Molecular Characterization of Plasmid pBM300 from *Bacillus megaterium* QM B1551

Muthusamy Kunnimalaiyaan* and Patricia S. Vary

Department of Biological Sciences, Northern Illinois University, De Kalb, Illinois 60115

Received 23 September 2004/Accepted 13 December 2004

Strain QM B1551 of *Bacillus megaterium* contains seven compatible plasmids: two small rolling circle plasmids and five theta-replicating plasmids with cross-hybridizing replicons. To expand our understanding of these plasmids, the replicon region (6.7 kb) from pBM300 was cloned, sequenced, and functionally characterized. Sequence analysis showed that the replication protein (RepM300) was highly homologous to two other plasmid Rep proteins of the same strain but to no other known proteins. Furthermore, the location of the replication origin was within the RepM300 coding region, and the origin contained three 12-base direct repeats. Deletion analysis of the replicon confirmed the role of the Rep protein and showed that open reading frame 2 (ORF2) was required for stability. However, the protein encoded by ORF2 is entirely different from the replicon stability proteins encoded by the other two replicons. The entire plasmid was isolated from the plasmid array by integrating a spectinomycin resistance gene and transforming a plasmidless strain, PV361. Complete sequencing showed that pBM300 was 26,300 bp long, had a G+C content of 35.2%, and contained 20 ORFs, two of which encoded proteins that had no similarity to other proteins in the database. The proteins encoded by the plasmid ORFs had similarity to proteins for mobilization and transfer, an integrase, a rifampin resistance protein, a cell wall hydrolase, glutathione synthase, and a biotin carboxylase. The similarities were to several gram-positive genera and a few gram-negative genera and archaea. *oriT* and *ssoT*-like regions were detected near two *mob* genes. These results suggest that pBM300 is a mobilizable hybrid plasmid that confers increased metabolic and germination ability on its host. Its replicon also helps define a new plasmid family.

Plasmids are important in the overall physiology and survival of many bacteria. They carry genes that may be involved in virulence, degradation of toxic compounds, and antibiotic and heavy metal resistance, as well as their own transfer among species or even genera. The concept of a “horizontal gene pool” is emerging as more plasmids are studied and sequenced (43). The study of plasmid replication is also very important in understanding the proteins involved in genomic as well as plasmid DNA replication. The DNA replication machinery could provide one of the richest targets for the development of new antibacterial agents by direct inhibition of DNA replication. Many gram-positive bacteria carry large plasmids, but only a few of the large plasmids have been studied with respect to the mechanism of replication (10, 15, 16, 25, 30, 35, 40, 43). In contrast, there have been many reports on the smaller rolling circle replicating plasmids, especially those from *Bacillus*, *Lactococcus*, *Streptococcus*, and *Staphylococcus* (8, 20, 21, 29). Several species of gram-positive bacteria carry multiple plasmids (10, 15, 23, 49), but most research on the large plasmids has been limited to partial sequencing and characterization of a few genes. Moreover, little attention has been focused on the interaction of multiple plasmids, the genes that such an array might carry, and their role in the cell. Recently, a few large plasmids have been completely sequenced, including the two virulence plasmids of *Bacillus anthracis*, pXO1 and pXO2 (30, 31), and the toxin-encoding plasmid of *Bacillus thuringiensis* (4). It is medically important to begin to investigate the genes

on these plasmids and to determine the extent of their transfer to other species.

The gram-positive bacterium *Bacillus megaterium* is of great industrial interest because of its production of vitamin B₁₂, several useful enzymes, and fungal antibiotics (46). It is an excellent cloning host because of its nonpathogenicity, its secretion of proteins, the absence of alkaline proteases, its ability to stably maintain plasmids, and its ability to produce intact, functional proteins (24, 36, 39, 46, 47). Our laboratory is interested in the role of plasmids in *B. megaterium* strain QM B1551 and their replication, stability, and possible horizontal transfer among the gram-positive bacteria. Many strains of *B. megaterium* carry 4 to 10 plasmids (46). The seven indigenous plasmids of strain QM B1551 comprise approximately 11% of the total cellular DNA (23). These plasmids are designated pBM100, pBM200, pBM300, pBM400, pBM500, pBM600, and pBM700, and their sizes range from 5.4 kb to over 165 kb. Previously, we reported sequencing and characterization of two of the replicons from the large plasmids of QM B1551 (23, 25, 40). These replicons cross-hybridize with all five of the large plasmids, suggesting that the five replicons are homologous. Indeed, the two replicons sequenced are more than 89% identical, which is very unusual for compatible plasmids. Of the five largest plasmids, pBM300 is present at the highest copy number (23). Characterization of this plasmid may lead to a better understanding of the similarity of these compatible plasmids. It may also facilitate construction of suitable expression vectors using the replicon. Hence, sequencing of one of the larger plasmids and annotation of its genes should yield important information concerning both the function of some of these genes and their possible transfer.

Here we describe characterization, separation, and sequenc-

* Corresponding author. Present address: Department of Surgery, University of Wisconsin Medical School, K4/638 CSC, 600 Highland Avenue, Madison, WI 53792. Phone: (608) 265-3749. Fax: (608) 263-8613. E-mail: kunni@surgery.wisc.edu.

