

Exercise: Searching Ensembl Fungi species

Links to be clicked shown in blue, text to be entered shown in red.

Navigate to fungi.ensembl.org. The number of the release may vary but you'll see a homepage similar to this:

The screenshot shows the Ensembl Fungi homepage. At the top, there is a search bar with "All species" selected and a "Go" button. Below the search bar, there is a placeholder text "e.g. NAT2 or alcohol*". On the left, there is a sidebar titled "All genomes" with a dropdown menu set to "Select a species". To the right of the sidebar, there is a section titled "Favourite genomes" featuring icons and names for *Saccharomyces cerevisiae* (R64-1-1), *Schizosaccharomyces pombe* (ASM294v2), *Aspergillus nidulans* (ASM1142v1), *Puccinia graminis* (ASM14992v1), *Magnaporthe oryzae* (MG8), and *Zymoseptoria tritici* (MG2). A link "View full list of all species" is located below the sidebar. On the right side of the page, there is a "What's New in Release 58" section with a list of updates, an "Ensembl Rapid Release" section with a note about new assemblies every two weeks, and an "Archive sites" section with links to previous releases. The overall layout is clean and organized, typical of a scientific bioinformatics database.

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

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

The screenshot shows the "Find a Species" search results for the genus *Bipolaris*. The search term "Bipolaris" is highlighted in the search input field. The results table has columns for Name, Classification, Taxon ID, Assembly, Accession, Variation database, Regulation database, Whole genome alignments, Other alignments, In peptide compara, and In pan-taxonomic compara. The table lists five entries for *Bipolaris* species: *Bipolaris maydis ATCC 4831*, *Bipolaris maydis C5*, *Bipolaris oryzae ATCC 44560*, *Bipolaris sorokiniana ND90Pr*, and *Bipolaris zeicola 26-R-13*. Each entry includes a thumbnail image of the species, its assembly name (e.g., CocheC4_1, CocheC5_3, Cochliobolus_miyakeanus_v1.0, Cocsai, Cochliobolus_carbonum_v1.0), and its GCA accession number. The "Download table" button is located at the top right of the table. A callout box points to the "Bipolaris" search term in the input field. Another callout box points to the "Data types available for each assembly" section at the bottom right of the table. The footer of the page includes links to various taxonomic groups and a note about the release date.

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*?

Search for features in this genome

About this genome assembly and sequence download

Information about this species

Links to example feature pages

Comparative genomics section

Download gene annotation

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?

Statistics about the genome assembly and annotation

Assembly	Cochliobolus miyabeanus v1.0, INSDC Assembly GCA_000523455.1
Database version	111.1
Golden Path Length	31,382,097
Genebuild by	JGI
Genebuild method	Import
Data source	INSDC

Gene counts

Coding genes	12,002
Gene transcripts	12,002

Hover over underlined terms with your mouse to see a description

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

The screenshot shows the ENA homepage with the logo and navigation menu. The main content area displays the assembly details for GCA_000523455.1. It includes a summary text about the family Pleosporaceae, a comment from JGI, and a detailed table of organismal characteristics. On the right, there is a sidebar with various download and view options.

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: [AMC001](#)

Exercise: Ensembl Fungi ‘Region in detail’ view

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

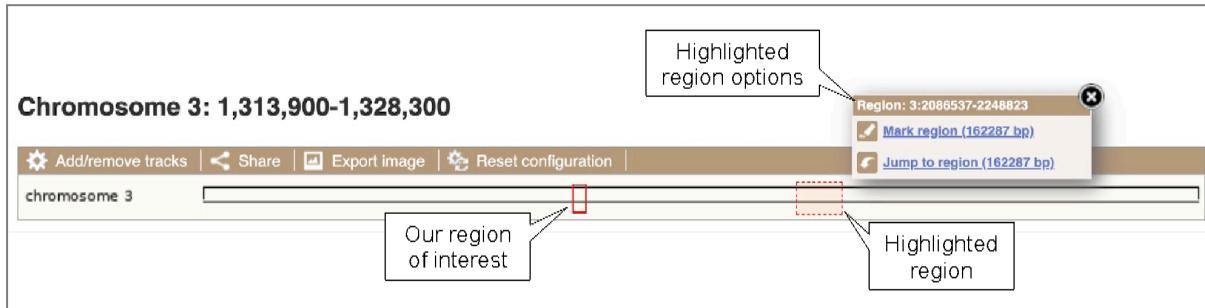
(a) Find *Zymoseptoria tritici* (assembly MG2), then type (or copy and paste) **3:1313900-1328300** into the search box. Press enter or click **Go** to jump directly to the **Region in detail** Page.

