JOURNAL OF BACTERIOLOGY, Jan. 1989, p. 581–584
0021-9193/89/010581-04$02.00/0

Vol. 171, No. 1

Duplication of the *tuf* Gene: a New Insight into the Phylogeny of Eubacteria

S. SELA,[1] D. YOGEV,[2] S. RAZIN,[2] AND H. BERCOVIER[1]*

Department of Clinical Microbiology[1] and Department of Membrane and Ultrastructure Research,[2] The Hebrew University—Hadassah Medical School, Jerusalem, 91010, Israel

Received 28 March 1988/Accepted 5 October 1988

The conservation and duplication of the *tuf* gene encoding the elongation factor EF-Tu were used to define phylogenetic relationships among eubacteria. When the *tufA* gene of *Escherichia coli* was used as a probe in hybridization experiments, duplicate *tuf* genes were found in gram-negative bacteria from three major phyla: purple bacteria, bacteroides, and cyanobacteria. Only a single copy of *tuf* was found in gram-positive bacteria, including mycobacteria and mycoplasmas. Gram-positive clostridia were found to carry two copies of *tuf*.

Genome duplication served as a remarkably effective means of gene amplification in evolution (16, 20). These amplifications were probably followed by rearrangements and mutations at the genome level under selective pressure. If one copy of a duplicate gene is sufficient for the survival of an organism, it would be assumed that the superfluous copy of the duplicate gene would rapidly diverge from the former. In fact, duplicate conserved genes with a high degree of nucleotide homology are rare in procaryotes (16). The *tuf* gene, which encodes the elongation factor EF-Tu, is duplicated in *Escherichia coli* (8, 9) as well as in other gram-negative bacteria thus far studied, whereas in the two gram-positive bacteria examined a single copy of the *tuf* gene was found (6). The two copies of the *E. coli tuf* gene are exceptional in that they differ from one another in the position of only 13 of about 1,200 nucleotides which make up the gene (1, 27). Filer et al. (4) reported evidence suggesting that the presence in *E. coli* of these two copies resulted from chromosomal duplication rather than from chromosomal gene transfer.

The near identity of the two copies of the *tuf* gene in *E. coli*, the presence of homologous sequences between the *E. coli tuf* gene and *tuf* genes of phylogenetically distant bacteria (5, 14), and the essential role of the elongation factor in bacterial translation raise the possibility of using the *tuf* gene as an effective marker in studies of bacterial phylogeny. The study presented here examines the number of *tuf* genes in bacterial species belonging to various phylogenetic branches of both gram-negative and gram-positive eubacteria (7, 23). An attempt is made to use the data to evaluate the significance of *tuf* gene duplication in the phylogeny of eubacteria and to compare the results with phylogenetic trees prepared on the basis of 5S rRNA (15) and 16S rRNA (23–26) sequences.

To determine the number of *tuf* genes in bacterial chromosomes, three probes were used. The first probe was ptufA, a plasmid which contains the *tufA* gene and its flanking regions from *E. coli* (18); the second was probe A, containing 600 base pairs from the 3′ end of the gene; and the third was probe B, containing 300 base pairs of the 5′ portion of the gene. Probes A and B were fragments derived from plasmids pT/G1 and pT1, respectively (4).

The bacterial species tested (Fig. 1) include representatives of four gram-negative superfamilies (3) and of the phyla (23) cyanobacteria, bacteroides, purple bacteria, and gram-positive bacteria (including mycoplasmas).

Bacterial DNAs were extracted and purified by the method of Marmur (13) or by a modification of this method (2). The DNAs were digested by restriction enzymes (Pharmacia) for 2 h at 37°C with the buffer mixtures recommended by the manufacturer. The digests were subjected to electrophoresis at 30 V for 18 h in 0.8% agarose (SeaKem) slab gels in 0.04 M Tris-acetate and 0.001 M EDTA-disodium salt. Gels were stained with ethidium bromide and photographed to visualize DNA bands. DNA fragments were transferred to nitrocellulose sheets by the method of Southern (19). Probes were labeled with [^{32}P]dCTP by using a multiprimer labeling kit (Amersham Corp.). Hybridization of probes with DNA fragments on the nitrocellulose sheets was performed as described by Maniatis et al. (12) in a solution containing 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 65°C for 18 h. The sheets were then washed three times, 30 min each, in 2× SSC plus 0.1% sodium dodecyl sulfate at room temperature and then once for 15 min in the same solution at 65°C. After being air dried, the nitrocellulose sheets were exposed at −70°C to Agfa Gevaert X-ray film with an intensifying screen.

