

2023 Fungal Pathogen Genomics

Pre-workshop module

Anatomy of homepages, Gene Pages & Genome Browsers

Table of Contents

Resource	Section	Page
<i>SGD</i>	Anatomy of the home page Gene pages	2 4
<i>CGD</i>	Anatomy of the home page Gene pages	12 14
<i>SGD & CGD</i>	Genome browser	22
<i>Ensembl Fungi</i>	Anatomy of the home page Gene pages and genome browser	27 29
<i>FungiDB</i>	Anatomy of the home page Exploring gene record pages Exploring records in JBrowse	51 52 61
<i>JGI MycoCosm</i>	Anatomy of MycoCosm	65

Anatomy of the *Saccharomyces* Genome Database (SGD) Homepage

The *Saccharomyces* Genome Database (**SGD**, <https://www.yeastgenome.org>) is the primary genetics and genomics resource for the budding yeast *Saccharomyces cerevisiae*. SGD provides public access to the complete *S. cerevisiae* genome sequence along with high-quality, manually curated information about budding yeast genes and gene products. SGD extracts and integrates multiple forms of data, such as gene function, mutant phenotypes, genetics & physical interactions, regulation, expression, post-translational modifications, and much more. These data are easily accessible through an easy-to-use search tool and intuitive gene summary pages, and are also integrated into a variety of bioinformatics tools to facilitate analysis, discovery, and experimental design.

The SGD website is designed for ease-of-use. The homepage (<https://www.yeastgenome.org>) can be broken down into three main components:

1. Main toolbar

The screenshot shows the SGD homepage. A red arrow points from the text '1. Main toolbar' to the top navigation bar. The navigation bar includes links for 'About', 'Blog', 'Download', 'Help', 'YeastMine', and social media icons for Twitter, Facebook, and LinkedIn. Below the navigation bar is a search box with the placeholder 'search: actin, kinase, glucose'. The main content area features a video thumbnail titled 'Rap1-GFP and Calcofluor White staining of stationary phase cells.' followed by a descriptive text about SGD and a 'Try this?' button.

2. Search box

The screenshot continues from the previous section. It shows the SGD homepage with the search box highlighted. To the right, there is an 'Outreach & Media' section featuring a 'Meetings' sidebar with events like the '31st VHYC Yeast Conference' and 'Fungal Pathogen Genomics'. Below this is a 'New & Noteworthy' section with articles such as 'Trouble with Triplets - April 06, 2018' and 'Apply Now for the 2018 Yeast Genetics and Genomics Course - March 16, 2018'. On the far right is a 'Tweets' sidebar from the @yeastgenome account, showing tweets about yeast research and outreach.

3. Outreach & Media

1. The **main toolbar** organizes SGD's key tools and resources into the five categories shown above. It is accessible from essentially all pages at SGD.
2. The **search box** is located on the right side of the main toolbar. It enables searching for key types of data, such as genes, phenotypes, Gene Ontology (GO) terms, references, and more. Hitting Enter/Return after typing in your query will take you to a results page of top hits, where you can filter your results. Typing into the search box without hitting Enter/Return populates a list of suggested hits.

3. The **outreach & media** section lists important meetings within the yeast community, posts on the SGD blog (<https://www.yeastgenome.org/blog>), and recent updates on the SGD Twitter page (<https://twitter.com/yeastgenome>).

Another notable feature of the SGD homepage is the collection of links located directly above the search box. These links provide quick access to SGD's vast help documentation and helpdesk email (sgd-helpdesk@lists.stanford.edu), as well as links to SGD's social media pages. The "Download" link takes you to a page where you can download SGD's curated data, the *S. cerevisiae* reference genome sequence, published datasets, and more. Lastly, the "YeastMine" link refers you to YeastMine, a powerful search & retrieval tool that can be used for even more sophisticated querying of *S. cerevisiae* data.

Searching SGD and Navigating Gene Pages

Use SGD's Faceted Search and Locus Summary page to explore gene-specific information about RER2.

- Find a gene involved in protein glycosylation.
- Open the SGD homepage (<https://yeastgenome.org>). Search for “glycosylation”.
- From the categories (**facets**) in the left column, select the category **Genes**. This filters the results to genes that have the keyword “glycosylation” somewhere in their summaries and annotations.

The screenshot shows the SGD homepage with a search bar containing "glycosylation". The left sidebar has a facet for "Genes" highlighted with an orange arrow. The main search results show 2,012 results for "glycosylation", with the first result being "glycosylation" itself, defined as the covalent attachment of carbohydrate residues to a substrate molecule. Below it is "protein glycosylation", defined as a protein modification process involving the addition of carbohydrate units to proteins. A link to "61 Associated Genes" is shown at the bottom.

- Select more facets to further filter your results. Since we're interested in protein glycosylation, find **Biological Process** in the left column and select **protein glycosylation (direct)**.

The screenshot shows the SGD Locus Summary page for RER2. It includes a facet for "Biological Process" highlighted with an orange arrow, showing results for "protein glycosylation" and "protein glycosylation (direct)". The main content area displays detailed information for three genes: ALG8 / YOR067C, ALG1 / YBR110W, and ALG13 / YGL047W, each with a "Gene" icon. The summary for ALG8 mentions its role in N-linked glycosylation and its similarity to Alg6p. The summary for ALG1 highlights its role in asparagine-linked glycosylation and its localization to the endoplasmic reticulum. The summary for ALG13 discusses its catalytic component in UDP-GlcNAc transferase.

- This filters for genes that are directly annotated to “protein glycosylation”. Terms without the “(direct)” suffix are annotated to either “protein glycosylation” or a more specific term, such as “protein N-linked glycosylation”.
- Let’s also filter for a specific enzymatic activity. Under **Molecular Function**, click on “Show more”. Find the term **dehydrodolichyl diphosphate synthase activity (direct)**.

calcium ion binding (direct)	3	ALG6 / YO
catalytic activity (direct)	3	Alpha 1,3 glucosidase; asparagine residue
dehydrodolichyl diphosphate synthase activity (direct)	3	name description ; molecular function; glycosylation /group; summary: ALG6; glycosylation
hydrolase activity, acting on glycosyl bonds (direct)	3	

- The results should now show 3 genes that have the following:
 - The keyword “glycosylation” somewhere in their summaries and annotations
 - A direct annotation to “protein glycosylation”
 - A direct annotation to “dehydrodolichyl diphosphate synthase activity”
- To see only the gene names (useful for many results) as shown in the figure, click on the **Wrapped** button above the list. The **Download** and **Analyze** buttons respectively allow you to save the list locally or send it to one of SGD’s tools for analysis. For now, click on RER2 to open its **Locus Summary page**.

3 results for **x "glycosylation"** **x protein glycosylation (direct)** **x Gene**
x dehydrodolichyl diphosphate synthase activity (direct)

Download **Analyze** **List** **Wrapped**

Genetic loci that are not mapped to the genome sequence will be excluded from the analysis list.

	NUS1	SRT1	RER2
transferase activity	3		
transferase activity (direct)	3		
transferase activity, transferring alkyl or aryl (other than methyl) groups (direct)	3		

Explore *S. cerevisiae* RER2 Locus Summary page.

You can scroll down and up the page, or you can jump to a specific section using the content table in upper left corner. Full pages for each category of data can be accessed via the top gray toolbar.

RER2 / YBR002C Overview

Standard Name: RER2¹
Systematic Name: YBR002C
SGD ID: SGD:S000000206
Feature Type: ORF, Verified
Description: Forms the dehydrodolichyl diphosphate synthase (DDS) complex with NUS1; major enzyme of polypropenol synthesis in both the endoplasmic reticulum (ER) and in lipid droplets; participates in ER protein sorting; human ortholog DHDDS functionally complements the heat sensitive growth defect of a ts allele, and is associated with retinitis pigmentosa^{2 3 4 5}
Name Description: Retention in the Endoplasmic Reticulum¹
Comparative Info: Integrated model organism details available at the [Alliance of Genome Resources](#) website

Sequence **Sequence Details**

- **Summaries:** What is known about this gene? (read the Description in Locus Overview, read summaries in Gene Ontology and Phenotype sections, and read the Summary Paragraph)

RER2 / YBR002C Gene Ontology i

[Gene Ontology Help ?](#)

Summary: Forms a dehydrodolichyl diphosphate synthase complex with NUS1; involved in dolichol biosynthesis and ER to Golgi vesicle-mediated transport

GO Slim Terms i: endomembrane system, transferase activity, Golgi vesicle transport, carbohydrate metabolic process, lipid metabolic process, protein glycosylation

RER2 / YBR002C Phenotype i

[Phenotype Help ?](#)

Summary: Non-essential gene; reduction of function causes abnormal ER, Golgi and vacuolar morphology and mislocalization of membrane proteins; null mutation results in severe growth defect

- **Gene Ontology:** Explore functional annotations on RER2 by visiting the Gene Ontology tab. What **biological processes** is RER2 involved in? Does Rer2p have any known **molecular function**, such as kinase activity? What **cellular components** does Rer2p localize to in the cell, and is it a member of any complexes?

Summary	Sequence	Protein	Gene Ontology	Phenotype	Disease	Interactions	Regulation	Expression	Literature
---------	----------	---------	---------------	-----------	---------	--------------	------------	------------	------------

RER2 / YBR002C
Gene Ontology Overview
 Manually Curated
 High-throughput
 Computational
 Shared Annotations

RER2 / YBR002C Gene Ontology i

[Gene Ontology Help ?](#)

Summary: Forms a dehydrodolichyl diphosphate synthase complex with NUS1; involved in dolichol biosynthesis and ER to Golgi vesicle-mediated transport

GO Slim Terms i: endomembrane system, transferase activity, Golgi vesicle transport, carbohydrate metabolic process, lipid metabolic process, protein glycosylation

[Download All Annotations \(.txt\)](#)

Manually Curated i

Date Last Reviewed: 2007-03-12

Biological Process 5 entries for 3 Gene Ontology terms

Qualifier	Gene Ontology Term	Annotation Extension	Evidence	Source	Assigned On	Reference
	protein glycosylation		IDA	SGD	2002-03-07	Sato M, et al. (1999) PMID:9858571

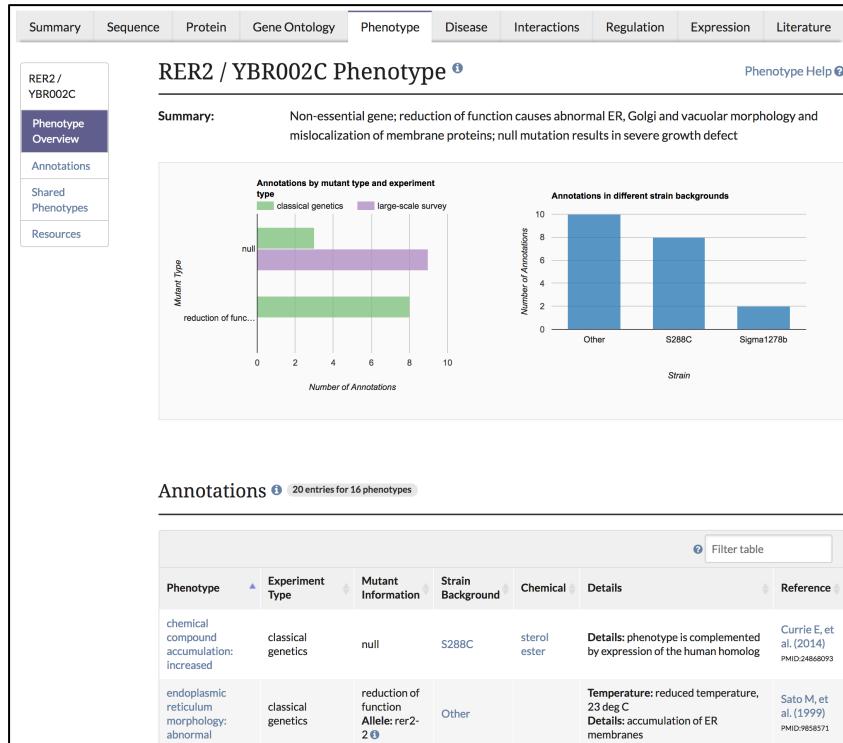
Current Locus Other Locus GO Term

Filter Genes by # of GO Terms Shared With RER2: 5 6 7 8 9 10 11 12

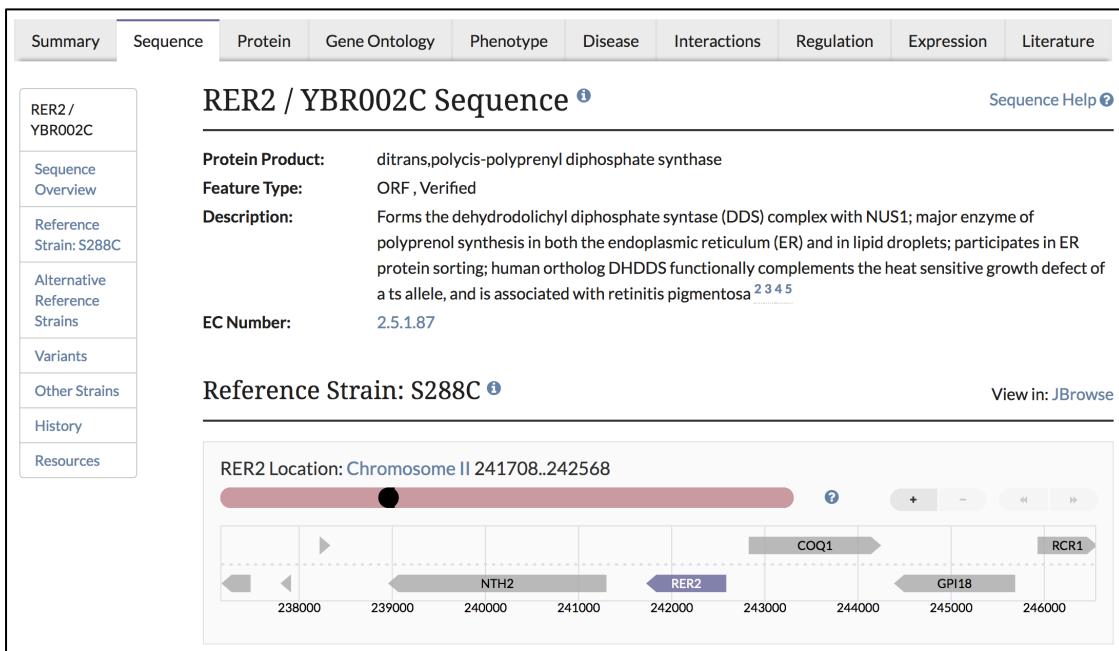
SGD 2019-03

- Scroll down the page and use the Shared Annotations diagram to find other genes that share the same biological processes.

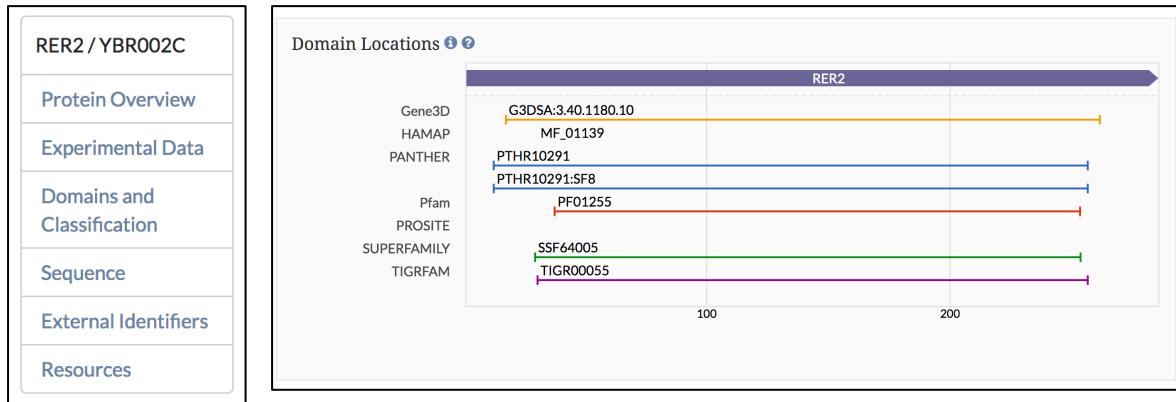
- **Phenotypes:** What details about the mutant phenotypes are available? See the Phenotype tab for information on mutant types, strain backgrounds, references. Based on the role of RER2 in ER to Golgi vesicle-mediated transport, do null mutants have phenotypes you would expect? Find other genes that share the same phenotypes by exploring the Network Diagram at the bottom of the page.



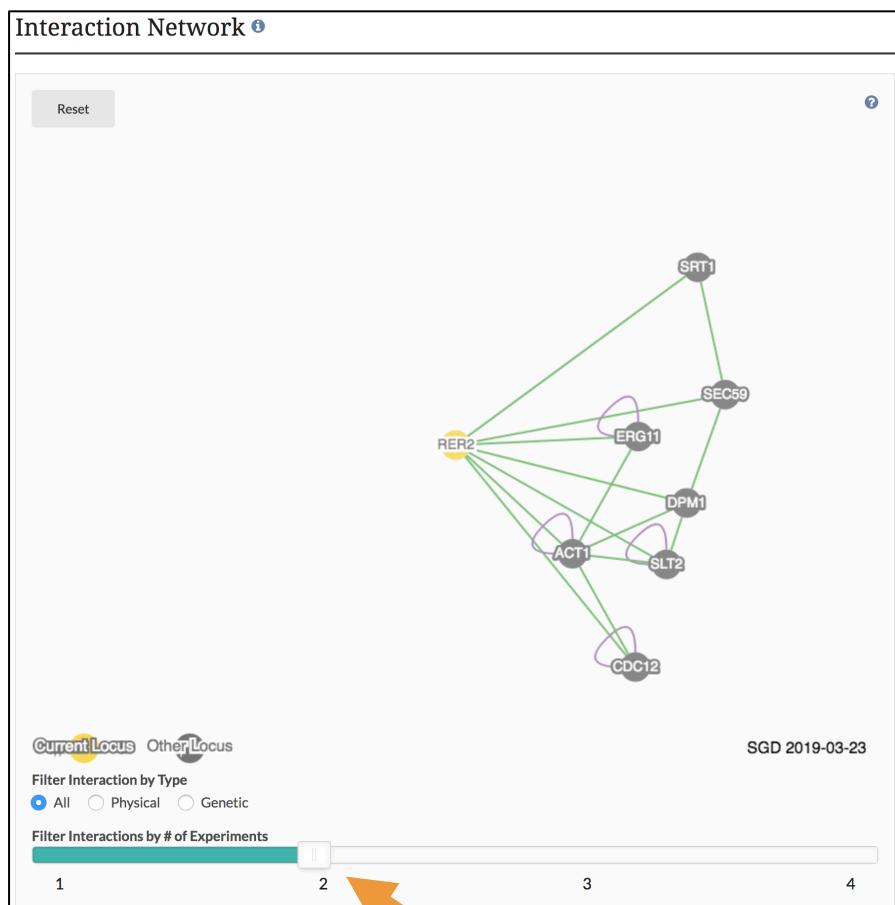
- **Sequence:** Visit the Sequence tab for RER2. What is the chromosomal location of RER2 and its neighboring genes? Note that the RER2 sequence can be downloaded here for the reference strain S288C and alternative strains.



- **Protein:** Visit the RER2 Protein tab. What is the Rer2p amino acid sequence? What is its half-life? What is the highest and lowest protein abundance listed for Rer2p? What protein domains does it have, and with which proteins does it share these domains? Is Rer2p post-translationally modified by ubiquitin? What is the calculated molecular weight and isoelectric point of this protein?



- **Interactions:** Go to the RER2 Interactions tab and look at the Annotations table. With which genes does RER2 have a genetic interaction? What about synthetic lethal interactions (hint: search the table for “synthetic lethal”)? Find the Interaction Network and set the # of experiments to 2 (see figure). Do any genetic interactors of RER2 also have a genetic interaction with each other?



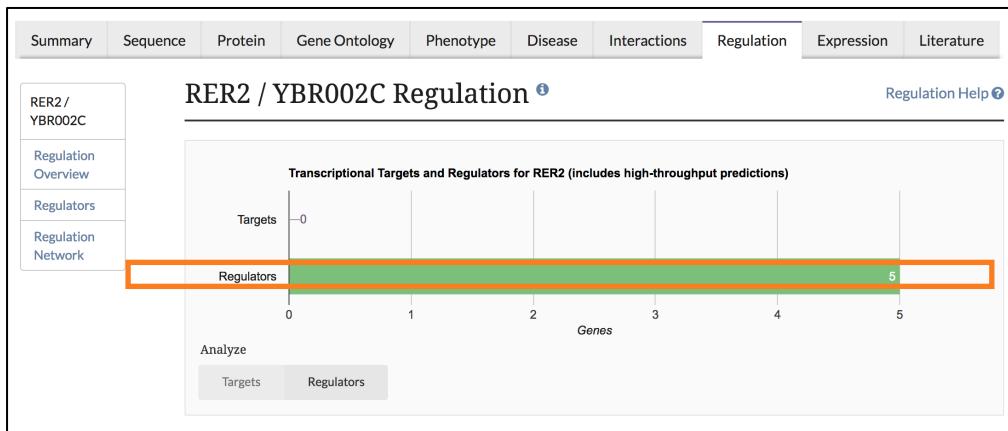
- **Disease:** Has yeast RER2 been used to study any human diseases? On the Disease tab, scroll to the bottom of the page and find the Shared Annotations network diagram. What other yeast genes have been used to study cancer? Do they have a human homolog?

Disease Ontology Term	Qualifier	Evidence	Source	Assigned On	Reference
cancer	ISS with DHDDS	SGD	2018-04-25	Hamza A, et al. (2015) PMID:26354769	
cancer	IGI with DHDDS	SGD	2018-04-25	Hamza A, et al. (2015) PMID:26354769	

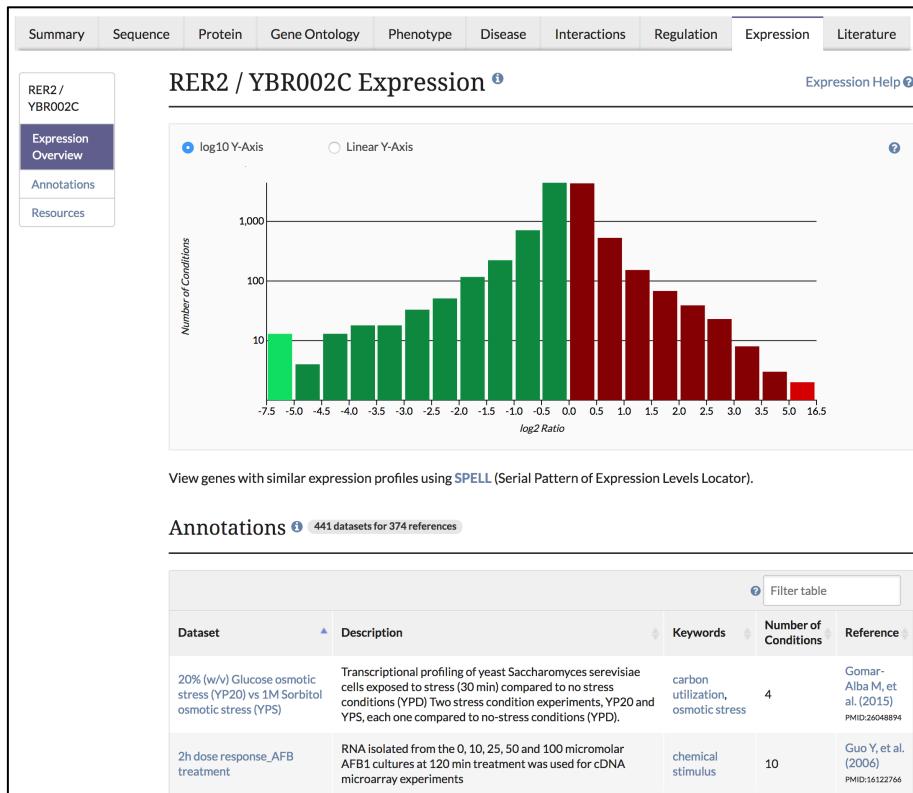
- **Homology:** Open the RER2 Homology tab. What are the homologous genes in other model organisms? Does the human homolog functionally complement the yeast mutant? What are the homologous genes in other fungi? Where can you find more information about these other fungal genes?

Species	Gene ID	Gene name	Source
<i>Caenorhabditis elegans</i>	WB:WBGene00044025	T01G1.4	Alliance
<i>Danio rerio</i>	ZFIN:ZDB-GENE-040426-2236	dhdds	Alliance
<i>Drosophila melanogaster</i>	FB:FBgn0029980	CG10778	Alliance
<i>Homo sapiens</i>	HGNC:20603	DHDDS	Alliance
<i>Mus musculus</i>	MGI:1914672	Dhdds	Alliance
<i>Rattus norvegicus</i>	RGD:13111560	Dhdds	Alliance

- **Regulation:** Open the RER2 Regulation tab. What regulatory relationships does RER2 have? Do any RER2 regulators regulate another (hint: see network diagram)?



- **Expression:** Go to the RER2 Expression tab. What factors affect the expression of RER2? The columns in the histogram indicate how many conditions result in a given increase/decrease in expression level of RER2 – click on a column to show the datasets, categories and references in the table below; hyperlinks lead to more details.



- **Literature:** Open the RER2 Literature tab. What reviews have been published that deal with RER2? Jump through the page using the content table.

