

Genomic Analysis Reveals *Mycoplasma pneumoniae* Repetitive Element 1-Mediated Recombination in a Clinical Isolate[∇]

Oxana Musatovova, T. R. Kannan, and Joel B. Baseman*

Department of Microbiology and Immunology, 7703 Floyd Curl Drive, The University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229-3900

Received 6 December 2007/Returned for modification 7 January 2008/Accepted 10 January 2008

Mycoplasmas are cell wall-less bacteria that evolved by drastic reduction of the genome size. Complete genome analysis of *Mycoplasma pneumoniae* revealed the presence of numerous copies of four distinct large *M. pneumoniae* repetitive elements (RepMPs). One copy each of RepMP2/3, RepMP4, and RepMP5 are localized within the P1 operon (MPN140 to MPN142 loci), and their involvement in sequence variation in adhesin P1 and adherence-related protein B/C has been documented. Here we analyzed a clinical strain of *M. pneumoniae* designated S1 isolated from a 1993 outbreak of respiratory infections in San Antonio, TX. Based on the type of RepMPs within the P1 operon, we classified clinical isolate S1 as type 2 with unique minor sequence variations. Hybridization with oligonucleotide arrays revealed sequence divergence in two previously unsuspected hypothetical genes (MPN137 and MPN138 loci). Closer inspection of this region revealed that the MPN137 and MPN138 loci harbored previously unrecognized unique RepMP1 sequences found only in *M. pneumoniae*. PCR and sequence analyses revealed a recombination event involving three RepMP1-containing genes that resulted in fusion of MPN137 and MPN138 reading frames and loss of all but a short fragment of another RepMP1-containing locus, MPN130. The multiple copies of unique RepMP1 elements spread throughout the chromosome could allow vast numbers of sequence variations in clinical strains. Comparisons of amino acid sequences showed the presence of leucine zipper motifs in MPN130 and MPN138 proteins in reference strain M129 and the absence of these motifs in the fused protein of S1. The presence of tandem leucine and other repeats points to possible regulatory functions of proteins encoded by RepMP1-containing genes.

Mycoplasma pneumoniae is a well-known human bacterial pathogen of the respiratory tract. The majority of clinical manifestations of *M. pneumoniae* infections range from tracheobronchitis and pharyngitis to pronounced signs of pneumonia and extrapulmonary pathologies. *M. pneumoniae* was identified in the early 1960s as the cause of primary atypical pneumonia, and some of the earliest documented outbreaks occurred in the military and in nonmilitary settings such as educational institutions, hospitals, and summer camps. In general, *M. pneumoniae* is among the most common causes of respiratory infections in school-age children and adults, and 20 to 30% of all community-acquired pneumonia cases are directly linked to *M. pneumoniae*. *M. pneumoniae* infections also lead to a spectrum of complicated sequelae, including asthma and other airway maladies, arthritis, pericarditis, central nervous system disorders, including encephalomyelitis, and autoimmune-associated pathologies (2, 39).

M. pneumoniae is among the smallest self-replicating prokaryotic pathogens and has a genome size of 816 kb; about 8% of the genome is composed of repetitive DNA elements (RepMPs) (16). The existence of a high number of RepMPs, along with the loss of many metabolic functions during the process of genome reduction, indicates the important role that multiple repeats are likely to play in the biology of *M. pneu-*

moniae (29). We and other workers have hypothesized that numerous copies of RepMPs form a pool of repetitive elements for homologous recombination which could provide *M. pneumoniae* with considerable biological resourcefulness (10, 29, 34–36, 43).

Three of the four RepMPs identified (RepMP2/3, RepMP4, and RepMP5) are in the P1 operon (31). This three-gene operon contains the gene encoding cytoadhesin P1 (MPN141 locus, also known as ORF5) followed by the MPN142 locus (also known as ORF6). Proteins encoded by the latter gene are required for P1 function and along with P1 likely constitute a major adhesin complex (40, 41). Two of the repetitive elements found in the P1 gene, RepMP2/3 and RepMP4, have been studied in depth (31, 34, 35). Analysis of the complete genome sequence of *M. pneumoniae* strain M129 revealed 10 copies of RepMP2/3 and 8 copies of RepMP4 distributed throughout the chromosome (16). Another large repetitive element, RepMP5, was identified in the proximal region of MPN142, and seven additional copies of RepMP5 are distributed throughout the remaining *M. pneumoniae* genome (16).

Sequence variations believed to be a result of RepMP2/3- and RepMP4-mediated recombination have been identified within the open reading frame of the P1 gene of clinical *M. pneumoniae* strains. Based on sequence divergence, *M. pneumoniae* strains belong to one of two groups; reference strain M129 is a representative of type 1, and strain FH is representative of type 2 (35). Interestingly, analysis of the repetitive region of ORF6 (RepMP5) also revealed sequence differences between the M129 and FH strains (30). Extensive analysis of 115 strains confirmed that this divergence is type specific, and

* Corresponding author. Mailing address: Department of Microbiology and Immunology, 7703 Floyd Curl Drive, The University of Texas Health Science Center at San Antonio, San Antonio, TX 78229-3900. Phone: (210) 567-3939. Fax: (210) 567-6491. E-mail: baseman@uthscsa.edu.

[∇] Published ahead of print on 22 January 2008.

TABLE 1. Genes compared in reference strain M129 and clinical strain S1 of *M. pneumoniae*

Locus	Protein	Position in M129 genome (nucleotides)
Adherence-related genes		
MNP140	DHH family phosphoesterases	179871–180845
MPN141	Adhesin P1	180858–185741
MPN142	Cytadherence related protein(s)	185747–189403
RepMP1-containing genes		
MPN037	Hypothetical protein	45770–46213
MPN094	Hypothetical protein	116287–116709
MPN100	Hypothetical protein	129458–130009
MPN130	Hypothetical protein	169042–169464
MPN137 ^a	Hypothetical protein	178143–177457
MPN138 ^a	Hypothetical protein	178892–178392
MPN139	Hypothetical protein	179620–179129
MPN204	Hypothetical protein	247655–248101
MPN283	Hypothetical protein	336479–336826
MPN368	Hypothetical protein	439220–439762
MPN410	Hypothetical protein	494694–495140
MPN465	Conserved hypothetical protein	569244–568645
MPN484	Hypothetical protein	588613–588302
MPN501	Hypothetical protein	608167–608757
MPN524	Hypothetical protein	646051–645545
MPN655	Hypothetical protein	780008–780622

^a RepMP1-containing gene identified in this study.

the corresponding repeats were designated RepMP5-1 and RepMP5-2 (9).

Among the large repetitive elements, RepMP1 is the only one not found within the P1 operon (43). Analysis of the complete genomic sequence of the M129 reference strain revealed 14 RepMP1 copies (RepMP1 genes) (Table 1) (16, 17, 43), several of which seem to encode parts of expressed proteins (R. Herrmann, personal communication). Since its first description in 1988 (43), RepMP1 has also been hypothesized to be involved in recombination (10), but this prediction has not been documented yet.

In the current study, we compared the genomes of two *M. pneumoniae* isolates and analyzed variations in repetitive sequences. While analysis of RepMP2/3, RepMP4, and RepMP5 within the cytoadhesin P1 operon revealed type 2 sequences, we discovered novel DNA sequence differences between clinical strain S1 and reference strain M129 that were consistent with RepMP1-mediated unique recombination, leading to sequence deletions and gene fusions. The resulting protein modifications could contribute to *M. pneumoniae* survival in the human respiratory tract.

