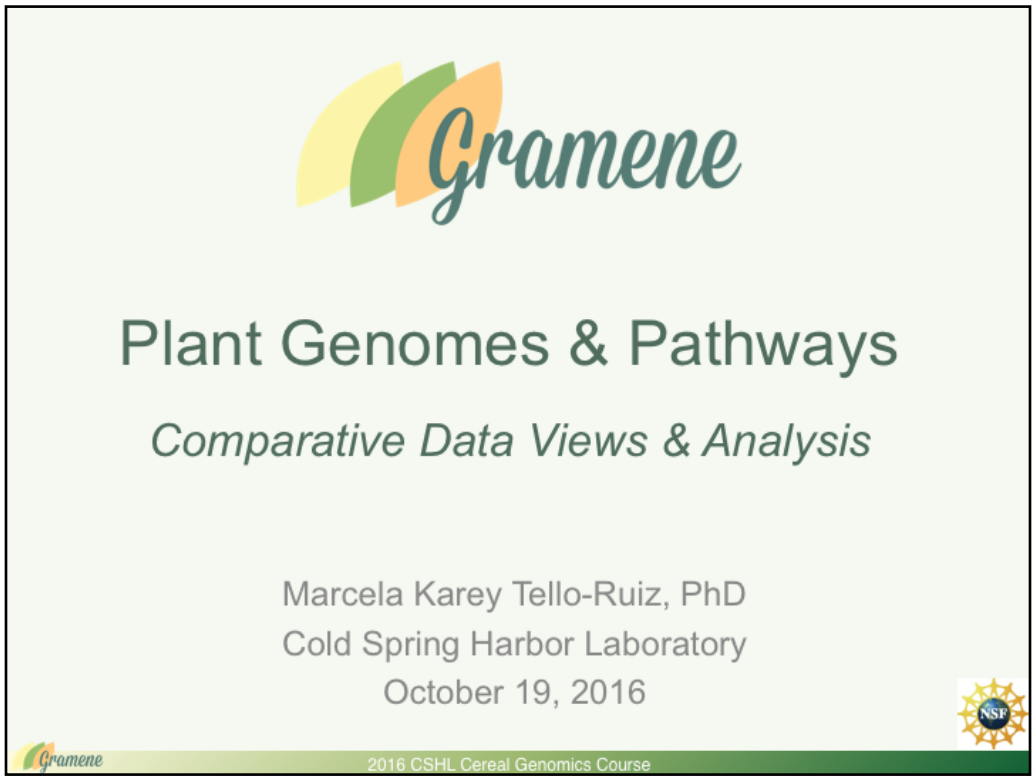
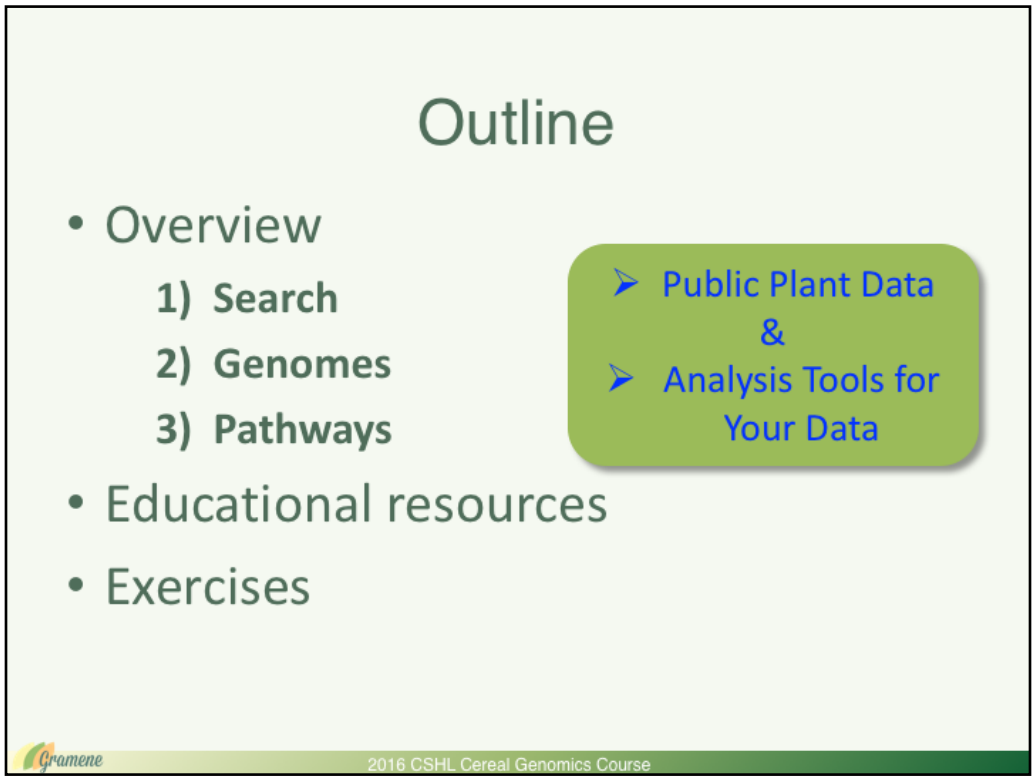

Gramene
Plant Genomes & Pathways
Comparative Data Views & Analysis
Marcela Karey Tello-Ruiz, PhD
Cold Spring Harbor Laboratory
October 19, 2016
NSF
Gramene
2016 CSHL Cereal Genomics Course



Genomes

Reference sequence assemblies & comparative genomics (synteny, WGAs)

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

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

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

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

Pathways

- Pathway browser
- Omics analysis
- Species comparison

Expression

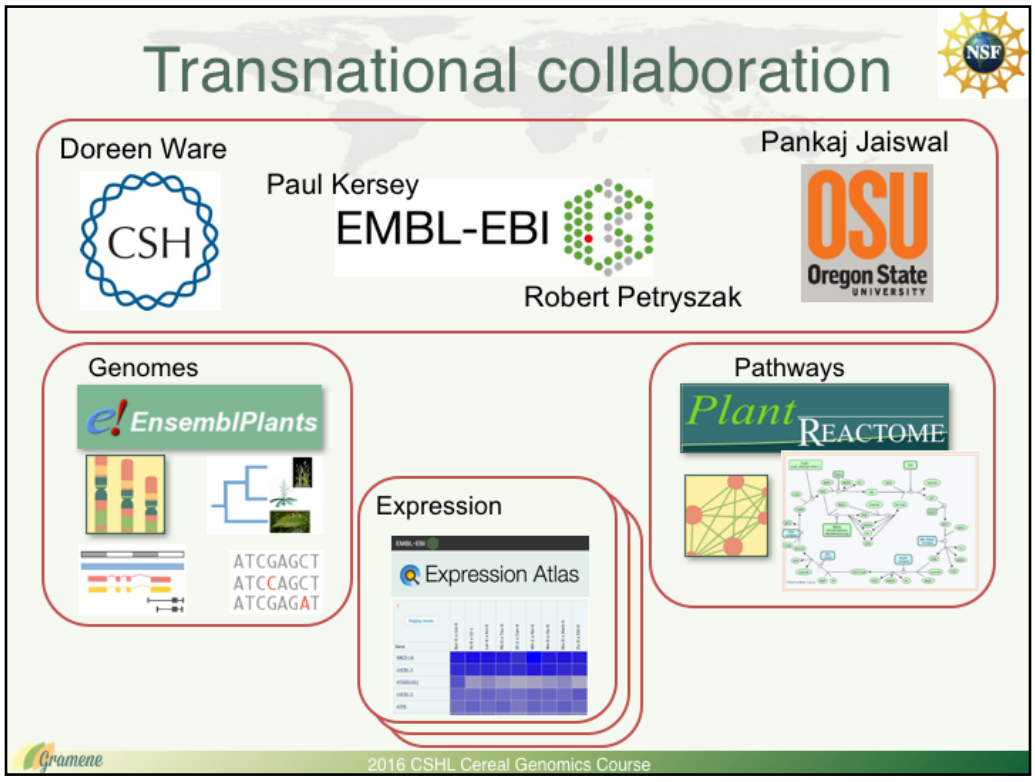
- ATLAS widget (genome browser & pathways views)

Videotutorials in the Gramene YouTube channel: https://goo.gl/ln9RLD

Use case: Explore functional variants in *cle18*

Search entry points

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



<Define Gramene>

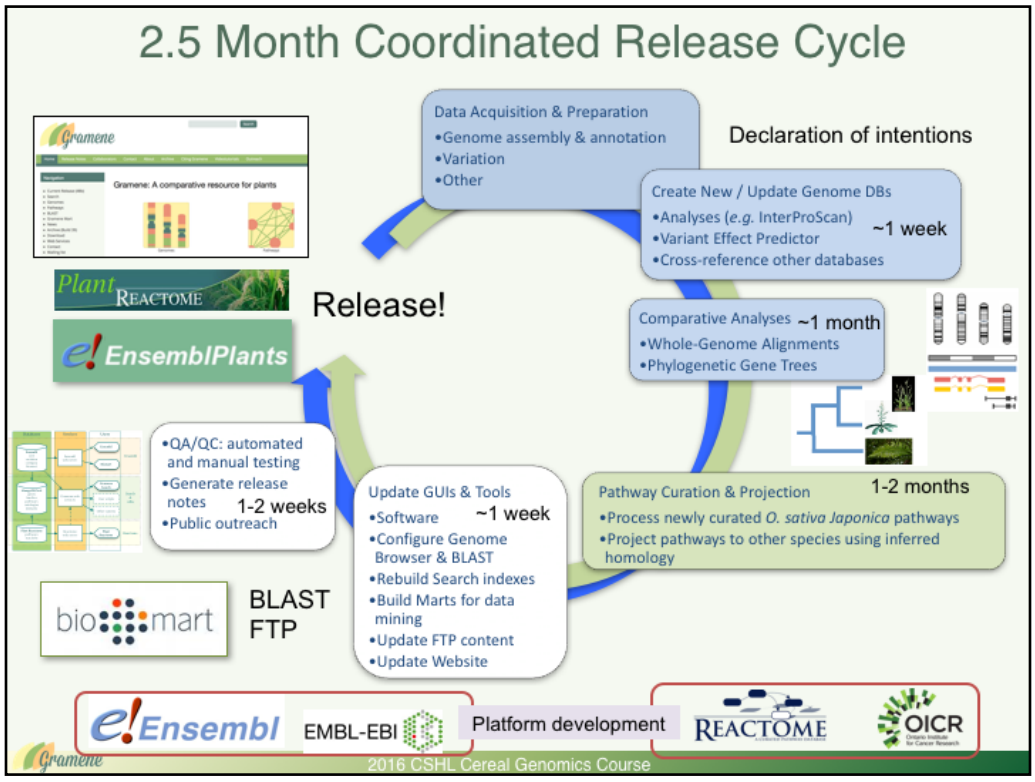
The Gramene project encompasses a set of best-in-class software tools that specialize in subsets of functionality.

Most significantly:

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

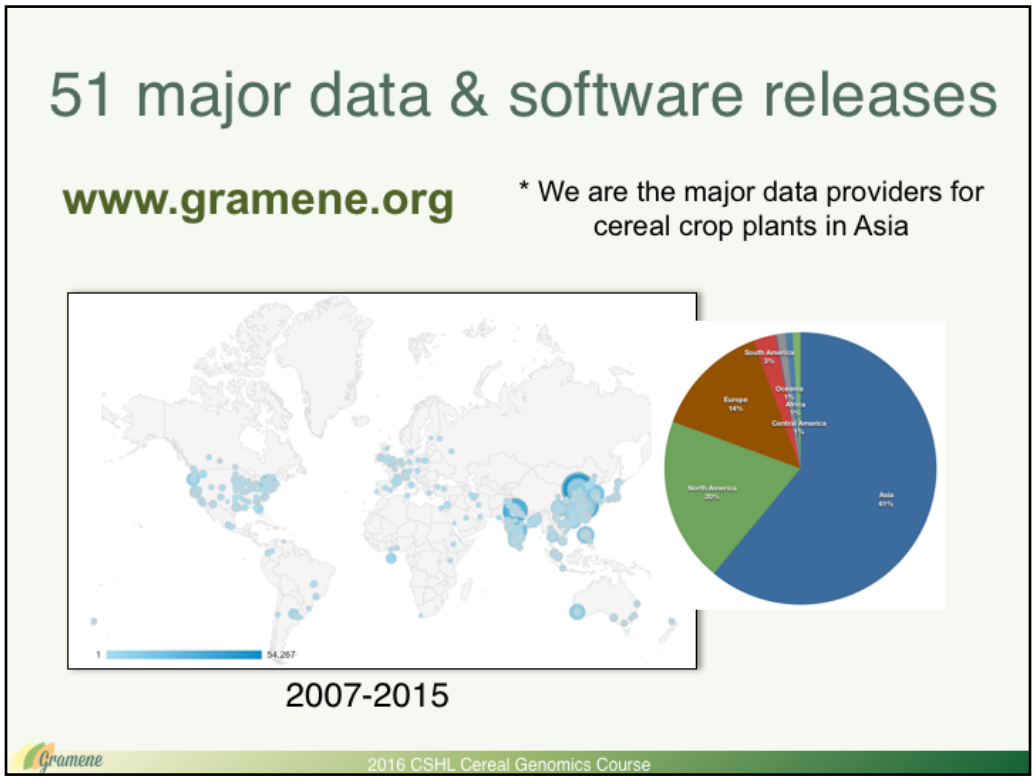
And also:

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



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

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

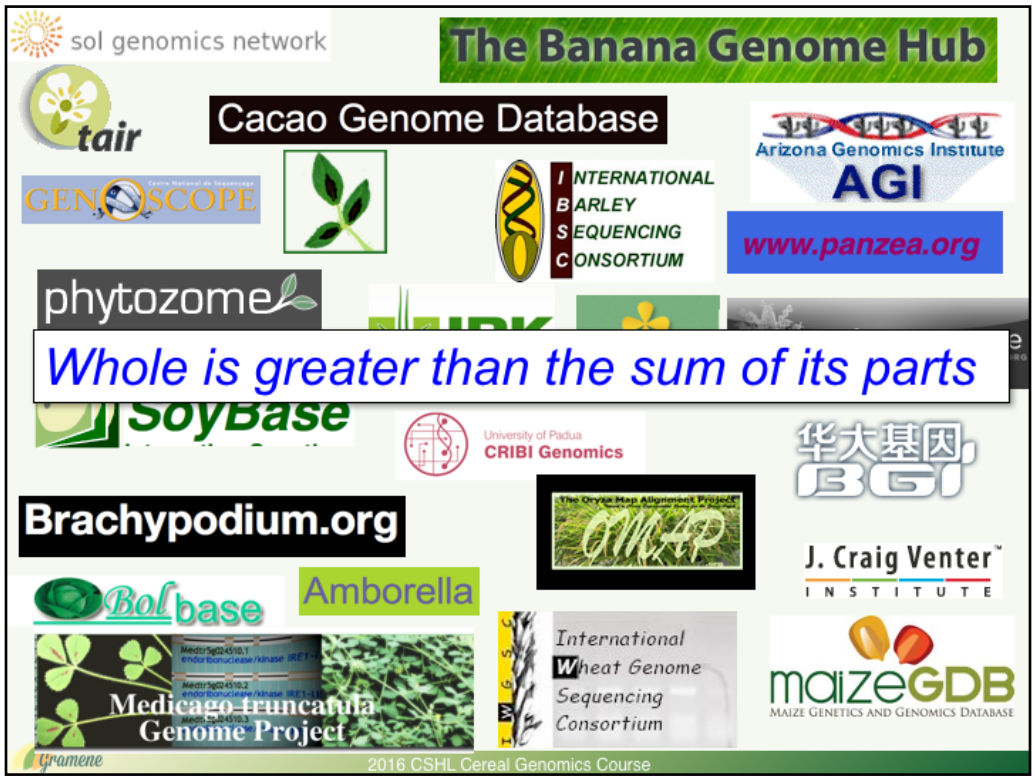


Other individuals/projects accessing Gramene using web services (API) not included in these counts

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

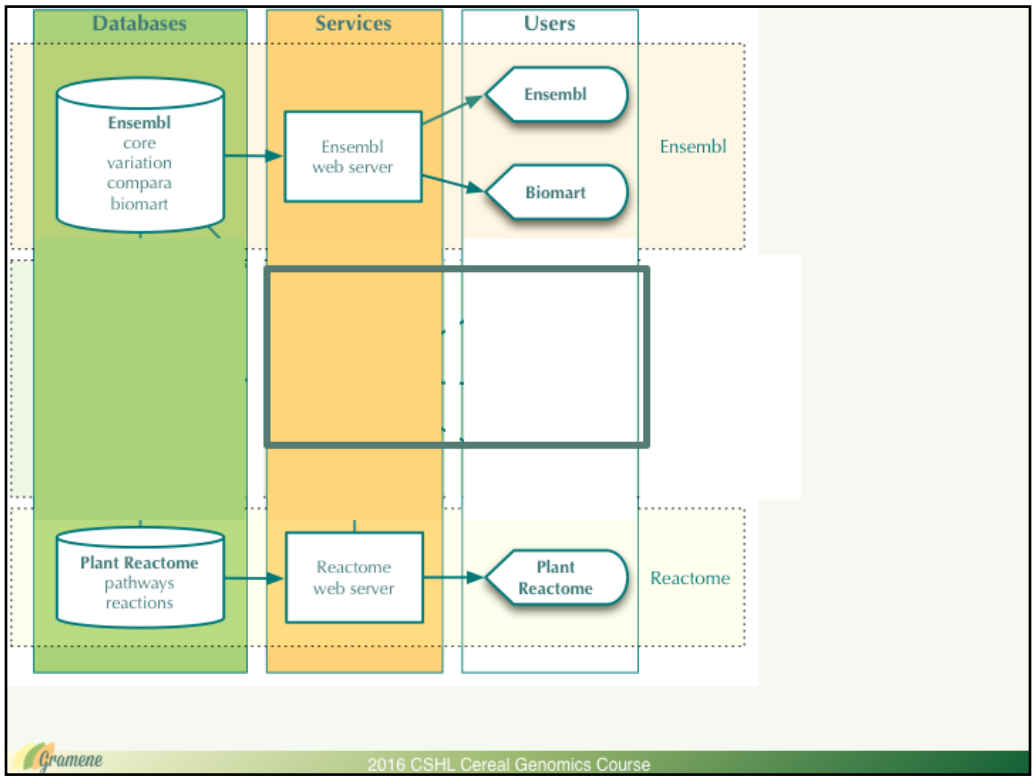
Pie chart: Top Countries - Visits% January 2016 – June 2016 by Google Analytics

Mapmundi 2007-2015



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

SEARCH



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

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

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

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

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

Goals

to enable complex questions to be asked of gramene's COMPARATIVE dataset *simply*

