**Complete nucleotide sequence and determination of the replication region
of the sporulation inhibiting plasmid p576 from *Bacillus pumilus* NRS5766**

Praveen K. Singh¹, Sandra Ballesterro-Beltrán¹, Gayetri Ramachandran and Wilfried J.J.
Meijer*

*Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Instituto de Biología
Molecular “Eladio Viñuela” (CSIC), Universidad Autónoma, Canto Blanco, 28049 Madrid,
Spain*

1: Contributed equally

*: Corresponding author.

Phone: (+34) 91 196 4515

Fax: (+34) 91 196 4420

email: wmeijer@cbm.uam.es

Abstract

Large plasmids, presumably replicating via the theta mechanism, have been identified in numerous gram positive bacteria. However, their characterization is rather poor and predominantly limited to those harbored by some (opportunistic) pathogenic bacteria. Here we determined the DNA sequence of the 43.3 kb plasmid p576 from *Bacillus pumilus* strain NRS576, being the first *B. pumilus* theta replicating plasmid sequenced. Plasmid p576 has a modular structure but strangely it does not seem to encode a Rep protein found on most theta replicating plasmids. However, an ~1 kb region was identified showing homology with the Rep-independent replication region of *Bacillus subtilis* plasmid pLS20, and we demonstrated that this region is sufficient for autonomous replication. The plasmid contains various large direct repeated sequences. A likely function could be attributed to at least 15 putative p576 genes. Some of these are predicted to be involved in stable maintenance of the plasmid; others are likely to encode proteins involved in conjugation. p576 also carries a *rap-phr* cassette whose possible function is discussed.

Keywords: *Bacillus pumilus*, plasmid, conjugation, DNA replication, sporulation

1. Introduction

Circular plasmids are commonly present in gram-positive and –negative bacteria. Based on their mode of replication they can be grouped into two classes. One group replicates via the rolling-circle mechanism and the other group via the theta mechanism. Generally, small (< 10 kb) and large plasmids replicate via the rolling circle and theta mechanism of replication, respectively (for review see, Meijer et al., 1998). Various large theta replicating plasmids from gram-negative bacteria have been studied extensively. However, although numerous large plasmids have been identified in gram-positive bacteria, they are rather poorly characterized with a few exceptions (Titok et al., 2003; Liroy et al., 2010). Many large plasmids carry conjugation genes permitting them to be transferred horizontally, sometimes even to distantly-related species. In addition, rolling circle replicating plasmids, which on their own cannot be transferred horizontally, can often be transferred (via the process named mobilization) when co-resident with a large conjugative theta type plasmid. Thus, together with bacteriophages and transposons, conjugative plasmids form part of a large prokaryotic mobile gene pool or “mobilome” that, in combination with (homologous) recombination, can enable rapid and profound prokaryotic genomic rearrangements (for review see, Frost et al., 2005; Gogarten and Townsend, 2005). This warrants a better understanding of gram-positive theta replicating plasmids. Our attention was drawn by a publication describing the presence of a naturally occurring 45 kb plasmid, p576, in the *Bacillus pumilus* strain NRS576 (Lovett and Bramucci, 1974). No complete or partial DNA sequence of any *B. pumilus* theta replicating plasmid is available. In addition, the presence of p576 appears to inhibit sporulation of its host (Lovett, 1973). As a first impetus to improve our knowledge on this gram-positive plasmid we determined its sequence. Analysis of the p576 sequence revealed that it has a modular structure, as generally observed for plasmids. A function could be attributed to at least 15 putative p576 genes. Some of these putative genes are predicted to be

involved in stable maintenance of the plasmid. Others, which are located in a putative large operon, are predicted to encode essential conjugation proteins. In addition, we identified a *rap-phr* cassette that might be involved in its associated phenotypic trait to inhibit sporulation. Finally, we determined the minimal replication region and show that it is sufficient to drive replication in *Bacillus subtilis*.

2. Materials and Methods

2.1. Bacterial strains, plasmids and growth conditions.

Bacterial strains and plasmids used in this study are listed in Table 1. *Escherichia coli* and *B. subtilis* strains were grown in Luria-Bertani (LB) medium or on 1.5% LB agar plates (Sambrook et al., 1989). When appropriate, media or agar plates were supplemented with the following antibiotics: ampicillin (100 µg/ml), erythromycin (1 and 150 µg/ml in *B. subtilis* and *E. coli*, respectively). Insert containing derivatives of pBluescript KS+ were selected on LB agar plates supplemented with IPTG and Xgal (5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside) to final concentrations of 1 mM and 40 µg/ml, respectively.

2.2 Plasmid p576 isolation and other DNA techniques.

Plasmid p576 used for cloning and direct sequencing was isolated by isopycnic cesium chloride ethidium bromide gradients (Sambrook et al., 1989). Other DNA techniques were performed using standard molecular methods (Sambrook et al., 1989).

2.3. Subcloning and sequence determination of plasmid p576

Two different approaches were used to clone p576 fragments in pBluescript KS+ vector. In the first approach, *Hind*III digested fragments of p576 were cloned in the corresponding site of pBluescript KS+. Five out of the nine p576 *Hind*III fragments were cloned. In the second approach, purified p576 DNA was partially digested with *Sau*3A and then fragments ranging in size between 0.6 and 2 kb were isolated from agarose gels. Next, these fragments were cloned in the pBluescript KS+ vector. To avoid cloning of multiple fragments in the same vector the following strategy was followed. The p576 *Sau*3A fragments were treated with Klenow enzyme in the presence of GTP and ATP nucleotides. Similarly, *Sa*I linearized pBluescript was treated with Klenow enzyme in the presence of nucleotides

CTP and TTP. Next, a ligation mixture of these partially filled-in fragments was used to transform competent *E. coli* XL1-Blue cells. Finally, plasmid DNA was isolated from 100 white transformants grown on IPTG/XGal/ampicillin complemented agar plates and the borders of the cloned inserts were sequenced using universal primers M13Rev (5'-CAGGAAACAGCTATGACC-3') and M13Fw-21 (5'-TGTAACGACGGCCAGT-3'). The DNA sequences obtained were then used to design primers (purchased from Isogen Life Science, The Netherlands) which were used in a subsequent round of DNA sequencing using Maxi-Prep purified p576 as template DNA. p576 was then sequenced to completeness in additional rounds by the primer walking strategy. Sequencing was performed by capillary electrophoresis on an ABI PRISM 3730xl equipment (Applied Biosystems) with BigDye Terminators v3.1.

