

Mining comparative plant data in Gramene



The project is supported by NSF IOS-1127112 and USDA

Marcela Karey Tello-Ruiz¹, Sharon Wei¹, Andrew Olson¹, Justin Preece², Parul Gupta², Sushma Naithani², Joshua Stein¹, Yinping Jiao¹, Bo Wang¹, Sunita Kumari¹, Young Koung Lee^{1,3}, Lifang Zhang¹, Demitri Muna¹, Dan Bolser⁴, Matthieu Muffato⁴, Peter D'Eustachio⁵, Irene Papatheodorou⁴, Paul J. Kersey³, Pankaj Jaiswal², Doreen Ware^{1,6}

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; ²Oregon State University, Corvallis, OR, USA; ³Wonkwang University, Iksan 54538, Korea; ⁴EMBL-EBI, Hinxton, UK, ⁵NYU School of Medicine, New York, NY, USA, ⁶USDA 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

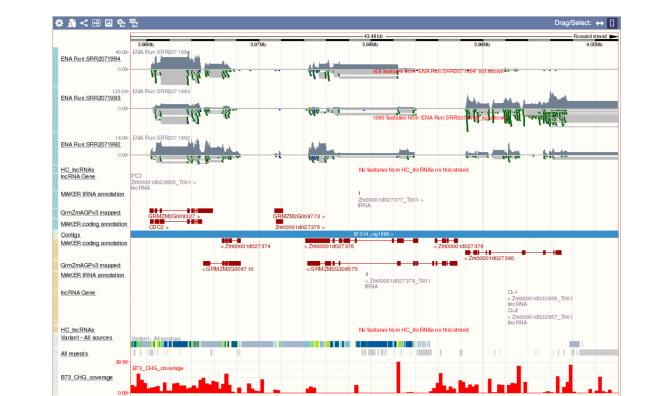
- 5 database releases: build 54 (July 2017); build 55 (Sept 2017); build 56 (Jan 2018); build 56b (Feb 2018); build 57 (May 2018)
- Ensembl software updates V89-92
- 53 plant reference assemblies including:
 - 9 new genomes:
 - <u>Cucumis sativus</u> (cucumber)
 - *Dioscorea rotundata* (white Guinea 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:

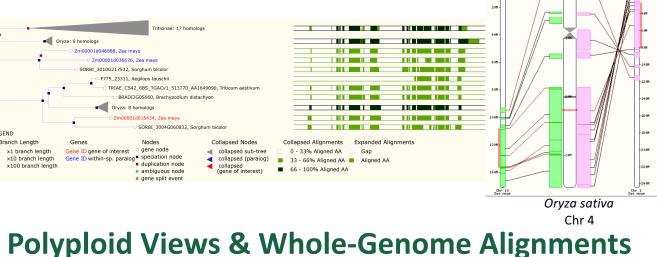
- Sorghum bicolor (sorghum [V3]
- Glycine max (soybean [V3]) Prunus persica (peach [V2])
- Hordeum vulgare (barley [V2])
- Updated gene models:

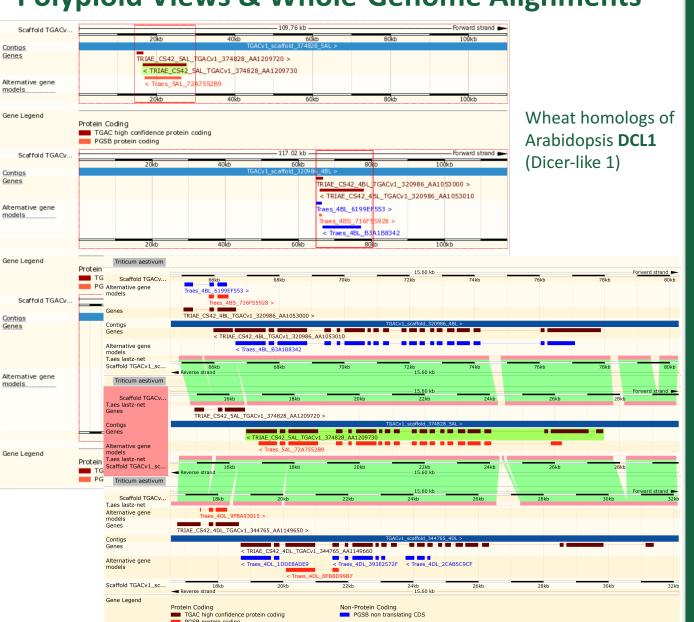
- Zea mays (corn [V4])
- Oryza sativa japonica (rice [IRGSP1]) New polyploid views for wheat
- New gene structural annotations: maize, barley & rice
- New functional annotations (e.g., InterPro domains)
- Updated protein gene trees
- New genetic variation to a total of 230.5 M variants:
- 13M SNPs *Arabidopsis 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)
- Over 1,600 public RNA-Seq studies from Track Hub New gene neighborhood views for local synteny
- New whole-genome alignments (WGAs) for maize vs sorghum, and barley vs rice and Brachypodium
- New ncRNA feature alignments across all plants
- New and updated functional genomics databases (i.e.,
- gene expression arrays probes) New maize transposon annotation (Jiao et al, 2017)

