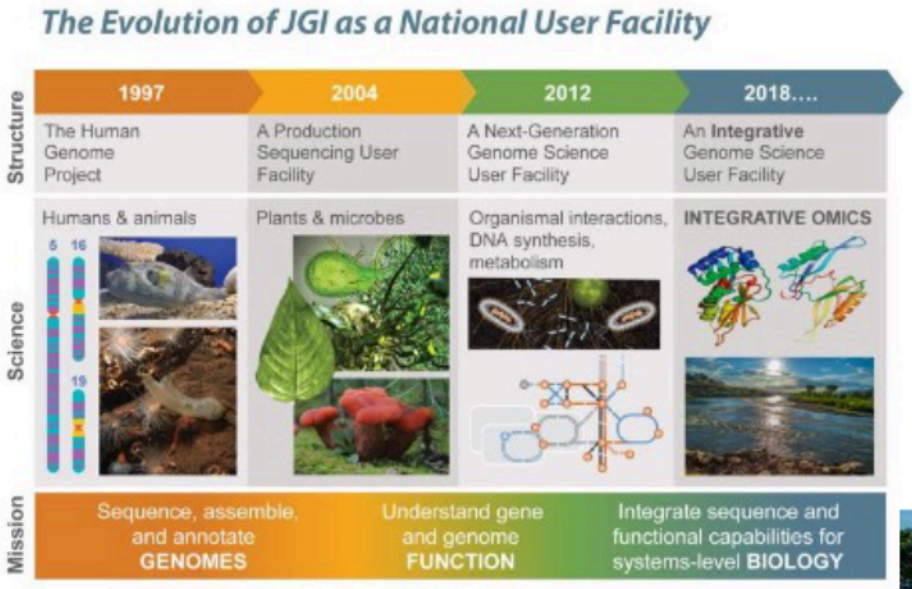
JGI JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY

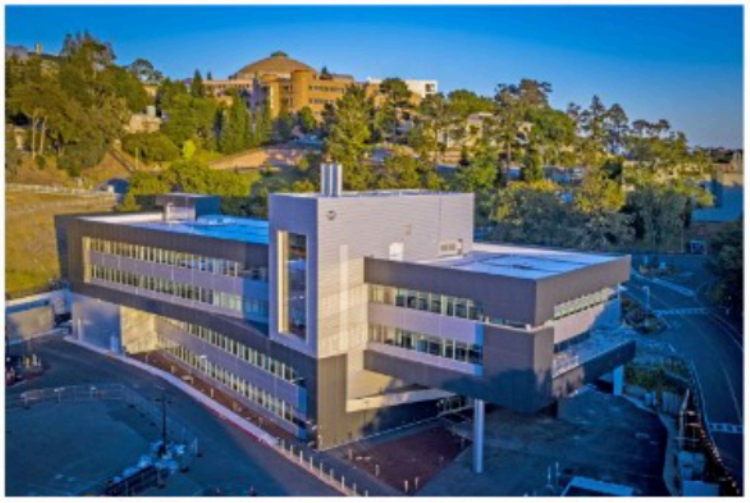
Introduction to JGI & MycoCosm

Richard D. Hayes, PhD
Data Scientist
Fungal & Algal Genomics Program

Joint Genome Institute

- 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

Igor Shabalov
ishabalov@lbl.gov
Lead Software developer

- Software developers
- Data scientists
- Postdoctoral scholars

mycocosm.jgi.doe.gov

JGI JOINT GENOME INSTITUTE

JGI MycoCosm
THE FUNGAL GENOMICS RESOURCE

JGI HOME | GENOME PORTAL | MYCOCOSM | PHYCOCOSM | LOGIN

Home | JGI Data Policy | Outreach | Video Tutorials | About

Genes of Unknown Function
Claim to characterize!
1000 Fungal Genomes project
Nominate New Species!
Genomic Encyclopedia of Fungi
Submit CSP proposal
Myco-Ed Genomics Initiative

Announcements

Jun 2-6, 2025
Fungal Pathogen Genomics Virtual Course: Fungal Pathogen Genomics

Jun 28-Jul 2, 2025
Mycological Society of America meeting
Madison, WI

Latest Additions

May 29, 2025
Russula ochroleuca BavarianF39 v1.0

May 29, 2025
Amanita aff. grandis KIS10 v1.0

May 29, 2025
Russula cyanoxantha BavarianF04 v1.0

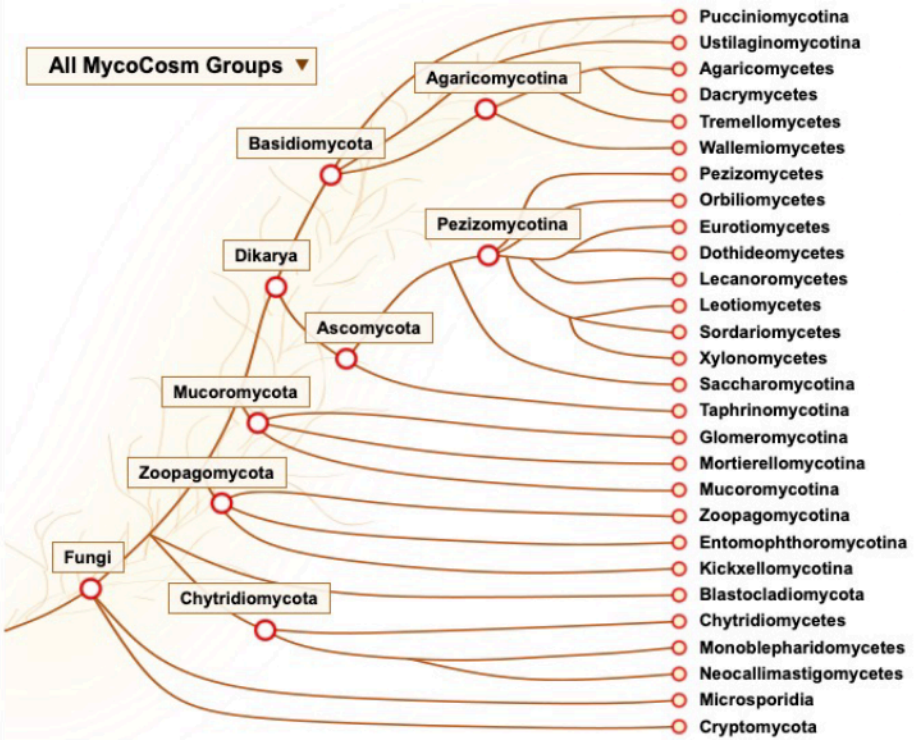
May 29, 2025
Amanita rubescens BavarianF08 v1.0

May 29, 2025
Elaphomyces granulatus Maridalen v1.0

May 01, 2025
Collariella robusta PSN 660 v1.0

more...

Search MycoCosm | Search for JGI Data



To use the tree navigation click a branch name and select an organism from the list.

For MycoCosm, please cite: Igor V. Grigoriev, Roman Nikitin, Sajeet Haridas, Alan Kuo, Robin Ohm, Robert Otillar, Robert Riley, Asaf Salamov, Xueling Zhao, Frank Korzeniewski, Tatyana Smirnova, Henrik Nordberg, Inna Dubchak, Igor Shabalov, MycoCosm portal: gearing up for 1000 fungal genomes, Nucleic Acids Research, Volume 42, Issue D1, 1 January 2014, Pages D699–D704, DOI:10.1093/nar/gkt1183

For JGI Fungal Program, please cite: Igor V. Grigoriev, Daniel Cullen, Stephen B. Goodwin, David Hibbett, Thomas W. Jeffries, Christian P. Kubicek, Cheryl Kuske, Jon K. Magnuson, Francis Martin, Joseph W. Spatafora, Adrian Tsang & Scott E. Baker (2011) Fueling the future with fungal genomics, Mycology, 2:3, 192-209, DOI:10.1080/21501203.2011.584577

Contact Us | Cite Us | Credits
JGI Data Policy | Disclaimer | Accessibility/Section 508

