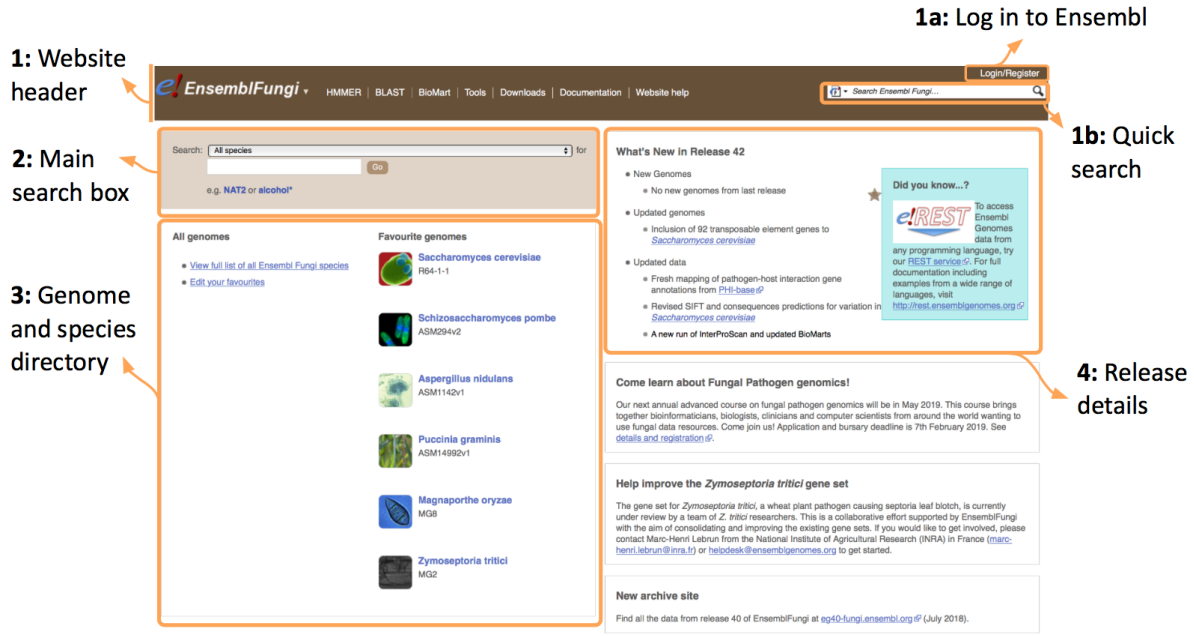
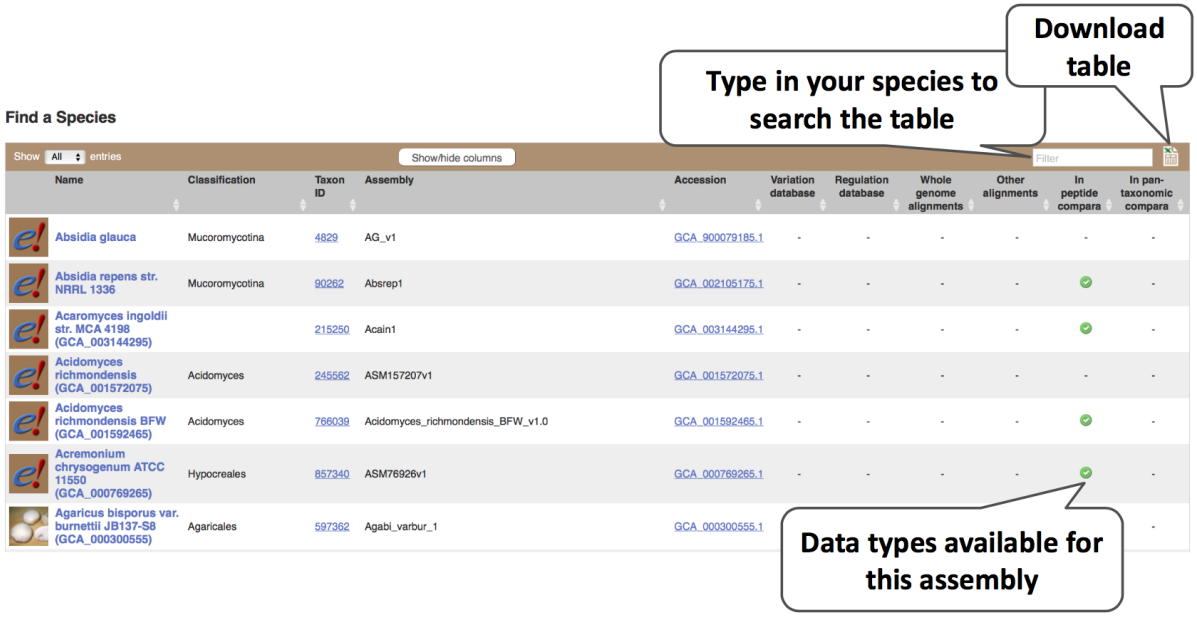
Demo: Ensembl Fungi species

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

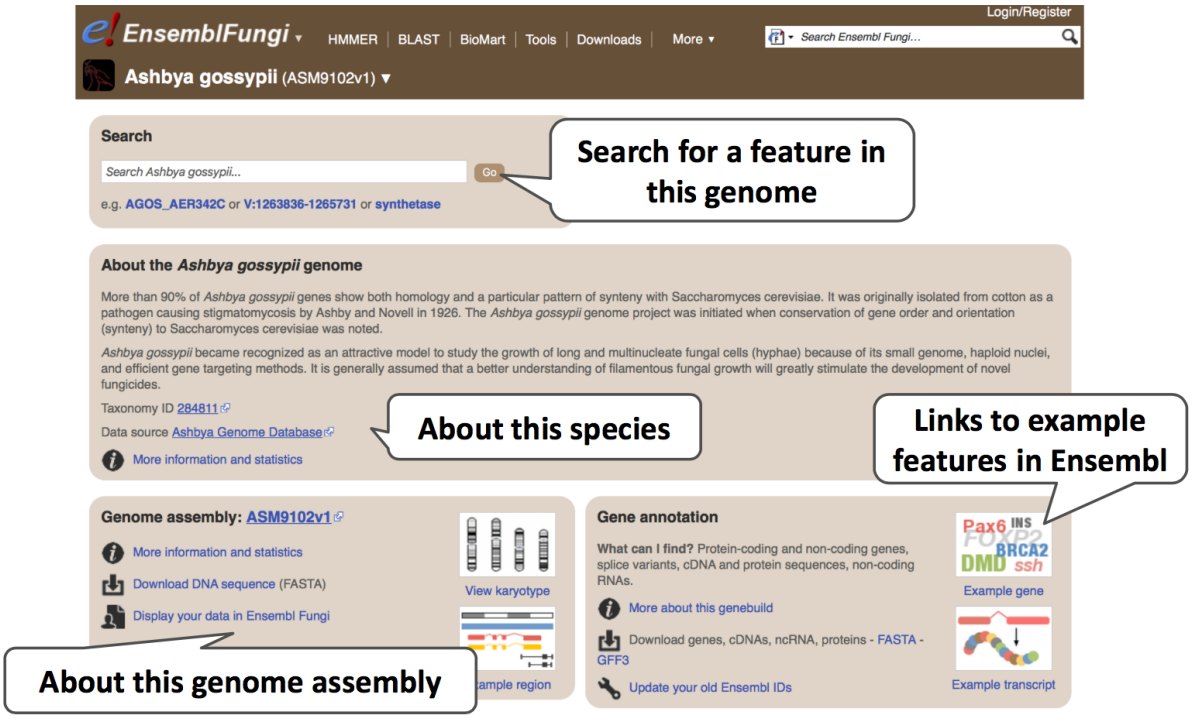
Navigate to fungi.ensembl.org. You'll see a homepage similar to this:



