JOURNAL OF BACTERIOLOGY, Nov. 1992, p. 7289–7296
0021-9193/92/227289-08$02.00/0

Vol. 174, No. 22

Construction of an *Eco*RI Restriction Map of *Mycoplasma pneumoniae* and Localization of Selected Genes

RAINER WENZEL, ELSBETH PIRKL, AND RICHARD HERRMANN*

Microbiology, ZMBH, University of Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, Germany

Received 30 March 1992/Accepted 17 July 1992

A restriction map of the genome of *Mycoplasma pneumoniae*, a small human pathogenic bacterium, was constructed by means of an ordered cosmid library which spans the complete bacterial chromosome. The positions of 143 endonuclease *Eco*RI restriction fragments were determined and aligned with the physical map. In addition, restriction sites for the rare-cutting enzymes *Xho*I (25 sites), *Apa*I (13 sites), *Not*I (2 sites), and *Sfi*I (2 sites) were included. The resulting map consists of 185 restriction sites, has a mean resolution of 4.4 kbp, and predicts a genome size of 809 kbp. In addition, several genes were identified and mapped to their respective genomic *Eco*RI restriction fragments.

Until recently, the construction of genetic maps and the determination of microbial genome sizes were time-consuming procedures. DNA size measurements were performed by renaturation kinetics or electron microscopy techniques (2, 6, 9, 18). Genetic linkage maps, like the one constructed for *Escherichia coli* (1), were assembled from analyses of conjugational recombination, complementation, and generalized transduction and by mapping of transposon insertions (for a review, see reference 31).

The introduction of pulsed-field gel electrophoresis (7, 46, 49) and various two-dimensional agarose gel systems (5, 44, 64) enabled the rapid construction of low-resolution physical maps of bacterial genomes with rare-cutting enzymes such as *Not*I, *Sfi*I, or *Avr*II (examples are given in references 3, 35, 48, and 50; for a review, see reference 49a). These strategies have the advantage of being fast and easy to perform. However, because of their low resolution, the information extracted from these physical maps is limited. Another more laborious approach uses ordered clone libraries for the construction of genomic restriction maps of more frequently cutting enzymes such as *Eco*RI, *Bam*HI, and *Hin*dIII. The best-known example of such an analysis is the eight-enzyme restriction map of the *E. coli* K-12 W3110 genome constructed by Kohara et al. (25).

Members of the class *Mollicutes*, i.e., bacteria lacking a cell wall, are known to possess the smallest genomes of all bacteria so far examined (28, 43). Recent studies showed that their genome sizes range from 600 to 1,800 kbp (13, 36, 39, 53, 63). Several physical maps have been established by means of pulsed-field gel electrophoresis or two-dimensional gel electrophoresis, (e.g., for *Mycoplasma mycoides* subsp. *mycoides* [40], *Mycoplasma* strain PG50 [41], *Mycoplasma capricolum* [62], *Mycoplasma mobile* [5], *Mycoplasma genitalium* [13], and *Ureaplasma urealyticum* [10]). *Mycoplasma pneumoniae*, a member of the class *Mollicutes* and the object of this study, causes primary atypical pneumonia in humans. The complete genome of this bacterium has been cloned into a set of 34 overlapping or adjacent cosmids, two λ phages, and one plasmid by using a fast chromosome walking technique (59). As a first step toward establishing a detailed physical map, an *Xho*I restriction map composed of 25 fragments was constructed with this cosmid clone collection (59). The genome size was determined by summation of the sizes of the individual unique *Eco*RI restriction fragments of the gene bank (this enzyme was used to construct the library) and of the genomic *Xho*I restriction fragments. These summations yielded values of 835 and 849 kbp. In an independent study by Krause and Mawn (27), the genome sizes were calculated to be 775 and 794 kbp, respectively. A physical map was also constructed and comprised 13 *Apa*I sites, 2 *Not*I sites, and 2 *Sfi*I sites. In addition, the position of four genes, coding for rRNA (*rrn*), deoxyribose-phosphate aldolase (*deoC*), a high-molecular-weight protein (*hmw*3) (26), and the adhesin P1 (23, 55), have been localized on this physical map.

In order to characterize the genome of *M. pneumoniae* more precisely, we decided to establish a detailed *Eco*RI restriction map. Our strategy consists of the following steps: (i) determination of number and size of all *M. pneumoniae* *Eco*RI restriction fragments represented by the cosmid collection; (ii) construction of *Eco*RI restriction maps of individual cosmid clones by a modification of the technique described by Smith and Birnstiel (51); and (iii) assembly of the *Eco*RI restriction map of the complete genome by alignment of the data from the individual cosmid clones.

MATERIALS AND METHODS

Bacterial strains, growth of *M. pneumoniae* M129 B18, and construction of the cosmid library are described elsewhere (57).

Restriction enzymes. Restriction endonucleases were purchased from Boehringer or New England Biolabs and used as recommended by the manufacturers.

Isolation of tRNA. About 5×10^8 cells of a stationary *M. pneumoniae* culture were harvested by centrifugation at 12,500 × *g* for 15 min at 4°C, washed once in phosphate-buffered saline, and resuspended in 900 μl of 50 mM Tris (pH 8)–50 mM EDTA–200 mM NaCl. Following incubation for 5 min at 90°C, 10 μl of diethylpyrocarbonate and 20 μl of 10% sodium dodecyl sulfate (SDS) were added to the suspension. To shear chromosomal DNA, the solution was passed several times through a 21-gauge needle. After phenol-chloroform extraction, nucleic acids were precipitated with ethanol and resuspended in 100 μl of denaturing gel loading buffer (80% formamide, 50 mM Tris borate [pH 8.3], 1 mM EDTA, 0.1% [wt/vol] xylene cyanol, 0.1% [wt/vol] bromophenol

* Corresponding author.

TABLE 1. Synthetic oligonucleotides

Designation	Sequence	Specificity
o.SP6TW	5'-CACATACGATTTAGGTGACAC-3'	SP6 promoter pcosRW2
o.RW7	5'-TTCAAAATAGAGTGTTGTGGG-3'	$tRNA^{Trp}$ T, M.p.[a]
o.P1-5	5'-GAGAATCCAAGTGGACTTGG-3'	P1 gene 5' end, M.p.
o.P1-3	5'-GGAGCCCCTGGCTTTGGTGG-3'	P1 gene 3' end, M.p.
o.P1-M2	5'-CCCCCACCACTAAGCACAC-3'	P1 gene RepMP2/3 region, M.p.
o.MP16S	5'-CTTTAGCAGGTAATGGCTAGAG-3'	16S rRNA, M.p.
o.28K5	5'-GGCATCAACAATAACGGCTAAGGC-3'	ORF4 5' end, M.p.
130 N1	5'-CCGGGGAATTCGGATCCATGAAATCGAAGCTAAAG-3'	ORF6 5' end, M.p.
o.P 30-1	5'-CAGTTTAAGCTTTCTTCGAGGTGG-3'	P30 gene 5' end, M.p.
o.P 30 2	5'-CCACCTCAACCCGGTATGGCGCCTC-3'	P30 gene 3' end, M.p.

