# 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/)

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

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

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.

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

Using 'Search for JGI Data' will allow searching the Genome Portal.

- Exercise: Find 'Lichen' on the Genome Portal (https://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. For example, the home page for *Aspergillus welwitschiae* is https://mycocosm.jgi.doe.gov/**Aspwel1**. In this case, the databaseID is **Aspwel1** (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.
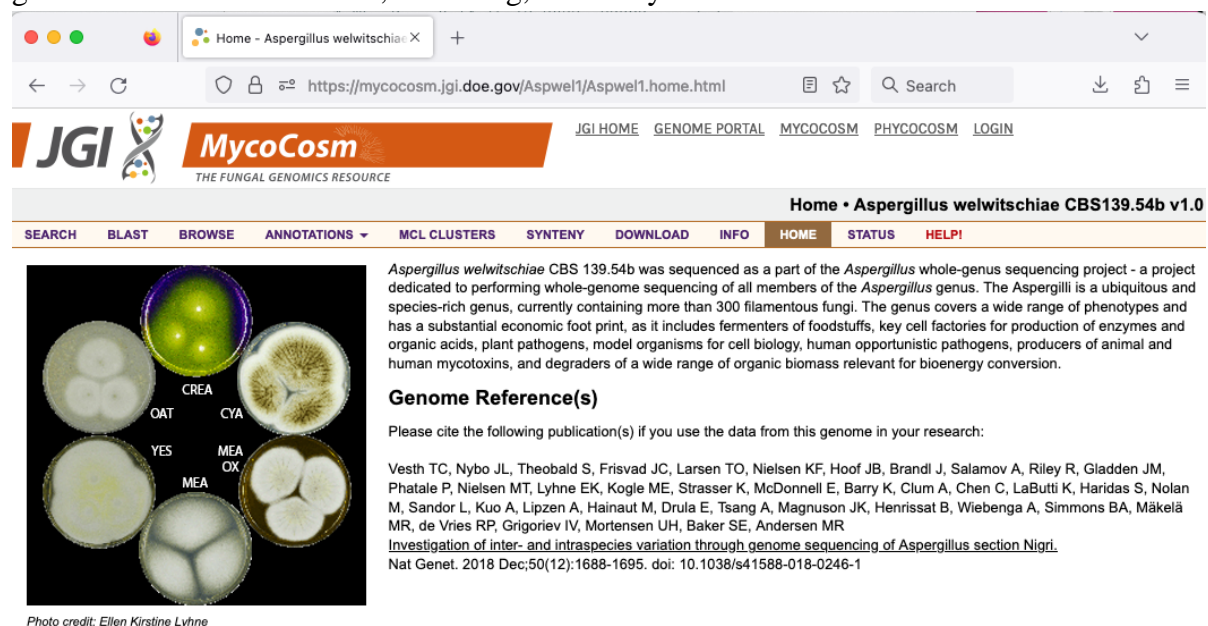
The navigation toolbar

At the top of the genome portal is the navigation toolbar with the following functions.
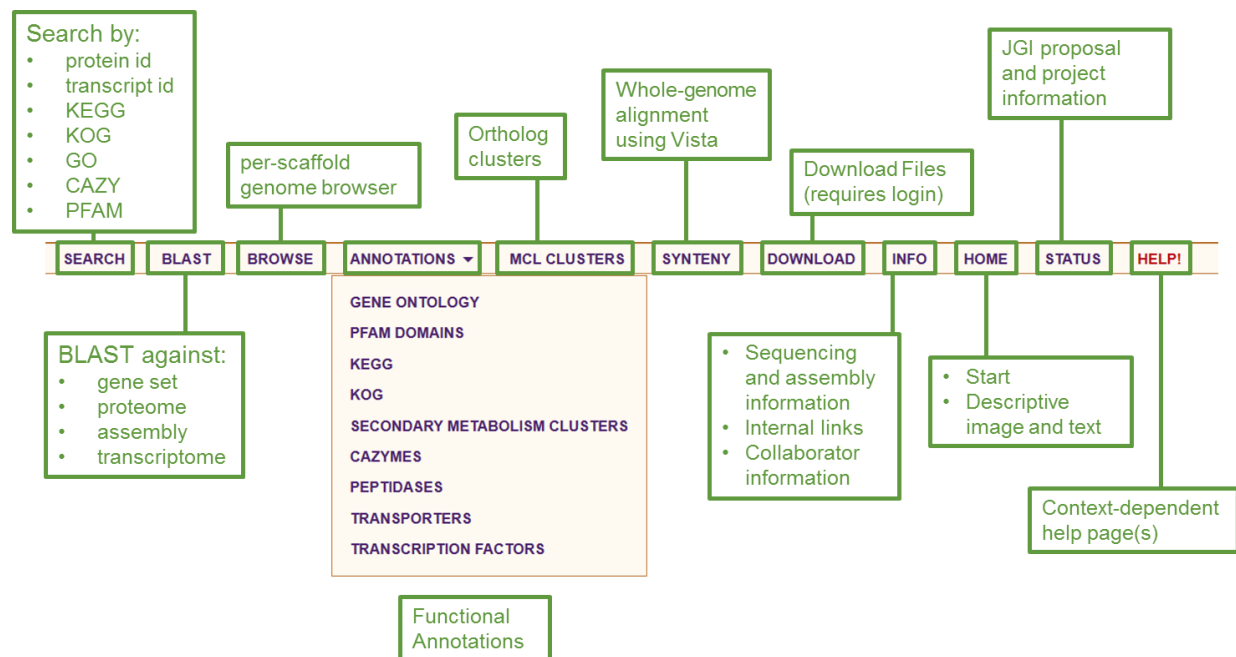
We will explore most of the tools in a later exercise, but for now, let us investigate the **synteny browser**, the **genome browser**, and **protein pages**.

