



# New Gramene Search Interfaces

**[search.gramene.org](http://search.gramene.org)**

**[data.gramene.org](http://data.gramene.org)**

Joseph Mulvaney and Andrew Olson  
Cold Spring Harbor Laboratory  
April 26, 2016



# Overview

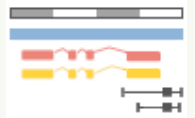
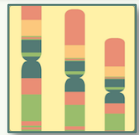
- Gramene
- Gramene Search
- Web service

# What is in Gramene?



## Genomes

*e!* EnsemblPlants



ATCGAGCT  
ATCCAGCT  
ATCGAGAT

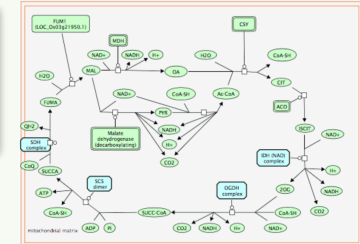
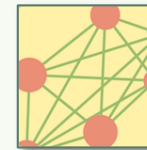
## Expression

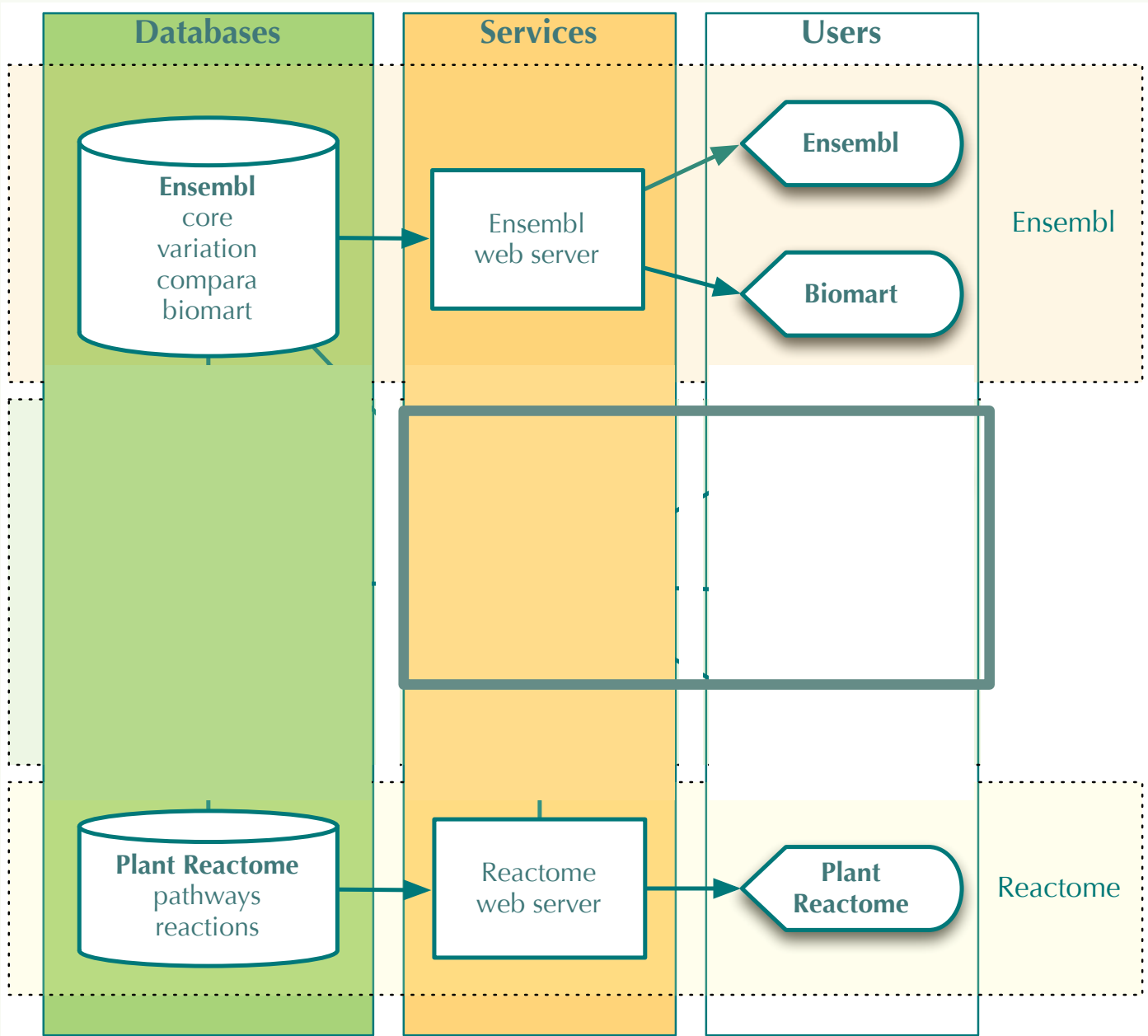
EMBL-EBI  
Expression Atlas

Gene	Arabidopsis thaliana	Brassica napus	Citrus sinensis	Helianthus annuus	Lycopersicon esculentum	Nicotiana glauca	Oryza sativa	Phaseolus vulgaris	Solanum tuberosum	Zea mays
RBC1L1A	High	Low	Low	Low	Low	Low	Low	Low	Low	Low
LHC1L3	High	High	High	High	High	High	High	High	High	High
ATG20L1	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
LHC1L2	High	High	High	High	High	High	High	High	High	High
ATPE	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low

