

## Exercise: Searching Ensembl Fungi species

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

Navigate to [fungi.ensembl.org](https://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. The top navigation bar includes links for HMMER, BLAST, BioMart, Tools, Downloads, Help & Docs, and Blog. A search bar is located in the top right corner. Below the search bar, there is a section for 'All genomes' with a dropdown menu set to 'All species'. To the right of this, there is a 'Favourite genomes' section with a list of species: Saccharomyces cerevisiae, Schizosaccharomyces pombe, Aspergillus nidulans, Puccinia graminis, Magnaporthe oryzae, and Zymoseptoria tritici. On the right side of the page, there is a 'What's New in Release 58' section and an 'Ensembl Rapid Release' section. At the bottom, there is an 'Archive sites' section.

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' page in Ensembl Fungi. The page features a table of genome assemblies for the genus *Bipolaris*. The 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 assemblies for *Bipolaris*: *Bipolaris maydis* ATCC 48331, *Bipolaris maydis* C5, *Bipolaris oryzae* ATCC 44560, *Bipolaris sorokiniana* ND90Pr, and *Bipolaris zeicola* 26-R-13. Annotations highlight the 'Download table' button, the 'Search table' input field, and the 'Data types available for each assembly' section.

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	<a href="#">665024</a>	CocheC4_1	<a href="#">GCA_000354255.1</a>	-	-	-	✓	-	-
<i>Bipolaris maydis</i> C5	Pleosporales	<a href="#">701091</a>	CocheC5_3	<a href="#">GCA_000338975.1</a>	-	-	-	✓	✓	-
<i>Bipolaris oryzae</i> ATCC 44560	Pleosporales	<a href="#">930090</a>	Cochliobolus_miyabeanus_v1.0	<a href="#">GCA_000523455.1</a>	-	-	-	✓	✓	-
<i>Bipolaris sorokiniana</i> ND90Pr	Pleosporales	<a href="#">665912</a>	Cocsa1	<a href="#">GCA_000338995.1</a>	-	-	-	✓	-	-
<i>Bipolaris zeicola</i> 26-R-13	Pleosporales	<a href="#">930089</a>	Cochliobolus_carbonum_v1.0	<a href="#">GCA_000523435.1</a>	-	-	-	✓	-	-

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**

Search for features in this genome

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

Information about this species

**Genome assembly: *Cochliobolus miyabeanus* v1.0**

About this genome assembly and sequence download

**Gene annotation**

Links to example feature pages

**Comparative genomics**

Comparative genomics section

**Variation**

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?

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

**Statistics**

Statistics about the genome assembly and annotation

**Assembly**

Gene annotation provider

**Summary**


Assembly	Cochliobolus miyabeanus v1.0, INSDC Assembly <a href="#">GCA_000523455.1</a>
Database version	111.1
Golden Path Length	31,362,097
Genebuild by	JGI
Genebuild method	Import
Data source	<a href="#">JGI</a>


**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?

**ENA**  
European Nucleotide Archive

Enter text search terms  Search 

Examples: histone, BN000065

GCA\_000523455.1  View 

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

Organism:	<a href="#">Bipolaris oryzae ATCC 44560</a>
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

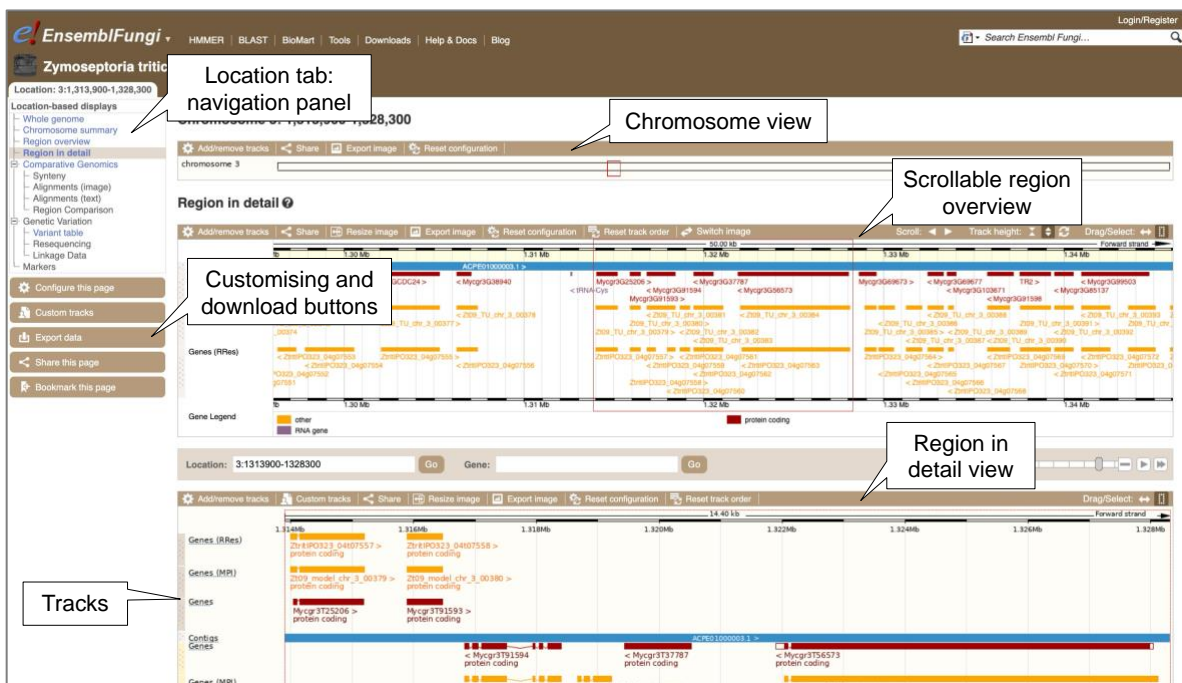
Start at the Ensembl Fungi homepage, [fungi.ensembl.org](http://fungi.ensembl.org). You can search for a region by typing it into a search box, but you have to specify the species.

