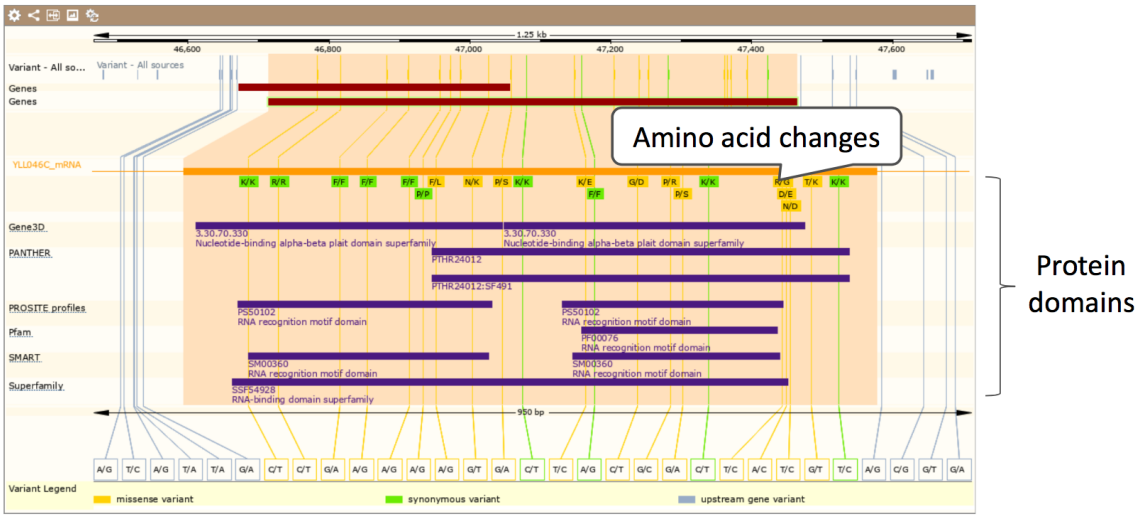
Demo: Exploring variants in Ensembl Fungi

In any of the sequence views shown in the Gene and Transcript tabs, you can view variants on the sequence. You can do this by clicking on Configure this page from any of these views.

Let's take a look at the Gene sequence view for *YLL046C*. This gene is a ribonuclease protein in *Saccharomyces cerevisiae*. Select Saccharomyces cerevisiae on the Ensembl Fungi homepage, search for YLL046C and go to the Variant image view.

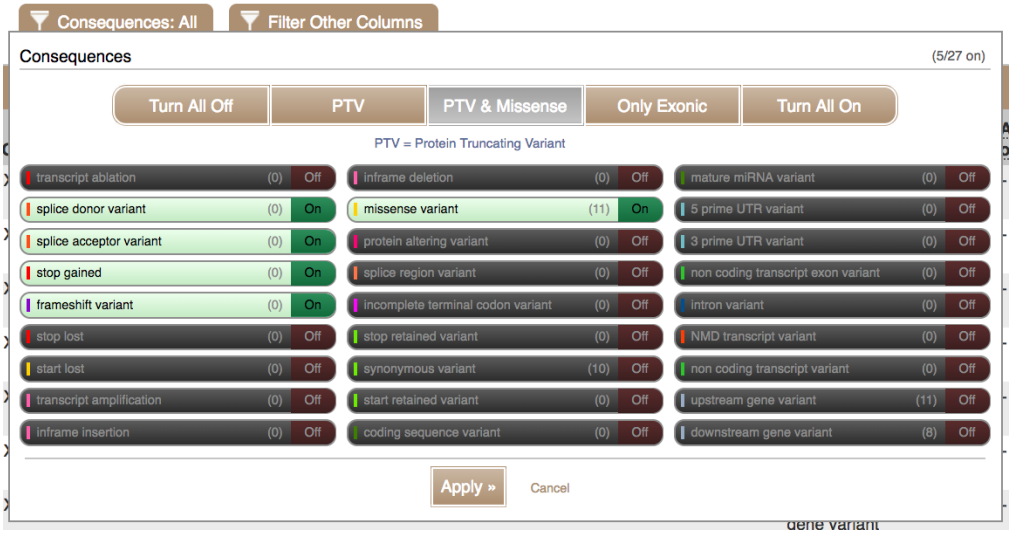
This view shows variants mapped to the gene structure and protein domains.