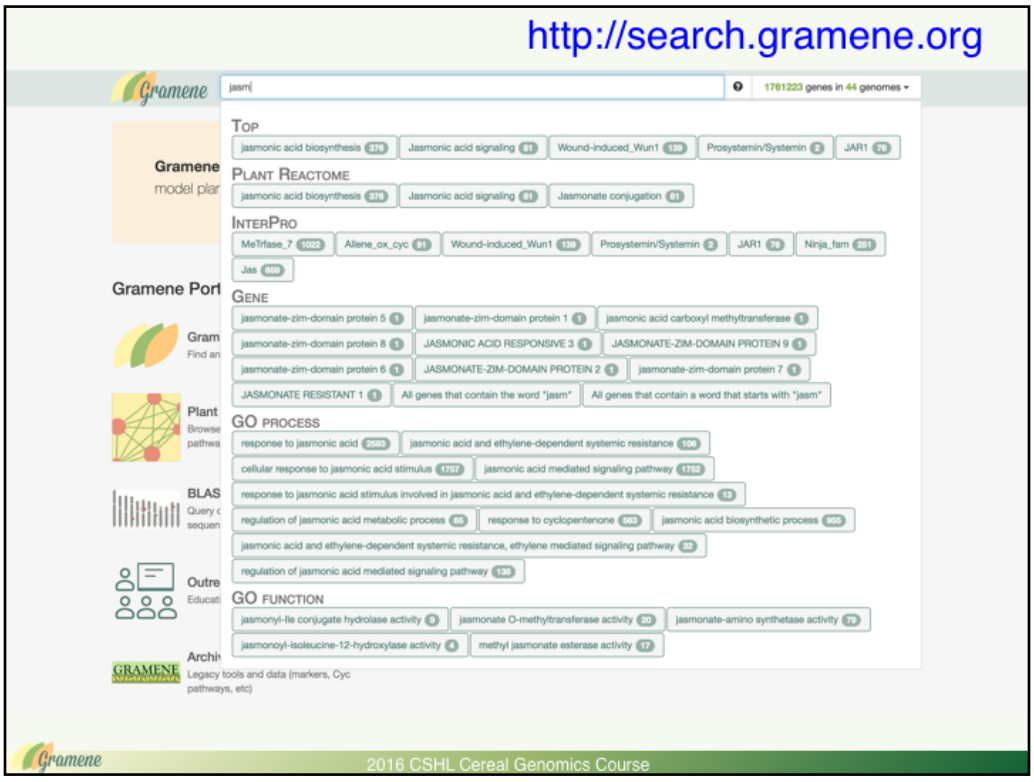
to encourage the asking of relatively broad questions and gaining meaningful insights for large result sets

to encorporate all gramene's datasets with those of relevant third parties

Key features of this application:

* Suggestions
* Filters

* Visualization



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

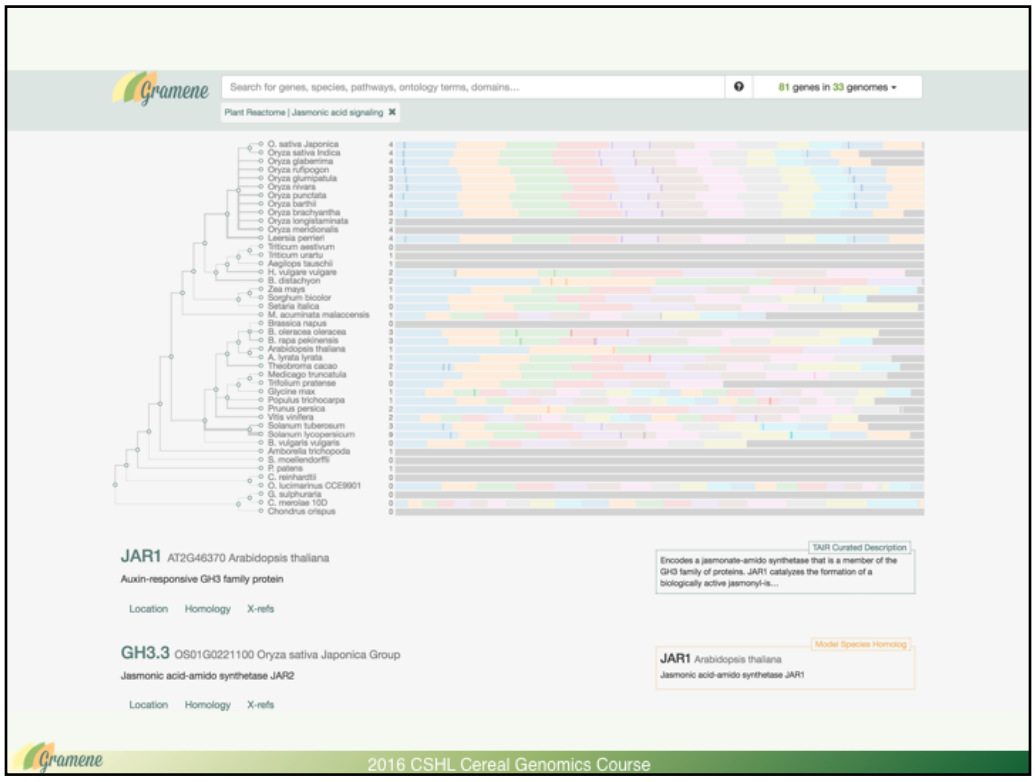
Focused search engines:

1) Genomes Browser
2) Plant Reactome

at search.gramene.org

Start typing jasmonic acid signaling

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



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

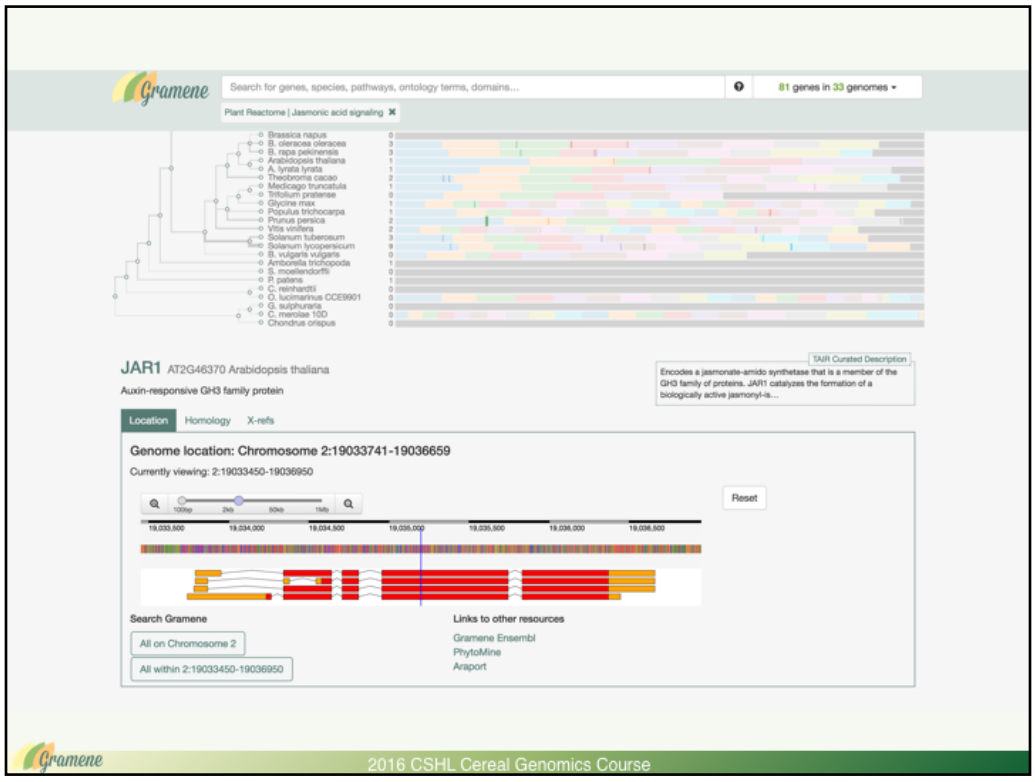
Taxagenomic distribution shows species tree (based on ncbi taxonomy)

Count of genes in each species

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

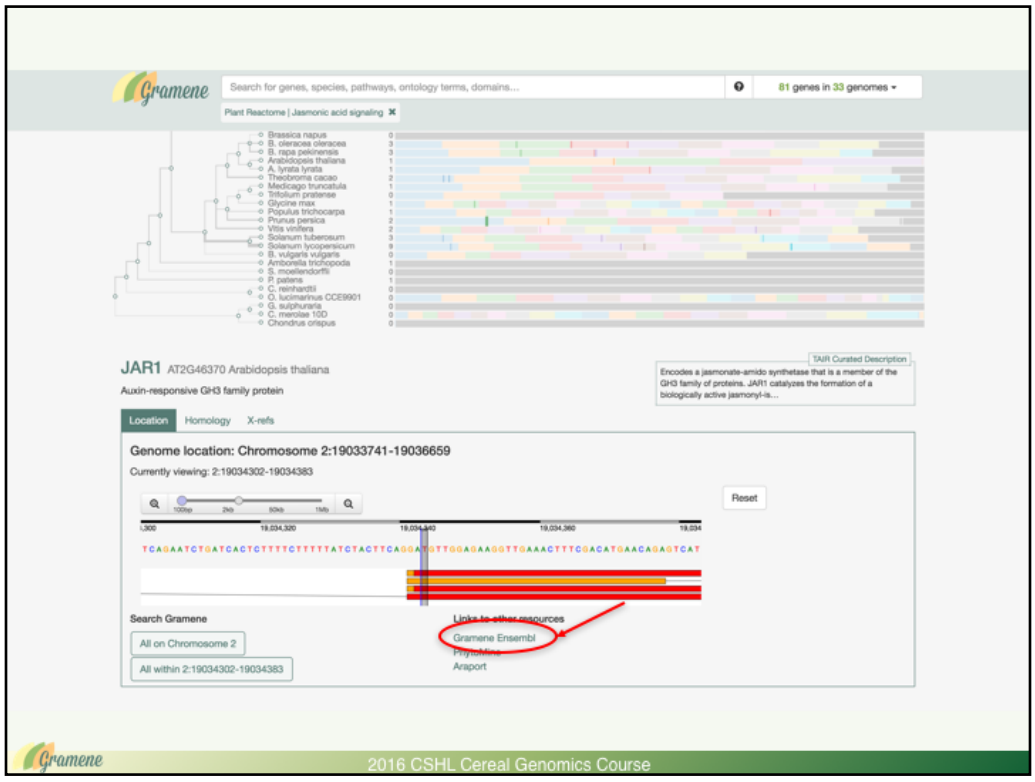
Below is a list view – arabidopsis results appear first because they tend to be best annotated

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

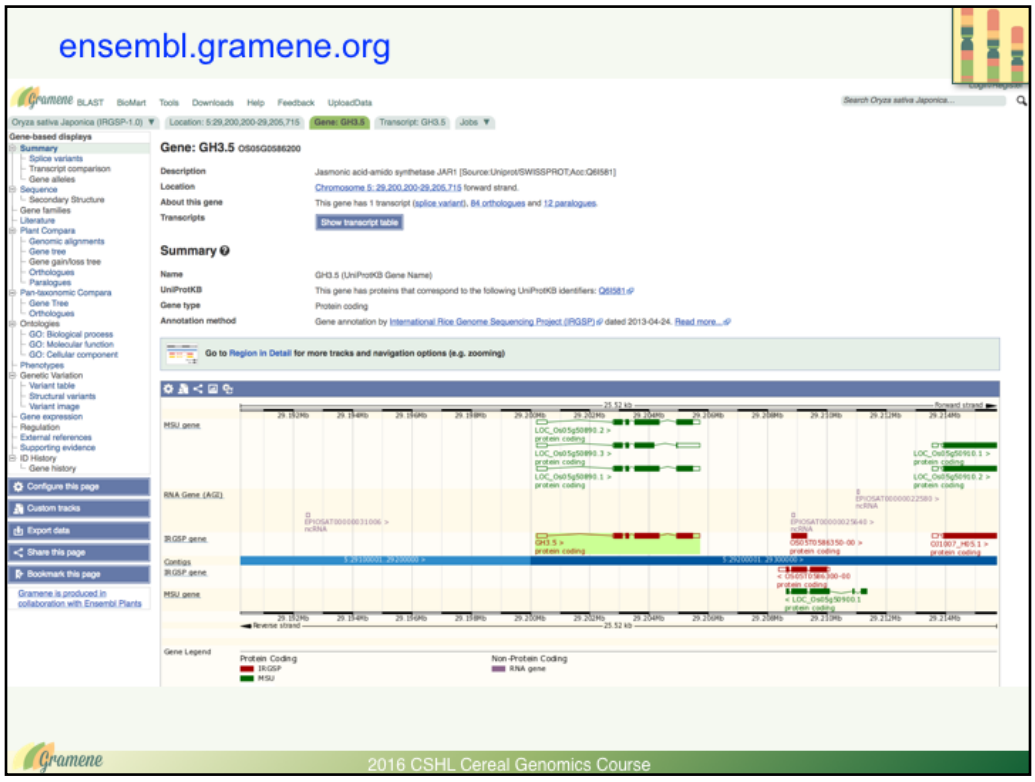


The location view shows a lightweight genome browser that lets you see the annotated gene structure.

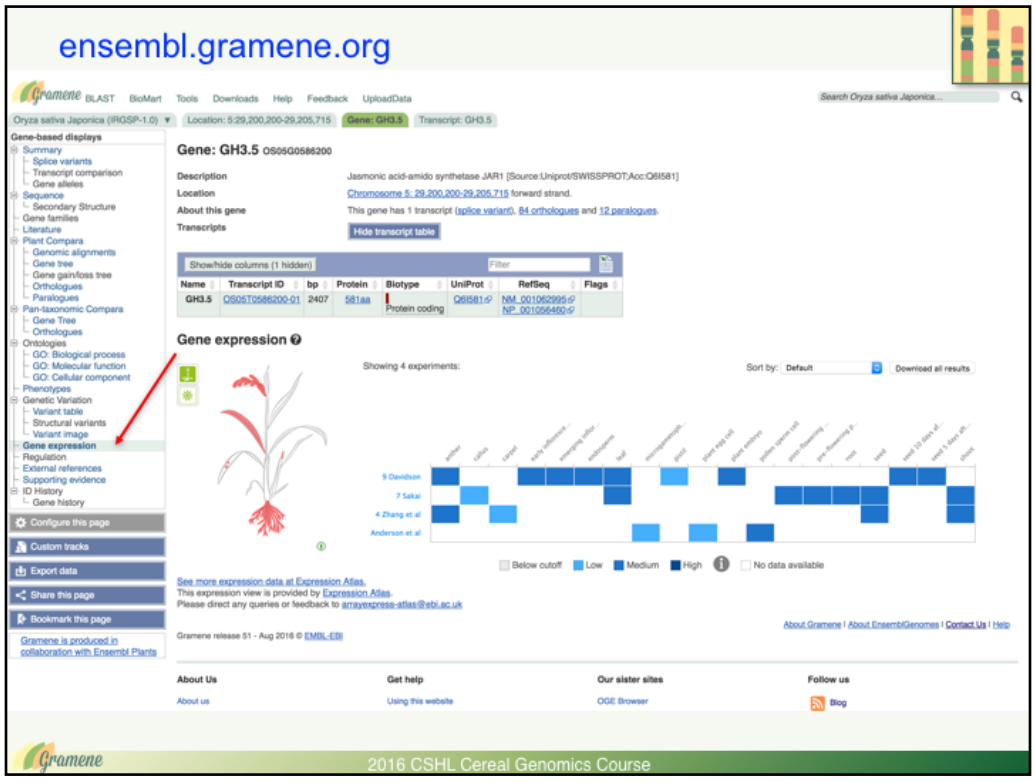
You can scroll and zoom in/out



Zoom to 100bp zoom level to see splice acceptor sequence and start codon

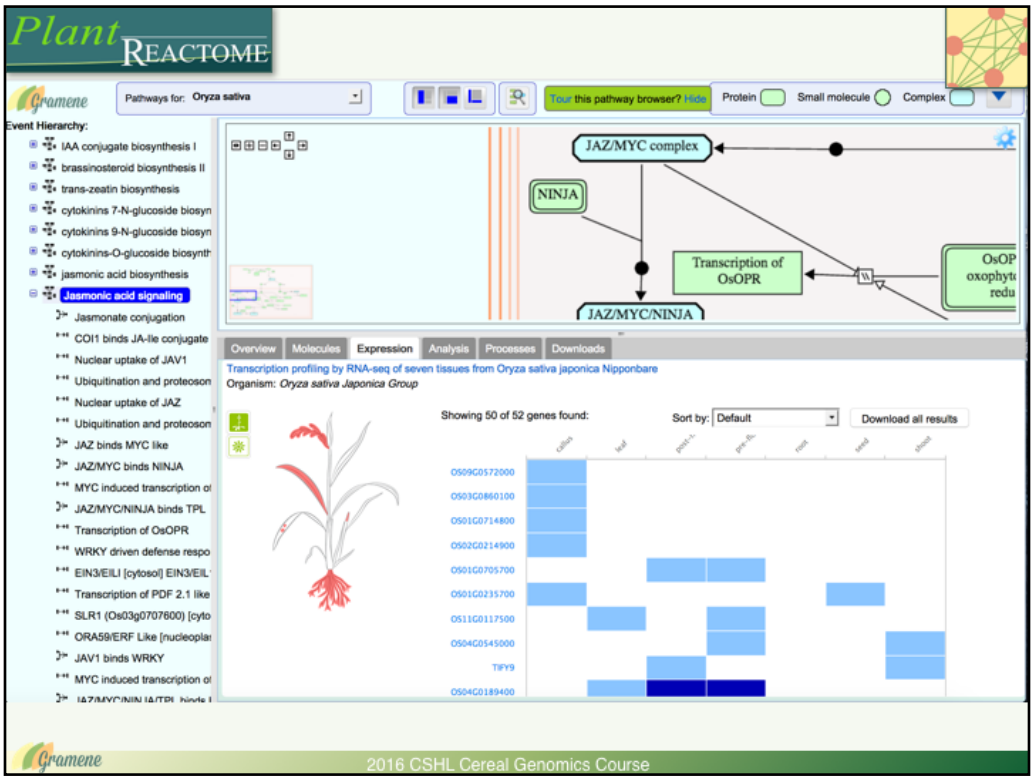


The ensembl gene page (for a different gene in the jasmonic signaling pathway)

