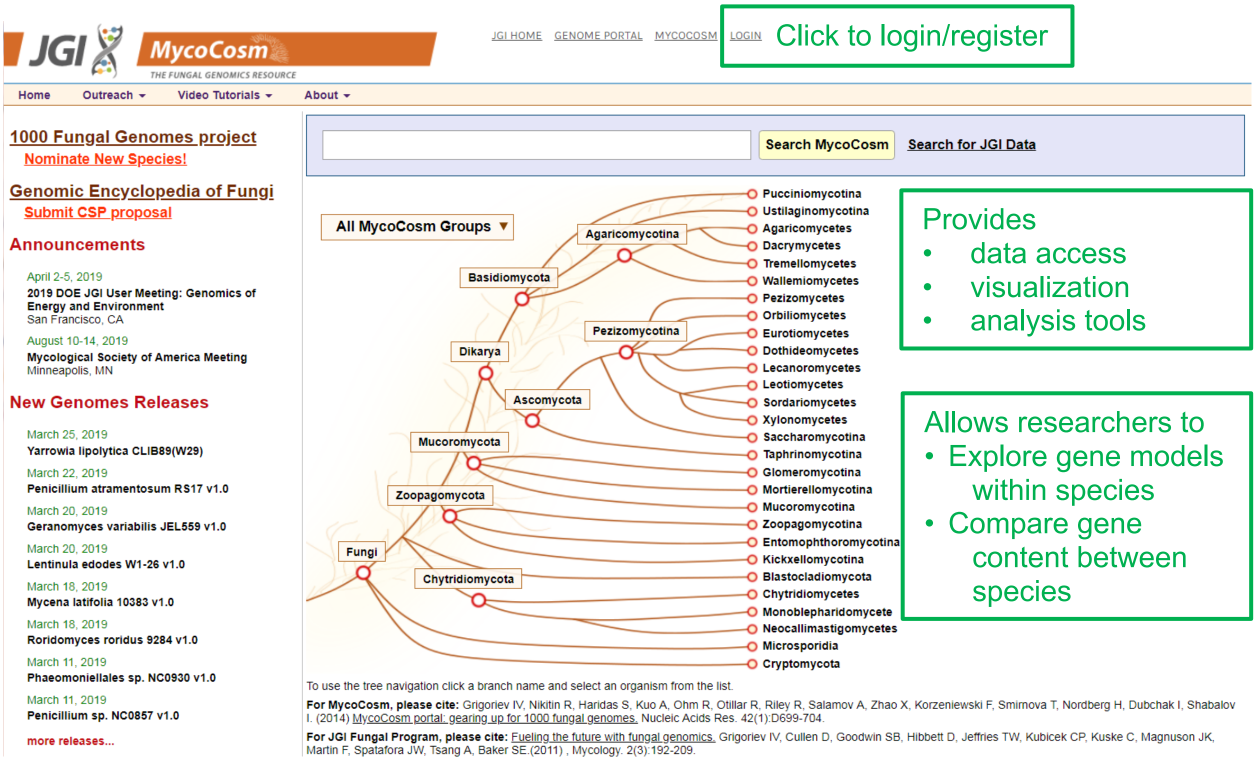
Introduction to MycoCosm

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/)



Click to login/register

Provides
- data access
- visualization
- analysis tools

Allows researchers to
- Explore gene models within species
- Compare gene content between species

1000 Fungal Genomes project
Nominate New Species!

Genomic Encyclopedia of Fungi
Submit CSP proposal

Announcements

April 2-5, 2019
2019 DOE JGI User Meeting: Genomics of Energy and Environment
San Francisco, CA

August 10-14, 2019
Mycological Society of America Meeting
Minneapolis, MN

New Genomes Releases

March 25, 2019
Yarrowia lipolytica CLIB89(W29)

March 22, 2019
Penicillium atramentosum RS17 v1.0

March 20, 2019
Geranomyces variabilis JEL559 v1.0

March 20, 2019
Lentinula edodes W1-26 v1.0

March 18, 2019
Mycena latifolia 10383 v1.0

March 18, 2019
Roridomyces roridus 9284 v1.0

March 11, 2019
Phaeomoniellales sp. NC0930 v1.0

March 11, 2019
Penicillium sp. NC0857 v1.0

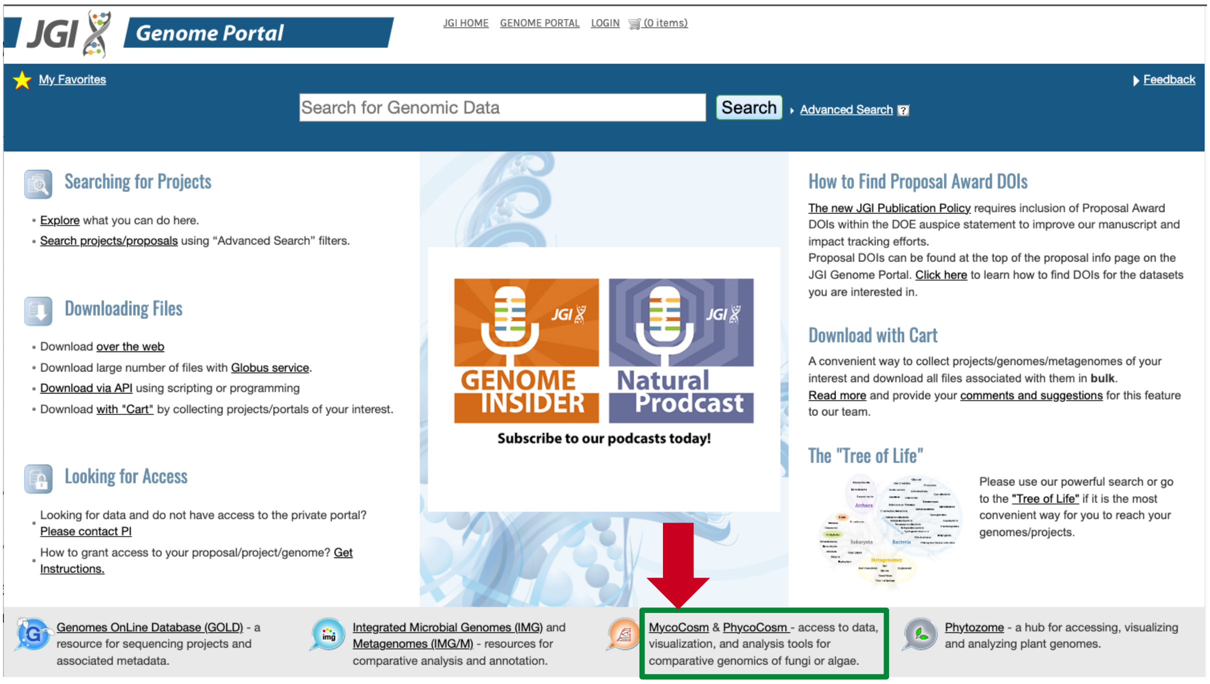
more releases...

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

For MycoCosm, please cite: Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. (2014) MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42(1):D699-704.

For JGI Fungal Program, please cite: Fueling the future with fungal genomics. Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, Kuske C, Magnuson JK, Martin F, Spatafora JW, Tsang A, Baker SE.(2011) , Mycology. 2(3):192-209.

You can also get to MycoCosm from the unified JGI genome portal (genome.jgi.doe.gov)



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.

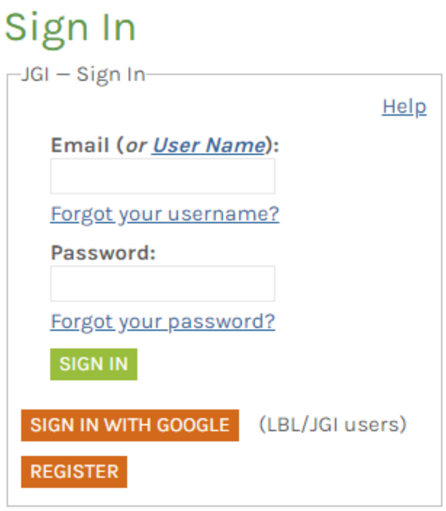
Genomes OnLine Database (GOLD) - a resource for sequencing projects and associated metadata.

