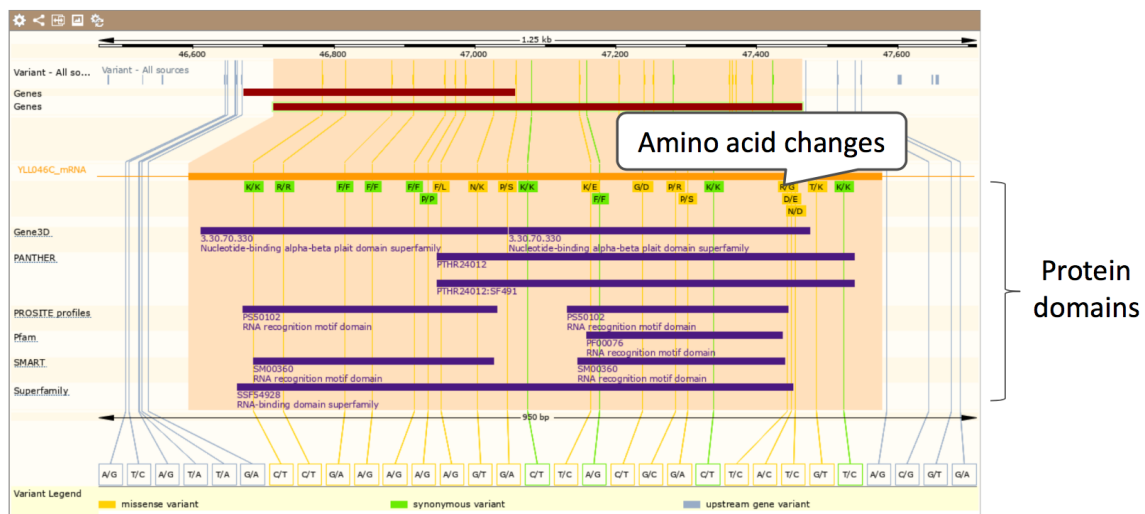


Demo: Exploring variants in Ensembl Fungi

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 occurrence 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 at 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](#).

Consequences: All Filter Other Columns

Consequences (5/27 on)

Turn All Off PTV PTV & Missense Only Exonic Turn All On

PTV = Protein Truncating Variant

transcript ablation (0) Off	inframe deletion (0) Off	mature miRNA variant (0) Off
splice donor variant (0) On	missense variant (11) On	5 prime UTR variant (0) Off
splice acceptor variant (0) On	protein altering variant (0) Off	3 prime UTR variant (0) Off
stop gained (0) On	splice region variant (0) Off	non coding transcript exon variant (0) Off
frameshift variant (0) On	incomplete terminal codon variant (0) Off	intron variant (0) Off
stop lost (0) Off	stop retained variant (0) Off	NMD transcript variant (0) Off
start lost (0) Off	synonymous variant (10) Off	non coding transcript variant (0) Off
transcript amplification (0) Off	start retained variant (0) Off	upstream gene variant (11) Off
inframe insertion (0) Off	coding sequence variant (0) Off	downstream gene variant (8) Off

Apply Cancel

gene variant

Filter SIFT: All Consequences: splice donor variant...(5/27) Filter Other Columns

Table filters

Variant ID	Chr: bp	Alleles	Class	Source	Evidence	Clin. Sig.	Conseq. Type	AA	AA co-ord	SIFT	Transcript
s12-46985	XII:46986	A/G	SNP	SGRP	-	-	missense variant	F/L	160	0	YLL046C_mRNA
s12-47025	XII:47026	G/T	SNP	SGRP	-	-	missense variant	N/K	146	0.01	YLL046C_mRNA
s12-47057	XII:47058	G/A	SNP	SGRP	-	-	missense variant	P/S	136	0.19	YLL046C_mRNA
s12-47145			SNP	SGRP	-	-	missense variant	K/E	106	1	YLL046C_mRNA
s12-47203			SNP	SGRP	-	-	missense variant	G/D	87	0.72	YLL046C_mRNA

Variant IDs/ link to variant tab

Protein function predictions

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.

Saccharomyces cerevisiae (R64-1-1) ▾

Location: XII:32,153-49,473 Gene: YLL046C Transcript: YLL046C_mRNA Variant: s12-46985

Variant tab

Variant displays

- Explore this variant
 - Genomic context
 - Genes and regulation
 - Flanking sequence
 - Genotype frequency
 - Phenotype data
 - Sample genotypes
 - Linkage disequilibrium
 - Phylogenetic context
 - Citations
 - 3D Protein model
- Configure this page
- Custom tracks
- Export data
- Share this page
- Bookmark this page

s12-46985 SNP

Most severe consequence: missense variant | See all predicted consequences

Alleles: A/G | Highest population MAF: 0.08

Location: Chromosome XII:46986 (forward strand) | VCF: XII 46986 s12-46985 A G

HGVS names: This variant has 5 HGVS names - Show

Original source: Variation features from SGRP, with Ensembl identifiers | About SGRP

About this variant: This variant overlaps 2 transcripts and has 3 sample genotypes.

