

# Introduction to MycoCosm

MycoCosm is a fungal genomics portal developed by the US Department of Energy Joint Genome Institute (JGI) to support integration, analysis, and dissemination of fungal genome sequences and related omics data by providing interactive web-based tools. To get started, navigate to the MycoCosm Homepage in your web browser (<https://mycocosm.jgi.doe.gov/>)

**MycoCosm**  
THE FUNGAL GENOMICS RESOURCE

JGI HOME GENOME PORTAL MYCOCOSM PHYCOCOSM LOGIN **Click to login/register**

Home 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**

March 12-17, 2024  
Fungal Genetics Conference  
Pacific Grove, CA  
June 2-7, 2024  
Fungal Pathogen Genomics Course  
Hinxton, UK  
June 9-13, 2024  
Mycological Society of America Meeting  
Markham, Canada  
September 30-October 4, 2024  
2024 JGI Genomics of Energy and Environment  
Walnut Creek, CA

**Latest Additions**

April 11, 2024  
Paraglomus occultum DAOM 240472  
April 09, 2024  
Rhizophagus irregularis 4401  
April 09, 2024  
Rhizophagus irregularis A1 v2.0  
April 09, 2024  
Rhizophagus irregularis DAOM197198 v2.0  
April 09, 2024  
Rhizophagus irregularis C2  
[more...](#)

**All MycoCosm Groups**

- Agaricomycotina
  - Pucciniomycotina
  - Ustilaginomycotina
  - Agaricomycetes
  - Dacrymycetes
  - Tremellomycetes
  - Walleriomycetes
  - Pezizomycetes
  - Orbiliomycetes
  - Eurotiomycetes
  - Dothideomycetes
  - Lecanoromycetes
  - Leotiomycetes
  - Sordariomycetes
  - Xylonomycetes
- Basidiomycota
  - Mucoromycota
  - Zoopagomycota
  - Chytridiomycota
- Ascomycota
  - Pezizomycotina
  - Mucoromycotina
  - Zoopagomycotina
  - Entomophthoromycotina
  - Kickxellomycotina
  - Blastocladiomycota
  - Chytridiomycetes
  - Monoblepharidomycetes
  - Neocallimastigomycetes
  - Microsporidia
  - Cryptomycota

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 Sminova, 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 Accessibility/Section 508  
Disclaimer Credits

© 1997-2024 The Regents of the University of California.

You can also get to MycoCosm from the unified JGI Genome Portal (<https://genome.jgi.doe.gov>)

**JGI Genome Portal**

JGI HOME GENOME PORTAL LOGIN 10 Items

My Favorites

Search for Genomic Data Search Advanced Search

**Searching for Projects**

- Explore what you can do here.
- Search projects/proposals using "Advanced Search" filters.

**Downloading Files**

- Download over the web
- Download large number of files with Globus service
- Download via API using scripting or programming
- Download with "Cart" by collecting projects/portals of your interest.

**Looking for Access**

- Looking for data and do not have access to the private portal? Please contact PI
- How to grant access to your proposal/project/genome? Get Instructions.

**How to Find Proposal Award DOIs**

The new JGI Publication Policy requires inclusion of Proposal Award DOIs within the DOE auspice statement to improve our manuscript and impact tracking efforts.

Proposal DOIs can be found at the top of the proposal info page on the JGI Genome Portal. Click here to learn how to find DOIs for the datasets you are interested in.

**Download with Cart**

A convenient way to collect projects/genomes/metagenomes of your interest and download all files associated with them in bulk. Read more and provide your comments and suggestions for this feature to our team.

**The "Tree of Life"**

Please use our powerful search or go to the "Tree of Life" if it is the most convenient way for you to reach your genomes/projects.

**Genomes OnLine Database (GOLD)** - a resource for sequencing projects and associated metadata.

**Integrated Microbial Genomes (IMG) and Metagenomes (IMG/M)** - resources for comparative analysis and annotation.

**MycoCosm & Phycosm** - access to data, visualization, and analysis tools for comparative genomics of fungi or algae.

**Phytozome** - a hub for accessing, visualizing and analyzing plant genomes.

Contact Us Cite Us Accessibility/Section 508  
Disclaimer Credits

© 1997-2024 The Regents of the University of California.

If you have an account, login now. If not, create an account by clicking 'LOGIN' and then 'REGISTER'.

## Sign In

JGI — Sign In [Help](#)

Email (or [User Name](#)):

[Forgot your username?](#)

Password:

[Forgot your password?](#)

SIGN IN

SIGN IN WITH GOOGLE (LBL/JGI users)

REGISTER

Currently MycoCosm hosts 2000+ genomes, including those sequenced by JGI in collaboration with the community, as well as those imported from the public domain or privately from collaborators. Most JGI-sequenced genomes are products of the Community Science Program (CSP) or the 1000 Fungal Genomes Project (1KFG). The MycoCosm page has links to these as well as other Programs and Projects. Nominations for genomes to 1KFG may be directly submitted from here!

**JGI**  
THE FUNGAL GENOMICS RESOURCE

**MycoCosm**

[JGI HOME](#) [GENOME PORTAL](#) [MYCOCOSM](#) [PHYCOCOSM](#) [LOGIN](#)

[Home](#) [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**  
March 12-17, 2024  
Fungal Genetics Conference  
Pacific Grove, CA  
June 2-7, 2024  
Fungal Pathogen Genomics Course  
Hinxton, UK  
June 9-13, 2024  
Mycological Society of America Meeting  
Markham, Canada  
September 30-October 4, 2024  
2024 JGI Genomics of Energy and Environment  
Walnut Creek, CA  
**Latest Additions**  
April 11, 2024  
Paraglomus occultum DAOM 240472  
April 09, 2024  
Rhizophagus irregularis 4401  
April 09, 2024  
Rhizophagus irregularis A1 v2.0  
April 09, 2024  
Rhizophagus irregularis DAOM197198 v2.0  
April 09, 2024  
Rhizophagus irregularis C2  
[more...](#)

**All MycoCosm Groups**

**Fungi** (2,554 genomes)

[Tree](#)  
[Search](#)  
[BLAST](#)  
[PFAM Domains](#)  
[Secondary Metabolism Clusters](#)  
[CAZymes](#)  
[Peptidases](#)  
[Transporters](#)  
[Transcription Factors](#)  
[MCL Clusters](#)  
[Geo Mapping](#)  
[Download](#)  
[Nominate new Species](#)