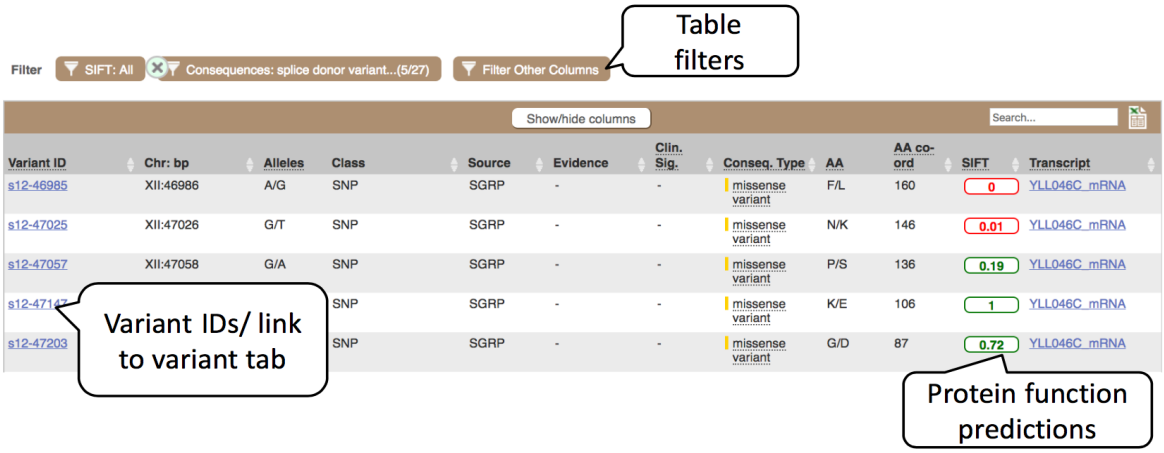


We can examine all variants and filter to see ones we are interested in using the variant table. Click on the Variant table link.

This table shows the variants in order of their occurence through the genome, and they are reported on the forward strand. The gene YLL046C is located on the reverse strand, so we are first shown variants downstream of the gene (starting a the 3' downstream region).

Let's filter the table to view variants that alter the protein sequence. Click on the Consequences:All button above the table. Click the option 'PTV and Missense' in the pop-up, click Apply. You can also filter by other columns such as variant Class.

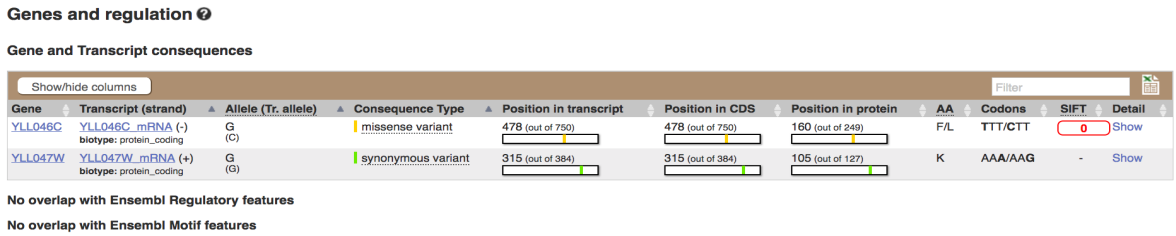




The SIFT scores predict the consequence of the variant on the function of the protein taking into account chemical changes and conservation of amino acids. Scores <0.05 and coloured red are ‘deleterious’ while scores >0.05 and coloured green are tolerated.

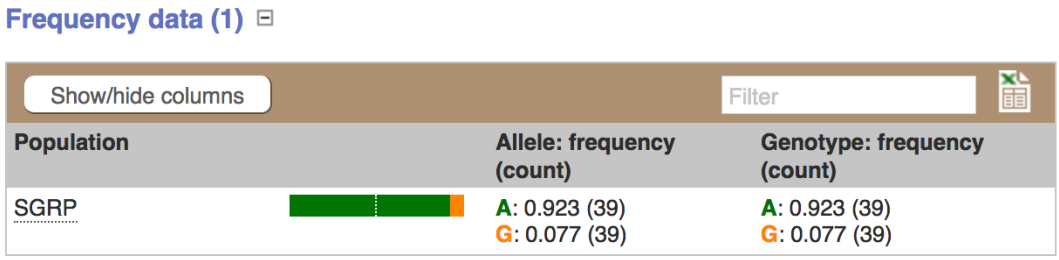
Let’s have a look at a specific variant. Click on the top result in the filtered table, or search for s12-46985. This will open up the variation tab.

The icons show you what information is available for this variant. Click on Genes and regulation, or follow the link at the left.



This variant overlaps two genes. It causes a change in the protein sequence (missense variant) in the YLL046C gene we were looking at, but it is a synonymous variant in the other gene, not causing a change in the amino acid. Only missense variants have SIFT scores, hence why this is missing for the other gene.

