Sequencing of a 65 kb region of the *Bacillus subtilis* genome containing the *lic* and *cel* loci, and creation of a 177 kb contig covering the *gnt*–*sacXY* region

Ken-ichi Yoshida, Katsuhiko Shindo, Hidetoshi Sano, Shin Seki, Miyuki Fujimura, Nobuo Yanai, Yasuhiko Miwa and Yasutaro Fujita

Author for correspondence: Yasutaro Fujita. Tel: +81 849 36 2111. Fax: +81 849 36 2023.
e-mail: yfujita@bt.fubt.fukuyama-u.ac.jp

Department of
Biotechnology, Faculty of
Engineering, Fukuyama
University, Higashimura-cho,
Fukuyama-shi, Hiroshima
729-02, Japan

Within the framework of an international project for the sequencing of the entire *Bacillus subtilis* genome, this paper communicates the sequencing of a chromosome region containing the *lic* and *cel* loci (65 kb), which creates a 177 kb contig covering the region from *gnt* to *sacXY*. This 65 kb region contains 64 ORFs (62 complete and two partial genes). The 14th, 15th and 17th genes correspond to *licT*, *licS* and *katE*, encoding the antiterminator for *licS* transcription, β -glucanase (lichenase) and catalase 2, respectively. The 11th, 30th, 36th, 39th, 41st, 45th–48th, 51st and 58th genes are designated *dead*, *pepT*, *galE*, *aldY*, *msmX*, *cydABCD*, *sigY* and *katX* because their products probably encode ATP-dependent RNA helicase, tripeptidase, UDP-glucose 4-epimerase, aldehyde dehydrogenase, multiple sugar-binding transport ATP-binding protein, the respective components of cytochrome *d* ubiquinol oxidase and ATP-binding cassette transporter, σ -factor of RNA polymerase and catalase, respectively. The 60th–64th genes are *celRABCD*, which are probably involved in cellobiose utilization. Gene organization and gene features in the *gnt*–*sacXY* region are discussed.

Keywords: *Bacillus subtilis*, genome sequencing, *lic* locus, *katE* locus, cellobiose operon

INTRODUCTION

Within the framework of an international project to sequence the entire *Bacillus subtilis* genome, we are responsible for the sequencing of a region between *gnt* (344°) and *sacXY* (333°). In our previous communications on this project, we reported the sequencing and gene features of the 36 kb region between the *gnt* and *iol* operons (Yoshida *et al.*, 1995b), the 15 kb region containing the *iol* operon (Yoshida *et al.*, 1994), the 23 kb region between the *iol* and *hut* operons (Yoshida *et al.*, 1995c), and the 29 kb region containing the *hut* operon and the *wapA* locus (Yoshida *et al.*, 1995a).

Here we describe the sequencing and gene features of the 65 kb region containing the *lic* and *cel* loci, creating a 177 kb contig from *gnt* to *sacXY*.

METHODS

Bacterial strains, phage and plasmids. *B. subtilis* strain 1A1 (*trpC2*), obtained from the *Bacillus* Genetic Stock Center (Ohio), was the internationally agreed sequencing strain. Phage vector, λ gt10, and *Escherichia coli* strain NM514 [*hsdR514 argH galE galX lyaB7 strA* (Hfr⁺)] were obtained from Stratagene. Plasmids pSOFT2 (Itaya & Tanaka, 1991) and pIC336, a pBluescript II SK[–] (Stratagene) derivative, were provided by M. Itaya (Mitsubishi Kasei Institute of Life Sciences, Tokyo) and D. Le Coq (Institut National Agronomique Paris-Grignon, France), respectively.

Preparation of a *B. subtilis* genomic λ gt10 library. A *B. subtilis* genomic library was prepared as follows. The chromosomal DNA of *B. subtilis* 1A1 was isolated and partially digested with *Sau3AI* to yield DNA fragments of approximately 7 kb. The *EcoRI* sites of the resulting *Sau3AI* DNA fragments were modified by *M.EcoRI* methylase (Takara). After filling in with the Klenow fragment of DNA polymerase I, the modified DNA fragments were ligated with *EcoRI* linkers and trimmed with *EcoRI*. Fragments ranging from 6 to 8 kb, which had been electrophoretically purified, were ligated to the arms of λ gt10

The GSDB, DDBJ, EMBL and NCBI accession number for the nucleotide sequence (65 143 bp) reported in this paper is D83026.

Gene	Endpoints (bp)		Size (aa)	Translational start*
<i>wapA</i> (N17G†)‡	1	529		
<i>yxxG</i> (N17H†)	594	1019	142	AAAtGAGGaGgaaat atg
<i>yxiF</i> (N17I†)	1348	1815	156	gcGGAGGGaatgcggctc atg
<i>yxiG</i> (N17J†)	2231	2644	138	AAaGAGGcagggg atg
<i>yxiH</i> (N17K†)	2770	3096	109	tAGGAGcataacca atg
<i>yxiI</i> (N17L†)	3639	4124	162	AgGGATGaGatt atg
<i>yxiJ</i>	4155	4448	98	AgGGAGagGatcaac gtg
<i>yxiK</i>	4467	4925	153	AAGGgGaaGatttt atg
<i>yxiL</i>	5037	5234	66	AAcGAGtTttctttctatatgaaagaag atg
<i>yxiM</i>	5420	6565	382	gAGGgGGAgtcttg atg
<i>deaD</i>	6650	8086	479	cAaGAGacaggagtttatac atg
<i>yxiO</i>	9394	8111	428	ggaGAGGgGctctg atg
<i>yxiP</i>	9601	10278	226	AgaGAGGAGacagagc atg
<i>licT</i>	10377	11207	277	ttGGAGGgacatgcc atg

Table 1 (cont.)

