The *yvsA*–*yvqA* (293°–289°) region of the *Bacillus subtilis* chromosome containing genes involved in metal ion uptake and a putative sigma factor

Anil Wipat,¹ S. Caroline Brignell,¹ B. Jane Guy,¹ Matthias Rose,³
Peter T. Emmerson² and Colin R. Harwood¹

Author for correspondence: Colin Harwood. Tel: +44 191 222 7708. Fax: +44 191 222 7736.
e-mail: colin.harwood@ncl.ac.uk

^{1,2} School of Microbiological, Immunological and Virological Sciences¹ and Department of Biochemistry & Genetics², University of Newcastle upon Tyne, Framlington Place, Newcastle upon Tyne NE2 4HH, UK

³ Johann Wolfgang Goethe-Universitaet Frankfurt, Institut für Mikrobiologie, Marie-Curie-Str., 9 Geb. N250, 60439 Frankfurt am Main, Germany

The region between *yvsA* (293°) and *yvqA* (289°) of the *Bacillus subtilis* chromosome has been sequenced within the framework of the *B. subtilis* 168 international sequencing programme. A primary analysis of the 42 ORFs identified in this 43 kb region is presented. The region included a high proportion of genes that did not show homology with genes in other bacteria. The identified ORFs showed homology to proteins involved in the transport of metal ions, two-component signal transducers, ATP-binding-cassette-type transporters and a sigma factor.

Keywords: *Bacillus subtilis*, genome sequencing, ABC transporter, metal ion utilization, two-component system

INTRODUCTION

Bacillus subtilis is one of the most genetically amenable bacteria and an important source of industrial enzymes. Furthermore, our understanding of the molecular biology of *B. subtilis* is second only to that of *Escherichia coli*. The sequencing of the entire *B. subtilis* genome has recently been completed by an international consortium involving European, Japanese and US research groups (Kunst *et al.*, 1997). With the complete genome sequence, *B. subtilis* is rapidly becoming a model organism for the study of Gram-positive bacteria. The genome sequence of *B. subtilis*, in conjunction with that of *E. coli* (Blattner *et al.*, 1997), will facilitate a detailed analysis of the evolution of two organisms that diverged more than a billion years ago, and the existence of a systematic functional analysis programme for *B. subtilis* (Harwood

& Wipat, 1996) means that the sequence data will be fully exploited.

We have sequenced the 43 kb region between *yvsA* (293°) and *yvqA* (289°). The region encodes genes involved in the utilization and uptake of metal ions, including at least three putative ATP-binding-cassette (ABC)-type transporters, three two-component systems, genes involved in germination and many genes of unidentified function. This study reports the primary analysis of the region and describes the main features of the ORFs which have not been previously sequenced.

METHODS

Bacterial strains, bacteriophages and plasmids. Chromosomal DNA for sequencing was prepared from *B. subtilis* strain 168 (Anagnostopoulos & Spizizen, 1961), kindly provided by C. Anagnostopoulos (INRA, Jouy-en-Josas, France). DNA was cloned into bacteriophage LambdaGEM-11 (Promega) or the general purpose vector pUC18 (Vieira & Messing, 1987), as described previously (Wipat *et al.*, 1996). *B. subtilis* and *E. coli* were grown in LB broth and on LB agar.

Cloning, generation of libraries and DNA preparation. A

Abbreviation: ABC, ATP-binding cassette.

The EMBL accession number for the sequence reported in this paper is AJ223978.

Fig. 1. Organization of the *B. subtilis* chromosome from *yvsA* to *yvqA* (293°–289°). The diagram is orientated in the direction of the replication complex (left to right). *NotI* sites are indicated on the diagram. No *SfiI* sites were found. Thick black arrows represent the putative ORFs, together with their direction of transcription, and are named according to the rules described by Demerec et al. (1968). Hatched arrows indicate genes which have been cloned and sequenced previously. The loop structures show putative ρ -independent transcription terminators and the small arrows above them indicate the direction of their termination.

lambda phage library containing chromosomal DNA from *B. subtilis* strain 168 was prepared as described previously (Wipat *et al.*, 1996). Specific chromosomal DNA was amplified for cloning and sequencing by PCR. *B. subtilis* chromosomal DNA, either cloned in lambda or as a long PCR product, was randomly fragmented with DNase I and cloned into pUC18, as described previously (Wipat *et al.*, 1996). All other DNA methods were as described previously (Wipat *et al.*, 1996).

