Mapping of the 150 kb *spoIIIC*–*pheA* region of the *Bacillus subtilis* chromosome using Long Accurate PCR and three yeast artificial chromosomes

Alexandre Bolotin, Alexei Sorokin and S. Dusko Ehrlich

Author for correspondence: Alexei Sorokin. Tel: +33 1 34 65 25 33. Fax: +33 1 34 65 25 21.
e-mail: sorokine@biotec.jouy.inra.fr

Laboratoire de Génétique Microbienne, Institut National de la Recherche Agronomique, Domaine de Vilvert, 78352 Jouy en Josas cedex, France

We constructed a PCR map of the 150 kb *spoIIIC*–*pheA* region of the *Bacillus subtilis* chromosome. It was established using known sequences of the *spoIIIC*, *blt*, *aadK*, *sacC*, *spoVB* and *pheA* loci and eight random sequence tags. The tags were generated using PFGE-purified DNA of yeast artificial chromosome (YAC) 11-17 from the yeast clone which carries the major part of this region. The ends of two other YACs were positioned on the map using total DNA extracted from yeast cells carrying them. The procedure allowed the placement of precisely known and new (putative) genes on the physical chromosome map and the generation of sufficient amounts of DNA for sequencing this region. Apart from allowing correction of the genetic map in this region, these results demonstrate how a collection of long segments of bacterial chromosome and Long Accurate PCR can be used for reliable high-resolution physical mapping of an extended chromosome area.

Keywords: *Bacillus subtilis*, genome sequencing, yeast artificial chromosome, Long Accurate PCR

The sequencing of the entire genome of *Bacillus subtilis* is greatly impaired by the difficulty in cloning its DNA in standard *Escherichia coli* vectors (Glaser *et al.*, 1993; Ogasawara *et al.*, 1994). This fact specifically hindered the *B. subtilis* genome project and obliged the participants to use PCR for closing uncloneable gaps (Glaser *et al.*, 1993; Ogasawara *et al.*, 1994; Sorokin *et al.*, 1993). To provide a general solution to this problem, a strategy based on the use of yeast artificial chromosomes (YACs) was developed in this laboratory (Sorokin *et al.*, 1996a). In this approach the YAC clones carrying long sequences of the *B. subtilis* genome (Azevedo *et al.*, 1993) are the source of DNA for cloning or the indicators of the region of interest by hybridization. The approach was greatly facilitated by applying Long Accurate PCR (LA PCR) protocols (Capuano *et al.*, 1996). The use of primer multiplexing extended the efficiency of the strategy to 200 kb regions (Sorokin *et al.*, 1996b). It was also shown that the error rate generated by LA PCR does not exceed 1 in 7000, which is an acceptable level for genome sequencing projects (Capuano *et al.*, 1996; Sorokin *et al.*, 1996b). We

report here data on LA PCR mapping of the 150 kb region between the *spoIIIC* and *pheA* loci of *B. subtilis* (for a recent version of the *B. subtilis* genetic map, see Anagnostopoulos *et al.*, 1993).

During the first step we used known sequences from the *SubtiList* database contigs (Moszer *et al.*, 1995; Perriere *et al.*, 1996) containing markers of *spoIIIC* (*SubtiList* contig SL227_1; 49552 bp), *blt* (SL230_1; 3723 bp), *aadK* (SL235_1; 1053 bp), *sacC* (SL233_1; 5164 bp), *spoVB* (SL238_1; 2459 bp) and *pheA* (SL240_1; 22685 bp). An extension of the SL233_1 contig up to 14021 bp was kindly provided by R. Mellado (personal communication). Primers corresponding to the ends of these contigs were synthesized (Table 1). The protocol used (*rTth* supplier's protocol, Perkin Elmer) worked up to 25 kb and allowed amplification of the *spoIIIC*–*blt*, *blt*–*aadK*, *aadK*–*sacC* and *spoVB*–*pheA* intergenic fragments, with *B. subtilis* chromosomal DNA as template and the corresponding primers (Fig. 1). Three YAC clones, 11-17, 15-141 and 11-229 from the collection of Azevedo *et al.* (1993) were positioned in this area. We hypothesized that the ends of the YAC insert are within the unamplified region. To locate these ends precisely, total chromosomal DNA of the corresponding yeast clones was prepared as

Abbreviations: LA PCR, Long Accurate PCR; YAC, yeast artificial chromosome.

