(www.gramene.org)

Mining Genomes With Gramene

The project is supported by NSF IOS-1127112 and USDA

Marcela Karey Tello-Ruiz[1], Sharon Wei[1], Andrew Olson[1], Justin Preece[2], Parul Gupta[2], Sushma Naithani[2], Joshua Stein[1], Yinping Jiao[1], Bo Wang[1], Sunita Kumari[1], Young Koung Lee[1,3], Lifang Zhang[1], Vivek Kumar[1], Demitri Muna[1],
Dan Bolser[4], Matthieu Muffato[4], Peter D'Eustachio[5], Irene Papatheodorou[4], Paul J. Kersey[3], Pankaj Jaiswal[2], Doreen Ware[1,6]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; [2]Oregon State University, Corvallis, OR, USA; [3]Wonkwang University, Iksan 54538, Korea; [4]EMBL-EBI, Hinxton, UK, [5]NYU School of Medicine, New York, NY, USA, [6]USDA ARS NAE Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA

Gramene is developed in collaboration with Ensembl Plants, and leverages the Ensembl & Reactome platforms

Specific Aim 1: GENOMES

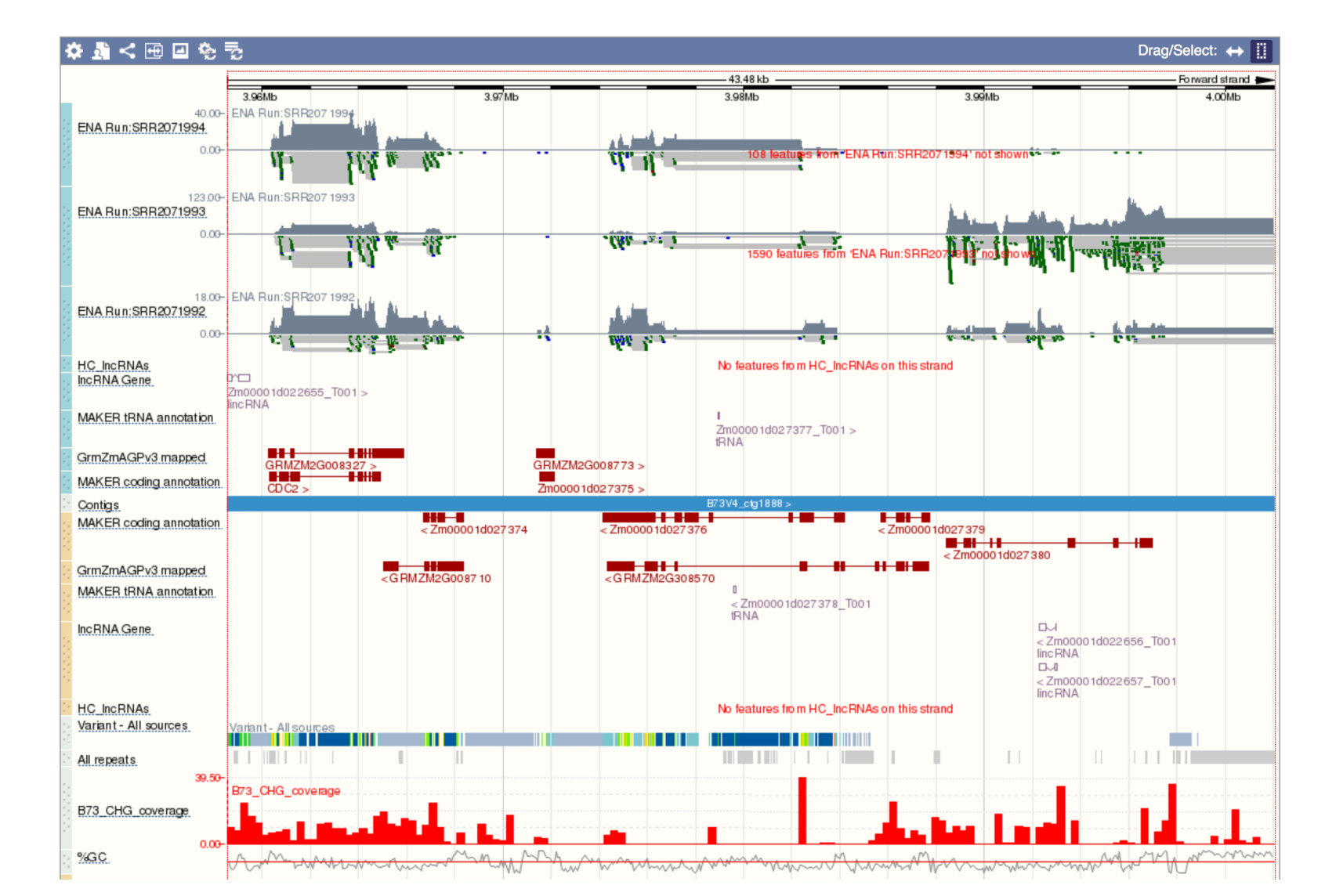
Year Highlights

- 5 database releases: build 53 (May 2017); build 54 (July 2017); build 55 (Sept 2017); build 56 (Jan 2018); build 56b (Feb 2018)
- Ensembl software updates V88-91
- 53 plant reference assemblies including:
  - ❖ 9 new genomes:
  - *Cucumis sativus* (cucumber)
  - *Dioscorea rotundata* (white Guinea yam)
  - *Gossypium raimondii* (cotton)
  - *Helianthus annuus* (common sunflower)
  - *Lupinus angustifolius* (blue lupin)
  - *Manihot esculenta* (cassava)
  - *Nicotiana attenuata* (coyote tobacco)
  - *Phaseolus vulgaris* (common bean)
  - *Corchorus capsularis* (Jute)
  - ❖ 4 updated genomes:
    - *Sorghum bicolor* (sorghum [V3]
    - *Glycine max* (soybean [V3])
    - *Prunus persica* (peach [V2])
    - *Hordeum vulgare* (barley [V2])
- New gene structural annotations: maize, barley & rice
- New functional annotations (e.g., InterPro domains)
- Updated protein gene trees
- New genetic variation: 13M SNPs *Arabidopsis thaliana* (1001 Genomes Project); 20,483 rice QTLs (Gramene archives & Q-TARO database) & 1,278 rice SSR/RFLPs (Gramene archives); 7.4 M EMS-induced wheat point mutations (Ksenia *et al*, 2017) to a total of 230.5 M variants
- Over 1,600 public RNA-Seq studies from Track Hub