DNA hybridization. DNA fragments were labelled by random primed labelling using the Rediprime system (Amersham). DNA hybridization was performed essentially following the procedures of Sambrook *et al.* (1989).

DNA sequencing. pUC18 clones containing *B. subtilis* DNA were sequenced using universal forward and reverse primers. Sequences were obtained using a model 877 Molecular Biology Labstation (Catalyst, Applied Biosystems) and a model 373A or 377A sequencer from Applied Biosystems. DNA sequence was determined on both strands.

Data handling and computer analysis. A Macintosh-based DNA sequence assembly and analysis program, SEQUENCHER (Gene Codes Corp.), was used to assemble the DNA sequence data. The DNA sequence was analysed for ORFs using the Lasergene (DNASTAR) software package. Amino acid homology searches were performed using the BLAST software (Altschul *et al.*, 1990) available on the World Wide Web.

RESULTS AND DISCUSSION

Sequencing and analysis of the region from *yvsA* to *yvqA*

The physical organization of the 42.8 kb region between *yvsA* and *yvqA*, located at approximately 290° on the chromosome, is shown in Fig. 1. Prior to these studies, contigs of 4.6 kb (*fhuD*–*fhuC*; Schneider & Hantke, 1993) and 5.9 kb (*gerAC*–*yuxN*; Miles & Guest, 1985; Zuberi *et al.*, 1987) had been sequenced. However, these regions were resequenced on both strands and some discrepancies indicated on the annotated sequence. The low density of previously mapped genes in this region meant that it was one of the last regions of the *B. subtilis* genome to be sequenced.

Probing a lambda phage library with *fhuD* and *citG* resulted in the identification of two non-overlapping lambda clones of approximately 11 kb that were sequenced in their entirety. Sequences at the ends of these clones facilitated the design of PCR primers to the regions flanking *fhuD* and *citG*. The remainder of the region was therefore sequenced using specific oligonucleotide primers and direct or inverse PCR products as templates.

Analysis of the region for ORFs of at least 60 amino acids in length and with appropriately located putative ribosome-binding sites and initiation codons resulted in the identification of 42 putative ORFs. These range in size from 61 residues (*yvrF*) to 617 residues (*yvgJ*), with a mean length of 306 residues. ORFs account for 90 % of the potential coding capacity of the DNA. The largest

non-coding region, 706 bp in length, is located between the divergent *yvsG* and *yvsH* genes. Sixty-three per cent of the ORFs are orientated in the same direction as the movement of the replication complex, which is slightly lower than that observed for other regions (Glaser *et al.*, 1993; Ogasawara *et al.*, 1994). This might reflect a higher proportion of genes that are expressed under conditions in which the chromosome is not replicating.

The sizes, putative ribosome-binding sites and initiation codons for the newly identified ORFs are shown in Table 1. Eleven putative *p*-independent transcription terminators were identified and their locations and sequences are shown in Fig. 1 and Table 2, respectively. In line with previous observations (Wipat *et al.*, 1996), ATG was the most common start codon (ATG, 70 %; TTG, 15 %; GTG, 15 %) whilst TAA was the most common stop codon (TAA, 51 %; TGA, 28 %; TAG, 21 %).

The number of genes of unknown function in this region is high and the functions of only 8 of the 43 putative genes have been confirmed experimentally. These are *fhuD*, *-B*, *-G* and *-C*, encoding components of a ferrichrome ABC transporter (Schneider & Hantke, 1993); *gerAA*, *-AB* and *-AC*, whose products form a complex at the inner spore membrane required for the germination response to L-alanine (Zuberi *et al.*, 1987); and *citG*, encoding fumarate hydratase (Miles & Guest, 1985).

Homology search and properties of newly identified ORFs

The BLAST algorithm (Altschul *et al.*, 1990) was used to search the protein databases for homologues of the newly identified ORFs. Table 3 shows the homologues identified up to 1 February 1998. The putative protein sequences of seven ORFs, *yvsA*, *yvrL*, *yvrJ*, *yvrF*, *yvqI*, *yvqG* and *yvqF*, show no significant homologies to any entries in the current databases. Two of these ORFs, *yvrJ* and *yvrF*, are contained entirely within, but would be transcribed in the opposite orientation to, genes *yvrI* and *yvrE*, respectively.

