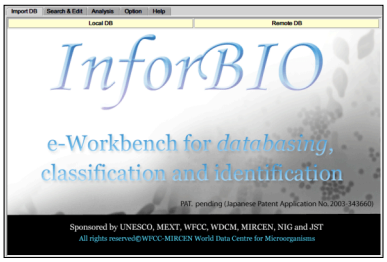
e-Workbench (InforBIO) for Barcoding of Life

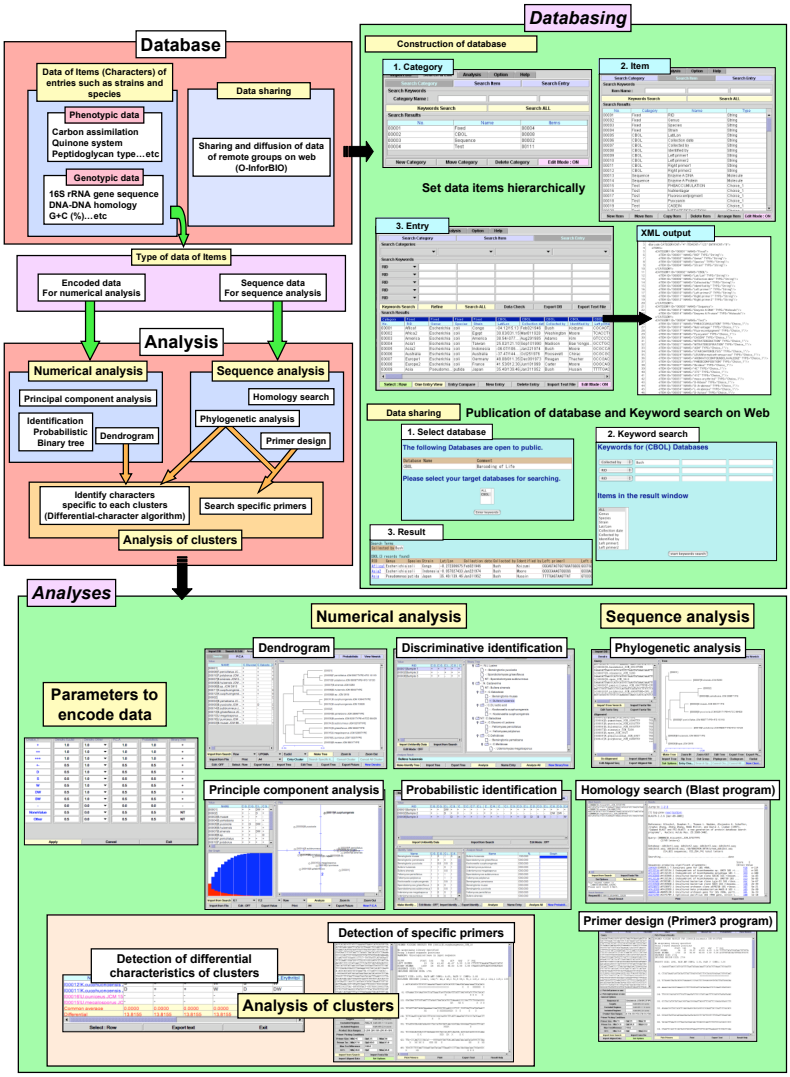
Hideaki Sugawara(1), Naoto Tanaka(2), Takashi Abe(1) and Satoru Miyazaki(3)

(1) Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan (2) Institute for Bioinformatics Research and Development, Japan Science and Technology Agency, 5-3 Yonbancho, Chiyoda-ku, Tokyo 102-8666, Japan (3) Tokyo University of Science, 2641 Yamazaki, Noda, Chiba 278-8510, Japan

Keywords: Database, Classification, Identification, Phylogenetic analysis, Data sharing

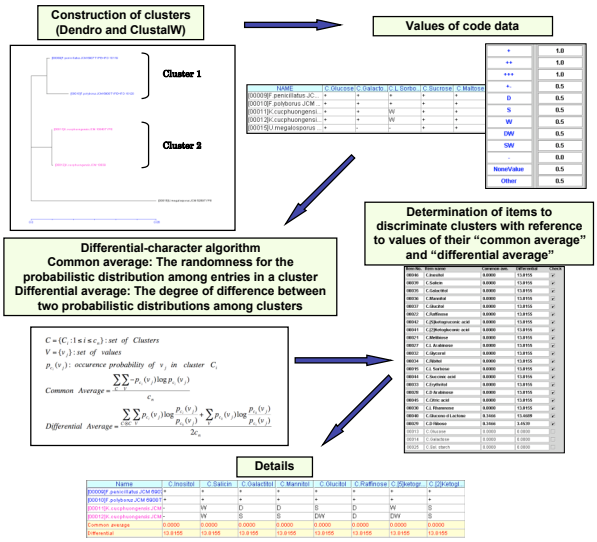


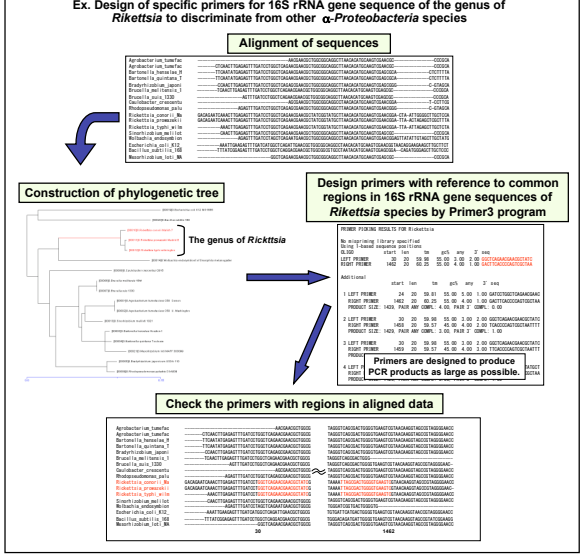
The functions of InforBIO include databasing of phenotypic and genotypic data, e.g. sequences of multiple genes, numerical analysis of phenotypic data and sequence data analysis. InforBIO uses an XML (eXtensible Markup Language) document for its main file in a way to realize a flexible database management. As analytical tools, InforBIO is equipped with Dendro analysis, Principal component analysis (PCA), discriminative and probabilistic identification. For sequence analysis, Blast, ClustalW and Primer3 are also added. Results of cluster analyses (Dendro analysis, PCA, and Clustal W) can be compared with each other graphically, and differential characteristics of clusters based on Dendro analysis and ClustalW can be also extracted from a database based on mutual information entropy. Additionally, we attempt to reflect the result of alignment by ClustalW to design specific primers by using Primer3. The e-Workbench is written in Java so that it is executable in Windows, Macintosh, Linux and UNIX machines and available free through the Internet to be run in a laptop computer. We have recently developed G-InforBIO for the comparative genomics and O-InforBIO to share and diffuse data of remote groups. The e-Workbench is available online at http://www.wdcm.org/. Therefore, groups of barcoding of life will be able to store and analyze their data at each site and also integrate their data for publication to the Internet.



Analysis of clusters

Determination of characteristics to discriminate clusters



Search specific primers of a cluster

Ex. Design of specific primers for 16S rRNA gene sequence of the genus of *Rikettsia* to discriminate from other α-*Proteobacteria* species