JOURNAL OF BACTERIOLOGY, Dec. 1994, p. 7703–7710
0021-9193/94/$04.00+0

Vol. 176, No. 24

# Phylogenetic Depth of S10 and *spc* Operons: Cloning and Sequencing of a Ribosomal Protein Gene Cluster from the Extremely Thermophilic Bacterium *Thermotoga maritima*

ANNA M. SANANGELANTONI,[1] MAURIZIO BOCCHETTA,[2] PIERO CAMMARANO,[2] AND ORSOLA TIBONI[1]*

*Dipartimento di Genetica e Microbiologia "A. Buzzati Traverso," Università di Pavia, 27100 Pavia,[1] and Dipartimento Biopatologia Umana, Sezione Biologia Cellulare, Istituto Pasteur–Fondazione Cenci-Bolognetti, Policlinico Umberto I, Università di Roma I, 00161 Rome,[2] Italy*

Received 16 March 1994/Accepted 17 October 1994

**A segment of *Thermotoga maritima* DNA spanning 6,613 bp downstream from the gene *tuf* for elongation factor Tu was sequenced by use of a chromosome walking strategy. The sequenced region comprised a string of 14 tightly linked open reading frames (ORFs) starting 50 bp downstream from *tuf*. The first 11 ORFs were identified as homologs of ribosomal protein genes *rps*10, *rpl*3, *rpl*4, *rpl*23, *rpl*2, *rps*19, *rpl*22, *rps*3, *rpl*16, *rpl*29, and *rps*17 (which in *Escherichia coli* constitute the S10 operon, in that order); the last three ORFs were homologous to genes *rpl*14, *rpl*24, and *rpl*5 (which in *E. coli* constitute the three promoter-proximal genes of the spectinomycin operon). The 14-gene string was preceded by putative −35 and −10 promoter sequences situated 5′ to gene *rps*10, within the 50-bp spacing between genes *tuf* and *rps*10; the same region exhibited a potential transcription termination signal for the upstream gene cluster (having *tuf* as the last gene) but displayed also the potential for formation of a hairpin loop hindering the terminator; this suggests that transcription of *rps*10 and downstream genes may start farther upstream. The similar organization of the sequenced *rp* genes in the deepest-branching bacterial phyla (*T. maritima*) and among *Archaea* has been interpreted as indicating that the S10-*spc* gene arrangement existed in the (last) common ancestor. The phylogenetic depth of the *Thermotoga* lineage was probed by use of r proteins as marker molecules: in all except one case (S3), *Proteobacteria* or the gram-positive bacteria, and not the genus *Thermotoga*, were the deepest-branching lineage; in only two cases, however, was the inferred branching order substantiated by bootstrap analysis.**

The chromosomal organization of the (eu)bacterial genes encoding protein components of the translational apparatus has been investigated in members of *Proteobacteria* (*Escherichia coli* [33]), low-G+C gram-positive bacteria (*Bacillus subtilis* [18] and *Mycoplasma capricolum* [39]), organelles (12, 13), and *Cyanophora paradoxa* cyanelles (37); less extensive data have been reported for the cyanobacteria (*Spirulina platensis* [43]), the deinococci (*Thermus aquaticus* [22]), and members of *Thermotogales* (*Thermotoga maritima* [31, 53]). Typically, in *E. coli* the genes for translational elongation factors EF-Tu (*tuf*) and EF-G (*fus*) and the genes encoding several ribosomal proteins (*rp*) constitute three contiguous transcription units termed the *str* (streptomycin), S10, and *spc* (spectinomycin) operons, in that order. The gene arrangement of the three clusters is basically invariant over the phylogenetic spectrum, from *E. coli* to the deinococci (representing the fourth deepest branching within the bacterial tree [40]). In fact only minor deviations from the *E. coli* theme have been observed in evolutionarily diverse members of *Bacteria*.

In order to investigate further the phylogenetic depth of the ribosomal protein gene organization among procaryotes, we undertook the examination of organisms closer than the deinococci to the root of the bacterial tree.

In this article we report the sequence of a 6.5-kbp DNA segment downstream from the *tuf* gene of *T. maritima*, an extreme thermophile representing the second deepest divergence in the bacterial domain on the basis of 16S rRNA-derived phylogenies (40, 56). Analysis of the sequenced region shows that the basic arrangement of the genes constituting the (*E. coli*) *str*, S10, and *spc* operons can be traced back to the *T. maritima* divergence, close to the (present) root of the bacterial tree. The placement of *T. maritima* in the bacterial domain has been investigated by use of the sequenced ribosomal proteins as phylogenetic marker molecules, and the robustness of the inferred topologies has been analyzed by bootstrapping.

## MATERIALS AND METHODS

**Bacterial strains and plasmids.** *T. maritima* MSB8 (DSM 3109) cells were kindly supplied by R. Huber (University of Regensburg, Regensburg, Federal Republic of Germany). Plasmid pTM6 (54), harboring the *tuf* gene of *T. maritima*, was used for preliminary sequencing. Plasmids pBR322 and pUC19 were used as the vectors, and *E. coli* HB101 was used as the host.

**Recombinant DNA techniques.** Chromosomal DNA from *T. maritima* was prepared as described elsewhere (54). Plasmid DNA was isolated and purified by use of Qiagen tips (Diagen) according to the manufacturer's instructions. Recovery of DNA fragments from low-melting-temperature agarose gels, ligation, transformation experiments, and inverse PCR were carried out as described by Sambrook et al. (42). Southern hybridization experiments were performed at 42°C in the presence of 50% formamide. Probes were radioactively labelled by the multiprime DNA labelling system with [α-$^{32}$P] dCTP (specific activity, 3,000 Ci/mmol; Amersham) and the

* Corresponding author. Mailing address: Dipartimento di Genetica e Microbiologia "A. Buzzati Traverso," Università di Pavia, via Abbiategrasso 207, 27100 Pavia, Italy.

Multiprime DNA Labelling System kit supplied by Amersham. Restriction endonuclease enzymes and T4 DNA ligase were obtained from Boehringer and used according to the manufacturer's instructions. Sequencing was done by the dideoxy chain termination method (44) with overlapping templates. DNA was radioactively labelled with $^{35}$S-dATP (specific activity > 1,000 Ci/mmol; Amersham) and primed as specified by the manufacturer with a $^{T7}$Sequencing kit (Pharmacia).

