

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.

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 Ensembl Fungi gene tree interface for *Saccharomyces cerevisiae* (R64-1-1). The top navigation bar includes the location (XVI:814,391-815,536), gene (TAZ1), and transcript (TAZ1). The left sidebar has a tree menu with options like Summary, Sequence, Fungal Compara, and Pan-taxonomic Compara. The 'Gene tree' option is highlighted. The main content area displays the gene tree summary statistics and a table of ontology terms.

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- Sequence
 - Secondary Structure
 - Gene families
 - Literature
- Fungal Compara
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Pan-taxonomic Compara
 - Gene Tree
 - Orthologues

Gene: TAZ1 YPR140W

Description

Location

About this gene

Transcripts

Gene tree ?

Gene tree

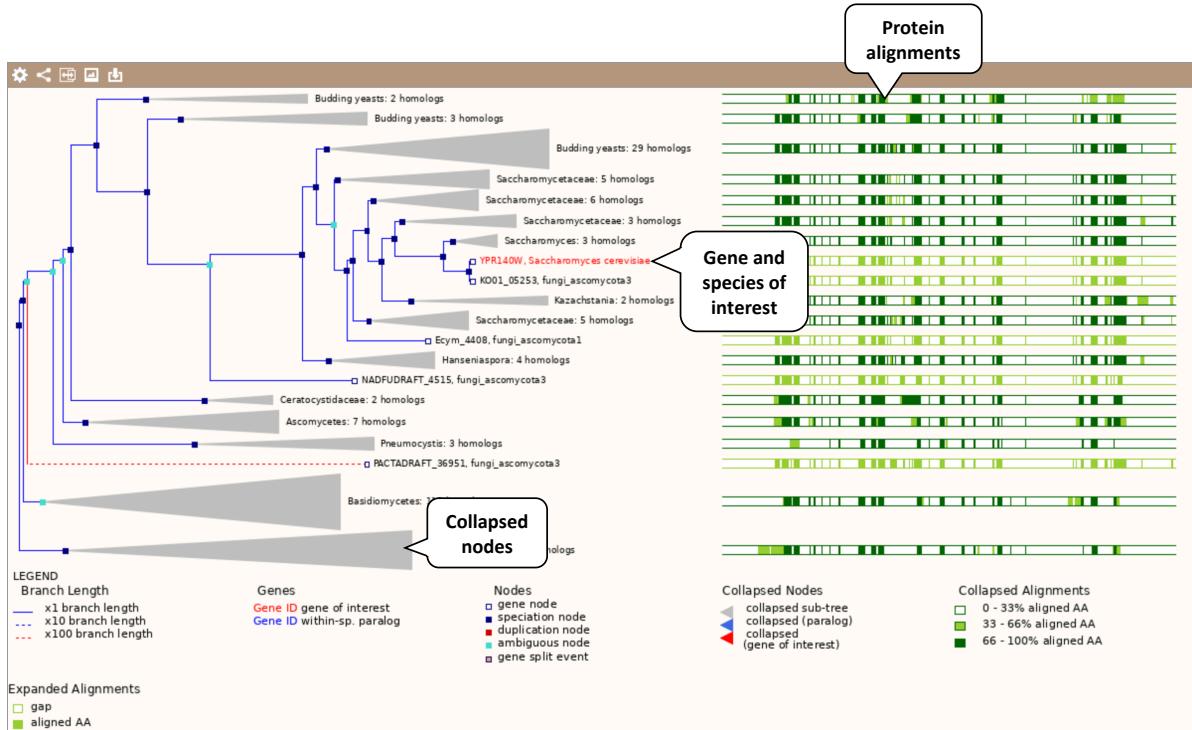
Summary statistics

Unique Gene Tree Stable ID

Filter Tree by Gene Ontology terms or InterPro protein domains

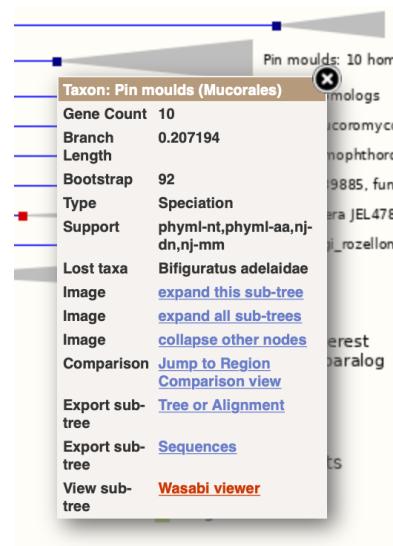
highlight	Accession	Description
182 members	GO:0003674	molecular_function
182 members	GO:0003824	catalytic activity
182 members	GO:0006629	lipid metabolic process
182 members	GO:0006644	phospholipid metabolic process
182 members	GO:0006793	phosphorus metabolic process
182 members	GO:0008798	phosphate-containing compound metabolic process
182 members	GO:0008150	biological_process
182 members	GO:0008152	metabolic process
182 members	GO:0009987	cellular process
182 members	GO:0016740	transferase activity

(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 ...

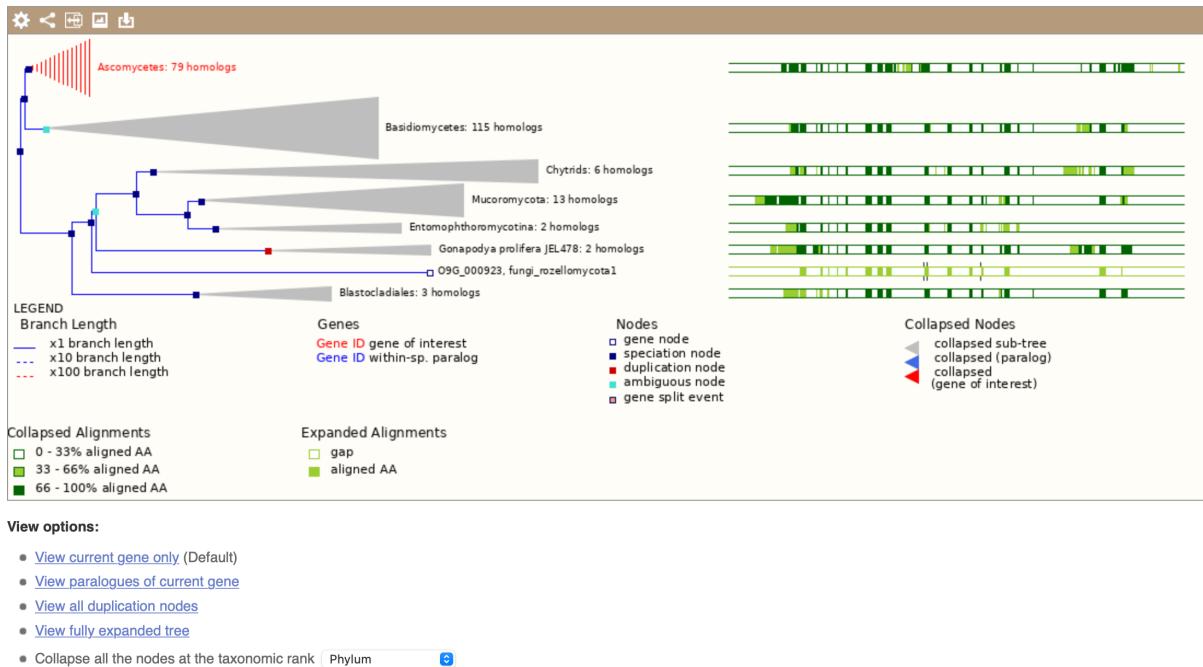
Ensembl Fungi Test Site release 51 - April 2021 © EBI

