

Variation data in Ensembl Plants

Hands-on workshop
July 2015

Dan Bolser
Denise Carvalho-Silva
Brandon Walts
EMBL-EBI

<http://plants.ensembl.org>



Overview of the next two hours...

Before tea

Session 1, Browser:

- Introduction
 - Ensembl Plants and plant variation data
- Hands-on
 - Variation in the Ensembl browser
- Displaying your data in Ensembl

After tea

Session 2, Tools:

- Variant Effect Predictor
 - Introduction and hands-on
- Mining variation data with BioMart
 - Introduction and hands-on

Resources

These slides:

<http://tinyurl.com/transplant2015>

Course booklet:

<http://tinyurl.com/transplant2015b>

Introduction: An overview of Ensembl Plants

Dan Bolser
Ensembl Plants project leader
EMBL-EBI

<http://plants.ensembl.org>



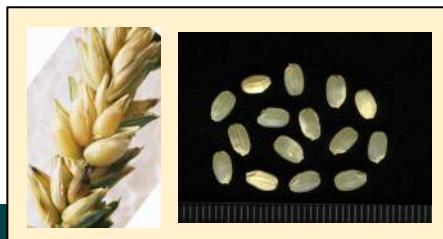
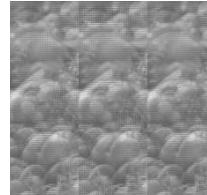
Introduction overview

Background:

- Ensembl Plants
 - History
 - Data



- Recent updates
 - Wheat
 - Tomato
 - Rice



Plant variation data:

- Plant variation data
- Variation in the Ensembl browser



- Tools for processing your own data



Ensembl Plants uses Ensembl technology

Ensembl:



- A platform for **genome browsing**, annotation and analysis.
- Has 'modules' for handling:
 - Genomic data, comparative genomics, variations, and epigenomics.
- Multiple points of access to data:
 - Browser-based application, Perl and REST APIs, direct access (MySQL), DAS (client and server), FTP, ..
 - Data mining tool. The logo for BioMart features the word "biomart" in a lowercase sans-serif font. To the left of the "b" is a graphic element consisting of a 3x3 grid of colored dots: red at the top, green in the middle, and orange at the bottom-right.
- Upload your own data and compare it to the reference seq. and annotation.

FREE

Ensembl was originally developed for vertebrate genomes, subsequently extended to non-vertebrate species:

- Ensembl → Ensembl Genomes → Ensembl Plants



**Ensembl is developed
jointly by the EBI and
the Wellcome Trust
Sanger Institute**

Currently 39 genomes in Ensembl Plants

<http://plants.ensembl.org>

Dicots in Ensembl Plants: 12 in 7 orders

 **Arabidopsis lyrata** GA

[JGI | 81972](#)

 **Arabidopsis thaliana** VP GA

[TAIR | 3702](#)

 **Brassica rapa** GA

[IVFCAAS | 51351](#)

 **Glycine max** GA

[JGI | 3847](#)

 **Medicago truncatula** GA

[IMGAG | 3880](#)

 **Populus trichocarpa** GA

[JGI | 3694](#)

 **Prunus persica** GA

[IPGI | 3760](#)

 **Solanum lycopersicum** PGA

[ITGSP | 4081](#)

 **Solanum tuberosum** GA

[PGSC | 4113](#)

 **Vitis vinifera** VP GA

[Genoscope | Vitis vinifera | 29760](#)

Brassicales

Fabales

Malpighiales

Rosales

Solanales

Vitales



Theobroma cacao (cocoa)

Data Source: [The Cacao Genome Database](#)

Malvales



Zingiberales



Poales

45-60



30-39
40-54

Monocots in
Ensembl Plants:
21 in 2 orders



...

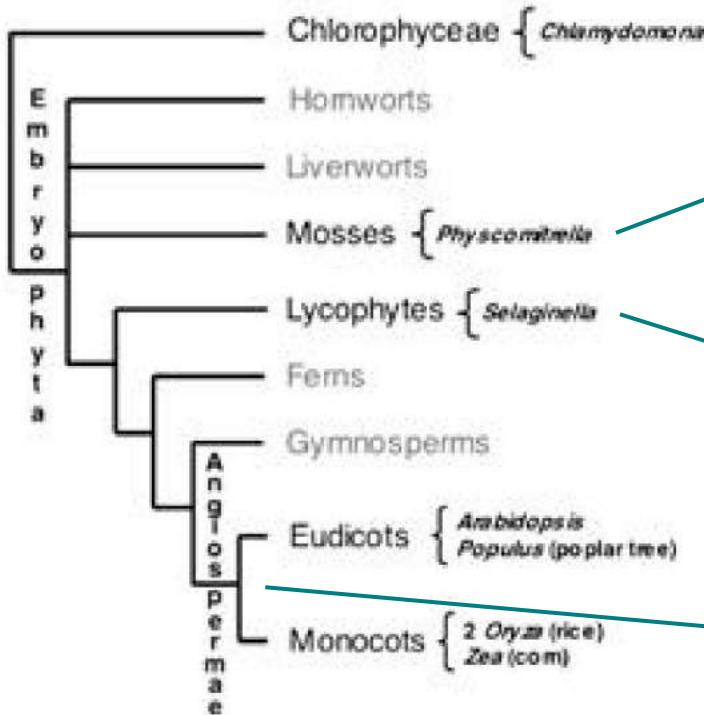


'Others': 6 in ~5 Classes

Rhodophyta



Cyanidioschyzon merolae P G
[CMGP | 280699](#)



Chlorophyta



Chlamydomonas reinhardtii P G A
[JGI | 3055](#)

Bryophyta



Physcomitrella patens P G A
[JGI | 145481](#)

Lycopodiophyta



Selaginella moellendorffii G A
[JGI | 88036](#)

Amborellales



Amborella trichopoda P G
[AGD | 13333](#)

What do we host?

Genomes are selected for inclusion based on:

- Availability of complete genome sequence
- Importance as model organisms
 - e.g. *Arabidopsis* and *Brachypodium*.
- Importance for agriculture
 - e.g. potato, sorghum, barley, rice, wheat and *Brassica*.
- Interest as evolutionary reference points
 - e.g the basal angiosperm, *Amborella trichopoda*, the aquatic alga *Chlamydomonas reinhardtii*, the moss *Physcomitrella patens* and the vascular non-seed spikemoss *Selaginella moellendorffii*.



Or all the above!

Ensembl | Log in | Register | BLAST/BLAT | Bedtools | Tools | Downloads | Help | Documentation | Metrics

Search: All species

1 e.g. mouse gene BRCA1 or test X:50000..200000 or breast cancer disease

Browsing a Genome

The Ensembl project generates databases for metazoans and other eukaryotic species, and makes this information available via the web.

Click on a genome name to go to the species' home page.

Popular genomes: [Log in to contribute](#) ([See list](#))

 Human
 Mouse
 Zebrafish

All genomes

- Search species

[View full list of Ensembl species](#)

One species are available in [GCF](#) and [Glossary](#)

Ensembl is a joint effort between [DSI](#), [EBI](#) and the [Wellcome Trust Sanger Institute](#) to build a reference system that predicts and maintains annotation in selected metazoan genomes.

New to Ensembl?

Did you know you can:

- [Learn how to use Ensembl](#) with our video tutorials and walk-throughs
- [Add your own data](#) to Ensembl
- [Annotate your own data](#) with our new Create Panel
- [Upload and analyze your data](#) with our new Data Importer
- [Search for a gene](#) or protein sequence using BLAST at EBI
- [Fetch all the data](#) you need for your research using the API
- [Download raw databases](#) via ZIP (FASTA, GFF, BED, and other formats)
- [Mine Ensembl data](#) with our new search interface and export sequences to tables in text, XML, or Excel format

Still got questions? Try the [FAQ](#) or [Contact us](#).

What's New in Ensembl 9 (8 August 2016)

- [New transcript models and displays \(Homo\)](#)
- [ENCODE tracks now available \(Homo\)](#)
- [Changes to model organism support](#)
- [Details of our updates, API changes, etc.](#)

[More news](#)



Ensembl Genomes

The Ensembl Genomes project generates genome databases for important species from across the Eukaryotic tree of life. For more information see the [Ensembl Genomes homepage](#). This page will compare the existing [Ensembl Bacteria](#) and [Ensembl Protists](#) databases.

Ensembl Genomes data is available through many of the same tools as Ensembl Data. Data can be accessed via:

- the [BioMart interface](#) ([Bacteria](#) and [Protists](#) sections), or
- [http://www.ensemblgenomes.org/](#) (the project homepage)
- [http://www.ncbi.nlm.nih.gov/taxonomy/](#) (the taxonomic database)
- the [Ensembl Bacteria](#) and [Ensembl Protists](#) (beta) databases
- the [Ensembl Genomes API](#)
- the [Ensembl Genomes](#) mailing list (subscribe (using [http://lists.ensembl.org/mailman/listinfo/ensembl-genomes](#)) or unsubscribe (using [http://lists.ensembl.org/mailman/listinfo/ensembl-genomes/unsubscribe](#)))
- the [Ensembl Genomes](#) GitHub repository (to copy the existence of "genome collections", i.e. the existence of many small genomes in a single Ensembl database (a database which has been adapted by Ensembl Bacteria, the Ensembl Protists and Ensembl Fungi teams to accommodate the needs of their respective communities). To access a collection, single genome details are published. Who can do this? Ensembl Genomes is a community effort, anyone can contribute to it. If you have a dataset you would like to contribute then contact the [Ensembl Genomes team](#) for more information about how to use Ensembl Genomes.

See the [FAQ](#) for more information about how to use Ensembl Genomes.

Ensembl

Search All species
e.g. *PCR* (OR or AND)
Go!

Popular genomes (Go to [to customize the list](#))

- Bacillus subtilis**
- Paramecium tetraurelia**
- Diplocystis oligosporum**
- Plasmodium vivax**

All genomes

[View genome details](#)
[View list of all Ensembl genomes](#)

What's in Release 5 (June 2010)

- New core databases for the diatoms *Thalassiosira weissflogii* and *Thalassiosira pseudonana* based on the [Phaeophyceae](#)
- Optimized peptide comparsion
- Improved phylogenetic tree and genome paralogs
- Virtual RefSeq
- Submissions to Ensembl
- Ensembl Protein version 5 (provides gene coverage for 1000s of species)

Links

Log In | Register | Help | Support | Contact | About | Feedback | Home | Search | Help

Ensembl Genomes

The Ensembl Genomes project generates genome databases for important species from across the Eukaryotic tree of life. For more information see the [Ensembl Genomes homepage](#). This page will compare the existing [Ensembl Bacteria](#) and [Ensembl Protists](#) databases.

