Sequencing of regions downstream of *addA* (98°) and *citG* (289°) in *Bacillus subtilis*

N. Médina, F. Vannier, B. Roche, S. Autret, A. Levine and S. J. Srr

Author for correspondence: S. J. Srr. Tel: +33 1 69 15 57 14. Fax: +33 1 69 15 78 08.
e-mail: seror@igmors.u-psud.fr

Institut de Gntique et
Microbiologie, URA CNRS
2225, Universit Paris XI
Btiment 409, 91405 Orsay
Cedex, France

The nucleotide sequence of 17.3 kbp downstream of *addA* (98°) on the *Bacillus subtilis* chromosome was determined. Twenty putative ORFs were identified. Three of them coincided with known *B. subtilis* genes, *addA*, *sbcD* and *wprA*. The product of four other ORFs showed similarity to SbcC of *Clostridium perfringens*, Coth of *B. subtilis*, 2-hydroxyhepta-2,4-diene-1,7-diolate isomerase of *Methanococcus jannaschi* and a putative ORF of *Pseudomonas syringae*. In addition, a sequence of 7.6 kbp downstream of *citG* (189°) was analysed. Among 10 putative ORFs identified, two coincided with known genes, *citG* and *mrgA*, whilst three showed homology with X86780, a sensory protein kinase of *Streptomyces hygroscopicus*, an alkaline phosphatase regulatory protein and a hypothetical protease, *YyxA*, of *B. subtilis*.

Keywords: *Bacillus subtilis*, genome sequencing, *addA* region, *citG* region

As participants in the European *Bacillus subtilis* genome sequencing project, we were responsible for the sequencing of the chromosomal region between *addA* and *argJ*. In this communication we report the determination and the properties of a 17.3 kbp sequence downstream of *addA* (98°). This completes our sequencing of the 54 kbp *addA*–*argJ* chromosomal region which proved to be much shorter than originally estimated. The sequencing strategy mainly involved Long Range PCR between *addA* and *ysp* (Roche *et al.*, 1997), shotgun analysis after nebulization and subcloning in pUC18, as described previously (Levine *et al.*, 1997). The gaps were filled by normal PCR. The sequences were analysed with an Applied Biosystems 373A DNA Sequencer equipped with long plates. This paper also describes the 7.6 kbp sequence downstream of *citG* (289°). Inverse PCR was used to isolate DNA fragments adjacent to *citG* (289°) and the nucleotide analysis of the new extremity allowed us to synthesize a new DNA fragment by normal PCR, using the chromosomal DNA as a template. Sequencing was achieved directly from the PCR fragments, on both strands, using the *Taq* Dye Primer/Terminator Cycle Sequencing Kits (Perkin Elmer) and an automatic sequencer (373A, Applied Biosystems).

Abbreviation: SD, Shine–Dalgarno.

The EMBL accession numbers for the sequences reported in this paper are Y09476 (98°) and Z93941 (289°).

The 17.3 kbp region (98°) downstream of *addA*

From the sequence data, putative ORFs were identified, considering as valid the longest possible ORFs starting with ATG, GTG or TTG and preceded by a Shine–Dalgarno (SD) sequence. Twenty putative ORFs were identified and are listed in Table 1. Suitable RBSs were found in all cases except ORF5 and ORF19. The initiation codon was ATG for 15 of the ORFs, GTG for two and TTG for the two others. Note that the direction of transcription and translation of almost half of the ORFs (9 out of the 20, 8 of the 9 being in a cluster) in this chromosomal region is in the opposite direction of DNA replication (Fig. 1).

The similarity analysis of the predicted gene products to known proteins is shown in Table 1. The ORF5 product showed 50% amino acid sequence identity with SbcC protein from *Clostridium perfringens* (GenPept no. X86484). The amino acid sequence of the ORF14 product showed 33% similarity with Coth of *B. subtilis* (Naclerio *et al.*, 1996) and that of ORF15 showed 48% identity with 2-hydroxyhepta-2,4-diene-1,7-diolate isomerase of *Methanococcus jannaschi* (Bult *et al.*, 1996). The ORF19 product showed 35% amino acid sequence identity with a putative ORF of *Pseudomonas syringae* (GenPept no. X91388). Three of the ORFs corresponded to genes already sequenced, *addA*, *sbcD* (Kooistra & Venema, 1991) and *wprA* (Margot & Karamata, 1996). The reported sequences for these genes were found to be identical to our sequence, except

Table 1. Putative ORFs in the region downstream of *addA* (98°) and similarity to other proteins

