
Cloning of the complete *Mycoplasma pneumoniae* genome

Rainer Wenzel and Richard Herrmann

Department of Microbiology, University of Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, FRG

Received July 25, 1989; Accepted 11 August, 1989

ABSTRACT

The complete genome of *Mycoplasma pneumoniae* was cloned in an ordered library consisting of 34 overlapping or adjacent cosmids, one plasmid and two λ phages. The genome size was determined by adding up the sizes of either the individual unique EcoRI restriction fragments of the gene bank or of the XhoI fragments of genomic *M. pneumoniae* DNA. The values from these calculations, 835 and 849 kbp, are in good agreement. An XhoI restriction map was constructed by identifying adjacent DNA fragments by probing with selected cosmid clones.

INTRODUCTION

During the last few years techniques which permit sequencing at a megabase scale have been developed and improved. Therefore the DNA sequence analysis of bacterial genomes now is a realistic project for smaller research teams to perform within a reasonable time.

Among the smallest bacterial genomes are those of the Mollicutes ranging from 700 - 1600 kbp (1). Mainly because of the relative small size of its genome - 800 kbp (1) when compared to 4.7 Mbp for *Escherichia coli* (2) - we have chosen *Mycoplasma pneumoniae* for defining the minimal genetic requirement of a self-replicating cell. *M. pneumoniae*, the causative agent of atypical pneumonia in humans, may also be useful for analyzing the interaction between a human pathogenic bacterial surface parasite and its eukaryotic host cell.

A genome of 800 kbp may code for as many as 600-800 proteins (3). It should be possible to map and identify a considerable proportion of the corresponding genes with known functions by techniques like cross hybridization using DNA or RNA probes derived from conserved bacterial genes, or random DNA sequencing

and computer aided search for homology to cloned and sequenced bacterial genes. In addition, the functions of certain genes can be elucidated by complementation of mutants by cloned M. pneumoniae DNA in homologous and heterologous systems (4). This publication describes the construction of a complete cosmid gene library of the M. pneumoniae genome in E. coli to provide the basis for comprehensive genome analysis including DNA sequence determination.

MATERIALS AND METHODS

Strains and plasmids, growth of M. pneumoniae, DNA isolation, radioactive labelling, cosmid cloning and mapping procedures were described elsewhere (5).

Restriction enzymes

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

Lambda cloning

M. pneumoniae DNA was digested to completion with endonuclease XhoI and ligated to XhoI cleaved λ ZAPII arms (6) under conditions recommended by the supplier (Stratagene cloning systems). In vitro packaging was performed using the Gigapack plus extract (Stratagene). E. coli XL1-Blue cells (6) were infected with the phage and selected on NZY plates containing X-Gal and IPTG.

Recovery of DNA fragments from agarose gels

Restriction fragments were eluted from agarose gels using the Gene clean kit (Bio 101 Inc.) according to the instructions of the supplier.

Gel electrophoresis and Southern blotting

Separation of high-molecular weight M.pneumoniae DNA was performed by field inversion gel electrophoresis (FIGE) using an EMBL CS130 FIGE/OFAGE apparatus with 1% agarose gels (15x 10cm). Approximately 5 μ g XhoI digested M.pneumoniae DNA was loaded into a 10cm wide slot (see Fig.5). The gel was run with a pulse time ranging from 6 to 7 seconds for 48h at 80V in 0.5x TBE (1x TBE is 90mM Tris, 90mM boric acid, 2mM EDTA) at 4°C.

Smaller restriction fragments (up to 40 kbp) were separated by direct electrophoresis on 0.4% agarose gels (23x 20cm). Approximately 25 μ g DNA was loaded into a 15cm wide slot. The gel was

run for 96h at 70mA in running buffer (36mM Tris, 10mM Na₂EDTA, 30mM NaH₂PO₄) at 4°C (7). The buffer was exchanged daily.