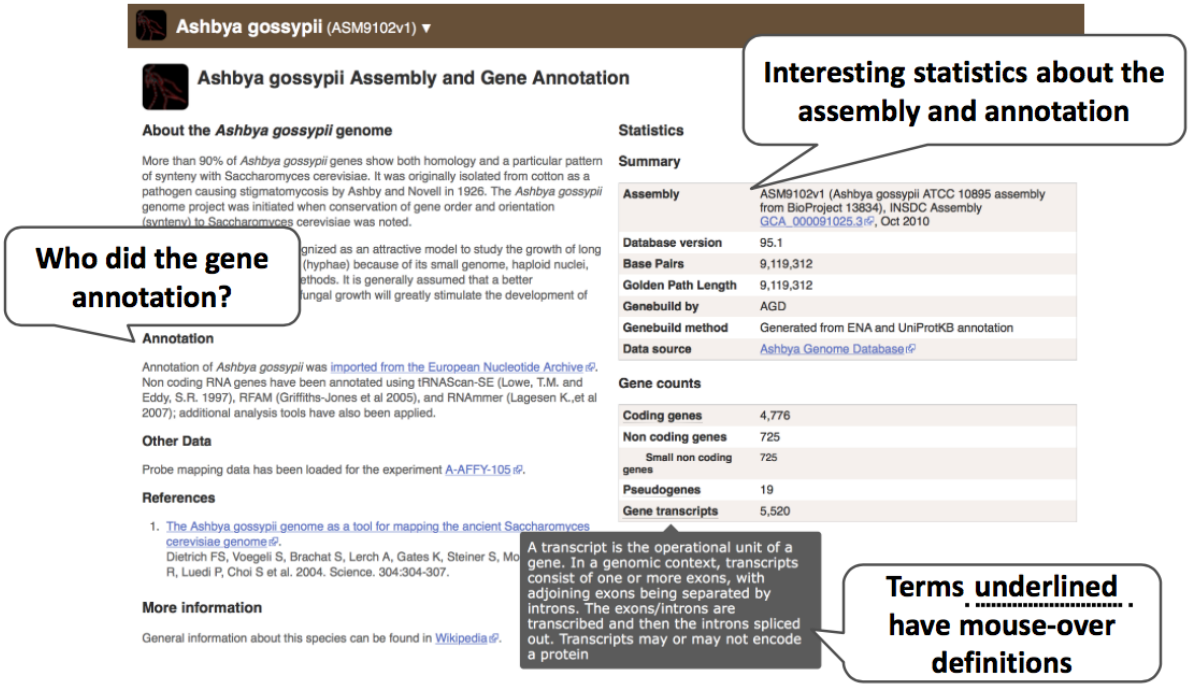
Click on 'View full list of all Ensembl Fungi species', which you can find in section **3: Genome and species directory** shown above.



Click on the latin name of your species of interest to go to the species homepage. We'll click on Ashbya gossypii.



To find out more about the genome assembly and gene annotation, click on More information and statistics.



Exercises: Searching Ensembl Fungi species

Exercise – Ustilago maydis

(a) Navigate to the species homepage for *Ustilago maydis*. What is the name of the genome assembly for *Ustilago maydis*?

(b) Click on More information and statistics. How long is the *Ustilago maydis* genome (in bp)? How many genes have been annotated?

Exercise - *Bipolaris* species

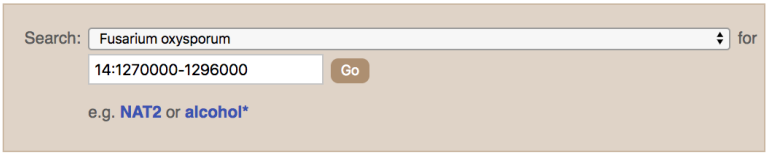
(a) How many genome assemblies are there for the genus *Bipolaris* in Ensembl Fungi?

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

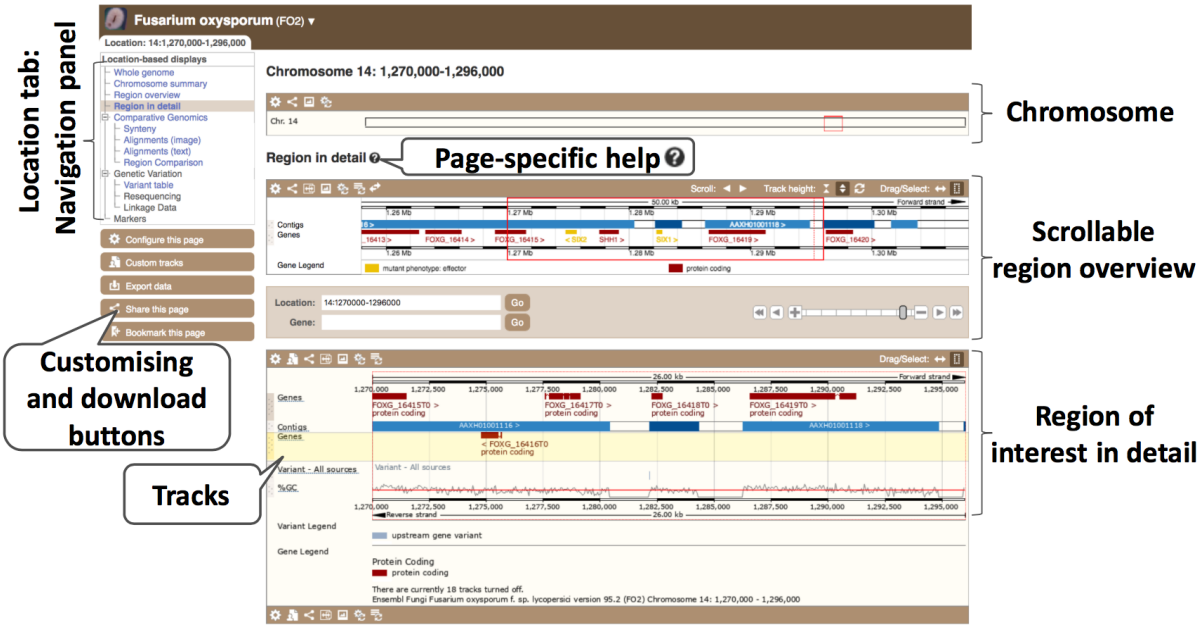
Demo: The Ensembl Fungi Region in detail view