**Fungi**

To use the tree

For MycoCosm  
Nordberg, Inna  
DOI:10.1093/mbe/mbz013

For JGI Fungal  
Joseph W. Spatola

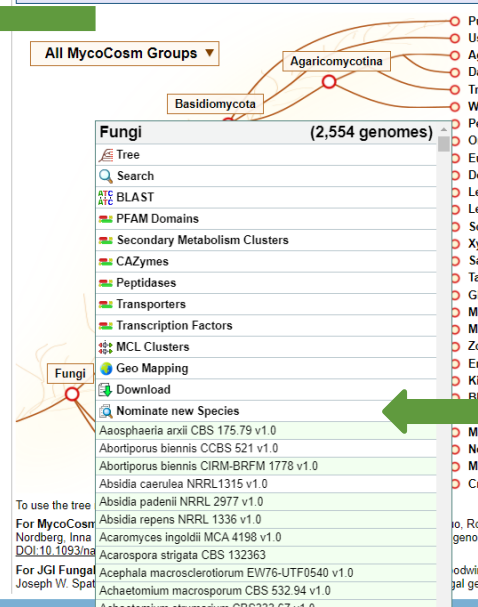
**Agaricomycotina**

- Pucciniomycotina
- Ustilaginomycotina
- Agaricomycetes
- Dacrymycetes
- Tremellomycetes
- Wallemiomycetes
- Pezizomycetes
- Orbiliomycetes
- Eurotiomycetes
- Dothideomycetes
- Lecanoromycetes
- Leotiomycetes
- Sordariomycetes
- Xylonomycetes
- Saccharomycotina
- Taphrinomycotina
- Glomeromycotina
- Mortierellomycotina
- Mucoromycotina
- Zoopagomycotina
- Entomophthoromycotina
- Kickxellomycotina
- Blastocladiomycota
- Blastocladiomycetes
- Monoblepharidomycetes
- Neocallimastigomycetes
- Microsporidia
- Cryptomycota

[Contact Us](#) [Cite Us](#) [Accessibility/Section 508](#)


[Disclaimer](#) [Credits](#)

© 1997-2024 The Regents of the University of California.



io. Robin Ohm, Robert Otillar, Robert Riley, Asaf Salamov, Xueling Zhao, Frank Korzeniewski, Tatyana Smirnova, Henrik  
genomes, Nucleic Acids Research, Volume 42, Issue D1, 1 January 2014, Pages D699–D704,

odwin, David Hibbett, Thomas W. Jeffries, Christian P. Kubicek, Cheryl Kuska, Jon K. Magnuson, Francis Martin,  
gal genomics, Mycology, 2:3, 192-209, DOI:10.1080/21501203.2011.584577



Office of  
Science

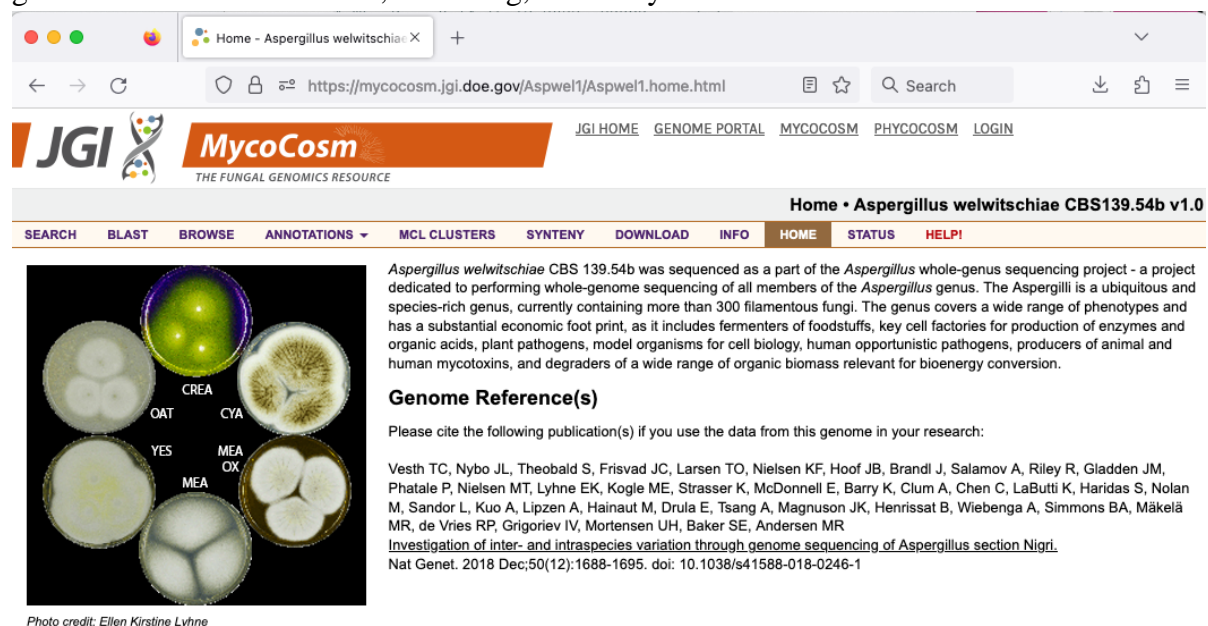
Search MycoCosm

[Search for JGI Data](#)

- Using ‘Search for JGI Data’ will allow searching the Genome Portal.

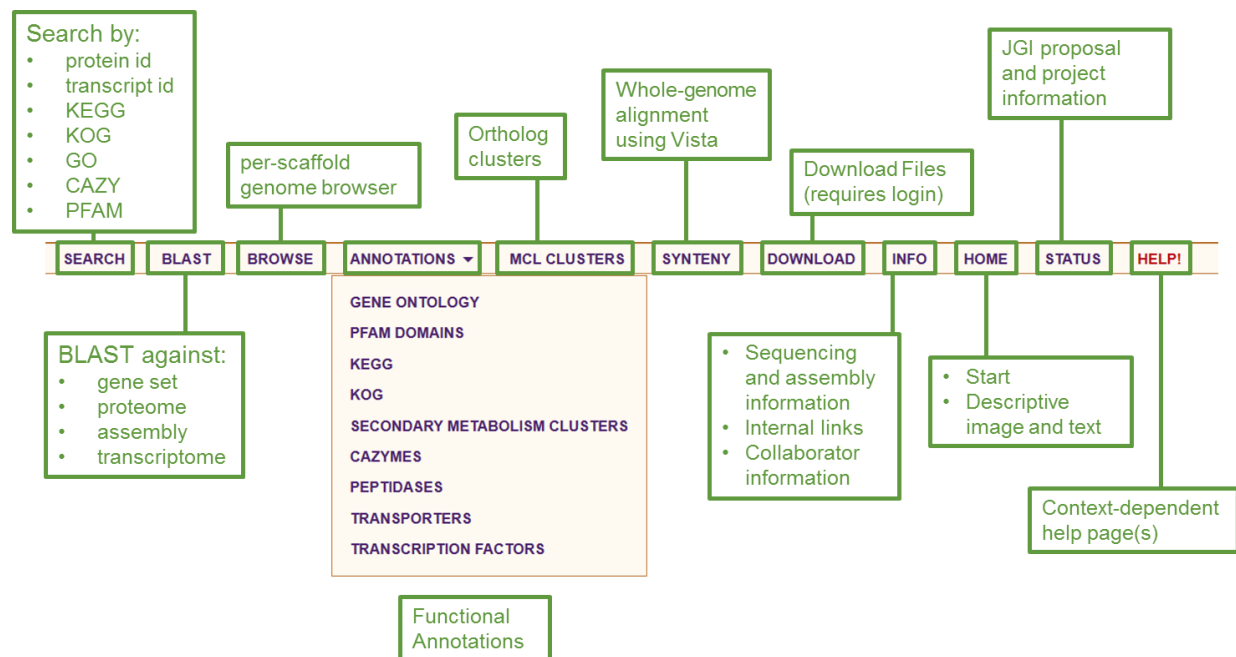
- Note: A global search ('Anything') for the keyword is the default search parameter at the JGI genome portal

Choose one of the *Aspergillus* genome portals you identified in the exercise above. This will take you to the genome portal for that species. The data for that genome is stored in a database, which can be identified using the url for the genome portal. For example, the home page for *Aspergillus welwitschiae* is <https://mycocosm.jgi.doe.gov/Aspwell1>. In this case, the databaseID is **Aspwell1** (i.e. the part right after mycocosm.jgi.doe.gov/). JGI uses these databaseIDs as shorthand for the portal in various places, including urls, download files, genome annotation browsers, clustering, and many others.