Following ethidium bromide staining, the gels were transferred to nitrocellulose (BA85, Schleicher and Schuell) or nylon membranes (Biodyne, Pall) according to the instructions of the suppliers. The filters were cut into 0.5cm wide strips for hybridization probing with the cosmids which were ³²P-labelled by nick-translation (ca. 10⁸ cpm/μg DNA).

Hybridization

After pre-hybridizing the nitrocellulose (or nylon) filters in 50% formamide, 5x SSC (1x SSC: 150 mM NaCl, 15 mM sodium citrate pH 7.2), 5x Denhardt's solution (1x Denhardt's solution: 0.02 % Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 0.1% SDS and 100μg/ml denatured herring sperm DNA at 37°C for at least 4 hours, the hybridization was performed in 50% formamide, 5x SSC, 5x Denhardt's, 0.1% SDS and 10% dextran sulphate by adding the appropriate ³²P-labelled probe and incubating at 37°C overnight. The filters were washed three times for 30min with 1x SSC, 0.1% SDS at 68°C. Autoradiographic signals were obtained by exposing X-ray films (Kodak) overnight to the filters.

Preparation of the high-molecular-weight size standard

40μg of EcoRI linearized DNA from pSC101 (8) or pUC18 (9) was ligated at a concentration of 500μg/ml overnight at 16°C with 6 Weiss units of T4 DNA ligase. In the case of pUC18 the ligation product was partially digested with EcoRI to reduce the size of the highly polymerized concatamers to the range of 2.6 - 40 kbp. Approximately 0.5μg of DNA was loaded into a 5mm slot.

RESULTS

Cloning of the complete genome

In a previous publication we reported the cloning of about 90% of the M. pneumoniae genome (5).

At that time, attempts to clone the residual 10% had failed despite the screening of several thousand clones, a number more than sufficient from a statistical point of view (10). Since then, extended screening has led to the isolation of three more overlapping cosmids, pcosMPK4, pcosMPG7 and pcosMPGT9, however the two clones at the end, pcosMPGT9 and pcosMPE7, were neither

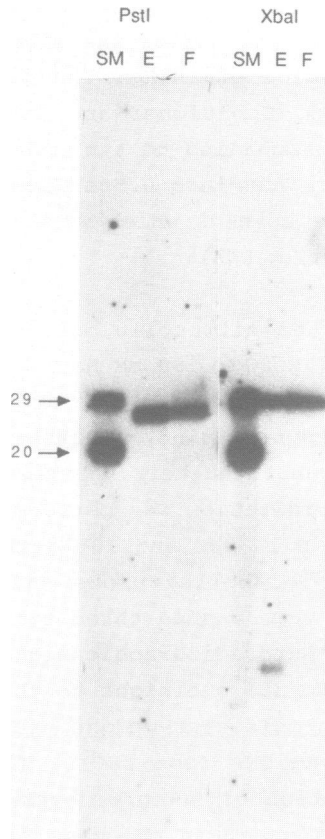


Figure 1: Restriction analysis of the region between pcosMPGT9 and pcosMPE7. *M. pneumoniae* DNA was digested with PstI and XbaI, separated on a 0.4% agarose gel, blotted and hybridized to probes which were specific for the ends of the cloned region. SM = size marker (λ EMBL3x EcoRI; the cos site containing fragments of 9 and 20 kbp form the 29 kbp band); E = E7E (subclone of E7, see Fig.3); F = XhoI restriction fragment V (specific for the end of pcosMPGT9, see Fig.3). Sizes are in kbp.

overlapping nor adjacent. At this point of the analysis it seemed unprofitable to continue just screening more clones. We decided rather to characterize the remaining gap between pcosMPGT9 and pcosMPE7 first, using restriction analysis of genomic *M. pneumoniae* DNA to find one or more restriction fragments which would link the terminal cosmids. To probe the genomic restriction fragments on Southern blots it was unsatisfactory to use the

