

Gramene 2023

**An online resource on comparative genomics in
plant reference genomes
&
pan genomes**

Sunita Kumari

Cold Spring Harbor Laboratory

Gramene Timeline (dates from publications)



- 2002** Pre-genomic era for crop genomes. Provided Rice maps projections to other species.
- 2006** Only rice had been sequenced and annotated. Provided syntenic relationships.
- 2007** Introduced pathways for functional annotation of rice genes
- 2009** **Formal collaboration with Ensembl Genomes – core databases and annotation**
- 2013** **Plant Reactome** portal featuring curated rice pathways
- 2016** Drupal CMS, Archive site, **EBI Expression Atlas**
- 2018** Integrated search engine, embedded views
- 2021-23** **Pan genome sites and CLIM Tools**



Sara Dyer



Pankaj Jaiswal



Irene Papatheodorou

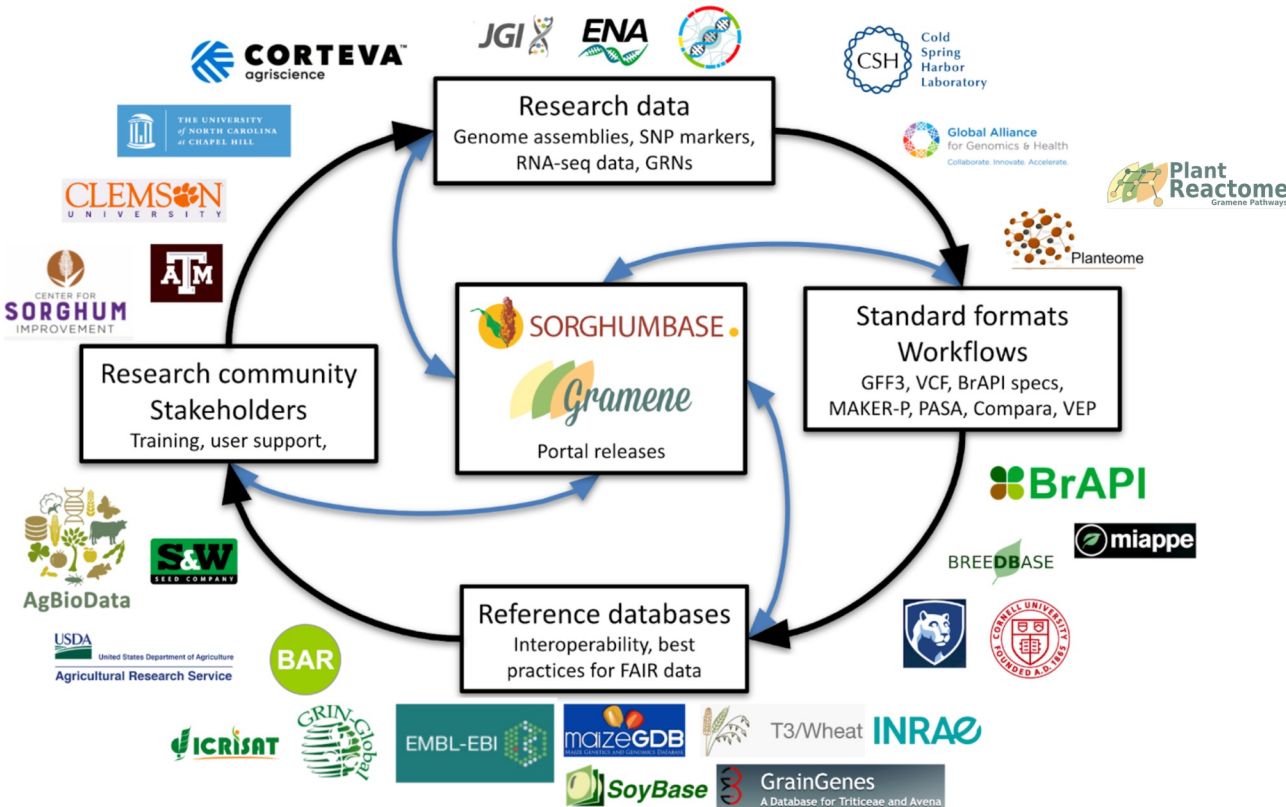


Sally Assmann



- **Findable:** Gramene datasets can be easily searched on the Gramene main site and its pan genome sites.
- **Accessible:** Gramene datasets do not require authentication or authorization and are publicly available.
- **Interoperable:** Gramene web services provide data in standard and machine readable formats that can be easily integrated into workflows.
- **Reusable:** Gramene includes datasets that are publicly available and do not require a subscription. So the data can be reused in any external apps.

