Gramene Exercises

March 2018

The Gramene database (**http://www.gramene.org**) is an integrated resource for comparative genomics and pathways in plants. The database provides researchers with valuable information on numerous crops and model species, enabling powerful functional comparisons across species.

In partnership with Ensembl Plants, we host genome browsers for 53 complete reference genomes (build 56b; for a current list see our release notes: http://gramene.org/release-notes). Each plant genome encompasses value-added annotations, gene-trees, and whole genome alignments. Evolutionary histories provided in phylogenetic gene trees classify orthologous and paralogous relationships as speciation and duplication events. Orthologous genes inform synteny maps that enable interspecies browsing across ancestral regions. Browsers from multiple species can be viewed simultaneously, with links showing homologous gene and whole-genome alignment mappings. SNP and structural diversity data, available for 9 reference genomes, are displayed in the context of gene annotation, showing functional consequences that can be assigned to individual accessions within the diversity panel. Genomic data include phenotypic, transcriptome profiling, and methylome data. Visual displays can be downloaded as high-resolution, publication-ready, image files. A fully integrated BLAST tool enables visualization of alignments within the browser. For data mining, our BioMart tool enables complex queries of sequence, annotation, homology and variation data, and provides an additional gateway into the genome browsers.

Gramene is driven by several platform infrastructures or modules that are linked to provide a unified user experience. The Genomes and Pathway modules enable species-specific and cross-species data downloads for discrete region(s), gene(s) or gene feature(s) via the Genome Browser, and pathway-centered downloads via the Pathways portal and Plant Reactome. The genome browser portal (http://ensembl.gramene.org) takes advantage of the Ensembl project’s infrastructure to provide an interface for exploring genome features, functional ontologies, variation data, and comparative phylogenomics. In addition, plant metabolic and regulatory pathways are available for cross-species analysis via the Plant Reactome (http://plantreactome.gramene.org). The Plant Reactome hosts 264 rice pathways (80% manually curated) and orthology-based projections of the rice reference pathways to 74 plant species. Through a collaboration with EBI-ATLAS, we now also display baseline gene expression on our browsers.

Gramene's archives (http://archive.gramene.org) host historical data and tools such as legacy databases (*e.g.*, genetic markers, comparative maps, curated genes and phenotype-associated variant alleles, proteins, ontologies). Our legacy BioCyc collection of pathway databases is hosted on a virtual server at CyVerse (http://pathway.iplantcollaborative.org).

In addition, project data is available for customizable downloads from the GrameneMart utility (http://ensembl.gramene.org/biomart), nucleotide and protein sequence alignments via BLAST (http://ensembl.gramene.org/Tools/Blast), bulk downloads via file transfer protocol (FTP) at Gramene (ftp://ftp.gramene.org/pub/gramene) and Ensembl Genomes (http://ensembl.gramene.org/info/website/ftp/index.html), and programmatic access via Ensembl's REST application programming interface (API) and public MySQL (http://www.gramene.org/web-services). The website, database, and its contents are being updated quarterly and updates can be followed from the Gramene news portal (http://www.gramene.org/blog), by browsing the site's release notes (http://www.gramene.org/release-notes), and through our social media: Facebook (https://www.facebook.com/Gramene) and Twitter (https://twitter.com/gramenedatabase).

The examples below provide sample queries to explore the Gramene website. We describe one of many possible ways to solve a given exercise and encourage you to discover other ingenious ways to solve them!

These and additional examples are also available on the Gramene's Outreach page (www.gramene.org/outreach).

EXERCISES

These exercises will illustrate the power of comparative plant genomics in research using the resources in Gramene.

Exercise 1. View a phylogenetic tree for a family of transcription factors

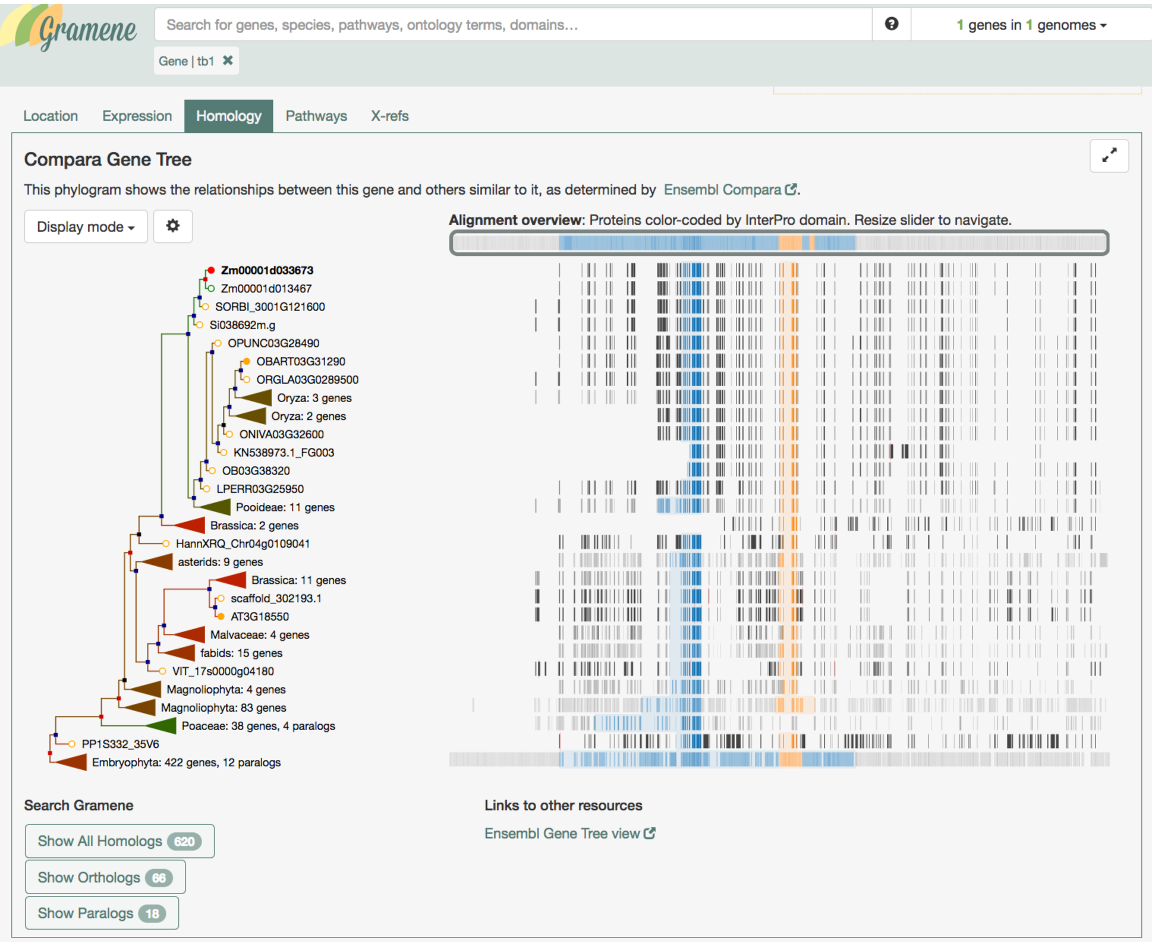
In this exercise, we will navigate a phylogenetic tree for plant genes in the TCP family of transcription factors (named after the first characterized protein members: maize TB1, snapdragon CYC, and rice PCF), highlight species-specific orthologs/paralogs with particular GO annotations in the tree. We will then proceed to generate lists of orthologs/paralogs and download both, images and tables with our results.

a. How many orthologs can you identify for maize TB1?

Hint: You may find the answer for this through different approaches. Gramene's new search will give you the quickest answer through a snapshot of the *tb1* (Zm00001d033673) gene tree. Other approaches are described in subsequent exercises.

1. Go to www.gramene.org. This is Gramene's homepage.
2. Enter TB1 in the search box. This will redirect you to search.gramene.org.
3. Find the maize *tb1* (Zm00001d033673) gene. Click on the "Homology" tab.

Answer: There are 66 orthologs of maize TB1 in the current Gramene build #56.



Compara Gene Tree

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

Alignment overview: Proteins color-coded by InterPro domain. Resize slider to navigate.

b. What is the most prominent TCP domain among members of the gene tree? How many maize genes have a TCP domain?

Note: By looking at the maize TB1 gene tree in Gramene's genome browser, 3 InterPro domains with TCP features appear to be shared among family members. IPR005333 is considered a "family" of protein domain as it encompasses TCP domains: IPR017887 and IPR017888.

Answer: Again, there are multiple ways to answer a question.

1) Via search.gramene.org:
   a) Go to the "Homology" tab of the Search results for maize *tb1* (see above).
   b) Click on the most prominent blue colored domain (IPR017887).
   c) Simple answer: IPR017887 - Transcription factor TCP subgroup.
2) Via the genome browser:
   a) From the "Homology" tab in your search results (see above), click on "Ensembl Gene Tree view" OR go to ensembl.gramene.org, search for maize TB1 and click on (Plant Compara) Gene Tree (EPIGT00820000103607).

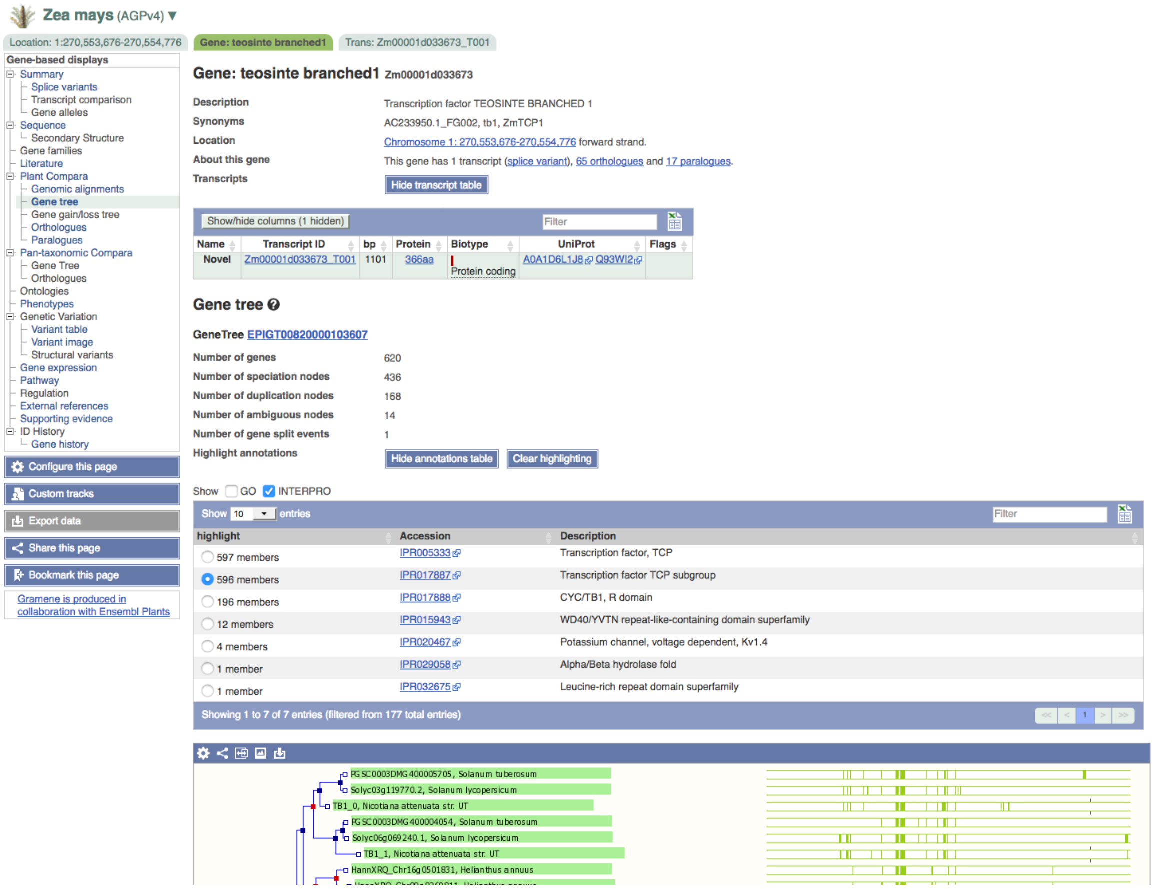
b) Select InterPro domains in the annotations table. By selecting an individual domain, all members that share it will be highlighted in the tree.