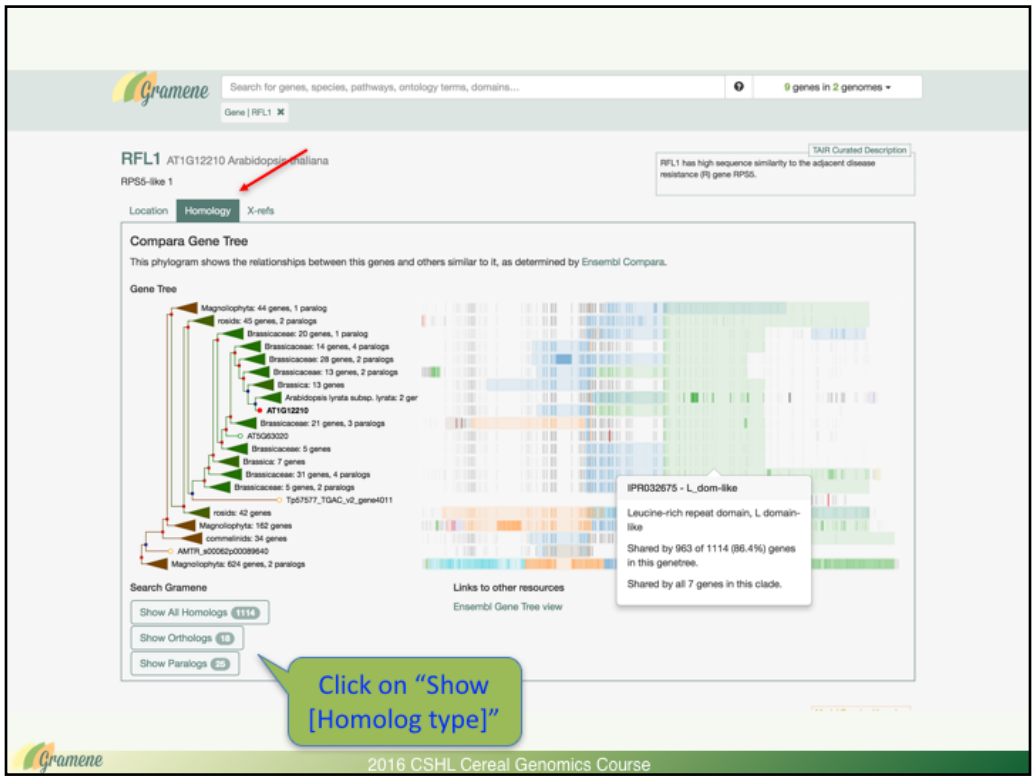


Click Gene expression in the sidebar

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

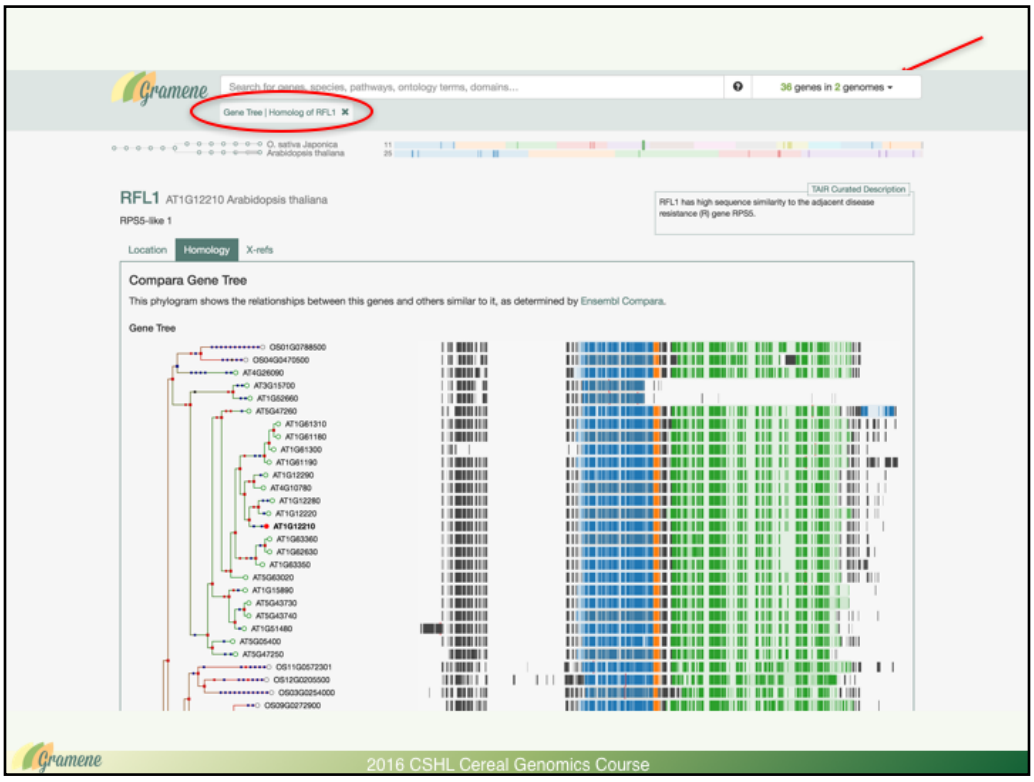


Same data at the pathway level available from Plant Reactome



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

Click on Show All Homologs to update the search filter.



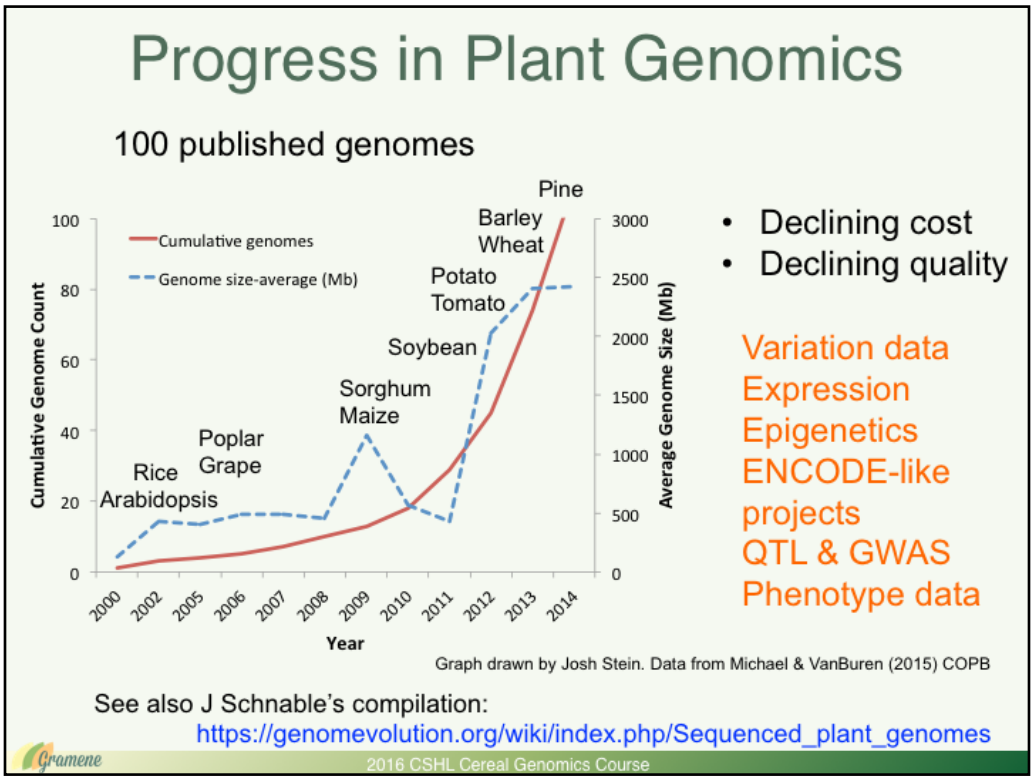
If you change the genomes you are searching to just rice and Arabidopsis, the search results are filtered down to the 36 rice and Arabidopsis genes in the gene family.

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

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

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

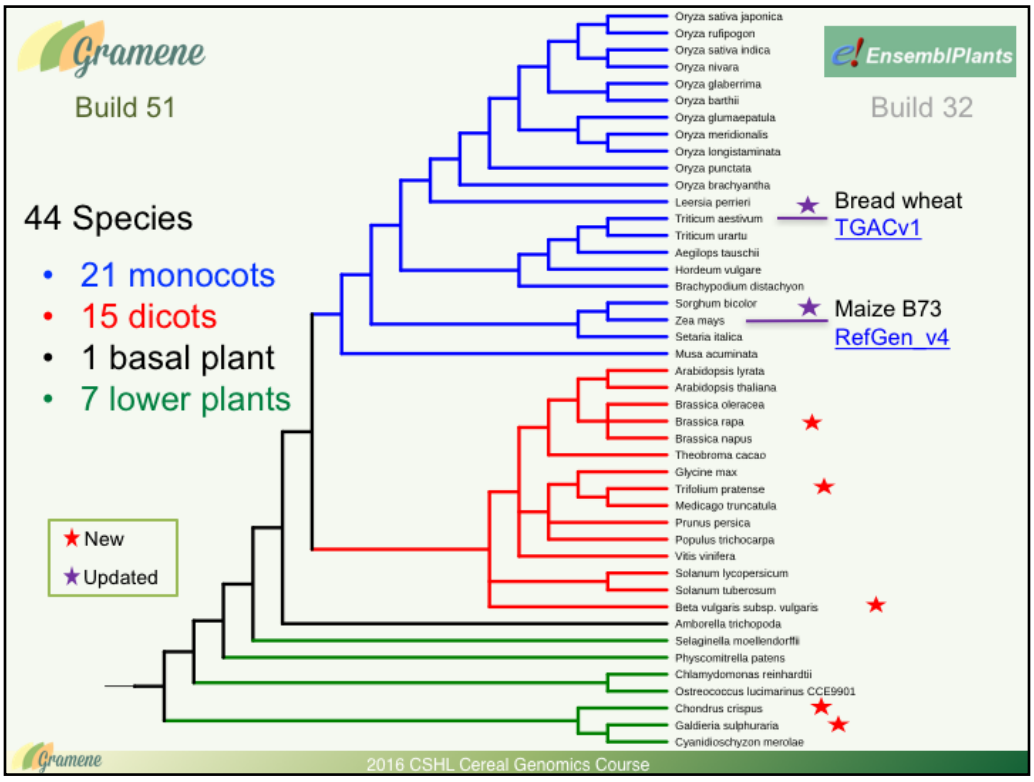
GENOMES

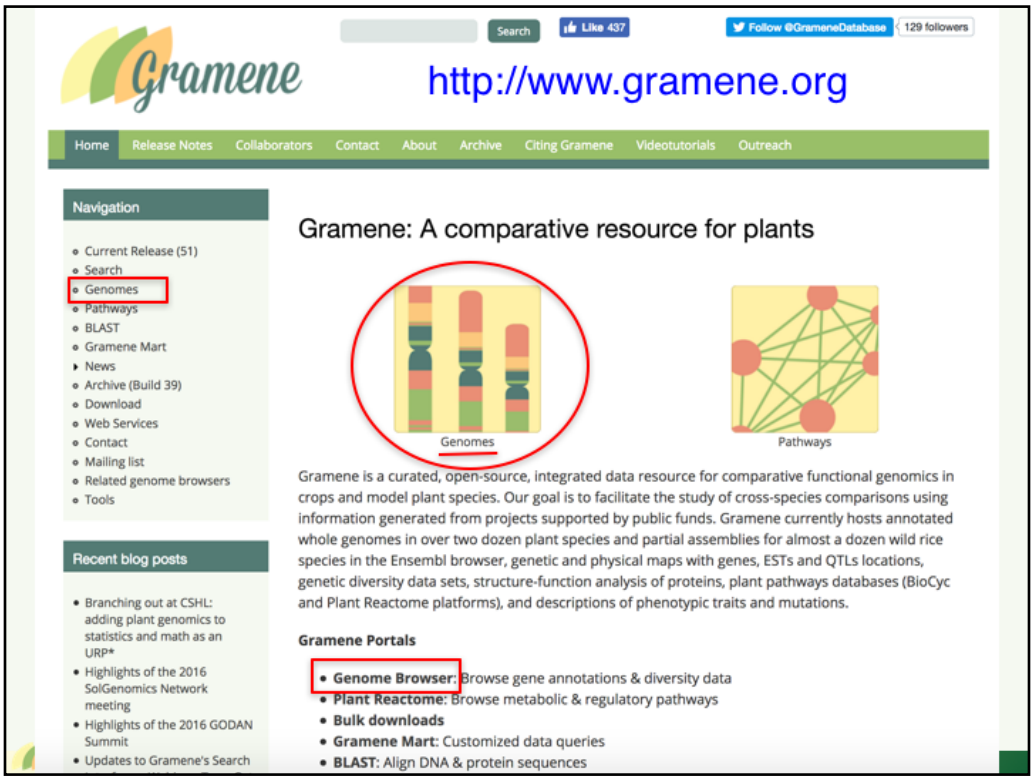


Progress in Plant Genomics

Gramene to the rescue

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

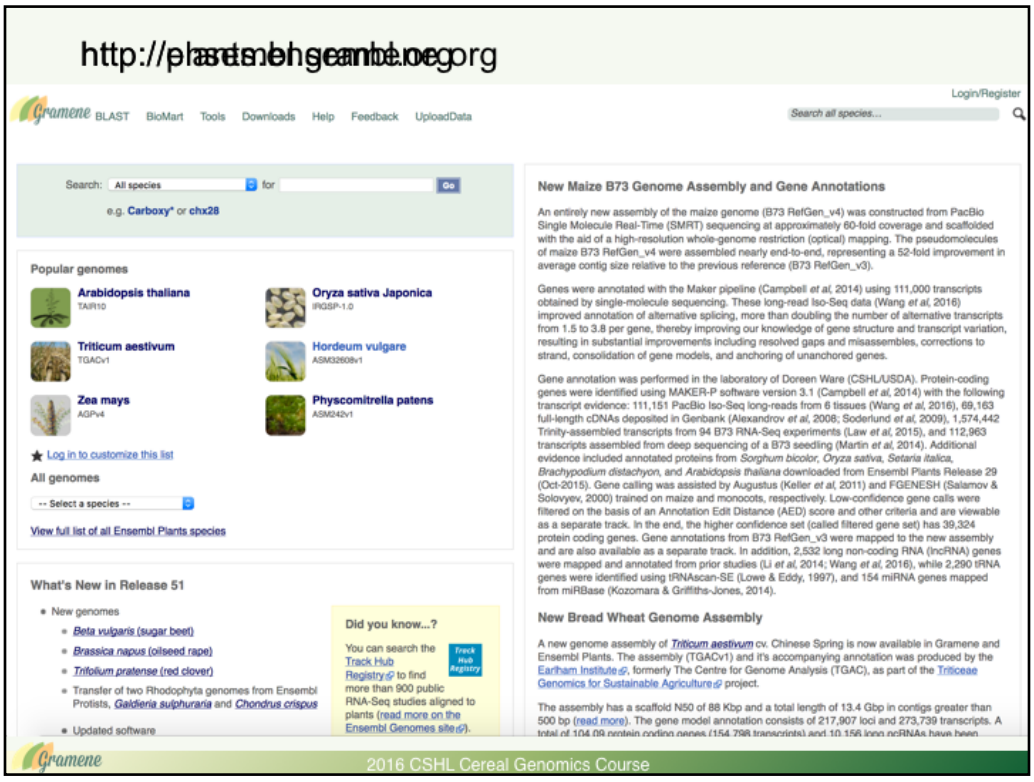

Gramene
Build 51
EnsemblPlants
Build 32
44 Species
21 monocots
15 dicots
1 basal plant
7 lower plants
New
Updated
Oryza sativa japonica
Oryza rufipogon
Oryza sativa indica
Oryza nivara
Oryza glaberrima
Oryza barthii
Oryza glumaepatula
Oryza meridionalis
Oryza longistaminata
Oryza punctata
Oryza brachyantha
Leersia perrieri
Triticum aestivum
Triticum urartu
Aegilops tauschii
Hordeum vulgare
Brachypodium distachyon
Sorghum bicolor
Zea mays
Setaria italica
Musa acuminata
Arabidopsis lyrata
Arabidopsis thaliana
Brassica oleracea
Brassica rapa
Brassica napus
Theobroma cacao
Glycine max
Trifolium pratense
Medicago truncatula
Prunus persica
Populus trichocarpa
Vitis vinifera
Solanum lycopersicum
Solanum tuberosum
Beta vulgaris subsp. vulgaris
Amborella trichopoda
Selaginella moellendorffii
Physcomitrella patens
Chlamydomonas reinhardtii
Ostreococcus lucimarinus CCE9901
Chondrus crispus
Galdieria sulphuraria
Cyanidioschyzon merolae
Bread wheat
TGACv1
Maize B73
RefGen_v4
Gramene
2016 CSHL Cereal Genomics Course



How to access Gramene's Genome Browser

Entry points

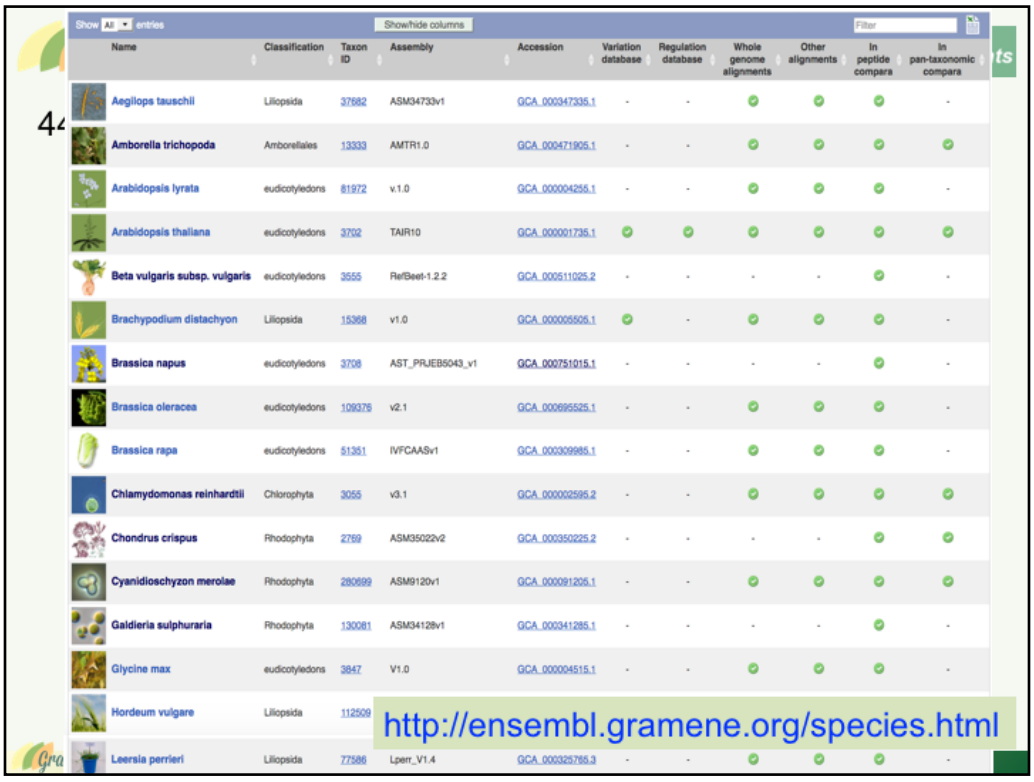
1) General Search =>
2) Genomes



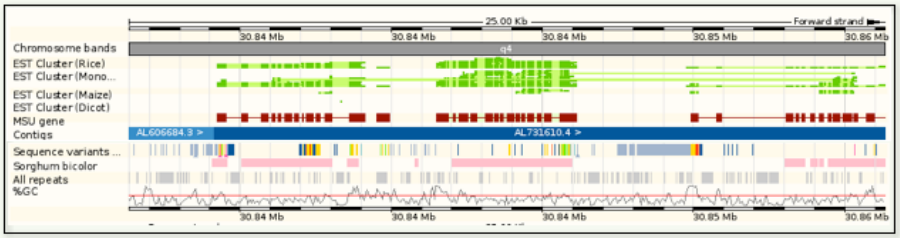
Gramene & Ensembl Plants shared interface

http://ensembl.gramene.org

http://plants.ensembl.org


Show All entries
Show/hide columns
Filter
Name
Classification
Taxon ID
Assembly
Accession
Variation database
Regulation database
Whole genome alignments
Other alignments
In peptide compara
In pan-taxonomic compara
Aegilops tauschii Liliopsida 37682 ASM34733v1 GCA_000347335.1
Amborella trichopoda Amborellales 13333 AMTR1.0 GCA_000471905.1
Arabidopsis lyrata eudicotyledons 81972 v.1.0 GCA_000004255.1
Arabidopsis thaliana eudicotyledons 3702 TAIR10 GCA_000001735.1
Beta vulgaris subsp. vulgaris eudicotyledons 3555 RefBeet-1.2.2 GCA_000511025.2
Brachypodium distachyon Liliopsida 15368 v1.0 GCA_000005505.1
Brassica napus eudicotyledons 3708 AST_PRJEB5043_v1 GCA_000751015.1
Brassica oleracea eudicotyledons 109376 v2.1 GCA_000695525.1
Brassica rapa eudicotyledons 51351 IVFCAASv1 GCA_000309985.1
Chlamydomonas reinhardtii Chlorophyta 3055 v3.1 GCA_000002595.2
Chondrus crispus Rhodophyta 2769 ASM35022v2 GCA_000350225.2
Cyanidioschyzon merolae Rhodophyta 280699 ASM9120v1 GCA_000091205.1
Galdieria sulphuraria Rhodophyta 130081 ASM34128v1 GCA_000341285.1
Glycine max eudicotyledons 3847 V1.0 GCA_000004515.1
Hordeum vulgare Liliopsida 112509
Leersia perrieri Liliopsida 77586 Lperr_V1.4 GCA_000325765.3
http://ensembl.gramene.org/species.html

