

Exercise: Ensembl Fungi gene trees and homologues

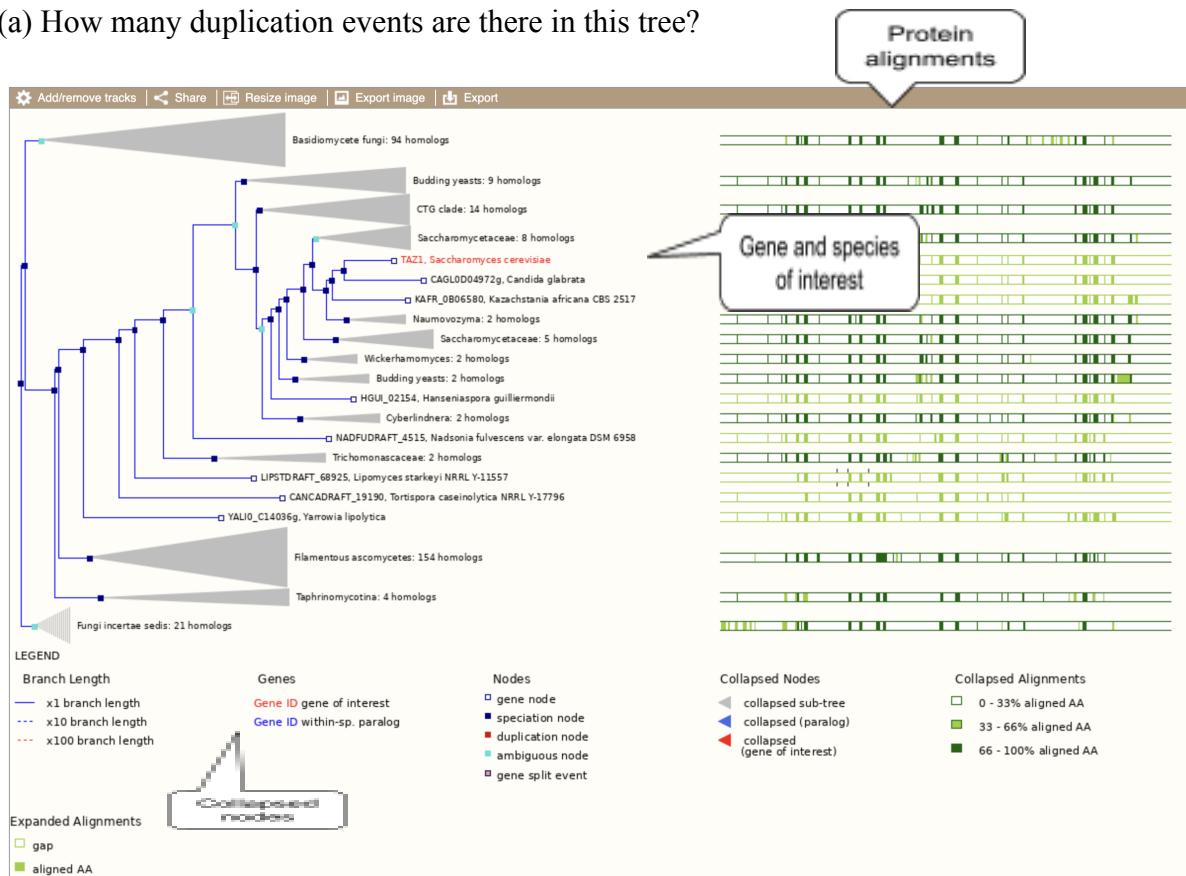
Let's look at the homologues of *Saccharomyces cerevisiae* TAZ1 (Gene stable ID: YPR140W). This gene is involved in stress response and conserved across different taxonomic domains. Search for the gene and go to the Gene tab.

The screenshot shows the Ensembl Fungi interface for *Saccharomyces cerevisiae* (R64-1-1). The top navigation bar includes the species logo, the species name, and the gene identifier TAZ1. The left sidebar contains a tree menu with categories like Gene-based displays, Sequence, Fungal Compara, and Pan-taxonomic Compara. Under Fungal Compara, the 'Gene tree' option is highlighted. The main content area displays the gene details for TAZ1, including its description, location, and links to about this gene and transcripts. A large button labeled 'Gene tree ?' is prominently displayed.