Three genes had paralogues in the *B. subtilis* chromosome. The *yvgN* gene product showed homology to that of *ytbE*, located in the *rrnB*–*dnaB* region. Both of these gene products showed similarities to plant-metabolite dehydrogenases, such as morphine dehydrogenase and aldose reductases, from a variety of organisms. The predicted product of the *yvgJ* gene showed a high level of similarity to the products of two other genes of unknown function, *yqgS* and *yfnI*. *yuxN* had been partially sequenced previously (Miles & Guest, 1985) and we extended the N-terminal region of the encoded polypeptide by 213 residues. Its product showed a weak level of similarity with the product of the *ycnC* gene (Yamane *et al.*, 1996).

Table 1. Main features of newly identified ORFs in the *yvsA*–*yvqA* region of the *B. subtilis* genome

ORF	Size (aa)	End points (nt)	Translation start*	Description/comments
<i>yvsA</i>	100	185–487	<u>AGaAGGaagactcc</u> ATG	No significant homologues
<i>yvgN</i>	276	646–1476	<u>AgGGtGGaGgttcaa</u> GTG	Plant-metabolite dehydrogenase
<i>yvgM</i>	186	2067–1507	<u>GctGGaGcatgg</u> ATG	Permease component of ABC transporter
<i>yvgL</i>	260	2953–2171	<u>AaGgGGTGcaatc</u> ATG	Molybdate-binding protein
<i>yvgK</i>	308	3064–3990	<u>AgGagaGTGatgatg</u> ATG	Nif-related protein homologue
<i>yvgJ</i>	617	5871–4018	<u>GAGGaGtcattatg</u> ATG	Hypothetical protein
<i>yvsG</i>	160	6453–5971	<u>AAGGAGaaGcgga</u> ATG	Hypothetical protein
<i>yvsH</i>	469	7160–8569	<u>AtGGAGGTactgga</u> ATG	Arginine/ornithine antiporter
<i>yvrP</i>	397	13009–14202	<u>AAaGAGGaGatcgt</u> GTG	ABC transporter accessory factor
<i>yvrO</i>	230	14186–14875	<u>AAGcAGGTtacagtca</u> ATG	ABC transporter component
<i>yvrN</i>	416	14847–16097	<u>ggaGAGtTGgtattag</u> ATG	Hypothetical protein
<i>yvrL</i>	136	16541–16131	<u>AAGGAGcTGaacca</u> GTG	No significant homologues
<i>yvrK</i>	385	17759–16602	<u>AAGGAGGaaacatttc</u> ATG	Hypothetical protein
<i>yvrJ</i>	86	18668–18408	<u>ttGtcGGgactccccta</u> ATG	No significant homologues
<i>yvrI</i>	190	18191–18763	<u>AgGGAGGcGgtttg</u> ATG	RNA polymerase σ^G factor homologue
<i>yvrH</i>	369	18767–19876	<u>AAGagGGaaaagtaagct</u> GTG	DNA-binding response regulator
<i>yvrG</i>	573	19894–21615	<u>AAGatGGaagttt</u> TTG	Sensor kinase homologue
<i>yvrF</i>	61	22102–21917	<u>gAGGtGGaGacttgatcc</u> TTG	No significant homologues
<i>yvrE</i>	292	21687–22565	<u>AAaGAGGTgaacac</u> ATG	RNA polymerase subunit
<i>yvrD</i>	263	23394–22603	<u>AAGGAatgattctt</u> ATG	Polyketide reductase homologue
<i>yvrC</i>	314	23792–24736	<u>tAGGAGGaaaagatt</u> ATG	Iron transport system binding protein
<i>yvrB</i>	353	24699–25760	<u>AAGGgGtcGaagaac</u> TTG	Haemin permease homologue
<i>yvrA</i>	442	25760–27088	<u>ggGGAGaaGcttatg</u> ATG	Ferrichrome transport ATP-binding protein
<i>yvqK</i>	193	27111–27692	<u>AAGGAGattcagggt</u> ATG	Hypothetical protein
<i>yvqJ</i>	422	27869–29137	<u>AcGGGgGGgctctgatgg</u> ATG	Macrolide efflux determinant
<i>yvqI</i>	126	29299–29679	<u>AAGGAGGatctgc</u> ATG	No significant homologues
<i>yvqH</i>	225	29706–30383	<u>tAGGAGGaattcagg</u> ATG	Membrane protein homologue
<i>yvqG</i>	240	30633–31355	<u>ccGcgGGacattgacagc</u> ATG	No significant homologues
<i>yvqF</i>	241	31389–32114	<u>AAGGAGGcGgacaccagga</u> ATG	No significant homologues
<i>yvqE</i>	360	32111–33193	<u>gtGGAtGTGaagtacgt</u> ATG	Histidine kinase homologue
<i>yvqC</i>	211	33171–33806	<u>AAGGAGagaacgaac</u> GTG	Transcription regulatory protein
<i>yvqB</i>	451	41830–40475	<u>cAGGAtGatgcatc</u> ATG	Sensory protein kinase homologue
<i>yvqA</i>	225	42504–41827	<u>AAGGAtGTGaagagcc</u> TTG	Response regulator homologue

* Underlined letters indicate ribosome-binding site 16S rRNA consensus; bold letters indicate start codons.

