ABSTRACT

CyanoBase provides internet access to the complete genomic information of the cyanobacterium Synechocystis sp. strain PCC6803. CyanoBase contains annotations to each protein-coding gene, deduced from the entire nucleotide sequence of the genome, gene classification lists, keywords and similarity search engines. The present paper describes a recent extension of CyanoBase, named CyanoMutants. CyanoMutants is a repository database of mutant information on PCC6803. Each entry contains a dataset which describes a gene identifier, mutant information, and an address for correspondence. Two closely-linked databases, CyanoBase and CyanoMutants, connect information obtained from computational analysis to experimental analysis resulting in the clarification of the functions of hypothetical genes of the cyanobacterial genome. CyanoMutants can be accessed at http://www.kazusa.or.jp/cyano/mutants/

INTRODUCTION

Cyanobacteria are prokaryotic microorganisms which carry a complete set of genes for oxygenic photosynthesis. Since their photosynthetic system resembles that of chloroplasts found in higher plants, cyanobacteria have been used as model organisms for the investigation of the apparatus and the mechanisms of oxygenic photosynthesis (reviewed in ref. 1). One of the unicellular cyanobacterial strains, Synechocystis sp. PCC6803, has been used to generate mutants by means of DNA recombination technology for two reasons: (i) its ability to be easily transformed and (ii) its ability to grow both photoautotrophically and photoheterotrophically, which allows for the isolation of disruption mutants of essential genes. Until now, a number of mutants with known mutations of the genes for various photosynthetic processes have been reported (3,4). In 1996, we reported a complete genomic sequence of Synechocystis sp. strain PCC6803 (5), in which 3168 potential protein-coding genes were deduced. Of 3168 potential protein-coding genes, 1722 were annotated as functionally-unassigned genes, which contained 1270 putative genes, 418 genes similar to EST of other genomes. To analyze the function of these genes, systematic gene disruption and characterization of the resulting mutants is thought to be one of the most promising strategies.

To post various PCC6803 genomic information in a user-friendly manner, we have created a web-based database, CyanoBase (6).CyanoBase provides the nucleotide sequence of the genome, annotation for protein-coding genes as well as a variety of tools to browse and extract genomic information. Here, we describe an extension of CyanoBase, a cumulative database for mutant information of PCC6803 genes, named CyanoMutants. The objective of CyanoMutants is to collect mutant information and facilitate understanding of the genes by connecting the structural and functional information of each gene.

CyanoMutants

CyanoMutants is a repository database which stores and provides mutant information through the WWW. An entry to CyanoMutants contains three essential sections; (i) identification of the mutated gene, (ii) information about phenotype and (iii) to whom correspondence should be addressed.

Each database entry links to a corresponding annotation in CyanoBase (Fig. 1). A user can follow the link and browse the annotation page of the protein-coding gene, which consists of information concerning the location of the genome, sequence retrieval links, similarity search results, and classification. When CyanoMutants stores a mutant for a protein-coding gene, the corresponding annotation page in CyanoBase shows a link to a page which provides mutant information in CyanoMutants. A link to a mutant page is not shown in the CyanoBase annotation page if mutant data for the gene has not been submitted to CyanoMutants.

DESCRIPTION

Adding a new mutant entry

A sample image of a submission page is shown in Figure 2. An entry is divided into the following sections.

(i) Information about the gene. In order to link to CyanoBase, an identifier of the mutated gene must be included. There are
optional input boxes for gene name, product name and the function of the product.

(ii) Mutant information. This section consists of a mutant name, a mutation type (interruption, deletion or site- or domain-directed mutation), phenotype if observed, phenotype details, segregation (complete or incomplete) and a storage type (DNA and/or Mutant).

(iii) Correspondence address and other information. Medline ID of publication (if available), correspondence (name, Email and web site address) and additional information.

Adding a new mutant from a WWW submission form has been simplified as much as possible in order to facilitate the process for the user. Although there are several input and check boxes on the page, only three sections are required to be completed in order to submit an entry, i.e., a gene identifier, phenotype (yes or no) and an address for correspondence (author’s name and Email). A researcher who wants to submit detailed data may use optional sections. Using the additional information box, any further format-free information such as an author’s postal address, whether it is distributable or not, and experimental details, can be included.

List and search

Information in the database is accessed in two ways; a list browser and a keyword search box. The list browser shows the first 20 entries sorted by the gene identifier as a default (Fig. 3). Sorting on the other keys can be performed by clicking ‘sort by’ links at the top of the columns. The identifier is a link to detailed information for each mutant. An entry of a mutant page shows each submitted data section listed in the list table. A search box allows a simple keyword search against the entire database of information including gene names, gene ids, researchers’ names and any text-based contents. Returned contents have the same format as the list browser described above. The searched keyword is shown by a highlight in the entry page.

As of September 1998, CyanoMutants contains 94 mutant entries, of which 62 have descriptions of phenotypes. The number of genes registered is expected to increase continuously since
extensive gene disruption experiments have been carried out since the release of the genome sequence of PCC6803.

ACKNOWLEDGEMENTS

We thank Drs T. Ogawa, M. Ikeuchi, T. Omata and W. Vermaas for helpful suggestions. This work was supported by the Kazusa DNA Research Institute Foundation.

REFERENCES