by limited proteolysis, of an archaebacterial citrate synthase into essentially a citryl-CoA hydrolase. Eur. J. Biochem. **208:**459–466.
220. **Lindmark, D. G.** 1980. Energy metabolism of the anaerobic protozoon *Giardia lamblia*. Mol. Biochem. Parasitol. **1:**1–12.
221. **Longstaff, M., C. A. Raines, E. M. McMorrow, J. W. Bradbeer, and T. A. Dyer.** 1989. Wheat phosphoglycerate kinase: evidence for recombination between the genes for the chloroplastic and cytosolic enzymes. Nucleic Acids Res. **17:**6569–6580.
222. **Lupas, A., F. Zühl, T. Tamura, S. Wolf, I. Nagy, R. De Mot, and W. Baumeister.** 1997. Eubacterial proteasomes. Mol. Biol. Rep. **24:**125–131.
223. **Lupas, A., P. Zwickl, and W. Baumeister.** 1994. Proteasome sequences in eubacteria. Trends Biochem. Sci. **19:**533–534.
224. **Margulis, L., and R. Guerreo.** 1991. Kingdoms in turmoil. New Sci. **129:**46–50.
225. **Markos, A., A. Miretsky, and M. Müller.** 1993. A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote, *Trichomonas vaginalis*. J. Mol. Evol. **37:**631–643.
226. **Marsh, C. L., C. I. Reich, R. B. Whitlock, and G. J. Olsen.** 1994. Transcription factor IID in the Archaea: sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA-binding protein of eukaryotes. Proc. Natl. Acad. Sci. USA **91:**4180–4184.
227. **Martin, W.** 1996. Is something wrong with the tree of life? Bioessays **18:**523–527.
228. **Martin, W.** Personal communication.
229. **Martin, W., H. Brinkman, C. Savonna, and R. Cerff.** 1993. Evidence for a chimeric nature of nuclear genomes: Eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proc. Natl. Acad. Sci. USA **90:**8692–8696.
230. **Maupin-Furlow, J. A., and J. G. Ferry.** 1995. A proteasome for the methanogenic archaeon *Methanosarcina thermophila*. J. Biol. Chem. **270:**28617–28622.

231. **May, B. P., and P. P. Dennis.** 1990. Unusual evolution of a superoxide dismutase-like gene from the extremely halophilic archaebacterium *Halobacterium cutirubrum*. J. Bacteriol. **172:**3725–3729.
232. **Mayr, E.** 1990. A natural system of organisms. Nature **348:**491.
233. **Mehta, P. K., T. I. Hale, and P. Christen.** 1993. Aminotransferases: demonstration of homology and division into evolutionary subgroups. Eur. J. Biochem. **241:**549–561.
234. **Meile, L., R. Stettler, R. Banholzer, M. Kotik, and T. Leisinger.** 1991. Tryptophan gene cluster of *Methanobacterium thermoautotrophicum* Marburg: molecular cloning and nucleotide sequence of a putative *trpEGCF-BAD* operon. J. Bacteriol. **173:**5017–5023.
235. **Michels, P. A. M., M. Marchand, L. Kohl, S. Allert, R. K. Wierenga, and F. R. Opperdoes.** 1991. The cytosolic and glycosomal isoenzymes of glyceraldehyde-3-phosphate dehydrogenase in *Trypanosoma brucei* have a distant evolutionary relationship. Eur. J. Biochem. **198:**421–428.
236. **Moras, D.** 1992. Structural and functional relationships between aminoacyl-tRNA synthetases. Trends Biochem. Sci. **17:**159–164.
237. **Morris, C. J., and J. N. Reeve.** 1988. Conservation of structure in the human gene encoding argininosuccinate synthetase and *argG* genes of the archaebacteria *Methanosacrina barkeri* MS and *Methanococcus vannielii*. J. Bacteriol. **170:**3125–3130.
238. **Muir, J. M., R. J. M. Russel, D. W. Hough, and M. J. Danson.** 1995. Citrate synthase from the hyperthermophilic Archaeon *Pyrococcus furiosus*. Protein Eng. **8:**583–592.