U.S. DEPARTMENT OF ENERGY | Office of Science

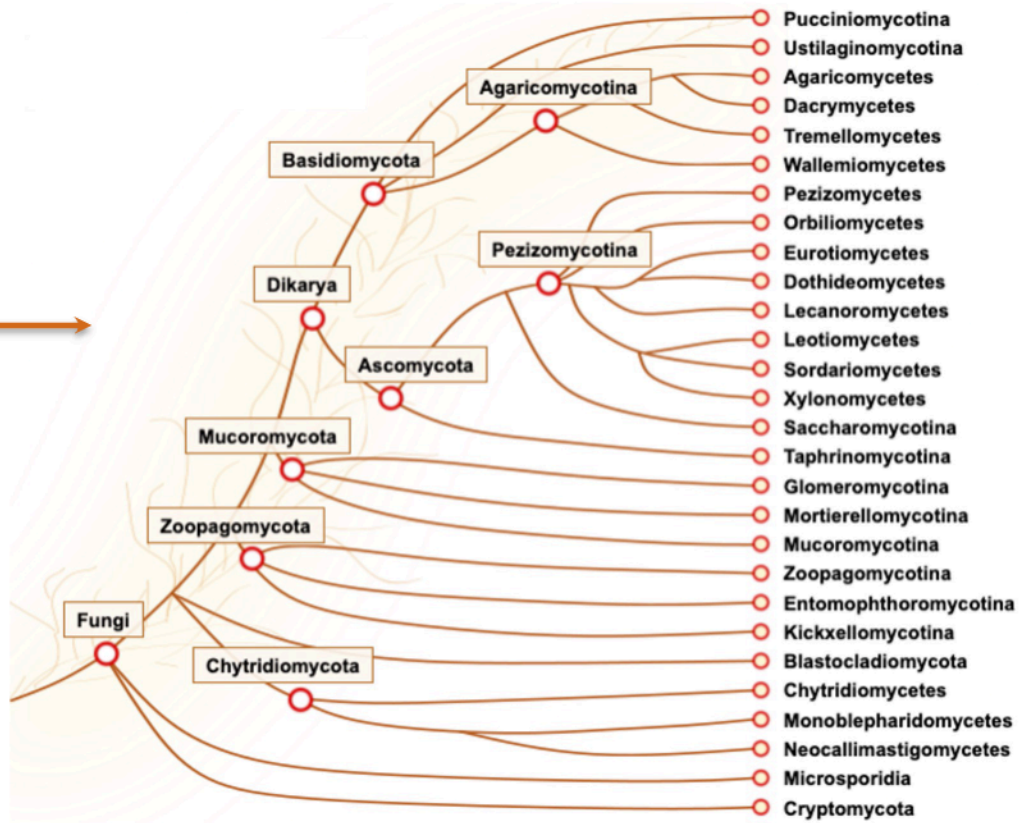
JGI Fungal Genomics – A growing field

2004:
1st basidiomycete genome published

2025:
2700+ genomes in MycoCosm

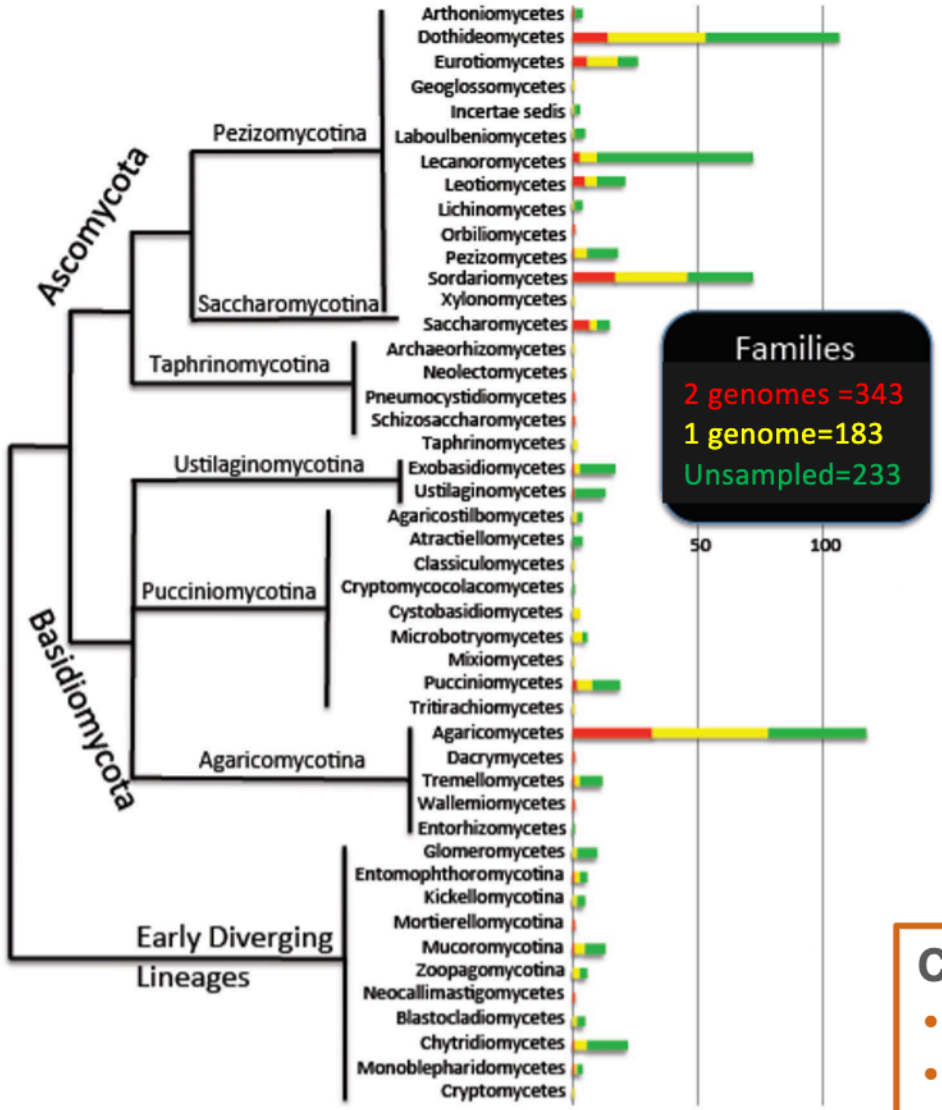
JGI MycoCosm
THE FUNGAL GENOMICS RESOURCE

White rot fungus
Phanerochaete chrysosporium



1000 Fungal Genomes Project

GOAL: Genome references for every family of fungi



Open to community

- 90+ participants
- 1000+ nominations
- 700+ sequenced

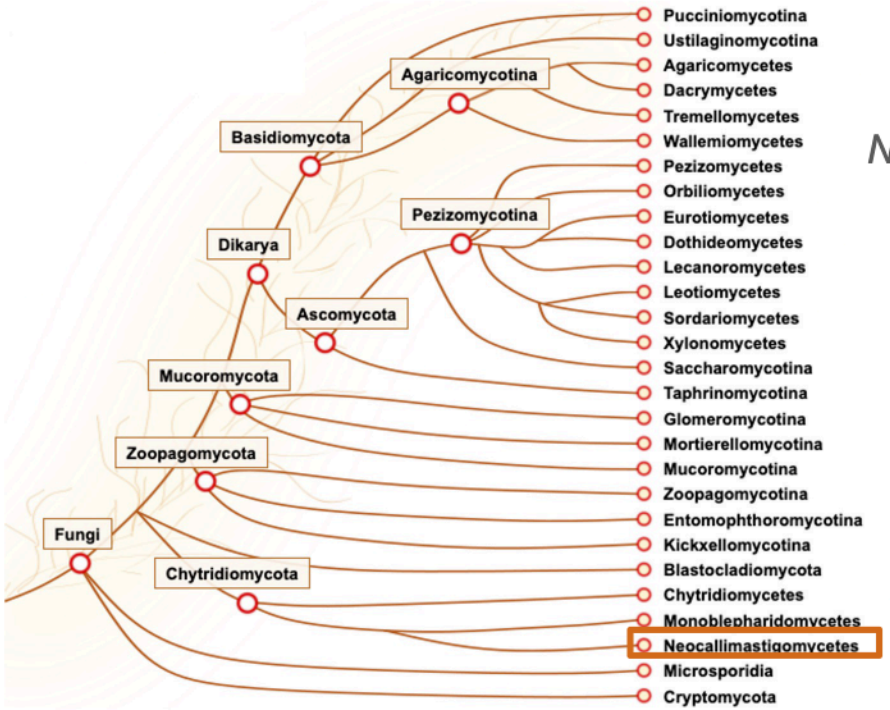
mycocosm.jgi.doe.gov/pages/fungi-1000-projects.jsf