RER2 / YBR002C Literature Literature Help 

Unique References: 56

Primary Literature 20 references 

[!\[\]\(fb581a674d20e658b53cf34efaf73ce7_img.jpg\) Download References \(nbib\)](#)

Sun S, et al. (2016) An extended set of yeast-based functional assays accurately identifies human disease mutations. *Genome Res* 26(5):670-80 PMID:26975778
[SGD Paper](#) [DOI full text](#) [PMC full text](#) [PubMed](#)

Hamza A, et al. (2015) Complementation of Yeast Genes with Human Genes as an Experimental Platform for Functional Testing of Human Genetic Variants. *Genetics* 201(3):1263-74 PMID:26354769
[SGD Paper](#) [DOI full text](#) [PMC full text](#) [PubMed](#)

Surmacz L, et al. (2015) Short-chain polyisoprenoids in the yeast *Saccharomyces cerevisiae* - New companions of the old guys. *Biochim Biophys Acta* 1851(10):1296-303 PMID:26143379
[SGD Paper](#) [DOI full text](#) [PubMed](#)

Currie E, et al. (2014) High confidence proteomic analysis of yeast LDs identifies additional droplet proteins and reveals connections to dolichol synthesis and sterol acetylation. *J Lipid Res* 55(7):1465-77 PMID:24868093
[SGD Paper](#) [DOI full text](#) [PMC full text](#) [PubMed](#)

Park EJ, et al. (2014) Mutation of Nogo-B receptor, a subunit of cis-prenyltransferase, causes a congenital disorder of glycosylation. *Cell Metab* 20(3):448-57 PMID:25066056
[SGD Paper](#) [DOI full text](#) [PMC full text](#) [PubMed](#)

Akhtar TA, et al. (2013) The tomato cis-prenyltransferase gene family. *Plant J* 73(4):640-52 PMID:23134568
[SGD Paper](#) [DOI full text](#) [PubMed](#)

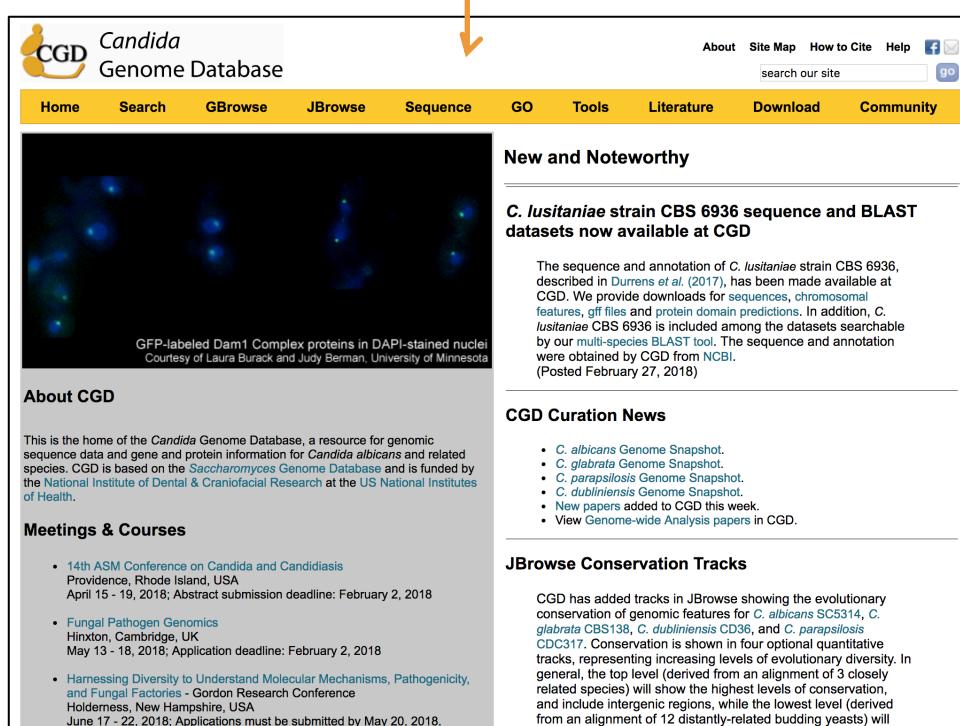


Anatomy of the *Candida* Genome Database (CGD) Homepage

The *Candida* Genome Database (CGD, <http://www.candidagenome.org>) is a freely available online resource that provides gene, protein and sequence information for multiple *Candida* species, along with web-based tools for accessing, analyzing and exploring these data. CGD contains information about genes and proteins; descriptions and classifications of their biological roles, molecular functions, and subcellular localizations; gene, protein, and chromosome sequence information; tools for analysis and comparison of sequences; and links to literature information.

The CGD homepage (<http://www.candidagenome.org>) is organized similarly to the SGD homepage. It can be broken down into three main sections:

1. Main toolbar



The screenshot shows the top navigation bar of the CGD homepage. On the left is the CGD logo and the text "Candida Genome Database". To the right are links for "About", "Site Map", "How to Cite", and "Help", followed by social media icons for Facebook and Twitter, and a search bar with a "go" button. Below the search bar is a "New and Noteworthy" section featuring a photograph of GFP-labeled Dam1 Complex proteins in DAPI-stained nuclei, with the caption "GFP-labeled Dam1 Complex proteins in DAPI-stained nuclei. Courtesy of Laura Burack and Judy Berman, University of Minnesota". The main menu below the toolbar includes "Home", "Search", "GBrowse", "JBrowse", "Sequence", "GO", "Tools", "Literature", "Download", and "Community".

2. Search box

3. News & Media

1. The **main toolbar** organizes CGD's key pages, tools, and resources into the categories shown above. Mousing over each category reveals a dropdown menu with options within each category. This toolbar is accessible from most CGD pages.
2. The **search box** is located above the main toolbar on the right side of the page. It enables searching for key types of data, including gene names, Gene Ontology (GO) terms, references, phenotypes, biochemical pathways, orthologs, and more. Hitting Enter/Return after typing in your query will take you to a disambiguation page, where you can select a category of data within various *Candida* species or other areas of CGD to narrow your search results. Typing into the search box without hitting Enter/Return populates a list of suggested hits.

3. The **news & media** section lists important updates for CGD users, such as curation projects, new data, and new tools. On the left side of the homepage, conferences and courses relevant to yeast researchers are listed.

In the upper right hand corner of the page, CGD also provides links to CGD's help documentation and helpdesk email (candida-curator@lists.stanford.edu), CGD citation information and publications, the CGD Facebook page, and a Site Map. The CGD Site Map issues a comprehensive listing of major CGD tools and resources, providing a helpful summary of the purpose of each page and tool.

Searching CGD and Navigating Gene Pages

Explore gene-specific information in CGD using Quick search and Locus Summary page for aquaporin (**AQY1**)

- Open the CGD home page (<http://www.candidagenome.org>).
- Enter your query into the "search our site" box above the banner. Looking for a particular gene, you can enter a standard gene name (**AQY1**), a systematic name (**CR_02920C**), an alias or a systematic name from previous genome assemblies (**orf19.2849, orf6.4943, CA2873**), or an identifier from some of the external resources (NCBI's Gene ID: **3642587**, as an example).

The screenshot shows the CGD homepage with a navigation bar at the top. The search bar contains the placeholder "search our site" and a "go" button. Below the search bar, there is a banner image showing GFP-labeled Dam1 Complex proteins in DAPI-stained nuclei. To the right of the banner, a news section titled "New and Noteworthy" is displayed, featuring a link to the *C. lusitaniae* strain CBS 6936 sequence and BLAST datasets. The "About CGD" section on the left provides information about the database's funding and partners. The "Meetings & Courses" section is also visible. On the right, the "CGD Curation News" section lists recent additions to the database.

The screenshot shows the "CGD Quick Search Result" page for the query "aqy1". The results are organized by species. For each species, it shows the number of hits for Gene names, Biochemical pathways, General Descriptions, and Phenotypes. The results for *Candida albicans* show 1 hit for Gene names and 1 hit for Ortholog or Best Hit. The results for *Candida glabrata* show 1 hit for Gene names and 3 hits for Ortholog or Best Hit. The results for *Candida parapsilosis* show 0 hits for all categories. The results for *Candida dubliniensis* show 0 hits for all categories.

- If your query produces multiple hits, such as a gene name that is used in several *Candida* species represented in CGD, you will get a "CGD Quick Search Result" page that lists the type and number of hits, general and broken down by species. Positive hits are hyperlinked to either their respective Locus Summary pages, or to an intermediate list of individual hits. We will explore other search options later, but for now, select **1 Gene names (gene name/alias/ORF name)** under "*Candida albicans* Search Results" to open the Locus Summary page.

Explore the *C. albicans* AQY1 Locus Summary page:

- On the **Summary** tab, visit the Description and GO Annotations section.

C. albicans AQY1/CR_02920C Summary

Last curated 2014-06-25
See the [Literature Guide](#) for most recent publications.

[Summary](#) [Locus History](#) [Literature](#) [Gene Ontology](#) [Phenotype](#) [Homologs](#) [Protein](#)

AQY1 BASIC INFORMATION [View References]

Standard Name	AQY1
Systematic Name, Reference Strain	CR_02920C_A (<i>C. albicans</i> SC5314)
Assembly 19/21 Identifier	orf19.2849
Alias	orf19.10368, IPF24496.1, IPF10705.1, Contig4-2389_0006, orf6.4943, CA2873, CaO19.2849, orf19.2849, CR_02920C_B, CR_02920C
Feature Type	ORF, Verified
Description	Aquaporin water channel; osmotic shock resistance, WT freeze tolerance; virulent in mice; flucytosine repressed; flow model/RPMI/Spider/rat catheter biofilm induced; required for RPMI biofilm formation; Bcr1-induced in a/a RPMI biofilms (1, 6, 7, 8, 9, 10, 11)
Literature	Literature Guide [View]
Allele Name	CR_02920C_B
Allelic Variation	Synonymous variation between alleles Sequence variation between alleles within 100 bp upstream of feature start coordinates
CUG Codons	CR_02920C_A: 0 CR_02920C_B: 0
Systematic Names Used in Other Strains	CAWG_G_01628 (<i>C. albicans</i> WO-1)
Orthologous genes in <i>Candida</i> species	<i>C. dubliniensis</i> CD36 Ortholog(s) : Cd36_27990 <i>C. auris</i> B4441 Ortholog(s) : B9J08_000928 <i>C. parapsilosis</i> CDC317 Ortholog(s) : CPAR2_800150 View ortholog cluster : 15 genes among 15 <i>Candida</i> -related species/strains
Ortholog(s) in non-CGD species	<i>A. nidulans</i> (AN10902) ; <i>N. crassa</i> (NCU08052) ; <i>S. cerevisiae</i> (AQY1)
Best hit(s) in non-CGD species	<i>S. pombe</i> (SPAC977.17)
Best hits in <i>Candida</i> species	<i>C. glabrata</i> CBS138 best hit(s) : CAGL0A01221g
JBrowse	

GO Annotations

[View all AQY1 GO evidence and references](#)

Molecular Function	<ul style="list-style-type: none"> water channel activity (ISS, IDA) water channel activity (IEA with <i>S. cerevisiae</i>: AQY1)
Biological Process	<ul style="list-style-type: none"> cellular response to freezing (IMP) cellular response to osmotic stress (IMP) single-species biofilm formation on inanimate substrate (IMP) ascospore formation (IEA with <i>S. cerevisiae</i>: AQY1) proteasome-mediated ubiquitin-dependent protein catabolic process (RCA) transmembrane transport (IEA with <i>S. cerevisiae</i>: AQY1) water transport (IEA with <i>S. cerevisiae</i>: AQY1)
Cellular Component	<ul style="list-style-type: none"> membrane (ISS) plasma membrane (IDA) plasma membrane (IEA with <i>S. cerevisiae</i>: AQY1)

[View all AQY1 Phenotype details and references](#)

Mutant Phenotype

Classical genetics

homozygous null

- biofilm formation: abnormal
- flocculation: normal
- freeze-thaw resistance: decreased
- hyphal growth: normal
- invasive growth: normal
- osmotic stress resistance: increased
- viable
- virulence: normal

- Scroll down to Locus Summary Notes.

LOCUS SUMMARY NOTES for AQY1 (Last Updated: 2011-06-20)

- AQY1 encodes an aquaporin water channel (1)
- it is required for wild-type tolerance of freezing (7)
- the mutant shows increased resistance to osmotic shock (1)
- its expression is induced during biofilm formation (9)
- the transcript is repressed by flucytosine (6)
- it is not required for virulence in a mouse model of systemic infection (1)

REFERENCES CITED ON THIS PAGE [View Complete Literature Guide for AQY1]

- 1) Carbrey JM, et al. (2001) Aquaporin in Candida: characterization of a functional water channel protein. *Yeast* 18(15):1391-6
- 2) CandidaDB
- 3) Berman J (2005) Mapping of ORFs in Assembly 4 to those in Assembly 19.
- 4) CGD (2005) CGD comparison of orf6 and orf19 predicted protein products.
- 5) Magliotti D, et al. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 35(Database issue):D26-31
- 6) Liu TT, et al. (2005) Genome-wide expression profiling of the response to azole, polyene, echinocandins, and pyrimidine antifungal agents in Candida albicans. *Antimicrob Agents Chemother* 49(1):2226-36
- 7) Tanghe A, et al. (2005) Aquaporin expression and freeze tolerance in Candida albicans. *Appl Environ Microbiol* 71(10):6434-7
- 8) Nett JE, et al. (2009) Time course global gene expression analysis of an in vivo Candida biofilm. *J Infect Dis* 200(2):307-13
- 9) Bonhommier J, et al. (2011) Contribution of the glycolytic flux and hypoxia adaptation to efficient biofilm formation by Candida albicans. *Mol Microbiol* 80(4):995-1013
- 10) Nobile CJ, et al. (2012) A recently evolved transcriptional network controls biofilm development in Candida albicans. *Cell* 148(1-2):126-38
- 11) Srikantha T, et al. (2013) Identification of genes upregulated by the transcription factor Bcr1 that are involved in impermeability, impenetrability, and drug resistance of Candida albicans a/alpha biofilms. *Eukaryot Cell* 12(6):875-88

[Return to CGD](#) [Send a Message to the CGD Curators](#)

CGD Copyright © 2004-2023 The Board of Trustees, Leland Stanford Junior University
Permission to use the information contained in this database was given by the researchers/institutes who contributed or published the information. Users of the database are solely responsible for compliance with any copyright restrictions, including those applying to the author abstracts. Documents from this server are provided "AS-IS" without any warranty, expressed or implied.
To cite CGD, please use the following reference: Skrzypek MS, Binkley J, Binkley G, Miyasato SR, Simison M, Sherlock G (2017). The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res* 45 (D1): D592-D596; see How to cite CGD.

- Open the **Gene Ontology** tab to see more information, including evidence and references. Hyperlinked phrases lead to definitions.

C. albicans AQY1/CR_02920C Gene Ontology Annotations [?](#)

[Summary](#) [Locus History](#) [Literature](#) [Gene Ontology](#) [Phenotype](#) [Homologs](#) [Protein](#)

This page displays GO annotations in different sections according to the methods used in the reference from which the annotation was made:

- Manually Curated GO Annotations: includes annotations based on published experiments or analyses that focus on specific genes.
- GO Annotations from High-throughput Experiments: includes annotations made from published experiments performed on a high-throughput or genome-wide basis.
- Computational GO Annotations: includes annotations that are predicted by computational methods (e.g., sequence similarity comparisons) and are not individually reviewed.

AQY1 Manually Curated GO Annotations*: Last Reviewed on: 2011-10-25 Molecular Function | Biological Process | Cellular Component

[Jump to: Top | Computational | High-throughput](#)

Manually Curated Molecular Function				
Annotation(s)	Reference(s)	Evidence	Assigned By	
water channel activity	Carbrey JM, et al. (2001) Aquaporin in Candida: characterization of a functional water channel protein. <i>Yeast</i> 18(15):1391-6	ISS : Inferred from Sequence or structural Similarity IDA : Inferred from Direct Assay Assigned on 2014-06-25	CGD	

Manually Curated Biological Process				
Annotation(s)	Reference(s)	Evidence	Assigned By	
cellular response to freezing	Tanghe A, et al. (2005) Aquaporin expression and freeze tolerance in Candida albicans. <i>Appl Environ Microbiol</i> 71(10):6434-7	IMP : Inferred from Mutant Phenotype Assigned on 2014-06-25	CGD	
cellular response to osmotic stress	Carbrey JM, et al. (2001) Aquaporin in Candida: characterization of a functional water channel protein. <i>Yeast</i> 18(15):1391-6	IMP : Inferred from Mutant Phenotype Assigned on 2014-06-25	CGD	
single-species biofilm formation on inanimate substrate	Srikantha T, et al. (2013) Identification of genes upregulated by the transcription factor Bcr1 that are involved in impermeability, impenetrability, and drug resistance of Candida albicans a/alpha biofilms. <i>Eukaryot Cell</i> 12(6):875-88	IMP : Inferred from Mutant Phenotype Assigned on 2014-06-25	CGD	

Manually Curated Cellular Component				
Annotation(s)	Reference(s)	Evidence	Assigned By	
membrane	Carbrey JM, et al. (2001) Aquaporin in Candida: characterization of a functional water channel protein. <i>Yeast</i> 18(15):1391-6	ISS : Inferred from Sequence or structural Similarity Assigned on 2014-06-25	CGD	

*Manually Curated GO annotations reflect our best understanding of the basic molecular function, biological process, and cellular component for this gene product. Manually Curated annotations are assigned by CGD curators based on published, small-scale experiments. Curators periodically review all Manually Curated GO annotations for accuracy and completeness. The "Last Reviewed on:" date at the top of this section indicates when these annotations were last reviewed.

AQY1 GO annotations from High-Throughput Experiments:** [Jump to: Top | Computational | Manually curated](#)

Cellular Component Annotations from High-Throughput Experiments				
Annotation(s)	Reference(s)	Evidence	Assigned By	
plasma membrane	Cabezon V, et al. (2009) Analysis of Candida albicans plasma membrane proteome. <i>Proteomics</i> 9(20):4770-86	IDA : Inferred from Direct Assay Assigned on 2014-06-25	CGD	

- What are the phenotypes caused by mutations in this gene? In the Summary tab, find the Mutant Phenotype section. Open the **Phenotype** tab for more details, including experimental settings and references.

C. albicans AQY1/CR_02920C_A Phenotypes

Summary Locus History Literature Gene Ontology **Phenotype** Protein

This page lists all curated single mutant phenotypes associated with AQY1. Click on a term in the phenotype column to see other genes associated with that term or [download all data](#). [Browse phenotype terms](#)

Jump to: Download Data

13 Single Mutant Phenotype(s) for AQY1/CR_02920C_A						
Experiment type	Mutant Information	Strain background	Phenotype	Details	Virulence Model	References
heterozygous diploid, classical genetics	Description: repressible Allele: aqy1::FRT/aqy1::FRT ADH1/adh1::pTET-AQY1-GFP::SATR	SC5314	biofilm formation: abnormal	Condition: RPMI medium 29 degrees C with rocking for 48 hr Details: mutant has minor defects in vertical orientation of hyphae and reduced extracellular matrix, complete permeability to Sypro Ruby and human PMNs and decreased fluconazole resistance		Srikantha T, et al. (2013) Identification of genes upregulated by the transcription factor Bcr1 that are involved in impermeability, impenetrability, and drug resistance of <i>Candida albicans</i> a/alpha biofilms. <i>Eukaryot Cell</i> 12(6):875-88 curation status: CURATED published status: Published Access Full Text
heterozygous diploid, classical genetics	Description: repressible Allele: ADH1/adh1::pTET-AQY1-GFP::SATR (tet repressed)	P37005	biofilm formation: abnormal	Condition: RPMI medium 29 degrees C with rocking for 48 hr Details: decreased Sypro Ruby and human PMN penetrability, decreased resistance to fluconazole		Srikantha T, et al. (2013) Identification of genes upregulated by the transcription factor Bcr1 that are involved in impermeability, impenetrability, and drug resistance of <i>Candida albicans</i> a/alpha biofilms. <i>Eukaryot Cell</i> 12(6):875-88 curation status: CURATED published status: Published Access Full Text
homozygous diploid, classical genetics	Description: null Allele: a/alpha aqy1::FRT/aqy1::FRT	SC5314	biofilm formation: abnormal	Condition: RPMI medium 29 degrees C with rocking for 48 hr Details: mutants have reduced vertical orientation of hyphae, an abnormal extracellular matrix, complete penetrability to Sypro Ruby and human PMNs and increased sensitivity to fluconazole		Srikantha T, et al. (2013) Identification of genes upregulated by the transcription factor Bcr1 that are involved in impermeability, impenetrability, and drug resistance of <i>Candida albicans</i> a/alpha biofilms. <i>Eukaryot Cell</i> 12(6):875-88 curation status: CURATED published status: Published Access Full Text
homozygous diploid, classical genetics	Description: null	SC5314	biofilm formation: abnormal	Details: forms biofilm with altered detachment properties		Sellam A, et al. (2009) A <i>Candida albicans</i> early stage biofilm detachment event in rich medium. <i>BMC Microbiol</i> 9:25 curation status: CURATED published status: Published Access Full Text
homozygous diploid, classical genetics	Description: null	CAI-4	flocculation: normal	Condition: liquid YPD medium		Carbrey JM, et al. (2001) Aquaporin in <i>Candida</i> : characterization of a functional water channel protein. <i>Yeast</i> 18(15):1391-6 curation status: CURATED published status: Published Access Full Text
heterozygous diploid, classical genetics	Description: repressible	Not recorded	freeze-thaw resistance: decreased			Tanghe A, et al. (2005) Aquaporin expression and freeze tolerance in <i>Candida albicans</i> . <i>Appl Environ Microbiol</i> 71(10):6434-7 curation status: CURATED published status: Published Access Full Text
homozygous diploid, classical genetics	Description: null	Not recorded	freeze-thaw resistance: decreased			Tanghe A, et al. (2005) Aquaporin expression and freeze tolerance in <i>Candida albicans</i> . <i>Appl Environ Microbiol</i> 71(10):6434-7 curation status: CURATED published status: Published Access Full Text
homozygous diploid, classical genetics	Description: null	CAI-4	hyphal growth: normal	Condition: solid Spider medium		Carbrey JM, et al. (2001) Aquaporin in <i>Candida</i> : characterization of a functional water channel protein. <i>Yeast</i> 18(15):1391-6 curation status: CURATED published status: Published Access Full Text
homozygous diploid, classical genetics	Description: null	CAI-4	invasive growth: normal	Condition: solid YPD medium		Carbrey JM, et al. (2001) Aquaporin in <i>Candida</i> : characterization of a functional water channel protein. <i>Yeast</i> 18(15):1391-6 curation status: CURATED published status: Published Access Full Text
homozygous diploid, classical genetics	Description: null	CAI-4	osmotic stress resistance: increased	Details: greater-than-wild type resistance to both hyper- and hypoosmotic shock		Carbrey JM, et al. (2001) Aquaporin in <i>Candida</i> : characterization of a functional water channel protein. <i>Yeast</i> 18(15):1391-6 curation status: CURATED published status: Published Access Full Text
heterozygous diploid, large-scale survey (haploinsufficient phenotype assay, barcode deletion set)	Description: null	SC5314	viable			Xu D, et al. (2007) Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in <i>Candida albicans</i> . <i>PLoS Pathog</i> 3(6):e92 curation status: CURATED published status: Published Web Supplement
homozygous diploid, classical genetics	Description: null	CAI-4	viable			Carbrey JM, et al. (2001) Aquaporin in <i>Candida</i> : characterization of a functional water channel protein. <i>Yeast</i> 18(15):1391-6 curation status: CURATED published status: Published Access Full Text
homozygous diploid, classical genetics	Description: null	CAI-4	virulence: normal	Details: organ colonization is also normal	mouse intravenous infection	Carbrey JM, et al. (2001) Aquaporin in <i>Candida</i> : characterization of a functional water channel protein. <i>Yeast</i> 18(15):1391-6 curation status: CURATED published status: Published Access Full Text

- What are the orthologs in other *Candida* species? Return to Summary tab and find Orthologous Genes and Best Hits among items listed in Basic Information section with links to their source databases. Open the Homologs tab to see a Phylogenetic Tree and multiple sequence alignments.

***C. albicans* AQY1/CR_02920C Homology Information**

?

Summary Locus History Literature Gene Ontology Phenotype Homologs Protein

AQY1 HOMOLOG INFORMATION

Ortholog Cluster From CGOB View CGOB cluster and synteny information

Download cluster sequence files:

- Proteins (multi-FASTA format)
- Coding (multi-FASTA format)
- Genomic (multi-FASTA format)
- Genomic +/- 1000 BP (multi-FASTA format)

Sequence ID	Organism	Source	Status
AQY1/CR_02920C_A	<i>Candida albicans</i> SC5314	CGD	VERIFIED
Cd36_27990	<i>Candida dubliniensis</i> CD36	CGD	UNCHARACTERIZED
CPAR2_800150	<i>Candida parapsilosis</i> CDC317	CGD	UNCHARACTERIZED
AQY1/YPR192W	<i>Saccharomyces cerevisiae</i> S288C	SGD	VERIFIED
CAWG_01628	<i>Candida albicans</i> WO-1	EnsemblFungi	
CLUG_01034	<i>Candida lusitaniae</i> ATCC 42720	EnsemblFungi	
CORT_0A00270	<i>Candida orthopsisilosis</i> Co 90-125	EnsemblFungi	
LELG_00007	<i>Lodderomyces elongisporus</i> NRRL YB-4239	EnsemblFungi	
DEHA2F27104g	<i>Debaryomyces hansenii</i> CBS767	EnsemblFungi	

Best hits in CGD species *C. glabrata* CBS138 best hit: [CAGL0A01221g](#)

Orthologs in fungal species *A. nidulans* ([AN10902](#)) ; *N. crassa* ([NCU08052](#))

Best hits in fungal species *S. pombe* ([SPAC977.17](#))

Reciprocal best hits in other species *M. musculus* ([AQP1](#)) ; *R. norvegicus* ([Aqp1](#)) ; *D. discoideum* ([wacA](#))

Phylogenetic Tree
Built with SEMPHY
Download tree files:

- Unrooted Tree (Newick format)
- Rooted Tree (Newick format)
- Rooted Tree (phyloXML format)
- Rooted, Annotated Tree (phyloXML format)

Tree rooted by midpoint; total tree length = 2.45 subs/site

0.2 subs/site

Protein Sequence Alignment
Built with MUSCLE
Download alignment files:

- Protein alignment (Multi-FASTA format)
- Protein alignment (ClustalW format)

Reference sequence (1): CR_02920C_A
Identities normalized by aligned length.
Colored by: identity >= 80% and property:
 Hydrophobic (A, I, L, M, V)
 Aromatic (F, W, Y)
 Polar (N, Q, S, T)
 Negative charge (D, E)
 Positive charge (H, K, R)
 Backbone change (G, P)
 Cysteine (C)

```

1 CR_02920C_A 100.0% -----MVAESSSIDNT-----PNDVEAQ
2 CAWG_01628 100.0% -----MVAESSSIDNT-----PNDVEAQ
3 Cd36_27990 98.9% -----MVAESSSIDNT-----ANDVEAQ
4 CPAR2_800150 81.6% -----
5 CORT_0A00270 79.3% -----MTVEAT---SP-----IDDIEQQ
6 LEGL_00007 80.4% -----MTAAGSIAEPT-----PNEIEAQ
7 DEHA2F27104g 19.4% -----MDSTLGSDSLPEKERTTIDSEGLNHRNPFRFEGENRLSPDLEAQ
8 CLUG_01034 75.0% -----MT-----ASDVEAQ
9 YPR192W 45.9% MSSNDSNNDTDKQHTRLDPT-----GVDDAYI
  
```

- Explore the **Protein** tab. What is known about the structure and physicochemical properties of this protein? How is it expected to fold?

C. albicans Aqy1p/Cr_02920cp Protein Information

?