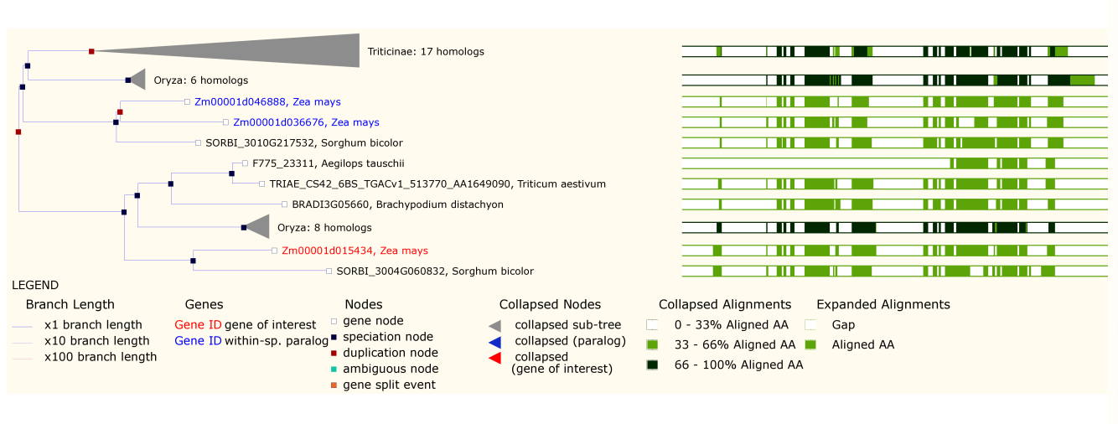
New Maize RefGenV4 Data

- Updated maize V4 gene models
- Updated protein domains from InterProScan
- Updated protein gene trees
- Gene neighborhood views for local synteny
- New whole-genome alignments (WGAs) between maize and sorghum support coding & non-coding sequence conservation
- New ncRNA feature alignments
- New transposon annotation (Jiao *et al*, 2017)
- Affy & Agilent probe arrays remapped
- New maize public RNA-Seq studies in Track Hub

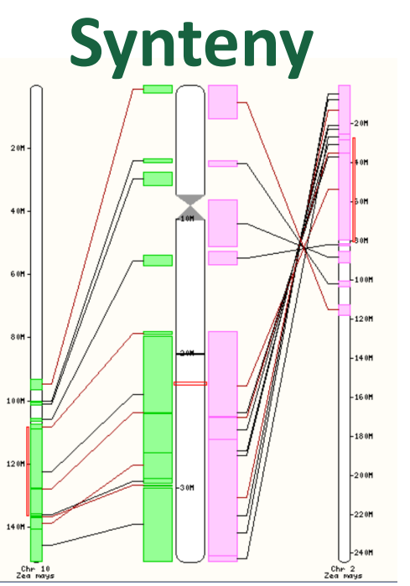
RNA-Seq Data from Public Track Hub Registry



Gene Family Trees



Synteny



Oryza sativa
Chr 4

Whole-Genome Alignments

Specific Aim 2: PATHWAYS & EXPRESSION

Year Highlights

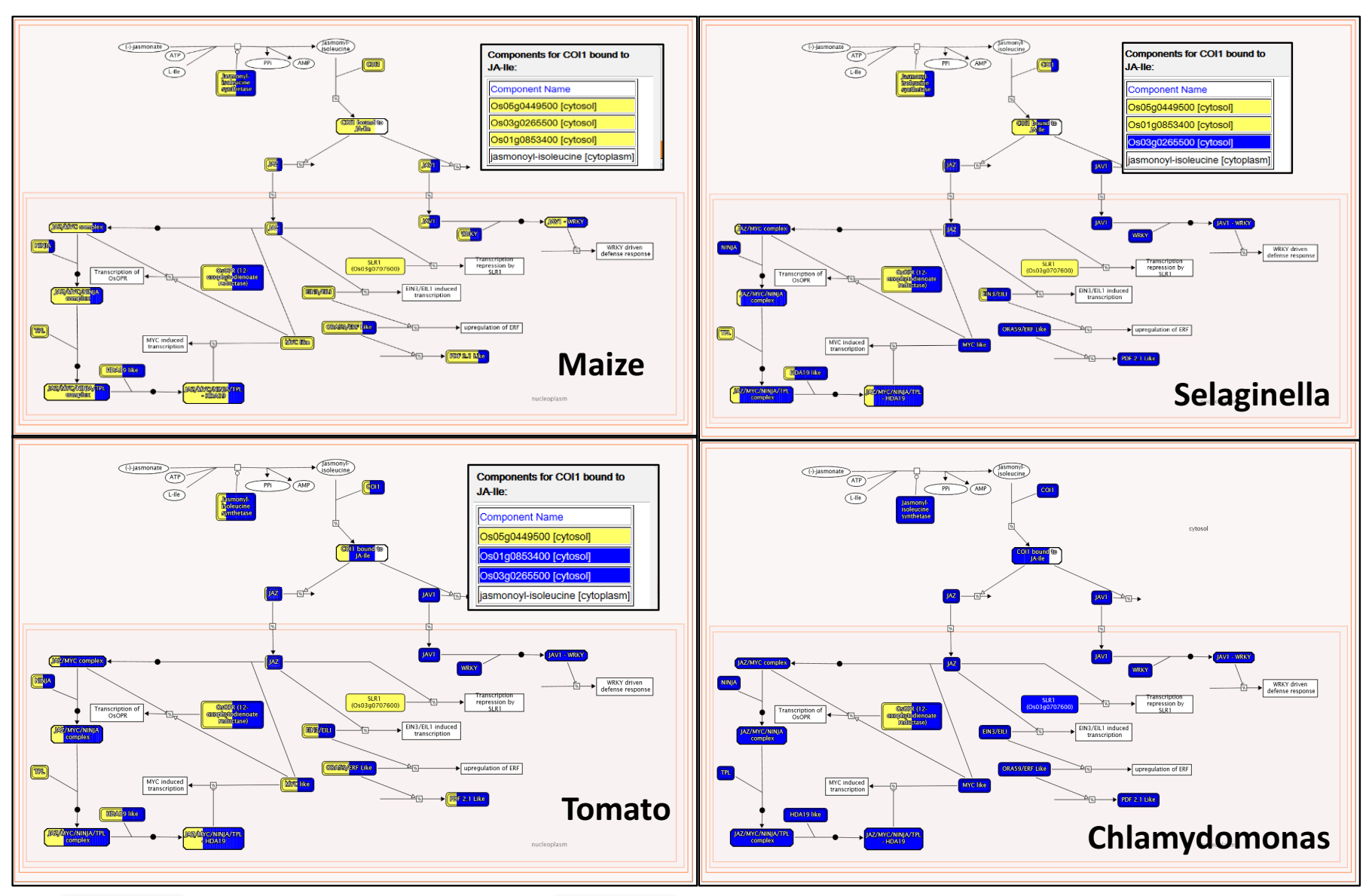
- 22 newly curated rice pathways to a total of 264 reference rice pathways
- Orthology-based projections to 74 plant species including maize allows pathway enrichment and inter-species comparison
- Baseline and differential expression data sets from 734 experiments and 18 reference plant species (Expression Atlas)