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

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

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

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:

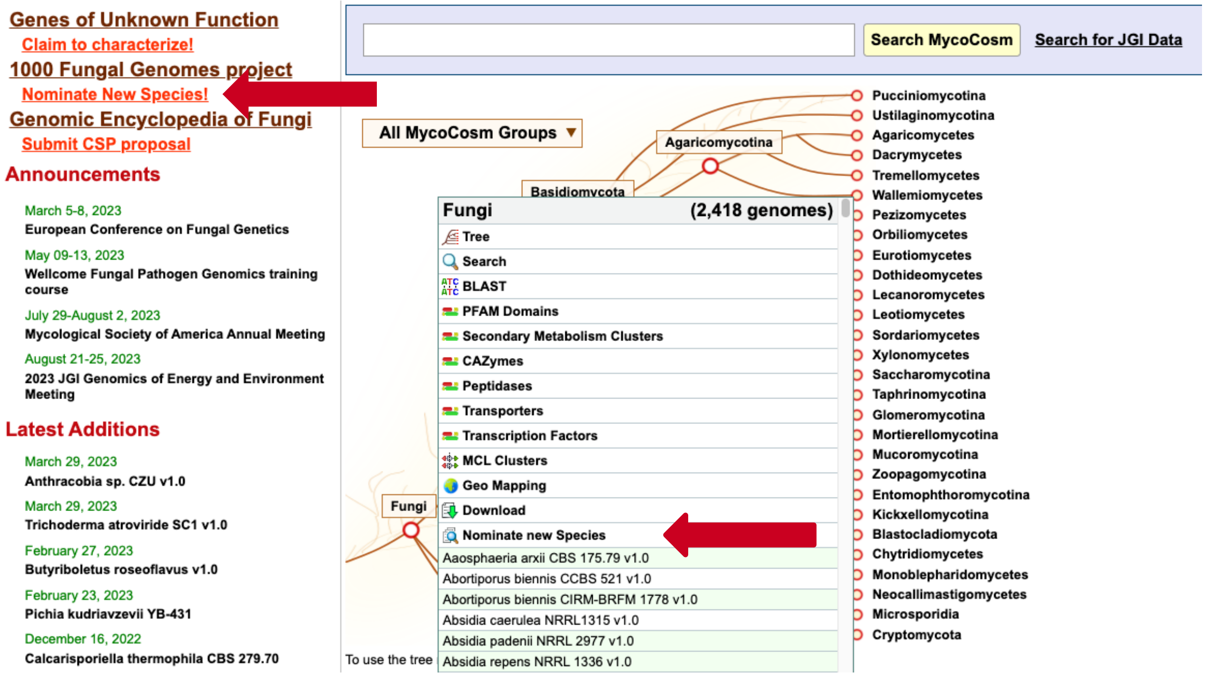
Forgot your password?

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!



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

Announcements

March 5-8, 2023
European Conference on Fungal Genetics

May 09-13, 2023
Wellcome Fungal Pathogen Genomics training course

July 29-August 2, 2023
Mycological Society of America Annual Meeting

August 21-25, 2023
2023 JGI Genomics of Energy and Environment Meeting

Latest Additions

March 29, 2023
Anthracobia sp. CZU v1.0

March 29, 2023
Trichoderma atroviride SC1 v1.0

February 27, 2023
Butyriboletus roseoflavus v1.0

February 23, 2023
Pichia kudriavzevii YB-431

December 16, 2022
Calcarisporiella thermophila CBS 279.70

Search MycoCosm | Search for JGI Data

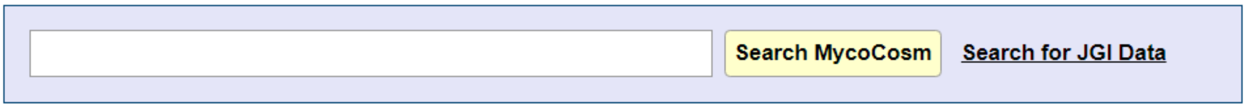
All MycoCosm Groups ▼

Fungi (2,418 genomes)
- Tree
- Search
- BLAST
- PFAM Domains
- Secondary Metabolism Clusters
- CAZymes
- Peptidases
- Transporters
- Transcription Factors
- MCL Clusters
- Geo Mapping
- Download
- Nominate new Species

Aaosphaeria arxii CBS 175.79 v1.0
Abortiporus biennis CCBS 521 v1.0
Abortiporus biennis CIRM-BRFM 1778 v1.0
Absidia caerulea NRRL1315 v1.0
Absidia padenii NRRL 2977 v1.0
Absidia repens NRRL 1336 v1.0

To use the tree

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



<u>Exercise:</u> Find 'Aspergillus' on MycoCosm using the default parameters.

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

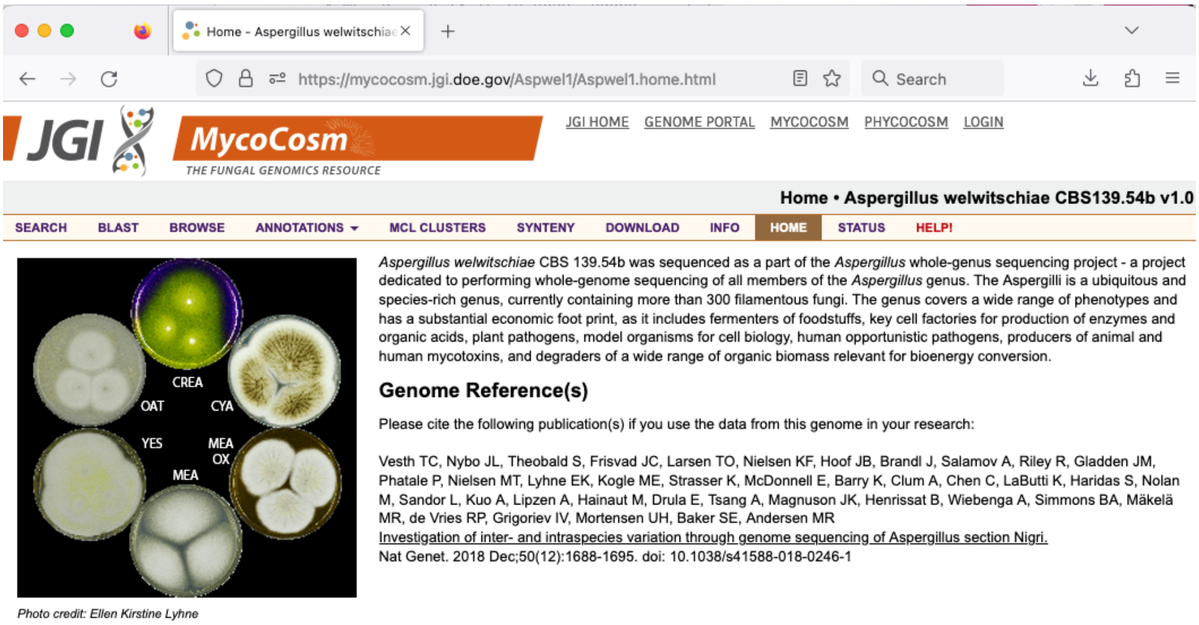
<u>Exercise:</u> Find 'Lichen' on genome.jgi.doe.gov using "Search for JGI Data". How does this compare to the default search on MycoCosm?

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. For example, 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.



Home • Aspergillus welwitschiae CBS139.54b v1.0

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)

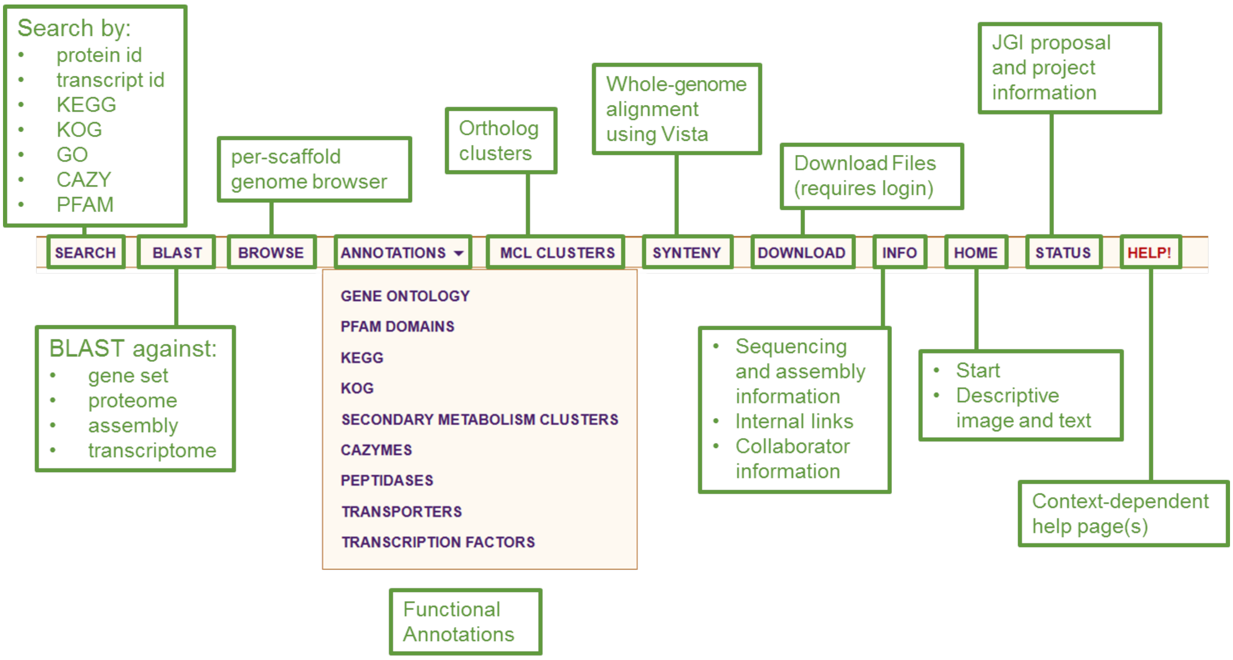
Please cite the following publication(s) if you use the data from this genome in your research:

Vesth 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, Kogle 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, Mäkelä MR, de Vries RP, Grigoriev IV, Mortensen UH, Baker SE, Andersen MR
<u>Investigation of inter- and intraspecies variation through genome sequencing of Aspergillus section Nigri.</u>
Nat Genet. 2018 Dec;50(12):1688-1695. doi: 10.1038/s41588-018-0246-1

