



Arabidopsis resources in Gramene

SNP data

(Explore, Upload & Analyze)

Website: www.gramene.org

Pankaj Jaiswal
Oregon State University

Gramene Outreach Webinar
July 14, 2015



Topics to be covered

Genetic variation data on SNPs including indels

1. Explore publicly available and mapped variants.
2. Upload variation data identified in your project.
3. Predict structural and functional effect(s) of variants identified in your project.

Talk is on Arabidopsis data but much of the functionality remains the same for any other species.

Explore existing variants

- Publicly available
- Mapped to reference genome
- Pre-analyzed by Gramene & Ensembl Plants
- Data is freely accessible to everyone

The Arabidopsis variation database contains data on ~14.2 million SNPs

- 1,179 strains by [Nordborg et al.](#) using the Affymetrix 250k Arabidopsis SNP chip
- BBSRC funded resequencing of 18 *Arabidopsis* lines.
- 392 strains from the [1001 Genomes Project](#):
 - 80 strains from the Cao pilot study
 - 132 strains from a study by the [Salk Institute](#)
 - 180 strains from a study by the [Nordborg group at GMI](#)

Start at Gramene Home

www.gramene.org

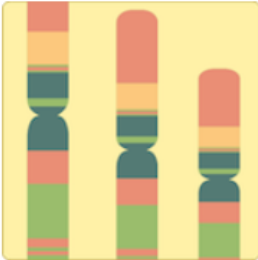
Navigation

- Current Release (45)
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- ▶ News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Mailing list
- Tools


Recent blog posts

- Fascination of Plants Day 2015

Gramene: A comparative resource for plants



Genomes



Pathways

Gramene is a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species. Our goal is to facilitate the study of cross-species comparisons using information generated from projects supported by public funds. Gramene currently hosts annotated whole genomes in over two dozen plant species and partial assemblies for almost a dozen wild rice species in the Ensembl browser, genetic and physical maps with genes, ESTs and QTLs locations, genetic diversity data sets, structure-function analysis of proteins, plant pathways databases (BioCyc and Plant Reactome platforms), and descriptions of phenotypic traits and mutations.

Sequenced plant genome dataset

<http://ensembl.gramene.org/info/about/species.html>

ensembl.gramene.org/genome_browser/index.html

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Search all species... Login/Register

Search: All species for

Go

e.g. Carboxy* or chx28

Popular genomes

- Arabidopsis thaliana** TAIR10
- Oryza sativa Japonica** IRGSP-1.0
- Triticum aestivum** IWGSC1.0+popseq
- Hordeum vulgare** 082214v1
- Zea mays** AGPv3
- Physcomitrella patens** ASM242v1

★ [Log in to customize this list](#)

Updated assembly for wheat

The [bread wheat assembly](#) in Gramene release 45 has been updated by ordering the existing [Chromosome Survey Sequence \(CSS\)](#) contigs into chromosomal pseudomolecules. This was done using anchoring data from high marker density population sequencing (POPSEQ) generated by the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben (Mascher *et al*, 2013) and processed at The Centre for Genome Analysis (TCGA) in Norwich.

The resulting set of chromosomes have been combined with the IWGSC's BAC by BAC assembly and annotation of chromosome 3B, generated by the [GDEC](#) group at INRA (Paux *et al*, 2008), which replaces the CSS-derived assembly and annotation of 3B.

The contig level gene models (PGSB - formerly known as MIPS version 2.2) were projected to the chromosome pseudomolecules using the [Assembly Contig Tool](#) (Wang *et al*, 2013) with the MIPS gene models as an annotation.

The assembly consists of 4,237 scaffolds, with cumulative length of 1,507,506,667 bp.

Every species has its own genome browser

Search results

tools.gramene.org/search?query=FRO1&fq=species~arabidopsis_thaliana

Gramene Search Ontologies Feedback

Species "Arabidopsis thaliana"

Showing 1 to 2 of 2 records.
Search took 0.194504 seconds.

GO

Arabisidopsis thaliana gene AT1G01590 (Probable ferric reduction oxidase 1 [Source:UniProtKB/Swiss-Prot;Acc:Q9LMM2])
AT1G01590.1.exon4 AT1G01590 A_84_P18906 **FRO1** 3702 Riboflavin_synthase-like_b-brl IPR017927
Fe_red_NAD-bd_6

Arabisidopsis thaliana gene AT5G67590 (NADH dehydrogenase [ubiquinone] iron-sulfur protein 4, mitochondrial [Source:UniProtKB/Swiss-Prot;Acc:Q9FJW4])
A_84_P15064 **FRO1** 3702 NADH_UbQ_FeS_4_mit IPR006885 AT5G67590 PO:0020100 PO:0007123 PO:0009010
PO:0009009 PO

Gene and transcript detail view

ensembl.gmane.org/Arabidopsis_thaliana/Gene/Summary?db=core;g=AT1G01590;r=1:214229-217304;t=AT1G01590.1

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Gene alleles
- Sequence
- Secondary Structure
- External references
- Regulation
- Ontology
- Plant Compara
- Pan-taxonomic Compara
- Phenotype
- Genetic Variation
- External data
- ID History

Gene: FRO1 AT1G01590

Description: Probable ferric reduction oxidase 1 [Source:UniProtKB/Swiss-Prot;Acc:Q9LMM2]

Location: Chromosome 1: 214,229-217,304 forward strand.

About this gene: This gene has 1 transcript ([splice variant](#)), [61 orthologues](#) and [17 paralogues](#).

Transcripts: [Hide transcript table](#)

| Name | Transcript ID | bp | Protein | Biotype | UniProt | RefSeq | Flags |
|------|---------------|------|---------|----------------|---------|------------------------|-------|
| FRO1 | AT1G01590.1 | 2115 | 704aa | Protein coding | Q9LMM2 | NM_100041 NP_171665 | |

Summary

Name: FRO1 (UniProtKB Gene Name)

UniProtKB: This gene has proteins that correspond to the following UniProtKB identifiers: [Q9LMM2](#)

Gene type: Protein coding

Annotation Method: Gene annotation by [TAIR](#) through a process of automatic and manual curation.

