

Microbial Diversity Information Facility: Bacteriology Insight Orienting System (BIOS)

Junko Shimura^{1,2}, Hideyuki Shimizu¹, Keiji Tsuruwaka¹, Yukimitsu Moritani¹,
Kenji Miyazaki^{1,3}, Akira Tsugita³ and Makoto M. Watanabe¹

¹National Inst. Environmental Studies, 13-6 Onogawa Tsukuba, Ibaraki 305-0053, Japan

²RIKEN, 2-1 Hirosawa Wako, Saitama 351-0198, Japan

³JIPID Protein Data Bank, 2669 Yamazaki, Noda, Chiba 278, Japan

Abstract

Global Biodiversity is common interest of humans for better health and sustainable development of the society. To provide access and analysis on microbial diversity information, Bacteriology Insight Orienting System (BIOS) has been developed. BIOS contains 6402 species and subspecies names of bacteria and archaea, 2606 names of cyanobacteria by March 2000. BIOS of which web based analytical tool provides windows to compare the results of phylogenetic analysis based on 16S rDNA sequence and the results of cluster analysis on proteome profiling. The sequence data and 2 dimensional gel electrophoresis analysis data were accumulated in BIOS database content for cyanobacteria reclassification and taxonomy. (BIOS URL: <http://www-sp2000ao.nies.go.jp/bios/index.html>).

Introduction

In the internet era there are growing numbers of information resources for researchers to use. To access the reliable data and knowledge for studying biodiversity is however, becoming more and more difficult due to uncontrolled and distributed nature of the information sources. The system to orient users towards the appropriate information which covers 1.75 million of known species of the world is desired by researchers and decision makers. To cover the relevant biodiversity information and to allow users searching those distributed data and knowledge as if it was a single integrated system, large scale thesauri and harmonizing gateway interface for different existing databases are necessary. Taking these necessities into account, Species 2000 programme called for participation of existing databases to create Catalog of Life (2). While, Global Biodiversity Information Facility (GBIF) has been discussed at OECD megascience forum and successive committees, that an international mechanism

to develop such desired system cored with catalog of life is needed (3-5).

We have developed a prototype information facility with interests in bacteriology as BIOS. BIOS is a challenging system to create an open ended virtual facility to see insight of bacteriology by submitting and using distributed data resources on internet.

Materials and Methods

Two relational databases (Oracle 8 running on Sun Ultra Sparc II Deskside Workgroup Server, GP7000) were set up by compiling existing and newly developed species information files. We compiled data from "List of Bacterial Names with Standing in Nomenclature" (1) (<http://www-sv.cict.fr/bacterio/>) supplemented with "Bacterial Nomenclature up-to-date" (<http://www.dsmz.de/bactnom/bactname.htm>) to build this database with an agreement. Cyanobacteria names controlled by botanical code (15,16) were newly made in digital form and stored in a database. Common gateway interface (CGI) program to send query to the two separate databases was developed in Perl language (ver. 5). We developed software to calculate the similarity value of 2-DE profile among tested species by simple matching of spots to an out-group species. This standardized similarity matrix was subjected to cluster analysis to deduce tree view of the relationship of the species. The cluster analysis for 2-DE spots was performed by unweighted pair-group method using arithmetic averages (UPGMA). The module to analyze sequences was used a public domain software, CLUSTAL W Multiple Sequence Alignment Program(ver.1.7,1997, <http://bioinformers.ebi.ac.uk/newsletter/archives/2/clustalw17.html>.)

Results and Discussion

BIOS was built by assembling two databases, one for bacteria/archaea database that are solely controlled by International Code of Nomenclature of Bacteria (ICNB) (6) and another one as separate database for cyanobacteria, which previously controlled by International Code of Botanical Nomenclature (ICBN) (7). The prototype information facility was designed to allow accessibility to relevant taxonomic data and to use taxonomic tools through WWW browser from remote clients. The system components in BIOS were shown in Fig. 1. Bacteria and archaea database and its CGI software to query by names provides the screen views shown in Fig. 2.

