Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 9892–9895, November 1993
Microbiology

Occurrence of fragmented 16S rRNA in an obligate bacterial endosymbiont of *Paramecium caudatum*

NINA SPRINGER*, WOLFGANG LUDWIG*†, RUDOLF AMANN*, HELMUT JOSEF SCHMIDT‡, HANS-DIETER GÖRTZ§, AND KARL-HEINZ SCHLEIFER*

*Lehrstuhl für Mikrobiologie, Technische Universität, Arcisstrasse 21, D-80290 Munich, Federal Republic of Germany; ‡Institut für Allgemeine Zoologie und Genetik, Universität Münster, Schlossplatz 5, D-48149 Münster, Federal Republic of Germany; and §Biologisches Institut, Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Federal Republic of Germany

Communicated by Carl R. Woese, June 10, 1993

ABSTRACT **The phylogenetic position of *Caedibacter caryophila*, a so far noncultured killer symbiont of *Paramecium caudatum*, was elucidated by comparative sequence analysis of *in vitro* amplified 16S rRNA genes (rDNA). *C. caryophila* is a member of the α subclass of the *Proteobacteria* phylum. Within this subclass *C. caryophila* is moderately related to *Holospora obtusa*, which is another obligate endosymbiont of *Paramecium caudatum*, and to *Rickettsia*. A 16S rRNA targeted specific hybridization probe was designed and used for *in situ* detection of *C. caryophila* within its host cell. Comparison of the 16S rDNA primary structure of *C. caryophila* with homologous sequences from other bacteria revealed an unusual insertion of 194 base pairs within the 5′-terminal part of the corresponding gene. The intervening sequence is not present in mature 16S rRNA of *C. caryophila*. It was demonstrated that *C. caryophila* contained fragmented 16S rRNA.**

Bacterial endosymbionts are often found in protozoa (1). Some bacterial endosymbionts of paramecia, including all members of the genus *Caedibacter*, are toxic for susceptible strains of paramecia (2). *Caedibacter* species are distinguished from other killing endosymbionts by their ability to produce unusual refractile inclusion bodies, so-called R bodies (2). R bodies are long (up to 20 μm) proteinaceous ribbons (approximately 0.5 μm wide and 1.3 μm thick) that are tightly rolled up within the bacterial cells. Usually fewer than 10% of the cells in any given population contain R bodies. These cells are called bright forms (acknowledging their refractility in phase-contrast microscopy), whereas R body-free cells are referred to as nonbright forms (3). As far as is known only the nonbright forms reproduce. They can change into bright forms by producing R bodies. The R bodies of *Caedibacter caryophila* can be morphologically distinguished from those of other *Caedibacter* species (4). *C. caryophila* can be found as an obligate endosymbiont mostly in the macronucleus of *Paramecium caudatum* (4), whereas the other *Caedibacter* species occur as cytoplasmic endosymbionts in *Paramecium biaurelia* and *Paramecium tetraurelia* (2, 6).

C. caryophila cells can be enriched by centrifugation (4) but so far have not been cultivated on artificial medium. In the present study the phylogenetic position of the species and pecularities of its 16S rRNA structure are shown.

MATERIALS AND METHODS

Purification of Nucleic Acids. Enrichment of *C. caryophila* (type strain 221, carried in *Paramecium caudatum* C221 ATCC 50168) cells was done as described (7). Genomic DNA was purified according to Schmidt *et al.* (7). Cellular RNA was extracted by the method of Oelmüller *et al.* (8).

Sequence Analysis. The 16S rRNA genes were amplified *in vitro* by the polymerase chain reaction (PCR) technique (9). The oligodeoxynucleotide primers were 5′-AGAGTTTGATYMTGGCTCAG-3′ (*Escherichia coli* positions 8–27; ref. 10) and 5′-AKAAAGGAGGTGATCC-3′ (*E. coli* positions 1529–1544). The amplified DNA was sequenced directly by using the T7Sequencing kit of Pharmacia. The oligonucleotide primers were obtained from MWG-Biotech Gesellschaft fuer angewandte Biotechnologie (Ebersberg, F.R.G.). Reverse transcriptase sequencing was done with the RNA sequencing kit of Boehringer Mannheim.