2.4. Sequence analysis

DNA sequences were analyzed and assembled using the Lasergene software package of DNASTAR, Inc (Madison, Wi, USA). Putative genes were initially identified using the PRODIGAL program (Hyatt et al., 2010) and subsequently checked manually. Additional DNA analyses were performed using Sci Ed Central Clone Manager Profession Suite (version 6.0) in combination with the online tools of Blast (www.blast.ncbi.nlm.nih.gov/Blast.cgi) and Expasy (www.expasy.ch/).

2.5. Cloning of the minimal replicon region of p576

PCR was used to amplify the p576 region showing 72% identity with the minimum replicon region of pLS20 using the proofreading-proficient Vent DNA polymerase (NEBiolabs). Purified p576 DNA was used as template DNA together with primers ori576_UP (5'-ccccggatcCAAGGGGGGGGACTTTACTATGA-3') and ori576_Down1 (5'-

tttggatccTGCGCCCCGGTATATTATCGTC-3'). The obtained PCR product was digested with *Bam*HI and cloned into the unique *Bam*HI site of pMTL23ET, resulting into vectors pNRS100E_A and –_B depending on the orientation of the insert. The correctness of the inserts was verified by sequence analysis using primers M13Fw-21 and EMTerSeq2 (5'-cccacagatacggtaaactagc-3').

3. Results and discussion

3.1 General features of plasmid p576

Plasmid p576 has a total size of 43,434 bp and its (G+C) content is 37.2%. Sequence analysis indicated that p576 contains 67 ORFs with putative ribosome-binding sites (Fig. 1). Position 1 of the plasmid was arbitrarily chosen and corresponds to the “A” of the ATG start codon of the first putative gene. Fifty three ORFs are transcribed rightwards. The fourteen leftward reading ORFs are distinguished from the rightward oriented ones by extension of the letter “c”. The deduced amino acid sequences of identified ORFs were compared by the online BLAST and Expasy tools (see Materials and Methods). A summary of the results obtained, which includes details on the putative translation start sites, is given in Table 2. Based on the identified features and ORFs, four functional modules were identified: (i) replication region, (ii) plasmid stability functions, (iii) *rap-phr*, and (iv) conjugative transfer module. In addition, p576 contains another region spanning 15 ORFs. The deduced protein sequences of most of these ORFs lack significant similarity with proteins present in available databases. Finally, plasmid p576 is characterized by the presence of various regions containing different types of direct repeated sequences. These modules/regions are discussed in more detail below.

3.2. The replication region.

Most large theta-replicating plasmids encode a protein, Rep, that is essential for plasmid DNA replication (Meijer, 1995; Weaver et al., 2009). Strikingly, none of the 67 identified putative genes of p576 was predicted to encode a replication protein. However, a nucleotide BLAST search resulted in the identification of an approximately 1 kb p576 region showing 72% identity with the replicon region of the ~ 60 kb *B. subtilis natto* plasmid pLS20 (Meijer et al., 1995). We cloned a 1,056 bp fragment of p576 encompassing this region into

the *E. coli* vector pMTL23E which contains a gram-positive selectable erythromycin resistance marker resulting in plasmid pNRS100E. Next, we used pMTL23E and pNRS100E plasmid DNA to transform competent *B. subtilis* 168 cells. Erythromycin resistant colonies were obtained only with pNRS100E. Analysis of ten transformants revealed that they all contained autonomously replicating plasmid pNRS100E (not shown). These results demonstrate that this region of p576 is sufficient to sustain autonomous replication in *B. subtilis* and hence corresponds to the replication region of p576.

The replication region of pLS20 was placed in a new class of theta replicons based on its unique features (Meijer et al., 1995). The alignment of the ~1 kb regions of p576 and pLS20 presented in Figure 2 shows that the identified p576 region shares most of the characteristics described for the pLS20 replicon. In both cases the regions (i) are situated in between divergently oriented putative genes, (ii) do not encode a typical Rep protein, (iii) are characterized by the presence of several regions of dyad symmetry, (iv) contain four (im)perfect DnaA boxes, and (v) contain an ~ 35 bp region with high A/T content (>80%). Moreover, the position of these features is conserved in both plasmids. Based on these similarities we propose that the replication region of p576 belongs to the same class as that of pLS20. As far as we know, no other plasmids have been described belonging to this class of theta replicons.

It is worth mentioning that there are two differences between the replication regions of p576 and pLS20 (Fig. 2). First, contrary to that of pLS20, the replication region of p576 does not contain a perfect DnaA box. And second, a functional bidirectional DNA replication terminator site, located at the right side of the pLS20 replicon (Meijer et al., 1996; Mohanty et al., 2001), is absent at a similar position or elsewhere on p576. The presence of a functional replication terminator site flanking the pLS20 origin region probably causes replication of this plasmid to occur unidirectionally whereas the absence of a terminator suggests that replication

of p576 proceeds bidirectionally. Support for this latter assumption is the observation that most of the putative genes located on the left and the right side of the origin region are divergently transcribed (see Fig. 1). As a result, the transcription and bidirectional replication machineries move mostly in the same direction thereby minimizing head-on collisions between them. It is generally believed that this is the reason why most of the highly active genes in prokaryotic genomes are co-transcribed with the direction of replication (Brewer, 1988).

