Sequencing and functional annotation of the *Bacillus subtilis* genes in the 200 kb *rrnB*–*dnaB* region

Alla Lapidus, Nathalie Galleron, 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 Genetique Microbienne, Institut National de la Recherche Agronomique, Domaine de Vilvert, 78352 Jouy-en-Josas cedex, France

The 200 kb region of the *Bacillus subtilis* chromosome spanning from 255 to 275° on the genetic map was sequenced. The strategy applied, based on use of yeast artificial chromosomes and multiplex Long Accurate PCR, proved to be very efficient for sequencing a large bacterial chromosome area. A total of 193 genes of this part of the chromosome was classified by level of knowledge and biological category of their functions. Five levels of gene function understanding are defined. These are: (i) experimental evidence is available of gene product or biological function; (ii) strong homology exists for the putative gene product with proteins from other organisms; (iii) some indication of the function can be derived from homologies with known proteins; (iv) the gene product can be clustered with hypothetical proteins; (v) no indication on the gene function exists. The percentage of detected genes in each category was: 20, 28, 20, 15 and 17, respectively. In the sequenced region, a high percentage of genes are implicated in transport and metabolic linking of glycolysis and the citric acid cycle. A functional connection of several genes from this region and the genes close to 140° in the chromosome was also observed.

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

INTRODUCTION

The chromosome region between *rrnB* and *dnaB* genetic locus was assigned to us for sequencing in the European *Bacillus subtilis* genome sequencing project (Kunst *et al.*, 1995). Our approach to the sequencing of this genome is essentially based on using a yeast artificial chromosome (YAC) collection of ordered chromosomal segments (Azevedo *et al.*, 1993). Sequencing of two YACs covering about 140 kb of contiguous area was recently reported (Sorokin *et al.*, 1996a; Capuano *et al.*, 1996). This paper reports further development of this approach and its application to another two YACs, containing in total about 200 kb of new sequence. We report functional annotation of genes found according to the results of homology search in databases.

Abbreviations: LR PCR, Long Range PCR; MLA PCR, multiplex Long Accurate PCR; YAC, yeast artificial chromosome.

The GenBank accession number for the sequence reported in this paper is AF008220.

METHODS

Strains and growth conditions. Collection of yeast clones containing ordered segments of the *B. subtilis* genome cloned in YAC was described earlier (Azevedo *et al.*, 1993). Yeast cells containing artificial chromosomes were grown as described (Azevedo *et al.*, 1993). *Escherichia coli* JJC 128F', *araD139* Δ (*ara-leu*)7696 *galE15 galK16* Δ (*lac*)X74 *hsdR hsdM*⁺ Str^R F'[*lacI*^q D(*lacZ*)M15 *traD36*] which is reproducibly electrotransformed by M13 DNA with an efficiency of 5×10^8 p.f.u. (μ g DNA)⁻¹, was used in YAC DNA cloning experiments. Electrocompetent cells of JJC128F' were prepared according to the protocol described by Dower *et al.* (1988) and stored at -80 °C. *E. coli* TG1 K-12 Δ (*lac-pro*) *supE thi hsdR/F' traD36 proAB lacI*^q Δ *lacZ* Δ M15 was used for propagation of M13 phages and for plasmid cloning experiments. Competent cells were prepared by 0.1 M CaCl₂ treatment of early exponential culture. *B. subtilis* 168 *trpC2* strain was provided by C. Anagnostopoulos (INRA, Jouy-en-Josas, France). This strain, considered as standard for systematic *B. subtilis* genome sequencing, was used for chromosomal DNA preparations. The standard medium for *E. coli* and *B. subtilis* was 2YT (Sambrook *et al.*, 1989).

Isolation of bacterial chromosomal DNA. *B. subtilis* chromo-

somal DNA was isolated as described by Sorokin *et al.* (1996b). Twenty-five millilitres overnight cultures in 2YT were harvested and treated with 10 mg lysozyme ml⁻¹ in 50 mM Tris/HCl, pH 8.0, 50 mM EDTA, 25% sucrose. SDS was added to a final concentration of 0.5% and proteinase K to 100 µg ml⁻¹. Incubation was carried out at 50 °C for 4 h and extraction was done twice with water-saturated phenol/chloroform (1:1, v/v), pH 8.0. After precipitation with 2 vols ethanol in 0.3 M NaOAc, pH 4.8, the DNA was removed with a glass rod and washed in 70% ethanol. The DNA was stored in water at 100 µg ml⁻¹.

Isolation of YAC DNA. Plugs containing yeast chromosomes were prepared as described by Anand *et al.* (1989). Electrophoresis was performed in 1.2% agarose gels in 0.3 × TBE at 10 V cm⁻¹ at 18 °C (1 × TBE = 90 mM Tris/borate, 2 mM EDTA). The switching interval was from 0.3 to 6 s in forward migration and 0.1 to 2 s in reverse migration for 20 h. After separation, YAC DNA was electroeluted in a dialysis membrane (Amicon) for 4 h at 10 V cm⁻¹. The solution was concentrated fivefold by butanol treatment and the DNA was precipitated with ethanol. Phenol/chloroform treatment was usually necessary to make DNA digestible by restriction enzymes. This protocol allowed us to obtain about 1–10 µg YAC DNA from 500 ml yeast cell culture. The contamination by yeast chromosomal DNA was estimated to be 20–50%.

Construction of M13 bank. DNA (0.5 to 10 µg) was partially digested by one of the restriction enzymes *AluI*, *TaqI*, *HpaII*, *HinPI* or *Sau3A*. The digested DNA was separated in agarose gel and segments between 500 and 1500 bp were purified as described by Sorokin *et al.* (1993). The purified DNA was ligated with 50 ng *SmaI*, *AccI* or *BamHI* cut and dephosphorylated M13mp19 vector in 20 µl ligase buffer (Boehringer), precipitated by 2-propanol using tRNA or glycogen as a carrier, rinsed with 70% ethanol, dried, dissolved in deionized water and used for electroporation. The yield of phage plaques was 10⁶–10⁷ (µg M13 DNA)⁻¹, with a ratio of white to blue clones from 1 to 10.

Selection of M13 clones for random sequencing. White M13 plaques were propagated for 5–15 h in 48-well plates in 2 × YT medium on TG1 cells as a host. After centrifugation in a GPKR centrifuge (Beckman), 150 µl phage supernatants were distributed in 96-well plates using a BIOMECH 1000 Laboratory workstation (Beckman). Fifty microlitre aliquots of the supernatants were used to prepare filters corresponding to the stock plates. PFGE-purified YAC DNA was ³²P-labelled (Random-Prime Labelling Kit; Boehringer) and then hybridized with the phage DNA on the filters.

Long-Range PCR fragment cloning. Cloning of LR PCR products (5–20 kb) was performed in M13 phages or pSGMU2 plasmid (Errington, 1986) as described previously (Capuano *et al.*, 1996). A bank of plasmids with β-galactosidase fusions was also constructed by using a plasmid pDT1. This plasmid was constructed by insertion of a fragment containing '–21' standard direct primer sequence and a *B. subtilis* ribosome-binding site sequence (...ttAGGAGGAtaaataATG...) into the plasmid pDIA5303 described previously (Sorokin *et al.*, 1995). The plasmid pDT1 allows construction of transcriptional fusions with β-galactosidase gene as a reporter. At the same time this plasmid allows selection of promoter-containing fragments in *E. coli*. To construct a bank of fusion plasmids, LR PCR products of 3 kb corresponding to regions of interest were treated by exonuclease *Bal31* (New England Biolabs) and Klenow polymerase (Boehringer), and were then ligated with *SmaI*-cut and dephosphorylated pDT1. After

Fig. 1. PCR map of the *rrnB*–*dnaB* region of the *B. subtilis* chromosome. Only the minimal map described in the text is shown. Known sequences longer than 1 kb are shown by hatched bars. Corresponding GenBank accession numbers and genetic loci names are shown above the bars. White rectangles correspond to LR PCR fragments synthesized by using appropriate primers, designated by small black arrows. Inserts in the YAC clones used to construct the map are shown below the scale line. Scale is in kb corresponding to GenBank accession number AF008220.

transformation into *E. coli* TG1 cells, blue colonies were selected on plates containing 50 µg ampicillin ml⁻¹ and 25 µg X-Gal ml⁻¹. Plasmid DNA was prepared from these cells as described by Sorokin *et al.* (1995) and used for sequencing.

Sequencing. ssDNA of M13 phages was prepared as described earlier (Sorokin *et al.*, 1993). For the reverse sequencing ds DNA of the inserts was prepared by PCR using primers: 5'-GTTTTTCCAGTCACGAC3' and 5'-GAGCGGATAACA-ATTTCAC3'. Plasmid DNA for sequencing was prepared as described (Sorokin *et al.*, 1995). PCR products used for sequencing with dye terminators were purified by the Wizard PCR* Preps kit (Promega) or agarose gel electrophoresis. Direct and reverse PCR sequencing was performed using ABI PRISM direct, reverse or Dye Terminator Sequencing Kit (Applied Biosystems) on the Perkin Elmer 9600 thermal cycler or the Catalyst station.