Family name	Genome Projects			
Incertae sedis	Ascomycete sp.	WGS		
	Calcarisporiella thermophila	WGS		
	Cerataphis brasiliensis yeast-like symbiont	WGS		
	Helminthosporium solani	WGS		
	Nilaparvata lugens yeast-like symbiont	WGS		
	Sclerotium cepivorum	WGS		
	* Stanjemonium griseum	WGS	Kathryn Bushley, University of Minnesota	
	Symbiotaphrina kochii	WGS	Joseph Spatafora, Oregon State University	
	Taxomyces andreanae	WGS		
				Nominate / Add to GOLD
Arthoniaceae	Arthonia rubrocincta	WGS		
				Nominate / Add to GOLD
Chrysothricaceae	None sequenced			Nominate / Add to GOLD
Roccellaceae	None sequenced			Nominate / Add to GOLD
Melaspileaceae	None sequenced			Nominate / Add to GOLD

CSP Proposals

- Joey Spatafora (Oregon State U)
- Jason Stajich (UC Riverside)
- Francis Martin (INRA-Nancy)

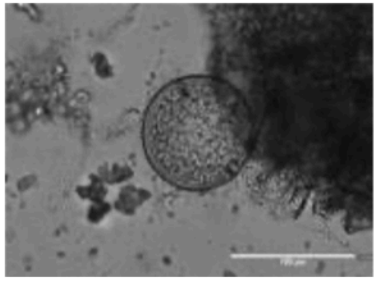
Neocallimastigomycetes (Gut Fungi) – powerful lignocellulose degraders



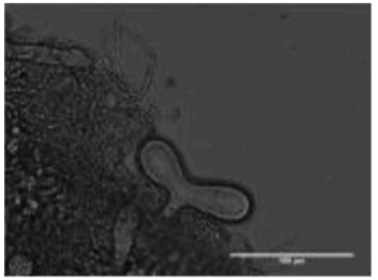
Genomic challenges

- AT rich: 17% GC
- Large: up to 190 Mb

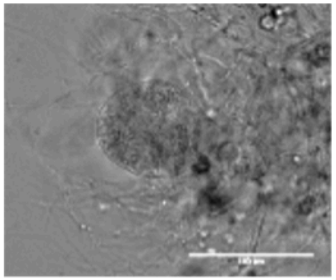
Neocallimastix californiae



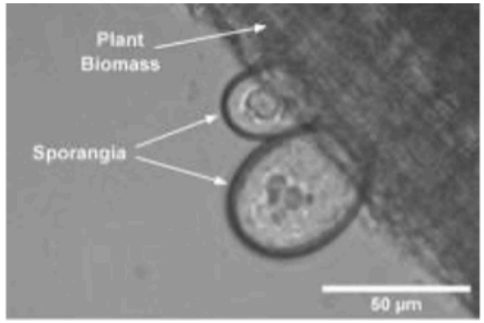
Piromyces finnis



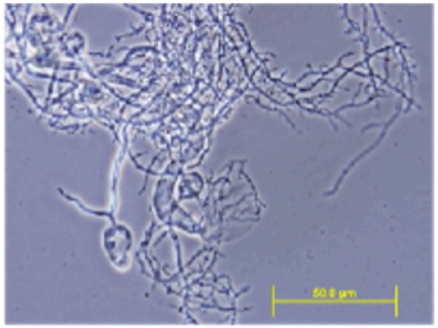
Anaeromyces robustus



Caecomyces churrovis



Orpinomyces sp.

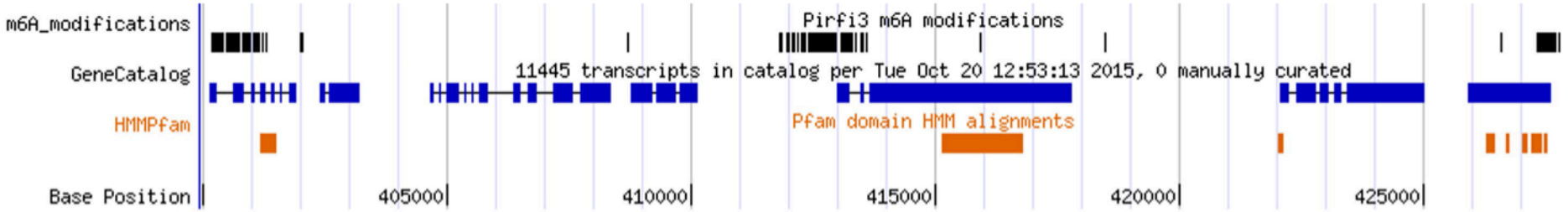


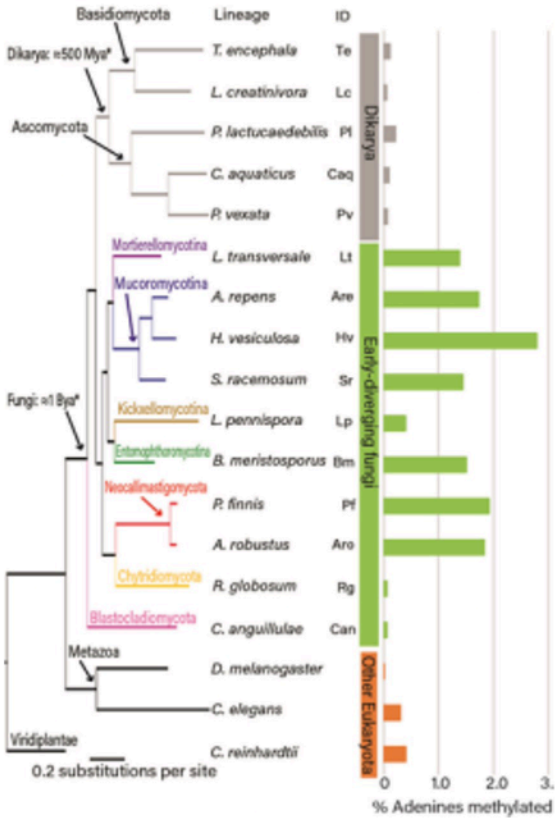
Solution

Illumina **PacBio**

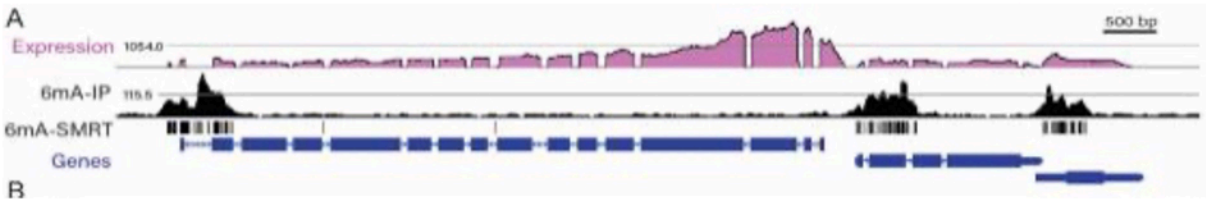
PI: Michelle O'Malley
University of California, Santa Barbara

Bonus data with PacBio: 6mA methylation in early fungi





- High 6mA levels in **early-diverging fungi**
- 6mA concentrates in **'islands'** at gene promoters
- Positive impact on **gene expression**

