

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 website interface. At the top, there is a search bar containing "Afu2g13260" with a magnifying glass icon and a "Run a new search" button. Below the search bar is a navigation menu with links: My Strategies, Searches, Tools, My Workspace, Data, About, Help, and Contact. The main content area is titled "Genes matching Afu2g13260". On the left, there is a "Filter results" panel with sections for "Genome" (Genes) and "Filter Gene fields" (External links, Gene ID). Below that is a "Filter organisms" section with a dropdown menu showing "Fungi" and "Ascomycota". The main results area displays two entries. The first entry is highlighted with a red circle labeled "1" and contains details: "Gene - Afu2g13260 Putative regulator of adherence, host cell interactions and virulence", "Gene name or symbol: medA", "Organism: Aspergillus fumigatus Af293", and "Fields matched: External links; Gene ID". The second entry is highlighted with a red circle labeled "2" and contains similar details. A blue button labeled "3" is located at the top right of the results area, with an arrow pointing to it from the text below.

- 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

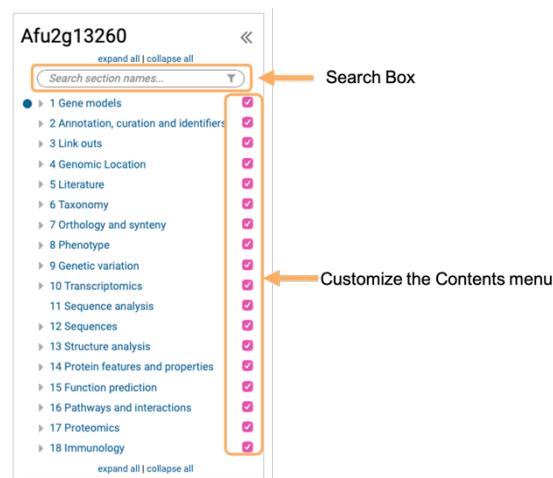
The top section of the gene record page provides a snapshot of the information available for this gene and also offers several shortcuts:

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 different button: 'Add to basket' (1), 'Add to favorites' (2), and 'Download Gene' (3). Below these buttons is the gene identifier 'Afu2g13260 Developmental regulator medA, putative'. To the right of the identifier is a 'Shortcuts' panel (5) containing links to Synteny, Alignments, SNPs, Transcriptomics, Protein Features, and Proteomics. Below the gene identifier, there is a summary of gene details: Name: medA, Gene Type: protein coding gene, Biotype Classification: protein_coding, Chromosome: 2, Location: Chr2_A_fumigatus_Af293:3,408,135..3,413,423(-), Species: Aspergillus fumigatus, Strain: Af293, and Status: Curated Reference Strain. There are also links to view comments or update annotations in Apollo. At the bottom left, it says 'Model Organism Database(s)' and 'CGD: C3_01180C_A'. Callout 4 points to the 'View 1 user comment, or add a comment' 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.

