

Fungal Pathogen Genomics 2025

Wellcome Connecting Science

Manuel Carbajo, Aleena Mushtaq, Jorge Batista da Rocha
Ensembl, EMBL-EBI

Monday, 2nd to Friday, 6th June 2025



Ensembl Fungi: Introduction

**Database introductions,
Fungal Pathogen Genomics 2025**

Jorge Batista da Rocha and Aleena Mushtaq
Ensembl Outreach



*e!*Ensembl

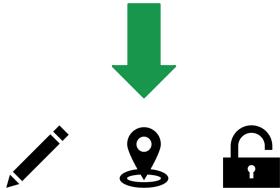
Why do we need genome browsers?

CGGCCTTGGGCTCCGCCTTCAGCTAAGACTTAACCTCCCTCCCAGCTGTCCCAGATGACGCCATCTGAAATTCTTGGAAACACGATCAC
TTAACGGAATATTGCTGTTTGGGAAGTGTTCAGCTGGGCACGCTGTATTGCCTACTTAAGCCCCTGGTAATTGCTGTATTGCGCTTCTGAA
CGAAGACATGCTGATGGATTACCAAGGCAGCGTTGGTCTCTAACTGGAGCCCTCTGTCCCCACTAGCCACGCGTCACTGGTTAGCGTGATTG
GAAACTAAATCGTATGAAAATCCTCTCTAGTCGACTAGCCACGTTCGAGTGCTTAATGTGGCTAGTGGCACCGGTTGGACAGCAC
GCTGTAAAATGTTCCCACAGTAAGCTGTTACCGTTCCAGGAGATGGACTGAATTAGAATTCAAACAAATTTCAGCGCTTCTGAA
GTTTACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAGTGCATTGGTCTTGTGACTTACCAAGCATTGGAGGAA
ATATCGTAGGTAAAATGCCTATTGGATCCAAGAGAGGCCAACATTGGAAATTGAAACACGCTGCAACAAAGCAGGTATTGACA
ATTTATATAACTTATAAATTACACCGAGAAAGTGTTCCTAAAAAATGCTGCTAAAAACCCAGTACGTACAGTGTGCTTAGAACCAT
AAACTGTTCCCTATGTGTGATAAAATCCAGTTAACACATAATCATCGTTGCAGGTTAACCATGATAAAATAGAACGTCTAGTGGATA
AAGAGGAAACTGGCCCTTGACTAGCAGTAGGAACAATTACTAACAAATCAGAACATTAATGTTACTTATGGCAGAAGTTGTCCA
TTGGTTTCAGTACTCCTTACTCTTAAATGATCTAGGACCCCCGGAGTGCTTTGTTATGTAGCTTACCATATTAGAAATTAAA
AAGAATTAAAGGCTGGCGTGGCTCACGCCTGTAATCCCAGCACTTGGAGGCCAGGTGGCGGATCATTGAGGCCAGAAGTTGA
GACCAGCCTGGCCAACATGGTGAACACCTATCTACTAAAAACAAAAATGCTGCGTGTGGTGGTGCCTGTAATCCCAGCA
ACGGGAGGTGGAGGCAGGAGAATCGCTGAACCCCTGGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCACTGCACTTAGCCTGGCC
TAGCATGACTCTGTCTAAAACAAACAAACAAAAACTAAGAATTAAAGTTAACTTACTAAAAATAATGAAAGCTAACCCATTGCA
TATTATCACAAACATTCTAGGAAAAAAACTTTGAAAACAAGTGAGTGGAAATAGTTTACATTGGCAGTTCTTTAATGTCTGGCT
AAATAGAGATAGCTGGATTCACTTATCTGTGCTAATCTGTTATTGGTAGAAGTATGTGAAAAAAATTAACCTCACGTTGAAAAAGGAA
ATATTAAATAGTTTCAGTTACTTTGGTATTTCTGTACTTGCATAGATTTCAGATCTAATAGATATAACCAGGGCTTT
CCATGTCGCAACATCATGCACTGATTGGAAAGATAGTGGTGTCTGAATTGCAATTGCATGCCGTTAAAGC

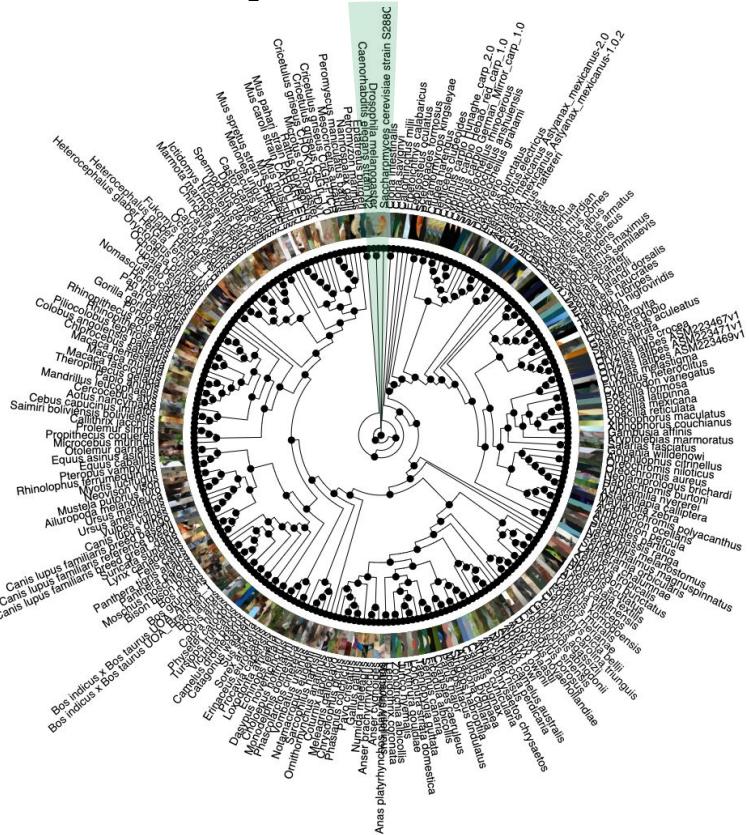
What is Ensembl?

Ensembl annotates and maps genomic features from genome sequences, and brings together information from a wide range of other databases in a single site.

```
AGTGCTTAATGTGGCTAGTGGCACCGGTTGGACAGCACAGCTGTAAAATGTTCCC  
ATCCTCACAGTAAGCTGTTACCGTTCCAGGAGATGGGACTGAATTAGAATTCAAAC  
AAATTTCCAGCGCTTCTGAGTTTACCTCAGTCACATAATAAGGAATGCATCCCT  
GTGTAAGTGCATTGGTCTTCTGTTGCAGACTTATTACCAAGCATTGGAGGA  
ATATCGTAGGTAAAAATGCCATTGGATCCAAGAGAGGCCAACATTTTGAAAT  
TTTTAAGACACGCTGCAACAAAGCAGGTATTGACAAATTTATATAACTTATAAAA  
TTACACCGAGAAAGTGTCTAAAAATGCTGCTAAAACCAGTACGTACAG  
TGTTGCTAGAACCATAAACTGTTCTATGTGTATAAATCCAGTTAACACAT  
AATCATCGTTGCAGGTTAACACATGATAAATAGAACGTCTAGTGGATAAAGA  
GGAAACTGGCCCCTGACTAGCAGTAGGAACAATTACTAACAAATC
```



Ensembl: vertebrate species



Ensembl Genomes: non-vertebrate species

EnsemblPlants

<https://plants.ensembl.org/>

Search: All species e.g. *Carboxy** or *ch22*

All genomes — Select a species — Favourite genomes

View full list of all species

Wheat assemblies
Ensembl Plants hosts the [last wheat assembly](#) from the IWGSC (PanDae v1.0), including:
• The IWGSC RefSeq v1.0 gene annotation, with links to wheat-specific transcript and protein data;
• 14 wheat chromosomes (Chr 1-4 and Chr 5-8);
• Alignment of ~20,720 high confidence genes from TIGACv1 annotation;
• Around 20K BOKS SNPs arrays from [CereDB](#)v9, including GTL lines in selected cases and Linkage Disequilibrium display; with GTL learning datasets;
• A collection of ~100,000 wheat SSRs from [WheatSSR](#), with links to Codon (coding regions) and Konka (coding regions and promoters);
• Chromosome specific KASP markers were added from the Nottingham BBSRC Wheat Research Centre;
• Inter-Homologous Variants (IHVs) between the A, B and D genome components;
• Chromosome specific KASP markers were added from the Nottingham BBSRC Wheat Research Centre;
• A collection of ~100,000 wheat SSRs from [WheatSSR](#) and [TIGAC](#);
• A collection of ~100,000 wheat SSRs from [TIGAC](#) mapped to the previous TIGAC v1 assembly, archived at [eGIG](#)-[wheat.ensembl.org](#);
• Phylogenetic tree enabled, allowing users to view alignments among multiple wheat accessions ([wheat.accessions](#)).

EnsemblMetazoa

<https://metazoa.ensembl.org/>

Search: All species e.g. *CPRD* or *chitin**

All genomes — Select a species — Favourite genomes

View full list of all species

What's New in Release 53
Updated data
• *Anopheline palps* (Piso apical, GCA_005009785.1)
• *Anopheline palps* (Piso apical, GCA_005009785.1)
• *Anopheline palps* (Monach long-horned beetle, GCA_000300025.2)
• *Bombyx mori* (Domestic silkworm, GCA_014002326.1)
• *Danaca pluvialis* (Monarch butterfly, GCA_000721995.1)
• *Glossina kibensis* (Tsetse fly, GCA_014006210.1)
• *Synopsis invicta* (Red ant, GCA_019827295.1)
• *Synopsis invicta* (Red ant, GCA_019827295.1)
• *Tetraponera damarensis* (Malaysian stingless fly, GCA_002237195.2)

• New species
• *Coleo cyprinodontella* (Southern house mosquito, GCA_013702785.1)
• *Acodes scapularis* (Blue-legged tick, GCA_019827852.2)
• *Dermacentor alpinus* (Ticks, GCA_01339741.1)

EnsemblBacteria

<https://bacteria.ensembl.org/>

Search for a gene Search for a genome
e.g. type esp-1 to find *Escherichia coli* or uridylate*
Archive sites
The following archive sites are available to access previous versions of data:
• Release 40, December 2020 [eGIG-bacteria.ensembl.org](#);
• Release 45, September 2019 [eGIG-bacteria.ensembl.org](#);
• Release 40, July 2018 [eGIG-bacteria.ensembl.org](#);
• Release 37, October 2017 [eGIG-bacteria.ensembl.org](#)

Ensembl Bacteria
Ensembl Bacteria is a browser for bacterial and archaeal genomes. These are drawn from the databases of the International Nucleotide Sequence Database Collaboration, the European Nucleotide Archive at the EBI, Genbank at the US National Center for Biotechnology Information and the Collection of Acinetobacter.

Data access
Data can be visualised through the Ensembl genome browser and accessed programmatically via our Perl and REST APIs. Data is also accessible via the Ensembl API, which provides a standard interface for querying genomes as defined by criteria set out by UniProt. See more details about the update in our [release notes](#).

EnsemblProtists

<https://protists.ensembl.org/>

Search: All species e.g. *PF397_8532500* or *cyst**

All genomes — Select a species — Favourite genomes

View full list of all species

What's New in Release 53
New assemblies with gene and protein annotation every two weeks.
Now, species that already exist on the site will continue to be updated with the full range of annotations.
The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as the Human Genome Project, the Vertebrate Genomes Project and the Earth Biodiversity Project. [Read the press release](#) or [our blog](#).

EnsemblFungi

<https://fungi.ensembl.org/>

Search: All species e.g. *NAT2* or *alcohol**

All genomes — Select a species — Favourite genomes

View full list of all species

What's New in Release 53
New assemblies with gene and protein annotation every two weeks.
Now, species that already exist on the site will continue to be updated with the full range of annotations.
The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as the Human Genome Project, the Vertebrate Genomes Project and the Earth Biodiversity Project. [Read the press release](#) or [our blog](#).

EnsemblCOVID-19

<https://covid-19.ensembl.org/>

SARS-CoV-2
Reference Genome
The complete genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolate Wuhan-Hu-1, release of the disease COVID-19-01.
Genome assembly: **ASMB5888v3** [GCA_005008053.3](#)
Sequencer: MN908947

View in Genome browser
• Example gene
• Example protein summary
• Whole genome

COVID-19 resources
The EnsemblGENCODE human protein coding gene set linked to COVID-19 can be found at [HumanCode19.GeneAnnotations](#).

Sequence and annotation data
• Download data

Assemblies and annotation



Genome assemblies



JOINT GENOME INSTITUTE
DEPARTMENT OF ENERGY



UNIVERSITY & RESEARCH

100 years
1918 - 2018

Data generators and
collaborators, in-house annotation



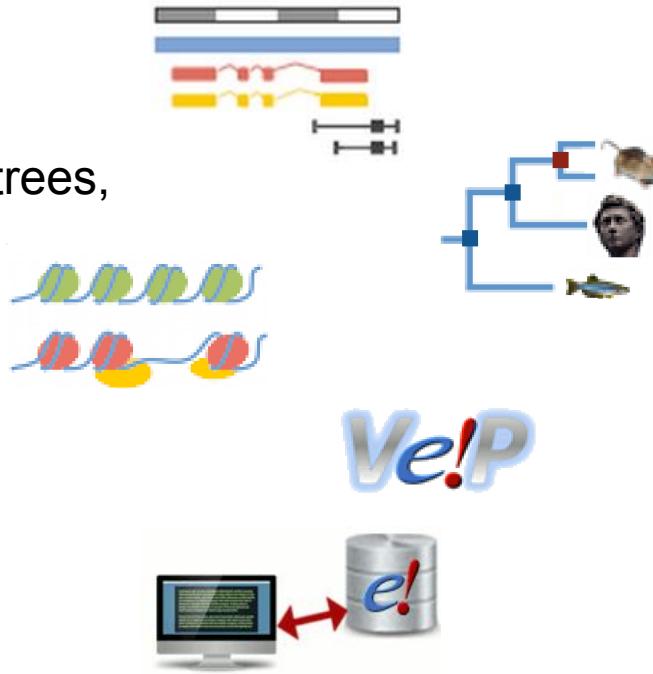
<https://fungi.ensembl.org/info/genome/index.html>

