Organization of the *Bacillus subtilis* 168 chromosome between *kdg* and the attachment site of the *SP β* prophage: use of Long Accurate PCR and yeast artificial chromosomes for sequencing

Véronique Capuano, Nathalie Galleron, Petar Pujic, Alexei Sorokin and S. Dusko Ehrlich

Author for correspondence: Alexei Sorokin. Tel: +33 1 34 65 25 33. Fax: +33 1 34 65 25 21.
e-mail: sorokine@biotec.jouy.inra.fr

Laboratoire de Génétique Microbienne, Institut National de la Recherche Agronomique, Domaine de Vilvert, 78352 Jouy en Josas cedex, France

Within the *Bacillus subtilis* genome sequencing project, the region between *lysA* and *ilvA* was assigned to our laboratory. In this report we present the sequence of the last 36 kb of this region, between the *kdg* operon and the attachment site of the *SP β* prophage. A two-step strategy was used for the sequencing. In the first step, total chromosomal DNA was cloned in phage M13-based vectors and the clones carrying inserts from the target region were identified by hybridization with a cognate yeast artificial chromosome (YAC) from our collection. Sequencing of the clones allowed us to establish a number of contigs. In the second step the contigs were mapped by Long Accurate (LA) PCR and the remaining gaps closed by sequencing of the PCR products. The level of sequence inaccuracy due to LA PCR errors appeared to be about 1 in 10 000, which does not affect significantly the final sequence quality. This two-step strategy is efficient and we suggest that it can be applied to sequencing of longer chromosomal regions. The 36 kb sequence contains 38 coding sequences (CDSs), 19 of which encode unknown proteins. Seven genetic loci already mapped in this region, *xpt*, *metB*, *ilvA*, *ilvD*, *thyB*, *dfrA* and *degR* were identified. Eleven CDSs were found to display significant similarities to known proteins from the data banks, suggesting possible functions for some of the novel genes: *cspD* may encode a cold shock protein; *bcsA*, the first bacterial homologue of chalcone synthase; *exol*, a 5' to 3' exonuclease, similar to that of DNA polymerase I of *Escherichia coli*; and *bsaA*, a stress-response-associated protein. The protein encoded by *ypIP* has homology with the transcriptional NifA-like regulators. The arrangement of the genes relative to possible promoters and terminators suggests 19 potential transcription units.

Keywords: *Bacillus subtilis*, genome sequencing, yeast artificial chromosome, Long Accurate PCR

INTRODUCTION

Two types of approach are presently being followed in the sequencing of small genomes. One is based on random collections of cloned genome segments. The segments are

Abbreviations: CDS, coding sequence; LA PCR, Long Accurate PCR; YAC, yeast artificial chromosome.

The GenBank accession number for the sequence reported in this paper is L77246.

sequenced to a high redundancy, typically greater than sixfold above the genome size, the resulting sequence is assembled into contigs and the gaps between the contigs filled by a variety of methods. This approach was used to sequence the complete genomes of *Haemophilus influenzae* and *Mycoplasma genitalium* (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995). Another approach is based on the use of ordered collections of genome segments. The segments are sequenced independently to completion and the final sequence is then assembled. This approach was used to

sequence several *Saccharomyces cerevisiae* chromosomes (Oliver *et al.*, 1992, Dujon *et al.*, 1994).

Within the *Bacillus subtilis* genome sequencing project we are following an approach based on the ordered collection of segments in yeast artificial chromosomes (YACs; Azevedo *et al.*, 1993). Sequencing of one YAC, 15-6B, carrying an insert of about 105 kb, involved: (i) cloning of YAC DNA in phage M13-based vectors and random sequencing of the clones; (ii) contig ordering, by combinatorial PCR and (iii) gap closure by direct sequencing of the PCR products (Sorokin *et al.*, 1996a). The resulting new sequence extends from *serA* to the *kdgT* gene and covers a part of the chromosome region assigned to our laboratory, between *lysA* and the attachment site of phage *SPβ* (*attSPβ*). Here we describe the sequencing of another YAC, YAC 10-9, from the same collection, that completes the sequence of this region. A somewhat different approach was used to sequence this YAC. Its DNA was purified by PFGE and used as a hybridization probe for screening a *B. subtilis* total DNA bank, established in an M13mp18 vector. The contigs formed by sequencing of the identified clones were mapped by Long Accurate PCR (LA PCR; Barnes, 1994). The use of LA PCR products as a source of DNA for sequencing has not been reported previously. We therefore estimated the error rate of DNA amplification, which is important for further utilization of the procedure in the framework of different sequencing projects, and found it to be not significant. Sequencing of the LA PCR products allowed us to extend the contiguous sequence of this region of the *B. subtilis* chromosome to the *attSPβ* site.

METHODS

Strains and growth conditions. *Bacillus subtilis* strain 168, originally provided by C. Anagnostopoulos (Jouy-en-Josas, France) was used as the chromosomal DNA donor. *Escherichia coli* JJC128F' (*araD139 Δ(ara-leu)7696 galE15 galK16 Δ(lac)Y74 hsdR⁻ hsdM⁺ strA F' [lacI^q lacZΔM15 traD36]*) or *E. coli* TG1 (*K-12 Δ(lac-pro) supE thi hsdR⁻ F' [lacI^q lacZΔM15 traD36]*) were used as recipient strains for M13mp18 (Messing & Vieira, 1982) or pSGMU2 (Errington, 1986) cloning and DNA preparation. Luria Broth medium was used for standard culture of *B. subtilis*. 2YT (Yeast Tryptone) medium was used for *E. coli*. Yeast cells containing artificial chromosomes were grown as described by Azevedo *et al.* (1993) and YAC 10-9 isolation was carried out as described by Sorokin *et al.* (1996a). This protocol allowed about 1 µg YAC DNA per 500 ml yeast cell culture to be obtained.

Genomic banks and screening. *B. subtilis* chromosomal DNA was partially digested by *AluI* or randomly fragmented, using pancreatic DNase I or nebulization. Randomized DNA was electrophoresed on agarose gels and fragments from 600 to 1500 bp were purified using the GeneClean kit (Bio 101). Ends of fragments generated by pancreatic DNase I or by nebulization were filled using the Klenow fragment of *E. coli* DNA polymerase I. DNA fragments were cloned into M13mp18 DNA linearized at the *SmaI* polylinker site and dephosphorylated. Yield of phage plaques was about 10⁷ (µg M13 DNA)⁻¹ for cloning of *AluI* or nebulized DNA fragments. Some 10% of the plaques were white on plates containing IPTG. The cloning of fragments generated by pancreatic DNase I yielded 10⁶ plaques (µg M13 DNA)⁻¹, 50% of which were white. Plaques were replicated on Biotyne A membranes (Pall) and hybridized

with 10 ng ³²P-labelled YAC 10-9 DNA using the Random Primed DNA Labeling Kit (Boehringer-Mannheim). Phages corresponding to the hybridizing plaques were propagated on *E. coli* strain JJC128F' for 15 h in 48-well plates. The following steps were carried out with the aid of a BIOMEC 1000 Laboratory workstation (Beckman). Phage suspensions were stored in 96-well plates and 15 µl of each were used to prepare the filters (Hybond N⁺, Amersham) for a second hybridization screening. Positive clones (410 were identified) were stored and used for sequencing. About 90% of the positive clones carried inserts from the target region, as deduced from the analysis of the sequence. The remaining clones carried inserts from other regions of the *B. subtilis* chromosome and about 50% had homology with known genes.

