JOURNAL OF BACTERIOLOGY, Oct. 1994, p. 5929–5937
0021-9193/94/$04.00+0

Vol. 176, No. 19

An Amplifiable DNA Region from the *Mycoplasma hyorhinis* Genome

GANG DENG AND MARK A. McINTOSH*

Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri—Columbia, Columbia, Missouri 65212

Received 18 January 1994/Accepted 4 April 1994

A novel amplifiable genomic region that displays variability in the number of tandem copies of a 1,368-bp DNA sequence (designated RS-2) was discovered among individual clonal derivatives within *Mycoplasma hyorhinis* broth-grown cell populations. Clonal isolates representing variant subpopulations from the original broth culture were of a single size variant, and although continued culture under a variety of growth conditions did not result in further amplification of RS-2, evidence for deletion events which reduced RS-2 copy number, presumably by homologous recombination, was obtained. RS-2 homologous sequences were identified in all *M. hyorhinis* strains tested, but only the tissue culture-derived strains GDL-1 and GDL-2 showed variability in genomic dosage. The RS-2 nucleotide sequence established that each tandem copy is flanked by direct repeats of a 20-bp sequence and suggested a possible mechanism for its original duplication as the initial phase of a genetic amplification process. The coding strand was defined by PCR amplification of a reverse transcriptase-generated cDNA, and its sequence revealed that RS-2 encodes a 456-residue internal, highly cysteine-rich domain of a larger *M. hyorhinis* protein whose coding sequence initiates and terminates in unique genomic sequences several hundred base pairs from RS-2 on either side of it. Changes in RS-2 copy number maintain the reading frame, and therefore the coding capacity, for this predicted size-variant protein.

Mycoplasmas are a diverse group of nearly 100 bacterial species that were suggested to have evolved monophyletically at a high rate (17, 35, 37) from the low-G+C-content gram-positive eubacterial phylogenetic branch (25). Their rapid evolutionary tempo reflects a high basal-level mutation rate (35, 37) and in several species may include a considerable frequency of more substantial genomic variabilities and rearrangements. Diversification among the mycoplasmas is apparent in various morphological, biochemical, and serological characteristics of the organisms (11), with individual strains within several mycoplasma species displaying significant genotypic, phenotypic, and antigenic variability. In *Mycoplasma hyorhinis*, recently isolated gene sequences (10, 29, 38) have identified important genetic structures expected to be significant factors in generating such diversity. For example, repetition of mobile genetic elements, such as insertion sequences and transposons, distributed throughout the genome, provides a fertile source of DNA homologies for recombinational genomic reordering (21). Although mycoplasmas have characteristically small genomes (24), several species, including *M. hyorhinis* (29), *M. hyopneumoniae* (10), *M. fermentans* (13), and *M. pulmonis* (3), contain multiple genomic copies of an insertion sequence related to IS*3*. It has not been established that interactions among such repetitive sequences are directly related to the high-frequency phenotypic variability within these species. However, their high copy number is likely to result in measurable recombinational variations, some of which might dictate cell survival under changing population pressures, and furthermore provides an attractive substrate for recombinational deletion of large genetic regions during the proposed (36) evolution of mycoplasmas by genetic attrition from their gram-positive progenitors.

* Corresponding author. Mailing address: Department of Molecular Microbiology and Immunology, University of Missouri School of Medicine, M613 Medical Sciences Building, Columbia, MO 65212. Phone: (314) 882-4133. Fax: (314) 882-4287. Electronic mail address: MAM@MEDSCI.MBP.MISSOURI.EDU.

Another example of genetic plasticity is associated with the high-frequency antigenic variation in several strains of *M. hyorhinis* that results from size and epitope heterogeneity of surface-exposed lipoproteins (26, 38). Examination of the repertoire of variable lipoprotein (*vlp*) genes from these strains revealed mosaic genetic structures that represent recombinatorial antigenic diversity-generating mechanisms.

This report describes a novel genetic variability discovered within a batch culture population of *M. hyorhinis* GDL-1. Variations in the copy number of a 1.4-kb DNA sequence were identified among individual isolates from the population. The characteristics of the sequence suggest that the variants were generated by recombination processes that resemble characterized prokaryotic gene amplification and reciprocal reduction events. Data are also presented to suggest that the amplifiable sequence encodes a large, internal polypeptide domain of a larger but size-variant protein. The high concentration of cysteine residues within this domain suggest multiple metal-binding motifs that are probably significant for the as yet unknown biological properties associated with these large proteins.

MATERIALS AND METHODS

Strains and culture conditions. The sources and culture conditions for *M. hyorhinis* GDL-1, GDL-2 (ATCC 23839), BTS-7 (M718-002-084), SK76, and PG29 have been described previously (5, 31, 34). *Escherichia coli* strains HB101 (4), JM101 (18), and DH5α (BRL Life Technologies, Inc., Gaithersburg, Md.) were grown with shaking at 37°C (unless otherwise noted) in Luria-Bertani medium (19). Recombinant plasmids in *E. coli* host strains were selected for by growth in Luria-Bertani medium supplemented with 50 μg of ampicillin per ml.

