Analysis of the *Bacillus subtilis* genome: cloning and nucleotide sequence of a 62 kb region between 275° (*rrnB*) and 284° (*pai*)

Bauke Oudega,¹ Gregory Koningstein,¹ Lúsa Rodrigues,¹ Maria de Sales Ramon,¹ Helmut Hilbert,² Andreas Düsterhöft,² Thomas M. Pohl³ and Thomas Weitzenegger³

Author for correspondence: Bauke Oudega. Tel. +31 20 4447177. Fax: +31 20 4447229.
e-mail: oudega@bio.vu.nl

¹ Department of Molecular Microbiology, Institute of Molecular Biological Sciences, BioCentrum Amsterdam, Faculty of Biology, Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

² Qiagen GmbH, Max-Volmer Strasse 4, D-40724 Hilden, Germany

³ Gesellschaft für Analyse-Technik und Consulting GmbH, Fritz-Arnold Strasse 23, D-78467 Konstanz, Germany

In the framework of the international project aimed at the sequencing of the *Bacillus subtilis* genome, five DNA fragments in the region between *rrnB* (275°) and *pai* (284°) were cloned by inverse and combinatorial long-range PCR and their nucleotide sequences were determined and analysed. Together these sequences constituted a contig of 62 229 bp. On the basis of the position of *NotI* and *SfiI* restriction sites, the orientation and order of known genetic markers was determined to be *pai* (284°) – *degQ* *comQ* *comP* *comAA* *comAB* – *pbpD* – *kapB* *kinB* *patB* – *mcpB* *tlpA* *mcpA* *tlpB* – *rrnB* (275°). Fifty-four ORFs were detected. Thirteen of these coincided with known *B. subtilis* genes, and 41 new ORFs were found. Of the predicted new gene products, 12 showed no significant similarity to other known proteins, whereas ten showed strong similarity to proteins of other organisms with unknown function. Nineteen predicted proteins showed strong similarity to known proteins of other organisms, for instance a Na⁺/H⁺ antiporter system of *Bacillus alcalophilus*, a sugar transport system found in *Mycoplasma genitalium*, NADH-dependent butanol dehydrogenase of *Clostridium acetobutylicum*, glucose-6-phosphate isomerase A of *B. subtilis*, exo-1,4- α -glucosidase activity of *Bacillus stearothermophilus* and L-rhamnose isomerase of *Escherichia coli*.

Keywords: *Bacillus subtilis*, genome sequencing, 275–284° region

INTRODUCTION

Within the framework of the international consortium sequencing the *Bacillus subtilis* genome, we are responsible for the sequencing of the chromosomal region between *rrnB* (275°) and *pai* (284°) (Kunst *et al.*, 1995). According to the genetic map of *B. subtilis* published previously (Anagnostopoulos *et al.*, 1993; Biaudef *et al.*, 1996) and the *SubtiList* database (Moszer *et al.*, 1995), this area of the *B. subtilis* genome contains several genes and DNA contigs that have been published. At 276°, a 2105 bp DNA sequence was found (Anagnostopoulos *et al.*, 1993) containing the *degA* gene (Bussey & Switzer, 1993). However, sequencing and analysis of a 10 kb DNA segment generated by inverse long-range PCR and containing the *degA* gene revealed that this genetic

marker is actually located at 99° on the genetic map (see also Biaudef *et al.*, 1996). Therefore, this information was not further used for this study. At 277° or at 280° (dependent on the genetic map used) a contig of 3771 bp was mapped containing the *patB*, *kinB* and *kapB* genes (Trach & Hoch, 1993). At 279° a contig of 7269 bp was mapped containing the *yuxH*, *degQ*, *comQ*, *comP*, *comAA* and *comAB* genes (Weinrauch *et al.*, 1989, 1990, 1991; Yang *et al.*, 1986). Also at 279° or at 281°, an 8877 bp region was mapped containing the *mcpB*, *tlpA*, *mcpA* and *tlpB* genes (Hanlon & Ordal, 1994). Furthermore, at 279° or at 281° a region of 1709 bp was mapped encoding the *ald* gene (Siranosian *et al.*, 1993). Finally, at 280° a contig of 3414 bp was sequenced containing the partial gene *yuxJ*, and the *pbpD* and *yuxK* genes (Popham & Setlow, 1994).