239. **Musgrave, D. R., K. M. Sandman, and J. N. Reeve.** 1991. DNA binding by the archaeal histone HMf in positive supercoiling. Proc. Natl. Acad. Sci. USA **88:**10397–10401.
240. **Nagel, G. M., and R. F. Doolittle.** 1991. Evolution and relatedness in two aminoacyl-tRNA synthetase families. Proc. Natl. Acad. Sci. USA **88:**8121–8125.
241. **Nagel, G. M., and R. F. Doolittle.** 1995. Phylogenetic analysis of the aminoacyl-tRNA synthetases. J. Mol. Evol. **40:**487–498.
242. **Neumann-Spallart, C., J. Jakowitsch, M. Kraus, M. Brandtner, H. J. Bohnert, W. Löffelhardt.** 1991. *rps10*, unreported for plastid DNAs, is located on the cyanelle genome of *Cyanophora paradoxa* and is cotranscribed with the *str* operon genes. Curr. Genet. **19:**313–315.
243. **Nichols, B. P.** 1996. Evolution of genes and enzymes of tryptophan biosynthesis, p. 2638–2648. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
244. **Nolling, J., and W. M. Vos.** 1992. Characterization of the archaeal, plasmid-encoded type II restriction-modification system MthTI from *Methanobacterium thermoformicicum* THF: homology to the bacterial NgoPII system from *Neisseria gonorrhoeae*. J. Bacteriol. **174:**5719–5726.
245. **Nyunoya, H., and C. J. Lusty.** 1983. The *carB* gene of *Escherichia coli*: a duplicated gene coding for the large subunit of carbamoyl-phosphate synthetase. Proc. Natl. Acad. Sci. USA **80:**4629–4633.
246. **Olsen, G. J., and C. R. Woese.** 1993. Ribosomal RNA: a key to phylogeny. FASEB J. **7:**113–123.
247. **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.
248. **Osteryoung, K. W., and E. Vierling.** 1995. Conserved cell and organelle division. Nature **376:**473–474.
249. **Ouzounis, C., and C. Sander.** 1992. TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes. Cell **71:**189–190.
250. **Palmer, J. D., and J. M. Logsdon, Jr.** 1991. The recent origins of introns. Curr. Opin. Genet. Dev. **1:**470–477.
251. **Palmer, J. R., and J. N. Reeve.** 1993. Structure and function of methanogen genes, p. 497–534. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The biochemistry of Archaea (Archaebacteria). Elsevier Biomedical Press, Amsterdam, The Netherlands.
252. **Patee, P. A., H. C. Lee, and J. P. Bannantine.** 1990. Genetic and physical mapping of *Staphylococcus aureus*, p. 42–56. *In* R. P. Novick (ed.), Molecular biology of the staphylococci. VCH Publishers, New York, N.Y.
253. **Paulus, T. J., and R. L. Switzer.** 1989. Characterization of pyrimidine-repressible and arginine-repressible carbamoyl phosphate synthetases from *Bacillus subtilis*. J. Bacteriol. **137:**82–91.
254. **Perler, F. B., D. G. Comb, W. E. Jack, L. S. Moran, B. Qiang, R. B. Kucera, J. Benner, B. E. Slatko, D. O. Nwankwo, S. K. Hempstead, C. K. S. Carlow, and H. Jannasch.** 1992. Intervening sequences in an Archaea DNA polymerase gene. Proc. Natl. Acad. Sci. USA **89:**5577–5581.
255. **Pesole, G., M. P. Bozzetti, C. Lanave, G. Preparata, and C. Saccone.** 1991. Glutamine synthetase gene evolution: a good molecular clock. Proc. Natl. Acad. Sci. USA **88:**522–526.
256. **Pisani, F. M., C. De Martino, and M. Rossi.** 1992. A DNA polymerase from the Archaeon *Sulfolobus solfataricus* show sequences similarity to family B DNA polymerases. Nucleic Acids Res. **20:**2711–2716.
257. **Possot, O., L. Sibold, and J. P. Aubert.** 1989. Nucleotide sequence and expression of the glutamine synthetase gene, *glnA* of the Archaea *Methanococcus voltae*. Res. Microbiol. **140:**355–371.