Southern and Northern Hybridizations. DNA and RNA preparations were subjected to agarose gel electrophoresis and the separated nucleic acids were subsequently transferred to Zeta-Probe membranes (Bio-Rad) as recommended by the manufacturer, using the vacuum blotting system of Pharmacia. The oligonucleotide probes were labeled with [γ-^{32}P]ATP (NEN) by T4 polynucleotide kinase (Boehringer Mannheim). Hybridizations using the 16S rRNA-specific probes 620R (5′-TTACTCACCCDTBYGC-3′) and Eub338 (11) were carried out at 48°C in 5× standard saline citrate (SSC; 1× SSC is 0.15 M NaCl/0.015 M sodium citrate, pH 7.0)/7% SDS/20 mM phosphate buffer overnight. The membranes were washed twice in 2× SSC/0.1% SDS at 48°C. Hybridizations with the insertion-specific probe CCIns (5′-GGCCCTTTCCTTCACCCAA-3′) were performed under the same conditions, but washing was done at 52°C. The membranes were stripped in 1× SSC/0.5% SDS at 80°C for reprobing.

***In Situ* Whole-Cell Hybridization.** *In situ* whole-cell hybridizations were carried out as described (12).

Phylogenetic Analysis. The *C. caryophila* 16S rRNA sequence was aligned with about 1300 homologous sequences of bacteria (13, 14). Phylogenetic trees were reconstructed by applying distance-matrix, parsimony, and maximum-likelihood methods. The corresponding computer programs were NEIGHBOR and DNAPARS of the PHYLIP package (15), as well as fastDNAml (14).

RESULTS

Sequence Analysis. A 1695-bp DNA fragment encoding 16S rRNA was amplified *in vitro* from purified *C. caryophila* DNA and directly sequenced (Fig. 1).¶

