

## Exercise: Searching Ensembl Fungi species

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

Navigate to [fungi.ensembl.org](http://fungi.ensembl.org). You'll see a homepage similar to this:

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

**Anatomy of the Ensembl Fungi homepage**

The screenshot shows the Ensembl Fungi homepage with various sections labeled for reference:

- 1: Website header**: Points to the EnsemblFungi logo and navigation links (HMMER, BLAST, BioMart, Tools, Downloads, Help & Docs, Blog).
- 2: Main search box**: Points to the main search bar with placeholder text "All species" and a "Go" button.
- 3: Genome and species directory**: Points to the "All genomes" section, which includes a dropdown menu "Select a species" and a list of favorite genomes: *Saccharomyces cerevisiae* (R64-1-1), *Schizosaccharomyces pombe* (ASM294v2), *Aspergillus nidulans* (ASM1142v1), *Puccinia graminis* (ASM14992v1), *Magnaporthe oryzae* (MG8), and *Zymoseptoria tritici*.
- 1a: Ensembl login**: Points to the "Log in/Register" link in the top right corner.
- 1b: Quick search**: Points to the search bar at the top right.
- 4: Release highlights**: Points to the "What's New in Release 56" section, which highlights updated data and protein features.
- Ensembl Rapid Release**: Points to the "Ensembl Rapid Release" section, which provides gene and protein annotation every two weeks.
- 5: Archive sites**: Points to the "Archive sites" section, which lists previous versions of the data available at [archive.ensembl.org/fungi](http://archive.ensembl.org/fungi).

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

Name	Classification	Taxon ID	Assembly	Accession	Variation database	Regulation database	Whole genome alignments	Other alignments	In peptide compara	In pan-taxonomic compara
<i>Bipolaris maydis</i> ATCC 48331	Pleosporales	665024	CochreC4_1	GCA_000354255.1	-	-	-	✓	-	-
<i>Bipolaris maydis</i> C5	Pleosporales	701081	CochreC5_3	GCA_000338975.1	-	-	-	✓	✓	-
<i>Bipolaris oryzae</i> ATCC 44560	Pleosporales	930090	Cochliobolus_miyabeanus_v1.0	GCA_000523455.1	-	-	-	✓	✓	-
<i>Bipolaris sorokiniana</i> ND90Pr	Pleosporales	665012	Cocca1	GCA_000338995.1	-	-	-	✓	-	-
<i>Bipolaris zeicola</i> 26-R-13	Pleosporales	930089	Cochliobolus_carbonum_v1.0	GCA_000523455.1	-	-	-	✓	-	-

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*?

**About *Bipolaris oryzae* ATCC 44560 (GCA\_000523455)**

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

[Data](#) and [issues](#) from [Wikidata](#), the free encyclopedia

Taxonomy ID: 500000-9

Data source: JGI

[More information and statistics](#)

**Genome assembly: Cochliobolus\_miyabeanus\_v1.0**

[More information and statistics](#)  
[Download DNA sequence \(FASTA\)](#)  
[Display your data in Ensembl Fungi](#)

**Comparative genomics**

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.  
[More about comparative analysis](#)  
[Phylogenetic overview of gene families](#)  
[Download alignments \(EMF\)](#)

**Gene annotation**

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.  
[More about this genome!](#)  
[Download genes, cDNAs, ncRNA, proteins - FASTA](#) - GRP3  
[Update your old Ensembl IDs](#)

**Variation**

This species currently has no variation database. However you can process your own variants using the Variant Effect Predictor:  
[Variant Effect Predictor](#)

**Download gene annotation**

**Links to example feature pages**

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

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

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

**Bipolaris oryzae ATCC 44560 (Cochliobolus\_miyabeanus\_v1.0) ▾**

**Bipolaris oryzae ATCC 44560 Assembly and Gene Annotation**

**About Bipolaris oryzae ATCC 44560 (GCA\_000523455)**

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

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

**Assembly**

The assembly presented is the Cochliobolus miyabeanus v1.0, INSDC Assembly GCA\_000523455.1. (ISDC-IP with

**Gene annotation provider**

**Annotation**

The annotation presented is derived from annotation submitted to ISDC-IP with the assembly accession GCA\_000523455.1, with additional non-coding genes derived from Bamm-IP. For more details, please visit [INSDC annotation import](#).

