

Exercise: Searching Ensembl Fungi species

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

Navigate to fungi.ensembl.org. You'll see a homepage similar to this:

The screenshot shows the 'Anatomy of the Ensembl Fungi homepage'. The page has a header with the Ensembl Fungi logo and navigation links like HMMER, BLAST, BioMart, Tools, Downloads, Help & Docs, and Blog. A search bar at the top right contains the placeholder 'Search Ensembl Fungi...'. To the left of the search bar, there's a 'Login/Register' button. The main content area is divided into several sections:

- 1: Website header**: Points to the Ensembl Fungi logo and navigation links.
- 2: Main search box**: Points to the search bar and placeholder text 'e.g. NAT2 or alcohol*'. Below it is a dropdown menu for 'All genomes' and a list of 'Favourite genomes' including *Saccharomyces cerevisiae* (R64-1-1), *Schizosaccharomyces pombe* (ASM224v2), *Aspergillus nidulans* (ASM114v1), *Puccinia graminis* (ASM1499v1), *Magnaporthe oryzae* (MG8), and *Zymoseptoria tritici* (MG2).
- 3: Genome and species directory**: Points to the 'View full list of all species' link in the genome dropdown.
- 4: Release details**: Points to a box titled 'What's New in Release 60' which lists updates for genomes, updated protein features, BioMarts, and pan-taxonomic gene trees.
- 5: Archive sites**: Points to a section titled 'Archive sites' which lists previous releases (49-53) and their URLs.
- 1a: Log in to Ensembl**: Points to the 'Login/Register' button.
- 1b: Quick search**: Points to the search bar.

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 table is titled 'Find a Species' and has a 'Download table' button. It includes a 'Search the table' input field and a 'Bipolaris' filter. The columns are: Name, Classification, Taxon ID, Assembly, Accession, Variation database, Regulation database, Whole genome alignments, Other alignments, In peptide compara, and In pan-taxonomic compara. The data rows are:

Name	Classification	Taxon ID	Assembly	Accession	Variation database	Regulation database	Whole genome alignments	Other alignments	In peptide compara	In pan-taxonomic compara
Bipolaris maydis ATCC 48331 (GCA_000354255)	Pleosporales	665024	CochxC4_1	GCA_000354255.1	-	-	-	-	-	-
Bipolaris maydis C5 (GCA_000338975)	Pleosporales	701091	CochxC5_3	GCA_000338975.1	-	-	-	-		-
Bipolaris oryzae ATCC 44560 (GCA_000523455)	Pleosporales	930090	Cochliobolus_miyabeanus_v1.0	GCA_000523455.1	-	-	-	-		-
Bipolaris sorokiniana ND90Pr (GCA_000338995)	Pleosporales	665912	Cocsa1	GCA_000338995.1	-	-	-	-		-
Bipolaris victoriae F13 (GCA_000527765)	Pleosporales	930091	Cochliobolus_victoriae_v1.0	GCA_000527765.1	-	-	-	-		-
Bipolaris zeicola 26-R-13 (GCA_000523435)	Pleosporales	930089	Cochliobolus_carbonum_v1.0	GCA_000523435.1	-	-	-	-		-

A callout box points to the 'Bipolaris' filter with the text 'Data types available for this assembly'.

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 *Bipolaris oryzae*?

The screenshot shows the Ensembl Fungi species homepage for *Bipolaris oryzae* ATCC 44560 (GCA_000523455). The top navigation bar includes links for HMMER, BLAST, BioMart, Tools, and More. A search bar at the top right contains the placeholder "Search Ensembl Fungi...". Below the header, a banner displays the species name and its genome assembly ID (GCA_000523455).

The main content area is organized into several sections:

- Search:** A search bar with a "Go" button and a placeholder "Search ...". Below it, a note says "e.g. COCMIDRAFT_61370 or KI963956:111801-112479 or synthetase". A callout bubble points to this area with the text "Search for a feature in this genome".
- About this species:** A large callout bubble containing information about the species, including its common name (*Cochliobolus miyabeanus*), its role as a causal agent of the Bengal famine, and its use as a biological weapon during World War II. It also links to Wikipedia and provides a Taxonomy ID (930090) and Data source (European Nucleotide Archive).
- Genome assembly:** Information about the genome assembly, including the name (*Cochliobolus miyabeanus v1.0*), download links for DNA sequence (FASTA) and data in Ensembl Fungi, and a link to more information and statistics.
- Comparative genomics:** Information about homologues, gene trees, and whole genome alignments across multiple species, with links to more about comparative analyses, phylogenetic overview of gene families, and download alignments (EMF).
- Gene annotation:** Information about protein-coding and non-coding genes, splice variants, cDNA and protein sequences, and non-coding RNAs. It includes a list of example genes (Pax6, INS, FOXP2, BRCA2, DMD, ssh), a link to more about this genebuild, download links for genes, cDNAs, ncRNA, protein FASTA - GFF3, and an update key, and a link to the Variant Effect Predictor (Ve!P). A callout bubble points to this area with the text "Links to example feature pages and gene annotation download".

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

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

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

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

Bipolaris oryzae ATCC 44560 (GCA_000523455) (Cochliobolus_miyabeanus_v1.0) ▾



Bipolaris oryzae ATCC 44560 (GCA_000523455) 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 assembly submitted to [INSDC](#) with the assembly accession [GCA_000523455.1](#).

Annotation

Who did the gene annotation?

The annotation presented is derived from annotation submitted to [INSDC](#) with the assembly accession [GCA_000523455.1](#), with additional non-coding genes derived from [Rfam](#). For more details, please visit [INSDC annotation import](#).