The information presented above indicates that the map position and orientation of the various known contigs in the 275–284° region is not at all clear. To clone and

The EMBL accession numbers for the sequences reported in this paper are Z93932–Z93938.

Fig. 1. Schematic representation of the 62229 bp contig. The upper part shows the position of various restriction endonuclease cleavage sites. The lower part shows the orientation of the contig, the theoretical position of the known DNA fragments based on the genetic map of Biaudet *et al.* (1996) (open bars), the determined location of the known DNA fragments after sequencing (hatched bars) and the position of the cloned DNA fragments obtained by long-range PCR techniques (black bars). Inverse PCR fragments: 1 (accession no. Z93932) and 5 (Z93938); combinatorial PCR fragments: 2 (Z93933), 3 (Z93937) and 4 (Z93934, Z93935 and Z93936).

sequence this region, we decided to use inverse and combinatorial long-range PCR techniques. In this paper we describe the cloning, nucleotide sequencing and gene features of a 62 kb region of the *B. subtilis* genome located between 275° and 284°.

METHODS

Bacterial strains and growth medium. *B. subtilis* strain 168 is used by all the groups participating in the *B. subtilis* chromosome sequencing project. We obtained this strain from Sierd Bron (Groningen, The Netherlands) and used it as a source of chromosomal DNA for all cloning and PCR experiments. *Escherichia coli* strain XL-1 Blue, pUC18/19 vectors and phage vectors M13mp18/19 (Messing, 1983) were used for shotgun cloning and preparation of DNA templates for sequencing. LB medium was used for standard cultures of *B. subtilis* and *E. coli* (Miller, 1972).

Cloning of DNA and PCR techniques. All basic recombinant DNA procedures, like, for instance, isolation and purification of chromosomal DNA and plasmid DNA, restriction enzyme digestion, agarose electrophoresis, ligation of DNA and bacterial transformation were carried out essentially as described by Sambrook *et al.* (1989).

For long-range PCR experiments the Expand Long Template PCR system (Boehringer Mannheim) or the GeneAmp XL PCR kit (Perkin Elmer) was used. Based on the nucleotide sequence of known genes in the region to be sequenced, primers of 22 or 23 nucleotides ($T_m \sim 66^\circ\text{C}$) were prepared. For direct combinatorial PCR experiments (Sorokin *et al.*, 1996) chromosomal DNA of *B. subtilis* was purified and used with the appropriate primers. For inverse PCR experiments chromosomal DNA was digested to completion with appropriate restriction enzymes not cutting in the known sequence. After dilution of the digestion mixture, DNA fragments were circularised by ligation. This ligation mixture was then used with the appropriate primers in long-range PCR experiments. In all cases, ten independent PCR runs were carried out. The resulting products of all ten PCR runs were

then mixed, the DNA fragment was purified and the ends of the fragment were sequenced using the PCR primers.

DNA sequencing. For sequencing of DNA a random sequencing strategy was used. Some of the DNA fragments to be sequenced were sonicated and fragments between 500 and 1000 bp in length were purified and treated with Klenow enzyme and T4 DNA polymerase. Blunt-ended fragments were then inserted into the *Sma*I site of pUC18 or pUC19 and amplified in *E. coli* XL-1 Blue cells. In an alternative approach, other DNA fragments were partially digested with *Sau*3A1, *Alu*I or *Rsa*I and DNA fragments between 500 and 1500 bp in length were purified and cloned in M13mp18 or in M13mp19. Over 600 clones were selected and sequenced, at least 10 clones for each kb of DNA. Single stranded DNA was isolated and sequenced with a DNA sequencer (model 373A, Perkin Elmer/Applied Biosystems) using a Taq Dye Primer Cycle Sequencing kit (Perkin Elmer/Applied Biosystems). Double-stranded DNA and PCR products were sequenced by using custom made sequencing primers and the Taq Dye Terminator Cycle Sequencing kit (Applied Biosystems). Each strand was sequenced at least twice. Synthesized oligonucleotide primers and the cycle sequencing protocol were used on the original long-range PCR fragments to complete the sequence. Regions of uncertainty were re-sequenced in both directions. The overall redundancy was 6–8.

Computer analysis. The following programs and packages were used for analysis and alignment of DNA sequences: GENEWORKS 2.3, DNA Strider 1.0, SEQUENCER 2.1 and the GCG package. Putative gene products were searched for similarity to sequences reported previously in non-redundant protein databases (NCBI BLASTX Search (<http://www.ncbi.nlm.nih.gov>; see also Gish & States, 1993; Altschul *et al.*, 1990).