TABLE 1. Strains and plasmids used in this study

Strain or plasmid	Description or genotype ^a	Source or reference
<i>B. megaterium</i> strains		
QM B1551	Wild type (7p+)	Laboratory strain
PV361	Plasmidless (7p-) QM B1551	39
PV634	PV361/pBM300::spc	This study
<i>Escherichia coli</i> DH5α	F ⁻ ϕ 80dlacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17 (r _K ⁻ m _K ⁺) phoA supE44 λ ⁻ thi-1 gyrA96 relA1	Bethesda Research Laboratory
Plasmids		
pDS48-20	Sau3A 6.7-kb replicon fragment in pBEST501	40
pBM300::spc	27.3-kb pBM300::Spc ^r	This study
pBEST501	Amp ^r in pGEM4, Nm ^r from pUB110 <i>Bacillus</i> integrative vector	19
pJM103	Amp ^r in pUC19, Cm ^r from pC194	32
pDG792	pMTL23 containing Km ^r cassette	17
pDG1726	pMTL23 containing Spc ^r cassette	17
pGEM-T-easy	Amp ^r F1 ori lacZ, for PCR cloning	Promega
pGEM7zf	Amp ^r F1 ori lacZ::PrT7, PrSP6, 3.0 kb	Promega
pYZ11	pBM100, Tc ^r , pUC19	25
pKM60	Δ402- <i>repM400</i> <i>orfB</i> in pJM103 (Cm ^r Km ^r)	25
pKM300	6.5-kb EcoRI replicon fragment from pBM300	This study
pKM307	HaeIII fragment of pKM300 at SmaI site of pJM103, <i>repM300</i> plus most of <i>orf2</i>	This study
pKM310	Minimal replicon of <i>repM300</i>	This study
pKM312	3.2-kb fragment in pGEM7zf	This study
pKM313	4-kb pE194 (Ts origin) EcoRI from pLTV1 in EcoRI site of pKM312	This study
pKM314	1-kb <i>spc</i> HindIII insertion into pKM313	This study
pKM315	Integrative vector for pBM300, Spc ^r Km ^r Amp ^r	This study
pKM317	<i>repM300</i> plus <i>orf2</i> in pGEM-T-easy, Spc ^r	This study

^a 7p+ and 7p-, presence and absence, respectively, of the seven indigenous plasmids.

ing of a third theta replicon, the 26.3-kb plasmid pBM300. The study of a large cryptic plasmid in a multiplasmid array with varying copy numbers presented several technical difficulties that had to be overcome. Nevertheless, the sequencing and initial annotation of pBM300 revealed possible genes for enhanced survival, metabolism, DNA plasticity, and transfer.

MATERIALS AND METHODS

Bacterial strains, growth, and transformation. The plasmids and strains used in this study are listed in Table 1. PV361, a plasmidless derivative of *B. megaterium* QM B1551 (41), was used as the host strain for replication studies. The growth conditions, the SNB and Luria-Bertani media, and transformation of *B. megaterium* and *Escherichia coli* have been described previously (40). The following antibiotic concentrations were used: for *B. megaterium*, 10 μg/ml chloramphenicol, 2 μg/ml neomycin, 100 μg/ml spectinomycin, and 5 μg/ml kanamycin; and for *E. coli*, 100 μg/ml ampicillin and 25 μg/ml kanamycin.

Plasmid DNA isolation and methods. Standard recombinant DNA methods described by Sambrook et al. (37) were used. Plasmid DNA and total cellular DNA were isolated as previously described (40). Biotinylated probes were made using a NEBlot Phototape kit (New England Biolabs), and hybridizations were carried out at 65°C according to the manufacturer's protocol. Standard PCRs were carried out under the following conditions: 95°C for 60 s, 45 to 50°C for 90 s, and 72°C for 90 s for 30 cycles using *Taq* DNA polymerase.

Isolation of the pBM300 replicon (clone III) and construction of subclones. Total plasmid DNA from QM B1551 was isolated, restricted with Sau3A, and ligated into the BamHI site of pBEST501 containing no replicon for *Bacillus*. The ligated material was transformed into PV361 with selection on SNB medium containing neomycin to isolate clones that could replicate. Plasmid DNA from the transformants was isolated, retransformed into DH5α, and selected on LB medium containing ampicillin. Restriction analysis was carried out on the transformants (data not shown), and one of the clones (6.7 kb) that could replicate in PV361 was designated pDS48-20 (David Stevenson, unpublished data). The selective concentration of neomycin in PV361 using pBEST-501 was not very effective, so a 6.4-kb EcoRI fragment from pDS48-20, designated clone III, was cloned at the EcoRI site of pJM103 to construct pKM300.

Subclones of the replicon region in clone III were constructed either by restriction digestion or by PCR. A 1.8-kb PCR fragment containing only the *rep* gene from clone III was cloned using primers 5'-TTTTTCCCTTCCTTTTAACTT-3' and 5'-CAAGTCCTACATCAAGTCCTAACTC-3' into a Kan^r derivative of pUC18 to form pKM310. Clone III was restricted with HaeIII, and the

3.5-kb fragment was cloned into pJM103 cut with SmaI to construct pKM307 with *repM300* and most of open reading frame 2 (ORF2). A 2.6-kb PCR product containing the *rep* gene plus all of ORF2 was cloned into pGEM-T-easy. Then a *spc* cassette from pDG1726 that is expressed in both *E. coli* and *Bacillus* was cut with EcoRV-StuI and was inserted into the ampicillin gene at the ScaI site to construct pKM317.

Construction of integrative plasmid pKM315. A 3.2-kb HpaI fragment specific for pBM300 was cloned into the SmaI site of pJM103. Then the fragment was removed by EcoRI-BamHI digestion and was ligated at the EcoRI-BamHI site of pGEM7zf to form pKM312. A schematic diagram of construction of the integrative plasmid, pKM315, is shown in Fig. 1. The pE194 temperature-sensitive replicon in pLTV1 (5) was recovered as an EcoRI fragment, ligated into the EcoRI site of pKM312, and amplified in DH5α. The resultant plasmid, pKM313, was linearized with HindIII, ligated to a 1-kb HindIII fragment containing a spectinomycin resistance gene cassette from pDG1726 (17), transformed into DH5α, and selected with spectinomycin.

