Syst. Biol. 50(4):497–512 , 2001

Genomic and Phylogenetic Perspectives on the Evolution of Prokaryotes

JAMES R. BROWN

Microbial Bioinformatics Department, Division of Bioinformatics, Glaxo SmithKline Pharmaceuticals, 1250 South Collegeville Road, P.O. Box 5089, UP1345, Collegeville, Pennsylvania 19426-0989 , USA; E-Mail: James_R_Brown@gsk.com

Abstract.—Prokaryotes have been at the forefront of the genome sequencing revolution. Many genomes have been completely sequenced, revealing much about bacterial and archaeal genome content and organization. Yet, a meaningful evolutionary picture of prokaryotes still eludes us. Much of the problem lies in understanding the mode and tempo of genome evolution. Here phenylalanyl-tRNA synthetase is used as an example of the complex interplay among lateral gene transfer, operon recombination, and gene recruitment in the evolution of some prokaryotic genes. Promising new approaches to genomic analyses, which could add to our understanding prokaryotic evolution and help in their classification, are discussed. [Archaea; comparative genomics; eubacteria; phenylalanyl-tRNA synthetase; universal tree.]

An understanding of the evolutionary relationships among prokaryotes—the most abundant and diverse life forms on the planet—is essential. The important impact of prokaryotes on human health and economy as well as their key role in eukaryotic evolution makes prokaryotic evolution a particularly compelling study subject. Yet, the determination of definitive relationships among various prokaryotic groups and of the specific steps in the transition from the prokaryotic to eukaryotic cell still eludes us.

In principle, metazoan systematics has the advantage of several different potential sources of collaborative evidence, including organism morphology, ontogeny, paleontology, and molecular evolution. However, studies of prokaryotic evolution are so reliant on the latter that advances in molecular techniques and DNA sequence databases could be used to define periods or eras of growth in the field.

THE PREMOLECULAR ERA

Although earlier observations of microbes abound, the Premolecular Era of prokaryotic evolution truly began roughly 60 years ago, with Chatton (1937) and Stanier and van Niel (1941) both suggesting that life could be subdivided into two fundamental cellular categories, prokaryotes and eukaryotes (summarized in Doolittle and Brown, 1994). The distinction between these two groups was subsequently refined as studies of cellular biology and genetics progressed, such that prokaryotes became universally distinguishable from eukaryotes on the basis of missing internal membranes (such as the nuclear membrane and endoplasmic recticulum), nuclear division by fission rather than mitosis, and the presence of a cell wall (Stanier and van Niel, 1962; Stanier, 1970).

The pinnacle of intellectual achievement in the Premolecular Era was the endosymbiont hypothesis, as formulated by Margulis (1970), which linked the origin of eukaryotic organelles (hence eukaryotes themselves) to specific groups of prokaryotes, the cyanobacteria and proteobacteria. Although the endosymbiotic origin of organelles is now widely accepted, the mechanisms and extent patterns of direct prokaryotic contributions to the eukaryotic cell are still being explored today. New and important expansions to the endosymbiosis hypothesis have been recently proposed (Doolittle, 1998; Martin and Müller, 1998; Moreira and López-Gracía, 1998). But until the late 1970s, the lack of complex morphological characters clearly hindered any further advances in understanding prokaryote evolution.

THE MOLECULAR ERA

The development of phylogenetic informative data sets heralds the birth of the Molecular Era. In the late 1970s, Woese and coworkers developed new molecular methods that challenged the fundamental prokaryote–eukaryote dichotomy. By digesting in vivo labeled 16S rRNA by using T1

ribonuclease and then cumulating and comparing catalogues of the resulting oligonucleotide "words", they were able to derive dendograms showing the relationships between different bacterial species. Analyses involving some unusual methanogenic bacteria revealed surprising and unique species clusters among prokaryotes (Fox et al., 1977). So deep was the split in the prokaryotes that Woese and Fox (1977) proposed to call the methanogens and their relatives "archaebacteria", a name relaying their distinctness from the true bacteria or "eubacteria" as well as contemporary preconceptions that these organisms might have thrived in the environmental conditions of a younger Earth.

Phylogenies based on rRNA molecules held the promise of acting as the ultimate scaffolding for complete hierarchical classifications of prokaryotes (Woese, 1987; Olsen et al., 1994) and eukaryotes (Cavalier-Smith, 1993) alike. In 1990, Woese, Kandler, and Wheelis formally proposed the replacement of the bipartite view of life with a new tripartite scheme based on three urkingdoms or domains: the Bacteria (formerly eubacteria), Archaea (formerly archaebacteria) and Eucarya (eukaryotes, still the more often used name). The rationale behind this revision came from a growing body of biochemical, genomic, and phylogenetic evidence that, viewed collectively, suggested the archaebacteria were unique and distinct from eukaryotes and eubacteria. Although most archaebacterialogists supported this view, other workers dissented in the elevation of two prokaryotic groups to be taxonomic equals of eukaryotes (reviewed in Doolittle and Brown, 1994; Brown and Doolittle, 1997).

At the center of the controversy surrounding the three domain concepts are the archaebacteria. According to rRNA trees, the Archaea include at least two well-defined subgroups, the kingdoms Crenarchaeota and Euryarchaeota (Woese, 1987), and possibly more (Barnes et al., 1994, 1996). Because early research focused on understanding archaeal adaptations to extreme environments, species of Crenarchaeota were initially characterized as being thermoacidophiles, whereas species of Euryarchaeota were a mix of hyperthermophiles, halophiles, methanogens, and even thermophilic methanogens. Recent surveys using the polymerase chain reaction amplification of rRNA sequences from microbes in the environment have made important advances in characterizing the biodiversity of microbes. The result has been the discovery of the pandemic distribution of archaeal and bacterial rRNA signatures in a wide range of mesophilic as well as "extremophilic" environments, thus breaking down the traditional metabolism-based classifications of the Archaea (i.e., Fuhrman et al., 1992; DeLong, 1992; DeLong et al., 1994; Stein and Simon, 1996).

Before complete genome sequences became available, considerable information had been accumulating on the comparative biochemistry and cellular and molecular biology of the Archaea (reviewed in Brown et al., 1989; Zillig, 1991; Danson, 1993; Kates et al., 1993; Keeling et al., 1994; Brown and Doolittle, 1997). At a cellular level, the Archaea can be defined as having only a few unique biochemical and genetic traits, whereas many of their characteristics were once thought to be solely bacterial or eukaryotic. Their most prominent novel features include isopranyl ether lipids, the absence of acyl ester lipids and fatty acid synthetase, modified tRNA molecules, and a split gene coding one of the RNA polymerase subunits. Also, Archaea use a variety of metabolic regimes, which deviate from known metabolic pathways of Bacteria and eukaryotes, and are not simply particular environmental adaptations.

Archaea and Bacteria are united in the "realm of prokaryotes" by their generally similar cell sizes, lack of a nuclear membrane and organelles, organization of certain genes into operons, common metabolic pathways (most often found among hyperthermophiles of both groups), and having a large circular chromosome, occasionally accompanied by one or more smaller circular DNA plasmids. However, Archaea and eukaryotes share important components of DNA replication, transcription, and translation. Surprisingly, few DNA replication proteins are homologous across all three domains, although the process of DNA strand elongation is functionally similar (Edgell and Doolittle, 1997). Many DNA replication/repair proteins are homologous between Archaea and eukaryotes but are completely absent in Bacteria. DNA scaffolding proteins in Archaea are like eukaryotic histones (Reeve et al., 1997). Transcriptional components are also strikingly similar between eukaryotes and the

Archaea, particularly the multisubunit structure of the main RNA polymerases (Langer et al., 1995). Although many components of translation such as aminoacyl-tRNA synthetases and elongation factors occur universally, Archaea and eukaryotes share several translation initiation factors that do not have orthologs in the Bacteria (Olsen and Woese, 1997; Kyprides and Woese, 1998).