The bacterial DNAs were digested with the following restriction enzymes: *Bam*HI, *Bgl*II, *Ban*II, *Cla*I, *Hin*dIII, *Hin*fI, *Pst*I, and *Sma*I. They were then hybridized with ptufA. To exclude the possibility that two *tuf* genes were located on the same fragment, only those restriction enzymes that revealed at least two hybridization bands were chosen for further analysis. The digested DNAs were electrophoresed, and duplicate Southern blots were prepared and hybridized with probe A or probe B.

The resulting hybridization patterns are shown in Fig. 1. Two or three hybridization bands were observed for each probe with the DNAs of all of the gram-negative bacteria tested (Fig. 1a). The presence of at least two hybridization bands revealed by each probe, each larger than 1.2 kilobase pairs (the size of the *E. coli tuf* gene [27]), indicates that the genome of all of the gram-negative bacteria tested carries two genes with sequences homologous to the *tuf* genes of *E. coli*. By contrast, hybridization of the digested DNAs of the gram-positive bacteria and the mycoplasmas tested yielded one or two hybridization bands with probe A and only one band with probe B (Fig. 1b). The presence of only one hybridization band revealed by probe B indicates that these bacteria contain only one *tuf* gene. In the cases of *Myco-*

* Corresponding author.

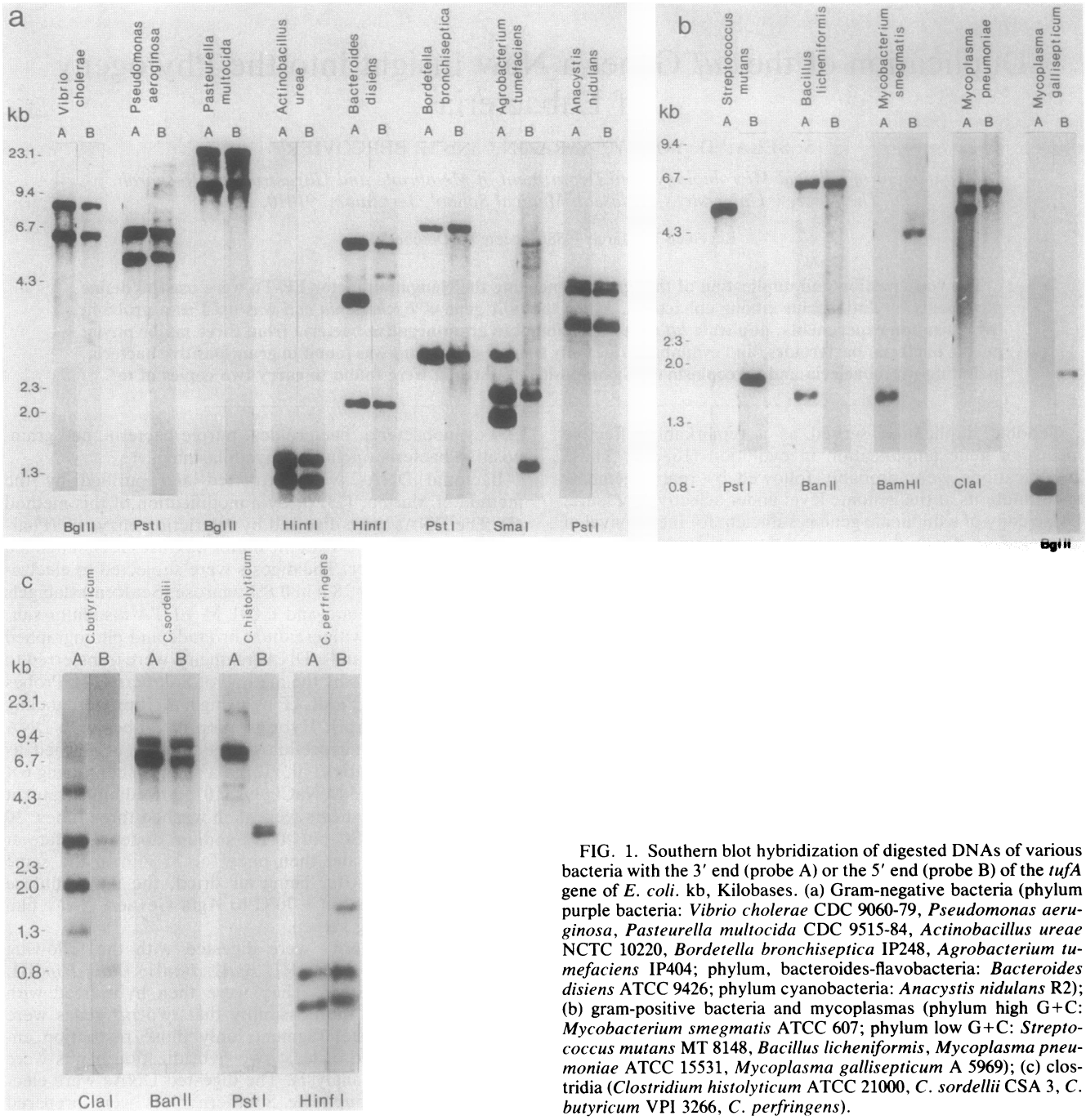