PCR procedures. Oligonucleotides were synthesized on DNA synthesizer 'Oligo 1000' (Beckman). For amplification of fragments shorter than 2 kb, standard PCR was carried out, using either M13 ssDNA clones or *B. subtilis* chromosomal DNA as template. LA PCR was carried out essentially as described by Cheng *et al.* (1994). Oligonucleotides of 20–23 bases, 12 of which were G or C residues, were used under the following reaction conditions: 20 mM Tricine, pH 8.7; 85 mM KOAc; 1 mM Mg(OAc)₂; 8% (v/v) glycerol; 2% (v/v) DMSO; 0.2 mM each dNTP (Perkin Elmer); 0.2 µM each primer; 2 U *rTth* polymerase (Perkin Elmer); 0.05 U Vent polymerase (Biolabs) and 0.1 µg *B. subtilis* chromosomal DNA prepared as described by Sorokin *et al.* (1996b). The final reaction volume was 50 µl. Cycling conditions were: 94 °C, 5 min; 12 cycles of 10 s at 94 °C and 12 min at 68 °C and 24 cycles increasing the extension time by 15 s each cycle.

PCR fragment subcloning. LA PCR reaction products, corresponding to the *xpt-degR*, *degR-thyB* or *thyB-attSPβ* interval were partially cut by *HpaII*, *TaqI* or *Sau3A*. The cleaved DNA was electrophoresed on agarose gels and the fragments of 500–1500 bp were purified and inserted into the *AclI* or *BamHI* site of M13mp18 or pSGMU2 (Errington, 1986). For direct sequencing and sometimes for cloning, the LA PCR products, obtained by mixing 10 independent reactions, were purified using a Wizard PCR Preps Kit (Promega) or, after electrophoresis on agarose gels, using GeneClean (Bio 101).

Sequencing. ssDNA of phage clones was prepared by the perchlorate method (Zimmermann *et al.*, 1989) as previously described (Sorokin *et al.*, 1993). Plasmid DNA was prepared as described by Sorokin *et al.* (1995). PCR with biotinylated primers (Dynal) carried out as specified by the supplier, was used to prepare DNA for reverse sequencing of the M13 phage inserts. Direct and reverse sequencing was carried out with Applied Biosystems PRISM and Reverse Sequencing kits, respectively, on the thermal cycler (Perkin Elmer) or Catalyst station (Applied Biosystems).

Computing. R. Staden's XBAP program (Dear & Staden, 1991) was used for gel assembly and generation of a consensus sequence. Sequence homologies were searched using FASTA (Pearson & Lipman, 1988) or BLAST (Altschul *et al.*, 1990) algorithms, realized on the NCBI e-mail server.

Nomenclature. The nomenclature for coding sequences (CDSs) is in accordance with the *B. subtilis* genome project as described by Sorokin *et al.* (1996a).

RESULTS AND DISCUSSION

Sequencing of the region between *kdg* and *attSPβ*

YAC 10-9 contains an insert which was estimated to be ~ 100 kb by PFGE (Azevedo *et al.*, 1993) and thus represents approximately 2.5% of the total *B. subtilis* genome. The

Fig. 1. Organization of the *kdg-attSP β* region of the *B. subtilis* 168 chromosome. (a) The upper continuous line represents the chromosome between *kdg* and *attSP β* . The genes are placed according to the LA PCR products obtained with primers corresponding to the relevant gene sequence. The LA PCR products are symbolized by double arrows and their size is noted. The number of contigs identified between *kdg* and *degR*, *degR* and *thyB*, and *thyB* and *attSP β* markers is indicated. The numbers in parentheses refer to the total contig length. (b) The dashed line represents a part of the YAC 10-9 insert. The box gives the scale in kb. Double arrows indicate the three LA PCR products used for subcloning. Bold arrows represent putative CDSs. Current CDS designations and suggested gene names are shown respectively below and above the scale bar. The putative ρ -independent transcription terminators are indicated by t.

YAC DNA was purified and used as a probe for hybridization with a total chromosomal DNA bank, constructed in the M13mp18 phage vector. Of the 60 000 phage clones tested, 410 were positive. This number is almost fourfold lower than expected from the ratio of the YAC to chromosome length, which may be due to several mutually non-exclusive factors: (i) the size of the YAC may have been overestimated; (ii) a cloning bias against the regions covered by the YAC may exist; and (iii) the hybridization procedure used may not have allowed detection of all clones homologous to the probe.

All hybridizing inserts were sequenced, resulting in a total sequence length of about 54 kb, distributed among 125 contigs. Homology searches with protein data banks allowed us to identify the contigs corresponding to *ilvA*, *metB*, *degR*, *thyB* and *xpt* genes and also the contigs carrying the *SP β* immunity region (*immSP β*). A contig which has homology with GenBank entries M81760 and M81762, corresponding to *attSP β* , was also identified. Oligonucleotides corresponding to these contigs and to the *kdgT* gene were designed and used for LA PCR amplification. Overlapping PCR products which covered

the region between *kdg* and *attSP β* were thus obtained, allowing construction of the genetic and physical map (Fig. 1a). The region between *attSP β* and *immSP β* was not amplified, which indicates that the distance between the two sites is longer than 25 kb, the maximal amplifiable fragment.

The LA PCR products were used as hybridization probes to attribute the 410 phage clones, and thus the sequenced contigs, to the intervals of the genetic map. The number of contigs and the length of the determined sequence in each of the three amplified 12 kb intervals is shown in Fig. 1(a). The total corresponding contig length was ~24.4 kb, that is ~68% of the size of this region, which indicates that the random sequencing had generated some 70% of the sequence of the YAC 10-9 insert. Since we did not intend to sequence the phage *SP β* region present in the YAC, the sequences of the clones containing phage fragments were not used.

To complete the sequencing of the area delimited by the *kdg* and *attSP β* markers, the three contiguous LA PCR products, *xpt-degR*, *degR-thyB* and *thyB-attSP β* , were

