

# Training materials

Ensembl materials are protected by a CC BY license



<http://creativecommons.org/licenses/by/4.0/>

If you wish to re-use these, please credit Ensembl for their creation

If you use Ensembl for your work, please cite our papers

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

# Plants and Pathogens in Ensembl

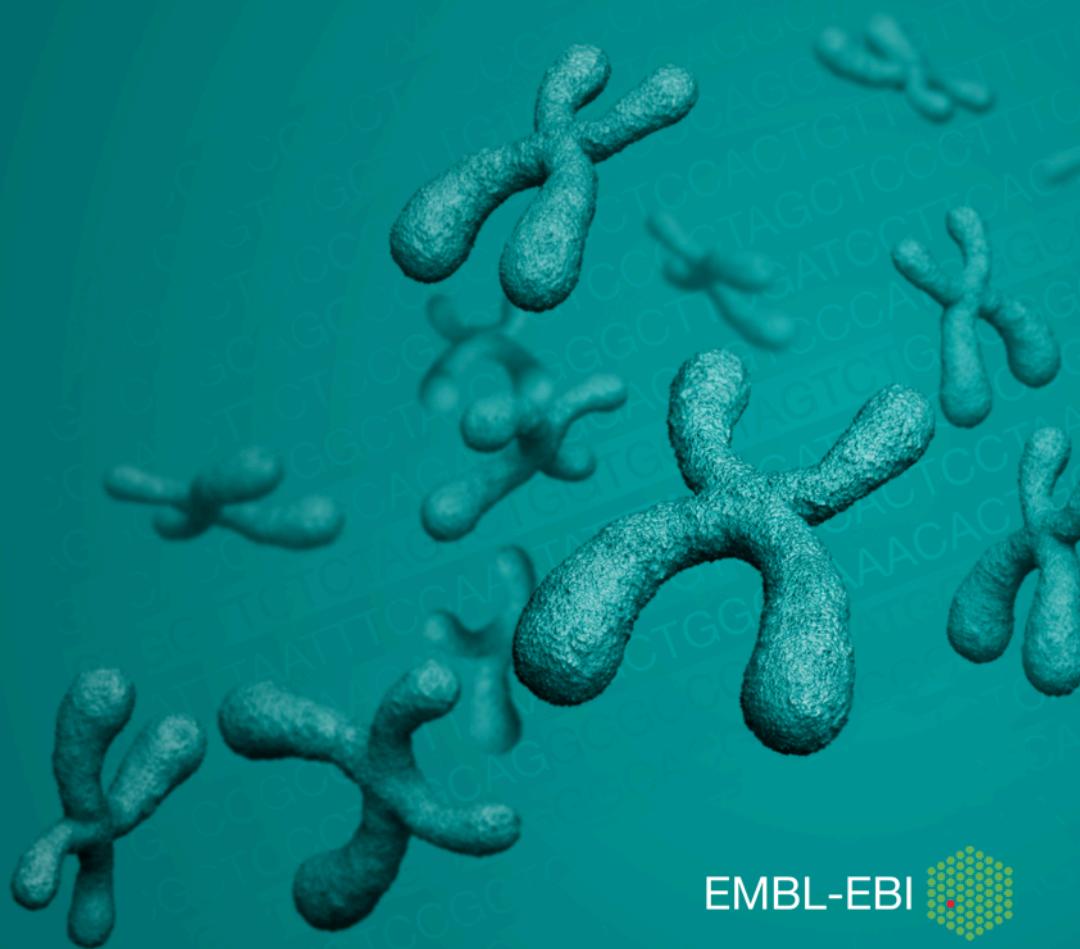
Bioinformatics of Plants and Plant pathogens

Hinxton, Cambridge

23-25 May 2016

**Denise Carvalho-Silva**

[ensemblgenomes.org](http://ensemblgenomes.org)  
@ensemblgenomes



# Today 10:30-12:30

- Introduction to Ensembl and Ensembl Genomes
- The Ensembl browser: live demo and hands-on
- Tools for omics analyses: live demo
- Wrap up



# Objectives: 2 hour session

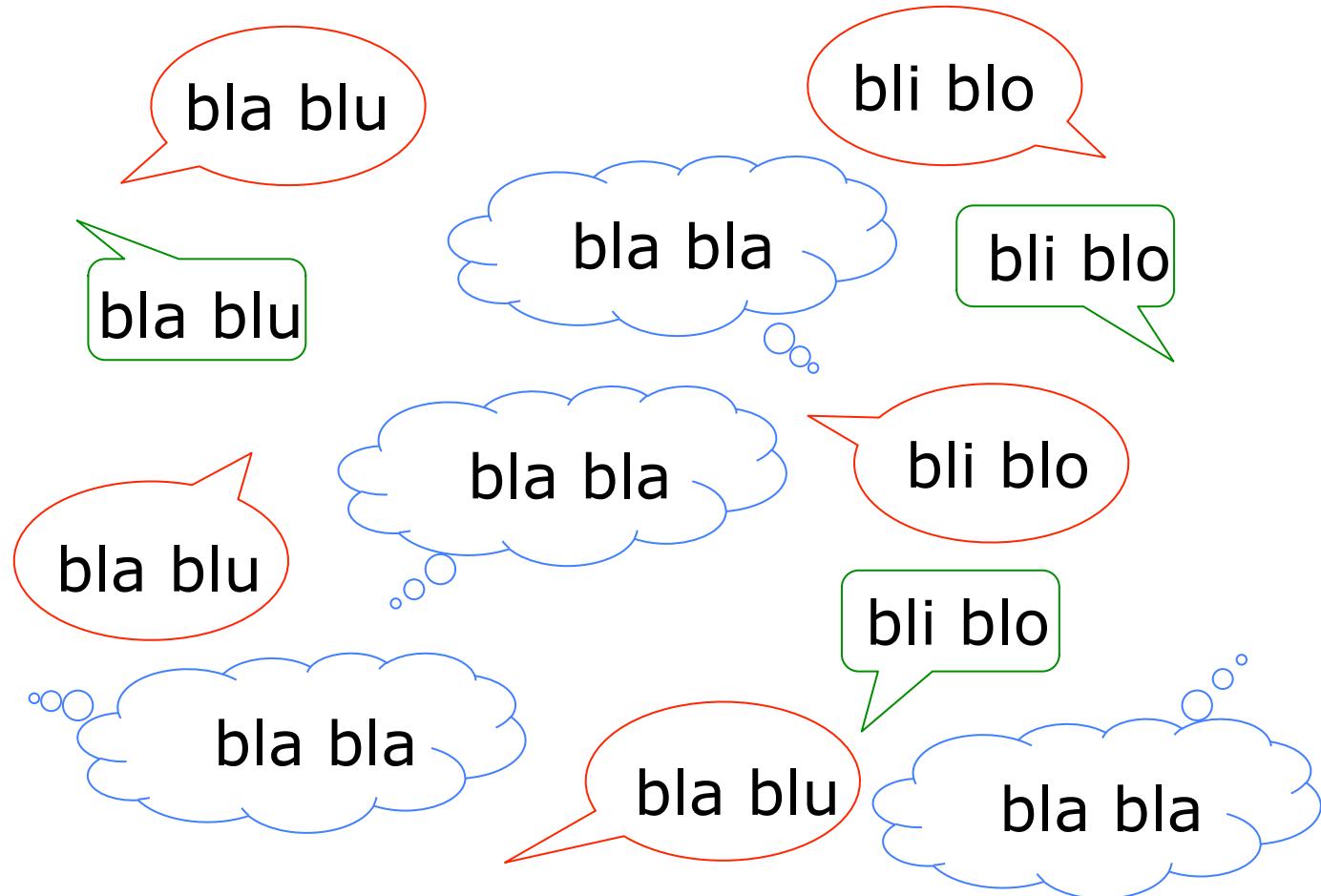
What is Ensembl?