**More information**

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

Ensembl Fungi release 56 - Feb 2023 © EMBL-EBI

**Statistics**

**Summary**

**Assembly** Cochliobolus miyabeanus v1.0, INSDC Assembly GCA\_000523455.1

**Database version** 109.1

**Golden Path Length** 31,362,097

**Genebuild by** JGI

**Genebuild method** Import

**Data source** JGI

**Gene counts**

**Coding genes** 12,002

**Gene transcripts** 12,002

Hover over underlined terms to see description

**ENA**  
European Nucleotide Archive

Enter text search terms  Search

Examples: histone, BN000065

GCA\_000523455.1

Examples: Taxon:9606, BN000065, PRJEB402

**Assembly: GCA\_000523455.1**

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

**Comment**

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

Show Less

**Organism:** Bipolaris oryzae ATCC 44560

**Accession:** GCA\_000523455

**Assembly Level:** scaffold

**Strain:** ATCC 44560

**Genome Representation:** full

**View:** XML

**Download:** XML

Sequence Report

WGS SET EMBL

WGS SET FASTA

**Navigation:** Show

**Additional Attributes:** Show

**BlobToolKit:** Hide

**Assembly Statistics:** Show

**WGS Sequence Set:** AMCO01

## Exercise: Ensembl Fungi ‘Region in detail’ view

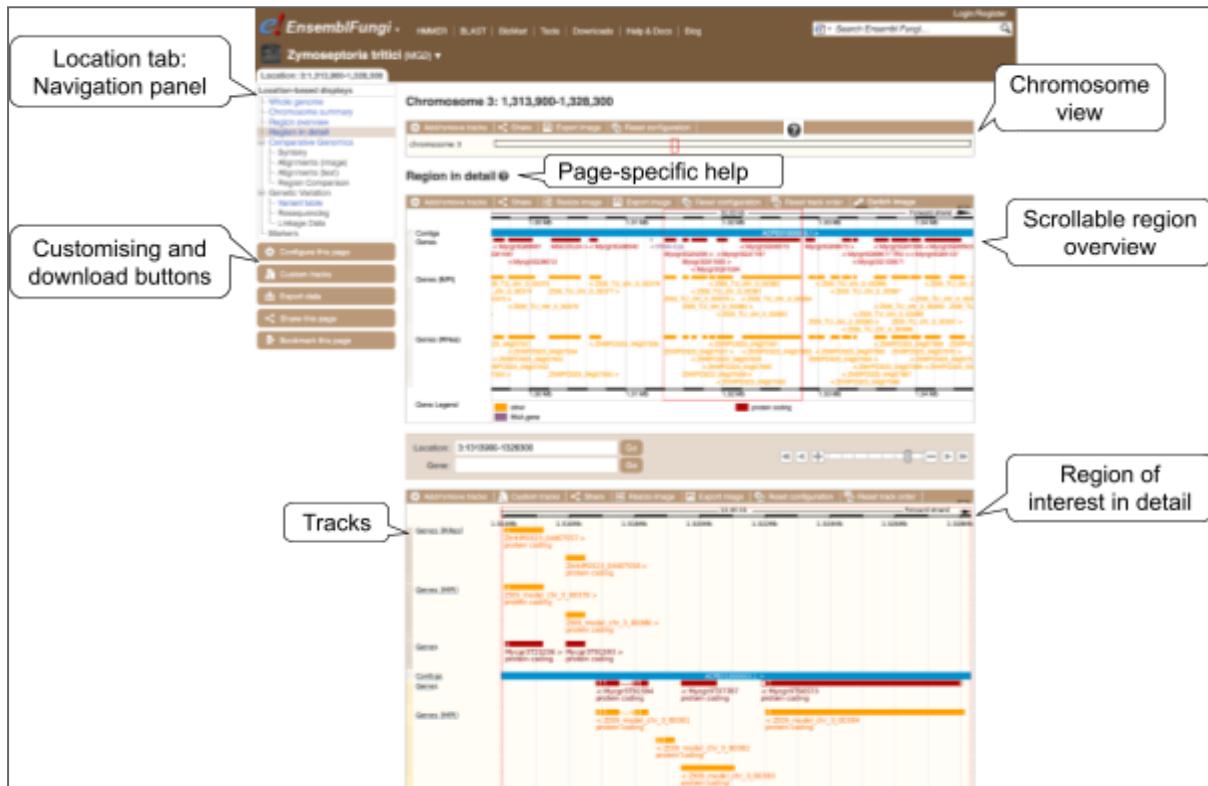
Start at the Ensembl Fungi front page, [fungi.ensembl.org](http://fungi.ensembl.org). You can search for a region by typing it into a search box, but you have to specify the species.

- (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.

The screenshot shows the Ensembl Fungi search interface. A search bar at the top contains "Search: Zymoseptoria tritici" and a dropdown menu "for". Below the search bar is a text input field containing "3:1313900-1328300" and a "Go" button. Below the input field is the text "e.g. NAT2 or alcohol\*".

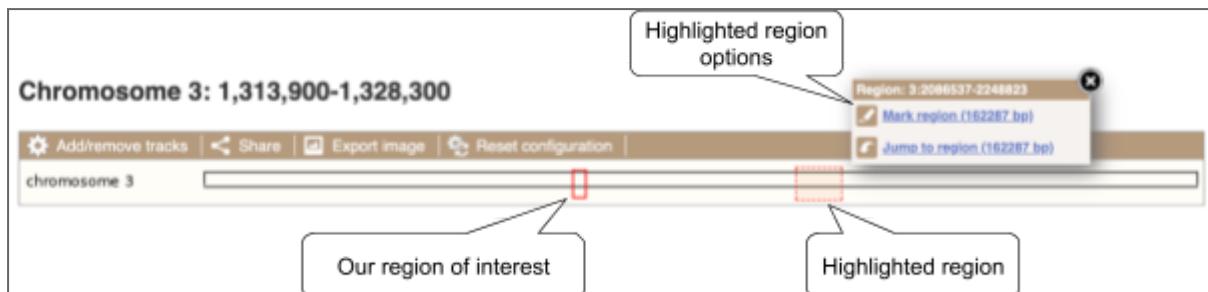
Click on the button to view page-specific help. The help pages provide links to Frequently Asked Questions, a Glossary, Video Tutorials, and a form to Contact Helpdesk. There is a help video on this page at <http://youtu.be/tTKEvgPUq94>.

