

# Introduction to MycoCosm

To get started, connect to the MycoCosm Homepage ([mycocosm.jgi.doe.gov](http://mycocosm.jgi.doe.gov))

The screenshot shows the MycoCosm homepage. At the top right is a green button labeled "Click to login/register". Below the header, there's a navigation bar with links to Home, Outreach, Video Tutorials, and About. On the left, there are sections for the "1000 Fungal Genomes project" (with a "Nominate New Species!" link), the "Genomic Encyclopedia of Fungi" (with a "Submit CSP proposal" link), Announcements (listing events like the 2019 DOE JGI User Meeting and the Mycological Society of America Meeting), and New Genomes Releases (listing various fungal species releases from March 2019). The central feature is a phylogenetic tree of fungi. A dropdown menu "All MycoCosm Groups" is open, showing major clades: Basidiomycota, Ascomycota, Mucoromycota, Chytridiomycota, Zoopagomycota, Pezizomycotina, Dikarya, and All MycoCosm Groups. To the right of the tree is a list of fungal groups. Two green boxes on the right side provide information: one about data access, visualization, and analysis tools; and another about exploring gene models within species and comparing gene content between species.

- Provides
  - data access
  - visualization
  - analysis tools
- Allows researchers to
  - Explore gene models within species
  - Compare gene content between species

You can also get to MycoCosm from the unified JGI genome portal ([genome.jgi.doe.gov](http://genome.jgi.doe.gov))

The screenshot shows the JGI Genome Portal homepage. At the top right is a "Feedback" link. Below the header, there's a "My Favorites" link, a search bar, and a "Search" button. The main content area has sections for "Searching for Projects", "Downloading Files", and "Looking for Access". In the center, there's a promotional banner for the "Virtual VEGA Symposium 2021". A large red arrow points from the "Virtual VEGA Symposium 2021" section down to a green box containing the text: "MycoCosm & PhycoCosm - access to data, visualization, and analysis tools for comparative genomics of fungi or algae." To the right of this box is a "What's New" section with a "My Favorites" update and a "Download with Cart" section. At the bottom, there are links to other databases: Genomes OnLine Database (GOLD), Integrated Microbial Genomes (IMG) and Metagenomes (IMG/M), and Phytozome.

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, both genomes sequenced by JGI in collaboration with the community, and genomes 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!

[1000 Fungal Genomes project](#)  
[Nominate New Species!](#)

[Genomic Encyclopedia of Fungi](#)  
[Submit CSP proposal](#)

**Announcements**

2021  
Fungal Pathogen Genomics Workshop  
Hinxton, UK

Mar 15-20, 2022  
31st Fungal Genetics Conference in Asilomar  
Pacific Grove, CA

August 30-September 1, 2021  
Virtual JGI User Meeting

**Latest Additions**

April 07, 2021  
*Moesziomyces antarcticus* JCM 10317

April 02, 2021  
*Phakopsora pachyrhizi* UFGV02 v2.1

April 02, 2021  
*Phakopsora pachyrhizi* K8108 v2.0

March 31, 2021  
*Alternaria* sp. UNIPAMPA017 v1.0

March 31, 2021  
*Phyllosticta citribraziliensis* CPC 17464 v1.0

[more...](#)

**Fungi** (2,017 genomes)

[Tree](#)  
[Search](#)  
[BLAST](#)

All Myc

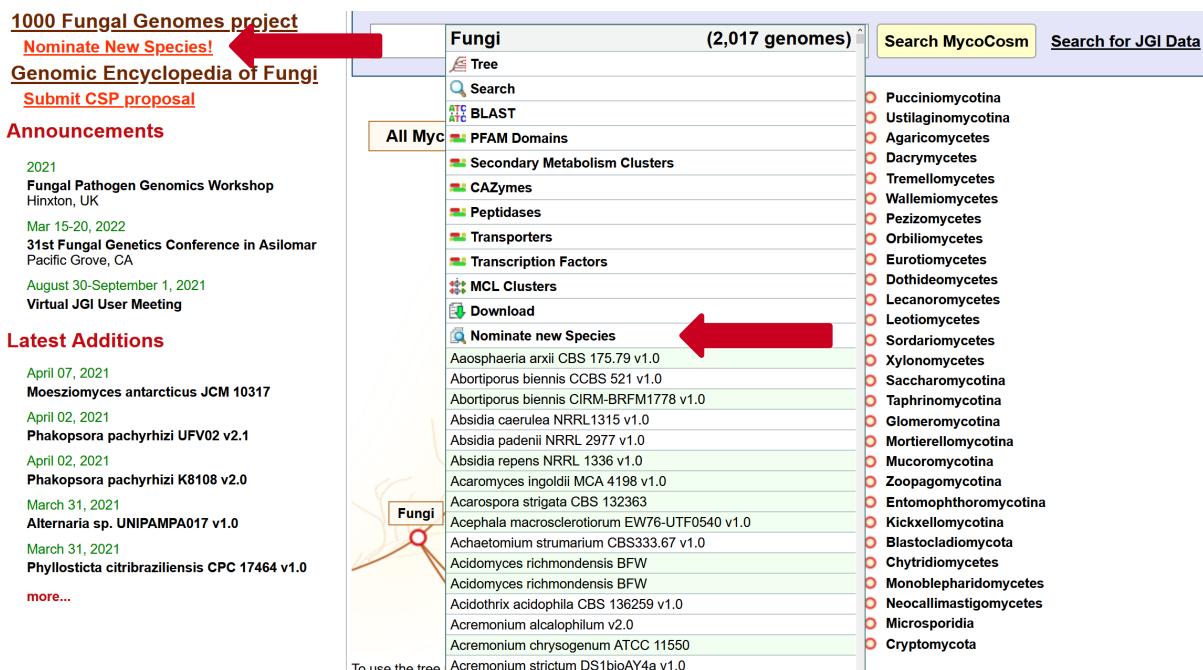
PFAM Domains  
Secondary Metabolism Clusters  
CAZymes  
Peptidases  
Transporters  
Transcription Factors  
MCL Clusters  
[Download](#)

[Nominate new Species](#)

Aaospheeria arxii CBS 175.79 v1.0  
Abortiporus biennis CCBS 521 v1.0  
Abortiporus biennis CIRM-BRFL1778 v1.0  
Absidia caerulea NRRL1315 v1.0  
Absidia padenii NRRL 2977 v1.0  
Absidia repens NRRL 1336 v1.0  
Acaromyces ingoldii MCA 4198 v1.0  
Acarospora strigata CBS 132363  
Acephala macrosclerotiorum EW76-UTF0540 v1.0  
Achaetomium strumarium CBS333.67 v1.0  
Acidomyces richmondensis BFW  
Acidomyces richmondensis BFW  
Acidothrix acidophila CBS 136259 v1.0  
Acremonium alcalophilum v2.0  
Acremonium chrysogenum ATCC 11550  
Acremonium strictum DS1bioAY4a v1.0

To use the tree

Search MyCoCosm Search for JGI Data



Searching for organism portals, groups and projects: The main search box on MycoCosm can be used to find organism genome portals using the default search parameters.