Newly curated pathways

- Reproductive structure development
- Anther and pollen development
- Vegetative structure development
- Primary root development
- Response to submergence
- Underwater shoot and internode elongation
- Response to phosphate deficiency
- Photorespiration
- Ureide biosynthesis
- Beta-alanine betaine biosynthesis
- Proline biosynthesis V from arginine
- Phospholipid biosynthesis I
- Sphingolipid metabolism
- Response to iron deficiency
- Mugineic acid biosynthesis
- Iron uptake and transport in root vascular system
- Proline degradation
- Glutamate degradation
- Phenylalanine degradation III
- Lysine degradation II
- Threonine degradation

Comparative pathway analysis

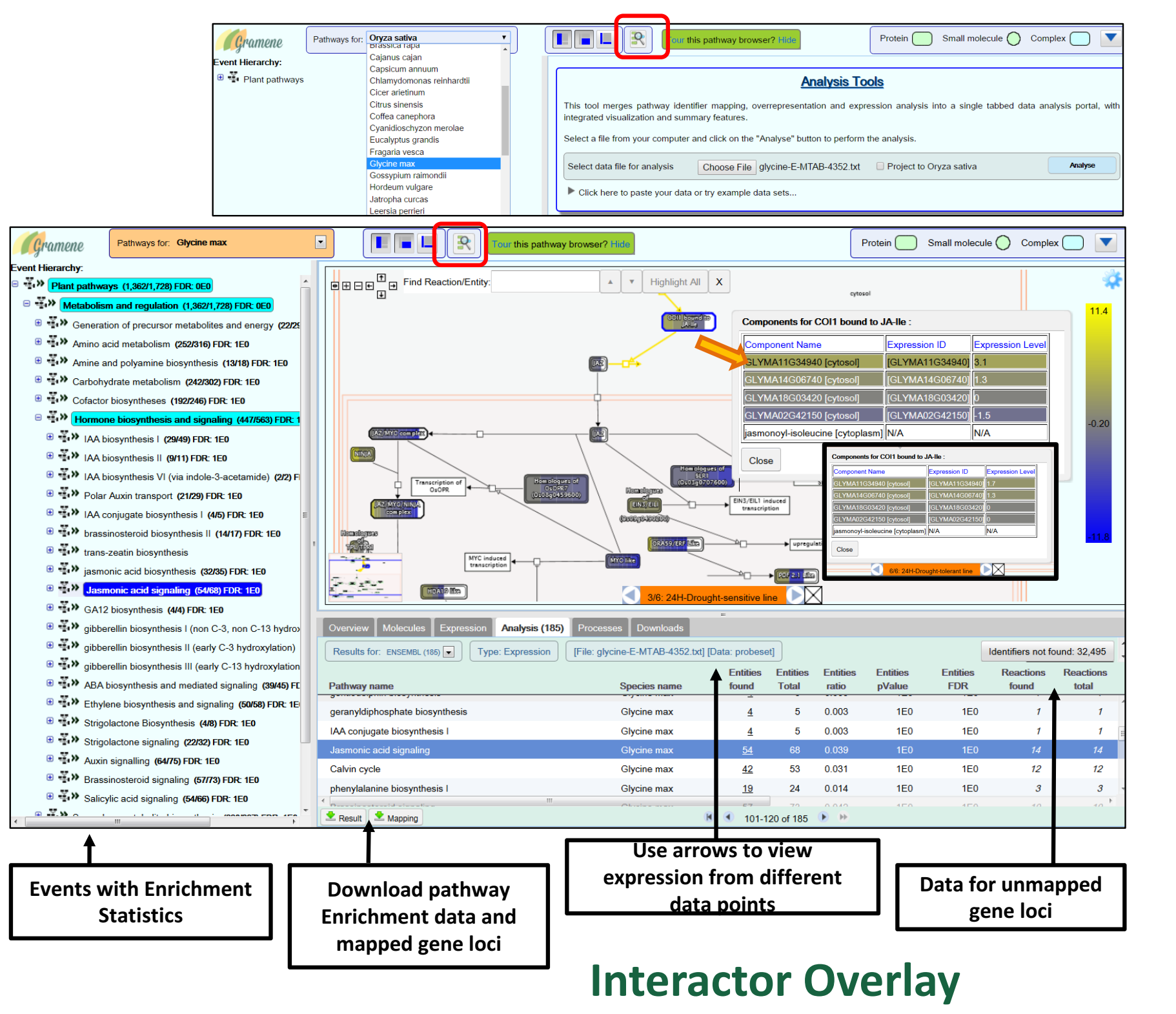
Example: JA signaling in rice vs 3 species



COLOR KEY 🟦 No homolog found 🟨 Match between rice and target species homologs
🟦🟨 If more than one gene (*e.g.* paralogs) is associated to a reaction, then only some (yellow) had ortholog match

http://plantreactome.gramene.org

OMICs data analysis tool



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.