Three new putative transporters were located in the region. *yvgM*, *yvgL* and *yvgK* genes appear to be in an operon of 2.5 kbp. Their products are likely to be involved in the transport of molybdenum. YvgM shows homology with the permease component of an ABC transporter protein from *Synechocystis* sp. (Kaneko *et al.*, 1995) and with ModB, a membrane protein of *E. coli* involved in the uptake of molybdenum by the products of the *modABCD* operon (Maupin-Furlow *et al.*, 1995). The putative product of *yvgL* shows a high level of similarity to ModA, the molybdenum-binding protein of *Synechocystis* sp., *E. coli* and *Haemophilus influenzae*. In common with the binding proteins of many other uptake pathways in Gram-positive bacteria, YvgL has a putative lipoprotein signal sequence that would anchor it to the *trans* side of the cytoplasmic membrane (Simonen & Palva, 1993). The putative

product encoded by *yvgK* shows a high level of similarity to a putative Nif-related protein from *Clostridium pasteurianum* (Chen *et al.*, 1986), a significant level of similarity to a hypothetical protein in the ferredoxin [2Fe–2S] gene 5' region of *C. pasteurianum* and to a molybdenum cofactor biosynthesis protein encoded by a *moeA* homologue of *Methanococcus jannaschii* (PIR accession number F64410).

Genes *yvrP*, *yvrO* and *yvrN* appear to be in an operon of approximately 3 kb encoding a putative ABC transporter. The *yvrP* gene product shows a low level of similarity to a putative ABC transporter accessory factor from *Lactobacillus sake* (Axelsson & Holck, 1995). o648 (YbjZ) from *E. coli* (Blattner *et al.*, 1997), which is 648 amino acids in length, shows similarity throughout its length to both YvrO and YvrN. In contrast, Ybba

Table 2. Putative *p*-independent transcription terminators in the *yvsA-yvqA* region of the *B. subtilis* genome

Gene	Putative <i>p</i> -independent transcription terminators*
<i>yvgN</i>	tctgttttAaTcAAAAAACTCCCGgttatGGGAGTTTTTTTtAgTgcgttttt
<i>yvgM</i>	aaaaacgcAcTaAAAAAACTCCCGataacgGGGAGTTTTTTTgAtTaaaacaga
<i>yvgK</i>	gaaAcgtAAAAAATCCCTCTtgatAGAGGGATTTTTTcatTgg
<i>yvgJ</i>	ccAatgAAAAAATCCCTCTatcaAGAGGGATTTTTTtTtc
<i>yvsH</i>	ataAtgAAAAAGACACCTGAtttgcTCAGGTGTCTTTTTtTata
<i>fhuD</i>	tgacTAAATAAAaCAAAAGAGCCGcCtGCCcCGGGCTCTTTTGcTTATTTAataag
<i>fhuC</i>	ctgTAtgAAAAAaCCCGGtTTgAAgCCGGGaTTTTTtTattc
<i>yvrM</i>	gCtAtgaAtAAAAAGAACGCATcCaAtTgGaATGCGTTCCTTTTTtTagcTgGg
<i>yvrL</i>	cCcAgctAaAAAAAGAACGCATcCcAaTtGgATGCGTTCCTTTTTTaTcaTaGc
<i>yvrE</i>	gtataAAAAAAACGCATCCGaacaCGGATGCGTTTTTTTTtatact
<i>yvrD</i>	agtatAAAAAAACGCATCCGgttCGGATGCGTTTTTTTTtatac
<i>yvqK</i>	gtgAaAAggAAAGCCGGCgTGTtcaaatACaTGCCGGCTTTtTTgTttt
<i>yvqJ</i>	aaaGggttttAAAACGCCATGCctcgtGCATGGCGTTTTtttTgtGcaa
<i>citG</i>	aggCgTAaTagGAAGAACGGCTGCTTtttAAGCAGCCGTTCTTCtaAaTAtGttg
<i>yvxN</i>	caaCaTAtTaGAAGAACGGCTGCTTaaaAAGCAGCCGTTCTTCctAtTAcGcct
<i>yvqB</i>	gcagTgccAAAATAGACTGtAgAtgtTtTgCAGTCTATTTTTtttAtgtg