**Search MyCoCosm** **Search for JGI Data**

Exercise: Find 'Aspergillus' on MycoCosm using the default parameters.

Additional search options are available using "Search for JGI Data".

Exercise: Find 'Lichen' on genome.jgi.doe.gov using "Search for JGI Data". How does this compare to the default search on MycoCosm?

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

For the next part, we will use a single genome portal to explore tools available in MycoCosm.

## Navigating a 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. Eg; the home page for *Aspergillus welwitschiae* is mycocosm.jgi.doe.gov/**Aspwel1**/Aspwel1.home.html. In this case, the databaseID is Aspwel1 (the part right after mycocosm.jgi.doe.gov/). JGI uses this databaseID as a stand in for the portal in urls, download files, genome annotation browsers, clustering etc.

The screenshot shows the MycoCosm genome portal for *Aspergillus welwitschiae*. The top navigation bar includes links for JGI HOME, GENOME PORTAL, MYCOCOSM, and LOGIN. The main content area displays a grid of six culture images labeled CREA, OAT, CYA, YES, MEA, and OX. To the right of the cultures, there is descriptive text about the sequencing project and the Aspergilli genus. Below this, a section titled 'Genome Reference(s)' lists publications and a photo credit to Elen Kirstine Lyhne.

Aspergillus welwitschiae CBS 139.54b was sequenced as a part of the *Aspergillus* whole-genus sequencing project - a project dedicated to performing whole-genome sequencing of all members of the *Aspergillus* genus. The Aspergilli is a ubiquitous and species-rich genus, currently containing more than 300 filamentous fungi. The genus covers a wide range of phenotypes and has a substantial economic foot print, as it includes fermenters of foodstuffs, key cell factories for production of enzymes and organic acids, plant pathogens, model organisms for cell biology, human opportunistic pathogens, producers of animal and human mycotoxins, and degraders of a wide range of organic biomass relevant for bioenergy conversion.

**Genome Reference(s)**

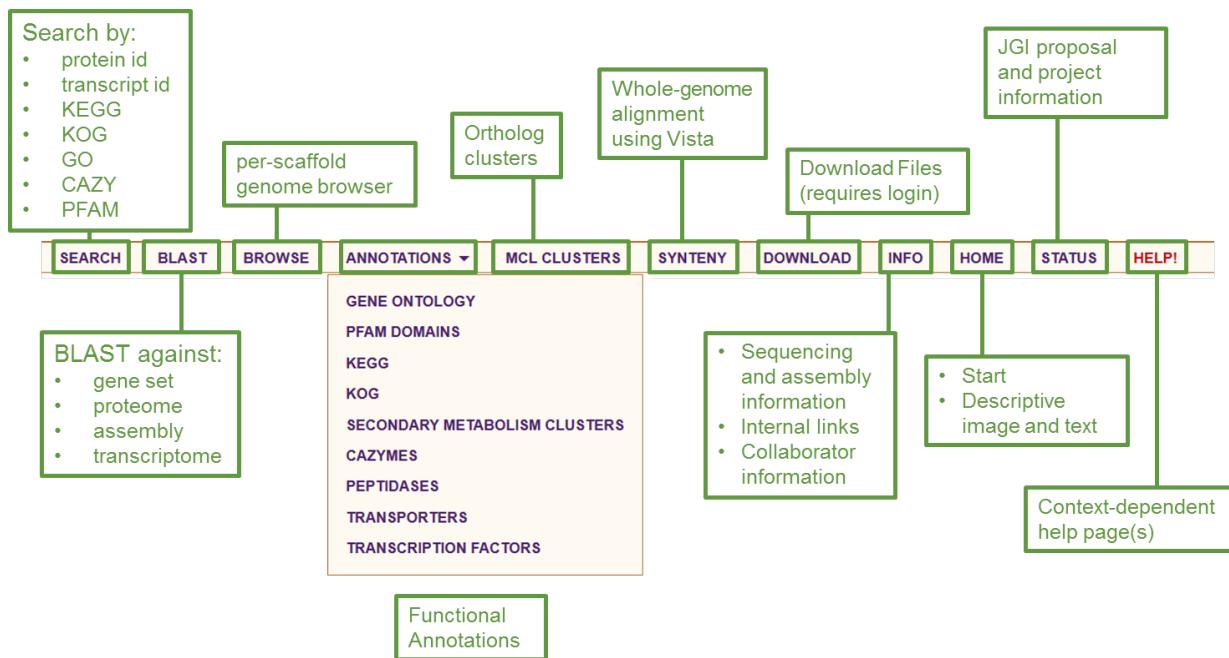
Please cite the following publication(s) if you use the data from this genome in your research:

Vest TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, Hoof JB, Brandl J, Salamov A, Riley R, Gladden JM, Phatale P, Nielsen MT, Lyhne EK, Kogte ME, Strasser K, McDonnell E, Barry K, Clum A, Chen C, LaButti K, Haridas S, Nolan M, Sandor L, Kuo A, Lipzen A, Hainaut M, Drula E, Tsang A, Magnuson JK, Henrissat B, Wiebenga A, Simmons BA, Makela MR, de Vries RP, Grigoriev IV, Mortensen UH, Baker SE, Andersen MR  
Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section Nigri. Nat Genet. 2018 Dec;50(12):1688-1695. doi: 10.1038/s41588-018-0246-1. Epub 2018 Oct 22.

