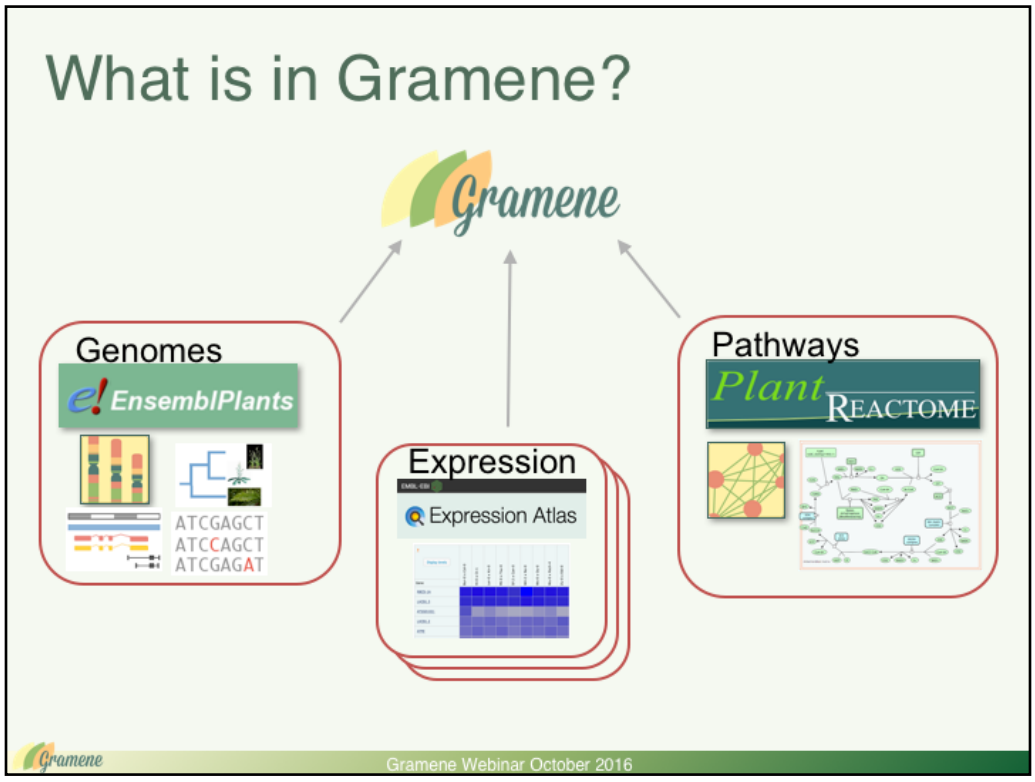
Gramene Search

search.gramene.org **data.gramene.org**

Andrew Olson
Cold Spring Harbor Laboratory
October 4, 2016

Overview

- The role of search in Gramene
- Demo
- Features in development



<Define Gramene>

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

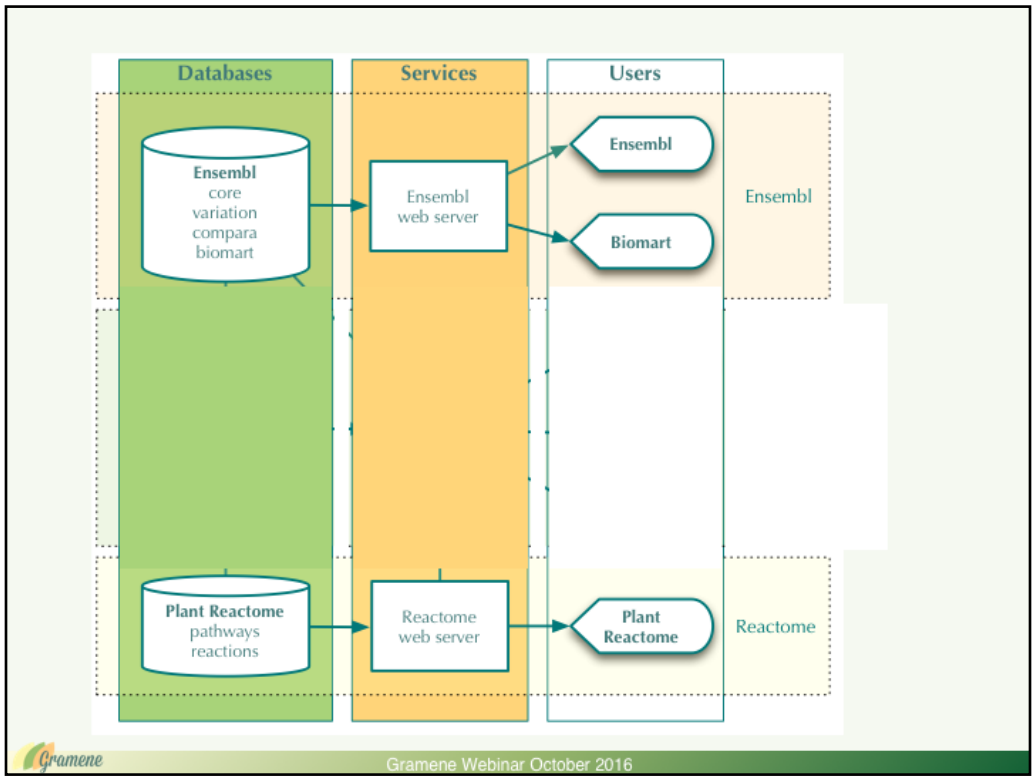
Most significantly:-

- Ensembl for Genomes and Comparative Genomics
- Plant Reactome for pathways and metabolic networks

And also:-

- ATLAS for expression

We also are working with databases like EVA – variation archive – to simplify integration of variation data.