**Sequence analysis.** The DNA sequences were analyzed with the DNA Strider 1.0 program. Homologies of the predicted proteins to proteins in the databases were searched with the PCGENE software package (Intellegenetics). Amino acid sequences were aligned with the MULTALIN (10) and CLUSTAL V (21) software packages with a gap penalty of 8.0. The alignments obtained by the two methods were corrected for obvious mismatches of signature sequences and economy of gapping. Minor discrepancies between the alignments obtained by the two programs were resolved subjectively.

**Phylogenetic methods.** Bootstrapped distance matrix and maximum-parsimony trees were constructed with the programs FITCH, PROTDIST, PROTPARS, SEQBOOT, and CONSENSE in the PHYLIP software package version 3.5c (17). Unless otherwise specified, amino acid sequence distance matrixes were generated by the Kimura distance option of Felsenstein's PROTDIST program, by use of selected blocks of the multiply aligned sequences. Corrected evolutionary distances ($D$) were also calculated from the fractional amino acid identities ($S$) by the equation $D = -\ln S$ (50) with the personal computer program SCORE (16).

**Nucleotide sequence accession number.** The complete sequence of the DNA region (6,613 bp) encompassing the 14 open reading frames (ORFs) and the upstream sequences discussed below has been deposited in the EMBL Data Library under accession number Z21677.

## RESULTS

**Cloning and sequencing.** In a previous work, we cloned a *T. maritima* 3.5-kbp *Bam*HI DNA fragment comprising the 3′ end of the *fus* gene, the entire *tuf* gene, and a region extending approximately 1,400 bp downstream from *tuf* (54).

Sequencing of the 1,400-bp DNA segment revealed a new ORF of 102 codons starting 50 bp downstream from the translation stop codon of *tuf*. The protein predicted from this ORF was 56% identical to the *E. coli* ribosomal protein S10. The S10-encoding gene (*rps*10) was immediately followed by two ORFs of 207 and 76 codons, respectively, whose predicted amino acid sequences displayed significant similarity to *E. coli* ribosomal protein L3 (49% identity) and to the N-terminal region of *E. coli* ribosomal protein L4.

The *Thermotoga* DNA region extending 3′ to the L4-encoding gene (*rpl*4) was then cloned by a chromosome walking strategy. In essence, five recombinant plasmids (pTM7 to pTM11) containing overlapping DNA fragments were obtained by use of the 3′ end of each insert as a probe. Sequence analysis of the inserts revealed the missing 3′ portion of gene *rpl*4 and a string of eight ORFs lying on the same DNA strand. The last ORF within the pTM11 insert (identified as a homolog of the *E. coli rps*17 gene) lacked the last few base pairs, and all attempts to clone the DNA region containing the translation stop codon by chromosome walking proved unsuccessful. This difficulty was circumvented by resorting to the inverse PCR that made possible the cloning of the 3′ end of gene *rps*17 and of three additional ORFs (plasmid pTM12).

A database search for homology of the predicted proteins to the primary structures of all ribosomal proteins available to date revealed significant similarities to the proteins listed in Table 1; from the similarities (31 to 70% amino acid identity with bacterial proteins), the 11 ORFs located downstream from *rps*10, *rpl*3, and *rpl*4 could be confidently identified as the *Thermotoga* homologs of genes *rpl*23, *rpl*2, *rps*19, *rpl*22, *rps*3, *rpl*16, *rpl*29, and *rps*17 (representing the eight promoter-distal genes of the *E. coli* S10 operon) and of genes *rpl*14, *rpl*24, and *rpl*5 (corresponding to the three promoter-proximal genes of the *E. coli spc* operon). In all except two cases (S10 and L29), the predicted *Thermotoga* proteins correspond to primary RNA-binding proteins which, in *E. coli*, associate directly and independently with the cognate domains of the related rRNA species.

The *Thermotoga* protein S10 had significant similarity to the rat protein designated S20 (29% amino acid sequence identity); proteins S17 and L22 were homologous to the rat proteins designated S11 and L17, respectively; L23 was homologous to the L25 proteins from *Trypanosoma brucei*, *Hansenula jadinii*, and *Saccharomyces cerevisiae* and to the rat protein designated L23a. In certain cases (proteins S3, L3, L4, and L22), the similarities between the predicted *Thermotoga* proteins and their putative archaeal and eucaryal homologs were slightly higher than that (20% identity) expected for two random sequences having the same amino acid composition (41); in one instance (*Saccharomyces pombe* L2; 179 residues), the eucaryal protein was remarkably shorter than the bacterial homologs (271 to 281 residues). In all these cases, however, true homology was confirmed by analysis of shared signature sequences (results not shown).

**Gene order and linkage.** The organization of the *rp* genes in the 6.5-kbp region of the *T. maritima* chromosome was essentially superimposable on that of the *E. coli* S10 and *spc* operons (Fig. 1; see also Fig. 3). Similarly to those of *E. coli*, most genes were separated by very short (only 4 to 20 bp) intergenic regions apparently devoid of transcription signals, and no intergenic spacing existed between genes *rpl*16 and *rpl*29 and between genes *rps*17 and *rpl*14 (highlighted in Fig. 1, boxes C and D); both the *rpl*16-*rpl*29 and the *rps*17-*rpl*14 junctions exhibited overlapping of the TGA stop codon of the upstream gene with the ATG start codon of the downstream one (. . . ATGA . . . .). A more impressive overlapping (29 bp, highlighted in Fig. 1, box B) existed between the 3′ end of gene *rpl*4 and the 5′ region of gene *rpl*23; a similar, though less extensive, overlapping at the *rpl*4-*rpl*23 junction occurs in *E. coli* (58) and *M. capricolum* (39).

Analysis of the noncoding region (highlighted in Fig. 1, box A) between genes *tuf* and *rps*10 (which in *E. coli* represent the promoter-distal gene of the *str* operon and the promoter-proximal gene of the S10 operon, respectively) revealed a putative rho-independent transcription termination signal consisting of a dyad-symmetrical structure followed by a T cluster; interestingly, this region also possessed the potential for formation of an alternative hairpin hindering the terminator. Furthermore (as Fig. 1, box A, shows), two motifs resembling authentic *T. maritima* −10 and −35 promoter elements (31) could be identified immediately upstream of the initiation codon of gene *rps*10.