More information

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

Statistics Summary

Statistics about the genome assembly and annotation

Assembly	Cochliobolus miyabeanus v1.0, INSDC Assembly GCA_000523455.1
Database version	103.1
Golden Path Length	31,362,097
Genebuild by	
Genebuild method	Import
Data source	European Nucleotide Archive

Gene counts

Coding genes	12,002
Gene transcripts	12,002

Terms underlined have mouse-over definitions

ENA
European Nucleotide Archive

Enter text search terms Search

Examples: histone, BN000065

GCA_000523455.1

Examples: Taxon:9606, BN000065, PRJEB402

Home | Submit ▾ | Search ▾ | Rulespace | About ▾ | Support ▾

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

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](#)

Organism:	Bipolaris oryzae ATCC 44560
Accession:	GCA_000523455
Assembly Level:	scaffold
Strain:	ATCC 44560
Genome Representation:	full

Exercise: Ensembl Fungi ‘Region in detail’ view

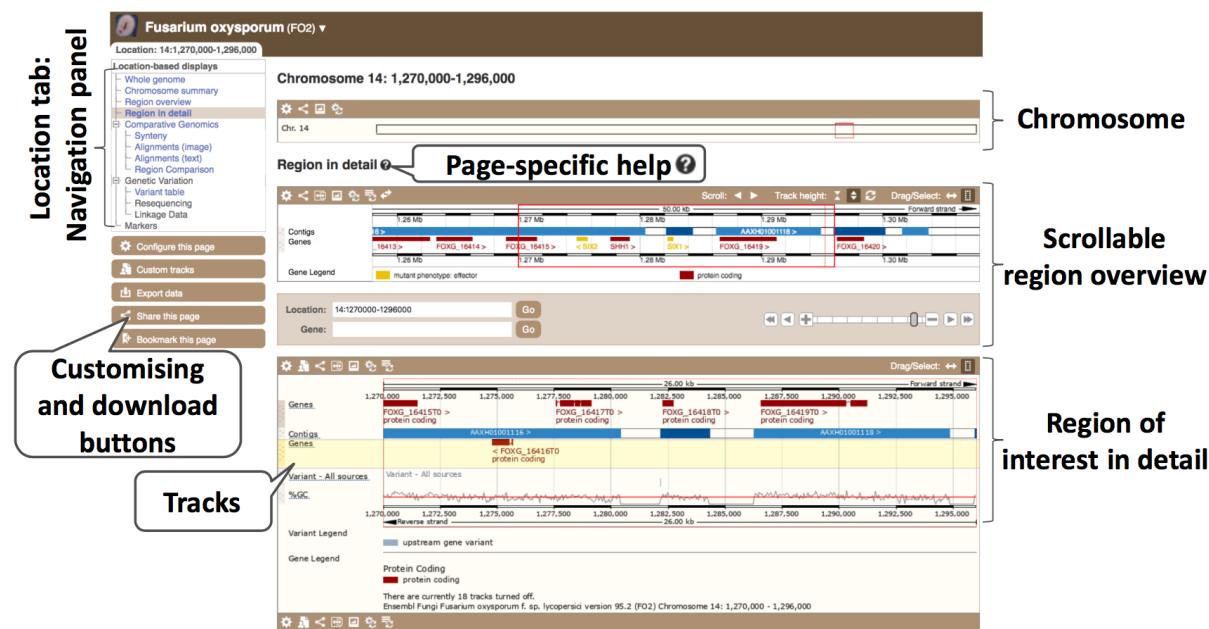
Start at the Ensembl Fungi front page, fungi.ensembl.org. You can search for a region by typing it into a search box, but you have to specify the species.

(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.

Search: Zymoseptoria tritici
Location: 3:1313900-1328300 **Go**
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/tKEvgPUq94>.

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



Let's look at each image on detail.

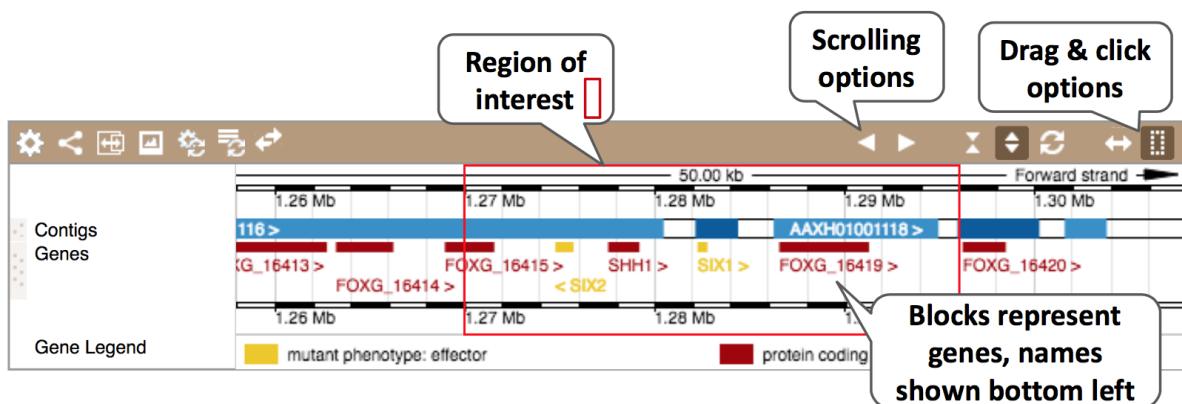
The first image shows the chromosome. You can jump to a different region by dragging out a box in this image. Drag out a box on the chromosome; a pop-up menu will appear.

Chromosome 14: 1,270,000-1,296,000

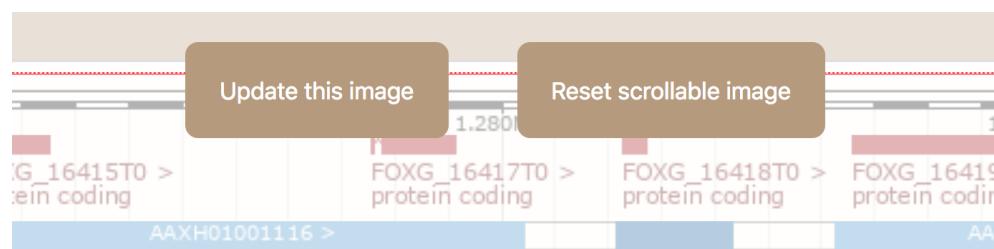


If you would like to move to the region, you could click on [Jump to region \(### bp\)](#). To highlight it, click on [Mark region \(### bp\)](#). For now, we'll close the pop-up by clicking on the X on the corner.

