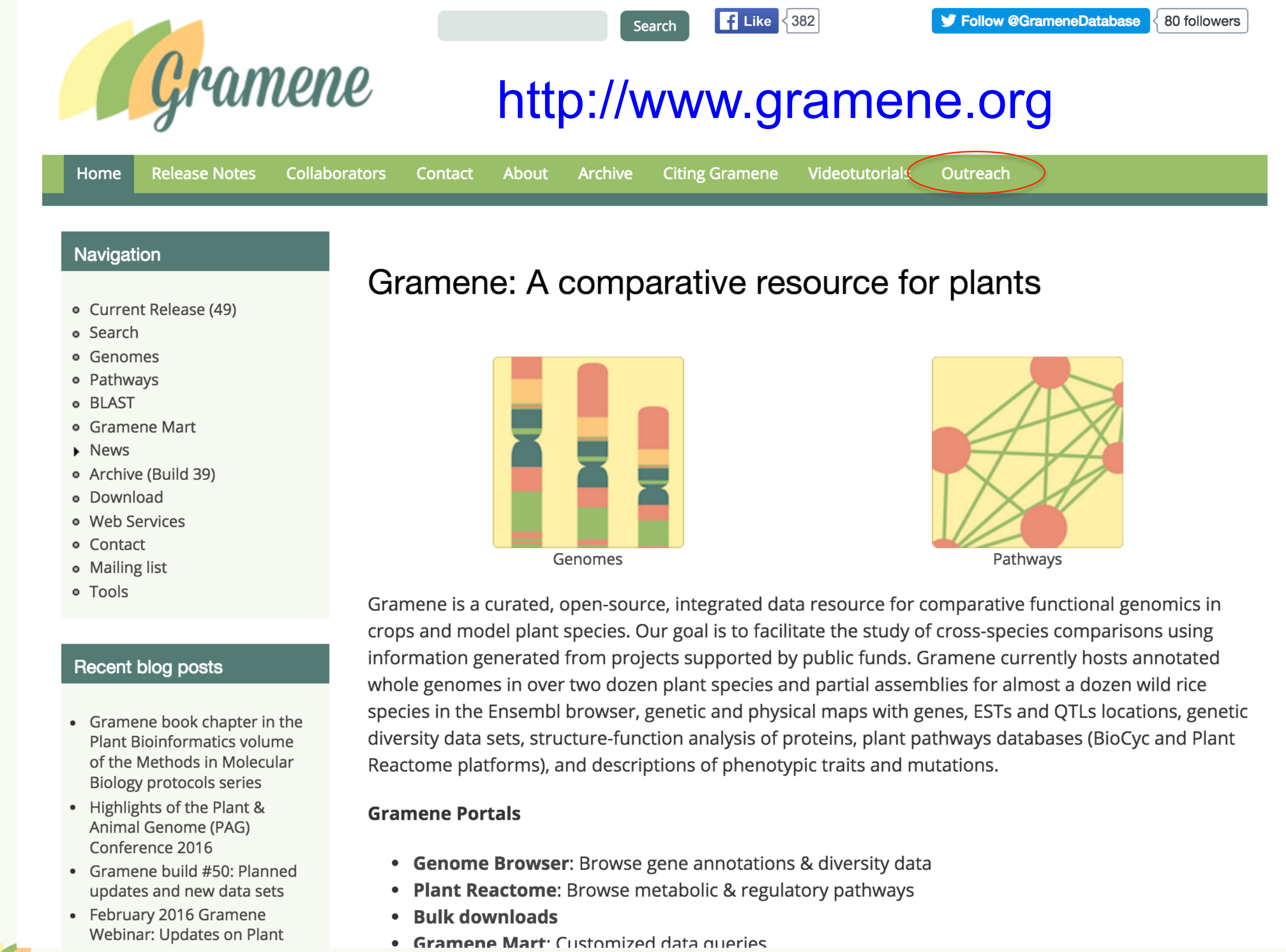
Gramene—Comparative Plant Genomics & Pathway Resources

www.gramene.org

Doreen Ware & Yinping Jiao
USDA ARS
Cold Spring Harbor Laboratory
March 17, 2016

Outline

- Background & new development
- Ensembl Browser: Genomes & Trees
- Reactome Browser: Pathways
  - Video
- Maize RefGen V4 Draft Browser
- Sample Exercises on your own



http://www.gramene.org

Navigation

- Current Release (49)
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Mailing list
- Tools

Recent blog posts

- Gramene book chapter in the Plant Bioinformatics volume of the Methods in Molecular Biology protocols series
- Highlights of the Plant & Animal Genome (PAG) Conference 2016
- Gramene build #50: Planned updates and new data sets
- February 2016 Gramene Webinar: Updates on Plant

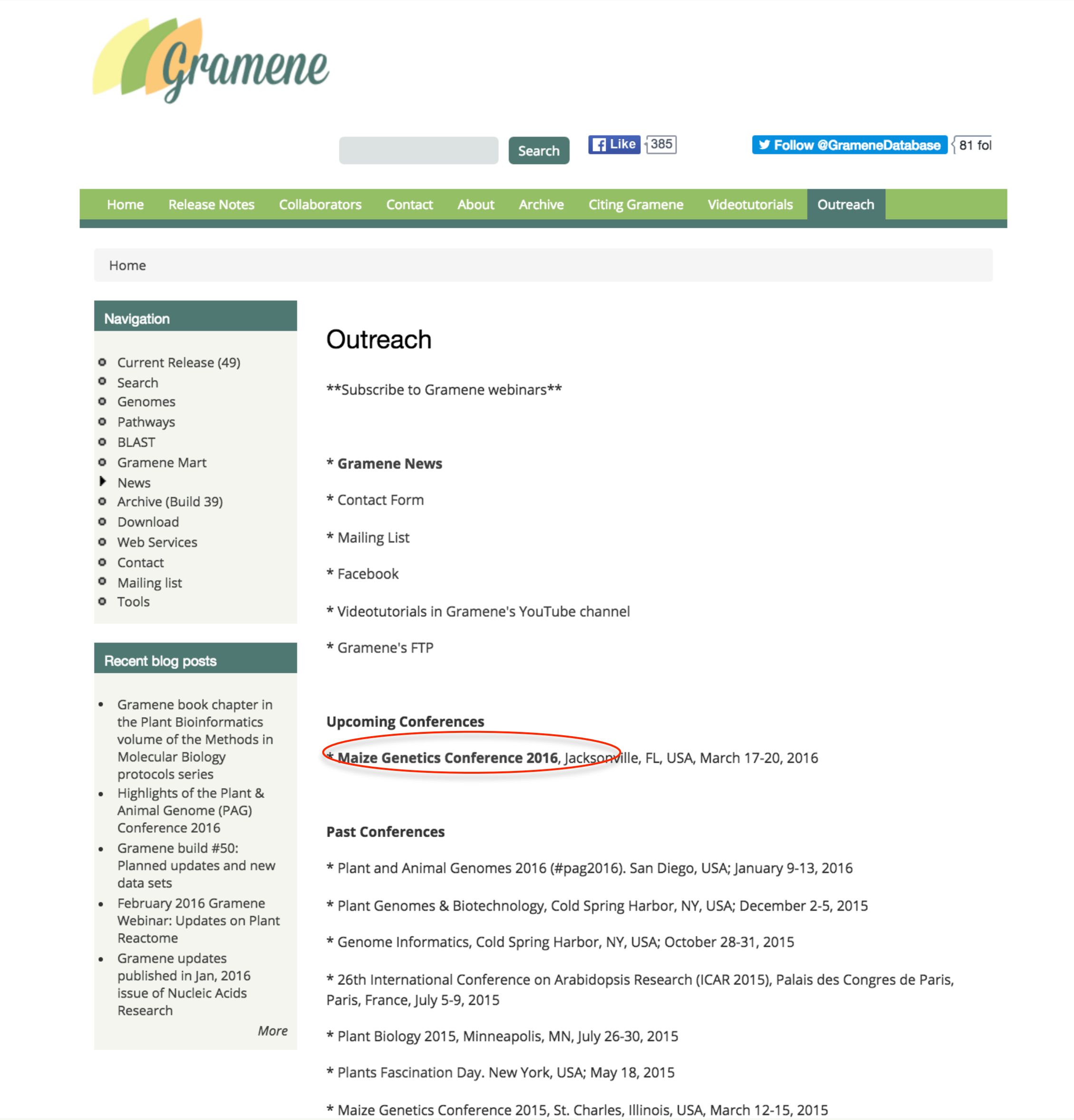
Gramene: A comparative resource for plants

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.

Gramene Portals

- **Genome Browser**: Browse gene annotations & diversity data
- **Plant Reactome**: Browse metabolic & regulatory pathways
- **Bulk downloads**
- **Gramene Mart**: Customized data queries

Outreach



- Gramene book chapter in the Plant Bioinformatics volume of the Methods in Molecular Biology protocols series
- Highlights of the Plant & Animal Genome (PAG) Conference 2016
- Gramene build #50: Planned updates and new data sets
- February 2016 Gramene Webinar: Updates on Plant Reactome
- Gramene updates published in Jan, 2016 issue of Nucleic Acids Research

Outreach

Subscribe to Gramene webinars

* **Gramene News**

* Contact Form

* Mailing List

* Facebook

* Videotutorials in Gramene's YouTube channel

* Gramene's FTP

Upcoming Conferences

* **Maize Genetics Conference 2016**, Jacksonville, FL, USA, March 17-20, 2016

Past Conferences

* Plant and Animal Genomes 2016 (#pag2016). San Diego, USA; January 9-13, 2016

* Plant Genomes & Biotechnology, Cold Spring Harbor, NY, USA; December 2-5, 2015

* Genome Informatics, Cold Spring Harbor, NY, USA; October 28-31, 2015

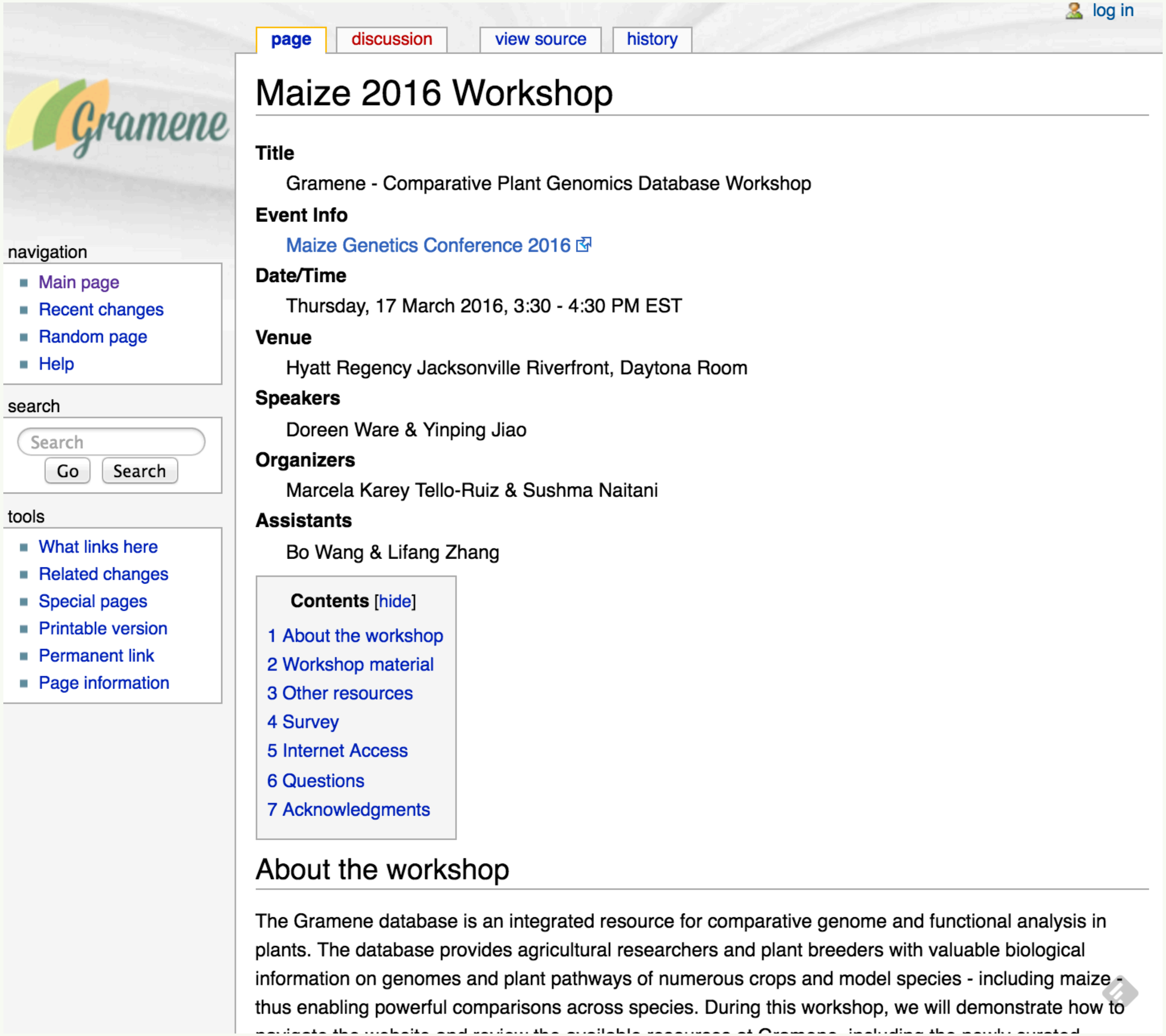
* 26th International Conference on Arabidopsis Research (ICAR 2015), Palais des Congres de Paris, Paris, France, July 5-9, 2015

* Plant Biology 2015, Minneapolis, MN, July 26-30, 2015

* Plants Fascination Day. New York, USA; May 18, 2015

* Maize Genetics Conference 2015, St. Charles, Illinois, USA, March 12-15, 2015

Outreach Wiki



Maize 2016 Workshop

Title
Gramene - Comparative Plant Genomics Database Workshop

Event Info
Maize Genetics Conference 2016

Date/Time
Thursday, 17 March 2016, 3:30 - 4:30 PM EST

Venue
Hyatt Regency Jacksonville Riverfront, Daytona Room

Speakers
Doreen Ware & Yinping Jiao

Organizers
Marcela Karey Tello-Ruiz & Sushma Naitani

Assistants
Bo Wang & Lifang Zhang

Contents [hide]

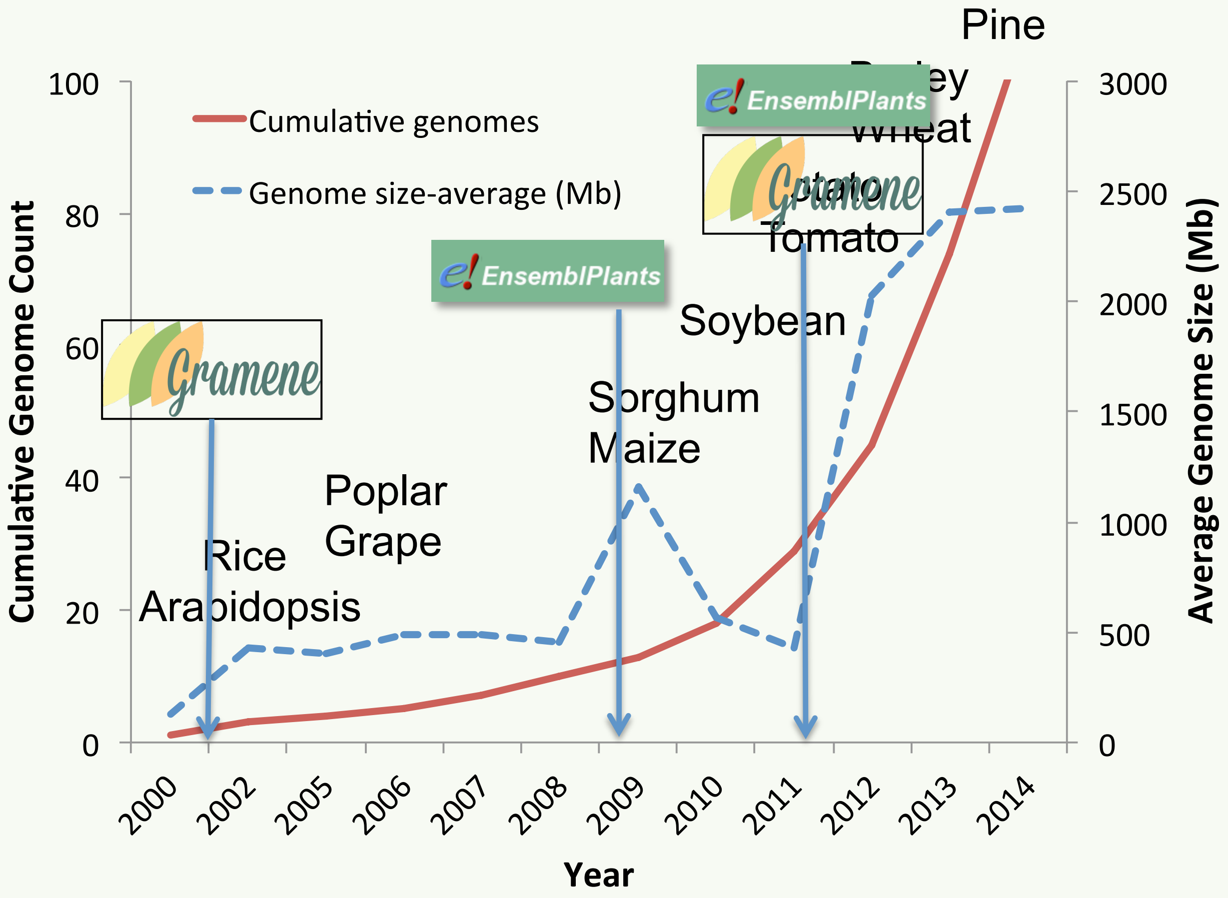
About the workshop

The Gramene database is an integrated resource for comparative genome and functional analysis in plants. The database provides agricultural researchers and plant breeders with valuable biological information on genomes and plant pathways of numerous crops and model species - including maize - thus enabling powerful comparisons across species. During this workshop, we will demonstrate how to

Progress in Plant Genomics

- Adoption of the Ensembl Platform for Plants
- Gramene Microfinance: Rice a small genome a foundation for knowledge for the other grasses

100 published genomes



Adapted from Michael & VanBuren (2015) COPB

- Declining cost
- Declining quality

Variation
Expression
ENCODE-like
Metabolomics
QTL & GWAS

sol genomics network

The Banana Genome Hub

tair

Cacao Genome Database

Arizona Genomics Institute
AGI

Centre National de Séquençage
GENOSCOPE

INTERNATIONAL
BARLEY
SEQUENCING
CONSORTIUM

www.panzea.org

phytozome

Whole is greater than the sum of its parts

SoyBase

University of Padua
CRIBI Genomics

华大基因
BGI

Brachypodium.org

The Oryza Map Alignment Project
OMAP

J. Craig Venter™
INSTITUTE

Bolbase

Amborella

Medtr5g024510.1
endoribonuclease/kinase IRE1-
Medtr5g024510.2
endoribonuclease/kinase IRE1-
Medtr5g024510.3
Medicago truncatula
Genome Project

IWGSC
International
Wheat Genome
Sequencing
Consortium

maizeGDB
MAIZE GENETICS AND GENOMICS DATABASE

Gramene Adds Value to Plant Genomes

Accessioned Genomes + Community Annotation + Variation

Annotation Pipelines

- Repeats/TE's
- Genes
- EST/cDNA
- InterPro domain
- Gene Ontology (GO)
- Variant Effect Prediction

Comparative Analysis

- Whole Genome Alignment
- Phylogenetic Gene Trees
- Ortholog/Paralog calling
- Synteny mapping

Pathway Curation & Projection

BLAST

Programmatic Access:

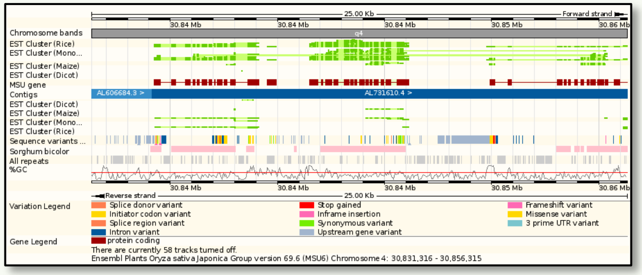
Gramene API

Ensembl API & RESTFUL interface

Reactome API & RESTFUL interface

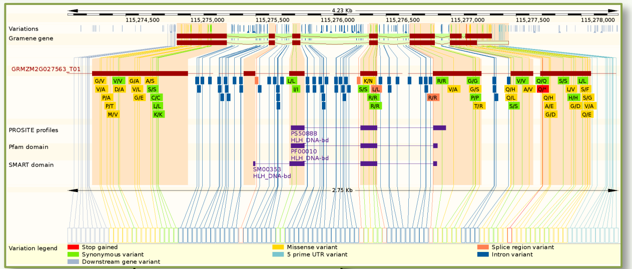
BioMart

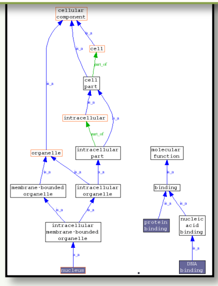
Public mysql server

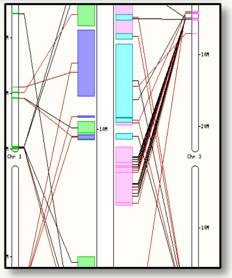


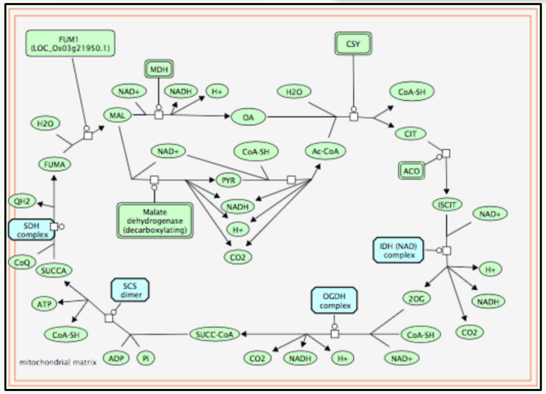
There are currently 58 tracks turned off.

Ensembl Plants Oryza sativa Japonica Group version 69.6 (MSU6) Chromosome 4: 30,831,316 - 30,856,315









Transnational collaboration

Doreen Ware

Paul Kersey

EMBL-EBI

Pankaj Jaiswal

Oregon State University

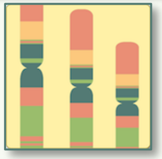
Robert Petryszak

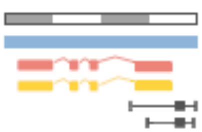
OICR Ontario Institute for Cancer Research

Lincoln Stein

Genomes

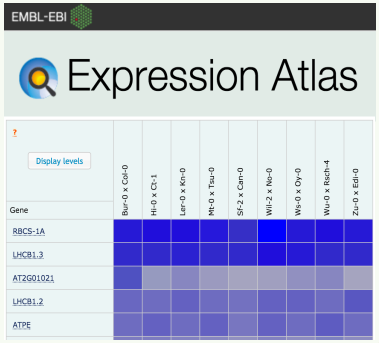
EnsemblPlants





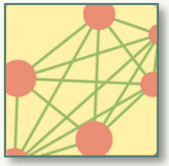
ATCGAGCT
ATCCAGCT
ATCGAGAT

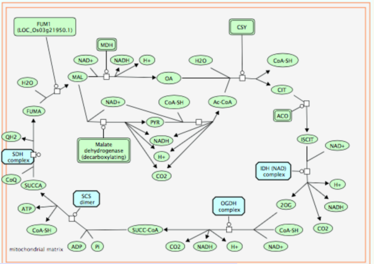
Expression



Pathways

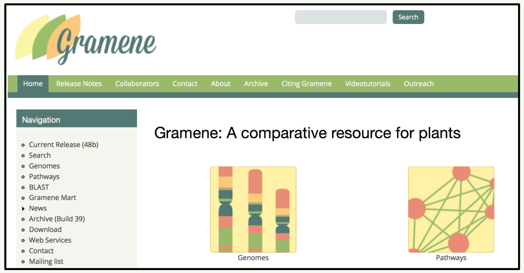
Plant Reactome





Reactome

2.5 Month Coordinated Release Cycle



Navigation

- Current Release (48b)
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Mailing list

Gramene: A comparative resource for plants

Plant Reactome

EnsemblPlants

Data Acquisition & Preparation

- Genome assembly & annotation
- Variation
- Other

Declaration of intentions

Create New / Update Genome DBs

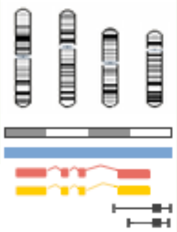
- Analyses (*e.g.* InterProScan)
- Variant Effect Predictor
- Cross-reference other databases

~1 week

Release!

Comparative Analyses ~1 month

- Whole-Genome Alignments
- Phylogenetic Gene Trees

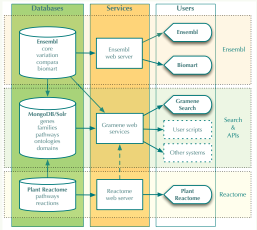


Gramene releases:

- 2015 June, Aug & Nov; 2016 Jan & Mar

Plant Reactome updates:

- 2015 Aug & Nov; 2016 Mar



- QA/QC: automated and manual testing
- Generate release notes
- Public outreach

1-2 weeks

Update GUIs & Tools

- Software
- Configure Genome Browser & BLAST
- Rebuild Search indexes
- Build Marts for data mining
- Update FTP content
- Update Website

~1-2 weeks

Pathway Curation & Projection

1-2 months

- Process newly curated *O. sativa Japonica* pathways
- Project pathways to other species using inferred homology

biomart

BLAST
FTP

Ensembl EMBL-EBI

Platform development

REACTOME
A CURATED PATHWAY DATABASE

OICR
Ontario Institute for Cancer Research

Gramene

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**, Jim Thomason, **Bo Wang**, **Sharon Wei**, Peter van Buren, **Marcela K. Tello-Ruiz** (Project Coordinator)

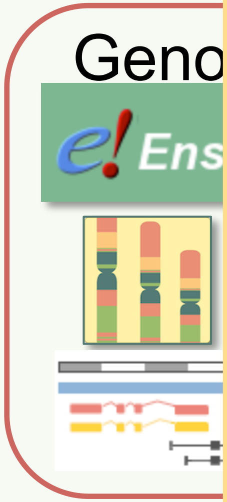
Pankaj Jaiswal, Co-PI (OSU)
Noor Al-Bader, Justin Elser, Matthew Geniza, Parul Gupta, **Justin Preece**, **Sushma Naithani**

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)

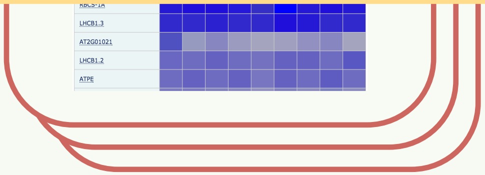
What is in Gramene?

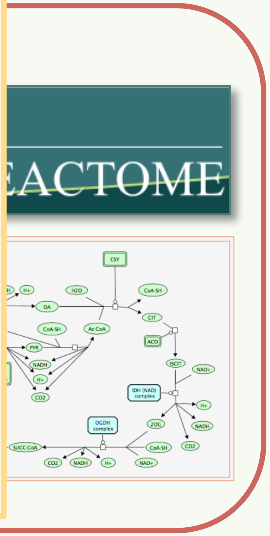


Best-in-class tools

but **siloed**

How to integrate?





