Characterization of the 23S and 5S rRNA Genes of *Coxiella burnetii* and Identification of an Intervening Sequence within the 23S rRNA Gene

GUY AFSETH,[†] YIN-YUAN MO,[‡] AND LOUIS P. MALLAVIA*

Department of Microbiology, Washington State University, Pullman, Washington 99164-4233

Received 6 September 1994/Accepted 10 March 1995

Characterization of the rRNA operon from the obligate intracellular bacterium *Coxiella burnetii* has determined the order of the rRNA genes to be 16S-23S-5S. A 444-bp intervening sequence (IVS) was identified to interrupt the 23S rRNA gene beginning at position 1176. The IVS is predicted to form a stem-loop structure formed by flanking inverted repeats, and the absence of intact 23S rRNA molecules suggests that the loop is removed. An open reading frame in the IVS has been identified that shows 70% similarity at the amino acid level to IVS open reading frames characterized from four species of *Leptospira*.

Generally, rRNA in prokaryotes is highly conserved in terms of both size and structure, with the rRNA operon usually consisting of 23S, 16S, and 5S components with genes of approximately 2,900, 1,600, and 120 bp in length, respectively (10). In eubacteria such as *Yersinia enterocolitica* and *Salmonella typhimurium*, intervening sequences (IVSs) of 90 to 100 bp in length have been observed to interrupt the 23S rRNA gene (2, 15). These IVSs, however, contain no open reading frames (ORFs). In contrast, for selected species of the genus *Leptospira*, the 23S rRNA gene is interrupted by an IVS which varies in size from 485 to 759 bp and contains an ORF coding for polypeptides that range from 111 to 133 amino acids in length. In both examples, the IVS forms a stem-loop structure which is removed without splicing from the mature product, resulting in a fragmented 23S rRNA molecule. The IVSs all have characteristics of a highly mobile genetic element, with inverted repeats flanking the insertion element sequence (12, 15). Interestingly, the IVSs are not found in all species or strains of these bacteria and also are not always in all copies of the rRNA operons within the same bacterium (2, 15). As a result of the sporadic appearance of IVSs coupled with their total absence in selected bacteria, a functional role for IVSs has yet to emerge.

Coxiella burnetii is an obligate intracellular bacterium which grows in the phagolysosomal vacuole of infected cells and is the etiologic agent of Q fever (1). During our characterization of the rRNA operon of *C. burnetii*, we have sequenced the 23S and 5S rRNA genes and identified an IVS within the 23S rRNA gene. This is the first identification of an IVS in the rRNA gene of an obligate intracellular bacterium.

The nucleotide sequences of both strands of the 23S and 5S rRNA genes were obtained from two separate plasmids to obtain the complete 23S rRNA gene. The first plasmid was pYM38, which was isolated from a *C. burnetii* lambda-ZAP II gene bank (provided by Jim Samuel, Texas A&M, College

Station). pYM38 is a 9.0-kb clone containing all but the first 160 bp of the complete 23S rRNA gene plus the 5S rRNA gene sequence and the 23S-5S intergenic spacer. To obtain the sequence for the 5' end of the 23S rRNA gene, a second plasmid was constructed by PCR by using a GeneAmp PCR kit (Perkin-Elmer, Norwalk, Conn.). The forward primer was Q16F (5' AGGCTAACCTTCGGGAG 3'), while the reverse primer was 23SR11 (5' TATGGCTGAACCTTCCA 3'). For the PCR, the times and temperatures for denaturing, annealing, and polymerizing were 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min, respectively. *C. burnetii* chromosomal DNA served as the template and was isolated by a method previously described (11). The amplified product, an 850-bp fragment, was purified from a 2% agarose gel by using the Gene Clean kit (Bio 101, La Jolla, Calif.), cloned into a TA cloning vector (Stratagene, La Jolla, Calif.), and transformed into an *Escherichia coli* DH5 α host (Bethesda Research Laboratories, Gaithersburg, Md.). The resulting plasmid was designated pGA100, and partial sequencing confirmed that it contained the forward portion of the 23S rRNA gene. The sequence of both pGA100 and pYM38 was obtained by using the dideoxy chain termination method (14) with Sequitide (New England Nuclear, Boston, Mass.) and Sequenase (United States Biochemical, Cleveland, Ohio). The plasmids were primed by using either pUC forward or reverse primers or custom-made sequencing primers (17-mers) prepared by DNA International (Lake Oswego, Oreg.) or Ransom Hill Bio-Science Inc. (Ramona, Calif.).