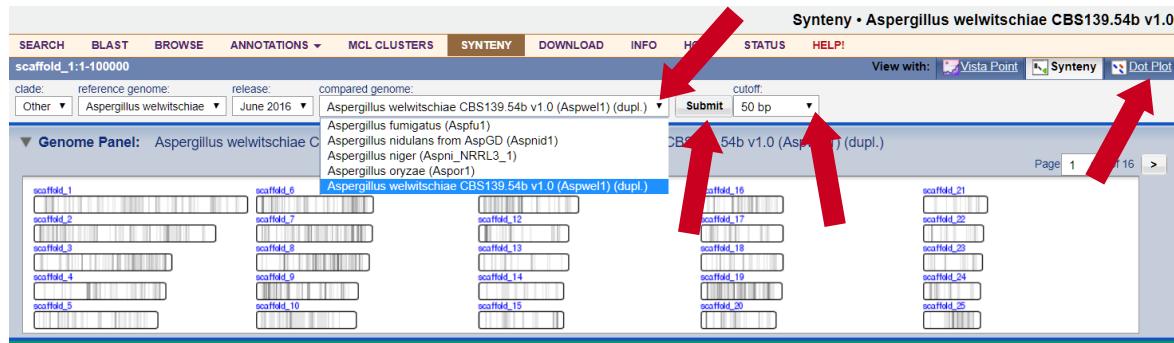
Photo credit: Elen Kirstine Lyhne

## The navigation toolbar

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



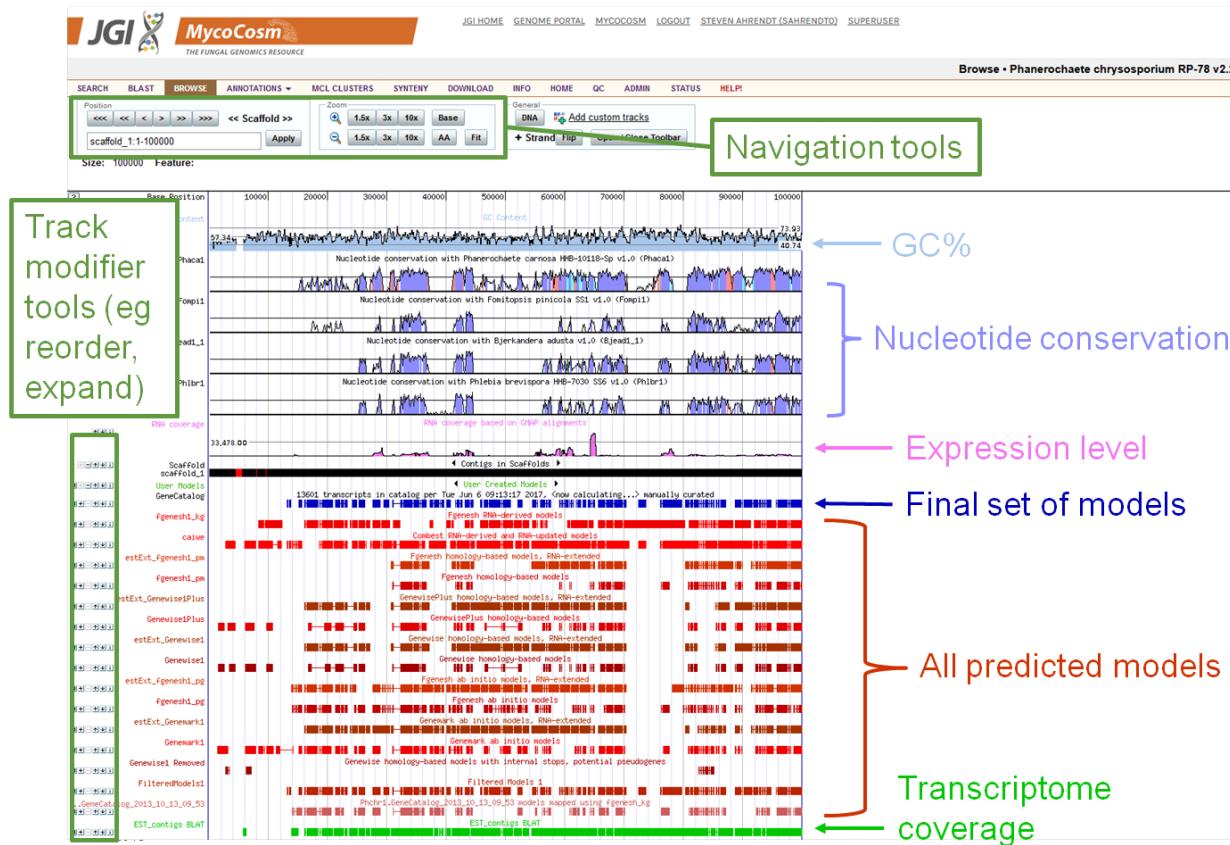
We will explore the most of the tools in a later exercise, but for now, let us see the synteny browser and the genome browser. The synteny browser allows you to look at genome synteny at the nucleotide level using various filters.



Check out the basic functions of the synteny browser by clicking on the buttons and pulldowns marked by the red 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.

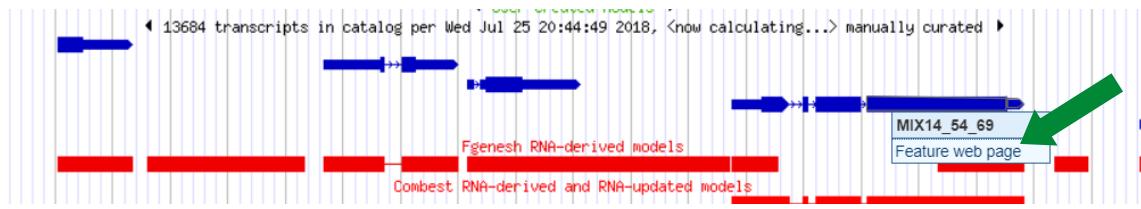
Now, we will explore the genome browser using the “BROWSE” tab. On the browser, you can navigate using available navigation tools and explore multiple layers of information we superimpose upon scaffolds. These include nucleotide conservation with closely related genomes, RNAseq expression data, Blastx results and more. 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. You can get additional help on how to take advantage of the toolbar 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 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 (ie ‘<<<’ and ‘>>>’) and zoom (ie +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.



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.

**Link back to genome browser**

**Gene details and functional annotation**

**Structural information**

**Functional domains**

**Blast results**

## EXERCISES:

1. Go to the test portal for this course at mycocosm.jgi.doe.gov/Rhoto1\_AP2687\_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.

- 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.
- 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 use this information in a later exercise on manual curation.

## MycoCosm Group Portals

In addition to single genome portals, MycoCosm also hosts groups of genomes in “Group portals”. These can be phylogenetic groups or manually curated groups such as “Plant Pathogens”. 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.

