Complete Physical Map of the *rfb* Gene Cluster Encoding Biosynthetic Enzymes for the O Antigen of *Salmonella* typhimurium LT2

HIMANSHU N. BRAHMBHATT, 1* PAUL WYK, 2 NEIL B. QUIGLEY, 1 AND PETER R. REEVES1

Department of Microbiology, University of Sydney, Sydney, New South Wales 2006, and Department of Microbiology and Immunology, University of Adelaide, Adelaide, South Australia 5001, Australia

Received 20 July 1987/Accepted 23 September 1987

Biosynthesis of the Salmonella typhimurium LT2 O antigen is encoded by genes which map in the rfb cluster. The cloning and restriction enzyme analysis of part of this cluster have been described previously (H. N. Brahmbhatt, N. B. Quigley, and P. R. Reeves, Mol. Gen. Genet. 203:172–176, 1986). The entire rfb gene cluster has now been cloned, and a detailed restriction enzyme map has been constructed which has enabled us to map the approximate positions of individual rfb genes.

Lipopolysaccharides form a major component of the outer membranes of gram-negative bacteria. They are composed of three parts: the innermost lipid A, the middle oligosaccharide core, and the outermost linear polymer of repeating oligosaccharide units, the O antigen (9). The O-specific polysaccharide shows high structural variation within the genus Salmonella (7, 9, 10), whereas the core and lipid A remain constant. This variation in O-antigen structure in the genus Salmonella provides the basis for a major serological typing scheme, and its chemical nature has been studied in great detail (5, 11). The genetic basis of O-specific polysaccharide biosynthesis has been studied almost exclusively by using Salmonella strains, generally with S. typhimurium LT2 as a tool. The genes coding for the biosynthesis and assembly of O antigen are located mostly in the rfb gene cluster at map position 42 (21). A partial gene order for this cluster has been established by using a series of mutants with deletions extending from his into rfb (17). The basic repeating oligosaccharide unit of the S. typhimurium O antigen is a tetramer composed of abequose, mannose, rhamnose, and galactose. This unit is assembled from appropriate nucleoside diphosphate derivatives of the monosaccharides as an undecaprenol phosphate-linked intermediate before polymerization, possible modification, and translocation to the core. In a previous paper (1), we reported the cloning and physical mapping of a part of the S. typhimurium LT2 rfb gene cluster and determined the approximate location of the rfbK and rfbM genes. In this paper, we report the cloning of the complete rfb gene cluster and give a detailed physical map. We have also localized all deletion endpoints which are known to lie in the rfb region, thereby giving the approximate positions of most of the known genes on the map.

MATERIALS AND METHODS

Bacteria, bacteriophages, and plasmids. S. typhimurium LT2 and Escherichia coli K-12 strains used are listed in Table 1. P9003 is wild type for rfb and is often referred to as wild type in the text. An S. typhimurium gene bank constructed in the vector $\lambda 1059$ (6) was a kind gift from Russel Maurer (15).

Enzymes and radiochemicals. Restriction enzymes, DNA polymerase I, bacterial alkaline phosphatase, and T4 DNA ligase were purchased from Promega Biotec, Madison, Wis.,

Boehringer Mannheim Biochemicals, Indianapolis, Ind., New England BioLabs, Inc., Beverly, Mass., and BRESA, Adelaide, Australia. [α -³²P]dCTP was obtained from BRESA and Amersham Corp., Arlington Heights, Ill.

DNA techniques. The methods used to prepare plasmid and high-molecular-weight chromosomal DNA to digest with restriction enzymes and to ligate restriction fragments were those described previously (1). Recombinant $\lambda 1059$ phage DNA was purified as described by Karn et al. (6). Competent cells were prepared and transformed with DNA by the method of Mandel and Higa (12). Agarose gel electrophoresis, radioactive labeling of DNA, in situ DNA hybridization, and autoradiography were performed as described by Maniatis et al. (13). EcoRI-digested bacteriophage SPP1 DNA (20), BglII-digested λ DNA, and MspI-digested pUC9 DNA were used as molecular size standards.

