JOURNAL OF BACTERIOLOGY, Feb. 1996, p. 917–921
0021-9193/96/$04.00+0

Vol. 178, No. 3

The Repetitive Element Rep MP 1 of *Mycoplasma pneumoniae* Exists as a Core Element within a Larger, Variable Repetitive Mosaic

MARK H. FORSYTH AND STEVEN J. GEARY*

Department of Pathobiology, The University of Connecticut, Storrs, Connecticut 06269-3089

Received 5 September 1995/Accepted 26 November 1995

The repetitive element Rep MP 1 has been previously described as a 300-bp element present within the chromosome of *Mycoplasma pneumoniae* in at least 10 copies. Sequence found flanking Rep MP 1 in λ clone 5B52 has been demonstrated to exist in multiple copies within the genome of *M. pneumoniae*. A *Hin*dIII *M. pneumoniae* fragment containing the λ MP 5B52 insert and its flanking sequences was cloned into pBluescript. Sequence comparisons of this clone, designated pMPH 624-20, with nucleotide database entries showed that in addition to the 300-bp Rep MP 1 element, flanking sequence blocks were conserved within several other cloned sequences. These short repeated sequences, approximately 80, 56, and 71 nucleotides long, are termed sRep A, sRep B, and sRep C, respectively. These sRep regions were used as probes in the selection of cloned sequences containing these repetitive flanking regions. Comparison of these sequences demonstrated that the 300-bp Rep MP 1 element is actually the core element within a larger, variable repetitive mosaic. The repetitive regions surrounding the core element are found in various combinations, arrangements, and distances from the core in a mosaic pattern. These newly identified portions of the mosaic do not exist independently of the core element. The core appears to be the only invariant portion of this repetitive mosaic.

Repetitive sequences have been found in a great variety of microorganisms, both prokaryotic and eukaryotic, such as *Candida albicans*, *Brucella* spp., *Leptospira interrogans*, *Neisseria* spp., *Pneumocystis carinii*, and *Spiroplasma* spp. (3, 8, 10, 13, 20, 25). The genomes of these organisms often possess several repeated sequences, some at low copy number, such as the genes encoding tRNA, rRNA, or insertion elements. Repeated genes can, in systems such as that of *Neisseria gonorrhoeae*, be used as a source of antigenic variation (7). They are also known to be responsible for duplications within the *Escherichia coli* genome (11). Reiterated DNA stretches from *Mycoplasma hyorhinis*, *M. hyopneumoniae*, and *M. pneumoniae* have also been characterized (21, 22). Varied and high-copy-number repeats may serve as homologous substrates for generalized recombination. It is widely believed that mycoplasmas arose because of reductive evolution, losing large portions of the genome throughout the process (23). Homologous stretches of DNA, such as these repetitive elements, could mediate the loss of genomic material by deletion of intervening sequences during recombination. This type of deletion has been documented in *Yersinia pestis*, where a large block of chromosomal DNA responsible for pigmentation has been lost because of recombination between flanking repetitive elements (5).

M. pneumoniae is the etiologic agent of primary atypical pneumonia of humans. The P1 gene, which encodes a protein involved in cytadherence, contains two repetitive elements. The first, termed Rep MP 2 (22), was initially reported to be 150 bp long. Later work has expanded the boundaries of this reiterated sequence to 1,793 bp, and the element has been renamed Rep MP 2/3 (16). A second repetitive element, Rep MP 4, is 1,545 bp long and located at the 5′ end of the coding region. The P1 gene is the central gene of a three-gene operon (open reading frame 4 [ORF4], P1, and ORF6). Within ORF6 of the P1 operon is another, distinct repetitive element 2,167 bp long, termed Rep MP 5 (16). This repetitive element is present in at least seven other locations within the *M. pneumoniae* genome. Several of these extragenic copies vary slightly in both size (16) and sequence (15). It is significant that large portions of the coding regions of these genes exist in multiple, extragenic copies. It has been proposed that rearrangements may occur placing different copies of a repetitive element within the coding region of a structural gene (15).