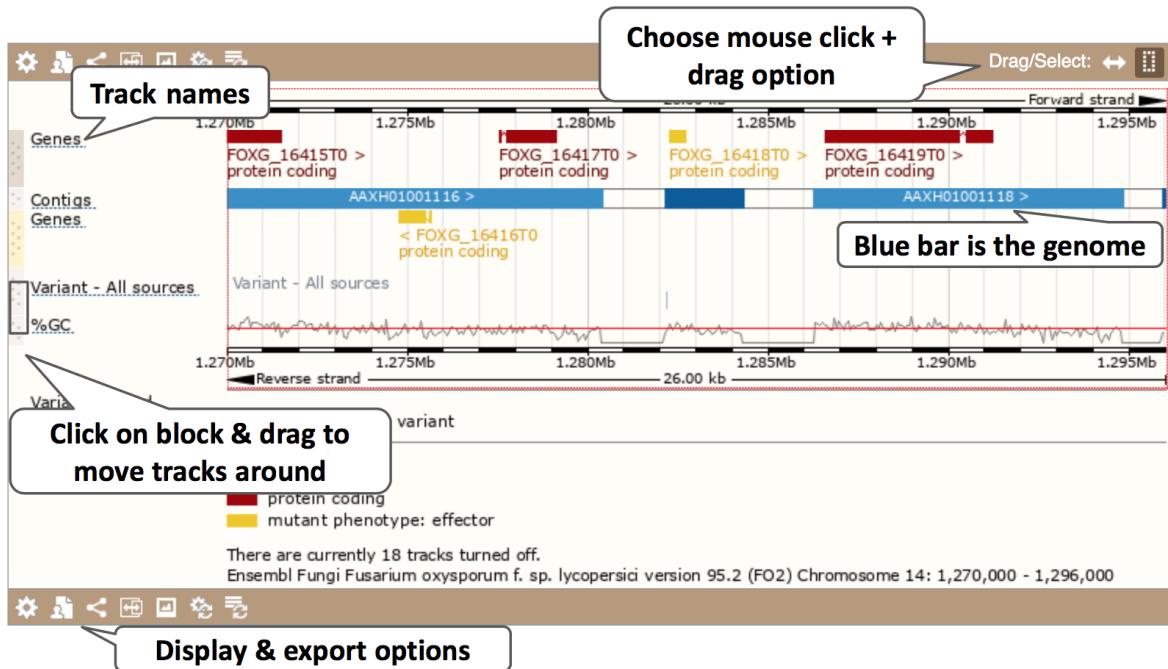
The second image shows a 50 kb region around our selected region. This view allows you to scroll back and forth along the chromosome.



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](#).



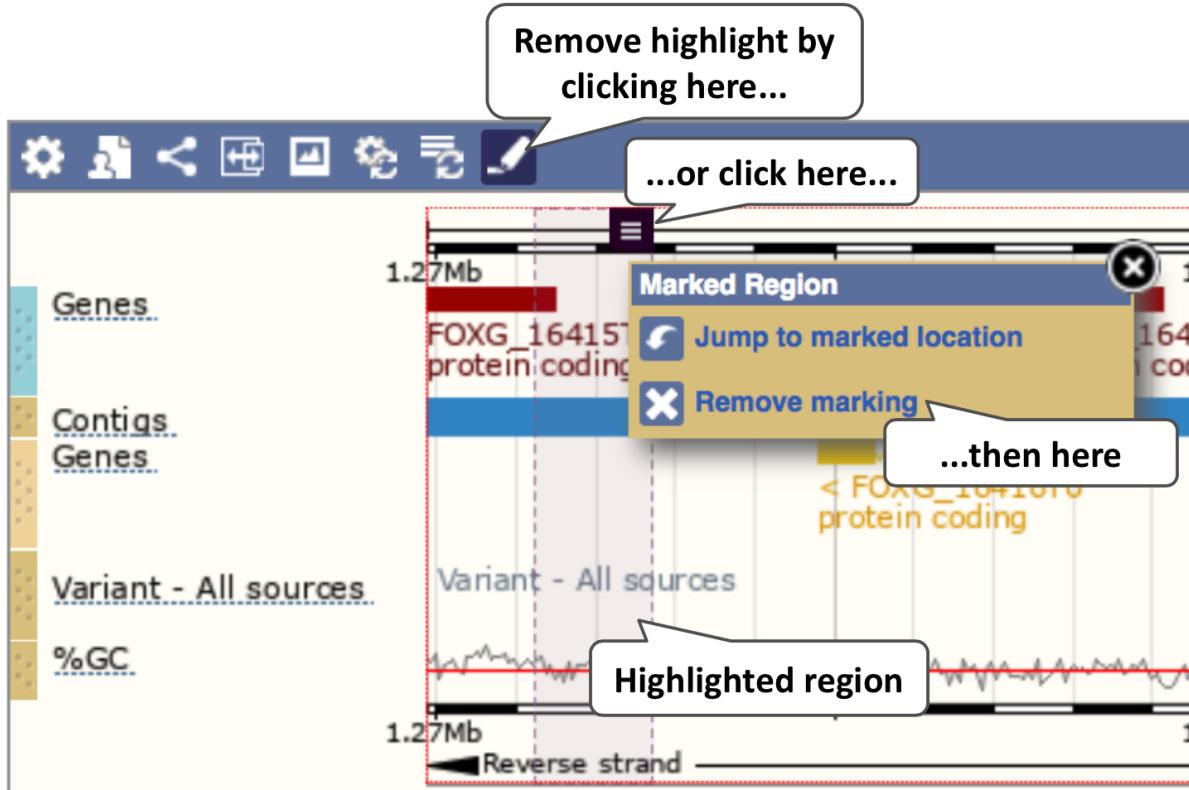
The third image is a detailed, configurable view of the region similar to this:



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.

