

## A Chimeric Disposition of the Elongation Factor Genes in *Rickettsia prowazekii*

ANN-CHRISTINE SYVÄNEN,<sup>†</sup> HALEH AMIRI, ASFAR JAMAL,  
SIV G. E. ANDERSSON, AND CHARLES G. KURLAND\*

Department of Molecular Biology, Biomedical Center,  
Uppsala University, S751 24 Uppsala, Sweden

Received 13 May 1996/Accepted 13 August 1996

**An exceptional disposition of the elongation factor genes is observed in *Rickettsia prowazekii*, in which there is only one *tuf* gene, which is distant from the lone *fus* gene. In contrast, the closely related bacterium *Agrobacterium tumefaciens* has the normal bacterial arrangement of two *tuf* genes, of which one is tightly linked to the *fus* gene. Analysis of the flanking sequences of the single *tuf* gene in *R. prowazekii* shows that it is preceded by two of the four tRNA genes located in the 5' region of the *Escherichia coli* *tufB* gene and that it is followed by *rpsJ* as well as associated ribosomal protein genes, which in *E. coli* are located downstream of the *tufA* gene. The *fus* gene is located within the *str* operon and is followed by one tRNA gene as well as by the genes *secE* and *nusG*, which are located in the 3' region of *tufB* in *E. coli*. This atypical disposition of genes suggests that intrachromosomal recombination between duplicated *tuf* genes has contributed to the evolution of the unique genomic architecture of *R. prowazekii*.**

Bacteria that are obligate parasites of eucaryotic cells, such as *Mycoplasma*, *Coxiella*, *Chlamydia*, and *Rickettsia* spp., are often relatively small, with genomes of the size of 1 Mb or less. The phylogenetic placement of these bacteria on the basis of their rRNA genes indicate that they are independent descendants of diverse bacteria with much larger genomes (40). In order to examine the hypothesis that reductive evolution of the genomes of obligate parasitic bacteria proceeds through intrachromosomal recombination at duplicated genes, we have compared the dispositions of the genes coding for the elongation factors in the alpha proteobacteria *Rickettsia prowazekii* and *Agrobacterium tumefaciens*.

These closely related bacteria differ drastically in their lifestyles and genomic architectures. *R. prowazekii* is an obligate intracellular parasite causing typhus in humans, and it has a single, circular genome of 1.1 Mb (16). In contrast, *A. tumefaciens* is a free-living soil bacterium that is also responsible for the development of crown gall and hairy root diseases of dicotyledonous plants (13, 63). It has a large, complex genome containing four replicons: two chromosomes of 3.0 and 2.1 Mb, a 450-kb cryptic plasmid, and the 200-kb Ti plasmid (1).

The gene for elongation factor Tu (EF-Tu) is present in two copies in the genome of *Escherichia coli* (18, 23) as well as in many other proteobacterial genomes (54). Furthermore, the two *tuf* genes are normally associated with characteristic flanking sequences. For example, in *E. coli*, *tufA* is a member of the *str* (streptomycin) operon, in which the genes are arranged in the order *rpsL*, *rpsG*, *fusA*, and *tufA*, whereas *tufB* is part of the *tufB* operon, in which four tRNA genes, *thrU*, *tyrU*, *glyT*, and *thrT*, are positioned at the 5' side of the *tufB* gene (3, 46, 62). Similar arrangements of the flanking sequences of *tuf* genes are common in eubacteria and in archaeobacteria (6, 25, 42). Such

widely preserved patterns of genomic disposition presumably reflect ancient ancestral patterns.

Despite the fact that phylogenetic reconstructions based on the genes for the EFs support a close relationship between *R. prowazekii* and *A. tumefaciens* (3a), we find that these genes are arranged in very different ways in the two bacteria. Thus, free-living *A. tumefaciens* apparently has a normal complement of duplicate *tuf* genes. In contrast, *R. prowazekii* has lost one *tuf* gene and the remaining *tuf* gene is not linked to *fusA*. An analysis of the flanking sequences of the single *tuf* gene in *R. prowazekii* shows that the sequences are shuffled versions of the left flanking and right flanking sequences normally associated with *tufB* and *tufA*, respectively. The data suggest that the evolution of the genome of *R. prowazekii* has involved an intrachromosomal recombination event in the duplicated *tuf* genes of a proteobacterial ancestor.

### MATERIALS AND METHODS

**Genomic DNA.** *A. tumefaciens* genomic DNA was purified from an overnight culture grown at 30°C in Luria broth (47). *R. prowazekii* genomic DNA as well as a genomic library constructed in the lambda Zap II cloning system (Stratagene, La Jolla, Calif.) was prepared as described previously (5).

**Primers.** Degenerate oligomers (Table 1) used for the amplification of the *tuf* and *fus* genes from *A. tumefaciens* and *R. prowazekii* were designed on the basis of conserved amino acid sequences of EF-Tu (30) and EF-G (9), taking into account the high G+C content of the *A. tumefaciens* genome (61) and the high A+T content of the *R. prowazekii* genome (52, 59). The primers for capture PCR were biotinylated at their 5' ends during synthesis with a biotinyl phosphoramidite reagent (Cruachem Ltd., Glasgow, Scotland).

**PCR amplification.** Single *A. tumefaciens* bacterial colonies and aliquots of the *R. prowazekii* lambda phage stock were suspended in 50 µl of 1% Triton X-100 and boiled for 15 min in preparation for PCR amplification. Five microliters of a 1:10 dilution of the *A. tumefaciens* lysate or an amount corresponding to about 10<sup>5</sup> PFU of the *R. prowazekii* lambda phage were used in each amplification reaction. In some cases, 0.1 ng of genomic DNA or 0.2 pg of plasmid DNA was used per PCR. The amplifications were carried out under standard conditions with 1 µM primers or 2 µM degenerate primers. The annealing temperature was usually 4 to 6°C below the melting temperature of the primer (calculated according to the 2 to 4°C rule [47]).

Reamplifications with nested primers were carried out with 10 µl of a 1:300 dilution of the primary PCR product. In some cases 10 µl of a 1:1,000 dilution of primary PCR products eluted from agarose electrophoresis gels was reamplified. The PCR products were purified for sequencing or labelling with Wizard PCR Prep resin (Promega Biotech) directly from the PCR mixture or after separation on 2% agarose gels according to the instructions of the supplier of the resin.

\* Corresponding author. Mailing address: Department of Molecular Biology, Uppsala University, Biomedical Center, Box 590, S751 24 Uppsala, Sweden. Phone: 18-174200. Fax: 18-557723. Electronic mail address: chuck@xray.bmc.uu.se.