Table 1. General features of the CDSs

CDS*	Endpoints (nt)†	Size (aa)	Translation start‡	PIR entry	Description	Match (%)§
<i>yprP</i> (<i>xpr</i>)	1426–2010	194	AcGGAGGTaGAcagg ATG	rteca	Adenine phosphoribosyl transferase (<i>E. coli</i>)	27 (141)
<i>yprQ</i> (<i>pbnX</i>)	2007–3323	438	AGGAGGTtcattc ATG	s34223	Uracil transport protein (<i>E. coli</i>)	26 (410)
<i>yprP</i> (<i>bcsA</i>)	3397–4494	365	AaGAGGTGATcgc ATG	ps0079	Chalcone synthase (plants)	26 (364)
<i>yprQ</i>	4498–5004	168	AAGGgGGcctgactc ATG			
<i>yprR</i>	5340–8921	1193	AAAGgGGttcagactt ATG			
<i>yprS</i>	8986–9243	85	AAGcGAGaTGAacgatta ATG			
<i>yprP</i> (<i>exoI</i>)	9564–10455	296	AAGGgaGTtatttcg ATG	jx0256	5' Exonuclease of PolI (<i>E. coli</i>)	36 (294)
<i>yprQ</i>	10633–11031	132	AAGGAGrTcata ATG	P36921 S13170	Cell wall enzyme (<i>Ent. faecalis</i>) Ribonuclease HI (<i>E. coli</i>)	41 (125) 34 (74)
<i>yprP</i>	11031–11720	229	AAGGAaGcttttcac TTG			
<i>yprP</i>	11803–12483	226	aatGGtaAGGTGATtga ATG			
<i>yprQ</i>	12476–12658	60	AcGGAGaTGAacgata TTG			
<i>degR</i>	13111–13293	60	AAGGgGaaccgacgt ATG	M15318 ¶	Degradative enzyme regulatory protein PrtR	100 (60)
<i>yprQ</i> (<i>cspD</i>)	13345–13545	66	AGGAGGaattaagcaat ATG	a45723	Cold-shock protein CspB (<i>B. subtilis</i>)	82 (66)
<i>yprP</i>	13788–14936	382	AGGTGAatttac TTG			
<i>metB</i>	15399–16073	224	AtGGAGGTGgcacg TTG	xyecm	Homoserine succinyl transferase (<i>E. coli</i>)	47 (221)
<i>yprP</i> (<i>bsaA</i>)	16416–16898	160	AGGAGGgacttac ATG	s33618	Salt-stress-associated protein (plants)	51 (159)
				S46121	Glutathione peroxidase homologue (<i>Sacch. cerevisiae</i>)	52 (159)
				qrecbe	B12 binding protein (<i>E. coli</i>)	48 (117)
<i>yprQ</i>	16914–17531	205	AAAGAGGgtgtttt GTG			
<i>yprR</i>	17591–18724	377	AGGAGGgacgac ATG	S31054	MpeV protein (<i>Synechococcus</i> sp.)	26 (120)
<i>ilvD</i>	19013–20689	558	AGGAGGatcacc ATG	dwecca	Dihydroxyacid dehydratase (<i>E. coli</i>)	51 (367)
<i>yprP</i>	20796–21230	144	actaGAGGgggcttttt ATG			
<i>yprP</i>	21527–22045	172	ggGGAGaaaGAacct ATG	s12697	Site-specific DNA methylase	25 (95)
<i>yprP</i>	22305–22916	203	AgaGAGGaataagcagtc ATG			
<i>yprQ</i>	22934–23467	177	AAGGAGaTgccattagac ATG			
<i>thyB</i>	23551–24345	264	AAGGAaGaaaaa ATG	sybstb	Thymidilate synthase B (<i>B. subtilis</i>)	99 (264)
<i>dfrA</i>	24342–24848	168	AAGGgGcggtcagcgt ATG	rdbsb	Dihydrofolate reductase (<i>B. subtilis</i>)	99 (168)
<i>yprP</i>	24849–25469	206	GGAGGAtttta TTG			
<i>yprQ</i>	25508–26149	213	AAGGAGaatgtac TTG			
<i>yprP</i>	26170–27165	331	AAGGAGccaattg ATG	rgecgg	Nitrogen assimilation regulatory protein (<i>E. coli</i>)	30 (217)
<i>ilvA</i>	27414–28682	422	AAGaaGGAatctgtac ATG	M58606	Threonine dehydratase	100 (422)
<i>yprP</i>	28768–29019	83	AAGAGGattgatagc TTG			
<i>yprQ</i>	29167–29748	193	AAGGAaGgatgtc ATG			
<i>yprR</i>	29826–30593	255	AAGGAGcaacaaga TTG			
<i>yprS</i>	30603–31166	187	agagtagGGTGacagc ATG			
<i>yprT</i>	31179–31373	64	AtaGAGGTGactgaaa ATG			
<i>yprP</i>	31445–32782	445	AAGGAGaaaag ATG			
<i>yprP</i>	32832–33257	141	AAtGGAGGTtcacaaat TTG	s14473 s16409	Regulatory protein MprA (<i>E. coli</i>) 17 kDa hypothetical protein (<i>E. coli</i>)	23 (141) 25 (120)
<i>yprP</i>	33389–33922	177	AAGGAGGcaccagt ATG	jc1317	Methionine S oxidoreductase (<i>E. coli</i>)	44 (160)
				jq0988	DNA-binding protein E4 (tomato)	42 (151)
				s02018	Regulatory protein PilB (<i>N. gonorrhoeae</i>)	35 (123)
<i>yprQ</i>	33922–34353	143	ttgGGgGGctaataa ATG	s02018	Regulatory protein PilB (<i>N. gonorrhoeae</i>)	55 (126)
<i>pipL</i> <i>attSPβ</i> **	34398–35021 35018				Interrupted by <i>SPβ</i> prophage	

* Nomenclature for the putative CDSs is described in the text.

† Corresponding to L77246 GenBank entry.

‡ The putative start codon is indicated in bold and nucleotides complementary to 16S RNA are indicated in upper case letters.

§ The number of amino acids over which the percentage match was determined is shown in parentheses.

|| SwissProt entry. ¶ GenBank entry. ** The position of the attachment site of *SPβ* is indicated.

used to construct M13 or pSGMU2 banks. About 300 clones were chosen at random and used as templates for ~ 500 sequencing reactions. Some 200 000 bases were thus read, and a consensus sequence of 25 kb established. A total of 21 point mutations was revealed in the cloned LA PCR products. The error rate of the amplification procedure was therefore estimated to be approximately 1 in 10 000 bp (21/200 000), which is far below the error rate of a standard sequencing run, and thus should not affect the quality of the final sequence. In all cases of sequence discrepancy, the correct sequence was determined by sequencing several independent clones and/or direct sequencing of the product mixture from independent LA PCR experiments. The gaps remaining in the sequence were filled by direct sequencing of the PCR products. The final reading redundancy of the 35.5 kb sequence between *kdg* and *attSP β* was about 6. It should be possible to significantly reduce this value, by replacing some or most of the shotgun sequencing of the LA PCR products by primer walking. All of the sequence was read at least once on each strand.

Identification of putative genes

CDSs larger than 60 codons, exhibiting ATG, TTG or GTG as start codon and which had an appropriately located upstream sequence complementary to the 3' end of 16S rRNA (McLaughlin *et al.*, 1981) were analysed. By these criteria, the region contains 38 CDSs and one truncated coding sequence, interrupted by the insertion of the *SP β* prophage. As reported in Table 1, ATG is the putative start codon of 28 CDSs, while TTG and GTG are the initiation codons for 9 and 1 CDSs, respectively. The genetic organization of the region is shown in Fig. 1(b). A transcriptional organization is also proposed, as deduced from the presence of inverted repeats followed by a T-rich sequence that could form a *p*-independent transcription terminator (Fig. 1b). The genes localized between the adjacent transcription terminators and transcribed in the same direction were considered to be in the same operon. We found 19 putative transcriptional units, 13 of which were co-directional with chromosome replication, a proportion similar to that observed by Zeigler *et al.* (1990). Sequence analysis indicated that 19 of the 38 proteins encoded by the CDSs were found to have homologues in the data banks, a value higher than the ~ 40% observed in many systematic sequencing projects (cf. Fleischmann *et al.*, 1995). The function of the genes that have homologues, and the transcriptional features of the region, are discussed below.