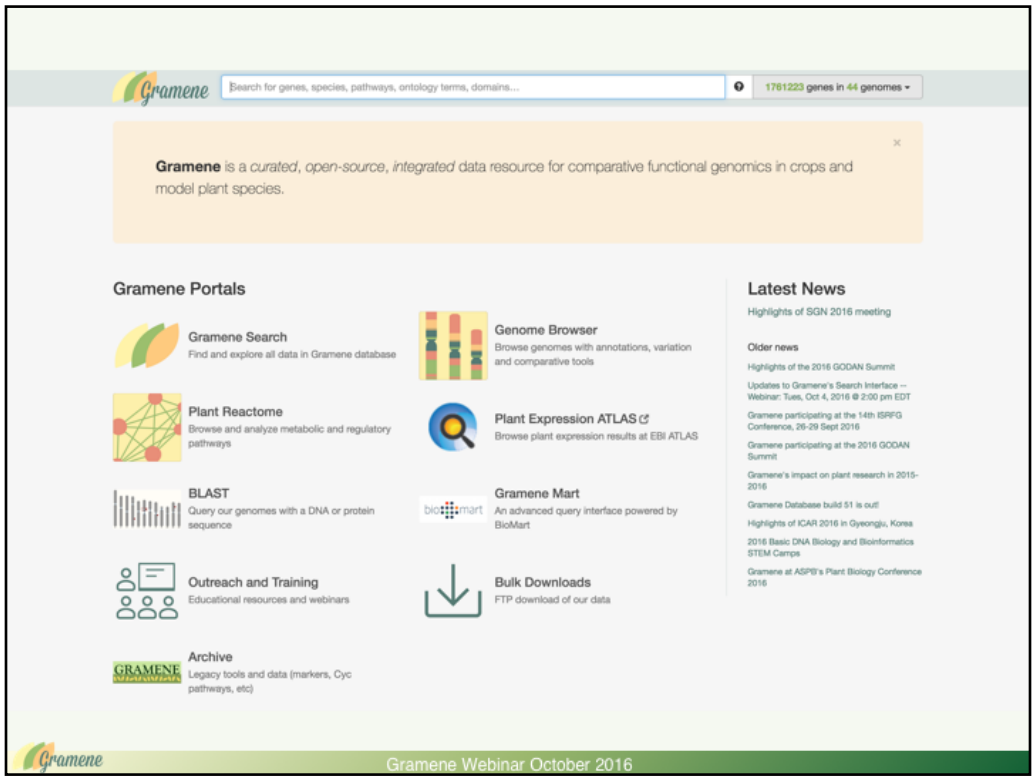


When you think about the architecture of these tools, you see siloes databases, services and user interfaces.

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

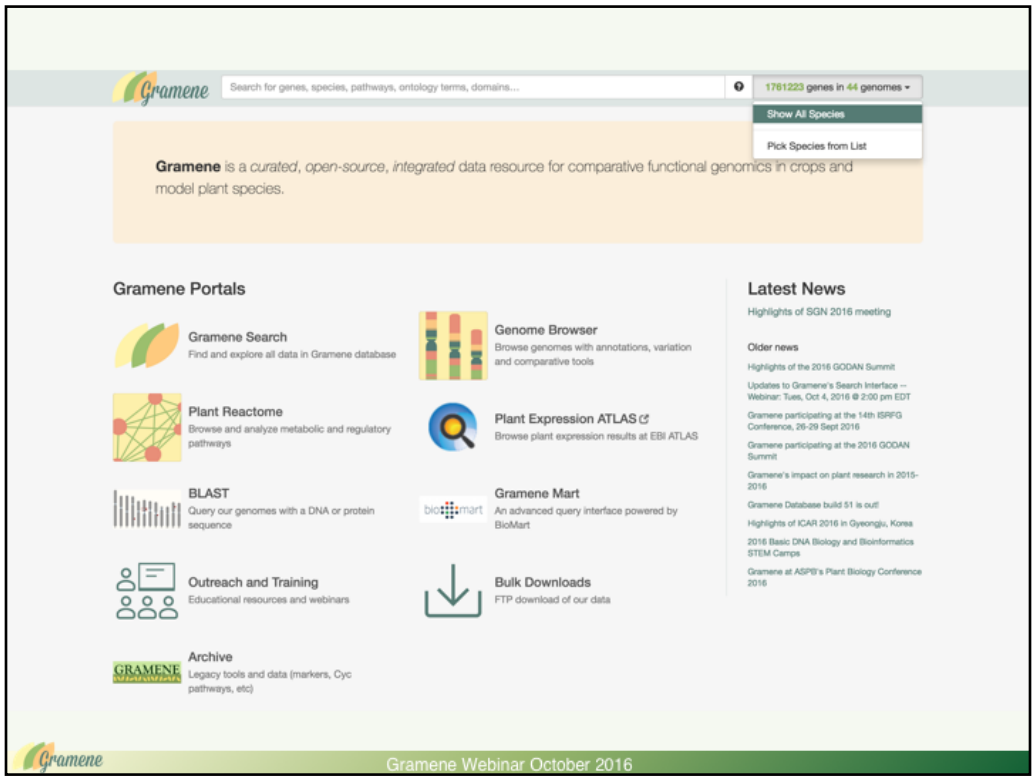
Goals of Search Interface

- Search all Gramene data
- Enable powerful, expressive searches
- Provide:-
  - useful summaries and visualizations
  - links to specialized tools
- REALLY FAST
- demo



