



Plant Genomes & Pathways

Comparative Data Views & Analysis

Marcela Karey Tello-Ruiz, PhD
Cold Spring Harbor Laboratory
October 19, 2016



2016 CSHL Cereal Genomics Course

Outline

- Overview
 - 1) Search
 - 2) Genomes
 - 3) Pathways
 - Educational resources
 - Exercises
- Public Plant Data &
 - Analysis Tools for Your Data



2016 CSHL Cereal Genomics Course

Genomes

Reference sequence assemblies & comparative genomics (synteny, WGs)

Gene annotations & gene-centered phylogenetic data (gene family trees, orthologs/paralogs)

Genetic & structural variation (SNPs, EMS-mutants, indels)

Download data (browser, Mart, FTP) & graphs (browser)

Upload & analyze your own data - Tools
(assembly/ID converter, etc.)

Pathways

Pathway browser

Omics analysis

Species comparison

Expression

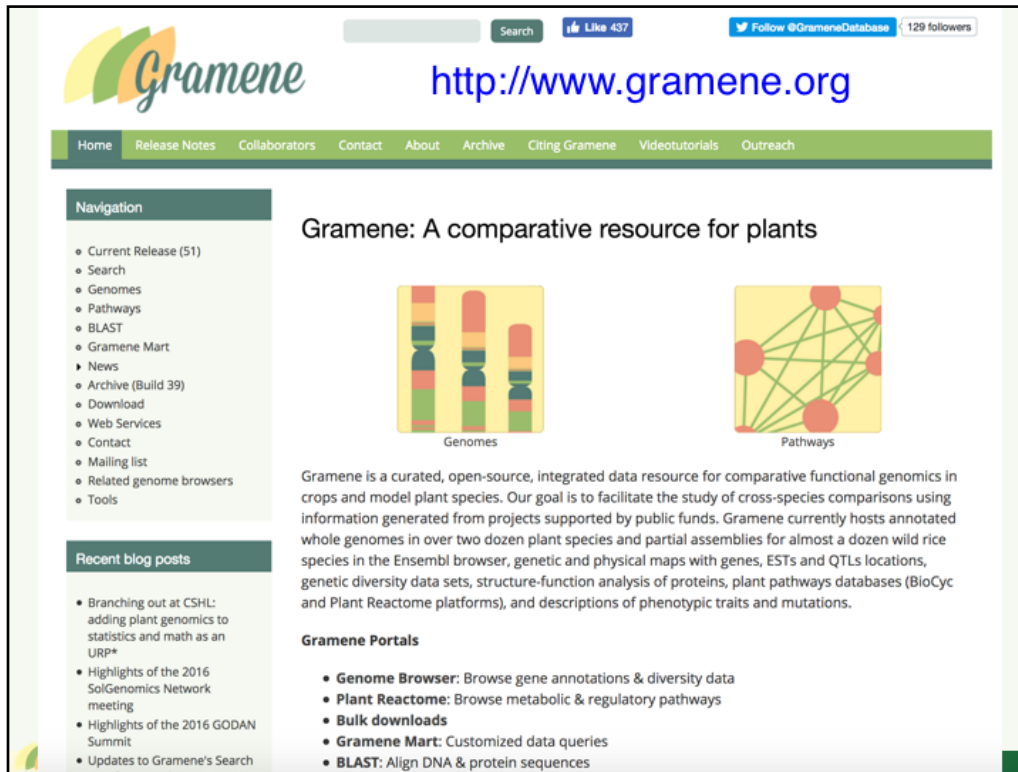
ATLAS widget (genome browser &
pathways views)

Videotutorials in the Gramene

YouTube channel:

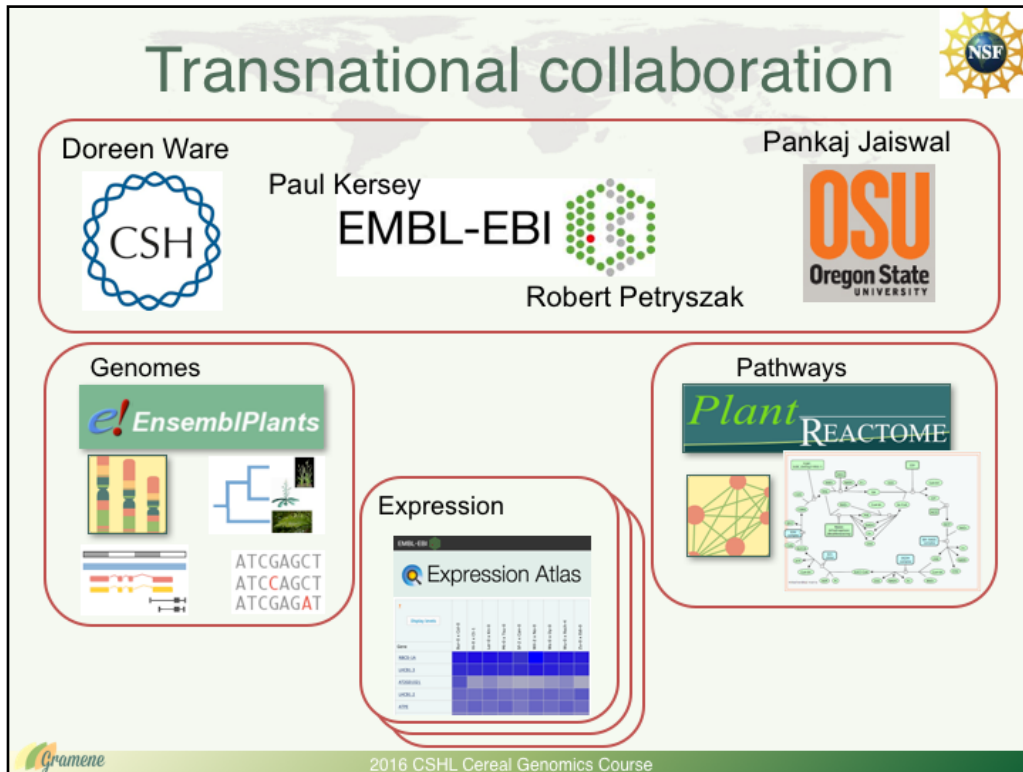
<https://goo.gl/ln9RLD>

Use case: Explore functional variants
in *cle18*



Search entry points

- 1) General Search => search.gramene.org
- 2) Genome Browser (ensembl style)
- 3) Plant Reactome (reactome style)



<Define Gramene>

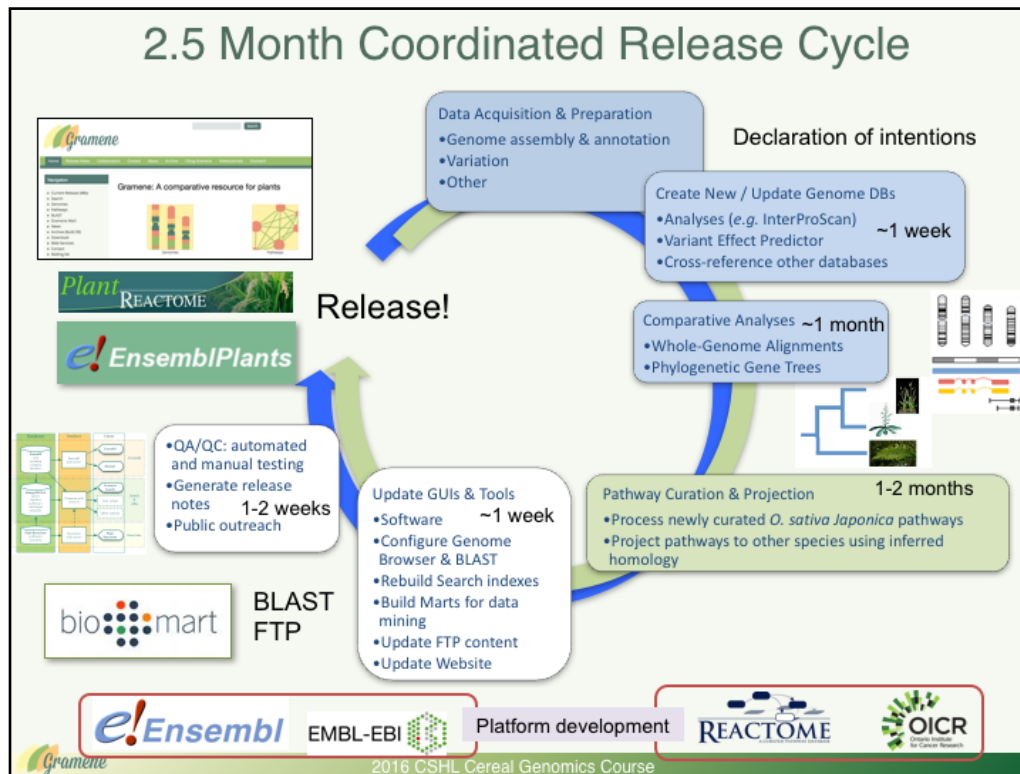
The Gramene project encompasses a set of best-in-class software tools that specialize in subsets of functionality.

Most significantly:

- Ensembl for Genomes, Gene Annotations and Comparative Genomics
- Plant Reactome for Pathways and Metabolic networks

And also:

- ATLAS for expression data in both, genomic and gene network context



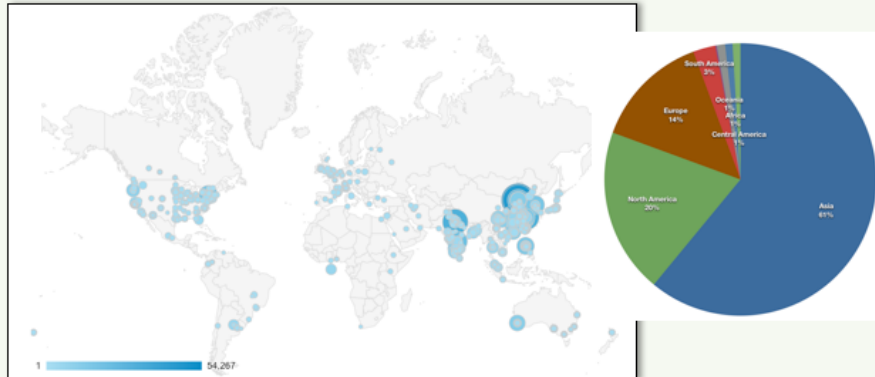
Data preparation is a continuous process. ~2 months before a release, genomes to be released or updated are frozen

We use EBI's Ensembl pipelines and schema to **build or update genomes** and the Reactome platform to **curate and portrait pathways**.

51 major data & software releases

www.gramene.org

* We are the major data providers for cereal crop plants in Asia



2007-2015



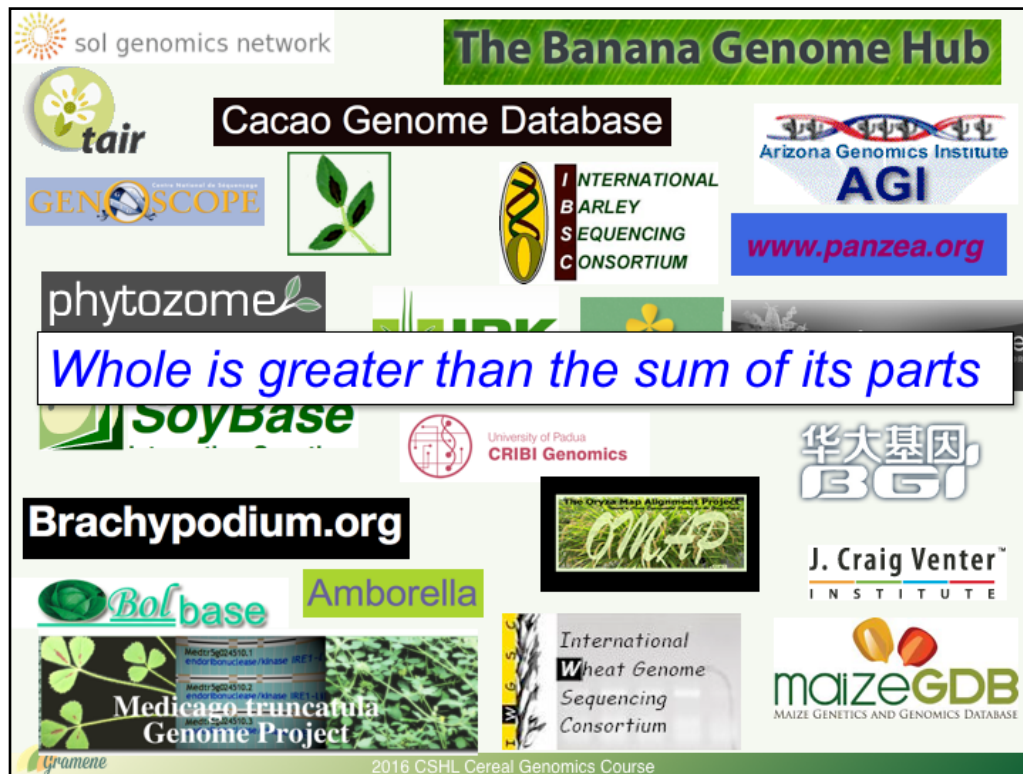
2016 CSHL Cereal Genomics Course

Other individuals/projects accessing Gramene using web services (API) not included in these counts

We get more hits from Asia because folks in Asia access Gramene primarily, because there is no good alternative. We are the major data providers for cereal crop plants.

Pie chart: Top Countries - Visits% January 2016 – June 2016 by Google Analytics

Mapmundi 2007-2015

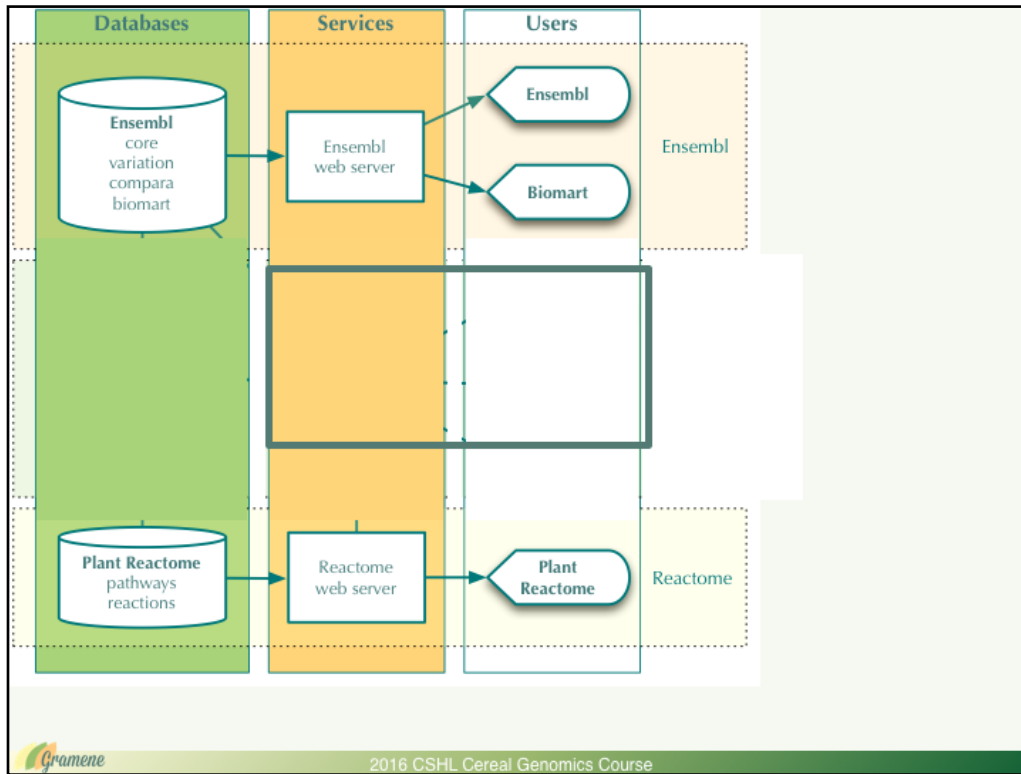


We owe this resource to the many publicly funded genome projects. Gramene has aggregated these data to develop a resource in which the whole is greater than the sum of its parts.

SEARCH



2016 CSHL Cereal Genomics Course



When you think about the architecture of these tools, you see siloes databases, services and user interfaces.

We have been working to build a data warehouse that combines these data and makes them searchable together.



Joe now at New York Genome Center & we're hiring!