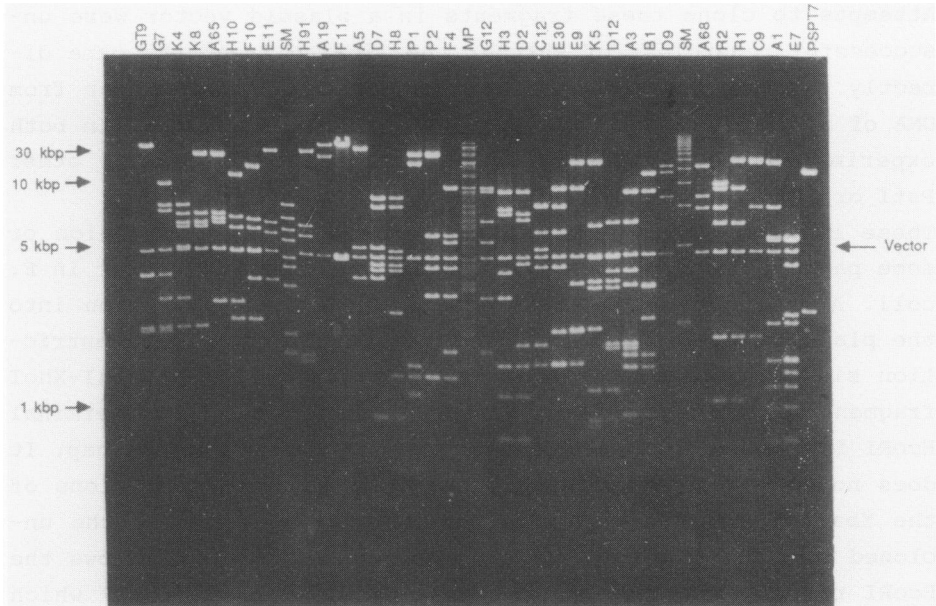


Figure 2: Restriction analysis of the 34 cosmids plus 1 plasmid (pSPT7) covering the complete genome of *Mycoplasma pneumoniae* except for 5.3 kbp cloned in λ (see Fig.3). The cosmids were digested with EcoRI and pSPT7 with XbaI and XhoI and the digests were separated on a 0.8% agarose gel. SM = size marker (left: λ BstEII; right: pUC18 polymers); MP = *M.pneumoniae* DNA x EcoRI

complete cosmids as probes since both contain repetitive elements (11) making it necessary to use DNA with sequences from only the gap proximal ends of the cosmid inserts.

For the end of pcosMPGT9, the probe used was an 8 kbp genomic DNA XhoI DNA fragment (fragment V, see below) which could be readily isolated from an agarose gel. It has been shown that the 6.7 kbp XhoI-EcoRI gap proximal fragment of pcosMPGT9 is part of this XhoI fragment and that the residual 1.3 kbp reaches the gap region. For a probe from pcosMPE7, its gap proximal 2.7 kbp EcoRI fragment was subcloned resulting in E7E. Southern blots of genomic DNA digested with several enzymes (BamHI, EcoRI, XhoI, HindIII, PstI and XbaI) were screened with these probes. Among them, a 27 kbp PstI fragment and a 29 kbp XbaI fragment reacted positively with both probes (Fig.1), strongly indicating that these fragments linked the terminal cosmids.

Attempts to clone these fragments in a plasmid vector were unsuccessful regardless of whether the ligated fragments came directly from a restriction digest of total genomic DNA or from DNA of appropriate size fractionated in an agarose gel. In both experiments positive controls came from the cloning of other PstI or XbaI fragments.

These results supported the conclusion that the gap region or some part of it might not be clonable in a plasmid vector in *E. coli*. Attempts to clone subfragments from the gap region into the plasmid pSB02 (12) which contains unique XbaI/XhoI restriction sites resulted in pSPT7 which contains a 13 kbp XbaI-XhoI fragment. This fragment overlaps 1.9 kbp of the 2.7 kbp terminal EcoRI fragment of pcosMPE7 and reaches 11.1 kbp into the gap. It does not bridge pcosMPGT9 and pcosMPE7. Since the positions of the XbaI restriction sites in the cosmids were known, the uncloned region could not be larger than 5.3 kbp. Fig.2 shows the EcoRI restriction pattern of the 34 cosmids and of pSPT7 which was XbaI/XhoI digested.