Isolation of nucleic acids and construction of recombinant plasmids. Basic molecular methods including plasmid isolation, ligation, transformation, end labeling, and DNA extraction from agarose gels were performed by procedures described by Sambrook et al. (27). Mycoplasma *Eco*RI genomic

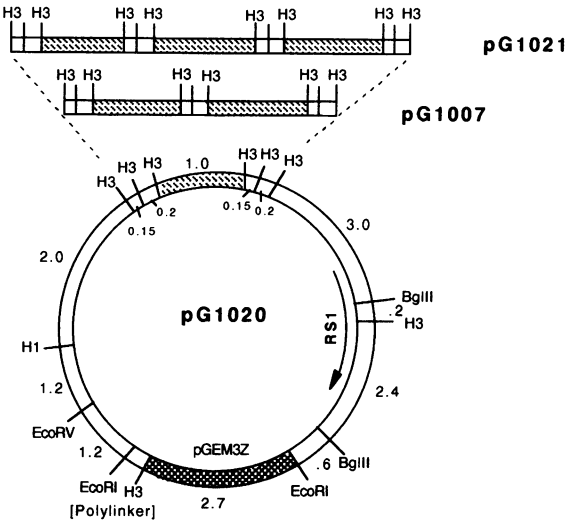

FIG. 1. Plasmid subclones containing the *M. hyorhinis* size-variant *Eco*RI fragment. Restriction endonuclease cleavage maps are given for the recombinant plasmids pG1020, pG1007, and pG1021 carrying size-variant *Eco*RI fragments from a recombinant phage library (30). The expanded areas of the last two plasmids depict the organization of the amplified DNA fragments. The positions of an IS-like element RS-1 and the vector sequences are shown for pG1020. The designation H3 represents *Hin*dIII sites, and H1 represents a *Hpa*I site. Other cleavage sites are designated by the appropriate restriction endonuclease. Sizes of DNA fragments are given in kilobases.

fragments from a library of recombinant phages (30) were subcloned into plasmid pGEM3Z (Promega, Madison, Wis.) to create recombinant plasmids pG1007, pG1020, and pG1021 (Fig. 1). The 1.0-, 0.2-, and 0.15-kb *Hin*dIII fragments from pG1020 were subcloned into pGEM3Z to create recombinants pG1020-1, pG1020-2, and pG1020-3, respectively. A 1.2-kb *Eco*RV-*Eco*RI fragment from pG1020 was subcloned into *Sma*I-*Eco*RI-digested pGEM3Z to generate pG1020-4.

DNA hybridization. Southern DNA transfers from agarose gels onto nitrocellulose filters were performed with a Vacublot apparatus (American Bionetics, Emeryville, Calif.) as specified by the manufacturer. Specific DNA probes, as either whole plasmids or isolated fragments, were labeled with [α-^{32}P]dATP (800 Ci/mmol; Dupont NEN Research Products, Boston, Mass.) by random oligonucleotide priming and extension with Klenow fragment (27). Probes were used at a specific activity of 10^6 cpm/ml of hybridization buffer. Hybridizations were performed as previously described (29) with a final wash stringency of 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 68°C.

Nucleotide sequence analysis. Double-stranded plasmid templates were isolated by the alkali lysis method (27) and purified by precipitation with 6.5% polyethylene glycol–0.8 M NaCl. Dideoxy sequencing reactions (28) were carried out by using modified bacteriophage T7 DNA polymerase (Sequenase version 2.0; United States Biochemical Corp., Cleveland, Ohio) with the reagents in the accompanying kit and [α-^{32}P] dATP. Oligonucleotide primers were produced at the University of Missouri DNA Core Facility. Sequencing reactions were electrophoresed on 8% polyacrylamide–7 M urea gels.

Computer-assisted analysis of nucleotide and protein sequence data was performed with the Microgenie sequence analysis program (23) distributed by Beckman Instruments, Palo Alto, Calif. DNA and protein homology searches used the FASTP and FASTA programs (15) with University of Wisconsin Genetics Computer Group software (8) to search the EMBL/GenBank Genetic Sequence and Swiss Prot databases.

Isolation of RNA. A log-phase *M. hyorhinis* GDL-1 culture was subcultured 1:2 in 500 ml of Hayflick's medium and grown for 24 h. Cells were harvested by centrifugation, washed once in 20 ml of phosphate-buffered saline, resuspended in 10 ml of 20 mM sodium acetate (pH 5.5)–0.5% sodium dodecyl sulfate–1 mM EDTA, and immediately subjected to two rounds of hot-phenol extraction. The aqueous phase was then extracted once with phenol-chloroform (1:1) and then with chloroform until a clear interface was obtained. RNA was precipitated by addition of 3 volumes of absolute ethanol and stored at −70°C until use.

Detection of RS-2 transcription by RT-PCR. *M. hyorhinis* GDL-1 RNA was first reverse transcribed into cDNA and then amplified by PCR. The procedure was modified from several methods (12, 20, 32). RNA was treated with 0.5 U of RNase-free DNase per μg, extracted with phenol-chloroform, and precipitated with ethanol to remove any contaminating DNA. Primers used in reverse transcription were RT-1 (5′-GTG TACTTCTTCAACAAGTGG-3′), a 21-bp sequence antisense to nucleotides 2127 to 2107 of the strand shown in Fig. 4, and RT-2 (5′-GGAATGTCATGAATGC-3′), a 16-bp sequence from nucleotides 1785 to 1800 of the strand shown in Fig. 4. Reactions were carried out at 42°C for 1 h with 10 to 20 μg of RNA in a total volume of 100 μl containing 2 U of avian myeloblastosis virus reverse transcriptase (RT) and 1× RT buffer (50 mM Tris-HCl [pH 8.0], 5 mM $MgCl_2$, 5 mM dithiothreitol, 50 mM KCl, 40 μM each deoxynucleoside triphosphate). The cDNA products were then amplified by 40 or 50 cycles of PCR. *Taq*I DNA polymerase was added after the initial denaturation to preserve its activity. Primers used for the RT-1 cDNA product (5′-CCCCTTCTAATGTTGC-3′ and 5′-GAGTTCTTAGATTCCC-3′) are located inside the RS-2 repeated domain and are separated by 340 bp. Primers for the RT-2 cDNA (5′-CAACAATTTCTCTACATC-3′ and 5′-GG AATGTGATGAATGC-3′) should yield a 260-bp product.

Nucleotide sequence accession number. The sequences described here have been assigned GenBank/EMBL accession number L11447.