It should be noticed that in *E. coli* the region (about 160 bp) between genes *rps*17 and *rpl*14 harbors promoter and transcription termination sequences, so that the two genes constitute the promoter-distal element of the S10 operon and the promoter-proximal element of the *spc* operon, respectively (33). These two genes overlapped in the *T. maritima* chromosome. As Fig. 1 shows, however, potential −35 and −10 promoter motifs (overlined in box D) starting 304 bp upstream from the

TABLE 1. Homologies between the predicted *T. maritima* ribosomal proteins and proteins from other sources

| Ribosomal protein, size[a] | Organism[b] | % Identity[c] | No. of aa | Reference |
|---|---|---|---|---|
| S3, 209 aa | **Eco** | 51 | 232 | 58 |
| | **Bst** | 50 | 217 | 28 |
| | **Mca** | 39 | 233 | 39 |
| | **Cpa** | 43 | 219 | 37 |
| | **Ech** | 41 | 218 | 12 |
| | Hha | 26 | 302 | 47 |
| | Hma | 26 | 304 | 2 |
| | Mva | 26 | 224 | 47 |
| | Xla | 20 | 246 | 14 |
| | Rat | 21 | 243 | 8 |
| | Hsa | 21 | 243 | 57 |
| S10, 102 aa | **Eco** | 56 | 103 | 58 |
| | **Mca** | 56 | 102 | 39 |
| | **Spl** | 65 | 96 | 43 |
| | **Cpa** | 60 | 104 | 38 |
| | Mva | 35 | 91 | 29 |
| | Tac | 40 | 104 | 52 |
| | Pwo | 40 | 102 | 11 |
| | Sac | 41 | 102 | 4 |
| | Rat (S20) | 29 | 120 | 9 |
| S17, 107 aa | **Eco** | 59 | 83 | 58 |
| | **Bst** | 57 | 86 | 19 |
| | **Bsu** | 52 | 86 | 18 |
| | **Mca** | 59 | 85 | 39 |
| | **Ctr** | 57 | 83 | 23 |
| | **Taq** | 51 | 105 | 22 |
| | **Cpa** | 44 | 88 | 37 |
| | Hma | 40 | 111 | 1 |
| | Mva | 35 | 109 | 3 |
| | Rat (S11) | 34 | 157 | 51 |
| S19, 95 aa | **Eco** | 55 | 91 | 58 |
| | **Bst** | 65 | 91 | 28 |
| | **Mca** | 59 | 88 | 39 |
| | **Cpa** | 57 | 92 | 15 |
| | **Ech** | 50 | 94 | 12 |
| | Hma | 30 | 140 | 2 |
| | Hha | 32 | 115 | 35 |
| L2, 276 aa | **Eco** | 61 | 272 | 58 |
| | **Bst** | 65 | 275 | 26 |
| | **Mca** | 53 | 281 | 39 |
| | **Cpa** | 58 | 275 | 15 |
| | **Ech** | 57 | 271 | 12 |
| | Hma | 41 | 239 | 2 |
| | Mva | 35 | 237 | 27 |
| | Spo | 31 | 179 | 55 |
| L3, 207 aa | **Eco** | 49 | 209 | 58 |
| | **Bst** | 54 | 215 | 20 |
| | **Mca** | 48 | 223 | 39 |
| | **Cpa** | 43 | 209 | 15 |
| | Hma | 32 | 337 | 2 |
| | Sce | 27 | 386 | 46 |
| | Rat | 24 | 402 | X62166[d] |
| L4, 235 aa | **Eco** | 41 | 209 | 58 |
| | **Bst** | 55 | 213 | 20 |
| | **Mca** | 43 | 223 | 39 |
| | Hma | 21 | 245 | 2 |
| L5, 184 aa | **Eco** | 55 | 178 | 7 |
| | **Bst** | 59 | 179 | 24 |
| | **Bsu** | 59 | 179 | 18 |
| | **Ctr** | 55 | 155 | 23 |
| | **Mca** | 56 | 180 | 39 |
| | **Taq** | 58 | 182 | 22 |
| | **Cpa** | 56 | 181 | 37 |
| | Hma | 33 | 176 | 45 |
| | Mva | 34 | 181 | 3 |
| L14, 122 aa | **Eco** | 62 | 123 | 7 |
| | **Bst** | 62 | 122 | 25 |
| | **Bsu** | 64 | 122 | 18 |
| | **Ctr** | 62 | 122 | 23 |
| | **Mca** | 62 | 122 | 39 |
| | **Taq** | 70 | 122 | 22 |
| | **Cpa** | 60 | 122 | 37 |
| | **Ech** | 60 | 121 | 13 |
| | Hma | 42 | 132 | 1 |
| | Mva | 38 | 132 | 3 |
| L16, 142 aa | **Eco** | 57 | 136 | 58 |
| | **Bsu** | 68 | 106 | 18 |
| | **Mca** | 62 | 137 | 39 |
| | **Cpa** | 61 | 136 | 37 |
| L22, 159 aa | **Eco** | 50 | 110 | 58 |
| | **Bst** | 53 | 113 | 28 |
| | **Mca** | 62 | 111 | 39 |
| | **Cpa** | 42 | 113 | 37 |
| | **Ech** | 43 | 113 | 12 |
| | Hma | 16 | 154 | 2 |
| | Hha | 18 | 155 | 35 |
| | Rat (L17) | 24 | 184 | 48 |
| | Hsa | 23 | 184 | 34 |
| L23, 100 aa | **Eco** | 44 | 100 | 58 |
| | **Bst** | 55 | 95 | 25 |
| | **Mca** | 41 | 94 | 39 |
| | **Ech** | 28 | 100 | 12 |
| | Hma | 39 | 84 | 2 |
| | Mva | 38 | 86 | 27 |
| | Tbr (L25) | 34 | 156 | 36 |
| | Sce (L25) | 34 | 137 | 30 |
| | Hja (L25) | 30 | 142 | 36 |
| | Rat | 31 | 157 | 49 |
| L24, 105 aa | **Eco** | 55 | 103 | 7 |
| | **Bst** | 60 | 103 | 25 |
| | **Mca** | 38 | 108 | 39 |
| | **Ctr** | 33 | 111 | 23 |
| | **Taq** | 54 | 110 | 22 |
| L29, 66 aa | **Eco** | 44 | 63 | 58 |
| | **Bst** | 50 | 66 | 25 |
| | **Bsu** | 46 | 66 | 18 |
| | **Mca** | 39 | 138 | 39 |
| | **Ctr** | 31 | 72 | 23 |
| | Hma | 32 | 70 | 2 |
| | Hha | 31 | 71 | 47 |

[a] aa, amino acids.