c) Detailed answer:

   i) 597 members have IPR005333 Transcription factor, TCP.

   ii) 596 members IPR017887 Transcription factor TCP subgroup

   iii) 196 members IPR017888 CYC/TB1, R domain



Gene: teosinte branched1 Zm00001d033673

Description	Transcription factor TEOSINTE BRANCHED 1
Synonyms	AC233950.1_FG002, tb1, ZmTCP1
Location	Chromosome 1: 270,553,676-270,554,776 forward strand.
About this gene	This gene has 1 transcript (splice variant), 65 orthologues and 17 paralogues.
Transcripts	Hide transcript table

Name	Transcript ID	bp	Protein	Biotype	UniProt	Flags
Novel	Zm00001d033673_T001	1101	366aa	Protein coding	A0A1D6L1J8 Q93WI2	

Gene tree

GeneTree EPIGT00820000103607

Number of genes	620
Number of speciation nodes	436
Number of duplication nodes	168
Number of ambiguous nodes	14
Number of gene split events	1
Highlight annotations	Hide annotations table Clear highlighting

Show ☐ GO ☑ INTERPRO

highlight	Accession	Description
○ 597 members	IPR005333	Transcription factor, TCP
● 596 members	IPR017887	Transcription factor TCP subgroup
○ 196 members	IPR017888	CYC/TB1, R domain
○ 12 members	IPR015943	WD40/YVTN repeat-like-containing domain superfamily
○ 4 members	IPR020467	Potassium channel, voltage dependent, Kv1.4
○ 1 member	IPR029058	Alpha/Beta hydrolase fold
○ 1 member	IPR032675	Leucine-rich repeat domain superfamily

Showing 1 to 7 of 7 entries (filtered from 177 total entries)

3) Customized data dump: Using the BioMart utility.

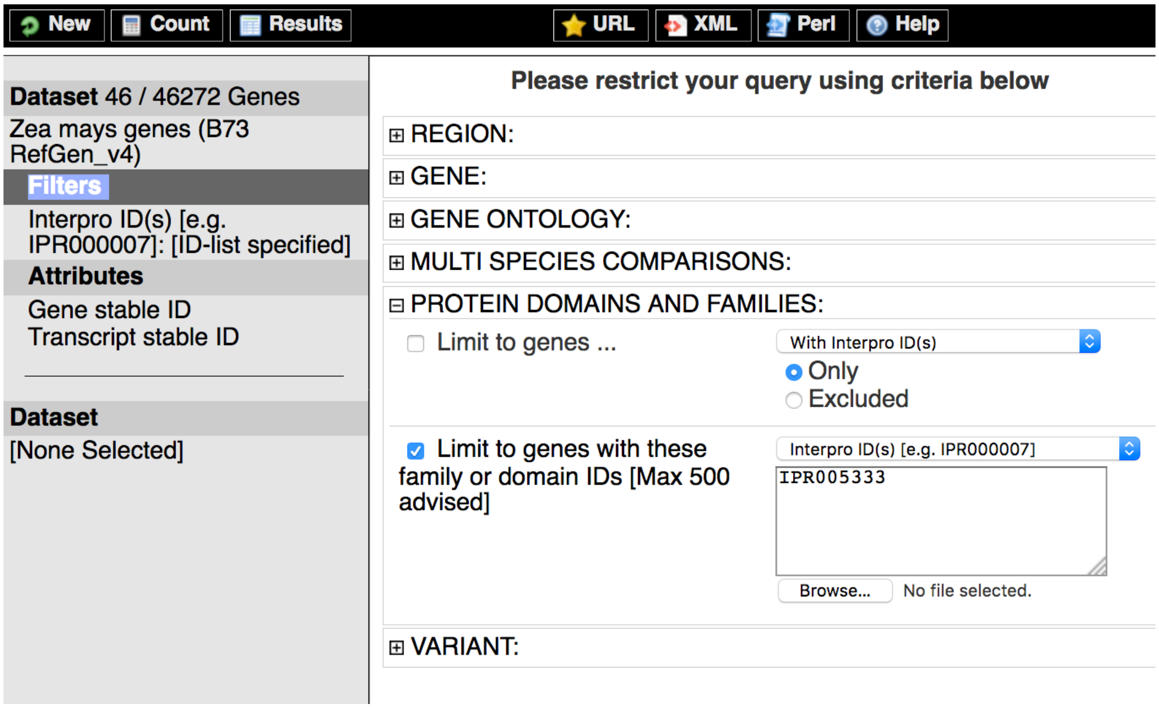
   a) Go to http://ensembl.gramene.org/biomart/martview.

   b) Select Database: “Plant Genes” and Dataset: “Zea mays genes”.

   c) Under “Protein Domains”, select “Limit to genes with these family or (InterPro) domain IDs” and enter “IPR005333”, “IPR017887” or “IPR017888”.

   d) Click on “Count”. Alternatively, under “Attributes” select the associated data (e.g., gene or transcript ID, position, sequence, variants, GO terms, etc.) that you would like to download for these genes.

Answer: There are 46 maize genes with IPR005333, 45 with IPR017887 and 4 with IPR017888.



Dataset 46 / 46272 Genes
Zea mays genes (B73 RefGen_v4)

Filters
Interpro ID(s) [e.g. IPR000007]: [ID-list specified]

Attributes
Gene stable ID
Transcript stable ID

Dataset
[None Selected]

Please restrict your query using criteria below

⊞ REGION:
⊞ GENE:
⊞ GENE ONTOLOGY:
⊞ MULTI SPECIES COMPARISONS:
⊟ PROTEIN DOMAINS AND FAMILIES:

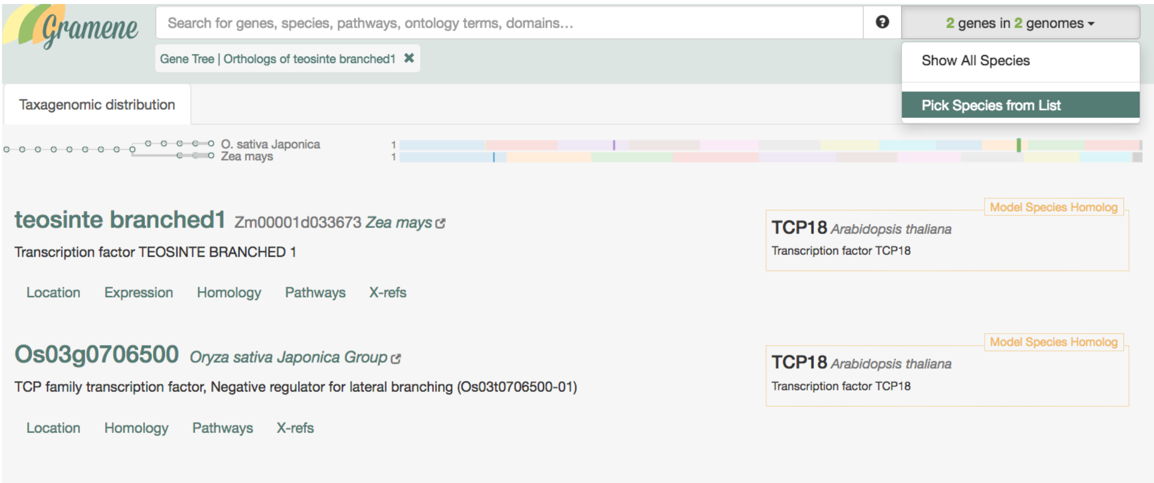
□ Limit to genes ... With Interpro ID(s) ● Only ○ Excluded

☑ Limit to genes with these family or domain IDs [Max 500 advised] Interpro ID(s) [e.g. IPR000007]
IPR005333
Browse... No file selected.

⊞ VARIANT:

c. You have learned 3 ways to find orthologs for a given gene (via Search, Genome Browser and BioMart). Can you identify the (*Japonica*) rice ortholog of the maize *tb1* gene? Wouldn't it be nice to highlight both genes in the TCP gene family tree?

1) Via Search:
   a) From the "Homology" tab in Search results (see above), select *Zea mays* and *Oryza sativa japonica* from the drop-down menu on the top right of the Search page.
   b) Click on "Show Orthologs".



Gene Tree | Orthologs of teosinte branched1 ✖

teosinte branched1 Zm00001d033673 *Zea mays*
Transcription factor TEOSINTE BRANCHED 1

TCP18 *Arabidopsis thaliana*
Transcription factor TCP18

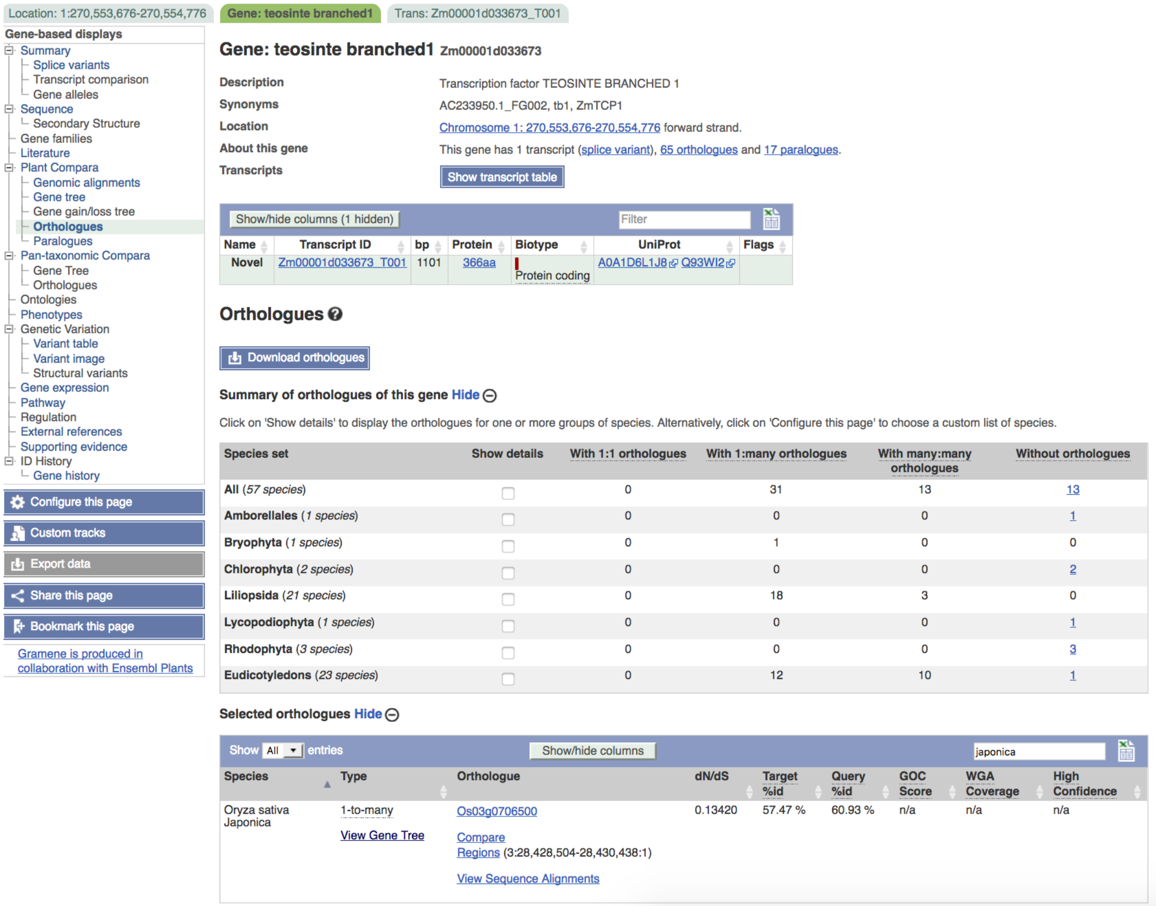
Os03g0706500 *Oryza sativa Japonica Group*
TCP family transcription factor, Negative regulator for lateral branching (Os03t0706500-01)

TCP18 *Arabidopsis thaliana*
Transcription factor TCP18

2) Via Genome Browser:

a) From the left side menu of the Gene Summary page or the Plant Compara Gene Tree view (see above), select the (Plant Compara) “Orthologues” option

b) Type “japonica” on the “Filter” box to show only rice orthologues in the results table.