What types of data can I get from Ensembl?

How to browse the Ensembl website?

How to connect with the Ensembl team





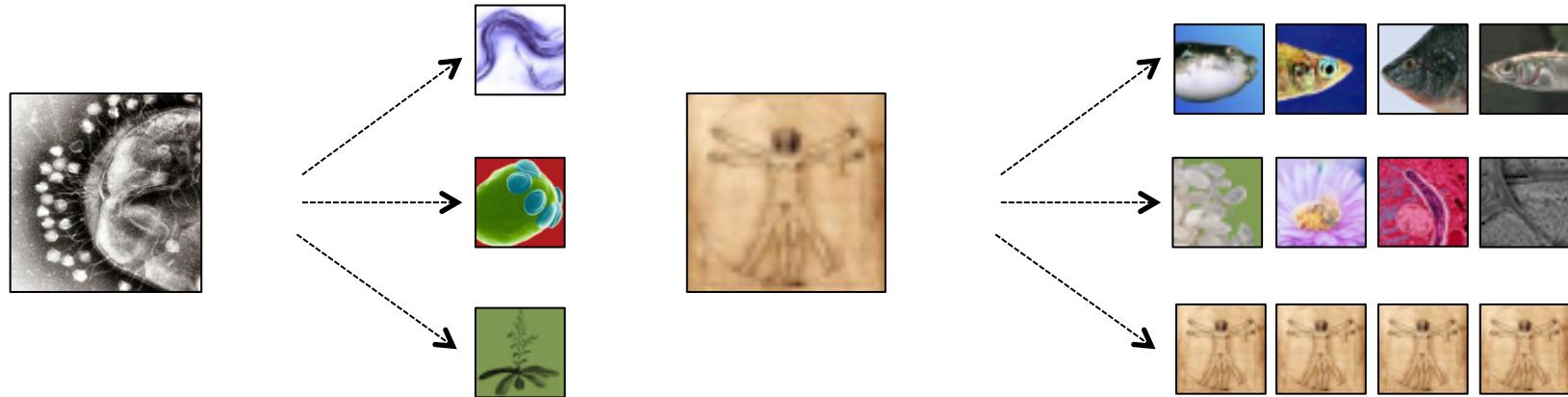
# Do We Need Genome Browsers?

YES

NO

1977: 1<sup>st</sup> genome to be sequenced (5 kb)

2000: draft human sequence (3 gb)