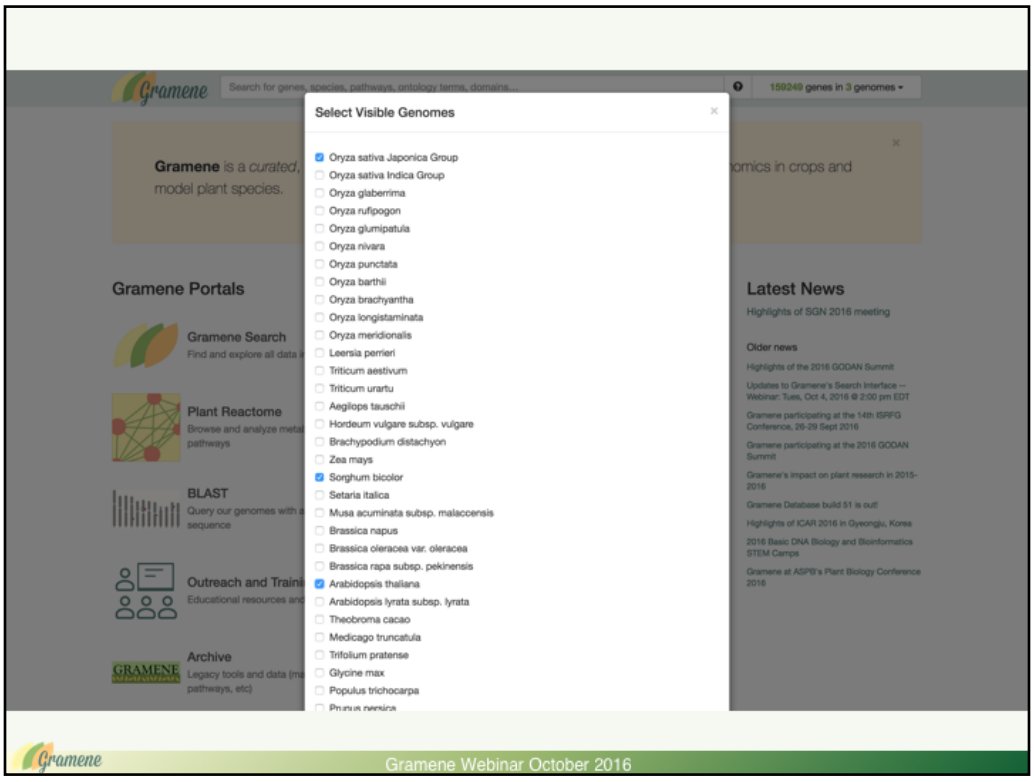
The highlighted region on the top right is the search status (currently all gene in all genomes)

Click on it to select genomes of interest



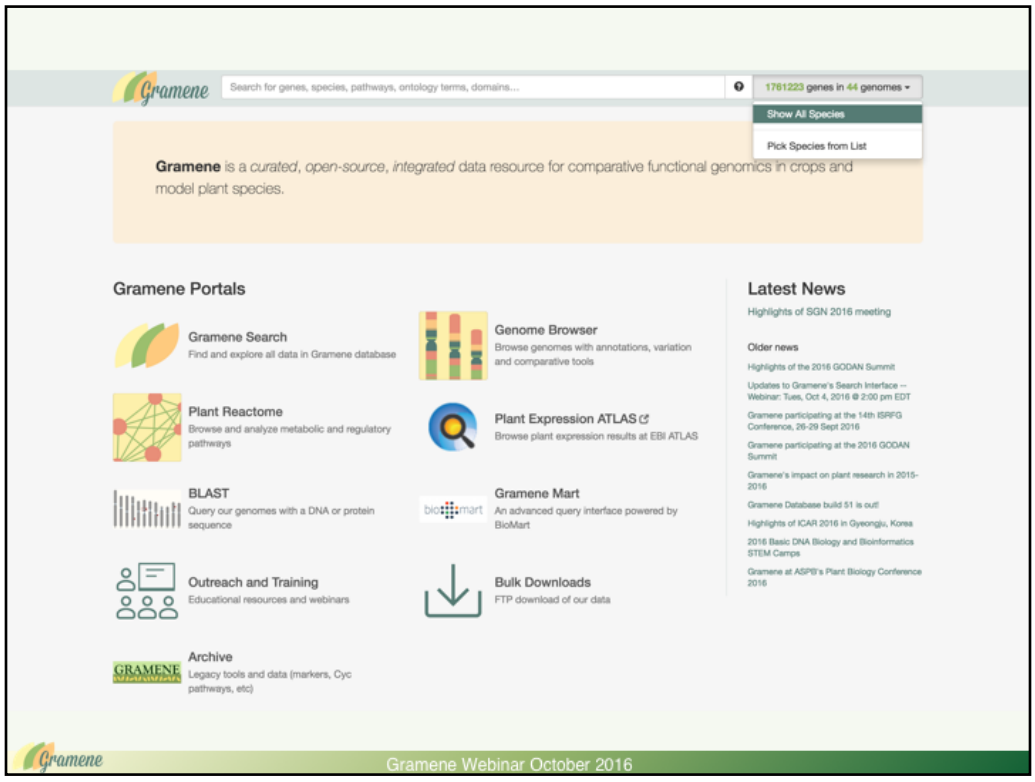
By default you search across all species

Or you can select a subset from a list



The list of species (in the same order as in the species tree)

Future plans include making this list easier to navigate by laying it out on the species tree, color coding clades such as monocots and dicots, and highlighting model organisms and crop species.