RNA-Seq Data from Public Track Hub Registry









Specific Aim 2: PATHWAYS & EXPRESSION

Auxin signalling (64/75) FDR: 1E0

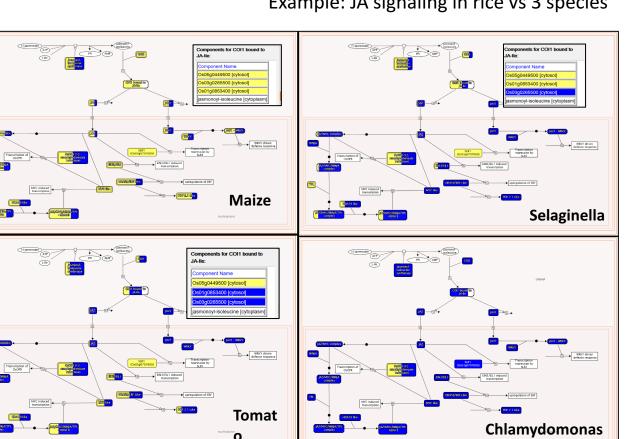
Download pathway

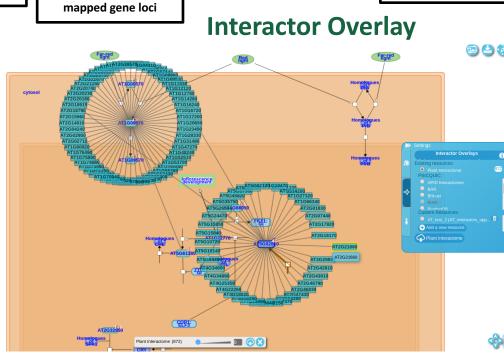
Year Highlights

- 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 780 experiments and >20 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



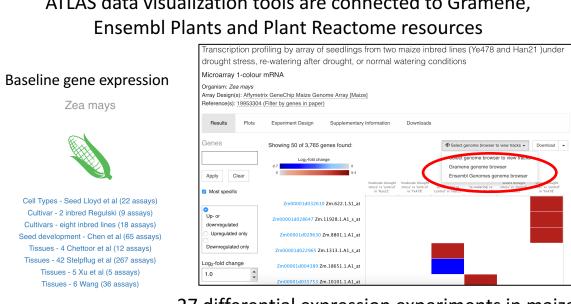


expression from different

http://plantreactome.gramene.org

OMICs data analysis tool

ATLAS data visualization tools are connected to Gramene,



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

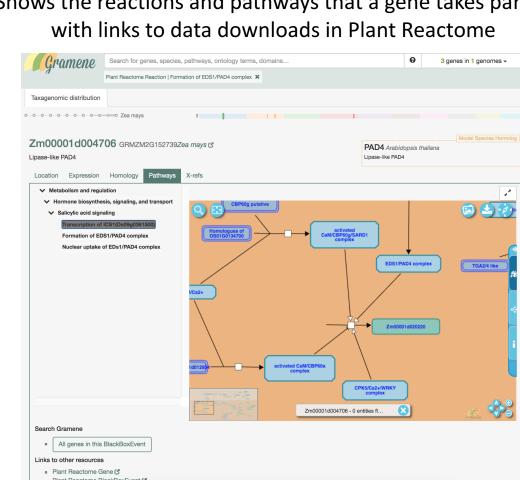
Powerful search interface http:www.gramene.org

