Bacillus subtilis as a Model for Bacterial Systems Biology

Colin R Harwood, *Newcastle University, Newcastle upon Tyne, UK*

Bacillus subtilis has been the subject of intense study for nearly six decades. Initially, the key drivers were: (i) the need of the food industry for a nonpathogenic model bacterium to study the characteristics of endospores, and (ii) the observation, in 1959, that *B. subtilis* strain 168 could be genetically manipulated by transformation. In the intervening period, *B. subtilis* 168 has become second only to *Escherichia coli* K-12 in terms of the detail with which aspects of its genetic, biochemistry and physiology is understood. For the foreseeable future, *B. subtilis* represents an eminently tractable model in which to integrate knowledge gained from the reductionist approach to biology towards an understanding of how this bacterium functions as a unitary system. This will require the application of various 'omics' (e.g. genomics, transcriptomics, proteomics, metabolomics), the increased application of high-throughput technologies and system modelling. The ultimate aim of an *in silico* model of *B. subtilis* is that it can accurately mimic or predict its behaviour in the environment.

Introduction

Bacillus subtilis is a member of the low $G + C\%$ Gram-positive genus *Bacillus* that has played a prominent role in the development of microbiology since the genus was first proposed by Ferdinand Cohn in 1872. Broadly speaking, the genus consists of aerobic endospore-forming rod-shaped bacteria with a saprophytic life style. With the principal exception of members of the *B. cereus* group (e.g. *B. anthracis, B. cereus* and *B. thuringiensis*), the genus comprises mainly nonpathogenic species that are active in soil and associated water sources. *B. subtilis* Cohn 1872 is the type species of the genus. The ability of *Bacillus* to form endospores was first recognized by Louis Pasteur in 1870 during his studies on the diseases of silk worms. However, the significance of spores in long-term survival and pathogenesis was only recognized in 1877, respectively by Cohn (*B. subtilis*) and Robert Koch (*B. anthracis*). From the 1920s to approximately 10 years ago, members of the genus were primarily defined according to their shape, ability to form spores and a requirement of oxygen for growth. However, this resulted in an unwieldy and very heterologous genus in which the $G + C\%$ composition varied from 33 to 67%. More recently, 16S ribosomal ribonucleic acid (rRNA) sequencing studies have allowed members of the genus to be reclassified into a number of more clearly defined genera. In addition to the remnants of the existing genus, these are: *Alicyclobacillus*, *Aneurinibacillus*, *Paenibacillus*, *Brevibacillus*, *Virgibacillus*, *Gracilibacillus*, *Geobacillus*, *Ureibacillus* and *Marinibacillus*.

B. subtilis and its close relatives are important sources of industrial enzymes, fine biochemicals, antibiotics and insecticides. *B. subtilis*, along with strains of *B. amyloliquefaciens*, *B. licheniformis* and *B.* (now *Geobacillus*) *stearothermophilus* account for more than half of the world's production of industrial enzymes. Two *Bacillus* enzymes dominate the market; alkaline protease (subtilisin) and α-amylase. Related enzymes from different sources often display different physical and enzymatic properties and this variation is increasingly exploited by the enzyme producers to tailor their products to specific industrial process. These commercially important species grow well in industrial fermenters and have a well-proven safety record. The high secretion capacity of well-developed commercial strains makes them prime candidates for the production of heterologous proteins such as therapeutic proteins and vaccines. In the laboratory, *B. subtilis* is able to grow in simple salts' media containing glucose or other simple sugars as sources of carbon, and ammonium salts or certain amino acids as sources of nitrogen. *B. subtilis* strain 168, the model organism on which most studies are performed, is a tryptophan auxotroph and this amino acid is required even in media with acid-hydrolysed casein as the main source of nitrogen. A widely used minimal medium for *B. subtilis* is the chemically defined Spizizen's minimal medium (Harwood and Cutting, 1999).

Although *B. subtilis* forms endospores in response to nutrient deprivation (e.g. carbon, nitrogen or phosphate starvation), special media and growth protocols are required for efficient sporulation. Schaeffer's sporulation medium is widely used for efficient spore production in the laboratory (Harwood and Cutting, 1999). The production of a pink pigment during sporulation provides a useful

visual indication of the formation of spores both in liquid and on solid media.

B. subtilis survives well on agar plates, either at room temperature or at 4°C, although it is recommended to sub-culture on a weekly basis. For long-term preservation, *B. subtilis* is stored as glycerol or lyophilized cultures. Strains that sporulate well (i.e. not sporulation mutants) may also be preserved as spore suspensions in sterile water at 4°C. Provided they have been extensively washed to remove nutrients which may otherwise cause them to germinate, spores are generally stable for many years. *B. subtilis* can be transported on the surface of freshly inoculated nutrient agar slopes or as a spore suspension spotted on sterilized filter paper discs and encased in aluminium foil.

Cultures of *B. subtilis* 168 are available from the *Bacillus* Genetic Stock Center (http://www.bgsc.org/) which has an extensive collection of mutant *B. subtilis* strains, bacterio-phages and plasmids. The collection also includes strains of *B. cereus, B. licheniformis, B. megaterium, B. pumilus, B. (now Geobacillus) stearothermophilus* and *B. thuringiensis.*

Genomics

The genome of *B. subtilis* strain 168 was sequenced by a joint European/Japanese consortium and the resulting sequence, completed in June 1997, was the first for both a Gram-positive bacterium and a differentiating bacterium (Kunst *et al.*, 1997). The genome is estimated to be 4 214 630 bp in length, comprising 4106 protein encoding genes, 86 transfer RNA (tRNA) genes, 30 rRNA genes and 3 stable RNA genes. These are organized into approximately 1500 operons that are controlled by some 200 regulatory proteins (Ishii *et al.*, 2001). Analysis of the *B. subtilis* chromosome has revealed high-level organization and relationships (Rocha *et al.*, 1999, 2001). Highly expressed genes and genes expressed during growth tend to be oriented in the direction of the replication fork, presumably to reduce conflicts between replication and transcription (Kunst *et al.*, 1997). The SubtiList database (http://genolist.pasteur.fr/SubtiList/) is dedicated to the genome of *B. subtilis* strain 168 and pro-vides a curated dataset of deoxyribonucleic acid (DNA) and protein sequences, relevant annotations and functional assignments.