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.

Find *Fusarium oxysporum*, then type (or copy and paste) 14:1270000-1296000 into the search box. Press enter or click Go to jump directly to the **Region in detail** Page.

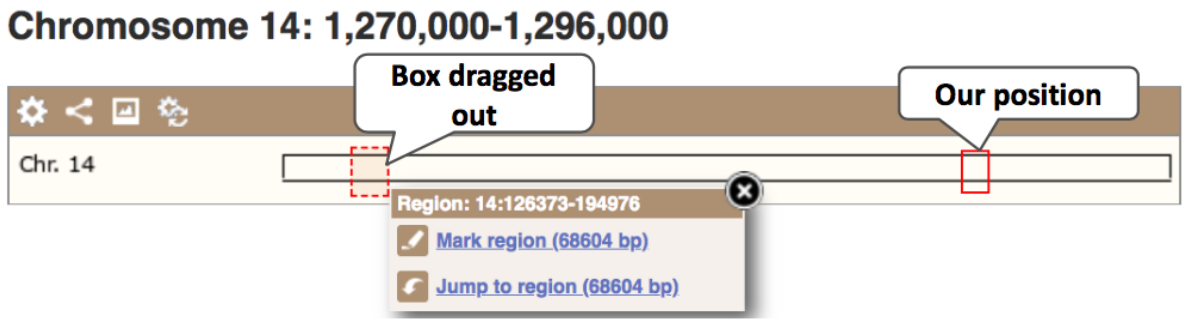


Click on the button to view page-specific help. The help pages provide links to Frequently Asked Questions, a Glossary, Video Tutorials, and a form to Contact HelpDesk. There is a help video on this page at http://youtu.be/tKEvgPUq94.



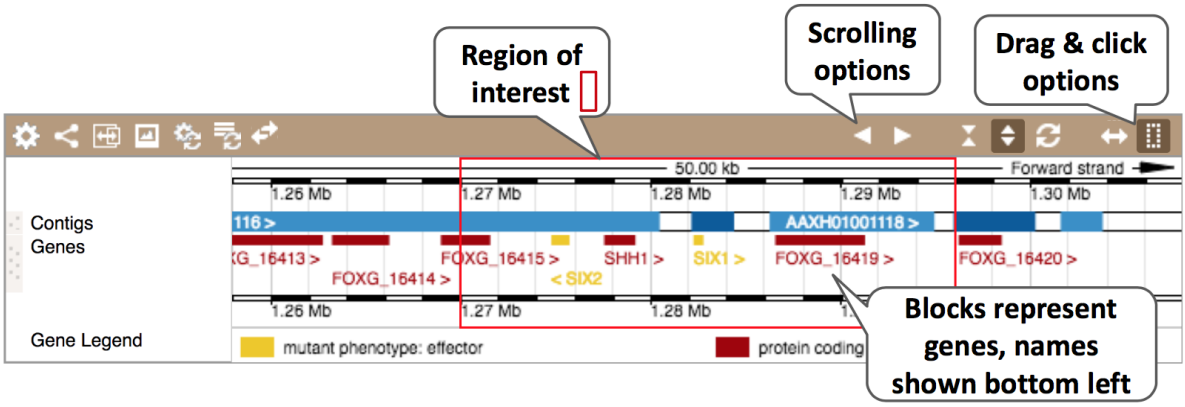
The Region in detail page is made up of three images, let's look at each one on detail.

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.



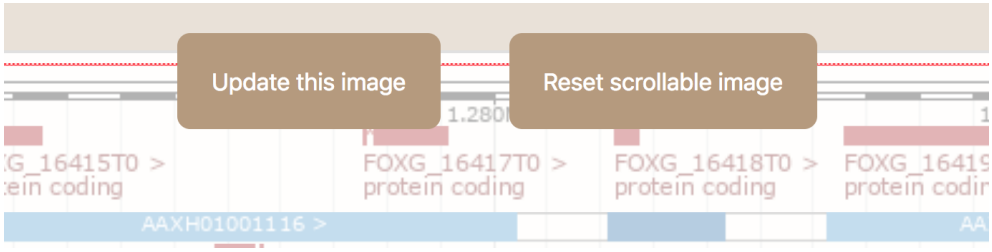
If you would like to move to the region, you could click on Jump to region (### bp). To highlight it, click on Mark region (### bp). For now, we'll close the pop-up by clicking on the X on the corner.