## Pathways

*Plant* REACTOME

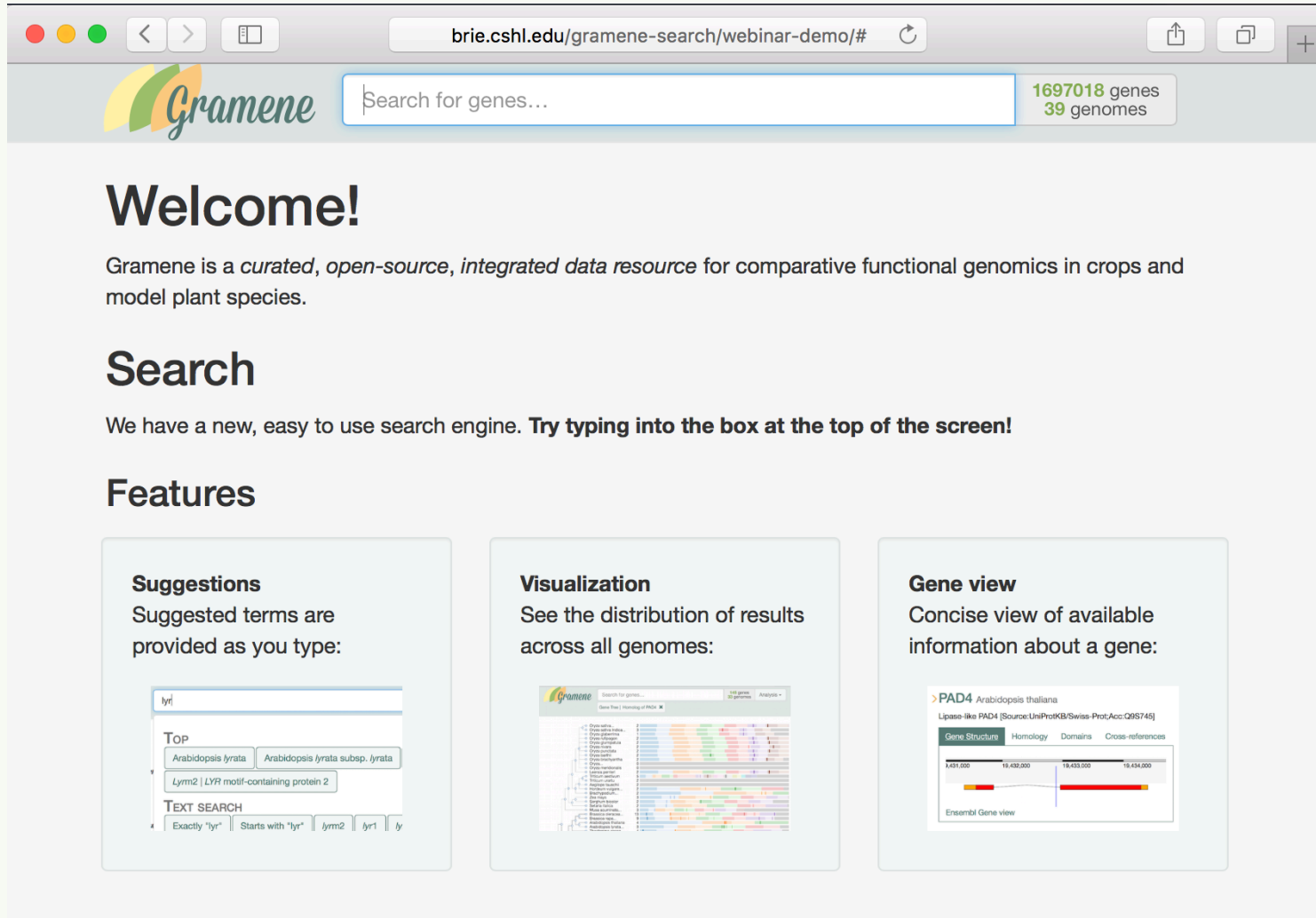




# Goals of Search Interface

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

# Gene Search



The screenshot shows a web browser window with the URL `brie.cshl.edu/gramene-search/webinar-demo/#`. The page features the Gramene logo, a search bar with the placeholder text "Search for genes...", and a status bar indicating "1697018 genes" and "39 genomes".

## Welcome!

Gramene is a *curated, open-source, integrated data resource* for comparative functional genomics in crops and model plant species.

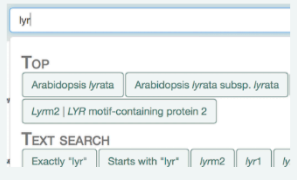
## Search

We have a new, easy to use search engine. **Try typing into the box at the top of the screen!**

## Features

### Suggestions


Suggested terms are provided as you type:



The suggestions dropdown for the query "lyr" shows a "TOP" section with "Arabidopsis lyrata" and "Arabidopsis lyrata subsp. lyrata", and a "TEXT SEARCH" section with "Lym2 | LYF1 motif-containing protein 2". Below the search bar are filters: "Exactly 'lyr'", "Starts with 'lyr'", "lym2", "lyr1", and "ly".

### Visualization

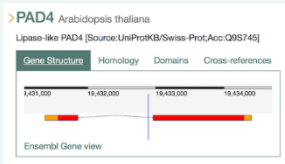
See the distribution of results across all genomes:



The visualization interface displays a heatmap where rows represent different genomes and columns represent search results. The colors in the heatmap indicate the presence and relative frequency of the search terms across the different species.

### Gene view

Concise view of available information about a gene:



The Gene view interface for "PAD4 Arabidopsis thaliana" shows a "Lipase-like PAD4 [Source:UniProtKB/Swiss-Prot;Acc:Q95745]" entry. It includes tabs for "Gene Structure", "Homology", "Domains", and "Cross-references". A genomic map shows the gene's location on a chromosome with coordinates from 1,491,000 to 19,434,000. Below the map is an "Ensembl Gene view" link.

# Gene Search

The screenshot shows the Gramene Gene Search interface. The search bar contains the text 'ara'. The results are categorized into several sections: TOP, TAXONOMY, INTERPRO, PLANT REACTOME, and GENE. Each category contains a list of terms with associated counts. A red box highlights an empty search result box in the INTERPRO section.

**Gramene** 1697018 genes  
39 genomes

ara

### TOP

- Arabidopsis thaliana 33602
- AraC-bd 8
- Arabinogalactan\_peptide 258
- Tscrpt\_reg\_HTH\_AraC\_N 1
- HTH\_AraC 37

### TAXONOMY

- Arabidopsis thaliana 33602
- Arabidopsis lyrata subsp. lyrata 32667
- Arabidopsis lyrata 32667
- Arabidopsis 66269

### INTERPRO

- ArAA\_b-elim\_lyase/Thr\_aldolase 90
- AraC-bd 8
- Arabinogalactan\_peptide 258
- Tscrpt\_reg\_HTH\_AraC\_N 1
- HTH\_AraC 37
- HTH\_AraC-typ\_CS 19
- ArAA\_hydroxylase\_Fe/CU\_BS 3
- AraC-type\_N 1
- ArAA\_permease 3
- Tscrpt\_reg\_HTH\_AraC-type 17

### PLANT REACTOME

- CMP-KDO biosynthesis II (from D-arabinose 5-phosphate) 92
- UDP-L-arabinose biosynthesis I (from UDP-xylose) 118
- UDP-arabinose 4-epimerase 118
- (1->4)-beta-D-xylan (n+1) [Golgi lumen]->arabinoxylan [Golgi lumen] 24