Large amounts of raw DNA sequence data

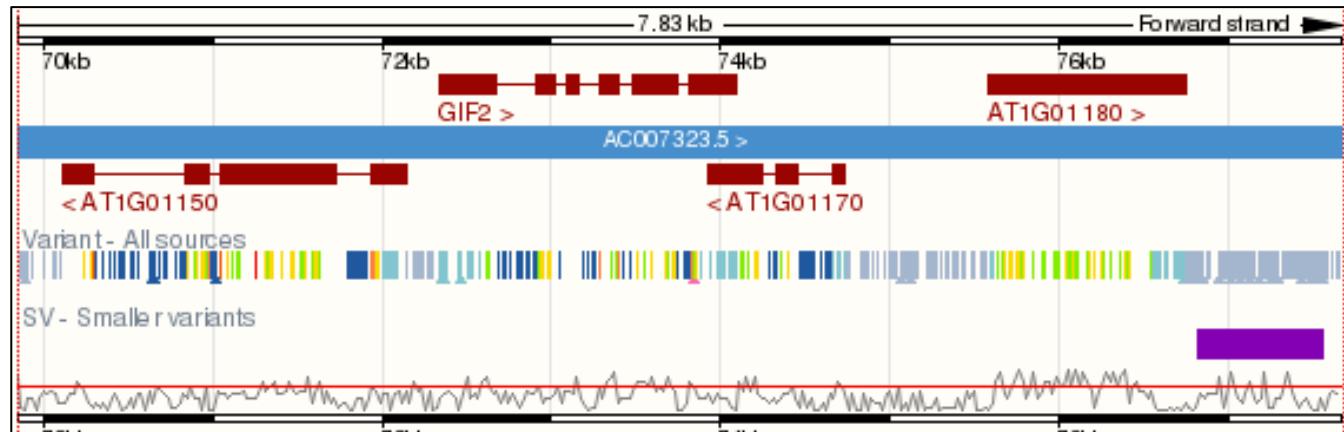
# Raw DNA sequence data

```
>chromosome:GRCh37:18:45357922:45457515:-1
GCCGGGAGGCAGGGCGGGCGTAGGCAAAGGGAGGTGGGAGGCAGGCGACTCCCCGCGCCCGCTCGC
CCCCCGGCCCTTCCCGCGGTGCTCGCCTCGTTCCCTCCGCTCCCTCCGCTTCCATACCCGCCCGC
GCGGCTTCGGCCGGCGTGCCTCGCCCTAACGGGCGGCTGGAGGCAGCAATCAGCAGGGCGGCAGGGTGCCAGC
CCCAGGGCTGCGCCGGCGAATCGCGGGGCCGCGGCCAGGGTGGCAGGGCGGTCTACCCGCGCGCGCGCG
GCGGAGAAGCAGCTGCCAGCCAGCAGCCCCGCCAGCCGCCGGAGGTGGTGGCGTGGCGCCGCCGGCG
CCGAGGGCGGAGGGCGGAAGCGGAGGTGGCTGGCGGGGGAGGGCGCGCCGTGCGGGCGGCCGTAGGGCTGCG
GGCGCGCGCCTGAGGGGAGGGAGGGCAGCGCGGGCGCGCCTCACCCCTCCTCCCCCGGGCGGCCGGCG
AGGCTCCCTCCCCCTCCCTCTCCTCCCCCTCCCCCTCCCCCTCTCTTCCCCTACCCCTCCCCCGCGCCGGCG
GCCGGCCGGGCCGGCCTGGGGCGGGGAAGACGGCGGCCGGAGTGTTCAGTTCCGCCTCCAATCGC
CCATTCCCCCTTCCCCCTCCCAGCCCCCTCCATCCCATCGGAAGAGGAAGGAACAAAGGTCCCAGCCCCCGG
ATCTGACGGGGCGGGACCTGGGCCACCTGCAGGTAAAGCCTGGCGCCCGCCTCCAGCTAGGGAAAGTGT
TTGCGTGCCTCCCGCGGGCGATGGGCCGTGTCACATGGCCGCTGCGGGTGGGGCTGGGTGTGGTGAGTT
CGGGGGCTGTGGGTGCGCCGCCGGCGTGCAGGGTTCGGGGCGGAGAGCGGGGAAGGGACGGGGCTCGGTTGC
ACTGCGCTGCCGCCAGGCTGACGGGGCGGGGCTGCCTGCGTCCCTCCCCCTCGCTGCTCACACTCCATAG
TTTGTGGTTCTGATTTTAAAGCGGAAGGCCAGAGCTCTGTTCAGGAGTTGTGGGAAGCCTTGTAGGGAC
GCGTAATACTTGCCTCCACTTTTTGGTCTCGTAACCTTGTAAACAGTGGTAGTCTCGGTTTCCAT
GGCTTGTGTACTCACAAACACGCACACAGCTGATAACTCTGTGATCATCCTCACTCACTGTAAAGTTGT
CTGCGGGTGGAGCCTATTAGTTTATCACACGCCCTTGGAAAGCCCTCGAAGTGTGATTCTCGTATTCAAACTT
GGTTTTAAAATTGCAACTTACTTGTCTCTGGAGGGGATATTGTCTTGGGTAGGTACTTTGTCTTTA
CTTATCCTTAAAATGCTCTTGATCCTAACAGAAAAAGAATGCCTGGTTGGTTATTCTTCTCTTG
ACCCTTAAAAGCAAATACCACCAAGTGTGTTGCAACCATACTTAAAAAGAAATTCCAGGTAAAACGAAT
TTGCAAGCAGCATTTCATCAAAGTCATGTCTGTTGCAATTGGACTATCTTAAATTGGTTGTGAA
ACTTTAAAACGACATGTGAAAATAATTGTTAATAAAATAACTTGGAAATAATTACATCTTTAGAAATGG
TCTTGAAATGTGATTAAATTCCAAAATTCTTTCTCTTCTGACTTCATATGAATTGAACCTGATAGTTTC
GTATGAATGAATCGCTGATAGGATGTTCTGAGCCTAGTAATAAAATTGCTTATTATGCTGAAATTTGC
TATTCTACTTAAACACCCCTTAAAAGTTCCATCTTACAGAGAAGTTAGTCAAATAGTTAAATGGAGTTCAAC
AGTTTAATACATGAATTAGTGAAGCAGATGAAGCAGAGATGGGAGAAGTAATTAAAATGGCTTCATAATGT
```

# Annotation: making sense

>chromosome:GRCh37:18:45357922:45457515:-1

GCCGGGAGGCGGGGCGGGCGTAGGCAAAGGGAGGTGGGGAGGCGGTGGCCGGCGACTCCCCGCGCCCCGCTCGC  
CCCCCGGCCCTTCCC CGGGT GCTCGGCCTCGTT CCTT CCTCGCTCCCTCCGTCTTCCATACCCGCCCGC  
GCGGCTTCGGCCGGCGTGCCCTCGCCCTAACGGCGGCTGGAGGCGCCAATCAGCGGGCGGCAGGGTGCCAGC  
CCC GGCGCTGCGCCGGCGAATCGGCGGGGCCAGGGTGGCAGGCAGGGTCTACCCCGCGCGGCCGCGGCG  
GCGGAGAAGCAGCTGCCAGCCAGCAGCCCCCAGCCAGCGCCGGAGGTGGGTGCCTGGCGCCGGCGGCCGG  
CCGAGGGCGGAGGGCGGAAGCGGAGGTGGGCTGGCGGGGGAGGGCGCGGCCGTGCGGGCGGCCGGTAGGGCTGCG  
GGCGCGCGCCTGAGGGGAGGAGGGCAGCGCGGGCGCGCGTCTCACCCCCCTCTTCCCCCGCGGGCGGCC  
AGGCTCCCTCCCCCTCCCCCTCCCTCTCCCTCCCCCTCCCCCTCTTCCCCCTACCCCTCCCAGCGCGCCGG  
GCCGGCCGGGCCCGGGCTGGGGCGGGCGGGAAAGACGGCGGCCGGAGTGTTCAGTTCCGCCTCCAATCGC  
CCATTCCYCTCTTCCCCCTCCAGCCCCCTCCATCCCRTCGGAAGAGGAAGGAACAAAGGTCCCGAACCCCCCGG  
ATCTGACGGGGCGGGACCTGGYGCCACCTTGCAGGTAAAGCCTGGCGCCMGCGGGCCKCCAGCTAGGAAAGTGT  
TTGYGTGCGTCCCGCGGGCGATGGCCGTGTCACATGGCCGCTGCGGGTGGGGCTGGGTGTGGTGTGAGTT