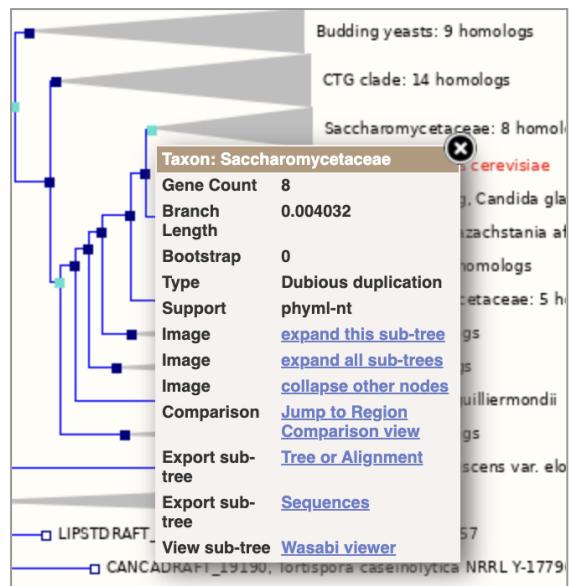
Click on **Fungal Compara: Gene tree**, which will display the current gene in the context of a phylogenetic tree used to determine orthologues and paralogues.

The screenshot shows the 'Gene tree' interface for the gene TAZ1. At the top, it displays the unique gene tree stable ID (EGT0105000064920) and various summary statistics: 327 genes, 295 speciation nodes, 13 duplication nodes, 18 ambiguous nodes, and 0 gene split events. Below this, there are checkboxes for filtering by InterPro and GO terms, and a dropdown menu to show 10 entries. A callout box points to the 'Filter tree by Gene Ontology (GO) terms or InterPro protein domains' link. The main area is a table listing 121 entries, each with a highlight status (327 members), accession number, and description. The descriptions include molecular function, catalytic activity, lipid metabolic process, phospholipid metabolic process, phosphorus metabolic process, phosphate-containing compound metabolic process, biological process, metabolic process, cellular process, and transferase activity. The table includes a filter input field and a navigation bar at the bottom.

(a) How many duplication events are there in this tree?



Funnels indicate collapsed nodes. Click on a node (coloured square) to get a pop-up. We can then see what type of node this is, some statistics and options to expand or export the sub-tree:



There are some quick filtering options below the image, where you can add paralogues, and quickly expand or collapse nodes:

View options:

- [View current gene only](#) (Default)
- [View paralogues of current gene](#)
- [View all duplication nodes](#)
- [View fully expanded tree](#)
- Collapse all the nodes at the taxonomic rank

Use the 'configure page' link in the left panel to see more options available from menus on individual tree nodes.

Ensembl Fungi release 56 - Feb 2023 © EMBL-EBI

About Us

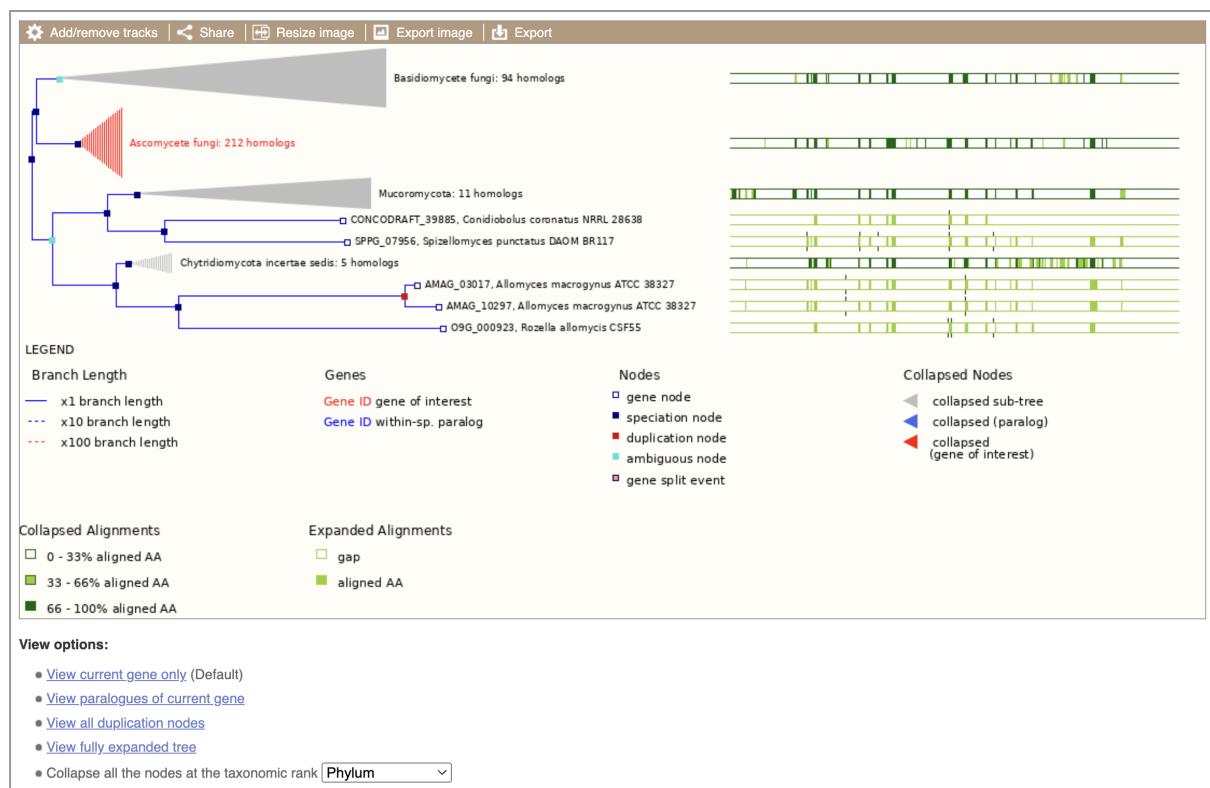
[About us](#)

