An Unusual Gene Containing a \textit{dnaJ} N-Terminal Box Flanks the Putative Origin of Replication of \textit{Mycoplasma genitalium}

CAMELLA C. BAILEY\textsuperscript{1} AND KENNETH F. BOTT\textsuperscript{1,2,*}

Department of Microbiology and Immunology\textsuperscript{1} and Curriculum of Genetics,\textsuperscript{2} University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Received 1 April 1994/Accepted 12 July 1994

Origins of replication are known to be highly conserved among widely divergent microbial species, with the gene order in those regions being \textit{dnaA-dnaN-recF-gyrB}. On the basis of sequence identities to entries in GenBank, the gene order of a 6-kb fragment of \textit{Mycoplasma genitalium} DNA was determined to be \textit{dnaN-orf311-gyrB-gyrA-serS}, which is structurally similar to the ancestral origin of replication. We have directly linked the \textit{dnaN} gene to the \textit{M. genitalium} \textit{dnaA} gene by PCR amplification. However, we found a novel open reading frame, designated \textit{orf311}, in place of an expected sequence encoding \textit{recF}. \textit{Orf311} contains a \textit{dnaJ} box motif at its N terminus, but it has no overall homology to any other protein or sequence in the database. We are unable to detect any \textit{recF} homolog in \textit{M. genitalium} by hybridization or during a random sequencing survey of the genome.

Mycoplasmas are significant pathogens of humans, animals, and plants. They are believed to be the smallest free-living organisms and are thought to have evolved from higher gram-positive bacteria through a loss of genetic material (27). We are characterizing \textit{Mycoplasma genitalium}, which contains a genome of less than 600 kbp, as a model for an organism containing only those genes which are essential for life (25).

Since no genetic system for this species exists and clonal selection of individual cells is difficult, a random sequencing approach was undertaken to survey the \textit{M. genitalium} genome (25). During this study, a collection of 291 independent clones was generated and partially sequenced. This examination of the chromosome allowed the identification of potential metabolic and biosynthetic pathways of \textit{M. genitalium} and provided the tools necessary to study genome organization.

A clone which showed sequence homology to the \textit{Escherichia coli} \textit{gyrA} gene was identified during this study. This clone was chosen for further characterization because of the relation of \textit{gyrA} genes to origins of replication in most bacterial genomes. On the basis of our own studies showing the linkage of the \textit{gyrA} and \textit{gyrB} genes in \textit{Mycoplasma pneumoniae} (7), we speculated that these genes would also be linked in \textit{M. genitalium} and that the replication origin of this organism might be similar to those of gram-positive species.

Comparison of chromosomal origins of DNA replication among the most common eubacteria (enteric bacteria, pseudomonads, and bacilli) has allowed the determination of several highly conserved structural features of this region (22, 30). On the basis of this comparison, Ogawara and Yoshikawa defined an ancestral origin lying near the \textit{dnaA} gene (23). Extended analyses of other species have shown them to fit into the same general pattern (23). The ancestral origin is not only conserved structurally, but the genes localized there are themselves highly conserved, and most of them encode essential proteins which are involved in DNA or RNA metabolism. The conserved gene order of the right side of this region is \textit{dnaA-dnaN-recF-gyrB} with the actual point of replication origin lying among \textit{dnaA} boxes directly to the left of the \textit{dnaA} gene. A comparison of the \textit{Bacillus subtilis} origin region with that of \textit{Pseudomonas putida} revealed that the order of more than 12 genes is conserved between the two species (23). The \textit{E. coli} origin region, the most extensively characterized of all such regions, also fits the pattern if one compensates for an inversion which places its actual point of origin (\textit{oriC}) away from the defined ancestral origin region (22). In gram-positive species the \textit{gyrA} gene is linked to the \textit{gyrB} gene; in gram-negative species the \textit{gyrA} gene is not usually associated with the origin region. In all species, the \textit{dnaA} gene encodes a protein which is involved in initiation of DNA replication; \textit{dnaN} encodes the \textit{B} subunit of DNA polymerase III. The exact function of the \textit{recF} gene product is not clear, but it is thought to be involved at some step in recombination and in DNA repair (1, 18). The \textit{gyrB} and \textit{gyrA} genes encode the subunits of DNA gyrase, an essential enzyme involved in putting negative supercoils into DNA (28).