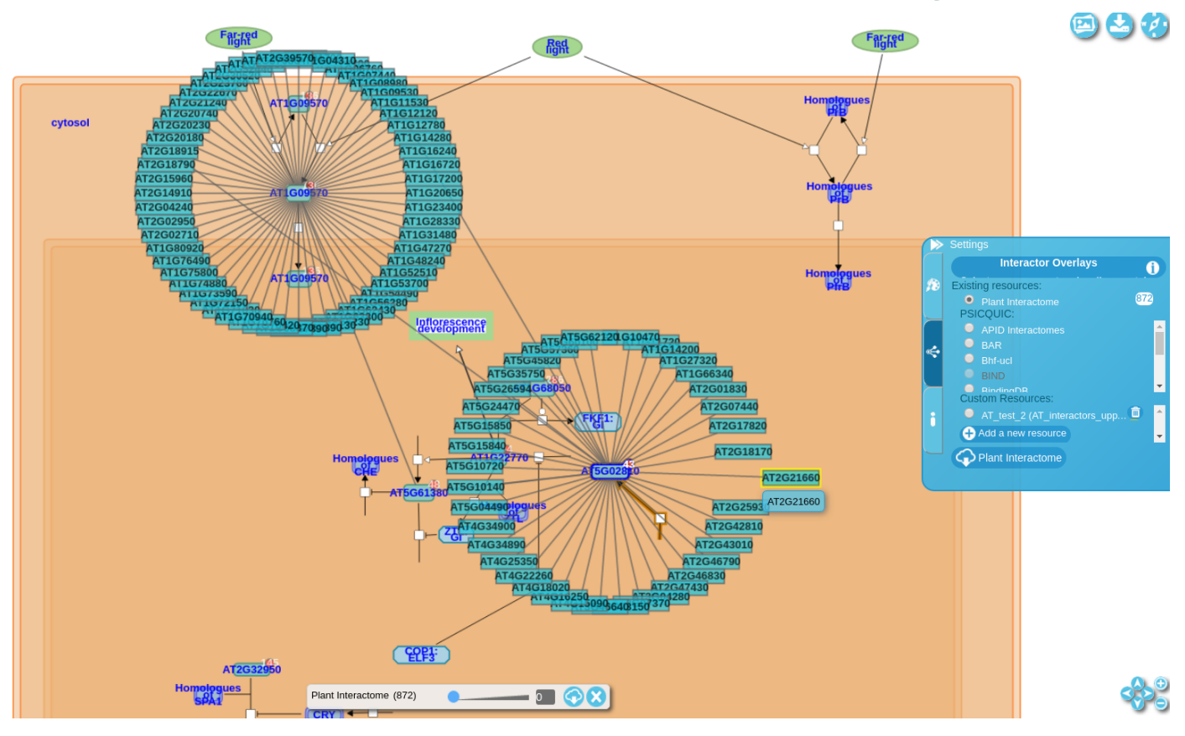
Events with Enrichment Statistics

Download pathway Enrichment data and mapped gene loci

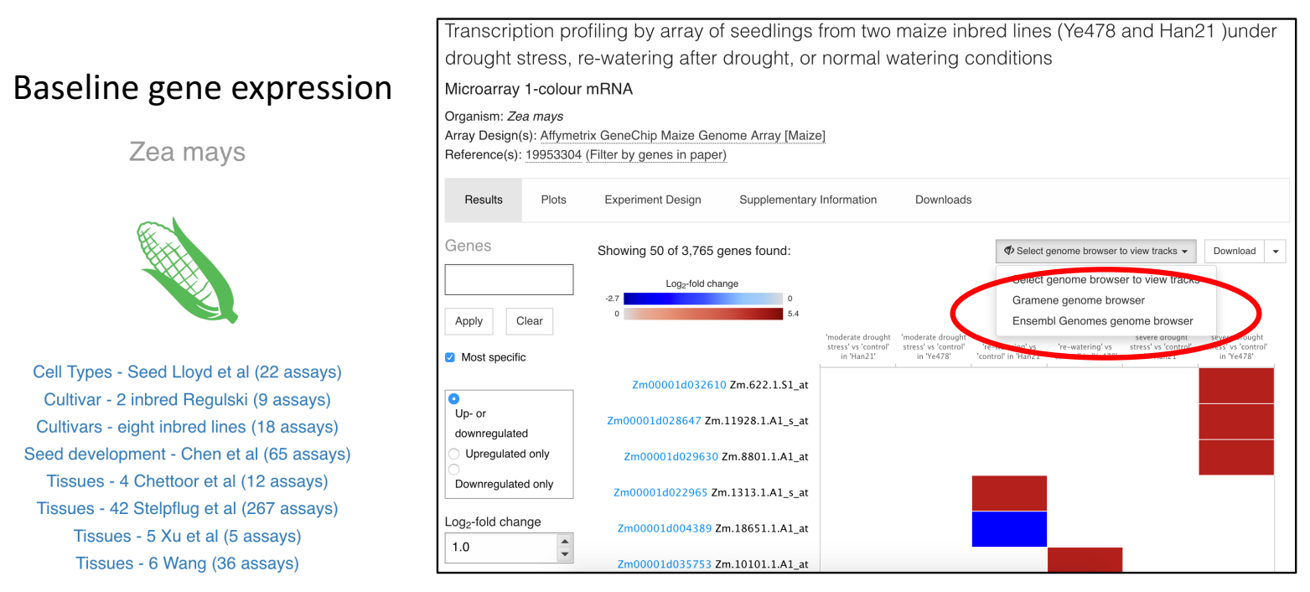
Use arrows to view expression from different data points

Data for unmapped gene loci

Interactor Overlay



ATLAS data visualization tools are connected to Gramene, Ensembl Plants and Plant Reactome resources



Baseline gene expression

Zea mays

Cell Types - Seed Lloyd et al (22 assays)
Cultivar - 2 inbred Regulski (9 assays)
Cultivars - eight inbred lines (18 assays)
Seed development - Chen et al (65 assays)
Tissues - 4 Chettoor et al (12 assays)
Tissues - 42 Stelpflug et al (267 assays)
Tissues - 5 Xu et al (5 assays)
Tissues - 6 Wang (36 assays)