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

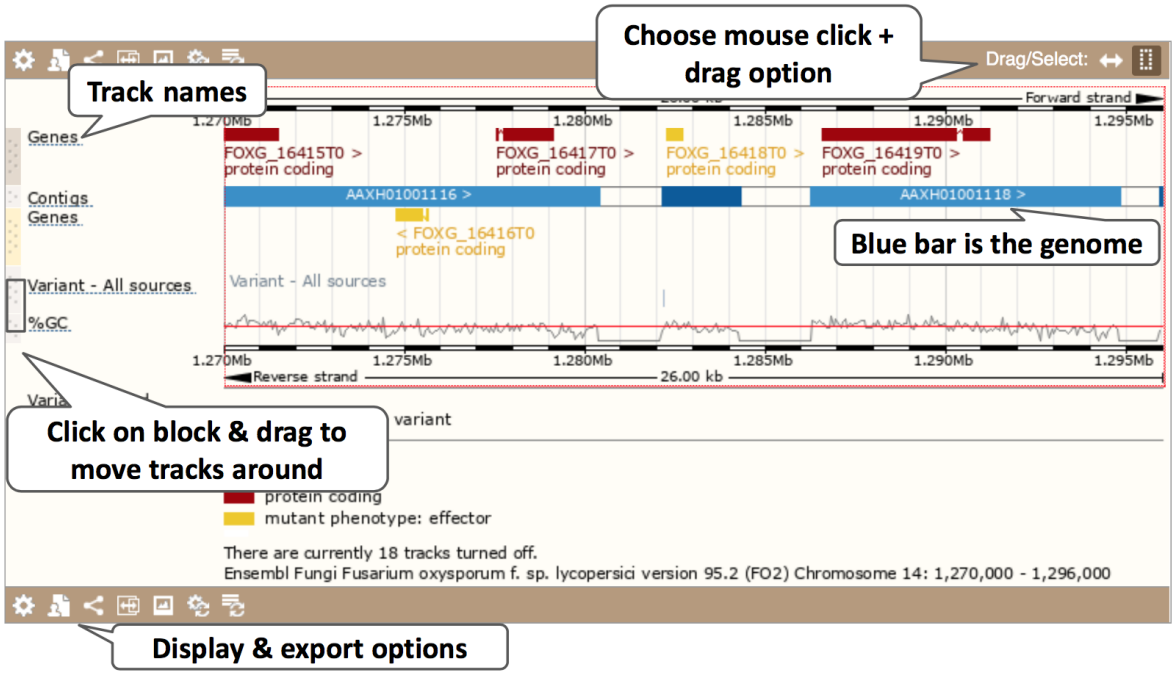


Click on the Drag/Select button 

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.

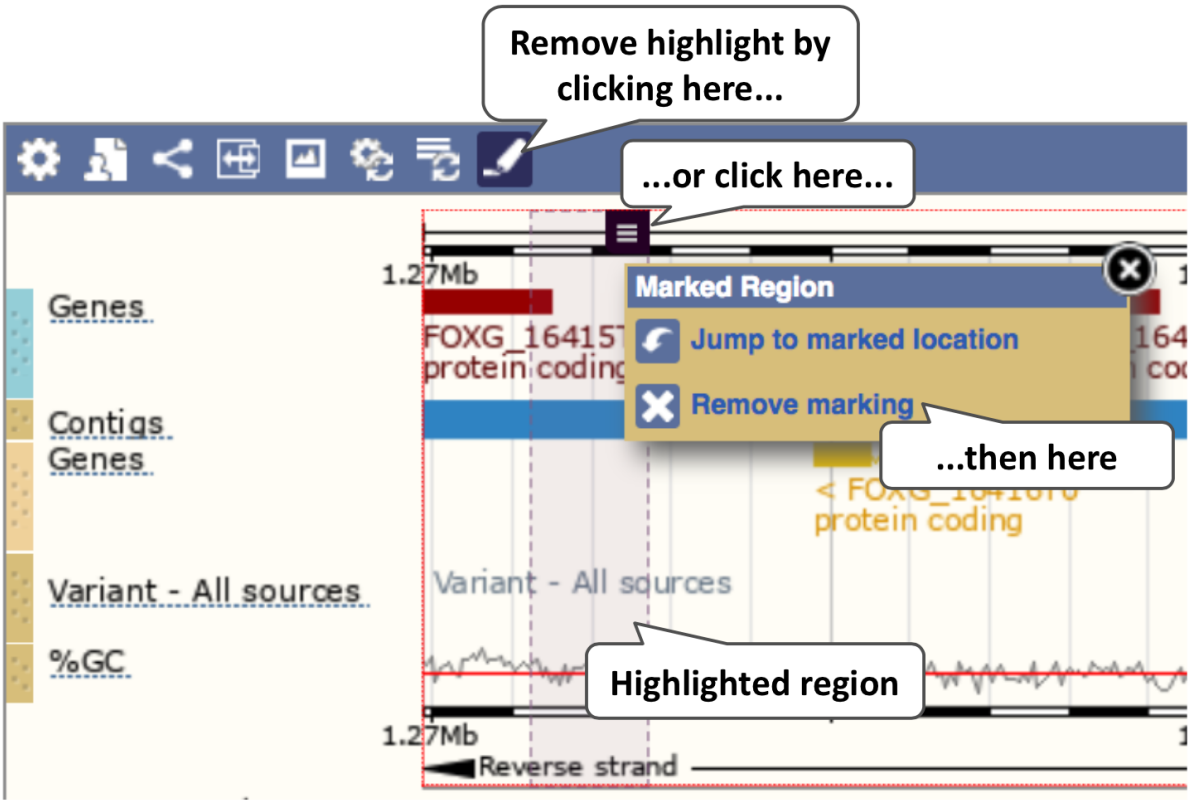


The third image is a detailed, configurable view of the region.



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.

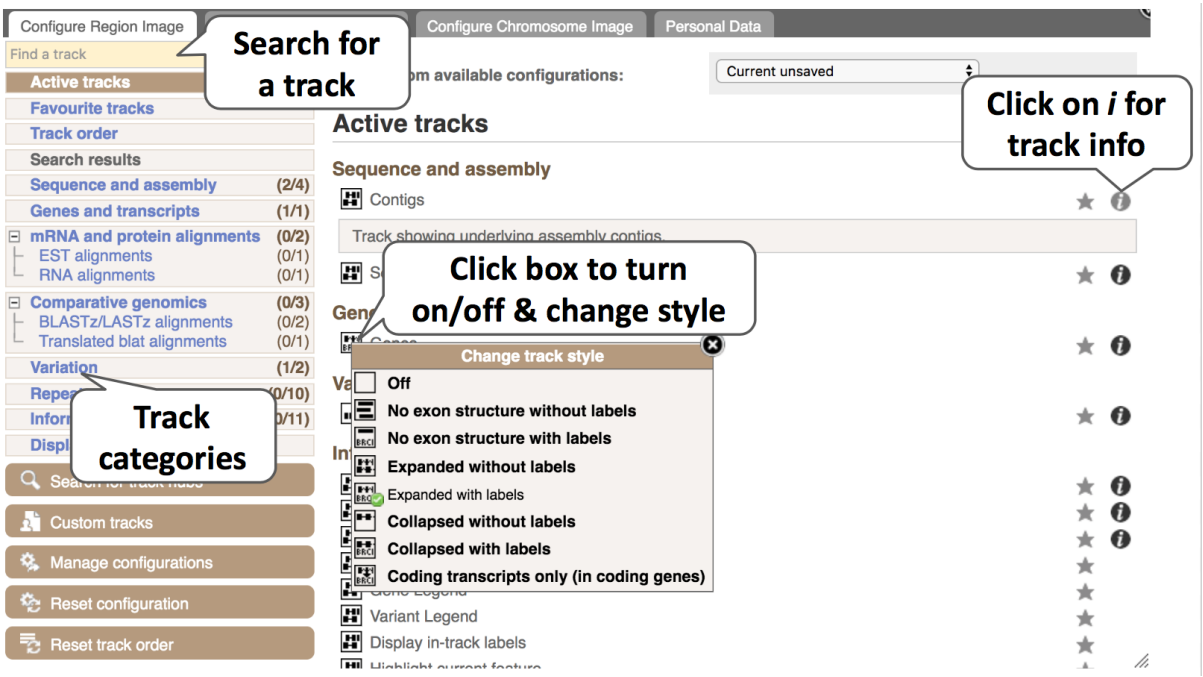
With the tool set to Select, drag out a box around an exon and choose Mark region.



