The 25°–36° region of the *Bacillus subtilis* chromosome: determination of the sequence of a 146 kb segment and identification of 113 genes

Kunio Yamane, Miyuki Kumano and Kenji Kurita

Author for correspondence: Kunio Yamane. Tel/Fax: +81 298 53 6680.
e-mail: kyamane@sakura.cc.tsukuba.ac.jp

Institute of Biological
Sciences, University of
Tsukuba, Tsukuba-shi,
Ibaraki 305, Japan

We determined a 146 kb contiguous sequence at the 25°–36° region of the *Bacillus subtilis* chromosome containing the *amyE*–*srfA* segment. Among the 113 ORFs identified, 33 are already known. Functions were assigned to 38 ORFs by a search of non-redundant protein sequence data banks and those of 16 ORFs were suggested through significant similarity with reported sequences. The amino acid sequences of 13 of the ORFs were similar to proteins of unknown function of *Escherichia coli*, *Haemophilus influenzae* and other species. We did not find similarities for 29 ORFs to any known proteins. The 146 kb region is rich in enzymes (35 ORFs) related to the metabolism of low molecular mass compounds and five genes for surfactin production occupy about 26 kb of the region.

Keywords: *Bacillus subtilis*, genome sequencing, *amyE*–*srfA* region

INTRODUCTION

To understand the biochemical mechanisms of living cells, all the genes of an organism should be determined. Several projects to sequence genomes have been initiated and the results of those of *Saccharomyces cerevisiae* (Oliver *et al.*, 1992; Dujon *et al.*, 1994; Murakami *et al.*, 1995), *Haemophilus influenzae* (Fleischman *et al.*, 1995) and *Synechocystis* (Kaneko *et al.*, 1996) have been reported. *Bacillus subtilis* is the most extensively characterized Gram-positive bacterium. This organism is important for basic research as a model for cell differentiation, as well as for application to the industry for the production of extracellular enzymes, nucleotides, vitamins and other compounds. As participants in the co-operative project between European and Japanese laboratories to determine the entire nucleotide sequence of the *B. subtilis* genome, we have been analysing a 200 kb region between 20° and 36° of the genetic and physical maps of the chromosome, which were published by Anagnostopoulos *et al.* (1993) and Itaya & Tanaka (1991). The genes for *amyE*, *outB*, *narB*, *srfA* and *gerK* are in this area. However, only a few genes of this region have been sequenced and these sequences have been reported piecemeal by our group and other groups.

In this paper, we describe a 146 kb contiguous nucleotide sequence around 25°–36°, of which about 80 kb is new sequence, and in which *opuA* (GenBank: U17292), *amyE* (V00101, K00563), *outB* (A26936), *tmrB* (D00349), *aroI* (D16321), *narB* (D30689), *srfA* (D30769, D13262, X70356, X77636) and pNEXT59 (D38161) are included.

METHODS

Bacterial strain, plasmids and phage. *B. subtilis* 168 (*trpC2*) from our laboratory stock was used as the standard strain for sequencing. Its chromosomal DNA was extracted by the phenol/pH 9 method of Saito & Miura (1963). Plasmids pTUBE24 and pTUBE1051 were constructed by inserting the *aroI* gene and the *amyE* gene, respectively, from the DNA of a recombinant temperate *B. subtilis* phage ϕ 11-AA18 containing the *amyE*–*aroI* 28 kb fragment (Yamazaki *et al.*, 1983), into a *B. subtilis*–*Escherichia coli* shuttle vector consisting of pUB110 and pBR322 (Otozai & Yamane, 1985; Nakane *et al.*, 1994). Plasmids pMMN159 and pMMN148, and λ DASH II clones λ -110, λ -193, λ -21, λ -333 and λ -74 have already been reported (Ogawa *et al.*, 1995b; Fujishima & Yamane, 1995; Fuma *et al.*, 1993; Akagawa *et al.*, 1995).

Sequence determination. We constructed overlapping libraries of short fragments obtained by sonic oscillation (30 W, 30 s) as described by Ogasawara *et al.* (1994) and Ogawa *et al.* (1995a). Nucleotides from the fragments of the libraries were determined using a Model 373 DNA Sequencer (Perkin Elmer) with the *Taq* Dye Primer Cycle Sequencing Kit and the *Taq* DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystems).

The GSD/EMBL/DBJ/NCBI accession number of the nucleotide sequence reported in this paper is D50453.

Preparation of DNA fragments of the 25°–36° region of the *B. subtilis* chromosome for sequencing

Analysis of tunicamycin-resistant mutations (*tmrA* and *tmrB*) (Hashiguchi *et al.*, 1986; Harada *et al.*, 1988) had indicated that the *amyE*-*aroI* region consists of a 5.0, a 2.3 and a 15 kb *Eco*RI fragment. To clone the three fragments, we prepared λ DASH II libraries of the *B. subtilis* chromosomal DNA after digestion with *Eco*RI. Two types of plaques were isolated from the libraries using 32 P-labelled pTUBE24 and pTUB4 containing *amyE* (Yamazaki *et al.*, 1983) as the probes. One hybridized only

