

# Database of Operons in *Bacillus subtilis*

Shujiro Okuda

okuda@kuicr.kyoto-u.ac.jp

Shuichi Kawashima

shuichi@kuicr.kyoto-u.ac.jp

Minoru Kanehisa

kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto  
611-0011, Japan

**Keywords:** operon, transcriptional unit, regulon, *Bacillus subtilis*

## 1 Introduction

The operon, a unit of transcription in prokaryotes, comprises a part of the complicated gene regulatory networks. Thus, the prediction of operons has been carried out at several laboratories [1, 2, 4]. However, for validation of the prediction method, a database of known operons is required with additional information about actual transcriptional units and the conditions when they are transcribed. In *Escherichia coli*, the RegulonDB [2] have been developed, which integrates biological knowledge on operons, regulons, promoters, terminators and so on [3], enabling researches to validate prediction operons. In contrast, an operon database of *B. subtilis* has not been developed. Hence, we have initiated organizing the information about operons and transcriptional units in literature and also in experiments.

## 2 Materials and Methods

### 2.1 Known Operons from Literature

We have collected known operon data in *B. subtilis* from literature. The data set was comprised of 111 operons including 366 ORFs. We define operons to include mono-transcribed units containing single ORFs. In our operon data set, there were 26 mono- and 85 multi-transcribed units (Table 1).

	Literature	BSORF
Transcriptional unit	111	467
Mono-transcribed	26	266
Multi-transcribed	85	201
ORFs	366	565

Table 1: Information of our database.

### 2.2 Transcriptional Units from BSORF

BSORF is a genome database about *B. subtilis* [5]. In the database, there are figures of transcriptional maps obtained by northern blotting experiments. We have translated the figure into sets of co-transcribed ORFs and stored them in a flat file. The file contained the ORF id, ORF name, transcriptional units and experimental conditions. The data comprises a total of 467 transcriptional units and 565 ORFs (Table 1).

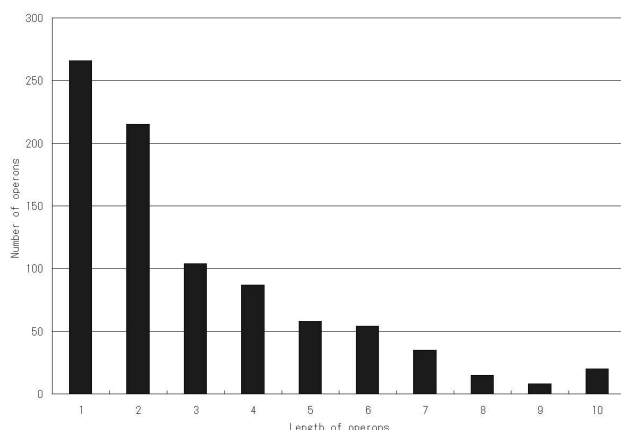


Figure 1: Distribution of the length of transcriptional units.

### 3 Results and Discussion

The distribution of the length of transcriptional units is shown in Figure 1. The length ranges from one to ten ORFs and mono-transcribed genes were most abundant. According to the northern blotting experiments, one operon identified from literature tend contain different transcriptional units with varying lengths, suggesting that operon are not always co-transcribed as a whole. We are currently analyzing microarray gene expression data in *B. subtilis* in the time-course experiments under different conditions and trying to predict operons based on the identification of co-expressed gene clusters along the chromosome. The database presented here will be useful to validate such predictions.

### Acknowledgments

This work was supported by the grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

### References

- [1] Sabatti, C., Rohlin, L., Oh, M-K., and Liao, J.C., Co-expression pattern from DNA microarray experiments as a tool for operon prediction, *Nucleic Acids Res.*, 30:2886–2893, 2002.
- [2] Salgado, H., Moreno-Hagelsieb, G., Smith, T.F., and Collado-Vides, J., Operons in *Escherichia coli*: Genomic analyses and predictions, *Proc. Natl. Acad. Sci. USA*, 97:6652–6657, 2000.
- [3] Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Millan-Zarate, D., Diaz-Peredo, E., Sanchez-Solano, F., Perez-Rueda, E., Bonavides-Martinez, C., and Collado-Vides, J., RegulonDB (version 3.2): Transcriptional regulation and operon organization in *Escherichia coli* K-12, *Nucleic Acids Res.*, 29:72–74, 2001.
- [4] Yada, T., Nakao, M., Totoki, Y., and Nakai, K., Modeling and predicting transcriptional units of *Escherichia coli* genes using hidden Markov models, *Bioinformatics*, 15:987–993, 1999.
- [5] <http://bacillus.genome.ad.jp/>