Let's look at population genetics. Either click on Genotype frequency in the left-hand menu.



Note that the frequencies reported here are correct but the counts (in brackets) are not!

Exercises: Exploring variants in Ensembl Fungi

Exercise – Missense variants in *Zymoseptoria tritici*

(a) View all the variants in the gene CYP-74. How many are predicted to be missense?

(b) Click on tmp_7_1448810_A_G to go to the variant tab. What is the genotype at this locus in the SRS383147.bam sample?

Exercise – Variation data in *Fusarium oxysporum*

(a) Select the Fusarium oxysporum genome and search for FOXG_13574T0 gene. One of its upstream variants is SNP tmp_10_6610. What are the possible alleles for this polymorphic position? Which one is on the reference genome?

(b) What is the most frequent allele at this position?

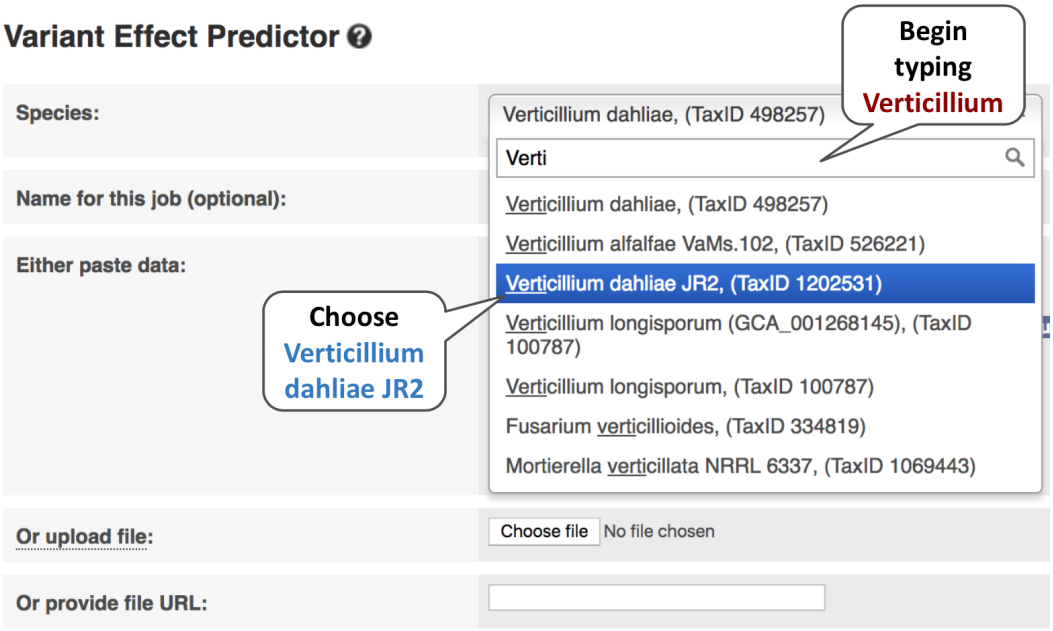
(c) Which individuals have got genotypes C|T and T|T?

Demo: The Ensembl Fungi Variant Effect Predictor (VEP)

We have identified four variants in *Verticillium dahliae* JR2: chromosome 5, C->G at 698711, G->T at 698935, G->A at 700313 and C->A at 701484.

We will use the Ensembl VEP to determine:
Have my variants already been annotated in Ensembl?
What genes are affected by my variants?
Do any of my variants affect gene regulation?