Oligonucleotide synthesis and standard PCR. Oligonucleotides were synthesized in the DNA synthesizer 'Oligo 1000' (Beckman). Standard PCR was performed by using M13 DNA or supernatants and *B. subtilis* or yeast chromosomal DNA in the conditions previously described (Sorokin *et al.*, 1993). Primers used for LR PCR were 20–22mers, chosen to contain 12 GC bases. Most of the sequences of primers, shown on Fig. 1, were presented previously (Sorokin *et al.*, 1996b). The others are: AG8, 5'-GGTGC GGACTTTGAATCTCGC3'; AH1, 5'-GAGCGATTCTCGTGCTGAAGG3'; AH2, 5'-

TGACGGGAACCTCTGTCGGAA3'; AH8, 5'ATGCCT-TGCGTTGCGGCATAG3'; BA1, 5'TTTCCTGACAGGC-CCGTCTTG3'; BA2, 5'CCTGAAATGATCGTCCACCGC3'; BB1, 5'ACATACCGATGCCATCGCCCA3'; BB2, 5'-TGTTCTCTATCCGCACAGCC3'; CH4, 5'GATATAC-CGTGCGGTCAAGCAG3'; CH5, 5'ACTCATCGGCATGATCAGCCG3'.

Multiplex Long Accurate PCR (MLA PCR). The conditions for the reaction were as described by Sorokin *et al.* (1996b): 20 mM Tricine, pH 8.7; 85 mM KOAc; 1 mM Mg(OAc)₂; 8% glycerol; 2% DMSO; 0.2 mM each dNTP; 0.2 µM each primer; 0.1 µg *B. subtilis* chromosomal DNA; 2 U rTth (Perkin Elmer); 0.05 U Vent polymerase (New England Biolabs). The final reaction volume was 50 µl. Cycling conditions: 94 °C, 5 min; 12 cycles of 10 s melting at 94 °C and 12 min annealing-polymerization-repair at 68 °C followed by 24 cycles with increasing the extension time 15 s for each cycle.

Computing. The program XBAP version 9.0 was used for gel assembling and consensus sequence generating (Dear & Staden, 1991). XNIP was used for sequence interpretation. Sequence homologies were searched using FASTA (Pearson & Lipman, 1988), contained in the GCG package, version 8.0, or BLAST (Altschul *et al.*, 1990) realized on the NCBI e-mail server (URL: <http://www.ncbi.nlm.nih.gov/BLAST/>).

Nomenclature. The nomenclature for putative ORFs is coordinated with the *B. subtilis* genome project as described by Sorokin *et al.* (1996a). Finally it was adopted to exactly coincide with gene names in the *SubtiList* database (Moszer *et al.*, 1995).

RESULTS AND DISCUSSION

Construction of a minimal PCR map of the *rrnB-dnaB* region of the *B. subtilis* chromosome

The physical map of the *rrnB-dnaB* region, relevant to our sequencing strategy, is shown on Fig. 1. Most of the region appeared to be cloned in two YACs, 15-132 and 11-105 (Azevedo *et al.*, 1993). Several sequences of this region were reported previously and are available from GenBank under the accession numbers: M10606 for the *rrnB* operon (Green *et al.*, 1985), Z25795 for the *glg* operon (Kiel *et al.*, 1994), M74183 for the *men* operon (Driscoll & Taber, 1992; Rowland *et al.*, 1995), L31845 for UDP-N-acetyl muramate-alanine ligase, X65945 for *aroA* (Bolotin *et al.*, 1995), M34719 for *ccpA* (Henkin *et al.*, 1991), L17309 for the *acu* operon (Grundy *et al.*, 1993b), M77668 for *tyrS* (Henkin *et al.*, 1992), M34718 for *rpsD* (Grundy & Henkin, 1990), L17320 for *ackA* (Grundy *et al.*, 1993a), M12620 for *sspA* (Connors *et al.*, 1986), U05257 for *citZCH* (Jin & Sonenshein, 1994), M23549 for *phoPR* (Seki *et al.*, 1988) and X04963 for *dnaB* (Ogasawara *et al.*, 1986). YAC15-132 carries an insert of ~ 175 kb as estimated by PFGE (Azevedo *et al.*, 1993). DNA from this YAC was purified, cloned in phage M13 and 500 clones were sequenced randomly. This yielded 42 representative contigs containing each at least two sequences from independent phage clones. We chose 32 representative contigs and applied the (MLA PCR) mapping strategy described previously (Sorokin *et al.*, 1996b). The PCR map (shown in Fig. 1) extending from Z25795 (*glg* operon) to L17320 (*ackA* gene), coordinates 15–160 kb, was thus constructed.

The segment extending between *citZCH* and *dnaB* was amplified by LR PCR using primers AB4 and AH6 which correspond to sequences with GenBank accession numbers U05257 and X04963 (Fig. 1, coordinates 200–229 kb). Partial sequencing of the insert in the plasmid pMP78-32, containing a part of the *polA* gene (Perego *et al.*, 1987), was used to generate sequenced tags and primers AB7 and AB8, corresponding to the *B. subtilis* *polA* gene.

The segment between the last YAC15-132 marker *ackA* and *citZCH* could not be amplified by LR PCR. However, it was covered by YAC11-105, since the DNA (PCR fragments or phage clones) of all the contigs mapped between primers AF3 and AB8 (Fig. 1, coordinates 125 and 205 kb) hybridized with a probe containing this YAC (not shown). We used YAC11-105 DNA, in the same way as 15-132, to generate sequenced tags within the gap. This allowed completion of the PCR map of the region shown on Fig. 1.

A map scale can be defined as the size of the mapped region, in kb, divided by the number of mapped contigs. A minimal map is then the one with the biggest scale. The scale of the YAC15-132 complete map (not shown) was 7.6 kb (175 kb/23 mapped tags) and that of the minimal map 13.4 kb (175 kb/13). The map of the lower scale (and eventually the nucleotide sequence) can be used for confirmation of those of the bigger scale. We found no discrepancies between the maps of different scales, which indicates that the approach used is reliable. This conclusion was confirmed independently by another approach. A 150 kb *spoIIIC-pheA* region of the *B. subtilis* chromosome constructed in a similar way was tested by comparing restriction patterns of LR PCR fragments and chromosomal DNA (Bolotin *et al.*, 1996). No discrepancy was found.

Sequencing of the *rrnB-dnaB* region

Sequencing strategy was based on using the LR PCR fragments shown in Fig. 1. The level of errors due to LR PCR was estimated previously to be about 10⁻⁴, by comparing sequences of M13 clones, obtained either by cloning YAC DNA or LR PCR products (Capuano *et al.*, 1996). We estimated the mutagenesis rate of LR PCR during sequencing of YAC15-132, by determining the number of mutations in the 67600 bp area between primers AG2-AC8 and AF8-AG8 (Fig. 1). The correct sequence was considered to be that occurring in the majority of clones. The mutation rate appeared to be 1 in 7000 bp (Sorokin *et al.*, 1996b). It can therefore be concluded that the error rate due to LR PCR mutagenesis is negligible compared to that of a standard sequence reaction, that is, usually, 1 in <400 bp for the most reliable parts of gel readings.

Two approaches for sequencing of the 5–18 kb LR PCR fragments were used. In the first, several fragments were mixed in equimolar quantities, their ends were randomized by restriction enzymes or Bal31 exonuclease, and cloned in plasmids or M13 phage as described in Methods. Phage or plasmid DNA from the clones was

Table 1. Experimentally defined gene functions in the *rrnB*–*dnaB* region of the *B. subtilis* chromosome

Genes are listed in the order of their position on the map in Fig. 2. Only GenBank numbers for the ORFs which were not resequenced are given. Endpoints are those in GenBank entry AF008220. Putative start codons are indicated in bold. The most direct evidence of the product activity or biological function of the gene is indicated. Abbreviations: SHCHC, 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate; OSB, O-succinylbenzoate; PEP, phosphoenolpyruvate; SAM, S-adenosyl-methionine; DAPA, 7,8-diaminopelargonic acid; 7-KAP, 7-keto-8-aminopelargonic acid; DAHP, 3-deoxy-D-arabino-heptulosonate-7-phosphate.