Detailed restriction enzyme mapping. A detailed restriction enzyme map was made for the *rfb* region. The strategy used in mapping each clone was as follows. The plasmid DNA

TABLE 1. Bacterial strains

Strain	Genotype
S. typhimurium LT2	
	hsdL6 trpC2 nml H1b fla-66 H2-enx
	rpsL120 xyl-1404ilv-452 metE551
	metA22 hsdSA29
P9073	$\dots \Delta (hisI-rfbK)642^a$
	$\dots \Delta$ (his D-rfb K) 1462 a
P9078	
P9033	
P9031	
P9034	
P9030	
P9074	
	$\dots \Delta$ (hisO-galF)809 a,b
P9051	
	$\ldots \Delta$ (hisO-metG)388 c
E. coli K-12	
P334 (C600)	thrl leu6 tonAl lacYl supE44 thi-l

^a Endpoint within rfb is shown in Fig. 1.

^{*} Corresponding author.

b Note that galF is within the rfb cluster deletion.

^c Note that this includes all of the *r/b* cluster. P9003 was obtained from I. Beacham, and other LT2 strains were obtained from H. Nikaido (17). C600 was from our collection.

was cleaved in single digests with each of 28 enzymes to determine the frequency of cleavage. Unique sites could be mapped in relation to vector polylinker sites, and for enzymes with two or three sites the two terminal locations could be mapped with respect to polylinker and cloning sites. The third site could then be mapped in relation to other defined sites. This rough map was then refined by doing a series of double digests to accurately measure the sizes of the small fragments between adjacent sites and to determine the order where it was ambiguous. Sufficient double digests were then done to enable us to see each fragment, defined by sites adjacent on the map, and to estimate its size. All such estimates were the average of at least two replicates with the most appropriate gels. The sizes of the larger fragments were also accurately measured to confirm the map as derived.

Deletion endpoint mapping. The deletion endpoints in *rfb* were mapped by analysis of the autoradiographs produced by Southern hybridization of various cloned *rfb* DNA probe fragments, against restriction enzyme digests of chromosomal DNA from the deletion strains. The restriction enzymes used were those for which sites were known to exist adjacent to or within the region spanned by the probe. Adjacent to every deletion DNA digest track, a track con-

taining wild-type DNA digested with the same restriction enzyme was included as a positive control. Comparison of band sizes in the test and control tracks on the autoradiographs defined specific restriction enzyme sites as lost or retained in the deletion strain DNA. The deletions run from within or to the left of his and into rfb (17), and hence the restriction enzyme sites to the right (distal to the his operon) of a deletion endpoint were expected to be present, and those to the left (proximal to the his operon) of this endpoint were expected to be absent.

Cosmid cloning. High-molecular-weight chromosomal DNA from strain P9003 was partially digested with Sau3A1 by the method of Maniatis et al. (13). Cosmid vector pcos2EMBL was used as described by Poustka et al. (18) for cosmid libraries, but without the use of phosphatase on the chromosomal DNA.

RESULTS

Cloning of the *rfb* gene cluster. The *MluI-HindIII* insert fragment (from approximately position 4.52 to position 5.76 [note that the *MluI* site is just outside the *rfb* region cloned]) of plasmid pPR259 (Fig. 1) (2) was used as a probe to screen