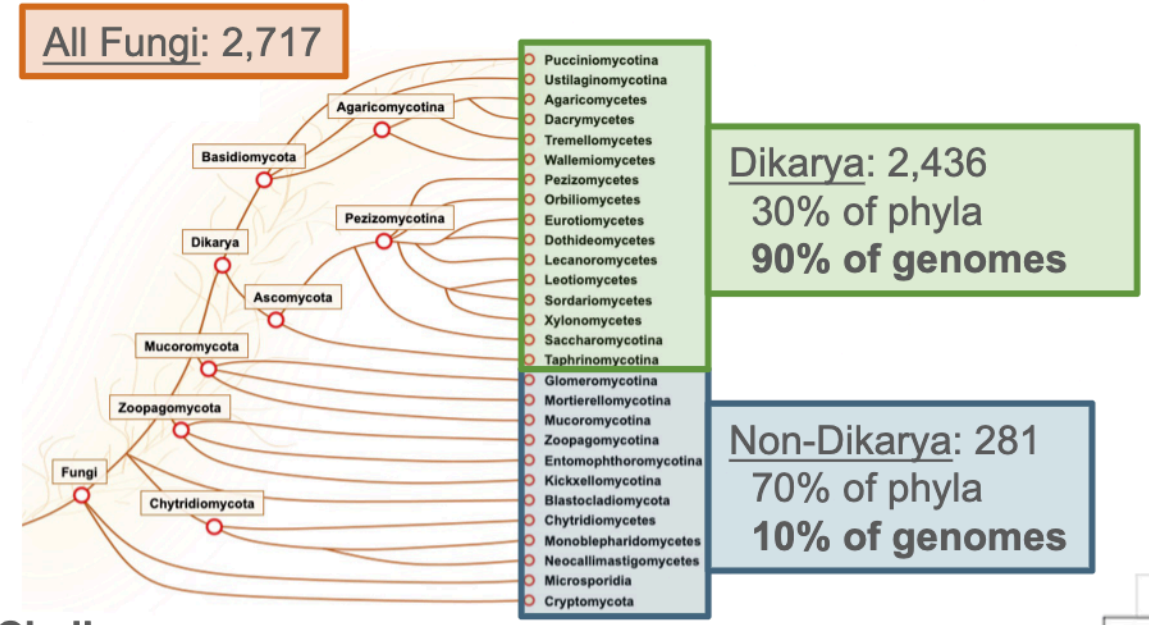


Stephen Mondo
sjmondo@lbl.gov

Mondo et al., 2017, *Nat. Genet.*

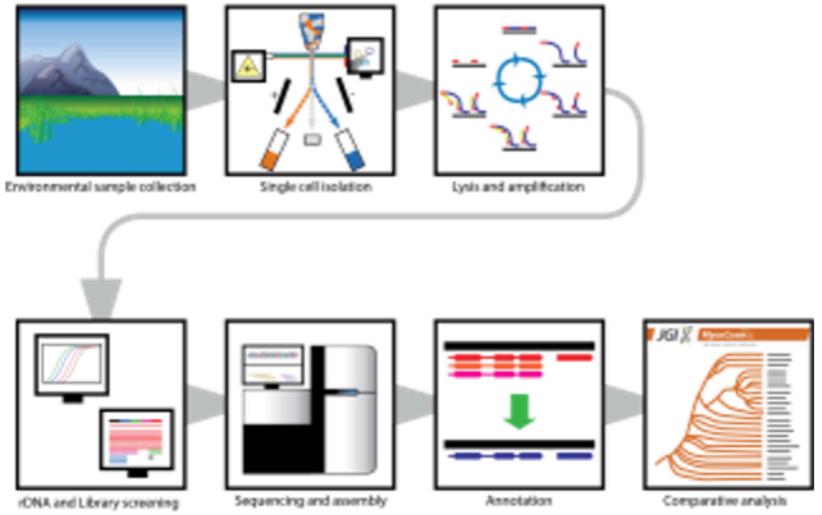
Meredith Blackwell (LSU)
Richard Dannenbaum (JGI)

Single cell genomics can address bias in Fungal genome sequencing



Dikarya: 2,436
30% of phyla
90% of genomes

Non-Dikarya: 281
70% of phyla
10% of genomes

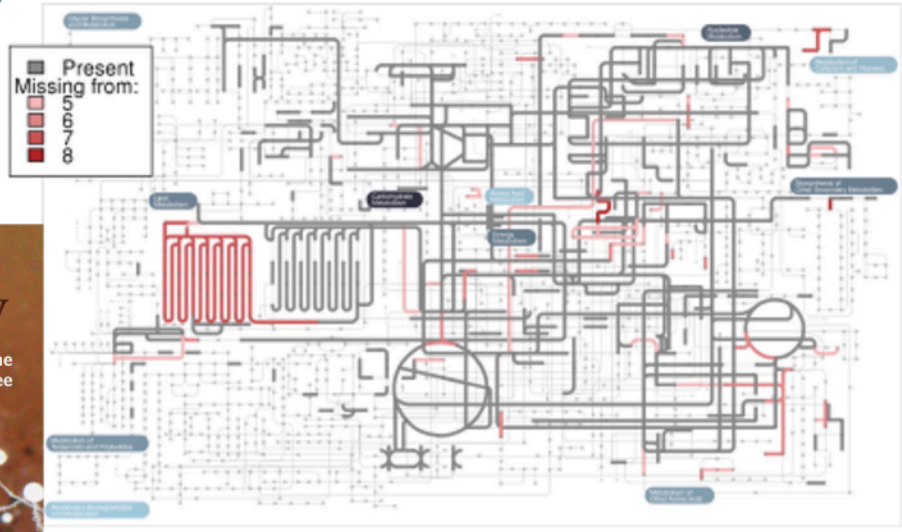


Challenges

- Early lineages not well sequenced due to cryptic biology
- Single-cell workflows for fungi need to be developed

Results

- First Zoopagomycotina genomes
- Single-cell genomes can resolve phylogeny
- Conserved metabolic losses reflect similar deficiencies among uncultivated fungi
- Novel proteases and secondary metabolism proteins in early lineages