Ensembl Genomes data is available through many of the same tools as Ensembl Data. Data can be accessed via:

- the [BioMart interface](#) ([Bacteria](#) and [Protists](#) sections), or
- [http://www.ensemblgenomes.org/](#) (the project homepage)
- [http://www.ncbi.nlm.nih.gov/taxonomy/](#) (the taxonomic database)
- the [Ensembl Bacteria](#) and [Ensembl Protists](#) (beta) databases
- the [Ensembl Genomes API](#)
- the [Ensembl Genomes](#) mailing list (subscribe (using [http://lists.ensembl.org/mailman/listinfo/ensembl-genomes](#)) or unsubscribe (using [http://lists.ensembl.org/mailman/listinfo/ensembl-genomes/unsubscribe](#)))
- the [Ensembl Genomes](#) GitHub repository (to copy the existence of "genome collections", i.e. the existence of many small genomes in a single Ensembl database (a database which has been adapted by Ensembl Bacteria, the Ensembl Protists and Ensembl Fungi teams to accommodate the needs of their respective communities). To access a collection, single genome details are published. Who can do this? Ensembl Genomes is a community effort, anyone can contribute to it. If you have a dataset you would like to contribute then contact the [Ensembl Genomes team](#) for more information about how to use Ensembl Genomes.

Feedback

Feedback forms

Feedback forms 1 includes the majority of these datasets: *Phaeophyceae*, *Chlorophyta* and *Thalessiosira pseudonana*.

Feedback forms 2 includes the remaining datasets: *Stramenopiles*, *Euglenozoa* (plus *Microsporidia* which are thought to have arisen from *Stramenopiles*), *Alveolates* (including *Apicomplexa*), *Amoebozoa*, *Forams* and *Other eukaryotes*. These datasets are included in Feedback forms 2 because they are composed with other members of Ensembl's Fungi, the core of which is *Yarrowia lipolytica*, while the other members of Ensembl's Fungi are composed with other members of Ensembl's Protists.

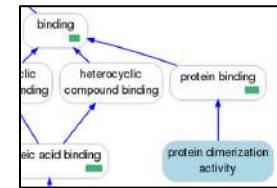
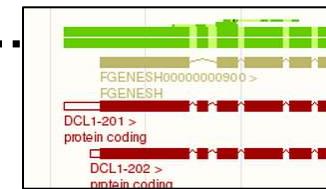
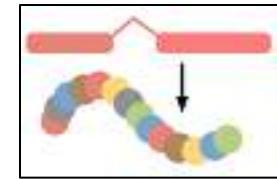


Ensembl and agriculture

- Scope of Ensembl Plants includes important crop and model species
- Ensembl Metazoa, Protists, Fungi and Bacteria include important agricultural pathogens, pests, pollinators, symbionts
- Important farm animals (cow, sheep, pig, chicken, etc.) are among the key species in vertebrate Ensembl
- See also, 
 - Integrates genome-scale data from important plant pathogen species with literature-curated information about the phenotypes of host infection.

Types of data in Ensembl (Ensembl Plants)

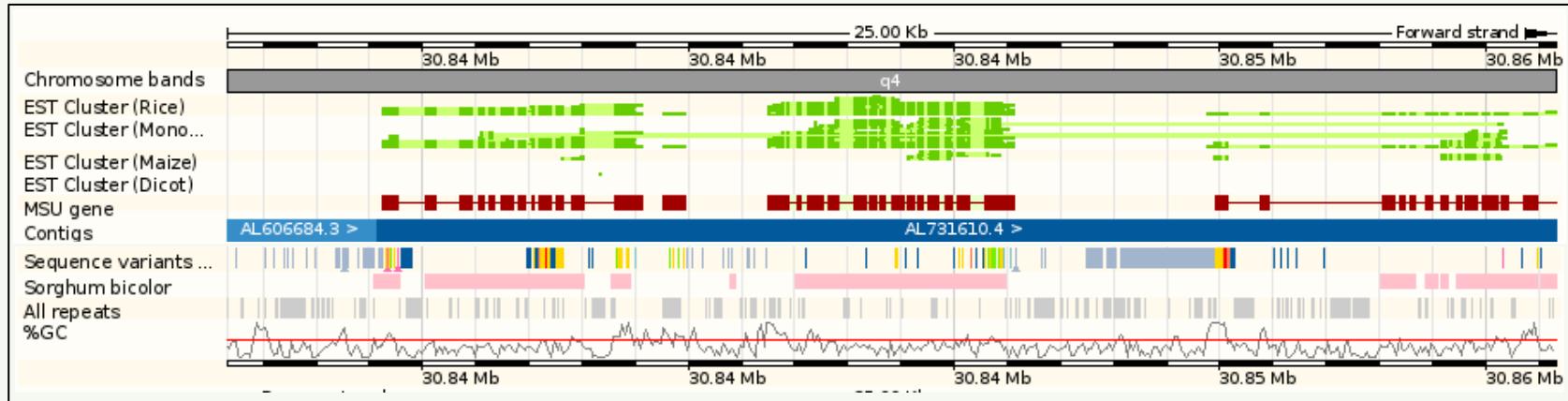
- Genomic sequence
- Gene, transcript, and protein annotations
- External references and ontology terms
- Mapped sequences: cDNAs, proteins, probes, BACs, repeats, markers, ...
- Variation data:
 - sequence variants
 - structural variants
- Comparative data:
 - gene trees, orthologues, paralogues
 - whole genome alignments and synteny



ATCGAGCT
ATCCAGCT
ATCGAGAT



Genome Browsers



- Assembly structure and sequence
- Genes
- Expressed regions
- Comparative alignments
- Genetic markers
- Variation
- Repeats & transposable elements
- Regulatory & Epigenetic marks

- Baseline Annotation
- Transposon discovery
 - Gene prediction
 - EST Alignment

Recent updates (Release 27)

<http://plants.ensembl.org>

Release 4 to 5 times a year (3 months)

Release 24 (Nov. 2014)

- New chromosome assembly for barley.
- Added 10 million wheat ‘Inter-homoeologous variations’ (IHVs).
- Added 70 million variations from the tomato 150 genomes project.

Release 25 (Jan. 2015)

- New chromosome assembly for wheat.
- More than 10 million new barley variations from 170 individuals.
- New whole genome alignments between *Brassica rapa* and *oleracea*.

Release 4 to 5 times a year (3 months)

Release 26 (Mar. 2015)

- Updated genome for *Medicago truncatula* and *Oryza meridionalis*.
- New RNA-Seq data for barley.
- New whole genome alignments between triticeae using ATAC.

Release 27 (Jun. 2015)

- Tomato genome updated to version 2.5
- New variation data from the wheat HapMap project

Funding (Ensembl Plants)

- Ensembl Genomes Funded by
 - EMBL
 - EU (INFRAVEC, Microme, **transPLANT**, **AllBio**)
 - BBSRC (PhytoPath, **wheat**, **barley** and midge sequencing, **UK-US collaboration**, RNACentral, **capital grant**)
 - Wellcome Trust (PomBase)
 - NIH/NIAID (VectorBase)
 - **NSF (Gramene collaboration)**
 - Bill and Melinda Gates Foundation (**wheat rust**)

Ensembl Genomes (Ensembl Plants)

- James Allen, Irina Armean, **Dan Bolser**, Mikkel Christensen, Paul Davies, **Christoph Grabmueller**, Kevin Howe, Malcolm Hinsley, Jay Humphrey, **Arnaud Kerhornou**, **Paul Kersey**, Julia Khobdova, Eugene Kulesha, Nick Langridge, Dan Lawson, Mark McDowall, Uma Maheswari, Gareth Maslen, Michael Nuhn, Chuang Kee Ong, Michael Paulini, Helder Pedro, Anton Petrov, Dan Staines, Mary Ann Tuli, **Brandon Walts**, Gary Williams
 - If you have a question that is not answered here, please contact me (dbolser@ebi.ac.uk) or our HelpDesk:
 - helpdesk@ensemblgenomes.org

People

- Arnaud Kerhornou
 - Most of the wheat work
- Brandon Walts
 - Barley projection
- Christoph Grabmüller
 - Projection pipeline
- Paul Kersey
 - Wheat inter-homoeologous variations
- The web team
 - Eugene Kulesha
 - Julia Khobova
 - Nick Langridge



On to variation...

Introduction to variation data in Ensembl Plants

Dan Bolser
Ensembl Plants project leader
EMBL-EBI

<http://plants.ensembl.org>



Introduction overview

Background:

- Ensembl Plants
 - History
 - Data



- Recent updates
 - Wheat
 - Tomato
 - Rice



Plant variation data:

- Plant variation data
- Variation in the Ensembl browser



- Tools for processing your own data



Current plant variation data

Species	Samps	Pops*	~Variations
<i>Arabidopsis thaliana</i>	1,610	7	14,000,000 SV: 14,000
<i>Brachypodium distachyon</i>	3	3	300,000
<i>Hordeum vulgare</i>	188	16	24,000,000
<i>Oryza glaberrima</i>	21	1	800,000
<i>Oryza indica</i>	19	19	5,000,000
<i>Oryza sativa</i>	416	21	6,000,000
<i>Solanum lycopersicum</i>	84	8	71,000,000
<i>Sorghum bicolor</i>	3	1	SV: 64,000
<i>Triticum aestivum</i>	248	1	700,000
<i>Vitis vinifera</i>	17	3	500,000
<i>Zea mays</i>	103	1	51,000,000

Some large upcoming variation projects:

- International Maize and Wheat Improvement Center (CIMMYT)
- BMAP (150 brassica)
- 3000 rice cultivars
- Barley GBS on exomes of ~2500 EMS mutants

* Pops can be true genetic populations or, more often, simply 'collections'.

