



Upload your data in Gramene

Website: www.gramene.org

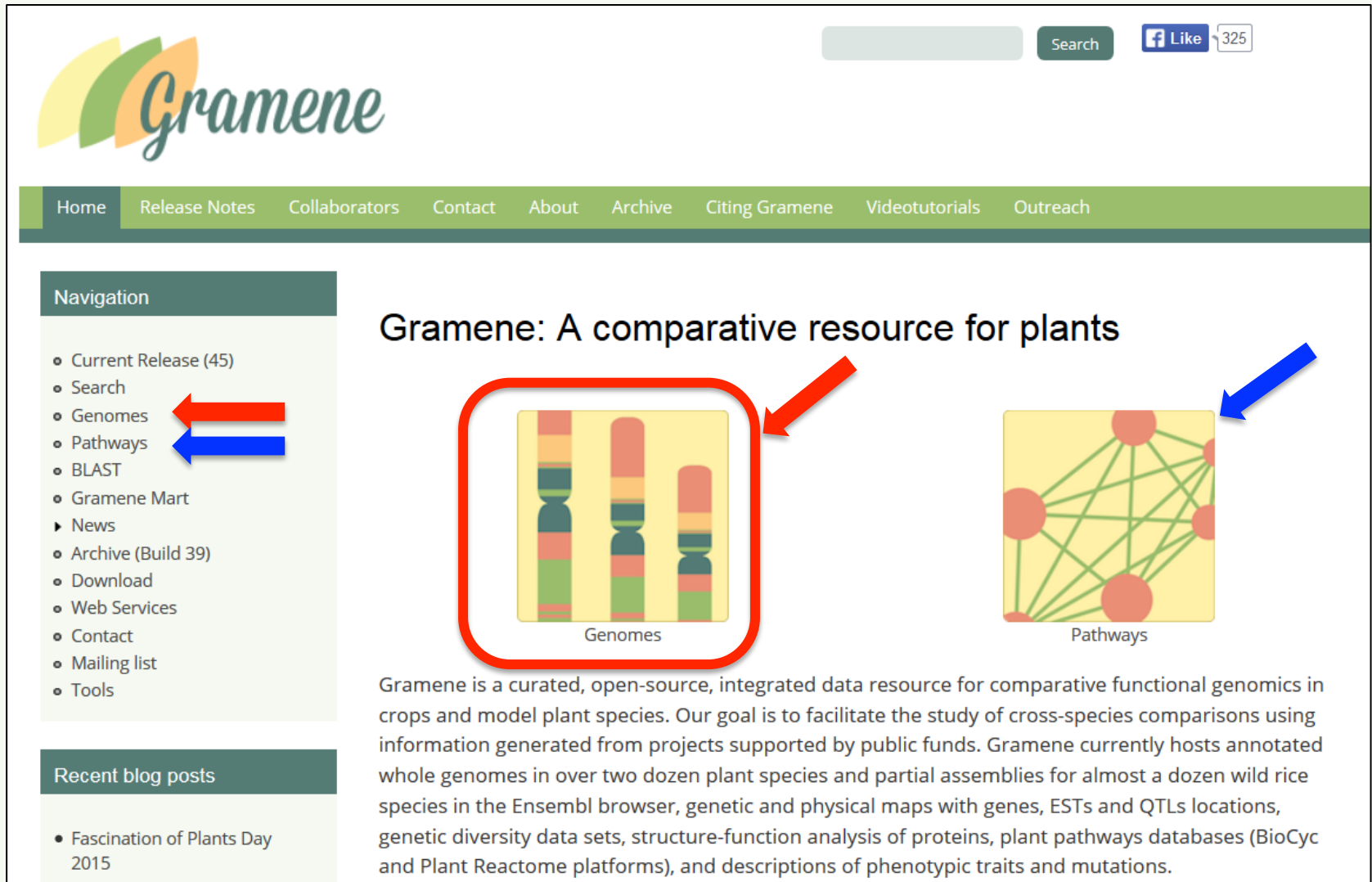
Pankaj Jaiswal
Oregon State University

Gramene Outreach Webinar
June 22, 2015



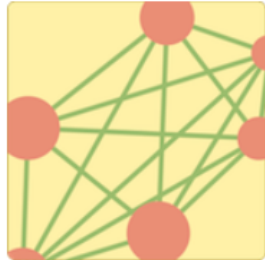
Start at Gramene Home

www.gramene.org



The screenshot shows the Gramene website homepage. At the top is the Gramene logo, a search bar, and a Facebook 'Like' button with 325 likes. Below this is a green navigation bar with links: Home, Release Notes, Collaborators, Contact, About, Archive, Citing Gramene, Videotutorials, and Outreach. On the left is a 'Navigation' sidebar with a list of links: Current Release (45), Search, Genomes, Pathways, BLAST, Gramene Mart, News, Archive (Build 39), Download, Web Services, Contact, Mailing list, and Tools. Two red arrows point to 'Genomes' and 'Pathways' in this list. Below the sidebar is a 'Recent blog posts' section with one post titled 'Fascination of Plants Day 2015'. The main content area features the heading 'Gramene: A comparative resource for plants'. Below this heading are two icons: 'Genomes', which is a red-outlined box containing three vertical bars of stacked colored segments, and 'Pathways', which is a network diagram with red nodes and green connecting lines. A red arrow points to the 'Genomes' icon, and a blue arrow points to the 'Pathways' icon. Below these icons is a paragraph of text describing Gramene as a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species.

Gramene: A comparative resource for plants



Genomes

Pathways

Gramene is a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species. Our goal is to facilitate the study of cross-species comparisons using information generated from projects supported by public funds. Gramene currently hosts annotated whole genomes in over two dozen plant species and partial assemblies for almost a dozen wild rice species in the Ensembl browser, genetic and physical maps with genes, ESTs and QTLs locations, genetic diversity data sets, structure-function analysis of proteins, plant pathways databases (BioCyc and Plant Reactome platforms), and descriptions of phenotypic traits and mutations.