Collections of molecular sequence data and advances in phylogenetic methodologies inevitably lead to testing the rRNA tree topology with other molecular markers and fuel the more ambitious agenda of deriving a universal tree of life rooted in the cenancestor (Fitch and Upper, 1987). Alternatively called the last common ancestor (LCA) or last universal common ancestor (LUCA), the cenancestor is the extinct cellular organism at the confluence of the lineages of contemporary Archaea, Bacteria, and Eucarya domains. Historical progress and recent frustrations in determining the topology of the universal tree have been described elsewhere (Brown and Doolittle, 1997; Doolittle, 1999). Briefly, phylogenies based on universal paralogous proteins (Gogarten et al., 1989; Iwabe et al., 1989; Brown and Doolittle, 1995; Lawson et al., 1996; Brown et al., 1997) position the root in the Bacteria, such that Archaea and eukaryotes would be considered sister groups—a view bolstered by biochemical similarities between the two groups (Woese et al., 1990).

However, the rooting of the universal tree has been attacked on both technical and philosophical grounds. Phylogenetic analyses using alternative methods and expanded data sets have raised questions about the rooting of the universal tree and the monophyly of the Archaea (Lake, 1988; Rivera and Lake, 1992; Baldauf et al., 1996). Philippe and coworkers (Lopez et al., 1999; Philippe and Forterre, 1999) have maintained that phylogenetic reconstruction of deep evolutionary events are untenable because of mutation saturation effects. Other problems include incidents of unequal mutation rates in different lineages, which could lead to attraction artifacts of long branches (Forterre and Philippe, 1999) and in more extreme cases result in unrecognizable gene paralogy at the primary sequence level while retaining three-dimensional structural relatedness (Gogarten and Oldendzenski, 1999).

Many trees based on single genes, although not uniquely rooted, have unreconcilable topologies with the rRNA universal tree (Smith et al., 1992; Brown and Doolittle, 1997). Problems with phylogenetic methods or hidden gene paralogy cannot solely explain the considerable extent of topological disagreement among prokaryotic trees. Many of these findings invoke scenarios of lateral gene transfer (LGT) among organisms with no contemporary zones of contact or genetic exchange. Comparative analyses of complete genome sequences are now moving LGT, with all of its ramifications to traditional species classification, into the forefront of evolutionary discussion (Hilario and Gogarten, 1993; Martin and Müller, 1998; Doolittle, 1999).

Besides the Archaea-Bacteria dichotomy, rRNA trees make several other predictions about the evolution of prokaryotes. Species tolerant of extreme thermal conditions (called thermophiles) occur as the most basal lineages in both Archaea and Bacteria clades (Woese et al., 1990; Pace, 1991). Among the Bacteria, rRNA trees place the extreme thermophilic genus, *Aquifex*, at the base (Burggraf et al., 1992). Given that the early earth was indeed a very warm place, some have speculated that life itself evolved high temperatures. However, phylogenetic reconstructions using protein genes disagree with the rRNA tree and place *Aquifex* elsewhere, sometimes with the proteobacteria, a largely mesophilic group (i.e., *Escherichia coli* [Klenk et al., 1999; Brown and Lupas, 1998]). The complete genome sequence of *A. aoelicus* was recently determined (Deckert et al., 1998); nonetheless, the evolutionary view of this organism remains confusing, although Bocchetta et al. (2000) found support for both the monophyly and basal position of bacterial thermophiles in trees based on highly conserved ribosomal proteins, elongation factors, and RNA polymerases. However, the lack of common strategies among thermophiles in stabilizing proteins at high temperatures suggests that adaptation to existence at high temperatures might also have evolved independently and recently in different lineages (Brown and Lupas, 1998). Although the paleontological record is scant, shapes of the earliest "prokaryotic" fossils, ~3,465 million years old, are suggestive of contemporary cyanobacteria, which are not thermophilic (Schopf, 1993). However, the

possibility that these early organisms were heat-tolerant or that they had thermophilic predecessors cannot be precluded.

A major factor causing poor resolution of nodes in single gene trees is the paucity of phylogenetically informative sites. To address this issue, Brown et al.(2001) used large combined alignments of 23 orthologous proteins, conserved across 45 species from all Domains, to construct highly robust universal trees. They found that, while individual protein trees were variable in their support of Domain integrity, trees based on combined protein datasets strongly supported separate monophyletic Domains. However, within the Domain Bacteria, Spirochaetes, rather the thermophiles *Aquifex* and *Thermotoga*, were placed as the earliest derived bacterial group. After the elimination from the combined protein alignment of 9 protein datasets, as likely candidates for LGT, thermophiles were placed as the earliest evolved bacterial lineages and all three Domains were monophyletic which is highly congruent with SSU rRNA tree topologies.

The Genomic Era

The hallmark of the Genomic Era was the publication of first complete genome sequences from a bacterium, *Haemophilus influenzae* (Fleischmann et al., 1995), and subsequently from an archaebacterium, *Methanococcus jannaschii* (Bult et al., 1996). Genomes from >40 different organisms have now been completely sequenced and nearly twice that number are currently in progress (see NCBI Genome at http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/bact.html and TIGR Microbial Data-base at http://www.tigr.org/tdb/mdb/mdb.html). However, the abundance of sequence data has resulted in a more, not less, confusing picture of prokaryote evolution. Many genes are highly organism-specific, for which searches of sequence databases reveal no ancestral relationships and few clues about biochemical function. More puzzling is the complete absence of some parts of a particular pathway whereas other components are clearly present. In some instances, one protein may take over the catalytic role of another unrelated protein, so-called nonorthologous gene replacement (Koonin et al., 1996).

Arguably, evolutionary approaches to whole genome data have not matured as fast as the DNA sequencing technology. Automated large-scale homology analyses (i.e., BLAST-based comparisons) can provide checklists of common proteins but subtleties such as gene duplications and fusion and fission events are often missed. Gene inventories can have erroneous annotations, which are amplified throughout public databases unless careful manual curation intervenes.

A prominent example of genome annotation pitfalls occurred in the publication of the human genome by the International Human Genome Sequencing Consortium (IHGSC, 2001). Based on the interpretation of significant BLAST (Altschul et al., 1997) hits, they made an extraordinary claim that as many as 113 vertebrate genes, some only found in humans, were the result of direct horizontal transfers from bacteria. However, two independent studies, using more thorough searches of public databases (i.e. the National Center for Biotechnology Information "EST others" database) and phylogenetic analyses, concluded that there was no compelling support for direct bacteria to vertebrate gene transfers (Salzburg et al., 2001; Stanhope et al., 2001).

Assuming proper annotation and elucidation of gene function, comparative genomic analyses still suggests that LGT was extensive in early cellular evolution (Ochman et al., 2000). Furthermore, LGT can affect all types of genes, including those essential to cell viability or integrated into complex pathways. The aminoacyl-tRNA synthetases are an example. In protein synthesis, aminoacyl-tRNA synthetases are responsible for the attachment of a tRNA to its cognate amino acid. As such, a specific aminoacyl-tRNA synthetase exists for each amino acid. On the basis of structure, biochemical activity, and sequence similarity, aminoacyl-tRNA synthetases can be divided nearly equally into two evolutionary distinct protein families known as class I and II. Despite their critical role in protein synthesis and ancient origins (without them, interpretation of the genetic code would be impossible), aminoacyl-tRNA synthetases have been extensively shuttled between genomes (Brown, 1998; Wolf et al., 1999; Woese et al., 2000). Phylogenetic trees suggest that isoleucyl-tRNA synthetases may have been transferred from an early eukaryote to bacteria as

a specific adaptation to resist a natural antibiotic compound (Brown et al., 1998). Glutaminyl-tRNA synthetase orthologs occur in various proteobacteria, *Deinococcus radiodurans* (a radiation-and desiccation-tolerant bacterium), and eukaryotes but not in other bacteria or the Archaea (Brown and Doolittle, 1999).

Even more suprising are the lysyl-tRNA synthetases, which not only cross species boundaries but also exist in both class I and II families. Until recently, all examples of lysyl-tRNA synthetase were class II-type enzymes. However, novel class I-type lysyl-tRNA synthetases were discovered in the Archaea and subsequently in bacterial Spirochaetes, both of which lack the more typical class II isoforms (Ibba et al., 1997). Whereas class I and II lysyl-tRNA synthetases do not share any sequence or structural similarity, class I lysyl-tRNA synthetases in Spirochaetes and the Archaea are clearly related. The mechanism or timing of the implicated LGT event is unclear. Spirochaetes are human parasites, whereas the Archaea are not known to be pathogenic; thus opportunities in a recent evolutionary time scale for genetic exchanges between these groups are highly unlikely.