Slightly less than 90% of the genome encodes proteins or stable RNA. Most of the remaining DNA appears to be structural or to be involved in gene regulation. At least 5% of the chromosome has been acquired by horizontal gene trans-fer (Rocha *et al.*, 2001). They include prophage SP*β* and defective prophages PBSX and the *skin* element, identified prior to the sequence determination, and seven additional prophage-like elements (Kunst *et al.*, 1997). Just over half of the genome is required for cell processes, intermediary metabolism and for macromolecular synthesis while a signi-ficant proportion of the remaining genome is required for growth and survival in the environment. *B. subtilis* can utilize a wide range of substrates and analysis of the genome reveals the presence of a large number of transporter proteins for the uptake or efflux of particular substrates (Saier *et al.*, 2002). These include six voltage-gated ion channel proteins, 185 secondary active transporters and more than 70 ABC trans-porters (or components thereof). *B. subtilis* senses the environment via 34 two-component signal transducers, one of which (YycF/G) is essential for growth.

Under conditions of nutrient deprivation or stress, *B. subtilis* initiates a series of responses designed to main-tain or restore growth. These include the secretion of enzymes for the utilization of macromolecules such as pro-teins and carbohydrates, chemotaxis and motility, transport systems, the induction of specific and general stress response proteins and genetic competence for the uptake of DNA from the environment. For the most part, these responses are transitional and, if they fail to restore growth, *B. subtilis* induces a complex differentiation pathway that leads to the production of the highly resistant endospore (Piggot and Hilbert, 2004). About 5% of the genome is dedicated to genes required for sporulation and germination, and the analysis of this primitive differentiation pathway has led to an understanding of many types of signalling (e.g. phosphorelay) and regulatory (e.g. sigma factors/antisigma/ anti-antisigma factors) pathways (Iber *et al.*, 2006).

The response of *B. subtilis* to stress is often mediated by alternative sigma factors that direct RNA polymerase to specific subsets of genes (Helmann and Moran, 2002). The principal sigma factor, Sigma-A (σ^A), exhibits a similar consensus sequence to that of *E. coli* σ^{70}. Of the remaining 16 sigma factors, four are required for sporulation (σ^E, σ^F, σ^G, σ^K), one for transition phase activities including sporulation initiation and competence (σ^H), one for the general stress response (σ^B), one for levanase and amino acid catabolism (σ^L), one for chemotaxis and motility (σ^D), one for salt resistance (σ^M), while the remaining seven sigma factors are members of the extracellular function (ECF) group (Asai *et al.*, 2003).

Competence and Recombination

The ability to introduce DNA into *B. subtilis* is central to this bacterium's extreme genetic amenability. The most widely used method for introducing DNA into *B. subtilis* is natural transformation, although protoplasts of *B. subtilis* and several other *Bacillus* species can be induced to take up DNA and regenerate their cell walls. Electrotransforma-tion is also possible, but usually results in low efficiencies. DNA can also be introduced in *B. subtilis* by transducing phages, although nowadays transduction is limited to very specific purposes.

The development of competence to take up DNA was first described for *B. subtilis* strain 168 in 1958. Competence develops naturally during transition from exponential to stationary phase. The size of DNA fragments that can be

taken up is about 20–30 kb (Dubnau, 1993). At saturating levels of DNA ($>1\,\mu g\,mL^{-1}$ of competent cells), transformation frequencies of up to 5% of the cells in a population can be achieved with homologous chromosomal DNA. Under these conditions, the cotransfer of unlinked genetic markers, a phenomenon referred to as congression, makes it possible to introduce nonselectable genes into the chromosome since approximately 1% of recombinants will also contain the required nonselected gene.

Transformation of plasmid DNA is possible, although the frequency is low compared to homologous chromosomal DNA; 0.001–0.01% for intact plasmids and one or two orders of magnitude lower for ligation mixtures. This is because transforming DNA is converted into a single-stranded form and then randomly fragmented prior to uptake by the cell. The plasmid is effectively reconstituted in the cell by homologous recombination and, consequently, only multimeric plasmid DNA (present in most plasmid preparations) or monomers containing internal repeats are efficiently transformed.

Plasmid-based Host/Vector Systems

The absence of native antibiotic resistance plasmids in early studies meant that *B. subtilis* vectors were originally developed using plasmids from other Gram-positive bacteria, such as *Staphylococcus aureus* and *Lactococcus lactis* (Janniere *et al*., 1993). The replication functions and/or antibiotic resistance genes from several of these plasmids are still widely used (e.g. pUB110, pC194, pE194 and pWVO1). More recently, vectors based on endogenous *Bacillus* plasmids have been developed (Meijer *et al*., 1998).

Plasmids derived from Gram-positive bacteria often replicate via a rolling-circle mode of replication. A characteristic of rolling-circle replication is the formation of single-stranded DNA intermediates that are subsequently converted into double-stranded DNA via a secondary single-strand origin (SSO) of replication. This mode of replication tends to lead to increased segregational and structural instability. The copy number of rolling-circle plasmids in *B. subtilis* can vary from about 5 to 200 per chromosome. A number of theta replicating plasmids have been developed as vectors for *B. subtilis,* including the native *B. subtilis* plasmid pLS20 (Meijer *et al*., 1995) and the enterococcal plasmid pAM81 (Janniere *et al*., 1993). Despite their increased stability over rolling-circle plasmids, they are not as widely used as cloning vectors. In *B. subtilis* plasmid vectors are now mainly used for overexpressing genes; for most other purposes (e.g. gene knockouts, complementation analysis, controlled gene expression), integrative plasmids are preferred because of their low copy number and increased stability.

B. subtilis integrative plasmids are able to replicate in *E. coli*, are selectable in both *E. coli* and *B. subtilis* but are not able to replicate in the latter. Instead they use homologous recombination between sequences as short as 150 bp to mediate efficient integration into the *B. subtilis* chromosome, either via a single (integrating the entire plasmid) or double (integrating a region between regions of homology) crossover recombination (**Figure 1**).
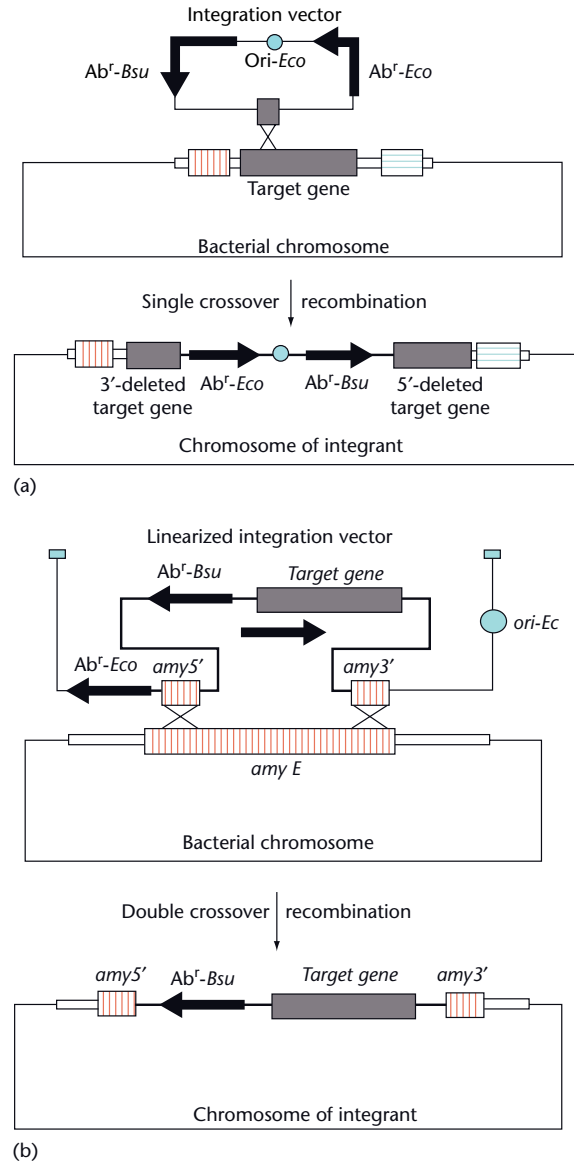