The complete sequence of the 23S and 5S rRNA genes, along with the 23S-5S spacer region, is shown in Fig. 1. The 23S rRNA gene is 3,329 bp long, including a 444-bp IVS, while the 5S rRNA gene is 117 bp long. The two rRNA genes are separated by a 44-bp spacer region, and both the 23S and 5S rRNA genes have approximately 80 and 78% identity to the rRNA genes of *E. coli*, respectively. The IVS in the 23S rRNA gene contains an ORF capable of coding for 106 amino acids. A comparison of the four IVSs from *Leptospira* species with the one seen in *C. burnetii* reveals 68 to 70% similarity at the deduced amino acid level, but not at the DNA sequence level (data not shown). This conservation suggests that maintenance of the amino acid sequence is under evolutionary control. If it were not, the amino acid sequence would be predicted to diverge at least twice as fast as the DNA sequence (12). Since these conserved ORFs would be part of the 23S primary tran-

* Corresponding author. Mailing address: Department of Microbiology, Washington State University, Pullman, WA 99164-4233. Phone: (509) 335-3323. Fax: (509) 335-3517. Electronic mail address: malla via@wsu.edu.

[†] Present address: The Perkin-Elmer Corporation, Applied Biosystems Division, Norwalk, CT 06859-0251.

[‡] Present address: Department of Tumor Cell Biology, St. Jude Children's Research Hospital, Memphis, TN 38101.