Oryza sativa japonica
Oryza rufipogon
Oryza sativa indica
Oryza nivara
Oryza glaberrima
Oryza barthii
Oryza glumaepatula
Oryza meridionalis
Oryza punctata
Oryza brachyantha
Leersia perrieri
Triticum aestivum AABBD
Triticum urartu AA
Aegilops tauschii DD
Hordeum vulgare
Brachypodium distachyon
Zea mays
Sorghum bicolor
Setaria italica
Musa acuminata

List of 39 sequenced plant genome dataset provided by Gramene

Every species has its own genome browser

Monocots

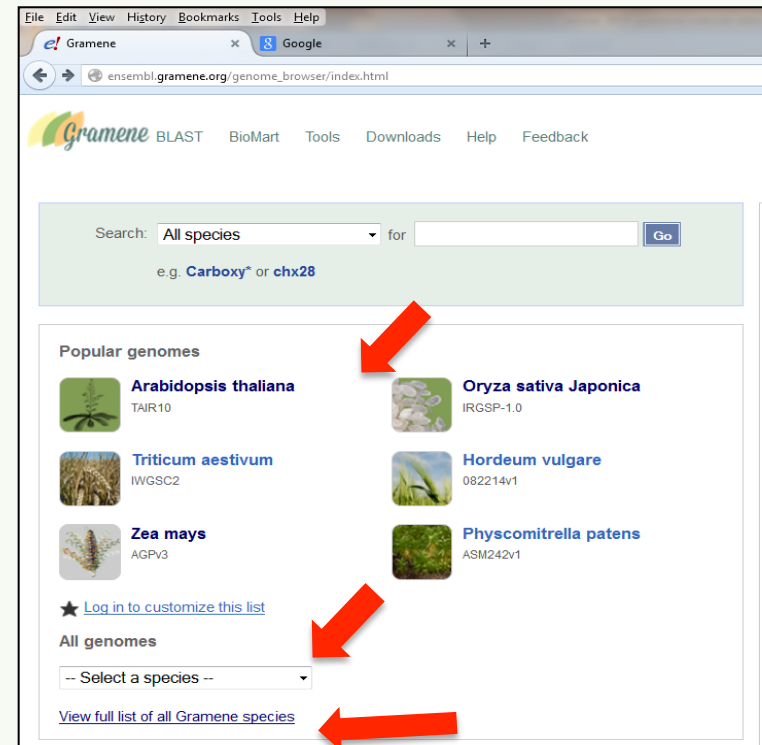
<http://ensembl.gramene.org/info/about/species.html>

Arabidopsis thaliana
Arabidopsis lyrata
Brassica oleracea
Brassica rapa
Glycine max
Medicago truncatula
Prunus persica
Populus trichocarpa
Vitis vinifera
Theobroma cacao
Solanum lycopersicum
Solanum tuberosum

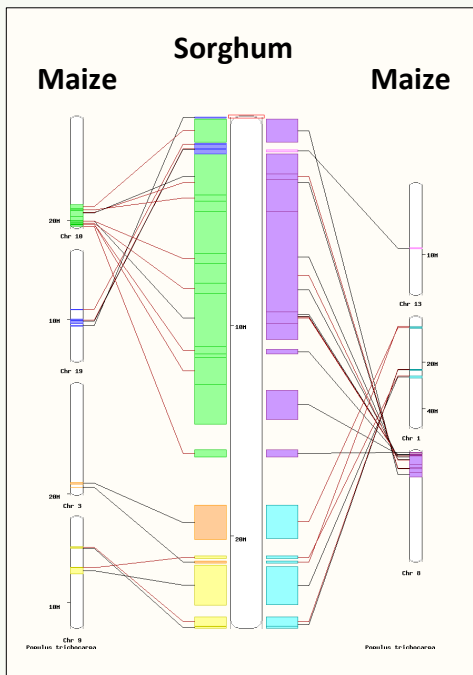
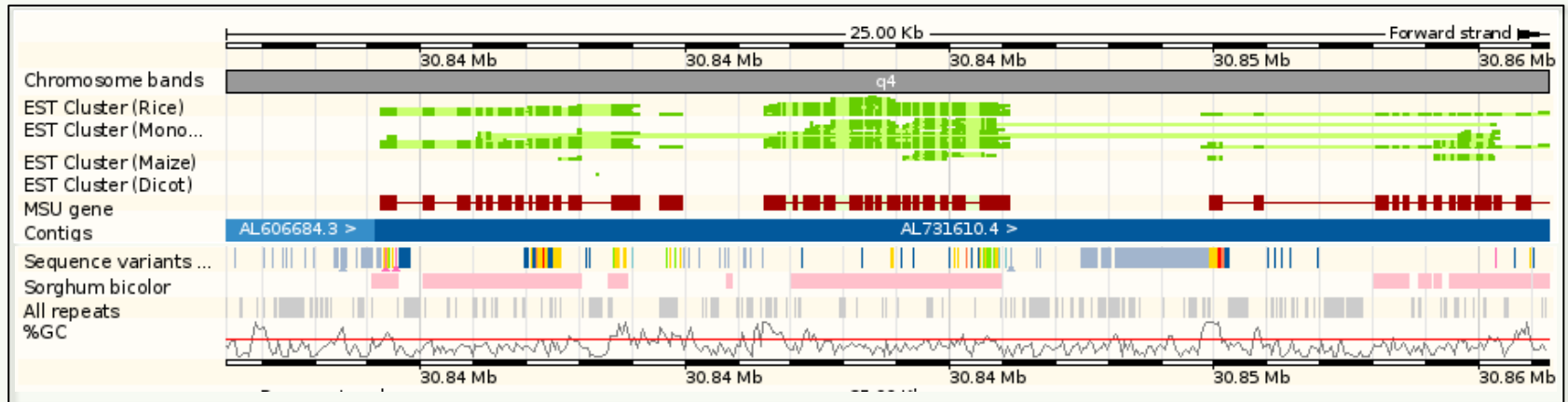
Dicots

Amborella trichopoda (Basal Angiosperm)

Selaginella moellendorffii (Lycopod)
Physcomitrella patens (Moss)
Chlamydomonas reinhardtii (green algae)
Ostreococcus lucimarinus (green algae)
Cyanidioschyzon merolae (red algae)