MATERIALS AND METHODS

***M. pneumoniae* strains and DNA isolation.** *M. pneumoniae* reference strain M129 and clinical strain S1, which was isolated from an acute outbreak of *M. pneumoniae* respiratory infections in San Antonio in 1993, were grown in SP-4 medium at 37°C for 72 to 96 h in tissue culture flasks as described previously (37). Chromosomal DNA was isolated using an Easy DNA kit as recommended by the manufacturer (Invitrogen).

Microarray and hybridization. 70-mers representing 689 open reading frames of strain M129 were designed and synthesized by Qiagen Operon, Inc., and slides

were printed by Microarrays Inc. For hybridization, 4- μ g portions of purified chromosomal DNA of M129 and S1 were labeled using random hexamers and the aminoallyl method as recommended by The Institute for Genomic Research (TIGR) (<http://pfgc.tigr.org/protocols/M009.pdf>). Generated probes, M129 DNA labeled with Cy3, and S1 DNA labeled with Cy5 were hybridized to the same slide. After posthybridization washes, images were collected with a GenePix 4000B microarray scanner and analyzed by using Aquity Rev4.

PCR amplification of *M. pneumoniae* chromosomal regions. Primers specific for selected loci were designed based on the available genomic sequence of reference strain M129 (Table 2). All PCR amplifications were performed using Platinum *Taq* High Fidelity DNA polymerase (Invitrogen) at appropriate annealing temperatures and with necessary extension times. Amplified products were separated on 1% agarose, and their sizes and specificities were evaluated.

PCR-restriction fragment length polymorphism analysis of the strain S1-P1 gene sequence. *M. pneumoniae* strains M129 and S1 were typed by performing a PCR-restriction fragment length polymorphism analysis based on the sequence divergence of the adhesin P1-encoding gene (7). Primers ADH1 (5'-CTGCCTTGTCCTCAAGTCCACT-3'), ADH2 (5'-AACCTTGTCGGGAAGAGCTG-3'), ADH3 (5'-CGAGTTTGCTGCTAACGAGT-3'), and ADH4 (5'-CTTGACTGATACCTGTGCGG-3') were used to amplify the ADH1-2 (~2,280 bp) and ADH3-4 (~2,580 bp) regions, respectively. Amplification reactions were performed using 50- μ l mixtures containing 100 ng of chromosomal DNA and a Platinum *Taq* High Fidelity DNA polymerase (Invitrogen). Regions of the P1-encoding gene were amplified using 30 cycles consisting of 94°C for 30 s, 55°C for 30 s, and 68°C for 3 min and were visualized on 1% agarose. Amplified products were then subjected to complete restriction digestion with HaeIII, HhaI, HpaII, and RsaI, and the restriction patterns generated were resolved on 2% agarose for comparison and analysis.

Sequencing. PCR products that were selected for sequencing were cloned into the pCRII vector (Invitrogen), and plasmid DNA was isolated using a QIAprep Spin kit (Qiagen Sciences). Sequencing was performed at the Department of Microbiology and Immunology Nucleic Acids Facility (University of Texas Health Science Center at San Antonio). Sequences were analyzed using the BLAST program available in the NCBI database.

Southern analysis of chromosomal DNA. Five-microgram portions of chromosomal DNA from strains M129 and S1 were digested to completion with EcoRI or HindIII (New England Biolabs, Inc.). The fragments generated were separated on 1% agarose and blotted on a Zeta-Probe membrane by capillary transfer as suggested by the manufacturer (Bio-Rad Laboratories). For hybridization, an MPN130-specific 70-mer was end labeled using the Primer Extension System and [γ -³²P]dATP as recommended by the manufacturer (Promega). Other hybridization probes were generated by PCR amplification of selected regions using 100 ng of M129 chromosomal DNA, corresponding primers (Table 2), and Platinum *Taq* DNA polymerase in buffer with [α -³²P]dATP. Amplified products were purified using a QIAquick PCR purification column (Qiagen Sciences), equal volumes of formamide were added, and probes were denatured by boiling them for 5 min and instantly cooled on ice. Prehybridization was carried out at 42°C for 4 h in a solution containing 50% formamide, 0.12 M Na₂HPO₄ (pH 7.2), 0.25 M NaCl, 7% sodium dodecyl sulfate, and 1 mM EDTA. Hybridization was performed in the same solution at 42°C overnight. Membranes were washed as suggested for the Zeta-Probe membrane.

RNA isolation and reverse transcription-PCR analysis. Total RNA was isolated from cultures grown to mid-log phase using Tri Reagent as recommended by the manufacturer (Sigma) and was quantified by determining the absorbance at 260 nm. Two micrograms of isolated RNA was subjected to DNase I treatment (Invitrogen) and reverse transcribed using a mixture of gene-specific reverse primers (2 μ M each) and the SuperScript First-Strand synthesis system for reverse transcription-PCR (Invitrogen). Amplification was performed using Platinum *Taq* DNA polymerase (Invitrogen), and products were separated on agarose alongside corresponding markers.

Nucleotide sequence accession number. Sequences of clinical strain S1 generated in this study have been deposited in the GenBank database under accession number EF470909.

RESULTS

Microarray hybridization analysis revealed a distinct pattern for clinical strain S1. An oligonucleotide microarray comprised of 70-mers, which represented 689 reading frames identified in the genome sequence of reference strain M129, was hybridized with M129 and S1 chromosomal DNA. As ex-

TABLE 2. Primers used in comparisons of *M. pneumoniae* strains M129 and S1

Primer ^a	Sequence	Size of product (bp)	Other characteristics
PCR amplification of RepMP1-containing and other genes			
MPN037F MPN037R	CCTAATCAATATAGGCACTGTTGG GCAGTTGATCCTCGTTGACA	569	
MPN094F MPN094R	GCTGAACCTTAGTTGGCAGCA AATTGCTGGTATTCTTTATTTTGAA	526	
MPN100F MPN100R	GCTCAAGCTAAGCTCCCAA CGGTGGATGGCTTTTATTT	714	
MPN129F MPN131R	TGATGATGTTAGGATCATTAAATTGCATTG GCTGATTAGGCGTAGTTCGCGAACCA	2,289	Size of product in S1, 946 bp
MPN138BF MPN138BR	GATTGAAACTGAGTTAAAGAGTCAGGG ATGAAAGCCGTGGGATCACG	2,586	Size of product in S1, 1,626 bp
MPN139F MPN139R	GGTGTACTTGGCTATTCATTGGTG GTCGACTTTTCTTCTAATTTGTCC	1,381	
MPN140F MPN140R	GCGTCGTGGTGGTACTTGTGAAGTGTCC GTCGATCAACTTTGTCCTCAACAA	1,662	
MPN142F MPN142R	CCGCACAGGTATCAGTCAAG TACCCCATCACTGTCCAA	1,979	
MPN204F MPN204R	GCTCAAGCTAAGCTCCCAA CGGTGGATGGCTTTTATTT	607	
MPN283F MPN283R	AATGGTATTGATCCCGTTG AATTTTCTCGCCCTGCTTTT	483	
MPN368F MPN368R	GCTGAACCTTAGTTGGCAGCA TCCTTGATAGGATGGCTTTTT	715	
MPN410F MPN410R	ATACGCCTTTGCGGTGTTAC AAAAATACCAACAGCTTGTTTTAAAG	597	
MPN465F MPN465R	TTAGCCGTCATCTTTCACC GCTGAACTTAGTTGGCAGCA	724	
MPN484F MPN484R	CATCAAAAGCCTGAATCGAA GATAAATAAAAAATACCAACAACACTGACA	423	
MPN501F MPN501R	CGTTGCTAATGCAAGCTCAA CCTTGGACAGATGGCTTTTT	767	
MPN524F MPN524R	GCTGAACCTTAGTTGGCAGCA CCTTGGGCAGATGGTTTTT	682	Size of product in S1, 661 bp
MPN655F MPN655R	AGCTTGATCAGTGTGTTGCCTA TGCCAACAAAATTGCAAAAA	754	
Hybridization probe for Southern analysis of MPN138-MPN137 and MPN130 regions			
ECOdelF ECOdelR	GCTGGAAGTTAAGGTCGATAAA GAGTTAATTTGTTGATTGCTCACC	847	Sequence deleted in this region of S1
ECOcomF ECOcomR	TGAAAAACAGGGCGAGCAAAT AACAGAATCAAGGCGACCTTCTA	223	Common sequence in M129 and S1
130OLIGO	TCCATGACATGGTAAAAAGACCAAAAAGGGCACTT TTTCAGGTTGATACATCGTTAATCCAGAGAACAA		MPN130-specific 70-mer
HINDdelF HINDdelR	GATTAATCAGTAAAGAGTCTGATATCAG CTCCACTAAATAAATTGACCGGC	652	Sequence deleted in this region of S1
HIND129F HIND129R	CAAAGCATCGATTGTAGAAGC ATGTAGAGATAAATTGAGCTCACT	403	Coding region of MPN129; shares no significant homology within M129 genome
Transcriptional analysis of MPN139 and MPN138/7			
139RTPCRF 139RTPCRR	CTGAGTTAAAGAGTCAGGGTCAAAGGATTCAAGTTG ACCAAAATCAGTCAAAGTGTTCGATTGA	191	
1387RTPCRF 1387RTPCRR	TTCCGTTTTATAACGAAAAGGAATTAACGAG CAAGGCAACCTTCTATTGAATCCAAGCGATTTTCCA	477	