Gene	Endpoints (bp)		Size (aa)	Translational start*
<i>licS</i> (<i>bglS</i>)	11 506	12 231	242	AAtGAaagGggaatgccaat atg
<i>yxiQ</i>	13 672	12 395	426	ggGGAGGTcatat catg
<i>katE</i>	13 873	15 930	686	ggGGAGcaGac atg
<i>yxiS</i>	15 983	16 270	96	AAGGAGcatgacaa atg
<i>yxiT</i>	16 418	16 657	80	AAGGAGtaGattgt atg
<i>yxiA</i>	18 324	17 134	397	tAGGAGtTGctatc atg
<i>yxiB</i>	18 940	19 785	282	AATGgGGTGtttttat catg
<i>yxiC</i>	19 942	21 303	454	AcGGgGGatgtgtgt gtg
<i>yxiD</i>	21 385	22 098	238	AAGGgGcatcaat atg
<i>yxiE</i>	22 098	22 745	216	gaaagGGTGaggcca agtg
<i>yxiF</i>	22 765	23 535	257	ggGaAGGTGtttgac atg
<i>yxiG</i>	24 724	23 723	334	AAGaAGGaattggacaac atg
<i>yxiH</i>	26 109	25 090	340	AAaGAGGcGaataac atg
<i>yxiI</i>	26 852	26 367	162	AAGGAGacttcagca atg
<i>yxiJ</i>	27 244	26 984	87	ttaGAGGgacagtttc ttg
<i>pepT</i>	28 998	27 769	410	AAGGAGaTGattggaa atg
<i>yxiL</i>	29 704	29 051	218	AAGGAGcaaaaag atg
<i>yxiM</i>	30 911	29 694	406	AAGaAGGaccagc atg
<i>yxiN</i>	31 402	30 932	157	AgaaAGGatttgatgac atg
<i>yxiO</i>	32 355	31 483	291	AAGGAGaTatttc atg
<i>yxkA</i>	32 485	32 988	168	AgaGAGGgattcat atg
<i>galE</i>	33 106	34 122	339	tAGGAGGgtttacatc atg
<i>yxkC</i>	34 841	34 203	213	tAGGAGGaaatata atg
<i>yxkD</i>	35 310	36 143	278	AAGGAGtggtttcgtg atg
<i>aldY</i>	37 645	36 191	485	AAGGAGGgtcatott atg
<i>yxkF</i>	37 829	38 719	297	AgaGAGGaGactct atg
<i>msmX</i>	38 843	39 937	365	AAGGgGGatgttag atg
<i>yxkH</i>	40 047	40 883	279	AAGcAGGaGaacg atg
<i>yxkI</i>	42 846	41 161	562	AgaaAGGaattctggc atg
<i>yxkJ</i>	44 320	42 971	450	ggGGAGGTttttca atg
<i>cydA</i>	44 876	46 279	468	AtGGAGGaGaaagc atg
<i>cydB</i>	46 266	47 279	338	cAGGAGGTatacc atg
<i>cydC</i>	47 282	48 982	567	AAGGAGcctatgacttatta atg
<i>cydD</i>	48 982	50 706	575	ctGGgGGaaagagc atg
<i>yxkO</i>	50 795	51 622	276	gAGGAGGgatagag atg
<i>yxIA</i>	53 013	51 643	457	AtGGAGGacgatgcatt g
<i>sigY</i>	53 225	53 758	178	AAGGgGGgatcacggtcatt g
<i>yxIC</i>	53 754	54 071	106	AAGGAGtgGatg atg
<i>yxID</i>	54 071	54 274	68	tgcGAGGTGaaacgcg atg
<i>yxIE</i>	54 274	54 459	62	AAGGgGGatcaagca atg
<i>yxIF</i>	54 469	55 353	295	AAGGtaagccgat atg
<i>yxIG</i>	55 353	56 129	259	AAGGtGGTGcacgc atg
<i>yxIH</i>	57 327	56 128	400	AAaGgGGTGgagagatac atg
<i>katX</i>	59 075	57 435	547	AAGGAGGgccaaccatt g
<i>yxIJ</i>	59 204	59 791	196	AAGGAGagGtcgattc gtg
<i>celR</i>	60 144	62 066	641	tAGGAGGTttttctc atg
<i>celA</i>	62 197	62 502	102	tAGGgGGTcttttac atg
<i>celB</i>	62 521	63 876	452	AgaGAGGgGcgggctat gtg
<i>celC</i>	63 898	64 227	110	AAGGAaGTGaatcgc gtg
<i>celD</i> †	64 227	65 143		AAaGgGGaGcaagcat atg

* Putative initiation codons are shown in boldface. The nucleotides conforming to the consensus Shine–Dalgarno sequence (AAGGAGGTG) are shown in uppercase.

† Former gene designations (Yoshida *et al.*, 1995a).

‡ Partial ORFs.

Table 2. Results of comparison with protein databases of 37 genes in the *lic-cel* region