[Using this website](#)

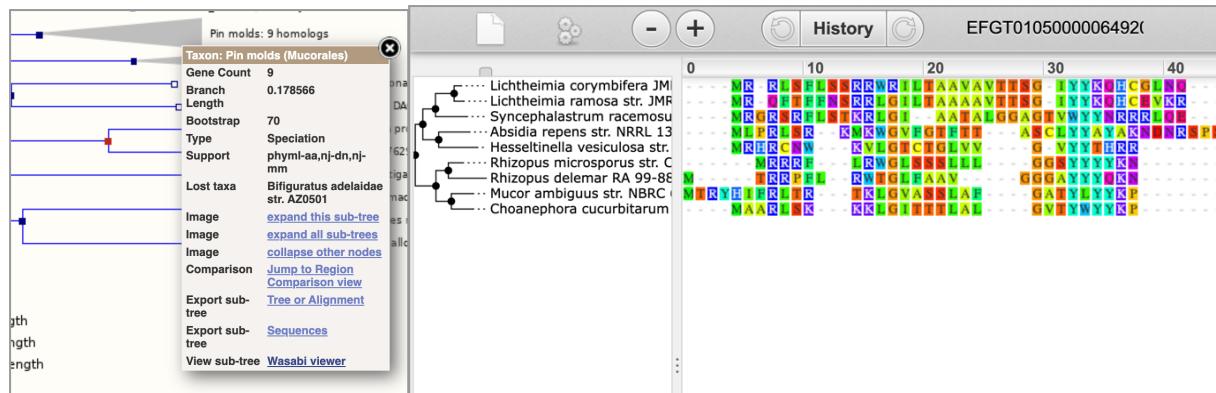
Our sister sites

[Ensembl](#)

(b) What is the Phylum with the highest number of *TAZ1* homologues?



(c) What is the bootstrap support of the pin moulds (*Mucorales*) Class? Can you display the sequence alignment of all the homologues in this Class (Hint: Use the Wasabi viewer)?



You can download the tree in a variety of formats. Click on the [Export](#) icon  in the bar at the top of the image to get a pop-up where you can choose your format. You can preview this file before you download.

We can look at homologues in the [Orthologues](#) and [Paralogues](#) pages, which can be accessed from the left-hand menu. If there are no orthologues or paralogues, then the name will be greyed out. Click on [Orthologues](#) to see the orthologues available.

Orthologues ?

Download orthologues

Hover over orthologue types of description

Summary of species with orthologues

Summary of orthologues of this gene [Hide](#)

Click on 'Show details' to display the orthologues for one or more groups of species. Alternatively, click on 'Configure this page' to choose a custom list of species.

Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
All (370 species)	<input checked="" type="checkbox"/>	298	14	0	58
Acidomyces (1 species)	<input type="checkbox"/>	1	0	0	0
Agaricales (21 species)	<input type="checkbox"/>	18	1	0	2
Atheliales (2 species)	<input type="checkbox"/>	1	1	0	0
Blastocladiales (1 species)	<input type="checkbox"/>	0	1	0	0
Boletales (9 species)	<input type="checkbox"/>	6	0	0	3
Botryosphaerales (2 species)	<input type="checkbox"/>	2	0	0	0
Cantharellales (3 species)	<input type="checkbox"/>	1	1	0	1

Select taxon of interest

Download table

Similarity metrics

Filter table

Orthologue details by species

Selected orthologues [Hide](#)