Go to [Region in Detail](#) for more tracks and navigation options (e.g. zooming)

23.08 kb Forward strand

TAIR: F22L4.11 > protein coding

ATMPK11 > protein coding

FRO2 > protein coding

FRO1 > protein coding

CYP86A4 > protein coding

Contigs: AC061957.3 >

TAIR: < ATGPAT4 < PIP1C

Variation: Comparison Image

ensembl.gemene.org/Arabidopsis_thaliana/Transcript/Population/Image?db=core;g=AT1G01590;r=1:214229-217304;t=AT1G01590.1

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Search Arabidopsis thaliana...

Arabidopsis thaliana Location: 1:214,229-217,304 Gene: FRO1 Transcript: FRO1

Transcript-based displays

- Summary
- Supporting evidence
- Sequence
 - Exons
 - cDNA
 - Protein
- External References
 - General identifiers
 - Oligo probes
- Ontology
 - GO: biological process (1)
 - GO: molecular function (2)
 - GO: cellular component (1)
- Genetic Variation
 - Variation table
 - Variation image
 - Population comparison
 - Comparison image**
- Protein Information
 - Protein summary
 - Domains & features
 - Variations
 - EBI protein data summary
 - Protein structure
- External data
 - Personal annotation
- ID History
 - Transcript history
 - Protein history

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

Transcript: FRO1 AT1G01590.1

Description Probable ferric reduction oxidase 1 [Source:UniProtKB/Swiss-Prot;Acc:Q9LMM2]

Location [Chromosome 1: 214,229-217,304](#) forward strand.

About this transcript This transcript has [8 exons](#), is annotated with [21 domains and features](#), is associated with [103 variations](#) and maps to [1 oligo probe](#).

Gene This transcript is a product of gene [AT1G01590](#) [Hide transcript table](#)

| Name | Transcript ID | bp | Protein | Biotype | UniProt | RefSeq | Flags |
|------|-----------------------------|------|-----------------------|----------------|------------------------|--|-------|
| FRO1 | AT1G01590.1 | 2115 | 704aa | Protein coding | Q9LMM2 | NM_100041 NP_171665 | |

Comparison image

Variations

Aa_0::Salk
Resequencing co...

Abd_0::Salk
Resequencing co...

Ag_0::Salk
Resequencing co...

Ak_1::Salk
Resequencing co...

Alst_1::Salk

ensembl.gramene.org/Arabidops

Gramene BLAST BioMart

Arabidopsis thaliana Location:

Transcript-based displays

- Summary
- Supporting evidence
- Sequence
 - Exons
 - cDNA
 - Protein
- External References
 - General identifiers
 - Oligo probes
- Ontology
 - GO: biological process (1)
 - GO: molecular function (2)
 - GO: cellular component (1)
- Genetic Variation
 - Variation table
 - Variation image
 - Population comparison
 - Comparison image**
- Protein Information
 - Protein summary
 - Domains & features
 - Variations
 - EBI protein data summary
 - Protein structure
- External data
 - Personal annotation
- ID History
 - Transcript history
 - Protein history

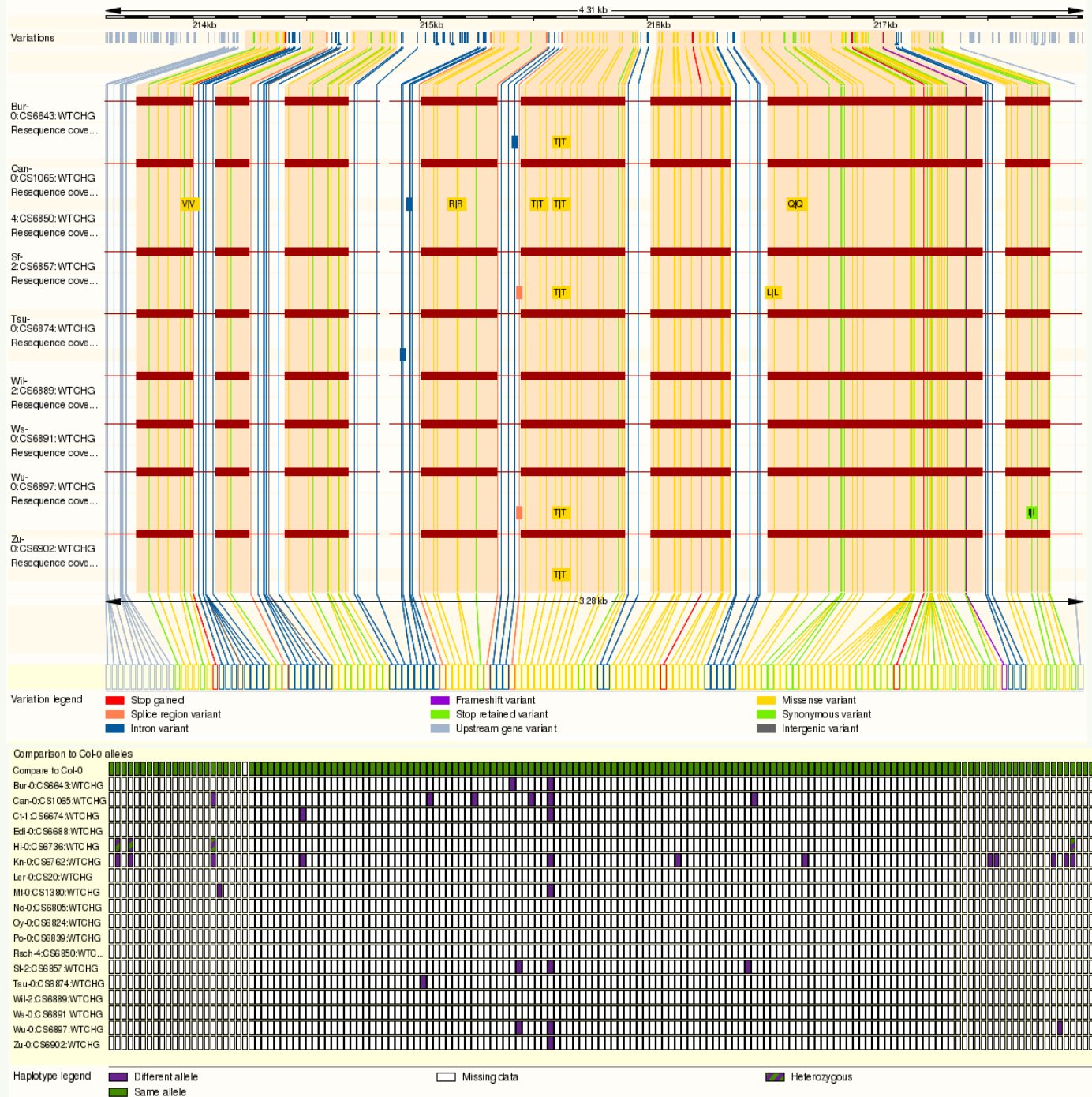
Configure this page

Manage your data

Export data

Share this page

Bookmark this page



Variation: Population Comparison

ensembl.gemene.org/Arabidopsis_thaliana/Transcript/Population?db=core;g=AT1G01590;r=1:214229-217304;t=AT1G01590.1

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Search Arabidopsis thaliana...