Gene	Similar protein (organism)*	Database: entry†	Percentage identity (over aa)
<i>deaD</i>	ATP-dependent RNA helicase DeaD (<i>Escherichia coli</i>) (<i>Klebsiella pneumoniae</i>)	SP:DEAD_ECOLI SP:DEAD_KLEPN	43.1 (390) 42.8 (390)
<i>yxjO</i>	Hypothetical 58.8 kDa protein in <i>GLK1-STE50</i> intergenic region (<i>Saccharomyces cerevisiae</i>)	SP:YCD8_YEAST	25.4 (197)
<i>licT</i>	β -Glucoside operon antiterminator (<i>Erwinia chrysanthemi</i>) (<i>Escherichia coli</i>)	SP:ARBG_ERWCH SP:BGLG_ECOLI	54.2 (273) 42.0 (276)
<i>licS</i>	β -Glucanase precursor (<i>Bacillus amyloliquefaciens</i>) (<i>Bacillus licheniformis</i>)	SP:GUB_BACAM SP:GUB_BACLI	90.4 (239) 84.5 (239)
<i>katE</i>	Catalase (<i>Bacillus firmus</i>) (<i>Escherichia coli</i>)	SP:CAT_BACFI SP:CATE_ECOLI	98.7 (525) 53.4 (684)
<i>yxjA</i>	Nucleoside permease NupC (<i>Bacillus subtilis</i>) (<i>Escherichia coli</i>)	SP:NUPC_BACSU SP:NUPC_ECOLI	37.2 (395) 34.7 (395)
<i>yxjB</i>	Hypothetical 30.4 kDa protein in <i>manZ-cspC</i> intergenic region (<i>Escherichia coli</i>) Mycinamicin-resistance protein MyrA (<i>Micromonospora griseorubida</i>)	SP:YEBH_ECOLI SP:MYRA_MICGR	30.8 (253) 28.9 (291)
<i>yxjD</i>	3-Oxoadipate CoA-transferase subunit A (<i>Pseudomonas putida</i>) Butyrate-acetoacetate CoA transferase subunit A (<i>ctfA</i>) homologue (<i>Haemophilus influenzae</i>)	SP:PCAI_PSEPU PIR:B64092	45.6 (228) 46.0 (200)
<i>yxjE</i>	3-Oxoacid CoA-transferase (<i>Sus scrofa domestica</i>) 3-Oxoadipate CoA-transferase subunit B (<i>Pseudomonas putida</i>)	PIR:A41771 SP:PCAJ_PSEPU	57.5 (214) 47.7 (199)
<i>yxjF</i>	Granaticin-producing polyketide synthase chain 5 (<i>Streptomyces violaceoruber</i>) Putative ketoacyl reductase (<i>Streptomyces coelicolor</i>) Ketoreductase (<i>Streptomyces griseus</i>)	SP:DHK1_STRVN SP:ACT3_STRCO PIR:D55587	36.4 (253) 34.5 (252) 32.4 (253)
<i>pepT</i>	Tripeptidase (<i>Lactococcus lactis</i> subsp. <i>cremoris</i>) (<i>Salmonella typhimurium</i>)	SP:PEPT_LACLC SP:PEPT_SALTY	52.2 (404) 55.2 (402)
<i>yxjL</i>	Transcriptional activator protein DegU (<i>Bacillus brevis</i>) Regulator protein VsrD (<i>Pseudomonas solanacearum</i>)	GP:BACDEGSU_2 PRF:2107259B	37.9 (227) 33.8 (213)
<i>yxjM</i>	Sensor protein UhpB (<i>Salmonella typhimurium</i>) Sensor protein VsrA (<i>Pseudomonas solanacearum</i>)	SP:UHPB_SALTY PIR:S41544	26.1 (314) 23.6 (271)
<i>yxjO</i>	Citrate synthase I repressor (<i>Bacillus subtilis</i>) Hypothetical protein (<i>Bacillus subtilis</i>)	SP:CITR_BACSU PIR:S39744	37.4 (265) 24.6 (285)
<i>galE</i>	UDP-glucose 4-epimerase (<i>Haemophilus influenzae</i>) (<i>Escherichia coli</i>)	SP:GALE_HAEIN SP:GALE_ECOLI	65.6 (337) 57.4 (336)
<i>yxkD</i>	Hypothetical protein JojC (<i>Bacillus subtilis</i>) Hypothetical 30.5 kDa protein in <i>ipi</i> 5'-region (<i>Bacillus subtilis</i>)	PRF:2111327A SP:YITB_BACSU	40.1 (202) 32.3 (269)
<i>aldY</i>	Aldehyde dehydrogenase (<i>Pseudomonas putida</i>) Benzaldehyde dehydrogenase (<i>Pseudomonas putida</i>)	GP:PSEPADA_1 GP:PPU15151_1	45.4 (485) 38.9 (476)
<i>yxkF</i>	Leucine-rich protein (<i>Streptococcus equisimilis</i>)	PIR:S39972	22.8 (263)
<i>msmX</i>	Multiple sugar-binding transport ATP-binding protein MsmK (<i>Streptococcus mutans</i>) Hypothetical protein (<i>Synechocystis</i> sp.)	SP:MSMK_STRMU GP:SSU32936_4	61.0 (377) 51.4 (366)

Table 2 (cont.)

Gene	Similar protein (organism)*	Database: entry†	Percentage identity (over aa)
<i>yxkH</i>	Hypothetical 27.6 kDa protein in <i>hpt-panD</i> intergenic region (<i>Escherichia coli</i>)	SP:YADE_ECOLI	26.4 (193)
<i>yxkJ</i>	Probable organic acid transport protein (<i>Streptococcus bovis</i>)	GP:SBU35658_1	38.3 (428)
	Citrate permease (<i>Leuconostoc mesenteroides</i>)	GP:LEUCITP_1	31.0 (426)
<i>cydA</i>	Cytochrome <i>d</i> ubiquinol oxidase subunit I (<i>Escherichia coli</i>)	SP:CYDA_ECOLI	45.1 (315)
	(<i>Azotobacter vinelandii</i>)	PIR:A38170	45.0 (309)
<i>cydB</i>	Cytochrome <i>d</i> ubiquinol oxidase subunit II (<i>Escherichia coli</i>)	SP:CYDB_ECOLI	38.3 (230)
	Cytochrome <i>d</i> ubiquinol oxidase subunit II homologue (<i>Haemophilus influenzae</i>)	PIR:D64181	38.0 (229)
<i>cydC</i>	AarD protein (<i>Providencia stuartii</i>)	GP:PSU30383_1	32.0 (562)
	ABC membrane transporter component (<i>Escherichia coli</i>)	SP:CYDD_ECOLI	30.9 (569)
<i>cydD</i>	ABC membrane transporter component (<i>Escherichia coli</i>)	SP:CYDC_ECOLI	32.2 (560)
	Transporter ATP-binding cassette homologue SurB (<i>Escherichia coli</i>)	PIR:B49454	29.8 (564)
<i>yxkO</i>	Hypothetical 54.7 kDa protein in <i>psd-amiB</i> intergenic region (<i>Escherichia coli</i>)	SP:YJEF_ECOLI	30.9 (272)
<i>yxlA</i>	Hypothetical 58.0 kDa protein in PET117-CEM1 intergenic region (<i>Saccharomyces cerevisiae</i>)	SP:YEP0_YEAST	23.4 (465)
	Purine-cytosine permease (<i>Saccharomyces cerevisiae</i>)	SP:FCY2_YEAST	24.2 (466)
<i>sigY</i>	RpoE (<i>Mycobacterium leprae</i>)	GP:MLU15180_30	28.5 (144)
	RNA polymerase sigma factor CarQ (<i>Mycobacterium xanthus</i>)	SP:CARQ_MYXXA	22.7 (163)
<i>yxlF</i>	<i>bcrA</i> gene product (<i>Bacillus licheniformis</i>)	GP:BACBCRABC_1	33.8 (302)
	Hypothetical protein (<i>Synechocystis</i> sp.)	GP:SYCSLRB_21	31.7 (306)
<i>katX</i>	Catalase (<i>Deinococcus radiodurans</i>)	GP:DEIPLSX_3	68.1 (511)
	(<i>Listeria seeligeri</i>)	PIR:A40367	52.4 (479)
<i>yxlJ</i>	N-Methylpurine DNA glycosylase (<i>Mus musculus</i>)	PRF:2105215A	34.8 (207)
	3-Methyladenine DNA glycosylase (<i>Arabidopsis thaliana</i>)	GP:ATMAG_1	33.0 (197)
<i>celR</i>	<i>sacPA</i> operon antiterminator (<i>Bacillus subtilis</i>)	SP:SACT_BACSU	21.3 (254)
	<i>sacB</i> operon antiterminator (<i>Bacillus subtilis</i>)	SP:SACY_BACSU	32.7 (101)
<i>celA</i>	Cellobiose phosphotransferase system CelA (<i>Bacillus stearothermophilus</i>)	PIR:B49898	59.2 (98)
	PTS system, cellobiose-specific II B component (<i>Escherichia coli</i>)	SP:PTCB_ECOLI	45.7 (94)
<i>celB</i>	Cellobiose phosphotransferase system CelB (<i>Bacillus stearothermophilus</i>)	PIR:C49898	65.7 (449)
	PTS system, lactose-specific II BC component (<i>Staphylococcus aureus</i>)	SP:PTLB_STAAU	30.9 (431)
	PTS system, cellobiose specific II C component (<i>Escherichia coli</i>)	SP:PTCC_ECOLI	30.5 (416)
<i>celC</i>	Cellobiose phosphotransferase system CelD (<i>Bacillus stearothermophilus</i>)	PIR:E49898	61.5 (104)
	Phosphotransferase system cellobiose-specific enzyme III CelC (<i>Escherichia coli</i>)	GP:ECOCLECI_1	40.0 (100)
<i>celD</i>	6-Phospho- β -glucosidase (<i>Escherichia coli</i>)	SP:CELF_ECOLI	56.8 (368)
	Probable 6-phospho- β -glucosidase (<i>Escherichia coli</i>)	SP:GLVG_ECOLI	36.9 (203)
<i>ywaA</i>	Branched chain amino acid transaminase homologue YJR148w (<i>Saccharomyces cerevisiae</i>)	PIR:S57177	41.2 (371)
	ECA39 gene production (<i>Homo sapiens</i>)	GP:HSU21551_1	41.2 (362)