Phylogeny. The nearly complete 16S rRNA primary structure from *C. caryophila* was aligned with homologous sequences of other bacteria. Different methods of tree reconstruction were applied on a variation of data sets. These data sets differed with regard to the selection of reference sequences as well as of alignment positions. The latter were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. X71837).

```
                                                                    620R
  1 agaguuugauymuggcucagAACGAACGCUGGCGGCACGCCUAACACAUGCAAGUCGAACGAGGGCAUUCUUUCGGGGAUGUUCCUAGUGGCGGACGGGU
101 GAGUAACGCGUGGGAAUCUGCCUGAUAGUGGGGGAUAACUGUUGGAAACGACAGCUAAUACCGCAUACGCUGGAGAGUagaaaagagaagcaggaaagga
                                                  CCIns
201 aagguaccgagggaaaucuuugguauuucaguuuucugauuuggugaagggaaaggccuugaaaaccuuuucguuuuccucaagagagaaguuggggaaa
301 gggugucaauucuuuccggauuucuucuggggagaggcuuugaaaaccuuucugcauuguuucuuuuuucucCCUCUCCAGGAAAGAUUUAUCGCUAUCA
401 GAUGAGCCCGCGUCAGAUUAGGUAGUUGGUGAGGUAAAGCUCACCAAGCCAAUGAUCUGUAGCUGGUCUGAGAGGAUGACCAGCCACACUGGGACUGAG
                  Eub338
501 ACACGGCCCAGACUCCUACGGGAGGCAGCAGUGGGGAAUAUUGGACAAUGGGGGCAACCCUGAUCCAGCGAUGCCGCGUGAAUGAAGAAGGCCUUAGGGU
601 UGUAAAGUUCUUUCACCCGUGACGAUAAUGACGGUAGCGGGAGAAGAAGCCCCGGCUAACUCCGUGCCAGCAGCCGCGGUAAGACGGAGGGGGCUAGCGU
701 UGUUCGGAAUGACUGGGCGUAAAGGGCGUGUAGGCUGUUGAUCAAGUCAAGCGUGAAAGCCUCGGGCUUAACCCGAGAACUGCGUUUGAUACUGGUUGAC
                    CC23a
801 UUGAGGACGAGAGAGGAAAGUGGAAUUCCGAGUGUAGAGGUGAAAUUCGUAGAUAUUCGGAGGAACAUCAGAGGCGAAGGCGGCUUUCUGGCUCGUAUCU
901 GACGCUGAGGCGCGAAAGCGUGGGGAGCAAACAGGAUUAGAUACCCUGGUAGUCCACGCUGUAAACGAUGAGUGCUAGACGUUGGGGGGUUCCCCUUCAG
1001 UGUCGCAGCUAACGCAUUAAGCACUCCGCCUGGGGAGUACGGUCGCAAGAUUAAAACUCAAAGGAAUUGACGGGGGCCCGCACAAGCGGUGGAGCAUGUG
1101 GUUUAAUUCGAAGCAACGCGCAGAACCUUACCAGCUCUUGACAUGGGGAUUAUGGAGAGGAGAGAUUCUUUCCUUCAGUUCGGCUGGAUCCCGCACAGGU
1201 GCUGCAUGGCUGUCGUCAGCUCGUGUCGUGAGAUGUUGGGUUAAGUCCCGCAACGAGCGCAACCCUCAUUUUAGUUGCCAUCAGGUUUGGCUGGGUACU
1301 CUAAAGAAACUGCCGGUGAUAAACCGGAGGAAGGUGGGGAUGACGUCAAGUCCUCAUGGCCCUUAUGGGCUGGGCUACACACGUGCUACAAUGGCGGUGA
1401 CAGAGAGGAGCAAAGGGGCGACCUGGAGCGAAUCUUUAAAAGCCGUCUCAGUUCGGAUUGCACUCUGCAACUCGAGUGCAUGAAGUUGGAAUCGCUAGUA
1501 AUCGCGGAUCAGCAUGCCGCGGUGAAUACGUUCACGGGCCUUGUACACACCGCCCGUCAUCACCAUGGGAGUUGAUUCAACCUUAAGGCGGUGAGCUAACC
1601 GAAAGGGAGCAGCCGACCACGGUGGGGUCAAUGACUGGGGUAAAGUCGUAACAAGGUAGCCGUAGGGGAACCUGCGGCUggaucaccucCuuunu
```

FIG. 1. Nucleotide sequence of the 16S rRNA gene of *C. caryophila*. The intervening sequence and the amplification primers are indicated by lowercase letters. The putative 5′ end of the insertion was not determined experimentally. The target sites of the hybridization probes are indicated by underlining.

included or excluded according to their degrees of conservation as determined from the complete data set or subsets from phylogenetic groupings. In all analyses *C. caryophila* could be assigned to the phylum of *Proteobacteria* (16) originally called purple bacteria (17). Within this phylum, *C. caryophila* belongs to the α subclass and is moderately related to *Holospora obtusa*, another endosymbiont found in the macronucleus of *Paramecium caudatum* (12). The overall 16S rRNA sequence similarity is 86% for the two species. The sequence data indicate a common origin of the two *Paramecium* endosymbionts and *Rickettsia* (Fig. 2).

The Intervening Sequence. Comparison with available homologous sequences from other bacteria revealed a stretch of an additional 194 bp which are unique within the *C. caryophila* 16S rRNA gene. The insertion is located within the part of the gene that encodes 16S rRNA regions involved in the formation of a helix (Figs. 3 and 4). The 5′- and 3′-terminal parts of the rRNA-like strand of the intervening sequence can be folded in a potential secondary structure that would extend the stem of a helix.