Phenylalanyl-tRNA Synthetase: A Case Study in Phylogeny and Evolutionary Genomics

Genome Organization of pheST Operon

Phenylalanyl-tRNA synthetase (PheRS), also a class II-type enzyme, provides another compelling example of the complexity of genome evolution. The phylogeny of this enzyme has been covered in earlier global analyses of tRNA synthetases (Wolf et al., 1999; Woese et al., 2000) and conserved bacterial proteins (Teichmann and Mitchison, 1999) from completely sequenced genomes; here, however, the evolution of PheRS is considered in more detail from the perspectives of genomic organization and gene duplication and with a broader sampling of taxa. PheRS is unique among the tRNA synthetases (besides glycyl-tRNA synthetase), in being a hetrotetramer enzyme comprising two α-subunits, which corresponds to the catalytic domains of other class II synthetases, and two larger β-subunits (Goldgur et al., 1997). In *E. coli*, the α-subunit and β-subunit are 327 and 785 amino acids long, respectively, and in all organisms are separately encoded by the genes *pheS* and *pheT* , respectively. (Note that the naming of these two genes in GenBank has been reversed for some eukaryotic and archaeal species. Here, the bacterial nomenclature, based on the *E. coli* genome annotation, will be used.) The two subunits are evolutionarily unrelated at the level of either primary amino acids sequence or three-dimensional structure. In the genomes of most bacteria and archaea, these two genes occur together in a simple operon, *pheST* (Table 1; Fig. 1). However, many other variations exist, including the separation of the two loci to distant parts of the genome and the inclusion of one or the other *phe* loci in an operon with seemingly metabolically unrelated genes. For example, *Streptococcus pneumoniae* has a small open reading frame (ORF) 169 amino acids long and of unknown function situated between the *pheS* and *pheT* genes. Sequence homology searches revealed that this ORF occurs in other Gram-positive cocci as well as *Bacillus subtilis*, where the most similar homolog is a DNA-binding protein involved in the negative control of sporulation and degradative-enzyme production (GenBank accession number P21340).

Closely related species do not necessarily have the same gene arrangements. For example, consider the spirochaetes *Borrelia burgdorferi* and *Treponema pallidum*. The typical *pheST* operon is present in *B. burgdorferi*, whereas the two genes are highly dispersed in *T. pallidum*, with *pheT* appearing to be cotranscribed with the gene *lon-1*, which codes for an ATPase-dependent protease. Dispersion of *pheS* and *pheT* to different parts of the genome would seem also to imply the co-evolution of mechanisms for coordinating gene expression, because in all species examined both gene products are essential for a functional PheRS. However, nothing is known about the regulation of unlinked *pheS* and *pheT* genes.

All living organisms have single genes coding for PheRS α- and β- subunits, but eukaryotes have an additional PheRS α-subunit gene that apparently codes for an isoform targeted to the mitochondria. No PheRS β-subunit mitochondrial isoforms exist. However, detailed structural analyses suggest that mitochondrial PheRS α-subunits are in part similar to bacterial β-subunits (Sanni

TABLE 1. Genomic organization of *pheS* and *pheT* genes. Groups of species correspond to tree shown in Figure 1, from top to bottom. Genes linked to either *pheS* or *pheT* or the *pheST* operon are connected by a dash. Linkage was defined as 30 nucleotides or less separating the stop to start codons of adjacent open reading frames (ORFs). Semicolons separating *pheS* and *pheT* genes (or gene clusters) indicates unlinked transcriptional units. ORFs of unknown function are indicated as "unk" with length in amino acids, molecular mass, or gene identification number given in parentheses.

Group	Species	Gene organization
Proteobacteria	*Campylobacter jejuni, Escherchia coli, Haemophilus influenzae, Helicobacter pylori 26695, H. pylori J 99, Pseudomonas aeruginosa, Vibrio cholerae, Yersinia pestis*	*pheST*
	Rickettsia prowazekii	*pheST*-dnaN
	Neisseria gonorrhoeae, N. meningitidis	*pheS*-hpaIIR-hpAIIM; *pheT*
Aquificales	*Aquifex aeolicus*	rplt-*pheS*-lepB-unk(39.5kd)-PGPA2; *pheT*
Low−G + C Gram-positives	*Bacillus subtilis, Clostridium acetobutylicum, Enterococcus faecalis, Staphylococcus aureus, Streptococcus pyogenes*	*pheST*
	Streptococcus pneumoniae	*pheS*-unk(169AA)-*pheT*
High−G + C Gram-positives	*Mycobacterium tuberculosis*	*pheST*
Thermotogales	*Thermotoga maritima*	*pheST*
Chlamydiales	*Chlamydia pneumoniae, C. trachomatis*	infc3-rl20-*pheS*-tRNASer; *pheT*-unk (CPN0595)-ada-oppC2-oppB2-oppA5
Cyanobacteria	*Synechocystis* sp. PCC6803	*pheS*; appA-*pheT*
Thermus/Deinococcus	*Deinococcus radiodurans*	*pheS*-unk(DR2355)-unk(DR2356)-*pheT*
Mycoplasma	*Mycoplasma genitalium, M. pneumoniae*	*pheST*
Spirochaetes	*Borrelia burgdorferi*	*pheST*
	Treponema pallidum	*pheS*; *pheT*-(lon-1)
Archaea	*Pyrococcus abyssi, P. furiosus, P. horikoshii*	*pheST*
	Aeropyrum pernix, Methanobacterium thermoautotrophicum, Methanococcus jannaschii	*pheS*; *pheT*
	Archaeoglobus fulgidus	*pheS*; *pheT*-bcpC1-glpF-unk(AF1427)
	Thermoplasma acidophilum	*pheS*; *pheT*-unk(MJ0304)

et al., 1991) and, in effect, a chimera of the two subunits (Diaz-Lazcoz et al., 1998).

In summary, organization of the PheRS genes into operon structures is highly species-specific. Despite the functional importance and universal occurrence of this enzyme, its encoding *pheS* and *pheT* genes have undergone a wide range of rearrangements multiple times over the course of prokaryotic evolution.

Phylogenetic Analyses of PheRS

In subsequent phylogenetic analyses, the following methodology was used. Homologous proteins were initially aligned by using the program CLUSTALW v1.7 (Thompson et al., 1994) with the BLOSUM62 (Henikoff and Henikoff, 1992) similarity matrix and gap-opening and extension penalties of 10.0 and 0.05, respectively. The multiple sequence alignments were further refined manually by using the program SEQLAB of the GCG v11.0 software package (Genetics Computer Group, Madison, WI), to delete most of the gapped regions before phylogenetic analysis. Multiple sequence alignments are available on request from the author.

Phylogenetic trees were constructed by neighbor-joining (N-J) and maximum parsimony (MP) methods for each set of alignments. N-J trees were based on pairwise distances between amino acid sequences by using the programs NEIGHBOR and PROTDIST of the PHYLIP 3.57c package (Felsenstein, 1993). The "Dayhoff" program option was invoked in the latter program, which estimates based on the Dayhoff 120 matrix (Dayhoff et al., 1972). The programs SEQBOOT and CONSENSE were used to estimate the confidence limits of branching points from 1,000 bootstrap replications. MP analysis was done with the software package PAUP* (Swofford, 1999). Given the large size of the dataset, it was not possible to exhaustively search for the total number of