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.

We can edit what we see on this page by clicking on the Configure this page menu at the left.

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.



Let's add some tracks to this image. Add:
Start/stop codons
Type II Transposons
Now click on the tick in the top right hand tick to save and close the menu. Alternatively, click anywhere outside of the menu. We can now see the tracks in the image.

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. 

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.

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.

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

Exercises: Ensembl Fungi Region in Detail

Region Exercise 1 – Exploring a *Coprinopsis cinerea okayama* region

(a) Go to the region 7:1400000-1425000 in *Coprinopsis cinerea okayama*.

(b) How many complete genes are found in this region? How many on the forward and how many on the reverse strand?

(c) Zoom in on the largest gene EFI27358. How many exons does this gene have?

(d) Export the genomic sequence for this region.

Region Exercise 2 – Exploring a region in *Zymoseptoria tritici*

(a) Go to the region 3:1310171-1318171 in *Zymoseptoria tritici* MG2.

(b) How many genes are annotated in this region? Are they all annotated by the same institute?

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

Region Exercise 3 - Exploring a region in *Schizosaccharomyces pombe*

(a) We're exploring the region surrounding the gene Tor1. Search *Schizosaccharomyces pombe* for the region II:3075647-3085541.

(b) Turn on all the tracks for Polyadenylation sites. Which studies show data for the gene Tor1? Click on the track names to see descriptions.

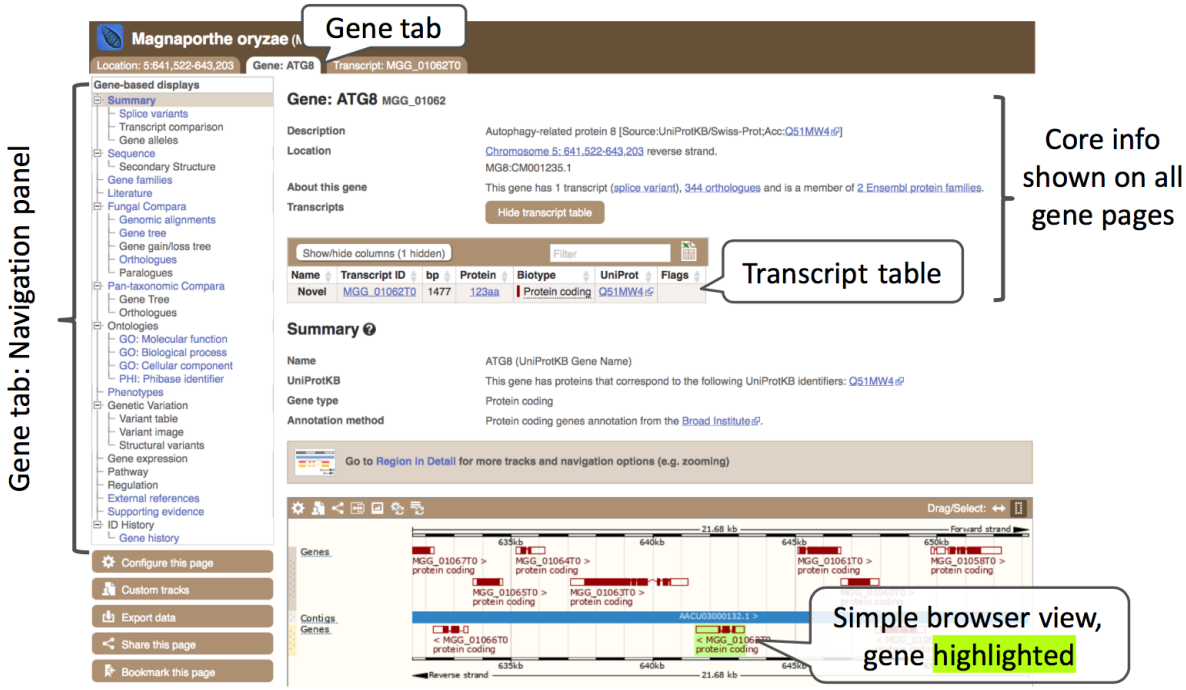
(c) Zoom into the 3' UTR for tor1 to see the peaks. What do you notice about the y-axis scales on the tracks? Can you change the Mata (2013) tracks to have the same Y-axis limits at the Schlackow (2013) tracks? (Hint: Click on the track name and explore the options in the pop-up).

Demo: The Ensembl Fungi gene tab

We're going to look at the gene ATG8 in *Magnaporthe oryzae*. This gene is involved in autophagy, and targeted silencing of this gene inhibits infection (further info in Wilson and Talbot, *Nature Reviews Microbiology* volume 7, pages 185–195 (2009)).