### GENE

- ARA 5
- ARA-1 5
- ARA-2 2
- ARA-4 7
- ARA2 5
- ARA3 8
- ARA4 2

**Welcome**  
Gramene is a *curated*, open access database of model plant species.

**Search**  
We have a new, easy to use search interface.

**Features**

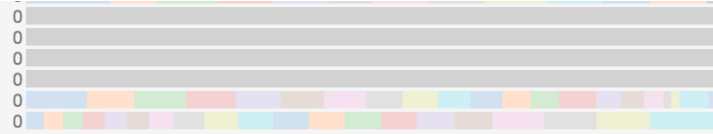
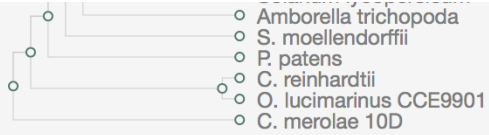
**Suggestions**  
Suggested terms are provided as you type:

lyr

**TOP**  
Arabidopsis lyrata Arabidopsis lyrata  
Lym2 | LYR motif-containing protein 2

**TEXT SEARCH**  
Exactly "lyr" Starts with "lyr" Lym2

**For Example**  
You can use the search interface to ask sophisticated questions of model plant genomes:



> **RABE1C** AT3G46060 Arabidopsis thaliana

RAB GTPase homolog 8A

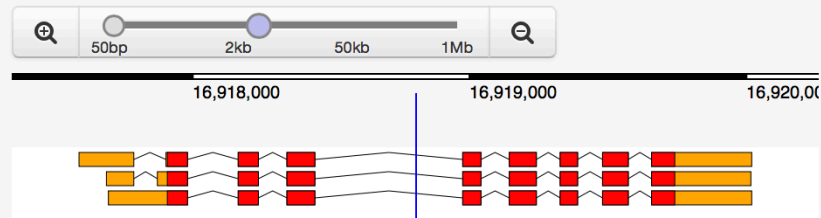
TAIR Curated Description

small GTP-binding protein (ara-3)

Location Homology X-refs

Genome location: Chromosome 3:16917589-16920018

Currently viewing: 3:16917346-16920261



Reset

Search Gramene

All on Chromosome 3

All within 3:16917346-16920261

Links to other resources

Gramene Ensembl  
PhytoMine  
Araport

> **Bra006690** Brassica rapa subsp. pekinensis

AT3G46060 (E=8e-097) ARA3 | ATRAB8A; GTP binding

Model Species Homolog

**AT5G59840** Arabidopsis thaliana

Ras-related small GTP-binding family protein

Location Homology X-refs

> **Bra038265** Brassica rapa subsp. pekinensis

AT3G46060 (E=2e-090) ARA3 | ATRAB8A; GTP binding

Model Species Homolog

**RABE1C** Arabidopsis thaliana

Ras-related protein RABE1c



# Composing Queries

You can add more filters to your search. For example:

Genes that are:-

- Involved in jasmonic acid biosynthesis
- In dicots

[Demo](#)

<http://brie.cshl.edu/gramene-search/webinar-demo/#>

# Exploring Gramene

Many plant genomes do not have good curated, descriptive gene annotations (and we've already run BLAST)

- We can work around that using
  - Arabidopsis thaliana, and
  - Gramene's homology information