RESULTS AND DISCUSSION

Five different DNA fragments were obtained in the region between 275° and 284° (Fig. 1). First a 5 kb inverse PCR fragment was obtained that extended the

Fig. 2. Organization of ORFs of the 62229 bp contig. The scale is given in kb. The positions of two *NotI* sites and one *SfiI* site are indicated; the positions of these sites in kb on the *B. subtilis* chromosome as published by Itaya & Tanaka (1991) are shown in parentheses. ORFs are indicated by thick black arrows. Arrows show the direction of transcription/translation. The names of the genes are indicated below the arrows; bold indicates previously published sequences, roman characters indicate newly discovered ORFs. Stem-loop structures representing *p*-independent terminators found by computer analyses are indicated. Promoter regions published previously are indicated by P.

7269 bp *comQPA* region. Following sequencing of this fragment, combinatorial PCR experiments yielded three DNA products. A 4 kb fragment connected the 3414 bp *pbpD* contig to the 3771 bp *kin-pat-kap* region. A 14 kb DNA fragment connected the extended *com* region to the *pbpD* contig, and a 16 kb DNA segment connected the *kin-pat-kap* contig to the 8877 bp *mcp-tlp* contig. Finally, a 6 kb inverse PCR fragment was obtained that extended the *mcp-tlp* region.

The nucleotide sequence of the PCR-generated DNA fragments was determined. The obtained sequences and

the known sequences were aligned and formed a large contig of 62229 bp. Combinatorial PCR experiments were then carried out to connect this 62 kb contig to the *rrnB* region at 275°, the *pai* marker at 284°, or to the *ald* region at 279°. However, these PCR experiments did not yield any product. This suggested that the size of the gaps between the 62 kb contig and the other region in this area is at least 20 kb, and that the *ald* marker may be outside the region.

Based on the PCR results and the sequence analysis of the 62 kb contig the order of the known genetic markers in the 275–284° region can be determined. In contrast to the data published before (*SubtiList* database; Anagnostopoulos *et al.*, 1993) the following order was found: –*degQ comQ comP comAA comAB –pbpD –kapB kinB patB –mcpB tlpA mcpA tlpB* – (see also Fig. 2). The 62 kb contig contains two *NotI* restriction sites, one at 5513 bp in the *comQPA* region and one at 40318 bp in *yugO* (Fig. 2). These sites correspond to the *NotI* sites at 3249 kb and 3222 kb, respectively, as published in the physical map of Itaya & Tanaka (1991). Furthermore, the 62 kb fragment contains one *SfiI* restriction site in *yufQ* at 15670 bp, corresponding to the *SfiI* at 3240 kb in the physical map of Itaya & Tanaka (1991). Based on the position of these restriction enzyme sites and the published *NotI* and *SfiI* restriction map of *B. subtilis* (Anagnostopoulos *et al.*, 1993), the orientation of the 62 kb contig could be deduced. The *com* region is near the *pai* marker at 284°, whereas the *mcp-tlp* region is closer to *rrnB* at 275° (Fig. 1). The 62 kb contig spans about 5.4° of the chromosome and is most probably located between 277° and 282.5°.

The sequences obtained were analysed for the location of ORFs (Figs 1 and 2, Table 1). In total, 54 ORFs were detected, starting from ATG, GTG or TTG and preceded by a putative RBS (Fig. 2 and Table 1). The predicted genes were designated according to the rules agreed upon by the international consortium sequencing the *B. subtilis* genome. The initiation codon for 44 ORFs is ATG, eight ORFs start with a GTG codon, and two ORFs start with TTG. All these putative genes had reasonable to good RBSs complementary to the 3' end of 16S rRNA of *B. subtilis*, which is 3'-UCUUUCCUCC-ACUAG (McLaughlin *et al.*, 1981), except for *yufT*, *yufN*, *yufK*, *yugE*, *yugH* and *yugU*, which showed rather poor similarity to the consensus RBS. These genes might be poorly expressed. About 87.6% of the DNA sequence encoded putative genes, whereas 12.4% was non-coding. Relatively large non-coding regions are located between *yufA* and *degQ* (about 800 bp), *yufC* and *yufD* (about 500 bp), *yufN* and *yufM* (about 500 bp), *yugI* and *yugJ* (about 520 bp) and *yugU* and *yugV* (about 520 bp). Of the 54 putative genes, 24 (45%) were transcribed in the direction of the origin of replication, and 30 (55%) were transcribed in the other direction.