Genome Browsers

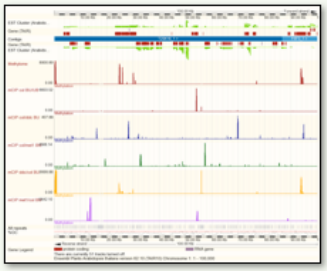


- Assembly structure and sequence
- Genes
- Expression
- Comparative alignments
- Genetic markers (SNPs, Indels, QTLs, SSRs)
- Variation
- Ontologies
- Repeats & transposable elements
- Regulatory & Epigenetic marks

Baseline Annotation

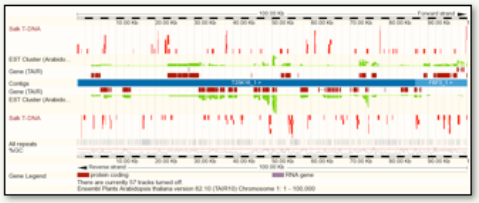
- Transposon discovery
- Gene prediction
- EST Alignment

Add Custom Tracks

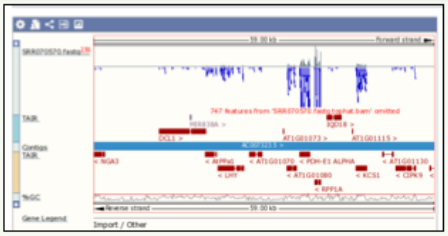


- Methylome (Ecker)
- Uploaded from an URL
- BED file format

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

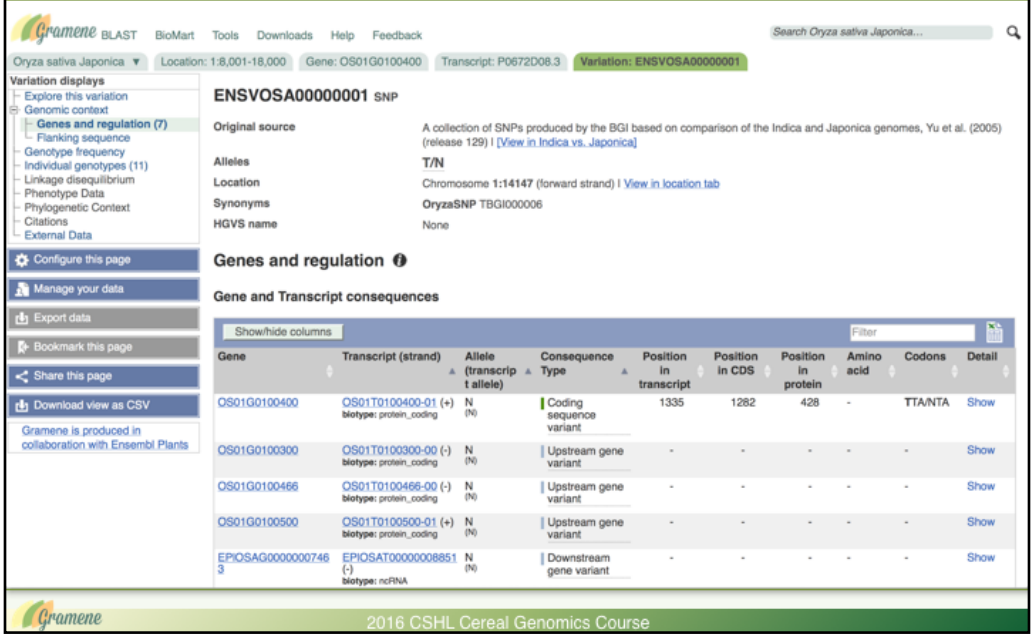


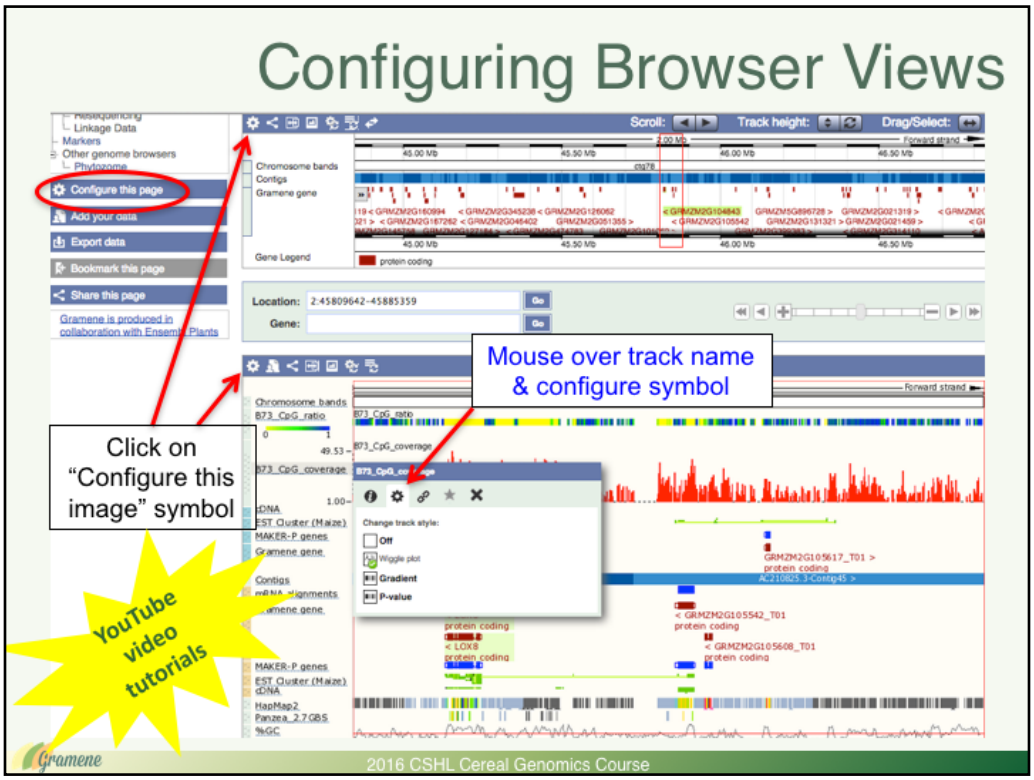
- Salk T-DNA lines
- Uploaded from my laptop
- GFF file format



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

Navigating Browsers





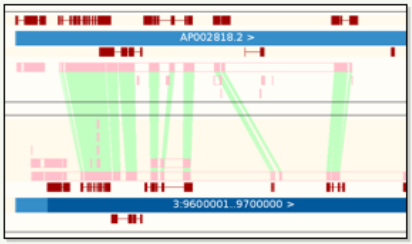
Collaboration with Ensembl Genomes (Paul Kersey)

Comparative Phylogenomics

Reconstructing evolutionary histories

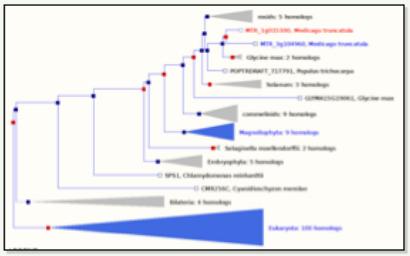
Whole Genome DNA Alignment

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



Phylogenetic Gene Trees

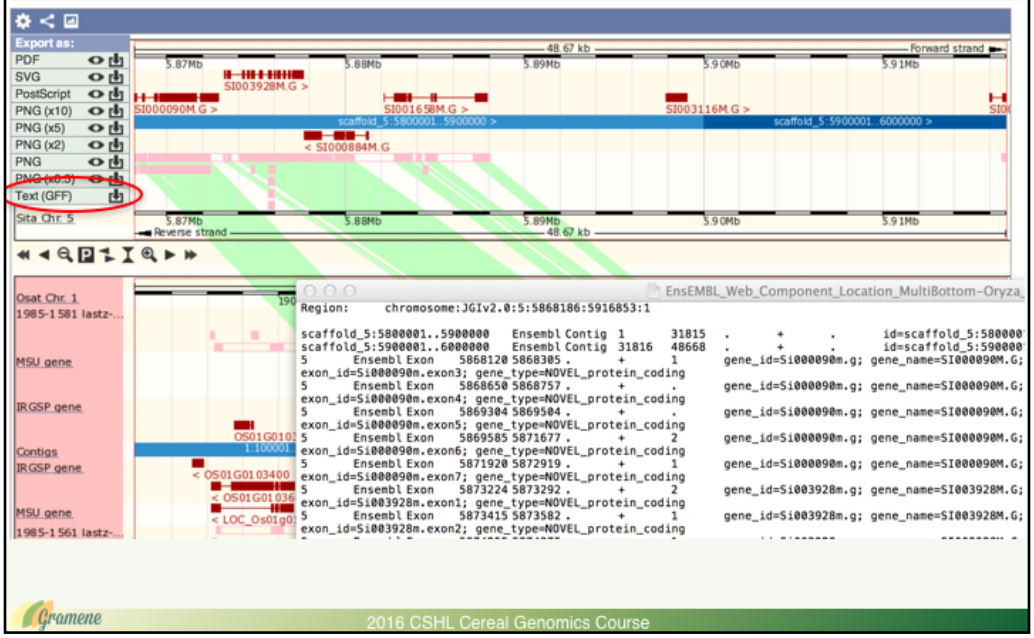
- Infers orthologs and paralogs
- Taxonomic dating
- dN, dS, dN/dS
- Oryza-centered trees



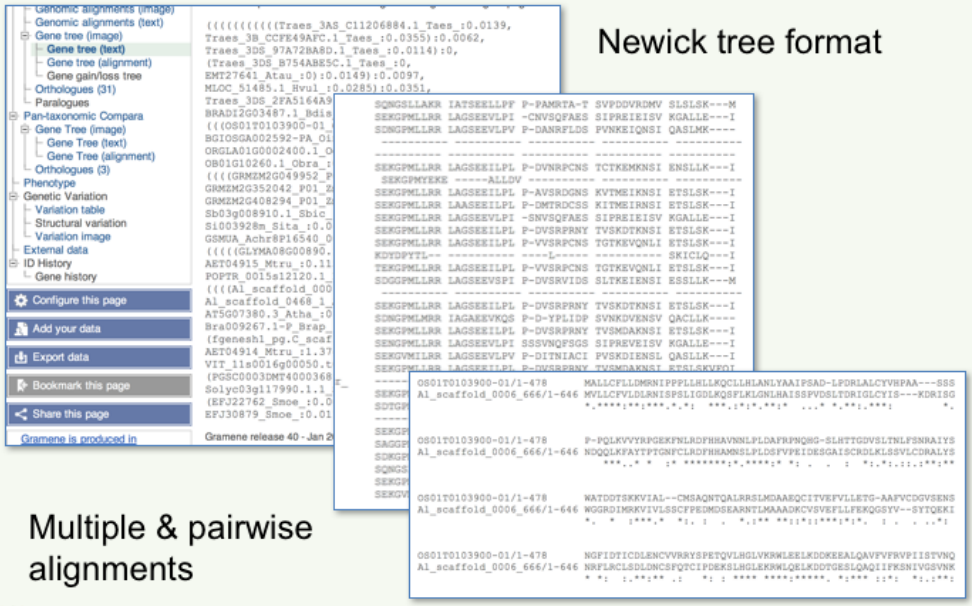
Vilella A.J., et al. (2008). *Genome Res.* doi:10.1101/gr.073585.107
Schwartz S et al., Genome Res.;13(1):103-7, Kent WJ et al., Proc Natl Acad Sci USA., 2003;100(20):11484-9

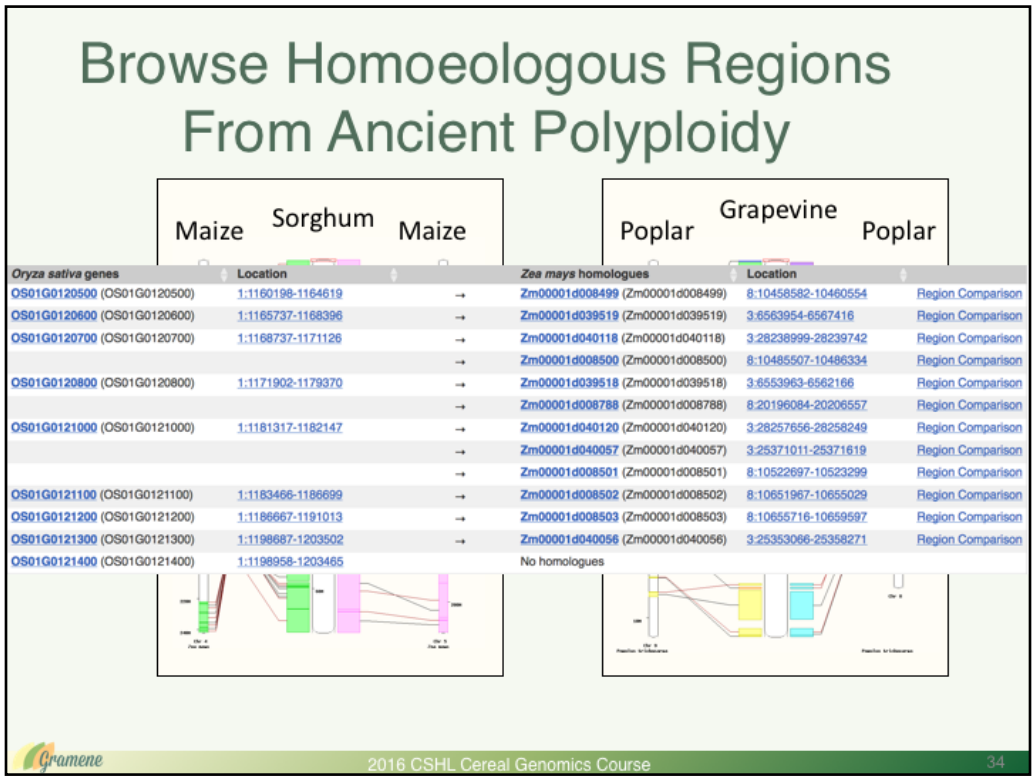
http://useast.ensembl.org/info/docs/compara/homology_method.html

Exporting Images & Data



Exporting Images & Data



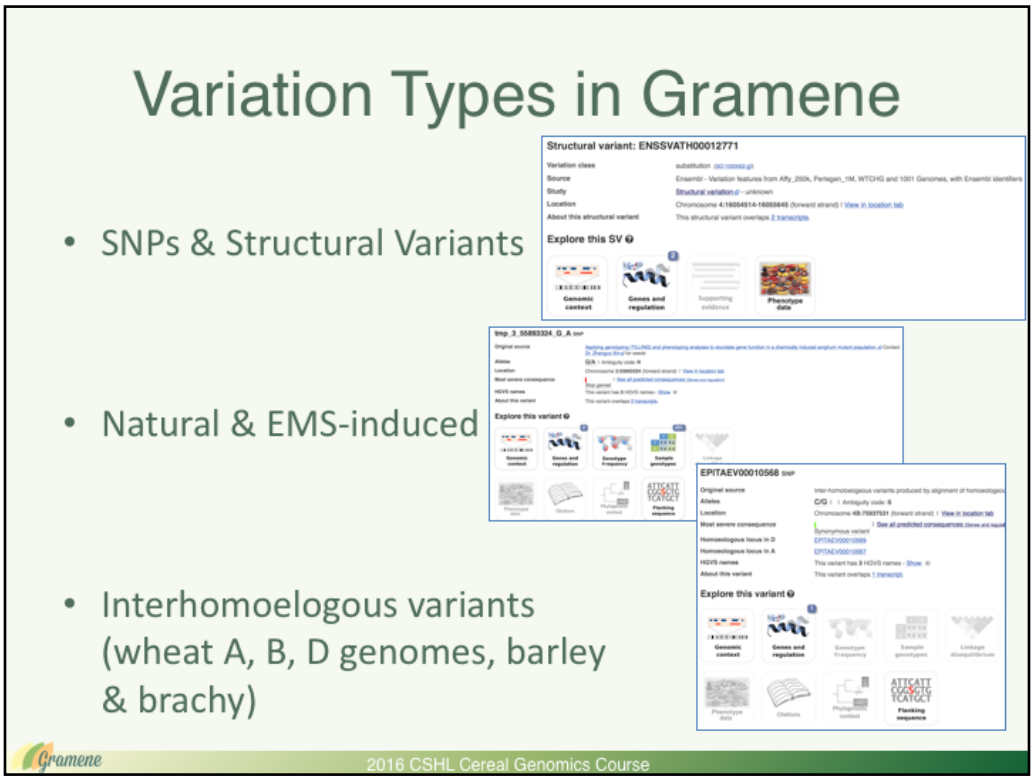


Synteny is based on Compara gene trees

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

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

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



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

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

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

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

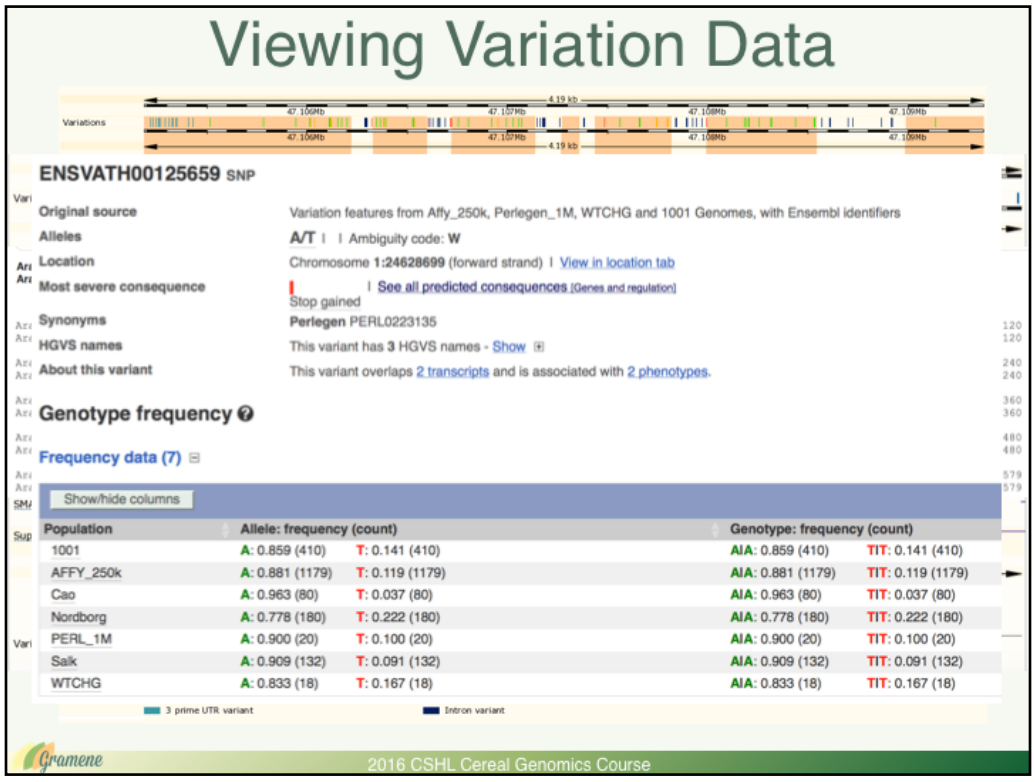
http://ensembl.gramene.org/Sorghum_bicolor/Variation/Explore?db=core;g=Sb03g028120;r=3:55892797-55900180;source=Sorghum_EMS_mutants;t=Sb03g028120.1;v=tmp_3_55893324_G_A;vdb=variation;vf=7315833