# The Ensembl projects



- launched in 1999



- Ensembl gene annotation
- Ensembl browser and tools



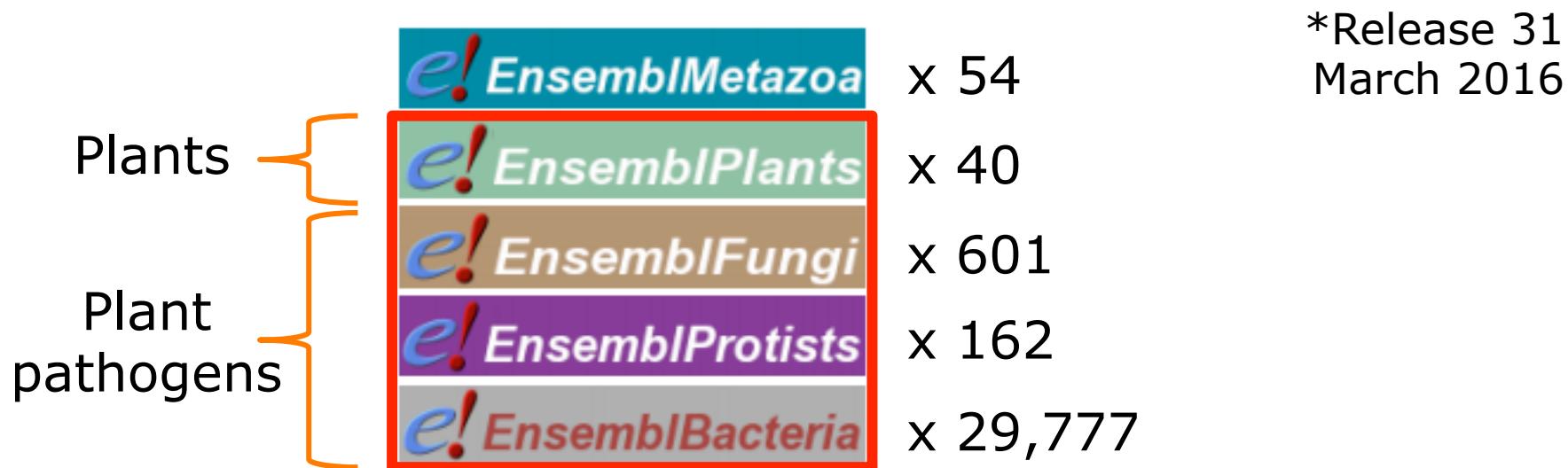
- launched in 2009



- community gene annotation
- Browsers: Plants, Fungi, etc...
- Tools

# Ensembl Genomes

Extends the use of Ensembl to other species  
Wider taxonomic range (v31\*, >30K genomes)



# Genomes in Ensembl Plants

	<i>Aegilops tauschii</i>		<i>Hordeum vulgare</i>		<i>Oryza nivara</i>		<i>Setaria italica</i>
	<i>Amborella trichopoda</i>		<i>Leersia perrieri</i>		<i>Oryza punctata</i>		<i>Solanum lycopersicum</i>
	<i>Arabidopsis lyrata</i>		<i>Medicago truncatula</i>		<i>Oryza rufipogon</i>		<i>Solanum tuberosum</i>
	<i>Arabidopsis thaliana</i>		<i>Musa acuminata</i>		<i>Oryza sativa Indica</i>		<i>Sorghum bicolor</i>
	<i>Brachypodium distachyon</i>		<i>Oryza barthii</i>		<i>Oryza sativa Japonica</i>		<i>Theobroma cacao</i>
	<i>Brassica oleracea</i>		<i>Oryza brachyantha</i>		<i>Ostreococcus lucimarinus</i>		<i>Triticum aestivum</i>
	<i>Brassica rapa</i>		<i>Oryza glaberrima</i>		<i>Physcomitrella patens</i>		<i>Triticum urartu</i>
	<i>Chlamydomonas reinhardtii</i>		<i>Oryza glumaepatula</i>		<i>Populus trichocarpa</i>		<i>Vitis vinifera</i>
	<i>Cyanidioschyzon merolae</i>		<i>Oryza longistaminata</i>		<i>Prunus persica</i>		<i>Zea mays</i>
	<i>Glycine max</i>		<i>Oryza meridionalis</i>		<i>Selaginella moellendorffii</i>		

# Plant pathogens genomes

 <i>Ashbya gossypii</i>	 <i>Dothistroma septosporum</i>	 <i>Magnaporthe oryzae</i>	 <i>Pyrenophora teres</i>	 <i>Ustilago maydis</i>
 <i>Aspergillus ruber</i>	 <i>Erysiphe necator</i>	 <i>Magnaporthe poae</i>	 <i>Pyrenophora tritici-repentis</i>	 <i>Verticillium alfalfae</i>
 <i>Bipolaris maydis</i>	 <i>Eutypa lata</i>	 <i>Marssonina brunnea</i>	 <i>Rhizoctonia solani</i>	 <i>Verticillium dahliae</i>
 <i>Bipolaris oryzae</i>	 <i>Fusarium fujikuroi</i>	 <i>Melampsora larici-populina</i>	 <i>Sclerotinia borealis</i>	 <i>Verticillium longisporum</i>
 <i>Bipolaris sorokiniana</i>	 <i>Fusarium graminearum</i>	 <i>Microbotryum violaceum</i>	 <i>Sclerotinia sclerotiorum</i>	 <i>Zymoseptoria brevis</i>
 <i>Bipolaris victoriae</i>	 <i>Fusarium langsethiae</i>	 <i>Mixia osmundae</i>	 <i>Setosphaeria turcica</i>	 <i>Zymoseptoria tritici</i>
 <i>Bipolaris zeicola</i>	 <i>Fusarium oxysporum</i>	 <i>Moniliophthora perniciosa</i>	 <i>Sphaerulina musiva</i>	
 <i>Blumeria graminis</i>	 <i>Fusarium pseudograminearum</i>	 <i>Moniliophthora roreri</i>	 <i>Sporisorium reilianum</i>	
 <i>Botrytis cinerea</i>	 <i>Fusarium solani</i>	 <i>Neofusicoccum parvum</i>	 <i>Sporisorium scitamineum</i>	
 <i>Claviceps purpurea</i>	 <i>Fusarium verticillioides</i>	 <i>Neonectria ditissima</i>	 <i>Stemphylium lycopersici</i>	
 <i>Colletotrichum fioriniae</i>	 <i>Gaeumannomyces graminis</i>	 <i>Penicillium digitatum</i>	 <i>Stereum hirsutum</i>	
 <i>Colletotrichum graminicola</i>	 <i>Geotrichum candidum</i>	 <i>Phaeosphaeria nodorum</i>	 <i>Thielaviopsis punctulata</i>	
 <i>Colletotrichum higginsianum</i>	 <i>Gloeophyllum trabeum</i>	 <i>Phanerochaete carnosa</i>	 <i>Togninia minima</i>	
 <i>Colletotrichum orbiculare</i>	 <i>Grosmannia clavigera</i>	 <i>Puccinia graminis f. sp. tritici</i>	 <i>Trametes cinnabarina</i>	
 <i>Colletotrichum sublineola</i>	 <i>Heterobasidion irregularare</i>	 <i>Puccinia sorghi</i>	 <i>Trichoderma virens</i>	
 <i>Colletrotrichum gloeosporioides</i>	 <i>Leptosphaeria maculans</i>	 <i>Puccinia striiformis</i>	 <i>Ustilaginoidea virens</i>	
	 <i>Macrohomina phaseolina</i>	 <i>Puccinia triticina</i>	 <i>Ustilago hordei</i>	