with pTUBE24 and the other with both. DNA cloned in the former type of recombinant phages contained the 5.0 kb fragment and that in the latter the 15 kb *EcoRI* fragment. The former type phage was designated λ -5.0. The latter, however, could not be amplified to provide enough DNA for sequencing. Therefore, we first determined the nucleotide sequences of the 5.0 kb *EcoRI* fragment in λ -5.0, pTUBE1051 and pTUBE24. We then prepared four DNA fragments F1 (2.0 kb), F2 (1.8 Kb), F3 (3.5 Kb) and F4 (3.0 kb) from around the *amyE-aroI* region using PCR (Saiki *et al.*, 1985) and inverse PCR (Triglia *et al.*, 1988) using region-specific 18-mer nucleotide primers and *ampliTag* DNA polymerase (Perkin Elmer) and sequenced them. F5 (10 kb) and F6 (4 kb) DNA fragments, between *aroI* and *narB* and between *nasA* and the 5' flanking region of *srfAA*, respectively, as well as five DNA fragments [F7 (8 kb), F8 (7 kb), F9 (16 kb), F10 (6 kb) and F11 (7 kb)] between *sfp* and pNEXT59, were prepared by PCR and inverse PCR using *rTth* DNA polymerase XL (Perkin Elmer) and region-specific 25-mer nucleotide primers (Fig. 1). This sequence determination indicated that the cloned fragment in pTUBE24 was formed by a 17103 bp deletion between position 1458 (T) and 18560 (C) within two direct repeat

Fig. 1. Cloned and sequenced regions for determining the 146 kb contiguous sequence at the 25°–36° region of the *B. subtilis* chromosome. Solid bars, cloned DNA regions in plasmids and λ DASH II vectors; dashed solid bar, the deleted region in pTUBE24; open bars, DNA regions prepared by PCR using region-specific primers; hatched bars, DNA regions prepared by inverse PCR using region-specific primers; shaded bar, the DNA region reported by Cosmina *et al.* (1993). Locations of the probe DNAs (*amyE*, *arol*, *narB*, pNEXT13, pNEXT61 and pNEXT59) are shown.

Table 1. Putative ORFs in the 146 kb sequence around the 25°–36° region of the *B. subtilis* chromosome

ORF	Endpoints (nt)	Size of product (aa/kDa)	SD consensus sequence (upper case) and initiation codon (bold)
<i>opuAA</i>	220 > 1473	418/46.4	GGAGGgaaaat ATG
<i>opuAB</i>	1478 > 2323	282/30.2	AGGAGGTgaaataag ATG
<i>opuAC</i>	2326 > 3204	293/32.2	AGGgGGAatgcctg ATG
<i>ambX</i>	3229 < 4395	389/42.9	AAGGAGGTatttcttacc ATG
<i>yegA</i>	4787 > 5974	396/41.5	AGGAtgga ATG
<i>yegB</i>	6093 > 6671	193/22.2	GGAGGgaaggtc ATG
<i>amyE</i>	6823 > 8802	660/72.8	AAGGAGTgtcaga ATG
<i>lctE</i>	8982 > 9944	321/34.9	AAGGAtgatta ATG
<i>yegC</i>	10180 > 11601	474/50.2	AAAGcGAtttctactgtt TG
<i>yegD</i>	12132 < 13193	354/39.0	AGGAGGTagaacaa ATG
<i>yegE</i>	13468 > 13770	101/11.3	AAAGcAGGgccg ATG
<i>yegF</i>	14000 > 14473	158/17.8	GGAtGGGctgtac ATG
<i>yegG</i>	14546 > 15304	253/29.2	GGAGGTggatca ATG
<i>yegH</i>	15342 < 16472	377/41.2	AGGGAtcgggtacgtatt TG
<i>yegI</i>	16855 > 17370	172/19.5	AGGtGGTtg ATG
<i>outB</i>	17513 > 18328	272/30.4	AGGAGGaatgattggg ATG
<i>tmrB</i>	18384 < 18974	197/22.7	GGAGGaatggcg ATG
<i>aroI</i>	19250 > 19807	186/21.9	GGAGGatttcgccgaag ATG
<i>yegJ</i>	20289 > 20672	128/15.2	GGcAGGTGatgatgtcaa ATG
<i>yegK</i>	20715 > 21686	324/36.4	GGAGGctgaa ATG
<i>cah</i>	21761 > 22714	318/35.8	AAAGGGGAataatcggaa ATG
<i>yegL</i>	22801 > 23580	260/30.9	GGAGGacgaatagat ATG
<i>yegM</i>	23774 > 24682	303/35.0	GAGGTtgga GTG
<i>yegN</i>	24702 > 26246	515/56.5	AAAGGAGagattatc ATG
<i>yegO</i>	26445 > 27791	449/48.4	GGAGGacgctct TG
<i>yegP</i>	27947 > 29179	411/47.8	AGGAGaaaaaggt ATG
<i>yegQ</i>	29222 < 30076	285/33.2	AAAGGGGTgattgg ATG
<i>yegR</i>	30084 < 30965	294/31.9	GGAGGTtttatt GTG
<i>yegS</i>	31068 < 31835	256/28.7	AAGGGAactccgctctgtc ATG
<i>yegT</i>	32554 > 33090	179/19.7	AAAGGGAaacgc GTG
<i>nasBE</i>	33128 < 34576	483/53.9	GGaAGGAtgac ATG
<i>nasBD</i>	34640 < 34957	106/12.1	AGGAGGaatggac ATG
<i>nasBC</i>	34992 < 37406	805/88.4	AAGGAGcggatcatctg ATG
<i>nasBB</i>	37531 < 39660	710/76.6	GAGGcatt TG
<i>nasBA</i>	39670 < 41979	770/84.8	AAAGGAGGTctgatac ATG
<i>nasA</i>	42159 > 43421	421/46.1	AAAGGGGAatctt ATG
<i>yciA</i>	43480 > 44394	305/34.7	AGGAGTtttctga ATG
<i>yciB</i>	44394 > 44975	194/22.1	AAGGAaGTggataaaat ATG
<i>yciC</i>	45287 > 46477	397/45.3	GGAGGcactatt ATG
<i>yckA</i>	46532 < 47209	226/25.2	AGGAcGTgatttgaatcggacgcc ATG
<i>yckB</i>	47222 < 47788	189/21.0	AAGGAGatgggcattgacggg ATG
<i>yckC</i>	48461 > 48913	151/17.0	GGAGtcaa ATG
<i>yckD</i>	48998 > 49327	110/12.8	GGAGGaataaaatgg ATG
<i>yckE</i>	49484 > 50914	477/55.1	AGGAGtgataaacact ATG
<i>comJ</i>	50957 < 51352	132/15.0	GGAGGaacagcaca TG
<i>comI</i>	51382 < 51690	103/11.5	AAAGAAatgatgaaca ATG
<i>ilpC</i>	51999 < 53717	573/61.8	GGAGGaatt ATG
<i>yckF</i>	53831 < 54385	185/20.0	AAGGAtaactccg ATG
<i>yckG</i>	54498 < 55022	175/19.0	AAAGGgaGTGgatcaaca ATG
<i>yckH</i>	55308 > 55613	102/11.1	AGaAGGAttaacgcttgca GTG
<i>srfAA</i>	56190 > 66953	3588/402	GGAGGTatgacaat ATG
<i>srfAB</i>	66969 > 77729	3587/402	GAGGTggcat ATG
<i>comS</i>	70105 > 70242	46/5.2	AAGGAGGagcagacgt TG
<i>srfAC</i>	77769 > 81590	1274/144	AAGGcGGgacgaaacat ATG
<i>srfAD</i>	81622 > 82347	242/27.6	AAAGGAGGagacagcca ATG
<i>ycx.A</i>	82451 > 83677	409/44.9	AAAGGcAGGTcgaaac ATG
<i>ycx.B</i>	83698 < 84252	185/22.0	GAGGTgtaggtttaccgt ATG