(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.



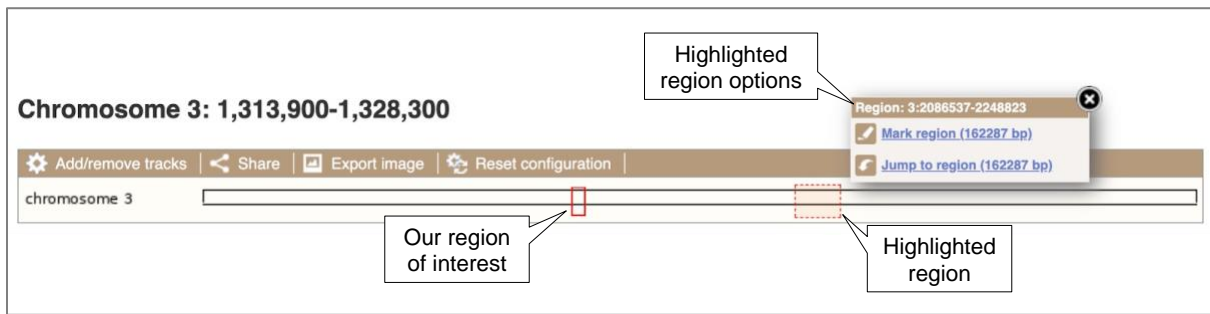
Click on the button  to open a page-specific help page. These help pages provide links to Frequently Asked Questions (FAQs), a glossary, video tutorials, and a form to contact the Ensembl helpdesk. You can find a help video on this page at <http://youtu.be/tKEvgPUq94>.

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



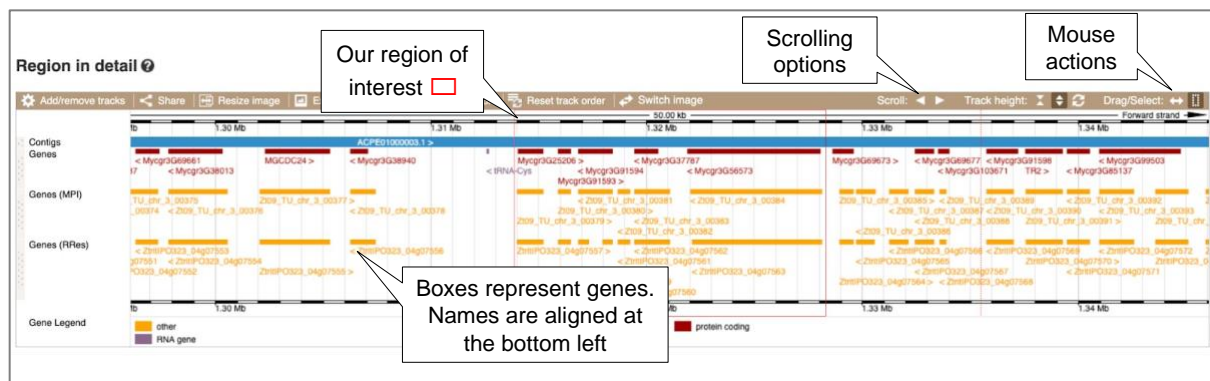
Let's look at each image in detail.


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.