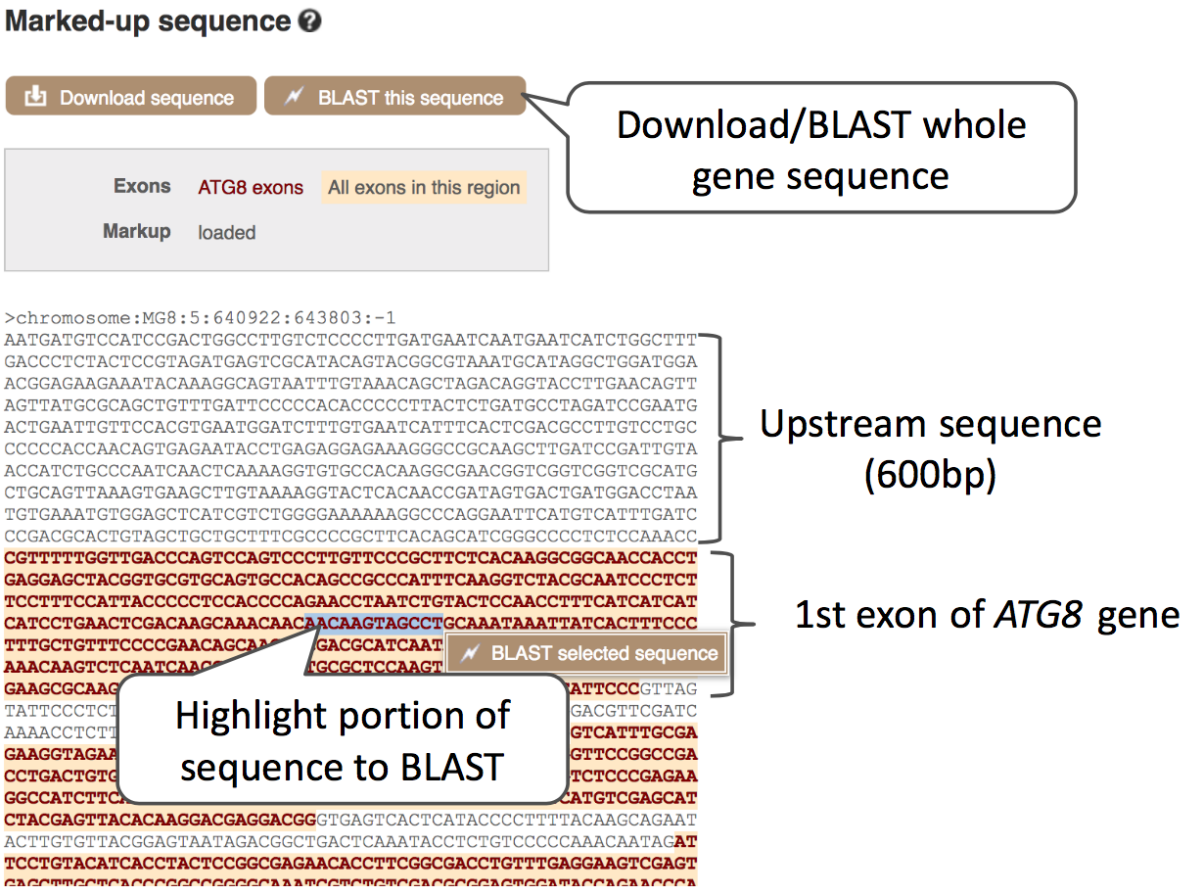
From fungi.ensembl.org, type *ATG8* into the Search for a gene search bar, click the drop-down menu and select Magnaporthe oryzae and click the Go button.

Click on the gene name ATG8 in the results. The **Gene tab** should open:

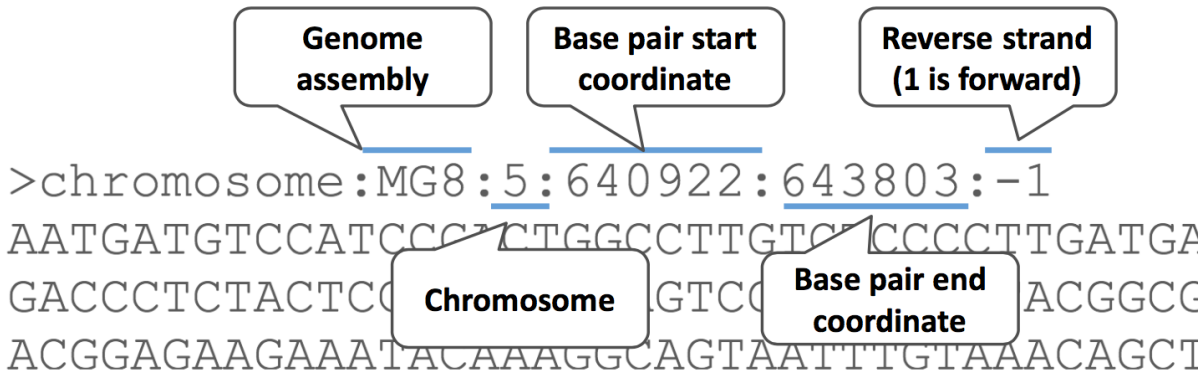


The *ATG8* gene is highlighted in green and in the centre of the display as it is the gene of interest.

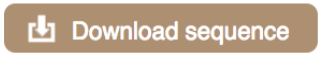
Let's walk through some of the links in the left hand navigation column. How can we view the genomic sequence? Click Sequence at the left of the page.