Goals of Search Interface

- Search all Gramene data
- Enable powerful, expressive searches
- Provide useful summaries and visualizations
- REALLY FAST

- Above all, SIMPLICITY

Composed Query

racea...

rapa...

dopsis thaliana 1

abidopsis lyrata... 1

Theobroma cacao 0

Glycine max 0

Medicago... 1

Populus trichocarpa 0

Prunus persica 0

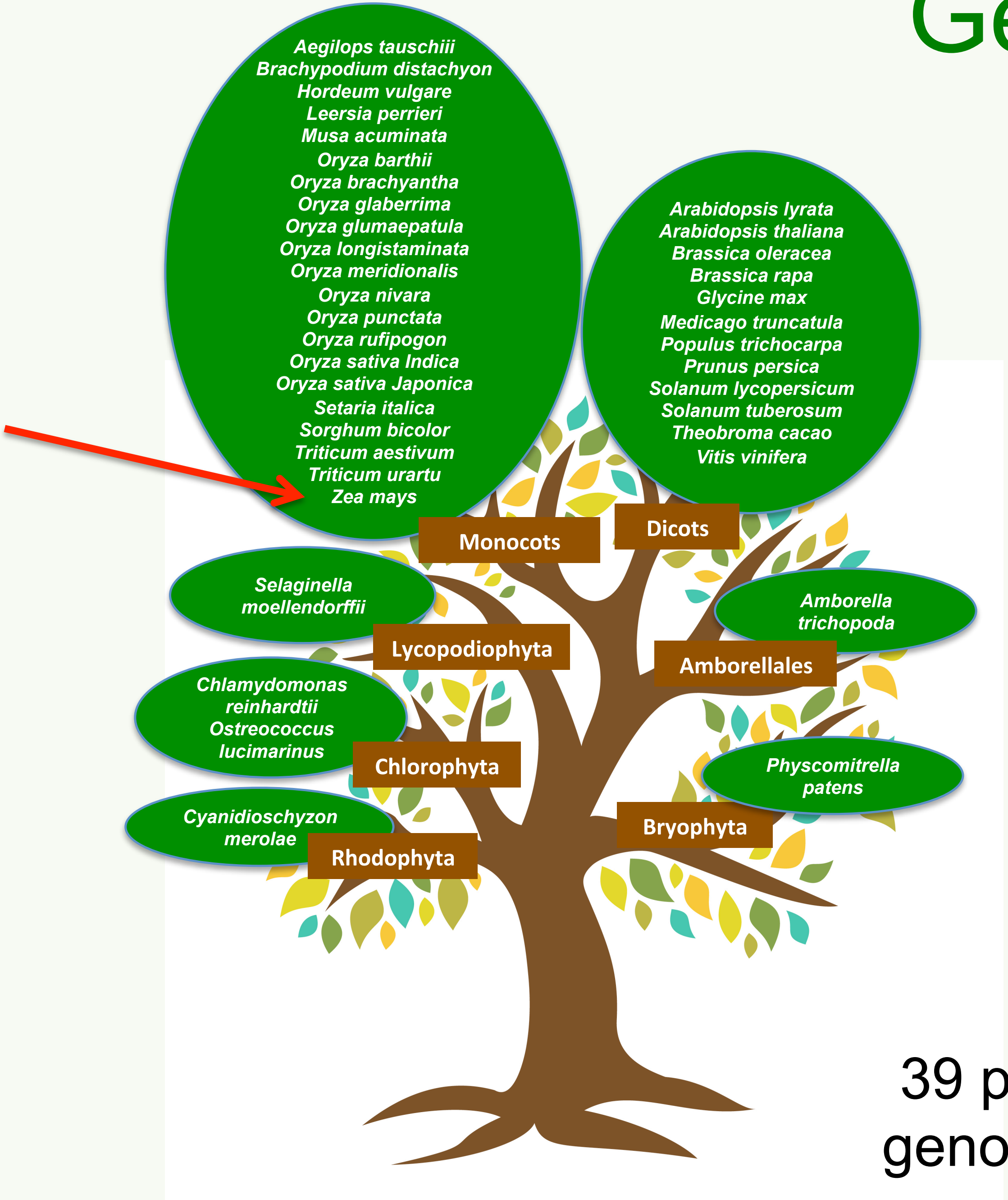
Vitis vinifera 1

Solanum tuberosum 1

Beta testers wanted!

Sign up at gramene.org/beta

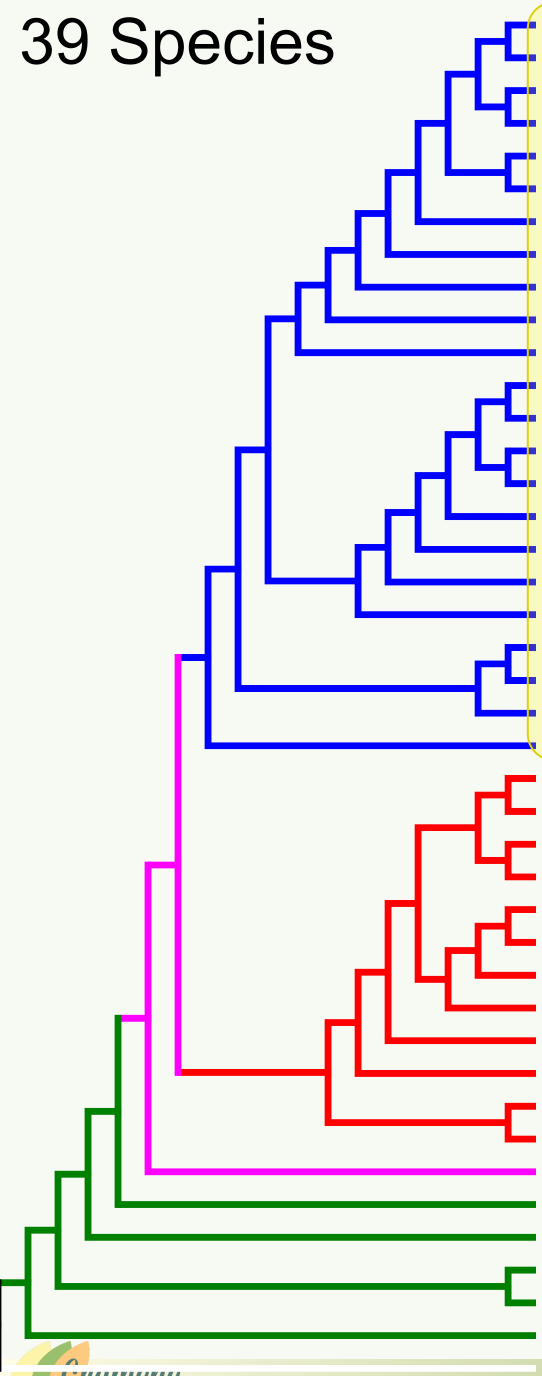
Genomes



39 plant reference genomes in build 49

Gramene Genome Portfolio

39 Species



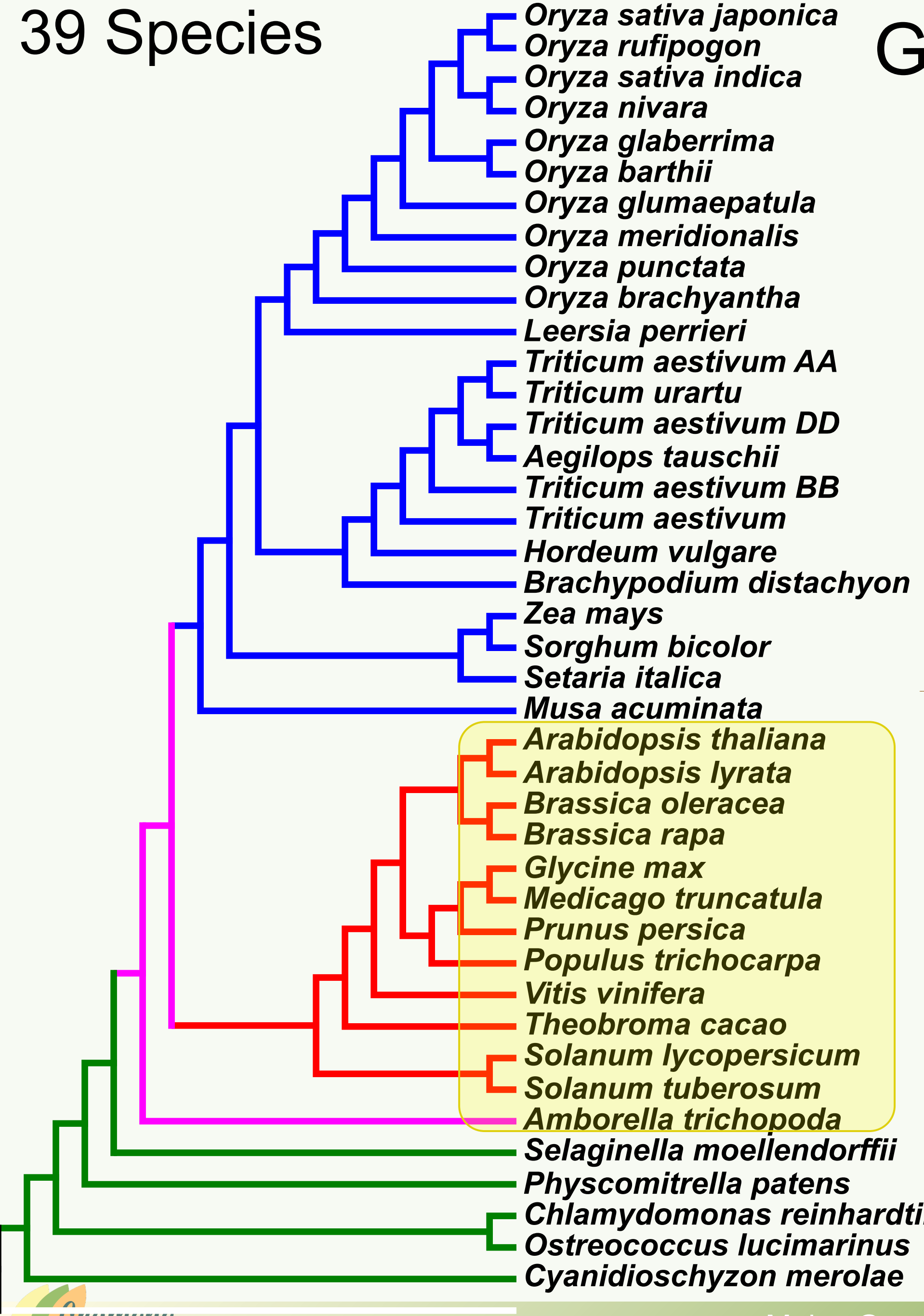
Oryza sativa japonica
Oryza rufipogon
Oryza sativa indica
Oryza nivara
Oryza glaberrima
Oryza barthii
Oryza glumaepatula
Oryza meridionalis
Oryza punctata
Oryza brachyantha
Leersia perrieri
Triticum aestivum AA
Triticum urartu
Triticum aestivum DD
Aegilops tauschii
Triticum aestivum BB
Triticum aestivum
Hordeum vulgare
Brachypodium distachyon
Zea mays
Sorghum bicolor
Setaria italica
Musa acuminata
Arabidopsis thaliana
Arabidopsis lyrata
Brassica oleracea
Brassica rapa
Glycine max
Medicago truncatula
Prunus persica
Populus trichocarpa
Vitis vinifera
Theobroma cacao
Solanum lycopersicum
Solanum tuberosum
Amborella trichopoda
Selaginella moellendorffii
Physcomitrella patens
Chlamydomonas reinhardtii
Ostreococcus lucimarinus
Cyanidioschyzon merolae

21 Monocots

Models
Crops
Progenitors

Gramene Genome Portfolio

39 Species



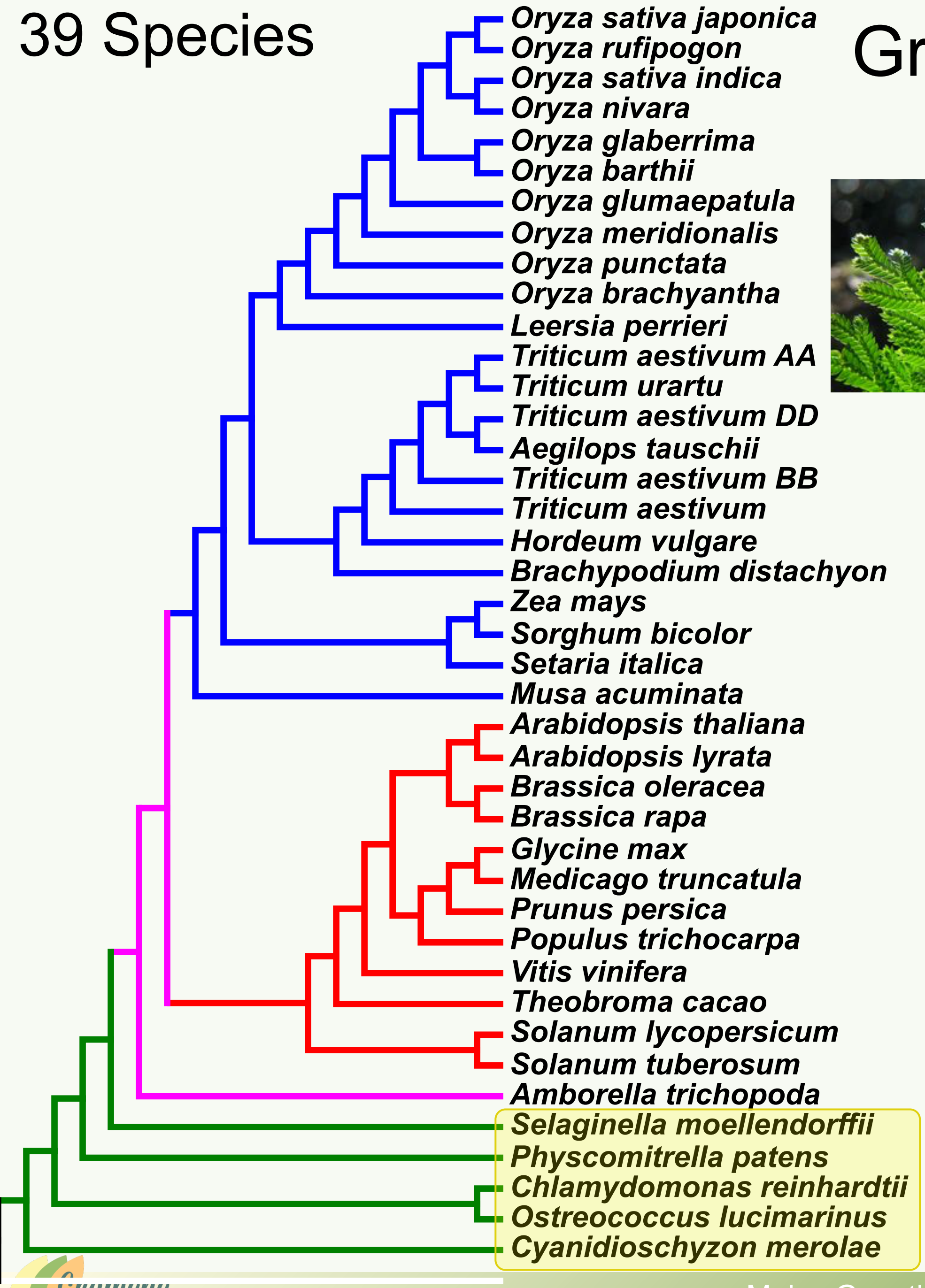
Oryza sativa japonica
Oryza rufipogon
Oryza sativa indica
Oryza nivara
Oryza glaberrima
Oryza barthii
Oryza glumaepatula
Oryza meridionalis
Oryza punctata
Oryza brachyantha
Leersia perrieri
Triticum aestivum AA
Triticum urartu
Triticum aestivum DD
Aegilops tauschii
Triticum aestivum BB
Triticum aestivum
Hordeum vulgare
Brachypodium distachyon
Zea mays
Sorghum bicolor
Setaria italica
Musa acuminata
Arabidopsis thaliana
Arabidopsis lyrata
Brassica oleracea
Brassica rapa
Glycine max
Medicago truncatula
Prunus persica
Populus trichocarpa
Vitis vinifera
Theobroma cacao
Solanum lycopersicum
Solanum tuberosum
Amborella trichopoda
Selaginella moellendorffii
Physcomitrella patens
Chlamydomonas reinhardtii
Ostreococcus lucimarinus
Cyanidioschyzon merolae

12 Core Eudicots

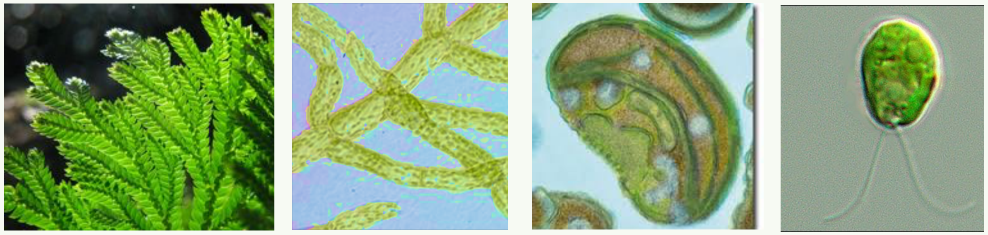
1 Basal angiosperm

Gramene Genome Portfolio

39 Species

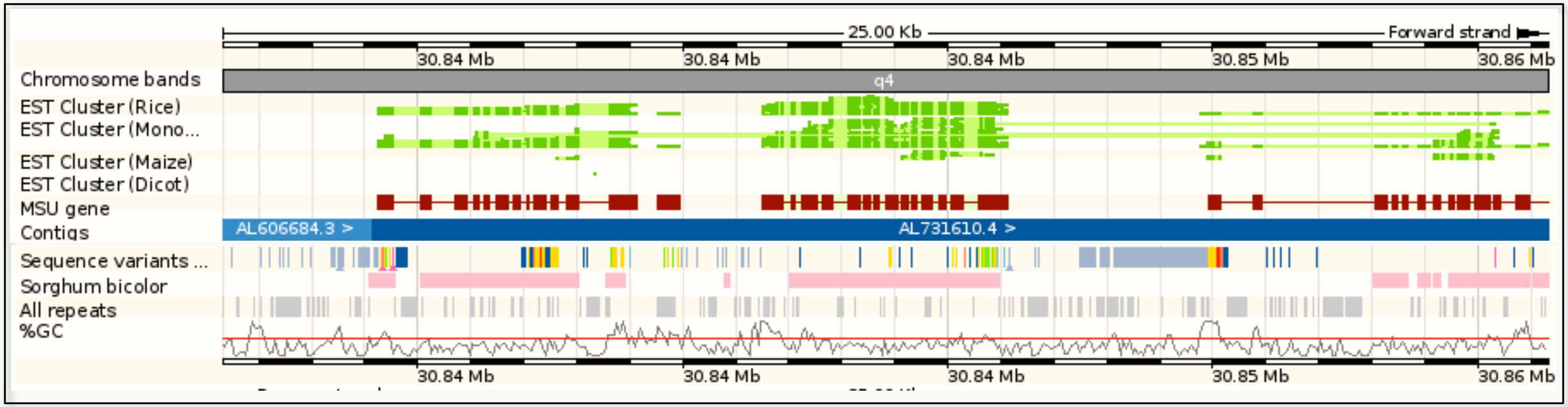


Oryza sativa japonica
Oryza rufipogon
Oryza sativa indica
Oryza nivara
Oryza glaberrima
Oryza barthii
Oryza glumaepatula
Oryza meridionalis
Oryza punctata
Oryza brachyantha
Leersia perrieri
Triticum aestivum AA
Triticum urartu
Triticum aestivum DD
Aegilops tauschii
Triticum aestivum BB
Triticum aestivum
Hordeum vulgare
Brachypodium distachyon
Zea mays
Sorghum bicolor
Setaria italica
Musa acuminata
Arabidopsis thaliana
Arabidopsis lyrata
Brassica oleracea
Brassica rapa
Glycine max
Medicago truncatula
Prunus persica
Populus trichocarpa
Vitis vinifera
Theobroma cacao
Solanum lycopersicum
Solanum tuberosum
Amborella trichopoda
Selaginella moellendorffii
Physcomitrella patens
Chlamydomonas reinhardtii
Ostreococcus lucimarinus
Cyanidioschyzon merolae



5 Lower plants

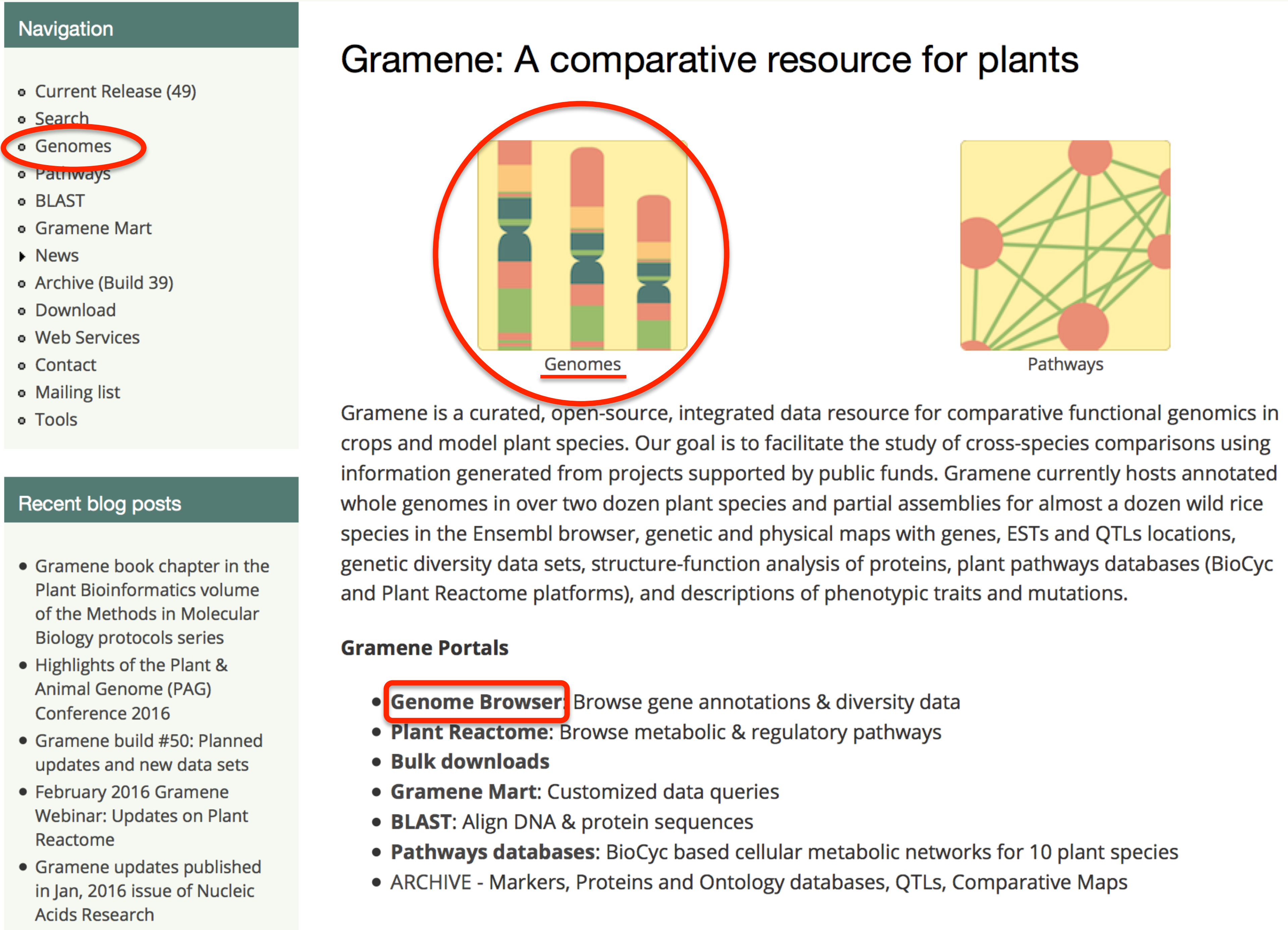
Genome Browsers



- Assembly structure and sequence
- Genes
- Expression
- Comparative alignments
- Genetic markers
- Variation
- Repeats & transposable elements
- Regulatory & Epigenetic marks

Baseline Annotation

- Transposon discovery
- Gene prediction
- EST Alignment



Navigation

- Current Release (49)
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Mailing list
- Tools

Recent blog posts

- Gramene book chapter in the Plant Bioinformatics volume of the Methods in Molecular Biology protocols series
- Highlights of the Plant & Animal Genome (PAG) Conference 2016
- Gramene build #50: Planned updates and new data sets
- February 2016 Gramene Webinar: Updates on Plant Reactome
- Gramene updates published in Jan, 2016 issue of Nucleic Acids Research

Gramene: A comparative resource for plants

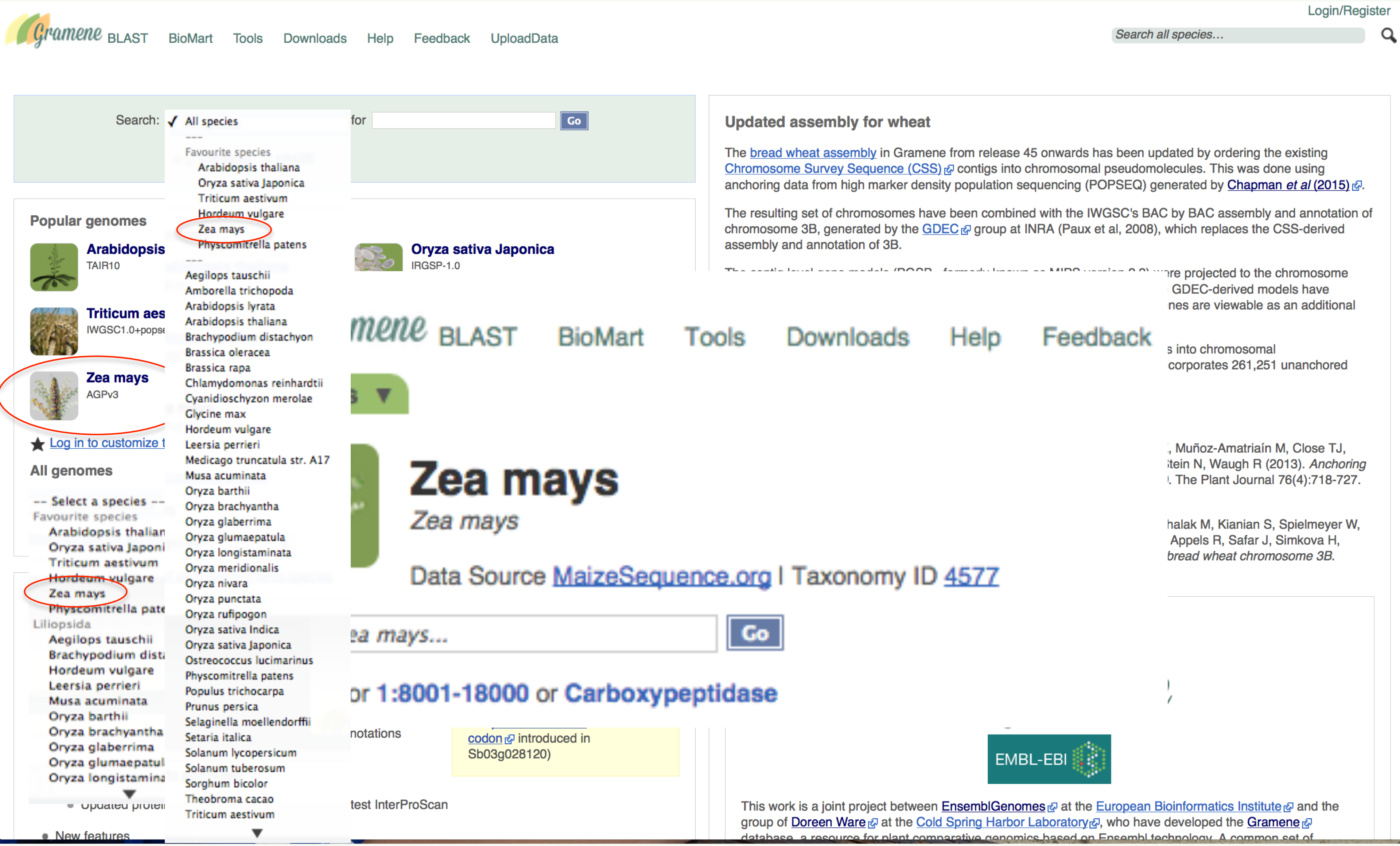
Gramene is a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species. Our goal is to facilitate the study of cross-species comparisons using information generated from projects supported by public funds. Gramene currently hosts annotated whole 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.

Gramene Portals

- **Genome Browser**: Browse gene annotations & diversity data
- **Plant Reactome**: Browse metabolic & regulatory pathways
- **Bulk downloads**
- **Gramene Mart**: Customized data queries
- **BLAST**: Align DNA & protein sequences
- **Pathways databases**: BioCyc based cellular metabolic networks for 10 plant species
- ARCHIVE - Markers, Proteins and Ontology databases, QTLs, Comparative Maps

Genomes

In collaboration with EnsemblGenomes



Gramene BLAST BioMart Tools Downloads Help Feedback UploadData

Login/Register

Search all species…

Search: ✓ All species for Go

Favourite species
- Arabidopsis thaliana
- Oryza sativa Japonica
- Triticum aestivum
- Hordeum vulgare
- Zea mays
- Physcomitrella patens

- Aegilops tauschii
- Amborella trichopoda
- Arabidopsis lyrata
- Arabidopsis thaliana
- Brachypodium distachyon
- Brassica oleracea
- Brassica rapa
- Chlamydomonas reinhardtii
- Cyanidioschyzon merolae
- Glycine max
- Hordeum vulgare
- Leersia perrieri
- Medicago truncatula str. A17
- Musa acuminata
- Oryza barthii
- Oryza brachyantha
- Oryza glaberrima
- Oryza glumaepatula
- Oryza longistaminata
- Oryza meridionalis
- Oryza nivara
- Oryza punctata
- Oryza rufipogon
- Oryza sativa Indica
- Oryza sativa Japonica
- Ostreococcus lucimarinus
- Physcomitrella patens
- Populus trichocarpa
- Prunus persica
- Selaginella moellendorffii
- Setaria italica
- Solanum lycopersicum
- Solanum tuberosum
- Sorghum bicolor
- Theobroma cacao
- Triticum aestivum

Popular genomes

Arabidopsis
TAIR10

Oryza sativa Japonica
IRGSP-1.0

Triticum aes
IWGSC1.0+popse

Zea mays
AGPv3

★ Log in to customize t

All genomes

-- Select a species --
Favourite species
- Arabidopsis thalian
- Oryza sativa Japoni
- Triticum aestivum
- Hordeum vulgare
- Zea mays
- Physcomitrella pate

Liliopsida
- Aegilops tauschii
- Brachypodium dist
- Hordeum vulgare
- Leersia perrieri
- Musa acuminata
- Oryza barthii
- Oryza brachyantha
- Oryza glaberrima
- Oryza glumaepatul
- Oryza longistamina

- Updated protei
- New features

Updated assembly for wheat

The bread wheat assembly in Gramene from release 45 onwards has been updated by ordering the existing Chromosome Survey Sequence (CSS) contigs into chromosomal pseudomolecules. This was done using anchoring data from high marker density population sequencing (POPSEQ) generated by Chapman *et al* (2015).