Table 1 (cont.)

ORF	Endpoints (nt)	Size of product (aa/kDa)	SD consensus sequence (upper case) and initiation codon (bold)
<i>ycxC</i>	84 309 < 85 244	312/34.0	AAAGGAGcccttcccatt ATG
<i>ycxD</i>	85 368 > 86 699	444/50.8	AGGgGGAGaaggc ATG
<i>sfp</i>	86 878 < 87 372	165/19.6	GGAGGatctaga ATG
<i>ycxI</i>	88 445 < 89 185	247/27.8	AAAGGAGGTtccgagt ATG
<i>ycxJ</i>	89 202 < 89 903	234/26.4	AAAtGGTcggcgaa ATG
<i>ycxK</i>	89 893 < 90 696	268/29.5	AAAGtGGGGAataag ATG
<i>yclA</i>	90 814 < 91 683	290/33.0	AAGGAGGcgctttct TTG
<i>yclB</i>	91 773 > 92 384	204/22.5	AGGAGtatgattgaa ATG
<i>yclC</i>	92 390 > 93 808	473/50.0	AAAGGAGGagcttgatc ATG
<i>yclD</i>	93 828 > 94 514	229/26.1	AAAGGAtGTtcgaa ATG
<i>yclE</i>	94 583 > 95 425	281/32.7	AAAGGgGGcgtgagaa ATG
<i>yclF</i>	95 471 < 96 946	492/53.3	AAAGGGTttgctgac ATG
<i>yclG</i>	97 226 > 98 977	584/62.8	AAGGAGGctcgaa TTG
<i>gerKA</i>	99 343 > 100 974	544/60.4	AAGGAaAGGTatttgt ATG
<i>gerKC</i>	100 967 > 102 187	407/45.8	AAGGGGAtcaga ATG
<i>gerKB</i>	102 215 > 103 333	373/41.8	AAAGGGGAaGTgacag ATG
<i>yclH</i>	103 444 < 104 121	226/25.2	AAaGAGGcggaacct ATG
<i>yclI</i>	104 140 < 105 597	486/51.9	GGAGGAacataa ATG
<i>yclJ</i>	105 810 > 106 490	227/26.5	GGAGGatata ATG
<i>yclK</i>	106 480 > 107 898	473/55.0	GGAGtaggggtataat TTG
<i>rapC</i>	108 064 > 109 209	382/45.6	GGAGAGGgtgaagga ATG
<i>phrC</i>	109 196 > 109 315	40/4.2	AAAGGAGtGaAGGTttat ATG
<i>yclM</i>	109 859 < 111 220	454/49.7	AGGAGatgtaaac ATG
<i>yclN</i>	111 605 > 112 552	316/34.9	AAGGAGGcagc ATG
<i>yclO</i>	112 548 > 113 492	315/35.5	AAGGAGaaaagcct ATG
<i>yclP</i>	113 489 > 114 244	252/28.5	AAAGGAGaataaatc ATG
<i>yclQ</i>	114 269 > 115 219	317/34.8	GAGGTgaggagc ATG
<i>ycnB</i>	115 272 < 116 687	472/50.8	AAAGaAGGgatatgta TTG
<i>ycnC</i>	116 710 < 117 585	292/33.5	AAAGaAGGAagccatgag ATG
<i>ycnD</i>	117 752 < 118 498	249/27.9	AAAGGAaGGgtgtctcat ATG
<i>ycnE</i>	118 518 < 118 802	95/10.9	GGAGGagatttac ATG
<i>ycnF</i>	119 261 < 120 697	479/55.2	AAGGAGaaactacaaa ATG
<i>ycnG</i>	120 804 > 122 111	436/48.6	AGGgGGAtattcac ATG
<i>ycnH</i>	122 183 > 123 568	462/50.3	AAAGGAatgttcaat ATG
<i>ycxE</i>	123 694 > 124 554	287/30.9	AAAGaAGGTtccccaca ATG
<i>gdb</i>	124 577 > 125 359	261/28.1	GGAGGAGGatgtat ATG
<i>ycnI</i>	125 410 < 126 021	204/22.1	AAGGAGtgtttacg TTG
<i>ycnJ</i>	126 037 < 127 659	541/59.9	AAGGtGGTgtctgagt ATG
<i>ycnK</i>	127 697 < 128 266	190/21.4	AAGagGGTgtatga ATG
<i>ycnL</i>	128 431 > 128 781	117/13.0	GAGGaataaccct ATG
<i>mltA</i>	128 957 > 130 786	610/65.5	GGAGGgtttatat ATG
<i>mltD</i>	130 842 > 131 939	366/40.5	GAGGTgaactgac ATG
<i>ycsA</i>	132 414 > 132 947	178/19.7	ATG
<i>sipU</i>	133 240 > 133 800	187/21.2	AAGGAag TTG
<i>ycsC</i>	134 107 > 134 550	148/17.5	AGGAccgGGcAGGctggcat GTG
<i>ycsD</i>	134 555 > 135 064	170/19.5	AAAGGAGatataaa ATG
<i>ycsE</i>	135 270 > 136 016	249/28.1	AAAGaGAGGTacatat ATG
<i>ycsF</i>	136 224 > 136 856	211/22.7	GGAGGgattct TTG
<i>ycsG</i>	137 259 > 138 221	321/40.3	AGGAaattgcgaat ATG
<i>ycsI</i>	138 248 > 139 018	257/28.1	GGAGcaattaacaaa ATG
<i>ycsJ</i>	139 065 > 140 798	578/53.8	GAGGTgagaaag ATG
<i>ycsO</i>	140 817 > 141 554	246/27.3	AAacgGGAtgaatg ATG
<i>ycsK</i>	141 632 > 142 270	213/23.6	GGAGGaaagaac ATG
<i>ycsL</i>	142 462 > 142 773	104/11.8	ATG
<i>pbpC</i>	143 132 > 145 135	668/74.4	AAGGcAGGgaaagtc ATG
<i>ycsN</i>	145 240 > 146 139	300/34.1	GAGGgatttgc ATG

