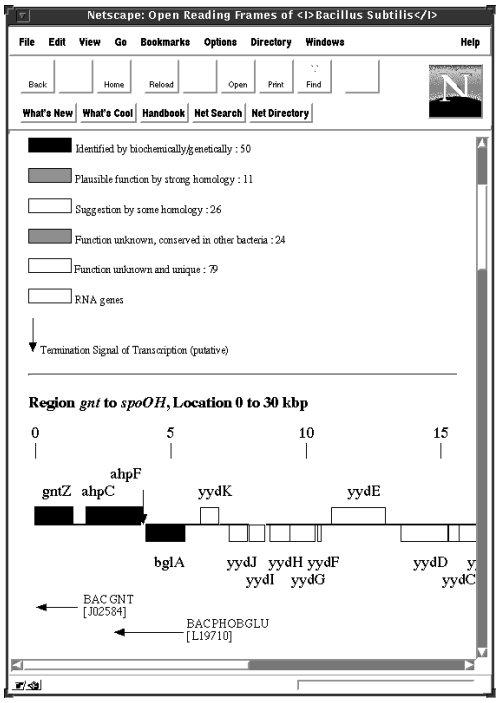
On annotating the functions of ORFs, every translated ORF sequence was applied to FASTA homology search. The top 20 results of FASTA search are visible using WWW. The results were summarized in order to check them easily at a glance, but of course raw results can be obtained. In addition, a user can get directly the contents of the entries that were found to be similar. The search subjects library was constructed of the released SWISS-PROT database and of the weekly updated SWISS-PROT database. The homology search will be applied every month and the BSORF homology entries will be updated monthly. The regular update of homology search will help the project members to annotate currently yet unknown ORF entries.

The sequences themselves are not directly available from the BSORF DB entries because these sequences are finally registered to the SWISS-PROT database which is easy to obtain. Actually, the ID names of the BSORF DB are basically common to those of SWISS-PROT.

There is another database for *Bacillus* genome project constructed in France. SubtiList[4] is a database system based on the relational data model and contains nucleotide sequences and other related information. Comparing with the SubtiList, the most characteristic point of our BSORF DB lies on its functional information. BSORF DB is not only a sequence database but rather a function database of *Bacillus Subtilis*. The database will be updated until all ORFs are functionally elucidated.

In the current scheme, each functional information is attributed to each ORF, but to tell the ideal, the function should also be described as a relationship or a network of concerning factors. One idea to illustrate the network of interactions is to represent them in a drawing. We are also planning to make links to the Kyoto Encyclopedia of Genes and Genomes (KEGG) [5] to show the illustration of functional networks.

In the workshop, we will present the current version of BSORF DB on WWW. The URL is `http://bacillus.genome.ad.jp/BSORF-DB.html` but it will also be accessible via the home page of the GenomeNet WWW server (`http://www.genome.ad.jp`).



Acknowledgement

We thank Pf. M. Kanehisa for providing us the on-line list of E.coli gene classification implemented in the KEGG. Dr. Y. Akiyama kindly allowed us to use his HTML code of a folding list to make the gene category list. Data entries of click-able map are made by the staffs of the Human Genome Center. Access to the public sequence databases is provided with the WebDBget system served on the GenomeNet.

References

[1] N. Ogasawara, S. Nakai & H. Yoshikawa, "Systematic Sequencing of the 180 Kilobase Region of the *Bacillus Subtilis* Chromosome Containing the Replication Origin", *DNA Research* **1**: 1 - 14 (1994).

[2] N. Ogasawara, Y. Fujita, Y. Kobayashi, Y. Sadaie, T. Tanaka, H. Takahashi, K. Yamane & H. Yoshikawa, "Systematic sequencing of the *Bacillus subtilis* genome: progress report of the Japanese group", *Microbiology* **141**: 257 - 259 (1995).

[3] M. Riley, "Functions of the gene products of *Escherichia coli*", *Microbiol. Rev.* **57**: 862 - 952 (1993).

[4] I. Moszer, P. Glaser & A. Danchin, "SubtiList: a relational database for the *Bacillus subtilis* genome", *Microbiology* **141**: 261 - 268 (1995).

[5] M. Kanehisa, unpublished. (KEGG will be accessible throuth the GenomeNet WWW server, `http://www.genome.ad.jp`)

A WWW Database of *Bacillus Subtilis* ORFs determined by the International Project of Sequencing *B. Subtilis* Genome

A. Ogiwara[1]
ogi@ims.u-tokyo.ac.jp

T. Takagi[1]
takagi@ims.u-tokyo.ac.jp

N. Ogasawara[2]
nogasawa@bs.aist-nara.ac.jp

[1] Human Genome Center,
Institute of Medical Science, The University of Tokyo
4-6-1 Shirokanedai, Minato-ku, Tokyo 108 Japan

[2] Graduate School of Biological Science,
Nara Institute of Science and Technology,
8916-5 Takayama-cho, Ikoma, Nara 630-01 Japan

The sequencing project of *Bacillus Subtilis* genome is being carried out under the international collaboration of Japan and EU[1, 2]. The project is planned to cover the whole *Bacillus* genome of 4,188 kbp in 1996. At least, Japanese groups will complete sequencing about 1.3 Mbp continuous area within this fiscal year.

In order to open the data to the public, we have made a database system using the World-Wide Web (WWW). As the sequences derived from the project will also be deposited to the public data banks such as EMBL, DDBJ or SWISS-PROT, the distinct aim of our database is to provide various kind of functional information on open reading frames (ORFs) identified from the determined sequences.

The *Bacillus Subtilis* ORF database (**BSORF DB**) consists of the following parts:

1. Determined nucleotide sequences of contiguous part of the *Bacillus* genome, and translated amino acid sequences of identified ORFs
2. Entries of ORF information on various functions and the classification
3. Lists of the result of FASTA homology search for each ORF sequence
4. Corresponding SWISS-PROT entries and other database entries

An entry of ORF information contains the description of products, its functions, and some basic properties like molecular weight, isoelectric points, sequence motifs, and initiation codon and SD sequence that determined the ORF. It also describes the position of the ORF in the physical map.

In the tentative version of WWW BSORF DB, we are planning to provide two routes to access the ORF information. The first route begins with the figure of the location map of ORFs. Using a WWW client software, a user can click the position of the ORF on the figure to get the ORF information entry. In this figure, positions of GenBank entries are also drawn with arrow marks, which are also click-able to obtain contents of the GenBank entries.

The second access method using the category list of gene products will be available in the near future. The list is being constructed based on Riley's list[3]. By designating the category of the gene product on the hierarchical list, a user will also be able to reach the ORF entry.

[1] 荻原 淳, 高木 利久：東京大学医科学研究所ヒトゲノム解析センター, 〒 108 東京都港区白金台 4-6-1
[2] 小笠原 直毅：奈良先端科学技術大学院大学バイオサイエンス研究科, 〒 630-01 奈良県生駒市高山町 8916-5