(a)

(b)

Figure 1 The use of integration vectors as genome management tools. Non-replicating integration vectors can be used to: (a) inactivate specific chromosomal genes via a single crossover recombination between regions of homology on the vector and chromosome; (b) deliver target genes to non-essential regions of the chromosome (here the *amyE* gene encoding α-amylase) via a double crossover recombination of the linearized vector. *Abbreviations*: Ab^r-*Bsu*, antibiotic resistance gene active in *B. subtilis*; Ab^r-*Eco*, antibiotic resistance gene active in *E. coli*; *amyE5'/amyE3'*, fragments of the *B. subtilis amyE* gene; Ori-*Eco,* origin of replication active in *E. coli*.

The pMUTIN series of integration/reporter gene vectors (Vagner *et al.*, 1998) were used to generate the *Bacillus* Functional Analysis (BFA) isogenic mutant collection in which virtually all genes of unknown function ('*y* genes') and the majority of genes of known function were inactivated. pMUTIN4 (~ 8.6 kbp) has the following properties (**Table 1**): (1) a ColE1 replication origin that is functional in *E. coli* but not *B. subtilis*; (2) an ampicillin resistance (Ap^R) gene functional in *E. coli* and an erythromycin/lincosamide resistance (Erm^R) gene functional in *B. subtilis*; (3) a *lacZ* reporter gene preceded by a *B. subtilis spoVG* ribosome binding site (RBS) (Perkins and Youngman, 1986) and stop codons in all three reading frames to avoid translational fusions with upstream genes; (4) an isopropyl β-D-l-thiogalactopyranoside (IPTG)-inducible P$_{spac}$ promoter with two strong upstream terminators (t_1t_2 from *rrnB*); and (5) a constitutively expressed *lacI* gene encoding the Lac repressor.

When DNA fragments (150–400 bp) containing an internal region of the target gene are cloned into pMUTIN4, a single crossover recombination (**Figure 1**) between the homologous regions on the vector and the host chromosome generates a 'knockout' mutant. The integration places the *lacZ* gene within the transcriptional unit of a 3'-truncated copy of the target gene, enabling it to monitor the activity of the target gene's promoter via the synthesis of β-galactosidase. The transcription of genes downstream and in the same operon (including a 5'-truncated copy of the target gene) can, if required, be induced via the vector-encoded P$_{spac}$ promoter. This reduces potential polar effects of the insertion, and allows pMUTIN4 to be used even when downstream genes are essential for growth. Because IPTG is a gratuitous inducer, the level of expression of P$_{spac}$ can be modulated by varying the IPTG concentration. If the target gene itself is essential, as determined by an inability to isolate knockout mutants irrespective of the inclusion of IPTG in the selection medium, its expression can be made conditional by generating an 'RBS-fusion' rather than knockout mutant. In this case the RBS and the 5'-end of the target gene are cloned into pMUTIN4, and integration results an intact copy of the target gene downstream of the P$_{spac}$ promoter, rendering the strain IPTG-dependent. More recently, pMUTIN10 (~ 3.5 kbp) has been constructed to be compatible with pMUTIN4 (**Table 1**). pMUTIN10 encodes spectinomycin (Spc^R) resistance for selection in both *E. coli* and *B. subtilis* and the P$_{spac}$ promoter is replaced with the xylose-inducible P$_{xyl}$ promoter. pMUTIN10 is considerably shorter than pMUTIN4 (8.6 kbp) by virtue of it not encoding the *lacI* and *lacZ* genes.

In some instances, for example when the target gene is too small to use a single crossover recombination, it is more appropriate to generate knockout mutants via a double-crossover recombination (**Figure 1b**).

The BFA mutant collection comprises about 2880 strains. BFA mutants were not generated in genes encoding components of the translation apparatus (e.g. rRNA or ribosomal protein genes) or genes encoded by the major prophage genomes (e.g. *skin* element, SPβ or PBSX). Additionally, it was not possible to construct mutants for seven genes of unknown function. Two hundred and sixty-seven genes were designated as essential on the basis of their IPTG dependency or reports from the literature, including 82 genes currently classified as being of unknown function (Kobayashi *et al.*, 2003). The data from the BFA programme are publicly available in dedicated databases: Micado (http://locus.jouy.inra.fr/cgi-bin/genmic/madbase/progs/madbase.operl/) and BSORF (http://bacillus.genome.jp/).

Complementation Analyses

Complementation analyses are best carried out using integration vectors that avoid copy-number effects. An effective approach is to use a combination of compatible expression vectors, one of which integrates at the site of the target gene (e.g. pMUTIN4 or pMUTIN10) and the other (e.g. pJPR1 or pAX01, see **Table 1**) at a region of the chromosome known to have little or no transcriptional activity of its own (e.g. *lacA* or *amyE* loci). For example, we have used this system to study the functional activities of three *B. anthracis* genes that show homology to *B. subtilis prsA,* the product of which is essential for growth (Williams *et al.*, 2003). Placing the *B. subtilis prsA* gene under the control of P$_{spac}$ promoter as a pMUTIN4 RBS-fusion mutant (see above) rendered the strain IPTG-dependent. When individual copies of the *B. anthracis prsA* homologues (*prsAA, prsAB* and *prsAC*) were introduced onto the chromosome at the *amyE* locus under the control of the xylose-inducible P$_{xyl}$ promoter, their ability to complement *B. subtilis prsA* were determined by their ability to grow in the absence of IPTG, but in a xylose-dependent manner (**Figure 2**).