* Upper case letters indicate inverted repeats.

from *H. influenzae* and ClpP of *Synechocystis* sp. are of a similar length and show strong similarity only to YvrO. The product of the *yvrN* gene shows similarity to putative membrane-spanning components of ABC transporters from *Synechocystis* sp. (slr0594), *B. subtilis* (YknZ, Kunst *et al.*, 1997) and the C-terminus of the *E. coli* o648 (Blattner *et al.*, 1997). Thus, while *E. coli* appears to have a single large protein equivalent to ORFs YvrO and YvrN, other organisms encode two smaller proteins.

The N-terminus of YvrN was identical to a previously sequenced polypeptide fragment, YziB, showing signal peptide activity (Smith *et al.*, 1988). Since this protein does not have a recognizable signal sequence, but has membrane-spanning domains, the previously observed signal peptide activity is presumably a consequence of its hydrophobic nature.

In summary, *yvrP*, *yvrO* and *yvrN* appear to encode the components of an ABC transporter in which YvrP is an accessory factor, YvrO is an ATP-binding protein and YvrN a membrane-spanning component. The ORFs of the latter two genes overlap and it is possible that a mechanism of translation slippage operates between these two genes to generate a single protein (Atkins *et al.*, 1991).

The *yvrC*, *yvrB*, *yvrA* and *yvqK* genes appear to be arranged in an operon of 3.9 kb in size encoding another putative ABC transporter. Comparison of the three proteins encoded by *yvrC*, *yvrB* and *yvrA* with the subunits of other ABC-type systems suggests that this is a system involved in metal ion transport. YvrC shows significant homology to iron transport binding proteins from numerous different bacteria. YvrB is homologous

to haemin permeases from bacterial species (Fleischmann *et al.*, 1995) and is probably membrane-located. YvrA shows homology to the product of the *fhuC* gene, an ATP-binding protein (Schneider & Hantke, 1993), which is located approximately 10 kb upstream of *yvrA*, and to other ATP-binding proteins. The product of the fourth putative ORF, YvqK, shows a good level of similarity to a hypothetical protein of *Mycobacterium leprae* (U15186), the function of which is unknown.

Three putative two-component signal transduction pathways were identified, in some cases located within operons of unknown functions. *yvqI*, *yvqH*, *yvqG*, *yvqF*, *yvqE* and *yvqC* appear to be arranged in a single transcriptional unit of 4.5 kb. The putative products of *yvqI*, *yvqG* and *yvqF* have no significant homologues in the current databases. The protein encoded by *yvqH* shows significant similarity to a 30 kDa protein (D64002) found in the chloroplast membranes of *Synechocystis* sp. The *yvqE* and *yvqC* gene products are homologous to a sensor protein kinase (U81487) of *Lactococcus lactis* and a transcription regulatory protein, DegU, of *B. subtilis* (Tanaka & Kawata, 1988). It is therefore likely that *yvqE* and *yvqC* encode a two-component signal transduction pathway.

The putative products of the *yvqB* and *yvqA* genes appear to make up another two-component signal transduction system. YvqB shows significant similarity to a sensor protein kinase from *Streptomyces hygroscopicus* (Schwecke *et al.*, 1995), whilst *yvqA* shows significant homology to the product of the *phoP* gene of *B. subtilis*, which is the response regulator component of the signal transduction system involved in regulating the response of the cells to phosphate starvation.

Table 3. Homologues of newly identified/sequenced ORFs in the *yvsA*–*yvqA* region of the *B. subtilis* genome

No significant homologues were found for *yvsA*, *yvrL*, *yvrJ*, *yvrF*, *yvqI*, *yvqG* and *yvqF*.