Table 2. Similarity of the predicted ORF products to other proteins in the data banks and identified genes

ORF	Description	BLAST score	Reference*
OpuAA	Glycine betaine/L-proline transport ATPase	2078	Kempf & Bremer (1995)
OpuAB	Glycine betaine/L-proline transmembrane protein	1394	Kempf & Bremer (1995)
OpuAC	Glycine betaine-binding protein precursor	1513	Kempf & Bremer (1995)
AmhX	Amidohydrolase	–	Kempf & Bremer (1996)
YcgA	Hypothetical protein HI10594 of <i>H. influenzae</i>	420	PIR: E64010
YcgB	Unknown	–	
AmyE	α -Amylase	3496	Yang <i>et al.</i> (1983)
		2946	Yamazaki <i>et al.</i> (1983)
LctE	L-Lactate dehydrogenase	1564	Hediger <i>et al.</i> (1986)
YcgC	L-Lactate permease of <i>E. coli</i>	575	PIR: A49904
YcgD	Multidrug resistance protein (EmrB) homologue of <i>H. influenzae</i>	297	PIR: A64101
YcgE	Unknown	–	
YcgF	Unknown	–	
YcgG	Unknown	–	
YcgH	Aromatic amino acids transport protein of <i>E. coli</i>	218	SP: AROP_ECOLI
YcgI	Unknown	–	
OutB	Spore outgrowth factor B	1396	Albertini <i>et al.</i> (1987)
TmrB	Tunicamycin resistance protein	1056	Harada <i>et al.</i> (1988)
AroI	Shikimate kinase	979	Nakane <i>et al.</i> (1994)
YcgJ	Unknown	–	
YcgK	Regulatory protein OxyR of <i>E. coli</i>	238	PIR: RGECOX
Cah	Cephalosporin C deacetylase	1656	Mitsushima <i>et al.</i> (1995)
YcgL	Unknown	–	
YcgM	Unknown	–	
YcgN	68 % identity to 1-pyrroline-5-carboxylate dehydrogenase of <i>B. subtilis</i> but different function	1908	PIR: S39731
YcgO	Proline permease of <i>E. coli</i>	736	GP: ECOPUTPA_1
YcgP	Hypothetical protein HI0093 of <i>H. influenzae</i>	128	PIR: E64142
YcgQ	Unknown	–	
YcgR	Unidentified protein of <i>E. coli</i>	115	GP: ECOUW67_81
YcgS	Unknown	–	
YcgT	Thioredoxin reductase of <i>Mycoplasma genitalium</i>	118	GPU: MGU39689_7
NasBE	S-Adenosyl-L-methionine:uroporphyrinogen III methyltransferase	2515	Ogawa <i>et al.</i> (1995b)
NasBD	Subunit of nitrite reductase	566	Ogawa <i>et al.</i> (1995b)
NasBC	Subunit of nitrite reductase	4158	Ogawa <i>et al.</i> (1995b)
NasBB	Catalytic subunit of nitrate reductase	3509	Ogawa <i>et al.</i> (1995b)
NasBA	Electron transfer subunit of nitrate reductase	3985	Ogawa <i>et al.</i> (1995b)
NasA	Nitrate transporter	2184	Ogawa <i>et al.</i> (1995b)
YciA	Unknown	–	
YciB	Unknown	–	
YciC	Nitrile hydratase region 3' hypothetical protein P47K of <i>Pseudomonas chlororaphis</i>	416	PIR: D42725
YckA	Glutamine permease of <i>H. influenzae</i>	295	PIR: H64181
YckB	FliY of <i>E. coli</i>	144	GP: ECU18539_3
YckC	Unknown	–	
YckD	Unknown	–	
YckE	β -Glucosidase of <i>Bacillus circulans</i>	457	PIR: A48969
ComJ	Competence-specific protein	690	Vosman <i>et al.</i> (1988)
ComI	DNA-entry nuclease	560	Vosman <i>et al.</i> (1988)
TlpC	Methyl-accepting chemotaxis protein of <i>B. subtilis</i>	2743	PRF: 2106276A Hanlon <i>et al.</i> (1994)
YckF	Unknown	–	