Species Specific Genome Browser

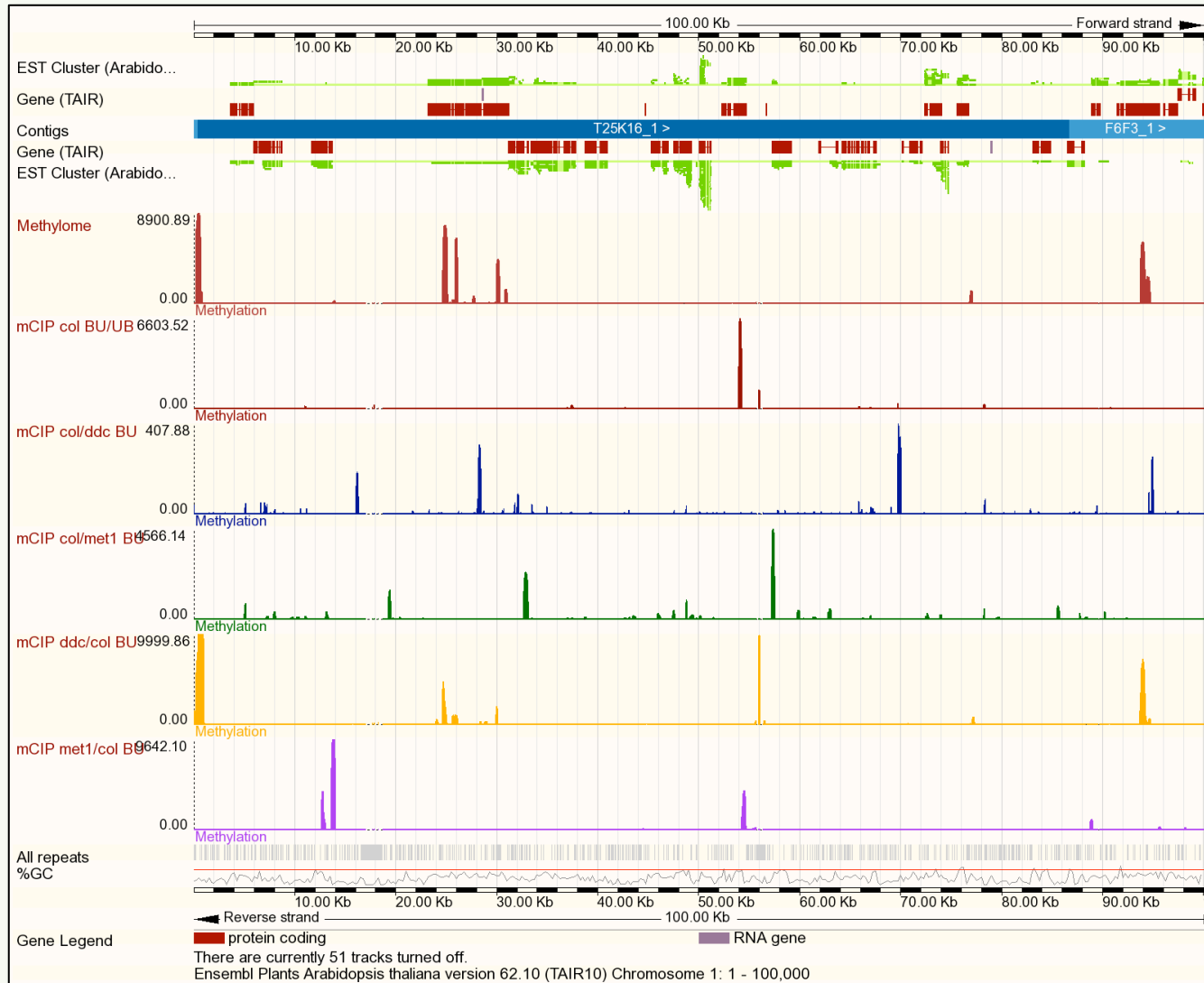


- Genome assembly structure and sequence
- Genes (protein coding and ncRNA)
- Expressed regions
- Comparative genome alignments
- Genetic markers
- Variation (SNPs)
- Repeats & transposable elements
- Regulatory & Epigenetic loci
- **DATA UPLOADED BY YOU FOR YOUR OWN VIEW**



EXAMPLE-1

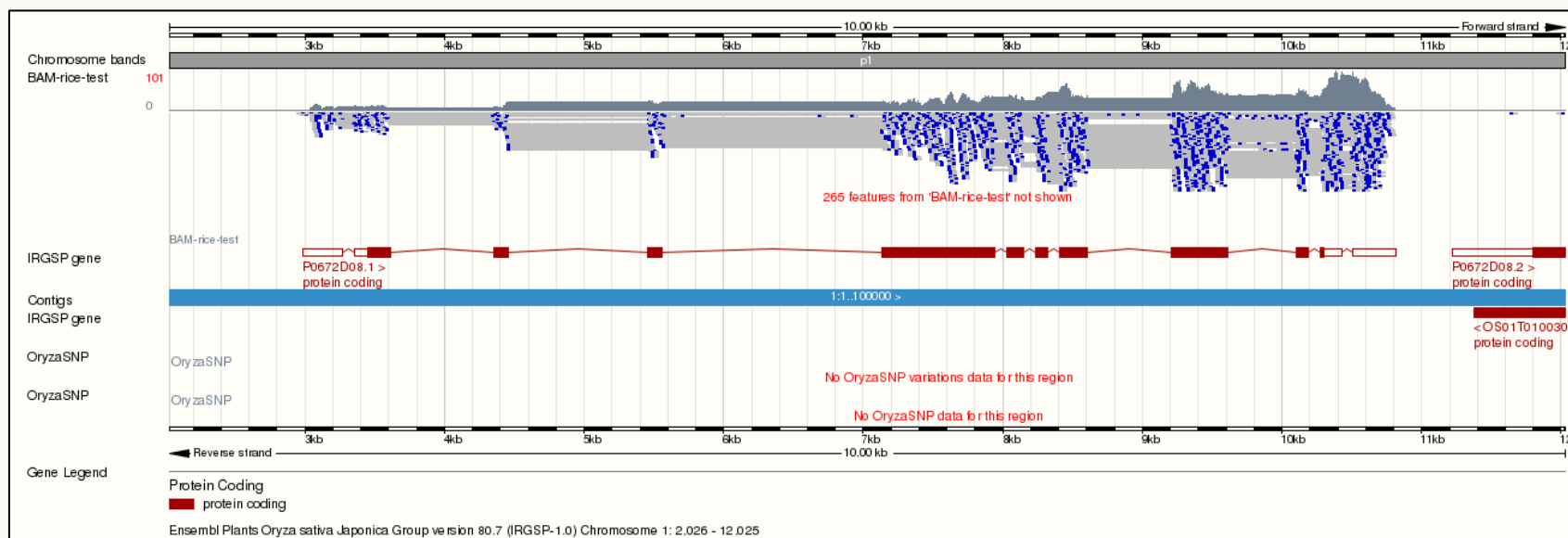
- E.g. Methylome/Expression data
- BED/BEDgraph file format
- Uploaded from my computer/URL