In the construction (described below) of an endonuclease XhoI restriction map of the *M. pneumoniae* genome based on the cosmid collection we correlated the XhoI fragments in genomic DNA digests with those in the cosmids. Only one, a 4 kbp fragment (fragment X, see below), could not be attributed to corresponding cosmids. Therefore this fragment was assumed to be located in the gap region between XhoI fragments A and V. This assumption was verified by probing Southern blots of XbaI and PstI digested *M. pneumoniae* DNA with the genomic 4 kbp XhoI fragment as a labelled probe. This probe recognized only the 27 kbp PstI and 29 kbp XbaI fragments.

Attempts to clone the XhoI fragments V and X into either a low or high copy number plasmid failed. These results, indicating that the fragments can not be cloned into an *E. coli* plasmid vector, suggest that the presence of fragment X and a 1.3 kbp part of fragment V in an *E. coli* cell is inhibitory to its continued growth. Therefore we used a vector system in which cell survival is not required and the chance of cloning the missing fragments should be greater. Using the λ ZAPII system (6), which has the advantage that its maximum cloning capacity of 10 kbp would ac-

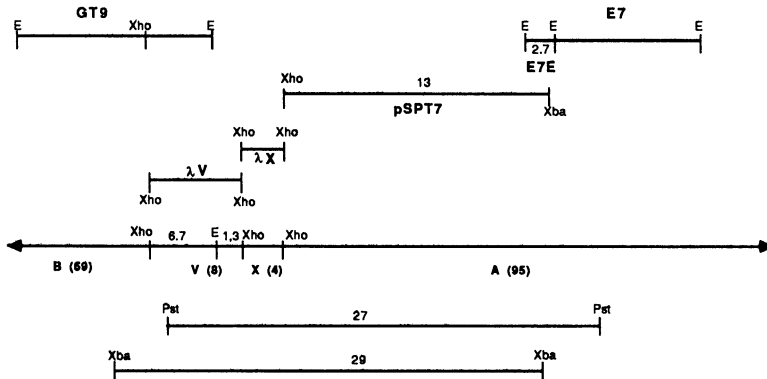


Figure 3: Restriction map of the pcosMPGT9/pcosMPE7 region. The numbers indicate the size of the fragments in kbp. The map shows the position of the λ clones λ V, λ X and of pSPT7. A, B, V and X represent the corresponding XhoI restriction fragments (see Table 1). E = EcoRI

cept only the four smaller XhoI fragments, it was possible to clone fragments V and X to give λ V and λ X. This closed the gap in the region between pcosMPGT9 and pSPT7. Fig.3 shows the restriction map and the position of the clones in the region of the original gap.

Fig.4 shows the final alignment of the 34 cosmids, one plasmid and two λ recombinants which cover the complete genome. Compared to the recently published cosmid map (5) one clone - pcosMPB11 - had to be eliminated since it proved to be a mixture of two DNA regions, one overlapping with pcosMPR2 in two EcoRI fragments and the other with pcosMPGT9 in three EcoRI fragments (see also discussion).

Size of the *M.pneumoniae* genome

Our data allow estimations of the size of the *M. pneumoniae* genome by calculation of the sum of the sizes for the fragments generated by restriction endonuclease EcoRI and XhoI digests.

For one calculation every unique EcoRI restriction fragment which has been cloned in the cosmid collection was defined and its size was measured by agarose gel electrophoresis using appropriate size markers. A total number of 145 fragments was counted from an agarose gel like in Fig.2. For an exact size determination, different agarose concentrations were applied: a