Gene symbol	GenBank entry	Endpoints (bp)	Size (aa)	Translation start	Gene product function	Biological function or gene category	Reference
<i>glgB</i>	Z25795	7473–9356	627	atgtcagaaaggatgattacATG	1,4- α -Glucan branching enzyme	Glycogen biosynthesis	Kiel <i>et al.</i> (1994)
<i>glgC</i>	Z25795	9353–10495	380	aagaggagagataaaagATG	ADPglucose pyrophosphorylase	Glycogen biosynthesis	Kiel <i>et al.</i> (1994)
<i>glgD</i>	Z25795	10519–11550	343	aagaggagcagagtcATG	Unknown	Glycogen biosynthesis	Kiel <i>et al.</i> (1994)
<i>glgA</i>	Z25795	11547–13001	484	aagaggagctgagtaacTTG	Starch synthase	Glycogen biosynthesis	Kiel <i>et al.</i> (1994)
<i>glgP</i>	Z25795	12988–15384	798	acaaggagrtggacgggATG	Glycogen phosphorylase	Glycogen biosynthesis	Kiel <i>et al.</i> (1994)
<i>cotS</i>	–	18373–19428	352	atagaagggtgaacagaATG	Spore-coat protein S	Spore maturation	Abe <i>et al.</i> (1995)
<i>menF</i>	–	25400–26815	472	attatggaggaatgtctcATG	Isochorismate synthase	Menaquinone biosynthesis	Rowland & Taber (1996)
<i>menD</i>	–	26812–28554	581	cattaggaggtgagggcgTTG	SHCHC synthase	Menaquinone biosynthesis	Palaniappan <i>et al.</i> (1994)
<i>menB</i>	–	29401–30216	272	ataaaggagggtcatcatATG	Dihydroxynaphthoate synthase	Menaquinone	Driscoll & Taber (1992)
<i>menE</i>	–	30307–31767	487	acctcggaagatgatcATG	OSB-CoA synthase	Menaquinone biosynthesis	Driscoll & Taber (1992)
<i>metE</i>	–	50089–51291	401	tgcaggaggaagcaaatcATG	SAM synthetase	Methyl transfer	Wabiko <i>et al.</i> (1988)
<i>leuS</i>	–	74073–76487	805	atcagaaggaggtttttacTTG	Leucine-tRNA synthetase	Protein biosynthesis	Vander Horn & Zahler (1992)
<i>bioW</i>	–	83703–84437	245	tcagaatgagggtctcaATG	Pimelyl-CoA synthetase	Biotin biosynthesis	Bower <i>et al.</i> (1996)
<i>bioA</i>	–	84427–85773	449	agagrtggagggaagatcATG	DAPA aminotransferase	Biotin biosynthesis	Bower <i>et al.</i> (1996)
<i>bioF</i>	–	85763–86932	390	ccacgaggttacgagccTTG	7-KAP synthetase	Biotin biosynthesis	Bower <i>et al.</i> (1996)
<i>bioD</i>	–	86944–87624	232	tcatttggagggttttttGTG	Dethiobiotin synthetase	Biotin biosynthesis	Bower <i>et al.</i> (1996)
<i>bioB</i>	–	87627–88634	336	atcaagrtgggggtatgagaATG	Biotin synthetase	Biotin biosynthesis	Bower <i>et al.</i> (1996)
<i>bioI</i>	–	88703–89890	396	cagaaggaggaataacacATG	Unknown	Biotin biosynthesis	Bower <i>et al.</i> (1996)
<i>opuD</i>	–	100760–102298c	513	aaaaattagagggtgagaaATG	Glycine betaine transporter	Osmoprotection	Kappes <i>et al.</i> (1996)
<i>aroA</i>	X65945	132601–133677	358	gggaaaggatgaaaaaATG	DAHP synthase-chorismate mutase	Aromatic amino acid biosynthesis	Bolotin <i>et al.</i> (1995)
<i>ccpA</i>	M34719	133953–134957	334	gtaaaaggagtgttttaggATG	LacI family transcription regulator	Carbon metabolism regulation	Henkin <i>et al.</i> (1991)
<i>acuA</i>	L17309	138398–139030c	210	gatgggacgggtgaattaGTG	Unknown	Acetoin catabolism	Grundy <i>et al.</i> (1993b)
<i>acuB</i>	L17309	137727–138371c	214	gacaaaagggaataaataATG	Unknown	Acetoin catabolism	Grundy <i>et al.</i> (1993b)
<i>acuC</i>	L17309	136567–137730c	387	agagcaaaaggatctgctATG	Unknown	Acetoin catabolism	Grundy <i>et al.</i> (1993b)
<i>acsA</i>	L17309	139191–140909	572	caaaggggagatgtttggcATG	Acetyl-CoA synthetase	Acetate metabolism	Grundy <i>et al.</i> (1993b)
<i>tyrS</i>	M77668	141251–142519	422	ttaaaaaaggagctttatcttATG	Tyrosine-tRNA synthetase	Protein biosynthesis	Henkin <i>et al.</i> (1992)
<i>rpsD</i>	M34718	142790–143392c	200	ccaaggaggagtcacattATG	Ribosomal protein S4	Protein biosynthesis	Grundy & Henkin (1990)
<i>sspA</i>	–	153471–153680	70	cacaaggaggtagacacATG	DNA-binding protein SspA	Protection of DNA in the spore	Tovar-Rojo & Setlow (1991)
<i>ackA</i>	–	162827–164014	396	agtataaaggcgtatcATG	Acetate kinase	Acetate metabolism	Grundy <i>et al.</i> (1993a)
<i>argG</i>	–	164781–165992	404	ttgagagggtgattttatcATG	Arginino succinate synthase	Arginine biosynthesis	Mountain <i>et al.</i> (1986)
<i>argH</i>	–	165989–167374	462	gaaggagcagattaaagcATG	Arginino succinate lyase	Arginine biosynthesis	Mountain <i>et al.</i> (1986)
<i>citZ</i>	–	196855–197973	373	aaaaggagatgttatatATG	Citrate synthase subunit II	Citric acid cycle	Jin & Sonenshein (1996)
<i>citC</i>	–	198137–199408	424	atactggagggttttttGTG	Isocitrate dehydrogenase	Citric acid cycle	Jin & Sonenshein (1994)
<i>citH</i>	–	199452–200390	313	ataaaggagaagagacATG	Malate dehydrogenase	Citric acid cycle	Jin <i>et al.</i> (1996)
<i>phoP</i>	–	200602–201324	241	tactggaggcacagcATG	Signal transduction regulator	Phosphate metabolism regulation	Hulett <i>et al.</i> (1994)
<i>phoR</i>	–	201317–203056	580	ctggaggagccaaaaatgaATG	Protein kinase	Phosphate metabolism regulation	Hulett <i>et al.</i> (1994)
<i>polA</i>	–	203300–205942	881	cttcttaggaggaattgaATG	DNA polymerase I	DNA replication and repair	Perego <i>et al.</i> (1987)
<i>dnaB</i>	X04963	213557–214975	472	ttcggaagggttgagaatgtATG	Unknown	DNA replication and repair	Ogasawara <i>et al.</i> (1986)
<i>dnaI</i>	X04963	215003–215938	311	tgcggggtaagggaatagATG	Unknown	DNA replication and repair	Bruand & Ehrlich (1995)

then sequenced by direct and reverse standard fluorescent sequencing primers. In the second approach the purified LR PCR products were sequenced directly by synthetic primers and fluorescent dideoxy nucleotides (the primer walking approach). We found the quality of the direct sequencing of the 5–18 kb PCR fragments was satisfactory only when F667Y modified *Taq* DNA polymerase was used (Tabor & Richardson, 1995). However, eventually, with a frequency of 1 in 2000 bp, there was a base call with primers corresponding to different DNA strands, although in both cases no

ambiguity was observed between single-strand readings. Cloned templates were used to establish the correct sequence. Sequencing by primer walking can be used, however, especially for DNA segments difficult to clone. In such cases the ambiguities were resolved by synthesizing two primers, each corresponding to one sequence, and using them for sequencing of LR PCR products, since only the correct primer yielded the readable sequence. We suggest that this may be a rather general way to establish the correct sequence in a number of cases. The reasons for the occurrence of

Table 2. Putative gene functions in the *rrnB-dnaB* region of the *B. subtilis* chromosomeSee the legend to Table 1. All ORFs were resequenced except *murC*, *ytxD* and *ytxE* (accession no. L31845).