* Gene products which exhibited significant similarity are listed (FASTA optimized score > 165).

† Abbreviations for the protein databases are as follows: PIR, NBRF-PIR protein sequence database; SP, SwissProt protein sequence database; GP, translated protein sequence database from the NCBI-GenBank nucleotide sequence database; PRF, Protein Research Foundation protein sequence database. Database entries of GP are followed by an underlined space and an ordinal number of each of the coding sequences.

(Stratagene) which had been cleaved with *Eco*RI and treated with calf intestine alkaline phosphatase (Stratagene). The ligation mixture was incubated with an *in vitro* packaging mixture (GIGAPACK II XL, Stratagene). Then, 3×10^8 independent phage particles were amplified by plaque formation on *E. coli* NM514 to yield a *B. subtilis* genomic library (5.2×10^{10} p.f.u. ml⁻¹).

Amplification of chromosomal segment by long PCR. A 30 kb chromosomal segment between *lic* and *cel* (Fig. 1) was amplified by long PCR (Cheng *et al.*, 1994) using the GeneAmp XL PCR Kit and AmpliWax PCR Gem 100 (Applied Biosystems) according to the manufacturer's protocol. For preparation of template DNA, cells of *B. subtilis* 1A1 were lysed by successive treatments with lysozyme and SDS, and the lysate was treated gently with phenol. DNA, precipitated with ethanol, was added as template at a concentration of 1 ng μ l⁻¹. The PCR was performed with a primer pair each containing 20 bases for hybridization (approximately 50% G + C content) in 18 tubes containing 100 μ l each of the reaction mixture, using 32 cycles (15 s at 95 °C, 12 min at 65 °C) with the Perkin Elmer GeneAmp PCR System 9600. The long PCR products from the 18 tubes (approximately 5 μ g each) were mixed, and purified by precipitation with polyethylene glycol [8% (w/v)].

Random sequencing of the target DNA fragments. Each target DNA fragment isolated from restriction enzyme digests of the inserts of λ DASH II and λ gt10 clones and the long PCR product was self-ligated, randomly fragmented with DNase I, and cloned into phage M13mp19 as described previously (Yoshida *et al.*, 1994). The PCR products of the inserts of M13mp19 recombinant clones derived from each fragment were sequenced with a DNA sequencer (373A; Applied Biosystems) using the Dye Primer Cycle Sequencing Kit (Applied Biosystems), and compiled (Yoshida *et al.*, 1994). The resulting sequence islands were aligned with reference to a physical map of each fragment. Gaps between sequence islands were filled in by sequencing PCR products prepared using oligonucleotide primers and chromosome DNA as template (Yoshida *et al.*, 1994).

Computer analysis. The DNA sequences determined were analysed with the GENETYX-MAC 6.2.0 program (Software Development Co.) and the GeneWorks 2.4 program (IntelliGenetics). Similarity search for putative gene products was performed with the FASTA program (Pearson & Lipman, 1988) using a GenomeNet FASTA server (Supercomputer Laboratory, Institute for Chemical Research, Kyoto University; e-mail: fasta@genome.ad.jp).

RESULTS AND DISCUSSION

Cloning (or amplification) and sequencing of the *lic*-*cel* region, creating a 177 kb contig spanning from *gnt* to *sacXY*

Several λ DASH II recombinant clones carrying the *wapA* locus were isolated and their inserts were physically mapped (Yoshida *et al.*, 1995a). To extend the cloned region towards *sacXY*, chromosomal walking was performed by repeated screenings of a *B. subtilis* genomic λ DASH II library (Yoshida *et al.*, 1994) with the region of the phage insert furthest from *hut* among the recombinant clones as a probe. This walking enabled us to isolate 15 recombinant clones carrying different unsequenced regions towards *sacXY*. Among these recombinant clones, two, E2.7-5 and E1.0-6, were selected for systematic

sequencing to extend towards *sacXY* by 20 kb. The locations of the inserts in these clones are shown in Fig. 1, together with their physical maps. Further chromosomal walking from the E1.0-6 clone towards *sacXY* was not successful in spite of several attempts. Fig. 1 also indicates the locations of the 8 kb and 3 kb fragments derived from the E2.7-5 clone and of the 15 kb fragment from the E1.0-6 clone. These fragments were subjected to random sequencing. Assembling the sequences of these fragments resulted in a 26183 bp contiguous sequence, which overlapped the 28954 bp sequence containing *hut* and *wapA* [GSDDB, DDBJ, EMBL and NCBI accession number D31856 (Yoshida *et al.*, 1995a)] by 3704 bp. This region was found to contain the *licS* (*bgIS*) (Murphy *et al.*, 1984) and *licT* (Schnetz *et al.*, 1996) genes, which encode β -glucanase (lichenase) and the antiterminator for *licS* expression, respectively, as well as the *katE* gene encoding a σ^B -dependent stress catalase (Engelmann *et al.*, 1995; K. Yoshida & Y. Fujita, unpublished results). This 26183 bp sequence, together with the published sequences (Yoshida *et al.*, 1994, 1995a, b, c), generated a 121 kb contig spanning from *gnt* to *lic*.