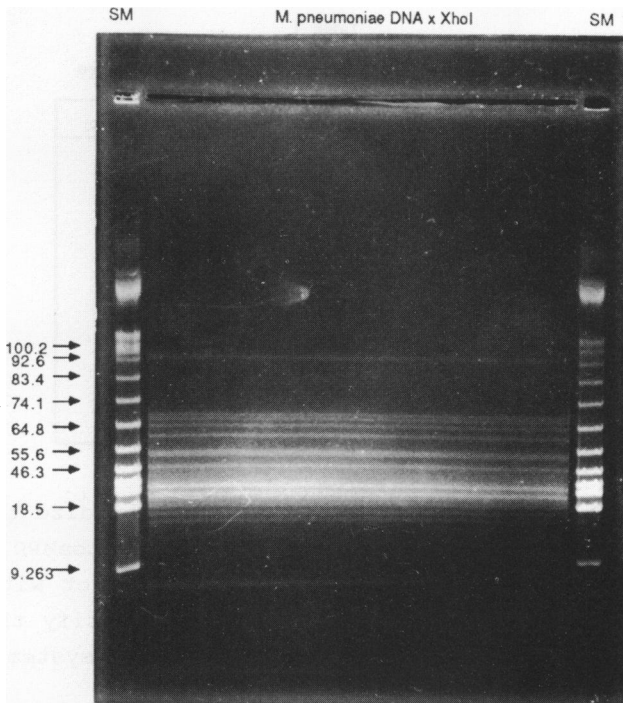


Figure 5: FAGE separation of large fragments from XhoI digested *M. pneumoniae* genomic DNA on a 1% agarose gel. After determining the size of the fragments, the gel was blotted and used for probing (Fig.6, part A). SM = size marker (pSC101 x EcoRI polymers). Sizes are in kbp.

Endonuclease XhoI restriction map of the *M. pneumoniae* genome

A XhoI restriction map was constructed mainly by means of the cosmid collection. Except for two cosmids (pcosMPE30, pcosMPD12), the XhoI sites were detected by comparing an EcoRI restriction digest with an EcoRI/XhoI double digest of the cosmids. The disappearance of an EcoRI fragment and the appearance of two smaller DNA fragments clearly demonstrated an XhoI site. In case of pcosMPE30 and pcosMPD12, nucleases BamHI and HindIII, respectively, were used because the EcoRI and EcoRI/XhoI restriction patterns were indistinguishable in a standard agarose gel.

Cosmids containing one or more XhoI sites were ^{32}P -labelled by nick-translation and used as probes for identifying adjacent XhoI fragments in a Southern blot of XhoI digested genomic *M.*

Table 1XhoI restriction fragments of *Mycoplasma pneumoniae*

Fragment	Size (kbp)	Fragment	Size (kbp)
A	95	N	24.5
B	69	O	24
C	68	P	22.5
D	66	Q	21.5
E	62	R	21
F	55	S	19.5
G	55	T	18
H	49	U	15
I	34	V	8
J	34	W	5
K	27.5	X	4
L	25	Y	1.3
M	25	Total	849

pneumoniae DNA. Cosmids with only one site hybridize to two XhoI fragments (for instance pcosMPD12, pcosMPD9, pcosMPD7 or pcosMPG12) whereas clones containing two sites react with 3 fragments, like pcosMPH3 (Fig.6). In order to identify the genomic XhoI fragments unambiguously, two different gel systems and different agarose concentrations had to be used.

In a number of hybridizations several lighter bands appear (compare the cosmids pcosMPR2, pcosMPP1, pcosMPA5, pcosMPA19 and pcosMPE7 in Fig.6). These cross-reactions are caused by one or more of at least five different repetitive DNA sequences (11,15, 16) which are present in the genome. In case of pcosMPB1 (two sites) only one strong band shows up. According to our interpretation this band represents two fragments of the same size. The missing third fragment is the smallest 1.3 kbp fragment Y which had run off the gel. Like all four fragments below 10 kbp, it was identified by direct comparison of XhoI digests of cosmids and genomic DNA in an agarose gel. Fig.7 shows the XhoI restriction map of the complete *M. pneumoniae* genome.