<sup>†</sup> Present address: Department of Human Molecular Genetics, National Public Health Institute, 00300 Helsinki, Finland.

TABLE 1. Degenerate primers used for PCR amplification of internal fragments of the *A. tumefaciens* and *R. prowazekii* genes encoding EF-Tu and EF-G

Primer <sup>a</sup>	Sequence <sup>b</sup>	Position <sup>c</sup>
AT-G-2	GT (CT) GGC (AG) CCCAT (AG) CG (AG) TCCAT	444-424
AT-G-3	CGAAACCGT (CT) TGGCG (CGT) CA (AG) GC	363-383
AT-G-4	GTC (AG) (CG) ACAGGGTGTC (AG) CC (CG) GT	1188-1168
AT-G-5	CTCGGC (CG) CTGGC (CG) TT (CT) AA (AG) AT	945-962
AT-G-6	GTTTC (AG) CG (AG) TAGGC (AG) AC (CT) TG	1469-1450
AT-G-7	GCCGG (CT) ATGGG (CT) GA (AG) CT (CG) CA	1366-1385
AT-G-8	GTTT (AG) ACCTTCATGAT (CT) GG (CT) TC	1840-1819
AT-G-17	AT (CT) (AT) (CG) (CG) GC (CG) CA (CT) AT (CT) GA (TC) GC	43-62
RP-Tu-1	GGTCA (CT) GT (AT) GATCA (CT) GGTAA (AG) AC	55-77
RP-Tu-2	CC (AT) ACAGTTC (GT) ACCACC (CT) TC (CT) C	1154-1133
RP-Tu-3	ACAGGT (AC) GAGT (AT) GAA (AC) G (AG) GG	685-704
RP-Tu-4	CCTCTTTC (AT) ACTC (GT) (AG) CC (AT) GT	704-685
RP-G-1	GATACTCCTGG (AT) CA (CT) GT (AT) GA (CT) TT	241-263
RP-G-2	TCTAC (CT) TTCAT (AT) ATAGG (CT) TC (AT) AG	1850-1828
RP-G-3	GGTATGGG (AT) GAACT (AT) CA (CT) CT (AT) GA	1354-1376
RP-G-4	TCAAG (AG) TG (AT) AGTTC (AT) CCCAT (AT) CC	1376-1354

<sup>a</sup> AT and RP denote *A. tumefaciens* and *R. prowazekii*. Primers of the sense orientation have odd numbers, and primers of the antisense orientation have even numbers.

<sup>b</sup> The degenerate nucleotides are given in parentheses.

<sup>c</sup> Nucleotide numbering is according to the respective sequences of the *tuf* and *fus* genes.

**Capture PCR.** The capture PCR method was applied to isolate the 5' and 3' ends and the noncoding flanking sequences of the *A. tumefaciens fus* and *tuf* genes as previously described (17) with the following modifications: *A. tumefaciens* genomic DNA (10 to 100 ng per digest) was cleaved with the restriction enzymes *AluI*, *RsaI*, *EcoRV*, *HaeIII*, *HincII*, and *DraI* and a blunt-ended oligonucleotide linker was ligated to the restriction fragments. The initial biotinylated primer extension products obtained from the digests were affinity captured on 0.25 mg of avidin-coated polystyrene microparticles (IDEXX Co., Portland, Maine) as previously described (57). A nested PCR primer complementary to a region of known sequence (Fig. 1) and a primer with the linker sequence (27) were used for amplification of one-fourth of the captured extension product.

**Plaque and Southern blot hybridization.** To serve as hybridization probes, 200 ng of *R. prowazekii tuf* or *fus* PCR products was labelled with [<sup>32</sup>P]dCTP to a specific activity of  $3 \times 10^7$  cpm/ $\mu$ g by extension of the PCR primers with Klenow DNA polymerase under standard conditions (47). Plaque hybridization filters were prepared from the *R. prowazekii* library, and hybridizations were carried out with  $5 \times 10^6$  cpm of the *fus*- or *tuf*-specific probes per  $\mu$ l (47). After secondary

screening, positive clones were excised in vivo according to the protocol of the supplier of the lambda ZAP II cloning system. Transformation and plasmid preparations were done according to standard methods (47). Southern blot hybridizations of *R. prowazekii* genomic and plasmid DNA were carried out with *R. prowazekii tuf* and *fus* probes as for the plaque hybridizations.

**Sequence analysis.** The double-stranded *A. tumefaciens* PCR products were sequenced by a chain termination protocol modified essentially as previously described (11). Plasmid DNA was isolated with Qiagen large-scale plasmid preps (KEBO, Stockholm, Sweden), and the DNA sequences of the inserts were determined by double-stranded dideoxy sequencing with modified T7 DNA polymerase (Sequenase) and fluorescent dATP (Pharmacia, Uppsala, Sweden). The products of the sequencing reactions were analyzed on an A.L.F. sequencer (Pharmacia).

The databanks were searched for homologous sequences with the BLAST program (2).

**Nucleotide sequence accession numbers.** The nucleotide sequences reported in this paper will appear in the EMBL, GenBank, and DDBJ nucleotide se-