Synteny

The synteny browser allows you to look at genome synteny at the nucleotide level using various filters.
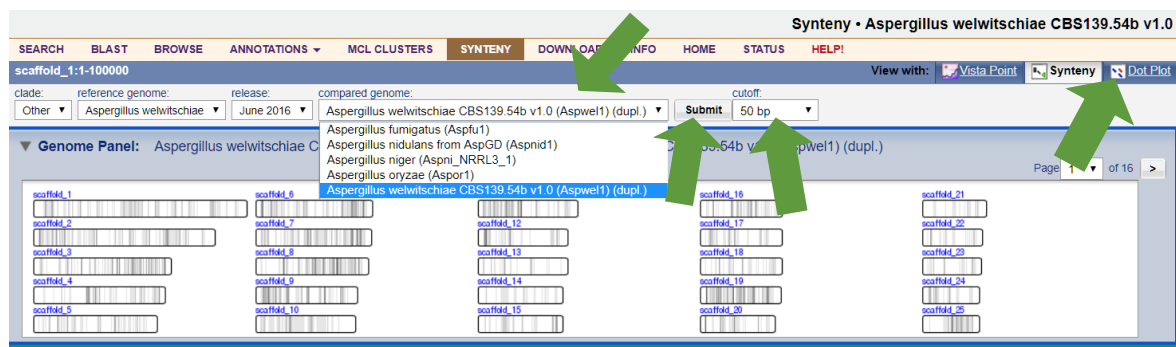
Check out the basic functions of the synteny browser by clicking on the buttons and pulldowns marked by the arrows in the image above. To return from the 'Dot Plot' view, click 'Synteny' next to it. Note that self-alignment (main diagonal) is already removed in the dot plot if the genome is compared 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).
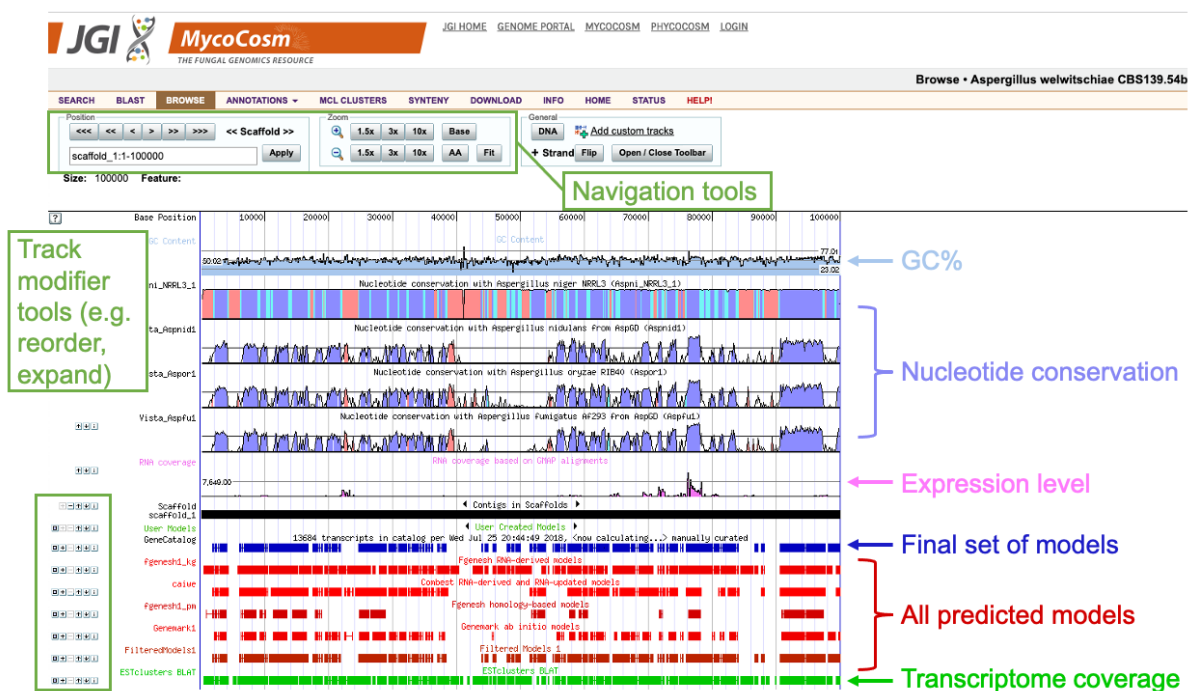
The nucleotide conservation tracks (based on vista alignment data) are a visual representation of alignments between two genomes. The peaks and valleys present on these curves represent percent conservation between aligned regions, and regions of high conservation are colored according whether that region 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.
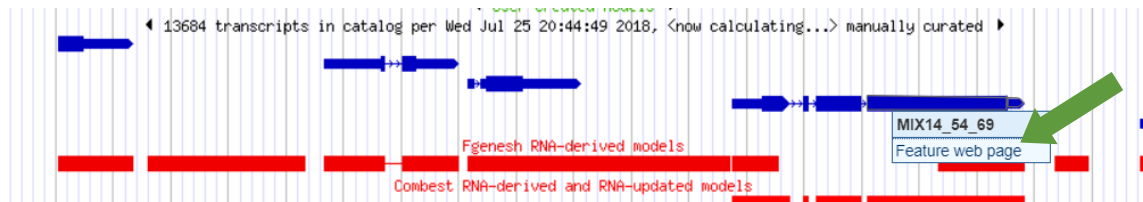
Protein page