FIG. 1. Southern blot hybridization of digested DNAs of various bacteria with the 3′ end (probe A) or the 5′ end (probe B) of the *tufA* gene of *E. coli*. kb, Kilobases. (a) Gram-negative bacteria (phylum purple bacteria: *Vibrio cholerae* CDC 9060-79, *Pseudomonas aeruginosa*, *Pasteurella multocida* CDC 9515-84, *Actinobacillus ureae* NCTC 10220, *Bordetella bronchiseptica* IP248, *Agrobacterium tumefaciens* IP404; phylum, bacteroides-flavobacteria: *Bacteroides disiens* ATCC 9426; phylum cyanobacteria: *Anacystis nidulans* R2); (b) gram-positive bacteria and mycoplasmas (phylum high G+C: *Mycobacterium smegmatis* ATCC 607; phylum low G+C: *Streptococcus mutans* MT 8148, *Bacillus licheniformis*, *Mycoplasma pneumoniae* ATCC 15531, *Mycoplasma gallisepticum* A 5969); (c) clostridia (*Clostridium histolyticum* ATCC 21000, *C. sordellii* CSA 3, *C. butyricum* VPI 3266, *C. perfringens*).

plasma pneumoniae digested by *Cla*I and *Bacillus licheniformis* digested by *Ban*II, the presence of two hybridization bands with probe A (Fig. 1b) can be explained by the existence of a restriction site for these enzymes within the 3′ part of the gene.

The hybridization patterns (Fig. 1c) obtained with digested DNAs from the gram-positive clostridia *Clostridium sordellii*, *C. butyricum*, *C. perfringens*, and *C. histolyticum* resembled those obtained with the DNAs of the gram-negative bacteria in indicating the presence of two *tuf* genes. Probe A revealed two fragments with *C. histolyticum* DNA which are very close in this figure. However, two defined bands were revealed by probe A with other enzymes (data not shown).

