

Modeling a relational database to a Web of cyanobacteria synchronized with the NCBI

SANTOS, V. C. A.¹; NASCIMENTO, W. E. da S.¹; KAWASAKI, R.S.¹; SCHNEIDER, M. P. C.²; COUTO, D.C.C.²

¹College Computing Institute of Exact and Natural Sciences - ICEN, Federal University of Pará – UFPA, Belém.
{vitor.cirilo3, walbertsilva5}@gmail.com, kawasaki@ufpa.br

²Laboratory of polymorphic DNA - LPDNA, Institute of Biological Sciences, Federal University of Pará - UFPA, Belém. {paula, danifc}@ufpa.br

Keywords: Modeling Database, Database biological, Cyanobacteria.

Contemporary biotechnology has explored the microbial potential to control diseases, promising source of drugs, enzymes, among many other products useful to man. In this context, cyanobacteria, photosynthetic prokaryotes that are able to survive in adverse conditions, are receiving increasing attention for its ability to produce a wide variety of biologically active natural products. In the age of genomes, the progress made in biotechnology with the next-generation sequencers is providing an exponential increase in the number of genomes stored in biological databases, making the task of handling such data problematic. This paper aims to centralize information on cyanobacteria using as reference data for modeling files .fasta and .gb to GenBank database provided by the National Center for Biotechnology Information (NCBI), and convert these structures to a relational model database data using DBDesigner 4 and PostgreSQL version 8.1.4, storing the information of cyanobacteria in a local server to enable the construction of an efficient database and automatically updated with public database that will be needed to integrate a Web System on cyanobacteria. This will allow a fast and reliable access to comparative genomic analysis of these organisms. The importance of system modeling is given to the fact of the database to provide integration across heterogeneous databases (GenBank, Swiss-Prot, RefSeq, UniProt, KEGG) and allow the implementation of a search system to automate the analysis and gene predictions on the data of cyanobacteria. Initially, we implemented a relational database of genomes synchronizing your information with the NCBI, based on an unpublished study of Dieval Guizelini. Currently, our database has 42 complete genomes of cyanobacteria obtained via ftp from GenBank. The steps of modeling and data mapping of genomes has already been made and will support the modeling phase of database of proteins and bioactive natural products, which will allow integration of three databases of cyanobacterial informations into a single web system. This project is part of the unpublished doctoral thesis of Danielle Couto, which proposes the use of Data Mining techniques to predict biological information into the Web System called Cyanobr.