ORF	Endpoints (bp)	Size (aa)	Translation start	PIR/SP entry	Description of PIR entry	FASTA score	Percentage identity (over aa)
<i>ytaA</i>	15962–17035c	358	taggagtaaaacaggaagATG	P46914	Spore coat protein S of <i>Bacillus subtilis</i>	433	25.9 (297)
<i>ytCB</i>	21161–22111c	317	tgtacgggagttggccgtcATG	S19784	dTDPglucose 4,6-dehydratase of <i>Streptomyces griseus</i>	371	29.0 (307)
<i>ytCA</i>	22108–23394c	419	gtaaaggagctcaggtgATG	JC2525	UDPglucose dehydrogenase of <i>Xanthomonas campestris</i>	534	36.7 (346)
<i>ytDA</i>	23566–24384	273	atgagctgaggggaaataccATG	a40650	UTPglucose-1-P uridylyltransferase of <i>Bacillus subtilis</i>	588	43.0 (272)
<i>ytFD</i>	31764–32879	372	gcgcgtaaaggagaaactgATG	a28630	Muconate cycloisomerase of <i>Pseudomonas putida</i>	260	24.7 (373)
<i>ytGA</i>	33159–34079	307	ggcgaaggaggagagaatATG	a43583	Adhesin B precursor of <i>Streptococcus sanguis</i>	414	27.4 (310)
<i>ytGB</i>	34098–34850	251	aaagaaaagggtggaggatATG	js0115	Iron (III) dicitrate transporter of <i>Escherichia coli</i>	315	32.0 (231)
<i>ytIB</i>	40233–40796	188	tagaagggaacaagaaATG	P17582	Carbonic anhydrase of <i>Escherichia coli</i>	174	25.0 (192)
<i>ytIC</i>	44948–45730c	261	aaataggaggcgccatcATG	s30894	Nitrate transport protein of <i>Synechococcus</i> sp.	417	36.2 (224)
<i>pkcA</i>	48003–49586c	528	aaaggaaggtttcatATG	P22259	Phosphoenolpyruvate carboxykinase of <i>Escherichia coli</i>	1622	45.0 (431)
<i>asnB</i>	51441–53339	633	tttaatttggaggcaatacATG	a36616	Asparagine synthase of <i>Escherichia coli</i>	390	23.1 (507)
<i>ytNA</i>	53475–54866	464	tttctaggggagagaagcATG	a39431	Phenylalanine transport protein of <i>Escherichia coli</i>	1334	43.6 (450)
<i>ytRA</i>	59877–60269	131	gaaagtaggggaagtaaATG	h31844	KorA protein of <i>Streptomyces griseus</i> plasmid pJJ101	141	28.9 (121)
				b36730	Repressor protein HutC of <i>Klebsiella pneumoniae</i>	137	32.6 (89)
<i>ytRB</i>	60262–61140	293	gtgaaggaggagcaaaaaATG	b45313	Putrescine transport protein PotG of <i>Escherichia coli</i>	313	32.3 (223)
<i>ytRE</i>	63142–63837	232	tgtgaaggaggagagaacATG	s03183	Glutamine transport protein GlnQ of <i>Escherichia coli</i>	388	43.9 (198)
				s03131	FtsE cell division protein of <i>Escherichia coli</i>	413	38.0 (221)
<i>ytSA</i>	65234–65929	232	aagtcggtgagtcgaTTG	a47210	Nitrate transduction regulator – chromatic adaptation protein RcaC of <i>Fremyella diplosiphon</i>	341	33.0 (224)
				a41965	Alkaline phosphatase regulator PhoR of <i>Escherichia coli</i>	284	30.8 (221)
<i>ytSB</i>	65922–66926	335	gaaggaagagacagttctATG	b42372	Signal sensor kinase KdpD of <i>Escherichia coli</i>	160	24.8 (250)
<i>ytSC</i>	67028–67789	254	tataaaggagactgctgaaATG	s03183	Glutamine transport protein GlnQ of <i>Escherichia coli</i>	338	31.6 (212)
<i>mleA</i>	76937–78235c	433	aaacattaaaggaggagcaatATG	a26571	α -Galactosidase <i>Escherichia coli</i>	1033	49.2 (441)
<i>amyC</i>	78255–79085c	277	aaagaaaggagacggaatgcATG	s37705	Maltose transport protein AmyC of <i>Thermoanaerobacterium thermosulfurigenes</i>	451	30.8 (266)
<i>amyD</i>	79082–79993c	304	cacaggagactgaaacaATG	s37704	Transport protein AmyD of <i>Thermoanaerobacterium thermosulfurigenes</i>	535	31.6 (288)
<i>msmE</i>	79986–81266c	427	acgatgataaggagaaaATG	b42400	Sugar-binding protein MsmE of <i>Streptococcus mutans</i>	249	22.1 (366)
<i>msmR</i>	81300–82334c	345	atagatgagaggagataccATG	s09205	LacI family transcription regulator EbgR of <i>Escherichia coli</i>	518	34.0 (335)
<i>ytCP</i>	90935–91795	287	gaggtaggagctatcgacattcaCTG	s03782	Glycerol-3-phosphate transport protein UgpE of <i>Escherichia coli</i>	262	22.2 (288)
<i>ytDQ</i>	93357–94091c	245	tgcggaaggctcagaagcTTG	a29625	Melibiose operon transcription regulatory protein of <i>Escherichia coli</i>	191	22.4 (143)
<i>ytbP</i>	107005–107715	237	ttttgagggtgagttTTG	s13584	Transport protein NosF of <i>Pseudomonas stutzeri</i>	324	35.4 (212)
<i>ytbP</i>	111728–112663c	312	tcagaaaggagaatgacaATG	syebac	Cysteine synthase of <i>Salmonella typhimurium</i>	547	45.8 (319)
<i>malS</i>	120565–122265	567	tcagaaaggagaaggcaggGTG	s55433	Malolactic enzyme homologue of <i>Bacillus subtilis</i>	1905	65 (563)
				s43057	Malolactic enzyme of <i>Lactococcus lactis</i>	1451	49.5 (546)
<i>ytbP</i>	124603–124926	108	aacaactggaggatattgaaATG	jq2242	Thioredoxin of <i>Arabidopsis thaliana</i>	186	31.0 (84)
<i>murC</i>	129638–130936	433	acgtttgttggaggtacaattATG	s40601	UDP-N-acetylmuramate-alanine ligase	337	29.9 (308)
<i>ytXD</i>	135020–135838	273	ttttctatggggagaatgATG	a42882	Motility protein A of <i>Bacillus subtilis</i>	498	31.2 (263)
<i>ytXE</i>	135828–136556	243	aaaaaaaggatctgttgcATG	b42882	Motility protein B of <i>Bacillus subtilis</i>	332	31.7 (183)
<i>ytbP</i>	146081–146704	208	agggaagggtgagccgATG	P43506	Transcriptional repressor BM3R1 of <i>Bacillus megaterium</i>	156	24.9 (169)
<i>braB</i>	149487–150824c	446	ttattaggaggattttttATG	s11497	Branched-chain amino acid transport protein BraB of <i>Pseudomonas aeruginosa</i>	1105	45.1 (432)
<i>nifS2</i>	151019–152164	382	cataaaggagtaaaaaATG	c41880	Nitrogenase stabilizer NifS of <i>Rhodospirillum rubrum</i>	510	33.2 (379)
<i>ytCL</i>	153853–155448	532	gggtgggagaaagtggagTTG	s39646	Acetate CoA ligase of <i>Bacillus subtilis</i>	767	34.1 (510)
<i>ytEL</i>	158078–159085	336	aaaaaaggaggagaaaaATG	prect4	Proteinase IV of <i>Escherichia coli</i>	324	35.5 (220)
<i>moaB</i>	164099–164611	171	acataaaaggagtgggccaATG	s31880	Molybdenum cofactor biosynthesis protein MoaB of <i>Escherichia coli</i>	295	41.3 (150)
<i>ytIL</i>	169260–170186c	309	aaaacaggaggagcacaTTG	a47126	LysR family transcription regulator of the <i>alsSD</i> operon of <i>Bacillus subtilis</i>	232	27.1 (192)
<i>ytMJ</i>	170889–171698	270	gagtgaggagccctATG	jkeccq	Glutamine-binding protein of <i>Escherichia coli</i>	291	29.6 (230)
<i>ytMK</i>	171712–172524	271	taaatggggcgatgagATG	jkebt	Lysine-arginine-ornithine-binding protein of <i>Salmonella typhimurium</i>	218	25.4 (236)
<i>ytML</i>	172546–173265	240	aaaagagaggagcagtgATG	qrecgp	Glutamine transport protein GlnP of <i>Escherichia coli</i>	373	35.5 (220)
<i>ytMM</i>	173295–174002	236	ggatgagggggagagaTTG	s55585	Octopine permease protein OccM of <i>Agrobacterium tumefaciens</i>	386	32.9 (225)
<i>hisP</i>	173999–174778	260	tttcaaggaggcattttATG	qrebpt	Histidine transport protein of <i>Salmonella typhimurium</i>	643	50.0 (248)
<i>ytNL</i>	175794–176075	94	tttaaggaggagacagagaATG	s54825	Glutaredoxin of <i>Ricinus communis</i> (castor bean)	92	30.4 (79)
<i>hipO</i>	178157–179407	417	gggaaggaggcggaacATG	s47321	Hippurate hydrolase of <i>Campylobacter jejuni</i>	583	35.1 (367)
<i>dnaE</i>	184508–187855	1116	gagaaaagggtgatagcATG	djcc3a	DNA polymerase III α -chain of <i>Escherichia coli</i>	1777	42.0 (920)
<i>ytTL</i>	189643–190431	263	tgttccgaaggcattATG	P43778	Acetyl-CoA carboxylase carboxyl-transferase subunit β of <i>Haemophilus influenzae</i>	937	47.2 (250)
<i>accA</i>	190416–191393	326	aaacaggaggtgacattgaGTG	a43452	Acetyl-CoA carboxylase of <i>Escherichia coli</i>	855	53.2 (312)
<i>pfk</i>	191577–192536	320	ctgagggtgaatggagaATG	kibsf	6-Phosphofructokinase of <i>Bacillus stearothermophilus</i>	1374	86.2 (319)
<i>pykA</i>	192579–194336	586	ttcagaaggagaagtgaacaaATG	s29783	Pyruvate kinase of <i>Bacillus stearothermophilus</i>	1987	72.8 (585)
<i>mutM</i>	205959–206795	279	acagagatagggaagtATG	s11489	Formamidopyrimidine-DNA glycosylase of <i>Bacillus firmus</i>	929	63.0 (279)
<i>ytbE</i>	208243–209085c	281	tttttaaggaggaaataacATG	s30383	Morphine dehydrogenase of <i>Pseudomonas putida</i>	614	43.3 (284)
<i>ytbD</i>	209109–210299c	397	tacaaggagggaacatcATG	P23910	Protein AraJ precursor of <i>Escherichia coli</i>	682	32.8 (360)
<i>gapB</i>	211069–212091	341	ttgataaggggtgccaaATG	js0164	Glyceraldehyde-3-P dehydrogenase of <i>Bacillus stearothermophilus</i>	1051	61.9 (333)

Table 3. Known protein motifs in the *rrnB*–*dnaB* region of the *B. subtilis* chromosome

See the legend to Table 1. All previously known ORFs listed in the Table were resequenced.