3.3. Plasmid stability functions.

Plasmids, per definition, do not carry essential genes for cell survival under standard conditions and as such are prone to be lost over time. In order to be maintained stably over time plasmids must either (i) provide a selective advantage to the cell, (ii) possess effective systems that ensure accurate segregation and/or systems that otherwise prevent plasmid loss, or (iii) possess functions allowing horizontal transfer of the plasmid that may compensate for possible vertical loss of the plasmid and even enhance spreading of the plasmid to other (related) species. These functions are especially important for plasmids having a low copy-number. Presently, we do not have an indication that p576 provides a selective advantage to its host. However, p576 does carry putative conjugation genes suggesting that the plasmid can be transferred horizontally (see below) and others that are likely to ensure its stable maintenance. Studies on low-copy plasmids have revealed the presence of three main different systems that contribute to their stable maintenance: (i) resolution of newly replicated plasmids (multimer resolution systems), (ii) true partitioning systems ensuring each daughter cell to receive at least one copy of the plasmid before cell division, and (iii) post-segregational killing (PSK) systems that function by killing plasmid-free segregants (Møller-Jensen et al.,

2000; Ebersbach and Gerdes, 2005). As outlined below plasmid p576 carries at least two of these systems.

Multimer resolution system: ORF16 of p576 is predicted to encode a protein belonging to resolvase and invertase subfamily of serine recombinases. Proteins belonging to this family catalyze intra-molecular recombinations leading either to the resolution or inversion depending on the orientation of the two cognate recombination sites on which they act. Resolution occurs when the two recombination sites are located in the same orientation and inversion when they are present in inverse orientation (Grindley et al., 2006). Hence it is likely that ORF16 of p576 encodes a resolvase that is responsible for resolving p576 dimers or higher order multimers, generated either by plasmid replication or homologous recombination, into monomers thereby contributing to the stable maintenance of the plasmid. Accordingly, ORF16 is named *res576*. The cis-acting recombination site of Res576 remains to be determined.

True partitioning system. True partitioning systems, also named ParAB-like partitioning systems, have been identified on various low-copy number plasmids, as well as on phage genomes and chromosomes (for review see, Ebersbach and Gerdes, 2005; Thanbichler and Shapiro, 2008). Generally, these are tripartite loci consisting of two genes that code for an actin-like ATPase (often named ParA) and a DNA-binding protein (often named ParB) and one or more centromere-like cis-acting sites (often named *parS*). ParB binds preferentially to *parS* and the actin-like ParA is responsible for physical plasmid segregation.

The deduced protein sequences of p576 ORFs 33c and 32c share a high level of similarity with the plasmid pLS20 ParA-type encoded ATPase Alp7A (74% identity) and the presumed ParB-type DNA binding protein AlpR (45% identity). And, as expected, the putative protein encoded by p576 ORF33c contains the characteristic ATPase domain patterns (not shown) as defined by Bork et al. (1992). Both, the p576 ORFs 33c-32c and pLS20 *alpA*-

R operon flank their respective replication region (see Fig. 2B). The pLS20 *alp7A-R* operon has recently been demonstrated to be required for plasmid stability (Derman et al., 2009). These authors also showed that Alp7A forms dynamic filaments in the cell that are able to treadmill. These data indicate that the *alp7A-R* partitioning mechanism resembles fundamentally the well-studied ParM system of *E. coli* plasmid R1; thus, plasmids are associated with dynamic filaments and the elongation of the filaments between the plasmids is responsible for their separation.

Based on the high level of similarity it is likely that the p576 ORFs 33c-32c form part of the p576 true partition system. Accordingly we denominate ORF33c and ORF32c *alp576A* and *alp576R*, respectively.

Post-segregational killing systems. These systems, which are also known as addiction or toxin-antitoxin systems, have been identified in numerous plasmids as well as in bacterial chromosomes (for review see, Ebersbach and Gerdes, 2005; Gerdes et al., 2005). PSK systems, which kill (or arrest growth of) plasmid-free cells that may arise after cell division, are composed of an operon of at least two essential genes encoding a stable toxin and an unstable antitoxin. In a plasmid-free cell the unstable antitoxin is rapidly degraded liberating the destructive activity of the toxin. PSK systems can be divided into two distinct classes: (i) those in which the toxin and antitoxin are both proteins, and (ii) those in which the antitoxin is an antisense RNA molecule that directly or indirectly inhibits translation of the toxin. Although the analysis of p576 did not reveal a clear PSK system, the deduced protein sequence of ORF15 shares significant similarity with toxins belonging to the PemK family.

3.4 *Rap-phr* module.

The deduced protein sequence of p576 ORF34 (370 residues) shares significant homology with numerous members of the Rap families of proteins. *Rap* genes generally form

part of a *rap-phr* operon that are found on bacterial genomes of most if not all *Bacillus* species as well on various rolling circle and theta replicating plasmids from *Bacilli* (Meijer et al., 1995; Meijer et al., 1998; Perego and Hoch, 2002). The *B. subtilis* genome contains 11 *rap* genes, seven of which are genetically coupled to a *phr* gene. For several *Rap-phr* cassettes it has been shown that they are involved in the regulation of developmental processes such as sporulation, competence development and exoprotease production (Perego et al., 1994; Smits et al., 2007). Rap proteins are characterized by the presence of tetratricopeptide repeat (TPR) domains and are shown to affect the DNA binding activity of specific response regulators either directly or indirectly. The downstream located *phr* genes encode small (~ 40 residues) precursors, which, after an export-import maturation process, inhibit the activity of their cognate Rap protein. Thus, after being exported by the secretory pathway the Phr peptides are further processed into mature pentapeptides, which are transported into the cell by an oligopeptide permease. Plasmid p576 ORF34 is followed by a small ORF (34 codons) that most probably encodes a precursor of a secreted Phr molecule. Based on this we designate ORF34 and ORF35 *rap576* and *phr576*, respectively.