Arabidopsis thaliana Location: 1:214,229-217,304 Gene: FRO1 Transcript: FRO1

Transcript-based displays

- Summary
- Supporting evidence
- Sequence
 - Exons
 - cDNA
 - Protein
- External References
 - General identifiers
 - Oligo probes
- Ontology
 - GO: biological process (1)
 - GO: molecular function (2)
 - GO: cellular component (1)
- Genetic Variation
 - Variation table
 - Variation image
 - Population comparison**
 - Comparison image
- Protein Information
 - Protein summary
 - Domains & features
 - Variations
 - EBI protein data summary
 - Protein structure
- External data
 - Personal annotation
 - ID History
 - Transcript history
 - Protein history

Transcript: FRO1 AT1G01590.1

Description Probable ferric reduction oxidase 1 [Source:UniProtKB/Swiss-Prot;Acc:Q9LMM2]

Location [Chromosome 1: 214,229-217,304](#) forward strand.

About this transcript This transcript has [8 exons](#), is annotated with [21 domains and features](#), is associated with [103 variations](#) and maps to [1 oligo probe](#).

Gene This transcript is a product of gene [AT1G01590](#) [Hide transcript table](#)

| Name | Transcript ID | bp | Protein | Biotype | UniProt | RefSeq | Flags |
|------|-----------------------------|------|-----------------------|----------------|------------------------|--|-------|
| FRO1 | AT1G01590.1 | 2115 | 704aa | Protein coding | Q9LMM2 | NM_100041 NP_171665 | |

Population comparison

Variations in [Bla_1::Salk](#):

| ID | Type | Chr: bp | Ref. allele | Strain genotype | Ambiguity | Transcript codon | CDS coord. | Amino acids | AA coord. | Class | Source | Validation |
|---------------------------------|--------------------|----------|-------------|-----------------|-----------|------------------|------------|-------------|-----------|-------|---------|------------|
| ENSVATH01002192 | Synonymous variant | 1:216978 | G | A A | A | GCA GCA | 1863 | A A | 621 | SNP | Ensembl | - |
| ENSVATH00001008 | Synonymous variant | 1:216654 | A | G G | G | GAG GAG | 1539 | E E | 513 | SNP | Ensembl | - |
| ENSVATH00000998 | Missense variant | 1:214841 | G | C C | C | ATC ATC | 426 | I I | 142 | SNP | Ensembl | - |
| ENSVATH01002191 | Missense variant | 1:216969 | G | T T | T | AAT AAT | 1854 | N N | 618 | SNP | Ensembl | - |
| ENSVATH01002194 | Missense variant | 1:217277 | C | G G | G | CAG CAG | 2088 | Q Q | 696 | SNP | Ensembl | - |
| ENSVATH00001012 | Synonymous variant | 1:217262 | T | G G | G | TCG TCG | 2073 | S S | 691 | SNP | Ensembl | - |
| ENSVATH00001005 | Missense variant | 1:215760 | G | C C | C | ACT ACT | 848 | T T | 283 | SNP | Ensembl | - |
| ENSVATH00001007 | Missense variant | 1:216233 | T | C C | C | ACC ACC | 1238 | T T | 413 | SNP | Ensembl | - |

Variation: SNP detail view

ensembl.gramene.org/Arabidopsis_thaliana/Variation/Explore?db=core;g=AT1G01590;r=1:214229-217304;source=Ensembl;t=AT1G01590 Search

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData Search Arabidopsis thaliana...

Arabidopsis thaliana Location: 1:214,229-217,304 Gene: FRO1 Transcript: FRO1 Variation: ENSVATH00000998

Variation displays

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

Configure this page

Manage your data

Export data

Share this page

Bookmark this page

Ensembl Plants is produced in collaboration with Gramene

ENSVATH00000998 SNP

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

Alleles **G/C** | Ambiguity code: **S**

Location Chromosome **1:214841** (forward strand) | [View in location tab](#)

Most severe consequence **Missense variant** | [See all predicted consequences \[Genes and regulation\]](#)


Synonyms **Perlegen** PERL0001424

HGVS names This variation has **3** HGVS names - click the plus to show
[1:g.214841G>C](#)
[AT1G01590.1.1:c.426G>C](#)
[AT1G01590.1.1:p.Met142Ile](#)

About this variant This variant overlaps [3 transcripts](#) and has [51 individual genotypes](#).

Explore this variation

- Genomic context
- Genes and regulation **3**
- Genotype frequency
- Individual genotypes **51**
- Linkage disequilibrium
- Phenotype data
- Citations
- Phylogenetic context
- Flanking sequence



Variation: Linkage disequilibrium view

ensembl.gramene.org/Arabidopsis_thaliana/Variation/HighLD?db=core;g=AT1G01590;r=1:214229-217304;t=AT1G01590.1;v=ENSVATH0

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Arabidopsis thaliana Location: 1:214,229-217,304 Gene: FRO1 Transcript: FRO1 Variation: ENSVATH00000998

Variation displays

- Explore this variation
- Genomic context
 - Genes and regulation
 - Flanking sequence
- Genotype frequency
- Individual genotypes
- Linkage disequilibrium**
- Phenotype Data
- Phylogenetic Context
- Citations
- External Data

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

Ensembl Plants is produced in collaboration with Gramene

ENSVATH00000998 SNP

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

Alleles **G/C** | Ambiguity code: **S**

Location Chromosome **1:214841** (forward strand) | [View in location tab](#)

Most severe consequence Missense variant | [See all predicted consequences \(Genes and regulation\)](#)

Synonyms **Perlegen** PERL0001424

HGVS names This variation has **3** HGVS names - click the plus to show
[1:g.214841G>C](#)
[AT1G01590.1:c.426G>C](#)
[AT1G01590.1:p.Met142Ile](#)

About this variant This variant overlaps [3 transcripts](#) and has [51 individual genotypes](#).