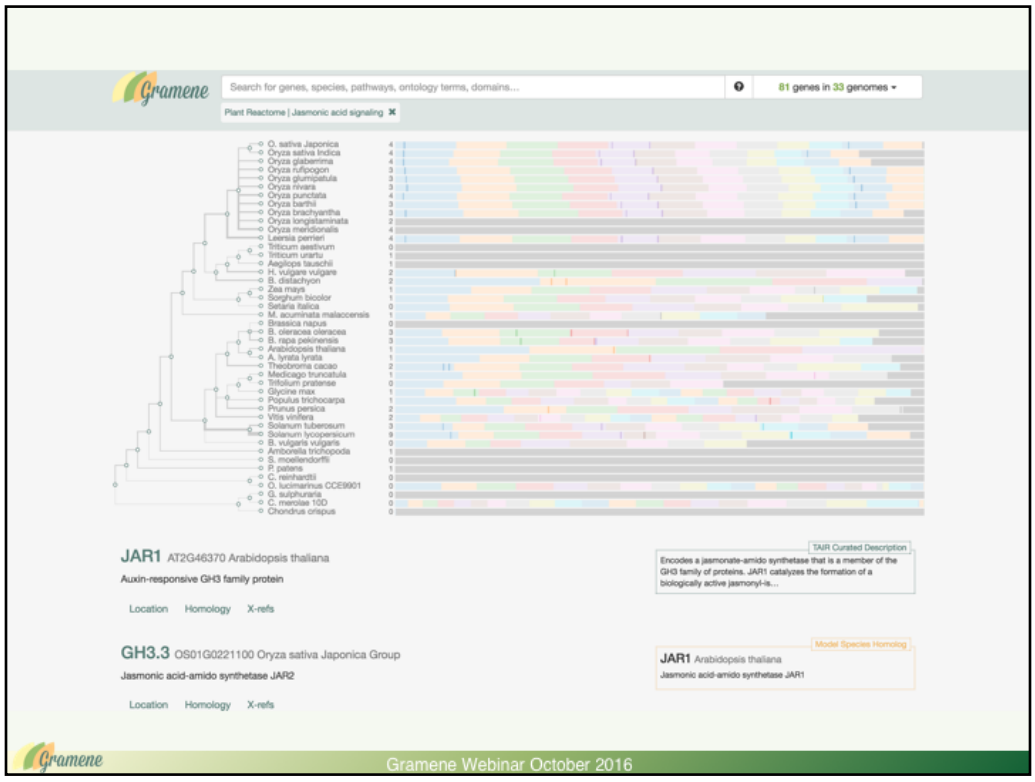


Revert back to all species

The little help icon loads a small guide

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

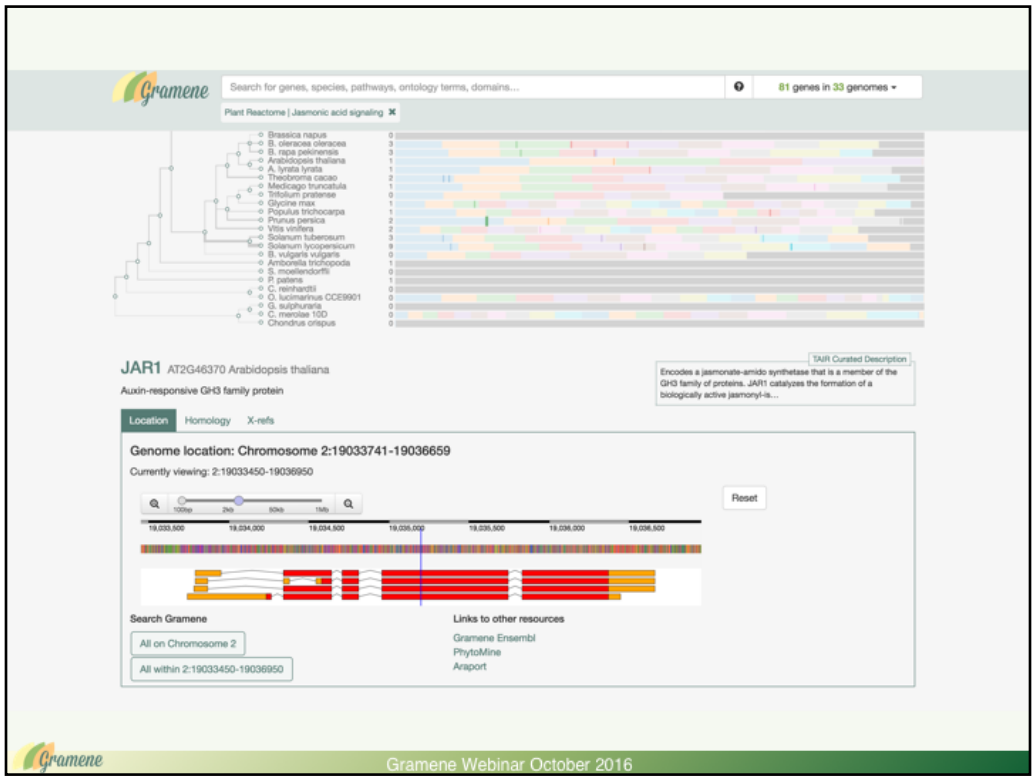
Taxagenomic distribution shows species tree (based on ncbi taxonomy)

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

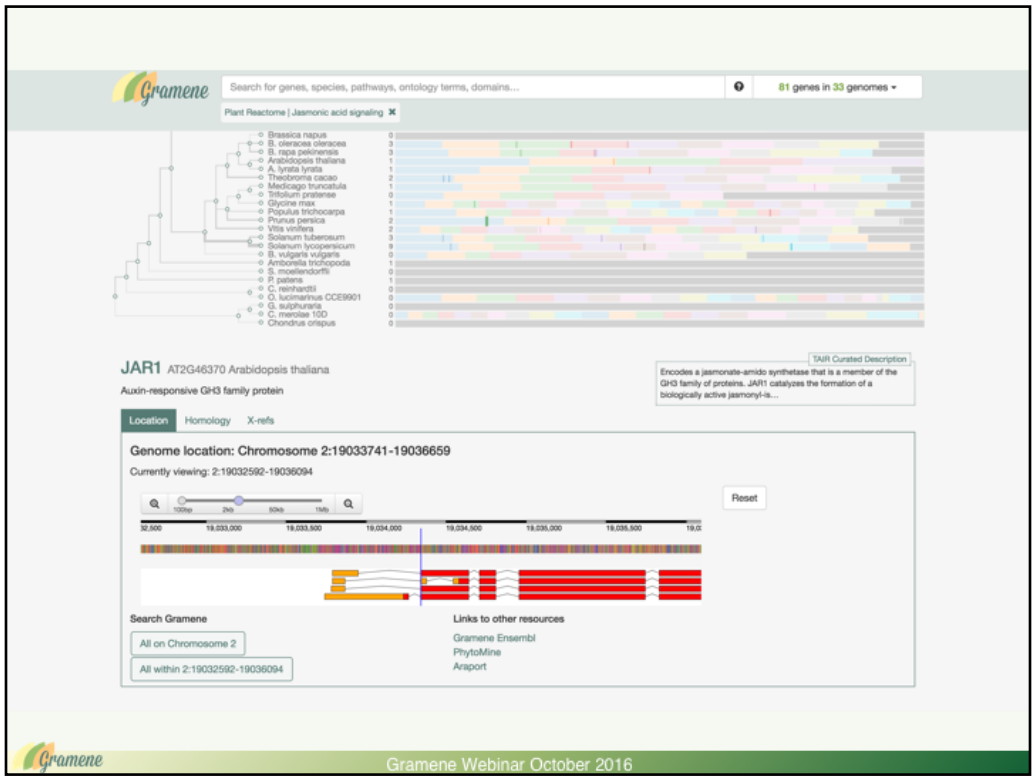
Below is a list view – arabidopsis results appear first because they tend to be best annotated