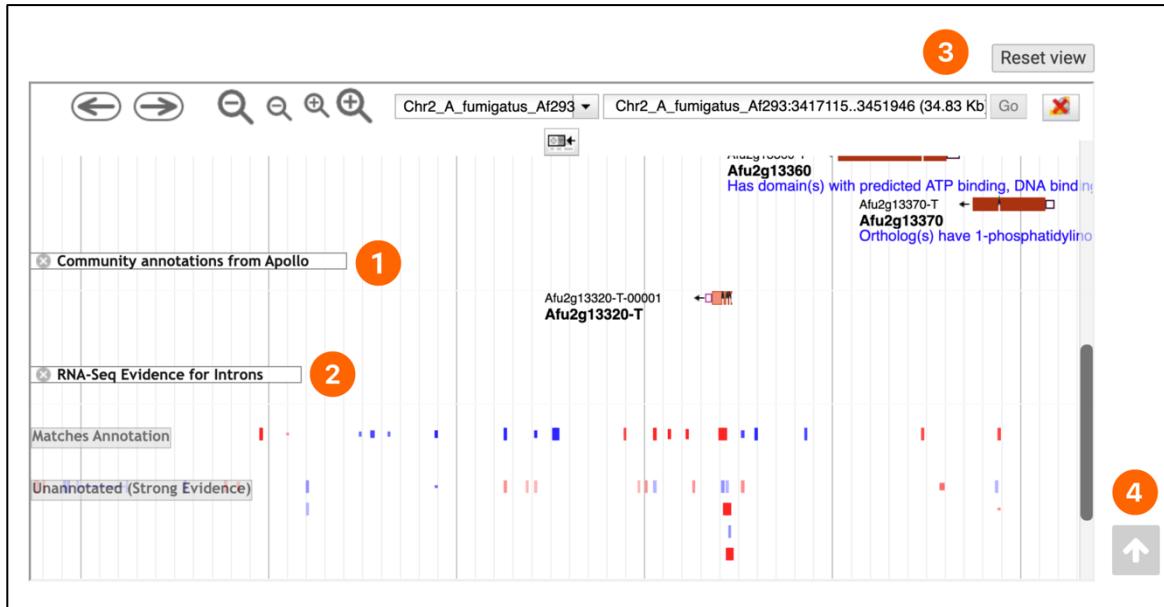
2. Explore the content of this gene record page. Below are the navigational highlights.

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.



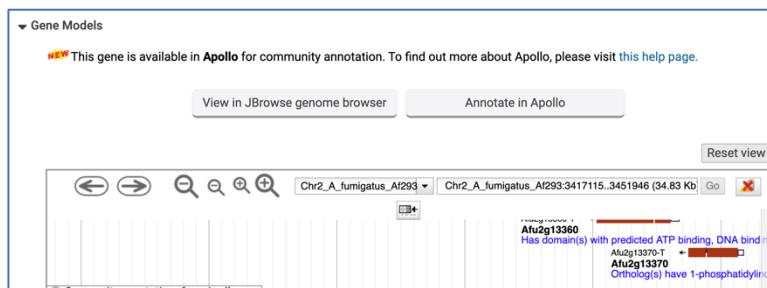
3. Explore the gene model section.

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):

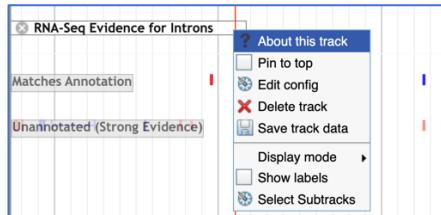
Intron Spanning Reads (ISR): The total number of uniquely mapped reads (all samples) which map across the junction and are on the appropriate strand. GSNAP uses splice site consensus sequences to determine the strand of the mapped read.

ISR per million (ISRPBM): 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 (ISRPBM)					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	100
	Delta hrmA normoxia	61	28.65	4.78	61.61

% of Most Abundant Intron (MAI): The percentage (ISRPBM of this junction / ISRPBM 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.CADAFAUBP00002828</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>UPI0000051EE09</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>CADAFAUAG00004866</td> <td>previous ID</td> <td>N/A</td> </tr> </tbody> </table>			Search this table...			Name/ID/Alias	Type	Source	746128.CADAFAUBP00002828	alternate ID	STRING	EAL93620.1	alternate ID	protein_id	Q4X0J5	alternate ID	Uniprot/SPTREMBL	UPI0000051EE09	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	CADAFAUAG00004866	previous ID	N/A
Search this table...																																			
Name/ID/Alias	Type	Source																																	
746128.CADAFAUBP00002828	alternate ID	STRING																																	
EAL93620.1	alternate ID	protein_id																																	
Q4X0J5	alternate ID	Uniprot/SPTREMBL																																	
UPI0000051EE09	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																																	
CADAFAUAG00004866	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 OOG_19894