Linkage disequilibrium

Links to linkage disequilibrium data by population

| Population | Description | Tags | Tagged by | Linked variations table | LD plot (image) | LD plot (table) |
|------------|--|------|-----------|-------------------------|----------------------|----------------------|
| Cao | Population of individuals/strains...(more) | - | - | Show | Show | Show |
| Nordborg | Population of individuals/strains...(more) | - | - | Show | Show | Show |
| Salk | Population of individuals/strains...(more) | - | - | Show | Show | Show |

Gramene release 46 - June 2015 © [EBI](#) [About Gramene](#) | [About EnsemblGenomes](#) | [Contact Us](#) | [Help](#)

Variation: Linkage disequilibrium view

ensembl.gramene.org/Arabidopsis_thaliana/Location/LD?db=core;focus=variation;g=AT1G01590;pop1=Nordborg;r=1:204841-224840

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Arabidopsis thaliana Location: 1:204,841-224,840 Gene: FRO1 Transcript: FRO1 Variation: ENSV

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
 - Alignments (image)
 - Alignments (text)
 - Region Comparison
 - Synteny
- Genetic Variation
 - Resequencing
 - Linkage Data**
 - Markers
- Other genome browsers
 - ARAPORT
 - Phytozome

Configure this page
Manage your data
Export data
Share this page
Bookmark this page
Select populations

Ensembl Plants is produced in collaboration with Gramene

Chromosome 1: 204,841-224,840

chromosome 1

Linkage Disequilibrium Data

Focus: Variant (Ensembl) [View variant]

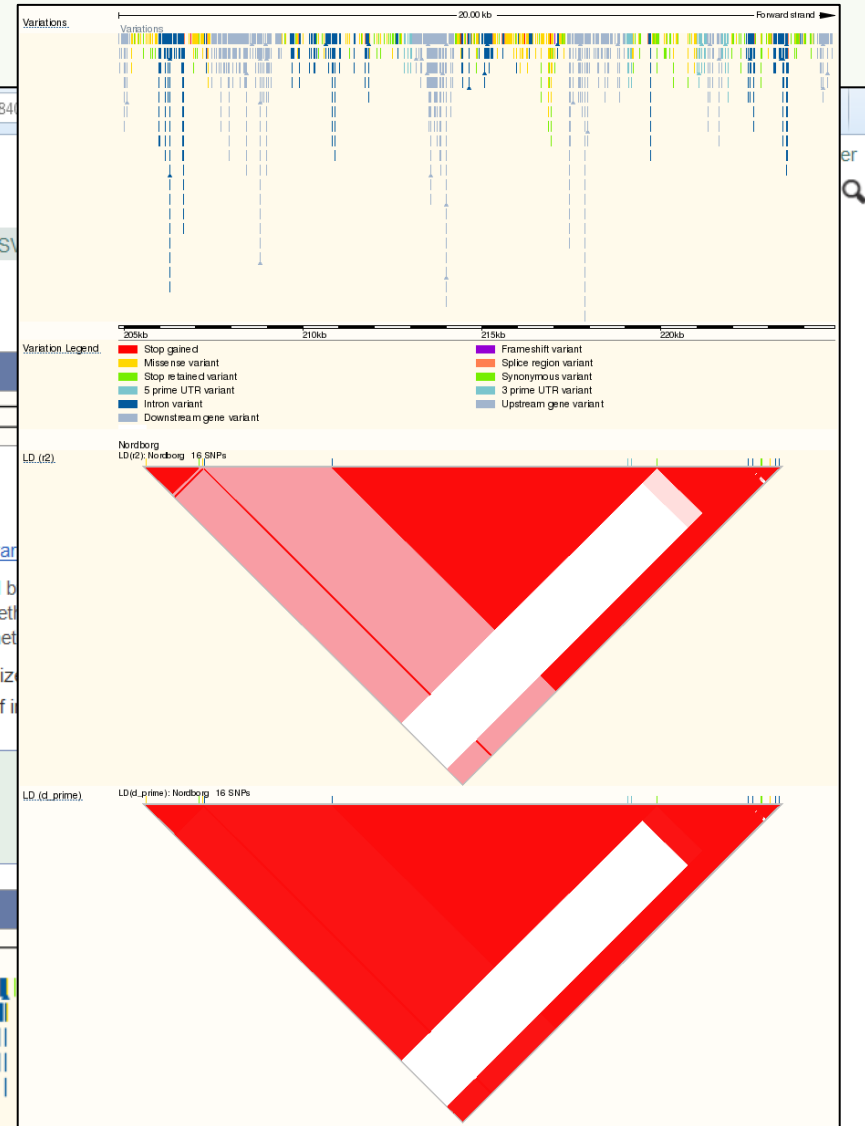
Prediction method: LD values were calculated by a sliding window. An established method was used to contribute to the double heterozygosity.

Populations: Nordborg [size: 16 SNPs]
Description: Population of individuals

Location: 1:204841-224840

Gene:

Variations



Upload SNP data from your projects

Data types & formats that you can upload

• Data types (Pre-analyzed)

- SNP Variations including allele information identified by
 - Genome sequencing
 - Genotype by sequencing
 - Genotyping chips
 - Transcriptome sequencing
 - Etc.