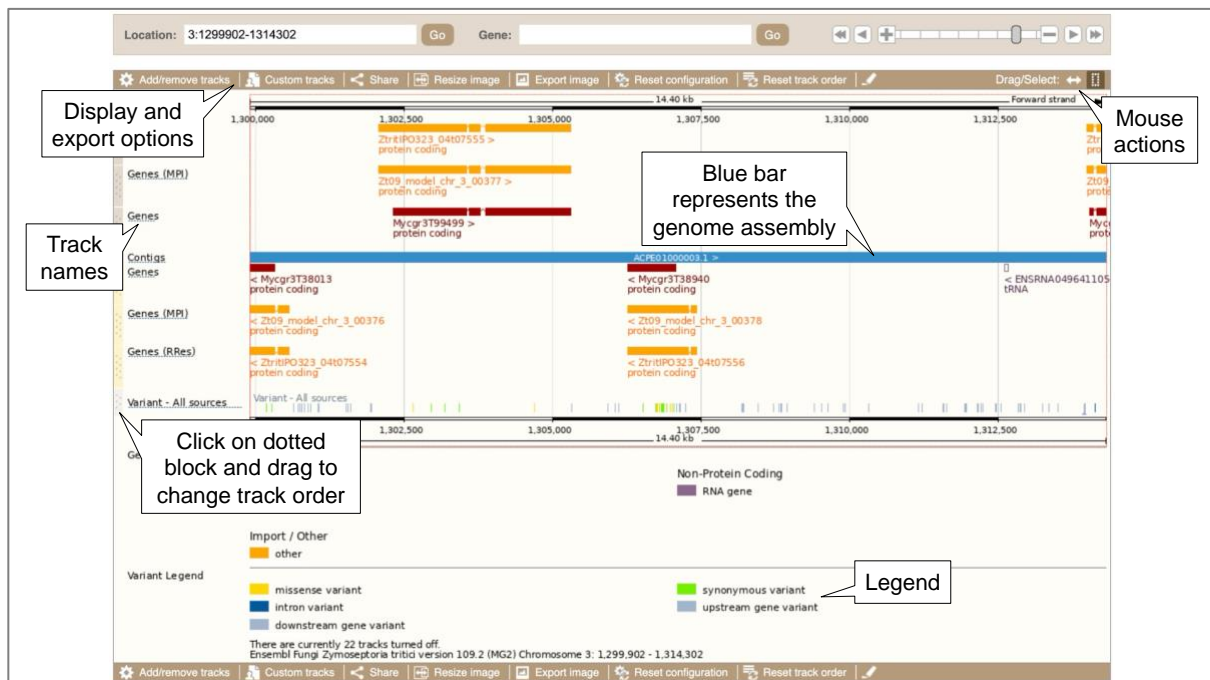


Click on the [Drag/Select](#) button  to change the action of your mouse click. Changing to the arrow allows you to scroll along the chromosome by clicking and dragging within the image. As you do this, you'll see the image below grey out and two buttons appear. Clicking on [Update this image](#) will 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:

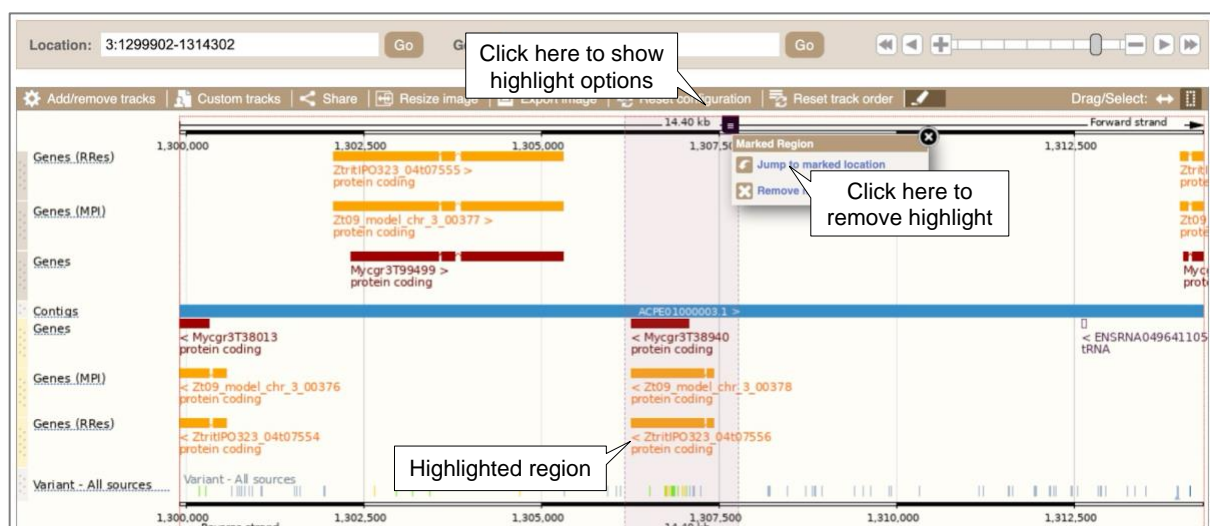




Genes are shown as transcripts with exons represented as boxes and introns shown as lines connecting the exons. Forward-stranded genes are shown above the genome assembly (Contigs track), while reverse-stranded genes are shown below.