ORF	Gene name	Endpoints (nt)	Size of product (aa)	SD consensus sequence (upper case) and initiation codon (bold)	Similar protein in database	BLAST probabilities
ORF1	<i>addA</i>	1 > 699	232 (partial)		ATP-dependent nuclease of <i>B. subtilis</i> (partial sequence)	Known gene
ORF2	<i>sbcD</i>	774 > 1514	245	AGAGAGGtgaagcatttg	SbcD of <i>B. subtilis</i>	Known gene
ORF3	<i>yirY</i>	1947 > 2816	288	AGAGGAGGatgcgaatg	None	—
ORF4	<i>yirZ</i>	3010 > 4626	537	AGGAACAGctttcacgcttg	None	—
ORF5	<i>yisA</i>	4824 > 5339	170	GAAAAAATGGcagatg	SbcC of <i>Clostridium perfringens</i>	2·8 e—44
ORF6	<i>yisB</i>	5353 > 5655	99	GGAGGAGcgatgatg	None	—
ORF7	<i>yisC</i>	5692 < 5919	76	AcAGAGActtcgataaagggtg	None	—
ORF8	<i>yisD</i>	5942 < 6343	134	GcGccAGAataatg	None	—
ORF9	<i>yisE</i>	6343 < 6519	59	GAAAGGAGAtgagcaagcatg	None	—
ORF10	<i>yisF</i>	6516 < 7133	204	GAAAGAGGTgaagaacatg	None	—
ORF11	<i>yisG</i>	7156 < 7389	78	AGGAGtgaggctcatg	None	—
ORF12	<i>yisH</i>	7404 < 7625	74	AGGAtGtGctcgatg	None	—
ORF13	<i>yisI</i>	8048 < 8218	57	GGAGAtAGcgtaatg	None	—
ORF14	<i>yisJ</i>	8364 < 9287	306	AGGAGGAaagaatg	CotH of <i>B. subtilis</i>	2·0 e—22
ORF15	<i>yisK</i>	9442 > 10347	300	GGAGGAataacatg	2-Hydroxyhepta-2,4-diene-1,7-diolate isomerase of <i>Methanococcus jannaschi</i>	9·5 e—63
ORF16	<i>yisL</i>	10463 > 10819	117	AGGGGGAAcatatg	None	—
ORF17	<i>wprA</i>	10987 > 13671	893	AGGAGGgataacatg	WprA, cell-wall-associated protease precursor of <i>B. subtilis</i>	Known gene
ORF18	<i>yisN</i>	13702 < 14289	194	GGAAAGAAaagagaggaatg	None	—
ORF19	<i>yisO</i>	14504 > 16279	590	GGAAAAAcAAAcGAtggacagaatg	Putative ORF of <i>Pseudomonas syringae</i>	2·5 e—59
ORF20	<i>yisP</i>	17121 > 17287	55 (partial)	GAAGAAGcagtg	Phytoene synthase of <i>Synechococcus</i> sp.	

Fig. 1. Physical map and gene organization of the 17·3 kbp region (98°), *addA*–*yisP*. The scale, in kbp, starts at the T residue of the *Av*II site in *addA* and stops at the G residue, nt 17287 in *yisP* (corresponding to nt 1 in Roche *et al.*, 1997). The numbers 1–20 indicate the ORFs with the arrows indicating the direction of transcription and translation. The positions of previously identified genes, e.g. *wprA*, are also indicated.

for one position in *sbcD* (an additional G at position 1332 in our sequence) and two positions in *wprA* (T at position 11012 and C at position 11026 in our sequence instead of C and A, respectively). From these results and our previous data (Levine *et al.*, 1997; Roche *et al.*, 1997)

we deduced that *addA* was located 53·5 kbp from *argJ*. Consequently, in agreement with these results, its position on the *B. subtilis* genetic map has been re-mapped at 98° (Biaudet *et al.*, 1996) instead of 86° indicated previously (Anagnostopoulos *et al.*, 1993).

Examination of the deduced amino acid sequences clearly predicts that YisL is a membrane protein with four transmembrane domains (TMDs), whilst WprA apparently carries a classical N-terminal signal sequence indicative of a secreted protein (Fig. 2). Indeed, other studies (Margot & Karamata, 1996) indicate that this is a cell-wall-associated protease. Two other proteins, YisA and YisC, are weakly predicted to form monotopic membrane proteins with a single TMD, but this would require detailed topological analysis to confirm.

The 7·6 kbp region (289°) downstream of *citG*

From the sequence data, all six possible reading frames were screened with the GENEMARK program (Borodowsky & McIninch, 1993). Based on starting codons ATG, GTG or TTG, 10 putative ORFs were detected. Except for ORF5, SD sequences were detectable in these

Fig. 2. Hydropathy analysis of potential membrane and secreted proteins.

genes (Table 2). Note that, also in this chromosomal region, the direction of transcription and translation of half of the ORFs is in the opposite direction of DNA replication (Fig. 3). The endpoints of the ORFs in the sequence, the size of the proteins encoded, putative SD sequences and initiation codons are shown in Table 2. We could not find any SD sequence before ORF5 but the BLAST analysis detected a

high degree of similarity with a transcriptional regulator (Table 2). Four of the ORFs found have already been sequenced. *citG* (ORF1) is the only one with a known function (Miles & Guest, 1985) and *mrgA* (ORF8) has been shown to be regulated by metals (Chen *et al.*, 1993). Three other ORFs (ORF4, ORF5 and ORF7) showed

Fig. 3. Physical map and gene organization of the 7.6 kbp region (289°), *citG*–*yirG*. Hatched bars indicate DNA regions prepared by inverse PCR using region-specific primers. Numbers represent ORFs with arrows indicating the direction of transcription.

possible homology with other proteins (Table 2). Interestingly, a BLASTN search allowed us to find a 130 bp region which seems to be a repetition of the 5' region of the 16S rRNA gene (*trpE26* mutant, Jarvis *et al.*, 1990).

yirC appears to encode a membrane protein as predicted from the deduced amino acid sequence (Fig. 2). *YirC* is predicted to form a single transmembrane domain (residues 38–58), giving the relatively rare (in bacteria) type III topology with the N terminus on the outside. This, however, would require confirmation by a specific topological analysis.