• Data Formats

- VCF (for variations)

Get more information on VCF format from

<http://tinyurl.com/snp-vcf>

For more information visit:

<http://ensembl.gramene.org/info/website/upload/index.html>

View of a typical VCF file with SNP data

```
##fileformat=VCFv4.0
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##FILTER=<ID=q25,Description="Quality below 25">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT 108
1 175877 . A T 40 PASS NS=1;DP=25 GT:GQ:DP 1:40:25
1 175878 . A C 40 PASS NS=1;DP=26 GT:GQ:DP 1:40:26
1 214406 . G T 40 PASS NS=1;DP=65 GT:GQ:DP 1:40:65
1 215105 . T A 40 PASS NS=1;DP=52 GT:GQ:DP 1:40:52
1 139701 . G A 40 PASS NS=1;DP=69 GT:GQ:DP 1:40:69
1 139732 . C A 40 PASS NS=1;DP=61 GT:GQ:DP 1:40:61
1 147110 . T C 40 PASS NS=1;DP=74 GT:GQ:DP 1:40:74
1 148345 . C T 40 PASS NS=1;DP=75 GT:GQ:DP 1:40:75
1 148357 . T C 40 PASS NS=1;DP=77 GT:GQ:DP 1:40:77
1 148601 . C G 40 PASS NS=1;DP=67 GT:GQ:DP 1:40:67
1 139983 . T C 38 PASS NS=1;DP=41 GT:GQ:DP 1:38:41
1 139991 . CT G 40 PASS NS=1;DP=45 GT:GQ:DP 1:40:45
1 140005 . C G 40 PASS NS=1;DP=59 GT:GQ:DP 1:40:59
1 140030 . C A 40 PASS NS=1;DP=64 GT:GQ:DP 1:40:64
2 1287854 . T C 12 q25 NS=1;DP=6 GT:GQ:DP 1:12:6
2 1288090 . A G 40 PASS NS=1;DP=46 GT:GQ:DP 1:40:46
2 1288284 . G C 40 PASS NS=1;DP=73 GT:GQ:DP 1:40:73
2 1288440 . T C 40 PASS NS=1;DP=52 GT:GQ:DP 1:40:52
2 1288804 . G C 40 PASS NS=1;DP=54 GT:GQ:DP 1:40:54
2 1288808 . A T 40 PASS NS=1;DP=49 GT:GQ:DP 1:40:49
2 1288905 . C T 40 PASS NS=1;DP=49 GT:GQ:DP 1:40:49
2 1288930 . C T 40 PASS NS=1;DP=56 GT:GQ:DP 1:40:56
2 1289238 . T G 5 q25 NS=1;DP=8 GT:GQ:DP 1:5:8
2 1289411 . T C 40 PASS NS=1;DP=28 GT:GQ:DP 1:40:28
3 10447781 . G A 38 PASS NS=1;DP=16 GT:GQ:DP 1:38:16
3 10447786 . G A 40 PASS NS=1;DP=16 GT:GQ:DP 1:40:16
3 10447826 . T C 40 PASS NS=1;DP=16 GT:GQ:DP 1:40:16
3 10447843 . C A 40 PASS NS=1;DP=16 GT:GQ:DP 1:40:16
3 10447854 . A C 40 PASS NS=1;DP=16 GT:GQ:DP 1:40:16
3 10447885 . C A 40 PASS NS=1;DP=16 GT:GQ:DP 1:40:16
4 2452185 . T C 40 PASS NS=1;DP=31 GT:GQ:DP 1:40:31
4 2452209 . A T 40 PASS NS=1;DP=31 GT:GQ:DP 1:40:31
4 2452212 . C G 40 PASS NS=1;DP=31 GT:GQ:DP 1:40:31
4 2452272 . A G 40 PASS NS=1;DP=31 GT:GQ:DP 1:40:31
4 2452299 . T G 32 PASS NS=1;DP=31 GT:GQ:DP 1:32:31
5 25221157 . C A 40 PASS NS=1;DP=16 GT:GQ:DP 1:40:16
```

Get more information on VCF format from <http://tinyurl.com/snp-vcf>

Location of your data

- Via URL
 - Especially for large files (filename.vcf)
 - Direct links to your data files
 - e.g. from iPlant data store or any file share server
- From your computer (20MB file size limit)
- Distributed Annotation System (DAS)

Uploaded data is saved temporarily to our file system and accessed by you via cookie, so only you can view it from the browser in which you uploaded it. For more information visit...

<http://ensembl.gramene.org/info/website/upload/index.html#access>

Access data upload tools

<http://ensembl.gramene.org/info/about/species.html>

The screenshot shows the Ensembl Gramene website interface. At the top, the navigation menu includes 'BLAST', 'BioMart', 'Tools', 'Downloads', 'Help', 'Feedback', and 'UploadData' (highlighted with a red box). A search bar on the right contains the text 'Search all species...'. Below the navigation, there is a search input field with a dropdown menu set to 'All species' and a 'Go' button. Below this, a text prompt reads 'e.g. Carboxy* or chx28'. The 'Popular genomes' section is located below the search area, featuring a grid of genome entries. A red arrow points to the 'Arabis thaliana' entry. The entries include:

- Arabis thaliana** (TAIR10)
- Oryza sativa Japonica** (IRGSP-1.0)
- Triticum aestivum** (IWGSC1.0+popseq)
- Hordeum vulgare** (082214v1)
- Zea mays** (AGPv3)
- Physcomitrella patens** (ASM242v1)

At the bottom of the 'Popular genomes' section, there is a link: '★ [Log in to customize this list](#)'. On the right side of the page, there is a section titled 'Updated assembly for wheat' with text describing updates to the bread wheat assembly and the resulting chromosome pseudomolecules. A large green box with black text is overlaid on the bottom right of the screenshot, containing the text: 'Every species has its own genome browser'.

Data Upload View

The screenshot shows the Gramene Data Upload View interface. On the left is a navigation menu under 'Personal Data' with 'Add your data' highlighted. The main area is titled 'Select File to Upload' and contains several form fields: 'Name for this data (optional):' with the value 'My-SNP-Data-2015'; 'Species:' with a dropdown menu showing 'Arabidopsis thaliana'; 'Assembly:' with the value 'TAIR10'; 'Data format:' with a dropdown menu showing 'VCF'; 'Type:' with radio buttons for 'Upload data (max 20MB)' (selected) and 'Attach via URL'; and 'Paste data:' with an empty text area. Below these fields are options to 'Or choose file:' (with a 'Browse...' button and a file name 'ath-mysnp.vcf') and 'Or provide file URL:' (with an empty text area). At the bottom is an 'Upload' button. Red arrows point to the 'Add your data' menu item, the 'Name for this data' field, the 'Species' dropdown, the 'Data format' dropdown, the 'Paste data' text area, the 'Browse...' button, and the 'Upload' button.