Add Custom Tracks

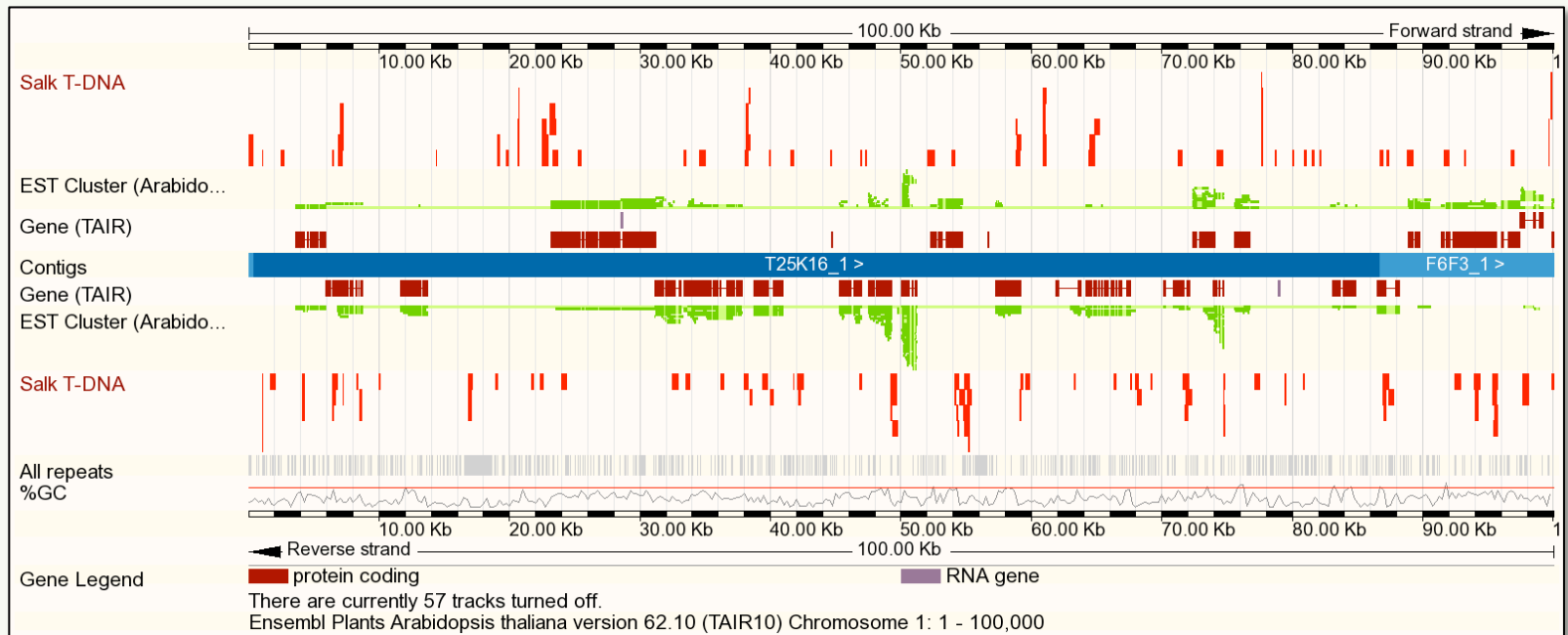
EXAMPLE-2

- NGS sequencing data aligned to the reference genome
- BAM format
- Uploaded from an URL



EXAMPLE-3

- T-DNA lines, markers, BLAST/BLAT alignments, etc.
- GFF/BED file format
- Uploaded from my computer/URL



Data types & formats that you can upload

• Data types (Pre-analyzed)

- Alignments/mappings of (non)sequenced features
- Expression data
- NGS
- Variations including allele information
- Known features
- Interactions

• Data Formats

- BED
- BedGraph
- GFF/GTF
- Pairwise interactions (WashU)
- PSL
- WIG
- BAM (NGS data)
- BigBed
- BigWig
- VCF (for variations)

For more information visit:

<http://ensembl.gramene.org/info/website/upload/index.html>

Location of my data?

- Via URL
 - Especially for large files e.g. GFF & BAM formats
 - Direct links to your data files
 - e.g. from iPlant data store or any file share server
- From your computer (20MB file size limit)
- Distributed Annotation System (DAS)

Uploaded data is saved temporarily to our file system and accessed by you via cookie, so only you can view it from the browser in which you uploaded it. For more information visit...

<http://ensembl.gramene.org/info/website/upload/index.html#access>

Expression Data Upload

EXAMPLE-1

- ✓ Microarray
- ✓ NGS: RNA-Seq
- ✓ Proteomics

Prepare Expression Data Files

Excel (all samples) to BED/BEDgraph format (one sample per file)

Chromosome	Gene start (bp)	Gene end (bp)	16DAP_Embryo	16DAP_Endosperm	V1_GH_Primary Root	R1_Anthers	V1_Pooled Leaves	GENE_ID
1	136307	138929	-107.314	-70.234	324.17603	-97.804	-48.824005	GRMZM2G093399
1	161143	161925	-67.8	883.9	-389.2	-173.8	-253.09999	AC177838.2_FG015
1	644611	645388	1021.984	-143.11601	-284.606	-327.146	-267.116	GRMZM2G333069
1	1001050	1004936	-1151.6	-2904.6	-2006.6	-1505.6	7568.4	GRMZM2G374812
1	1564786	1566049	3550.394	-1405.976	194.394	-1115.106	-1223.706	GRMZM2G097229
1	2219065	2223784	-1313.022	-2027.862	-2010.772	7202.578	-1850.922	GRMZM2G410990
1	2464882	2467458	434.948	-155.432	-34.35201	-113.532	-131.632	GRMZM2G028870
1	2734521	2739053	-879.74	-659.74	-3170.04	-1790.74	6500.26	GRMZM2G085872
1	2740244	2744906	-635.63	-835.83	-1178.08	-531.13	3180.67	GRMZM2G086242
1	2751021	2753499	-604.254	-585.534	-617.364	1617.876	189.27602	GRMZM2G086269

```
# 300 base wide bar graph, autoScale is on by default == graphing
# limits will dynamically change to always show full range of data
# in viewing window, priority = 20 positions this as the second graph
# Note, zero-relative, half-open coordinate system in use for bedGraph format
track type=bedGraph name="Anther" description="Expression" visibility=full color=600,300,0 altColor=0,100,200 priority=20
1 136307 138929 -97.804 GRMZM2G093399
1 161143 161925 -173.8 AC177838.2_FG015
1 644611 645388 -327.146 GRMZM2G333069
1 1001050 1004936 -1505.6 GRMZM2G374812
1 1564786 1566049 -1115.106 GRMZM2G097229
1 2219065 2223784 7202.578 GRMZM2G410990
1 2464882 2467458 -113.532 GRMZM2G028870
1 2734521 2739053 -1790.74 GRMZM2G085872
1 2740244 2744906 -531.13 GRMZM2G086242
```

Anther sample data file

```
# 300 base wide bar graph, autoScale is on by default == graphing
# limits will dynamically change to always show full range of data
# in viewing window, priority = 20 positions this as the second graph
# Note, zero-relative, half-open coordinate system in use for bedGraph format
track type=bedGraph name="V1-leaf" description="Expression" visibility=full color=400,300,0 altColor=0,100,200 priority=20
1 136307 138929 -48.824005 GRMZM2G093399
1 161143 161925 -253.09999 AC177838.2_FG015
1 644611 645388 -267.116 GRMZM2G333069
1 1001050 1004936 7568.4 GRMZM2G374812
1 1564786 1566049 -1223.706 GRMZM2G097229
1 2219065 2223784 -1850.922 GRMZM2G410990
1 2464882 2467458 -131.632 GRMZM2G028870
1 2734521 2739053 6500.26 GRMZM2G085872
1 2740244 2744906 3180.67 GRMZM2G086242
```

V1-leaf sample data file

Start at Gramene Home

www.gramene.org



Search

 Like 325

HomeRelease NotesCollaboratorsContactAboutArchiveCiting GrameneVideotutorialsOutreach

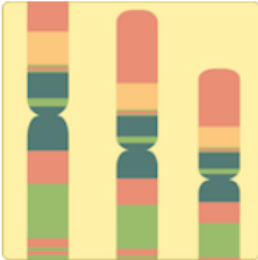
Navigation

- Current Release (45)
- Search
- Genomes 
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Mailing list
- Tools

Recent blog posts

- Fascination of Plants Day 2015

Gramene: A comparative resource for plants



Genomes



Pathways

Gramene is a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species. Our goal is to facilitate the study of cross-species comparisons using information generated from projects supported by public funds. Gramene currently hosts annotated whole genomes in over two dozen plant species and partial assemblies for almost a dozen wild rice species in the Ensembl browser, genetic and physical maps with genes, ESTs and QTLs locations, genetic diversity data sets, structure-function analysis of proteins, plant pathways databases (BioCyc and Plant Reactome platforms), and descriptions of phenotypic traits and mutations.