Table 2 (cont.)

ORF	Description	BLAST score	Reference*
YckG	Unknown	—	
YckH	Hypothetical 17·6 kDa protein in <i>rplI-cpdB</i> intergenic region of <i>E. coli</i>	162	SP: YTFH_ECOLI
SrfAA	Surfactin production	18 547	Fuma <i>et al.</i> (1993) Cosmina <i>et al.</i> (1993)
SrfAB	Surfactin production	18 410	Fuma <i>et al.</i> (1993) Cosmina <i>et al.</i> (1993)
ComS	Regulation of genetic competence	241	D'Souza <i>et al.</i> (1994)
SrfAC	Surfactin production	6593	Cosmina <i>et al.</i> (1993)
SrfAD	Surfactin production	1275	Cosmina <i>et al.</i> (1993)
YcxA	K05B2.5 gene product of <i>Caenorhabditis elegans</i>	120	GP: CELK05B2_5
YcxB	Unknown	—	
YcxC	Hypothetical protein of <i>Buchnera aphidicola</i>	115	PIR: S36430
YcxD	<i>mocR</i> gene product of <i>Rhizobium meliloti</i>	128	PIR: 2106332F
Sfp	Surfactin production	861	Nakano <i>et al.</i> (1992)
YckI	Glutamine transport ATP-binding protein GlnQ of <i>E. coli</i>	497	SP: GLNQ_ECOLI Rodriguez & Grandi (1995)
YckJ	Glutamine transport system permease protein GlnP of <i>E. coli</i>	225	SP: GLNP_ECOLI Rodriguez & Grandi (1995)
YckK	Glutamine-binding periplasmic protein precursor GlnH of <i>E. coli</i>	150	SP: GLNH_ECOLI Rodriguez & Grandi (1995)
YclA	<i>als</i> operon regulatory protein AlsR of <i>B. subtilis</i>	191	SP: ALSR_BACSU
YclB	Phenylacrylic acid decarboxylase PAD1 of yeast	550	SP: PAD1_YEAST
YclC	Hypothetical protein of <i>E. coli</i>	297	SP: YIGC_ECOLI
YclD	Unknown	—	
YclE	Unknown	—	
YclF	Di-tripeptide transporter Dtp of <i>Lactococcus lactis</i>	515	SP: DTPT_LACLA
YclG	Unknown	—	
GerKA	Spore germination protein GerKA	2687	Irie <i>et al.</i> (1993)
GerKC	Spore germination protein GerKC	2105	Irie <i>et al.</i> (1993)
GerKB	Spore germination protein GerKB	1867	Irie <i>et al.</i> (1993)
YclH	Hypothetical protein in a rapamycin synthesis gene cluster of <i>Streptomyces hygroscopicus</i>	298	GP: SHGCPIR_4
YclI	Hypothetical protein in a rapamycin synthesis gene cluster of <i>Streptomyces hygroscopicus</i>	269	GP: SHGCPIR_5
YclJ	Alkaline phosphatase synthesis transcriptional regulatory protein PhoP of <i>B. subtilis</i>	165	SP: PHOP_BACSU
YclK	Alkaline phosphatase synthesis sensor protein PhoR of <i>B. subtilis</i>	104	SP: PHOR_BACSU
RapC	Aspartate phosphatase response regulator	—	Perego <i>et al.</i> (1996)
PhrC	Phosphatase regulator	—	Perego <i>et al.</i> (1996)
YclM	Aspartokinase 2 α and β subunits LysC of <i>B. subtilis</i>	227	SP: AK2_BACSU
YclN	Ferric anguibactin transport system permease protein FatD of <i>Vibrio anguillarum</i>	697	SP: FATD_VIBAN
YclO	Ferric anguibactin transport system permease protein FatC of <i>Vibrio anguillarum</i>	454	SP: FATC_VIBAN
YclP	Iron(III) dicitrate transport ATP-binding protein FecE of <i>E. coli</i>	208	SP: FECE_ECOLI
YclQ	Ferric anguibactin-binding protein precursor FatB of <i>Vibrio anguillarum</i>	232	SP: FATB_VIBAN
YcnB	Multidrug resistance protein B EmrB of <i>E. coli</i>	209	SP: EMRB_ECOLI
YcnC	Unknown	—	
YcnD	NADPH-flavin oxidoreductase Frp of <i>Vibrio harveyi</i>	175	PRF: 2014289A
YcnE	Unknown	—	
YcnF	Regulatory protein MocR of <i>Rhizobium meliloti</i>	219	PIR: S43169
YcnG	4-Aminobutyrate aminotransferase GabT of <i>E. coli</i>	589	SP: GABT_ECOLI

Table 2 (cont.)

