Genomic variation data in the Gramene Ensembl genome browser

www.gramene.org **plants.ensembl.org**

Marcela Karey Tello-Ruiz, PhD
Cold Spring Harbor Laboratory
March 29, 2016

Outline

- Variation data in Gramene
  - Types
  - Species
  - Databases (VEP - Sequence Ontology)
- Genome browser
  - Get a gene/transcript sequence showing variants & links
  - Get a table with all SNP variants within a gene/transcript
  - Visualize SVs, configure & export image
  - Download a sequence alignment showing conserved regions and variants
  - Explore variation: genomic context, genes & regulation, genotype frequency, sample genotypes, phenotypes, etc.
  - Upload a VCF
  - Tools: Variant Effect Predictor & Mart
- Stay in touch!
- Exercises

Variation Types in Gramene

- SNPs vs Structural Variants

[12 species] [Arabidopsis & Sorghum]

- Natural vs EMS-induced

[12 species] [Sorghum]

- Interhomoelogous variants (wheat A, B, D genomes, barley, brachy)

[Wheat]

Seq alignments for barley & brachy

Variation Data

445K SNPs in
16,718 lines

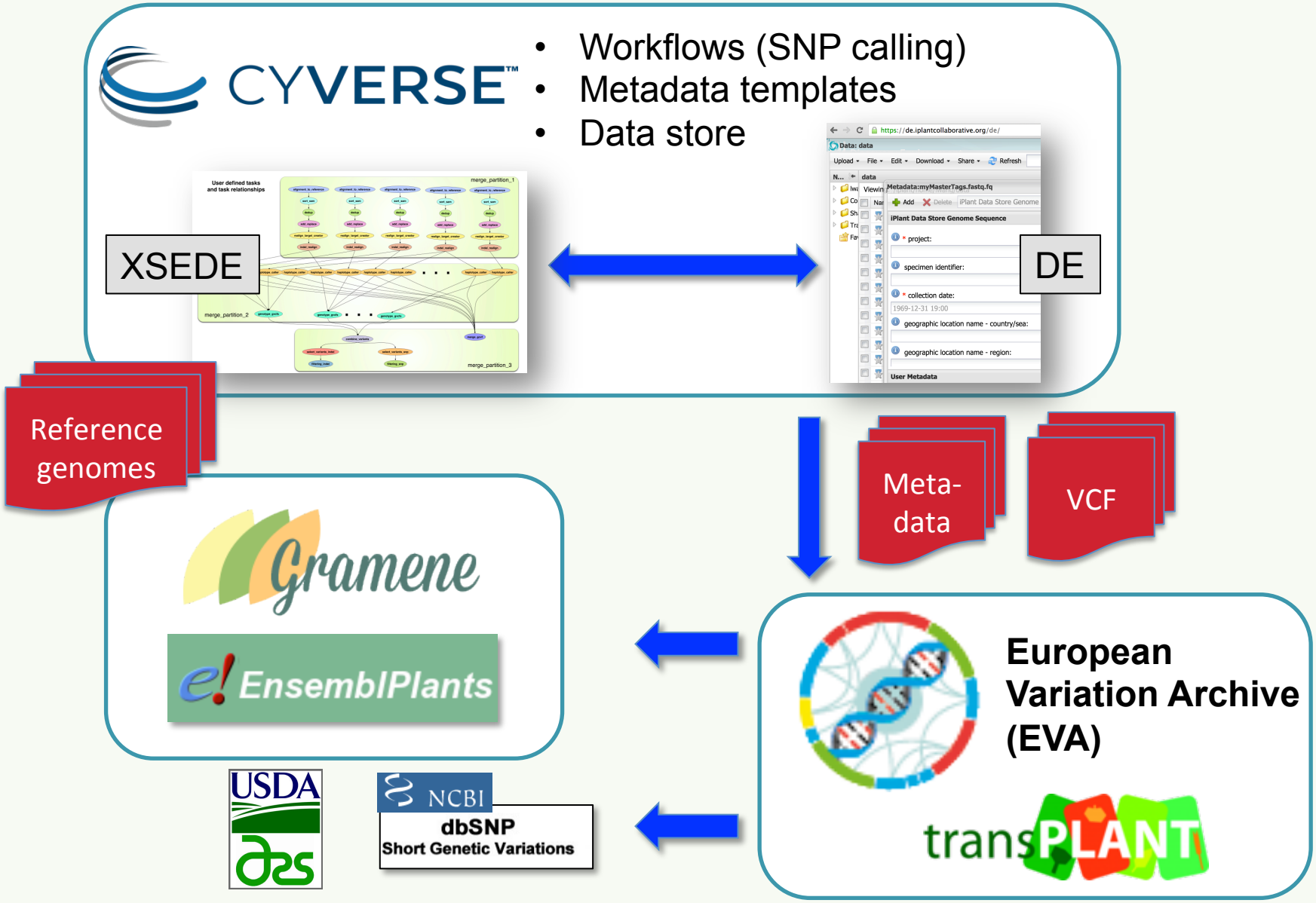
- **Arabidopsis** –1001 Genomes (3 studies), Horton *et al* , BB
  collaboration, Atwell *et al* (phenotypes)
- **Maize** – HapMap 1 & 2 (Gore *et* al & Chia *et al*), Panzea 2.7 GBS
- **Wheat** – HapMap (Jordan *et al*), CerealsDB, Inter-homeologous
- **Rice** [*Oryza sativa* Japonica & I *atula*, *O. glaberrima*, *O. barthii*] – dbSNP (Yu *et al*), M *l*, Duitama *et al* (b50) &
  3000 Genomes IRRI (coming

71M SNPs in
84 lines

- **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

Variation: tractable data & metadata exchange



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

Search | Like 386 | Follow @GrameneDatabase 84 followers

http://www.gramene.org

Home | Release Notes | Collaborators | Contact | About | Archive | Citing Gramene | Videotutorials | Outreach

Plant Variation in Gramene Webinar: Tues, Mar 29, 2016 @ 2:00 pm EDT

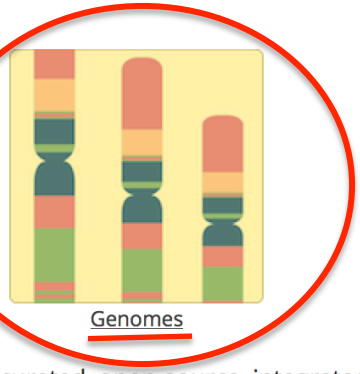
Navigation

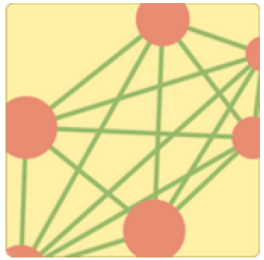
- Current Release (49)
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Mailing list
- Related genome browsers
- Tools

Recent blog posts

- Plant Variation in Gramene Webinar: Tues, Mar 29, 2016 @ 2:00 pm EDT
- Gramene Workshop at the Maize Genetics Conference 2016
- Gramene book chapter in the Plant Bioinformatics volume

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**

http://plantsembl.org

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
082214v1

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).

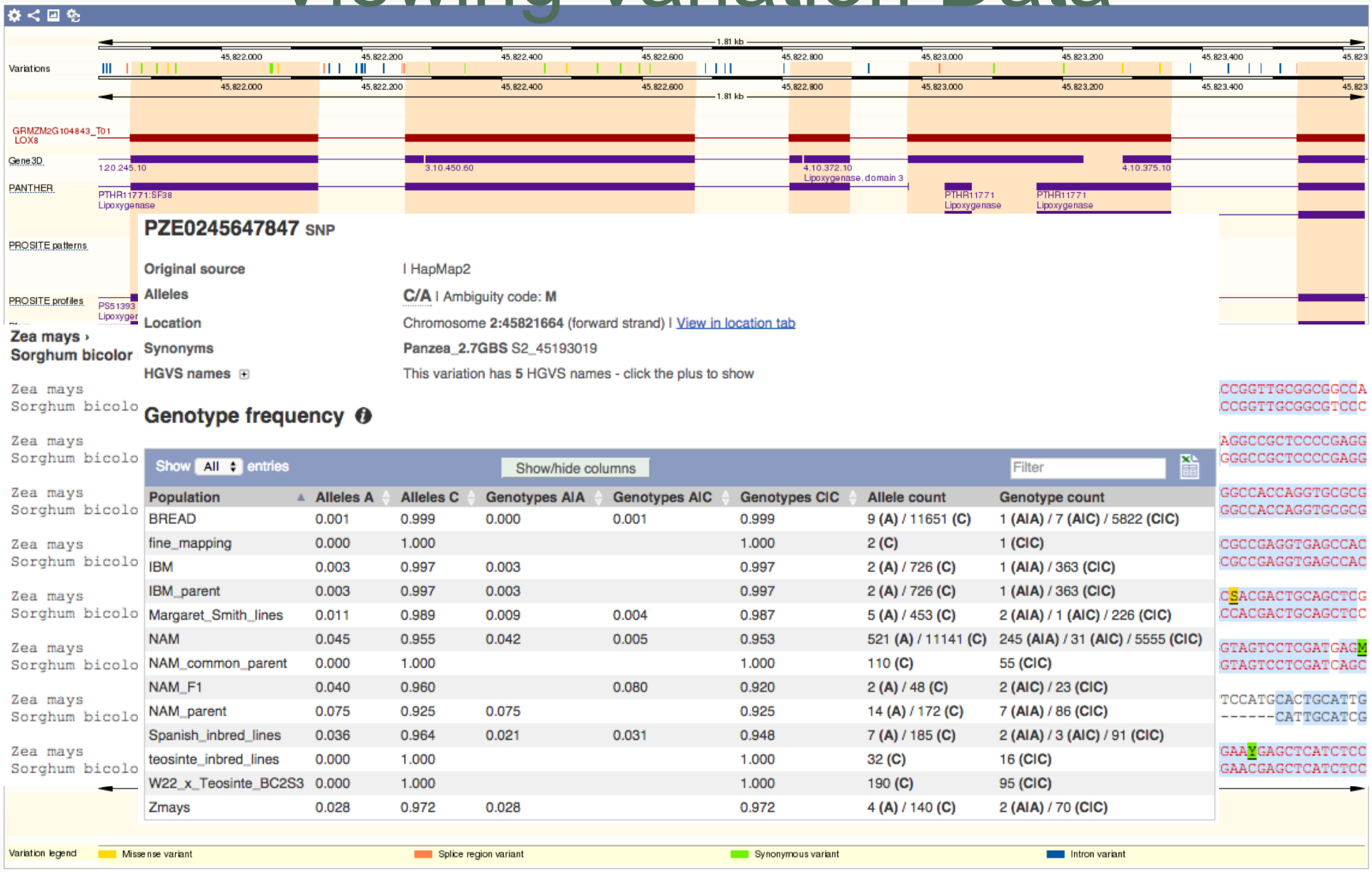
The resulting set of chromosomes have been combined with the IWGSC's BAC by BAC assembly and annotation of chromosome 3B, generated by the GDEC group at INRA (Paux et al, 2008), which replaces the CSS-derived assembly and annotation of 3B.