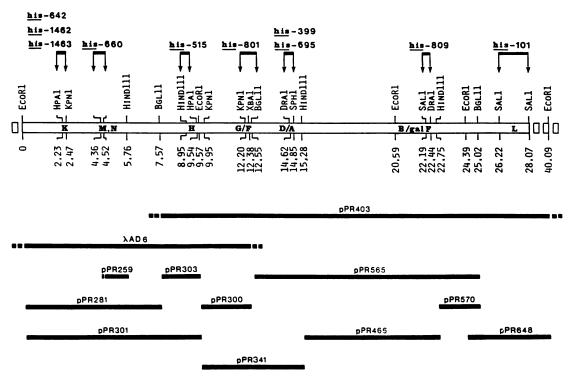


FIG. 1. Location of his-rfb deletion endpoints within the rfb cluster of S. typhimurium LT2. The his and gnd genes are located to the left of the region shown, and the deletions all extend from within or to the left of his through gnd, into rfb. The names of specific deletions (see also Table 1) are positioned above short intervals of rfb DNA. These intervals are defined by specific restriction sites indicated by paired arrows and contain the rfb endpoints of the deletions indicated. Other restriction sites shown were used for subcloning rfb DNA segments and are discussed in the text. The rfb endpoint of the his-660 deletion was mapped previously (1). The rfb genes known to be interrupted by each deletion (17) are in boldface type on the thick open bar representing the chromosome. Genes separated by a slash are those for which the relative order has not been determined. Heavy lines represent specific rfb DNA segments which have been subcloned into lambda or plasmid vectors. Plasmids pPR259 and pPR281 have been described previously (1); other fragments were subcloned as follows. λAD6 was selected from a λ1059 bank by using the MluI-HindIII (Fig. 1) (2) fragment of pPR259 as a probe; pPR300 has an EcoRI-XbaI fragment from λAD6 in pJRD158b (2, 4); pPR301 has an EcoRI fragment from λAD6 in pEMBL8 (3); pPR340 (not shown) has the 11.02 EcoRI fragment (positions 9.57 to 20.59) of pPR403 in pEMBL8; pPR341 has an EcoRI-HindIII fragment (positions 9.57 to 20.59) of pPR340 in pPR328 (19); pPR403 is a cosmid clone selected from the pcos2EMBL bank; pPR465 has a HindIII fragment from pPR403 in pUC9 (23); pPR556 has a BglII fragment from pPR403 in pJRD158b; pPR570 has the HindIII-BglII fragment from pPR565 in pUC9; and pPR648 has the EcoRI fragment (15.7 kb) from pPR403 in pKO1 (16). The units are in kilobases, and position 0 is arbitrarily chosen to signify the his-proximal end of rfb. The exact location of the rfb cluster terminus has not been defined, but is thought to lie close to position 0 (1).

100 BRAHMBHATT ET AL. J. BACTERIOL.

the $\lambda 1059$ bank of S. typhimurium DNA by plaque hybridization. Phages from several positive plaques were purified, and their DNA was isolated, nick translated, and used in dot blot hybridization against chromosomal DNA from all deletion strains (Table 1) to identify which cloned DNA extended farthest into the rfb region. The isolate selected for further analysis, $\lambda AD6$, hybridized to DNA from the wild-type strain P9003 and to DNA from strains carrying deletions which extend to the rfbF and rfbG genes (Fig. 1). The insert in $\lambda AD6$ presumably carries rfbK, rfbM, rfbN, rfbH, and possibly rfbG and rfbF also. However, we do not have a functional test for the expression of these Salmonella genes by this lambda clone.

Plasmids pPR300 and pPR301 (Fig. 1) were obtained from λAD6, and pPR403 was selected from the cosmid bank by using a *Kpn*I fragment (9.95 to 12.20) of pPR300 as probe. Several cosmid clones hybridized with this probe, and preliminary analysis showed that pPR403 extended farthest to the right: plasmids pPR341, pPR465, pPR565, pPR570, and pPR648 were obtained from it in a series of steps (Fig. 1).

Sufficient restriction map analysis was done as the subcloning proceeded to facilitate the subcloning itself and to start the mapping of deletion endpoints (see below) so that progress along the *rfb* gene cluster could be monitored. The deletion endpoint data also indicated that *rfb* did not extend beyond the region covered by pPR648.

Southern blot analysis of pPR403 and wild-type chromosome DNA, with pPR341, pPR465, and pPR570 as probes, was carried out to confirm that the general pattern of *EcoRI* sites was the same on both clone and chromosome and hence that there had been no major structural rearrangement during cloning. This was further confirmed for a variety of sites during the analysis of deletion endpoints (see below), which involved extensive Southern blot analysis of chromosomal DNA, with all results being consistent with the absence of structural rearrangements.

Detailed restriction enzyme analysis of the rfb gene cluster. We have already published a detailed restriction enzyme map of rfb DNA between positions 0 and 7.57 (1). The complete physical map presented in Fig. 2 was established up to position 25.02 by performing detailed restriction enzyme analysis on plasmids pPR341, pPR465, and pPR570 (see Materials and Methods for details of the mapping procedure). Plasmids pPR303 and pPR648 were analyzed only on a preliminary basis, since they carry DNA which is either being sequenced in our laboratory (pPR303) or lies predominantly beyond the last deletion endpoint (his operon distal), that of $\Delta(his-rfb)101$, known to lie in or beyond rfbL.