Photo credit: Ellen Kirstine Lyhne

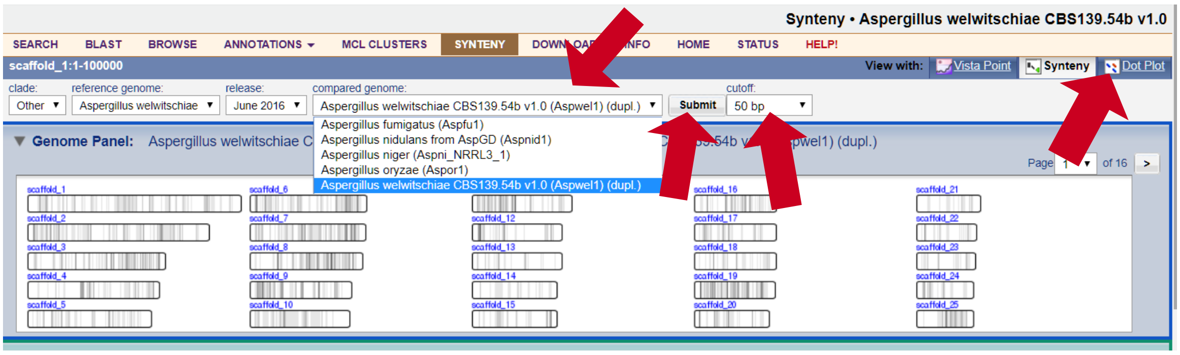
The navigation toolbar

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



Synteny

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.


Synteny • Aspergillus welwitschiae CBS139.54b v1.0

scaffold_1:1-100000

clade: Other
reference genome: Aspergillus welwitschiae
release: June 2016
compared genome: Aspergillus welwitschiae CBS139.54b v1.0 (Aspwel1) (dupl.)
cutoff: 50 bp

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.

Genome browser

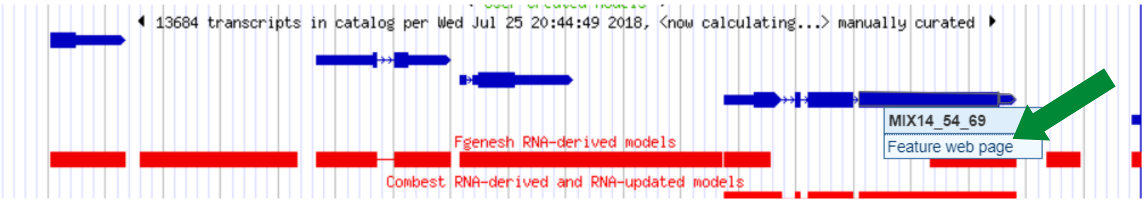
Now, we will explore the genome browser using the “BROWSE” tab. On the browser, you can navigate using available navigation tools and explore multiple layers of information we superimpose upon scaffolds. These include nucleotide conservation with closely related genomes, RNAseq expression data, Blastx 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).

scaffold_1:1-100000

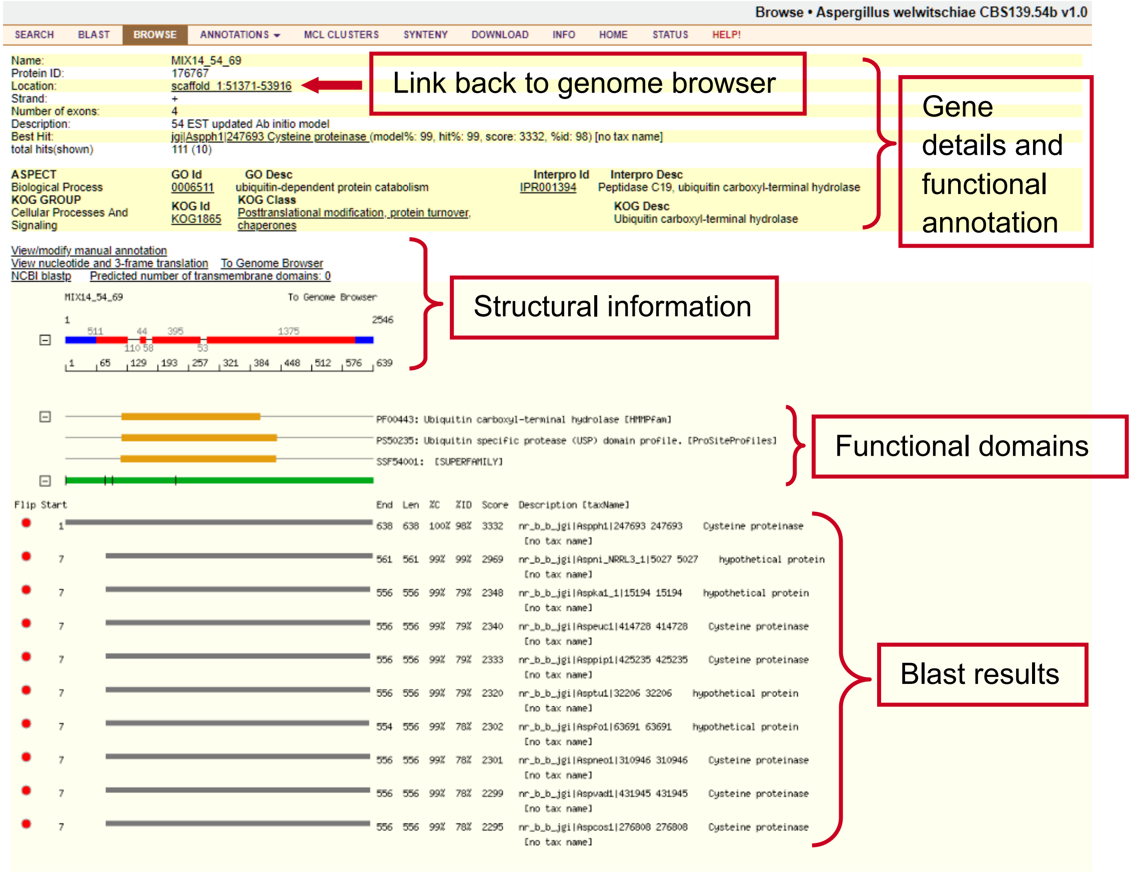
Size: 100000 Feature:

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.



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.



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|Aspph1|247693 Cysteine proteinase (model%: 99, hit%: 99, score: 3332, %id: 98) [no tax name]
total hits(shown) 111 (10)

ASPECT	GO Id	GO Desc	Interpro Id	Interpro Desc
Biological Process	0006511	ubiquitin-dependent protein catabolism	IPR001394	Peptidase C19, ubiquitin carboxyl-terminal hydrolase

KOG GROUP	KOG Id	KOG Class	KOG Desc
Cellular Processes And Signaling	KOG1865	Posttranslational modification, protein turnover, chaperones	Ubiquitin carboxyl-terminal hydrolase

View/modify manual annotation
View nucleotide and 3-frame translation To Genome Browser
NCBI blastp Predicted number of transmembrane domains: 0

Flip	Start	End	Len	%C	%ID	Score	Description [taxName]
●	1	638	638	100%	98%	3332	nr_b_b_jgi|Aspph1|247693 247693 Cysteine proteinase [no tax name]
●	7	561	561	99%	99%	2969	nr_b_b_jgi|Aspni_NRRL3_1|5027 5027 hypothetical protein [no tax name]
●	7	556	556	99%	79%	2348	nr_b_b_jgi|Aspka1_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|Asppip1|425235 425235 Cysteine proteinase [no tax name]
●	7	556	556	99%	79%	2320	nr_b_b_jgi|Asptu1|32206 32206 hypothetical protein [no tax name]
●	7	554	556	99%	78%	2302	nr_b_b_jgi|Aspfo1|63691 63691 hypothetical protein [no tax name]
●	7	556	556	99%	78%	2301	nr_b_b_jgi|Aspneo1|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|Aspcos1|276808 276808 Cysteine proteinase [no tax name]