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-Amatriaín 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, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S

Viewing Variation Data



Population	Alleles A	Alleles C	Genotypes AIA	Genotypes AIC	Genotypes CIC	Allele count	Genotype count
BREAD	0.001	0.999	0.000	0.001	0.999	9 (A) / 11651 (C)	1 (AIA) / 7 (AIC) / 5822 (CIC)
fine_mapping	0.000	1.000			1.000	2 (C)	1 (CIC)
IBM	0.003	0.997	0.003		0.997	2 (A) / 726 (C)	1 (AIA) / 363 (CIC)
IBM_parent	0.003	0.997	0.003		0.997	2 (A) / 726 (C)	1 (AIA) / 363 (CIC)
Margaret_Smith_lines	0.011	0.989	0.009	0.004	0.987	5 (A) / 453 (C)	2 (AIA) / 1 (AIC) / 226 (CIC)
NAM	0.045	0.955	0.042	0.005	0.953	521 (A) / 11141 (C)	245 (AIA) / 31 (AIC) / 5555 (CIC)
NAM_common_parent	0.000	1.000			1.000	110 (C)	55 (CIC)
NAM_F1	0.040	0.960		0.080	0.920	2 (A) / 48 (C)	2 (AIC) / 23 (CIC)
NAM_parent	0.075	0.925	0.075		0.925	14 (A) / 172 (C)	7 (AIA) / 86 (CIC)
Spanish_inbred_lines	0.036	0.964	0.021	0.031	0.948	7 (A) / 185 (C)	2 (AIA) / 3 (AIC) / 91 (CIC)
teosinte_inbred_lines	0.000	1.000			1.000	32 (C)	16 (CIC)
W22_x_Teosinte_BC2S3	0.000	1.000			1.000	190 (C)	95 (CIC)
Zmays	0.028	0.972	0.028		0.972	4 (A) / 140 (C)	2 (AIA) / 70 (CIC)

Gene Example: *cle18*

nature genetics

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

Whole-genome sequencing of multiple *Arabidopsis thaliana* populations

Jun Cao[1,8], Korbinian Schneeberger[1,2,8], Stephan Ossowski[1,3,4,8], Torsten Günther[5,8], Sebastian Bender[1], Joffrey Fitz[1], Daniel Koenig[1], Christa Lanz[1], Oliver Stegle[6], Christoph Lippert[6], Xi Wang[1], Felix Ott[1], Jonas Müller[1], Carlos Alonso-Blanco[7], Karsten Borgwardt[6], Karl J Schmid[5] & Detlef Weigel[1]

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.

2 VEPs in CLE18 (Cao et al, 2011)

SNPs with predicted drastic effects on gene function.

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

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
082214v1

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-Amatriaín 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, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S

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: Select a species...

CLE18 [AT1G66145]

Description	CLAVATA3/ESR (CLE)-related protein 18 [Source:UniProtKB/Swiss-Prot;Acc:Q3ECH
Gene ID	AT1G66145
Species	*Arabidopsis thaliana*
Location	1:24628530-24628835
Synonyms	CLAVATA3/ESR RELATED 18
Gene trees	EPlGT00140000030760 (Plant Compara)

CLE18 [scaffold_201322.1]

Description	Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:D7KTK6]
Gene ID	scaffold_201322.1
Species	*Arabidopsis lyrata*
Location	2:10746223-10746565
Gene trees	EPlGT00140000030760 (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	EPlGT00180000045708 (Plant Compara)

Gramene release 49 - 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
- 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
- Pan-taxonomic Compara
  - Gene Tree
  - Orthologues
- Phenotype
- Genetic Variation
  - Variant table
  - Structural variants
  - Variant image
- External data
  - Gene expression
  - Personal annotation
- ID History
  - Gene 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:Q3ECH9]

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

Show/hide columns (1 hidden) Filter

Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags
CLE18	AT1G66145.1	306	101aa	Protein coding	Q3ECH9	NM_148631 NP_683472	

Summary

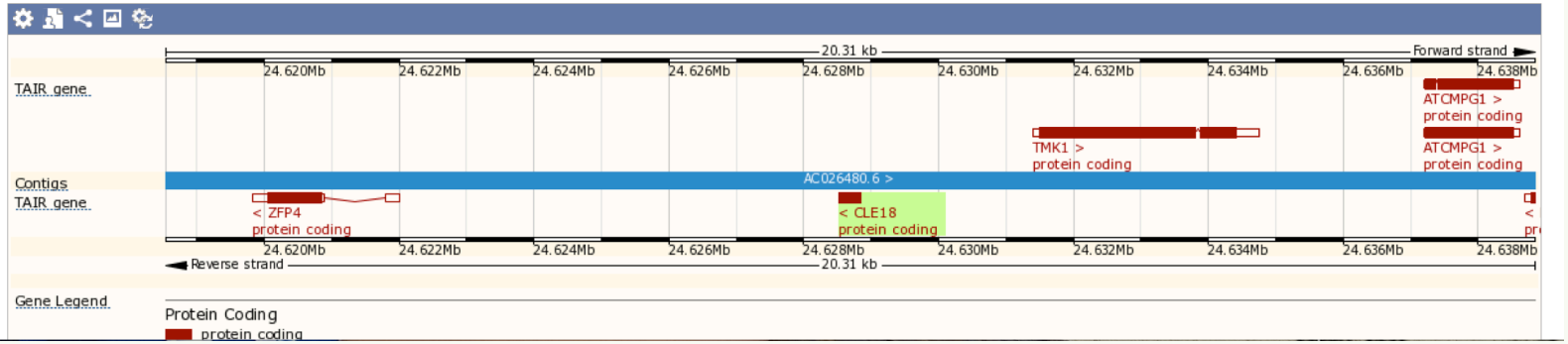
Name CLE18 (UniProtKB Gene Name)

UniProtKB This gene has proteins that correspond to the following UniProtKB identifiers: Q3ECH9

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 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
- Pan-taxonomic Compara
  - Gene Tree
  - Orthologues
- Phenotype
- Genetic Variation
  - Variant table
  - Structural variants
  - Variant image
- External data
  - Gene expression
  - Personal annotation
- ID History
  - Gene 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:UniProtK
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

```
>chromosome:TAIR10:1:24627930:24629435:-1
TAACAAGAGATTTATGTTCGAAACCAATCTGTCACAATATTAGATCCATCGCAAAAATAT
TTCCGTAAAGTGATATTCCACAAAATTGCTTCACAAAGGCTAAATTTTCAAATCATTGGG
TTTATTTTCATGAAAATATGAAATTCTATAAATGTTTTATGATTTAGAAAATGTCTACCA
TATATGAACGAACAAATTAACTACATAGATGAGATTTAAATAAAATTTTCAAAAATGAGA
TTTTGTTACCATGAAATAAAGAATAGCCTTATGAAGATTTCAAAACGAAAAATTGAAACT
AAATTTAGTCGCAACACAATTATTTTATGGTCAAATAAGTAAAAACAAATTAAAAATGAG
ATTACCATGAAATAATTCGGGAAAACTTTTTTTTTTTGGGTCAAACAATTGGGGTTCAGT
TAAAAGGGTCAAGTACATATCTTCTAATATACTCAAGTCAATTGCAATTGACATTTGATT
AGCTTCCCTTAATTAAGATTAAAATTCCACTATAAAAATAAAAAGGAGGAAACAAGAGTG
TAGCCAACTAAAAAAATCTATACAAAAACAATCAAAAAGCAAAAAATAAGGATCAGAAAG
ATGCATTTGTTAAAAGGTGGTGTGGTATTAATCATAACGCTAATCTTGTTCCTAATAACG
TCGTCTATTGTTGCTATTCGTGAAGACCCATCGTTGATTGGAGTAGACAGACAGATCCCA
ACAGGTCCTGATCCTTTACATAATCCTCCTCAGCCTTCACCAAAACATCATCATTGGATC
GGAGTCGAGGAAAACAATATCGACCGGTCTTGGAATTATGTGGATTATGAATCTCATCAT
GCACATAGTCCTATACATAACTCTCCTGAGCCTGCACCATTATATCGTCATTTGATTGGA
GTCTAGAACAAATATATATATATATATATATTTAATCATCTACATGCAGATATATTTTTT
AGTATGAGAATGAGTTTGGAGTGGTAAAACCGTAATATGTAGATTCCATTAGTTTGCGTA
TACAGATATAGTTTGTATTAGAGATTGACATAGATAATCGAAAGCTTTGTTAGCAGAAAA
AAAAACAGATTGATTATGACCTTATAAACTATTTAAATTCGAATGCAAAAATATGCACAC
TTCTTCTACAATTAAAGAAAAATGGGCCTTGGGTCAATATGTATGCAAACAACATAAGTG
CACACACAAAAAACCAATGCTAATAGAGTAATCGCCATTGATCAATTGCATGCAAGCACA
TATATATGGATTTCATTTTTATTTTTTTTGTAGGATACTTAGTTGGAAAATGATGATCAA
AGTTGTGATTCTATGTTTACAAAACCACTTTTGATTAGAAAGTCTGGTCATGACTCATGA
GCCCAACGAGTTGAAATATGGGCTTAGGGTCCCTAAAATGTTTAAGCTCATCTCATATGA
AAATAGATCCCCAAACATAACTAGGTCCTCAATTTTTTCAATCTTTAAAATATGCTCTCA
TTTTCA
```

Configure (Sequence) page: Show variants

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Login/R

Search Arabidopsis thaliana...

Configure Page | Personal Data

Display options

Save configuration as...

Load configuration

Reset configuration

Display options

5' Flanking sequence (upstream):	600 * (Maximum of 1000000)
3' Flanking sequence (downstream):	600 * (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
Hide variants longer than 10bp:	✓
Hide variants by frequency (MAF):	Don't hide
Filter variants by consequence type:	No filter 3 prime UTR variant 5 prime UTR variant Coding sequence variant Downstream gene variant
Line numbering:	Relative to this sequence
Display pop-up information on mouseover:	Yes

Fields marked * are required

1

2

3

Export data

Share this page

Bookmark this page

TATATATGGATTTCATTTTTATTTTTTTTGTAGGATACTTAGTTGGAAAATGATGATCAA
AGTTGTGATTCTATGTTTACAAAACCACTTTTGATTAGAAAGTCTGGTCATGACTCATGA
GCCCAACGAGTTGAAATATGGGCTTAGGGTCCCTAAAATGTTTAAGCTCATCTCATATGA
AAATAGATCCCCAAACATAACTAGGTCCTCAATTTTTTCAATCTTTAAAATATGCTCTCA
TTTTCA

Mouse hover a variant to reveal details

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
- Pan-taxonomic Compara
  - Gene Tree
  - Orthologues
- Phenotype
- Genetic Variation
  - Variant table
  - Structural variants
  - Variant image
- 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