An anomaly which arose in this analysis was that the cosmid digests gave a total of 26 XhoI sites, including the one between λV and λX , whereas the genomic DNA digests showed only 25. XhoI digests of cosmids pcosMPA68 and pcosMPD9 showed a 4 kbp XhoI fragment which does not have its counterpart in the genomic digest. After isolating the fragment from an agarose gel and using

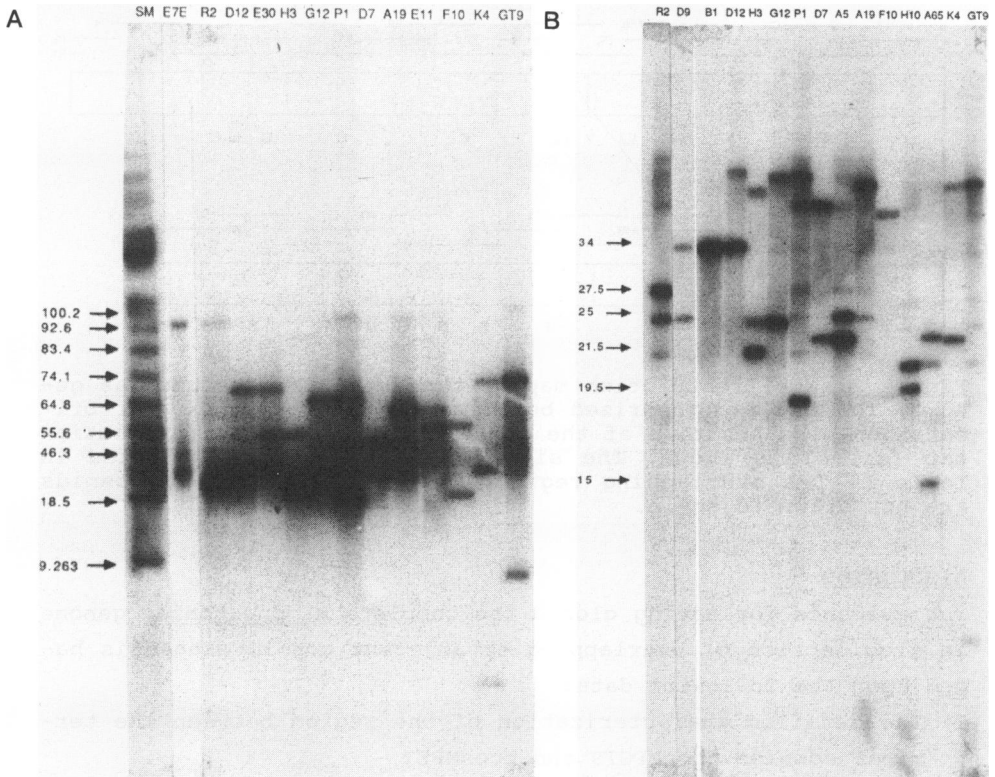


Figure 6: Southern blots of *M. pneumoniae* DNA digested with XhoI. In order to separate fragments larger than 40 kbp, a field inversion gel (FIGE) was run (part A). Fragments smaller than 40 kbp were separated by the gel system described in material and methods (part B). Both filters were cut into stripes and hybridized to a set of selected ^{32}P -labelled cosmids. The resulting XhoI restriction map is shown in Figure 7. SM = size marker (pSC101xEcoRI polymers). Sizes are in kbp.

it as a ^{32}P -labelled probe in a Southern blot against XhoI restricted *M. pneumoniae* DNA, it reacted only with the I/J double band (see Fig.7). This indicates strongly that either I or J possesses an additional XhoI site, which is protected in genomic DNA probably by methylation in vivo.

Our approach to use aligned cosmids from the gene library for constructing a restriction map appeared to be fast and simple and yielded unambiguous results.