ORF	BLAST Score	Genes encoding homologous products	Function/activity of homologue	Reference/database accession number
<i>yvgN</i>	9.8e ⁻¹⁴⁵	AF008220 of <i>B. subtilis</i>	Putative morphine dehydrogenase	Lapidus <i>et al.</i> (1997)
<i>yvgM</i>	1.2e ⁻⁴⁸	sl0739 of <i>Synechocystis</i> sp. strain PCC6803	ABC transporter	Kaneko <i>et al.</i> (1995)
<i>yvgL</i>	3.4e ⁻⁶⁰	sl0738 of <i>Synechocystis</i> sp. strain PCC6803	Molybdate-binding periplasmic protein	Kaneko <i>et al.</i> (1995)
<i>yvgK</i>	6.8e ⁻⁷⁸	M21537 of <i>Clostridium pasteurianum</i>	Putative Nif-related protein	Chen <i>et al.</i> (1986)
<i>yvgJ</i>	3.4e ⁻²³⁶	<i>yqgS</i> of <i>B. subtilis</i>	Hypothetical protein	P54496
<i>yvsG</i>	4.1e ⁻¹⁴	o200 of <i>Escherichia coli</i> K-12	Hypothetical protein	Blattner <i>et al.</i> (1997)
<i>yvsH</i>	1.7e ⁻¹¹⁰	X97768 of <i>Clostridium perfringens</i>	Arginine/ornithine antiporter	Ohtani <i>et al.</i> (1997)
<i>yvrP</i>	3.6e ⁻⁷	<i>sapE</i> of <i>Lactobacillus sake</i> strain Lb706 plasmid pLSA60	Possible ABC transporter accessory factor	Axelsson & Holck (1995)
<i>yvrO</i>	3.2e ⁻⁶³	slr0544 of <i>Synechocystis</i> sp. strain PCC6803	ABC transporter	Kaneko <i>et al.</i> (1995)
<i>yvrN</i>	3.3e ⁻¹²⁰	<i>yziB</i> of <i>B. subtilis</i>	Hypothetical protein	Smith <i>et al.</i> (1988)
	1.3e ⁻³⁴	D90917 of <i>Synechocystis</i> sp.	Hypothetical protein	Kaneko <i>et al.</i> (1995)
<i>yvrK</i>	1.5e ⁻⁹⁷	D90907 of <i>Synechocystis</i> sp. strain PCC6803	Hypothetical protein	Kaneko <i>et al.</i> (1995)
<i>yvrI</i>	6.6e ⁻⁷	<i>rpsG</i> of <i>B. subtilis</i>	RNA polymerase σ^G factor	Karmazyn-Campelli <i>et al.</i> (1989)
<i>yvrH</i>	2.0e ⁻⁵⁷	<i>drvA</i> of <i>Thermotoga maritima</i>	DNA-binding response regulator	Lee & Stock (1996)
<i>yvrG</i>	2.5e ⁻²⁸	<i>rseE</i> of <i>B. subtilis</i>	Sensor protein	Sorokin <i>et al.</i> (1993)
<i>yvrE</i>	7.8e ⁻⁶¹	<i>rpoX</i> of <i>Sulfolobus acidocaldarius</i>	Putative DNA-directed RNA polymerase subunit	U05664
<i>yvrD</i>	2.8e ⁻⁴⁷	<i>snoD</i> of <i>Streptomyces nogalater</i>	Polyketide ketoreductase	Ylihonko <i>et al.</i> (1996)
<i>yvrC</i>	3.4e ⁻¹²	E64310 of <i>Methanococcus jannaschii</i>	Iron transport system binding protein	Bult <i>et al.</i> (1996)
<i>yvrB</i>	2.2e ⁻⁴⁶	<i>hemU</i> of <i>Haemophilus influenzae</i>	Haemin permease	Fleischmann <i>et al.</i> (1995)
<i>yvrA</i>	3.0e ⁻⁵⁵	<i>fhuC</i> of <i>B. subtilis</i>	Ferrichrome transport ATP-binding protein	X93092
<i>yvqK</i>	1.0e ⁻³⁷	U471A of <i>Mycobacterium leprae</i>	Hypothetical protein	U15186
<i>yvqJ</i>	4.7e ⁻¹⁸	<i>mefE</i> of <i>Streptococcus pneumoniae</i>	Macrolide-efflux determinant	Tait-Kamradt <i>et al.</i> (1997)
<i>yvqH</i>	1.6e ⁻¹⁴	D64002 of <i>Synechocystis</i> sp. strain PCC6803	Chloroplast-membrane-associated 30 kDa protein	Kaneko <i>et al.</i> (1995)
<i>yvqE</i>	6.6e ⁻⁴¹	U81487 of <i>Lactococcus lactis</i> subsp. <i>cremoris</i>	Histidine kinase	O'Connell-Motherway <i>et al.</i> (1997)
<i>yvqC</i>	5.9e ⁻⁵⁰	<i>degU</i> of <i>B. subtilis</i>	Transcription regulatory protein	Tanaka & Kawata (1988)
<i>yuxN</i>	4.9e ⁻⁴¹	<i>yuxN</i> of <i>B. subtilis</i>	Hypothetical protein	Miles & Guest (1985)
<i>yvqB</i>	4e ⁻¹⁹	X86780 of <i>Streptomyces hygroscopicus</i>	Sensory protein kinase	Schwecke <i>et al.</i> (1995)
<i>yvqA</i>	1.6e ⁻⁴	<i>phoP</i> of <i>B. subtilis</i>	Response regulator, phosphate starvation	Lee & Hulett (1992)