Gene: CLE18 AT1G66145

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

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

Exons CLE18 exons | All exons in this region

Variants Coding sequence | Flagged variant | Missense | Stop gained | Synonymous | Upstream

```
>chromosome:TAIR10:1:24627930:24629435:-1
   1 TAACAARAGATTKATGTTCGAWACCAATCTGTCACAATATTAGATCCATCGCAAAAATAT   60 7: ENSVATH13741852; 13: ENSVATH05076765; 22: ENSVATH05076764;
  61 TTCCGTAAAGTGATATTCCACAAAATTGCTTCACAAWGGCTAAHTTTTCRRATCATTGGG  120 71: ENSVATH01479132; 97: ENSVATH05076763; 104: ENSVATH05076762; 110: ENSVATH05076761
 121 TTTATTTTCATGAAAATATGAAAKTCTATAAATGTTTTATGATTTAGAAAATGTCTAYCA  180 144: ENSVATH01479131; 178: ENSVATH13741850;
 181 TATATGAASGAACAARTTAACTACATAGATGAGATTTAAAKAAAAWTTTCAAAATGAGA  240 189: ENSVATH05076760; 196: ENSVATH13741849; 221: ENSVATH13741848; 226: ENSVATH014791
 241 TTYWGTTACCATRAAATAAAGAATAGCCTTATGAAGATTTCAAAACGAAAAATTGAAACT  300 243: ENSVATH13741847; 244: ENSVATH05076759; 253: ENSVATH13741846;
 301 AAATTTAGTCGCAACACAATTATTTWATGGTMAARTMAGTAAAAACAAATTAAAAATGAG  360 326: ENSVATH07680648; 332: ENSVATH13741845; 335: ENSVATH13741844; 337: ENSVATH050767
 361 ATTACCAYGAAATAATTCGGGAAAACTWTTTTTTTYKGGTCAAACAAKTKGGGTTCAGT  420 368: ENSVATH13741843; 388: ENSVATH13741842; 390: ENSVATH01479129; 397: ENSVATH050767
 421 TAAAAGGGTCAAGYAMAWAYYTTYTAATATACTCAARTYWATTGCAWTTGACATTTGATT  480 434: ENSVATH05076755; 436: ENSVATH13741840; 438: ENSVATH13741839; 440: ENSVATH137418
 481 AGCTTCCCTTAATTAAGATTAAAAYTCCASTRTAMAAATAAAAAGGAGGAAMCAAGAGTG  540 505: ENSVATH05076753; 510: ENSVATH13741832; 512: ENSVATH13741831; 515: ENSVATH050767
 541 TAGSCAACTAARAAAATCTRTACAAAAACAATCAAAARGYAAAAAWTAAGGAWCARAAAG  600 544: ENSVATH05076750; 552: ENSVATH05076749; 560: ENSVATH14426198; 578: ENSVATH137418
 601 ATGCATTTGTYAAAAGGTGGTGTGKTATTAATCATAACGCTAAYCTTKTTCCTAATRACG  660 611: ENSVATH13741827; 625: ENSVATH05076747; 637: ENSVATH00125660; 644: ENSVATH050767
 661 TCGTCWAWTGTTGCTATTCGTGAAGAYCCATCGTTGATWGGAGTMGACAGACAGATCCCA  720 666: ENSVATH05076743; 668: ENSVATH05076742; 687: ENSVATH05076741; 699: ENSVATH050767
 721 ACAGGTCCTGATCCTTYACAYAATCCTCCTYAGCCTTCACCAAAACATCATCATTGKATC  780 737: ENSVATH00125659; 741: ENSVATH05076738; 751: ENSVATH05076737; 777: ENSVATH001256
 781 GRAGTCGAGGAAAACAATATCGRCCGGTCTTGGAATTATGTKGATTATGAATCTCAYMAT  840 782: ENSVATH05076736; 803: ENSVATH01479127; 822: ENSVATH14426196; 837: ENSVATH050767
 841 GCACATAGTCCTATWCATAACTCTCCTGAGC AT1G66145.1.exon1 TCATTTSATTGGA  900 855: ENSVATH13741825; 872: ENSVATH00125657; 887: ENSVATH13741824; 894: ENSVATH137418
 901 GTCTAGAACWAATRWMTATATWTATATATA C/T AGATATATWTKTT  960 910: ENSVATH13741822; 914: ENSVATH13741821; 915: ENSVATH13741820; 915: ENSVATH155219
 961 AKTATGAGAATKAGTTTGGAGTGGTAAAACCGTAATATGTAGATTCCATTAGTTTGCGTA 1020 962: ENSVATH13741813; 970: ENSVATH04071268; 970: ENSVATH08710284; 972: ENSVATH137418
1021 TACAGATATAGTTTGTATTAGAGMTTGACATAGATAATCGAAAGCWTTGWWAGCAGAAAA 1080 1044: ENSVATH14426174; 1066: ENSVATH13741811; 1068: ENSVATH00125656; 1070: ENSVATH13
1081 AAAAACAGATTGATTATGACCTTATAAAYTATWKAAAWTCSAATGCAAAAATATGCACAC 1140 1083: ENSVATH08710283; 1085: ENSVATH04071267; 1085: ENSVATH08710282; 1109: ENSVATH13
1141 TTCTTCTHCAATTAAAGAAAAATGGGCCTTGGGTCAATATKTATGCAAACAACATAAGTG 1200 1148: ENSVATH05076730; 1181: ENSVATH08710281; 1182: ENSVATH04071266;
1201 CACACACAAAAAACCARTGCYAATAGAGTAATCGCMATTGATCAATTGCATGCAAGCACA 1260 1217: ENSVATH05076729; 1221: ENSVATH05076728; 1236: ENSVATH13741807;
1261 WATATATGGATTTCATTTTTAWTTTTTTKGWARRATRCWTAGTTGRAAAATGATGATCAA 1320 1261: ENSVATH13741806; 1281: ENSVATH07680646; 1282: ENSVATH07680645; 1289: ENSVATH13
1321 AGTTGTGATTCYATGTTTACAAAACCACKTTTGATTAGAAAGKCTGGTCATGMCTCATGA 1380 1332: ENSVATH05076725; 1349: ENSVATH01479125; 1363: ENSVATH05076724; 1373: ENSVATH00
1381 GCYCAACGARTTGAAATATGGGCTTAGGGTCCCTMAAWTKTTTAWGCTCATCTCATATGA 1440 1383: ENSVATH13741800; 1390: ENSVATH05076723; 1415: ENSVATH05076722; 1418: ENSVATH05
1441 AAATAGATCCCCAAACATAACTAGGTCCTCAATTTTTTCAATCTTTAAAATATGCTCTCA 1500
```

ensembl.gramene.org/Arabidopsis_thaliana/ZMenu/TextSequence?db=core;factorytype=Location;g=AT1G66145;r=1:24628530-24628835;t=AT1G66145.1;vf=5429980

Variant table (Gene & Transcript)

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
- 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
- Pan-taxonomic Compara
  - Gene Tree
  - Orthologues
- Phenotype
- Genetic Variation
  - Variant table
  - Structural variants
  - Variant image
- External data
  - Gene expression
  - Personal annotation
- ID History
  - Gene 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:Q3ECH9]

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

Variant table

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

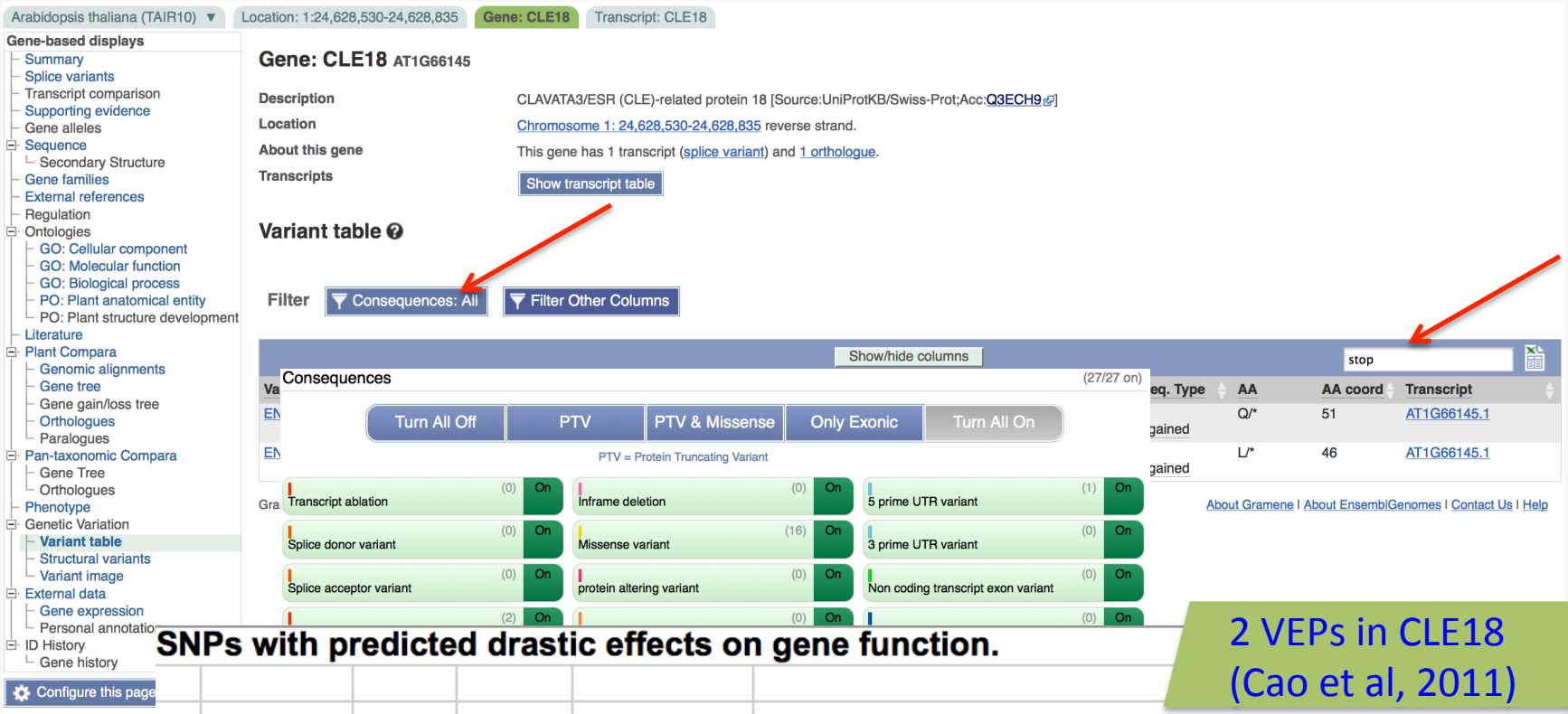
Filter: Consequences: All | Filter Other Columns

Show/hide columns: Variant ID, Chr: bp, Alleles, Class, Source, Evidence, Clin. Sig., Conseq. Type, AA, AA coord, Transcript

Search...