e!Ensembl

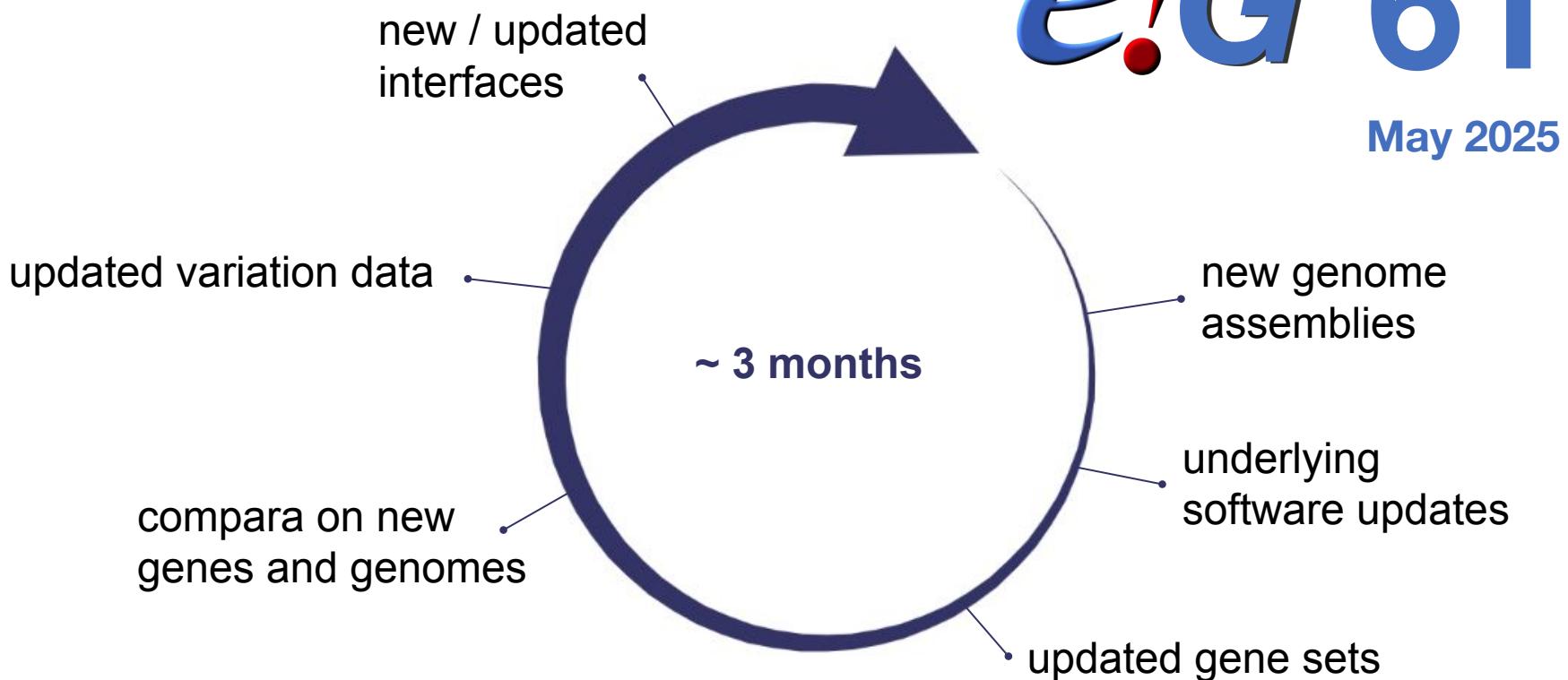
EMBL-EBI

Ensembl Fungi features

- Genome assemblies and gene builds
- Variation data and Ensembl VEP
- Comparative genomics (alignments, gene trees, homologues)
- Cross-species molecular interactions
- Microarray data
- TrackHubs
- BioMart (database query)
- Display your own data
- Programmatic access via APIs
- Completely open-source (FTP, GitHub)



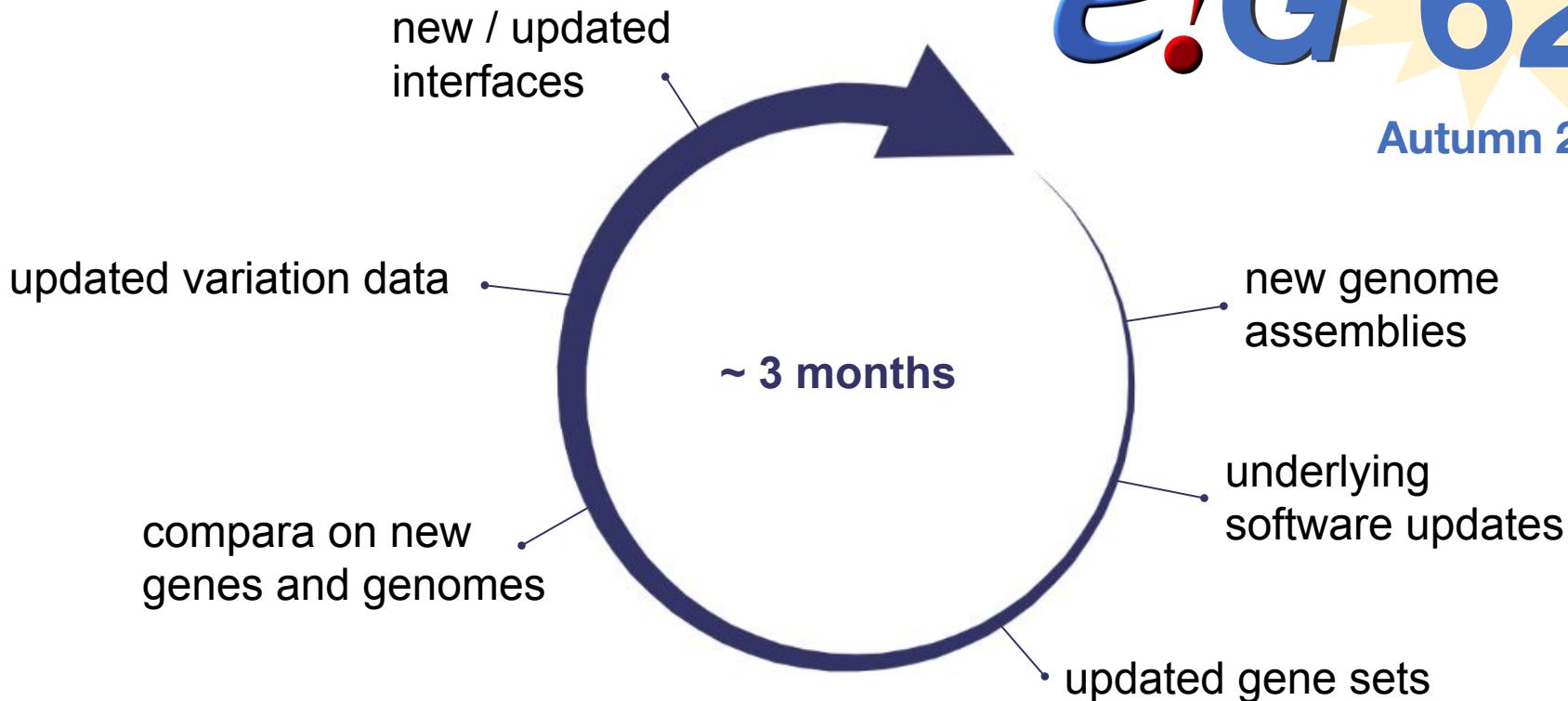
Ensembl Release Cycle



e!G 61

May 2025

Ensembl Release Cycle



Ensembl Rapid Release

- Released every two weeks
- Genome assemblies and gene builds
- BLAST
- Limited genetic variation data
- Comparative analyses gradually rolled out across all species
- No BioMart



The new Ensembl genome browser: Ensembl Beta

- Ensembl Rapid Release has been replaced by the new Ensembl website beta.ensembl.org
- The new site has equivalent functionality as Rapid Release, with additional species and richer functionality coming online in the near future
- Highly responsive and modern user interface
- Users will have the ability to browse multiple species of interest simultaneously
- All Ensembl sites in a single platform

[About the ENSEMBL project](#)

ENSEMBL

Genome data & annotation

[About using Ensembl](#)

Species selector



Create & manage your own
species list

Genome browser



Look at genes & transcripts
in their genomic context

Entity viewer



Get gene & transcript
information

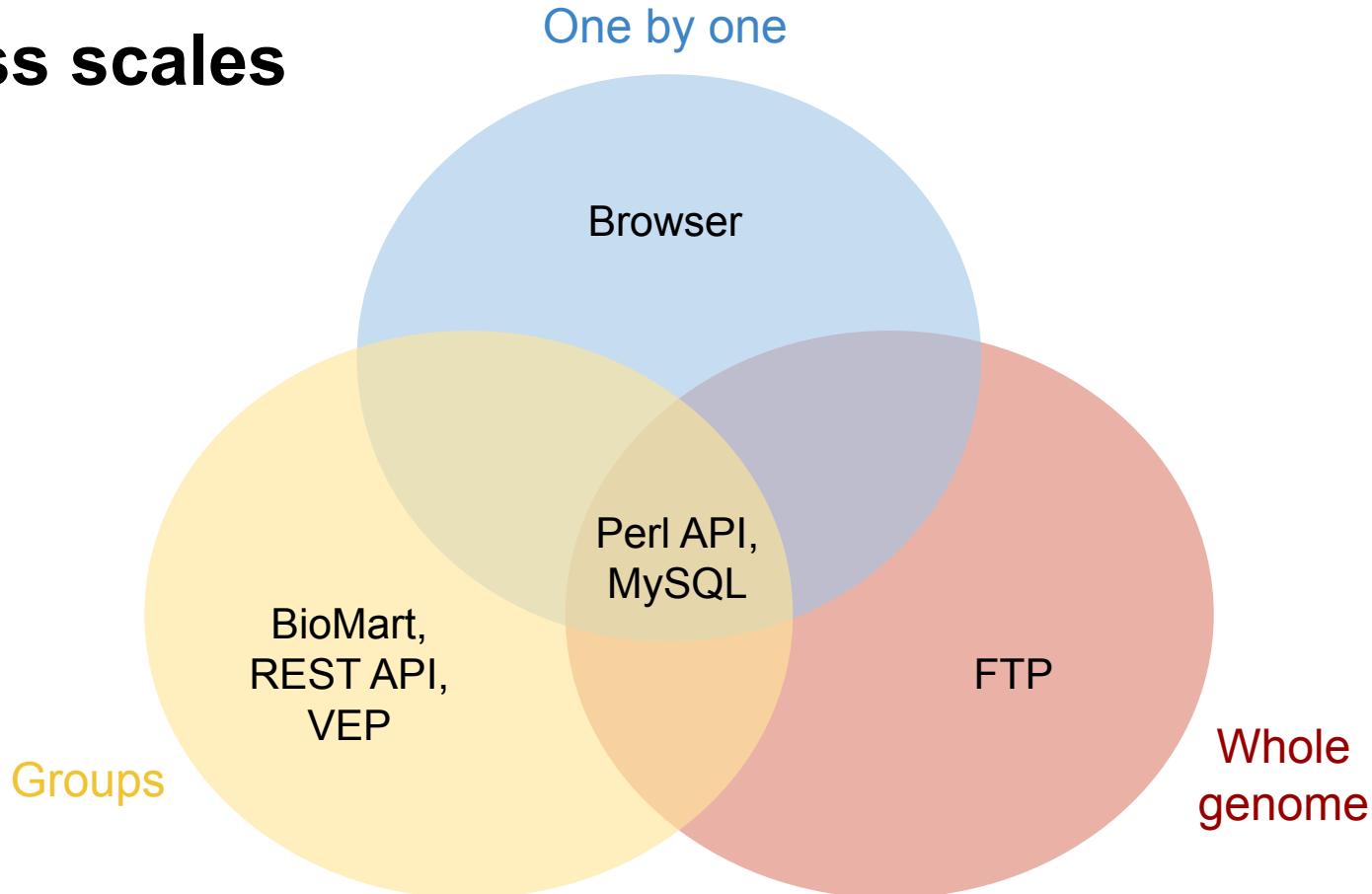


ENSEMBL EMBL-EBI

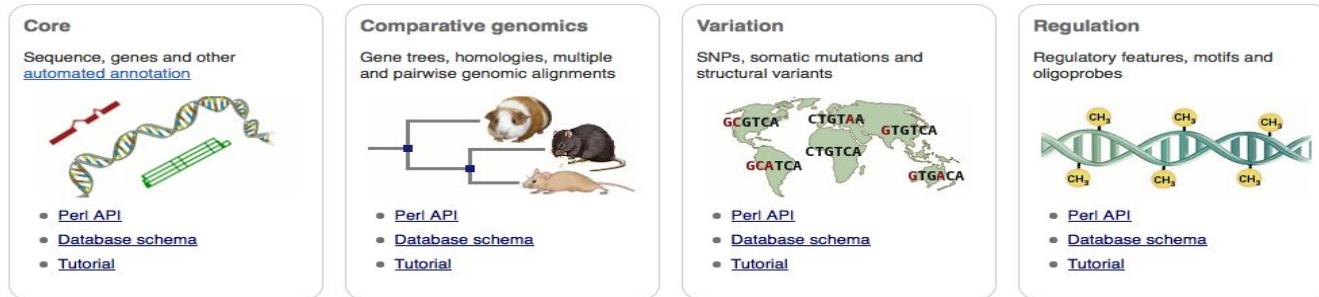
Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

[Ensembl blog](#)