About Us

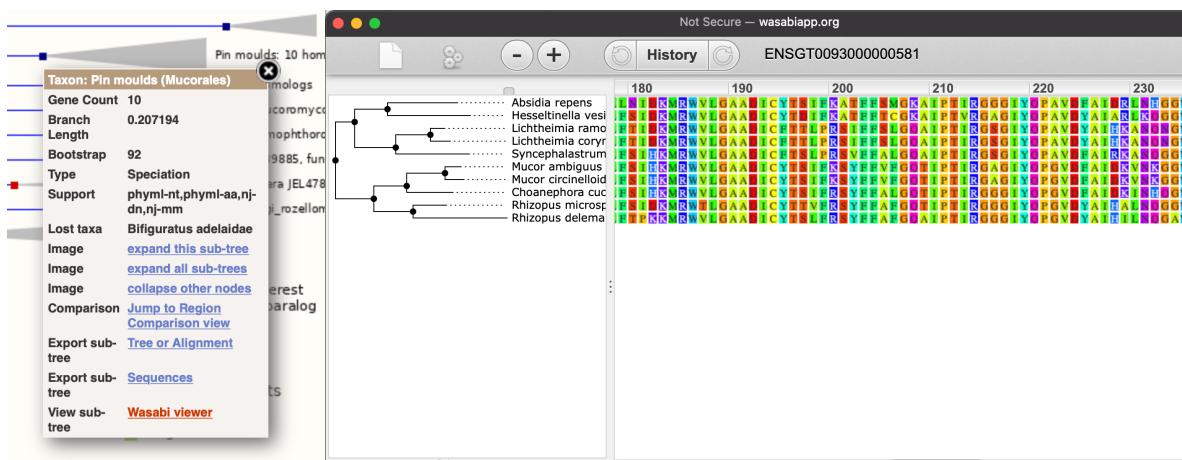
✓ -- Select a rank--

Species
Genus
Family
Order
Class
Phylum
Kingdom

(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. From the pop-up node menu above you can click to [Export sub-tree: Tree or Alignment / Sequences](#). Alternatively, click on the

download 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.

The screenshot shows a user interface for sequence alignment formats. At the top, there are fields for 'File name:' (TAZ1_gene_tree) and 'File format:' (FASTA). Below these are buttons for 'Preview', 'Download', and 'Download Compressed'. A 'Guide to file formats' section provides examples for CLUSTALW, FASTA, and MEGA.

- CLUSTALW:**

```
CLUSTAL W(1.81) multiple sequence
homo_sapiens/1-465588 CCTCAGGAC
pan_troglodytes/1-465588 *****
```
- FASTA:**

```
>homo_sapiens/1-464308
CCTCAGGACCGAGCCAAACCAACCCAGAT
CCCAGTCCTTGGACTGCCTCTGGGGCC
TGGGACAGAGAGAACACACAGCTGGCTC
AGGGGCCCTGTGGGGTTAGATCAAA
CCCCCTATGATGATGATGATGATGATGAT
CCAGGCTCTGTGAAAMAGGTTGGCTGTTG
AGGAAGACCGTGGCTCTGGCTCTGTTT
AAAGATGGGGGGTGGGTGGATTCTCTT
GGGAGAGGGAAAGAGAAAGGGCCCTGGG
```
- Mega:**

```
#mega
!Title: ProjectedMultiAlign;
!Format datatype=dna identical=.

#homo_sap CCTCAGGAC GACGGCAAAC
#pan_trop ..C..... .
#homo_sap CCCAGTCCT TCGACTGCCT
#pan_trop ..... .
#homo_sap TGGGACAGAG AGAGAACAC
```

Below the guide, sections for MSF and Newick are shown, each with their respective alignment formats.

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](#)

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 the search' to change the search parameters.

Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
All (487 species)	<input checked="" type="checkbox"/>	200	10	0	277
Acidomyces (1 species)	<input type="checkbox"/>	0	0	0	1
Agaricales (24 species)	<input type="checkbox"/>	20	1	0	3
Atheliales (2 species)	<input type="checkbox"/>		1	0	0
Blastocladiales (2 species)	<input type="checkbox"/>		1	0	0
Boletales (10 species)	<input type="checkbox"/>	0	0	0	4
Botryosphaerales (4 species)	<input type="checkbox"/>	0	0	0	4

Summary of species with orthologues

Choose a taxon of interest

Orthologue types: mouse over for definitions

Selected orthologues [Hide](#)

Show All entries Show/hide columns

Species	Type	Orthologue	Link to the orthologue gene tab	Similarity metrics	Filter table	Download table		
[<i>Candida</i>] auris str. 6684	1-to-1	QG37_0516	View Gene Tree	Target %id 49.33 %	Query %id 48.29 %	GOC Score n/a	WGA Coverage n/a	High Confidence n/a
[<i>Candida</i>] glabrata str. CBS138	1-to-1	CAGL0D04972g	View Gene Tree	62.23 %	60.10 %	n/a	n/a	n/a
<i>Absidia</i> repens str. NRRL 1336	1-to-1	BCR42DRAFT_4057	View Gene Tree	22.66 %	16.54 %	n/a	n/a	n/a

