



# 3<sup>rd</sup> Annual Grape RCN

June 2-5, 2011

Joshua Stein  
Cold Spring Harbor Laboratory (CSHL)  
Doreen Ware Lab

GRAMENE *Home*

[Search](#) [Genomes](#) [Species](#) [Download](#) [Resources](#) [About](#) [Help](#) [Feedback](#)

Release #31

May 2010  
[Release notes](#)

News

- [Scheduled downtime](#)  
Gramene will be offline for a few hours in the evening of Saturday, October 30, in order to perform...
- [Plant variation Mart now available](#)  
Gramene is happy to offer a new data set in our BioMart for variation data. This is a mirror of the...
- [Tassel version 3.0 release](#)  
For over a year, the Tassel software has been undergoing major design improvements leading to the...

[Subscribe to RSS Feed](#)

Have Questions...?

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

[Outreach calendar](#)  
[Presentation materials](#)

Gramene is a curated, open-source, data resource for comparative genome analysis in the grasses. Our goal is to facilitate the study of cross-species homology relationships using information derived from public projects involved in genomic and EST sequencing, protein structure and function analysis, genetic and physical mapping, interpretation of biochemical pathways, gene and QTL localization and descriptions of phenotypic characters and mutations.

Explore Gramene

Quick Search

Genomes

Genetic Diversity

Pathways

Proteins

Genes

Ontologies

Markers

Comparative Maps

QTL

BLAST

Gramene Mart

Species Pages

SPECIES:

# Metabolic Pathways in Plants

Pankaj Jaiswal  
Oregon State U.



GrapeCyc coming this fall!

**GRAMENE Home**

Search Genomes Species Download Resources About Help Feedback

Pathways Home | Search Pathways | Omics Viewer | Omics Validator | Downloads | Help | Tutorial | FAQs | Release notes

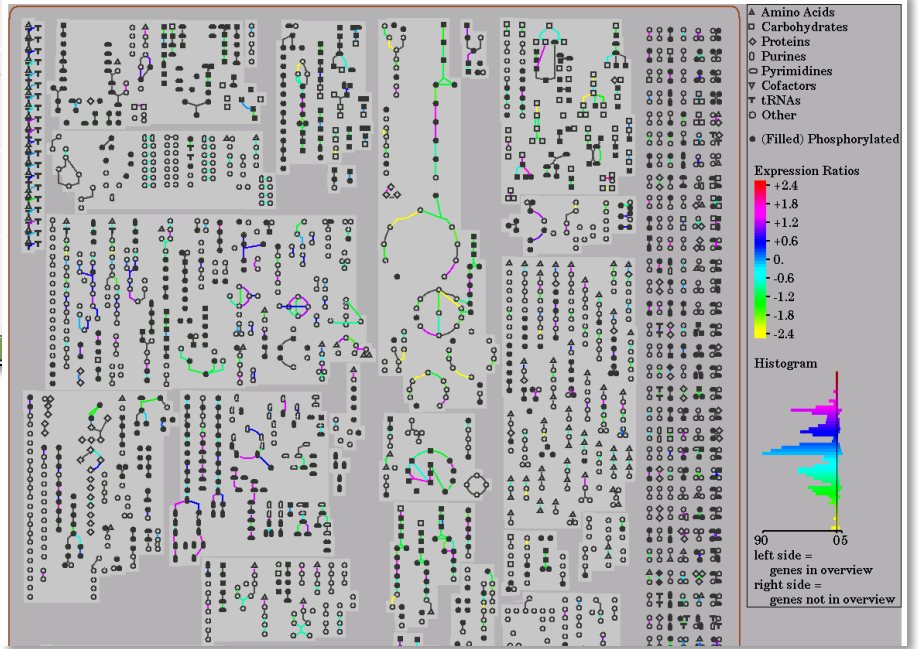
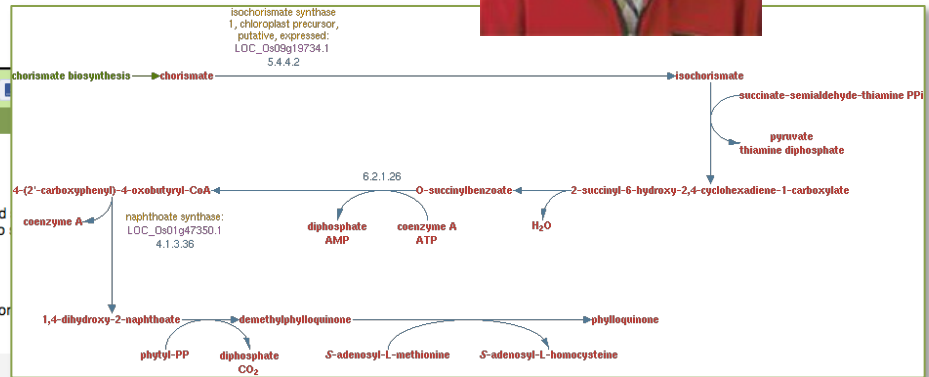
### Plant Metabolic Pathways