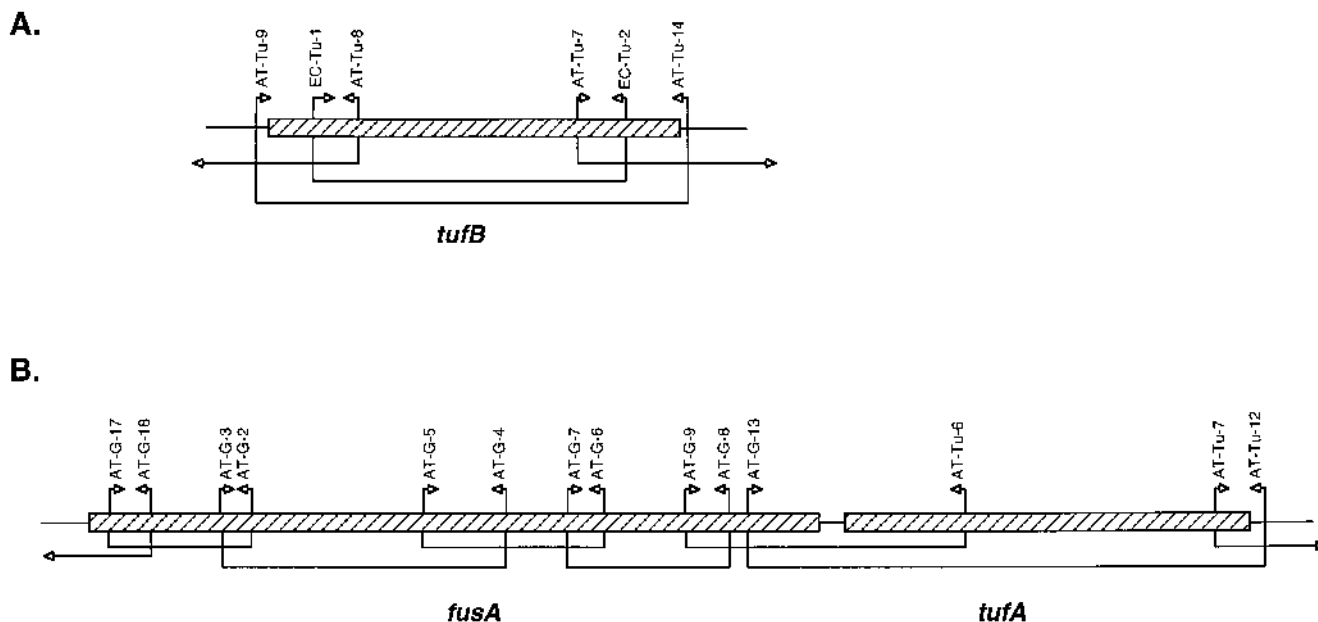


FIG. 1. Schematic representation of primers used for PCR amplifications of *tufB* (A), *fusA*, *tufA*, and their noncoding flanking sequences in *A. tumefaciens* (B). AT, *A. tumefaciens*; EC, *E. coli*.

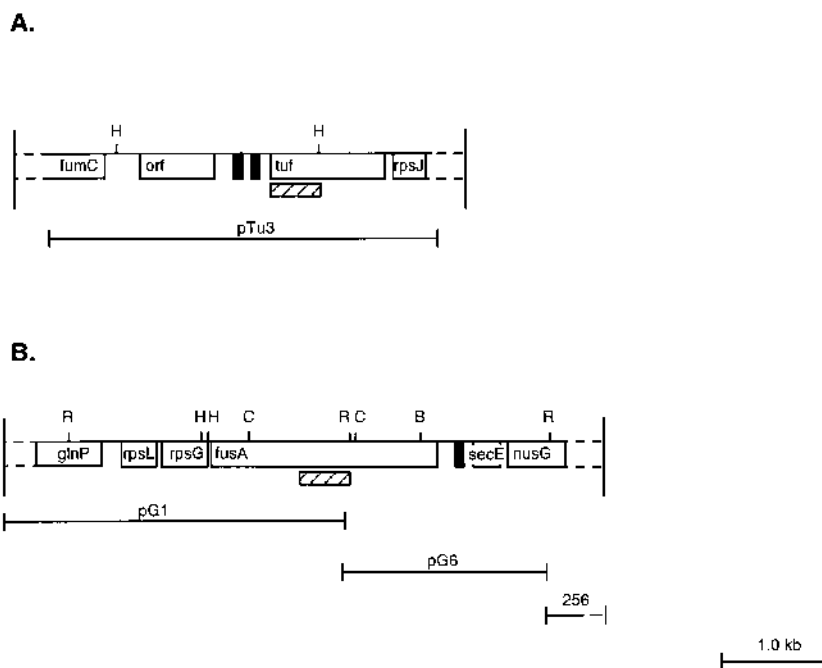


FIG. 2. Organization of the region surrounding the *tuf* gene (A) and the *fus* gene (B) in *R. prowazekii*. Lines indicate noncoding regions; boxes indicate coding regions. Filled boxes indicate tRNA genes. Abbreviations used to indicate restriction enzyme sites are as follows: B, *Bam*HI; R, *Eco*RI; C, *Hinc*II, and H, *Hind*III. The lengths of the sequenced DNA clones are indicated. The sizes and locations of the probes used in the Southern blot analysis (Fig. 4) are shown as hatched boxes below the *tuf* and *fus* genes.

quence databases under the accession numbers X99674 (Fig. 1A), X99673 (Fig. 1B), and Z54170 (Fig. 2A). Figure 2B shows schematically the nucleotide sequence assigned EMBL accession number Z54171 in combination with the sequence of Wood and Winkler, EMBL accession number U02603.

## RESULTS

**EF genes of *A. tumefaciens* and *R. prowazekii*.** Our strategy was to isolate the *A. tumefaciens* *tufA*, *tufB*, and *fus* genes for sequence determination directly from bacterial colonies (Fig. 1). First, we used PCR primer sequences based on conserved sequences of the corresponding genes in other bacteria (Table 1) to amplify internal fragments of the genes. The sequence information from these fragments then allowed us to design specific primers for amplification and sequencing of the 5' and 3' ends as well as of the noncoding sequences flanking the *tuf* and *fus* genes by a method called capture PCR (27). Finally, primers with sequences based on the 5' and 3' noncoding sequences flanking the *tuf* gene allowed us to amplify and sequence *tufA* and *tufB* separately. The *tufA* and *tufB* genes differ from each other by five nucleotides, none of which causes an amino acid change. It is evident from the successful PCR amplifications of *A. tumefaciens* genomic DNA with primers from the 3' end of the *fusA* gene in combination with primers from the *tufA* gene that these two genes are close to each other on the genome (Fig. 1B), as is normal in proteobacteria.

The *R. prowazekii* *tuf* and *fus* genes were isolated from a genomic library by plaque hybridization with probes specific for the respective genes. The probes for screening the *R. prowazekii* library were prepared by PCR amplification of the lambda phage stock with sets of nested degenerate primers (Table 1). Fifteen clones were identified by the *tuf* probe, and eight clones were identified by the *fus* probe. A restriction digest analysis suggested that all clones identified with the *tuf* probe were derived from the same genomic environment (data

not shown). Plasmids pTu3, pG1, and pG6 were chosen for sequence analysis (Fig. 2).

*R. prowazekii* has a genomic G+C content of 29% (52, 59) compared with a genomic G+C content of 57 to 63% in *A. tumefaciens* (56). The nucleotide sequences of *fus* and *tuf* in the two organisms reflect these biases, particularly at synonymous third codon positions (GC3<sub>s</sub>). In *R. prowazekii* the *fus* and *tuf* genes have GC3<sub>s</sub> values of 14 and 16%, respectively, close to the average GC3<sub>s</sub> value of 17% calculated from the known sequences of 27 protein-coding genes from *R. prowazekii* (4). In contrast, *tuf* and *fus* in *A. tumefaciens* have GC3<sub>s</sub> values of 69%, reflecting the higher G+C content of this organism. Despite the difference in G+C contents, the gene products of *tuf* and *fus* from *R. prowazekii* and *A. tumefaciens* are clustered in phylogenetic reconstructions on the basis of the sequences of EFs Tu and G from a variety of bacterial species (3a). The deduced amino acid sequences of the *tuf* and *fus* genes in *A. tumefaciens* and *R. prowazekii* are compared in Fig. 3A and B, respectively. The amino acid sequences for the EF-Tus from the two bacteria are 78% identical, while the two EF-Gs are 69% identical.