Species	Type	Orthologue	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence
Absidia repens str. NRRL 1336	1-to-1	BCR42DRAFT_405738 View Gene Tree View Sequence Alignments	23.74 %	17.32 %	n/a	n/a	No
Acaromyces ingoldii str. MCA 4198	1-to-1	FA10DRAFT_281454 View Gene Tree View Sequence Alignments	24.93 %	n/a	n/a	n/a	No
Acidomyces richmondensis BFW	1-to-1	M433DRAFT_12235 View Gene Tree View Sequence Alignments	27.41 %	20.25 %	n/a	n/a	Yes
Acromonium chrysogenum ATCC 11550	1-to-1	ACRE_050350 View Gene Tree View Sequence Alignments	32.27 %	n/a	n/a	n/a	Yes
Agaricus bisporus var. burrettii JB137-S8	1-to-1	AGABI1DRAFT_91626 View Gene Tree View Sequence Alignments	29.32 %	23.62 %	n/a	n/a	No

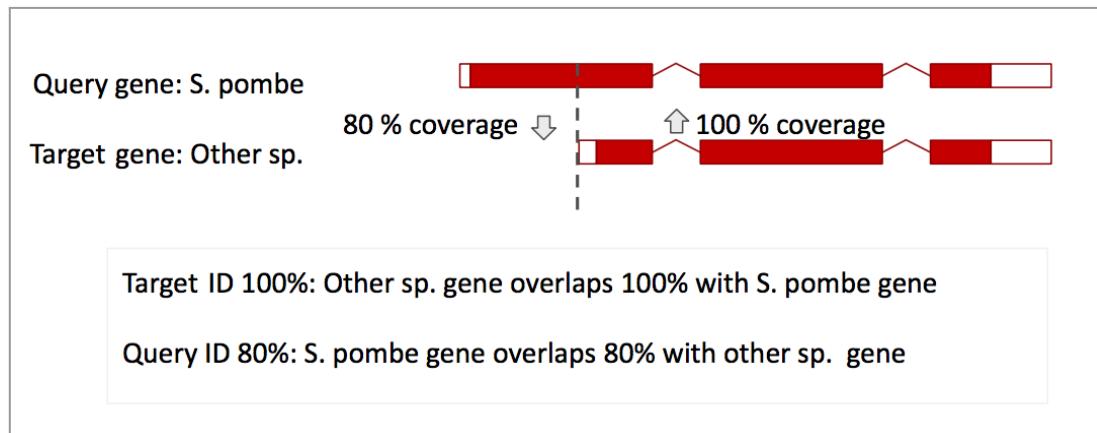
Link to orthologue gene tab

View region comparison of orthologues

View protein of cDNA sequence alignment

(d) What is the difference between Target %id and Query %id? (*Hint: Mouse over*)

The sequence identity is reported in two ways, Target %id is how much of the orthologue or ‘target gene’ overlaps with the query gene, or our *S. cerevisiae* gene. The Query %id is the inverse of this. For example:



Scroll to the bottom of the page to see a list of the species that do not have any orthologues with *TAZ1* in *Saccharomyces cerevisiae*... there's a lot!

Species without orthologues

58 species are not shown in the table above because they don't have any orthologue with YPR140W.

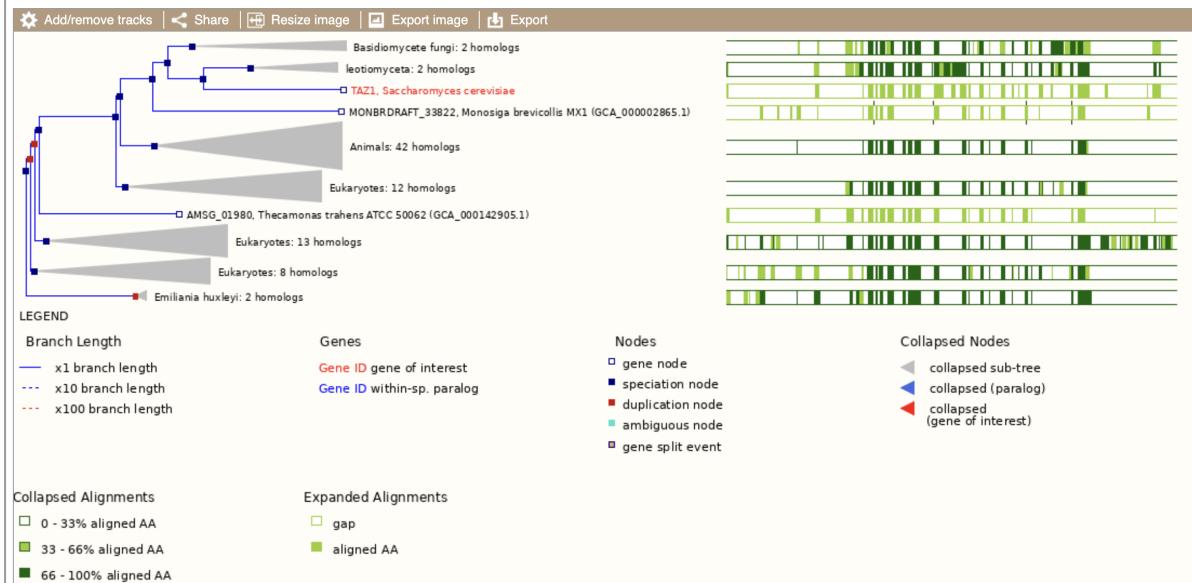
- *Amphiamblys* sp. WSBS2006
- *Anncalilia algerae* PRA339
- *Aspergillus flavus* NRRL3357
- *Aspergillus nidulans*
- *Aspergillus terreus* NIH2624
- *Batrachochytrium dendrobatidis* JEL423
- *Bifiguratus adelaideae* str. AZ0501