The amino acid sequences of the newly discovered putative gene products were searched for similarity to sequences reported previously in non-redundant protein

Table 1. Putative ORFs in the 62 kb chromosomal fragment

ORF	Endpoints (nt)	Product (aa)	RBS (bold); initiation codon (<i>italic, bold</i>)
<i>yufA</i>	186 > 1412	409	TTGTAAGGGGGATAAAATCAATG
<i>degQ</i>	2220 > 2357	46	AAAAAGGAGTGTGGAAACGATG
<i>comQ</i>	2545 > 3441	299	TGTCCGGATCAAGGAGAAAATG
<i>comP</i>	3614 > 5920	769	ATTAATAGGTGGATTAATATTG
<i>comAA</i>	6004 > 6645	214	GAGTAAAAGGGAGGAAAACATG
<i>comAB</i>	6667 > 7044	126	AATGTTGGGGGGTGTAGAGATG
<i>yufB</i>	7089 < 7460	124	AAAGGAGAGATTATCGAAAATG
<i>yufC</i>	7447 < 7728	94	TACAGGAGGTGAGCCGGTGATG
<i>yufD</i>	8249 < 9688	480	ATGAGGGGAAATGATCAACATG
<i>yufV</i>	9684 < 10022	113	CGATTGGAGAGGAAGAATAATG
<i>yufU</i>	10025 < 10453	143	AAAGAGGAGGGGAAAAGCGGTG
<i>yufT</i>	10449 < 12770	774	GGCAAAATACGCAAAAAGAGTG
<i>yufS</i>	13982 > 13294	71	AAGGAAAGGAGCTGGTCATGTG
<i>yufR</i>	13336 < 14679	448	ATACATGGGGGTGTATGACATG
<i>yufQ</i>	14836 < 15792	319	GAAAAAGGGGGCGAACTAAGTG
<i>yufP</i>	15796 < 16839	348	AAGGAAGCGGGGAAAGCAAATG
<i>yufO</i>	16835 < 18364	510	ATAAAGGAGCGGTCAATCAATG
<i>yufN</i>	18443 < 19492	350	ACGAAAAATCGGATTGGCGATG
<i>yufM</i>	19994 < 20698	235	GAACCGAAGGAGGAAAATCATG
<i>yufL</i>	20694 < 22292	533	AAACGTAAGGAAGTTTTATATG
<i>yufK</i>	22413 > 23024	204	GGAATAATTTATCTTGTCATG
<i>yugD</i>	23233 < 23643	137	GGATCGGAGGGGATATGGAGTG
<i>pbpD</i>	23667 < 25538	624	GAATTTAGGAGAAAAGAGAGTG
<i>yugC</i>	25634 < 26701	356	CAAACGGCTTTGTGCAAAGGTG
<i>yugB</i>	27019 > 27633	205	GGAAAGGCAGGTGTAAAAGGTG
<i>kapB</i>	27666 < 28049	128	AAATGCGAGGTGATGACAGATG
<i>kinB</i>	28098 < 29384	429	ACGAAAGGATTTACACGAATG
<i>patB</i>	29513 < 30673	387	TACATGAGGGGGTAAACAGAATG
<i>yugE</i>	30754 > 31017	88	AATTTTATTATAAGCGGTGATG
<i>yugF</i>	31052 < 31870	273	AAAAAGGGAGTGACACAGAATG
<i>yugG</i>	32226 > 32516	97	ACATGAGGGCGTTACGGCCATG
<i>yugH</i>	32520 > 33590	357	AATCGTGGTGTCACCGTAAATG
<i>yugI</i>	33888 > 34277	130	TAACTTAGGAGTGTTATGAATG
<i>yugJ</i>	34824 > 35984	387	AATCTAAGGAGGAGATGTCATG
<i>yugK</i>	36091 > 37260	390	ATCATTTGGAGGTTTACTATG
<i>yugL</i>	37375 > 38727	451	GTTATGGAGGGACAGCAGCATG
<i>yugM</i>	38789 > 39157	123	TAAAATCAGAAAGGTGCTGATG
<i>yugN</i>	39220 > 39621	137	TGAAAAAAGGGGCATTCTGATG
<i>yugO</i>	39812 < 40936	375	TGGAGAAATCCGGTTTTATATG
<i>yugP</i>	41288 > 41962	225	AACAAAGGATGATGAAAAAATG
<i>yugS</i>	41993 > 43297	435	AGGTGGCCCCCTTACCATTTATG
<i>yugT</i>	43424 > 45085	554	ACTGGGAAGGAGTGGCTGTATG
<i>yugU</i>	45178 > 45603	142	ACACAGAACGAATTAAGCGATG
<i>yugV</i>	46134 < 46868	245	AAAACAAGGGGGGCTAAACATG
<i>mcpB</i>	47026 > 49011	662	ACCTGAAGGAGCAAGTGACATG
<i>tlpA</i>	49191 > 51176	662	AAAGAAGGTTCAAAAAGGAAATG
<i>mcpA</i>	51302 > 53287	662	AAAAAAGGAGAAATGAATGATG
<i>tlpB</i>	53404 > 55389	662	AAAGGAGGAGTCAACGAGAATG
<i>yulA</i>	55530 > 57596	689	GTATGGAGGAATATCAAATATG
<i>yulB</i>	57656 > 58429	258	ATGAGGAAGGGCGTTACCCATG
<i>yulC</i>	58437 > 59891	485	TCGAAGAGAGGGTGAAAGGATG
<i>yulD</i>	59912 > 60223	104	GATGAAGAGGTGAAAAGGTTTG
<i>yulE</i>	60252 > 61523	424	GATAAAGGGAGAGAGCCGTATG
<i>yulF</i>	61570 < 62238	223	TACAAAAGAAACAGAAGAGATG