Thus, the spectinomycin gene was inserted into the pBM300-specific fragment (Fig. 1). The kanamycin gene cassette (StuI-SmaI from pDG792), selective in both *E. coli* and *B. megaterium*, was then inserted into the ampicillin coding region at the ScaI site of pKM314, to obtain pKM315. The final integrative plasmid specific for pBM300, pKM315, was then amplified in DH5α with kanamycin selection.

Sequence analysis. To analyze the replicon of pBM300, the 6.7-kb replicon fragment from pDS48-20 was first sequenced by using universal primers and primer walking. Custom primers were synthesized both by MWG Biotech (North Carolina) and by the Northern Illinois University Core DNA Synthesis and Sequencing Facility. The entire 26.3-kb plasmid was then sequenced by constructing PstI and EcoRI libraries in pJM103 using pBM300::spc DNA. Plasmid DNA from the libraries was sequenced in both directions by primer walking with a series of custom primers and universal primers. The complete sequence was compiled by overlapping the fragments between the libraries.

For computer homology searches we utilized the BLAST series of programs (1) provided by the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) and GeneMark (<http://opal.biology.gatech.edu/GeneMark/hmmchoice.html>). DNA and amino acid sequences were analyzed using PCGENE (Intelligenetics, Inc.), as well as the following online tools: MultiAlin (<http://prodes.toulouse.inra.fr/multalin/multalin.html>) for multiple alignments; EINHATED (<http://bioweb.pasteur.fr/seqanal/interfaces/einverted.html>) for inverted repeats; REPuter (<http://bibiserv.techfak.uni-bielefeld.de/reputer/>) for direct and inverted repeats; and Promoter Prediction by Neural Network (<http://www.fruitfly.org/seq-tools/promoter.html>). Access to several internet sites was obtained through A. Kropinski's site at Queen's University, Ontario, Canada (<http://molbiol-tools.ca>).

FIG. 1. Construction of pKM315, an integrative vector for pBM300. See the text for a description.

Stability of the replicon. Subclones of the 6.7-kb fragment containing the replicon with various flanking DNA sequences were tested every 6 h for stability in PV361 for over 100 generations, as previously described (25).

Nucleotide sequence accession number. The sequence data have been deposited in the GenBank databases under accession number AY725176.

RESULTS AND DISCUSSION

Isolation and localization of the replicon from plasmid pBM300. A 6.7-kb *Sau3A* fragment (clone III) that could replicate in PV361 had previously been isolated, amplified, and designated pDS48-20 (Stevenson, unpublished data), as described in Materials and Methods. A 6.4-kb *EcoRI* fragment from pDS48-20 was then cloned at the *EcoRI* site of pJM103 to construct pKM300. Both plasmid pDS48-20 and plasmid pKM300 were used to sequence clone III in both directions.

Due to cross hybridization of the replicon region with four other large theta plasmids (40), PCR was carried out at the distal (specific) end of the 6.7-kb fragment to identify the source plasmid for the clone III fragment. The resulting PCR fragment was labeled and hybridized to QM B1551, QM B1551::spc, PV203, and PV634 DNA, as shown in Fig. 2A. The probe hybridized only with the 26.3-kb pBM300 plasmid (Fig. 2D), and there was no homology to the other plasmids. Therefore, we concluded that the clone III fragment contained a replicon from pBM300.

Sequence analysis and characterization of the replicon. The clone III replicon region was sequenced in both directions, and a schematic diagram is shown in Fig. 3. The six ORFs included the replication protein RepM300 that is highly similar (>80%) at both the nucleotide and amino acid sequence levels to the

FIG. 2. Identification of pBM300 as the source plasmid for clone III by Southern hybridization and separation of pBM300 into PV361. (A) Agarose gel electrophoresis of plasmids. QM B1551, the wild-type strain, contains all seven plasmids; QM B1551/26::spc contains plasmids isolated after *spc* gene insertion; the PV203 strain does not contain plasmid pBM300 (used as a negative control); and PV634 contains only pBM300::spc. The DNA was transferred to a nitrocellulose membrane. Then the membrane was hybridized with different probes, as indicated above panels B to E, sequentially after the old probe was stripped. (B to E). Hybridization with plasmids and derivatives from *B. megaterium* QM B1551 with the following specific probes: pBM500 (clone I) (40) (B), pBM400 (clone II) (38) (C), a PCR fragment specific for the clone III probe (distal to the *rep* gene in pBM300) (D), and the *spc* gene probe (E). Note that the blots in panels B to E were derived from the gel shown in panel A.

other described theta replication proteins of QM B1551, RepM400 (clone I) and RepM500 (clone II) (25, 40). The region immediately downstream of *repM300* contains an ORF (ORF2) that is not present in the other QM B1551 replicons and has no sequence homology to ORFs encoding any other proteins in the database. A typical theta iteron-type origin (*ori300*) (11) was observed in the gene coding for the replication protein (*repM300*), as has been found for *repM500*, *repM400*, and *repM700* (25, 40; Kunnimalaiyaan and Vary, unpublished data). The location of an origin in an open read-

FIG. 3. Diagram of the 6.7-kb clone III fragment. The six ORFs, including *repM300* and ORF2 and a possible glutathione synthase gene, are included. Also shown is an alignment of three of the iterons from three of the theta plasmids (clones I to III) and the replication analysis of deletion clones. The numbers above the ORFs indicate the sizes of the proteins encoded by the ORFs (in amino acids). CI, clone I; CII, clone II; CIII, clone III.

ing frame is similar to the locations of the phage λ origin of replication, the pT181 initiator (22), and *Staphylococcus* plasmids pSK1, pSK41, and pI9789::Tn552 (12). However, unlike the proteins encoded by these ORFs, RepM300 has no homology to any of the replication proteins reported previously or in databases other than those encoded by the large theta plasmids of wild-type strain QM B1551.