Click on the button to open a page-specific help page. These help pages provide links to Frequently Asked Questions (FAQs), a glossary, video tutorials, and a form to contact the Ensembl helpdesk. You can find 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:

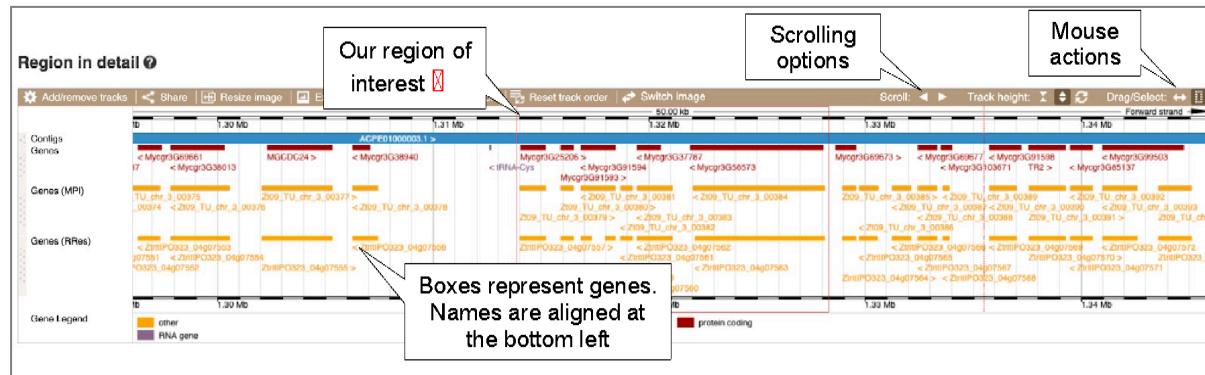
Let's look at each image in detail.

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

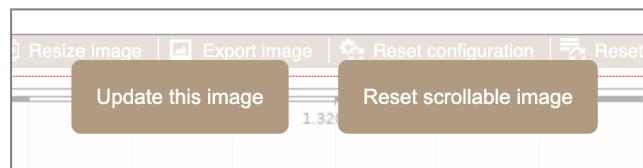


If you would like to move to the region, you could click on [Jump to region \(### bp\)](#). To highlight it, click on [Mark region \(### bp\)](#). For now, we'll close the pop-up by clicking on the [X](#) in the top right-hand corner.

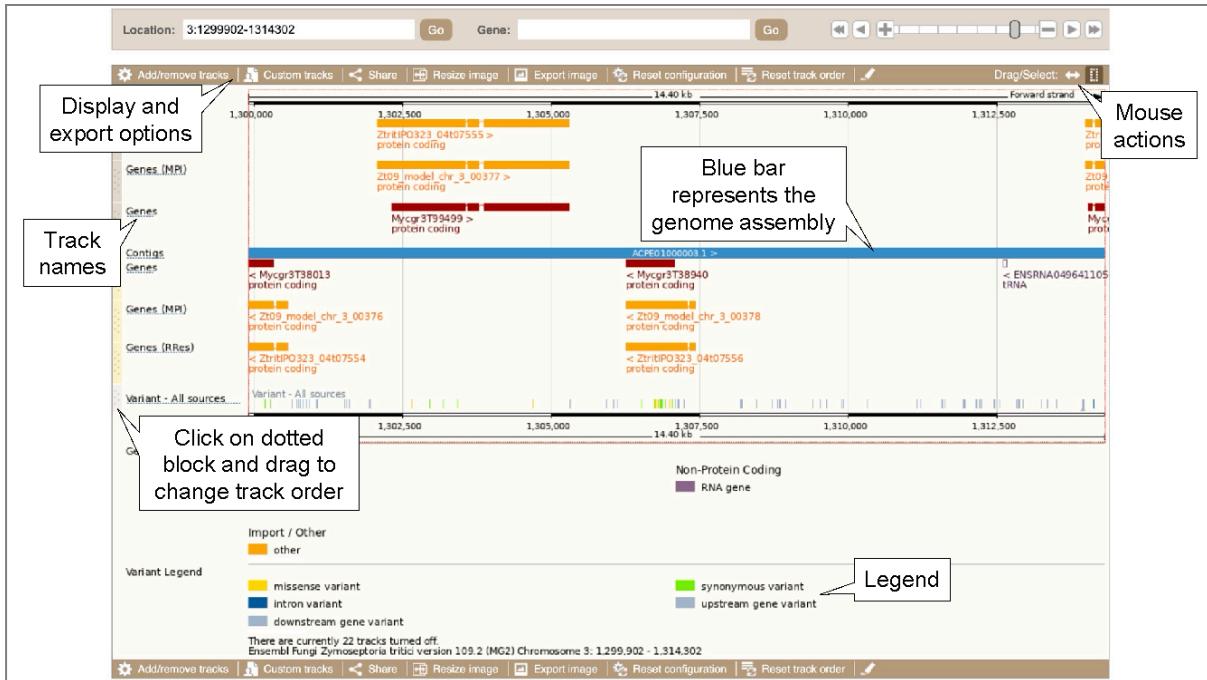
The second image is the region overview, showing a 50 kb region around our selected region. This view allows you to scroll back and forth along the chromosome.