Gene: teosinte branched1 Zm00001d033673

Description	Transcription factor TEOSINTE BRANCHED 1
Synonyms	AC233950.1_FG002, tb1, ZmTCP1
Location	Chromosome 1: 270,553,676-270,554,776 forward strand.
About this gene	This gene has 1 transcript (splice variant), 65 orthologues and 17 paralogues.
Transcripts	Show transcript table

Show/hide columns (1 hidden) Filter

Name	Transcript ID	bp	Protein	Biotype	UniProt	Flags
Novel	Zm00001d033673_T001	1101	366aa	Protein coding	A0A1D6L1J8 Q93WI2	

Orthologues

Download orthologues

Summary of orthologues of this gene Hide ⊖

Click on 'Show details' to display the orthologues for one or more groups of species. Alternatively, click on 'Configure this page' to choose a custom list of species.

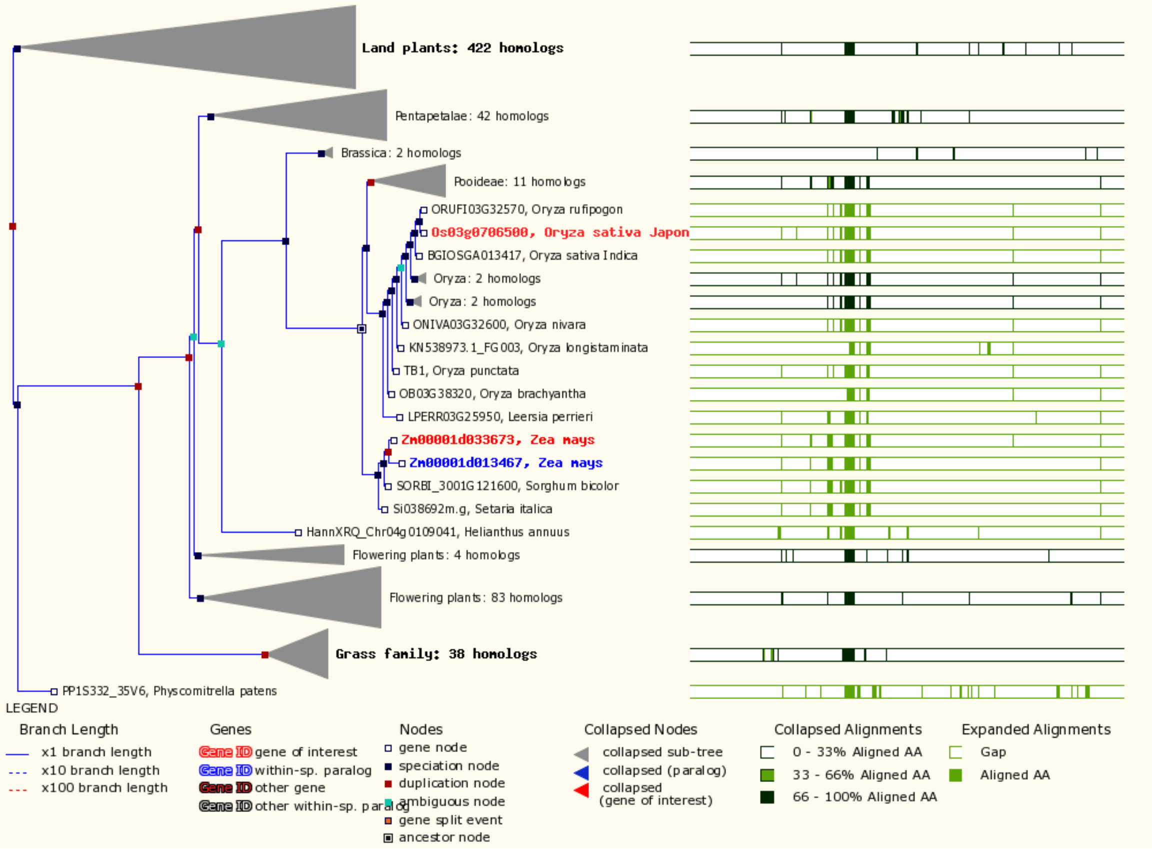
Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
All (*57 species*)	□	0	31	13	13
Amborellales (*1 species*)	□	0	0	0	1
Bryophyta (*1 species*)	□	0	1	0	0
Chlorophyta (*2 species*)	□	0	0	0	2
Liliopsida (*21 species*)	□	0	18	3	0
Lycopodiophyta (*1 species*)	□	0	0	0	1
Rhodophyta (*3 species*)	□	0	0	0	3
Eudicotyledons (*23 species*)	□	0	12	10	1

Selected orthologues Hide ⊖

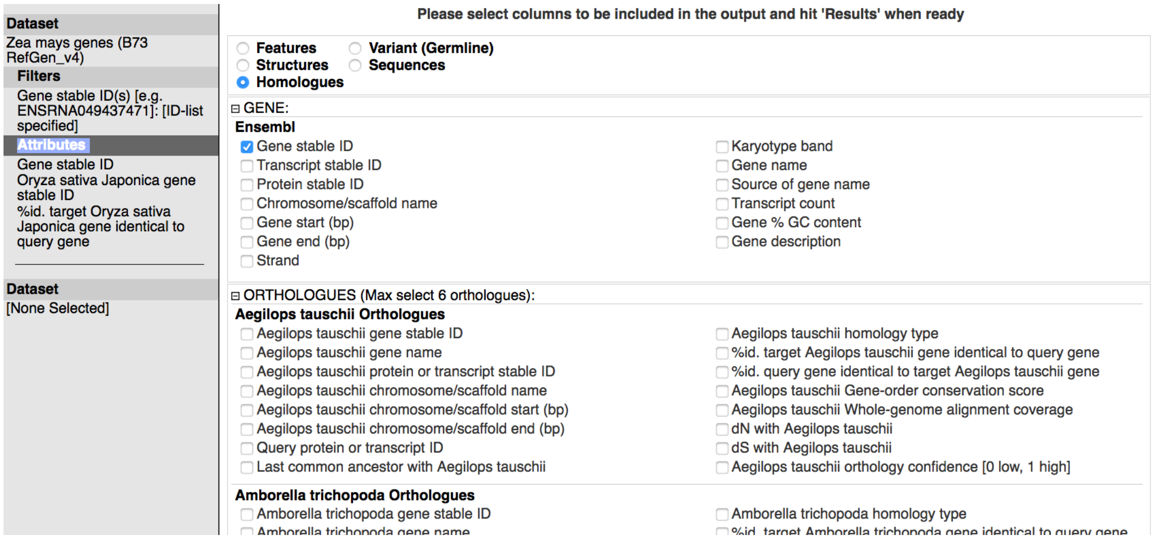
Show All entries Show/hide columns japonica

Species	Type	Orthologue	dN/dS	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence
Oryza sativa Japonica	1-to-many View Gene Tree	Os03g0706500 Compare Regions (3:28,428,504-28,430,438:1) View Sequence Alignments	0.13420	57.47 %	60.93 %	n/a	n/a	n/a

c) Click on the “View Gene Tree” link for the rice orthologue.


Land plants: 422 homologs
Pentapetalae: 42 homologs
Brassica: 2 homologs
Pooideae: 11 homologs
ORUFI03G32570, Oryza rufipogon
Os03g0706500, Oryza sativa Japon
BGIOSGA013417, Oryza sativa Indica
Oryza: 2 homologs
Oryza: 2 homologs
ONIVA03G32600, Oryza nivara
KN538973.1_FG003, Oryza longistaminata
TB1, Oryza punctata
OB03G38320, Oryza brachyantha
LPERR03G25950, Leersia perrieri
Zm00001d033673, Zea mays
Zm00001d013467, Zea mays
SORBI_3001G121600, Sorghum bicolor
Si038692m.g, Setaria italica
HannXRQ_Chr04g0109041, Helianthus annuus
Flowering plants: 4 homologs
Flowering plants: 83 homologs
Grass family: 38 homologs
PP1S332_35V6, Physcomitrella patens
LEGEND
Branch Length
x1 branch length
x10 branch length
x100 branch length
Genes
Gene ID gene of interest
Gene ID within-sp. paralog
Gene ID other gene
Gene ID other within-sp. paralog
Nodes
gene node
speciation node
duplication node
ambiguous node
gene split event
ancestor node
Collapsed Nodes
collapsed sub-tree
collapsed (paralog)
collapsed (gene of interest)
Collapsed Alignments
0 - 33% Aligned AA
33 - 66% Aligned AA
66 - 100% Aligned AA
Expanded Alignments
Gap
Aligned AA

3) Via BioMart:
   a) From the "*Zea Mays* genes" data set in BioMart (see above), under the "Gene" filter, select "ID list limit".
   b) Enter "Zm00001d033673" as the "Gene stable ID" for maize *tb1*.
   c) Under "Attributes", select "Homologs".
   d) From the "Homologs" attributes form, under "Gene Attributes" select "Gene stable ID", and under "Orthologs" select "Oryza sativa Japonica gene stable ID" and any additional data desired (e.g., % identity).
   e) Click on "Results". Customize how to view and export your results.