Rep MP 1 was initially described as a 300-bp element, and at least 10 copies have been detected and mapped to various locations on the chromosome of *M. pneumoniae* by using an ordered cosmid library (22). Unlike the other characterized repetitive elements of this species, Rep MP 1 has not been conclusively demonstrated to be within a structural gene; however, a sequence containing approximately 100 bp of Rep MP 1 has been expressed as a fusion protein in *E. coli*. Antiserum against this fusion protein reacts with an approximately 30-kDa protein on the membrane of *M. pneumoniae* (14). Data from our laboratory also suggest that a copy of Rep MP 1 is a portion of a structural gene. The fusion protein expressed from the *M. pneumoniae* λ clone designated 5B52 reacts with monospecific antiserum to a purified 32-kDa membrane protein of *M. pneumoniae* (6). Sequence analysis has demonstrated the presence of 160 bp of Rep MP 1 as a portion of this ORF. Additionally, data from our laboratory obtained by using reverse transcriptase PCR of total *M. pneumoniae* RNA with primers specific to Rep MP 1 confirm that at least one copy of Rep MP 1 is transcriptionally active (data not shown).

Repetitive character of the 3′ flanking sequence of Rep MP 1. The *Eco*RI insert of λ 5B52 was labeled by nick translation with [α-^{32}P]dATP and a Nick Translation System (Gibco/BRL) by following the manufacturer's protocol and used to probe genomic digests of *M. pneumoniae* DNA. The results revealed the presence of between 10 and 12 genomic copies of this sequence (Fig. 1B, blot 1). The differing signal intensities of the various bands are probably due to the multiple copies of

* Corresponding author. Mailing address: Dept. of Pathobiology, Box U-89, University of Connecticut, 61 North Eagleville Road, Storrs, CT 06269. Phone: (203) 486-0835. Fax: (203) 486-2794.

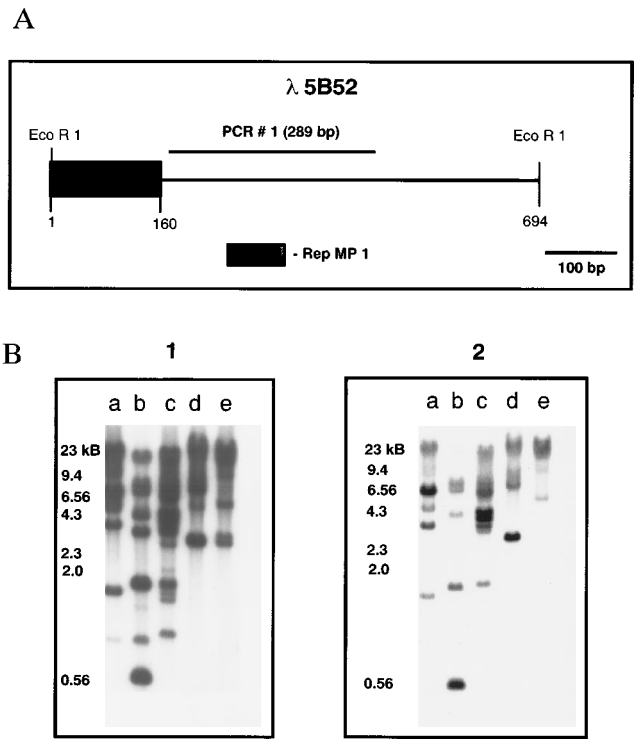

FIG. 1. (A) Diagram of λ 5B52. The salient features of this sequence are represented. The 3′ 160 nucleotides of the 300-bp Rep MP 1 element are shown. PCR #1 was generated with primers A and B (see Fig. 2). (B) Two identical Southern hybridizations of *M. pneumoniae* PI1428 genomic DNA digested with *Cla*I (a), *Eco*RI (b), *Hin*dIII (c), *Sal*I (d), and *Pst*I (e). Southern blot 1 was hybridized with the 694-bp *Eco*RI insert of λ 5B52. Southern blot 2 was probed with PCR #1.