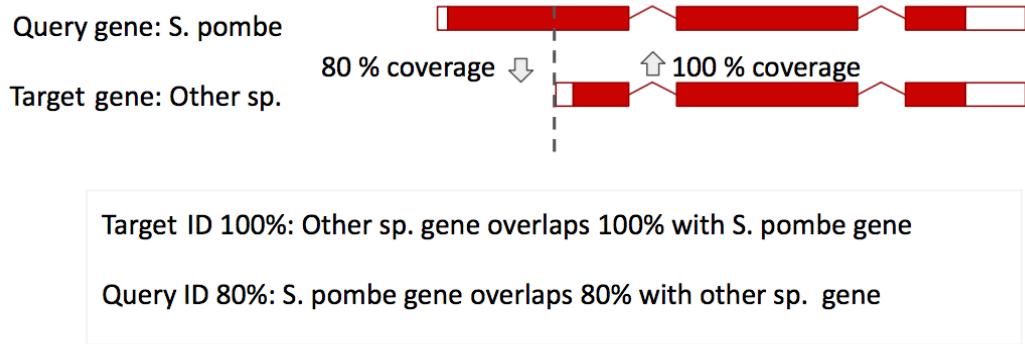
Orthologue details per species

Link to the Region Comparison with the aligned homologues side

View the protein or cDNA 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:



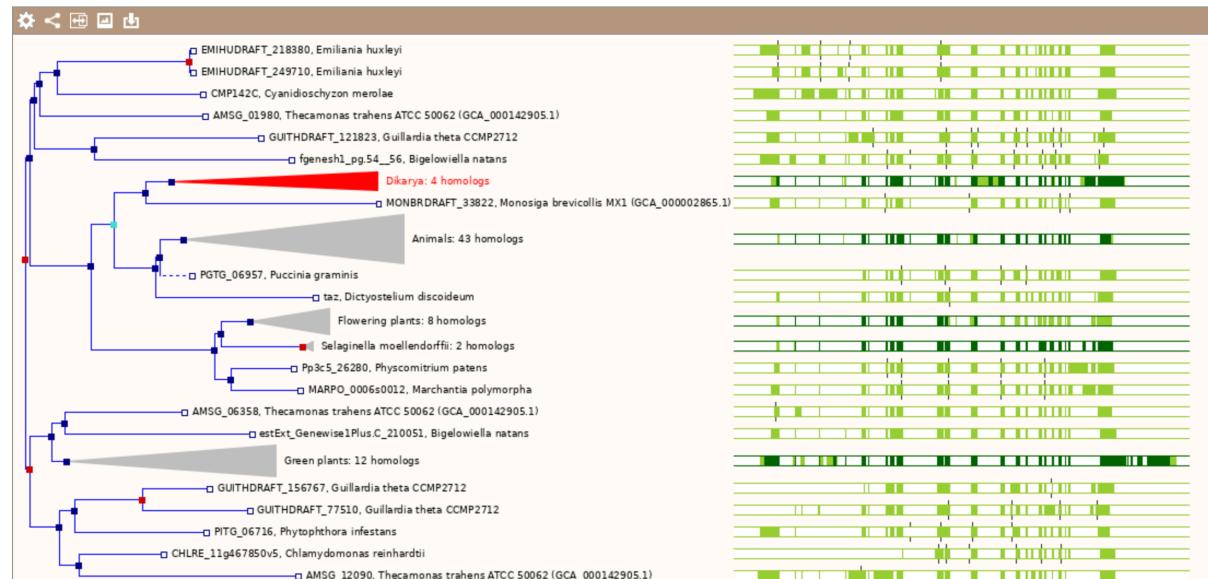
Scroll to the bottom of the page to see a list of the species that do not have any orthologues with *TAZ1* in *Saccharomyces cerevisiae*... there's a lot!