Personal Data

- Login
- Register
- Lost Password
- Custom Data
 - Add your data**
 - Attach DAS
 - Manage Data
 - Features on Karyotype
- Manage Configurations
 - Configurations for this page
 - All configurations
 - Configuration sets
- Online Tools
 - Variant Effect Predictor
 - Assembly Converter
 - ID History Converter
 - Data Slicer
- Help

Select File to Upload

Add a custom track

Name for this data (optional):

Species:

Assembly: TAIR10

Data format:

[Help on supported formats, display types, etc](#)

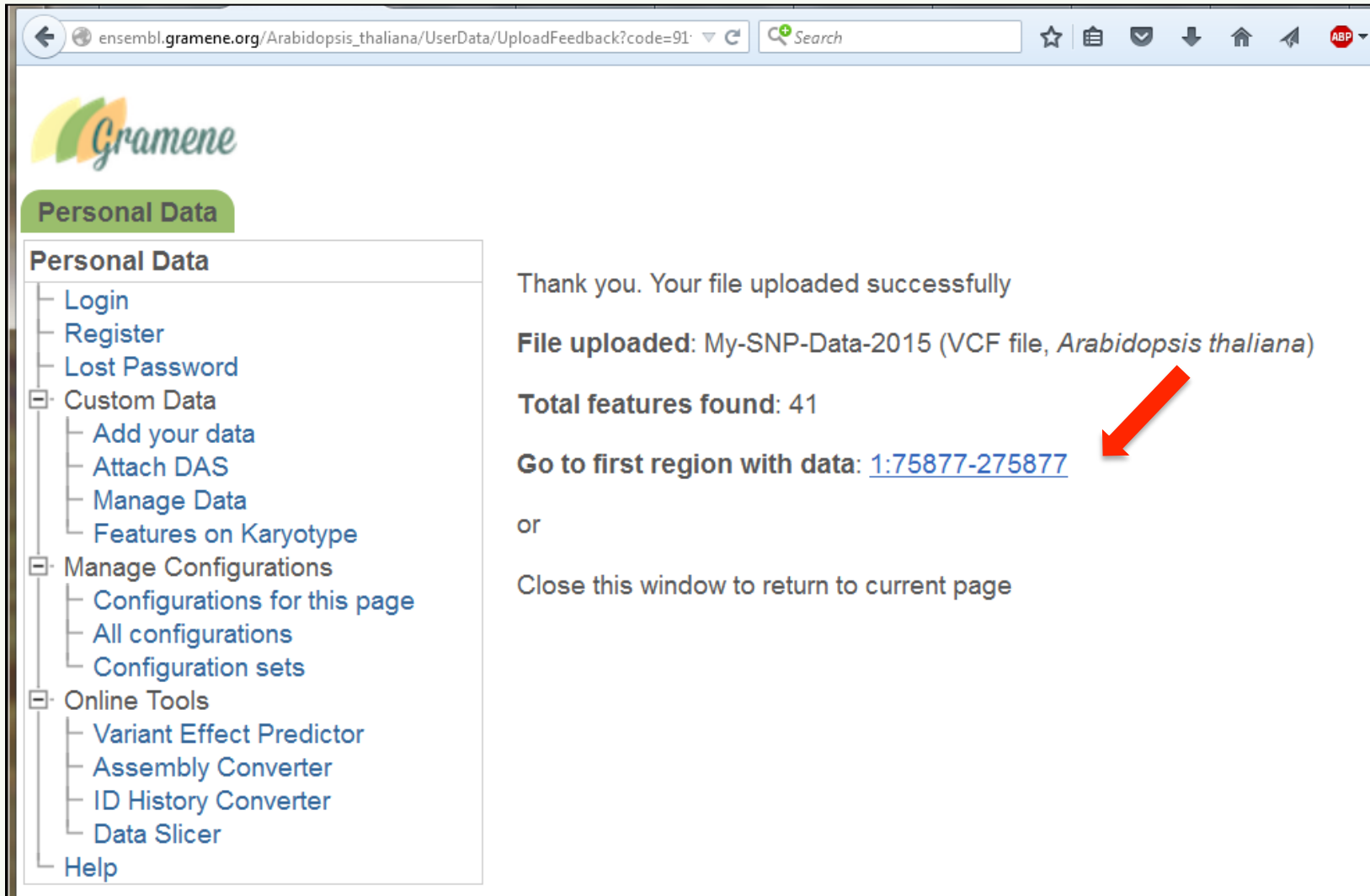
Type:
 Upload data (max 20MB)
 Attach via URL

Paste data:

Or choose file:


Or provide file URL:

Data Upload Results View



ensembl.gramene.org/Arabidopsis_thaliana/UserData/UploadFeedback?code=91

Search



Personal Data

Personal Data

- Login
- Register
- Lost Password
- [-] Custom Data
 - Add your data
 - Attach DAS
 - Manage Data
 - Features on Karyotype
- [-] Manage Configurations
 - Configurations for this page
 - All configurations
 - Configuration sets
- [-] Online Tools
 - Variant Effect Predictor
 - Assembly Converter
 - ID History Converter
 - Data Slicer
- Help

Thank you. Your file uploaded successfully


File uploaded: My-SNP-Data-2015 (VCF file, *Arabidopsis thaliana*)

Total features found: 41

Go to first region with data: [1:75877-275877](#)