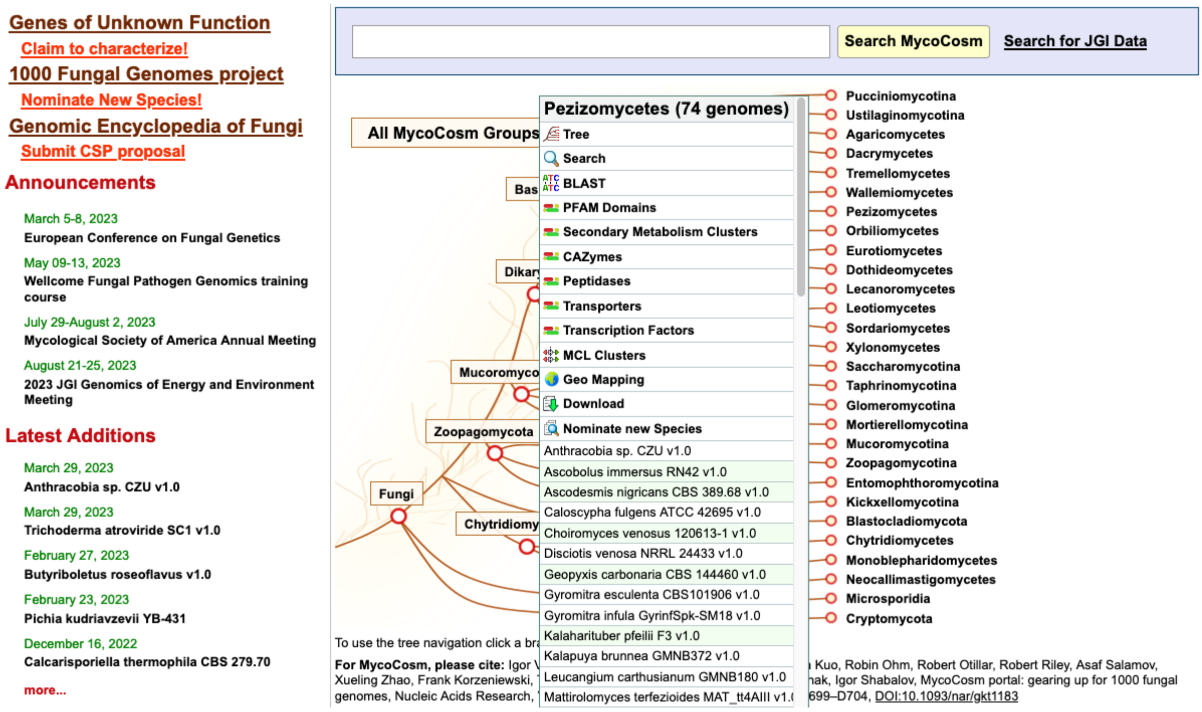
EXERCISES:

1. Go to the test portal for this course at mycocosm.jgi.doe.gov/Rhoto1_APA2687_1/
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 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.



Announcements

March 5-8, 2023
European Conference on Fungal Genetics

May 09-13, 2023
Wellcome Fungal Pathogen Genomics training course

July 29-August 2, 2023
Mycological Society of America Annual Meeting

August 21-25, 2023
2023 JGI Genomics of Energy and Environment Meeting

Latest Additions

March 29, 2023
Anthracobia sp. CZU v1.0

March 29, 2023
Trichoderma atroviride SC1 v1.0

February 27, 2023
Butyriboletus roseoflavus v1.0

February 23, 2023
Pichia kudriavzevii YB-431

December 16, 2022
Calcarisporiella thermophila CBS 279.70

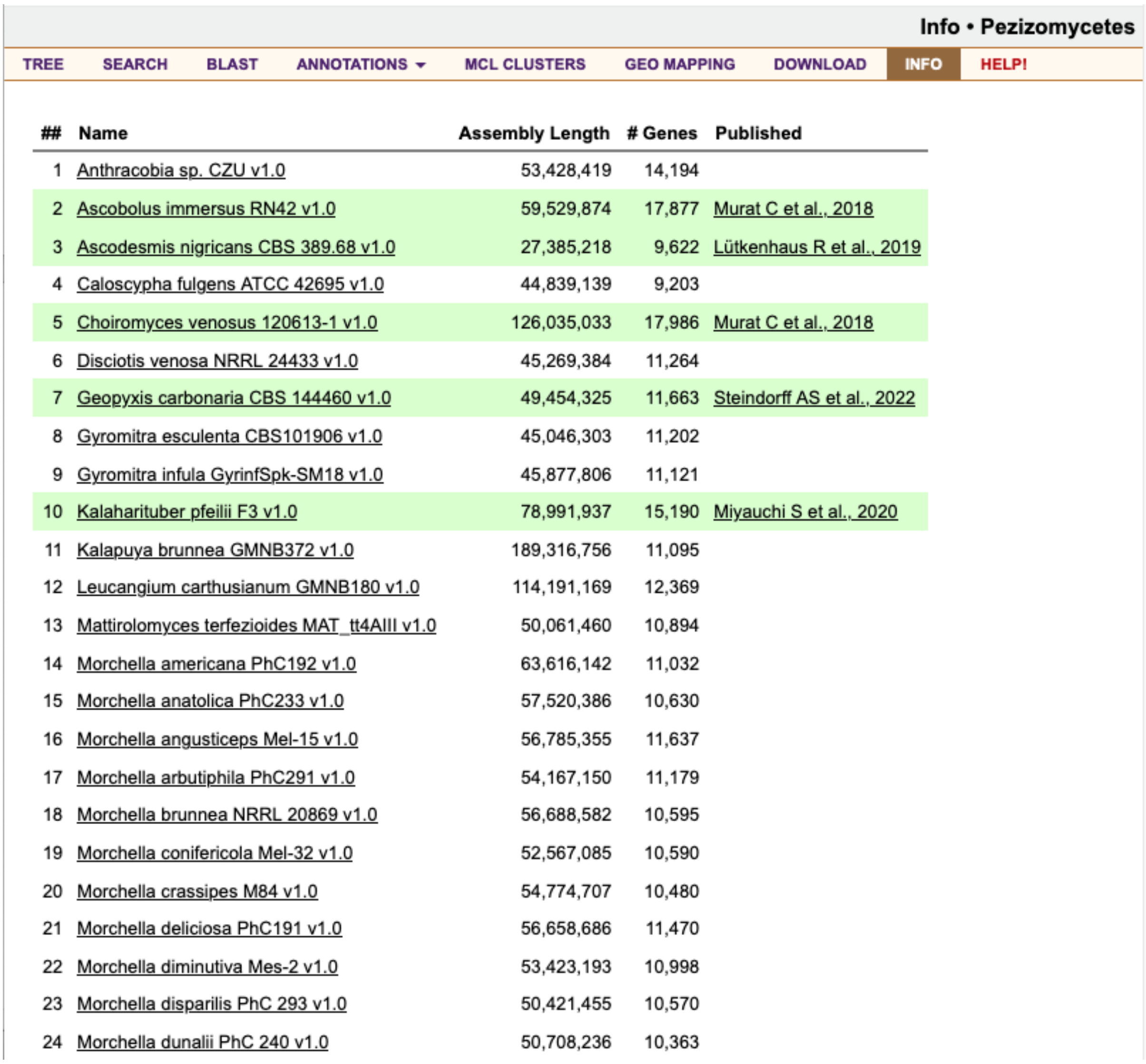
Anthracobia sp. CZU v1.0
Ascobolus immersus RN42 v1.0
Ascodesmis nigricans CBS 389.68 v1.0
Caloscypha fulgens ATCC 42695 v1.0
Choiromyces venosus 120613-1 v1.0
Disciotis venosa NRRL 24433 v1.0
Geopyxis carbonaria CBS 144460 v1.0
Gyromitra esculenta CBS101906 v1.0
Gyromitra infula GyrinfSpk-SM18 v1.0
Kalaharituber pfeilii F3 v1.0
Kalapuya brunnea GMNB372 v1.0
Leucangium carthusianum GMNB180 v1.0
Mattirolomyces terfezioides MAT_tt4AIII v1.

To use the tree navigation click a br

For MycoCosm, please cite: Igor V Kuo, Robin Ohm, Robert Otillar, Robert Riley, Asaf Salamov,
Xueling Zhao, Frank Korzeniewski, hak, Igor Shabalov, MycoCosm portal: gearing up for 1000 fungal
genomes, Nucleic Acids Research, 699–D704, DOI:10.1093/nar/gkt1183

Anatomy of a MycoCosm Group Portal

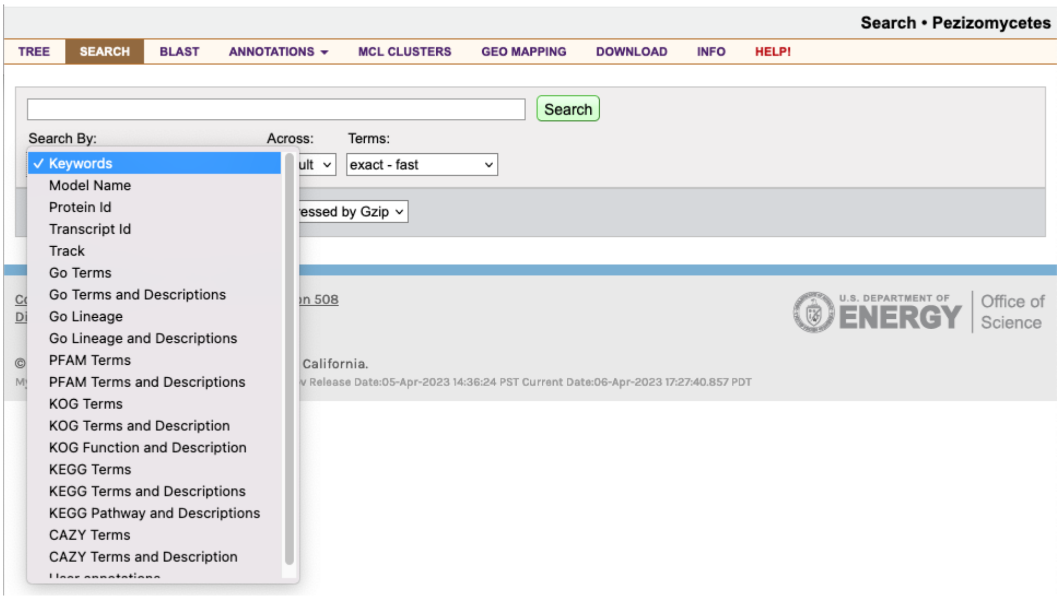
The default view of a Group Portal shows a hyperlinked list of all of the member genomes, basic statistics of each genome, and the primary associated publication, if any. Genome names are linked to their Genome Portals, and publications to PubMed. The Pezizomycetes Group currently consists of 74 genomes, 25 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.