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

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

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

Seq alignments for barley & brachy



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

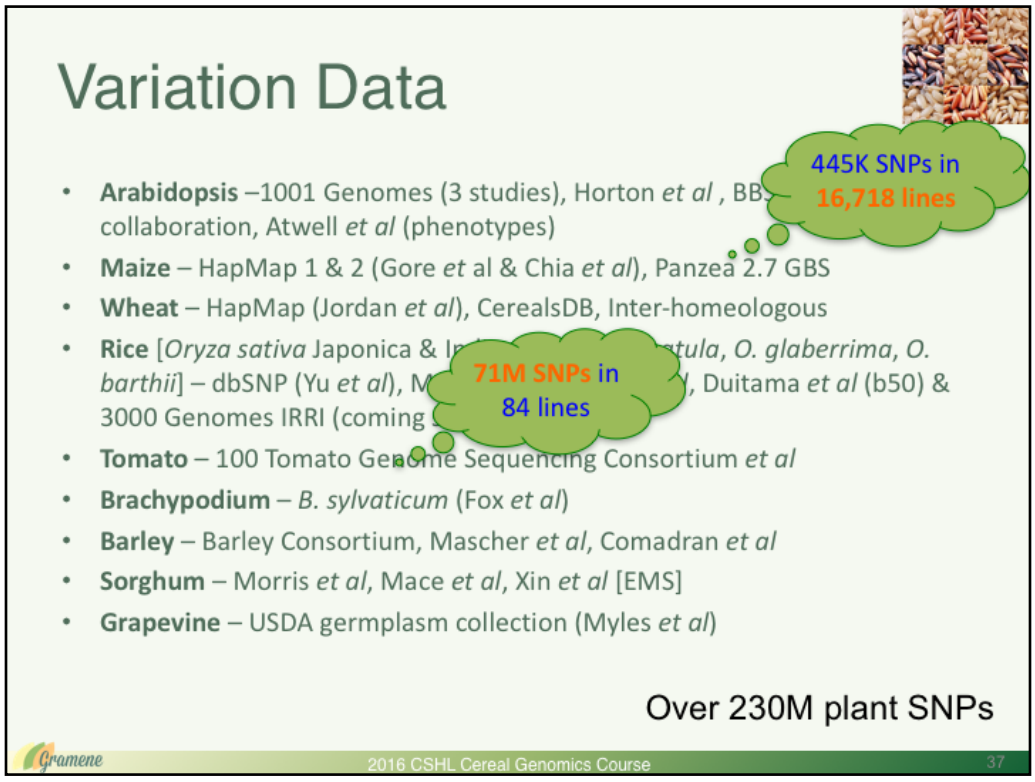
To get the color coded sequence alignment (if available for this region against desired species – not available for example region between maize and sorghum):

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

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

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

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



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

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

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

27,869,011 OS japonica SNPs in EG31

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

USDA grape collection: 17 strains

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

Computed using 1) Compara pipeline for detecting protein orthology,

and 2) supporting WGAs to find sequence diffs at the nucleotide level

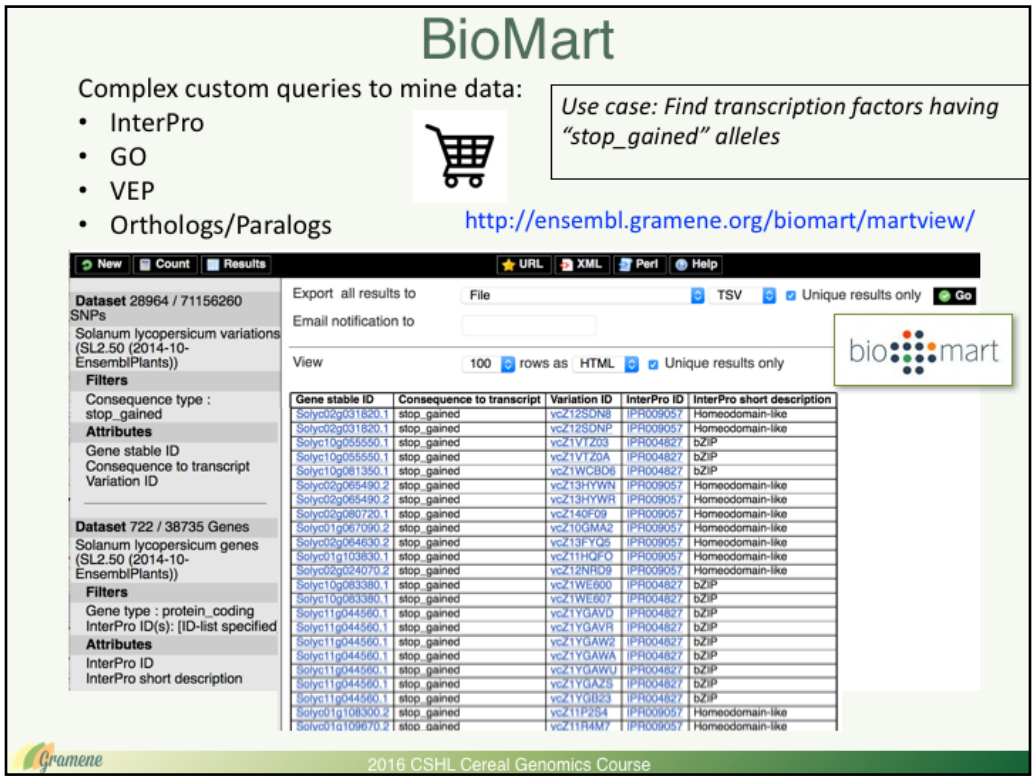
http://www.vegkitchen.com/tips/cooking-with-whole-grains/attachment/ge-rice-threatens-biodiversity/

Ensembl Diversity Databases

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

Example VEP SO classifications

intron_variant
missense_variant
synonymous_variant
stop_gained
stop_lost
splice_donor_variant
splice_acceptor_variant
initiator_codon_variant
stop_retained_variant



IPR001005

IPR006447

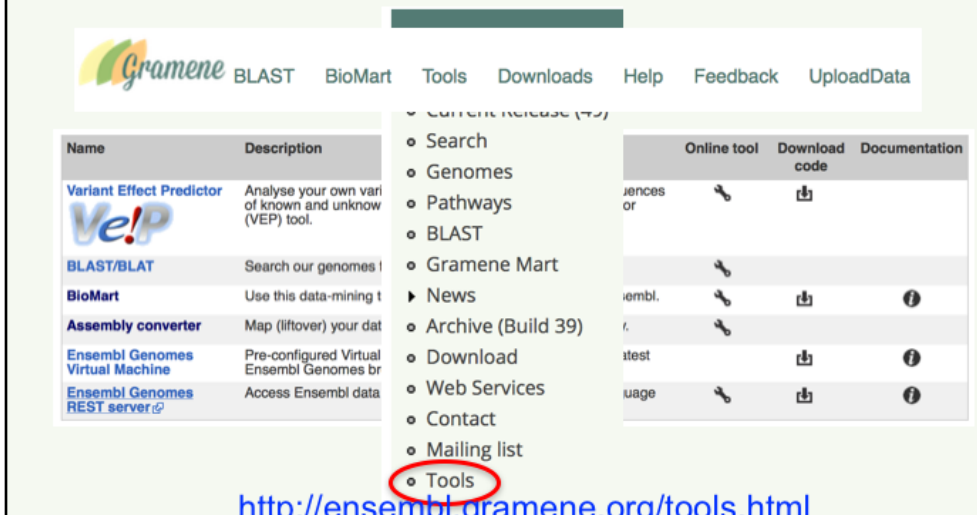
IPR009057

IPR004827

IPR011598

IPR009057

Tools



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

FTP site

Gramene FTP site

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

Ensembl Plants FTP server

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

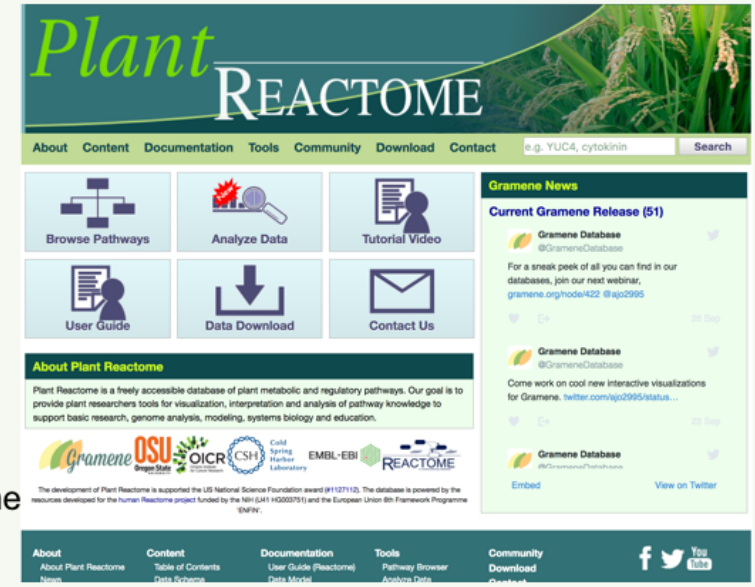
- DNA, cDNA and Protein (FASTA)
- Genomic DNA sequence (EMBL and GenBank flatfile dumps)
- Ensembl core databases and specific features (MySQL and simple text format)
- Species-centered data including comparative genomics annotations (TSV)
- Gene annotations (GTF and GFF3)
- Variation data (GVF and VCF)

http://gramene.org/ftp-download

PATHWAYS

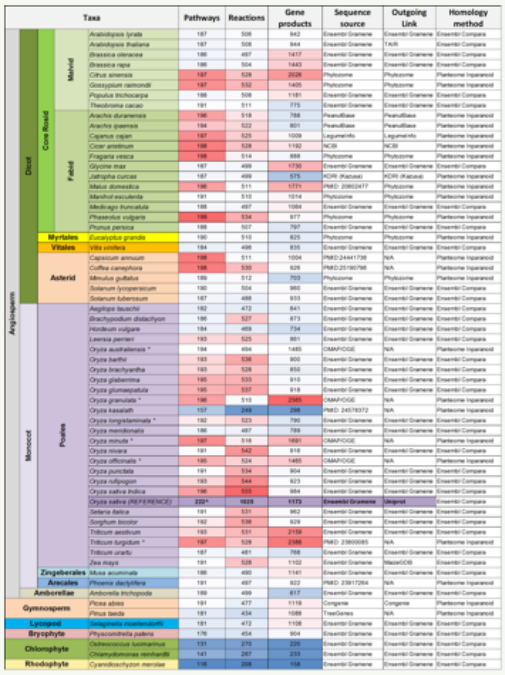
Pathways Platform

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



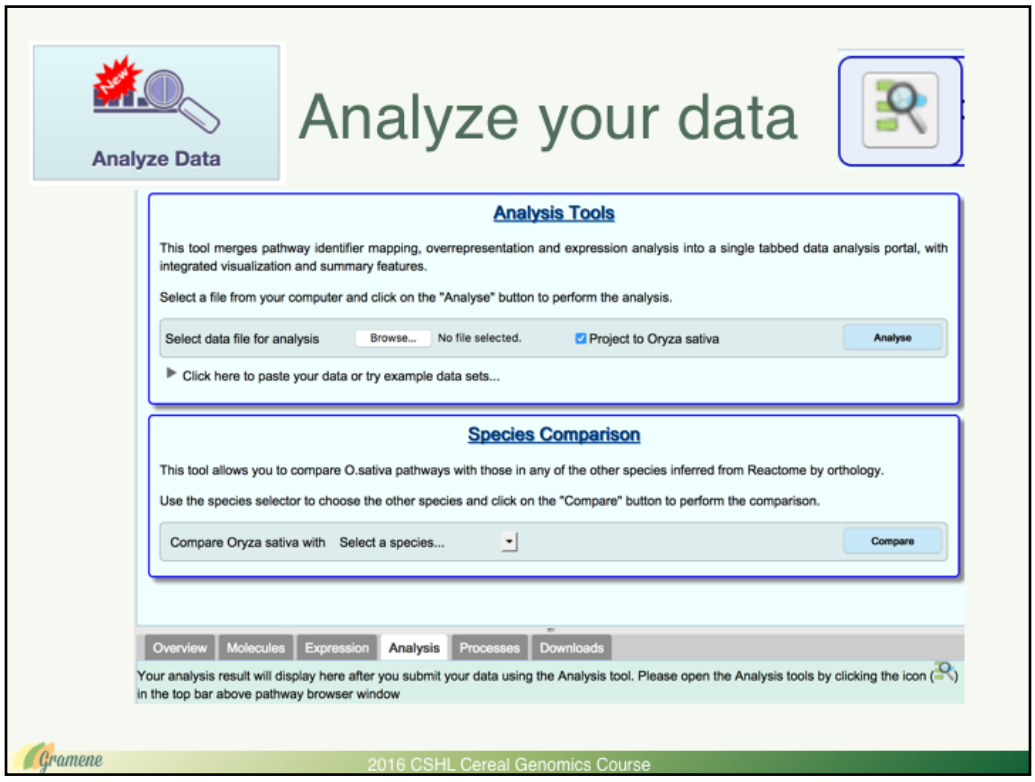
Orthology-based pathway projections

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



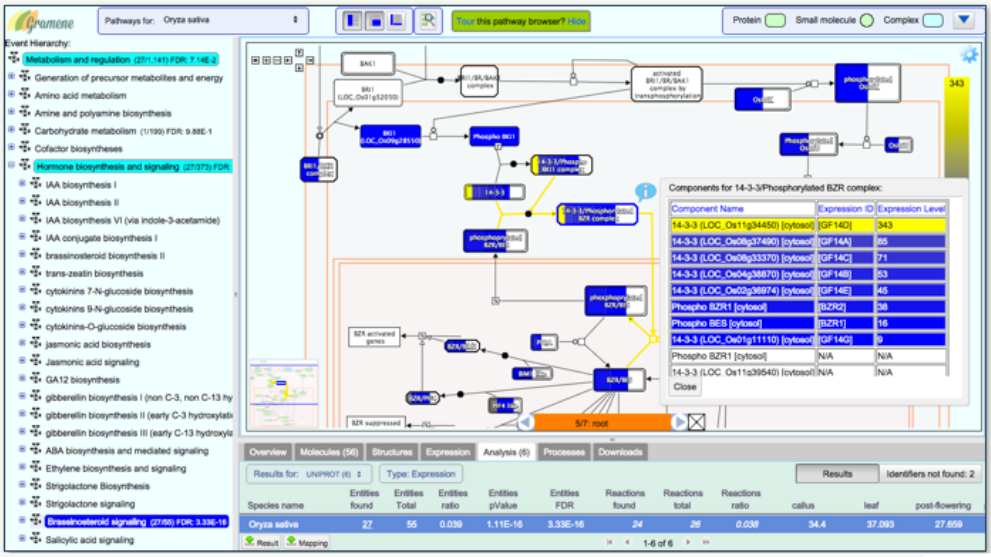
	Taxa	Pathways	Reactions	Gene products	Sequence source	Outgoing Link	Homology method
Angiosperm	Arabidopsis lyrata	187	506	942	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Arabidopsis thaliana	187	508	944	Ensembl Gramene	TAIR	Ensembl Compara
	Brassica oleracea	186	497	1417	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Brassica rapa	186	504	1443	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Citrus sinensis	197	528	2029	Phytozome	Phytozome	Plantsome Inparanoid
	Gossypium raimondii	187	532	1405	Phytozome	Phytozome	Plantsome Inparanoid
	Populus trichocarpa	188	508	1181	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Theobroma cacao	191	511	775	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Arachis duranensis	196	518	788	PeanutBase	PeanutBase	Plantsome Inparanoid
	Arachis ipaensis	194	522	801	PeanutBase	PeanutBase	Plantsome Inparanoid
	Cajanus cajan	197	525	1009	LegumeInfo	LegumeInfo	Plantsome Inparanoid
	Cicer arietinum	198	528	1192	NCBI	NCBI	Plantsome Inparanoid
	Fragaria vesca	186	514	888	Phytozome	Phytozome	Plantsome Inparanoid
	Glycine max	187	499	1736	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Jatropha curcas	187	499	575	KDRI (Kazusa)	KDRI (Kazusa)	Plantsome Inparanoid
	Malus domestica	196	511	1771	PMID: 20801477	Phytozome	Plantsome Inparanoid
	Manihot esculenta	191	510	1014	Phytozome	Phytozome	Plantsome Inparanoid
	Medicago truncatula	188	497	1084	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Phaseolus vulgaris	198	534	977	Phytozome	Phytozome	Plantsome Inparanoid
	Prunus persica	188	507	797	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Eucalyptus grandis	190	510	825	Phytozome	Phytozome	Plantsome Inparanoid
	Vitis vinifera	184	496	835	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Capsicum annuum	198	511	1004	PMID:24441736	N/A	Plantsome Inparanoid
	Coffea canephora	196	530	928	PMID:25190796	N/A	Plantsome Inparanoid
	Mimulus guttatus	189	512	703	Phytozome	Phytozome	Plantsome Inparanoid
	Solanum lycopersicum	190	504	960	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Solanum tuberosum	187	486	933	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Aegilops tauschii	182	472	841	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Brachypodium distachyon	186	527	873	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Hordeum vulgare	184	469	734	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Leersia perrieri	193	525	861	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza australiensis *	194	494	1485	OMAPYDSE	N/A	Plantsome Inparanoid
	Oryza barthii	193	536	900	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza brachyantha	193	528	850	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza glaberrima	195	533	910	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza glumaepatula	195	537	918	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza granulata *	196	510	2065	OMAPYDSE	N/A	Plantsome Inparanoid
	Oryza kasalath	157	242	298	PMID: 24578372	N/A	Plantsome Inparanoid
	Oryza longistaminata *	192	523	790	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza meridionalis	186	487	789	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza minuta *	197	516	1691	OMAPYDSE	N/A	Plantsome Inparanoid
	Oryza nivara	191	542	816	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza officinalis *	195	524	1485	OMAPYDSE	N/A	Plantsome Inparanoid
	Oryza punctata	191	534	904	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza rufipogon	193	544	923	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza sativa indica	196	535	984	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Oryza sativa (REFERENCE)	222^	1028	1173	Ensembl Gramene	Uniprot	Ensembl Compara
	Setaria italica	191	531	962	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Sorghum bicolor	192	536	929	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Triticum aestivum	193	531	2159	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Triticum turgidum *	197	528	2388	PMID: 23800085	N/A	Plantsome Inparanoid
	Triticum urartu	187	481	768	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Zea mays	191	528	1102	Ensembl Gramene	MaizeGDB	Ensembl Compara
	Musa acuminata	186	490	1141	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Phoenix dactylifera	191	497	922	PMID: 23917264	N/A	Plantsome Inparanoid
	Amborella trichopoda	189	499	817	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
Gymnosperm	Picea abies	191	477	1119	Congenie	Congenie	Plantsome Inparanoid
	Pinus taeda	191	434	1088	TreeGenes	N/A	Plantsome Inparanoid
Lycopod	Selaginella moellendorffii	181	472	1138	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
Bryophyte	Physcomitrella patens	176	454	904	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
Chlorophyte	Ostreococcus lucimarinus	131	270	229	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
	Chlamydomonas reinhardtii	141	287	233	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
Rhodophyte	Cyanidioschyzon merolae	116	208	158	Ensembl Gramene	Ensembl Gramene	Ensembl Compara