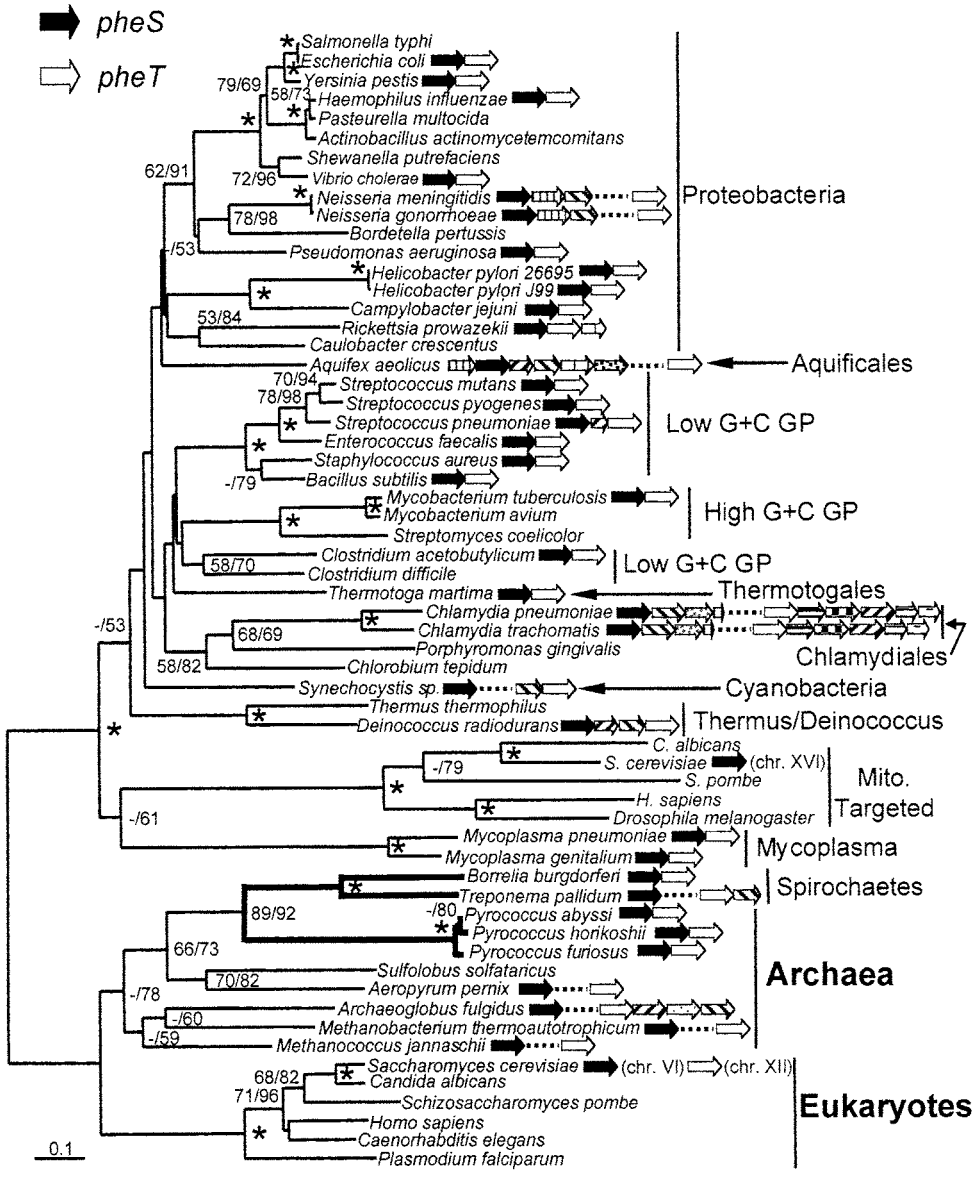

FIGURE 1. Neighbor-joining tree of phenylalanyl-tRNA synthetase α-subunits. The organization of phenylalanyl-tRNA synthetase α- (*pheS*) and β- (*pheT*) subunit genes are shown as solid and blank arrows, respectively, for complete and partially sequenced genomes. Genes located within 30 bp of *pheS* or *pheT* are indicated by arrows with different shadings. Note: Similar shading patterns do not indicate similar genes (other than *pheS* and *pheT*), except among the two *Neisseria* sp. and two *Chlamydia* sp. The identities of the genes in the operons are given in Table 1, in order from top to bottom. For *Saccharomyces cerevisiae*, the chromosome on which the gene occurs is identified. The neighbor-joining method was used as implemented by the program NEIGHBOR of the PHYLIP 3.57c package (Felsenstein, 1993). Thick lines indicate evidence for a proposed lateral gene transfer event from the archaeal genus Pyrococcus to the bacterial group Spirochaetes. The scale bar represents 0.1 expected amino acid replacements per site, as estimated by the program PROTDIST using the Dayhoff PAM substitution matrix. Numbers at the branching points represent the percent occurrence in 1,000 random bootstrap replications of maximum parsimony/neighbor-joining methods, respectively. Nodes where bootstrap in both methods were $\geq$90% are labeled with an asterisk (*). Values <50% are not shown or are indicated by a dash (-). GP, Gram-positive bacteria; Mito. Targeted, mitochondria-targeted isoforms encoded in the eukaryotic nuclear genome.

minimal length trees. Instead, the numbers and lengths of minimal trees were estimated from 100 replicate random heuristic searches, but confidence limits of branch points were estimated by 1,000 bootstrap replications.

Multiple sequence alignments of PheRS α- and β-subunits were trimmed for gaps and unalignable regions down to 212 and 393 amino acid positions, respectively. Phylogenetic trees were constructed for both PheRS α- and β-subunits by MP and N-J methods (Figs. 1 and 2). The tree topologies derived with either method were highly congruent; further, the overall topologies for the two subunit trees were generally similar but with some differences (as discussed below). In 100 random replicate searches, MP analysis detected 12 trees (minimal length 2,774 steps) for the PheRS α-subunit and 3 trees (4,014 steps) for the PheRS β-subunit. The equivalent trees differed mainly in arrangements of terminal taxa but the main branch points were highly consistent.

PheRS α- and β-subunit trees showed two main monophyletic groups (i.e., each originating from a single ancestral branch, assuming a midpoint rooting for the entire tree). The first cluster consists of only bacteria-like proteins, including mitochondria targeted PheRS α-subunits. Subgroups of bacteria, such as Gram-positive bacteria and proteobacteria, are evident in the trees, although

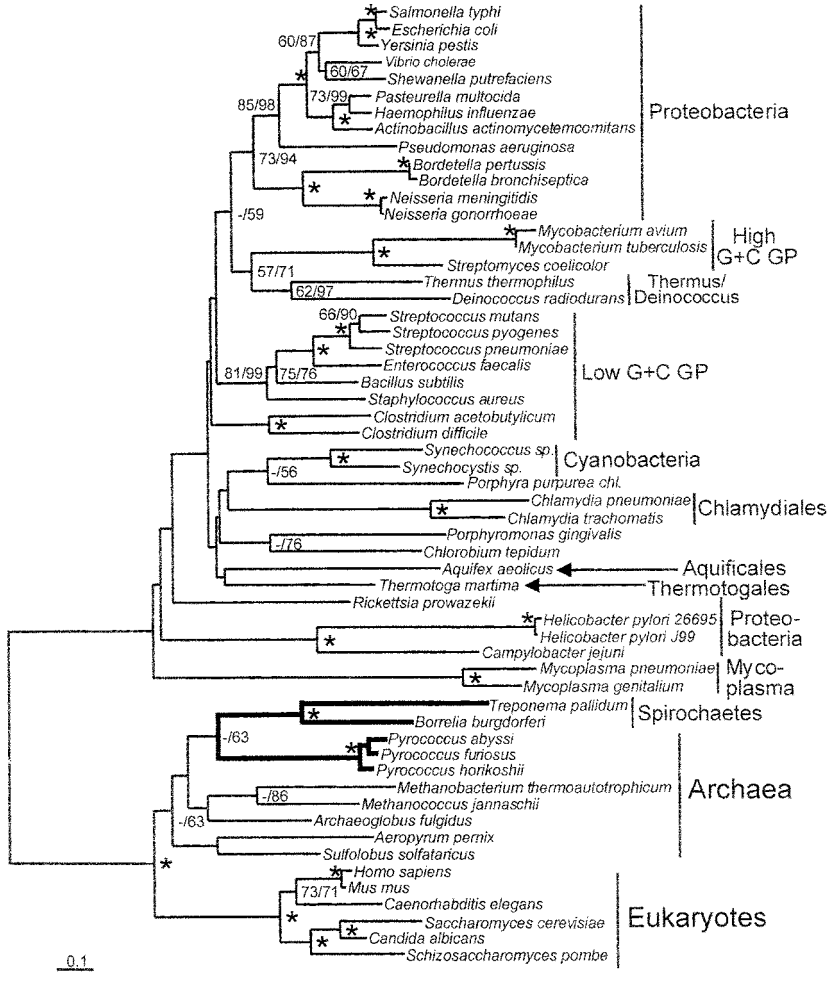

FIGURE 2. Neighbor-joining tree of phenylalanyl-tRNA synthetase β-subunits. Tree construction methods and nomenclature are as given in Figure 1.

the branching orders among subgroups are not well resolved—a fundamental problem in nearly all bacterial phylogenies. Some subgroups are inconsistent with respect to proposed monophyly in small subunit rRNA trees (Olsen et al., 1994). For example, *Campylobacter jejuni* and the two strains of *Helicobacter pylori* (J99 and 26695) do not cluster with other proteobacteria. Mitochondria targeted isoforms of PheRS α-subunits do not cluster with the proposed endosymbiont lineage (the proteobacteria); rather, they fall near the base of the bacterial cluster. This could be an artifact of longer branch lengths.