Our data combined with previously reported data (6, 21) indicate that duplication of the *tuf* gene is common among gram-negative bacteria, no matter how distant they are on the various phylogenetic trees proposed (15, 23). In fact, among the gram-negative bacteria that we examined were species representative of the three main phyla described by Woese (23), i.e., the purple bacteria, the flavobacteria-bacteroides, and the cyanobacteria. Our data for the gram-positive bacteria are more puzzling. According to oligonucleotide sequence catalogs of the 16S rRNA (7, 23) and to 5S rRNA sequence analyses (11, 15, 17), gram-positive bacteria are divided into two major subdivisions: the low-G+C (guanine-plus-cytosine content of the DNA) subdivision,

which includes the genera *Bacillus*, *Streptococcus*, and *Clostridium* and mycoplasmas (class, *Mollicutes*); and the high-G+C subdivision, which includes the genera *Mycobacterium*, *Streptomyces*, and *Actinomyces*. Only one *tuf* gene was found in the genome of *Streptococcus mutans*, *B. licheniformis*, and the two wall-less mycoplasmas, *Mycoplasma gallisepticum* and *Mycoplasma pneumoniae*, all representative of the low-G+C subdivision. One *tuf* gene was also found in the DNA of *Mycobacterium smegmatis* (Fig. 1b) and in seven other *Mycobacterium* species (data not shown), representatives of the high-G+C subdivision. Our findings corroborate those of Filer and Furano (6), who found one copy of the *tuf* gene in *Bacillus subtilis* (low G+C) and in *Micrococcus luteus* (high G+C). Surprisingly, all four clostridia that we studied were found to carry two *tuf* genes, like gram-negative bacteria. The clostridia we tested belong to group I of the genus *Clostridium* as defined by Johnson and Francis (10). Despite the heterogeneity of this genus, its members were included in the gram-positive low-G+C subdivision described by Woese et al. (26).

If we accept that the duplication of the *tuf* gene did not result from gene transfer, as was shown for *E. coli* (4), and assume that convergence of base sequences is unlikely to be the reason for the homologies found among the *tuf* genes of evolutionarily distant bacteria, then the *tuf* gene can be considered a very ancient gene with highly conserved sequences reminiscent of the ancestral parent gene. Therefore, deep links among eubacteria can be deduced from the conservation of homologous sequences of the *tuf* gene and from its presence as a single or a duplicated gene.

According to the assumptions stated above, gram-negative bacteria and the clostridia diverged very early from a common ancestor in which the *tuf* gene was already duplicated. Since we were able to detect only one *tuf* gene with sequences homologous to the *E. coli tuf* gene in gram-positive bacteria other than clostridia, it could mean that one of the two gene copies diverged to such an extent that it does not share any more homologous sequences with the *tufA* probe. We can assume, therefore, that the low-G+C anaerobic clostridia are the most ancient gram-positive bacteria, as was already suggested (26). Our results support the ideas of Woese (23), who states that "in contrast to the high G+C group, the low G+C gram-positive bacteria form a phylogenetically deep, and therefore presumably ancient, cluster."

This article reports the use of a phylogenetic marker, the *tuf* gene, different from the extensively studied rRNA genes. The *tuf* gene is an adequate gene for phylogenetic measurement because it is ubiquitous and fulfils a highly constrained function where neutral mutations are unlikely to be retained; however, the duplication of the gene in many bacteria and the fact that only one copy is apparently necessary for the bacterial translation system (1, 22) would be expected to provide the superfluous copy with enough latitude for divergence, which could be used to measure more recent evolutionary processes.

We thank R. Davidpour for her technical assistance; A. V. Furano for providing plasmids; D. J. Brenner, J. L. Johnson, R. Barletta, M. R. Popoff, and E. Padan for bacterial DNA; D. Filer for his helpful consultation; and N. Citri and T. G. Sacks for critically reading the manuscript.