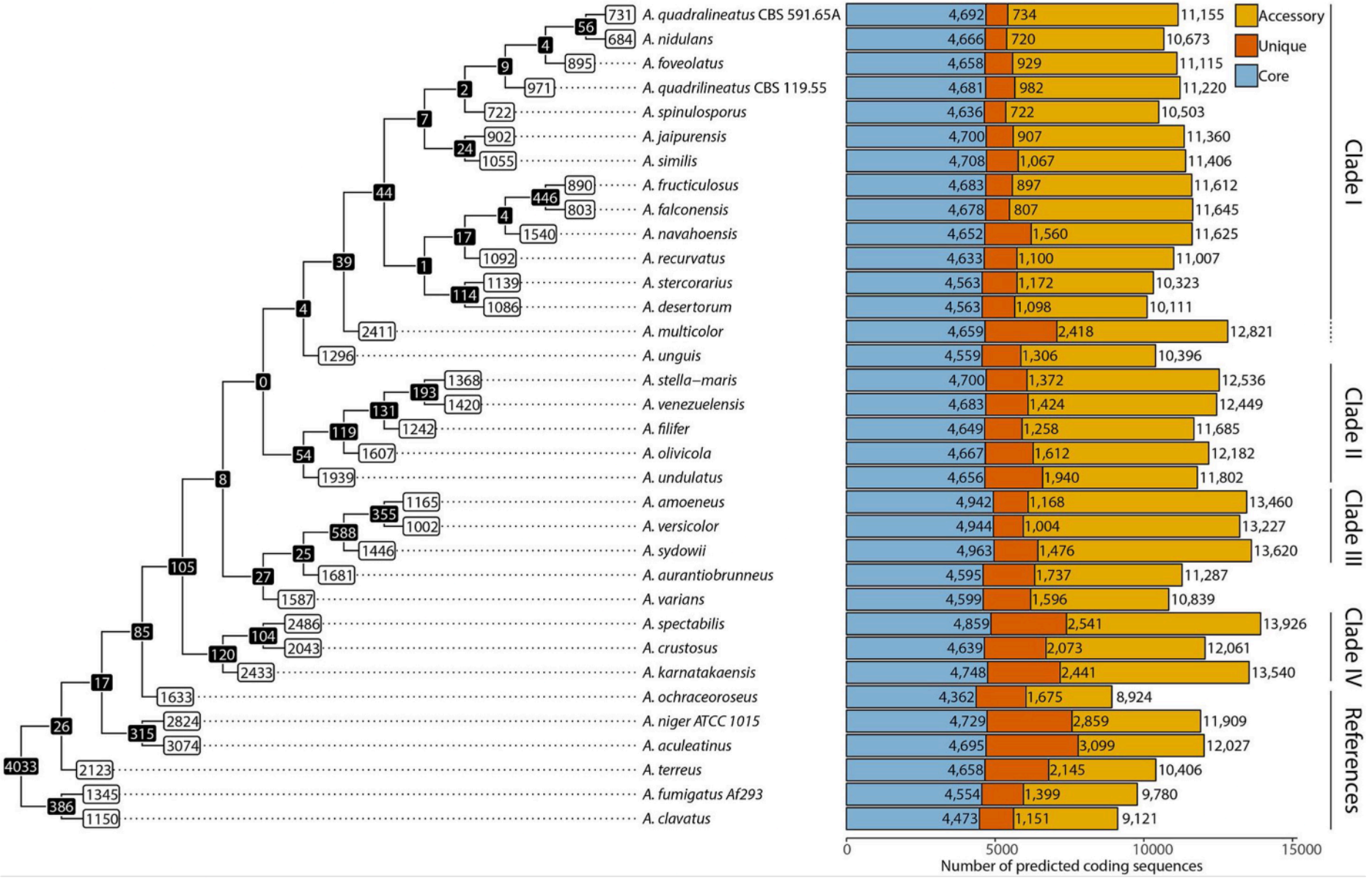


Steven Ahrendt
sahrendt0@lbl.gov

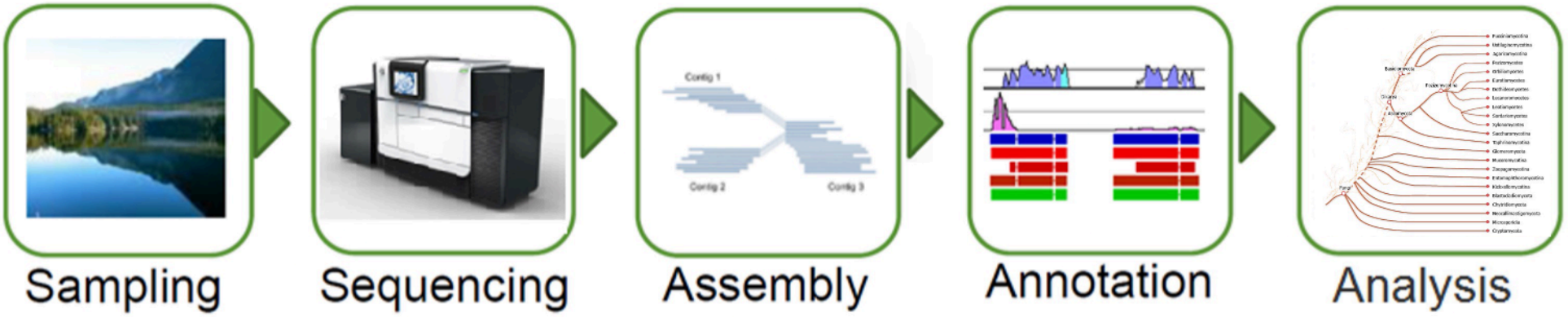
PI: Tim James
University of Michigan

Ahrendt et al., 2018, Nat. Microbio.

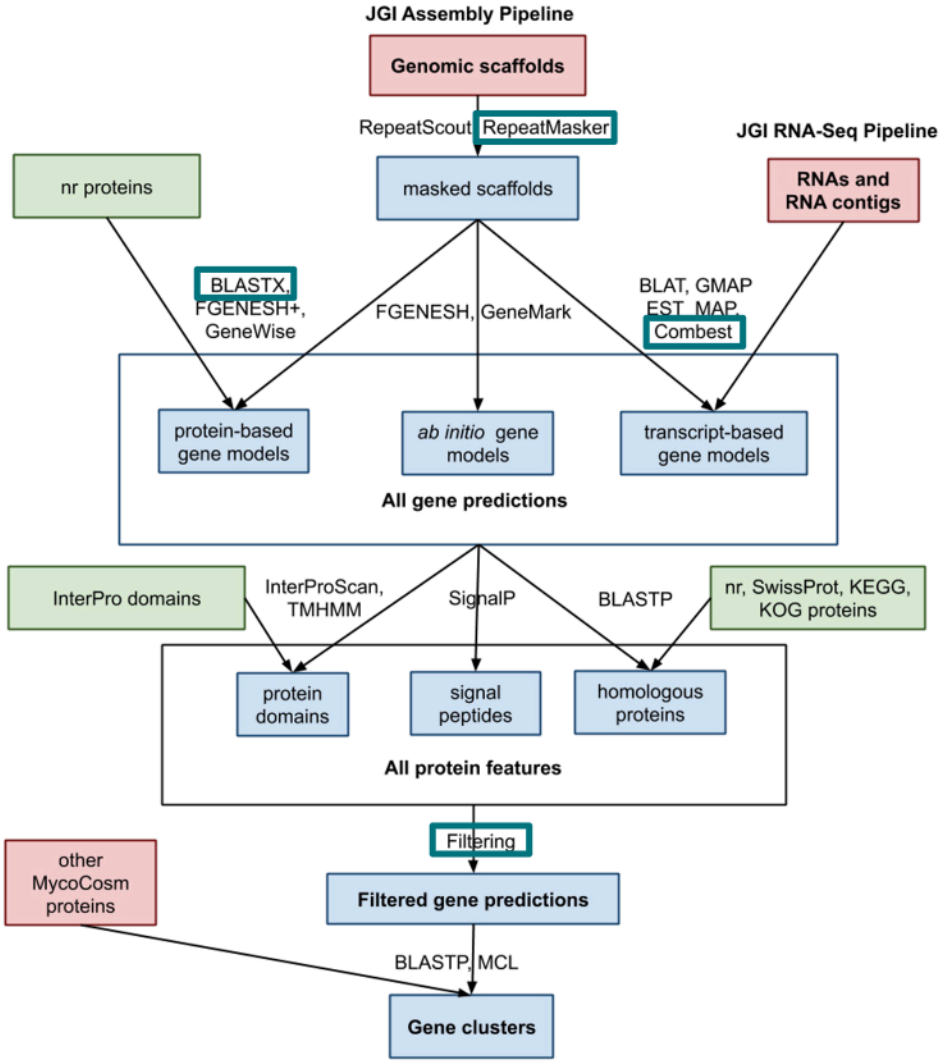
Whole genus sequencing for pan-genomics



JGI project flow



Eukaryotic annotation pipeline



Fungal pipeline

- Uses JGI assembly and RNA-Seq data
- Repeat masking uses fungal subset of RepBase library
- Three flavors of gene prediction:
  - homology
  - ab-initio
  - transcript-based
- Functional annotation:
  - homology (nr)
  - InterPro / PFAM
  - GO, KEGG, etc
- Heuristic, weighted filtering method to select best models

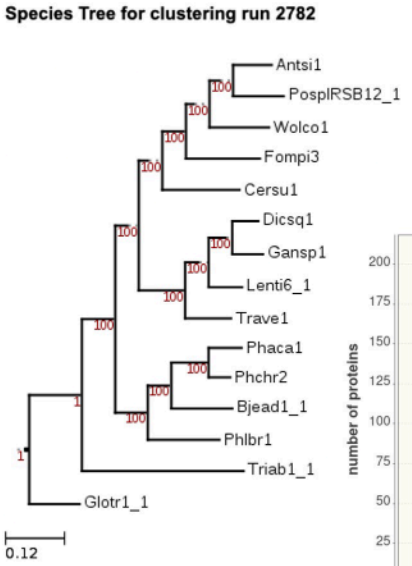
Repeat Masking
Data Mapping
Gene Prediction
Annotation

Algal pipeline changes

- RepeatMasker uses full RepBase library
- BlastX target includes MMETSP
- Modeling has relaxed intron-length and UTR cutoffs
- Modified filtering to prioritize completeness