The second cluster contains eukaryotic, archaeal, and Spirochaete PheRS subunits, distributed as two further subclusters. Eukaryotes form one monophyletic subcluster with strong statistical support. Although relatively few eukaryotic PheRS subunit sequences are available, the differentiation of major groups such as fungi and animals can be discerned. The second subcluster consists of the Archaea and Spirochaetes, which is in agreement with earlier analyses (Teichmann and Mitchison, 1999; Wolf et al., 1999; Woese et al., 2000). Although the statistical support for monophyly of the Archaea/Spirochaete subcluster is not strong, the association of Spirochaetes with *Pyrococcus* sp. is strongly supported in the α-subunit tree and marginally favored in the β-subunit tree. In addition, all archaeal, eukaryotic, and Spirochaete PheRS α-subunits share two insertion synapomorphies: a short, conserved insertion (27 amino acids long, starting at position E162 in *Saccharomyces cerevisiae*) and an N-terminal extension ~100 amino acids long. Collectively, these findings provide additional evidence for past LGT between Spirochaetes and Archaea.

A Conserved Partial PheRS β-Subunit Gene in Gram-Positive Bacteria

The genomic organization and phylogeny of PheRS subunits alone are enough to suggest that the contemporary picture of this simple system probably evolved through complex series of events involving LGT (hence the Spirochaete/Archaea similarities), protein fusions (mitochondrial PheRS α-subunits), and chromosomal recombination (dispersal or reconstitution—or both—of the *pheST* operon). Additional PheRS evolutionary complexity can be found in the Gram-positive bacteria, which have coopted part of the β-subunit for some, yet to be determined, purpose. Increasingly, tRNA synthetase-like domains have been found in proteins that either have a role other than the aminoacylation of tRNAs in protein synthesis or have an unknown function (Schimmel and de Pouplana, 2000). In bacteria, the PheRS β-subunit extends beyond the N- and C-termini of archaeal, eukaryotic, and Spirochaete orthologs (conversely, the α-subunit is larger in these groups relative to the remainder of bacteria). However, a substantial portion of the bacteria-specific N-terminal extension sequence is duplicated as a separate gene locus in the genomes of low–G + C Gram-positive bacteria (Fig. 3). This putative coding region, labeled as the hypothetical gene *YtpR* in *Bacillus subtilis* (Kunst et al., 1997) and called here pFRS (for partial PheRS), occurs in all sequenced Gram-positive bacteria genomes to date except the high–G + C Gram-positive organism, *Mycobacterium tuberculosis*. Gram-positive bacterial PheRS β-subunits are ~800 amino acids long, whereas the predicted pFRS ORFs are considerably shorter, 198 to 231 residues. In the *B. subtilis* chromosome, the translation start sites for the genes *pheT* (2928.5 kb location) and YtpR (3052.4 kb location) are separated by >100 kb of sequence (Kunst et al., 1997). Extensive homology searches did not reveal any further examples of pFRS in other species of Bacteria or the Archaea or eukaryotes.

Multiple sequence alignment of bacterial pFRS and PheRS β-subunits shows that the two proteins share regions of extensive amino acid conservation (Fig. 3). Conceptual translations of pFRS genes suggest that their putative initiator f-Met residue occurs between 44 to 55 amino acids upstream of those of PheRS β-subunits. Comparisons with the three-dimensional structure determined for *Thermus thermophilus* PheRS suggests that putative pFRS ORFs terminate near the middle of the β3 domain (Goldgur et al., 1997). Regions mapping to β-sheets in *T. thermophilus* PheRS β-subunit appear to have the most amino acid conservation between PheRS and pFRS. Although the function of pFRS is unknown, its high degree of amino acid sequence conservation and consistent length suggest it is unlikely to be a pseudogene.

FIGURE 3. Multiple sequence alignment of all partial phenylalanyl-tRNA synthetase β-subunits (pF) from Gram-positive bacteria with partial sequences of true phenylalanyl-tRNA synthetase β-subunits (P) from selected species. Species name abbreviations used in this figure are given in Figure 4. Shown are the locations of α-helices (h) and β-sheets (s) in the structure determined for *Thermus thermophilus* phenylalanyl-tRNA synthetase. Residues conserved in >50% of both pF and P sequences are bold font. Upper and lower sequence numbers correspond to the number of amino acids in pF from *Mycoplasma pneumoniae* (Mp) and P from *T. thermophilus* (Tt), respectively.

In *T. thermophilus*, PheRS has been shown to have DNA-binding specificity (Lechler and Kreutzer, 1998). The PheRS β-subunit is a modular protein with as many as eight different domains (Lechler and Kreutzer, 1997). Interestingly, large portions of the PheRS β-subunit, domains β1 through β5, do not appear to be essential for aminoacylation. Furthermore, putative DNA-binding activity has been localized to two regions with helix-loop-helix motifs in domains β1 and β5. Although pFRS overlaps with PheRS domain β1, the putative helix-loop-helix DNA-binding motif of PheRS does not seem to be conserved (Fig. 3).

PSI-BLAST (Altschul et al., 1997) homology searches using *B. subtilis* pFRS as the query sequence reached convergence (after 3 iterations) on three non-class II tRNA synthetases: the C-terminal region of class I methionyl-tRNA synthetase, the human protein endothelial-monocyte-activating polypeptide II (EMAP II; GenBank accession number B55053), and yeast GU4 nucleic-binding protein, Arc1p (GenBank accession number P46672). Homologies among the latter group of proteins were observed by Simos et al. (1996), who also determined that Arc1p has dual binding specificity for methionyl- and glutaminyl-tRNA synthetases and for their cognate tRNAs. Thus Arc1p might serve as a cofactor in aminoacylation by bringing together tRNAs and aminoacyl-tRNA synthetases. Whether or not pFRS plays a similar role awaits determination through further experiments involving assays for nucleotide-binding activity, protein–protein interactions, and aminoacylation enhancement.

Phylogenetic analyses suggest that the divergence between bacterial PheRS β-subunit and pFRS is ancient (Fig. 4). Although restricted in distribution to the Gram-positive bacteria, pFRS does not appear to have evolved through a specific gene duplication of the PheRS β-subunit in that lineage. Instead, phylogenetic trees suggest that the two protein families diverged very early in bacterial evolution. The timing of a bacterial PheRS β-subunit/pFRS split relative to an archaeal or eukaryotic outgroup cannot be determined because they have no region of sequence overlap with pFRS. Gram-positive bacteria thus either uniquely acquired the pFRS gene early in their evolution or singularly retained, among all living organisms, the vestal genomic artifact of an ancient gene duplication event involving the PheRS β-subunit gene. Alternatively, pFRS could have been one half of a gene duplication in Gram-positive bacteria but soon rapidly diverged away from true PheRS β-subunits.