The minimal replicon was identified by construction of clones of various lengths flanking *repM300* (Fig. 3) by either deletion or PCR as described in Materials and Methods. Replication of these clones showed that the fragment that could still replicate in PV361 was about 1.8 kb long (pKM310) and contained only a putative promoter and *repM300*. The functioning of the minimal replicon, with no other direct repeats, strongly suggested that the origin was indeed in the coding region of the initiation protein.

The origin in pBM300, *ori300*, contains only three identical 12-base direct repeats, and both the size and number contrast with the size and number for the iterons reported in plasmids in gram-negative bacteria and in other gram-positive bacteria, such as *Lactococcus* (6, 11). Normally, iterons are the site of binding of Rep initiation proteins and are located either upstream or downstream of the Rep protein gene. They usually have larger (17- to 22-bp) direct repeats, and there are usually several of them (8, 18, 28). Previously, we reported that clone I (pBM400) and clone II (pBM500) direct repeats are able to confer incompatibility for their own replicons but not for each other, which defined them as iterons (25). These results are unusual since the iterons are very similar, as shown in the alignment in Fig. 3. While clone I has four repeats, the first three are identical to the three repeats in clone III and differ by only one base from the repeats in clone II. There are no other direct repeats in the 6.7-kb clone III fragment. Therefore, it is probable that the very similar direct repeats of RepM300 act as iterons. The mechanism by which the homologous Rep proteins are able to distinguish between similar iterons is not surprising since the pT181 family of plasmids required only a stretch of six amino acids to determine specificity (21, 27).

ORF2 is required for the stability of the replicon. The stabilities of the various subclones obtained from the clone III fragment were also tested (Fig. 4). Plasmid pKM310 containing *repM300* without ORF2 was 100% stable for up to 50 generations, and then the stability decreased rapidly. However, pKM307, which contained *repM300* plus most of *orf2*, was 95% stable for about 100 generations. As expected, the 6.4-kb fragment of clone III in pKM300 was 100% stable for 100 generations. As a positive control we used pKM60, a clone II stable replicon that was stable for 100 generations. Plasmid pKM317, which contained the complete *orf2*, was more than 95% stable (data not shown). The results of the stability experiments suggest that ORF2 is required for the stability of pBM300. Surprisingly, the pKM317 plasmid is not as stable as pKM300. Further research is needed to determine the sequence or protein that confers the additional stability.

Insertion of the *spc* gene into pBM300 and its separation in PV361. Separation of a plasmid with no known selectable marker from an array of seven plasmids with various copy numbers presents a unique challenge. Previous attempts to label the plasmid with a transposon failed. Over 100 randomly cured strains retained one or two small, high-copy-number plas-

FIG. 4. Stability of the pBM300 replicon in PV361. PV361 containing pKM300 (whole replicon region of clone III), pKM310 (*repM300* alone, minimal replicon), pKM307 (*repM300* and most of ORF2), or pKM60 (*repM400* plus *orfB*) was grown in the absence of antibiotic selection. The percentages of cells maintaining the plasmid after different numbers of generations were determined as described in Materials and Methods.

mids and/or one or two of the large plasmids (Vary and Weiland, unpublished). Therefore, the integrative vector pKM315 containing the pE194 temperature-sensitive replicon was constructed as described in Materials and Methods, and wild-type strain QM B1551 was transformed, with selection for kanamycin resistance. A QM B1551/pKM315 transformant was purified and grown at 42°C for three subcultures at 6-h intervals without selection and then plated on SNB medium and replica plated on both SNB medium containing kanamycin and SNB medium containing spectinomycin. Growth on spectinomycin and an inability to grow on kanamycin indicated loss of pKM315 and the presence of a double crossover between the pBM300-specific fragment containing the spectinomycin gene and the indigenous pBM300 plasmid. A *Spc^r Kan^s* colony was saved as QM B1551/pBM300::*spc*, and the plasmid was isolated and separated on an agarose gel; then the 26-kb DNA region was recovered from the gel. The recovered DNA was transformed into PV361 to obtain PV634 carrying only pBM300::*spc*. The absence of all other plasmids in strain PV634 was verified first by PCR, using plasmid-specific primers for clone I (pBM500) and clone II (pBM400) and for the two smallest plasmids, which did not produce any product (data not shown), and then by Southern hybridization, using specific probes for pBM500 (Fig. 2B) (40) and pBM400 (Fig. 2C) (25), a PCR fragment specific for the clone III probe (Fig. 2D), and *spc* (Fig. 2E). As Fig. 2 shows, only the *spc* and pBM300-specific probes hybridized with PV634. Thus, PV634 contained only plasmid pBM300::*spc*. Interestingly, the amount of isolated plasmid pBM300 DNA from strain PV634 was increased compared to the wild-type QM B1551 plasmid DNA. The increase in the band intensity may have been due to the use of the antibiotic spectinomycin when the culture was grown. Fur-

FIG. 5. Schematic diagram of plasmid pBM300. ORFs are indicated by both open and shaded arrows. PstI, EcoRI, and BamHI enzyme sites are indicated for reference. Dark grey arrows, ORF encoding RepM; medium grey arrows, ORFs encoding predicted proteins with similarity to proteins having known functions; light grey arrows, ORFs encoding predicted proteins with similarity to proteins having unknown functions; open arrows, ORFs encoding proteins with no similarity.

thermore, the band migrating above the 26.3-kb plasmid in PV364 may have been the nicked or relaxed DNA of plasmid pBM300.

Sequencing of the entire pBM300 plasmid. Sequencing of the entire plasmid was done as described in Materials and Methods. Computer analysis revealed that plasmid pBM300 is 26,300 bp long (without *spc*) and has a G+C content of 35.2%, which is lower than that of the *B. megaterium* genome (37 to 39%) (33). A diagram of pBM300 is shown in Fig. 5. The results of a computer analysis of predicted ORFs and the similarity to other known gene products are shown in Table 2. There are 20 ORFs, and two of these ORFs have no homology to any proteins found in the GenBank database. Of the 20 ORFs, 12 (including *repM300*) are on one strand, and 8 are on the other strand. To include possible ORFs, two or more of the following criteria had to be met: Fickett scores of >70 (PC GENE analysis), putative Shine-Dalgarno sites, the presence of a putative promoter, and homology with proteins in the databases.