Summary Locus History Literature Gene Ontology Phenotype Homologs Protein

Aqy1p PROTEIN INFORMATION [View References]

Standard Name	Aqy1p ¹
Systematic Name	Cr_02920cp_a
Allele Name	Cr_02920cp_b
Description	Aquaporin water channel; osmotic shock resistance, WT freeze tolerance; virulent in mice; flucytosine repressed; flow model/RPMI/Spider/rat catheter biofilm induced; required for RPMI biofilm formation; Bcr1-induced in a/a RPMI biofilms (1, 2, 3, 4, 5, 6, 7)

Experimental Observations

Structural Information

Top PDB Hit: **2W2E**

Source
Quality
Structure

Predicted Structure from AlphaFold ([Link-out](#))

[View all PDB BLAST hits](#)
[View Top PDB Hit Information \(Link-out\)](#)
Komagataella pastoris; 1.15 Angstrom Crystal Structure Of P.Pastoris Aquaporin, Aqy1, In A Closed Conformation At Ph 3.5
68% identical to Aqy1p; P-value = 3.00e-98 ([View Alignment](#))
Click on image to access PDB interactive viewer ([Link-out](#))

Sequence Detail
[Download](#) in FASTA format)

Length = 273 aa; MW = 29.3 kDa; [Physicochemical Properties Page](#)

```

1 MVAESSSSIDN TPNDVVEAQRP VYERKYDDSV NVSPLKNHMI AFLGEFFGTF
51 IFLWVAFVPIA QIANQDPTIP DKGSQDPMQLI MISPGFGFVG MMGVPMFPRV
101 SGGNLNPATV ITLVLQAQAVP PIRGLPFMVA QMIAGMAAAG AASAMTPGRI
151 AFTNGLGGGA SKARGVFLEA FGTCLILCITV LMMAVEKSRA TFMAPFVIGI
201 SLFLGHLLIV YYTGAGLNPNA RSFGPCVAAR SFPVYHWIYW VGPILGSVIA
251 FAIWKIFKIL KYETCNPGQD SDA*

```

Homologs [BLAST Aqy1p against other *Candida* sequences](#)

- What is the chromosomal location of AQY1 and its neighboring genes? From the **Summary** tab, view the **Sequence Information** section and retrieve sequence with or without flanking sequence and/or introns.

Sequence Information

Ca22chrRA_C_albicans_SC5314:668602 to 667781 | [JBrowse](#)
Note: this feature is encoded on the Crick strand.

Last Update	Coordinates: 2016-01-21 Sequence: 2014-06-24		
Subfeature Details	Relative Coordinates	Chromosomal Coordinates	Most Recent Update
CDS 1 to 822	668,602 to 667,781	2016-01-21	2014-06-24

Retrieve Sequences [DNA + flanking intergenic seq for CR_02920C_A](#) [View](#)

Sequence Analysis Tools [C. albicans SC5314 Assembly 22](#) [View](#)

Maps & Displays [Flanking Features Table](#) [View](#)

Allele Location
Allele CR_02920C_B

Ca22chrRB_C_albicans_SC5314:668545 to 667724 | [JBrowse](#)
Note: this feature is encoded on the Crick strand.

Last Update	Coordinates: 2016-01-21 Sequence: 2014-06-24		
Subfeature Details	Relative Coordinates	Chromosomal Coordinates	Most Recent Update
CDS 1 to 822	668,545 to 667,724	2016-01-21	2014-06-24

- To analyze the AQY1 sequence, in the Sequence Information section, open the Sequence Analysis Tools pull-down menu to run BLAST, design primers, and get restriction maps.

Sequence Information ▾

Ca22chrRA_C_albicans_SC5314:668602 to 667781 | [GBrowse](#)
Note: this feature is encoded on the Crick strand.

Last Update	Coordinates: 2016-01-21 Sequence: 2014-06-24		
Subfeature Details	Relative Coordinates	Chromosomal Coordinates	Most Recent Update
CDS	1 to 822	668,602 to 667,781	Coordinates Sequence 2016-01-21 2014-06-24
Retrieve Sequences		-- C. albicans SC5314 Assembly 22 --	<input type="button" value="View"/>
Sequence Analysis Tools		-- C. albicans SC5314 Assembly 22 --	<input type="button" value="View"/>
Maps & Displays		Fanking Features Table	<input type="button" value="View"/>

- What literature is available on AQY1? View references at the bottom of AQY1 Locus Summary page and click on Complete Literature Guide, or simply open the **Literature** tab. To filter out papers that deal with a specific topic, such as mutants and phenotypes, click on that topic using the menu on the left side of the page.

C. albicans AQY1/CR_02920C Literature Guide

[Summary](#) [Locus History](#) [Literature](#) [Gene Ontology](#) [Phenotype](#) [Homologs](#) [Protein](#)

Other names in use for *C. albicans* AQY1: orf19.10368, IPF24496.1, IPF10705.1, Contig4-2389_0006, orf6.4943, CA2873, CaO19.2849, orf19.2849, CR_02920C_B, CR_02920C, CAWG_01628, CR_02920C_A

This page displays all the papers associated with *C. albicans* AQY1 in CGD, along with all the literature topics those papers address. Click on a topic on the left to see the papers that address it.

AQY1 LITERATURE TOPICS

Virulence-related

- Animal Model
- Sensitivity/response to drugs/other treatments

Related Genes/Proteins

- Cross-species Expression
- Fungal Related Genes/Proteins

Nucleic Acid Information

- DNA/RNA Sequence Features
- RNA Levels and Processing

Research Aids and Literature

- Strains/Constructs

Proteome-wide Analysis

- Large-scale protein detection

Gene Product Information

- Protein Domains/ Motifs
- Protein Physical Properties
- Substrates/Ligands/Cofactors

Life Cycle

- Biofilms

Curated Literature

- Alias
- Reviews
- List of all Curated References

Regulation

- Transcriptional Regulation

Genetics/Cell Biology

- Cell Growth and Metabolism
- Cellular Location

Function/Process

- Genetic Interactions

Mutants/Phenotypes

- Signal Transduction

Related Species

- *Candida albicans*

Genome-wide Analysis

- Genome-wide Analysis
- Genomic expression study
- Large-scale phenotype analysis

AQY1 Literature Curation Summary

Curated References for AQY1: 16

References Not Yet Curated: 0

References for Curation: 0

Number of Other Genes referred to in AQY1 Literature: 3401

Date of last curation: 2014-06-25

Date of last PubMed Search: 2017-04-02

Reference	Species	Other Genes Addressed
Srikantha T, et al. (2013) Identification of genes upregulated by the transcription factor Bcr1 that are involved in impermeability, impenetrability, and drug resistance of <i>Candida albicans</i> α/α biofilms. <i>Eukaryot Cell</i> 12(6):875-88	<i>C. albicans</i>	BCR1 BRG1 C1_05890W_A CHK1 CR_06500C_A CSA1 CSA2 EAP1 ECE1 GCA1 GCA2 GSL1 HAP3 HWP1 MORE
Bonhomme J, et al. (2011) Contribution of the glycolytic flux and hypoxia adaptation to efficient biofilm formation by <i>Candida albicans</i> . <i>Mol Microbiol</i> 80(4):995-1013	<i>C. albicans</i>	AAF1 ABC1 ACS1 ADH1 ADH5 ADK1 AHP1 ALP1 ALS3 ALT1 AMS1 ARE2 ARG1 ARG4 MORE
Vylkova S, et al. (2011) The fungal pathogen <i>Candida albicans</i> autoinduces hyphal morphogenesis by raising extracellular pH. <i>MBio</i> 2(3):e00055-11	<i>C. albicans</i>	ACH1 ACS1 ARG1 ARG3 ARG4 AT01 AT010 AT02 AT05 AT06 AT07 AT09 C2_02650C_A C4_06910W_A MORE
Synnott JM, et al. (2010) Regulation of the hypoxic response in <i>Candida albicans</i> . <i>Eukaryot Cell</i> 9(11):1734-46	<i>C. albicans</i>	AAH1 ABP1 ADH1 ALS4 ARE2 ASR2 AT02 BCR1 BI02 BMT3 C1_00160C_A C1_01360C_A C1_02700C_A C1_03510C_A MORE
Cabezon V, et al. (2009) Analysis of <i>Candida albicans</i> plasma membrane proteome. <i>Proteomics</i> 9(20):4770-86	<i>C. albicans</i>	ADH1 AFG3 AGC1 AL1 AL01 AO2 ATXP1 ATP17 ATP18 ATP19 ATP20 ATP20 ATP3 ATP4 MORE
Cottier F and Muhlschlegel FA (2009) Sensing the environment: response of <i>Candida albicans</i> to the X factor. <i>FEMS Microbiol Lett</i> 295(1):1-9	<i>C. albicans</i>	ACE2 BUD2 CAG1 CAN1 CCH1 CEK1 CPH1 ICYR1 CFZ1 EFG1 FIG1 GAP1 GAP2 PAT1 MORE
Hua X, et al. (2009) Morphogenic and genetic differences between <i>Candida albicans</i> strains are associated with keratomycosis virulence. <i>Mol Vis</i> 15:1476-84	<i>C. albicans</i>	AAP1 AA21 ACB1 AC02 ADAEC AHP1 ALD6 ALS1 ALS2 ALS4 JLS7 AM02 AOX2 APE3 MORE
Nett JE, et al. (2009) Time course global gene expression analysis of an in vivo <i>Candida</i> biofilm. <i>J Infect Dis</i> 200(2):307-13	<i>C. albicans</i>	AAH1 AGP2 ALD6 ALS1 AOX2 ARG1 ARG6 BGL2 C1_10570C_A CAN1 CAT8 CDC21 CDG1 CDR2 MORE
Sellam A, et al. (2009) A <i>Candida albicans</i> early stage biofilm detachment event in rich medium. <i>BMC Microbiol</i> 9:25	<i>C. albicans</i>	ALS1 ALS3 AMS1 BCR1 CW8H HS2P21 HWP1 MKC1 PGA13 PSA2 YWP1
Trunk K, et al. (2009) Depiction of the collin Cdc53p induces morphogenetic changes in <i>Candida albicans</i> . <i>Eukaryot Cell</i> 8(5):756-67	<i>C. albicans</i>	AOX2 ARG1 C4_05610C_A C4_06910W_A CDC53 CFI2 CR_09140C_A GLO1 HSP12 MET1 PHO112 STE11
Xu D, et al. (2007) Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in <i>Candida albicans</i> . <i>PLoS Pathog</i> 3(6):e92	<i>C. albicans</i>	AAH1 AAP1 AA1 AT21 ABC1 ABD1 ABP140 ABZ1 ACC1 ACF2 ACO1 ACO2 ACP12 ACS2 MORE

SGD/CGD JBrowse Genome Browser

SGD and CGD both provide the genome browsing tool **JBrowse** to enable easy exploration of yeast genomes. JBrowse enables quick scrolling through genomic features and visualization of experimental information from large-scale studies in the form of **data tracks**. In this exercise, we will use JBrowse to visualize the location of genes related to galactose catabolism and use data tracks to visualize how these genes are transcriptionally regulated.

Accessing JBrowse

- You can access CGD's JBrowse genome browser in the following locations:
- From the home page (www.candidagenome.org) toolbar menu for **JBrowse**.

The screenshot shows the homepage of the Candida Genome Database (CGD). At the top, there is a navigation bar with links for Home, Search, GBrowse, JBrowse (which is highlighted with a purple arrow), Sequence, GO, Tools, Literature, Download, and Community. Below the navigation bar, there is a main content area featuring a blue background image of a yeast cell. To the left of the image, there is a sidebar with the text "Extracellular alkalization by wild-type C. albicans" and a link to a PLoS paper. On the right side of the image, there is a list of Candida species: *C. albicans*, *C. auris*, *C. dubliniensis*, *C. glabrata*, and *C. parapsilosis*. Further down the page, there are sections for "New and Noteworthy" and "CANDIDA AND CANDIDIASIS 2021". The "CANDIDA AND CANDIDIASIS 2021" section contains a brief description of the meeting and its deadline. At the bottom of the page, there is a "CGD Curation News" section with a single item: "• *C. albicans* Genome Snapshot".

- From any Locus Summary page, by clicking on the JBrowse image link in the Basic Information section.
- CGD JBrowse provides *C. albicans*, *auris*, *dubliniensis*, *glabrata*, and *parapsilosis*.
- You can access the SGD's JBrowse genome browser in the following locations:
 - From the home page (www.yeastgenome.org), by opening the Sequence menu in the top purple toolbar and selecting **Genome Browser**.
 - From any Locus Summary page, by selecting **View in JBrowse** under Sequence
 - Or by following this link: <https://browse.yeastgenome.org>

Analyzing transcriptional regulation of galactose catabolism

Using SGD's JBrowse genome browser, analyze the transcriptional regulation of **GAL10**.

- In the JBrowse window, enter **GAL10** into the search box in the navigation bar on top and press **Go**. Multiple results will be listed, but all refer to the same gene.

Name	Location	Show	Go
GAL10	chrII:276253..278352	<input type="button" value="Show"/>	<input type="button" value="Go"/>
GAL10	chrII:276253..278352	<input type="button" value="Show"/>	<input type="button" value="Go"/>
GAL10	chrII:276253..278352	<input type="button" value="Show"/>	<input type="button" value="Go"/>

- Click on the GAL10 red feature bar to see an overview of GAL10 sequence data. What is the chromosomal location, strand, and sequence of this gene?
- What genes are upstream and downstream of GAL10? Zoom in/out using the magnifying glass icons in the navigation bar, or double-click on an empty spot in the browser. Move the viewing window left/right by using the arrow buttons on the navigation bar, the arrow keys on your keyboard, or by clicking the screen and dragging with your mouse.
- Notice that GAL10 shares its promoter region with the neighboring gene, GAL1, which is located on the opposite strand and transcribed in the opposite direction. Zoom in on the shared promoter by holding down the shift button on your keyboard and dragging over the region with your mouse.

- What transcription factors bind to the GAL1-10 promoter? Add a track with transcription factor binding data to the browser window:

- Press the **Select tracks** button in the upper left corner.
- On the left side of the menu, click on **transcription** (under **Category**).
- In the list of tracks, check the box next to the track that has **MacIssac** in the “First author” column and **TF_ChIP_ChIP** in Track column (you can sort each column by clicking on its header). Click on “**Back to browser**” in the upper left corner.

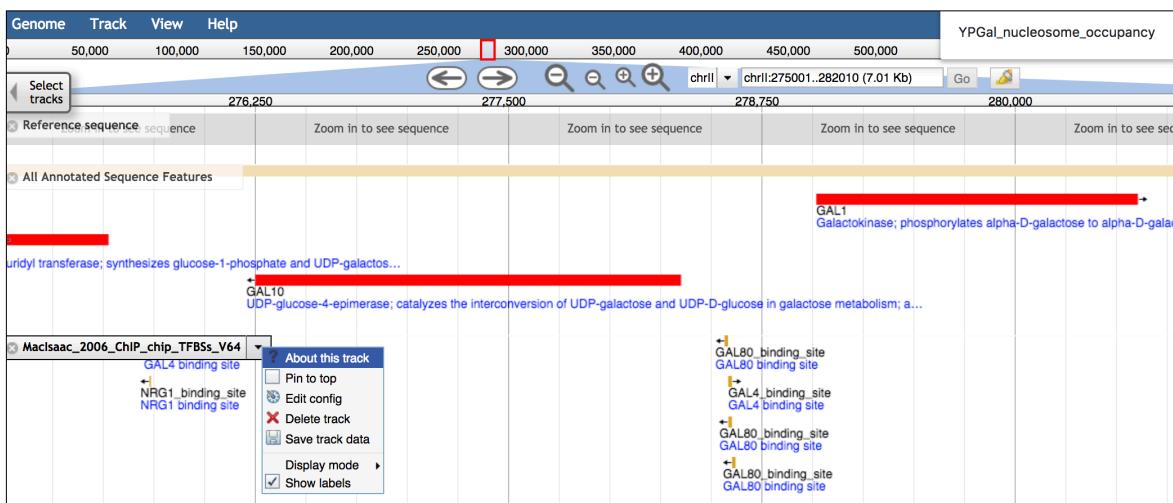
Select Tracks

My Tracks		Back to browser		Clear All Filters		Contains text		67 matching tracks					Help	
Currently Active Recently Used														
Assay Term Name		PMD	First author	Lab PI	Lab	Assay Term Name	Biosample Term Name	Strain background	Category	GBrowse Category	Track			
42 ChIP-chip assay		<input type="checkbox"/>	15905473	Zhang	Fred S. Dietrich	Duke University	Serial Analysis of Gene Expression (SAGE)	polyA RNA extract	W303	transcription	...	Transcription_start_sites		
1 ChIP-seq assay		<input checked="" type="checkbox"/>	16522208	MacIssac	Ernest Fraenkel	MIT	ChIP-chip assay	DNA extract	W303	transcription	transcription recombination	TF_ChIP_ChIP		
5 Chromatin immunoprecipitation with exon-microarray sequencing assay (ChIP-exo)		<input type="checkbox"/>	16569694	David	Lars M. Steinmetz	EMBL	transcription profiling by tiling array assay	polyA RNA extract	S288C	transcription	RNA expression profiling	Transcribed_regions_polyA_RNA		
4 RNA-seq assay		<input type="checkbox"/>	16569694	David	Lars M. Steinmetz	EMBL	transcription profiling by tiling array assay	RNA extract	S288C	transcription	RNA expression profiling	Transcribed_regions_total_RNA		
8 Serial Analysis of Gene Expression (SAGE)		<input type="checkbox"/>	17067396	Xu	Simon Tavare	Cancer Research UK Cambridge Institute	ChIP-chip assay	DNA extract	W303	transcription	RNA expression profiling	Known_ARs_identified		
7 transcription profiling by tiling array assay		<input type="checkbox"/>	17067396	Xu	Simon Tavare	Cancer Research UK Cambridge Institute	ChIP-chip assay	DNA extract	W303	transcription	RNA expression profiling	Known_Predicted_A_CSs		
10 (no data)		<input type="checkbox"/>	17067396	Xu	Simon Tavare	Cancer Research UK Cambridge Institute	ChIP-chip assay	DNA extract	W303	transcription	RNA expression profiling	Known_Predicted_ARs		
1 DNA replication, recombination and repair		<input type="checkbox"/>	17067396	Xu	Simon Tavare	Cancer Research UK Cambridge Institute	ChIP-chip assay	DNA extract	W303	transcription	RNA expression profiling	MCM2_ChIP_chip_binding		
11 DNA replication, recombination and repair		<input type="checkbox"/>	17067396	Xu	Simon Tavare	Cancer Research UK Cambridge Institute	ChIP-chip assay	DNA extract	W303	transcription	RNA expression profiling	ORC2_ChIP_chip_binding		
16 DNA repair		<input type="checkbox"/>	17157256	Steinmetz	David A. Brow	University of Wisconsin	ChIP-chip assay	DNA extract	Other	transcription	transcription regulation	Poll_occupancy_WT		
1 Reptile sequence		<input type="checkbox"/>	17157256	Steinmetz	David A. Brow	University of Wisconsin	ChIP-chip assay	DNA extract	Other	transcription	transcription regulation	Poll_occupancy_sen1		
1 carbon utilization		<input type="checkbox"/>	18628399	Ghavi-Helm	Julie Soutourina	CEA	ChIP-chip assay	DNA extract	S288C	transcription	transcription regulation	Dst1_RNA_PollIII_SC_L_16C_ChIP_chip		
44 chromatin organization		<input type="checkbox"/>	18628399	Ghavi-Helm	Julie Soutourina	CEA	ChIP-chip assay	DNA extract	S288C	transcription	transcription regulation	Dst1_RNA_PollIII_SC_L_16C_ChIP_chip		
1 chromatin organization/transcription		<input type="checkbox"/>										WT_RNA_Poll_VPD		
49 histone modification		<input type="checkbox"/>												
14 tRNA processing		<input type="checkbox"/>												
1 mitotic cell cycle		<input type="checkbox"/>												
17 stress(stress heat shock carbon utilization nutrient utilization osmotic stress oxidative stress) phosphorus utilization		<input type="checkbox"/>												
67 transcription		<input type="checkbox"/>												
2 translation regulation		<input type="checkbox"/>												
2 David		<input type="checkbox"/>												
14 ChIP Data		<input type="checkbox"/>												

- In the main browser window, a new data track for the MacIsaac dataset will be shown.

Click on the binding sites for **GAL4** or **GAL80** for more information about the sites.

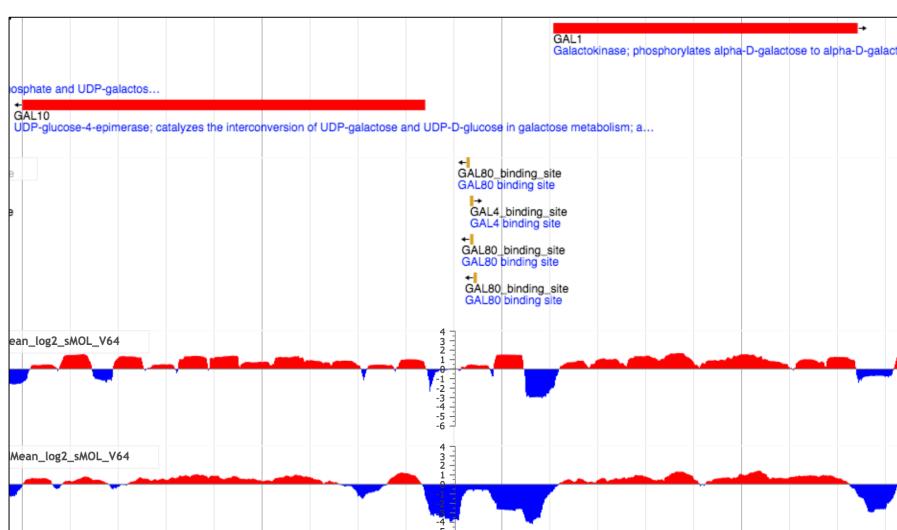
To learn more about the track itself (techniques, experimental design, reference), hover your mouse cursor over the track name and select **About this track** from the pull-down menu.



- What is the nucleosome occupancy around the GAL1-10 promoter and how does it change during growth on galactose? Add tracks with nucleosome occupancy data:
 - Click on **Select tracks** button again and then **Clear All Filters**
 - Under Category, select **chromatin organization** and filter tracks by typing **nucleosome** in “Contains text” search box
 - Check the boxes next to First author: **Kaplan**, Track:
YPD_nucleosome_occupancy_map_dMean_log2_sMOL and
YPGal_nucleosome_occupancy_map_dMean_log2_sMOL. Exit the tracks tab.

Select Tracks										Help				
<input checked="" type="radio"/> My Tracks <input type="radio"/> Currently Active <input type="radio"/> Recently Used		<input type="button" value="Back to browser"/> <input type="button" value="Clear All Filters"/>		Contains text	nucleosome					X 13 matching tracks				
Assay Term Name <ul style="list-style-type: none"> 1 ChIP-seq 3 DNA sequencing 8 micrococcal nuclease digestion followed by high throughput sequencing assay 1 micrococcal nuclease digestion followed by tiling array assay 					PMID	First author	Lab PI	Lab	Assay Term Name	Biosample Term Name	Strain background	Category	GBrowse Category	Track
Category <ul style="list-style-type: none"> 13 chromatin organization 					17392789	Albert	Frank Pugh	Penn State	ChIP-seq assay	DNA extract	S288C	chromatin organization	chromatin structure	H2A2_Nucleosome_positions
First author <ul style="list-style-type: none"> 1 Albert 3 Field 7 Kaplan 1 Lee 1 Mavrich 					17873876	Lee	Corey Nislow	UBC	micrococcal nuclease digestion followed by tiling array assay	DNA extract	S288C	chromatin organization	chromatin structure	Predicted_nucleosome_occupancy_less_o_model
Lab PI <ul style="list-style-type: none"> 1 Corey Nislow 10 Eran Segal 2 Frank Pugh 					18550805	Mavrich	Frank Pugh	Penn State	micrococcal nuclease digestion followed by high throughput sequencing assay	DNA extract	S288C	chromatin organization	chromatin structure	N3H4_Nucleosome_positions
PMID <ul style="list-style-type: none"> 1 17392789 1 17873876 1 18550805 3 18989395 7 19092803 					18989395	Field	Eran Segal	Weizmann Institute of Science	DNA sequencing	DNA extract	Other	chromatin organization	chromatin structure	predicted_average_nucleosome_occupancy
Strain background <ul style="list-style-type: none"> 10 Other 3 S288C 					18989395	Field	Eran Segal	Weizmann Institute of Science	micrococcal nuclease digestion followed by high throughput sequencing assay	DNA sequencing	Other	chromatin organization	chromatin structure	summarized_nucleosome_occupancy
					19092803	Kaplan	Eran Segal	Weizmann Institute of Science	micrococcal nuclease digestion followed by high throughput sequencing assay	DNA extract	Other	chromatin organization	chromatin structure	in vitro_nucleosome_occupancy_map_dMean_log2_sMOL
					19092803	Kaplan	Eran Segal	Weizmann Institute of Science	micrococcal nuclease digestion followed by high throughput sequencing assay	DNA extract	Other	chromatin organization	chromatin structure	YEPD_nucleosome_occupancy_map_dMean_log2_sMOL
					19092803	Kaplan	Eran Segal	Weizmann Institute of Science	micrococcal nuclease digestion followed by high throughput sequencing assay	DNA extract	Other	chromatin organization	chromatin structure	YPGal_nucleosome_occupancy_map_dMean_log2_sMOL
									micrococcal nuclease digestion					predicted_average

- Look for differences in nucleosome occupancy between the galactose condition and the YPD condition. Given that GAL1 and GAL10 function in galactose catabolism, do the nucleosome occupancy tracks suggest something about the regulation of GAL1 and GAL10?



To save the current display, or to share it with colleagues, simply copy and save the browser URL.

Anatomy of the Ensembl Fungi homepage

The screenshot shows the Ensembl Fungi homepage with various sections highlighted by orange boxes and arrows:

- 1: Website header**: Points to the top navigation bar.
- 2: Main search box**: Points to the main search input field.
- 3: Genome and species directory**: Points to the genome selection and species list area.
- 4: Release highlights**: Points to the "What's New in Release 56" section.
- 5: Archive sites**: Points to the "Archive sites" section at the bottom right.
- 1a: Ensembl login**: Points to the "Login/Register" button in the top right corner.
- 1b: Quick search**: Points to the search bar in the top right corner.

1: Website header

2: Main search box

3: Genome and species directory

4: Release highlights

5: Archive sites

1a: Ensembl login

1b: Quick search

1: Website header

This section is present on all pages as you navigate around Ensembl Fungi. It contains quick links to commonly used tools, help and documentation pages and download options.

- **1a: Log into your Ensembl account**

This is part of the header. By logging into Ensembl you will be able to save your queries on the Ensembl tools, such as the VEP to return to at a later date. You can also upload your own data or custom track-hubs to your account. You can also edit which species appear in section 3: Genome and species directory.

- **1b: Quick search**

This is part of the header. This enables you to quickly search all of Ensembl Fungi from whatever page you are on. You can search for a species, a gene, a region or the help and documentation.

2: Main search box

This enables you to search the whole of Ensembl Fungi, you can specify the species or strain you are interested in from the drop-down menu, or simply search across all species.

3: Genome and species directory

Here are listed the most popular species in Ensembl. If you create an account with Ensembl Fungi (See Section 1a), you can edit which species appear here. You can also click the link to view a table listing all of the genomes available.

4: Release details

A new version of Ensembl Fungi is released approximately every 3 months. This is where we add new data or update existing data, this section details the highlights of the current release. It is important to know which version you are working with as if you return to Ensembl at a later date the data may have changed.

5: Archive sites

This section lists available Ensembl Fungi web archives frozen in time in case you wanted to go back to previous releases.

Anatomy of the Ensembl Fungi gene page

1: Selected genome

2: Location and feature tabs

3: Tab-specific pages

4: Customisation and download options

5a: Feature overview

5b: Transcript table

6: Main display

1: Selected genome

This will show the species binomial Latin name and the strain/isolate identifier. Many species have more than one genome assembly. Clicking on this will take you to the information page for the genome assembly.