The Path Ahead—Discerning Evolutionary Mode and Tempo in Prokaryotes

Comparative analyses of genomes suggest that LGT has occurred extensively among and within prokaryotes and eukaryotes (see Koonin et al., 1997; Doolittle, 1999). Some examples are reverse gyrase genes from hyperthermophilic Archaea to Bacteria (Forterre et al., 2000), catalase-peroxidase genes between Archaea and pathogenic proteobacteria (Faguy and Doolittle, 2000), and the entire mevalonate pathway responsible for isoprenoid biosynthesis among eukaryotes, Archaea, and Gram-positive coccal bacteria (Doolittle and Logsdon, 1998; Boucher and Doolittle, 2000; Wilding et al., 2000).

However, the overall patterns or "rules" governing LGT events remain obscure. On the basis of LGT, Lake and colleagues have suggested that genes can be divided into two categories, informational and operational genes (Rivera et al., 1998). Informational genes, which include the central components of DNA replication, transcription, and translation, are less likely to be transferred between genomes than are operational genes involved with cell metabolism. The fact that informational gene products have qualitatively more complex interactions might restrict their opportunities for genetic exchange and fixation (Jain et al., 1999). However, as discussed above, such is clearly not the case for aminoacyl-tRNA synthetases. Alternatively, networks of genetic interactions at the base of the universal tree must have been so intense as to render useless the concept of a single cellular ancestor for contemporary lineages (Hilario and Gogarten, 1993; Woese, 1998).

The extent to which LGT upsets the basic tenants of the universal tree is still hotly debated. Based on the genomic content of orthologous genes among 13 completely sequenced genomes, a phylogenetic tree emerged that was strikingly similar to

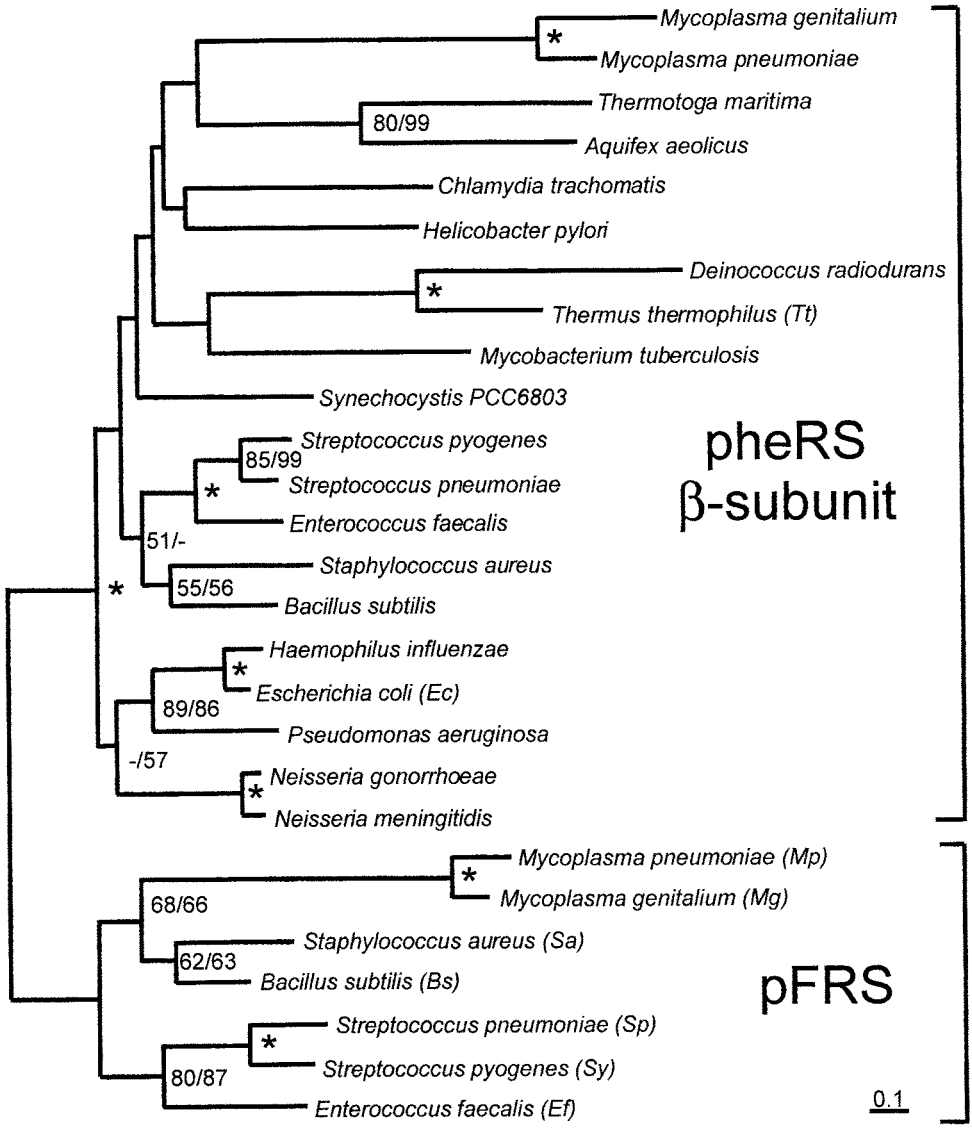

FIGURE 4. Phylogenetic tree of partial phenylalanyl-tRNA synthetase β-subunits (pFRS) from Gram-positive bacteria as well as N-terminal sequences of bacterial phenylalanyl-tRNA synthetase β-subunits (PheRS). Multiple sequence alignments for the phylogenetic analysis of pFRS included only those regions (91 amino acids long) showing strong sequence homology. Species name abbreviations where given are those used in Figure 3. Tree construction methods are as described for Figure 1.

the rRNA tree (Snel et al., 1999). A recent comparative genome analysis points to 351 Archaea-specific "phylogenetic footprints" or combinations of genes uniquely shared by two or more archaeal species but not found in Bacteria or eukaryotes (Graham et al., 2000). However, combinations of catalytic functional orthologs might be actually fewer because both hyperthermophilic Archaea and Bacteria tend to have more split genes than mesophilic species do (Snel et al., 2000).

Such controversies will either be resolved or amplified as genomes from more taxa are sequenced. The present bacterial genome database is biased towards pathogenic bacteria, some of which are highly DNA competent, such as the respiratory tract infection pathogens *Haemophilus influenzae and Streptococcus pneumoniae*. The sampling of archaeal

genomes is biased towards species living in extreme environments, especially high temperatures. Fluidity of genome content might be a positive adaptation towards survival under stressful conditions. Eukaryotic genomes sequenced thus far are from model genetic organisms, which are, relative to their positions in the universal rRNA tree, a closely related group. Even some single-cell protists, such as the malaria agent *Plasmodium falciparum*, whose genome will be soon entirely sequenced, is highly diverged from amitochondriate groups, such as the diplomonad *Giardia lamblia* (Cavalier-Smith, 1993). The genetic diversity of eukaryotes has yet to be fully explored.

Several promising new approaches to the mining of genomic data have recently emerged that could help further elucidate patterns and processes in genome evolution. Protein phylogenetic profiling can be used to assign function to unknown proteins by comparing metabolic pathway enzymes across multiple genomes (Pellegrini et al., 1999). According to this method, missing enzymes or "holes" in common pathways are identified in different genomes. Focused homology and motif-searching techniques identify candidates for the missing genes in the pools of hypothetical proteins and test them empirically.

Although gene order is not always conserved across large evolutionary distances, signals from some loci are sufficient to extract information on function by using "guilt-by-association" logic (Dandekar et al., 1998; Overbeek et al., 1999). Genes organized in operons or close linkage groups might be especially predisposed to LGT because an entire pathway can be more readily fixed in the recipient genome than can individual, unassociated genes (Lawrence and Roth, 1996; Lawrence, 1997). Therefore, genes encoding for unknown proteins might be involved in functions related to genes located nearby on the chromosome. However, as demonstrated in the PheRS example above, the degree or extent of linkages between functionally interacting proteins can vary markedly from species to species. Many proteins have undergone fusion events, in which two interacting subunits, once coded by separate genes, have merged into a single polypeptide, or fission events, in which the reverse has happened (Snel et al., 2000). For example, separate *E. coli* gyrase B and gyrase A subunits are homologous to the N- and C-terminal regions of yeast topoisomerase II, respectively. Finding such matches supports prediction of protein–protein interactions from the genomic sequences (Marcotte et al., 1999).