**One gene encoding EF-Tu in *R. prowazekii*.** Since *R. prowazekii* has a small-sized genome with a single copy of each of the rRNA genes (5, 41), we attempted to determine whether the gene encoding EF Tu is present in only one or two copies per genome. For this purpose, a Southern blot analysis of *R. prowazekii* genomic DNA digested with the enzymes *Hind*III, *Hinc*II, and *Eco*RI was performed. The *tuf*-specific probe was a 501-bp PCR product (nucleotides 34 to 535) amplified from genomic DNA (Fig. 2A). To confirm the identity of the probe, it was first hybridized to fragments obtained by *Hind*III digestions of pTu3 (Fig. 4A). As expected, the *tuf* probe hybridized to the 2-kb *Hind*III fragment generated from pTu3 (Fig. 4A and B). In the genomic digests, the *tuf* probe hybridized to a

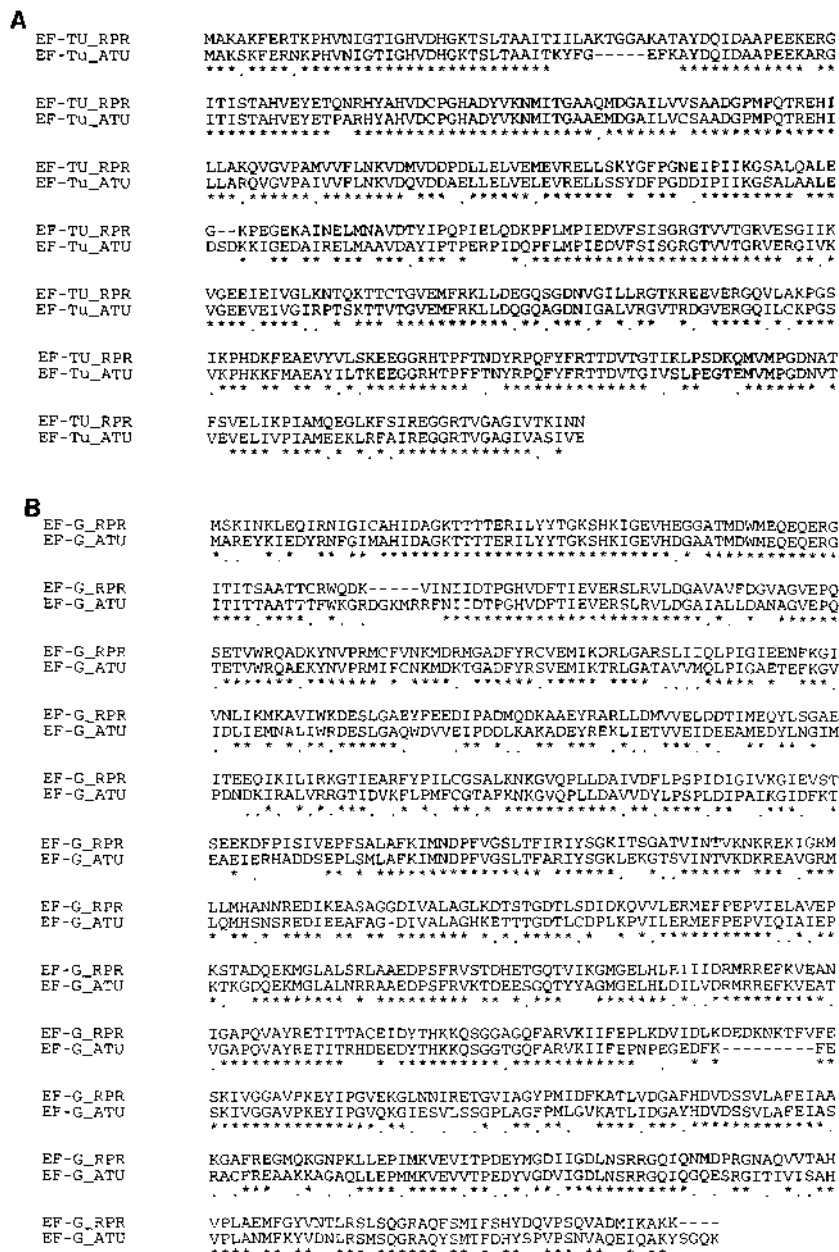


FIG. 3. Comparison of the deduced amino acid sequences of EF-Tu (A) and EF-G (B) in *A. tumefaciens* (ATU) and *R. prowazekii* (RPR). Symbols beneath the aligned sequences indicate identical residues (\*) and sites with conservative replacements (·).

single *Hind*III fragment, equivalent in size to that obtained by digestion of plasmid pTu3 (Fig. 4B). This fragment is too short to accommodate two separate *tuf* genes, which suggests that the *tuf* gene is present as a single copy in the *R. prowazekii* genome. Indeed, the *tuf* probe also hybridized to single fragments in the *Hinc*II and *Eco*RI digests. We conclude that the gene encoding EF-Tu is represented in the *R. prowazekii* genome only once.

**The *fus* gene and flanking sequences in *R. prowazekii*.** Since the genes encoding EFs Tu and G are known to be linked in many bacterial species, we examined the possibility that these two genes are also linked in *R. prowazekii*. For this purpose, we performed a Southern blot analysis comparing the genomic restriction patterns of the two genes. The *tuf*-specific probe was

as described above. The *fus*-specific probe was a 473-bp (nucleotides 912 to 1385) PCR product amplified from genomic DNA (Fig. 2B). This probe was found to hybridize to a 1-kb fragment from the *Hinc*II genomic digest (Fig. 4C). The size of this fragment is similar to that of the *Hinc*II-*Eco*RI fragment obtained from plasmid pG1, which was used as a positive control. Sequence data predict that a string of 2,787 nucleotides overlaps the sequence of the *fus* probe and is bordered by two *Eco*RI sites (Fig. 2B). This is in good agreement with the observed value of 2.8 kb for the fragment responding to the *fus* probe but different from that of the genomic *Eco*RI fragment responding to the *tuf* probe. In the genomic *Hind*III digest, the *fus* probe hybridized to a 4.2-kb fragment (Fig. 4C). This fragment is expected to contain the complete *fus* gene as well as 2.2