Sequenced plant genome dataset

File Edit View History Bookmarks Tools Help

Gramene Google

ensembl.gramene.org/genome_browser/index.html

Gramene BLAST BioMart Tools Downloads Help Feedback

Search: All species for Go

e.g. Carboxy* or chx28

Popular genomes

Arabidopsis thaliana
TAIR10

Oryza sativa Japonica
IRGSP-1.0

Triticum aestivum
IWGSC2

Hordeum vulgare
082214v1

Zea mays
AGPv3

Physcomitrella patens
ASM242v1

★ [Log in to customize this list](#)

All genomes

-- Select a species --

[View full list of all Gramene species](#)

<http://ensembl.gramene.org/info/about/species.html>

Every species has its own genome browser

Uploading Data on Genome Browser

The screenshot shows the Ensembl Genome Browser interface for *Zea mays* (maize). The browser window has tabs for 'Zea mays - Ensembl Geno...' and 'Google'. The address bar shows 'ensembl.gramene.org/Zea_mays/Info/Index'. The page features a navigation bar with links like BLAST, BioMart, Tools, Downloads, Help, and Feedback. A search bar is located in the top right corner. The main content area is titled 'Zea mays' and includes a search bar with a red arrow pointing to it. Below the search bar, there is a section 'About Zea mays' with a paragraph of text. To the right of the 'About Zea mays' section is a 'What's New in Zea mays release 79' banner. Below the 'About Zea mays' section is a 'Genome assembly' section with a red box highlighting the 'Display your data in Gramene' option. To the right of the 'Genome assembly' section is a 'Gene annotation' section with a red box highlighting the 'Example gene' option.

Zea mays
Zea mays
Data Source [Maize Genome Project](#) | Taxonomy ID [4577](#)

Search Zea mays...

e.g. [bt1](#) or [1:8001-18000](#) or [Carboxypeptidase](#)

About Zea mays

Zea mays (maize) has the highest world-wide production of all grain crops, yielding 875 million tonnes in 2012 (<http://faostat.fao.org/>). Although a food staple in many regions of the world, most is used for animal feed and ethanol fuel. Maize was domesticated from wild teosinte in Central America and its cultivation spread throughout the Americas by Pre-Columbian civilizations. In addition to its economic value, maize is an important model organism for studies in plant genetics, physiology, and development. It has a large genome of about 2.4 gigabases with a haploid chromosome number of 10 (Schnable *et al*, 2009; Zhang *et al*, 2009). Maize is distinguished from other grasses in that its genome arose from an ancient tetraploidy event unique to its lineage.

[More information and statistics](#)

Genome assembly: [AGPv3](#)

[More information and statistics](#)

[Download DNA sequence \(FASTA\)](#)

[Convert your data to \[AGPv3\]\(#\) coordinates](#)

[Display your data in Gramene](#)

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

[More about this genebuild](#)

[Download genes, cDNAs, ncRNA, proteins - FASTA - GFF3](#)

[Update your old Ensembl IDs](#)

Example gene

Example transcript

Uploading Data on Genome Browser

The screenshot displays the Zea mays Genome Browser interface. The top navigation bar shows the location: 1:2,726,328-2,761,602. The main view shows Chromosome 1 with a detailed region view of 500.00 kb. The left sidebar contains a 'Location-based displays' menu with options like 'Whole genome', 'Chromosome summary', 'Region overview', 'Region in detail', 'Comparative Genomics', 'Genetic Variation', 'Resequencing', 'Linkage Data', 'Markers', 'Other genome browsers', and 'Phytozome'. The 'Region in detail' section is highlighted with a red box. Below this, a 'Configure this page' button is also highlighted with a red box. The 'Add a custom track' dialog box is open, showing options to 'Name for this data (optional)', 'Species' (Zea mays), 'Assembly' (AGPv3), 'Data format' (bedGraph), 'Type' (Upload data (max 20MB) selected), and 'Paste data'. The 'Upload' button is highlighted with a red arrow. The 'Add your data' option in the left sidebar is also highlighted with a red box.

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail**
- Comparative Genomics
 - Alignments (image)
 - Alignments (text)
 - Region Comparison
 - Synteny
- Genetic Variation
 - Resequencing
 - Linkage Data
- Markers
- Other genome browsers
- Phytozome

Configure this page

- Manage your data
- Export data
- Share this page
- Bookmark this page
- [Ensembl Plants is produced in collaboration with Gramene](#)

Region in detail

Chromosome bands

Contigs

Game

Add a custom track

Name for this data (optional):

Species: Zea mays

Assembly: AGPv3

Data format: bedGraph

[Help on supported formats, display types](#)

Type:
☒ Upload data (max 20MB)
☐ Attach via URL

Paste data:

Or choose file: Browse... No file selected.

Or provide file URL:

Upload

Open the Data Tracks

Configure Region Image | Configure Overview Image | Configure Chromosome Image | Personal Data

Active tracks

Find a track

Your data (5/5)

- Custom zm_16_DAP_endosperm.bed
- Custom zm_anther.bed
- Custom zm_v1-leaf.bed
- Custom zm_v1-root.bed
- Custom zm_16_DAP_embryo.bed

Sequence and assembly

- Contigs
- Sequence

Genes and transcripts

- Gramene Genes

Information and decorations

- Chromosome bands
- Scale bar
- Ruler
- Gene Legend
- Display in-track labels
- Information