Comparative genomics search interface built on top of data.gramene.org

We call it search but it does a lot more... **Filtering**

Goals

- to enable complex questions to be asked of gramene's COMPARATIVE dataset *simply*
- to encourage the asking of relatively broad questions and gaining meaningful insights for large result sets
- to incorporate all gramene's datasets with those of relevant third parties

Key features of this application:

- * Suggestions
- * Filters

* Visualization

<http://search.gramene.org>

Gramene
model plant

Gramene Port

Gram Find an

Plant Browse pathways

BLAS Query & sequence

Outreach Educate

Archives Legacy tools and data (markers, Cyt pathways, etc)

Search: 1761223 genes in 44 genomes

TOP

jasmonic acid biosynthesis 376 Jasmonic acid signaling 81 Wound-induced_Wun1 139 Prosystemin/Systemin 2 JAR1 76

PLANT REACTOME

jasmonic acid biosynthesis 376 Jasmonic acid signaling 81 Jasmonate conjugation 81

INTERPRO

MeTfase_7 1022 Allene_ox_cyc 81 Wound-induced_Wun1 139 Prosystemin/Systemin 2 JAR1 76 Ninja_fam 261

GENE

jasmonate-zim-domain protein 5 1 jasmonate-zim-domain protein 1 1 jasmonic acid carboxyl methyltransferase 1

jasmonate-zim-domain protein 8 1 JASMONIC ACID RESPONSIVE 3 1 JASMONATE-ZIM-DOMAIN PROTEIN 9 1

jasmonate-zim-domain protein 6 1 JASMONATE-ZIM-DOMAIN PROTEIN 2 1 jasmonate-zim-domain protein 7 1

JASMONATE RESISTANT 1 1 All genes that contain the word "jas" All genes that contain a word that starts with "jas"

GO PROCESS

response to jasmonic acid 4285 jasmonic acid and ethylene-dependent systemic resistance 428

cellular response to jasmonic acid stimulus 1707 jasmonic acid mediated signaling pathway 1702

response to jasmonic acid stimulus involved in jasmonic acid and ethylene-dependent systemic resistance 12

regulation of jasmonic acid metabolic process 65 response to cyclopentenone 203 jasmonic acid biosynthetic process 805

jasmonic acid and ethylene-dependent systemic resistance, ethylene mediated signaling pathway 22

regulation of jasmonic acid mediated signaling pathway 136

GO FUNCTION

jasmonyl-lc conjugate hydrolase activity 4 jasmonate O-methyltransferase activity 25 jasmonate-amino synthetase activity 22

jasmonyl-isoleucine-12-hydroxylase activity 4 methyl jasmonate esterase activity 17

2016 CSHL Cereal Genomics Course

Gramene General Search – user redirected to => search.gramene.org

Focused search engines:

- 1) Genomes Browser
- 2) Plant Reactome

at search.gramene.org

Start typing jasmonic acid signaling

Suggested filters appear grouped by category, with top scoring matches displayed first



Note the selected suggestion becomes displayed below the search box, numbers in the search status area are updated

Taxagenomic distribution shows species tree (based on ncbi taxonomy)

Count of genes in each species

Linear view of each genome – color by chromosome. Unanchored portions of a genome are gray and don't show the position of search results

Below is a list view – arabidopsis results appear first because they tend to be best annotated

Non-arabidopsis genes have a model species homolog to provide some hints of annotation to a gene

Gramene Search for genes, species, pathways, ontology terms, domains... 81 genes in 33 genomes

Plant Reactome | Jasmonic acid signaling

Brassica napus
B. oleracea
B. rapa
Arabidopsis thaliana
A. lyrata
Theobroma cacao
Medicago truncatula
Trifolium pratense
Glycine max
Populus trichocarpa
Pinus pinaster
Vitis vinifera
Solanum tuberosum
Solanum lycopersicum
B. vulgaris
Antirrhinum majus
S. moellendorffii
P. sativum
C. reinhardtii
O. lucimarinus
G. sulphurea
C. merriamii
Chondrus crispus

JAR1 AT2G46370 Arabidopsis thaliana
Auxin-responsive GHD family protein

Location Homology X-refs

Genome location: Chromosome 2:19033741-19036659
Currently viewing: 2:19033450-19036950

19,030,000 19,034,000 19,038,000 19,042,000 19,046,000 19,050,000

Search Gramene
All on Chromosome 2
All within 2:19033450-19036950

Links to other resources
Gramene Ensembl
PhytoMine
Arasport

JAR1 Curated Description
Encodes a jasmonate-amido synthetase that is a member of the GHD family of proteins. JAR1 catalyzes the formation of a biologically active jasmonyl-lysine...

Gramene 2016 CSHL Cereal Genomics Course

The location view shows a lightweight genome browser that lets you see the annotated gene structure.

You can scroll and zoom in/out

ensembl.gramene.org

Gene-based displays

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

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

Gene: GH3.5 O50G0586200

Description: Jasmonic acid-amido synthetase JAR1 [Source: UniProt/SwissProt; Acc: Q9581]

Location: Chromosome 5: 29,200,200-29,205,715 forward strand

About this gene: This gene has 1 transcript (splice variant), 84 orthologues and 12 paralogues

Transcripts: [Show transcript table](#)

Summary

Name: GH3.5 (UniProtKB Gene Name)

UniProtKB: This gene has proteins that correspond to the following UniProtKB identifiers: [Q9581](#)

Gene type: Protein coding

Annotation method: Gene annotation by International Rice Genome Sequencing Project (IRGSP) dated 2013-04-24. [Read more...](#)

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

MSU gene

RNA Gene (AGE)

BLASP gene

Genes

BLASP gene

MSU gene

Gene Legend

- Protein Coding
 - MSU
- Non-Protein Coding
 - RNA gene

The ensembl gene page (for a different gene in the jasmonic signaling pathway)

ensembl.gramene.org

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData Search Oryza sativa Japonica...

Oryza sativa Japonica (IROSP-1.0) Location: 5:29,200,200-29,205,715 Gene: GH3.5 Transcript: GH3.5

Gene-based displays

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

Gene: GH3.5 OS05G0586200

Description: Jasmonic acid-amido synthetase JAR1 [Source:UniProt/SWISSPROT;Acc:Q6581]

Location: Chromosome 5: 29,200,200-29,205,715 forward strand.

About this gene: This gene has 1 transcript (splice variant), 84 orthologues and 12 paralogues.

Transcripts: [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags
GH3.5	OS05G0586200-01	2407	581aa	Protein coding	Q6581	NM_001062995	NP_001056460

Gene expression

Showing 4 experiments: Sort by: Default Download all results

See more expression data at Expression Atlas. This expression view is provided by Expression Atlas. Please direct any queries or feedback to expression-atlas@ebi.ac.uk

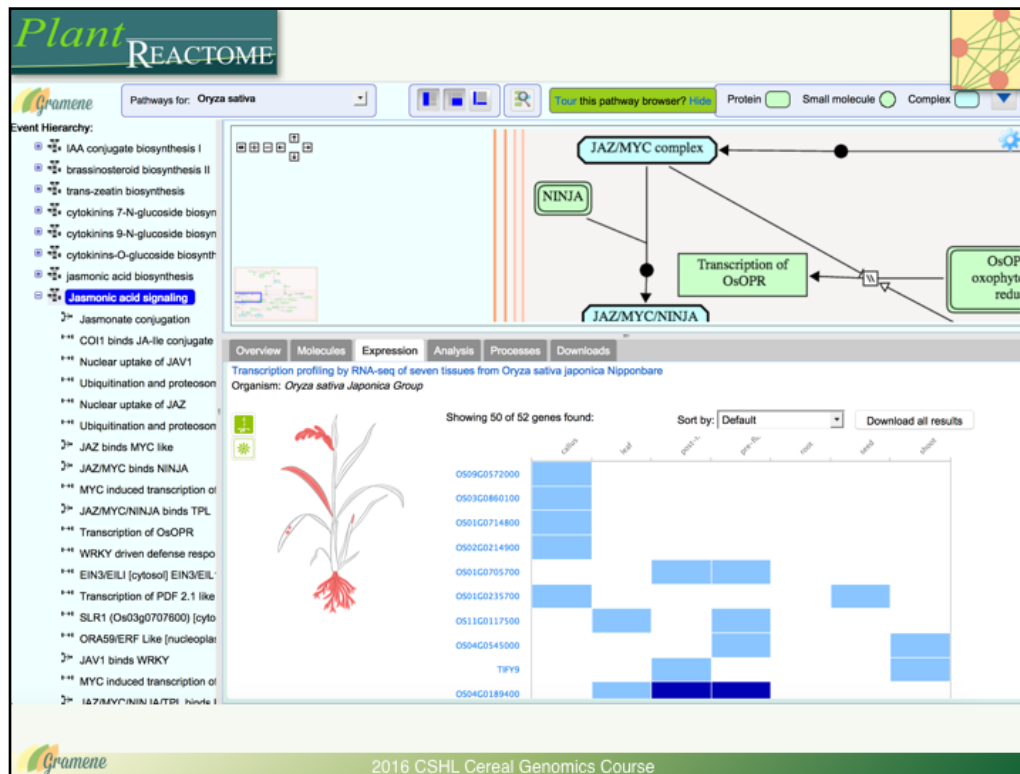
Gramene release 51 - Aug 2016 © EMBL-EBI

About Us Get help Our sister sites Follow us

2016 CSHL Cereal Genomics Course

Click Gene expression in the sidebar

This expression atlas view widget is something we'd like to incorporate into the gramene search results



Same data at the pathway level available from Plant Reactome

Gramene Search for genes, species, pathways, ontology terms, domains... 9 genes in 2 genomes

Gene | RFL1 X

RFL1 AT1G12210 Arabidopsis thaliana
RPS5-like 1

Location Homology X-refs

TAIR Curated Description
RFL1 has high sequence similarity to the adjacent disease resistance (R) gene RPS5.

Compare Gene Tree
This phylogram shows the relationships between this gene and others similar to it, as determined by Ensembl Compara.

Gene Tree

Phylogram showing relationships between RFL1 and other genes. The tree is rooted and shows clades with gene counts and paralog numbers. A tooltip for IPR032675 - L_dom-like shows it is shared by 963 of 1114 (86.4%) genes in this gene tree and by all 7 genes in this clade.

Search Gramene
Show All Homologs 1114
Show Orthologs 13
Show Paralogs 25

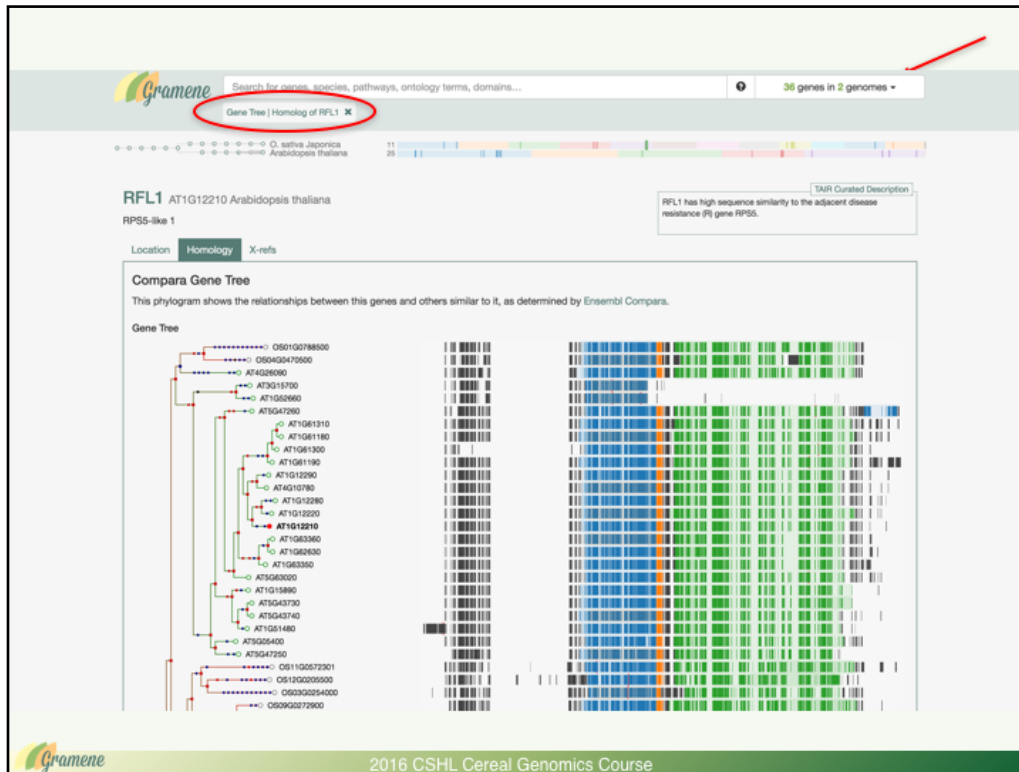
Click on "Show [Homolog type]"

Links to other resources
Ensembl Gene Tree view

Gramene 2016 CSHL Cereal Genomics Course

Popover on an InterPro domain shows how many genes in the gene tree are annotated with this domain, and for collapsed clades it shows the number of genes in the clade with the domain

Click on Show All Homologs to update the search filter.



If you change the genomes you are searching to just rice and Arabidopsis, the search results are filtered down to the 36 rice and Arabidopsis genes in the gene family.

This special filter also prunes the species tree and gene tree views to hide branches that don't lead to Arabidopsis or rice.

It is now tractable to visualize the evolutionary relationships between a pair of species.

We would like your feedback!

RFL1 AT1G12210 Arabidopsis thaliana
RPS5-like 1


Location Homology X-ref

Compare to Arabidopsis
This phylogram Show paralogs

Gene Tree

Search beta testers wanted!

Sign up at **gramene.org/beta**



Gramene

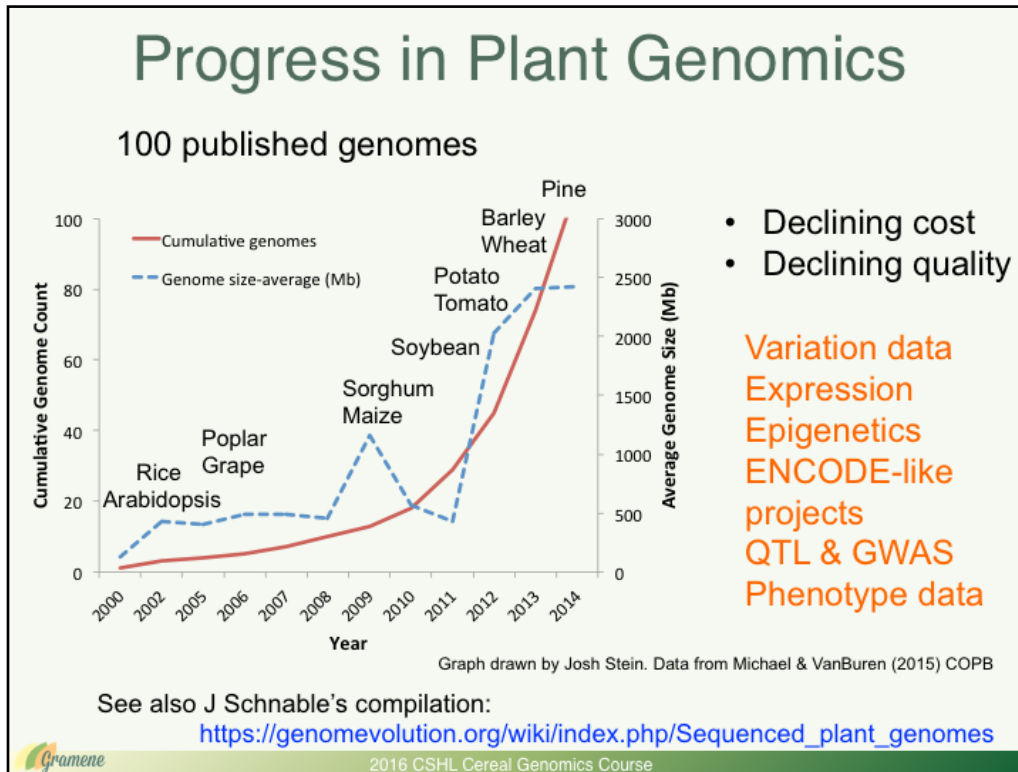
2016 CSHL Cereal Genomics Course

Is this useful information? What other type of info would you like to see in an inset?