Variant ID	Chr: bp	Alleles	Class	Source	Conseq. Type	AA	AA coord	Transcript
ENSVATH05076606	1:24623536	T/G	SNP	Ensembl	Downstream gene variant	-	-	AT1G66145.1
ENSVATH05076607	1:24623549	C/G	SNP	Ensembl	Downstream gene variant	-	-	AT1G66145.1
ENSVATH13741706	1:24623552	C/G	SNP	Ensembl	Downstream gene variant	-	-	AT1G66145.1
ENSVATH01479067	1:24623558	G/A	SNP	Ensembl	Downstream gene variant	-	-	AT1G66145.1
ENSVATH05076608	1:24623581	A/C	SNP	Ensembl	Downstream gene variant	-	-	AT1G66145.1
ENSVATH05076609	1:24623600	T/A	SNP	Ensembl	Downstream gene variant	-	-	AT1G66145.1
ENSVATH01479068	1:24623609	A/T	SNP	Ensembl	Downstream gene variant	-	-	AT1G66145.1

Filter consequences

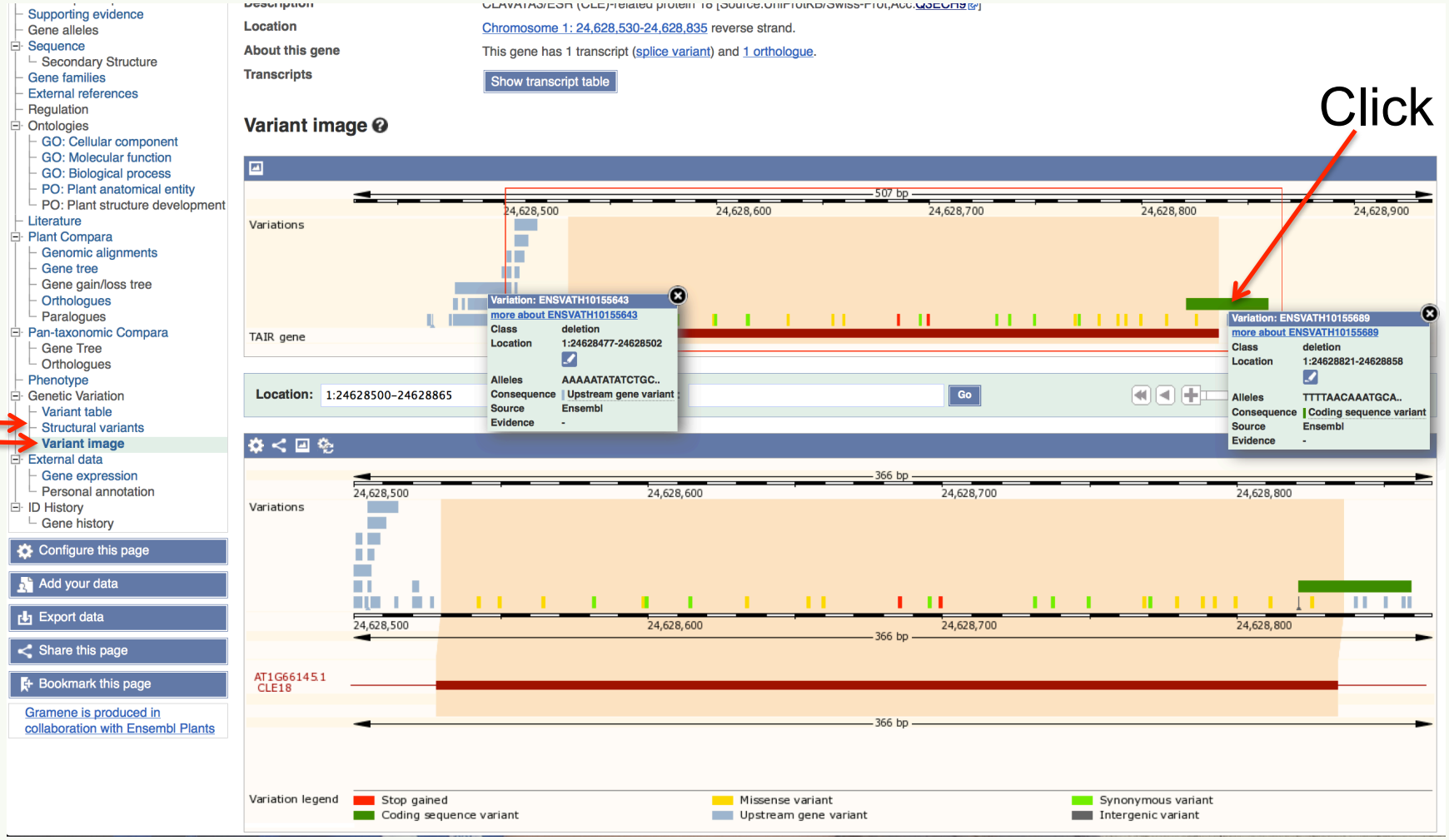


SNPs with predicted drastic effects on gene function.

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

2 VEPs in CLE18
(Cao et al, 2011)

Structural variants



Go to Transcript page

Login/Registe

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Search Arabidopsis thaliana…

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

Transcript-based displays

- Summary
- Supporting evidence
- Sequence
  - Exons
  - cDNA
  - Protein
- External 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

Gramene is produced in collaboration with Ensembl Plants

Transcript: CLE18 AT1G66145.1

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

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

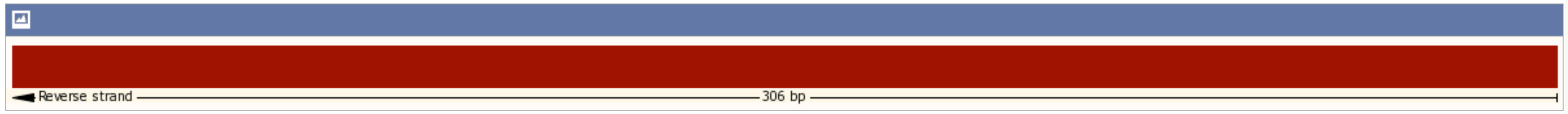
About this transcript This transcript has 1 exon, is annotated with 3 domains and features, is associated with 24 variations and maps to 1 oligo probe.

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	NM_148631 NP_683472	

Summary



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

Uniprot This transcript corresponds to the following Uniprot identifiers: Q3ECH9

Type Known protein coding

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

Gramene release 49 - Jan 2016 © EBI

About Gramene | About EnsemblGenomes | Contact Us | Help

Transcript Sequence: Exons

Login/Register

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Search Arabidopsis thaliana...

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

Transcript-based displays

- Summary
- Supporting evidence
- Sequence
  - Exons
  - cDNA
  - Protein
- External 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

Gramene is produced in collaboration with Ensembl Plants

Transcript: CLE18 AT1G66145.1

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

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	NM_148631 NP_683472	

Exons

Download sequence

Exons/ Introns: Translated sequence, Flanking sequence

Variants: Coding sequence, Missense, Stop gained, Synonymous

Show/hide columns Filter

No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence						.....aaaaaatctatacaaaaacaatcaaaaagcaaaaaataaggatcagaaag
1	AT1G66145.1.exon1	24,628,835	24,628,530	0	0	306	ATGCATTTGTTAAAAGGTGGTGTGGTATTAATCATAACGCTAATCTTGTTCCTAATAACG TCGTCTATTGTTGCTATTCGTGAAGACCCATCGTTGATTGGAGTAGACAGACAGATCCCA ACAGGTCCTGATCCTTTACATAATCCTCCTCAGCCTTCACCAAAACATCATCATTGGATC GGAGTCGAGGAAAACAATATCGACCGGTCTTGGAATTATGTGGATTATGAATCTCATCAT GCACATAGTCCTATACATAACTCTCCTGAGCCTGCACCATTATATCGTCATTTGATTGGA GTCTAG
	3' downstream sequence						aacaaatatatatatatatatatatttaatcatctacatgcagatatatt.....

Transcript Sequence: cDNA

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

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

Transcript-based displays

- Summary
- Supporting evidence
- Sequence
  - Exons
  - cDNA
  - Protein
- External 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

1

2

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

Transcript: CLE18 AT1G66145.1

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

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	NM_148631 NP_683472	

cDNA sequence

Download sequence | BLAST this sequence

Codons	Alternating codons	Alternating codons	
Exons	Alternating exons	Alternating exons	
Variants	Missense	Stop gained	Synonymous

```
                Y             K          *        Y   K         R
  1 ATGCATTTGTTAAAAGGTGGTGTGGTATTAATCATAACGCTAATCTTGTTCCTAATAAC
  1 ATGCATTTGTTAAAAGGTGGTGTGGTATTAATCATAACGCTAATCTTGTTCCTAATAAC
  1 -M--H--L--L--K--G--G--V--V--L--I--I--T--L--I--L--F--L--I--T

        W W                Y            W      M
 61 TCGTCTATTGTTGCTATTCGTGAAGACCCATCGTTGATTGGAGTAGACAGACAGATCCC
 61 TCGTCTATTGTTGCTATTCGTGAAGACCCATCGTTGATTGGAGTAGACAGACAGATCCC
 21 -S--S--I--V--A--I--R--E--D--P--S--L--I--G--V--D--R--Q--I--I

                W   Y         Y                          K
121 ACAGGTCCTGATCCTTTACATAATCCTCCTCAGCCTTCACCAAAACATCATCATTGGATC
121 ACAGGTCCTGATCCTTTACATAATCCTCCTCAGCCTTCACCAAAACATCATCATTGGATC
 41 -T--G--P--D--P--L--H--N--P--P--Q--P--S--P--K--H--H--H--W--I-

     R                R                K                YM
181 GGAGTCGAGGAAAACAATATCGACCGGTCTTGGAATTATGTGGATTATGAATCTCATCAT
```

```
Y                                                    K
TCAGCCTTCACCAAAACATCATCATTGG
TC
--                                                   -W-

TT                                                   CAT
TT                                                   CAT
--                                                   -H-

GC                                                   ATT
GC                                                   ATT
--P--A--P--L--Y--R--H--L--I-
```