Table 2. Similarity of predicted ORF products to proteins of other organisms

Product	Similar protein in database	Identity observed	BLAST score
YufA	No significant similarities	—	—
YufB	Unknown protein of <i>Sinorhizobium (Rhizobium) meliloti</i>	39–46 %, entire length	116
YufC	PhaF gene product of <i>Sinorhizobium meliloti</i>	24 %, entire length	88
YufD	PhaD gene product of <i>Sinorhizobium meliloti</i>	30–57 %, entire length	249
	NADH-ubiquinone oxidoreductase chain 13 of <i>Paracoccus denitrificans</i>	21–29 %, C-terminal part	262
YufV	Na ⁺ /H ⁺ antiporter system ORF3 of <i>Bacillus alcalophilus</i>	55 %, entire length	305
YufU	Na ⁺ /H ⁺ antiporter system ORF2 of <i>Bacillus alcalophilus</i>	40 %, entire length	263
YufT	Na ⁺ /H ⁺ antiporter system of <i>Bacillus alcalophilus</i>	41–66 %, entire length	1100
YufS	No significant similarities	—	—
YufR	L-Malate permease of <i>Streptococcus bovis</i>	49–59 %, entire length	642
YufQ	Hypothetical protein homologue MG121 of <i>Mycoplasma genitalium</i>	31–51 %, entire length	113
YufP	Ribose transport system permease homologue of <i>Mycoplasma genitalium</i>	26–38 %, N-terminal part	106
YufO	Methylgalactoside permease ATP-binding protein of <i>Mycoplasma genitalium</i>	27–50 %, entire length	305
YufN	CD4 ⁺ T-cell-stimulating antigen of <i>Listeria monocytogenes</i>	55 %, entire length	385
YufM	Hypothetical 26.2 kDa protein of <i>Bacillus megaterium</i> , response regulator	possibly 44–50 %, entire length	339
YufL	Hypothetical 60.6 kDa protein of <i>Escherichia coli</i> , possibly sensor kinase	33 %, entire length	447
YufK	No significant similarities	—	—
YugD	No significant similarities	—	—
YugC	Hypothetical 43.9 kDa protein of <i>Escherichia coli</i> , probably integral membrane protein	29–40 %, entire length	255
YugB	No significant similarities	—	—
YugE	No significant similarities	—	—
YugF	2-Hydroxy-6-oxohepta-2,4-dienoatehydrolase of <i>Synechocystis</i> sp.	28–34 %, entire length	156
YugG	No significant similarities	—	—
YugH	HisC homologue of <i>Bacillus subtilis</i>	38–66 %, entire length	322
	Aspartate aminotransferase of <i>Bacillus stearothermophilus</i>	33–60 %, entire length	233
YugI	Hypothetical 14.2 kDa protein of <i>Bacillus subtilis</i>	57–63 %, entire length	186
YugJ	NADH-dependent butanol dehydrogenase of <i>Clostridium acetobutylicum</i>	32–49 %, entire length	802
YugK	NADH-dependent butanol dehydrogenase of <i>Clostridium acetobutylicum</i>	38–52 %, entire length	839
YugL	Glucose-6-phosphate isomerase A of <i>Bacillus stearothermophilus</i>	82 %, entire length	1887
YugM	No significant similarities	—	—
YugN	No significant similarities	—	—
YugO	No significant similarities	—	—
YugP	No significant similarities	—	—
YugS	YqhB, unknown protein of <i>Bacillus subtilis</i>	39–49 %, entire length	553
YugT	Exo-1,4- α -glucosidase of <i>Bacillus stearothermophilus</i>	64 %, entire length	2024
YugU	Hypothetical protein of <i>Methanococcus jannaschii</i>	50–60 %, entire length	196
YugV	No significant similarities	—	—
YulA	Short-chain alcohol dehydrogenase family of <i>Synechocystis</i> sp.	26–58 %, entire length	305
YulB	Glucocitol operon repressor, regulatory protein GutR of <i>Escherichia coli</i>	26–45 %, entire length	132
YulC	Rhamnulokinase of <i>Escherichia coli</i>	33–45 %, entire length	314
YulD	Hypothetical 12.3 kDa protein of <i>Escherichia coli</i>	55 %, entire length	320
YulE	L-Rhamnose isomerase of <i>Escherichia coli</i>	56–61 %, entire length	1000
YulF	Hypothetical 24.8 kDa protein of <i>Escherichia coli</i>	52 %, entire length	327