ACKNOWLEDGEMENTS

N. Médina and F. Vannier contributed equally to the results described in this paper. We are grateful to A. Wipat (University of Newcastle) for preparing the *PstI* inverse PCR fragment. We are pleased to acknowledge support from

Table 2. Putative ORFs in the *citG*–*yirG* region (289°) and similarity to other proteins

ORF	Gene name	Endpoints (nt)	Size of product (aa)	SD consensus sequence (upper case) and initiation codon (bold)	Similar protein in database	BLAST probabilities
ORF1	<i>citG</i>	2>721	239		Known gene, partial sequence	–
ORF2	<i>yuxN</i>	756<1472	238	ttcGAAAGAAacgcttctctctctatg	Putative gene previously sequenced	–
ORF3	<i>yirB</i>	1799>1963	54	gaggaatcaaagtGGAGaaagtg	None	–
ORF4	<i>yirC</i>	1991<2959	322	aatGGAGAGaaagtg	GenBank no. X86780, sensory protein kinase of <i>Streptomyces hygroscopicus</i>	8.7e–21
ORF5	<i>yirD</i>	3343<3993	216	(no SD found) ccatttatctagtgaagtg	Alkaline phosphatase regulatory protein, <i>phoP</i> product of <i>B. subtilis</i>	2.4e–42
ORF6	<i>yirE</i>	4299>4664	121	AGGAGtgaagaacatg	None	–
ORF7	<i>yirF</i>	5001>5675	224	aagtGcGAGccttggtg	Hypothetical protease YyxA of <i>B. subtilis</i>	2.8e–57
ORF8	<i>mrgA</i>	5719<6180	154	aGGAGGAcgttatcttctatg	Known gene	–
ORF9	<i>yuxA</i>	6270<7112	281	aatAGGAGGggttagatg	Putative gene previously sequenced	–
ORF10	<i>yirG</i>	7245>7646	134	gaaGGAAAGGGttttgtgatg	None	–

CNRS, University Paris XI, the EU Biotechnology programme (BIO2-CT94-2011) and GREG (contract 118/96).

REFERENCES

- Anagnostopoulos, C., Piggot, P. J. & Hoch, J. A. (1993). The genetic map of *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, pp. 425–461. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Biaudet, V., Samson, F., Anagnostopoulos, C., Ehrlich, S. D. & Bessires, P. (1996). Computerized genetic map of *Bacillus subtilis*. *Microbiology* **142**, 2669–2729.
- Borodowsky, M. & McIninch, J. D. (1993). Genemark: parallel recognition for both strands. *Comput Chem* **17**, 123–133.
- Bult, C. J., White, O., Olsen, G. S. and 37 other authors (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschi*. *Science* **273**, 1058–1073.
- Chen, L., James, L. P. & Helmann, J. D. (1993). Metalloregulation in *Bacillus subtilis*: isolation and characterization of two genes differentially repressed by metal ions. *J Bacteriol* **175**, 5428–5437.
- Jarvis, E. D., Cheng, S. & Rudner, R. (1990). Genetic structure and DNA sequences at junctions involved in the rearrangements of *Bacillus subtilis* strains carrying the *trpE26* mutation. *Genetics* **126**, 785–797.
- Kooistra, J. & Venema, G. (1991). Cloning, sequencing, and expression of *Bacillus subtilis* genes involved in ATP-dependent nuclease synthesis. *J Bacteriol* **173**, 3644–3665.
- Levine, A., Vannier, F., Roche, B., Autret, S., Mavel, D. & Sror, S. J. (1997). A 10.3 kbp segment from *nprB* to *argJ* at the 102 region of the *Bacillus subtilis* chromosome. *Microbiology* **143**, 175–177.
- Margot, P. & Karamata, D. (1996). The *wprA* gene of *Bacillus subtilis* 168, expressed during exponential growth, encodes a cell-wall-associated protease. *Microbiology* **142**, 3437–3444.
- 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.
- Naclerio, G., Baccigalupi, L., Zilhao, R., De Felice, M. & Ricca, E. (1996). *Bacillus subtilis* spore coat assembly requires *cotH* gene expression. *J Bacteriol* **178**, 4375–4380.
- Roche, B., Autret, S., Levine, A., Vannier, F., Medina, N. & Sror, S. J. (1997). A *Bacillus subtilis* chromosome segment at the 100 to 102 position encoding 11 membrane proteins. *Microbiology* **143**, 3309–3312.

Received 1 May 1997; revised 3 July 1997; accepted 14 July 1997.