A Database of *B. subtilis* Promoters and Transcription Factors

Takahiro Ishii
ishii@imcb.osaka-u.ac.jp

Kenta Nakai
nakai@imcb.osaka-u.ac.jp

Institute for Molecular and Cellular Biology, Osaka University
1-3 Yamada-oka, Suita, Osaka 565-0871, Japan

1 Introduction

Although the number of bacteria with their entire genomic sequence known is increasing, it is essential to study well-known genomes for understanding the 'blueprint' of bacteria more precisely. For this purpose, *E. coli* and *B. subtilis* are especially suitable because of their long history of research. Among various information coded in bacterial genomes, we are interested in the analysis of transcriptional regulation network. One of our ultimate goal is to predict the expression condition of given ORFs from their upstream sequences. For example, we have developed a prediction system of sigma-dependency of ORFs found in *B. subtilis* [6] and in *E. coli* (unpublished). However, to understand the detailed mechanism of transcription regulation, the knowledge of other transcription factors is also crucial. For *E. coli*, there is such a database called RegulonDB [3] but there is no databases containing comprehensive information of transcription in *B. subtilis*. Thus, we constructed a database which will be useful not only for our theoretical study but also for the works of experimental researchers.

2 Database

Basically, our database has inter-related components for promoters and for transcription factors. The former is a collection of 313 characterized promoters and its most important information is the positions of known *cis*-elements in each promoter. These positions are shown both in a table format and in a picture. In contrast, the transcription factor module is a set of entries corresponding to the names of 76 transcription factors including 10 sigma factors. Its main feature is the collection of known binding sites for each factor (consensus pattern is highlighted in red if any). The information of experimental techniques used for the characterization of each site is also noted. The main source of our database is the literature survey of 298 references but the benefit of other pioneering works such as the compilation by Helmann [4] is also used. Since the database is constructed on WWW, all references cited are linked to the PubMed database in NCBI, enabling the one-click check of their abstracts. In addition, the entries are also linked to other databases such as SWISS-PROT [1]. Currently, our database is released to limited users for beta-testing. One who wants to use it before its official release should ask the authors by e-mail.

3 Analyses

We took some statistics on *B. subtilis* σ^A-dependent promoters and compared them with those on *E. coli* σ^{70}-dependent promoters [2] to see if there are some specific features of *B. subtilis* promoters. The ratio of repressible promoters and activatable ones is about 6:4, which is similar to the value of *E. coli*. But it appears that the number of promoters with repeated homologous binding sites is significantly fewer in *B. subtilis* than in *E. coli.* In addition, we noticed that, unlike *E. coli* promoters, *B. subtilis* promoters can have some binding regions of activators even at the downstream regions of transcriptional initiation sites.

The amino acid similarity between transcription factors was also examined. We aligned 33 of them based on the regions corresponding to the helix-turn-helix motif. It was found that the classification of factors based on our alignment does not precisely correspond to the traditional family classification, described in SubtiList [5]. Since the traditional classification might be biased from historical reasons, further studies seems necessary to reconsider the classification.

Anyway, we hope that our database will become a useful resource for researchers of bacterial transcription as well as for our own study. We plan to release our database free of charge before long.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas, Genome Science, from the Ministry of Education, Science, and Culture of Japan. KN was also supported by Special Coordination Funds for Promoting Science and Technology from Science and Technology Agency and by Research for the Future Program of JSPS.

References

[1] Bairoch, A, and Apweiler, R., The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998., *Nucleic Acids Res.*, 26:38-42, 1998.

[2] Gralla, J.D., and Collado-Vides, J., Organization and function of transcription regulatory elements, In Neidhardt, F.C. *et al.* (eds), *Cellular and Molecular Biology: Eschericha coli and Salmonella*, 2nd edn. American Society for Microbiology, Washington, DC, 1232–1245, 1996.

[3] Huerta, A.M., Salgado, H., Thieffry, D., and Collado-Vides, J., RegulonDB: a database on transcriptional regulation in *Escherichia coli*, *Nucleic Acids Res.*, 26:55–59, 1998.

[4] Helmann, J.D., Compilation and analysis of *Bacillus subtilis* σ^A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA, *Nucleic Acids Res.*, 23:2351–2360, 1995.

[5] Moszer, I., Glaser, P., and Danchin, A., SubtiList: a relational database for the *Bacillus subtilis* genome, *Microbiology*, 141:261–268, 1995.

[6] Yada, T., Totoki, Y., Ishii, T., and Nakai, K., Functional prediction of *B. subtilis* genes from their regulatory sequences, *Proc. Intell. Syst. Mol. Biol.*, 5:354–357, 1997.