or

Close this window to return to current page

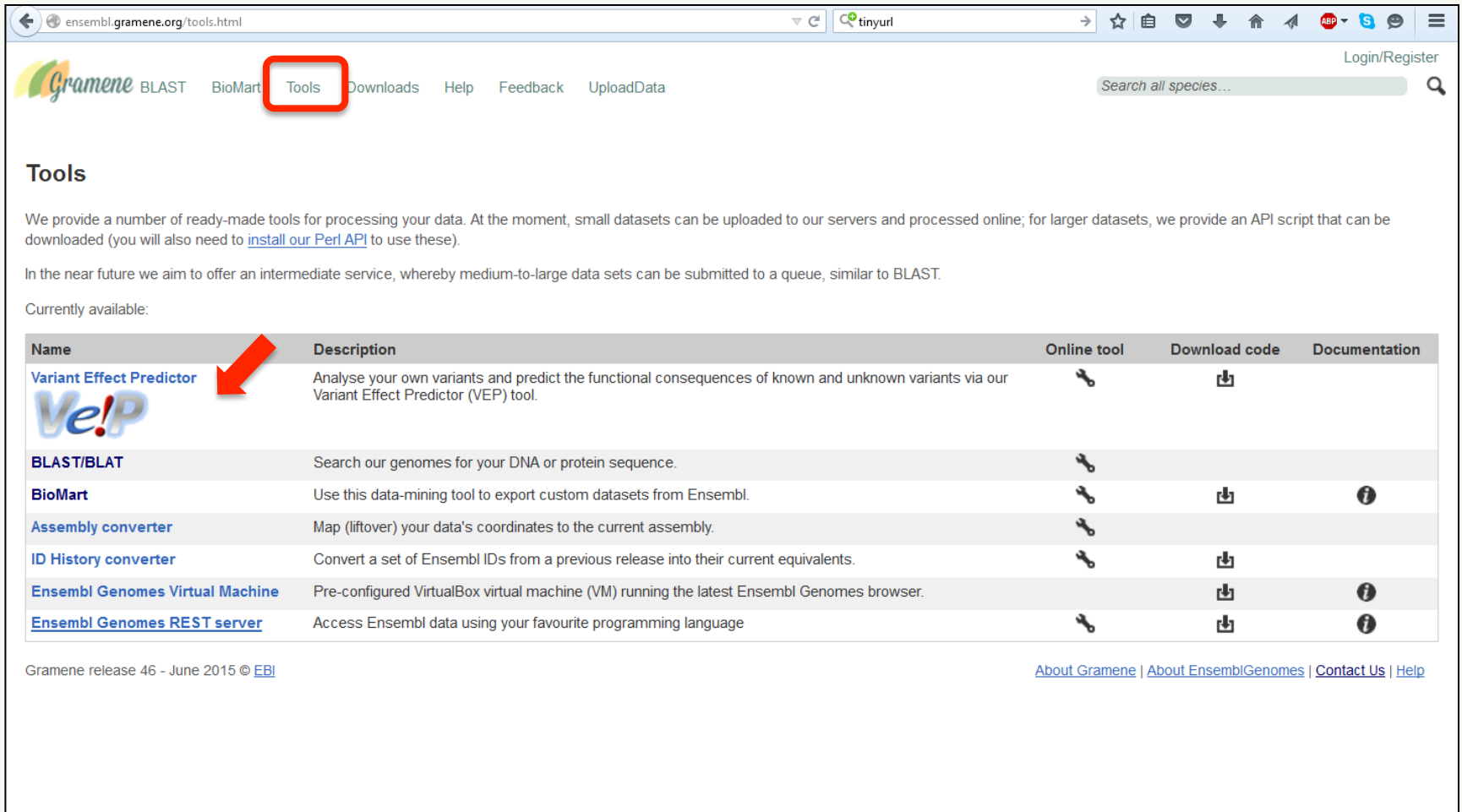


Variant Effect Predictor (VEP) tool

Used for predicting effects of variants on structure & function of gene products

Keep your SNP data file in exactly the same VCF file format

Access to VEP tool



ensembl.gramene.org/tools.html

Gramene BLAST BioMart **Tools** Downloads Help Feedback UploadData

Login/Register
















Search all species...

Tools

We provide a number of ready-made tools for processing your data. At the moment, small datasets can be uploaded to our servers and processed online; for larger datasets, we provide an API script that can be downloaded (you will also need to [install our Perl API](#) to use these).

In the near future we aim to offer an intermediate service, whereby medium-to-large data sets can be submitted to a queue, similar to BLAST.

Currently available:

| 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 Genomes Virtual Machine | Pre-configured VirtualBox virtual machine (VM) running the latest Ensembl Genomes browser. | |  |  |
| Ensembl Genomes REST server | Access Ensembl data using your favourite programming language |  |  |  |

Gramene release 46 - June 2015 © EBI

[About Gramene](#) | [About EnsemblGenomes](#) | [Contact Us](#) | [Help](#)

Upload SNP data using the VEP tool

The screenshot shows the VEP tool interface with several red arrows pointing to specific fields: the Species dropdown (Arabisopsis thaliana: TAIR10), the Name for this data field (My-data-01), the Paste data field (containing VCF-style variant data), the Input file format dropdown (set to VCF), the Type of consequences to display dropdown (set to Sequence Ontology terms), the Check for existing co-located variants dropdown (set to Yes), and the Next button.

Personal Data

- Manage Configurations
 - Configurations for this page
 - All configurations
 - Configuration sets
- Online Tools
 - Variant Effect Predictor**
 - Assembly Converter
 - ID History Converter
 - Data Slicer
- Help

Tools

- We provide script th...
- In the ne...
- Current...

Name

- Variant...
- BLAST
- BioMart
- Assem...
- ID Hist...
- Ensembl
- Machin...
- Ensembl

Search all species...

Ve!P

This tool takes a list of variant positions and alleles, and predicts the effects of each of these on overlapping transcripts and regulatory regions annotated in Ensembl. The tool accepts substitutions, insertions and deletions as input, see [data formats](#).

Upload is limited to 750 variants; lines after the limit will be ignored. Users with more than 750 variations can split files into smaller chunks, use the standalone [perl script](#) or the [variation API](#). See also [full documentation](#)

NB: Ensembl now by default uses Sequence Ontology terms to describe variation consequences. See [this page](#) for details

Input file

Species: Arabisopsis thaliana: TAIR10

Name for this data (optional): My-data-01

Paste data:

```
T      A      40      PASS
NS=1;DP=33      GT:GQ:DP      1:40:33
5      25721437      .
C      T      40      PASS
NS=1;DP=21      GT:GQ:DP      40:21
```

Upload file: Browse... No file selected.

or provide file URL:

Input file format: Ensembl default (dropdown menu with options: Ensembl default, VCF)

Options

Transcript database to use:

- Ensembl transcripts
- RefSeq and other transcripts

Type of consequences to display: Sequence Ontology terms

Check for existing co-located variants: Yes

Return results for variants in coding regions only:

Show Ensembl protein identifiers where available:

Show HGVS identifiers for variants where available: Coding and protein sequence

Missense SNP predictions

SIFT predictions: Prediction and score

Next >

VEP Results

ensembl.gemene.org/tools.html

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Search all species...

