

Gramene: Exploring Function through Comparative Genomics and Network Analysis

Doreen H. Ware, Ph.D.
United States Department of Agriculture ARS
Cold Spring Harbor Laboratory

Gramene II

Pankaj Jaiswal, Crispin Taylor, Lincoln Stein, Paul Kersey, Helen Parkinson, Doreen Ware

June 2012- 2016

Aim 1. Establish reference data for plant genomes and comparative annotation.

- 1.1 Establish an integrated reference genome resource.
- 1.2 Visualization of Epigenomic, Genotypic, and Functionally Phenotypic Diversity.
- 1.3 Representation of the pan-genome.
- 1.4 Comparative genomics: analysis of plant genomes and visualization informed by evolutionary histories
- 1.5 Stewardship of the maize B73 genome and assembly for one update



Aim 2. Establish integrated gene network analysis for plants (**maize, rice, arabidopsis**)

- 2.1 Functional annotation of core model organisms for reconstruction of metabolic and signaling pathway networks.
- 2.2 Meta data analysis and annotation of the gene expression datasets.



Aim 3. Integrate new and existing visualization/analysis tools for exploring emerging genomic data

- 3.1 Build a high-capacity data warehouse that promotes flexible implementation of software modules.
- 3.2 Improve visual interfaces and site navigability.
- 3.3 Improve community data access.



Aim 4. Transform the community through communication and training opportunities

- 4.1 Gramene users (researchers and general public interested in plant biology).
- 4.2 Data annotation, exchange and format standardization workshops (collaborators and experts).
- 4.3 Collaboration with the American Society of Plant Biologists (ASPB) publishing group.
- 4.4 Annotation jamborees (young investigators and students).
- 4.5 Public Lecture series at Cold Spring Harbor Laboratory.
- 4.6 Summer internships for undergraduate students and faculty.



Gramene II

- Aim 1. Establish reference data for plant genomes and comparative annotation. **(Ensembl)**
- 1.1 Establish an integrated reference genome resource.
 - 1.2 Visualization of Epigenomic, Genotypic, and Functionally Phenotypic Diversity.
 - 1.3 Representation of the pan-genome.
 - 1.4 Comparative genomics: analysis of plant genomes and visualization informed by evolutionary histories
- Aim 2. Establish integrated gene network analysis for plants (**arabidopsis, maize, rice,)**
- 2.1 Functional annotation of core model organisms for reconstruction of metabolic and signaling pathway networks. **(Reactome)**
 - 2.2 Meta data analysis and annotation of the gene expression datasets. **(ATLAS)**
- Aim 3. Integrate new and existing visualization/analysis tools for exploring genomic data
- 3.3 Improve community data access. **(Biomart)**
- Aim 4. Transform the community through communication and training opportunities
- 4.3 Collaboration with the American Society of Plant Biologists (ASPB) publishing group.

The logo for Ensembl, featuring a stylized 'e!' in blue followed by the word 'Ensembl' in a bold, blue, sans-serif font.The logo for ATLAS, with the word 'ATLAS' in a bold, red, sans-serif font.

Gramene/Ensembl Adds Value

Standard Analyses and Visualization

Accessioned
Genomes +
Community
Annotation

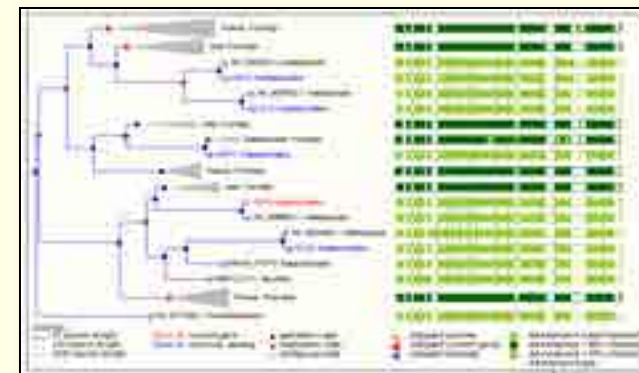
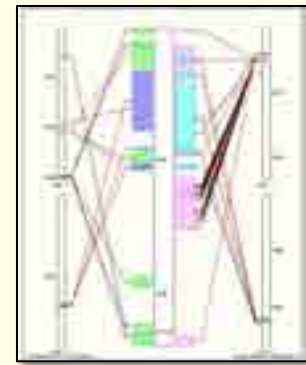
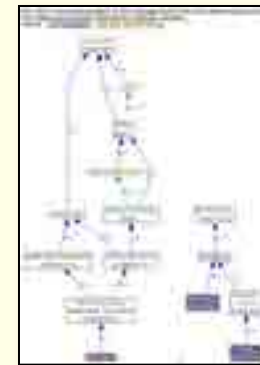
Primary sources:

- GenBank
- JGI
- Sol Genomics
- Genoscope
- TAIR
- MSU
- Etc.