2: Page tabs

Tabs allow you to see the gene information at different scales and to change between views easily. The ‘Gene’ tab is selected and highlighted in white and shows specific information about the gene. The ‘Location’ tab will show the gene information at a larger scale, showing surrounding features in the genome, such as regulatory features and neighbouring genes. The ‘Transcript’ tab will show you information at a finer scale, specific to gene transcript, such as protein products and cDNA.

3: Gene-specific pages

This is the left-hand navigation panel for the gene-specific views. Here are listed pages that show gene information grouped by data type. This will stay the same for all gene pages for all genomes, but the selection will be different on the location and transcript tabs.

4: Customisation options

These buttons are present in every tab. The ‘Configure this page’ button will open a range of options to add data displays to a page, and the options change depending on the tab and page you are on. ‘Custom tracks’ allows you to add tracks from the Track-hub registry and your own data to the page you are on. ‘Export data’ allows you to download the data on the page in a range of different data formats. ‘Share this page’ will give you a stable URL which will link to the current page, but also include information about any extra data you have added using the Configure this Page or custom tracks options to share with a colleague. ‘Bookmark this page’ will save this page if you have an Ensembl account.

5: Feature overview including the transcript table

These will be present on all pages in the Gene and Transcript tabs. You can hide or show the transcript table by clicking on the brown button above the table

6: Main display

Main display presents data specific to the chosen page. Toggling between gene-specific pages will update this view.

Exercise: Searching Ensembl Fungi species

Clickable links shown in blue, text to be entered shown in red.

Navigate to fungi.ensembl.org. You'll see a homepage similar to this:

Click on 'View full list of all species', which you can find in section 3: Genome and species directory shown above.

Anatomy of the Ensembl Fungi homepage

The screenshot shows the Ensembl Fungi homepage with various sections highlighted by orange boxes and labeled with numbers 1 through 5:

- 1: Website header**: Points to the top navigation bar with links to HMMER, BLAST, BioMart, Tools, Downloads, Help & Docs, and Blog.
- 2: Main search box**: Points to the main search input field with placeholder text "All species" and a "Go" button.
- 3: Genome and species directory**: Points to the "All genomes" section, which includes a dropdown menu "Select a species" and a list of favorite genomes: *Saccharomyces cerevisiae* R64-1-1, *Schizosaccharomyces pombe* ASM294v2, *Aspergillus nidulans* ASM1142v1, *Puccinia graminis* ASM14992v1, *Magnaporthe oryzae* MG8, and *Zymoseptoria tritici*.
- 1a: Ensembl login**: Points to the "Log in/Register" link in the top right corner.
- 1b: Quick search**: Points to the search bar with placeholder text "Search Ensembl Fungi..." and a magnifying glass icon.
- 4: Release highlights**: Points to the "What's New in Release 56" section, which lists updates for genomes, updated protein features, BioMarts, and pan-taxonomic gene trees.
- Ensembl Rapid Release**: Points to the "Ensembl Rapid Release" section, which provides new assemblies every two weeks and notes that existing species will be updated with full range of annotations.
- 5: Archive sites**: Points to the "Archive sites" section, which lists previous versions of data available at archive.ensembl.org, specifically Release 52 (December 2011) and 53 (April 2012).

(a) How many genome assemblies are there for the genus *Bipolaris* in Ensembl Fungi?

Click on the Latin name of your species of interest to go to the species homepage.

(b) Navigate to the species homepage for *Bipolaris oryzae*. What is the name of the genome assembly for *B. oryzae*?

To find out more about the genome assembly and gene annotation, click on [More information and statistics](#).

(c) How long is the *B. oryzae* genome? How many genes have been annotated?

(d) What is the INSDC accession number for *B. oryzae*? What institute submitted the data to INSDC?

Bipolaris oryzae ATCC 44560 Assembly and Gene Annotation

About Bipolaris oryzae ATCC 44560 (GCA_000523455)

Cochliobolus miyabeanus (formerly known as *Helminthosporium oryzae*) is a fungus that causes brown spot disease in rice. This disease was the causal agent of the Bengal famine of 1943. It was used by the USA as a biological weapon in Japan during World War II.

[Text](#) and [Image](#) from Wikipedia, the free encyclopaedia.

Assembly

The assembly presented is the Cochliobolus miyabeanus v1.0, INSDC Assembly GCA_000523455.1. It was derived from [INSDC](#) with the assembly accession [GCA_000523455.1](#).

Annotation

The annotation presented is derived from annotation submitted to [INSDC](#) with the assembly accession [GCA_000523455.1](#), with additional non-coding genes derived from [Rfam](#). For more details, please visit [INSDC annotation import](#).

More information

General information about this species can be found in [Wikipedia](#).

Ensembl Fungi release 56 - Feb 2023 © [EMBL-EBI](#)

Statistics

Summary

Assembly

Cochliobolus miyabeanus v1.0, INSDC Assembly	
GCA_000523455.1	
Database version	109.1
Golden Path Length	31,362,097
Genebuild by	JGI
Genebuild method	Import
Data source	JGI

Gene counts

Coding genes	12,002
Gene transcripts	12,002

Hover over underlined terms to see description

ENA
European Nucleotide Archive

Enter text search terms Search

Examples: histone, BN000065

GCA_000523455.1

Examples: Taxon:9606, BN000065, PRJEB402

Assembly: GCA_000523455.1

The family Pleosporaceae, represents a group of diverse plant pathogens, and also includes saprobic taxa. Sequencing the proposed Cochliobolus genomes will strengthen comparative genomic approaches. Through sequencing efforts, the individual biology of these additional Cochliobolus species will be better understood so that breeders can make a more informed decision about the deployment of resistance. The advanced knowledge of pathogenesis will serve to enhance breeding efforts.

Comment

URL -- <http://genome.jgi.doe.gov/Cocmi1~JGI> Project ID: 403761~The DNA was provided by Turgeon, B. Gillian(bgt1@cornell.edu)~The strain is available from ATCC culture collection (44560)~Assembly and annotation done by JGI~The JGI and collaborators endorse the principles for the distribution and use of large scale sequencing data adopted by the larger genome sequencing community and urge users of this data to follow them. It is our intention to publish the work of this project in a timely fashion and we welcome collaborative interaction on the project and analysis.~(<http://www.genome.gov/page.cfm?pageID=10506376>)

Show Less

Organism: [Bipolaris oryzae ATCC 44560](#)

Accession: [GCA_000523455](#)

Assembly Level: scaffold

Strain: ATCC 44560

Genome Representation: full

View: XML

Download: XML

Sequence Report

WGS SET EMBL

WGS SET FASTA

Navigation: Show

Additional Attributes: Show

BlobToolKit: Hide

Assembly Statistics: Show

WGS Sequence Set: [AMCO01](#)

Exercise: Ensembl Fungi ‘Region in detail’ view

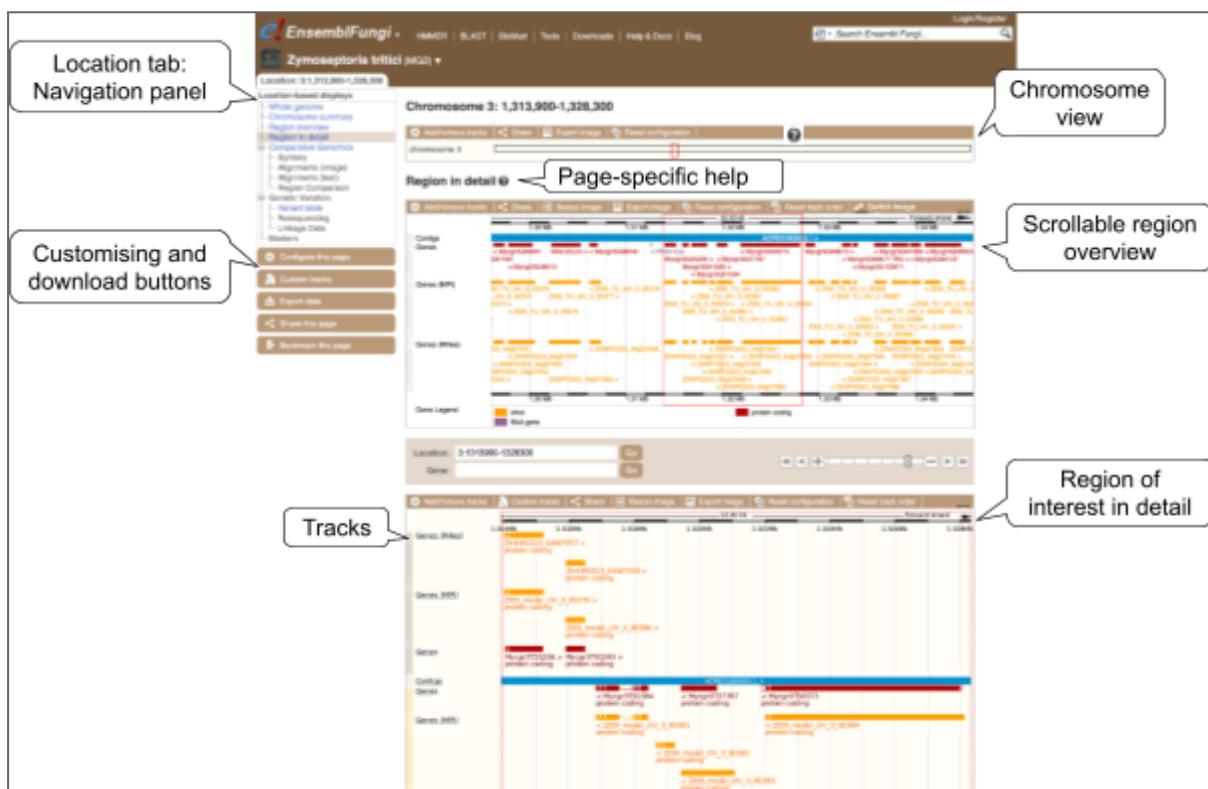
Start at the Ensembl Fungi front page, fungi.ensembl.org. You can search for a region by typing it into a search box, but you have to specify the species.

- (a) Find *Zymoseptoria tritici* (assembly MG2), then type (or copy and paste) **3:1313900-1328300** into the search box. Press enter or click **Go** to jump directly to the **Region in detail** Page.

Search: Zymoseptoria tritici
3:1313900-1328300 Go
e.g. **NAT2** or **alcohol***

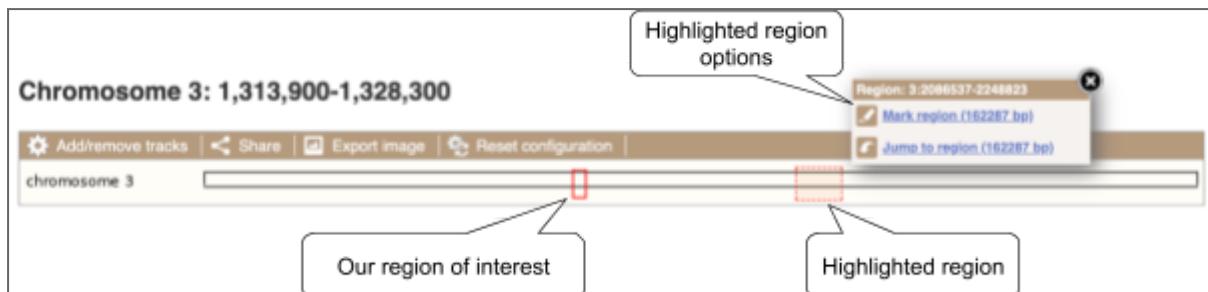
Click on the button to view page-specific help. The help pages provide links to Frequently Asked Questions, a Glossary, Video Tutorials, and a form to Contact Helpdesk. There is a help video on this page at <http://youtu.be/tTKEvgPUq94>.

The Region in detail page is made up of three images, similar to this:



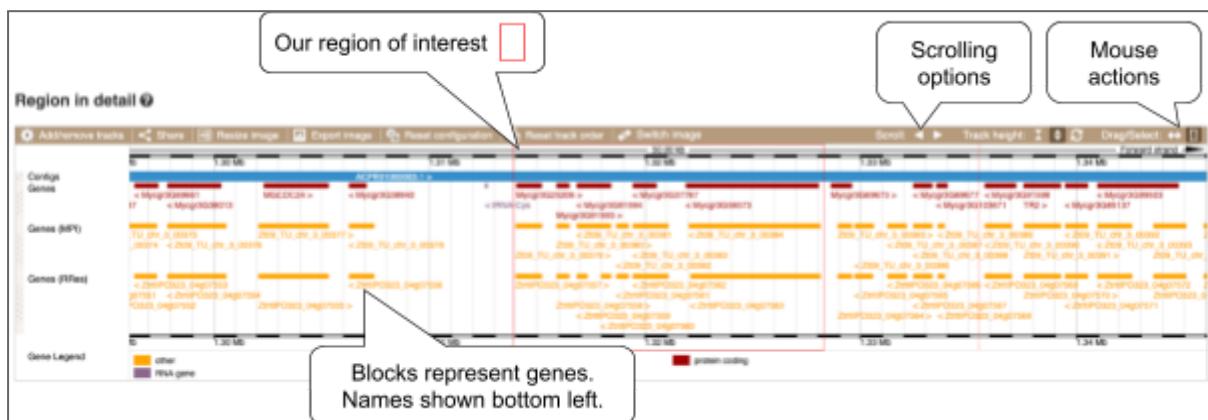
Let's look at each image in detail.

The first image shows the chromosome. You can jump to a different region by dragging out a box in this image. Drag out a box on the chromosome; a pop-up menu will appear.

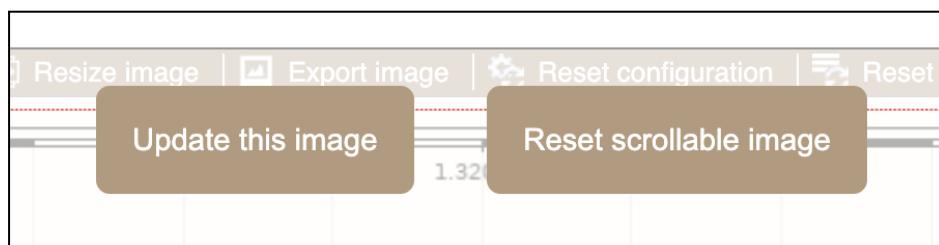


If you would like to move to the region, you could click on [Jump to region \(### bp\)](#). To highlight it, click on [Mark region \(### bp\)](#). For now, we'll close the pop-up by clicking on the [X](#) on the corner.

The second image shows a 50 kb region around our selected region. This view allows you to scroll back and forth along the chromosome.



Click on the [Drag/Select](#) button  to change the action of your mouse click. Now you can scroll along the chromosome by clicking and dragging within the image. As you do this, you'll see the image below grey out and two blue buttons appear. Clicking on [Update this image](#) would jump the lower image to the region central to the scrollable image. We want to go back to where we started, so we'll click on [Reset scrollable image](#).



The third image is a detailed, configurable view of the region similar to this:



Genes are shown as transcripts with exons represented as boxes and introns shown as lines connecting exons. Forward stranded genes are shown above the genome assembly (Contigs track), while reverse stranded genes are shown below.

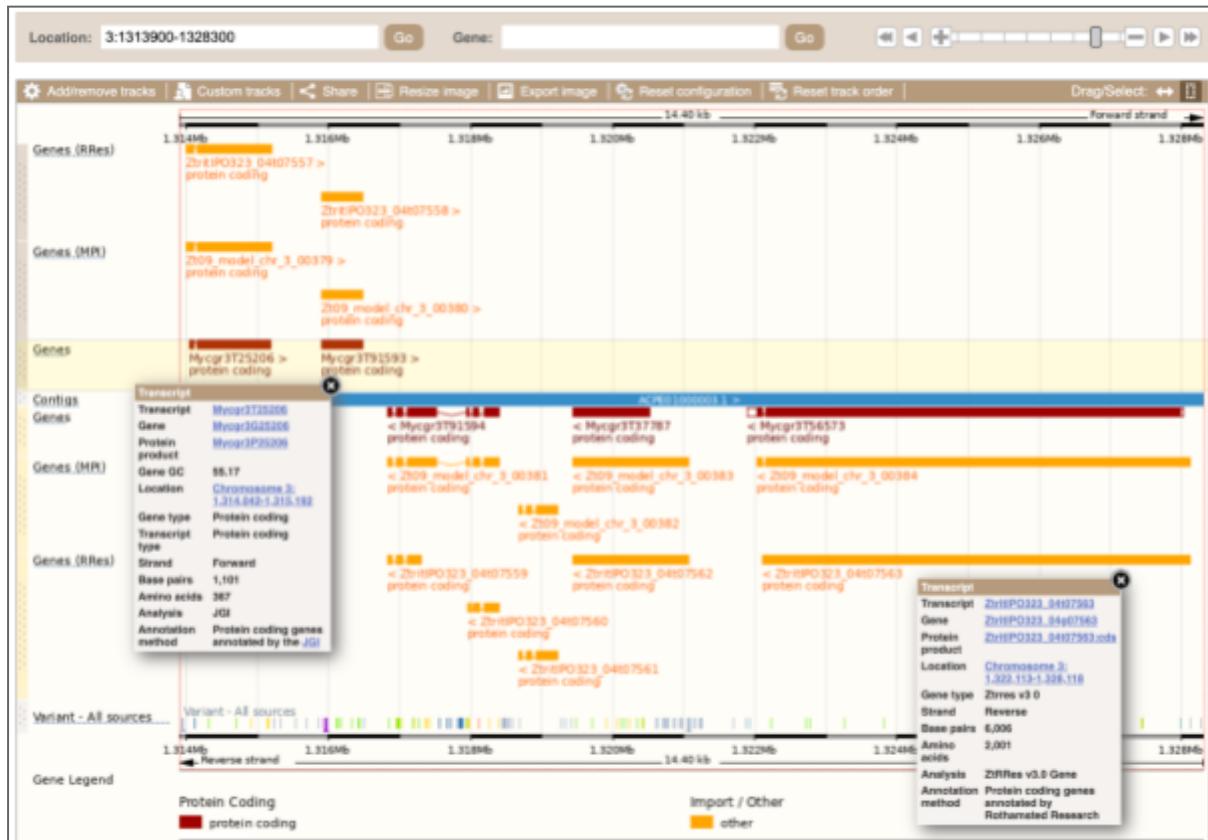
Click on the **Drag>Select** option at the top or bottom right to switch mouse action. On **Drag**, you can click and drag left or right to move along the genome, the page will reload when you drop the mouse button. On **Select** you can drag out a box to highlight or zoom in on a region of interest.

With the tool set to **Select**, drag out a box around an exon and choose **Mark region**.



The highlight will remain in place if you zoom in and out or move around the region. This allows you to keep track of regions or features of interest.

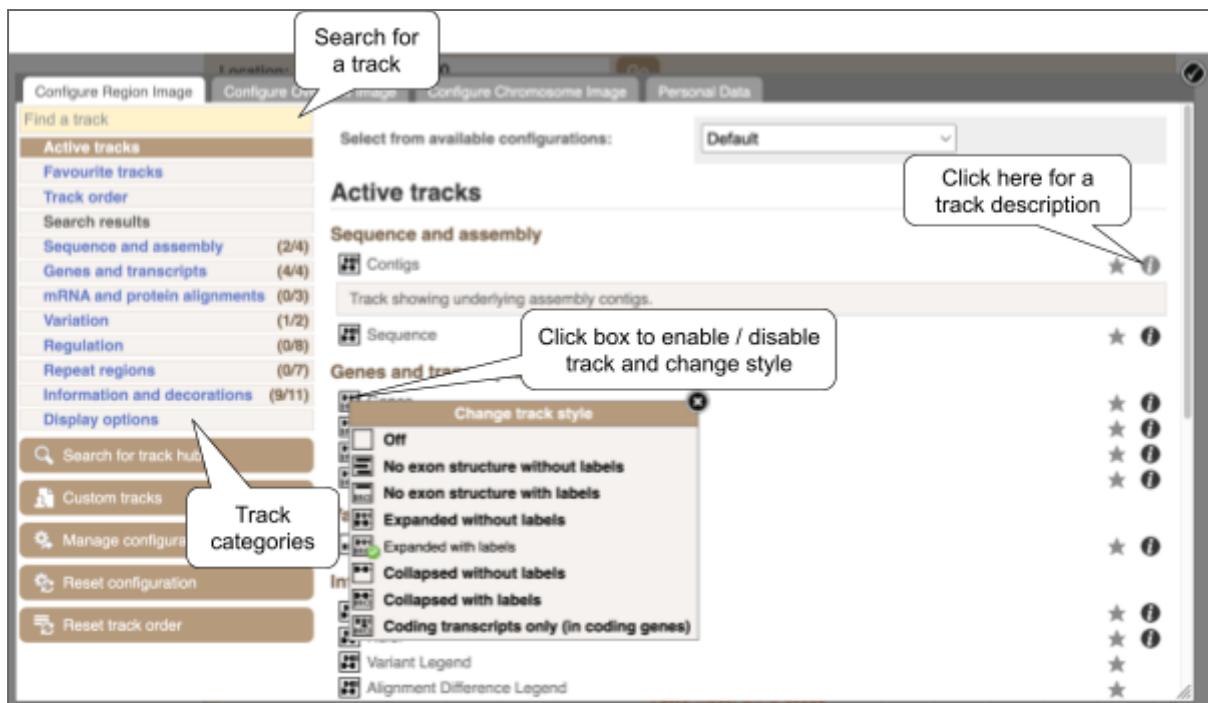
(b) How many genes are annotated in the current region? How many on the forward and how many on the reverse strand? Are they all annotated by the same institute?



We can edit what we see on this page by clicking on the [Configure this page](#) menu at the left.

 [Configure this page](#)

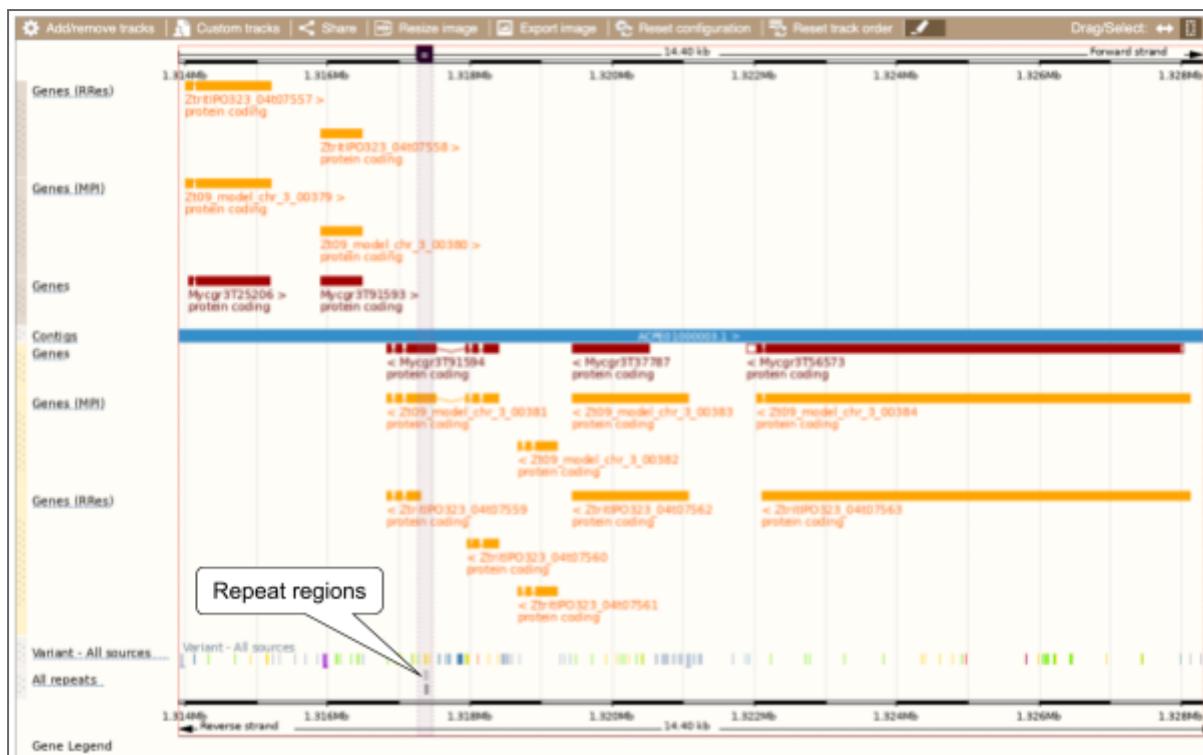
This will open a menu that allows you to change the image. You can put some tracks on in different styles; more details are in this FAQ: <http://www.ensembl.org/Help/Faq?id=335>.



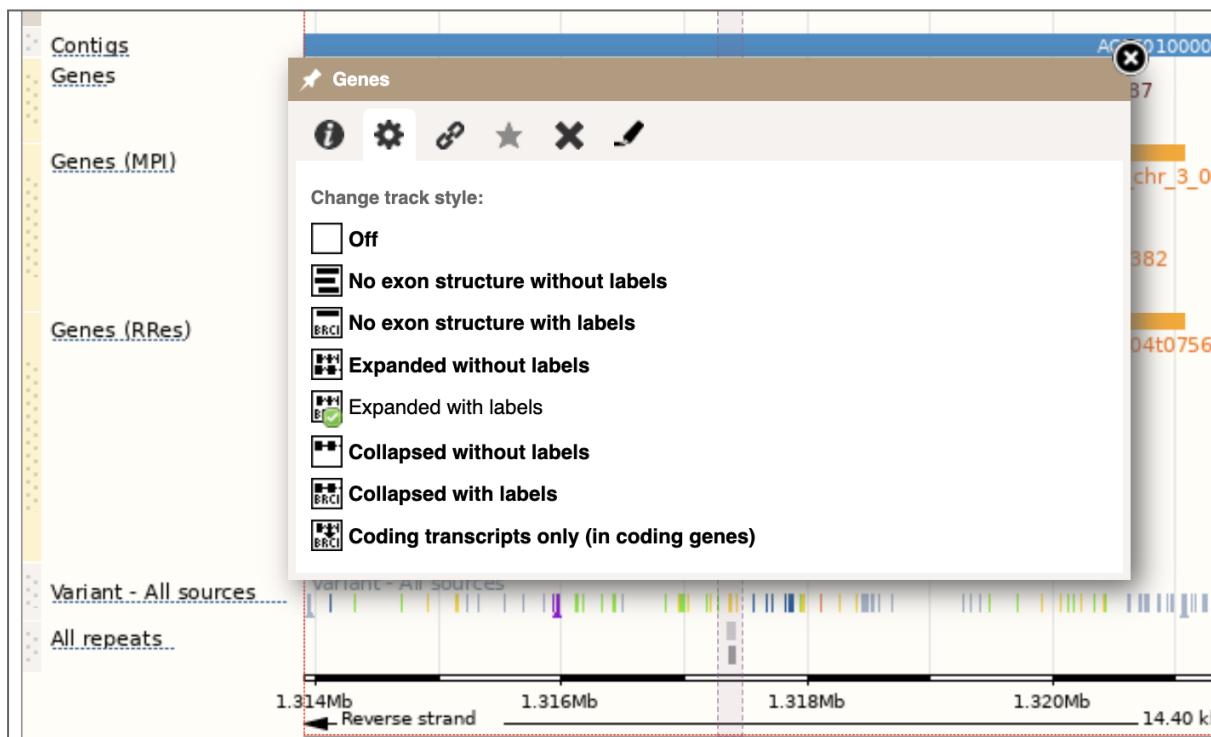
You can add a track to the image by clicking on the box and choosing track style. Click on the tick in the top right hand tick to save and close the menu. Alternatively, click anywhere outside of the menu.