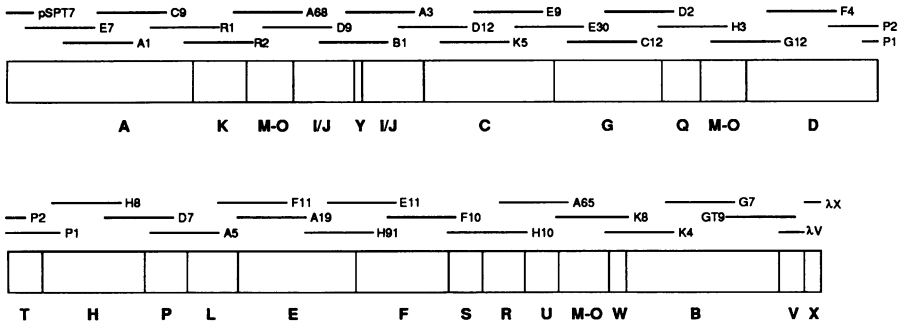


Figure 7: XhoI restriction map of the *Mycoplasma pneumoniae* genome. The map is linearized between fragments A and X. For orientation, the position of the clones relative to the restriction map (sites) is shown. The size of the fragments is listed in Table 1. The overlapping regions between neighbouring cosmids are not drawn to scale.

DISCUSSION

The evidence for having cloned the complete *M. pneumoniae* genome in a collection of overlapping or adjacent cosmid clones is based upon the following data:

- 1) the detailed characterization of the region between the terminal cosmids pcosMPGT9 and pcosMPE7,
- 2) the endonuclease XhoI restriction map and the identification of the corresponding restriction sites within the cosmids and
- 3) the consistency of the two independent genome size determinations, one based on data from the cloned DNA and the other on data from genomic DNA only.

The reasons for the difficulties in cloning the region between the terminal cosmids are still unknown. Since not only high copy number plasmids but also low copy number vectors like pSC101 failed to replicate with the two XhoI fragments we assume that they might code for sequences toxic to growth of *E. coli*. This assumption is strongly supported by the fact that these fragments could be cloned in phage lambda for which replication is not dependent on survival of the host bacterium which is usually lysed as a result of infection. DNA sequence analysis of the region in question may help to explain the cloning difficulties.

As noted above, we had to delete the clone pcosMPB11 from our previously published map (5), since it contained two DNA seg-

ments from different locations in the genome. It was one of three cosmid clones - which we reported in our previous paper - as not having other, independently isolated, identical or overlapping clones. In retrospect, the best argument for having cloned a contiguous DNA fragment instead of a mixture of fragments is the isolation of other identical or extensively overlapping clones.

The introduction of pulsed field gel electrophoresis (PFGE) (17) gives a powerful method for the size determination of complex genomes. Originally, the size of various Mollicutes genomes has been determined by electron microscopy (18) or by DNA renaturation kinetics (1). The results indicated two size classes of about 800 kbp and 1600 kbp length. Recently, Pyle et al. (19) published results based on PFGE showing larger genome sizes than anticipated for species in the former class, for instance 1140 kbp for Mycoplasma hyopneumoniae, J strain or 1240 kbp for Mycoplasma mycoides, subsp. mycoides, Y strain.

The discrepancies have not yet been resolved, though recently the validity of PFGE for AT-rich DNA sequences (as most of the Mollicutes genomes are) has been questioned by Maniloff and Poddar (20,21). In case of M. pneumoniae our size determination agrees fairly well with the value of 735 kbp derived from renaturation kinetics. It should be mentioned that, as an exception among the Mollicutes, M. pneumoniae has a GC-content of about 41% (22). Therefore M. pneumoniae DNA should behave "normally" in a PFGE analysis. The size of the M. pneumoniae genome measured as a single, or two, fragments in PFGE was in the range of 800-850 kbp (L. Pyle and L. Finch, pers. communication) in good agreement with our data.

The genome size determination by addition of the sizes of individual restriction fragments is straight forward. Potential sources of error could arise for calculations based on the restriction analysis of the cloned cosmids. Firstly, fragments in overlapping clones showing a very similar size might not be recognized as being different (for instance the largest EcoRI fragments in pcosMPG12 and pcosMPF4 (Fig.2) are not identical). Secondly, it is sometimes difficult to differentiate between a double or a triplet band (for instance the heavy band in pcosMP-

E7 (Fig.2) just above the vector and the band in pcosMPD12 containing the vector band are triplets). Smaller EcoRI fragments (<300 bp) have been neglected but they are insignificant in the overall size calculation.