GENOMES

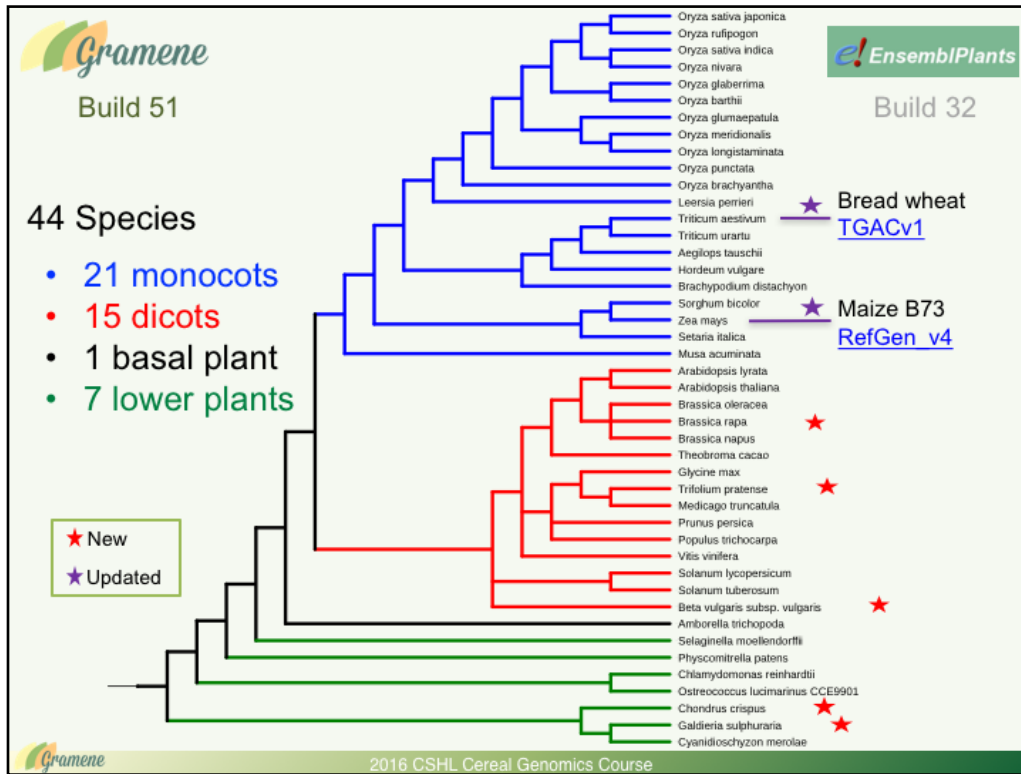



2016 CSHL Cereal Genomics Course



Progress in Plant Genomics
 Gramene to the rescue

- Over 100 published genomes
 [https://genomevolution.org/wiki/index.php/Sequenced_plant_genomes by J Schnable]
- Subset in COPB table by Michael & VanBuren (2015):
 - 63% crops
 - 76% dicots, 19% monocots & 5% lower plants




<http://www.gramene.org>

[Home](#)
[Release Notes](#)
[Collaborators](#)
[Contact](#)
[About](#)
[Archive](#)
[Citing Gramene](#)
[Videotutorials](#)
[Outreach](#)

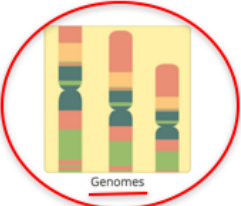
Navigation

- Current Release (51)
- Search
- **Genomes**
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Mailing list
- Related genome browsers
- Tools

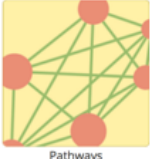
Recent blog posts

- Branching out at CSHL: adding plant genomics to statistics and math as an URP*
- Highlights of the 2016 SolGenomics Network meeting
- Highlights of the 2016 GODAN Summit
- Updates to Gramene's Search

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.

Gramene Portals

- **Genome Browser:** Browse gene annotations & diversity data
- **Plant Reactome:** Browse metabolic & regulatory pathways
- **Bulk downloads**
- **Gramene Mart:** Customized data queries
- **BLAST:** Align DNA & protein sequences

How to access Gramene's Genome Browser

Entry points

- 1) General Search =>
- 2) Genomes

http://plants.ensembl.org

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Search all species...

Search: All species for e.g. Carboxy* or chx28

Popular genomes

- Arabidopsis thaliana** TAIR10
- Oryza sativa Japonica** IRGSP-1.0
- Triticum aestivum** TGACv1
- Hordeum vulgare** ASM3200v1
- Zea mays** AGPv4
- Physcomitrella patens** ASM242v1

★ Log in to customize this list

All genomes

-- Select a species --

View full list of all Ensembl Plants species

What's New in Release 51

- New genomes
 - Beta vulgaris* (sugar beet)
 - Brassica napus* (oilseed rape)
 - Trifolium pratense* (red clover)
- Transfer of two Rhodophyta genomes from Ensembl Protists, *Galdieria sulphuraria* and *Chondrus crispus*
- Updated software

Did you know...?

You can search the Track Hub Registry to find more than 900 public RNA-Seq studies aligned to plants (read more on the Ensembl Genomes site).

Track Hub Registry

New Maize B73 Genome Assembly and Gene Annotations

An entirely new assembly of the maize genome (B73 RefGen_v4) was constructed from PacBio Single Molecule Real-Time (SMRT) sequencing at approximately 60-fold coverage and scaffolded with the aid of a high-resolution whole-genome restriction (optical) mapping. The pseudomolecules of maize B73 RefGen_v4 were assembled nearly end-to-end, representing a 52-fold improvement in average contig size relative to the previous reference (B73 RefGen_v3).

Genes were annotated with the Maker pipeline (Campbell et al. 2014) using 111,000 transcripts obtained by single-molecule sequencing. These long-read Iso-Seq data (Wang et al. 2016) improved annotation of alternative splicing, more than doubling the number of alternative transcripts from 1.5 to 3.8 per gene, thereby improving our knowledge of gene structure and transcript variation, resulting in substantial improvements including resolved gaps and misassemblies, corrections to strand, consolidation of gene models, and anchoring of unanchored genes.

Gene annotation was performed in the laboratory of Doreen Ware (CSHL/USDA). Protein-coding genes were identified using MAKER-P software version 3.1 (Campbell et al. 2014) with the following transcript evidence: 111,151 PacBio Iso-Seq long-reads from 6 tissues (Wang et al. 2016), 69,163 full-length cDNAs deposited in Genbank (Alexandrov et al. 2008; Soderlund et al. 2009), 1,574,442 Trinity-assembled transcripts from 94 B73 RNA-Seq experiments (Law et al. 2015), and 112,963 transcripts assembled from deep sequencing of a B73 seedling (Marlin et al. 2014). Additional evidence included annotated proteins from *Sorghum bicolor*, *Oryza sativa*, *Setaria italica*, *Brachypodium distachyon*, and *Arabidopsis thaliana* downloaded from Ensembl Plants Release 29 (Oct-2015). Gene calling was assisted by Augustus (Keller et al. 2011) and FGENESH (Salamov & Solovyev, 2000) trained on maize and monocots, respectively. Low-confidence gene calls were filtered on the basis of an Annotation Edit Distance (AED) score and other criteria and are viewable as a separate track. In the end, the higher confidence set (called filtered gene set) has 39,324 protein coding genes. Gene annotations from B73 RefGen_v3 were mapped to the new assembly and are also available as a separate track. In addition, 2,532 long non-coding RNA (lncRNA) genes were mapped and annotated from prior studies (Li et al. 2014; Wang et al. 2016), while 2,290 tRNA genes were identified using tRNAscan-SE (Lowe & Eddy, 1997), and 154 miRNA genes mapped from miRBase (Kozomara & Griffiths-Jones, 2014).

New Bread Wheat Genome Assembly

A new genome assembly of *Triticum aestivum* cv. Chinese Spring is now available in Gramene and Ensembl Plants. The assembly (TGACv1) and its accompanying annotation was produced by the Earlham Institute, formerly The Centre for Genome Analysis (TGAC), as part of the [Triticaceae Genomics for Sustainable Agriculture](#) project.

The assembly has a scaffold N50 of 88 Kbp and a total length of 13.4 Gbp in contigs greater than 500 bp (read more). The gene model annotation consists of 217,907 loci and 273,739 transcripts. A total of 104,086 protein coding genes, 1154,798 transcripts and 10,156 lncRNAs have been

















Gramene 2016 CSHL Cereal Genomics Course

Gramene & Ensembl Plants shared interface

http://ensembl.gramene.org

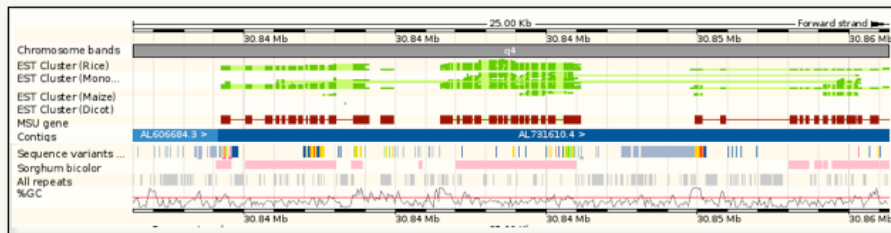
http://plants.ensembl.org

42

Show All entries										
Show/Hide columns										
Filter										
Name	Classification	Taxon ID	Assembly	Accession	Variation database	Regulation database	Whole genome alignments	Other alignments	In peptide comparison	In pan-taxonomic comparison
 Aegilops tauschii	Liliopsida	37882	ASM34733v1	GCA_000347335.1	-	-	✓	✓	✓	-
 Amborella trichopoda	Amborellales	13333	AMTR1.0	GCA_000471905.1	-	-	✓	✓	✓	✓
 Arabidopsis lyrata	eudicotyledons	81972	v1.0	GCA_000004295.1	-	-	✓	✓	✓	-
 Arabidopsis thaliana	eudicotyledons	3702	TAIR10	GCA_000001735.1	✓	✓	✓	✓	✓	✓
 Beta vulgaris subsp. vulgaris	eudicotyledons	3555	RefBeet-1.2.2	GCA_000511025.2	-	-	-	-	✓	-
 Brachypodium distachyon	Liliopsida	15368	v1.0	GCA_000005505.1	✓	-	✓	✓	✓	-
 Brassica napus	eudicotyledons	3708	AST_PRUEB5043_v1	GCA_000751015.1	-	-	-	-	✓	-
 Brassica oleracea	eudicotyledons	109379	v2.1	GCA_000695525.1	-	-	✓	✓	✓	-
 Brassica rapa	eudicotyledons	51351	IVFCAASv1	GCA_000309985.1	-	-	✓	✓	✓	-
 Chlamydomonas reinhardtii	Chlorophyta	3055	v3.1	GCA_000002595.2	-	-	✓	✓	✓	✓
 Chondrus crispus	Rhodophyta	2769	ASM35022v2	GCA_000350225.2	-	-	-	-	✓	✓
 Cyanidioschyzon merolae	Rhodophyta	280699	ASM9120v1	GCA_000091205.1	-	-	✓	✓	✓	✓
 Galdieria sulphuraria	Rhodophyta	130081	ASM34128v1	GCA_000341285.1	-	-	-	-	✓	-
 Glycine max	eudicotyledons	3847	V1.0	GCA_000004515.1	-	-	✓	✓	✓	-
 Hordeum vulgare	Liliopsida	112509								
 Leersia perrieri	Liliopsida	77586	Lperri_V1.4	GCA_000325785.3	-	-	✓	✓	✓	-

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

Genome Browsers



- Assembly structure and sequence
- Genes
- Expression
- Comparative alignments
- Genetic markers (SNPs, Indels, QTLs, SSRs)
- Variation
- Ontologies
- Repeats & transposable elements
- Regulatory & Epigenetic marks

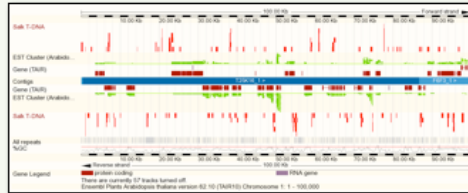
Baseline Annotation

- Transposon discovery
- Gene prediction
- EST Alignment

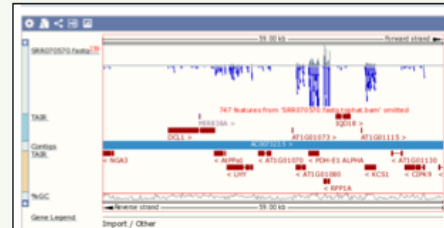
Add Custom Tracks



- Methylome (Ecker)
- Uploaded from an URL
- BED file format



- Salk T-DNA lines
- Uploaded from my laptop
- GFF file format



- RNA-seq
- Uploaded from iPlant URL
- BAM alignment file

GFF
BED
BAM
VCF
bedGraph
Gbrowse
PSL
WIG
GTF
BigBed
BigWig
TrackHub



2016 CSHL Cereal Genomics Course

Navigating Browsers

Gramene BLAST BioMart Tools Downloads Help Feedback

Search *Oryza sativa Japonica*...

Oryza sativa Japonica Location: 1,8,001-18,000 Gene: OS01G0100400 Transcript: P0672D08.3 Variation: ENSVOSA000000001

Variation displays

- Explore this variation
- Genomic context
- Genes and regulation (7)**
 - Flanking sequence
 - Genotype frequency
 - Individual genotypes (11)
 - Linkage disequilibrium
 - Phenotype Data
 - Phylogenetic Context
 - Citations
 - External Data
- Configure this page
- Manage your data
- Export data
- Bookmark this page
- Share this page
- Download view as CSV

Gramene is produced in collaboration with Ensembl Plants

ENSVOSA000000001 SNP

Original source A collection of SNPs produced by the BGI based on comparison of the Indica and Japonica genomes, Yu et al. (2005) (release 129) | [View in Indica vs. Japonica](#)

Alleles T/N

Location Chromosome 1:14147 (forward strand) | [View in location tab](#)

Synonyms OryzaSNP TBGI000006

HGVs name None

Genes and regulation

Gene and Transcript consequences

Gene	Transcript (strand)	Allele (transcript allele)	Consequence Type	Position in transcript	Position in CDS	Position in protein	Amino acid	Codons	Detail
OS01G0100400	OS01T0100400-01 (+) biotype: protein_coding	N (N)	Coding sequence variant	1335	1282	428	-	TTA/TAA	Show
OS01G0100300	OS01T0100300-00 (-) biotype: protein_coding	N (N)	Upstream gene variant	-	-	-	-	-	Show
OS01G0100466	OS01T0100466-00 (-) biotype: protein_coding	N (N)	Upstream gene variant	-	-	-	-	-	Show
OS01G0100500	OS01T0100500-01 (+) biotype: protein_coding	N (N)	Upstream gene variant	-	-	-	-	-	Show
EPIOSAG0000000746 3	EPIOSAT00000008851 (-) biotype: ncRNA	N (N)	Downstream gene variant	-	-	-	-	-	Show

Gramene 2016 CSHL Cereal Genomics Course

Configuring Browser Views

The screenshot displays the Gramene genome browser interface. On the left sidebar, a red circle highlights the 'Configure this page' option. A red arrow points from this option to a gear icon in the top track's header. Another red arrow points from a text box 'Click on "Configure this image" symbol' to the same gear icon. A third red arrow points from a text box 'Mouse over track name & configure symbol' to a gear icon in the 'B73_CpG_ratio' track header. A yellow starburst in the bottom left corner contains the text 'YouTube video tutorials'. The main view shows several tracks including 'Chromosome bands', 'Contigs', 'Gramene gene', 'Gene Legend', 'Location: 2:45809642-45885359', 'Gene:', 'B73_CpG_ratio', 'B73_CpG_coverage', 'cDNA', 'EST Cluster (Maize)', 'MAKER-P genes', 'Gramene gene', 'Contigs', 'maRNA alignments', 'Gramene gene', 'MAKER-P genes', 'EST Cluster (Maize)', 'cDNA', 'HapMap2', 'Panzea_2.7GBS', and '%GC'. A 'Change track style' dialog box is open over the 'B73_CpG_ratio' track, showing options for 'Wiggle plot', 'Gradient', and 'P-value'. The bottom of the slide features the Gramene logo and the text '2016 CSHL Cereal Genomics Course'.

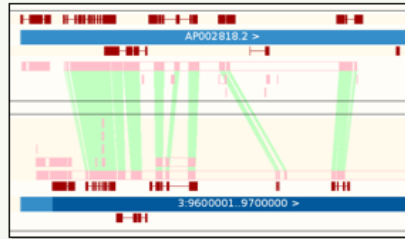
Collaboration with Ensembl Genomes (Paul Kersey)

Comparative Phylogenomics

Reconstructing evolutionary histories

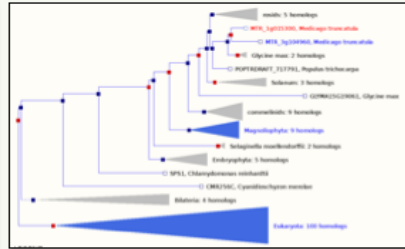
Whole Genome DNA Alignment

- Pairwise alignments
- B/LASTZ-CHAIN-NET method



Phylogenetic Gene Trees

- Infers orthologs and paralogs
- Taxonomic dating
- dN, dS, dN/dS
- Oryza-centered trees



Vilella A.J., et al. (2008). *Genome Res.* doi:10.1101/gr.073585.107
Schwartz S et al., *Genome Res.*;13(1):103-7, Kent WJ et al., *Proc Natl Acad Sci USA.*, 2003;100(20):11484-9

http://useast.ensembl.org/info/docs/compara/homology_method.html



2016 CSHL Cereal Genomics Course

Exporting Images & Data

The screenshot displays the Ensembl genome browser interface. On the left, the 'Export as:' menu is open, showing various file formats. The 'PNG (x0.5)' option is highlighted with a red circle. Below the menu, the genomic track view shows a region of the genome with various features, including exons and introns, color-coded by strand and type. The bottom panel shows a detailed view of the genomic region, including the 'Region:' field and a list of genomic features with their coordinates and descriptions.