Analysis Pipelines

- RepeatMasking
- Baseline gene prediction
- dbXref
- InterPro domain
- Gene Ontology (GO)
- Variant Effect Prediction
- Whole Genome Alignment
- Phylogenetic Gene Trees
- Synteny mapping

Browser Displays

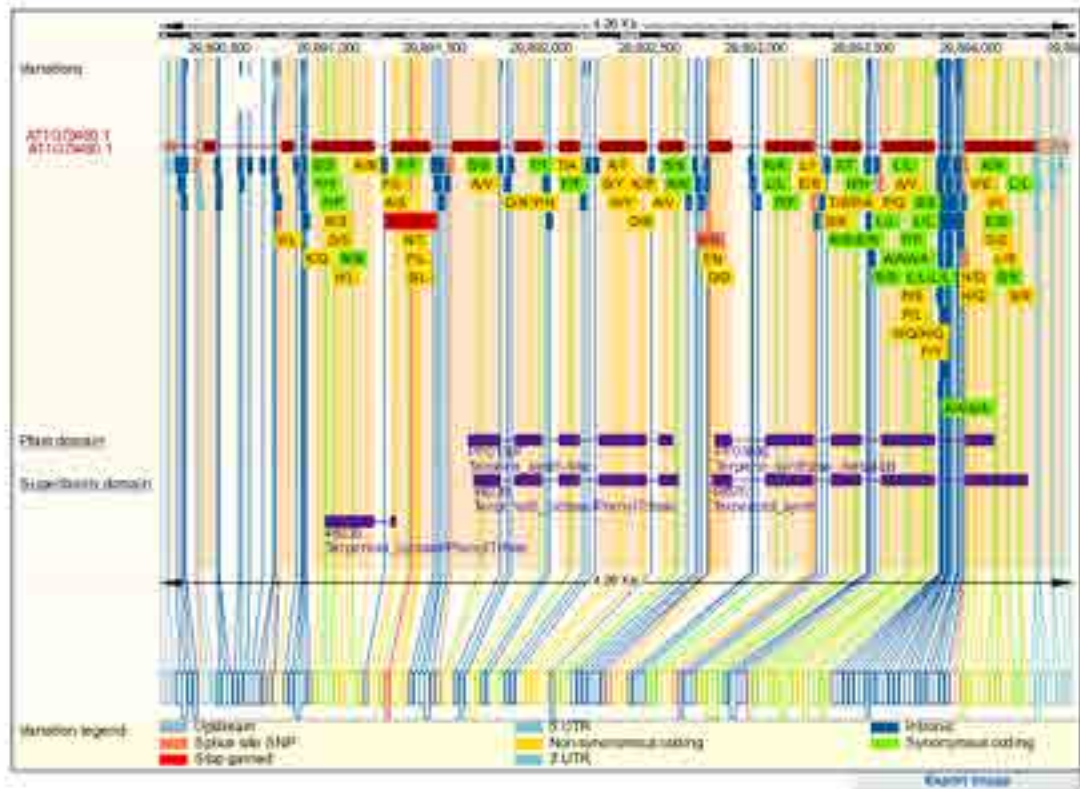


Gene Variation Image

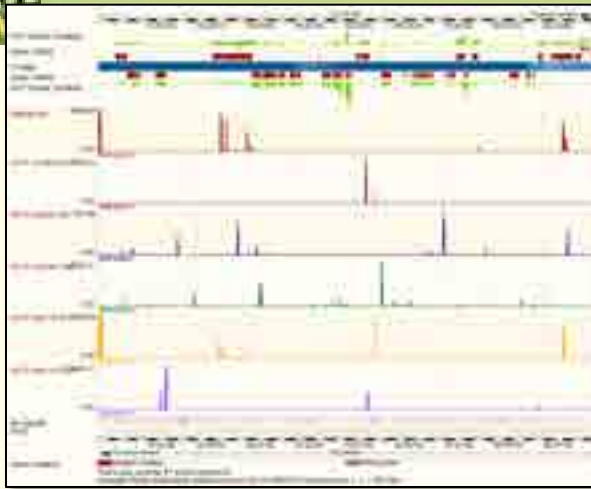
Shows functional consequence relative to gene and protein domain structure