Dataset
Zea mays genes (B73 RefGen_v4)
Filters
Gene stable ID(s) [e.g. ENSRNA049437471]: [ID-list specified]
Attributes
Gene stable ID
Oryza sativa Japonica gene stable ID
%id. target Oryza sativa Japonica gene identical to query gene

Dataset
[None Selected]

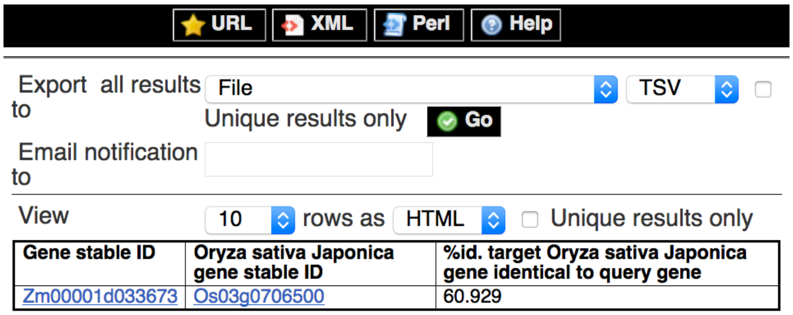
Please select columns to be included in the output and hit 'Results' when ready

○ **Features** ○ **Variant (Germline)**
○ **Structures** ○ **Sequences**
● **Homologues**

⊟ GENE:
Ensembl
☑ Gene stable ID
□ Transcript stable ID
□ Protein stable ID
□ Chromosome/scaffold name
□ Gene start (bp)
□ Gene end (bp)
□ Strand
□ Karyotype band
□ Gene name
□ Source of gene name
□ Transcript count
□ Gene % GC content
□ Gene description

⊟ ORTHOLOGUES (Max select 6 orthologues):
Aegilops tauschii Orthologues
□ Aegilops tauschii gene stable ID
□ Aegilops tauschii gene name
□ Aegilops tauschii protein or transcript stable ID
□ Aegilops tauschii chromosome/scaffold name
□ Aegilops tauschii chromosome/scaffold start (bp)
□ Aegilops tauschii chromosome/scaffold end (bp)
□ Query protein or transcript ID
□ Last common ancestor with Aegilops tauschii
□ Aegilops tauschii homology type
□ %id. target Aegilops tauschii gene identical to query gene
□ %id. query gene identical to target Aegilops tauschii gene
□ Aegilops tauschii Gene-order conservation score
□ Aegilops tauschii Whole-genome alignment coverage
□ dN with Aegilops tauschii
□ dS with Aegilops tauschii
□ Aegilops tauschii orthology confidence [0 low, 1 high]

Amborella trichopoda Orthologues
□ Amborella trichopoda gene stable ID
□ Amborella trichopoda gene name
□ Amborella trichopoda homology type
□ %id. target Amborella trichopoda gene identical to query gene



URL | XML | Perl | Help

Export all results to File TSV □
Unique results only Go

Email notification to

View 10 rows as HTML □ Unique results only

Gene stable ID	Oryza sativa Japonica gene stable ID	%id. target Oryza sativa Japonica gene identical to query gene
Zm00001d033673	Os03g0706500	60.929

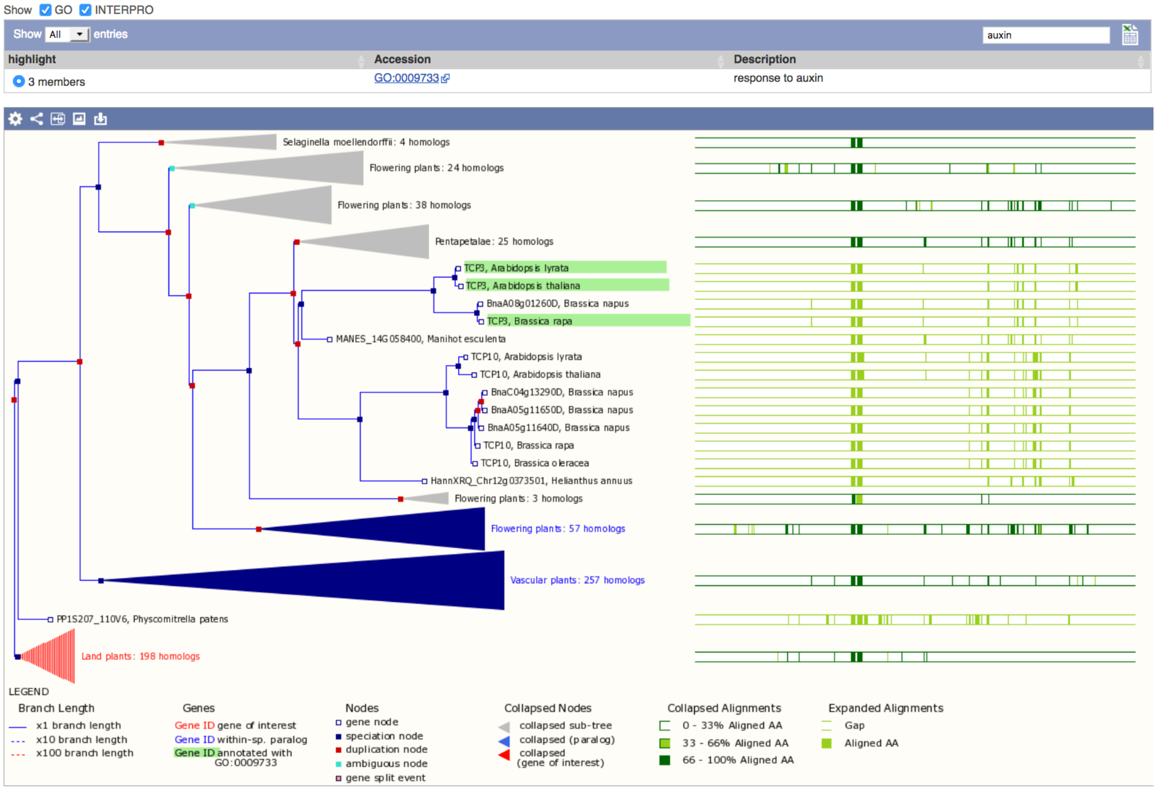
Answer: *Os03g0706500* (IRGSP1) or *LOC_Os03g49880*.

d. Identify genes in the tree that have been associated with auxin response.

Hint: GO:0009733 is the GO term identifier for “response to auxin”.

1. From the Plant Compara Gene Tree view (see above), enter the term “auxin” in the Filter box to identify GO or InterPro term(s) for auxin response.
2. Select GO:0009733

Answer: There are 3 genes encoding TCP3 in *Arabidopsis thaliana*, *A. lyrata* and *Brassica rapa* in the tree that have been associated with response to auxin.


Show ☑ GO ☑ INTERPRO
Show All entries
auxin

highlight	Accession	Description
● 3 members	GO:0009733	response to auxin

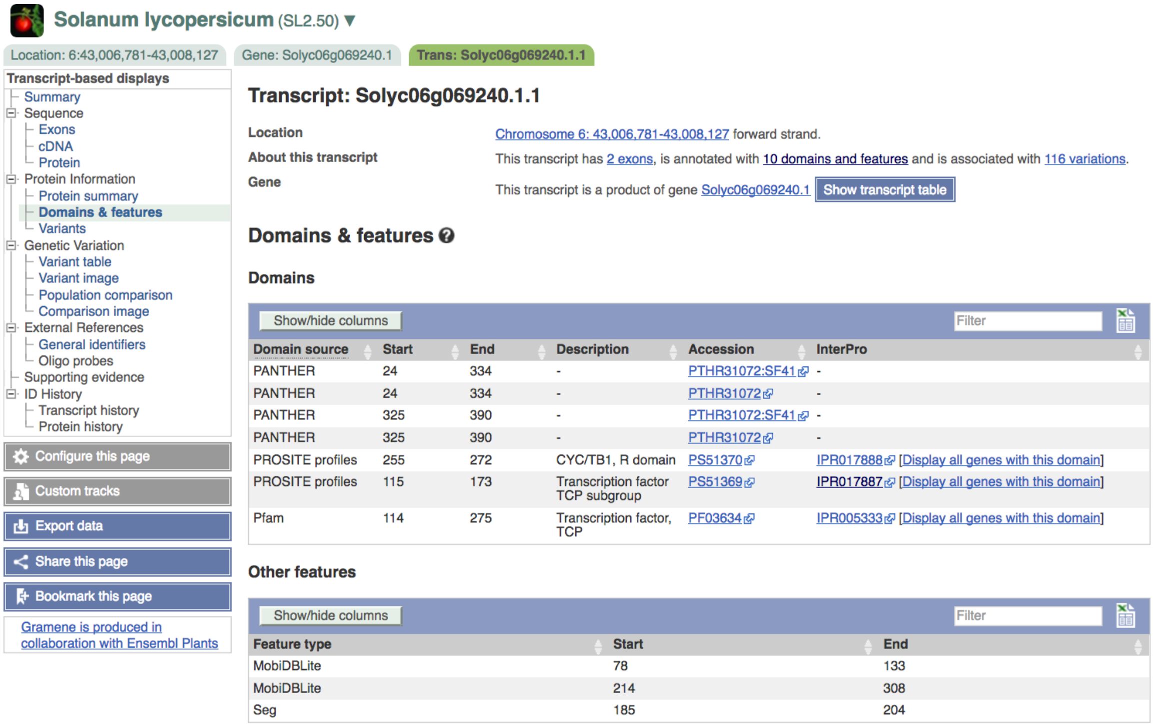
Selaginella moellendorffii: 4 homologs
Flowering plants: 24 homologs
Flowering plants: 38 homologs
Pentapetalae: 25 homologs
TCP3, Arabidopsis lyrata
TCP3, Arabidopsis thaliana
BnaA08g01260D, Brassica napus
TCP3, Brassica rapa
MANES_14G058400, Manihot esculenta
TCP10, Arabidopsis lyrata
TCP10, Arabidopsis thaliana
BnaC04g13290D, Brassica napus
BnaA05g11650D, Brassica napus
BnaA05g11640D, Brassica napus
TCP10, Brassica rapa
TCP10, Brassica oleracea
HannXRQ_Chr12g0373501, Helianthus annuus
Flowering plants: 3 homologs
Flowering plants: 57 homologs
Vascular plants: 257 homologs
PP1S207_110V6, Physcomitrella patens
Land plants: 198 homologs
LEGEND
Branch Length
x1 branch length
x10 branch length
x100 branch length
Genes
Gene ID gene of interest
Gene ID within-sp. paralog
Gene ID annotated with GO:0009733
Nodes
gene node
speciation node
duplication node
ambiguous node
gene split event
Collapsed Nodes
collapsed sub-tree
collapsed (paralog)
collapsed (gene of interest)
Collapsed Alignments
0 - 33% Aligned AA
33 - 66% Aligned AA
66 - 100% Aligned AA
Expanded Alignments
Gap
Aligned AA

Exercise 2. Identify tomato transcription factors within the TCP gene family with a SNP that results in a truncated peptide.

Note: The SNP will introduce a stop codon(*) resulting in a truncated protein product.

Hints:

1) Via Search and/or Genome Browser:
    a) Go to the Transcript page of Solyc06g069240.1, a tomato (*Solanum lycopersicum*) ortholog of maize *tb1* (proceed as we did above to identify the rice ortholog of maize TB1).
    b) Select “Domains & features” from the page’s left side menu.
    c) Find the IPR017887 domain (transcription factor TCP subgroup) and click on “Display all genes with this domain”.
    d) Copy the resulting gene list and use it as input to mine for tomato variants with a “stop_gained” as functional consequence in the Gramene Mart (see #2 ahead).