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

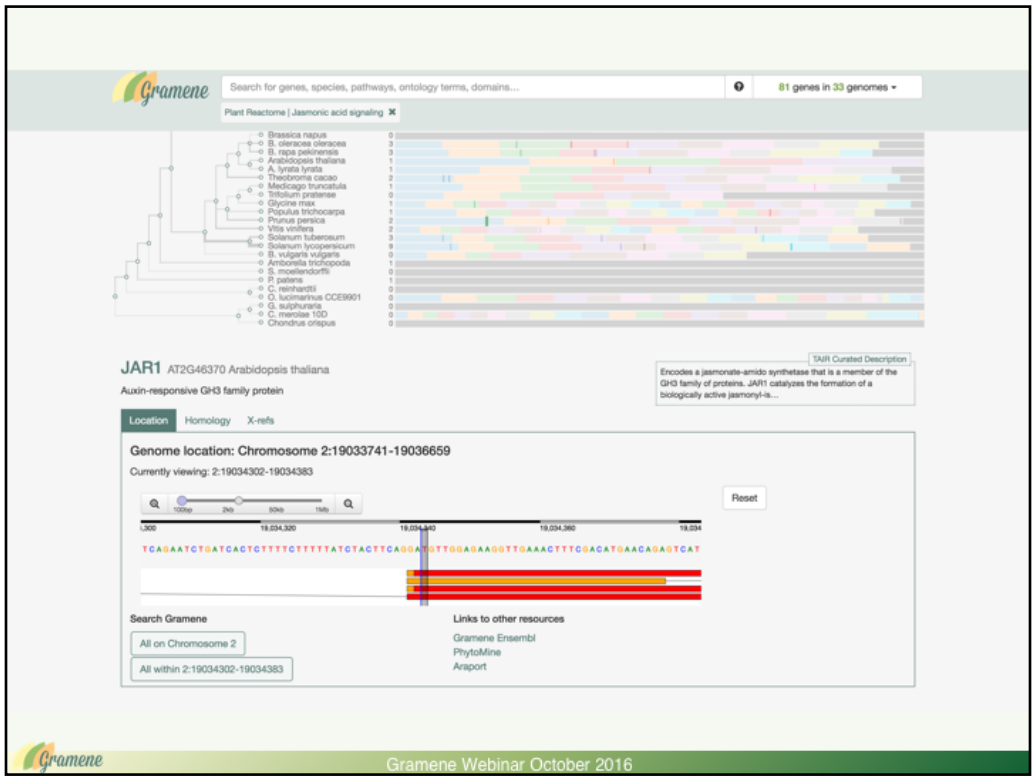


The location view shows a lightweight genome browser that lets you see the annotated gene structure.