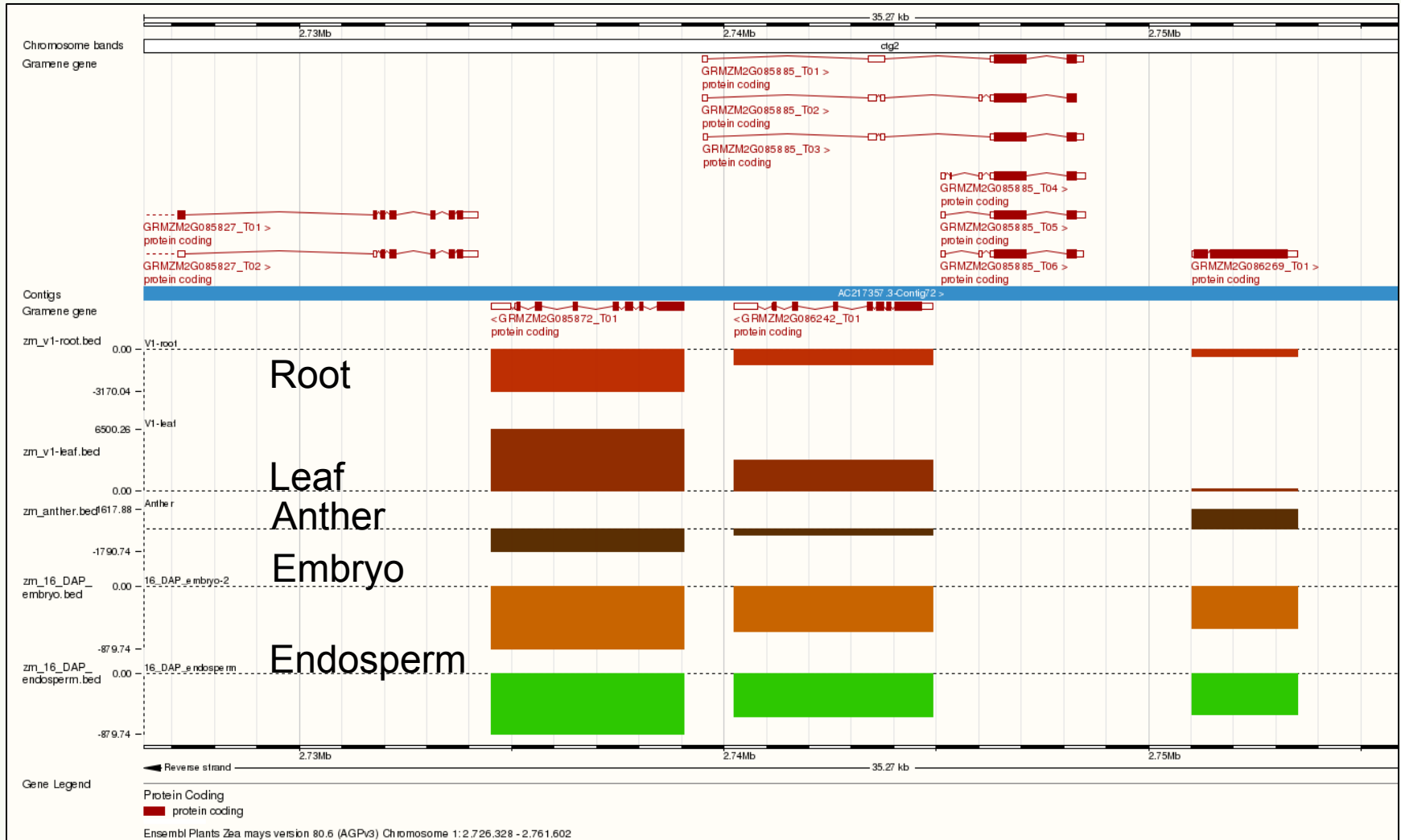
Key

- Track style
- External External data
- Forward strand
- Custom User-added track

Left Sidebar:

- Active tracks
- Favourite tracks
- Track order
- Search results
- Your data (5/5)
- Sequence and assembly (2/9)
 - Sequence (2/4)
 - GRC alignments (0/5)
- Genes and transcripts (1/4)
 - Genes (1/3)
 - Prediction transcripts (0/1)
- mRNA and protein alignments (0/9)
- Variation (0/3)
- Regulation (0/8)
- Comparative genomics (0/21)
- Oligo probes (0/1)
- Repeat regions (0/19)
- Information and decorations (6/11)
- Display options
- Save configuration as...
- Load configuration
- Reset configuration
- Reset track order
- Add your data

Uploaded Expression Data on Maize Genome Browser



NGS Data Upload

EXAMPLE-2

RNA-Seq read alignments: BAM format

The screenshot shows the Genomescope web interface. The sidebar on the left contains the following links: Login, Register, Lost Password, Custom Data (expanded), Add your data, Attach DAS, Manage Data, Features on Karyotype, Manage Configurations (expanded), Configurations for this page, All configurations, Configuration sets, Online Tools (expanded), Variant Effect Predictor, Assembly Converter, ID History Converter, Region Report, Data Slicer, and Help. The main content area has tabs for 'Configure Region Image', 'Configure Overview Image', 'Configure Chromosome Image', and 'Personal Data'. Under the 'Personal Data' tab, the 'Add a custom track' section contains the following fields: 'Name for this data (optional):' with the value 'My BAM file', 'Species:' with the value 'Oryza sativa Japonica', 'Assembly:' with the value 'IRGSP-1.0', and 'Data format:' with the value 'BAM'. Below these fields is a link 'Help on supported formats, display types, etc.' and a 'Provide file URL:' field with the value '0B-50D4BFBF00F6/PS5hr3_s.bam'. An 'Attach' button is located below the URL field. Four red arrows point to the 'Add your data' link, the 'Name for this data' field, the 'Species' dropdown, and the 'Provide file URL' field.

Configure Region Image | Configure Overview Image | Configure Chromosome Image | Personal Data

Login
Register
Lost Password
Custom Data
 Add your data
 Attach DAS
 Manage Data
 Features on Karyotype
Manage Configurations
 Configurations for this page
 All configurations
 Configuration sets
Online Tools
 Variant Effect Predictor
 Assembly Converter
 ID History Converter
 Region Report
 Data Slicer
Help

Add a custom track

Name for this data (optional):

Species:

Assembly:

Data format:

[Help on supported formats, display types, etc.](#)

Provide file URL:

Example data:

http://de.iplantcollaborative.org/dl/d/67DEFD9E-84EF-43F8-BB0B-50D4BFBF00F6/PS5hr3_s.bam