Localization of deletion endpoints. Plasmid pPR648 was

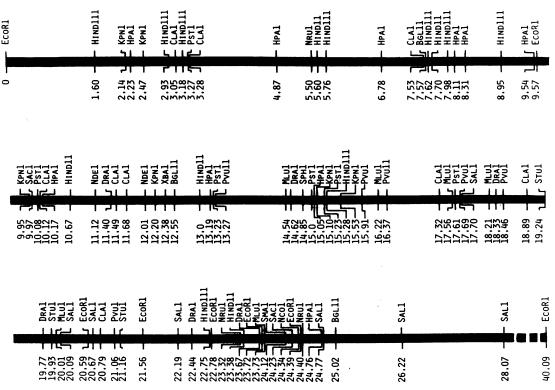


FIG. 2. Detailed restriction enzyme map of the rfb gene cluster of S. typhimurium LT2 from positions 0 to 25.02 (units in kilobases). The map from positions 0 to 7.57 has been previously published (1). The HindIII site at position 2.93 was previously mapped erroneously at position 1.35 (1). Plasmids pPR303, pPR341, pPR465, and pPR570 (Fig. 1) were digested with a set of 28 restriction enzymes which have 6-base-pair recognition sites. These enzymes were Apal, Aval, BamHI, BgIII, ClaI, DraI, EcoRI, HincII, HindIII, HpaI, KpnI, MluI, NaeI, NcoI, NdeI, NruI, PstI, PvuI, PvuII, SacI, SacII, SalI, ScaI, SmaI, SphI, StuI, XbaI, and XhoI. Digested DNAs were electrophoresed in 0.8%, 1.2%, or 2.0% agarose gels to resolve fragments of 2.73 to 8.0 kb, 0.36 to 2.73 kb, and 0.08 to 0.36 kb, respectively. Single digests were used to position unique sites relative to known vector sites and to identify enzymes with more than three sites in the mapping plasmids. Recognition sites for these enzymes (HincII and NruI in the interval 15.28 to 22.75) were not mapped. Double digests were used to map sites which appeared two or three times in a cloned fragment and to determine the order of ambiguous preliminary position assignments or to measure accurately the sizes of small fragments. There are no recognition sites for the enzymes AvaI, BamHI, NaeI, SacII, and ScaI between positions 9.57 and 25.02. Only the SalI and EcoRI sites between positions 25.02 and 40.09 have been mapped.

used as a probe for Southern hybridization analysis against Bg/II-cleaved chromosomal DNA from all of the his-rfb deletion strains listed in Table 1, with P9003 included as a positive control. The probe showed hybridization to DNA from P9003 and all of the deletion strains other than that carrying $\Delta(his-metG)388$. The deletions known to end within rfb thus all terminate before the end of pPR648 at position 40.09, whereas the deletion in P9029, known to extend from within his through rfb into metG, terminates beyond position 40.09 (Fig. 1). Precise localization of the deletion endpoints was done by Southern hybridization analysis as described in Materials and Methods. A series of probes and enzymes were used to define the endpoints, and in Fig. 3 we present the autoradiographs for some of these experiments which localized seven of the endpoints to short intervals between adjacent restriction sites. The location of each deletion endpoint is marked in Fig. 1, together with the two determined previously (1). The probe DNA used in localizing each of the deletion endpoints is specified in Fig. 3. All the deletion endpoints are shown as lying between two restriction enzyme sites, of which only the right-hand one is conserved. In each case, several sites to the right of the deletion endpoint were also shown to be present and several sites to the left were shown to be absent.

DISCUSSION

Salmonella O antigens constitute one of the best-documented examples of polymorphism involving a surface polysaccharide. The structural analysis of the lipopolysaccharides of different groups (9) and the demonstration that nearly all the genes determining O-antigen biosynthesis were essentially in the one rfb cluster (17) showed that this variation could be analyzed in detail at the genetic level. However, owing largely to the instability of strains with point mutations and the absence of good selection for the rfb region, further elucidation of the genetic basis of this variation has been limited. Recently, we reported the cloning and physical mapping of part of the rfb region of S. typhimurium (1). We have now cloned the complete rfb gene cluster and established a detailed restriction enzyme map, thus laying the foundation for an analysis of the organization of the cluster and determination of the genetic basis for the variation in the O antigen of Salmonella spp. We have also located all of the available deletion endpoints known to lie in the rfb locus, thereby providing an approximate position for the genes located near each endpoint (17). The deletions all lack a group of contiguous genes (17), and the last gene deleted (i.e., distal to the his operon) is indicated in Fig. 1. It