LITERATURE CITED

1. **An, G., and J. D. Friesen.** 1980. The nucleotide sequence of *tuf*B and four nearby tRNA structural genes of *Escherichia coli*. Gene **12**:33–39.
2. **Bercovier, H., O. Kafri, and S. Sela.** 1986. *Mycobacteria* possess a surprisingly small number of ribosomal RNA genes in relation to the size of their genome. Biochem. Biophys. Res. Commun. **136**:1136–1141.
3. **De Ley, J.** 1983. Phylogenetic relationships between gram-negative bacteria, p. 205–210. *In* H. Leclerc (ed.), Gram negative bacteria of medical and public health importance; taxonomy-identification-applications, vol. 114. Institut National de la Santé, et de la Recherche Médicale, Paris.
4. **Filer, D., R. Dhar, and A. V. Furano.** 1981. The conservation of DNA sequences over very long periods of evolutionary time. Evidence against intergenic chromosomal transfer as an explanation for the presence of *E. coli tuf* gene sequences in taxonomically-unrelated prokaryotes. Eur. J. Biochem. **120**:69–77.
5. **Filer, D., and A. V. Furano.** 1980. Portions of the gene encoding elongation factor Tu are highly conserved in prokaryotes. J. Biol. Chem. **255**:728–734.
6. **Filer, D., and A. V. Furano.** 1981. Duplication of the *tuf* gene, which encodes peptide chain elongation factor Tu, is widespread in gram-negative bacteria. J. Bacteriol. **148**:1006–1011.
7. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, G. J. Lewis, D. A. Stahl, K. R. Leuhrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **200**:457–463.
8. **Furano, A. V.** 1978. Direct demonstration of duplicate *tuf* genes in enteric bacteria. Proc. Natl. Acad. Sci. USA **75**:3104–3108.
9. **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.
10. **Johnson, J. L., and B. S. Francis.** 1975. Taxonomy of the clostridia: ribosomal ribonucleic acid homologies among the species. J. Gen. Microbiol. **88**:229–244.
11. **Hori, H., and S. Osawa.** 1979. Evolutionary change in 5S RNA secondary structure and a phylogenetic tree of 54 5S rRNA species. Proc. Natl. Acad. Sci. USA **76**:381–385.
12. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. **3**:208–218.
14. **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. **167**:4770–4777.
15. **Park, Y. H., H. Hori, K. I. Suzuki, S. Osawa, and K. Komagato.** 1987. Phylogenetic analysis of the coryneform bacteria by 5S rRNA sequences. J. Bacteriol. **169**:1801–1806.
16. **Riley, M., and A. Anilionis.** 1978. Evolution of the bacterial genome. Annu. Rev. Microbiol. **32**:519–560.
17. **Rogers, M., J. Simmons, R. T. Walker, W. G. Weisburg, C. R. Woese, R. S. Tanner, I. M. Robinson, D. A. Stahl, G. Olsen, R. H. Lench, and J. Maniloff.** 1985. Construction of the mycoplasma evolutionary tree from 5S—rRNA sequence data. Proc. Natl. Acad. Sci. USA **82**:1160–1164.
18. **Shibuya, M., H. Nashimoto, and Y. Kaziro.** 1979. Cloning of an *Eco*RI fragment carrying *E. coli tuf A* gene. Mol. Gen. Genet. **170**:231–234.
19. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
20. **Sparrow, A. H., and A. F. Nauman.** 1976. Evolution of genome size by DNA doublings. Science **192**:524–529.
21. **Tiboni, O., G. Di Pasquale, and O. Ciferri.** 1984. Two *tuf* genes in the cyanobacterium *Spirulina platensis*. J. Bacteriol. **159**: 407–409.
22. **Van der Meick, P. H., E. Vijgenboom, M. Dicke, and L. Bosch.** 1982. Regulation of the expression of *tuf*A and *tuf*B, the two genes coding for the elongation factor EF-Tu in *Escherichia coli*. FEBS Lett. **134**:325–330.
23. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51**:221–271.

24. **Woese, C. R., J. Maniloff, and L. B. Zablen.** 1980. Phylogenetic analysis of the mycoplasmas. Proc. Natl. Acad. Sci. USA **77:** 494–498.
25. **Woese, C. R., E. Stackebrandt, and W. Ludwig.** 1985. What are mycoplasmas: the relationship of tempo and mode in bacterial evolution. J. Mol. Evol. **21:**305–316.
26. **Woese, C. R., E. Stackebrandt, T. J. Macke, and E. F. Fox.** 1985. A phylogenetic definition of the major eubacterial taxa. Syst. Appl. Microbiol. **6:**143–151.
27. **Yakota, T., H. Sugisaki, M. Takanami, and Y. Kaziro.** 1980. The nucleotide sequence of the cloned *tuf* gene of *E. coli*. Gene **12:** 25–31.