Construction of Markerless Gene Deletions

The construction of deletion mutants using the double crossover recombination system (see above) leads to the replacement of all or part of the target gene with an antibiotic resistance gene. This can be a disadvantage when multiple deletions are required or for strains that are to be used in commercial production systems in which the presence of antibiotic resistance genes is discouraged. Systems have therefore been devised to facilitate so-called 'clean' gene deletions that produce strains which are not labelled with antibiotic resistance genes.

One such system is based on derivatives of the broad-range lactococcal plasmid pWV01, which lacks the *repA* gene required for the initiation of replication (Leenhouts *et al.*, 1996). Consequently, pWV01-based knockout vectors need to be constructed in helper strains that provide RepA *in trans*. Regions flanking the target gene are cloned into the

Table 1 Knockout, reporter, expression and affinity tag vectors used for *B. subtilis*

| Vector | Function | | | | Integration site | SCO/DCO | Promoter | Reporter | Tag | Selection agent | |
	Kout	Rpt	Exp	Tag						*Eco*	*Bsu*
pMUTIN1-4	√	√	√		Target gene	SCO	P_{spac}	*lacZ*		Amp^R	Erm^R
pMUTIN10	√		√		Target gene	SCO	P_{xyl}			Spec^R	Spec^R
pMUTIN-FLAG				√	Target gene	SCO	P_{spac}		FLAG	Amp^R	Erm^R
pMUTIN-HA				√	Target gene	SCO	P_{spac}		HA	Amp^R	Erm^R
pMUTIN-cMyc				√	Target gene	SCO	P_{spac}		cMyc	Amp^R	Erm^R
pMUTIN-GFP+		√			Target gene	SCO	P_{spac}	*gfp*		Amp^R	Erm^R
pMUTIN-YFP		√			Target gene	SCO	P_{spac}	*yfp*		Amp^R	Erm^R
pMUTIN-CFP		√			Target gene	SCO	P_{spac}	*cfp*		Amp^R	Erm^R
pJPR1			√		*amyE*	DCO	P_{xyl}			Amp^R	Cm^R
pAX01			√		*lacA*	DCO	P_{xyl}			Amp^R	Erm^R
pX			√		*amyE*	DCO	P_{xyl}			Amp^R	Cm^R
pA-spac			√		*lacA*	DCO	P_{spac}			Amp^R	Erm^R
pBgaB		√			*amyE*	DCO		*bgaB*		Amp^R	Neo^R
pORI240	√				Target gene	SCO				Tc^R	Tc^R
											lacZ
pTOPO-bac.DifCAT	√				Target gene	DCO				Kan^R	Cm^R
										Amp^R	

Kout, gene knockout; Rpt, reporter; Exp, expression; Tag, affinity tag (FLAG = DYKDDDDK; HA = YPYDVPDYA; cMyc = CEQKLISEEDL); SCO, single crossover recombination; DCO, double crossover recombination; Amp^R, ampicillin resistance; Cm^R, chloramphenicol resistance; Erm^R, erythromycin resistance; Kan^R, kanamycin resistance; Neo^R, neomycin resistance; Spec^R, spectinomycin resistance; Tc^R, tetracycline resistance.
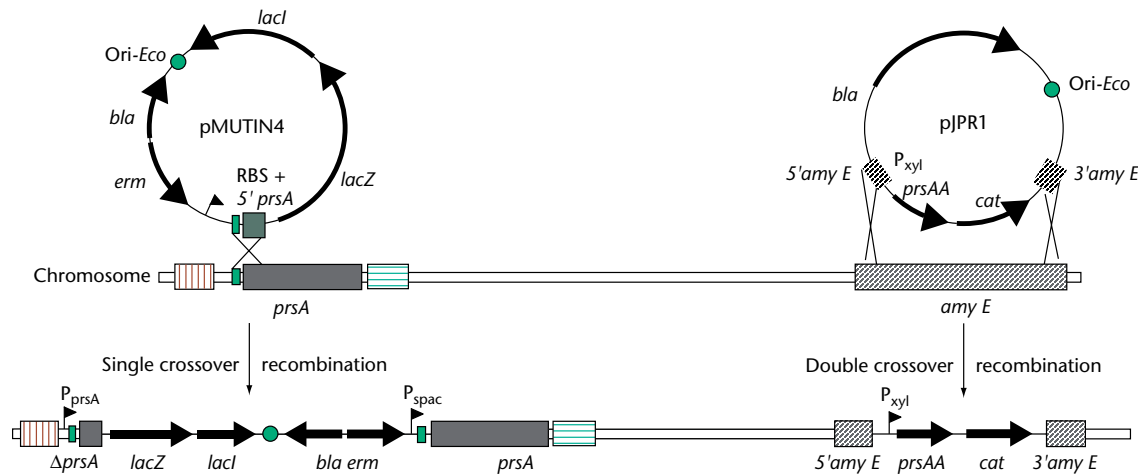
Figure 2 Complementation analysis in *B. subtilis* of the PrsA homologues of *B. subtilis* and *B. anthracis* using compatible chromosomal integrational expression vectors. *Abbreviations: amyE5′/amyE3′*, fragments of the *B. subtilis amyE* gene; *bla*, ampicillin resistance gene active in *E. coli*; *cat*, chloramphenicol resistance gene active in *B. subtilis*; *erm*, erythromycin resistance gene active in *B. subtilis*; *lacI*, encoding the lactose repressor of *E. coli*; *lacZ*, encoding β-galactosidase of *E. coli*; Ori-*Eco*, origin of replication active in *E. coli*; *prsA*, *B. subtilis* gene encoding PrsA; *prsAA*, *B. anthracis* gene encoding PrsAA; P_{prsA}, native promoter of the *B. subtilis prsA* gene; P_{xyl}, xylose-inducible promoter; RBS+5′*prsA*, fragment carrying the ribosome binding site and 5′ sequences of the *B. subtilis prsA* gene.

vector pORI240 (**Figure 3**) that encodes tetracycline resistance and *lacZ* reporter genes, both of which are active in *B. subtilis*. After transformation into *B. subtilis*, integrants resulting from a single crossover recombination between one of the flanking regions on the vector and the homologous region on the chromosome are selected in the presence of tetracycline. The resulting integrants contain two copies of the flanking regions. A subsequent excision event involving the same flanking region as the original integration event restores the wild-type genotype. However, an excision event involving the other flanking regions results in the deletion of all of the intervening sequences. Excision events are relatively rare and, in the absence of a positive selection marker, cells in which excision have taken place are monitored on plates containing X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) since they produce white colonies against a background of predominantly blue integrant colonies. In theory, $\sim 50\%$ of the white colonies should be the required deletion mutants and these are identified by a diagnostic polymerase chain reaction (PRC) across the flanking region.