[b] Bacteria: Eco, *E. coli*; Bst, *Bacillus stearothermophilus*; Bsu, *B. subtilis*; Mca, *M. capricolum*; Ctr, *Chlamydia trachomatis*; Taq, *T. aquaticus*; Spl, *S. platensis*; Cpa, *C. paradoxa* cyanelles; Ech, *Euglena gracilis* chloroplasts. Archaea: Hha, *Halobacterium halobium*; Hma, *H. (Haloarcula) marismortui*; Mva, *M. vannielii*; Tac, *Thermoplasma acidophilum*; Pwo, *Pyrococcus woesei*; Sac, *Sulfolobus acidocaldarius*. Eucarya: Xla, *Xenopus laevis*; Hsa, *Homo sapiens*; Spo, *S. pombe*; Sce, *S. cerevisiae*; Tbr, *T. brucei*; Hja, *H. jadinii*.

[c] Sequence identities were obtained with the program SCORE on the basis of multiple alignments obtained with the aid of the CLUSTAL V program.

[d] Sequence accession number in EMBL Data Library, deposited by Y. Kuwano and I. G. Wool.

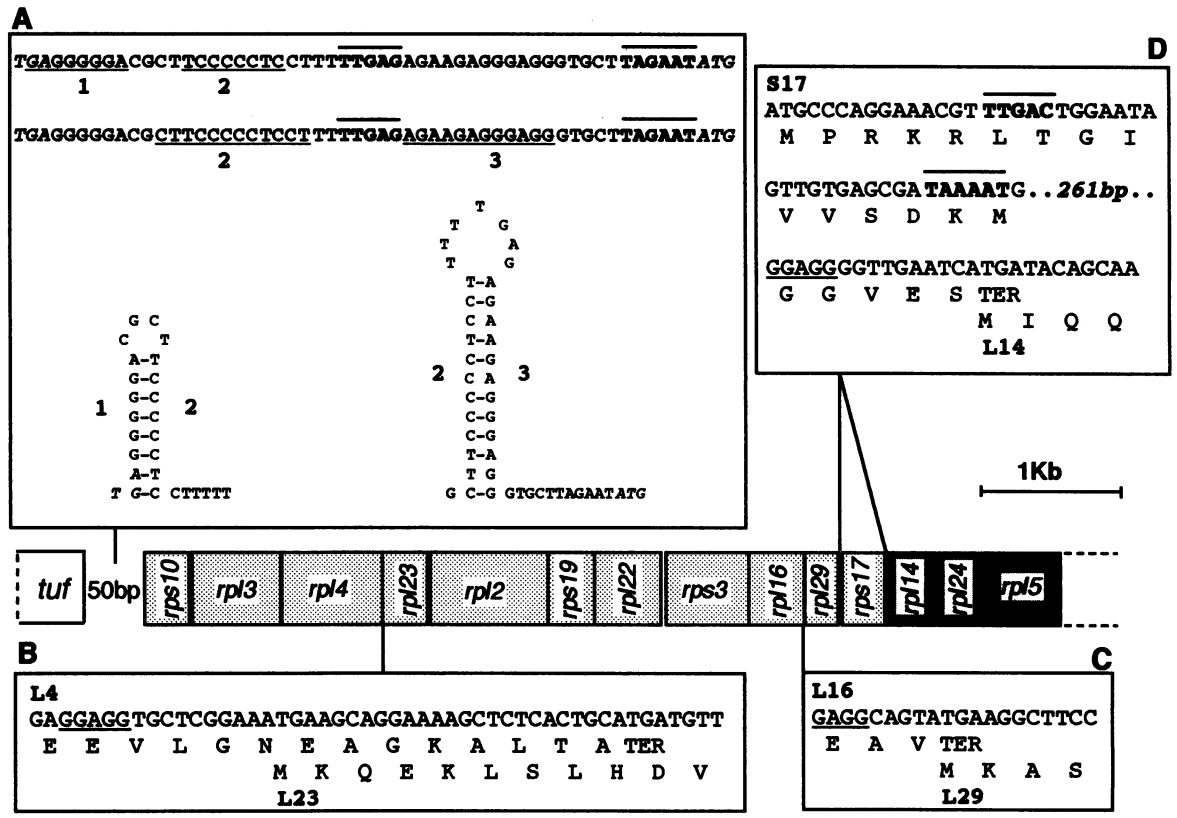


FIG. 1. Gene organization in the sequenced 6,613-bp fragment of the *T. maritima* genome. (A) Predicted secondary structure motifs in the 50-bp region between genes *tuf* and *rps*10. The inverted repeats are numbered and underlined. (B) Nucleotide sequence overlap at the *rpl*4-*rpl*23 junction. (C) Nucleotide sequence overlap at the *rpl*16-*rpl*29 junction. (D) Nucleotide sequence overlap at the *rps*17-*rpl*14 junction. In boxes B, C, and D, underlined sequences indicate ribosome binding sites; overlapping amino acid sequences are shown below the nucleotide sequences. Overlined sequences indicate potential promoter motifs; TER indicates a termination codon.

initiation codon of *rpl*14, well within the coding sequence of gene *rps*17, could be identified (see Discussion).

Lastly, most *rp* genes were preceded by the putative Shine-Dalgarno sequence $GGAGG_{OH}$, generally starting at position −13 from the initiation codon; as Fig. 1 (boxes B to D) shows, the $GGAGG_{OH}$ sequences for genes *rpl*23, *rpl*29, and *rpl*14 were situated within the 3′ coding regions of the respective upstream genes.

**Phylogenetic analysis of ribosomal protein sequences.** To probe the phylogenetic depth of the *Thermotogales* further (53), the *T. maritima rp* sequences were multiply aligned with the available homologs from archaea and/or eucarya and from other bacteria. Positions that were considered reliably alignable between all bacterial sequences were extracted from the global alignments, and evolutionary sequence distances were calculated from these selected data sets using the Kimura correction (see Materials and Methods). Complete and restricted alignments are available upon request from Piero Cammarano.

Outgroup-rooted bacterial phylogenies based on the evolutionary distances are exemplified in Fig. 2, showing Fitch-Margoliash trees inferred from three of the longer proteins. The complete topologies of all *rp* trees for which outgroups were available are described in nested notation in Table 2, together with the bootstrap confirmation values attached to the deepest branch in the bacterial cluster and to the branch separating the bacterial cluster from the outgroup species.

As Table 2 shows, *Bacteria* generally formed a monophyletic and holophyletic grouping rigorously segregated (99 to 100% bootstrap confirmation) from the outgroups, and *Archaea* constituted a coherent cluster distinct from the *Eucarya* (S3, L22, and S10); in only one case (protein L23) was the segregation of the bacterial sequences from the archaeal ones tenuous (48% bootstrap confirmation), and the two archaeal sequences (*Methanococcus vannielii* and *Halobacterium marismortui*) did not share a common origin.