1-----> Start of 23S	
1	TGGAATCATTTAGGGTTGCATGGTCAAGTGAATAAGCGCATATGGTGGATGCCTTGGCAGTAAGAGGCGATGAAGGACGTAGTAGCCTGCGATAACTTCG 100
101	GGGAGCTGGCAAACAAGCGATGATCCGGACATTTCCGAATGGGGAACCAACCTTTCGAGGTTATGTATACTGAATTCATAGGTATACGAGGCGAACCT 200
201	GGGGAACTGAAACATCTAAGTACCCAGAGGAAAAGAAATCAACCAGATTCCGTAGTAGCGGCGAGCGAAAGAACAGCCAGTTACTAAATCATTATTT 300
301	GTTCTAGCAGAATGTTCTGGAAGTTTCAGCCATAGCGGGTGATAGCCGGTACCCGAAAGAGTAAATAATGTGGGTAACGATGAGTAGGTCGGGACACGT 400
401	GGTATCTTGACTGAACATGGGGGACCATCTCCAAGGTAAATACTCTTACTGACCGATAGCGAACCGTAGCCGTGAGGAAAGGTGAAAAAGAACCCC 500
501	GGCGAGGGAGTGAATAGAACCTGAAACCGTATGCGTACAAGCTGTAGGAGCATTTCTCGGAAATGTGACTGCGTACCTTTTGATAATGGGTACGCGA 600
601	CTTACTTGTTGTAGCGAGCTTAACCGTCTAGGGGAGCGTAGGGAACCGAGTCCGAAATGGCGTTTGTAGTTGCAACGAGTAGACCGAACCCGAAGCGAT 700
701	CTATCTATGGCAGGGTGAAGGTACAGTAACACTGACTGGAGGCCGAACCACTAATGTGAAAAAGTTAGGGGATGAGCTGTGGATAGGAGTGAAAGGCTA 800
801	ATCAAGCTCGGAGATAGCTGTTCTCTCGAAAGCTATTTAGGTAGCGCTCGTGTATGACTCTTGGGGTAGAGCACTGTTTCGGCTAGGGGGCCATCC 900
901	CGGCCACCAAACCGAGGCAAACTCCGAATACCAAGAAGTTTTAGCAGCGGAGACACTGCTGGGTGATAAGGTCCGTGGTGAAAGGAACGCCAGATC 1000
1001	GCCAGATCGCAGCTAAGGTCAAATACACGTAGTGAACAGTGTGGGAAGGCTCAGACAGCCAGACCTTGGCTTAGAAGCAGCCATCCTTTAAGAAA 1100
1101	GCCTAATAGCTCACTGGTCGAGTCGCTCGCGGAAGATTTAACGGGGCTCAAATGTGTACCGAAGCTGCGG CATCAGAAGACAGATGACAGATGACA 1200
1201	GAAATATGAAAAAGAAATATCTAGCTTTGAAGATTTAGTGGTTTTTCAGAAAGCGTATAAGCTTTCTTTAAAGCTTCATAAACTAACTTTAAGTATGCC 1300
1	M K K E I S S F E D L V V F Q K A Y K L S L K L H K L T L S M P 33
1301	TAAAGAAGAGCAATATGGATTGGCAAGCCAAATCCGTCAAGCAAGTAAATCAATTTGCGCGAATATTGCTGAAGGGTTTGGAAAGCAGAGTGTTCACG 1400
34	K E E Q Y G L A S Q I R Q A S K S I C A N I A E G F G K Q S V S T 66
1401	GCTGAATTTAAGCGATTATTTTAAATGCAATGGGTTCCGCGAGATGAGTGCCTGTGTGGTTACGATATTGTTTGATTTAGGTTATTTGAATGAAAAA 1500
67	A E F K R F I L I A M G S A D E M R V W L R Y C F D L G Y L N E K T 99
1501	CCTGGGTAGGACTTCCGCAATACTATCAGGATATTGCGAAATCGTAAAGGGTTTGCACAAAGCTGGAGTTAATTTTTATCTTCTGCTCAGTCTTC 1600
101	W V G L P Q * 133
1601	TGCCATCTGTCTTCTGAT GGGTAGAGGAGCGTTCTGTAACCTGTGAAGGTATCGAGAGGTTTCTGGAGGTATCAGAAGTGCGAATGCTGACATAAAGT 1700
1701	AACGATAATGTGGGTGAAAAACCCACACGCCGAAAGTCTAAGGTTTCTCGCCGTTAATCCGAGGGTGAGTCGGCCCTAAGGGAGGAGAAATCGTAGT 1800
1801	CGATGGGAACCGGGTTAATATTTCCCGTACTTTATAATACTCGATGGGGGGGACGGAGAAGGCTAGGTCAGCCCCCGATGGTTGTGGTGGGCTAGTCA 1900
1901	GCCACCGATGTGTGTGTGTTAAGTGTGTGTAGAAAGGATCTTCAAATCCGGAACCTCAATTCGAGACATGATGACGAGTACGAACTGTTCTGTGCAAAAG 2000
2001	TGATTGATGCCACGACTTCCAGGAAAAGTCCCTAAGCTTCAGGTATTGTAAACCGTACTATAAACCGACACAGGTGGACAGGTAGAGAATACCAAGGCG 2100
2101	CTTGAGAGAACTTGGGTGAAGGAACTAGGCAAAATGGCACCGTAACCTCGGGAGAAGGTGCGCCCTTGTAAGTGAAGTCCCTTGCAGTCTGAGCTGAAAG 2200
2201	GTTGAGATACCAAGGTGGCTGCGAAAGACTAGTTTACTAAAAACAGCAACTCTGCAAACTCGTAAGAGAGCGTAATAGGGTGTGAGCGCTGCCCGGTGC 2300
2301	CGGAAAGGTTAAGTGTGGGTTAGCCCTCGGGCGAAGCTCTTGATCGAAGCCCCGGTAAACGGGCGGCCGTAACATAAAATTAACCGTTGTAGTTACGT 2400
2401	AAATAAACTTGGCTATATGTGGAACCTCCGAGTATCTCCAGTACTCATCTCAACTAAAATTGAGGTAGTAAAAATCTGCGAGGTGACAGTACCCGCGG 2500
2501	CTAGACGGAAGACCCCGTGAACCTTTACTACAGCTTTACACTGAACCTTTGAATATGCCTGTGTAGGATAGGTGGGAGGCTTTGAAGCTGGAGCGTAGC 2600
2601	TCCAGTGGAGCCAACCTTGAATACCAACCTGGCGTGTGAGGTTCTAACCTGTCCCGTAATCCGGGTGCGGAACAGTGTATGGTGGGTAGTTTGACT 2700
2701	GGGGCGGTCTCTCTAAAGAGTAACGGAGGAGTACAAAGTACCCTCAGCACGGTCGGAATCGTGATTGTGTGCAAGGCATAAGGTGCTTGACTGC 2800
2801	GAGACGTACAAGTCGAGCAGGTACGAAAGTAGGTCTTAGTGATCCGGTGTCTTTATGGAAGGGCACGCTCAACGGATAAAAGGACTCCGGGATAACA 2900
2901	GGCTGATACCAACCAAGAGTTTATATCGACGGTGGTGTGGCACCTCGATGTGCGCTCATCACATCCTGGGGCTGTAGCCGGTCCCAAGGGTATGGCTG 3000
3001	TTCGCCATTTAAAGTGTACGTGAGCTGGGTTTGAACGTGTGAGACAGTTCGGTCCCTATCTGCCGTTGGGCGTTGGAGATTGAGAGGAGCTGCTCC 3100
3101	TAGTACGAGAGACCGAGTGGACGTACCTCTGGTGTCCGGTTGTACGCCAGTGGCATTTCGCCGGTAGCTAAGTACGGACGGGATAACCGCTGAAAG 3200
3201	CATCTAAGCGGGAAACCCCTCAAGATGAGATCTCCGGACCTTTAAGGTCTAAAGATTCTGTTGAAGACGACAAGTTGATAGGCAGGGTGTGGAAGTCT 3300
End of 23S <----	
3301	CAGTAATGAGTGAAGCTAACCTGTACTAATTAATCGTGCAGTCTGACTATGTAAACCTAAATGGTTTCTATTGCCTTTACTAATTTTACCGGTTTCTG 3400
----> Start of 5S	
3401	GCGGCATAGCGAGCGGGCCCCACCTGCTTCATTCCGAACTCAGAAGTGAAACGCTTAAGGCCCGATGATAGTGTGGGTCTCCATGCGAAAGTAGGT 3500
3501	CACCGCCAGTTT 3512
End of 5S<-	