^a The same primers were used to amplify MPN100 and MPN204, and products were extracted from 1% agarose.

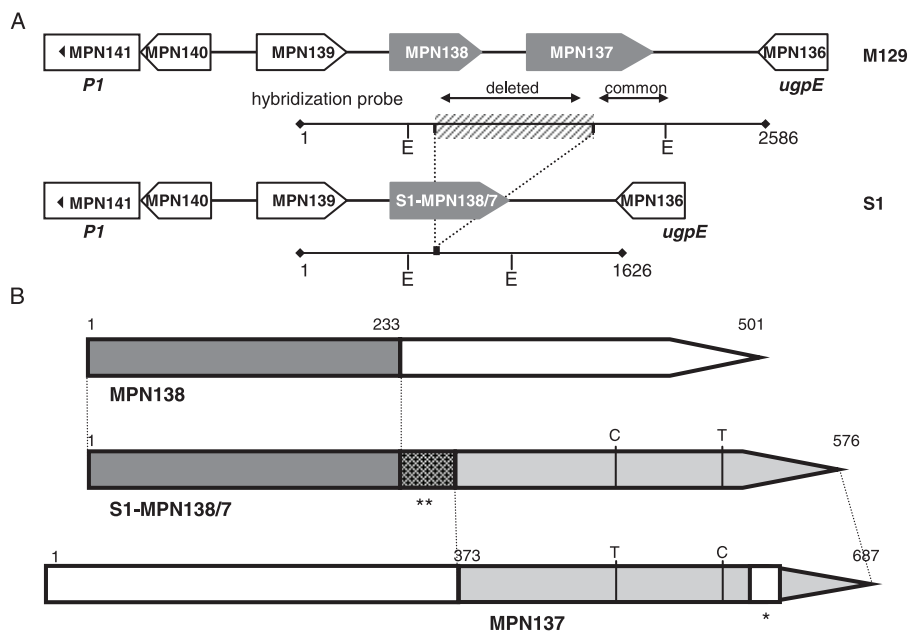


FIG. 1. (A) Comparison of MPN136-MPN141 regions in *M. pneumoniae* strains M129 and S1. The reference strain M129 region contains the MPN139, MPN138, and MPN137 loci. The three-gene operon containing the adhesin P1 gene (MPN141) is located upstream of the MPN139 locus, and the gene encoding a putative permease (*ugpE* or MPN136) is located immediately downstream of the MPN137 locus. The MPN138/7 fused reading frame (striped bar, deleted sequence) present in clinical strain S1 is surrounded by conserved sequences. The exact positions of PCR-amplified regions and their sizes are indicated. EcoRI restriction sites (E) are indicated together with the positions of two hybridization probes used in Southern analysis (designated “deleted” and “common”). (B) DNA sequence homologies between S1-MPN138/7 and M129-MPN137 and M129-MPN138 reading frames. The fused S1-MPN138/7 region contains the 5' end of MPN138 and the 3' end of MPN137. Two nucleotide transitions are indicated, as are 21 nucleotides missing in S1-MPN138/7 (one asterisk). The position of 49 novel nucleotides is indicated by two asterisks.

pected, all oligonucleotides hybridized with labeled M129 chromosomal DNA. In contrast, the hybridization pattern for S1 chromosomal DNA revealed no hybridization with oligonucleotides representing the MPN137 locus (nucleotides 178143 to 177457) and the MPN138 locus (nucleotides 178892 to 178392), which encode proteins with unknown functions.

The *mpn137* and *mpn138* genes are located on the minus strand and are preceded by the MPN139 locus on the same strand (Fig. 1A). These three genes are separated by intergenic regions (236 and 248 nucleotides, respectively). The closest downstream gene is located on the plus strand and encodes a putative permease protein (*ugpE* or MPN136). Both MPN136- and MPN139-specific oligonucleotides hybridized with S1 chromosomal DNA using the microarray. Primers corresponding to the region from nucleotide 176838 to nucleotide 179381 (MPN138BF and MPN138BR [Table 2]) were designed to amplify a 2,544-bp product encompassing the entire MPN137 and MPN138 loci, together with partial regions of MPN136 and MPN139 (Fig. 1). The amplicon generated for reference strain M129 was the expected size (~2.5 kb), whereas the PCR product generated for clinical strain S1 was ~1 kb shorter. The sequence amplified from M129 matched the chromosomal region reported by TIGR; that is, the complete coding regions of MPN137 (687 bp) and MPN138 (501 bp) were flanked by 42 bp of MPN136 and 253 bp of MPN139, and all coding regions were separated by corresponding intergenic sequences. The sequence generated for S1 (1,626 nucleotides) revealed conserved sequences of the MPN136 and MPN139 loci and connected intergenic sequences. Additionally, nucleotides 1 to 233

of MPN138 were completely conserved (Fig. 1B); similarly, nucleotides 373 to 687 of MPN137 were highly conserved except for two nucleotide transitions (T⁴⁹³ to C and C⁵⁷⁵ to T) and deletion of 21 nucleotides (at positions 595 to 615). Importantly, a stretch of 888 nucleotides between MPN138 and MPN137 partial sequences (nucleotides 177771 to 178660) was replaced by 49 nucleotides not originally detected in this region of reference strain M129. As a result, the fusion of two reading frames (M129-MPN138 and M129-MPN137) yielded a new 576-bp gene (S1-MPN138/7) (Fig. 1B).