## The navigation toolbar

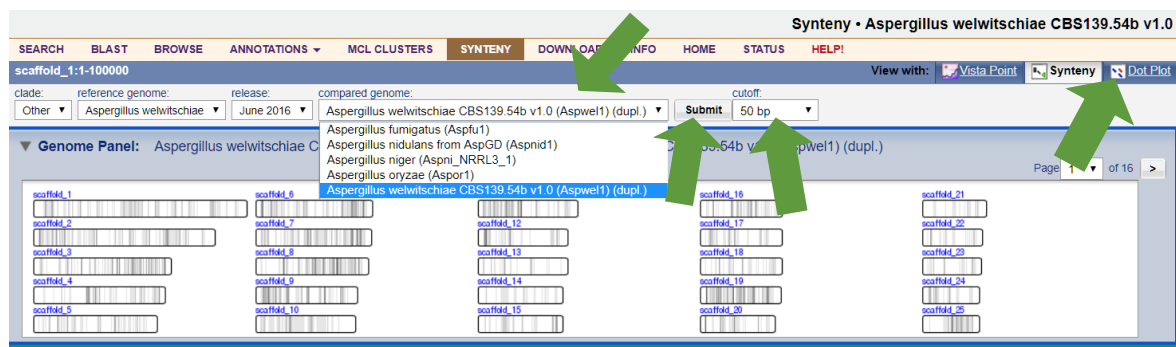
At the top of the genome portal is the navigation toolbar with the following functions.



We will explore most of the tools in a later exercise, but for now, let us investigate the **synteny browser**, the **genome browser**, and **protein pages**.

## Synteny

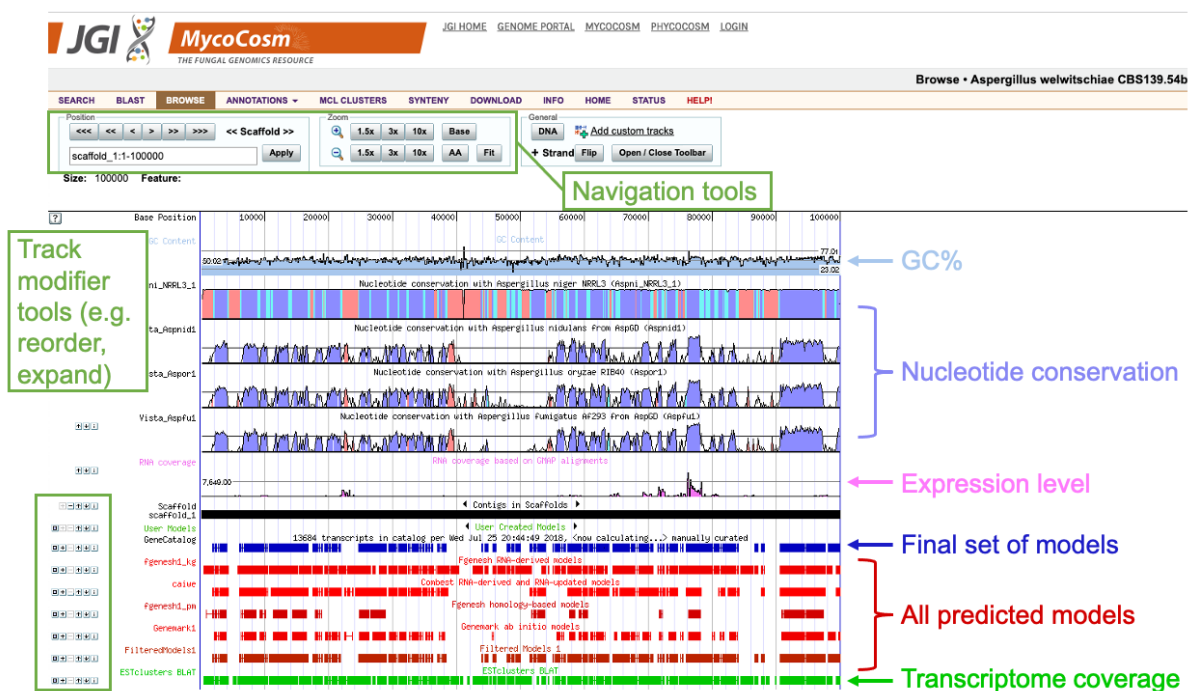
The synteny browser allows you to look at genome synteny at the nucleotide level using various filters.



Check out the basic functions of the synteny browser by clicking on the buttons and pulldowns marked by the arrows in the image above. To return from the 'Dot Plot' view, click 'Synteny' next to it. Note that self-alignment (main diagonal) is already removed in the dot plot if the genome is compared with itself, thus allowing you to observe other regions of duplication. We will explore this feature further in a later tutorial.

## Genome browser

Now, we will explore the genome browser using the ‘BROWSE’ tab. On the browser, you can navigate using available navigation tools and explore multiple layers (or “tracks”) of information we superimpose upon scaffolds. These include GC%, nucleotide conservation with closely related genomes, RNAseq expression data and transcriptome support, homology support, a total set of predicted gene models, and a single final set of models. Certain portals may also include multi-omics related tracks, such as for ChIPseq. With multiple sources of information visible simultaneously, identifying potentially interesting regions, or locations which need further improvement through manual curation, is relatively straightforward. You can also customize your view by clicking ‘Open/Close Toolbar’ and updating track settings. With the Toolbar open, you can get additional help by clicking the ‘?’ button, or close the toolbar by clicking the ‘X’ button (both of these are next to the refresh button).

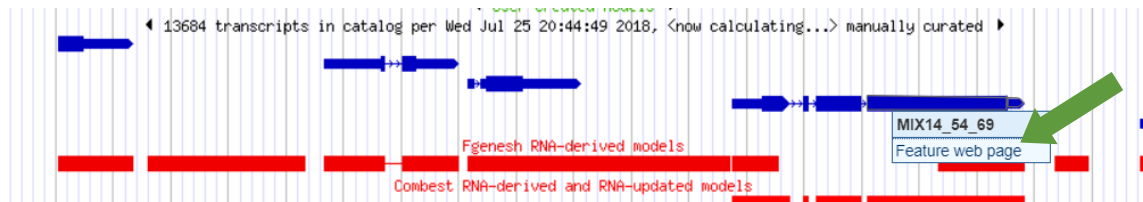


The nucleotide conservation tracks (based on vista alignment data) are a visual representation of alignments between two genomes. The peaks and valleys present on these curves represent percent conservation between aligned regions, and regions of high conservation are colored according to whether that region corresponds to an exon (dark blue), UTR (light blue), or non-coding (pink) region in the aligned genome. You can click on any of these tracks to get more information/download data.