However (Table 2), different ribosomal proteins delivered individual phylogenetic behaviors and most trees were virtually devoid of resolution confirmable by bootstrap analysis: in all except one case (protein S3), the *Proteobacteria* (*E. coli*) and the gram-positive bacteria, not the *Thermotogales*, were the deepest bacterial offshoots and in only 2 (L2 and S17) of 12 trees were these branching orders substantiated by bootstrapping.

Importantly, essentially identical branch patterns and similarly low bootstrap confirmation levels (<90%) were obtained when the same data sets giving the tree topologies described in Table 2 were analyzed by the Felsenstein's protein parsimony method (17), which is based on the minimum number of nucleotide substitutions needed to evolve the protein sequences on a phylogeny with neglect of synonymous changes (results not shown).

## DISCUSSION

**Organization of ribosomal protein genes.** The chromosomal organization (order and linkage) of the ribosomal protein

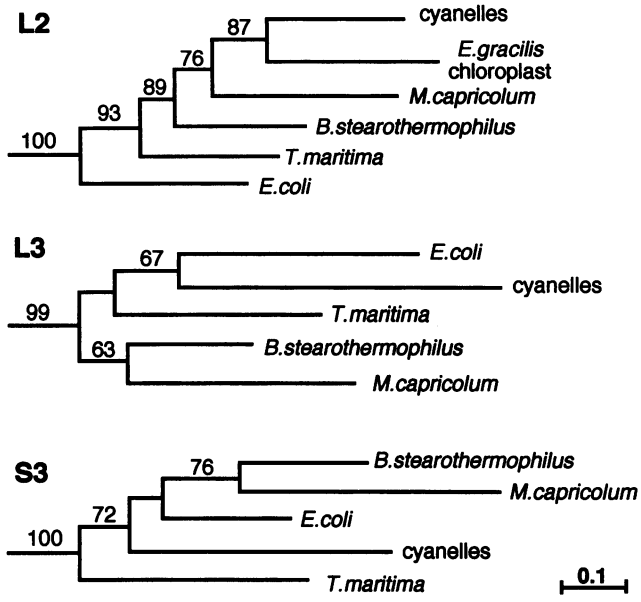


FIG. 2. Distance matrix trees (program FITCH) inferred from the indicated ribosomal proteins on the basis of Kimura distances (program PROTDIST). For concision, the branch patterns of the outgroup sequences are not shown; complete topologies are given in Table 2. Numbers near internal branches refer to the bootstrap replications (out of 100 resampling) confirming the grouping of the species to the right of the branch. Bootstrap values of ≤50 are not shown. The scale bar is in units of nucleotide substitutions per sequence position. The evolutionary distance between any two species is given by summing the lengths of the horizontal connecting branches.

genes constituting the *E. coli str*, S10, and *spc* operons (Fig. 3) is basically conserved among *Bacteria* (18, 22, 33, 37). Differences in detail, however, do exist. First, depending on the organisms, *rps*10 is the promoter-distal unit of the *str* operon (as in cyanelles [38] and *S. platensis* [43]) or the promoter-proximal gene of the S10 operon (most *Bacteria*). Second, in certain organisms (*M. capricolum* and *T. aquaticus*) the *rp* genes corresponding to the *E. coli* S10 and *spc* operons are not resolved into two clusters. Last, the spacing between genes *tuf* and *rps*10 is subject to considerable variation (up to 15,000 bp in *E. coli*).

The sequence data in the present report show that the arrangements of the *E. coli* S10 and *spc* operons can be traced back to the second deepest divergence of the bacterial tree, represented by *T. maritima* (Fig. 3). Unlike the situation in *E. coli*, however, the *tuf* and *rps*10 genes are separated by a stretch of only 50 bp comprising (i) a dyad-symmetrical structure followed by a T cluster typical of rho-independent terminators and (ii) a promoter-like motif resembling other *T. maritima* −10 and −35 consensus elements (31) located immediately upstream of the initiation codon of the *rps*10 gene; intriguingly, the potential for formation of an alternative hairpin loop suggests that, under physiological constraints, transcription may well originate farther upstream. Also, unlike *E. coli*, the 3′ end of gene *rps*17 overlaps with the 5′ end of gene *rpl*14, suggesting that, in *T. maritima*, all of the sequenced *rp* genes form a single cluster. However, sequence elements resembling an authentic *T. maritima* promoter (31) are found within the coding region of *rps*17; therefore, the possibility that the string comprising genes *rpl*14, *rpl*24, and *rpl*5 is transcribed independently from the upstream genes exists.

As the organizations of the S10 and *spc* operons is also conserved in *Archaea* (1, 3), it seems likely that the arrangement (though not necessarily the coordinate transcription) of the ribosomal protein genes existed in the last common ancestor represented as the root of the global phylogenetic tree (56).

**Evolution of ribosomal protein genes.** Phylogenies inferred from 16S rRNAs (40, 56) and translational elongation factors EF-Tu(1α) and EF-G(2) (see reference 6) all concur in the notion that the (hyperthermophilic) *Thermotogales* are a particularly deep offshoot in the bacterial tree, deeper than the green nonsulfur bacteria, although less so than the *Aquificales*, which represent the second (known) hyperthermophilic lineage within *Bacteria* (5). As the earliest archaeal lineages are also dominated by hyperthermophiles, it has been argued that the most recent common ancestor of extant life forms was hyperthermophilic, possibly reflecting the environmental conditions of the early earth (5).

TABLE 2. Topologies of Kimura distance matrix trees from ribosomal protein sequences