Closer analysis of nucleotide sequences revealed RepMP1 repeats and their involvement in recombination. Previously, a copy of the RepMP1 repeat was identified within MPN139 (nucleotides 179373 to 179647) (16, 43). Interestingly, sequence analysis of the amplified region revealed the presence of two previously unidentified core elements of RepMP1 in strain M129 at nucleotides 178646 to 178920 and 177876 to 178163, located within the 5' end and adjacent upstream sequences of MPN138 and MPN137, respectively (Fig. 2). Originally, RepMP1 was described as a 300-bp element (43), but later this core sequence was found to be flanked by short repeated sequences designated A, B, and C. These short sequences do not exist independent of the core RepMP1 element and are arranged in various combinations, forming a mosaic pattern (10). We also detected these RepMP1-related short repeated sequences flanking two newly identified core elements (Fig. 2). In the fused MPN138/7 gene of clinical strain S1, the region between MPN138 RepMP1 and MPN137 RepMP1 (including the entire MPN137 RepMP1 sequence

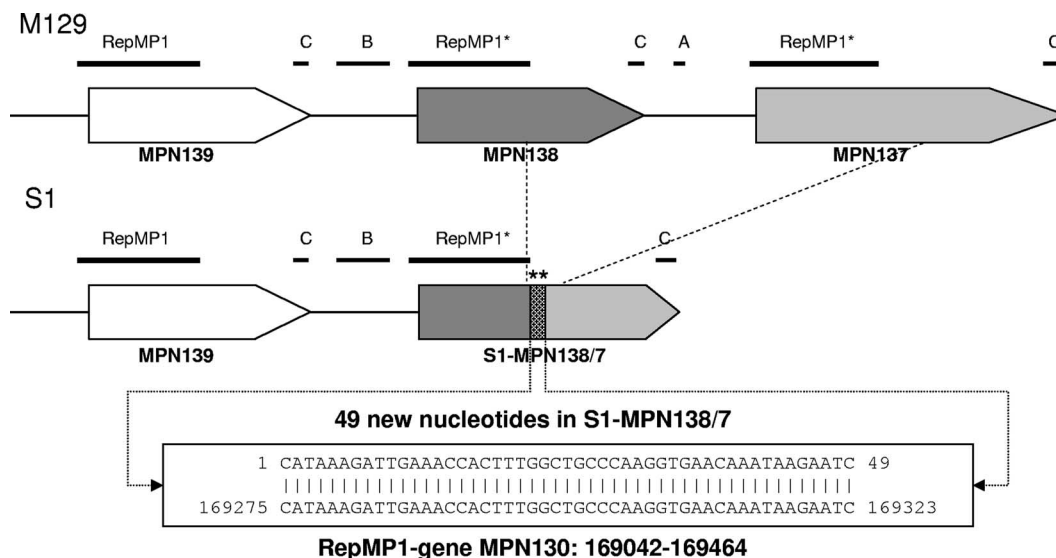


FIG. 2. Comparison of RepMP1 elements in the MPN139, MPN138, and MPN137 regions in strains M129 and S1. For M129, the positions of all three RepMP1 core sequences are indicated (asterisks indicate repeats identified in this study). Short repeats (A, B, and C) accompanying RepMP1 core elements are also indicated. For S1, only two copies of the RepMP1 core repeat element were found. The deleted region is indicated by dotted lines. The MPN138/7 fused reading frame contains a proximal region of MPN138, a distal region of MPN137, and a central region consisting of 49 nucleotides previously not found in either locus (two asterisks). BLAST analysis revealed a perfect match with the RepMP1 core element of the MPN130 locus.

and only the last 14 nucleotides of MPN138 RepMP1) was missing and replaced by a stretch consisting of 49 novel nucleotides (Fig. 1B and 2). To determine the origin of these 49 nucleotides, we performed a BLAST analysis with the available genome sequence of *M. pneumoniae* M129, which revealed homologies with several other genes that all belonged to the RepMP1 family (Table 1).

To further characterize the RepMP1 family, we amplified all 14 remaining RepMP1-containing genes of clinical strain S1. The amplified products were sequenced and compared with reference strain M129 sequences. As shown in Table 3, most of

the S1 RepMP1 genes exhibited only minor sequence differences; the only exceptions were MPN524 and MPN130. S1 MPN524 showed overall high homology with the homologous gene in M129 (only two nucleotide transversions were identified, T²⁶⁵ to G and G³⁶⁶ to T) but had lost 21 nucleotides (nucleotides 395 to 415 of M129-MPN524). We observed the most significant difference when the MPN130 region was analyzed. Primers were designed to amplify a 2,289-bp region (nucleotides 167990 to 170278) (Fig. 3). When the sequences were compared using agarose, the amplicon that originated from S1 was ~700 bp smaller than the PCR product generated

TABLE 3. Sequence variations detected in coding regions of RepMP1-containing genes in clinical strain S1

Locus	Nucleotide difference(s) ^a	Amino acid difference(s)
MPN037	1 (A ¹⁴⁵ to C)	1 (T ⁴⁹ to P)
MPN094	1 (G ³³¹ to A)	1 (E ¹¹¹ to K)
MPN100	2 (C ¹³² to T*, G ⁴³⁴ to T)	1 (G ¹⁴⁵ to V)
MPN130	Missing	
MPN137	Distal portion of MPN138/7 fused gene (nucleotides 373 to 594 and 616 to 687 remain); 2 (T ⁴⁹³ to C*, C ⁵⁷⁵ to T)	Amino acids 125 to 228 remain; amino acids 203 to 209 are missing; 1 (R ¹⁹² to C ¹⁶²)
MPN138	Proximal portion of MPN138/7 (nucleotides 1 to 233 remain)	Amino acids 1 to 78 remain
MPN139	1 (C ⁴⁰² to T*)	0
MPN204	1 (A ¹⁶⁶ to C)	1 (T ⁵⁶ to P)
MPN283	1 (G ³¹⁰ to A)	1 (E ¹⁰⁴ to K)
MPN368	3 (G ²⁵³ to A*, G ²⁸⁶ to A, G ⁴⁴⁰ to A)	2 (E ⁹⁶ to K, G ¹⁴⁷ to D)
MPN410	1 (A ⁵⁷ to G*)	0
MPN465	0	0
MPN484	0	0
MPN501	0	0
MPN524	2 (T ²⁶⁵ to G, G ³⁶⁶ to T); nucleotides 395 to 415 are missing	2 (V ⁸⁹ to K, K ¹²² to N); amino acids 132 to 138 are missing
MPN655	6 (T ¹¹⁴ to C*, C ²⁰³ to A, C ²⁰⁸ to A, C ²¹⁰ to T, G ²¹⁹ to C*, A ⁵¹⁷ to G)	3 (L ⁶⁸ to P, P ⁷⁰ to T, S ¹⁷³ to G)

^a An asterisk indicates a silent mutation.

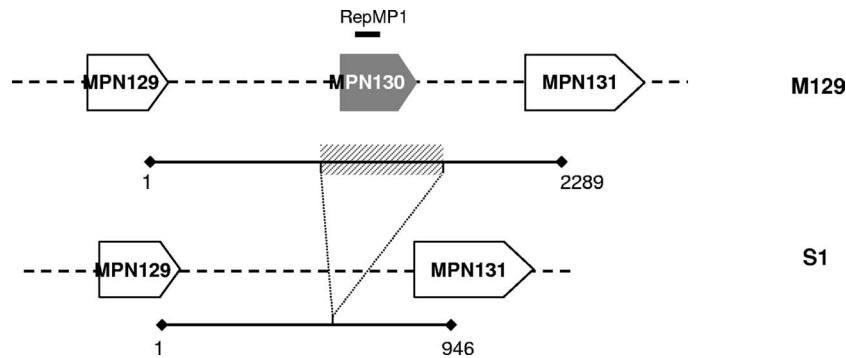


FIG. 3. Characterization of PCR-amplified MPN129-MPN131 regions in *M. pneumoniae* strains M129 and S1. The position of the RepMP1 core element within M129-MPN130 is indicated. In clinical strain S1, the MPN130 locus is deleted (the 680-bp missing region is indicated by a striped bar). The exact positions of PCR-amplified regions and their sizes are indicated.

from M129. Sequencing revealed a 680-bp deletion encompassing the entire MPN130 locus (423 bp), 116 adjacent nucleotides of the upstream region, and 141 adjacent nucleotides of the downstream region. Most importantly, among the RepMP1-containing genes, MPN130 was the only gene that showed 100% identity with the 49 nucleotides of S1-MPN138/7 in a BLAST analysis.