[a] M.p., *M. pneumoniae*.

blue) (32). Separation of low-molecular-weight RNAs is achieved by electrophoresis on 6% polyacrylamide gels containing urea (32). The tRNAs were identified by using an *E. coli* tRNA mixture as a size marker and were eluted from the gel by standard procedures (32).

Oligonucleotides. The oligonucleotides (Table 1) were synthesized by a Biosystems 380A machine according to the phosphoramidite method with a solid carrier (8). Oligonucleotides were purified as described by Ferretti et al. (15).

Gel electrophoresis and Southern blotting. The restricted DNA was analyzed on 0.4 to 1.5% agarose gels. Following ethidium bromide staining, the gels were transferred to nitrocellulose (BA85; Schleicher & Schuell) or nylon membranes (Biodyne; Pall) according to the instructions of the suppliers.

Dot blots of the cosmid collection were prepared by spotting 2 μl of the appropriate cosmid DNA solutions (10 μg/ml) onto nitrocellulose filters. The DNA was denatured in 1.5 M NaCl–0.5 M NaOH and subsequently neutralized in 1.5 M NaCl–0.5 M Tris-HCl (pH 8.0). After being dried and baked at 80°C for 2 h, the filters were used for hybridization.

Radioactive labelling of nucleic acids. Nick translation of DNA with [α-^{32}P]dATP was done with the Amersham International kit. Reactions were carried out for 90 min at 15°C and stopped by the addition of EDTA and phenol. DNA was separated from unincorporated nucleotides by Sephadex G-100 gel filtration and denatured by heating to 90°C for 5 min before use. 5' Labelling of oligonucleotides and dephosphorylated tRNA was done by the method of Maxam and Gilbert (33).

Preparation of the high-molecular-weight size standard. Forty micrograms of *Eco*RI-linearized pUC18 DNA (56) was ligated at a concentration of 500 μg/ml overnight at 16°C with 6 Weiss units (56a) of T4 DNA ligase. In some cases it was necessary to digest the ligation product partially with *Eco*RI to reduce the size of the highly polymerized concatamers to the range of 2.6 to 40 kbp. Approximately 0.5 μg of DNA was loaded into a 5-mm slot of an agarose gel.

Hybridization. Standard procedure. After the nitrocellulose (or nylon) filters were prehybridized in 50% formamide–5× SSC (1× SSC is 150 mM NaCl plus 15 mM sodium citrate [pH 7.2])–5× Denhardt's solution (1× Denhardt's solution is 0.02% Ficoll, 0.02% polyvinylpyrrolidine, and 0.02% bovine serum albumin)–0.1% SDS–100 μg of denatured herring sperm DNA per ml at 37°C for at least 4 h, hybridization was performed with 50% formamide–5× SSC–5× Denhardt's solution–0.1% SDS by adding the appropriate ^{32}P-labelled probe and incubating at 37°C overnight. The filters were washed three times for 30 min with 1× SSC–0.1% SDS at 68°C. Autoradiographic signals were obtained by exposing X-ray films (Kodak) overnight to the filters. When 5' ^{32}P-labelled oligonucleotides were used as probes, prehybridization and hybridization steps were performed with 5× SSC–5× Denhardt's solution at 37°C. The wash steps (two washes for 15 min at 37°C) were performed with 5× SSC–0.1% SDS.

Low-stringency hybridizations. For the identification of genes with probes derived from species other than *M. pneumoniae*, the protocol was slightly modified in order to enable cross hybridizations between conserved DNA regions. Low-stringency conditions were achieved by variation of the formamide concentration (between 20 and 40%) in the hybridization solution, reduction of hybridization temperature to 30°C, and use of washing solutions containing high salt concentraitons (2× SSC–0.1% SDS) combined with wash temperatures of 50°C.

***Eco*RI restriction mapping of the cosmid collection.** Individual cosmid clones were linearized by digestion with endonuclease *Sfi*I (see Fig. 3). Linearized cosmid DNA (0.75 μg) was then digested partially with 2 U of *Eco*RI in the presence of either 5.5 or 11 μg of sonicated calf thymus DNA in a total volume of 50 μl. Ten-microliter samples were taken after 0.5, 1, 2, 4, and 8 min and loaded onto a 0.6% agarose gel. Plasmid pUC18 multimers served as size markers. After electrophoresis, the gel was blotted onto nylon filters and hybridized against radioactively labelled pcosRW2 DNA.

Computer analysis. Computer analyses of DNA sequences were performed with the program package HUSAR (Heidelberg Unix Sequence Analysis Resources Release 2.0) at the German Cancer Research Center, Heidelberg, Germany, and data bank searches were carried out by using the GenBank and EMBL DNA libraries and the Dayhoff protein sequence library.

Nucleotide sequence accession numbers. The DNA sequence data were submitted to the EMBL data library. The accession numbers are X67651 for the *M. pneumoniae rplP* gene for 50S ribosomal protein L16, X67652 for the *M. pneumoniae rpsC* gene for 30S ribosomal protein S3, and X67653 for the *M. pneumoniae atpA* gene for ATP synthase F1 sector, α-subunit.

RESULTS

Size determination of all *Eco*RI restriction fragments. In a previous publication, we reported the determination of the genome size of *M. pneumoniae* by summation of the sizes of all unique *Eco*RI restriction fragments present in the cosmid collection (59). Since these data are essential for construction of the *Eco*RI map, we describe here the experimental basis for this calculation. Agarose (0.4 and 1.5%) gels were