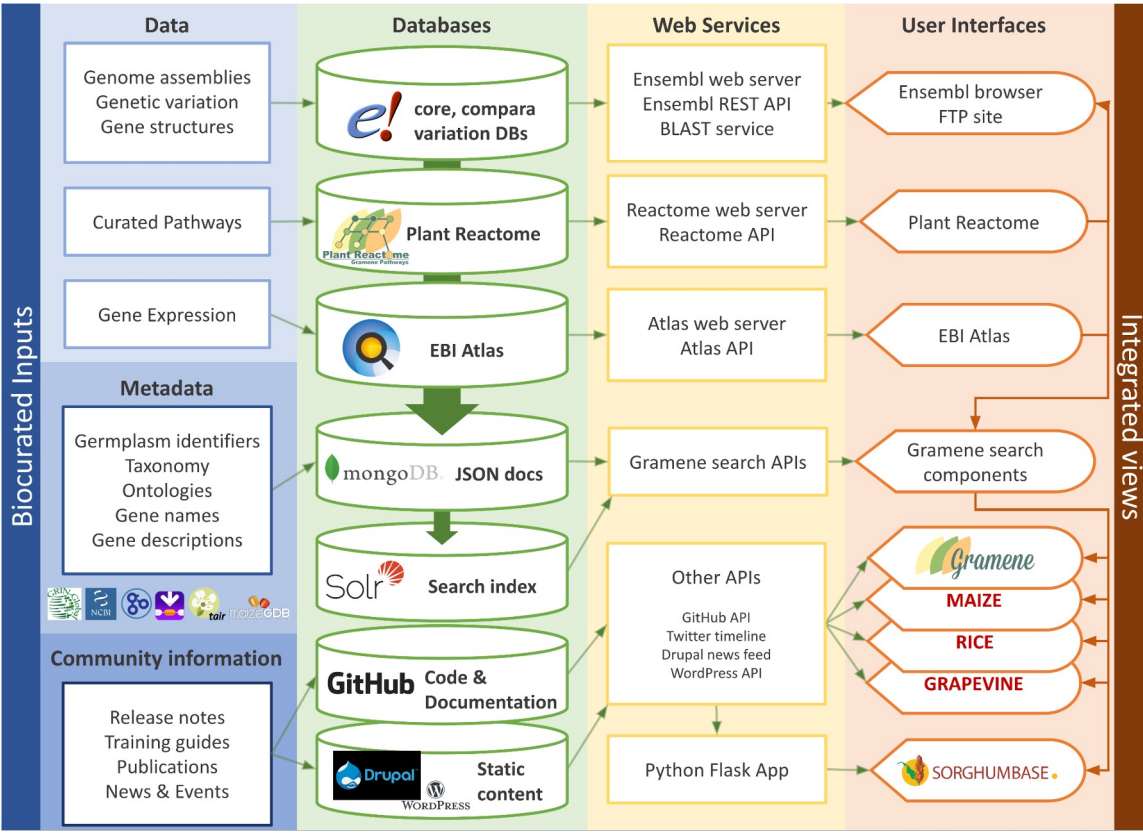
Key Collaborators



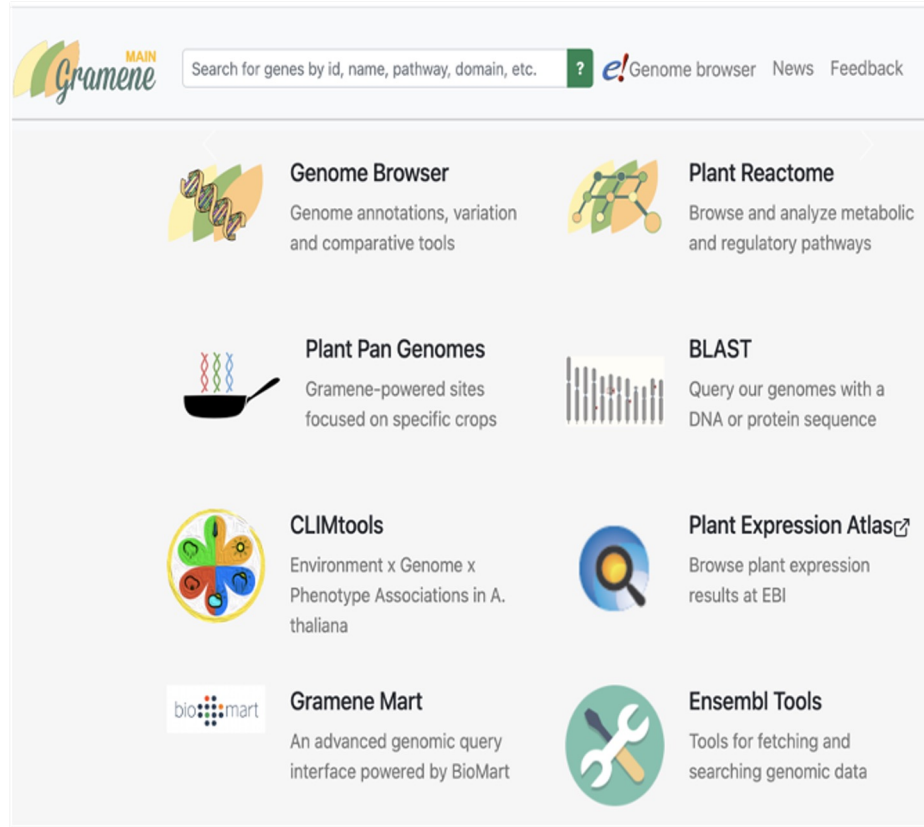
Follow FAIR Principles

1. Integrated Genomic Resources
2. Identify and apply standard formats
3. Support Interoperability with reference databases
4. Provide training and support to the research community and stakeholders

How we process inputs into integrated views



- **Ensembl “core” database**
 - Load genome and gene models
 - Protein annotation pipeline (InterProScan)
 - Select canonical transcripts (TRaCE)
- **Ensembl “variation” database**
 - Load populations and variants
 - Variant Effect Prediction (VEP) pipeline
- **Ensembl “compara” database**
 - Build genetrees (incremental update)
 - Whole genome alignments
 - Synteny maps
- **Deploy Ensembl genome browser, tools, API**
- **Build search index**
- **Local versions of EBI services**
 - BLAST, Search (EBEYE)
- **Prepare files for FTP site**
- **Release notes, news, email announcements**
- **Deploy search interface / homepage**



The screenshot shows the Gramene Main Site homepage. At the top left is the Gramene logo with the word "MAIN" above it. To the right is a search bar with the text "Search for genes by id, name, pathway, domain, etc." and a green button with a question mark. Further right are links for "e! Genome browser", "News", and "Feedback". Below the search bar is a grid of eight tool cards, each with an icon, a title, and a brief description.