Recently, several different investigators have shown that some species do not follow the conventional ancestral gene order in their respective origin regions. Zweiger and Shapiro have demonstrated that the origin region of \textit{Caulobacter crescentus} and its \textit{dnaA} gene are removed from the linked \textit{dnaN-recF-gyrB} genes (37). There appears to be a DNA insert containing a \textit{dnaJ} operon and a \textit{parA} gene in this region (37). A second departure from the ancestral gene order is seen in \textit{Mycoplasma capricolum}, for which Miyata et al. show that the gene order to the right of \textit{dnaA} is \textit{dnaN-orfL3-ksgA} (19).

Here we report the chromosomal organization and partial nucleotide sequence of the putative \textit{M. genitalium} origin and show it to have a gene order, on the basis of DNA database homologies, of \textit{dnaN-orf311-gyrB-gyrA-serS}. By using the PCR, we were able to show linkage between the \textit{dnaA} gene and the \textit{dnaN} gene along with its associated gene cluster. We discuss the unexpected presence of \textit{orf311}, an open reading frame (ORF) containing a \textit{dnaJ} box motif which is substituted within this highly conserved region, and consider the necessity of modifying the classic origin structure.

**MATERIALS AND METHODS**

**Clones and sequencing.** Gyrase clone Contig 34 contains a 5-kb EcoRI fragment of \textit{M. genitalium} in the vector pUC118 (26). An overlapping chromosomal \textit{BamHI} fragment was
isolated from genomic digests and identified by Southern hybridization (33) and DNA sequence identity to Contig 34. This 2.5-kb BamHI fragment was ligated into BamHI-digested pGEM Blue vector (Promega, Madison, Wis.), and the resulting plasmid was named pCBI. A third overlapping fragment, a 9.0-kb EcoRI chromosomal fragment cloned into pBlueScript (Stratagene, La Jolla, Calif.) as pF3-52, was identified by colony blot hybridization using the 0.5-kb EcoRI-BamHI end of pCBI as a probe. A collection of deletion clones of the original Contig 34, made in both orientations, was generated by the mini Tn3 mutagenesis method (9). DNA isolation was performed by using Wizard Magic Prep according to the manufacturer's instructions (Promega). Sequencing was performed with Sequenase (U.S. Biochemical Corp., Cleveland, Ohio) by the dideoxynucleotide method (29) with an oligonucleotide primer designed for sequences in the β-lactamase gene of the mini Tn3 transposon (9). Contig clones were ligated and extended onto overlapping clones by directed sequencing using a series of synthesized oligonucleotide primers. Sequence in both directions was generated from overlapping clones. Computer analysis of the data was facilitated by the use of the Wisconsin Genetics Computer Group computer program package (version 7.0; Genetics Computer Group, Inc., Madison, Wis.) running on a UNIVAC System (10).

DNA hybridizations. Genomic M. genitalium DNA was digested with various restriction enzymes, electrophoresed in 0.8% agarose, and prepared for Southern hybridization by a modification of the procedure of Smith and Summers (31). Agarose gels were exposed to short-wave UV for 10 min on a UV box (Fotodyne Inc., Hartland, Wis.), treated with 1.5 M NaCl-0.5 M NaOH for 30 min to denature double-stranded DNA, and neutralized in 1.5 M CHCOONa (ammonium acetate)-20 mM NaOH for 60 min. The DNA was transferred to Magnagraph membranes (Schleicher & Schuell, Keene, N.H.) by capillary transfer in neutralization solution. DNA was UV cross-linked to the membranes in a Stratalinker 2400 (Stratagene) according to the manufacturer's instructions. Purified inserts from plasmid DNA for use as probes were released from the vector by using the appropriate restriction enzymes (Boehringer Mannheim Corp., Indianapolis, Ind.), electrophoresed in 1% agarose in 1× TAE (0.04 M Tris acetate, 0.002 M EDTA) gel purified by using GENECLEAN (Bio 101, La Jolla, Calif.), digested with Sau3A to generate small fragments, and labeled enzymatically by using the GENIUS random priming kit according to the manufacturer's instructions (Boehringer Mannheim). Hybridizations were performed at 65°C; and the following washes were performed: two room temperature washes for 5 min in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)-0.1% sodium dodecyl sulfate (SDS) and two washes at 65°C for 15 min in 0.1× SSC-0.1% SDS. Fragments to which the probes hybridized were detected according to the manufacturer's instructions by using alkaline phosphatase-conjugated antidigoxigenin Fab fragments (1:5,000 in buffer B; Boehringer Mannheim) and LUMIPHOS (Boehringer Mannheim) and then were exposed to XAR film (X-Omat AR; Eastman Kodak Co., Rochester, N.Y.) for varying amounts of time at 37°C with no screens.