Southern analysis confirmed deletion within the MPN138-MPN137 and MPN130 regions. Using the sequence data, we set out to determine whether the missing regions were translocated elsewhere within the genome of S1 or were entirely deleted. Analysis of the M129 DNA sequence revealed that the EcoRI restriction sites closest to the MPN137-MPN138 missing region should produce a 1,435-bp fragment, while in the case of S1 this fragment was predicted to be only 575 bp long (Fig. 1). Chromosomal DNAs of M129 and S1 were individually digested with EcoRI and blotted (Fig. 4). We amplified a 847-bp radiolabeled probe representing the missing region in S1-MPN138/7 (labeled "deleted" in Fig. 1). Hybridization under stringent conditions revealed the presence of a ~1.4-kb fragment in M129. However, a lack of hybridization with S1 chromosomal DNA indicated the absence of this region (Fig. 4B). Additional hybridization of the same membrane was performed using a 223-bp probe (labeled "common" in Fig. 1) representing the sequence within the same EcoRI fragment that is shared by both M129 and S1. This probe hybridized to the same ~1.4-kb fragment in M129 and to the predicted ~580-bp fragment in S1 (data not shown). An additional fragment hybridized under the same conditions with the second probe, which is not surprising because this "common" sequence shares homology with numerous other regions of the *M. pneumoniae* chromosome.

Southern analysis targeting the MPN130 region was designed to test for the presence of the MPN130 coding region within the S1 genome. Based on the available sequence, HindIII restriction of M129 chromosomal DNA should produce a 3,932-bp MPN130-containing fragment, while restriction of S1 chromosomal DNA should yield a 3,252-bp fragment (a smaller restriction fragment because of deletion of a 680-bp region). Southern analysis with an MPN130 oligomer indeed revealed a ~4-kb band in the M129 hybridization pattern and the absence of this band in the S1 hybridization pattern. Additionally, ~7.5-kb band hybridized with the 70-mer in both

strains. This band corresponded to the predicted fragment containing the MPN465 gene (the locus contained a nucleotide stretch that shared 97% homology with the MPN130 oligomer) and very possibly to the same gene of S1. This result explained why MPN130 was detected in the microarray hybridization pattern of S1. Additional hybridization of HindIII-restricted chromosomal DNA with the entire missing region revealed the same ~4-kb fragment in M129 and the absence of a strongly hybridizing fragment in S1, as well as several weaker hybridizing bands in both strains (data not shown). This was expected as the MPN130 region shares homology with numerous other RepMP1 genes. Lastly, when the coding region of MPN129 (Table 2), which was predicted to share no significant homology with other regions of the M129 chromosome, was ampli-

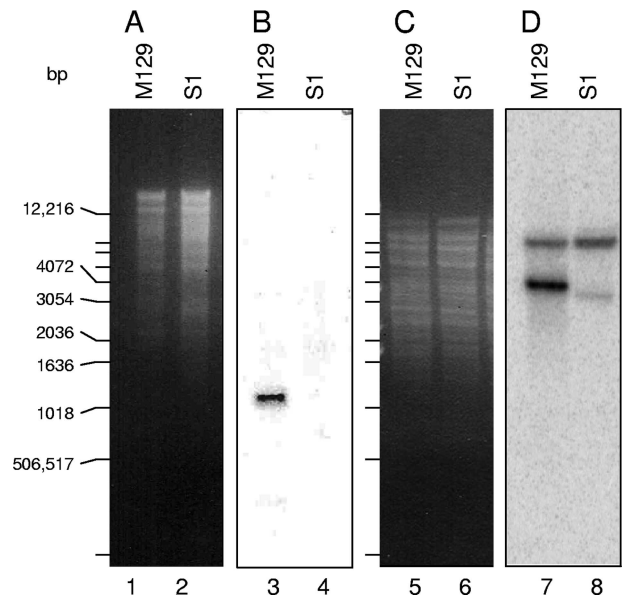


FIG. 4. Southern analysis of strains M129 and S1. Chromosomal DNA isolated from M129 (lanes 1, 3, 5, and 7) and S1 (lanes 2, 4, 6, and 8) was digested to completion with EcoRI and HindIII, and the fragments generated were separated on 1% agarose gels (A and C). After the restricted DNA was blotted on membranes, Southern hybridization was performed using the deleted portion of the MPN138/7 region as a probe (B), as well as a ^{32}P -labeled MPN130 oligomer (D). A 1-kb DNA ladder was used to estimate the sizes of fragments.

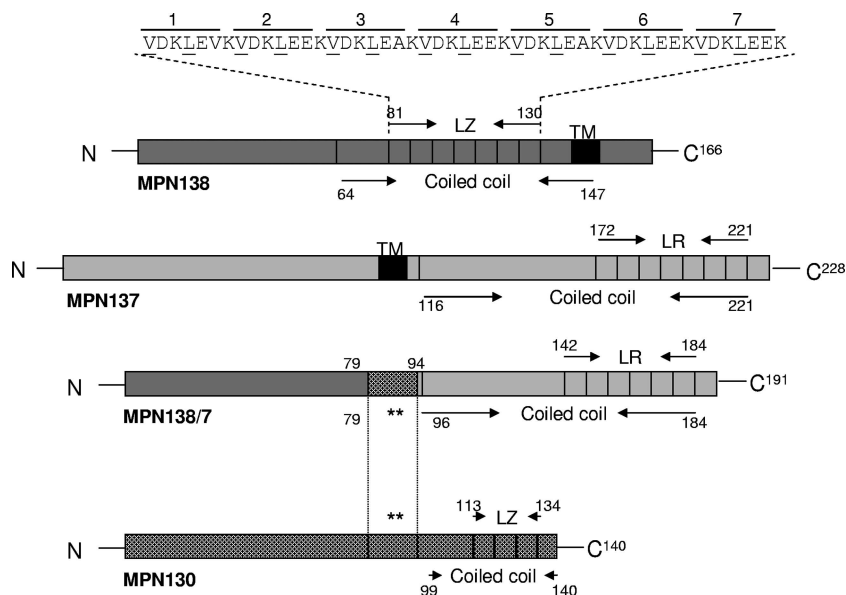


FIG. 5. Predicted secondary structure and leucine-containing tandem repeats in modified proteins encoded by genes containing RepMP1. Regions of coiled coils within the MPN138, MPN137, MPN138/7, and MPN130 proteins are indicated. The positions of LZs in the MPN138 and MPN130 proteins are also indicated. In the LZ motif in the MPN138 protein the positions of leucine and hydrophobic residues (both underlined) are indicated. The locations of LRs in the MPN138 and MPN138/7 proteins are indicated. The positions of transmembrane domains (TM) predicted by dense alignment surface analysis in the MPN137 and MPN138 proteins are also shown. As shown, only the indicated 16-amino-acid region of the MPN130 protein (two asterisks) is retained in S1.

fied and hybridized with HindIII-restricted DNA, an expected single restriction fragment, at ~4 and ~3.5 kb, was detected in the M129 and S1 hybridization patterns, respectively (data not shown). The hybridization patterns confirmed that 680 bp was deleted from the S1 genome.