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



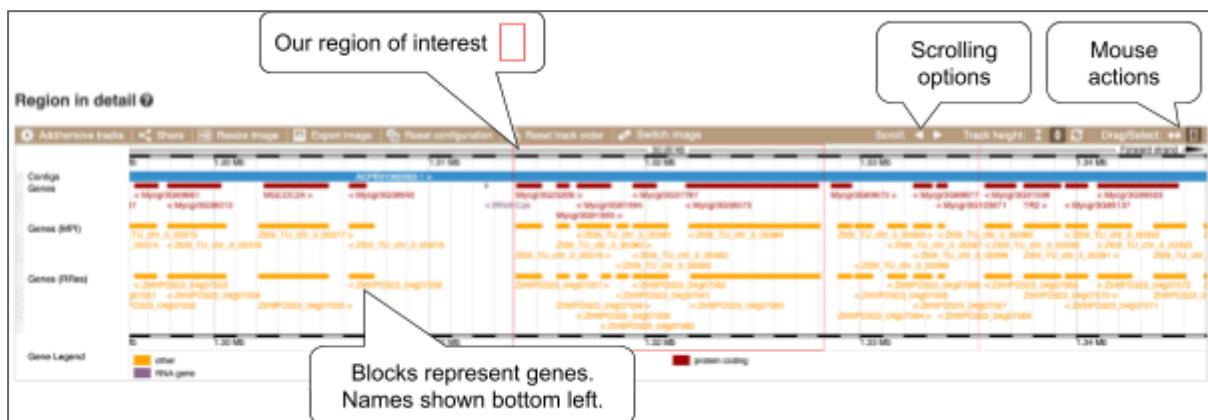
Let's look at each image in detail.

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

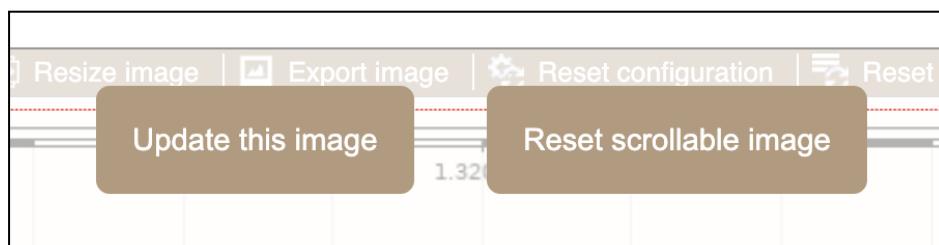


If you would like to move to the region, you could click on [Jump to region \(### bp\)](#). To highlight it, click on [Mark region \(### bp\)](#). For now, we'll close the pop-up by clicking on the [X](#) on the corner.

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



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



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



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

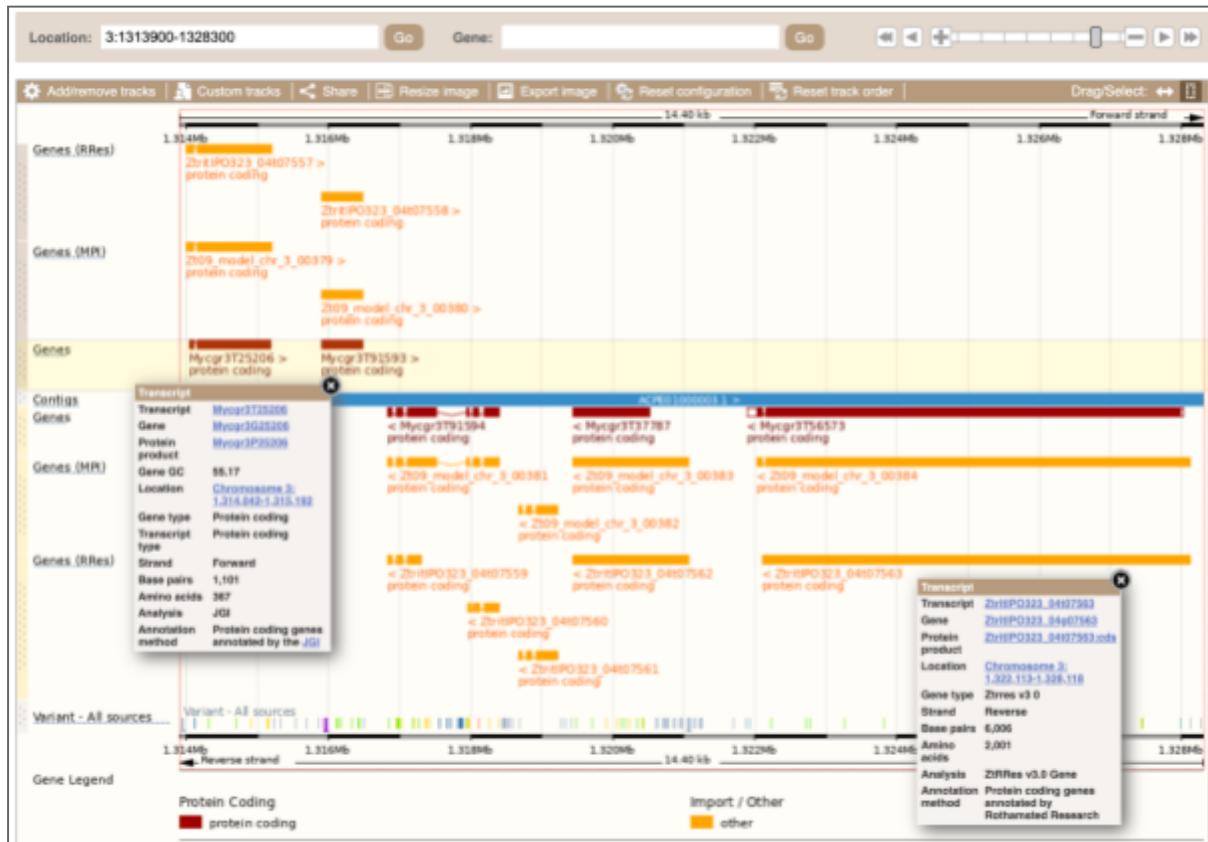
Click on the **Drag>Select** option at the top or bottom right to switch mouse action. On **Drag**, you can click and drag left or right to move along the genome, the page will reload when you drop the mouse button. On **Select** you can drag out a box to highlight or zoom in on a region of interest.