Tool	Description
Genome Browser	Genome annotations, variation and comparative tools
Plant Reactome	Browse and analyze metabolic and regulatory pathways
Plant Pan Genomes	Gramene-powered sites focused on specific crops
BLAST	Query our genomes with a DNA or protein sequence
CLIMtools	Environment x Genome x Phenotype Associations in A. thaliana
Plant Expression Atlas	Browse plant expression results at EBI
Gramene Mart	An advanced genomic query interface powered by BioMart
Ensembl Tools	Tools for fetching and searching genomic data

Genome Browser - Genome Annotation, Variation and comparative tools

Ensembl Tools - Tools for fetching and searching genomic data

Gramene Mart - advanced query interface to download

CLIM Tools - E x G x P associations in Arabidopsis and Rice.

Plant Pan Genomes - Gramene powered sites

Plant Expression Atlas - Plant expression results at EBI

Plant Reactome - Analyze curated metabolic pathways of rice and other genomes

Gramene Main Site - The Website view

ensembl.gramene.org/index.html

BLAST | BioMart | Tools | Downloads | Help

Search for genomic regions, genes, transcripts, variants, phenotypes etc.

Search: for

All genomes

- ▼ -- Select a species --
- Favourite
 - Arabidopsis thaliana
 - Oryza sativa Japonica Group
 - Triticum aestivum
 - Hordeum vulgare
 - Zea mays
- Liliopsida
 - Aegilops tauschii
 - Ananas comosus
 - Asparagus officinalis
 - Brachypodium distachyon
 - Digitaria exilis
 - Dioscorea rotundata
 - Echinochloa crus-galli
 - Eragrostis curvula
 - Eragrostis tef
 - Hordeum vulgare
 - Hordeum vulgare cultivars
 - Leersia perrieri

Favourite genomes


- Arabidopsis thaliana**
TAIR10
- Oryza sativa Japonica** C
IRGSP-1.0
- Triticum aestivum**
IWGSC
- Hordeum vulgare**
MorexV3_pseudomolecules
- Zea mays**
Zm-B73-REFERENCE-NAM

- Access to 128 Plant Genomes; Gene models with functional & structural annotations.
- Protein-based gene trees provide access to orthologs and paralogs
- whole-genome alignments between each genome and a reference monocot and dicot genomes & synteny maps.
- Genetic variation amounts to >238 M markers including SNPs, QTLs, SSR/RFLPs and EMS-induced variants.

Tools




We provide a number of ready-made tools for processing both our data and yours. We routinely delete results from our servers after 10 days, but if you have an [ensembl account](#) you will be able to save the results indefinitely.

Processing your data

Name	Description
Variant Effect Predictor 	Analyse your own variants and predict the functional consequences of known and unknown variants via our Variant Effect Predictor (VEP) tool.
HMMER	Quickly search our genomes for your protein sequence.
BLAST/BLAT	Search our genomes for your DNA or protein sequence.
Assembly Converter	Map (liftover) your data's coordinates to the current assembly.
ID History Converter	Convert a set of Ensembl IDs from a previous release into their current equivalents.

* For larger datasets we provide an API script that can be downloaded (you will also need to install our Perl API, below, to run the script).

Accessing Ensembl Plants data

Name	Description		
BioMart	Use this data-mining tool to export custom datasets from Ensembl Plants.	Ensembl Plants BioMart	
Ensembl Perl API	Programmatic access to all Ensembl data using simple Perl scripts	GitHub or FTP download (current release only)	
Ensembl Genomes REST server	Access Ensembl data using your favourite programming language		

Ensembl Tools For Processing Your Data

- Variant Effect Predictor
- HMMER
- BLAST/BLAT
- Assembly Converter
- ID History Converter

- In this section
- Control Panel
 - Find a Data Display
 - Adding Custom Tracks**
 - BED File Format
 - Large File Formats
 - Pairwise Interactions File Format
 - GFF/GTF File Format
 - GFF3 File Format
 - PSL File Format
 - Variation File Format
 - WIG File Format
 - Coordinate Systems for custom tracks
 - Track Hubs
 - Adding Track Hubs
 - Track Hub Support in Ensembl
 - RNA-Seq studies in plants
 - Glossary
 - Supported browsers

Using this website

Find a Data Display

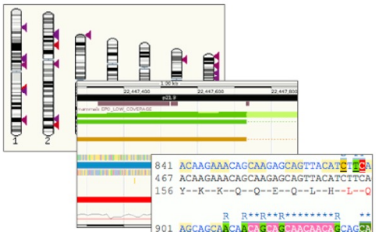
Not sure how to find the data visualisation you need? Choose the type of data you are interested in, then browse a selection of relevant visualisations!