Access scales



Use the Ensembl infrastructure



bio  **mart**

Genome
browser



Bulk data retrieval
+ bespoke queries



Programmatic
data retrieval



Variant effect
prediction

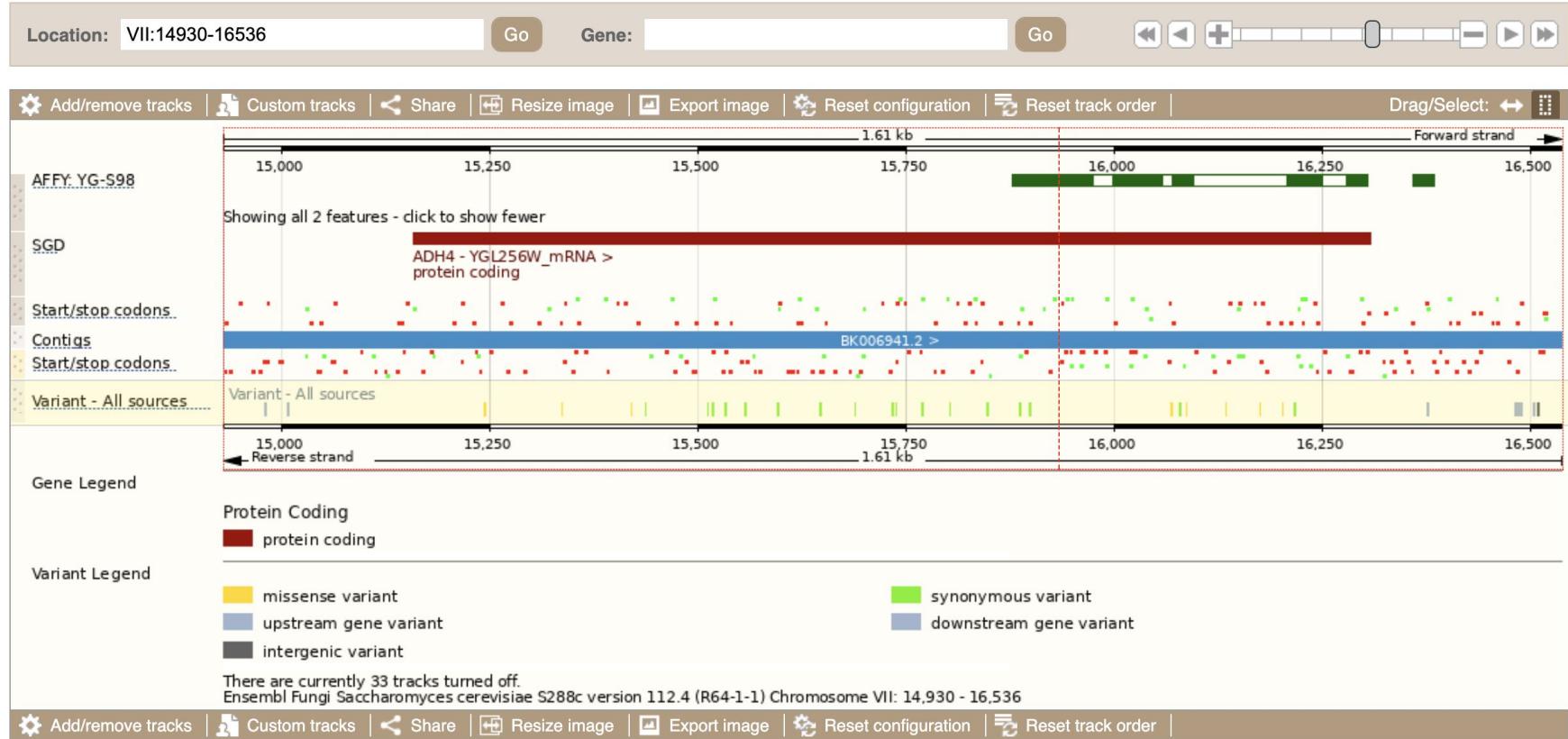
The Track Hub Registry

Community gene
annotation

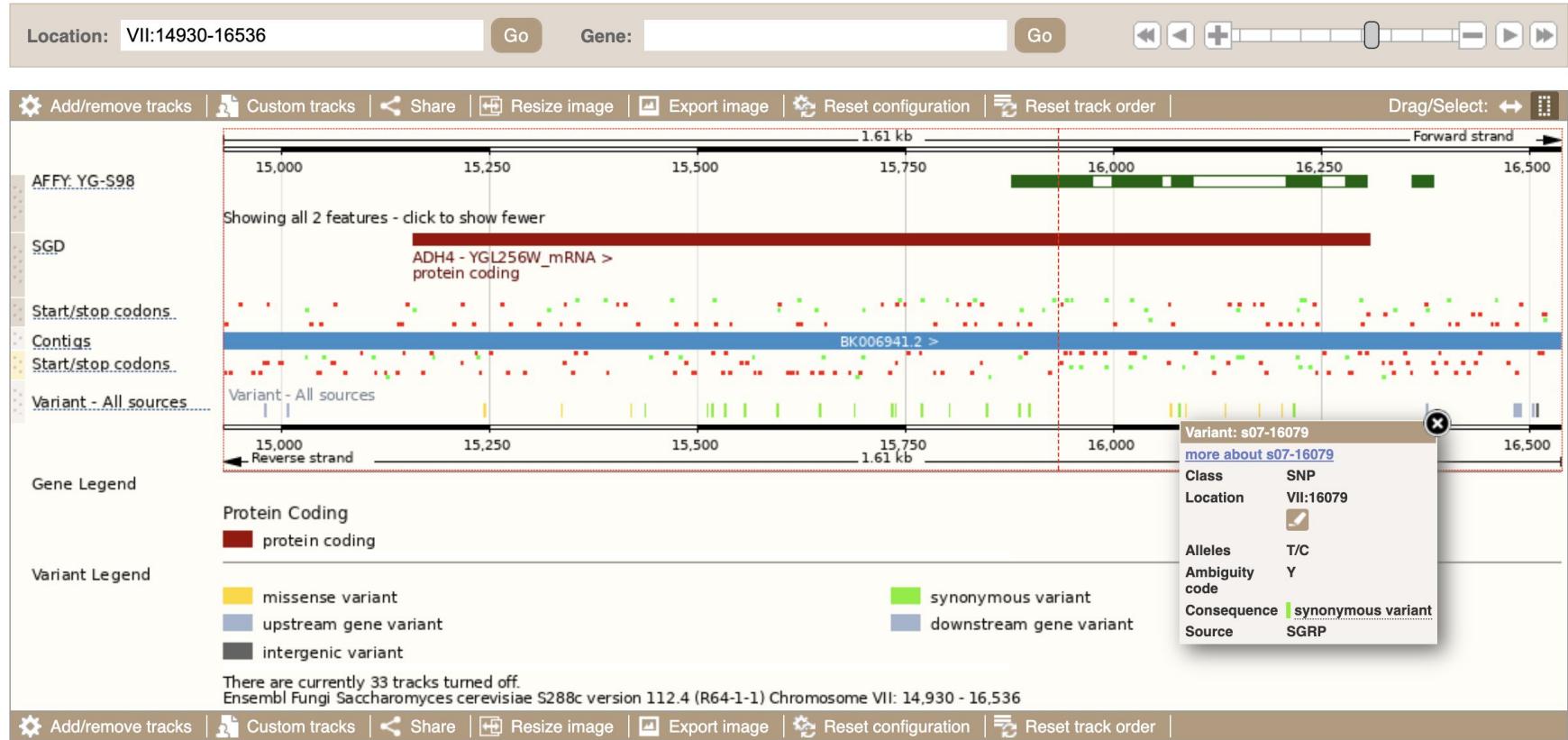
Display external
RNA-Seq data

Training

Visualising genomic regions



Visualising genomic regions




 Saccharomyces cerevisiae (R64-1-1) ▾

Location: VII:15,159-16,307

Gene: ADH4

Transcript: ADH4

Variant: s07-14981

Gene-based displays

⊖ Summary

- Splice variants
- Transcript comparison
- Gene alleles

⊖ Sequence

- Secondary Structure
- Gene families
- Literature

⊖ Fungal Compara

- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues

⊖ Pan-taxonomic Compara

- Gene Tree
- Orthologues

⊖ Ontologies

- GO: Cellular component
- GO: Biological process
- GO: Molecular function
- PHL: Phibase identifier

⊖ Phenotypes

⊖ Genetic Variation

- Variant table
- Variant image

- Structural variants

⊖ Gene expression

Pathway

Molecular interactions

Regulation

External references

Supporting evidence

⊖ ID History

- Gene history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Gene: ADH4 YGL256W

Description

Alcohol dehydrogenase isoenzyme type IV; dimeric enzyme demonstrated to be zinc-dependent despite sequence similarity to iron-activated alcohol dehydrogenases; transcription is induced in response to zinc deficiency [Source:SGD;Acc:S000003225]

Gene Synonyms

NRC465, ZRG5

Location

Chromosome VII: 15,159-16,307 forward strand.

About this gene

This gene has 1 transcript ([splice variant](#)), 306 orthologues and is a member of [1 Ensembl protein family](#).

Transcripts

[Show transcript table](#)

Summary ⓘ

Name

[ADH4](#) (SGD gene name)

UniProtKB

This gene has proteins that correspond to the following UniProtKB identifiers: [P10127](#)

Gene type

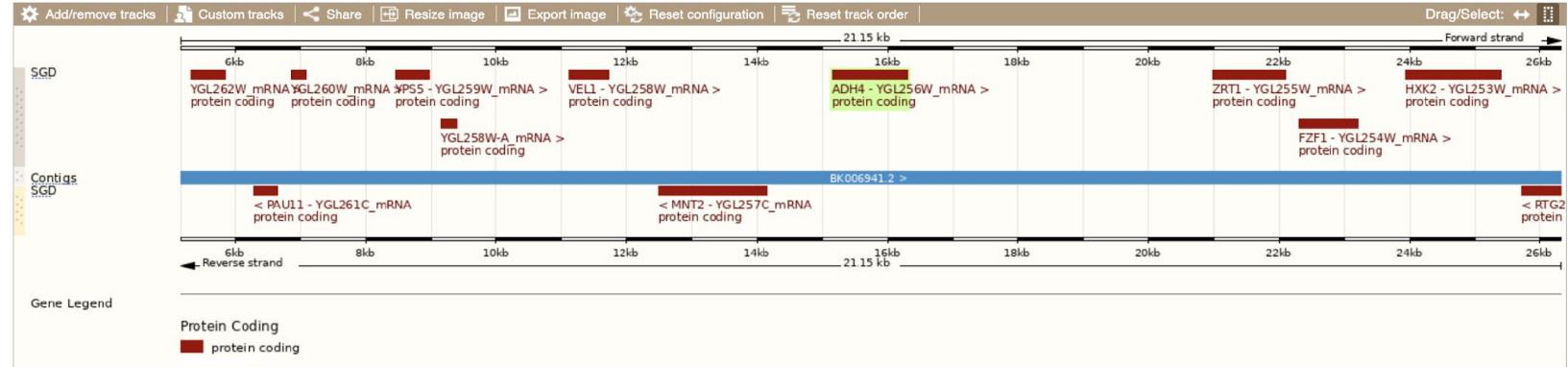
Protein coding

Annotation method

Annotation imported from SGD

[Go to Region in Detail](#) for more tracks and navigation options (e.g. zooming)[Add/remove tracks](#)[Custom tracks](#)[Share](#)[Resize image](#)[Export image](#)[Reset configuration](#)[Reset track order](#)

Drag>Select: ↕




 Saccharomyces cerevisiae (R64-1-1) ▾

Location: VII:15,159-16,307

Gene: ADH4

Transcript: ADH4

Variant: s07-14981

Gene-based displays

↳ Summary

- Splice variants
- Transcript comparison
- Gene alleles

↳ Sequence

- Secondary Structure
- Gene families
- Literature

↳ Fungal Compara

- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues

↳ Pan-taxonomic Compara

- Gene Tree
- Orthologues

↳ Ontologies

- GO: Cellular component
- GO: Biological process
- GO: Molecular function
- PHI: Phibase identifier

↳ Phenotypes

↳ Genetic Variation

- Variant table
- Variant image
- Structural variants

– Gene expression

– Pathway

– Molecular interactions

– Regulation

– External references

– Supporting evidence

↳ ID History

↳ Gene history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Gene: ADH4 YGL256W

Description

Alcohol dehydrogenase isoenzyme type IV; dimeric enzyme demonstrated to be zinc-dependent despite sequence similarity to iron-activated alcohol dehydrogenases; transcription is induced in response to zinc deficiency [Source:SGD;Acc:S000003225]

Gene Synonyms

NRC465, ZRG5

Location

Chromosome VII: 15,159-16,307 forward strand.

R64-1-1:BK006941.2

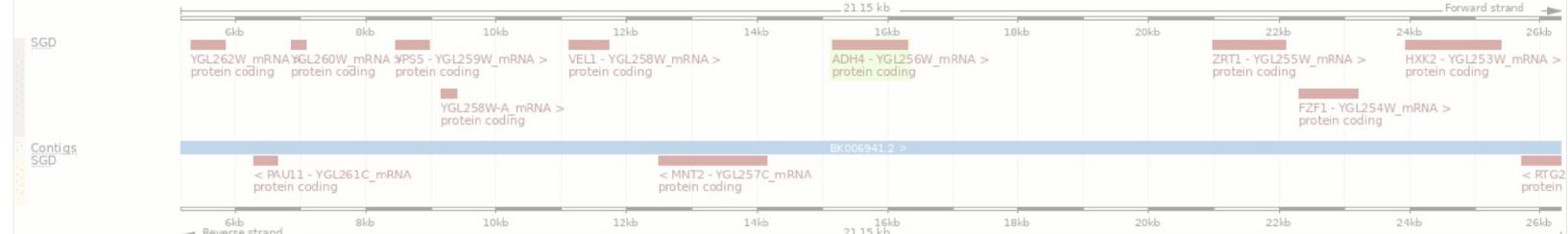
About this gene

This gene has 1 transcript ([splice variant](#)), 306 orthologues and is a member of [1 Ensembl protein family](#).

Transcripts

[Show transcript table](#)Summary 
[HMMER](#) | [BLAST](#) | [BioMart](#) | [Tools](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

Go to Region In Detail for more tracks and navigation options (e.g. zooming)

[Add/remove tracks](#) | [Custom tracks](#) | [Share](#) | [Resize image](#) | [Export image](#) | [Reset configuration](#) | [Reset track order](#)
Drag>Select: Forward strand 

Protein Coding

 protein coding


 Saccharomyces cerevisiae (R64-1-1) ▾

Location: VII:15,159-16,307

Gene: ADH4

Transcript: ADH4

Variant: s07-14981

Gene-based displays

⊖ Summary

- ⊖ Splice variants
- ⊖ Transcript comparison
- Gene alleles

⊖ Sequence

- ⊖ Secondary Structure
- Gene families
- Literature

⊖ Fungal Compara

- ⊖ Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues

⊖ Pan-taxonomic

- ⊖ Gene Tree
- Orthologues

⊖ Ontologies

- ⊖ GO: Cellular
- GO: Biologic
- GO: Molecular

⊖ Phenotypes

- ⊖ Genetic Variatio
- Variant table
- Variant image

⊖ Structural variants

- Gene expression

Pathway

- Molecular interactions

Regulation

- External references

Supporting evidence

⊖ ID History

- Gene history

Gene: ADH4 YGL256W

Description

Alcohol dehydrogenase isoenzyme type IV; dimeric enzyme demonstrated to be zinc-dependent despite sequence similarity to iron-activated alcohol dehydrogenases; transcription is induced in response to zinc deficiency [Source:SGD;Acc:S000003225]

Gene Synonyms

NRC465, ZRG5

Location

Chromosome VII: 15,159-16,307 forward strand.

R64-1-1:BK006941.2

About this gene

This gene has 1 transcript (splice variant), 306 orthologues and is a member of 1 Ensembl protein family.