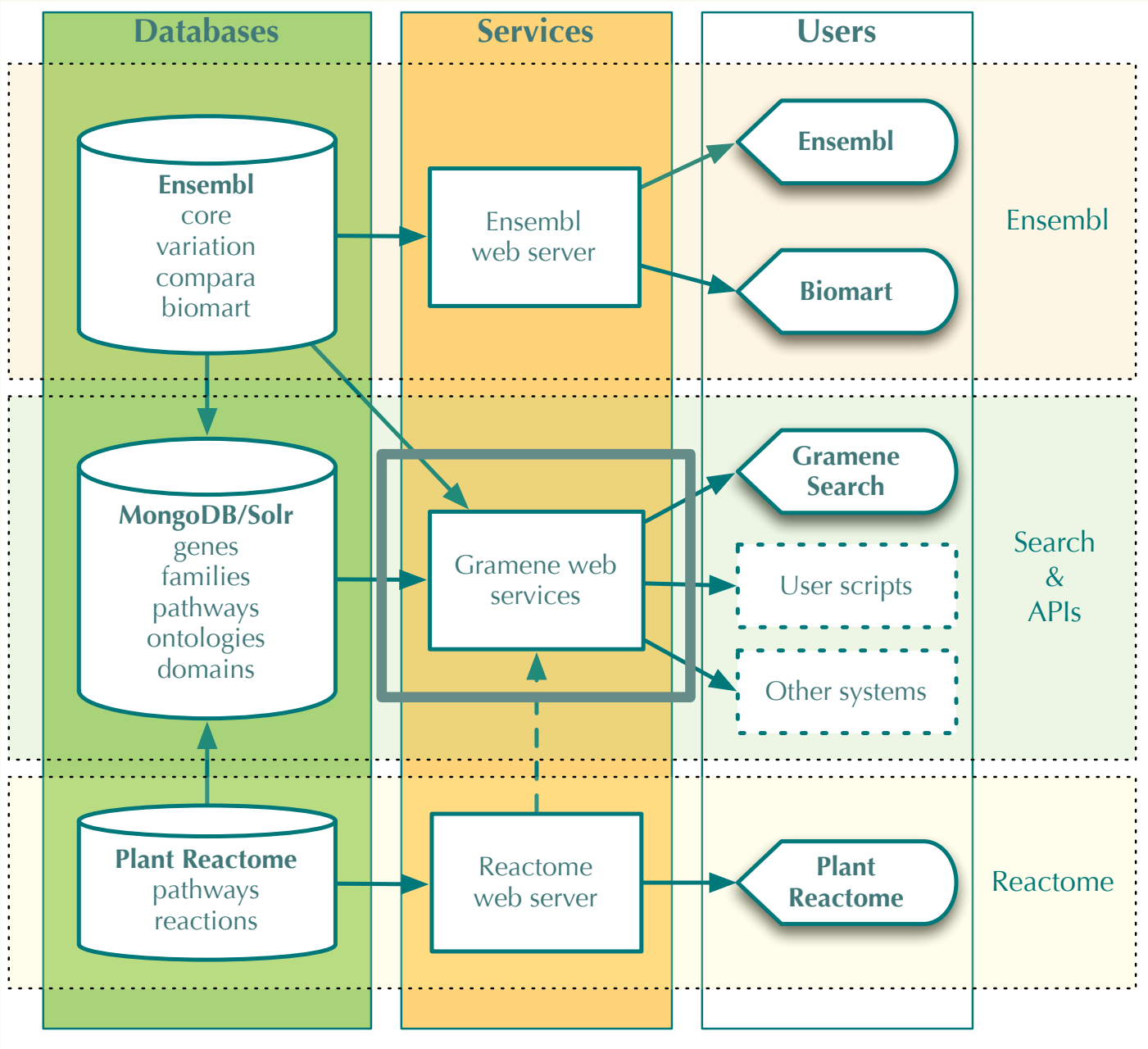
## Demo

<http://brie.cshl.edu/gramene-search/webinar-demo/#>

# Web Service API

Underpinning this search is a simple and powerful API





# Web service

The screenshot shows a web browser window displaying the Swagger API documentation for the Gramene API v50. The browser's address bar shows the URL `data.gramene.org/v50/docs/?url=/v50/swagger#/`. The Swagger interface has a green header with the Swagger logo, the URL `http://data.gramene.org/v50/swagger`, an `api_key` input field, and an `Explore` button.

## Gramene API for release 50

**API Documentation** [Show/Hide](#) | [List Operations](#) | [Expand Operations](#)

**Data access** [Show/Hide](#) | [List Operations](#) | [Expand Operations](#)

GET	/genes	MongoDB
GET	/genetrees	MongoDB
GET	/maps	MongoDB
GET	/domains	MongoDB
GET	/taxonomy	MongoDB
GET	/GO	MongoDB
GET	/PO	MongoDB
GET	/pathways	MongoDB
GET	{collection}	MongoDB

**Search** [Show/Hide](#) | [List Operations](#) | [Expand Operations](#)

[ BASE URL: /v50 , API VERSION: 0.50.11 ] [VALID](#) [{...}](#)