The *Bacillus anthracis* virulence plasmid pXO1 carries a *rap-phr* cassette and it has been shown that this cassette regulates initiation of sporulation of this organism (Bongiorni et al., 2006). It is possible that the *rap-phr* cassette present on p576 is involved in the plasmid-mediated inhibition of sporulation of its host. Finally, it is worth mentioning that (i) Rap576 shares highest level of similarity with the putative Rap protein encoded by the *B. subtilis* plasmid pLS20 (41% identity, Meijer et al., 1995), and (ii) that the *rap-phr* cassettes of p576 and pLS20 both flank the replication region (see Fig. 2B).

3.5. Conjugative transfer module.

Bacterial conjugation is a highly specific process by which DNA is transferred from a donor to a recipient cell via specialized multiprotein complexes, named Type IV secretion systems (T4SS) (for recent review on T4SS see, Alvarez-Martinez and Christie, 2009). Many large theta replicating plasmids carry a functional conjugation system, thereby making them efficient vehicles for horizontal gene spread. Conjugative plasmids and Integrative Conjugative Elements (ICE, aka conjugative transposons) are considered to be one of the major reasons for the increase in antibiotic-resistant bacteria. Because of this, most studies on conjugative plasmids are related to clinically relevant bacteria. Extensive studies have been performed especially on the conjugation systems of a few gram-negative related plasmids and on the *Agrobacterium tumefaciens* Ti plasmid. In comparison, information on gram-positive plasmid-located conjugation systems is relatively scarce (for review see, Grohmann et al., 2003; Clewell and Francia, 2004). Whereas diverse mechanisms are responsible for proper expression of different plasmid-encoded T4SSs, the overall process of conjugative DNA transfer is similar for all plasmids and can be divided into three biochemical reactions: DNA substrate processing, substrate recruitment, and translocation. Plasmid p576 contains a putative large operon (ORF37 to ORF11 [or possibly ORF17]) encoding probably all essential proteins required for a functional conjugation apparatus, strongly indicating that p576 is a conjugative plasmid. A succinct overview of some of the main features of gram positive conjugative T4SSs is given below including the putative conjugation genes identified on p576. More extensive overviews of plasmid conjugation are described elsewhere (Clewell and Francia, 2004; Grohmann et al., 2003; Alvarez-Martinez and Christie, 2009).

Processing of the DNA substrate for conjugation is initiated by formation of a relaxosome, i.e. the binding of a relaxase and one or more accessory factors to the origin-of-transfer (*oriT*). The relaxase then produces a site-specific nick in the transferred DNA strand.

After nicking, the relaxase remains bound to the 5' end of the nicked DNA, and “delivers” the substrate to the translocation channel by direct interaction with a conserved ATPase named Type IV coupling protein (T4CP, aka TraG/TrwB/VirD4 family of coupling proteins). In the recipient cell, the relaxase catalyzes the recircularization of the transferred single-stranded DNA molecule and may also participate in double-strand DNA synthesis. A putative relaxase, is encoded by p576 ORF10. The *oriT* of p576 remains to be identified. The T4CP physically interacts with the translocation channel through which the DNA and some proteins are transported from the donor to the recipient cell. Thus, the T4CP forms an interface between the relaxome and the translocation channel. A putative T4CP protein is encoded by p576 ORF63. The translocation channel is comprised of the mating-pair formation (mpf) proteins, which can be divided in distinct functional classes. First, energetic subunits. Besides T4CP protein, many T4SSs encode two additional ATPases (homologs of the *A. tumefaciens* VirB4 and VirB11 ATPases) that energize early steps of the machine biogenesis or substrate transfer. Although most gram-positive T4SS studied so far lack a homolog of *virB11*, they are present on conjugative plasmids from a group of *Bacillus* species. Plasmid p576 is predicted to encode a VirB11 homolog (ORF53) as well as a VirB4 homolog (ORF3). Second, peptidoglycan hydrolase. These extracellularly located cell wall degrading proteins are believed to facilitate the assembly of the T4SS channel through localized degradation of the cell wall. A putative membrane-localized cell wall hydrolyzing enzyme is encoded by p576 ORF5. Third, other channel and scaffolding subunits. These proteins, several of which contain one or more transmembrane spanning domains, contribute in various ways to channel formation and activity. Various ORFs in the putative large conjugation operon of p576 are predicted to encode proteins containing one or more transmembrane spanning domains (see Fig. 1). Probably (some of) these encode channel and scaffold subunits of the conjugation channel.

3.6. Other putative p576 genes.

Most of the putative proteins encoded by p576 ORFs 18c to 31c do not show significant similarity with proteins in available databases and hence their function is unknown. A noteworthy exception is ORF28c whose deduced protein sequence shows significant similarity with ArdC type of antirestriction proteins. Antirestriction proteins are inhibitors of type I restriction-modification enzymes and are found on various conjugative plasmids, ICEs and bacteriophages (for review see, Zavilgelsky and Rastorguev, 2009). The structure of antirestriction proteins mimics the B-form of DNA. The N-terminal region of ArdC-type proteins, including putative p576 protein p28c, shows homology with TraC1 primases. Evidence exists that antirestriction proteins are transferred along with the ssDNA molecule into the recipient cell to protect the plasmid DNA from digestion by type I restriction enzymes. Based on these data, it is plausible that ORF28c and possibly surrounding ORFs are part of the p576 conjugation system.

3.7. Direct repeated sequences.

Analysis of the p576 DNA sequence revealed the presence of three types of direct repeats (DR, see Fig. 1 and Table 3). Versions of the type I DR are located upstream ORF12, ORF22c, ORF34 (*rap576*) and in between the convergently oriented ORFs 25 and 26c. These DR consist of a 47-50 bp sequence which is repeated between 2.5 (upstream ORF12) and 8.5 times (upstream *rap576*). Currently, the function of these direct repeats is unknown. The second type of DR concerns a 200 bp sequence that is repeated four times. Each of the 200 bp repeats contains an ORF (ORF57A-D) coding for a putative 45 residue membrane protein (except the last repeat which would encode a 48 residue membrane protein) that is preceded

by a potential RBS and followed by a putative Rho-independent transcriptional terminator. In the third type, the DRs are located within ORFs 61 and 64. In the case of ORF61, a 36 bp sequence is repeated 20 times, implying that the encoded protein would contain 20-fold repeat of a 12 amino acid motif. A similar situation occurs in ORF64. In this case a 13 times repeated sequence of 48 bp is immediately followed by a 13 times repeated sequence of 42 bp. Consequently, the encoded protein would contain two consecutive motifs of 16 and 14 amino acids, each being repeated 13 times. So far, the function of putative genes 61 and 64 is unknown.