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

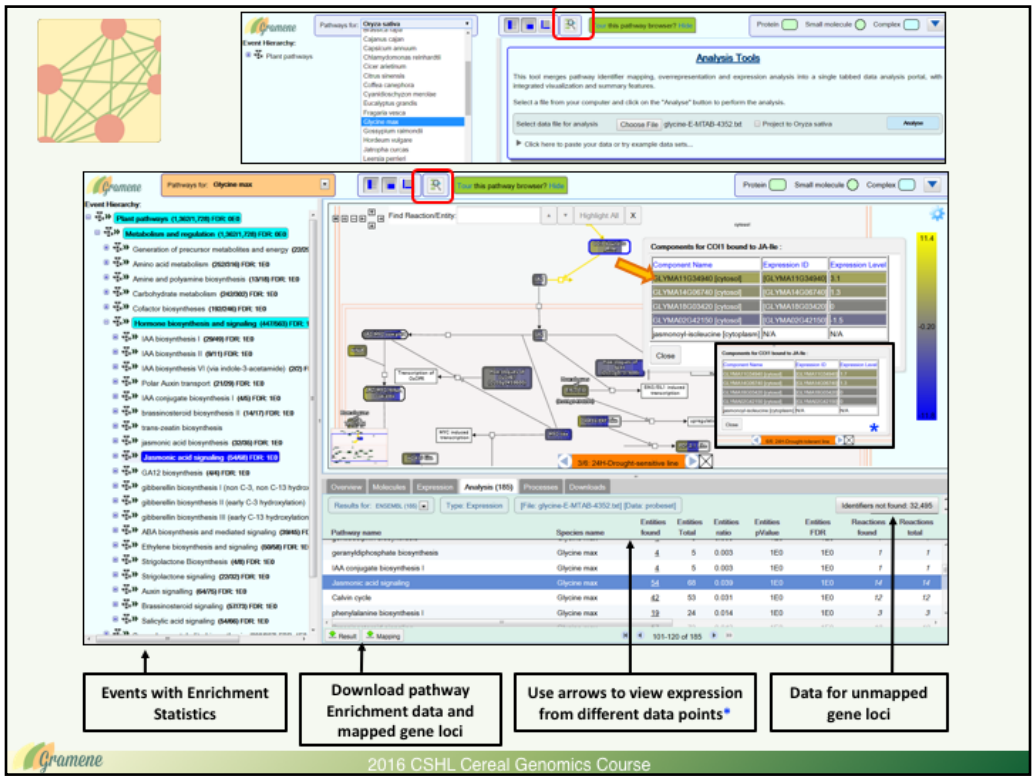

Analyze Data
Analyze your data
Analysis Tools
This tool merges pathway identifier mapping, overrepresentation and expression analysis into a single tabbed data analysis portal, with integrated visualization and summary features.
Select a file from your computer and click on the "Analyse" button to perform the analysis.
Select data file for analysis
Browse...
No file selected.
Project to Oryza sativa
Analyse
Click here to paste your data or try example data sets...
Species Comparison
This tool allows you to compare O.sativa pathways with those in any of the other species inferred from Reactome by orthology.
Use the species selector to choose the other species and click on the "Compare" button to perform the comparison.
Compare Oryza sativa with
Select a species...
Compare
Overview
Molecules
Expression
Analysis
Processes
Downloads
Your analysis result will display here after you submit your data using the Analysis tool. Please open the Analysis tools by clicking the icon ( ) in the top bar above pathway browser window
Gramene
2016 CSHL Cereal Genomics Course

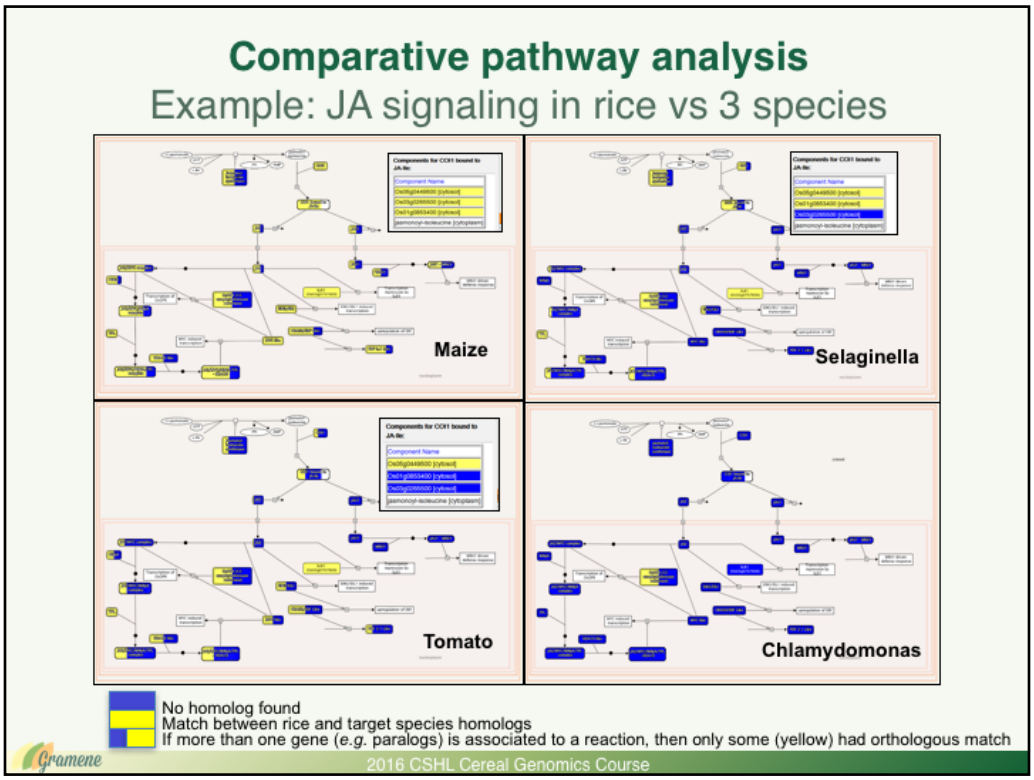
plantreactome.gramene.org

Tools for enrichment analysis and species comparison



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

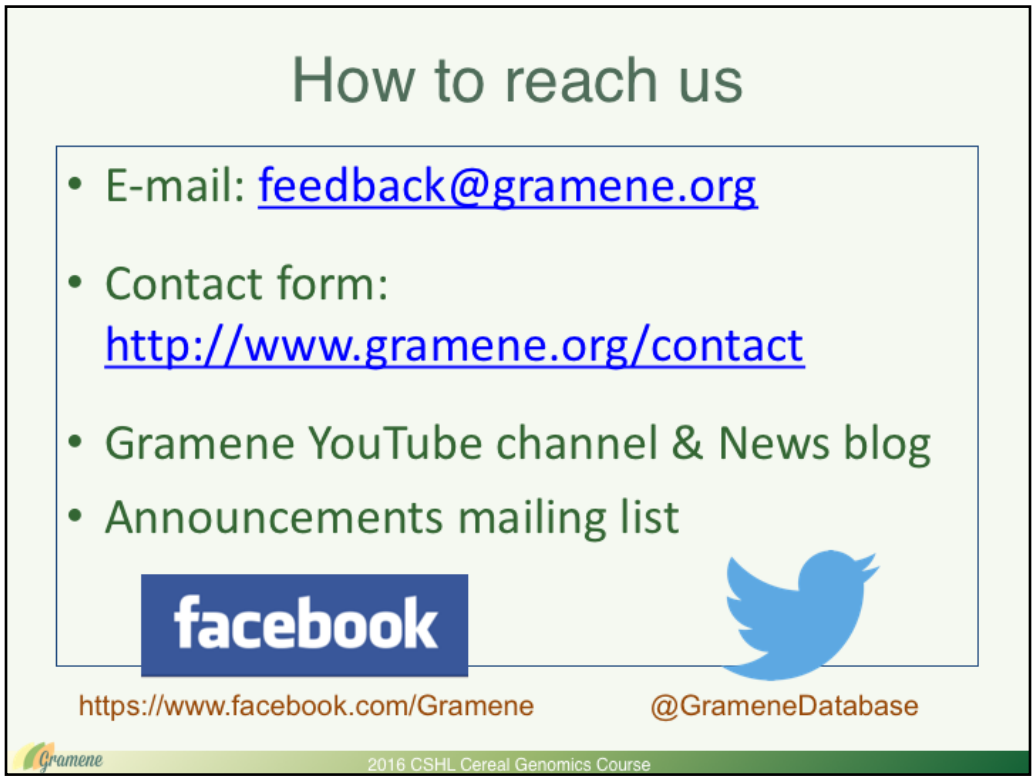

Events with Enrichment Statistics
Download pathway Enrichment data and mapped gene loci
Use arrows to view expression from different data points*
Data for unmapped gene loci
2016 CSHL Cereal Genomics Course



Species comparison

EDUCATIONAL RESOURCES

http://www.gramene.org/outreach



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

Team of very talented scientists and developers. Thanks!

Gramene - Exploring Function through Comparative Genomics and Network Analysis

NSF IOS 1127112 (2011- 2017)

Doreen Ware, PI (USDA-ARS, CSHL)
Michael Campbell, Kapeel Chougule, Yinping Jiao, Sunita Kumari, Joe Mulvaney, Andrew Olson, Joshua Stein, Marcela K. Tello-Ruiz, Jim Thomason, Peter van Buren, Bo Wang, Sharon Wei

Pankaj Jaiswal, Co-PI (OSU)
Noor Al-Bader, Justin Elser, Matthew Geniza, Parul Gupta, Sushma Naithani, Justin Preece
Paul Kersey / Robert 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)

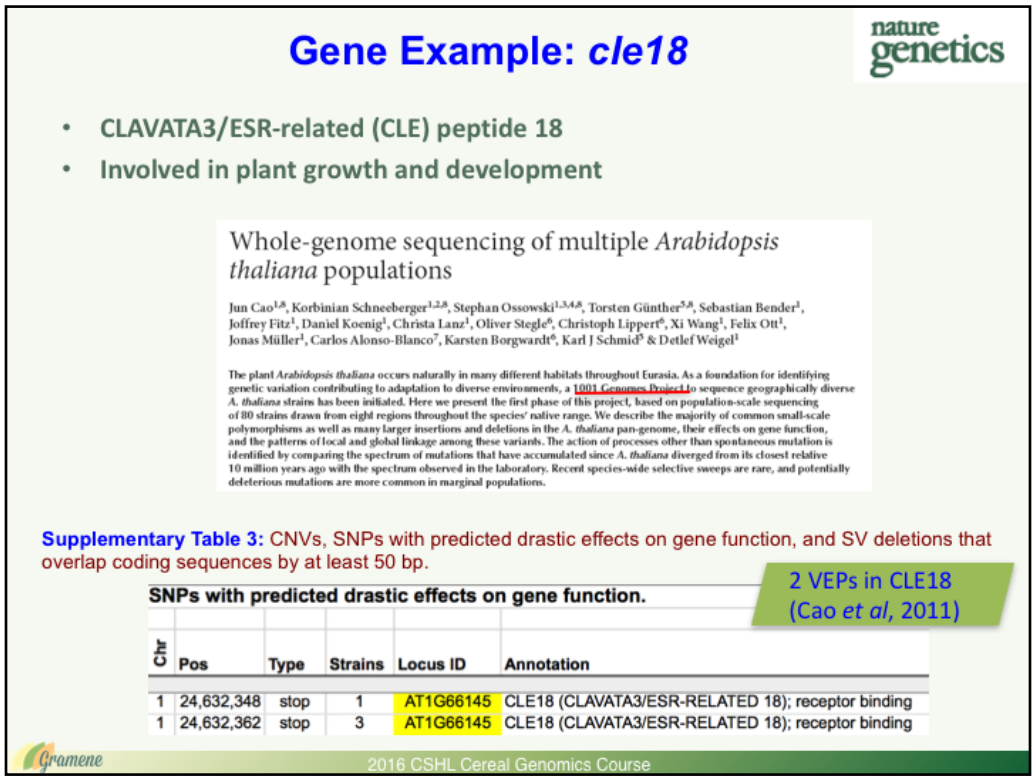
Crispin Taylor (ASPB)
Patty Lockhart; Weijia Xu (TACC), Amit Gupta(TACC)


Gramene

2016 CSHL Cereal Genomics Course

Use Case

EXPLORING CLE18



CLE 18 (CLAVATA3/ESR-related)

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

A different gene:

CLV3

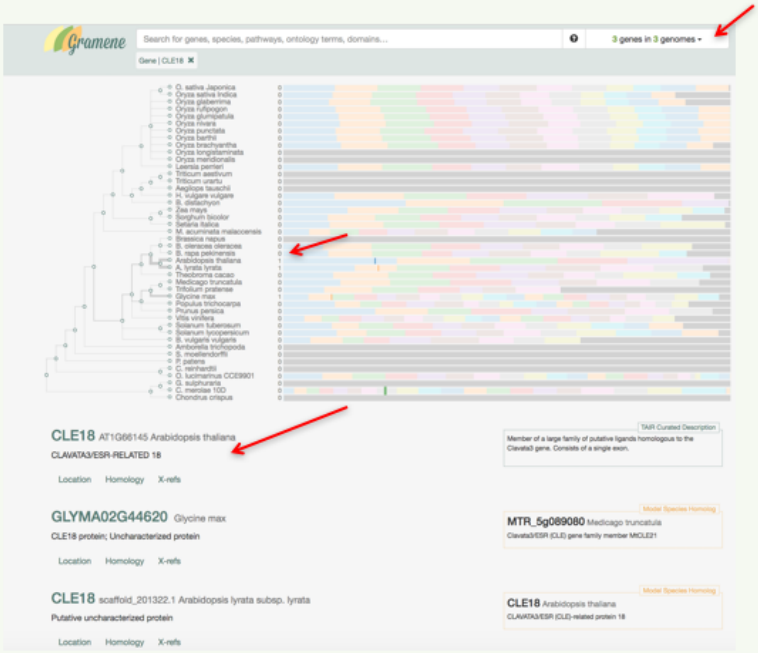
AT2G27250

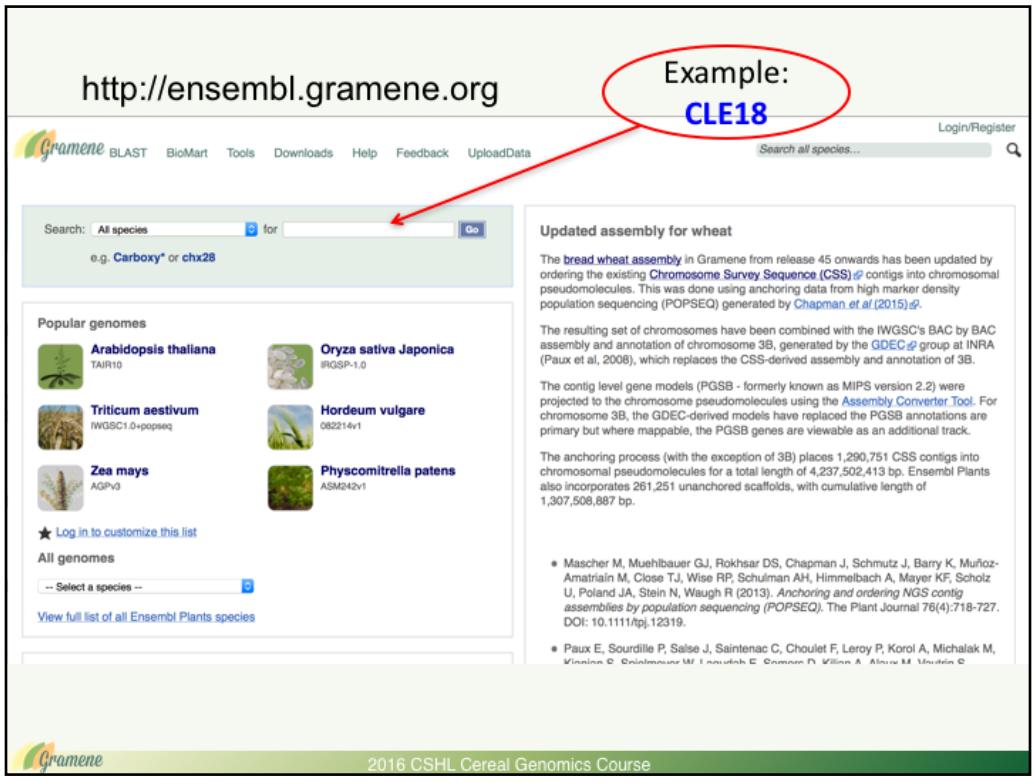
Contributes to meristem structure and identity in both, shoot and flower

stop-gained

ENSVATH00125659

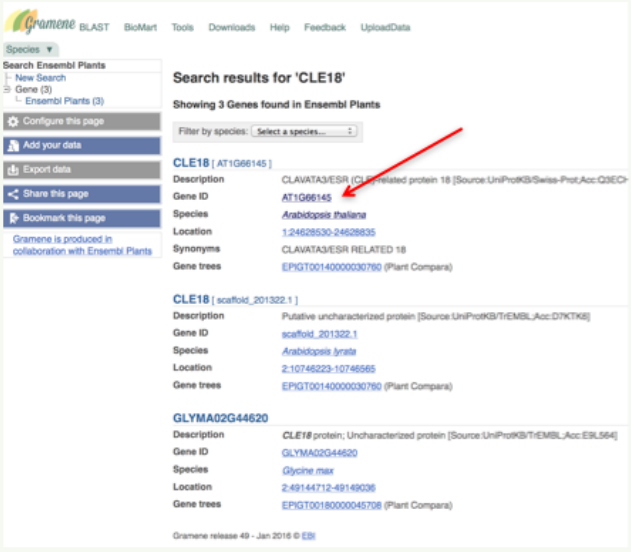
Search results



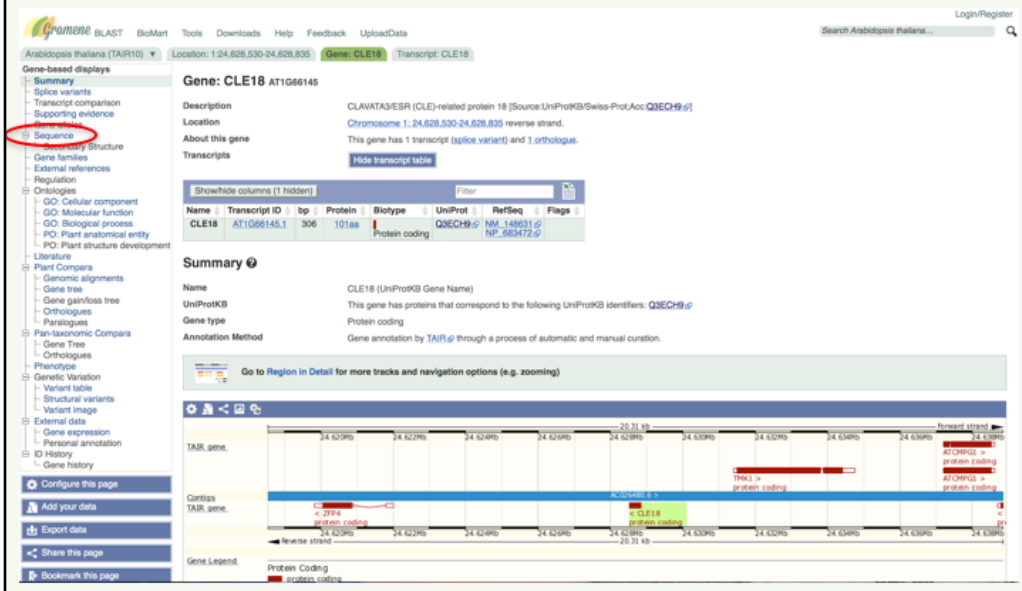


Plant CLAVATA3/ESR-related (CLE) peptides have diverse roles in plant growth and development

