Abstract 92

COMPREHENSIVE MICROBIAL GENOME DISPLAY AND ANALYSIS

Rajesh Laik[1], Frank Larimer[2], Doug Hyatt[2], Miriam Land[2], Richard Mural[2], Morey Parang[2], Manesh Shah[2], Jay Snoddy[2] and Ed Uberbacher[2]
[1]BIT Mesra, Department of Biotechnology, Ranchi, India
[2]Computational Biosciences and Toxicology and Risk Analysis Sections, Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

We are now representing all completed microbial genomes in the Genome Channel and the Genome Catalog, providing comprehensive sequence-based views of genomes from a full genome display, to the nucleotide sequence level. We have developed a tool for comparative multiple genome analysis that provides automated, regularly updated, comprehensive annotation of microbial genomes using consistent methodology for gene calling and feature recognition. The visual genome browser currently represents ca. 51,000 Microbial GRAIL gene models as well as providing over 45,000 GenBank gene models. Precomputed BEAUTY searches are provided for all gene models, with links to original source material as well as links to additional search engines. Comprehensive representation of microbial genomes will require deeper annotation of structural features, including operon and regulon organization, promoter and ribosome binding site recognition, repressor and activator binding site calling, transcription terminators, and other functional elements. Sensor development is in progress to provide access to these features. Linkage and integration of the gene/protein/function catalog to phylogenetic, structural, and metabolic relationships are being developed. A draft analysis pipeline has been constructed to provide annotation for the microbial sequencing projects being carried out at the Joint Genome Institute. The pipeline is being applied to annotating the Nitrosomonas europaea and Prochlorococcus marinus genomes currently being sequenced. Multiple gene callers (currently Generation, Glimmer and Critica) are used to construct a candidate gene model set. The conceptual translations of these gene models are used to generate similarity search results and protein family relationships; from these results a metabolic framework is constructed and functional roles are assigned. Simple repeats, complex repeats, tRNA genes and other structural RNA genes are also identified. Annotation summaries are made available through the JGI Microbial Sequencing web site; in addition, draft results are being integrated into the interactive display schemes of the Genome Channel/Catalog