The browser can be a useful tool for identifying potentially interesting genes, or exploring the gene neighborhood surrounding a gene of interest. For example, let's zoom in on one region of scaffold 1 to take a closer look. In the navigation tools, replace 'scaffold\_1:1-100000' with 'scaffold\_1:352500-362500' and press 'Apply' to zoom in. Alternatively, you can scroll (i.e. '<<<' and '>>>') and zoom (i.e. '+1.5x, +3x, or +10x) using the available navigation tools. You can also expand any track by clicking on it. Try this now for the 'GeneCatalog', which



will allow you view this track with increased resolution. Click on any gene in the GeneCatalog and click on 'Feature web page'. This will take you to the protein page.



## Protein page

On the protein page, detailed information is available regarding gene structure, functional annotation, homology to other organisms, etc. Almost everything on this page is click-able, bringing you to detailed descriptions of specific functional annotations. The top part of the page describes gene details, functional annotations and information on gene structure/domains present. Below this, BLAST alignments and scores (length, score, % identity, etc) are shown. Clicking on the alignment will allow you to view individual alignments, or clicking on the hit name will bring you to a page describing that protein in detail. Below you can select and retrieve protein sequences of either this protein, or any BLAST hit.

Browse • *Aspergillus welwitschiae* CBS139.54b v1.0

SEARCH BLAST BROWSE ANNOTATIONS MCL CLUSTERS SYNTENY DOWNLOAD INFO HOME STATUS HELP!

Name: MIX14\_54\_69  
Protein ID: 176767  
Location: scaffold\_1:51371-53916  
Strand: +  
Number of exons: 4  
Description: 54 EST updated Ab initio model  
Best Hit: jgi|Aspgh1|247693 Cysteine proteinase (model%: 99, hit%: 99, score: 3332, %id: 98) [no tax name]  
total hits (shown): 111 (10)

ASPECT: Biological Process, KOG GROUP, Cellular Processes And Signaling

GO Id: 0006511, KOG Id: KOG1865

GO Desc: ubiquitin-dependent protein catabolism, KOG Class: Posttranslational modification, protein turnover, chaperones

Interpro Id: IPR001394

Interpro Desc: Peptidase C19, ubiquitin carboxyl-terminal hydrolase, KOG Desc: Ubiquitin carboxyl-terminal hydrolase

View/modify manual annotation  
View nucleotide and 3-frame translation  
NCBI blastx Predicted number of transmembrane domains: 0

MIX14\_54\_69 To Genome Browser

1 111 44 395 1375 2546

1 65 129 193 257 321 384 448 512 576 639

PF00443: Ubiquitin carboxyl-terminal hydrolase [HMPFfam]  
PS50235: Ubiquitin specific protease (USP) domain profile. [ProSiteProfiles]  
SSF54001: [SUPERFAMILY]

Flip Start End Len ZC ZID Score Description [taxName]

1 638 638 100% 98% 3332 nr\_b\_b\_jgi|Aspgh1|247693 247693 Cysteine proteinase [no tax name]

7 561 561 99% 99% 2969 nr\_b\_b\_jgi|Aspni\_NRR3\_1|5027 5027 hypothetical protein [no tax name]

7 556 556 99% 79% 2348 nr\_b\_b\_jgi|Aspkal\_1|15194 15194 hypothetical protein [no tax name]

7 556 556 99% 79% 2340 nr\_b\_b\_jgi|Aspeuc1|414728 414728 Cysteine proteinase [no tax name]

7 556 556 99% 79% 2333 nr\_b\_b\_jgi|Asppil1|425235 425235 Cysteine proteinase [no tax name]

7 556 556 99% 79% 2320 nr\_b\_b\_jgi|Asptul1|32206 32206 hypothetical protein [no tax name]

7 554 556 99% 78% 2302 nr\_b\_b\_jgi|Aspfol1|63691 63691 hypothetical protein [no tax name]

7 556 556 99% 78% 2301 nr\_b\_b\_jgi|Aspneol1|310946 310946 Cysteine proteinase [no tax name]

7 556 556 99% 78% 2299 nr\_b\_b\_jgi|Aspvad1|431945 431945 Cysteine proteinase [no tax name]

7 556 556 99% 78% 2295 nr\_b\_b\_jgi|Asposci1|276808 276808 Cysteine proteinase [no tax name]

## EXERCISES:

1. Go to the test portal for this course at [mycocosm.jgi.doe.gov/Rhoto1\\_APA2687\\_1/](http://mycocosm.jgi.doe.gov/Rhoto1_APA2687_1/)

2. Hide all gene model tracks except the Gene Catalog and make all SNP tracks visible on the browser. Identify genes that show SNP variation within coding sequence and explore the protein pages of those genes.
3. Using the navigation tools to scroll, compare the RNA coverage tracks (from different conditions) to identify additional genes that show different expression profiles and identify what those genes are.
4. Using the data displayed on the browser, identify and take note of regions which need manual curation. Note: curation privileges were granted only to this specific genome portal. We will learn more about manual curation tools in this workshop.

## MycoCosm Group Portals

In addition to single genome portals, MycoCosm also hosts groups of genomes in ‘Group portals.’ These groups are either based on phylogeny (‘PhyloGroups,’ e.g. ‘Pezizomycetes’) in an automatic way, or on a common feature such as a lifestyle (‘EcoGroups,’ e.g. ‘Plant Pathogens’) in a more curated effort. We will look at phylogenetic groups in this exercise by starting at the MycoCosm home page.

Browse the cartoon phylogeny by clicking on a node or leaf. Each node and leaf has a menu of links to the associated Group Portal, several Group tools, and each member Genome Portal. The Pezizomycetes are an important class of Ascomycota that includes truffles and morels, and both saprobes and mycorrhizal symbionts. Let’s find out more about them by clicking on the ‘Pezizomycetes’ leaf, and selecting the top ‘Pezizomycetes’ link in the menu.

The screenshot shows the MycoCosm website interface. At the top, there is a navigation bar with links: JGI HOME, GENOME PORTAL, MYCOCOSM, PHYCOCOSM, and LOGIN. Below this is a secondary navigation bar with links: Home, Outreach, Video Tutorials, and About. The main content area is divided into several sections:

- Genes of Unknown Function:** Includes links like 'Claim to characterize!', '1000 Fungal Genomes project', 'Nominate New Species!', 'Genomic Encyclopedia of Fungi', 'Submit CSP proposal', and 'Myco-Ed Genomics Initiative'.
- Announcements:** Lists recent events such as the 'Fungal Genetics Conference' (March 12-17, 2024) and the 'Fungal Pathogen Genomics Course' (June 2-7, 2024).
- Latest Additions:** Lists newly added genomes, including *Paraglomus occultum* DAOM 240472 and *Rhizophagus irregularis* 4401.
- Cartoon Phylogeny:** A tree diagram showing the relationships between various fungal groups. The 'Pezizomycetes' group is highlighted with a green box and a green arrow. A dropdown menu is open for 'Pezizomycetes (76 genomes)', showing a list of member genomes and their accession numbers.
- Search and Tools:** A search bar at the top right allows users to search for MycoCosm data or JGI data. Below the phylogeny, there are links to various tools like BLAST, PFAM Domains, Secondary Metabolism Clusters, and more.