Transcripts

[Show transcript table](#)

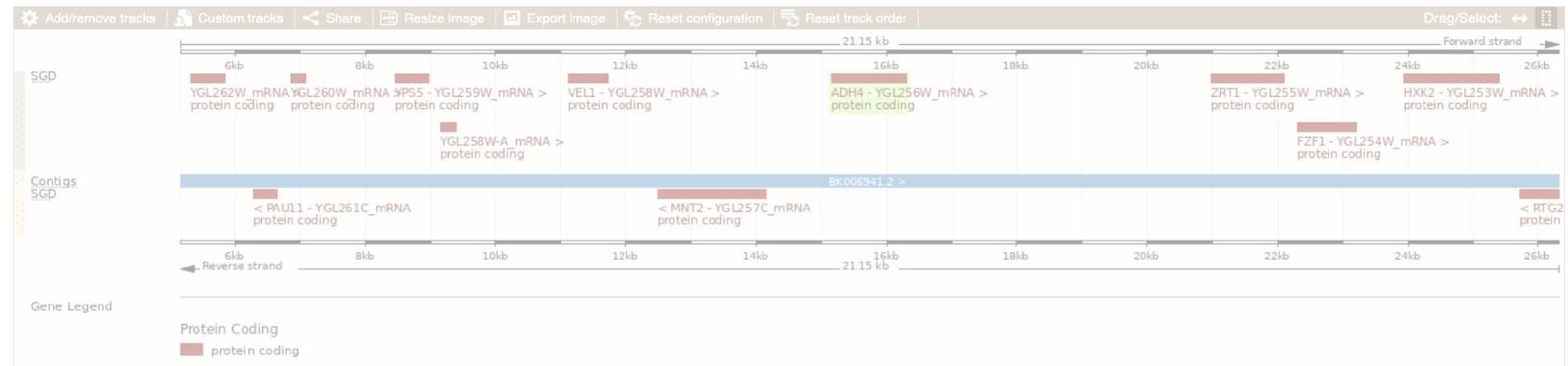
Saccharomyces cerevisiae (R64-1-1) ▾

Location: VII:15,159-16,307

Gene: ADH4

Transcript: ADH4

Variant: s07-14981



Location: VII:15,159-16,307 Gene: ADH4 Transcript: ADH4 Variant: s07-14981

Gene-based displays

- ⊖ Summary
 - └ Splice variants
 - └ Transcript comparison
 - └ Gene alleles
- ⊖ Sequence
 - └ Secondary Structure
 - └ Gene families
 - └ Literature
- ⊖ Fungal Compara
 - └ Genomic alignments
 - └ Gene tree
 - └ Gene gain/loss tree
 - └ Orthologues
 - └ Paralogues
- ⊖ Pan-taxonomic Compara
 - └ Gene Tree
 - └ Orthologues
- ⊖ Ontologies
 - └ GO: Cellular component
 - └ GO: Biological process
 - └ GO: Molecular function
 - └ PHI: Phibase identifier
- ⊖ Phenotypes
- ⊖ Genetic Variation
 - └ Variant table
 - └ Variant image
 - └ Structural variants
- ⊖ Gene expression
- ⊖ Pathway
- ⊖ Molecular interactions
- ⊖ Regulation
- ⊖ External references
- ⊖ Supporting evidence
- ⊖ ID History
- └ Gene history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Location: VII:15,159-16,307 Gene: ADH4 Transcript: ADH4 Variant: s07-14981

Gene: ADH4 YGL256W

Description Alcohol dehydrogenase (E1) subunit 4; involved in zinc deficiency [Source: SGD]

Gene synonyms NRC465, ZRG5

Location Chromosome VII: 15,159-16,307 R64-1-1:BK006941.2

About this gene This gene has 1 transcript

[Show transcript table](#)

Summary ⓘ

Name	ADH4 ⓘ (SGD gene name)
UniProtKB	This gene has proteins in UniProtKB
Gene type	Protein coding
Annotation method	Annotation imported from SGD

[Go to Region in Detail for more tracks and navigation](#)

[Add/remove tracks](#) | [Custom tracks](#) | [Share](#) | [Reset](#)

SGD

6kb 8kb

YGL262W_mRNA -> YGL260W_mRNA < PPS5 -> protein coding protein coding protein coding

6kb 8kb

< PAU11 -> YGL261C_mRNA protein coding

6kb 8kb

< RTG2 -> YGL254W_mRNA > protein coding

Reverse strand

Gene Legend

Protein Coding

protein coding

Location: VII:15,159-16,307 Gene: ADH4

Gene-based displays

Summary

- └ Splice variants
- └ Transcript comparison
- └ Gene alleles

Sequence

- └ Secondary Structure
- └ Gene families
- └ Literature

Fungal Compara

- └ Genomic alignments
- └ Gene tree
- └ Gene gain/loss tree
- └ Orthologues
- └ Paralogues

Pan-taxonomic Compara

- └ Gene Tree
- └ Orthologues

Ontologies

- └ GO: Cellular component
- └ GO: Biological process
- └ GO: Molecular function
- └ PHI: Phibase identifier

Phenotypes

Genetic Variation

- └ Variant table
- └ Variant image
- └ Structural variants

Gene expression

Pathway

Molecular interactions

Regulation

External references

Supporting evidence

ID History

- └ Gene history

Gene

Description

despite sequence similarity to iron-activated alcohol dehydrogenases; transcription is induced in response to

Gene Structure

Location

About this gene

Transcript

Summary ⓘ

Name

UniProtKB

Gene type

Annotation method

SGD

18kb 20kb 22kb 24kb 26kb

Forward strand

YGL256W_mRNA > protein coding

ZRT1 -> YGL255W_mRNA > protein coding

HXK2 -> YGL253W_mRNA > protein coding

FZF1 -> YGL254W_mRNA > protein coding

< RTG2 -> protein coding

Drag>Select:

SGD

18kb 20kb 22kb 24kb 26kb

Reverse strand

SGD

18kb 20kb 22kb 24kb 26kb

< RTG2 -> protein coding

Saccharomyces cerevisiae (R64-1-1) ▾

Location: VII:15,159-16,307

Gene: ADH4

Transcript: ADH4

Variant: s07-14981

Gene-based displays

⊖ Summary

- Splice variants
- Transcript comparison
- Gene alleles

⊖ Sequence

- Secondary Structure
- Gene families
- Literature

⊖ Fungal Compara

- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues

⊖ Pan-taxonomic Compara

- Gene Tree
- Orthologues

⊖ Ontologies

- GO: Cellular component
- GO: Biological process
- GO: Molecular function

⊖ PHiB

- Phibase identifier

⊖ Phenotypes

⊖ Genetic Variation

- Variant table
- Variant image

– Structural variants

– Gene expression

– Pathway

– Molecular interactions

– Regulation

– External references

– Supporting evidence

⊖ ID History

- Gene history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Gene: ADH4 YGL256W

Description

Alcohol dehydrogenase isoenzyme type IV; dimeric enzyme demonstrated to be zinc-dependent despite sequence similarity to iron-activated alcohol dehydrogenases; transcription is induced in response to zinc deficiency [Source:SGD;Acc:S000003225]

Gene Synonyms

NRC465, ZRC5

Location

Chromosome VII

R64-1-1-BK006941.2

About this gene

This gene has

Transcripts

Show transcripts

Summary

Name

ADH4

UniProtKB

This gene has

Gene type

Protein coding

Annotation method

Annotation in

Go to Region in Detail for more tracks and analysis

Add/remove tracks

Custom tracks

Share this page

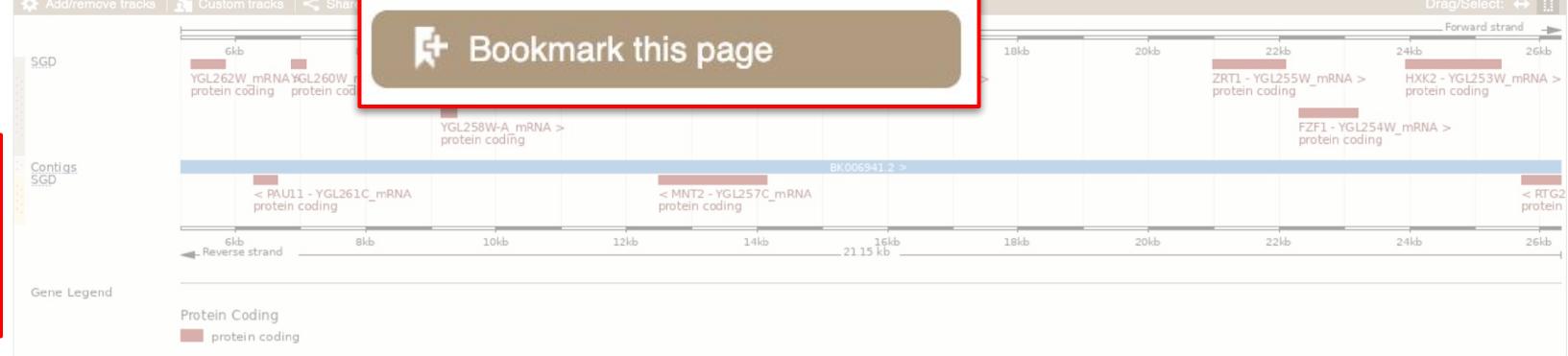
Configure this page

Custom tracks

Export data

Share this page

Bookmark this page



What can you use Ensembl Fungi for?

- Visualise genomic regions and map custom annotations using tracks
- Learn more about genes of interest
- Learn more about variants of interest and annotate your own variants
- Investigate orthologues to explore behaviour in other pathogens
- Search for information in external databases

Fungal Pathogen Genomics 2025: Ensembl Fungi

Mon, 2 June

Introduction to
Database Queries

Mon, 2 June

Transcriptomics &
Proteomics

Tue, 4 June

SNPs & Variants

Wed, 5 June

Comparative
Genomics &
Orthology

- | | | | |
|---|--|---|---|
| <ul style="list-style-type: none">• Bulk Ensembl data download• Molecular interactions | <ul style="list-style-type: none">• Public RNAseq data• Visualise your own data | <ul style="list-style-type: none">• Variation data• Ensembl Variant Effect Predictor (VEP) | <ul style="list-style-type: none">• Gene trees• Whole-genome alignments (WGAs) |
|---|--|---|---|

Ensembl staff at this course



Manuel Carbajo

Ensembl Microbes

Aleena Mushtaq

Ensembl

Jorge Batista da

Rocha

Special thanks:

Nishadi De Silva

Outreach

Ensembl Outreach

Leaving

Ensembl Microbes

Publications

- List of Ensembl publications:
<https://www.ensembl.org/info/about/publications.html>
- Topic-specific articles mentioned throughout the workshop

Dyer SC, Austine-Orimoloye O, Azov AG, Barba M, et al. [Ensembl 2025](#). Nucleic Acids Research. 2024 Dec;53(D1):D948–D957. DOI: 10.1093/nar/gkae1071. PMID: 39656687 PMCID: PMC11701638

Yates AD, Allen J, Amode RM, et al. [Ensembl Genomes 2022: an expanding genome resource for non-vertebrates](#). Nucleic Acids Research. 2022 Jan;50(D1):D996-D1003. DOI: 10.1093/nar/gkab1007. PMID: 34791415; PMCID: PMC8728113.

<https://www.ensembl.org/info/about/publications.html>

Acknowledgements and funding



UK Research
and Innovation



National Human
Genome Research
Institute (NHGRI)



National Institutes of
Health (NIH)



Funded by the European Union
Horizon 2020
Horizon Europe Programme

Ensembl Fungi: BioMart

Introduction to database queries,
Fungal Pathogen Genomics 2025

Jorge Batista da Rocha and Aleena Mushtaq
Ensembl Outreach



Ensembl BioMart

The 4-step principle

Step 1: Dataset

Step 2: Filters

Step 3: Attributes

Step 4: Results

Choose the database and species.

Narrow down the dataset.

Specify your output and what to print on your table.

Choose the format of your results and export.

Biomart Interface

The screenshot shows the Ensembl Fungi Biomart interface. At the top, there is a navigation bar with links for BLAST, BioMart, FTP, Docs & FAQs, URL, XML, Perl, and Help. Below the navigation bar, there is a search bar and a dropdown menu. On the left side, there is a sidebar with a 'Dataset' section that currently says '[None selected]'. In the center, there is a large white area with a dropdown menu labeled '- CHOOSE DATABASE -'.

Biomart Interface - choose datasets

EnsemblFungi

New Count Results

BLAST | BioMart | FTP | Docs & FAQs

URL XML Perl Help

Dataset

[None selected]

Ensembl Fungi Genes 61

✓ - CHOOSE DATASET -

- Aphanomyces astaci genes (GCA000520075v1)
- Aphanomyces invadans genes (GCA000520115v1)
- Ashbya gossypii genes (ASM9102v1)
- Aspergillus clavatus NRRL 1 genes (ASM271v1)
- Aspergillus fischeri NRRL 181 genes (ASM14964v1)
- Aspergillus flavus NRRL3357 genes (JCVI-afl1-v2.0)
- Aspergillus fumigatus A1163 genes (ASM15014v1)
- Aspergillus fumigatus Af293 genes (ASM265v1)
- Aspergillus nidulans genes (ASM1142v1)
- Aspergillus niger genes (ASM285v2)
- Aspergillus oryzae RIB40 genes (ASM18445v3)
- Aspergillus terreus NIH2624 genes (ASM14961v1)
- Beauveria bassiana genes (ASM168263v1)
- Blumeria graminis genes (EF2)
- Botrytis cinerea B05.10 genes (ASM83294v1)
- Candida albicans genes (GCA000182965v3)
- Candida auris genes (GCA002759435v2)
- Candida duboishaeumulonis genes (GCA002926085v1)
- Candida glabrata genes (GCA000002545v2)

Biomart Interface - choose datasets

EnsemblFungi

New Count Results

BLAST | BioMart | FTP | Docs & FAQs

URL XML Perl Help

Dataset
Saccharomyces cerevisiae genes (R64-1-1)

Filters
[None selected]

Attributes
Gene stable ID
Transcript stable ID

Dataset
[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION:

Chromosome/scaffold

Coordinates

Start: 1

End: 10000000

Multiple regions (Chr:Start:End:Strand) [Max 500 advised]
e.g. 1:100:10000:-1, 1:100000:200000:1

GENE:

PATHOGEN PHENOTYPES (PHI-BASE):

GENE ONTOLOGY:

Ensembl Genomes release 61 - April 2025 (c) EBI



https://github.com/WCSCourses/Fungal_Pathogen_Genomics_2025

EMBL-EBI

Biomart Interface - choose attributes

EnsemblFungi

New Count Results

BLAST | BioMart | FTP | Docs & FAQs

URL XML Perl Help

Please select columns to be included in the output and hit 'Results' when ready

Features Variant (Germline)
 Structures Sequences
 Homologues (Max select 6 orthologues)

GENE:

Ensembl

Gene stable ID Transcript length (including UTRs and CDS)
 Transcript stable ID Ensembl Canonical
 Gene start (bp) Gene name
 Gene end (bp) Source of gene name
 Protein stable ID Transcript name
 Exon stable ID Source of transcript name
 Gene description Transcript count
 Chromosome/scaffold name Gene % GC content
 Gene start (bp) Gene type
 Gene end (bp) Transcript type
 Strand Source (gene)
 Karyotype band Source (transcript)
 Transcript start (bp) Gene Synonym
 Transcript end (bp)
 Transcription start site (TSS)

EXTERNAL:

PROTEIN DOMAINS AND FAMILIES:

Dataset 126 / 7127 Genes
Saccharomyces cerevisiae genes (R64-1-1)
Filters
Chromosome/scaffold: I
Attributes
Gene stable ID
Transcript stable ID
Gene start (bp)
Gene end (bp)
Gene name
Dataset [None Selected]

Biomart Interface - see results

e!EnsemblFungi

New Count Results

BLAST | BioMart | FTP | Docs & FAQs

★ URL | XML | Perl | Help

Dataset 126 / 7127 Genes
Saccharomyces cerevisiae genes (R64-1-1)

Filters
Chromosome/scaffold: I

Attributes
Gene stable ID
Transcript stable ID
Gene start (bp)
Gene end (bp)
Gene name

Dataset
[None Selected]

Export all results to File TSV Unique results only Go

Email notification to

View 10 rows as HTML Unique results only

Gene stable ID	Transcript stable ID	Gene start (bp)	Gene end (bp)	Gene name
YAL040C	YAL040C mRNA	65778	67520	CLN3
YAL007C	YAL007C mRNA	137698	138345	ERP2
HRA1	HRA1 ncRNA	99305	99868	
YAR028W	YAR028W mRNA	184892	185596	
YAL064C-A	YAL064C-A mRNA	13363	13743	TDA8
YAL012W	YAL012W mRNA	130799	131983	CYS3
YAL058W	YAL058W mRNA	37464	38972	CNE1
YAR068W	YAR068W mRNA	222406	222891	
YAL044C	YAL044C mRNA	57950	58462	GCV3
YAR066W	YAR066W mRNA	221049	221660	

Ensembl Fungi: Molecular interactions

Introduction to database queries,
Fungal Pathogen Genomics 2025

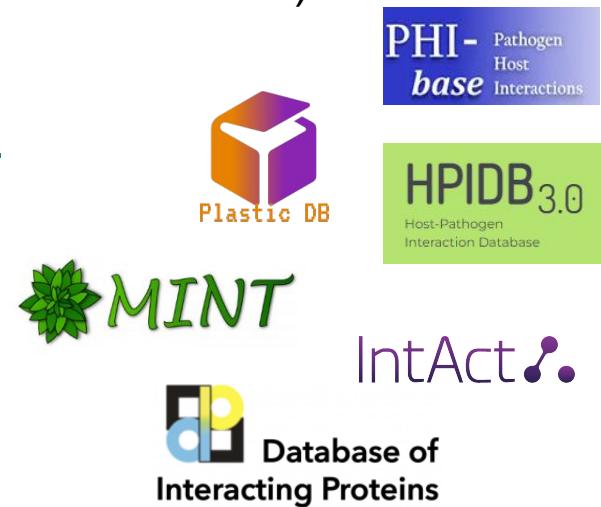
Manuel Carbajo
Ensembl Outreach



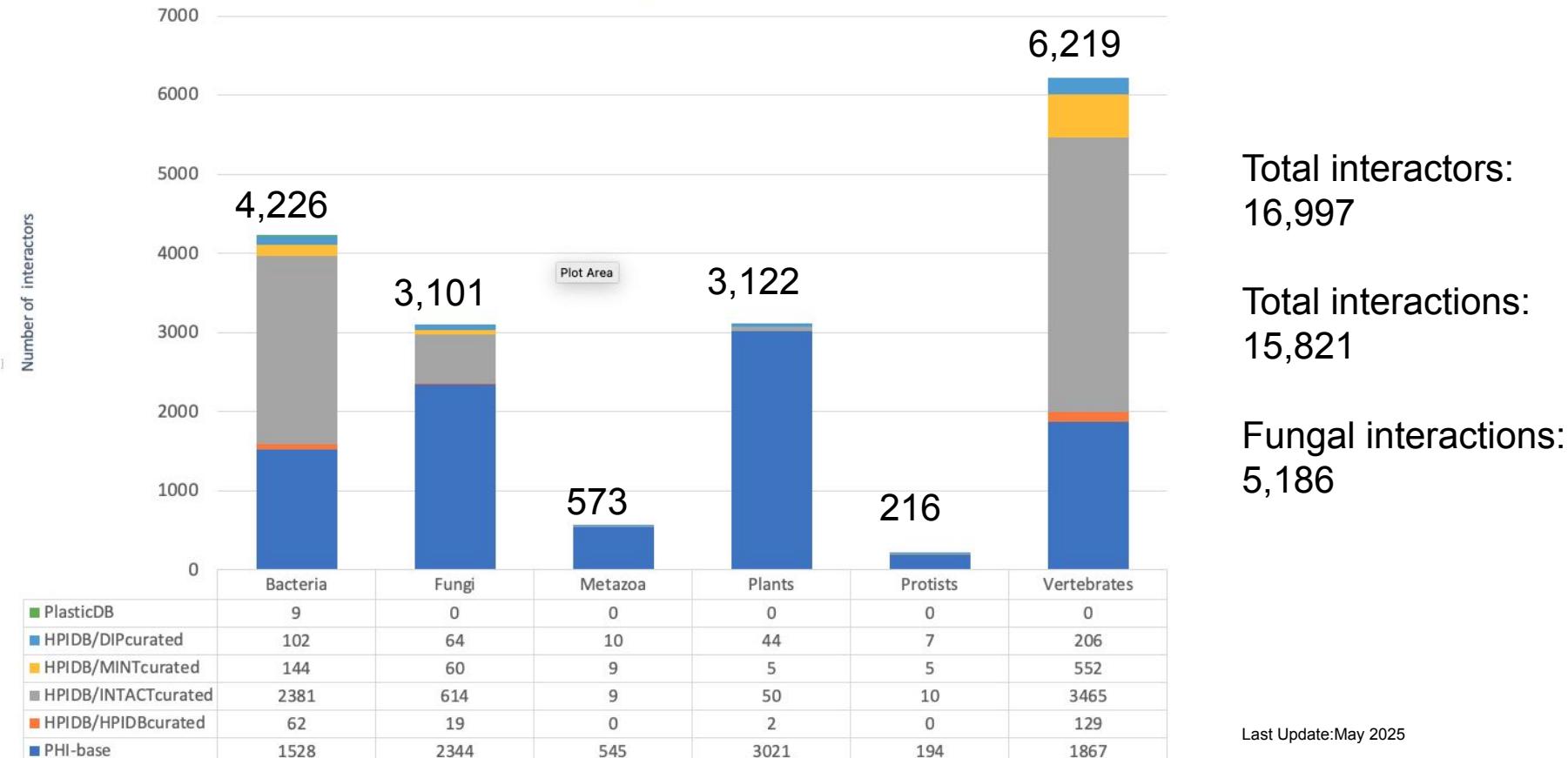
Cross-species molecular interactions

A new Ensembl resource:

- 2 molecules + experimentally verified + published + curated
- Cross-species (species-species OR species-synthetic molecules)
- Imported from external interaction databases:
[DIP](#), [HPIDB](#), [IntAct](#), [MINT](#), [PHI-base](#), [PlasticDB](#)
- Conservative import of exact protein/strain matches
- Data interlaced with other Ensembl resources



Unique interactors per division and source database



Ensembl staff at this course



Manuel Carbajo

Ensembl Microbes

Aleena Mushtaq

Ensembl

Outreach

Jorge Batista da

Rocha

Ensembl Outreach

Special thanks:

Nishadi De Silva

Leaving

Ensembl Microbes

Accessing molecular interactions

The screenshot shows a gene page for *Zymoseptoria tritici* (Mycorrhizal). The main content area displays the gene details for *Mycorrhizal* (Mycorrhizal), including its chromosomal location (Chromosome 11, 105,887,894-105,887,905) and its role in MUCORRHIZAL 1. Below this, there is a 'Molecular Interactions' section. This section includes a table of interactions with other genes and a detailed view of one interaction. The detailed view shows the interaction type as 'Protein-protein interaction', involving *Zymoseptoria tritici* (Mycorrhizal) and *Phoma exigua*. The table lists various experimental assays such as 'Protein-protein interaction', 'Protein-DNA interaction', 'Protein-RNA interaction', 'Pathogen protein modification', 'Pathogen nucleic acid level assay', 'Pathogen experimental strain', and 'Other experimental assay'. Each row in the table provides a link to the assay details.

Target	Gene ID	Interaction	Strain
<i>Zymoseptoria tritici</i> (Mycorrhizal)	Mycorrhizal	protein	<i>Phoma exigua</i>

Target	Gene ID	interaction	Strain
<i>Zymoseptoria tritici</i> (Mycorrhizal)	Mycorrhizal	protein	<i>Phoma exigua</i>

Detailed information for the first interaction:

Target	Gene ID	interaction	Strain
<i>Zymoseptoria tritici</i> (Mycorrhizal)	Mycorrhizal	protein	<i>Phoma exigua</i>

Interaction type: Protein-protein interaction
Interaction phenotype: PHYSIO-00000000
Strain name: PHYSIO-00000000
Pathogen protein modification: gene knockout; full
PHI base high level assay: 0.0000
Pathogen experimental strain: PT0000
Other experimental assay: no assay found

Detailed information for the second interaction:

Target	Gene ID	interaction	Strain
<i>Zymoseptoria tritici</i> (Mycorrhizal)	Mycorrhizal	protein	<i>Phoma exigua</i>

Interaction type: Protein-protein interaction
Interaction phenotype: PHYSIO-00000000
Strain name: PHYSIO-00000000
Pathogen protein modification: gene knockout; full
PHI base high level assay: 0.0000
Pathogen experimental strain: PT0000
Other experimental assay: no assay found

Accessing molecular interactions (browser)

EnsemblFungi ▾ HMMER | BLAST | BioMart | Tools | Downloads | Help & Docs | Blog

Phaeosphaeria nodorum (ASM14691v1) ▾

Location: supercont1.101,101,626-1,102,347 Gene: SNOG_20078 Transcript: SNOT_20078

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- Sequence
 - Secondary Structure
 - Gene families
 - Literature
- Fungal Compara
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Pan-taxonomic Compara
 - Gene Tree
 - Orthologues
- Ontologies
 - GO: Biological process
 - GO: Molecular function
 - GO: Cellular component
 - PHI: Phibase identifier
- Phenotypes
- Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
 - Gene expression
 - Pathway
- Molecular interactions
 - Regulation
 - External references
 - Supporting evidence
- ID History
 - Gene history

Gene: SNOG_20078

Description Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:[A9JX75|e](#)]

Location SuperContig supercont1.101,101,626-1,102,347 reverse strand.
ASM14691v1.CH445334.1

About this gene This gene has 1 transcript ([splice variant](#)).

Transcripts [Show transcript table](#)

Molecular interactions Cross-species interactions imported from PHI-base, HPIDB and PlasticDB with exact matches to proteins in Ensembl.

This species		Interacts with		Show metadata
Species	Gene ID	Interactor	Identifier	Source DB
Phaeosphaeria nodorum	SNOG_20078	protein	uniprot:A9JX75	
SN15				

Species	Gene ID	Interactor	Identifier	Source DB
Triticum aestivum	UNDETERMINED	protein	UNDETERMINED	PHI-base
Triticum aestivum	TraesCS1B02G004100	protein	uniprot:W5AB81	PHI-base
Triticum aestivum	TraesCS1B02G004100	protein	uniprot:W5AB81	PHI-base

Ensembl Fungi release 56 - Feb 2023 © EMBL-EBI

About Us

- About us
- Contact us
- Citing Ensembl Genomes

Get help

- Using this website
- Documentation
- Adding custom tracks

Our sister sites

- Ensembl
- Ensembl Bacteria
- Ensembl Plants
- Ensembl Protists
- Ensembl Metazoa

Follow us

- [Blog](#)
- [Twitter](#)

[Configure this page](#)

[Custom tracks](#)

[Export data](#)

[Share this page](#)

Accessing molecular interactions (REST API)

Swagger
Supported by SMARTBEAR

https://interactions.rest.ensembl.org/?format=openapi

Explore

API interactions doc v1

[Base URL: interactions.rest.ensembl.org/]
<https://interactions.rest.ensembl.org/?format=openapi>

Ensembl Interactions REST API
A list of endpoints to provide access to the Ensembl molecular interactions database.
We welcome your suggestions to add endpoints to fit new use case scenarios. Please write to helpdesk@ensemblgenomes.org

Schemes
HTTPS ▾

Authorize

Filter by tag

display_by_gene

GET /display_by_gene/{ens_stbl_id} display_by_gene_read ▾

ensembl_gene

GET /ensembl_gene ensembl_gene_list ▾

GET /ensembl_gene/ensembl_name ensembl_gene_ensembl_name_list ▾

Accessing molecular interactions

- API REST endpoints: <https://interactions.rest.ensembl.org/interaction>

?species_A=zymoseptoria tritici

&species_B=Triticum aestivum

&meta_key=disease

Filters

Field filters

Interaction id:

interactor name:

ensembl stable id:

species A:
zymoseptoria tritici

species B:
triticum aestivum

Source db:

meta value:

meta key:
disease

Submit



```
[  
  {  
    "interaction_id": 18174,  
    "interactor_1": "uniprot:F9WR1",  
    "interactor_2": "uniprot:UNDETERMINED_PHI:4966_Triticum_aestivum",  
    "ensembl_gene_1": "Mycgr3G53658",  
    "ensembl_gene_2": "UNDETERMINED_PHI:4966_Triticum_aestivum",  
    "species_1": "zymoseptoria tritici",  
    "species_2": "Triticum aestivum",  
    "doi": "26092798",  
    "source_db": "PHI-base"  
  },  
  {  
    "interaction_id": 18254,  
    "interactor_1": "uniprot:F9X056",  
    "interactor_2": "uniprot:UNDETERMINED_PHI:2442_Triticum_aestivum",  
    "ensembl_gene_1": "Mycgr3G88451",  
    "ensembl_gene_2": "UNDETERMINED_PHI:2442_Triticum_aestivum",  
    "species_1": "zymoseptoria tritici",  
    "species_2": "Triticum aestivum",  
    "doi": "19522561",  
    "source_db": "PHI-base"  
  },  
  {  
    "interaction_id": 18277,  
    "interactor_1": "uniprot:F9X7N2",  
    "interactor_2": "uniprot:UNDETERMINED_PHI:4943_Triticum_aestivum",  
    "ensembl_gene_1": "Mycgr3G85040",  
    "ensembl_gene_2": "UNDETERMINED_PHI:4943_Triticum_aestivum",  
    "species_1": "zymoseptoria tritici",  
    "species_2": "Triticum aestivum",  
    "doi": "26092796",  
    "source_db": "PHI-base"  
  }  
]
```

Ensembl Fungi: Custom tracks

Transcriptomics & proteomics
Fungal Pathogen Genomics 2023

Jorge Batista da Rocha and Aleena Mushtaq
Ensembl Outreach

Sunday, 2nd June 2023



Track Hubs

- The [Track Hub registry](#) is a way of publicly sharing large amounts of data
- Many large collaborative projects make their data available in sets of tracks organised as a single hub
- You can create a Track Hub containing your own data to share with the public and collaborators
- You can search for data in the Track Hub registry and view this data directly on Ensembl

Visualise your own data

Upload data:

- Data saved by Ensembl
- 20 MB limit (not possible for large files)

Attach remote file:

- URL-based (HTTP or FTP)
- No size-limit

Data visualisation formats

Genes and features

- GBrowse
- GFF / GTF
- BigBed

Sequence alignments

- PSL
- BAM
- CRAM

Continuous values

- BedGraph
- WIG
- BigWig

Variants

- VCF

Where to host URLs for larger files?