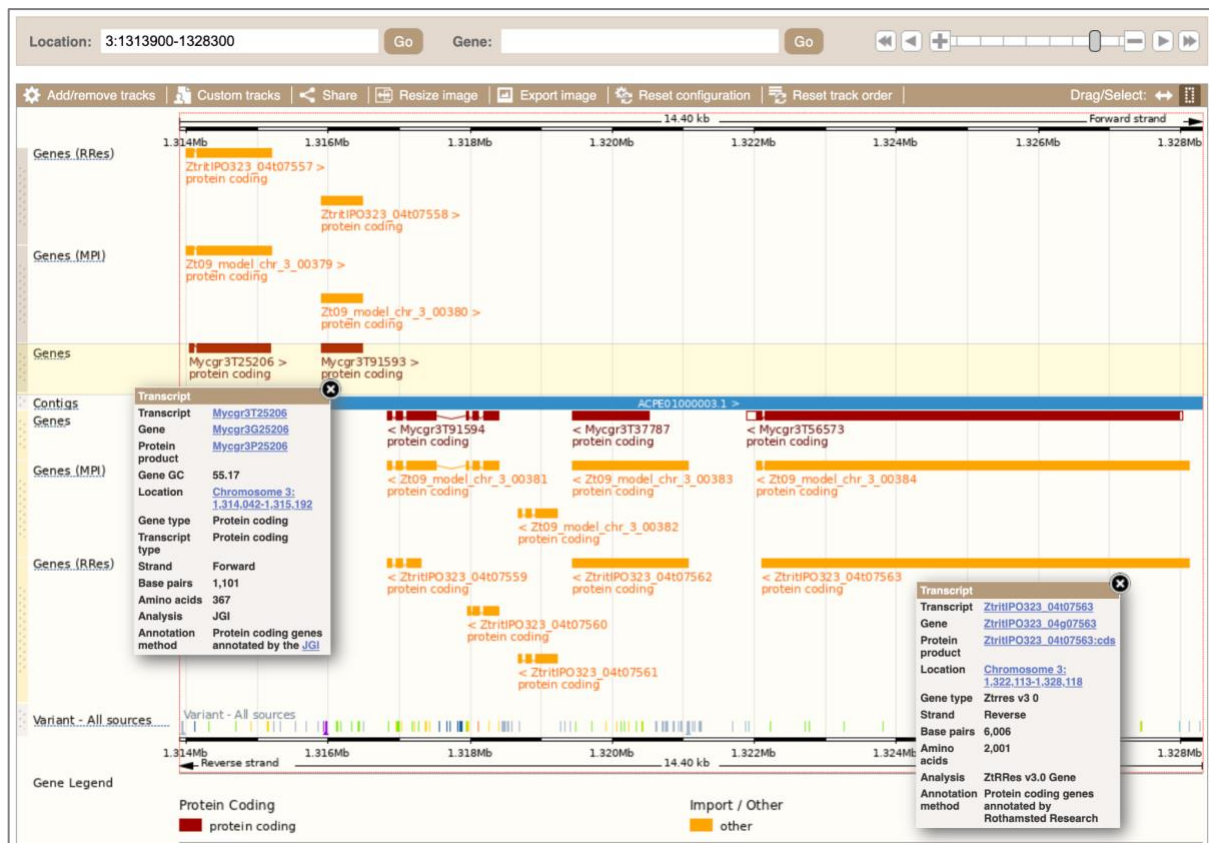
Click on the [Drag/Select](#) option at the top or bottom right to switch your mouse action. On [Drag](#), you can click and drag sideways to move along the genome, the page will reload when you drop the mouse button. The [Select](#) option allows you to drag out a box to highlight or zoom in on a region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

Change your mouse action to [Select](#), drag out a box around an exon and choose [Mark region](#).



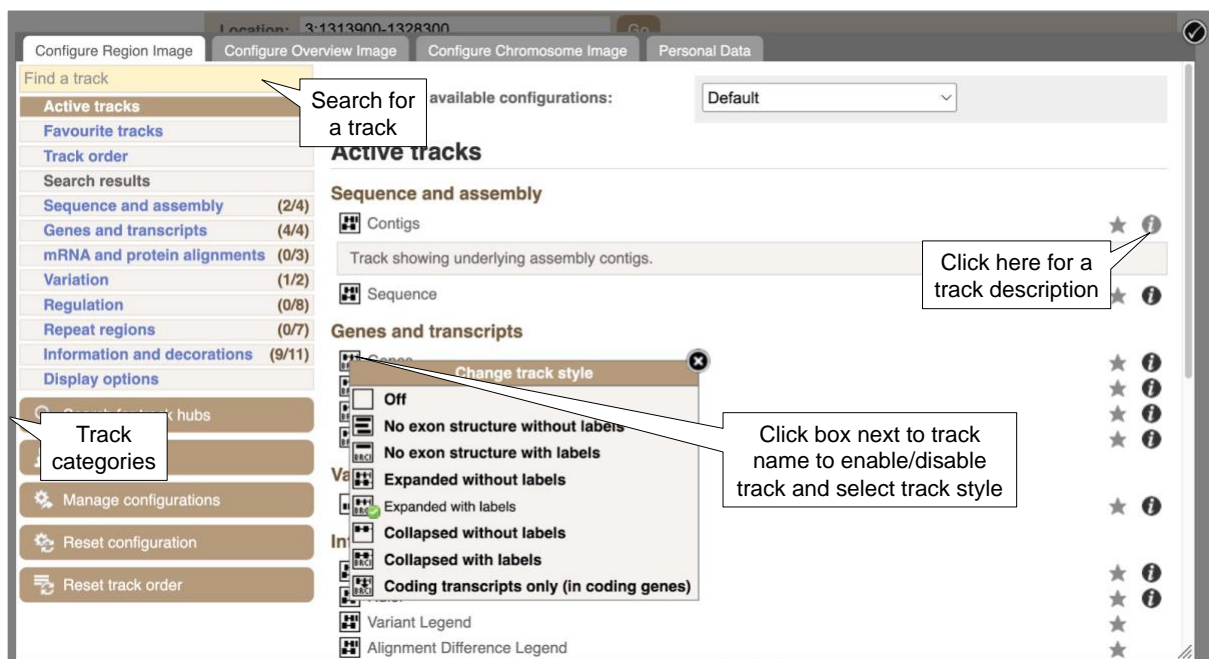
The highlight will remain in place if you zoom in and out or move around the region. This allows you to keep track of regions or features of interest.