At the bottom of the page, there is a footer with contact information, a disclaimer, and a copyright notice: © 1997-2024 The Regents of the University of California.

## Anatomy of a MycoCosm Group Portal

The default view of a Group Portal shows a hyperlinked list of all of the member genomes, basic statistics of each genome, and the primary associated publication, if any. Genome names are linked to their Genome Portals, and publications to PubMed. The Pezizomycetes Group currently consists of 76 genomes, 25 of which are published.

Info • Pezizomycetes

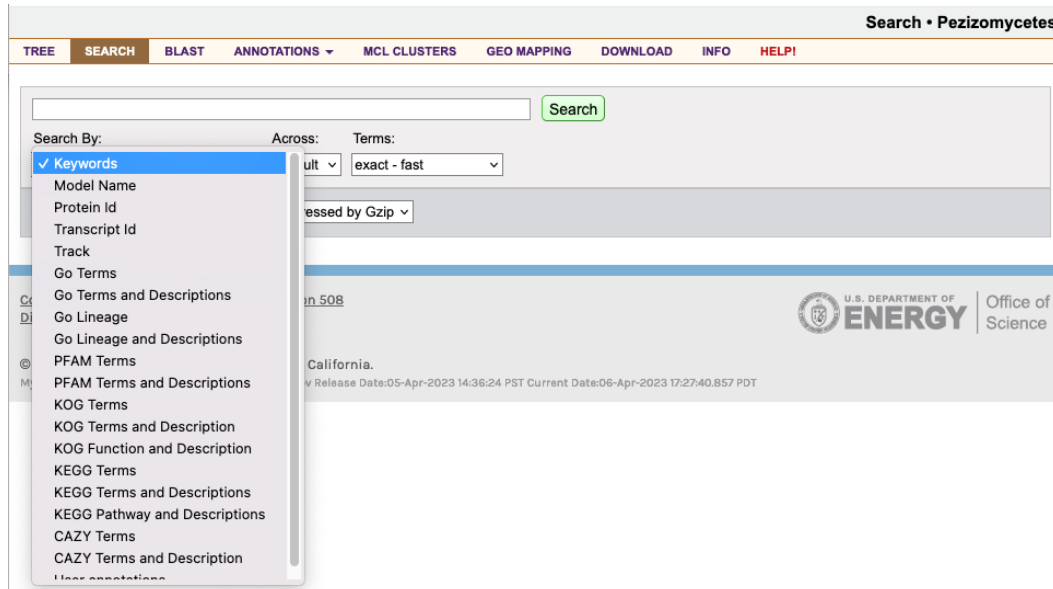
TREE	SEARCH	BLAST	ANNOTATIONS ▼	MCL CLUSTERS	GEO MAPPING	DOWNLOAD	INFO	HELP!
##	Name	Assembly Length	# Genes	Published				
1	<a href="#">Anthracobia sp. CZU v1.0</a>	53,428,419	14,194					
2	<a href="#">Ascobolus immersus RN42 v1.0</a>	59,529,874	17,877	<a href="#">Murat C et al., 2018</a>				
3	<a href="#">Ascodesmis nigricans CBS 389.68 v1.0</a>	27,385,218	9,622	<a href="#">Lütkenhaus R et al., 2019</a>				
4	<a href="#">Caloscypha fulgens ATCC 42695 v1.0</a>	44,839,139	9,203					
5	<a href="#">Choiromyces venosus 120613-1 v1.0</a>	126,035,033	17,986	<a href="#">Murat C et al., 2018</a>				
6	<a href="#">Disciotis venosa NRRL 24433 v1.0</a>	45,269,384	11,264					
7	<a href="#">Geopyxis carbonaria CBS 144460 v1.0</a>	49,454,325	11,663	<a href="#">Steindorff AS et al., 2022</a>				
8	<a href="#">Gyromitra esculenta CBS101906 v1.0</a>	45,046,303	11,202					
9	<a href="#">Gyromitra infula GyrinfSpk-SM18 v1.0</a>	45,877,806	11,121					
10	<a href="#">Kalaharituber pfeilii F3 v1.0</a>	78,991,937	15,190	<a href="#">Miyauchi S et al., 2020</a>				
11	<a href="#">Kalapuya brunnea GMNB372 v1.0</a>	189,316,756	11,095					
12	<a href="#">Leucangium carthusianum GMNB180 v1.0</a>	114,191,169	12,369					
13	<a href="#">Mattirolomyces terfezioides MAT_tt4AIII v1.0</a>	50,061,460	10,894					
14	<a href="#">Morchella americana PhC 192 v1.0</a>	63,616,142	11,032					
15	<a href="#">Morchella anatolica PhC 233 v1.0</a>	57,520,386	10,630					
16	<a href="#">Morchella angusticeps Mel-15 v1.0</a>	56,785,355	11,637					
17	<a href="#">Morchella arbutiphila PhC 291 v1.0</a>	54,167,150	11,179					
18	<a href="#">Morchella brunnea JGI1270818 v1.0</a>	54,761,096	11,196					
19	<a href="#">Morchella brunnea NRRL 20869 v1.0</a>	56,688,582	10,595					
20	<a href="#">Morchella conifericola Mel-32 v1.0</a>	52,567,085	10,590					
21	<a href="#">Morchella crassipes M84 v1.0</a>	54,774,707	10,480					
22	<a href="#">Morchella deliciosa PhC 191 v1.0</a>	56,658,686	11,470					
23	<a href="#">Morchella diminutiva Mes-2 v1.0</a>	53,423,193	10,998					
24	<a href="#">Morchella disparilis PhC 293 v1.0</a>	50,421,455	10,570					
25	<a href="#">Morchella dunalii PhC 240 v1.0</a>	50,708,236	10,363					
26	<a href="#">Morchella eohespera Mel-19 v1.0</a>	52,040,391	10,683					
27	<a href="#">Morchella esculenta PhC 158 v1.0</a>	51,151,225	10,777					
28	<a href="#">Morchella eximia CBS 144461 v1.0</a>	47,574,696	11,510					

Note the navigation tool bar at the top of the Group Page. Here it contains many of the same viewing, searching, and browsing functions as in the single genome portal, however with the notable absence of home, synteny, and genome browser tools (as these do not make sense for group portals). You can similarly find the context-dependent ‘HELP!’ function, that provides useful information specific to the current page. The ‘DOWNLOAD’ function is also available, from which you may, if registered and logged in, download files from all genomes in this Group, including genomic and transcriptomic assemblies, as well as structural and functional annotations.



Click on SEARCH to try the text-based search tool for all genomes in this group.

The default search is for keywords in the Default Analysis Track (the accepted gene model set) of each genome in the group. Pulldown menus allow choices other than the default settings.



Multiple search terms are combined using a logical 'AND', meaning that a search for the terms 'pheromone receptor' will return only results containing both 'pheromone' and 'receptor'. Try it! The basic anatomy of search output is shown below.