Species without orthologues	
277 species are not shown in the table above because they don't have any orthologue with YPR140W.	
●	Acidomyces richmondensis BFW (GCA_001592465)
●	Acremonium chrysogenum ATCC 11550 (GCA_000769265)
●	Alternaria alternata str. SRC1IrK2f (GCA_001642055)
●	Amorphotheca resinae ATCC 22711 (GCA_003019875)
●	Amphiamblys sp. WSBS2006 (GCA_001875675)
●	Anncalbia algerae PRA339 (GCA_000385875)
●	Ascochyta rabiei str. ArDII (GCA_001630375)

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 look at the Pan-taxonomic tree with nodes collapsed at the Kingdom rank.

GeneTree EGGT00050000021121

Number of genes	87
Number of speciation nodes	65
Number of duplication nodes	15
Number of ambiguous nodes	4
Number of gene split events	2



Click on Pan-taxonomic Compara > Orthologues now.

Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
All (174 species)	<input checked="" type="checkbox"/>	45	12	0	117
Vertebrates (12 species)	<input type="checkbox"/>	11	1	0	0
Metazoa (24 species)	<input type="checkbox"/>	17	6	0	1
Plants (11 species)	<input type="checkbox"/>	7	4	0	0
Fungi (6 species)	<input type="checkbox"/>	4	0	0	2
Protists (13 species)	<input type="checkbox"/>	6	1	0	6
Bacteria (84 species)	<input type="checkbox"/>	0	0	0	84
Archaea (24 species)	<input type="checkbox"/>	0	0	0	24

Selected orthologues [Hide](#) 

Show All entries		Show/hide columns						Filter
Species	Type	Orthologue	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence	
Aedes aegypti (LVP_AGWG)	1-to-1	AAEL001564	24.49 %	18.90 %	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	24.13 %	18.11 %	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 (487 species)	<input checked="" type="checkbox"/>	200	10	0	277

Pan Compara:

Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
All (174 species)	<input checked="" type="checkbox"/>	45	12	0	117

(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?

Show All entries		Show/hide columns						human
Species	Type	Orthologue	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence	
Human	1-to-1	TAZ (ENSG00000102125)	23.97 %	18.37 %	n/a	n/a	No	
		View Gene Tree	X:154,411,524-154,421,726:1					
			View Sequence Alignments					
Pediculus humanus	1-to-1	PHUM309640	59 %	n/a	n/a	n/a	No	
		View Gene Tree	DS235308:45,836-47,144:-1					
			View Sequence Alignments					

Orthologue Alignment



Type: 1-to-1 orthologues

Species	Gene ID	Peptide ID	Peptide length	% identity (Protein)	% coverage	Genomic location
Saccharomyces cerevisiae	YPR140W	YPR140W	381 aa	18 %	66 %	XVI:814391-815536
Human	ENSG00000102125	ENSP00000469981	292 aa	23 %	86 %	X:154411524-154421726

CLUSTAL W (1.81) multiple sequence alignment

YPR140W/1-381 MSFRDVLERGDEFLE-AYPRRS---PLWRFLSYSTSLLTFG-VSKLLLFTCYNVKLNGFE
ENSP0000469981/1-292 -----MPLHVKWPFPAPVPLT--WT-LASSVVMGLVGTYSCFWTKYMNHLTVHNRE
.....::..*: . * . * * : * . : . * * : * : ..:::.. *

Additional Exercise 1 - *Zymoseptoria* orthologues

Exploring an orthologue that we identified using BioMart, exercise 2. We identified six genes with a pathogenic phenotype of ‘Loss of pathogenicity’ in *Zymoseptoria tritici*. We then found a single gene low confidence gene 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 [CNC06590](#) to go to the gene page.

Search:  for



(a) Does this gene in *C. neoformans* have a PHI-Base annotation?

 **Cryptococcus neoformans var. neoformans JEC21 (ASM9104v1)**

Location: 3:1,927,422-1,929,917 Gene: CNC06590 Transcript: AAW42642 Jobs ▾

Gene-based displays

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

Gene: CNC06590

Description mitogen-activated protein kinase, putative

Location Chromosome 3: 1,927,422-1,929,917 forward strand. ASM9104v1:AE017343.1

About this gene This gene has 1 transcript (splice variant), 358 orthologues, 1 parologue and is a member of 2 Ensembl protein families.