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

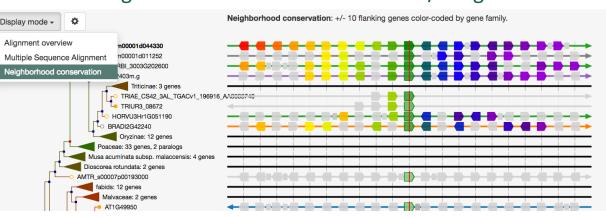
Shows inferred evolutionary histories from Compara, integrated with functional domain information from InterPro.

> Overview Links to other resources Show All Homologs 251

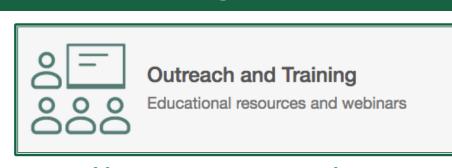
Multi-alignment view for viewing amino acid conservation



Neighborhood conservation mode +/-10 genes



Specific Aim 4: OUTREACH



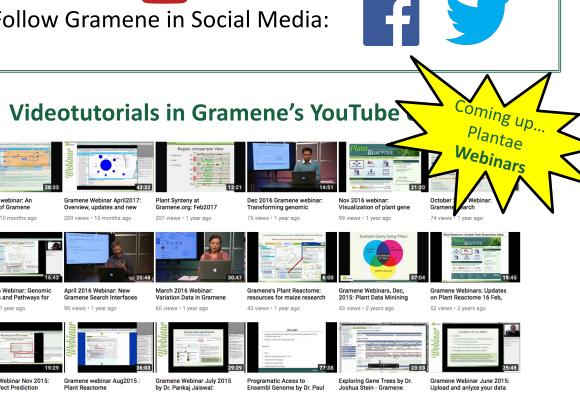
http://www.gramene.org/outreach

Find here links to: Download a quick start Gramene pamphlet and other

e-learning materials: ftp://ftp.gramene.org/pub/gramene/outreach

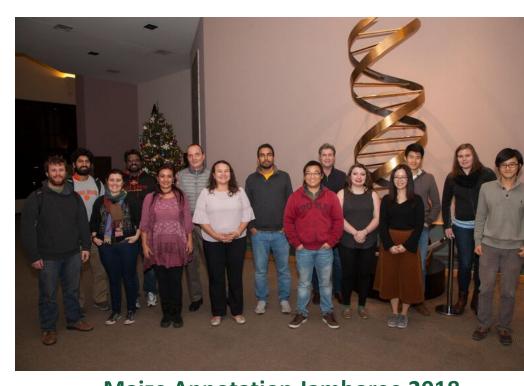
Subscribe to Gramene's mailing list

Gramene's **You** Tube channel to Follow Gramene in Social Media:



Selected outreach activities:

- Gene annotation jamboree in collaboration with the MaizeCODE project. Recruiting interested PUIs
- Coordination on data and metadata formats, standards and best practices with the AgBioData Consortium
- Bioinformatics resources outreach booths with AgBioData Consortium
- User workshops & developers hackathons DNA & bioinformatics summer camps



Maize Annotation Jamboree 2018



Contact us! feedback@gramene.org

Recent publications

- Tello-Ruiz et al (2018). Gramene 2018: unifying comparative genomics and pathway resources for plant research. Nucleic Acids Res. 46(D1):D1181.
- Wang et al (2018). A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing. Genome Res.
- Jiao et al (2018) Efficient Identification of Causal Mutations through Sequencing of Bulked F₂ 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...
- Jiao et al (2017). Improved maize reference genome with singlemolecule technologies. Nature 546(7659):524. Bukowski et al (2017). Construction of the third generation Zea mays
- haplotype map. Gigascience. Naithani et al (2017). Plant Reactome: a resource for plant pathways and comparative analysis. Nucleic Acids Res, 45(D1): p. D1029.
- Kersey et al (2018). Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Research

Papatheodorou et al (2018). Expression Atlas: gene and protein

expression across multiple studies and organisms. Nucleic Acids Res.

46(D1):D246.