Let's add some tracks to this image.

(c) Turn on the **All repeats** track. Are there any repeat regions identified in this region? Do they overlap any of the genes?



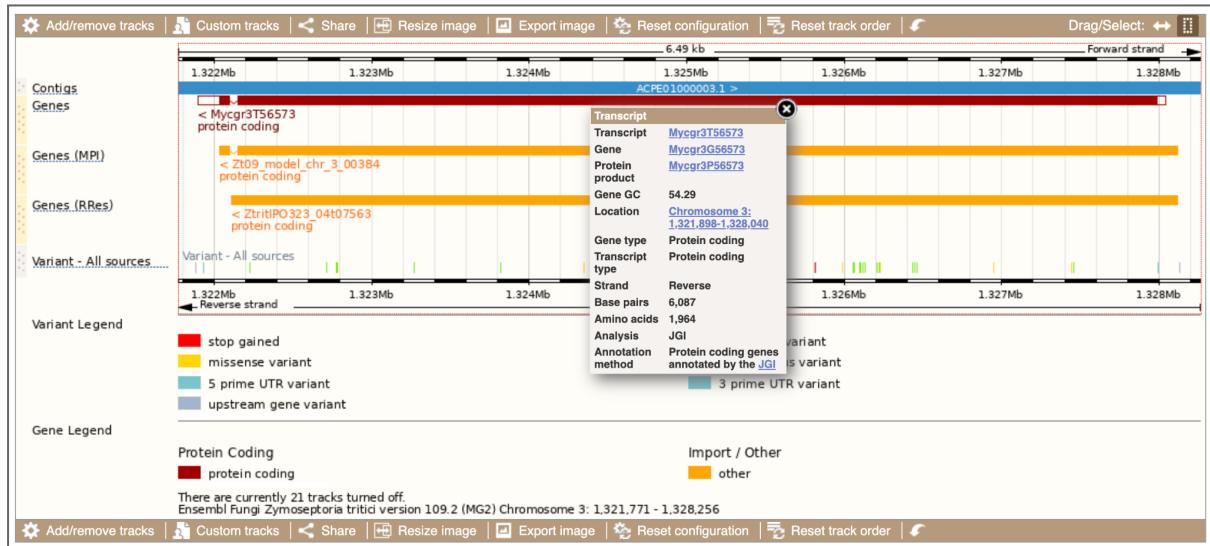
We can also change the way the tracks appear by clicking on the track name to open a menu.



We can move tracks around by clicking and dragging on the coloured dotted block/bar to the left of the track name.

E.g.

(d) Zoom in on the largest transcript [Mycgr3T565738](#). How many exons does this gene have?



Now that you've got the view how you want it, you might like to show something you've found to a colleague or collaborator. Click on the [Share this page](#) button to generate a link.

Share this page

Email the link to someone else, so that they can see the same view as you, including all the tracks you've added. These links contain the Ensembl release number, so if a new release or even assembly comes out, your link will just take you to the archive site for the release it was made on.

(d) Export the genomic sequence for this region

Export data

Export data

Logout 

Export Configuration - Feature List

Location to export: chromosome:MG2:3:1321771:1328256:1

Output: FASTA sequence *

Select location: 3 * 1321771 * 1328256 * 1

5' Flanking sequence (upstream): 0 * (Maximum of 1000000)

3' Flanking sequence (downstream): 0 * (Maximum of 1000000)

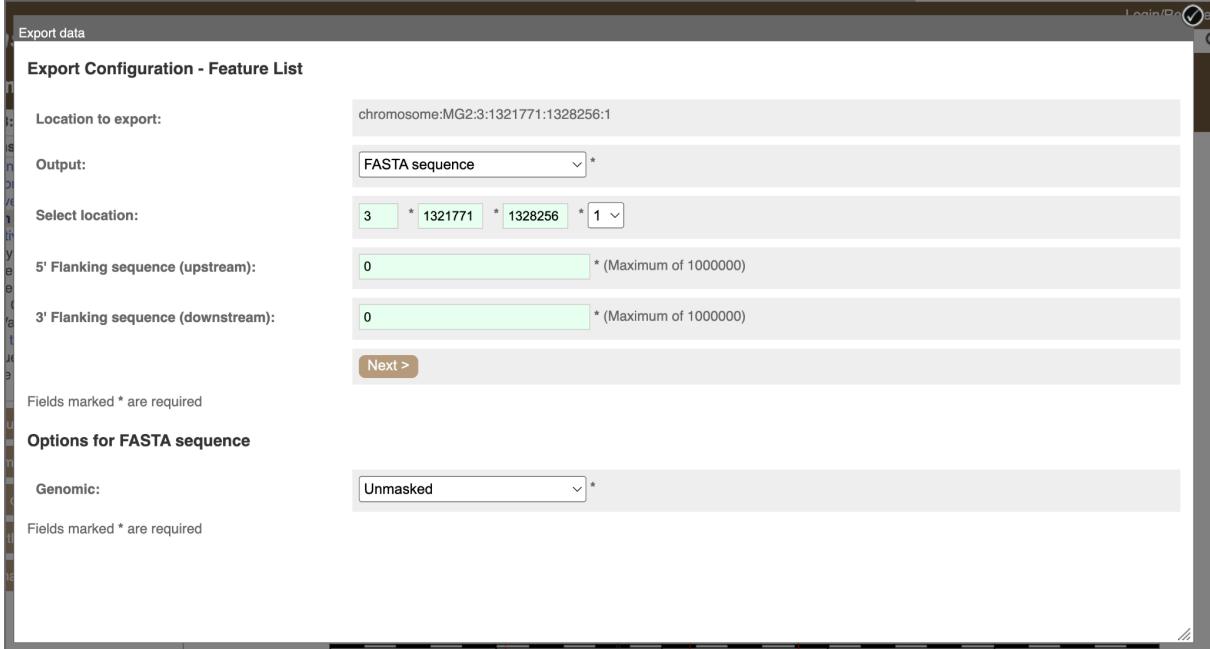
Next >

Fields marked * are required

Options for FASTA sequence

Genomic: Unmasked *

Fields marked * are required



To return this to the default view, go to [Configure this page](#) and select [Reset configuration](#) at the bottom of the menu.

Exercise: Ensembl Fungi gene and transcript tabs

We're going to look at the gene *LEUC* in *Zymoseptoria tritici* (assembly MG2). This gene is involved in leucine biosynthetic process.

From fungi.ensembl.org, type *LEUC* into the main search box, click the drop-down menu and select *Z. tritici* and click the Go button.

Click on the gene ID [Mycgr3G103221](#) in the results. The **Gene tab** should open, similar to this:

Anatomy of the Ensembl Fungi gene page

1: Selected genome

2: Location and feature tabs

3: Tab-specific pages

4: Customisation and download options

5a: Feature overview

5b: Transcript table

6: Main display

The *LEUC* gene is highlighted in green and in the centre of the display as it is the gene of interest.

(a) On which chromosome and which strand of the genome is this gene located?

Gene: LEUC Mycgr3G103221

Description

3-isopropylmalate dehydrogenase [Source:UniProtKB/Swiss-Prot;Acc:[Q9Y897](#)]

Location

[Chromosome 2: 1,581,329-1,582,830](#) reverse strand.

MG2:ACPE01000002.1

Let's walk through some of the links in the left hand navigation column. How can we view the genomic sequence? Click **Sequence** at the left of the page.

Marked-up sequence ②

[Download sequence](#)

[BLAST this sequence](#)

Download / BLAST sequence

Exons LEUC exons All exons in this region

Markup loaded

```
>chromosome:MG2:2:1580729:1583430:-1
GTAGCGAATGCTTCCAGGCTGATGTCGGAGCTGGCAGTTGCGAAGTAGCTCGTCCAG
GGTCAAGATCTCGTCAGAGTGGAGTGCCTGCGCTTACGGCGACTCGGGATCCAA
TATGTCGCAATCCACCGGGCTTGTGGCGAATTCTCGCGCTGAATGCAAGCGCTG
ATTCAGTTTCGAGGCGATCTGGTCTCAATCGATTGACTCGGTTGAGTGTCAAGGTGAG
CGCCAGTGCTGCCATTGCTGGAGGCTGCACTTCTGCTGATGGTCGCC
CGCGAAGTCCAGTAAAGGAAGATGCACTGGCGATCCCAGACCTGCG
TTCCCACGAAGCGGTTGAAAGCATGGCACTGCCATCCCAGACCTGCG
CTAGATATCCATGCGATGTTGAGTAAGAGATGAGGGAGTTCTGGGA
GTGGATGGTGGAACCTTGCTCCCCGGATGTGTCGGAGGGCAGCGAAG
ATCTCGACCGCTCTCCAATCCACGTAACAAATTCTTGGCAGAACATGACCATACAAAC
AATGAAATCGTCGAAAAATGCCACTTACAACATCGTCGTTTGGAGTGAGTGCCTA
TGCCCCGGCAATTGAAATCATGCTGATGTCGATTCAAGGTGATCACTGCCGCTCTGAAGT
AAGCAGCTCTTACCGTACTCACTCGTAGCAGCTCAAGCTAACGATTACAGGTCACC
GCCGAGGCCTGAAGGTATGCCATACAAGTAC
AAAGCATAACAGGTGCTCGACGTGATCGACAAC
AACCAACCTGCTCGGTTGGCTACGTTGCTCC
AACCGAAGCAGGCCTCGATCGATGCCACGGCGAGCC
CTGCCAACAGCAGCCGACGCCGTGATCCCTGGAGCCAT
GCAAAGTCCGTCCAGAGCAAGGCATCCGCGCTCCG
ACCTGCGCCATGCTTCTCGCGCTGAGAGCTCGTCAGAAGACTT
AAGTCTGCCGGCGCTCAACTCAACATCGTCGTGAGCTGACAGC
GTGAGCGACCGAGGACGACGGCTCCGGTACCGAGTGGACACGGAACCCCTACTCGCGCG
CTGAGATCGAGCGTGTGGCACGACTGGCGGTTCTGGCTTGGCTGAAGACCCGCCGT
```

Upstream sequence
(600bp)

First exon of
LEUC gene

Highlight region to BLAST
specific sequence

[BLAST selected sequence](#)

The sequence is shown in FASTA format. Take a look at the FASTA header:

```
>chromosome:MG2:2:1580729:1583430:-1
```

The FASTA header follows this format:

Genome assembly:Chromosome:Base pair start coordinate:Base pair end coordinate:Strand
Forward strand denoted by 1, and reverse strand by -1.

Exons are highlighted within the genomic sequence. If you click on [Configure this page](#) you can change display options, and for species with variation databases you can highlight variants on this view.

(b) Use the Configure this page option to show variants on the sequence and also the line numbering relative to the coordinate system, to this view.

- Are all exons shown in this display part of the *LEUC* gene? How can you tell?

	Exons	LEUC exons	All exons in this region			
Variants	3 prime UTR	Intronic	Missense	Stop gained	Synonymous	Upstream
Markup	loaded					
1582050	GTCCGGTCTGGTCGCTCGACAAGGCGAACGT CATGGCTACAAGCAGACTGTGGCGCAAGA					1581991
1581990	CCGTGACGGATGTGTTGCCAACGAGTTCCCGCAGTTGAAGATCGGGCATCACCTCATTG					1581931
1581930	ATTCCGCGGCCGATGTTGATGGTC AAGAACCCGAGAGCGTTGAACCGAGTCATCGTCACTA					1581871
1581870	GCAATTGTTCGGCATATCATTAGT GACGAGGCGAGCGTCATTCAGGATC K CTGGTT					1581811
1581810	TGTTGCCCAAGTGCAGTTGACTGCCTCGCCGGATGGCAAGAGCAAGTGCATGGCATT					1581751
1581750	ACGAGCCGATTACCGTATGTACAATGTTGCTTGATGCGAGGTCATTGTTGCTGACAA					1581691
1581690	CCATCAG GCTCTGCACCAGACATCAGCGCAAGGGTATCGTCAACCCCGTCGCCATGATC					1581631
1581630	TTGTCTTGGGCATGATGTGCAAGTACTCTCTCCAGCAGCCGAGCTGGCCAAGAAGATC					1581571
1581570	GACGAGGCCGTGAGGAACGTAATTGAGAAGGGTATCAACACTGCCGATATTGGTGGTTCG					1581511
1581510	GCCAAGACTGCCGAGGTGGGTGA X GCTATTGCGAAGGAGCTCGAGGCCCTGCTGAAGTAA					1581451
1581450	ACGGTGCAACTKGGTACTGG S TACACTCGCGTTCGCGGAGTGTGGACAAGAGGATGGC					1581391
1581390	GCAAGGTCAAACAAAAGTTGTAAGTGAAGCTTCTGGG I TTGAACATAAX X TCTCTAT					1581331
1581330	AGAGGAACCTTGATGCACGAAATGGAACAAACCTGGTCGTAATGATCACGAAGGT CAC					1581271
1581270	CATCGACCTAGAGATTGCCCTCGCCGGTGCCGGTGCCTGCTTGTATGCTATCACAGAGA					1581211
1581210	CAGTTCAGGACAGGTCTCGCTCGTTGT C TCCTACTGCCACAGTGGCGCAACGGTAT					1581151
1581150	CCACCGCCCTCTCCACCACTGCTCTCACCGATCTCCAACGTCGGTCCCACACTCCA					1581091
1581090	CCTCCAACCTCTCCAACCACACCTTCGCCACTGCC T TCACCTTGATCTCAGCCGG					1581031
1581030	GCACCTGGAAATGCCAGCGCATGCAGCCTCGTCAACTCCGCCCTCAGCCTCTCCCT					1580971
1580970	CCACCTCGTCGACTTCTCCTCTCCGTCTGCACCATTCTCA S AGCAGGCATTTCG					1580911
1580910	GCCTCGGATTGGACCCACACAAATCCTCCCCAAACTGCCCTCAAGAAACATACACAACC					1580851
1580850	TCCCCAACCTTCCTCAGGACTCACATTGCATGCCATTGGGATTCTGCCCTTCCCAT					1580791
1580790	CCCCATTGACACCATTTCTCGACTGGGT C AT T YCAAAGTGTGCGCATGTTGTC					1580731
1580730	GG					1580729

- Can you find the Stop Gained mutation? What letter is it represented by? What nucleotides does it stand for? (these are [IUPAC ambiguity codes](#))

	Exons	LEUC exons	All exons in this region			
Variants	3 prime UTR	Intronic	Missense	Stop gained	Synonymous	Upstream
Markup	loaded					
1582590	AAAGCATACAG GTGCTCGACGTGATCGACAACAGCAATGCCGACGTTCACTTCAACATT C					1582531
1582530	AACCAACACTGCTCGGTGGC GTACGTTGCTCCGACCCACCTTCCGAAAAGCCAGGGCT					1582471
1582470	AACCGAACG AGCAGCGCTG M T CGATGCCAACGGCGAGCCTCTAACAGACGAAGC ACTGCCG					1582411
1582410	CTGCCAAAS CAGCCGACGCCGTGATCCTCGGGAGCCATGGCGGCCAAATGGGGCACGG					1582351
1582350	GCAAAGTCCGTCCAGAGCAAGGCATCCTCGCGCTCCGAAAGAAATGGGCACATATGGAA					1582291
1582290	ACCTGCGCCCATGCTTCTCGCGTCTGAGAGCCTCGTCAAGACTCTCCATTGAAGGAGG					1582231

- Which exon does the stop gained mutation fall in?

You can download this sequence by clicking in the [Download sequence](#) button above the sequence.

This will open a dialogue box that allows you to pick between plain [FASTA sequence](#), or [sequence in RTF](#), which includes all the coloured annotations and can be opened in a word processor. This button is available for all sequence views.

Download sequence

File name: Zymoseptoria_tritici_LEUC_sequence

-- Choose Format --

- FASTA
- RTF (Word-compatible)

Preview Download Download Compressed

Settings

Sequences to export:

- Select/deselect all
- cDNA (transcripts)
- Coding sequences (CDS)
- Amino acid sequences
- 5' UTRs
- 3' UTRs
- Exons
- Introns
- Genomic sequence

5' Flanking sequence (upstream): 600 * (Maximum of 1000000)

3' Flanking sequence (downstream): 600 * (Maximum of 1000000)

(c) Export this sequence in RTF format

If we are interested in finding out about gene functions, the Gene Ontology (GO) annotations can tell us where the protein is located, the biological processes it is involved in and its molecular function.

(d) What biological processes have been associated with *LEUC*?

Click on [GO: Biological process](#). This page shows all linked GO annotations, some of these are linked as GO terms are hierarchical. For example, if you click on 'leucine biosynthetic process' GO accession number, you will be taken to the GO pages, which shows that this is a child term to 'cellular amino acid biosynthetic process' which is also shown on the GO pages in Ensembl.

GO: Biological process ?

Show/hide columns						Filter
Accession	Term	Evidence	Annotation source	Mapped using	Transcript IDs	
GO:0006097	glyoxylate cycle	IEA	Ensembl	Propagated from Saccharomyces_cerevisiae YCL018W by orthology	Mycgr3T103221	<ul style="list-style-type: none"> Search BioMart View on karyotype
GO:0008652	cellular amino acid biosynthetic process	IEA	UniProt	UniProtKB/TrEMBL:F9X2A6 MYCGM	Mycgr3T103221	<ul style="list-style-type: none"> Search BioMart View on karyotype
GO:0009082	branched-chain amino acid biosynthetic process	IEA	UniProt	UniProtKB/TrEMBL:F9X2A6 MYCGM	Mycgr3T103221	<ul style="list-style-type: none"> Search BioMart View on karyotype
GO:0009098	leucine biosynthetic process	IEA		InterPro:Isopropylmalate_DH, UniProtKB/TrEMBL:F9X2A6 MYCGM	Mycgr3T103221	<ul style="list-style-type: none"> Search BioMart View on karyotype
Inferred from Electronic Annotation						

For some pathogenic species in Ensembl Fungi we have Pathogen-Host Interactions (PHI-base) annotations which can be found under ‘Ontologies’: [PHI: Phibase identifier](#) link in the left-hand menu, e.g. [ATG8](#) gene in *Magnaporthe oryzae* (MG8) has four entries here, listed by the PHI-base ID. Click on the link to go to the PHI-base website to view more information about this annotation.

PHI: Phibase identifier

Accession	Term	Evidence	Annotation source	Mapped using	Transcript IDs																
PHI:2061	2061	ND	Sequence Publications:19115483		MGG_01062T0	<ul style="list-style-type: none"> Search BioMart View on karyotype 															
PHI:2076	2076	ND	Sequence Publications:10717456		MGG_01062T0	<ul style="list-style-type: none"> Search BioMart 															
PHI:2139	2139			<table border="1"> <thead> <tr> <th>Pathogen Gene</th> <th>Mutant Phenotype</th> <th>Pathogen Species</th> <th>Disease</th> <th>Host Species</th> </tr> </thead> <tbody> <tr> <td>Gph1</td> <td>loss of pathogenicity</td> <td><i>Magnaporthe oryzae</i></td> <td>Rice Blast</td> <td><i>Hordeum vulgare</i> (related: barley)</td> </tr> <tr> <td>ATG8</td> <td>loss of pathogenicity</td> <td><i>Magnaporthe oryzae</i></td> <td>Rice Blast</td> <td><i>Hordeum vulgare</i> (related: barley)</td> </tr> </tbody> </table>	Pathogen Gene	Mutant Phenotype	Pathogen Species	Disease	Host Species	Gph1	loss of pathogenicity	<i>Magnaporthe oryzae</i>	Rice Blast	<i>Hordeum vulgare</i> (related: barley)	ATG8	loss of pathogenicity	<i>Magnaporthe oryzae</i>	Rice Blast	<i>Hordeum vulgare</i> (related: barley)		
Pathogen Gene	Mutant Phenotype	Pathogen Species	Disease	Host Species																	
Gph1	loss of pathogenicity	<i>Magnaporthe oryzae</i>	Rice Blast	<i>Hordeum vulgare</i> (related: barley)																	
ATG8	loss of pathogenicity	<i>Magnaporthe oryzae</i>	Rice Blast	<i>Hordeum vulgare</i> (related: barley)																	
PHI:768	768			<table border="1"> <thead> <tr> <th>Pathogen Gene</th> <th>Allele</th> <th>Pathogen</th> <th>Host</th> </tr> </thead> <tbody> <tr> <td>Gene:ATGB PHI-base entry:Phi:2061 Gene ID:XP_368182 Protein ID: Q51MW4 Sequence strain:70-15</td> <td></td> <td>Pathogen species:<i>Magnaporthe oryzae</i> Pathogen ID: 318829 Pathogen strain:B157</td> <td>Host species:<i>Hordeum vulgare</i> (related: barley) Host classification:Monocots Host ID: 4513 Host strain:subsp. <i>vulgare</i> (related: domesticated barley) Tissue:leaf</td> </tr> </tbody> </table>	Pathogen Gene	Allele	Pathogen	Host	Gene:ATGB PHI-base entry:Phi:2061 Gene ID:XP_368182 Protein ID: Q51MW4 Sequence strain:70-15		Pathogen species: <i>Magnaporthe oryzae</i> Pathogen ID: 318829 Pathogen strain:B157	Host species: <i>Hordeum vulgare</i> (related: barley) Host classification:Monocots Host ID: 4513 Host strain:subsp. <i>vulgare</i> (related: domesticated barley) Tissue:leaf									
Pathogen Gene	Allele	Pathogen	Host																		
Gene:ATGB PHI-base entry:Phi:2061 Gene ID:XP_368182 Protein ID: Q51MW4 Sequence strain:70-15		Pathogen species: <i>Magnaporthe oryzae</i> Pathogen ID: 318829 Pathogen strain:B157	Host species: <i>Hordeum vulgare</i> (related: barley) Host classification:Monocots Host ID: 4513 Host strain:subsp. <i>vulgare</i> (related: domesticated barley) Tissue:leaf																		
				<table border="1"> <thead> <tr> <th>Reference</th> <th>Comments</th> <th>PHI Phenotype</th> <th>Pathogen Phenotype</th> </tr> </thead> <tbody> <tr> <td>Pmid:19115483 Ref source:Pubmed Year:2009 Author reference:Yi Zhen Deng</td> <td></td> <td>Phenotype:loss of pathogenicity Disease name:Rice Blast Tissue:leaf Host response:light induction of hypersensitive reaction Experimental technique:Gene deletion: full</td> <td></td> </tr> </tbody> </table>	Reference	Comments	PHI Phenotype	Pathogen Phenotype	Pmid:19115483 Ref source:Pubmed Year:2009 Author reference:Yi Zhen Deng		Phenotype:loss of pathogenicity Disease name:Rice Blast Tissue:leaf Host response:light induction of hypersensitive reaction Experimental technique:Gene deletion: full										
Reference	Comments	PHI Phenotype	Pathogen Phenotype																		
Pmid:19115483 Ref source:Pubmed Year:2009 Author reference:Yi Zhen Deng		Phenotype:loss of pathogenicity Disease name:Rice Blast Tissue:leaf Host response:light induction of hypersensitive reaction Experimental technique:Gene deletion: full																			
				<table border="1"> <thead> <tr> <th>+ ATGB</th> <th>loss of pathogenicity</th> <th><i>Magnaporthe oryzae</i></th> <th>Rice Blast</th> <th><i>Hordeum vulgare</i> (related: barley)</th> </tr> </thead> </table>	+ ATGB	loss of pathogenicity	<i>Magnaporthe oryzae</i>	Rice Blast	<i>Hordeum vulgare</i> (related: barley)												
+ ATGB	loss of pathogenicity	<i>Magnaporthe oryzae</i>	Rice Blast	<i>Hordeum vulgare</i> (related: barley)																	

Back to the *LEUC* gene in *Z. tritici* (MG2). Let’s explore the transcript tab now. Many genes have multiple transcripts which can be seen in the transcript table. Click on

Show transcript table

We can go to the transcript tab either by clicking on the transcript ID, [Mycgr3T103221](#), in the table, or on the transcript tab at the top of the page.

Gene: LEUC Mycgr3G103221

Description	3-isopropylmalate dehydrogenase [Source:UniProtKB/Swiss-Prot;Acc: Q9Y897]																							
Location	Chromosome 2: 1,581,329-1,582,830 reverse strand. MG2:ACPE01000002.1																							
About this gene	This gene has 1 transcript (splice variant), 334 orthologues and 7 paralogues .																							
Transcripts	Show transcript table <table border="1"> <thead> <tr> <th>Name</th> <th>Transcript ID</th> <th>bp</th> <th>Protein</th> <th>Biotype</th> <th>UniProt</th> <th>RefSeq</th> <th>Flags</th> </tr> </thead> <tbody> <tr> <td>-</td> <td>Mycgr3T103221</td> <td>1237</td> <td>365aa</td> <td>Protein coding</td> <td>F9X2A6</td> <td>-</td> <td>Ensembl Canonical</td> </tr> </tbody> </table>								Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags	-	Mycgr3T103221	1237	365aa	Protein coding	F9X2A6	-	Ensembl Canonical
Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags																	
-	Mycgr3T103221	1237	365aa	Protein coding	F9X2A6	-	Ensembl Canonical																	

You are now in the Transcript tab on the summary page. Some summary information about the number of exons, length etc is shown at the bottom of the page under the diagram.

[e! Ensembl Fungi](#) • HMMER | BLAST | BioMart | Tools | Downloads | Help & Docs | Blog | Login/Register | [Search Ensembl Fungi...](#)

Zymoseptoria tritici (MG2) • Location: 2:1,581,329-1,582,830 • Gene: LEUC • Transcript: Mycgr3T103221

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- Protein Information
 - Protein summary
 - Domains & features
 - Variants
 - PDB 3D protein model
 - AlphaFold predicted model
- Genetic Variation
 - Variant table
 - Variant image
 - Population comparison
 - Comparison image
- External References
 - General identifiers
 - Oligo probes
- Supporting evidence
- ID History
 - Transcript history
 - Protein history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Transcript: Mycgr3T103221

Description: 3-isopropylmalate dehydrogenase [Source:UniProtKB/Swiss-Prot;Acc:[Q9Y897](#)]

Location: [Chromosome 2: 1,581,329-1,582,830](#) reverse strand.

About this transcript: This transcript has [6 exons](#), is annotated with [10 domains and features](#) and is associated with [125 variant alleles](#).

Gene: This transcript is a product of gene [Mycgr3G103221](#) [Show transcript table](#)

Summary

Export image

Exons: 6, Coding exons: 6, Transcript length: 1,237 bps, Translation length: 365 residues

Version: Mycgr3T103221

Type: Protein coding

Annotation Method: Protein coding genes annotated by the [JGI](#)

Ensembl Fungi release 56 - Feb 2023 © [EMBL-EBI](#)

About Us

[About us](#)

[Contact us](#)

Get help

[Using this website](#)

[Documentation](#)

Our sister sites

[Ensembl](#)

[Ensembl Bacteria](#)

Follow us

[Blog](#)

[Twitter](#)

(e) How many exons does this transcript have? Which one is the longest?