Acknowledgements

The authors are grateful to the Genomics Unit “Antonia Martín Gallardo” of the “Fundación Parque Científico Madrid (FPCM)” for their excellent help with sequencing. This work was supported by the Grant BFU2008-04034/BMC from the Spanish Ministry of Science and Innovation to W.J.J.M., and an institutional grant from Fundación Ramón Areces to the Centro de Biología Molecular ‘Severo Ochoa’. P.K.R. is holder of a predoctoral fellowship from the Spanish Research Council.

References

- Alvarez-Martinez,C.E. and Christie,P.J. 2009. biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev* 73: 775-808.
- Bongiorni,C., Stoessel,R., Shoemaker,D., and Perego,M. 2006. Rap phosphatases of virulence plasmid pXO1 inhibits *Bacillus anthracis* sporulation. *J. Bacteriol.* 188: 487-498.
- Bork,P., Sander,C., and Valencia,A. 1992. An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc. Natl. Acad. Sci. USA* 89: 7290-7294.
- Brewer,B.J. 1988. When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. *Cell* 53: 679-686.
- Bullock,W.O., Fernandez,J.M., and Short,J.M. 1987. XL1-blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with Beta-galactosidase selection. *Biotechniques* 5: 376-379.
- Clewell,D.B. and Francia,M.V. 2004. Conjugation in gram-positive bacteria. *In* Funnell,B.E. and Philips,G.J., eds. *Plasmid biology*. ASM Press, Washington DC, pp 227-256.
- Derman,A.I., Becker,E.C., Truong,B.D., Fujioka,A., Tucey,T.M., Erb,M.L., Patterson,P.C., and Pogliano,J. 2009. Phylogenetic analysis identifies many uncharacterized actin-like proteins (Alps) in bacteria: regulated polymerization, dynamic instability and treadmilling in Alp7A. *Mol. Microbiol.* 73: 534-552.
- Ebersbach,G. and Gerdes,K. 2005. Plasmid segregation mechanisms. *Annu. Rev. Genet.* 39: 453-479.

- Frost,L.S., Leplae,R., Summers,A.O., and Toussaint,A. 2005. Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3: 722-732.
- Gerdes,K., Christensen,S.K., and Løbner-Olesen,A. 2005. Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Microbiol.* 3: 371-382.
- Gogarten,J.P. and Townsend,J.P. 2005. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3: 679-687.
- Grindley,N.D., Whiteson,K.L., and Rice,P.A. 2006. Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* 75: 567-605.
- Grohmann,E., Muth,G., and Espinosa,M. 2003. Conjugative plasmid transfer in gram-positive bacteria. *Microbiol. Mol. Biol. Rev* 67: 277-301.
- Hyatt,D., Chen,G.L., LoCascio,P.F., Land,M.L., Larimer,F.W., and Hauser,L.J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11: 130.
- Lioy,V.S., Pratto,F., de la Hoz,A.B., Ayora,S., and Alonso,J.C. 2010. Plasmid pSM19035, a model to study stable maintenance in firmicutes. *Plasmid* 64: 1-17.
- Lovett,P.S. 1973. Plasmid in *Bacillus pumilus* and the enhanced sporulation of plasmid-negative variants. *J. Bacteriol.* 115: 291-298.
- Lovett,P.S. and Bramucci,M.G. 1974. Biochemical studies of two *Bacillus pumilus* plasmids. *J. Bacteriol.* 120: 488-494.
- Meijer,W.J.J. 1995. Replication and maintenance of plasmids in *Bacillus subtilis*.

- Meijer, W.J.J., de Boer, A., van Tongeren, S., Venema, G., and Bron, S. 1995. Characterization of the replication region of the *Bacillus subtilis* plasmid pLS20: a novel type of replicon. *Nucleic Acids Res.* 23: 3214-3223.
- Meijer, W.J.J., Smith, M., Wake, R.G., de Boer, A.L., Venema, G., and Bron, S. 1996. Identification and characterization of a novel type of replication terminator with bidirectional activity on the *Bacillus subtilis* theta plasmid pLS20. *Mol. Microbiol.* 19: 1295-1306.
- Meijer, W.J.J., Wisman, G.B.A., Terpstra, P., Thorsted, P.B., Thomas, C.M., Holsappel, S., Venema, G., and Bron, S. 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.
- Mohanty, B.K., Bussiere, D.E., Sahoo, T., Pai, K.S., Meijer, W.J.J., Bron, S., and Bastia, D. 2001. Structural and functional analysis of a bipolar replication terminus: Implications for the origin of polarity of fork arrest. *J. Biol. Chem.* 276: 13160-13168.
- Møller-Jensen, J., Jensen, R.B., and Gerdes, K. 2000. Plasmid and chromosome segregation in prokaryotes. *Trends Microbiol.* 8: 313-320.
- Perego, M., Hanstein, C., Welsh, K.M., Djavakhishvili, T., Glaser, P., and Hoch, J.A. 1994. Multiple protein aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in *Bacillus subtilis*. *Cell* 79: 1047-1055.
- Perego, M. and Hoch, J.A. 2002. Two-component systems, phosphorelays, and regulation of their activities by phosphatases. In Sonenshein, A.L., Hoch, J.A., and Losick, R., eds. *Bacillus subtilis* and its closest relatives: from genes to cells. American society for microbiology press, Washington, DC, pp 473-481.