Export as:

- PDF
- SVG
- PostScript
- PNG (x10)
- PNG (x5)
- PNG (x2)
- PNG
- PNG (x0.5)**
- Text (GFF)

Region: chromosome:JGIv2.0:5:5868186:5916853:1

Genomic Features:

Feature	Coordinates	Description
scaffold_5:5800001..5900000	Ensembl Contig 1	31815
scaffold_5:5900001..6000000	Ensembl Contig 31816	48668
5	Ensembl Exon	5868120 5868305
exon_id=Si000090m.exon3	gene_type=NOVEL_protein_coding	1
5	Ensembl Exon	5868650 5868757
exon_id=Si000090m.exon4	gene_type=NOVEL_protein_coding	1
5	Ensembl Exon	5869384 5869584
exon_id=Si000090m.exon5	gene_type=NOVEL_protein_coding	2
5	Ensembl Exon	5869585 5871677
exon_id=Si000090m.exon6	gene_type=NOVEL_protein_coding	2
5	Ensembl Exon	5871920 5872919
exon_id=Si000090m.exon7	gene_type=NOVEL_protein_coding	1
5	Ensembl Exon	5873224 5873292
exon_id=Si003928m.exon1	gene_type=NOVEL_protein_coding	2
5	Ensembl Exon	5873415 5873582
exon_id=Si003928m.exon2	gene_type=NOVEL_protein_coding	1

Exporting Images & Data

Newick tree format

Genomic alignments (image)

Gene tree (image)

Gene tree (text)

Gene tree (alignment)

Gene gain/loss tree

Orthologues (31)

Paralogues

Phylogenetic Comparison

Gene Tree (image)

Gene Tree (text)

Orthologues (3)

Phenotype

Genetic Variation

Variation table

Structural variation

Variation image

External data

ID History

Gene history

Configure this page

Add your data

Export data

Bookmark this page

Share this page

Continue is produced in

```

((((((((((Taeas_3AS_C11206884.1_Taes_0.0139,
Taeas_3B_CCFE49AFC.1_Taes_0.0355):0.0062,
Taeas_3DS_97A72ABED.1_Taes_0.0114):0,
(Taeas_3DS_B754ABESC.1_Taes_0,
EMT27641_Atau_0):0.0149):0.0097,
MLOC_51485.1_Hvul_0.0285):0.0351,
Taeas_3DS_2FAS164A9)
BRAD12G03487.1_Bdia
(((OS01T0103900-01_
BG10SGA002592-PA_01
OBGLA01G0002400_1_0
OB01G10240.1_Obra_1
(((GRMZM2G049952_P
GRMZM2G0352042_P01_2
GRMZM2G048294_P01_2
SD039008910.1_Sb1c
SI003928n_Sita_0.0
GEMIA_Achr8P16540_0
(((GLYMA08G00890.
AET04915_Mtru_0.11
POPR_0015a12120_1
(((AI_scaffold_000
AI_scaffold_0468_1
AT5G07380.3_Atha_0
Bra009267.1-P_Brap
(Igeneshl_pg.C_scaf
AET04914_Mtru_11.37
VIT_11a0016g00050.t
(PGSC0003DMT400368
Solyce03g117990.1.1
(EFJ22762_Smo_0.0
EFJ30879_Smo_0.01
Gramene release 40 - Jan 2015

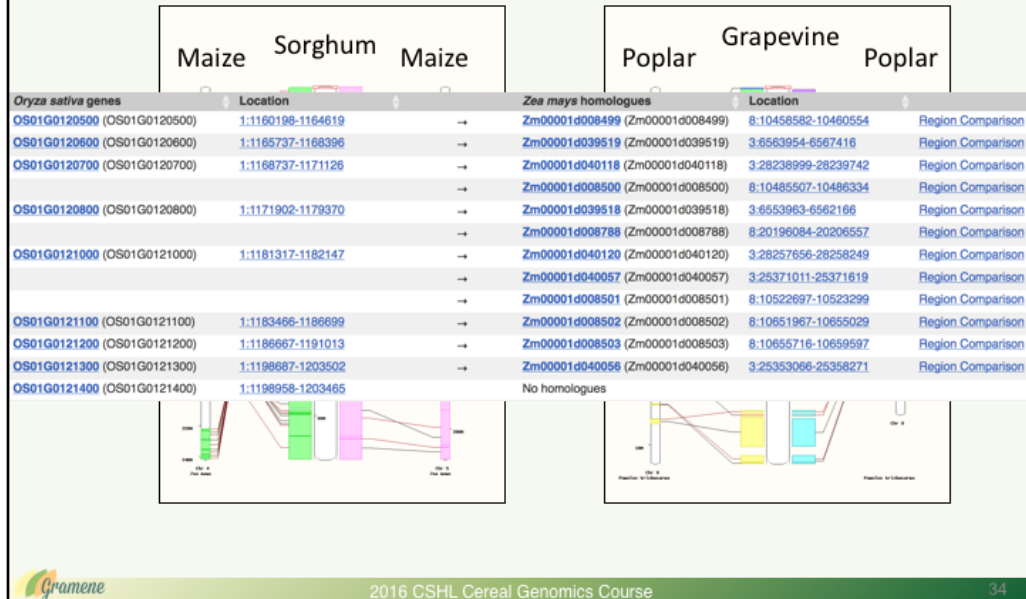
```

Multiple & pairwise
alignments



2016 CSHL Cereal Genomics Course

Browse Homoeologous Regions From Ancient Polyploidy



Synteny is based on Compara gene trees

http://ensembl.gramene.org/Vitis_vinifera/Location/Synteny?otherspecies=Populus_trichocarpa&r=13%3A1-1000

http://ensembl.gramene.org/Sorghum_bicolor/Location/Synteny?otherspecies=Zea_mays&r=4%3A1-1000

http://ensembl.gramene.org/Oryza_sativa/Location/Synteny?db=core&r=1%3A1160000-1231000&g=OS01G0120500&t=OS01T0120500-01&otherspecies=Zea_mays

Variation Types in Gramene

- SNPs & Structural Variants
- Natural & EMS-induced
- Interhomoeologous variants
(wheat A, B, D genomes, barley & brachy)

The image displays three screenshots of the Gramene Variation Explorer interface, illustrating different types of genetic variation:

- Structural variant: ENSSVATH00012771**: This screenshot shows details for a structural variant, including its location on chromosome 4L (1058414-1058845) and its overlap with the *GrS1* gene. It provides links to explore the variant's genomic context, gene and regulation, and supporting evidence.
- tmp_3_5589324_G_A**: This screenshot shows details for a natural SNP. It includes information about the original source (Ensembl), the variant's location on chromosome 3B (5589324), and its overlap with the *GrS1* gene. It also provides links to explore the variant's genomic context, gene and regulation, and supporting evidence.
- EPITAEV00010568**: This screenshot shows details for an EMS-induced SNP. It includes information about the original source (Ensembl), the variant's location on chromosome 4B (7593731), and its overlap with the *GrS1* gene. It also provides links to explore the variant's genomic context, gene and regulation, and supporting evidence.

Single-nucleotide level vs SVs = range from tens to millions of base pairs in size and include insertions, deletions, inversions, translocations and CNPs

Inter-homoeologous variants. Reported as seq diffs where 1-to-1 homoeology btw bread wheat component genomes. 1st computed using Compara pipeline 4 detecting protein orthology, then parsing supporting WGs to find nt seq diffs

EMS-induced: a [premature STOP codon](#) introduced in Sb03g028120 (ems =ethyl methanesulfonate)

http://ensembl.gramene.org/Triticum_aestivum/Variation/Explore?r=4B:75937031-75938031;v=EPITAEV00010568;vdb=variation;vf=6737904

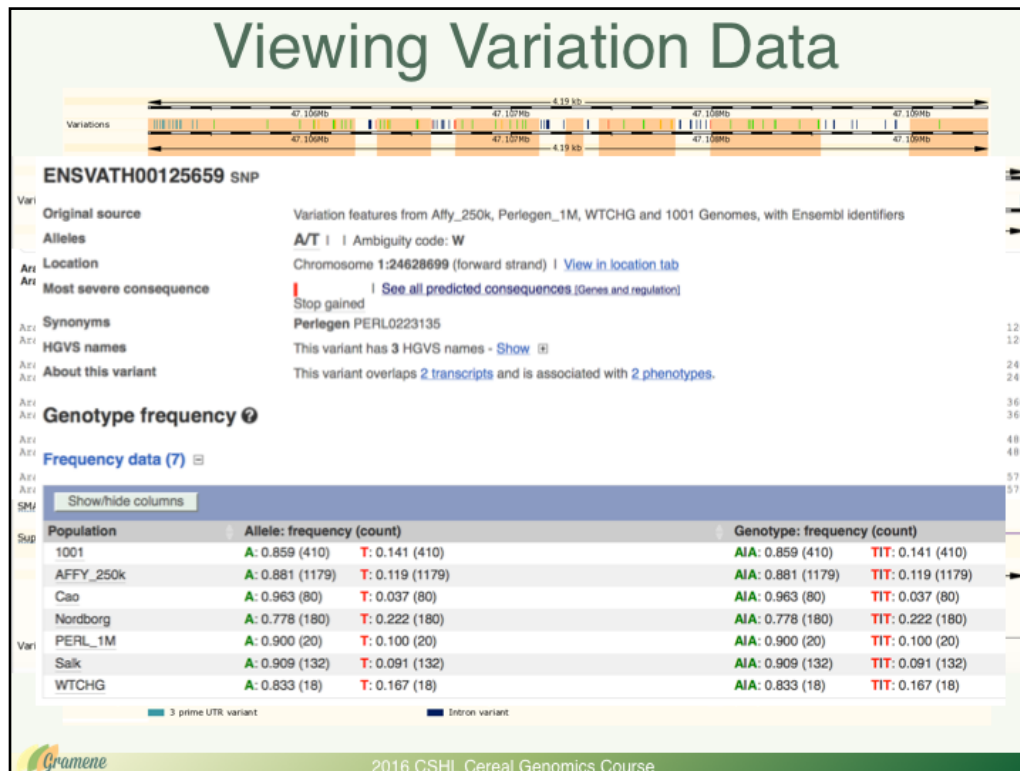
http://ensembl.gramene.org/Sorghum_bicolor/Variation/Explore?db=core;g=Sb03g028120;r=3:55892797-55900180;source=Sorghum_EMS_mutants;t=Sb03g028120.1;v=tmp_3_55893324_G_A;vdb=variation;vf=7315833

As part of the wheat genome analysis, we have aligned a set of [Triticum aestivum](#) (bread wheat) homoeologous SNPs (SNPs between the component A, B, and D genomes of wheat) against the *Brachypodium distachyon* genome. SNPs have been classified into two groups, 1) SNPs that differ between the A and D genomes (where the B genome is

unknown) and, 2) SNPs that are the same between the A and D genomes, but differ in B (Brencheley et al, 2012).

The wheat sequence alignments and the projected homoeologous SNPs are available as tracks under the "Wheat SNPs and alignments" section of the "Configure This page" menu

Seq alignments for barley & brachy



http://ensembl.gramene.org/Zea_mays/Transcript/Variation_Transcript/image?db=core;g=Zm00001d003533;r=2:47104443-47110654;t=Zm00001d003533_T001

To get the color coded sequence alignment (if available for this region against desired species – not available for example region between maize and sorghum):

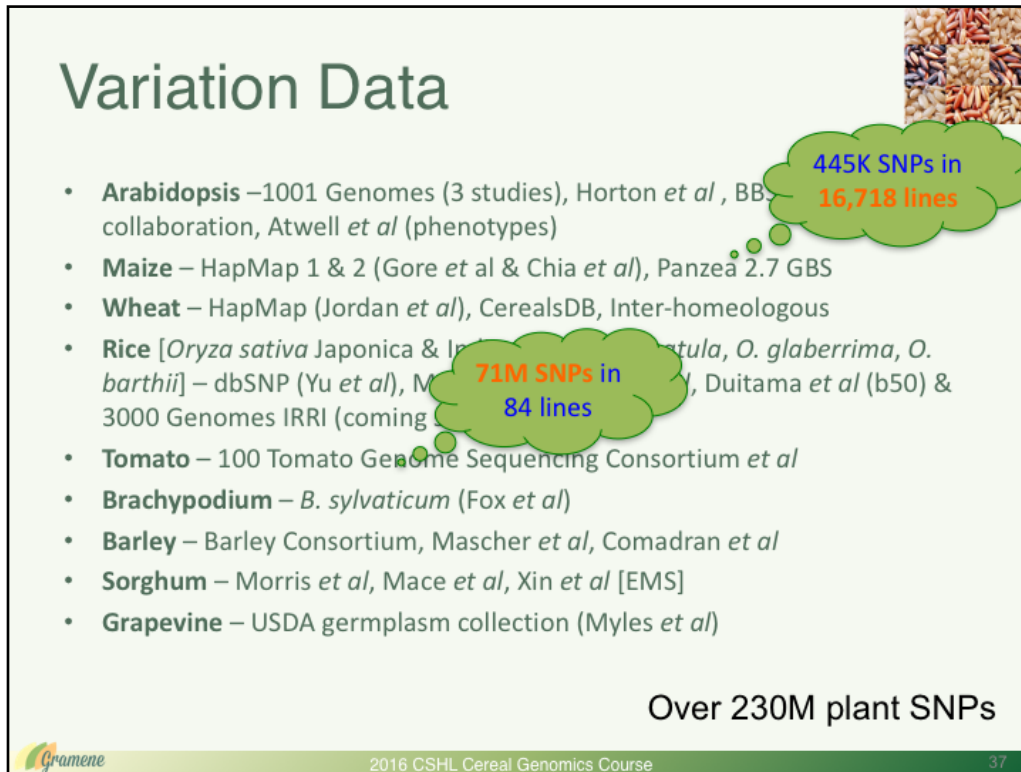
Go to Location tab -> Select Alignments (text) -> “Configure this page”
→ “Show variants” (to color code variants)

To get more data on a particular variant, go to the Transcript tab -> Select Variant Table -> Select SNP of interest (e.g., one with stop gained or affecting a splice junction)

http://ensembl.gramene.org/Zea_mays/Transcript/Variation_Transcript/Table?db=core;g=Zm00001d003533;r=2:47104443-47110654;t=Zm00001d003533_T001

http://ensembl.gramene.org/Zea_mays/Variation/Explore?db=core;g=Zm00001d003533;r=2:47104443-47110654;t=Zm00001d003533_T001;v=PZE0245647646;vdb=variation;vf=8694234

Variation Data



- **Arabidopsis** – 1001 Genomes (3 studies), Horton *et al*, BB collaboration, Atwell *et al* (phenotypes)
- **Maize** – HapMap 1 & 2 (Gore *et al* & Chia *et al*), Panzea 2.7 GBS
- **Wheat** – HapMap (Jordan *et al*), CerealsDB, Inter-homeologous
- **Rice** [*Oryza sativa* Japonica & Indica, *O. sativa*, *O. glaberrima*, *O. barthii*] – dbSNP (Yu *et al*), M (Duitama *et al* (b50) & 3000 Genomes IRRI (coming))
- **Tomato** – 100 Tomato Genome Sequencing Consortium *et al*
- **Brachypodium** – *B. sylvaticum* (Fox *et al*)
- **Barley** – Barley Consortium, Mascher *et al*, Comadran *et al*
- **Sorghum** – Morris *et al*, Mace *et al*, Xin *et al* [EMS]
- **Grapevine** – USDA germplasm collection (Myles *et al*)

Over 230M plant SNPs

Gramene 2016 CSHL Cereal Genomics Course 37

A. Thaliana ~1,600 strains. Phenotype data added from a GWAS study of 107 phenotypes in 95 inbred lines by Atwell *et al*

Maize HapMap2 – 55M SNPs & x 103 pre-domesticated and domesticated varieties

Panzea 2.7 GBS - 719,472 SNPs x 16,718 maize and teosinte lines, grouped in 14 overlapping populations

27,869,011 OS japonica SNPs in EG31

Duitama set consisted of **26,012,124** SNPs in 104 lines (exclude chrM, C, Sy & Un)

USDA grape collection: 17 strains

Wheat homoeologous SNPs: >10M variant features (insertions, deletions and substitutions) identified. 1-to-1 homoeology relationship between genes on the different bread wheat component genomes

Computed using 1) Compara pipeline for detecting protein orthology,

and 2) supporting WGAs to find sequence diffs at the nucleotide level
<http://www.vegkitchen.com/tips/cooking-with-whole-grains/attachment/ge-rice-threatens-biodiversity/>

Ensembl Diversity Databases



- Loaded using standard VCF & metadata files
 - Each SNP/Variant allele assigned stable id
- Obtain individual genotypes & population frequencies
- Variant Effect Prediction (VEP)
 - Classifies SNPs based on predicted effect on transcript using Sequence Ontology (SO) terms => Controlled vocabulary & standards support complex data mining