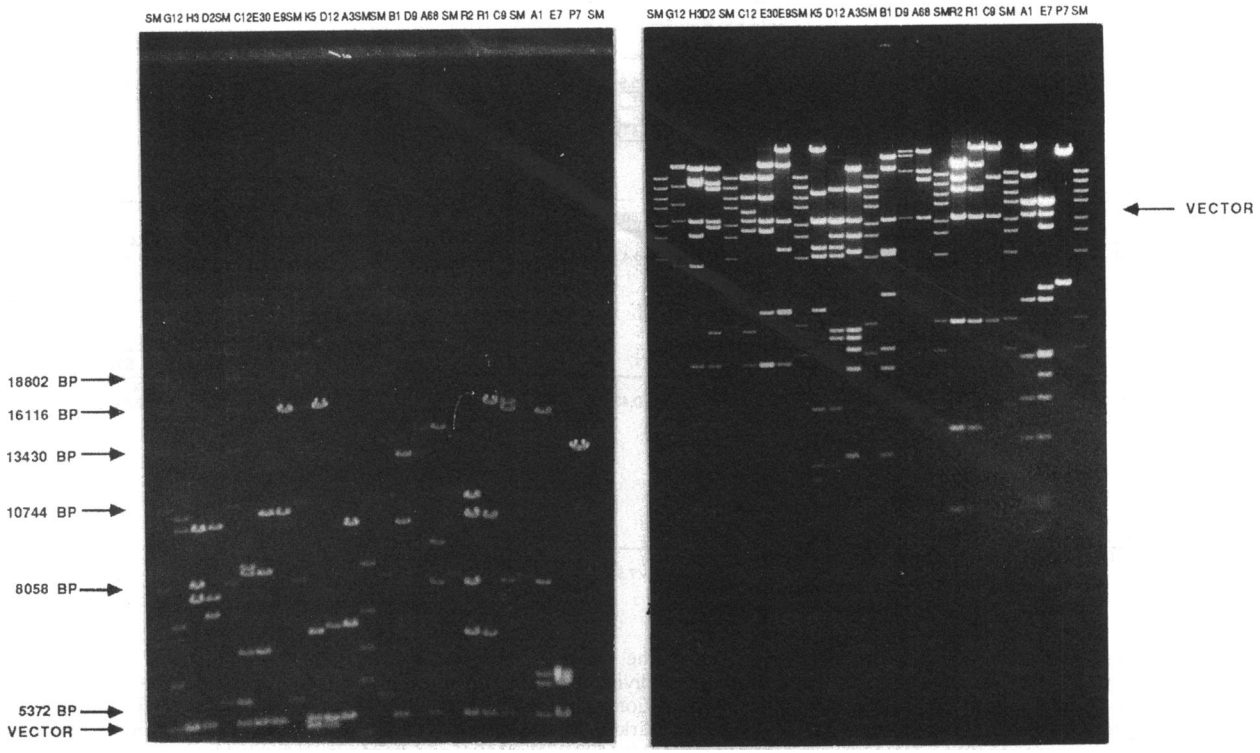

FIG. 1. Size determination of *M. pneumoniae Eco*RI restriction fragments, represented by 17 *Eco*RI digested cosmids and plasmid pSPT7 (P7) (57, 59). The fragments were separated on 0.4% (left) and 1.5% (right) agarose gels. SM, size marker (left, pUC18 polymers [monomer, 2,686 bp]; right, phage λ DNA × *Bst*EII [8,454, 7,242, 6,369, 5,686, 4,822, 4,324, 3,675, 2,323, 1,929, 1,371, 1,264, 702, 224, and 117 bp]).

used to resolve *Eco*RI restriction fragments of the cosmid library. An example of such a gel is shown in Fig. 1.

Figure 2 shows the size distribution of all fragments larger than 300 bp in the cosmid library. Fragments smaller than 300 bp were neglected. As described recently (59), several other clones, including a 16.5-kbp *Eco*RI fragment cloned in the plasmid pSPT7 and the two phage λ recombinants λX and λV, link the terminal cosmid clones pcosMPE7 and pcosMPGT9 and encompass the complete array of *Eco*RI fragments.

***Eco*RI restriction mapping of the individual cosmid clones and assembly of the genomic *Eco*RI map.** Our method is a modification of the technique originally introduced by Smith and Birnstiel (51). A brief description of it is shown in Fig. 3. Instead of end-labelled fragments, the partially digested cosmids were blotted and probed with radioactively labelled vector (pcosRW2) DNA. An example is shown in Fig. 4. The sizes of various DNA intermediates were determined by using linear polymers of pUC18 as size markers. The difference in sizes of two neighboring intermediates was evaluated by comparing them with the *Eco*RI fragments produced by complete *Eco*RI digestion of the cosmid in question. Such a comparison was not always possible either because the fragment sizes were too large to be accurately estimated or because there were too many *Eco*RI fragments with similar lengths in a cosmid. Generally, these problems could be solved by conventional restriction analysis with enzymes such as *Bam*HI or *Hin*dIII in combination with *Eco*RI. Another way to analyze the regions in question was to perform restriction mapping from the other end (T7) of the cloned fragment, since two *Not*I restriction sites flank the cloned DNA in pcosRW2 (Fig. 3). In all cases in which no additional *Not*I site is present in the cloned DNA, the insert could be excised as one contiguous fragment. If the terminal *Eco*RI fragments are known from parallel analyses from overlapping cosmids, then they can be used as probes in hybridizations against Southern blots of the partially *Eco*RI-digested *Not*I insert fragment.

The cosmids pcosMPP1/P2 and pcosMPGT9, which each contained one *Sfi*I site in the cloned region, caused problems in the restriction analysis. In pcosMPP1/P2, the *Sfi*I site is present in the T7-terminal 20.1-kbp *Eco*RI fragment of both clones. This means that the fragment size is reduced by an *Sfi*I-*Eco*RI subfragment in a partial *Eco*RI restriction analysis. In the case of pcosMPGT9, the *Sfi*I site in the SP6-terminal 25-kbp *Eco*RI fragment cuts the cloned DNA away from the vector. The *Eco*RI restriction map of this cosmid was established by data derived from other studies in which repetitive DNA elements in *M. pneumoniae* were analyzed (45). The final map is shown in Fig. 2.

Ambiguities remain in the pcosMPK5/E9/E30 region because of the unfavorable position of the cloned inserts with respect to the SP6 and T7 promoters, making it difficult to define the order of some fragments. In pcosMPK5, the order of the 6.6- and 0.87-kbp fragments could not be established. An 0.48-kbp *Eco*RI fragment in pcosMPE9/E30 has three possible locations. These ambiguities are indicated by brackets in Fig. 2.

In the cosmid clone pcosMPG7/GT9, two *Eco*RI fragments with very similar sizes are present which are indistinguishable in our analysis. Therefore, the two possible positions of these fragments are shown (Fig. 2).

The map also contains a 0.1-kbp *Eco*RI fragment located between the 5.7- and 2.7-kbp *Eco*RI fragments in