Gene-based displays

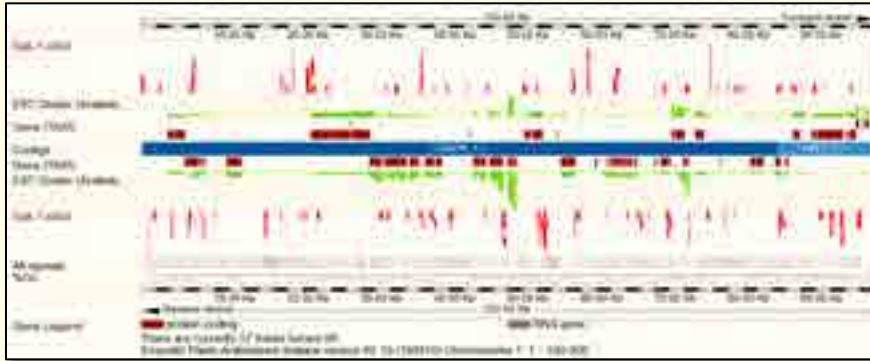
- Gene summary
- Splice variants (1)
- Supporting evidence
- Sequence
- External references
- Regulation
- Expression
- Plant Compara
 - Genomic alignments
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (56)
 - Paralogues (2)
 - Protein families
- Pan-taxonomic Compara
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (15)
 - Paralogues (2)
 - Protein families (1)
- Genetic Variation
 - Variation Table
 - Variation Image**
- External Data
- ID History
 - Gene history



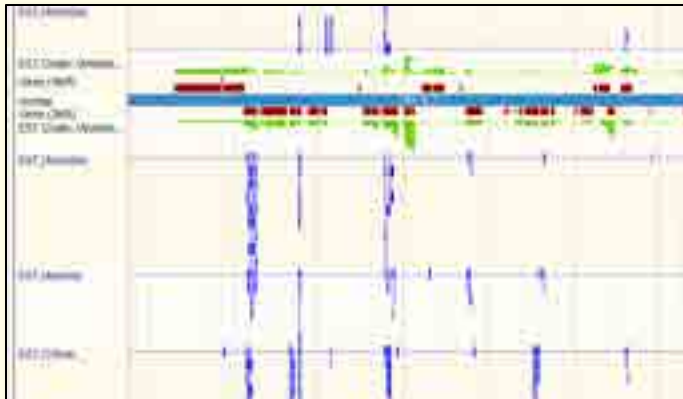
Add Custom Tracks



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



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



- EST alignments from non-model plants
- DAS: Distributed Annotation system
- Protocol for sharing 3rd party data
- DAS Registry

Multi-species View

A. lyrata

Arabidopsis

Grapevine

Arabidopsis

Poplar

GRAMENE

Arabidopsis thaliana (TAIR10)

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
 - Alignments (image) (9)
 - Alignments (text) (9)
 - Multi-species view (8) ↗
 - Synteny (3)
- Genetic Variation
 - Resequencing (18)
 - Linkage Data
- Markers
- Other genome browsers
 - TAIR
 - Phytozome

Configure this page

Manage your data

Export data

Bookmark this page

Select species

Select species

Tip

Click on the plus and minus buttons to select or deselect options.

Selected options can be reordered by dragging to a different position in the list.

Selected species

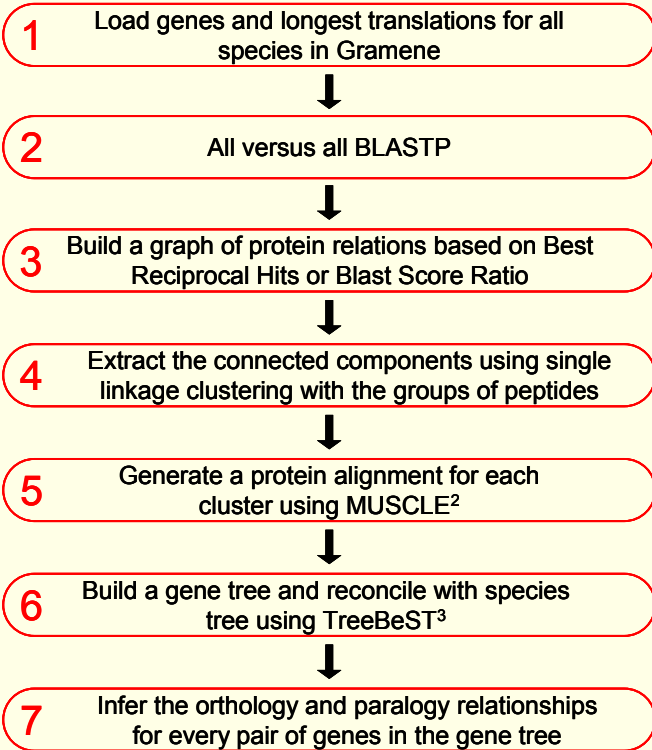
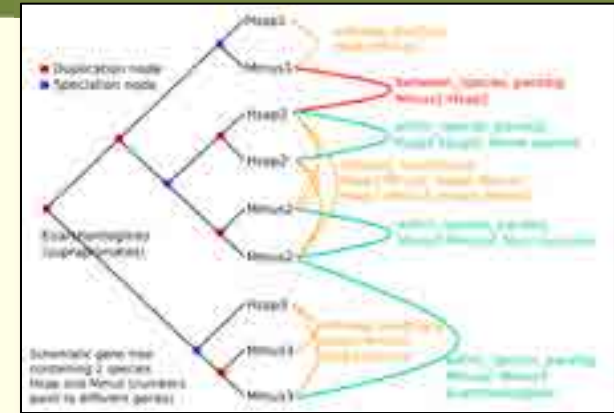
- Arabidopsis lyrata
- Populus trichocarpa