It was extremely difficult to bridge an approximately 40 kb gap between the 121 kb contig and the 97 kb sequence containing *sacXY* and *sacPA* (Glaser *et al.*, 1993), by chromosomal walking using λ DASH II as a vector; the gap size was estimated from the physical map of the *B. subtilis* chromosome (Itaya & Tanaka, 1991). Meanwhile, D. Le Coq provided us with plasmid pIC336, carrying an insert which hybridized with yeast artificial chromosome (YAC) clones (15-8 and 11-418) of an ordered collection of *B. subtilis* DNA segments (Azevedo *et al.*, 1993) (D. Le Coq, personal communication), indicating that this insert is located close to *sacXY*. However, a short stretch of the sequence in this insert (by D. Le Coq, personal communication) was neither in the 97 kb sequence (Glaser *et al.*, 1993) nor in the 121 kb contig from *gnt* to *katE*, suggesting that this insert is located in the 40 kb gap region. However, no positive λ DASH II recombinant clones were obtained when our λ DASH II library was screened with this plasmid. On the assumption that λ gt10, which can carry a shorter insert (less than 7 kb) than λ DASH II (approximately 15 kb), might be a more suitable vector, we constructed another *B. subtilis* genomic library using λ gt10 as a vector. When the λ gt10 library was screened with pIC336, several positive recombinant clones were obtained, the inserts of which were physically mapped and aligned. Among them, two overlapping recombinant clones (GT-2 and GT-3) covering a 12 kb region (Fig. 1) were selected for random sequencing. Connection of the sequences from these clones (7 and 6 kb) resulted in an 11698 bp sequence, 5180 bp of which coincided with nucleotides 7–5186 of a 6863 bp sequence determined by P. Glaser and others (unpublished; accession number Z49992) (Fig. 1). This overlapping region contained a putative operon (*celABCD*), which appears to be involved in cellobiose utilization, and its regulator (*celR*) (M. Hecker, personal communication). As the 6863 bp sequence partially overlapped the 97 kb sequence (Glaser *et al.*, 1993), the

Table 3. 164 genes found in the *gnt-sacXY* region (177 kb)

Region (length)	Reference (accession number)	Genes*
<i>gnt-iol</i> (36 kb)	Yoshida <i>et al.</i> (1995b) (D45242)	<i>gntZ-gntP-gntK-gntR-yxaA(S14A)-yxaB(S14B)-yxaC(S14CR)-yxaD(S14D)-yxaE(S14E)-yxaF(S14F)-yxaG(S14G)-yxaH(S14H)-yxaI(S14IR)-yxaJ(S14J)-yxaK(S14K)-yxaL(S14L)-yxaM(S14MR)-yxaN(VE7AR)-yxbA(VE7BR)-yxbB(VE7CR)-yxbC(VE7D)-yxbD(VE7E)-aldX(VE7FR)-yxbF(E3A)-yxbG(E3BR)-yxcA(E3C)-btpG(SS92A)-yxcC(SS92BR)-yxcD(SS92CR)-yxcE(SS92DR)-iolS(SS92ER)-iolR(SS92FR)-iolA(E83A)</i>
<i>iol</i> (15 kb)	Yoshida <i>et al.</i> (1994) (D14399)	<i>iolB(E83B)-iolC(E83C)-iolD(E83D)-iolE(E83E)-iolF(E83F)-iolG(E83G, idb)-iolH(B65A)-iolI(B65B)-iolJ(B65C)-yxdJ(B65D)-yxdK(B65E)-yxdL(B65F)</i>
<i>dra</i> (23 kb)	Yoshida <i>et al.</i> (1995c) (D45912)	<i>yxdL(B65F)-yxdM(B65G)-yxeA-yxeB-yxeC-yxeD-yxeE-yxeF-yxeG-yxeH-yxeI-yxeJ-yxeK-yxeL-yxeM-yxeN-yxeO-yxeP-yxeQ-yxeR-yxxB-deoR-dra-nupC-pdp</i>
<i>hut-wapA</i> (29 kb)	Yoshida <i>et al.</i> (1995a) (D31856)	<i>hutM(EE57D)-hutG(EE57C)-hutI(EE57B)-hutU(EE57A)-hutH-hutP-yxiA(J3A)-yxiB(J3B)-yxiC(J3C)-yxiD(J3D)-yxxD(N17A)-yxxE(N17B)-bglP(N17C, sytA)-bglH(N17D)-yxiE(N17E)-yxxF(N17F)-wapA-yxxG(N17H)-yxiF(N17I)-yxiG(N17J)-yxiH(N17K)</i>
<i>lic-cel</i> (65 kb)	This work (D83026)	<i>yxxG</i> to <i>celC</i> (62 genes)
<i>cel</i> † (7 kb)	Glaser <i>et al.</i> (Z49992)	<i>celR-celA-celB-celC-celD-ywaA(ipa-0r)</i>
<i>sacA</i> † (97 kb)	Glaser <i>et al.</i> (1993) (X73124)	<i>dltE(ipa-1r)-dltD(ipa-2r)-dltC(ipa-3r)-dltB(ipa-4r)-dltA(ipa-5r)-ywaB(ipa-6d)-ywaC(ipa-7d)-ywaD(ipa-8r)-tyrZ-ywaE(ipa-10r)-ywaF(ipa-11d)-gspA(ipa-12d)-sacX-sacY</i> (78 genes)

* The genes shown in boldface are transcribed in the direction opposite to the movement of the replication fork.

† Also refer to 'SubtiList version 1.5' (Moszer *et al.*, 1995) regarding the function and designation of the genes in the *cel* and *tyrZ-sacA* region.

sequencing of 11 698 bp generated a 29 kb contig spanning *cel* to *sacXY*.