Besides the RepM300 coding region, several other possible replication or plasmid transfer genes and regions are present on the plasmid. Upstream of the replicon there is a transfer origin-like sequence (*oriT*) with a possible nick site and similarity to *oriT* regions in plasmids pIP501 (pVA1702), RSF1010, and pTF1 (48) (Fig. 6). Wang and Macrina (48) demonstrated that the *oriT*-like sites of RSF1010 and pTF1 are required for conjugal mobilization in *E. coli*. It is interesting that the *B. megaterium oriT* sequence is similar to those from some gram-negative bacterial plasmids (RS1010 and pTF1). Two *oriTs* have been reported on plasmid pAD1 of *Enterobacter faecalis* and have been functionally characterized (13, 14), whereas a single *oriT* was found on plasmid pBM300 by sequence analysis. A small ORF, ORF20, with similarity to MobC of *E. faecalis* pAD1 is present between the *oriT*-like structure and *repM300*. It is similar to ORF3 of plasmid pMRC01 of *Lactococcus lactis* (10) and to an ORF on plasmid RSF1010 (9). Upstream from *oriT* is a large ORF, ORF19, which encodes a predicted protein that has similarity to the

TABLE 2. Plasmid pBM300 predicted ORFs and homologies

ORF	Position		Protein size (amino acids)	Fickett value	Possible Shine-Dalgarno sequence	Gene product or homologue	Organism	No. positive/ no. examined (%)	E value
	Stand	bp							
ORF1	+	108–1394	428	98	AAGGAGAAAGATGAACAGATG	RepM400 RepM500 Stability	<i>B. megaterium</i> <i>B. megaterium</i> <i>B. megaterium</i> (pBM300)	387/428 (90) 385/419 (91)	0.0 0.0
ORF2	+	1600–2139	179	92	AATCCAGGGGGAGCACATGTG	Glutathione synthase S6 modification		159/318 (50)	6e ⁻³⁷
ORF3	+	3562–4593	342	77	AAGGAAAGGATGGTTTATG	Hypothetical protein		65/154 (42)	0.06
ORF4	+	4754–5290	178	77	AGGATAGGATGATAGATG	Biotin carboxylase		113/225 (50)	1e ⁻¹⁷
ORF5	+	5287–5988	233	92	AACGTGAGTACATTCATCTATG	Hypothetical protein	<i>Clostridium acetobutylicum</i>	148/221 (66)	5e ⁻⁴⁸
ORF6	+	6768–7524	252	77	AAGGAGATGAGAAAGT	Hypothetical protein	<i>Magnetospirillum magnetotacticum</i>	56/127 (44)	0.12
ORF7	+	7882–8316	144	100	AGAAAGGAGGTGATAAAATG	Collagen-like protein	<i>Bacillus cereus</i>	123/203 (60)	2e ⁻⁵⁶
ORF8	+	9059–10138	359	100	AAAGAAAGTGAATTTATG	Mobilization protein	<i>Bacillus subtilis</i> pTA1015	303/488 (61)	7e ⁻⁹³
ORF9	+	11520–12938	472	100	GAAATGGGGATGAGGGATG	Rifampin-ADP-ribosyl transferase	<i>M. smegmatis</i>	92/103 (89)	8e ⁻³⁴
ORF10	–	13623–14039	138	40	AAGATTAAGGGGTGTGCTATATG		<i>P. aeruginosa</i>	90/133 (67)	2e ⁻³¹
ORF11	+	14470–14736	88	98	TTGAAAGGTGGTGGCATTATG	Hypothetical protein	None		
ORF12	–	14830–15399	189	92	AACTATTAGTGGGGGCTAAATG	Hypothetical protein	<i>Bacillus sphaericus</i> (pLPIG)	99/191 (51)	4e ⁻¹⁸
ORF13	–	16020–16730	236	92	ATAAGATTGGAGCCCAATG	Conserved hypothetical protein	<i>B. anthracis</i> (pXO1-68)	105/214 (47)	1e ⁻¹⁰
ORF14	+	16920–17792	266	77	AGTATGGGGAAATTTAAATAATG	Integrase	<i>Listeria</i> phageU15	149/261 (57)	2e ⁻³⁵
ORF15	+	19896–20597	233	92	AACGACAGGATAGATCAATTATG	Hypothetical protein	<i>B. anthracis</i>	161/212 (75)	1e ⁻⁶⁴
ORF16	+	20476–20724	82	98	CGGACAGATTCAGAGAAAGATG	Hypothetical protein	<i>Plasmodium falciparum</i>	41/79 (51)	0.054
ORF17	+	21547–21984	145	92	GGTAGAGGTGATAAAAATG	Cell wall hydrolase	<i>B. anthracis</i>	98/141 (69)	7e ⁻³⁹
ORF18	–	22553–22846	97	40	TTGAGGTGAAAAATATG	Hypothetical protein	<i>O. iheynsis</i>	41/73 (56)	9e ⁻¹³
ORF19	–	23093–25084	663	98	AAGAAAGGAGGAAACGAATCATG	MobA/TraA (nicking)	<i>L. lactis</i> (pMRCO1)	271/563 (48)	1e ⁻⁵⁷
ORF20	+	25462–25773	103	98	AGGAGATGAGGATGFGGGTGTG	Hypothetical protein	<i>E. faecalis</i> <i>L. lactis</i> (pMRCO1)	45/73 (61) 37/57 (64)	6e ⁻⁰⁸ 1e ⁻⁰⁴

FIG. 6. Alignment of *oriT*-like sequences of pBM300 and other plasmids. The dashed arrows indicate palindromic sequences; the asterisks indicate homologous sequences in these plasmids.