ORF	Description	BLAST score	Reference*
YcnH	Succinate semialdehyde dehydrogenase GabD of <i>E. coli</i>	963	SP: GABD_ECOLI
YcxE	Hypothetical 30.5 kDa protein <i>gdb1</i> 5' region of <i>Bacillus megaterium</i>	594	SP: YCXE_BACME
Gdh	Glucose dehydrogenase	1153	Lampel <i>et al.</i> (1986)
YcnI	Unknown	–	
YcnJ	Copper export protein PcoD of <i>E. coli</i>	68	PIR: S52256
YcnK	Unknown	–	
YcnL	Unknown	–	
MltA	Mannitol transport protein of <i>Bacillus stearothermophilus</i>	1332	GP: BSU18943_1
MltD	Mannitol-1-phosphate dehydrogenase of <i>Streptococcus mutans</i>	358	SP: MTLD_STRMU
YcsA	Tartrate dehydrogenase of <i>Pseudomonas putida</i>	160	PIR: S48460
SipU	Signal peptidase	–	J. M. van Dijl (personal communication)
YcsC	Unknown	–	
YcsD	Unknown	–	
YcsE	Hypothetical 29.7 kDa protein in <i>ihpA</i> – <i>gyrB</i> of <i>E. coli</i>	182	SP: YIDA_ECOLI
YcsF	Lactam utilization protein LamB homologue of <i>H. influenzae</i>	314	PIR: H64138
YcsG	Branched chain amino acid transport system II carrier protein BraB homologue of <i>H. influenzae</i>	1247	PIR: G64138
YcsI	Unidentified protein of <i>Schizosaccharomyces pombe</i>	245	PIR: S55479
YcsJ	Hypothetical protein HI1731 of <i>H. influenzae</i>	399	PIR: C64041
YcsO	Pectin degradation repressor KdgR of <i>Erwinia chrysanthemi</i>	233	PIR: S17713
YcsK	Unknown	–	
YcsL	Unknown	–	
PbpC	Penicillin-binding protein III (PBP3)	–	C. E. Buchanan (personal communication)
YcsN	Aryl-alcohol dehydrogenase of <i>Phanerochaete chrysosporium</i>	100	Murray <i>et al.</i> (1996) PIR: A55449

* SP, SwissProt; GP, GenBank; GPU, GenBank-UPD.

sequences of ¹⁴⁵¹ATCCTTCTGCACAGGAGGTGAA¹⁴⁷² at *opuA.A* and ¹⁸⁵⁵⁴ATCCTTCAACACAGCGGTCAA¹⁸⁵⁷⁴ (identical nucleotide residues are underlined) at *tmrB*, which is located in front of *aroI*.

Assignment of ORFs

Putative ORFs in the 25°–36° region were searched in all six possible translation frames. Based on the length of a sequence starting from ATG, GTG or TTG and preceded by a Shine–Dalgarno (SD) sequence, we identified 113 putative ORFs (Tables 1 and 2). Among them, 53 had been reported previously. The initiation codon for 95 ORFs was ATG, for 5 GTG and for 13 TTG. No suitable SD sequences were found for *ycsA* and *ycsL*. The orientation of the transcription and translation of 75 ORFs was identical to the movement of the replication fork. The organization of the 113 ORFs is summarized in Fig. 2.

Identification of previously reported genes

Among the 113 putative ORFs, the functions of the products of 33 genes have been reported previously.

These genes are shown in Tables 1 and 2, and Fig. 2. The predicted amino acid sequence of LctE showed 97.5% identity with that of the lactate dehydrogenase of *B. subtilis*, which was determined by amino acid sequencing analysis by Hediger *et al.* (1986). We predicted that LctE consisted of 321 aa. We compared this with the reported sequence of 316 aa. There are two consecutive AUG codons at the translation initiation site preceding the SD sequence. The translation initiator f-Met will be processed out during the maturation of the enzyme. Four amino acid residues at the COOH terminus (position 316–319) were absent in the determined sequence. In addition, 4 aa residues were different.

The genes for surfactin synthesis (*srfAA*, *srfAB*, *srfAC* and *srfAD*) and a regulator of its production (*sfp*) occupied about 26 kb. The *ycsG* and *ycsH* genes described in our previous paper (Akagawa *et al.*, 1995) composed a single gene. Therefore it was designated as *ycsG*.

Similarity of the gene products to known proteins

The amino acid sequences of the putative product of each ORF were compared with the non-redundant protein data

Fig. 2. Organization of the chromosomal region from 25° to 36°. Genes shown above the scale are those whose transcription and translation are in the same direction as the movement of the replication fork, and those shown below are transcribed in the opposite direction. ■, functions of gene products already known, except for *nasBD* and *comS*, the functions of which are known but are shown by white boxes; ▨, the products show high similarity to known proteins in the data banks (BLAST scores above 400); ▩, considerable similarity (BLAST scores between 200 and 399); ▪, very low similarity (BLAST scores below 199); □, gene products with no similarity to known proteins.

banks and the results are summarized in Table 2. Among the 80 putative genes searched, except for genes whose functions were already known, the predicted amino acid sequences of 16 were highly similar (BLAST score higher than 400) to known proteins in the data banks, those of 18 showed considerable similarity (BLAST scores between 400 and 200) and those of 17 very low similarity (BLAST scores less than 200) to known proteins. Among the 51 gene products with similarity to known proteins in the data banks, four products (YcgR, YckH, YclC and YcsE) were similar to proteins of unknown function of *E. coli*, three (YcgA, YcgP and YcsJ) to proteins of unknown function of *H. influenzae* and six (YcxA, YcxC, YclH, YclI, YcxE and YcsI) to unidentified proteins of other organisms. Twenty-nine gene products showed no similarity to known proteins in the data banks. Therefore, the functions of 42 gene products remain unknown.