Example VEP SO classifications

intron_variant
missense_variant
synonymous_variant
stop_gained
stop_lost
splice_donor_variant
splice_acceptor_variant
initiator_codon_variant
stop_retained_variant

BioMart

Complex custom queries to mine data:

- InterPro
- GO
- VEP
- Orthologs/Paralogs



Use case: Find transcription factors having "stop_gained" alleles

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

New Count Results

Dataset 28964 / 71156260 SNPs

Solanum lycopersicum variations (SL2.50 (2014-10-EnsemblPlants))

Filters

Consequence type : stop_gained

Attributes

Gene stable ID

Consequence to transcript

Variation ID

Dataset 722 / 38735 Genes

Solanum lycopersicum genes (SL2.50 (2014-10-EnsemblPlants))

Filters

Gene type : protein_coding

InterPro ID(s): (ID-list specified)

Attributes

InterPro ID

InterPro short description

Export all results to

File

TSV

Unique results only

Go

Email notification to

View

100 rows as


HTML

Unique results only












Gene stable ID	Consequence to transcript	Variation ID	InterPro ID	InterPro short description
Solyc02g031820.1	stop_gained	vcZ12SDN8	IPR009057	Homeodomain-like
Solyc02g031820.1	stop_gained	vcZ12SDN8	IPR009057	Homeodomain-like
Solyc10g055550.1	stop_gained	vcZ1VTZ03	IPR004827	bZIP
Solyc10g055550.1	stop_gained	vcZ1VTZ03	IPR004827	bZIP
Solyc10g081350.1	stop_gained	vcZ1WCB06	IPR004827	bZIP
Solyc02g065490.2	stop_gained	vcZ13HYWN	IPR009057	Homeodomain-like
Solyc02g065490.2	stop_gained	vcZ13HYWN	IPR009057	Homeodomain-like
Solyc02g060720.1	stop_gained	vcZ140F09	IPR009057	Homeodomain-like
Solyc01g067090.2	stop_gained	vcZ10GMA2	IPR009057	Homeodomain-like
Solyc02g064630.2	stop_gained	vcZ13FY05	IPR009057	Homeodomain-like
Solyc01g103830.1	stop_gained	vcZ11HQF0	IPR009057	Homeodomain-like
Solyc02g020470.2	stop_gained	vcZ12NHD9	IPR009057	Homeodomain-like
Solyc10g083380.1	stop_gained	vcZ1WE600	IPR004827	bZIP
Solyc10g083380.1	stop_gained	vcZ1WE600	IPR004827	bZIP
Solyc11g044560.1	stop_gained	vcZ1YGAVD	IPR004827	bZIP
Solyc11g044560.1	stop_gained	vcZ1YGAVR	IPR004827	bZIP
Solyc11g044560.1	stop_gained	vcZ1YGAW2	IPR004827	bZIP
Solyc11g044560.1	stop_gained	vcZ1YGAWA	IPR004827	bZIP
Solyc11g044560.1	stop_gained	vcZ1YGAWJ	IPR004827	bZIP
Solyc11g044560.1	stop_gained	vcZ1YGAW5	IPR004827	bZIP
Solyc11g044560.1	stop_gained	vcZ1YGB23	IPR004827	bZIP
Solyc01g108300.2	stop_gained	vcZ11P254	IPR009057	Homeodomain-like
Solyc01g109670.2	stop_gained	vcZ11R4M7	IPR009057	Homeodomain-like

IPR001005
IPR006447
IPR009057
IPR004827
IPR011598
IPR009057

Tools

Name	Description
Variant Effect Predictor 	Analyse your own vari of known and unknow (VEP) tool.
BLAST/BLAT	Search our genomes
BioMart	Use this data-mining t
Assembly converter	Map (liftover) your dat
Ensembl Genomes Virtual Machine	Pre-configured Virtual Ensembl Genomes br
Ensembl Genomes REST server	Access Ensembl data

- **Current release (42)**
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Mailing list
- **Tools**

	Online tool	Download code	Documentation
Genomes			
Pathways			
BLAST			
Genomes			

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

FTP site



Gramene FTP site

In addition to data produced in collaboration with Ensembl Plants (see below), Gramene provides legacy data from the MaizeSequence.Org website, and in-house developed BioCyc pathway databases.

Ensembl Plants FTP server

Data (and file formats) available for the plant reference genomes in Gramene and Ensembl Plants include:

- DNA, cDNA and Protein (FASTA)
- Genomic DNA sequence (EMBL and GenBank flatfile dumps)
- Ensembl core databases and specific features (MySQL and simple text format)
- Species-centered data including comparative genomics annotations (TSV)
- Gene annotations (GTF and GFF3)
- Variation data (GVF and VCF)

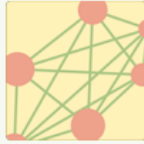
<http://gramene.org/ftp-download>



PATHWAYS



2016 CSHL Cereal Genomics Course



- Metabolic & signaling pathways for crops & plant models
- Pathway enrichment & inter-species comparison
- ATLAS baseline & differential expression

Pathways Platform

The screenshot shows the Plant REACTOME website. The header features the 'Plant REACTOME' logo over a background image of green plants. Below the header is a navigation bar with links: About, Content, Documentation, Tools, Community, Download, and Contact. A search bar on the right contains the text 'e.g. YUC4, cytokinin' and a 'Search' button. The main content area is divided into several sections:

- Navigation Icons:** A grid of six icons with labels: 'Browse Pathways' (hierarchy icon), 'Analyze Data' (microscope icon), 'Tutorial Video' (person with screen icon), 'User Guide' (document icon), 'Data Download' (download arrow icon), and 'Contact Us' (envelope icon).
- About Plant Reactome:** A text block stating: 'Plant Reactome is a freely accessible database of plant metabolic and regulatory pathways. Our goal is to provide plant researchers tools for visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education.'
- Logos:** A row of logos for partner institutions: Gramene, OSU, OICR, CSH, Cold Spring Harbor Laboratory, EMBL-EBI, and REACTOME.
- Footer:** A dark green bar with links for 'About', 'Content', 'Documentation', 'Tools', 'Community', and 'Download'. It also includes social media icons for Facebook, Twitter, and YouTube.
- Right Sidebar:** A section titled 'Gramene News' with the sub-header 'Current Gramene Release (51)'. It features three tweets from @GrameneDatabase, each with an 'Embed' and 'View on Twitter' link.



2016 CSHL Cereal Genomics Course

Orthology-based pathway projections


- 241 curated reference pathways (*O. sativa japonica*)
- Orthologous projections for 62 species

	Taxa	Pathways	Reactions	Gene products	Sequence source	Outgoing Link	Homology method
Dicot	Maloid	<i>Arabidopsis lyrata</i>	187	508	947	Ensembl Genes	Ensembl Genes
		<i>Arabidopsis thaliana</i>	187	508	944	Ensembl Genes	Ensembl Genes
		<i>Brassica oleracea</i>	186	497	1412	Ensembl Genes	Ensembl Genes
		<i>Brassica napus</i>	186	504	1443	Ensembl Genes	Ensembl Genes
		<i>Citrus sinensis</i>	187	538	2026	Phytozome	Phytozome
	Core Eudicot	<i>Medicago truncatula</i>	186	536	1465	Phytozome	Phytozome
		<i>Nepenthes distichophylla</i>	186	508	1181	Ensembl Genes	Ensembl Genes
		<i>Neonotia flexilis</i>	181	511	775	Ensembl Genes	Ensembl Genes
		<i>Populus deltoides</i>	186	518	788	Phytozome	Phytozome
		<i>Populus nigra</i>	184	522	801	Phytozome	Phytozome
Angiosperm	Foloid	<i>Cyperus setiger</i>	187	525	1059	Ensembl Genes	Ensembl Genes
		<i>Eleusine indica</i>	186	528	1182	Phytozome	Phytozome
		<i>Phragmites australis</i>	186	514	889	Phytozome	Phytozome
		<i>Setaria italica</i>	187	499	1736	Ensembl Genes	Ensembl Genes
		<i>Sorghum bicolor</i>	187	499	975	NCBI RefSeq	Phytozome
	Myrtaceae	<i>Myrtus communis</i>	186	511	1771	Phytozome	Phytozome
		<i>Myrtus laevis</i>	186	511	1771	Phytozome	Phytozome
		<i>Myrtus pimenta</i>	186	511	1771	Phytozome	Phytozome
		<i>Myrtus sp.</i>	186	511	1771	Phytozome	Phytozome
		<i>Myrtus sp.</i>	186	511	1771	Phytozome	Phytozome
Monocot	Poaceae	<i>Hordeum vulgare</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Lolium perenne</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Pennisetum glaucum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Triticum aestivum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Zea mays</i>	186	527	873	Ensembl Genes	Ensembl Genes
	Gramineae	<i>Brachypodium distachyon</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Hordeum vulgare</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Lolium perenne</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Pennisetum glaucum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Triticum aestivum</i>	186	527	873	Ensembl Genes	Ensembl Genes
Gymnosperm	Pinaceae	<i>Pinus taeda</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Pinus strobus</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Pinus sylvestris</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Pinus sp.</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Pinus sp.</i>	186	527	873	Ensembl Genes	Ensembl Genes
	Coniferales	<i>Abies balsamea</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Juniperus communis</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Pinus taeda</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Pinus strobus</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Pinus sylvestris</i>	186	527	873	Ensembl Genes	Ensembl Genes
Lycopodium	Lycopodiaceae	<i>Lycopodium obscurum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Lycopodium obscurum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Lycopodium obscurum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Lycopodium obscurum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Lycopodium obscurum</i>	186	527	873	Ensembl Genes	Ensembl Genes
	Selaginellales	<i>Selaginella selaginoides</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Selaginella selaginoides</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Selaginella selaginoides</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Selaginella selaginoides</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Selaginella selaginoides</i>	186	527	873	Ensembl Genes	Ensembl Genes
Bryophyte	Bryophyta	<i>Physcomitrella patens</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Physcomitrella patens</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Physcomitrella patens</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Physcomitrella patens</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Physcomitrella patens</i>	186	527	873	Ensembl Genes	Ensembl Genes
	Marchantiophyta	<i>Marchantia polymorpha</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Marchantia polymorpha</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Marchantia polymorpha</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Marchantia polymorpha</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Marchantia polymorpha</i>	186	527	873	Ensembl Genes	Ensembl Genes
Chlorophyte	Chlorophyta	<i>Chlamydomonas reinhardtii</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Chlamydomonas reinhardtii</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Chlamydomonas reinhardtii</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Chlamydomonas reinhardtii</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Chlamydomonas reinhardtii</i>	186	527	873	Ensembl Genes	Ensembl Genes
	Rhodophyta	<i>Rhodospirillum rubrum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Rhodospirillum rubrum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Rhodospirillum rubrum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Rhodospirillum rubrum</i>	186	527	873	Ensembl Genes	Ensembl Genes
		<i>Rhodospirillum rubrum</i>	186	527	873	Ensembl Genes	Ensembl Genes

<http://plantreactome.gramene.org/pages/content/release-summary>




2016 CSHL Cereal Genomics Course



Analyze Data

Analyze your data



Analysis Tools

This tool merges pathway identifier mapping, overrepresentation and expression analysis into a single tabbed data analysis portal, with integrated visualization and summary features.

Select a file from your computer and click on the "Analyse" button to perform the analysis.

Select data file for analysis
Browse... No file selected.
☒ Project to Oryza sativa
Analyse

► [Click here to paste your data or try example data sets...](#)


Species Comparison

This tool allows you to compare O.sativa pathways with those in any of the other species inferred from Reactome by orthology.

Use the species selector to choose the other species and click on the "Compare" button to perform the comparison.

Compare Oryza sativa with
Select a species...
Compare

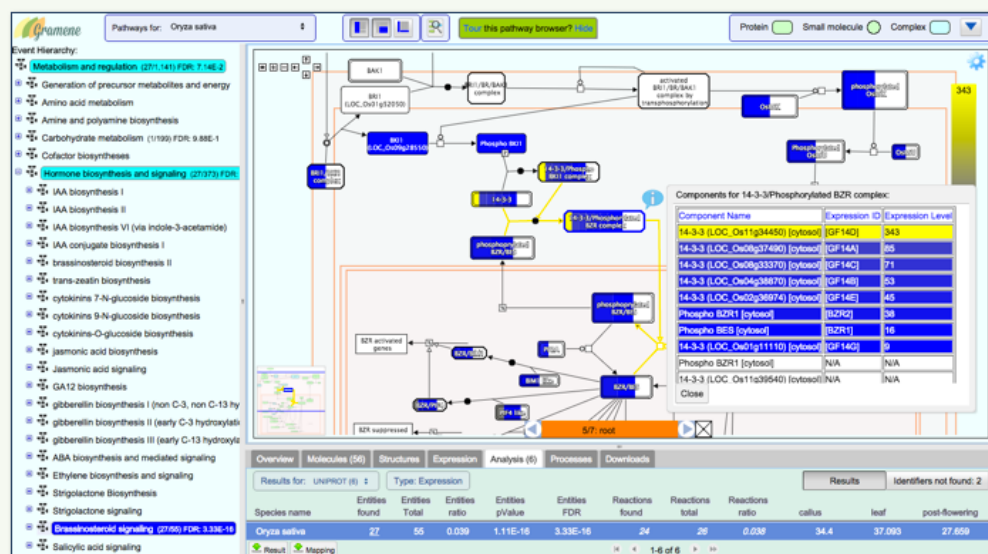
Overview
Molecules
Expression
Analysis
Processes
Downloads

Your analysis result will display here after you submit your data using the Analysis tool. Please open the Analysis tools by clicking the icon  in the top bar above pathway browser window



plantreactome.gramene.org

Tools for enrichment analysis and species comparison



Expression overlay

Sample data drawn from EBI Expression Atlas 7-tissue baseline expression experiments (Petryszak et al., NAR, Oct 2015)



2016 CSSE Global Genomics Course

The screenshot displays the KEGG Pathway Analysis Tools interface. The top section shows a list of pathways for 'Oryza sativa' (rice), with 'Glycolysis' selected. The 'Analysis Tools' section provides instructions on how to use the tool, including a 'Select data file for analysis' button and a 'Choose File' button. The main area shows a metabolic pathway map for 'Glycolysis' (G00001). The map includes various enzymes and metabolites, with a color scale indicating expression levels from 0.00 to 1.00. A table on the right lists 'Components for C01H bound to JA-IR' with columns for Component Name, Expression ID, and Expression Level. Below the map, a table titled 'Results for: ENDOG108' provides a summary of pathway enrichment data.