Species:

Feature type:

- Genes
- Genomic locations
- Variants

Identifier:



Upload Data

Adding Custom Tracks:

- BED File Format
- Large File Formats
- Pairwise Interactions File Format
- GFF/GTF File Format
- GFF3 File Format
- PSL File Format
- Variation File Format
- WIG File Format

[See answer →](#)

- How do I convert IDs? I have ENSG... IDs and I would like HGNC symbols and EntrezGene IDs along with matching Affymetrix platform HC G110 probes.

[See answer →](#)

[Download a sequence or region](#)

Export data

CAGATGAT
AAATGTTCT
AAGAAGCA
CTGTCTGTC
ATAAAGAA
AGTGATACT

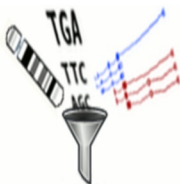


Click on the 'Export data' button in the lefthand menu of most pages to export:

- FASTA sequence
- GTF or GFF features

...and more!

[Customise your download](#)



Custom datasets can be retrieved using the BioMart data-mining tool.

You may find exploring this web-based query tool easier than extracting information direct from our databases.

[Fetch data programmatically](#)



Write your own Perl scripts to retrieve small-to-medium datasets. All our data, as well as added functionality, is available through the Ensembl Perl API.

Use the API to retrieve gene and transcript sets, fetch alignments between sequences, compare allele frequencies and much more!

You can also use our [REST API](#) to retrieve data to process in the programming language of your choice.

[Download databases & software](#)



All of our data and software, including pipelines and web code, is available free.

- [Download data via FTP](#)
- [Ensembl pipeline on GitHub](#)
- [Set up your own Ensembl Plants website](#)

Download data:

- Export data
- Customize download using Biomart
- Fetch data programmatically using REST APIs
- Download data via FTP Or Github

Gramene Mart - Data Export



ensembl.gramene.org/biomart/martview/21213f50a7508d59d3542a78df0acb8e

BLAST **BioMart** Tools | Downloads | Help | Feedback | UploadData Search Gramene...

New Count Results URL XML Perl Help

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

Dataset
Arabidopsis thaliana genes (TAIR10)

Filters
[None selected]

Attributes
Gene stable ID
Transcript stable ID

Attributes
 Features
 Structures
 Homologues (Max select 6 orthologues)
 Variant (Germline)
 Sequences

GENE:
 EXTERNAL:
GO

External References (max 3)
 BioGRID Interaction data, The General Repository for Interaction Datasets
 ChEMBL ID
 EntrezGene transcript name ID
 European Nucleotide Archive ID
 RefSeq DNA ID
 RefSeq mRNA ID
 RefSeq ncRNA ID
 RefSeq peptide ID
 RNAcentral ID

Step 1: Dataset **Step 2: Filters** **Step 3: Attributes** **Step 4: Results**

Choose the database and species.

Narrow down the dataset.

Specify your output and what to print on your table.

Choose the format of your results and export.

Biological queries made easy.

Query multiple things (gene / variants) at once:

- ID conversions
- gene locations
- download sequences
- Export large amounts of data



CLIMtools: Environment × Genome × Phenotype Associations



AraCLIM V2

Interactive spatial analysis of local Arabidopsis environments



GenoCLIM V2

Genetic variation associated with environmental variables



CLIMGeno V2

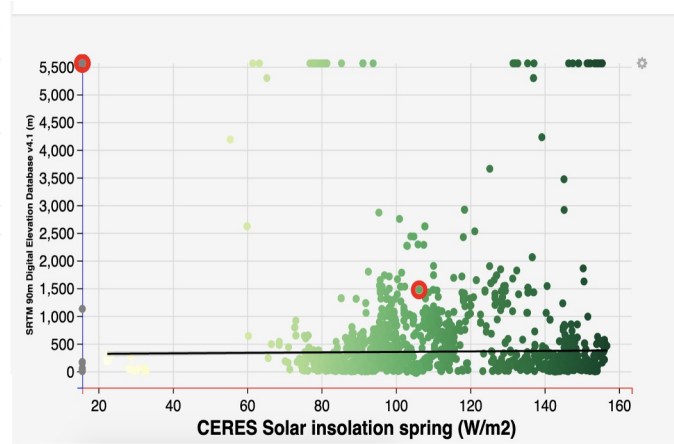
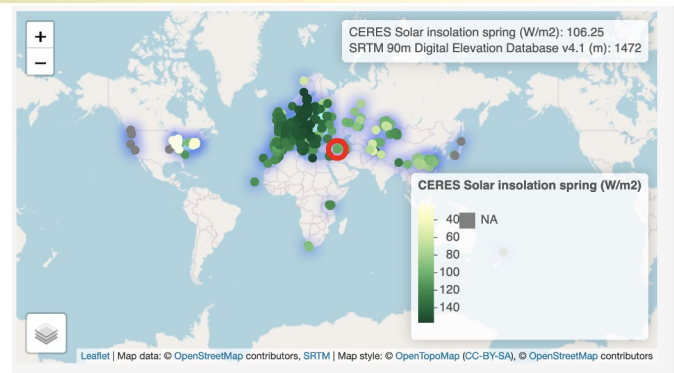
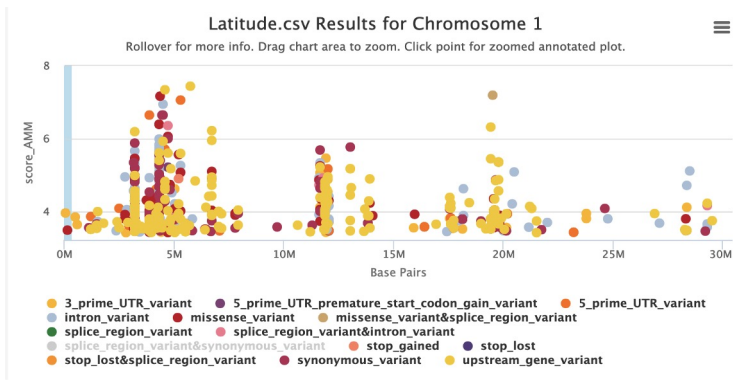
Environmental variation associated with genes or variants