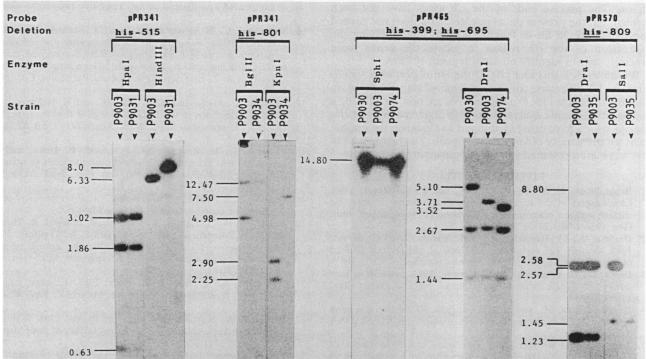


FIG. 3. Southern hybridization analysis of deletion endpoints. Data are shown only for the two restriction enzymes which cleave nearest to, and on either side of, each deletion endpoint. A control track containing wild-type DNA digested with the same enzyme is also shown for each case. The probe DNA used is indicated in the top line, and the sizes of the bands are in kilobases. Note that the 8.8-kb band in the Sall digest of P9035 is very faint owing to the short region of homology with probe DNA. In several cases we can interpret the data only if we assume that fragments with only a short region of homology with probe DNA are not detected. For example, in the his-515 deletion endpoint, the expected 9.84-kb HpaI fragment (between positions 15.05 and 24.75) is not seen, but the detection of the 0.63-kb HpaI fragment (between positions 9.54 and 10.17) from P9031 DNA shows that the HpaI site at 9.54 is conserved. At the same time, the presence of an 8.0-kb HindIII fragment from P9031 instead of the 6.33-kb fragment (positions 8.95 to 15.28) seen in the P9003 HindIII (gostion 8.95) and HpaI (position 9.54) sites. Other fragments having limited overlap with the probe DNA which are not detected include a 7.48-kb KpnI fragment (positions 2.47 to 9.95) in the P9003-KpnI track for the his-801 deletion endpoint; a 0.25-kb KpnI fragment (positions 15.23 to 15.53) in the P9003 and P9034 KpnI digests for the his-801 deletion endpoint; and a 1.23-kb DraI fragment (positions 22.44 to 23.67) in the P9003, P9030, and P9074 DraI digests for the his-399 and his-695 deletion endpoints.

102 BRAHMBHATT ET AL. J. BACTERIOL.

is interesting that the three deletions ending in or just beyond rfbK all end within the same 0.24-kilobase (kb) region, although they are known to include at least two different endpoints at the other end (Table 1). A similar result is seen with the two endpoints terminating in or just beyond rfbD or rfbA, both endpoints lying in the same 0.66-kb region. It remains to be seen whether the endpoints in each cluster are at specific sites within these two regions. The deletion endpoints given here provide only an approximate estimate for the locations of the genes. Each deletion must end in or beyond the gene indicated as the deletion endpoint in Fig. 1: each gene could therefore lie anywhere between that point and the deletion endpoint to the left. Several other unmapped enzyme functions are expected to lie in the rfb locus; for example, only one of the four transferase genes (rhamnosyltransferase, rfbN) has been definitively located (8), and although Yuasa et al. (24) have suggested that the galactosyltransferase gene is at the left-hand end of the rfb cluster closest to his, where several point mutations have been identified, the other two are unmapped. Furthermore, rfbH is the genetic symbol for CDP-abequose synthetase, which by analogy with Yersinia pseudotuberculosis types II, III, and V (14), may well comprise three or more enzymes and hence genes. Two large gaps have been found between the rfb endpoints of deletions his-695 and his-809 and that of his-101, but only a few known genes are located here, indicating that further functions may be located in this region. The precise ends of the rfb locus have not been defined, but the present data indicate that it does not exceed approximately 30 kb in length. We are now studying the expression of the rfb region to locate the genes more precisely and to identify possible regulatory sites.

We undertook this analysis of the *rfb* region of *S. typhimurium* LT2 because of the polymorphism exhibited in *Salmonella* species for O antigens. We are now conducting a comparative genetic analysis of the *rfb* regions of *Salmonella* species which have other O antigens to elucidate the genetic basis for the variety of known O antigens. The first results of this work are presented in the accompanying paper (22).