258. **Potter, S., P. Durovic, and P. P. Dennis.** 1995. Ribosomal RNA precursor processing by a eukaryotic U3 small nucleolar RNA-like molecule in an Archaeon. Science **268:**1056–1060.
259. **Potter, S., and L. A. Fothergill-Gilmore.** 1992. Purification and properties of pyruvate kinase from *Thermoplasma acidophilum*. FEMS Microbiol. Lett. **73:**235–239.
260. **Prangishvili, D., and H.-P. Klenk.** 1993. The gene for a 74 kDa DNA polymerase from the Archaeon *Sulfolobus solfataricus*. Syst. Appl. Microbiol. **16:**165–171.
261. **Preston, C. M., K. Y. Wu, T. F. Molinski, and E. F. DeLong.** 1996. A psychrophilic crenarchaeon inhabits a marine sponge: *Crenarchaeum symbiosum* gen. nov., sp. nov. Proc. Natl. Acad. Sci. USA **93:**6241–6246.
262. **Pruss, B., H. E. Meyer, and A. W. Holldorf.** 1993. Characterization of the glyceraldehyde 3-phosphate dehydrogenase from the extremely halophilic archaebacterium *Haloarcula vallismortis*. Arch. Microbiol. **160:**5–11.
263. **Pühler, G., H. Leffers, F. Gropp, P. Palm, H.-P. Klenk, F. Lottspeich, R. A. Garrett, and W. Zillig.** 1989. Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc. Natl. Acad. Sci. USA **86:**4569–4573.
264. **Pühler, G., F. Lottspeich, and W. Zillig.** 1989. Organization and nucleotide sequences of the genes encoding the large subunits A, B and C of the DNA-dependent RNA polymerase of the archaebacterium *Sulfolobus acidocaldarius*. Nucleic Acids Res. **17:**4517–4534.
265. **Pühler, G., F. Pitzer, P. Zwickl, and W. Baumeister.** 1994. Proteasomes: multisubunit common to *Thermoplasma* and eukaryotes. Syst. Appl. Microbiol. **16:**234–241.
266. **Qureshi, S. A., B. Khoo, P. Baumann, and S. P. Jackson.** 1995. Molecular cloning of the transcription factor TFIIB homolog from *Sulfolobus shibatae*. Proc. Natl. Acad. Sci. USA **92:**6077–6081.
267. **Ragan, M. A., J. M. Logsdon Jr., C. W. Sensen, R. L. Charlebois, and W. F. Doolittle.** 1996. An archaebacterial homolog of pelota, a meiotic cell division protein in eukaryotes. FEMS Microbiol. Lett. **144:**151–155.
268. **Ramírez, C., A. K. E. Köpke, D.-C. Yang, T. Boeckh, and A. T. Matheson.** 1993. The structure, function and evolution of archaeal ribosomes, p. 439–466. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The biochemistry of Archaea (Archaebacteria). Elsevier Biomedical Press, Amsterdam, The Netherlands.
269. **Ramírez, C., L. C. Shimmin, P. Leggatt, and A. T. Matheson.** 1994. Structure and transcription of the L11-L1-L10-L12 ribosomal protein gene operon from the extreme thermophilic Archaeon, *Sulfolobus acidocaldarius*. J. Mol. Biol. **244:**242–249.
270. **Rashid, N., M. Morikawa, K. Nagahisa, S. Kanaya, and T. Imanaka.** 1997. Characterization of a RecA/Rad51 homologue from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1. Nucleic Acids Res. **25:**719–726.
271. **Rensing, S. A., and U.-G. Maier.** 1994. The SecY protein family: comparative analysis and phylogenetic relationships. Mol. Phylogenet. Evol. **3:**187–191.
272. **Reeve, J. N.** 1992. Molecular biology of methanogens. Annu. Rev. Microbiol. **46:**165–191.
273. **Reeves, J. N., K. Sandman, and C. J. Daniels.** 1997. Archaeal histones, nucleosomes and transcription initiation. Cell **89:**999–1002.
274. **Reeves, R. E., L. G. Warren, B. Suskind, and H. S. Lo.** 1977. An energy-conserving pyruvate-to-acetate pathway in *Entamoeba histolytica*: pyruvate synthase and a new acetate thiokinase. J. Biol. Chem. **252:**726–731.