The 146 kb region is rich in genes of enzymes related to the metabolism of low molecular mass compounds (35 gene products out of 112). Nineteen gene products seemed to function in the transport of amino acids, organic acids, sugars and ions. Three ABC transporters for glycine betaine/L-proline (OpuAA, OpuAB and OpuAC) (Kempf & Bremer, 1995), glutamine (YckI, YckJ and YckK) (Rodriguez & Grandi, 1995) and ferric ions (YclN,

YclO, YclP and YclQ) were found. Furthermore, permeases and transport proteins for L-lactate (YcgC), aromatic amino acids (YcgH), proline (YcgO), nitrate (NasA), glutamine (YckA), di-tripeptides (YclF), copper export (YcnJ), mannitol (MltA) and branched amino acids (YcsG) were predicted.

Ten gene products seem to be regulatory proteins (YcgK, ComS, YcxD, Sfp, YclA, YclJ, YclK, YclL, YcnF and YcsO). YclJ has similarity to PhoP of *B. subtilis* (40% identity) and to PhoB of *E. coli* (41%), and YclK to PhoR of *B. subtilis* (23%) and of *E. coli* (23%). PhoP and PhoR of *B. subtilis* (Seki *et al.*, 1988), and PhoB and PhoR of *E. coli* (Makino *et al.*, 1986) are two component signal transduction regulators of the phosphate regulon. Additionally, four genes for drug resistance [multidrug resistance (*ycgD* and *ycnB*), tunicamycin (*tmrB*) and cephalosporin C (*cah*)], four for spore outgrowth (*outB*, *gerKA*, *gerKB* and *gerKC*) and three for genetic competence (*comJ*, *comI* and *comS*) were located in the region. No genes related to spore formation, except for *gdb*, were found.

Location of the *tmrA* mutation

The *tmrA*7 mutation induces α -amylase hyperproduction by amplifying a 16 kb unit fragment including *amyE* and

tmrB, but not *aroI*, on the *B. subtilis* chromosome (Hashiguchi *et al.*, 1986). The 5' end of the amplification unit corresponds to position 3332 (G residue) at *ambX* and the 3' end of the unit to 18993 (A residue) at the upstream region of *tmrB* (Fig. 2). Therefore, a segment of 15662 bp has been amplified by the *tmrA7* mutation. This amplification might have been induced by an abnormal recombination of the chromosome during mutagenesis as a rare event and tunicamycin resistance seems to be acquired by *tmrB* amplification, which was caused by the abnormal recombination event.