[1000 Fungal Genomes project](#)

[Nominate New Species!](#)

[Genomic Encyclopedia of Fungi](#)

[Submit CSP proposal](#)

**Announcements**

- 2021
- [Fungal Pathogen Genomics Workshop](#)  
Hinxton, UK
- Mar 15-20, 2022
- [31st Fungal Genetics Conference in Asilomar](#)  
Pacific Grove, CA
- August 30-September 1, 2021
- [Virtual JGI User Meeting](#)

**Latest Additions**

- April 07, 2021  
[Moesziomyces antarcticus JCM 10317](#)
- April 02, 2021  
[Phakopsora pachyrhizi UFGV02 v2.1](#)
- April 02, 2021  
[Phakopsora pachyrhizi K8108 v2.0](#)
- March 31, 2021  
[Alternaria sp. UNIPAMPA017 v1.0](#)
- March 31, 2021  
[Phyllosticta citribraziliensis CPC 17464 v1.0](#)
- [more...](#)

Pezizomycetes (56 genomes)

Tree

Search

ATC BLAST

PFAM Domains

Secondary Metabolism Clusters

CAZymes

Peptidases

Transporters

Transcription Factors

MCL Clusters

Download

Nominate new Species

Ascodolus immersus RN42 v1.0

Ascodesmis nigricans CBS 389.68 v1.0

Caloscypha fulgens ATCC 120613-1 v1.0

Choiromyces venosus NRRL 24433 v1.0

Disciotis venosa NRRL 24433 v1.0

Geopixys carbonaria DOB1671 v1.0

Gyromitra esculenta CBS101906 v1.0

Gyromitra infusa GyrimSpk-SM18 v1.0

Kaharituber felleii F3 v1.0

Leucangium carthusianum GMNBB180 v1.0

Morchella angusticeps Mel-15 v1.0

Morchella arbutiphila Phc291 v1.0

Morchella brunnea Mel-22 v1.0

Morchella diminutiva Mes-2 v1.0

Morchella disparilis Phc 293 v1.0

Morchella eohespera Mel-19 v1.0

Search MycoCosm

Search for JGI Data

Pucciniomycotina

Ustilaginomycotina

Agaricomycetes

Dacrymycetes

Tremellomycetes

Walemiomycetes

Pezizomycetes

Orbiliomycetes

Eurotiomycetes

Dothideomycetes

Lecanoromycetes

Leotiomycetes

Sordariomycetes

Xylinomycetes

Saccharomycotina

Taphrinomycotina

Glomeromycotina

Mortierellomycotina

Mucoromycotina

Zoopagomycotina

Entomophthoromycotina

Kickxellomycotina

Blastocladiomycota

Chytridiomycetes

Monoblepharidomycetes

Neocallimastigomycetes

Microsporidia

Cryptomycota

## 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 56 genomes, 19 of which are published. The group also shows the main navigation bar with the same function as in the single genome portal, with the notable absence of home, synteny and genome browser, since these do not make sense for group portals.

Group Name: Pezizomycetes

| ## | Name   | Assembly Length | # Genes | Published                 |
|----|--|-----------------|---------|---------------------------|
| 1  | <a href="#">Ascobolus immersus RN42 v1.0</a>         | 59,529,874      | 17,877  | Murat C et al., 2018      |
| 2  | <a href="#">Ascodesmis nigricans CBS 389.68 v1.0</a> | 27,385,218      | 9,622   | Lütkenhaus R et al., 2019 |
| 3  | <a href="#">Caloscypha fulgens ATCC 42695 v1.0</a>   | 44,839,139      | 9,203   |                           |
| 4  | <a href="#">Choiromyces venosus 120613-1 v1.0</a>    | 126,035,033     | 17,986  | Murat C et al., 2018      |
| 5  | <a href="#">Disciotis venosa NRRL 24433 v1.0</a>     | 45,269,384      | 11,264  |                           |
| 6  | <a href="#">Geopyxis carbonaria DOB1671 v1.0</a>     | 49,454,325      | 11,663  |                           |
| 7  | <a href="#">Gyromitra esculenta CBS101906 v1.0</a>   | 45,046,303      | 11,202  |                           |
| 8  | <a href="#">Gyromitra infula GyrintSpk-SM18 v1.0</a> | 45,877,806      | 11,121  |                           |
| 9  | <a href="#">Kalaharituber pfeillii F3 v1.0</a>       | 78,991,937      | 15,190  | Miyauchi S et al., 2020   |
| 10 | <a href="#">Leucangium carthusianum GMNB180 v1.0</a> | 114,191,169     | 12,369  |                           |
| 11 | <a href="#">Morchella angusticeps Mel-15 v1.0</a>    | 56,785,355      | 11,637  |                           |
| 12 | <a href="#">Morchella arbutiphila PhC291 v1.0</a>    | 54,167,150      | 11,179  |                           |
| 13 | <a href="#">Morchella brunnea Mel-22 v1.0</a>        | 56,688,582      | 10,595  |                           |
| 14 | <a href="#">Morchella diminutiva Mes-2 v1.0</a>      | 53,423,193      | 10,998  |                           |
| 15 | <a href="#">Morchella disparilis PhC 293 v1.0</a>    | 50,421,455      | 10,570  |                           |
| 16 | <a href="#">Morchella eohespera Mel-19 v1.0</a>      | 52,040,391      | 10,683  |                           |
| 17 | <a href="#">Morchella fluvialis PhC 265 v3.0</a>     | 55,013,795      | 11,113  |                           |
| 18 | <a href="#">Morchella hispaniolensis Mel-18 v1.0</a> | 62,728,251      | 12,099  |                           |
| 19 | <a href="#">Morchella importuna CCBAS932 v1.0</a>    | 48,213,273      | 11,600  | Murat C et al., 2018      |
| 20 | <a href="#">Morchella importuna SCYDJ1-A1 v1.0</a>   | 48,802,685      | 11,971  | Tan H et al., 2019        |
| 21 | <a href="#">Morchella kakiicolor PhC 280 v3.0</a>    | 68,805,960      | 10,759  |                           |
| 22 | <a href="#">Morchella palazonii PhC 149 v1.0</a>     | 53,912,323      | 11,272  |                           |
| 23 | <a href="#">Morchella peruviana NRRL 66754 v1.0</a>  | 63,365,966      | 10,593  |                           |

Note the top Menu bar. The Menu bar is found on all Group and Genome Portals, and offers appropriate tools (“Items”) for viewing, searching, and browsing genomic data. The default Group Portal Item is INFO. The Item to the right is a context-dependent HELP! that provides useful information specific to the current page. The Item to the left of INFO is DOWNLOAD, from which you may, if registered and logged in, download files of the genomic and transcriptomic assemblies, and structural and functional annotations, of all genomes of this Group. The SEARCH item provides a text-based search tool for all genomes in this group. Click on SEARCH to try it out.