Click on the [Drag>Select](#) button to change the action of your mouse click. Changing to the arrow allows you to 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 buttons appear. Clicking on [Update this image](#) will jump the lower image to the region you have selected above. We want to go back to where we started, so we'll click on [Reset scrollable image](#).



The third image is the region in detail view. It is a detailed, configurable view of your selected region similar to this:



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

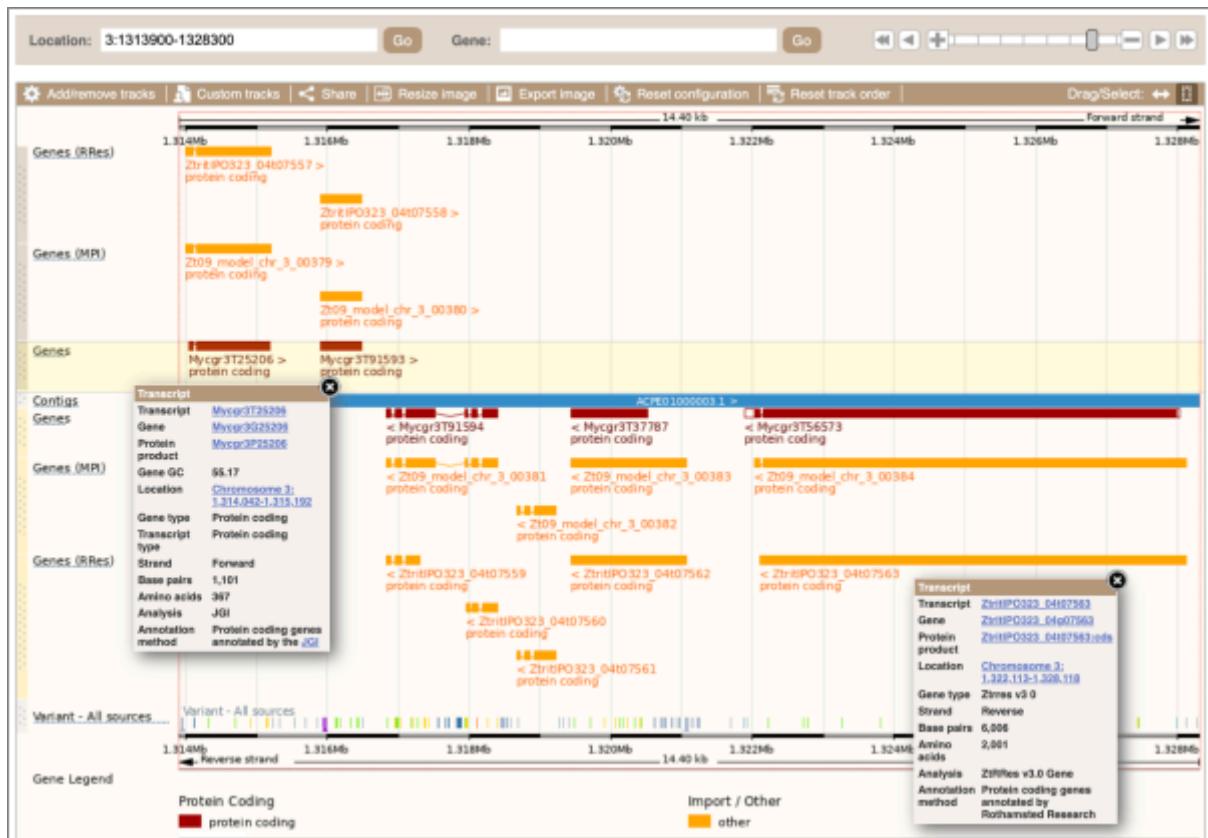
Click on the **Drag/Select** option at the top or bottom right to switch your mouse action. On **Drag**, you can click and drag sideways to move along the genome, the page will reload when you drop the mouse button. The **Select** option allows you to drag out a box to highlight or zoom in on a region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

Change your mouse action 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.