Fragment Designation	Primer Designations and Sequence
PCR #1	A 5′ TAAGAATCCAGGGTGAGCAAAT
	B 5′ TCCTCCACTAAATAAATTGACCG
PCR #2	C 5′ TTGGCAGCACTACTGCAAGTTG
	D 5′ ATTTGCTCACCCTGGATTCTTA
PCR#3	E 5′ TTGTGTTAAAAAAACGAGCC
	F 5′ AGGTAAGAGGCAGACAGCTC
PCR #4	G 5′ AGAACTTCAAATCGAGCAAA
	H 5′ AAGACTAATTGCTGGTATTC

FIG. 2. Primers used to generate the various probes used throughout this study. Designations of fragments used as probes and their lengths are shown. Locations of the various regions amplified are shown schematically in Fig. 1A and 3A.

the target sequence present in several, but not all, positive restriction fragments. Sequence analysis of this 694-bp insert with MacDNASIS (Hitachi) and database searches with BLAST (1) demonstrated that 160 bp of the 300-bp repetitive element Rep MP 1 were present within this cloned insert, suggesting that this repeat was responsible, at least in part, for the hybridization pattern.

PCR fragment #1 (Fig. 1A) is a 289-bp fragment flanking Rep MP 1 on the 3′ side which was designed to select an overlapping clone. All PCRs were 100 μl and contained 2.5 U of AmpliTaq DNA Polymerase (Perkin-Elmer Cetus); 250 μM each dATP, dCTP, dGTP, and dTTP; 2 mM $MgCl_2$, 400 ng of each primer, and 100 ng of template DNA. All DNA templates were cloned sequences, i.e., λ or plasmid clones, to eliminate artifacts which may arise from genomic templates. The reaction conditions were as follows. After polymerase addition, 30 cycles of 94°C for 1 min and annealing at 5°C below the thermal melt of each primer set for 1 min and 72°C for 1.5 min. After completion of 30 cycles, a final extension was performed at 72°C for 10 min.

PCR #1, labeled and used as a probe, identified a 4.3-kb *Hin*dIII clone designated pMPH624-20, from an *M. pneumoniae* PI1428 pBluescript SK+ library. This clone contains the original λMP5B52 insert, as well as the flanking regions. Both strands of an internal 1,936-bp section of pMPH624-20 were sequenced by dideoxy reactions (17) with a modified T7 DNA polymerase (Sequenase; U.S. Biochemical). The pMPH624-20 sequence is available in the GenBank database (accession number U08987).

Southern hybridization of *M. pneumoniae* genomic DNA with PCR #1 indicates the presence of between five and seven copies of this sequence within the genome (Fig. 1B, blot 2). This is significant considering that the sequence within this probe is outside the recognized boundaries of Rep MP 1.

All prehybridizations of Southern blots were done at 42°C in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–5× Denhardt's solution–100 μg of salmon sperm DNA per ml–10 mM EDTA–0.1% sodium dodecyl sulfate (SDS). All hybridizations were performed at 42°C in prehybridization buffer with formamide added to a final concentration of 45%. High-stringency posthybridization washes were done as follows: two 3-min washes in 2× SSC–0.1% SDS at 25°C, two 3-min washes in 0.2× SSC–0.1% SDS at 25°C, and one wash in 0.16× SSC at 50°C for 15 min.

Repetitive character of the 5′ flanking sequence of Rep MP 1. Analysis of the 5′ flanking region of the Rep MP 1 in clone pMPH624-20 demonstrated that this region has repetitive character as well. PCR #3 is a 139-bp fragment flanking Rep MP 1 on the 5′ side. Hybridization of *M. pneumoniae* genomic digests with PCR #3 (primers E and F in Fig. 2) showed that this sequence hybridizes with as many as seven fragments (Fig. 3B). Comparison of the hybridization pattern of probe PCR #3 with that of probe PCR #2 (the Rep MP 1 sequence [primers C and D in Fig. 2]) suggests that these two sequences are linked in many of these genomic fragments (Fig. 3B).

pMPH624-20 contains novel short repeats. Homology between pMPH624-20 and four additional cloned sequences of *M. pneumoniae* (pSPT6, pSPT13, pF41 [22], and Mycp1C [18]) was found upon BLAST search (1). Much of this homology was