275. **Reiter, W.-D., U. Hüdepohl, and W. Zillig.** 1990. Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection *in vitro*. Proc. Natl. Acad. Sci. USA **87:**9509–9513.
276. **Ribas de Pouplana, L., M. Furgier, C. L. Quinn, and P. Schimmel.** 1996. Evidence that two present-day components needed for the genetic code appeared after nucleated cells separated from eubacteria. Proc. Natl. Acad. Sci. USA **93:**166–170.
277. **Riley, M.** 1993. Functions of the gene products of *Escherichia coli*. Microbiol. Rev. **57:**862–952.
278. **Rivera, M. C., and J. A. Lake.** 1992. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science **257:**74–76.
279. **Robb, F. T., and R. Weiss.** Unpublished data.
280. **Rochefort, D. A., and D. R. Benson.** 1990. Molecular cloning, sequencing and expression of the glutamine synthetase II (*glnII*) gene from the actinomycete root nodule symbiont *Frankia* sp. strain CpI1. J. Bacteriol. **172:**5335–5342.
281. **Roger, A. J., and J. R. Brown.** 1996. A chimeric origin for eukaryotes re-examined. Trends Biochem. Sci. **21:**370–371.
282. **Roger, A. J., C. G. Clark, and W. F. Doolittle.** 1996. A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. Proc. Natl. Acad. Sci. USA **93:**14618–14622.
283. **Rowlands, T., P. Baumann, and S. P. Jackson.** 1994. The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. Science **264:**1326–1329.
284. **Russell, A. G., M. Moniz de Sa, and P. Dennis.** 1997. A U3-like small nucleolar RNA in Archaea: retraction. Science **277:**1189.
285. **Russell, R. J. M., D. W. Hough, M. J. Danson, and G. L. Taylor.** 1994. The

crystal structure of citrate synthase from the thermophilic Archaeon, *Thermoplasma acidophilum*. Structure **2:**1157–1167.
286. **Saier, M. H., Jr., M. J. Fagan, C. Hoischen, and J. Reizer.** 1993. Transport mechanisms. p. 133–156. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
287. **Sanangelantoni, A. M., and O. Tiboni.** 1993. The chromosomal location of genes for elongation factor Tu and ribosomal protein S10 in the cyanobacterium *Spirulina platensis* provides clues to the ancestral organization of the *str* and S10 operons in prokaryotes. J. Gen. Microbiol. **139:**2579–2584.
288. **Sanangelantoni, A. M., D. Barbarini, G. Di Pasquale, P. Cammarano, and O. Tiboni.** 1990. Cloning and sequencing of an archaeal glutamine synthetase gene: phylogenetic implications. Mol. Gen. Genet. **221:**187–194.
289. **Sandler, S. J., L. H. Satin, H. S. Samra, and A. J. Clark.** 1996. *recA*-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res. **24:**2125–2132.
290. **Sandman, K., J. A. Krzycki, B. Dobrinski, R. Lurz, and J. N. Reeve.** 1990. HMf, a DNA-binding protein isolated from the hyperthermophilic Archaeon *Methanothermus fervidus*, is most closely related to histones. Proc. Natl. Acad. Sci. USA **87:**5788–5791.
291. **Sarich, V. M., and A. C. Wilson.** 1973. Generation time and genomic evolution in primates. Science **179:**1144–1147.
292. **Schäfer, T., K. B. Xavier, H. Santos, and P. Schönheit.** 1994. Glucose fermentation to acetate and alanine in resting cell suspensions of *Pyrococcus furiosus*: proposal of a novel glycolytic pathway based on $^{13}C$ labelling data and enzyme activities. FEMS Microbiol. Lett. **121:**107–114.
293. **Schofield, J. P.** 1993. Molecular studies on an ancient gene encoding for carbamoyl-phosphate synthetase. Clin. Sci. **84:**119–128.
294. **Schönheit, P., and T. Schäfer.** 1995. Metabolism of hyperthermophiles. World J. Microbiol. Biotechnol. **11:**26–75.
295. **Shatters, R. G., and M. L. Kahn.** 1989. Glutamine synthetase II in *Rhizobium*: reexamination of the proposed horizontal transfer of DNA from eukaryotes to prokaryotes. J. Mol. Evol. **29:**422–428.
