

Introduction to JGI & MycoCosm

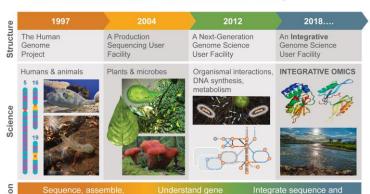
Steven Ahrendt, PhD (he/him)
Data Scientist
Fungal & Algal Genomics Program



Joint Genome Institute



The Evolution of JGI as a National User Facility



functional capabilities for



- US Department of Energy User Facility
- Located at Lawrence Berkeley National Laboratory (LBNL) in Berkeley, CA
- Leading the integration and application of genomics for energy and environmental research

Science programs:

- Fungal & Algal Program
- Plant Program
- Metagenome Program
- Microbial Program
- DNA Synthesis Science Program
- Metabolomics Program
- Secondary Metabolites

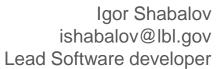


The Fungal & Algal Team





Igor Grigoriev ivgrigoriev@lbl.gov Fungal & Algal Program Lead





- Software developers
- Data scientists
- Postdoctoral scholars
- Interns

























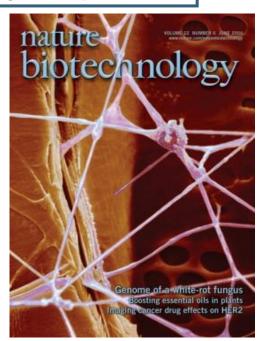


JGI Fungal Genomics – A growing field



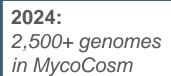
2004:

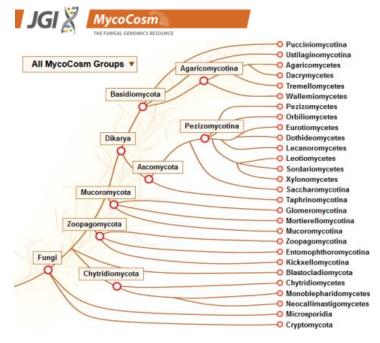
1st basidiomycete genome published



White rot fungus

Phanerochaete chrysosporium





JGI Community Science Program











Short & long reads

Metabolomics



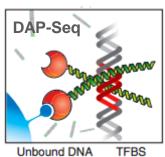
	Proposal Call Type	Review Frequency	Next Submission Deadline	Next Review Date
•	CSP Annual	Annual	Spring 2025	August 2024
	FICUS JGI-EMSL	Annual	Spring 2025	June 27-28 2024
	CSP New Investigator	Annual	Oct 4, 2024	Dec 2024
	CSP Functional Genomics	Annual	Jan 30, 2025	March 2025

for Proposals

Calls



Bioinformatics



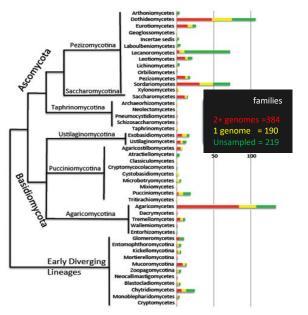
MODOLOGICAL

TFBS

1000 Fungal Genomes Project



Goal: Produce a reference genome for every family of fungi

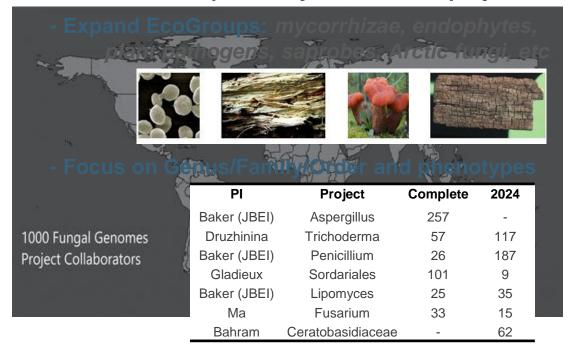


105 participants: 1080 total projects 220 await DNA/RNA; 140 in progress

720 completed: 318 published

NEXT: Toward 10,000 fungal genomes

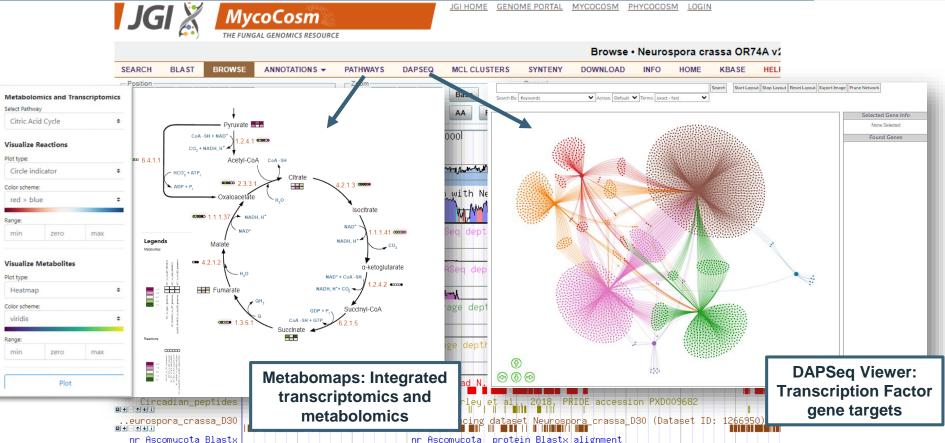
- 3-5 references per family within 1KFG project