Saccharomyces cerevisiae is part of Pan-compara, which compares a subset of fungal species with species from other taxa, such as plants, bacteria and vertebrates. Go to [Pan-taxonomic Compara > Gene Tree](#). Let's have a look at the Pan-taxonomic tree with nodes collapsed at the Kingdom rank.

Gene Tree

GeneTree EGGT0005000021121

Number of genes	84
Number of speciation nodes	64
Number of duplication nodes	16
Number of ambiguous nodes	3
Number of gene split events	0



Click on Pan-taxonomic Compara > Orthologues now.

Orthologues

Download orthologues

Summary of orthologues of this gene [Hide](#)

Click on 'Show details' to display the orthologues for one or more groups of species. Alternatively, click on 'Configure this page' to choose a custom list of species.

Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
All (370 species)	<input checked="" type="checkbox"/>	4	0	0	366
Acidomyces (1 species)	<input type="checkbox"/>	0	0	0	1
Agaricales (21 species)	<input type="checkbox"/>	0	0	0	21
Atheliales (2 species)	<input type="checkbox"/>	0	0	0	2
Blastocladiales (1 species)	<input type="checkbox"/>	0	0	0	1
Boletales (9 species)	<input type="checkbox"/>	0	0	0	9

Selected orthologues [Hide](#)

Species	Type	Orthologue	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence
Aedes aegypti (LVP_AGWG)	1-to-1	AAEL001564	25.85 %	19.95 %	n/a	n/a	No
	View Gene Tree	2:21,496,991-21,541,309:-1					
	View Sequence Alignments						
Amborella trichopoda	1-to-1	AMTR_s00022p00068080	23.08 %	17.32 %	n/a	n/a	No
	View Gene Tree	AmTr_v1.0_scaffold00022:710,032-717,504:-1					
	View Sequence Alignments						

(d) How many species with predicted orthologues for this gene are there in Fungal Compara? What about in Pan-compara?

Fungal Compara:

Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
All (370 species)	<input checked="" type="checkbox"/>	298	14	0	58

Pan Compara:

Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
All (370 species)	<input checked="" type="checkbox"/>	4	0	0	366

(e) How many animal orthologues are there? Does this number agree with the Pan-taxonomic tree above? (*Hint: Click the 'Show details' box for Vertebrates and Metazoa, and count the number of orthologues in the table below*).

(f) Filter the second table to view the human orthologue. How much sequence identity does the human protein have to the *Saccharomyces cerevisiae* one? Is it a high confidence homology? Click on the [View Sequence Alignment](#) link in the Orthologue column to [View Protein Alignment](#) in Clustal W format. Does it support your conclusions?

Selected orthologues [Hide](#)

Show All entries		Show/hide columns		human				
Species	Type	Orthologue		Target %id	Query %id	GOC Score	WGA Coverage	High Confidence
Human	1-to-1	TAFazzin (ENSG00000102125)	View Sequence Alignments	24.66 %	18.90 %	n/a	n/a	No
		X:154,411,524-154,421,726:1						
			Orthologue Alignment					
Pediculus humanus	1-to-1	PHUM309640	View Protein Alignment	18.90 %	n/a	n/a	n/a	No
		DS235308:45,836-47,144:-1	View cDNA Alignment					
			View Sequence Alignments					

Orthologue Alignment

 Download homology

Type: 1-to-1 orthologues

Species	Gene ID	Peptide ID	Peptide length	% identity (Protein)	% coverage	Genomic location
Saccharomyces cerevisiae	YPR140W	YPR140W	381 aa	18 %	65 %	XVI:814391-815536
Human	ENSG00000102125	ENSP00000469981	292 aa	24 %	85 %	X:154411524-154421726