The resulting set of chromosomes have been combined with the IWGSC's BAC by BAC assembly and annotation of chromosome 3B, generated by the GDEC group at INRA (Paux et al, 2008), which replaces the CSS-derived assembly and annotation of 3B.

The contig-level gene models (PGSB - formerly known as MIPS version 2.2) were projected to the chromosome
GDEC-derived models have
nes are viewable as an additional

s into chromosomal
corporates 261,251 unanchored

, Muñoz-Amatriaín M, Close TJ,
tein N, Waugh R (2013). *Anchoring*
. The Plant Journal 76(4):718-727.

halak M, Kianian S, Spielmeyer W,
Appels R, Safar J, Simkova H,
bread wheat chromosome 3B.

mene BLAST BioMart Tools Downloads Help Feedback

Zea mays

Zea mays

Data Source MaizeSequence.org | Taxonomy ID 4577

a mays… Go

or 1:8001-18000 or **Carboxypeptidase**

notations

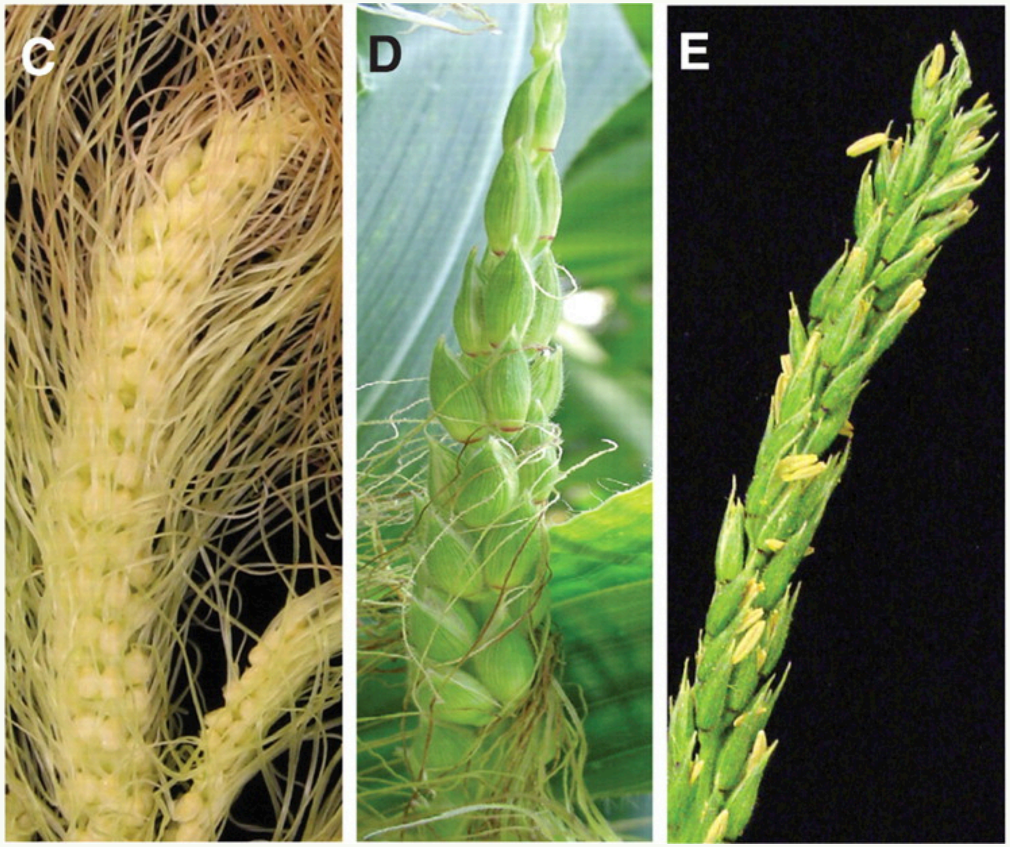
codon introduced in Sb03g028120)

test InterProScan

EMBL-EBI

This work is a joint project between EnsemblGenomes at the European Bioinformatics Institute and the group of Doreen Ware at the Cold Spring Harbor Laboratory, who have developed the Gramene database, a resource for plant comparative genomics based on Ensembl technology. A common set of

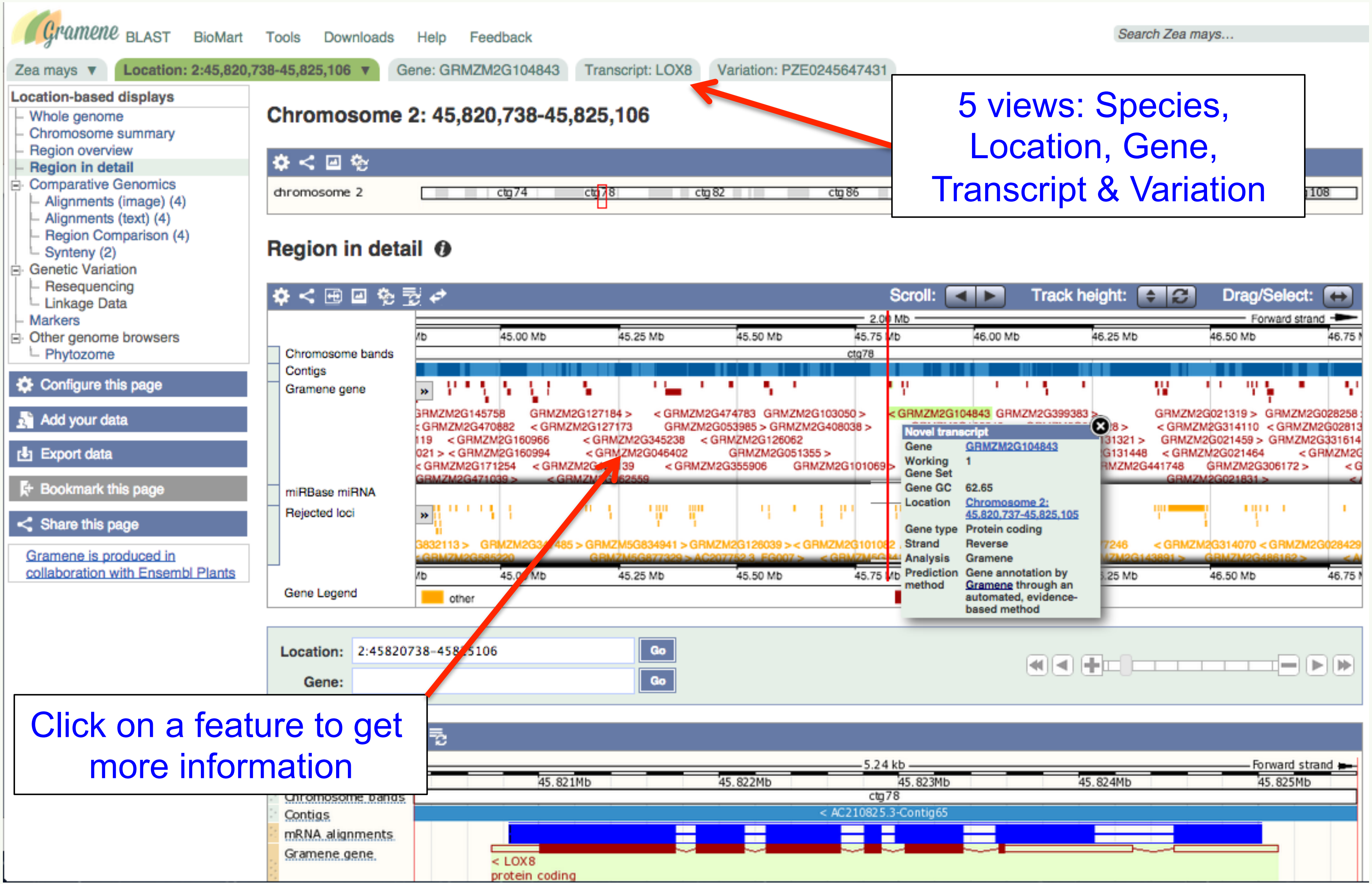
Example: *tasselseed1* (*ts1*) encodes LOX8



Acosta *et al* (2009). Science 323:262.
(**C**) Blank-treated mutant *ts1* tassel. (**D**) JA-treated *ts1* tassel. (**E**) JA-treated *ts2* tassel.

- Acosta *et al* (2009) positionally cloned & functionally characterized *ts1*
- *ts1* encodes LOX8
- LOX8 is a lipoxygenase affecting jasmonic acid (JA) signaling in sex determination

Navigating Ensembl Browsers



Gene	GRMZM2G104843
Working Gene Set	1
Gene GC	62.65
Location	Chromosome 2: 45,820,737-45,825,105
Gene type	Protein coding
Strand	Reverse
Analysis	Gramene
Prediction method	Gene annotation by Gramene through an automated, evidence-based method

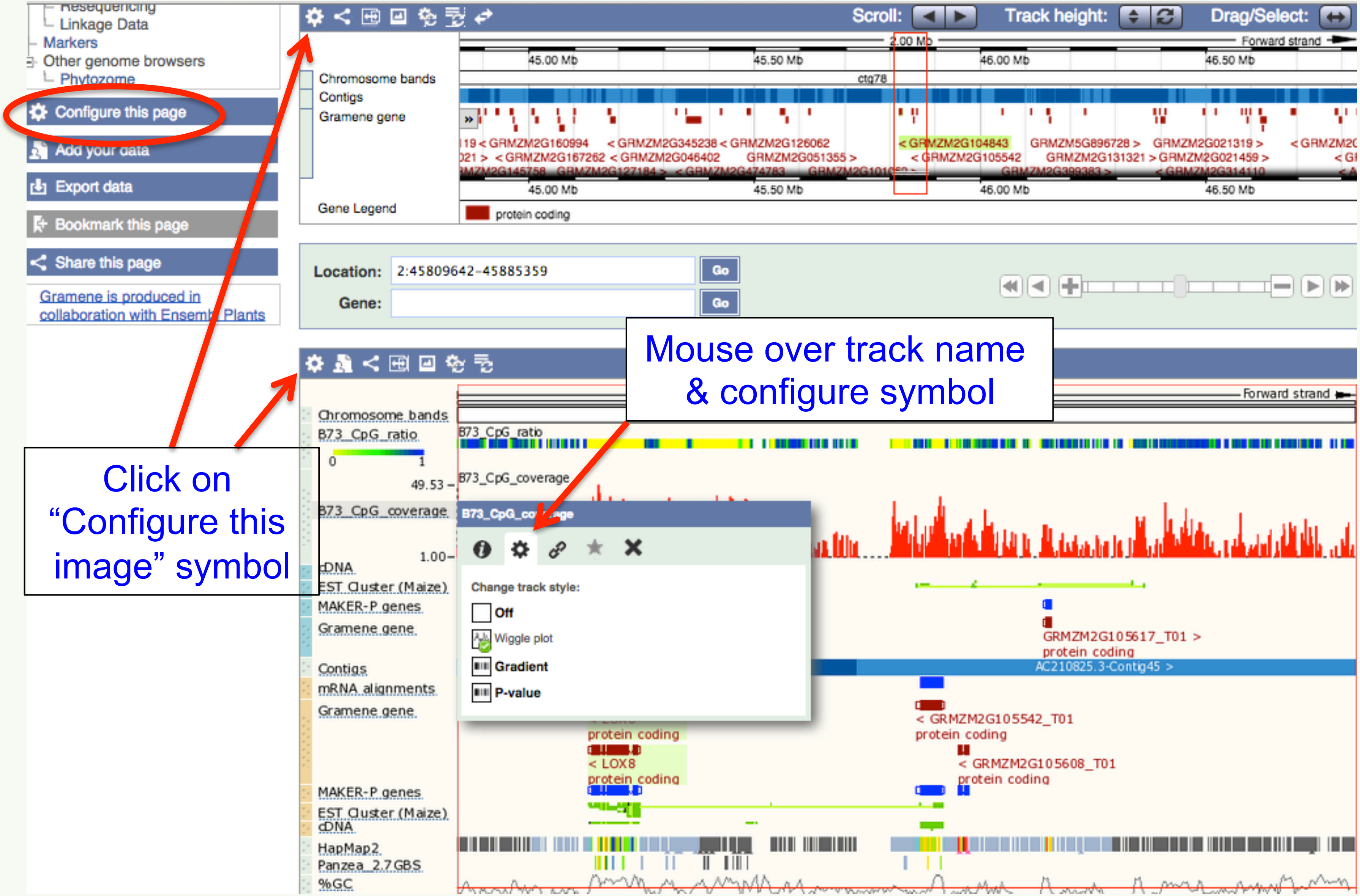
Location: 2:45820738-45825106 Go

Gene: Go

5 views: Species, Location, Gene, Transcript & Variation

Click on a feature to get more information

Configuring Browser Views



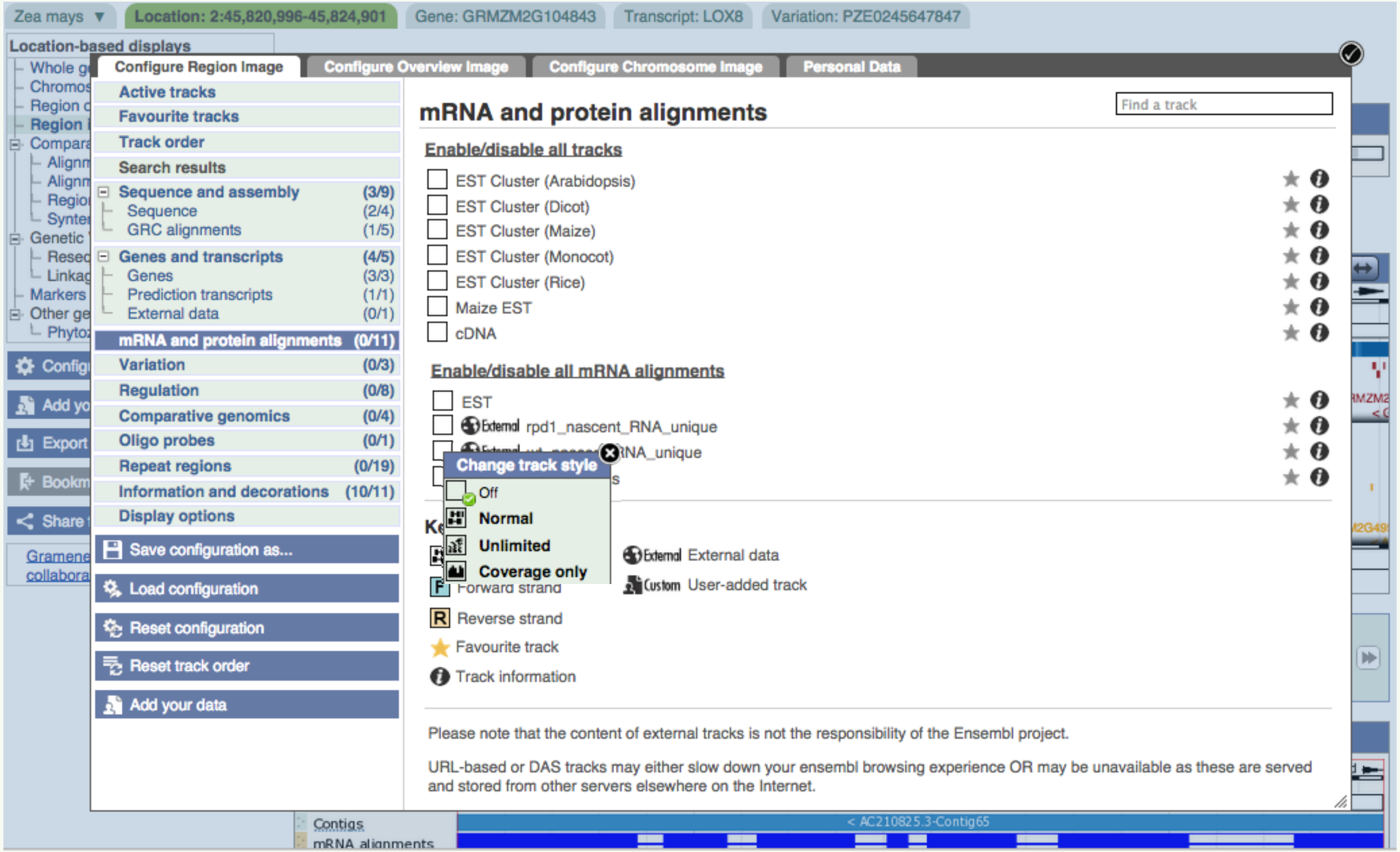
Location: 2:45809642-45885359

Gene:

Mouse over track name & configure symbol

Click on “Configure this image” symbol

Select tracks to show on the Genome browser



Zea mays ▼ | Location: 2:45,820,996-45,824,901 | Gene: GRMZM2G104843 | Transcript: LOX8 | Variation: PZE0245647847

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

Active tracks	
Favourite tracks	
Track order	
Search results	
⊟ Sequence and assembly	(3/9)
Sequence	(2/4)
GRC alignments	(1/5)
⊟ Genes and transcripts	(4/5)
Genes	(3/3)
Prediction transcripts	(1/1)
External data	(0/1)
mRNA and protein alignments	(0/11)
Variation	(0/3)
Regulation	(0/8)
Comparative genomics	(0/4)
Oligo probes	(0/1)
Repeat regions	(0/19)
Information and decorations	(10/11)
Display options	

Save configuration as...

Load configuration

Reset configuration

Reset track order

Add your data

mRNA and protein alignments

Find a track

Enable/disable all tracks

- ☐ EST Cluster (Arabidopsis)
- ☐ EST Cluster (Dicot)
- ☐ EST Cluster (Maize)
- ☐ EST Cluster (Monocot)
- ☐ EST Cluster (Rice)
- ☐ Maize EST
- ☐ cDNA

Enable/disable all mRNA alignments

- ☐ EST
- ☐ External rpd1_nascent_RNA_unique
- ☐ External wt_nascent_RNA_unique
- ☐ …s

Change track style

- Off
- Normal
- Unlimited
- Coverage only

Key

- F Forward strand
- R Reverse strand
- ★ Favourite track
- 🛈 Track information
- External External data
- Custom User-added track

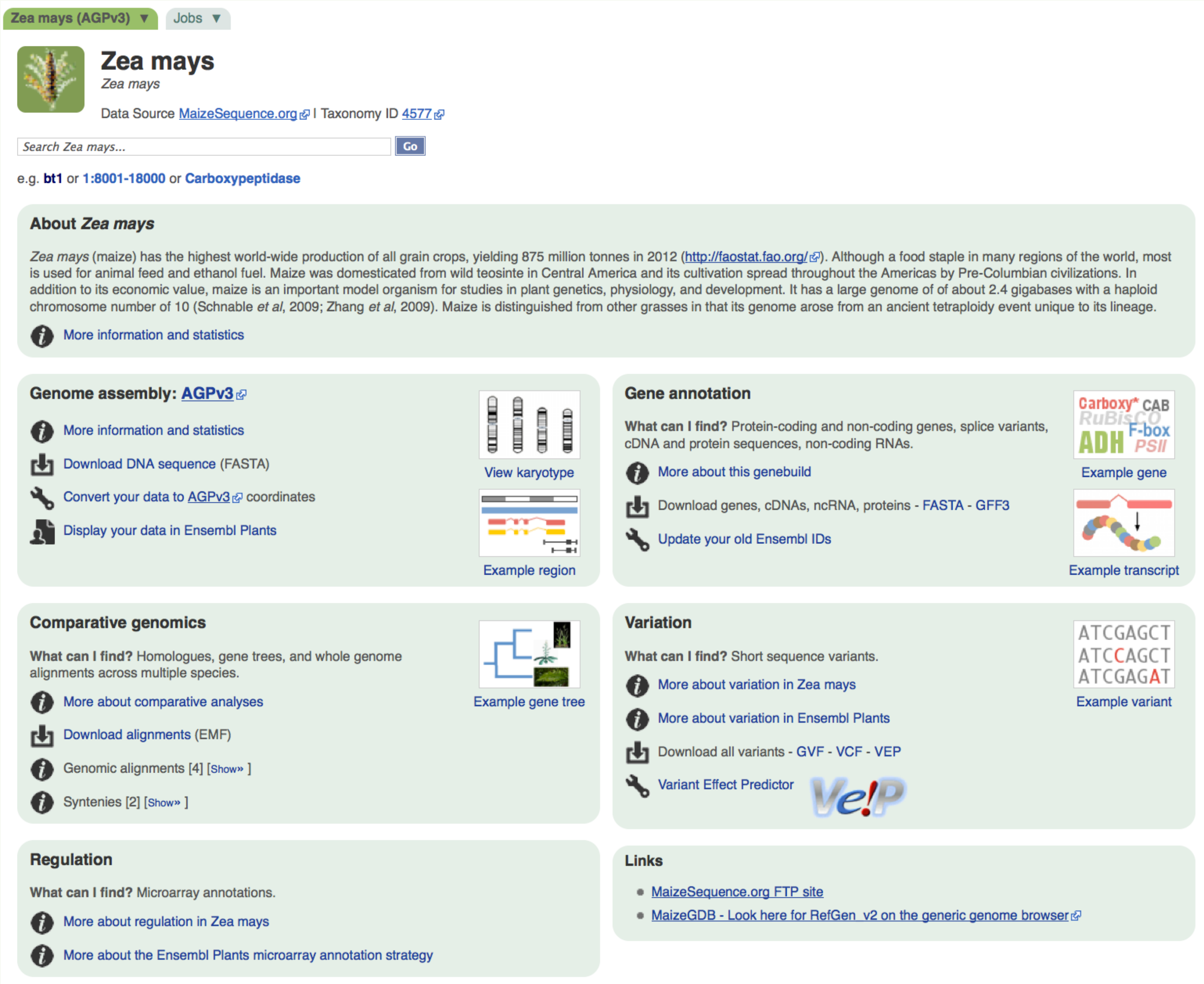
Please note that the content of external tracks is not the responsibility of the Ensembl project.

URL-based or DAS tracks may either slow down your ensembl browsing experience OR may be unavailable as these are served and stored from other servers elsewhere on the Internet.

Maize V3 data (genome browser)

- RefGen v3 assembly & gene annotation
- MAKER-P gene models
- Variation: HapMap2 & Panzea 2.7 GBS
- B73 & Mo17 methylome (Regulski *et al*, 2013)
- Nascent transcriptomes (Erhard *et al*, 2015)
- Long non-coding RNAs (Li *et al*, 2014)
- GeneChip maize genome array
- 186 orthology-based projected pathways in Plant Reactome (506 reactions & 925 gene products)

Maize V3



Zea mays

Zea mays

Data Source MaizeSequence.org | 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 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 Ensembl Plants

View karyotype

Example region

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

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

More about comparative analyses

Download alignments (EMF)

Genomic alignments [4] [Show»]

Syntenies [2] [Show»]

Example gene tree

Variation

What can I find? Short sequence variants.

More about variation in Zea mays

More about variation in Ensembl Plants

Download all variants - GVF - VCF - VEP

Variant Effect Predictor

Example variant

Regulation

What can I find? Microarray annotations.

More about regulation in Zea mays

More about the Ensembl Plants microarray annotation strategy

Links

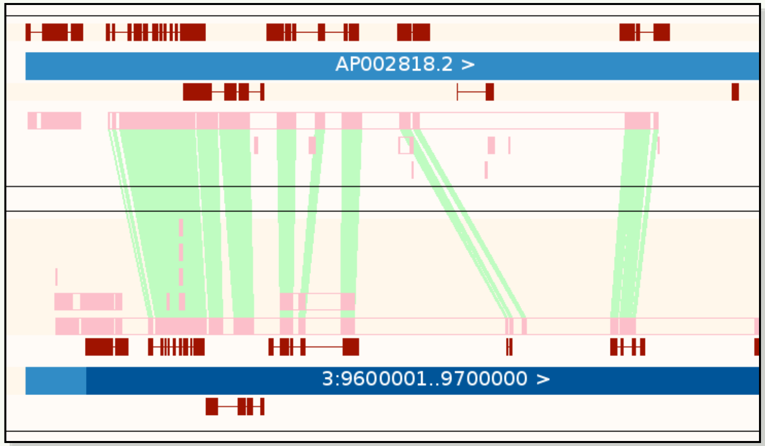
- MaizeSequence.org FTP site
- MaizeGDB - Look here for RefGen_v2 on the generic genome browser

Comparative Phylogenomics

Reconstructing evolutionary histories

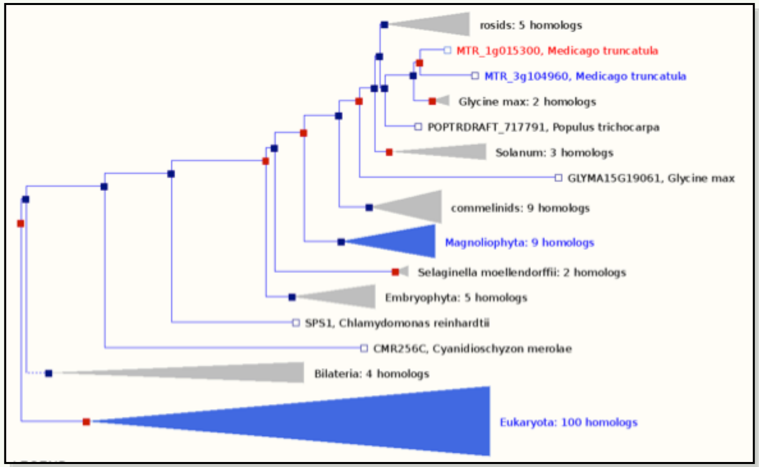
Whole Genome DNA Alignment

- Pairwise alignments
- B/LASTZ-CHAIN-NET method



Phylogenetic Gene Trees

- Infers orthologs and paralogs
- Taxonomic dating
- Synteny maps



Vilella A.J., et al. (2008). *Genome Res.* doi:10.1101/gr.073585.107
Schwartz S et al., Genome Res.;13(1):103-7, Kent WJ et al., Proc Natl Acad Sci USA., 2003;100(20):11484-9

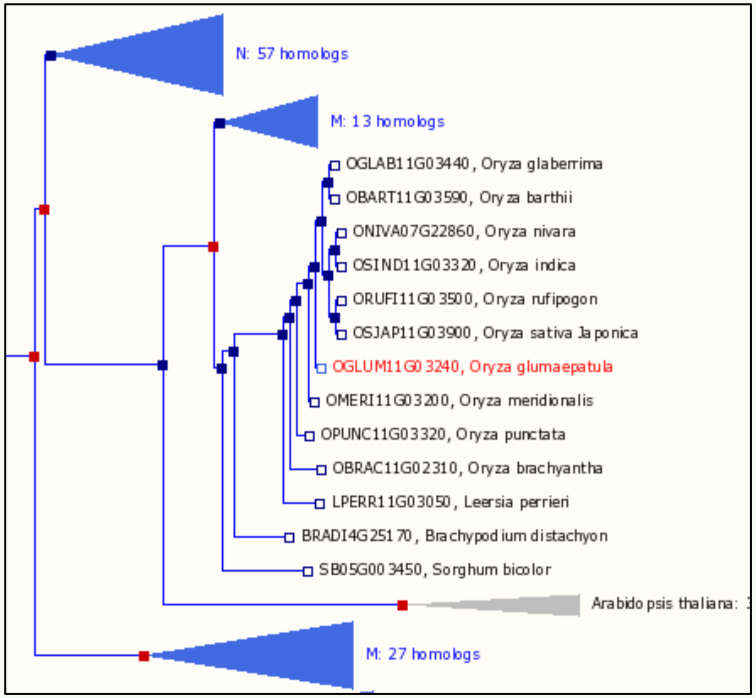
http://useast.ensembl.org/info/docs/compara/homology_method.html

Ensembl Compara Gene Trees

Sharon Wei (CSHL)

Phylogenetic Gene Trees

- Gene “family” clusters
- Protein & CDS MSA
- 5 Different tree-build methods
  - ML-protein-WAG, ML-phyml-codon(HKY), NJ-codon-p-distance, NJ-codon-dS, NJ-codon-dN
- Consensified by TreeBest
- Infers orthologs and paralogs
- Taxonomic dating