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

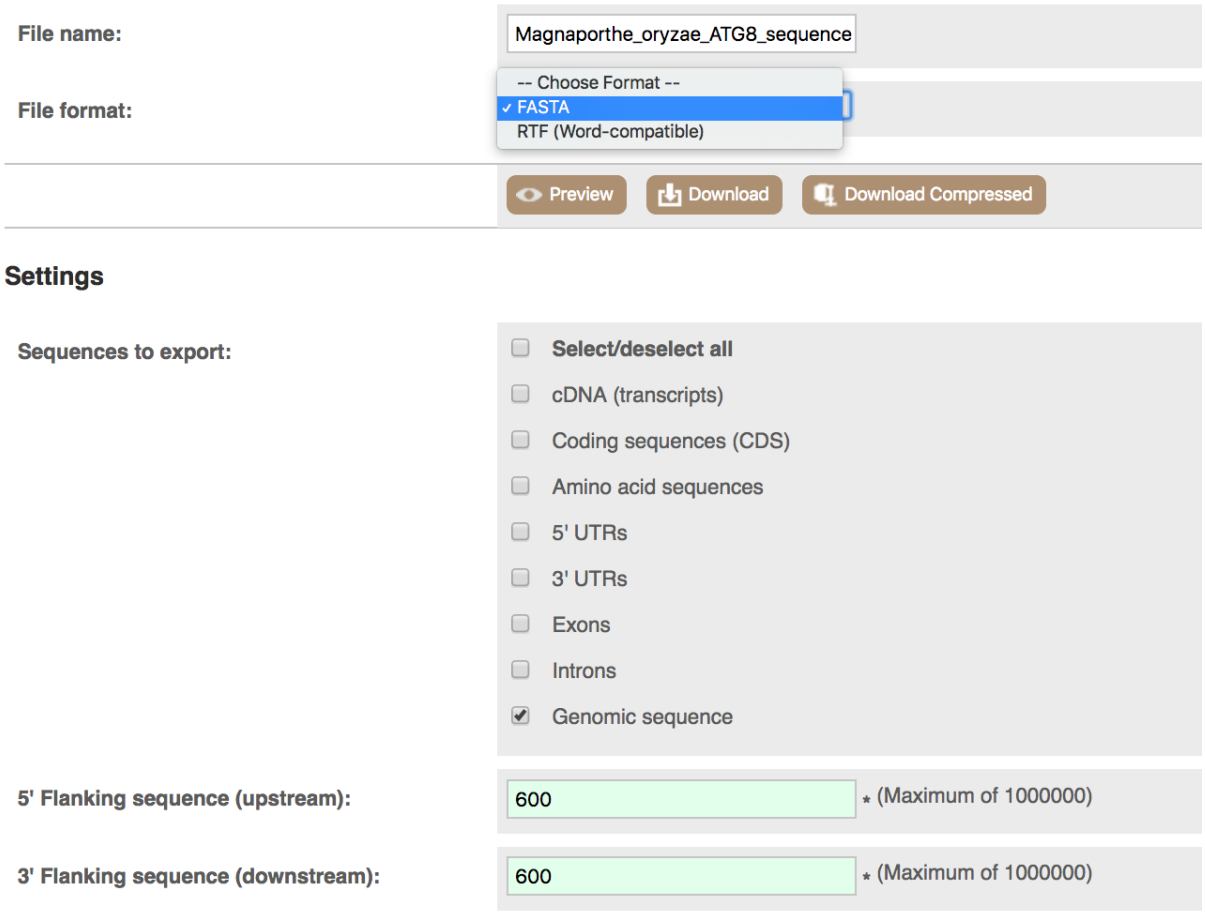


Exons are highlighted within the genomic sequence. If you click on Configure this page you can change display options, and for species with variation databases you can highlight variants on this view.

You can download this sequence by clicking in the 

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.



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 it's molecular function.

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 'Positive regulation of macroautophagy' you will be taken to the GO pages, which shows that this is a child term to 'Autophagy' which is also shown on the GO pages in Ensembl.

GO: Biological process

Accession	Term	Evidence	Annotation source	Mapped using	Transcript IDs	
GO:0006914	autophagy	IEA	UniProtKB/Swiss-Prot:ATG8_MAGO7		MGG_01062T0	• Search BioMart • View on karyotype
GO:0009405	pathogenesis	IMP	UniProtKB/Swiss-Prot:ATG8_MAGO7		MGG_01062T0	• Search BioMart • View on karyotype
GO:0015031	protein transport	IEA	UniProtKB/Swiss-Prot:ATG8_MAGO7		MGG_01062T0	• Search BioMart • View on karyotype
GO:0016239	positive regulation of macroautophagy	IMP	UniProtKB/Swiss-Prot:ATG8_MAGO7		MGG_01062T0	• Search BioMart • View on karyotype
GO:0048102	autophagic cell death	IMP	UniProtKB/Swiss-Prot:ATG8_MAGO7		MGG_01062T0	• Search BioMart • View on karyotype

Inferred from Mutant Phenotype

For some pathogenic species in Ensembl Fungi we have Pathogen-Host Interactions (PHI-base) annotations. Click on the PHI: Phibase identifier link in the left-hand menu.

There are four results 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	• Search BioMart • View on karyotype
PHI:2076	2076	ND	Sequence		MGG_01062T0	• Search
PHI:2139	2139					
PHI:768	768					

Pathogen Gene	Mutant Phenotype	Pathogen Species	Disease	Host Species
Gph1	loss of pathogenicity	Magnaporthe oryzae	Rice Blast	Hordeum vulgare (related: barley)
ATG8	loss of pathogenicity	Magnaporthe oryzae	Rice Blast	Hordeum vulgare (related: barley)

Pathogen Gene | Allele
Gene:ATG8
PHI-base entry:PHI:2061
Gene ID:XP_368182
Protein ID: Q51MW4
Sequence strain:70-15

Pathogen
Pathogen species:Magnaporthe oryzae
Pathogen ID: 318829
Pathogen strain:B157

Host
Host species:Hordeum vulgare (related: barley)
Host classification:Monocots
Host ID: 4513
Host strain:subsp. vulgare (related: domesticated barley)
Tissue:leaf

Reference | Comments
Pmid:19115483
Ref source:Pubmed
Year:2009
Author reference:Yi Zhen Deng

PHI Phenotype | Pathogen Phenotype
Phenotype:loss of pathogenicity
Disease name:Rice Blast
Tissue:leaf
Host response:slight induction of hypersensitive reaction
Experimental technique:Gene deletion: full

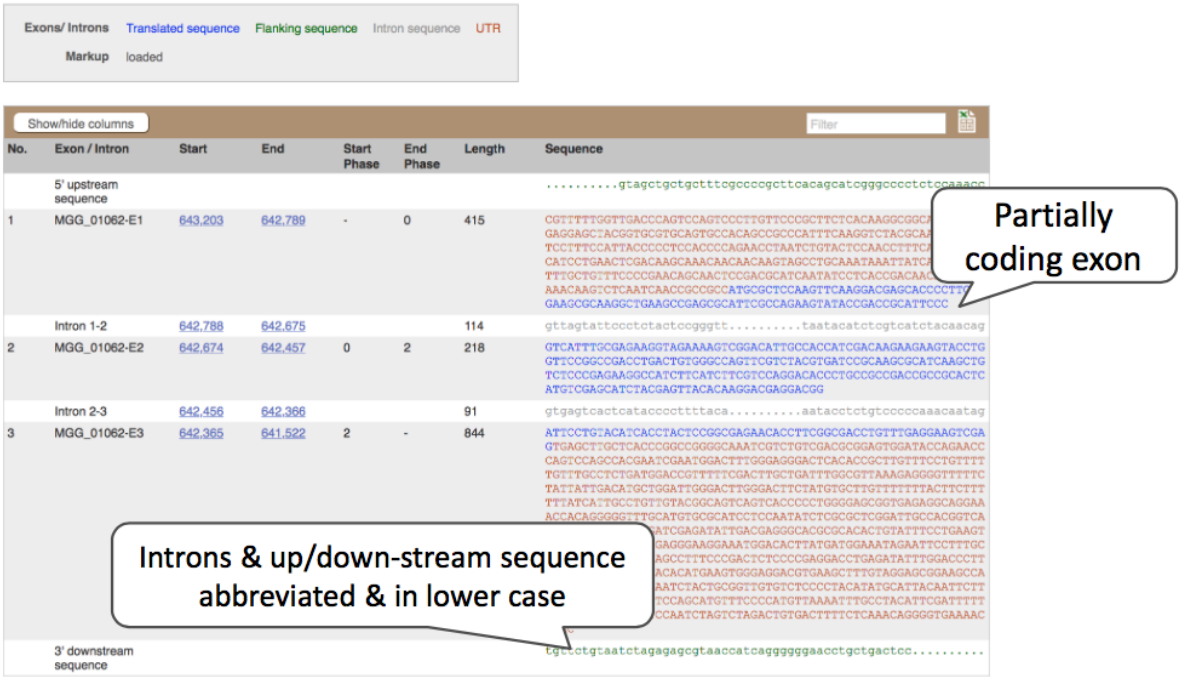
ATG8	loss of pathogenicity	Magnaporthe oryzae	Rice Blast	Hordeum vulgare (related: barley)

Demo: The Ensembl Fungi transcript tab

Many genes have multiple transcripts which can be seen in the transcript table. Click on Show transcript table .