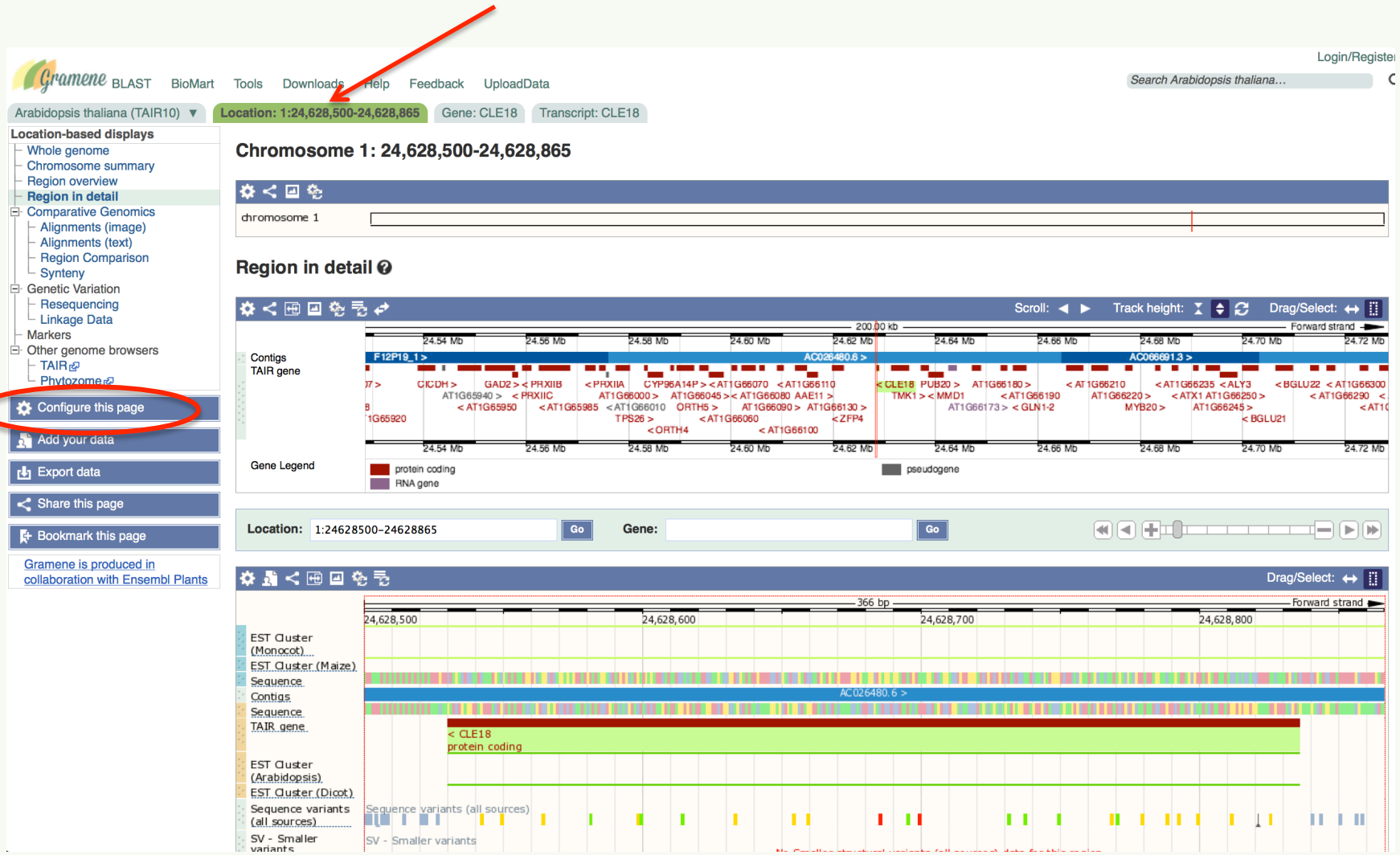
Variation: ENSVATH05076737

Position	1:24628685
Alleles	G/A
cDNA position	151
Protein position	51
Amino acids	Q/*
Codons	Cag/Tag
Consequences	Stop gained

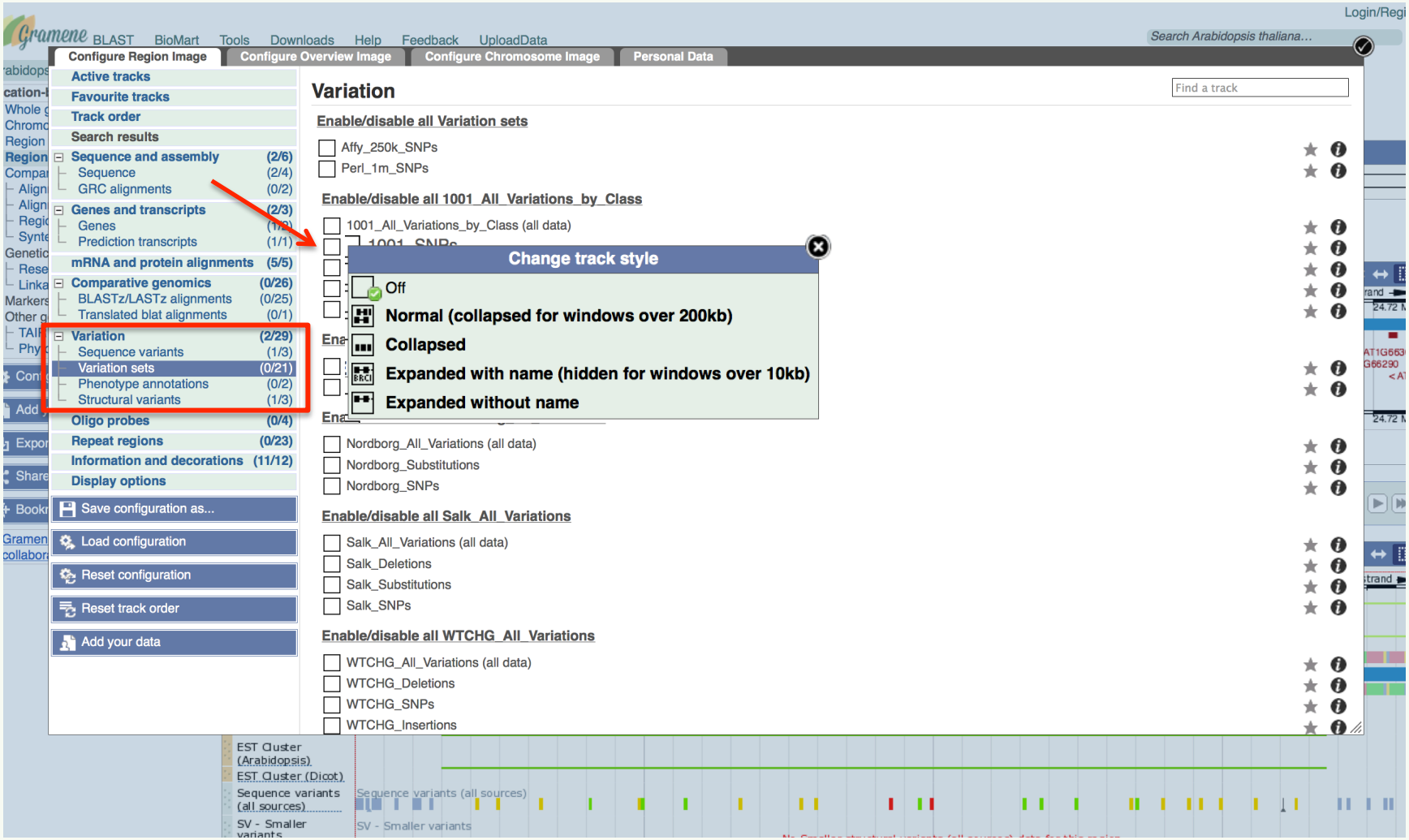
Explore this variant

Gene/Transcript Locations

Go to location page



Configure (location) page: Show variation



Location: show variants for a sequence alignment

1

Arabidopsis thaliana (TAIR10) ▼ | Location: 1:24,628,469-24,628,896 | Gene: CLE18 | Transcript: CLE18

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

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

3

Chromosome 1: 24,628,469-24,628,896

chromosome 1

Alignments (text)

2

Alignment: Arabidopsis lyrata - blastz | Go

-- Select an alignment --
Pairwise alignments
- Amborella trichopoda - translated blat
- Arabidopsis lyrata - blastz
- Brachypodium distachyon - blastz
- Brassica oleracea - lastz
- Brassica rapa - lastz
- Chlamydomonas reinhardtii - blastz
- Cyanidioschyzon merolae - lastz
- Glycine max - blastz
- Medicago truncatula str. A17 - lastz
- Musa acuminata - lastz
- Oryza brachyantha - lastz
- Oryza glaberrima - blastz
- Oryza longistaminata - lastz
- Oryza meridionalis - lastz
- Oryza sativa Indica - blastz
- Oryza sativa Japonica - lastz
- Physcomitrella patens - blastz
- Populus trichocarpa - lastz

Location: | Gene: | Go

Species

No tree is c

Arabidopsis 396:1
Arabidopsis

```
Arabidopsis thaliana  TCATACTAAAAAATATATCTGCATGTAGATGATTAAATATATATATATATATATATTTGTTCTAGACTCCAATCAAATGACGATATAATGGTGCAGGCTCAGGAGAGTTATGTATAGGAC
Arabidopsis lyrata    TCATTTTTAAAAA-ATATCTGCATGTAGATGATTAAAAATATAAATCTATATTTTTTTTTCTAGACTCCAATCAAATGACGATATAATGGTGAAGGCTGAGGAGCACTATGTATAGGTT

Arabidopsis thaliana  TATGT--GCATGATGAGATTCATAATCCACATAATTCCAAGACCGGTCGATATTGTTTTCCTCGACTCCGATCCAATGATGATGTTTTGGTGAAGGCTGAGGAGGATTATG---------
Arabidopsis lyrata    TATGTAGGGATGAGGAGAATCATAATCCACATAATTCCAAGACCGTTCGATATTTTTTTCCTCGACTCCGATCCAATGACGATGTCCTGGTGAAGGCTGATGAGGATTATGATGAGGCGC

Arabidopsis thaliana  ------TAAAGGATCAGGACCTGTTGGGATCTGTCTGTCTACTCCAATCAACGATGGGTCTTCACGAATAGCAACAATAGACGACGTTATTAGGAACAAGATTAGCGTTATGATTAATAC
Arabidopsis lyrata    ATTATGTAAAAGATCCGGACCTGTTGGGATCTTTCTCTCGACTCCAATCAAAGATGGGACTTCAAGAATAGCAGCCACCGACGACGTTATTAGGAAAAAGATTAGTGTATTGATTAGTGC

Arabidopsis thaliana  CACA-----------------------------------------------CCACCTTTTAA----------CAAA--------TGCATCTTTCTG--------------ATCCTTATT
Arabidopsis lyrata    TAAAATTAGGCCTCAAAAATAATTTTTATTATAAAAATAGACCTCATTTTTTCTAACTTTTAAGATTTAGTCTCAAATTTTGAATTTTATCTATTTGTTATTGCAAAAATAAGCCTTATT

Arabidopsis thaliana  TTTTG--------------------CTTTTTGA--------------------------------TTGTTTTTGTATAGA--TTTTTTTAGTTGGCTAC
Arabidopsis lyrata    TTCTGAAAATTTAAGAGTTGGTCTCATTTTTAAATTTTATCTACAAAATCCATTAAAAAAAATTCTTGTTTCTTTGTAAACTTTTTTTTAGTTGACT--
```

Gramene release 49 - Jan 2016 © EBI

About Gramene | About EnsemblGenomes | Contact Us | Help

ensembl.gramene.org/Arabidopsis_thaliana/Location/Compara_Alignments?align=8654;db=core;g=AT1G66145;r=1:24628469-24628896;t=AT1G66145.1

Configure (Alignments) page

BLAST BioMart Tools Downloads Help Feedback UploadData

Search Arabidopsi

Configure Alignments | Configure Chromosome Image | Personal Data

Display options

Save configuration as...