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.

Search • Pezizomyctes

TREE SEARCH BLAST ANNOTATIONS MCL CLUSTERS DOWNLOAD INFO HELP!

Search By: Across: Terms:

Keywords Default exact - fast

compressed by Gzip

/Section 508

U.S. DEPARTMENT OF ENERGY Office of Science

of California.  
-1.nercsc.gov Release Date:01-Apr-2021 17:15:23 PST Current Date:09-Apr-2021 10:04:24.35 PDT

Search By: Across: Terms:

Keywords Default exact - fast

compressed by Gzip

/Section 508

U.S. DEPARTMENT OF ENERGY Office of Science

of California.  
-1.nercsc.gov Release Date:01-Apr-2021 17:15:23 PST Current Date:09-Apr-2021 10:04:24.35 PDT

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

**Links to external databases**

**Manual annotations from users**

| Gene  | Gene Ontology  | Annotations   | User Annotations   |
|---|--|---|--|
| Portal: Phchr2<br>Portal Name: Phanerochaete chrysosporium RP-78 v2.2<br>Protein id: 2912118<br>Transcript id: 2912333<br>Location: scaffold_10_52485-92959 (+)<br>Model Name: e_gwv1.10.150.1<br>Track: Genewise1Plus                                      | GO:0005199 • structural constituent of cell wall<br>GO:0005618 • cell wall | PF01185 • Fungal hydrophobin<br>IPR001338 • Hydrophobin |  |
| Portal: Phchr2<br>Portal Name: Phanerochaete chrysosporium RP-78 v2.2<br>Protein id: 3021085<br>Transcript id: 3021300<br>Location: scaffold_6_1957160-1957677 (+)<br>Model Name: gmt5302_g<br>Track: Genemark1   | GO:0005199 • structural constituent of cell wall<br>GO:0005618 • cell wall | PF01185 • Fungal hydrophobin<br>IPR001338 • Hydrophobin |  |
| Portal: Phchr2<br>Portal Name: Phanerochaete chrysosporium RP-78 v2.2<br>Protein id: 3029111<br>Transcript id: 3029326<br>Location: scaffold_10_94606-95056 (+)<br>Model Name: gmt7328_g<br>Track: Genemark1  | GO:0005199 • structural constituent of cell wall<br>GO:0005618 • cell wall | PF01185 • Fungal hydrophobin<br>IPR001338 • Hydrophobin |  |
| Portal: Phchr2<br>Portal Name: Phanerochaete chrysosporium RP-78 v2.2<br>Protein id: 3030340<br>Transcript id: 3030555<br>Location: scaffold_12_679129-679642 (+)<br>Model Name: gmt8557_g<br>Track: Genemark1  | GO:0005199 • structural constituent of cell wall<br>GO:0005618 • cell wall | PF01185 • Fungal hydrophobin<br>IPR001338 • Hydrophobin |  |
| Portal: Phchr2<br>Portal Name: Phanerochaete chrysosporium RP-78 v2.2<br>Protein id: 2997<br>Transcript id: 2997<br>Location: scaffold_4_886000-886439 (+)<br>Model Name: Phchr1GeneSh1_pg_C_scaffold_4000256<br>Track: Phchr1_GeneCatalog_2013_10_13_09_53 | GO:0005199 • structural constituent of cell wall<br>GO:0005618 • cell wall | PF01185 • Fungal hydrophobin<br>IPR001338 • Hydrophobin | Name: Pch12 ( <a href="#">Ursula Kues</a> , 2011-08-15)<br>Description: class I hydrophobin ( <a href="#">Ursula Kues</a> , 2011-08-15)<br>Define: class I hydrophobin ( <a href="#">Ursula Kues</a> , 2011-08-15) |

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.

Another way to search all Pezizomycetes genomes at once is to do a Blast search, using the Menu bar's BLAST Item. There are also multiple ways to browse the Group:

- 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 Pfams.
- CAZYMES – carbohydrate-active enzyme annotations done by the CAZy team ([www.cazy.org](http://www.cazy.org)).
- PEPTIDASES – assignments made by Blastp against the highly curated MEROPS database ([www.ebi.ac.uk/merops](http://www.ebi.ac.uk/merops)).
- TRANSPORTERS - assignments made by Blastp against the highly curated Transporter Classification Database ([www.tcdb.org](http://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 it to go to the PEPTIDASES browser. 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 portalId; the portalId is short and often obscure, but the species and gene model set may be revealed by hovering over the portalId. 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.