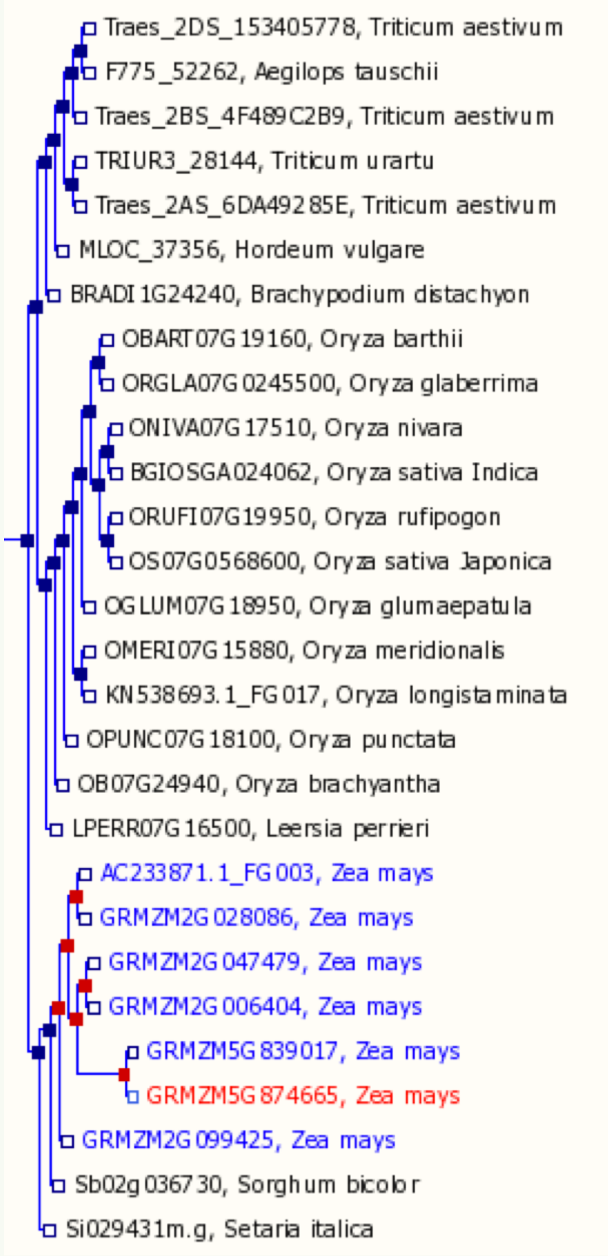
Vilella A.J., et al. (2008). *Genome Res.* doi:10.1101/gr.073585.107

http://useast.ensembl.org/info/docs/compara/homology_method.html

Heng Li. TreeBest: https://github.com/muffato/treebest

Maize-specific gene expansions

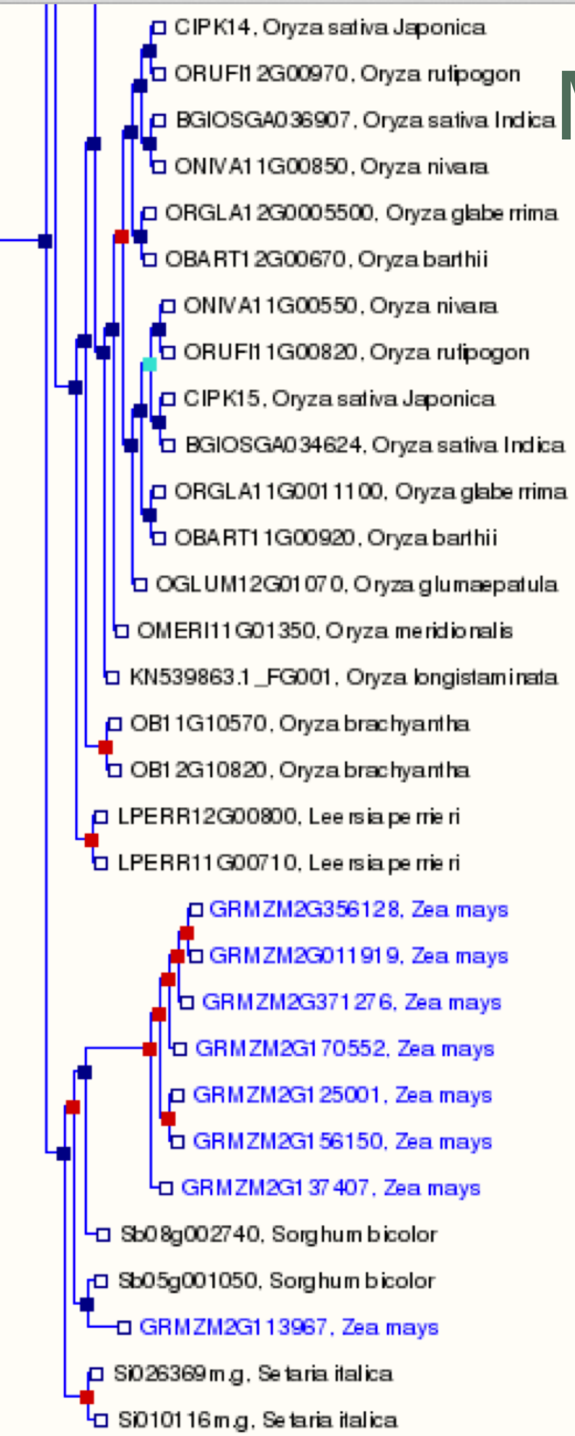
Example 1



- Calcium-dependent (calmodulin-like) protein kinase
- Orthologous group is single-copy in other grass species
- Expanded to 7 unlinked copies in maize
- Possible functions in stress response, growth and development

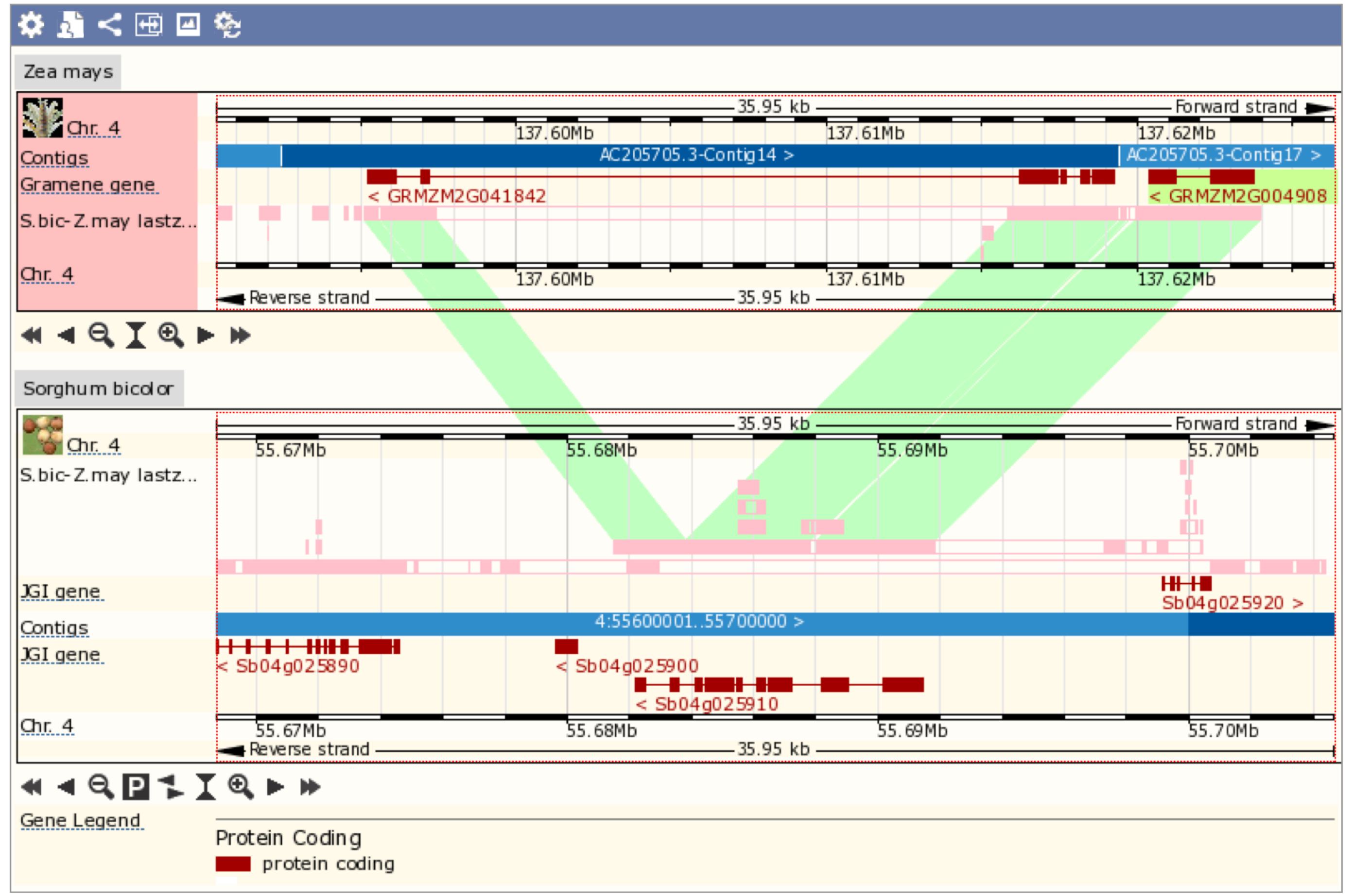
Maize-specific gene expansions

Example 2



- CBL-interacting protein kinase
- Orthologous group is low copy in other grasses
- Expanded to 7 unlinked copies in maize
- Arabidopsis ortholog SOS2 involved in Na$^+$ and K$^+$ homeostasis

Split gene model



Mis-annotated gene model spans 2 contigs

Exporting Images & Data

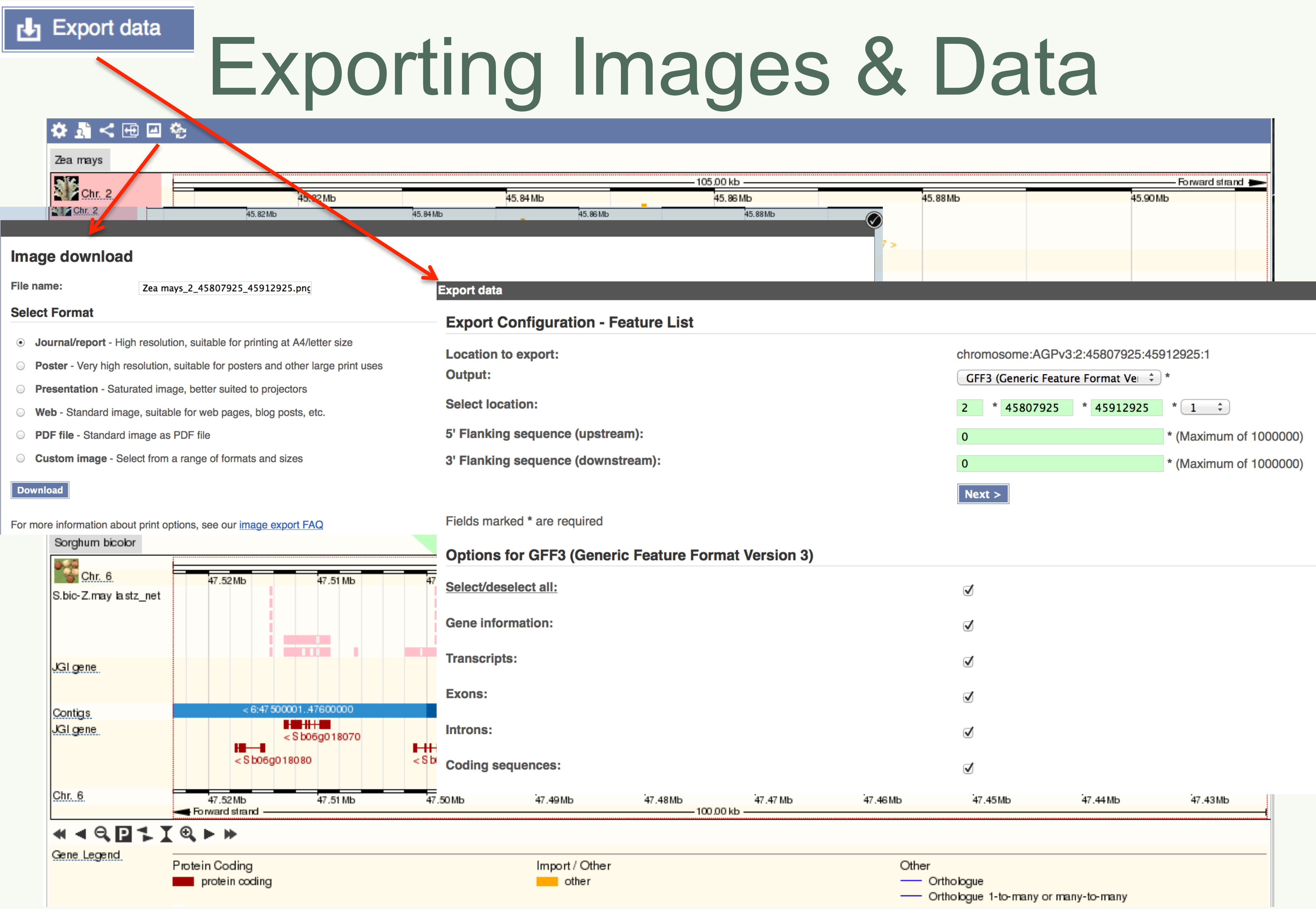


Image download

File name: Zea mays_2_45807925_45912925.png

Select Format

- ● **Journal/report** - High resolution, suitable for printing at A4/letter size
- ○ **Poster** - Very high resolution, suitable for posters and other large print uses
- ○ **Presentation** - Saturated image, better suited to projectors
- ○ **Web** - Standard image, suitable for web pages, blog posts, etc.
- ○ **PDF file** - Standard image as PDF file
- ○ **Custom image** - Select from a range of formats and sizes

Download

For more information about print options, see our image export FAQ

Export data

Export Configuration - Feature List

Location to export: chromosome:AGPv3:2:45807925:45912925:1

Output: GFF3 (Generic Feature Format Ve *

Select location: 2 * 45807925 * 45912925 * 1

5' Flanking sequence (upstream): 0 * (Maximum of 1000000)

3' Flanking sequence (downstream): 0 * (Maximum of 1000000)

Next >

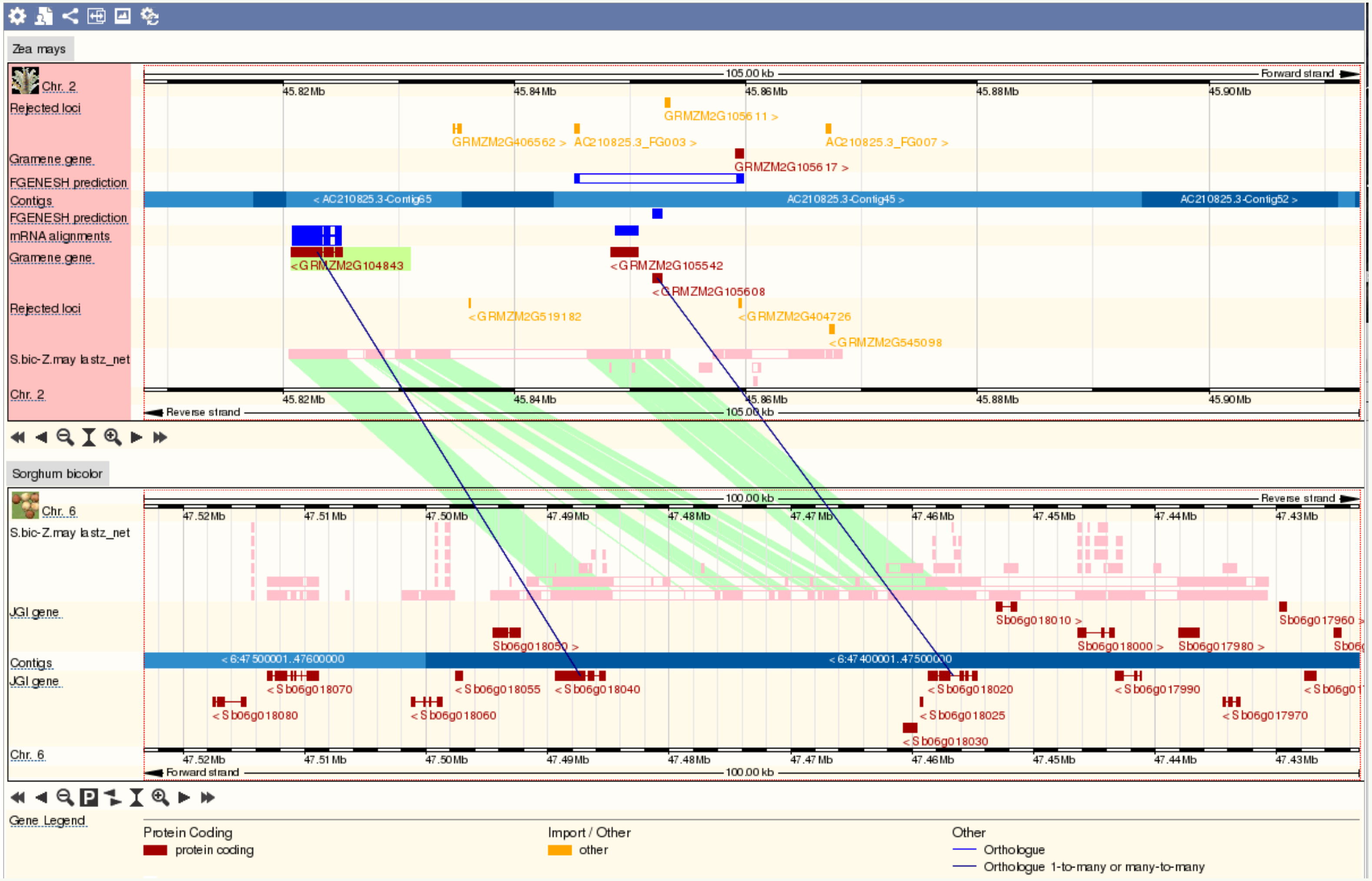
Fields marked * are required

Options for GFF3 (Generic Feature Format Version 3)

Select/deselect all:	☑
Gene information:	☑
Transcripts:	☑
Exons:	☑
Introns:	☑
Coding sequences:	☑

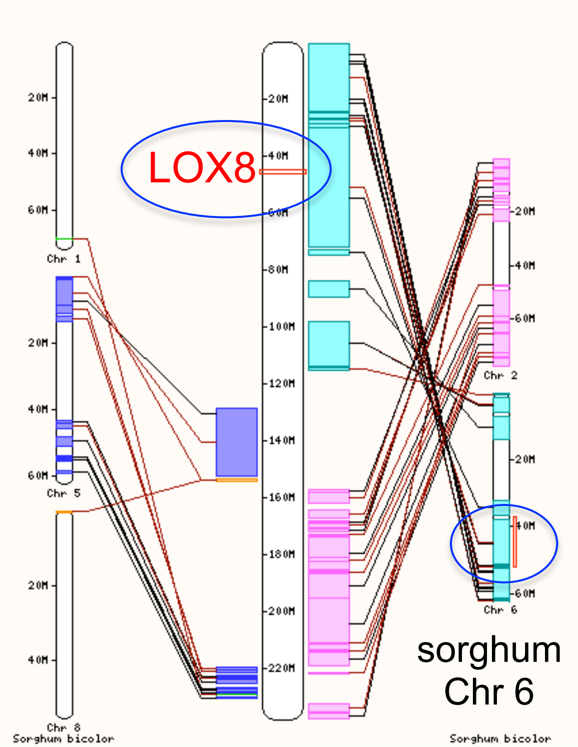
Whole Genome Alignments

- Structural variation
- Conserved non-coding

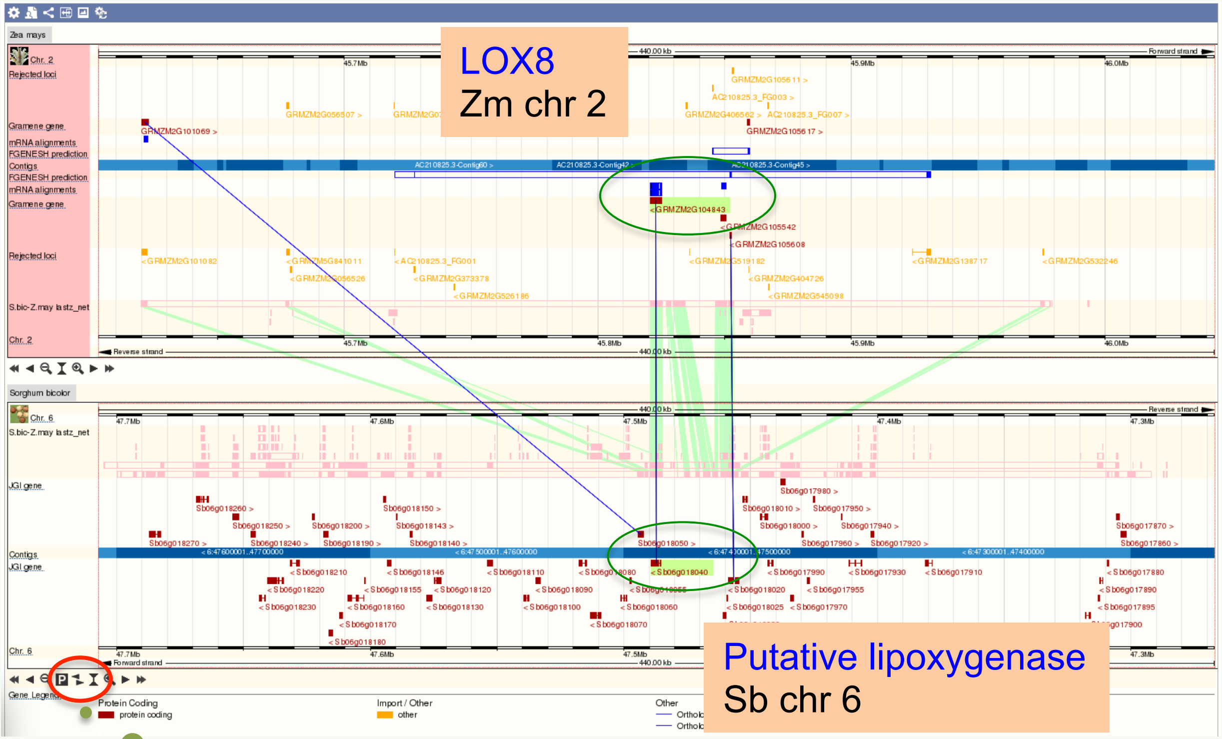


Synteny & orthologous region comparison

maize
Chr 2



sorghum
Chr 6

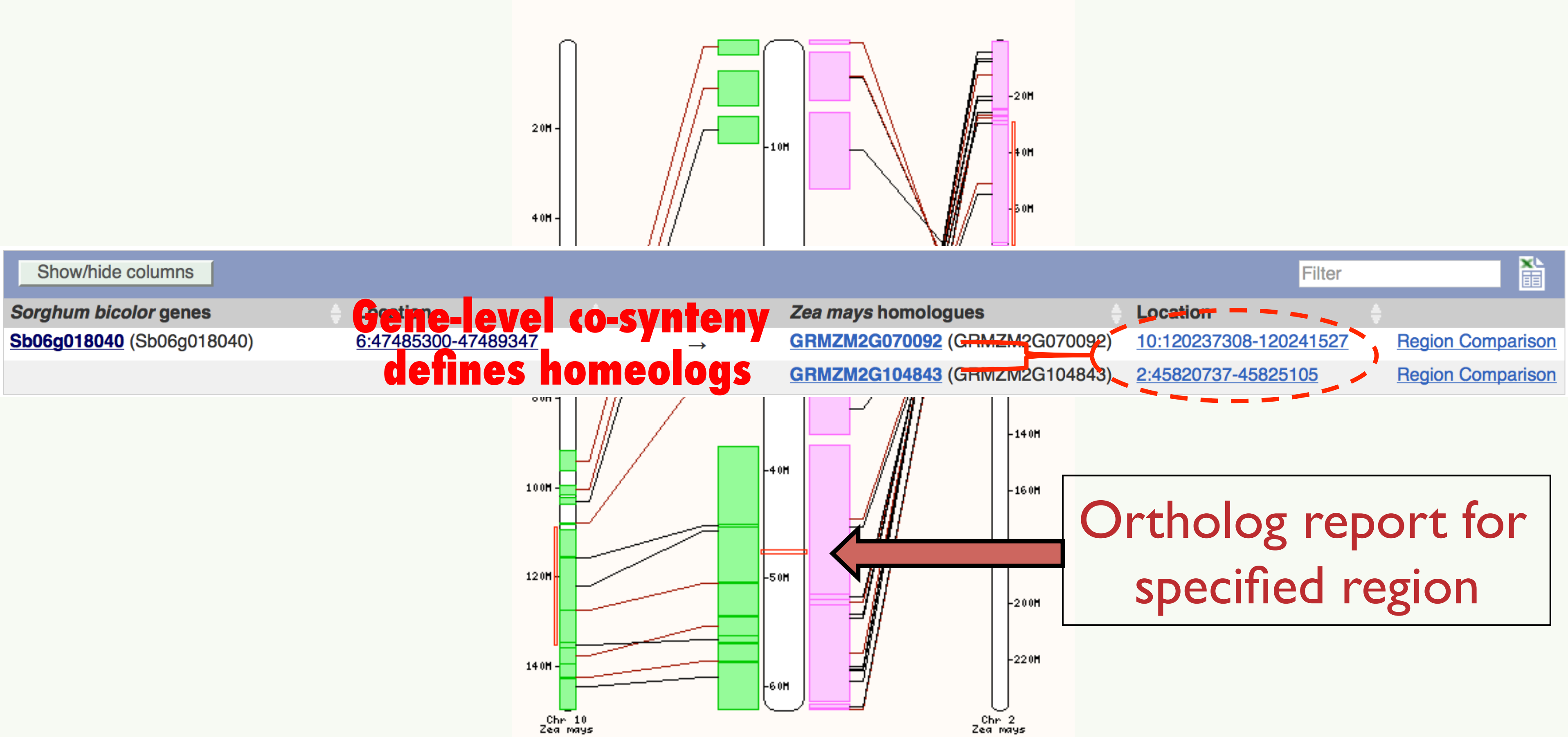


LOX8
Zm chr 2

Putative lipoxygenase
Sb chr 6

Click to invert strand orientation

Browse Homoeologous Regions From Ancient Polyploidy



Sorghum bicolor genes	Location		*Zea mays* homologues	Location	
Sb06g018040 (Sb06g018040)	6:47485300-47489347	→	**GRMZM2G070092** (GRMZM2G070092)	10:120237308-120241527	Region Comparison
			GRMZM2G104843 (GRMZM2G104843)	2:45820737-45825105	Region Comparison

Gene-level co-synteny defines homeologs

Ortholog report for specified region

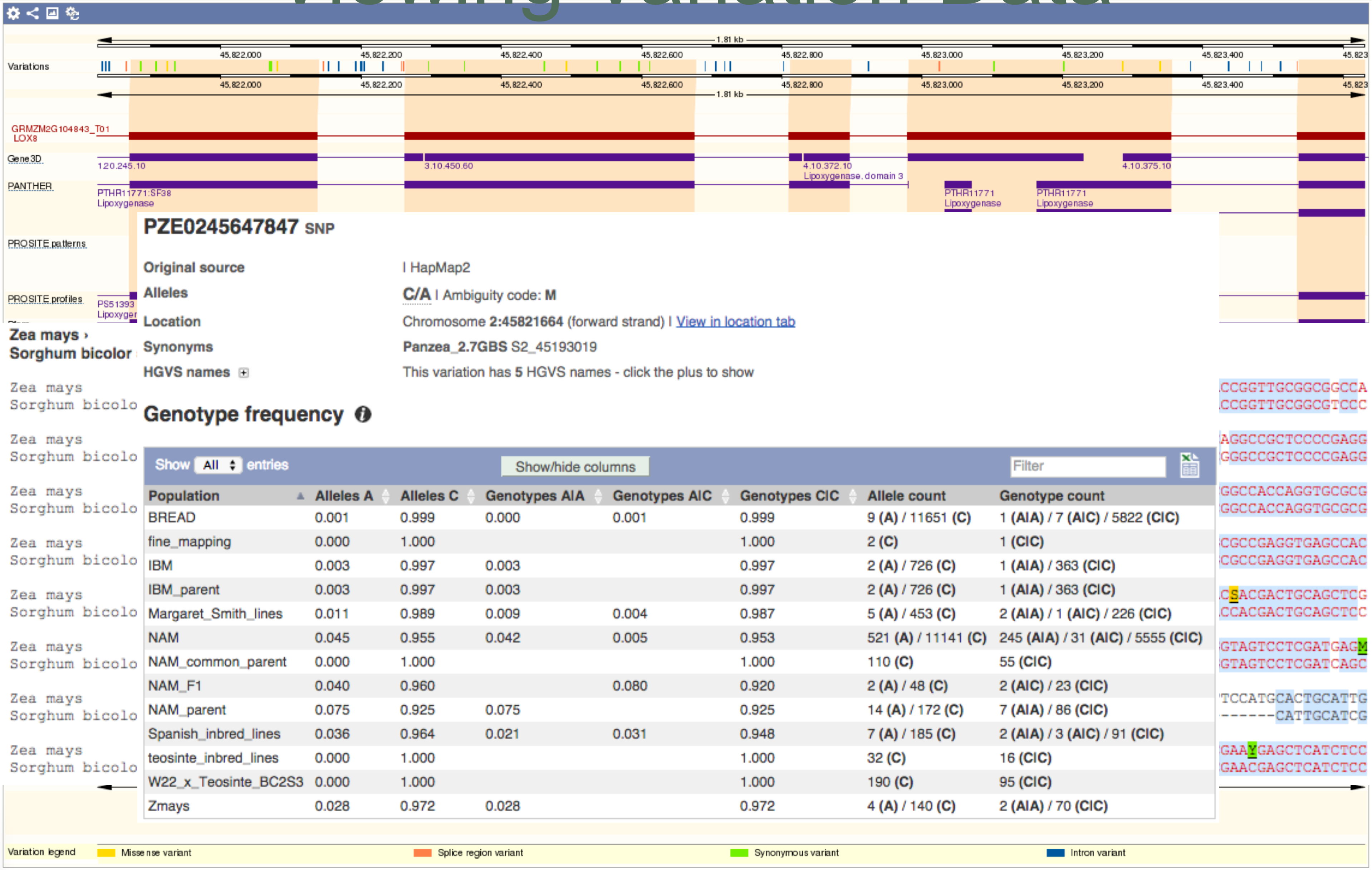
Chr6
Sorghum

Ensembl Diversity Databases

- Variant Effect Prediction (VEP)
  - Classifies SNPs based on predicted effect on transcript using Sequence Ontology terms
- Display individual genotypes and population frequencies

intron_variant
missense_variant
synonymous_variant
stop_gained
stop_lost
splice_donor_variant
splice_acceptor_variant
initiator_codon_variant
stop_retained_variant