Gene functions in the *xpt-attSP β* region

ypaP-ypaQ (xpt-pbuX). The xanthine phosphoribosyl transferase gene (*xpt*), mutations which confer resistance to purine analogues, maps in this region (Saxild & Nygaard, 1987) and probably corresponds to *ypaP*. This hypothesis was confirmed by sequencing of the corresponding mutations (EMBL/GenBank/DDBJ accession

no. X83878). The other gene of this operon, *ypaQ*, probably encodes a xanthine transport protein.

ypbP (bcsA). Since *ypbP* is the first bacterial gene related to plant chalcone synthase (Table 1) we propose the name *bcsA* (bacterial chalcone synthase A). Chalcone synthase catalyses the first step in the branch of phenylpropanoid metabolism leading to the synthesis of the ubiquitous flavonoid pigments and UV protectants. This pathway also leads to the synthesis of the isoflavonoid antibiotic phytoalexin, in response to wounding or microbial attack, which inhibits microbial growth by disrupting the cell membrane (for review, see Phillips & Kapulnik, 1995). In plants, different external stimuli, such as fungal elicitors or UV illumination, can induce the expression of chalcone synthase activity (Ryder *et al.*, 1987). It is conceivable that BcsA is involved in bacterial adaptive and protective responses to environmental stress.

ypcP (exon). The product encoded by *ypcP* has > 35% identity with the N-terminal domain of DNA polymerase I (PolI) from *E. coli* and other species. The small product of partial PolI proteolysis (residues 1–300) carries the 5'–3' exonuclease activity while the large fragment contains 3'–5' exonuclease and DNA polymerase activities. Genetically encoded separated functions of PolI-type polymerases are known for bacteriophage T5, in which the gene denoted D15 encodes a 5'–3' exonuclease (Leavitt & Ito, 1989). D15 exonuclease is thought to be part of the bacteriophage T5 transcription–replication enzyme complex and may be involved in nick translation or in accumulation of ssDNA (Sayers & Eckstein, 1991). PolI has been purified from *B. subtilis* and the exonuclease activity was reported to be absent (Okazaki & Kornberg, 1964). However, the *polA* gene that encodes PolI was sequenced in our laboratory and the intrinsic 5'–3' exonuclease domain has been found (unpublished results). The exonuclease function of the YpcP protein might therefore indicate the redundant nature of this activity in *B. subtilis*.

ypdQ. The *ypdQ* gene encodes a protein which is 40% identical to a cell wall enzyme (*ebcB* gene product) from *Enterococcus faecalis* (Bensing & Dunny, 1993). Conjugation in *Ent. faecalis* is thought to be mediated by an aggregation substance which may interact with a surface protein (binding substance) of the recipient cell. It was proposed that the *ebc* locus is related to the synthesis or assembly of the binding substance and that EbcB may act as a cell wall hydrolase. The 70 aa N-terminal region of YpdQ also exhibits similarity to the N-terminal region of ribonuclease HI from *E. coli*. Conserved amino acids, presumed to be at the catalytic site of RNase HI (Itaya & Kondo, 1991), are also present in YpdQ.

degR. The *degR* gene, involved in the regulation of transcription of the extracellular degradative enzyme genes, known also as *prrR*, has been identified and sequenced earlier by Yang *et al.* (1987).

ypfQ (cspD). *ypfQ* encodes a protein which is 82% identical to the major cold-shock protein CspB from *B. subtilis*

Fig. 2. Alignment of amino acid sequences. (a) Comparison of the *B. subtilis* CspD (Bs CspD) and the *B. subtilis* cold-shock protein B (Bs CspB). (b) Comparison of the *B. subtilis* *ypgP* gene product (Bs YpgP) and the *E. coli* B12-binding protein (Ec BtuE) and *Sacch. cerevisiae* glutathione peroxidase homologue (Sc GSH). (c) Comparison of the *B. subtilis* *yplP* gene product (Bs YpIP) and bacterial transcriptional activators *Klebsiella pneumoniae* NifA (Kp NifA) and NtrC (Kp NtrC), *B. subtilis* LevR (Bs LevR) and RocR (Bs RocR). (d) Comparison of the *B. subtilis* *yppP* and *ypqQ* gene products (Bs YppP and Bs YppQ), the tomato DNA-binding E4 protein (E4) and the *N. gonorrhoeae* PilB protein (Ng PilB). Black boxes indicate identical amino acids.

(Willinsky *et al.*, 1992). A gene encoding a protein homologous to CspB and named *cspC* has also been detected in *B. subtilis* (GenBank accession no. D25230). We therefore propose to designate *yplQ* as *cspD*. As shown in Fig. 2(a), the CspD protein has only 12 amino acid differences from CspB, all of which are in the non-structured regions of the protein (Schindelin *et al.*, 1993; Schnuchel *et al.*, 1993). It is thus very likely that CspD can substitute for CspB. The limits of the role of CspB have

been investigated by using a *B. subtilis* *cspB* mutant (Willinsky *et al.*, 1992). It was noted that the viability of the mutant is drastically affected at a freezing temperature, but cryotolerance increases when the cells are pre-incubated for 2 h at 10 °C before freezing. It was therefore suggested that genes other than *cspB* are involved in this tolerance. Probable candidates are *cspC* and *cspD*. It should be noted that the sequence 5' ATTGG, named the Y-box and thought to be recognized by CspB (Schröder *et*

Fig. 3. Alignment of the *B. subtilis* 168 genetic and CDS maps between *spoVAF* and *attSPβ*. The CDS map is compiled from Sorokin *et al.* (1993, 1996a) and the results of this study. Experimentally confirmed gene functions are shown above the scale bar. The corresponding references are compiled in Table 2.

al., 1995), is present 274 bp upstream of the initiation codon of *csdD*.

metB. The *metB* gene, which is involved in methionine biosynthesis, has been mapped in this part of the chromosome (Anagnostopoulos *et al.*, 1993). The entire 36 kb region contains only one CDS encoding a protein of the methionine biosynthetic pathway. It is thus highly probable that this CDS, homologous to homoserine succinyl transferase of *E. coli*, corresponds to the *metB* gene. The occurrence of a putative terminator downstream from *metB* suggests that it is transcribed as a monocistronic unit in the direction from the terminus to the origin of replication. Almost perfect –10 (TATAAT) and –35 (TTGAAA) promoter sequences are present 38 bp upstream from the translational start codon of *metB*. In *E. coli* and *S. typhimurium*, transcription of the *metB*

counterpart is negatively regulated by the MetR protein (Mares *et al.*, 1992). In the putative regulatory region of the *B. subtilis metB* gene we did not detect any sequence resembling the *E. coli* MetR binding site.

ypgP (bsaA). The translational start codon of *ypgP* is separated by 341 bp from that of the *metB* gene and the two genes are transcribed divergently. The first gene of the *ypg* operon, *ypgP*, has very high similarity (more than 50%, Table 1) with plant salt-stress-associated proteins. The corresponding homologue in the orange is encoded by *csaA* (citrus salt-stress-associated). The very high similarity of these proteins over their whole length strongly suggests that *ypgP* is involved in a stress response in *B. subtilis*. We therefore propose the designation *bsaA* (bacillus stress-associated) for this gene. This is the first example of a potential *csaA* counterpart found in bacteria, although one is known in yeast (PIR accession no. S46121). The similarity of these proteins to glutathione peroxidases may suggest that free radical levels increase during the salt stress response (Holland *et al.*, 1993). The third CDS of this operon (*ypgR*) shares homology with the *mpeV* gene product from cyanobacteria. The homology is, however, restricted to only a part of the two proteins, which suggests that they may have a conserved domain rather than a similar function.