Solanum lycopersicum (SL2.50) ▼

Location: 6:43,006,781-43,008,127 | Gene: Solyc06g069240.1 | Trans: Solyc06g069240.1.1

Transcript-based displays
- Summary
- Sequence
  - Exons
  - cDNA
  - Protein
- Protein Information
  - Protein summary
  - Domains & features
  - Variants
- Genetic Variation
  - Variant table
  - Variant image
  - Population comparison
  - Comparison image
- External References
  - General identifiers
  - Oligo probes
- Supporting evidence
- ID History
  - Transcript history
  - Protein history

Configure this page
Custom tracks
Export data
Share this page
Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

Transcript: Solyc06g069240.1.1

Location	Chromosome 6: 43,006,781-43,008,127 forward strand.
About this transcript	This transcript has 2 exons, is annotated with 10 domains and features and is associated with 116 variations.
Gene	This transcript is a product of gene Solyc06g069240.1 Show transcript table

Domains & features

Domains

Show/hide columns Filter

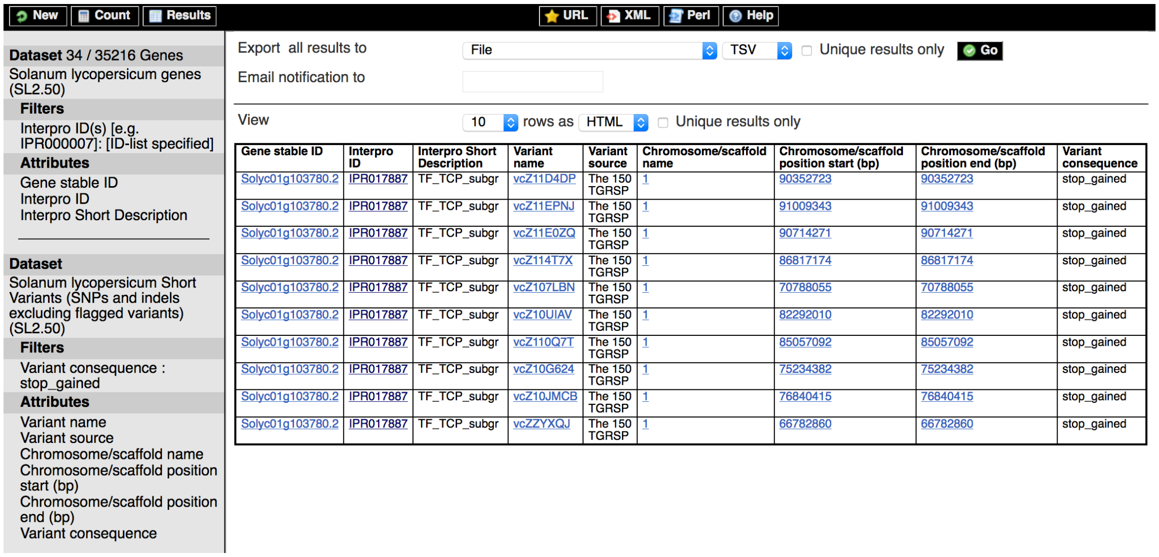
Domain source	Start	End	Description	Accession	InterPro
PANTHER	24	334	-	PTHR31072:SF41	-
PANTHER	24	334	-	PTHR31072	-
PANTHER	325	390	-	PTHR31072:SF41	-
PANTHER	325	390	-	PTHR31072	-
PROSITE profiles	255	272	CYC/TB1, R domain	PS51370	IPR017888 [Display all genes with this domain]
PROSITE profiles	115	173	Transcription factor TCP subgroup	PS51369	IPR017887 [Display all genes with this domain]
Pfam	114	275	Transcription factor, TCP	PF03634	IPR005333 [Display all genes with this domain]

Other features

Show/hide columns Filter

Feature type	Start	End
MobiDBLite	78	133
MobiDBLite	214	308
Seg	185	204

2) Via BioMart: First use IPR017887 (TCP domain) to filter the tomato genes data set. Then select tomato variations as a second data set and under Filters, use "stop_gained" as "Consequence type".



Dataset 34 / 35216 Genes
Solanum lycopersicum genes (SL2.50)
Filters
Interpro ID(s) [e.g. IPR000007]: [ID-list specified]
Attributes
Gene stable ID
Interpro ID
Interpro Short Description

Dataset
Solanum lycopersicum Short Variants (SNPs and indels excluding flagged variants) (SL2.50)
Filters
Variant consequence : stop_gained
Attributes
Variant name
Variant source
Chromosome/scaffold name
Chromosome/scaffold position start (bp)
Chromosome/scaffold position end (bp)
Variant consequence

Export all results to File TSV □ Unique results only Go

Email notification to

View 10 rows as HTML □ Unique results only

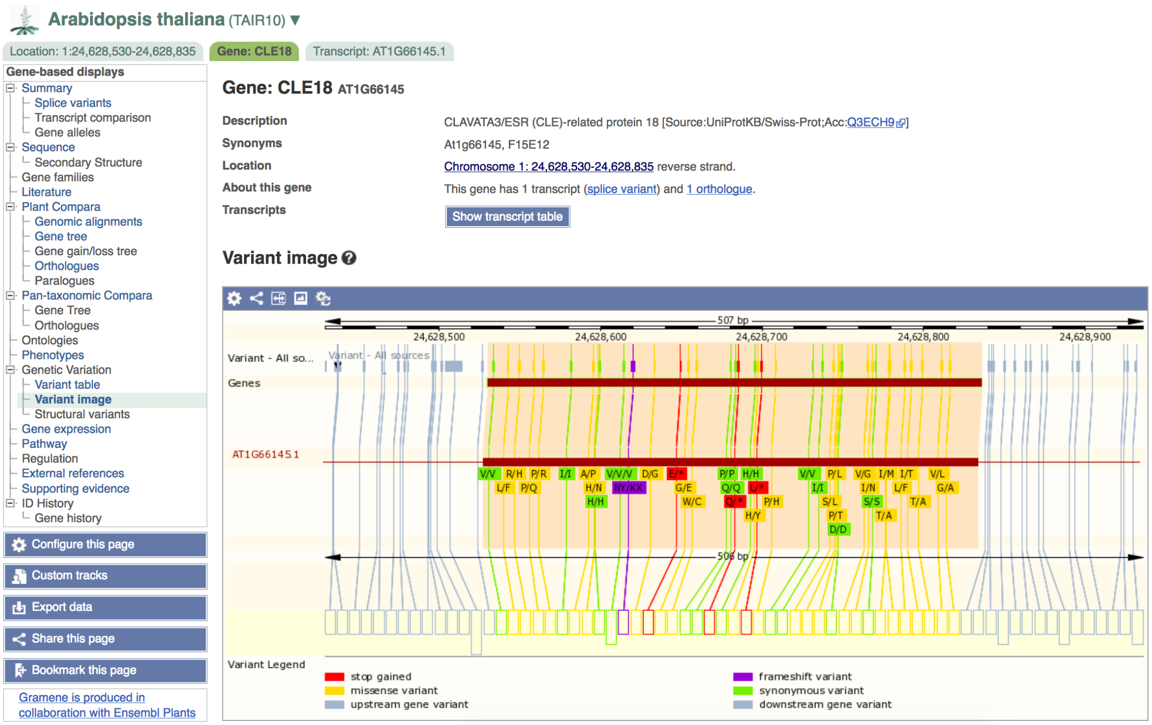
Gene stable ID	Interpro ID	Interpro Short Description	Variant name	Variant source	Chromosome/scaffold name	Chromosome/scaffold position start (bp)	Chromosome/scaffold position end (bp)	Variant consequence
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZ11D4DP	The 150 TGRSP	1	90352723	90352723	stop_gained
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZ11EPNJ	The 150 TGRSP	1	91009343	91009343	stop_gained
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZ11E0ZQ	The 150 TGRSP	1	90714271	90714271	stop_gained
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZ114T7X	The 150 TGRSP	1	86817174	86817174	stop_gained
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZ107LBN	The 150 TGRSP	1	70788055	70788055	stop_gained
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZ10UIAV	The 150 TGRSP	1	82292010	82292010	stop_gained
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZ110Q7T	The 150 TGRSP	1	85057092	85057092	stop_gained
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZ10G624	The 150 TGRSP	1	75234382	75234382	stop_gained
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZ10JMCB	The 150 TGRSP	1	76840415	76840415	stop_gained
Solyc01g103780.2	IPR017887	TF_TCP_subgr	vcZZYXQJ	The 150 TGRSP	1	66782860	66782860	stop_gained

Exercise 3. Explore the genetic variation associated with a gene

We will now explore genetic variants along the Arabidopsis *cle18* gene to find 2 SNPs reported to have drastic functional consequences for the CLE18 peptide. CLE18 is a CLAVATA3/ESR-related (CLE) peptide with diverse roles in plant growth and development. Two functional consequences were described by Cao *et al* (2011) [Nature Genetics].

a. Visualize the genetic variants for this gene
b. Are there any stop codons introduced (nonsense or stop gained variants) in this gene? Compare your findings with Supplementary Table 3 of Cao *et al* (2011)
c. Are there any transcript-specific variants for this gene?
d. Download a subset of the variants (*e.g.*, those that introduce an amino acid change in the protein)

To visualize the genetic variants in CLE18 (AT1G66145), simply select the "Variant image" view from the left menu bar of the corresponding gene page or go directly to http://ensembl.gramene.org/Arabidopsis_thaliana/Gene/Variation_Gene/Image?g=AT1G66145;r=1:24628530-24628835;t=AT1G66145.1



Gene: CLE18 AT1G66145

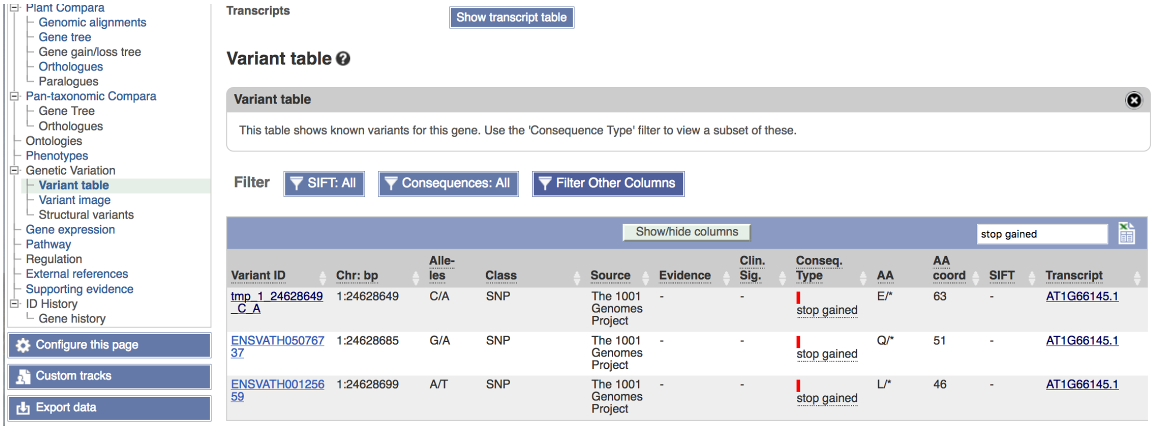
Description: CLAVATA3/ESR (CLE)-related protein 18 [Source:UniProtKB/Swiss-Prot;Acc:Q3ECH9]
Synonyms: At1g66145, F15E12
Location: Chromosome 1: 24,628,530-24,628,835 reverse strand.
About this gene: This gene has 1 transcript (splice variant) and 1 orthologue.
Transcripts: Show transcript table

Variant image

🟥 stop gained
🟨 missense variant
▨ upstream gene variant
🟪 frameshift variant
🟩 synonymous variant
▨ downstream gene variant

Answers:

There are three SNP variants that introduce stop codons (stop gained) in the CLE18 gene (TAIR10/AraPort11). Find these by going to the "Variant table" view of the CLE18 gene (http://ensembl.gramene.org/Arabidopsis_thaliana/Gene/Variation_Gene/Table?g=AT1G66145;r=1:24628530-24628835;t=AT1G66145.1) and filtering all variants using the word "stop gained" as shown in the figure below.