Transcription profiling by array of seedlings from two maize inbred lines (Ye478 and Han21)under drought stress, re-watering after drought, or normal watering conditions

Microarray 1-colour mRNA

Organism: *Zea mays*

Array Design(s): Affymetrix GeneChip Maize Genome Array [Maize]

Reference(s): 19953304 (Filter by genes in paper)

27 differential expression experiments in maize

Specific Aim 3: INFRASTRUCTURE

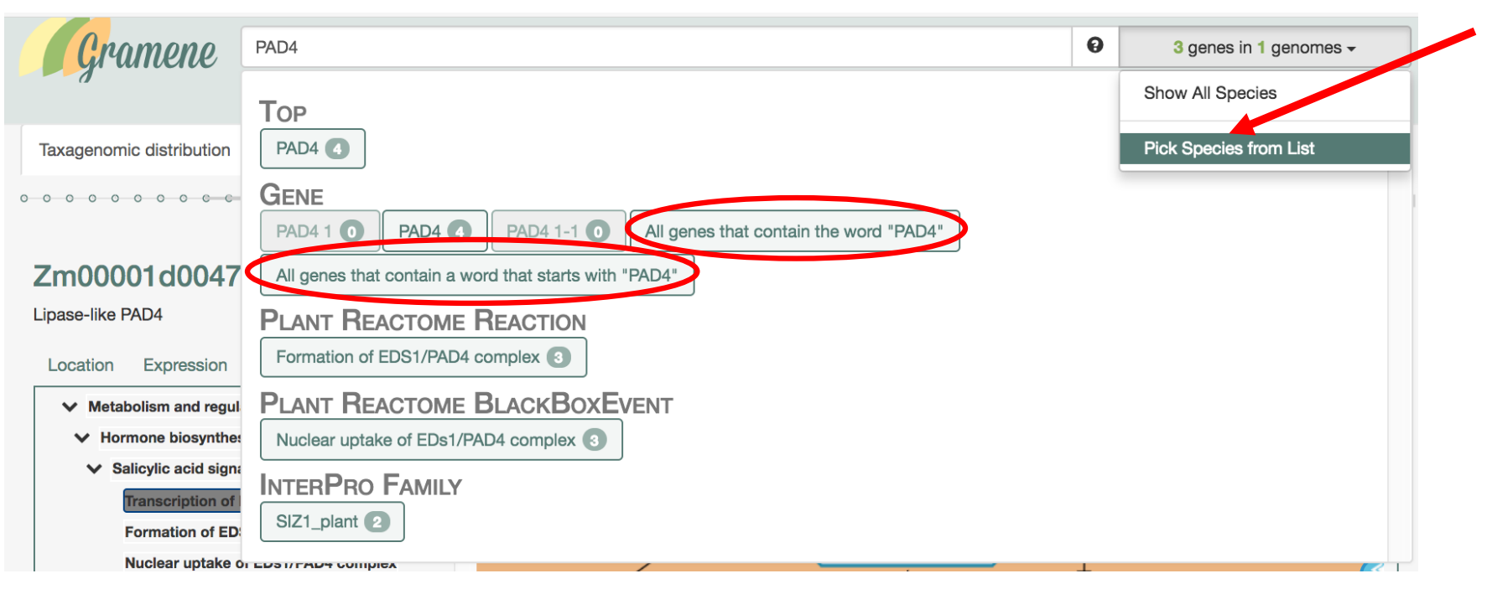
Year Highlights

- The Gramene Search Interface became Gramene's new front page http://www.gramene.org
- New pathways view in Gramene Search
- New neighborhood conservation mode in the Homology tab of the results of a query