Sambrook,J., Fritsch,E.F., and Maniatis,T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Smits,W.K., Bongiorno,C., Veening,J.W., Hamoen,L.W., Kuipers,O.P., and Perego,M. 2007. Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in *Bacillus subtilis*. Mol. Microbiol. 65: 103-120.

Thanbichler,M. and Shapiro,L. 2008. Getting organized. How bacterial cells move proteins and DNA. Nat. Rev. Microbiol. 6: 28-40.

Titok,M.A., Chapuis,J., Selezneva,Y.V., Lagodich,A.V., Prokulevich,V.A., Ehrlich,S.D., and Jannire,L. 2003. *Bacillus subtilis* soil isolates: plasmid replicon analysis and construction of a new theta-replicating vector. Plasmid 49: 53-62.

Weaver,K.E., Kwong,S.M., Firth,N., and Francia,M.V. 2009. The RepA_N replicons of Gram-positive bacteria: a family of broadly distributed but narrow host range plasmids. Plasmid 61: 94-109.

Zavilgelsky,G.B. and Rastorguev,S.M. 2009. Antirestriction proteins ArdA and Ocr as efficient inhibitors of type I restriction-modification enzymes. Mol. Biol. 43: 241-248.

Legends to Figures

Fig. 1. Genetic organization of plasmid p576. The position of the replication region (origin) is indicated by a rectangular, those of ORFs and their direction of transcription are indicated with arrows. ORFs coding for putative membrane proteins are shown in brown, and those predicted to encode for proteins to be exported by the sec-dependent pathway are shown in blue. Hairpins indicate stem-loop structures which may constitute potential Rho-independent terminators. The different types of large Direct Repeats are indicated with roman numbered boxes (type I, red; type II, light green; type III, grey).

Fig. 2. Similarities between the replication regions of plasmids p576 and pLS20 from *B. pumilis* NRS576 and *B. subtilis* natto UM4, respectively. (A) Alignment of the DNA sequences corresponding to the minimal replication regions of p576 (576) and pLS20 (LS20). Sequences shown correspond to p576 positions 19,974-20,907 and positions 491-1,452 of pLS20 accession number U26059. Identical sequences are shown by dots. The ~35 bp A/T rich region (>80%) is indicated in grey, inverted repeated sequences are indicated by arrows. Imperfect and perfect DnaA boxes, which are boxed, are marked with $\approx$ DnaA and DnaA, respectively. (B) Schematic representation of p576 and pLS20 plasmid regions encompassing their replication regions. In both plasmids the replication region is flanked at the left and right side by an *alp* and *rap* gene, respectively. The conserved positions on both plasmids of the following features are indicated: (i) A/T rich region, grey box marked A/T; (ii) dyad symmetry (numbered as in Fig. 2A), hairpins; and (iii) imperfect DnaA boxes, open triangles. The differences between p576 and pLS20 are indicated as follows. The single consensus DnaA box sequence in pLS20 is indicated by a filled triangle below the line. The region upstream the *rap* gene of p576 contains several direct repeated sequences (indicated by

>>>>), and that of pLS20 contains a bidirectional replication terminator (indicated by a bow tie symbol).

Table 1
Bacterial strains and plasmids used in this study

Strains	Properties	reference/source
<i>B. pumilus</i> NRS576	Natural isolate harbouring plasmid p576	(Lovett, 1973), BGSC*
<i>B. subtilis</i> 168	<i>trpC2</i>	BGSC*
<i>E. coli</i> XL1-Blue	endA1 gyrA96(nal ^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB ⁺ lacI ^q Δ(lacZ)M15] hsdR17(r _K ⁻ m _K ⁺)	(Bullock et al., 1987)
Plasmids		
pBluescript KS+	High-copy <i>E. coli</i> cloning vector	Stratagene (USA)
pMTL23ET	pMTL23 derivative containing erythromycin resistance gene from pE194 (cloned in <i>Cla</i> I site) and <i>E. coli rrnB</i> terminator (cloned in <i>Eco</i> RI site)	Laboratory stock
pNRS100E_A/B	pMTL23ET containing 1054 bp minimal replication region from p576 (positions 19975-21028). A and B correspond to different orientations of the insert	This study

*, BGSC: Bacillus Genetic Stock Center, Department of Biochemistry, The Ohio State University, Columbus, OH; USA.

Table 2
Identification of genes and ORFs on plasmid p576

ORF	(start - stop)	Size (aa)	Putative function and/or most relevant homology	E-value	Putative RBS
			- Micro organism (amino acid identity %) - Putative conserved domain (PCD)		
1	1-306	101	hypothetical protein LmonocyFSL_00185 / <i>Listeria monocytogenes</i> FSL J1-208 (33%)	2e-6	AAAGGAAT-8-GTG
2	299-868	189	hypothetical protein EET96815.1/ <i>Enterococcus faecalis</i> T1 (44%)	5e-36	AAAGGTGA-6-ATG
3	881-2776	631	T4SS VirB4 ATPase hypothetical protein EET96816.1 <i>Enterococcus faecalis</i> T1 (50%) PCD: TraC-F-type	2e-176	CTAGGAGG-8-GTG
4	2778-3371	197	hypothetical membrane protein		GAAGGAGG-8-TTG
5	3364-4473	369	cell wall hydrolase putative lipoprotein ZP_05294309.1 <i>Listeria monocytogenes</i> FSL J1-208 (54%) PCD: - Lysozyme like superfamily - NLPC60 superfamily	1e-98	AAAGGAAT-8-ATG
6	4490-5290	266	hypothetical protein LmonocyFSL_00155 <i>Listeria monocytogenes</i> FSL J1-208 (26%)	7e-18	AAAGGAGA-7-TTG
7	5287-5568	93	hypothetical protein		AAAGGAAG-17-TTG
8	5570-5812	80	hypothetical protein EET97059.1 <i>Enterococcus faecalis</i> T2 (33%)	3e-4	AAAGGGAG-17-ATG
9	5820-6239	139	hypothetical protein LmonocyFSL_00140 <i>Listeria monocytogenes</i> FSL J1-208 (31%)	1e-4	GAAGGATG-12-ATG
10	6239-7447	402	Relaxase hypothetical protein LmonocyFSL_00135 <i>Listeria monocytogenes</i> FSL J1-208 (50%)	7e-113	CAAGAAGG-12-GTG
11	7478-8839	453	Putative DNA primase hypothetical protein EET97064.1 <i>Enterococcus faecalis</i> T2 (37%)	4e-76	ACTGGTGG-6-TTG
12	9026-9343	105	hypothetical protein EDZ49916.1 <i>Bacillus cereus</i> AH1134 (28%)	2e-8	TATAGAGG-9-ATG
13	9334-9504	56	hypothetical protein		ATGGGTGG-5-ATG
14	9497-9643	48	hypothetical protein		AAAGGACG-7-ATG
15	9644-10096	150	hypothetical protein Dtox_0798 Desulfotomaculum acetoxidans DSM 771 (33%) PCD: PemK superfamily	1e-6	CAAGGAAA-8-GTG
16	10230-10835	201	Recombinase/resolvase Recombinase EEK64355.1 <i>Bacillus cereus</i> MM3 (68%)	2e-78	AAAGGTGG-7-ATG