ACKNOWLEDGEMENTS

The authors thank N. Foster for critically reading the manuscript. This study was supported by a Grant-in-Aid for Creative Research (Human Genome Project) from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- Akagawa, E., Kurita, K., Sugawara, T., Nakamura, K., Kasahara, Y., Ogasawara, N. & Yamane, K. (1995). Determination of a 17 484 bp nucleotide sequence around the 39° region of the *Bacillus subtilis* chromosome and similarity analysis of the products of putative ORFs. *Microbiology* **141**, 3241–3245.
- Albertini, A. M., Caramori, T., Henner, D., Ferrari, E. & Galizzi, A. (1987). Nucleotide sequence of the *outB* locus of *Bacillus subtilis* and regulation of its expression. *J Bacteriol* **169**, 1480–1484.
- 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.
- Cosmina, P., Rodriguez, F., de Ferra, F., Grandi, G., Perego, M., Venema, G. & van Sinderen, D. (1993). Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Mol Microbiol* **8**, 821–831.
- D'Souza, C., Nakano, M. M. & Zuber, P. (1994). Identification of *comS*, a gene of the *srfA* operon that regulates the establishment of genetic competence in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **91**, 9397–9401.
- Dujon, B. and others (1994). Complete DNA sequence of yeast chromosome XI. *Nature* **369**, 371–378.
- Fleischmann, R. D. and others (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
- Fujishima, Y. & Yamane, K. (1995). A 10 kb nucleotide sequence at the 5' flanking region (32°) of *srfA* of the *Bacillus subtilis* chromosome. *Microbiology* **141**, 277–279.
- Fuma, S., Fujishima, Y., Corbell, N., D'Souza, C., Nakano, M. M., Zuber, P. & Yamane, K. (1993). Nucleotide sequence of 5' portion of *srfA* that contains the region required for competence establishment in *Bacillus subtilis*. *Nucleic Acids Res* **21**, 93–97.
- Hanlon, D. W., Rosario, M. M. L., Ordal, G. W., Venema, G. & van Sinderen, D. (1994). Identification of TlpC, a novel 62 kDa MCP-like protein from *Bacillus subtilis*. *Microbiology* **140**, 1847–1854.
- Harada, S., Yoda, K., Mori, M., Tanimoto, A., Furusato, T., Yamane, K., Takatsuki, A., Yamasaki, M. & Tamura, G. (1988). The gene responsible for tunicamycin resistance, *tmrB*, in *Bacillus subtilis*. *Agric Biol Chem* **52**, 1785–1789.
- Hashiguchi, K., Tanimoto, A., Nomura, S., Yamane, K., Yoda, K., Harada, S., Mori, M., Furusato, T., Takatsuki, A., Yamasaki, M. & Tamura, G. (1986). Amplification of the *amyE*–*tmrB* region on the chromosome in tunicamycin-resistant cells of *Bacillus subtilis*. *Mol Gen Genet* **204**, 36–43.
- Hediger, M. A., Frank, G. & Zuber, H. (1986). The primary structure of the mesophilic lactate dehydrogenase from *Bacillus subtilis*. *Biol Chem Hoppe-Seyler* **367**, 891–903.
- Irie, R., Fujita, Y. & Okamoto, T. (1993). Cloning and sequencing the *gerK* spore germination gene of *Bacillus subtilis* 168. *J Gen Appl Microbiol* **39**, 453–456.
- 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.
- Kaneko, T. and others (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* **3**, 109–136.
- Kempf, B. & Bremer, E. (1995). OpuA, an osmotically regulated binding protein-dependent transport system for the osmoprotectant glycine betaine in *Bacillus subtilis*. *J Biol Chem* **270**, 16701–16713.
- Kempf, B. & Bremer, E. (1996). A novel amidohydase gene from *Bacillus subtilis*: cloning, DNA-sequencing analysis and map position of *ambX*. *FEMS Microbiol Lett* **141**, 129–137.
- Lampel, K. A., Uratani, B., Chaudhry, G. R., Ramaley, R. F. & Rudikoff, S. (1986). Characterization of the developmentally regulated *Bacillus subtilis* glucose dehydrogenase gene. *J Bacteriol* **166**, 238–243.
- Makino, K., Shinagawa, H., Amemura, M. & Nakata, A. (1986). Nucleotide sequence of the *phoR* gene, a regulatory gene for the phosphate regulon of *Escherichia coli*. *J Mol Biol* **192**, 549–556.
- Mitsushima, K., Takimoto, A., Sonoyama, T. & Yagi, S. (1995). Gene cloning, nucleotide sequence, and expression of a cephalosporin C deacetylase from *Bacillus subtilis*. *Appl Environ Microbiol* **61**, 2224–2229.
- Murakami, Y. and others (1995). Analysis of the nucleotide sequence of the chromosome VI from *Saccharomyces cerevisiae*. *Nat Genet* **10**, 261–268.
- Murray, T., Popham, D. & Setlow, P. (1996). Identification and characterization of pbpC, the gene encoding *Bacillus subtilis* PBP3. *Proceedings of the 12th International Spore Conference*.
- Nakane, A., Ogawa, K., Nakamura, K. & Yamane, K. (1994). Nucleotide sequence of the shikimate kinase gene (*aroI*) of *Bacillus subtilis*. *J Ferment Bioeng* **77**, 312–314.
- Nakano, M. M., Corbell, N., Besson, J. & Zuber, P. (1992). Isolation and characterization of *sfp*: a gene that functions in the production of lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. *Mol Gen Genet* **232**, 313–321.
- Ogasawara, N., Nakai, S. & Yoshikawa, H. (1994). Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. *DNA Res* **1**, 1–14.
- Ogawa, K., Akagawa, E., Nakamura, K. & Yamane, K. (1995a). Determination of a 21 548 bp nucleotide sequence around the 24° region of the *Bacillus subtilis* chromosome. *Microbiology* **141**, 269–275.
- Ogawa, K., Akagawa, E., Yamane, E., Sun, Z.-W., LaCelle, M., Zuber, P. & Nakano, M. M. (1995b). The *nasB* operon and *nasA* gene are required for nitrate/nitrite assimilation in *Bacillus subtilis*. *J Bacteriol* **177**, 1409–1413.

- Oliver, S. G. and others (1992).** The complete DNA sequence of yeast chromosome III. *Nature* **357**, 38–46.
- Otozai, K. & Yamane, K. (1985).** Instability of a chimeric plasmid containing *Bacillus subtilis* *aroI*⁺ *tmrB* genes, pBR322 and pUB110, during *B. subtilis* transformation. *Agric Biol Chem* **49**, 351–357.
- Perego, M., Glaser, P. & Hoch, J. A. (1996).** Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in *Bacillus subtilis*. *Mol Microbiol* **19**, 1151–1157.
- Rodriguez, F. & Grandi, G. (1995).** An operon encoding a novel ABC-type transport system in *Bacillus subtilis*. *Microbiology* **141**, 1781–1784.
- Saiki, R. K., Sharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985).** Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354.
- Saito, H. & Miura, K. (1963).** Preparation of transforming deoxyribonucleic acid by phenol treatment. *Bichim Biophys Acta* **72**, 619–629.
- Seki, T., Yoshikawa, H., Takahashi, H. & Saito, H. (1988).** Nucleotide sequence of the *Bacillus subtilis* *phoR* gene. *J Bacteriol* **170**, 5935–5938.
- Triglia, T., Peterson, M. G. & Kemp, D. J. (1988).** A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res* **16**, 8186.
- Vosman, B., Kuiken, G., Kooistra, J. & Venema, G. (1988).** Transformation in *Bacillus subtilis*: involvement of the 17 kilodalton DNA-entry nuclease and the competence specific 18 kilodalton protein. *J Bacteriol* **170**, 3703–3710.
- Yamazaki, H., Ohmura, K., Nakayama, A., Takeichi, Y., Otozai, K., Yamasaki, M., Tamura, H. & Yamane, K. (1983).** α -Amylase genes (*amyR2* and *amyE*) from an α -amylase hyperproducing *Bacillus subtilis*: molecular cloning and nucleotide sequence. *J Bacteriol* **156**, 327–337.
- Yang, M., Galizzi, A. & Henner, D. (1983).** Nucleotide sequence of the amylase gene from *Bacillus subtilis*. *Nucleic Acids Res* **11**, 237–249.
-
- Received 18 April 1996; revised 26 June 1996; accepted 8 July 1996.