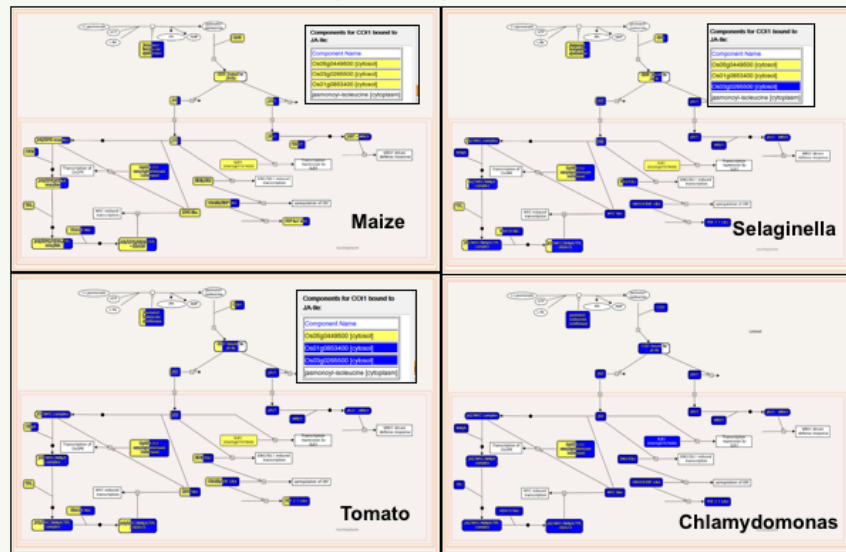
Pathway name	Species name	Enrichment ratio	Enrichment pValue	Enrichment FDR	Enrichment Reactions	Enrichment total
Glycolysis	Glycine max	6	0.003	100	100	1
Geranylphosphate biosynthesis	Glycine max	6	0.003	100	100	1
IAA conjugate biosynthesis I	Glycine max	54	0.000	100	100	14
Ascorbic acid signaling	Glycine max	62	0.001	100	100	12
Calvin cycle	Glycine max	33	0.014	100	100	3

Annotations at the bottom of the screenshot:

- Events with Enrichment Statistics
- Download pathway Enrichment data and mapped gene loci
- Use arrows to view expression from different data points*
- Data for unmapped gene loci

Comparative pathway analysis

Example: JA signaling in rice vs 3 species



Species comparison

EDUCATIONAL RESOURCES



2016 CSHL Cereal Genomics Course



Videotutorials





Gramene Webinar OCT 2016: Gramene Search by Andrew
5 views • 1 day ago



Gramene Webinar Jun 2016: Tomato genomic resources an...
17 views • 3 months ago



April 2016 webinar: New Gramene Search Interfaces
50 views • 5 months ago



March 2016 Webinar: Variation Data in Gramene
34 views • 6 months ago



Gramene's Plant Reactome: resources for maize research
25 views • 7 months ago



Example Query Using Filters
27:54



Plant Reactome Analysis Tools Gramene Portal
19:45



Gramene Webinar Nov 2015: Variant Effect Prediction analy...
81 views • 10 months ago



Gramene webinar Aug 2015: Plant Reactome
181 views • 1 year ago



Gramene Webinar July 2015 by Dr. Pankaj Jaiswal: Arabidopsis
45 views • 1 year ago



Programatic Access to Ensembl Genome by Dr. Paul Kersey
28 views • 1 year ago



Exploring Gene Trees by Dr. Joshua Stein - Gramene Webin...
38 views • 1 year ago



Gramene Webinar June 2015: Upload and analyze your data by
42 views • 1 year ago



Browsing & Comparing Genomes using the Gramene Ensembl
72 views • 1 year ago



Maize Data and Resources by Dr. Marcela Karey Tello-Ruiz
69 views • 1 year ago



Rice Data and Resources in Gramene Database
105 views • 1 year ago



Plant Reactome @ PAG 2015
171 views • 1 year ago



Introduction to the Gramene Website - Nov 2014 Webinar
70 views • 1 year ago



Introduction to Gramene
35 views • 1 year ago



Plant Reactome final
18 views • 1 year ago

<https://goo.gl/In9RLD>



2016 CSHL Cereal Genomics Course

<http://www.gramene.org/outreach>

How to reach us

- E-mail: feedback@gramene.org
- Contact form:
<http://www.gramene.org/contact>
- Gramene YouTube channel & News blog
- Announcements mailing list

facebook



<https://www.facebook.com/Gramene>

[@GrameneDatabase](https://twitter.com/GrameneDatabase)



2016 CSHL Cereal Genomics Course

<http://brie4.cshl.edu/mailman/listinfo/gramene-announce>



Team of very talented scientists and developers. Thanks!

Gramene - Exploring Function through Comparative
Genomics and Network Analysis
NSF IOS 1127112 (2011- 2017)

Doreen Ware, PI (USDA-ARS, CSHL)

Michael Campbell, Kapeel Chougule, Yinping Jiao, Sunita Kumari, Joe Mulvaney, Andrew Olson, Joshua Stein, Marcela K. Tello-Ruiz, Jim Thomason, Peter van Buren, Bo Wang, Sharon Wei

Pankaj Jaiswal, Co-PI (OSU)

Noor Al-Bader, Justin Elser, Matthew Geniza, Parul Gupta, Sushma Naithani, Justin Preece

Paul Kersey / Robert Petryszk (EMBL-EBI)

Dan Bolser, Christopher Grabmuller, Chuang Kee Ong, Dan Staines, Brandon Walts / Elisabet Barrera, Maria Keays, Oliver Mannion, Nuno Fonseca, Laura Huerta Martinez

Lincoln Stein (OICR)

Peter D' Eustachio (NYU); Guanming Wu, Robin Haw, Joel Weiser, Sheldon McKay; Antonio Fabregat (EBI)

Crispin Taylor (ASPB)

Patty Lockhart; Weijia Xu (TACC), Amit Gupta(TACC)



2016 CSHL Cereal Genomics Course

Use Case

EXPLORING CLE18



2016 CSHL Cereal Genomics Course

Gene Example: *cle18*

- CLAVATA3/ESR-related (CLE) peptide 18
- Involved in plant growth and development

Whole-genome sequencing of multiple *Arabidopsis thaliana* populations

Jun Cao^{1,2}, Korbinian Schneeberger^{1,2,3}, Stephan Ossowski^{1,3,4,5}, Torsten Günther^{2,3}, Sebastian Bender¹, Joffrey Fitz¹, Daniel Koenig¹, Christa Lanz¹, Oliver Stegle⁶, Christoph Lippert⁶, Xi Wang¹, Felix Ott¹, Jonas Müller¹, Carlos Alonso-Blanco⁷, Karsten Borgwardt⁶, Karl J Schmid³ & Detlef Weigel¹

The plant *Arabidopsis thaliana* occurs naturally in many different habitats throughout Eurasia. As a foundation for identifying genetic variation contributing to adaptation to diverse environments, a [1001 Genomes Project](#) to sequence geographically diverse *A. thaliana* strains has been initiated. Here we present the first phase of this project, based on population-scale sequencing of 80 strains drawn from eight regions throughout the species' native range. We describe the majority of common small-scale polymorphisms as well as many larger insertions and deletions in the *A. thaliana* pan-genome, their effects on gene function, and the patterns of local and global linkage among these variants. The action of processes other than spontaneous mutation is identified by comparing the spectrum of mutations that have accumulated since *A. thaliana* diverged from its closest relative 10 million years ago with the spectrum observed in the laboratory. Recent species-wide selective sweeps are rare, and potentially deleterious mutations are more common in marginal populations.

Supplementary Table 3: CNVs, SNPs with predicted drastic effects on gene function, and SV deletions that overlap coding sequences by at least 50 bp.

SNPs with predicted drastic effects on gene function.

2 VEPs in CLE18
(Cao *et al*, 2011)

Chr	Pos	Type	Strains	Locus ID	Annotation
1	24,632,348	stop	1	AT1G66145	CLE18 (CLAVATA3/ESR-RELATED 18); receptor binding
1	24,632,362	stop	3	AT1G66145	CLE18 (CLAVATA3/ESR-RELATED 18); receptor binding

CLE 18 (CLAVATA3/ESR-related)

Plant CLAVATA3/ESR-related (CLE) peptides have diverse roles in plant growth and development

A different gene:

CLV3

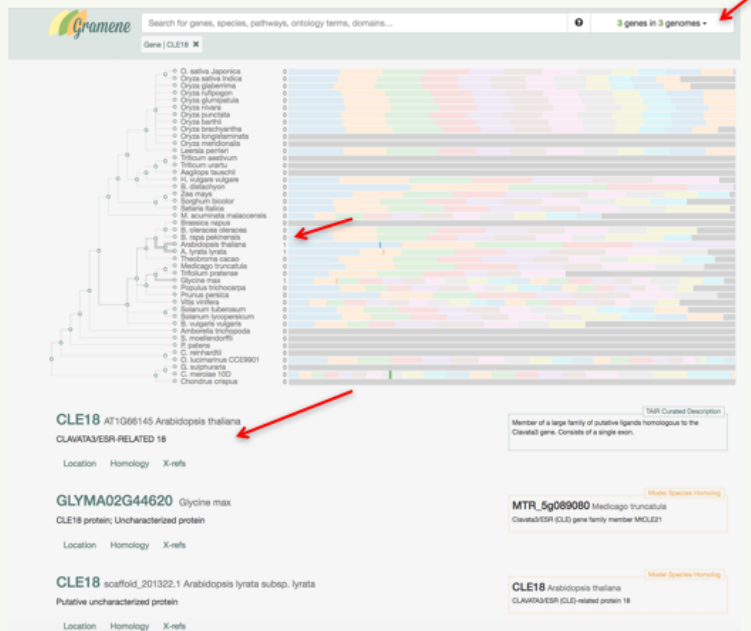
[AT2G27250](#)

Contributes to meristem structure and identity in both, shoot and flower

stop-gained

ENSVATH00125659

Search results







http://ensembl.gramene.org

Example: **CLE18**

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData Login/Register Search all species...

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 HG2214v1
 Zea mays AGPv3	 Physcomitrella patens ASM242v1

★ Log in to customize this list

All genomes
-- Select a species --

[View full list of all Ensembl Plants species](#)

Updated assembly for wheat

The **bread wheat assembly** in Gramene from release 45 onwards 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 [Chapman et al \(2015\)](#).

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 Converter Tool](#). For chromosome 3B, the GDEC-derived models have replaced the PGSB annotations are primary but where mappable, the PGSB genes are viewable as an additional track.

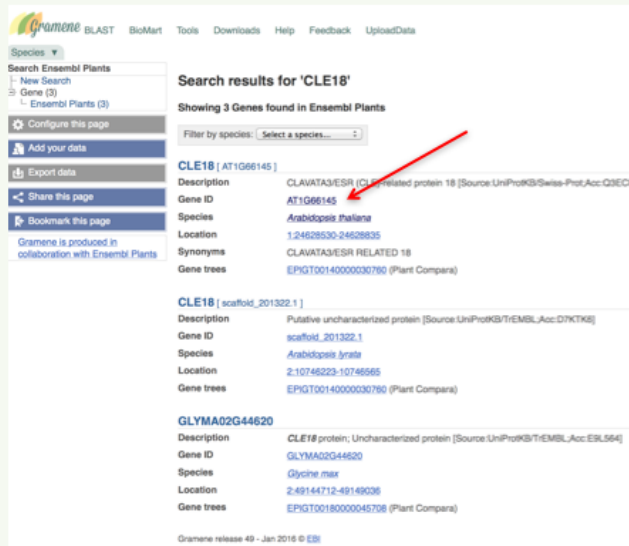
The anchoring process (with the exception of 3B) places 1,290,751 CSS contigs into chromosomal pseudomolecules for a total length of 4,237,502,413 bp. Ensembl Plants also incorporates 261,251 unanchored scaffolds, with cumulative length of 1,307,508,887 bp.

- Mascher M, Muehlbauer GJ, Rokhsar DS, Chapman J, Schmutz J, Barry K, Muñoz-Amatrián M, Close TJ, Wise RP, Schulman AH, Himmelbach A, Mayer KF, Scholz U, Poland JA, Stein N, Waugh R (2013). *Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ)*. The Plant Journal 76(4):718-727. DOI: 10.1111/tpj.12319.
- Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Viret R, Schumacher M, Lecomte E, Simeon D, Vilain A, Aury M, Yvert G

Gramene 2016 CSHL Cereal Genomics Course

Plant CLAVATA3/ESR-related (CLE) peptides have diverse roles in plant growth and development

Search results



Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Species ▼

Search Ensembl Plants

- New Search
- Gene (3)
- Ensembl Plants (3)

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

Search results for 'CLE18'

Showing 3 Genes found in Ensembl Plants

Filter by species:

CLE18 [AT1G66145]	
Description	CLAVATA3/ESR (CLE) related protein 18 [Source:UniProtKB/Swiss-Prot;Acc:Q2E0C9]
Gene ID	AT1G66145
Species	Arabidopsis thaliana
Location	1:24628530-24628635
Synonyms	CLAVATA3/ESR RELATED 18
Gene trees	EPHGT00140000030780 (Plant Compara)
CLE18 [scaffold_201322.1]	
Description	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:D7YKTK8]
Gene ID	scaffold_201322.1
Species	Arabidopsis lyrata
Location	2:10745223-10745555
Gene trees	EPHGT00140000030780 (Plant Compara)
GLYMA02G44620	
Description	CLE18 protein; Uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:E9L564]
Gene ID	GLYMA02G44620
Species	Glycine max
Location	2:49144712-49149036
Gene trees	EPHGT00180000045708 (Plant Compara)

Gramene release 43 - Jan 2016 © EBI

CLE18 gene page

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData Login/Register

Search Arabidopsis thaliana...

Arabidopsis thaliana (TAIR10) Location: 1-24,628,530-24,628,835 **Gene: CLE18** Transcript: CLE18

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Sequence**
- Secondary structure
- Gene families
- External references
- Regulation
- Orthologs
- GO: Cellular component
- GO: Molecular function
- GO: Biological process
- PO: Plant anatomical entity
- PO: Plant structure development
- Literature
- Plant Compara
- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues
- Phenotype
- Genetic Variation
- Variant table
- Structural variants
- Variant image
- External data
- Gene expression
- Personal annotation
- ID History

Gene: CLE18 AT1G66145

Description: CLAVATA3/ESR (CLE)-related protein 18 [Source:UniProtKB/Swiss-Prot;Acc:Q3ECH9.6]

Location: Chromosome 1: 24,628,530-24,628,835 reverse strand.

About this gene: This gene has 1 transcript (splice variant) and 1 orthologue.

Transcripts: [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags
CLE18	AT1G66145.1	306	101aa	Protein coding	Q3ECH9.6	NM_146931.2 NP_663472.2	

Summary

Name: CLE18 (UniProtKB Gene Name)

UniProtKB: This gene has proteins that correspond to the following UniProtKB identifiers: [Q3ECH9.6](#)

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)

Gene Legend: Protein Coding

Gene sequence

Arabidopsis thaliana (TAIR10) Location: 1:24,628,530-24,628,835 Gene: CLE18 Transcript: CLE18

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Gene alleles
- Sequence**
- Secondary structure
- Gene families
- External references
- Regulation
- Ontologies
 - GO: Cellular component
 - GO: Molecular function
 - GO: Biological process
 - PO: Plant anatomical entity
 - PO: Plant structure development
- Literature
- Plant Compara
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Par-taxonomic Compara
 - Gene Tree
 - Orthologues
- Phenotype
- Genetic Variation
 - Variant table
 - Structural variants
 - Variant image
- External data
 - Gene expression
 - Personal annotation
 - ID History (click history)
- Configure this page**
- Add your data
- Export data
- Share this page
- Bookmark this page

Gene: CLE18 AT1G66145

Description CLAVATA3/ESR (CLE)-related protein 18 [Source:UniProt]

Location Chromosome 1: 24,628,530-24,628,835 reverse strand.

About this gene This gene has 1 transcript (splice variant) and 1 orthologue.

Transcripts [Show transcript table](#)

Marked-up sequence

[Download sequence](#) [BLAST this sequence](#)

Exons CLE18 exons All exons in this region

```
>chromosome1:TAIR10:1:24627930:24629435:1
TAACAAGAGATTATGTTGGAAACCAATCTGTCAATATTAGATCATCGCAAAATAT
TTCCGTAAAGTGAATTCACAAATTCCTTCACAAAGCTAAATTTTCAATCATTTGGG
TTTATTTCCTGAAATTTGAAATCTATTAATCTTTTATGATTAGAAATGTTTACCA
TATATGAACGAACAAATTAACATAGATGAGATTAAATAAAATTTTCAAAATGAGA
TTTGTTCATGAAATAAAGATAGCTTATGAGATTTCAAAGGAAATTTGAACCT
AAATTTGTGCGACCACTATTTTATGCTCAATAGTAAACCAATTTAAATGAG
ATTACATGAATAATTCGGGAACCTTTTGTGGTCAAACTTGGGTTGAGT
TAAAGGGTCAAGTACATCTCTTAATATATCTCAAGTCAATTCGAATGAGATTGATT
AGCTTCTTAAATAGATTAATAATCTCACTATAAAATAAAGGGGAAACAGAGGTG
TAGCCAACTAAAAATCTATACAAAACCAATCAAAAGCAAAATAAGATCAGAAAG
ATGCAATTTTAAAGGTGGTGGTATTAATCAAGCTAACTCTGTGCTCAATACG
TCTCTATTTGCTATTTGAGAGACCAATTTGAGTGGATGACAGACATGCA
ACAGGTCTGATCTTTACATAATCTCTCTCAGCTTCACCAAAACATCATTTGATC
GGAGTGGAGAAACATATCGACCGCTCTGGATATATGAGATTATGATCTCATCAT
GACAGATGCTTACAGAACTCTCTGAGCTGACAGATATATGATGATTTGATGAG
GTCTAGACAAATATATATATATATATATATATATATATATATATATATATATAT
AGATGAGATGATTTGGAGTGTAAACCTTAATATATATATATATATATATATATAT
TACGATGATGATTTGATGAGATGATATATATATATATATATATATATATATATAT
AAAAACAGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
TTCTTCTACAAATTAAGAAATATGGGCTTGGGTCATATGATGATGATGATGATGAT
CACACCAAAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
AGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
GCCCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
AAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
TTTTCA
```

Configure (Sequence) page: Show variants

The screenshot shows the 'Configure (Sequence) page' in the Gramene interface. The 'Display options' section is active, showing various settings for sequence display. Red arrows are used to highlight specific features:

- Arrow 1 points to the 'Show variants' dropdown menu, which is currently set to 'Yes and show links'.
- Arrow 2 points to the 'Line numbering' dropdown menu, which is currently set to 'Relative to this sequence'.
- Arrow 3 points to the 'Close' button (a small 'X' icon) in the top right corner of the configuration panel.