| Protein family | No. of selected sites/total no. of sites[a] | Tree topology[b] | Bootstrap confirmation[c] | |
|---|---|---|---|---|
| | | | a | b |
| L2 | 267/288 | ((**Eco**,(Tma,(Bst,(Mca,(Cpa,Ech)))))),(*Spo*,(*Mva*,*Hma*))) | 93 | 100 |
| L4 | 200/256 | ((**Bst**,(Mca,(Tma,Eco))),*Hma*) | <50 | 100 |
| L3 | 191/408 | ((((**Bst**,**Mca**),(Tma,(Eco,Cpa))),*Hma*),(*Sce*,*rat*)) | <50 | 99 |
| L5 | 175/205 | ((**Eco**,((Mca,(Bst,Bsu)),((Taq,Cpa),(Tma,Ctr)))),(*Mva*,*Hma*)) | 81 | 100 |
| S3 | 163/345 | (((**Tma**,(Cpa,(Eco,(Bst,Mca)))),(*Xla*,(*rat*,*Hsa*))),(*Mva*,(*Hha*,*Hma*))) | 72 | 100 |
| L14 | 121/137 | ((**Mca**,((Bst,Bsu),(Ctr,(Tma,(Eco,(Taq,(Cpa,Ech))))))),(*Mva*,*Hma*)) | 60 | 100 |
| L22 | 104/186 | ((**Bst**,(Tma,((Mca,Cpa),(Eco,Ech)))),((*Hsa*,*rat*),(*Hha*,*Hma*))) | <50 | 100 |
| L23 | 96/179 | ((**Eco**,((Bst,Tma),(Mca,Ech))),(*Hma*,(*Mva*,(*Tbr*,(*rat*,(*Sce*,*Hja*)))))) | <50 | 48 |
| S17 | 94/175 | ((**Eco**,((Cpa,Ctr),(Tma,(Taq,(Mca,(Bst,Bsu)))))),(*Mva*,*Hha*)) | 92 | 99 |
| S10 | 91/122 | ((**Mca**,(Eco,(Tma,(Spl,Cpa)))),((((*Pwo*,*Mva*),*Sac*),*Tac*),*rat*)) | <50 | 100 |
| S19 | 85/140 | ((**Eco**,(Cpa,(Ech,(Mca,(Bst,Tma))))),(*Hha*,*Hma*)) | 79 | 100 |
| L29 | 61/141 | (((**Bst**,**Bsu**),(Tma,(Mca,(Eco,Ctr)))),(*Hha*,*Hma*)) | 82 | 100 |

[a] Number of selected positions over by the total number of positions in the alignments.

[b] Species abbreviations are as in Table 1. The earliest-branching lineage(s) within the bacterial grouping are given in boldface; outgroup (archaeal and/or eucaryal) species used for rooting the bacterial trees are shown in italics.

[c] a, bootstrap confirmation values attached to the deepest node in the bacterial tree, giving the species shown in boldface as the earliest branching lineage. b, bootstrap values attached to the branch separating *Bacteria* from outgroup species. Nested notations can be translated into the corresponding trees with the PHYLIP 3.5c CONSENSE program.

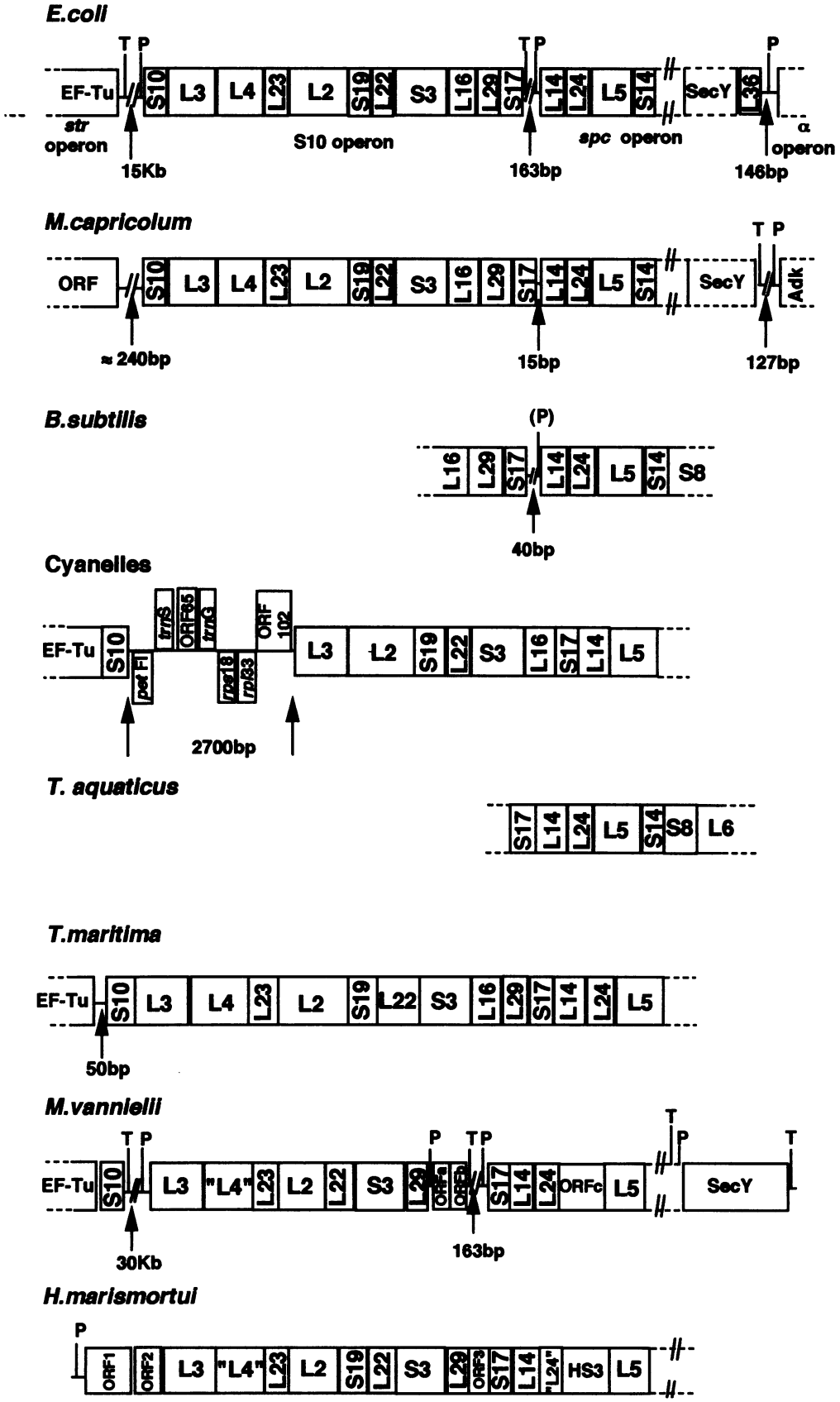


FIG. 3. Organization of *str* (represented only by the EF-Tu-encoding gene), S10, and *spc* gene clusters in different microorganisms. P and T indicate promoter and terminator sequences, respectively. The arrows indicate intergenic spaces.