 Ensembl Fungi

# Plant pathogens genomes

	<i>Albugo candida</i>		<i>Phytophthora ramorum</i>
	<i>Albugo laibachii</i>		<i>Phytophthora sojae</i>
	<i>Hyaloperonospora arabidopsis</i>		<i>Pythium aphanidermatum</i>
	<i>Phytomonas sp.</i>		<i>Pythium arrhenomanes</i>
	<i>Phytophthora infestans</i>		<i>Pythium irregularare</i>
	<i>Phytophthora kernoviae</i>		<i>Pythium iwayamai</i>
	<i>Phytophthora lateralis</i>		<i>Pythium ultimum</i>
	<i>Phytophthora parasitica</i>		<i>Pythium vexans</i>

 EnsemblProtists

 EnsemblBacteria

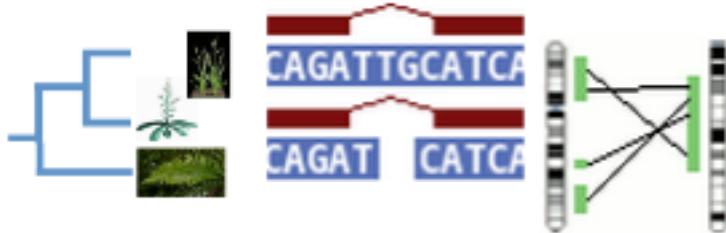
	<i>Agrobacterium vitis</i>
	<i>Burkholderia glumae</i>
	<i>Erwinia amylovora</i>
	<i>Pantoea stewartii</i>
	<i>Pectobacterium carotovorum</i>
	<i>Pseudomonas syringae</i> pv. <i>tabaci</i>
	<i>Pseudomonas syringae</i> pv. <i>tomato</i>
	<i>Ralstonia solanacearum</i>
	<i>Xanthomonas axonopodis</i> pv. <i>citri</i>
	<i>Xanthomonas campestris</i> pv. <i>campestris</i>
	<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i>
	<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>

# Ensembl features

## Gene models



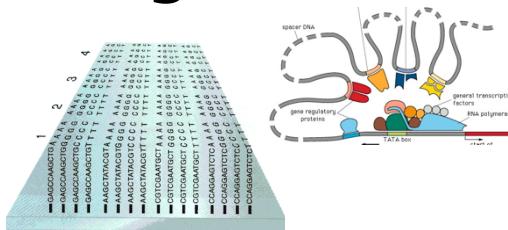
## Comparative Genomics



## Variation



## Regulation



## Toolkit



## Programmatic access



Display your data in Ensembl

## Custom data display



## Free code and data



# Browser walkthrough

The *Arabidopsis* floral homeotic gene APETALA1 (AP1) encodes a putative TF that plays a role in the development of the floral meristem (<https://www.wikigenes.org/e/gene/e/843244.html>).

## Popular genomes



**Arabidopsis  
thaliana**  
TAIR10



**Triticum  
aestivum**  
IWGSC1+popseq



**Zea mays**  
AGPv3



**Oryza sativa  
Japonica**  
IRGSP-1.0



**Hordeum  
vulgare**  
ASM32608v1



**Physcomitrella  
patens**  
ASM242v1

**Arabidopsis thaliana (TAIR10)** ▾



**Arabidopsis thaliana**

*Arabidopsis thaliana*

Data Source [TAIR](#) | Taxonomy ID [3702](#)

Go

e.g. [PAD4](#) or [1:8001-18000](#) or [Carboxypeptidase](#)



Arabidopsis thaliana ▾ Location: 1:25,982,330-25,986,313 Gene: AP1 Transcript: AP1



The **Location** tab → genomic views

- How long is the region where AP1 is annotated?
- Are there oligo probes mapping to this region?
- Which types of variants have been mapped to the last exon of the API1 gene?
- Can I share this view with a colleague/collaborator?

The screenshot shows the Ensembl Plants interface. At the top, there's a green header bar with the logo 'e! EnsemblPlants' on the left. To its right are links for 'Sequence Search', 'BLAST', 'BioMart', 'Tools', and 'Downloads'. Below the header, a blue navigation bar contains the species 'Arabidopsis thaliana' with a dropdown arrow, the location 'Location: 1:25,982,330-25,986,313', the gene identifier 'Gene: AP1' (which is highlighted in red), and the transcript identifier 'Transcript: AP1'. A large red arrow points downwards from the word 'Gene' in the navigation bar to the corresponding tab on the main content area.



The **Gene** tab → gene related views

- Can I view the genomic sequence of my gene
- Can I export the sequence with its variants in RTF?
- Are there orthologues of this gene in other species?
- Which cellular compartment is the product of AP1 found?

The screenshot shows the EnsemblPlants interface for *Arabidopsis thaliana*. The top navigation bar includes links for Sequence Search, BLAST, BioMart, and Downloads. Below the navigation is a search bar with the location 1:25,982,330-25,986,313, gene AP1, and transcript AP1. A large red arrow points to the 'Transcript: AP1' button.



The **Transcript** tab → transcript related views:

- Which domains have been mapped to the AP1 protein?
- Does the AP1 transcript match sequences from other databases?
- Can you find the cDNA sequence of the transcript?
- Can you BLAST this sequence to find out if it's unique?



# Tutorial

DREB6 is a TF on chromosome 3 of **bread wheat**, which plays an important role in tolerance to drought.

1. How many coding genes have been annotated in the latest assembly of the wheat genome?
2. What is the location of DREB6? How many alternatively transcripts have been annotated? Does any transcript match UniProt proteins? How similar are they?
3. Does the region of this gene align to other regions of the wheat hexaploid genome?