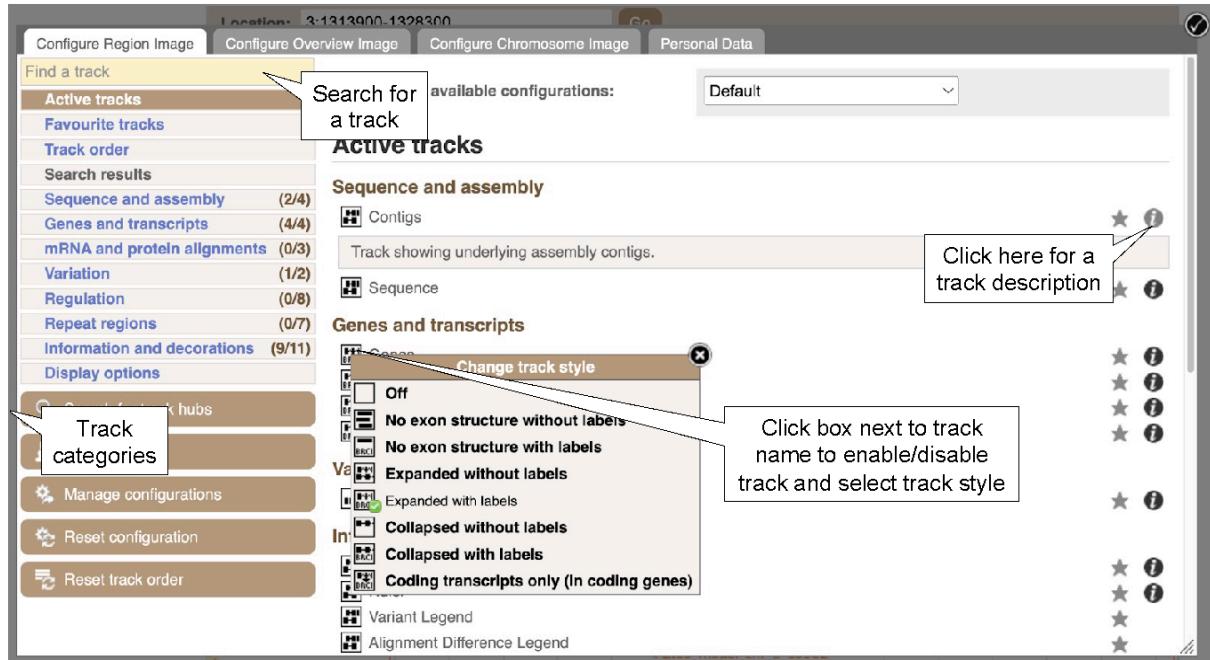
(b) How many genes are annotated in the current region? How many are on the forward and how many are on the reverse strand? Are they all annotated by the same institute?



We can edit what we see on this page by clicking on the [Configure this page](#) button located on the left-hand side.

This will open a menu that allows you to change the image. You can put some tracks on in different styles. You can read more details in this FAQ:

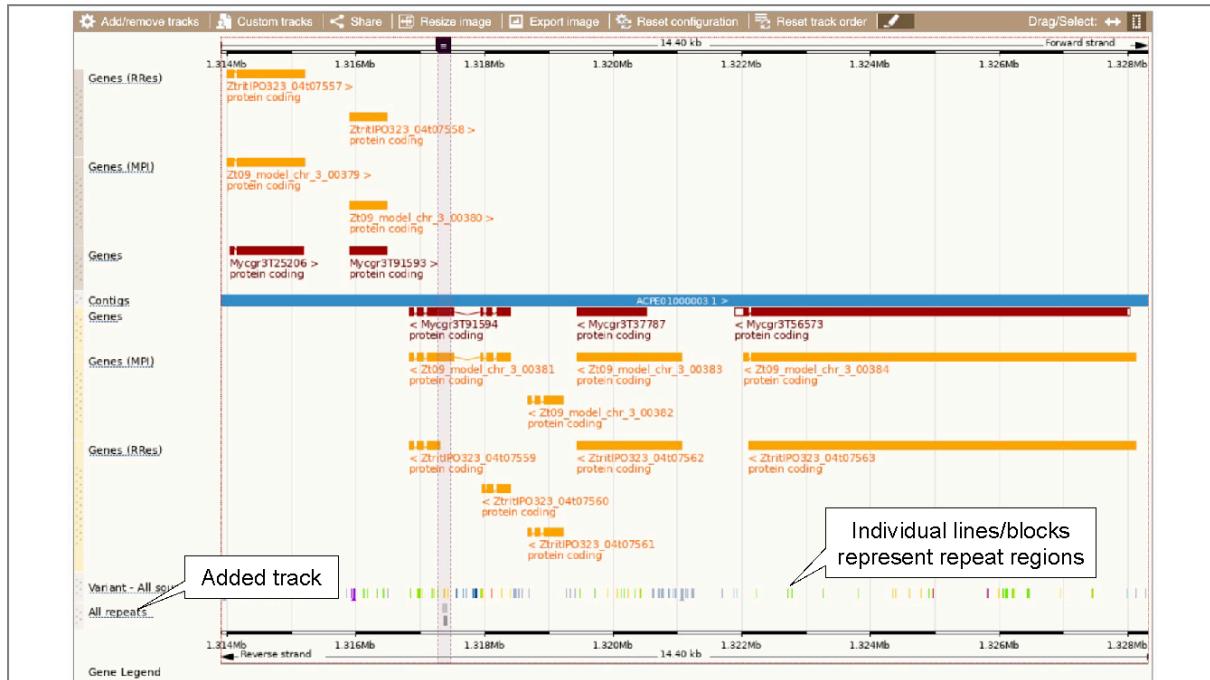
<https://www.ensembl.org/Help/Faq?id=335>.