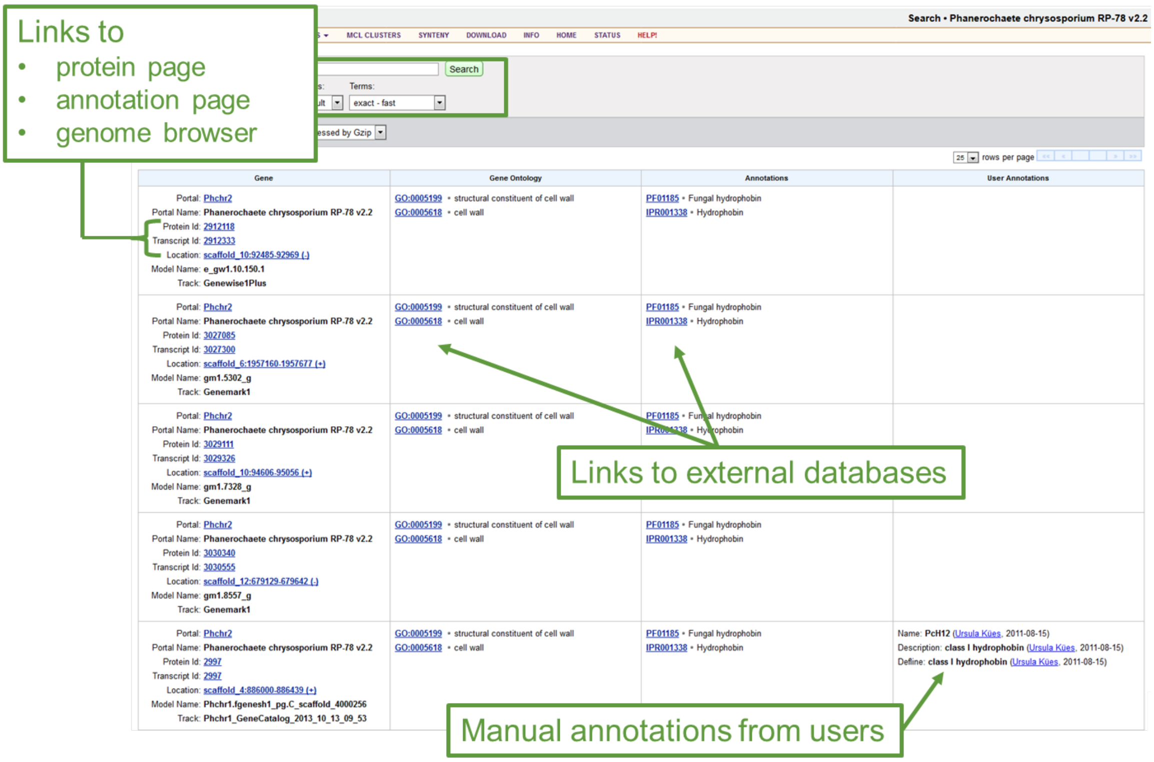
##	Name	Assembly Length	# Genes	Published
1	Anthracobia sp. CZU v1.0	53,428,419	14,194	
2	Ascobolus immersus RN42 v1.0	59,529,874	17,877	Murat C et al., 2018
3	Ascodesmis nigricans CBS 389.68 v1.0	27,385,218	9,622	Lütkenhaus R et al., 2019
4	Caloscypha fulgens ATCC 42695 v1.0	44,839,139	9,203	
5	Choiromyces 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 PhC192 v1.0	63,616,142	11,032	
15	Morchella anatolica PhC233 v1.0	57,520,386	10,630	
16	Morchella angusticeps Mel-15 v1.0	56,785,355	11,637	
17	Morchella arbutiphila PhC291 v1.0	54,167,150	11,179	
18	Morchella brunnea NRRL 20869 v1.0	56,688,582	10,595	
19	Morchella conifericola Mel-32 v1.0	52,567,085	10,590	
20	Morchella crassipes M84 v1.0	54,774,707	10,480	
21	Morchella deliciosa PhC191 v1.0	56,658,686	11,470	
22	Morchella diminutiva Mes-2 v1.0	53,423,193	10,998	
23	Morchella disparilis PhC 293 v1.0	50,421,455	10,570	
24	Morchella dunalii PhC 240 v1.0	50,708,236	10,363	

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.



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.



Gene	Gene Ontology	Annotations	User Annotations
Portal: Phchr2 Portal Name: Phanerochaete chrysosporium RP-78 v2.2 Protein Id: 2912118 Transcript Id: 2912333 Location: scaffold_10:92485-92969 (-) Model Name: e_gw1.10.150.1 Track: Genewise1Plus	GO:0005199 • structural constituent of cell wall GO:0005618 • cell wall	PF01185 • Fungal hydrophobin IPR001338 • Hydrophobin	
Portal: Phchr2 Portal Name: Phanerochaete chrysosporium RP-78 v2.2 Protein Id: 3027085 Transcript Id: 3027300 Location: scaffold_6:1957160-1957677 (+) Model Name: gm1.5302_g Track: Genemark1	GO:0005199 • structural constituent of cell wall GO:0005618 • cell wall	PF01185 • Fungal hydrophobin IPR001338 • Hydrophobin	
Portal: Phchr2 Portal Name: Phanerochaete chrysosporium RP-78 v2.2 Protein Id: 3029111 Transcript Id: 3029326 Location: scaffold_10:94606-95056 (+) Model Name: gm1.7328_g Track: Genemark1	GO:0005199 • structural constituent of cell wall GO:0005618 • cell wall	PF01185 • Fungal hydrophobin IPR001338 • Hydrophobin	
Portal: Phchr2 Portal Name: Phanerochaete chrysosporium RP-78 v2.2 Protein Id: 3030340 Transcript Id: 3030555 Location: scaffold_12:679129-679642 (-) Model Name: gm1.8557_g Track: Genemark1	GO:0005199 • structural constituent of cell wall GO:0005618 • cell wall	PF01185 • Fungal hydrophobin IPR001338 • Hydrophobin	
Portal: Phchr2 Portal Name: Phanerochaete chrysosporium RP-78 v2.2 Protein Id: 2997 Transcript Id: 2997 Location: scaffold_4:886000-886439 (+) Model Name: Phchr1.fgenesh1_pg.C_scaffold_4000256 Track: Phchr1_GeneCatalog_2013_10_13_09_53	GO:0005199 • structural constituent of cell wall GO:0005618 • cell wall	PF01185 • Fungal hydrophobin IPR001338 • Hydrophobin	Name: PcH12 (Ursula Kües, 2011-08-15) Description: class I hydrophobin (Ursula Kües, 2011-08-15) Define: class I hydrophobin (Ursula Kües, 2011-08-15)

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

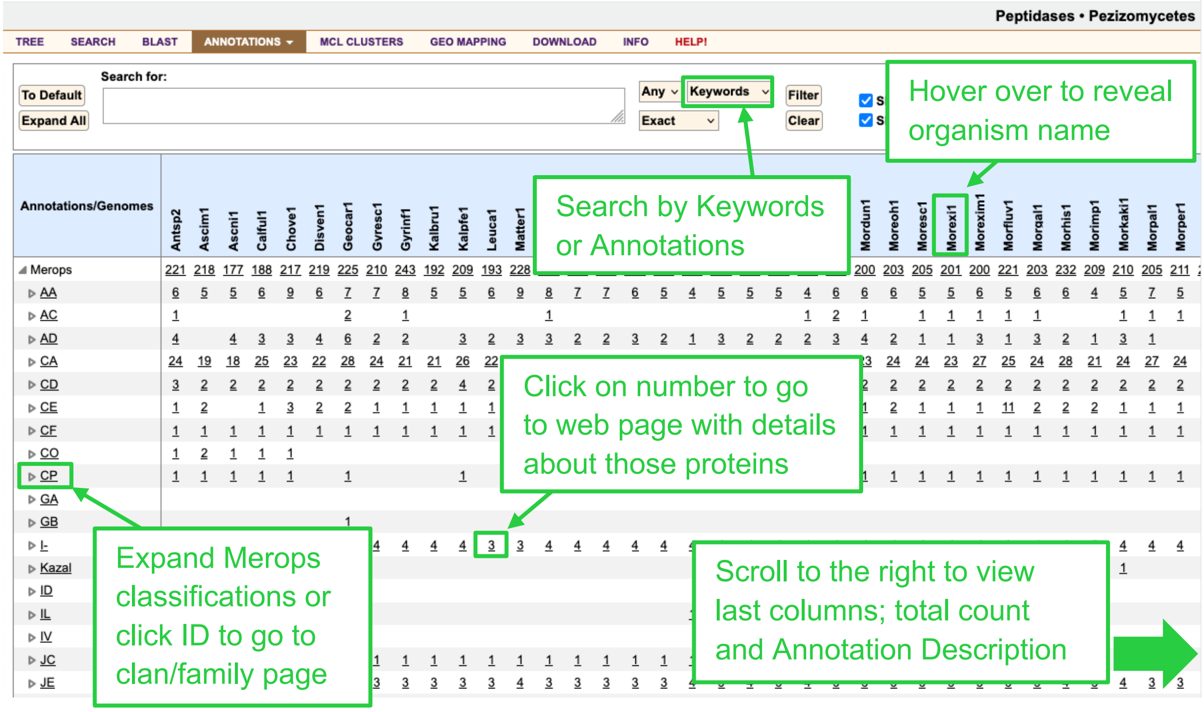
Exercise: Use the "Lichens" group identified in the previous exercise for this exercise (mycocosm.jgi.doe.gov/lichens). How many genes in these genomes have Pfam annotations?