RESULTS

Identification and characterization of RS-2. Studies of the genomic distribution of individual copies of the *M. hyorhinis* insertion sequence RS-1 led to the discovery of a size-variant *Eco*RI fragment reflecting a unique restriction fragment length polymorphism among subpopulations of the broth-grown culture represented in this particular genomic library (30). Physical mapping experiments indicated that the size heterogeneity within this particular *Eco*RI fragment was associated with the tandem repetition of sequences including all or parts of 0.15-, 0.20-, and 1.0-kb *Hin*dIII fragments located approximately 2 kb from the RS-1 sequence positioned on this fragment (Fig. 1). Three plasmid clones (pG1020, pG1007, and pG1021) carrying 12.3-, 13.65-, and 15.0-kb variants, respectively, of this *Eco*RI fragment were used for detailed molecular analysis. The directly repeating arrangement of this 1.35-kb RS-2 region suggested that it might have arisen from reciprocal recombinational amplification and reduction events, which could occur in the *M. hyorhinis* genome in response to uncharacterized selective pressure or may have been generated upon introduction into *E. coli*. Therefore, the structural organization of this genomic region within the *M. hyorhinis* chromosome was investigated by hybridization of *Eco*RI-digested chromosomal

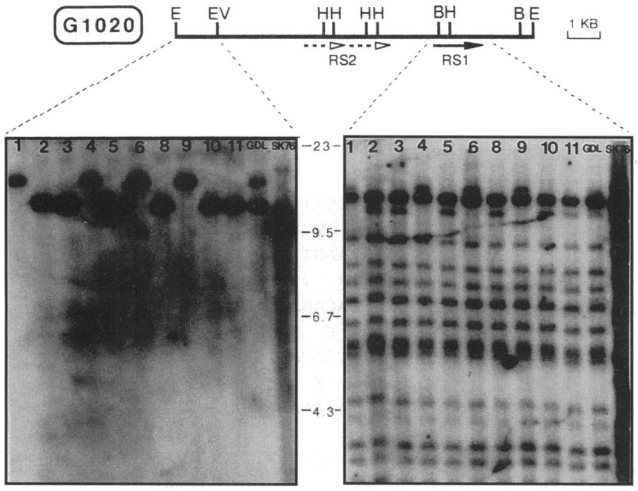

FIG. 2. Restriction fragment size variation associated with RS-2 among *M. hyorhinis* isolates. The restriction map of the pG1020 mycoplasma *Eco*RI fragment is depicted at the top (E, *Eco*RI; EV, *Eco*RV; H, *Hin*dIII; B, *Bgl*II). Probes for Southern hybridization are the 1.2-kb *Eco*RI-*Eco*RV fragment (left panel) and a 1.8-kb *Xba*I-*Hin*dIII probe (right panel) containing a copy of RS-1 from a previously characterized recombinant plasmid pJ125 (10). Lanes 1 to 6 and 8 to 11 represent chromosomal DNA from individual colony isolates of *M. hyorhinis* GDL-1, and lanes GDL and SK76 are from batch-passaged cultures of strains GDL-1 and SK76, respectively. The migration of λ *Hin*dIII size markers is designated in kilobases between the panels.

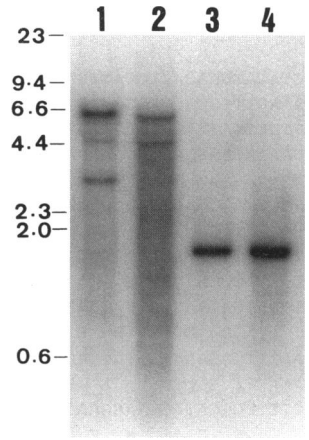

FIG. 3. Hybridization to RS-2 sequences from *M. hyorhinis* strains. Chromosomal DNA from batch cultures of *M. hyorhinis* GDL-1 (lane 1), GDL-2 (lane 2), SK76 (lane 3), and PG29 (lane 4) was digested with *Nhe*I, and the resulting fragments were separated on a 1% agarose gel. The gel was vacuum blotted onto nitrocellulose, which was dried and hybridized with the 1.0-kb *Hin*dIII probe containing only RS-2 sequences (Fig. 1). The migration of λ *Hin*dIII size markers is shown on the left.

DNA isolated from a continually passaged broth culture of *M. hyorhinis* GDL-1 with either a 1.2-kb *Eco*RI-*Eco*RV fragment (from one end of the 12.3-kb *Eco*RI fragment from pG1020) or an RS-1 specific probe (Fig. 2, lanes GDL). The former probe (left panel) hybridized intensely to *Eco*RI fragments equivalent in size to those characterized for pG1020 and pG1021, and upon further exposure, a fragment similar in size to that of pG1007 was also identified (data not shown), suggesting that all three variant forms of this genomic region were present in this cell population. Identical results were obtained with the 1.0-kb *Hin*dIII probe from within RS-2 (data not shown). The RS-1 probe (right panel) identified multiple *Eco*RI fragments, including two equivalent to those of pG1020 and pG1021.

Chromosomal DNAs from individually selected colonies isolated on agar plates from the original broth culture were shown to contain only one form of this variant genomic region when probed with the 1.2-kb *Eco*RI-*Eco*RV fragment (Fig. 2, lanes 1 to 6 and 8 to 11, left panel). These fragments were also clearly discernible among the multiple fragments detected with the RS-1 probe (right panel). A very limited survey of clonal populations derived from any one of the isolates containing the smaller, G1020-sized fragment did not reveal significant subsequent variations that might be generated from amplification in RS-2 copy number within these fragments (data not shown). Continued subculture of populations harboring the larger, G1021-sized fragment did, however, give rise to the low-frequency appearance of the smaller forms, presumably as a result of reduction in RS-2 copy number by simple deletion between homologous sequences.