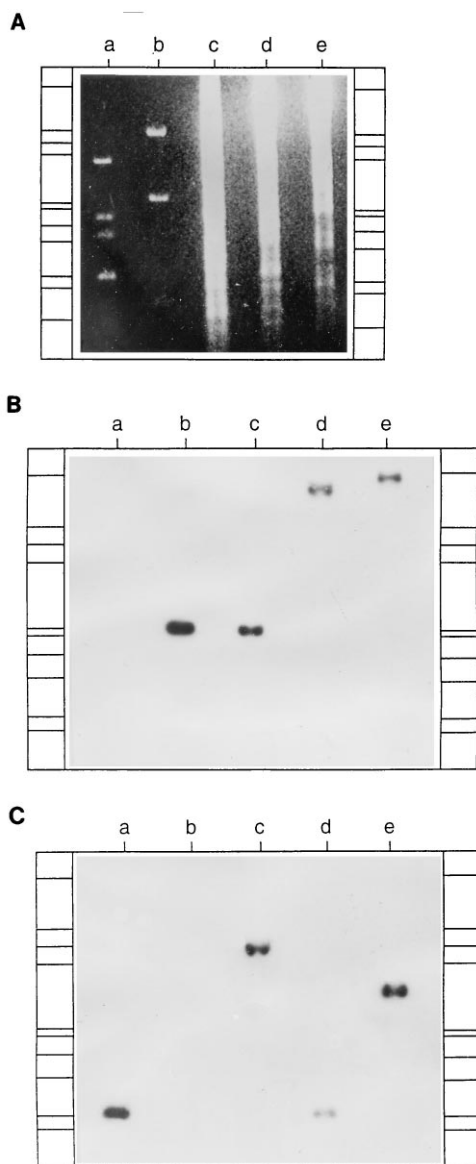


FIG. 4. Restriction site analysis of the region of the *R. prowazekii* chromosome containing the *tuf* and *fus* genes. (A) The agarose gel before blotting. (B) Hybridization signals generated with the *tuf* probe. (C) Hybridization signals generated with the *fus* probe. Lane a, clone pG1 digested with *HincII* and *EcoRI*; lane b, clone pTu3 digested with *HindIII*; lanes c to e, single digestions of chromosomal DNA with *HindIII*, *HincII*, and *EcoRI*, respectively. Molecular markers are products of lambda DNA digested with the enzymes *HindIII* and *EcoRI* and had the following band sizes (in kilobases): 21.2, 5.1, 5.0, 4.3, 3.5, 2.0, 1.9, 1.6, 1.3, 1.0, 0.8, and 0.6.

kb of sequence downstream of the *fus* gene. The *tuf* probe did not hybridize to a *HindIII* fragment that matches the one containing the *fus* gene (Fig. 4B). We conclude that *tuf* is not located downstream of *fus*.

In order to test this conclusion, we determined the nucleotide sequence of a 1.2-kb region downstream of *fus* in clone pG6. No *tuf*-specific sequences were detected. Instead, we noted a short sequence of less than 100 bp with a G+C content of more than 40%, which we identified as the gene encoding tRNA<sup>Trp</sup> (Fig. 2B). Further downstream we observed two complete open reading frames (ORFs). The first ORF was found to encode a protein of 66 amino acids that have significant

sequence similarity (30% amino acid identity) with the carboxy-terminal region of the *E. coli* SecE protein. The *secE* gene product is an integral cytoplasmic membrane protein that spans the *E. coli* inner membrane three times (51). Since the *R. prowazekii* SecE product lacks the first and second membrane-spanning regions characteristic of *E. coli*, it seems likely that it spans the membrane only once. This is not unusual, however, because organisms as diverse as *Bacillus subtilis*, *Staphylococcus carnosus*, *Thermotoga maritima*, and *Thermus thermophilus* have been found to encode SecE proteins in the size range of 60 to 70 amino acids (20, 24, 29, 33).

The second ORF downstream of the gene encoding tRNA<sup>Trp</sup> was found to be a homolog of the *E. coli nusG* gene, whose product has been implicated in transcription antitermination (28, 32). The nucleotide sequence of the *R. prowazekii nusG* gene predicts a protein of 178 amino acids whose sequence identity to the corresponding *E. coli* protein is 41%.

**The *tuf* gene and flanking sequences in *R. prowazekii*.** The *tufB* operon in *E. coli* contains the genes for four tRNA species (*thrU*, *tyrU*, *glyT*, and *thrT*) 100 bp upstream of the gene for EF Tu (Fig. 5). Since the single *tuf* gene in *R. prowazekii* is not located within the *str* operon, we investigated the 5' flanking sequences of the *tuf* gene for the presence of tRNA genes. Indeed, two tRNA structural genes, identified as tRNA<sup>Tyr</sup> and tRNA<sup>Gly</sup>, are located immediately 5' of *tuf*, with a 103-bp spacer between the two tRNA gene and 82 bp separating *glyT* from *tuf* (Fig. 2A). A G+C profile of a 2-kb region upstream of *tuf* identified two potential protein-coding genes located on the strand opposite the tRNA genes and the *tuf* gene (Fig. 2A). One of the two ORFs corresponds to a gene product from *E. coli* encoded by *fumC*. The other ORF did not correspond to any previously identified genes in the databanks.

Finally, sequence analysis of the 3' flanking region of the *tuf* gene identified a gene encoding the ribosomal protein S10 as well as the first 69 bases of a gene encoding the ribosomal protein L3.

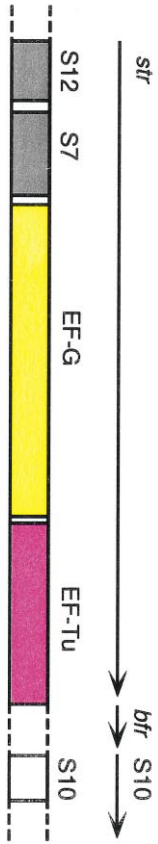
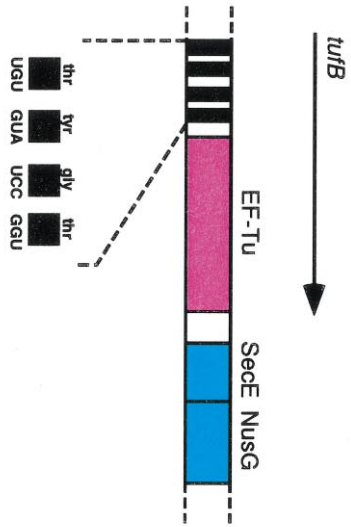
## DISCUSSION

We have compared the disposition of the EF genes in the genomes of *A. tumefaciens* and *R. prowazekii*. The free-living *A. tumefaciens* has two *tuf* genes, one of which is close to *fus*, in accordance with the arrangement of genes within the *str* operon (8, 34, 39, 42). In contrast, we find that the obligate endocellular parasite *R. prowazekii* has an arrangement that is exceptional for proteobacteria: one single *tuf* gene distant from the *fus* gene. Since we do not know the transcriptional units of these genes, we will refer to their genomic neighborhoods as motifs rather than operons.