(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.

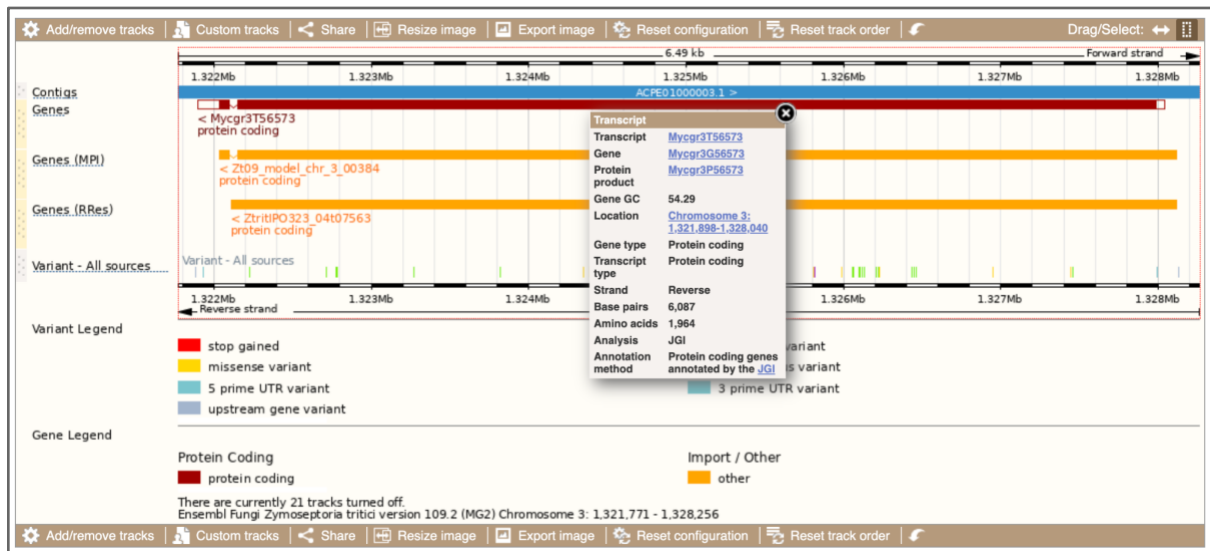
This will open a menu that allows you to change the image. You can put some tracks on in different styles. You can read more details in this FAQ: <https://www.ensembl.org/Help/Faq?id=335>.

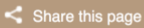







(d) Zoom in on the largest transcript [Mycgr3T565573](#). How many exons does this gene have?



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.