- Environmental genomics: MAGs, forays, Myco-Ed

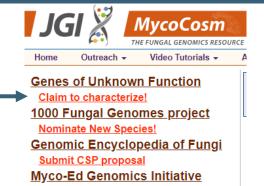
Integrative Multi-omics

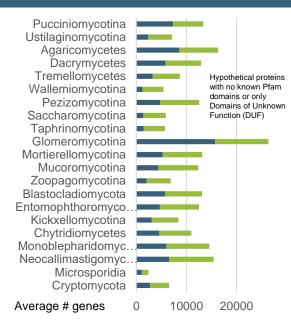


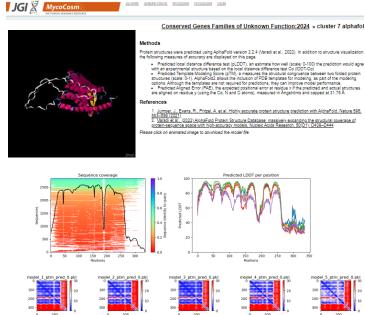


Conserved Genes of Unknown Function









Total Unique PFAM Domains: 80

Total Unique Uniprot: 126

Total Genes: 162,087 Total Unique Species: 1,753 Total Annotated Genes: 0 Total Unique PDB:66

Updated: 2024-03-08

Conserved Genes Families of Unknown Function: 202

## Gene	s Expressed Genes %	Genes with Phenotypes	Unique Species	Proteins with PFAM Domains	PFAM Domains Count	Protein PFAM Domains	Uniprot HMM Hint	PDB HMM Hint	AlphaFold pLDDT	Foldseek PDB Hint	Conserved In	User Curated Models	Avg. Protein Length
1 1,64	1 84	1	461	0	0				<u>78</u>		Fungi	0	244
2 1,34	<u>8</u> 50	1	726	2	2	• 2OG-Fe dioxygenase:1 • 2OG-Fe(II) oxygenase superfamily:1	A0A093V829_TALMA	<u>6NIE</u>	88		Universal	0	356

Myco-Ed Genomics



Mycological Curriculum for Education and Discovery

- A Course-Based Undergraduate
 Research Experience (CURE) that
 consists of coordinated hands-on
 experiments across fungal biology
 teaching labs with the goals of:
 - Training students and professors in laboratory techniques and data analyses
 - Compiling original data for improving fungal genomic resources
 - Empowering students through isolation and genome sequencing of new fungal species