CLUSTAL W (1.81) multiple sequence alignment

YPR140W/1-381 MSFRDVL-----ERGDEFLEAYPRRS---PLWRFLSYSTSLLTGVSKLLLFTCYNV
ENSP00000469981/1-292 -----MPLHVKW-----PFP---AVPPLTWTLASSVVMGLVGTYSFCFWTKYMNHL
.:*: * : * . * . * : * : :

YPR140W/1-381 KLNQFEKLETALERSKRENRGLMTVMNHMSMVDDPLVVATLPYKLFTSLDNIRWSLGAHN
ENSP00000469981/1-292 TVHNREVLYELIEK-RGPATPLITVSNHQSCMDDPHLGILKLRHIWNLKLMRWTPAAAD
.::.. * * ;*: : *;** ** * ;*** :*. * : : .*. ;**: .* :

YPR140W/1-381 ICFQNKFANFFSLGQVLSTER-----FGVGPQGS
ENSP00000469981/1-292 ICFTKELHSHFFSLGKCVPCRGAEFFQAENEKGKVLDTGRHMPGAGKRREKGDGVYQKG
*** :; :;*****: :.. *

Additional Exercise 1 - *Zymoseptoria* orthologues

Exploring an orthologue that we identified using BioMart (additional exercise 1). We identified 18 genes associated with the GO term detoxification in *Zymoseptoria tritici*. We then found a single high confidence orthologue in *Cryptococcus neoformans* which we will now explore further.

Search for CNC06590 in *Cryptococcus neoformans* var. *neoformans* JEC21 to go to the gene page. Click on the gene ID [CNM01690](#) to go to the gene page.

Search: for

e.g. [NAT2](#) or [alcohol*](#)

(a) Does this gene in *C. neoformans* have a UniProtKB-Gene Ontology annotation?

Location: 13:510,531-512,507 Gene: CNM01690 Transcript: AAW46801

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- Sequence
 - Secondary Structure
 - Gene families
- Literature
- Fungal Compara
 - Genomic alignments
 - Gene tree
 - Gene gainloss tree
 - Orthologues
 - Paralogues
- Pan-taxonomic Compara
 - Gene Tree
 - Orthologues
- Ontologies
 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process
 - PHi: Phibase identifier
- Phenotypes
- Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
- Gene expression
- Pathway
- Molecular interactions
- Regulation
- External references
- Supporting evidence
- ID History

Gene: CNM01690

Description: formaldehyde dehydrogenase (glutathione), putative

Location: Chromosome 13: 510,531-512,507 reverse strand.

About this gene: This gene has 1 transcript ([splice variant](#)), [259 orthologues](#) and [27 paralogues](#).

Transcripts: [Show transcript table](#)

External references

This gene corresponds to the following database identifiers:

External database	Database identifier
NCBI gene (formerly Entrezgene)	3255046 [View all locations]
UniGene	Ene_7413 [View all locations]
UniProtKB-Gene Ontology Annotation	Q9K7P9 [View all locations]

The following database identifiers correspond to the transcripts of this gene:

Transcript ID
AAW46801

Ensembl Fungi release 56 - Feb 2023 © EMBL-EBI

(b) Find the *Z. tritici* orthologue in the [Orthologues](#) page and view a protein alignment.

Selected orthologues [Hide](#)

Species	Type	Orthologue	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence
Zymoseptoria tritici	1-to-1	MHGOG1 (Mycgr3G76502)	84.59 %	82.74 %	n/a	n/a	No

[View Gene Tree](#) 10:1,108,132-1,110,166:-1
[View Sequence Alignments](#)

[Orthologue Alignment](#) [View Protein Alignment](#) [View cDNA Alignment](#)

(c) At which end of the protein (N- or C-terminus) does the alignment between these two genes become worse?

