Creating a Critical Mass of Data for Genome Annotation and Comparative Analysis

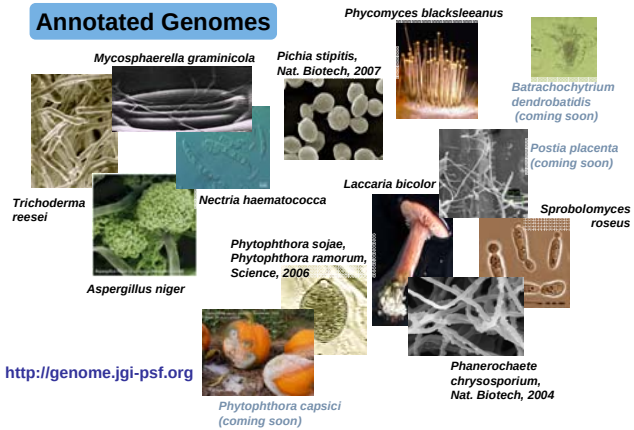
Igor Grigoriev, Andrea Aerts, Alan Kuo, Asaf Salamov, Diego Martinez, Jean Challacombe, Thomas Brettin, Erika Lindquist, Harris Shapiro, Gregory Werner, Susan Lucas, Jane Grimwood, Jeremy Schmutz, Paul Richardson, James Bristow*

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

Abstract

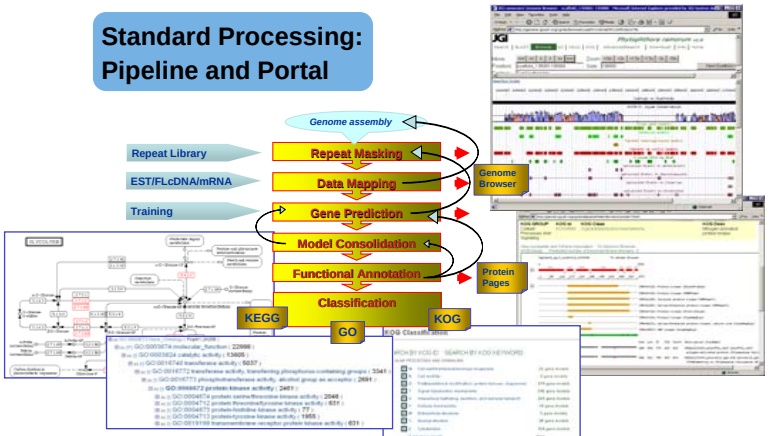
The DOE Joint Genome Institute contributes to the rapidly growing number of fungal genomes available to the research community through high-throughput sequencing and annotation. We sequenced and annotated 9 fungal and 2 oomycete genomes (genome.jgi.doe.gov) and expect to nearly double this number in 2007.
The broad and diverse set of genomes covers major branches of the fungal phylogenetic tree and samples some of the branches at higher depth, which allows analysis of differences between clades and within species on the genomic level. In addition to several basidiomycete and ascomycete genomes, we have annotated the genome of the zygomycete *Phycomyces blakesleeanus* and are currently annotating the chytridiomycete *Batrachochytrium dendrobatidis*.
Many of these genomes are either finished or are in finishing, which enables effective analysis of genome organization and evolution, such as, for example, a conditionally dispensable chromosome in *Nectria haematococca* MPVI. Genome projects triggered development of additional resources of ESTs, proteomics and microarray data to assist with genome annotation and expression analysis. User Communities with over 100 biologists around the world actively participate in analysis and manual curation of predicted genes and functions using the JGI Genome Portal.
Creating a critical mass of data improves genome annotation and enables efficient comparative analysis to address a broad spectrum of biological questions and to find important industrial applications.

Annotated Genomes



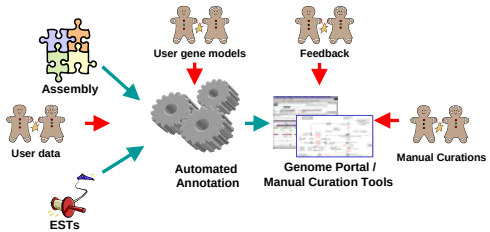
http://genome.jgi-psf.org

Standard Processing: Pipeline and Portal



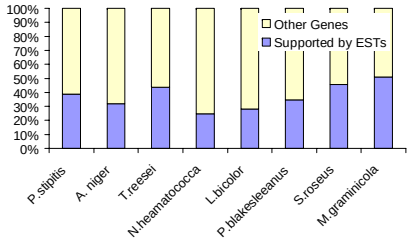
JGI Annotation Pipeline **provides a standard framework for uniform annotation of a diverse set of genomes**

JGI Genome Portal **provide visualization tools for genome analysis and manual curation, in which we actively involve** ***User Communities***

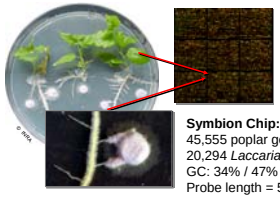


Additional Genome Driven Resources

ESTs Support 25-50% of Predicted Genes

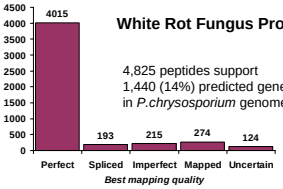


Microarrays: Analysis of Interactome of Symbiosis



Symbion Chip:
45,555 poplar genes +
20,294 *Laccaria bicolor* genes
GC: 34% / 47% (for *Poplar/L.bicolor*)
Probe length = 50-70/ 34-44 bases

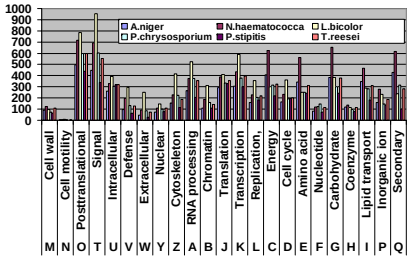
White Rot Fungus Proteomics



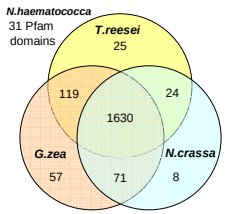
4,825 peptides support
1,440 (14%) predicted genes
in *P.chrysosporium* genome.

Validation of Annotations using Comparative Genomics

Functional Comparative Analysis



Over- and under-represented KOG categories in different fungi

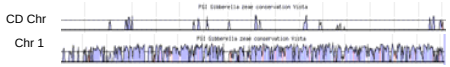


Over- and under-represented PFAM domains

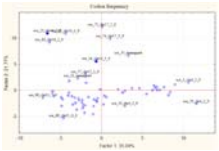
Core Gene Set Shared between the Genomes

Clustered with PhiGs (P.Dehal and J.Boore, 2006)

VISTA Genome Conservation



Codon Usage Analysis

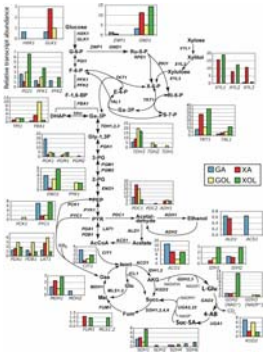


Conditionally Dispensable (CD) chromosome in contrast to other chromosomes in *Nectria haematococca* MPVI shows significantly lower genome conservation and shows bias in codon usage

Metaboic Reconstruction

Central metabolic pathway regulation in *P. stipitis:*
Cells grown on minimal defined medium under four conditions: glucose aerobic (GA), xylose aerobic (XA), glucose oxygen limited (GOL) and xylose oxygen-limited (XOL).

Jeffries et al., (2007)
Nature Biotechnology



US DOE Joint Genome Institute

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 Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.