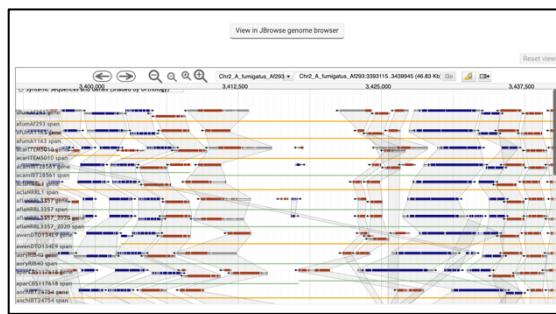
Orthologs and Paralogs within EuPathDB

Clustel Use search box to limit to a species of interest

Protein cluster	ID	Gene	Organism	Product	Is paralogic?	Has comments?
CHE_02079		Coccioides immitis	HGS84	transcriptional regulator Medusa	yes	no
CIM_00070		Coccioides immitis	RS	transcriptional regulator	yes	no
CIM_00088		Coccioides immitis	RS	medium	yes	no
CIM_00090		Coccioides pseudotropicalis	PMSCC	medium	yes	no
CPTCBL_0088270		Coccioides pseudotropicalis	CT35	hypothetical protein	yes	no
CPSG_05687		Coccioides pseudotropicalis	delta S6Wng	hypothetical protein	yes	no
CPSG_05688		Coccioides pseudotropicalis	delta S6Wng	transcription factor mediX	yes	no
PAAS_11119		Coccidioides immitis	Pt01	transcriptional regulator	yes	no
PABG_00024		Paracoccidioides brasiliensis	PMD	hypothetical protein	yes	no
PAGL_10218		Paracoccidioides brasiliensis	PMD	hypothetical protein	yes	no

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

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



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

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

Aspergillus fumigatus transcription factor K0 collection										Data sets	
No data available											
Pht-base curated phenotypes										Download	Data sets
Pht-base entry	Essential gene	Multiple mutations	Pathogen species	Pathogen strain	Host species	Host strain	Tissue	Mutant phenotype	Disease		
Pht2651	no	no	Aspergillus fumigatus	Af293		N/A	larva	reduced virulence	Acute pulmonary aspergillosis	No	Yes

Phenotype (qualities or directionality + entity or biological process)												Download	Data sets
Search this table...													
Modification	Allele	Phenotype	Further Information	PubMed	ChEBI Annotation Extension								
null mutant	medA-hypR	abnormal colony color	Details: delay in brown pigmentation	19889083	N/A								
null mutant	medA-hypR	decreased amount biological adhesion	Condition: fibronectin coated wells	19889083	N/A								
null mutant	medA-hypR	decreased amount conidium formation	N/A	19889083	N/A								
null mutant	medA-hypR	decreased amount virulence	Virulence model: immunosuppressed mouse peritonitis infection	19889083	N/A								
null mutant	medA-hypR	decreased amount virulence	Virulence model: insect infection (Galleria mellonella larvae)	19889083	N/A								
null mutant	medA-hypR	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.

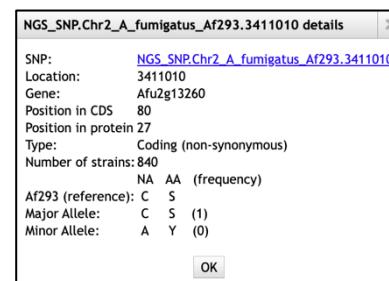
Left click brings up a pop-up window containing more information about a particular SNP:

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: AGCCGATCCGTCTGCCGCATTGCCAAGGAGCAGCAGTGCTCAAGGAAGAAAGGGCA
SNP context (gene strand): TGCCCTTTCTTCTTGAGCACTGCTGCTCTTGGCAAATGCAGGCACGGATCGGCT

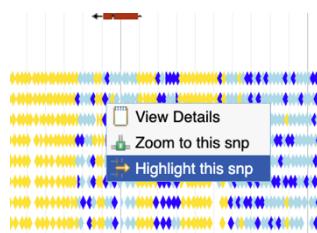


Major allele is the most common allele in the studied population/isolates.