296. **Shimmin, L. C., and P. P. Dennis.** 1989. Characterization of the L11, L1, L10 and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium *Halobacterium cutirubrum*. EMBO J. **8:**1225–1235.
297. **Shimmin, L. C., C. Ramirez, A. T. Matheson, and P. P. Dennis.** 1989. Sequence alignment and evolutionary comparison of the L10 equivalent and L12 equivalent ribosomal proteins from archaebacteria, eubacteria and eucaryotes. J. Mol. Evol. **29:**448–462.
298. **Sibold, L., and M. Henriquet.** 1988. Cloning of the *trp* genes from the archaebacterium *Methanococcus voltae*: nucleotide sequence of the *trpBA* genes. Mol. Gen. Genet. **214:**439–450.
299. **Siebers, B., and R. Hensel.** 1993. Glucose catabolism of the hyperthermophilic archaeon *Thermoproteus tenax*. FEMS Microbiol. Lett. **111:**1–8.
300. **Slesarev, A. I., K. O. Stetter, J. A. Lake, M. Gellert, R. Krah, and S. A. Kozyavkin.** 1993. DNA topoisomerase V is a relative of eukaryotic topoisomerase I from a hyperthermophilic prokaryote. Nature **364:**735–736.
301. **Smith, D. W., and R. F. Doolittle.** 1992. A comparison of evolutionary rates of the two major kinds of superoxide dismutase. J. Mol. Evol. **34:**175–184.
302. **Smith, M. W., D.-F. Feng, and R. F. Doolittle.** 1992. Evolution by acquisition: the case for horizontal gene transfers. Trends Biochem. Sci. **17:**489–493.
303. **Sogin, M. L.** 1989. Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. Am. Zool. **29:**487–499.
304. **Sogin, M. L.** 1991. Early evolution and the origin of eukaryotes. Curr. Opin. Genet. Dev. **1:**457–463.
305. **Sorokine, I., K. Ben-Mahrez, M. Nakayama, and M. Kohiyama.** 1991. Exonuclease activities associated with DNA polymerases α and β of the archaebacterium *Halobacterium halobium*. Eur. J. Biochem. **197:**781–784.
306. **Stanier, R. Y.** 1970. Some aspects of the biology of cells and their possible evolutionary significance. Symp. Soc. Gen. Microbiol. **20:**1–38.
307. **Stanier, R. Y., and C. B. van Niel.** 1941. The main outlines of bacterial classification. J. Bacteriol. **42:**437–466.
308. **Stanier, R. Y., and C. B. van Niel.** 1962. The concept of a bacterium. Arch. Microbiol. **42:**17–35.
309. **Starich, M. R., K. Sandman, J. N. Reeve, and M. F. Summers.** 1996. NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. J. Mol. Biol. **255:**187–203.
310. **Stein, J. L., and M. I. Simon.** 1996. Archaeal ubiquity. Proc. Natl. Acad. Sci. USA **93:**6228–6230.
311. **Stöffler, G., and M. Stöffler-Meilicke.** 1986. Electron microscopy of archaebacterial ribosomes. Syst. Appl. Microbiol. **7:**123–130.
312. **Stöffler-Meilicke, M., C. Böhme, O. Strobel, A. Böck, and G. Stöffler.** 1986. Structure of ribosomal subunits of *M. vannielii*: ribosomal morphology as a phylogenetic marker. Science **231:**1306–1308.
313. **Sumi, M., M. H. Sato, K. Denda, T. Date, and M. Yoshida.** 1992. A DNA fragment homologous to $F_1$-ATPase β subunit amplified from genomic DNA of *Methanosarcina barkeri*: indication of an archaebacterial F-type ATPase. FEBS Lett. **314:**207–210.
314. **Sutherland, K. J., C. M. Henneke, P. Towner, D. W. Hough, and M. J. Danson.** 1990. Citrate synthase from the thermophilic archaebacterium *Thermoplasma acidophilum*: cloning and sequencing of the gene. Eur. J. Biochem. **194:**839–844.
315. **Takao, M., T. Kobayashi, A. Oikawa, and A. Yasui.** 1989. Tandem arrangement of photolyase and superoxide dismutase genes in *Halobacterium halobium*. J. Bacteriol. **171:**6323–6329.