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.

EXERCISES:

1. Go to the test portal for this course at 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.

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.

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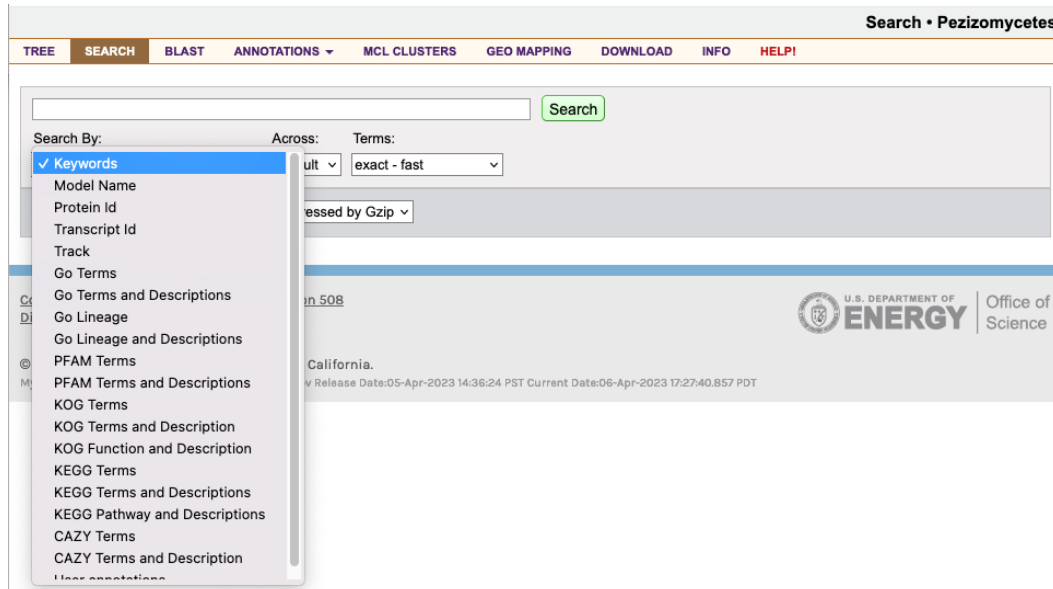
Multiple search terms are combined using a logical 'AND', meaning that a search for the terms 'pheromone receptor' will return only results containing both 'pheromone' and 'receptor'. Try it! The basic anatomy of search output is shown below.