Other strains of *M. hyorhinis* also contained sequences which hybridized to the RS-2-specific probe. The porcine-derived strain SK76 showed only a single stringently detectable *Eco*RI fragment slightly smaller than that from pG1020 (Fig. 2, lanes SK76). Other porcine derivatives, PG29 and BTS-7, gave similar results (data not shown). However, a broth culture of strain GDL-2, a continuously passaged tissue culture derivative (5), showed multiple hybridizing fragments very similar to those of GDL-1 and indicative of variations in the RS-2 copy number within individual subpopulations from this culture. Chromosomal DNAs from batch culture GDL-1 and GDL-2 populations, digested with *Nhe*I, which cleaves in unique sequences on either side of (but not within) RS-2, show multiple size-variant fragments containing RS-2 (Fig. 3). The SK76 and PG29 chromosomes, on the other hand, harbor a 1.4-kb-smaller RS-2-containing *Nhe*I fragment, which by size calculations contains only one copy of RS-2. No sequences hybridizing to RS-2 under stringent conditions were detected in several other mycoplasma species tested, including *M. bovis*, *M. flocculare*, *M. hyopneumoniae*, and *M. fermentans*.

Nucleotide sequence and key features of RS-2. The structural details of the size-variant RS-2 region were deduced from nucleotide sequence analysis of a contiguous 3,968-bp segment from the central region of pG1020 (Fig. 4). The strand depicted is that shown to be transcribed (see below). This analysis revealed two directly repeated sequences of 1,348 bp separated by a 20-bp "cassette" sequence that is also found flanking either side of these two copies of RS-2 (Fig. 4 and 5A). This configuration is analogous to that identified for recombinational gene amplification or reduction phenomena in other organisms (6, 9, 16) and suggests that the 20-bp cassettes represent the sites at which an initial recombinational duplication takes place. This predicted amplification or reduction mechanism is further supported by the analysis of RS-2 sequence organization in pG1007 (Fig. 5B) and pG1021 (Fig. 5C), deduced by restriction mapping and quantitation of restriction fragment copy number and confirmed by nucleotide sequence analysis. There are three tandemly repeated identical copies of RS-2 in pG1007 and four in pG1021, each separated by individual copies of the cassette sequence and with single copies at either boundary of the RS-2-amplified region.

The sequences of the three 20-bp cassettes from pG1020 are also shown in Fig. 5A; all three sequences are characterized by