316. **Takao, M., A. Yasui, and A. Oikawa.** 1991. Unique characteristics of superoxide dismutase of a strictly anaerobic archaebacterium *Methanobacterium thermoautotrophicum*. J. Biol. Chem. **266:**14151–14154.
317. **Tannich, E., I. Bruchhaus, R. D. Walter, and R. D. Horstmann.** 1991. Pathogenic and nonpathogenic *Entamoeba histolytica*: identification and molecular cloning of an iron-containing superoxide dismutase. Mol. Biochem. Parasitol. **49:**61–72.
318. **Thomm, M.** 1996. Archaeal transcription factors and their role in transcription initiation. FEMS Microbiol. Rev. **18:**159–71.
319. **Tiboni, O., P. Cammarano, and A. M. Sanangelantoni.** 1993. Cloning and sequencing of the gene encoding glutamine synthetase I from the archaeon *Pyrococcus woesei*: anomalous phylogenies inferred from analysis of archaeal and bacterial glutamine synthetase I sequences. J. Bacteriol. **175:**2961–2969.
320. **Tiboni, O., R. Cantoni, R. Creti, P. Cammarano, and A. M. Sanangelantoni.** 1991. Phylogenetic depth of *Thermotoga maritima* inferred from analysis of the *fus* gene: amino acid sequence of elongation factor G and organization of the *Thermotoga str* operon. J. Mol. Evol. **33:**142–151.
321. **Trotta, C. R., F. Miao, E. A. Arn, S. W. Stevens, C. K. Ho, R. Rauhut, and J. N. Abelson.** 1997. The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. Cell **89:**849–858.
322. **Tsutsumi, S., K. Denda, K. Yokoyama, T. Oshima, T. Date, and M. Yoshida.** 1991. Molecular cloning of genes encoding major subunits of a eubacterial V-type ATPase from *Thermus thermophilus*. Biochim. Biophys. Acta **1098:**13–20.
323. **Tutino, M. L., G. Scarano, G. Marino, G. Sannia, and M. V. Cubellis.** 1993. Tryptophan biosynthesis genes *trpEGC* in the thermoacidophilic archaebacterium *Sulfolobus solfataricus*. J. Bacteriol. **175:**299–302.
324. **Verkamp, E., A. M. Kumar, A. Lloyd, O. Martins, N. Stange-Thomann, and D. Söll.** 1995. Glutamyl-tRNA as an intermediate in glutamate conversions, p. 545–550. *In* D. Söll and U. RajBhandary (ed.), tRNA: structure, biosynthesis, and function. American Society for Microbiology, Washington, D.C.
325. **Viale, A. M., and A. K. Arakaki.** 1994. The chaperone connection to the origins of eukaryotic organelles. FEBS Lett. **341:**146–151.
326. **Wang, J. C., P. R. Caron, and R. A. Kim.** 1990. The role of DNA topoisomerases in recombination and genome stability: a double-edged sword? Cell **10:**403–406.
327. **Wang, X., and J. Lutkenhaus.** 1996. FtsZ ring: the eubacterial division apparatus conserved in archaebacteria. Mol. Microbiol. **21:**313–319.
328. **Watson, R. J., and V. K. Rastogi.** 1993. Cloning and nucleotide sequencing of *Rhizobium meliloti* aminotransferase genes: an aspartate aminotransferase required for symbiotic nitrogen fixation is atypical. J. Bacteriol. **175:**1919–1928.
329. **Weng, M., C. A. Makaroff, and H. Zalkin.** 1986. Nucleotide sequence of *Escherichia coli pyrG* encoding CTP synthetase. J. Biol. Chem. **261:**5568–5574.
330. **Wenzel, T., and W. Baumeister.** 1993. Thermoplasma acidophilum proteasomes degrades partially unfolded and ubiquitin-associated proteins. FEBS Lett. **326:**215–218.
331. **Wettach, J., H. P. Gohl, H. Tschochner, and M. Thomm.** 1995. Functional interaction of yeast and human TATA-binding proteins with an archaeal RNA polymerase and promoter. Proc. Natl. Acad. Sci. USA **92:**472–476.