Compara Gene Trees

Reconstructing evolutionary histories

- Gene Trees for 23 plants plus human, Ciona, fly, worm, & yeast
- Infers orthologs and paralogs by reconciling gene tree with input species tree
- Taxonomic dating



Taxon: Poaceae

Gene Count: 10

Branch Length: 0.120504

Bootstrap: 100

Type: Duplications (percentage): 3.697

Image: expand this sub-tree

Image: expand all sub-trees

Image: collapse other nodes

Comparison: Jump to Multi-species view

View Sub-tree: Alignment: FASTA

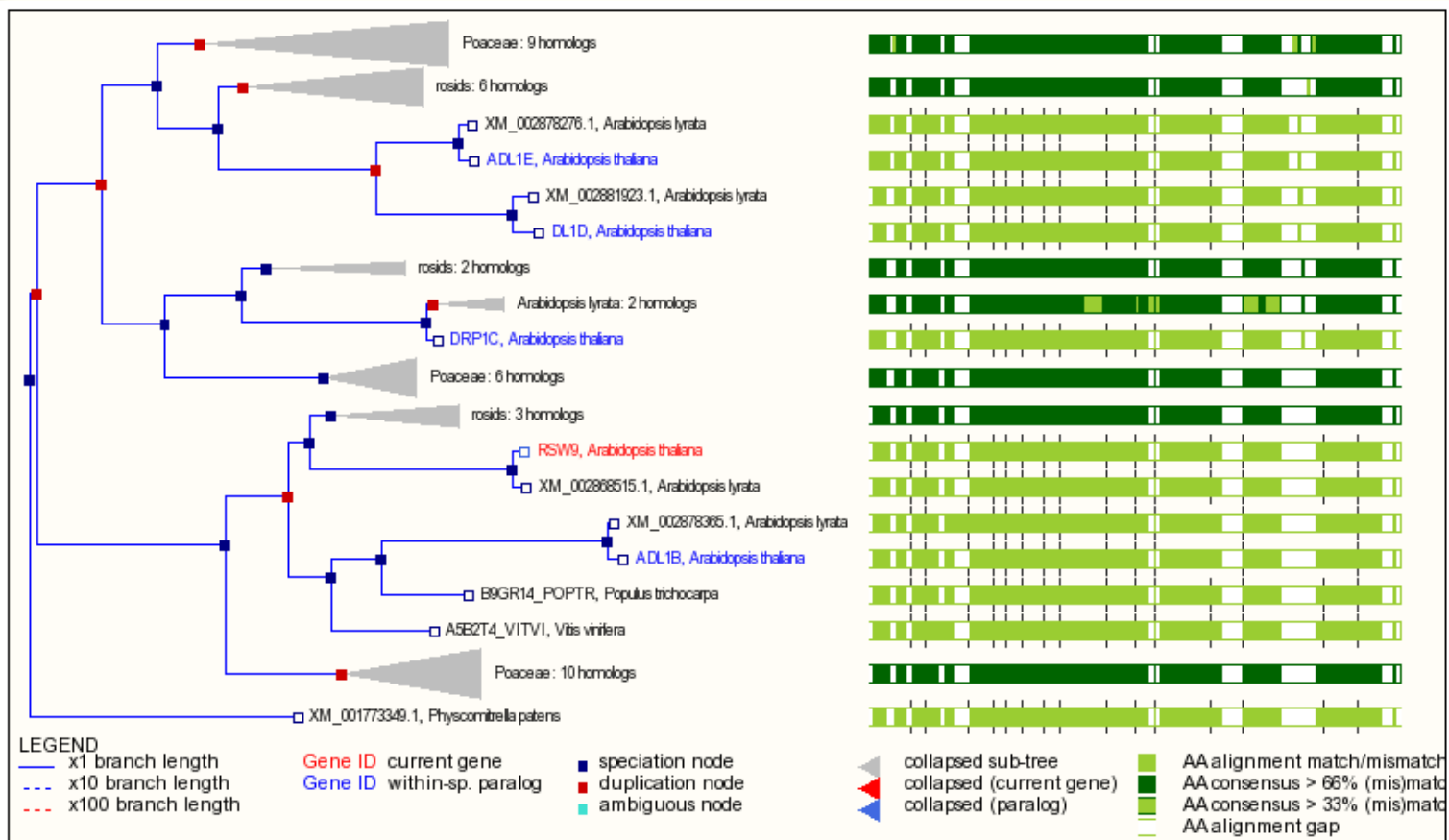
View Sub-tree: Tree: New Navigation

View Sub-tree: Expand for Sub-view

Tree Viewer

Gene-based displays

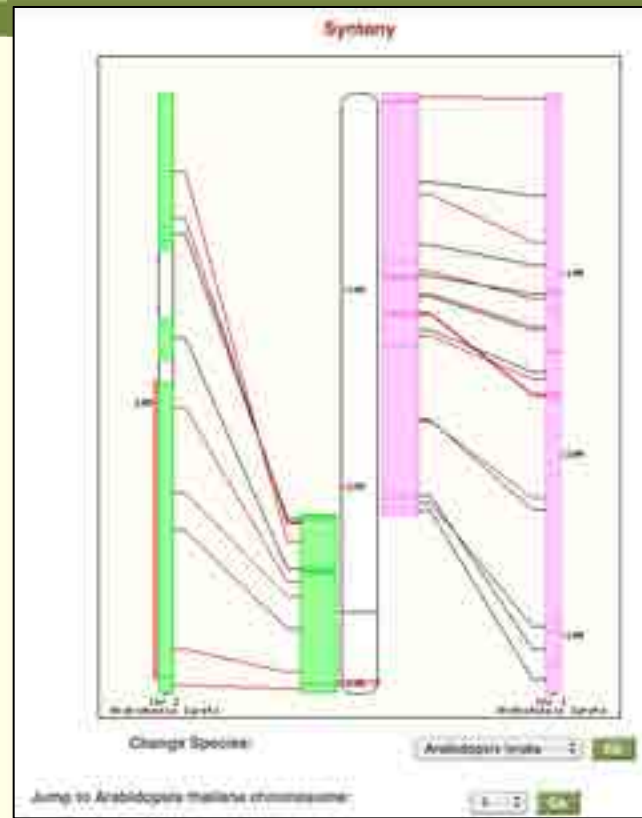
- Gene summary
- Splice variants (3)
- Supporting evidence
- Sequence
- External references
- Regulation
- Expression
- Plant Compara
 - Genomic alignments
 - Gene Tree (image)**
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (15)
 - Paralogues (4)
 - Protein families
- Pan-taxonomic Compara
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (94)
 - Paralogues (10)
 - Protein families (1)
- Genetic Variation
 - Variation Table
 - Variation Image
 - External Data
- ID History
 - Gene history



■ Speciation node = ortholog
 ■ Duplication node = paralog

Synteny View

- Arabidopsis vs *A. lyrata*,
Navigate to other genome
- Ortholog browser
- Link to multi-species view