yvrI, *yvrH*, *yvrG* and *yvrE* form a putative operon of approximately 4 kb. The product of *yvrI* shows a low but significant similarity with RNA polymerase sigma factors from a number of micro-organisms, with the forespore-specific *B. subtilis* SigG being the most similar (Karmazyn-Campelli *et al.*, 1989). The putative product of *yvrH* shows a good level of similarity to a DNA-binding response regulator, DrrA, of *Thermotoga*

maritima (Lee & Stock, 1996), whilst the product of *yvrG* shows similarity to a sensor kinase protein, ResE, of *B. subtilis* (Sorokin *et al.*, 1993). YvrH and YvrG are likely to form a two-component signal transduction pathway for the regulation of gene expression. The product of the final putative ORF, *yvrE*, shows a good level of similarity to RpoX (P46218) of *Sulfolobus acidocaldarius*, which is a subunit of a DNA-directed

RNA polymerase (U05664). Interestingly, YvrE also shows good similarity to the senescence marker protein-30 (SMP30) (Fujita *et al.*, 1992) from a number of mammals.

ACKNOWLEDGEMENTS

This work was funded by the European Commission under the Biotechnology Programme, contract numbers BIO2-CT93-0272 and BIO2-CT96-0655. We also wish to acknowledge the technical support of R. M. Liddell of the Molecular Biology Facility, University of Newcastle upon Tyne.