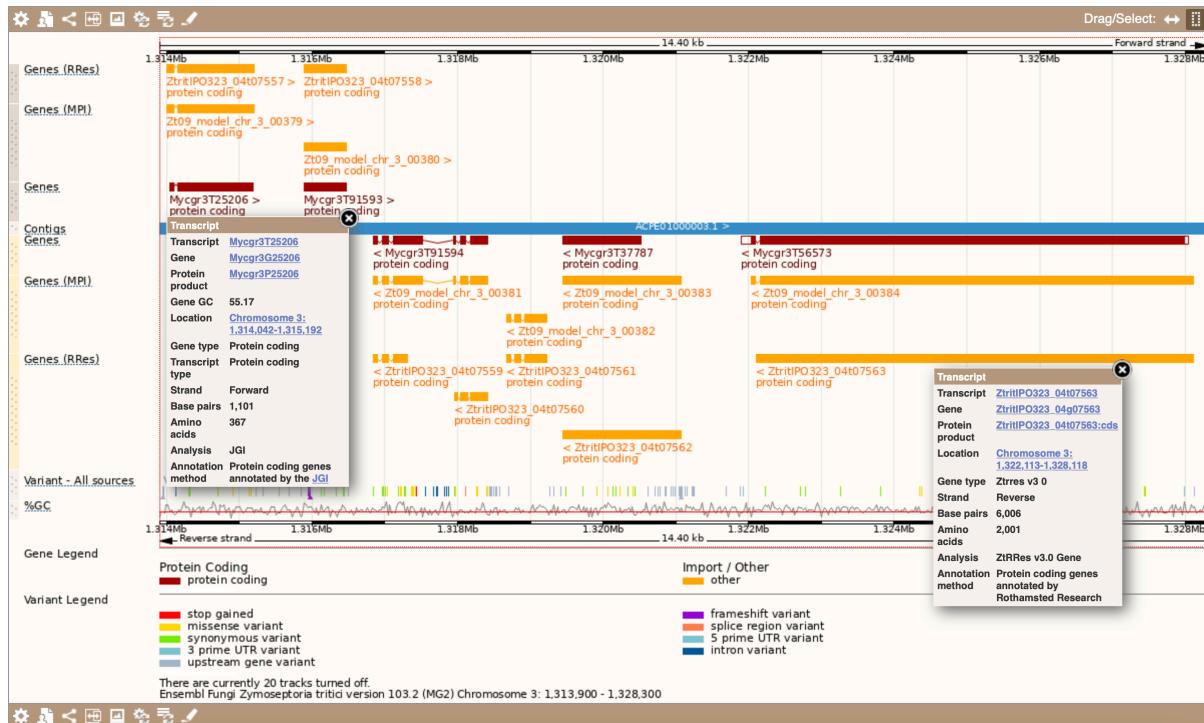
Click on the **Drag>Select** option at the top or bottom right to switch mouse action. On **Drag**, you can click and drag left or right to move along the genome, the page will reload when you drop the mouse button. On **Select** you can drag out a box to highlight or zoom in on a region of interest.

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.

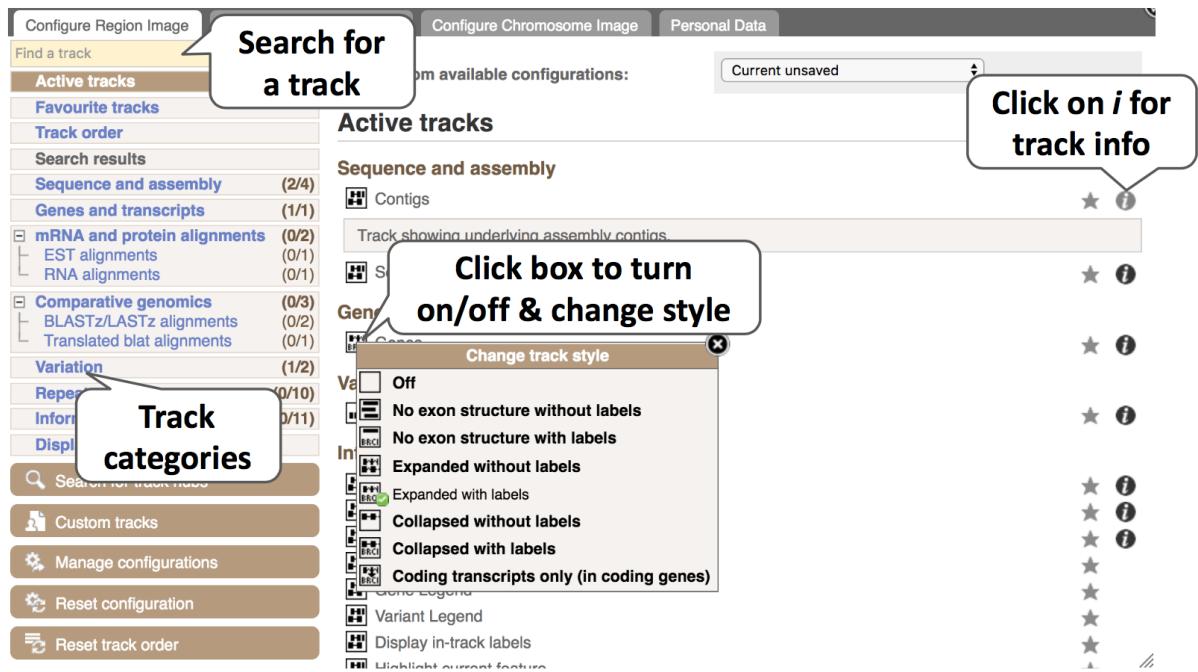
(b) How many genes are annotated in the current region? How many on the forward and how many on the reverse strand? Are they all annotated by the same institute?



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

 [Configure this page](#)

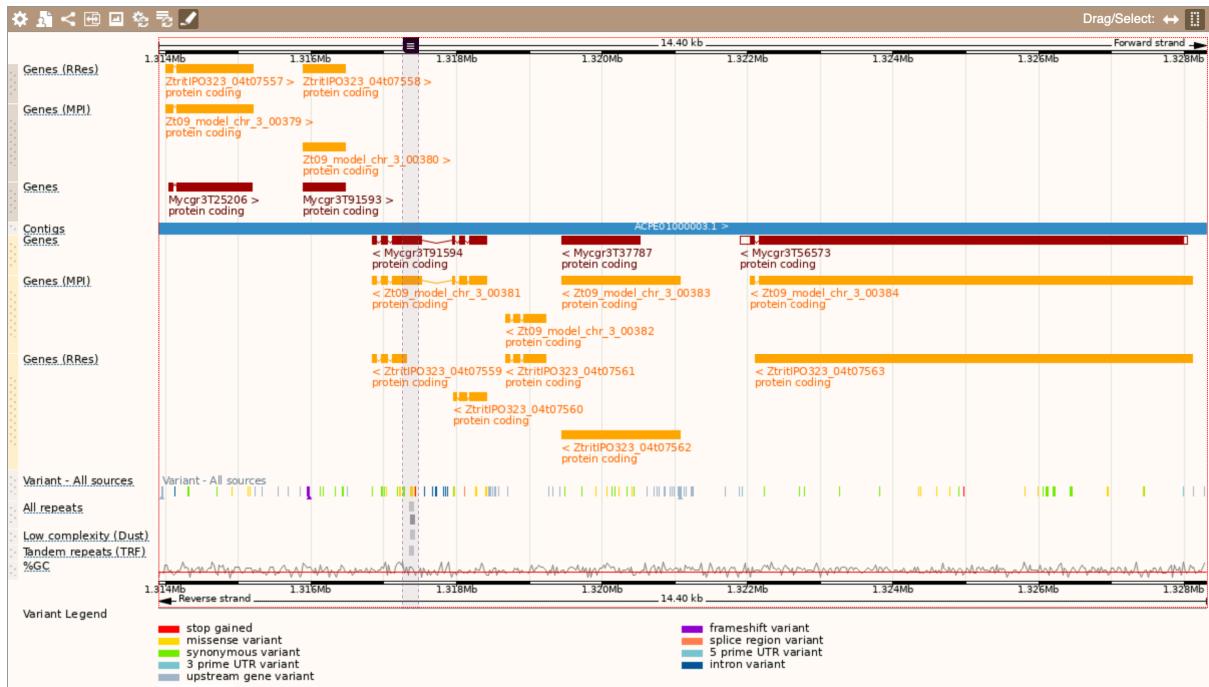
This will open a menu that allows you to change the image. You can put some tracks on in different styles; more details are in this FAQ: <http://www.ensembl.org/Help/Faq?id=335>.