PCR. M. genitalium genomic DNA was isolated as previously described by Peterson et al. (25). Individual 100-μl reaction mixtures, each with 1.0 ng of DNA contained 200 μM each deoxynucleoside triphosphate, 2.5 U of Taq polymerase, 100 pmol of each primer, and 1× PCR buffer, which was supplied in the GENEMAP kit (Perkin-Elmer Cetus, Norwalk, Conn.). Two 24-mer oligonucleotide primers specific to M. genitalium dnaA and dnaN gene sequences were generated. Conditions for the 30 cycles of PCR were as follows: denaturation for 30 s at 94°C, annealing for 45 s at 50°C, and extension for 2 min at 72°C. PCR products were electrophoresed on 1.0% agarose in 1× TAE to determine their sizes. The primer sequences were as follows: dnaA1, 5′-ATTGATGCAAGAATGATFITCT-3′ and dnaN, 5′-GAAATTTCTCACAAGAATTTC-3′.

Nucleotide sequence accession number. The accession number of the submitted sequence from this study is U09251.

RESULTS AND DISCUSSION

Gene organization of a region of the M. genitalium chromosome. The clone Contig 34, which showed significant homology to the E. coli gyrA gene, was identified in a random sequence analysis of the M. genitalium genome (26). The insert of this clone was sequenced in both directions by using a mini Tn3 transposon sequencing approach (9). Chromosomal restriction fragments, pCBI and pF3-52, which overlap the original 5-kb EcoRI fragment were cloned, and sequence in both directions was obtained by using synthetic oligonucleotides.

The 6,140 nucleotides (nt) of contiguous DNA sequence from these three overlapping clones is diagrammed in Fig. 1, within the contiguous sequence, five ORFs have been identified. Two, which show sequence homology to gyrB and gyrA genes, are complete; two others, which show homology to dnaN and serS genes, have been partially sequenced. We did not detect an analog to the recF gene between the gyrB and dnaN genes; instead we found orf311, a complete ORF with no overall homology to entries in the database. This ORF does however contain a DNA box motif at its N terminus (3). The gene order of this M. genitalium chromosomal fragment, on the basis of significant sequence identities with entries in GenBank, is dnaN-orf311-gyrB-gyrA-serS (shown in Fig. 1).

We predict that this chromosomal segment is located near the origin of replication because of the highly conserved gene organization and structure of these regions in other bacterial species (23). Sequences generated from this project overlap six submissions to GenBank from the random sequencing project previously reported (accession numbers in parentheses) (25): sf10a (U02187), sf7a (U02199), sg4a (U02211), esa3 (U01696), sg4 (U02210), and Contig 34 (X61533).

DNA sequence analysis of the gyrB and gyrA genes of M. genitalium. The M. genitalium gyrB gene is 1,950 nt long and encodes a putative protein product of 650 amino acids (aa). This protein shows 90.6% identity with the M. pneumoniae gyrB translation product and 67.7 and 70.1% similarity to the E. coli and B. subtilis proteins, respectively, as determined by using the Genetics Computer Group alignment program GAP. Structurally, the 650-aa mycoplasma protein is more like the Bacillus and Staphylococcus subunits, as it does not contain the large insert present in the E. coli protein. Long stretches of homology in the N-terminal regions of GyrB proteins which have been shown to be involved in ATP binding are conserved in the M. genitalium protein (34), as are all the regions described by Huang (14) as being conserved in all type II topoisomerases. The putative GyrB proteins of M. genitalium and M. pneumoniae contain similar 10-aa inserts which are not present among the twelve GyrB proteins from other species that were compared (alignment examples shown in Fig. 2). These inserts, which are so far unique to the two evolutionarily related species M. genitalium and M. pneumoniae, may serve as good molecular markers for better assessment in the phylogenetic analysis of members of the class Mollicutes.

