

Systematic sequencing of the *Bacillus subtilis* genome: progress report of the Japanese group

Naotake Ogasawara,¹ Yasutaro Fujita,² Yasuo Kobayashi,³
Yoshito Sadaie,⁴ Teruo Tanaka,⁵ Hideo Takahashi,⁶ Kunio Yamane⁷
and Hiroshi Yoshikawa¹

Author for correspondence: Naotake Ogasawara. Tel: +81 7437 2 5430. Fax: +81 7437 2 5439.
e-mail: nogasawa@bs.aist-nara.ac.jp

¹Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara 630-01, Japan

²Faculty of Engineering, Fukuyama University, Fukuyama, Hiroshima 729-02, Japan

³Faculty of Agriculture, Tokyo University of Agriculture and Technology, Fuchu, Tokyo 183, Japan

⁴National Institute of Genetics, Mishima, Shizuoka 411, Japan

⁵School of Marine Science and Technology, Tokai University, Shimizu, Shizuoka 424, Japan

⁶Institute of Molecular and Cellular Biosciences, Tokyo University, Bunkyo-Ku, Tokyo 113, Japan

⁷Institute of Biological Sciences, Tsukuba University, Tsukuba, Ibaraki 305, Japan

Keywords: *Bacillus subtilis*, genome project, genome sequencing

Introduction

We have undertaken the systematic sequencing of the *Bacillus subtilis* genome as part of an international co-operative project between the European Community and Japan. The Japanese Human Genome Program was officially initiated in 1991 supported by a research grant from Monbusho (The Ministry of Education, Science and Culture of Japan). The program comprises three projects: (1) genetic and physical mapping of the human genome; (2) systematic analysis of human cDNA; and (3) development of analytical methods. The *B. subtilis* project was included in the third project as a model system for a large sequencing project. Six groups are now actively engaged in sequencing, and Tanaka's group, who has constructed a physical map of the *B. subtilis* genome (Itaya & Tanaka, 1991), is engaged in the development of new tools for genome manipulation. Some papers that resulted from the project have already been published (Miyao *et al.*, 1992; Fuma *et al.*, 1993; Ogasawara *et al.*, 1994; Nakane *et al.*, 1994; Yoshida *et al.*, 1994); others appear in this issue (Ogawa *et al.*, 1995; Fujishima & Yamane, 1995; Yoshida *et al.*, 1995; Takemaru *et al.*, 1995).

Cloning

Five contiguous sections totalling 1 Mb and one separate section of 0.2 Mb were assigned to the six Japanese groups. In addition, a new section of 0.1 Mb was recently assigned to another Japanese group (Jun-ichi Sekiguchi,

Faculty of Textile Science & Technology, Shinshu University, Ueda, Nagano 386, Japan). Thus we are now responsible for the sequencing of a 1.3 Mb region (Fig. 1). We started by isolating λ linking clones covering each assigned region, using genes already sequenced and mapped, and *NotI* and *SfiI* linking clones isolated by Itaya & Tanaka (1991) as probes. It should be noted that the *NotI* and *SfiI* linking clones proved very effective markers for cloning the assigned regions, since their location on the genome is known very precisely. We have now cloned regions covering 700 kb in total, as shown in Fig. 1. However, during the isolation of λ clones, many regions were found to be difficult to clone into the λ phage vectors. We have overcome this difficulty mainly by using the inverse PCR method.

Sequence determination

Random sequencing of λ phage inserts (Ogasawara *et al.*, 1994) has been adopted by most of the Japanese groups. A library of random overlapping sub-clones in M13 or pUC vectors was constructed after DNaseI digestion of λ phage inserts. Inserts of sub-clones were amplified directly from plaques or colonies by PCR, and their sequences determined by an automatic dye sequencer. Sequence information from random sub-clones corresponding to about four times the length of the initial fragment was usually sufficient to cover most of the target fragment. The remaining gap regions were filled using region-specific primers. To avoid errors in the number of bases,