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



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

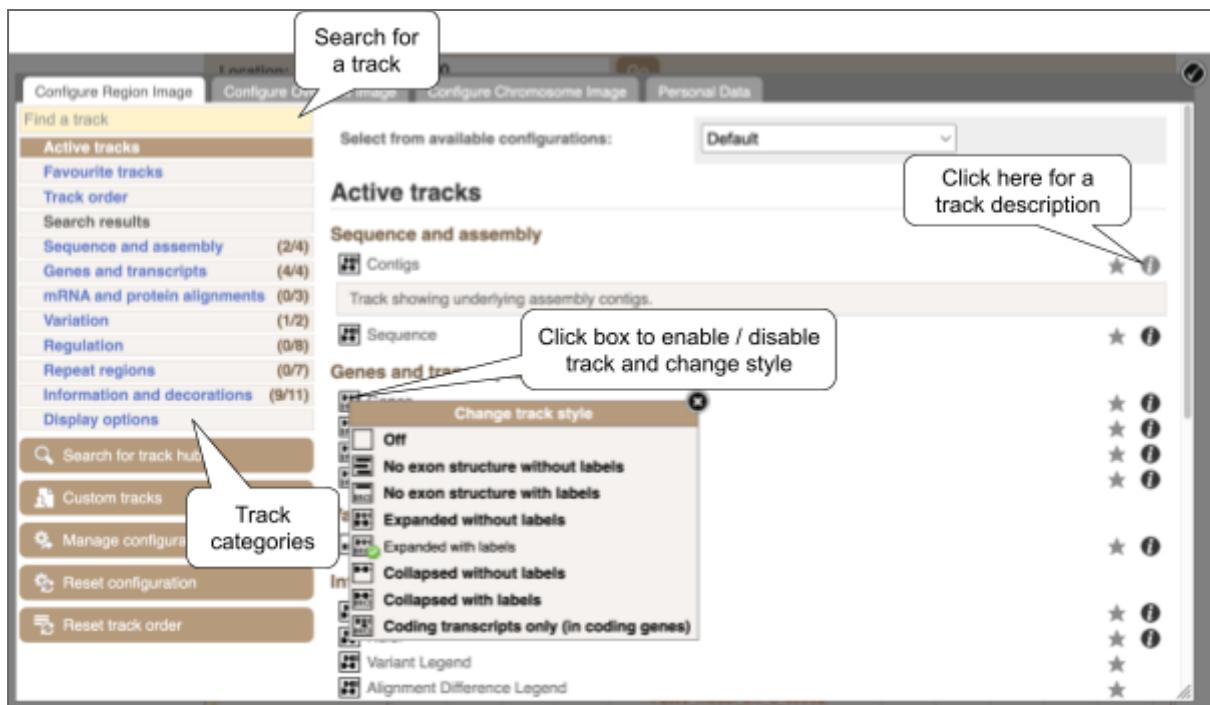
(b) How many genes are annotated in the current region? How many on the forward and how many on the reverse strand? Are they all annotated by the same institute?



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

 [Configure this page](#)

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



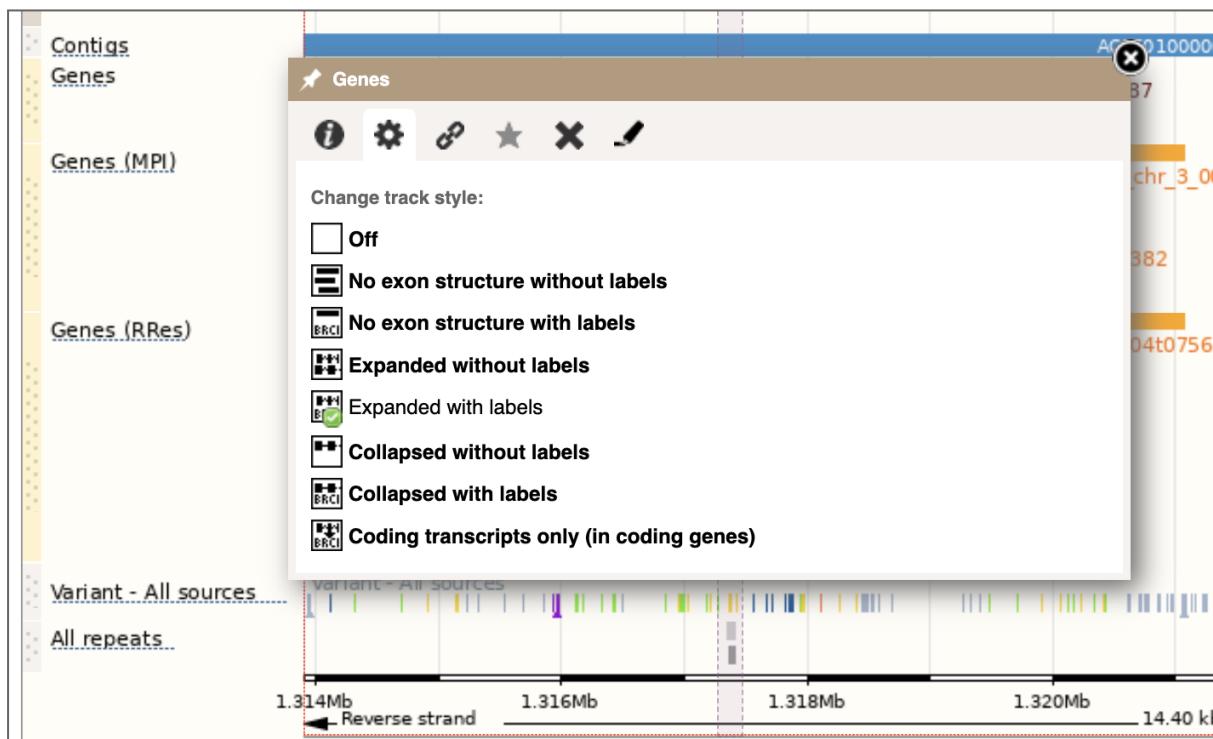
You can add a track to the image by clicking on the box and choosing track style. Click on the tick in the top right hand tick to save and close the menu. Alternatively, click anywhere outside of the menu.