Links to

- protein page
- annotation page
- genome browser

Gene	Gene Ontology	Annotations	User Annotations
Portal: <a href="#">Phch2</a> Portal Name: <a href="#">Phanerochaete chrysosporium RP-78 v2.2</a> Protein Id: <a href="#">2912118</a> Transcript Id: <a href="#">2912333</a> Location: <a href="#">scaffold_10:92485-92969 (-)</a> Model Name: <a href="#">e_gpv1.10.150.1</a> Track: <a href="#">Genewise1Plus</a>	<a href="#">GO:0005199</a> • structural constituent of cell wall <a href="#">GO:0005618</a> • cell wall	<a href="#">PF01185</a> • Fungal hydrophobin <a href="#">IPR001338</a> • Hydrophobin	
Portal: <a href="#">Phch2</a> Portal Name: <a href="#">Phanerochaete chrysosporium RP-78 v2.2</a> Protein Id: <a href="#">3027085</a> Transcript Id: <a href="#">3027200</a> Location: <a href="#">scaffold_6:1957160-1957577 (+)</a> Model Name: <a href="#">gm1.5302_g</a> Track: <a href="#">Genemark1</a>	<a href="#">GO:0005199</a> • structural constituent of cell wall <a href="#">GO:0005618</a> • cell wall	<a href="#">PF01185</a> • Fungal hydrophobin <a href="#">IPR001338</a> • Hydrophobin	
Portal: <a href="#">Phch2</a> Portal Name: <a href="#">Phanerochaete chrysosporium RP-78 v2.2</a> Protein Id: <a href="#">3028111</a> Transcript Id: <a href="#">3028326</a> Location: <a href="#">scaffold_10:94606-95056 (+)</a> Model Name: <a href="#">gm1.7328_g</a> Track: <a href="#">Genemark1</a>	<a href="#">GO:0005199</a> • structural constituent of cell wall <a href="#">GO:0005618</a> • cell wall	<a href="#">PF01185</a> • Fungal hydrophobin <a href="#">IPR001338</a> • Hydrophobin	
Portal: <a href="#">Phch2</a> Portal Name: <a href="#">Phanerochaete chrysosporium RP-78 v2.2</a> Protein Id: <a href="#">3030340</a> Transcript Id: <a href="#">3030555</a> Location: <a href="#">scaffold_12:679129-679642 (-)</a> Model Name: <a href="#">gm1.8557_g</a> Track: <a href="#">Genemark1</a>	<a href="#">GO:0005199</a> • structural constituent of cell wall <a href="#">GO:0005618</a> • cell wall	<a href="#">PF01185</a> • Fungal hydrophobin <a href="#">IPR001338</a> • Hydrophobin	
Portal: <a href="#">Phch2</a> Portal Name: <a href="#">Phanerochaete chrysosporium RP-78 v2.2</a> Protein Id: <a href="#">2997</a> Transcript Id: <a href="#">2997</a> Location: <a href="#">scaffold_4:886000-886439 (-)</a> Model Name: <a href="#">Phch1.1geneset1_pg_C_scaffold_4000296</a> Track: <a href="#">Phch1_GeneCatalog_2013_10_13_09_53</a>	<a href="#">GO:0005199</a> • structural constituent of cell wall <a href="#">GO:0005618</a> • cell wall	<a href="#">PF01185</a> • Fungal hydrophobin <a href="#">IPR001338</a> • Hydrophobin	Name: <a href="#">PcH12 (Ursula Kues, 2011-08-15)</a> Description: <a href="#">class I hydrophobin (Ursula Kues, 2011-08-15)</a> Define: <a href="#">class I hydrophobin (Ursula Kues, 2011-08-15)</a>

The search output provides links to the genome portals, genome browser, protein pages, and functional annotations.

Exercise: Use the 'Lichens' group identified in the previous exercise for this exercise (mycocosm.jgi.doe.gov/lichens). How many genes in these genomes have Pfam annotations?

Hint: You can find this using the prefix 'HMMPfam' common to all Pfam annotations in MycoCosm.

There are multiple additional ways to investigate the Group. For example, the navigation bar's BLAST function will perform a Blast search against all genomes at once. Additional ANNOTATION functions include:

- MCL CLUSTERS – draft gene families clustered using Blastp and MCL
- PFAM DOMAINS - identified using InterProScan
- SECONDARY METABOLISM CLUSTERS – predicted secondary metabolism gene clusters based on Pfam domains
- CAZYMES – carbohydrate-active enzyme annotations done by the CAZy team (www.cazy.org)
- PEPTIDASES – assignments made by Blastp against the highly curated MEROPS database (www.ebi.ac.uk/merops)
- TRANSPORTERS - assignments made by Blastp against the highly curated Transporter Classification Database (www.tcdb.org)
- TRANSCRIPTION FACTORS - assignments made by Blastp against an in-house curated database

We will explore a few of these in subsequent exercises. Now let's look a little more closely at PEPTIDASES. Click on the tab in the menu bar to go to the PEPTIDASES page. Each row represents a Merops clan or family, and may be expanded or collapsed; clicking on the clan or family goes to the cognate Merops web page. Each column represents a genome and has a unique databaseID; the databaseID is short and often obscure, but the species and gene model set may be revealed by hovering over the databaseID. The number in each cell is the number of proteins in that genome with the assigned Merops classification. Clicking on it will present details about those proteins. For large groups like this one, you may need to scroll to the right to view the rest of the table.

Peptidases • Pezizomycetes

TREE SEARCH BLAST ANNOTATIONS MCL CLUSTERS GEO MAPPING DOWNLOAD INFO HELP!

Search for:  Any **Keywords** Filter ☒ S ☒ S  
Expand All Exact Clear

Annotations/Genomes

Antsp2 Ascim1 Ascni1 Calful1 Chove1 Disven1 Geocar1 Gyresc1 Gyrinf1 Kalbru1 Kalpfe1 Leuca1 Matter1 Mordun1 Moreoh1 Moresc1 Moresim1 Morflu1 Morqal1 Morhis1 Morimp1 Morhaki1 Morpal1 Morper1

Merops

AA 6 5 5 6 9 6 7 7 8 5 6 9 8 7 7 6 5 4 5 5 4 6 6 5 5 6 5 6 4 5 7 5  
AC 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1  
AD 4 4 3 3 4 6 2 2 3 2 3 3 2 2 3 2 1 3 2 2 2 3 4 2 1 1 3 1 3 2 1 3 1  
CA 24 19 18 25 23 22 28 24 21 21 26 22 3 24 24 23 27 25 24 28 21 24 27 24  
CD 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
CE 1 2 1 3 2 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1  
CF 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
CO 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
CP 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
GA 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
GB 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
L 4 4 4 4 4 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4  
Kazal 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
ID 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
IL 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
IV 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
JC 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
JE 3 3 3 3 3 4 3 3 3 3 3 4 3 3 3 3 3 4 3 3 3 3 3 3 3 3 3 3 3 3

Search by Keywords or Annotations

Click on number to go to web page with details about those proteins

Expand Merops classifications or click ID to go to clan/family page

Scroll to the right to view last columns; total count and Annotation Description

Let's search for a specific type of protein by entering 'subtilase' in the 'Search for:' textbox and clicking on 'Filter'. This search by 'Keywords' will look for matches to domain descriptions. Now only subtilase-related clans and families are shown (and inhibitors thereof). Note that the S08.115 peptidases are found only in a handful of organisms, mostly saprobes. To search by the specific Merops ID, enter 'S08.115' in the Search box, change the dropdown selection to 'Annotation', and click 'Filter'. The right-most column of numbers are totals for that Merops row, so let us click on the '10' to learn more about these proteases.