ORF	Endpoints (bp)	Size (aa)	Translation start	PIR/SP entry	Description of PIR entry	FASTA score	Percentage identity (over aa)
<i>ytaB</i>	15 415–15 882	156	atacatgaggagggtgcaaatATG	a53405	Benzodiazepine receptor 1 of <i>Mus. musculus</i>	180	28.9 (142)
<i>ytxN</i>	17 225–18 358	378	aggaatggaggctatgATG	s27131 c41317	Transcription factor SPT14 of <i>Saccharomyces cerevisiae</i> Probable LPS N-acetylglucosaminyltransferase of <i>Salmonella typhimurium</i>	141 238	16.9 (249) 18.9 (375)
<i>ytcC</i>	19 935–21 158c	408	tgatcagggggaataaaaATG	s27131 c41317	Transcription factor SPT14 of <i>Saccharomyces cerevisiae</i> Probable LPS N-acetylglucosaminyltransferase of <i>Salmonella typhimurium</i>	169 234	22.3 (251) 21.4 (374)
<i>yteA</i>	24 566–25 111c	182	agagaacgaggcgatcacaTTG	s45214	Dosage-dependent <i>dnaK</i> suppressor protein of <i>Escherichia coli</i>	113	40.0 (55)
<i>ytxM</i>	28 542–29 366	275	gctgtgcgggaagtgaaaaaaATG	P23133	2-Hydroxy-6-oxo-2,4-heptadienoate hydrolase of <i>Pseudomonas putida</i>	265	26.0 (288)
<i>ythB</i>	37 271–38 311c	347	tggaaggaagatcgtaacaATG	I30010	NADH-dehydrogenase of <i>Sauroleishmania tarentolae</i> mitochondrion	128	18.7 (182)
<i>ythA</i>	38 355–39 671c	439	attgttgatgattatgTTG	E64181	Cytochrome oxidase d subunit I homologue of <i>Haemophilus influenzae</i>	404	29.3 (311)
<i>ytiA</i>	39 891–40 139	83	gattttatgaaggagatacccATG	B64091	Ribosomal protein L31 homologue of <i>Haemophilus influenzae</i>	117	28.8 (66)
<i>ytkA</i>	41 742–42 179	146	cgatgaagggtgatggcATG	a41335	Extracellular metalloproteinase precursor of <i>Bacillus polymyxa</i>	84	19.6 (138)
<i>ytlD</i>	44 161–44 973c	271	ratggaaggagctgaattctcTTG	s16649	<i>dciA</i> protein of <i>Bacillus subtilis</i>	162	23.1 (208)
<i>ytlA</i>	46 057–46 746c	230	taaggaggtgtctatGTG	P42722	RUBISCO operon transcriptional regulator of <i>Alcaligenes eutrophus</i>	191	27.5 (191)
<i>ytmA</i>	46 897–47 670	258	aaaaggaggcggtgtrTTG	P27487	Dipeptidyl peptidase IV of human	177	23.5 (200)
<i>ytoA</i>	55 111–55 626c	172	aagaaggatgaatcttATG	P40882	Ferrityochelin-binding protein of <i>Pseudomonas aeruginosa</i>	396	35.8 (165)
<i>ytpA</i>	55 675–56 454	260	gtaggagtggaagatgATG	jh0655	Macrotetralide-resistance protein of <i>Streptomyces griseus</i>	125	19.7 (218)
<i>ytpA</i>	58 149–59 117c	323	aaggaggacatgaattgtgATG	P19206	Biotin synthetase of <i>Bacillus sphaericus</i>	161	25.7 (222)
<i>ytrC</i>	61 134–62 120	329	ttggagggaagcggaATG	s13585	<i>nosY</i> protein precursor of <i>Pseudomonas stutzeri</i>	127	19.7 (249)
<i>ytrD</i>	62 150–63 127	326	ttggagggaagtgATG	s044749	Cytochrome-c oxidase of <i>Rattus norvegicus</i> (mitochondrion)	111	20.6 (272)
<i>ytsD</i>	67 779–69 719	647	ctgtttaggcggggtgcagcATG	a00442	NADH dehydrogenase of <i>Aspergillus amstelodami</i> mitochondrion	158	18.5 (297)
<i>yttA</i>	69 756–70 502c	247	agaaaaagggttcattATG	s05568	IgA receptor precursor of <i>Streptococcus pyogenes</i>	112	23.2 (181)
<i>yttB</i>	70 691–71 884	398	aagaaggaggatgagATG	a03508	Tetracycline-resistance protein of <i>Escherichia coli</i> pBR322 plasmid	177	18.7 (353)
<i>ytvA</i>	72 121–72 906c	262	cggcctaagggggattatATG	s42866 s31136	Protein kinase of <i>Mesembryanthemum crystallinum</i> Regulatory protein NifL of <i>Azotobacter vinelandii</i>	288 186	42.4 (118) 29.6 (135)
<i>yteQ</i>	96 383–96 856	164	aacgtagacggaggcTTG	s32571	Membrane protein LacF of <i>Agrobacterium radiobacter</i>	214	30.4 (148)
<i>yteU</i>	99 805–100 473	223	aaaaaggaggggcgccATG	s52972	NADH dehydrogenase of <i>Apis mellifera</i> mitochondrion	132	20.9 (172)
<i>ytpP</i>	103 951–105 585	545	aatgggtgatgaagaacATG	Q00758	SpoVB protein of <i>Bacillus subtilis</i>	477	22.3 (537)
<i>ytpP</i>	110 303–111 694	464	rtgagggaaggagatcattATG	P45494	XAA-His dipeptidase of <i>Lactobacillus delbrueckii</i>	1014	35.1 (459)
<i>ytlR</i>	114 352–115 281	310	taagcaggtgagaagATG	P39074	BMRU protein of <i>Bacillus subtilis</i>	235	24.4 (299)
<i>amyX</i>	115 307–117 463	719	tttgaggagagagcttcgATG	s38801	Pullulanase precursor of <i>Klebsiella pneumoniae</i>	364	37.9 (190)
<i>ytpR</i>	125 766–126 371	202	agaaggatttttacttATG	syecfb	Phenylalanine-tRNA ligase of <i>Escherichia coli</i>	223	43.2 (132)
<i>ytpT</i>	127 280–129 388	703	gaagagcaaacaccgctgagATG	s09411	<i>spolIIE</i> protein of <i>Bacillus subtilis</i>	1272	49.8 (562)
<i>ytpP</i>	147 704–149 392	563	cagcaaggggcgctcattATG	s35760	<i>fcrA</i> protein precursor of <i>Streptococcus pyogenes</i>	155	19.6 (331)
<i>ytdI</i>	157 089–157 892c	268	aaagggggaatgggatATG	s51850	POS5 protein of <i>Saccharomyces cerevisiae</i>	135	18.6 (215)
<i>ytlG</i>	160 926–161 429	168	acaaaaggagggaattataATG	jc2570	tagD protein of <i>Vibrio cholerae</i>	283	38.9 (157)
<i>ytxK</i>	161 490–162 479	330	atggcaaggaggaacacaatATG	s26426	Type II specific deoxyribonuclease Eco571 of <i>Escherichia coli</i>	132	26.0 (73)
<i>ytkK</i>	167 697–168 464	256	gctttgggaggttatcatGTG	b42147 s36090	3-Ketoacyl-ACP reductase of <i>Escherichia coli</i> Glucose-1-dehydrogenase of <i>Bacillus subtilis</i>	319 235	34.9 (249) 27.6 (243)
<i>ytmO</i>	174 775–175 779	335	taaaggagaaggtgatATG	c42951 s39699	Luciferase of <i>Xenorhabdus lumines</i> Hypothetical protein of <i>Bacillus subtilis</i>	190 638	18.9 (185) 40.1 (332)
<i>ytnJ</i>	176 072–177 400	443	aaaggaggcaaaagATG	P54989	Nitrotriacetate monooxygenase component A of <i>Chelatobacter heintzii</i>	1056	38.3 (444)
<i>ribR</i>	177 447–178 139	231	gagagaggaggctgacttttTTG	P54575	RibC protein of <i>Bacillus subtilis</i>	262	40.0 (125)
<i>ytlQ</i>	182 275–183 216	314	taaaggagatcaaaATG	P22746	MGP A protein of <i>Mycoplasma genitalium</i>	492	30.4 (309)
<i>ytsJ</i>	187 992–189 224	411	aaggaggtgttaggagaATG	s23435	Malate dehydrogenase (NADP ⁺) of <i>Anas platyrhynchos</i>	183	34.1 (176)

problematic places in LR PCR direct sequencing are not known, and may be worthy of a thorough study.