The disposition of genes around the two *tuf* genes in *E. coli* is summarized in Fig. 5. Here, the gene order within the *str* operon is *rpsL*, *rpsG*, *fusA*, and *tufA* (42). *str* motifs are identically organized in a phylogenetically broad sample of other bacterial species (8, 34, 39, 58, 60). The same four genes are also arranged in the same order in archaea (7, 25). Furthermore, the gene encoding ribosomal protein S10 (*rpsJ*) is closely linked to the *tuf* gene in many species (22, 38, 48, 64). Taken together, these data suggest that the location of the *tuf* gene within the *str* motif represents an ancient ancestral organization.

In *E. coli*, the second *tuf* gene is found in the *tufB* operon, which features four tRNA genes (*thrU*, *tyrU*, *glyT*, and *thrT*) 100 bp upstream of *tufB* (3) (Fig. 5). The *tufB* operon is followed by the two genes *secE* and *nusG* (14). A *tufB*-like operon structure is also found in *Thermus thermophilus*, in which *secE* and *nusG* are located downstream of *tufB* (20, 50, 55).

# Escherichia coli



# Rickettsia prowazekii

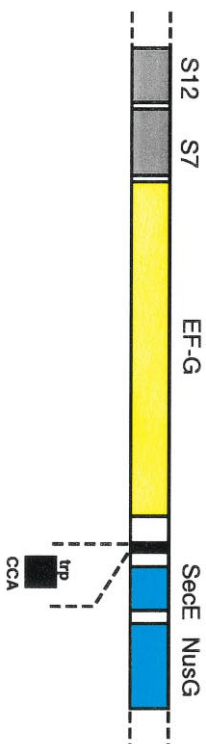
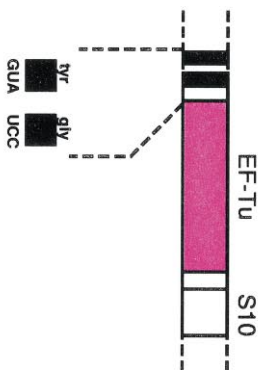


FIG. 5. Comparison of the chromosomal regions containing the genes encoding EFs Tu and G in *E. coli* and *R. prowazekii*.

In contrast, a single copy of the *tuf* gene located within the *str* operon seems to be the rule in the gram-positive lineages. In *B. subtilis* and *Bacillus licheniformis*, *secE* and *nusG* are preceded by the ribosomal protein gene *rpmG* (15, 24, 55). This pattern of genes is also observed in *Thermotoga maritima*, which contains a single *tuf* gene located within the *str* operon (29). We know of two other gram-negative eubacteria with single-copy *tuf* genes: the obligate intracellular parasite *Chlamydia trachomatis* (19) and the cyanobacterium *Anacystis nidulans* (35). In *Anacystis nidulans* the *tuf* gene is located within the *str* operon, in accordance with the general pattern. The arrangement of the EF genes in *C. trachomatis* remains to be determined. Bearing these exceptions in mind, in all other gram-negative species that have been studied there are two copies of the *tuf* gene (54). The genomic configurations of the EF genes in these bacteria can be compared with those we have observed in *R. prowazekii*.

First, two tRNA structural genes for tRNA<sup>Tyr</sup> and tRNA<sup>Gly</sup> are located immediately upstream of *tuf* in *R. prowazekii*, but the two other tRNA genes found in the same position in the *tufB* operon of *E. coli* (*thrU* and *thrT*) are missing from this region. A screen for tRNA genes in *R. prowazekii* has shown that *thrT* is located elsewhere in the genome, but *thrU* remains to be identified (21a). In addition, analysis of the downstream sequences flanking the *tuf* gene identify the genes for ribosomal proteins S10 and L4. Finally, the genes encoding tRNA<sup>Trp</sup> as well as *secE* and *nusG* have been observed within the sequence flanking *fus* on the 3' side. The sequences upstream of *fus* contain genes for the ribosomal proteins S12 and S7 (61a), as are found in modern proteobacteria.

Thus, in *R. prowazekii* the upstream sequences of the *tuf* gene are reminiscent of the string immediately upstream of *tufB* in *E. coli* while the downstream sequence is reminiscent of a sequence more distantly downstream from *tufA* in *E. coli* (Fig. 5). Additionally, in *R. prowazekii* we find that the region located upstream of *tufA* in *E. coli* is linked to a sequence that corresponds to the downstream region of *tufB* in *E. coli*. We interpret these findings to mean that an ancestor of *R. prowazekii* once had two *tuf* genes organized more or less like those in most present-day proteobacterial genomes. Further, we suggest that intrachromosomal recombination between these two ancestral *tuf* genes generated the atypical disposition of genes flanking the *tuf* and *fus* genes in *R. prowazekii*. A crossover event between the duplicated *tuf* genes oriented in this way could lead to an exchange between the ancestral flanking sequences without deletion of the intervening sequences (53). At this point there are insufficient data to determine whether the rearrangements and the deletions in these particular sequences occurred in the same or in serial events. Slightly more complicated scenarios can generate these rearrangements and deletions as well. For example, since tRNA genes are known to be hotspots for recombination events (10, 12, 21, 31, 36, 37, 43–45, 49), it is possible that the tRNA genes located in the region of analysis contributed to the observed rearrangement events.

In comparison, it is interesting that *Mycoplasma genitalium*, with a genome size of only 580 kb, also contains a partial *str* operon lacking the terminal *tuf* gene (17). However, in this species the genomic neighborhood of the single *tuf* gene provides no indication of a putative rearrangement mechanism. Since most gram-positive bacteria carry only a single copy of the *tuf* gene, it seems unlikely that the mechanisms of rearrangement are similar in the two species.

We have previously reported data concerning the simplification and rearrangement of the rRNA genes during the evolution of *R. prowazekii* (5). It is precisely these kinds of mosaic

patterns for which we searched when we embarked on an analysis of the *R. prowazekii* genome. Our expectations were based on the notion that obligate parasitic bacteria, such as *R. prowazekii*, are evolving under constraints much like those governing the evolution of organelle genomes. Here, dependence on host functions, selection for effective propagation and the influence of Muller's ratchet are all expected to promote an extensive reduction in genome size through the agency of intragenomic recombination and rearrangement (26). The interpretation of the findings presented in this paper is consistent with the idea that the small-sized *R. prowazekii* genome has evolved from an ancestral genome that once contained more genes than it has today.