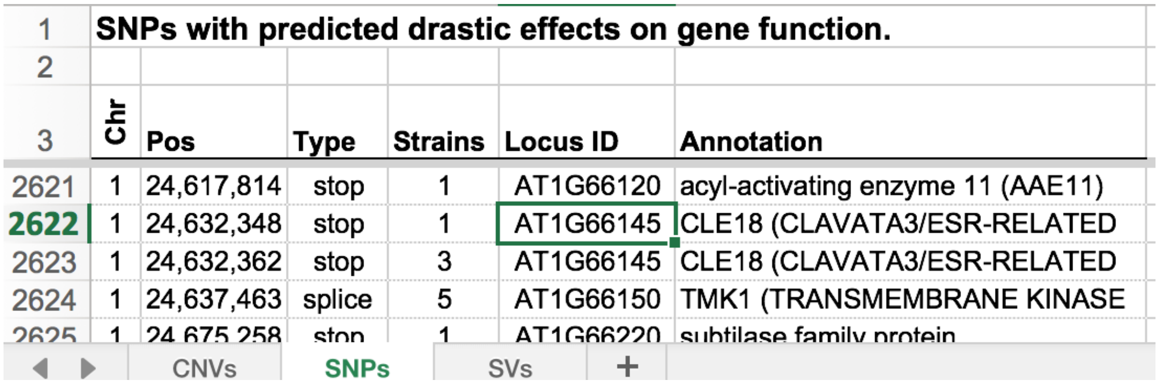
Variant table

Variant table

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

Variant ID	Chr: bp	Alle-les	Class	Source	Evidence	Clin. Sig.	Conseq. Type	AA	AA coord	SIFT	Transcript
tmp_1_24628649_C_A	1:24628649	C/A	SNP	The 1001 Genomes Project	-	-	stop gained	E/*	63	-	AT1G66145.1
ENSVATH05076737	1:24628685	G/A	SNP	The 1001 Genomes Project	-	-	stop gained	Q/*	51	-	AT1G66145.1
ENSVATH00125659	1:24628699	A/T	SNP	The 1001 Genomes Project	-	-	stop gained	L/*	46	-	AT1G66145.1

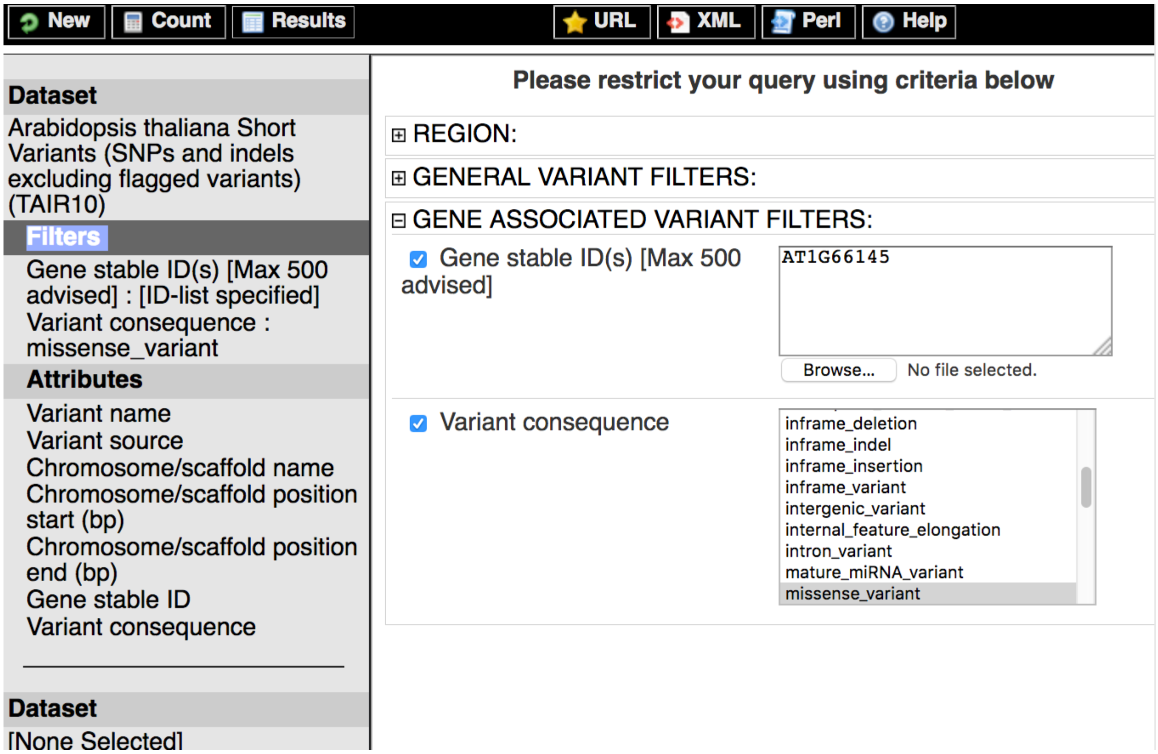
Here is the fragment of Supplementary Table 3 in Cao et al (2011) that lists the SNPs with predicted drastic effects in CLE18 (AT1G66145). The genomic coordinates clearly differ for these SNPs (there is no change when converted to TAIR9), however the relative distance between the two variants reported in the 2011 study is 14 bp, the exact same distance between ENSVATH05076737 and ENSVATH00125659.


CNVs
SNPs
SVs
+

1	**SNPs with predicted drastic effects on gene function.**					
2						
3	**Chr**	**Pos**	**Type**	**Strains**	**Locus ID**	**Annotation**
2621	1	24,617,814	stop	1	AT1G66120	acyl-activating enzyme 11 (AAE11)
2622	1	24,632,348	stop	1	AT1G66145	CLE18 (CLAVATA3/ESR-RELATED
2623	1	24,632,362	stop	3	AT1G66145	CLE18 (CLAVATA3/ESR-RELATED
2624	1	24,637,463	splice	5	AT1G66150	TMK1 (TRANSMEMBRANE KINASE
2625	1	24,675,258	stop	1	AT1G66220	subtilase family protein

No transcript-specific variants are seen in *cle18* because – so far – a single transcript has been identified for this gene.

To download a subset of the variants (*e.g.*, those that introduce an amino acid change in the protein), use the Gramene Mart to identify CLE18 variants with a specific predicted functional consequence (missense in this case).



Dataset

Arabidopsis thaliana Short Variants (SNPs and indels excluding flagged variants) (TAIR10)

Filters

Gene stable ID(s) [Max 500 advised] : [ID-list specified]
Variant consequence : missense_variant

Attributes

Variant name
Variant source
Chromosome/scaffold name
Chromosome/scaffold position start (bp)
Chromosome/scaffold position end (bp)
Gene stable ID
Variant consequence

Dataset

[None Selected]

Please restrict your query using criteria below

⊞ REGION:

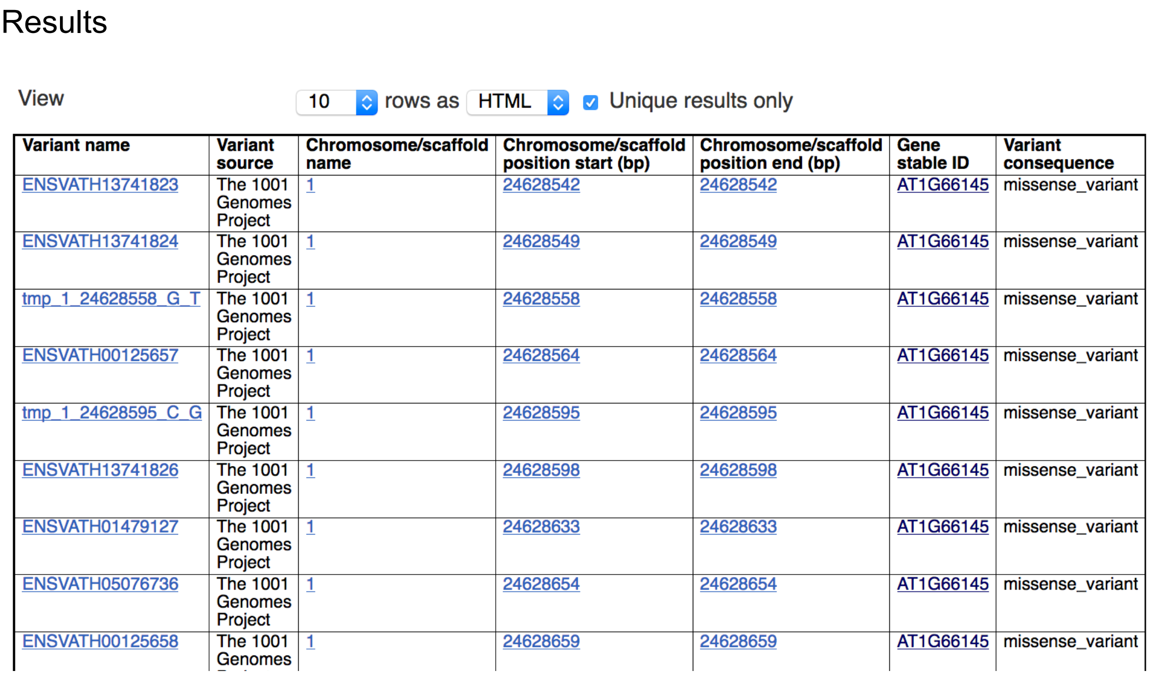
⊞ GENERAL VARIANT FILTERS:

⊟ GENE ASSOCIATED VARIANT FILTERS:

☑ Gene stable ID(s) [Max 500 advised]: AT1G66145

Browse... No file selected.

☑ Variant consequence



Results

View 10 rows as HTML ☑ Unique results only

Variant name	Variant source	Chromosome/scaffold name	Chromosome/scaffold position start (bp)	Chromosome/scaffold position end (bp)	Gene stable ID	Variant consequence
ENSVATH13741823	The 1001 Genomes Project	1	24628542	24628542	AT1G66145	missense_variant
ENSVATH13741824	The 1001 Genomes Project	1	24628549	24628549	AT1G66145	missense_variant
tmp_1_24628558_G_T	The 1001 Genomes Project	1	24628558	24628558	AT1G66145	missense_variant
ENSVATH00125657	The 1001 Genomes Project	1	24628564	24628564	AT1G66145	missense_variant
tmp_1_24628595_C_G	The 1001 Genomes Project	1	24628595	24628595	AT1G66145	missense_variant
ENSVATH13741826	The 1001 Genomes Project	1	24628598	24628598	AT1G66145	missense_variant
ENSVATH01479127	The 1001 Genomes Project	1	24628633	24628633	AT1G66145	missense_variant
ENSVATH05076736	The 1001 Genomes Project	1	24628654	24628654	AT1G66145	missense_variant
ENSVATH00125658	The 1001 Genomes	1	24628659	24628659	AT1G66145	missense_variant

Note: In addition to the Ensembl "Tools" for genomic analysis (e.g., BLAST, BioMart, Assembly Converter, Variant Effect Predictor), other genetic analysis (*e.g.*, Simple Sequence Repeat Identification Tool or SRIT) tools can be accessed through Gramene's archival Diversity pages at http://archive.gramene.org/diversity/tools.html