Viewing Variation Data



PZE0245647847 SNP

Original source | HapMap2

Alleles C/A | Ambiguity code: **M**

Location Chromosome **2:45821664** (forward strand) | View in location tab

Synonyms **Panzea_2.7GBS** S2_45193019

HGVS names ⊞ This variation has **5** HGVS names - click the plus to show

Genotype frequency

Show All entries — Show/hide columns — Filter

Population	Alleles A	Alleles C	Genotypes A|A	Genotypes A|C	Genotypes C|C	Allele count	Genotype count
BREAD	0.001	0.999	0.000	0.001	0.999	9 **(A)** / 11651 **(C)**	1 **(A|A)** / 7 **(A|C)** / 5822 **(C|C)**
fine_mapping	0.000	1.000			1.000	2 **(C)**	1 **(C|C)**
IBM	0.003	0.997	0.003		0.997	2 **(A)** / 726 **(C)**	1 **(A|A)** / 363 **(C|C)**
IBM_parent	0.003	0.997	0.003		0.997	2 **(A)** / 726 **(C)**	1 **(A|A)** / 363 **(C|C)**
Margaret_Smith_lines	0.011	0.989	0.009	0.004	0.987	5 **(A)** / 453 **(C)**	2 **(A|A)** / 1 **(A|C)** / 226 **(C|C)**
NAM	0.045	0.955	0.042	0.005	0.953	521 **(A)** / 11141 **(C)**	245 **(A|A)** / 31 **(A|C)** / 5555 **(C|C)**
NAM_common_parent	0.000	1.000			1.000	110 **(C)**	55 **(C|C)**
NAM_F1	0.040	0.960		0.080	0.920	2 **(A)** / 48 **(C)**	2 **(A|C)** / 23 **(C|C)**
NAM_parent	0.075	0.925	0.075		0.925	14 **(A)** / 172 **(C)**	7 **(A|A)** / 86 **(C|C)**
Spanish_inbred_lines	0.036	0.964	0.021	0.031	0.948	7 **(A)** / 185 **(C)**	2 **(A|A)** / 3 **(A|C)** / 91 **(C|C)**
teosinte_inbred_lines	0.000	1.000			1.000	32 **(C)**	16 **(C|C)**
W22_x_Teosinte_BC2S3	0.000	1.000			1.000	190 **(C)**	95 **(C|C)**
Zmays	0.028	0.972	0.028		0.972	4 **(A)** / 140 **(C)**	2 **(A|A)** / 70 **(C|C)**

Variation Data

- Arabidopsis –1001 Genomes
- Maize – HapMap 1 & 2, Panzea GBS
- Wheat – HapMap (Jordan et al), CerealsDB
- Rice – dbSNP, McNally, Zhao, 3000 IRRI (coming soon)
- Tomato – 100 Genomes
- Brachypodium
- Barley
- Sorghum
- Grapevine
- Oryza glaberrima

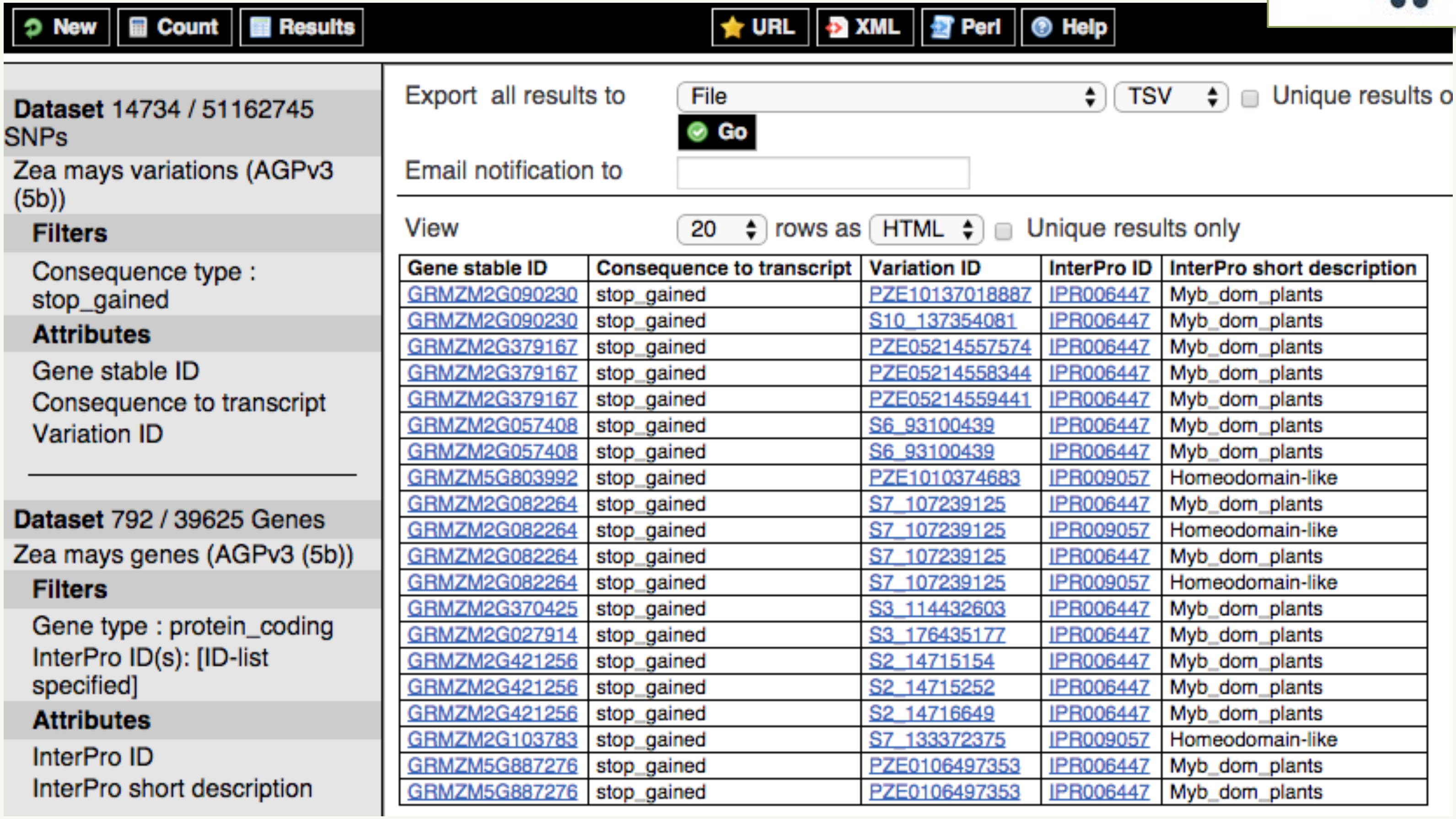
Gramene Mart

Complex custom queries to mine data:

- InterPro
- GO
- VEP
- Orthologs/Paralogs

Use case: Find transcription factors having "stop_gained" alleles

bio mart



Dataset 14734 / 51162745 SNPs

Zea mays variations (AGPv3 (5b))

Filters

Consequence type : stop_gained

Attributes

Gene stable ID
Consequence to transcript
Variation ID

Dataset 792 / 39625 Genes

Zea mays genes (AGPv3 (5b))

Filters

Gene type : protein_coding
InterPro ID(s): [ID-list specified]

Attributes

InterPro ID
InterPro short description

Export all results to: File, TSV ☐ Unique results o

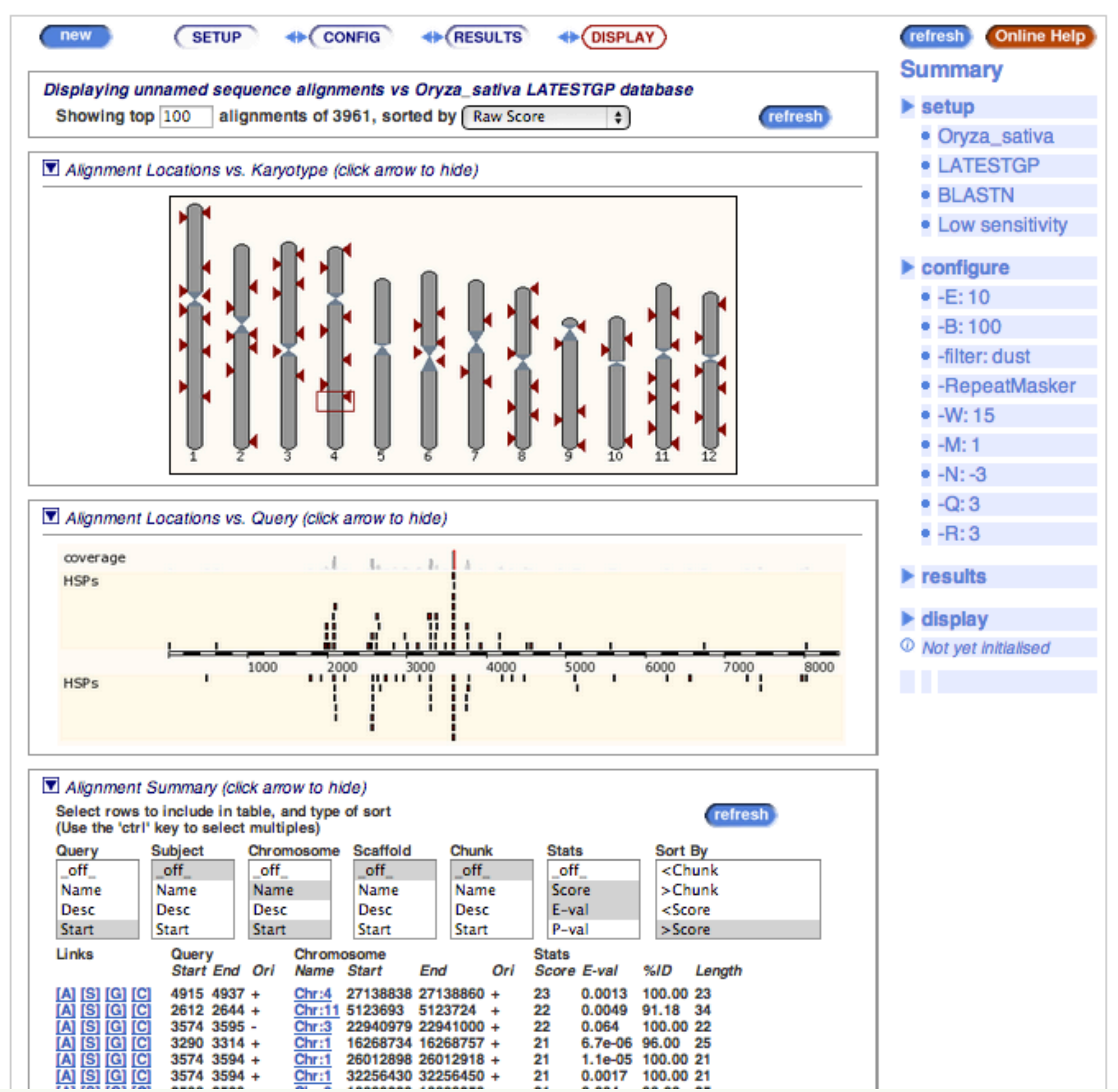
Email notification to

View 20 rows as HTML ☐ Unique results only

Gene stable ID	Consequence to transcript	Variation ID	InterPro ID	InterPro short description
GRMZM2G090230	stop_gained	PZE10137018887	IPR006447	Myb_dom_plants
GRMZM2G090230	stop_gained	S10_137354081	IPR006447	Myb_dom_plants
GRMZM2G379167	stop_gained	PZE05214557574	IPR006447	Myb_dom_plants
GRMZM2G379167	stop_gained	PZE05214558344	IPR006447	Myb_dom_plants
GRMZM2G379167	stop_gained	PZE05214559441	IPR006447	Myb_dom_plants
GRMZM2G057408	stop_gained	S6_93100439	IPR006447	Myb_dom_plants
GRMZM2G057408	stop_gained	S6_93100439	IPR006447	Myb_dom_plants
GRMZM5G803992	stop_gained	PZE1010374683	IPR009057	Homeodomain-like
GRMZM2G082264	stop_gained	S7_107239125	IPR006447	Myb_dom_plants
GRMZM2G082264	stop_gained	S7_107239125	IPR009057	Homeodomain-like
GRMZM2G082264	stop_gained	S7_107239125	IPR006447	Myb_dom_plants
GRMZM2G082264	stop_gained	S7_107239125	IPR009057	Homeodomain-like
GRMZM2G370425	stop_gained	S3_114432603	IPR006447	Myb_dom_plants
GRMZM2G027914	stop_gained	S3_176435177	IPR006447	Myb_dom_plants
GRMZM2G421256	stop_gained	S2_14715154	IPR006447	Myb_dom_plants
GRMZM2G421256	stop_gained	S2_14715252	IPR006447	Myb_dom_plants
GRMZM2G421256	stop_gained	S2_14716649	IPR006447	Myb_dom_plants
GRMZM2G103783	stop_gained	S7_133372375	IPR009057	Homeodomain-like
GRMZM5G887276	stop_gained	PZE0106497353	IPR006447	Myb_dom_plants
GRMZM5G887276	stop_gained	PZE0106497353	IPR006447	Myb_dom_plants

http://www.gramene.org/biomart/martview/

BLAST



Displaying unnamed sequence alignments vs Oryza_sativa LATESTGP database

Showing top 100 alignments of 3961, sorted by Raw Score

Alignment Locations vs. Karyotype (click arrow to hide)

Alignment Locations vs. Query (click arrow to hide)

Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples)

Query	Subject	Chromosome	Scaffold	Chunk	Stats	Sort By
off	_off_	_off_	_off_	_off_	_off_	<Chunk
Name	Name	Name	Name	Name	Score	>Chunk
Desc	Desc	Desc	Desc	Desc	E-val	<Score
Start	Start	Start	Start	Start	P-val	>Score

Links	Query			Chromosome				Stats			
	Start	*End*	*Ori*	*Name*	*Start*	*End*	*Ori*	*Score*	*E-val*	*%ID*	*Length*
[A] [S] [G] [C]	4915	4937	+	Chr:4	27138838	27138860	+	23	0.0013	100.00	23
[A] [S] [G] [C]	2612	2644	+	Chr:11	5123693	5123724	+	22	0.0049	91.18	34
[A] [S] [G] [C]	3574	3595	-	Chr:3	22940979	22941000	+	22	0.064	100.00	22
[A] [S] [G] [C]	3290	3314	+	Chr:1	16268734	16268757	+	21	6.7e-06	96.00	25
[A] [S] [G] [C]	3574	3594	+	Chr:1	26012898	26012918	+	21	1.1e-05	100.00	21
[A] [S] [G] [C]	3574	3594	+	Chr:1	32256430	32256450	+	21	0.0017	100.00	21

Summary

setup
- Oryza_sativa
- LATESTGP
- BLASTN
- Low sensitivity

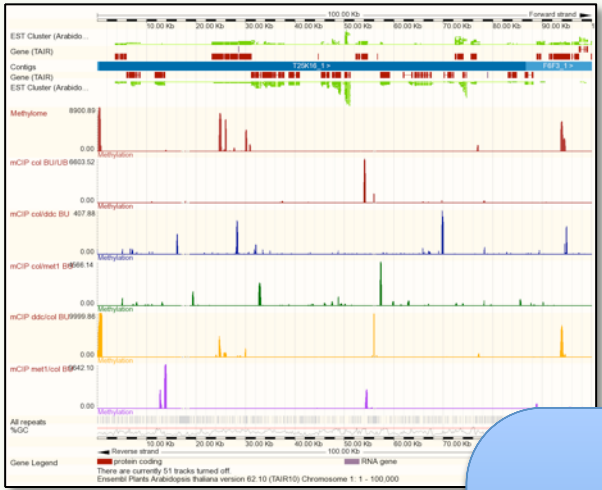
configure
- -E: 10
- -B: 100
- -filter: dust
- -RepeatMasker
- -W: 15
- -M: 1
- -N: -3
- -Q: 3
- -R: 3

results

display

Not yet initialised

Add Custom Tracks



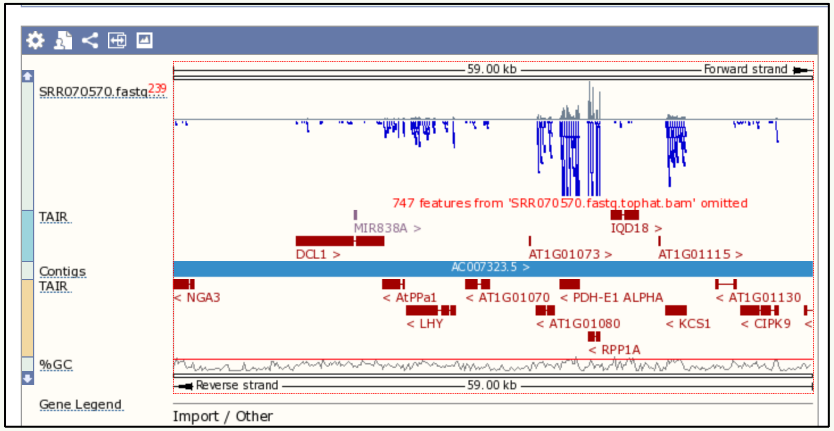
There are currently 51 tracks turned off.
Ensembl Plants Arabidopsis thaliana version 62.10 (TAIR10) Chromosome 1: 1 - 100,000

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

There are currently 57 tracks turned off.
Ensembl Plants Arabidopsis thaliana version 62.10 (TAIR10) Chromosome 1: 1 - 100,000

- NA lines
- d from my laptop
- format

What are the example data sets that you would like to have for Maize?



747 features from 'SRR070570.fastq.tophat.bam' omitted

- RNA-seq
- Uploaded from iPlant URL
- BAM alignment file

GFF
BED
BAM
VCF
bedGraph
Gbrowse
PSL
WIG
GTF
BigBed
BigWig
TrackHub

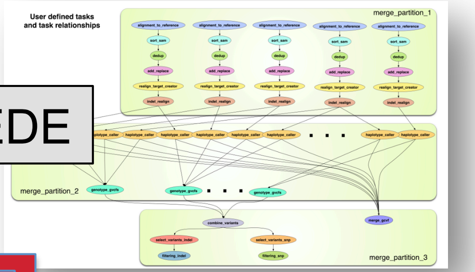
Gramene & the Informatics Ecosystem

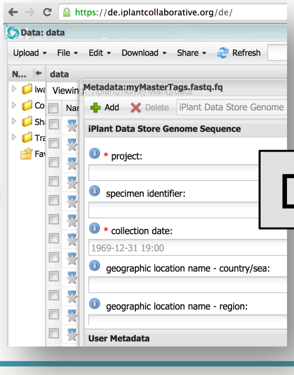
Variation: tractable data & metadata exchange

iPlant Collaborative™

- Workflows (SNP calling)
- Metadata templates
- Data store

XSEDE





Metadata:myMasterTags.fastq.fq

iPlant Data Store Genome Sequence

* project:

specimen identifier:

* collection date: 1969-12-31 19:00

geographic location name - country/sea:

geographic location name - region:

User Metadata

DE

Liya Wang

Marcela Tello-Ruiz

Reference genomes

Meta-data

VCF

Gramene

EnsemblPlants

European Variation Archive (EVA)

transPLANT

USDA ARS

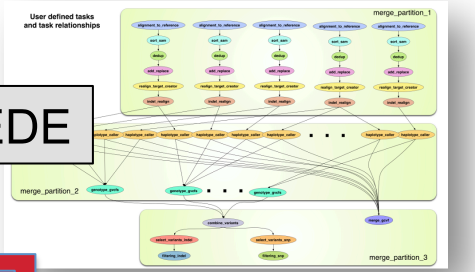
NCBI dbSNP Short Genetic Variations

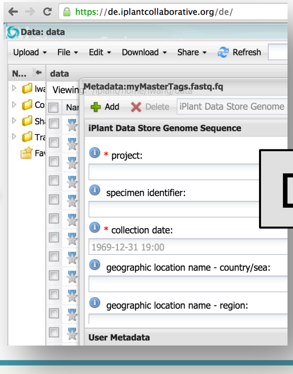
Variation: tractable data & metadata exchange

CYVERSE™

- Workflows (SNP calling)
- Metadata templates
- Data store

XSEDE





Metadata:myMasterTags.fastq.fq

iPlant Data Store Genome Sequence

* project:

specimen identifier:

* collection date: 1969-12-31 19:00

geographic location name - country/sea:

geographic location name - region:

User Metadata

DE

Liya Wang

Marcela Tello-Ruiz

Reference genomes

Meta-data

VCF

Gramene

EnsemblPlants

European Variation Archive (EVA)

transPLANT

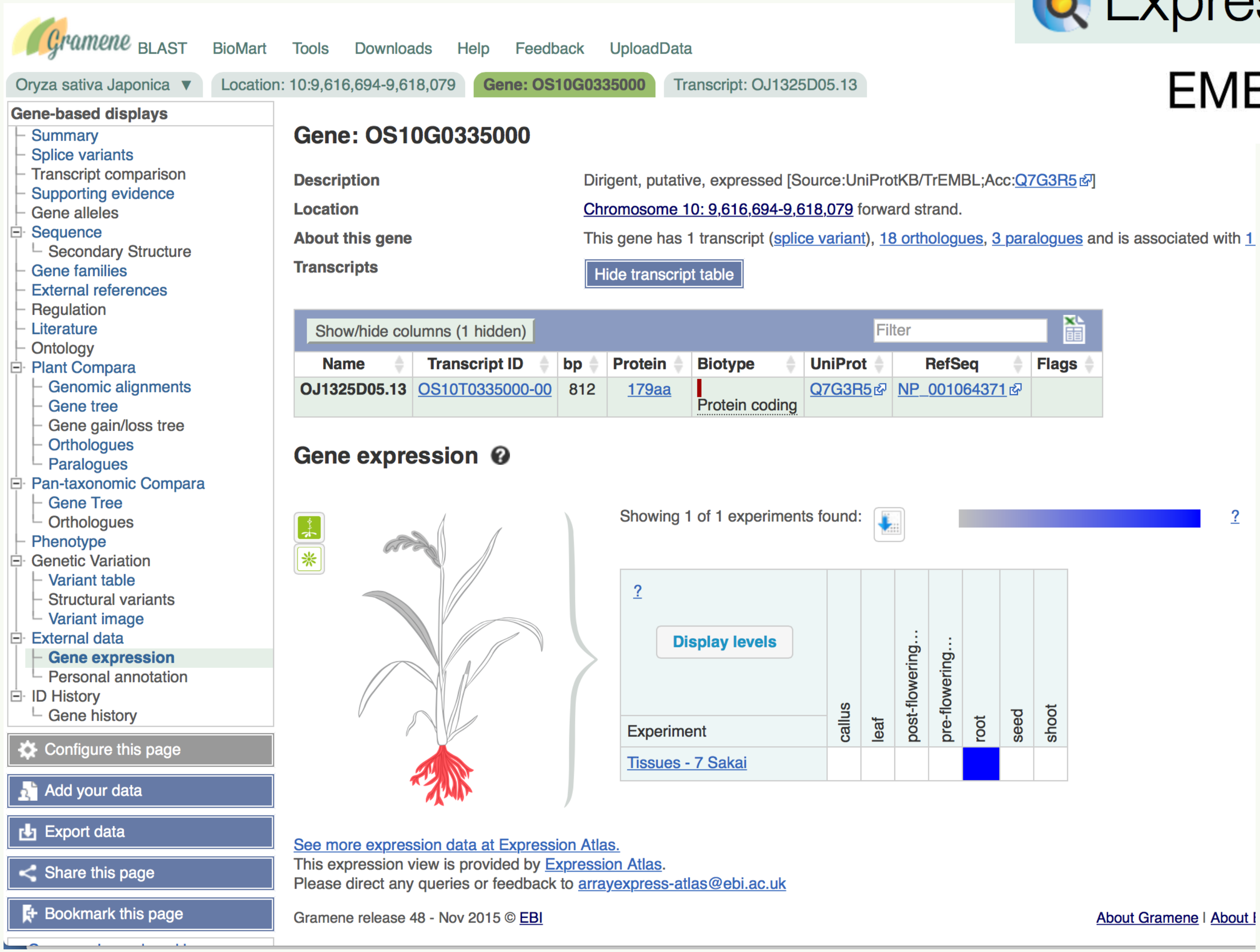
USDA ARS

NCBI dbSNP Short Genetic Variations

Expression data from Atlas

Expression Atlas

EMBL-EBI



Location: 10:9,616,694-9,618,079

Gene: OS10G0335000

Transcript: OJ1325D05.13

Gene: OS10G0335000

Description	Dirigent, putative, expressed [Source:UniProtKB/TrEMBL;Acc:Q7G3R5]
Location	Chromosome 10: 9,616,694-9,618,079 forward strand.
About this gene	This gene has 1 transcript (splice variant), 18 orthologues, 3 paralogues and is associated with 1
Transcripts	Hide transcript table

Name	Transcript ID	bp	Protein	Biotype	UniProt	RefSeq	Flags
OJ1325D05.13	OS10T0335000-00	812	179aa	Protein coding	Q7G3R5	NP_001064371	

Gene expression

Showing 1 of 1 experiments found:

Experiment	callus	leaf	post-flowering...	pre-flowering...	root	seed	shoot
Tissues - 7 Sakai					🟦		

See more expression data at Expression Atlas.
This expression view is provided by Expression Atlas.
Please direct any queries or feedback to arrayexpress-atlas@ebi.ac.uk