#### ACKNOWLEDGMENTS

This work was supported by grants from the Natural Sciences Research Council, Swedish Cancer Society, and the Knud and Alice Wallenberg Foundation.

#### REFERENCES

- Allardent-Servent, A., S. Michaux-Charachon, E. Jumas-Bilak, L. Karyan, and M. Ramuz. 1993. Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. *J. Bacteriol.* **175**:7869–7874.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- An, G., and J. D. Friesen. 1980. The nucleotide sequence of *tufB* and four nearby tRNA structural genes of *Escherichia coli*. *Gene* **12**:33–39.
- Andersson, S. G. E., and C. G. Kurland. Unpublished data.
- Andersson, S. G. E., and P. M. Sharp. Codon usage and base composition in *Rickettsia prowazekii*. *J. Mol. Evol.*, in press.
- Andersson, S. G. E., A. Zomorodipour, H. Winkler, and C. G. Kurland. 1995. Unusual organization of the rRNA genes in *Rickettsia prowazekii*. *J. Bacteriol.* **177**:4171–4175.
- Auer, J., K. Lechner, and A. Böck. 1989. Gene organization and structure of two transcriptional units from *Methanococcus* coding for ribosomal proteins and elongation factors. *Can. J. Microbiol.* **35**:200–204.
- Auer, J., G. Spicker, L. Mayerhofer, G. Puhler, and A. Böck. 1991. Organization and nucleotide sequence of a gene cluster comprising the translational elongation factor EF-1 alpha from *Sulfolobus acidocaldarius*. *Syst. Appl. Microbiol.* **14**:14–22.
- Buttarelli, F. R., R. A. Caloger, O. Tiboni, C. O. Gualersi, and C. L. Pon. 1989. Characterization of the *str* operon from *Spirulina platensis* and their evolutionary relationship to those of other prokaryotes. *Mol. Gen. Genet.* **217**:97–104.
- 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.
- Campbell, A. 1992. Chromosomal insertion sites for phages and plasmids. *J. Bacteriol.* **174**:7495–7499.
- Casanova, J.-L., C. Pannetier, C. Jaulin, and P. Kourilsky. 1990. Optimal conditions for directly sequencing double-stranded PCR products with Sequenase. *Nucleic Acids Res.* **18**:4028.
- Chalker, D. J., and S. Sandmeyer. 1990. Transfer RNA genes are genomic targets for de novo transposition of the yeast retrotransposon TY3. *Genetics* **126**:837–850.
- Citovsky, V. C., G. McLean, E. Greene, E. Howard, G. Kuldau, Y. Thorntson, J. Zupan, and P. Zambryski. 1991. Agrobacterium-plant cell interaction: induction of vir genes and T-DNA transfer, p. 169–198. In D. P. S. Verma (ed.), *Molecular signals in plant-microbe communications*. CRC Press, Inc., Boca Raton, Fla.
- Downing, W. L., S. L. Sullivan, M. E. Gottesman, and P. P. Dennis. 1990. Sequence and transcriptional pattern of the essential *Escherichia coli* *secE-nusG* operon. *J. Bacteriol.* **172**:1621–1627.
- Dubnau, E., J. Weir, G. Nair, L. Carter, III, C. Moran, Jr., and I. Smith. 1988. *Bacillus* sporulation gene *spo0H* codes for  $\sigma^{30}$  ( $\sigma^H$ ). *J. Bacteriol.* **170**:1054–1062.
- Eremeeva, E. M., V. Roux, and D. Raoult. 1993. Determination of genome size and restriction pattern polymorphism of *Rickettsia prowazekii* and *Rickettsia typhi* by pulsed field gel electrophoresis. *FEMS Microbiol. Lett.* **112**:105–112.
- Fraser, C. M., J. D. Gocayne, O. White, M. Adams, R. Clayton, et al. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**:397–403.
- Furano, A. V. 1978. Direct demonstration of duplicate *tuf* genes in enteric bacteria. *Proc. Natl. Acad. Sci. USA* **75**:3104–3108.
- Goldstein, B. P., G. Zaffaroni, O. Tiboni, B. Amiri, and M. Denaro. 1989.