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 investigate 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 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.



Hover over to reveal organism name

Search by Keywords or Annotations

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

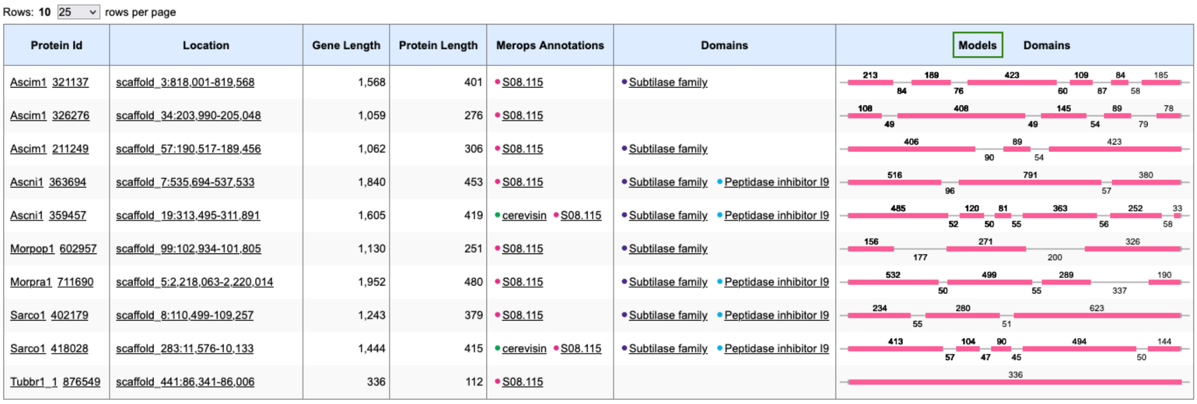
Expand Merops classifications or click ID to go to clan/family page

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

Annotations/Genomes	Antsp2	Ascim1	Ascni1	Calfu1	Chove1	Disven1	Geocar1	Gyresc1	Gyrinf1	Kalbru1	Kalpfe1	Leuca1	Matter1
Merops	221	218	177	188	217	219	225	210	243	192	209	193	228
AA	6	5	5	6	9	6	7	7	8	5	5	6	9
AC	1						2		1				
AD	4		4	3	3	4	6	2	2		3	2	3
CA	24	19	18	25	23	22	28	24	21	21	26	22	[illegible]
CD	3	2	2	2	2	2	2	2	2	2	4	2	[illegible]
CE	1	2		1	3	2	2	1	1	1	1	1	[illegible]
CF	1	1	1	1	1	1	1	1	1	1	1	1	[illegible]
CO	1	2	1	1	1								[illegible]
CP	1	1	1	1	1		1				1		[illegible]
GA													
GB							1						
I-	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4	4	4	4	3	3
Kazal	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]						
ID	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]						
IL	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]						
IV	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]						
JC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	1	1	1	1	1
JE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3	3	3	3	3	4

Annotations/Genomes	Mordun1	Moreoh1	Moresc1	Morexi1	Morexim1	Morfluv1	Morgal1	Morhis1	Morimp1	Morkaki1	Morpal1	Morper1
Merops	200	203	205	201	200	221	203	232	209	210	205	211
AA	6	6	5	5	6	5	6	6	4	5	7	5
AC	1		1	1	1	1	1			1	1	1
AD	4	2	1	1	3	1	3	2	1	3	1	
CA	[illegible]	24	24	23	27	25	24	28	21	24	27	24
CD	[illegible]	2	2	2	2	2	2	2	2	2	2	2
CE	[illegible]	2	1	1	1	11	2	2	2	1	1	1
CF	[illegible]	1	1	1	1	1	1	1	1	1	1	1
CO												
CP	[illegible]	1	1	1	1	1	1	1	1	1	1	1
GA												
GB												
I-	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4	4	4
Kazal	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1		
ID	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
IL	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
IV	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
JC	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
JE	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4	3	3

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.

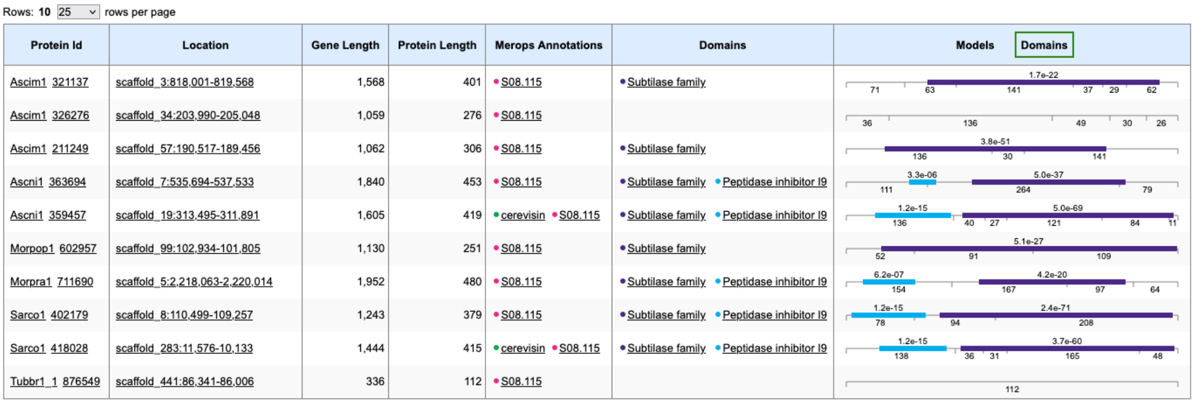


Rows: **10** 25 rows per page

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	213 84 189 76 423 60 109 87 84 58 185
Ascim1 326276	scaffold_34:203,990-205,048	1,059	276	• S08.115		108 49 408 49 145 54 89 79 78
Ascim1 211249	scaffold_57:190,517-189,456	1,062	306	• S08.115	• Subtilase family	406 90 89 54 423
Ascni1 363694	scaffold_7:535,694-537,533	1,840	453	• S08.115	• Subtilase family • Peptidase inhibitor I9	516 96 791 57 380
Ascni1 359457	scaffold_19:313,495-311,891	1,605	419	• cerevisin • S08.115	• Subtilase family • Peptidase inhibitor I9	485 52 120 50 81 55 363 56 252 58 33
Morpop1 602957	scaffold_99:102,934-101,805	1,130	251	• S08.115	• Subtilase family	156 177 271 200 326
Morpra1 711690	scaffold_5:2,218,063-2,220,014	1,952	480	• S08.115	• Subtilase family • Peptidase inhibitor I9	532 50 499 55 289 337 190
Sarco1 402179	scaffold_8:110,499-109,257	1,243	379	• S08.115	• Subtilase family • Peptidase inhibitor I9	234 55 280 51 623
Sarco1 418028	scaffold_283:11,576-10,133	1,444	415	• cerevisin • S08.115	• Subtilase family • Peptidase inhibitor I9	413 57 104 47 90 45 494 50 144
Tubbr1_1 876549	scaffold_441:86,341-86,006	336	112	• S08.115		336

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.

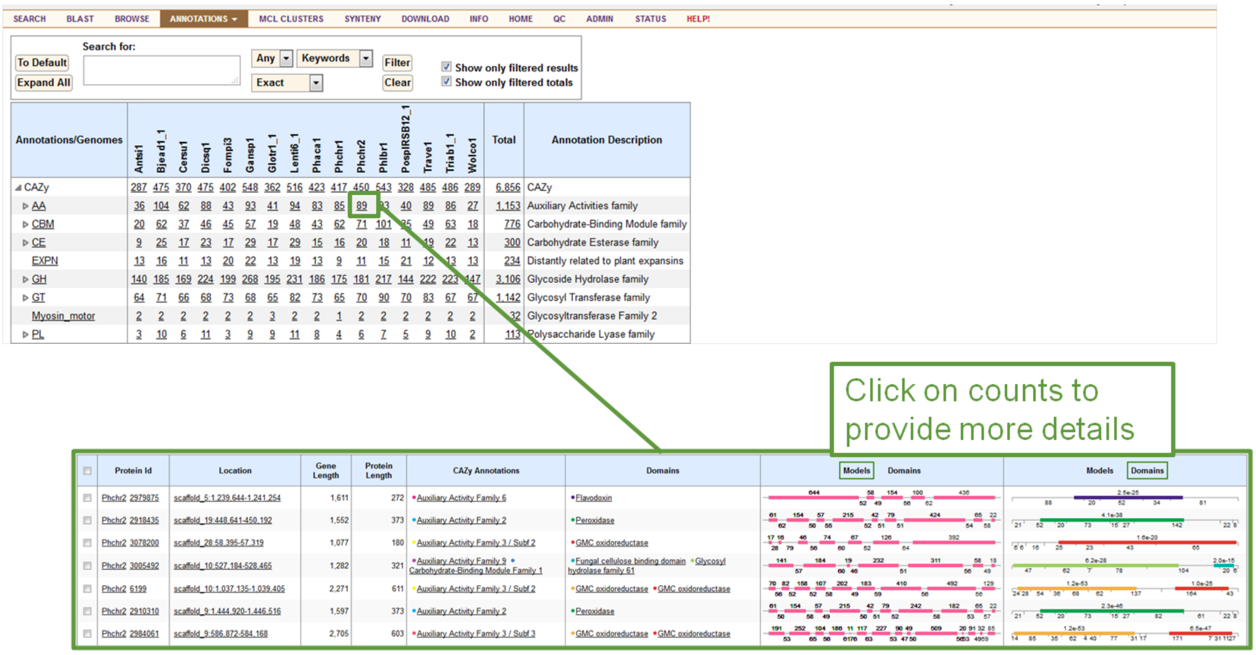
Click on ‘Domains’ to view the domain structure of the proteins. As is often the case with subtilases, the gene product is a pre-protein with a peptidase inhibitor domain that likely must be cleaved off before the mature proteinase can function as such.