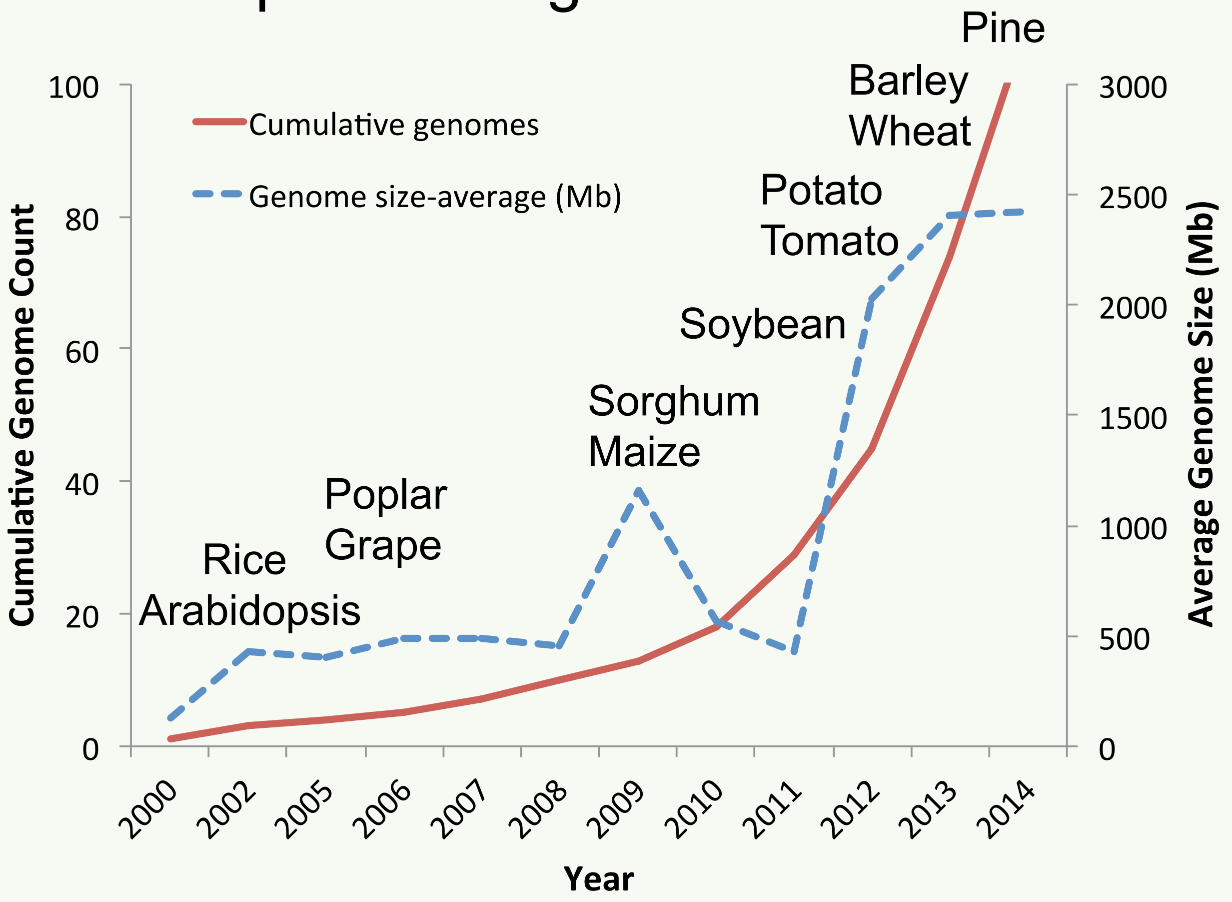
Gramene release 48 - Nov 2015 © EBI

About Gramene | About

Scaling to support the Progress in Plant Genomics

Adoption of the Ensembl Platform for Plants

100 published genomes



Adapted from Michael & VanBuren (2015) COPB

- Declining cost
- Declining quality

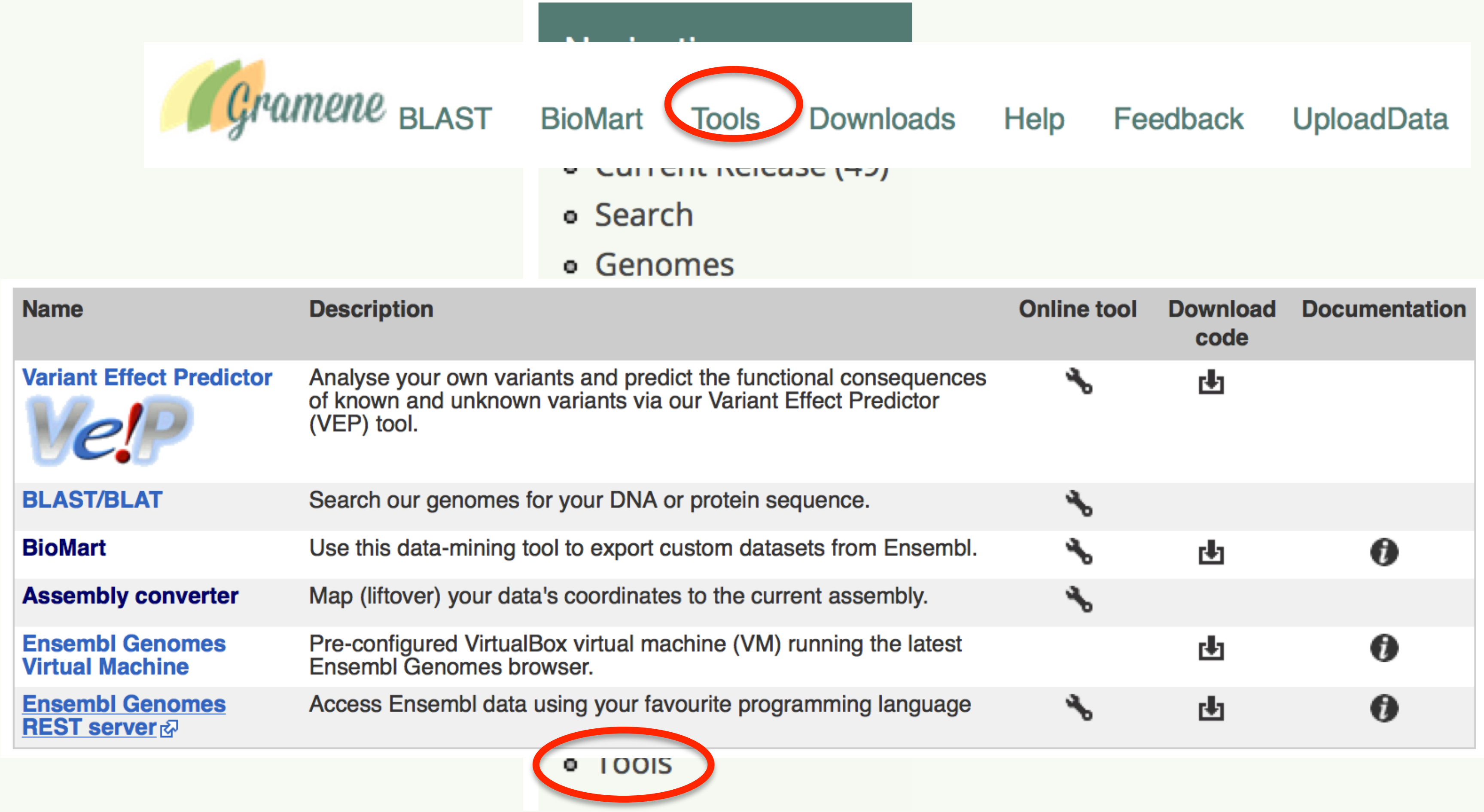
✓ Variation
✓ Expression

ENCODE-like
Metabolomics
QTL & GWAS

Maize highlights 2015-2016

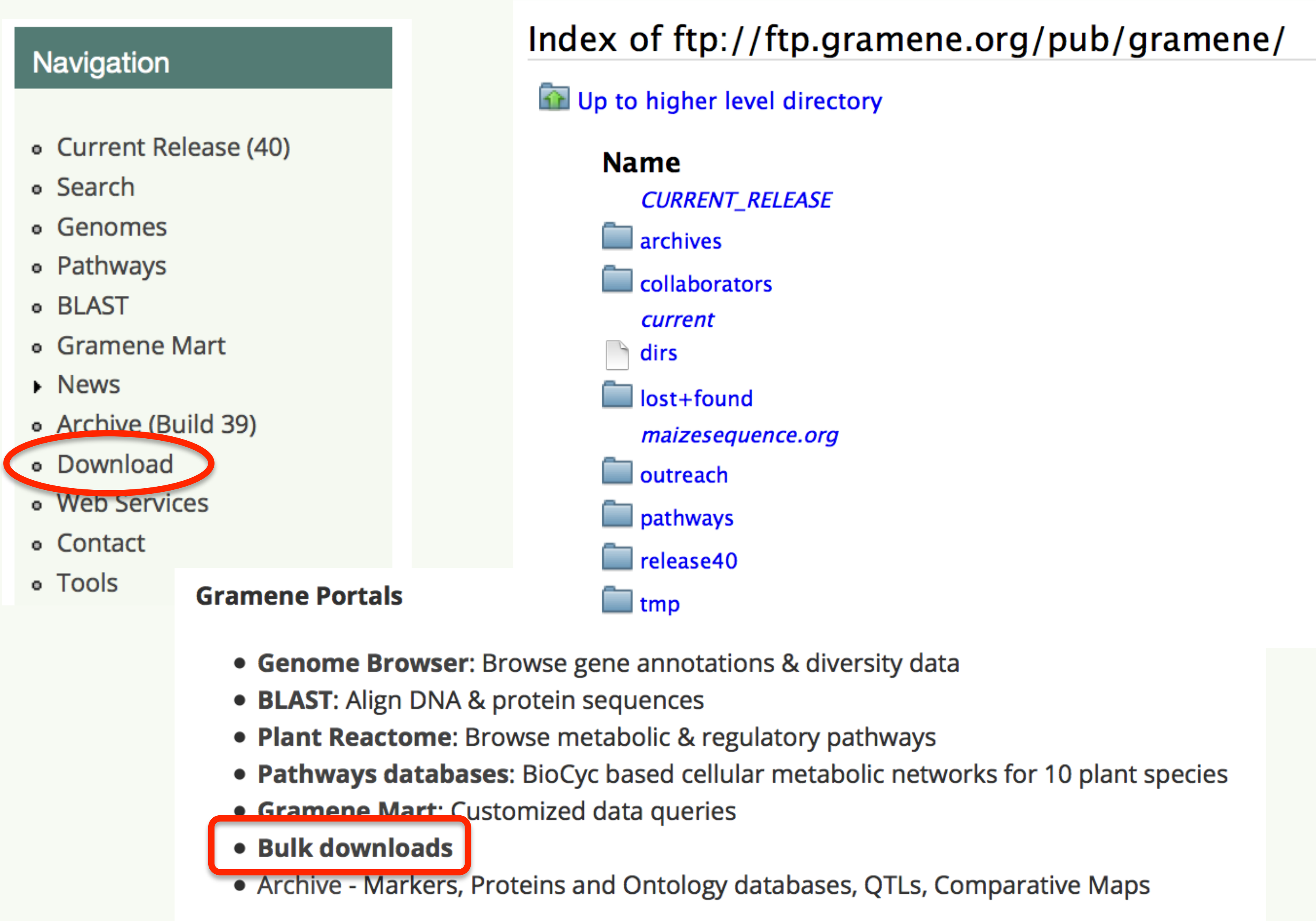
- Completed a new assembly of maize, B73 RefGen_v4 by de novo assembly of maize B73 using ~65X PacBio long reads
- Genome annotation of the new v4 maize reference genomes including protein-coding gene, non-coding gene and repeat elements
- Collaborations with MaizeGDB, NCBI and the Maize Nomenclature Group to release the maize V4 reference genome
- Gene annotation of 3 maize inbred lines (W22, Mo17 & CML247)
- Collaborate with EVA & CYVERSE on Variation prototype using the Maize HapMap2
- Collaborate with ATLAS on protoype of the maize expression data

Tools



Name	Description	Online tool	Download code	Documentation
Variant Effect Predictor VE!P	Analyse your own variants and predict the functional consequences of known and unknown variants via our Variant Effect Predictor (VEP) tool.	🔧	⤓	
BLAST/BLAT	Search our genomes for your DNA or protein sequence.	🔧		
BioMart	Use this data-mining tool to export custom datasets from Ensembl.	🔧	⤓	ℹ
Assembly converter	Map (liftover) your data's coordinates to the current assembly.	🔧		
Ensembl Genomes Virtual Machine	Pre-configured VirtualBox virtual machine (VM) running the latest Ensembl Genomes browser.		⤓	ℹ
Ensembl Genomes REST server	Access Ensembl data using your favourite programming language	🔧	⤓	ℹ

FTP site



Navigation

- Current Release (40)
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Tools

Index of ftp://ftp.gramene.org/pub/gramene/

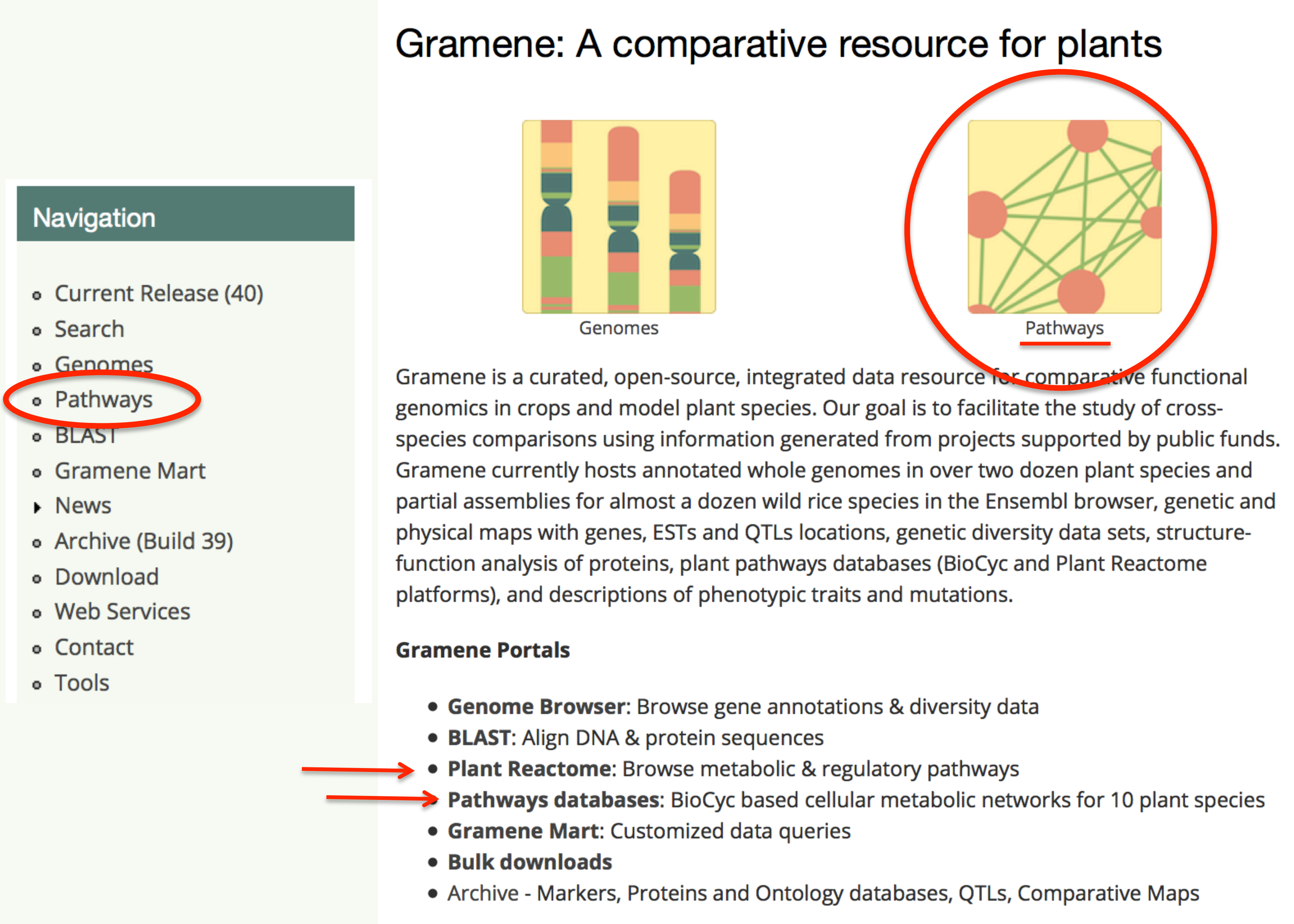
Up to higher level directory

Name
CURRENT_RELEASE
archives
collaborators
current
dirs
lost+found
maizesequence.org
outreach
pathways
release40
tmp

Gramene Portals

- **Genome Browser**: Browse gene annotations & diversity data
- **BLAST**: Align DNA & protein sequences
- **Plant Reactome**: Browse metabolic & regulatory pathways
- **Pathways databases**: BioCyc based cellular metabolic networks for 10 plant species
- **Gramene Mart**: Customized data queries
- **Bulk downloads**
- Archive - Markers, Proteins and Ontology databases, QTLs, Comparative Maps

Gramene Pathway Resources



Gramene: A comparative resource for plants

Navigation

- Current Release (40)
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Tools

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.

Gramene Portals

- **Genome Browser**: Browse gene annotations & diversity data
- **BLAST**: Align DNA & protein sequences
- **Plant Reactome**: Browse metabolic & regulatory pathways
- **Pathways databases**: BioCyc based cellular metabolic networks for 10 plant species
- **Gramene Mart**: Customized data queries
- **Bulk downloads**
- Archive - Markers, Proteins and Ontology databases, QTLs, Comparative Maps

Plant Reactome is a Resource for Comparative Plant Pathway Analysis

http://plantreactome.gramene.org

About Plant Reactome

Plant Reactome is a freely accessible database of plant metabolic and regulatory pathways. Our goal is to provide plant researchers tools for visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education.

The development of Plant Reactome is supported the US National Science Foundation award (#1127112). The database is powered by the resources developed for the human Reactome project funded by the NIH (U41 HG003751) and the European Union 6th Framework Programme 'ENFIN'.

Gramene News

Current Gramene Release (49)

Gramene Database @GrameneDatabase 5h

Inviting everyone for The Plant Reactome Webinar. Tue Feb 16, 2016, at 2:00pm EST/11:00am PST. Register cshl.cmail20.com/t/t-l-drlukrt-...

Gramene Database @GrameneDatabase 8 Feb

Feb 16, 2016: Gramene Webinar: Updates on Plant Reactome Database
more info and registration link gramene.org/node/387

Gramene Database @GrameneDatabase 8 Feb

gramene.org/node/387

Tweet to @GrameneDatabase

Plant Reactome hosts:

- Curated metabolic & regulatory pathways in rice
- Orthology based pathway projections for 58 plant species.
- Provides tools for OMICS data analysis and visualization
- cross-species Pathway comparisons

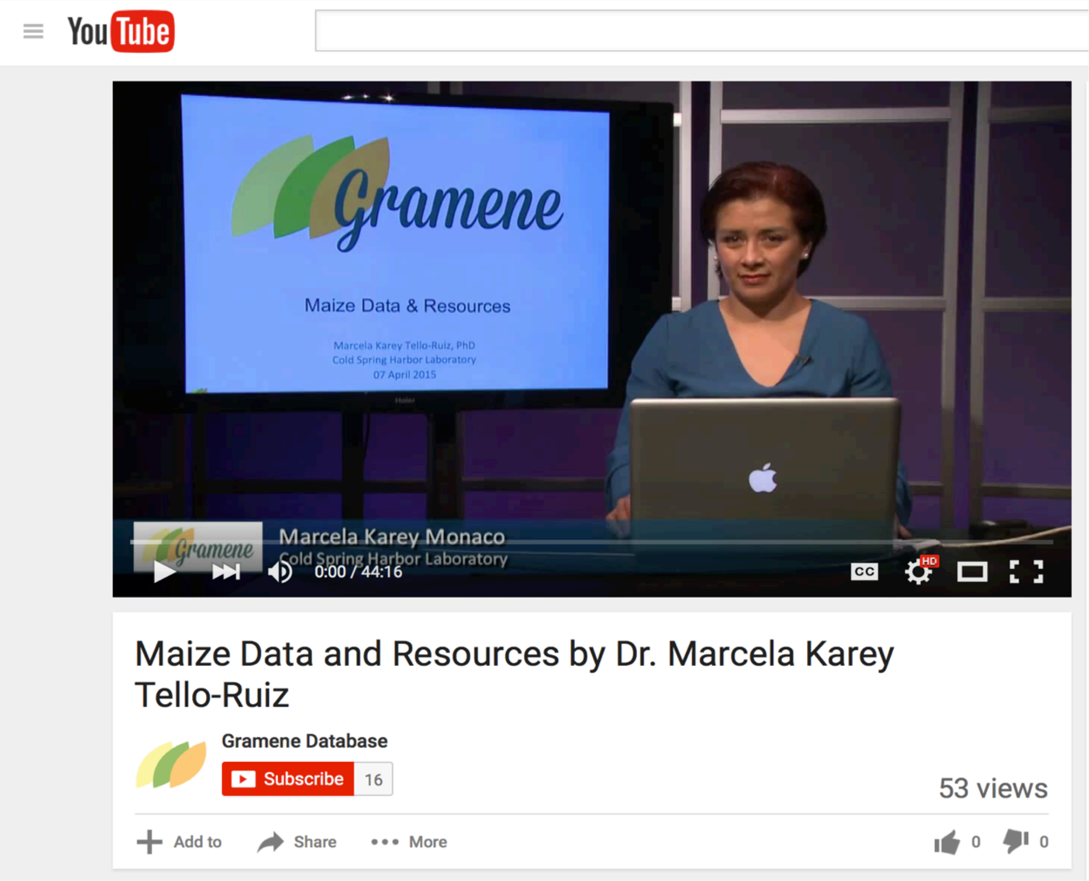
Reactome Model on Gramene Youtube channel

- https://www.youtube.com/watch?v=UmrQUClm3jc&feature=youtu.be

Future plans: Pathways

- Double the curated number of pathways in reference (rice) species rice and projected species
- Include more plant species

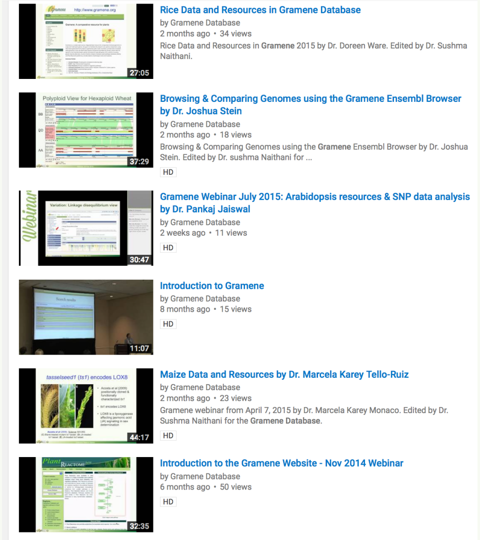
Gramene Database @ YouTube



Maize Data and Resources by Dr. Marcela Karey Tello-Ruiz

Gramene Database

53 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.

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 ...
HD

Gramene Webinar July 2015: Arabidopsis resources & SNP data analysis by Dr. Pankaj Jaiswal
by Gramene Database
2 weeks ago • 11 views
HD

Introduction to Gramene
by Gramene Database
8 months ago • 15 views
HD

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**.
HD

Introduction to the Gramene Website - Nov 2014 Webinar
by Gramene Database
6 months ago • 50 views
HD

- 12 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)

Requests to the community

- Beta testers of the new Gramene search
  gramene.org/beta
- Review pathways for accuracy & completeness
- Prioritize pathway curation (experimental evidence based)
- Webinar requests & training data sets from the Maize Community

What's in Gramene build 50?

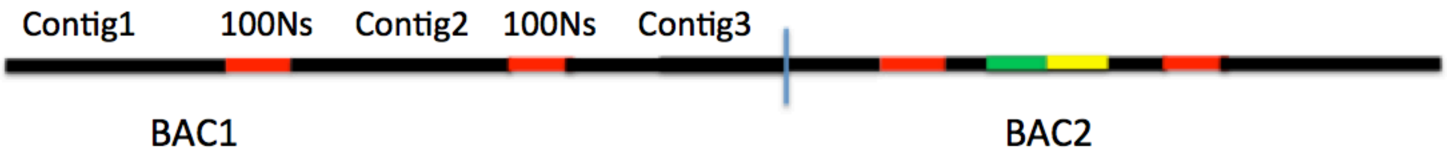
- 3 new genomes: *Trifolium pratense* (red clover), *Brassica napus* (rapeseed) & *Beta vulgaris* (beet)
- New rice variation data
- New WGAs for wheat
- New curated rice pathways
- Projected pathways for 3 new species: *Jatropha curcas, Manihot esculenta, Eucalyptus grandis & Mimulus guttatus*
- New detailed views -genetree, pathway, atlas - for search results

Maize B73 RefGen_v4 @Gramene

Yinping Jiao

Cold Spring Harbor Laboratory

Status of RefGen_v3



Assembly status

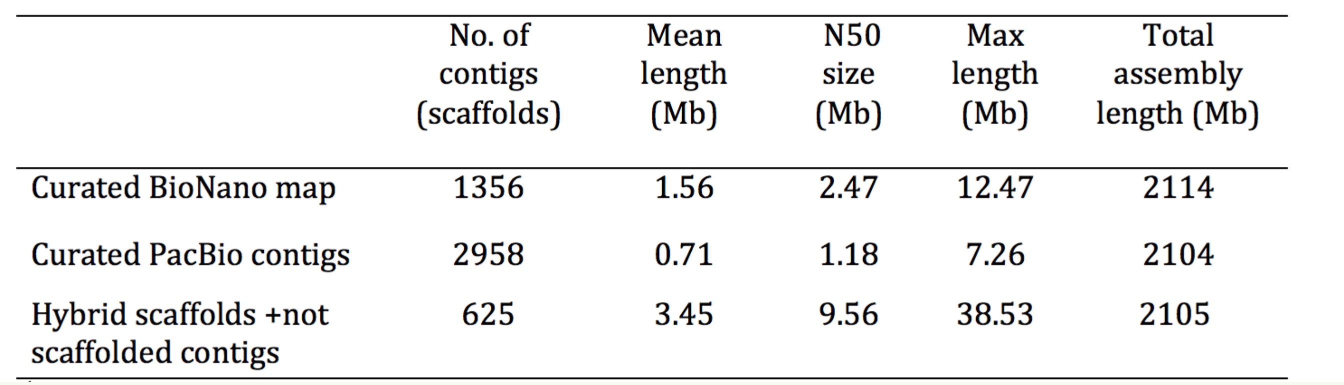
- Total bases: 2,066,912,289 – **missing ~200Mb**
- **No. of contigs** : 138,936
- **No. of gaps** estimated from psuedomolecule: 125,077
- Lots of the contigs were **not ordered or orientated**
- Assembly **errors within contigs**

Main issues:

- ***Order and orientation of contigs***
- ***Gaps***
- ***Incomprehensive annotation of alternative splicing***

Summary of RefGen_v4

- **De novo assembly with PacBio long reads and BioNano map**



	No. of contigs (scaffolds)	Mean length (Mb)	N50 size (Mb)	Max length (Mb)	Total assembly length (Mb)
Curated BioNano map	1356	1.56	2.47	12.47	2114
Curated PacBio contigs	2958	0.71	1.18	7.26	2104
Hybrid scaffolds +not scaffolded contigs	625	3.45	9.56	38.53	2105

- **Pseudomolecule construction**
  - total size: 2.10Gb
  - 2,523 gaps left : 1,115 gaps with estimated size from BioNano map
- **Gene annotation**
  - MAKER protein coding: 48,934
  - v3 genes missed from Maker projected to v4: 269
  - MAKER tRNA: 2,290:
  - lncRNA: 2,561

More details will present at the workshop at 5pm

Nomenclature

- Following Maize genetics Nomenclature
- **Genome:**

  <assembly_version_code> = <assembly code>, <version code>

  B73 RefGen_v4 - Zm00001d

- **Annotation:**

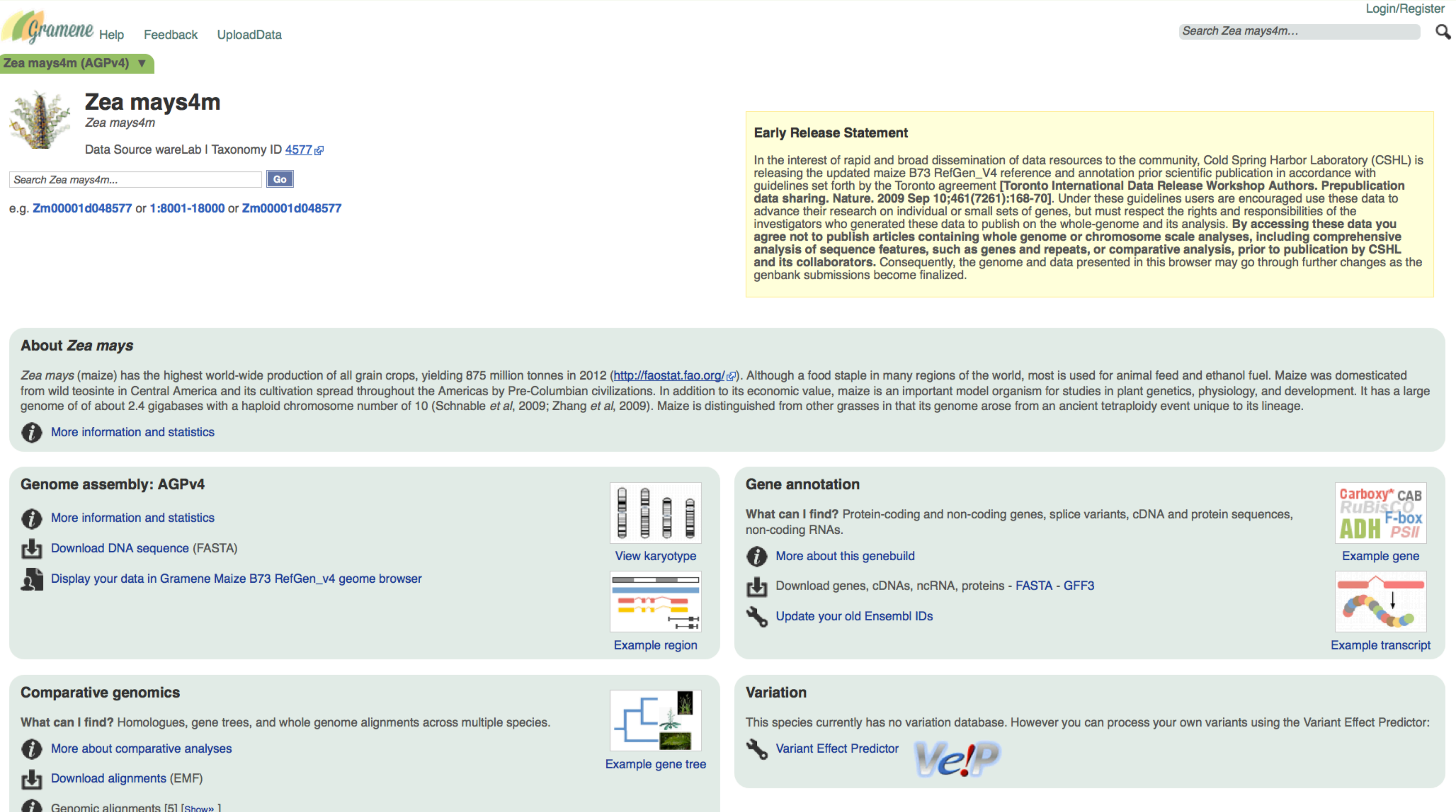
  <gene model> = <species id>, <assembly_code> , <version_code>, <six digits>

  Gene: Zm00001d000001

  Transcript: Zm00001d000001_T001

Coding and noncoding genes are numbered the same and this information is not encoded in the ID. This information will be included in metadata as "Biotype"

B73 RefGen_v4 @Gramene
http://maizev4.gramene.org/



Zea mays4m

Zea mays4m

Data Source wareLab | Taxonomy ID 4577

e.g. Zm00001d048577 or 1:8001-18000 or Zm00001d048577

Early Release Statement

In the interest of rapid and broad dissemination of data resources to the community, Cold Spring Harbor Laboratory (CSHL) is releasing the updated maize B73 RefGen_V4 reference and annotation prior scientific publication in accordance with guidelines set forth by the Toronto agreement **[Toronto International Data Release Workshop Authors. Prepublication data sharing. Nature. 2009 Sep 10;461(7261):168-70]**. Under these guidelines users are encouraged use these data to advance their research on individual or small sets of genes, but must respect the rights and responsibilities of the investigators who generated these data to publish on the whole-genome and its analysis. **By accessing these data you agree not to publish articles containing whole genome or chromosome scale analyses, including comprehensive analysis of sequence features, such as genes and repeats, or comparative analysis, prior to publication by CSHL and its collaborators.** Consequently, the genome and data presented in this browser may go through further changes as the genbank submissions become finalized.