In the *gnt-sacXY* region, only a 30 kb gap between *lic* and *cel* remained to be sequenced. The unsequenced region was amplified by long PCR with a primer pair designed to generate a product flanked by *EcoRI* sites. This product was digested with *EcoRI* to produce 13, 9 and 5 kb fragments and several short fragments (less than 1 kb). The three longer fragments were subjected to random sequencing, and the gaps were filled in by direct sequencing of the PCR products. The resulting sequence (30 557 bp), together with the 121 kb and 29 kb contigs, created a 177 kb contig from *gnt* to *sacXY*.

Genes found in the *lic-cel* region

The current sequencing of 26 183 bp, 11 698 bp and 30 557 bp generated a contig of 65 143 bp of the *lic-cel* region. As shown in Fig. 1 and Table 1, computer analysis of the 65 143 bp sequence revealed 62 complete (*yxxG* to *celC*) and two partial (*wapA* and *celD*) ORFs. The *N17H*, *N17I*, *N17J*, *N17K* and *N17L* genes reported previously (Yoshida *et al.*, 1995a) were renamed *yxxG* and *yxiFGHI*, respectively. The *deaD*, *pepT*, *galE*, *aldY*, *msmX*, *cydABCD*, *sigY* and *katX* genes were named according to the high similarity of their products to known proteins in sequence databases (Table 2). The endpoints of these genes in the sequence, and their size, putative Shine-

Dalgarno sequence and initiation codon, are shown in Table 1. Of the 63 genes (*yxxG* to *celD*), 53 were predicted to use ATG as initiation codon, though *yxiJ*, *yxiL*, *sigY* and *katX* might use TTG, and *yxiJ*, *yxiCE*, *yxiJ* and *celBD* might use GTG. This indicates that GTG is preferred to TTG as initiation codon in the *lic-cel* region, in contrast to a relative frequency reported recently (Glaser *et al.*, 1993; Ogasawara *et al.*, 1994). The direction of transcription of 45 of the 64 genes is the same as that of the movement of the replication fork. This is not well consistent with the other recent observations that transcription tends to be co-directional with replication in regions, such as *lic-cel*, located relatively close to *oriC* (Glaser *et al.*, 1993; Ogasawara *et al.*, 1994; Yoshida *et al.*, 1994, 1995a, c; Zeigler & Dean, 1990).

Similarity search and function of the gene products in the *lic-cel* region

We searched for similarity of the 62 complete gene products (*YxxG* to *CelC*, Table 1) and the *CelD* and *YwaA* proteins (P. Glaser and others, accession number Z49992), with proteins present in a non-redundant protein sequence database (Institute for Chemical Research, Kyoto University), using the FASTA program to obtain the FASTA optimized scores (Pearson & Lipman, 1988). The latest similarity search was performed in December 1995. As shown in Table 2, this search revealed that 37 of the 64

Table 4. Features of 164 genes in the *gnt-sacXY* region

Description		No./164
Transcription		
Co-directional with replication	Yes	108
	No	56
Translation		
Initiation codon usage	ATG	137
	TTG	15
	GTG	12
Size (aa)	< 99	12
	100–199	34
	200–299	40
	300–399	31
	400–499	29
	500–599	11
	600–699	5
	> 700	2
Function*		
Function identified		39
Tentative function†		19
Similarity‡		
To known proteins		28
To functionally unknown proteins		11
No significant similarity		67

* Similarity searches were performed currently and previously (Yoshida *et al.*, 1994, 1995a, b, c).

† Tentative functions were deduced by similarity searches.

‡ FASTA optimized score > 200.

gene products exhibited significant similarity to known proteins.

The DeaD protein showed very high similarity to RNA helicases of various organisms. This protein actually contained the D-E-A-D box (Schmid & Linder, 1992) in its amino acid sequence. The YxjD and YxjE proteins exhibited high similarity to 3-oxoadipate CoA-transferase subunits A and B, respectively, of *Pseudomonas putida*, suggesting that they might be the respective subunits of a transferase of CoA to 3-ketoacylate. The resulting 3-ketoacylate CoA might be reduced by the successive gene product (YxjF), which showed similarity to keto-reductases of various organisms. The PepT protein appears to be a tripeptidase.

The YxjL protein showed high similarity to many proteins belonging to the OmpR subfamily of response regulators of the two-component regulatory systems of bacteria (Stock *et al.*, 1989). The YxjM protein exhibited similarity to histidine protein kinases of the two-component regulatory systems (Stock *et al.*, 1989). The *yxjM* gene is adjacent to *yxjL*, so their gene products seem to constitute a two-component regulatory system, though their function remains to be determined. The GalE, AldY and MsmX proteins are probably a UDP-glucose 4-epimerase, an aldehyde dehydrogenase, and a multiple sugar-binding transport ATP-binding protein, respectively; we pre-

viously found another aldehyde dehydrogenase gene (*VE7FR* was now designated *aldX*; Yoshida *et al.*, 1995b) (Table 3).

The CydABCD proteins exhibited high similarity to *E. coli* cytochrome *d* ubiquinol oxidase subunits I and II (CydA and CydB), and the components of a membrane transporter of *E. coli* belonging to the ABC family (Higgins *et al.*, 1990) (CydD and CydC), respectively. These four genes probably constitute an operon because we found a σ^A -dependent promoter-like sequence immediately upstream of *cydA* and a ρ -independent terminator-like sequence immediately downstream of *cydD*, and because these four genes are adjacent to each other. In contrast, the *E. coli* *cydAB* and *cydDC* genes constitute two separate operons (Fang & Gennis, 1993; Poole *et al.*, 1993). The *sigY* gene was so designated because its product showed similarity to some sigma factors; *sigX* has already been used for the *orfX20* gene in the *spoVA-serA* region of the chromosome (Sorokin *et al.*, 1993; Lonetto *et al.*, 1994). In addition to the *katE* gene, encoding a σ^B -dependent catalase (Engelmann *et al.*, 1995), we found another gene which probably encodes a catalase (*katX*).

Glaser and co-workers cloned and sequenced a 7 kb region containing the probable *cel* operon (*celABCD*) involved in cellobiose utilization (unpublished; accession number Z49992). We also sequenced this *cel* operon cloned in λ gt10 independently, though our clone (GT-3) did not contain the entire *celD* gene (Fig. 1). The CelAB proteins probably encode the phosphoenolpyruvate-dependent phosphotransferase system (PTS) enzyme II B and C components, respectively (Parker & Hall, 1990; Reizer *et al.*, 1990; Lai & Ingram, 1993). The CelC protein probably encodes the PTS system enzyme III. The CelD protein was considered to be a 6-phospho- β -glucosidase. Interestingly, the CelR protein showed significant similarity to the *B. subtilis* SacT and SacY proteins, which are the antiterminators for *sacPA* and *sacB*, respectively (Débarbouillé *et al.*, 1990; Crutz *et al.*, 1990). We found a σ^A -dependent promoter-like sequence immediately upstream of *celR*, and ρ -independent terminator-like sequences immediately downstream of *celD* as well as *celR*. The *celABCD* genes are closely adjacent to each other. Thus, we predict that the *cel* operon might include not only the *celABCD* genes but also the *celR* gene, and that the antiterminator (CelR) might regulate this operon positively through interaction with the terminator between *celR* and *celA*. Hence, transcript analysis of the *cel* operon for elucidation of its organization is awaited.