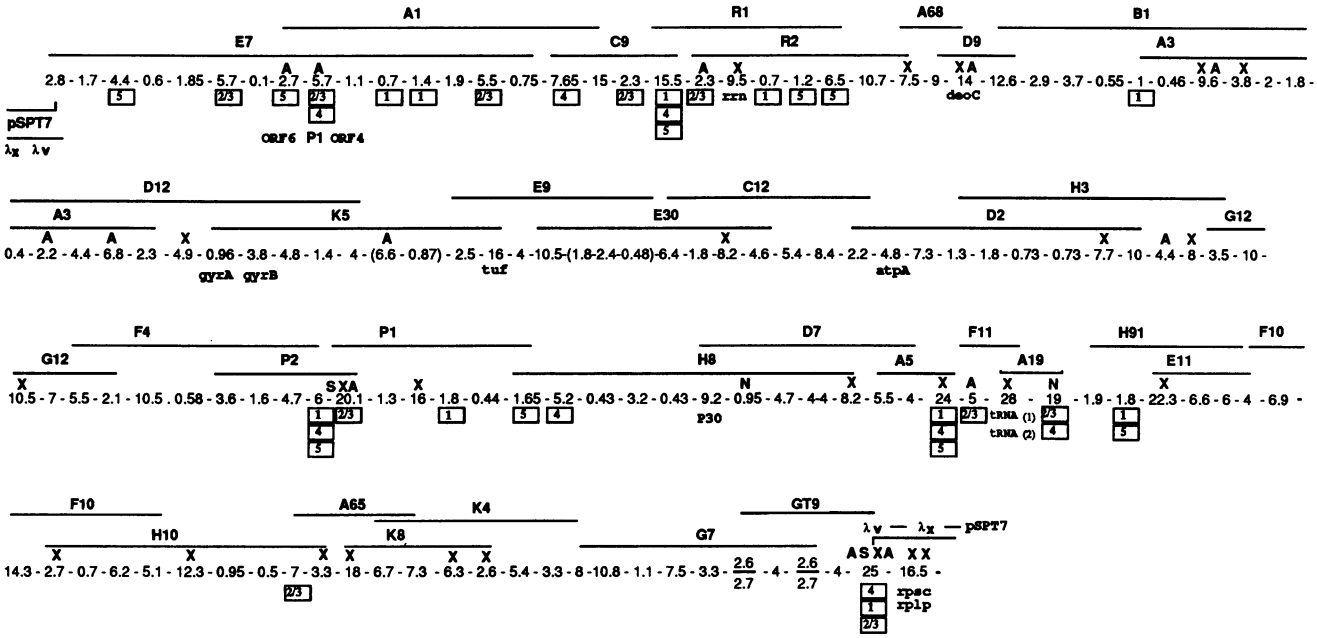

FIG. 2. *Eco*RI restriction map of the *M. pneumoniae* genome. The map was linearized between the 16.5-kbp (pSPT7) and the 2.8-kbp *Eco*RI (pcosMPE7) fragments. The numbers indicate the sizes of individual fragments (in kilobase pairs). For orientation, positions of the cosmid clones from the library are shown (horizontal bars). Recognition sites for endonucleases *Xho*I (X), *Apa*I (A), *Not*I (N), and *Sfi*I (S) are indicated. Restriction fragments containing identified genes are marked accordingly (for details, see text and Table 2). For completeness, the positions of four different repetitive DNA sequences (RepMP1, RepMP2/3, RepMP4, and RepMP5; indicated by boxed 1, 2/3, 4, and 5, respectively) are included (45).

pcosMPA1/E7 which is known from a published 10-kbp DNA sequence of this region (23). Furthermore, sequence analysis of certain repetitive DNA elements, called RepMP1 (58), revealed seven 85-bp *Eco*RI fragments in the cosmids pcosMPR1/R2, -P1, -P2/F4, -A1/E7, -H91, -GT9, and -C9 which were not located on the map.

The restriction sites for the enzymes *Xho*I (25 sites) *Apa*I (13 sites), *Sfi*I (2 sites), and *Not*I (2 sites), which have been published already, were mapped onto this *Eco*RI map by double digest of individual cosmids with these endonucleases. A total of 150 *Eco*RI fragments, including the seven 85-bp *Eco*RI fragments, were counted, and a genome size of 809 kbp was determined. This calculation differs by 26 kbp from the former determination, which was based on the comparison of the cosmid's *Eco*RI restriction pattern. This size difference probably arises from duplicating detection of several fragments that are located on overlapping regions of neighboring cosmids.

Identification and mapping of *M. pneumoniae* genes. In order to construct a genetic map, all *M. pneumoniae* genes which have been cloned and sequenced so far were localized on the *Eco*RI map by using either the cloned genes or oligonucleotides as specific probes in cross hybridization experiments. An alternative approach tested whether heterologous cloned genes could be useful in identifying the corresponding *M. pneumoniae* genes on a routine basis.

Gene probes derived from *M. pneumoniae* (sequences) were initially hybridized against dot blots of the complete cosmid library. Positive clones were characterized in more detail by Southern blotting of the endonuclease *Eco*RI-digested cosmid(s) in question. With *M. pneumoniae* probes all of the positive signals were unambiguous, and a further confirmation by DNA sequencing was unnecessary. However, this procedure had to be modified for analyses with heterologous gene probes as follows. Hybridization to genomic Southern blots of *Eco*RI-digested *M. pneumoniae* DNA was performed under various low-stringency hybridization conditions (see Materials and Methods). If positive signals (i.e., one or two cross-reacting *Eco*RI fragments) were obtained, the cosmid library was screened and the corresponding cosmid was identified. After subcloning, the region in question had to be sequenced. Finally, computer analyses of the DNA sequences were performed with respect to open reading frames, amino acid sequences, and homologies to known protein sequences of other bacterial species present in data bases; the existence of the counterpart gene in *M. pneumoniae* was considered to be proved when significant homologies, at least 40% identity, were detected.

Table 2 summarizes all of the genes which have been mapped so far, and Fig. 2 shows their positions on the *Eco*RI restriction map. Most of the analyses done with homologous gene probes were straightforward and need no further explanation. But some others are briefly described below.

P1 operon. The P1 gene that encodes the major adhesin protein P1 (20, 21, 55) is flanked by two open reading frames, ORF4 and ORF6 (23), which have coding capacities for a 28- and a 130-kDa protein. Substantial regions of the P1 gene (RepMP2/3 and RepMP4) and of ORF6 (RepMP5) appear as repetitive DNA sequences (12, 45, 54, 58, 60). In order to map these genes, it was necessary to use oligonucleotides which hybridized only to the unique regions, such as o.P1-5 and o.P1-3, which are specific for the 5′ and 3′ ends of the P1 gene, respectively, o.130 N1, with a specificity for ORF6, and o.28K5, with a specificity for ORF4. An example of a probe that gives false positives is o.P1-M2, which is derived