Orthologue alignment

 Download homology

Type: 1-to-1 orthologues

Species	Gene ID	Peptide ID	Peptide length	% identity (Protein)	% coverage	Genomic location
Cryptococcus neoformans var. neoformans JEC21	CNC06590	AAW42642	365 aa	82 %	97 %	3:1927422-1929917
Zymoseptoria tritici	Mycgr3G76502	Mycgr3P76502	357 aa	84 %	99 %	10:1108132-1110166
CLUSTAL W (1.81) multiple sequence alignment						
AAW42642/1-365 Mycgr3P76502/1-357		MADEFVKLSIFGTVFEVTTTRYVDLQPVGGMGAFGLVCSAKDQLSCTSVAIKKIMKPFPSTPVLM MAEFVRAQIFGTTFEITTSRYTDLQPVGGMGAFGLVCSAKDQLTGQAVAVKKIMKPFPSTPVLM ***:***: .****.*;*:***.*****:*****:*****:*****:*****:*****:*****:*****				
AAW42642/1-365 Mycgr3P76502/1-357		SKRTYRELKLLKHILRHENIISLSDIFISPLEDIYFVTTELLGTDLHRLLTSPLEKQFQIY SKRTYRELKLLKHILKHENVISLSDIFISPLEDIYFVTTELLGTDLHRLLTSPLEKQFQIY *****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****				
AAW42642/1-365 Mycgr3P76502/1-357		FLYQILRGLKYVHSAGVVHRLDKPSNILVNENCDLKICDFGLARIQDPQMTGYVSTRYYR FLYQILRGLKYVHSAGVVHRLDKPSNILVNENCDLKICDFGLARIQDPQMTGYVSTRYYR *****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****				
AAW42642/1-365 Mycgr3P76502/1-357		APEIMLTWQKYDVAVDIWSTGCIFAEMLEGKPLFPKGDKDHVNQFSIITELLGTPPPDDVIQT APEIMLTWQKYDVEDIWSACCIPAEMLEGKPLFPKGDKDHVNQFSIITDLLGTPPPDDVIST *****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****				
AAW42642/1-365 Mycgr3P76502/1-357		IASENTLRFVQSLPKREKVPFSTKFPNADPVSLDLLEKMLVFDPRTRISAAEGLAHEYLA ICSENTLRFVOSLPKRERQPLKNKFKNADPQAIELLERMLVFDPRKRVKAGEALADPYLS *.*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****				
AAW42642/1-365 Mycgr3P76502/1-357		PYHDPTDEPVAAEVFDWSFNDADLPVDTWKVMMYSEILDFHNLGDISQNE--AEGPVVTGE PYHDPTDEPEAEERKDWSFNDADLPVDTWKIMMYSEILDYHNVDS-ANNGEGQE--NGG *****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****				
AAW42642/1-365 Mycgr3P76502/1-357		VPAAPAS A-----				

Additional Exercise 2 - Mushroom genes

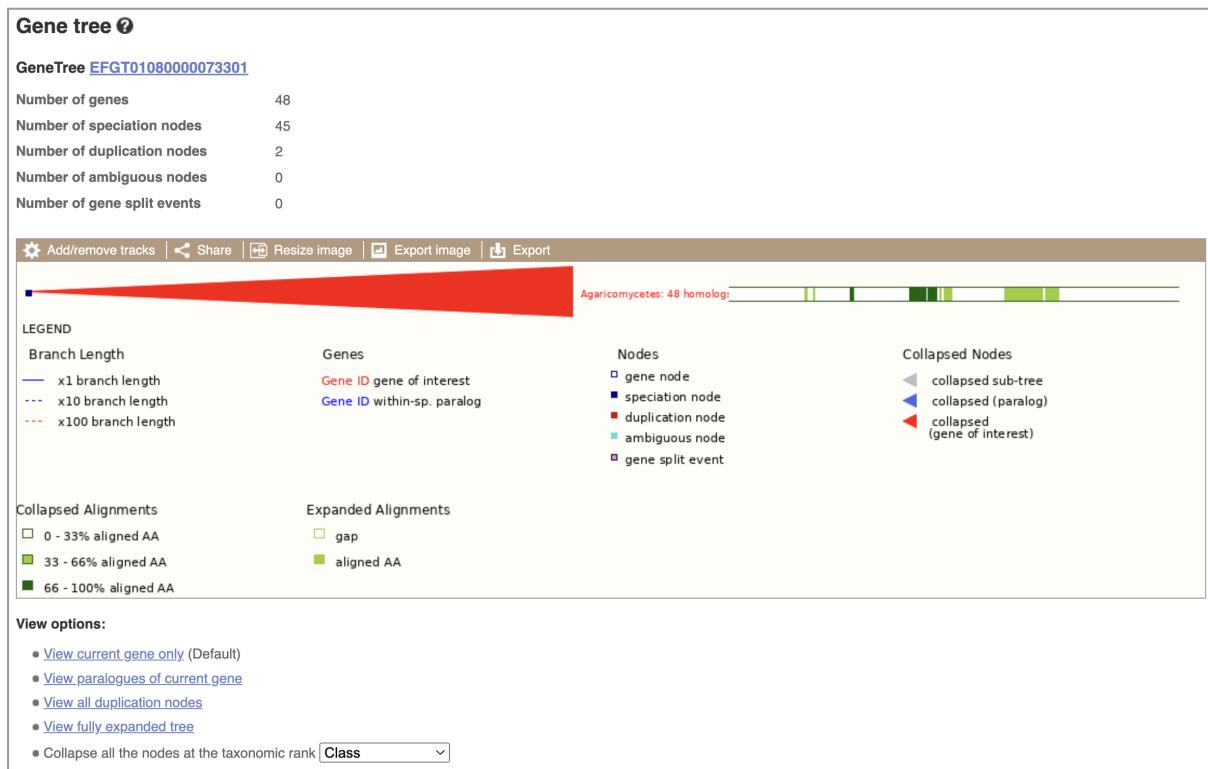
We're going to take a look at the gene [CC1G_05700](#) in *Coprinopsis cinerea* okayama7#130.

