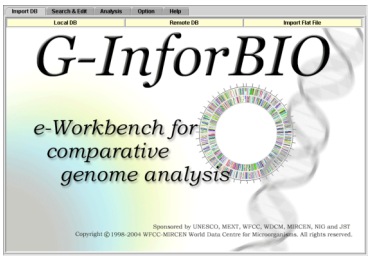
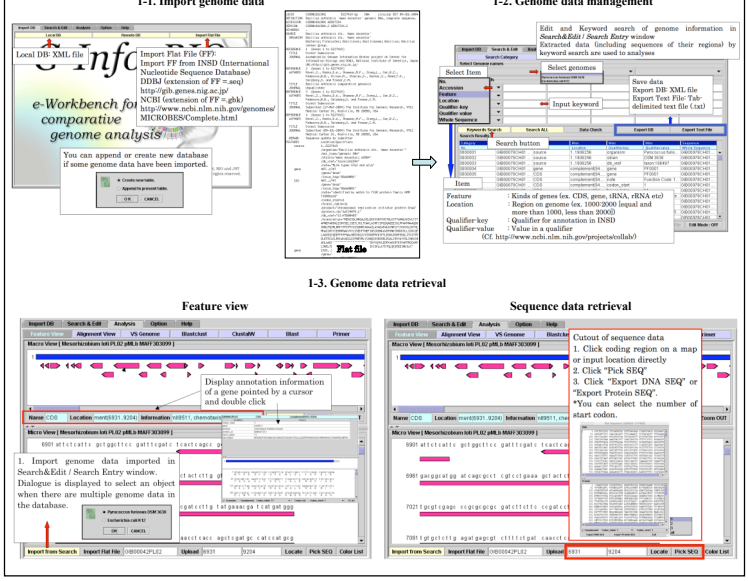
G-InforBIO, an integrated suite for comparative genomics

Naoto Tanaka, Satoru Miyazaki, and Hideaki Sugawara
Center for Information Biology and DNA Data Bank of Japan
National Institute of Genetics
Shizuoka 411-8540, Japan



We have developed a suite (G-InforBIO) for management and analysis of genomic data published in the International Nucleotide Sequence Database (DDBJ/EMBL/GenBank) and also local data in research laboratories. Major functions of G-InforBIO are database management, data retrieval, sequence data analysis and the visualization of data correlation. The database accommodates multiple flat files of entries registered in INSD and also tab-delimited files of local data. A subset of the database can be retrieved for the further analysis and graphical representation. G-InforBIO includes such programs for sequence data analysis as Blast, ClustalW, MegaBlast and Primer3. The analysis can be applied to the subset of the G-InforBIO database and data imported from databases outside. We, especially, contrived the viewer (called Alignment Viewer) to interpret the result of MegaBlast for the comparison of two genomes. The viewer will support for us to infer the dynamics of genome evolution and reorganization. G-InforBIO will be downloadable from WFCC-MIRCEN World Data Centre for Microorganisms site (http://www.wdcm.org/)

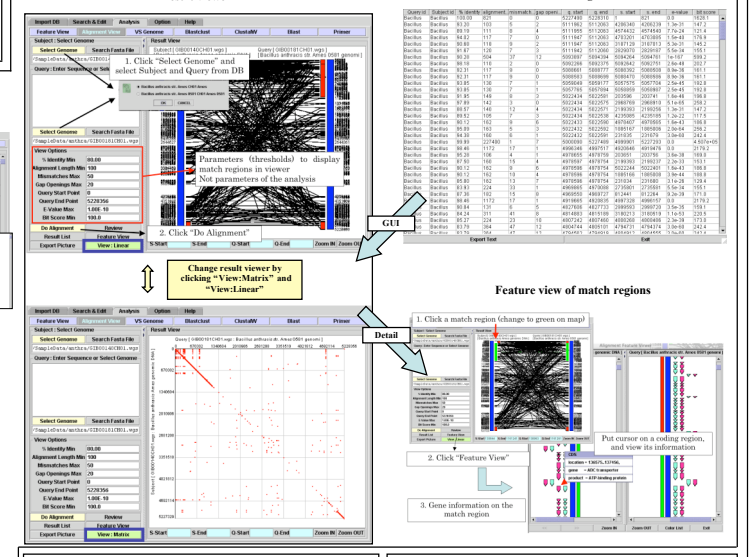
1. Genome data management and retrieval



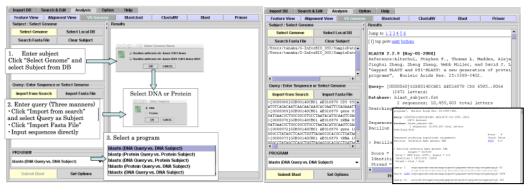
2. Sequence data analysis

2-1. Homology analysis

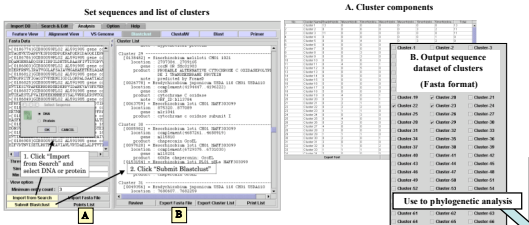
2-1-1. Alignment view: Blast to genome against genome (Megablast)



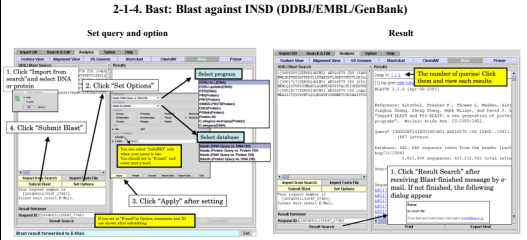
2-1-2. VS genome: Blast against multiple genomes



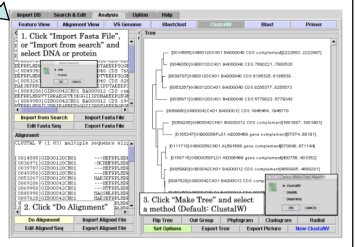
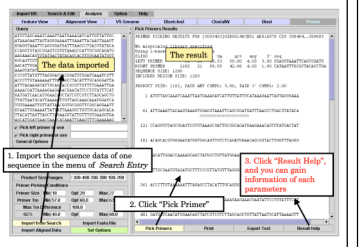
2-1-3. Blastclust: Blast score-based single-linkage clustering for nucleotide and amino acid sequences



2-1-4. Bast: Blast against INSD (DDBJ/EMBL/GenBank)



2-2. Phylogenetic analysis



2-3. Primer design