LITERATURE CITED

- 1. Brahmbhatt, H. N., N. B. Quigley, and P. R. Reeves. 1986. Cloning part of the region encoding biosynthetic enzymes for surface antigen (O-antigen) of Salmonella typhimurium. Mol. Gen. Genet. 203:172–176.
- Davison, J., M. Heusterspreute, M. Merchez, and F. Brunel. 1984. Vectors with restriction-site banks. I. pJRD158, a 3903-bp plasmid containing 28 unique cloning sites. Gene 28:311-318.
- Dente, L., G. Cesareni, and R. Cortese. 1984. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. 11:1645-1655.
- Heusterspreute, M., and J. Davison. 1984. Restriction site bank vectors. II. DNA sequence analysis of plasmid pJRD158. DNA 3:259-268.
- Jann, K., and B. Jann. 1984. Structure and biosynthesis of O-antigen, p. 138-186. In E. T. Rietschel (ed.), Handbook of endotoxin, vol. 1. Elsevier Scientific Publishers, Amsterdam.
- Karn, J., S. Brenner, and G. Cesareni. 1980. Novel bacteriophage cloning vector. Proc. Natl. Acad. Sci. USA 77:5172– 5176.
- 7. Kauffmann, F. 1966. Enterobacteriaceae. Ejnar Munksgaard,

Copenhagen.

- Levinthal, M., and H. Nikaido. 1969. Consequences of deletion mutations joining two operons of opposite polarity. J. Mol. Biol. 42:511-520.
- Luderitz, O., K. Jann, and R. Wheat. 1968. Somatic and capsular antigens of Gram-negative bacteria, p. 105-277. In M. Florkin and E. H. Stotz (ed.), Comprehensive biochemistry. Extracellular and supporting structures, vol. 26A. Elsevier Publishing Co., Amsterdam.
- Makela, P. H., and M. Rhen. 1985. Antigenic variation in enteric bacteria, p. 75-92. In T. K. Korhonen, E. A. Dawes, and P. H. Makela (ed.), Enterobacterial surface antigens; methods for molecular characterization. Elsevier Science Publishers, Amsterdam.
- 11. Makela, P. H., and B. A. D. Stocker. 1984. Genetics of lipopoly-saccharide, p. 75-92. *In* E. T. Rietschel (ed.), Handbook of endotoxin, vol. 1. Elsevier Science Publishers, Amsterdam.
- Mandel, M., and A. Higa. 1970. Calcium dependent bacteriophage DNA infection. J. Mol. Biol. 53:159-162.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual, p. 163 and 468. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Matsuhashi, S., and J. L. Strominger. 1967. Enzmatic synthesis
 of cytidine diphosphate 3,6-dideoxyhexoses. IV. Partial resolution of the enzymes required for reduction of cytidine diphosphate 4-keto-6-deoxy-p-glucose and isolation of a new intermediate. J. Biol. Chem. 242:3494-3500.
- Maurer, R., B. C. Osmond, E. Shekhtman, A. Wong, and D. Botstein. 1984. Functional interchangeability of DNA replication genes in Salmonella typhimurium and Escherichia coli demonstrated by a general complementation procedure. Genetics 108:1-23.
- McKenney, K., H. Shimatake, D. Court, U. Schmeissner, and M. Rosenberg. 1981. A system to study promoter and terminator signals recognized by *Escherichia coli* RNA polymerase, p. 383. In J. Chirikjian and T. Papas (ed.), Gene amplification and analysis, vol. 2. Elsevier/North-Holland Publishing Co., New York.
- Nikaido, H., M. Levinthal, K. Nikaido, and K. Nakane. 1967.
 Extended deletions in the histidine-rough-B region of the Salmonella chromosome. Proc. Natl. Acad. Sci. USA 57:1825–1832.
- Poustka, A., H. Rackwitz, A. Frischauf, B. Hohn, and H. Lehrach. 1984. Selective isolation of cosmid clones of homologous recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 81:4129-4133.
- Quigley, N. B., and P. R. Reeves. 1987. Chloramphenicol resistance vector based on pUC9. Plasmid 17:54-57.
- Ratcliff, S. W., J. Luh, A. T. Ganesan, B. Behrens, R. Thompson, M. A. Montenegro, G. Morelli, and T. A. Trautner. 1979.
 The genome of *Bacillus subtillis* SPP1: the arrangement of restriction endonuclease generated fragments. Mol. Gen. Genet. 168:165-172.
- Sanderson, K. E., and J. R. Roth. 1983. Linkage map of Salmonella typhimurium, edition VI. Microbiol. Rev. 47:410– 453
- Verma, N. K., N. B. Quigley, and P. R. Reeves. 1988. O-antigen variation in Salmonella spp.: rfb gene clusters of three strains.
 J. Bacteriol. 170:103-107.
- 23. Vieira, J., and J. Messing. 1982. The pUC9 plasmids, an M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268.
- Yuasa, R., M. Levintha, and H. Nikaido. 1969. Biosynthesis of cell wall lipopolysaccharide in mutants of *Salmonella*. J. Bacteriol. 100:433

 –444.