TraA or Mob proteins encoded by various plasmids of gram-positive bacteria, including pIP501 and pMRC01 (10, 48). In addition, there is another possible Mob coding region (ORF9) whose product exhibits amino acid similarity to the products of ORFs on rolling circle plasmids pTA1015, pTA1040, and pTA1060 (29) and p1414 of *Bacillus subtilis* (44). Adjacent to this gene is a possible single strand origin (*ss*o) on the opposite strand whose sequence and position are similar to those of the same *B. subtilis* plasmids (Fig. 7). The *ss*o gene also has homology to one of the rolling circle plasmids of the QM B1551 plasmid array, pBM200 (Y. Zhou et al., unpublished data). The presence of two possible *mob* genes and a *traA* gene strongly suggests that this plasmid is mobilizable. Interestingly, there is another gene that may influence the plasticity of the pBM300 DNA. The ORF14 product has significant similarity to a putative viral integrase from *Listeria monocytogenes* bacteriophage U153 (26). However, the ORF14 product contains only 266 amino acids, whereas other integrases contain from 450 to 500 amino acids.

Among the other coding regions on pBM300, ORF10 has similarity to the rifampin-ADP-ribosyl transferase proteins in *Mycobacterium smegmatis* and *Pseudomonas aeruginosa* (34, 45). Rifampin resistance is rarely found on a plasmid in clinical isolates. The resistance in these isolates is normally conferred by missense mutations in the RNA polymerase gene *rpoB* (45). However, a plasmid-mediated rifampin resistance gene was found recently in *Klebsiella pneumoniae* (2). Preliminary tests on the function of ORF10 did not show observable differences in rifampin resistance between the parental strain QM B1551 and the plasmidless strain PV361. Its presence in a nonpathogen is also puzzling.

Several coding regions that may be involved in cell differentiation and metabolism were also present. The ORF17 product had significant similarity to the cell wall hydrolase CwlJ encoded by the genomes of *B. subtilis* and *Bacillus halodurans* (42). Such hydrolases are important in germination. In fact, a CwlJ gene has also been found on pBM400 in QM B1551 (38). The plasmid-borne hydrolase genes may help explain the efficient, synchronized germination of this species observed by many laboratories. A possible glutathione synthase S6 ribosomal modifying enzyme (ORF3, GshB) was present, and the highest level of similarity was the level of similarity to an enzyme encoded by the recently sequenced genome of the deep-sea extremely halotolerant and alkaliphilic species *Oceanobacillus iheynsis*, a microorganism related to *Bacillus* (42). In addition, a putative biotin carboxylase (encoded by ORF5)

FIG. 7. Alignment of putative *sso* sequences from various plasmids.

with an acetyl-coenzyme A carboxylase catalysis domain, similar to an enzyme from *Streptomyces coelicolor* A3(2) (3, 7), was present, and the gene was translationally coupled to an upstream gene, ORF4, with an unknown function in either organism. Interestingly, there was a gene in pBM300 that seems to be conserved in most gram-positive bacterial plasmids. ORF2C, encoding a hypothetical DNA binding protein, has also been found in *B. subtilis* pTA1040 and pTA1060, in *B. anthracis* pXO2-68 (29, 30, 44), and in the other sequenced plasmids of QM B1551 (25, 38, 40; Vary and Kunnimalaiyaan, unpublished data), and the protein is similar to Ptr encoded by p1414 (44). Eleven other possible coding regions had similarity to regions encoding hypothetical proteins with no known functions in various genera and plasmids and are listed in Table 2. Two of these genes (ORF15 and ORF16) have considerable overlap (122 bp).

There is also significant secondary structure on pBM300, including structural features such as large repeats and hairpin loops. A region with two long direct repeats of 171 bases and an almost half-inverted repeat on both sides is present in ORF8 (Fig. 5). In addition, there are two direct repeats of 87 nucleotides at positions 18476 to 18562 and 19552 to 19638 (77% identity) flanking a 990-bp region with no ORF. There is also a 47-bp inverted repeat sequence at positions 8775 to 8821 and 22356 to 22402. There are 13 hairpin structures present which are more than 30 nucleotides long. These hairpin structures are at positions 3318 to 3377, 10149 to 10191, 11094 to 11146, 11211 to 11247, 14181 to 14220, 15863 to 15909, 16902 to 16937, 18626 to 18687, 18724 to 18771, 20820 to 20861, 21246 to 21289, 22277 to 22316, and 22990 to 23023. Many large direct and inverted repeats were also observed on pBM400 (38). Interestingly, there is a 125-nucleotide inverted repeat adjacent to the putative integrase gene and *ssoT* (Fig.

5). The function of these large repeats and the inverted repeat in plasmids is not known.

In summary, sequencing of one of the large, high-copy-number plasmids revealed a theta replication protein gene highly homologous to genes in the other compatible plasmids previously described for this strain, including three very similar iterons. In addition, plasmid pBM300 has a gene that encodes a protein that has a plasmid stability function that is downstream from the *repM300* gene yet is entirely different than the other sequenced theta replicons of the plasmids from QM B1551. The presence of possible *tra* and *mob* genes and *oriT* and *sso* regions for rolling circle plasmid replication suggests that pBM300 is mobilizable under some conditions. Genes for putative enzymes, such as a cell wall hydrolase, biotin carboxylase, and glutamine synthase, probably enhance the survival of *B. megaterium* under some conditions. Furthermore, the lack of any similarity of the replication protein, the stability gene, and the iteron sequences to those in the databases strongly suggests that this replicon and the compatible plasmid theta replicons in strain QM B1551 are unique and may form a new class of compatible replicons. In fact, the very high homology of the Rep proteins and iterons, coupled with an obvious ability to exist within the same cell, is both intriguing and very unusual. Therefore, our laboratory has long had an interest in the biology of the indigenous plasmids of *B. megaterium* strain QM B1551.