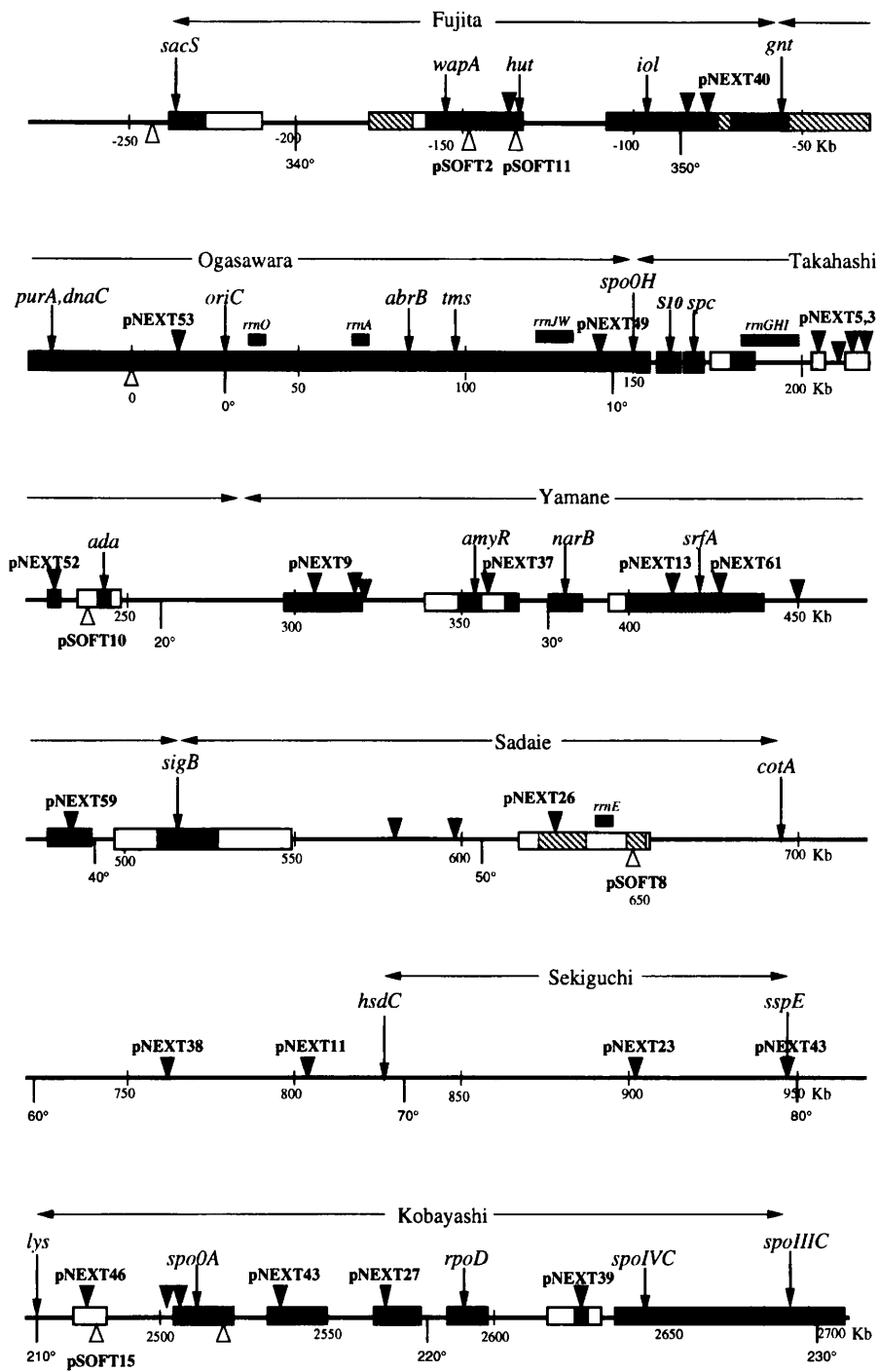


Fig. 1. Present status of cloning and sequencing of the *B. subtilis* genome by the Japanese groups. The cloned regions are shown by boxes (total 700 kb) on the physical map (Itaya & Tanaka, 1991) of the sections of the *B. subtilis* genome assigned to the Japanese groups. Sequences of regions shown in black have been determined during this study (total 500 kb) and those of shaded regions are near to completion (total 65 kb). The sequences of the hatched regions were reported by other groups (total 30 kb). Open and filled triangles indicate *Sfi*I and *Not*I sites, respectively. Positions of *Sfi*I and *Not*I linking clones (pSOFT and pNEXT; Itaya & Tanaka, 1991) are indicated, together with some genetic markers.

we routinely determined both strands even when readings from one strand seemed to be clear and without errors. In addition, we inspected every chart of raw data from the sequencer and used sequence outputs only from regions where base peaks were well separated. For further quality control, we kept the raw data to re-check the process of sequence determination.

Present status of the Japanese project

Fig. 1 shows the present status of the sequence determination of the regions assigned to the Japanese groups.

Approximately 500 kb has been sequenced during this project, and the remaining region is estimated to be about 800 kb. About 370 kb of the determined sequence has already been deposited in the GenBank/EMBL/ DDBJ data bank with accession numbers D10602 (*murG*, 1324 bp), D13262 (*srfA*, 20535 bp), D26185 (*oriC* region, 180136 bp), D16312 (*aroI*, 911 bp), D14399 (*iol* region, 14974 bp), D30808 (24° region, 21548 bp), D30762 (*srfAA* region, 10000 bp), D31856 (*hut-wapA*, 28954 bp), D32216 (*skin* element, 48031 bp), D38161 (39° region, 15065 bp), D31629 (*gnt* region, 18204 bp) and D30689 (*narB* region, 11604 bp).