Let's add some tracks to this image.

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



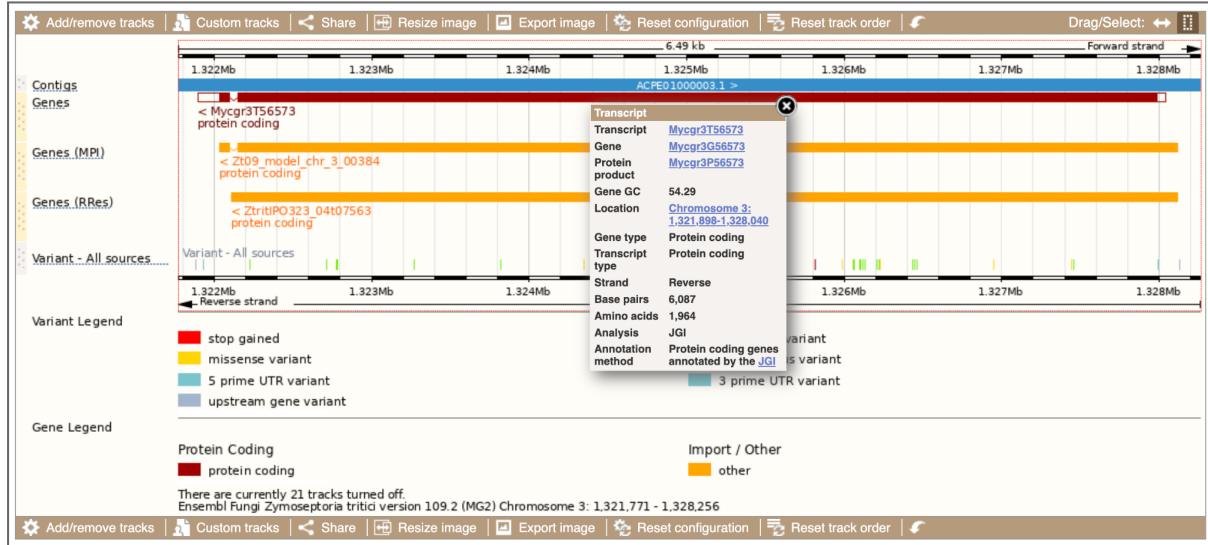
We can also change the way the tracks appear by clicking on the track name to open a menu.



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

E.g.

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



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

[Share this page](#)

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

(d) Export the genomic sequence for this region

[Export data](#)