Load configuration

Reset configuration

Display options

Number of base pairs per row:	120 bps
Strand:	Forward
Additional exons to display:	Core exons
Orientation of additional exons:	Display exons in both orientation
Show variants:	Yes and show links
Hide variants longer than 10bp:	☑
Hide variants by frequency (MAF):	Don't hide
Filter variants by consequence type:	No filter 3 prime UTR variant 5 prime UTR variant Coding sequence variant Downstream gene variant
Line numbering:	Relative to this sequence
Codons:	Do not show codons
Show conservation regions:	☑
Mark alignment start/end:	☐

```
Arabidopsis thaliana     481 TTTTR-------------------CYTTTTGA-------------------------------TTGTTTTTGTAYAGA--TTTTYTTAGTTGSCTAC
Arabidopsis lyrata       481 TTCTGAAAATTTAAGAGTTGGTCTCATTTTTAAATTTTATCTACAAAATCCATTAAAAAAAATTCTTGTTTCTTTGTAAACTTTTTTTAGTTGACT--
```

Sequence alignment with variants

Arabidopsis thaliana (TAIR10) ▼ | Location: 1:24,628,469-24,628,896 | Gene: CLE18 | Transcript: CLE18

Location-based displays

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

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

Chromosome 1: 24,628,469-24,628,896

chromosome 1

Alignments (text)

Alignment: Arabidopsis lyrata - blastz Go

Location: 1:24628469-24628896 Go **Gene:** Go

Download alignment

Species Tree

No tree is drawn for pairwise alignments

Arabidopsis thaliana › chromosome:TAIR10:1:24628469:24628896:1
Arabidopsis lyrata › chromosome:v.1.0:2:10733693:10733740:1
chromosome:v.1.0:2:10746220:10746523:1
chromosome:v.1.0:2:10746587:10746810:1

```
Arabidopsis thaliana      1 TCATAMTAAMAWATATATCTGCATGTAGATRAWWAAATATATATAWATATAKWYATTWGTTCTAGACTCCAATSAAATGAYGATATAATGGTGCASGCTCAGGAGAGTTATGWATAGGAC   120
Arabidopsis lyrata        1 TCATTTTTAAAAA-ATATCTGCATGTAGATGATTAAAAATATAAATCTATATTTTTTTTTCTAGACTCCAATCAAATGACGATATAATGGTGAAGGCTGAGGAGCACTATGTATAGGTT   120

Arabidopsis thaliana    121 TATGT--GCATKRTGAGATTCATAATCMACATAATTCCAAGACCGGYCGATATTGTTTTCCTCGACTYCGATMCAATGATGATGTTTTGGTGAAGGCTRAGGAGGATTRTG---------   240
Arabidopsis lyrata      121 TATGTAGGGATGAGGAGAATCATAATCCACATAATTCCAAGACCGTTCGATATTTTTTCCTCGACTCCGATCCAATGACGATGTCCTGGTGAAGGCTGATGAGGATTATGATGAGGCGC   240

Arabidopsis thaliana    241 ------TWAAGGATCAGGACCTGTTGGGATCTGTCTGTCKACTCCWATCAACGATGGRTCTTCACGAATAGCAACAWTWGACGACGTYATTAGGAAMAAGRTTAGCGTTATGATTAATAM   360
Arabidopsis lyrata      241 ATTATGTAAAAGATCCGGACCTGTTGGGATCTTTCTCTCGACTCCAATCAAAGATGGGACTTCAAGAATAGCAGCCACCGACGACGTTATTAGGAAAAAGATTAGTGTATTGATTAGTGC   360

Arabidopsis thaliana    361 CACA-----------------------------------------------CCACCTTTTRA----------CAAA--------TGCATCTTTYTG--------------WTCCTTAWT   480
Arabidopsis lyrata      361 TAAAATTAGGCCTCAAAAATAATTTTTATTATAAAAATAGACCTCATTTTTTCTAACTTTTAAGATTTAGTCTCAAATTTTGAATTTTATCTATTTGTTATTGCAAAAATAAGCCTTATT   480

Arabidopsis thaliana    481 TTTTR--------------------CYTTTTGA-------------------------------TTGTTTTTGTAYAGA--TTTTYTTAGTTGSCTAC                          579
Arabidopsis lyrata      481 TTCTGAAAATTTAAGAGTTGGTCTCATTTTTAAATTTTATCTACAAAATCCATTAAAAAAAATTCTTGTTTCTTTGTAAACTTTTTTTTAGTTGACT--                          579
```

Transcript Sequence: cDNA

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

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

Transcript-based displays

- Summary
- Supporting evidence
- Sequence
  - Exons
  - cDNA
  - Protein
- External 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

Gramene is produced in collaboration with Ensembl Plants

Transcript: CLE18 AT1G66145.1

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

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	NM_148631 NP_683472	

cDNA sequence

Download sequence | BLAST this sequence

Codons	Alternating codons	Alternating codons	
Exons	Alternating exons	Alternating exons	
Variants	Missense	Stop gained	Synonymous

```
                Y               K           *       Y   K           R
  1 ATGCATTTGTTAAAAGGTGGTGTGGTATTAATCATAACGCTAATCTTGTTCCTAATAACG
  1 ATGCATTTGTTAAAAGGTGGTGTGGTATTAATCATAACGCTAATCTTGTTCCTAATAACG
  1 -M--H--L--L--K--G--G--V--V--L--I--I--T--L--I--L--F--L--I--T-

          W W                 Y           W       M
 61 TCGTCTATTGTTGCTATTCGTGAAGACCCATCGTTGATTGGAGTAGACAGACAGATCCCA
 61 TCGTCTATTGTTGCTATTCGTGAAGACCCATCGTTGATTGGAGTAGACAGACAGATCCCA
 21 -S--S--I--V--A--I--R--E--D--P--S--I--I--G--V--D--R--Q--I--P-

                    W   Y           Y                           K
121 ACAGGTCCTGATCCTTTACATAATCCTCCTCAGCCTTCACCAAAACATCATCATTGGATC
121 ACAGGTCCTGATCCTTTACATAATCCTCCTCAGCCTTCACCAAAACATCATCATTGGATC
 41 -T--G--P--D--P--L--H--N--P--P--Q--P--S--P--K--H--H--H--W--I-

     R                   R                   K               YM
181 GGAGTCGAGGAAAACAATATCGACCGGTCTTGGAATTATGTGGATTATGAATCTCATCAT
```

Variation: ENSVATH05076737

Position	1:24628685
Alleles	G/A
cDNA position	151
Protein position	51
Amino acids	Q/*
Codons	Cag/Tag
Consequences	Stop gained

Explore this variant

Gene/Transcript Locations

Click

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	Variation features from Affy_250k, Perlegen_1M, WTCHG and 1001 Genc
Alleles	A/T | | Ambiguity code: **W**
Location	Chromosome **1:24628699** (forward strand) | View in location tab
Most severe consequence	| See all predicted consequences [Genes and regulation] 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.

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

Explore this variant

Genomic context | Genes and regulation (2) | Genotype frequency | Sample genotypes (200) | Linkage disequilibrium

Phenotype data (2) | 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

Genomic context

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

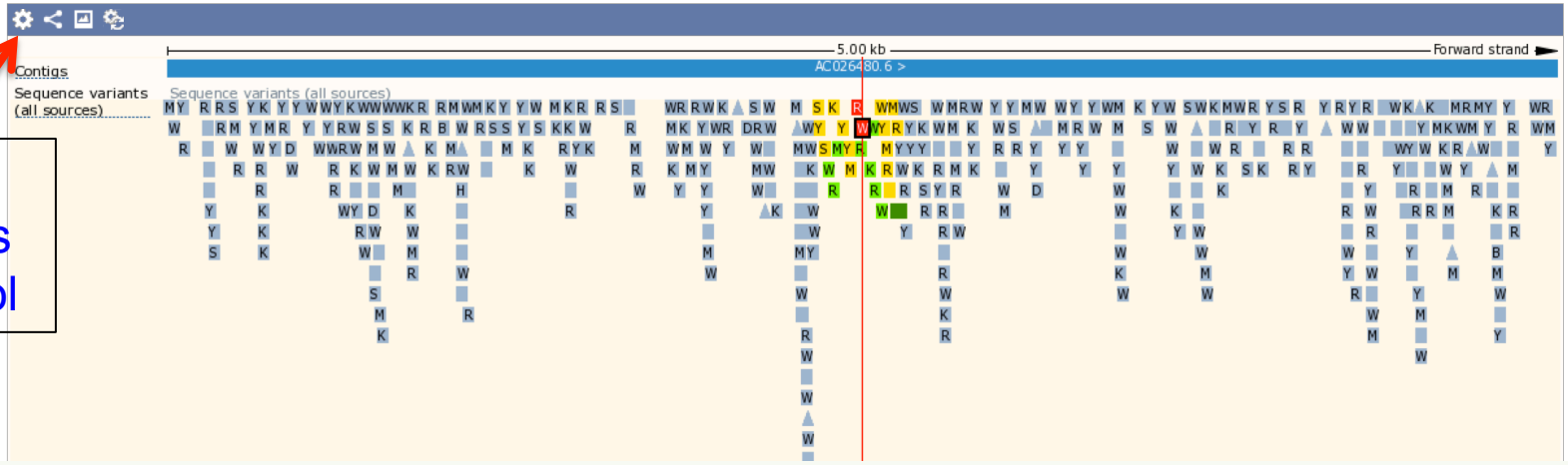
Configure this page
Add your data
Export data
Share this page
Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

ENSVATH00125659 SNP

Original source	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	| See all predicted consequences [Genes and regulation] Stop gained
Synonyms	**Perlegen** PERL0223135
HGVS names	This variant has **3** HGVS names - Hide 1:g.24628699A>T AT1G66145.1.1:c.137T>A AT1G66145.1.1:p.Leu46X
About this variant	This variant overlaps 2 transcripts and is associated with 2 phenotypes.

Context



Click on "Configure this image" 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
  - Genes and regulation
  - Flanking sequence
- Genotype frequency
- Sample genotypes
- Linkage disequilibrium
- Phenotype Data
- Phylogenetic Context
- Citations
- External

Config | Add yo | Export | Share | Bookm

Gramene collabora

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	| See all predicted consequences [Genes and regulation] Stop gained

AT1G66145	AT1G66145.1 (-) **biotype:** protein_coding	T (A)	Stop gained	137	137	46	L/*	TTA/TAA	Hide
AT1G66150	AT1G66150.1 (+) **biotype:** protein_coding	T (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
Allele (transcript)	A
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)
Consequence (Old Ensembl term)	STOP_GAINED
HGVS 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

Detail
Show
Show

Flanking sequence

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, 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	| See all predicted consequences [Genes and regulation] 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.

Flanking sequence

Download sequence | BLAST this sequence

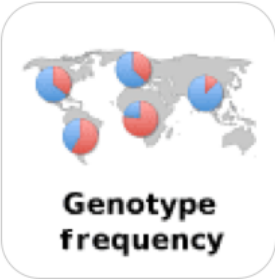
Flanking sequence

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

```
TGCATATTTTTGCATTSGAWTTTMWATARTTTATAAGGTCATAATCAATCTGTTTTTTTT
TCTGCTWWCAAWGCTTTCGATTATCTATGTCAAKCTCTAATACAAACTATATCTGTATAC
GCAAACTAATGGAATCTACATATTACGGTTTTACCACTCCAAACTMATTCTCATAMTAAM
AWATATATCTGCATGTAGATRAWWAAATATATATAWATATAKWYATTWGTTCTAGACTCC
AATSAAATGAYGATATAATGGTGCASGCTCAGGAGAGTTATGWATAGGACTATGTGCATK
RTGAGATTCATAATCMACATAATTCCAAGACCGGYCGATATTGTTTTCCTCGACTYCGAT
MCAATGATGATGTTTTGGTGAAGGCTRAGGAGGATTRTGTWAAGGATCAGGACCTGTTGG
GATCTGTCTGTCKACTCCWATCAACGATGGRTCTTCACGAATAGCAACAWTWGACGACGT
YATTAGGAAMAAGRTTAGCGTTATGATTAATAMCACACCACCTTTTRACAAATGCATCTT
TYTGWTCCTTAWTTTTTRCYTTTTGATTGTTTTTGTAYAGATTTTYTTAGTTGSCTACAC
TCTTGKTTCCTCCTTTTTATTTKTAYASTGGARTTTTAATCTTAATTAAGGGAAGCTAAT
CAAATGTCAAWTGCAATWRAYTTGAGTATATTARAARRTWTKTRCTTGACCCTTTTAACT
GAACCCMAMTTGTTTGACCMRAAAAAAAAWAGTTTCCCGAATTATTTCRTGGTAATCTC
ATTTTTAATTTGTTTTTACTK
```

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
- 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, 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	| See all predicted consequences [Genes and regulation] 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.