# Toolkit for omics analyses\*

## Access Ensembl data



## Process your data



 Display your data in Ensembl

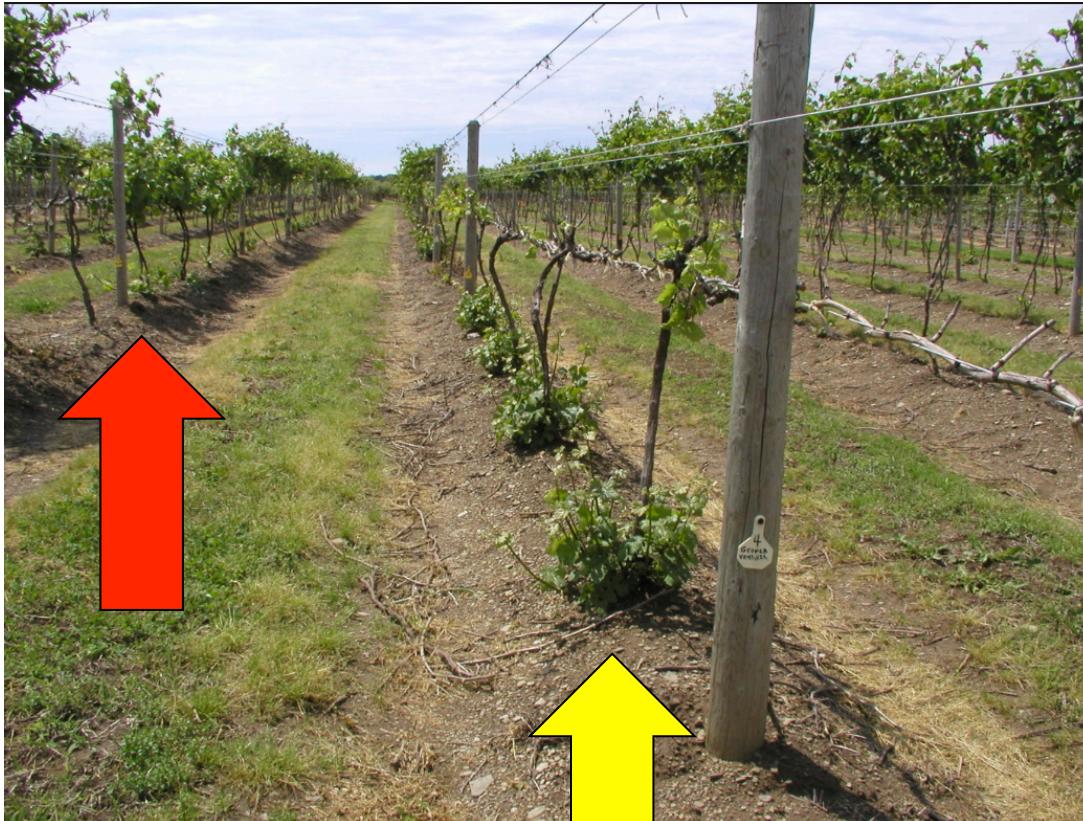
Assembly Converter 

- \* [plants.ensembl.org/tools.html](http://plants.ensembl.org/tools.html)
- \* [fungi.ensembl.org/tools.html](http://fungi.ensembl.org/tools.html)

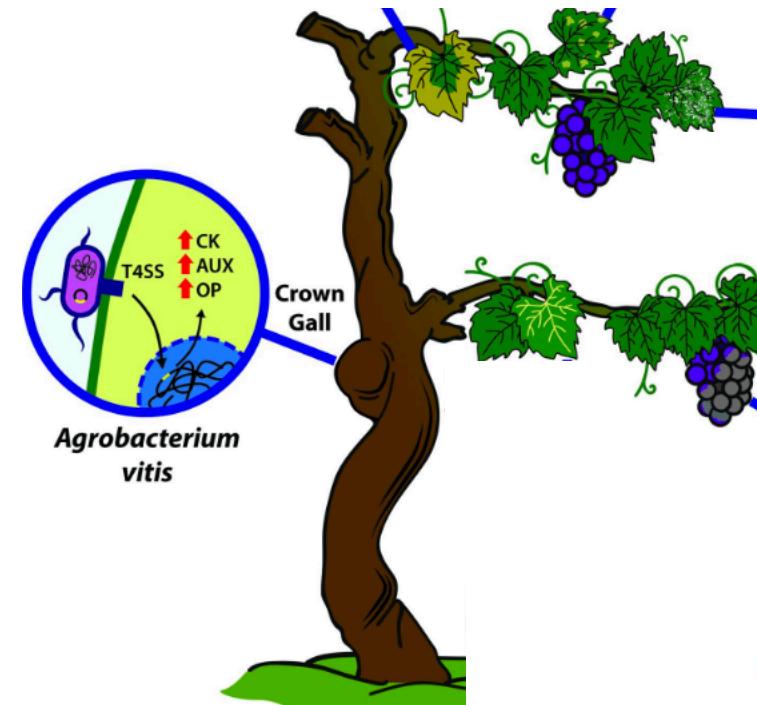
# **Ve!P** : before we find out

- What is it?
- What does it do?
- Where can I find it?

# A use case: vineyards

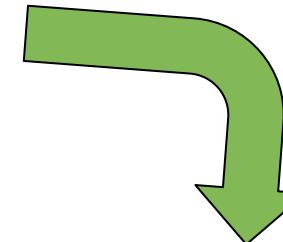


healthy  
X  
unhealthy



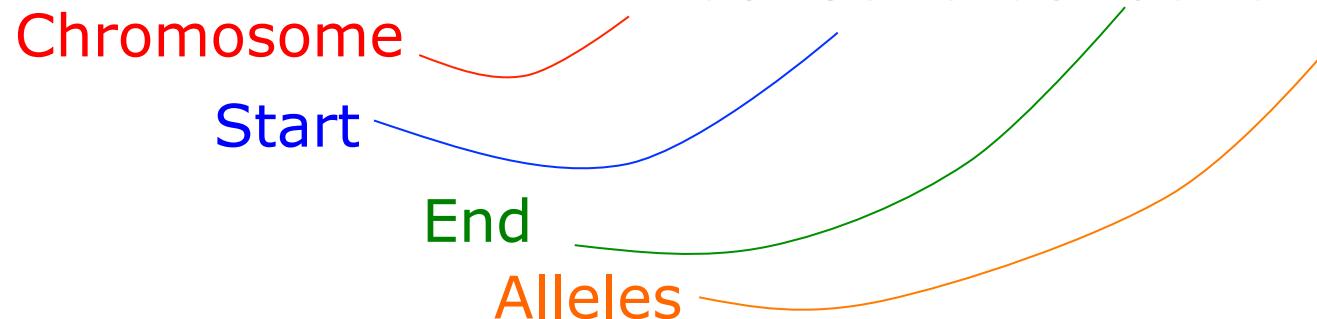
# My sequencing experiments

healthy grapes  
X  
unhealthy grapes



Positions in the genome  
vary between H and U

5 18122 18122 C/T  
4 16295 16295 T/G  
1 12841 12841 G/-  
19 6848593 6848593 T/C



# Can I annotate these variants?

**Yes, you can!**



- Variant Effect Predictor
- Annotate variants (SNPs, CNVs, indels)
- Available for species in Ensembl Genomes
- Make your own cache: FASTA & GTF files\*

\*[http://www.ensembl.org/info/docs/tools/vep/script/vep\\_cache.html#gtf](http://www.ensembl.org/info/docs/tools/vep/script/vep_cache.html#gtf)

# Predict the effect of variants

- Variant Effect Predictor



- Different input formats e.g.