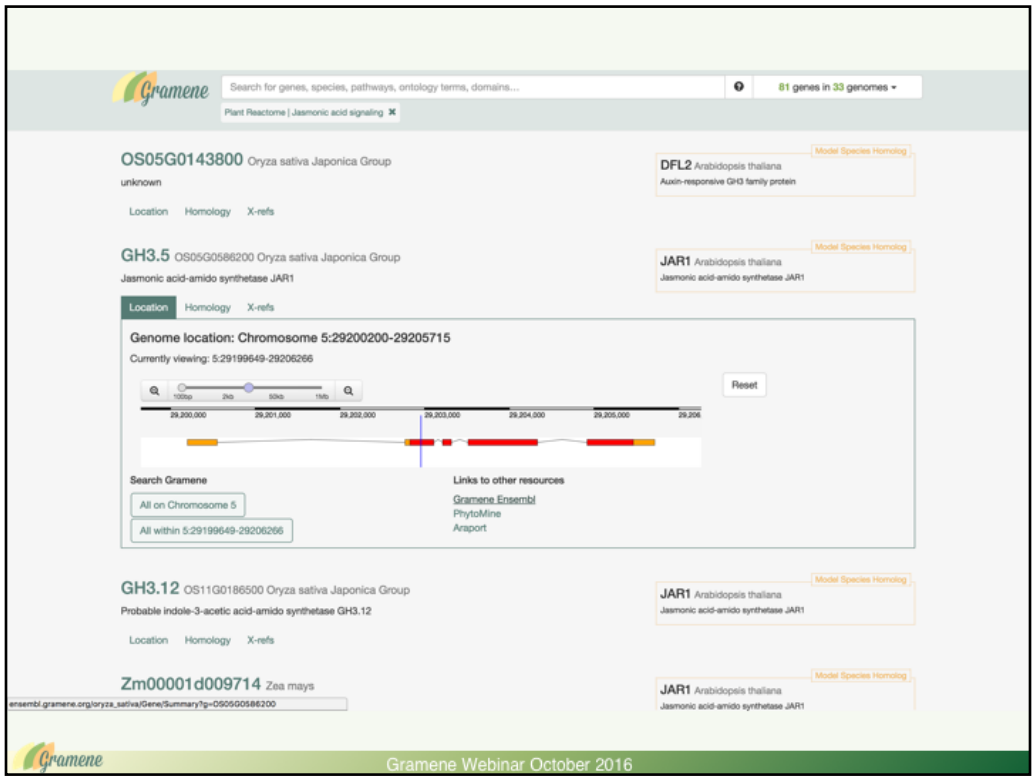
You can scroll and zoom in/out



Scroll to exon junction

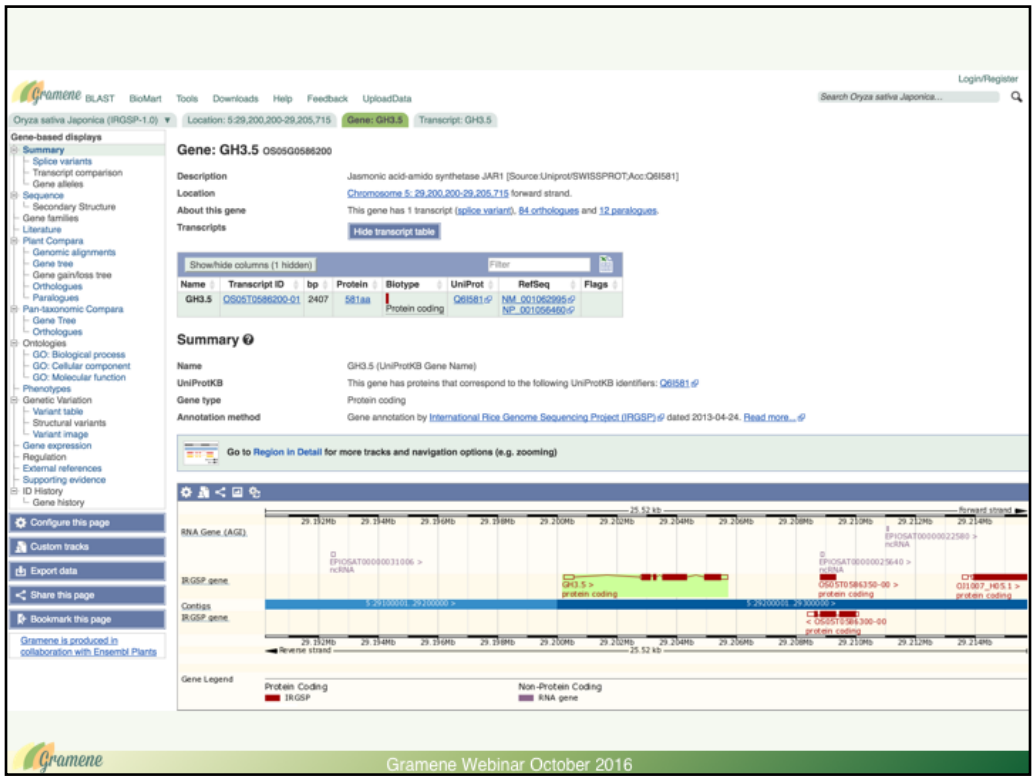


Zoom to 100bp zoom level to see splice acceptor sequence and start codon

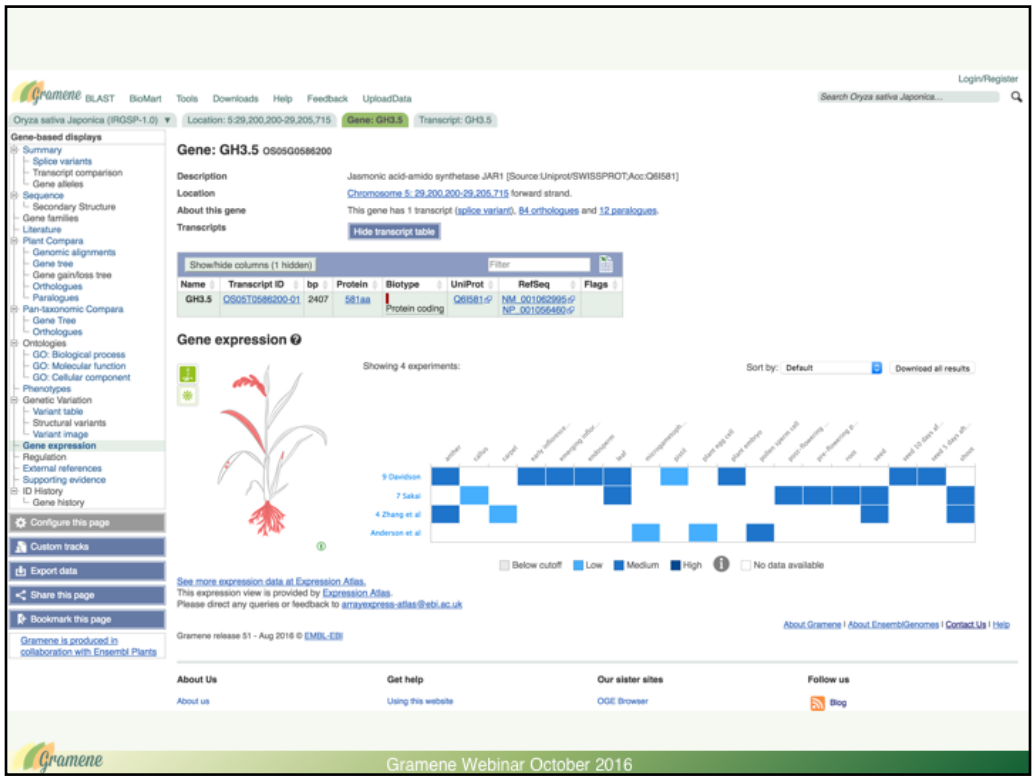


Expanded location tab on rice gene GH3.5

Click Gramene Ensembl link to go the gene page

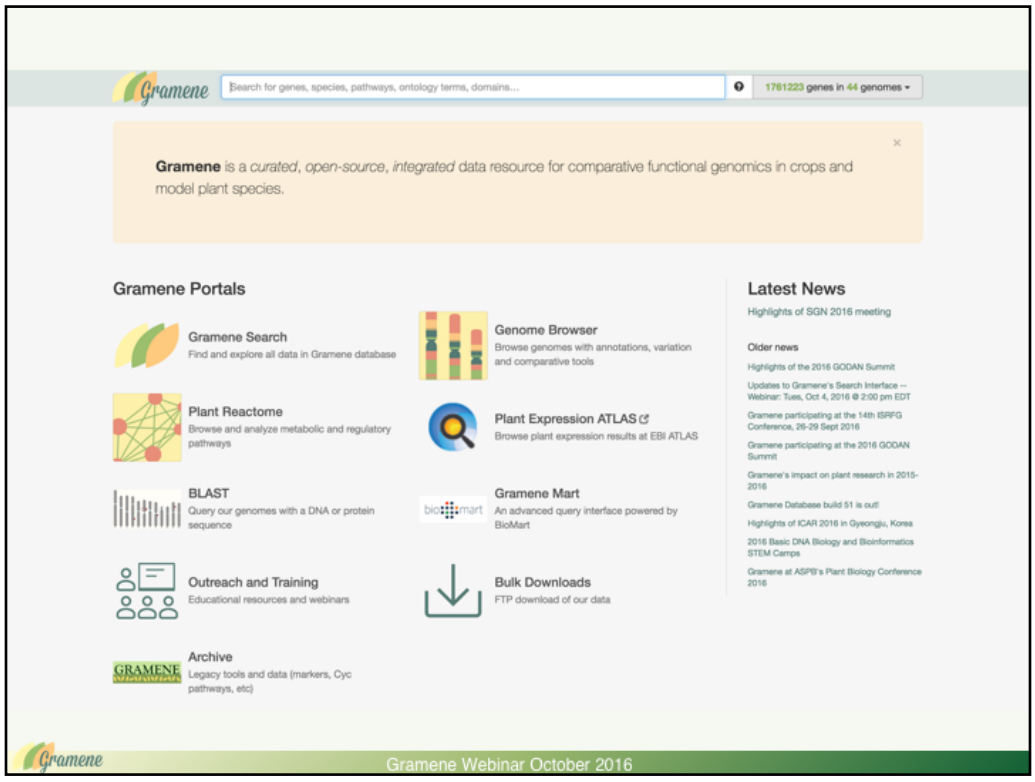