The left hand navigation column provides several options for the transcript. Click on the [Exons](#) link.

You may want to change the display (for example, to show more flanking sequence, or to show full introns). In order to do so click on [Configure this page](#) and change the display options accordingly.

Now click on the [cDNA](#) link to see the spliced transcript sequence.

cDNA sequence

[Download sequence](#)[BLAST this sequence](#)

Codons	Alternating codons	Alternating codons		
Exons	An exon	Another exon		
Variants	3 prime UTR	Missense	Stop gained	Synonymous
Other	UTR			
Markup	loaded			

- Variants are filtered by consequence type

1	AACTGAATCGTCGAAAAATGCCACTTACAACATCGTCGTGGAGGTGATCACTGCG	60
ATGCCACTTACAACATCGTCGTGGAGGTGATCACTGCG	43
M--P--T--Y--N--I--V--V--F--G--G--D--H--C--	14
61	GTCCTGAAGTCACCGCCGAGGCCCTGAAGGTGCTCGACGTGATCGACAACAGCAATGCCG	120
44	GTCCTGAAGTCACCGCCGAGGCCCTGAAGGTGCTCGACGTGATCGACAACAGCAATGCCG	103
15	G--P--E--V--T--A--E--A--L--K--V--L--D--V--I--D--N--S--N--A--	34
121	ACGTTCACTTCAACATTCAACCACACCTGCTCGGTGGCGCCTCGATCGATGCCACGGCG	180
104	ACGTTCACTTCAACATTCAACCACACCTGCTCGGTGGCGCCTCGATCGATGCCACGGCG	163
35	D--V--H--F--N--I--Q--P--H--L--L--G--G--A--S--I--D--A--H--G--	54

UnTranslated Regions (UTRs) are highlighted in dark yellow, codons are indicated by alternating light yellow highlight, and exon sequence is shown in black or blue letters to show exon divides.

We can look at the protein sequence in more detail, finding domains and structural information. Click on [Protein summary](#) to view domains from SignalP, Pfam, PROSITE, Superfamily, InterPro, and more.

(f) What domains can be found in the protein product of this transcript? How many different domain prediction methods agree with each of these domains?

- Can you see the stop gained mutation we saw in (b) here?
- Will this variant cause the deletion of an entire protein domain?
- Which one(s)?



Clicking on **Domains & features** shows a table of this information.

Next, follow the **General identifiers** link at the left.

This page shows information from other databases such as ENA, UniProtKB, INSDC and others, that match to the Ensembl transcript and protein.

General identifiers

This transcript corresponds to the following database identifiers:

External database	Database identifier
European Nucleotide Archive	AF156181 [align] [view all locations] CM001197 [align] [view all locations]
INSDC protein ID	AAD40111.1 [align] [view all locations] EGP90559.1 [align] [view all locations] EGP90559.1 [align] [view all locations]
RefSeq DNA	XM_003855535.1 [align] [view all locations]
RefSeq peptide	XP_003855583.1 [align] [view all locations]
STRING	1047171.Mycgr3P103221 [align] [view all locations]
UniParc	UPI000012E54F [align] [view all locations]
UniProtKB/TrEMBL	F9X2A6_MYCGM [align] Putative uncharacterized protein [view all locations]

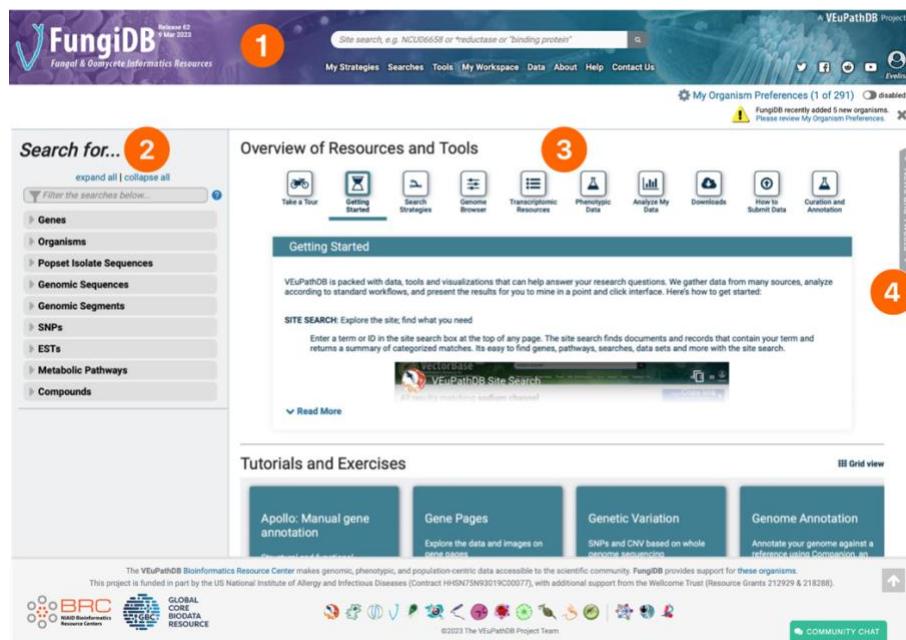
Anatomy of the FungiDB main page

FungiDB (<https://fungidb.org/>) is a free online informatics resource for exploration of fungal and oomycete omics scale data. It provides an easy-to-use, interactive interface to explore genomes, annotation, functional data (transcriptomics or proteomics), metabolic pathways and results from numerous genome-wide analyses (i.e. InterPro scan, signal peptide and transmembrane domain predictions, orthology, etc.). FungiDB also contains an expanding number of genomes from species spanning the Oomycetes and Fungi groups including but not limited to plant, animal, and human pathogens.

FungiDB is a component of the Eukaryotic Pathogen, Vector and Host Informatics Resource (VEuPathDB.org) that enables browsing, querying, and mining of genomic-scale datasets across diverse groups of organisms including hosts (HostDB.org), invertebrate vectors of human disease, pathogenic and non-pathogenic species, and environmental and epidemiological studies (ClinEpiDB.org).

- 1 The **Banner** contains a) release cycle information on the left, b) site search and main menu with access to nearly all searches, tools, and records in FungiDB (located centrally), and access to the personal account and social media (on the right).

My Organism Preferences is located under the banner and is a filtering that help restrict database searches to a list of organisms you are interested in. The feature can be easily disabled/enabled at any time from a button located under the header on every page.



- 2 The **Left panel** provides easy access to a list of categorised searches for genes and other types of data. It offers the same search options as the “Searches” within the main menu in the banner.
- 3 The central **Overview of Resources & Tools section** offers expandable cards to help you get started quickly, plus additional education resources containing in-depth tutorials and exercises.
- 4 The expandable News & Tweets section on the right contains information about new datasets, tools, features, and Twitter updates.

At the bottom of the main page, there are link outs to other VEuPathDB Sites (e.g., the HostDB database contains host-specific data from host-fungal pathogen studies) and the community chat app.

Exploring the gene record page

Learning objectives

- Become familiar with gene page structure and content
- Navigate to and from the gene pages

1. Use the site search to navigate to the gene record page of [Afu2g13260](#), which is a gene known to be important for the virulence of *Aspergillus fumigatus*.

The screenshot shows the FungiDB search results page. At the top, there is a search bar containing "Afu2g13260" with a magnifying glass icon and a yellow arrow pointing to it. Below the search bar is a navigation menu with links: My Strategies, Searches, Tools, My Workspace, Data, About, Help, and Contact. A "Run a new search" button is also present.

The main content area is titled "Genes matching Afu2g13260" and displays "1 - 1 of 1". On the left, there is a "Filter results" panel with a "Genome" section showing "Genes" (count 1) and a "Filter Gene fields" section with checkboxes for "External links" and "Gene ID". Below these are "Filter organisms" sections for "Fungi" and "Ascomycota".

The right side of the page shows the details for the single result: "Gene - Afu2g13260 Putative regulator of adherence, host cell interactions and virulence". It lists the "Gene name or symbol: medA" and the "Organism: Aspergillus fumigatus Af293". A note indicates "Fields matched: External links; Gene ID". There is also a "Run a new search" button with a blue arrow pointing to it.

- 1 The panel on the left provides a summary of all record types that match Afu2g13260.
- 2 Click on the gene link to navigate to the gene record page for Afu2g13260.
- 3 Clicking on this button will transform your search into a search strategy.
Note: If the button is shaded/inactive, limit your search to a single data type using the Filter results panel on the left.

Gene page components

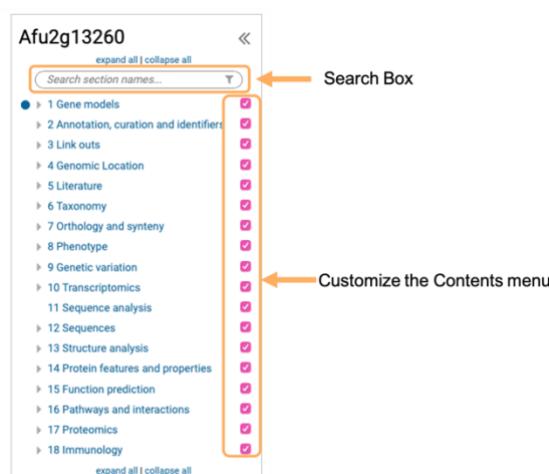
The top section of the gene record page provides a snapshot of the information available for this gene and also offers several shortcuts:

The screenshot shows the gene record page for Afu2g13260. At the top, there are three orange circles labeled 1, 2, and 3, each pointing to a button: 'Add to basket' (with a shopping cart icon), 'Add to favorites' (with a star icon), and 'Download Gene' (with a download icon). Below these buttons is the gene name 'Afu2g13260 Developmental regulator medA, putative'. To the right of the gene name is a 'Shortcuts' panel labeled 5, which contains links to Synteny, Alignments, SNPs, Transcriptomics, Protein Features, and Proteomics. Below the gene name, there is a section for 'Also see Afu2g13260 in the Genome Browser or Protein Browser'. On the left side of the page, there are sections for 'Name', 'Gene Type', 'Biotype Classification', 'Chromosome', 'Location', 'Species', 'Strain', 'Status', and 'Model Organism Database(s)'. At the bottom left, there is a link to 'View and update community annotations in Apollo'. A fourth orange circle labeled 4 points to the 'View and update community annotations in Apollo' link.

- 1 Add to basket: Save to basket if you want to download gene-specific information for selected genes.
- 2 Add to favorites: Saves genes in the private My favorites section, where you can add notes or keep track of your project.
- 3 Download Gene: Redirects to a download options page where gene records can be exported at text, FASTA, and GFF3 formats.
- 4 Submit a comment or annotate gene in Apollo, a web-based structural and functional gene annotation platform.
- 5 Shortcuts panel provides quick access to the selected section within the gene record page.

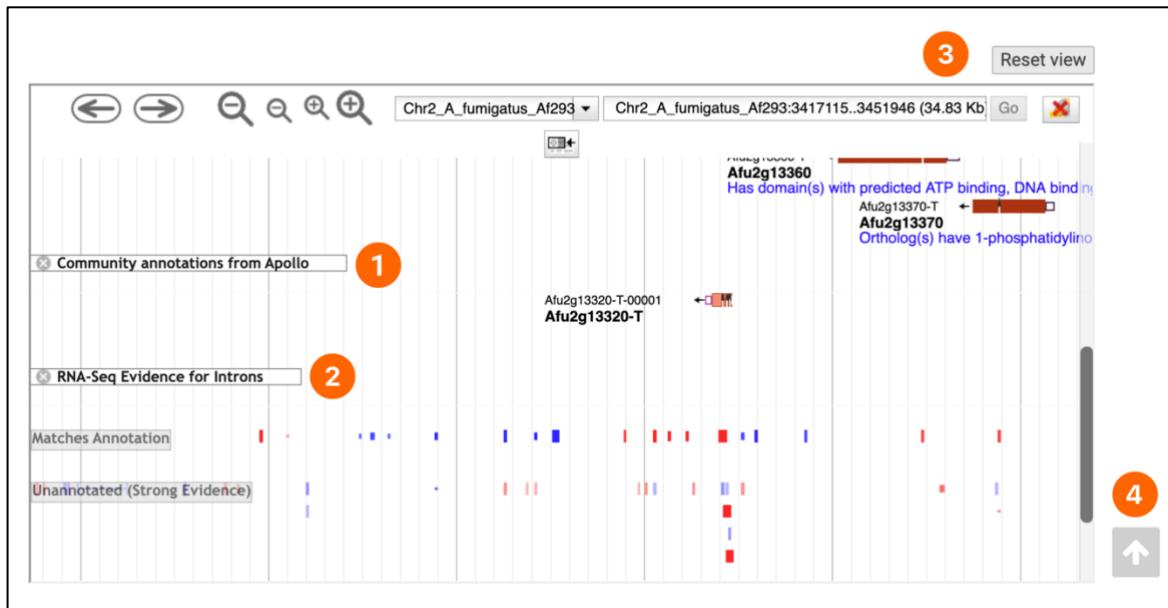
2. Explore the content of this gene record page. Below are the navigational highlights.

The **Contents** section contains a list of links to various sections in the gene record page and is searchable. Sub-sections of the Contents menu can be hidden by checking the box to the right.



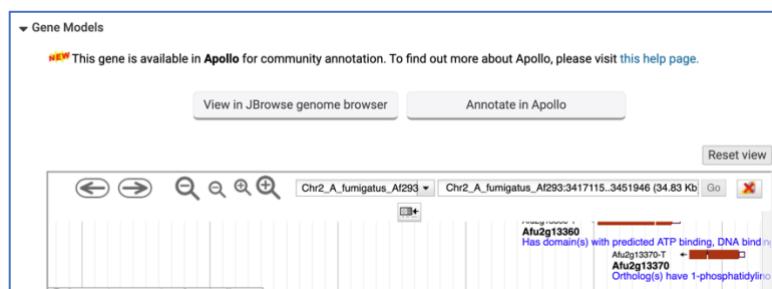
3. Explore the gene model section.

The **Gene Models** section is the first section of the gene record page, and it contains information about the structure of the gene (e.g. exon count, transcript number, annotated UTRs, community gene annotation in Apollo (1), introns (2), etc.) displayed within the genome browse JBrowse wrapper.



- 1 Community annotation from Apollo provides the latest annotation updates (within 24-48hrs after the changes were submitted to the curation team in Apollo)
- 2 Introns that are matching transcript annotation for which there is an abundance of supporting data from aligned RNA-Seq reads are displayed in bold colours (Blue for a forward gene and red for a gene located on the reverse strand).
- 3 If you navigated away from the gene of interest while scrolling, click on the “Reset view” button to return to the default position within the JBrowse wrapper.
- 4 Click this button to navigate to the top of the gene record page.

The “View in JBrowse genome browser” and “Annotate in Apollo” buttons open in separate tabs. In JBrowse, you can activate additional tracks and build custom evidence views. In Apollo, you can modify and create new genes to improve the genome annotation.



- Click on one of the introns within the **RNA-Seq Evidence for Introns** track. The pop-up window contains a table showing all experiments and samples that provide evidence for this intron junction (generated by in-house automated pipelines):

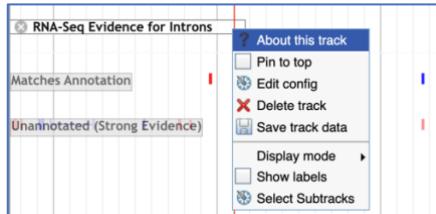
Intron Spanning Reads (ISR): The total number of uniquely mapped reads (all samples) which map across the junction and are on the appropriate strand. GSNAp uses splice site consensus sequences to determine the strand of the mapped read.

ISR per million (ISRPm): Intron Spanning Reads Per Million intron spanning reads and thus represents a normalized count of unique reads.

Chr2_A_fumigatus_Af293_3434523_3434611_0 details				
Intron Junction Details				
Intron Location	Chr2_A_fumigatus_Af293:3434523..3434611 (+ strand)			
Intron Spanning Reads (ISR)	7027			
ISR per million (ISRPm)	2140.85			
Gene assignment	Afu2g13295 - annotated intron			
% of Most Abundant Intron (MAI)	100			
Sample Details				
Experiment	Sample	Unique ISRPm	ISR/Cov	% MAI
Gene expression in WT, hrmA deletion, hrmA OE, hrmA_REV, EVOL under hypoxia and normoxia conditions	Delta hrmA hypoxia	64	33.07	5.61
	Delta hrmA normoxia	61	28.65	4.78
				61.61

% of Most Abundant Intron (MAI): The percentage (ISRPm of this junction / ISRPm of maximum junction for this gene) of this junction over the maximum for this gene.

Note that the tracks within the JBrowse wrapper have a drop-down menu for further track customization:



4. Explore other contents within the page.

The **Annotation, curation and identifiers** section offers alternate product descriptions, previous identifiers, and aliases, and is populated using data from internal curation, other fungal resources (*e.g.* AspGD, Ensembl, *etc.*), or user-submitted data (user comments).

2 Annotation, curation and identifiers																																			
Community annotations from Apollo																																			
Product Descriptions																																			
Alternate Product Descriptions																																			
Gene Name or Symbol																																			
Names, Previous Identifiers, and Aliases																																			
<table border="1"> <thead> <tr> <th colspan="3">Search this table...</th> </tr> <tr> <th>Name/ID/Alias</th> <th>Type</th> <th>Source</th> </tr> </thead> <tbody> <tr> <td>746128.CADAFUBP00002828</td> <td>alternate ID</td> <td>STRING</td> </tr> <tr> <td>EAL93620.1</td> <td>alternate ID</td> <td>protein_id</td> </tr> <tr> <td>Q4X0J5</td> <td>alternate ID</td> <td>Uniprot/SPTREMBL</td> </tr> <tr> <td>UPI000051EE09</td> <td>alternate ID</td> <td>UniParc</td> </tr> <tr> <td>XM_750565.1</td> <td>alternate ID</td> <td>RefSeq_dna</td> </tr> <tr> <td>XP_755658.1</td> <td>alternate ID</td> <td>RefSeq_peptide</td> </tr> <tr> <td>medA</td> <td>name</td> <td>N/A</td> </tr> <tr> <td>AFUA_2G13260</td> <td>previous ID</td> <td>N/A</td> </tr> <tr> <td>CADA/FUAG00004866</td> <td>previous ID</td> <td>N/A</td> </tr> </tbody> </table>			Search this table...			Name/ID/Alias	Type	Source	746128.CADAFUBP00002828	alternate ID	STRING	EAL93620.1	alternate ID	protein_id	Q4X0J5	alternate ID	Uniprot/SPTREMBL	UPI000051EE09	alternate ID	UniParc	XM_750565.1	alternate ID	RefSeq_dna	XP_755658.1	alternate ID	RefSeq_peptide	medA	name	N/A	AFUA_2G13260	previous ID	N/A	CADA/FUAG00004866	previous ID	N/A
Search this table...																																			
Name/ID/Alias	Type	Source																																	
746128.CADAFUBP00002828	alternate ID	STRING																																	
EAL93620.1	alternate ID	protein_id																																	
Q4X0J5	alternate ID	Uniprot/SPTREMBL																																	
UPI000051EE09	alternate ID	UniParc																																	
XM_750565.1	alternate ID	RefSeq_dna																																	
XP_755658.1	alternate ID	RefSeq_peptide																																	
medA	name	N/A																																	
AFUA_2G13260	previous ID	N/A																																	
CADA/FUAG00004866	previous ID	N/A																																	

The **Link outs** section offers redirection to other resources (*e.g.*, CGD, Ensembl, MycoCosm, *etc.*).

The **Orthology and synteny** section provides a table of Orthologs and Paralogs within FungiDB produced by OrthoMCL (www.orthomcl.org).

The table has a search box for creating a custom display of orthologs and also deploy the ClustalOmega analysis. The output of this tool can be used to build phylogenetic trees (*e.g.* iTOL).

Orthology and synteny

Ortholog Group: CDB_19803

Showing 8 of 125 rows

Candida

Use search box to limit to a species of interest

Protein cluster	Gene	Organism	Product	In synteny	has commands
CM0_23707	Coccidioides immitis HG28.4		transcription regulator	yes	no
CM0_26073	Coccidioides immitis RS		transcription regulator	yes	no
CP02_09848	Coccidioides pseudotropicalis		transcription factor	yes	no
CP07_08670	Coccidioides pseudotropicalis		hypothetical protein	yes	no
CP05_08547	Coccidioides pseudotropicalis		hypothetical protein	yes	no
CP05_08548	Coccidioides pseudotropicalis		transcription factor motif	yes	no
CP05_08549	Coccidioides pseudotropicalis		transcriptional regulator	yes	no
PA00_17198	Paracoccidioides lutzii PMP1		hypothetical protein	yes	no
PA00_26320	Paracoccidioides brasiliensis		hypothetical protein	yes	no
PA00_26321	Paracoccidioides brasiliensis		hypothetical protein	yes	no

Run clustal Omega for selected genes: Check All

Run ClustalOmega to generate protein sequence alignments and a .dnd file for building phylogenetic trees

Clustal Omega 1.2.3 Multiple Sequence Alignments

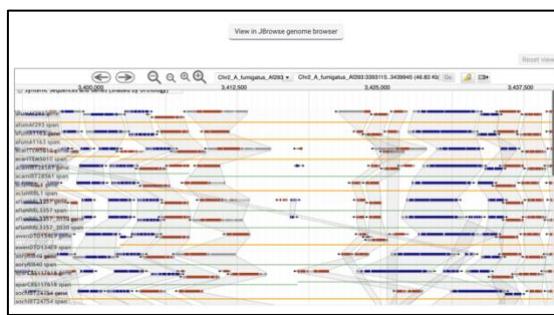
.dnd file

```

{
  "PAM_12199-t30_1-p1": 0.0101626
  ...
  "PAM_05304-t30_1-p1": 0.00135501
  ...
  "PAM_12215-t30_1-p1": 0.00135501
  ...
  "PAM_05304-t30_1-p1": 0.00135501
  ...
  "PAM_12199-t30_1-p1": 0.271152
}

```

The **Orthology and Synteny** section also contains synteny graphs in JBrowse:



In the screenshot above, the synteny genes are highlighted in grey.

The **Phenotype** section offers curated information, including annotations from the Pathogen-Host Interactions database, COFUND project (selected transcription factors knockouts) and other sources.

Aspergillus fumigatus transcription factor KO collection

No data available

PHI-base curated phenotypes

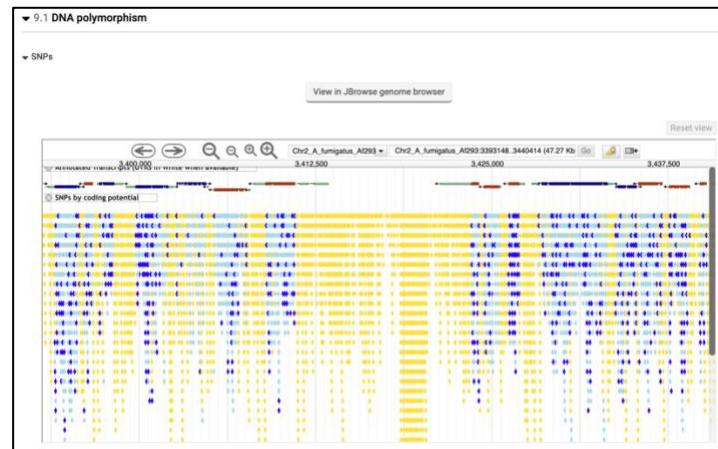
PHI-base entry	Essential gene	Multiple mutations	Pathogen species	Pathogen strain	Host species	Host strain	Tissue	Mutant phenotype	Disease
PHI:2661	no	no	Aspergillus fumigatus	Af293	Galleria mellonella (related greater wax moth)	N/A	larva	reduced virulence	Invasive pulmonary aspergillosis

Search this table...

Modification	Allele	Phenotype	Further Information	PubMed	ChemBiol Annotation Extension
null mutant	medA::hygR	abnormal colony color	Details:delay in brown pigmentation	19889083	N/A
null mutant	medA::hygR	decreased amount biological adhesion	Condition:fibronectin coated wells	19889083	N/A
null mutant	medA::hygR	decreased amount conidium formation	N/A	19889083	N/A
null mutant	medA::hygR	decreased amount virulence	Virulence model:immunosuppressed mouse pulmonary infection	19889083	N/A
null mutant	medA::hygR	decreased amount virulence	Virulence model:insect infection (Galleria mellonella)	19889083	N/A
null mutant	medA::hygR	increased amount cell growth	Details:slightly larger conidia and conidiophores	19889083	N/A

The **Genetic variation** section summarizes integrated SNP data for a given region and classifies SNPs based on the effect on gene function:

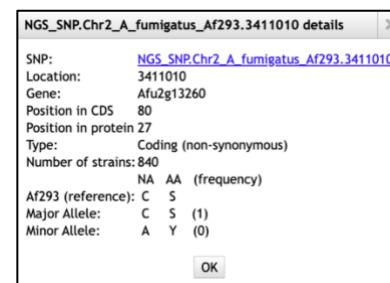
- noncoding (yellow diamonds)
- non-synonymous (dark blue)
- synonymous (light blue)
- nonsense (red)



Note that you can interact with the SNP records by using left and right clicking options on your mouse/touch pad.

Left click brings up a pop-up window containing more information about a particular SNP:

The SNP record linked in blue is linked to the SNP record page, which contains summary of the SNP across different isolates and samples.



Add to basket Add to favorites Download SNP

SNP: NGS_SNP.Chr2_A_fumigatus_Af293.3411010

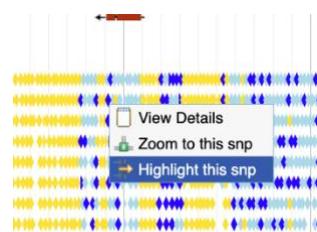
Organism: Aspergillus fumigatus Af293
Location: Chr2_A_fumigatus_Af293: 3,411,010
Type: coding
Number of Strains: 840
Gene ID: Afu2g13260
Gene Strand: reverse
Major Allele: G (1)
Minor Allele: T (0)
Distinct Allele Count: 2
Reference Allele: G
Reference Product: S 27
Allele (gene strand): C
SNP context: AGCCGATCCGTCTGCCCTGCATTTGCCAAAGGAGCAGCAGTGCTCAAGGAAGAAGAGGGCA
SNP context (gene strand): TGCCCTCTTCTCTTGAGCACTGCTGCTCTTTGCCAAATGCAGGCAGACGGATCGGCT

Major allele is the most common allele in the studied population/isolates.

Minor allele frequency is the frequency of the second most common allele. Minor allele

frequency is useful when looking for rare variants or disease-causing SNPs. However, there can be exceptions when minor allele frequency increases under selective pressure (e.g., development of drug resistance).