# Web service

Response Content Type

### Parameters

Parameter	Value	Description	Parameter Type	Data Type
q	<input type="text" value="pad4"/>	string for free text search	query	string
rows	<input type="text"/>	number of rows to return (default: 20)	query	integer
start	<input type="text"/>	offset into results list	query	integer
fl	<input type="text" value="Provide multiple values in new lines."/>	list of fields to return	query	Array[string]
wt	<input type="text" value=""/>	response format, if tab is specified fl must also be, otherwise JSON documents returned.	query	string
idList	<input type="text" value="Provide multiple values in new lines."/>	a comma separated list of document ids	query	Array[string]
bedFeature	<input type="text" value="gene (default)"/>	type of feature to export	query	string
bedCombiner	<input type="text" value="canonical (default)"/>	method for extracting features from multiple transcripts	query	string
taxon_id	<input type="text"/>	NCBI taxonomy id	query	integer
db_type	<input type="text"/>	set to core if you don't care about otherfeatures	query	string

# Web service

Try it out! [Hide Response](#)

**Curl**

```
curl -X GET --header 'Accept: application/json' 'http://data.gamene.org/v50/genes?q=pad4&bedFeature=gene&bedCombiner=canonical'
```

**Request URL**

```
http://data.gamene.org/v50/genes?q=pad4&bedFeature=gene&bedCombiner=canonical
```

**Response Body**

```
[
  {
    "_id": "Bo7g102600",
    "name": "Bo7g102600",
    "description": null,
    "biotype": "protein_coding",
    "taxon_id": 109376,
    "system_name": "brassica_oleracea",
    "db_type": "core",
    "gene_idx": 40993,
    "location": {
      "region": "C7",
      "start": 39678647,
      "end": 39682835,
      "strand": -1,
      "map": "GCA_000695525.1"
    },
    "xrefs": [
      {
        "db": "UniParc",
```

**Response Code**

```
200
```

**Response Headers**

# Web service (BED file)

q	<input type="text"/>	string for free text search	query	string
rows	<input type="text"/>	number of rows to return (default: 20)	query	integer
start	<input type="text"/>	offset into results list	query	integer
fl	<input type="text" value="Provide multiple values in new lines."/>	list of fields to return	query	Array[string]
wt	<input type="text" value="bed"/>	response format, if <code>tab</code> is specified <code>fl</code> must also be, otherwise JSON documents returned.	query	string
idList	<input type="text" value="Provide multiple values in new lines."/>	a comma separated list of document ids	query	Array[string]
bedFeature	<input type="text" value="transcript"/>	type of feature to export	query	string
bedCombiner	<input type="text" value="canonical (default)"/>	method for extracting features from multiple transcripts	query	string
taxon_id	<input type="text" value="4577"/>	NCBI taxonomy id	query	integer
db_type	<input type="text" value="core"/>	set to core if you don't care about otherfeatures	query	string

[Hide Response](#)



# Web service (BED file)

Try it out! [Hide Response](#)

### Curl

```
curl -X GET --header 'Accept: text/tab-separated-values' 'http://data.gamene.org/v50/genes?wt=bed&bedFeature=transcri'
```

### Request URL

```
http://data.gamene.org/v50/genes?wt=bed&bedFeature=transcript&bedCombiner=canonical&taxon_id=4577&db_type=core
```

### Response Body

1	4854	9652	GRMZM2G059865_T01	0	-	5127	9519	0	9	335,66,119,158,156,159
1	9882	10387	GRMZM5G888250_T01	0	-	9887	10117	0	1	506 0
1	109519	111769	GRMZM2G093344_T01	0	-	111111	111727	0	4	157,177,374,304 0,157,
1	136307	138929	GRMZM2G093399_T01	0	+	138579	138899	0	4	328,295,86,378 0,328,
1	144361	144657	GRMZM5G809743_T01	0	+	144361	144657	0	1	297 0
1	144957	145646	GRMZM5G833153_T01	0	+	144957	145646	0	1	690 0
1	146265	147500	GRMZM2G306216_T01	0	-	146265	147500	0	4	8,261,58,78 0,8,26
1	161143	161925	AC177838.2_FGT015	0	+	161143	161925	0	1	783 0
1	234779	235890	GRMZM2G104572_T01	0	+	235319	235816	0	1	1112 0
1	243464	243742	GRMZM5G822187_T01	0	+	243544	243726	0	1	279 0
1	257026	257371	GRMZM5G815900_T01	0	+	257120	257371	0	1	346 0
1	264665	270895	GRMZM2G023794_T02	0	-	264665	270895	0	6	292,440,505,281,264,19
1	272525	274767	GRMZM2G023863_T02	0	+	272592	274767	0	6	73,158,119,66,269,96
1	371975	375908	GRMZM2G137648_T01	0	-	374004	375878	0	1	3934 0
1	372538	374866	GRMZM2G137697_T01	0	+	373661	373897	0	2	1684,259 0,1684