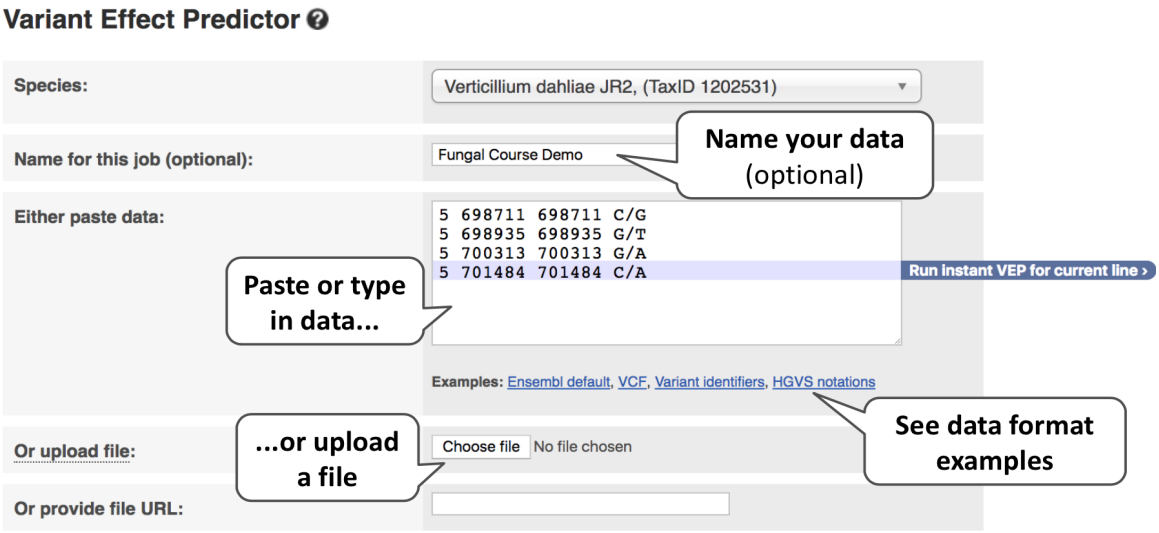
Click on Tools in the top brown bar from any Ensembl Fungi page, then Variant Effect Predictor to open the input form.



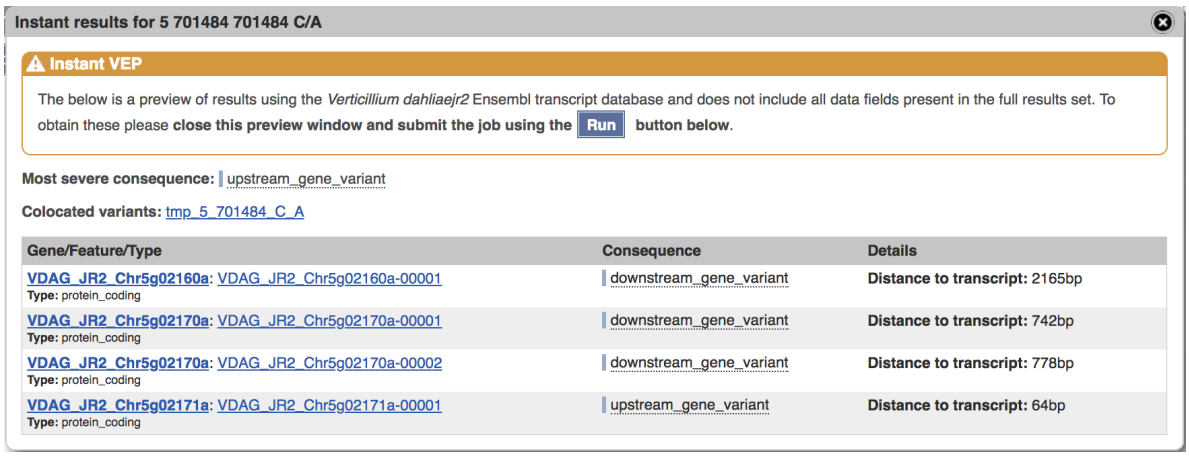
The data is in the format:
Chromosome Start End alleles (reference/mutation) strand name

Put the following into the Paste data box:
5 698711 698711 C/G
5 698935 698935 G/T
5 700313 700313 G/A
5 701484 701484 C/A

The VEP will automatically detect that the data is in Ensembl default format.