SD
P3 START
+1

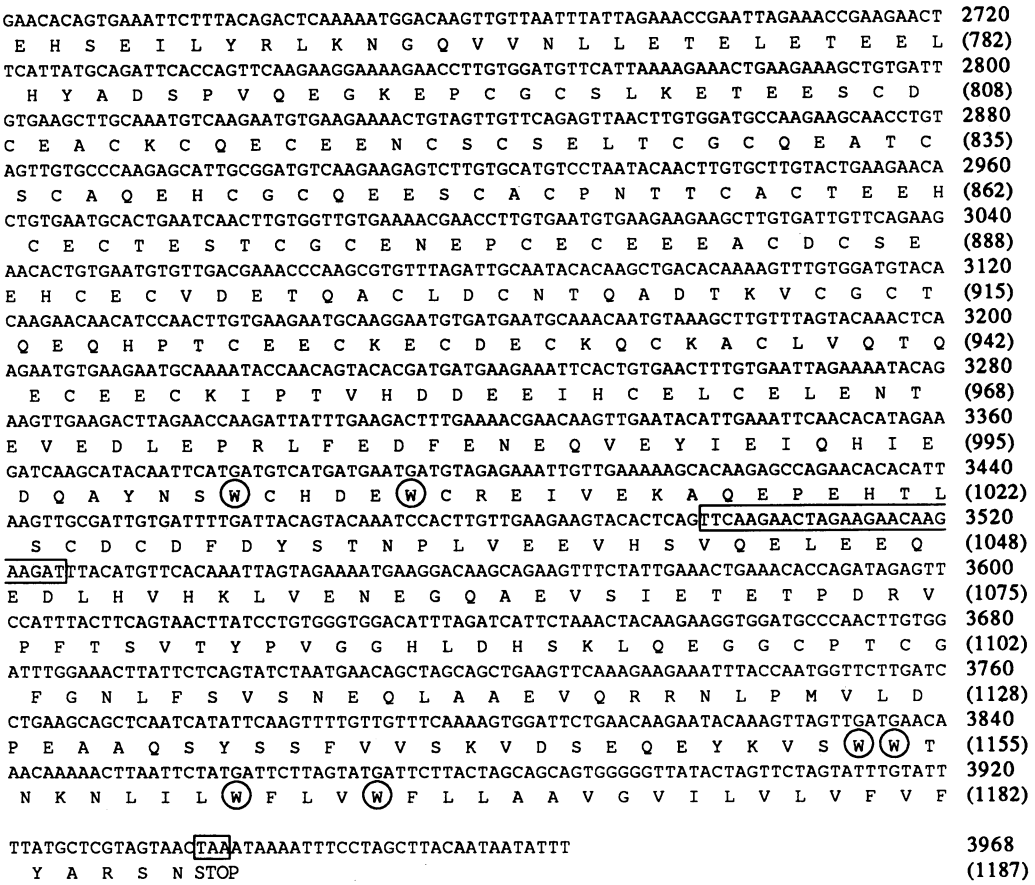

FIG. 4. Nucleotide sequence of the *M. hyorhinis* GDL-1 genomic region containing RS-2. A 3,986-bp segment from the central region of pG1020 was sequenced to completion on both strands. The sequence of the strand corresponding to an identified mRNA product (see Fig. 6) is shown 5′ to 3′, and the nucleotide positions are given at the right end of each sequence line. The single large P3 ORF is denoted by the single-letter amino acid code centered under each codon, and its residue numbers are given in parentheses at the right end of each amino acid sequence line. The putative P3 Met translational start codon (boxed), its probable ribosome-binding site (underlined and noted as SD), and the TAA stop codon (boxed) are indicated. Eight in-frame TGA codons encoding tryptophan are circled. Multiple 20-bp cassette sequences demarcating two identical 1,348-bp RS-2 sequences are noted by long boxes; the most 3′ cassette has 26 bp. Nucleotide sequence analysis from the similar regions of pG1007 and pG1021 revealed three and four identical copies of RS-2, respectively, and the corresponding increases in cassette numbers flanking these copies. The surrounding unique sequences representing possible P3 coding sequences were also unchanged (data not shown). The circled G residue upstream of the P3 start codon is the most likely transcriptional start site (+1) determined by preliminary primer extension analysis (data not shown).

two copies of the hexanucleotide sequence 5′-CAAGAA-3′ separated by 3 nucleotides, except in the farthest downstream copy (as shown in Fig. 4 and 5), where an extra 6 nucleotides, 5′-GAAGAA-3′, are present. The 20-bp cassettes from pG1007 and pG1021 have identical features, with the downstream cassette having 26 nucleotides and all upstream copies having the identical 20 nucleotides.

RS-2 transcription. To determine which RS-2 DNA strand is transcribed in mycoplasmas, we attempted to detect mRNA products from either strand as a potential coding template. It occurred to us that since RS-2 amplification might be induced by some yet unknown growth factor(s), which might also induce RS-2 transcription, it could be difficult to detect transcription products until the factor is identified. Attempts to identify the mRNAs from either strand by standard Northern (RNA) hybridization failed to detect any signal, perhaps because of low levels of transcription under standard broth culture conditions and in the absence of any known induction signals. Subsequently, PCR amplification was performed following an RT reaction with specific primers for either strand. This method allows any cDNA synthesized during the RT reaction to be amplified by PCR when additional corresponding primers for either template are added. Under identical conditions with aliquots of the same mycoplasma RNA preparations, only the reaction with the RT-1 primer (specific to the strand shown in Fig. 4) produced a cDNA that could be amplified. With 40 cycles of PCR amplification of the cDNA products, followed by agarose gel electrophoresis, a specific band at the predicted size (340 bp) was observed in only that reaction, whereas no product was produced in the control reaction without RT (Fig. 6), confirming that this product did not result from contamination of the RNA samples with chromosomal DNA. The expected 340-bp product strongly hybridizes with an RS-2-specific probe (data not shown). No products were observed with the RT-2 and PCR primers specific for the other strand. These results strongly suggested that the only transcript in *M. hyorhinis* from the RS-2 genetic region, although at a relatively low level under the growth conditions used in this experiment, is the strand shown in Fig. 4.

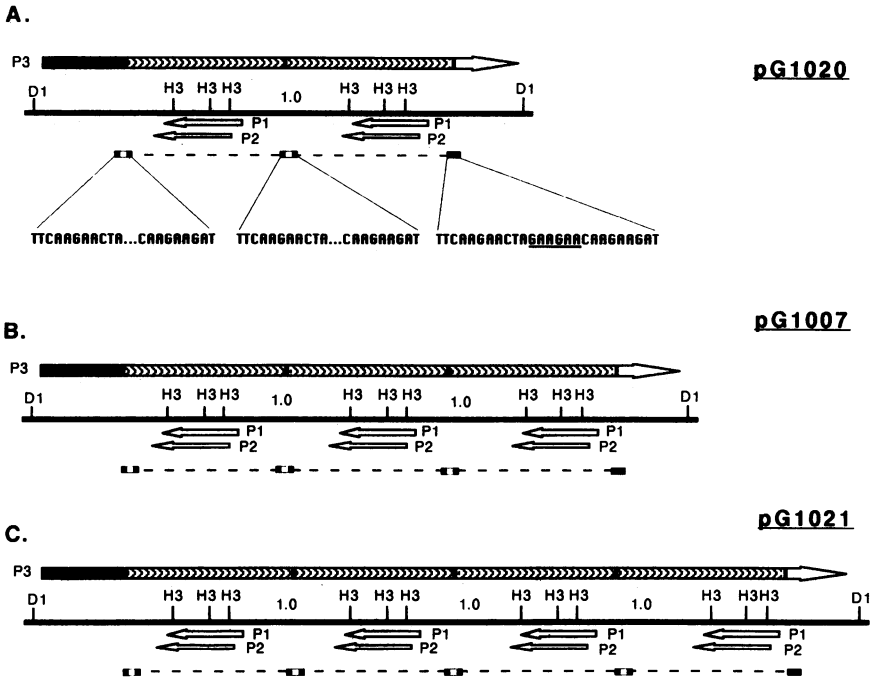

FIG. 5. Schematic representation of structural features (deduced from nucleotide sequences) of the RS-2 regions of strain GDL-1 size-variant populations. Restriction maps for the size-variant subclones are depicted. H3, *Hin*dIII; D1, *Dra*I. The dotted line below each map represents the extent of the 1,348-bp RS-2 sequences; ▯ and ■, represent 20- and 26-bp cassette sequences, respectively, at junctions between RS-2 copies and between RS-2 and the flanking genomic sequences. The sequences of these cassettes are expanded below the corresponding boxes in panel A. The larger hatched arrows above each map designate the direction and extent of the coding sequence for P3, with the solid and open regions signifying unique coding sequences at the 5′ and 3′ ends, respectively, and the stippled regions signifying RS-2-encoded redundant sequences. The small white arrows under each map designate the direction and extent of overlapping P1 and P2 ORFs deduced from the DNA sequence of the complementary strand to encode potential polypeptides of 20 and 18 kDa, respectively. Evidence including the lack of a detectable transcript, poor translation signals, and no significant homologies to database proteins suggests that they are not produced in mycoplasmas.