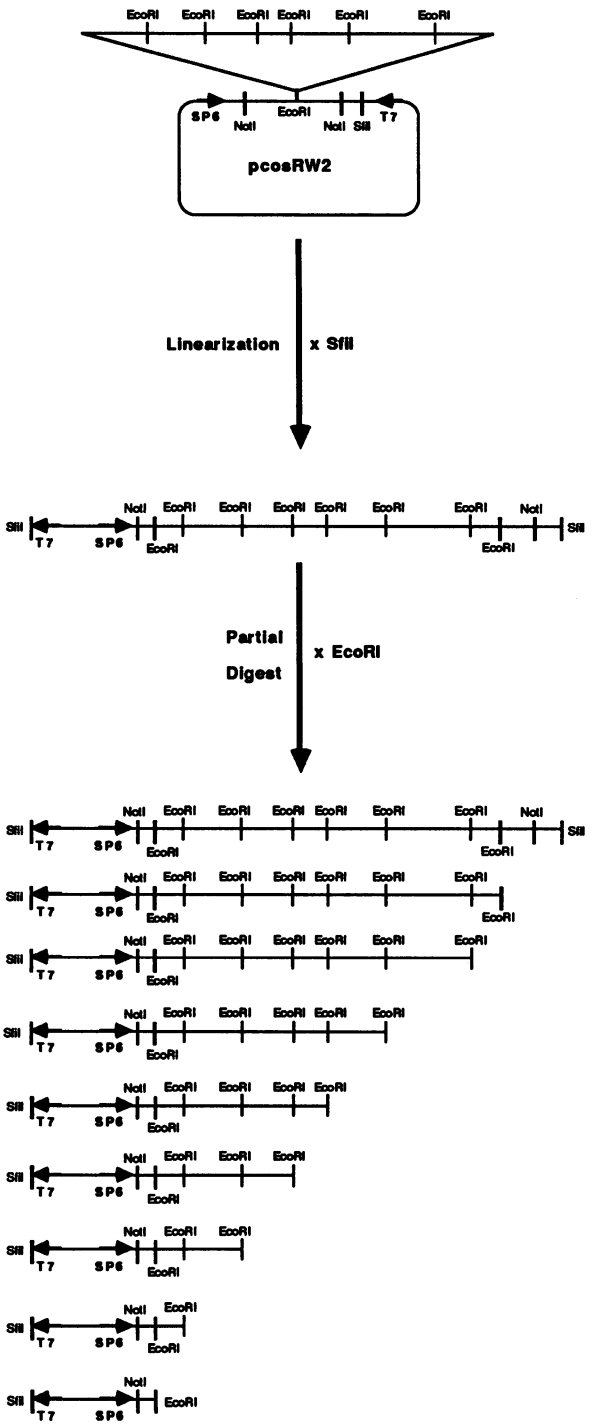

FIG. 3. Schematic drawing of the principle of the *Eco*RI restriction mapping of individual cosmid clones. Construction of the cloning vector pcosRW2 has been described previously (57). *Eco*RI restriction fragments produced by the partial digest are ordered according to size. Only those fragments which hybridize to radioactively labelled vector DNA are shown. SP6 and T7 indicate the positions of the bacteriophage promoters.

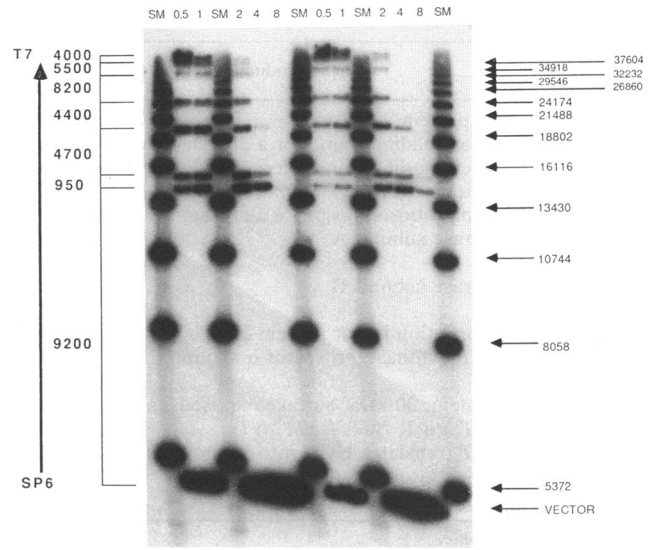

FIG. 4. *Eco*RI restriction mapping of cosmid pcosMPD7. The partially *Eco*RI-digested DNA was separated on a 0.6% agarose gel and blotted onto nylon. The filter was hybridized with ^{32}P-labelled pcosRW2 DNA. This vector also cross-reacts with pUC18 polymers which serve as size standards (numbers on the right). For an exact size determination, less-exposed X-ray films were used. SM, size marker; the numbers at the top represent reaction times (in minutes) at which the samples were taken. The numbers on the left indicate the order of *Eco*RI restriction fragments in pcosMPD7; SP6 and T7 positions are indicated. All sizes are in base pairs.

from a repetitive region of the P1 gene (RepMP2/3) and reacts with many *Eco*RI fragments in a Southern blot (Table 1 and Fig. 2).

tRNAs. tRNAs were identified by hybridizing radioactively labelled total tRNA preparations from *M. pneumoniae* cells (see Materials and Methods) to dot blots of the cosmid collection. Several cosmids appeared to be positive. A tRNA gene cluster coding for five tRNAs and the single tRNA with the anti-codon UCA gene were mapped to the 28-kbp *Eco*RI fragment of the cosmid pcosMPA19/F11. The position of the single tRNA cluster was confirmed by subcloning and DNA sequencing (47). The $tRNA^{Trp}$(UCA) gene was located by hybridization with the sequence-specific oligonucleotide o.RW7 (22). The other presumptive tRNA loci are not shown on the map. The final confirmation awaits completion by DNA sequencing of the hybridization-positive *Eco*RI fragments.

ATPase operon (*atp*). An example for the use of a heterologous probe to identify the corresponding *M. pneumoniae* is the F1Fo-ATPase. The probe was a cloned fragment of the ATPase operon from *Mycoplasma* strain PG50, an operon which has been cloned and partially sequenced (42).

Under our hybridization conditions (40% formamide–5× SSC at 37°C), a strong positive signal appeared with cosmid pcosMPD2 and was mapped to a 4.8-kbp *Eco*RI fragment. DNA sequence analysis of about 450 nucleotides from one DNA strand of the cloned fragment and a computer-assisted analysis of the data revealed an open reading frame with an amino acid sequence showing significant identity (50%) to a region of *atpA* of *Mycoplasma* strain PG50 and of *E. coli* (24) (Fig. 5).

Additional experiments were performed with heterologous gene probes to identify other genes which are widely spread

TABLE 2. *M. pneumoniae* genes localized onto the *Eco*RI map

Gene	Gene product	Map position (cosmid/ *Eco*RI fragment, kbp)	Probe (organism)	Reference and/or source
P1	P1 protein, adhesin	pcosMPE7/5.7	o.P1-5, o.P1-3, o.P1-M2 (M.p.[a])	21, 55
ORF4	Open reading frame for a 28-kDa protein	pcosMPE7/1.1	o.28 K5 (M.p.)	23
ORF6	Open reading frame for a 130-kDa protein	pcosMPE7/2.7	o.13ON 1 (M.p.)	23
rrn	Unique rRNA operon	pcosMPR2/9.5	o.MP16S (M.p.)	19
deoC	Deoxyribose-phosphate aldolase	pcosMPD9/14	Plasmid p11-6 (M.p.)	P.-C. Hu (30)
gyrA	Gyrase subunit A	pcosMPK5/0.96	Personal communication, K. F. Bott (M.p.)	11
gyrB	Gyrase subunit B	pcosMPK5/3.8	Personal communication, K. F. Bott (M.p.)	11
tuf	Elongation factor protein EF-Tu	pcosMPK5/16	Plasmid (M.p.)	S. Razin (65)
atpA	ATP synthase Fo sector α-subunit	pcosMPD2/4.8	Plasmid pMYC 1193 (strain G50)	C. Christiansen (42)
P30	Protein, 30 kDa, surface exposed	pcosMPH8/9.2	o.30-1, o.30-2 (M.p.)	14
tRNA cluster	*tyrT; gluT, lysT, leuT, glyT*	pcosMPF11/28	Total tRNA; sequencing (M.p.)	47
tRNA	*trpT* translates UGA into tryptophan	pcosMPF11/28	o.RW7	22
rplP	50S ribosomal subunit protein L16	λ v/8; pcosMPGT9/25	Random sequencing	
rpsC	30S ribosomal subunit protein S3	λ v/8	Random sequencing	