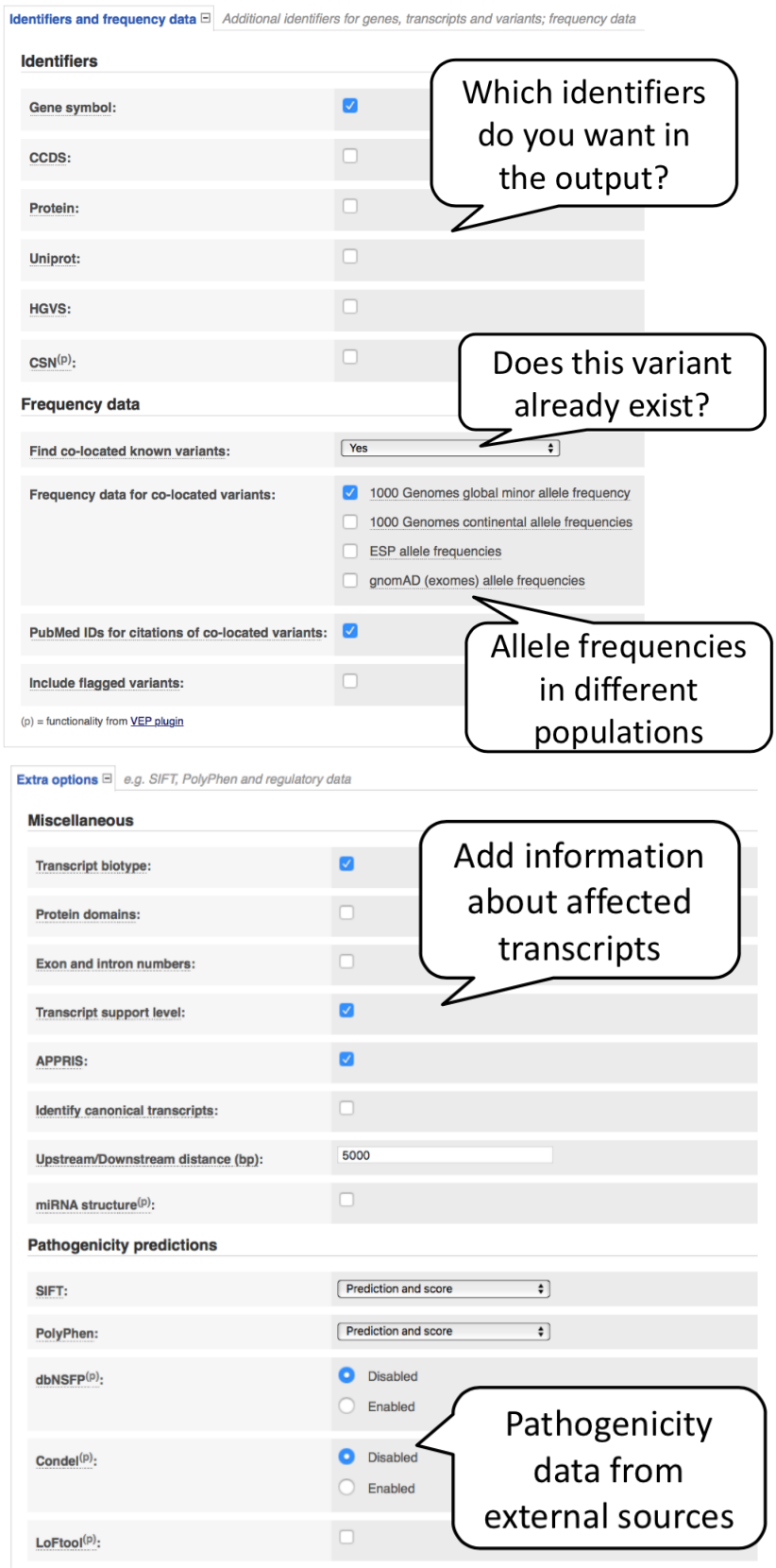


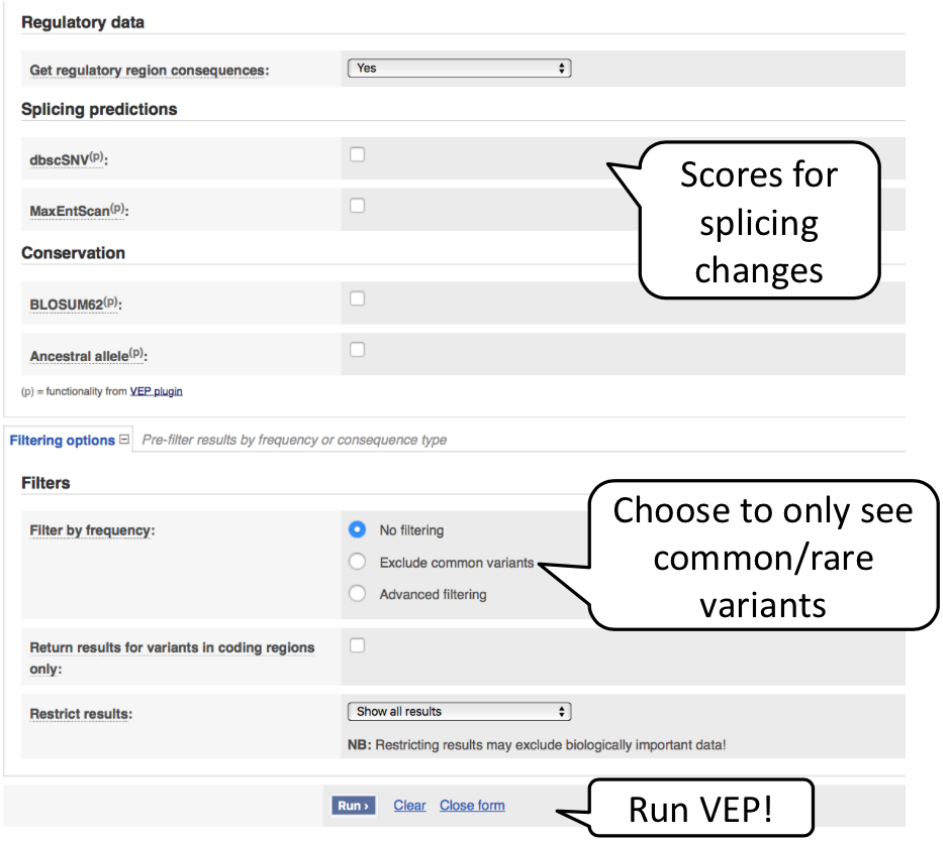
Clicking on the 'Run instant VEP for current line' will generate a pop-up with summarised results for that individual variant.