Pathways section in the Gramene databases is home for RiceCyc, MaizeCyc, BrachyCyc and SorghumCyc, the pathway databases for rice, maize, *Bracypodium*, and provides mirrors of pathway databases from *Arabidopsis*, tomato, potato, coffee, *Medicago*, *E. coli*, and the MetaCyc and PlantCyc reference databases. In addition to the database allows users to find genes mapped to respective reactions and pathways and draw inetrspecific comparison between the pathways.

#### Pathways Browse and Other Options




Click on the species specific links such as **browse** to go through the list of pathways; **summary** to get a summarized overview. Click on the **more info** link to learn more pathway database.

<b>RiceCyc</b> ver 3.1 <i>Oryza sativa japonica</i> Strain: Nipponbare <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>	<b>AraCyc*</b> ver 8.0 <i>Arabidopsis thaliana</i> Strain: Columbia <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>	<b>EcoCyc*</b> ver 15.0 <i>Escherichia coli</i> Strain: K-12 MG1655 <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>
<b>SorghumCyc</b> ver 1.0.1.1 <i>Sorghum bicolor</i> Strain: BTx623 <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>	<b>MediCyc*</b> ver 1.0.1 <i>Medicago truncatula</i> , Barreloclover Strain: n/a <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>	<b>MetaCyc*</b> ver 15.0 Reference Pathway Database Strain: not applicable <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>
<b>MaizeCyc**</b> ver 1.2.1.1 beta <i>Zea mize</i> Strain: B73 <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>	<b>PoplarCyc*</b> ver 3.0 <i>Populus trichocarpa</i> (and other Populus species and hybrids) Strain: not applicable <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>	<b>PlantCyc*</b> ver 5.0 Plant Metabolic Pathway Database Strain: not applicable <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>
<b>BrachyCyc**</b> ver 1.2.1.1 beta <i>Brachypodium distachyon</i> Strain: Bd21 <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>	<b>PotatoCyc*</b> ver 1.0.1.1 <i>Solanum tuberosum</i> , Potato Strain: n/a <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>	
	<b>CoffeaCyc*</b> ver 1.1.1 <i>Coffea canephora</i> , Coffee Strain: n/a <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More info</a>	
	<b>Lycocyc*</b> ver 2.0.1.1	



# Genetic Diversity

## Current Diversity Panels

- 
**Rice**
  - 160,000 SNPs x 21 varieties (incl. Nipponbare ref.) from OryzaSNP, MSU6.
- 
**Maize**
  - 1.6 million SNPs x 27 NAM founder lines from Panzea, AGPv1.
- 
**Arabidopsis**
  - 2010 Project SNP Discovery: 637,522 SNPs x 21 ecotypes (incl. Col-0 ref.), TAIR9.
  - 2010 Project 250K SNP chip genotypes v3.04, 214,000 SNPs x 1179 ecotypes, TAIR9.
  - 1001 Genomes/WTCHG SNPs from dbSNP, 2.7 million SNPs, 17 ecotypes, TAIR9.