Recently, the *ysbC* gene of *Lactococcus lactis* has been shown to encode a membrane-associated transporter for the uptake of orotate, a precursor of pyrimidine biosynthesis. Since *B. subtilis* lacks this transporter, it is not able to transport orotate or its toxic analogue, fluoro-orotate. Transformants of *B. subtilis* encoding *ysbC* are not able to grow on minimal agar plates supplemented with fluoro-orotate, indicating that the absence or removal of this gene can be used as an efficient counterselectable marker for this bacterium.

An alternative efficient method for isolating markerless deletions has been developed using a cassette containing a chloramphenicol resistance gene flanked by *B. subtilis dif* sites (Bloor and Cranenburgh, 2006). Natural *dif* sites are located in the terminus region of the chromosome and are essential for resolving chromosome dimers prior to cell division. This involves an intermolecular recombination across tandemly repeated *dif* sites by Xer-like recombinases (XerC/XerD in *E. coli* and RipX/CodV in *B. subtilis*). Plasmid pTOPO-bac.DifCAT contains a cassette comprising a chloramphenicol gene (*cat*) flanked by 28-bp *B. subtilis dif* sites. After amplification, this cassette is cloned into an *E. coli* cloning vector (e.g. pCR21.-TOPO, Invitrogen) between ~ 0.4-kbp sequences flanking the region on the chromosome to be deleted. After transformation into *B. subtilis*, integrants resulting from a double crossover recombination are selected on chloramphenicol plates. Propagation of the integrants in nonselective liquid medium results in an excision frequency across the *dif* sites of about 4% per generation. As a result of the excision, the target gene is replaced with a single copy of the 28-bp *dif* site.

Markerless deletion technology was used in a *B. subtilis* genome minimalization project in which two prophages (SPβ, PBSX), three prophage-like regions and the large (~ 76 kbp) polyketide synthesis (*pks*) operon were deleted, resulting in a reduction of the genome by $\sim 8\%$ and the elimination of 332 genes (Westers *et al.*, 2003).

Transcriptional Studies and Transcriptomics

Transcriptomics aims to catalogue the mRNA expression profile of cells in responses to the prevailing growth conditions, to transitions from one growth condition to another and in response to intrinsic and extrinsic stresses.

Transcriptomics uses a variety of tools such as reporter gene technology, Northern blotting, primer extension mapping, quantitative real-time PCR and DNA arrays, and all have been applied to the analysis of *B. subtilis*. Transcriptome analysis is complex since the RNA content (messenger RNA (mRNA), tRNA, rRNA and various other species) of a cell is dependent on the physical and chemical environment, its stage in the growth, cell or differentiation cycles and the patterns of transcription and RNA processing. Analysis of the transcriptome can be used to determine: transcription start and end points; mRNA expression profiles in response to environmental changes and challenges; the activity of regulatory proteins and the identity of their cognate target genes; high-level regulatory networks.

Reporter genes

E. coli lacZ has been used extensively for transcriptional and translational reporter genes in *B. subtilis* (see **Table 1**).

When used as a transcriptional reporter, the native *E. coli*-derived RBS is replaced with the one derived from *B. subtilis*. As with *E. coli,* the *lacZ* reporter can be used to detect gene expression on solid media using the chromogenic substrate X-gal and by microscopy using the fluorogenic substrate fluorescein-di-*β*-D-galactopyranoside (Lewis *et al.*, 1994). In liquid cultures the assay of Miller is most commonly used to assay *β*-galactosidase activity (Harwood and Cutting, 1990). The *lacZ* reporter is not suitable for studying heat-shock gene expression, since *E. coli β*-galactosidase is degraded rapidly under these conditions. A widely used heat-stable reporter is the *β*-galactosidase-encoding *bgaB* gene of *B.* (now *Geobacillus) stearothermophilus*. The *xylE* gene from *Pseudomonas putida*, specifying catechol-2, 3-dioxygenase, is a useful reporter gene in *Bacillus* sp. for analysing expression from strong promoters. Expression is measured in extracts spectrophotometrically or in colonies after spraying with the substrate catechol.

Reporter systems based on light-producing enzymes or fluorescent proteins have been extensively used in
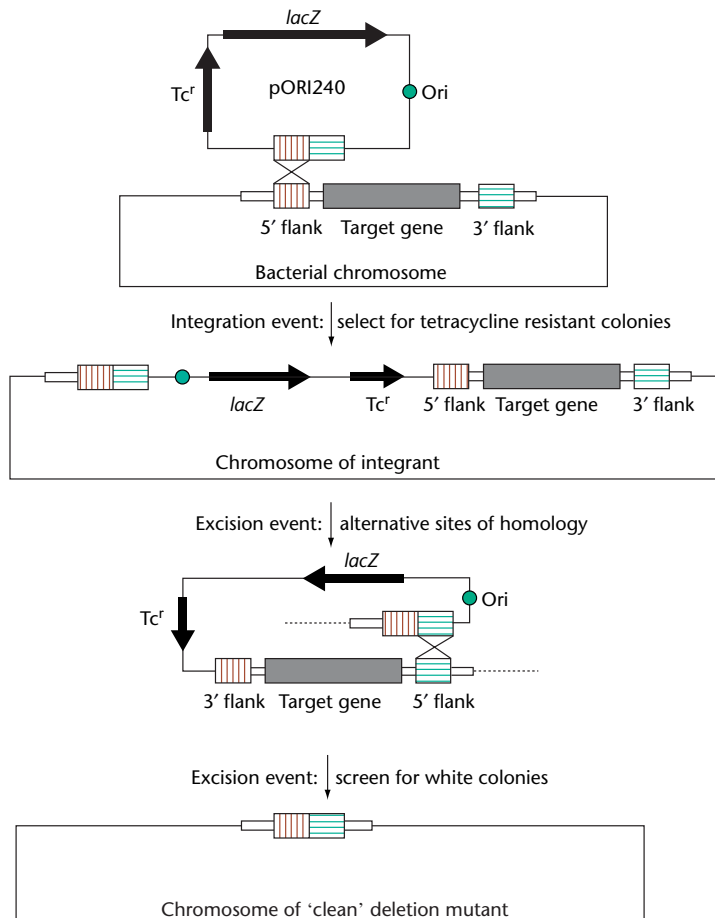

Figure 3   The generation of 'clean' chromosome deletions using the pVWO1-based vector pORI240. *Abbreviations: lacZ*, encoding *β*-galactosidase from *E. coli*; Tc[r], tetracycline resistance gene.