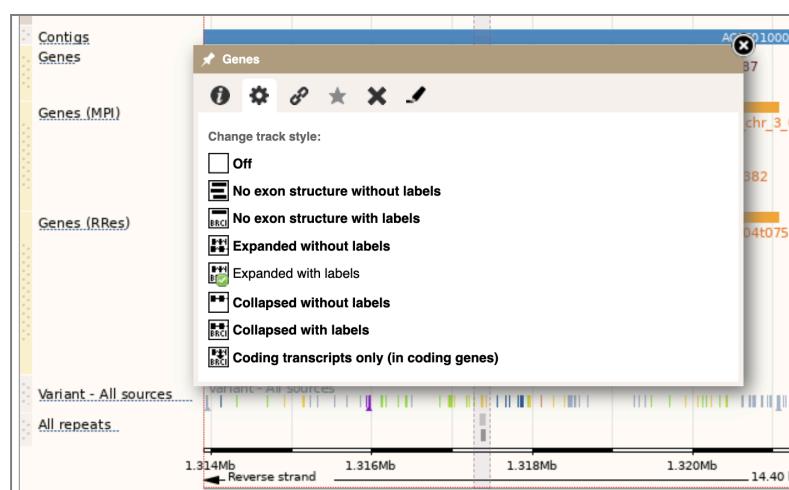
You can add a track to the image by clicking on the box to the left of the track name and selecting a track style from the pop-up menu. Click on the tick in the top right-hand corner to save and close the menu. Alternatively, click anywhere outside of the menu.

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?

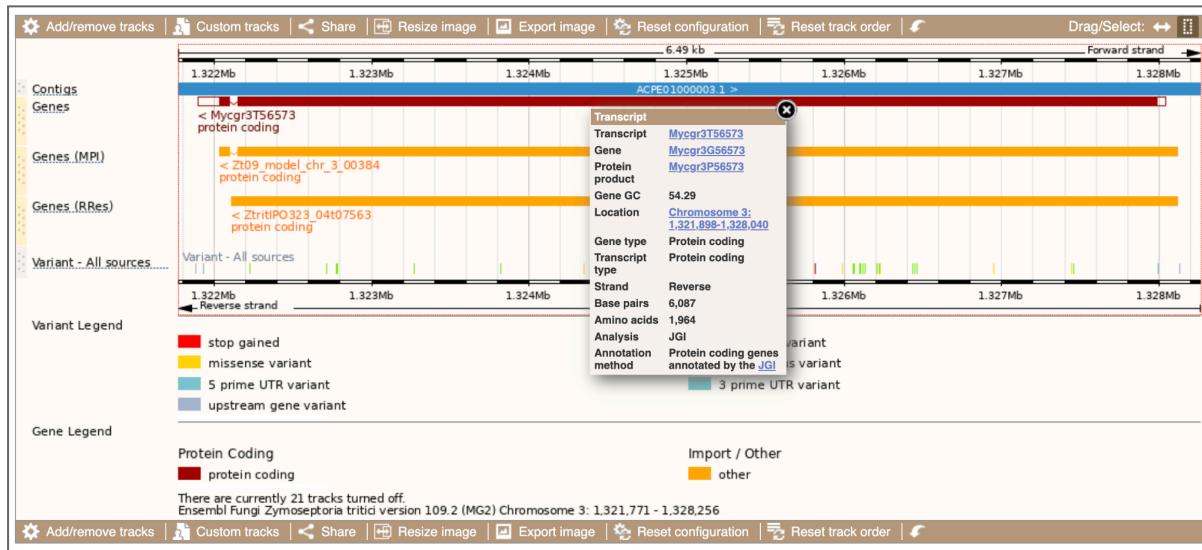


We can also change the way the tracks appear by clicking on the track name to open a menu.



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

(d) Zoom in on the largest transcript [Mycgr3T565573](#). 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 located either at the top of the image, or in the left-hand panel to generate a link.

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 by clicking the [Export data](#) button located either at the top of the image or in the left-hand panel.

To return the region image to the default view, click on [Reset configuration](#) at the top of the region in detail image.

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 the leucine biosynthetic process.

From fungi.ensembl.org, type *LEUC* into the main search box, click the drop-down menu, 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:

Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags
yfh7	SPAC227.14.1	2407	235aa	Protein coding	Q9UTCS	-	Ensembl Canonical

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 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 ?