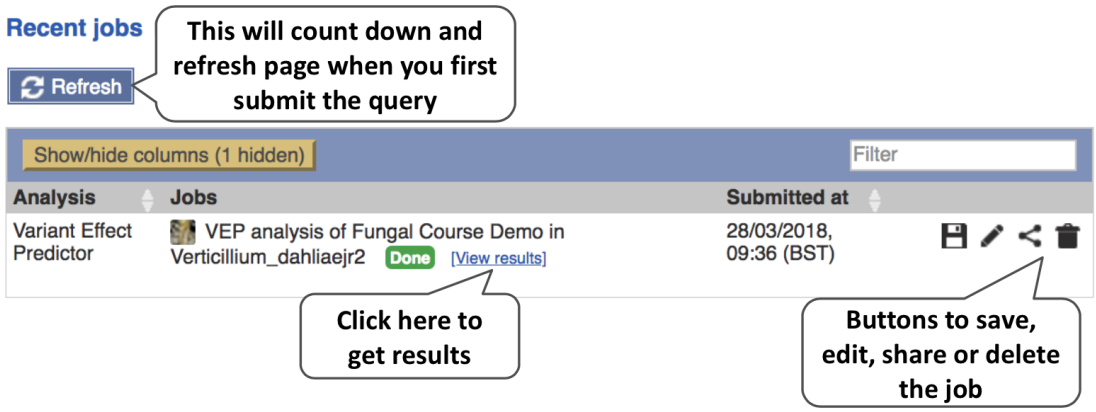
There are further options that you can choose for your output. These are categorised as Identifiers and frequency data, Filtering options and Extra options. Let's open all the menus and take a look.

Please note that the options displayed are the same across all species, including human, so a number of the options may not be available for fungi.


Identifiers and frequency data
Additional identifiers for genes, transcripts and variants; frequency data
Identifiers
Gene symbol:
CCDS:
Protein:
Uniprot:
HGVS:
CSN(p):
Which identifiers do you want in the output?
Frequency data
Find co-located known variants:
Yes
Does this variant already exist?
Frequency data for co-located variants:
1000 Genomes global minor allele frequency
1000 Genomes continental allele frequencies
ESP allele frequencies
gnomAD (exomes) allele frequencies
PubMed IDs for citations of co-located variants:
Include flagged variants:
Allele frequencies in different populations
(p) = functionality from VEP plugin
Extra options
e.g. SIFT, PolyPhen and regulatory data
Miscellaneous
Transcript biotype:
Protein domains:
Exon and intron numbers:
Transcript support level:
APPRIS:
Identify canonical transcripts:
Upstream/Downstream distance (bp):
5000
miRNA structure(p):
Add information about affected transcripts
Pathogenicity predictions
SIFT:
Prediction and score
PolyPhen:
Prediction and score
dbNSFP(p):
Disabled
Enabled
Condel(p):
Disabled
Enabled
LoFtool(p):
Pathogenicity data from external sources