Table 1. Primer structures

Primer	Contig	Sequence
AA3	<i>sacC</i>	ATC ACT GGA GCC TGT CCA GAG
AA7	<i>spoVB</i>	TAC CGT CAA GAA GCT GGC GGA
AA8	<i>pheA</i>	AAG CCT CTG GCG GGA TTA CG
AB3	pYAC4	GGT GAT GTC GGC GAT ATA GGC
AC1	New	GAA GAT GTG ATC CGC GTC TGC
AC2	pYAC4	ATA CTC TCG GTA GCC AAG TTG G
AC5	<i>spoIIIC</i>	CAG TTT GCA GTC CCT CGA TG
AD8	<i>blt</i>	AGA GCT GGG ATA GGG GGG AC
AD9	<i>blt</i>	GCC CTC ATT CAG CAC CTT GC
AF4	New	ACG GAG GAG AAC GTT CTC CC
AF8	New	GGG AGA ACG TTC TCC TCC GT
AG4	<i>sacC</i>	GGG GCA TGA TTG TAA CAT ACA CCA
BA3	<i>aadK</i>	GGT CAC TTT GGA AGG GTC ACG
BA4	<i>aadK</i>	GAG ATG CGA TAT GCC AGG CC
BA6	<i>spoVB</i>	CTG ATC GCG ACC GGC AGA CC

Fig. 1. Minimal PCR map of the *spoIIIC*–*pheA* region of the *B. subtilis* chromosome showing the LA PCR products covering the region. Hatched bars correspond to the sequenced areas. LA PCR fragments are shown in boxes with the names of corresponding primers (small arrows). The scale is in kb. Zero point corresponds to the position at 2690 kb on the physical chromosome map (Itaya, 1993). YAC clones are shown by dashed lines. *NotI* sites with their position on the map of Itaya are indicated.

described previously for *B. subtilis* (Sorokin *et al.*, 1996b), replacing lysozyme by the yeast lytic enzyme of *Arthro-bacter luteus* (ICN Biochemicals). Primers, corresponding

Fig. 2. Comparison of the pattern of Southern hybridization of the *B. subtilis* chromosome (lanes 1, 3, 5) and the AA7 × AA8 LA PCR product (lanes 2, 4, 6). Cleavage was with *PstI* (lanes 1, 2), *EcoRI* (lanes 3, 4) and *HindIII* (lanes 5, 6).

to the vector parts of these YACs (AB3 and AC2, Table 1), were tested in LA PCR reactions against primers AC5, AD8, AD9, BA3, BA4, AA3, AG4, BA6, AA7 and AA8, which are homologous to the *B. subtilis* chromosome, using DNA from the yeast clones as template. This allowed us to position the ends of the YAC clones on the map and to extend the map by amplifying the AB3 × BA6 fragment (Fig. 1). The remaining area appeared to be covered by YAC 11-17. PFGE-purified DNA of this YAC was cut by *AluI* and cloned in M13mp18, as described previously (Sorokin *et al.*, 1996a). Four hundred and eighty M13 clones were hybridized with the purified

Table 2. New genes in the *spoIIIC-pheA* region

Map position (kb)	Tag size (bp)	SwissProt entry	Description	Translation frame/probability
77	1127	sp P39312 CYCA_ECOLI	D-Serine-alanine/glycine transporter	+2/4.2e ⁻¹⁰
		sp P24207 PHEP_ECOLI	Phenylalanine-specific permease	+2/2.9e ⁻⁹
		sp P38102 YCAR_PSEAE	Hypothetical 23.3 kDa protein	-3/1.4e ⁻⁸
97	872	sp P24247 PFS_ECOLI	PFS protein	-2/2.0e ⁻¹³
		sp P37887 CYSK_BACSU	Cysteine synthase	-1/6.8e ⁻¹²
		sp P32582 CBS_YEAST	Cystathione β -synthase	-1/1.0e ⁻¹⁰
115	811	sp P42399 YCKA_BACSU	Probable amino acid ABC transporter	+3/4.4e ⁻³¹
		sp P41074 GLTJ_ECOLI	Glutamate/aspartate transporter	+3/2.9e ⁻¹⁹
		sp P45090 ARTQ_HAEIN	Arginine transport system permease	+3/1.5e ⁻¹⁷
120	822	sp P44551 YCFB_HAEIN	Hypothetical protein HI0174	+2/1.2e ⁻⁴⁴

YAC DNA and several fragments from the region, amplified by LA PCR. One hundred and twenty M13 clones gave a positive signal with YAC DNA and were not homologous to the LA PCR fragments. Sequencing of 29 randomly chosen clones gave 14 new unique tags corresponding to the as yet unmapped region. The remaining clones yielded *B. subtilis* sequences from this region that had been previously determined or sequences from the pYAC4 vector sequence. Finally, 14 primers matching the sequences of eight of the tags allowed us to complete the mapping and to derive the 'minimal PCR map' shown in Fig. 1.

The whole 130 kb region of unknown sequence was amplified by seven LA PCR reactions (data not shown). The amplified fragments were tested for identity with the physical map of the *B. subtilis* chromosome. For this purpose, Southern hybridization of the LA PCR products and *B. subtilis* chromosomal DNA, cut with *Pst*I, *Eco*RI or *Hind*III restriction enzymes was performed in parallel. Analysis of the AA7 $\times$ AA8 fragment is shown on Fig. 2 as an example. No difference between the pattern of the PCR fragments and the chromosomal DNA was detected, apart from the restriction products containing the amplification primer sequences. This shows that the LA PCR fragments have the same structure as the chromosome. The size of the *Not*I fragment detected in this area corresponded perfectly to that mapped by Itaya (fragment 17N, 100 kb; Itaya, 1993). The sequences of mapped tags were screened for homology using the BLASTX algorithm (Altschul *et al.*, 1990). We found that the encoded proteins have homologies to amino acid transporters, cysteine synthase and to three hypothetical proteins, the latter apparently conserved in bacteria (Table 2). Six new putative genes can therefore be placed on the map of this area. It should also be mentioned that the approach described here allowed us to position precisely the genes located previously in this part of the chromosome. This is particularly relevant to the *aadK* (Ohmiya *et al.*, 1989) and *blt* loci, the latter shown to be the same as *acfA* (Ahmed *et al.*, 1995). The method applied here is therefore very powerful for the mapping of sequenced genes and can be