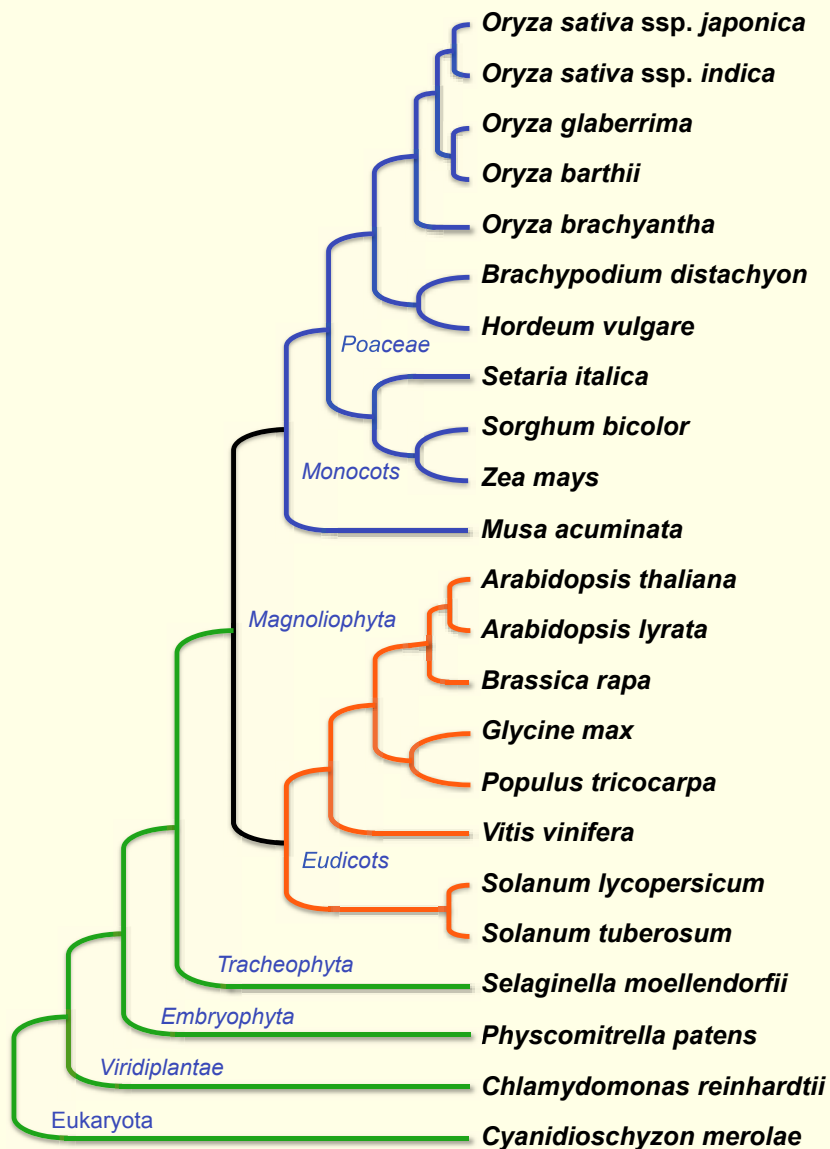


Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
 - Alignments (image) (9)
 - Alignments (text) (9)
 - Multi-species view (8)
 - Synteny (3)**
- Genetic Variation
 - Resequencing (18)
 - Linkage Data
- Markers
- Other genome browsers
 - TAIR
 - Phytozome

← 15 upstream genes	Navigate homology	15 downstream genes →
Arabidopsis thaliana Query		
SP... A0M54 (AT1G7940) 1,209,611-1,209,630	→	Arabidopsis lyrata homology X0_00081728.1 (lyrata_202914.1) 2,302,112-2,302,029
E8B115 (AT1G7900) 1,209,633-1,209,643	→	X0_00081184.1 2,302,012-2,302,080
ATGCT5 (AT1G7900) 1,209,632-1,209,632	→	X0_00081184.1 (lyrata_202914.1) 2,302,012-2,302,080
C2P79C (AT1G7900) 1,209,634-1,209,634	→	X0_00081184.1 2,302,014-2,302,014
AT1G7900 (AT1G7900) 1,209,631-1,209,631	→	X0_00081184.1 (lyrata_202914.1) 2,302,012-2,302,080
AT1G7900 (AT1G7900) 1,209,632-1,209,632	→	X0_00081184.1 (lyrata_202914.1) 2,302,012-2,302,080
AT1G7900 (AT1G7900) 1,209,633-1,209,633	→	X0_00081184.1 (lyrata_202914.1) 2,302,013-2,302,080
ATGCT5 (AT1G7900) 1,209,632-1,209,632	→	X0_00081184.1 (lyrata_202914.1) 2,302,012-2,302,080
A0GCT5 (AT1G7940) 1,209,632-1,209,632	→	X0_00081728.1 (lyrata_202914.1) 2,302,112-2,302,029
AT1G7900 (AT1G7900) 1,209,633-1,209,633	→	X0_00081184.1 (lyrata_202914.1) 2,302,013-2,302,080
AT1G7900 (AT1G7900) 1,209,634-1,209,634	→	X0_00081184.1 (lyrata_202914.1) 2,302,014-2,302,080

23 Complete Genomes



- Crops
- Models
- Phylogenetic significance

Gramene II


- Aim 1. Establish reference data for plant genomes and comparative annotation. (**Ensembl**)
- 1.1 Establish an integrated reference genome resource.
 - 1.2 Visualization of Epigenomic, Genotypic, and Functionally Phenotypic Diversity.
 - 1.3 Representation of the pan-genome.
 - 1.4 Comparative genomics: analysis of plant genomes and visualization informed by evolutionary histories
- Aim 2. Establish integrated gene network analysis for plants (**arabidopsis, maize, rice,**)
- 2.1 Functional annotation of core model organisms for reconstruction of metabolic and signaling pathway networks. (**Reactome**)
 - 2.2 Meta data analysis and annotation of the gene expression datasets. (**ATLAS**)
- Aim 3. Integrate new and existing visualization/analysis tools for exploring genomic data
- 3.3 Improve community data access. (**Biomart**)
- Aim 4. Transform the community through communication and training opportunities
- 4.3 Collaboration with the American Society of Plant Biologists (ASPB) publishing group.