Transcripts Hide transcript table

Show/hide columns (1 hidden)								Filter
Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags	
-	AAW42642	1986	365aa	Protein coding	P0CP68, P0CP69, Q56R42	XM_569949.1, XP_569949.1		

PHI: Phibase identifier

Accession	Term	Evidence	Annotation source	Mapped using	Transcript IDs
PHI:497	497	ND	Sequence Publications:15728721	AAW42642	<ul style="list-style-type: none"> Search BioMart View on karyotype

Ensembl Fungi release 50 - February 2021 © EMBL-EBI

(b) Is it the same as that in *Z. tritici* (loss of pathogenicity)? What host species was this linked to? Is function of this gene in pathogenicity conserved between the two species?

	Pathogen Gene	Mutant Phenotype	Pathogen Species	Disease	Host Species
+	HOG1	reduced virulence	Cryptococcus neoformans	Cryptococcosis	Mus musculus (related: house mouse)

	Pathogen Gene	Mutant Phenotype	Pathogen Species	Disease	Host Species
+	MgHog1	loss of pathogenicity	Zymoseptoria tritici	Septoria leaf blotch	Triticum (related: wheat)

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

Show All entries Show/hide columns Zymoseptoria tritici

Species	Type	Orthologue	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence
Zymoseptoria tritici	1-to-1	MGHOG1 (Mycgr3G76502)	84.03 %	82.19 %	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				

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

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	MADFVKLSIFGTVFVFTTRYVDLQPVGGMAGFLVCSAKDQLSGTSVAIKKIMKPFSTPVL MAEFVRAQIFGTTFEITSRYTDLQPVGGMAGFLVCSAKDQLTGAVAVKKIMKPFSTPVL ***:***: .****.**:***.*****:*****:*****:***:*****:*****
AAW42642/1-365 Mycgr3P76502/1-357	SKRTYRELKLKHLRHENIISLDIFISPLEDIYFVTTELLGTDLHRLLTSPLEKQFIQY SKRTYRELKLKHLKHENVISLDIFISPLEDIYFVTTELLGTDLHRLLTSPLEKQFIQY *****:*****:*****:*****:*****:*****:*****
AAW42642/1-365 Mycgr3P76502/1-357	FLYQILRGLKYVHSAGVVHDLKPSNIVLNENCDLKICDFGLARIQDPQMTGYVSTRYYR FLYQILRGLKYVHSAGVVHDLKPSNIVLNENCDLKICDFGLARIQDPQMTGYVSTRYYR *****:*****:*****:*****:*****:*****
AAW42642/1-365 Mycgr3P76502/1-357	APEIMLTWQKYDVAVDIWSTGCIFAEMLEGKPLFPKGDKDHVNQFSIITELLGTPDDVIQT APEIMLTWQKYDVVEVDIWSAGCIFAEMLEGKPLFPKGDKDHVNQFSIITDLLGTPDDVIST *****:*****:*****:*****:*****:*****:*****:*****:*****:*****
AAW42642/1-365 Mycgr3P76502/1-357	IASENTLRFVQSLPKREKVFPSTKFPNADPVSLDLERMLVFPRTRISAEGLAHEYLA ICSENTLRFVQSLPKRERQPLKNFKNAQPQAIELLERMLVFDPRKRVKAGEALADPYLS *:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****
AAW42642/1-365 Mycgr3P76502/1-357	PYHDPTDEPVAAEVFDWSFNADLPVDTWKVMMYSEILDHFNLGDISQ-NEAEGPVITGEV PYHDPTDEPEAEKFDWSDNADLPVDTWKIMMYSEILDYHNDS-ANN---GEG--- *****:*****:*****:*****:*****:*****:*****:*****:*****:*****:*****
AAW42642/1-365 Mycgr3P76502/1-357	PAAPAS- -QENGGA ..