Powerful search interface

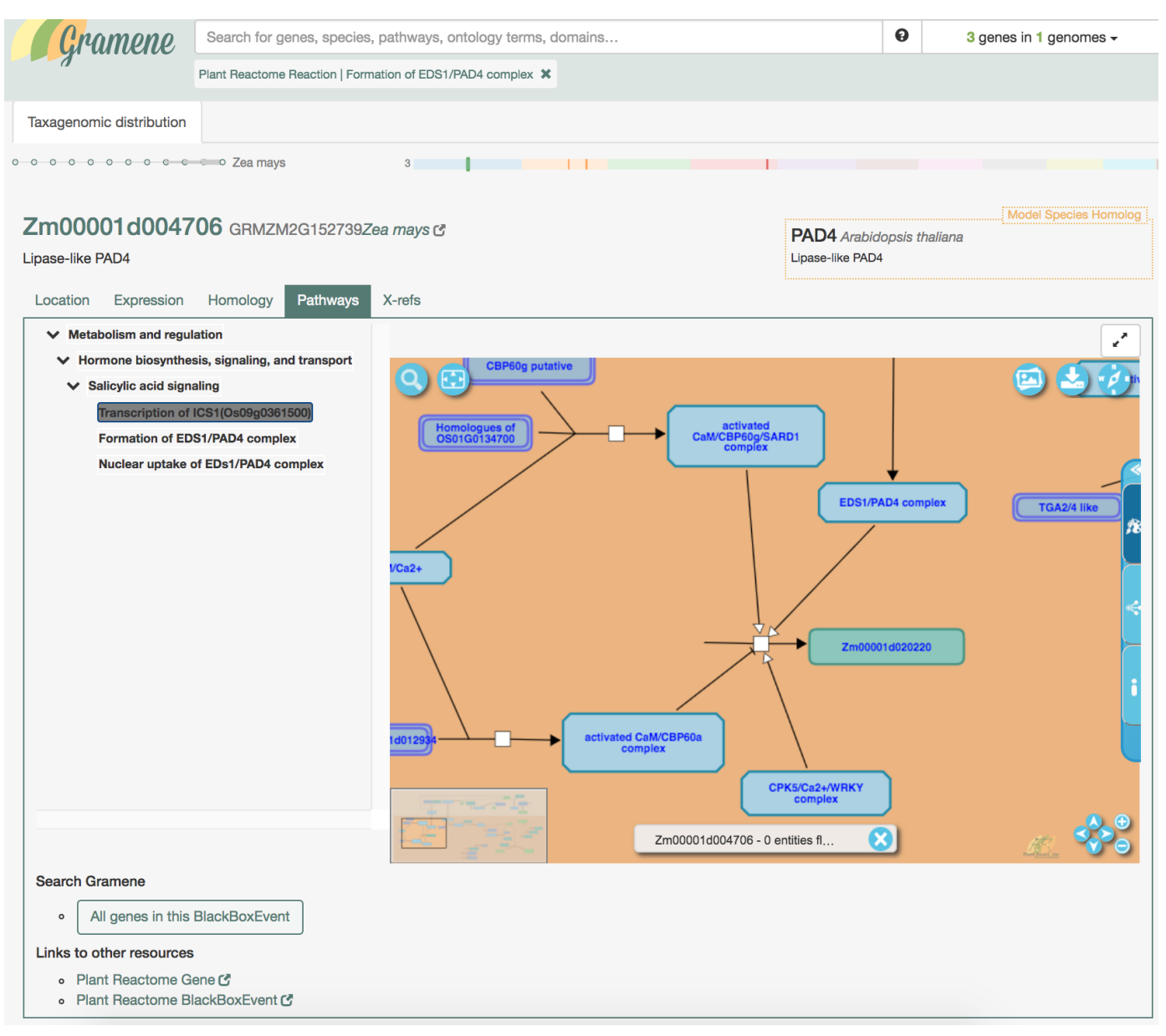
http:www.gramene.org

Summarize results by category, suggest text-based queries, filter by species



Pathways view

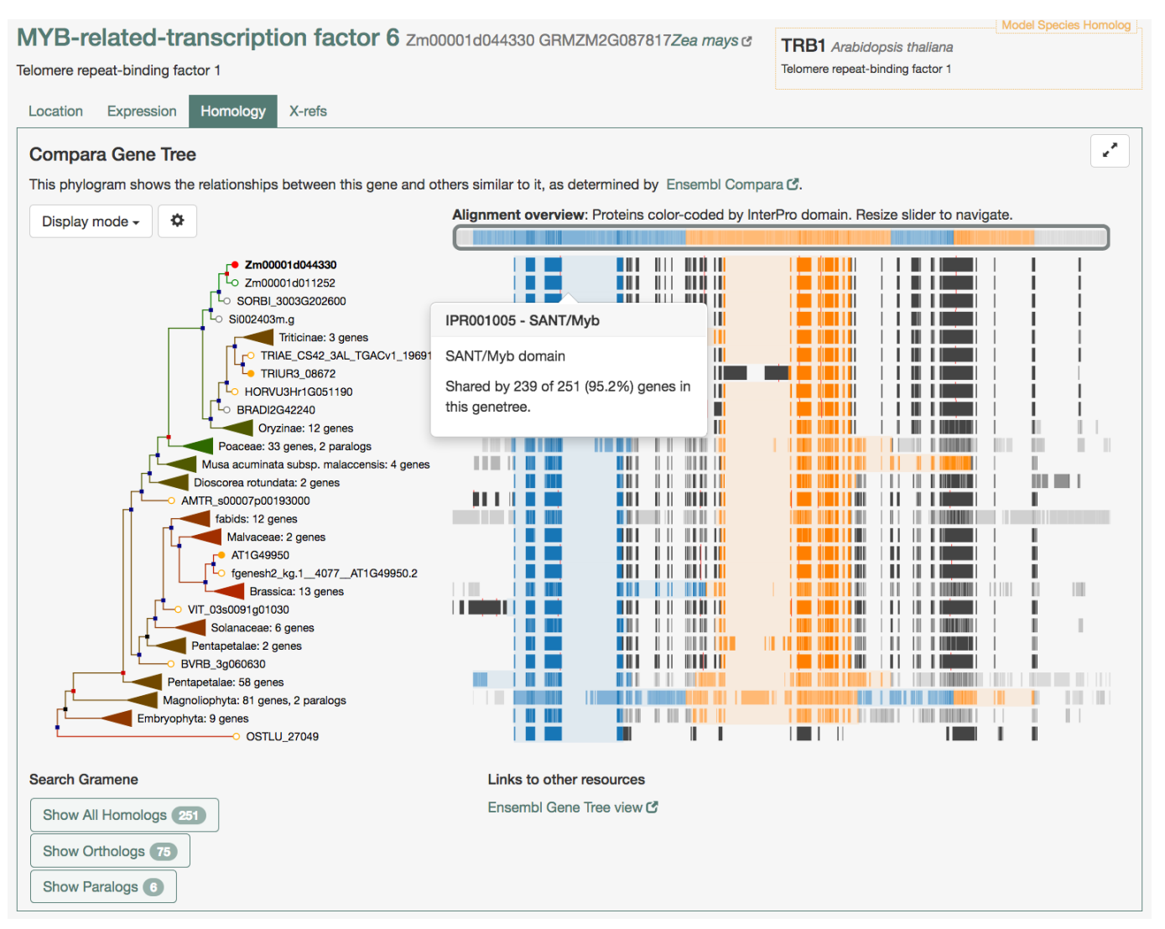
Shows the reactions and pathways that a gene takes part in with links to data downloads in Plant Reactome