T-CLIM V2

Association between gene expression and environmental variables

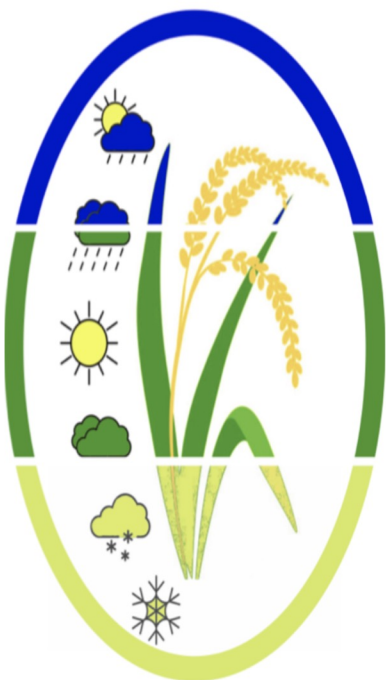
RiboSNitch prediction
SNPs that change RNA secondary structure between the reference & the alternative allele.



Ferrero-Serrano *et al* (2022) 10.1186/s13059-022-02656-4



OryzaCLIMtools



GenoCLIM

What is the GxE of my gene of interest?



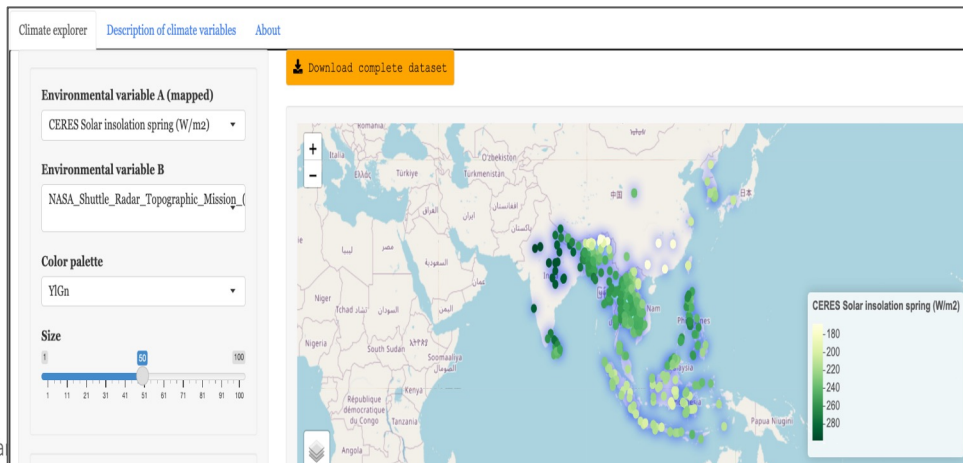
OryzaCLIM

What is the local environment of sequenced landraces?



CLIMGeno Japonica / Indica

What is the GxE for my environmental variable of interest?



- 413 geo-environmental variables
- 658 indica and 283 Japonica landraces fully sequenced by the 3K Rice Genome Project
- RiboSNitch prediction
- Climate GWAS

Crop specific portals built on the Gramene infrastructure



Comparative plant genomics focused on maize varieties



26 maize (R3, March 2023)



Comparative plant genomics focused on rice varieties



28 rice (R6, Jan 2023)



Comparative plant genomics focused on grapevine varieties



18 grapevine (R3, May 2022)



Genomic resources for the sorghum research community



29 sorghum (R6, June 2023)

Common reference genomes (8)

- *Maize B73*
- *Sorghum BTx623*
- *Rice Nipponbare*
- *Grapevine PN40024*
- *Populus trichocarpa*
- *Arabidopsis thaliana*
- *Selaginella moellendorffii*
- *Chlamydomonas reinhardtii*

Search for lox9 on Maize Pan Genome



Search for genes by name, domain, etc. ? e! Genome browser News Guides Feedback

1 genes in 1 genomes **Filters** AND Genes | lox9

Location View

lox9 Zm00001eb005920 [Zea mays B73 v5](#) Model Species Homolog

GRMZM2G017616, Zm00001d027893, lipoxygenase9

LOX3 *Arabidopsis thaliana*
Lipoxygenase 3, chloroplastic

Location Expression Homology Pathways Papers Xrefs

Currently viewing: 1:16842852-16851091

200bp 5kb 50kb 1Mb

Search Filters

Links to other resources

- [Ensembl Browser](#)
- [MaizeGDB](#)

Homology view (Alignments overview with interpro domains)

Shows inferred evolutionary histories from Compara, integrated with functional domain information from InterPro.

lox9 Zm00001eb005920 [Zea mays B73 v5](#)