Additional Exercise 2 - Mushroom genes

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

Search:

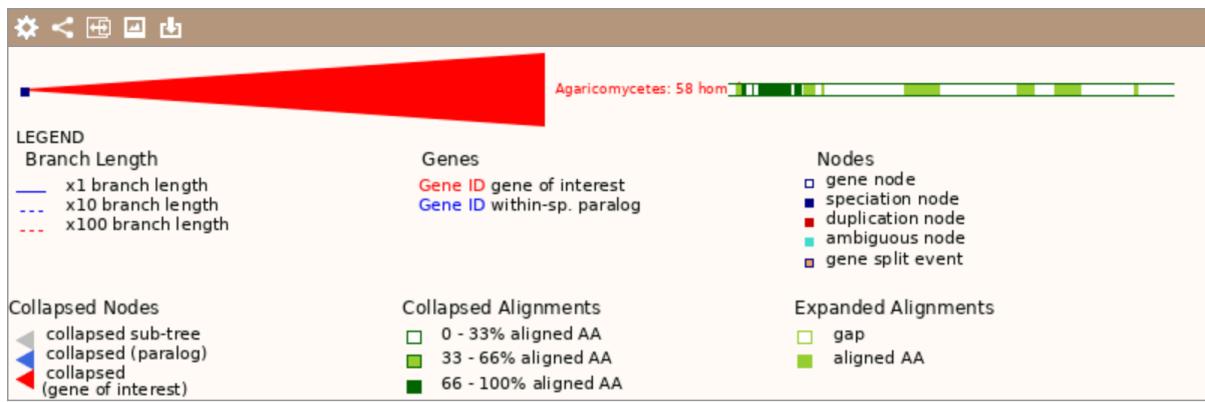
for

Go

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?



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 [Class](#)

(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 57 orthologues .
Transcripts	Hide transcript table

Show/hide columns (1 hidden)		Filter					
Name	Transcript ID	bp	Protein	Biotype	UniProt	Flags	
-	EAU90162	1389	462aa	Protein coding	A8N9X3		

(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?

Show All entries								Show/hide columns	Amanita thiersii
Species	Type	Orthologue	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence		
Amanita thiersii Skay4041 (GCA_002554575)	1-to-1 View Gene Tree	AMATHDRAFT_122148 KZ301993:102,546-102,928:-1	41.82 %	9.96 %	n/a	n/a	No		
View Sequence Alignments									
<div style="border: 1px solid black; padding: 5px; display: inline-block;"> Orthologue Alignment View Protein Alignment View cDNA Alignment </div>									

CLUSTAL W (1.81) multiple sequence alignment

EAU90162/1-462 -----MRVLLHDTQMNLEKFGSHVEALISNVKETSQELRKTSSTFEQEHDKLLG
PFH51030/1-110 PLTPLDKNATSMRVLHDTQANFEKFSTRVDNFNLNGLAETKSEINLVKSLFERGQETLTN
***** *:**** :*: ::..: **.*:.. .** ***: ::.* .

EAU90162/1-462 DIIIDLV-NRSQKQLQSSIGSPAQSALDAFMNKVELRLESLDQRDLAMQAFNQTHSQALQ
PFH51030/1-110 DIIDL-VNRCSQIQLKTLGSPAQSAGMEQLSKDINQRLDCLDKRLDAIQTV-----
***** **.*:.*:*****::: : ..: ***.***:***:***.

EAU90162/1-462 TQIQAIQNLQAQQNLILNAVTPLLPLQLSPPLPRLAPSTS LANPSQTQRTDASSQTIEKR
PFH51030/1-110 -----

EAU90162/1-462 QPSYHQETLRKRQRVDSDIQEISPPLPLPGSAQKKRRIESPRSVQKPSLELTQRLFPSSS
PFH51030/1-110 -----

EAU90162/1-462 PDLIKYSTDSEGPKVNRERSA PLVTPRRPLQDLFPFPGSNQRSVSKRPMPPSSTRL
PFH51030/1-110 -----

EAU90162/1-462 VGPGKSATPGPSRVGAESRAALARRPLIKPLAIAPLAFSSTS KTPVHISNFTP KPVTPS
PFH51030/1-110 -----

EAU90162/1-462 LRNAVAGEGRALKIAQTPQVLKNERMTSQAAKNTTMPPAGMVSLRSSTITTATAKPT
PFH51030/1-110 -----

EAU90162/1-462 SNTPRFGPEANKPLL RAPTNNNGPRPLQERMKEPVREGRRFIPLVDTDDEDDSD
PFH51030/1-110 -----