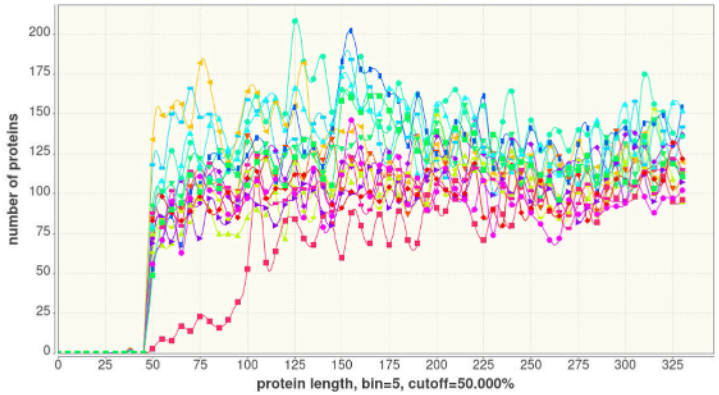
Manual Validation and QC

- **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



Validation

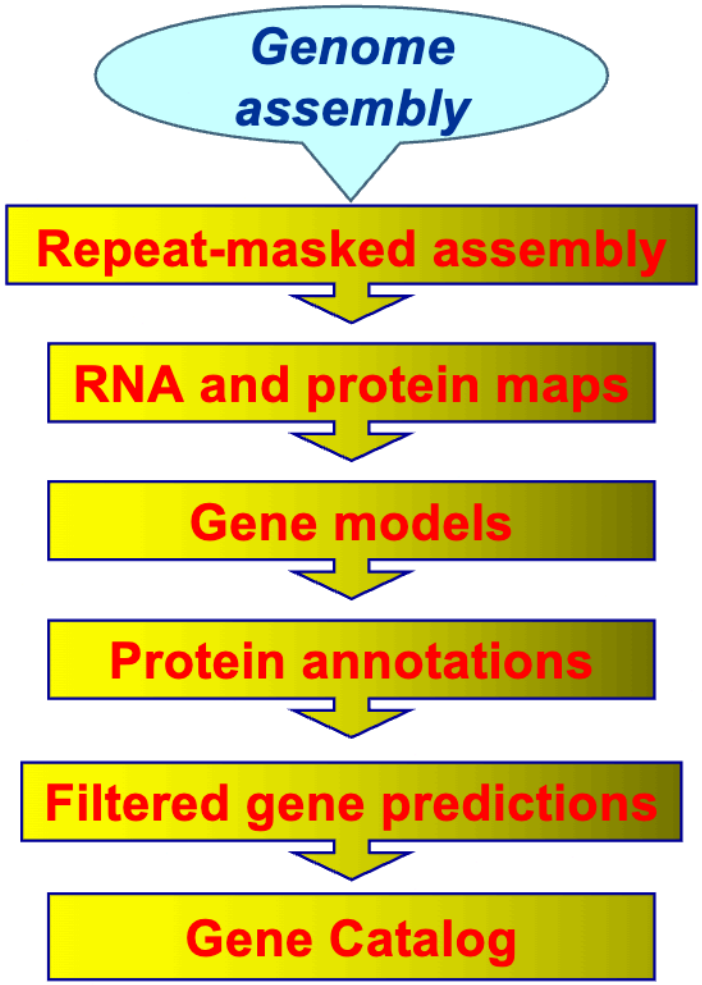
Release



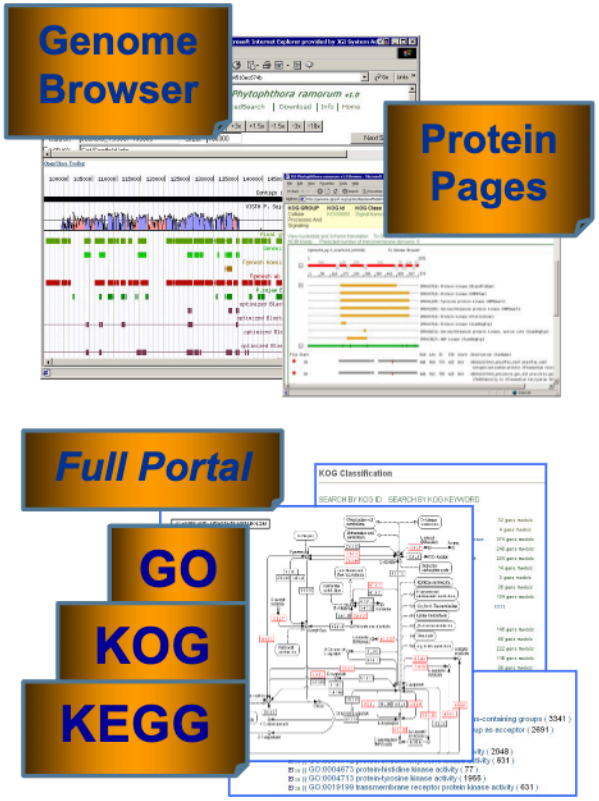
Gene Models Length and Structure

	Phchr2	Triab1_1	Phlbr1	Gansp1	Bjead1_1	Glotr1_1	Phaca1
	FilteredModels1	FilteredModels1	FilteredModels1	FilteredModels1	FilteredModels2	FilteredModels1	FilteredModels4
# genes	13,602	14,978	16,170	12,910	15,473	11,846	13,937
Protein length (median)	332	333	329	355	334	347	313
Exon Length Median	157	147	140	148	153	146	159
Gene length (median)	1,416	1,461	1,347	1,541	1,424	1,500	1,448
Transcript length (median)	1,184	1,159	1,058	1,182	1,168	1,191	1,190
Intron length (median)	54	59	55	60	53	55	55
# spliced genes	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)	4	4	4	4	4	4	4
Exons per gene (median)	4	4	4	5	4	5	4
Gene Density (Mbp)	386.98	NO DATA	323.63	326.65	362.11	318.60	301.06