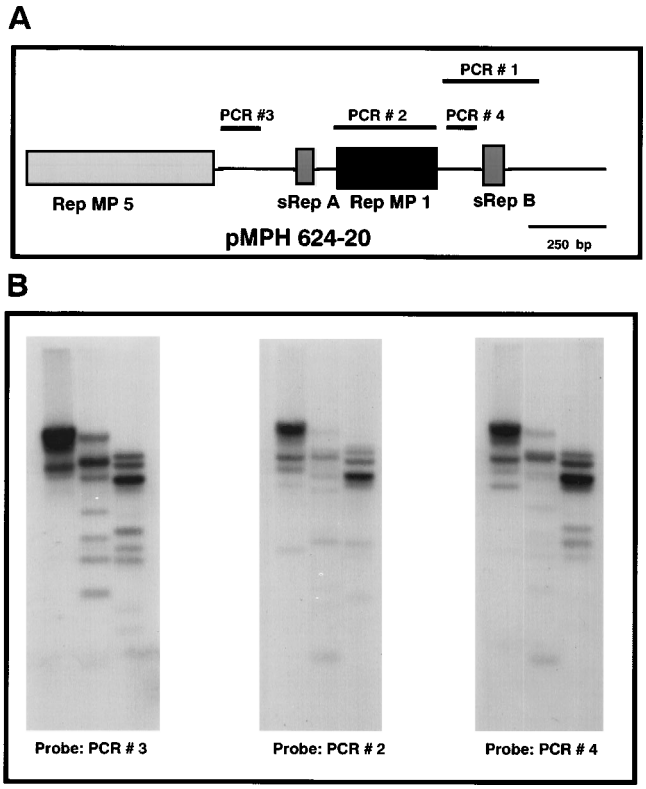

FIG. 3. (A) Drawing of pMPH624-20 showing locations of the various PCR products used as probes in panel B. The location of PCR #1, containing the sRep B sequence, is shown for comparison. An autoradiograph of a Southern hybridization using PCR #1 as a probe is shown in Fig. 1B, blot 2. (B) Autoradiographs from identical Southern blots of *M. pneumoniae* PI1428 digested with *Cla*I, *Eco*RI, or *Hin*dIII. Each blot was probed with the PCR product indicated under the individual autoradiograph. The locations of the various PCR probes in the cloned sequence, pMPH624-20, are shown in panel A.

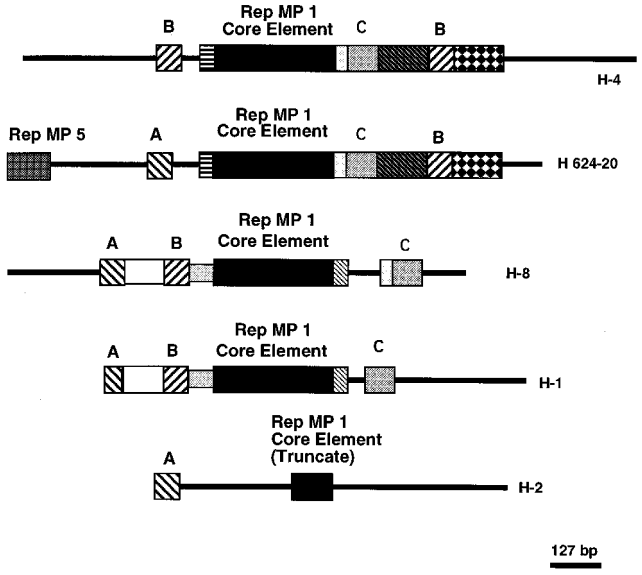

FIG. 4. Rep MP 1 repetitive mosaic. Nucleotide sequences with similar shaded boxes share significant (78 to 95%) homology. Sequences represented by lines share little or no homology with other sequences within this series. A, sRep A; B, sRep B; C, sRep C. Nucleotide sequences of each of the fragments are available through GenBank; their accession numbers are U08987 (H624-20), U35011 (H-1), U35012 (H-2), U35013 (H-4), and U35014 (H-8).