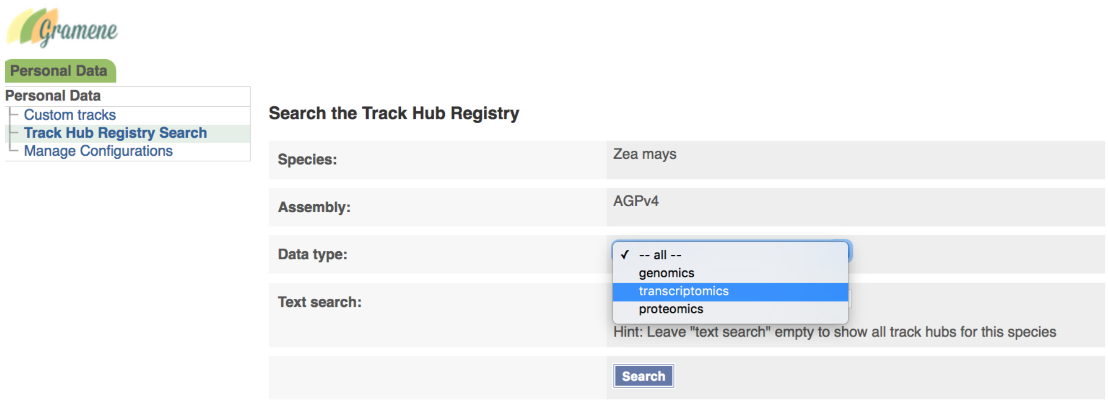
Exercise 4. Upload, visualize and share your own data into a new genome browser track

The Ensembl genome browser allows users to upload their own data and visualize it on a custom track and/or analyze it with various tools like the BLAST sequence alignment tool, the Assembly Converter tool, the Variant Effect Prediction (VEP) tool, etc. Data may be formatted in various file formats including FASTA, GFF, GTF, BED, BAM, VCF, bedGraph, gbrowse, PSL, WIG, BigBed, BigWig, and TrackHub. Some data like GFF annotations may be directly uploaded from a local machine. Large data files like BED/BAM alignments or BigWig graphic display configurations need to be uploaded onto a local server that is accessible to the browser via an URL. Another way to share third-party data is via a DAS (Distributed Annotation System) registry, which would need to be set up by a software engineer.

The test data sets that are available for upload and visualization for this exercise have been preloaded onto a local server that is publicly accessible: http://data.gramene.org/public/Zea_mays4m/methylome. The data consists of BAM alignments and CpG methylation for B73 & Mo17 maize lines used in the study by Regulski *et al* (2013) [Genome Research 23:1651] and were used to create expression tracks in Gramene build 56.

1) Go to a genomic region of your choice (Location View, e.g., http://ensembl.gramene.org/Zea_mays/Location/View?r=1:113755522-113782806;db=core;time=1521250985769.769). Click on "Custom tracks" for pop-up window to appear.

Note: This pop-up window also allows you to load public RNA-Seq data from ***The Track Hub!*** *Simply click on "Track Hub Registry Search" instead and proceed to find the type of omics data for the species you are interested in, as so:*



Search the Track Hub Registry

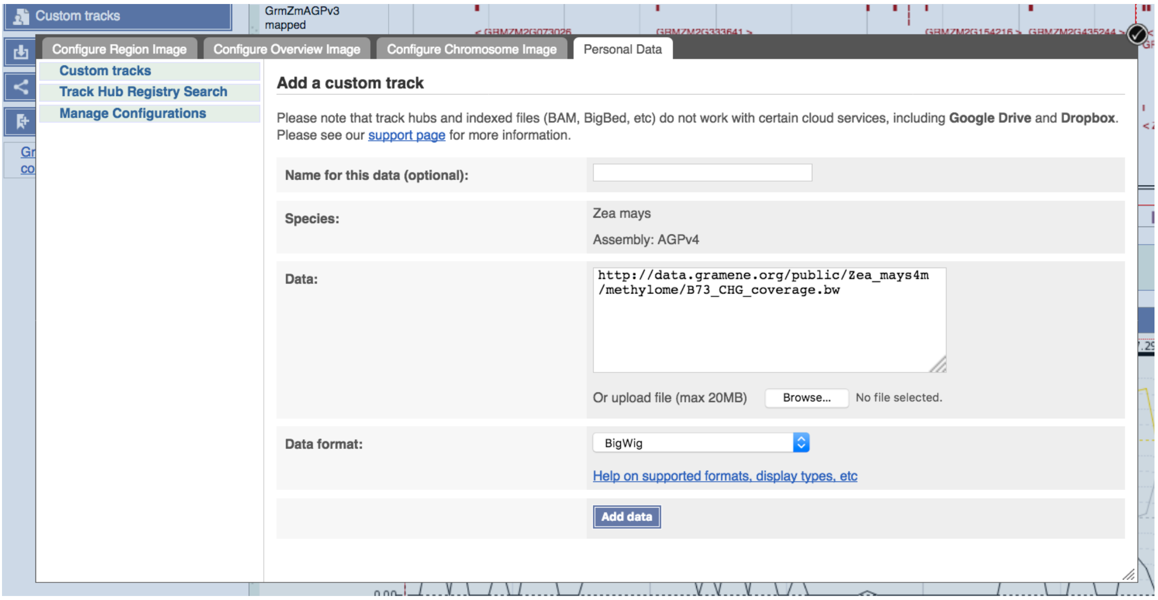
Species:	Zea mays
Assembly:	AGPv4
Data type:	✓ -- all -- genomics transcriptomics proteomics
Text search:	Hint: Leave "text search" empty to show all track hubs for this species
	Search

2) To add a custom track, paste custom data in valid format or enter the corresponding URL (see methylome data files in local server above) in the

“Data” field (*e.g.*, http://data.gramene.org/public/Zea_mays4m/methylome/B73_CHG.bw) and click on “Add data”.

Custom tracks → Add a custom track or Add more data

3) Click on the check mark (top right corner) to close the window. You may need to navigate to a region that contains your data to visualize it on the screen and make sure that your custom track is turned on.

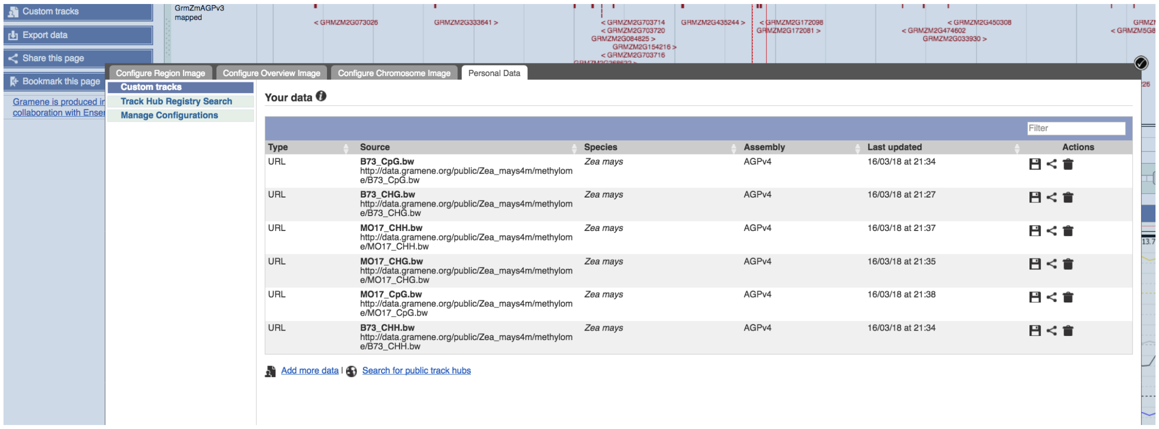


Add a custom track

Please note that track hubs and indexed files (BAM, BigBed, etc) do not work with certain cloud services, including **Google Drive** and **Dropbox**. Please see our support page for more information.

Name for this data (optional):	
Species:	Zea mays Assembly: AGPv4
Data:	http://data.gramene.org/public/Zea_mays4m/methylome/B73_CHG_coverage.bw Or upload file (max 20MB) Browse... No file selected.
Data format:	BigWig Help on supported formats, display types, etc

Custom data sets loaded

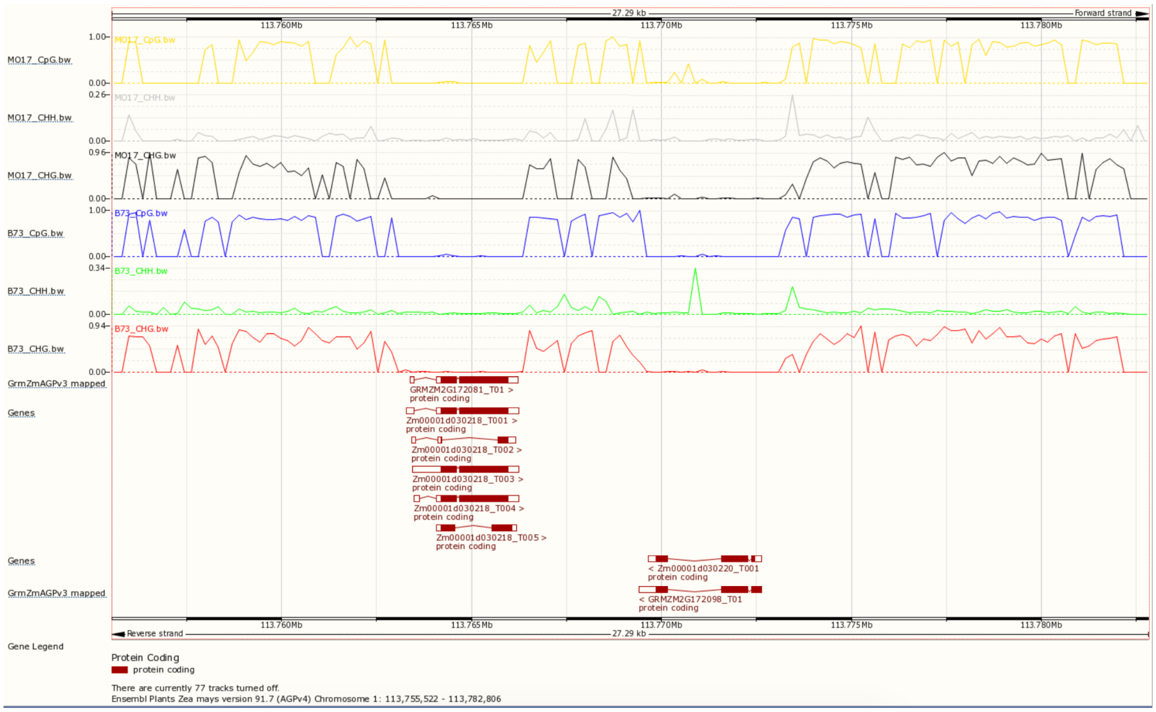


Your data

Type	Source	Species	Assembly	Last updated	Actions
URL	**B73_CpG.bw** http://data.gramene.org/public/Zea_mays4m/methylome/B73_CpG.bw	*Zea mays*	AGPv4	16/03/18 at 21:34	
URL	**B73_CHG.bw** http://data.gramene.org/public/Zea_mays4m/methylome/B73_CHG.bw	*Zea mays*	AGPv4	16/03/18 at 21:27	
URL	**MO17_CHH.bw** http://data.gramene.org/public/Zea_mays4m/methylome/MO17_CHH.bw	*Zea mays*	AGPv4	16/03/18 at 21:37	
URL	**MO17_CHG.bw** http://data.gramene.org/public/Zea_mays4m/methylome/MO17_CHG.bw	*Zea mays*	AGPv4	16/03/18 at 21:35	
URL	**MO17_CpG.bw** http://data.gramene.org/public/Zea_mays4m/methylome/MO17_CpG.bw	*Zea mays*	AGPv4	16/03/18 at 21:38	
URL	**B73_CHH.bw** http://data.gramene.org/public/Zea_mays4m/methylome/B73_CHH.bw	*Zea mays*	AGPv4	16/03/18 at 21:34	