thyB and dfrA. Thymidylate synthase, specified in *B. subtilis* by the *thyA* and *thyB* genes, catalyses the synthesis of dTMP from dUMP, while dihydrofolate reductase (*dfrA*) is involved in the reduction of dihydrofolate to tetrahydrofolate. The *thyB* and *dfrA* genes are closely linked and were cloned and characterized by Iwakura *et al.* (1988). Our translated sequence differs by one amino acid from both the ThyB (aa 144) and DfrA (aa 165) sequences of that study. Verification of the sequencing gels indicated no error in our sequence. A strain difference could possibly account for the discrepancy.

ilvD and ilvA. The genes *ilvA* and *ilvD*, which are involved in the synthesis of isoleucine, leucine and valine, have been genetically mapped in this part of the *B. subtilis* chromosome (Barat *et al.*, 1965). The *B. subtilis ilvA* gene has been cloned and sequenced (GenBank accession no. M58606). This sequence is identical to ours over 1.6 kb, but the 5'-terminal 130 bp are completely different. We believe that our sequence is correct, since it was determined from two different templates – the chromosomal DNA cloned in M13 phage and the LA PCR product, which was sequenced directly. The *ilvD* gene was identified by the homology of the encoded product to the corresponding *E. coli* enzyme (Table 1). In the species closest to *B. subtilis* from which *ilv* genes have been sequenced, *Lactococcus lactis*, as well as in *E. coli* and other enterobacteriaceae, *ilvA* and *ilvD* belong to the same operon (Godon *et al.*, 1992; Renault *et al.*, 1995; Driver & Lawther, 1985). In *B. subtilis*, however, the *ilvA* and *ilvD* genes are separated by 10 kb.

ypIP. The N-terminal part of the *ypIP* gene product is homologous to the central domain of the NifA and NtrC proteins, which control nitrogen assimilation in Gram-negative bacteria (Drummond *et al.*, 1986). These proteins

Table 2. Experimentally defined gene functions in the *lysA-attSP β* region of the *B. subtilis* chromosome

Gene*	Gene product function	Biological function or gene category	Reference†
<i>lysA</i>	DAP decarboxylase	Lysine biosynthesis	Yamamoto <i>et al.</i> (1991)
<i>ppiB</i>	Peptidyl-prolyl-isomerase	Unknown	Herrler <i>et al.</i> (1994)
<i>sipS</i>	Signal peptidase	Protein secretion	van Dijl <i>et al.</i> (1992)
<i>ribG</i>	Deaminase	Riboflavin biosynthesis	Mironov <i>et al.</i> (1994)
<i>ribB</i>	Riboflavin synthase (α -subunit)	Riboflavin biosynthesis	Schott <i>et al.</i> (1990)
<i>ribA</i>	GTP cyclohydrolase II	Riboflavin biosynthesis	Mironov <i>et al.</i> (1994)
<i>ribH</i>	Lumazine synthase (β -subunit)	Riboflavin biosynthesis	Ludwig <i>et al.</i> (1987)
<i>ribT</i>	Reductase	Riboflavin biosynthesis	Mironov <i>et al.</i> (1994)
<i>dacB</i>	D,D-Carboxypeptidase	Cell wall biogenesis	Buchanan & Ling (1992)
<i>spmA</i>	Unknown	Spore maturation	Popham <i>et al.</i> (1995)
<i>spmB</i>	Unknown	Spore maturation	Popham <i>et al.</i> (1995)
<i>resD</i>	Signal transduction regulator	Respiration regulation	Sun <i>et al.</i> (1996)
<i>resE</i>	Protein kinase	Respiration regulation	Sun <i>et al.</i> (1996)
<i>aroC</i>	3-Dehydroquinase	Aromatic amino acid biosynthesis	Sorokin <i>et al.</i> (1993)
<i>serA</i>	Phosphoglycerate dehydrogenase	Serine biosynthesis	Sorokin <i>et al.</i> (1993)
<i>gpsA</i>	Glycerol-3-phosphate dehydrogenase	Membrane biosynthesis	Morbidoni <i>et al.</i> (1995)
<i>spoIV_A</i>	Spore coat assembly protein	Spore maturation	Driks <i>et al.</i> (1994)
<i>dbpA</i>	DNA-binding histone-like protein	Chromosome morphogenesis	Micka & Marahiel (1992)
<i>mtrA</i>	GTP cyclohydrolase I	Folic acid biosynthesis	Babitzke <i>et al.</i> (1992)
<i>mtrB</i>	RNA-binding transcription regulator	Tryptophan biosynthesis regulation	Babitzke <i>et al.</i> (1992)
<i>gerC1</i>	Unknown	Spore germination	Yazdi & Moir (1990)
<i>gerC2</i>	Unknown	Spore germination	Yazdi & Moir (1990)
<i>gerC3</i>	Unknown	Spore germination	Yazdi & Moir (1990)
<i>cheR</i>	Chemotactic methyltransferase	Chemotaxis	Ordal <i>et al.</i> (1985)
<i>aroH</i>	Chorismate mutase, isozymes 1 and 2	Aromatic amino acid biosynthesis	Gray <i>et al.</i> (1990)
<i>trpE</i>	Anthranilate synthase	Tryptophan biosynthesis	Henner <i>et al.</i> (1984)
<i>trpD</i>	Anthranilate phosphoribosyltransferase	Tryptophan biosynthesis	Henner <i>et al.</i> (1984)
<i>trpC</i>	Indol-3-glycerolphosphate synthase	Tryptophan biosynthesis	Henner <i>et al.</i> (1984)
<i>trpF</i>	Phosphoribosylanthranilate isomerase	Tryptophan biosynthesis	Henner <i>et al.</i> (1984)
<i>trpB</i>	Tryptophan synthase β	Tryptophan biosynthesis	Henner <i>et al.</i> (1984)
<i>trpA</i>	Tryptophan synthase α	Tryptophan biosynthesis	Henner <i>et al.</i> (1984)
<i>hisH</i>	Histidinol-phosphate aminotransferase	Histidine biosynthesis	Henner <i>et al.</i> (1986)
<i>tyrA</i>	Prephenate dehydrogenase	Tyrosine biosynthesis	Henner <i>et al.</i> (1986)
<i>aroE</i>	3-Enolpyruvylshikimate-phosphate synthase	Aromatic amino acid biosynthesis	Henner <i>et al.</i> (1986)
<i>qcrA</i>	Rieske iron-sulphur protein	Aerobic respiration	Yu <i>et al.</i> (1995)
<i>qcrB</i>	Cytochrome <i>bc</i> complex <i>b</i> domain	Aerobic respiration	Yu <i>et al.</i> (1995)
<i>qcrC</i>	Cytochrome <i>bc</i> complex <i>c</i> domain	Aerobic respiration	Yu <i>et al.</i> (1995)
<i>birA</i>	Transcription regulator	Biotin biosynthesis regulation	Bower <i>et al.</i> (1995)
<i>aspB</i>	Aspartate aminotransferase	Aspartate biosynthesis	Hoch & Mathews (1972)
<i>dnaD</i>	Unknown	DNA replication	Bruand <i>et al.</i> (1995)
<i>ponA</i>	Transglycosylase and transpeptidase	Cell wall biogenesis	Popham & Setlow (1995)
<i>cotD</i>	Spore coat protein	Spore maturation	Donovan <i>et al.</i> (1987)
<i>rnpB</i>	RNase P RNA component	tRNA processing	Reich <i>et al.</i> (1986)
<i>xpt</i>	Xanthine phosphoribosyltransferase	Purine nucleotide biosynthesis	Saxild & Nygaard (1987)
<i>degR</i>	Unknown	Protein secretion regulation	Tanaka <i>et al.</i> (1987)
<i>metB</i>	Homoserine succinyl transferase	Methionine biosynthesis	Barat <i>et al.</i> (1965)
<i>ilvD</i>	Dihydroxyacid dehydratase	Isoleucine biosynthesis	Barat <i>et al.</i> (1965)
<i>thyB</i>	Thymidylate synthetase B	Thymidine biosynthesis	Iwakura <i>et al.</i> (1988)
<i>dfrA</i>	Dihydrofolate reductase	Folic acid metabolism	Iwakura <i>et al.</i> (1988)
<i>ilvA</i>	Threonine dehydratase	Isoleucine biosynthesis	Barat <i>et al.</i> (1965)