5 18122 18122 C/T

variant IDs

**BIOINFORMATICS APPLICATIONS NOTE** Vol. 26 no. 16 2010, pages 2069–2070 doi:10.1093/bioinformatics/btq330

Advance Access publication June 18, 2010

**Databases and ontologies**

**Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor**

William McLaren<sup>1,\*</sup>, Bethan Pritchard<sup>2</sup>, Daniel Rios<sup>1</sup>, Yuan Chen<sup>1</sup>, Paul Flicek<sup>1</sup> and Fiona Cunningham<sup>1,\*</sup>

<sup>1</sup>European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD and  
<sup>2</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

PMID: 20562413

Web interface



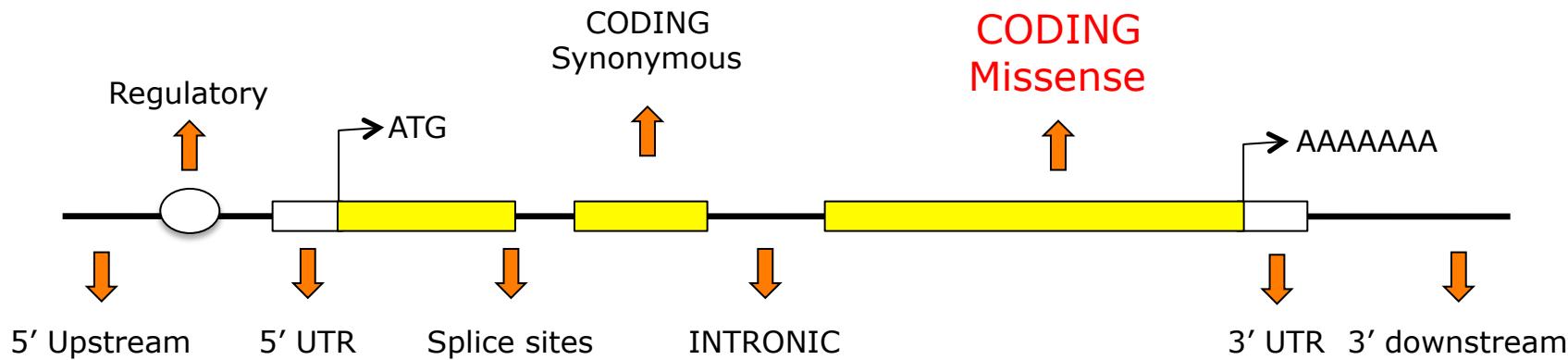
Perl script



REST API



# Map variants on transcripts



Identify the transcripts that overlap my variants and predict the consequence of these on transcripts using the **Ve!P**

# Consequence terms for variants

* SO term	SO description
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron
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
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
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript
initiator_codon_variant	A codon variant that changes at least one base of the first codon of a transcript
inframe_insertion	An inframe non synonymous variant that inserts bases into the coding sequence
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence
missense_variant	A sequence variant, where the change may be longer than 3 bases, and at least one base of a codon is changed resulting in a codon that encodes for a different amino acid
transcript_amplification	A feature amplification of a region containing a transcript
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
incomplete_terminal_codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed
synonymous_variant	A sequence variant where there is no resulting change to the encoded amino acid
stop_retained_variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains
coding_sequence_variant	A sequence variant that changes the coding sequence
mature_miRNA_variant	A transcript variant located with the sequence of the mature miRNA
5_prime_UTR_variant	A UTR variant of the 5' UTR
3_prime_UTR_variant	A UTR variant of the 3' UTR
intron_variant	A transcript variant occurring within an intron
NMD_transcript_variant	A variant in a transcript that is the target of NMD
non_coding_exon_variant	A sequence variant that changes non-coding exon sequence
nc_transcript_variant	A transcript variant of a non coding RNA
upstream_gene_variant	A sequence variant located 5' of a gene
downstream_gene_variant	A sequence variant located 3' of a gene

[http://www.ensembl.org/info/genome/variation/predicted\\_data.html#consequence\\_type\\_table](http://www.ensembl.org/info/genome/variation/predicted_data.html#consequence_type_table)

\* defined by the Sequence Ontology (SO) project

# Inputting data into the Ve!P

**e! EnsemblProtists** ▾ BLAST | BioMart | Tools | Downloads | Documentation | Website help

Plasmodium falciparum (ASM276v1) ▾ VEP

**Web Tools**

- Web Tools
  - BLAST
  - Variant Effect Predictor**
  - Assembly Converter
  - ID History Converter

**Configure this page**

**Add your data**

**Export data**

**Share this page**

**Bookmark this page**

## Variant Effect Predictor

Click for help (opens in new window)

Plasmodium falciparum

Name for this data (optional):

Either paste data:

Chromosome Start End Alleles (REF/ALT) Strand

9 20083 20083 G/T 1  
7 36133 36133 A/C 1  
8 22373 22373 C/-1

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

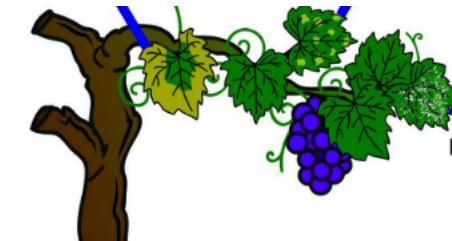
Instant results for first variant >

Or upload file: Choose File No file chosen

Instant results for first variant >

# My sequencing experiments

healthy grapes  
X  
unhealthy grapes



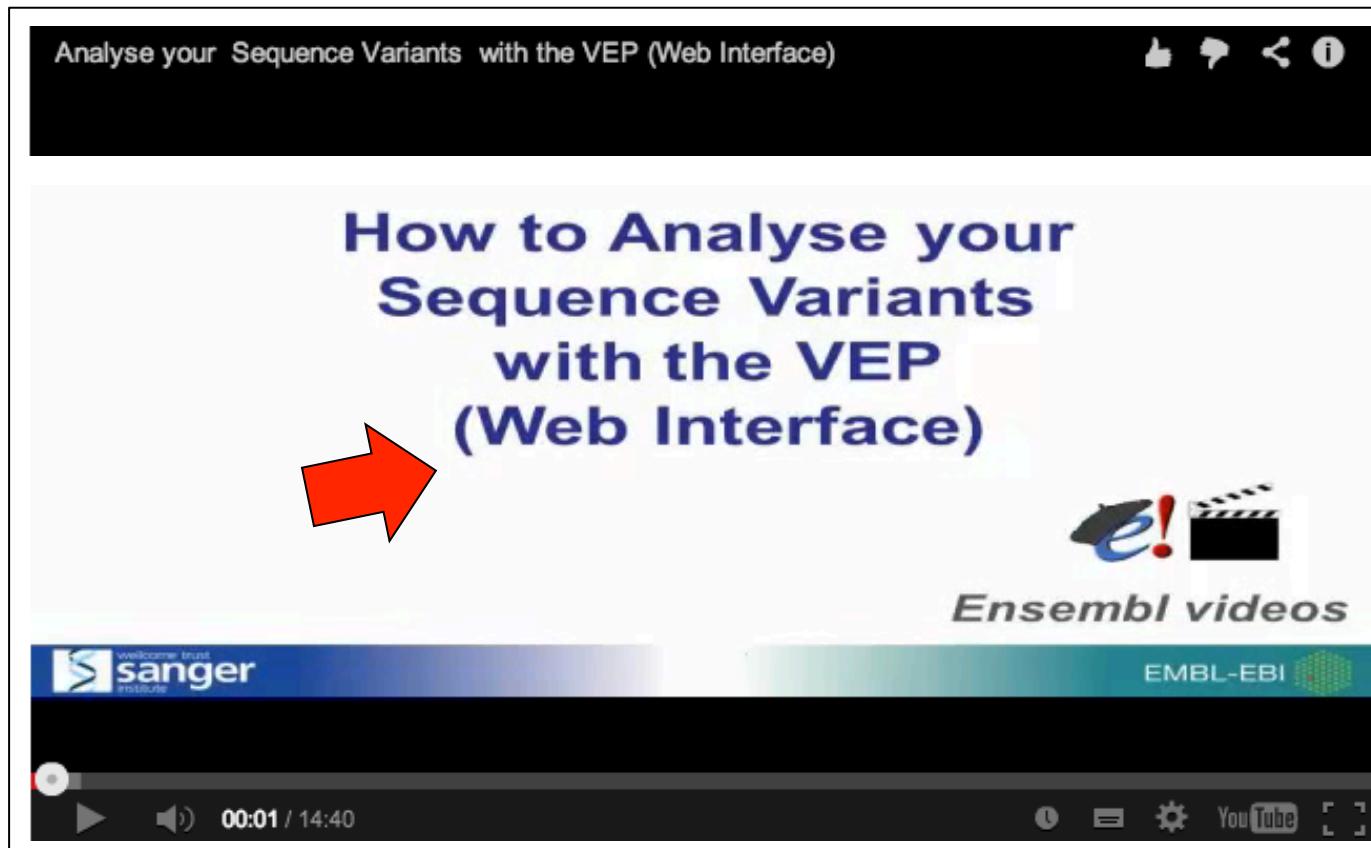
5 18122 18122 C/T  
4 16295 16295 T/G  
1 12841 12841 G/-  
19 6848593 6848593 T/C

- Do these variants map to genes and/or transcripts?
- What is the effect of those variants?
- Are those variants known in public databases?



# LIVE DEMO

# VEP video



<http://tinyurl.com/vep-video>

# Wrap up

## Plants and pathogens: Ensembl Genomes

Genome annotation, genetic variants, omics analyses and more

Biomart

R

Ensembl.org

Ensembl API

MySQL

REST



Oh Yes!  
And all is 100% free

# Ensembl Genomes paper

D574–D580 *Nucleic Acids Research*, 2016, Vol. 44, Database issue  
doi: 10.1093/nar/gkv1209

Published online 17 November 2015

## Ensembl Genomes 2016: more genomes, more complexity

Paul Julian Kersey<sup>1,\*</sup>, James E. Allen<sup>1</sup>, Irina Armean<sup>1</sup>, Sanjay Boddu<sup>1</sup>, Bruce J. Bolt<sup>1</sup>, Denise Carvalho-Silva<sup>1</sup>, Mikkel Christensen<sup>1</sup>, Paul Davis<sup>1</sup>, Lee J. Falin<sup>1</sup>, Christoph Grabmueller<sup>1</sup>, Jay Humphrey<sup>1</sup>, Arnaud Kerhornou<sup>1</sup>, Julia Khobova<sup>1</sup>, Naveen K. Aranganathan<sup>1</sup>, Nicholas Langridge<sup>1</sup>, Ernesto Lowy<sup>1</sup>, Mark D. McDowall<sup>1</sup>, Uma Maheswari<sup>1</sup>, Michael Nuhn<sup>1</sup>, Chuang Kee Ong<sup>1</sup>, Bert Overduin<sup>1</sup>, Michael Paulini<sup>1</sup>, Helder Pedro<sup>1</sup>, Emily Perry<sup>1</sup>, Giulietta Spudich<sup>1</sup>, Electra Tapanari<sup>1</sup>, Brandon Walts<sup>1</sup>, Gareth Williams<sup>1</sup>, Marcela Tello-Ruiz<sup>1</sup>, Joshua Stein<sup>2</sup>, Sharon Wei<sup>2</sup>, Doreen Ware<sup>2,3</sup>, Daniel M. Bolser<sup>1</sup>, Kevin L. Howe<sup>1</sup>, Eugene Kulesha<sup>1</sup>, Daniel Lawson<sup>1</sup>, Gareth Maslen<sup>1</sup> and Daniel M. Staines<sup>1</sup>

<sup>1</sup>The European Molecular Biology Laboratory, The European Bioinformatics Institute, The Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK, <sup>2</sup>Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA and <sup>3</sup>USDA-ARS NAA Plant, Soil and Nutrition Laboratory Research Unit, Cornell University, Ithaca, NY 14853, USA

# Connect with the team



[helpdesk@ensemblgenomes.org](mailto:helpdesk@ensemblgenomes.org)

# Training materials

Ensembl materials are protected by a CC BY license



<http://creativecommons.org/licenses/by/4.0/>

If you wish to re-use these, please credit Ensembl for their creation

If you use Ensembl for your work, please cite our papers

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