Characterization of gene organization and gene features in the *gnt-sacXY* region (177 kb)

We were able to create a contig of 177 kb covering the *B. subtilis* chromosome *gnt-sacXY* region. The 164 genes found in this region are listed in Table 3. Many of these genes were renamed according to the international agreement of the *B. subtilis* genome project, following the rules of Demerec *et al.* (1968). The *SS92A* gene was renamed *htpG* because its product showed very high similarity to *E. coli* HtpG (Yoshida *et al.*, 1995b). The

Table 5. Classification of the 86 genes whose function was identified or suggested

Category	Gene*	<i>gnt-sacXY</i> No./86 (%)	<i>E. coli</i> No./1720 (%)
I. Intermediary metabolism	<u><i>gntZ gntK gntR aldX yxbG iolS iolR</i></u> <u><i>iolA iolB iolC iolD iolE iolG</i></u> <u><i>iolH iolI iolJ yxdJ yxeK yxeP deoR</i></u> <u><i>dra pdp hutG hutI hutU hutH hutP</i></u> <u><i>bglH licT licS katE yxjL yxjM</i></u> <u><i>yxjO aldY cydA cydB cydC cydD</i></u> <u><i>sigY katX celR celD sacY sacX</i></u>	45 (52)	375 (22)
II. Biosynthesis of small molecules	<i>yxaN galE ywaA</i>	3 (3)	319 (19)
III. Macromolecule metabolism	<u><i>deaD yxjD yxjE yxjF pepT yxjI</i></u> <u><i>ywaD tyrZ</i></u>	8 (9)	419 (24)
IV. Cell structure	<i>yxeQ wapA dltE dltD dltC dltB</i> <i>dltA</i>	7 (8)	141 (8)
V. Cellular processes	<u><i>gntP yxcC iolF yxdL yxeB yxeM</i></u> <u><i>yxeN yxeO yxeR nupC hutM bglP</i></u> <u><i>yxjA msmX yxkJ yxjF celA celB celC</i></u>	19 (22)	304 (18)
VI. Other functions	<i>yxaM htpG yxeI gspA</i>	4 (5)	162 (9)

* The 86 genes (function identified, tentative function, and similarity to known proteins; Table 4), are classified according to the categories described by Riley (1993). Genes whose functions were identified experimentally are shown in bold face. Genes whose tentative functions were deduced by similarity searches are underlined.

SS92ER, *SS92FR*, *E83ABCDEFGF* and *B65ABC* genes were renamed *iolSR* and *iolABCDEFGHJ*, respectively, because they were found to be involved in *myo*-inositol catabolism (K. Yoshida & Y. Fujita, unpublished).

Some features of the 164 genes are summarized in Table 4. Of the 164 genes, 127 are transcribed in the same orientation as that of the movement of the replication fork, which is consistent with recent observations that transcription tends to be co-directional with replication in the regions, like the *gnt-sacXY* region, located relatively close to *oriC* (Zeigler & Dean, 1990; Glaser *et al.*, 1993; Ogasawara *et al.*, 1994). In the *gnt-sacXY* region, this tendency was prominent in the genes in the *iol-wapA* region (Yoshida *et al.*, 1994, 1995a, c), but it was not clearly applicable to the genes in the *gnt-ol* and *wapA-sacXY* regions (Yoshida *et al.*, 1995b; this work) (Fig. 1, Table 3). We do not know the reason why this co-directionality fluctuates considerably from region to region, independently of their relative location with respect to *oriC*.

Table 4 also shows the relative initiation codon usage (ATG, TTG and GTG) of the 164 genes found in the *gnt-sacXY* region and the molecular size distribution of their products. Initiation codon usage indicates that the genes in the *gnt-sacXY* region utilize initiation codons with a frequency similar to those reported recently (Glaser *et al.*, 1993; Ogasawara *et al.*, 1994). The molecular size distribution also indicates that the size of most gene products is between 100 and 499 amino acids, with the

most frequent population in the size range from 200 to 299 (Table 4). This molecular size distribution is also similar to that reported recently (Glaser *et al.*, 1993; Ogasawara *et al.*, 1994).

The function of 39 of the 164 genes of the *gnt-sacXY* region was identified experimentally (Table 4). From similarity searches of their gene products, the function of 19 genes was tentatively identified despite the lack of experimental evidence, and that of 28 genes was suggested. The function of 78 genes remained unknown, though 11 products of these genes showed similarity to proteins of unknown function (Table 4). In the *gnt-sacXY* region, there are several catabolic operons such as *gnt*, *iol*, *dra-nupC-pdp*, *hut* and *cel* involved in gluconate, *myo*-inositol, deoxyribonucleoside (Saxild *et al.*, 1996), histidine and cellobiose utilization, respectively. We found in this region two gene pairs (*yxdJK* and *yxjLM*), products of which are likely to constitute two-component regulatory systems (Stock *et al.*, 1989), and five ATP-binding transport proteins [members of ABC family (Higgins *et al.*, 1990); *YxdL*, *YxeO*, *MsmX* and *CydCD*].

We tried to classify the 86 genes, the function of which was identified or suggested, according to the classification of the *E. coli* genes (Riley, 1993) (Table 5). It was one of the notable features of the *gnt-sacXY* region that there was no gene involved in sporulation, germination and competence. Thus, we adopted the categories for the *E. coli* genes to classify the 86 *B. subtilis* genes. As shown in

Table 5, a striking functional characteristic of the *gnt-sacXY* region is that 52% of the genes in this region are likely to be involved in intermediary metabolism (category I), particularly in degradation, in contrast with only 22% of the 1720 *E. coli* genes distributed in the whole chromosome classified into this category. The percentages of the genes involved in biosynthesis of small molecules and macromolecule metabolism (categories II and III) are 3 and 9, respectively, which are very low in comparison with those in the 1720 *E. coli* genes (19 and 24%) (Table 5). These facts imply that the *B. subtilis* chromosome might encode genes not randomly but rather in an orderly way according to their functions; in other words, it tends to be functionally differentiated, depending on the relative location of each gene within the chromosome. This striking feature of the *B. subtilis* chromosome may be clearer when its entire sequencing is completed in the near future.