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.

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

Search for:

To Default   Expand All   [Any ▾] [Keywords ▾] Filter   Exact ▾   Clear

Hover over to reveal organism

Search by Keywords or Annotations

Annotations/Genomes	Antsp2	Ascim1	Ascni1	Calfu1	Chove1	Disven1	Geocar1	Gvresc1	Gyrinf1	Kalbru1	Kalpfe1	Leuca1	Matter1		Mordun1	Moreoh1	Moresc1	Morexi1	Morexim1	Morfluv1	Morgal1	Morhis1	Morimp1	Morkaki1	Morpal1	Morper1
◢ Merops	221	218	177	188	217	219	225	210	243	192	209	193	228		200	203	205	201	200	221	203	232	209	210	205	211
▷ AA	6	5	5	6	9	6	7	7	8	5	5	6	9		8	7	7	6	5	4	5	5	5	4	6	6	5	5	6	5	6	4	5	7	5
▷ AC	1				2		1				1						1		2	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	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		23	24	24	23	27	25	24	28	21	24	27	24
▷ CD	3	2	2	2	2	2	2	2	2	2	4	2		2	2	2	2	2	2	2	2	2	2	2	2
▷ CE	1	2		1	3	2	2	1	1	1	1			2	1	1	1	11	2	2	2	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
▷ CO	1	2	1	1	1																					
▷ CP	1	1	1	1	1		1			1				1	1	1	1	1	1	1	1	1	1	1	1
▷ GA																										
▷ GB							1																			
▷ I-						4	4	4	4	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	4	5	4	4	4
▷ Kazal															1	1				1		1
▷ ID																										
▷ IL													1													
▷ IV																										
▷ JC					1	1	1	1	1	1	1	1	1	1	1											1	2	1
▷ JE					3	3	3	3	3	4	3	3	3	3	4											4	3	3

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	scaffold_3:818,001-819,568	1,568	401	● S08.115	● Subtilase family	
Ascim1 326276	scaffold_34:203,990-205,048	1,059	276	● S08.115		
Ascim1 211249	scaffold_57:190,517-189,456	1,062	306	● S08.115	● Subtilase family	
Ascni1 363694	scaffold_7:535,694-537,533	1,840	453	● S08.115	● Subtilase family ● Peptidase inhibitor I9	
Ascni1 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		

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.

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

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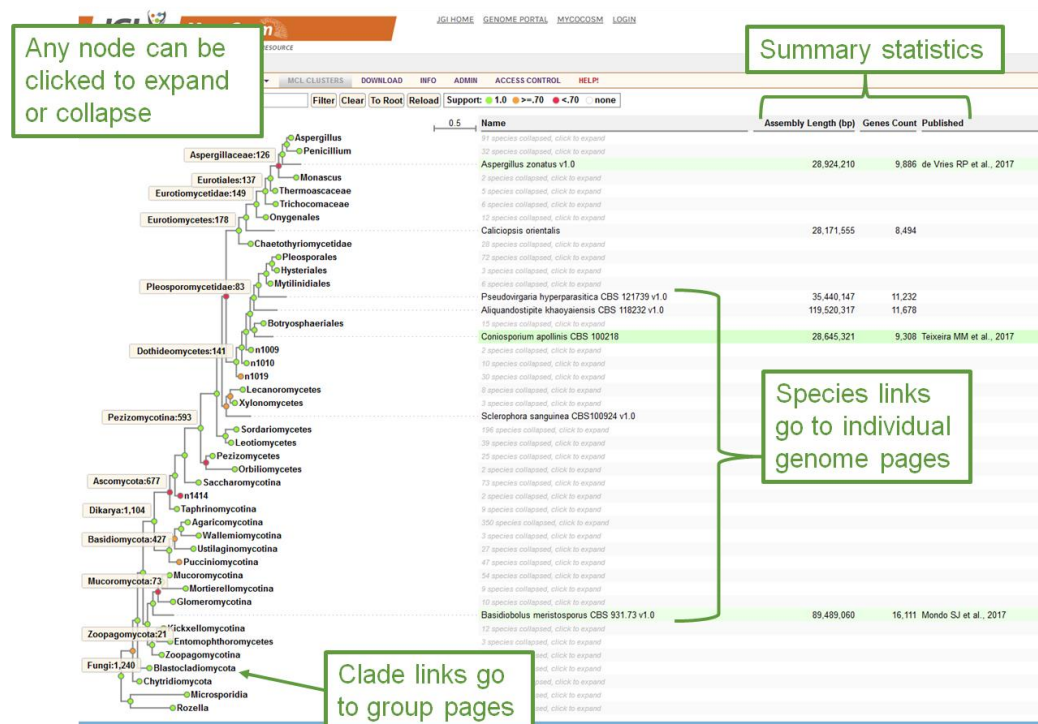
Let's take a look at a larger tree, the Eurotiomycetes. The main driver for the growth in Eurotiomycetes is JGI's Aspergillus Whole-Genus Sequencing Project, with the goal of sequencing ~300 species of Aspergillus.