The M. genitalium gyrA gene is 2,508 nt long and encodes a protein of 836 aa which shows 65.2 and 67.5% homology with
the E. coli and B. subtilis proteins, respectively. As expected, 
the mycoplasmal GyrA protein is structurally more similar to 
the GyrA proteins of the gram-positive species, B. subtilis and 
Staphylococcus aureus, than to that of the gram-negative E. 
coli. In a manner similar to that of the GyrB subunit, those 
regions which have been demonstrated to be functionally 
active, e.g., the DNA binding site at Tyr-122 (E. coli numbering), 
are conserved in the M. genitalium subunit, as are the 
regions described by Huang for all type II topoisomerases (14).

The mycoplasmal GyrA proteins have a unique N-terminal 
leader which is not seen in the GyrA proteins of other species 
(Fig. 3). The two M. genitalium gyrase genes are separated by 14 nt, 
with a change in the translational reading frame between the 
two genes. This arrangement is similar to the arrangement in 
M. pneumoniae, in which there is also a change in the reading 
frame. **FIG. 1.** Schematic diagram showing the arrangement of the M. genitalium chromosomal fragments from which overlapping clones were 
generated as described in the text. ORFs, determined by analysis of the contiguous sequence generated from these clones, are labeled according to 
homologies with GenBank entries. E, EcoRI; B, BamHI; X, XhoI.

**FIG. 2.** PILEUP analysis of the regions of the GyrB proteins of 
several species which show the unique mycoplasmal insertion (in bold). The first amino acid corresponds to the E. coli protein aa 234. GyrB 
(gyrB) sequences are from the following species (accession numbers for 
the sequences are in parentheses): Ecogyra, E. coli (X04341); 
Ngogyra, Neisseria gonorrhoeae (M59981); Bsgyra, B. subtilis (X02369); 
Stagyra, S. aureus (X71437); Hlfgyra, Haloflexa sp. (M38373); Mpgyra, 
M. pneumoniae (X53555); Mggyra, M. genitalium (U09251); and 
Scgyra, Spiroplasma citri (Z19108).

**FIG. 3.** PILEUP analysis of the GyrA proteins from several differ-
ent species. *, the residue corresponding to Ser-83 of the E. coli 
protein, at which mutations confer nalidixic acid resistance. It is not 
known whether the wild-type mycoplasmal proteins actually confer 
resistance. 0, the tyrosine at the active site of the protein. Unique 
N-terminal leader sequences of the mycoplasmal proteins are in bold. 
GyrA (gyrA) sequences are from the following species (accession 
numbers are in parentheses): Mpgyra, M. pneumoniae (X53555); 
Mggyra, M. genitalium (U09251); Bsgyra, B. subtilis (X02369); Stagyra, 
S. aureus (X71437); and Ecogyra, E. coli (Y00544).
frame but the two genes overlap by 1 nt (7). We could identify no consensus promoter region between the \( \text{gyrB} \) and \( \text{gyrA} \) coding regions, suggesting that the two genes are transcribed as a polycistronic message, as is the case in most of the gram-positive species examined (5, 7). However, in most other species there is no change in reading frame between the two genes (5). A putative promoter (based on the \( E. coli \) \( \sigma^70 \) consensus sequence) upstream of the \( \text{gyrB} \) gene was identified. This promoter has a consensus \(-10 \) box (TATAAT at nt 1254 of the submitted sequence) but no apparent consensus \(-35 \) region. There is a putative factor-independent transcriptional terminator downstream of the \( \text{gyrA} \) gene within the coding sequence for \( \text{serS} \). A putative promoter for the \( \text{serS} \) gene which lies within the \( \text{gyrA} \) sequence contains only a weak consensus \(-10 \) box (two mismatches) and, again, no apparent \(-35 \) box. It must be noted, however, that there is potential difficulty in the identification of mycoplasmal promoter regions, since a consensus sequence has not yet been reported for these species, as relatively few complete genes have been sequenced.