* Given in the order of their position on the map in Fig. 3.

† The most direct evidence of the product or biological function.

belong to the so-called NtrC/NifA family of transcription activators, which includes gene products involved in many different physiological functions, such as pilin formation (PilR), catabolism of xylene (XylR) or C4 dicarboxylate transport (DctD). In *B. subtilis*, the two examples known until now are involved in fructose and sucrose hydrolysis (LevR, Martin-Verstaete *et al.*, 1994) and in arginine catabolism (RocR, Calogero *et al.*, 1994). All the genes from the family contain a highly conserved domain, revealed by multiple alignments (Fig. 2c). This domain is supposed to be essential for formation of open promoter complexes with RNA polymerase (Popham *et al.*, 1989). It is possible that *yplP* encodes a protein which belongs to this family of regulators.

***ypoP*.** The *ypoP* gene product exhibits homology over the whole length with the MprA protein and with a hypothetical 17 kDa protein from *E. coli* (Table 1). MprA is thought to act as a transcriptional regulator involved in the osmoinduction of the *proU* operon (Del Castillo *et al.*, 1991). The 17 kDa polypeptide is encoded by an ORF located downstream of the *pap* and *prs* gene clusters, required for the formation of P pili in *E. coli* (Marklund *et al.*, 1992). However, a deletion mutation in the gene encoding the 17 kDa protein does not affect pilus formation.

***yppP* and *yppQ*.** These two CDSs, which are transcribed divergently from *ypoP*, have a strong homology with the same protein, the *pilB* gene product of *Neisseria gonorrhoeae* (Taha *et al.*, 1988). The YppP sequence aligns with PilB from amino acids 227 to 347 while YppQ aligns with PilB from amino acids 390 to 515 (Fig. 2d). PilB acts negatively *in trans* on the pilin promoter. It has been shown that the 280 N-terminal residues of PilB are sufficient for repressor activity. We propose that PilB of *N. gonorrhoeae* might be a bifunctional protein which interacts with DNA via its N-terminal domain. In both *N. gonorrhoeae* and *B. subtilis*, the *pilB* and *yppP* genes are located upstream of a transcription regulator gene and are transcribed divergently from it. It is tempting to suggest that *ypoP*, *yppP* and *yppQ* may have a function related to the same physiological process. Inactivation of expression of pilin genes in *N. gonorrhoeae* results in strains that exhibit marked competence deficiency. While in *B. subtilis* no structure resembling a pilus has been observed, the late competence *comG* protein is similar to a type IV *N. gonorrhoeae* pilus protein (Albano *et al.*, 1989). It might be interesting to test whether inactivation of YpoP, YppP or YppQ affects *B. subtilis* competence.

***yqpP* (*pipL*).** This CDS is interrupted by the *SP β* prophage. Obviously, its integrity is not essential for *B. subtilis*, since the corresponding protein is produced only when the phage is excised. We have designated this CDS *pipL* (Phage-Interrupted Protein near the Left attachment site).

Conclusions

The complete nucleotide sequence of the 35.5 kb *kdg-attSP β* region was determined in this study. A contiguous sequence of 153 kb, extending between the *lysA* gene and

the *SP β* prophage, and corresponding to the region assigned to our laboratory within the *B. subtilis* genome sequencing project, was thus completed. The genetic and CDS organization of this region, the latter based on current and previous work (Sorokin *et al.*, 1993, 1996a) are summarized in Fig. 3. The gene functions from this region which have been defined experimentally, and the corresponding references, are listed in Table 2. The sequencing strategy used here relied on LA PCR. The procedure has proved to be very powerful for ordering of the contigs formed by random sequencing, and for generating DNA both for subcloning and for direct sequencing. The sequence errors due to LA PCR are not significant. We conclude that LA PCR products can be used directly within genome sequencing projects.

REFERENCES

- Albano, M., Breitling, R. & Dubnau, D. (1989). Nucleotide sequence and genetic organization of the *Bacillus subtilis comG* operon. *J Bacteriol* **171**, 5386–5404.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J Mol Biol* **215**, 403–410.
- Anagnostopoulos, C., Piggot, P. J. & Hoch, J. A. (1993). The genetic map of *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 425–461. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Azevedo, V., Alvarez, E., Zumstein, E., Damiani, G., Sgarrella, V., Ehrlich, S. D. & Serró, P. (1993). An ordered collection of *Bacillus subtilis* DNA segments cloned in yeast artificial chromosomes. *Proc Natl Acad Sci USA* **90**, 6047–6051.
- Babitzke, P., Gollnick, P. & Yanofsky, C. (1992). The *mtrAB* operon of *Bacillus subtilis* encodes GTP cyclohydrolase I (MtrA), an enzyme involved in folic acid biosynthesis, and MtrB, a regulator of tryptophan biosynthesis. *J Bacteriol* **174**, 2059–2064.
- Barat, M., Anagnostopoulos, C. & Schneider, A.-M. (1965). Linkage relationships of genes controlling isoleucine, valine, and leucine biosynthesis in *Bacillus subtilis*. *J Bacteriol* **90**, 357–369.
- Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and yield from λ bacteriophage templates. *Proc Natl Acad Sci USA* **91**, 2216–2220.
- Bensing, B. A. & Dunny, G. M. (1993). Cloning and molecular analysis of gene affecting expression of binding substance, the recipient-encoded receptor(s) mediated mating aggregate formation in *Enterococcus faecalis*. *J Bacteriol* **175**, 7421–7429.
- Bower, S., Perkins, J., Yocum, R. R., Serró, P., Sorokin, A., Rahaim, P., Howitt, C. L., Prasad, N., Ehrlich, S. D. & Pero, J. (1995). Cloning and characterization of the *Bacillus subtilis birA* gene encoding a repressor of the biotin operon. *J Bacteriol* **177**, 2572–2575.
- Bruand, C., Sorokin, A., Serró, P. & Ehrlich, S. D. (1995). Nucleotide sequence of the *Bacillus subtilis dnaD* gene. *Microbiology* **141**, 321–322.
- Buchanan, C. E. & Ling, M.-L. (1992). Isolation and sequence analysis of *dacB*, which encodes a sporulation-specific penicillin-binding protein in *Bacillus subtilis*. *J Bacteriol* **174**, 1717–1725.
- Calogero, S., Gardan, R., Glaser, P., Schweizer, J., Rapoport, G. & Débarbouillé, M. (1994). RocR, a novel regulator controlling arginine utilization in *Bacillus subtilis*, belongs to NtrC/NifA family of transcription activators. *J Bacteriol* **176**, 1234–1241.