In contrast to the phylogenies based on rRNA and elongation factors, the present analysis, in which 12 ribosomal proteins (mostly representing primary RNA-binding proteins) were used, is inconclusive with respect to the placement of *T. maritima*. In all except one case (protein S3), both sequence distance and parsimony methods (not shown) predict that mesophiles (*Proteobacteria* or the gram-positive bacteria), not the genus *T. maritima*, are the deepest-branching lineages in the bacterial tree. In only two instances, however, was the inferred branch pattern rigorously defined by bootstrapping; in these cases (L2 and S17), *E. coli* was given as the deepest-branching lineage in more than 90% of bootstrap resamplings (Table 2).

A similar result (32) has been reported recently for the ribosomal protein families L1, L10, L11, and L12 (129 to 267 residues). Here too, both parsimony and sequence distance methods indicated mesophiles (gram-positive bacteria, cyanobacteria, or α-purple bacteria) as being phylogenetically more deeply separated than (or as deeply separated as) the *Thermotoga* lineage; also, similar to those discussed in the present report, none of the inferred topologies was robust by the bootstrap criterion.

This situation indicates that a substantial fraction of the alignment positions do not support the topology inferred from the original sample. This cannot be accounted for simply by the relatively small size of the ribosomal proteins or/and errors in the alignments. First, most *rp* sequences could be aligned with relatively few uncertainties, and the ambiguously aligned positions were not used in construction of phylogenetic trees. Second, the deepest branch in the bacterial tree was given with 92% bootstrap confidence in the S17-based tree (94 sites) while being insignificantly supported in trees based on the L3 and L4 proteins (191 and 200 sites, respectively). A likely possibility is that compensatory (adaptive) mutations occur at several sequence sites because of concerted evolution with interacting (rRNA and protein) components; one such case is illustrated by the coevolution of protein L23 and the interacting 23S rRNA domain(s) (36). These (adaptive) mutations would add to the neutrally induced ones and deliver altered phylogenetic information. On the whole, our results and those of Liao and Dennis (31) indicate that ribosomal proteins are poor chronometers with which to infer organismal relations. However, conflicts between the robust S17- and L2-derived trees and the trees derived from 16S rRNA and EF sequences remain unexplained.

## ACKNOWLEDGMENTS

This work was supported by grants from MURST and from a CNR grant, Progetto Finalizzato Biotecnologie e Biostrumentazione, Sottoprogetto Bioingegneria Molecolare e Cellulare.

## REFERENCES

1. **Arndt, E.** 1990. Nucleotide sequence of four genes encoding ribosomal proteins from the "S10 and spectinomycin" operon equivalent region in the archaebacterium *Halobacterium marismortui*. FEBS Lett. **267:**193–198.
2. **Arndt, E., W. Krömer, and T. Hatakeyama.** 1990. Organization and nucleotide sequence of a gene cluster coding for eight ribosomal proteins in the archaebacterium *Halobacterium marismortui*. J. Biol. Chem. **265:**3034–3039.
3. **Auer, J., G. Spicker, and A. Böck.** 1989. Organization and structure of the *Methanococcus* transcriptional unit homologous to the *Escherichia coli* "spectinomycin operon." J. Mol. Biol. **209:**21–36.
4. **Auer, J., G. Spicker, L. Mayerhofer, G. Pülher, 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.
5. **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.
6. **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.
7. **Cerretti, D. P., D. Dean, G. R. Davis, D. M. Bedwell, and M. Nomura.** 1983. The *spc* ribosomal protein operon of *Escherichia coli*: sequence and cotranscription of the ribosomal protein genes and a protein export gene. Nucleic Acids Res. **11:**2599–2616.
8. **Chan, Y.-L., K. R. Devi, J. Olvera, and I. G. Wool.** 1990. The primary structure of rat ribosomal protein S3. Arch. Biochem. Biophys. **283:**546–550.
9. **Chan, Y.-L., and I. G. Wool.** 1990. The primary structure of rat ribosomal protein S20. Biochim. Biophys. Acta **1049:**93–95.
10. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16:**10881–10890.
11. **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 ultrathermophilic archaeote *Pyrococcus woesei*: phylogenetic implications. J. Mol. Evol. **33:**332–342.
12. **Cristopher, D. A., J. C. Cusham, C. A. Price, and R. B. Hallick.** 1988. Organization of ribosomal protein genes *rpl*23, *rpl*2, *rps*19, *rpl*22 and *rps*3 on the *Euglena gracilis* chloroplast genome. Curr. Genet. **14:**275–286.
13. **Cristopher, D. A., and R. B. Hallick.** 1989. *Euglena gracilis* chloroplast ribosomal protein operon: a new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. Nucleic Acids Res. **17:**7591–7608.
14. **Di Cristina, M., R. Menard, and P. Pierandrei-Amaldi.** 1991. *Xenopus laevis* ribosomal protein S1a cDNA sequence. Nucleic Acids Res. **19:**1943.
15. **Evrard, J. L., M. Kuntz, and J. H. Weil.** 1990. The nucleotide sequence of five ribosomal protein genes from the cyanelles of *Cyanophora paradoxa*: implications concerning the phylogenetic relationship between cyanelles and chloroplasts. J. Mol. Evol. **30:** 16–25.
16. **Felsani, A.** Unpublished data.
17. **Felsenstein, J.** 1990. Phylogenetic inference program (PHYLIP) manual version 3.3. University of Washington, Seattle.
18. **Henkin, T. M., S. H. Moon, L. C. Mattheakis, and M. Nomura.** 1989. Cloning and analysis of the *spc* ribosomal protein operon of *Bacillus subtilis*: comparison with the *spc* operon of *Escherichia coli*. Nucleic Acids Res. **17:**7469–7486.
19. **Herfurth, E., H. Hirano, and B. Wittmann-Liebold.** 1991. The amino acid sequences of the *Bacillus stearothermophilus* ribosomal proteins S17 and S21 and their comparison to homologous proteins of other ribosomes. Biol. Chem. Hoppe-Seyler **372:**955–961.
20. **Herwig, S., V. Kruft, and B. Wittmann-Liebold.** 1992. Primary structures of ribosomal proteins L3 and L4 from *Bacillus stearothermophilus*. Eur. J. Biochem. **207:**877–885.
21. **Higgins, D. G., E. J. Bleasby, and R. Fuchs.** 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8:**189–191.
22. **Jahn, O., R. K. Hartmann, and V. A. Erdmann.** 1991. Analysis of the *spc* ribosomal protein operon of *Thermus aquaticus*. Eur. J. Biochem. **197:**733–740.
23. **Kaul, R., G. J. Gray, N. R. Koehncke, and L. Gu.** 1992. Cloning and sequence analysis of the *Chlamydia trachomatis spc* ribosomal protein gene cluster. J. Bacteriol. **174:**1205–1212.
24. **Kimura, J., and M. Kimura.** 1987. The complete amino acid sequence of the 5S rRNA binding proteins L5 and L18 from the moderate thermophile *Bacillus stearothermophilus* ribosome. FEBS Lett. **210:**85–90.
25. **Kimura, M., J. Kimura, and K. Ashman.** 1985. The complete primary structure of ribosomal proteins L1, L14, L15, L23, L24 and L29 from *Bacillus stearothermophilus*. Eur. J. Biochem. **150:** 491–497.
26. **Kimura, M., J. Kimura, and K. Watanabe.** 1985. The primary