Fused MPN138-MPN137 region is transcribed in clinical strain S1. Transcription of S1-MPN138/7 was confirmed by reverse transcription-PCR. Interestingly, the two genes are separated by a rather long intergenic region, and a search for potential transcription attenuators in the sequences preceding all coding regions revealed the presence of a terminatorlike structure upstream of S1-MPN138/7 (position -189) (24). Nevertheless, primer extension did not yield any transcriptional start upstream of MPN138/7, which together with the generation of a bicistronic *mpn139-mpn138/7* transcript (971 bp containing 236 bp of intergenic sequence) demonstrated that MPN139 and MPN138/7 were cotranscribed (data not shown).

Analysis of amino acid sequences encoded by MPN137, MPN138, and MPN138/7 revealed the presence of direct tandem repeats and a leucine zipper (LZ). Proteins encoded by RepMP1-containing genes are small, ranging from 103 amino acids long (MPN484 protein) to 228 amino acids long (MPN137 protein) (molecular masses, 11.9 and 26.8 kDa, respectively). Based on sequence analyses, the MPN138/7 fused protein has an N-terminal 78-amino-acid region of the MPN138 protein (Fig. 5), and amino acids 95 to 191 of the MPN138/7 protein are almost identical to amino acids 125 to 228 of the MPN137 protein, except for a seven-amino-acid deletion from amino acid 190 to amino acid 196 (VEGRLLDS) and substitution of C for R at position 199 (Fig. 5). MPN138/7 amino acids 79 to 94 were derived from the MPN130 protein

and exactly match regions found in two other RepMP1-containing proteins (MPN094 and MPN524). Interestingly, a dense alignment surface analysis (6) revealed the absence of transmembrane regions in the MPN138/7 protein that were observed in both the MPN138 protein (amino acids 140 to 149) and the MPN137 protein (amino acids 102 to 111) (Fig. 5).

Leucine repeat (LR) motifs were identified in the deduced MPN137, MPN138, and MPN138/7 amino acid sequences. The MPN137 and MPN138 proteins each have seven LRs, and deletion of a seven-amino-acid stretch resulted in six LRs in the MPN138/7 protein (Fig. 5). Computer analysis revealed that the C-terminal region of all three proteins assumed a predominant alpha-helical structure with an LZ-like motif distributed with, on average, six intervening amino acid residues in the helices (Fig. 5). This LR structure resembles the LZ motif that consists of heptad repeats of leucine residues over a span of 20 to 30 amino acids (23). Further computer analysis revealed only one LZ in the MPN138 protein (positions 81 to 130) (Fig. 5). Sequence analysis also identified an LZ motif in the MPN130 protein (positions 113 to 134) that is missing in S1 (Fig. 5).

P1 operon repetitive elements of clinical strain S1 are type 2 specific. Next we evaluated clinical isolate S1 to determine sequence variations within RepMP2/3, RepMP4, and RepMP5 in the P1 operon. We amplified P1 genes (MPN141) from both strains using primers ADH1 and ADH4 (5) and sequenced both products. The sequence generated for S1 was 33 nucleotides longer than the corresponding sequence of M129.

Next the S1-MPN141 sequence was aligned with available sequences of other type 2 clinical strains, including strains Mac (accession number AF290001), Mp1842 (accession number AF290002), FH, and TW7-5 (34). The sequences compared

exhibited high degrees of identity, including regions of RepMP2/3 and RepMP4. Based on the alignments, we identified six unique nucleotides in the S1-MPN141 gene: G⁵³² (transition of A), G⁶³² (transition of A), T²⁵⁴² (transition of C), GA⁴⁸³⁹ (transition of AG), and G⁴⁸⁵⁰ (transition of A). Further, five unique S1 amino acid mutations were predicted (E to G²¹¹; S to L⁸⁴⁵; S to D¹⁶¹⁵; and at positions 965 and 966, which are unique in all available type 2 P1 sequences, methionine and aspartic acid, respectively).

Another individual variation was insertion of nine nucleotides (three AGT triplets) after T¹⁹⁵¹. This insertion of three AGT triplets resulted in three additional serine residues added to a seven-serine stretch identified in all type 2 strains analyzed here, as well as in the M129 reference strain (type 1). A similar stretch of 10 serine residues was previously observed only in the protein encoded by the P1 gene of type 1 strain Mp4817 (accession number AF290000).

Primers MPN142F (complementary to ADH4) and MPN142R (Table 2) were used for PCR amplification of the MPN142 (ORF6) proximal region containing RepMP5. The amplicon generated for S1 appeared to be ~200 bp smaller when it was run alongside the M129 product on a 1% agarose gel. Sequencing confirmed that the PCR product generated from S1 was 201 bp shorter than the product amplified from M129 and completely matched the corresponding region of type 2 isolate FH (accession number Z33395) and previously published data on the size of the RepMP5 region in numerous type 2 strains (9, 30).

High level of conservation was observed for remaining reading frames of the region analyzed. We sequenced a 9,721-nucleotide region for S1 (in M129, this region corresponds to 10,747 nucleotides from nucleotide 176838 to nucleotide 187584). Within this region we identified a 680-bp deletion that resulted in the loss of MPN130. As a result of another detected deletion, a new reading frame (fused MPN138/7 loci) was identified in S1. Other than these described rearrangements, the sequenced regions (the MPN129, MPN131-MPN136, and MPN139-MPN142 loci and corresponding intergenic regions) were highly conserved, and only minor DNA variations were detected.

DISCUSSION

Like other microorganisms, mycoplasmas have the ability to alter their surface molecules by recombination, which is mediated by multiple copies of specific RepMPs. Previous studies by us (33–36) and other workers (4, 7, 8, 20, 31) demonstrated the presence of one copy of three distinct RepMPs within the cytoadhesin P1 operon of *M. pneumoniae*. Since protein P1 is a major adhesin that also stimulates a strong host immune response, sequence variations resulting from RepMP-mediated recombination could contribute to *M. pneumoniae* disease pathogenesis. Restriction fragment patterns associated with P1 gene sequence heterogeneity led to the classification of *M. pneumoniae* strains into two distinct groups (type 1 and type 2) (35). Analysis of numerous clinical strains confirmed that these two groups occur worldwide, regardless of their geographic origin (4, 5, 32, 33, 35, 36). As more clinical strains were analyzed, a rather low rate of sequence variation within the P1 gene was observed, although RepMP2/3- and RepMP4-medi-

ated variations were reported for individual isolates (7–9, 20). However, none of these studies revealed the exact mechanism of RepMP-mediated recombination.

Sequence analysis of the S1-P1 gene (MPN141) and the proximal region of MPN142 (ORF6) revealed *M. pneumoniae* type 2-specific sequences. Minor sequence divergence within the S1 MPN141 gene translated into five amino acids unique to this strain. Additionally, although S1-MPN141 is highly homologous to P1 genes of other type 2 strains, it encodes a stretch of 10 serine residues previously observed only in type 1 *M. pneumoniae* strain Mp4817 (accession number AF290000). The occurrence of short tandem repeats (for example, AGT), believed to be responsible for sequence variability through slipped-strain mispairing, is surprisingly limited in *M. pneumoniae* (38), and such repeats were only found within the MPN141 gene.