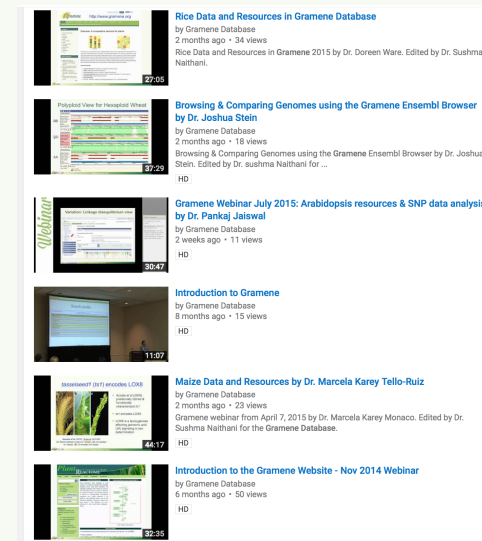
Beta testers  
wanted!

Sign up at [gramene.org/  
beta](http://gramene.org/beta)

# Gramene Database @ YouTube



The video player shows a man in an orange hoodie sitting at a desk with a laptop. Behind him is a large screen displaying a presentation slide. The slide has the Gramene and EnsemblPlants logos at the top. The main text on the slide reads: "Browsing & Comparing Genomes using the Gramene Ensembl Browser". Below this, it lists the websites "www.gramene.org" and "www.plants.ensembl.org", the name "Joshua Stein", and "Cold Spring Harbor Laboratory". Logos for CSH, Gramene, and EMBL-EBI are at the bottom of the slide. The video player interface includes a play button, a progress bar at 0:30 / 37:28, and a "Subscribe" button for the "Gramene Database" channel, which has 16 subscribers. The video title is "Browsing & Comparing Genomes using the Gramene Ensembl Browser by Dr. Joshua Stein" and it has 43 views.



- Rice Data and Resources in Gramene Database**  
by Gramene Database  
2 months ago • 34 views  
Rice Data and Resources in Gramene 2015 by Dr. Doreen Ware. Edited by Dr. Sushma Naithani. 27:06
- Browsing & Comparing Genomes using the Gramene Ensembl Browser by Dr. Joshua Stein**  
by Gramene Database  
2 months ago • 18 views  
Browsing & Comparing Genomes using the Gramene Ensembl Browser by Dr. Joshua Stein. Edited by Dr. Sushma Naithani for ... 37:28
- Gramene Webinar July 2015: Arabidopsis resources & SNP data analysis**  
by Dr. Pankaj Jaiswal  
by Gramene Database  
2 weeks ago • 11 views  
HD 38:47
- Introduction to Gramene**  
by Gramene Database  
8 months ago • 15 views  
HD 11:07
- Maize Data and Resources by Dr. Marcela Karey Tello-Ruiz**  
by Gramene Database  
2 months ago • 23 views  
Gramene webinar from April 7, 2015 by Dr. Marcela Karey Monaco. Edited by Dr. Sushma Naithani for the Gramene Database. 44:17
- Introduction to the Gramene Website - Nov 2014 Webinar**  
by Gramene Database  
6 months ago • 50 views  
HD 32:35

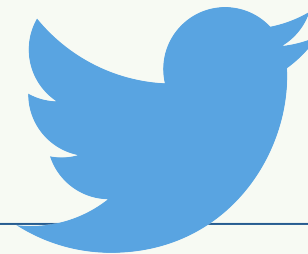
- 13 live webinars & 14 recorded video-tutorials
- Master mailing list of over 1,100 plant researchers
- Engaging specific plant communities (Arabidopsis, rice, maize, etc.)
- Create a robust presence for Gramene in the Plantae platform (~30,000 plant scientists globally)

# How to reach us

- E-mail: [webinars@gramene.org](mailto:webinars@gramene.org)
- Contact form:  
<http://www.gramene.org/contact>
- Gramene YouTube channel
- Announcements mailing list

The Facebook logo, consisting of the word "facebook" in white lowercase letters on a dark blue rectangular background.

<https://www.facebook.com/Gramene>



[@GrameneDatabase](https://twitter.com/GrameneDatabase)

# 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 Petryszuk** (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)