Southern and Northern hybridizations were performed to clarify whether the intervening sequence is present in the mature 16S rRNA. Almost complete 16S rRNAs from *C. caryophila*, as well as from *Pseudomonas diminuta* and *E. coli* as references, were amplified *in vitro* and used as targets in Southern hybridizations to the insertion-specific probe CCIns and to the probes 620R and Eub338 (11). The latter two probes are specific for conserved 16S rRNA regions adjacent to the 5′ and 3′ termini of the insertion. The target sites of the probes are marked in Fig. 1. The corresponding Northern hybridizations were carried out using crude RNA preparations from the same organisms. The insertion-specific probe hybridized only to the amplified 16S rDNA from *C. caryo-*

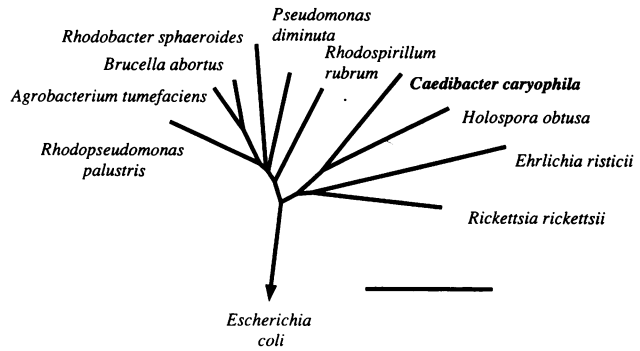

FIG. 2. Phylogenetic tree reflecting the relationships of *C. caryophila* and selected references. The tree was reconstructed by using a maximum-likelihood method as implemented in the fastDNAml program (14). Only positions that have the same composition in at least 50% of all available 16S rRNA sequences from the α class of *Proteobacteria* were included. The bar indicates 10% estimated sequence divergence.

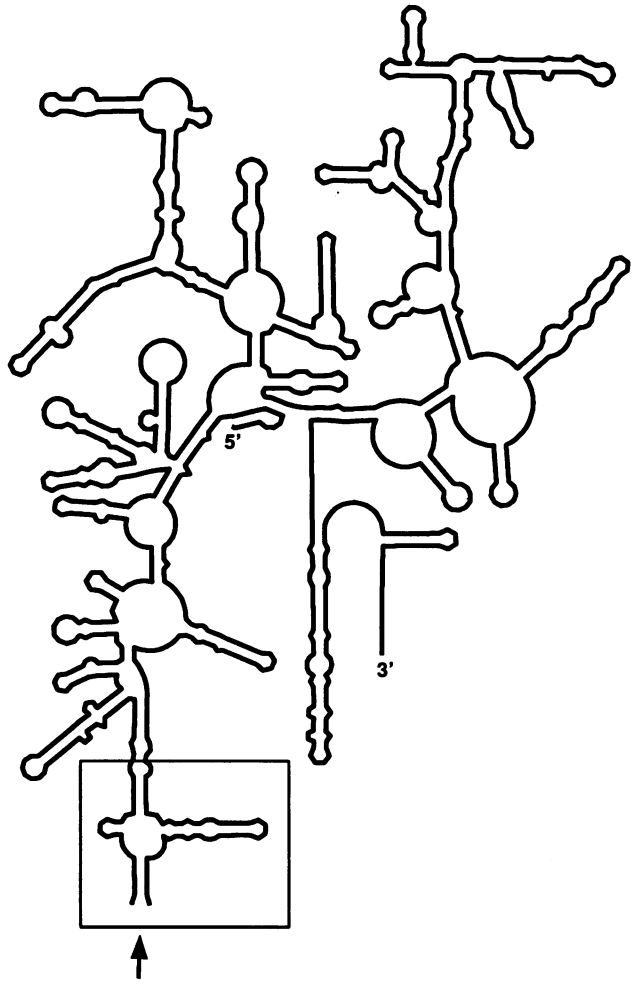

FIG. 3. Schematic secondary structure model based on the 16S rRNA sequence of *C. caryophila*. The region shown in Fig. 4 is boxed. Arrow indicates the position of the insertion.