PCD: Serine Recombinase family, Resolvase and Invertase subfamily.					
17	10832-11323	163	Nuclease Thermonuclease GB ADE69862.1 <i>Bacillus megaterium</i> QM B1551 (51%)	7e-35	AAAAGAGA-9-ATG
PCD: Staphylococcal nuclease homologues. (SNc), Micrococcal nuclease (thermonuclease) homologs					
18c	11659-11393	88	hypothetical protein Plarl_00478 <i>Paenibacillus larvae</i> subsp. <i>larvae</i> BRL-230010 (33%)	5e-8	ACGGGAGG-8-ATG
19c	11961-11719	80	hypothetical protein		GCAGGAGG-7-ATG
20c	12259-11993	88	hypothetical protein		AAAGGAGA-7-ATG
21c	12402-12328	24	hypothetical protein		AAAGGAGA-14-ATG
22c	12772-12437	111	hypothetical protein SPBc2p016 <i>Bacillus</i> phage SPBc2 (39%)	3e-12	TTTGGAGG-8-ATG
PCD: YodD superfamily, Members of this family are functionally uncharacterized. However it has been predicted that these proteins are functionally equivalent to the UmuD subunit of polymerase V from gram-negative bacteria					
23c	14149-13304	281	hypothetical protein		TAAGGAGT-8-ATG
24	14471-14941	156	putative lipoprotein gb: EDW20705.1 <i>Bacillus pumilus</i> ATCC 7061 (78%)]	4e-67	TAGGAAGG-10-TTG
25	14944-15078	44	hypothetical protein BAT_1626 <i>Bacillus pumilus</i> ATCC 7061 (88%)	2e-5	TATGGACA-7-ATG
26c	15733-15527	68	hypothetical protein bthur0002_59040 <i>Bacillus thuringiensis</i> Bt407 (44%)	1e-2	AAAAGAGG-9-ATG
27c	15938-15726	70	hypothetical protein BALH_p0035 <i>Bacillus thuringiensis</i> str. Al Hakam (50%)	2e-5	GAAGGAGT-8-ATG
28c	16862-16023	279	ArdC-type Antirestriction protein hypothetical protein BpOF4_21454 <i>Bacillus pseudofirmus</i> OF4 (65%)	2e-101	AGAGGAAG-8-ATG
PCD: COG4227 (aa1-279: Antirestriction protein [DNA replication, recombination, and repair])					
29	17475-17594	39	hypothetical protein		CCAGGAGG-8-ATG
30c	17848-17591	85	hypothetical protein		AGGGGAGT-7-ATG
31c	18221-18018	67	hypothetical protein		TGAGGGGG-7-ATG
32c	18627-18241	128	Alp576R (plasmid maintenance: True partitioning) Alp7R gb: ACU27364.1	1e-20	GATGTAGG-8-ATG

			<i>Bacillus subtilis</i> natto plasmid pLS20 (45%)		
33c	19816-18647	389	Alp576A (plasmid maintenance: True partitioning, actin-like ATPase) Alp7A GB: ACU27363.1 <i>Bacillus subtilis</i> natto plasmid pLS20 (73%)	2e-170	AAAGGAGA-9-ATG
34	21583-22695	370	Rap (Response regulator Aspartate Phosphatase) <i>Bacillus subtilis</i> plasmid pLS20 (41%)	4e-74	TTAGGAGG-10-ATG
35	22695-22799	34	Putative phr protein		ATTGGAGG-10-ATG
36c	23233-22850	127	Xre tyep transcriptional regulator DNA-binding protein gb: ABP52080.1 <i>Bacillus subtilis</i> plasmid p19 (35%) PCD: Helix-turn-helix XRE-family like proteins. Prokaryotic DNA binding proteins belonging to the xenobiotic response element family of transcriptional regulators.	3e-12	ATATGAGG-8-TTG
37	23639-24157	172	conserved hypothetical protein gb: EEN76786.1 <i>Listeria grayi</i> DSM 20601] (31%)	9e-10	GAAGGAGA-9-GTG
38	24493-25521	342	conserved hypothetical protein gb: EEN76782.1 <i>Listeria grayi</i> DSM 20601 (25%)	1e-07	AAAGGTGT-7-ATG
39	25521-25685	54	hypothetical membrane protein		TTGGGAGT-10-ATG
40	25712-25897	61	hypothetical membrane protein		AAAGGTTG-8-ATG
41	25881-26009	42	hypothetical protein		TATGGTGT-4-ATG
42	25933-26127	64	hypothetical membrane protein		AAAGGAAA-8-ATG
43	26191-26505	104	hypothetical protein		AAAGGAGT-8-TTG
44	26492-26686	64	hypothetical membrane protein		AAAGGATG-10-ATG
45	26768-26962	64	hypothetical membrane protein		GTGAGAGG-13-ATG
46	27008-27427	139	hypothetical membrane protein		TTTGGAGG-7-TTG
47	27443-27664	73	hypothetical membrane protein		TAATGAGG-10-ATG
48	27748-30045	765	hypothetical protein LM5578_p32 <i>Listeria monocytogenes</i> 08-5578 (45%)	1e-180	AAAGGAGG-14-ATG
49	30042-30197	51	hypothetical protein		TAAGGAGT-8-GTG
50	30265-30663	132	Single Strand DNA binding protein single-strand DNA-binding protein gb:	8e-38	AAAGGTGA-7-ATG