ACKNOWLEDGMENTS

We thank Scott Grayburn for primer synthesis and automated sequencing. We also thank David M. Stevenson for constructing pDS48-20, Michael D. Scholle for many useful discussions, Mustafa Hussein for verification of the ORF analysis, and Barbara Ball for graphics.

This research was supported in part by grant 1R15GM49440-01 from the National Institute of Health and by a gift from Abbott Laboratories.

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Meyers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Arlet, G., D. Nadjar, J. L. Herrmann, J. L. Donay, P. H. Lagrange, and A. Philippon. 2001. Plasmid-mediated rifampin resistance encoded by an *arr-2*-like gene cassette in *Klebsiella pneumoniae* producing an ACC-1 class C beta-lactamase. *Antimicrob. Agents Chemother.* **45**:2971–2972.
- Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabinowitz, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**:141–147.
- Berry, C., S. O'Neil, E. Ben Dov, A. F. Jones, L. Murphy, M. A. Quail, M. T. Holden, D. Harris, A. Zaritsky, and J. Parkhill. 2002. Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. *israelensis*. *Appl. Environ. Microbiol.* **68**:5082–5095.
- Camilli, A., D. Portnoy, and P. Youngman. 1990. Insertional mutagenesis of *Listeria monocytogenes* with a novel Tn917 derivative that allows direct cloning of DNA flanking transposon insertions. *J. Bacteriol.* **172**:3738–3744.
- Chatteraj, D. K. 2000. Control of plasmid DNA replication by iterons: no longer paradoxical. *Mol. Microbiol.* **37**:467–476.
- Cronan, J. E., Jr., and G. L. Waldrop. 2002. Multi-subunit acetyl-CoA carboxylases. *Prog. Lipid Res.* **41**:407–435.
- del Solar, G., R. Giraldo, J. Ruiz-Echevarria, M. Espinosa, and R. Diaz-Orejas. 1998. Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**:434–464.
- Derbyshire, K. M., and N. S. Willems. 1987. Mobilization of the non-conjugative plasmid RSF1010: a genetic analysis of its origin of transfer. *Mol. Gen. Genet.* **206**:154–160.
- Dougherty, B. A., C. Hill, J. F. Weidman, D. R. Richardson, J. C. Venter, and R. P. Ross. 1998. Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasmid pMRC01 from *Lactococcus lactis* DPC3147. *Mol. Microbiol.* **29**:1029–1038.
- Filutowicz, M., S. Dellis, I. Levchenko, M. Urh, F. Wu, and D. York. 1994. Regulation of replication of an iteron-containing DNA molecule. *Prog. Nucleic Acid Res.* **48**:239–273.
- Firth, N., S. Apisiridej, T. Berg, B. A. O'Rourke, S. Curnock, K. G. Dyke, and R. A. Skurray. 2000. Replication of staphylococcal multiresistance plasmids. *J. Bacteriol.* **182**:2170–2178.
- Francia, M. V., and D. B. Clewell. 2002. Transfer origins in the conjugative *Enterococcus faecalis* plasmids pAD1 and pAM373: identification of the pAD1 nic site, a specific relaxase and a possible TraG-like protein. *Mol. Microbiol.* **45**:375–395.
- Francia, M. V., W. Haas, R. Wirth, E. Samberger, A. Muscholl-Silberhorn, M. S. Gilmore, Y. Ike, K. E. Weaver, F. Y. An, and D. B. Clewell. 2001. Completion of the nucleotide sequence of the *Enterococcus faecalis* conjugative virulence plasmid pAD1 and identification of a second transfer origin. *Plasmid* **46**:117–127.
- Gravesen, A., J. Josephsen, A. von Wright, and F. K. Vogensen. 1995. Characterization of the replicon from the lactococcal theta-replicating plasmid pJW563. *Plasmid* **34**:105–118.
- Gravesen, A., A. von Wright, J. Josephsen, and F. K. Vogensen. 1997. Replication regions of two pairs of incompatible lactococcal theta-replicating plasmids. *Plasmid* **38**:115–127.
- Guerout-Fleury, A. M., K. Shazand, N. Frandsen, and P. Stragier. 1995. Antibiotic-resistance cassettes for *Bacillus subtilis*. *Gene* **167**:335–336.
- Helinski, D. R., A. E. Toukdarian, and R. P. Novick. 1996. Replication control and other stable maintenance mechanisms of plasmids, p. 2295–2324. In F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.
- Itaya, M., K. Kondo, and T. Tanaka. 1989. A neomycin resistance gene cassette selectable in a single copy state in the *Bacillus subtilis* chromosome. *Nucleic Acids Res.* **17**:4410.
- Janniere, L., A. Gruss, and S. Ehrlich. 1993. Plasmids, p. 625–645. In A. Sonenshein, J. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, D.C.
- Khan, S. A. 1997. Rolling circle replication of bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **61**:442–455.
- Khan, S. A., G. K. Adler, and R. P. Novick. 1982. Functional origin of replication of pT181 plasmid DNA is contained within a 168-base-pair segment. *Proc. Natl. Acad. Sci. USA* **79**:4580–4584.
- Kieselburg, M. K., M. Weickert, and P. S. Vary. 1984. Analysis of resident and transformant plasmids in *Bacillus megaterium*. *BioTechnology* **2**:254–259.
- Kim, H., D. H. Ahn, K. H. Jung, and M. Y. Pack. 1997. Adsorption of *Bacillus subtilis* endo-beta-1,4-glucanase to cellulosic materials. *Biochem. Mol. Biol. Int.* **41**:665–677.