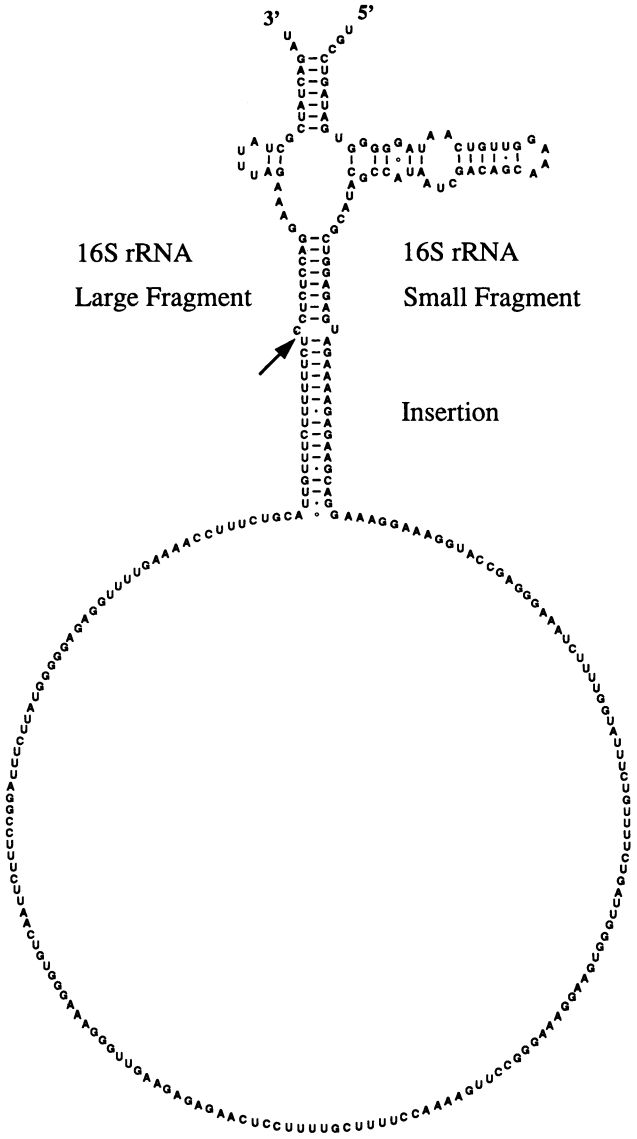

FIG. 4. Partial secondary structure of the predicted 16S rRNA precursor molecule. Arrow indicates the 5′ end of the large 16S rRNA fragment.

phila, but no hybridization could be detected with the RNA samples (Fig. 5 *d* and *h*). Probes 620R and Eub338 hybridized to all amplified rDNAs and to apparently intact 16S rRNAs of *Pseudomonas diminuta* and *E. coli* (Fig. 4 *b*, *c*, *f*, and *g*). However, in the case of *C. caryophila* RNA, probe 620R hybridized to a small rRNA fragment of about 180 bases, whereas probe Eub338 hybridizes to the large rRNA fragment of about 1320 bases (Fig. 4 *f* and *g*). These experiments prove that the 16S rRNA molecule is fragmented and that the intervening sequence is not present in mature RNA.

Cytosine-373 (Fig. 1) was determined as the 5′ terminus of the large 16S rRNA fragment by reverse transcriptase sequencing and is marked in Fig. 4.

Probe Design. A 16S rRNA-targeted specific hybridization probe for *C. caryophila* was designed after sequence comparison. The sequence of the probe CC23a (5′-TTCCACTTTCCTCTCTCG-3′) is complementary to a 16S rRNA region homologous to bases 658–675 of *E. coli* 16S rRNA.

***In Situ* Detection.** The results of whole-cell hybridizations are shown in Fig. 6. Infected paramecia were simultaneously treated with the tetramethylrhodamine-labeled *C. caryophila* specific probe CC23a and the fluorescein-labeled probe