Right click provides more options for JBrowse view:



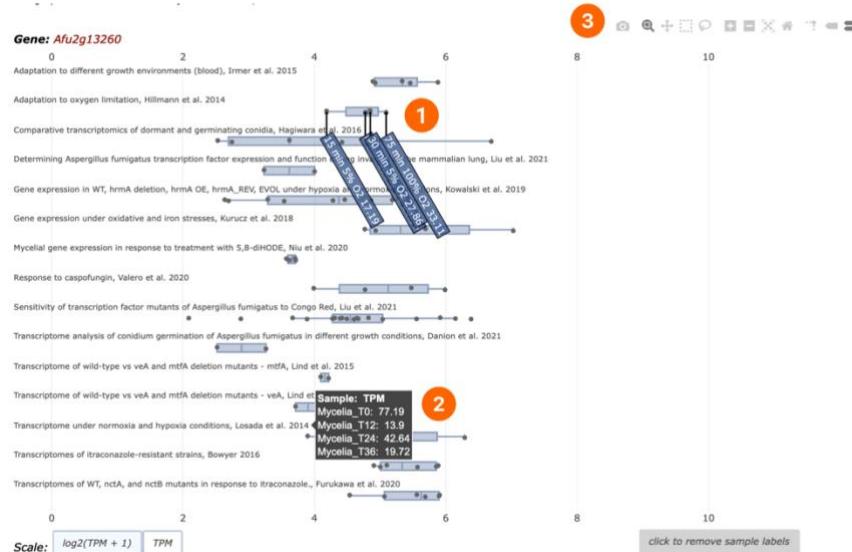
The **Transcriptomics** section (RNA-Seq and microarray data).

The Transcript Expression Summary section provides a big picture of gene expression across different samples and experiments, and helps identify experiments in which the current gene is highly regulated.

▼ RNA-Seq Transcription Summary 

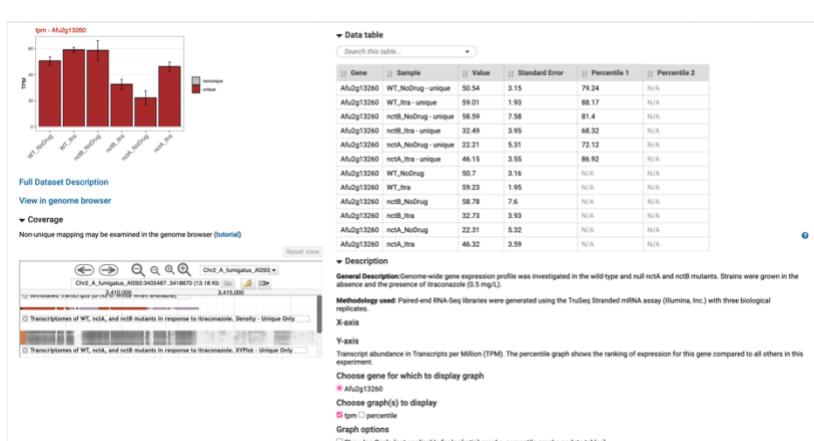
▼ Summary of expression values. Each row represents a distinct RNA-Seq experiment. [Click to read more...](#)

- Each black dot represents expression in a sample. A boxplot is shown, with the box representing the median and upper/lower quartiles, and the whiskers representing the max/min values (or 1.5 times the interquartile range, in which case values beyond the whiskers are considered outliers).
- Hover over the experiment name to show a table of sample names and values.
- Click on a boxplot to show labels and values for each sample in an experiment. Click again on the boxplot to hide labels and values. A button at the bottom-right removes labels and values for all experiments.
- Use the toggle button at the bottom-left to switch from Log Scale to Linear Scale.
- Log Scale values are $\log_2(\text{TPM} + 1)$ for these reasons:
 - TPM+1: to de-emphasize low noisy TPM values (i.e., <1).
 - log2: so that each unit on the x-axis represents a 2-fold difference.
- Navigation buttons appear at the above-right when hovering over the graph. If the buttons do not appear, reload the page.
- Zoom in with the navigation button or click and drag within the graph. Zoom out with the navigation button or double-click within the graph.
- This graph was created with Plotly. [Get more help at their website.](#)

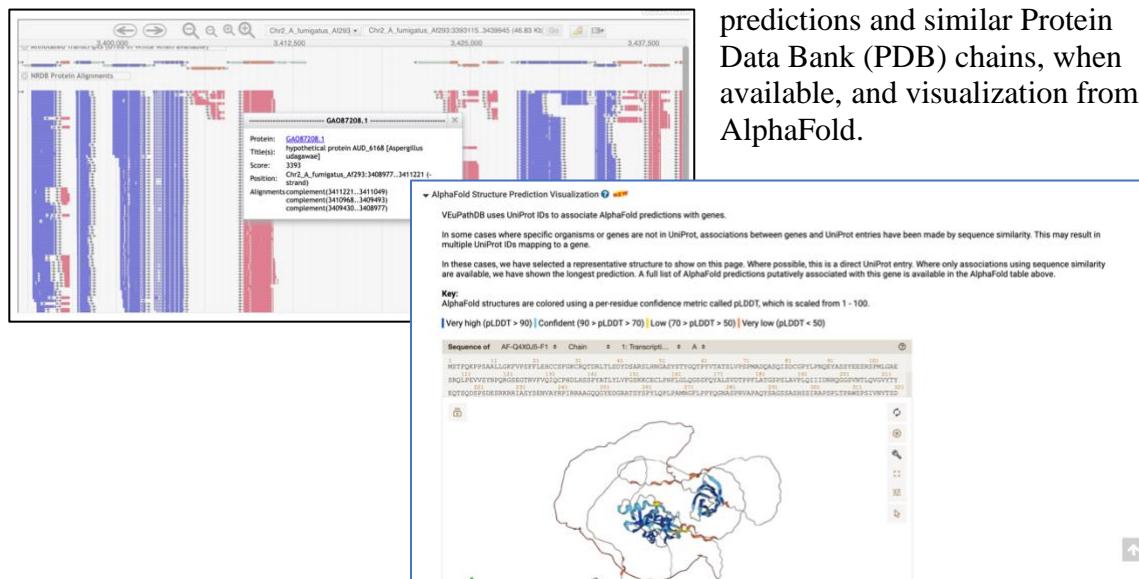


- 1 Clicking on the box plot will bring up sample labels.
- 2 Hovering over the experiments will display sample names.
- 3 The buttons above the summary graph provide additional options (e.g., download data in PNG, zoom, pan, etc.).

The **Transcript Expression** section, which is located under the RNA-Seq summary section, can be expanded to view the expression graph (TPM), data table, full dataset description, coverage plots, a link to the dataset in JBrowse.



The **Sequence analysis, Sequences, and Structure analysis** sections offer sequence information (DNA, RNA, and protein), an interactive summary of EST alignments and BLAT hits against the nonredundant protein sequence database (NRDB), 3D structure



The **Protein features and properties** section provides access to information about protein domains identifications (InterPro), signal and transmembrane predictions graphics, BLASTP hits, and other tools that can be deployed directly from the gene record page using the amino acid sequence of interest:

14 Protein features and properties

- ▶ Attributes and Protein Browser
- ▶ BLASTP (protein-protein BLAST)
- ▶ GPI anchor prediction: big-PI Predictor
- ▶ InterPro Domains
- ▶ InterProScan: Run on EBI site
- ▶ MitoProt
- ▶ STRING: functional protein association networks
- ▶ WoLF PSORT

The **Function Prediction** section features Gene Ontology (GO) assignments that have been either downloaded from Gene Ontology databases and manually curated by FungiDB. GO Slim terms, Enzyme Commission (EC) numbers with links to EC number and GO terms description and relevant publications are available as well.

Gene ontology provides statements for describing the functions of genes along with three aspects: Molecular function, Biological process, and Cellular component, and it is organized in hierarchies. The GO terms table above provides GO IDs and terms associated with a particular gene and additional metadata that are available for these associations such as source, evidence code (e.g. IDA, IMP, etc.) and reference (PubMed ID).

There are three classes of GO terms in FungiDB:

1. Automatically assigned by InterPro2GO
2. Assigned by FungiDB curators
3. Obtained from external resources such as AspGD, MIPS, and others

predictions and similar Protein Data Bank (PDB) chains, when available, and visualization from AlphaFold.

The **Pathways and interactions** section provides information about metabolic pathways loaded from the KEGG and MetaCyc repositories. Genes are linked to individual pathways by EC numbers when this data is available and clicking on any of the links redirects to an interactive metabolic pathway viewer where the user can explore individual reactions or export all known data for a given pathway. EC numbers are inferred by the OrthoMCL pipeline automatically. As with any automated analysis, use caution when interpreting this data.

The **Proteomics** section is populated when Mass Spec evidence data and phosphoproteomics datasets become available in FungiDB.

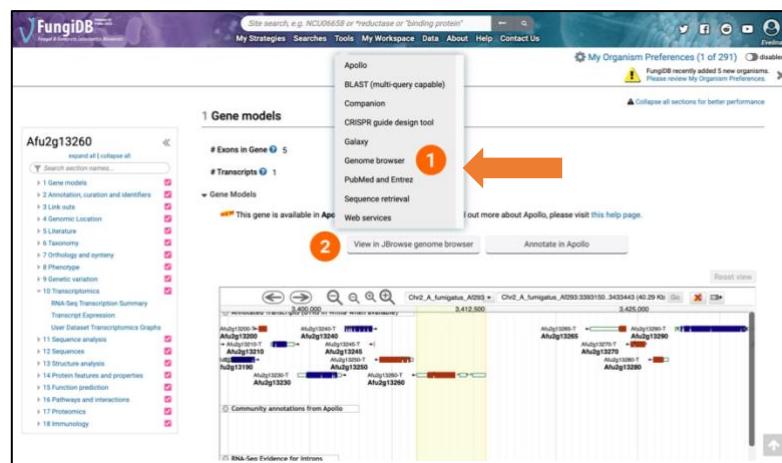
Exploring records in JBrowse

Learning objectives

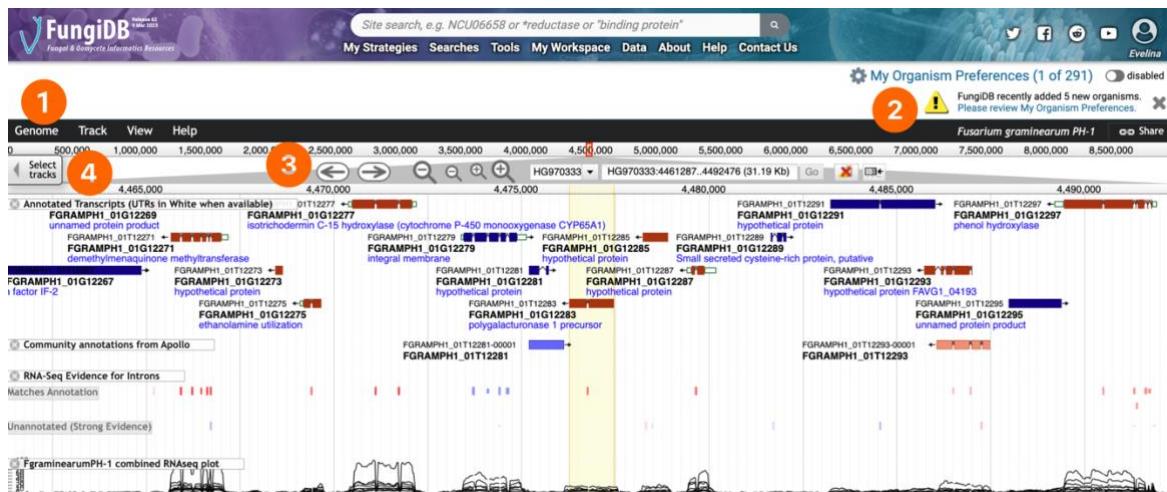
- Become familiar with JBrowse layout and navigation menus

JBrowse can be accessed from the main menu and also gene record pages.

Accessing JBrowse from gene record pages will pre-select the genome automatically. If navigating directly from the main page, the genome of choice can be chosen under the Genome tab.

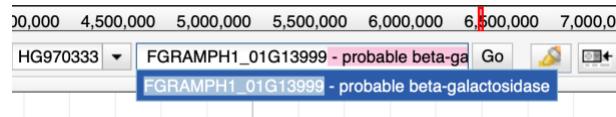


1. Navigate to the gene **FGRAMPH1_01T12283** in JBrowse.



- 1 The main menu provides options to select genomes, upload sequence or tracks, customize JBrowse view, and access several help articles.
- 2 Current genome selection id indicated on the right. The Share link generate a custom URL specific to the JBrowse session & tracks loaded.
- 3 The navigation bar contains zooming (magnifying glass icons), panning (left/right arrows) and highlighting (yellow highlighter) buttons, reference sequence selector (drop down with sequences from the selected genome sorted by length), a text box to search for features such as gene ids, and overview bar showing the location of the region in view.
- 4 Select additional tracks (transcriptomics, SNPs, sequence, ploidy, etc.) to customise your JBrowse view.

If you navigated away from the gene, you can type **FGRAMPH1_01T12283** directly in the location search box and select the highlighted match to return to your original position.



- Displaying more information about a gene

Left-clicking on a gene of interest will display a pop-up window that provides additional information about gene classification, position, orthology, links to JBrowse and gene record pages, as well CDS and protein sequences.

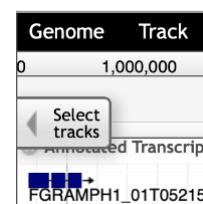
- Highlighting gene for easy tracking

Right-click (or control click) on the gene feature to display the context menu. The context menu offers gene highlighting and also options to display the gene record page and left-click details menu described above.



- Practice activating tracks in JBrowse by clicking on the Select tracks tab and choosing to display the following tracks:

- Intron Evidence (if not selected by default already)
- RNA-Seq (mycelium and 6days post-infection of wheat with Fusarium (6dpi))
- Chip-Seq tracks of chromatin marks in WT and kmt6 mutant (KMT6 is Histone H3 K27 Methyltransferase)
- Syntenic Sequences and Genes (Shaded by Orthology)
- To view the RNA-Seq data, select the Transcriptomics category and:
 - click on the “**Transcriptomic analysis during vegetative and infectious growth of Fusarium graminearum PH-1**” dataset
 - Select **unique coverage** tracks only



- Use the filter “Contains text” to search for and select two tracks: **infected 6dpi** and **mycelia**

The screenshot shows the JBrowse interface with two tracks selected:

- infected 6dpi**: Transcriptomic analysis during vegetative and infectious growth of *Fusarium graminearum* PH-1 - infected 6dpi (unique) Coverage.
- mycelia**: Transcriptomic analysis during vegetative and infectious growth of *Fusarium graminearum* PH-1 - mycelia (unique) Coverage.

- Clear your search by clicking on the “Clear All Filters” button.
- Select **ChIP-Seq** subcategory and select H3K27me3 and H3K4me3 methylation marks in kmt6 deletion mutant and WT in low nitrogen growth conditions (low). The exact track names are provided below. Use the “Contains text” filter window to search for the tracks.

For kmt6 deletion mutant:

- H3K27me3_DELkmt6_neoR_low-R2
- H3K4me2_DELkmt6_neoR_low-R2

For WT:

- H3K27me3_WT_low
- H3K4me2_WT_low

- Activate the **Syntenic Sequences and Genes (Shaded by Orthology)** track, which is located under the **Comparative Genomics** category, **Orthology and Synteny** subcategory but can be also searched for in the “Contains text” filter box at the top.

The 'Select Tracks' dialog box shows the following settings:

- Category: Comparative Genomics
- Subcategory: Orthology and Synteny
- Track Type: Segments
- RNASeq Alignment: (no data)
- RNASeq Strand: (no data)

Name	Category	Subcategory
Syntenic Sequences and Genes (Shaded by Orthology)	Comparative Genomics	Orthology and Synteny

- Click on the **Back to browser** button to return to JBrowse.

- Customize the JBrowse syntenic view to display *Fusarium* species only.

Hint: Click on the drop-down menu in the Syntenic Sequences and Genes (Shaded by Orthology) track, choose Select Subtracks, unselect current tracks, use the filter to identify “Fusarium” and then click on the “Save” button at the bottom.

The JBrowse interface shows the 'Syntenic Sequences and Genes (Shaded by Orthology)' track. The 'About this track' dropdown is open, and the 'Select Subtracks' button is highlighted. To the right, a table lists the tracks available for selection:

Label	Species	Kingdom	Genus	Phylum	Class	Order	Taxon
fgmH-1 gene	<i>Fusarium</i> graminearum	Fungi	<i>Fusarium</i>	Ascomycota	Sordariomycetes	gene	<i>Fusarium</i> graminearum PH-1
fgmH-1 span	<i>Fusarium</i> graminearum	Fungi	<i>Fusarium</i>	Ascomycota	Sordariomycetes	span	<i>Fusarium</i> graminearum PH-1
fgmH-1M58289 gene	<i>Fusarium</i> fujikuroi	Fungi	<i>Fusarium</i>	Ascomycota	Sordariomycetes	gene	<i>Fusarium</i> fujikuroi M58289
fgmH-1M58289 span	<i>Fusarium</i> fujikuroi	Fungi	<i>Fusarium</i>	Ascomycota	Sordariomycetes	span	<i>Fusarium</i> fujikuroi M58289
oxyFOSC3a gene	<i>Fusarium</i> oxysporum	Fungi	<i>Fusarium</i>	Ascomycota	Sordariomycetes	gene	<i>Fusarium</i> oxysporum F07
oxyFOSC3a span	<i>Fusarium</i> oxysporum	Fungi	<i>Fusarium</i>	Ascomycota	Sordariomycetes	span	<i>Fusarium</i> oxysporum F07
oxyFo47 gene	<i>Fusarium</i> fujikuroi	Fungi	<i>Fusarium</i>	Ascomycota	Sordariomycetes	gene	<i>Fusarium</i> fujikuroi
oxyFo47 span	<i>Fusarium</i> fujikuroi	Fungi	<i>Fusarium</i>	Ascomycota	Sordariomycetes	span	<i>Fusarium</i> fujikuroi

- Re-arrange tracks within JBrowse in the following order:
 - Pin the highlighted gene and gene model track to the top.

The JBrowse interface shows the re-arranged tracks. The 'fgmH-1' gene and 'FGM448-1_0112281 hypothetical protein' track are pinned to the top of the view. The 'About this track' dropdown is open, and the 'Pin to top' option is selected.

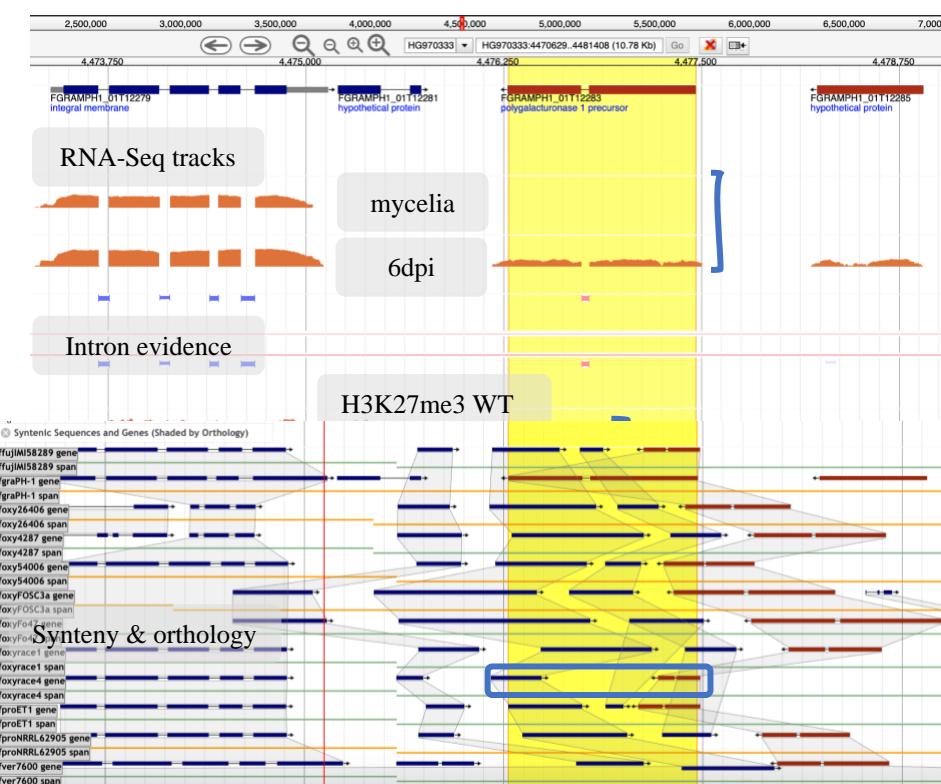
Note: you can use the drag and drop function in JBrowse to rearrange tracks in the order that helps you to evaluate the data.

In *Fusarium* and other fungi, H3K4me2 and H3K27me3 are found in large, mutually exclusive, gene-rich blocks of the genome. About one-third of the *F. graminearum* genome is associated with H3K27me3 when the fungus is grown in a minimal medium with low nitrogen.

Kmt6 encodes a DNA methylase that deposits repressive chromatin mark H3K27me3. *kmt6* mutants in *Fusarium* and other species demonstrate reorganization of chromatin marks (e.g., absence of the repressive chromatin mark H3K27me3) and upregulation of genes that are normally suppressed.

Having this information in hand and using JBrowse track view, answer the following questions:

- Do you observe *FGRAMPH1_01T12283* expression in both samples (mycelia and 6 days post-infection)?
- Does this gene show de-regulation of repressive chromatin marks (H3K27me3) in the *kmt6* mutant? Would you expect the expression of this gene to be up-regulated or down-regulated in the *kmt6* mutant?
- Is this gene expressed in mycelia?
- Is this gene conserved in *Fusarium* species?
- How would you generate a unique URL to this JBrowse view?



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/>)

The screenshot shows the MycoCosm homepage. At the top, there's a navigation bar with links to JGI HOME, GENOME PORTAL, MYCOCOSM, LOGIN, and a green button labeled "Click to login/register". Below the navigation is a search bar with two dropdowns: "Search MycoCosm" and "Search for JGI Data". The main content area features a large phylogenetic tree of fungi. Labels on the tree include: All MycoCosm Groups, Basidiomycota, Dikarya, Ascomycota, Mucoromycota, Zoopagomycota, Chytridiomycota, Agaricomycotina, Pezizomycotina, and various smaller clades like Tremellomycetes, Orbiliomycetes, Dothideomycetes, Lecanoromycetes, Leotiomycetes, Sordariomycetes, Xylophagomycetes, Saccharomycotina, Taphrinomycotina, Glomeromycotina, Mortierellomycotina, Mucoromycotina, Zoopagomycotina, Entomophthoromycotina, Kickxellomycotina, Blastocladiomycota, Chytridiomycetes, Monoblepharidomycete, Neocallimastigomycetes, Microsporidia, and Cryptomycota. To the right of the tree, there are two callout boxes. The top box, titled "Provides", lists three bullet points: "data access", "visualization", and "analysis tools". The bottom box, titled "Allows researchers to", lists two bullet points: "Explore gene models within species" and "Compare gene content between species". At the bottom of the page, there's a note about citation and a link to the JGI Fungal Program.

Provides

- data access
- visualization
- analysis tools

Allows researchers to

- Explore gene models within species
- Compare gene content between species

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, Otlilar 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, Spatha 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)

The screenshot shows the JGI Genome Portal homepage. At the top, there's a navigation bar with links to JGI HOME, GENOME PORTAL, LOGIN, and a link to "0 items". Below the navigation is a search bar with a "Search" button and a "Advanced Search" link. The main content area has several sections: "Searching for Projects" (with links to explore projects/proposals), "Downloading Files" (with links to download over the web, via Globus service, via API, and with Cart), "Looking for Access" (with links to contact PIs and get instructions), "GENOME INSIDER" and "Natural Prodcast" (with a link to subscribe to podcasts), and "The 'Tree of Life'" (with a link to the tree of life). A large red arrow points from the "The 'Tree of Life'" section down to a callout box at the bottom right. This callout box contains the text: "MycoCosm & PhycoCosm - access to data, visualization, and analysis tools for comparative genomics of fungi or algae." At the bottom of the page, there are links to Genomes OnLine Database (GOLD), Integrated Microbial Genomes (IMG) and Metagenomes (IMG/M), and Phytozome.

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.

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:

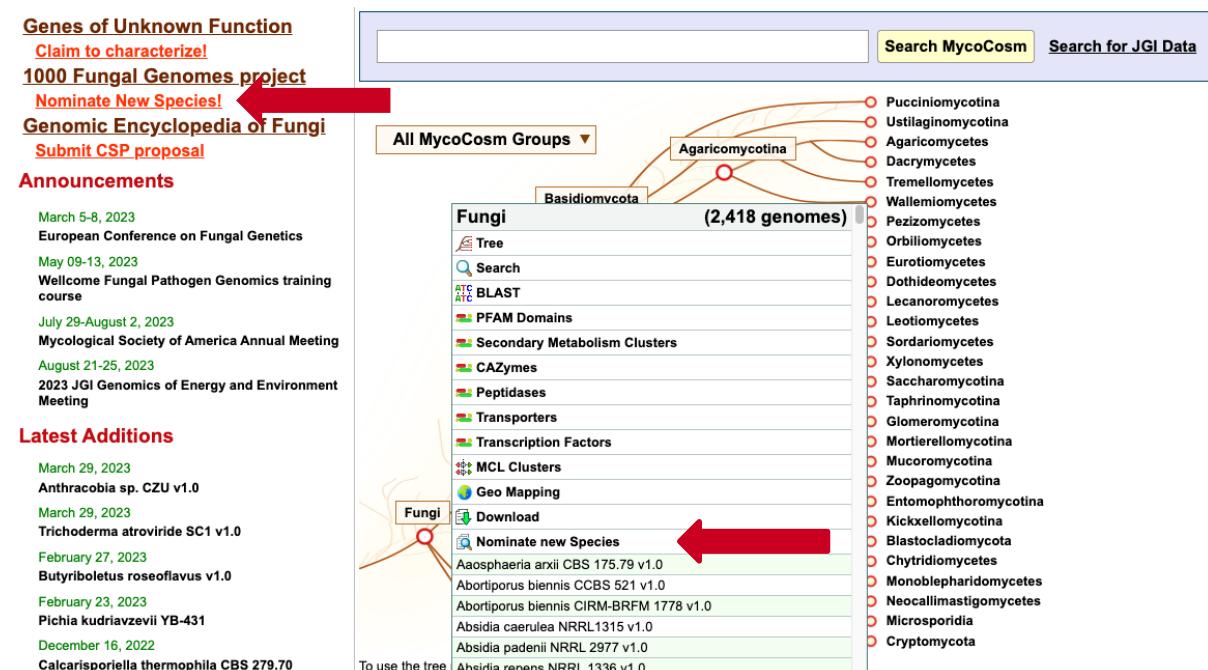
[Forgot your password?](#)

SIGN IN

SIGN IN WITH GOOGLE (LBL/JGI users)

REGISTER

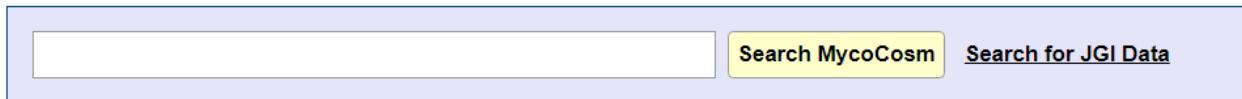
Currently MycoCosm hosts 2000+ genomes, both genomes sequenced by JGI in collaboration with the community, and genomes imported from the public domain or privately from collaborators. Most JGI-sequenced genomes are products of the Community Science Program (CSP) or the 1000 Fungal Genomes Project (1kFG). 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!