- Determination of the number of *tuf* genes in *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. FEMS Microbiol. Lett. **60**:305–310.
20. Heinrich, T., W. Schröder, V. A. Erdman, and R. K. Hartman. 1993. Identification of the gene encoding transcription factor *nusG* of *Thermus thermophilus*. J. Bacteriol. **174**:7859–7863.
  21. Howe, C. J., R. F. Barker, C. M. Bowman, and T. A. Dyer. 1988. Common features of three inversions in wheat chloroplast DNA. Curr. Genet. **13**:343–349.
  - 21a. Jamal, A. Unpublished data.
  22. Jaskunas, S. R., A. M. Fallon, and M. Nomura. 1977. Identification of organization of ribosomal protein genes of *Escherichia coli* carried by lambda-dafus2 transducing phage. J. Biol. Chem. **252**:7323–7336.
  23. Jaskunas, S. R., L. Lindhal, M. Nomura, and R. R. Burgess. 1975. Identification of two copies of the gene for elongation factor EF-Tu in *E. coli*. Nature (London) **257**:458–462.
  24. Jeong, S. M., H. Yoshikawa, and H. Takahashi. 1993. Isolation and characterization of the *secE* homologue gene of *Bacillus subtilis*. Mol. Microbiol. **10**:133–142.
  25. Keeling, P. J., R. L. Charlebios, and W. F. Doolittle. 1994. Archaeobacterial genomes: eubacterial form and eukaryotic content. Curr. Opin. Genet. Dev. **4**:816–822.
  26. Kurland, C. G. 1992. Evolution of mitochondrial genomes and the genetic code. Bioessays **14**:709–714.
  27. Lagerström, M., J. Parik, H. Malmgren, J. Stewart, U. Pettersson, and U. Landegren. 1991. Capture PCR: efficient amplification of DNA fragments adjacent to known sequence in human and YAC DNA. PCR Methods Appl. **1**:111–119.
  28. Li, J., R. Horwitz, S. McCracken, and J. Greenblatt. 1992. NusG, a new *Escherichia coli* elongation factor involved in transcriptional antitermination by the N protein of phage lambda. J. Biol. Chem. **267**:6012–6019.
  29. 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.
  30. Ludwig, W., M. Weizenegger, D. Betzl, E. Leidel, T. Lenz, A. Ludvigsen, D. Möllenhoff, P. Wenzig, and K. H. Schleifer. 1990. Complete nucleotide sequences of seven eubacterial genes coding for elongation factor Tu: functional, structural and phylogenetic evaluations. Arch. Microbiol. **153**:241–247.
  31. Marschalek, R., T. Brechner, E. Amon-Böhm, and T. Dingermann. 1989. Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. Science **244**:1493–1496.
  32. Mason, S. W., and J. Greenblatt. 1991. Assembly of transcription elongation complexes containing the N protein of phage lambda and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. Genes Dev. **5**:1504–1512.
  33. Meens, J., M. Klose, and R. Freudl. 1993. The *Staphylococcus carnosus secE* gene: cloning, nucleotide sequence, and functional characterization in *Escherichia coli secE* mutant strains. FEMS Microbiol. Lett. **117**:113–119.
  34. Meng, B. Y., K. Shinozaki, and S. Masahiro. 1989. Genes for the ribosomal proteins S12 and S7 and elongation factors EF-G and EF-Tu of the cyanobacterium *Anacystis nidulans*: structural homology between 16S rRNA and S7 mRNA. Mol. Gen. Genet. **216**:25–30.
  35. Mickel, F. S., and L. L. Spremulli. 1986. Organization of the genes for protein synthesis elongation factors Tu and G in the cyanobacterium *Anacystis nidulans*. J. Bacteriol. **166**:78–82.
  36. Moritz, C., and W. M. Brown. 1986. Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science **233**:1425–1427.
  37. Nelson, I., F. Degoul, B. Obermaier-Kusser, N. Romero, C. Borrone, C. Marsac, J. L. Vaysierre, K. Gerbitz, M. Fardeau, F. Ponsot, and P. Lesienne. 1989. Mapping of heteroplasmic mitochondrial DNA deletions in Kearns-Sayre syndrome. Nucleic Acids Res. **17**:8117–8124.
  38. Neumann-Spallart, C., J. Jakowitsch, M. Kraus, M. Brandtner, H. J. Bohnert, and W. Löffelhardt. 1991. *rps10*, unreported for plastid DNAs, is located on the cyanelle genome of *Cyanophora paradoxa* and is cotranscribed with the *str* operon genes. Curr. Genet. **19**:313–315.
  39. Ohama, T., F. Yamao, A. Muto, and S. Osawa. 1987. Organization and codon usage of the streptomycin operon in *Micrococcus luteus*, a bacterium with a high genomic G+C content. J. Bacteriol. **169**:4770–4777.
  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. Pang, H., and H. H. Winkler. 1993. Copy number of the 16S rRNA gene in *Rickettsia prowazekii*. J. Bacteriol. **175**:3893–3896.
  42. Post, L. E., and M. Nomura. 1980. DNA sequences from the *str* operon of *Escherichia coli*. J. Biol. Chem. **255**:4660–4666.
  43. Quigley, F., and J. H. Weil. 1985. Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. Curr. Genet. **9**:495–503.
  44. Reiter, W. D., P. Palm, and Y. Siobhan. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17**:1907–1914.
  45. Ritter, A., G. Blum, L. Emödy, M. Kerenyi, A. Böck, B. Neuhierl, W. Rabsch, F. Scheutzi, and J. Hacker. 1995. tRNA genes and pathogenicity islands: influence on virulence and metabolic properties of uropathogenic *Escherichia coli*. Mol. Microbiol. **17**:109–121.
  46. Rossi, J. J., and Y. Landy. 1979. Structure and organization of the two tRNA<sup>Tyr</sup> gene clusters on the *E. coli* chromosome. Cell **16**:523–534.
  47. 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.
  48. 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.
  49. Sandmeyer, S. B., V. W. Bilanchone, D. J. Clark, P. Morcos, G. Carle, and G. M. Brodeur. 1988. Sigma elements are position-specific for many different yeast tRNA genes. Nucleic Acids Res. **16**:1499–1515.
  50. Satoh, M., T. Tanaka, A. Kushiro, T. Hakoshima, and K.-I. Tomita. 1991. Molecular cloning, nucleotide sequence and expression of the *tufB* gene encoding elongation factor Tu from *Thermus thermophilus* HB8. FEBS Lett. **288**:98–100.
  51. Schatz, P. J., K. L. Bieker, K. M. Otteman, T. J. Silhavy, and J. Beckwith. 1991. The *secE* gene encodes an integral membrane protein required for protein export in *Escherichia coli*. EMBO J. **10**:1749–1757.
  52. Schramek, S. 1972. Deoxyribonucleic acid base composition of members of the typhus group of rickettsiae. Acta Virol. **16**:447.
  53. Segall, A., J. M. Mahan, and J. R. Roth. 1988. Rearrangement of the bacterial chromosome: forbidden inversions. Science **241**:1314–1318.
  54. Sela, S., D. Yogeve, S. Razin, and H. Bercovier. 1989. Duplication of the *tuf* gene: a new insight into the phylogeny of eubacteria. J. Bacteriol. **171**:581–584.
  55. Sharp, P. M. 1994. Identification of genes encoding ribosomal protein L33 from *Bacillus licheniformis*, *Thermus thermophilus* and *Thermotoga maritima*. Gene **139**:135–136.
  56. Stackebrandt, E., and W. Liesack. 1993. Nucleic acids and classification, p. 151–194. In M. Goodfellow and A. G. O'Donell (ed.), Handbook of new bacterial systematics. Academic Press, London.
  57. Syvänen, A.-C., and H. Söderlund. 1993. Quantification of polymerase chain reaction products by affinity-based collection. Methods Enzymol. **218**:474–490.
  58. Tiboni, O., R. Cantoni, R. Creti, P. Camarano, 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.
  59. Tyeryar, F. J., E. Weiss, D. B. Millar, F. M. Bozeman, and R. A. Ormsbee. 1973. DNA base composition of *Rickettsiae*. Science **180**:415–417.
  60. Vijgenboom, E., L. P. Woudt, P. W. H. Heinstra, K. Rietveld, J. van Haarlem, J. P. van Wezel, S. Shochat, and L. Bosch. 1994. Three *tuf*-like genes in the kirromycin producer *Streptomyces ramocissimus*. Microbiology **140**:983–998.
  61. Wada, K., Y. Wada, F. Ishibashi, T. Gjobori, and T. Ikemura. 1992. Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res. **20**:2111–2118.
  - 61a. Wood, D. A., and H. H. Winkler. EMBL accession number U02603.
  62. Yamamoto, M., and M. Nomura. 1979. Organization of genes for transcription and translation in the *rif* region of the *Escherichia coli* chromosome. J. Bacteriol. **137**:584–594.
  63. Zambryski, P., J. Tempe, and J. Schell. 1989. Transfer and function of T-DNA genes from *Agrobacterium* Ti and Ri plasmids in plants. Cell **56**:193–201.
  64. Zurawski, G., and S. M. Zurawski. 1985. Structure of the *Escherichia coli* S10 ribosomal protein operon. Nucleic Acids Res. **13**:4521–4526.