| Annotations/Genomes |  | Asim1 | Aser1 | Cat1 | Chov1 | Gecarl | Gyresci | Kalpfe1 | Mero1 | Molimp1 | Morsny1 | Pezech1 | Pezsub1 | Pleme1 | Pyrc01 | Pyrdom1 | Pyrom1 | Sarc01 | Tetra2 | Tetral1 | Tirniv1 | Thryb1 | Tirpa1 | Tubae1 | Tubb01 | Tubma1 | Tubme1/2                                      | Wilm1 | Total |
|---------------------|--|-------|-------|------|-------|--------|---------|---------|-------|---------|---------|---------|---------|--------|--------|---------|--------|--------|--------|---------|---------|--------|--------|--------|--------|--------|---|-------|-------|
| ↳ Merops            |  | 218   | 177   | 188  | 217   | 225    | 210     | 209     | 206   | 209     | 197     | 176     | 218     | 351    | 219    | 229     | 219    | 258    | 189    | 185     | 176     | 225    | 220    | 198    | 195    | 184    | 217   | 5,515 |       |
| ↳ AA                |  | 5     | 5     | 6    | 9     | 7      | 7       | 5       | 4     | 4       | 5       | 6       | 5       | 5      | 21     | 6       | 5      | 9      | 4      | 3       | 5       | 8      | 5      | 5      | 3      | 4      | 7   | 158   |       |
| ↳ AC                |  |       |       |      |       |        |         |         |       |         |         |         |         |        |        |         |        |        |        |         |         |        |        |        |        |        |   | 6     |       |
| ↳ AD                |  | 4     | 3     | 3    | 6     | 2      | 3       | 1       | 1     | 2       | 3       | 5       | 9       | 7      | 7      | 7       | 7      | 4      | 4      | 4       | 4       | 3      | 1      | 2      | 4      |        | 96  |       |       |
| ↳ CA                |  | 19    | 18    | 25   | 23    | 28     | 24      | 26      | 22    | 21      | 22      | 23      | 22      | 27     | 27     | 28      | 25     | 21     | 20     | 28      | 24      | 23     | 23     | 20     | 23     | 29     | 23  | 614   |       |
| ↳ CD                |  | 2     | 2     | 2    | 2     | 2      | 2       | 4       | 2     | 2       | 2       | 2       | 2       | 3      | 9      | 8       | 7      | 3      | 2      | 2       | 3       | 2      | 1      | 3      | 2      | 2      | 2   | 75    |       |
| ↳ CE                |  | 2     | 1     | 3    | 2     | 1      | 1       | 1       | 2     | 1       | 2       | 1       | 2       | 1      | 2      | 2       | 1      | 2      | 2      | 1       | 3       | 1      | 1      | 2      | 1      | 3      | 41  |       |       |
| ↳ 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      | 24  |       |       |
| ↳ CO                |  | 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      | 13  |       |       |
| ↳ 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      | 20  |       |       |
| ↳ GA                |  |       |       |      |       |        |         |         |       |         |         |         |         |        |        |         |        |        |        |         |         |        |        |        |        |        | 4   |       |       |
| ↳ GB                |  |       |       |      |       |        |         |         |       |         |         |         |         |        |        |         |        |        |        |         |         |        |        |        |        |        | 1   |       |       |
| ↳ L                 |  | 3     | 3     | 3    | 3     | 3      | 4       | 4       | 4     | 5       | 4       | 5       | 3       | 3      | 3      | 4       | 5      | 4      | 3      | 4       | 4       | 3      | 3      | 3      | 2      | 4      | 93  |       |       |
| ↳ ID                |  |       |       |      |       |        |         |         |       |         |         |         |         |        |        |         |        |        |        |         |         |        |        |        |        |        | 4   |       |       |
| ↳ IL                |  |       |       |      |       |        |         |         |       |         |         |         |         |        |        |         |        |        |        |         |         |        |        |        |        |        | 1   |       |       |
| ↳ IV                |  |       |       |      |       |        |         |         |       |         |         |         |         |        |        |         |        |        |        |         |         |        |        |        |        |        | 4   |       |       |
| ↳ JC                |  | 2     | 3     | 1    | 1     | 1      | 1       | 1       | 1     | 1       | 2       | 2       | 1       | 1      | 1      | 1       | 1      | 1      | 2      | 2       | 1       | 2      | 1      | 1      | 1      | 1      | 32  |       |       |
| ↳ JE                |  | 3     | 1     | 3    | 3     | 4      | 3       | 3       | 4     | 3       | 3       | 4       | 6       | 3      | 3      | 3       | 2      | 1      | 1      | 2       | 4       | 4      | 2      | 3      | 2      | 3      | 76  |       |       |
| ↳ M-                |  | 2     | 1     | 5    | 1     | 1      | 2       | 1       | 2     | 2       | 2       | 1       | 1       | 7      | 2      | 2       | 2      | 2      | 1      | 1       | 1       | 2      | 1      | 1      | 1      | 1      | 47  |       |       |
| ↳ zincin            |  | 15    | 6     | 12   | 12    | 12     | 25      | 15      | 16    | 18      | 12      | 13      | 17      | 21     | 13     | 17      | 16     | 14     | 10     | 10      | 10      | 17     | 15     | 15     | 13     | 10     | 14  | 368   |       |
| ↳ MC                |  | 2     | 1     | 1    | 1     | 1      | 1       | 1       | 2     | 1       | 1       | 1       | 2       | 2      | 2      | 2       | 1      | 1      | 1      | 1       | 1       | 1      | 1      | 1      | 1      | 1      | 27  |       |       |
| ↳ MD                |  |       |       |      |       |        |         |         |       |         |         |         |         |        |        |         |        |        |        |         |         |        |        |        |        |        | 5   |       |       |
| ↳ ME                |  | 2     | 1     | 1    | 1     | 2      | 1       | 3       |       | 1       | 1       | 1       | 2       | 1      | 1      | 1       | 2      | 1      | 2      | 1       | 2       | 1      | 1      | 1      | 1      | 2      | 34  |       |       |
|                     |  |       |       |      |       |        |         |         |       |         |         |         |         |        |        |         |        |        |        |         |         |        |        |        |        |        | Water nucleophile; water bound by single zinc |       |       |

Let's search for a common class of proteases by entering 'subtilase' in the 'Search for:' textbox and clicking on 'Filter'. 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. The right-most column of numbers are totals for that Merops row, so let us click on the ‘9’ to learn more about these proteases.

| Merops » S08.115             |                                |             |                |                     |   |                |
|------------------------------|--------------------------------|-------------|----------------|---------------------|---|----------------|
| Rows: 9   25   rows per page |                                |             |                |                     |   |                |
| Protein Id                   | Location                       | Gene Length | Protein Length | Merops Annotations  | Domains                                   | Models Domains |
| Ascm1_326276                 | scaffold_34:203,990-205,048    | 1,059       | 276            | •S08.115            |   |                |
| Ascm1_321137                 | scaffold_3:818,001-819,568     | 1,568       | 401            | •S08.115            | •Subtilase family                         |                |
| 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 |                |
| 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            |   |                |

The list of proteins shows basic information of the underlying gene models. The left-most column is the portalId 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 appropriate pages in Merops and Pfam. 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.

| Merops » S08.115             |                                |             |                |                     |   |                |
|------------------------------|--------------------------------|-------------|----------------|---------------------|---|----------------|
| Rows: 9   25   rows per page |                                |             |                |                     |   |                |
| Protein Id                   | Location                       | Gene Length | Protein Length | Merops Annotations  | Domains                                   | Models Domains |
| Ascm1_326276                 | scaffold_34:203,990-205,048    | 1,059       | 276            | •S08.115            |   |                |
| Ascm1_321137                 | scaffold_3:818,001-819,568     | 1,568       | 401            | •S08.115            | •Subtilase family                         |                |
| 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 |                |
| 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. Eg; the tools in CAZYMES are shown below.