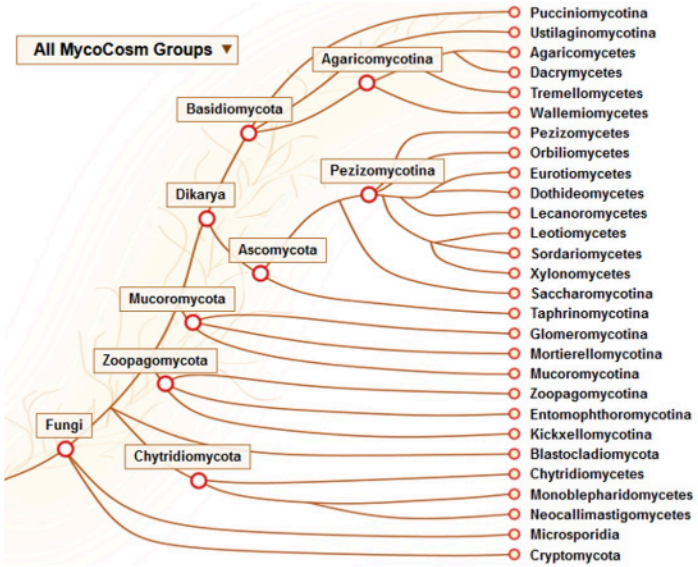
Annotation pipeline and portal



Annotation Pipeline



Genome Portal



MycoCosm

External genomes in MycoCosm

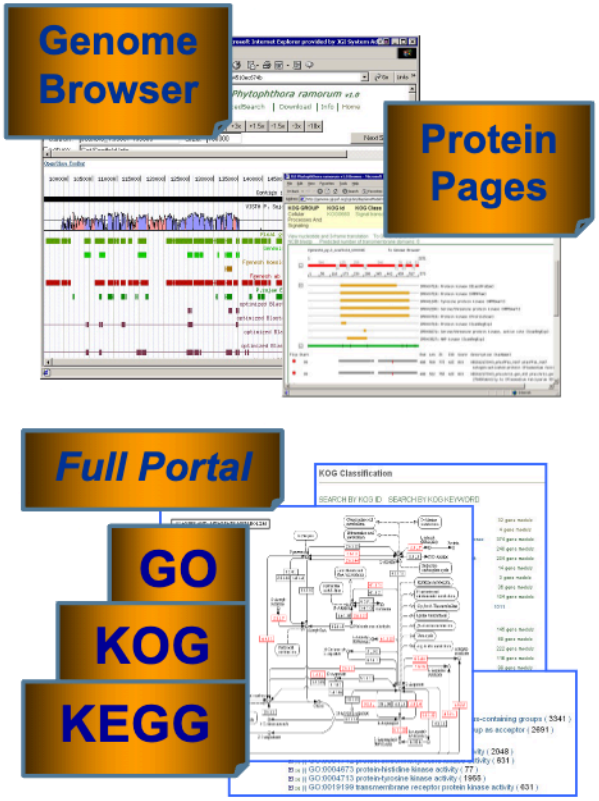
External genome assembly and annotation

Protein annotations

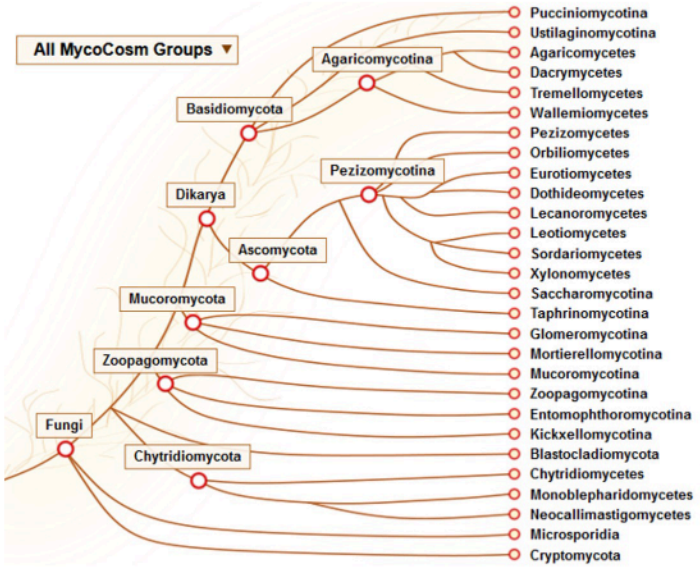
Filtered gene predictions*

Gene Catalog

Annotation Pipeline



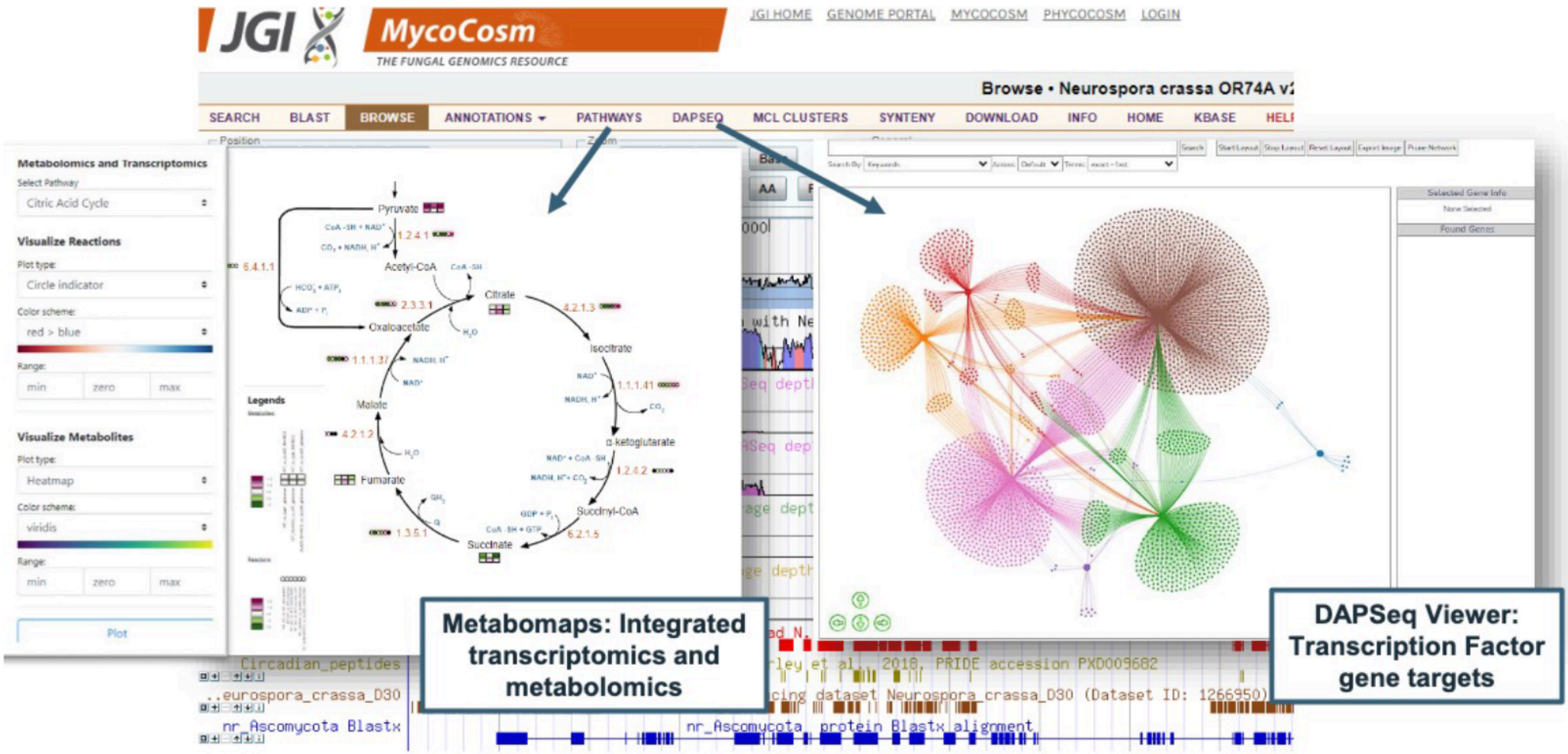
Genome Portal



MycoCosm

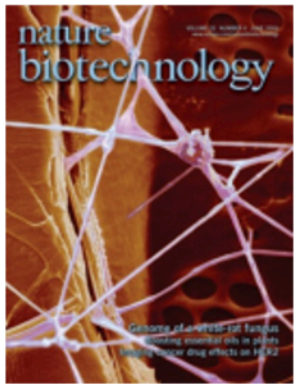
New tools on the horizon

Neurospora crassa OR74A v2.0 is our multi-omics showcase genome portal



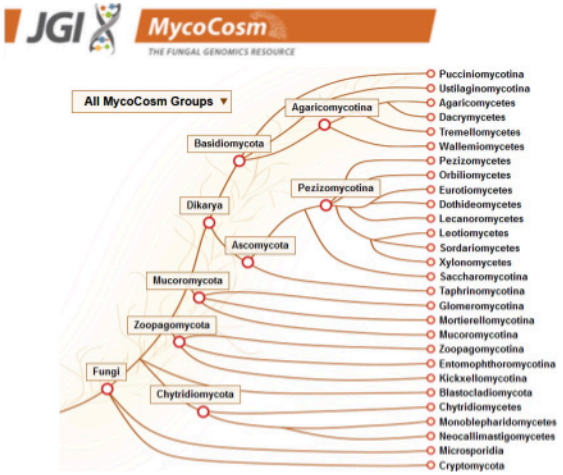
https://mycocosm.jgi.doe.gov/Neucr2/Neucr2.home.html