Despite only minor sequence divergence within the adhesin P1 operon of S1, comparisons of the complete genome sequences by oligonucleotide array analysis identified new inter-strain variations involving the MPN130, MPN137, and MPN138 loci. Using sequence analysis, we discovered an 888-nucleotide deletion and replacement of these nucleotides with 49 nucleotides in this region, which originated from MPN130, leading to generation of a new fused reading frame, S1-MPN138/7. An explanation of the origin of these rearrangements became possible when we identified two previously unrecognized RepMP1 elements (Fig. 2).

Until now, there has been no direct evidence for sequence variation involving the RepMP1 repeat, although, based on the presence of numerous repeats, the possibility of recombination has been suggested previously (10). By comparing two *M. pneumoniae* strains, we discovered gene sequence divergence and amino acid modifications that resulted from such RepMP1-related rearrangements. In this process, large DNA fragments were lost, and a novel fused coding region (S1-MPN138/7) that contains portions of three genes (MPN138, MPN137, and MPN130) was formed.

A direct consequence of the rearrangement demonstrated is the alteration of protein patterns. While reference strain M129 possesses MPN130 (16.5 kDa), MPN138 (19.5 kDa), and MPN137 (26.8 kDa) proteins, clinical isolate S1 has a single fused protein (predicted molecular mass, 22.5 kDa). Our findings agree with size variations of RepMP1 proteins observed in Western blot analyses of different strains with specific anti-RepMP1 antibodies (R. Herrmann, personal communication). The MPN137 and MPN138 proteins were recently identified in a total protein extract of *M. pneumoniae* strain FH by proteome analysis (19). We demonstrated transcription of S1-MPN138/7, and transcription of MPN130, MPN138, and MPN137 in M129 strain was described previously (42). The rearrangements described above also have significant effects on the secondary structure of proteins that lead to predictable alterations in structure and function properties. For example, unlike the MPN138 and MPN137 proteins, no transmembrane domain is predicted in the protein encoded by the fused MPN138/7 loci.

The existence of repetitive elements as portions of the coding regions of structural genes enables generation of constant and variable antigenic domains. Intergenic recombination-mediated protein domain variations have been reported previ-

ously for the Arp protein (an immunoglobulin A receptor in the M protein family) of *Streptococcus pyogenes*. The recombination event resulted in novel N-terminal sequences and retention of constant signal and C-terminal sequences (14). Interestingly, comparisons of proteins encoded by genes containing RepMP1 revealed that the first 78 amino acids of the MPN138 protein shared high homology with N-terminal regions of several other members of this protein family (MPN094, MPN100, MPN130, MPN139, MPN204, MPN368, MPN410, and MPN501 proteins). As we demonstrate here, recombination created linkages between the N-terminal region (78 amino acids of the MPN138 protein) and the C-terminal region of another protein (MPN137 protein) through a 16-amino-acid stretch originating from yet a third protein (MPN130 protein). Fusions of two genes resulting in proteins that contained novel combinations of functional domains were described previously. In many cases in eukaryotes, these translocations involve transcriptional factors (1, 11). However, in *M. pneumoniae* the functional domains of the proteins encoded by genes containing RepMP1 involved in recombination have not been characterized yet.

We identified an LR structure that resembles the LZ motif (18, 23, 25) within the deduced MPN130, MPN137, MPN138, and S1-MPN138/7 amino acid sequences. Also, analysis identified motifs present in the MPN130 and MPN138 proteins as LZs. Initially, LZs were proposed to be characteristic of DNA-binding proteins (12, 22, 23), and this view has been extended to other proteins that form oligomers (3). In addition, LZ-harboring proteins are involved in regulation and cytoskeleton formation (21).

In *M. pneumoniae*, LZs have previously been observed in high-molecular-weight proteins like cytoskeletal protein HMW2 (MPN310) (21) and the MPN474 protein (27). Although most of the RepMP1-containing were predicted to result in coiled-coil structures, none of them (including the proteins with the LZ motif) has been identified in the *M. pneumoniae* cytoskeletal fraction, where other high-molecular-weight LZ proteins (the MPN310 and MPN474 proteins) were found (21, 27, 28).

Recently, a mutant with a mutation in the RepMP1/DUF16 MPN524 protein (a protein which shares homology with the fused region of the MPN138/7 protein) was reported to have a satellite growth-altered phenotype and reduced glass binding (13). Like most of the proteins encoded by genes containing RepMP1, the MPN524 protein is also predicted to form a coiled-coil structure in its C-terminal region. The lack of known regulatory elements in *M. pneumoniae* and the absence of these small LZ proteins in the cytoskeletal fraction (28) suggest that these proteins could be involved in regulation of cytoskeletal or membrane proteins directly or indirectly.

The fusion and generation of new repeats in the MPN138/7 protein without the putative transmembrane domains could help the organism overcome the host immune response against proteins, as observed for the *Mycoplasma hominis* surface-variable, LZ-containing p50 protein (15). Variations in the number of direct repeats can also influence antibody recognition. Cytoskeleton-associated protein P65 (MPN309) was found to contain short direct repeats, and comparisons of M129 P65 and FH P65 revealed differences in the number of five-amino-acid repeats (26). Similarly, we observed seven-

amino-acid deletions in the S1-MPN138/7 and S1-MPN524 proteins.

Since at least 16 of the proteins encoded by genes containing RepMP1 (including the MPN137 and MPN138 proteins) have been found in *M. pneumoniae*, the existence of other RepMP1-mediated variations in different strains appears to be very likely. Further analysis of proteins encoded by genes containing RepMP1 and their variations could reveal how this protein family affects the biology and virulence of *M. pneumoniae*, providing additional insights into how these smallest prokaryotes adapt to and survive in the hostile host environment.

ACKNOWLEDGMENTS

We thank Daniele Provenzano for his involvement in establishing the microarray and Richard Herrmann for sharing his unpublished results and his comments and suggestions. We thank Rose Garza for her assistance with finalizing the manuscript.

This work was supported by National Institutes of Health grants AI045737 and AI070412.