Variation Data

- SNPs and structural variation
- Natural and structured populations



Species	Variants	Source	Studies
<i>Oryza sativa</i> <i>Group japonica</i>	3,332,525	160K SNPs x 20 accessions 1311 SNPs x 395 accessions NCBI dbSNP	McNally et al. (2009). PNAS 106:12273-12278 Zhao et al. 2010. PLoS ONE. 5:e10780
<i>Oryza sativa</i> <i>Group indica</i>	4,747,883	NCBI dbSNP	
<i>Zea mays</i>	50,719,843	HapMap1: NAM founder lines HapMap2: pre-domesticated & domesticated lines	Gore et al. 2009. Science 326:1115-1117. Chia et al. 2013. Genet 44:803-807.
<i>Arabidopsis thaliana</i>	14,234,197 SV: 13,667	250K SNPs x 1179 accessions 1001 genomes project: 411 resequenced accessions	Atwell et al. 2010. Nature. 465:627-631.
<i>Brachypodium distachyon</i>	327,988	3 accessions of Brachypodium sylvaticum	Fox et al. (2013) Applications in Plant Sciences 1 (3):1200011. 2013
<i>Vitis vinifera</i>	457,404	Resequencing USDA germplasm collection	Myles S, et al. 2010. PLoS ONE. 5:e8219.
<i>Hordeum vulgare</i>	12,994,003	Resequencing 4 accessions plus wild barely	The International Barley Genome Sequencing Consortium. 2012. Nature 491, 711–716
<i>Oryza glaberrima</i>	7,704,409	Resequenced 20 accessions African rice & wild progenitor	Oryza Genome Evolution project
<i>Sorghum bicolor</i>	SV: 64,507	Structural variants from Database of Genomic Variants archive (dGVA)	Zheng et al. 2011. Genome Biol.. 12:R114.

Arabidopsis



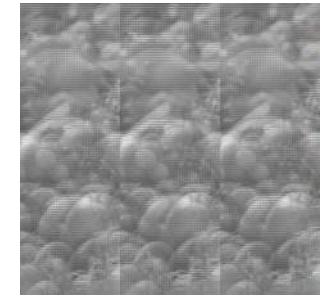
Currently:

- SNPs for 1,179 strains on the Affy 250k SNP chip.
- Resequencing data for 18 Arabidopsis lines.
- A total of 392 strains from the 1001 Genomes Project:
 - 80 strains from the Cao pilot study.
 - 132 strains from the Salk Institute.
 - 180 strains from the Nordborg group at GMI.
- Phenotype data from a GWAS study of 107 phenotypes in 95 inbred lines.

Coming ‘soon’:

- The ‘final’ 1001 data
 - ~1,300 strains and 13,000,000 genotypes.
 - The ‘full imputed’ matrix for ~2,000 strains.

Tomato



Currently:

- Data from resequencing 84 tomato accessions
 - Accessions include cultivated wild relatives representative of the *Lycopersicon*, *Arcanum*, *Eriopersicon* and *Neolycopersicon* groups.
 - The variation data has been submitted to the ENA with accession ERP004618, and has been locus-level accessioned using the transPLANT variation archive.

Potentially:

- Data from the 350 genomes study...

Wheat



Currently:

- SNP data for ~725,000 loci across ~250 individuals provided by CerealsDB.
- ~1.5 million loci from the Wheat Hapmap project generated by resequencing 62 diverse wheat lines.
- A total of ~2,000,000 inter-homoeologous variants called from whole genome alignments between the A, B and D component genomes.

Coming ‘soon’:

- Exome capture of ~2,000 EMS mutants.

A ‘typical’ variation view in Ensembl

Gene-based displays

- Gene summary
- Splice variants (1)
- Supporting evidence
- Sequence
- External references (4)
- Regulation
- Plants Compara
 - Genomic alignments (7)
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (11)
 - Paralogues (3)
- Pan-taxonomic Compara
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (2)
 - Paralogues (3)
 - Protein families (0)
- Genetic Variation
 - Variation Table
 - Variation Image** (selected)
 - External Data
- Personal annotation
- ID History
 - Gene history

- Configure this page
- Manage your data
- Export data
- Bookmark this page

Ensembl Plants is produced in collaboration with Gramene

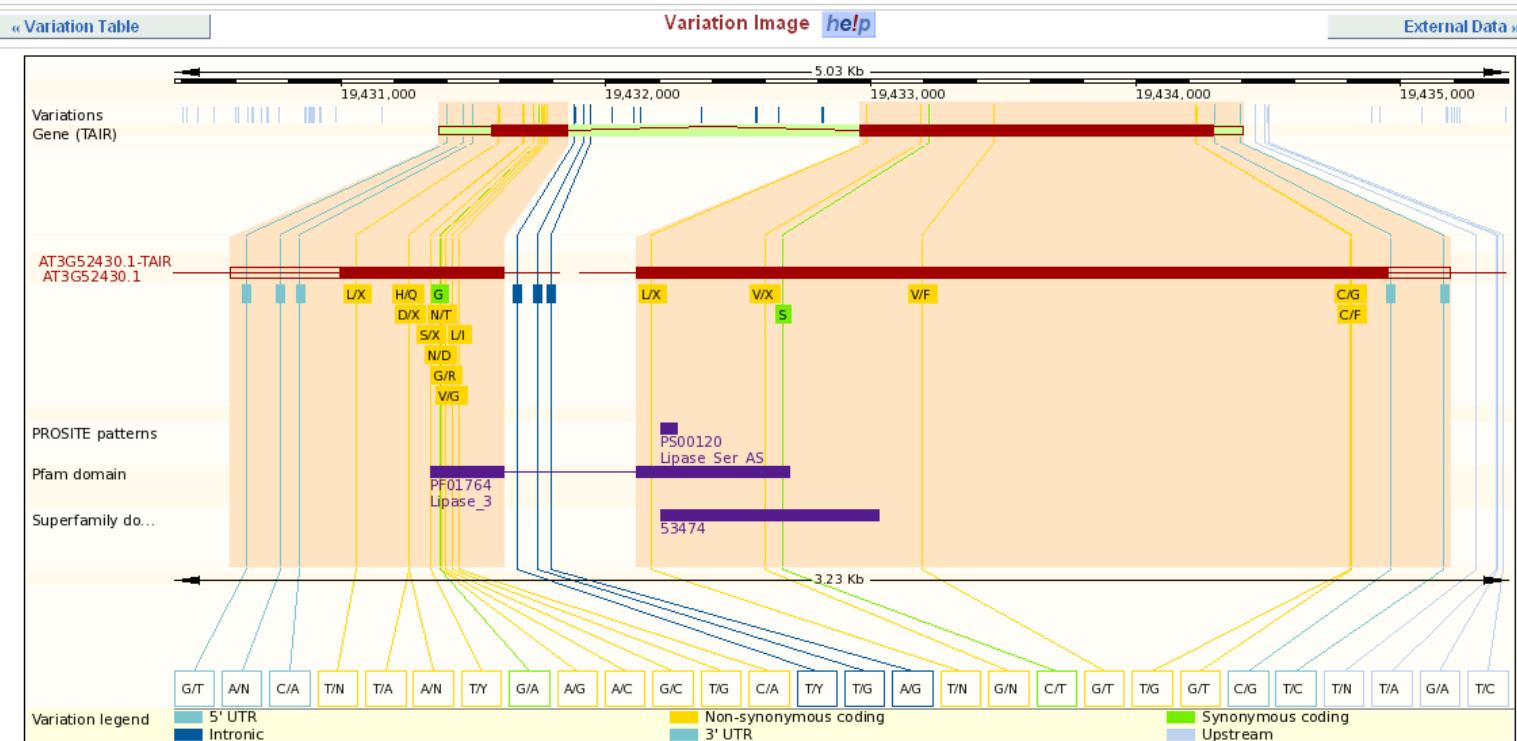
DB built by NASC

Gene: PAD4 (AT3G52430-TAIR-G)

PAD4 (PHYTOALEXIN DEFICIENT 4); lipase/ protein binding / triacylglycerol lipase; Encodes a lipase-like gene that is important for salicylic acid signaling and function in resistance (R) gene-mediated and basal plant disease resistance. PAD4 can interact directly with EDS1, another disease resistance signaling protein. Expressed at elevated level in response to green peach aphid (GPA) feeding, and modulates the GPA feeding-induced leaf senescence through a mechanism that doesn't require camalexin synthesis and salicylic acid (SA) signaling. Source: TAIR PAD4

Location Chromosome 3: 19,431,371-19,434,403 forward strand.

Transcripts There is 1 transcript in this gene: [show transcripts](#)



Viewing Variation Data

Variations
Gramene gene

GRMZM2G027563_T01

G/V V/V G/A A/S
V/A D/A V/L S/S
UL WI

Population genetics ⓘ

Show/hide columns

Population	Alleles A	Alleles G	Genotypes AIA	Genotypes AGA
Zmays	0.147	0.853	0.147	0.853