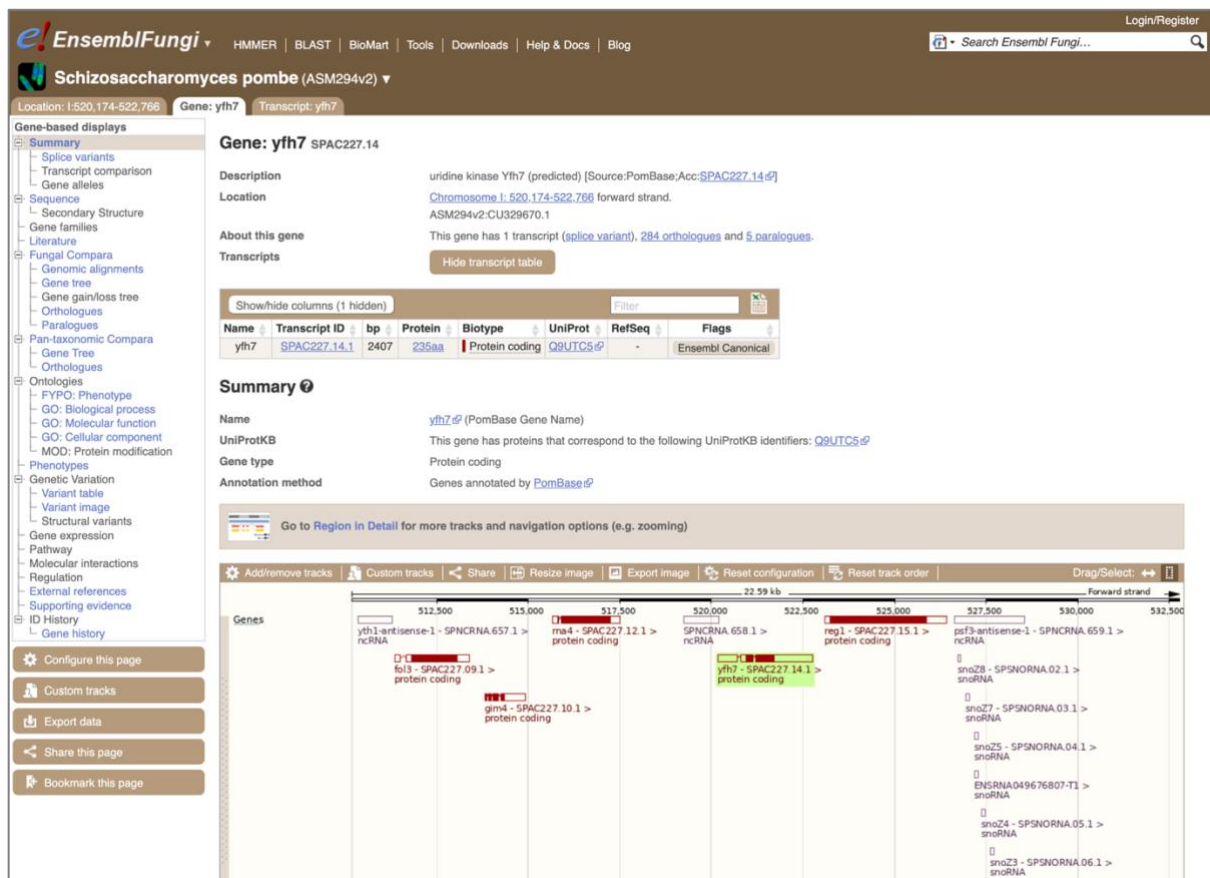
To return the region image to the default view, click on [Reset configuration](#) at the top of the region in detail image.

## 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](https://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:



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?

<b>Gene: LEUC</b> Mycgr3G103221	
<b>Description</b>	3-isopropylmalate dehydrogenase [Source:UniProtKB/Swiss-Prot;Acc: <a href="#">Q9Y897</a>
<b>Location</b>	<a href="#">Chromosome 2: 1,581,329-1,582,830</a> 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

Download or BLAST sequence

Exons LEUC exons All exons in this region

Markup loaded

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

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					

<

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

AACCACACCTGCTCGGTGGCGTACGTTTGCTCCGACCACCTTTCGGAAAAGCCCAGGGCT

1582471

1582470

AACCGAAGCAGGCCTTGMTCGATGCCACGGCGAGCCTCTAACAGACGAAGCACTCGCCG

1582411

1582410

CTGCCAAAACAGCCGACGCGTGATCCTCGGAGCCATCGGCGGTCCCAAATGGGGCACGG

1582351

1582350

GCAAAGTCCGTCCAGAGCAAGGCATCCTGCGCCTCCGCAAAGAAATGGGCACATATGGAA

1582291

1582290

ACCTGCGCCCATGCTTCTTCGCGTCTGAGAGCCTCGTCAAGACTTCTCCATTGAAGGAGG

1582231

Which exon does the stop gained mutation fall in?

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

[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.

**Download sequence**

File name: Zymoseptoria\_tritici\_LEUC\_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.

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*?

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		
<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><a href="#">Search BioMart</a></li> <li><a href="#">View on karyotype</a></li> </ul>	
<a href="#">GO:0008652</a>	cellular amino acid biosynthetic process	IEA	UniProt	<a href="#">UniProtKB/TrEMBL:F9X2A6</a> <a href="#">MYCGM</a>	<a href="#">Mycgr3T103221</a>	<ul style="list-style-type: none"> <li><a href="#">Search BioMart</a></li> <li><a href="#">View on karyotype</a></li> </ul>	
<a href="#">GO:0009082</a>	branched-chain amino acid biosynthetic process	IEA	UniProt	<a href="#">UniProtKB/TrEMBL:F9X2A6</a> <a href="#">MYCGM</a>	<a href="#">Mycgr3T103221</a>	<ul style="list-style-type: none"> <li><a href="#">Search BioMart</a></li> <li><a href="#">View on karyotype</a></li> </ul>	
<a href="#">GO:0009098</a>	leucine biosynthetic process	IEA		<a href="#">InterPro:isopropylmalate_DH</a> , <a href="#">UniProtKB/TrEMBL:F9X2A6</a> <a href="#">MYCGM</a>	<a href="#">Mycgr3T103221</a>	<ul style="list-style-type: none"> <li><a href="#">Search BioMart</a></li> <li><a href="#">View on karyotype</a></li> </ul>	