used, for example, for pathogenic or nutritionally fastidious micro-organisms, where genetic approaches are very restricted.

A few successful attempts were recently presented on the direct use of YAC or P1 phage vector collections for initiating sequencing projects (Vaudin *et al.*, 1995; Nurminsky & Hartl, 1996). The experiments reported from our laboratory revealed the difficulties encountered in the final stages of YAC-based strategies (Sorokin *et al.*, 1996a; Capuano *et al.*, 1996). LA PCR (Barnes, 1994; Cheng *et al.*, 1994) applied in the manner presented here and elsewhere (Sorokin *et al.*, 1996b) allowed us to scale down YAC sequencing to 15–25 kb projects, and provides an efficient solution for at least 200 kb scale projects.

ACKNOWLEDGEMENTS

We thank Dr R. Mellado for providing the unpublished sequence, Nathalie Galleron for help with the M13 bank construction and Dr Alla Lapidus for advice in using LA PCR. This work was supported in part by EEC grant BIO2-CT93-0272.

REFERENCES

- Ahmed, M., Lyass, L., Markham, P. N., Taylor, S. S., Vázquez-Laslop, N. & Neyfakh, A. R. (1995). Two highly similar multidrug transporters of *Bacillus subtilis* whose expression is differentially regulated. *J. Bacteriol.* **177**, 3904–3910.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410.
- Anagnostopoulos, C., Piggot, P. J. & Hoch, J. A. (1993). The genetic map of *Bacillus subtilis*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 425–461. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Azevedo, V., Alvarez, E., Zumstein, E., Damiani, G., Sgarbetta, V., Ehrlich, S. D. & Serrão, P. (1993). An ordered collection of *Bacillus subtilis* DNA segments cloned in yeast artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **90**, 6047–6051.

- Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and yield from λ bacteriophage templates. *Proc Natl Acad Sci USA* **91**, 2216–2220.
- Capuano, V., Galleron, N., Pujic, P., Sorokin, A. & Ehrlich, S. D. (1996). Organization of the *Bacillus subtilis* 168 chromosome between *kedg* and the attachment site of the *SP β* prophage: use of Long Accurate PCR and yeast artificial chromosomes for sequencing. *Microbiology* **142**, 3005–3015.
- Cheng, S., Fockler, C., Barnes, W. M. & Higuchi, R. (1994). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc Natl Acad Sci USA* **91**, 5695–5699.
- 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.
- Itaya, M. (1993). Physical map of the *Bacillus subtilis* 168 chromosome. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 463–472. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Moszer, I., Glaser, P. & Danchin, A. (1995). *SubtiList*: a relational database for the *Bacillus subtilis* genome. *Microbiology* **141**, 261–268.
- Nurminsky, D. I. & Hartl, D. L. (1996). Sequence scanning: a method for rapid sequence acquisition from large-fragment DNA clones. *Proc Natl Acad Sci USA* **93**, 1694–1698.
- 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.
- Ohmiya, K., Tanaka, T., Noguchi, N., O'Hara, K. & Kono, M. (1989). Nucleotide sequence of the chromosomal gene coding for the aminoglycoside 6-adenylyl-transferase from *Bacillus subtilis* Marburg 168. *Gene* **78**, 377–378.
- Perriere, G., Moszer, I. & Gojobori, T. (1996). *NRSUB*: a non-redundant database for *Bacillus subtilis*. *Nucleic Acids Res* **24**, 41–45.
- 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.
- Sorokin, A., Azevedo, V., Zumstein, E., Galleron, N., Ehrlich, S. D. & Serror, P. (1996a). Sequence analysis of the *Bacillus subtilis* chromosome region between the *serA* and *kedg* loci cloned in a yeast artificial chromosome. *Microbiology* **142**, 2005–2016.
- Sorokin, A., Lapidus, L., Capuano, V., Galleron, N., Pujic, P. & Ehrlich, S. D. (1996b). A new approach using Multiplex Long Accurate PCR and yeast artificial chromosomes for bacterial chromosome mapping and sequencing. *Genome Res* **6**, 448–453.
- Vaudin, M., Roopra, A., Hillier, L., Brinkman, R., Sulston, J., Wilson, R. K. & Waterston, R. H. (1995). The construction and analysis of M13 libraries prepared from YAC DNA. *Nucleic Acids Res* **23**, 670–674.

Received 18 April 1996; revised 29 June 1996; accepted 1 July 1996.