Export data

Logout 

### Export Configuration - Feature List

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

Output: FASTA sequence \*

Select location: 3 \* 1321771 \* 1328256 \* 1

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

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

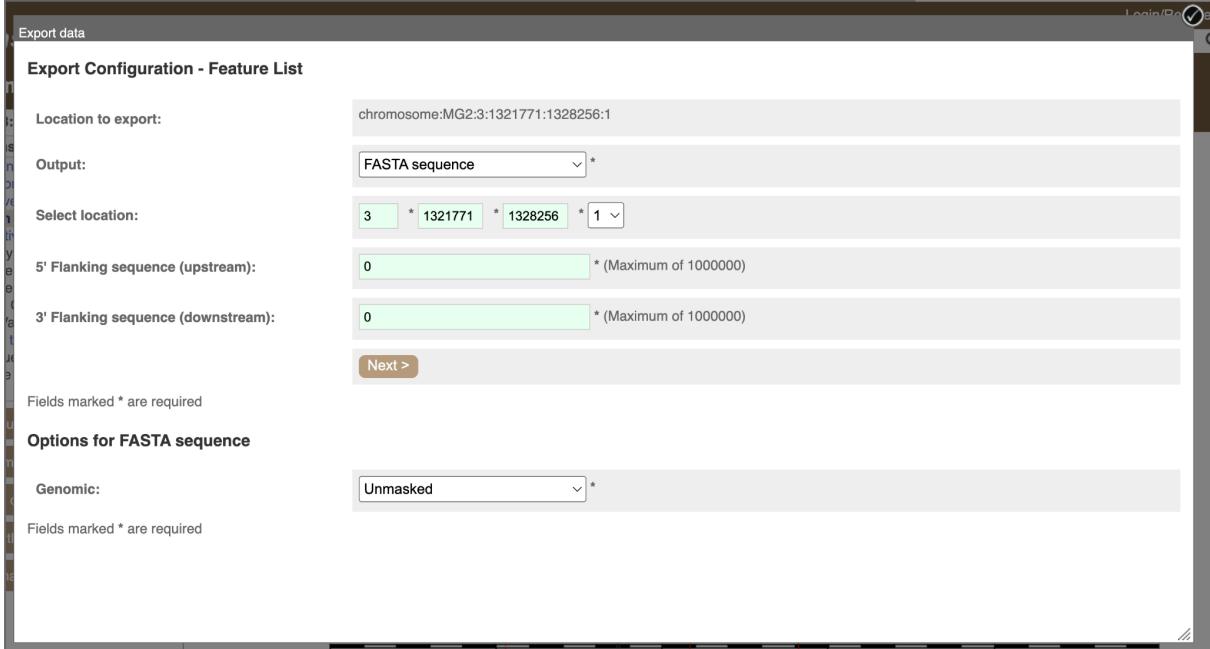
**Next >**

Fields marked \* are required

#### Options for FASTA sequence

Genomic: Unmasked \*

Fields marked \* are required



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

## Exercise: Ensembl Fungi gene and transcript tabs

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

From [fungi.ensembl.org](https://fungi.ensembl.org), type *LEUC* into the main search box, click the drop-down menu and select *Z. tritici* and click the Go button.

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

1: Selected genome

2: Location and feature tabs

3: Tab-specific pages

4: Customisation and download options

5a: Feature overview

5b: Transcript table

6: Main display

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

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

### Gene: LEUC Mycgr3G103221

#### Description

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

#### Location

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

MG2:ACPE01000002.1

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

## Marked-up sequence ②

[Download sequence](#)[BLAST this sequence](#)[Download / BLAST sequence](#)[Exons](#) [LEUC exons](#) [All exons in this region](#)

Markup loaded

```
>chromosome:MG2:2:1580729:1583430:-1
GTAGCGAATGCTTCCAGGCTGATGTCGGAGCTGGCAGTTGCGAAGTAGCTCGTCCAG
GGTCAAGATCTCGTCAGAGTGGAGTGCCTGCGCTTACGGCGACTCGGGATCCAA
TATGTCGCAATCCACCGGGCTTGTGGCGAATTCTCGCGCTGAATGCAAGCGCTG
ATTCAGTTTCGAGGCGATCTGGTCTCAATCGATTGACTCGGTTGAGTGTCAAGGTGAG
CGCCAGTGCTGCCATTGCTGGAGGCTGCACTTCTGCTGATGGTCGCC
CGCGAAGTCCAGTAAAGGAAGATGCACTGGCGATCCCAGACCTGCG
TTCCCACGAAGCGGTTGAAAGCATTCGCACTGCCATCCCAGACCTGCG
CTAGATATCCATGCGATGTTGAGTAAGAGATGAGGGAGTTCTGGGA
GTGGATGGTGGAACCTTGCTCCCCGGATGTGTCGGAGGGCAGCGAAG
ATCTCGACCGCTCTCCAATCCACGTAACAAATTCTTGGCAGAACATGACCATACAAAC
AACTGAATCGTCGAAAAATGCCACTTACAACATCGTCGTTTGGAGTGAGTGCCTA
TGCCCCCTGCAATTGAAATCATGCTGATGTCGATTCAAGGTGATCACTGCCGTCTGAAGT
AAGCAGCTCTTACCGTACTCACTCGTAGCAGCTCAAGCTAACGATTACAGGTCACC
GCCGAGGCCTGAAGGTATGCCATACAAGTAC
AAAGCATAACAGGTGCTCGACGTGATCGACAAC
AACCACACCTGCTCGGTGGCTACGTTTGTCC
AACCAGAGCAGGCCTCGATCGATGCCACGGCGAGCC
CTGCCAACAGCAGCCGACGCCGTGATCCCTGGAGCCAT
GCAAAGTCCGTCCAGAGCAAGGCATCCGTGCCCTCG
ACCTGCGCCATGCTTCTCGCGCTGAGAGCTCGTCAAGACTT
AAAGTCTGCCCGGGCGTCAACTCAACATCGTCGTGAGCTGACAGC
GTGAGCGCACCGAGGACGACGGCTCCGGTACCGAGTGGACACGGAAACCTACTCGCGCG
CTGAGATCGAGCGTGTGGCACGACTGGCGGTTCTGGCTTGGCTGAAGACCCGCCGT
```

[Upstream sequence \(600bp\)](#)[First exon of LEUC gene](#)[Highlight region to BLAST specific sequence](#)[BLAST selected sequence](#)

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

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

The FASTA header follows this format:

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

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

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

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

	Exons	LEUC exons	All exons in this region			
Variants	3 prime UTR	Intronic	Missense	Stop gained	Synonymous	Upstream
Markup	loaded					
1582050	<b>GTCCGGTCTGGTCGCTCGACAAGGCGAACGT</b> CATGGCTACAAGCAGACTGTGGCGCAAGA					1581991
1581990	<b>CCGTGACGGATGTGTTGCCAACGAGTTCCCGCAGTTGAAGATCGGGCATCACCTCATTG</b>					1581931
1581930	<b>ATTCCCGGGCGATGTTGATGGTC</b> AAGAACCCGAGAGCGTTGAACCGAGTCATCGTCACTA					1581871
1581870	<b>GCAATTGTTGGCGATATCATTAGT</b> GACGAGGCAGCGTCATTCCAGGATCKTGGTT					1581811
1581810	<b>TGTTGCCCAAGTGCAGTTGACTGCCTCGCCGGATGGCAAGAGCAAGTGCATGGCATT</b>					1581751
1581750	<b>ACGAGCCGATTACCGTATGTACAATGTTGCTTGATGCGAGGTCATTGTTGCTGACAA</b>					1581691
1581690	CCATCAG <b>GCTCTGCACCAGACATCAGCGCAAGGGTATCGTCAACCCCGTCCGATGATC</b>					1581631
1581630	<b>TTGTCTTGGGCATGATGTGCAAGTACTCTCTCCAGCAGCCCAGCTGGCCAAGAAGATC</b>					1581571
1581570	<b>GACGAGGGCGTGGAGAACGTGATTGAGAAGGGTATCAACACTGCCGATATTGGTGGTTCG</b>					1581511
1581510	<b>GCCAAGACTGCCGAGGTGGGTGA</b> Y <b>GCTATTGCGAAGGAGCTCGAGGCCCTGCTGAAGTAA</b>					1581451
1581450	<b>ACGGTGCAACTKGGTACTGG</b> S <b>TACACTCGCGTTCGCGGAGTGTGGACAAGAGGATGGC</b>					1581391
1581390	<b>GCAAGGTCAAACAAAAGTTGTAAGTGAAGCTTCTGGG</b> T <b>GGAACATAAX</b> T <b>TCTCTAT</b>					1581331
1581330	<b>AGAGGAACCTTGATGACGACAAATGGAACAAACCCCTGGGCGTAATGATCACGAAGGT</b> CAC					1581271
1581270	CATCGACCTAGAGATTGCCCTCGCCGGTGCCGGTGCCTGATGCTATCACAGAGA					1581211
1581210	CAGTTCAGGACAGGTCTCGCTCGTTGTCT <b>TCCTACTGCCACAGTGGCGCAACGGT</b> TAT					1581151
1581150	CCACCGCCCTCTCCACCACTGCTCTACCCGATCTCCAAACGTCCGGTCCCACACTCCA					1581091
1581090	CCTCCAACCTCTCCAACACACCTTCGCCACTGCC <b>CTTGCCTT</b> R <b>TCCACCTTGATCTCAGCCGG</b>					1581031
1581030	GCACCTGGAAATGCCAGCGCATGCAGCCTCGTCAACTCCGCCCTCAGCCTCTCCCTCT					1580971
1580970	CCACCTCGTCGACTTCTCCTCTCCGCTCGACCATTCTCAASAGCAGGCATTTGCG					1580911
1580910	GCCTCGGATTGGACCCACACAAATCCTCCCCAAACTGCCCTCAAGAAACATACACAACC					1580851
1580850	TCCCCAACCTTCCTCAGGACTCACATTGCATGCCATTGGGATTCTGCCCTTCCCAT					1580791
1580790	CCCCATTGACACCATTTCTCGACTGGGT <b>CATCT</b> Y <b>CAAAGTGTGCGCATGTTGTC</b>					1580731
1580730	GG					1580729

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

	Exons	LEUC exons	All exons in this region			
Variants	3 prime UTR	Intronic	Missense	Stop gained	Synonymous	Upstream
Markup	loaded					
1582590	AAAGCATACAG <b>GTGCTCGACGTGATCGACAACAGCAATGCCGACGTTCACTTCAACATT</b> C					1582531
1582530	<b>AACCAACACTGCTCGGTGGC</b> GTACGTTGCTCCGACCCACCTTCCGAAAAGCCAGGGCT					1582471
1582470	AACCGAACGAG <b>GCCTM</b> GMT <b>CGATGCCACGGCGAGCCTCTAACAGACGAAGC</b> ACTGCCG					1582411
1582410	<b>CTGCCAAAS</b> CAGCCGACGCCGTGATCCTCGGGAGCCATGGCGGTTCCCAAATGGGGCACGG					1582351
1582350	<b>GCAAAGTCCGTCCAGAGCAAGGCATCCTCGCCCTCCGCAAAGAAATGGGCACATATGGAA</b>					1582291
1582290	<b>ACCTGCGCCCATGCTTCTCGCGTCTGAGAGCCTCGTCAAGACTCTCCATTGAAGGAGG</b>					1582231

- Which exon does the stop gained mutation fall in?

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

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

Download sequence

File name: Zymoseptoria\_tritici\_LEUC\_sequence

-- Choose Format --

FASTA

RTF (Word-compatible)

Preview  Download  Download Compressed

**Settings**

Sequences to export:

Select/deselect all

cDNA (transcripts)

Coding sequences (CDS)