[a] M.p., *M. pneumoniae*.

among bacteria. Examples include genes such as the *dnaA* protein gene (16), the heat shock protein gene *hsp*90 (4, 29), and the *cheW* gene product, which plays a role in chemotaxis (52). The probes derived from *Salmonella typhimurium* (*cheW*) and *E. coli* (*dnaA* and *htpG*). In these hybridization experiments, either too many cross-reacting fragments were detected at low stringency or none were detected at more stringent hybridization conditions; in the case of the *htpG* gene, a weak signal appeared at low stringency (37°C, 20% formamide–5× SSC). DNA sequence analysis however, revealed a region with some limited homology to the probe at the DNA level but no open reading frame with significant homology to *hsp*90.

Ribosomal protein genes. Finally, genes for ribosomal proteins S3 and L16 were identified by sequence analysis (one strand only) of a region which initially was difficult to clone into *E. coli* plasmids or cosmids (57). Deduced amino acid sequences (181 amino acids) showed identity of 49 to 64% (depending on the analyzed region) between the *M. pneumoniae* and the *M. capricolum* (37) and *E. coli* (66) ribosomal proteins S3 and L16 (data not shown).

DISCUSSION

The construction of a physical *Eco*RI restriction map of the *M. pneumoniae* genome with a cloned library in combination with the fast and simple mapping technique of Smith and Birnstiel (51) was easy to perform and yielded clear and reliable results. In most cases, no additional information had to be provided by other types of experiments. A total of 185 restriction sites (143 *Eco*RI, 25 *Xho*I, 13 *Apa*I, 2 *Not*I, and 2 *Sfi*I sites) have been mapped with a mean resolution of 4.4 kbp and with a calculated genome size of 809 kbp. If the 143 *Eco*RI sites are counted separately, a mean fragment length of 5.7 kbp is determined. This value is only fairly consistent with the theoretical value of 3,086 bp calculated for a genome with a GC content of 41% (34), even considering that fragments smaller than 300 bp were neglected.

The *Eco*RI restriction map serves as a basis for the construction of a genetic one. This was achieved by the identification of several genes and their locations to particular restriction fragments. Our selection of genes was based mainly on three criteria. The first criterion is the relevance of

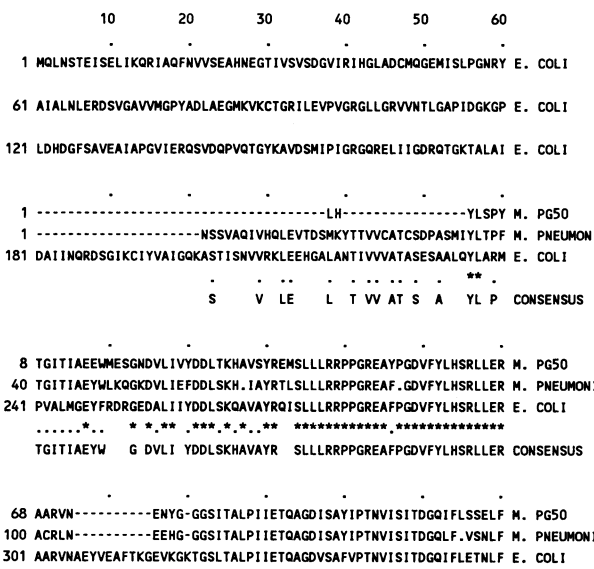

FIG. 5. Partial amino acid homology between the genes for the α-subunit of the ATPase from *M. pneumoniae*, *Mycoplasma* strain PG50 (42), and *E. coli* (24). For *Mycoplasma* strain PG50, only partial sequence data were available. Identical amino acids are indicated by asterisks; dots indicate identity between two of three amino acids. Homology is present across the complete sequenced DNA region of *M. pneumoniae* which in *E. coli* is represented by amino acid positions 202 to 363.

certain genes with respect to evolutionary or systematic relationship studies within the rather heterogeneous class *Mollicutes* (e.g., the elongation factor Tu). The second aspect concerns the possibility of comparing the *M. pneumoniae* map with other bacterial genetic maps. This means that only those genes which have previously been characterized can be taken into consideration. Examples include rRNA operons or ribosomal proteins. The last aspect is a more practical one and concerns gene probes which are available from related species. There is an additional difficulty in analyzing *M. pneumoniae* since this species, as an exception among members of the class *Mollicutes*, has a relatively high GC content of 41%, compared with the average GC range of members of the class *Mollicutes* of 24 to 30% (43). Therefore, with gene probes from other *Mycoplasma* species, the advantages arising from the close relationship are countered by the different codon usages of these extremely AT-rich organisms. It may be more promising to use probes from less-related species, such as *Bacillus subtilis*, which have similar GC contents. This was shown by Colman et al. (11), who used a *B. subtilis* gene probe for the identification of the *gyrA* and *gyrB* genes, which code for the gyrase subunits A and B, respectively.

The failure to detect *dnaA*, *htpG*, and *cheW* with heterologous probes does not indicate that these genes are absent from *M. pneumoniae*. It only illustrates a negative result that is based on cross hybridization difficulties with heterologous probes. In fact, Fujita et al. (17) succeeded in cloning the *dnaA* gene from *M. capricolum* by using degenerate synthetic oligonucleotides which were derived from sequences of regions conserved among different bacterial *dnaA* genes.

In summary, we concluded that the cross hybridization strategy with heterologous probes is not the method of choice for fast identification and localization of a larger number of genes for the purpose of constructing a genetic map. Because of the reduced stringency in the hybridization conditions, we frequently obtained several signals or even unique signals that turned out to be falsely positive. Therefore, it is essential in any instance to sequence the DNA of the fragment in question. The faster way of identifying genes is to sequence DNA regions by a shotgun strategy followed by the identification of possible coding regions and the search for amino acid homologies to known genes by computational analysis. Mycoplasmas, which are known to have small genomes, appear to be especially well suited for this strategy since their chromosomes should not contain too much unnecessary information (38). As far as *M. pneumoniae* is concerned, where a complete ordered cosmid library exists, sequencing of the genome in an ordered fashion is a promising alternative to a shotgun approach.