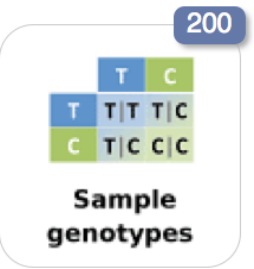
Genotype frequency

Frequency data (7)

Show/hide columns

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

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
  - 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, 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	| See all predicted consequences [Genes and regulation] 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.

Sample genotypes

Search for a sample: Search (e.g. NA18507)

Summary of genotypes by population (7)

Show/hide columns | Filter

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)

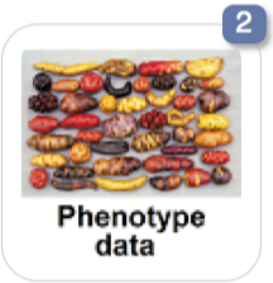
[back to top]

Genotypes for PERL_1M

Show/hide columns | Filter

Sample	Genotype (forward strand)	Description
Br-0:CS22678:PERL_1M	TTT	-
Ts-1:CS22692: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
- 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** I I Ambiguity code: **W**
Location	Chromosome **1:24628699** (forward strand) I View in location tab
Most severe consequence	I I See all predicted consequences [Genes and regulation] 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.

Phenotype Data

Significant association(s)

Show/hide columns | Filter

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 MgSO4 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_avrRpm1_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 MgSO4 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_avrRpm1_32	CLE18, AT1G66130, AT1G66173, AT1G66180, AT1G66110, AT1G66120, AT1G66160, ZFP4, TMK1, MMD1	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 | Filter

Gene	HGNC name	Position
AT1G66145	CLE18	Overlaps variant

Summary

Ensembl quick reference cards

Ensembl variations quick reference card

EMBL-EBI wellcome trust sanger institute

- The genome is shown as a blue bar, labeled with contig names
- Configure this page to turn on variation tracks
- Variations are coloured according to their consequence type

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

- Zoom into a region

- Conserved sequences
- Transcripts for the *TAGAP* gene are aligned to the genome

- Gene and transcript locations
- Allele frequencies for populations
- Associated phenotypes

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

Follow the blue link to open the variation tab

- Jump to the location tab

Click a variation for a detailed pop-up

Variation: rs1738074

rs1738074 properties

bp	6:159465977
Status	-
Class	SNP
Ambiguity code	Y
Mapweight	1
Alleles	T/C
Source	ENSEMBL, Illumina_CytoSNP12v1, Illumina_Human660W-quad, dbSNP, Illumina_Human1M-duoV3
Type	5PRIME_UTR

www.ensembl.org

Contact us at helpdesk@ensembl.org

Reference: Flicek, P. et al. Ensembl 2011 (2011) Nucleic Acids Research 39:D800-D806

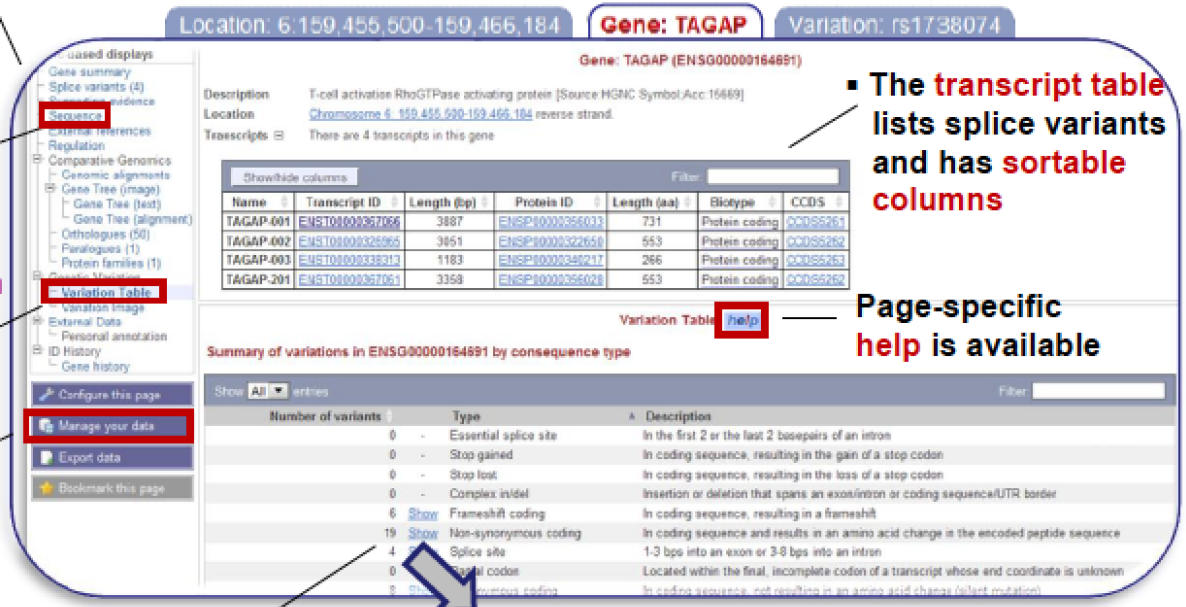
Reference: BMC Bioinformatics – Ensembl Update 2010 http://www.biomedcentral.com/series/ENSEMBL2010

Tutorials: http://www.ensembl.org/info/website/tutorials

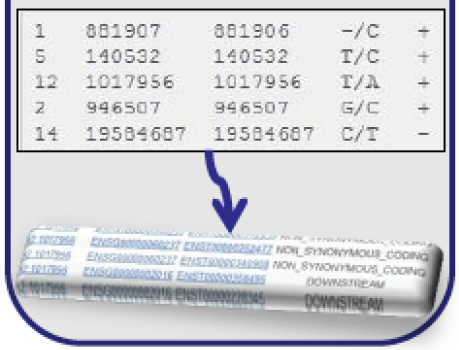
Ensembl variations quick reference card

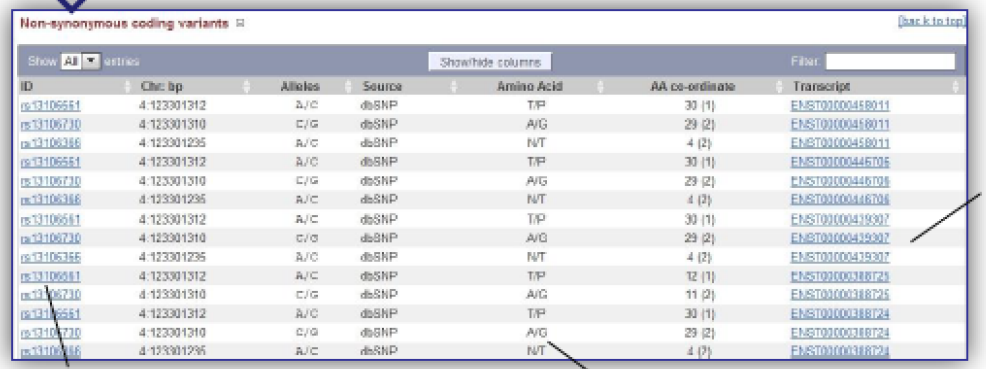
www.ensembl.org

- Contextual navigation for a gene
- View variations in the genomic sequence
- The variation table shows variations for a gene
- Use the variant effect predictor
- Search for a gene
- The transcript table lists splice variants and has sortable columns
- Page-specific help is available
- Show a table for one variation type
- View variations on the cDNA sequence in the transcript tab
- Click a variation ID to open the variation tab
- Possible amino acids at a position
- Ensembl transcripts (splice variants) are listed



Click 'Tools' for the Variant Effect Predictor





version 1.0

Gramene Mart

Complex custom queries to mine data:

- InterPro
- GO
- VEP
- Orthologs/Paralogs

Use case: Find transcription factors having "stop_gained" alleles

New | Count | Results | URL | XML | Perl | Help

Dataset 14734 / 51162745 SNPs

Zea mays variations (AGPv3 (5b))

Filters

Consequence type : stop_gained

Attributes

Gene stable ID
Consequence to transcript
Variation ID

Dataset 792 / 39625 Genes

Zea mays genes (AGPv3 (5b))

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 o
Go

Email notification to

View 20 rows as HTML Unique results only

Gene stable ID	Consequence to transcript	Variation ID	InterPro ID	InterPro short description
GRMZM2G090230	stop_gained	PZE10137018887	IPR006447	Myb_dom_plants
GRMZM2G090230	stop_gained	S10_137354081	IPR006447	Myb_dom_plants
GRMZM2G379167	stop_gained	PZE05214557574	IPR006447	Myb_dom_plants
GRMZM2G379167	stop_gained	PZE05214558344	IPR006447	Myb_dom_plants
GRMZM2G379167	stop_gained	PZE05214559441	IPR006447	Myb_dom_plants
GRMZM2G057408	stop_gained	S6_93100439	IPR006447	Myb_dom_plants
GRMZM2G057408	stop_gained	S6_93100439	IPR006447	Myb_dom_plants
GRMZM5G803992	stop_gained	PZE1010374683	IPR009057	Homeodomain-like
GRMZM2G082264	stop_gained	S7_107239125	IPR006447	Myb_dom_plants
GRMZM2G082264	stop_gained	S7_107239125	IPR009057	Homeodomain-like
GRMZM2G082264	stop_gained	S7_107239125	IPR006447	Myb_dom_plants
GRMZM2G082264	stop_gained	S7_107239125	IPR009057	Homeodomain-like
GRMZM2G370425	stop_gained	S3_114432603	IPR006447	Myb_dom_plants
GRMZM2G027914	stop_gained	S3_176435177	IPR006447	Myb_dom_plants
GRMZM2G421256	stop_gained	S2_14715154	IPR006447	Myb_dom_plants
GRMZM2G421256	stop_gained	S2_14715252	IPR006447	Myb_dom_plants
GRMZM2G421256	stop_gained	S2_14716649	IPR006447	Myb_dom_plants
GRMZM2G103783	stop_gained	S7_133372375	IPR009057	Homeodomain-like
GRMZM5G887276	stop_gained	PZE0106497353	IPR006447	Myb_dom_plants
GRMZM5G887276	stop_gained	PZE0106497353	IPR006447	Myb_dom_plants

http://www.gramene.org/biomart/martview/

Tools: Variant Effector Predictor

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

- Search
- Genomes

Name	Description	Online tool	Download code	Documentation
Variant Effect Predictor VeP	Analyse your own variants and predict the functional consequences of known and unknown variants via our Variant Effect Predictor (VEP) tool.			
BLAST/BLAT	Search our genomes for your DNA or protein sequence.			
BioMart	Use this data-mining tool to export custom datasets from Ensembl.			
Assembly converter	Map (liftover) your data's coordinates to the current assembly.			
Ensembl Genomes Virtual Machine	Pre-configured VirtualBox virtual machine (VM) running the latest Ensembl Genomes browser.			
Ensembl Genomes REST server	Access Ensembl data using your favourite programming language			

- Tools

Variant Effect Predictor

Oryza sativa Japonica (IRGSP-1.0) ▼ VEP ▼

Web Tools
- Web Tools
  - BLAST
  - Variant Effect Predictor
  - Assembly Converter

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

Gramene is produced in collaboration with Ensembl Plants

Variant Effect Predictor

Species: Oryza sativa Japonica
Assembly: IRGSP-1.0

Name for this data (optional):

Either paste data:

```
1 3452 . T A . . .
3 5087 . C A . . .
2 630 . CA C . . .
```

Examples: Ensembl default, VCF, Variant identifiers, HGVS notations, Pileup

Instant results for first variant ›

Instant results for 1 3452 . T A . . .