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

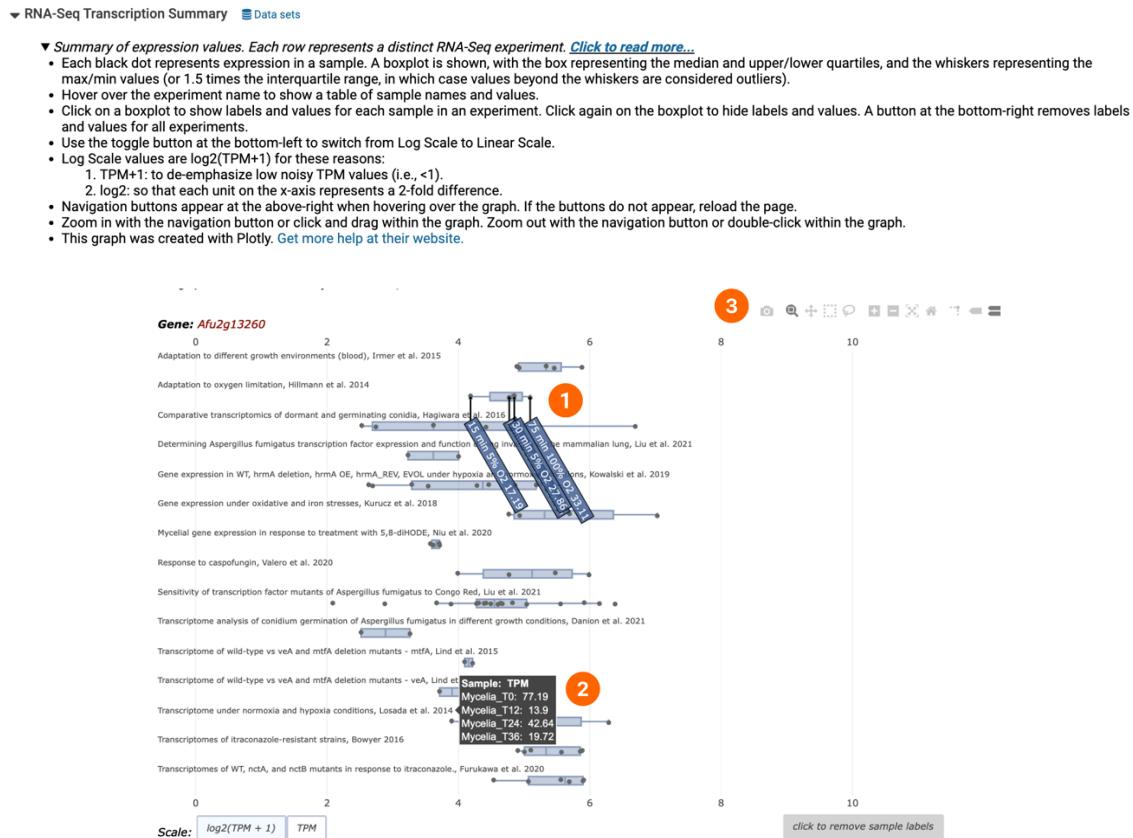
frequency is useful when looking for rare variants or disease-causing SNPs. However, there can be exceptions when minor allele frequency increases under selective pressure (e.g., development of drug resistance).