due to the presence of Rep MP 1 in these previously characterized clones. In each of these four clones, two short regions, approximately 80 and 56 nucleotides long, were found to share homology with pMPH624-20. While these previously characterized clones are largely homologous to one another throughout their length, their homology with pMPH624-20 (exclusive of Rep MP 1) lies only in these two short elements, which we have termed sRep A (80 bp) and sRep B (56 bp). The positions of sRep A and sRep B in the cloned sequence of pMPH624-20 are shown in Fig. 3A. More striking than the conserved nature of these short repeats is the different positioning of these homologous sequence blocks within pMPH624-20 with respect to the previously characterized portions of the chromosome.

While these short repeated sequences clearly contribute to the repetitive nature of the sequences surrounding Rep MP 1, they are not solely responsible for this repetitive nature. This was demonstrated through hybridization with two PCR probes designed to exclude Rep MP 1, sRep A, and sRep B (PCR #3 and PCR #4 in Fig. 3). When these probes were used to determine the copy number within the *M. pneumoniae* genome, they were found to be highly reiterated sequences as well (Fig. 3B). The hybridization patterns of these probes and a probe specific for the 300-bp Rep MP 1 element (PCR #2) are quite similar to one another. This suggests that although the flanking sequences are clearly repetitive, they do not appear to exist independently of Rep MP 1.

Rep MP 1 is a conserved element within a larger, variable repetitive mosaic. To examine the extended nature of the Rep MP 1 sequence, we selected clones from a *Hin*dIII-pBluescript SK (Statagene, LaJolla, Calif.) *M. pneumoniae* library. Clones pMPH215-1 (H-1), pMPH215-2 (H-2), pMPH215-4 (H-4), and pMPH215-8 (H-8) (Fig. 4) were selected by using probes made from the sequences flanking Rep MP 1 on both sides in cloned sequence pMPH624-20. Although the probes did not contain the Rep MP 1 sequence, each of these selected clones does contain Rep MP 1. This is further evidence that these repetitive sequences found flanking Rep MP 1 appear to exist only in association with the 300-bp Rep MP 1 element, which we now consider to be the core element within a repetitive mosaic.

Careful examination of the sequences found in each of the clones characterized in this study, shown schematically in Fig. 4, revealed a complex pattern of conserved blocks of sequences flanking the core element. It is not simply the conserved nature of these homologous blocks which is significant. These flanking sequences are found in some, but not all, members of this series. Upon examination of the relative distances of these repeated sequences from the core element, we found that several of the repeated blocks exist in different locations and in a variety of arrangements relative to one another. In addition to sRep A and sRep B, a third short repeat was found in several copies at various locations relative to the core element. This block, 71 bp long, is designated sRep C (Fig. 4). We have demonstrated that at least one of these conserved, repeated blocks (sRep B) can exist on either side of the core element and, in at least one case, on both sides of the core (Fig. 4, H-4).

The single unifying feature of this variable repetitive mosaic is the presence of the core element. However, while existing in each clone, the core element in clone H-2 is truncated to a length of 100 bp. The remaining 200 bp have apparently been deleted. While the mosaic pattern described in this report is suggestive of genomic rearrangements as the genesis of this variable pattern, the deletion of approximately 67% of the core in H-2 is perhaps the best indication to date of a genomic rearrangement involving Rep MP 1.

Although there is no indication that any of the repetitive

elements found within this mosaic exist as mobile elements within the chromosome, the various combinations and arrangements of these interspersed repeat and intervening sequences suggest that rearrangements of the chromosome may have occurred. The presence of homologous sequence blocks within the same mosaic unit (sRep B copies in H-4) suggests that homologous recombination may occur within the same mosaic unit.