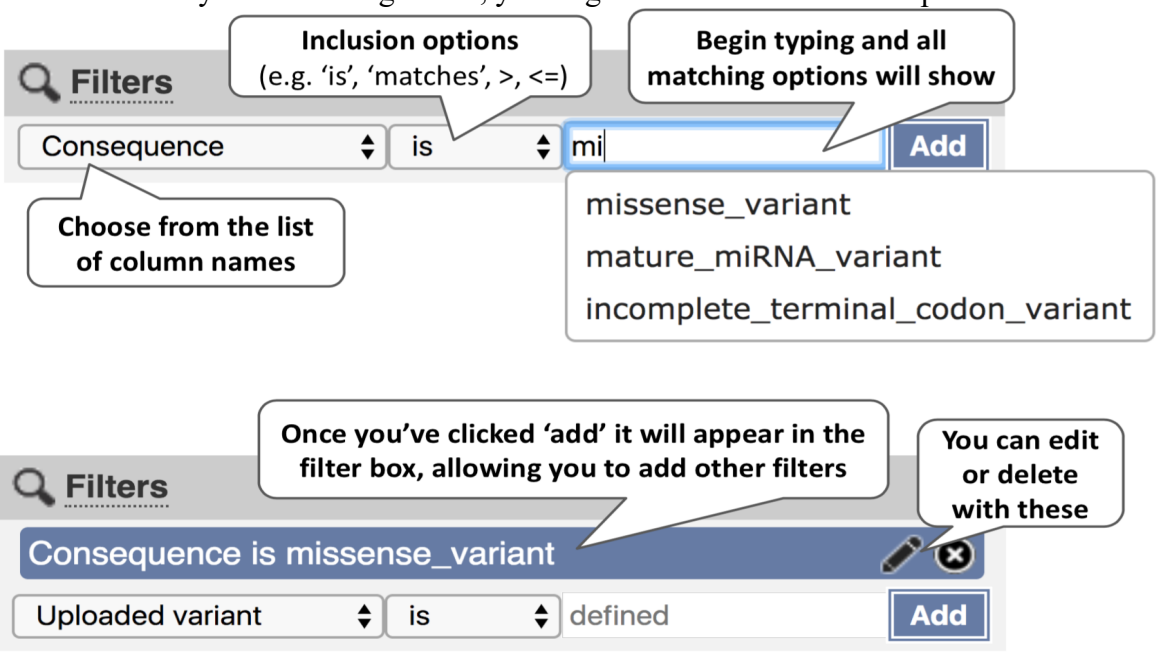
Hover over the options to see definitions. When you've selected everything you need, scroll right to the bottom and click Run.

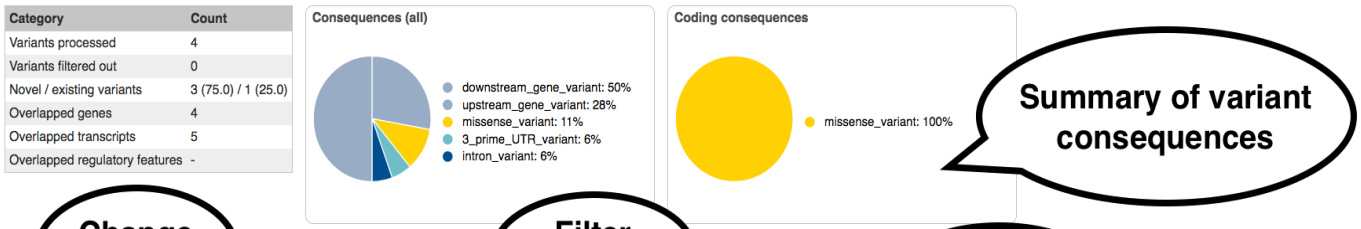


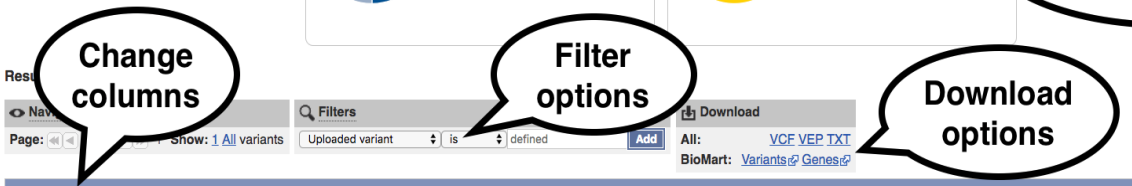
The display will show you the status of your job. It will say Queued, then automatically switch to Done when the job is done, you do not need to refresh the page. You can edit or discard your job at this time. If you have submitted multiple jobs, they will all appear here.

Click View results once your job is done. In your results you will see a graphical summary of your data, as well as a table of your results.

If you have a large table, you might want to use the filter option.