Fungal-Algal Genomics Parallels



1st fungal genome

~50 fungal genomes



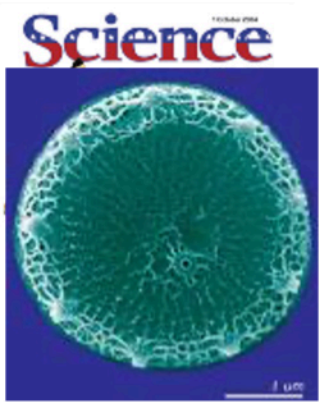
2700+ fungal genomes

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

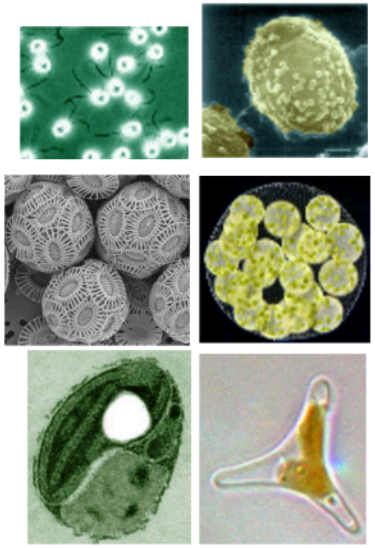
10000 fungal genomes

2004 **2009** **2025** **2030**

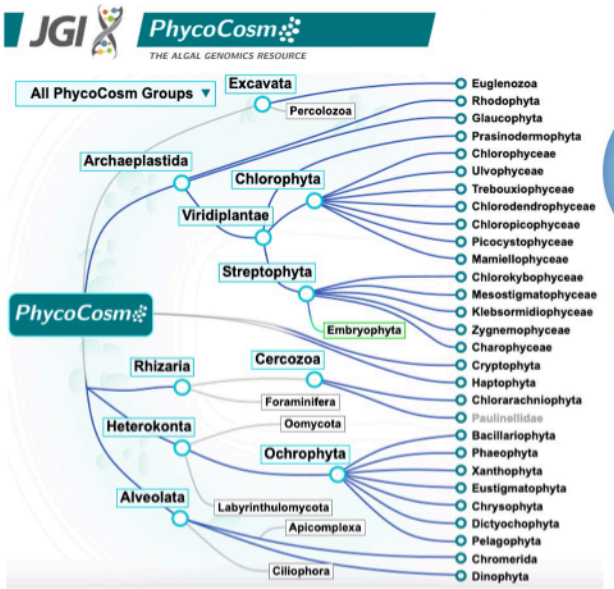
1st algal genome



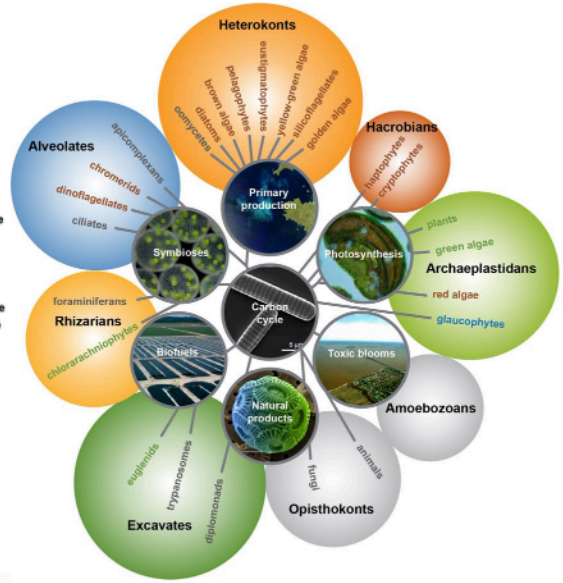
~10 algal genomes



250+ algal genomes



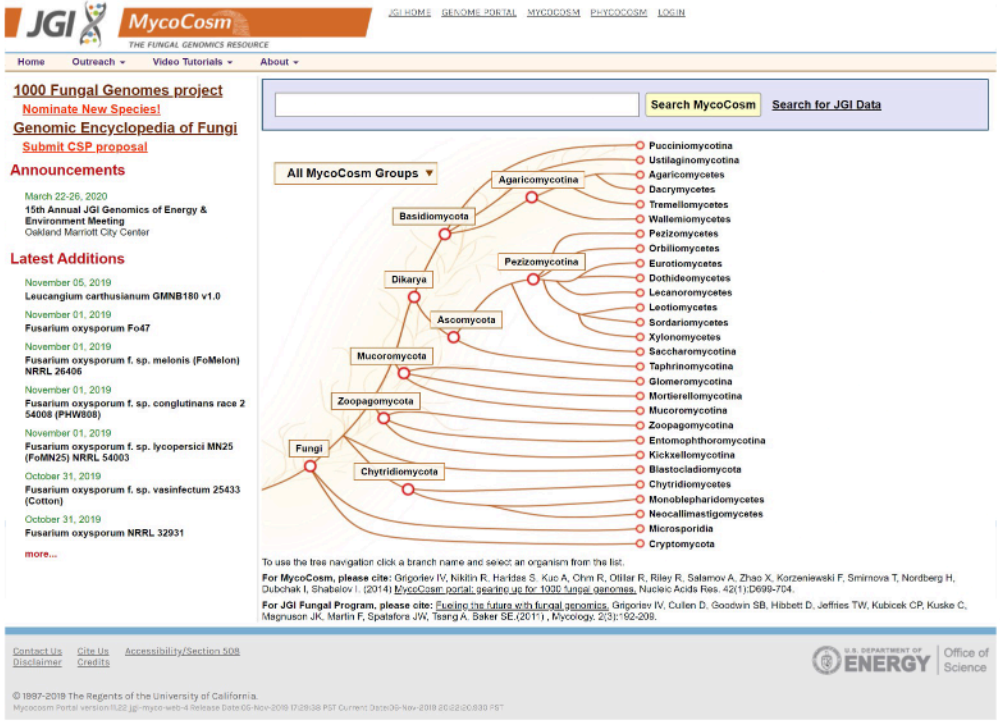
1000 algal genomes?

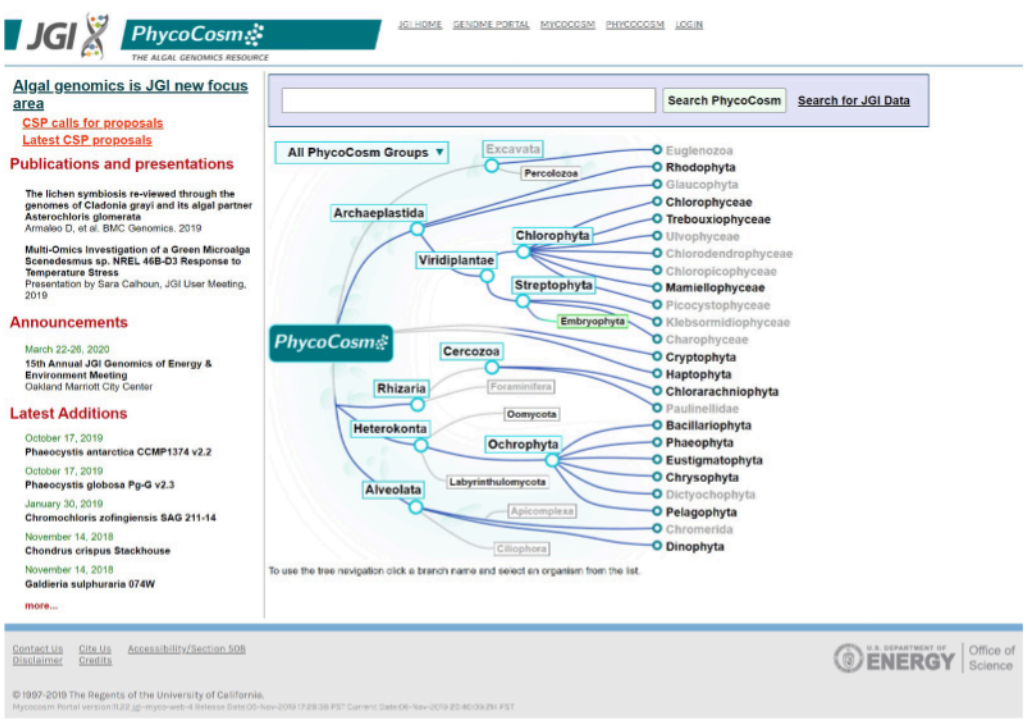


JGI's Fungal & Algal Genomics Resources

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

JGI Calls for Proposals

Integrative Genomics Building
Berkeley, CA, USA

https://jgi.doe.gov/work-with-us/proposals

Returning Spring 2026

Community Science Program (CSP) annual call

- Large-scale proposals in bioenergy and biogeochemistry (up to 10Tbp)
- Sequencing, synthesis, metabolomics

Facilities Integrating Collaborations for User Science (FICUS) Joint JGI-Environmental Molecular Sciences Laboratory (EMSL) annual call:

- EMSL imaging and mass spec plus JGI capabilities (up to 3Tbp)

Ongoing

CSP Functional Genomics biannual call

- Large scale synthesis proposals (up to 1.5Mbp)
- Proposals due ***January 29, 2026***

https://jgi.doe.gov/user-programs/program-info/csp-overview/csp-functional-genomics/

•CSP New Investigator biannual call

- Investigators who have never before worked with JGI.
- Small scale proposals (up to 3Tbp)
- Proposals due ***September 9, 2025***

https://jgi.doe.gov/user-programs/program-info/csp-overview/csp-new-investigator/