The advantage of the calculation from the endonuclease XhoI genomic restriction analysis is the limited number of fragments so that they can be distinguished for counting on a gel. An observation that sometimes the larger fragments seem not to be present in equimolar amounts as judged from the staining is very probably due to mechanical shearing of large DNA fragments (> 50 kbp). This would be minimized by release and digestion of the DNA in agarose blocks (17).

Size determination for fragments larger than 40 kbp requires use of PFGE gels because of the rapidly decreasing resolution in conventional gel systems. In both systems it is also indispensable to have polymeric size markers from sequenced DNA to give many reference points at close and defined distance. Even in the worst case the error is then less than the difference between two reference points. Therefore we believe that the error in our calculation is less than 5% (or about 40 kbp). For a more precise determination, the DNA sequencing of the complete genome will be necessary.

ACKNOWLEDGMENT

The excellent technical work of Elsbeth Pirkl is gratefully acknowledged. We thank L. Finch for critical reading of the manuscript. This work was supported by the BMFT (BCT 0381/5).

REFERENCES

1. Bak, A.L., Black, F.T., Christiansen, C., Freundt, E.A. (1969) *Nature* **224**: 1209-1210
2. Kohara, Y., Akiyama, K., Isono, K. (1987) *Cell* **50**: 495-508
3. Morowitz, H.J. (1984) *Isr. J. Med. Sci.* **20**: 750-753
4. Loechel, S., Inamine, J.M., Hu, P.C. (1989) *Nucl. Acids Res.* **17**: 801
5. Wenzel, R. and Herrmann, R. (1988) *Nucl. Acids Res.* **16**: 8323-8336
6. Short, J.M., Fernandez, J.M., Sorge, J.A., Huse, W.D. (1988) *Nucl. Acids Res.* **16**: 7583-7600
7. Darai, G., Zöller, L., Matz, B., Delius, H., Speck, P.T., Flügel, R.M. (1982) *J. Bacteriol.* **150**: 788-794
8. Bernardi, G. and Bernardi, F. (1984) *Nucl. Acids Res.* **12**: 9415-9426

-
9. Yanisch-Perron, C., Vieira, J., Messing, J. (1985) *Gene* 33: 103-119
 10. Clarke, L. and Carbon, J. (1979) *Meth. Enzymol.* 68: 396-408
 11. Wenzel, R. and Herrmann, R. (1988) *Nucl. Acids Res.* 16: 8337-8350
 12. Nassal, M., Mogi, T., Karnik, S.S., Khorana, H.G. (1987) *J. Biol. Chem.* 262: 9264-9270
 13. McClelland, M., Jones, R., Patel, Y., Nelson, M. (1987) *Nucl. Acids Res.* 15: 5985-6005
 14. Carle, G.F., Frank, M., Olson, M.V. (1986) *Science* 232: 65-68
 15. Su, C.J., Chavoya, A., Baseman, J.B. (1988) *Infect. Immun.* 56: 3157-3161
 16. Wenzel, R., Pirkl, E., Ruland, K., Herrmann, R. (1989) *Zentralbl. Bakteriolog. Mikrobiolog. Hyg. 1 Abt. Orig. A* (in press)
 17. Schwartz, D.C. and Cantor, C.R. (1984) *Cell* 37: 67-75
 18. Bode, H.R., Morowitz, H.J. (1967) *J. Mol. Biol.* 23: 191-199
 19. Pyle, L.E., Corcoran, L.N., Cocks, B.G., Bergemann, A.D., Whitley, J.C., Finch, L.R. (1988). *Nucl. Acids Res.* 16: 6015-6025
 20. Maniloff, J. (1989) *Nucl. Acids Res.* 17: 1268
 21. Poddar, S.K. and Maniloff, J. (1989) *Nucl. Acids Res.* 17: 2889-2895
 22. Carle, P., Saillard, C., Bove, J.M. (1983) Determination of guanine plus cytosine content of DNA. In: Razin, S. and Tully, J.G. (eds) *Methods in Mycoplasmaology, Vol. I.* Academic Press, Inc. New York, pp301-308