Rows: 10 25 rows per page

Protein Id	Location	Gene Length	Protein Length	Merops Annotations	Domains	Models	Domains
Ascim1 321137	<a href="#">scaffold_3:818,001-819,568</a>	1,568	401	S08.115	Subtilase family		
Ascim1 326276	<a href="#">scaffold_34:203,990-205,048</a>	1,059	276	S08.115	Subtilase family		
Ascim1 211249	<a href="#">scaffold_57:190,517-189,456</a>	1,062	306	S08.115	Subtilase family		
Ascni1 363694	<a href="#">scaffold_7:535,694-537,533</a>	1,840	453	S08.115	Subtilase family Peptidase inhibitor I9		
Ascni1 359457	<a href="#">scaffold_19:313,495-311,891</a>	1,605	419	cerevisin S08.115	Subtilase family Peptidase inhibitor I9		
Morpop1 602957	<a href="#">scaffold_99:102,934-101,805</a>	1,130	251	S08.115	Subtilase family		
Morpra1 711690	<a href="#">scaffold_5:2,218,063-2,220,014</a>	1,952	480	S08.115	Subtilase family Peptidase inhibitor I9		
Sarco1 402179	<a href="#">scaffold_8:110,499-109,257</a>	1,243	379	S08.115	Subtilase family Peptidase inhibitor I9		
Sarco1 418028	<a href="#">scaffold_283:11,576-10,133</a>	1,444	415	cerevisin S08.115	Subtilase family Peptidase inhibitor I9		
Tubbr1 1 876549	<a href="#">scaffold_441:86,341-86,006</a>	336	112	S08.115	Subtilase family		

The list of proteins shows basic information of the underlying gene models. The left-most column is the databaseID and proteinId, which together provide a unique identifier for the gene model in MycoCosm. The next 5 columns are self-explanatory, but note that Location is linked to the Browse view of that Genome Portal, and Merops and Domains are linked to the corresponding page on Merops and Pfam domains on EBI's InterPro site. The right-most column is by default in 'Models' mode and shows the exon-intron structure of each gene.

Click on ‘Domains’ to view the domain structure of the proteins. As is often the case with subtilases, the gene product is a pre-protein with a peptidase inhibitor domain that likely must be cleaved off before the mature proteinase can function as such.

Rows: 10 25 rows per page

Protein Id	Location	Gene Length	Protein Length	Merops Annotations	Domains	Models Domains
Ascm1_321137	scaffold_3:818,001-819,568	1,568	401	• S08.115	• Subtilase family	
Ascm1_326276	scaffold_34:203,990-205,048	1,059	276	• S08.115		
Ascm1_211249	scaffold_57:190,517-189,456	1,062	306	• S08.115	• Subtilase family	
Ascn1_363694	scaffold_7:535,694-537,533	1,840	453	• S08.115	• Subtilase family • Peptidase inhibitor I9	
Ascn1_359457	scaffold_19:313,495-311,891	1,605	419	• cerevisin • S08.115	• Subtilase family • Peptidase inhibitor I9	
Morpop1_602957	scaffold_99:102,934-101,805	1,130	251	• S08.115	• Subtilase family	
Morpra1_711690	scaffold_5:2,218,063-2,220,014	1,952	480	• S08.115	• Subtilase family • Peptidase inhibitor I9	
Sarco1_402179	scaffold_8:110,499-109,257	1,243	379	• S08.115	• Subtilase family • Peptidase inhibitor I9	
Sarco1_418028	scaffold_283:11,576-10,133	1,444	415	• cerevisin • S08.115	• Subtilase family • Peptidase inhibitor I9	
Tubbr1_1_876549	scaffold_441:86,341-86,006	336	112	• S08.115		

A similar set of tools is available for other functional annotations. For example, the tools in CAZYMES are shown below.

Annotations/Genomes	Ambr1	Bread_1	Cesut	Dicet1	Fomp3	Gamp1	Glor1_1	Lent6_1	Phae1	Phch1	Phch2	Phbr1	PeplRSB2_1	Trav1	Trnb1_1	Wolcor1	Total	Annotation Description
CAZy	287	475	370	475	402	549	362	516	423	417	450	543	328	485	486	289	6,856	CAZy
AA	36	104	62	88	43	93	41	94	83	85	89	83	40	89	86	27	1,153	Auxiliary Activities family
CBM	20	62	37	46	45	57	19	48	43	62	71	191	95	49	63	18	776	Carbohydrate-Binding Module family
CE	9	25	17	23	17	29	17	29	15	16	20	18	11	19	22	13	300	Carbohydrate Esterase family
EXP	13	16	11	13	20	22	13	19	13	9	11	15	21	12	13	13	234	Distantly related to plant expansins
GH	140	185	169	224	199	268	195	231	186	175	181	217	144	222	223	147	3,106	Glycoside Hydrolase family
GT	64	71	66	68	73	68	65	82	73	65	70	90	70	83	67	67	1,142	Glycosyl Transferase family
Myosin_motor	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	32	Glycosyltransferase Family 2
PL	3	10	6	11	3	9	9	11	8	4	6	7	5	9	10	2	113	Polysaccharide Lyase family

Click on counts to provide more details

Protein Id	Location	Gene Length	Protein Length	CAZy Annotations	Domains	Models Domains	Models Domains
Phch2_297975	scaffold_5:1,229,644-1,241,254	1,611	272	• Auxiliary Activity Family 5	• Flaxodioxin		
Phch2_2918435	scaffold_19:448,641-450,192	1,552	373	• Auxiliary Activity Family 2	• Peroxidase		
Phch2_3073209	scaffold_28:58,395-57,319	1,077	180	• Auxiliary Activity Family 3 / Subf 2	• GMC oxidoreductase		
Phch2_3005492	scaffold_10:527,184-528,465	1,282	321	• Auxiliary Activity Family 3 • Carbohydrate-Binding Module Family 1	• Fungal cellulose binding domain • Glycosyl hydrolase family 51		
Phch2_6139	scaffold_10:1,037,135-1,039,405	2,271	611	• Auxiliary Activity Family 3 / Subf 2	• GMC oxidoreductase • GMC oxidoreductase		
Phch2_2910210	scaffold_5:1,444,920-1,446,516	1,597	373	• Auxiliary Activity Family 2	• Peroxidase		
Phch2_2984061	scaffold_9:586,872-584,168	2,705	603	• Auxiliary Activity Family 3 / Subf 3	• GMC oxidoreductase • GMC oxidoreductase		