Although the regions of the *M. pneumoniae* chromosome characterized in these five clones contain no ORFs that could encode a protein larger than approximately 16 kDa or with significant homology to any known protein, the existence of repetitive elements as portions of coding regions of structural genes leads to the possibility of constant and variable domains of antigens. This has been proposed, and suggestive evidence has been presented (2, 15). There is no direct evidence to date of protein variation in mycoplasmas due to repetitive elements; however, the Arp protein, an immunoglobulin A receptor in the M protein family of *Streptococcus pyogenes*, is an example of a protein in which variation of domains has occurred because of intergenic recombination. This has resulted in novel N-terminal sequences with retention of constant signal and C-terminal sequences (9). Repetitive elements have also been proposed as segments controlling gene regulation both by differential translation and by increasing the half-life of mRNA (12, 19). A stem-and-loop structure potentially resulting from the small repetitive extragenic palindrome (REP) in *E. coli* may accomplish these functions even though it has not been shown to terminate transcription.

A potential role for repetitive elements in homologous or generalized recombination cannot be ignored since the presence of high-copy-number segments of homologous DNA could be a ready supply of substrates for recombinative events, and this has indeed been proposed within the class *Mollicutes* (4, 15). Rearrangements of the genome that would occur as a result of homologous recombination could explain the *M. pneumoniae* mosaic of constituent blocks within the repetitive element Rep MP 1 characterized in this work.

It has recently been determined that recombination between directly repeated sequences can occur via either *recA*-dependent or *recA*-independent mechanisms, depending on the distance separating the homologous sequences and the size of the homologous regions (24). It was shown that there was virtually no *recA* dependence when the repeat sequence sizes were as short as 14 bp or as long as 300 bp. Increasing the repeat sequence size to greater than 300 bp gradually increased *recA*-dependent recombination. The length of the intervening sequence between two repeats also determines the *recA* dependence of recombination. Increasing the distance between adjacent repeats from 100 to 3,872 bp gradually increased *recA*-dependent recombination. Recombination, either *recA* dependent or independent, may also mediate rearrangements involving separate Rep MP 1 mosaic units or within the same mosaic unit based on the presence of homologous repeats within the same unit. Such recombinative events may be responsible not only for the variable pattern in Rep MP 1 but also for loss of genomic material through the deletion of intervening sequences.

It is clear that Rep MP 1 exists as a mosaic consisting of a core element within a larger, variable repetitive sequence. The Rep MP 1 mosaic is variable in both its length and the arrangement of the blocks that make up the mosaic. On the basis of direct sequence comparisons and hybridization of various portions of the mosaic with genomic DNA, our current minimal estimate of the potential length of the Rep MP 1 mosaic is greater than 1 kb.

The expansion of this repetitive sequence within the limited chromosome of *M. pneumoniae* presents several problems, as well as many possibilities. A previous estimate of the percentage of the genome of *M. pneumoniae* represented by repetitive elements was reported to be approximately 7% (16). With a genome so limited, it could be argued that there is little room for superfluous DNA. While it seems likely that these interspersed repetitive elements play a significant role in the functioning of these prokaryotes, they may not be the driving force of chromosomal rearrangements but may instead be physical indicators of rearrangements that have previously occurred. A current focus of our research is the identification and characterization of other regions of the chromosome containing copies of these novel sequences. We are currently assessing the potential of Rep MP 1 to act as a substrate in homologous recombination. Only in this manner can we begin to understand the complex arrangements of these elements and hope to establish their function(s). It is now critically important to our understanding of the biology of these organisms that a function(s) be assigned to these sequences.

This work was funded through the University of Connecticut Research Foundation (440114) and the USDA Agricultural Experiment Station (no. CONS00640).