GRMZM2G017616, Zm00001d027893, lipoxygenase9

LOX3 *Arabidopsis thaliana*
Lipoxygenase 3, chloroplastic

Location Expression **Homology** Pathways Papers Xrefs

Compara Gene Tree

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

Display Mode Color Scheme

- Alignment overview **Zm00001eb005920**
- Multiple Sequence Alignment 017616v3 27893
- Neighborhood conservation 0509 3140

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

Search Filters

Gene Tree | Homologs **566** | Gene Tree | Orthologs **45** | Gene Tree | Paralogs **12**

Homology view (Multiple Sequence Alignment - Amino acid view)



Inferred evolutionary histories from Compara, integrated with functional domain information from InterPro.

lox9 Zm00001eb005920 [Zea mays B73 v5](#)

GRMZM2G017616, Zm00001d027893, lipoxigenase9

LOX3 *Arabidopsis thaliana*
Lipoxygenase 3, chloroplastic

Location Expression **Homology** Pathways Papers Xrefs

Compara Gene Tree

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

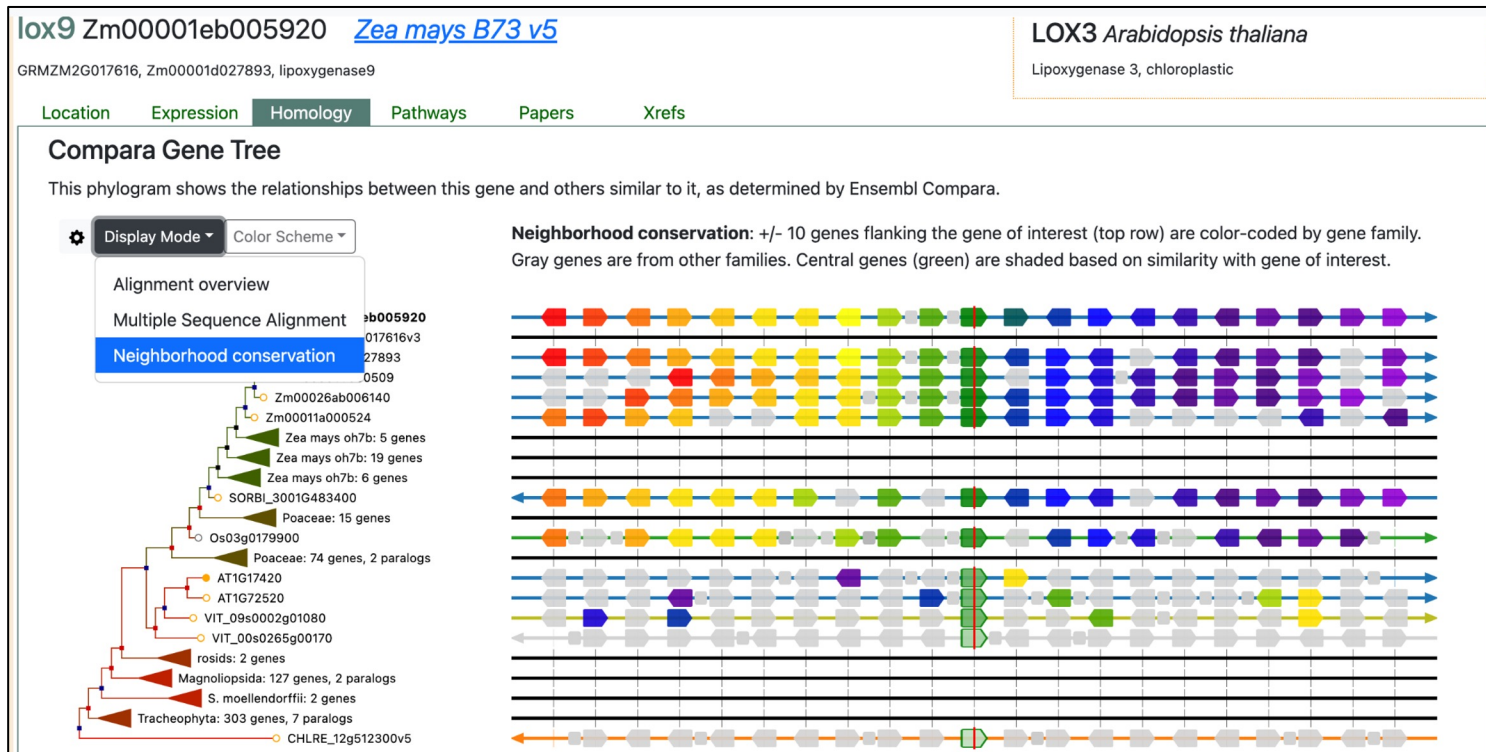
Display Mode Color Scheme

- Alignment overview
- Multiple Sequence Alignment**
- Neighborhood conservation

Multiple Sequence Alignment: Amino acid MSA. Drag slider to reposition.

Homology view (Neighborhood conservation view)

Inferred evolutionary histories from Compara, integrated with functional domain information from InterPro.