The logo for Ensembl, featuring a stylized 'e!' in blue followed by the word 'Ensembl' in a blue sans-serif font.The logo for ATLAS, with the word 'ATLAS' in a bold, red, sans-serif font.

What is Reactome

- A curated open source and open access pathway database (www.reactome.org)
- >1400 pathways on human metabolism, signaling, and gene regulation etc.
- Human is the Reference species with extension to other species based on gene orthology and inferred reactions

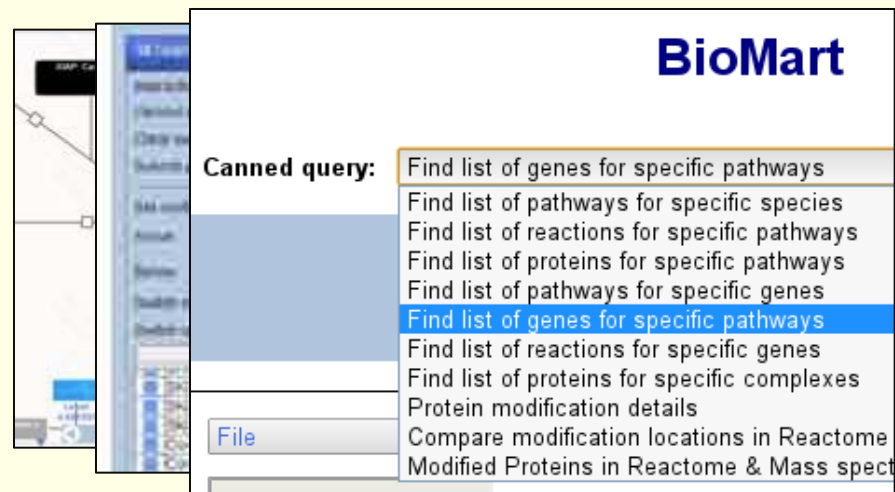


- Data sources
 - RiceCyc 3.1 
 - Rice projections from Reactome (*H. sapiens*)
- “Cyc” Conversion process
 - Exported rice data from RiceCyc 3.1 in BioPAX format
 - Ran “Pathway Exchange” data conversion utility (Java)
 - Performed additional mappings and refinements
 - Updated ChEBI references for Reference Molecules
 - UniProt review of the *O. s. japonica* proteome
 - Pathway layout illustration in Reactome Curator Tool

Upcoming developments

- Include additional metabolic and regulatory rice pathways
- Enable more Reactome analysis tools:

- Gene expression overlay
- Interaction analysis
- Enrichment analysis
- Reactome BioMart
- Species comparison



- Bring **Arabidopsis** into the Plant Reactome as a reference species
 - Update the existing AraCyc Reactome data conversion
 - Perform regulatory pathway projections from Reactome (*H. sapiens*)

Acknowledgements



Oregon State University

Pankaj Jaiswal (PI)

Palitha Dharmawardhana (Curation and coordination)

Sushma Naithani (Curation and outreach)

Vindhya Amarasinghe (Curation)

Justin Preece (Software development and database management)

Dylan Beorchia (Undergraduate student curator)

Kindra Amoss (Undergraduate student curator)



Collaborators

Reactome Project

Ontario Institute for Cancer Research

Lincoln Stein (Reactome PI)

Guanming Wu (Software and data management)

Robin Haw (Web servers and outreach)



NYU

Peter D'Eustachio (Curation mentor and outreach)

Cold Spring Harbor Laboratory

Doreen Ware (PI)

Marcela K Monaco (Gramene coordination and outreach)



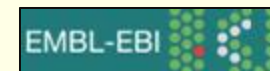
European Bioinformatics Institute

Henning Hermjakob (Reactome and IntAct PI)

David Croft (Software)



Species-wide data



Maize

Carolyn Lawrence (MaizeGDB)

Andrew Hanson (Vitamin-B pathways)

Arabidopsis

Nick Provar (BAR PI)

Hardeep Nahal (Curation)



Gramene II

- Aim 1. Establish reference data for plant genomes and comparative annotation. **(Ensembl)**
- 1.1 Establish an integrated reference genome resource.
 - 1.2 Visualization of Epigenomic, Genotypic, and Functionally Phenotypic Diversity.
 - 1.3 Representation of the pan-genome.
 - 1.4 Comparative genomics: analysis of plant genomes and visualization informed by evolutionary histories
- Aim 2. Establish integrated gene network analysis for plants **(arabidopsis, maize, rice,)**
- 2.1 Functional annotation of core model organisms for reconstruction of metabolic and signaling pathway networks. **(Reactome)**
 - 2.2 Meta data analysis and annotation of the gene expression datasets. **(ATLAS)**
- Aim 3. Integrate new and existing visualization/analysis tools for exploring genomic data
- 3.3 Improve community data access. **(Biomart)**
- Aim 4. Transform the community through communication and training opportunities
- 4.3 Collaboration with the American Society of Plant Biologists (ASPB) publishing group.