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.

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). This genome has also been imported by FungiDB, and again this page has a link to the corresponding FungiDB page. Try it!

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.

MycoCosm
THE FUNGAL GENOMICS RESOURCE

Home | Outreach ▾ | **Video Tutorials ▾** | About ▾

MycoCosm: A comparative platform for fungal genomics
MycoCosm: Single genome tools
MycoCosm: Comparative Genomics Tools

MycoCosm Introduction (WebM)
MycoCosm Introduction (MP4)
MycoCosm Introduction (Theora)
MycoCosm Introduction (Old)
Genome Browser (WebM)
Genome Browser (MP4)
Genome Browser (Theora)
Genome Browser (Old)
Protein Page (WebM)
Protein Page (MP4)
Protein Page (Theora)
Protein Page (Old)
Functional Annotation Browsers: KOG/GO/KEGG (WebM)
Functional Annotation Browsers: KOG/GO/KEGG (MP4)
Functional Annotation Browsers: KOG/GO/KEGG (Theora)
Functional Annotation Browsers: KOG/GO/KEGG (Old)

Genes of Unknow
Claim to characteriz
1000 Fungal Geno
Nominate New Spec
Genomic Encyclop
Submit CSP proposa
Myco-Ed Genomic

Announcements

March 12-17, 2024
Fungal Genetics Confere
Pacific Grove, CA

June 2-7, 2024
Fungal Pathogen Genom
Hinxton, UK

June 9-13, 2024
Mycological Society of A
Markham, Canada

September 30-October 4, 2
2024 JGI Genomics of Er
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...

Search MycoCosm | **Search for JGI Data**
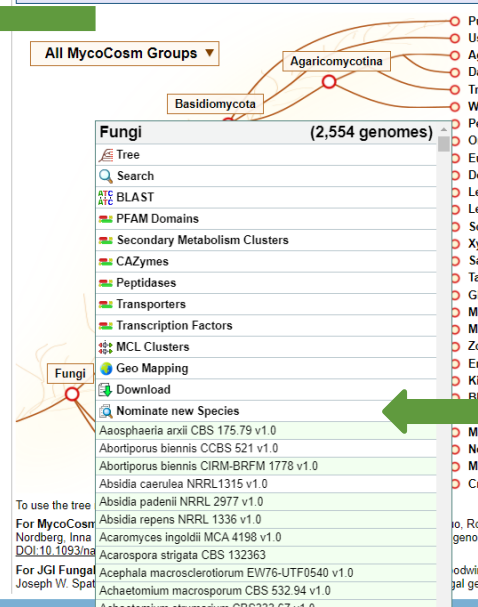
Agaricomycotina
- Pucciniomycotina
- Ustilaginomycotina
- Agaricomycetes
- Dacrymycetes
- Tremellomycetes
- Wallemiomycetes

Pezizomycotina
- Pezizomycetes
- Orbiliomycetes
- Eurotiomycetes
- Dothideomycetes
- Lecanoromycetes
- Leotiomycetes
- Sordariomycetes
- Xylonomycetes

Ascomycota
- Saccharomycotina
- Taphrinomycotina
- Glomeromycotina
- Mortierellomycotina
- Mucoromycotina
- Zoopagomycotina
- Entomophthoromycotina
- Kickxellomycotina
- Blastocladiomycota

Fungi
Chytridiomycota
- 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 Smirnova, Henrik Nordberg, Inna Dubchak, Igor Shabalov, MycoCosm portal: gearing up for 1000 fungal genomes, Nucleic Acids Research, Volume 42, Issue D1, 1 January 2014, Pages D699–D704, DOI:10.1093/nar/gkt1183

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

U.S. DEPARTMENT OF **ENERGY** | Office of Science