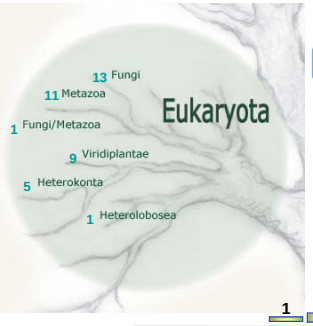
Eukaryotic Genome Annotation and Analysis @ JGI

Igor Grigoriev, Andrea Aerts, Frank Korzeniewski, Alan Kuo, Asaf Salamov, Robert Otillar, Xueling Zhao, Kemin Zhou*

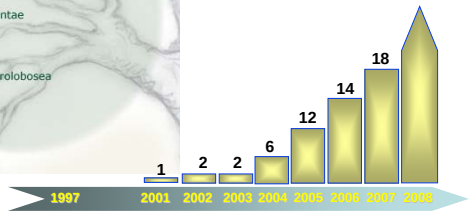
DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
*Correspondence: ivgrigoriev@lbl.gov

Abstract

Over 40 eukaryotic genomes have been annotated using JGI Annotation Pipeline, which combines several gene prediction and annotation tools. Predicted genes and annotations are accessible via interactive visualization tools integrated into JGI Genome Portal. The diverse set of annotated genomes represents the major branches of the tree of life and provides basis for comparative genomics analysis. This comparative analysis as well as analysis of available experimental data (ESTs, microarrays, proteomics) enable validation of predicted genes. In addition, community-driven manual curation of predicted genes and functions in annotated genomes further improves their quality.

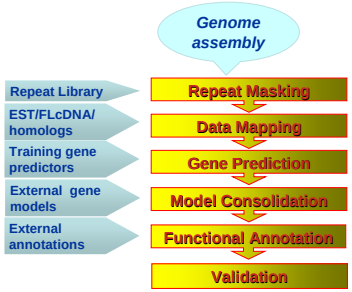


Annotated JGI Genomes

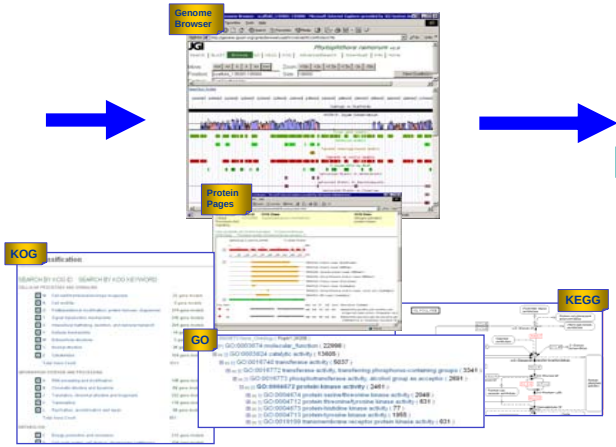


Annotation Process

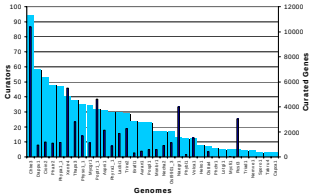
Annotation Pipeline



Genome Portal



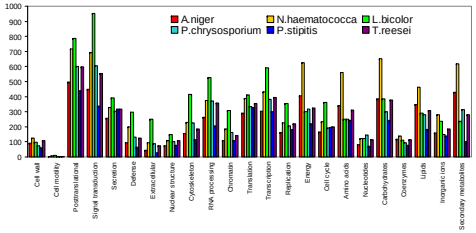
Manual Curation



800+ registered curators, 25+ jamborees

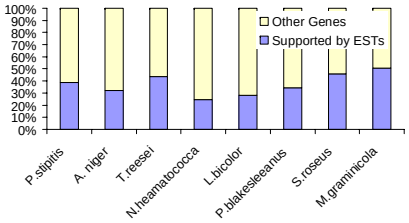
Validation using Comparative Genomics and Experimental data

Functional Comparative Analysis



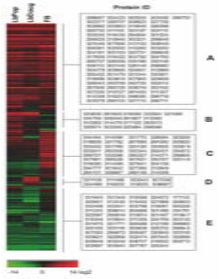
Comparative analysis of gene counts in different KOG categories of several genomes reveals expandes and reduced gene families

EST Support



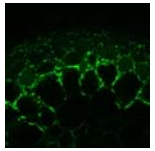
On average,1/3 of predicted genes are supported by in-house Sanger ESTs. Pyrosequencing (454FLX) can increase this number to 2/3 or higher

Microarray Expression



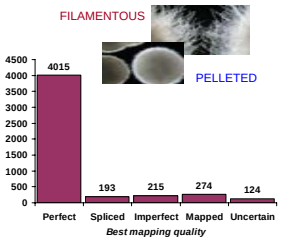
Laccaria bicolor microarray demonstrated expression of 79.4% of predicted genes including small secreted proteins differentially expressed symbiotic tissues and fruiting bodies (B-C). Martin *et al., Nature 2008*

Immunolocalization



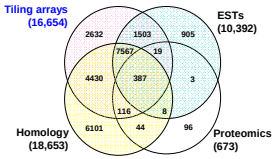
Immunolocalization of MISSP7, a Mycorrhiza-induced Small Secreted Protein of Laccaria bicolor
Martin *et al., Nature 2008*

Proteomics



1,440 (14%) *Phanerochaette chrysosporium* predicted genes were supported with 4,825 peptides (Baker *et al*, PNNL).

Multiple Lines of Evidence



77% of 30,940 predicted genes in *Daphnia pulex* are supported by at least one line of evidence (Colbourne et al)

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396