If your file is >20 MB in size, you will need to attach it by URL. You can create a URL using:

- Institute FTP site
- [CyVerse](#)
- [FigShare](#)

Note that Google Drive will not work and GitHub has proven unreliable.

Exercise: Attaching Track Hubs to Ensembl Fungi

The **RNA-Seq alignment hub SRP062592** track in the TrackHub Registry is currently unavailable. Instead, please use the URL below (this is the same file):

<https://ftp.sra.ebi.ac.uk/vol1/analysis/ERZ492/ERZ492288/SRR2174493.cram>

Ensembl Fungi: Ensembl VEP

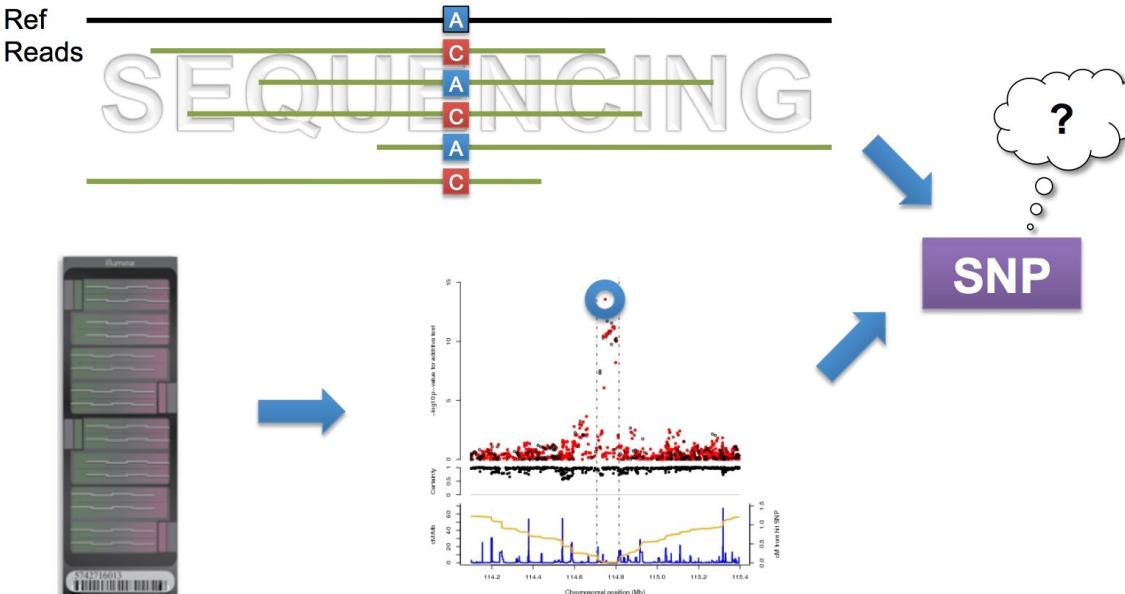
**SNPs & Variants
Fungal Pathogen Genomics**

Jorge Batista da Rocha and Aleena Mushtaq
Ensembl Outreach



What is the Variant Effect Predictor (VEP) for?

A tool to predict and annotate the **functional consequences** of variants (SNPs, insertions, deletions, CNVs, or structural variants).

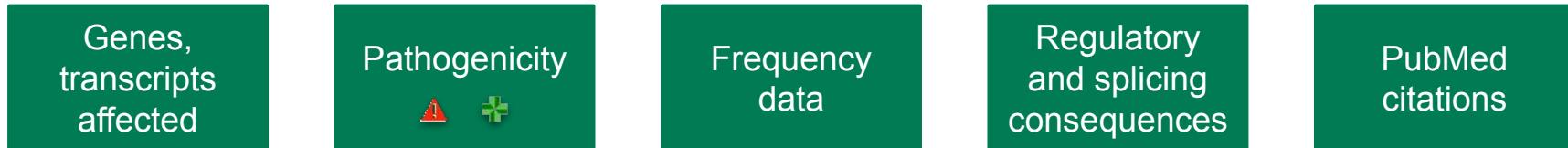


What can you do with the VEP?

Data input:



VEP output:



Where does the data come from?

The Ensembl variation process

Variant import

Quality control

Linked data

Ensembl analysis

Input formats

Ensembl default

VEP input:

```
#chr      start    end allele strand id
1   65568    65568   A/C 1
2   265023   265023  C/T 1
3   319781   319781  A/- 1
```

VCF:

```
#chr      pos id  ref alt qual    filt    info
1   65568   .   A   C   .   .   .
2   265023   .   C   T   .   .   .
3   319780   .   GA  G   .   .   .
```

Input formats

HGVS notation:

ENST00000641515.2:c.4A>C
ENST00000272065.10:c.43+16C>T
ENST00000256509.7:c.5del
9:g.22125504G>C

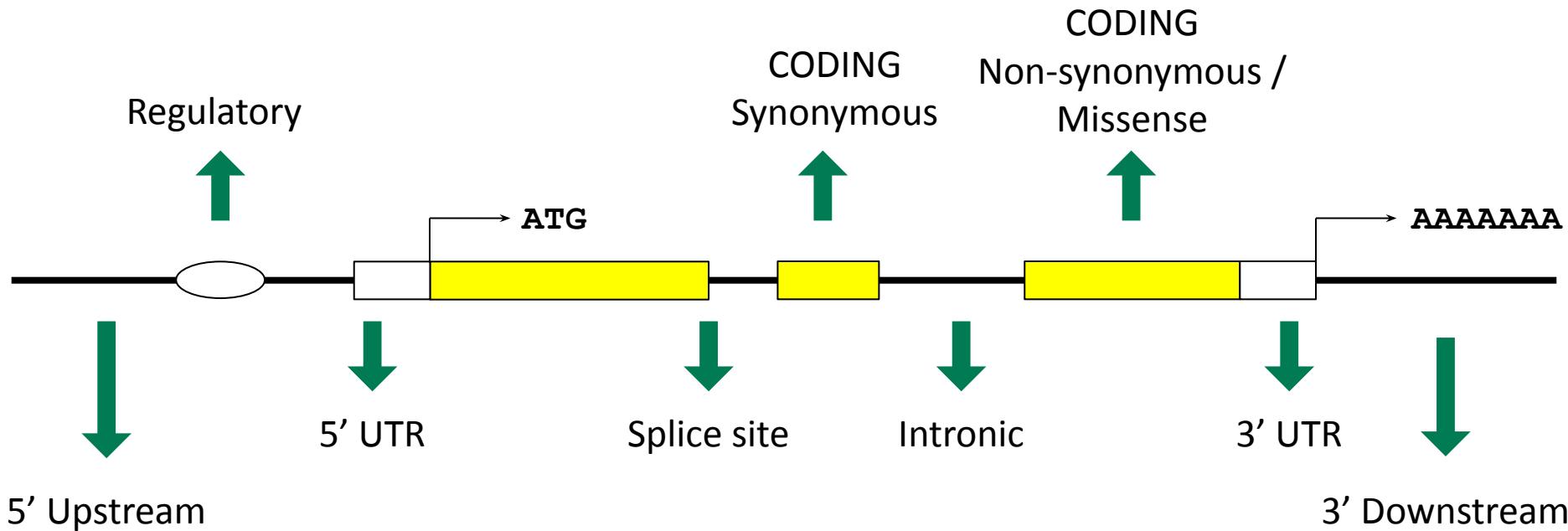
Variant IDs:

rs1156485833
COSM327779
rs867704559

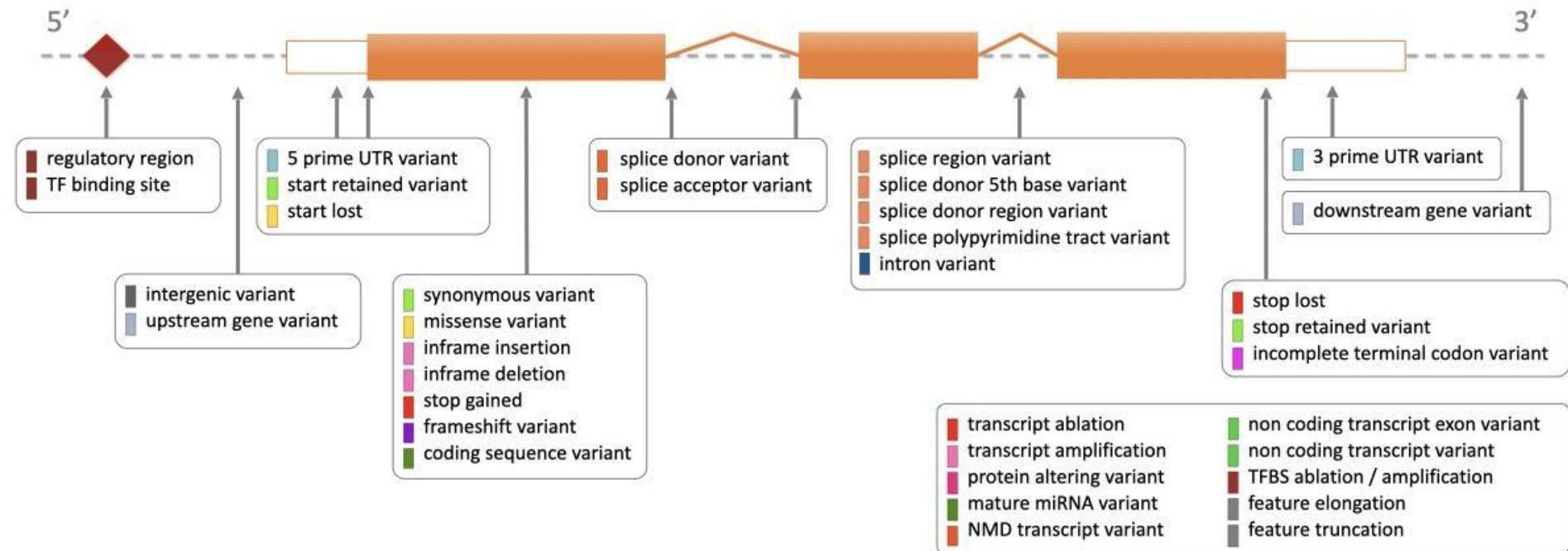
SPDI:

NC_000001.11:65567:A:C
NC_000002.12:265022:C:T
NC_000003.12:319780:A:

Variation consequences



Sequence ontology (SO)



Sequence ontology (SO)

SO term	SO description	SO accession	Display term	IMPACT
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	SO_0001893	Transcript ablation	HIGH
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron	SO_0001574	Splice acceptor variant	HIGH
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron	SO_0001575	Splice donor variant	HIGH
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO_0001587	Stop gained	HIGH
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO_0001588	Frameshift variant	HIGH
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO_0001578	Stop lost	HIGH
start_lost	A codon variant that changes at least one base of the canonical start codon	SO_0002012	Start lost	HIGH
transcript_amplification	A feature amplification of a region containing a transcript	SO_0001889	Transcript amplification	HIGH
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence	SO_0001821	Inframe insertion	MODERATE
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO_0001822	Inframe deletion	MODERATE
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO_0001583	Missense variant	MODERATE
protein_altering_variant	A sequence variant which is predicted to change the protein encoded in the coding sequence	SO_0001818	Protein altering variant	MODERATE
splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron	SO_0001630	Splice region variant	LOW
splice_donor_5th_base_variant	A sequence variant that causes a change at the 5th base pair after the start of the intron in the orientation of the transcript	SO_0001787	Splice donor 5th base variant	LOW
splice_donor_region_variant	A sequence variant that falls in the region between the 3rd and 6th base after splice junction (5' end of intron)	SO_0002170	Splice donor region variant	LOW
splice_polyprymidine_tract_variant	A sequence variant that falls in the polyprymidine tract at 3' end of intron between 17 and 3 bases from the end (acceptor -3 to acceptor -17)	SO_0002169	Splice polyprymidine tract variant	LOW
incomplete_terminal_codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed	SO_0001626	Incomplete terminal codon variant	LOW
start_retained_variant	A sequence variant where at least one base in the start codon is changed, but the start remains	SO_0002019	Start retained variant	LOW
stop_retained_variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains	SO_0001567	Stop retained variant	LOW
synonymous_variant	A sequence variant where there is no resulting change to the encoded amino acid	SO_0001819	Synonymous variant	LOW
coding_sequence_variant	A sequence variant that changes the coding sequence	SO_0001580	Coding sequence variant	MODIFIER
mature_miRNA_variant	A transcript variant located with the sequence of the mature miRNA	SO_0001620	Mature miRNA variant	MODIFIER
5_prime_UTR_variant	A UTR variant of the 5' UTR	SO_0001623	5 prime UTR variant	MODIFIER
3_prime_UTR_variant	A UTR variant of the 3' UTR	SO_0001624	3 prime UTR variant	MODIFIER
non_coding_transcript_exon_variant	A sequence variant that changes non-coding exon sequence in a non-coding transcript	SO_0001792	Non coding transcript exon variant	MODIFIER
intron_variant	A transcript variant occurring within an intron	SO_0001627	Intron variant	MODIFIER
NMD_transcript_variant	A variant in a transcript that is the target of NMD	SO_0001621	NMD transcript variant	MODIFIER
non_coding_transcript_variant	A transcript variant of a non coding RNA gene	SO_0001619	Non coding transcript variant	MODIFIER
upstream_gene_variant	A sequence variant located 5' of a gene	SO_0001631	Upstream gene variant	MODIFIER
downstream_gene_variant	A sequence variant located 3' of a gene	SO_0001632	Downstream gene variant	MODIFIER
TFBS_ablation	A feature ablation whereby the deleted region includes a transcription factor binding site	SO_0001895	TFBS ablation	MODIFIER
TFBS_amplification	A feature amplification of a region containing a transcription factor binding site	SO_0001892	TFBS amplification	MODIFIER
TF_binding_site_variant	A sequence variant located within a transcription factor binding site	SO_0001782	TF binding site variant	MODIFIER
regulatory_region_ablation	A feature ablation whereby the deleted region includes a regulatory region	SO_0001894	Regulatory region ablation	MODIFIER
regulatory_region_amplification	A feature amplification of a region containing a regulatory region	SO_0001891	Regulatory region amplification	MODIFIER
feature_elongation	A sequence variant that causes the extension of a genomic feature, with regard to the reference sequence	SO_0001907	Feature elongation	MODIFIER
regulatory_region_variant	A sequence variant located within a regulatory region	SO_0001566	Regulatory region variant	MODIFIER
feature_truncation	A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence	SO_0001906	Feature truncation	MODIFIER
intergenic_variant	A sequence variant located in the intergenic region, between genes	SO_0001628	Intergenic variant	MODIFIER