B. subtilis. An advantage of the latter is that the uptake of substrates by the host cell is not required for activity. The *luxAB* genes of *Vibrio harvei* encode luciferase which emits light when exposed to a suitable substrate (e.g. decanal). A chromosomally-located *luxAB* gene fusion has been used as a reporter for tracking *B. subtilis* in soil. However, the most versatile reporter is green fluorescent protein (GFP), specified by a gene that was isolated from the jellyfish *Aequorea victoria*. GFP has been used extensively in *B. subtilis* for the site- and compartment-specific location of proteins (Carballido-Lopez and Errington, 2003; Meile *et al.*, 2006). The wild-type protein is excited at 395 nm and emits green light at 590 nm. However, mutagenesis of *gfp* has generated cyan (*cfp*) and yellow (*yfp*) variants that facilitate multichannel labelling experiments.

Regulatory proteins

A wide variety of transcription regulatory proteins have been identified in *B. subtilis* on the basis of multiple protein alignments and hidden Markov modelling. Data on *B. subtilis* promoters and transcription factors are available at the DBTBS website (http://dbtbs.hgc.jp/) and the classes of regulatory proteins are summarized in **Table 2**.

Transcriptome analysis

The completion of the *B. subtilis* genome sequence has provided the opportunity to develop whole-genome DNA arrays to analyse the expression profiles. Extensive expression studies have been published using oligonucleotide-based microarrays (Affymetrix GeneChip technology) and ORFmer-based arrays. In addition, at-line measurements of specific mRNA species that represent event markers for monitoring commercial bioprocesses have been determined using an electronic DNA chip (Jürgen *et al.*, 2005). Null mutations have been widely used to identify the regulators and members of specific regulons and such analyses have covered a wide range of aspects of central metabolism (e.g. riboflavin, biotin and thiamine synthesis), stress responses (e.g. sigB-dependent general stress pathway), catabolite repression and sporulation. DNA arrays have also been used to survey rates of mRNA decay on a genomic scale in early stationary-phase cultures of *B. subtilis*.

Proteomics

Proteome analysis facilitates the grouping of genes into regulatory classes, and provides information of post-translational modifications and protein stability. A theoretical two-dimensional protein map has been constructed from the calculated pI and molecular mass values of each of the polypeptides encoded by the genome of *B. subtilis* strain 168 (Bernhardt *et al.*, 2001). A plot of the pI-frequency distribution shows a bimodal distribution, with 65% of the

Table 2 Classes of transcription regulators in *Bacillus subtilis*

Classes of transcription factor	Number
Sigma factors	
Sigma-70	9
Sigma-54	1
Sigma-70 ECF	7
Helix-turn-helix	
MarR	23
LacI	11
GntR	21
LysR (HTH-1)	20
ArsR (HTH_5)	7
DeoR	7
AraC (HTH_2)	12
GerE	10
Crp	2
Xre (HTH_3)	17
MerR	10
TetR	19
AsnC	7
LexA	1
HTH_6	2
Other families	
C family regulators	13
Response regulators	35
Fur	3
Sigma-54-related factors	5
Bgl-antiterminators	8
Cold shock-dependent (CSD)	3
IclR	1
GreA/B	1
Fe-dependent repressors	1
HrcA	1
Arg repressor	1
Not assigned	35

Source: Data extracted from Ishii *et al.* (2001)

proteins apparently having an acid pI: YvcF is the most acidic (calculated pI value of 2.79) and RpmH the most basic (pI 13.0). A high proportion of secreted proteins have a pI close to 7.0, which may facilitate their passage through the acidic cell wall (Jensen *et al.*, 2000). The largest protein, the polyketide synthetase PksN, has a molecular mass of 609 058 Da.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) has been used to catalogue the polypeptides synthesized by *B. subtilis* under various growth and stress regimes and the relationship between cell physiology and the proteome of *B. subtilis* has been extensively reviewed by Hecker and his colleagues (Hecker and Völker, 2004). 2D-PAGE has been of particular value for the identification of components

of stimulons (proteins encoded by genes induced by a specific physical or chemical stress) or regulons (proteins encoded by gene induced by a specific regulatory protein). In recent years, 2D-PAGE protocols have been modified and adapted to detect secreted proteins and proteomic signature associated with stress and starvation responses.

Proteins may be radiolabelled (^{35}S-methionine) or stained with dyes such as Coomassie Brilliant Blue or Sypro Ruby. Digitized images can be combined as dual-colour images that provide additional information. For example, images visualized by a staining technique represent accumulated proteins, while protein pulse-labelled with ^{35}S-methionine show proteins synthesized during the pulse period. If the resulting images are respectively false-coloured green and red, when the two images are combined, proteins that are accumulated and synthesized in growing cells are coloured yellow. However, red- or green-coloured proteins identify proteins that are either induced or repressed, respectively, in response to a specific stimulus or change in growth (Hecker and Völker, 2004).

The resulting data has been stored as coordinates that allow individual polypeptides from independent gels to be compared and expression profiles to be constructed. The data, together with clickable representative gels, are available at the Sub2D website at: http://microbio2.biologie.uni-greifswald.de:8880/sub2d.htm. Sub2D includes data on house-keeping proteins, secreted proteins, proteins induced following phosphate starvation, heat shock, oxidative shock, ethanol stress as well as the effects of null mutations in genes encoding specific regulatory proteins. Sub2D is crosslinked to other *B. subtilis* databases, including SubtiList, BSORF, Micado, as well as SwissProt and MEDLINE.

Currently, 745 proteins have been identified by two-dimensional gel electrophoresis together with an additional 473 proteins identified by a semi-gel-based analysis of membrane proteins. This means that almost half of the ~ 2500 genes expected to be expressed in growing *B. subtilis* cells have been demonstrated at the protein level. However, in order to provide a more complete understanding of cellular physiology and adaptation to environmental changes or stress, proteins showing an alteration in expression level are of primary interest. To this end, gel-free multidimensional liquid chromatography-based techniques, linked to high-resolution mass spectrometry (e.g. isobaric tagging for relative and absolute quantitation (iTRAQ), Fourier transform (FT-MS)), are being developed (Wolff *et al.*, 2006). Preliminary analyses indicate that additional datasets are obtained by the gel-free approach.