The ensembl gene page



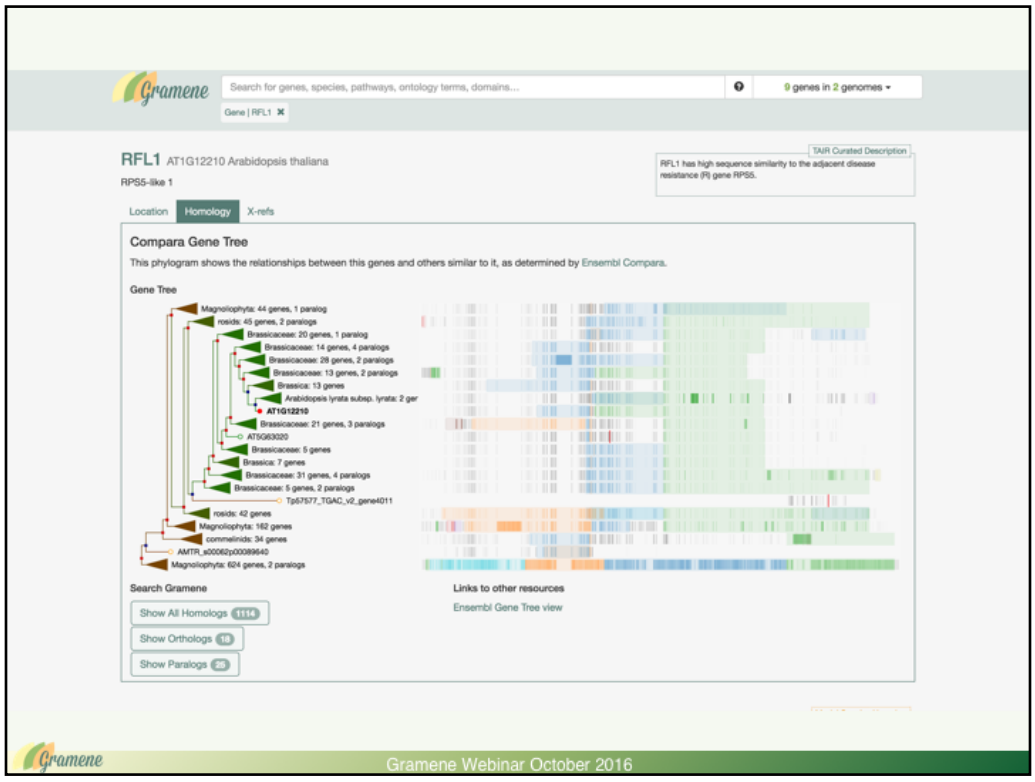
Click Gene expression in the sidebar

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



Back to the search homepage to do a new search for RFL1

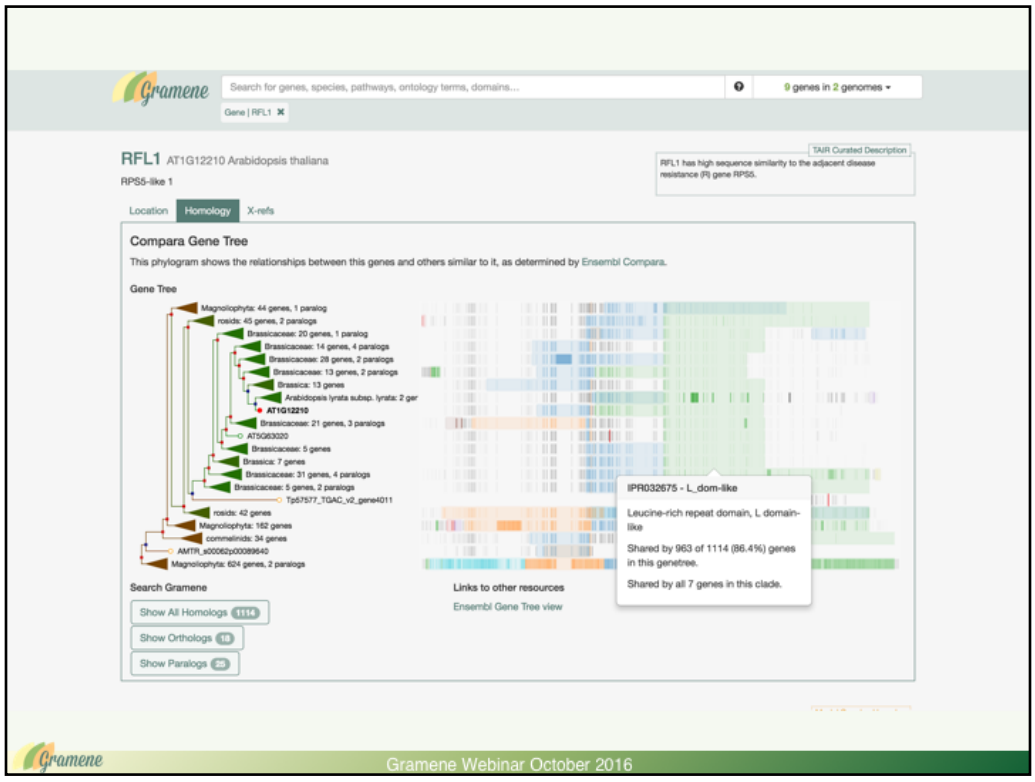
Note only a few genes are annotated with this term (9 genes in 2 genomes)



Click the homolog tab to see the gene tree (1114 genes!)