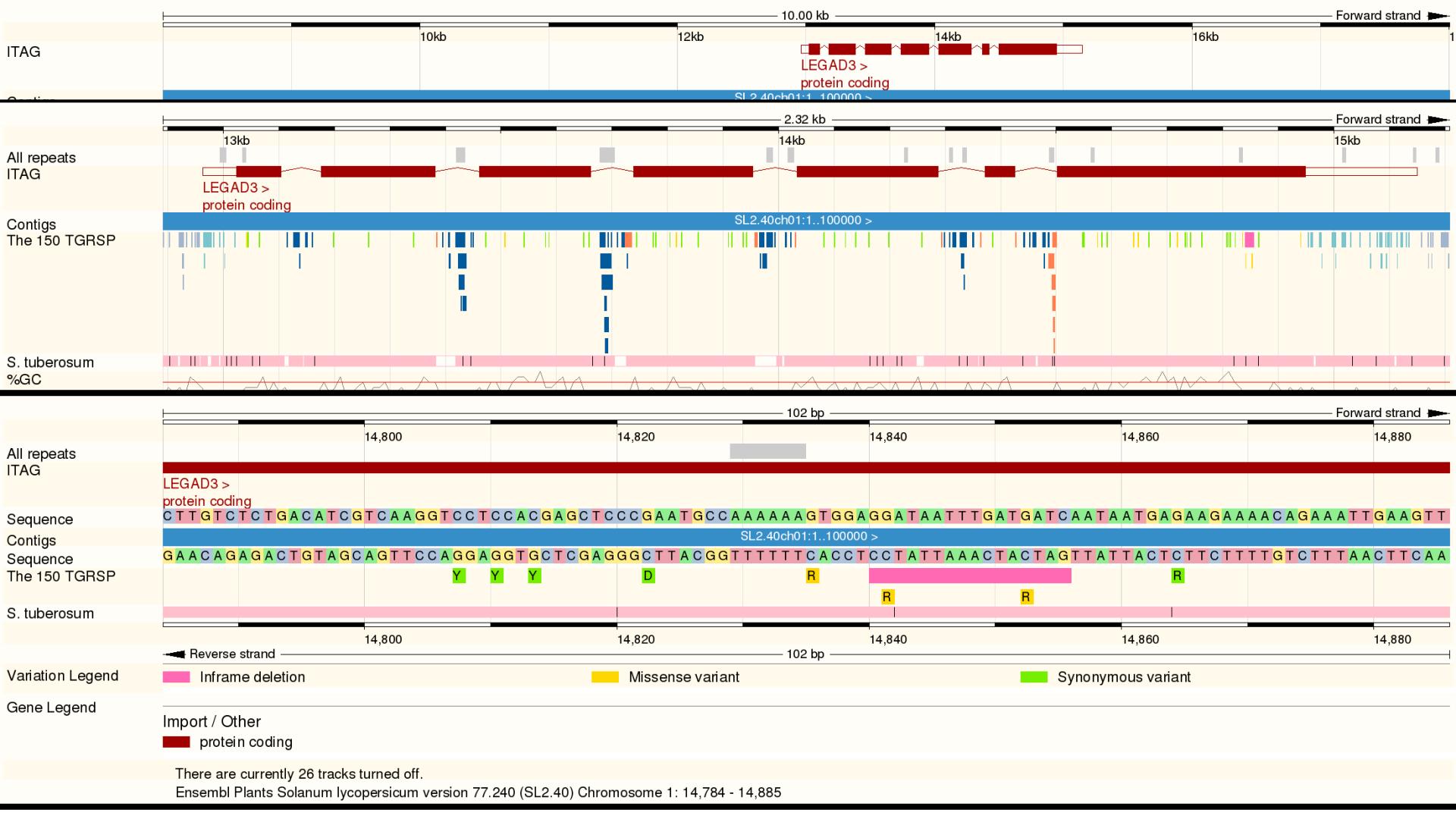
Genotypes for Zmays ⓘ

Show All entries Show/hide c

Individual	Genotype(forward)
B73:MZ (U)	GIG
B97:MZ (U)	AIA
BKN015:MZ (U)	GIG
BKN016:MZ (U)	GIG
BKN017:MZ (U)	GIG
BKN018:MZ (U)	GIG
BKN022:MZ (U)	GIG
BKN023:MZ (U)	AIA
BKN025:MZ (U)	GIG
BKN027:MZ (U)	GIG
BKN029:MZ (U)	GIG
BKN031:MZ (U)	GIG
BKN032:MZ (U)	GIG
BKN033:MZ (U)	GIG
BKN034:MZ (U)	AIA
BKN035:MZ (U)	GIG
BKN040:MZ (U)	GIG
CAU5003:MZ (U)	AIA
CAUCHANG72:MZ (U)	AIA
CAUMO17:MZ (U)	AIA
CML103:MZ (U)	GIG
CML228:MZ (U)	GIG

Zea_mays	Sorghum_bicolor	Sequence
7:115277533	7:115277533	TACCGA-GTACTAGTAAACAACTAGC@GCCGCC-CATTATAAATAGC
2: 59255525	2: 59255525	TTCTGATGGAGTAGTAAAACAAATAGCCGCCGCCCCGTATATAATAGC
7:115277420	7:115277420	C---CTTCCTCACTCTTAT---CCAG@CAGCTCCCTTG-----TAACCC
2: 59255406	2: 59255406	CAACCTTCTCTCAGCTTAGCAGCCAGCCTTCTTGAACACTTAACCC
7:115277311	7:115277311	ATGGCTGCCGCCAGCAGATCTCCTCGACGAC@TGCCGCCGCCAGCAGCAG
2: 59255286	2: 59255286	ATGGCCGCCGCCGCCAGATCTCCCTCGACGACCTGCGCCGCCAGCAGCAG
7:115277194	7:115277194	CCGGAACATGGCATCCGTGGGGAGGGAAAGGCAGCCGGATGGCCCGGCG
2: 59255166	2: 59255166	CCGGAGCTGGCATCCGCCGGAGGGAAAGGCAGCCGGATGGCCGCCGCC
7:115277083	7:115277083	CAGGGAGAGGAGAAATCCCGGGCAGCGCTTCGGGGCTCTTGTGGGCG
2: 59254658	2: 59254658	CAGGGAG---GAGAACCGCCGGC---GTCTTCGGGGTTCTTCGCCGG
7:115276969	7:115276969	ATGTGTACATACGATTTACATAGCTTAATGCTTGTAAGAACCGAGCC
2: 59254546	2: 59254546	ATCGTACAT----ACGTACATGCCTTAATGCTTG-TG-GCACACATGCC
7:115276853	7:115276853	CCGGGCCAGGGTTCCGGTGGAGGAGGCAGAGGAGGTGGACGCCGGCG
2: 59254435	2: 59254435	AAGGGCCAAGACTATGCGGAGGATGCGAAGT-----CCCCGGCG
7:115276733	7:115276733	GTACGGGCCGGCGGGCAGGGCAGGCCACCGGACCCCCACAGTATCGGGAGC
2: 59254324	2: 59254324	GAACGGCGTGGCGCCAGGGCACCACAGCATCGCGGAGC
7:115276640	7:115276640	TAATTGCGCTCTCGTCTC---CTATTAGTCACGTCG---AGTAGTTAA
2: 59254204	2: 59254204	TAATTGCCGCTCTCGCTCTCCAAATTAGTCACGTCGTGAGCAGTTG
7:115276542	7:115276542	CTGTCGTTTTCTTGGCTCGTTGAATTGTTCTCTCGTCCCAGTC
2: 59254084	2: 59254084	CTGTTGTTTCTTGGAT---TTGGATTGTTCTCC---CGTCCCGACTC
7:115276455	7:115276455	CGTCGTCGCCGTTTGCAGCTCGTAGAGAGAG@ATCGCGGAGAGGATG
2: 59253969	2: 59253969	CGTCGTCGCCGTTTGCAGCTCCGAGGGAGGATCGCGGAGAGGATG
7:115276335	7:115276335	ACG-----AGAGAGGCAGATCGATTCAAAGGCCTAGGGTTT
2: 59253862	2: 59253862	ACGCTGCTTGAGAAAGACGAGAGATTATT-----GCCTAGGGTTT

Looking at some tomato variation data...



Arabidopsis thaliana ▾

Location: 4:16,056,194-16,057,194

Variation: ENSVATH00550254

Variation displays

- Explore this variation
- Genomic context
 - Genes and regulation (18)
 - Flanking sequence
- Genotype frequency
- Individual genotypes (289)
- Linkage disequilibrium
- Phenotype Data (49)
- Phylogenetic Context
- Citations
- External Data

Configure this page

Add your data

Export data

Bookmark this page

Share this page

Ensembl Plants is produced in
collaboration with Gramene**ENSVATH00550254 SNP**

Original source

Variation features from Affy_250k, Perlegen_1M, WTCHG and 1001 Genomes, with Ensembl identifiers | Ensembl

Alleles

C/N/T

Location

Chromosome 4:16056694 (forward strand) | [View in location tab](#)

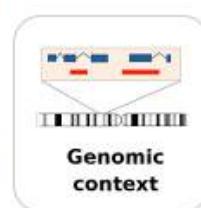
Co-located

with Ensembl [ENSVATH09775155](#) (C/-)

Synonyms

This variation has 2 synonyms - click the plus to show

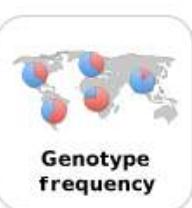
HGVS name

[4:g.16056694C>T](#)**Explore this variation** 

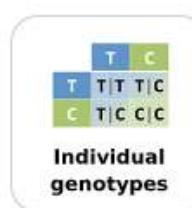
Genomic context



Genes and regulation



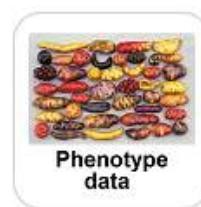
Genotype frequency



Individual genotypes



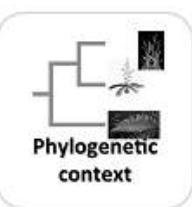
Linkage disequilibrium



Phenotype data



Citations



Phylogenetic context

ATTCAATT CGGS~~G~~TG TCATGCT

Flanking sequence

Using the website

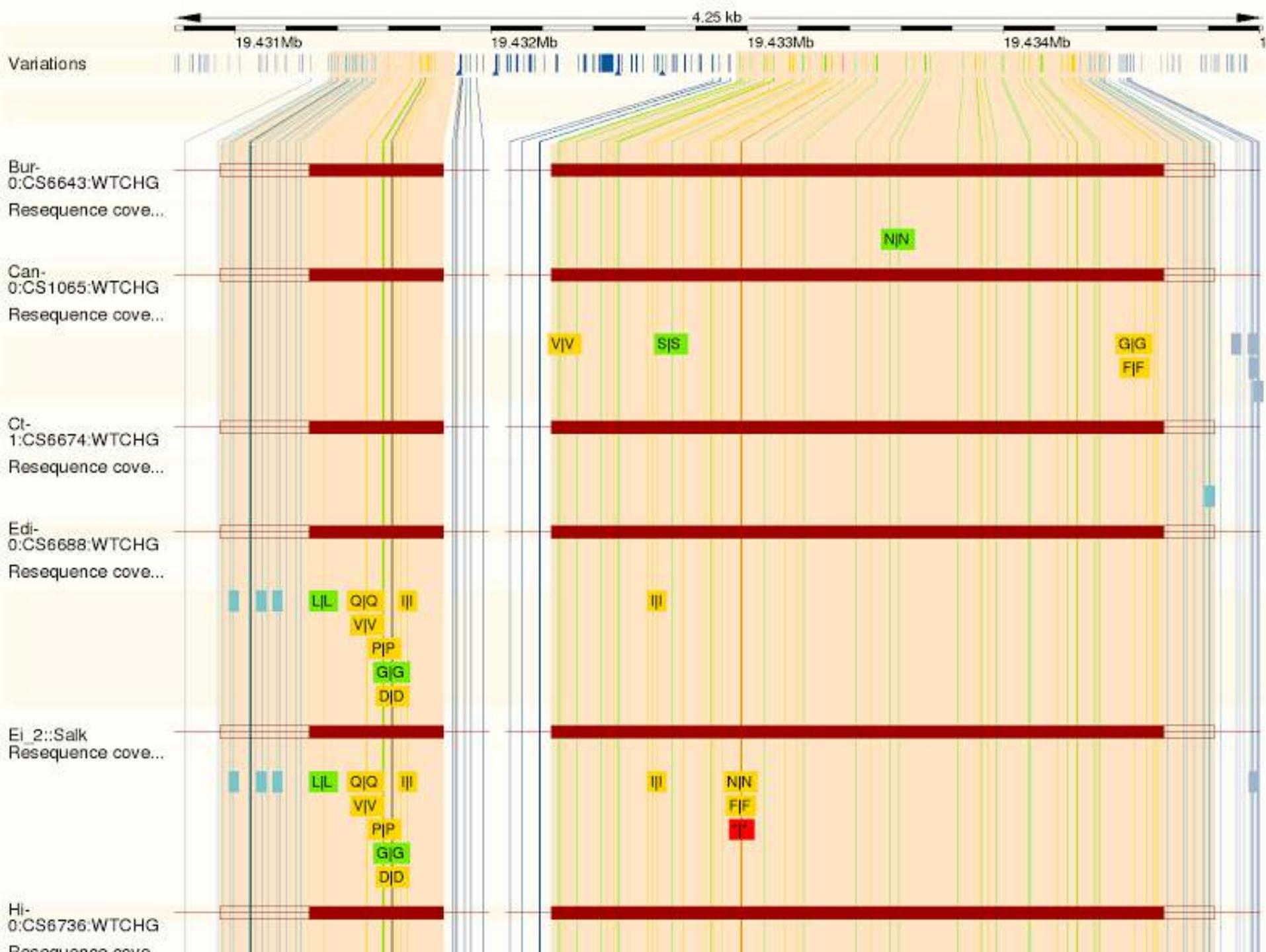
- Video: [Browsing SNPs and CNVs in Ensembl](#)
- Video: [Clip: Genome Variation](#)
- Video: [BioMart: Variation IDs to HGNC Symbols](#)
- Exercise: [Genomes and SNPs in Malaria](#)