Most severe consequence: missense_variant

Gene/Feature/Type	Consequence	Details
OS01G0100100: OS01T0100100-01 **Type:** protein_coding	missense_variant	**Amino acids:** S/T
EPIOSAG00000002326: EPIOSAT00000003714 **Type:** ncRNA	upstream_gene_variant	**Distance to transcript:** 692bp

Note: the above is a preview of results using the *Oryza sativa* Ensembl transcript database and does not include all data fields present in the full results set. Please submit the job using the Run button below to obtain these.

Or upload file: Browse... No file selected.

Or provide file URL:

Identifiers and frequency data *Additional identifiers for genes, transcripts and variants; frequency data*

Extra options *e.g. SIFT, PolyPhen and regulatory data*

Filtering options *Pre-filter results by frequency or consequence type*

Run › Clear

Recent jobs

You have no jobs currently running or recently completed.

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

ecies

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)

Archive b39: markers & QTLs

archive.gramene.org is an instance of Gramene's build 39 which was the last to support the following legacy datasets:

- QTL data
- Comparative maps
- Gramene Markers Database
- Gramene Proteins Database
- Gramene Ontologies Database

- **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
- **Pathways databases**: BioCyc based cellular metabolic networks for 10 plant species
- ARCHIVE - Markers, Proteins and Ontology databases, QTLs, Comparative Maps

- Tools

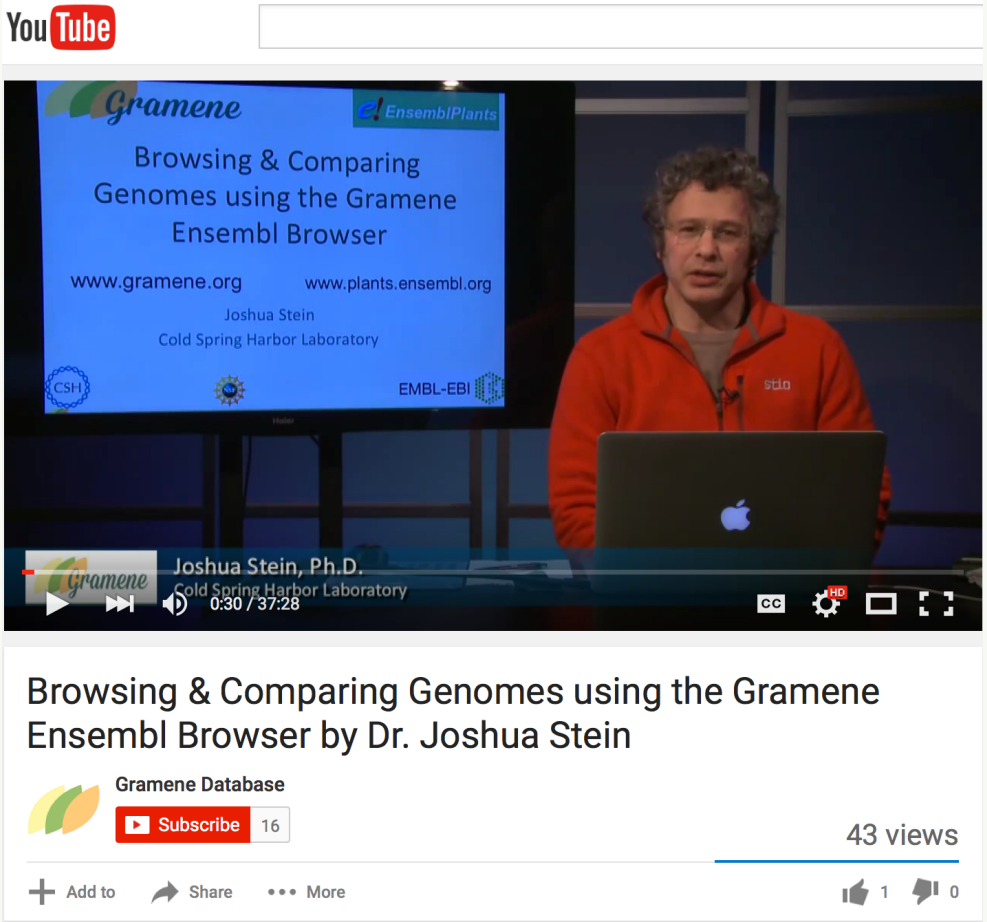
Ensembl videos & tutorials

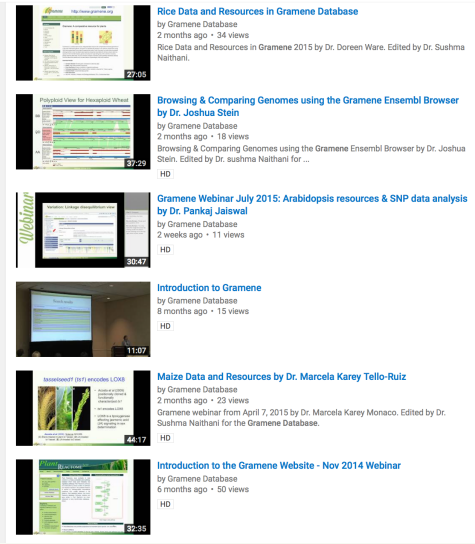
- Video: Browsing SNPs and CNVs in Ensembl
  [http://ensembl.gramene.org/Help/Movie?id=208]

- Demo: Structural variation for a region
  [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]

Gramene Database @ YouTube





- 12 live webinars & 14 recorded video-tutorials
- Master mailing list of over 1,100 plant researchers
- Engaging specific plant communities (Arabidopsis, rice, maize, etc.)
- Create a robust presence for Gramene in the Plantae platform (~30,000 plant scientists globally)

How to reach us

- E-mail: webinars@gramene.org
- Contact form: http://www.gramene.org/contact
- Gramene YouTube channel
- Announcements mailing list

https://www.facebook.com/Gramene

@GrameneDatabase

Gramene - Exploring Function through Comparative Genomics and Network Analysis

NSF IOS 1127112 (2011- 2017)

CyVerse: Liya Wang

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, Justin Preece, Sushma Naithani

Paul Kersey / **Robert Petryszyk** (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)

EVA: Gary Sanders

Variation Data

- Single base (SNP, inter-homoelogous) & structural
- Natural & EMS-induced

Species	Variants	Source	Studies
Arabidopsis thaliana	**28,468,394** **SV: 27,334**	250K SNPs x 1179 accessions (Affy chip) 18 resequenced accessions (BBSRC-funded collaboration) 392 accessions (1001 genomes project)	Cao et al (2011). Nat Genet 43:956 Horton et al (2012). Nat Genet 44:212 Clark et al (2007). Science. 317:338 Atwell et al (2010). Nature. 465:627
Brachypodium distachyon	**655,976**	394K SNPs x 3 populations of *B. sylvaticum*	Fox et al (2013). Appl in Plant Sci 1(3):1200011
Hordeum vulgare	**38,949,078**	4 resequenced accessions plus wild barely & RNA-Seq (Illumina HiSeq 2000) from embryos of 9 cultivars & Morex 5M SNPs from 90 Morex x Barke; 6M SNPs from 84 Oregon Wolfe 2,600 SNPs (iSelect 9k chip)	The International Barley Genome Sequencing Consortium (2012). Nature 491: 711 Mascher et al (2013.) Plant J. 76:718 Comadran et al (2012). Nat Genet 44:1388
Oryza glaberrima	**7,172,036**	Resequenced 20 O. glaberrima accessions & 19 accessions of wild progenitor *O. barthii*	Oryza Genome Evolution project
Oryza glumaepatula	**4,901,344**	4.9M SNPs x 7 accessions (OGE)	Oryza Genome Evolution project
Oryza sativa ssp japonica	**5,512,746**	160K SNPs x 20 accessions 1311 SNPs x 395 accessions 4M Japonica vs Indica SNPs (BGI) SNPs from OMAP alignments	OryzaSNP: McNally et al (2009). PNAS 106:12273-12278 Zhao et al (2010). PLoS ONE 5:e10780 NCBI dbSNP: Yu et al (2005). PLoS Biol. 3:e38 Oryza Map Alignment Project
Oryza sativa ssp indica	**5,418,373**	4M Japonica vs Indica SNPs (BGI) SNPs from OMAP alignments	NCBI dbSNP: Yu et al (2005). PLoS Biol. 3:e38 Oryza Map Alignment Project
Solanum lycopersicum	**71,156,260**	WGS of 84 tomato accessions & related species	100 Tomato Genome Sequencing Consortium (2014). The Plant Journal 80:136-148
Sorghum bicolor	**6,707,777** **EMS: 1,481,453** **SV: 32,246**	265K SNPs in 378 SAP lines 6.5M SNPs in 45 *S. bicolor* & 2 *S. propinquum* lines 1.5M EMS-induced mutations from 252 families 32K SVs from dGVA	Morris et al (2013). PNAS 110:453-458 Mace et al (2013). Nat Commun 4:2320 Xin et al (2008). BMC Plant Biol 8:103 dGVA: Zheng et al (2011). Genome Biol 12:R114
Triticum aestivum	**12,220,924**	1.57M SNPs in 62 lines (Wheat HapMap) 725K non-redundant SNP from CerealsDB (Axiom, iSelect & KASP array sets) 10.3M inter-homoeologous variants (genome browser)	Jordan et al (2015). Genome Biology 16:48 CerealsDB Plants Ensembl
Vitis vinifera	**457,404**	Resequencing USDA germplasm collection	Myles et al (2010). PLoS ONE 5:e8219
Zea mays	**51,162,745**	HapMap1: NAM founder lines HapMap2: 55M SNPs in 104 accessions Panzea 2.7 GBS (719K SNPs in 16,718 accessions)	Gore et al (2009). Science 326:1115-1117 Chia et al (2013). Genet 44:803-807. Panzea.org

Exercise 1 – Explore the genetic variation around a gene

Explore the genetic variation for the lcyE gene and identify the nonsynonymous substitution associated with provitamin A accumulation in the kernel by Harjes et al (2008) [Science 319:330]. This variant is also known as SNP216 for its position relative to the start of the gene's coding sequence.

1. Visualize the genetic variants for this gene
2. Are there any new stop codons introduced (nonsense variants) in this gene?
3. Are there any transcript-specific variants for this gene?
4. Download a subset of the variants (e.g., those that introduce an amino acid change in the protein)
5. Find the missense variant (nonsynonymous substitution) associated with accumulation of provitamin A in the study according to Harjes et al (2008). What are the genomic coordinates for this variant?

Exercise 2: Gramene Mart

Identify all the transcription factors that have a SNP variant that introduces a stop codon resulting in a truncated peptide

Hint: Use the Gene Ontology or AmiGO database to find the GO term for "transcription factor activity, transcription factor binding" [GO:0000989].

Another hint: The description of the functional consequence for a stop codon introduced by a SNP variant is: "stop_gained"