Protein-coding capacity of RS-2 and flanking sequences. Computer analysis of the translational capacity for the RS-2 nucleotide sequences revealed an extensive open reading frame (ORF) on the depicted strand (Fig. 4) that is maintained in all size variants and that therefore encodes potentially size-variant protein products, which we have arbitrarily designated P3. The P3 ORF initiates following a strong ribosome-binding site (5′-AGGAG-3′) at an AUG codon located in unique sequences 388 bp upstream from the most 5′ cassette sequence at the boundary of RS-2. This ORF extends entirely across the RS-2 region and is maintained in frame no matter how many copies of RS-2 are present. It terminates with a UAA codon located 411 bp downstream of the most 3′ cassette. Therefore, genetic amplification or reduction in RS-2 copy number, which maintains the appropriate reading frame, results in expansion or contraction of the final product size in multiples of an internal 456-amino-acid domain. The P3 product from pG1020 contains 1,187 residues with a calculated molecular weight of 134,938; the pG1007 product is 1,643 amino acids in length with a molecular weight of 186,544; and the predicted pG1021 P3 product of 2,099 amino acids would have a molecular weight of 238,149. If the coding sequence contained only one copy of RS-2, as is apparent in the porcine-derived strains SK76 and PG29, the potential P3 product would contain 731 residues with an estimated molecular weight of 83,332.

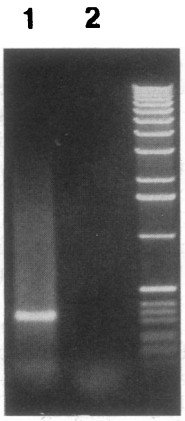

FIG. 6. RS-2 transcription detected by PCR. RNA prepared from *M. hyorhinis* GDL-1 was first reverse transcribed into cDNA and then amplified by PCR. Conditions were described in Materials and Methods. Lanes: 1, P3-specific PCR products from cDNA made from *M. hyorhinis* RNA with avian myeloblastosis virus RT; 2, products from a similar reaction without RT added. The size markers (right lane) are the 1-kb DNA ladder (BRL Life Technologies, Inc.).

Analysis of the P3 ORF for codons specifying tryptophan (Trp) reveals two UGA codons in each copy of RS-2 and another four UGA codons in the 3′ unique sequences of P3. However, the only Trp codon in the 5′ unique sequences is specified by the universal UGG.

The most striking feature of the 456-residue peptide specified by RS-2 is its extremely high cysteine (Cys) content of 61

residues or 13.4% of the total. The unique domains contain another five Cys residues, making the overall Cys content around 12%.

DISCUSSION

This study describes the identification and sequence of a novel region of the *M. hyorhinis* GDL-1 genome that contains up to four tandem copies of the 1.4-kb RS-2 DNA sequence. Evidence is presented to suggest that the copy number of RS-2 varies among individual isolates of a GDL-1 cell population and may be influenced by undetermined cell culture conditions. The variable region was defined as a 1,368-bp sequence that is absolutely conserved in all tandem copies examined. The repeated sequence is open for translation in one reading frame along its entire length no matter how many copies; it is therefore predicted to encode a 456-amino-acid internal cysteine-rich domain of a large but size-variable mycoplasma protein, which has translational initiation and termination signals located in unique sequences flanking the RS-2 element. In addition, the 5′-terminal 20 nucleotides of RS-2 are repeated once again at the distal 3′ end of the entire variable region, although always with an additional, centrally located 6 nucleotides. The overall structure of RS-2, therefore, which is made up of tandem repetitions of a 1,348-bp sequence flanked by short, directly repeating sequences (in this case, 20-bp cassettes), is similar to previously characterized examples of prokaryotic and eukaryotic genes capable of reciprocal amplification and reduction events affecting their ultimate copy number; these alterations usually become apparent in response to the presence or absence of selective growth pressures in cell populations (9, 16, 22, 33). Such events in this case, however, are predicted to alter the size of the mycoplasma P3 protein by changing the copy number of a large internal coding domain.

In the several examples of prokaryotic gene amplification investigated in some molecular detail (9, 22, 33), it has been suggested that these events occur in two defined phases. In the first, a direct repeat of the sequence is generated, perhaps by molecular processes such as transposition or illegitimate recombination. The second phase generates longer arrays of directly repeated sequences, several hundred copies in some instances, by a process that probably involves unequal recombination catalyzed by RecA-like recombinases.

The finding in microbes that many, if not all, of these amplification events occur at short, repeated sequences (9, 33) suggests that the initial sequence duplication in the amplification process is mediated by the short sequence repeat (33). In the case of the RS-2 sequences, all copy number variants are flanked by identical 20-bp cassettes, suggesting that the generation of these short duplicated sequences initiated the amplification process. Each cassette contains two direct copies of the hexanucleotide sequence CAAGAA separated by the trimer CTA; however, the distal cassette in each variation always contains an additional GAAGAA sequence located between the trimer and the downstream hexanucleotide motif. The molecular architecture of the standard 20-bp sequence suggests that the hexanucleotide motif may represent the binding site for the recombination machinery that created the original RS-2 duplication. Moreover, the aberrant distal cassette may reflect a slippage or mismatch in a replicative process that generated this particular copy of the cassette. The fact that this aberration was found in all RS-2 variants examined in this GDL-1 population indicates that it was involved in, or created as a result of, the original phase I RS-2 duplication. Subsequent RecA-mediated phase II multiplications would not involve this distal cassette.