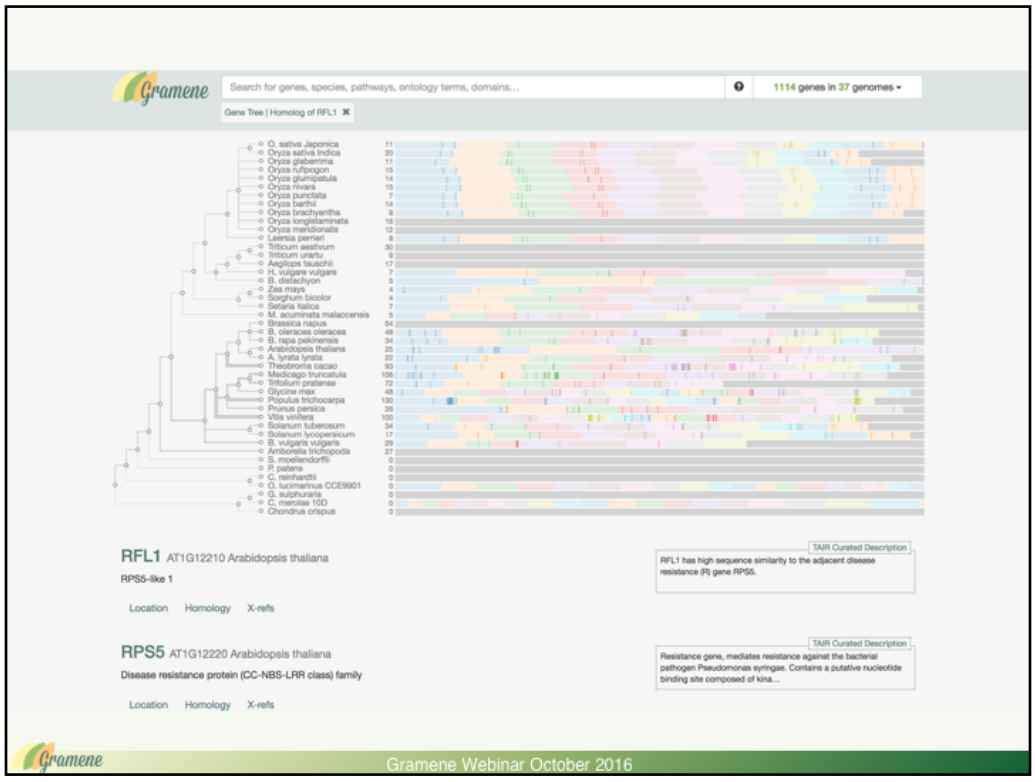
The tree branches are opened up to show the current gene of interest (RFL1)

The right hand side of the display is a view of the multiple species alignment. Colors correspond to interpro domains.



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



The taxagenomic distribution shows that there was a large expansion in the dicots relative to the monocots

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.

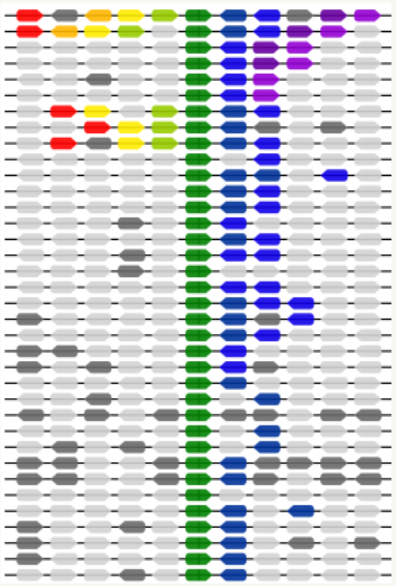
It is now tractable to visualize the evolutionary relationships between a pair of species

Features in development

- Download search results
- Link from BLAST results to Gramene search
- Improve UI for selecting genomes of interest
- Customize filter logic
- Index Atlas expression data for search
- Atlas and Plant Reactome viewers
- Gene family neighborhood conservation viewer

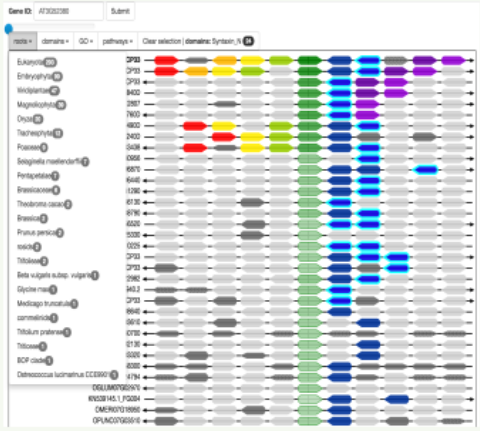
Neighborhood conservation

- Given a query gene and a set of orthologous genes, display the neighboring genes and assign colors based on the gene families of the neighbors of the query gene.
- It can show aspects of gene neighborhoods that are conserved across a gene family



Extend basic functionality

- Use shape to distinguish biotypes
- Use shading in center column to show % identity to query gene
- Add arrows to indicate direction of neighborhood
- Highlight genes by property



demo

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, Justin Preece, Sushma Naithani

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)


We're hiring!

https://cshl.peopleadmin.com/postings/10733