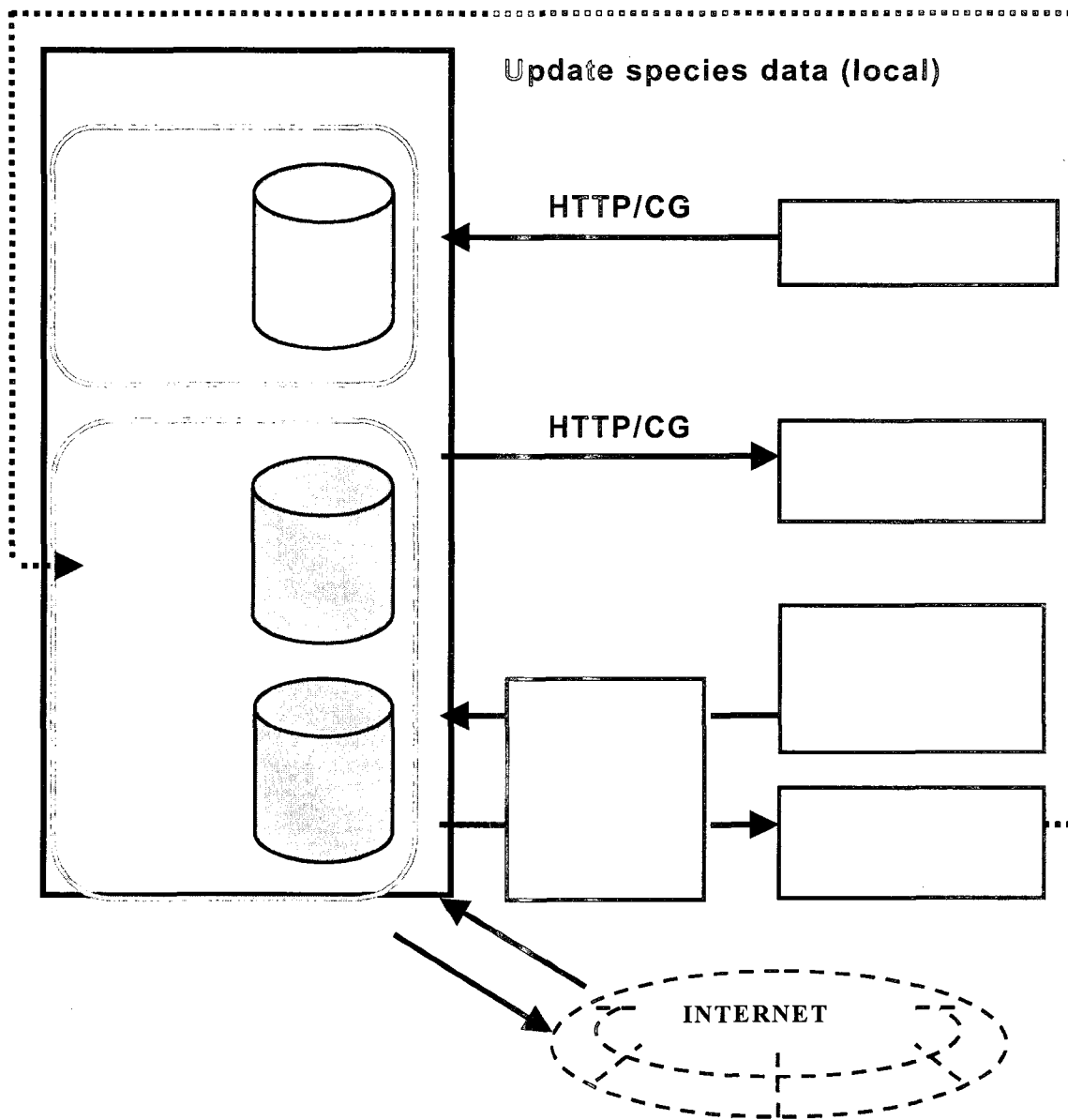


Fig.1 System Components of BIOS and Data Flow

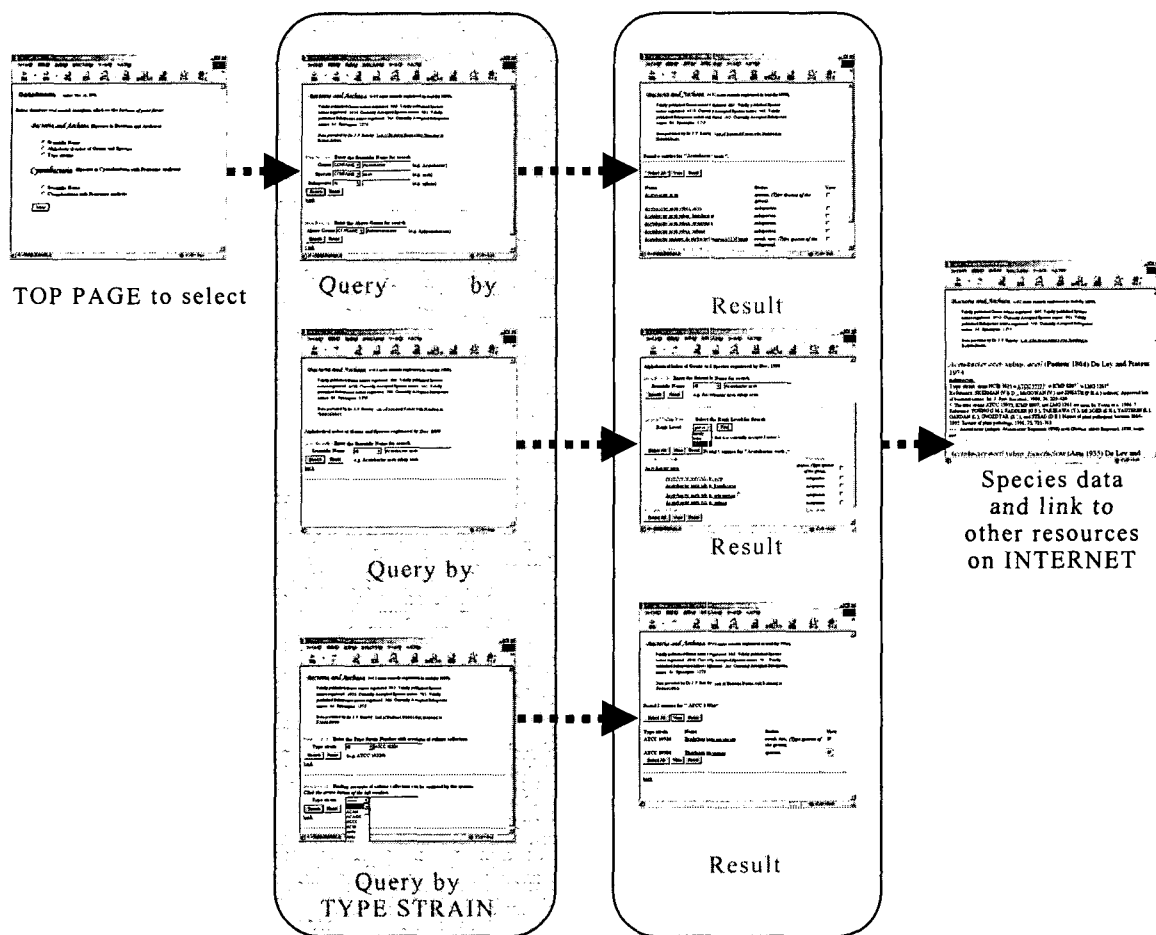


Fig. 2. A Database for Bacteria and Archaea Species

The window on top of the middle column in Fig. 2 is a search screen by species names. Second window provides a screen to search with scientific name that indicates rank such as “order”, “family”, “genus”. The species and subspecies on the same level of the hierarchy are shown in alphabetical order. The data is shown with flags indicating whether or not the name is currently accepted. The flag indicator in the alphabetical index of species is useful to reuse this database to further integrate into global biodiversity information systems with other existing databases for plants, animals etc. in future. Users can also send type strain number to retrieve the valid species name and through the interface of BIOS currently accepted name for the strain can be reached. History of changing species name due to new combination can also be found in the contents of the database and it is indicated with arrow on the screen of WWW browser. Other

synonyms than new combination including illegitimate names and very old names such as publication before 1980 are also listed in “synonym”.