Show/hide columns

Uploaded variant	Location	Allele	Consequence	Impact	Symbol	Gene	Feature type	Feature	Biotype	Exon	Intron	HGVSc	HGVSp	cDNA position	CDS position	Protein position	Amino acids	Codons	Existing variant	Dis to tra
5_698711_C/G	5:698711-698711	G	downstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02150a	Transcript	VDAG_JR2_Chr5g02150a-00001	protein_coding	-	-	-	-	-	-	-	-	-	-	46
5_698711_C/G	5:698711-698711	G	intron_variant	MODIFIER	-	VDAG_JR2_Chr5g02160a	Transcript	VDAG_JR2_Chr5g02160a-00001	protein_coding	-	7/7	-	-	-	-	-	-	-	-	-
5_698711_C/G	5:698711-698711	G	upstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02170a	Transcript	VDAG_JR2_Chr5g02170a-00001	protein_coding	-	-	-	-	-	-	-	-	-	-	13
5_698711_C/G	5:698711-698711	G	upstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02170a	Transcript	VDAG_JR2_Chr5g02170a-00002	protein_coding	-	-	-	-	-	-	-	-	-	-	13
5_698711_C/G	5:698711-698711	G	downstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02171a	Transcript	VDAG_JR2_Chr5g02171a-00001	protein_coding	-	-	-	-	-	-	-	-	-	-	19
5_698935_G/T	5:698935-698935	T	downstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02150a	Transcript	VDAG_JR2_Chr5g02150a-00001	protein_coding	-	-	-	-	-	-	-	-	-	-	48
5_698935_G/T	5:698935-698935	T	3_prime_UTR_variant	MODIFIER	-	VDAG_JR2_Chr5g02160a	Transcript	VDAG_JR2_Chr5g02160a-00001	protein_coding	8/8	-	-	-	1679	-	-	-	-	-	-
5_698935_G/T	5:698935-698935	T	upstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02170a	Transcript	VDAG_JR2_Chr5g02170a-00001	protein_coding	-	-	-	-	-	-	-	-	-	-	11
5_698935_G/T	5:698935-698935	T	upstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02170a	Transcript	VDAG_JR2_Chr5g02170a-00002	protein_coding	-	-	-	-	-	-	-	-	-	-	11
5_698935_G/T	5:698935-698935	T	downstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02171a	Transcript	VDAG_JR2_Chr5g02171a-00001	protein_coding	-	-	-	-	-	-	-	-	-	-	17
5_700313_G/A	5:700313-700313	A	downstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02160a	Transcript	VDAG_JR2_Chr5g02160a-00001	protein_coding	-	-	-	-	-	-	-	-	-	-	99
5_700313_G/A	5:700313-700313	A	missense_variant	MODERATE	-	VDAG_JR2_Chr5g02170a	Transcript	VDAG_JR2_Chr5g02170a-00001	protein_coding	2/2	-	-	-	155	52	18	A/T	GCC/ACC	-	-
5_700313_G/A	5:700313-700313	A	missense_variant	MODERATE	-	VDAG_JR2_Chr5g02170a	Transcript	VDAG_JR2_Chr5g02170a-00002	protein_coding	2/2	-	-	-	161	52	18	A/T	GCC/ACC	-	-
5_700313_G/A	5:700313-700313	A	downstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02171a	Transcript	VDAG_JR2_Chr5g02171a-00001	protein_coding	-	-	-	-	-	-	-	-	-	-	34
5_701484_C/A	5:701484-701484	A	downstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02160a	Transcript	VDAG_JR2_Chr5g02160a-00001	protein_coding	-	-	-	-	-	-	-	-	-	tmp_5_701484_C_A	21
5_701484_C/A	5:701484-701484	A	downstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02170a	Transcript	VDAG_JR2_Chr5g02170a-00001	protein_coding	-	-	-	-	-	-	-	-	-	tmp_5_701484_C_A	74
5_701484_C/A	5:701484-701484	A	downstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02170a	Transcript	VDAG_JR2_Chr5g02170a-00002	protein_coding	-	-	-	-	-	-	-	-	-	tmp_5_701484_C_A	77
5_701484_C/A	5:701484-701484	A	upstream_gene_variant	MODIFIER	-	VDAG_JR2_Chr5g02171a	Transcript	VDAG_JR2_Chr5g02171a-00001	protein_coding	-	-	-	-	-	-	-	-	-	tmp_5_701484_C_A	64

Existing variants

Exercise: The Ensembl Fungi Variant Effect Predictor (VEP)

On the course file page, you will find a VCF file labelled VEP_exercise.vcf. This is a small subset of the outcome of *Puccinia graminis Ug99* whole genome sequencing and variant calling experiment.

Run the file through the VEP and determine:

(a) How many genes and transcripts are overlapped by variants in this file?

(b) Do any of the variants change the amino acid sequences of any proteins? What genes? What is the amino acid change? (Hint: use the filters above the table to filter by consequences)