The 'Display options' section includes the following settings:

- 5' Flanking sequence (upstream): 500 (Maximum of 1000000)
- 3' Flanking sequence (downstream): 500 (Maximum of 1000000)
- Number of base pairs per row: 60 bps
- Additional exons to display: Core exons
- Orientation of additional exons: Display exons in both orientation
- Show variants: Yes and show links (indicated by arrow 1)
- Hide variants longer than 10bp: ☒
- Hide variants by frequency (MAF): ☐ Don't hide
- Filter variants by consequence type: No filter (indicated by arrow 2)
- Line numbering: Relative to this sequence (indicated by arrow 2)
- Display pop-up information on mouseover: Yes

The bottom of the page shows a sequence viewer with the following sequence:

```
TATATATGATTCATTTTATTTTGTAGGATCTAGTGGAAATGATCA  
ACTTGTATCTATGTTCAAAACCTTTGATTTGAAATGATCTAGTCA  
GGCAACGATTTGAATATGGCTTATGGTCCCAAAATTTTAACTCATCTATGA  
AAATGAGTGGCAACATAGCTAGTGTCTATTTTGAATCTTTTAAATATGCTCA  
TTTCA
```

[illegible]

Variant table (Gene & Transcript)

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData Login/Register

Arabidopsis thaliana (TAIR10) Location: 1:24,628,530-24,628,835 Gene: CLE18 Transcript: CLE18 Search Arabidopsis thaliana...

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Gene alleles
- Sequence
- Secondary Structure
- Gene families
- External references
- Regulation
- Ontologies
- GO: Cellular component
- GO: Molecular function
- GO: Biological process
- PO: Plant anatomical entity
- PO: Plant structure development
- Literature
- Plant Compara
- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogs
- Phenotype
- Phylogenetic Compara
- Gene tree
- Orthologues
- Phenotype
- Genetic Variation
- Variant table**
- Structural variants
- Variant image
- External data
- Gene expression
- Personal annotation
- GO History

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gene: CLE18 AT1G66145

Description CLAVATA3/ESR (CLE)-related protein 18 [Source:UniProtKB/Swiss-Prot;Acc:Q20CH9.6]

Location Chromosome 1: 24,628,530-24,628,835 reverse strand.

About this gene This gene has 1 transcript (splice variant) and 1 orthologue.

Transcripts [Show transcript table](#)

Variant table

This table shows known variants for this gene. Use the 'Consequence Type' filter to view a subset of these.

Filter Consequences: All Filter Other Columns

Variant ID	Chr: bp	Alleles	Class	Conseq. Type	AA	AA coord	Transcript
ENSVATH05079608	1:24623536	T/G	SNP	Downstream gene variant	-	-	AT1G66145.1
ENSVATH05079607	1:24623549	C/G	SNP	Downstream gene variant	-	-	AT1G66145.1
ENSVATH13741706	1:24623552	C/G	SNP	Downstream gene variant	-	-	AT1G66145.1
ENSVATH01479067	1:24623558	G/A	SNP	Downstream gene variant	-	-	AT1G66145.1
ENSVATH05079608	1:24623581	A/C	SNP	Downstream gene variant	-	-	AT1G66145.1
ENSVATH05079609	1:24623600	T/A	SNP	Downstream gene variant	-	-	AT1G66145.1
ENSVATH01479068	1:24623609	A/T	SNP	Downstream gene variant	-	-	AT1G66145.1

Filter consequences

Arabidopsis thaliana (TAIR10) Location: 1:24,628,530-24,628,835 Gene: CLE18 Transcript: CLE18

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Supporting evidence
- Gene allies
- Sequence
 - Secondary Structure
- Gene families
- External references
- Regulation
- Orthologs
 - GO: Cellular component
 - GO: Molecular function
 - GO: Biological process
 - PO: Plant anatomical entity
 - PO: Plant structure development
- Literature
- Plant Compara
 - Genomic alignments
 - Gene tree
 - Gene gain-loss tree
 - Orthologs
 - Paralogs
 - Pan-taxonomic Compara
 - Gene Tree
 - Orthologs
- Phenotype
- Genetic Variation
 - Variant table
 - Structural variants
 - Variant image
 - Variant data
- Gene expression
- Personal annotation
- History
- Gene history

Gene: CLE18 AT1G66145

Description: CLAVATA3/ESR (CLE)-related protein 18 [Source UniProtKB/Swiss-Prot; Acc: Q2EDH9.2]

Location: Chromosome 1: 24,628,530-24,628,835 reverse strand.

About this gene: This gene has 1 transcript (splice variant) and 1 orthologous.

Transcripts: [Show transcript table](#)

Variant table

Filter: **Consequences: All** [Filter Other Columns](#)

Consequences: 27/27 vns

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

PTV = Protein Truncating Variant

Transcript ablation (2) Inframe deletion (2) 3 prime UTR variant (2) 5 prime UTR variant (2) Splice donor variant (2) Missense variant (14) 3 prime UTR variant (2) Splice acceptor variant (2) protein altering variant (2) Non-coding transcript exon variant (2)

es, Type AA AA coord Transcript

gained G* 51 AT1G66145.1

gained L* 48 AT1G66145.1

About Gramene | About EnsemblGenomes | Contact Us | Help

SNPs with predicted drastic effects on gene function.

2 VEPs in CLE18 (Cao et al, 2011)

Chr	Pos	Type	Strains	Locus ID	Annotation
1	24,632,348	stop	1	AT1G66145	CLE18 (CLAVATA3/ESR-RELATED 18); receptor binding
1	24,632,362	stop	3	AT1G66145	CLE18 (CLAVATA3/ESR-RELATED 18); receptor binding



Structural variants

- Supporting evidence
- Gene aliases
- Sequences
 - Secondary Structure
- Gene families
- External references
- Regulation
- Ontologies
 - GO: Cellular component
 - GO: Molecular function
 - GO: Biological process
 - PO: Plant anatomical entity
 - PO: Plant structure development
- Literature
- Plant Compare
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
- Phenotype
- Genetic Variation
 - Variant table
 - Structural variants
- External data
 - Gene expression
 - Personal annotation
 - ID history
 - Gene history
- Configure this page
- Add your data
- Export data
- Share this page
- Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

Location: **Chromosome 1: 24,628,530-24,628,835 reverse strand.**
This gene has 1 transcript (splice variant) and 1 orthologue.

Transcripts

[Show transcript table](#)

Variant image

Click

Location: 1:24628500-24628865

Class: deletion

Location: 1:24628500-24628865

Aliases: TTTTACGAATGCA

Consequence: Coding sequence variant

Source: Ensembl

Evidence: -

Variations

ATCGA1451 CUE18

366 bp

366 bp

Variant legend

- Stop gained
- Coding sequence variant
- Missense variant
- Upstream gene variant
- Synonymous variant
- Intergenic variant

Go to Transcript page

Grainene BLAST BioMart Tools Downloads Help Feedback UploadData [Search Arabidopsis thaliana...](#) [Login/Registe](#)

Arabidopsis thaliana (TAIR10) Location: 1:24,628,530-24,628,855 Gene: CLE18 Transcript: CLE18

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- Sequence references
 - General identifiers
 - Oligo probes
- Genetic Variation
 - Variant table
 - Variant image
 - Population comparison
 - Comparison image
- Protein information
 - Protein summary
 - Domains & features
 - Variants
- External data
 - Personal annotation
 - ID History
 - Transcript history
 - Protein history

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Grainene is produced in collaboration with Ensembl Plants

Transcript: CLE18 AT1G66145.1

Description CLAVATAS1/ESR (CLE)-related protein 18 [Source: UniProtKB/Swiss-Prot; Acc: Q3ECH9.6]

Location Chromosome 1: 24,628,530-24,628,835 reverse strand.

About this transcript This transcript has 1 exon, is annotated with 3 domains and features, is associated with 24 variations and maps to 1 oligo probe.

Gene This transcript is a product of gene AT1G66145 [Hide transcript table](#)

Show/hide columns (1 hidden) Filter

Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags
CLE18	AT1G66145.1	306	101aa	Protein coding	Q3ECH9.6	NM_148931.6 NP_663472.6	

Summary

Reverse strand

Statistics Exons: 1 Coding exons: 1 Transcript length: 306 bps Translation length: 101 residues

UniProt This transcript corresponds to the following UniProt identifiers: [Q3ECH9.6](#)

Type Known protein coding

Annotation Method Gene annotation by TAIR through a process of automatic and manual curation.

Grainene release 49 - Jan 2018 © EBI [About Grainene](#) | [About Ensembl Genomes](#) | [Contact Us](#) | [Help](#)

[illegible]

Transcript Sequence: cDNA

[illegible]

Go to location page



Configure (location) page: Show variation

The screenshot displays the Gramene BLAST interface, specifically the 'Configure Region Image' page. The left sidebar lists various tracks, with 'Variation' highlighted. A 'Change track style' dialog box is open, showing options for 'Normal (collapsed for windows over 200kb)', 'Collapsed', 'Expanded with name (hidden for windows over 10kb)', and 'Expanded without name'. The 'Expanded with name' option is selected. The main panel shows the 'Variation' track configuration, including checkboxes for 'Enable/disable all Variation sets', 'Enable/disable all 1001_AIL_Variations_by_Class', 'Enable/disable all Salk_All_Variations', and 'Enable/disable all WTCG_All_Variations'. The bottom of the interface shows a genomic track with various annotations and a scale bar.

Configure (Alignments) page

The screenshot shows the 'Display options' panel in the Ensembl genome browser. The panel is divided into several sections with various settings:

- Number of base pairs per row:** 120 bps
- Strand:** Forward
- Core exons:** Yes
- Display exons in both orientation:** Yes (indicated by a red arrow)
- Yes and show links:** Yes (checked, indicated by a red arrow)
- Don't hide:** Yes (checked)
- No filter:** Selected (options: 3 prime UTR variant, 5 prime UTR variant, Coding sequence variant, Downstream gene variant)
- Line numbering:** Relative to this sequence (indicated by a red arrow)
- Codons:** Do not show codons (indicated by a red arrow)
- Show conservation regions:** Yes (checked)
- Mark alignment start/end:** No (unchecked)

The background shows the Ensembl genome browser interface with the 'Display options' panel open on the left side. The main content area displays the genomic track for the Arabidopsis thaliana gene, showing the gene structure with exons and introns, and the sequence alignment for the selected gene.

[illegible]

Transcript Sequence: cDNA

[illegible]

Explore this variant

Arabidopsis thaliana (TAIR10)
Location: 1:24,628,500-24,628,865
Gene: CLE18
Transcript: CLE18
Variant: ENSVATH00125659

Variant displays

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

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

ENSVATH00125659 SNP

Original source

Alleles: A/T

Location: Chromosome 1:24628699 (forward strand)

Most severe consequence: Stop gained

Synonyms: Perlegen PERL0223135

HGVs names: This variant has 3 HGVs names - Show

About this variant: This variant overlaps 2 transcripts and is associated with 2 phenotypes.

Variation features from Affy_250k, Perlegen_1M, WTCHG and 1000 Genomes

Ambiguity code: W

See all predicted consequences (genes and regulatory)

Explore this variant

Genomic context

Genes and regulation

Genotype frequency

Sample genotypes

Linkage disequilibrium

Phenotype data

Citations

Phylogenetic context

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

Test your own variants with the Variant Effect Predictor

Programmatic access

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

Reference materials

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

5 views: Species, Location, Gene, Transcript & Variation

Correspondence between icons and left bar menu

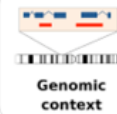
Not grayed out have data

Videos

ENSVATH00125659

http://ensembl.gramene.org/Arabidopsis_thaliana/Variation/Explore?r=1:24628199-24629199;v=ENSVATH00125659;vdb=variation;vf=125659

Genomic context



Arabidopsis thaliana (TAIR10) Location: 1:24,628,500-24,628,865 Gene: CLE18 Transcript: CLE18 Variant: ENSVATH00125659 SNP

Variant displays

- Explore this variant
- Genomic context**
- Genes and regulation
- Transcript sequence
- Genotype frequency
- Sample phenotypes
- Linkage disequilibrium
- Phenotype Data
- Phylogenetic Context
- Citations
- External Data

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

Context

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

Alleles: A/T | Ambiguity code: W

Location: Chromosome 1:24628699 (forward strand) | View in location tab

Most severe consequence: Stop gained | See all predicted consequences (Genes and regulation)

Synonyms: Perlegen PERL0223135

HQVS names: This variant has 3 HQVS names - Hide

1:g.24628699A>T
AT1G66145.1:c.1377>A
AT1G66145.1:p.Leu46X

About this variant: This variant overlaps 2 transcripts and is associated with 2 phenotypes.

Click on "Configure this page" symbol

Genes & regulation



Arabidopsis thaliana (TAIR10) Location: 1:24,628,500-24,628,865 Gene: CLE18 Transcript: CLE18 Variant: ENSVATH00125659

Variant displays

- Explore this variant
- Genomic context
- Flanking sequence
- Sequence frequency
- Sample genotypes
- Linkage disequilibrium
- Phenotype Data
- Phylogenetic Context

ENSVATH00125659 SNP

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

Alleles: A/T | Ambiguity code: W

Location: Chromosome 1:24628699 (forward strand) | [View in location tab](#)

Most severe consequence: Stop gained | [See all predicted consequences \(Genes and regulation\)](#)

Clonotype	AT1G66145	AT1G66145.1 (-)	T	Stop gained	137	137	46	L*	TTA/TAA	Hide
Contig	AT1G66150	AT1G66150.1 (+)	T	Upstream gene variant	-	-	-	-	-	Show

No overlap with Ensembl Regulatory features

No overlap with Ensembl Motif features

Consequence detail for ENSVATH00125659 (T) in AT1G66145.1 [\[back to top\]](#)

Variation name	ENSVATH00125659	
Gene	AT1G66145	
Transcript	AT1G66145.1	
Protein	AT1G66145.1	
Allele (variation)	T	Show
Allele (transcript)	A	Show
Consequence (SO term)	Stop gained - A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript (SO:0001587.6)	Show
Consequence (Old Ensembl term)	STOP_GAINED	
HGV names	AT1G66145.1.1 c.137T>A	
	AT1G66145.1.1 p.Leu46X	
Exon	AT1G66145.1.exon1 (1 of 1, length 306)	
Position in exon	137	
Overlapping protein domains	-	
Variants in same codon	-	
Context	Show in transcript	

Gramene

2016 CSHL Cereal Genomics Course

Flanking sequence

ATTCATT
CGGSGTG
TCATGCT
Flanking
sequence

Arabidopsis thaliana (TAIR10) Location: 1:24,628,199-24,629,199 Variant: ENSVATH00125659

Variant displays

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

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

Gramene is produced in collaboration with Ensembl Plants

ENSVATH00125659 SNP

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

Alleles: A/T | Ambiguity code: W

Location: Chromosome 1:24628699 (forward strand) | View in location tab

Most severe consequence: Stop gained | See all predicted consequences (Genes and regulators)

Synonyms: Perlegen PERL0223135

HQVS names: This variant has 3 HQVS names - Show

About this variant: This variant overlaps 2 transcripts and is associated with 2 phenotypes.

Flanking sequence

Download sequence BLAST this sequence

Flanking sequence

The sequence below is from the reference genome flanking the variant location. The variant is shown in red text. Neighbouring variants are shown with highlighted letters and ambiguity codes. To change the display of the flanking sequence (e.g. hide the other variants, change the length of the flanking sequence), use the "Configure this page" link on the left.

Variants: Flagged variant Focus variant Missense Stop gained Synonymous Upstream

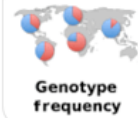
```

TGCATATTTTGCATTTTATATTTTATAGGTCATATCAATCTTTT
TCTGCTTTTCTTTTGGATTTCTATGTCATTTCTATACAACTATCTGTATAC
GCAACTATATGGAATCTACATATACGTTTACACCTCCAACTTTCTCATATTT
ATTTATATTCATATGATGATTTTATATATATTTTCTAGACTCTT
AATTAATGATATATATGTTTCTTATAGAGAGTTATCTAGACATATGCA
TTGAGATTCATATCTCATATTTCCAGACCGGTCATATTTTCTCTGACTT
CAATGATGATTTTGGTCAAGGCTTGGAGGATTTTGAAGGATCAGGACTTTGG
GATCTGCTGTTTCTCTGACGATTTCTTACAAATAGGCAATTTGACGAGGT
ATTAGGAATTAATGATTTATGATTATTTACACACACCTTTTCAATGATCTTT
TTTCTCTTTTCTTTTCTTTGATTTTGTATAGATTTTATGTTTCTAGAC
TCTTTCTCTCTCTTTTATTTTCTTTTATCTTATATAGGGAGCTAT
CAATCTCATTTGCAATTTTGAATATATTAATTTCTTGACCTTTAATCT
GAACTCTTTTGTGATTTTAAATTTTCTTTTCTGAAATTTCTGTAATCTC
ATTTTAAATTTTCTTTTACT
    
```

Gramene

2016 CSHL Cereal Genomics Course

Genotype frequency



Arabidopsis thaliana (TAIR10) Location: 1:24,628,500-24,628,865 Gene: CLE18 Transcript: CLE18 Variant: ENSVATH00125659

Variant displays

- Explore this variant
- Genomic context
- Genes and regulation
- Flanking sequence
- Genotype frequency**
- Sample genotypes
- Orthologous
- Phenotype Data
- Phylogenetic Context
- Citations
- External Data

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

ENSVATH00125659 SNP

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

Alleles: A/T | Ambiguity code: W

Location: Chromosome 1:24628699 (forward strand) | View in location tab

Most severe consequence: Stop gained | See all predicted consequences (Genes and regulation)

Synonyms: Perlegen PERL0223135

HGVS names: This variant has 3 HGVS names - Show

About this variant: This variant overlaps 2 transcripts and is associated with 2 phenotypes.