Right click provides more options for JBrowse view:



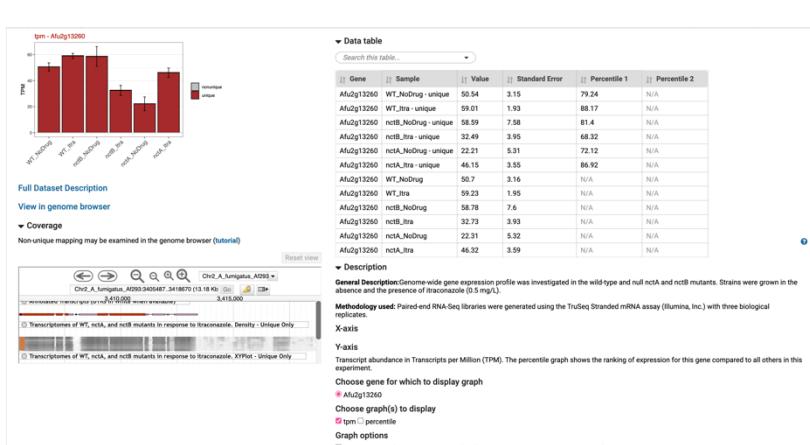
The **Transcriptomics** section (RNA-Seq and microarray data).

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.



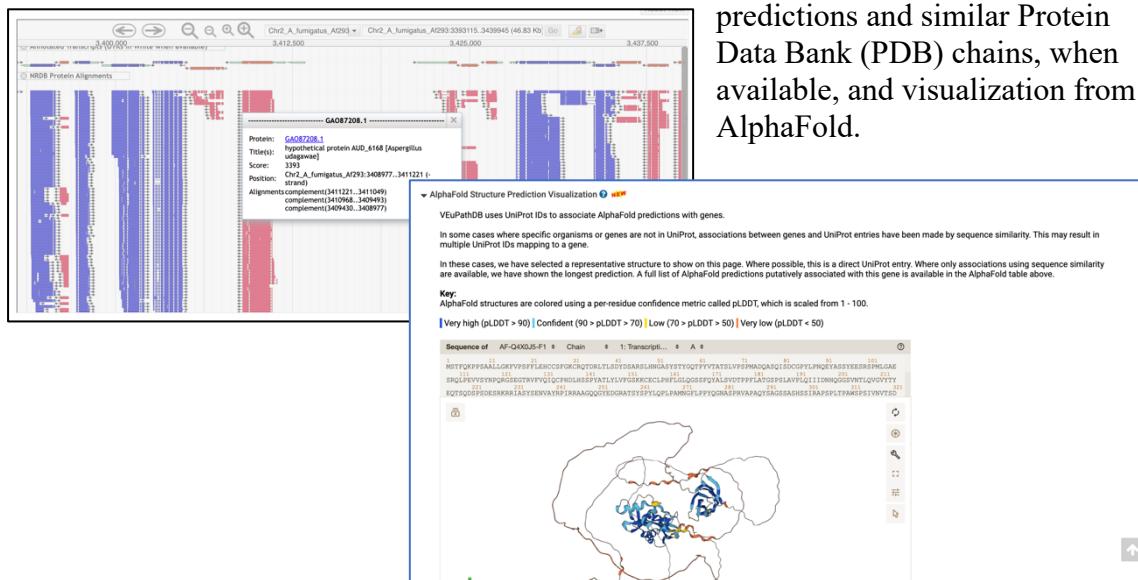
- 1 Clicking on the box plot will bring up sample labels.
- 2 Hovering over the experiments will display sample names.
- 3 The buttons above the summary graph provide additional options (e.g., download data in PNG, zoom, pan, etc.).

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

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



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

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.

Accessing JBrowse

JBrowse can be accessed from the main menu and 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.

- Navigate to JBrowse from the gene record page of [FGRAMPH1_01T12283](#) to learn the JBrowse environment.

Menu bar includes the Genome menu, Track menu, View menu, Help menu and the Sharing link.

Current genome is indicated above the tracks

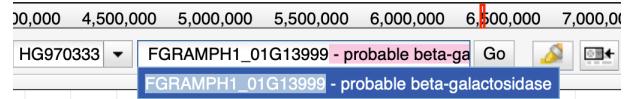
Customised JBrowse views can be shared via a URL generated by clicking on the *Share* button

Genome view area

Select additional tracks to view (transcriptomics, SNPs, etc.)

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 which shows the location of the region in view.