FIG. 1. Nucleotide sequence of the 23S and 5S rRNA genes of *C. burnetii*. The beginnings and ends of the 23S and 5S genes are denoted by a single line and a dashed arrow above the sequence. The IVS is underlined, with the translation of the ORF appearing below. The IVS has at each end inverted repeats which form a double-helix stem 28 nucleotides in length, while the remainder of the sequence forms no stable structure. The inverted repeats appear in boldface type.

script, they would be abundant in the cell. While a purine-rich region precedes the start of the ORF, no conserved Shine-Dalgarno sequence is observed; thus, the efficiency of translation of the predicted polypeptide may be low.

The discovery of conserved ORFs in the IVSs of such evolutionarily distant bacteria as *C. burnetii* and *Leptospira* spp. (17, 18) raises additional questions as to their origin and function. In other bacteria, the IVSs identified were seemingly random pieces of DNA which were excised from the mature rRNA (2, 15). Conservation of the ORF's putative amino acid sequence suggests that it may function as something other than a spacer between the inverted repeats of the IVS. While the IVS may code for a protein which is required for initial insertion of the IVS, such as a transposase, no homology was found between these putative amino acid sequences and any known proteins. In addition, how two distinct bacteria such as *Coxiella* and *Leptospira* spp. came to possess such similar IVSs is not known. It is possible that because of separate unknown evolutionary constraints, insertion events have occurred with maintenance of the IVSs in some bacteria. However, it is more likely that lateral transfer of the IVS occurred between *Coxiella* and *Leptospira* spp. some time after their evolutionary divergence. *Leptospira* spp. have been shown to be capable of horizontal transfer of IVSs (13). It is unlikely that ORFs of clear common ancestry would occur in such divergent species as *Coxiella* and *Leptospira* spp. In addition, the intragenus variation in the amino acid sequence is as much as two-thirds of the interdivision difference, suggesting that the element is mobile.