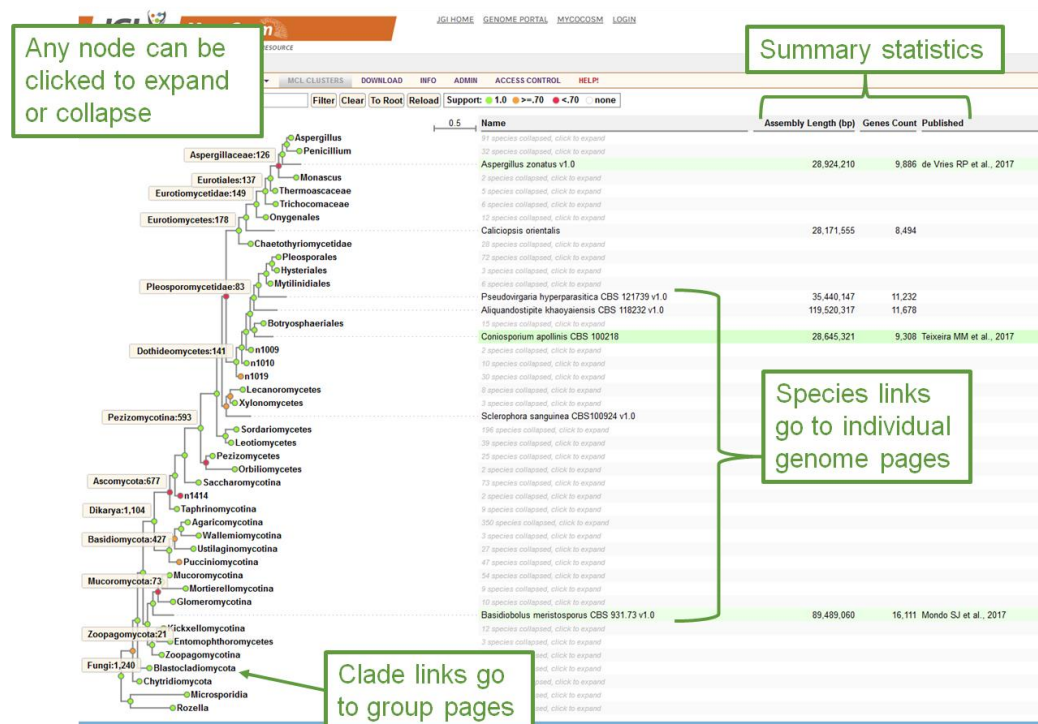
So far, we have browsed the Group based on common (or uncommon) functional annotation. Another way to browse is based on phylogeny, and may be accessed through the Menu bar’s Tree Item. Click on it:

JGI HOME GENOME PORTAL MYCOCOSM PHYCOCOSM LOGIN

Info • Pezizomycetes

TREE SEARCH BLAST ANNOTATIONS MCL CLUSTERS GEO MAPPING DOWNLOAD INFO HELPI

Unlike the cartoon on the MycoCosm home page, this tree, and others associated with each Group, is calculated by JGI using the clusters mentioned above, followed by MAFFT alignment of presumptive orthologs and FastTree modeling. Sometimes unexpected branchings appear that contradict other resources such as NCBI Taxonomy, but not in this case. The nodes are selectable, collapsible, and expandable. The table to the right of the tree recapitulates the Group data we saw earlier. The basic tools available in the tree view are shown below using the whole MycoCosm fungal tree as an example.



Let's take a look at a larger tree, the Eurotiomycetes. The main driver for the growth in Eurotiomycetes is JGI's *Aspergillus* Whole-Genus Sequencing Project, with the goal of sequencing ~300 species of *Aspergillus*.

Click on the MycoCosm link at the top of the page, above the Menu bar, to go back to the MycoCosm home page. From there, go to the Eurotiomycetes Group Portal. Or even better, on the leaf's menu click on 'Search' rather than the Group link, to go straight to the Group's Search tool. This way we don't have to wait for the entire list of genomes to load. Next, look for the known *Aspergillus nidulans* septin. Enter the GeneID "AN8182" into the Search box. We get a single gene, as expected.



**Search • Eurotiomycetes**

---

**TREE   SEARCH   BLAST   ANNOTATIONS ▼   GEO MAPPING   DOWNLOAD   INFO   HELP!**

---

Search By:    Across:    Terms:

---

as    compressed by

---

Total genes found: 1 25 rows per page

Gene	Gene Ontology	Annotations	User Annotations
Portal: <a href="#">Aspnid1</a> Portal Name: <b>Aspergillus nidulans</b> Protein ID: <b>1633</b> Transcript ID: <a href="#">1813</a> Location: <a href="#">ChrII_A_nidulans_FGSC_A4:1046574-1048647</a> Model Name: <b>AN8182</b> Track: <b>AspGD_genes</b>	<a href="#">GO:0004182</a> • obsolete carboxypeptidase A activity <a href="#">GO:0005515</a> • protein binding <a href="#">GO:0005525</a> • GTP binding <a href="#">GO:0006508</a> • proteolysis <a href="#">GO:0007049</a> • cell cycle <a href="#">GO:0008270</a> • zinc ion binding <a href="#">GO:0031105</a> • septin complex	<b>KOG2655</b> • Septin family protein (P-loop GTPase) <b>PF00735</b> • Septin <b>IPR000038</b> • missing_ipr000038 <b>IPR016491</b> • <b>IPR000834</b> •	

Click on the Protein ID 1633. This show the protein page for this gene and protein, which is part of the Aspnid1 Genome Portal. We will see more of the protein page in later exercises. *A. nidulans* was not sequenced at JGI. Instead this genome was imported from AspGD ([www.aspergillusgenome.org](http://www.aspergillusgenome.org)). This genome has also been imported by FungiDB, and again this page has a link to the corresponding FungiDB page. Try it!

Name:	AN8182
Protein ID:	1633
Location:	<a href="#">ChrII_A_nidulans_FGSC_A4:1046574-1048647</a>
Strand:	-
Number of exons:	7
Description:	
FungiDB:	<a href="#">Link to AN8182 in FungiDB</a>
AspGD:	<a href="#">Link to AN8182 in the Aspergillus Genome Database (AspGD)</a>
Best Hit:	<a href="#">gil259480955 tpelCBF74054.1 TPA: Septin Fragment</a> [Source:UniProtKB/TrEMBL;Acc:Q9C1M2][ <i>Aspergillus nidulans</i> FGSC A4](model%: 99, hit%: 100, score: 1918, %id: 100) [ <i>Aspergillus nidulans</i> FGSC A4]
total hits(shown)	347 (10)

ASPECT	GO Id	GO Desc	Interpro Id	Interpro Desc
Molecular Function	0005515	protein binding	IPR016491	Septin
	0005525	GTP binding	IPR000038	Cell division/GTP binding protein
	0004182	carboxypeptidase A activity	IPR000834	Peptidase M14, carboxypeptidase A
	0008270	zinc ion binding	IPR000834	Peptidase M14, carboxypeptidase A
Biological Process	0006508	proteolysis and peptidolysis	IPR000834	Peptidase M14, carboxypeptidase A
	0007049	cell cycle	IPR000038	Cell division/GTP binding protein
<b>KOG GROUP</b>	<b>KOG Id</b>	<b>KOG Class</b>		<b>KOG Desc</b>
Metabolism	KOG2655	Cell cycle control, cell division, chromosome partitioning		Septin family protein (P-loop GTPase)

Finally, for additional resources outside of this workshop, there are video tutorials available to the public on the MycoCosm homepage that walkthrough different features of the portal.

