Lawrence Berkeley National Laboratory
Recent Work

Title
Eukaryotic Genome Annotation and Analysis @ JGI

Permalink
https://escholarship.org/uc/item/51v0r5sh

Authors
Grigoriev, Igor
Aerts, Andrea
Korzeniewski, Frank
et al.

Publication Date
2008-03-26

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
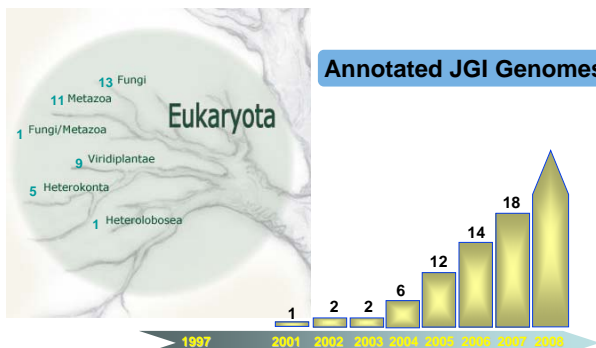*Correspondence: ivgrigoriev@lbl.gov
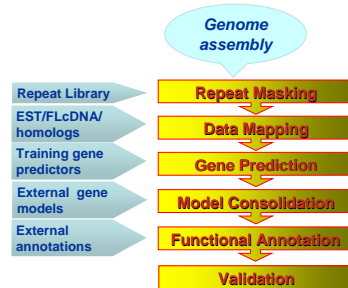
Abstract

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.
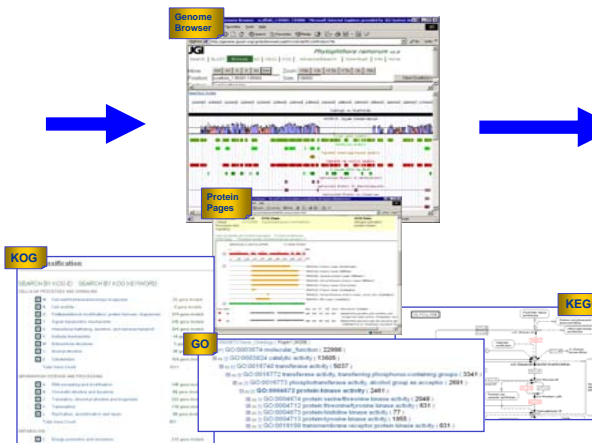
Annotated JGI Genomes

Eukaryota
- 13 Fungi
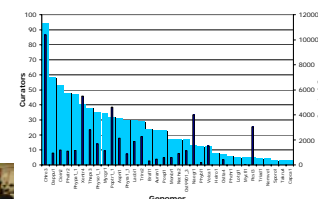- 11 Metazoa
- 1 Fungi/Metazoa
- 9 Viridiplantae
- 5 Heterokonta
- 1 Heterolobosea

Genomes annotated by year: 1997: 1; 2001: 2; 2002: 2; 2003: 6; 2004: 12; 2005: 14; 2006: 18

Annotation Process

Annotation Pipeline

Genome assembly

- Repeat Library → Repeat Masking
- EST/FLcDNA/homologs → Data Mapping
- Training gene predictors → Gene Prediction
- External gene models → Model Consolidation
- External annotations → Functional Annotation
- Validation

Genome Portal

Genome Browser, Protein Pages, KOG, GO, KEGG

Manual Curation

Functional Annotation, Structural Annotation
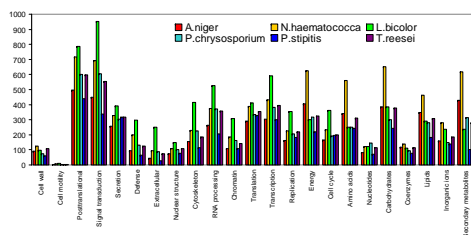
800+ registered curators, 25+ jamborees

Validation using Comparative Genomics and Experimental data

Functional Comparative Analysis

A.niger, N.haematococca, L.bicolor, P.chrysosporium, P.stipitis, T.reesei

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

EST Support

Other Genes / Supported by ESTs — P.stipitis, A.niger, T.reesei, N.heamatococca, L.bicolor, P.blakesleeanus, S.roseus, M.graminicola
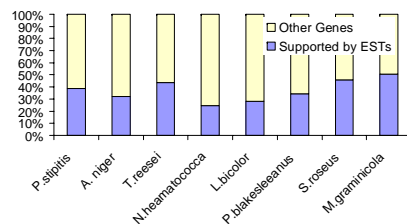
On average,1/3 of predicted genes are supported by in-house Sanger ESTs. Pyrosequencing (454FLX) can increase this number to 2/3 or higher
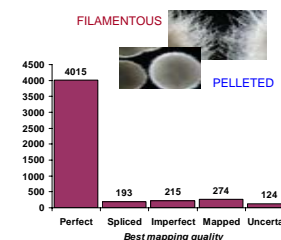
Microarray Expression
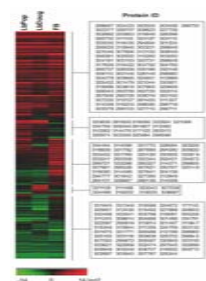
Laccaria bicolor microarray demonstrated expression of 79.4% of predicted genes including small secreted proteins differentially expressed symbiotic tissues and fruiting bodies (B-C). Martin *et al., Nature* 2008

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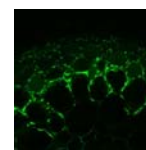
Immunolocalization of MISSP7, a Mycorrhiza-induced Small Secreted Protein of Laccaria bicolor
Martin *et al., Nature* 2008

Proteomics

FILAMENTOUS / PELLETED
Best mapping quality — Perfect: 4015; Spliced: 193; Imperfect: 215; Mapped: 274; Uncertain: 124

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

Multiple Lines of Evidence

Tiling arrays (16,654); ESTs (10,392); Homology (18,653); Proteomics (673)
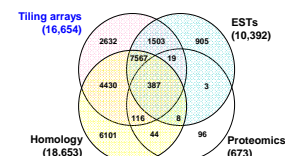
77% of 30,940 predicted genes in *Daphnia pulex* are supported by at least one line of evidence (Colbourne et al)

US DOE Joint Genome Institute