In addition to the appearance of an ORF, a large inverted repeat was identified to flank the *C. burnetii* IVS. The inverted repeat seen in *C. burnetii* is 28 nucleotides long and may form a stable double-helix structure containing only three mismatches (Fig. 1). The remainder of the IVS is predicted to form no stable secondary structure and likely forms a loop of 388 nucleotides in length. In previously detected IVSs of *Salmonella* (2), *Yersinia* (15), and *Leptospira* (12) spp., extensive inverted repeats flanking the sequence of the insertion element at the boundaries of the 23S sequence were also observed. In *Salmonella* (2) and *Leptospira* (12) spp., RNase III was shown to remove the IVS from the mature 23S rRNA by cutting at both ends of the terminal repeats without religation of the 23S rRNA ends. The double helix formed by the inverted repeats remained intact, but since the RNA was not spliced, it fragmented upon denaturation (2, 7, 8, 12). If the 23S rRNA molecule is fragmented in *C. burnetii* this would indicate that the IVS is removed from the mature product. To examine total bacterial nucleic acids, 1.3×10^9 and 2.6×10^9 *C. burnetii* cells were resuspended in 50 μ l of phosphate-buffered saline (PBS) each (5). For *E. coli*, an amount equivalent to 1.0 unit of optical density at 540 nm was resuspended in 50 μ l of PBS. An equal volume of 10% sodium dodecyl sulfate was added to each bacterial suspension, and the mixtures were incubated at 65°C for 1 h. Nucleic acids were separated by electrophoresis in a 1.5% agarose gel (3 h at 100 V) in 1 $\times$ TBE (11). Examination of the RNA from *C. burnetii* Nine Mile reveals no intact 23S rRNA molecule (Fig. 2). A comparison of the rRNA molecules of *C. burnetii* and *E. coli* shows that both contain an intact 16S rRNA molecule, with *C. burnetii*'s 16S rRNA appearing slightly larger than *E. coli*'s. *E. coli* displays a single band representing an intact 23S rRNA molecule. *C. burnetii* has no corresponding 23S band but instead has two bands migrating near the region of 16S rRNA. Taken together, the sizes of the two small molecules (approximately 1.7 and 1.2 kb) correspond to the length of an intact *E. coli* 23S rRNA molecule (2.9 kb). When the *C. burnetii* 23S rRNA gene was radiolabeled and used as a probe,

FIG. 2. Examination of rRNAs from *C. burnetii* and *E. coli*. The positions of *E. coli* 16S and 23S rRNA bands are labeled on the left. The positions of *C. burnetii* 16S and the fragmented 23S rRNA molecules are labeled on the right. Lane a, *E. coli* DH5 α ; lane b, 1.3×10^9 *C. burnetii* Nine Mile RSA 493 cells; lane c, 2.6×10^9 *C. burnetii* Nine Mile RSA 493 cells.

it hybridized to the ~1.7- and 1.2-kb molecules, confirming that they are 23S rRNA sequences (data not shown).

A number of fragmented 23S rRNAs have been identified in a broad range of eubacteria such as *Rhodobacter*, *Anacystis*, *Micrococcus*, *Brucella*, and *Agrobacterium* spp. although whether they are due to excision of IVSs has yet to be determined for all cases (12). The mechanism by which these sequences have inserted into the rRNA genes of such a broad range of bacteria is not known. It is possible that the IVSs are remnants of errors made during replication that were never fully removed, or perhaps they are the footprints left by the insertion and subsequent excision of a transposable element (2). Such an event often leaves inverted repeats at the flanks of the sequence which is characteristic of the IVSs identified to date.

Of interest is not only how these IVSs were introduced, but also why they persist in some bacteria and not others. One suggestion is that these IVSs may offer some selective advantage to the bacteria (15). Since cleavage sites seem to occur only in the 23S rRNA and not in the 16S rRNA, fragmentation of the 23S rRNA molecule must offer some flexibility which does not exist for the 16S rRNA molecule (8).