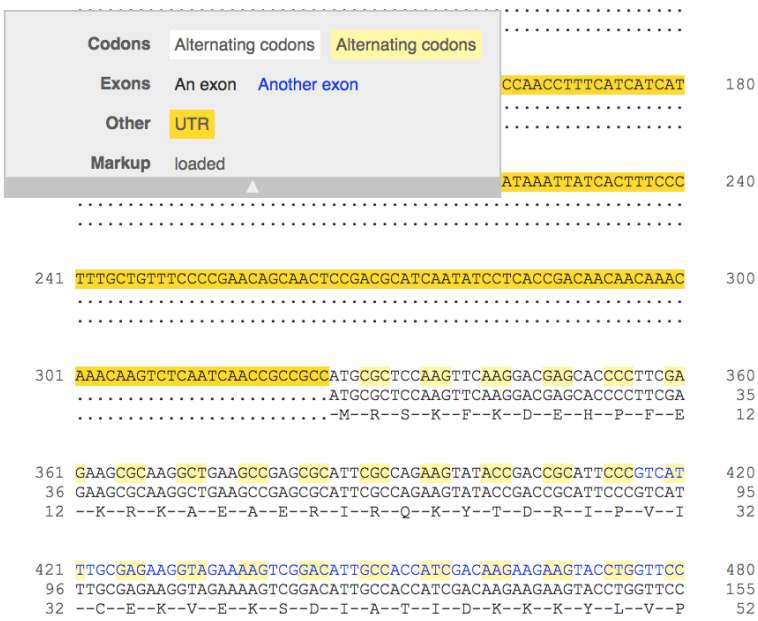
We can go to the transcript tab either by clicking on the transcript ID, MGG_01062T0, in the table, or on the transcript tab at the top of the page. You are now in the Transcript tab on the summary page. Some summary information about the number of exons, length etc is shown at the bottom of the page under the diagram.

The left hand navigation column provides several options for the transcript. Click on the Exons link.



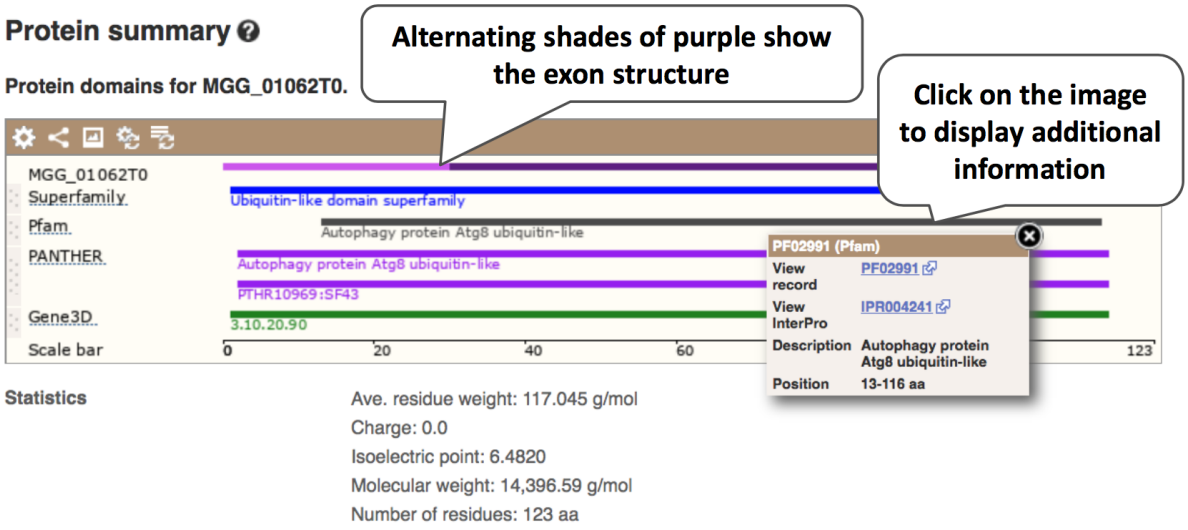
You may want to change the display (for example, to show more flanking sequence, or to show full introns). In order to do so click on Configure this page and change the display options accordingly.

Now click on the cDNA link to see the spliced transcript sequence.



UnTranslated Regions (UTRs) are highlighted in dark yellow, codons are highlighted in light yellow, 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.



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:

Filter

External database	Database identifier
European Nucleotide Archive	CH476836 [align] [view all locations] CM001235 [align] [view all locations]
INSDC protein ID	EDK02260.1 [align] [view all locations] EHA48293.1 [align] [view all locations]
RefSeq mRNA predicted	XM_368182.1 [Target %id: 25; Query %id: 100] [align] 70-15 autophagy protein 8 (MGG_01062) partial mRNA [view all locations]
RefSeq peptide predicted	XP_368182.1 [Target %id: 100; Query %id: 100] [align] autophagy protein 8 [view all locations]
UniParc	UPI000021B9B9 [view all locations]
UniProtKB/Swiss-Prot	ATG8_MAGO7 [Target %id: 100; Query %id: 100] [align] Autophagy-related protein 8 [view all locations]

Exercises: Ensembl Fungi Genes and transcripts

Exercise – Exploring the *Zymoseptoria tritici LEUC* gene

(a) Search Ensembl Fungi for the *LEUC* gene in *Zymoseptoria tritici* MG2. On which chromosome and which strand of the genome is this gene located?

(b) What biological processes have been associated with LEUC?

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

- Are all exons shown in this display part of the LEUC gene? How can you tell?
- Can you find the Stop Gained mutation? What letter is it represented by, looking at the letters and alleles of these surrounding variants what do you think this letter represents (these are IUPAC ambiguity codes)
- Which exon does the stop gained mutation fall in?
- Export this sequence in RTF format

(d) Click to go to the transcript tab by clicking on the transcript ID Mycgr3T103221, and click to view the Protein summary page.

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

Exercise – Exploring a *Trichoderma reesei* gene

Find the *Trichoderma reesei* genome and search for the gene TRIREDRAFT_5868.

(a) What are the molecular functions of this gene?

(b) Go to the transcript tab for the transcript. How many exons does it have? Which one is the longest?

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