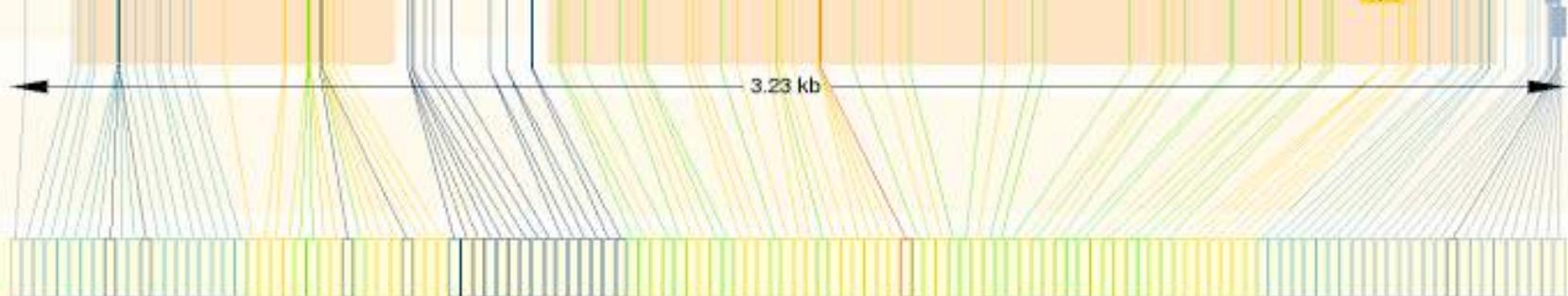
Analysing your data**Programmatic access**

- Tutorial: [Accessing variation data with the Variation API](#)

Reference materials

- [Ensembl variation documentation portal](#)
- [Ensembl variation data description](#)
- [Variation Quick Reference card](#)





Variation legend

- █ Stop gained
- 5 prime UTR variant
- Upstream gene variant
- Missense variant
- 3 prime UTR variant
- Intergenic variant
- Synonymous variant
- Intron variant

Comparison to Col-0 alleles

Compare to Col-0

Bur-0:CS6643:WTCHG
Can-0:CS1065:WTCHG
Ct-1:CS6674:WTCHG
Edi-0:CS6688:WTCHG
El_2::Salk
Hi-0:CS6736:WTCHG
Kn-0:CS6762:WTCHG
Ler-0:CS20:WTCHG
Mt-0:CS1380:WTCHG
No-0:CS6805:WTCHG
Oy-0:CS6824:WTCHG
Po-0:CS6839:WTCHG
Rsch-4:CS6850:WTCHG
Sf-2:CS6857:WTCHG
Sg_1::Salk
Tsu-0:CS6874:WTCHG
Uk_1::Salk
Wil-2:CS6889:WTCHG
Ws-0:CS6891:WTCHG
Wu-0:CS6897:WTCHG
Zu-0:CS6902:WTCHG

Haplotype legend

- Different allele
- Same allele
- Heterozygous
- Missing data

EXERCISES

Variation in the Ensembl Browser

Exercise 1 – Exploring a SNP in Arabidopsis

The arabidopsis [ATCDSP32](#) gene is a chloroplastic, 32 kDa, drought-induced stress protein that is proposed to participate in cell redox homeostasis ([GO:0045454](#)).

- A. How many variants have been identified in the gene that can cause a change in the protein sequence?
- B. What is the ID of the variant that change the residue 60 from Alanine to Threonine? What is the location of this SNP in the Arabidopsis genome? What are its possible alleles?
- C. Download the flanking sequence of this SNP in RTF (Rich Text Format). Can you change how much flanking sequence is displayed on the browser?
- D. Does this SNP cause a change at the amino acid level for other genes or transcripts?
- E. What is the most frequent genotype at this locus in the ‘1001 Population’?

Exercise 2 – Variation data in the tomato (*S. lycopersicum*) genome

Nearly every aspect of plant biology is dependent on cytochrome P450 enzymes, including metabolism of secondary metabolites in tomato involved in fruit development and ripening.

- A. Find a cytochrome P450 gene in tomato known as Solyc02g085360.2 and go to its Location tab. Can you add the data track that shows variation data from the ‘150 Tomato Genome ReSequencing Project’ (TGRSP)?
- B. Zoom in around the last exon of this gene. What are the different types of variants seen in that region? What is the location of the only inframe deletion mapped in the region?
- C. Click on the splice region variant showed in that view. Why does Ensembl Plants put the G allele first in the string (G/A)?

Exercise 3 – Missense variants in the bread wheat genome

The bread wheat transcript Traes_2AL_5C7E76139.1 is involved in calcium ion binding ([GO:0005509](#)). Around 45 variants have been mapped to this transcript.

- A. What are the two types of variants annotated in this transcript?
- B. Are there any variants predicted to be deleterious? Which amino acid residue is affected and what are the possible amino acids in that position?
- C. What are the different sources of EPITAEV06993600 and BA00258972?

Displaying your data in Ensembl Plants

Dan Bolser

Denise Carvalho-Silva

Brandon Walts

EMBL-EBI

<http://plants.ensembl.org>



Visualise your own data

Upload data:

- Data saved on server
- 5 MB limit
 - Large file formats?

Attach remote files:

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

User added tracks:

- Can be saved or shared
- Only trivial security, do not use for sensitive data!

Upload formats:

- | | |
|------------|------------------------|
| • BED | genes / features |
| • Gbrowse | genes / features |
| • GFF/GTF | genes / features |
| • PSL | sequence alignments |
| • WIG | continuous-valued data |
| • BedGraph | continuous-valued data |
| • TrackHub | collections of tracks |

Attach formats:

- | | |
|----------|------------------------|
| • BigBed | genes / features |
| • BAM | sequence alignments |
| • BigWig | continuous-valued data |
| • VCF | variants |

Live demo

Variant Effect Predictor

Dan Bolser

Denise Carvalho-Silva

Brandon Walts

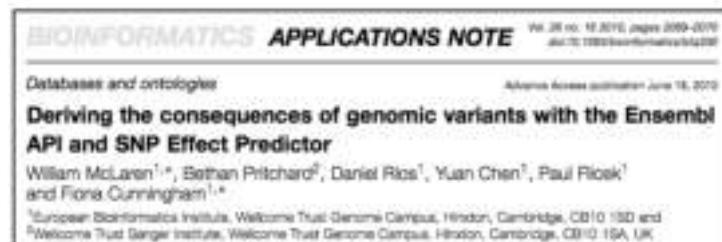
EMBL-EBI

<http://plants.ensembl.org>



Variant Effect Predictor

- Maps variants on Ensembl transcripts and predicts their consequences
- Different input formats
 - Ensembl default
 - list of rs IDs
 - VCF
 - Pileup
 - HGVS
- Versions: Rest API, Perl script and online interface

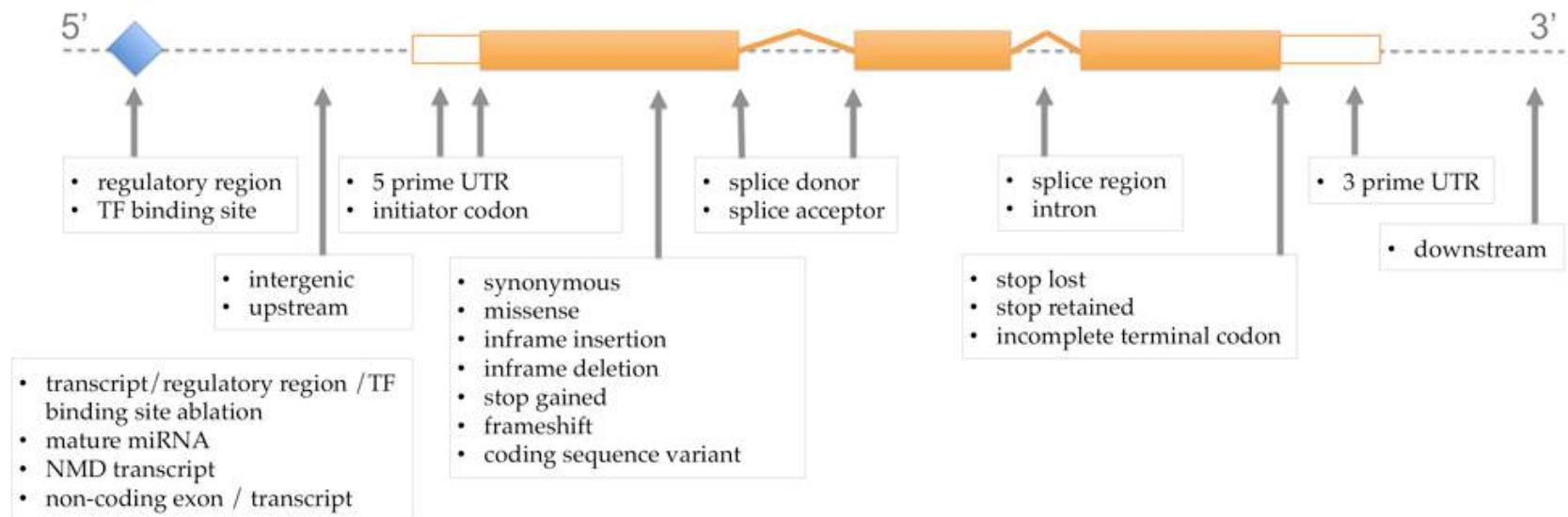


PMID: 20562413



Variant Effect Predictor (VEP)