			ABV64363.1 Bacillus pumilus SAFR-032 (72%) PCD:SSB_OBF		
51	30682-30909	75	hypothetical protein		AAATGAGG-9-ATG
52	30915-31307	130	hypothetical protein LM5578_p34 <i>Listeria monocytogenes</i> 08-5578 (40%)	1e-20	TAATGAGG-8-ATG
53	31304-32470	388	T4SS VirB11 ATPase Type II/IV secretion system protein gb:EAL09741.1 <i>Listeria monocytogenes</i> str. 4b H7858 (42%) PCD: P.loop NTPase superfamily, VirB11	7e-79	AAAGGAGG-7-GTG
54	32483-33289	268	hypothetical protein gb: EAL09742.1 <i>Listeria monocytogenes</i> str. 4b H7858 (43%)	2e-60	TTAGGAGG-8-TTG
55	33286-33990	234	hypothetical protein LmonF_05293 <i>Listeria monocytogenes</i> FSL F2-515 (36%)	2e-41	TATGAAGG-7-ATG
56	34006-34368	120	hypothetical protein gb: EEN76767.1 <i>Listeria grayi</i> DSM 20601 (41%)	2e-22	AAAGGAAG-8-ATG
57A	34400-34537	45	hypothetical membrane protein		AAAGGAGG-8-ATG
57B	34600-34737	45	hypothetical membrane protein		AAGTGAGG-8-ATG
57C	34800-34937	45	hypothetical membrane protein		AAGTGAAG-8-ATG
57D	35000-35146	48	hypothetical membrane protein		AAGTGAGG-8-ATG
58	35158-35376	72	hypothetical protein hypothetical protein		TTAGGAGG-7-TTG
59	35400-35576	58	hypothetical protein		TCAGGTGA-5-ATG
60	35598-35828	76	hypothetical protein LmonocyFSL_00035 <i>Listeria monocytogenes</i> FSL J1-208 (48%)	5e-10	ATGGAAGG-12-ATG
61	35844-37271	475	enterotoxin gb: EDX58891.1 <i>Bacillus cereus</i> W (46%)	9e-18	AAAGGAGC-7-ATG
62	37283-38317	344	hypothetical protein		ATGAAAGG-7-ATG
63	38367-40652	761	T4SS Coupling Protein, TraG-VirD4 ATPase predicted membrane protein gb: EET96811.1 <i>Enterococcus faecalis</i> T1 (46%) PCD:Non-specific hits: TraG_VirD4, TrwB, TrwB_AAD-bind, sxt_traD, Prk13850, Superfamilies: P-loop NTPase superfamily Multi-domains: TraG, VirD4, TraD_F-	0	ATTTGAGG-7-TTG

64	40655-4	927	type, TraG-Ti. predicted protein EET97051.1 <i>Enterococcus faecalis</i> T2 (39%)	1e-86	CAAGGAGT-6-ATG
----	---------	-----	--	-------	----------------

Table 3
Characteristics of Direct Repeats

Type	name	CDS/ (gene n°)	positions	Length (bp)	N° of repeats
I	IA	No	8,862-8,984	49	2.5
	IB	No	12,822-13,140	49/50	6.5
	IC	No	15,127-15,479	47	7.5
	ID	No	21,034-21,442	48	8.5
II		Yes (57A-D)	34,583 – 35,283	200	3.5
III	IIIA	Yes (61)	36,285 – 37,005	36	20
	IIIB	Yes (64)	42,246 – 42,870	48	13
	IIIC	Yes (64)	42,686 – 43,211	42	12.5
Type	name	Consensus sequence (5' - 3')			
I	IA(*)	gtggccgggaaa-gatggccgcagctgaagcg-agtaccagggcgagcac			
	IB(+)	gtggcc-gaaaaagatggccgggctgaagcaaagtaccagggcgagcac			
	IC(!)	gtggcc-gaaaaagatggccgcagctgaagc-aagtaccaggg--cagcac			
	ID(#)	gtggccggaaaaag-tggccgggatcgttgccaagtaccaggg--cagcac			
II	(\$)	Aaagtgaggtgataactatgttagatcagaacattgaccgattgttatgggtgatcgctgc			
		p57A M L D Q N I D R L L W V I A			
		tgtcgggaattgtaggggcaatgatccatacatcagatatggaaatgaccgaactgttaaa			
		A V G I V G A M I H T S D M E M T E L L			
III	IIIA(\$)	aacagtattcaacggcatgaaagctaaaatctaaggaacgtgtgaggaataatagacac			
		K T V F N G M K A K I *			
		gttccttttttgttggccg			
		gccagggcagcagctgaagcgaaagcagctgaagag			
III	IIIB(£)	A R A A A E A K A A E E			
		acggagcaccaggcaagccgggagccggagaaacaccactaggcggag			
		N G A P G K P G A G E T P L G G			
		IIIC(@) ctgaacgcagcaccaagaagcaatagtggcagcagcgacca			
III	IIIC(@)	L N A A P R S N S G S S A P			

Number of deviations from consensus sequence: (*), 0 in 123 bp; (+), 7 in 326 bp; (!), 13 in 352 bp, (#), 0 in 395 bp; (\$), 19 in 720 bp; (£), 6 in 701 bp; (£), 16 in 622 bp; (@), 0 in 525 bp.

Fig 1.

Fig. 2A

Fig. 2B