REFERENCES

- Barr, F. G., N. Galili, J. Holick, J. A. Biegel, G. Rovera, and B. S. Emanuel. 1993. Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. *Nat. Genet.* **3**:113–117.
- Baseman, J. B., and J. G. Tully. 1997. Mycoplasmas: sophisticated, reemerging, and burdened by their notoriety. *Emerg. Infect. Dis.* **3**:21–32.
- Buckland, R., and F. Wild. 1989. Leucine zipper motif extends. *Nature* **338**:547.
- Cousin, A., B. de Barbeyrac, A. Charron, H. Renadin, and C. Bebear. 1994. Analysis of RFLPs of amplified cytohesin P1 gene for epidemiological study of *Mycoplasma pneumoniae*. *IOM Lett.* **3**:494–495.
- Cousin-Allery, A., A. Charron, B. de Barbeyrac, G. Fremy, J. Skov Jensen, H. Renadin, and C. Bebear. 2000. Molecular typing of *Mycoplasma pneumoniae* strains by PCR-based methods and pulsed-field gel electrophoresis. Application to French and Danish isolates. *Epidemiol. Infect.* **124**:103–111.
- Cserzo, M., E. Wallin, I. Simon, G. von Heijne, and A. Elofsson. 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.* **10**:673–676.
- Dorigo-Zetsma, J. W., J. Dankert, and S. A. Zaat. 2000. Genotyping of *Mycoplasma pneumoniae* clinical isolates reveals eight P1 subtypes within two genomic groups. *J. Clin. Microbiol.* **38**:965–970.
- Dorigo-Zetsma, J. W., B. Wilbrink, J. Dankert, and S. A. Zaat. 2001. *Mycoplasma pneumoniae* P1 type 1- and type 2-specific sequences within the P1 cytohesin gene of individual strains. *Infect. Immun.* **69**:5612–5618.
- Dumke, R., I. Catrein, E. Pirkil, R. Herrmann, and E. Jacobs. 2003. Subtyping of *Mycoplasma pneumoniae* isolates based on extended genome sequencing and on expression profiles. *Int. J. Med. Microbiol.* **292**:513–525.
- Forsyth, M. H., and S. J. Geary. 1996. The repetitive element Rep MP 1 of *Mycoplasma pneumoniae* exists as a core element within a larger, variable repetitive mosaic. *J. Bacteriol.* **178**:917–921.
- Galili, N., R. J. Davis, W. J. Fredericks, S. Mukhopadhyay, F. J. Rauscher III, B. S. Emanuel, G. Rovera, and F. G. Barr. 1993. Fusion of a fork head domain gene to PAX3 in the solid tumour alveolar rhabdomyosarcoma. *Nat. Genet.* **5**:230–235.
- Gentz, R., F. J. Rauscher III, C. Abate, and T. Curran. 1989. Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. *Science* **243**:1695–1699.
- Hasselbring, B. M., C. A. Page, E. S. Sheppard, and D. C. Krause. 2006. Transposon mutagenesis identifies genes associated with *Mycoplasma pneumoniae* gliding motility. *J. Bacteriol.* **188**:6335–6345.
- Heden, L. O., and G. Lindahl. 1993. Conserved and variable regions in protein Arp, the IgA receptor of *Streptococcus pyogenes*. *J. Gen. Microbiol.* **139**:2067–2074.
- Henrich, B., A. Kitzerow, R. C. Feldmann, H. Schaal, and U. Hadding. 1996. Repetitive elements of the *Mycoplasma hominis* adhesin p50 can be differentiated by monoclonal antibodies. *Infect. Immun.* **64**:4027–4034.
- Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkil, B. C. Li, and R. Herrmann. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**:4420–4449.
- Himmelreich, R., H. Plagens, H. Hilbert, B. Reiner, and R. Herrmann. 1997. Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res.* **25**:701–712.
- Hu, J. C., E. K. O'Shea, P. S. Kim, and R. T. Sauer. 1990. Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science* **250**:1400–1403.

19. Jaffe, J. D., H. C. Berg, and G. M. Church. 2004. Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* **4**:59–77.
20. Kenri, T., R. Taniguchi, Y. Sasaki, N. Okazaki, M. Narita, K. Izumikawa, M. Umetsu, and T. Sasaki. 1999. Identification of a new variable sequence in the P1 cytoadhesin gene of *Mycoplasma pneumoniae*: evidence for the generation of antigenic variation by DNA recombination between repetitive sequences. *Infect. Immun.* **67**:4557–4562.
21. Krause, D. C. 1996. *Mycoplasma pneumoniae* cytoadherence: unravelling the tie that binds. *Mol. Microbiol.* **20**:247–253.
22. Landschulz, W. H., P. F. Johnson, and S. L. McKnight. 1989. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science* **243**:1681–1688.
23. Landschulz, W. H., P. F. Johnson, and S. L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**:1759–1764.
24. Merino, E., and C. Yanofsky. 2005. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet.* **21**:260–264.
25. O'Shea, E. K., R. Rutkowski, and P. S. Kim. 1989. Evidence that the leucine zipper is a coiled coil. *Science* **243**:538–542.
26. Proft, T., H. Hilbert, G. Layh-Schmitt, and R. Herrmann. 1995. The proline-rich P65 protein of *Mycoplasma pneumoniae* is a component of the Triton X-100-insoluble fraction and exhibits size polymorphism in the strains M129 and FH. *J. Bacteriol.* **177**:3370–3378.
27. Regula, J. T., G. Boguth, A. Gorg, J. Hegermann, F. Mayer, R. Frank, and R. Herrmann. 2001. Defining the mycoplasma 'cytoskeleton': the protein composition of the Triton X-100 insoluble fraction of the bacterium *Mycoplasma pneumoniae* determined by 2-D gel electrophoresis and mass spectrometry. *Microbiology* **147**:1045–1057.
28. Regula, J. T., B. Ueberle, G. Boguth, A. Gorg, M. Schnolzer, R. Herrmann, and R. Frank. 2000. Towards a two-dimensional proteome map of *Mycoplasma pneumoniae*. *Electrophoresis* **21**:3765–3780.
29. Rocha, E. P., and A. Blanchard. 2002. Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution. *Nucleic Acids Res.* **30**:2031–2042.
30. Ruland, K., R. Himmelreich, and R. Herrmann. 1994. Sequence divergence in the ORF6 gene of *Mycoplasma pneumoniae*. *J. Bacteriol.* **176**:5202–5209.
31. Ruland, K., R. Wenzel, and R. Herrmann. 1990. Analysis of three different repeated DNA elements present in the P1 operon of *Mycoplasma pneumoniae*: size, number and distribution on the genome. *Nucleic Acids Res.* **18**:6311–6317.
32. Sasaki, T., T. Kenri, N. Okazaki, M. Iseki, R. Yamashita, M. Shintani, Y. Sasaki, and M. Yayoshi. 1996. Epidemiological study of *Mycoplasma pneumoniae* infections in Japan based on PCR-restriction fragment length polymorphism of the P1 cytoadhesin gene. *J. Clin. Microbiol.* **34**:447–449.
33. Su, C. J., A. Chavoya, and J. B. Baseman. 1988. Regions of *Mycoplasma pneumoniae* cytoadhesin P1 structural gene exist as multiple copies. *Infect. Immun.* **56**:3157–3161.
34. Su, C. J., A. Chavoya, S. F. Dallo, and J. B. Baseman. 1990. Sequence divergence of the cytoadhesin gene of *Mycoplasma pneumoniae*. *Infect. Immun.* **58**:2669–2674.
35. Su, C. J., S. F. Dallo, and J. B. Baseman. 1990. Molecular distinctions among clinical isolates of *Mycoplasma pneumoniae*. *J. Clin. Microbiol.* **28**:1538–1540.
36. Su, C. J., S. F. Dallo, A. Chavoya, and J. B. Baseman. 1993. Possible origin of sequence divergence in the P1 cytoadhesin gene of *Mycoplasma pneumoniae*. *Infect. Immun.* **61**:816–822.
37. Su, C. J., V. V. Tryon, and J. B. Baseman. 1987. Cloning and sequence analysis of cytoadhesin P1 gene from *Mycoplasma pneumoniae*. *Infect. Immun.* **55**:3023–3029.
38. van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh. 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* **62**:275–293.
39. Waites, K. B., and D. F. Talkington. 2004. *Mycoplasma pneumoniae* and its role as a human pathogen. *Clin. Microbiol. Rev.* **17**:697–728.
40. Waldo, R. H., III, J. L. Jordan, and D. C. Krause. 2005. Identification and complementation of a mutation associated with loss of *Mycoplasma pneumoniae* virulence-specific proteins B and C. *J. Bacteriol.* **187**:747–751.
41. Waldo, R. H., III, and D. C. Krause. 2006. Synthesis, stability, and function of cytoadhesin P1 and accessory protein B/C complex of *Mycoplasma pneumoniae*. *J. Bacteriol.* **188**:569–575.
42. Weiner, J., III, R. Herrmann, and G. F. Browning. 2000. Transcription in *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **28**:4488–4496.
43. Wenzel, R., and R. Herrmann. 1988. Repetitive DNA sequences in *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **16**:8337–8350.

Editor: S. R. Blanke