The complete sequence, determined on both strands with a mean redundancy 6 and minimum of 4, spans primers AG2–AA8 (coordinates 13 and 130 kb, Fig. 1) and AC7–AH6 (coordinates 133 and 213 kb, Fig. 1). We resequenced the regions that were already in the databases (GenBank accession nos M74183, L17320, M12620, U05257 and M23549) and detected 37 discrepancies. Some of these may correspond to sequencing errors and some to strain differences. Our sequence corresponds to the *B. subtilis* 168 *trpC2* strain, with the

error rate estimated to be less than one mistake in 7000 bp. Most of the ambiguities (23) were detected in the sequence M74183, encoding a part of the menaquinone biosynthesis operon. All places where discrepancies have been detected were sequenced at least three times on each strand and the tracks were carefully analysed. Total sequence between the 5' end of primer AG2 and the 3' end of primer AH6 contains 197 815 bp. These include a contiguous 14025 bp sequence between AA8 and AC7, which we did not resequence. Only one reading was done (position near 137 kb, Fig. 1) to confirm the join between sequences M34719 and L17309 from GenBank.

Table 4. Hypothetical and new proteins in the *rrnB*–*dnaB* region of the *B. subtilis* chromosome

See the legend to Table 1. All ORFs were resequenced except *ytgG*, *ytgH* and *ytgJ* (accession no. L31845).

ORF	Endpoints (bp)	Size (aa)	Translation start	PIR/SP entry	Description of PIR entry	FASTA score	Percentage identity (over aa)
<i>ytgO</i>	19430–19861	144	tcgggagcgcaattgacATG	–	No homologies	<100	–
<i>ytgC</i>	34856–36163	436	taaggagtgagcgcaATG	g64149	Hypothetical protein HI0359 of <i>Haemophilus influenzae</i>	377	31.9 (282)
<i>ytgD</i>	36153–37040	296	gagaaggacagatcgcaaATG	E64063	Hypothetical membrane protein of <i>Haemophilus influenzae</i>	290	24.0 (279)
<i>ytgE</i>	37055–37222c	56	tttataaaggaggggATG	–	No homologies	<80	–
<i>ytjA</i>	40793–41020c	76	cagaaaggtgtctgaagaacATG	JQ0730	Hypothetical 9.3 kDa protein of <i>Proteus mirabilis</i>	277	63.9 (72)
<i>ytjB</i>	41149–41622	158	aagggagagtgccattATG	g64008	Hypothetical protein HI0491 of <i>Haemophilus influenzae</i>	273	37.9 (145)
<i>ytjK</i>	42442–42879	146	caaggaggagatgccATG	e64171	Hypothetical protein HI1349 of <i>Haemophilus influenzae</i>	259	28.4 (141)
<i>ytjC</i>	43045–43449	135	aaggatgttgggataaATG	s43905	Hypothetical protein D of <i>Clostridium perfringens</i>	176	26.4 (110)
<i>ytjD</i>	43658–44134	159	ggatttgaggataaacATG	–	No homologies	<80	–
<i>ytjB</i>	45743–46060c	106	atggcgaaacaaagctatTTG	–	No homologies	<60	–
<i>ytmB</i>	47722–47964	81	aaaggggatttcacTTG	–	No homologies	<70	–
<i>ytjB</i>	56475–57578	368	tagtaggggtgcatgtTTG	–	No homologies	<80	–
<i>ytjB</i>	57568–58152c	195	ttgaggaggaatcagcgtATG	–	No homologies	<70	–
<i>ytjF</i>	63827–65137	437	aaaggaggaattaccgtTTG	h64035	Hypothetical protein HI1548 of <i>Haemophilus influenzae</i>	160	19.2 (203)
<i>ytjB</i>	73311–73646	112	tttaggaaggtgatgATG	–	No homologies	<80	–
<i>ytjF</i>	76709–76915	69	gaggtagcggagggcATG	–	No homologies	<90	–
<i>ytjP</i>	82552–83451	300	aaggaaaggtgatggATG	–	No homologies	<90	–
<i>ytjQ</i>	90123–90728	202	aattatgaggagctgTTG	–	No homologies	<80	–
<i>ytjQ</i>	91848–93170	441	aatggaggggtctttgaTTG	P37966	Lipoprotein LPLA precursor of <i>Bacillus subtilis</i>	944	35.3 (442)
<i>ytjP</i>	94130–95674c	515	tgaagrtgagtttatgaaATG	–	No homologies	<80	–
<i>ytjP</i>	95890–96270	127	aaaggaggctaggaaaaATG	–	No homologies	<90	–
<i>ytjR</i>	96864–97985	374	aaggatttttaggaaagggATG	b40364	Hypothetical protein 1 (<i>picA</i> 5' region) of <i>Agrobacterium tumefaciens</i>	273	30.0 (230)
<i>yteS</i>	98006–98485	160	aaggatcgtacggctTTG	–	No homologies	<70	–
<i>yteT</i>	98497–99783	429	atagaggtgaggataTTG	–	No homologies	<80	–
<i>ytjP</i>	102487–103704c	319	ggaagaggatcacggcggATG	s47712	Hypothetical protein f400 of <i>Escherichia coli</i>	270	32.0 (250)
<i>ytjQ</i>	107712–108869	386	ttcaaggtgatcggtcATG	s30687	Hypothetical protein o716 of <i>Escherichia coli</i>	132	18.6 (215)
<i>ytjP</i>	108908–110206c	433	aattggagatactATG	e64143	Hypothetical protein HI0125 of <i>Haemophilus influenzae</i>	753	34.3 (432)
<i>ytjP</i>	112812–113363	184	gggagrtgcagacggATG	s43916	Hypothetical protein 3 of <i>Bacillus stearothermophilus</i>	381	33.7 (181)
<i>ytjQ</i>	113389–114303	305	cggagrtgagaaacaaacATG	s43916	Hypothetical protein 3 of <i>Bacillus stearothermophilus</i>	152	41.8 (67)
<i>ytjP</i>	117631–118440	270	ttagaagaatgatgggTTG	–	No homologies	<80	–
<i>ytjQ</i>	118926–119567	214	atagaggtgtcaatcatATG	f64148	Hypothetical protein HI0340 of <i>Haemophilus influenzae</i>	229	28.5 (172)
<i>ytjP</i>	119713–120483	257	tatggacggagggcccATG	–	No homologies	<100	–
<i>ytjP</i>	122794–123867	358	tagaaatgaggggaaaaATG	s40842	Hypothetical protein f356 of <i>Escherichia coli</i>	656	37.8 (352)
<i>ytjQ</i>	123922–124368c	149	ataaggagggatttttattATG	–	No homologies	<70	–
<i>ytjQ</i>	124941–125750	270	aaaggaggacatcggtTTG	–	No homologies	<70	–
<i>ytjS</i>	126531–127325	265	aaggaggttagagataaattATG	–	No homologies	<70	–
<i>ytjG</i>	131098–131529	144	agcaaggaggtttgagATG	–	No homologies	<95	–
<i>ytjH</i>	131560–132015	152	aaggagggaataatgATG	–	No homologies	<90	–
<i>ytjL</i>	132039–132365	109	gggaaggaaagcgttgatcATG	–	No homologies	<90	–
<i>ytjP</i>	143687–145426	580	tttaggtgataaaaaATG	s33422	Hypothetical protein 183 of <i>Streptomyces ambofaciens</i>	188	27.1 (170)
<i>ytjP</i>	145464–145700c	79	tttcggcagaggggtTTG	s37891	hypothetical protein YKL069w of <i>Saccharomyces cerevisiae</i>	154	41.1 (73)
<i>ytjP</i>	146701–147507c	269	tttcattgggttgacatcaATG	d47754	orf13 3' to <i>hisE</i> – <i>Lactococcus lactis</i>	151	27.7 (177)
<i>ytjJ</i>	152168–153244	359	atgatgaggtagaacATG	P47612	Hypothetical protein MG372 of <i>Mycoplasma genitalium</i>	546	30.9 (363)
<i>ytjJ</i>	155468–157057	530	aaagagagtgaacgtaATG	–	No homologies	<80	–
<i>ytjE</i>	159098–159592	165	gtaggaggggaacaataATG	–	No homologies	<85	–
<i>ytjI</i>	159667–160347	227	tagaggaagtgaagaatATG	–	No homologies	<75	–
<i>ytjJ</i>	160361–160816	152	gtttaaggaggaaagcgtatATG	–	No homologies	<70	–
<i>ytjL</i>	168527–168829	101	ttgtgacaggtgaaagaaATG	–	No homologies	<85	–
<i>ytjM</i>	170339–170875	179	gaagaggagtgcacaaaATG	P54951	Hypothetical 19.0 kDa protein of <i>Bacillus subtilis</i>	559	45.6 (158)
<i>ytjM</i>	179426–180328	301	atatagatgggtgttctgATG	P29942	Hypothetical 27.4 kDa protein of <i>Paracoccus denitrificans</i>	167	22.4 (277)
<i>ytjO</i>	180504–181823	440	tttagaggtgaaagcTTG	s22196	Hypothetical protein 3 of <i>Desulfurolobus ambivalens</i>	148	30.4 (112)
<i>ytjP</i>	181842–182144	101	cgatggaggatttcattATG	–	No homologies	<90	–
<i>ytjR</i>	183527–184030c	168	cccaggagcatctttcATG	–	No homologies	<85	–
<i>ytjV</i>	194842–195960c	372	acgagaggaggacgtttttGTG	e64004	Hypothetical protein HI0237 of <i>Haemophilus influenzae</i>	113	22.7 (163)
<i>ytjO</i>	196057–196521	155	tttaaggaggagtgccatGTG	b64020	Hypothetical protein HI1074 of <i>Haemophilus influenzae</i>	261	34.3 (143)
<i>ytjF</i>	206967–207446	160	aggaagaaacggatatgcaaATG	–	No homologies	<115	–
<i>ytjG</i>	207608–208201	198	aaaaagagtggtgatcTTG	a64161	Hypothetical protein HI0890 of <i>Haemophilus influenzae</i>	383	35.6 (191)
<i>ytjD</i>	210483–210863	127	tgtagaggagaaatggtATG	P42406	Hypothetical protein YckH of <i>Bacillus subtilis</i>	455	56.7 (97)
<i>ytjF</i>	212324–212710	129	gttaaggggggtccATG	–	No homologies	<75	–
<i>ytjG</i>	212984–213442	153	tagttggagactgattttaTTG	b64162	Hypothetical protein HI0943 of <i>Haemophilus influenzae</i>	371	47.6 (147)