Homology view

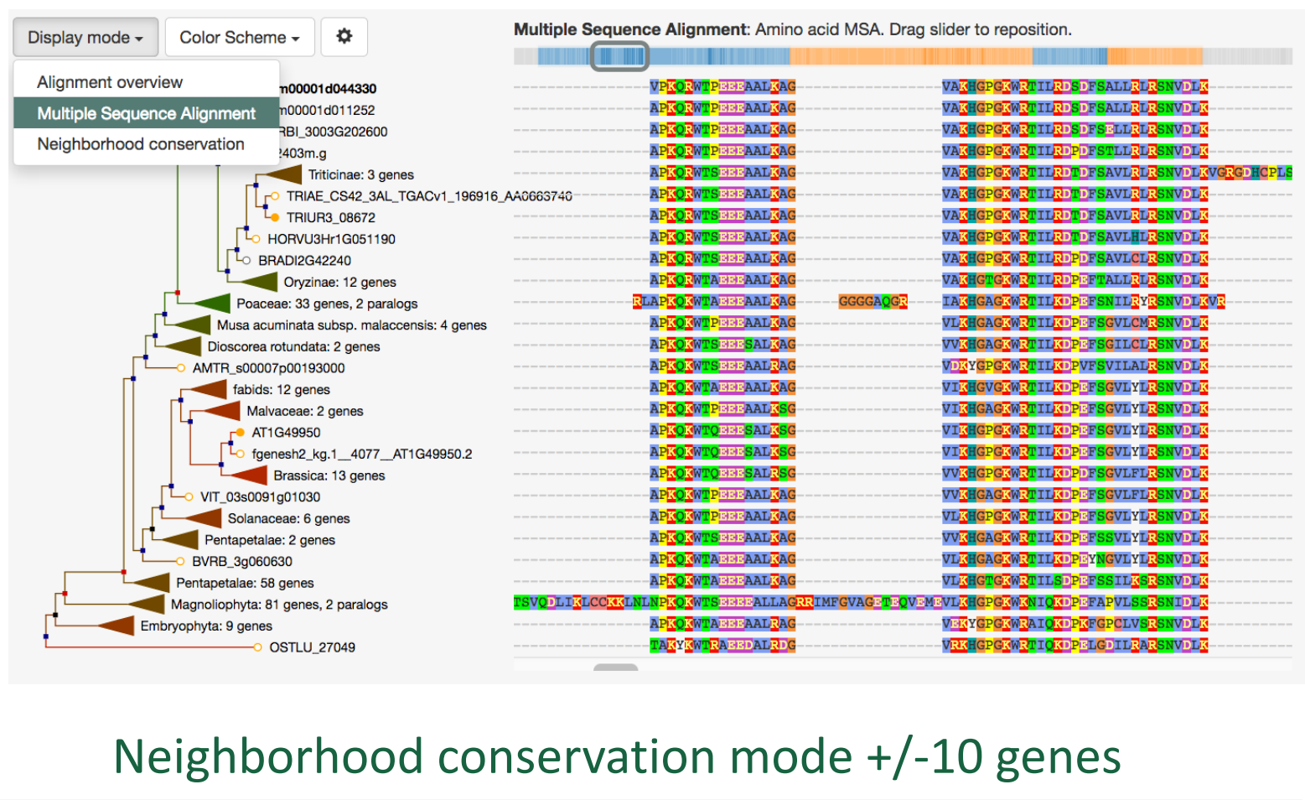
Shows inferred evolutionary histories from Compara, integrated with functional domain information from InterPro.

Overview

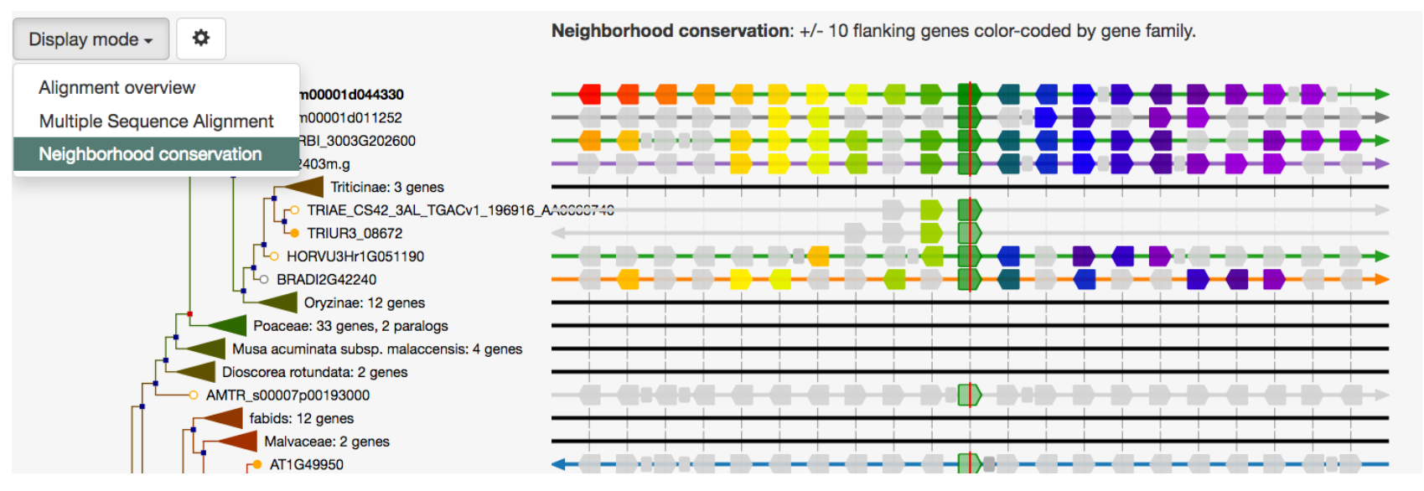


This phylogram shows the relationships between this gene and others similar to it, as determined by Ensembl Compara.

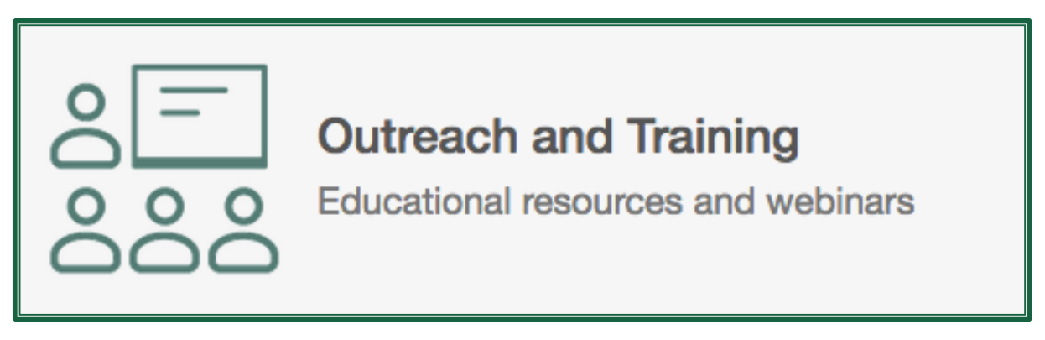
Multi-alignment view for viewing amino acid conservation