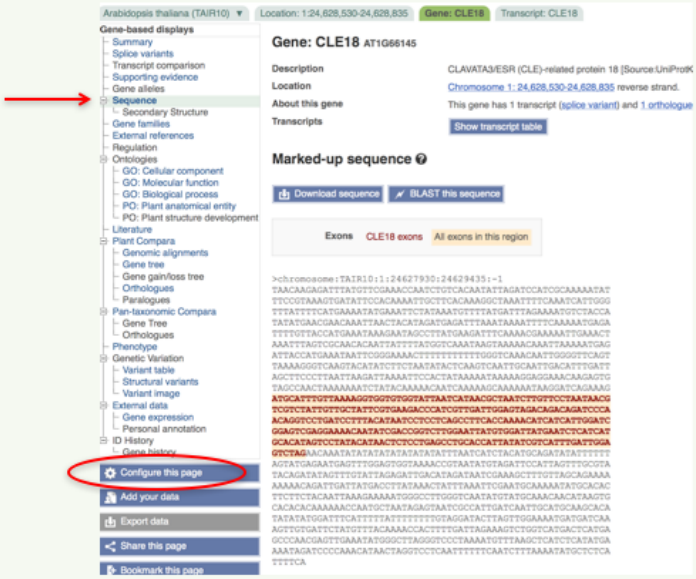
Search results



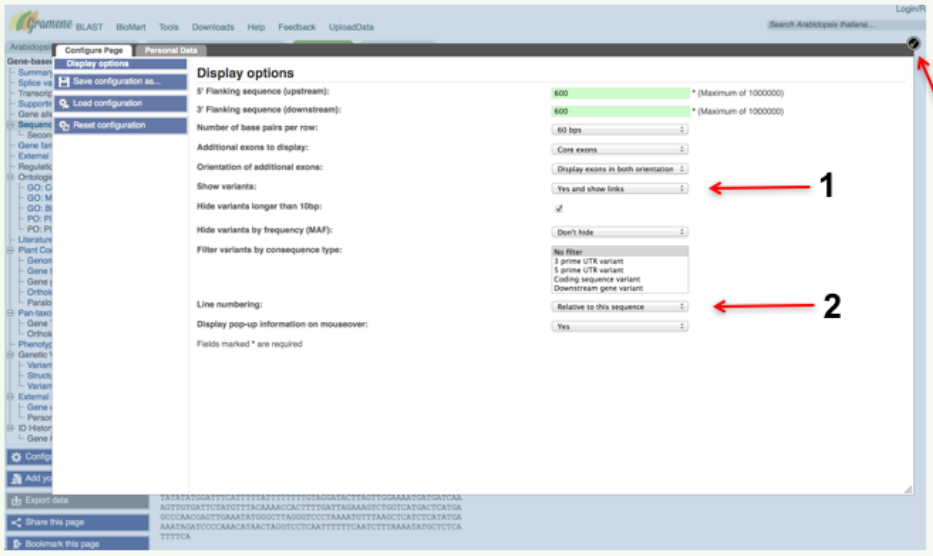
CLE18 gene page



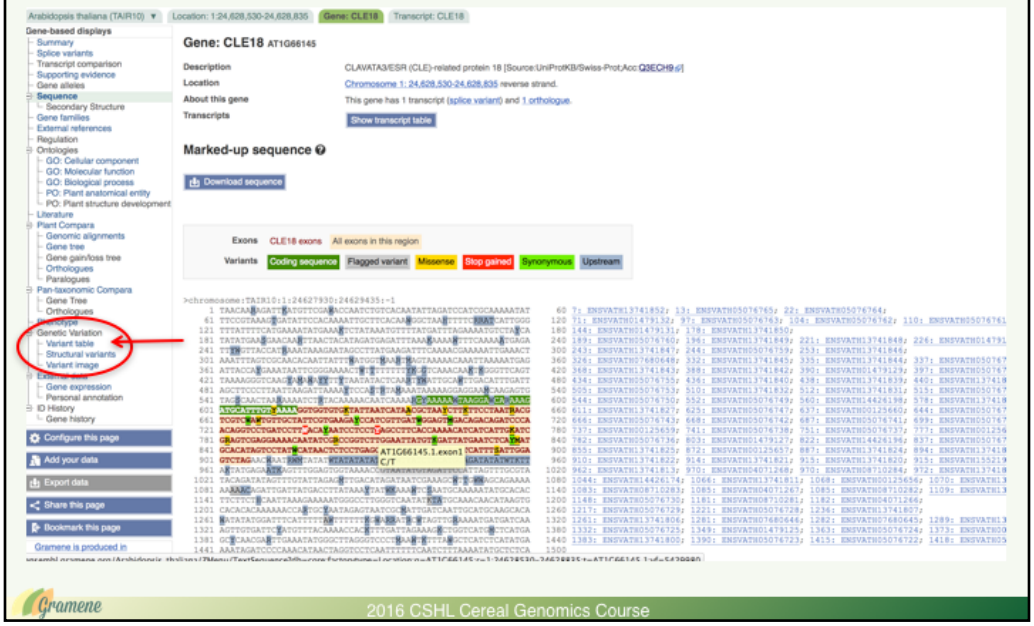
Gene sequence



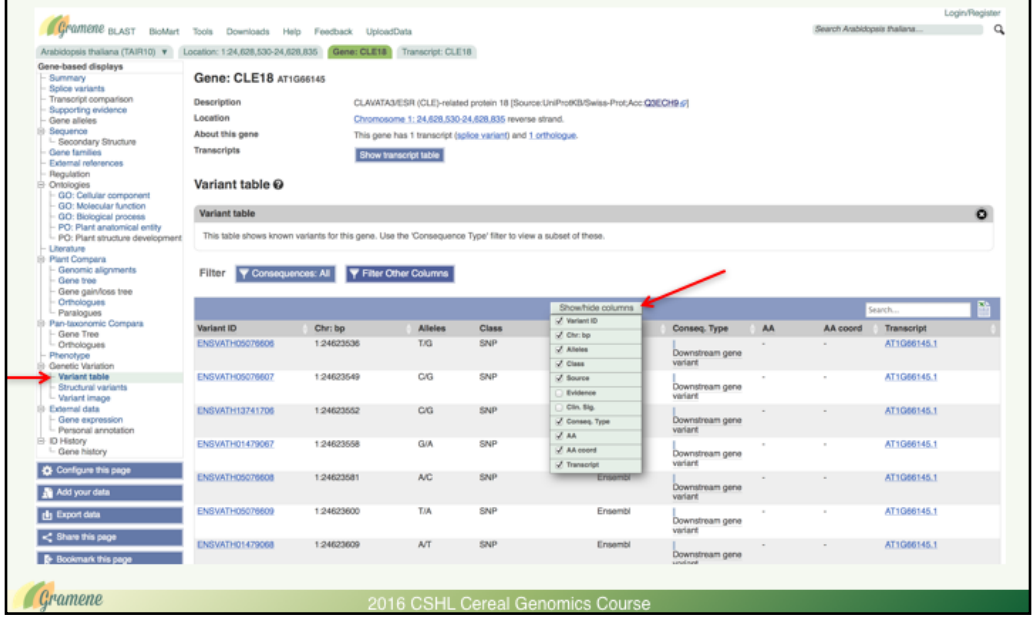
Configure (Sequence) page: Show variants



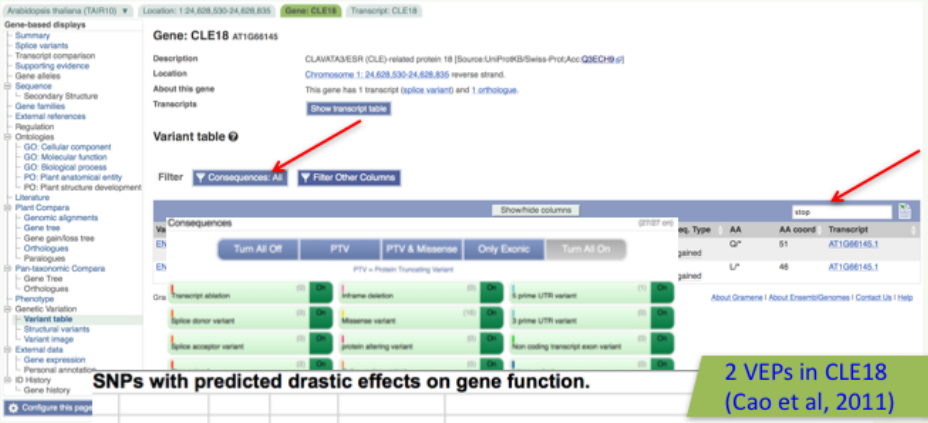
Mouse hover a variant to reveal details



Variant table (Gene & Transcript)



Filter consequences

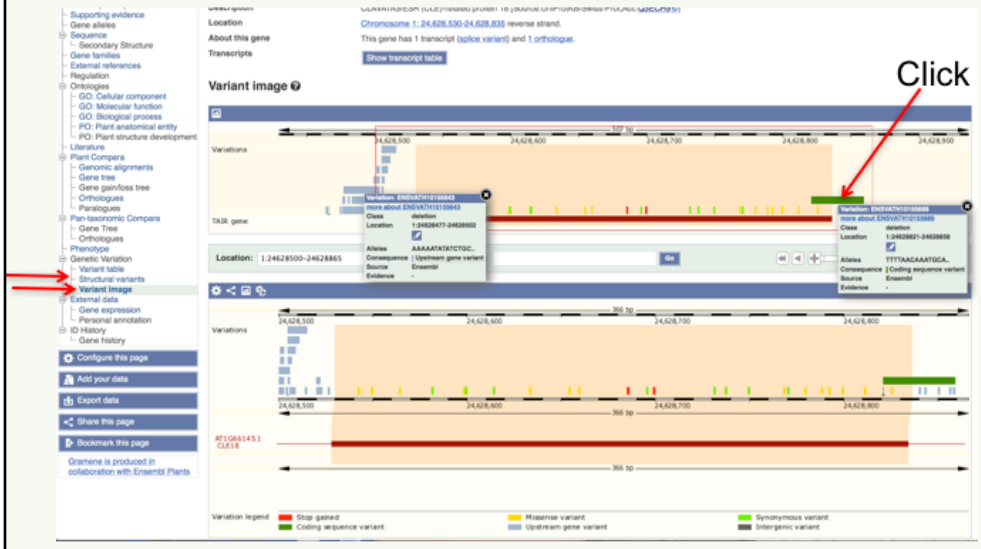


2 VEPs in CLE18 (Cao et al, 2011)

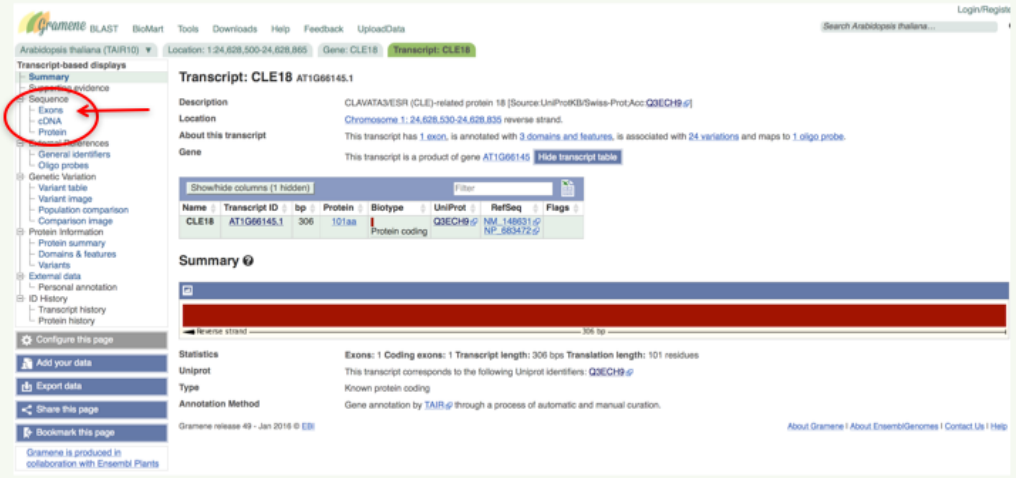
SNPs with predicted drastic effects on gene function.

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

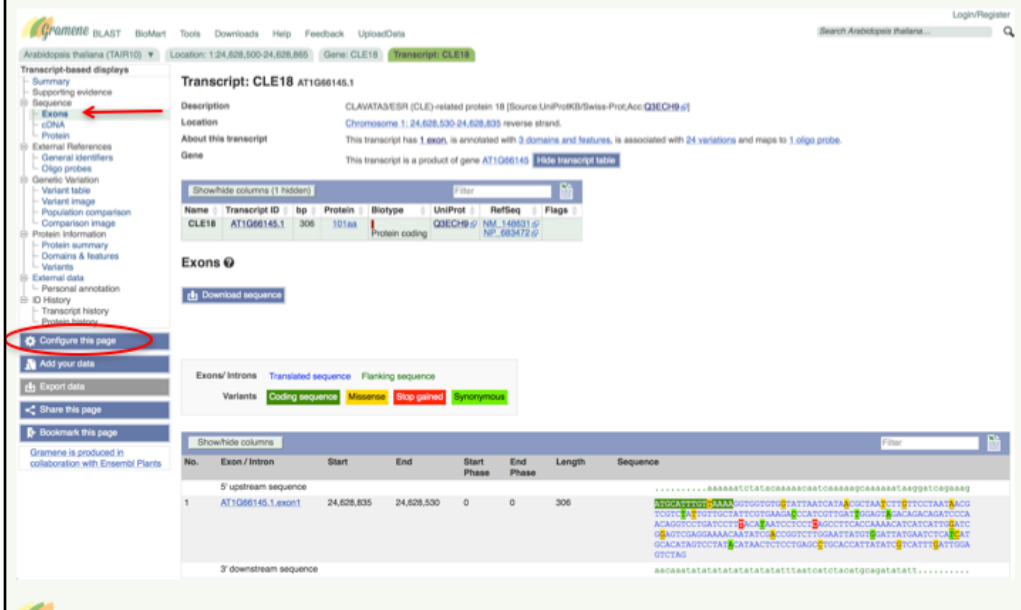
Structural variants



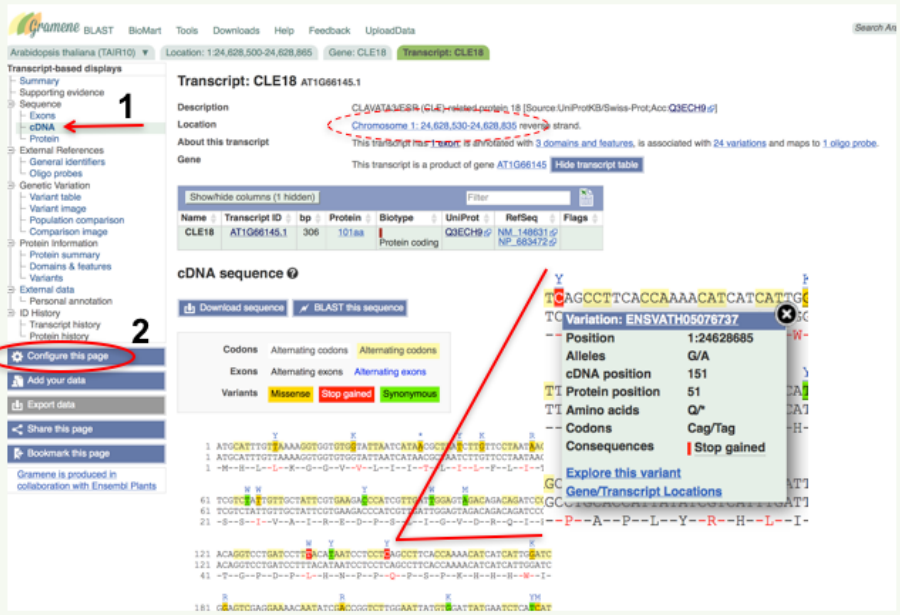
Go to Transcript page



Transcript Sequence: Exons

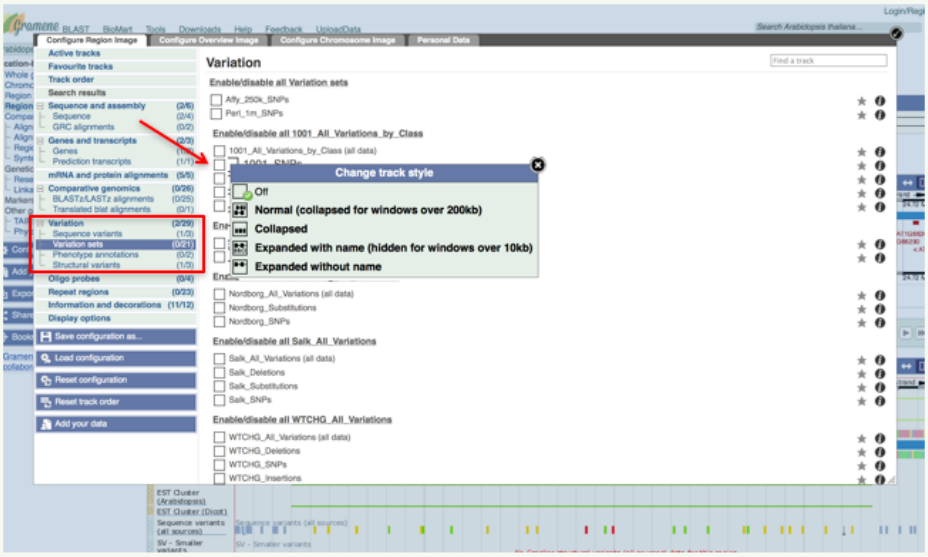


Transcript Sequence: cDNA

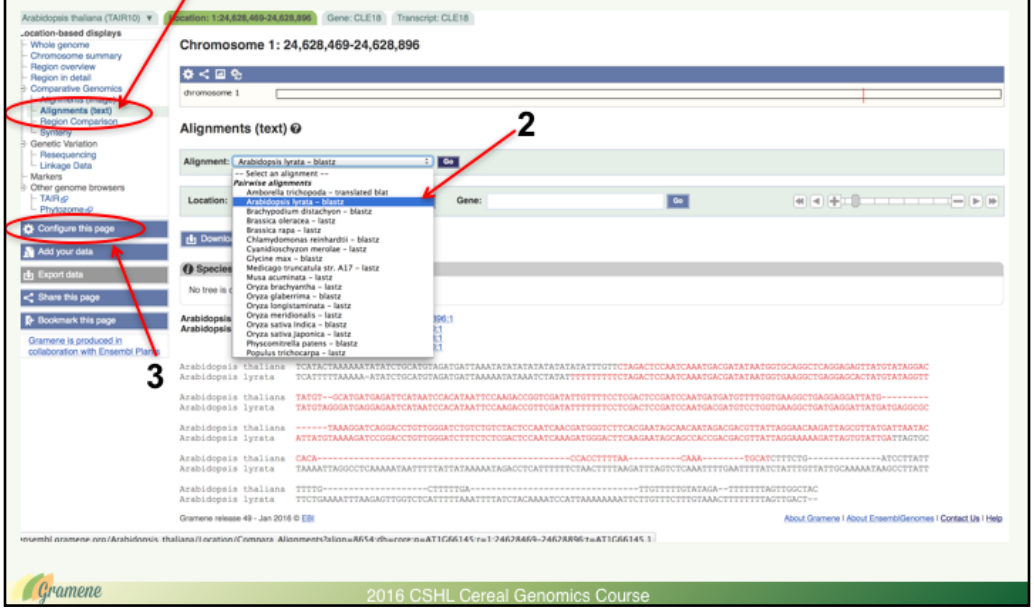


Go to location page

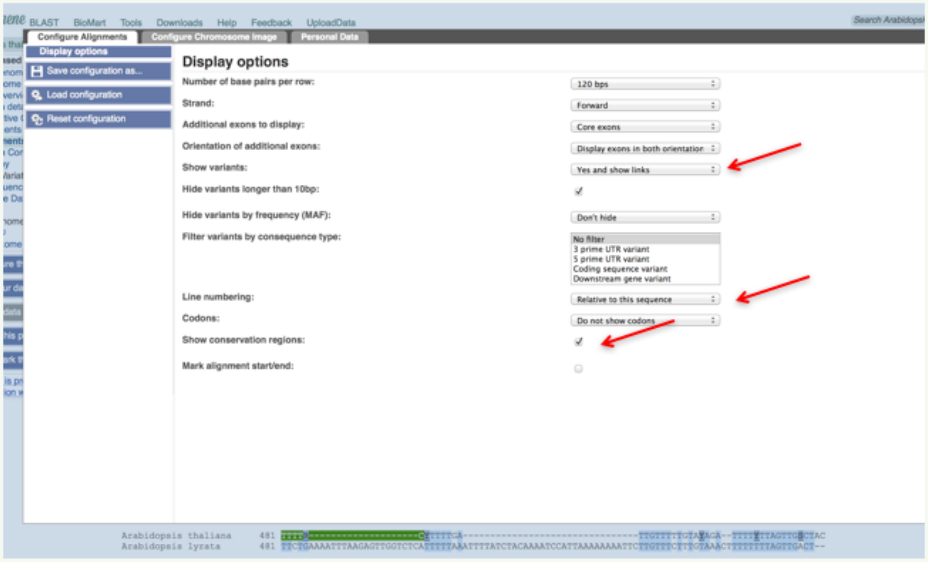
Configure (location) page: Show variation