Search: for

e.g. [NAT2](#) or [alcohol](#)*

From the gene tab, click to view the [Gene tree](#). At the bottom of the image click to collapse all the nodes at the taxonomic rank of [Class](#).

(a) What do you notice about the types of fungi shown in the gene tree?



(b) Does this match with what you would expect from the gene description? (*Hint: Agaricomycetes class belongs to the Basidiomycota phylum*)

Gene: CC1G_05700

Description	basidiospore development protein
Location	Chromosome 7: 2,117,260-2,118,876 forward strand. CC3:AACS02000007.1
About this gene	This gene has 1 transcript (splice variant) and 47 orthologues .
Transcripts	Hide transcript table

Show/hide columns (1 hidden)		Filter	Export		
Name	Transcript ID	bp	Protein	Biotype	Flags
-	EAU90162	1389	462aa	Protein coding	Ensembl Canonical

(c) Based on the protein alignment shown at the right, can you predict which end of the gene/protein is most conserved?



(e) Click to view the [Orthologues](#) page. In the Selected orthologues table, find the entry for the species *Amanita thiersii* and click to view a protein alignment. Does this support your conclusion about the conserved region of the gene/protein?

Selected orthologues Hide								
Show All entries Amanita thiersii 								
Show/hide columns								
Species	Type	Orthologue	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence	
Amanita thiersii Skay4041	1-to-1 View Gene Tree	AMATHDRAFT_122148 KZ301993:102,546-102,928:-1 View Sequence Alignments	42.73 %	10.17 %	n/a	n/a	No	

[Orthologue Alignment](#) [View Protein Alignment](#) [View cDNA Alignment](#)

Orthologue alignment ②

Download homology

Type: 1-to-1 orthologues

Species	Gene ID	Peptide ID	Peptide length	% identity (Protein)	% coverage	Genomic location
Coprinopsis cinerea okayama7#130	CC1G_05700	EAU90162	462 aa	10 %	21 %	7:2117260-2118876
Amanita thiersii Skay4041	AMATHDRAFT_122148	PFH51030	110 aa	42 %	90 %	KZ301993:102546-102928

CLUSTAL W (1.81) multiple sequence alignment

```

EAU90162/1-462      -----MRVLLHDQTQMNLKFGSHVEALISNVKETSQELRKTSSTFEQEHDKLLG
PFH51030/1-110      PLTFLDKNATSMRVLHLDTQANFEKFSTRVNDFNLGAETKESEINLVKSLSLFERGOETLTN
                     ***** *;**** :*:: :...: **..*: ..* **: ::.* .

EAU90162/1-462      DIIDLVNRCSQIQLGSPPAQASGMQLSKDINQLDCLDKRLDAIQTV-----
PFH51030/1-110      *****: .*.*: .:*****: ;:: : .::: **:.*:*****: ;.

EAU90162/1-462      QIQAQNLLQQQNLLINAVTPLLLQLPQLPRLAPSTSLANFNSQTQRTDASSQTIEKRO
PFH51030/1-110      ----

EAU90162/1-462      PSYHQETLRRQRVRVDSDIQEISPCKPLPGSAQKKRRIESPRSVQKPSLELTQRLFPSSSP
PFH51030/1-110      ----

EAU90162/1-462      DLIKYSTDSEGPKTPQVNERSAPIVTPRRPLQDLFPFFPGSNQRSVSKRPMPPSSTRLV
PFH51030/1-110      ----

EAU90162/1-462      GPGKSATPGPSRVGAESRAALARPLIKPLAIAPLAFSSTSKTPVHISNFTPKPVTPSL
PFH51030/1-110      ----

EAU90162/1-462      RNAVAGEGRALKIAQTPQVLKNERMTSQAAKNNTMPPAGMVSLRSSTTTATAKPTS
PFH51030/1-110      ----

EAU90162/1-462      NTPRFGEPEANKPPLLRAPTNNNGPRPLQERMKEPVREGRRFIPLVDTDDDEDDSD
PFH51030/1-110      ----

```