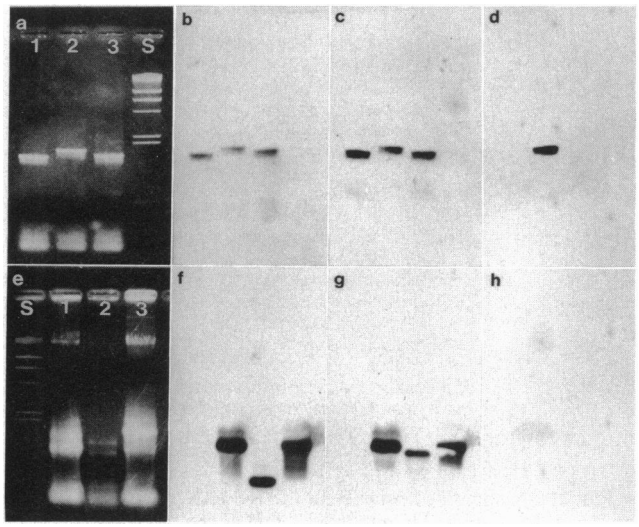

FIG. 5. Gel electrophoretic separation of *in vitro* amplified 16S rRNA genes (*a*) and crude rRNA preparations (*e*) from *Pseudomonas diminuta* (lanes 1), *C. caryophila* (lanes 2) and *E. coli* (lanes 3). Southern (*b–d*) and Northern (*f–h*) blots were hybridized to the 16S rRNA-specific probes 620R (*b* and *f*) and Eub338 (*c* and *g*) and to the insertion-specific probe CCIns (*d* and *h*). Lanes S, DNA molecular weight markers: *Hin*dIII-digested λ DNA.

Eub338 (11). The latter probe is complementary to a 16S rRNA region which is invariant in all bacteria analyzed so far. A phase-contrast micrograph of infected paramecia is shown in comparison with the corresponding epifluorescence micrographs. The bacterial probe Eub338 detected target sequences within the nuclei and food vacuoles, whereas the *C. caryophila*-specific probe CC23a reacted only with the bacteria present in the macronuclei. *In situ* hybridization with fluorescently labeled insertion-specific probe CCIns did not result in detectable signals.

DISCUSSION

The occurrence of bacterial endosymbionts in protozoa has been known for nearly a century. However, most of them cannot be grown in pure cultures so far. Therefore, the phylogenetic affiliations of these interesting organisms remained undetectable until very recently. Nowadays, the combined application of comparative sequence analysis (17) of *in vitro* amplified rRNA genes and whole-cell hybridization with (taxon-) specific probes (11, 18) allows phylogenetic analyses as well as *in situ* detection of uncultured bacteria. The moderate but distinct relationship of *C. caryophila* and *Holospora obtusa* is of special interest in that both endosymbionts share the same eukaryotic host. The two species are able to reproduce within the same host cell simultaneously but they can also inhabit host cells independently (19). The organisms are described as obligate endosymbionts, and nothing is known about persistence or potential reproduction outside the host cell. Therefore, the relatively deep branching of the *C. caryophila* and *Holospora obtusa* lineages may indicate that the endosymbiotic way of life evolved early in the history of these organisms. The answer to this interesting question has to await detection and phylogenetic analyses of additional endosymbiotic or nonendosymbiotic relatives. In this context, it is of further interest that the sequence data may indicate a common origin of the obligate cell-parasitic *Rickettsia* and the *Paramecium* endosymbionts. However, endosymbiotic behavior is not restricted to the α subclass of *Proteobacteria*. *Sarcobium lyticum*, an obligate intracellular parasite of small amoebae, phylogenetically has to be regarded as a *Legionella* species (5) belonging to the γ subclass.