ACKNOWLEDGMENTS

We thank M. Nassal, R. Frank, and B. Borchert for synthesis of the oligonucleotides; K. F. Bott, C. Christiansen, P.-C. Hu, O. Rasmussen, S. Razin, and J. Stock for supplying the gene probes and unpublished information; L. Runkel for critically reading the manuscript; and S. Nazareth for preparing the manuscript.

This work was supported by the Bundesministerium für Forschung und Technologie (BCT 0381/5).

REFERENCES

1. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
2. **Bak, A. L., F. T. Black, C. Christiansen, and E. A. Freundt.** 1969. Genome size of mycoplasmal DNA. Nature (London) **224**:1209–1210.
3. **Bancroft, I., C. P. Wolk, and E. V. Oren.** 1989. Physical and genetic maps of the genome of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. J. Bacteriol. **171**: 5940–5948.
4. **Bardwell, J. C. A., and E. A. Craig.** 1987. Eukaryotic Mr 83,000 heat shock protein has a homologue in Escherichia coli. Proc. Natl. Acad. Sci. USA **84**:5177–5181.
5. **Bautsch, W.** 1988. Rapid physical mapping of the *Mycoplasma mobile* genome by two-dimensional field inversion gel electrophoresis techniques. Nucleic Acids Res. **16**:11461–11467.
6. **Bode, H. R., and H. J. Morowitz.** 1967. Size and structure of the *Mycoplasma hominis* H39 chromosome. J. Mol. Biol. **23**:191–199.
7. **Cantor, C. R., C. L. Smith, and M. K. Mathew.** 1988. Pulsed-field gel electrophoresis of very large DNA molecules. Annu. Rev. Biophys. Biophys. Chem. **17**:287–304.
8. **Caruthers, M. H.** 1982. New methods for synthesizing deoxyoligonucleotides. Genet. Eng. **4**:1–16.
9. **Chow, L. T.** 1977. Sequence arrangement of the *Escherichia coli* chromosome and of putative insertion sequences, as revealed by electron microscopic heteroduplex studies. J. Mol. Biol. **113**:611–621.
10. **Cocks, B. G., L. E. Pyle, and L. R. Finch.** 1989. A physical map of the genome of *Ureaplasma urealyticum* 960T with ribosomal RNA loci. Nucleic Acids Res. **17**:6713–6719.
11. **Colman, S. D., P.-C. Hu, and K. F. Bott.** 1990. *Mycoplasma pneumoniae* DNA gyrase genes. Mol. Microbiol. **4**:1129–1134.
12. **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.
13. **Colman, S. D., P.-C. Hu, W. Litaker, and K. F. Bott.** 1990. A physical map of the *Mycoplasma genitalium* genome. Mol. Microbiol. **4**:683–687.
14. **Dallo, S. F., A. Chavoya, and J. B. Baseman.** 1990. Characterization of the gene for a 30-kilodalton adhesin-related protein of *Mycoplasma pneumoniae*. Infect. Immun. **58**:4163–4165.
15. **Ferretti, L., S. S. Karnik, H. G. Khorana, M. Nassal, and D. D. Oprian.** 1986. Total synthesis of a gene for bovine rhodopsin. Proc. Natl. Acad. Sci. USA **83**:599–603.
16. **Fujita, M. O., H. Yoshikawa, and N. Ogasawara.** 1989. Structure of the *dnaA* region of *Pseudomonas putida*: conservation among three bacteria, *Bacillus subtilis*, *Escherichia coli* and *P. putida*. Mol. Gen. Genet. **215**:381–387.
17. **Fujita, M. O., H. Yoshikawa, and N. Ogasawara.** 1992. Structure of the *dnaA* and *DnaA*-box region in the *Mycoplasma capricolum* chromosome: conservation and variations in the course of evolution. Gene **110**:17–23.
18. **Gillis, M., J. de Ley, and M. de Cleene.** 1970. The determination of molecular weight of bacterial genome DNA from renaturation rates. Eur. J. Biochem. **12**:143–153.
19. **Göbel, U. B., A. Geiser, and E. J. Stanbridge.** 1987. Oligonucleotide probes complementary to variable regions of ribosomal RNA discriminate between Mycoplasma species. J. Gen. Microbiol. **133**:1969–1974.
20. **Hu, P.-C., R. M. Cole, Y. S. Huang, J. A. Graham, D. E. Gardner, A. M. Collier, and W. A. Clyde, Jr.** 1982. *Mycoplasma pneumoniae* infection: role of a surface protein in the attachment organelle. Science **216**:313–315.
21. **Inamine, J. M., T. P. Denny, S. Loechel, U. Schaper, C. H. Huang, K. F. Bott, and P.-C. Hu.** 1988. Nucleotide sequence of the P1 attachment-protein gene of *Mycoplasma pneumoniae*. Gene **64**:217–229.
22. **Inamine, J. M., K.-C. Ho, S. Loechel, and P.-C. Hu.** 1990. Evidence that UGA is read as a tryptophan codon rather than as a stop codon by *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, and *Mycoplasma gallisepticum*. J. Bacteriol. **172**:504–506.
23. **Inamine, J. M., S. Loechel, and P.-C. Hu.** 1988. Analysis of the nucleotide sequence of the P1 operon of *Mycoplasma pneumoniae*. Gene **73**:175–183.
24. **Kanazawa, H., T. Kayano, K. Mabuchi, and M. Futai.** 1981. Nucleotide sequence of the genes coding for alpha, beta and gamma subunits of the proton-translocating ATPase of *Escherichia coli*. Biochem. Biophys. Res. Commun. **103**:604–612.

25. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
26. **Krause, D. C., D. K. Leith, R. M. Wilson, and J. B. Baseman.** 1982. Identification of *Mycoplasma pneumoniae* proteins associated with hemadsorption and virulence. Infect. Immun. **35**: 809–817.
27. **Krause, D. C., and C. B. Mawn.** 1990. Physical analysis and mapping of the *Mycoplasma pneumoniae* chromosome. J. Bacteriol. **172**:4790–4797.
28. **Krawiec, S., and M. Riley.** 1990. Organization of the bacterial chromosome. Microbiol. Rev. **54**:502–539.
29. **Lindquist, S., and E. A. Craig.** 1986. The heat-shock proteins. Annu. Rev. Biochem. **55**:1151–1191.
30. **Loechel, S., J. Inamine, and P.-C. Hu.** 1989. Nucleotide sequence of the deoC gene of *Mycoplasma pneumoniae*. Nucleic Acids Res. **17**:801.
31. **Low, K. B.** 1987. Mapping techniques and determination of chromosome size, p. 1184–1189. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
32. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
34. **McClelland, M.** 1988. Recognition sequences of type II restriction systems are constrained by the G+C content of host genomes. Nucleic Acids Res. **16**:2283–2294.
35. **McClelland, M., J. Jones, Y. Patel, and M. Nelson.** 1987. Restriction endonucleases for pulsed field mapping of bacterial genomes. Nucleic Acids Res. **15**:5985–6005.
36. **Neimark, H. C., and C. S. Lange.** 1990. Pulse-field electrophoresis indicates full-length *Mycoplasma* chromosomes range widely in size. Nucleic Acids Res. **18**:5443–5448.
37. **Ohkubo, S., A. Muto, Y. Kawauchi, F. Yamao, and S. Osawa.** 1987. The ribosomal protein gene cluster of *Mycoplasma capricolum*. Mol. Gen. Genet. **210**:314–322.
38. **Peterson, S. N., N. Schramm, P.-C. Hu, K. F. Bott, and C. A. Hutchison.** 1991. A random sequencing approach for placing markers on the physical map of *Mycoplasma genitalium*. Nucleic Acids Res. **19**:6027–6031.
39. **Pyle, L. E., L. N. Corcoran, B. G. Cocks, A. D. Bergemann, J. C. Whitley, and L. R. Finch.** 1988. Pulsed-field electrophoresis indicates larger-than-expected sizes for *Mycoplasma* genomes. Nucleic Acids Res. **16**:6015–6025.
40. **Pyle, L. E., and L. R. Finch.** 1988. A physical map of the genome of *Mycoplasma mycoides* subspecies mycoides Y with some functional loci. Nucleic Acids Res. **16**:6027–6039.
41. **Pyle, L. E., T. Taylor, and L. R. Finch.** 1990. Genomic maps of some strains within the *Mycoplasma mycoides* cluster. J. Bacteriol. **172**:7265–7268.
42. **Rasmussen, O. F., and C. Christiansen.** 1990. A 23 kb region of the *Mycoplasma* PG50 genome with three identified genetic structures, p. 315–323. *In* G. Stanek, G. H. Cassell, J. G. Tully, and R. F. Whitcomb (ed.), Recent advances in mycoplasmology. Proceedings of the 7th Congress of the International Organization for Mycoplasmology, Baden near Vienna, 1988. Suppl. 20. Fischer Verlag, Stuttgart, Germany.
43. **Razin, S.** 1985. Molecular biology and genetics of mycoplasmas (*Mollicutes*). Microbiol. Rev. **49**:419–455.
44. **Römling, U., D. Grothues, W. Bautsch, and B. Tümmler.** 1989. A physical genome map of *Pseudomonas aeruginosa* PAO. EMBO J. **8**:4081–4089.
45. **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.
46. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNA by pulsed-field gel electrophoresis. Cell **37**:67–75.
47. **Simoneau, P., R. Wenzel, R. Herrmann, and P.-C. Hu.** 1990. Nucleotide sequence of a tRNA cluster from *Mycoplasma pneumoniae*. Nucleic Acids Res. **18**:2814.
48. **Sitzmann, J., and A. Klein.** 1991. Physical and genetic map of the *Methanococcus voltae* chromosome. Mol. Microbiol. **5**:505–513.
49. **Smith, C. L., and C. R. Cantor.** 1986. Pulsed-field gel electrophoresis of large DNA molecules. Nature (London) **319**:701–702.

49a. **Smith, C. L., and G. Condemine.** 1990. New approaches for physical mapping of small genomes. J. Bacteriol. **172**:1167–1172.

50. **Smith, C. L., J. G. Econome, A. Schutt, S. Klco, and C. R. Cantor.** 1987. A physical map of the *Escherichia coli* K12 genome. Science **236**:1448–1453.
51. **Smith, H. O., and M. L. Birnstiel.** 1976. A simple method for DNA restriction site mapping. Nucleic Acids Res. **3**:2387–2398.
52. **Stock, A., J. Mottonen, T. Chen, and J. Stock.** 1987. Identification of a possible nucleotide binding site in CheW, a protein required for sensory transduction in bacterial chemotaxis. J. Biol. Chem. **262**:535–537.
53. **Su, C. J., and J. B. Baseman.** 1990. Genome size of *Mycoplasma genitalium*. J. Bacteriol. **172**:4705–4707.
54. **Su, C.-J., A. Chavoya, and J. B. Baseman.** 1988. Regions of *Mycoplasma pneumoniae* cytadhesin P1 structural gene exist as multiple copies. Infect. Immun. **56**:3157–3161.
55. **Su, C. J., V. V. Tryon, and J. B. Baseman.** 1987. Cloning and sequence analysis of cytadhesin P1 gene from *Mycoplasma pneumoniae*. Infect. Immun. **55**:3023–3029.
56. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.

56a. **Weiss, B., A. Jacquemin-Sablon, T. R. Live, G. C. Fareed, and C. C. Richardson.** 1968. Enzymatic breakage and joining of deoxyribonucleic acid. VI. Further purification and properties of polynucleotide ligase from *Escherichia coli* infected with bacteriophage T4. J. Biol. Chem. **243**:4543–4551.

57. **Wenzel, R., and R. Herrmann.** 1988. Physical mapping of the *Mycoplasma pneumoniae* genome. Nucleic Acids Res. **16**:8323–8336.
58. **Wenzel, R., and R. Herrmann.** 1988. Repetitive DNA elements in *Mycoplasma pneumoniae*. Nucleic Acids Res. **16**:8337–8350.
59. **Wenzel, R., and R. Herrmann.** 1989. Cloning of the complete *Mycoplasma pneumoniae* genome. Nucleic Acids Res. **17**:7029–7043.
60. **Wenzel, R., E. Pirkl, K. Ruland, and R. Herrmann.** 1990. Multiple copies of *Mycoplasma pneumoniae* gene P1 sequences. Zentralbl. Bakteriol. Mikrobiol. Hyg. Suppl. **20**:193–198.
61. **Whitley, J. C., and L. R. Finch.** 1989. Location of sites of transposon Tn*916* insertion in the *Mycoplasma mycoides* genome. J. Bacteriol. **171**:6870–6872.
62. **Whitley, J. C., A. Muto, and L. R. Finch.** 1991. A physical map for *Mycoplasma capricolum* Cal. kid with loci for all known tRNA species. Nucleic Acids Res. **19**:399–400.
63. **Ye, F., F. Laigret, J. C. Whitley, C. Citti, L. R. Finch, P. Carle, J. Renaudin, and J. M. Bové.** 1992. A physical and genetic map of the *Spiroplasma citri* genome. Nucleic Acids Res. **20**:1559–1565.
64. **Yee, T., and M. Inouye.** 1982. Two-dimensional DNA electrophoresis applied to the study of DNA methylation and the analysis of genome size in *Myxococcus xanthus*. J. Mol. Biol. **154**:181–196.
65. **Yogev, D., S. Sela, H. Bercovier, and S. Razin.** 1990. Nucleotide sequence and codon usage of the elongation factor Tu(EF-Tu) gene from *Mycoplasma pneumoniae*. Mol. Microbiol. **4**:1303–1310.
66. **Zurawski, G., and S. M. Zurawski.** 1985. Structure of the *Escherichia coli* S10 ribosomal protein operon. Nucleic Acids Res. **13**:4521–4526.