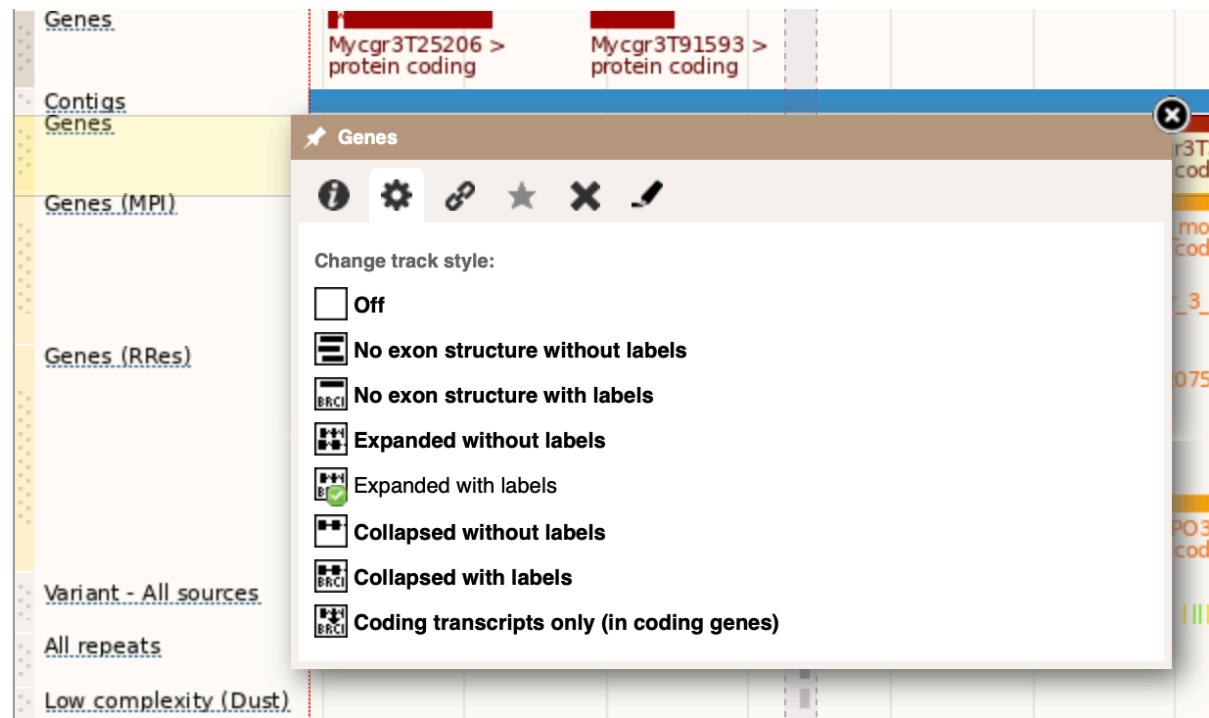
You can add a track to the image by clicking on the box and choosing track style. Click on the tick in the top right hand tick to save and close the menu. Alternatively, click anywhere outside of the menu.

Let's add some tracks to this image.

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



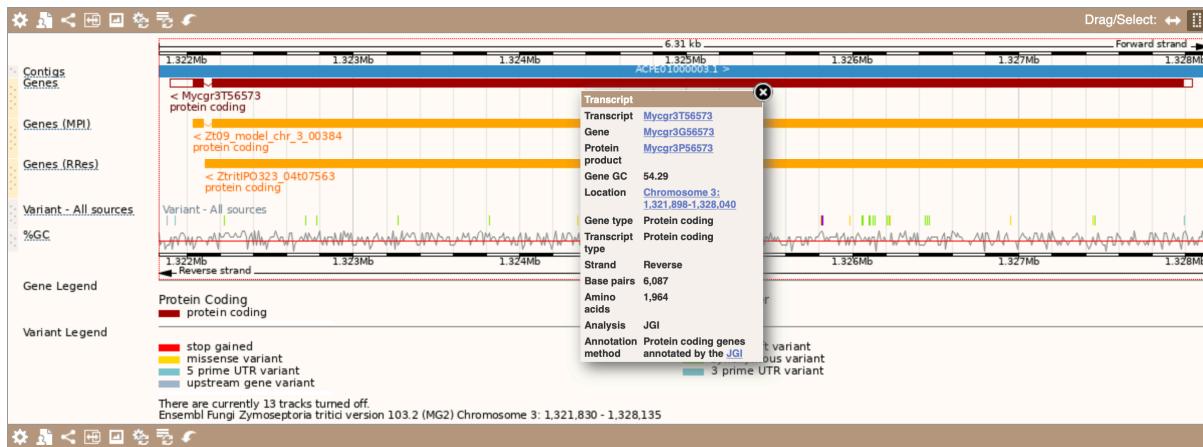
We can also change the way the tracks appear by clicking on the track name to open a menu.