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](#)

<b>Description</b>	3-isopropylmalate dehydrogenase [Source:UniProtKB/Swiss-Prot;Acc: <a href="#">Q9Y897</a> ]
<b>Location</b>	<a href="#">Chromosome 2: 1,581,329-1,582,830</a> reverse strand. MG2:ACPE01000002.1
<b>About this gene</b>	This gene has 1 transcript ( <a href="#">splice variant</a> ), <a href="#">334 orthologues</a> and <a href="#">7 paralogues</a> .
<b>Transcripts</b>	<a href="#">Show transcript table</a>

Show/hide columns (1 hidden)							Filter	
Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags	
-	<a href="#">Mycgr3T103221</a>	1237	<a href="#">365aa</a>	Protein coding	<a href="#">F9X2A6</a>	-	Ensembl Canonical	

**Ensembl Fungi** | HMMER | BLAST | BioMart | Tools | Downloads | Help & Docs | Blog

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  
**Location** 3-isopropylmalate dehydrogenase [Source:UniProtKB/Swiss-Prot;Acc:Q8Y897.6]

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

About this transcript

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

### Summary ?

< 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](#)

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

Exons/Introns Translated sequence Flanking sequence Intron sequence UTR

Download sequence

Download sequence

Partially coding exon

Introns and flanking sequence in lower case font

No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence						.....ctctccaatccaggttaacaaattctttggccgaacactcaacatacaac
1	Myo93E103221.1	1,582,830	1,582,783	-	1	48	ACCTCAATCTGCGAAAATGCCCACTTACACATCTGCTCTTGAG
	Intron 1-2	1,582,782	1,582,733			50	gtgagtgctgatcgcttcgctccaattggaatcgtgatgtgctg
2	Myo93E103221.2	1,582,732	1,582,713	1	0	20	GTGATCCTCGCGTCTGA
	Intron 2-3	1,582,712	1,582,657			56	gtaaagagctgtcttaagctactca.....agagctcaaaag
3	Myo93E103221.3	1,582,656	1,582,636	0	0	21	GTACCCCGGAGCGGCTGAAG
	Intron 3-4	1,582,712	1,582,657			56	gtaaagagctgtcttaagctactca.....agagctcaaaag
3	Myo93E103221.3	1,582,656	1,582,636	0	0	21	GTACCCCGGAGCGGCTGAAG
	Intron 3-4	1,582,635	1,582,580			56	gtatgctgaacagatattctgtctc.....gaaggaacactgcaaaagctatag
4	Myo93E103221.4	1,582,579	1,582,511	0	0	69	GTCTCGAGCTGATGCACACACACATGCCAGCTTCACTTCAACATCAACACACCTGCTCGCTGC
	Intron 4-5	1,582,510	1,582,460			51	gtacggttgctcgacacacattcttg.....aaagcagaggtgtcgaagcaga
5	Myo93E103221.5	1,582,459	1,582,388	-	+	704	GGCTCTTGATGCCCGCGGCGACCTCTTACAGAGAGAGAGCTCCGCTCGCTCAAAACGCGACGCGCTGATCTTGGAGCGCATCGCGGTGCCAAATGGCGACGCGCAACCTCGTCCAGAGCGCTTCCGCTCGCTGAGAGCTCTTCCATTGAGAGAGAGCTTCCGCGCGGCTCAACTTCAACATGCTGCTGATGCTGACAGCTTGGCTACTTCTGGGAGCGACGAGAGCGGCTTCTGCTACCGATGAGACGAGACCTTCTCGCGCTGATGATCTGCTGCGACAGCTGCGCGGTCTTGGCTTGGCTGAGAGACCGCGCTGCTCGCTGTGGTCTGCAAGAGCGACCTGATGCTGATGCTGACAGCAGCTTGGCGCAGACCGTGAAGGATATGTTGCTGCTCAACGCTTCCGCTAGTTGAGAGATGGAGCTACTTATTTTCCGGGTATGTTTGGCTCAAGACCGGAGCGGTGAGAGGAGCTGATCTGCTCACTAGAGATTTGTTGGGATATCATATGTCAGAGGCGAGCTCATTTCCAGGATCTCTGGGTTTGTGCCAATCCGAGTTTCTGCTCGGATGCGAGAGCGAGCTGCAATTCATCTTACGCGCATTCACG
	Intron 5-6	1,581,735	1,581,684			52	gtatgtaacaattgtgcttgatgag.....gtatgatttgctgctgcaaacatcag
6	Myo93E103221.6	1,581,683	1,581,329	1	-	355	GGCTCTGACAGACATCATCGCGCAGGGTATGCTGCAACCGCTGCCATGATCTTGTCTTGGCATATGTCAGATCTCTCTGACACCGCGAGCTCGGCCAAGAGATGCAAGAGCTCGTAGAGACCTGATGAGAGTATCAACACTTGCATATTTGTTGGTTCGGCAGAGCTTCCGAGGCTGGTGTGATATGTCAGAGAGCTCGAGGCTCTGAGATGATAAAGCTGCAATCTTATCTC.....ACTCTGCTTCCCGAGTGTGCGCAGAGAGTGTGGAGTCCAAACAAAGATTGAGTAGAAGCTTTCTCGGCGCAACATATCTCTCTATAG
	3' downstream sequence						aggaacttctgacgcgcaaatggaacacccctgggtctgatatgacg.....