Gene functions in the *rrnB*–*dnaB* region

Functional annotation of genes, described here, is based essentially on the properties of the encoded proteins. The approach to determine the gene functions is based on searching for the homologous proteins, and reveals

paralogous or orthologous genes. The term paralogous refers to genes encoding homologous proteins with different biological functions (but often identical biochemical functions) within an organism, whereas orthologous refers to genes having the same biological function in different organisms (Fitch, 1970). The

The biological function is defined here as knowledge of genetic, biochemical or physiological conditions when the properly regulated expression of the gene is essential. The biological function is indicated more precisely, when possible, by assigning the gene to a cellular process for which such conditions can be well-defined, such as a metabolite biosynthesis, citric acid cycle, protein biosynthesis, etc. (Table 1). The results of homology searches usually indicate the biochemical function of a gene product, which does not necessarily reveal the biological function of the gene. There are 55 genes in this region encoding proteins for which biochemical (or sometimes even biological) function can be assigned from homology to known proteins (Table 2). The third class groups 39 genes, for which neither biochemical nor biological function can be postulated from homology to known proteins (Table 3). We refer to these as encoding proteins with known motifs. Finally, 60 genes encode proteins with unknown function (Table 4), 28 of which

Table 5. Functional specialization of the *rrnB-dnaB* region

Gene function categories	<i>H. influenzae</i>		<i>M. genitalium</i>		<i>B. subtilis</i> ORF database		<i>B. subtilis</i> <i>rrnB-dnaB</i> region	
	ORFs	%	ORFs	%	ORFs	%	ORFs	%
Amino acid biosynthesis	68	6.7	1	0.3	50	5.7	5	5.3
Biosynthesis of cofactors	54	5.4	5	1.6	42	4.8	15	15.7
Cell envelope	84	8.3	17	5.3	46	5.3	3	3.2
Cellular processes	53	5.3	21	6.6	14	1.6	3	3.2
Central intermediary metabolism	30	3.0	6	1.9	43	4.9	10	10.6
Energy metabolism	112	11.0	31	9.7	100	11.4	11	11.7
Fatty acid and phospholipids	25	2.5	6	1.9	6	0.7	3	3.2
Purines and pyrimidines, nucleotides and nucleosides	53	5.3	19	6.0	44	5.0	3	3.2
Regulatory functions	64	6.4	7	2.2	97	11.1	11	11.7
Replication	87	8.6	32	10.0	36	4.1	5	5.3
Transcription	27	2.7	12	3.8	11	1.3	0	0
Translation	141	14.0	101	31.8	85	9.7	4	4.2
Transport	123	12.2	34	10.7	73	8.3	19	20.2
Other categories	93	9.2	17	8.2	225	25.8	2	2.1
Total classified ORFs	1014		309		872		94	

have homology to hypothetical proteins from other organisms. The ORF organization of the sequenced area is shown in Fig. 2.

Summarizing these data, we found 193 ORFs in the *rrnB-dnaB* area. Five levels of gene function knowledge were recognized, and the percentage of the genes belonging to each is 20, 28, 20, 15 and 17, respectively.

We calculated the percentage of genes, from Tables 1 and 2, belonging to different biological functional categories, as proposed by Riley (1993), and compared it with analogous distributions for completely sequenced genomes of *Haemophilus influenzae* and *Mycoplasma genitalium* (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995). The *B. subtilis* ORF classification suggested by Dr N. Ogasawara (URL: <http://bacillus.tokyo-center.genome.ad.jp:8008/BSORF-DB.html>) was also used in this comparison (Table 5). One feature of the sequenced *rrnB-dnaB* region seems to be apparent. The genes implicated in central intermediary, energy and fatty acid metabolism, together with transporters, in this region comprise 46% of 94 ORFs listed in Tables 1 and 2. This figure is almost twofold higher than the corresponding percentage for longer genome regions in organisms listed in Table 5. As described previously, parts of the *B. subtilis* genome seem to be specialized in carrying the functionally linked genes (Daniel & Errington, 1993; Lazarevic *et al.*, 1996; Mizuno *et al.*, 1996). If this is the case, we can suggest that the genes in the 200 kb *rrnB-dnaB* region are specialized in performing the metabolic link between the glycolytic pathway and the citric acid cycle. The activity of many transporters detected in this area should presumably be involved in modulating this metabolism. It is noteworthy that the

ccpA gene, encoding the principal catabolic regulator of *B. subtilis*, is also localized here (see Fig. 2). On the other hand, the *rrnB-dnaB* region seems to be almost completely devoid of sporulation genes ('Other categories' in Table 5).

The gene *ytpT* encodes a protein highly homologous to *spolIIE* gene product which was shown to be involved in the chromosome transfer between mother and daughter cell during sporulation (Wu & Errington, 1994). The function of *ytpT* is not known. In this respect it is interesting to mention apparent functional similarity between several other genes in the vicinity of *spolIIE* (~135°) and the *rrnB-dnaB* region. The genes *pdhABCD*, encoding pyruvate dehydrogenase complex, and *pycA*, encoding pyruvate carboxylase, map at 125°. They provide enzymic activities responsible for the entry of pyruvate into the Krebs cycle. The genes *pckA*, *pykA*, *ackA*, *acsA*, *citZ*, *citH*, *malS*, listed in Tables 1 and 2, encode activities dealing with the same metabolites. *fruB*, which encodes 1-phosphofructokinase (125°), is closely related to 6-phosphofructokinase encoded by *pfk* (Table 2). *polC*, which encodes subunits of DNA polymerase III, was mapped near 145°. We found another subunit of this polymerase encoded by *dnaE* (Table 2). Very close to it there is also *polA*, which encodes another DNA-synthesizing enzyme (Table 1). The last couple of genes which may be regarded as functionally similar are *codY*, mapping at 145°, which encodes a recently characterized nutritional regulator (Slack *et al.*, 1995), and *ccpA*, the principal catabolite regulator, which was known previously (Table 1). The significance of this functional relation between the genes at 140° and 270° is not clear. It is possible that, for

more efficient functioning *in vivo*, the corresponding proteins should work in some complexes and their positions on the chromosome provide a possibility of close spatial location of their synthesis. In this respect it is interesting to mention that if the circular genome map is divided by the axis of symmetry based on this similarity, the chromosome will be crossed by that axis near the origin (25°) and terminus (205°) of replication.

ACKNOWLEDGEMENTS

We highly acknowledge advice from Dr Veronique Capuano on LR PCR and the help of Dr Alexander Bolotin and Petar Pujic on plasmid sequencing. This work was supported in part by grants from GREG, the French Genome Project (decision 21) and the EEC (BIO2-CT93-0272).