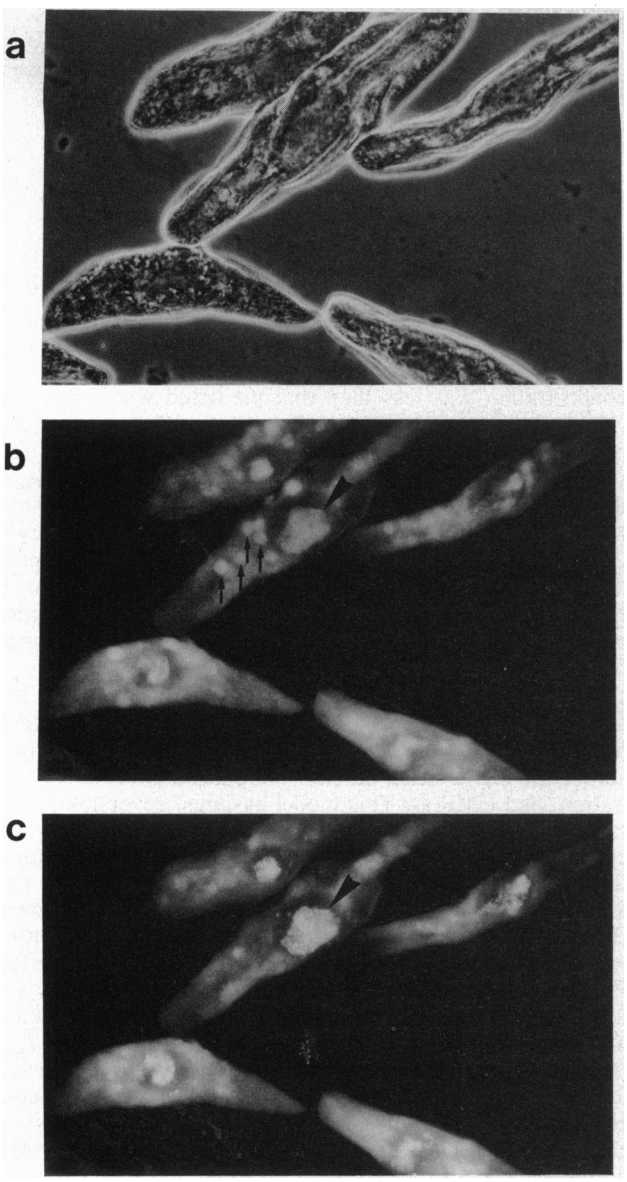

FIG. 6. *In situ* detection of *C. caryophila* within *Paramecium caudatum*. Identical microscopic fields (*a*, phase contrast) are shown after simultaneous hybridization of cells with the bacterial probe Eub338 labeled with fluorescein (*b*) and the *C. caryophila*-specific probe CC23a labeled with tetramethylrhodamine (*c*). Thick arrowhead, macronucleus; thin arrows, food vacuoles. (×290.)

Large stable and nonstable insertions with lengths of more than 100 bases have been described for bacterial 23S rDNAs (20–23). All known insertions are located within rather variable regions of the genes. The rRNA-like strands can be folded to form a potential secondary structure. In general, the insertions within proteobacterial 23S rDNAs known so far are removed during rRNA processing. The maturation processes have been analyzed for *Salmonella* (20). The processed rRNAs remain fragmented within the ribosome, and the intervening precursor rRNA parts are degraded. However, a characteristic insertion is maintained in mature 23S rRNAs of Gram-positive bacteria with a high DNA G+C content (22). Comparable large intervening sequences have not been described for bacterial 16S rRNAs. The 194-bp insertion found within the 16S rRNA genes of *C. caryophila* appears to be removed during processing of the precursor rRNA. The mature 16S rRNA is fragmented and the lengths of the fragments are as expected from the sequence data. The insertion fragment is apparently not only excised but also degraded, since it cannot be detected by Northern or *in situ* hybridization to the insertion-specific probe CCIns, whereas the shorter 5′ fragment is readily detected by the 16S rRNA-specific probe 620R (Fig. 5). The insertion within the predicted primary structure of the potential precursor rRNA can be folded into several alternative secondary structures (not shown). The 5′- and 3′-terminal parts can be arranged as an extension of a helix (Figs. 3 and 4). The extended helical element contains unpaired bases. The 5′ end of the longer 16S rRNA fragment was determined by reverse transcriptase sequencing and is located within this unpaired region. It has been shown by Burgin *et al.* (20) that RNase III is involved in the processing of the intervening sequences of *Salmonella* 23S rRNA precursors. RNase III substrate sites always occur in duplex stems (24). Therefore the extended helix 10 containing unpaired bases might be regarded as analogous to the RNase III processing sites described for the *Salmonella* 23S rRNA insertions.