Metabolomics and Fluxomics

B. subtilis has been used as a model to study metabolic networks and flux. These studies have mainly focused on a fluxome analysis of the central metabolism of glucose-grown cells (Fischer and Sauer, 2005). ^{13}C tracer experi-ments have emerged as the key technology for these studies, facilitating the identification of network topology and the quantitation of molecular fluxes. The main advantage of flux analysis is that it allows the operation of metabolic networks to be visualized in a dynamic manner via quantification of reaction velocities. These studies have shown that the Emden–Meyerhof–Parnas (EMP) pathway is the main pathway for glucose catabolism in *B. subtilis*. ^{13}C tracer experiments have been used to determine, in relation to lethality, flexibility and optimality, the influence of knockout mutants in the major functional gene categories on the metabolic state of *B. subtilis* at the level of reaction rates and the distribution of fluxes through alternative pathways (Fischer and Sauer, 2005). These have revealed that *B. subtilis* exhibits a suboptimal performance metabolism on its preferred substrate (i.e. glucose) because it invests significantly in a standby mode in anticipation of changes in the prevailing environmental conditions. This is likely to be a typical strategy for soil-living organisms continually subjected to fast/feast conditions.

Integrative Systems Biology

In recent years, advances in cell-imaging technology has begun to reveal information about the cell cycle, chromosome segregation and the precise topology of the macromolecular components of *B. subtilis* in both vegetative cells and endospores (Dworkin and Losick, 2005). These advances have included the identification of helically arranged actin-like proteins, such as MreB and homologues Mbl and MreBH, that provide *B. subtilis* with a shape-determining cytoskeleton that influence the positioning of the chromosome replication and cell wall synthesis machineries (Leaver and Errington, 2005). These studies point to ways in which the extensive knowledges of the biochemistry, regulatory network analysis and physiology of *B. subtilis* can be integrated with topological events and structures associated with the cell cycle (e.g. cell wall growth, cell division and chromosome segregation). Developing bioinformatical and modelling tools provide the prospect of constructing testable models that are able to predict the behaviour of complex subcellular components. Such subcomponent models will, hopefully, prove to be precursors to whole-cell systems in which the behaviour of an entire organism can be predicted *in silico* and tested *in vivo* in an iterative cyclical process.

The *Bacillus* Cell Factory

Bacillus species are an important source of industrial enzymes, fine biochemicals and antibiotics and the ease with which they can be grown and their well-proven safety, has also made them prime candidates for the production of heterologous proteins (**Figure 4**). Fermentation from
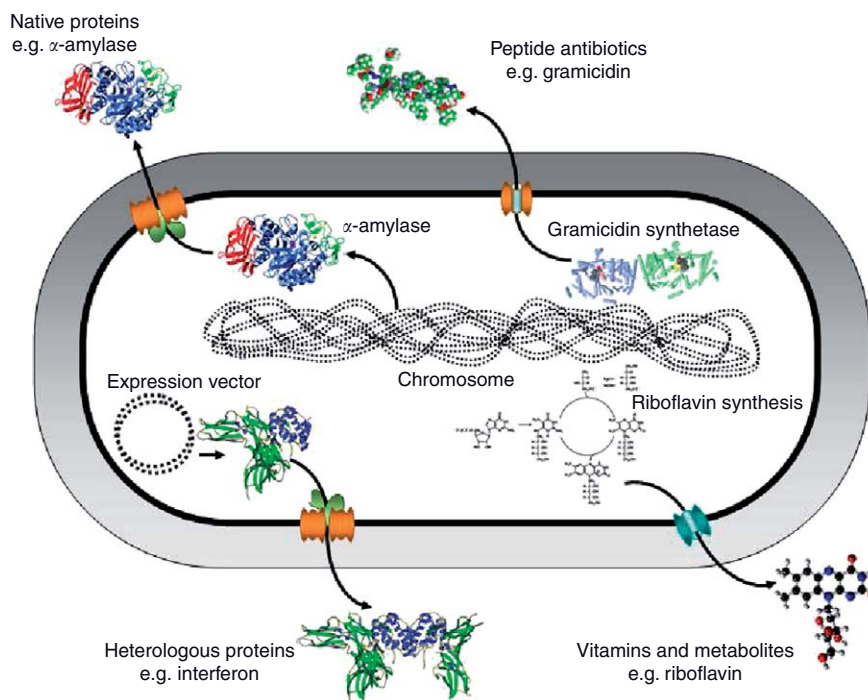
Figure 4 Showing the main commercial products synthesized in *B. subtilis* and its close relatives. *Bacillus* species are major producers of industrial enzymes (e.g. α-amylases, serine proteases), vitamins and food supplements (e.g. riboflavin, biotin, nucleosides), peptide antibiotics (e.g. gramicidin, polymyxin, tyrocidine) and heterologous proteins.

Bacillus accounts for about one-half of the world's production of industrial enzymes, the main classes of which are serine proteases (e.g. Subtilisins), α-amylases, β-glucanases and penicillin acylases. Despite the high-level secretion of certain native enzymes ($>10 \, \mathrm{gL^{-1}}$), attempts to use *B. subtilis* for the production of heterologous proteins have met with limited success. While extracellular proteins from close relatives can be produced at high concentrations, the yield of proteins from unrelated species, including eukaryotes, remains relatively low. This is, at least in part, due to the production of at least seven extracellular proteases and to incompatibilities with the *Bacillus* protein secretion pathway which is being actively engineered to improve productivity (Sarvas *et al.*, 2004).

Bacillus species are used for the production of a number of primary metabolites for the food and healthcare industries. These include the production of the nucleotides for use as flavour enhancers and vitamins such as biotin, folic acid and riboflavin. These pathways have, and continue to be, the focus of metabolic engineers who have attempted to remove the metabolic bottleneck from the relevant pathways.

Bacillus species also produce a variety of peptide antibiotics to improve their survival prospects in their natural environment. These are short (<20 residues) peptides which are synthesized within multienzyme complexes (peptide synthetases) that range in size from 100 to 600 kDa. A considerable amount of information is available about the activities of peptide synthetases and they are being engineered with a view to using them to produce *in silico*-designed peptides. Natural peptides with antibiotic activity (mainly against other Gram-positive bacteria) include: Gramicidin-S, a cyclic decapeptide from *B. brevis*; Bacitracin, a branched cyclic dodecapeptide produced by *B. licheniformis*; and Surfactin, a powerful surfactant with antibiotic activity produced by most strains of *B. subtilis*. A minority of the *Bacillus* peptide antibiotics are synthesized on ribosomes, but then subject to extensive post-translational modifications. Their products are usually somewhat larger than those produced by the peptide synthetases and include subtilin, a 32-residue lantibiotic produced by *B. subtilis* that shows antibacterial and antitumour activity.

References

Asai K, Yamaguchi H, Kang C-M *et al.* (2003) DNA microarray analysis of *Bacillus subtilis* sigma factors of extracytoplasmic function family. *FEMS Microbiology Letters* **220**: 155–160.

Bernhardt J, Völker U, Völker A *et al.* (1997) Specific and general stress proteins in *Bacillus subtilis* – A two dimensional protein electrophoretic study. *Microbiology* **143**: 999–1017.