Species with variation data

Name	Classification	Taxon ID	Assembly
 Fusarium graminearum str. PH-1	Fungi	5518	RR1
 Fusarium oxysporum	Hypocreales	426428	FO2
 Puccinia graminis	Pucciniales	418459	ASM14992v1
 Puccinia graminis Ug99	Pucciniales	1040990	v1
 Saccharomyces cerevisiae	Saccharomycetales	559292	R64-1-1
 Schizosaccharomyces pombe	Schizosaccharomycetales	284812	ASM294v2
 Verticillium dahliae JR2	Glomerellales	1202531	VDAG_JR2v.4.0
 Zymoseptoria tritici	Capnodiales	336722	MG2

Use the Ensembl VEP



Web interface

- Point-and-click interface
- Suits smaller volumes of data

[Documentation](#)

[Launch the web interface](#)



Standalone perl script

- More options, more flexibility
- For large volumes of data

[Documentation](#)

[Download latest version](#)



REST API

- Language-independent API
- Simple URL-based queries
- GET single variants, POST many

[Documentation](#)

Ensembl VEP command-line

- Can handle genome-wide data
- Can work 100% offline
- A lot of options and plugins
- Use your own genome (custom annotations)

Ensembl VEP plugins

- Plugins add extra functionality to VEP
- They may extend, filter or manipulate the output of VEP
- Plugins may make use of external data or code
- Available on the web interface and command line



Ensembl Fungi: Compara

**Comparative genomics & orthology
Fungal Pathogen Genomics 2025**

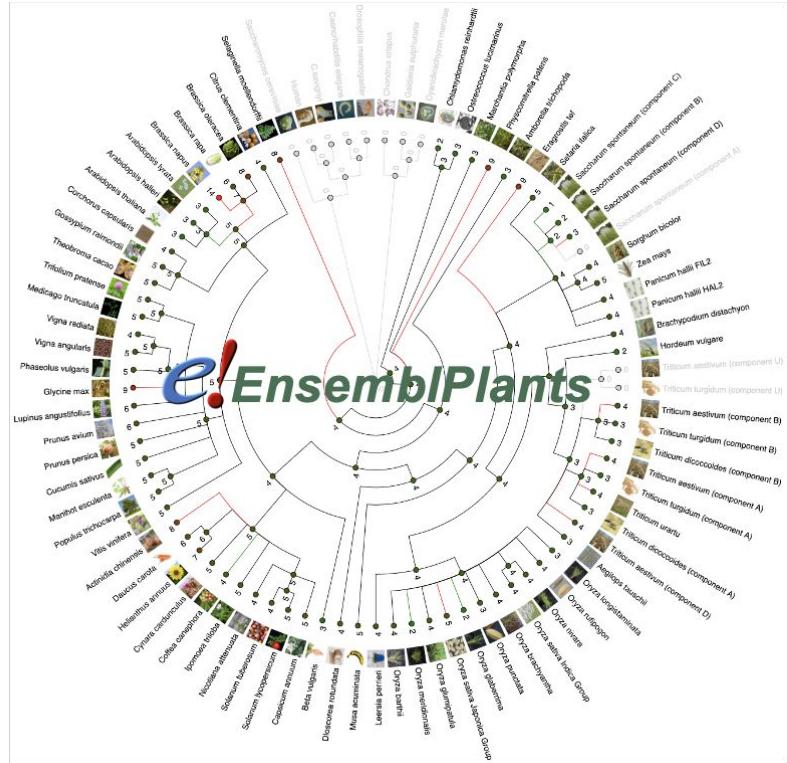
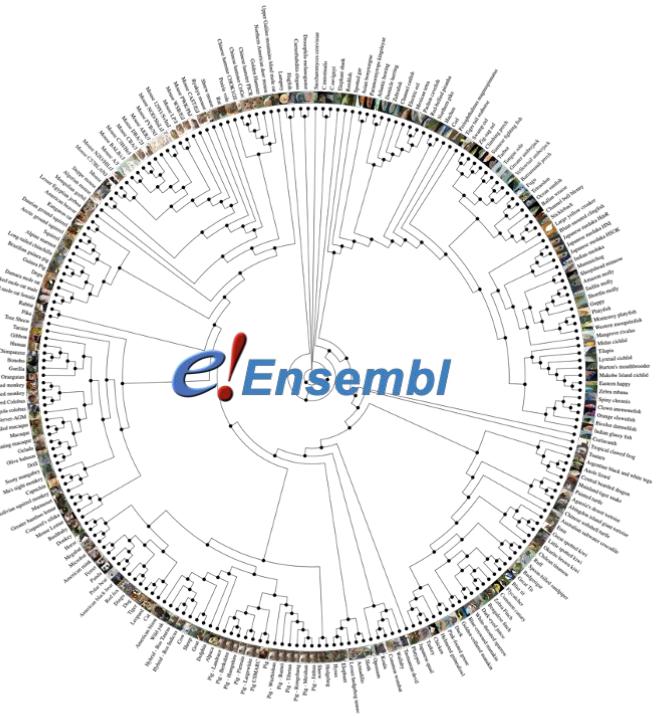
Jorge Batista da Rocha and Aleena Mushtaq
Ensembl Outreach



Overview

- Comparative genomics:
 - Gene trees
 - Homology predictions
- Whole-genome alignments (WGA):
 - Pairwise alignments
 - Shared synteny

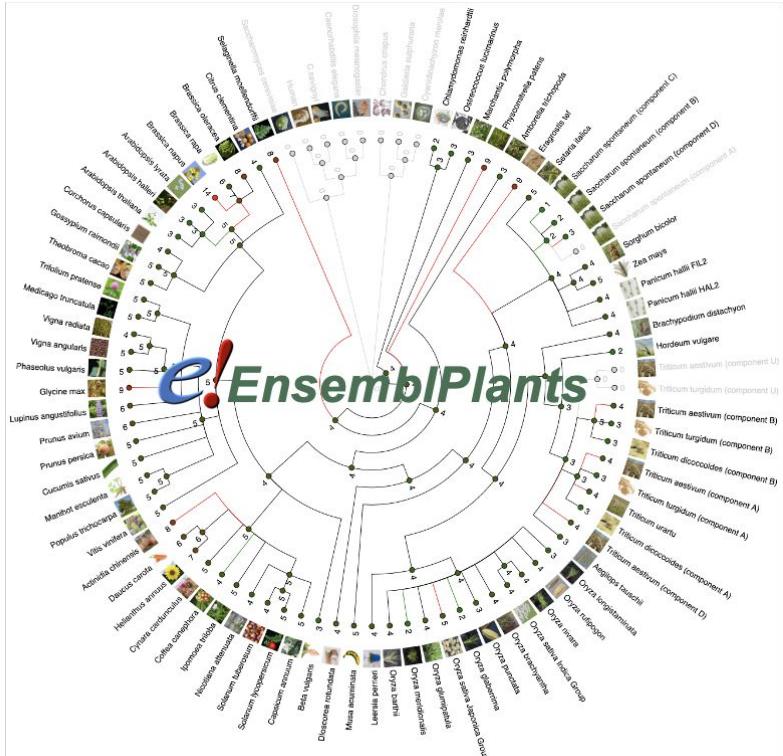
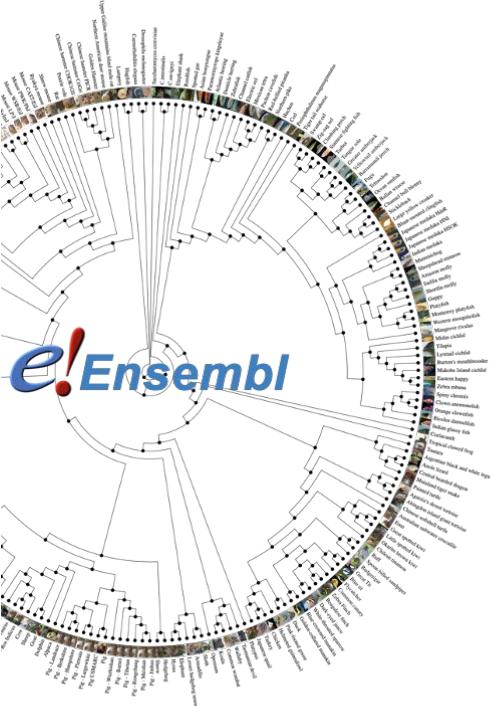
Comparative analysis by taxa



Comparative analysis by taxa

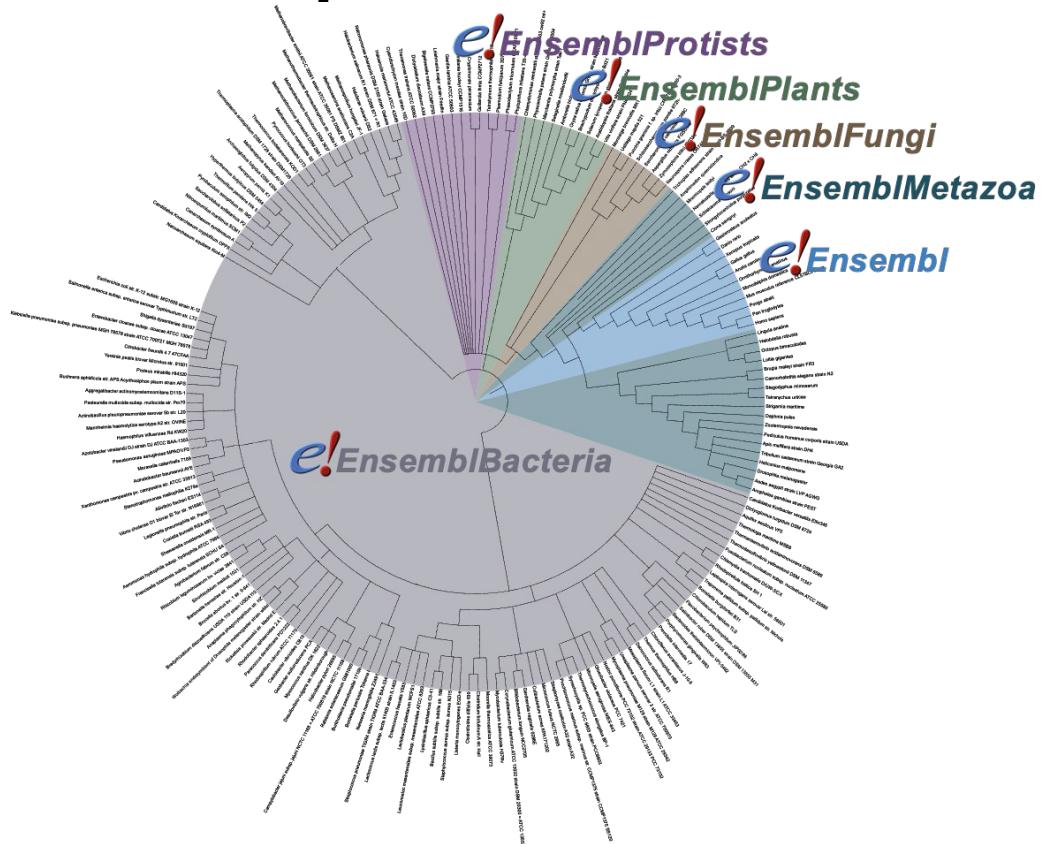
Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- Sequence
 - Secondary Structure
 - Gene families
 - Literature
- Fungal Compara
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Pan-taxonomic Compara
 - Gene Tree
 - Orthologues
- Ontologies
 - GO: Biological process
 - GO: Molecular function
 - GO: Cellular component
 - PHI: Phibase identifier
- Phenotypes
- Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
- Gene expression
- Pathway
- Molecular interactions
- Regulation
- External references
- Supporting evidence
- ID History
 - Gene history



https://github.com/WCSCourses/Fungal_Pathogen_Genomics_2025

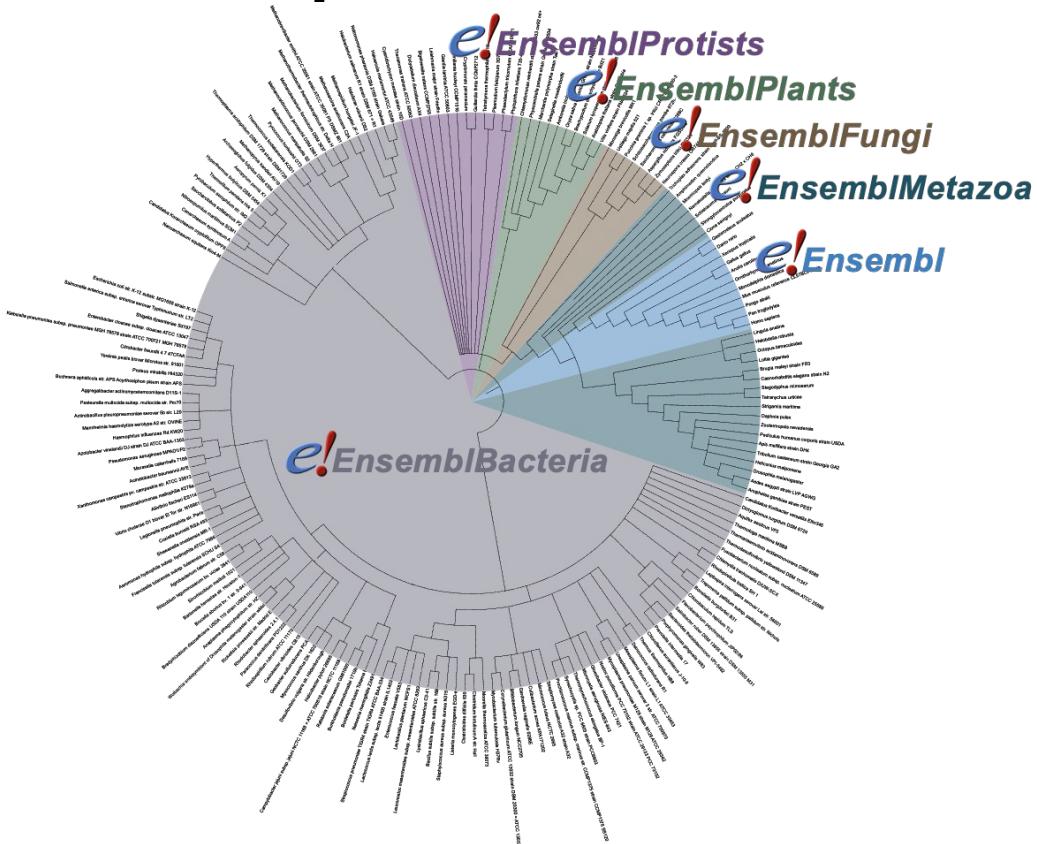
Pan-taxonomic compara



Pan-taxonomic compara

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- Sequence
 - Secondary Structure
 - Gene families
 - Literature
- Fungal Compara
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Pan-taxonomic Compara
 - Gene Tree
 - Orthologues
- Ontologies
 - GO: Biological process
 - GO: Molecular function
 - GO: Cellular component
 - PHI: Phibase identifier
- Phenotypes
- Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
- Gene expression
- Pathway
- Molecular interactions
- Regulation
- External references
- Supporting evidence
- ID History
 - Gene history