Intervening sequences in bacterial 16S rRNAs which are removed during processing may occur more often than is known so far. There are no indications from the current data set of about 1300 bacterial 16S rRNA sequences, but many of these sequences lack terminal parts, and insertions may have been overlooked.

This work was supported by grants from Deutsche Forschungsgemeinschaft, the European Economic Community [BIOT-CT91-0294 (SSMA)], and Fonds der Chemie.

1. Ball, G. H. (1969) in *Research in Protozoology*, ed. Chen, T.-T. (Pergamon, New York), pp. 565–718.
2. Quackenbusch, R. L., Cox, B. J. & Kanabrocki, J. A. (1986) in *Extrachromosomal Elements in Lower Eukaryotes*, eds. Wickner, R. B., Hinnebusch, A., Lambowitz, A. M., Consalus, I. C. & Hollaender, A. (Plenum, New York), pp. 265–278.
3. Preer, J. R. & Preer, L. B. (1984) in *Bergey's Manual of Systematic Bacteriology*, ed. Krieg, N. R. (Williams & Wilkins, Baltimore), Vol. 1, pp. 795–942.
4. Schmidt, H. J., Görtz, H.-D. & Quackenbusch, R. L. (1987) *Int. J. Syst. Microbiol.* **37**, 459–462.
5. Springer, N., Ludwig, W., Drozanski, V., Amann, R. & Schleifer, K. H. (1992) *FEMS Microbiol. Lett.* **96**, 199–202.
6. Preer, J. R., Preer, L. B. & Jurand, A. (1974) *Bacteriol. Rev.* **38**, 113–163.
7. Schmidt, H. J., Görtz, H.-D., Pond, F. R. & Quackenbusch, R. L. (1987) *Cell Res.* **174**, 49–57.
8. Oelmüller, U., Krüger, N., Steinbühl, A. & Friedrich, C. G. (1990) *J. Microbiol. Methods* **11**, 73–78.
9. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. G. & Ehrlich, E. H. (1988) *Science* **99**, 487–491.
10. Brosius, J., Dull, T. J., Sleeter, D. D. & Noller, H. F. (1981) *J. Mol. Biol.* **148**, 107–127.
11. Amann, R., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R. & Stahl, D. (1990) *Appl. Environ. Microbiol.* **56**, 1919–1925.
12. Amann, R., Springer, N., Ludwig, W., Görtz, H. D. & Schleifer, K. H. (1991) *Nature (London)* **351**, 161–164.
13. De Rijk, P., Neefs, J. M., Van de Peer, Y. & De Wachter, R. (1992) *Nucleic Acids Res.* **20**, 2075–2089.
14. Olsen, G. J., Overbeek, R., Larsen, N., Marsh, T. L., McCaughey, M. J., Maciukenas, M. A., Kuan, W. M., Macke, T. J., Xing, Y. & Woese, C. R. (1992) *Nucleic Acids Res.* **20**, Suppl., 2199–2200.
15. Felsenstein, J. (1982) *Quart. Rev. Biol.* **57**, 379–404.
16. Stackebrandt, E., Murray, R. G. E. & Trüper, H. G. (1988) *Int. J. Syst. Bacteriol.* **38**, 321–325.
17. Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221–271.
18. Delong, E. F., Wickham, G. S. & Pace, N. R. (1989) *Science* **143**, 1360–1363.
19. Preer, L. B. (1969) *J. Protozool.* **16**, 570–578.
20. Burgin, A. B., Parodos, K., Lane, D. J. & Pace, N. R. (1990) *Cell* **60**, 405–414.
21. Dryden, S. C. & Kaplan, S. (1990) *Nucleic Acids Res.* **18**, 7267–7277.
22. Roller, C., Ludwig, W. & Schleifer, K. H. (1992) *J. Gen. Microbiol.* **138**, 1167–1175.
23. Skurnik, M. & Toivanen, P. (1991) *Mol. Microbiol.* **5**, 585–593.
24. Robertson, H. D. (1982) *Cell* **30**, 669–672.