We are grateful to Timothy S. Gorton and Stephanie Goh for many helpful discussions.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Colman, S. D., P. C. Hu, and K. F. Bott.** 1990. Prevalence of novel repeat sequences in and around the P1 operon in the genome of *Mycoplasma pneumoniae*. Gene **87:**91–96.
3. **Correia, F. F., S. Inouye, and M. Inouye.** 1986. A 26 base pair repetitive sequence specific for *Neisseria gonorrhoeae* and *Neisseria meningitidis* genomic DNA. J. Bacteriol. **167:**1009–1015.
4. **Deng, G., and M. A. McIntosh.** 1994. An amplifiable DNA region from the *Mycoplasma hyorhinis* genome. J. Bacteriol. **176:**5929–5937.
5. **Fetherston, J. D., P. Schuetze, and R. D. Perry.** 1992. Loss of the pigmentation phenotype in *Yersinia pestis* is due to spontaneous deletion of 102 kb of chromosomal DNA which is flanked by a repetitive element. Mol. Microbiol. **6**(18)**:**2693–2704.
6. **Geary, S. J., and M. F. Forsyth.** 1991. Molecular characterization of the 32 kD binding protein of *Mycoplasma pneumoniae*, abstr. G-30, p. 138. *In* Abstracts of the 91st General Meeting of the American Society for Microbiology 1991. American Society for Microbiology, Washington, D.C.
7. **Haas, R., and T. F. Meyer.** 1986. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. Cell **44:**107–115.
8. **Halling, S. M., and E. S. Zehr.** 1990. Polymorphism in *Brucella* spp. due to highly repeated DNA. J. Bacteriol. **172:**6637–6640.
9. **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.
10. **Iwaguchi, S. I., M. Homma, H. Chibana, and K. Tanaka.** 1992. Isolation and characterization of a repeated sequence (RPS1) of *Candida albicans*. J. Gen. Microbiol. **138:**1893–1900.
11. **Lin, R. J., M. Capage, and C. W. Hill.** 1984. A repetitive DNA sequence, rhs, responsible for duplications within the *Escherichia coli* K-12 chromosome. J. Mol. Biol. **177:**1–18.
12. **Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles, and C. F. Higgins.** 1987. Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell **48:**297–310.
13. **Nur, I., D. J. LeBlanc, and J. G. Tully.** 1987. Short, interspersed, and repetitive DNA sequences in *Spiroplasma* species. Plasmid **17:**110–116.
14. **Proft, T., and R. Herrmann.** 1994. Identification and characterization of hitherto unknown *Mycoplasma pneumoniae* proteins. Mol. Microbiol. **13**(2)**:** 337–348.
15. **Ruland, K., R. Himmelreich, and R. Herrmann.** 1994. Sequence divergence in the ORF6 gene of *Mycoplasma pneumoniae*. J. Bacteriol. **176:**5202–5209.
16. **Ruland, K., R. Wenzel, and R. Herrmann.** 1990. Analysis of three different repeated elements present in the P1 operon of *Mycoplasma pneumoniae*: size, number, and distribution on the genome. Nucleic Acids Res. **18:**(21)6311–6317.
17. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

18. **Schaper, U., J. S. Chapman, and P. C. Hu.** 1987. Preliminary indication of unusual codon usage in the DNA coding sequence of the attachment protein of *Mycoplasma pneumoniae*. Isr. J. Med. Sci. **23:**361–367.
19. **Stern, M. J., E. Prossnitz, and G. F. L. Ames.** 1988. Role of the intercistronic region in post-transcriptional control of gene expression in the histidine transport operon of *Salmonella typhimurium*: involvement of REP sequences. Mol. Microbiol. **2:**141–152.
20. **Stringer, S., S. T. Hong, D. Giuntoli, and J. R. Striger.** 1991. Repeated DNA in *Pneumocystis carinii*. J. Clin. Microbiol. **29:**1194–1201.
21. **Taylor, M. A., R. V. Ferrel, K. S. Wise, and M. A. McIntosh.** 1988. Reiterated DNA sequences defining genomic diversity within the species *Mycoplasma hyorhinis*. Mol. Microbiol. **2:**665–672.
22. **Wenzel, R., and R. Herrmann.** 1988. Repetitive DNA sequences in *Mycoplasma pneumoniae*. Nucleic Acids Res. **16**(17)**:**8337–8350.
23. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
24. **Xin, B., and L. F. Liu.** 1994. *recA*-independent and *recA*-dependent intramolecular plasmid recombination. Differential homology requirement and distance effect. J. Mol. Biol. **235:**414–423.
25. **Zuerner, R. L., and C. A. Bolin.** 1988. Repetitive sequence element cloned from *Leptospira interrogans* serovar hardjo type hardjo-bovis provides a sensitive diagnostic probe for bovine leptospirosis. J. Clin. Microbiol. **26:**2495–2500.