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

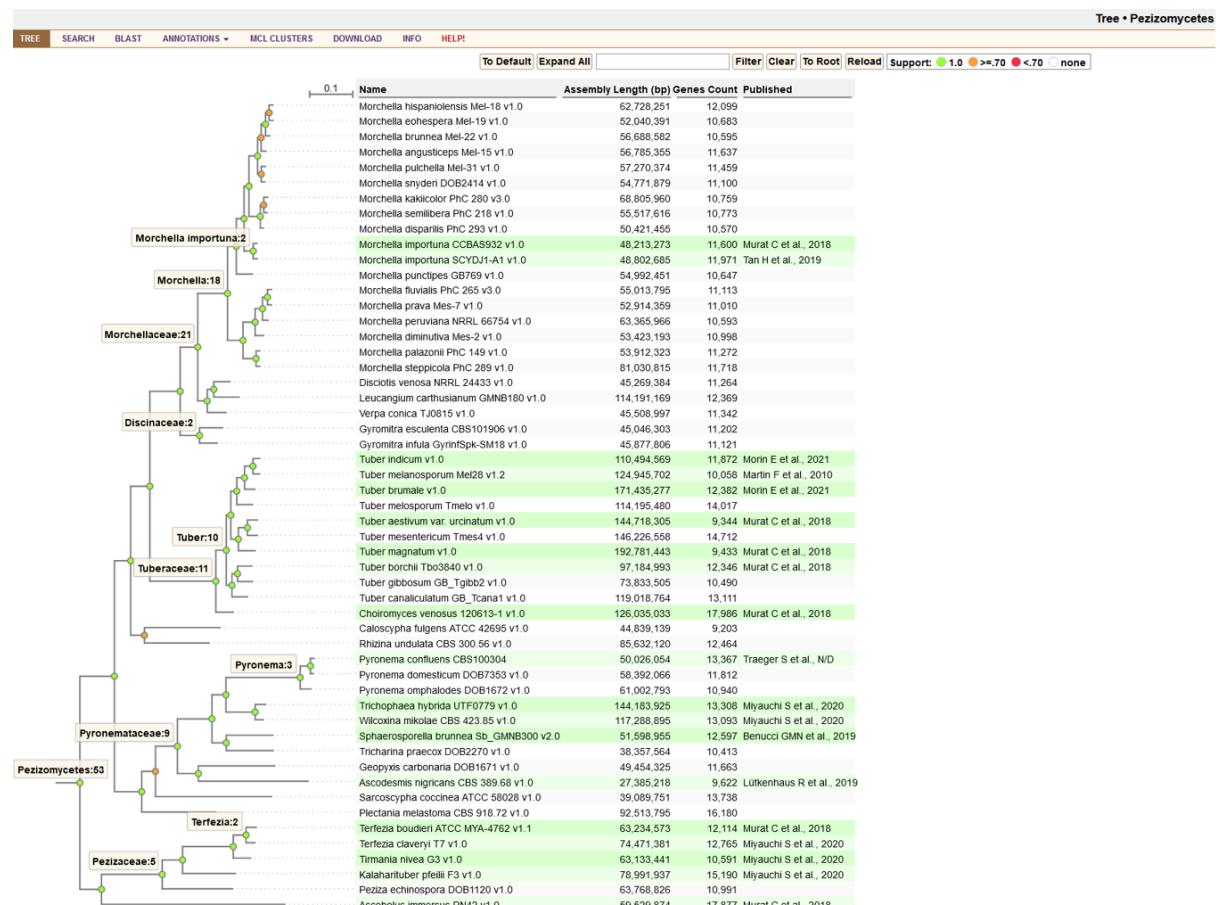
Search for: To Default Any Keywords Filter Show only filtered results  
Exact Clear Show only filtered totals  
Expand All

| Annotations/Genomes | Annotation Description |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |                     |       |     |       |                              |                                      |
|---------------------|------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|-------|-----|-------|------------------------------|--------------------------------------|
|                     | Annotations/Genomes    | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Annotations/Genomes | Total |     |       |                              |                                      |
| ↳ CAZy              | 287                    | 475                 | 370                 | 475                 | 402                 | 548                 | 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                  | 93                  | 40                  | 89    | 86  | 27    | 1,153                        | Auxiliary Activities family          |
| ↳ CBM               | 20                     | 62                  | 37                  | 46                  | 45                  | 57                  | 19                  | 48                  | 43                  | 62                  | 71                  | 103                 | 45                  | 49    | 63  | 18    | 776                          | Carbohydrate-Binding Module family   |
| ↳ CE                | 9                      | 25                  | 17                  | 23                  | 17                  | 29                  | 19                  | 25                  | 16                  | 20                  | 18                  | 11                  | 19                  | 22    | 13  | 300   | Carbohydrate Esterase family |                                      |
| EXPN                | 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                  | 70                  | 83                  | 67    | 67  | 1,142 | Glycosyl Transferase family  |                                      |
| Myosin_motor        | 2                      | 2                   | 2                   | 2                   | 2                   | 2                   | 3                   | 2                   | 2                   | 1                   | 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          |