User interface to use cyanobacteria database on WWW is shown in Fig. 3. There is at the moment only one window to send query by species name. The link to preserved materials are will be set in future.

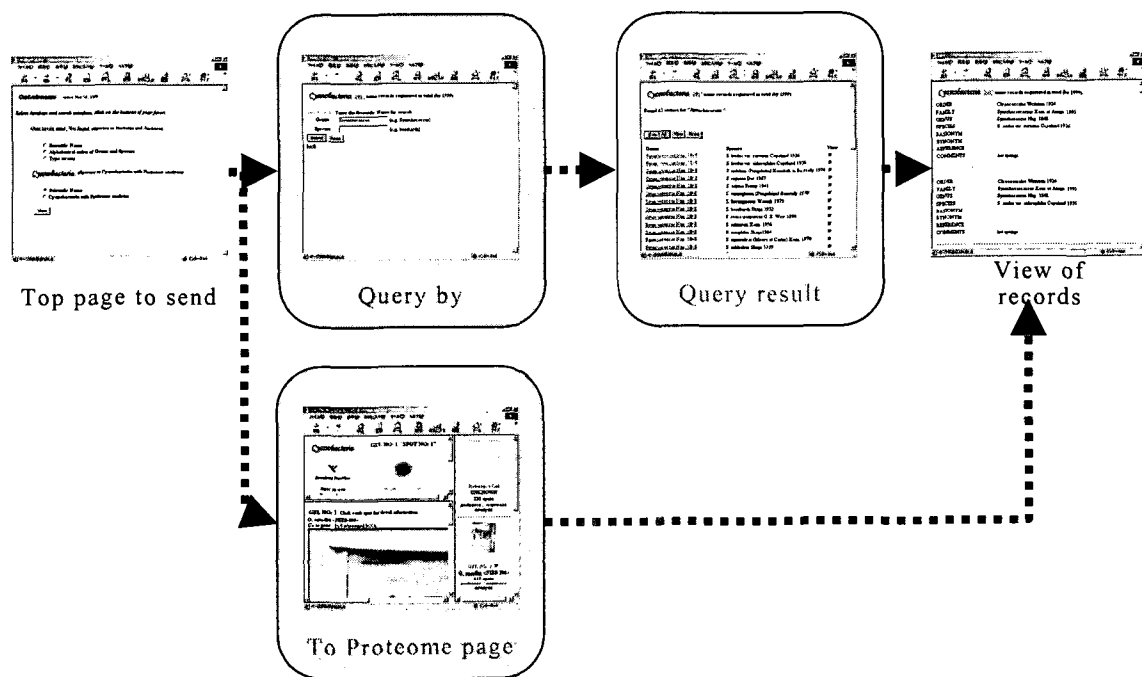


Fig. 3. A Database for Cyanobacteria Species

To facilitate to subject cyanobacteria into bacteriological system in its taxonomy, we also developed tools to analyze phylogeny based on sequence information and proteome data. In the evolution of life, genetic mutations are successively transferred from an ancestor to descendants. Substitution, insertion, inversion, deletion, duplication and crossing over of genes are concerned to contribute to the divergence of species. Comparing mutations in a common gene sequence has been thought that it can trace the evolutionary path of the organisms as deductive geneological tree. However, studies have shown that sequences of a single gene from different species would sometimes not result in a consistent phylogenetic tree when different genes or different algorithm were applied to the analysis

(14,11). These phylogenetically deduced group could reveal that the grouping is discrepant from the morphological, physiological or biochemical grouping of the subjected species (8,10,12). This evidence implies that comparing only one gene does not fulfill the required information to analyze the evolution of the species in general.