Login/Register

Personal Data

Login

Register

Lost Password

Custom Data

- Add your data
- Attach DAS
- Manage Data
- Features on Karyotype

Manage Configurations

- Configurations for this page
- All configurations
- Configuration sets

Online Tools

- Variant Effect Predictor
- Assembly Converter
- ID History Converter
- Data Slicer

Help

Variant Effect Predictor

Please select the format you would like your output in:

- [HTML](#)
- [Text](#)

or view a sample SNP in [Region in Detail](#)

(You can also view or download your converted file from 'Manage Your Data')

Documentation

Gramene release 46 - June 2015 © EBI

[About Gramene](#) | [About EnsemblGenomes](#) | [Contact Us](#) | [Help](#)

VEP Results

ensembl.gramene.org/tools.html

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Personal Data

Variant Effect Predictor Results:

[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 | Position in protein | Amino acid change | Codon change | Co-located Variation | Extra |
|--------------------|--------------------------|--------|---------------------------|-----------------------------|--------------|---------------------------------------|------------------|-----------------|---------------------|-------------------|--------------|---------------------------------|---|
| - | 1:175878 | C | AT1G01480 | AT1G01480.1 | Transcript | missense_variant | 97 | 17 | 6 | K/T | aAa/aCa | ENSVATH04502568 | ENSP=AT1G01480.1; STRAND=1; IMPACT=MODERATE; HGVS=AT1G01480.1:c.17A> HGVS=AT1G01480.1:p.Lys61 |
| - | 1:175877 | T | AT1G01480 | AT1G01480.1 | Transcript | stop_gained | 96 | 16 | 6 | K/* | Aaa/Taa | ENSVATH04502567 | ENSP=AT1G01480.1; STRAND=1; IMPACT=HIGH; HGVS=AT1G01480.1:c.16A> HGVS=AT1G01480.1:p.Lys61 |
| - | 1:214406 | T | AT1G01590 | AT1G01590.1 | Transcript | stop_gained, splice_region_variant | 178 | 178 | 60 | G/* | Gga/Tga | ENSVATH00000996 | ENSP=AT1G01590.1; STRAND=1; IMPACT=HIGH; HGVS=AT1G01590.1:c.178G> HGVS=AT1G01590.1:p.Gly60 |
| - | 1:148601 | G | AT1G01390 | AT1G01390.1 | Transcript | synonymous_variant | 1206 | 1161 | 387 | A | gcG/gcC | ENSVATH00000627 | ENSP=AT1G01390.1; STRAND=-1; IMPACT=LOW; HGVS=AT1G01390.1:c.1161C> HGVS=AT1G01390.1:c.1161C> |
| - | 1:148357 | C | AT1G01390 | AT1G01390.1 | Transcript | missense_variant | 1450 | 1405 | 469 | T/A | Acg/Gcg | ENSVATH00000626 | ENSP=AT1G01390.1; STRAND=-1; IMPACT=MODERATE; HGVS=AT1G01390.1:c.1405A> |

View VEP Results on the Genome Browser

plants.ensembl.org/Arabidopsis_thaliana/Location/View?r=1:214404-214408

Region in detail

Contigs
TAIR

Gene Legend
protein coding RNA gene

Location: 1:214404-214408 Gene: FRO1

Variation: EN SVATH00000996
more about EN SVATH00000996
Class SNP
Location 1:214406
Alleles G/A/T
Ambiguity code D
Consequence Stop gained
Evidence -
Source Perlegen, Ensembl

default -
Start 214406
End 214406
Strand -
Type stop_gained,splice_region_variant,upstream_gene_variant,downstream_gene_variant

Blocks show the locations of G/C base pairs.

Ensembl Plants is produced in collaboration with Gramene

View VEP Results on the Genome Browser

The screenshot displays the Ensembl genome browser interface for Arabidopsis thaliana. The browser address bar shows the URL: ensembl.gramene.org/Arabidopsis_thaliana/Location/View?db=core;r=1:214386-214423;time=1436840679. The main content area is divided into several sections:

- Configuration Region Image**: Includes options for Active tracks, Favourite tracks, Track order, Search results, Your data (4/4), Sequence and assembly (2/6), Genes and transcripts (0/3), mRNA and protein alignments (5/5), Variation (0/29), Comparative genomics (0/43), Oligo probes (0/4), Repeat regions (0/13), Information and decorations (10/11), and Display options.
- Configuration Overview Image**: Includes a search bar labeled "Find a track".
- Configuration Chromosome Image**: Includes a "Personal Data" section.
- Personal Data**: Includes sections for "Enable/disable all Sequence variants", "Enable/disable all Variation sets", "Enable/disable all 1001_All Variations by Class", "Enable/disable all Cao_All variations", "Enable/disable all Nordborg_All Variations", and "Enable/disable all Salk_All Variations".

Two red arrows point to the "Variation" track in the left sidebar and the "Enable/disable all Sequence variants" checkbox in the Personal Data section. A red circle highlights a checkmark icon in the top right corner of the interface.

Legend: ■ protein coding ■ Fgenesh

Ensembl Plants Arabidopsis thaliana version 80.10 (TAIR10) Chromosome 1: 214,386 - 214,423



Gramene

BLAST

BioMart

Tools

Downloads

Help

Feedback

For Questions and Feedback, visit
<http://tools.gramene.org/feedback>



Search

Home

Release Notes

Collaborators

Contact

About

Archive

Citing Gramene

Videotutorials

Outreach

For previous webinar recordings and
video tutorials, visit
<http://tinyurl.com/gramene-tutorial>

Acknowledgements



Funded by the National Science Foundation's
Plant Genome Research Program award IOS#1127112

Doreen Ware, PI (USDA ARS, CSHL)

Kapeel Cougule, Yinping Jiao, Sunita Kumari, Marcela K. Monaco, Andrew Olson, Joshua Stein, Jim Thomason, Peter van Buren, Sharon Wei, Bo Wang

Pankaj Jaiswal, Co-PI (OSU)

Vindhya Amarasinghe, Justin Elser, Sushma Naithani, Justin Preece

Paul Kersey / Robert Petrysyk (EMBL-EBI)

Dan Bolser, Arnaud Kerhornou, Eugene Kulesha, Dan Staines, Brandon Walts / Elisabet Barrera, Maria Keays, Oliver Mannion

Lincoln Stein (OICR)

Robin Haw; Peter D' Eustachio (NYU); Guanming Wu; David Croft (EBI)

Crispin Taylor (ASPB)

Patty Lockhart