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 Configuration - Feature List

Location to export:

chromosome:MG2:3:1321830:1328135:1

Output:

FASTA sequence



*

Select location:

3 *

1321830 *

1328135 *

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

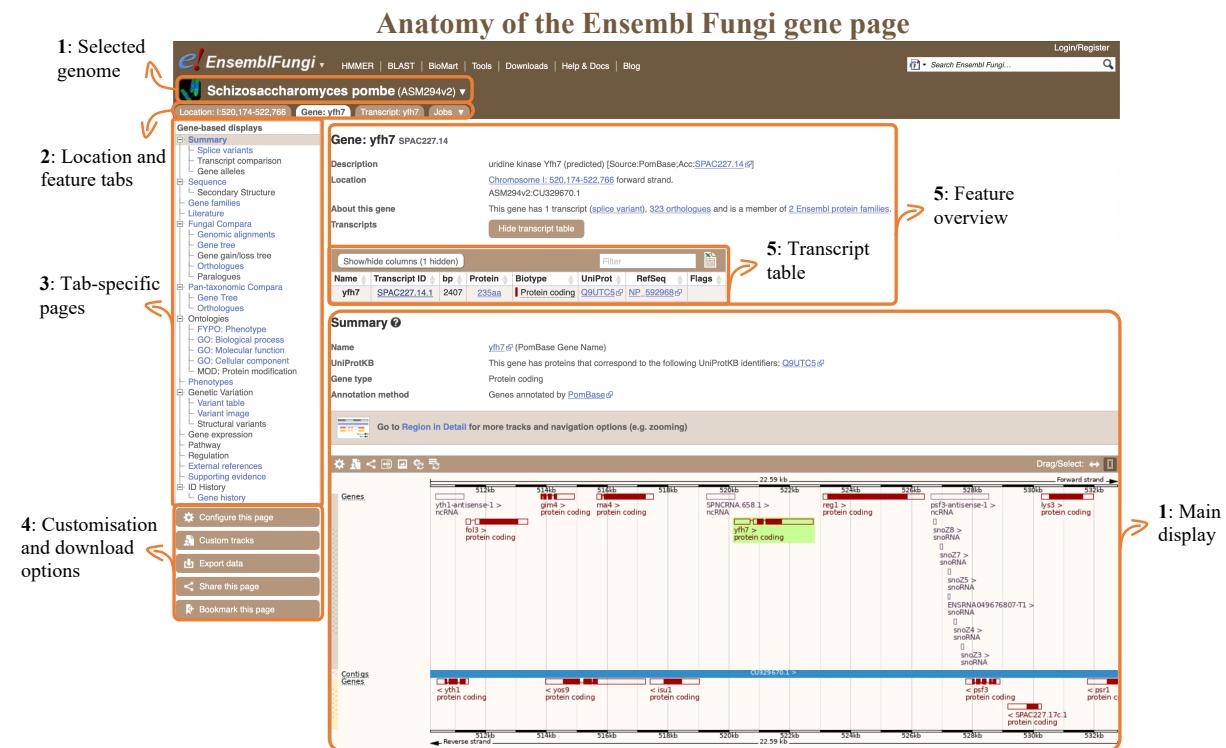
To return this to the default view, go to [Configure this page](#) and select [Reset configuration](#) at the bottom of the menu.

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, type *LEUC* into the main search box, click the drop-down menu and select *Zymoseptoria tritici* and click the Go button.

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



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 whole genome sequence

Exons LEUC exons All exons in this region

Markup loaded

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

GTAGCGAATGCTTCCAGGCTGATGTCGGAGCTGGCAGTTGCGAAGTAGCTCGTCCAG
GGTCAAGATCTCGTCAGAGTGGAGTGCCTGCGCTTACGGGCGACCTGCGGATCCAA
TATGTCGCAATCCACCGGGCTTGTGGCGCAATTCTCGCGCTGAATGCAAGCGGCTG
ATTCAAGTTTGAGGCCATCTGGCTTCATCGATTGACTGCGGTTGAGTGTCAAGGTGAG
CGCCAGTGTGCCATTGCTGGAGGCTCGATTCCTGCTGATGGTGCCTGTTGTGCAGC
CGCG
TTCC
CTAG
GTGG
ATCTCGA
First exon of LEUC gene
Upstream sequence (600bp)
GATGCACTGCCATCCGAGACCTGCGGAGGCGGAAGAG
CATTCGACTGCCATCCGAGACCTGCGGAGGCGGAAGAG
GAGTAAGAGATGAGGGAGTTCTGGGGAGAAAAGAACGAC
CCCCGGATGTGTCCGGAGGGGCAGCGAAGTCGGAGAACGA
AACTGAATCGTCGAAAAATGCCCACTTACAACATCGTCGTTGGAGGTGAGTGCCTA
TGCCCTTGCCAATTGAAATCATGCTGATGTCGATTCAAGTGATCACTGCGGTCTGAAGT
AAGCAGCTCCTTACCGTACTCACTCGTAGCGACTCCAAGCTAACGATTACAGGTCACC
GCCGAGGGCGCTGAAGGTATGCCATACAAGTATCTGCTCAATAAACAGCAACACTGAC
AAAGCATACAGGTGCTCGACGTGATCGACAACAGCAATGCCACGTTCACTTCAACATTC
AACCACACCTGCTCGGTGGC GTACGTTGCTCCGACCACCTTCGGAAAAGCCCAGGGCT
AACCGAAGCAGGCCTCGATCGATGCCACGGGAGCCTCTAACAGACGAAGCACTCCCG
CTGCCAAAGCAGCCGACGGCTGATCCTCGGAGCCATGGCGGTCCCAGGGCACGG
GCAAAGTCCGTCCA GAGCAAGGCATCCGCGCTCG
ACCTGCGCCCATGCTT CGCGTCTGAGAGCTCG
AAGTCTGCCGCC
GTGAGCGCACCO
Highlight portion of sequence to BLAST
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		
1582110			1582110
1582050	GTCCGGTCTGGTCGCTCGACAAGCGAACGTCATGGCTACAAGCAGACTGTGGCGCAAGA		1581991
1581990	CCGTGACGGATGTGTTGCCAACGAGTTCCCGCAGTTGAAGATCGGGCATCACCTCATTG		1581931
1581930	ATTCCCGGGCGATGTTGATGGTCAAGAACCCGAGAGCGTTGAACGGAGTCATCGTCACTA		1581871
1581870	GCAATTGTTGCCGATATCATTTAGTGACGAGGCAGCGTCATTCCAGGATCKTGGGTT		1581811
1581810	TGTTGCCCGAGTGGCAGTTGACTGCCCTGCCCGGATGCCAACGAGCAAGTGCAATGGCATT		1581751
1581750	ACGAGCCGATTCAACGGTATGACATGAAATGTTGCTTGTGAGGTGCATTGTCGACAA		1581691
1581690	CCATCAGGCTCTGCACCAGACATCAGCGGCAAGGGTATCGTCAACCCCGTGCCCATGATC		1581631
1581630	TTGTCTTGGGCATGATGTGCAAGTACTCTCTCAGCAGCCGAGCTGCCAACAGAGATC		1581571
1581570	GACGAGGCCGTGAGGAACGTAATTGAGAACGGGTATCAACACTGCCGATATTGGTGGTT		1581511
1581510	GCCAAGACTGCCGAGGTGGGTGAYGCTATTGCGAACGGAGCTCGAGGCCCTGCTGAAGTAA		1581451
1581450	ACGGTCAACTKGGTACTGGSTACACTCGCCTCGCGAGTGTGATGGACAAAGAGGATGGC		1581391
1581390	GCAAGGTCAAAACAAAAGTTGATGAAAGCTTCTGGGTGGAACATAAYTCTCTAT		1581331
1581330	AGAGGAACCTGATGCACGCAAATGGAACAAACCTGGTCGAATGATCACGAAGGT		1581271
1581270	CATCGACCTAGAGATTGCCCTCGCCGGTGCCTGCGGTCCTTGATGCTATCACAGAGA		1581211
1581210	CAGTCAGGACAGGTCTCGCTCGTTGTCRTCTTACTGCCACAGTGGCGAACGGTAT		1581151
1581150	CCACCGCCCTCTCCACCACCTGCTCTCACCGATCTCCAAACGTCCGGTCCCACACTCCA		1581091
1581090	CCTCCAACCTCTCAACCACACCTTCGCCACTGCCTRCCCACCTTGATCTCAGCCGG		1581031
1581030	GCACTGGAATGCCAGCGCATGCAGCCTCGTCAACTCCGCCCTCAGCCTCTCCCT		1580971
1580970	CCACCTCGTCGACTTCTCTCTCCGTCTGCCATTCTCCASAGCAGGCATCTCG		1580911
1580910	GCCTCGCGATTGGACCCACACAATCTCCCCAAACTGCGCTCAAGAAACATACACAACC		1580851
1580850	TCCCCAACCTTCCTCAGGACTCACATTGCATGCCGATTCTGCCCTCTCCCTCCAT		1580791
1580790	CCCCATTGACACCACCTTCCTCGACTGGGTCATYCAAAGTGTGATCGTGCCATGTTGTC		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		
1582650	GCCGAGGCCGTGAAGGTATGCCATACAAGTATCTCTGCTCAATAAACAAGCAACACTGAC		1582591
1582590	AAAGCATACAGGTGCTCGACGTGATCGACAACAGCAATGCCGACGTTCACTTCAACATTC		1582531
1582530	AACCACACCTGCTCGGTGGCTGTACGTTGCTCCGACCACCTTTCGAAAAGCCAGGGC		1582471
1582470	AACCGAAGCAGGCCTTGTMCGATTGCCCACGGCGGACCCTTAACAGACGAAGCACTCGCCG		1582411
1582410	CTGCCAAAACSAGCCGACCCGTGATCCTCGGAGCCATCGGGGGTCCAAATGGGGCACGG		1582351