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 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: AGPv4

More information and statistics

Download DNA sequence (FASTA)

Display your data in Gramene Maize B73 RefGen_v4 geome browser

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

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

More about comparative analyses

Download alignments (EMF)

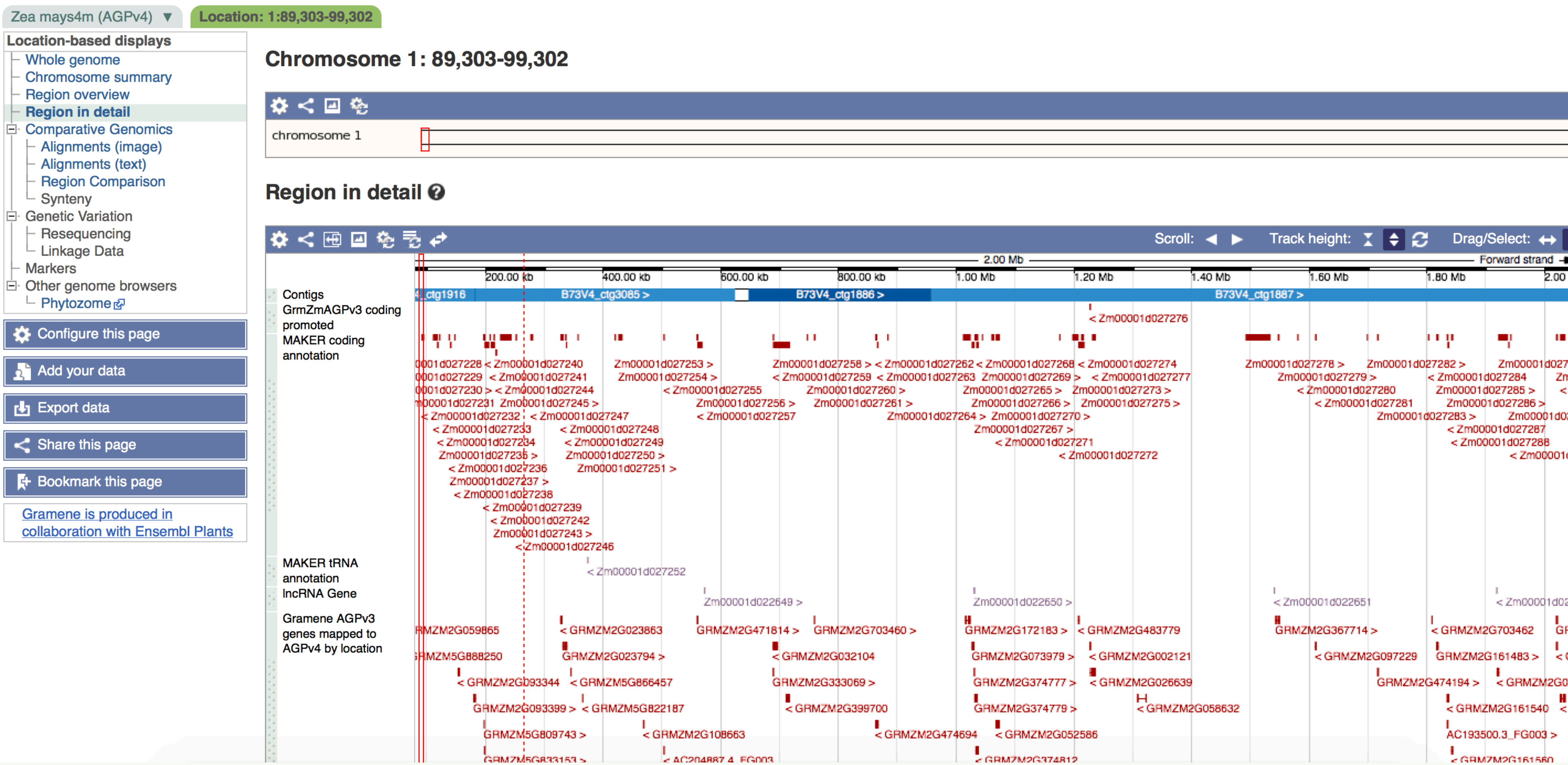
Genomic alignments [5] [Show»]

Variation

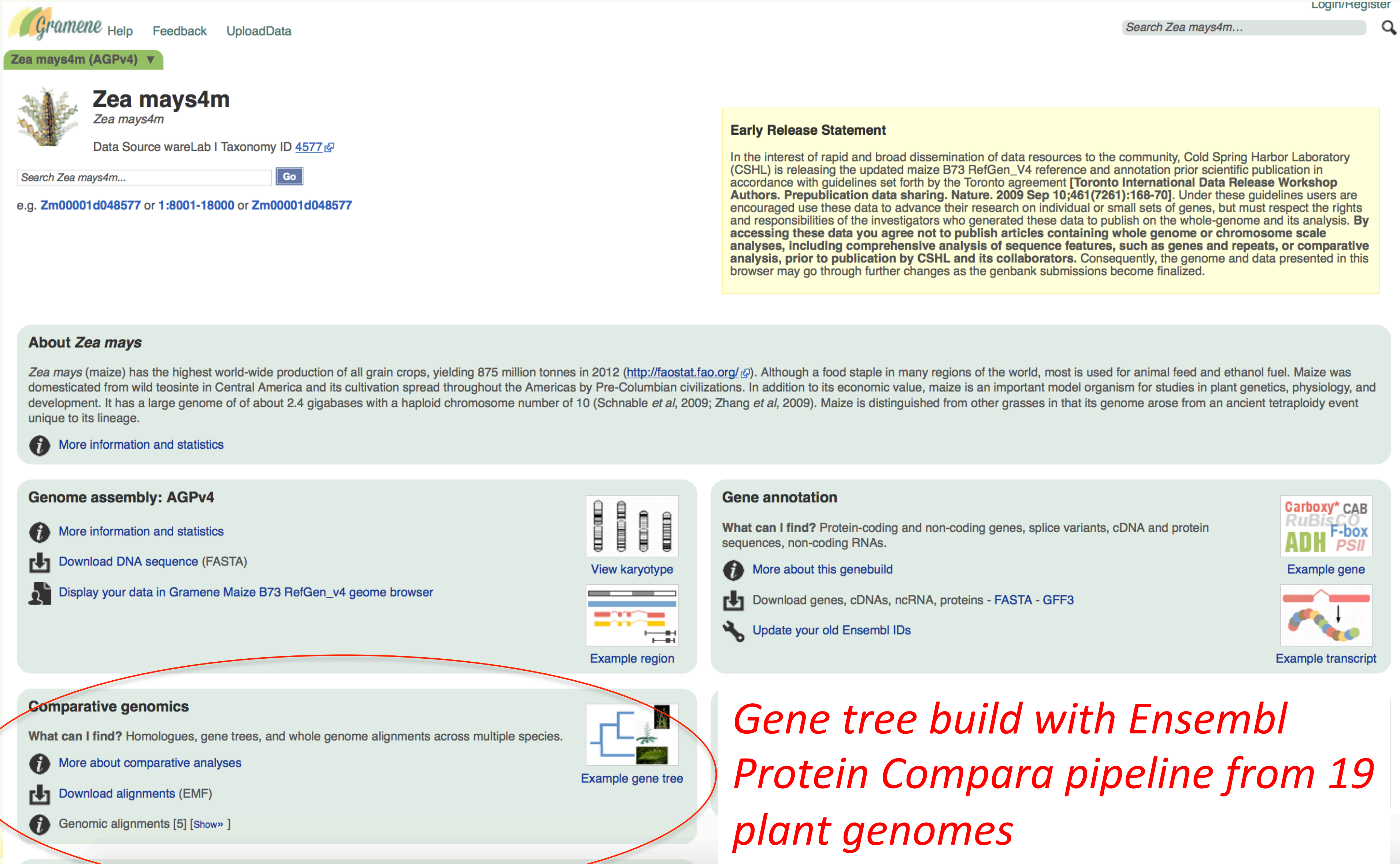
This species currently has no variation database. However you can process your own variants using the Variant Effect Predictor:

Variant Effect Predictor

V4 genome browser



Comparative genomics



Zea mays4m

Zea mays4m

Data Source wareLab | Taxonomy ID 4577

e.g. **Zm00001d048577** or **1:8001-18000** or **Zm00001d048577**

Early Release Statement

In the interest of rapid and broad dissemination of data resources to the community, Cold Spring Harbor Laboratory (CSHL) is releasing the updated maize B73 RefGen_V4 reference and annotation prior scientific publication in accordance with guidelines set forth by the Toronto agreement **[Toronto International Data Release Workshop Authors. Prepublication data sharing. Nature. 2009 Sep 10;461(7261):168-70]**. Under these guidelines users are encouraged use these data to advance their research on individual or small sets of genes, but must respect the rights and responsibilities of the investigators who generated these data to publish on the whole-genome and its analysis. **By accessing these data you agree not to publish articles containing whole genome or chromosome scale analyses, including comprehensive analysis of sequence features, such as genes and repeats, or comparative analysis, prior to publication by CSHL and its collaborators.** Consequently, the genome and data presented in this browser may go through further changes as the genbank submissions become finalized.

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 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: AGPv4

More information and statistics

Download DNA sequence (FASTA)

Display your data in Gramene Maize B73 RefGen_v4 geome browser

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

Comparative genomics

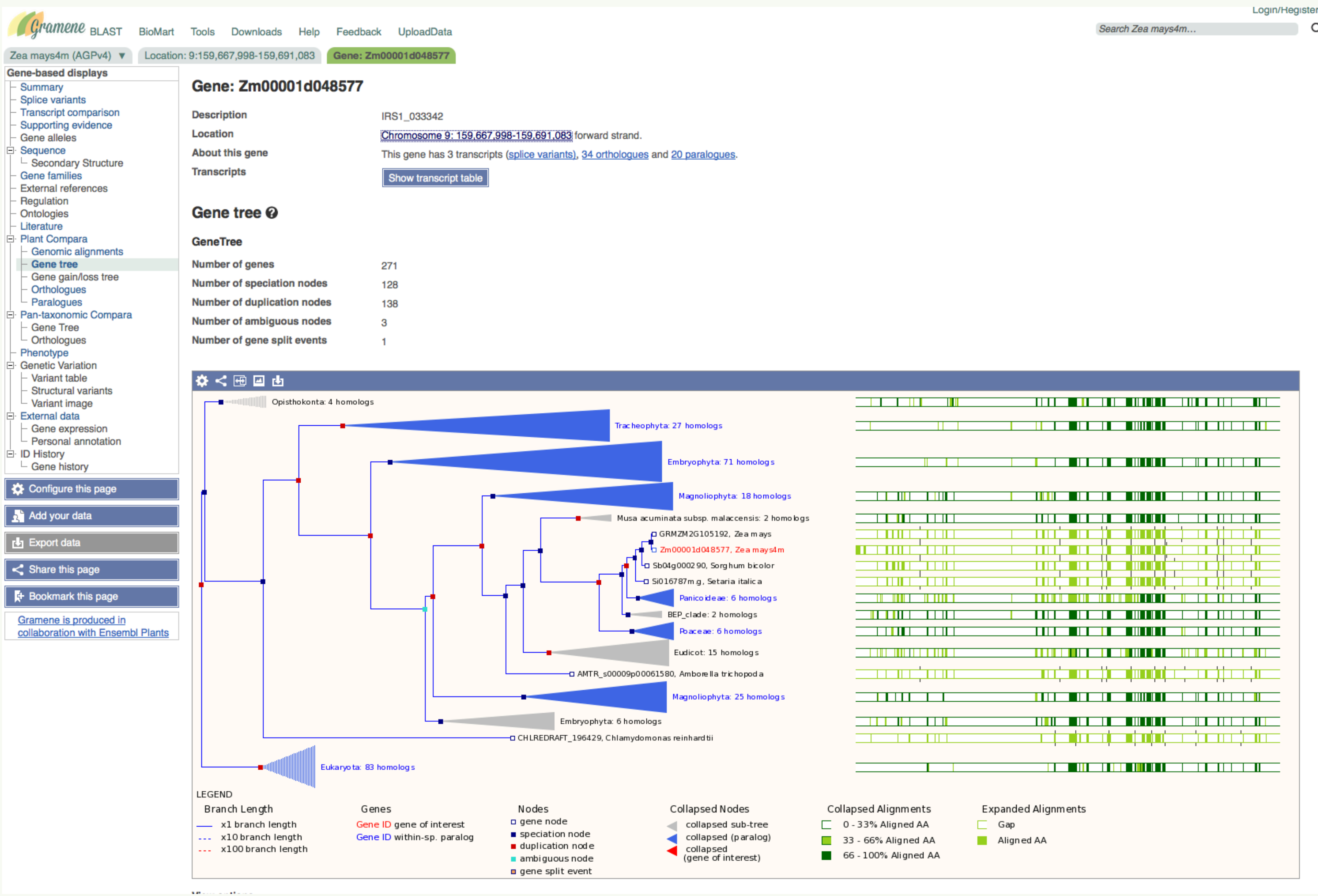
What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

More about comparative analyses

Download alignments (EMF)

Genomic alignments [5] [Show»]

Gene tree build with Ensembl Protein Compara pipeline from 19 plant genomes



Gene: Zm00001d048577

Description	IRS1_033342
Location	Chromosome 9: 159,667,998-159,691,083 forward strand.
About this gene	This gene has 3 transcripts (splice variants), 34 orthologues and 20 paralogues.
Transcripts	Show transcript table

Gene tree

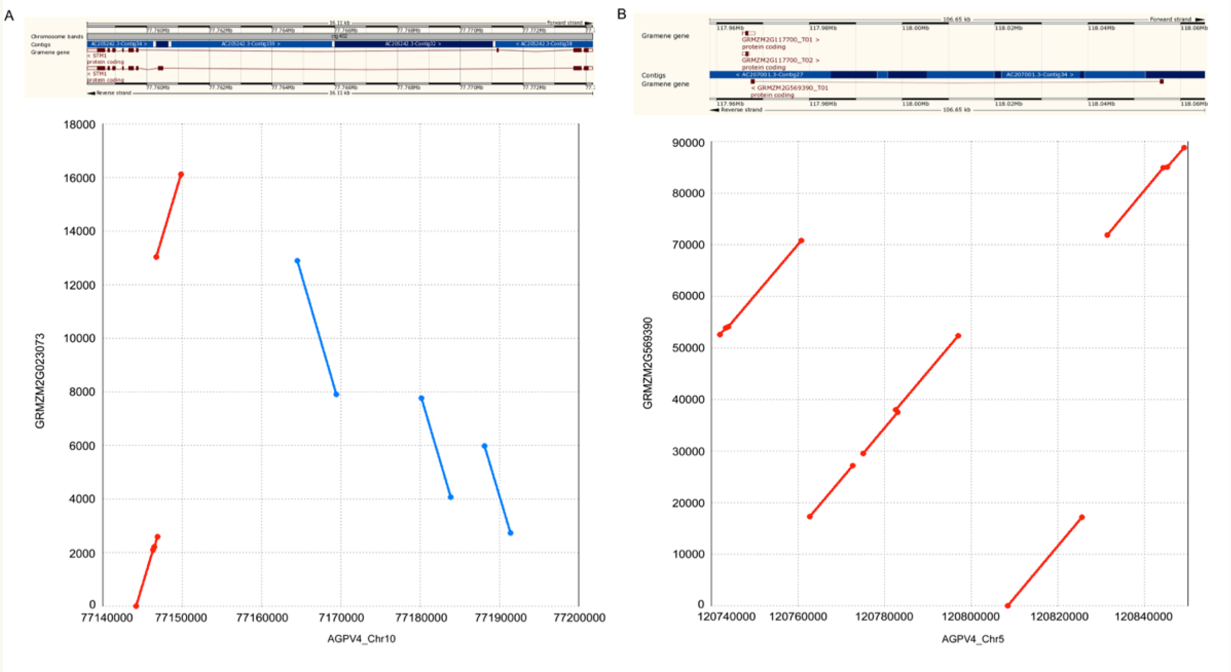
GeneTree

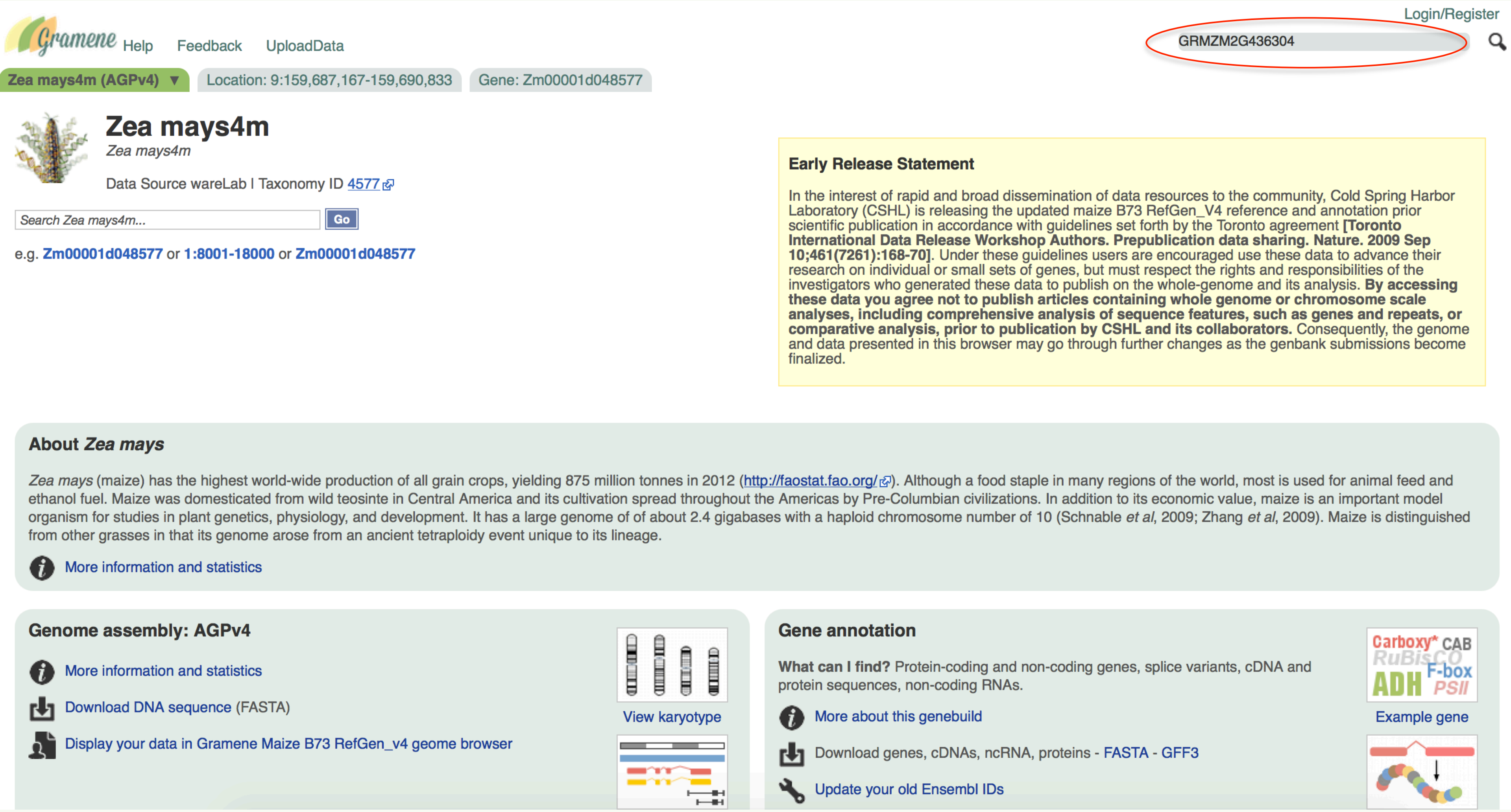
Number of genes	271
Number of speciation nodes	128
Number of duplication nodes	138
Number of ambiguous nodes	3
Number of gene split events	1

Whole Genome Alignments between Zea mays RefGen_v4 and Arabidopsis thaliana, Oryza sativa Japonica, Sorghum bicolor, Setaria italica, Vitis vinifera respectively

Search from v3 gene in v4 assembly

- *36,725 (94%) of 39,301 nuclear genes could be mapped to V4 without sequence change*
- *2,461 gene with sequence changes not included*
- *115 gene missed from v4*





GRMZM2G436304

Zea mays4m

Zea mays4m

Data Source wareLab | Taxonomy ID 4577

Search Zea mays4m... **Go**

e.g. **Zm00001d048577** or **1:8001-18000** or **Zm00001d048577**

Early Release Statement

In the interest of rapid and broad dissemination of data resources to the community, Cold Spring Harbor Laboratory (CSHL) is releasing the updated maize B73 RefGen_V4 reference and annotation prior scientific publication in accordance with guidelines set forth by the Toronto agreement **[Toronto International Data Release Workshop Authors. Prepublication data sharing. Nature. 2009 Sep 10;461(7261):168-70]**. Under these guidelines users are encouraged use these data to advance their research on individual or small sets of genes, but must respect the rights and responsibilities of the investigators who generated these data to publish on the whole-genome and its analysis. **By accessing these data you agree not to publish articles containing whole genome or chromosome scale analyses, including comprehensive analysis of sequence features, such as genes and repeats, or comparative analysis, prior to publication by CSHL and its collaborators.** Consequently, the genome and data presented in this browser may go through further changes as the genbank submissions become finalized.

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 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: AGPv4

- More information and statistics
- Download DNA sequence (FASTA)
- Display your data in Gramene Maize B73 RefGen_v4 geome browser

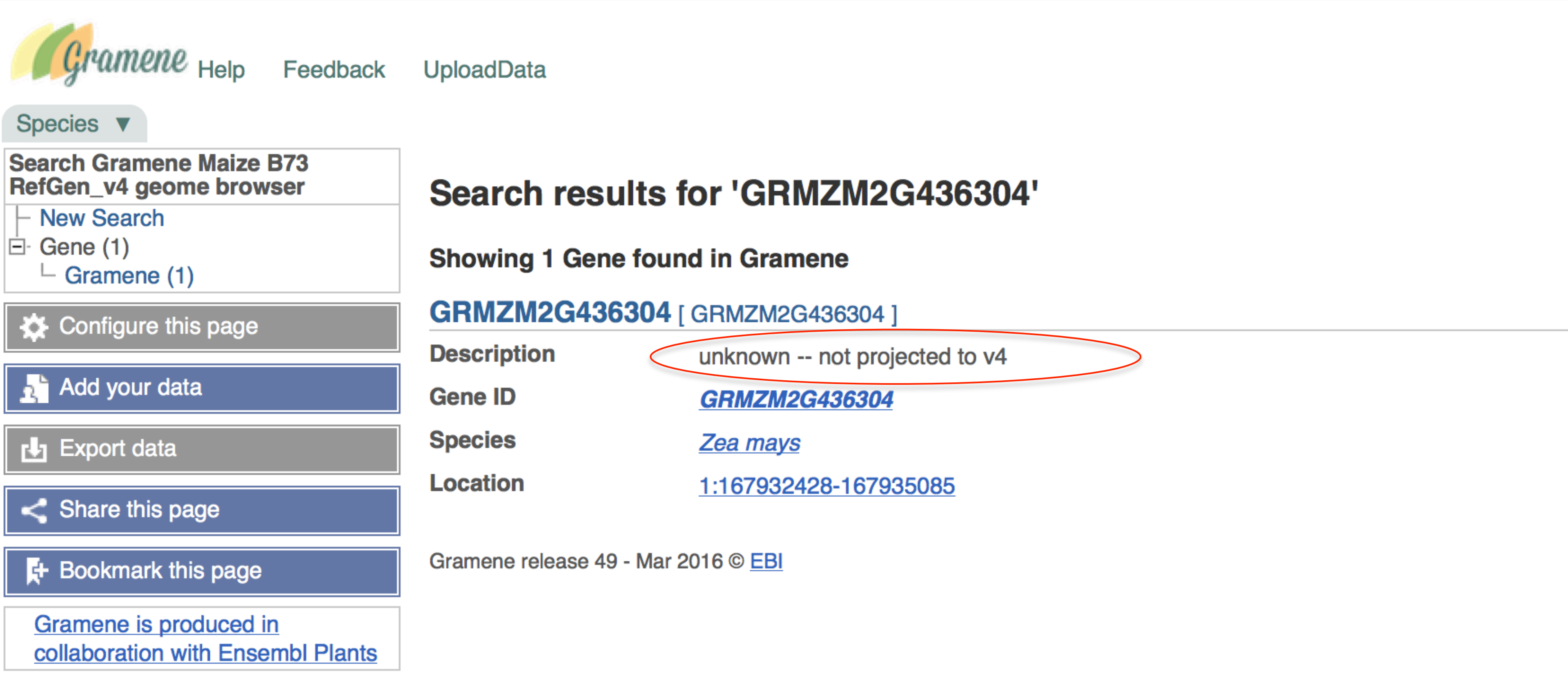
View karyotype

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



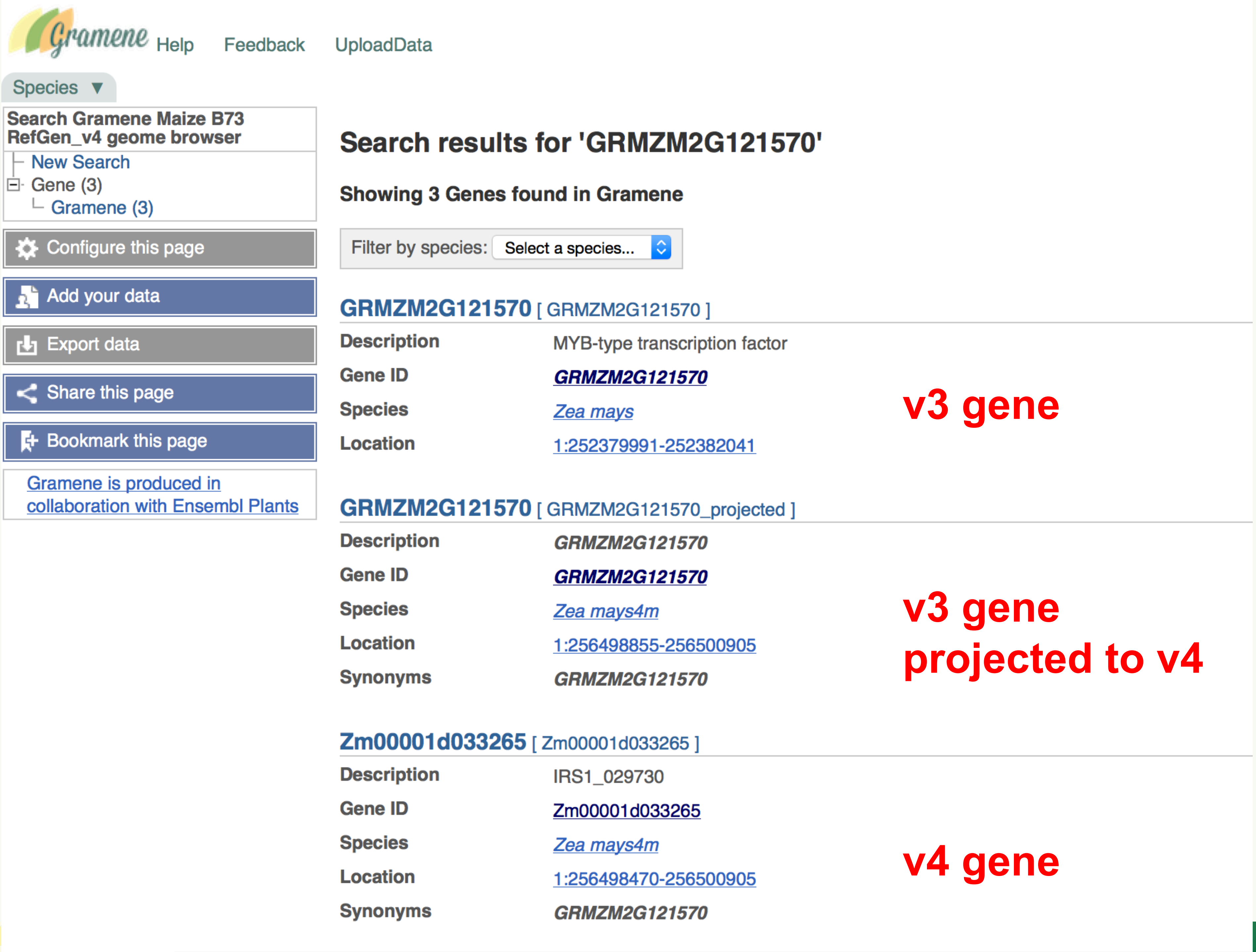
Search results for 'GRMZM2G436304'

Showing 1 Gene found in Gramene

GRMZM2G436304 [GRMZM2G436304]

Description	unknown -- not projected to v4
Gene ID	*GRMZM2G436304*
Species	*Zea mays*
Location	1:167932428-167935085

Gramene release 49 - Mar 2016 © EBI



Search results for 'GRMZM2G121570'

Showing 3 Genes found in Gramene

Filter by species: Select a species...