Finally, previous phylogenetic analysis of *C. burnetii* based on sequence alignments of the 16S gene placed it in the gamma subdivision of the purple bacteria (16, 17). By using sequence alignment programs (3, 4, 6, 9) with the *C. burnetii* 23S sequence, the insertion of an IVS was not found to affect its placement in the gamma subdivision of the purple bacteria.

Nucleotide sequence accession number. The DNA sequences of the 23S and 5S portions of the *C. burnetii* rRNA operon are available in GenBank under accession number X79704.

We thank Steve Thompson and Shirley Schmitt for assistance with the figures and phylogenetic studies. The software cited is part of the VADMS Center, a campus-wide computer resource at Washington State University. We thank Michael Konkel and Mike Minnick for supplying experimental procedures and Matt Nilles for critical reading of the manuscript.

This work was supported by grant AI20190 from the National Institutes of Health (NIAID).

REFERENCES

1. Baca, O. G., and D. Paretsky. 1983. Q fever and *Coxiella burnetii*: a model for host-parasite interactions. *Microbiol. Rev.* **47**:127-149.
2. Burgin, A. B., K. Parodos, D. J. Lane, and N. R. Pace. 1990. The excision of intervening sequences from *Salmonella* 23S ribosomal RNA. *Cell* **60**:405-414.
3. Felsenstein, J. 1993. PHYLIP-Phylogeny inference package, version 3.5c. University of Washington, Seattle.
4. Fitch, V. M., and E. Margoliash. 1967. Construction of phylogenetic trees: a method based on mutational distances as estimated from cytochrome c

- sequences is of general applicability. *Science* **155**:279–284.
5. **Hendrix, L., and L. P. Mallavia.** 1984. Active transport of proline by *Coxiella burnetii*. *J. Gen. Microbiol.* **130**:2857–2863.
 6. **Higgins, D. G., and P. M. Sharp.** 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Comput. Appl. Biosci.* **5**:151–153.
 7. **Hsu, D., M. Pan, Y. C. Zee, and R. B. LeFebvre.** 1990. Unique ribosome structure of *Leptospira interrogans* is composed of four rRNA components. *J. Bacteriol.* **172**:3478–3480.
 8. **Hsu, D., Y. C. Zee, J. Ingraham, and L. Shih.** 1992. Diversity of cleavage patterns of *Salmonella* 23S rRNA. *J. Gen. Microbiol.* **138**:199–203.
 9. **Kimura, M.** 1980. A simple model for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequence. *J. Mol. Evol.* **16**:111–120.
 10. **King, T. C., R. Sirdeskumukh, and D. Schlessinger.** 1986. Nucleolytic processing of ribonucleic acid transcripts in procaryotes. *Microbiol. Rev.* **50**:428–451.
 11. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 12. **Ralph, D., and M. McClelland.** 1993. Intervening sequence with conserved open reading frame in eubacterial 23S rRNA genes. *Proc. Natl. Acad. Sci. USA* **90**:6864–6868.
 13. **Ralph, D., and M. McClelland.** 1994. Phylogenetic evidence for horizontal transfer of an intervening sequence between species in a spirochete genus. *J. Bacteriol.* **176**:5982–5987.
 14. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
 15. **Skurnik, M., and P. Toivanen.** 1991. Intervening sequences (IVSs) in the 23S ribosomal RNA genes of pathogenic *Yersinia enterocolitica* strains. The IVSs in *Y. enterocolitica* and *Salmonella typhimurium* have a common origin. *Mol. Microbiol.* **5**:585–593.
 16. **Stein, A., N. A. Saunders, A. G. Taylor, and D. Rauolt.** 1993. Phylogenetic homogeneity of *Coxiella burnetii* strains as determined by 16S ribosomal RNA sequencing. *FEMS Microbiol. Lett.* **113**:339–344.
 17. **Weisburg, W. G., M. E. Dobson, J. E. Samuel, G. A. Dasch, L. P. Mallavia, O. G. Baca, L. Mandelco, J. E. Sechrest, E. Weiss, and C. R. Woese.** 1989. Phylogenetic diversity of the rickettsiae. *J. Bacteriol.* **171**:4202–4206.
 18. **Woese, C. R.** 1987. Bacterial evolution. *Microbiol. Rev.* **51**:221–271.