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

Exons LEUC exons All exons in this region

Markup loaded

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

GTAGCGAATGTCCTCCAGGCTGATGTCGGAGCTGGCGAGTTTGCAGAGTAGCTCGTCCAG
GGTCAGATCTCGTCAGAGTGGAGTGGCGTGGCCTCACGGGACCTGCGGATCCAA
TATGTCGGAATCACCAGGCTTGTTGGCGAATTCTCGCCCTGAATGCAAGGGCTG
ATTCACTTCCAGCGATCTGGCTTCATCATGACTCGGCTTGACTGCAAGGTGAG
CCCCAGTCGCACTAACAGAAAGATGCACTGGGCTCTCCCATACTGACTGCACTGTTG
TTCCCACAGCGGTTGAAGCATTGCACTGCCATCCGAGACCTGCGGGAGGCGGAAGAG
CTAGATATCCATCGATGTTGAGTAAGAGATGAGGGAGTTCTGGGAGAAAAGAACGAC
TGGATGTTGAAACCTTGGCTCCCGGATGTCCGGAGGGCAGCAAAGTCGGAGAACGA
AAGTCGACCGCTCTCCAATCCACGTAACAAATCTITGCCAGAACATCAACCATACA
AACATGAATCGTCGAAAGAACATGCCCACCTTACAACATCGTCGTTGGAG GTGAGTGGCTA
TGCCCTTGGCAATTGAATCATCTGCTGATCGATTCAAG**TGCAATCACTGGGCTCTGAA** GT
AAGCAGCTCCCTTACCCACTCACTCGTACGGACTCCAAGCTAACGATTACAG**GTCACC**
GCCGAGGGCTGAAG GATGCCATACAAGTATCTGTCATAAAACAAGCAACACTGAC
CTCGACGTTGATGCAACAGCAATGCCGACGTTCACTTCACATT
AAGTGTGCGGGGCTCAACTTCACATGTCCTGAGCTGACAG**GTGAGTGGCT**
CTGAGCCACGGAGGACGAGGGCTCCGGCTACGGAGCTGGACACGGAAACCTACTCGGG
CTGAGATCGAGCGTGTGGCACGACTGGGGGTTCTGGCTGAAGACCCGGCGT

[Download or BLAST sequence](#)

Upstream sequence (600bp)

First exon in LEUC gene

Highlight region to BLAST specific sequence region

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
The forward strand denoted by 1, and the reverse strand by -1.

Exons are highlighted within the genomic sequence. If you click on [Configure this page](#) in the left-hand panel, you can change display options. 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 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	GTCCGGTCTGGTCGCTCGACAAGGCAGAACGTATGGCTACAAGCAGACTGTGGCGCAAGA					1581991
1581990	CCGTGACGGATGTGTTGCCAACGAGTTCCCGCAGTTGAAGATCGGGCATCACCTATTG					1581931
1581930	ATTCCGGCGATGTGATGGTCAGAAACCCGAGAGCGTTGAACGGAGTCATCGTCACTA					1581871
1581870	GCAATTGTTCGGCATATCATTAGTGACGAGGCAGCGTCACTCCAGGATCKTGGGTT					1581811
1581810	TGTTGCCCAAGTGCAGTTGACTGCCCTGCCGGATGCCAAGAGCAAGTGCATGGCATT					1581751
1581750	ACGAGCCGATTCACGGTATGTACAATGTTGCTTGATGCGAGGTGCAATTGTTGCTGACAA					1581691
1581690	CCATCAGGCTCTGCACCAGACATCAGCGCAAGGGTATCGTCAACCCCGTCCCATGATC					1581631
1581630	TTGTCTTGGGATGATGTGCAAGTACTCTCTCCAGCAGCCGAGCTGGCCAAGAAGATC					1581571
1581570	GACGAGGCCGTGAGGAACGTTGATTGAGAAGGGTATCAACACTGCCGATATTGGTGGTTCG					1581511
1581510	GCCAAGACTGCCGAGGTGGGTGAYGCTATTGCGAAGGAGCTCGAGGCCCTGCTGAAGTAA					1581451
1581450	ACGGTGCAACTKGGTACTGGTACACTCGCGTTCGCGAGTGTGGACAAGAGGATGGC					1581391
1581390	GCAAGGTCAAACAAAAGTTGAGATGAAAGCTTCTGGGTTGGAACATAAYTCTCTAT					1581331
1581330	AGAGGAACCTGATGCAACGCAAATGGAACAAACCCCTGGGTCGAATGATCACGAAGGTAC					1581271
1581270	CATGACACTAGAGATGCCCTGCCGGTGCCGGTGCCTGCTTGTATGCTATCACAGAGA					1581211
1581210	CAGTTCAGGACAGGTCCTCGCTCGTTGTCRTCCTACTGCCACAGTGGCGCAACGGTAT					1581151
1581150	CCACCGCCCTCTCCACCACTGCTCTCACCGATCTCCAACGTCGGTTCCCACACTCCA					1581091
1581090	CCTCCAACCTCTCCAACACACCTTCGCCACTGCCRTTCACCTTGTATCTCAGCCGG					1581031
1581030	GCACTGGAATGCCAGCGCATGCCCTCGTCAACTGCCGCCCTCAGCCTTCTCCCTCT					1580971
1580970	CCACCTCGTCGACTTCTCCTCTCCGTCTGCACCAATTCTCASAAGCAGCATCTCG					1580911
1580910	GCCTCGGATTGGACCCACACAATCCTCCCCAAACTGCGCTTCAAGAACATACACAACC					1580851
1580850	TCCCCAACCTTCCTCAGGACTCACATTGCGATGCCATTGCCCTCTCCCTCCAT					1580791
1580790	CCCCATTGACACCATTCCCTCGACTGGTCATCTYCAAAGTGATCGTGCATGTTGTC					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					