structure of ribosomal protein L2 from *Bacillus stearothermophilus*. Eur. J. Biochem. **153**:289–297.
27. **Köpke, A. K. E., and B. Wittmann-Liebold.** 1989. Comparative studies of ribosomal proteins and their genes from *Methanococcus vannielii* and other organisms. Can. J. Microbiol. **35**:11–20.
28. **Kromer, W. J., T. Hatakeyama, and M. Kimura.** 1990. Nucleotide sequence of *Bacillus stearothermophilus* ribosomal protein genes: part of the ribosomal S10 operon. Biol. Chem. Hoppe-Seyler **371**: 631–636.
29. **Lechner, K., G. Heller, and A. Böck.** 1989. Organization and nucleotide sequence of a transcriptional unit of *Methanococcus vannielii* comprising genes for protein synthesis elongation factors and ribosomal proteins. J. Mol. Evol. **29**:20–27.
30. **Leer, R. J., M. M. C. van Raamsdonk-Duin, M. J. M. Hagendoorn, W. H. Mager, and R. J. Planta.** 1984. Structural comparison of yeast ribosomal protein genes. Nucleic Acids Res. **12**:6685–6700.
31. **Liao, D., and P. P. Dennis.** 1992. The organization and expression of essential transcription translation component genes in the extremely thermophilic eubacterium *Thermotoga maritima*. J. Biol. Chem. **267**:22787–22797.
32. **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.
33. **Lindahl, L., F. Sor, R. H. Archer, M. Nomura, and J. M. Zengel.** 1990. Transcriptional organization of the S10, *spc* and α operons of *Escherichia coli*. Biochim. Biophys. Acta **1050**:337–342.
34. **Mager, D. L., and J. D. Freeman.** 1990. A human gene related to the ribosomal protein L23 gene of *Halobacterium marismortui*. Nucleic Acids Res. **18**:5301.
35. **Mankin, A. S.** 1989. The nucleotide sequence of the genes coding for the S19 and L22 equivalent ribosomal proteins from *Halobacterium halobium*. FEBS Lett. **246**:13–16.
36. **Metzenberg, S., C. Joblet, P. Verspieren, and N. Agabian.** 1993. Ribosomal protein L25 from *Trypanosoma brucei*: phylogeny and molecular co-evolution of an rRNA-binding protein and its rRNA binding site. Nucleic Acids Res. **21**:4936–4940.
37. **Michalowski, C. B., B. Pfanzagl, W. Löffelhardt, and H. J. Bohnert.** 1990. The cyanelle S10 *spc* ribosomal protein gene operon from *Cyanophora paradoxa*. Mol. Gen. Genet. **224**:222–231.
38. **Neumann-Spallart, C., J. Jakowitsch, M. Kraus, M. Brandtner, H. J. Bohnert, and W. Löffelhardt.** 1991. *rps*10, 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.
39. **Ohkubo, S., A. Muto, Y. Kawauchi, F. Yamao, and S. Osawa.** 1987. The ribosomal protein gene cluster of *Mycoplasma capricolum*. Mol. Gen. Genet. **210**:314–322.
40. **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.
41. **Palm, P.** Unpublished data.
42. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **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.
44. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**: 5463–5467.
45. **Scholzen, T., and E. Arndt.** 1991. Organization and nucleotide sequence of ten ribosomal protein genes from the region equivalent to the spectinomycin operon in the archaebacterium *Halobacterium marismortui*. Mol. Gen. Genet. **228**:70–80.
46. **Schultz, L. D., and J. D. Friesen.** 1983. Nucleotide sequence of the *tcm1* gene (ribosomal protein L3) of *Saccharomyces cerevisiae*. J. Bacteriol. **155**:8–14.
47. **Spiridonova, V. A., A. S. Akhmanova, V. K. Kagramanova, A. K. E. Köpke, and A. S. Mankin.** 1989. Ribosomal protein gene cluster of *Halobacterium halobium*: nucleotide sequence of the genes coding for S3 and L29 equivalent ribosomal proteins. Can. J. Microbiol. **35**:153–159.
48. **Suzuki, K., and I. G. Wool.** 1991. The primary structure of rat ribosomal protein L17. Biochem. Biophys. Res. Commun. **178**: 322–328.
49. **Suzuki, K., and I. G. Wool.** 1993. The primary structure of rat ribosomal protein L23a. J. Biol. Chem. **268**:2755–2761.
50. **Swofford, D. L., and G. J. Olsen.** 1990. Phylogeny reconstruction, p. 411–436. *In* D. M. Hillis and C. Moritz (ed.), Molecular systematics. Sinauer Associates Inc., Sunderland, Mass.
51. **Tanaka, T., Y. Kuwano, K. Ishikawa, and K. Ogata.** 1985. Nucleotide sequence of cloned cDNA specific for rat ribosomal protein S11. J. Biol. Chem. **260**:6329–6333.
52. **Tesch, A., and F. Klink.** 1992. Nucleotide sequence of the gene coding for ribosomal protein S10 from the archaeum *Thermoplasma acidophilum*. Nucleic Acids Res. **20**:4090.
53. **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.
54. **Tiboni, O., A. M. Sanangelantoni, P. Cammarano, L. Cimino, G. Di Pasquale, and S. Sora.** 1989. Expression in *Escherichia coli* of the *tuf* gene from the extremely thermophilic eubacterium *Thermotoga maritima*: purification of the *Thermotoga* elongation factor Tu by thermal denaturation of the mesophile host cell proteins. Syst. Appl. Microbiol. **12**:127–133.
55. **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.
56. **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.
57. **Zhang, X. T., Y.-M. Tan, and Y. H. Tan.** 1990. Isolation of a cDNA encoding human 40S ribosomal protein s3. Nucleic Acids Res. **18**: 6689.
58. **Zurawski, G., and S. M. Zurawski.** 1985. Structure of the *Escherichia coli* S10 ribosomal protein operon. Nucleic Acids Res. **13**: 4521–4526.