Genotype frequency

Frequency data (7)

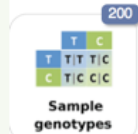
Show/hide columns

Population	Allele: frequency (count)		Genotype: frequency (count)	
1001	A: 0.859 (410)	T: 0.141 (410)	AA: 0.859 (410)	TT: 0.141 (410)
AFFY_250k	A: 0.881 (1179)	T: 0.119 (1179)	AA: 0.881 (1179)	TT: 0.119 (1179)
Cao	A: 0.963 (80)	T: 0.037 (80)	AA: 0.963 (80)	TT: 0.037 (80)
Nordborg	A: 0.778 (180)	T: 0.222 (180)	AA: 0.778 (180)	TT: 0.222 (180)
PERL_1M	A: 0.900 (20)	T: 0.100 (20)	AA: 0.900 (20)	TT: 0.100 (20)
Salik	A: 0.909 (132)	T: 0.091 (132)	AA: 0.909 (132)	TT: 0.091 (132)
WTCHG	A: 0.833 (18)	T: 0.167 (18)	AA: 0.833 (18)	TT: 0.167 (18)

Gramene

2016 CSHL Cereal Genomics Course

Sample genotypes



Arabidopsis thaliana (TAIR10) Location: 1:24,628,500-24,628,865 Gene: CLE18 Transcript: CLE18 Variant: ENSVATH00125659

Variant displays

- Explore this variant
- Genomic context
- Flanking sequence
- Genotype frequency
- Sample genotypes**
- Linkage disequilibrium
- Phenotype Data
- Protein structure
- Citations
- External Data

ENSVATH00125659 SNP

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

Alleles: A/T | Ambiguity code: W

Location: Chromosome 1:24628699 (forward strand) | View in location tab

Most severe consequence: Stop gained | See all predicted consequences (Genes and regions)

Synonyms: Perlegen PERL0223135

HGVIS names: This variant has 3 HGVIS names - Show

About this variant: This variant overlaps 2 transcripts and is associated with 2 phenotypes.

Sample genotypes

Search for a sample: Search (e.g. NA10851)

Summary of genotypes by population (7)

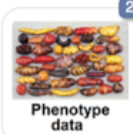
Number of genotypes	Population	Description
58 Show	1001	Population of individuals/strains for 1001 Genomes study, Weigel et al (2011)
140 Show	AFFY_250k	Population of individuals/strains for Affymetrix_250k study, Nordborg et al... (more)
3 Show	Cao	Population of individuals/strains for MPI study, Cao et al (2010)
40 Show	Nordborg	Population of individuals/strains for GMI Nordborg study, Long et al (2011)
2 Hide	PERL_1M	Population of individuals/strains for Perlegen_1M study
12 Show	Salk	Population of individuals/strains for Salk study, Schmitz et al (2010)
3 Show	WTCHG	Population of individuals/strains for WTCHG study, Mott and Kover (2010)

Genotypes for PERL_1M

Sample	Genotype (forward strand)	Description
IN-0-032078 PERL_1M	TTT	-
7b-1-032082 PERL_1M	TTT	-



Phenotypes



Arabidopsis thaliana (TAIR10) Location: 1:24,628,199-24,629,199 Variant: ENSVATH00125659

Variant displays

- Explore this variant
- Genomic context
- Genes and regulation
- Flanking sequence
- Genotype frequency
- Sample genotypes
- Linkage disequilibrium
- Phenotype Data**
- Phylogenetic Context
- Clonotypes
- External Data

ENSVATH00125659 SNP

Original source: Variation features from Affy_250k, Perlegen_1M, WTC10 and 1001 Genomes, with Ensembl identifiers

Alleles: A/T | Ambiguity code: W

Location: Chromosome 1:24628699 (forward strand) | View in location tab

Most severe consequence: Stop gained | See all predicted consequences (Genes and regulation)

Synonyms: Perlegen PERL0223135

HQVS names: This variant has 3 HQVS names - Show X

About this variant: This variant overlaps 2 transcripts and is associated with 2 phenotypes.

Phenotype Data

Significant association(s)

Show/hide columns

Phenotype, disease and trait	Source(s)	Study	Reported gene(s)	Associated allele	Statistics	Genomic Locations
Following inoculation of two leaves per plant with 0.1 ml of 10 ⁸ cfu/ml bacteria in 10 mM MgSO ₄ buffer using a blunt-tipped syringe. Leaf collapse was scored at 20 hrs and again at 24 hrs after inoculation [20C and 12 hrs daylight]	Nordborg	EmmaTrans_svrPpmt_29	-	T	p-value: 0.003020299	View on Karyotype
Following inoculation of two leaves per plant with 0.1 ml of 10 ⁸ cfu/ml bacteria in 10 mM MgSO ₄ buffer using a blunt-tipped syringe. Leaf collapse was scored at 20 hrs and again at 24 hrs after inoculation [20C and 12 hrs daylight]	Nordborg	EmmaTrans_svrPpmt_32	CLE18, AT1G06130, AT1G06173, AT1G06180, AT1G06110, AT1G06120, AT1G06160, ZFP4, TMK1, MUG1	T	p-value: 0.003020299	View on Karyotype

Genes in this region

The following gene(s) in the region of this variant might have associated phenotype data:

Show/hide columns

Gene	HGNC name	Position
AT1G06145	CLE18	Overlap variant



Ensembl variations quick reference card

www.ensembl.org → **Search for a gene**

Gene: TAGAP Location: 6:159,455,500-159,466,184 Variation: rs1738074

Click 'Tools' for the Variant Effect Predictor

The transcript table lists splice variants and has sortable columns

Page-specific help is available

Use the variant effect predictor

Show a table for one variation type

View variations on the cDNA sequence in the transcript tab

Ensembl transcripts (splice variants) are listed

Click a variation ID to open the variation tab

Possible amino acids at a position

version 1.0

Contextual navigation for a gene

View variations in the genomic sequence

The variation table shows variations for a gene

Summary of variations in ENSG00000164891 by consequence type

Consequence	Type	Description
0	Essential splice site	In the first 2 or the last 2 transcripts of an exon
0	Stop gained	In coding sequence, resulting in the gain of a stop codon
0	Stop lost	In coding sequence, resulting in the loss of a stop codon
0	Complex insert	Insertion or deletion that spans an exon-intron boundary
0	Frame shift	In coding sequence, resulting in a frame shift
0	Non-synonymous coding	In coding sequence and results in an amino acid change in the encoded peptide sequence
4	Splice site	1-2 bp into an exon or 3-6 bp into an intron
0	Other	Located within the first, incomplete codon of a transcript whose end coordinate is unknown

Transcript table

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Splices	CCDS
TAGAP-001	ENST00000262086	3887	ENSP00000262086	771	Protein coding	CCDS1242.1
TAGAP-002	ENST00000262087	3913	ENSP00000262087	813	Protein coding	CCDS1243.1
TAGAP-003	ENST00000262088	1189	ENSP00000262088	296	Protein coding	CCDS1244.1
TAGAP-004	ENST00000262089	3968	ENSP00000262089	813	Protein coding	CCDS1245.1

Variant table

Position	Ref	Alt	Effect
159,455,500	A	G	Essential splice site
159,455,501	A	G	Essential splice site
159,455,502	A	G	Essential splice site
159,455,503	A	G	Essential splice site
159,455,504	A	G	Essential splice site
159,455,505	A	G	Essential splice site
159,455,506	A	G	Essential splice site
159,455,507	A	G	Essential splice site
159,455,508	A	G	Essential splice site
159,455,509	A	G	Essential splice site
159,455,510	A	G	Essential splice site
159,455,511	A	G	Essential splice site
159,455,512	A	G	Essential splice site
159,455,513	A	G	Essential splice site
159,455,514	A	G	Essential splice site
159,455,515	A	G	Essential splice site
159,455,516	A	G	Essential splice site
159,455,517	A	G	Essential splice site
159,455,518	A	G	Essential splice site
159,455,519	A	G	Essential splice site
159,455,520	A	G	Essential splice site
159,455,521	A	G	Essential splice site
159,455,522	A	G	Essential splice site
159,455,523	A	G	Essential splice site
159,455,524	A	G	Essential splice site
159,455,525	A	G	Essential splice site
159,455,526	A	G	Essential splice site
159,455,527	A	G	Essential splice site
159,455,528	A	G	Essential splice site
159,455,529	A	G	Essential splice site
159,455,530	A	G	Essential splice site
159,455,531	A	G	Essential splice site
159,455,532	A	G	Essential splice site
159,455,533	A	G	Essential splice site
159,455,534	A	G	Essential splice site
159,455,535	A	G	Essential splice site
159,455,536	A	G	Essential splice site
159,455,537	A	G	Essential splice site
159,455,538	A	G	Essential splice site
159,455,539	A	G	Essential splice site
159,455,540	A	G	Essential splice site
159,455,541	A	G	Essential splice site
159,455,542	A	G	Essential splice site
159,455,543	A	G	Essential splice site
159,455,544	A	G	Essential splice site
159,455,545	A	G	Essential splice site
159,455,546	A	G	Essential splice site
159,455,547	A	G	Essential splice site
159,455,548	A	G	Essential splice site
159,455,549	A	G	Essential splice site
159,455,550	A	G	Essential splice site
159,455,551	A	G	Essential splice site
159,455,552	A	G	Essential splice site
159,455,553	A	G	Essential splice site
159,455,554	A	G	Essential splice site
159,455,555	A	G	Essential splice site
159,455,556	A	G	Essential splice site
159,455,557	A	G	Essential splice site
159,455,558	A	G	Essential splice site
159,455,559	A	G	Essential splice site
159,455,560	A	G	Essential splice site
159,455,561	A	G	Essential splice site
159,455,562	A	G	Essential splice site
159,455,563	A	G	Essential splice site
159,455,564	A	G	Essential splice site
159,455,565	A	G	Essential splice site
159,455,566	A	G	Essential splice site
159,455,567	A	G	Essential splice site
159,455,568	A	G	Essential splice site
159,455,569	A	G	Essential splice site
159,455,570	A	G	Essential splice site
159,455,571	A	G	Essential splice site
159,455,572	A	G	Essential splice site
159,455,573	A	G	Essential splice site
159,455,574	A	G	Essential splice site
159,455,575	A	G	Essential splice site
159,455,576	A	G	Essential splice site
159,455,577	A	G	Essential splice site
159,455,578	A	G	Essential splice site
159,455,579	A	G	Essential splice site
159,455,580	A	G	Essential splice site
159,455,581	A	G	Essential splice site
159,455,582	A	G	Essential splice site
159,455,583	A	G	Essential splice site
159,455,584	A	G	Essential splice site
159,455,585	A	G	Essential splice site
159,455,586	A	G	Essential splice site
159,455,587	A	G	Essential splice site
159,455,588	A	G	Essential splice site
159,455,589	A	G	Essential splice site
159,455,590	A	G	Essential splice site
159,455,591	A	G	Essential splice site
159,455,592	A	G	Essential splice site
159,455,593	A	G	Essential splice site
159,455,594	A	G	Essential splice site
159,455,595	A	G	Essential splice site
159,455,596	A	G	Essential splice site
159,455,597	A	G	Essential splice site
159,455,598	A	G	Essential splice site
159,455,599	A	G	Essential splice site
159,455,600	A	G	Essential splice site

Non-synonymous coding variants

Position	Ref	Alt	Effect
159,455,500	A	G	Essential splice site
159,455,501	A	G	Essential splice site
159,455,502	A	G	Essential splice site
159,455,503	A	G	Essential splice site
159,455,504	A	G	Essential splice site
159,455,505	A	G	Essential splice site
159,455,506	A	G	Essential splice site
159,455,507	A	G	Essential splice site
159,455,508	A	G	Essential splice site
159,455,509	A	G	Essential splice site
159,455,510	A	G	Essential splice site
159,455,511	A	G	Essential splice site
159,455,512	A	G	Essential splice site
159,455,513	A	G	Essential splice site
159,455,514	A	G	Essential splice site
159,455,515	A	G	Essential splice site
159,455,516	A	G	Essential splice site
159,455,517	A	G	Essential splice site
159,455,518	A	G	Essential splice site
159,455,519	A	G	Essential splice site
159,455,520	A	G	Essential splice site
159,455,521	A	G	Essential splice site
159,455,522	A	G	Essential splice site
159,455,523	A	G	Essential splice site
159,455,524	A	G	Essential splice site
159,455,525	A	G	Essential splice site
159,455,526	A	G	Essential splice site
159,455,527	A	G	Essential splice site
159,455,528	A	G	Essential splice site
159,455,529	A	G	Essential splice site
159,455,530	A	G	Essential splice site
159,455,531	A	G	Essential splice site
159,455,532	A	G	Essential splice site
159,455,533	A	G	Essential splice site
159,455,534	A	G	Essential splice site
159,455,535	A	G	Essential splice site
159,455,536	A	G	Essential splice site
159,455,537	A	G	Essential splice site
159,455,538	A	G	Essential splice site
159,455,539	A	G	Essential splice site
159,455,540	A	G	Essential splice site
159,455,541	A	G	Essential splice site
159,455,542	A	G	Essential splice site
159,455,543	A	G	Essential splice site
159,455,544	A	G	Essential splice site
159,455,545	A	G	Essential splice site
159,455,546	A	G	Essential splice site
159,455,547	A	G	Essential splice site
159,455,548	A	G	Essential splice site
159,455,549	A	G	Essential splice site
159,455,550	A	G	Essential splice site
159,455,551	A	G	Essential splice site
159,455,552	A	G	Essential splice site
159,455,553	A	G	Essential splice site
159,455,554	A	G	Essential splice site
159,455,555	A	G	Essential splice site
159,455,556	A	G	Essential splice site
159,455,557	A	G	Essential splice site
159,455,558	A	G	Essential splice site
159,455,559	A	G	Essential splice site
159,455,560	A	G	Essential splice site
159,455,561	A	G	Essential splice site
159,455,562	A	G	Essential splice site
159,455,563	A	G	Essential splice site
159,455,564	A	G	Essential splice site
159,455,565	A	G	Essential splice site
159,455,566	A	G	Essential splice site
159,455,567	A	G	Essential splice site
159,455,568	A	G	Essential splice site
159,455,569	A	G	Essential splice site
159,455,570	A	G	Essential splice site
159,455,571	A	G	Essential splice site
159,455,572	A	G	Essential splice site
159,455,573	A	G	Essential splice site
159,455,574	A	G	Essential splice site
159,455,575	A	G	Essential splice site
159,455,576	A	G	Essential splice site
159,455,577	A	G	Essential splice site
159,455,578	A	G	Essential splice site
159,455,579	A	G	Essential splice site
159,455,580	A	G	Essential splice site
159,455,581	A	G	Essential splice site
159,455,582	A	G	Essential splice site
159,455,583	A	G	Essential splice site
159,455,584	A	G	Essential splice site
159,455,585	A	G	Essential splice site
159,455,586	A	G	Essential splice site
159,455,587	A	G	Essential splice site
159,455,588	A	G	Essential splice site
159,455,589	A	G	Essential splice site
159,455,590	A	G	Essential splice site
159,455,591	A	G	Essential splice site
159,455,592	A	G	Essential splice site
159,455,593	A	G	Essential splice site
159,455,594	A	G	Essential splice site
159,455,595	A	G	Essential splice site
159,455,596	A	G	Essential splice site
159,455,597	A	G	Essential splice site
159,455,598	A	G	Essential splice site
159,455,599	A	G	Essential splice site
159,455,600	A	G	Essential splice site

Ensembl transcripts (splice variants) are listed

Click a variation ID to open the variation tab

Possible amino acids at a position

Ensembl videos & tutorials

- Video: [Browsing SNPs and CNVs in Ensembl](http://ensembl.gramene.org/Help/Movie?id=208)
[<http://ensembl.gramene.org/Help/Movie?id=208>]
- Demo: [Structural variation for a region](http://ensembl.gramene.org/Help/Movie?id=316)
[<http://ensembl.gramene.org/Help/Movie?id=316>]
- Programmatic access tutorial: [Accessing variation data with the Variation API](http://ensembl.gramene.org/info/docs/api/variation/variation_tutorial.html)
[http://ensembl.gramene.org/info/docs/api/variation/variation_tutorial.html]