- Cheng, S., Fockler, C., Barnes, W. M. & Higuchi, R. (1994). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc Natl Acad Sci USA* **91**, 5695–5699.
- Dear, S. & Staden, R. (1991). A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res* **19**, 3907–3911.
- Del Castillo, I., Gonzles-Pastor, J. E., San Milln & Moreno, F. (1991). Nucleotide sequence of the *Escherichia coli* regulatory gene *mprA* and construction and characterisation of *mprA* deficient mutants. *J Bacteriol* **173**, 3924–3929.
- van Dijl, J. M., de Jong, A., Vehmaanper, J., Venema, G. & Bron, S. (1992). Signal peptidase I of *Bacillus subtilis*: patterns of conserved amino acids in prokaryotic and eukaryotic type I signal peptidases. *EMBO J* **11**, 2819–2828.
- Donovan, W., Zheng, L., Sandman, K. & Losick, R. (1987). Genes encoding spore coat polypeptides from *Bacillus subtilis*. *J Mol Biol* **196**, 1–10.
- Driks, A., Roels, S., Beall, B., Moran, C. P., Jr & Losick, R. (1994). Subcellular localization of proteins involved in the assembly of the spore coat of *Bacillus subtilis*. *Genes Dev* **8**, 234–244.
- Driver, R. P. & Lawther, R. P. (1985). Restriction endonuclease analysis of *ihvGEDA* operon of members of the family *Enterobacteriaceae*. *J Bacteriol* **162**, 1317–1319.
- Drummond, A. M., Whitty, P. & Wootton, J. (1986). Sequence and domain relationships of *ntrC* and *nifA* from *Klebsiella pneumoniae*: homologies to other regulatory proteins. *EMBO J* **5**, 441–447.
- Dujon, B. and others (1994). Complete DNA sequence of yeast chromosome XI. *Nature* **369**, 371–378.
- Errington, J. (1986). A general method for fusion of the *Escherichia coli lacZ* gene to chromosomal genes in *Bacillus subtilis*. *J Gen Microbiol* **132**, 2953–2961.
- Fleischmann, R. D. and others (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
- Fraser, C. M. and others (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science*, **270**, 397–403.
- Godon, J. J., Chopin, M. C. & Ehrlich, S. D. (1992). Branched-chain amino acid biosynthesis genes in *Lactococcus lactis*. *J Bacteriol* **174**, 6580–6589.
- Gray, J. V., Golinelli-Pimpaneau, G. & Knowles, J. R. (1990). Monofunctional chorismate mutase from *Bacillus subtilis*: purification of the protein, molecular cloning of the gene, and overexpression of the gene product in *Escherichia coli*. *Biochemistry* **29**, 376–383.
- Henner, D. J., Band, L. & Shimotsu, H. (1984). Nucleotide sequence of the *Bacillus subtilis* tryptophan operon. *Gene* **34**, 169–177.
- Henner, D. J., Band, L., Flaggs, G. & Chen, E. (1986). The organization and nucleotide sequence of the *Bacillus subtilis* *hisH*, *tyrA* and *aroE* genes. *Gene* **49**, 147–152.
- Herrler, M., Bang, H. & Marahiel, M. A. (1994). Cloning and characterization of *ppiB*, a *Bacillus subtilis* gene which encodes a cyclosporin A-sensitive peptidyl-prolyl *cis-trans* isomerase. *Mol Microbiol* **11**, 1073–1083.
- Hoch, J. & Mathews, J. (1972). Genetic studies in *Bacillus subtilis*. In *Spores V*, pp. 113–116. Edited by H. O. Halvorson, R. Hanson & L. L. Campbell. Washington, DC: American Society for Microbiology.
- Holland, D., Ben-Hayyim, G., Foltin, Z., Camoin, L., Strosberg, A. D. & Eshdat, Y. (1993). Molecular organisation of salt-stress associated protein in citrons: protein and cDNA peroxidase. *Plant Mol Biol* **21**, 923–927.
- Itaya, M. & Kondo, K. (1991). Molecular cloning of a ribonuclease H (RNase HI) gene from an extreme thermophile *Thermus thermophilus* HB8: a thermostable RNase HI can functionally replace the *Escherichia coli* enzyme *in vivo*. *Nucleic Acids Res* **16**, 4443–4449.
- Iwakura, M., Kawata, M., Tsuda, K. & Tanaka, T. (1988). Nucleotide sequences of the thymidilate synthase B and dihydrofolate reductase genes contained in one *Bacillus subtilis* operon. *Gene* **64**, 9–20.
- Leavitt, M. C. & Ito, J. (1989). T5 DNA polymerase: structural-functional relationships to other DNA polymerases. *Proc Natl Acad Sci USA* **86**, 4465–4469.
- Ludwig, H. C., Lottspeich, F., Henschen, A. & Bacher, A. (1987). Heavy riboflavin synthase of *Bacillus subtilis*: primary structure of the β subunit. *J Biol Chem* **262**, 1016–1021.
- McLaughlin, J. R., Murray, C. L. & Rabinowitz, J. C. (1981). Unique features in the ribosome-binding site sequence of the Gram-positive *Staphylococcus aureus* β -lactamase gene. *J Biol Chem* **256**, 11183–11291.
- Mares, R., Urbanowski, M. L. & Stauffer, G. V. (1992). Regulation of the *Salmonella typhimurium metA* gene by the MetR protein and homocysteine. *J Bacteriol* **174**, 390–397.
- Marklund, B. I., Tennent, J. M., Garcia, E., Hamers, A., Baga M., Lindberg, F., Gaastra, W. & Normark, S. (1992). Horizontal gene transfer of the *Escherichia coli* *pap* and *prs* pili operon as a mechanism for the development of tissue-specific adhesive properties. *Mol Microbiol* **6**, 2225–2245.
- Martin-Verstaete, I., Dbarbouill, M., Klier, A. & Rapoport, G. (1994). Interactions of wild-type and truncated LevR of *Bacillus subtilis* with the upstream activating sequence of the levanase operon. *J Mol Biol* **241**, 178–192.
- Messing, J. & Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**, 269–276.
- Micka, B. & Marahiel, M. A. (1992). The DNA-binding protein HBsu is essential for normal growth and development in *Bacillus subtilis*. *Biochimie* **74**, 641–650.
- Mironov, V. N., Kraev, A. S., Chikindas, M. L., Chernov, B. K., Stepanov, A. I. & Skryabin, K. G. (1994). Functional organization of the riboflavin biosynthesis operon from *Bacillus subtilis* SHgw. *Mol Gen Genet* **242**, 201–208.
- Morbidoni, H. R., de Mendoza, D. & Cronan, J. E., Jr (1995). Synthesis of *sn*-glycerol 3-phosphate, a key precursor of membrane lipids in *Bacillus subtilis*. *J Bacteriol* **177**, 5899–5905.
- Okasaki, R. & Kornberg, A. (1964). Enzymatic synthesis of deoxyribonucleic acid. XV. Purification and properties of a polymerase from *Bacillus subtilis*. *J Biol Chem* **239**, 259–268.
- Oliver, S. G. and others (1992). The complete DNA sequence of yeast chromosome III. *Nature* **357**, 38–46.
- Ordal, G. W., Parker, H. M. & Kirby, J. R. (1985). Complementation and characterization of chemotaxis mutants of *Bacillus subtilis*. *J Bacteriol* **164**, 802–810.
- Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.
- Phillips, D. A. & Kapulnik, Y. (1995). Plant isoflavonoid, pathogens and symbionts. *Trends Microbiol* **2**, 58–64.
- Popham, D. L. & Setlow, P. (1995). Cloning, nucleotide sequence and mutagenesis of the *Bacillus subtilis* *ponA* operon, which codes for penicillin-binding protein (PBP) 1 and a PBP-related factor. *J Bacteriol* **177**, 326–335.
- Popham, D. L., Szeto, D., Keener, J. & Kustu, S. (1989). Function of