- 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.

File name: Magnaporthe_oryzae_ATG8_sequence

File format: -- 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

Accession	Term	Evidence	Annotation source	Mapped using	Transcript IDs	
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		UniProtKB/TrEMBL:F9X2A6_MYCGM	Mycgr3T103221	<ul style="list-style-type: none"> Search BioMart View on karyotype
GO:0055114	oxidation-reduction process	IEA	UniProt	UniProtKB/TrEMBL:F9X2A6_MYCGM	Mycgr3T103221	<ul style="list-style-type: none"> Search BioMart View on karyotype
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																
PHI:2061	2061	ND	Sequence Publications:19115483		MGG_01062T0	<ul style="list-style-type: none"> Search BioMart View on karyotype 															
PHI:2076	2076	ND	Sequence Publications:10717456		MGG_01062T0	<ul style="list-style-type: none"> Search BioMart 															
PHI:2139	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)																	
PHI:768	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> </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									
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																		
				<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>Ref source:Pubmed Year:2009 Author reference:Yi Zhen Deng</td> <td>Phenotypic loss of pathogenicity Disease name:Rice Blast Tissue:leaf Host response:slight induction of hypersensitive reaction Experimental technique:Gene deletion: full</td> <td></td> </tr> </tbody> </table>	Reference	Comments	PHI Phenotype	Pathogen Phenotype	Pmid:19115483	Ref source:Pubmed Year:2009 Author reference:Yi Zhen Deng	Phenotypic loss of pathogenicity Disease name:Rice Blast Tissue:leaf Host response:slight induction of hypersensitive reaction Experimental technique:Gene deletion: full										
Reference	Comments	PHI Phenotype	Pathogen Phenotype																		
Pmid:19115483	Ref source:Pubmed Year:2009 Author reference:Yi Zhen Deng	Phenotypic loss of pathogenicity Disease name:Rice Blast Tissue:leaf Host response:slight induction of hypersensitive reaction Experimental technique:Gene deletion: full																			
				<table border="1"> <thead> <tr> <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 *Zymoseptoria 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: 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), 171 orthologues and 2 paralogues .
Transcripts	Hide transcript table

Show/hide columns (1 hidden)								Filter	X
Name	Transcript ID	bp	Protein	Biotype	UniProt	Flags			
-	Mycgr3T103221	1237	365aa	Protein coding	F9X2A6				

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.

The screenshot shows the Ensembl Fungi transcript summary page for Mycgr3T103221. The main content area includes a summary table with the following data:

Name	Transcript ID	bp	Protein	Biotype	UniProt	Flags
-	Mycgr3T103221	1237	365aa	Protein coding	F9X2A6	

Below the table, there is a large genomic track visualization showing the transcript structure across 1.50 kb. The track indicates 6 coding exons and a total transcript length of 1,237 bp.

(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.