databases. The results are summarized in Table 2. Twelve out of 41 predicted ORFs not published before showed no significant similarity to other known gene products. Ten other predicted ORFs showed significant homology to unknown or hypothetical proteins of various organisms. The function of these products remains to be studied. The putative gene product YufT showed strong similarity to a Na⁺/H⁺ antiporter system

of *Bacillus alcalophilus* (Hamamoto *et al.*, 1994), as did YufV and YufU. These three gene products might constitute a transport system similar to that found in *B. alcalophilus*. YufP and YufO showed reasonable homology to a sugar transport system found in *Mycoplasma genitalium* (Fraser *et al.*, 1995). YugJ and YugK might constitute a NADH-dependent butanol dehydrogenase system as suggested by the high homology found to

similar proteins in *Clostridium acetobutylicum* which show a similar gene organization (Walter *et al.*, 1992). A high homology score was also observed for YugL. Based on the homology of this protein to glucose-6-phosphate isomerase A of *Bacillus stearothermophilus* (Tao *et al.*, 1989), YugL might be glucose-6-phosphate isomerase A of *B. subtilis*. The YugT gene product showed very strong homology to exo-1,4- α -glucosidase of *B. stearothermophilus* (Takii *et al.*, 1996), which strongly suggests that YugT possesses similar enzymic activity. Finally, a strong homology was observed for YulE. This protein showed significant similarity to L-rhamnose isomerase of *E. coli* (Moralejo *et al.*, 1993) and might therefore function in rhamnose metabolism in *B. subtilis*.