The screenshot shows the MycoCosm homepage with several key features highlighted:

- Navigation Links:** Includes "Genes of Unknown Function", "Claim to characterize!", "1000 Fungal Genomes project", "Nominate New Species!", "Genomic Encyclopedia of Fungi", and "Submit CSP proposal".
- Announcements:** Lists events like "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:** Lists new species added in March 2023, including *Anthracobia* sp. CZU v1.0, *Trichoderma atroviride* SC1 v1.0, *Butyriboletus roseoflavus* v1.0, *Pichia kudriavzevii* YB-431, *Calcarisporiella thermophila* CBS 279.70, *Aaosphearia arxii* CBS 175.79 v1.0, *Abortiporus biennis* CCB5 521 v1.0, *Abortiporus biennis* CIRM-BRFM 1778 v1.0, *Absidia caerulea* NRRL1315 v1.0, *Absidia padenii* NRRL 2977 v1.0, and *Absidia repens* NRRL 1336 v1.0.
- Search Interface:** Features a search bar, "Search MycoCosm", and "Search for JGI Data".
- Genome Tree and Statistics:** A large tree diagram shows the distribution of 2,418 genomes across various MycoCosm groups. The groups listed on the right include Pucciniomycotina, Ustilaginomycotina, Agaricomycetes, Dacrymycetes, Tremellomycetes, Wallomycetes, Pezizomycetes, Orbiliomycetes, Eurotiomycetes, Dothideomycetes, Lecanoromycetes, Leotiomycetes, Sordariomycetes, Xylinomycetes, Saccharomycotina, Taphrinomycotina, Glomeromycotina, Mortierellomycotina, Mucoromycotina, Zoopagomycotina, Entomophthoromycotina, Kickxellomycotina, Blastocladiomycota, Chytridiomycetes, Monoblepharidomycetes, Neocallimastigomycetes, Microsporidia, and Cryptomycota.

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.

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

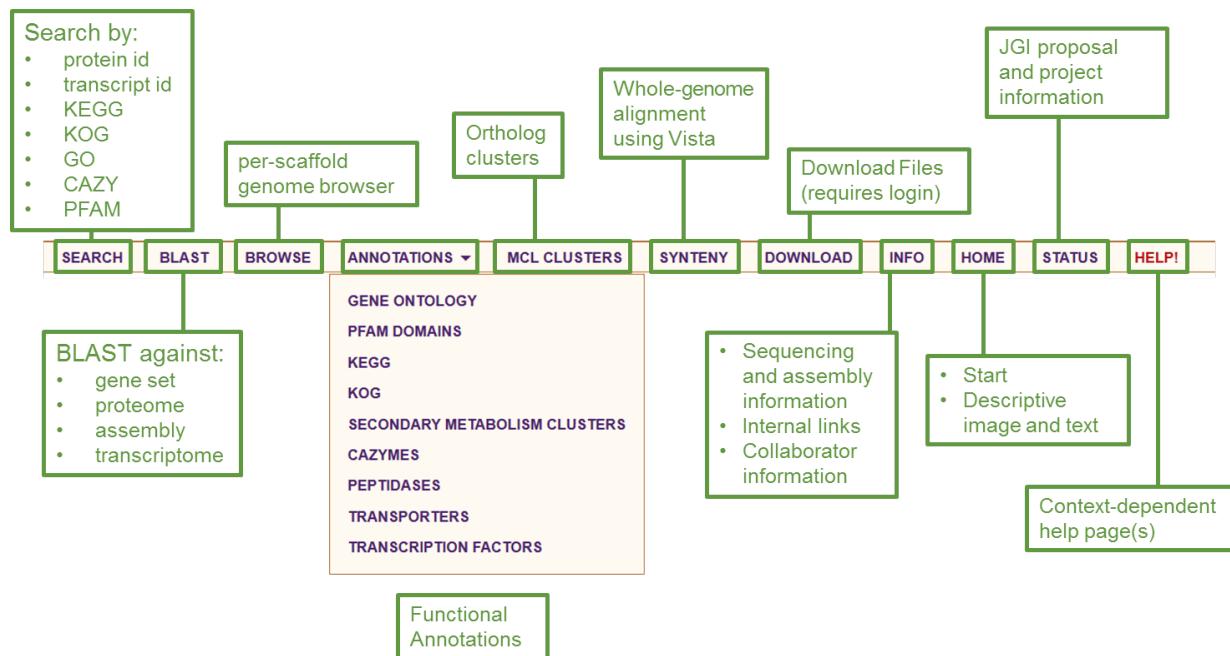
Choose one of the *Aspergillus* genome portals you identified in the exercise above. This will take you to the genome portal for that species. The data for that genome is stored in a database, which can be identified using the url for the genome portal. For example, the home page for *Aspergillus welwitschiae* 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.

A screenshot of a web browser displaying the MycoCosm genome portal for *Aspergillus welwitschiae*. The URL in the address bar is https://mycocosm.jgi.doe.gov/Aspwel1/Aspwel1.home.html. The page features a header with the JGI logo, MycoCosm logo, and navigation links for JGI HOME, GENOME PORTAL, MYCOCOSM, PHYCOCOSM, and LOGIN. Below the header is a banner for 'THE FUNGAL GENOMICS RESOURCE'. The main content area is titled 'Home • Aspergillus welwitschiae CBS139.54b v1.0'. On the left, there is a grid of six circular culture images labeled CREA, OAT, CYA, YES, MEA, and MEA OX. To the right of the cultures is a text box containing a paragraph about the genome project. Below this is a section titled 'Genome Reference(s)' with a link to a publication. At the bottom of the page, there is a photo credit: 'Photo credit: Ellen Kirstine Lyhne'.

Photo credit: Ellen Kirstine Lyhne

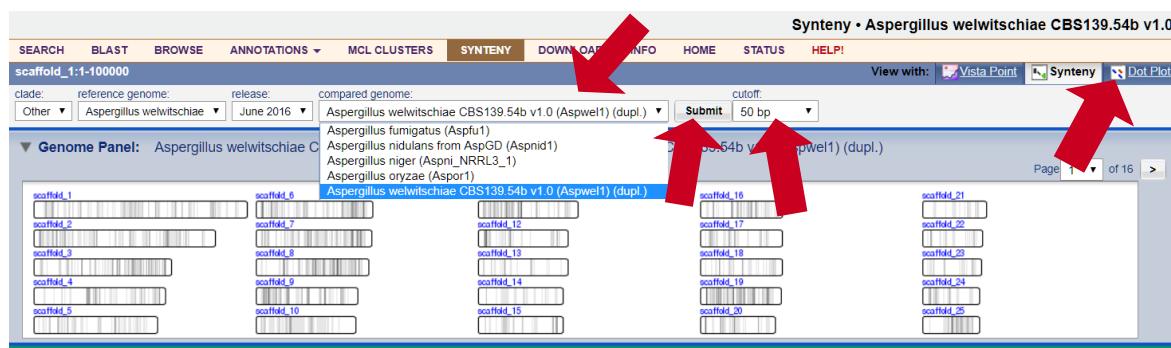
The navigation toolbar

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.

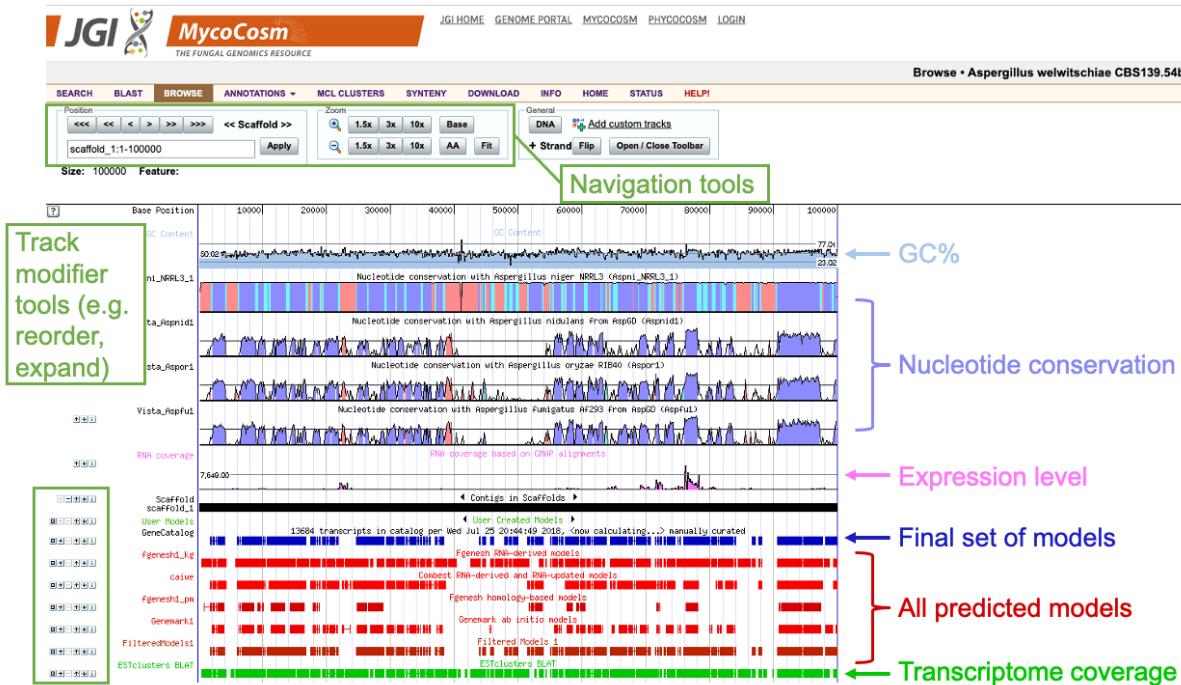


Check out the basic functions of the synteny browser by clicking on the buttons and pulldowns marked by the red arrows in the image above. To return from the ‘Dot Plot’ view, click ‘Synteny’ next to it. Note that self-alignment (main diagonal) is already removed in the dot plot if the genome is compared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

Genome browser

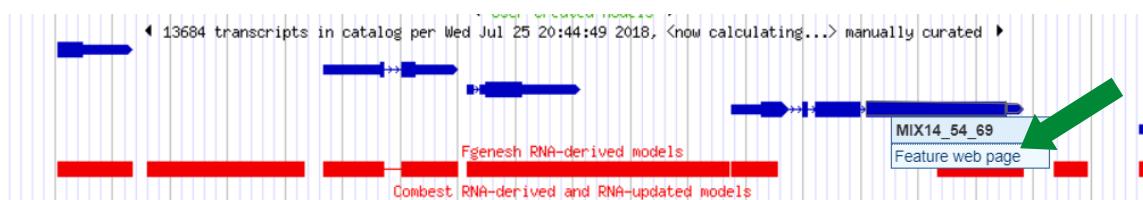
Now, we will explore the genome browser using the “BROWSE” tab. On the browser, you can navigate using available navigation tools and explore multiple layers of information we superimpose upon scaffolds. These include nucleotide conservation with closely related genomes, RNAseq expression data, Blastx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



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.

Link back to genome browser

Gene details and functional annotation

Structural information

Functional domains

Blast results

The screenshot displays the following sections:

- Gene details and functional annotation:** Shows gene ID (MIX14_54_69), location (scaffold 1:51371-53916), strand (+), number of exons (4), and a detailed description of the EST updated Ab initio model. It also lists ASPECT, GO Id (0006511), GO Desc (ubiquitin-dependent protein catabolism), KOG Group, KOG Id (KOG1865), KOG Class (Posttranslational modification, protein turnover, chaperones), Interpro Id (IPR001394), Interpro Desc (Peptidase C19, ubiquitin carboxyl-terminal hydrolase), and KOG Desc (Ubiquitin carboxyl-terminal hydrolase).
- Structural information:** Displays the protein sequence (MIX14_54_69) with its length (2546). A red box highlights the first 110 amino acids (aa 1-110).
- Functional domains:** Lists three domain families: PF00443: Ubiquitin carboxyl-terminal hydrolase [HMPFam], PS50235: Ubiquitin specific protease (USP) domain profile, [ProSiteProfiles], and SSF54001: [SUPERFAMILY].
- Blast results:** Shows a table of BLAST hits against the Aspergillus database. The columns include Flip Start, End, Len, ZC, Score, and Description. Most hits are Cysteine proteinases (no tax name).

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.
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.

The screenshot shows the MycoCosm Group Portal for the Pezizomycetes group. The main interface features a phylogenetic tree on the left with various fungal groups labeled: Basidiomycota, Dikarya, Mucoromycota, Zoopagomycota, Fungi, and Chytridiomycota. A dropdown menu titled "All MycoCosm Groups" is open over the tree, showing options like Tree, Search, ATC BLAST, PFAM Domains, Secondary Metabolism Clusters, CAZymes, Peptidases, Transporters, Transcription Factors, MCL Clusters, Geo Mapping, Download, and Nominate new Species. To the right of the tree is a detailed list of 74 Pezizomycetes genomes, each with a small circular icon and a link. At the bottom of the list is a note: "To use the tree navigation click a branch". Below the tree, there's a section for "Latest Additions" with a list of recent publications and genome entries. At the very bottom, there's a citation note: "For MycoCosm, please cite: Igor V. Kuo, Robin Ohm, Robert Otillar, Robert Riley, Asaf Salamov, Xueling Zhao, Frank Korzeniewski, Nucleic Acids Research, Mattirolomyces terfezioides MAT_tt4AII v1.1" followed by a DOI link.

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.

Info • Pezizomycetes								
TREE	SEARCH	BLAST	ANNOTATIONS ▾	MCL CLUSTERS	GEO MAPPING	DOWNLOAD	INFO	HELP!
##	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 GyrimSpk-SM18 v1.0			45,877,806	11,121			
10	Kalaharituber pfeillii 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.

Search • Pezizomyces

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

Search By: Across: Terms: ult exact - fast

Selected by Gzip

in 508

U.S. DEPARTMENT OF ENERGY Office of Science

California.

v Release Date:05-Apr-2023 14:36:24 PST Current Date:06-Apr-2023 17:27:40 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

Gene	Gene Ontology	Annotations	User Annotations
Portal: Phchr2 Portal Name: Phanerochaete chrysosporium RP-78 v2.2 Protein ID: 2912119 Transcript ID: 2912333 Location: scaffold_10:92485..92993 (+) 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: 3027200 Location: scaffold_6:1957160..1957677 (+) Model Name: gm1.597.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: gm1.732.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: gm1.857.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.genehm1_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 Kies , 2011-08-15) Description: class I hydrophobin (Ursula Kies , 2011-08-15) Define: class I hydrophobin (Ursula Kies , 2011-08-15)

Links to external databases

Manual annotations from users

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.

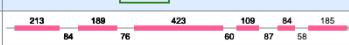
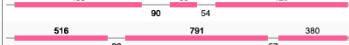
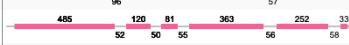
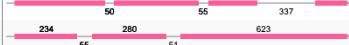
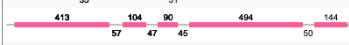
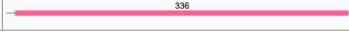
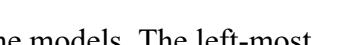
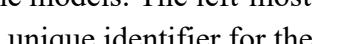
The screenshot shows a data grid with the following structure:

Annotations/Genomes	Antsp2	Ascm1	Ascm1	Cef1	Chov1	Disven1	Geocar1	Gresc1	Gyrinf1	Kalbru1	Kalife1	Luca1	Matter1	Mordum1	Moroh1	Moresc1	Moresc1	Moresc1	Morfluv1	Morlari1	Morris1	Morimp1	Mortak1	Morpai1	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	210	211	
↳ AA	6	5	5	6	9	6	7	7	8	5	5	6	9	8	7	7	6	6	6	5	6	5	4	5	7	5
↳ AC	1								2	1							1	2	1	1	1	1	1	1	1	1
↳ AD	4	4	3	3	4	6	2	2	3	2	3	3	2	2	2	2	3	4	2	1	1	3	1	3	2	1
↳ CA	24	19	18	25	23	22	28	24	21	21	26	22														
↳ CD	3	2	2	2	2	2	2	2	2	2	2	4	2													
↳ CE	1	2		1	3	2	2	1	1	1	1	1														
↳ CF	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																				
↳ GA																										
↳ GB																										
↳ L																										
↳ Kazal																										
↳ ID																										
↳ IL																										
↳ IV																										
↳ JC																										
↳ JE																										

Annotations and controls visible in the interface include:

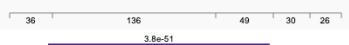
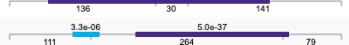
- Search bar: "Search for:" with dropdown options "Keywords" (selected), "Exact", and "Clear".
- Checkboxes: "S" and "S" (disabled).
- Help text: "Hover over to reveal organism name".
- Annotations: "Search by Keywords or Annotations".
- Interactions: "Click on number to go to web page with details about those proteins".
- Navigation: "Expand Merops classifications or click ID to go to clan/family page".
- Information: "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
Ascm1_321137	scaffold_3:818,001-819,568	1,568	401	•S08.115	•Subtilase family	
Ascm1_326276	scaffold_34:203,990-205,048	1,059	276	•S08.115		
Ascm1_211249	scaffold_57:190,517-189,456	1,062	306	•S08.115	•Subtilase family	
Ascn1_363694	scaffold_7:535,694-537,533	1,840	453	•S08.115	•Subtilase family •Peptidase inhibitor I9	
Ascn1_359457	scaffold_19:313,495-311,891	1,605	419	•cerevisin •S08.115	•Subtilase family •Peptidase inhibitor I9	
Morpop1_602957	scaffold_99:102,934-101,805	1,130	251	•S08.115	•Subtilase family	
Morpra1_711690	scaffold_5:2,218,063-2,220,014	1,952	480	•S08.115	•Subtilase family •Peptidase inhibitor I9	
Sarco1_402179	scaffold_8:110,499-109,257	1,243	379	•S08.115	•Subtilase family •Peptidase inhibitor I9	
Sarco1_418028	scaffold_283:11,576-10,133	1,444	415	•cerevisin •S08.115	•Subtilase family •Peptidase inhibitor I9	
Tubb1_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.

Rows: 10 25 ▾ rows per page						
Protein Id	Location	Gene Length	Protein Length	Merops Annotations	Domains	Models Domains
Ascm1_321137	scaffold_3:818,001-819,568	1,568	401	•S08.115	•Subtilase family	
Ascm1_326276	scaffold_34:203,990-205,048	1,059	276	•S08.115		
Ascm1_211249	scaffold_57:190,517-189,456	1,062	306	•S08.115	•Subtilase family	
Ascn1_363694	scaffold_7:535,694-537,533	1,840	453	•S08.115	•Subtilase family •Peptidase inhibitor I9	
Ascn1_359457	scaffold_19:313,495-311,891	1,605	419	•cerevisin •S08.115	•Subtilase family •Peptidase inhibitor I9	
Morpop1_602957	scaffold_99:102,934-101,805	1,130	251	•S08.115	•Subtilase family	
Morpra1_711690	scaffold_5:2,218,063-2,220,014	1,952	480	•S08.115	•Subtilase family •Peptidase inhibitor I9	
Sarco1_402179	scaffold_8:110,499-109,257	1,243	379	•S08.115	•Subtilase family •Peptidase inhibitor I9	
Sarco1_418028	scaffold_283:11,576-10,133	1,444	415	•cerevisin •S08.115	•Subtilase family •Peptidase inhibitor I9	
Tubb1_1_876549	scaffold_441:86,341-86,006	336	112	•S08.115		

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

The screenshot shows the CAZymes interface with a search bar at the top. Below it is a table of annotations grouped by genome. A green arrow points from the table to a callout box containing the text: "Click on counts to provide more details". Another green arrow points from the table to a detailed view of a protein's domains and models.

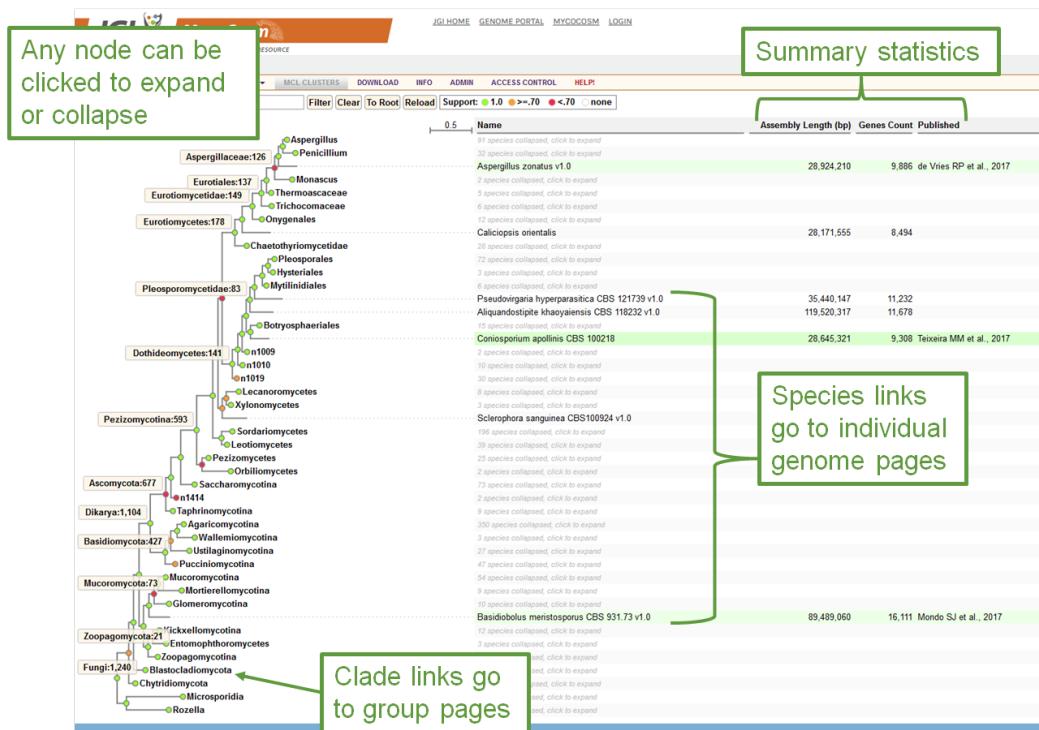
Annotations/Genomes	Antis1	Bjed1_1	Cors1	Dicat1	Fmp3	Gamp1	Giot1_1	Lntg6_1	Phca1	Pheb1	Pheb2	Pheb3	Pheb4	Pheb5	Pheb6_1	Trav1	Trie1_1	Wole1	Total	Annotation Description
CAZY	287	475	370	475	402	548	362	516	423	417	450	543	328	485	488	289		6,856	CAZy	
AA	36	104	62	88	43	93	41	94	83	85	89	93	40	89	86	27		1,153	Auxiliary Activities family	
CBM	20	62	37	45	45	57	19	48	43	62	71	101	45	49	63	18		776	Carbohydrate-Binding Module family	
CE	9	25	17	23	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	149	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	
GT	2	2	2	2	2	3	2	2	1	2	2	2	2	2	2	2		32	Glycosyltransferase Family 2	
Myosin_motor																		113	Myosin Motor family	
PL	3	10	6	11	3	9	9	11	8	4	6	7	5	9	10	2				

Protein Id	Location	Gene Length	Protein Length	CAZy Annotations	Domains	Models	Domains	Models	Domains				
Phchr2_2870875	scaffold_5_1:239,644-1,241,254	1,611	272	▲Auxiliary Activity Family 6	●Flavodoxin	644	28	154	100	436			
Phchr2_2319435	scaffold_19:448,641-450,192	1,552	373	▲Auxiliary Activity Family 2	●Peroxidase	81	154	57	219	52	49	56	62
Phchr2_3073200	scaffold_28:58,395-57,319	1,077	180	▲Auxiliary Activity Family 3 / Subf 2	●GMC oxidoreductase	17	10	46	74	67	120	79	424
Phchr2_3000492	scaffold_10:527,184-528,465	1,282	321	▲Auxiliary Activity Family 3 + Carbohydrate-Binding Module Family 1	●Fungal cellulose binding domain ●Glycosyl hydrolase family 61	141	184	19	233	392	64	55	58
Phchr2_61591	scaffold_10:1,037,135-1,039,405	2,271	611	▲Auxiliary Activity Family 3 / Subf 2	●GMC oxidoreductase ●GMC oxidoreductase	70	82	158	197	262	183	51	128
Phchr2_2810319	scaffold_9:1,444,920-1,446,516	1,597	373	▲Auxiliary Activity Family 2	●Peroxidase	61	154	57	215	42	79	242	182
Phchr2_2984061	scaffold_9:5,986,872-584,168	2,705	603	▲Auxiliary Activity Family 3 / Subf 3	●GMC oxidoreductase ●GMC oxidoreductase	66	52	58	49	59	410	56	49

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:

The screenshot shows the MycoCosm home page with a navigation bar. A red arrow points to the "TREE" button in the menu. The page title is "Info • Pezizomycetes".

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, 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.

The search results page for 'Search • Eurotiomycetes' shows a table with the following data:

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

Total genes found: 1

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), 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:	Chrl 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 pe 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 0005525 0004182 0008270	protein binding GTP binding carboxypeptidase A activity zinc ion binding	IPR016491 IPR000038 IPR000834 IPR000834	Septin Cell division/GTP binding protein Peptidase M14, carboxypeptidase A Peptidase M14, carboxypeptidase A
Biological Process	0006508 0007049	proteolysis and peptidolysis cell cycle	IPR000834 IPR000038	Peptidase M14, carboxypeptidase A 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.

The screenshot shows the MycoCosm homepage. At the top, there's a navigation bar with links to JGI HOME, GENOME PORTAL, MYCOCOSM, PHYCOCOSM, and LOGIN. Below the navigation bar, there's a main menu with sections like Genes of Unknown, 1000 Fungal Genomes, Genomic Encyclopedia, Announcements, and Latest Additions. A green arrow points to the 'Video Tutorials' link in the main menu. To the right of the menu is a large phylogenetic tree of Fungi. The tree starts with 'Fungi' at the root, which branches into 'Chytridiomycota' and 'Ascomycota'. 'Ascomycota' further branches into 'Basidiomycota' and 'Ascomycotina'. 'Ascomycotina' branches into 'Pezizomycotina' and 'Agaricomycotina'. 'Pezizomycotina' includes 'Oomycota' and 'Sordariomycetes'. 'Agaricomycotina' includes 'Tremellomycetes', 'Walemiomycetes', 'Pezizomycetes', 'Orbilliomycetes', 'Eurotiomycetes', 'Dothideomycetes', 'Lecanoromycetes', 'Leotiomyces', 'Sordariomycetes', 'Xylonomycetes', 'Saccharomycotina', 'Taphrinomycotina', 'Glomeromycotina', 'Mortierellomycotina', 'Mucoromycotina', 'Zoopagomycotina', 'Entomophthoromycotina', 'Kickxellomycotina', 'Blastocladiomycota', 'Chytridiomycetes', 'Monoblepharidomycetes', 'Neocallimastigomycetes', 'Microsporidia', and 'Cryptomycota'. There are also links for 'Search MycoCosm' and 'Search for JGI'.