REFERENCES

- Abe, A., Koide, H., Kohno, T. & Watabe, K. (1995). A *Bacillus subtilis* spore coat polypeptide gene, *cotS*. *Microbiology* **141**, 1433–1442.
- 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.
- Anand, R., Villasante, A. & Tyler-Smith, C. (1989). Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed-field electrophoresis. *Nucleic Acids Res* **17**, 3425–3433.
- Azevedo, V., Alvarez, E., Zumstein, E., Damiani, G., Sgaramella, V., Ehrlich, S. D. & Serror, P. (1993). An ordered collection of *Bacillus subtilis* DNA segments cloned in yeast artificial chromosomes. *Proc Natl Acad Sci USA* **90**, 6047–6051.
- Bolotin, A., Khazak, V., Stoyanova, N., Ratmanova, K., Yomantas, Y. & Kozlov, Y. (1995). Identical amino acid sequence of the *aroA(G)* gene products of *Bacillus subtilis* 168 and *B. subtilis* Marburg strain. *Microbiology* **141**, 2219–2222.
- Bolotin, A., Sorokin, A. & Ehrlich, S. D. (1996). Mapping of the 150 kb *spoIIIC-pheA* region of the *Bacillus subtilis* chromosome using Long Accurate PCR and three yeast artificial chromosomes. *Microbiology* **142**, 3017–3020.
- Bower, S., Perkins, J., Yocum, R. R., Howitt, C. L., Rahaim, P. & Pero, J. (1996). Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon. *J Bacteriol* **178**, 4122–4130.
- Bruand, C. & Ehrlich, S. D. (1995). The *Bacillus subtilis* *dnaL* gene is part of the *dnaB* operon. *Microbiology* **141**, 1199–1200.
- Capuano, V., Galleron, N., Pujic, P., Sorokin, A. & Ehrlich, S. D. (1996). Organization of the *Bacillus subtilis* 168 chromosome between *kdg* and the attachment site of the *SPβ* prophage: use of Long Accurate PCR and yeast artificial chromosomes for sequencing. *Microbiology* **142**, 3005–3015.
- Connors, M. J., Mason, J. M. & Setlow, P. (1986). Cloning and nucleotide sequencing of genes for three small acid-soluble proteins from *Bacillus subtilis* spores. *J Bacteriol* **166**, 417–425.
- Daniel, R. A. & Errington, A. (1993). DNA sequence of the *murE-murD* region of *Bacillus subtilis* 168. *J Gen Microbiol* **139**, 361–370.
- Dear, S. & Staden, R. (1991). A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res* **19**, 3907–3911.
- Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988). High-efficiency transformation of *E. coli* by high-voltage electroporation. *Nucleic Acids Res* **16**, 6127–6145.
- Driscoll, J. R. & Taber, H. W. (1992). Sequence organization and regulation of the *Bacillus subtilis* *menBE* operon. *J Bacteriol* **174**, 5063–5071.
- 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.
- Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Syst Zool* **270**, 445–446.
- Fleischmann, R. D., Adams, M. D., White, O. & 37 other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
- Fraser, C. M., Gocayne, J. D., White, O. & 26 other authors (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403.
- Green, C. J., Stewart, G. C., Hollis, M. A., Vold, B. S. & Bott, K. F. (1985). Nucleotide sequence of the *Bacillus subtilis* ribosomal RNA operon, *rrnB*. *Gene* **37**, 261–266.
- Grundy, F. J. & Henkin, T. M. (1990). Cloning and analysis of the *Bacillus subtilis* *rpsD* gene, encoding ribosomal protein S4. *J Bacteriol* **172**, 6372–6379.
- Grundy, F. J., Waters, D. A., Allen, H. G. & Henkin, T. M. (1993a). Regulation of the *Bacillus subtilis* acetate kinase gene by CcpA. *J Bacteriol* **175**, 7348–7355.
- Grundy, F. J., Waters, D. A., Takova, T. Y. & Henkin, T. M. (1993b). Identification of genes involved in utilization of acetate and acetoin in *Bacillus subtilis*. *Mol Microbiol* **10**, 259–271.
- Henkin, T. M., Grundy, F. J., Nicholson, W. L. & Chambliss, G. H. (1991). Catabolite repression of α -amylase gene expression in *Bacillus subtilis* involves a *trans*-acting gene product homologous to the *Escherichia coli* *lacI* and *galR* repressors. *Mol Microbiol* **5**, 575–584.
- Henkin, T. M., Glass, B. L. & Grundy, F. J. (1992). Analysis of the *Bacillus subtilis* *tyrS* gene: conservation of a regulatory sequence in multiple tRNA synthetase genes. *J Bacteriol* **174**, 1299–1306.
- Hulett, F. M., Lee, J., Shi, L., Sun, G., Chesnut, R., Sharkova, E., Duggan, M. F. & Kapp, N. (1994). Sequential action of two-component genetic switches regulates the PHO regulon in *Bacillus subtilis*. *J Bacteriol* **176**, 1348–1358.
- Jin, S. & Sonenshein, A. L. (1994). Identification of two distinct *Bacillus subtilis* citrate synthase genes. *J Bacteriol* **178**, 4669–4679.
- Jin, S. & Sonenshein, A. L. (1996). Characterization of the major citrate synthase of *Bacillus subtilis*. *J Bacteriol* **176**, 3658–3660.
- Jin, S., de Jesus-Berrios, M. & Sonenshein, A. L. (1996). A *Bacillus subtilis* malate dehydrogenase gene. *J Bacteriol* **178**, 560–563.
- Kappes, R. M., Kempf, B. & Bremer, E. (1996). Three transport systems for the osmoprotectant glycine betaine operate in *Bacillus subtilis*: characterization of OpuD. *J Bacteriol* **178**, 5071–5079.
- Kiel, J. A., Boels, J. M., Beldman, G. & Venema, G. (1994). Glycogen in *Bacillus subtilis*: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation. *Mol Microbiol* **11**, 203–218.
- Kunst, F., Vassaroti, A. & Danchin, A. (1995). Organisation of the European *Bacillus subtilis* genome sequencing project. *Microbiology* **141**, 249–255.
- Lazarevic, V., Maüel, C., Soldo, B., Freymond, P.-P., Margot, P. & Karamata, D. (1996). Sequence analysis of the 308° to 311° segment of the *Bacillus subtilis* 168 chromosome, a region devoted to cell wall metabolism, containing non-coding grey holes which reveal chromosomal rearrangements. *Microbiology* **141**, 329–335.
- Mizuno, M., Mazuda, S., Takemaru, K., Hosono, S., Sato, T., Takeuchi, M. & Kobayashi, Y. (1996). Systematic sequencing of the

283 kb 210°–232° region of the *Bacillus subtilis* genome containing the *skin* element and many sporulation genes. *Microbiology* **142**, 3103–3111.

Moszer, I., Glaser, P. & Danchin, A. (1995). *SubtiList*: a relational database for the *Bacillus subtilis* genome. *Microbiology* **141**, 261–268.

Mountain, A., McChesney, J., Smith, M. C. M. & Baumberg, S. (1986). Gene sequence encoding early enzymes of arginine synthesis within a cluster in *Bacillus subtilis*, as revealed by cloning in *Escherichia coli*. *J Bacteriol* **165**, 1026–1028.

Ogasawara, N., Moriya, S., Mazza, P. G. & Yoshikawa, H. (1986). Nucleotide sequence and organization of *dnaB* gene and neighboring genes on the *Bacillus subtilis* chromosome. *Nucleic Acids Res* **14**, 9989–9999.

Palaniappan, C., Taber, H. & Meganathan, R. (1994). Biosynthesis of O-succinylbenzoic acid in *Bacillus subtilis*: identification of *menD* mutants and evidence against the involvement of the α -ketoglutarate dehydrogenase complex. *J Bacteriol* **176**, 2648–2653.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.

Perego, M., Ferrari, E., Bassi, M. T., Galizzi, A. & Mazza, P. (1987). Molecular cloning of *Bacillus subtilis* genes involved in DNA metabolism. *Mol Gen Genet* **209**, 8–14.

Riley, M. (1993). Functions of gene products of *Escherichia coli*. *Annu Rev Microbiol* **32**, 862–952.

Rowland, B. & Taber, H. (1996). Duplicate isochorismate synthase genes of *Bacillus subtilis*: regulation and involvement in the biosyntheses of menaquinone and 2,3-dihydroxybenzoate. *J Bacteriol* **178**, 854–861.

Rowland, B., Hill, K., Miller, P., Driscoll, J. & Taber, H. (1995). Structural organization of a *Bacillus subtilis* operon encoding menaquinone biosynthetic enzymes. *Gene* **167**, 105–109.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Seki, T., Yoshikawa, H., Takahashi, H. & Saito, H. (1988). Nucleotide sequence of the *Bacillus subtilis* *phoR* gene. *J Bacteriol* **170**, 5935–5938.

Slack, F. J., Serror, P., Joyce, E. & Sonenshein, A. L. (1995). A gene

required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. *Mol Microbiol* **15**, 689–702.

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 homologs 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.

Tabor, S. & Richardson, C. C. (1995). A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxynucleotides. *Proc Natl Acad Sci USA* **92**, 6339–6343.

Tatusov, R. L., Mushegian, A. R., Bork, P., Brown, N. P., Hayes, W. S., Borodovsky, M., Rudd, K. E. & Koonin, E. V. (1996). Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr Biol* **6**, 279–291.

Tovar-Rojo, F. & Setlow, P. (1991). Effects of mutant small, acid-soluble spore proteins from *Bacillus subtilis* on DNA *in vivo* and *in vitro*. *J Bacteriol* **173**, 4827–4835.

Vander Horn, P. B. & Zahler, S. A. (1992). Cloning and nucleotide sequence of the leucyl-tRNA synthetase gene of *Bacillus subtilis*. *J Bacteriol* **174**, 3928–3935.

Wabiko, H., Ochi, K., Nguyen, D. M., Allen, E. R. & Freeze, E. (1988). Genetic mapping and physiological consequences of *metE* mutations of *Bacillus subtilis*. *J Bacteriol* **170**, 2705–2710.

Wu, L. J. & Errington, J. (1994). *Bacillus subtilis* SpoIIIE protein required for DNA segregation during asymmetric cell division. *Science* **264**, 572–575.

Received 20 June 1997; accepted 17 July 1997.