At present, there is no direct evidence to define a molecular mechanism responsible for this gene amplification event. As in most other prokaryotic examples, these unscheduled events occur at a low frequency in the population, providing few if any molecular intermediates that might define the details of the process. Mixed-cell populations of *M. hyorhinis* GDL-1 were shown to contain variant subpopulations with regard to the RS-2 copy number. Clonal isolates from this population were of a single-copy-number variant and were relatively stable throughout multiple passages in broth. For example, the low-copy-number variant, containing two tandem RS-2 copies, did not give rise to higher- or lower-copy-number variants at any detectable frequency through many tens of generations in laboratory culture. Modification of the culture medium to include variable concentrations of metal ions, including zinc, copper, and thallium, was attempted to enrich for possible amplification events since the multiplicity of cysteine-rich metal-binding motifs in the P3 polypeptide suggested a significant metal-binding capacity. However, these modifications did not induce or select any detectable higher-gene-dosage variants. Similar passages of the high-copy-number subclone, containing four tandem RS-2 copies, did not generate any detectable variants with a higher copy number; however, Southern hybridization with whole-population chromosomal DNA did reveal the low-frequency appearance of genomic fragments containing two or three RS-2 copies. These presumably reflect reduced-copy variants generated by RecA-directed deletion events. The present limited survey of *M. hyorhinis* growth parameters therefore does not provide any substantial clues to explain the derivation of the higher-copy-number subpopulations in our original GDL-1 stock population, a stock culture that has been maintained at −70°C in the laboratory for many years. It does suggest, however, that in the absence of whatever selective growth conditions favor higher RS-2 copy number, continuous passage in standard broth culture will result in a stable population of low-copy-number derivatives.

A survey of three other *M. hyorhinis* strains revealed a similar variability in RS-2 copy number in a stock population of strain GDL-2, but cultures of strains SK76 and PG29 contained only a single variant equivalent to an RS-2 copy number of one. To date, no clonal isolates of GDL-1 with only one copy of RS-2 have been identified. From the standpoint of RS-2 variation, it might be significant that both strains GDL-1 and GDL-2 were isolated as tissue culture contaminants and have been maintained as high-passage type strains by reference laboratories. Strains SK76 and PG29, on the other hand, were isolated from infected swine (5) and have not been passaged extensively. Since changes in RS-2 dosage may reflect previous exposure of the strains to favorable conditions for given subpopulations, these strain variations suggest that previous growth conditions in association with tissue culture cell lines or during multiple laboratory passages enriched for higher-copy-number subpopulations. While we have not yet identified RS-2 changes in response to multiple growth cycles in standard broth culture or upon exposure to certain metal ions, it remains to be tested whether alternative growth conditions or infection of tissue culture cell lines can lead to an enrichment of a particular variant subpopulation.

Analysis of the nucleotide sequence of RS-2 and its flanking genomic region revealed an extensive ORF on the DNA strand shown by RT-PCR to be expressed as an mRNA in *M. hyorhinis* cells. This P3 coding region initiates and terminates outside the RS-2 repeat sequence but is always maintained in frame in each of the RS-2 variants examined. As a result, changes in RS-2 dosage result in size variants of P3 containing

multiples of an internal, highly cysteine-rich, 456-amino-acid domain. The distribution of Cys residues within this domain resembles classic motifs found in several characterized metal-binding proteins. For example, the residues 202 to 221 and 488 to 507 (as depicted in Fig. 4) contain characteristic zinc-finger motifs, —C-X_2-C-X_{12}-C-X_2-C—, found in a variety of molecules including many DNA-binding proteins (1, 2). The region between residues 466 and 481 contains five tandem copies of the motif —C-X_2-C—, typical of cysteine-rich sites in metal-binding proteins like metallothionein and rubredoxin and in metal-containing enzymes like liver alcohol dehydrogenase (1). The most dramatic cysteine-rich region is the 10 overlapping copies of the motif —C-X-C-X_4-C-X-C— between residues 365 and 437. This motif has been characterized in aspartate carbamoyltransferase and ferredoxin (1), but the multiple-copy configuration is more commonly found in structural rather than catalytic sites of zinc ligands such as alcohol dehydrogenase and metallothionein. This motif is also found in cysteine-rich domains of the integrin superfamily of eukaryotic lymphocyte receptors (14).

Antibodies generated against a recombinant form of P3 purified from *E. coli* have identified multiple forms of the authentic protein expressed in *M. hyorhinis* populations (7). Examination of the cellular location and biochemical characteristics of these polypeptides may provide some useful information to explain not only the potentially interesting function of this unique mycoplasma protein but also which biological parameters might influence changes in the dosage of its internal coding domain. Although at present the only clue to such parameters focuses attention on the tissue culture cell association of strains GDL-1 and GDL-2, these studies indicate that all *M. hyorhinis* strains examined to date have conserved the gene encoding P3 and that amplifications or reductions in RS-2 copy number maintain its coding frame. These observations suggest that the P3 protein plays an important role in the biology of *M. hyorhinis*.

ACKNOWLEDGMENTS

We thank K. S. Wise for technical advice and helpful discussions during the course of these studies.

This work was supported by grant AR35587 from the National Institutes of Health.