- Predicts functional consequences of known and unknown variants
- For substitutions, insertions, deletions and structural variants
- Web interface (for up to 750 variants), standalone Perl script, Perl API and REST API



psbO, Oxygen-evolving enhancer protein 1

Traes_2BL_4B8B77E73.1

Variations

Residue	Variation ID	Type	Evidence	Alleles	Ambig. code	Residues	Codons	SIFT
35	BA00804152.2	Synonymous variant		G/A	R	V	GTC, GTT	-
48	BA00039145	Missense variant		T/C	Y	D, G	GAC, GGC	0.03
103	BS00179246.2	Missense variant		C/A	M	G, C	GGC, TGC	0
119	BA00233383.2	Synonymous variant		G/A	R	P	CCC, CCT	-
124	BA00897028.2	Synonymous variant		C/T	Y	K	AAG, AAA	-
136	BS00178628.2	Synonymous variant		G/A	R	S	TCC, TCT	-
139	BA00227489	Synonymous variant		G/C	S	V	GTC, GTG	-
154	BA00269810.2	Synonymous variant		C/T	Y	K	AAG, AAA	-
165	BA00829191	Missense variant		G/T	K	L, I	CTT, ATT	0.11
172	BA00058887.2	Synonymous variant		G/C	S	L	CTC, CTG	-
181	BA00863963	Missense variant		T/C	Y	K, R	AAG, AGG	0.27
190	BA00326027	Synonymous variant		A/G	R	Y	TAT, TAC	-
192	BA00121558	Synonymous variant		G/A	R	A	GCC, GCT	-
250	BA00796319	Synonymous variant		G/C	S	A	GCC, GCG	-
253	BA00078190	Synonymous variant		T/C	Y	L	CTA, CTG	-
264	BA00706748	Synonymous variant		A/G	R	L	TTG, CTG	-
276	BA00323812	Synonymous variant		T/C	Y	T	ACA, ACG	-
286	BS00116511	Missense variant		C/T	Y	S, N	AGC, AAC	0
295	BA00298030.2	Synonymous variant		G/A	R	G	GGC, GGT	-

Ensembl tools

<http://www.ensembl.org/tools.html>

Name	Description	Online tool	Download code	Documentation
Variant Effect Predictor 	Analyse your own variants and predict the functional consequences of known and unknown variants via our Variant Effect Predictor (VEP) tool.			
BLAST/BLAT	Search our genomes for your DNA or protein sequence.			
BioMart	Use this data-mining tool to export custom datasets from Ensembl.			
Assembly converter	Map (liftover) your data's coordinates to the current assembly.			
ID History converter	Convert a set of Ensembl IDs from a previous release into their current equivalents.			
Ensembl Virtual Machine	VirtualBox virtual Machine with Ubuntu desktop and pre-configured with the latest Ensembl API plus Variant Effect Predictor (VEP). NB: download is >1 GB	-		

<http://www.ensembl.org/vep>



Inputting data into Ve!P

Human (GRCh37) ▾ Jobs

Tools

- Web Tools
- Variant Effect Predictor

Configure this page

Add your data

Export data

Bookmark this page

Share this page

Variant Effect Predictor ⓘ

New VEP job:

Input

Species: Human (Homo sapiens)

Assembly: GRCh37

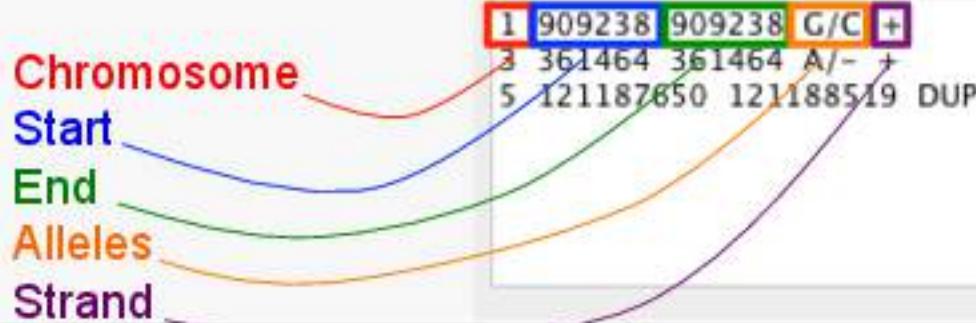
Name for this data (optional):

Input file format (details): Ensembl default

Either paste data:

Chromosome Start End Alleles Strand

1 909238 909238 G/C +
3 361464 361464 A/- +
5 121187650 121188519 DUP



EXERCISE

Displaying your data in Ensembl Plants

Exercise 4 – The Variant Effect Predictor in the bread wheat genome

An analysis of 5,000 individuals from two different populations of bread wheat (*T. aestivum*) has identified thousands of polymorphic loci. See a list of a few of them on the right:

Can you use the VEP tool to answer the following?

- A. Which genes and transcripts do these variants map to?
- B. Which consequence types can be found for these variants?
 - o Do any of them cause a change at the amino acid level?

chr 2D, genomic coordinate 89551917,
alleles G/A, forward strand

chr 2D, genomic coordinate 148408765,
alleles G/T, forward strand

chr 3D, genomic coordinate 113574123,
alleles C/A, forward strand

chr 3D, genomic coordinate 93827883,
alleles G/A, forward strand

chr 3B, genomic coordinate 727928129,
alleles C/T, forward strand

chr 3B, genomic coordinate 736734474,
alleles C/T, forward strand

chr 6A, genomic coordinate 196872409,
alleles T/G, forward strand

chr 6A, genomic coordinate 196153918,
alleles A/G, forward strand

chr 6A, genomic coordinate 196774882,
alleles G/C, forward strand



Analyse your Sequence Variants with the VEP (Web Interface)



How to Analyse your Sequence Variants with the VEP (Web Interface)



Ensembl videos



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

Ensembl Plants Tools: BioMart

Denise Carvalho-Silva
Ensembl Outreach Officer
EMBL-EBI

<http://plants.ensembl.org>



Outline

- Definitions
- The principle: 4 steps
- Tutorial: simple query in Arabidopsis
- Find Ensembl BioMart and BioMart elsewhere
- Sophisticated platforms: mart services, APIs, etc...
- Exercises

What is BioMart?

- Free service for easy retrieval of Ensembl data
- Data export tool with little/no programming required
- Complex queries with a few mouse clicks
- Output formats (.xls, .csv, fasta, tsv, html)

The four-step principle

DATA

FILTERS

ATTRIBUTES

RESULTS

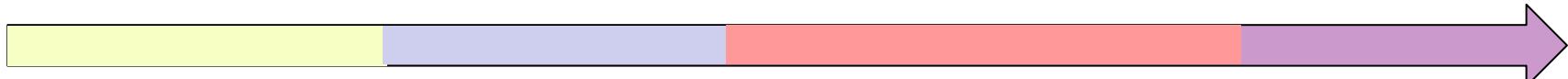
Database

Dataset

IDs
Regions
Domains
Expression

Homologs
Sequences
Features
Structures

Tables
Fasta



Choosing the data

The screenshot shows the Ensembl Plants homepage. A red box highlights the 'Dataset' button in the top left. A dropdown menu titled '✓ - CHOOSE DATABASE -' is open, showing 'Plant Mart' (which is selected) and 'Plant Variation Mart'. Another dropdown menu titled '✓ - CHOOSE DATASET -' is also open, listing various plant datasets. The 'Aegilops tauschii genes (ASM34733v1 (2013-12-BGI))' dataset is currently selected.

- ✓ - CHOOSE DATABASE -
 - Plant Mart
 - Plant Variation Mart
- ✓ - CHOOSE DATASET -
 - Aegilops tauschii genes (ASM34733v1 (2013-12-BGI))
 - Amborella trichopoda genes (AMTR1.0 (2014-01-AGD))
 - Arabidopsis lyrata genes (v.1.0 (2008-12-Araly1.0))
 - Arabidopsis thaliana genes (TAIR10 (2010-09-TAIR10))
 - Brachypodium distachyon genes (v1.0 (2010-02-Brachy1.2))
 - Brassica oleracea genes (v2.1 (v2.1))
 - Brassica rapa genes (IVFCAAASv1 (bra_v1.01_SP2010_01))
 - Chlamydomonas reinhardtii genes (v3.1 (2007-11-ENA))
 - Cyanidioschyzon merolae genes (ASM9120v1 (2008-11-ENA))
 - Glycine max genes (V1.0 (GI-Glyma-1.1))
 - Hordeum vulgare genes (IBSC_1.0 (IBSC_1.0))

Database and dataset

Selecting the filters

The screenshot shows the Ensembl BioMart interface. On the left, there's a sidebar with sections for Dataset (Arabidopsis lyrata genes (v.1.0) (2008-12-Araly1.0)), Filters (highlighted with a red box), and Attributes (Gene stable ID, Transcript stable ID). The main area has a header "Please restrict your query using criteria below" and a list of filter categories: REGION, GENE, GENE ONTOLOGY, PLANT ONTOLOGY, MULTI-SPECIES COMPARISONS, and PROTEIN DOMAINS.