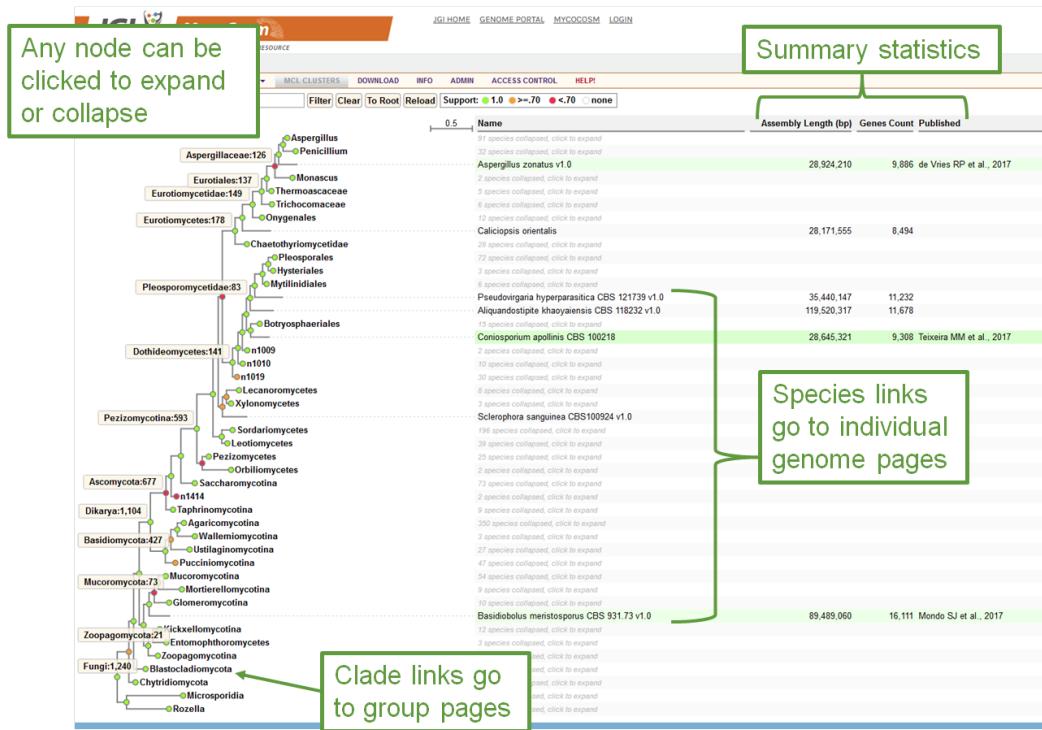
Click on counts to provide more details

| Protein Id     | Location                         | Gene Length | Protein Length | CAZy Annotations  |  |             | Domains | Models Domains |         | Models | Domains |
|----------------|----------------------------------|-------------|----------------|---|--|-------------|---------|----------------|---------|--------|---------|
|                |                                  |             |                | Annotations   | Annotations  | Annotations |         | Models         | Domains |        |         |
| Pchcr2_2978675 | sccaffold_5_1.239.644_1.241.254  | 1,611       | 272            | •Auxiliary Activity Family 6  | •Flavodoxin  |             | 644     | 52             | 154     | 100    | 435     |
| Pchcr2_2918435 | sccaffold_19.448.641-450.192     | 1,552       | 373            | •Auxiliary Activity Family 2  | •Peroxidase  |             | 61      | 154            | 57      | 219    | 42      |
| Pchcr2_2678200 | sccaffold_26.53.395-57.319       | 1,077       | 180            | •Auxiliary Activity Family 3 / Subf 2                               | •GMC oxidoreductase  |             | 50      | 42             | 79      | 51     | 424     |
| Pchcr2_3005492 | sccaffold_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 |             | 17      | 10             | 40      | 74     | 67      |
| Pchcr2_6199    | sccaffold_10.1.037.135-1.039.405 | 2,271       | 611            | •Auxiliary Activity Family 3 / Subf 2                               | •GMC oxidoreductase •GMC oxidoreductase                        |             | 141     | 87             | 184     | 60     | 49      |
| Pchcr2_2910312 | sccaffold_9.1.444.920-1.446.516  | 1,597       | 373            | •Auxiliary Activity Family 2  | •Peroxidase  |             | 28      | 79             | 56      | 62     | 64      |
| Pchcr2_2984661 | sccaffold_9.586.872-584.168      | 2,705       | 603            | •Auxiliary Activity Family 3 / Subf 3                               | •GMC oxidoreductase •GMC oxidoreductase                        |             | 28      | 95             | 100     | 111    | 109     |

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:



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.



Now that we are comfortable with a smaller Group, let's take a look at a larger tree, the Eurotiomycetes, with 353 genomes. The main driver for its 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 list of 353 genomes to load. Next enter the known *Aspergillus nidulans* septin GeneID AN8182 into the Search box. We get a single gene, as expected.

Search • Eurotiomycetes

**TREE**   **SEARCH**   **BLAST**   **ANNOTATIONS** ▾   **DOWNLOAD**   **INFO**   **HELP!**

AN8182     

Search By:      Across:      Terms:  
 Keywords      Default      exact - fast

as CSV      compressed by Gzip

Total genes found: 1      25 rows per page      << < > >>

| Gene   | Gene Ontology  | Annotations   | User Annotations |
|--|--|---|------------------|
| Portal: <a href="#">Aspnid1</a><br>Portal Name: <a href="#">Aspergillus nidulans</a><br>Protein Id: <a href="#">1633</a> ←<br>Transcript Id: <a href="#">1813</a><br>Location: <a href="#">ChrII_A_nidulans_FGSC_A4:1046574-1048647(-)</a><br>Model Name: <a href="#">AN8182</a><br>Track: <a href="#">AspGD_genes</a> | <a href="#">GO:0004182</a> • obsolete carboxypeptidase A activity<br><a href="#">GO:0005515</a> • protein binding<br><a href="#">GO:0005525</a> • GTP binding<br><a href="#">GO:0006508</a> • proteolysis<br><a href="#">GO:0007049</a> • cell cycle<br><a href="#">GO:0008270</a> • zinc ion binding<br><a href="#">GO:0031105</a> • septicin complex | <a href="#">KOG2655</a> • Septin family protein (P-loop GTPase)<br><a href="#">PF00735</a> • Septin<br><a href="#">IPR000038</a> • missing_ipr000038<br><a href="#">IPR016491</a> • Septin<br><a href="#">IPR000834</a> • Peptidase_M14 |                  |

Click on the Protein Id 1633. This shows 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)), and this page has a link to the cognate AspGD page. 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:          | gi 25940053 tel CBT74054 TPA_Sepin_Fragment [Source:UniProtKB/TrEMBL;Acc:Q9C1M2] [Aspergillus nidulans FGSC A4] (model%: 99, hit%: 100, score: 1918, %id: 100) [Aspergillus nidulans FGSC A4] |  |                                       |                                   |
| total hits(shown)  | 347 (10)  |  |                                       |                                   |
| <b>ASPECT</b>      | <b>GO Id</b>  | <b>GO Desc</b>   | <b>Interpro Id</b>                    | <b>Interpro Desc</b>              |
| Molecular Function | 0004182   | carboxypeptidase A activity                                | IPR000834                             | Peptidase M14, carboxypeptidase A |
|                    | 0005515   | protein binding  | IPR016491                             | Septin                            |
|                    | 0005525   | GTP binding  | IPR000038                             | Cell division/GTP binding protein |
|                    | 0008270   | zinc ion binding   | IPR000834                             | Peptidase M14, carboxypeptidase A |
| Biological Process | 0007049   | cell cycle   | IPR000038                             | Cell division/GTP binding protein |
|                    | 0006508   | proteolysis and peptidolysis                               | IPR000834                             | Peptidase M14, carboxypeptidase A |
| <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) |                                   |