By applying some or all of these new methodologies to evolutionary directed questions, a deeper understanding of the mode and tempo of prokaryotic genome evolution will emerge. Certainly, the physiological and genetic diversity of prokaryotes, and their tremendous adaptations to extreme environments, will continue to challenge our thinking about the boundaries of life. There is also the issue of whether or not a robust prokaryotic phylogeny can ever be constructed. Trees based on rRNA molecules have been the mainstay of universal and prokaryotic phylogenetic studies for >30 years. However, are rRNA markers the best, or most suitable, benchmark of prokaryotic evolution? Or could new systematic approaches based on either whole genome contents or multiple protein genes (selected for apparent resilience to LGT events) provide a better phylogenetic scaffolding for hanging our hypotheses concerning earlier cellular evolution? In this respect, unraveling the mode and tempo of prokaryotic evolution is one of the most challenging and important arenas of research for systematic biologists.

Acknowledgments

I thank the Society for Systematic Biology and the Glaxo SmithKline Division of Bioinformatics for their support of this symposium. Matt Kane is gratefully acknowledged for his efforts in organizing this symposium as well as his editorial handling of this manuscript.

References

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHÄFFER, J. ZHANG, Z. ZHANG, W. MILLER, AND D. J. LIPMAN. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

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.

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.

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.

BOCCHETTA, M., S. GRIBALDO, A. SANNANGELANTONI, AND P. CAMMARANO. 2000. Phylogenetic depth of the bacterial genera Aquifex and Thermotoga inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. J. Mol. Evol. 50:366–380.

BOUCHER, Y., AND W. F. DOOLITTLE. 2000. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. Mol. Microbiol. 37:703–716.

BROWN, J. R. 1998. Aminoacyl-tRNA synthetases: Evolution of a troubled family. Pages 217–230 *in* Thermophiles—the keys to molecular evolution and the origin of life? (J. Wiegel and M. Adams, eds.). Taylor & Francis Group, London.

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.

BROWN, J. R., AND W. F. DOOLITTLE. 1997. Archaea and the prokaryote to eukaryotes transition. Microbiol. Mol. Biol. Rev. 61:456–502.

BROWN, J. R., AND W. F. DOOLITTLE. 1999. Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. J. Mol. Evol. 49:485–495.

BROWN, J. R., C. J. DOUADY, M. J. ITALIA, W. E. MARSHALL, AND M. J. STANHOPE. 2001. Universal trees based on large combined protein sequence data sets. Nature Genetics 28 (in press).

BROWN, J. R., AND A. N. LUPAS. 1998. What makes a thermophile? Trends Microbiol. 6:349–351.

BROWN, J. R., F. T. ROBB, R. WEISS, AND W. F. DOOLITTLE. 1997. Evidence for the early divergence of tyrosyl- and tyrptophanyl-tRNA synthetases. J. Mol. Evol. 45:9–16.

BROWN, J. R., J. ZHANG, AND J. E. HODGSON. 1998. A bacterial antibiotic resistance gene with eukaryotic origins. Curr. Biol. 8:R365–R367.

BROWN, J. W., C. J. DANIELS, AND J. N. REEVE. 1989. Gene structure, organization and expression in archaebacteria. CRC Crit. Rev. Microbiol. 16:287–338.

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, ET AL. 1996. Complete genome sequence of the methanogenic Archaeon, *Methanococcus jannaschii*. Science 273:1058–1073.

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.

CAVALIER-SMITH, T. 1993. Kingdom protozoa and its 18 phyla. Microbiol. Rev. 57:953–994.

CHATTON, E. 1937. Titres et travauz scientifiques. Setes, Sottano, Italy.

DANDEKAR, T., B. SNEL, M. HUYNEN, AND P. BORK. 1998. Conservation of gene order: A fingerprint of proteins that physically interact. Trends Biochem. Sci. 23:324–328.

DANSON, M. J. 1993. Central metabolism of the Archaea. Pages 1–24 *in* The biochemistry of Archaea (Archaebacteria) (M. Kates, D. J. Kushner, and A. T. Matheson, eds.). Elsevier, Amsterdam.

DAYHOFF, M. O., R. V. ECK, AND C. M. PARK. 1972. A model of evolutionary change in proteins. Pages 89–99 *in* Atlas of protein sequence and structure, volume 5 (M. O.Dayhoff, ed.). National Biomedical Research Foundation, Washington.

DECKERT, G., P. V. WARREN, T. GAASTERLAND, W. G. YOUNG, A. L. LENOX, D. GRAHAM, R. OVERBEEK, M. SNEAD, M. KELLER, M. AUJAY, ET AL. 1998. The *Aquifex aeolicus* genome. Nature 392:353–358.

DELONG, E. F. 1992. Archaea in coastal marine environments. Proc. Natl. Acad. Sci. USA 89:5685–5689.

DELONG, E. F., K. Y. WU, B. B. PRÉZELIN, AND R. V. M. JOVINE. 1994. High abundance of Archaea in Antarctic marine picoplankton. Nature 371:695–697.

DIAZ-LAZCOZ, Y., J.-C. AUDE, P. NITSCHKÉ, H. CHIAPELLO, C. LANDÈS-DEVAUCHELLE, AND J.-L. RISLER. 1998. Evolution of genes, evolution of species: The case of aminoacyl-tRNA synthetases. Mol. Biol. Evol. 15:1548–1561.

DOOLITTLE, W. F. 1998. A paradigm gets shifty. Nature 392:15–16.

DOOLITTLE, W. F. 1999. Phylogentic classification and the universal tree. Science 284:2124–2128.

DOOLITTLE, W. F., AND J. R. BROWN. 1994. Tempo, mode, the progenote and the universal root. Proc. Natl. Acad. Sci. USA 91:6721–6728.

DOOLITTLE, W. F., AND J. M. LOGSDON, JR. 1998. Archaeal genomics: Do archaea have a mixed heritage? Curr. Biol. 8:R209–R211.

EDGELL, D. R., AND W. F. DOOLITTLE. 1997. Archaea and the origin(s) of DNA replication proteins. Cell 89:995–998.

FAGUY, D. M., AND W. F. DOOLITTLE. 2000. Horizontal transfer of catalase-peroxidase genes between Archaea and pathogenic bacteria. Trends Genet. 16: 196–197.

FELSENSTEIN, J. 1993. PHYLIP(Phylogeny Inference Package) version 3.57c. Distributed by the author: http://evolution.genetics.washington.edu/phylip.html, Department of Genetics, Univ. of Washington, Seattle.

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.

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, ET AL. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269: 496–512.

FORTERRE, P., C. BOUTHIER DE LA TOUR, H. PHILIPPE, AND M. DUGUET. 2000. Reverse gyrase from thermophiles: Probable transfer of a thermoadaptation trait from Archaea to Bacteria. Trends Genet. 16:152–154.

FORTERRE, P., AND H. PHILIPPE. 1999. Where is the root of the universal tree of life? BioEssays 21:871–879.

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.

FUHRMAN, J. A., K. MCALLUM, AND A. A. DAVIS. 1992. Novel major archaebacterial group from marine plankton. Nature 356:148–149.

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, ET AL. 1989. Evolution of the vacuolar H^+-ATPase: Implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA 86:6661–6665.

GOGARTEN, J. P., AND L. OLENDZENSKI. 1999. Orthologs, paralogs and genome comparisons. Curr. Opin. Genet. Dev. 9:630–636.

GOLDGUR, Y., L. MOSYAK, L. RESHETNIKOVA, V. ANKILOVA, O. LAVRIK, S. KHODYREVA, AND M. SAFRO. 1997. The crystal structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus* complexed with cognate tRNAPhe. Structure 5:59–68.

GRAHAM, D. E., R. OVERBEEK, G. J. OLSEN, AND C. R. WOESE. 2000. An archaeal genomic signature. Proc. Natl. Acad. Sci. USA 97:3304–3308.

HENIKOFF, S., AND J. G. HENIKOFF. 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 89:10915–10919.

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.

IBBA, M., S. MORGAN, A. W. CURNOW, D. R. PRIDMORE, U. C. VOTHKNECHT, W. GARDNER, W. LIN, C. R. WOESE, AND D. SÖLL. 1997. A euryarchaeal lysyl-tRNA synthetase: Resemblance to class I synthetases. Science 278:1119–1122.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

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.