REFERENCES

- 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.
- Biaudet, V., Samson, F., Anagnostopoulos, C., Ehrlich, S. D. & Bessires, P. (1996). Computerized genetic map of *Bacillus subtilis*. *Microbiology* 142, 2669–2729.
- Bussey, L. B. & Switzer, R. L. (1993). The *degA* gene product accelerates degradation of *Bacillus subtilis* phosphoribosylpyrophosphate amidotransferase in *Escherichia coli*. *J Bacteriol* 175, 6348–6353.
- Fraser, C. M., Gocayne, J. D., White, O. & 26 other authors (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397–403.
- Gish, W. & States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nat Genet* 3, 266–272.
- Hamamoto, T., Hashimoto, M., Hino, M., Kitada, M., Seto, Y., Kudo, T. & Horikoshi, K. (1994). Characterization of a gene responsible for the Na⁺/H⁺ antiporter system of alkalophilic *Bacillus* species strain C-125. *Mol Microbiol* 14, 939–946.
- Hanlon, D. W. & Ordal, G. W. (1994). Cloning and characterization of genes encoding methyl-accepting chemotaxis proteins in *Bacillus subtilis*. *J Biol Chem* 269, 14038–14046.
- Itaya, M. & Tanaka, T. (1991). Complete physical map of the *Bacillus subtilis* 168 chromosome constructed by a gene-directed mutagenesis method. *J Mol Biol* 220, 631–648.
- Kunst, F., Vassarotti, A. & Danchin, A. (1995). Organization of the European *Bacillus subtilis* genome sequence project. *Microbiology* 141, 249–255.
- McLaughlin, J. R., Murray, C. L. & Rabinovich, J. C. (1981). Unique features in the ribosome-binding site sequence of the Gram-positive *Staphylococcus aureus* β -lactamase gene. *J Biol Chem* 256, 11283–11291.
- Messing, J. (1983). New M13 vectors for cloning. *Methods Enzymol* 101, 20–78.
- Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Moralejo, P., Egan, S. M., Hidalgo, E. & Aguilar, J. (1993). Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*. *J Bacteriol* 175, 5585–5594.
- Moszer, I., Glaser, P. & Danchin, A. (1995). *SubtiList*: a relational database for the *Bacillus subtilis* genome. *Microbiology* 141, 261–268.
- Popham, D. L. & Setlow, P. (1994). Cloning, nucleotide sequence, mutagenesis, and mapping of the *Bacillus subtilis* *pbpD* gene, which codes for penicillin-binding protein 4. *J Bacteriol* 176, 7197–7205.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Siranosian, K. J., Ireton, K. & Grossman, A. D. (1993). Alanine dehydrogenase (*ald*) is required for normal sporulation in *Bacillus subtilis*. *J Bacteriol* 175, 6789–6796.
- Sorokin, A., Azevedo, V., Zumstein, E., Galleron, N., Ehrlich, S. D. & Serror, P. (1996). Sequence of the *Bacillus subtilis* chromosome region between the *serA* and *kdg* loci cloned in a yeast artificial chromosome. *Microbiology* 142, 2005–2016.
- Takii, Y., Takahashi, K., Yamamoto, K., Sogabe, Y. & Suzuki, Y. (1996). *Bacillus stearothermophilus* ATCC12016 α -glucosidase specific for α -1,4 bonds of maltosaccharides and alpha-glucans shows high amino acid sequence similarities to seven α -D-glucosylhydrolases with different substrate specificity. *Appl Microbiol Biotechnol* 44, 629–634.
- Tao, W. F., Wang, L. L., Shen, R. Q. & Sheng, Z. J. (1989). Complete nucleotide sequence of two phosphoglucosomerase isozymes from *Bacillus stearothermophilus*. *Nucleic Acids Res* 17, 10107–10108.
- Trach, K. A. & Hoch, J. A. (1993). Multisensory activation of the phosphorelay initiation sporulation in *Bacillus subtilis*: identification and sequence of the protein kinase of the alternate pathway. *Mol Microbiol* 8, 69–79.
- Walter, K. A., Bennet, G. N. & Papoutsakis, E. T. (1992). Molecular characterization of two *Clostridium acetobutylicum* ATCC 824 butanol dehydrogenase isozyme genes. *J Bacteriol* 174, 7149–7158.
- Weinrauch, Y., Guillen, N. & Dubnau, D. A. (1989). Sequence and transcription mapping of *Bacillus subtilis* competence genes *comB* and *comA*, one of which is related to a family of bacterial regulatory determinants. *J Bacteriol* 171, 5362–5375.
- Weinrauch, Y., Penchev, R., Dubnau, E., Smith, I. & Dubnau, D. (1990). A *Bacillus subtilis* regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two-component signal-transduction systems. *Genes Dev* 4, 860–872.
- Weinrauch, Y., Msadek, T., Kunst, F. & Dubnau, D. A. (1991). Sequence and properties of *comQ*, a new competence regulatory gene of *Bacillus subtilis*. *J Bacteriol* 173, 5685–5693.
- Yang, M., Ferrari, E., Chen, E. & Henner, D. J. (1986). Identification of the pleiotropic *sacQ* gene of *Bacillus subtilis*. *J Bacteriol* 166, 113–119.

Received 31 January 1997; revised 14 April 1997; accepted 21 April 1997.