Amino acid sequences

5' UTRs

3' UTRs

Exons

Introns

Genomic sequence

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

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

(c) Export this sequence in RTF format

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

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

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

## GO: Biological process ?

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

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

### PHI: Phibase identifier

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

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

Show transcript table

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

## Gene: LEUC Mycgr3G103221

Description	3-isopropylmalate dehydrogenase [Source:UniProtKB/Swiss-Prot;Acc: <a href="#">Q9Y897</a> ]																																		
Location	<a href="#">Chromosome 2: 1,581,329-1,582,830</a> reverse strand. MG2:ACPE01000002.1																																		
About this gene	This gene has 1 transcript ( <a href="#">splice variant</a> ), <a href="#">334 orthologues</a> and <a href="#">7 paralogues</a> .																																		
Transcripts	<a href="#">Show transcript table</a>																																		
<table border="1"> <thead> <tr> <th colspan="8">Show/hide columns (1 hidden)</th> <th>Filter</th> </tr> <tr> <th>Name</th> <th>Transcript ID</th> <th>bp</th> <th>Protein</th> <th>Biotype</th> <th>UniProt</th> <th>RefSeq</th> <th>Flags</th> <th></th> </tr> </thead> <tbody> <tr> <td>-</td> <td><a href="#">Mycgr3T103221</a></td> <td>1237</td> <td>365aa</td> <td>Protein coding</td> <td><a href="#">F9X2A6</a></td> <td>-</td> <td>Ensembl Canonical</td> <td></td> </tr> </tbody> </table>									Show/hide columns (1 hidden)								Filter	Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags		-	<a href="#">Mycgr3T103221</a>	1237	365aa	Protein coding	<a href="#">F9X2A6</a>	-	Ensembl Canonical	
Show/hide columns (1 hidden)								Filter																											
Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags																												
-	<a href="#">Mycgr3T103221</a>	1237	365aa	Protein coding	<a href="#">F9X2A6</a>	-	Ensembl Canonical																												

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

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

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

**Transcript-based displays**

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

**Configure this page**

**Custom tracks**

**Export data**

**Share this page**

**Bookmark this page**

**Transcript: Mycgr3T103221**

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

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

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

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

**Summary**

Export image

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

Version: Mycgr3T103221

Type: Protein coding

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

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

**About Us**

[About us](#)

[Contact us](#)

**Get help**

[Using this website](#)

[Documentation](#)

**Our sister sites**

[Ensembl](#)

[Ensembl Bacteria](#)

**Follow us**

[Blog](#)

[Twitter](#)

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

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

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

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

## cDNA sequence

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

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

- Variants are filtered by consequence type

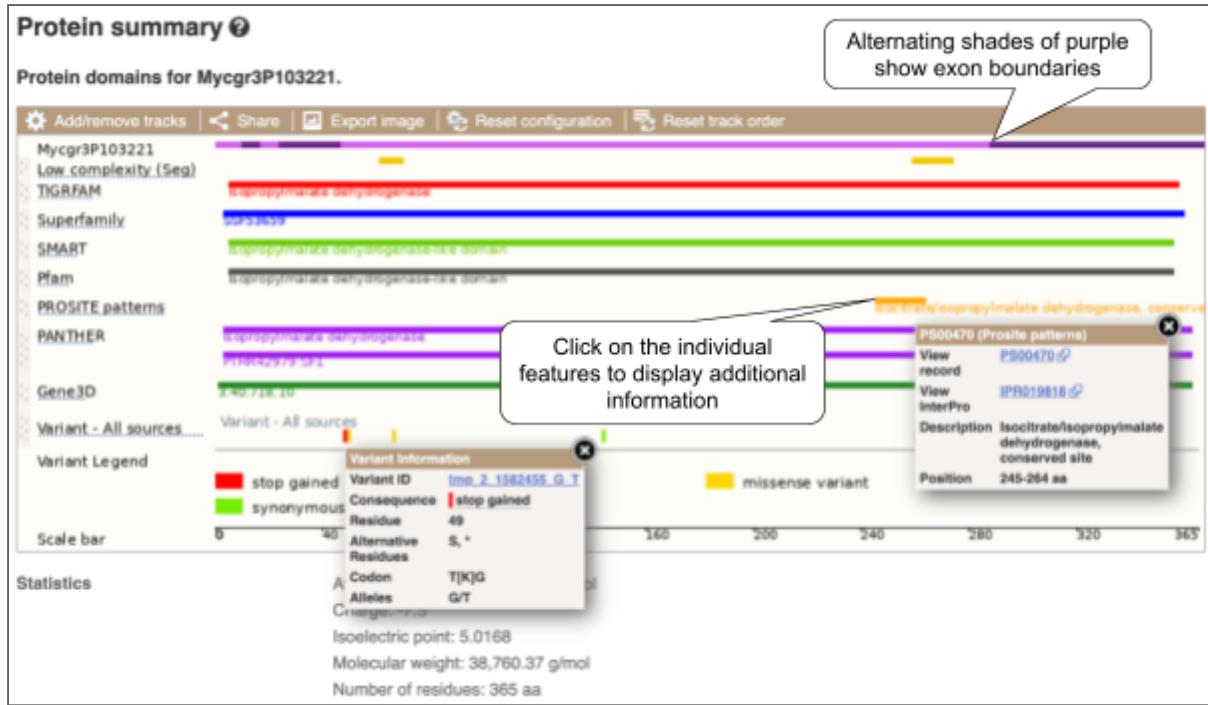
1	AACTGAATCGTCGAAAAATGCCACTTACAACATCGTCGTGGAGGTGATCACTGCG	60
	.....ATGCCACTTACAACATCGTCGTGGAGGTGATCACTGCG	43
	.....M--P--T--Y--N--I--V--V--F--G--G--D--H--C--	14
61	GTCCTGAAGTCACCGCCGAGGCCGCTGAAGGTGCTCGACGTGATCGACAACAGCAATGCCG	120
44	GTCCTGAAGTCACCGCCGAGGCCGCTGAAGGTGCTCGACGTGATCGACAACAGCAATGCCG	103
15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

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

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

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

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



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

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

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

## General identifiers

This transcript corresponds to the following database identifiers:

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