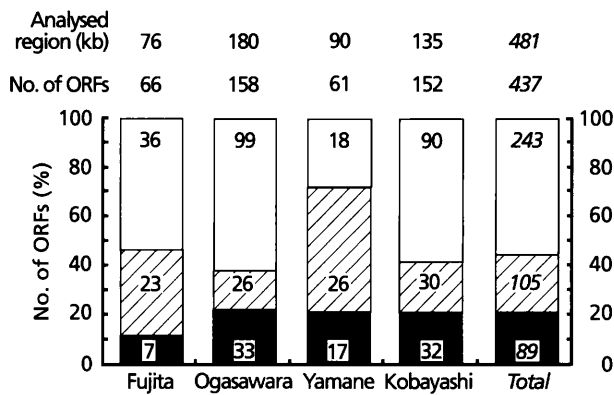


Fig. 2. Identification and analysis of the functions of ORFs in the sequenced regions. □, unknown; ▨, suggested; ■, known.

Each step of the shotgun sequencing method (preparation of a random library, sequencing of randomly selected inserts and filling of gaps by region-specific primers) requires about 2 or 3 weeks for sequencing a λ insert of 10–15 kb. Consequently, we are able to complete the sequencing of such an insert in about 2 months. It is now not difficult for a small team of two to three persons to sequence 50 kb per year. We therefore expect that the sequencing of the regions assigned to Japan will be near completion in 2 years, as each of the seven groups is expected to determine a 50 kb sequence per year.

Four groups, Ogasawara, Kobayashi, Yamane and Fujita, have determined and analysed a fairly large amount of sequence: 180, 135, 90 and 76 kb, respectively. Analysis of these sequences revealed 437 putative ORFs in total, including 89 ORFs already characterized genetically or biochemically (Fig. 2). The functions of 105 additional ORFs were suggested by the significant similarities of their products to known protein sequences in the current databases. The interesting new identifications include: the entire primary structures of the *but*, *iol*, *nar-nir* and *pts* operons; genes related to macromolecular synthesis, i.e., three ribosomal protein genes, two aminoacyl-tRNA synthetase genes and genes homologous to *Escherichia coli* *dnaB*, *ssb*, *xth* and *mfd*; cell division genes homologous to *E. coli* *ftsE*, *ftsH* and *pbp2*; and three new sets of genes belonging to the family of two-component regulatory

systems. However, no information on the functions of the remaining 243 ORFs was obtained through a first step database search with stringent criteria. A more extensive search for functional similarities will be necessary. We are also planning a systematic analysis of ORF functions by experimental means.

References

- Fujishima, Y. & Yamane, K. (1995). A 10 kb sequence at the 5' flanking region (32°) of *srf.AA* of the *Bacillus subtilis* chromosome. *Microbiology* **141**, 277–279.
- Fuma, S., Fujishima, Y., Corbell, N., D'Souza, C., Nakano, M., Zuber, P. & Yamane, K. (1993). Nucleotide sequence of 5' portion of *srfA* that contains the region required for competence establishment in *Bacillus subtilis*. *Nucleic Acids Res* **21**, 93–97.
- 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.
- Miyao, A., Yoshimura, A., Sato, T., Theeragool, G. & Kobayashi, Y. (1992). Sequence of the *Bacillus subtilis* homolog of the *Escherichia coli* cell division gene *murG*. *Gene* **118**, 147–148.
- Nakane, A., Ogawa, K., Nakamura, K. & Yamane, K. (1994). Nucleotide sequence of the shikimate kinase gene (*aroI*) of *Bacillus subtilis*. *J Ferment Technol* **77**, 312–314.
- Ogasawara, N., Nakai, S. & Yoshikawa, H. (1994). Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. *DNA Res* **1**, 1–14.
- Ogawa, K., Akagawa, E., Nakamura, K. & Yamane, K. (1995). Determination of a 21 548 bp nucleotide sequence around the 24° region of the *Bacillus subtilis* chromosome. *Microbiology* **141**, 269–275.
- Takemaru, K., Mizuno, M., Sato, T., Takeuchi, M. & Kobayashi, Y. (1995). Complete nucleotide sequence of a *skin* element, which is excised by DNA rearrangement during sporulation in *Bacillus subtilis*. *Microbiology* **141**, 323–327.
- 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. (1995). Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the *but* and *wapA* loci. *Microbiology* **141**, 337–343.

Received 13 July 1994; revised 13 September 1994; accepted 4 October 1994.