Rows: **10** 25 rows per page

Protein Id	Location	Gene Length	Protein Length	Merops Annotations	Domains	Models **Domains**
Ascim1 321137	scaffold_3:818,001-819,568	1,568	401	• S08.115	• Subtilase family	71 63 141 1.7e-22 37 29 62
Ascim1 326276	scaffold_34:203,990-205,048	1,059	276	• S08.115		36 136 49 30 26
Ascim1 211249	scaffold_57:190,517-189,456	1,062	306	• S08.115	• Subtilase family	136 3.8e-51 30 141
Ascni1 363694	scaffold_7:535,694-537,533	1,840	453	• S08.115	• Subtilase family • Peptidase inhibitor I9	111 3.3e-06 264 5.0e-37 79
Ascni1 359457	scaffold_19:313,495-311,891	1,605	419	• cerevisin • S08.115	• Subtilase family • Peptidase inhibitor I9	136 1.2e-15 40 27 121 5.0e-69 84 11
Morpop1 602957	scaffold_99:102,934-101,805	1,130	251	• S08.115	• Subtilase family	52 91 5.1e-27 109
Morpra1 711690	scaffold_5:2,218,063-2,220,014	1,952	480	• S08.115	• Subtilase family • Peptidase inhibitor I9	6.2e-07 154 167 4.2e-20 97 64
Sarco1 402179	scaffold_8:110,499-109,257	1,243	379	• S08.115	• Subtilase family • Peptidase inhibitor I9	78 1.2e-15 94 2.4e-71 208
Sarco1 418028	scaffold_283:11,576-10,133	1,444	415	• cerevisin • S08.115	• Subtilase family • Peptidase inhibitor I9	138 1.2e-15 36 31 3.7e-60 165 48
Tubbr1_1 876549	scaffold_441:86,341-86,006	336	112	• S08.115		112

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



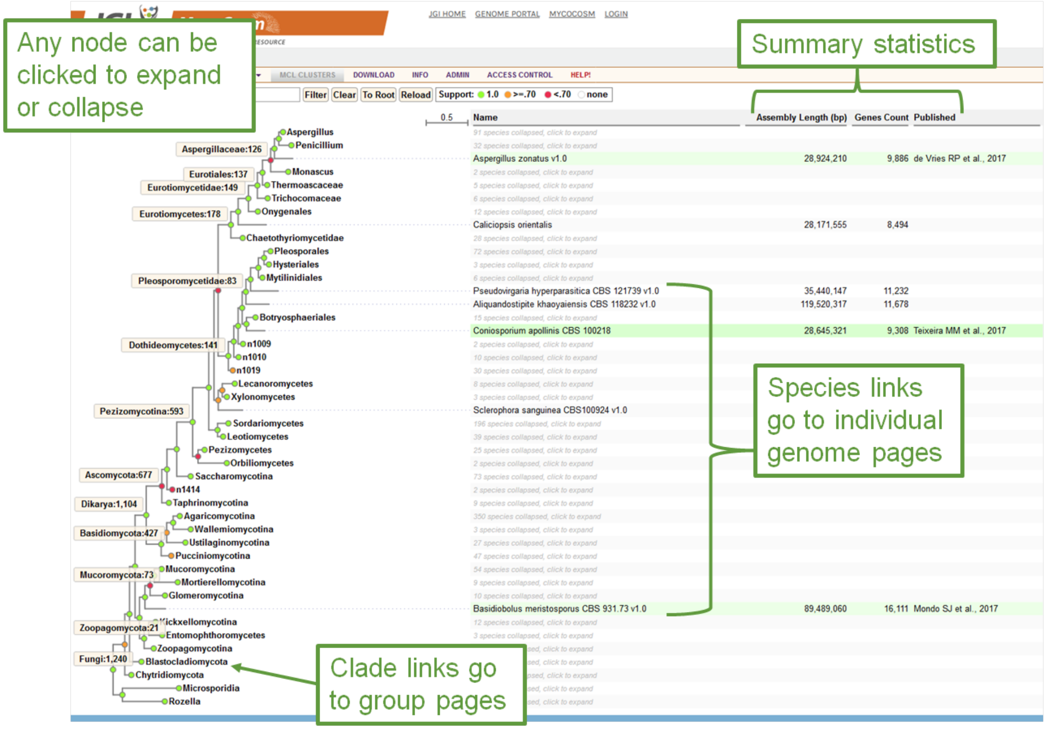
Annotations/Genomes	Antsi1	Bjead1_1	Cersu1	Dicsq1	Fompi3	Gansp1	Glotr1_1	Lenti6_1	Phaca1	Phchr1	Phchr2	Phlbr1	PosplRSB12_1	Trave1	Triab1_1	Wolco1	Total	Annotation Description
CAZy	287	475	370	475	402	548	362	516	423	417	450	543	328	485	486	289	6,856	CAZy
AA	36	104	62	88	43	93	41	94	83	85	89	93	40	89	86	27	1,153	Auxiliary Activities family
CBM	20	62	37	46	45	57	19	48	43	62	71	101	35	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
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	90	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
☐	Phchr2 2979875	scaffold_5:1,239,644-1,241,254	1,611	272	• Auxiliary Activity Family 6	• Flavodoxin
☐	Phchr2 2918435	scaffold_19:448,641-450,192	1,552	373	• Auxiliary Activity Family 2	• Peroxidase
☐	Phchr2 3078200	scaffold_28:58,395-57,319	1,077	180	• Auxiliary Activity Family 3 / Subf 2	• GMC oxidoreductase
☐	Phchr2 3005492	scaffold_10:527,184-528,465	1,282	321	• Auxiliary Activity Family 9 • Carbohydrate-Binding Module Family 1	• Fungal cellulose binding domain • Glycosyl hydrolase family 61
☐	Phchr2 6199	scaffold_10:1,037,135-1,039,405	2,271	611	• Auxiliary Activity Family 3 / Subf 2	• GMC oxidoreductase • GMC oxidoreductase
☐	Phchr2 2910310	scaffold_9:1,444,920-1,446,516	1,597	373	• Auxiliary Activity Family 2	• Peroxidase
☐	Phchr2 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:

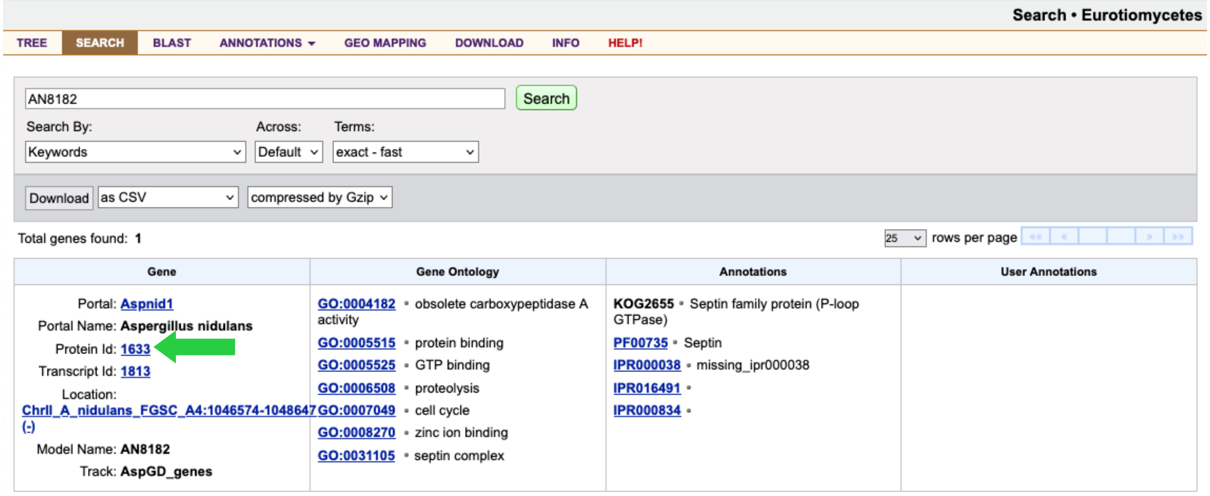
Unlike the cartoon on the MycoCosm home page, this tree, and others associated with each Group, is calculated by JGI using the clusters mentioned above, followed by MAFFT alignment of presumptive orthologs and FastTree modeling. Sometimes unexpected branchings appear that contradict other resources such 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.