[Download sequence](#)

Download sequence

Exons/ Introns		Translated sequence		Flanking sequence		Intron sequence	UTR
Variants		3 prime UTR	Missense	Stop gained	Synonymous		
Markup loaded							
Show All entries							
Show/hide columns							
No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
5' upstream sequence							
1	Mycgr3E103221.1	1,582,830	1,582,783	-	1	48	AACTGAATCGTCGAAAATGCCCACTTACAACATCGTCGCTTGGAG
	Intron 1-2	1,582,782	1,582,733			50	gtgagtgcgtatgcccttgcattttga..... gtatgtcgattcag
2	Mycgr3E103221.2	1,582,732	1,582,713	1	0	20	GTGATCACTGCCGTCTGAA
	Intron 2-3	1,582,712	1,582,657			56	gtaaggcgtccctttacgtact..... tacag
3	Mycgr3E103221.3	1,582,656	1,582,636	0	0	21	GTCACCGCCGAGGCCTGAAAG
	Intron 3-4	1,582,635	1,582,580			56	gtatgccatacagatatctgtctc..... caagcaacactgacaaagcatacag
4	Mycgr3E103221.4	1,582,579	1,582,511	0	0	69	GTGCTCGACGTATCGACAAACGCAATGCCGACGCTTACTTCACATTCAACCACACTG
	Intron 4-5	1,582,510	1,582,460			51	gtacgttgcgtccgaccacccccc..... aaaagcccgaggctaacccggacag
5	Mycgr3E103221.5	1,582,459	1,581,736	0	1	724	GCTCATACTCGATGCCAACGGCGAGCTCTAACAGACGAAGCCTGCCGTGCGAAAGCA GCCGAGCCGTGATCCTCGGAGGCCATCGGGTCCCAAATGGGCACGGCAAAGTCCGT CCAGACCAAGGCATCTGGCCCTCCGCAAAGAAATGGCACATATGGAAACCTGGCC TGCTTCTTCGGTCTGAGACCTCGTCAAGACTTCTCATTAAGGAGGAAGTCTGGCC GGGTCAACTCAACATCGTCCTGAGCTGACAGGTGGCATCTACTTCGGTGA GAGGACGAGGGCTCGGCTACGGAGTGGACACGGAAACCTACTCGGGCCTGAGATCGAG CGTGTGGCACGACTGGCGGGTTCTTGGCTTGGCTGAAGACCCGCCGTGGCGGTCTGG TCGCTCGACAAGGCGAACGTCATGGCTAACAGCAGACTGTCGGCAAGACCGTGACGGAT GTGTTGCGAACAGGTTCCCGCAGTTGAAGATGGGCATCACCTCATTTGGCGCG ATGTTGATGGTCAGAACCCGAGGGCTTGAAGGGAGTCATCGTCACTAGCAATTGTT GGCGATATCATAGTACGAGGGGAGCGTCACTTCAGGATCCTGGTTGGCGGAG GGCAGTTGACTGCCCTGCCGATGCCAAGGCAAGTGAATGCCATTACGAGGCCATT CAGC
	Intron 5-6	1,581,735	1,581,684			52	gtatgtacaatgttgtcttgatcg..... gtgcattgtgtcgacaaccatcg
6	Mycgr3E103221.6	1,581,683	1,581,329	1	-	355	GCTCTGCACCAAGACATCAGCGCAAGGGTATCGCAACCCCGTCGCCATGATTTCTT TGGGCATGATGTCAGACTCTCTCCAGCAGGGGAGCTGGCCAAGAAGATCGACGGAG CCGTGAGGAACCGTGAATGGAGAAGGTATCAACACTCGCGATATTGGTGGTTCGCCAAGA CTGCGCAGGGGGTGA AACTGGTACTGGTACACTCGCGTTCGGGAGTGATCGGACAAGGGATGGCGCAAGGT CAAAGAACAGTTGAGATGAAAGCTTCTGGCGAACATAATCTCTATAG
3' downstream sequence							

Introns and flanking sequence in lower case

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.

[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

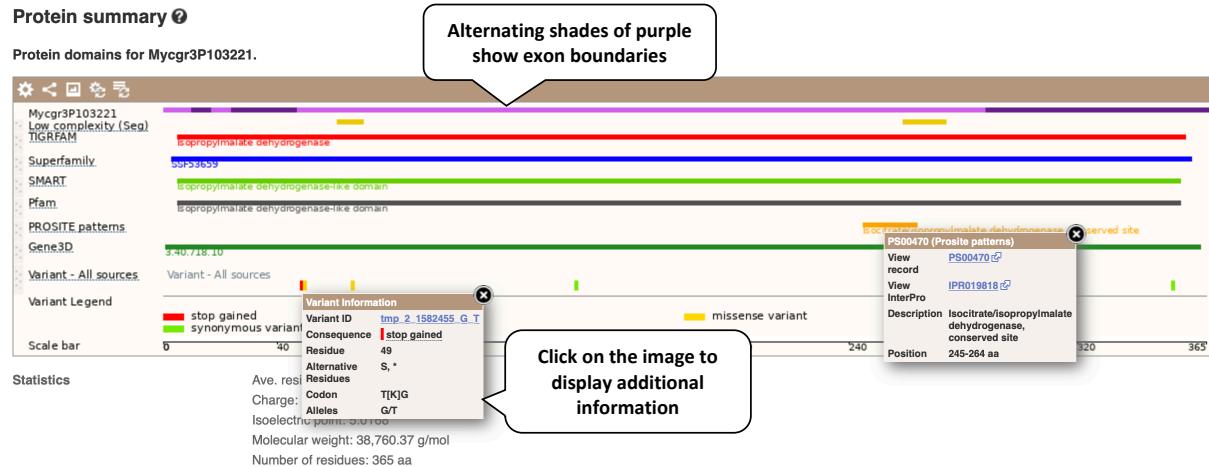
1	AACTGAATCGTCGAAAAATGCCACTTACAACATCGTCGTCTTGAGG	GTGATCACTGCG	60
	ATGCCACTTACAACATCGTCGTCTTGAGG	GTGATCACTGCG
	M--P--T--Y--N--I--V--V--F--G--G--D--H--C--	14
61	GTCCTGAA	GTCACCGCCGAGGGCCTGAAGGTGCTCGACGTGATCGACAACAGCAATGCCG	120
44	GTCCTGAA	GTCACCGCCGAGGGCCTGAAGGTGCTCGACGTGATCGACAACAGCAATGCCG	103
15	G--P--E--V--T--A--E--A--L--K--V--L--D--V--I--D--N--S--N--A--		34
121	ACGTTCACTTCAACATTCAACCACACCTGCTCGGTGGCGCCT	CGATCGATGCCCACGGCG	180
104	ACGTTCACTTCAACATTCAACCACACCTGCTCGGTGGCGCCTCGATCGATGCCCACGGCG		163
35	D--V--H--F--N--I--Q--P--H--L--L--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	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]
KEGG Pathway and Enzyme	00290+1.1.1.85 [align] Valine, leucine and isoleucine biosynthesis [view all locations] 00660+1.1.1.85 [align] C5-Branched dibasic acid metabolism [view all locations]
UniParc	UPI000012E54F [align] [view all locations]
UniProtKB/TrEMBL	F9X2A6_MYCGM [align] [Target %id: 100; Query %id: 100] 3-isopropylmalate dehydrogenase [view all locations]