Expression View (gene expression panel in pan genome browser)

lox9 Zm00001eb005920 [Zea mays B73 v5](#) Model Species Homolog

GRMZM2G017616, Zm00001d027893, lipoxigenase9

LOX3 *Arabidopsis thaliana*
Lipoxygenase 3, chloroplastic

Location Expression Homology Pathways Papers Xrefs

All Studies [Reference Study \(all paralogs\)](#) **All Studies**

Showing 9 experiments: ↓ By experiment type Filters Download

Studies

- T [Walley et al.](#)
- T [6 Wang](#)
- T [5 Xu et al.](#)
- T [Shuai Wang et al., 2014 - Organism part](#)
- T [Koch et al., 2016 - Organism part - 20](#)
- T [Koch et al., 2016 - Organism part - 14](#)
- T [4 Chettoor et al.](#)

Tissues & dev stages

- 4 mm from tip of ear primordium
- basal endosperm transfer layer
- ear
- embryo 20 days after pollination
- embryo sac
- endosperm
- endosperm crown
- growth zone of leaf 8
- internode 6-7
- leaf mesophyll
- mature female spikelet
- mature pollen
- pericarp and aleurone
- plant ovule
- primary root
- root cortex
- root elongation zone
- shoot
- style
- tassel
- vegetative meristem

EBI-Atlas data visualization tools are connected to Gramene, Ensembl Plants & Plant Reactome resources

Expression View (gene expression panel in pan genome browser)

lox9 Zm00001eb005920 [Zea mays B73 v5](#)

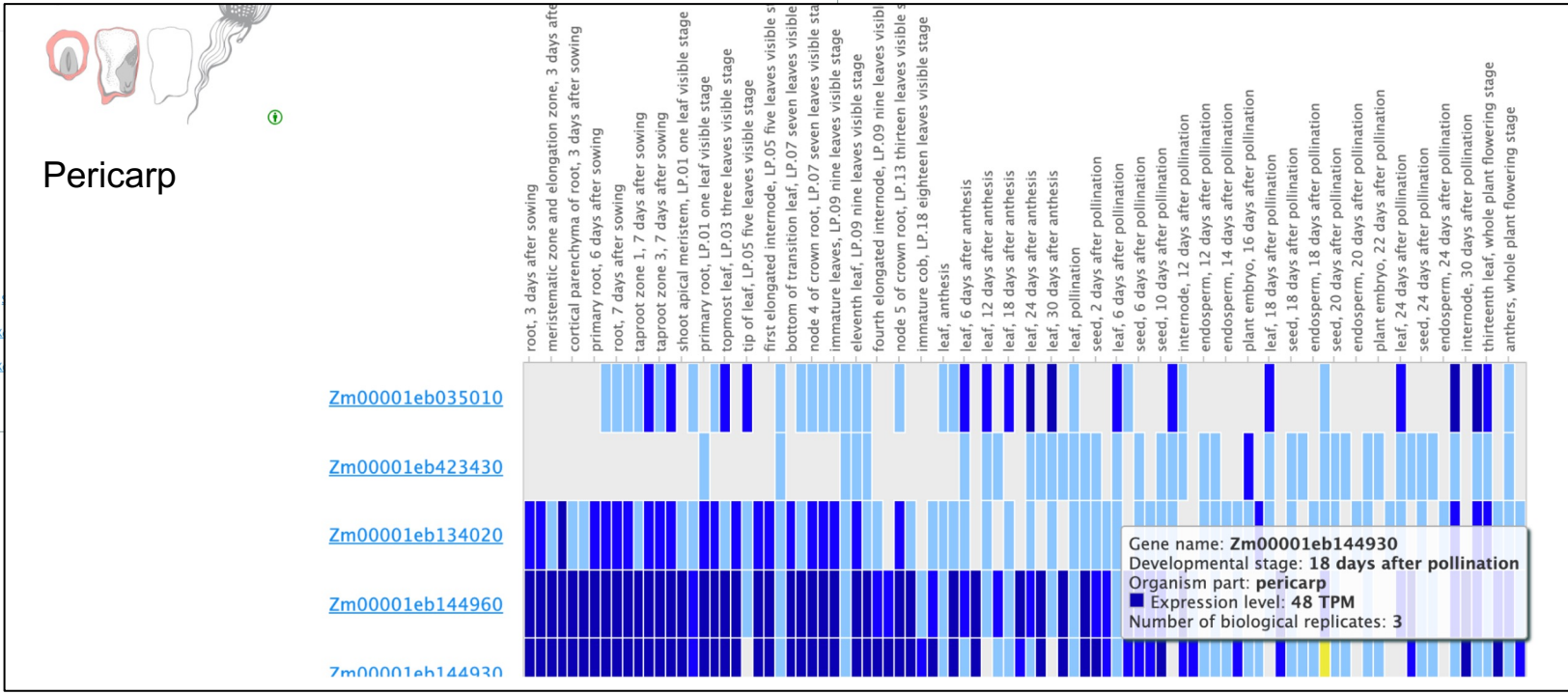
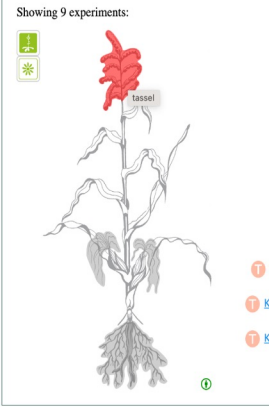
LOX3 *Arabidopsis thaliana*
Lipoxygenase 3, chloroplastic

Reference Study (all paralogs)

GRMZM2G017616, Zm00001d027893, lipoxygenase9

Location Expression Homology Pathways Papers Xrefs

All Studies [Reference Study \(all paralogs\)](#)



Pathway View

Shows the reactions & pathways and offers links to data downloads in Plant Reactome

lox9 Zm00001eb005920 [Zea mays](#)

GRMZM2G017616, Zm00001d027893, lipoxygenase9

Location Expression Homology **Pathways** Papers Xrefs

Metabolism and regulation
Hormone signaling, transport, and metabolism
Jasmonic acid biosynthesis
O2+linolenate->13-HPOT
Secondary metabolism
13-LOX and 13-HPL pathway
O2+linolenate->13-HPOT

JA Biosynthesis and 13 LOX/13HPL

Search for a term, e.g. plen...