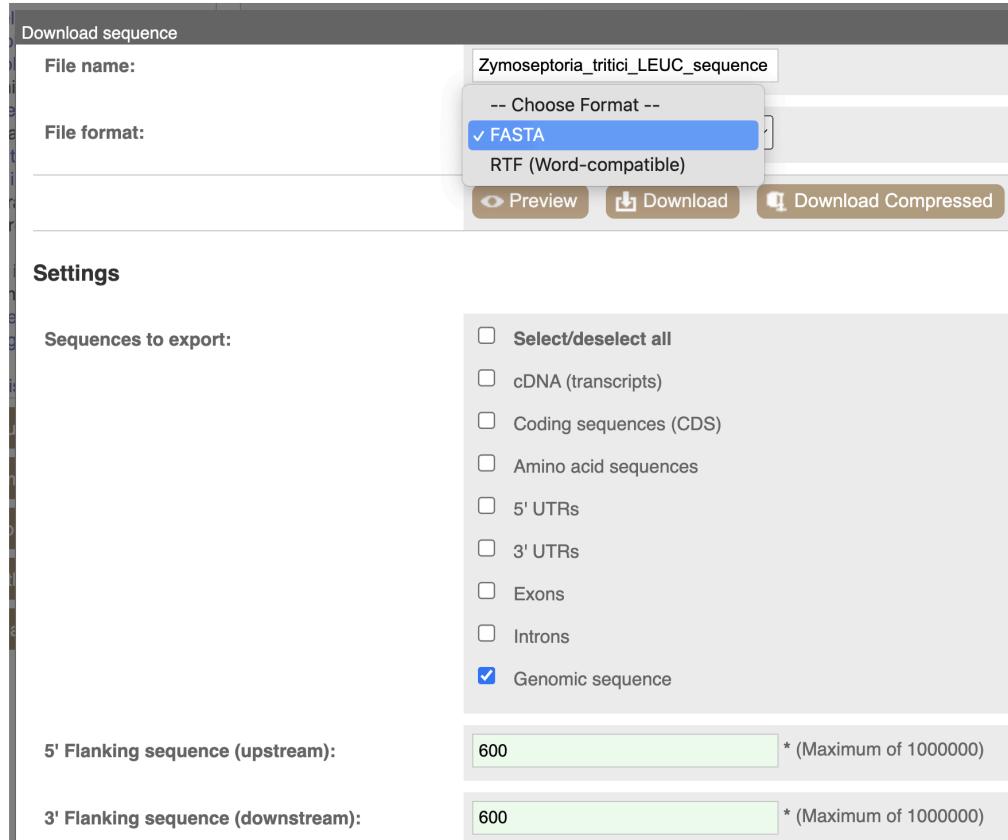
1582590	AAAGCATACAGGTGCTCGACGTGATCGACAACAGCAATGCCGACGTTCACTTCAACATTC					1582531
1582530	AACCAACACCTGCTCGTGGC GTACGTTGCTCCGACCACTTCCGAAAAGCCAGGGCT					1582471
1582470	AACCGAACGAGGCCTGMTCGATGCCAACGGCGAGCCTCTAACAGACGAAGCACTGCCG					1582411
1582410	CTGCCAAAS CAGCCGACCCGTATCCTCGGAGCCATGGCGTCCCAAATGGGGCACGG					1582351
1582350	GCAAAGTCCGTCCAGAGCAAGGCATCTGCCCTCCGCAAAGAAATGGGCACATATGGAA					1582291
1582290	ACCTGCGCCCATGCTTCTCGCGTCTGAGAGCCTCGTCAAGACTCTCCATTGAAGGAGG					1582231

Which exon does the stop gained mutation fall in?

You can download this sequence by clicking on the [Download sequence](#) button

 Download sequence

This will open a pop-up menu that allows you to pick between plain **FASTA sequence**, or **sequence in rich text format (RTF)**, which includes all the coloured annotations and can be opened in a word processor. This button is available for all sequence views.



(c) Export this sequence in RTF.

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. You can read more about GO terms here:

<https://geneontology.org/docs/ontology-documentation/>.

