

# CyanoBase: The Genome Database for *Synechocystis* sp. strain PCC6803

Yasukazu Nakamura      Nobuyuki Miyajima  
ynakamu@kazusa.or.jp    miyajima@kazusa.or.jp  
Makoto Hirosawa        Takakazu Kaneko  
hirosawa@kazusa.or.jp    kaneko@kazusa.or.jp  
Satoshi Tabata  
tabata@kazusa.or.jp

Kazusa DNA Research Institute  
1532-3 Yana-Uchino, Kisarazu, Chiba 292 Japan

## Introduction

The cyanobacterium carries a complete set of genes for oxygenic photosynthesis, which is the most fundamental life process on the earth. This organism is also interesting from an evolutionary viewpoint, for it was born in a very ancient age and has survived in various environments. Chloroplast is believed to have evolved from cyanobacterial ancestors which developed an endosymbiotic relationship with a eukaryotic host cell.

The strain we have chosen for the genome analysis was the unicellular *Synechocystis* sp. strain PCC 6803, with the advantage of its transformable characteristics. We launched a project of sequencing the whole genome of the strain in October 1994. 3.57 Mb of nucleotide sequence was determined in May 1996. We deduced 3166 protein genes and 49 RNA genes based on similarity search results and computational analysis from the complete nucleotide sequence [1].

## Overview

In order to open the data to the scientific community, we have constructed a database system on the World Wide Web. As the annotated sequence data have been also registered in international DNA databanks (DDBJ/EMBL/GenBank), the aim of the database is to provide detailed information in user-friendly interface including Java-based clickable maps.

The URL of CyanoBase:

<http://www.kazusa.or.jp/cyano/cyano.html>

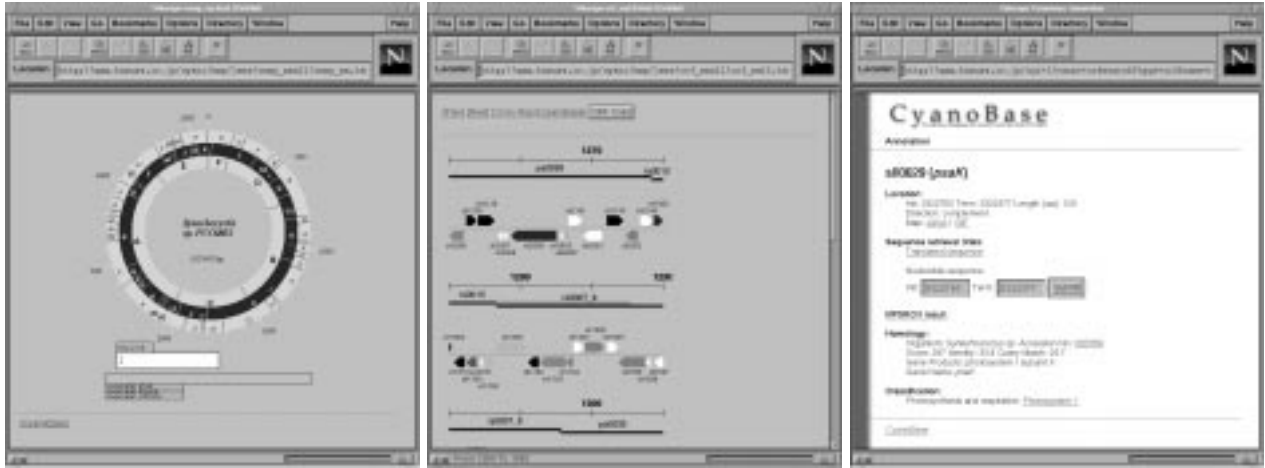


Figure 1: Sample images of CyanoBase

A user of the database can access full-annotated data of each gene through three ways; clickable maps (Java or GIF), a gene classification list and a keyword search engine. An entry of each gene has reverse-links for physical maps and the category list. Therefore, a user can access both locationally and classificationally neighboring genes by simply opening the link. An entry of information of a predicted protein gene contains 1) location data, 2) links for nucleotide or translated sequence, 3) a link and summary for similarity search result and 4) classification of deduced product.

Other services are also linked from the main page i.e. query form for similarity search, bibliography and a ftp link to distribute sequence or annotation data files etc. Any additional data to be published by us will be available through the database.

## Acknowledgements

We thank to Dr. Goto and KEGG project team for providing a browser for “Gene Category List” which is originally developed for KEGG by Dr. Akiyama, Institute for Chemical Research, Kyoto University. This work was supported by the Kazusa DNA Research Institute Foundation.

## References

- [1] T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda and S. Tabata “Sequence analysis of the genome of the unicellular Cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions” *DNA Res.*, Vol. 3, pp. 109–136, 1996.