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

```

1  AACTGAATCGTCGAAAAATGCCCACTTACAACATCGTCGCTTTGGAGGTGATCACTGCG 60
   .....ATGCCCACTTACAACATCGTCGCTTTGGAGGTGATCACTGCG 43
   .....-M--P--T--Y--N--I--V--V--F--G--G--D--H--C-- 14

61  GTCCTGAAGTCACCGCCGAGGCGCTGAAGGTGCTCGACGTGATCGACAACAGCAATGCCG 120
44  GTCCTGAAGTCACCGCCGAGGCGCTGAAGGTGCTCGACGTGATCGACAACAGCAATGCCG 103
15  G--P--E--V--T--A--E--A--L--K--V--L--D--V--I--D--N--S--N--A-- 34

121  ACGTTCACCTTCAACATTCAACCACACCTGCTCGGTGGCGCCTGATCGATGCCACGGCG 180
104  ACGTTCACCTTCAACATTCAACCACACCTGCTCGGTGGCGCCTGATCGATGCCACGGCG 163
35  D--V--H--F--N--I--Q--P--H--L--L--G--G--A--S--I--D--A--H--G-- 54

```

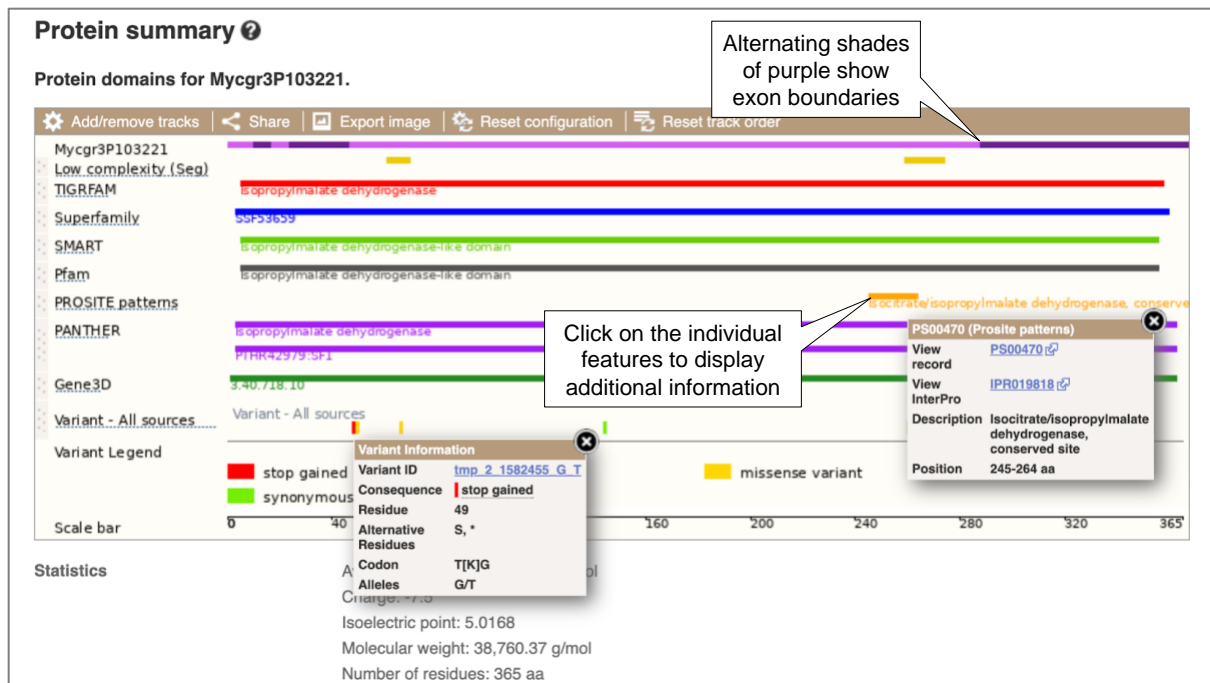
UnTranslated Regions (UTRs) are highlighted in dark yellow, codons are indicated by alternating light yellow highlights, and exon sequences are shown in alternating black and blue letters.

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	<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> [view all locations]
UniParc	<a href="#">UPI000012E54F</a> [view all locations]
UniProtKB/TrEMBL	<a href="#">F9X2A6_MYCGM</a> [align] Putative uncharacterized protein [view all locations]