Click “Count” to see if BioMart is reading the input data

Limit your data set
(information that you know)

Picking the attributes

New Count Results URL

Dataset 696 / 32667 Genes
Arabidopsis lyrata genes (v.1.0)
(2008-12-Araly1.0))

Filters
Chromosome/scaffold: 2
Gene type : protein_coding
Transmembrane domains :
Only

Attributes
Peptide
Gene stable ID
Transcript stable ID

Please select columns to be included in the output ready

Features Homologs
 Structures Sequences
 Variation

+ SEQUENCES:

+ HEADER INFORMATION:

Determine output columns
(information you want to know)

The different attributes

New Count Results URL

Dataset 696 / 32667

Arabidopsis lyrata genome (2008-12-Araly1.0)

Filters

Chromosome/scaffold

Features Homologs
 Structures Variation
 Transcript Event Sequences

GENE:

EXON:

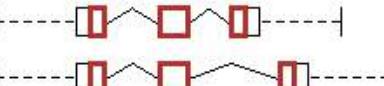
Exon Information

Exon Chr Start (bp)
 Exon Chr End (bp)
 Constitutive Exon
 Exon Rank in Transcript
 phase
 cDNA coding start

F S T

Please select columns to be included in the output

Features Homologs
 Structures Variation
 Transcript Event Sequences

SEQUENCES:
Sequences (max 1)


Unspliced (Transcript)
 Unspliced (Gene)
 Flank (Transcript)
 Flank (Gene)
 Flank-coding region (Transcript)
 Flank-coding region (Gene)

GENE:

GERMLINE VARIATION INFORMATION:

5' UTR 3' UTR

Features Homologs
 Structures Variation
 Transcript Event Sequences

TRANSCRIPT EVENT: (See PMID: [18978772](#) for type code key)

Ensembl

Ensembl Gene ID
 Ensembl Transcript ID

Splicing Event

Gene Name With Corresponding Event
 Chromosome Name
 Event Type
 Event Name

PARALOGS:

Getting the results

New Count Results

Dataset 696 / 32667 Genes
Arabidopsis lyrata genes (v.1.0 (2008-12-Araly1.0))

Filters

Chromosome/scaffold: 2
Gene type : protein_coding
Transmembrane
Only

Attribute

Peptide
Gene stable ID
Transcript stable ID

Dataset

Export all results to File FASTA Unique results only Go

Email notification to

View 10 rows as FASTA Unique results only

>Al_scaffold_0002_166|Al_scaffold_0002_166
>Al_scaffold_0002_1055|Al_scaffold_0002_1055
>Al_scaffold_0002_1070|Al_scaffold_0002_1070
>Al_scaffold_0002_1120|Al_scaffold_0002_1120
>Al_scaffold_0002_1134|Al_scaffold_0002_1134
>Al_scaffold_0002_1139|Al_scaffold_0002_1139
>Al_scaffold_0002_1178|Al_scaffold_0002_1178
>Al_scaffold_0002_1183|Al_scaffold_0002_1183

100 rows as HTML Unique results only

ID	Transcript stable ID	Peptide stable ID	Gene stable ID
Al_scaffold_0002_1015	9522259	9524695	
Al_scaffold_0002_1031	9507763	9508226	
Al_scaffold_0002_1055	9706008	9706423	
Al_scaffold_0002_1070	9801743	9802045	
Al_scaffold_0002_1120	10200845	10204513	ARK2
Al_scaffold_0002_1134	10313374	10313818	
Al_scaffold_0002_1139	10366414	10391298	
Al_scaffold_0002_1178	10778009	10779173	
Al_scaffold_0002_1183	10816424	10817970	

For the full table: click View “ALL” rows or “Go”

click “Unique results only”

Tables/sequences

A genome-wide transcriptional analysis using *Arabidopsis thaliana* Affymetrix gene chips determined plant responses to phosphate deprivation

Julie Misson*, Kashchandra G. Raghothama†, Ajay Jain†, Juliette Jouhet‡, Maryse A. Block‡, Richard Bligny‡,
Philippe Ortet§, Audrey Creff*, Shauna Somerville¶, Norbert Rolland‡, Patrick Doumas||, Philippe Nacry||,
Luis Herrera-Estrella**, Laurent Nussaume*, and Marie-Christine Thibaud*††

11934–11939 | PNAS | August 16, 2005 | vol. 102 | no. 33

www.pnas.org/cgi/doi/10.1073/pnas.0505266102

Selected ATH1 probes (Affy chip)

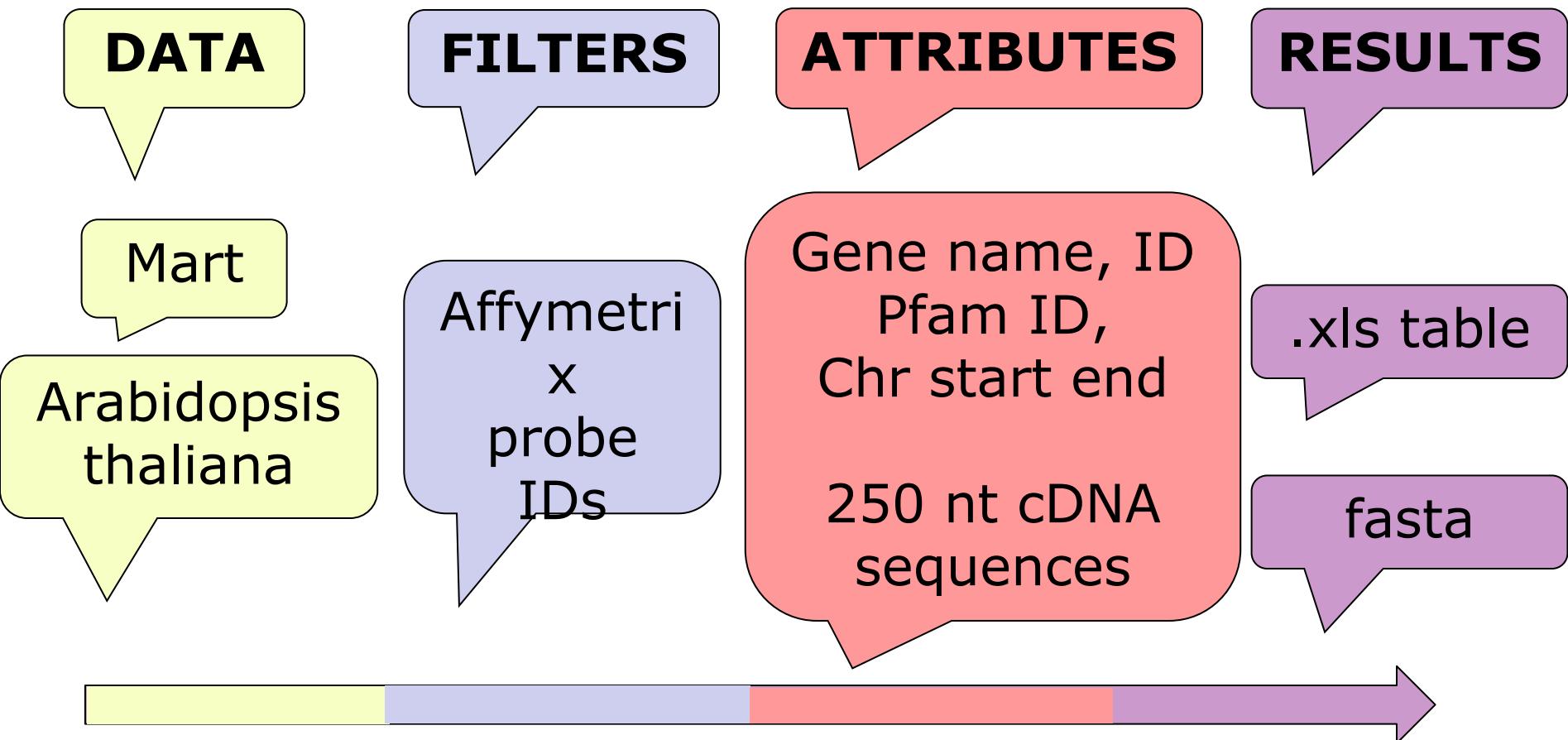
259842_at, 251193_at, 259303_at, 252534_at, 266957_at, 257891_at, 263593_at,
266372_at, 265342_at, 254011_at, 260623_at, 262238_at, 264118_at, 256910_at,
263846_at, 249996_at, 248094_at, 267361_at, 246275_at, 258034_at, 248622_at,
263483_at, 254250_at, 257964_at

Tutorial: BioMart

For this list of probes (Affymetrix array Arabidopsis), can I use BioMart to retrieve

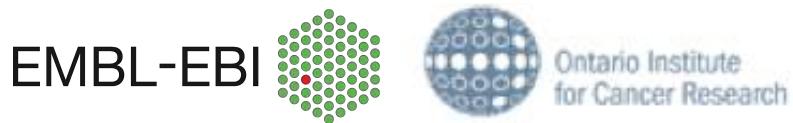
- a table with names and IDs, Pfam IDs, and location of the genes the probes map to?
- a fasta (sequence) file with 250 nt upstream of the cDNA sequences of those genes?

The four-step principle



Live demo

BioMart started at Ensembl



BioMart Central Portal

Home

IDENTIFIER SEARCH

Examples: KRAS, ENSG00000146648

TOOLS

Gene retrieval

Variant retrieval

Sequence retrieval

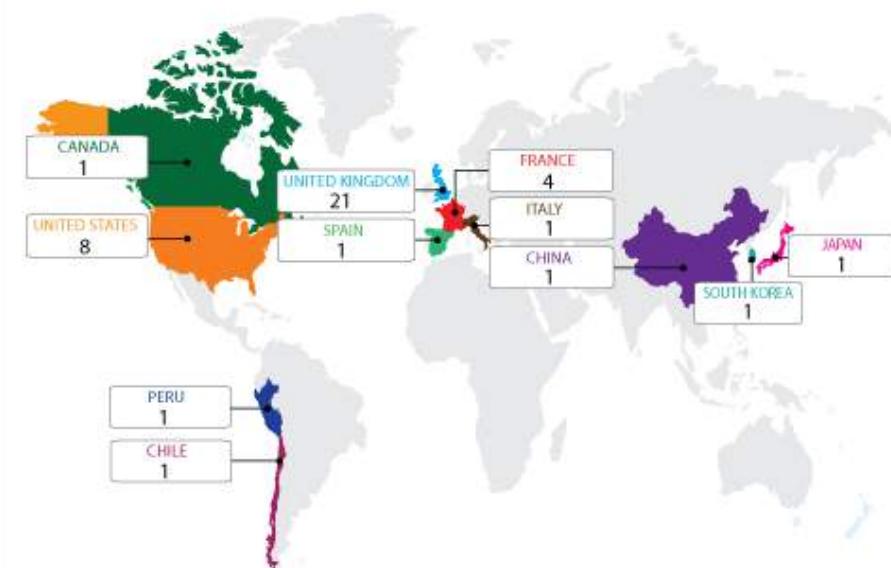
ID converter

Ensembl
COSMIC
VEGA
MGI
EnsemblGenomes
Gramene

Rfam
PRIDE
UniProt
InterPro
Reactome
and many more...