Bernhardt J, Buttner K, Coppee JY *et al.* (2001) The contribution of the European Community consortium to the two-dimensional protein index of *Bacillus subtilis*. In Schumann W, Ehrlich SD and Ogasawara N (eds) *Functional Analysis of Bacterial Genes*. Chichester: Wiley.

Bloor AE and Cranenburgh RM (2006) An efficient method of selectable marker gene excision by Xer recombination for gene replacement in bacterial chromosomes. *Applied and Environmental Microbiology* **72**: 2520–2525.

Carballido-Lopez R and Errington J (2003) The bacterial cytoskeleton: *in vivo* dynamics of the actin-like protein Mbl of *Bacillus subtilis*. *Developmental Cell* **4**: 19–28.

Dubnau D (1993) Genetic exchange and homologous recombination. In Sonenshein AL, Hoch JA and Losick R (eds) *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, pp. 555–584. Washington DC: American Society for Microbiology.

Dworkin J and Losick R (2005) Developmental commitment in a bacterium. *Cell* **121**: 401–409.

Espinosa M, del Solar G, Rojo F and Alonso JC (1995) Plasmid rolling circle replication and its control. *FEMS Microbiology Letters* **130**: 111–120.

Eymann C, Mach H, Harwood CR and Hecker M (1996) Phosphate starvation-inducible protein in *Bacillus subtilis* a two dimensional protein electrophoretic study. *Microbiology* **142**: 3163–3170.

Fischer E and Sauer U (2005) Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nature Genetics* **37**: 636–640.

Harwood CR and Cutting SM (1990) *Molecular Biological Methods for Bacillus*. Chichester: Wiley.

Hecker M and Völker U (2004) Towards a comprehensive understanding of *Bacillus subtili*s cell physiology by physiological proteomics. *Proteomics* **4**: 3727–3750.

Helmann JD and Moran CP (2002) RNA polymerase and sigma factors. In Sonenshein AL, Hock JA and Losick R (eds) *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*. Washington DC: American Society for Microbiology.

Ishii T, Yoshida Ki, Terai G, Fujita Y and Nakai K (2001) DBTBS: a database of *Bacillus subtilis* promoters and transcription factors. *Nucleic Acids Research* **29**: 278–280.

Janniere L, Gruss A and Ehrlich SD (1993) Plasmids. In Sonenshein AL, Hock JA and Losick R (eds) *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, pp. 625–644. Washington DC: American Society for Microbiology.

Jensen CL, Stephenson K, Jørgensen ST and Harwood CR (2000) Cell-associated degradation affects yield of secreted engineered and heterologous proteins in the *Bacillus subtilis* expression system. *Microbiology* **146**: 2583–2594.

Jürgen B, Barken KB, Tobisch S *et al*. (2005) Application of an electric DNA-chip for the expression analysis of bioprocess-relevant marker genes of *Bacillus subtilis*. *Biotechnology and Bioengineering* **92**: 299–307.

Kobayashi K, Ehrlich SD, Albertini A *et al*. (2003) Essential *Bacillus subtilis* genes. *Proceedings of the National Academy of Sciences of the USA* **100**: 4678–4683.

Kunst F, Ogasawara N, Moszer I *et al*. (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249–256.

Leaver M and Errington J (2005) Roles for MreC and MreD proteins in helical growth of the cylindrical cell wall in *Bacillus subtilis*. *Molecular Microbiology* **57**: 1196–1209.

Leenhouts K, Buist G, Bolhuis A *et al*. (1996) A general system for generating unlabelled gene replacements in bacterial chromosomes. *Molecular and General Genetics* **253**: 217–224.

Meijer WJJ, de Boer A, van Tongeren S, Venema G and Bron S (1995) Characterization of the replication region of the *Bacillus subtilis* pLS20: a novel type of replicon. *Nucleic Acids Research* **23**: 3214–3223.

Meijer WJJ, Wisman GBA, Terpstra P *et al*. (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 Microbiology Reviews* **21**: 337–368.

Meile J-C, Wu LJ, Ehrlich SD, Errington J and Noirot P (2006) Systematic localisation of proteins fused to the green fluorescent protein in *Bacillus subtilis*: Identification of new proteins at the DNA replication factory. *Proteomics* **6**: 2135–2146.

Piggot PJ and Hilbert DW (2004) Sporulation of *Bacillus subtilis*. *Current Opinion in Microbiology* **7**: 579–586.

Rocha E, Moszer I, Klaerr-Blanchard M *et al*. (2001) *In silico* genome analysis. In Schumann W, Ehrlich SD and Ogasawara N (eds) *Functional Analysis of Bacterial Genes*, pp. 6–19. Chichester: Wiley.

Rocha EPC, Danchin A and Viari A (1999) Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms. *Molecular Biology and Evolution* **16**: 1219–1230.

Saier MHJ, Goldman SR, Maile RR *et al*. (2002) Overall transport capabilities of *Bacillus subtilis*. In Sonenshein AL, Hock JA and Losick R (eds) *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*, pp. 113–128. Washington DC: American Society for Microbiology.

Sarvas M, Harwood CR, Bron S and van Dijl JM (2004) Post-translocational folding of secretory proteins in Gram-positive bacteria. *Biochimica et Biophysica Acta (BBA) – Molecular Cell Research* **1694**: 311–327.

Vagner V, Dervyn E and Ehrlich SD (1998) A vector for systematic gene inactivation in *Bacillus subtilis*. *Microbiology* **144**: 3097–3104.

Westers H, Dorenbos R, van Dijl JM *et al*. (2003) Genome engineering reveals large dispensable regions in *Bacillus subtilis*. *Molecular Biology and Evolution* **20**: 2076–2090.

Williams RC, Rees ML, Jacobs MF *et al*. (2003) Production of *Bacillus anthracis* protective antigen is dependent on the extracellular chaperone, PrsA. *Journal of Biological Chemistry* **278**: 18056–18062.

Wolff S, Otto A, Albrecht D *et al*. (2006) Gel-free and gel-based proteomics in *Bacillus subtilis*: a comparative study. *Molecular and Cellular Proteomics* **5**: 1183–1192.

Further Reading

Sonenshein AL, Hoch JA and Losick R (eds) (2002) *Bacillus subtilis* and it closest relatives: From genes to cells. Washington DC: ASM Press.

Schumann W, Ehrlich SD and Ogasawara N (eds) (2001) Functional analsysis of bacterial genes: A practical approach. Chichester: Wiley.

Thompson DK and Zhou J (2004) The functional genomics of model organisms: addressing old questions from a new perspective. In: Zhou J, Thompson DK, Xu Y and Tiedje JM (eds) *Microbial Functional Genomics*. New Jersey: Wiley.