VCF format

Explore this variant

- Genomic context
- Genes and regulation
- Flanking sequence
- Population genetics
- Phenotype
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations

Variation icons (these go to the same places as the links in the left hand navigation panel)

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

Genes and regulation

Gene and Transcript consequences

Gene	Transcript (strand)	Allele (Tr. allele)	Consequence Type	Position in transcript	Position in CDS	Position in protein	AA	Codons	SIFT	Detail
YLL046C	YLL046C_mRNA (-)	G (C)	missense variant	478 (out of 750)	478 (out of 750)	160 (out of 249)	F/L	TTT/CTT	0	Show
YLL047W	YLL047W_mRNA (+)	G (G)	synonymous variant	315 (out of 384)	315 (out of 384)	105 (out of 127)	K	AAA/AAG	-	Show

No overlap with Ensembl Regulatory features

No overlap with Ensembl Motif features

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.

Frequency data (1)

Population	Allele: frequency (count)	Genotype: frequency (count)
SGRP	A: 0.923 (39) G: 0.077 (39)	A: 0.923 (39) G: 0.077 (39)

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.

Variant Effect Predictor ⓘ

Species:

Name for this job (optional):

Either paste data:

Choose
Verticillium
dahliae JR2

Begin typing
Verticillium

Verticillium dahliae, (TaxID 498257)
Verti
Verticillium dahliae, (TaxID 498257)
Verticillium alfalfae VaMs.102, (TaxID 526221)
Verticillium dahliae JR2, (TaxID 1202531)
Verticillium longisporum (GCA_001268145), (TaxID 100787)
Verticillium longisporum, (TaxID 100787)
Fusarium verticillioides, (TaxID 334819)
Mortierella verticillata NRRL 6337, (TaxID 1069443)

Or upload file: Choose file No file chosen

Or provide file URL:

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.

Variant Effect Predictor

Species: Verticillium dahliae JR2, (TaxID 1202531)

Name for this job (optional): Fungal Course Demo Name your data (optional)

Either paste data:

Paste or type in data...

```
5 698711 698711 C/G
5 698935 698935 G/T
5 700313 700313 G/A
5 701484 701484 C/A
```

Run instant VEP for current line >

Examples: [Ensembl default](#), [VEP](#), [Variant identifiers](#), [HGVS notations](#)

Or upload file: Choose file No file chosen ...or upload a file

Or provide file URL:

See data format examples

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

Instant results for 5 701484 701484 C/A

Instant VEP

The below is a preview of results using the *Verticillium dahliae* JR2 Ensembl transcript database and does not include all data fields present in the full results set. To obtain these please close this preview window and submit the job using the **Run** button below.

Most severe consequence: [upstream_gene_variant](#)

Colocated variants: [tmp_5_701484_C_A](#)

Gene/Feature/Type	Consequence	Details
VDAG_JR2_Chr5g02160a: VDAG_JR2_Chr5g02160a-00001 Type: protein_coding	downstream_gene_variant	Distance to transcript: 2165bp
VDAG_JR2_Chr5g02170a: VDAG_JR2_Chr5g02170a-00001 Type: protein_coding	downstream_gene_variant	Distance to transcript: 742bp
VDAG_JR2_Chr5g02170a: VDAG_JR2_Chr5g02170a-00002 Type: protein_coding	downstream_gene_variant	Distance to transcript: 778bp
VDAG_JR2_Chr5g02171a: VDAG_JR2_Chr5g02171a-00001 Type: protein_coding	upstream_gene_variant	Distance to transcript: 64bp

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

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:



Does this variant already exist?

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

Regulatory data

Get regulatory region consequences:

Splicing predictions

dbcsNV^(p): ☐

MaxEntScan^(p): ☐

Conservation

BLOSUM62^(p): ☐

Ancestral allele^(p): ☐

(p) = functionality from VEP plugin

Filtering options Pre-filter results by frequency or consequence type

Filters

Filter by frequency: ☒ No filtering ☐ Exclude common variants ☐ Advanced filtering

Return results for variants in coding regions only: ☐

Restrict results:

NB: Restricting results may exclude biologically important data!

[Run >](#) [Clear](#) [Close form](#)

Scores for splicing changes

Choose to only see common/rare variants

Run VEP!

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

Recent jobs

[Refresh](#) This will count down and refresh page when you first submit the query

Analysis	Jobs	Submitted at	
Variant Effect Predictor	VEP analysis of Fungal Course Demo in Verticillium_dahliaejr2 Done View results	28/03/2018, 09:36 (BST)	

Click here to get results

Buttons to save, edit, share or delete the job

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.

Filters

Inclusion options
(e.g. 'is', 'matches', '>', '<=')

Begin typing and all matching options will show

Consequence is mi **Add**



Choose from the list of column names

- missense_variant
- mature_miRNA_variant
- incomplete_terminal_codon_variant

Filters

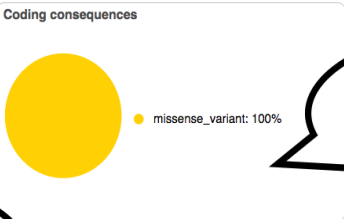
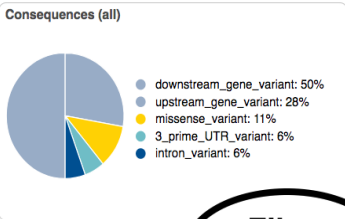
Once you've clicked 'add' it will appear in the filter box, allowing you to add other filters

You can edit or delete with these

Consequence is missense_variant  

Uploaded variant is defined **Add**

Category	Count
Variants processed	4
Variants filtered out	0
Novel / existing variants	3 (75.0) / 1 (25.0)
Overlapped genes	4
Overlapped transcripts	5
Overlapped regulatory features	-



Summary of variant consequences

Change columns

Filter options

Download options

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)