Model Species Homolog
LOX4 *Arabidopsis thaliana*
PLAT/LH2 domain-containing lipoxygenase family protein

Zm00001eb005920 - 0 entit... Interactors: Exclude

reactome

Search Filters
All genes in this Reaction
Links to other resources
• [Plant Reactome Reaction](#)

336 curated Rice pathways
Orthology-based projections
for 120 species allow inter-
species comparisons

<http://plantreactome.gramene.org>



Curated Publications (latest addition)



lox9 Zm00001eb005920 [Zea mays B73 v5](#)

GRMZM2G017616, Zm00001d027893, lipoxygenase9

Model Species Homolog

LOX3 *Arabidopsis thaliana*


Lipoxygenase 3, chloroplastic

[Location](#) [Expression](#) [Homology](#) [Pathways](#) **[Papers](#)** [Xrefs](#)

Curated publications

This gene has been described in the literature:

PubMed link	Curation source	Title/Description
19817685	geneRIF	The oxidation of 20:2 and 22:2 by 9-LOX afforded low yields of racemic 11-, 12-, 14-, and 15-hydroperoxides.
20673209	geneRIF	Lox9 was stable form pH 4.5 to 9.5 with a maximum activity at pH7.5.

- Targeted news feeds for Gramene sites
- Release notes and user guides
-  video-tutorials - 50 Videos
- Conferences: PAG, ASPB, etc.
- AgBioData: standards and best practices
- **Contact us for training, workshops, webinars and support**

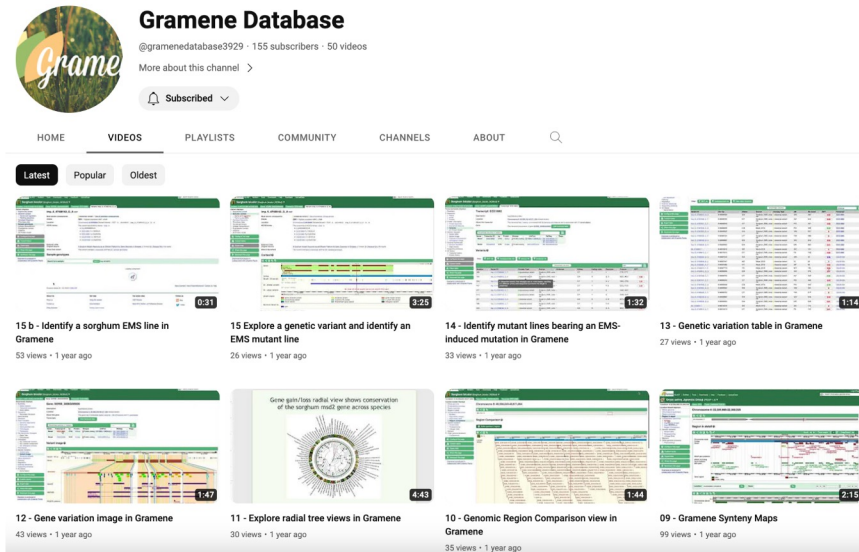
<https://gramene.org/feedback>

 Follow @GrameneDatabase



Gramene Database

@gramenedatabase3929 · 154 subscribers · 50 videos



The screenshot shows the YouTube channel page for 'Gramene Database'. The channel name is 'Gramene Database' with the handle '@gramenedatabase3929', 155 subscribers, and 50 videos. The page is set to 'Subscribed'. Below the channel header are navigation tabs for HOME, VIDEOS, PLAYLISTS, COMMUNITY, CHANNELS, and ABOUT. The video list is sorted by 'Latest' and includes the following items:

- 15 b - Identify a sorghum EMS line in Gramene (53 views • 1 year ago)
- 15 Explore a genetic variant and identify an EMS mutant line (26 views • 1 year ago)
- 14 - Identify mutant lines bearing an EMS-induced mutation in Gramene (33 views • 1 year ago)
- 13 - Genetic variation table in Gramene (27 views • 1 year ago)
- 12 - Gene variation image in Gramene (43 views • 1 year ago)
- 11 - Explore radial tree views in Gramene (30 views • 1 year ago)
- 10 - Genomic Region Comparison view in Gramene (35 views • 1 year ago)
- 09 - Gramene Synteny Maps (99 views • 1 year ago)

Acknowledgements



Ware Lab 2023



Posters @ ASPB 2023:

- #500-08: **Gramene** - Janeen Braynen
- #500-24: NUE GRN - Janeen Braynen
- #700-09: **SorghumBase** - Nick Gladman
- #800-64: GRAS TFs - Nick Gladman
- #800-56: **DAP-seq** - Audrey Fahey
- #500-22: **AgBioData** – Sunita Kumari
- #1100-19: QPSI - Sunita Kumari



We gratefully acknowledge support from
USDA-ARS-8062-21000-041-00D