REFERENCES

- Anagnostopoulos, C. & Spizizen, J. (1961). Requirements for transformation in *Bacillus subtilis*. *J Bacteriol* **81**, 741–746.
- 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.
- Atkins, J. F., Weiss, R. B., Thompson, S. & Gesteland, R. F. (1991). Toward a genetic dissection of the basis of triplet decoding, and its natural subversion: programmed reading frame shifts and hops. *Annu Rev Genet* **25**, 201–228.
- Axelsson, L. & Holck, A. (1995). The genes involved in production of and immunity to sakacin A, a bacteriocin from *Lactobacillus sake* Lb706. *J Bacteriol* **177**, 2125–2137.
- Blattner, F. R., Plunkett, G., III, Bloch, C. A. & 14 other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462.
- Bult, C. J., White, O., Olsen, G. J. & 37 other authors (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073.
- Chen, K. C., Chen, J. S. & Johnson, J. L. (1986). Structural features of multiple *nifH*-like sequences and very biased codon usage in nitrogenase genes of *Clostridium pasteurianum*. *J Bacteriol* **166**, 162–172.
- Demerec, M., Adelberg, E. A., Clark, A. J. & Hartman, P. E. (1968). Proposal for a uniform nomenclature in bacterial genetics. *J Gen Microbiol* **50**, 1–14.
- Fleischmann, R. D., Adams, M. D., White, O. & 37 other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
- Fujita, T., Shirasawa, T., Uchida, K. & Maruyama, N. (1992). Isolation of cDNA clone encoding rat senescence marker protein-30 (SMP30) and its tissue distribution. *Biochim Biophys Acta* **1132**, 297–305.
- Glaser, P., Kunst, F., Arnaud, M. & 14 other authors (1993). *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325° to 333°. *Mol Microbiol* **10**, 371–384.
- Harwood, C. R. & Wipat, A. (1996). Sequencing and functional analysis of the genome of *Bacillus subtilis* strain 168. *FEBS Lett* **389**, 84–87.
- Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. & Tabata, S. (1995). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Res* **2**, 153–166.
- Karmazyn-Campelli, C., Bonamy, C., Savelli, B. & Stragier, P. (1989). Tandem genes encoding sigma-factors for consecutive steps of development in *Bacillus subtilis*. *Genes Dev* **3**, 150–157.
- Kunst, F., Ogasawara, N., Moszer, I. & 148 other authors. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256.
- Lapidus, A., Galleron, N., Sorokin, A. & Ehrlich, D. (1997). Sequencing and functional annotation of the *Bacillus subtilis* genes in the 200 kb *rrnB-dnaB* region. *Microbiology* **143**, 3431–3441.
- Lee, J. W. & Hulett, F. M. (1992). Nucleotide sequence of the *phoP* gene encoding PhoP, the response regulator of the phosphate regulon of *Bacillus subtilis*. *Nucleic Acids Res* **20**, 5848.
- Lee, P. J. & Stock, A. M. (1996). Characterization of the genes and proteins of a two-component system from the hyperthermophilic bacterium *Thermotoga maritima*. *J Bacteriol* **178**, 5579–5585.
- Maupin-Furlow, J. A., Rosentel, J. K., Lee, J. H., Deppenmeier, U., Gunsalus, R. P. & Shanmugam, K. T. (1995). Genetic analysis of the *modABCD* (molybdate transport) operon of *Escherichia coli*. *J Bacteriol* **177**, 4851–4856.
- Miles, J. S. & Guest, J. R. (1985). Complete nucleotide sequence of the fumarase gene (*citG*) of *Bacillus subtilis* 168. *Nucleic Acids Res* **13**, 131–140.
- O'Connell-Motherway, M., Fitzgerald, G. F. & van Sinderen, D. (1997). Cloning and sequence analysis of histidine protein kinases isolated from *Lactococcus lactis* MG1363. *Appl Environ Microbiol* **63**, 2454–2459.
- Ogasawara, N., Nakai, S. & Yoshikawa, N. (1994). Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. *DNA Res* **1**, 1–14.
- Ohtani, K., Bando, M., Swe, T., Banu, S., Oe, M., Hayashi, H. & Shimizu, T. (1997). Collagenase gene (*colA*) is located in the 3'-flanking region of the perfringolysin O (*pfoA*) locus in *Clostridium perfringens*. *FEMS Microbiol Lett* **146**, 155–159.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Schneider, R. & Hantke, K. (1993). Iron-hydroxamate uptake systems in *Bacillus subtilis*: identification of a lipoprotein as part of a binding protein-dependent transport system. *Mol Microbiol* **8**, 111–121.
- Schwecke, T., Aparicio, J. F., Molnar, I. & 10 other authors (1995). The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc Natl Acad Sci USA* **92**, 7839–7843.
- Simonen, M. & Palva, I. (1993). Protein secretion in *Bacillus* species. *Microbiol Rev* **57**, 109–137.
- Smith, H., de Jong, A., Bron, S. & Venema, G. (1988). Characterization of signal-sequence-coding regions selected from the *Bacillus subtilis* chromosome. *Gene* **70**, 351–361.
- 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.
- Tait-Kamradt, A., Clancy, J., Cronan, M., DibHajj, F., Wondrack, L., Yuan, W. & Sutcliffe, J. (1997). *mefE* is necessary for the erythromycin-resistant M phenotype in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* **41**, 2251–2255.
- Tanaka, T. & Kawata, M. (1988). Cloning and characterization of *Bacillus subtilis iep*, which has positive and negative effects on production of extracellular proteases. *J Bacteriol* **170**, 3593–3600.

Vieira, J. & Messing, J. (1987). Production of single-stranded plasmid DNA. *Methods Enzymol* **153**, 3–15.

Wipat, A., Carter, N., Brignell, S. C., Guy, B. J., Piper, K., Sanders, J., Emmerson, P. T. & Harwood, C. R. (1996). The *dnaB*–*pheA* (256°–240°) region of the *Bacillus subtilis* chromosome containing genes responsible for stress responses, the utilization of plant cell walls and primary metabolism. *Microbiology* **142**, 3067–3078.

Yamane, K., Kumano, M. & Kurita, K. (1996). The 25°–36° region of the *Bacillus subtilis* chromosome: determination of the sequence of a 146 kb segment and identification of 113 genes. *Microbiology* **142**, 3047–3056.

Ylihonko, K., Tuikkanen, J., Jussila, S., Cong, L. & Mantsala, P. (1996). A gene cluster involved in nogalamycin biosynthesis from *Streptomyces nogalater*: sequence analysis and complementation of early-block mutations in the anthracycline pathway. *Mol Gen Genet* **251**, 113–120.

Zuberi, A. R., Moir, A. & Feavers, I. M. (1987). The nucleotide sequence and gene organisation of the *gerA* spore germination operon of *Bacillus subtilis* 168. *Gene* **51**, 1–11.

Received 7 November 1997; revised 6 February 1998; accepted 16 February 1998.