Name	Assembly Length (bp)	Genes Count	Published
91 species collapsed, click to expand			
32 species collapsed, click to expand			
Aspergillus zonatus v1.0	28,924,210	9,886	de Vries RP et al., 2017
2 species collapsed, click to expand			
5 species collapsed, click to expand			
6 species collapsed, click to expand			
12 species collapsed, click to expand			
Caliciopsis orientalis	28,171,555	8,494	
28 species collapsed, click to expand			
72 species collapsed, click to expand			
3 species collapsed, click to expand			
6 species collapsed, click to expand			
Pseudovirgaria hyperparasitica CBS 121739 v1.0	35,440,147	11,232	
Aliquandostipite khaoyaiensis CBS 118232 v1.0	119,520,317	11,678	
15 species collapsed, click to expand			
Coniosporium apollinis CBS 100218	28,645,321	9,308	Teixeira MM et al., 2017
2 species collapsed, click to expand			
10 species collapsed, click to expand			
30 species collapsed, click to expand			
8 species collapsed, click to expand			
3 species collapsed, click to expand			
Sclerophora sanguinea CBS100924 v1.0			
196 species collapsed, click to expand			
39 species collapsed, click to expand			
25 species collapsed, click to expand			
2 species collapsed, click to expand			
73 species collapsed, click to expand			
2 species collapsed, click to expand			
9 species collapsed, click to expand			
350 species collapsed, click to expand			
3 species collapsed, click to expand			
27 species collapsed, click to expand			
47 species collapsed, click to expand			
54 species collapsed, click to expand			
9 species collapsed, click to expand			
10 species collapsed, click to expand			
Basidiobolus meristosporus CBS 931.73 v1.0	89,489,060	16,111	Mondo SJ et al., 2017
12 species collapsed, click to expand			
3 species collapsed, click to expand			

Let's take a look at a larger tree, the Eurotiomycetes, with 414 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 414 genomes to load. Next enter the known *Aspergillus nidulans* septin GeneID AN8182 into the Search box. We get a single gene, as expected.



AN8182

Search By: Keywords

Across: Default

Terms: exact - fast

as CSV

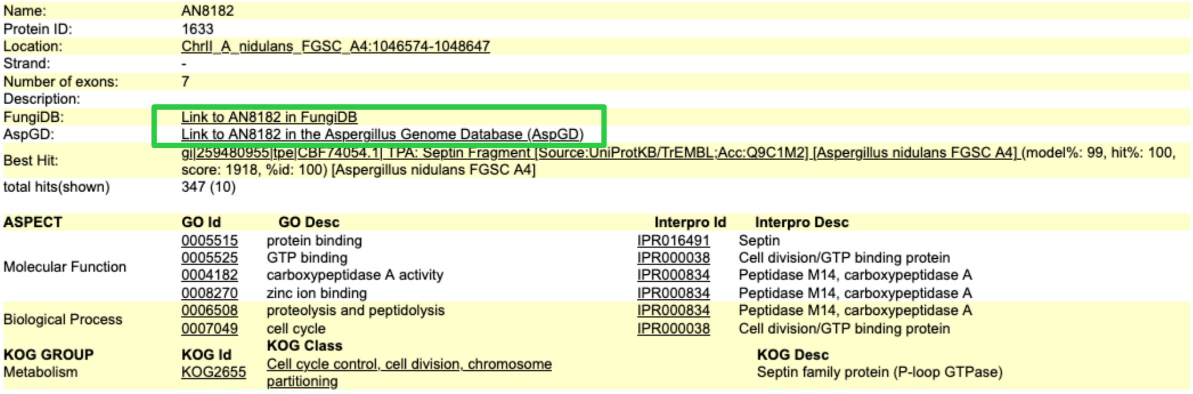
compressed by Gzip

Total genes found: **1**

25 rows per page

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

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), 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!

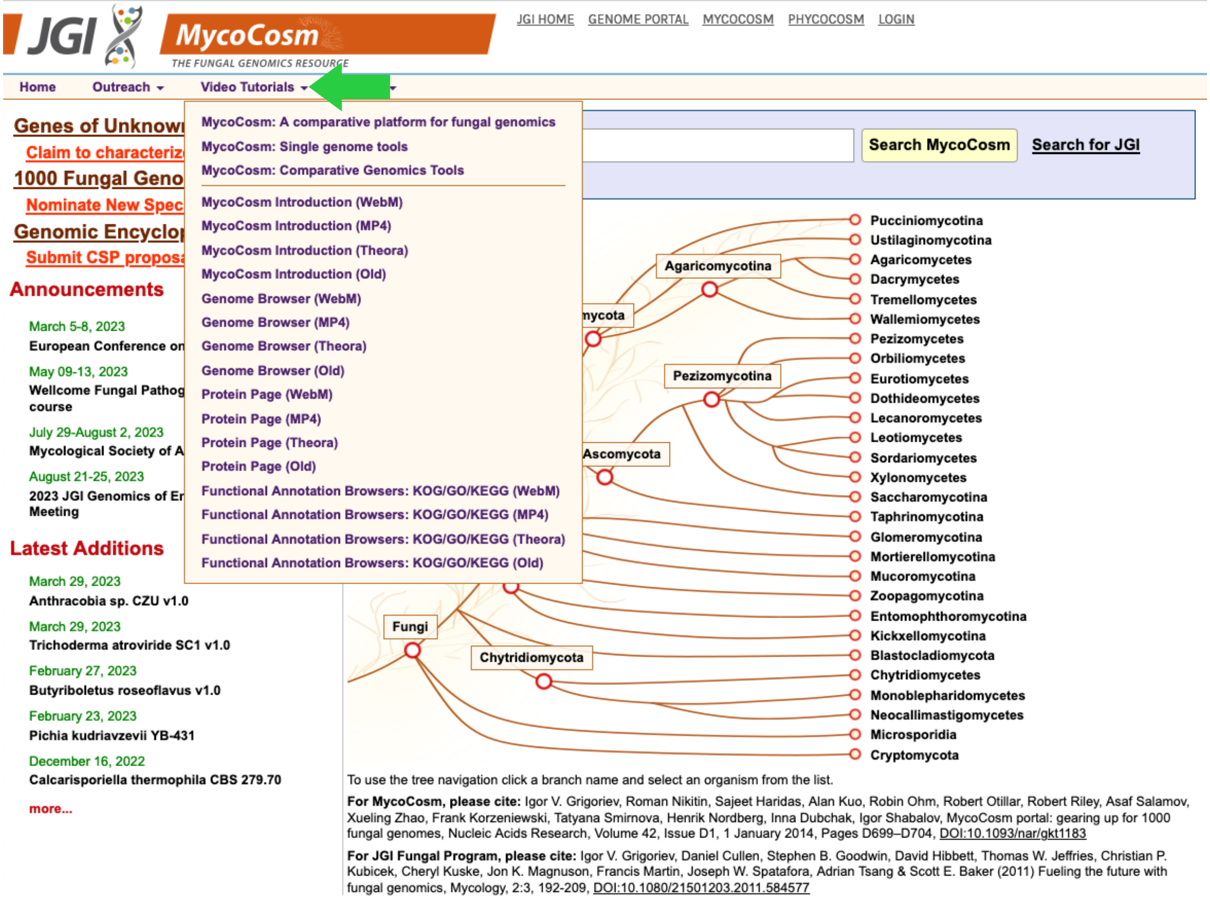


Name:	AN8182
Protein ID:	1633
Location:	ChrII_A_nidulans_FGSC_A4:1046574-1048647
Strand:	-
Number of exons:	7
Description:	
FungiDB:	Link to AN8182 in FungiDB
AspGD:	Link to AN8182 in the Aspergillus Genome Database (AspGD)
Best Hit:	gi|259480955|tpe|CBF74054.1| TPA: Septin Fragment [Source:UniProtKB/TrEMBL;Acc:Q9C1M2] [Aspergillus nidulans FGSC A4] (model%: 99, hit%: 100, score: 1918, %id: 100) [Aspergillus nidulans FGSC A4]
total hits(shown)	347 (10)

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

KOG GROUP	**KOG Id**	**KOG Class**	**KOG Desc**
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.



Genes of Unknow
Claim to characteriz
1000 Fungal Geno
Nominate New Spec
Genomic Encyclo
Submit CSP propos

Announcements

March 5-8, 2023
European Conference on

May 09-13, 2023
Wellcome Fungal Pathog
course

July 29-August 2, 2023
Mycological Society of A

August 21-25, 2023
2023 JGI Genomics of E
Meeting

Latest Additions

March 29, 2023
Anthracobia sp. CZU v1.0

March 29, 2023
Trichoderma atroviride SC1 v1.0

February 27, 2023
Butyriboletus roseoflavus v1.0

February 23, 2023
Pichia kudriavzevii YB-431

December 16, 2022
Calcarisporiella thermophila CBS 279.70

more...

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