The logo for Ensembl, featuring a stylized 'e!' in blue and the word 'Ensembl' in a bold, blue, sans-serif font.The logo for ATLAS, with the word 'ATLAS' in a bold, red, sans-serif font.

Community Access

- Ensembl/Genomes
- Public MySQL server
- BioMart
- DAS
- Galaxy
- iPlant images (late 2013)

Gramene II

- Aim 1. Establish reference data for plant genomes and comparative annotation. (**Ensembl**)
- 1.1 Establish an integrated reference genome resource.
 - 1.2 Visualization of Epigenomic, Genotypic, and Functionally Phenotypic Diversity.
 - 1.3 Representation of the pan-genome.
 - 1.4 Comparative genomics: analysis of plant genomes and visualization informed by evolutionary histories
- Aim 2. Establish integrated gene network analysis for plants (**arabidopsis, maize, rice,**)
- 2.1 Functional annotation of core model organisms for reconstruction of metabolic and signaling pathway networks. (**Reactome**)
 - 2.2 Meta data analysis and annotation of the gene expression datasets. (**ATLAS**)
- Aim 3. Integrate new and existing visualization/analysis tools for exploring genomic data
- 3.3 Improve community data access. (**Biomart**)
- Aim 4. Transform the community through communication and training opportunities
- 4.3 Collaboration with the American Society of Plant Biologists (ASPB) publishing group.
Metabolic pathway and Networks at time of publication

The logo for Ensembl, featuring a blue exclamation mark followed by the word "Ensembl" in blue.The logo for ATLAS, with the word "ATLAS" in red, bold, sans-serif capital letters.

Funding & Partnerships



EMBL-EBI



Gramene is funded by NSF Grant 1127112: *Gramene - Exploring Function through Comparative Genomics and Network Analysis*

Release #36

June 2012

[View this release](#)

News

- **Gramene docks under for Hurricane Sandy. Now back up and getting ready to release build 38**
Last week, Hurricane Sandy battered the mid-Atlantic region causing coastal flooding, knocking down...
- **SSRs at Gramene**
Questions about simple sequence repeats (SSRs) make up a significant portion of the questions we get...
- **Rice QTL genome positions**
A user wrote in asking about the rice QTL positions we have projected onto the *O. sativa japonica*...

[Subscribe to RSS Feed](#)

Have Questions...?

- [Quick Search Help](#)
- Ask questions through [Feedback](#) or [Email](#).
- See [FAQ](#).

[Outreach calendar](#)

[Presentation materials](#)

Gramene is a central, open-access, data resource for comparative genome analysis in the grasses. Our goal is to facilitate the study of cross-species homology relationships using information derived from public projects. Contact us at [gramene@gramene.org](#) or [2027](#).

Explore Gramene

Quick Search



Gramene Workshop

Tuesday, January 15, 2013 @ California Room

3:50 PM - [Introduction to the Gramene Website](#)

Ken Youens-Clark, Cold Spring Harbor Laboratory

4:10 PM - [Plant Reactome](#)

Palitha Dharmawardhana & Justin Preece, Oregon State University

4:40 PM - [Plant Pathways & Gene Expression Analysis in Gramene](#)

Sushma Naithani, Oregon State University

5:10 PM - [Browsing and Comparing Genomes Using the Gramene
Ensembl Browser](#)

Joshua Stein, Cold Spring Harbor Laboratory

5:30 PM - [Analyzing the Barley genome](#)

Paul J. Kersey, EMBL-EBI

Gramene at PAG XXI

Gramene Project Workshop

Tuesday, January 15, 2013

3:50 PM-6:00 PM

Room California

Arabidopsis Community Resources For Comparative and Network Analyses

IAIC workshop, Doreen Ware

Monday, January 14, 2013 @ 1:30 PM

Pacific Salon 6-7 (2nd Floor)

Booth #425

Plant Genome Bioinformatics Outreach Booth

Posters

P7879: Plant Reactome: Metabolic
and Regulatory Networks for Plants

P0942: Gramene Build 36: A
Resource for Comparative Plant
Genomics