GRMZM2G121570 [GRMZM2G121570]

Description	MYB-type transcription factor
Gene ID	***GRMZM2G121570***
Species	*Zea mays*
Location	1:252379991-252382041

v3 gene

GRMZM2G121570 [GRMZM2G121570_projected]

Description	***GRMZM2G121570***
Gene ID	***GRMZM2G121570***
Species	*Zea mays4m*
Location	1:256498855-256500905
Synonyms	***GRMZM2G121570***

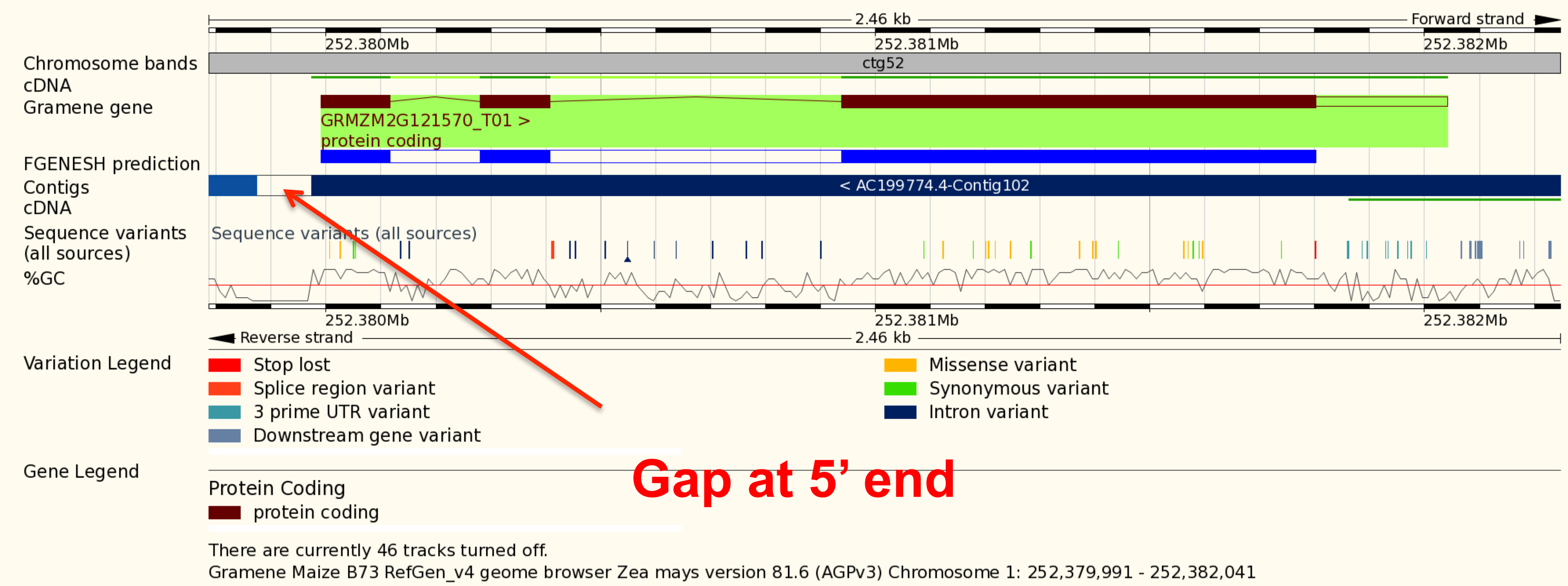
v3 gene projected to v4

Zm00001d033265 [Zm00001d033265]

Description	IRS1_029730
Gene ID	Zm00001d033265
Species	*Zea mays4m*
Location	1:256498470-256500905
Synonyms	***GRMZM2G121570***

v4 gene

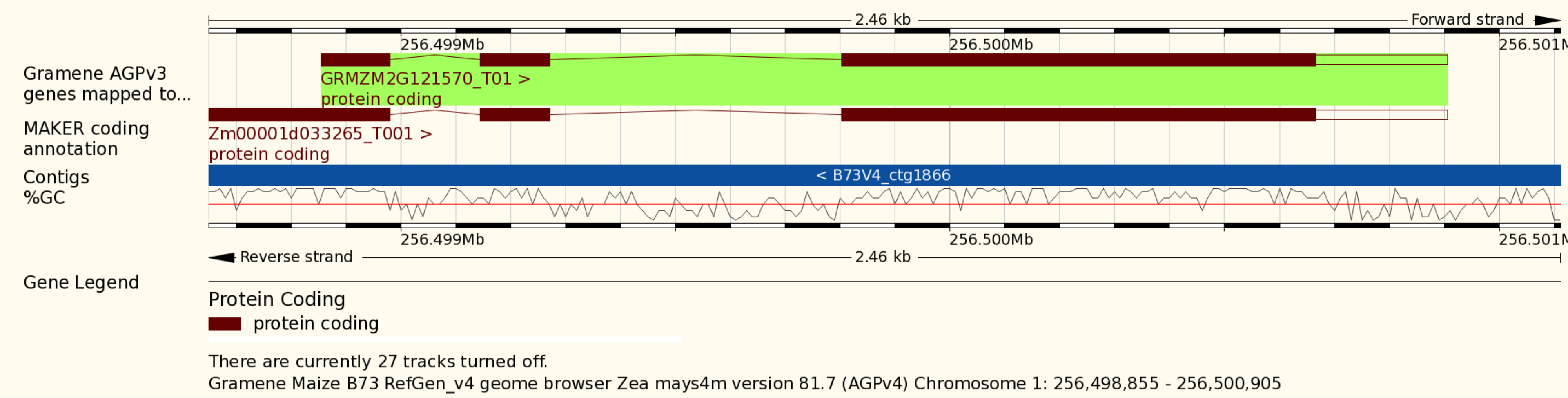
Incomplete gene model in v3



There are currently 46 tracks turned off.
Gramene Maize B73 RefGen_v4 geome browser Zea mays version 81.6 (AGPv3) Chromosome 1: 252,379,991 - 252,382,041

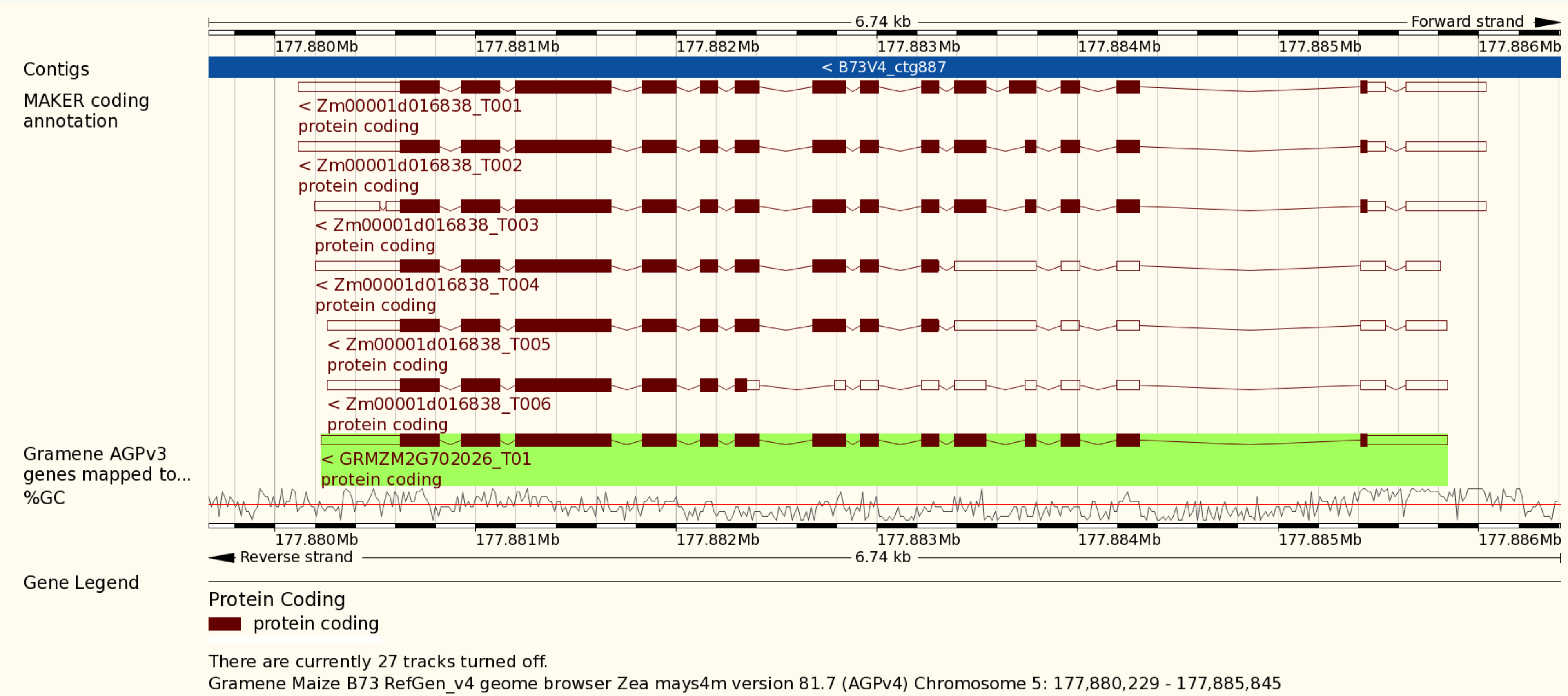
Gap at 5' end

Gap filled: complete gene model in v4



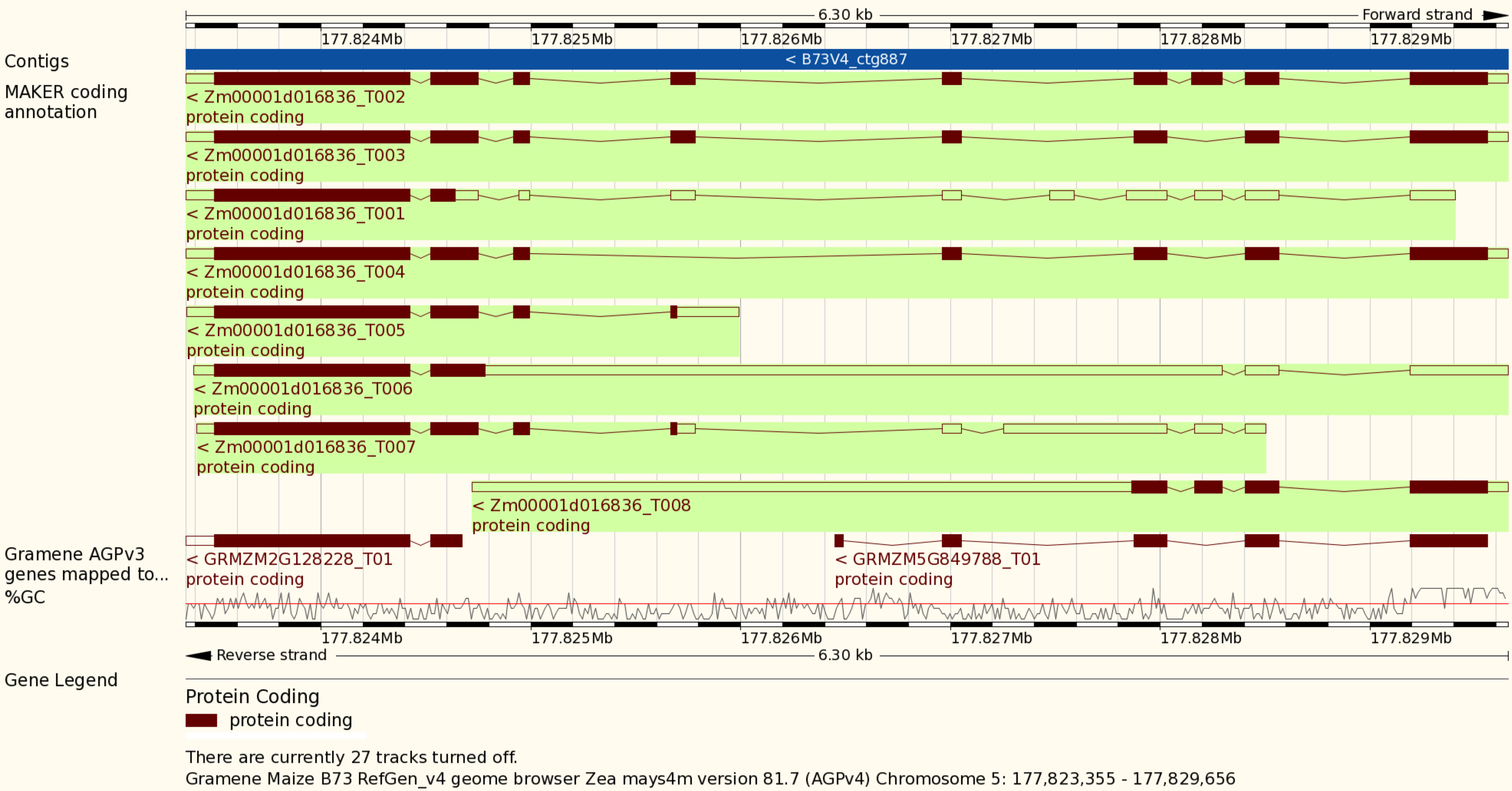
There are currently 27 tracks turned off.
Gramene Maize B73 RefGen_v4 geome browser Zea mays4m version 81.7 (AGPv4) Chromosome 1: 256,498,855 - 256,500,905

Single locus with the same model and new transcripts



There are currently 27 tracks turned off.
Gramene Maize B73 RefGen_v4 geome browser Zea mays4m version 81.7 (AGPv4) Chromosome 5: 177,880,229 - 177,885,845

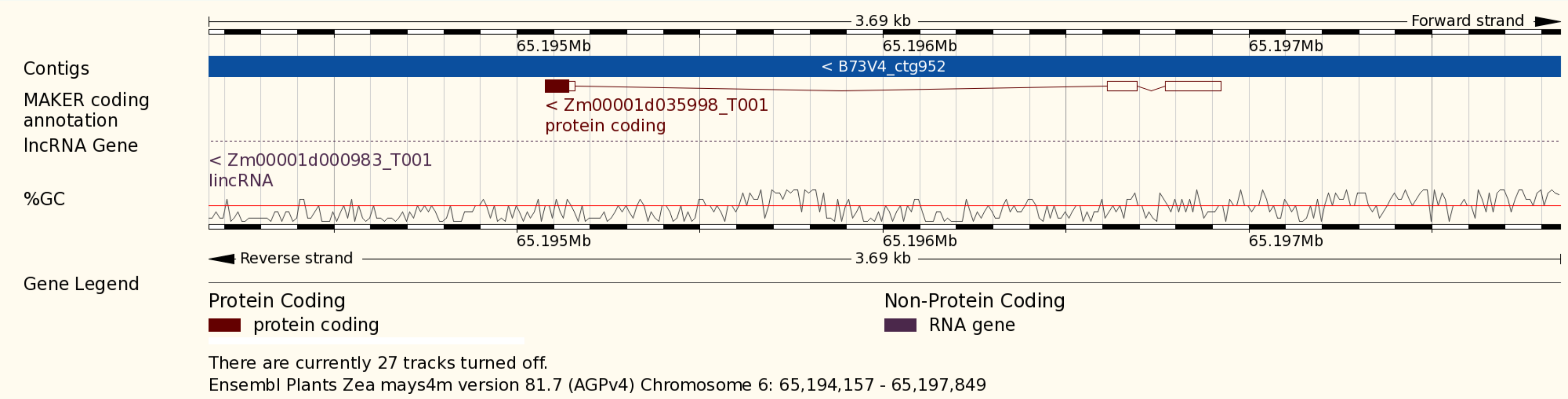
Two v3 gene models merged in v4



There are currently 27 tracks turned off.

Gramene Maize B73 RefGen_v4 geome browser Zea mays4m version 81.7 (AGPv4) Chromosome 5: 177,823,355 - 177,829,656

New gene model



There are currently 27 tracks turned off.

Ensembl Plants Zea mays4m version 81.7 (AGPv4) Chromosome 6: 65,194,157 - 65,197,849

Coming soon…

- Gene annotation evidence: full-length cDNA, RNA-seq…
- Regulatory data: expression, methylation ...
- Variation: GBS 2.7, Hapmap3.2…
- Tools: Genome assembly converter, BioMart…

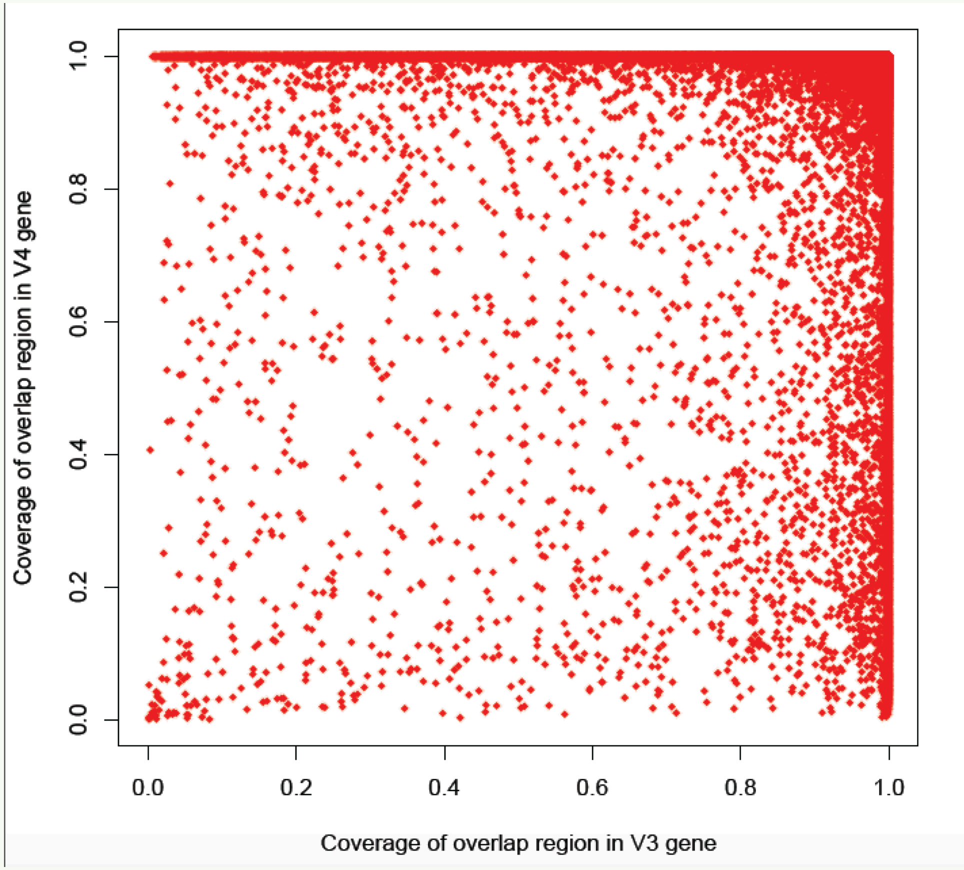
How to reach us

- E-mail: webinars@gramene.org
- Contact form: http://www.gramene.org/contact
- Announcements mailing list

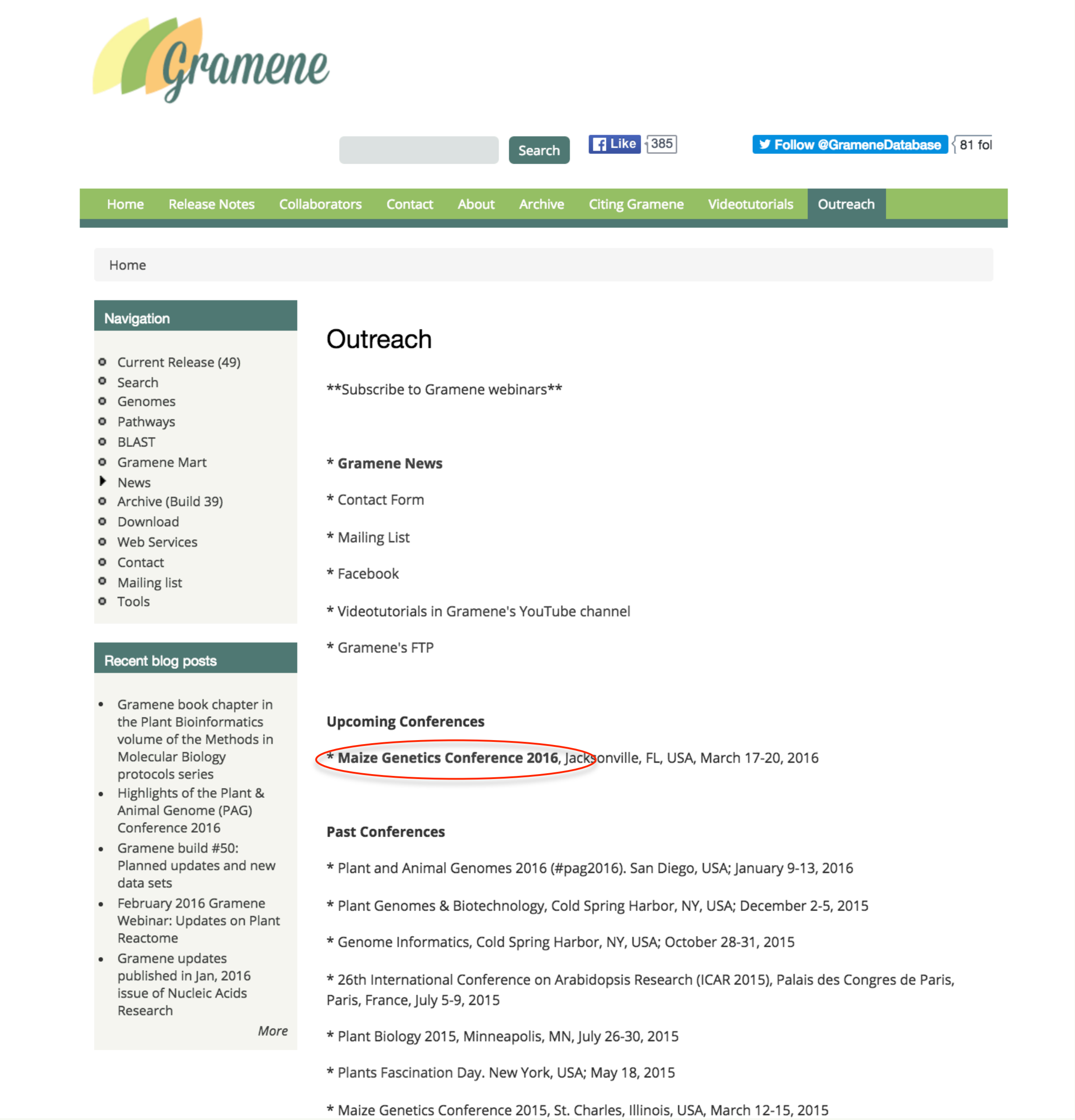
https://www.facebook.com/Gramene

@GrameneDatabase

Overlap between v3 FGS and v4 gene WGS



- Cutoff: Coverage >=0.5 in one version
- 31,474 V3 genes : 30,925 V4 genes
- 631 V3 genes were split
- 1,097 V4 genes covered more than one V3 genes
- 28,348 pairs of genes are 1-to-1



Navigation

- Current Release (49)
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Mailing list
- Tools

Recent blog posts

- Gramene book chapter in the Plant Bioinformatics volume of the Methods in Molecular Biology protocols series
- Highlights of the Plant & Animal Genome (PAG) Conference 2016
- Gramene build #50: Planned updates and new data sets
- February 2016 Gramene Webinar: Updates on Plant Reactome
- Gramene updates published in Jan, 2016 issue of Nucleic Acids Research

More

Outreach

Subscribe to Gramene webinars

* **Gramene News**

* Contact Form

* Mailing List

* Facebook

* Videotutorials in Gramene's YouTube channel

* Gramene's FTP

Upcoming Conferences

* **Maize Genetics Conference 2016**, Jacksonville, FL, USA, March 17-20, 2016

Past Conferences

* Plant and Animal Genomes 2016 (#pag2016). San Diego, USA; January 9-13, 2016

* Plant Genomes & Biotechnology, Cold Spring Harbor, NY, USA; December 2-5, 2015

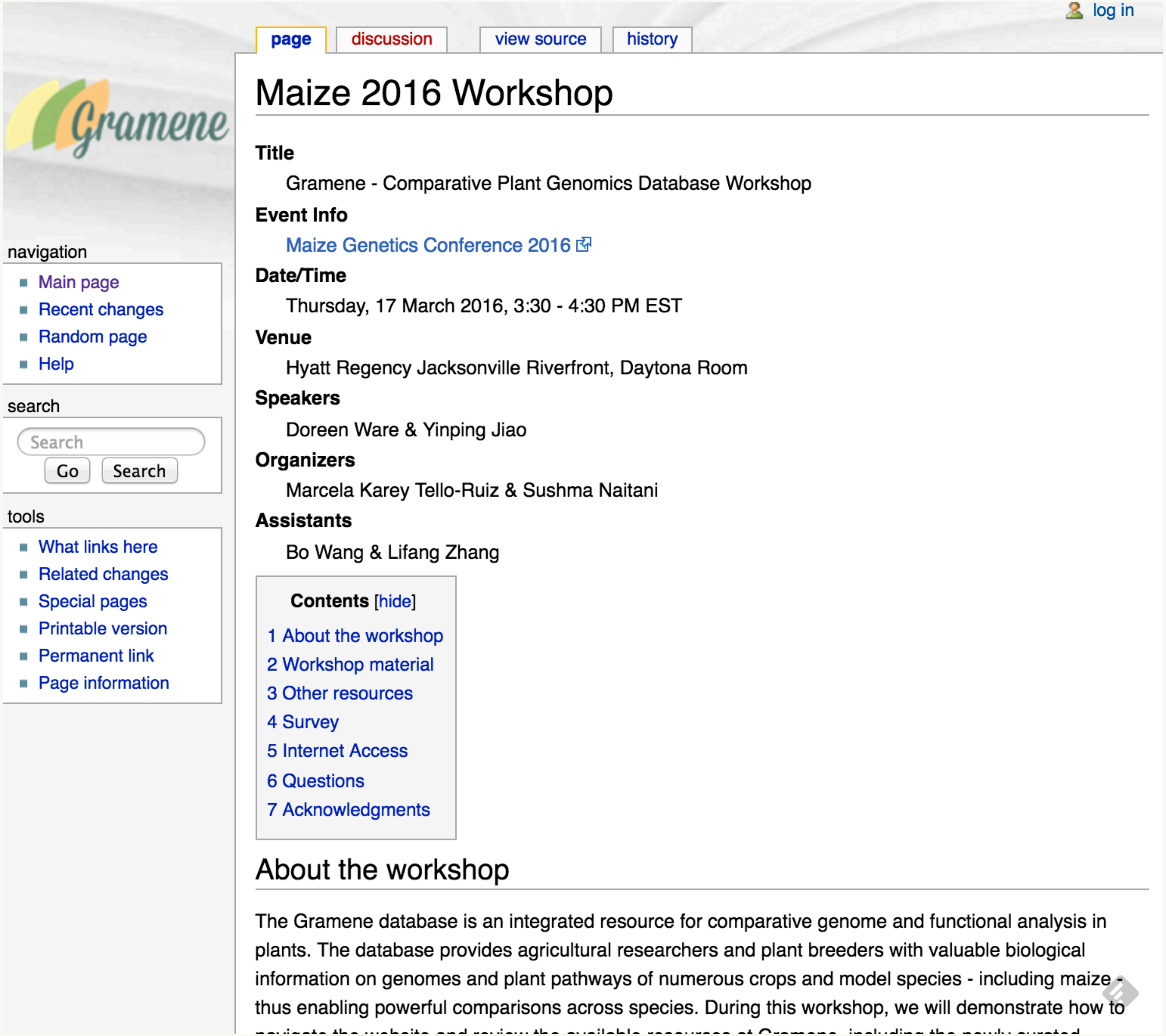
* Genome Informatics, Cold Spring Harbor, NY, USA; October 28-31, 2015

* 26th International Conference on Arabidopsis Research (ICAR 2015), Palais des Congres de Paris, Paris, France, July 5-9, 2015

* Plant Biology 2015, Minneapolis, MN, July 26-30, 2015

* Plants Fascination Day. New York, USA; May 18, 2015

* Maize Genetics Conference 2015, St. Charles, Illinois, USA, March 12-15, 2015



Maize 2016 Workshop

Title

Gramene - Comparative Plant Genomics Database Workshop

Event Info

Maize Genetics Conference 2016

Date/Time

Thursday, 17 March 2016, 3:30 - 4:30 PM EST

Venue

Hyatt Regency Jacksonville Riverfront, Daytona Room

Speakers

Doreen Ware & Yinping Jiao

Organizers

Marcela Karey Tello-Ruiz & Sushma Naitani

Assistants

Bo Wang & Lifang Zhang

Contents [hide]

About the workshop

The Gramene database is an integrated resource for comparative genome and functional analysis in plants. The database provides agricultural researchers and plant breeders with valuable biological information on genomes and plant pathways of numerous crops and model species - including maize - thus enabling powerful comparisons across species. During this workshop, we will demonstrate how to

How to reach us

- E-mail: webinars@gramene.org
- Contact form: http://www.gramene.org/contact
- Announcements mailing list

https://www.facebook.com/Gramene

@GrameneDatabase