Peterson et al. placed Contig 34 unambiguously on the physical map of \( M. genitalium \) near the X2-X3 intersection, by virtue of an \( \text{XhoI} \) site within the \( \text{gyrA} \) gene sequence (8, 26). We have also localized this clone by hybridizing it to one end of cosmids 5-30 from the collection of Lucier et al. (17; also data not shown).

The tandem arrangement of the two gyrase genes in \( M. genitalium \) is similar to that of gram-positive bacteria (5, 7, 16), thus providing support for the evolutionary relatedness of mycoplasmas to gram-positive bacteria, as first suggested by Woese et al. on the basis of an rRNA sequence comparison (35).

**DNA sequences of the genes surrounding the \( \text{gyrB- gyrA} \) genes.** On the basis of the highly conserved gene organization of the replication origin regions in other organisms (23), we expected to find a \( \text{recF} \) gene analog upstream of the \( \text{gyrB} \) gene in \( M. genitalium \). Instead, at 86 nt upstream and in the same reading frame as the \( \text{gyrB} \) gene we found an ORF of 933 nt which potentially encodes a protein product of 311 aa (orf311). orf311 is similar in size to the \( \text{recF} \) genes in other organisms that have been described previously (32), but it shows no overall sequence identity to any \( \text{recF} \) gene or protein in GenBank. In separate experiments, we used the cloned \( B. subtilis \) \( \text{recF} \) gene derived from plasmid pAK1 (16) as a hybridization probe against restriction digests of \( M. genitalium \) chromosomal DNA but were unable to detect any hybridization signal even under low-stringency hybridization conditions (data not shown). These results imply that a \( \text{recF} \) gene is absent in \( M. genitalium \) or, alternatively, that a sufficient difference in sequence exists to make detection by hybridization less efficient. It was reported by Dybvig et al. that the \( B. subtilis \) \( \text{recA} \) gene failed to identify its analog in mycoplasmas by hybridization. These authors later showed that these species actually contain a \( \text{recA} \) gene homolog by using PCR amplification with degenerative oligonucleotide primers (11).

When the MOTIFS algorithm is used, the N terminus of the predicted translation product of orf311 does show homology to one conserved region of a heat shock protein, DnaJ, and to a class of proteins which share a DnaJ box motif (3). DnaJ is a prominent player in the classic bacterial heat shock response (20); functional homologs contain three highly conserved structural motifs: the \( \text{N-terminal} \) 70 aa, which is called the DnaJ box motif; a glycine-rich region; and a series of Cys-X-Cys repeats (6). Clearly, the presence of orf311 is unique among the origin regions that have been described. A family of proteins containing the DnaJ box motif has been defined (3). Experiments with eukaryotic members of this protein family have allowed some speculation as to the function of this particular domain. The DnaJ box domain is thought to interact with the Hsp70 family of proteins (analogs of the \( E. coli \) dnaK gene) to stimulate Hsp70 ATP hydrolysis. DnaJ homologs may be required for the regulated dissociation of polypeptides from Hsp70 that would otherwise be stable; they may catalyze specific protein assembly and disassembly events within the cell (6). These proteins can have the DnaJ box domain at the N terminus, as it is in \( E. coli \) DnaJ or elsewhere. We believe that orf311 is not the functional dnaJ gene homolog of \( M. genitalium \) because clone esd11 (accession number U01724) from Peterson et al. (25) shows homology to dnaJ over its entire sequence and the putative protein which it encodes shares the other conserved regions of functional DnaJ homologs which Orf311 does not. Orf311 shows a low level of homology to esd11 except in the N-terminal region. esd11 has been mapped elsewhere on the physical map of the chromosome (13).