*Note: 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. You can also double click on the gene to re-centre your view.*

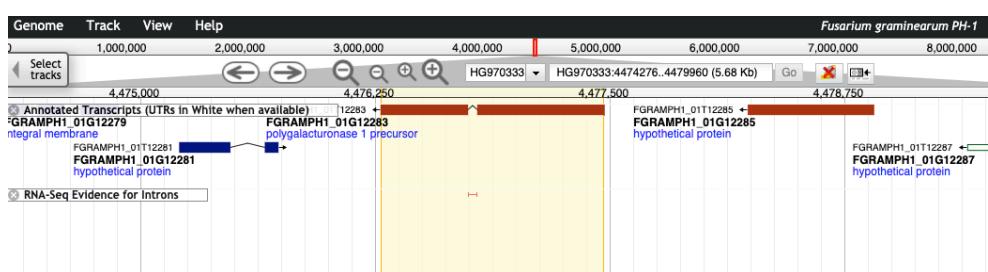
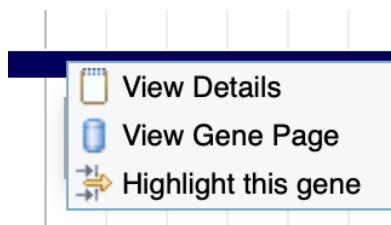


- Display more information about this gene and the coordinates for CDS.

Left-click on the gene and a pop-up window will display the coordinates of CDS, 3' and 5' UTRs (when available), etc. You can also export CDS and protein sequences or navigate to the gene record page.

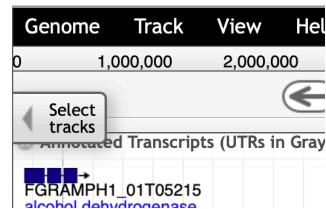
- Highlight this gene for easy tracking.

Right-click (or control click) on the gene feature to display the context menu which provides quick links to highlighting, navigation to the gene record page). You can also get the more details by selecting the View Details option, which will bring up the pop-up box shown above.



- Activate the following tracks to learn more about this gene (*hint: click on the Select tracks tab on the left*):
 - 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 activate tracks, click on the **Select tracks** tab on the left of the screen.



- Next, select the Transcriptomics category and click on the “Transcriptomic analysis during vegetative and infectious growth of *Fusarium graminearum* PH-1” dataset and search for two tracks - **infected 6dpi** and **mycelia**, and then select **unique coverage** tracks only for both conditions/samples.

<input checked="" type="checkbox"/> Transcriptomic analysis during vegetative and infectious growth of <i>Fusarium graminearum</i> PH-1 - infected 6dpi (unique) Coverage	Transcriptomics	RNA-Seq	Transcriptomic analysis during vegetative and infectious growth of <i>Fusarium graminearum</i> PH-1	Coverage	unique	not strand specific
---	-----------------	---------	---	----------	--------	---------------------

<input checked="" type="checkbox"/> Transcriptomic analysis during vegetative and infectious growth of <i>Fusarium graminearum</i> PH-1 - mycelia (unique) Coverage	Transcriptomics	RNA-Seq	Transcriptomic analysis during vegetative and infectious growth of <i>Fusarium graminearum</i> PH-1	Coverage	unique	not strand specific
---	-----------------	---------	---	----------	--------	---------------------

- Clear your search by clicking on the “Clear All Filters” button and activate ChIP-Seq tracks to select H3K27me3 and H3K4me3 methylation marks in WT and kmt6 deletion mutant in low nitrogen growth conditions (low). Search for the following text in the “Contains text” filter to select the “unique coverage” tracks.
 - kmt6 mutant:
 - H3K27me3_DELkmt6_neoR_low-R2
 - H3K4me2_DELkmt6_neoR_low-R2

WT:

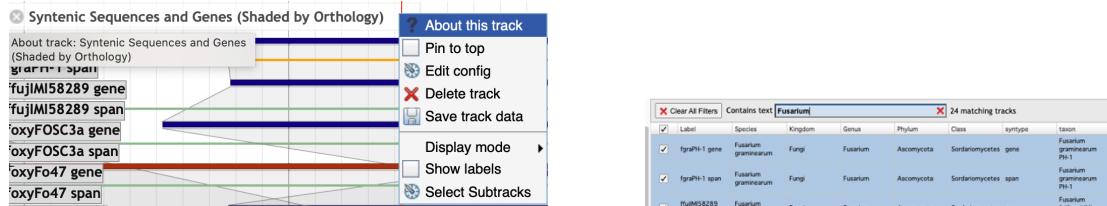
 - H3K27me3_WT_low
 - H3K4me2_WT_low
- Next, activate the **Syntenic Sequences and Genes (Shaded by Orthology)** track, which is located under the **Comparative Genomics** category, **Orthology and Synteny** subcategory but can be also searched for in the **Contains text** search window at the top (*hint: you can also use the filter*).

Select Tracks			
My Tracks		Back to browser Clear All Filters Contains text	
Category	Name	Category	Subcategory
Comparative Genomics	<input checked="" type="checkbox"/> Syntenic Sequences and Genes (Shaded by Orthology)	Comparative Genomics	Orthology and Synteny
Orthology and Synteny			
(no data)			
Segments			
RNASeq Alignment			
(no data)			
RNASeq Strand			
(no data)			

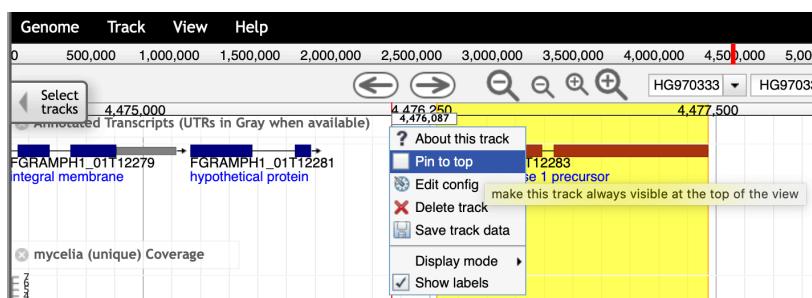
- Click on the **Back to browser** button
- Customise JBrowse view and organise tracks

- 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.*



- Pin the highlighted gene and gene model track to the top.



- Now, you can re-organise your tracks by drag and drop. Here is the order in which the tracks are shown below:
 - Gene model (pinned to top)
 - RNA-Seq evidence: mycelia
 - RNA-Seq evidence: 6dpi
 - Intron Evidence 1
 - Intron evidence 2
 - ChiP-Seq H3K27 WT
 - ChiP-Seq H3K4me2 WT H3K27
 - ChiP-Seq H3K27 kmt6
 - ChiP-Seq H3K4me2 kmt6

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.

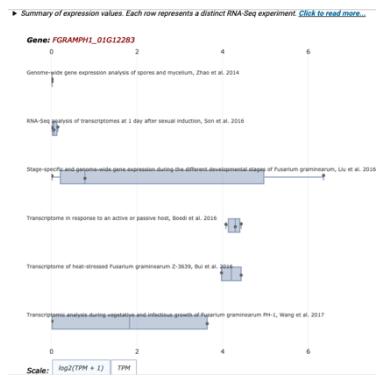
Kmt6 encodes a DNA methylase that deposits repressive chromatin mark H3K27me3. *kmt6* mutants in *Fusarium* and other species demonstrate reorganization of chromatic and upregulation of genes that are normally suppressed.

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? How would you go about testing if it is not expressed in this stage of development or is an artefact?

Hint: take a look at the transcriptomics summary graph on the gene record page or other transcriptomic tracks



- Is there evidence of alternative splicing?
- Is this gene conserved in *Fusarium* species?
- Let's say you would like to share your JBrowse view with a colleague. How would you go about getting the link?