ACKNOWLEDGEMENTS

We thank the following undergraduate students for their help with the experiments: T. Ihori, M. Nakatsuka, M. Ogimoto, K. Shiraishi, R. Yatagai, Y. Yamada, A. Yamamoto and H. Yoshimura. We are also grateful to D. Le Coq for providing plasmid pIC336, and communicating his unpublished results. We also thank M. Hecker for his personal communication. This work was supported by a Grant-in-Aid for Creative Basic Research on 'Human Genome Analysis' from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- Azevedo, V., Alvarez, E., Zumstein, E., Damiani, G., Sgaramea, 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.
- Cheng, S., Chang, S.-Y., Gravitt, P. & Respass, R. (1994). Long PCR. *Nature* **369**, 684–685.
- Crutz, A. M., Steinmetz, M., Aymerich, S., Richter, R. & Le Coq, D. (1990). Induction of levansucrase in *Bacillus subtilis*: an anti-termination mechanism negatively controlled by the phosphotransferase system. *J Bacteriol* **172**, 1043–1050.
- Débarbouillé, M., Arnaud, M., Fouet, A., Klier, A. & Rapoport, G. (1990). The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators. *J Bacteriol* **172**, 3966–3973.
- Demerec, M., Adelberg, E. A., Clark, A. J. & Hartman, P. E. (1968). A proposal for a uniform nomenclature in bacterial genetics. *J Gen Microbiol* **50**, 1–14.
- Engelmann, S., Lindner, C. & Hecker, M. (1995). Cloning, nucleotide sequence, and regulation of *katE* encoding a σ^B -dependent catalase in *Bacillus subtilis*. *J Bacteriol* **177**, 5598–5605.
- Fang, H. & Gennis, R. B. (1993). Identification of the transcriptional start site of the *cyd* operon from *Escherichia coli*. *FEMS Microbiol Lett* **108**, 237–242.
- Glaser, P., Kunst, F., Arnaud, M., Coudart, M.-P., Gonzales, W., Hullo, M.-F., Ionescu, M., Lubochinsky, B., Marcelino, L., Moszer, I., Presecan, E., Santana, M., Schneider, E., Schweizer, J., Vertes, A., Rapoport, G. & Danchin, A. (1993). *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325° to 333°. *Mol Microbiol* **10**, 371–384.
- Higgins, C. F., Hyde, S. C., Mimmack, M. M., Gileadi, U., Gill, D. R. & Gallagher, M. P. (1990). Binding protein-dependent transport systems. *J Bioenerg Biomembr* **22**, 571–592.
- 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.
- Lai, X. & Ingram, L. O. (1993). Cloning and sequencing of a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. *J Bacteriol* **175**, 6441–6450.
- Lonetto, M. A., Brown, K. L., Rudd, K. E. & Buttner, M. J. (1994). Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci USA*, **91**, 7573–7577.
- Moszer, I., Glaser, P. & Danchin, A. (1995). *SubtiList*: a relational database for the *Bacillus subtilis* genome. *Microbiology* **141**, 261–268.
- Murphy, N., McConnell, D. J. & Cantwell, B. A. (1984). The DNA sequence of the gene and genetic control sites for the excreted *B. subtilis* enzyme β -glucanase. *Nucleic Acids Res* **12**, 5355–5367.
- 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.
- Parker, L. L. & Hall, B. G. (1990). Characterization and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. *Genetics* **124**, 455–471.
- Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.
- Poole, R. K., Hatch, L., Cleeter, M. W., Gibson, F., Cox, G. B. & Wu, G. (1993). Cytochrome *bd* biosynthesis in *Escherichia coli*: the sequences of the *cydC* and *cydD* genes suggest that they encode the components of an ABC membrane transporter. *Mol Microbiol* **10**, 421–430.
- Reizer, J., Reizer, A. & Saier, M. H., Jr (1990). The cellobiose permease of *Escherichia coli* consists of three proteins and is homologous to the lactose permease of *Staphylococcus aureus*. *Res Microbiol* **141**, 1061–1067.
- Riley, M. (1993). Functions of the gene products of *Escherichia coli*. *Microbiol Rev* **57**, 862–952.
- Saxild, H. H., Andersen, L. N. & Hammer, K. (1996). *dra-nupC-pdp* operon of *Bacillus subtilis*: nucleotide sequence, induction by deoxyribonucleosides, and transcriptional regulation by the *deoR*-encoded DeoR repressor protein. *J Bacteriol* **178**, 424–434.
- Schmid, S. R. & Linder, P. (1992). D-E-A-D protein family of putative RNA helicases. *Mol Microbiol* **6**, 283–291.
- Schnetz, K., Stülke, J., Gertz, S., Krüger, S., Krieg, M., Hecker, M. & Rak, B. (1996). LicT, a *Bacillus subtilis* transcriptional anti-terminator protein of the BglG family. *J Bacteriol* **178**, 1971–1979.
- 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.
- Stock, J. B., Ninfa, A. & Stock, A. M. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol Rev* **53**, 450–490.
- Yoshida, K., Sano, H., Miwa, Y., Ogasawara, N. & Fujita, Y. (1994). Cloning and nucleotide sequencing of a 15 kb region of the *Bacillus subtilis* genome containing the *iol* operon. *Microbiology* **140**, 2289–2298.
- Yoshida, K., Sano, H., Seki, S., Oda, M., Fujimura, M. & Fujita, Y.

(1995a). Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the *hut* and *wapA* loci. *Microbiology* **141**, 337–343.

Yoshida, K., Seki, S., Fujimura, M., Miwa, Y. & Fujita, Y. (1995b). Cloning and sequencing of a 36-kb region of the *Bacillus subtilis* genome between the *gnt* and *iol* operons. *DNA Res* **2**, 61–69.

Yoshida, K., Fujimura, M., Yanai, N. & Fujita, Y. (1995c). Cloning

and sequencing of a 23-kb region of the *Bacillus subtilis* genome between the *iol* and *hut* operons. *DNA Res* **2**, 295–301.

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

.....
Received 4 March 1996; revised 20 May 1996; accepted 28 May 1996.