a bacterial activator protein that binds to the transcriptional enhancers. *Science* **243**, 629–635.

Popham, D. L., Illades-Aniar, B. & Setlow, P. (1995). The *Bacillus subtilis* *dacB* gene, encoding penicillin-binding protein 5, is part of a three-gene operon required for proper spore cortex synthesis and spore core dehydration. *J Bacteriol* **177**, 4721–4729.

Reich, C., Gardiner, K. J., Olsen, G. J., Pace, B., Marsh, T. L. & Pace, N. R. (1986). The RNA component of the *Bacillus subtilis* RNase P. Sequence, activity and partial secondary structure. *J Biol Chem* **261**, 7888–7893.

Renault, P., Godon, J. J., Goupil, N., Delorme, C. & Ehrlich, S. D. (1995). Metabolic operon in lactococci. *Dev Biol Stand* **85**, 431–441.

Ryder, T. B., Hedrick, S. A., Belt, J. N., Liang, X., Clouse, S. D. & Lamb, C. J. (1987). Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. *Mol Gen Genet* **210**, 219–233.

Saxild, H. H. & Nygaard, P. (1987). Genetic and physiological characterization of *Bacillus subtilis* mutants resistant to purine analogs. *J Bacteriol* **169**, 2977–2983.

Sayers, J. R. & Eckstein, F. (1991). A single-strand specific endonuclease activity copurifies with overexpressed T5 D15 exonuclease. *Nucleic Acids Res* **19**, 4127–4132.

Schindelin, H., Marahiel, M. A. & Heinemann, U. (1993). Universal nucleic acid binding domain revealed by crystal structure of the *Bacillus subtilis* major cold shock protein. *Nature* **364**, 164–168.

Schnuchel, A., Wilschek, R., Czisch, M., Herrier, M., Willinsky, G., Graumann, P., Marahiel, M. A. & Holak, T. A. (1993). Structure in solution of the major cold shock protein from *Bacillus subtilis*. *Nature* **364**, 169–171.

Schott, K., Kellermann, J., Lottspeich, F. & Bacher, A. (1990). Riboflavin synthases of *Bacillus subtilis*: purification and amino acid sequence of the α subunit. *J Biol Chem* **265**, 4204–4209.

Schröder, K., Graumann, P., Schnuchel, A., Holak, T. A. & Marahiel, M. A. (1995). Mutational analysis of the putative nucleic acid-binding surface of the cold-shock domain, CspB, revealed an essential role of aromatic and basic residues in binding of the single-stranded DNA containing the Y-box motif. *Mol Microbiol* **16**, 699–708.

Sorokin, A., Zumstein, E., Azevedo, V., Ehrlich S. D. & Serror, P. (1993). The organization of the *Bacillus subtilis* 168 chromosome region between the *spoVA* and *serA* genetic loci, based on sequence data. *Mol Microbiol* **10**, 385–395.

Sorokin, A., Serror, P., Pujic, P., Azevedo, V. & Ehrlich, S. D. (1995). The *Bacillus subtilis* chromosome region encoding homologues of

the *Escherichia coli* *mssA* and *rpsA* gene products. *Microbiology* **141**, 311–319.

Sorokin, A., Azevedo, V., Zumstein, E., Galleron, N., Ehrlich, S. D. & Serror, P. (1996a). Sequence analysis of the *Bacillus subtilis* chromosome region between the *serA* and *kdg* loci cloned in a yeast artificial chromosome. *Microbiology* **142**, 2005–2016.

Sorokin, A., Lapidus, L., Capuano, V., Galleron, N., Pujic, P. & Ehrlich, S. D. (1996b). A new approach using Multiplex Long Accurate PCR and yeast artificial chromosomes for bacterial chromosome mapping and sequencing. *Genome Res* **6**, 448–453.

Sun, G., Sharkova, E., Chesnut, R., Birkey, S., Duggan, M. F., Sorokin, A., Pujic, P., Ehrlich, S. D. & Hulett, M. F. (1996). Regulators of aerobic and anaerobic respiration in *Bacillus subtilis*. *J Bacteriol* **178**, 1374–1385.

Taha, M. K., So, M., Seifert, H. S., Billyard, E. & Marchal, C. (1988). Pilin expression in *Neisseria gonorrhoeae* is under both positive and negative transcriptional control. *EMBO J* **7**, 4367–4378.

Tanaka, T., Kawata, M., Nagami, Y. & Uchiyama, H. (1987). *prtR* enhances the mRNA level of the *Bacillus subtilis* extracellular proteases. *J Bacteriol* **169**, 3044–3050.

Willinsky, G., Bang, H., Fisher, G. & Marahiel, M. A. (1992). Characterization of *cspB*, a *Bacillus subtilis* inducible cold shock gene affecting cell viability at low temperature. *J Bacteriol* **174**, 6326–6335.

Yamamoto, J., Shimizu, M. & Yamane, K. (1991). Molecular cloning and analysis of nucleotide sequence of the *Bacillus subtilis* *lysA* gene region using *B. subtilis* phage vectors and a multi-copy plasmid, pUB110. *Agric Biol Chem* **55**, 1615–1626.

Yang, M., Shimotsu, H., Ferrari, E. & Henner, D. J. (1987). Characterization and mapping of *Bacillus subtilis* *ptrR* gene. *J Bacteriol* **169**, 434–437.

Yazdi, M. A. & Moir, A. (1990). Characterization and cloning of the *gerC* locus of *Bacillus subtilis*. *J Gen Microbiol* **136**, 1335–1342.

Yu, J., Hederstedt, L. & Piggot, P. (1995). The cytochrome *bc* complex (menaquinone:cytochrome *c* reductase) in *Bacillus subtilis* has a non-traditional subunit organisation. *J Bacteriol* **177**, 6751–6760.

Zeigler, D. R. & Dean, D. H. (1990). Orientation of the genes in the *Bacillus subtilis* chromosome. *Genetics* **125**, 703–708.

Zimmermann, J., Voss, H., Kristensen, T., Schwager, C., Stegemann, J., Erfle, H. & Ansorge, W. (1989). Automated preparation and purification of M13 templates for DNA sequencing. *Methods Mol Cell Biol* **1**, 29–34.

Received 18 April 1996; revised 11 June 1996; accepted 24 June 1996.