Upstream of the \( M. genitalium \) orf311 and overlapping it by 1 nt is a partial ORF whose predicted translation product shows a high level of homology to the C-terminal ends of the DnaN proteins of \( E. coli \), \( B. subtilis \), and Borrelia burgdorferi. Overlapping the \( 3' \) end of the \( \text{gyrA} \) gene by 27 nt is another putative gene for which we have only a partial sequence. The predicted protein product of this ORF shows similarity to SerS, a serine tRNA synthetase. In \( B. subtilis \) the \( \text{gyrA} \) and \( \text{serS} \) genes have been sequenced (up to 16), a PBP 5 analog would also be expected to exist. The guaB gene, which encodes an IMP dehydrogenase enzyme that is involved in the conversion of IMP to GMP (36), would have no use in \( M. genitalium \), since there is no evidence for de novo purine biosynthesis in that species (12, 25). PBP 5 genes encode small penicillin-binding proteins which have a speculated role in regulation of peptidoglycan cross-linking and in the cortex synthesis that occurs during sporulation (2). Since there is no cell wall in \( M. genitalium \) and no evidence for the existence of any genes relating to this function (25), a PBP 5 analog would also be dispensable.

**Physical linkage of dnaN and its associated gene cluster to the \( M. genitalium \) dnaA gene and the relationship of these genes to a putative origin region.** Using hybridization, we mapped clone sh3a (accession number U02229 [25]), which has homology to the dnaA gene of \( E. coli \), to the S2-X2 region of the chromosome. We also placed it on cosmids 5-30, which contains the gyrase genes (data not shown). Cosmid 5-30 contains about 35 kb of \( M. genitalium \) DNA upstream from the \( \text{gyrA} \) gene. We have directly linked the dnaN gene of \( M. genitalium \) to its dnaA gene by PCR. We successfully amplified a 1.6-kb product from both the \( M. genitalium \) chromosome and cosmids 5-30 DNA by using specific primers designed from the dnaN and dnaA gene sequences (data not shown). These data support the linkage of this \( M. genitalium \) chromosomal region to a putative origin, and because of the nature of primer construction, confirm that both dnaA and dnaN are transcribed in the same direction. The chromosomal origin of replication of \( M. capricolum \), which was defined by its biological function, is linked to the dnaA gene of that organism (19).

Recently, the origin region of the \( C. crescentus \) chromosome was shown to have a gene order altered from the ancestral gene order, with the presence of a dnaKJ operon and a purA gene between dnaA and the linked dnaN-recF-gyrB genes (37).
The gram-negative species *Buchnera aphidicola* (15) lacks a coding sequence between its *dnaN* and *gyrB* genes. *B. burgdorferi* (24) has the unusual gene order *dnaN-dnaA-gyrB-gyrA* in this region. For yet another example of divergence from the ancestral gene order, Miyata et al. have shown that in *M. capricolum*, a mycoplasma which causes disease in swine, the origin region has the gene order *dnaA-dnaN-orfL3-ksgA-orfL5* (19). It has not been shown whether these three species lack a *recF* completely or if it is just removed from this region on their respective chromosomes. We predict that *M. pneumoniae* has an ORF similar to *orfS1I* upstream of its *gyrB* gene on the basis of the partial sequence data available for that species (7). Now that species whose gene orders in origin regions diverge from the ancestral gene order are being recognized, it may be necessary to redefine the consensus gene order of this essential region. Among species of which the suspected origin has been linked to a biologically functioning origin of replication, most recently *C. crescentus* (37) and *M. capricolum* (19), the only conserved feature in this region is the presence of a *dnaA* gene, although some of the other highly conserved genes (i.e., *dnaN*, *recF*, and *gyrB*) are found within a close proximity.

We have identified a novel gene within the putative *M. genitalium* origin of replication. Further characterization of this region is required to confirm whether it is the functional origin or not. We are currently extending our analysis of this chromosomal region by examining the surrounding genes.

**ACKNOWLEDGMENTS**

We thank Ping-chuan Hu for generously growing cultures of *M. genitalium* from which DNA was isolated and Tom Lucier for his donation of *csm-53* and plasmid pGFP-52. We are also indebted to Clyde A. Hutchison III for the original gyrase clone Contig 34. We are grateful to Scott Peterson for helpful discussions and critical reading of the manuscript and to Christine Doyle for technical assistance and help with database submissions.

This research was supported by NIH grant AI33161 to K.F.B.

**REFERENCES**