Neighborhood conservation mode +/-10 genes



Specific Aim 4: OUTREACH



Outreach and Training

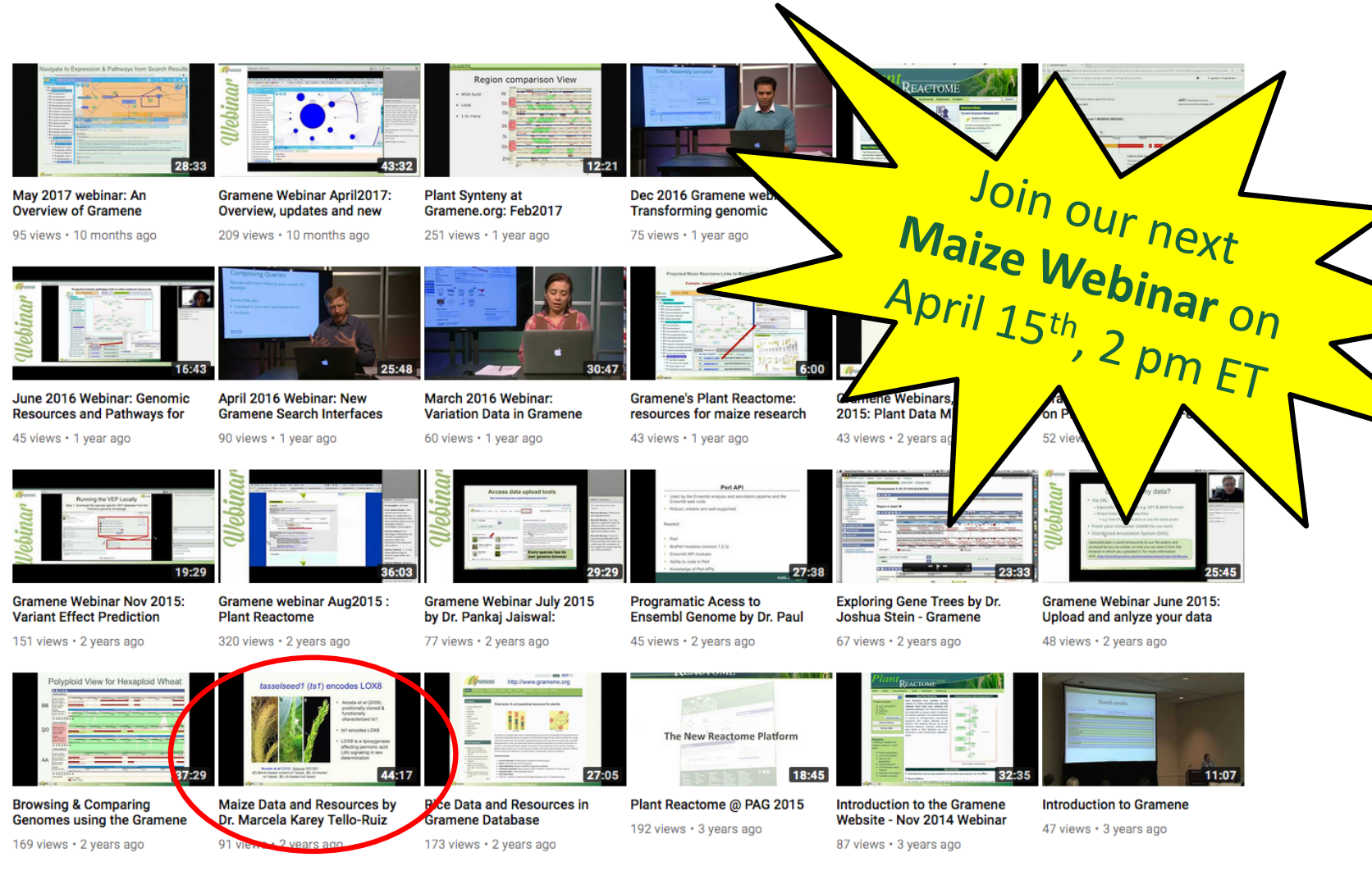
Educational resources and webinars

http://www.gramene.org/outreach

Find here links to:

- Download a quick start Gramene pamphlet and other e-learning materials: ftp://ftp.gramene.org/pub/gramene/outreach
- Subscribe to Gramene's mailing list
- Gramene's YouTube channel to
- Follow Gramene in Social Media:

Videotutorials in Gramene's YouTube channel



Join our next **Maize Webinar** on April 15th, 2 pm ET

Maize-specific outreach activities:

- Coordination on standards & maize reference assembly releases with MaizeGDB
- Maize annotation jamboree in collaboration with the MaizeCODE project
- Coordination of Maize HapMap4 with NSF # & European Variation Archive
- Bioinformatics resources outreach booths

Collaborative meeting between Gramene & MaizeGDB at Cold Spring Harbor, NY

Are there targeted training needs for your project or university?
Are you a scientist looking for short visit stays?

Contact us! **feedback@gramene.org**

Recent publications

- **Jiao *et al*** (2017). Improved maize reference genome with single-molecule technologies. Nature 546(7659):524-527.
- **Bukowski *et al*** (2017). Construction of the third generation Zea mays haplotype map. **Gigascience.**
- **Tello-Ruiz *et al*** (2018). Gramene 2018: unifying comparative genomics and pathway resources for plant research. Nucleic Acids Res. 46(D1):D1181-D1189.
- **Jiao *et al*** (2018) Efficient Identification of Causal Mutations through Sequencing of Bulked F_2 from Two Allelic Bloomless Mutants of *Sorghum bicolor*. Front Plant Sci.8:2267.
- **Stein *et al*** (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. Nat Genet. 50(2):285-296.
- **Naithani *et al*** (2017). Plant Reactome: a resource for plant pathways and comparative analysis. Nucleic Acids Res, 45(D1): p. D1029-D1039.
- **Kersey *et al*** (2018). Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Research* 46:D802–D808.
- **Papatheodorou *et al*** (2018). Expression Atlas: gene and protein expression across multiple studies and organisms. Nucleic Acids Res. 46(D1):D246-D251.