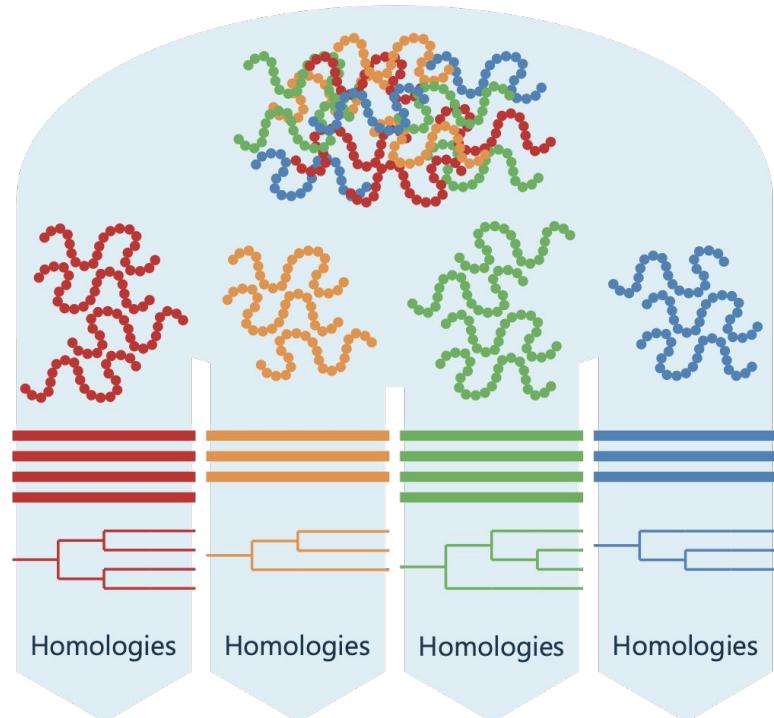
https://github.com/WCSCourses/Fungal_Pathogen_Genomics_2025

Pan-taxonomic compara

Name	Classification	Taxon ID	Assembly
 Aspergillus nidulans	Eurotiales	227321	ASM1142v1
 Neurospora crassa	Sordariales	367110	NC12
 Puccinia graminis	Pucciniales	418459	ASM14992v1
 Saccharomyces cerevisiae	Saccharomycetales	559292	R64-1-1
 Schizosaccharomyces pombe	Schizosaccharomycetales	284812	ASM294v2
 Ustilago maydis	Ustilaginales	237631	Umaydis521_2.0
 Zymoseptoria tritici	Capnodiales	336722	MG2

Gene / protein trees

1. Representative translation of each gene from all species
2. All-vs-all HMM search to classify into families or clustering
3. Multiple protein alignment
4. Phylogenetic tree for each aligned cluster and reconciliation against NCBI taxonomy
5. Ortho- / parologue inference



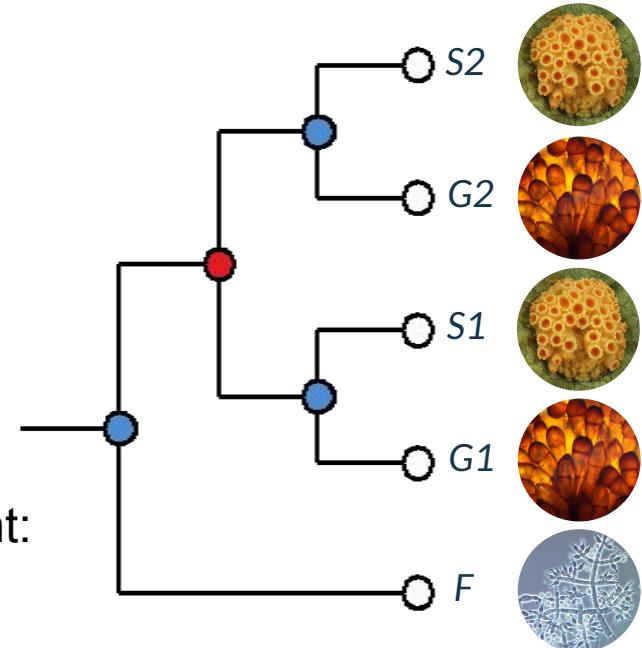
Homology

Orthologues:

- Genes emerged through a **speciation** event:
 - e.g. G1 and S1; G2 and F; S2 and F
- 1-to-1: G1 and S1
- 1-to-many: F and S1, S2, G1, G2

Paralogues:

- Genes emerged through a **duplication** event:
 - e.g. G1 and G2, S1 and S2
- Within species: G1 and G2
- Between species: G1 and S2



Whole-genome alignments (WGA)

- To identify highly conserved regions
 - Sequences that evolve slowly
 - Regions likely to be functional
 - Both coding and non-coding
- To support problematic gene predictions
- To define syntenic regions

Pairwise WGA

Pairwise WGA with BLASTZ / LASTZ and tBLAT:

- Related species vs each other:
Ashbya, Fusarium, Magnaporthe, Puccinia, Saccharomyces, Schizosaccharomyces, Trichoderma



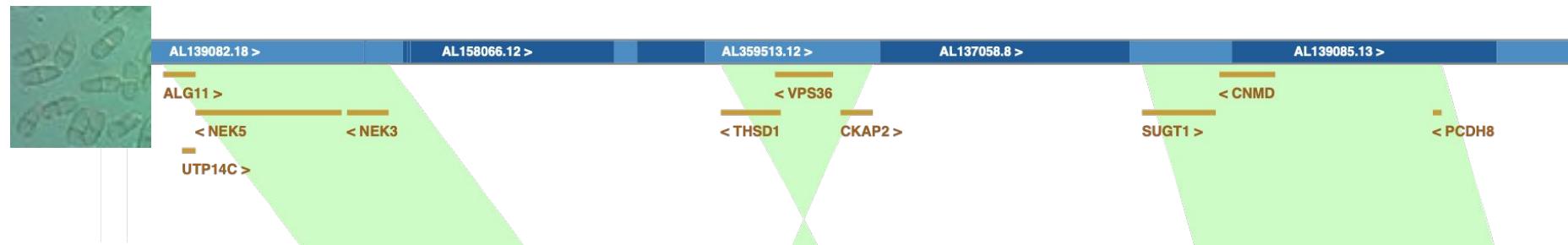
https://fungi.ensembl.org/info/genome/compara/compara_analyses.html

https://github.com/WCSCourses/Fungal_Pathogen_Genomics_2025

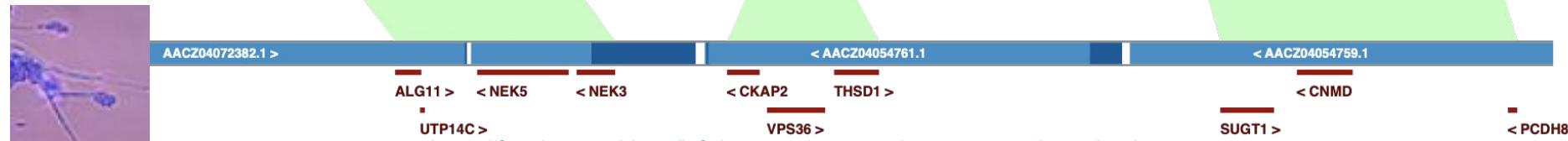
Shared synteny

Conserved order of aligned homologous genomic blocks between species (irrespective of orientation):

Fungus A



Fungus B



https://fungi.ensembl.org/info/genome/compara/compara_analyses.html

https://github.com/WCSCourses/Fungal_Pathogen_Genomics_2025

Types of data

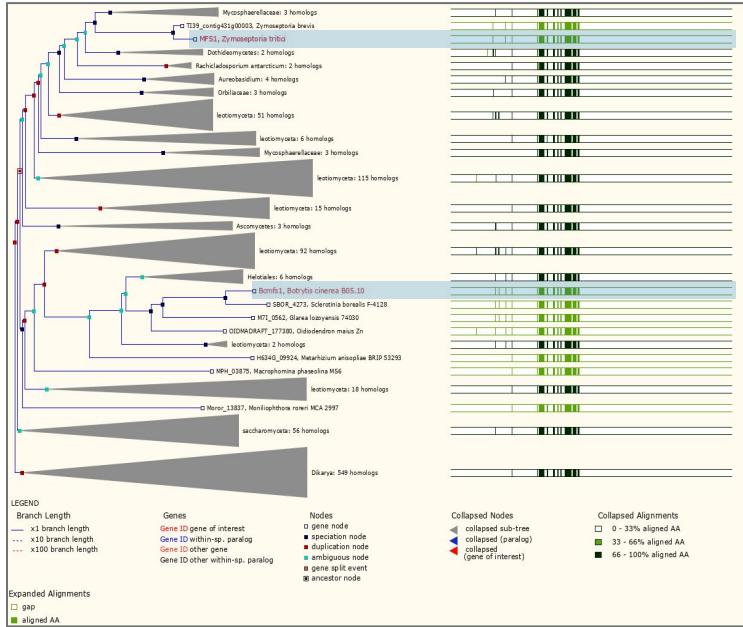
Gene-based resources
(under the **Gene** tab):

- Phylogenetic trees and tree-inferred homology
 - Gene / protein trees
 - ncRNA trees
- Stable ID mapping
- Protein families

Sequence-based resources
(under the **Location** tab):

- Whole-genome alignments
 - Ancestral sequences
 - Age of base
 - Conservation scores and constrained elements
 - Syntenies

Explore behaviour in other pathogens



e! EnsemblFungi



Botrytis cinerea infections

Pest Manag Sci. 2013 Jun;69(6):667-78. doi: 10.1002/ps.3506. Epub 2013 Apr 10.

French vineyards provide information that opens ways for effective resistance management of *Botrytis cinerea* (grey mould).

Walker A
Appl Environ Microbiol. 2013 Jan;79(1):159-67. doi: 10.1128/AEM.02655-12. Epub 2012 Oct 19.

Gray mold populations in german strawberry fields are resistant to multiple fungicides and dominated by a novel clade closely related to *Botrytis cinerea*.

Appl Environ Microbiol. 2002 Oct;68(10):4996-5004.

Bcmfs1, a novel major facilitator superfamily transporter from *Botrytis cinerea*, provides tolerance towards the natural toxic compounds camptothecin and cercosporin and towards fungicides.

Hayashi K¹, Schoonbeek HJ, De Waard MA.

e! Ensembl

https://github.com/WCSCourses/Fungal_Pathogen_Genomics_2025

EMBL-EBI

Recommend us

Browser course

One day course on the Ensembl browser, aimed at wet-lab scientists.

REST API course

Half day course on the Ensembl REST API, aimed at bioinformaticians.

Train the trainer course

One day course on delivering the Ensembl browser course.

We can teach an Ensembl course at any institute for free (plus trainers' expenses in high income countries).

helpdesk@ensembl.org

<https://training.ensembl.org/hosting>



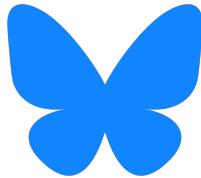
Reach out to us



[/Ensembl.org](https://www.facebook.com/Ensembl.org)



[/EnsemblGenomeBrowser](https://www.linkedin.com/company/ensemblgenomeweb)



@ensembl.bsky.social



@Ensembl



www.ensembl.info

Help and documentation

Courses online <https://www.ebi.ac.uk/training/>

Tutorials <https://www.ensembl.org/info/website/tutorials/>

Flash animations <https://www.youtube.com/user/EnsemblHelpdesk>

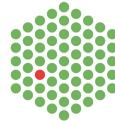
<https://youku.com/u/UMzM1NjkzMTI0>

Job opportunities <https://www.ebi.ac.uk/careers>

Email helpdesk@ensembl.org

Mailing lists dev@ensembl.org

announce@ensembl.org



Publications

- List of Ensembl publications:
<https://www.ensembl.org/info/about/publications.html>
- Topic-specific articles mentioned throughout the workshop

Dyer SC, Austine-Orimoloye O, Azov AG, Barba M, et al. [Ensembl 2025](#). Nucleic Acids Research. 2024 Dec;53(D1):D948–D957. DOI: 10.1093/nar/gkae1071. PMID: 39656687 PMCID: PMC11701638

Yates AD, Allen J, Amode RM, et al. [Ensembl Genomes 2022: an expanding genome resource for non-vertebrates](#). Nucleic Acids Research. 2022 Jan;50(D1):D996-D1003. DOI: 10.1093/nar/gkab1007. PMID: 34791415; PMCID: PMC8728113.

<https://www.ensembl.org/info/about/publications.html>

https://github.com/WCSCourses/Fungal_Pathogen_Genomics_2025

Ensembl staff at this course



Manuel Carbajo

Ensembl Microbes

Aleena Mushtaq

Ensembl

Outreach

Jorge Batista da

Rocha

Ensembl Outreach

Special thanks:

Nishadi De Silva

Leaving

Ensembl Microbes

Acknowledgements and funding



UK Research
and Innovation



National Human
Genome Research
Institute (NHGRI)



National Institutes of
Health (NIH)



Funded by the European Union
Horizon 2020
Horizon Europe Programme

Training materials

- Ensembl training materials are protected by a CC BY license:
<https://creativecommons.org/licenses/by/4.0/>
- If you wish to re-use these materials, please credit Ensembl for their creation.
- If you use Ensembl for your work, please cite our papers:
<https://www.ensembl.org/info/about/publications.html>