JAIN, R., M. C. RIVERA, AND J. A. LAKE. 1999. Horizontal gene transfer among genomes: The complexity hypothesis. Proc. Natl. Acad. Sci. USA 96:3801–3806.

KATES, M., D. J. KUSHNER, AND A. T. MATHESON. 1993. The biochemistry of Archaea (Archaebacteria). Elsevier, Amsterdam.

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.

KLENK, H. P., T. D. MEIER, P. DUROVIC, V. SCHWASS, F. LOTTSPEICH, P. P. DENNIS, AND W. ZILLIG. 1999. RNA polymerase of *Aquifex pyrophilus*: Implications for the evolution of the bacterial rpoBC operon and extremely thermophilic bacteria. J. Mol. Evol. 48:528–541.

KOONIN, E. V., A. R. MUSHEGIAN, AND P. BORK. 1996. Non-orthologous gene displacement. Trends Genet. 12:334–336.

KOONIN, E. V., A. R. MUSHEGIAN, M. Y. GALPERIN, AND D. R. WALKER. 1997. Comparison of archaeal and bacterial genomes: Computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Mol. Microbiol. 25:619–637.

KUNST, F., N. OGASAWARA, I. MOSZER, A. M. ALBERTINI, G. ALLONI, V. AZEVEDO, M. G. BERTERO, P. BESSIERES, A. BOLOTIN, S. BORCHERT, ET AL. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature 390:249–256.

KYRPIDES, N. C., AND C. R. WOESE. 1998. Universally conserved translation initiation factors. Proc. Natl. Acad. Sci. USA 95:224–228.

LAKE, J. A. 1988. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. Nature 331:184–186.

LANGER, D., J. HAIN, P. THURIAUX, AND W. ZILLIG. 1995. Transcription in Archaea: Similarity to that in Eucarya. Proc. Natl. Acad. Sci. USA 92:5768–5772.

LAWRENCE, J. G. 1997. Selfish operons and speciation by gene transfer. Trends Microbiol. 5:355–359.

LAWRENCE, J. G., AND J. R. ROTH. 1996. Selfish operons—horizontal transfer may drive the evolution of gene clusters. Genetics 143:1843–1860.

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.

LECHLER, A., AND R. KREUTZER. 1997. Domains of phenylalanyl-tRNA synthetase from *Thermus thermophilus* required for aminoacylation. FEBS Lett. 420:139–142.

LECHLER, A., AND R. KREUTZER. 1998. The phenylalanyl-tRNA synthetase specifically binds DNA. J. Mol. Biol. 278:897–901.

LOPEZ, P., P. FORTERRE, AND H. PHILIPPE. 1999. The root of the tree of life in the light of the covarion model. J. Mol. Evol. 49:496–508.

MARCOTTE, E. M., M. PELLEGRINI, H.-L. NG, D. W. RICE, T. O. YEATES, AND D. EISENBERG. 1999. Detecting protein function and protein–protein interactions from genome sequences. Science 285:751–753.

MARGULIS, L. 1970. Origin of eukaryotic cells. Yale Univ. Press, New Haven, Connecticut.

MARTIN, W., AND M. MÜLLER. 1998. The hydrogen hypothesis for the first eukaryote. Nature 392:37–41.

MOREIRA, D., AND P. LÓPEZ-GRACÍA. 1998. Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. J. Mol. Evol. 47:517–530.

OCHMAN, H., J. G. LAWRENCE, AND E. A. GROISMAN. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304.

OLSEN, G. J., AND C. R. WOESE. 1997. Archaeal genomics—an overview. Cell 89:991–994.

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.

OVERBEEK, R., M. FONSTEIN, M. D'SOUZA, G. D. PUSCH, AND N. MALTSEY. 1999. The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. USA 96:2896–2901.

PACE, N. R. 1991. Origin of life—facing up to the physical setting. Cell 65:531–533.

PELLEGRINI, M., E. M. MARCOTTE, M. J. THOMPSON, D. EISENBERG, AND T. O. YEATES. 1999. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. Proc. Natl. Acad. Sci. USA 96:4285–4288.

PHILIPPE, H., AND P. FORTERRE. 1999. The rooting of the universal tree of life is not reliable. J. Mol. Evol. 49:509–523.

REEVE, J. N., K. SANDMAN, AND C. J. DANIELS. 1997. Archaeal histones, nucleosomes and transcription initiation. Cell 89:999–1002.

RIVERA, M. C., R. JAIN, J. E. MOORE, AND J. A. LAKE. 1998. Genomic evidence for two functionally distinct gene classes. Proc. Natl. Acad. Sci. USA 95:6239–6244.

RIVERA, M. C., AND J. A. LAKE. 1992. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science 257:74–76.

SALZBERG, S. L., O. WHITE, J. Peterson, AND J. A. EISEN. 2001. Microbial genes in the human genome: Lateral transfer or gene loss? Science 292:1903–1906.

SANNI, A., P. WALTER, Y. BOULANGER, J. P. EBEL, AND F. FASIOLO. 1991. Evolution of aminoacyl-tRNA synthetase quaternary structure and activity: *Saccharomyces cerevisiae* mitochondrial phenylalanyl-tRNA synthetase. Proc. Natl. Acad. Sci. USA 88:8387–8391.

SCHIMMEL, P., AND L. R. DE POUPLANA. 2000. Footprints of aminoacyl-tRNA synthetases are everywhere. Trends Biochem. Sci. 25:207–209.

SCHOPF, J. W. 1993. Microfossils of the early Archaen apex chart: New evidence of the antiquity of life. Science 260:640–646.

SIMOS, G., A. SEGREF, F. FASIOLO, K. HELLMUTH, A. SHEVCHENKO, M. MANN, AND E. C. HURT. 1996. The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl- and glutamyl-tRNA synthetases. EMBO J. 15:5437–5448.

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.

SNEL, B., P. BORK, AND M. A. HUYNEN. 1999. Genome phylogeny based on gene content. Nat. Genet. 21: 108–110.

STANHOPE, M. J., A. N. LUPAS, M. J. ITALIA, K. K. KORETKE, C. VOLKER, AND J. R. BROWN. 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. Nature 411:940–944.

STANIER, R.Y. 1970. Some aspects of the biology of cells and their possible evolutionary significance. Symp. Soc. Gen. Microbiol. 20:1–38.

STANIER, R.Y., AND C. B. VAN NIEL. 1941. The main outlines of bacterial classification. J. Bacteriol. 42:437–466.

STANIER, R.Y., AND C. B. VAN NIEL. 1962. The concept of a bacterium. Arch. Mikrobiol. 42:17–35.

STEIN, J. L., AND M. I. SIMON. 1996. Archaeal ubiquity. Proc. Natl. Acad. Sci. USA 93:6228–6230.

SWOFFORD, D. L. 1999. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

TEICHMANN, S. A., AND G. MITCHISON. 1999. Is there a phylogenetic signal in prokaryote proteins? J. Mol. Evol. 49:98–107.

THOMPSON, J. D., D. G. HIGGINS, AND T. J. GIBSON. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

WILDING, E. I., J. R. BROWN, A. BRYANT, A. F. CHALKER, D. HOLMES, K. INGRAHAM, C. Y. SO, M. ROSENBERG, AND M. N. GWYNN. 2000. Identification, evolution and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in Gram-positive cocci. J. Bacteriol. 182:4319–4327.

WOESE, C. R. 1987. Bacterial evolution. Microbiol. Rev. 51:221–271.

WOESE, C. R. 1998. The universal ancestor. Proc. Natl. Acad. Sci. USA 51:221–271.

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.

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

WOESE, C. R., G. J. OLSEN, M. IBBA, AND D. SÖLL. 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. 64:202–236.

WOLF, Y. I., L. ARAVIND, N. V. GRISHIN, AND E. V. KOONIN. 1999. Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Res. 9:689–710.

ZILLIG, W. 1991. Comparative biochemistry of Archaea and Bacteria. Curr. Opin. Genet. Dev. 1:457–463.

Received 16 July 2000; accepted 14 August 2000
Associate Editor: M. Kane