1) Isolate environmental fungi

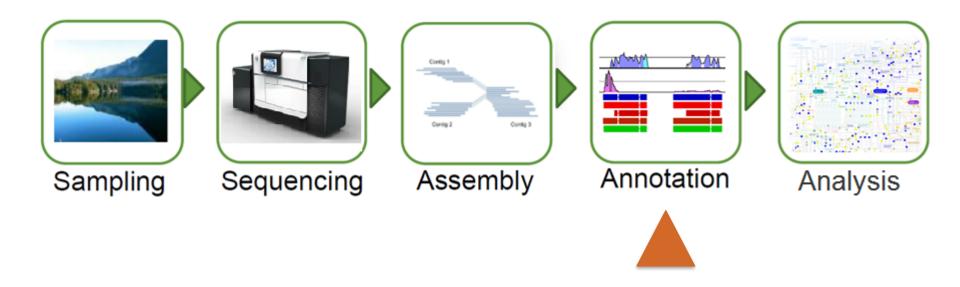




2) Conduct standardized growth experiments

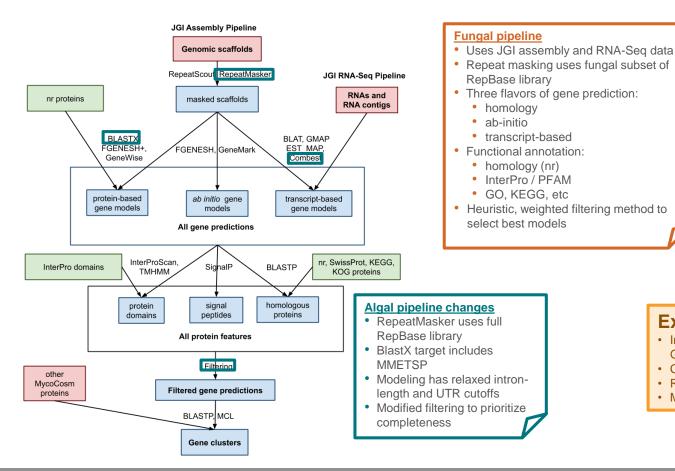
JGI project flow

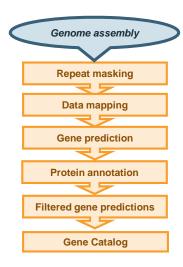




Eukaryotic annotation pipeline







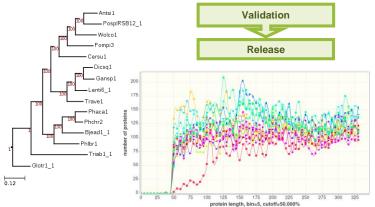
External Genomes

- Import annotations (collaborators, GenBank)
- Can apply filtering
- · Removal of TEs, alleles, isoforms, etc.
- Make comparable with internal genomes

Manual Validation and QC



- Species tree
 - phylogenetic neighborhood
- Genome assembly
 - size, N50/L50, repeats, ploidy
- Gene model sources
 - different algorithms contribute models
- Protein length distribution
 - short models removed
- Gene model lengths & structure
 - introns, exons, UTRs
- Gene model support
 - blast, PFAM, RNAseq
- Completeness
 - CEGMA, BUSCO



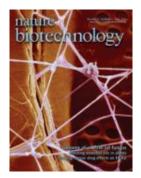
Gene Models Length and Structure

Species Tree for clustering run 2782

		Phchr2	Triab1_1	Phibr1	Gansp1	Bjead1_1	Glotr1_1	Phaca1
		FilteredModels1	FilteredModels1	FilteredModels1	FilteredModels1	FilteredModels2	FilteredModels1	FilteredModels4
# genes	4—°	13,602	14,978	16,170	12,910	15,473	11,846	13,937
Protein length (median)	<u></u>	332	333	329	355	334	347	313
Exon Length Median	<u></u>	157	147	140	148	153	146	159
Gene length (median)	<u></u>	1,416	1,461	1,347	1,541	1,424	1,500	1,448
Transcript length (median)	<u></u>	1,184	1,159	1,058	1,182	1,168	1,191	1,190
Intron length (median)	<u></u>	54	59	55	60	53	55	55
#spliced genes	<u></u>	12,066	13,359	14,453	11,812	13,958	10,764	11,242
% spliced		88.71	89.19	89.38	91.49	90.21	90.87	80.66
Introns per spliced gene (median)	<u></u>	4	4	4	4	4	4	4
Exons per gene (median)	#-•	4	4	4	5	4	5	4
Gene Density (Mbp)	њ.:	386.98	NO DATA	323.63	326.65	362.11	318.60	301.06

Fungal-Algal Genomics Parallels

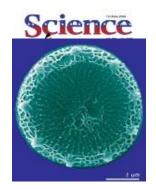




1st fungal genome

2004

1st algal genome





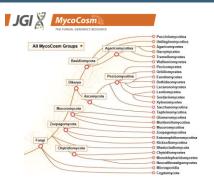
~50 fungal genomes

~10 algal genomes

To angular government

2009

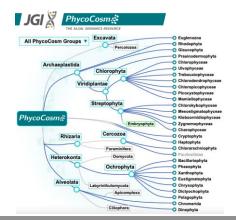




2,500+ fungal genomes

2022

150+ algal genomes



- pan-genomes
- multi-omes
- modeling
- synthesis

10,000 Fungal genomes

2030

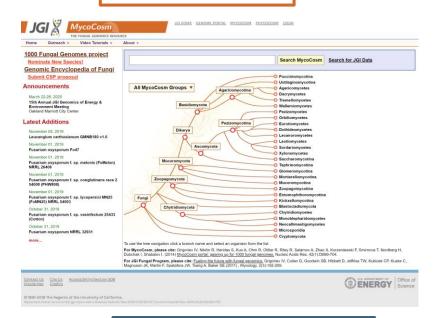
1,000+ algal genomes



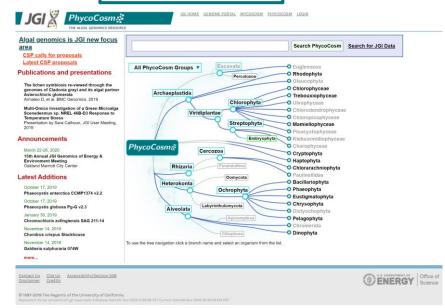
JGI's Fungal & Algal Genomics Resources



mycocosm.jgi.doe.gov



phycocosm.jgi.doe.gov



Genome portals provide:

- data access
- visualization
- analysis tools

Allow researchers to:

- Explore gene models within species
- Compare gene content between species

Connect with us!



Joint Genome Institute



https://twitter.com/jgi http://www.youtube.com/user/JointGenomeInstitute https://www.linkedin.com/company/joint-genome-institute







https://jgi.doe.gov/category/podcasts/