332. **White, B. N., and S. T. Bradley.** 1972. Further codon assignments in an extremely halophilic bacterium using a cell-free protein-synthesizing system and a ribosomal binding assay. Can. J. Biochem. **50:**600–609.
333. **Winefield, C. S., K. J. F. Farnden, P. H. S. Reynolds, and C. J. Marshall.** 1995. Evolutionary analysis of aspartate aminotransferases. J. Mol. Evol. **40:**455–463.
334. **Wittmann, H. G.** 1986. Structure of ribosomes, p. 1–27. *In* B. Hardesty and G. Kramer (ed.), Structure, function, and genetics of ribosomes. Springer-Verlag, New York, N.Y.
335. **Wittmann-Liebold, B., A. K. E. Köpke, E. Arndt, W. Krömer, T. Hatakeyama, and H.-G. Wittmann.** 1990. Sequence comparison and evolution of ribosomal proteins and their genes, p. 598–616. *In* W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger, and J. R. Warner (ed.), The ribosome: structure, function and evolution. American Society for Microbiology, Washington, D.C.
336. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
337. **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA **51:**221–271.
338. **Woese, C. R., and G. E. Fox.** 1977. The concept of cellular evolution. J. Mol. Evol. **10:**1–6.
339. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eu-*

*carya*. Proc. Natl. Acad. Sci. USA **87:**4576–4579.
340. **Yajima, H., H. Inoue, A. Oikawa, and A. Yasui.** 1991. Cloning and functional characterization of a eucaryotic DNA photolyase gene from *Neurospora crassa*. Nucleic Acids Res. **19:**5359–5362.
341. **Yasui, A., A. P. M. Eker, S. Yasuhira, H. Yajima, T. Kobayashi, M. Takao, and A. Oikawa.** 1994. A new class of DNA photolyases present in various organisms including a placental mammals. EMBO J. **13:**6143–6151.
342. **Zillig, W.** 1991. Comparative biochemistry of Archaea and Bacteria. Curr. Opin. Genet. Dev. **1:**457–463.
343. **Zillig, W., H.-P. Klenk, P. Palm, H. Leffers, G. Pühler, and R. A. Garrett.** 1989. Did eukaryotes originate by a fusion event? Endocytobiosis Cell Res. **6:**1–25.
344. **Zillig, W., P. Palm, H.-P. Klenk, G. Pühler, F. Gropp, and C. Schleper.** 1991. Phylogeny of DNA-dependent RNA polymerases: testimony for the origin of eukaryotes, p. 321–332. *In* F. Rodriguez-Valera (ed.), General and applied aspects of halophilic microorganisms. Plenum Press, New York, N.Y.
345. **Zillig, W., P. Palm, H.-P. Klenk, D. Langer, U. Hüdepohl, J. Hain, M. Lanzendörfer, and I. Holz.** 1993. Transcription in Archaea, p. 367–391. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The biochemistry of Archaea (Archaebacteria). Elsevier Biomedical Press, Amsterdam, The Netherlands.
346. **Zusman, T., I. Rosenshine, G. Boehm, R. Jaenicke, B. Leskiw, and M. Mevarech.** 1989. Dihydrofolate reductase of the extremely halophilic archaebacterium *Halobacterium volcanii*. J. Biol. Chem. **264:**18878–18883.
347. **Zwickl, P., S. Fabry, C. Bogedain, A. Haas, and R. Hensel.** 1990. Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaebacterium *Pyrococcus woesei*: characterization of the enzyme, cloning and sequencing of the gene, and expression in *Escherichia coli*. J. Bacteriol. **172:**4329–4338.
348. **Zwickl, P., A. Grizwa, G. Pühler, B. Dahlmann, F. Lottspeich, and W. Baumeister.** 1992. Primary structure of the *Thermoplasma* proteasome and its implications for the structure, function, and evolution of the multicatalytic proteinase. Biochemistry **31:**964–972.
349. **Zwickl, P., F. Lottspeich, B. Dahlmann, and W. Baumeister.** 1991. Cloning and sequencing of the gene encoding the large (α-) subunit of the proteasome from *Thermoplasms acidophilum*. FEBS Lett. **278:**217–221.