BIO MART CENTRAL PORTAL

Databases: 41



Find BioMart



<http://plants.ensembl.org/biomart/martview/>

central.biomart.org

More sophisticated platforms

- BioMart  queries: MartService
www.biomart.org/martservice.html
- APIs: PERL, Java, Web Services
- Third party softwares



taverna.org.uk



galaxyproject.org

Ensembl BioMarts



Database, Vol. 2011, Article ID bar030, doi:10.1093/database/bar030

Original article

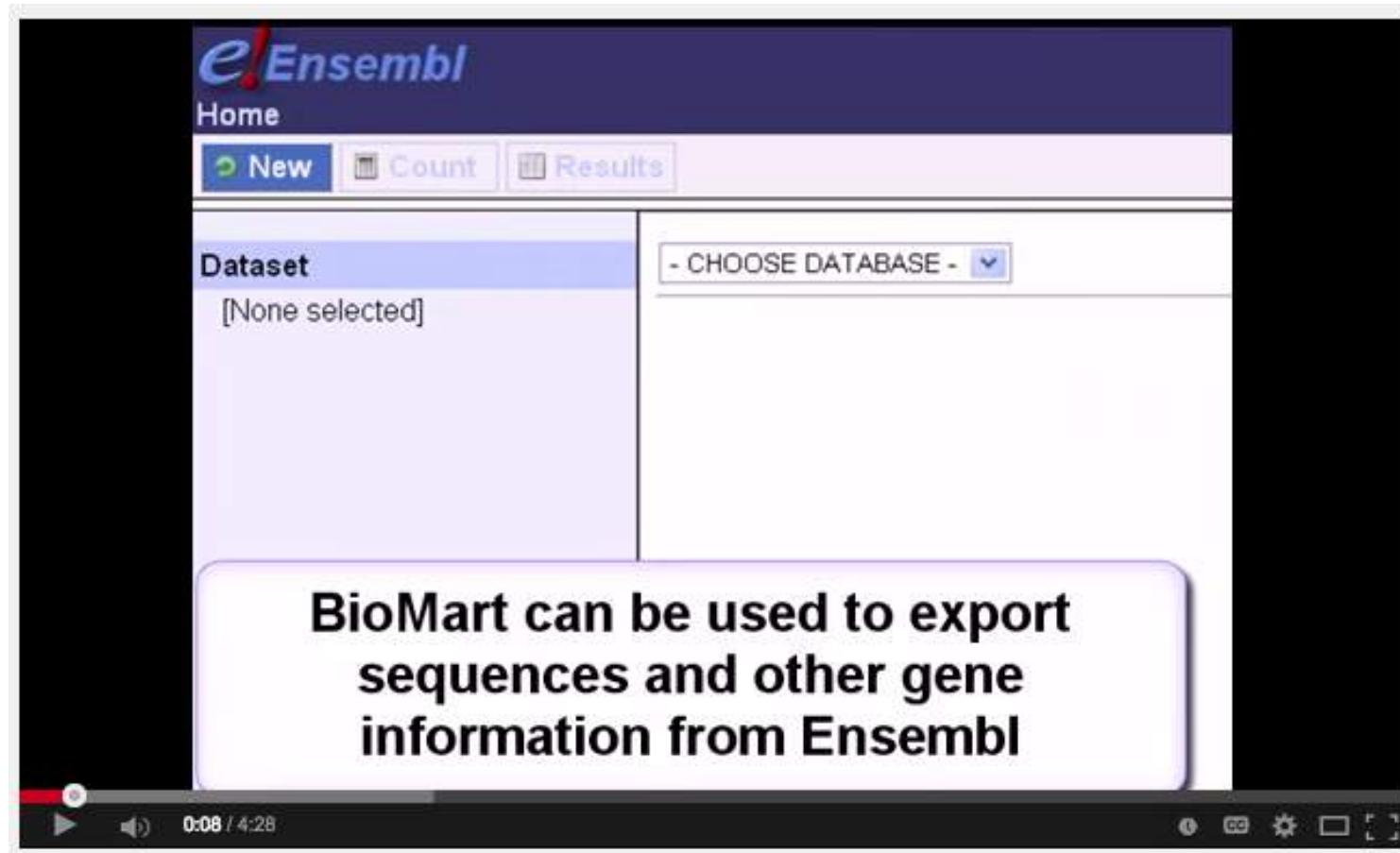
Ensembl BioMarts: a hub for data retrieval across taxonomic space

Rhoda J. Kinsella^{1,*}, Andreas Kähäri¹, Syed Haider², Jorge Zamora¹, Glenn Proctor¹, Giulietta Spudich¹, Jeff Almeida-King¹, Daniel Staines¹, Paul Derwent¹, Arnaud Kerhornou¹, Paul Kersey¹ and Paul Flicek^{1,*}

BioMart



video



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

Connect with Ensembl Genomes



helpdesk@ensemblgenomes.org

twitter.com/ensemblgenomes

EXERCISE

Mining variation data with BioMart

Exercise 5 – Retrieve a list of SNPs from tomato (*S. lycopersicum*)

The region between coordinates 21,394,819 and 21,397,868 on chromosome 6 in tomato contains a gene involved in oxidation-reduction process (GO: 0055114).

Can you use BioMart to retrieve all the SNPs that cause a change at the amino acid level of this gene (those SNPs are known as missense variants) including their IDs and possible alleles?

Exercise 6 – Open-ended BioMart

- Filter for some ‘interesting variations’ in a ‘region of interest’...



Export data and gather some ‘interesting’ statistics.

ANSWERS

Answer 1 – Exploring a SNP in Arabidopsis

- A. Search for the ATCDSP32 gene the Arabidopsis page in Ensembl Plants. On the left hand side menu of the Gene tab, click on Genetic variation to find out the number of variants that cause a change at the amino acid level of the ATCDSP32 protein. Those are known as missense variants. There are 16 of them.
- B. Next to the ‘Missense variant’ under the column Type, click on ‘Show’. Look for the residue 60, where a change from A to T is reported. The ID of this variant is ENSVATH05153232, located at position 28549171 on chromosome 1. The two possible alleles at this locus are C and T.
- C. Click on the link ENSVATH05153232. Then click on ‘Flanking sequence’ in the left hand side menu. Now click on ‘Download sequence’ and select the Rich Text Format (RTF). If you want to change how much flanking sequence is displayed on the browser, go back to the Flanking sequence page, click in the Configuration page and change the length of the sequence. The default settings is 400 bp.
- D. Click on ‘Genes and regulation’ to find out this SNP does not cause a change at the amino acid level for any other genes or transcripts in that genome. The other consequence types for this SNP are Downstream gene variant and Upstream gene variant.
- E. Click on ‘Genotype frequency’ in the left hand side menu to find out the most frequent genotype at this locus in the ‘1001 Population’ is CC (homozygous for allele C, the reference allele).

Answer 2 – Variation data in the tomato (*S. lycopersicum*) genome

- A. Go to the tomato page in the Ensembl Plants Genome Browser and search for Solyc02g085360.2 and click on the Location link in the results page. Configure the page and look for the data track named ‘The 150 TGRSP variations’.
- B. Zoom in around the last exon of this gene by drawing a box in the respective region. Please note the gene is on the reverse strand, so the last exon will be on the left hand side of that image. The types of variants seen in that region are 3 primer UTR variants, missense variants, synonymous variants and a few others. Look for the Variation legend of the different colour codes. The location of the only inframe deletion mapped in the region is 2:42861257-42861272. Just click on the pink box that denotes inframe deletion for a pop-up box with additional information on that variant including location.
- C. The splice region variant is coded in orange. Its reported alleles are G/A, G being the reference allele, and A the alternative allele.

Answer 3 - Missense variants in the bread wheat genome

- A. Search for Traes_2AL_5C7E76139.1 in bread wheat, click on the Gene ID in the results page and then on the transcript ID in the Transcript tab. Now click on ‘Variations’ to find out the two types are ‘missense variants’ and ‘synonymous variants’.
- B. One missense variants are predicted to be deleterious according to SIFT. The possible amino acid residues are D (Aspartic acid) and E (Glutamic acid).
- C. This missense variant has been reported by two different sources (inter-homoeologous variants from Ensembl Genomes and variants from CerealsDB) and gets two IDs, i.e. EPITAEV06993600 and BA00258972 as it’s a polymorphism within the different components of the hexaploid genome and between different populations of bread wheat.

Answer 4: The VEP tool, Variant Effect Predictor

Go the plants.ensembl.org and click on Tools on the top banner of the page. Click on VEP and select the species T. aestivum. Type in the following input data:

Scroll down in that page and click Next. Click on 'HTML' and you will see a table like this:

2D	89551917	89551917	G/A
2D	148408765	148408765	G/T
3D	113574123	113574123	C/A
3D	93827883	93827883	G/A
3B	727928129	727928129	C/T
3B	736734474	736734474	C/T
6A	196872409	196872409	T/G
6A	196153918	196153918	A/G
6A	196774882	196774882	G/C

Answer 4: The VEP tool, Variant Effect Predictor

EnsemblPlants - Personal Data Login/Register

Download text version

Show All entries Show/hide columns Filter

Uploaded Variation	Location	Allele	Gene	Feature	Feature type	Consequence	Position in cDNA	Position in CDS
en3_727928129_C/T	3B:727928129	T	TRAES3BF091000060CFD_g	TRAES3BF091000060CFD_t1	Transcript	missense_variant	1336	1336
me_113574123_C/A	3D:113574123	A	Traes_3DL_48F92563F	Traes_3DL_48F92563F.1	Transcript	missense_variant	235	400
ia_196774882_G/C	6A:196774882	C	Traes_6AL_77C81F092	Traes_6AL_77C81F092.1	Transcript	splice_region_variant, intron_variant	-	-
AS_736734474_C/T	3B:736734474	T	TRAES3BF052300070CFD_g	TRAES3BF052300070CFD_t1	Transcript	upstream_gene_variant	-	-
AS_93827883_G/A	3D:93827883	A	Traes_3DL_55D7A9A27	Traes_3DL_55D7A9A27.1	Transcript	downstream_gene_variant	-	-
M_148408765_G/T	2D:148408765	T	Traes_2DL_776B9B786	Traes_2DL_776B9B786.1	Transcript	stop_gained	853	770
se_89551917_G/A	2D:89551917	A	Traes_2DL_45F0B52C7	Traes_2DL_45F0B52C7.1	Transcript	missense_variant	567	400
H_736734474_C/T	3B:736734474	T	TRAES3BF052300060CFD_g	TRAES3BF052300060CFD_t1	Transcript	missense_variant	167	167
gi_93827883_G/A	3D:93827883	A	Traes_3DL_661D28EB1	Traes_3DL_661D28EB1.1	Transcript	missense_variant	851	770
se_196153918_A/G	6A:196153918	G	Traes_6AL_AC74E1A75	Traes_6AL_AC74E1A75.2	Transcript	stop_lost	555	555
se_196872409_T/G	6A:196872409	G	Traes_6AL_BBPF24273	Traes_6AL_BBPF24273.2	Transcript	missense_variant	1775	1775

Answer 4: The VEP tool, Variant Effect Predictor

- A. Several genes and transcripts have been annotated in the region where these variants map to, e.g.

Traes_2DL_776B9B786 / Traes_2DL_776B9B786.1

Traes_3DL_48F92563F / Traes_3DL_48F92563F.1

Traes_6AL_AC74E1A75 / Traes_6AL_AC74E1A75.2

- B. The consequences for these variants are missense, splice region, stop loss, downstream gene, among others. The missense variants do cause a change at the amino acid level.