Location: show variants for a sequence alignment

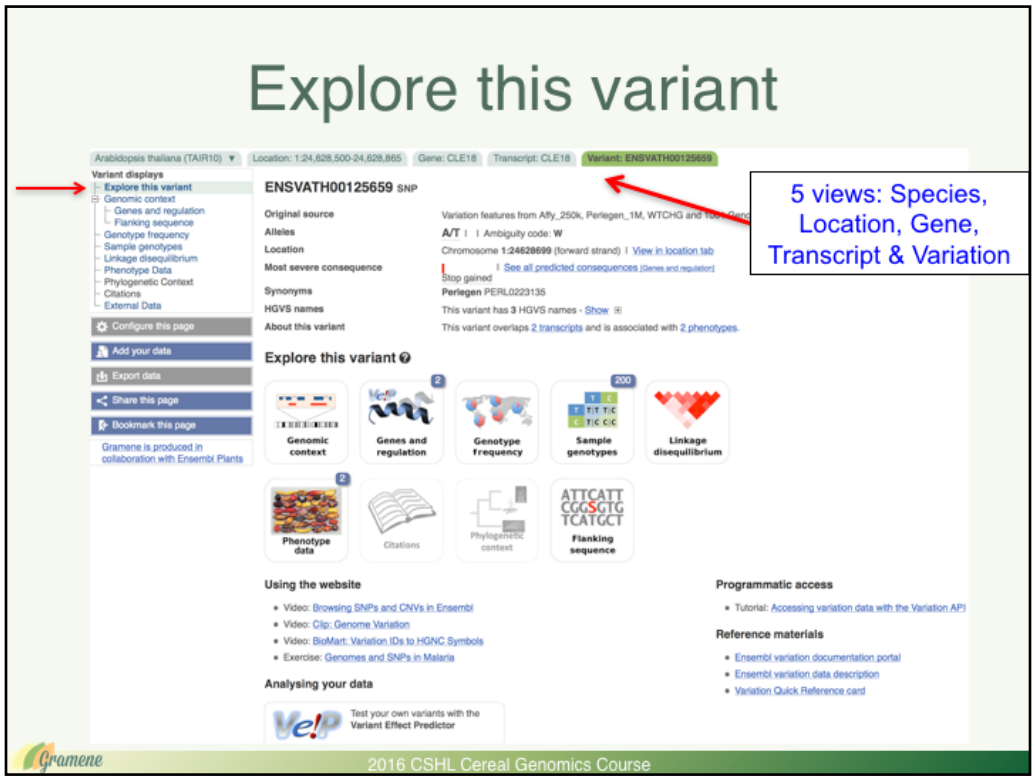


Configure (Alignments) page



Sequence alignment with variants

Transcript Sequence: cDNA



Correspondence between icons and left bar menu

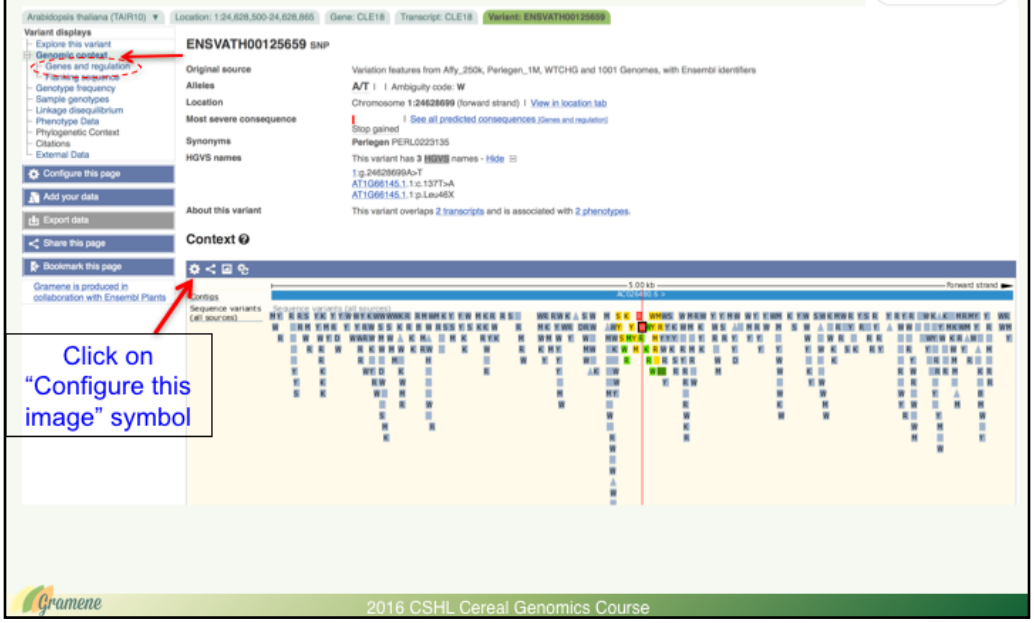
Not grayed out have data

Videos

ENSVATH00125659

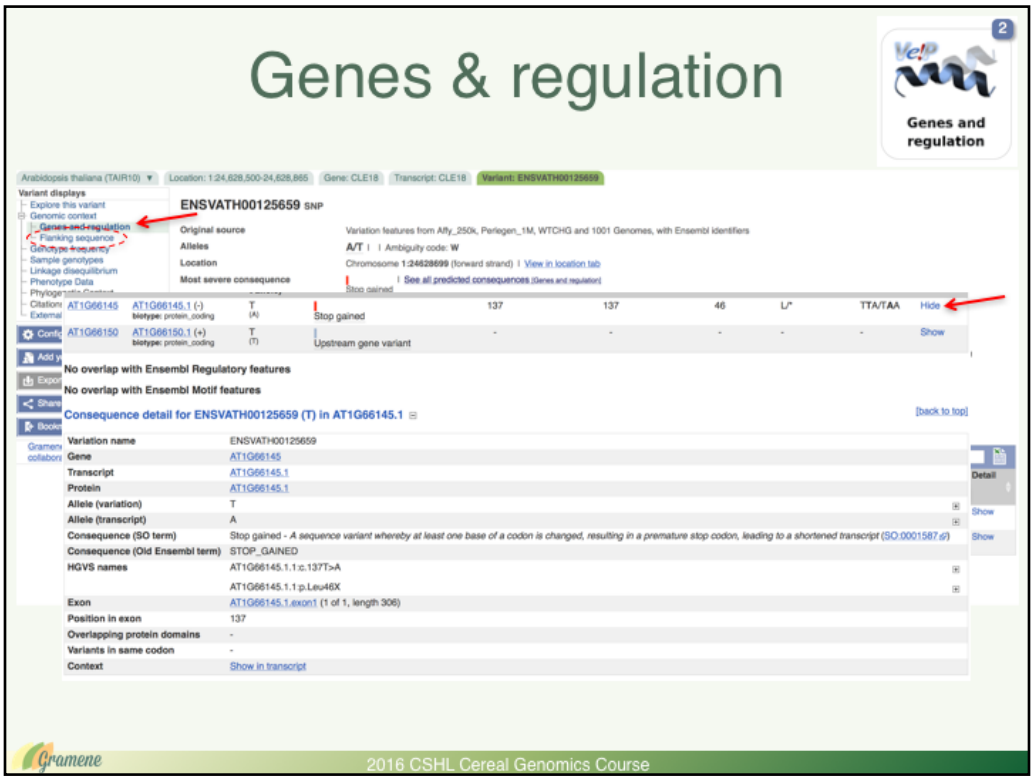
http://ensembl.gramene.org/Arabidopsis_thaliana/Variation/Explore?r=1:24628199-24629199;v=ENSVATH00125659;vdb=variation;vf=125659

Genomic context

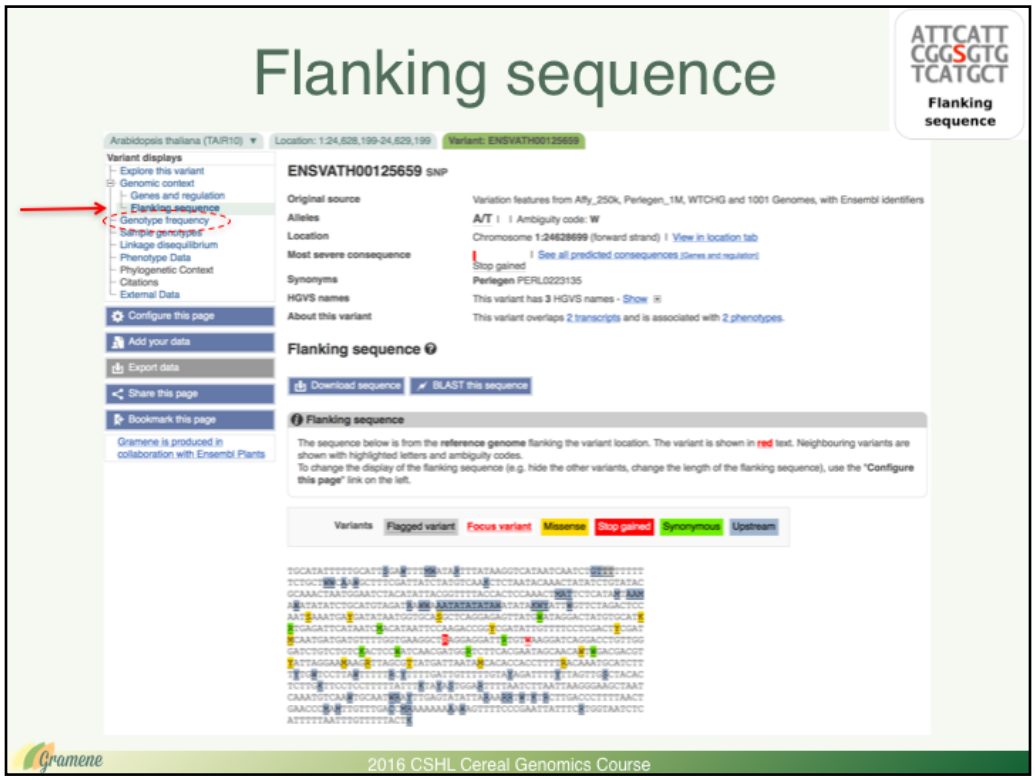


Click on "Configure this image" symbol

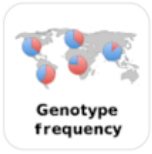
Genes & regulation



Flanking sequence



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.

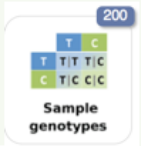
Genotype frequency

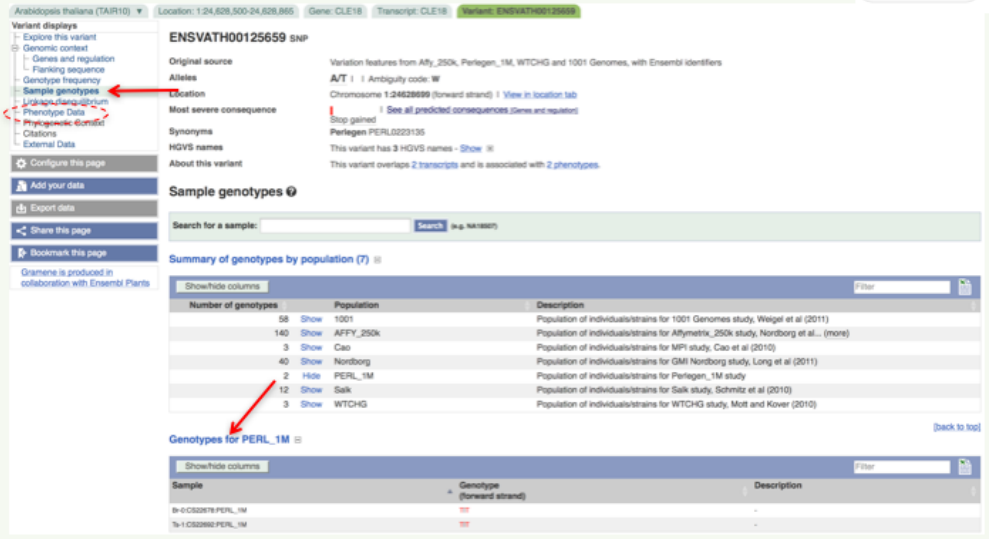
Frequency data (7)

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





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** ! ! 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.

Phenotype Data

Significant association(s)

Show/hide columns — Filter

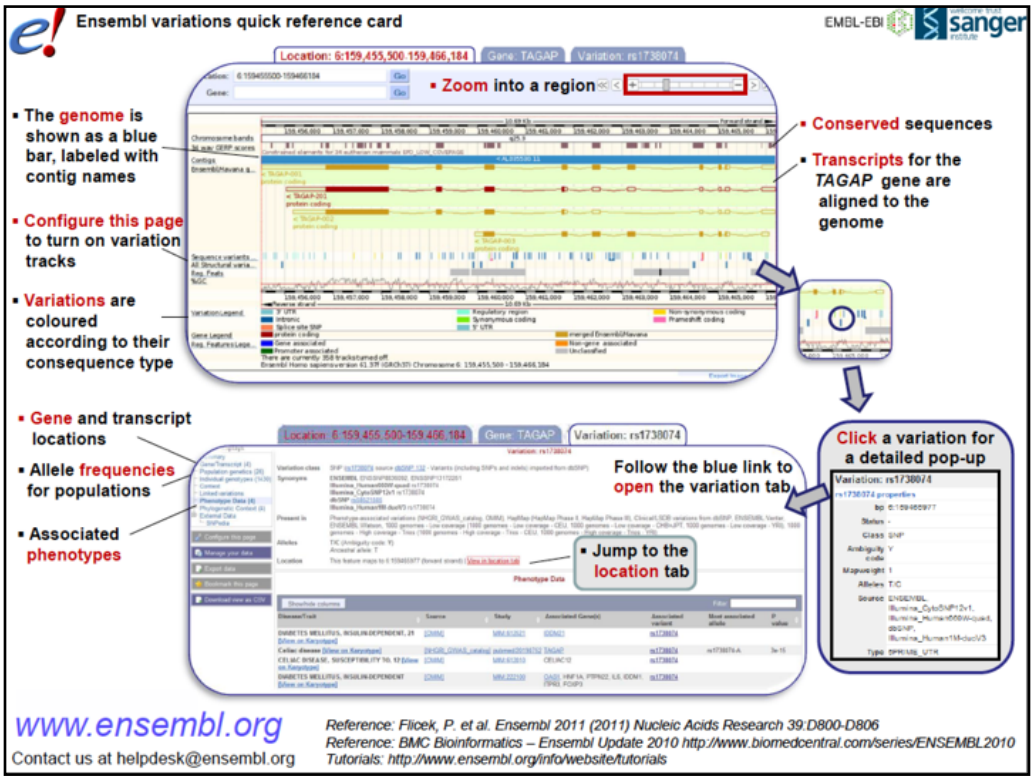
Phenotype, disease and trait	Source(s)	Study	Reported gene(s)	Associated allele	Statistics	Genomic Locations
Following inoculation of two leaves per plant with 0.1 ml of 10 -8 cfu/ml bacteria in 10 mM MgSO4 buffer using a blunt-tipped syringe. Leaf collapse was scored at 20 hrs and again at 24 hrs after 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

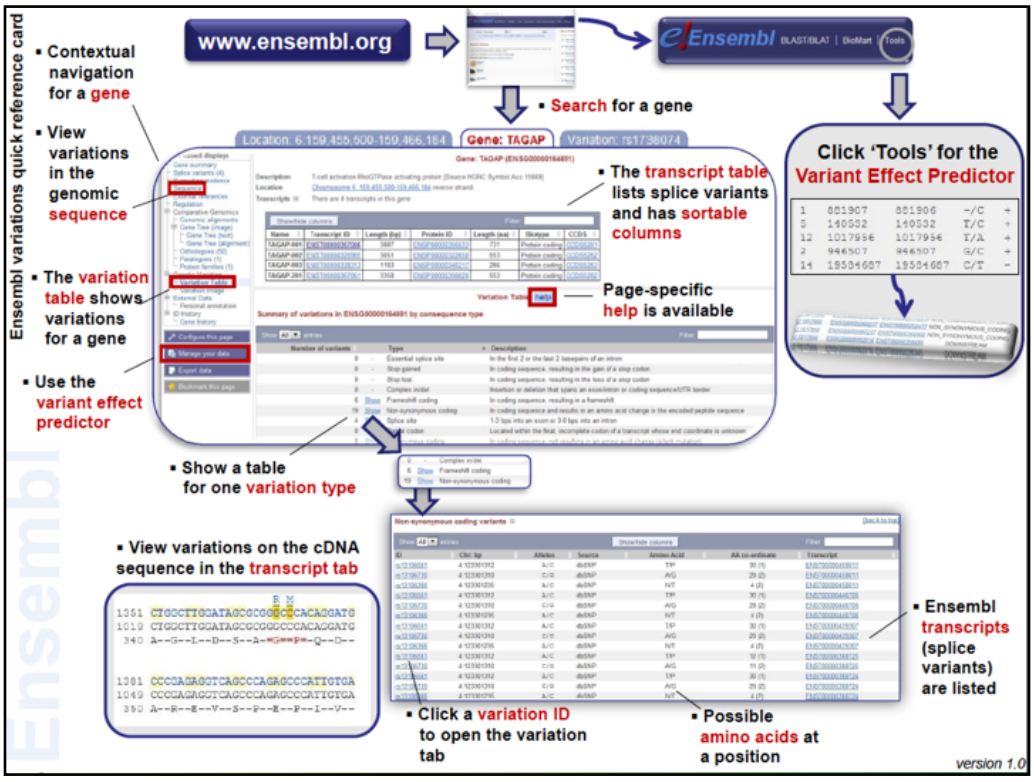
Genes in this region

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


Ensembl variations quick reference card
EMBL-EBI
sanger
Location: 6:159,455,500:159,466,184
Gene: TAGAP
Variation: rs1738074
Zoom into a region
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
Conserved sequences
Transcripts for the TAGAP gene are aligned to the genome
Gene and transcript locations
Allele frequencies for populations
Associated phenotypes
Follow the blue link to open the variation tab
Jump to the location tab
Click a variation for a detailed pop-up
Variation: rs1738074
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
Ensembl BLAST/BLAT | BioMart | Tools
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
Location: 6:159,455,500-159,466,184
Gene: TAGAP
Variation: rs138074
The transcript table lists splice variants and has sortable columns
Page-specific help is available
Click 'Tools' for the Variant Effect Predictor
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
version 1.0

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]