REFERENCES

1. **Berg, J. M.** 1986. Potential metal-binding domains in nucleic acid binding proteins. Science **232**:485–487.
2. **Berg, J. M.** 1990. Zinc fingers and other metal-binding domains. J. Biol. Chem. **265**:6513–6516.
3. **Bhugra, B., and K. Dybvig.** 1993. Identification and characterization of IS*1138*, a transposable element from *Mycoplasma pulmonis* that belongs to the IS*3* family. Mol. Microbiol. **7**:577–584.
4. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41**:459–474.
5. **Boyer, M. J., and K. S. Wise.** 1989. Lipid-modified surface protein antigens expressing size variation within the species *Mycoplasma hyorhinis*. Infect. Immun. **57**:245–254.
6. **Cross, M., and R. Renkawitz.** 1990. Repetitive sequence involvement in the duplication and divergence of mouse lysozyme genes. EMBO J. **9**:1283–1288.
7. **Deng, G., K. S. Wise, and M. A. McIntosh.** 1994. Unpublished data.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
9. **Edlund, T., and S. Normark.** 1981. Recombination between short DNA homologies causes tandem duplication. Nature (London) **292**:269–271.
10. **Ferrell, R. V., M. B. Heidari, K. S. Wise, and M. A. McIntosh.** 1989. A *Mycoplasma* genetic element resembling prokaryotic insertion sequences. Mol. Microbiol. **3**:957–967.
11. **Freundt, E. A., and D. G. Edward.** 1979. Classification and taxonomy, p. 1–41. *In* M. F. Barile and S. Razin (ed.), The mycoplasmas. Academic Press, Inc., New York.
12. **Fuqua, S. A. W., S. Fitzgerald, and W. L. McGuire.** 1990. A simple polymerase chain reaction method for detection and cloning of low-abundance transcripts. BioTechniques **9**:206–211.
13. **Hu, W. S., R. Y.-H. Wang, R.-S. Liou, J. W.-K. Shih, and S.-C. Lo.** 1990. Identification of an insertion-sequence-like genetic element in the newly recognized human pathogen *Mycoplasma incognitus*. Gene **93**:67–72.
14. **Larson, R. S., and T. A. Springer.** 1990. Structure and function of leukocyte integrins. Immunol. Rev. **114**:181–217.
15. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
16. **Maeda, N., and O. Smithies.** 1986. The evolution of multigene families: human haptoglobin genes. Annu. Rev. Genet. **20**:81–108.
17. **Maniloff, J.** 1992. Phylogeny of mycoplasmas, p. 549–560. *In* J. Maniloff, R. N. McElhaney, L. R. Finch, and J. B. Baseman (ed.), Mycoplasmas: molecular biology and pathogenesis. American Society for Microbiology, Washington, D.C.
18. **Messing, J., R. Crea, and P. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
19. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Noonan, K. E., C. Beck, T. A. Holzmayer, J. E. Chin, J. S. Wunder, J. L. Andrulis, A. F. Gazdar, C. L. Willman, B. Griffith, D. D. VonHoff, and I. B. Roninson.** 1990. Quantitative analysis of MDR1 (multi-drug resistance) gene expression in human tumors by polymerase chain reaction. Proc. Natl. Acad. Sci. USA **87**:7160–7164.
21. **Petes, T. D., and C. W. Hill.** 1988. Recombination between repeated genes in microorganisms. Annu. Rev. Genet. **22**:147–168.
22. **Petit, M.-A., J. M. Mesas, P. Noirot, F. Morel-Deville, and S. D. Ehrilich.** 1992. Induction of DNA amplification in the *Bacillus subtilis* chromosome. EMBO J. **11**:1317–1326.
23. **Queen, C., and L. J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. **12**:581–599.
24. **Razin, S.** 1992. Mycoplasma taxonomy and ecology, p. 3–22. *In* J. Maniloff, R. N. McElhaney, L. R. Finch, and J. B. Baseman (ed.), Mycoplasmas: molecular biology and pathogenesis. American Society for Microbiology, Washington, D.C.
25. **Rogers, M. J., J. Simmons, R. T. Walker, W. G. Weisburg, C. R. Woese, R. S. Tanner, I. M. Robinson, D. A. Stahl, G. Olsen, R. H. Leach, and J. Maniloff.** 1985. Construction of the mycoplasma evolutionary tree from 5S rRNA sequence data. Proc. Natl. Acad. Sci. USA **82**:1160–1164.
26. **Rosengarten, R., and K. S. Wise.** 1991. The Vlp system of *Mycoplasma hyorhinis*: combinatorial expression of distinct size variant lipoproteins generating high-frequency surface antigenic variation. J. Bacteriol. **173**:4782–4793.
27. **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.
28. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
29. **Taylor, M. A., R. V. Ferrell, K. S. Wise, and M. A. McIntosh.** 1988. Reiterated DNA sequences defining genomic diversity within the species *Mycoplasma hyorhinis*. Mol. Microbiol. **2**:665–672.
30. **Taylor, M. A., M. A. McIntosh, J. Robbins, and K. S. Wise.** 1983. Cloned genomic DNA sequences from *Mycoplasma hyorhinis* encoding antigens expressed in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80**:4154–4158.
31. **Taylor, M. A., K. S. Wise, and M. A. McIntosh.** 1985. Selective detection of *Mycoplasma hyorhinis* using cloned genomic DNA fragments. Infect. Immun. **47**:827–830.
32. **Wang, A. M., M. V. Doyle, and D. F. Mark.** 1989. Quantitation of mRNA by the polymerase chain reaction. Proc. Natl. Acad. Sci. USA **86**:9717–9721.

33. **Whoriskey, S. K., V.-H. Nghiem, P.-M. Leong, J.-M. Masson, and J. H. Miller.** 1987. Genetic rearrangements and gene amplification in *Escherichia coli*: DNA sequences at the junctures of amplified gene fusions. Genes Dev. **1:**227–237.
34. **Wise, K. S., and M. F. Kim.** 1987. Major membrane surface proteins of *Mycoplasma hyopneumoniae* selectively modified by covalently bound lipid. J. Bacteriol. **169:**5546–5555.
35. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
36. **Woese, C. R., J. Maniloff, and L. B. Zablen.** 1980. Phylogenetic analysis of the mycoplasmas. Proc. Natl. Acad. Sci. USA **77:**494–498.
37. **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.
38. **Yogev, D., R. Rosengarten, R. Watson-McKown, and K. S. Wise.** 1991. Molecular basis of *Mycoplasma* surface antigenic variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5′ regulatory sequences. EMBO J. **10:**4069–4079.