RNA-Seq read alignments: BAM format

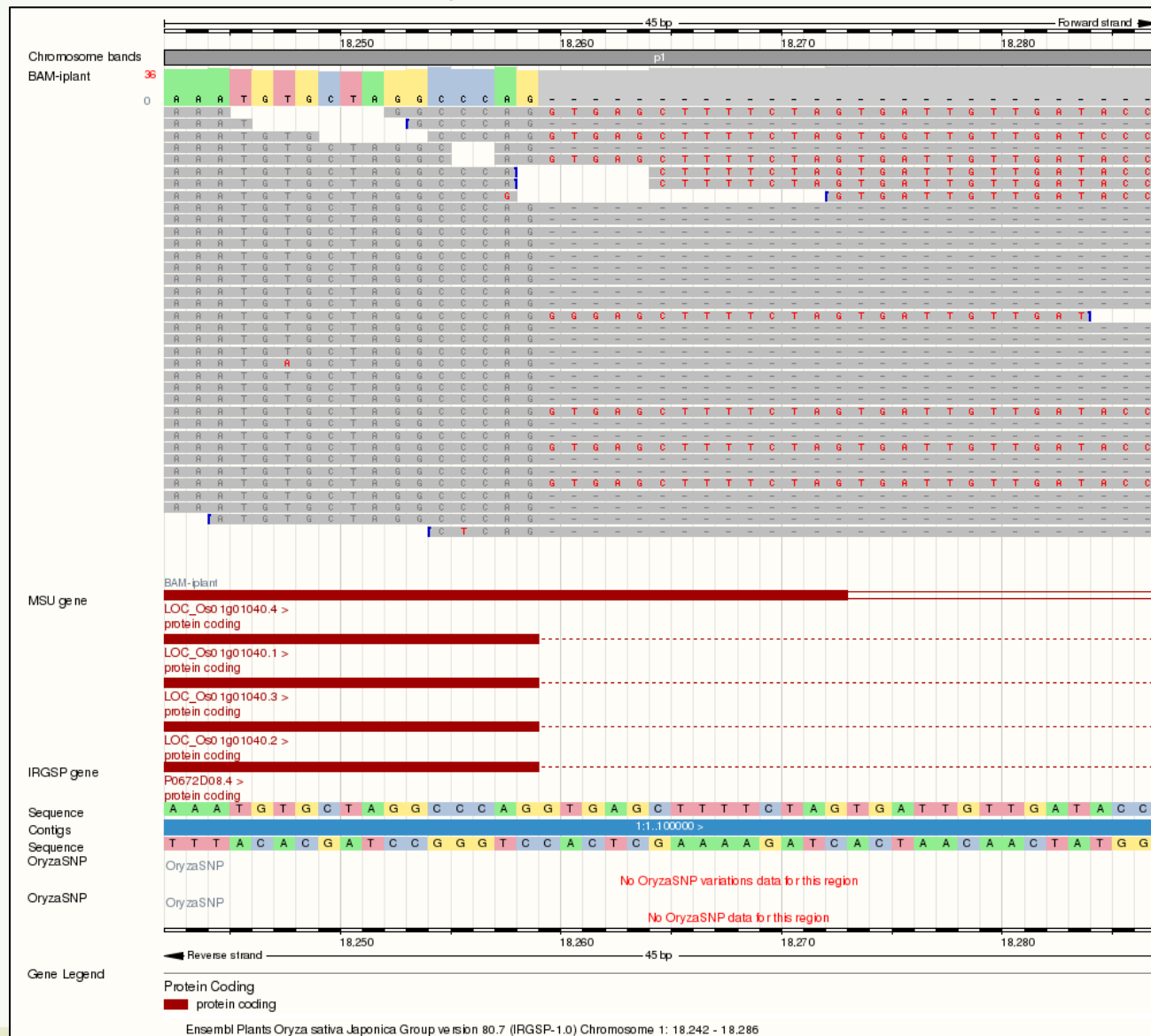
Example data: http://de.iplantcollaborative.org/dl/d/67DEFD9E-84EF-43F8-BB0B-50D4BFB500F6/PS5hr3_s.bam



Options to
navigate or
zoom-in/out

RNA-Seq read alignments: BAM format

Example data: http://de.iplantcollaborative.org/dl/d/67DEFD9E-84EF-43F8-BB0B-50D4BFBF00F6/PS5hr3_s.bam



Sequenced/Non-sequenced Mapping/Alignment Data Upload

EXAMPLE-3

Feature Mapping/Alignment: GFF format

Example: http://de.iplantcollaborative.org/dl/d/1D05C5A5-5C1F-48F5-AC12-4D38C8958B61/Salk_T-DNA.gff

#Data source: http://natural.salk.edu/database/tdnaexpress/T-DNA.SALK								
#Arabidopsis thaliana TAIR10 assembly								
chr1	SALK	T-DNA	1	303	e-168	+	.	ID=SALK_001127.50.70.x
chr1	SALK	T-DNA	1	303	e-171	+	.	ID=SALK_001127.52.15.x
chr1	SALK	T-DNA	32	303	e-152	+	.	ID=SALK_001127.52.15.x
chr1	SALK	T-DNA	1066	1097	7e-07	-	.	ID=SALK_082138.17.20.x
chr1	SALK	T-DNA	1066	1097	6e-07	+	.	ID=SALK_114475.16.50.x
chr1	SALK	T-DNA	1067	1093	3e-06	-	.	ID=SALK_065399.25.40.x
chr1	SALK	T-DNA	1073	1097	6e-05	-	.	ID=SALK_117416.15.55.n
chr1	SALK	T-DNA	1075	1099	6e-05	-	.	ID=SALK_132061.15.90.x
chr1	SALK	T-DNA	1076	1100	6e-05	-	.	ID=SALK_117013.15.75.n
chr1	SALK	T-DNA	1676	2070	0.0	-	.	ID=SALK_047276.52.80.x
chr1	SALK	T-DNA	2461	2688	4e-50	+	.	ID=SALK_005597.33.20.x
chr1	SALK	T-DNA	4101	4298	e-108	-	.	ID=SALK_074723.37.35.x
chr1	SALK	T-DNA	4101	4298	e-104	-	.	ID=SALK_128569.34.95.x
chr1	SALK	T-DNA	4101	4298	e-104	-	.	ID=SALK_128571.33.50.x

Personal Data

Login

Register

Lost Password

Custom Data

Add your data

Attach DAS

Manage Data

Features on Karyotype

Manage Configurations

Configurations for this page

All configurations

Configuration sets

Online Tools

Variant Effect Predictor

Assembly Converter

ID History Converter

Data Slicer

Help

Add a custom track

Name for this data (optional):

Test-SALK-T-DNA

Species:

Arabidopsis thaliana

Assembly:

TAIR10

Data format:

GFF

Type:

☐ Upload data (max 20MB)