Cyanobacteria has recently been studied in 16S rRNA/DNA or 16S to 23S internal transcribed spacer sequences. Again, evidences showed that monophyletic group could include independent taxonomic lineages which were based on their phenotypic characters (9).

New approach to deduce the evolutionary effected gene product to find reasonably explaining the divergence of species among cyanobacteria is therefore strongly desired. Two dimensional gel electrophoresis (2-DE) is an approach to visualize the most of gene product of the cell. Proteome analysis on *Synechocystis* sp. strain PCC6803 has successfully made a linkage between genome information and 2-DE spots (12). This suggests that 2-DE patterns may represent the whole genome that produce functioning proteins. Comparing the profiles on 2-DE between species and detecting the particular spot that is found in only particular group of the species can be count as an evolution marker. This marker is thought as a taxonomic marker to lead to new insights into relationship of species. We developed software to calculate the similarity value of 2-DE profile among tested species by simple matching of spots to an out-group species. This standardized similarity matrix was subjected to cluster analysis to deduce tree view of the relationship of the species. By using this prototype system, we have analyzed proteome of *Oscillatoria agardhii*, *Microcystis aeruginosa*, an *Anabaena affinis*. The performance of the system was reasonably fast to develop this database as open ended sharing system on high speed information network.

Acknowledgements

Authors gratefully thank to Prof. Jean P. Euzeby, Prof. Erko Stackebrandt, Dr. Norbert Weiss, Dr. Manfred Kracht and Dr. Shoichiro Suda for their kind data submission and suggestions for development of BIOS.

References

1. Euzeby, J.P. 1997. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* 47:590-592
2. (<http://www.sp2000.org>)

3. (http://www.oecd.org//dsti/sti/s_t/ms/prod/birepfin.pdf)
4. (http://www.oecd.org//dsti/sti/s_t/ms/prod/gbif.htm)
5. (<http://www.york.biosis.org/gbif/>)
6. *International Code of Nomenclature of Bacteria*. 1992 (1990 Revision)
7. *International Code of Botanical Nomenclature*. 1972
8. Neilan, B. A., D. Jacobs, T. Del Dot, L. L. Blackall, P. R. Hawkins, P. T. Cox, and A. E. Goodman. 1997. rRNA sequences and evolutionary relationships among toxic and nontoxic cyanobacteria of the genus *Microcystis*. *Int. J. Syst. Bacteriol.* **47**: 693-697
9. Otsuka S., S. Suda, R Li., M. Watanabe, H. Oyaizu., S. Matsumoto and M. M. Watanabe. 1999. Phylogenetic relationships between toxic and non-toxic strains of the genus *Microcystis* based on 16S to 23S internal transcribed spacer sequence. *FEMS Microbiol. Lett.* **172**:15-21
10. Rudi, K., O. M. Skulberg and K. S. Jakobsen. 1998. Evolution of cyanobacteria by exchange of genetic material among phylogenetically related strains., *J. Bacteriol.* **180**: 3453-3461
11. Russo, C. A., N. Takezaki, and M. Nei. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny., *Mol Biol Evol* **13**:525-536
12. Sazuka, T., M. Yamaguchi and O. Ohhara. 1999. Cyano2Dbase updated: linkage of 234 protein spots to corresponding genes through N-terminal microsequencing. *Electrophoresis* **20**:2160-2171
13. Stackebrandt E., F. A. Rainey and N. Ward-Rainey. 1996. Anoxygenic phototrophy across the phylogenetic spectrum: current understanding and future perspectives. *Arch. Microbiol.* **166**:211-223
14. Takezaki N. and M. Nei. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**:389-399
15. Umezaki, I. and M. Watanabe. 1994. Enumeration of the cyanophyta (blue-green algae) of Japan. 1. Chroococcales and Oscillatoriales.. *Jpn. J. Phycol.* **42**:175-219
16. Umezaki, I. and M. Watanabe. 1994. Enumeration of the Cyanophyta (blue-green algae) of Japan. 2. Nostocales and Stigonematales. *Jpn. J. Phycol.* **42**:301-324d