Add more data | Search for public track hubs

Custom tracks of methylation signatures data

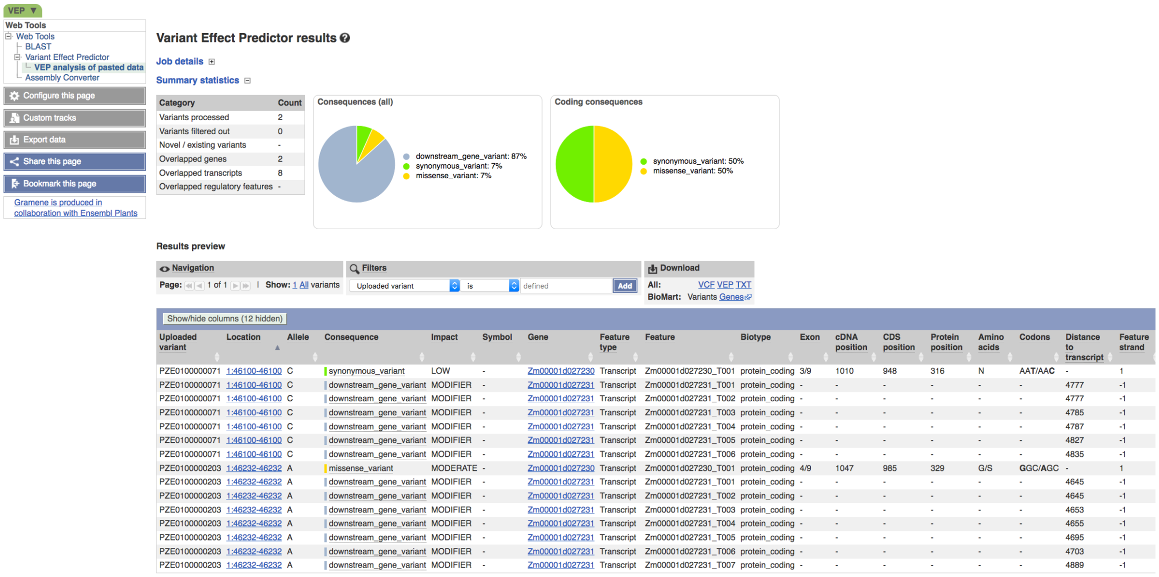


There are currently 77 tracks turned off.
Ensembl Plants Zea mays version 91.7 (AGPv4) Chromosome 1: 113,755,522 - 113,782,806

Predict functional effects in custom SNP data sets using the VEP tool

Copy/paste the following sample maize VCF in the VEP Tool (http://ensembl.gramene.org/Zea_mays/Tools/VEP?db=core) to visualize the predicted functional consequence for the SNPs listed. Results are shown below.

```
##fileformat=VCFv4.0
##fileDate=20161018
##source=MaizeHapMapMockUp
##reference=RefGenv4
##phasing=no
##INFO=<ID=MQ,Number=.,Type=Float,Description="RMS mapping quality">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
#CHROM  POS     ID      REF     ALT     QUAL    FILTER  INFO    FORMAT
B73:MZ  M97:MZ   MKN009:MZ       MKN010:MZ       MKN011:MZ
1       46100     PZE0100000071   T       C       .       PASS    MQ=92   GT      0/0
0/0     0/0     0/0     0/0
1       46232     PZE0100000203   G       A       .       PASS    MQ=91   GT      0/0
0/0     0/0     0/0     0/0
```



VEP ▼

Web Tools
- Web Tools
  - BLAST
  - Variant Effect Predictor
    - VEP analysis of pasted data
  - Assembly Converter

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

Variant Effect Predictor results

Job details ⊞

Summary statistics ⊟

Category	Count
Variants processed	2
Variants filtered out	0
Novel / existing variants	-
Overlapped genes	2
Overlapped transcripts	8
Overlapped regulatory features	-

Consequences (all)

- downstream_gene_variant: 87%
- synonymous_variant: 7%
- missense_variant: 7%

Coding consequences

- synonymous_variant: 50%
- missense_variant: 50%

Results preview

Navigation
Page: 1 of 1 | Show: 1 All variants

Filters
Uploaded variant | is | defined | Add

Download
All: VCF VEP TXT
BioMart: Variants Genes

Show/hide columns (12 hidden)

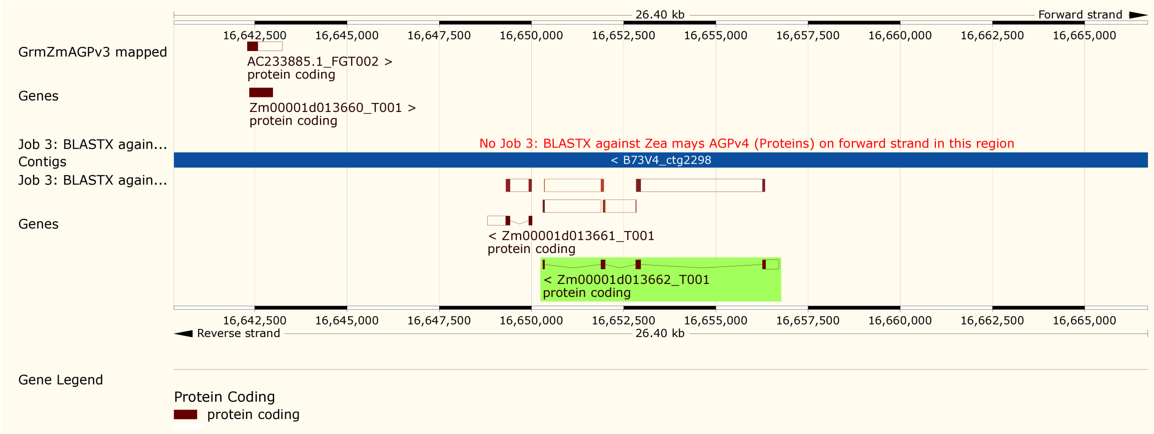
Uploaded variant	Location	Allele	Consequence	Impact	Symbol	Gene	Feature type	Feature	Biotype	Exon	cDNA position	CDS position	Protein position	Amino acids	Codons	Distance to transcript	Feature strand
PZE0100000071	1:46100-46100	C	synonymous_variant	LOW	-	Zm00001d027230	Transcript	Zm00001d027230_T001	protein_coding	3/9	1010	948	316	N	AAT/AAC	-	1
PZE0100000071	1:46100-46100	C	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T001	protein_coding	-	-	-	-	-	-	4777	-1
PZE0100000071	1:46100-46100	C	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T002	protein_coding	-	-	-	-	-	-	4777	-1
PZE0100000071	1:46100-46100	C	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T003	protein_coding	-	-	-	-	-	-	4785	-1
PZE0100000071	1:46100-46100	C	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T004	protein_coding	-	-	-	-	-	-	4787	-1
PZE0100000071	1:46100-46100	C	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T005	protein_coding	-	-	-	-	-	-	4827	-1
PZE0100000071	1:46100-46100	C	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T006	protein_coding	-	-	-	-	-	-	4835	-1
PZE0100000203	1:46232-46232	A	missense_variant	MODERATE	-	Zm00001d027230	Transcript	Zm00001d027230_T001	protein_coding	4/9	1047	985	329	G/S	GGC/AGC	-	1
PZE0100000203	1:46232-46232	A	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T001	protein_coding	-	-	-	-	-	-	4645	-1
PZE0100000203	1:46232-46232	A	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T002	protein_coding	-	-	-	-	-	-	4645	-1
PZE0100000203	1:46232-46232	A	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T003	protein_coding	-	-	-	-	-	-	4653	-1
PZE0100000203	1:46232-46232	A	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T004	protein_coding	-	-	-	-	-	-	4655	-1
PZE0100000203	1:46232-46232	A	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T005	protein_coding	-	-	-	-	-	-	4695	-1
PZE0100000203	1:46232-46232	A	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T006	protein_coding	-	-	-	-	-	-	4703	-1
PZE0100000203	1:46232-46232	A	downstream_gene_variant	MODIFIER	-	Zm00001d027231	Transcript	Zm00001d027231_T007	protein_coding	-	-	-	-	-	-	4889	-1

Exercise 5. BLAST a sequence. Determine synteny for a genomic region. Convert coordinates between different genome assemblies.

In this exercise, we will identify orthologues of a species whose reference genome is not available in Gramene via BLASTX and find corresponding synteny blocks in other species.

a. Use the nucleotide sequence of the *Sorghum virgatum* Sh1 gene taken from Lin *et al* (2012) [Nature Genetics 44:720] to identify orthologous genes in *Sorghum bicolor* and maize.

Answer:
The best hit in Sorghum bicolor (orthologue) is SORBI_3001G152901
The best hit in Zea mays (orthologue) maps to the C2C2-YABBY-transcription factor 6 (Zm00001d013662) gene region, which appears to be a “split gene” (annotation artifact) together with Zm00001d013661.



Work on your own to answer the following. We are always willing to help! ☺
E-mail us at feedback@gramene.org

b. Highlight the orthologues in two of those species in the tree as you learned in Exercise 1.

c. Download the genetic variation for one of the maize Sh1 orthologues as you learned in Exercise 2. How many nonsense substitutions can you find in this gene?

d. Lin *et al* (2012) also provide RefGen_v2 coordinates for maize shattering QTLs in Supplementary Table 5. Identify synteny blocks for the intervals at maize chromosomal regions (RefGen_v2) chr1: 259,223,260 - 261,622,457 and chr5: 15,806,322 - 16,428,681 in rice and sorghum. Download the synteny images that you generate.

Note: You will need to first use the Assembly converter tool to map the QTL intervals to RefGen_v4 coordinates.

e. Download all the genes for a given synteny block. Can you identify a *Sh1* orthologous (YABBY-like) gene in it?

f. Compare your results with those in *Lin et al* (2012)

>*S. virgatum* Sh1 CDS

```
ATGTCGGCCCAGCAGATCGCGCCGGTGCCGGAGCATGTGTGCTACGTGCA
CTGCAACTTCTGCAATACAATTCTCGCGGTCAGTGTCCCGAGTCACAGCAT
GCTGAACATCGTGACAGTCCGTTGTGGGCACTGCACTAGCCTGCTGTCAGT
GAACTTGAGAGGACTCCTCCAATCACTCCCTGTCCAGAATCACTACTCGCA
GGAGAATAATTTCAAGGTCCAAAATTTCAGCTTTACTGAAAACTACCCTGAG
TATGCACCTTCGTCTTCGAAATACCGCATGCCAACGATGTTGTCAGCAAAAG
GTGATCTGGATCATATGCTGCACGTGCGTGGTAAGCTCCAGAGAAGAGGCA
ACGTGTTCCTTCAGCATATAACAGATTTATTAAGGAAGAGATACGAAGGATT
AAAGCAAGCAACCCAGACATAAGCCACAGGGAAGCCTTCAGCACTGCAGCA
AAAAATTGGGCACATTTTCCAAACATTCATTTTGGACTAGGGCCCTATGAAA
GTAGCAACAAGCTTGATGAGGCCATTGGTGCAACGGGCCATCCCCAGAAAG
```

TCCAAGATCTCTACTAA