☒ Attach via URL

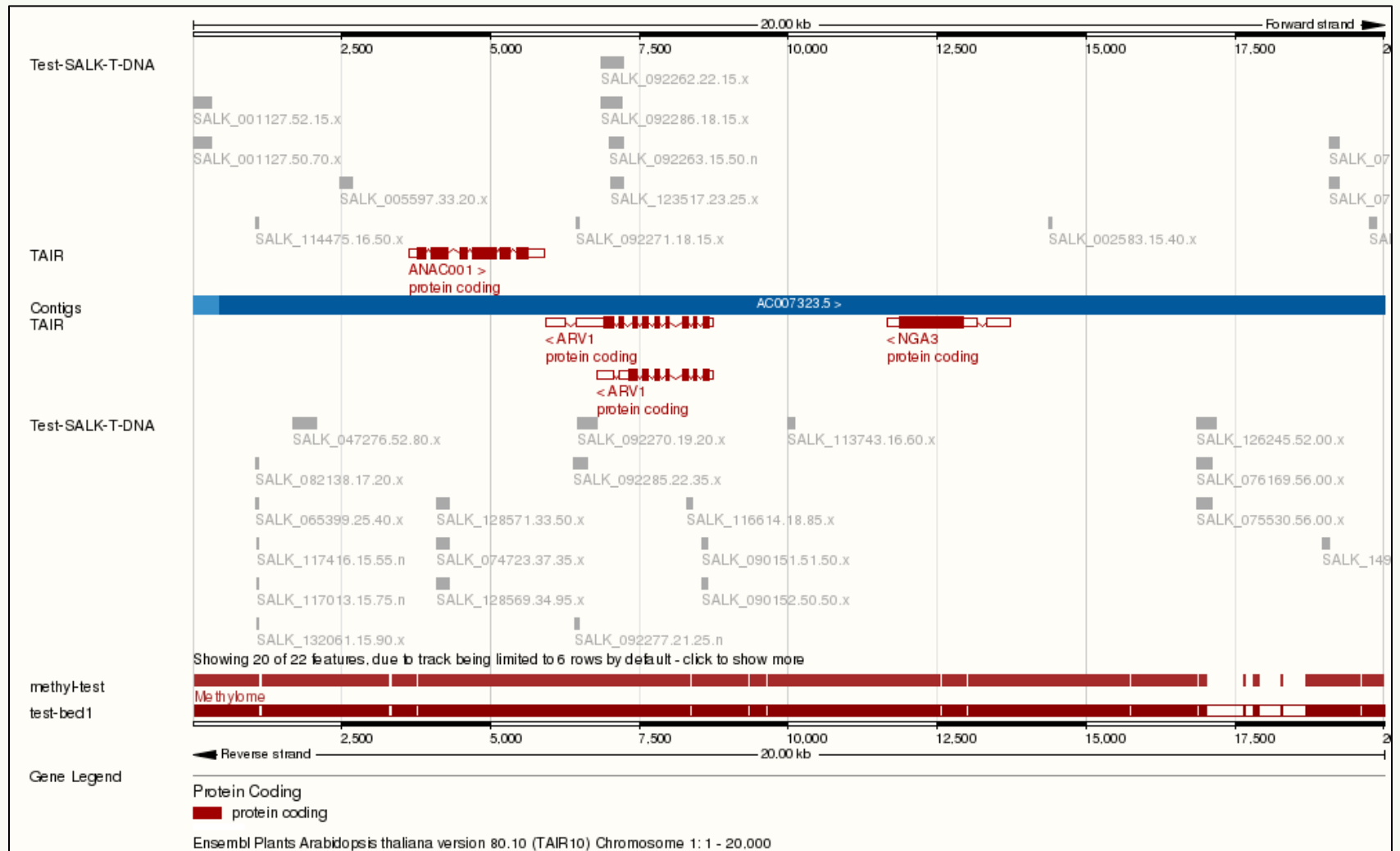
Provide file URL:

12-4D38C8958B61/Salk_T-DNA.gff

Attach

Sequenced Feature Alignments: GFF format

Example: http://de.iplantcollaborative.org/dl/d/1D05C5A5-5C1F-48F5-AC12-4D38C8958B61/Salk_T-DNA.gff



Manage your data on Gramene

Personal Data

- Login
- Register
- Lost Password
- Custom Data
 - Add your data
 - Attach DAS
 - Manage Data**
 - Features on Karyotype
- Manage Configurations
 - Configurations for this page
 - All configurations
 - Configuration sets
- Online Tools
 - Variant Effect Predictor
 - Assembly Converter
 - ID History Converter
 - Data Slicer
- Help

Help

You can rename your uploads and attached URLs by clicking on their current name in the Source column

[Help on supported formats, display types, etc](#)

Your data

Show **All** entries

Type	Source	Species	Assembly	Last updated	Actions
URL	Test-SALK-T-DNA View sample location http://de.iplantcollaborative.org/dl/d/1D05C5A5-5C1F-48F5-AC12-4D38C8958B61/Salk_T-DNA.gff	<i>Arabidopsis thaliana</i>	TAIR10	21/06/15 at 00:57	
URL	Mt-RNAseq-BAM http://de.iplantcollaborative.org/dl/d/67DEFD9E-84EF-43F8-BB0B-50D4BFBF00F6/PS5hr3_s.bam	<i>Zea mays</i>	AGPv3	21/06/15 at 00:42	
Upload	zm_v1-root.bed View sample location BED file	<i>Zea mays</i>	AGPv3	20/06/15 at 23:54	
Upload	zm_anther.bed View sample location BED file	<i>Zea mays</i>	AGPv3	20/06/15 at 23:49	
Upload	zm_v1-leaf.bed View sample location BED file	<i>Zea mays</i>	AGPv3	20/06/15 at 23:47	
Upload	zm_16_DAP_endosperm.bed View sample location BED file	<i>Zea mays</i>	AGPv3	20/06/15 at 23:43	
Upload	zm_16_DAP_embryo.bed View sample location	<i>Zea mays</i>	AGPv3	20/06/15 at 23:37	

Basic steps for data upload and visualization

1. Prepare your data files
 - a. More info at <http://ensembl.gramene.org/info/website/upload/index.html>
2. Visit the Gramene database home (www.gramene.org)
3. Go to the genome section
4. Select species
 - a. On species home page click on 'Display your own data' OR
 - b. Go to the genome region and click on 'Add your data'
5. Select options from upload window
6. Once the data is loaded, go to the desired region of the genome of the respective species genome to view



BLAST

BioMart

Tools

Downloads

Help

Feedback

For Questions and Feedback, visit
<http://tools.gramene.org/feedback>



Search

Home

Release Notes

Collaborators

Contact

About

Archive

Citing Gramene

Videotutorials

Outreach

For previous webinar recordings and
video tutorials, visit
<http://tinyurl.com/gramene-tutorial>

July 14, 2015

Next Gramene Webinar

SNP data
(Explore, Upload & Analyze)

Acknowledgements



**Funded by the National Science Foundation's
Plant Genome Research Program award IOS#1127112**

Doreen Ware, PI (USDA ARS, CSHL)

Kapeel Cougule, Yinping Jiao, Sunita Kumari, Marcela K. Monaco, Andrew Olson, Joshua Stein, Jim Thomason, Peter van Buren, Sharon Wei, Bo Wang

Pankaj Jaiswal, Co-PI (OSU)

Vindhya Amarasinghe, Justin Elser, Sushma Naithani, Justin Preece

Paul Kersey / Robert Petrysyk (EMBL-EBI)

Dan Bolser, Arnaud Kerhornou, Eugene Kulesha, Dan Staines, Brandon Walts / Elisabet Barrera, Maria Keays, Oliver Mannion

Lincoln Stein (OICR)

Robin Haw; Peter D' Eustachio (NYU); Guanming Wu; David Croft (EBI)

Crispin Taylor (ASPB)

Patty Lockhart