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

(NB - a [known bug](#) is causing GO links to break, if you encounter this, open [this link to the e!112 archive](#) in a new tab)

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 the 'leucine biosynthetic process' GO accession number [GO:0009098](#), 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 <i>Saccharomyces_cerevisiae</i> 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						

Let's explore the **Transcript** tab now. Many genes have multiple transcripts which can be seen in the transcript table. Click on the **Show transcript table** button

[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

Show/hide columns (1 hidden)				Filter			
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 summary diagram.

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

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 |

< Mycgr3T103221 protein coding

Reverse strand 1.50 kb

Statistics: 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 sequences, 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 in the navigation column on the left 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

<pre> 1 AACTGAATCGTCGAAAATGCCCACTTACAACATCGTGTCTTGAGGTGATCACTGCGATGCCCACTTACAACATCGTGTCTTGAGGTGATCACTGCG-M--P--T--Y--N--I--V--F--G--G--D--H--C-- </pre>	60 43 14
<pre> 61 GTCCTGAA GTCACCGCCGAGGGCGCTGAAGGTGCTCGACGTGATCGACAAACAGCAATGCCG 44 GTCCTGAA GTCACCGCCGAGGGCGCTGAAGGTGCTCGACGTGATCGACAAACAGCAATGCCG 15 G--P--E--V--T--A--E--A--L--K--V--L--D--V--I--D--N--S--N--A-- </pre>	120 103 34
<pre> 121 ACGTTCACTTCAACATTCAACCACACCTGCTCGGTGGC GCCT GA TCGATGCCACGGCG 104 ACGTTCACTTCAACATTCAACCACACCTGCTCGGTGGC CTGATGCCACGGCG 35 D--V--H--F--N--I--Q--P--H--L--L--G--G--A--S--I--D--A--H--G-- </pre>	180 163 54

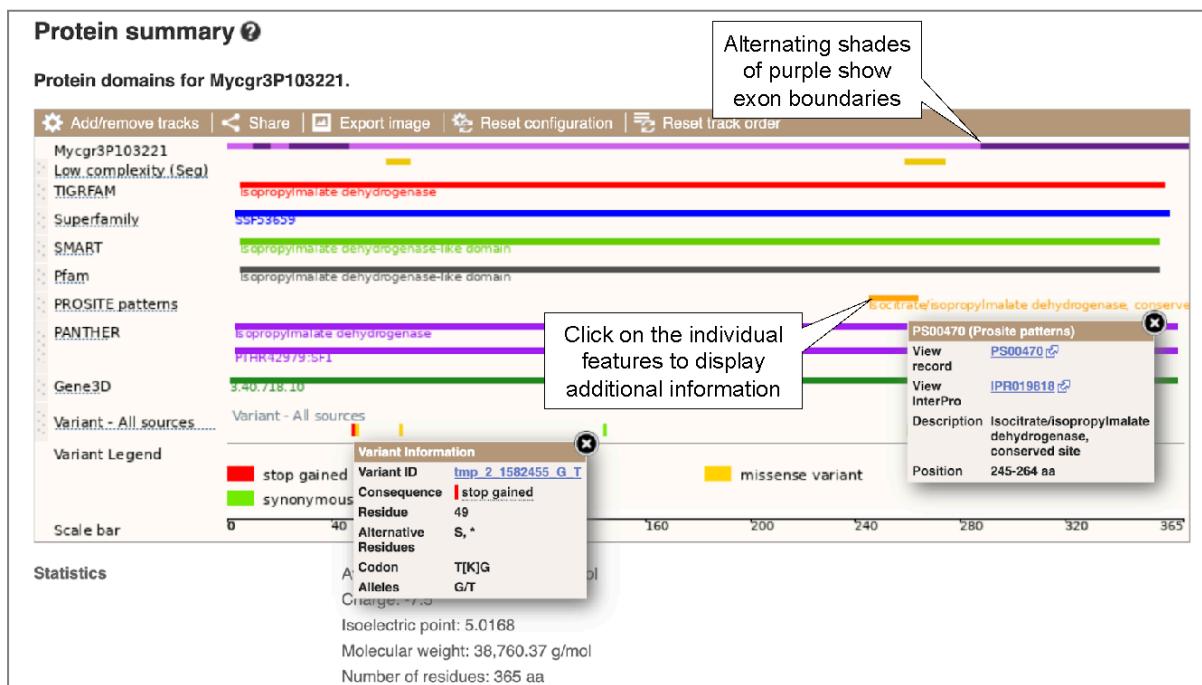
UnTranslated Regions (UTRs) are highlighted in dark yellow, codons are indicated by alternating light yellow highlights, and exon sequences are shown in alternating black and blue letters.

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 exercise (b) here?

Will this variant cause the deletion of an entire protein domain? If so, which one(s)?



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

Next, follow the **General identifiers** link in the navigation column on 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]