- Kunnimalaiyaan, M., D. M. Stevenson, Y. Zhou, and P. S. Vary. 2001. Analysis of the replicon region and identification of an rRNA operon on pBM400 of *Bacillus megaterium* QM B1551. *Mol. Microbiol.* **39**:1010–1021.
- Lauer, P., M. Y. Chow, M. J. Loessner, D. A. Portnoy, and R. Calendar. 2002. Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. *J. Bacteriol.* **184**:4177–4186.
- Li, Y. F., T. Hayashi, and Y. Terawaki. 1998. Functional domains of Rts1 and P1 RepA proteins for initiation of replication. *Plasmid* **40**:140–149.
- Liu, L., and E. J. Hansen. 1999. Structural analysis of plasmid pLQ510 from *Moraxella catarrhalis* E22. *Plasmid* **42**:150–153.
- Meijer, W. J., G. B. Wisman, P. Terpstra, P. B. Thorsted, C. M. Thomas, S. Holsappel, G. Venema, and S. Bron. 1998. Rolling-circle plasmids from *Bacillus subtilis*: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from gram-positive bacteria. *FEMS Microbiol. Rev.* **21**:337–368.
- Okinaka, R. T., K. Clound, O. Hampton, A. R. Hoffmaster, K. K. Hill, P. Keim, T. M. Koehler, G. Lamke, S. Kumano, J. Mahillon, D. Manter, Y. Martinez, D. Ricke, R. Svensson, and P. J. Jackson. 1999. Sequence and organization of pX01, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *J. Bacteriol.* **181**:6509–6515.
- Pannucci, J., R. T. Okinaka, E. Williams, R. Sabin, L. O. Ticknor, and C. R. Kuske. 2002. DNA sequence conservation between the *Bacillus anthracis* pXO2 plasmid and genomic sequence from closely related bacteria. *BMC Genomics* **3**:34.
- Perego, M., G. B. Spiegelman, and J. A. Hoch. 1988. Structure of the gene for the transition state regulator, *abrB*: regulator synthesis is controlled by the *spo0A* sporulation gene in *Bacillus subtilis*. *Mol. Microbiol.* **2**:689–699.
- Priest, F. G. 1981. DNA homology in the genus *Bacillus*, p. 33–57. In R. C. W. Berkeley and M. Godfellow (ed.), *The aerobic endospore-forming bacteria*. Academic Press, New York, N.Y.
- Quan, S., H. Venter, and E. R. Dabbs. 1997. Ribosylative inactivation of rifampin by *Mycobacterium smegmatis* is a principal contributor to its low susceptibility to this antibiotic. *Antimicrob. Agents Chemother.* **41**:2456–2460.
- Read, T. D., S. L. Salzberg, M. Pop, M. Shumway, L. Umayam, L. Jiang, E. Holtzapple, J. D. Busch, K. L. Smith, J. M. Schupp, D. Solomon, P. Keim, and C. M. Fraser. 2002. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* **296**:2028–2033.
- Rygus, T., and W. Hillen. 1991. Inducible high-level expression of heterologous genes in *Bacillus megaterium* using the regulatory elements of the xylose-utilization operon. *Appl. Microbiol. Biotechnol.* **35**:594–599.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
- Scholle, M. D., C. A. White, M. Kunnimalaiyaan, and P. S. Vary. 2003. Sequencing and characterization of pBM400 from *Bacillus megaterium* QM B1551. *Appl. Environ. Microbiol.* **69**:6888–6898.
- Shivakumar, A. G., R. I. Vanags, D. R. Willcox, L. Katz, P. S. Vary, and J. L. Fox. 1989. Gene dosage effect on the expression of the delta-endotoxin genes of *Bacillus thuringiensis* subsp. *kurstaki* in *Bacillus subtilis* and *Bacillus megaterium*. *Gene* **79**:21–31.
- Stevenson, D. M., M. Kunnimalaiyaan, K. Müller, and P. S. Vary. 1998. Characterization of a theta plasmid replicon with homology to all four large plasmids of *Bacillus megaterium* QM B1551. *Plasmid* **40**:175–189.
- Sussman, M. D., P. S. Vary, C. Hartman, and P. Setlow. 1988. Integration and mapping of *Bacillus megaterium* genes which code for small, acid-soluble spore proteins and their protease. *J. Bacteriol.* **170**:4942–4945.
- Takami, H., K. Nakasone, Y. Takaki, G. Maeno, R. Sasaki, N. Masui, F. Fuji, C. Hirama, Y. Nakamura, N. Ogasawara, S. Kuhara, and K. Horikoshi. 2000. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res.* **28**:4317–4331.
- Thomas, C. M. 2000. The horizontal gene pool. Bacterial plasmids and gene spread. Harwood Academic Publishers, Amsterdam, The Netherlands.
- Thorsted, P. B., C. M. Thomas, E. U. Poluektova, and A. A. Prozorov. 1999. Complete sequence of *Bacillus subtilis* plasmid p1414 and comparison with seven other plasmid types found in Russian soil isolates of *Bacillus subtilis*. *Plasmid* **41**:274–281.
- Tribuddharat, C., and M. Fennewald. 1999. Integrin-mediated rifampin resistance in *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **43**:960–962.
- Vary, P. S. 1994. Prime time for *Bacillus megaterium*. *Microbiology* **140**:1001–1013.
- Von Tersch, M. A., and H. L. Robbins. 1990. Efficient cloning in *Bacillus megaterium*: comparison to *Bacillus subtilis* and *Escherichia coli* cloning hosts. *FEMS Microbiol. Lett.* **70**:305–310.
- Wang, A. J., and F. L. Macrina. 1995. Streptococcal plasmid pIP501 has a functional *oriT* site. *J. Bacteriol.* **177**:4199–4206.
- Wang, T. T., and B. H. Lee. 1997. Plasmids in *Lactobacillus*. *Crit. Rev. Biotechnol.* **17**:227–272.