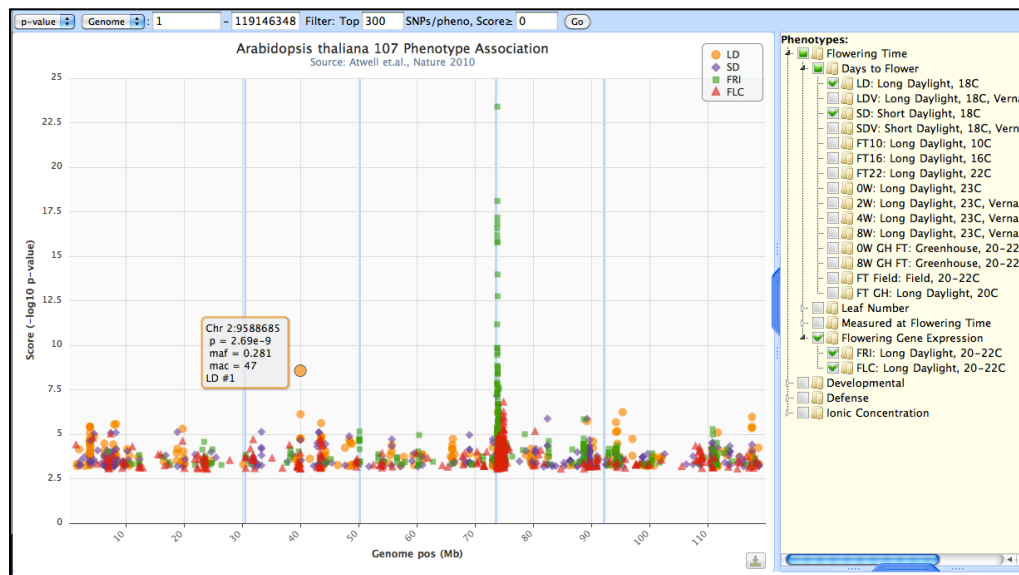


Susan McCouch

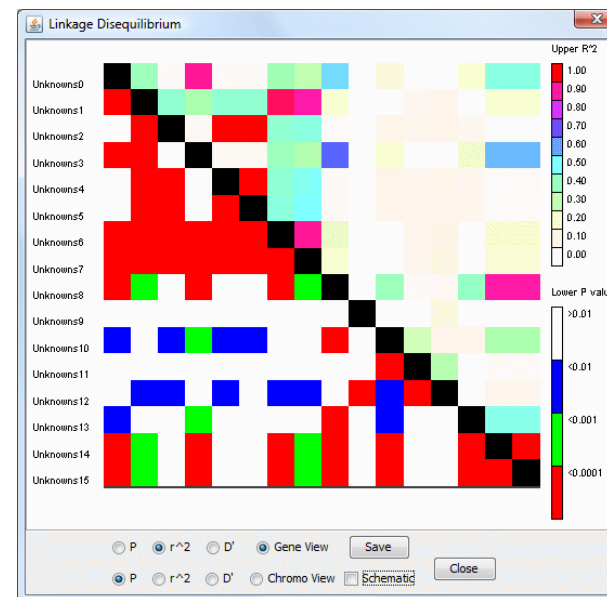


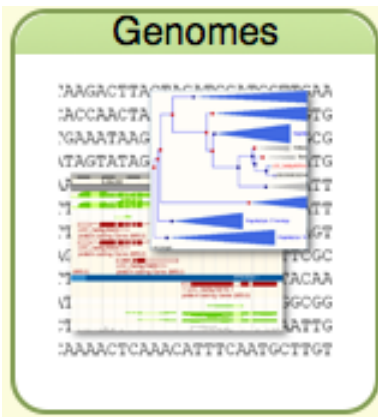
Ed Buckler

## GWAS Viewer



## Tassel





# Genomes



Doreen Ware  
CSHL

- Browser and visualizer
- Annotation
- Comparative genome analysis

Group	Species	Assembly	Annotation
Grasses	<i>Oryza sativa</i> ssp. japonica	MSU 6.0	MSU 6.0
	<i>Oryza sativa</i> ssp. indica	BGI-2005	BGI GLEAN 2008
	<i>Oryza glaberrima</i>	Oglab_v1	MIPS_v1
	<i>Brachypodium distachyon</i>	Brachy1.0	Brachy1.2
	<i>Zea mays</i>	RefGen_v2	CSHL 5a
	<i>Sorghum bicolor</i>	Sbi1	Sbi1.4
Eudicots	<i>Vitis vinifera</i>	IGGP_12X	Genoscope 2010 (v0)
	<i>Populus trichocarpa</i>	JGI 2.0	JGI 2.0
	<i>Arabidopsis thaliana</i>	TAIR10	TAIR10
	<i>Arabidopsis lyrata</i>	Araly1.2	Araly1.2
Basal land plant	<i>Physcomitrella patens</i>	v1.1	v1.1
Partial genomes	<i>Oryza brachyantha</i> (FF) 3S	brachyantha_454_AGP July 2010	CSHL_v2.1
	<i>Oryza glaberrima</i> (AA) 3S	BAC_Sanger_2009, Sep 2009	CSHL_v2.1
	<i>Oryza officinalis</i> (CC) 3S	Officinalis_3S Sep 2009	CSHL_v2.1
	<i>Oryza punctata</i> (BB) 3S	Punctata_3S Sep 2009	CSHL_v2.1
	<i>Oryza minuta</i> (BBCC) 3S	Minuta_CC_3S Sep 2009	CSHL_v2.1
	<i>Oryza barthii</i> (AA) 3S	BAC_pool_2008	CSHL_v2.1
	<i>Oryza nivara</i> (AA) 3S	nivara_454_AGP July 2010	CSHL_v2.1
	<i>Oryza rufipogon</i> (AA) 3S	rufipogon_454_AGP July 2010	CSHL_v2.1

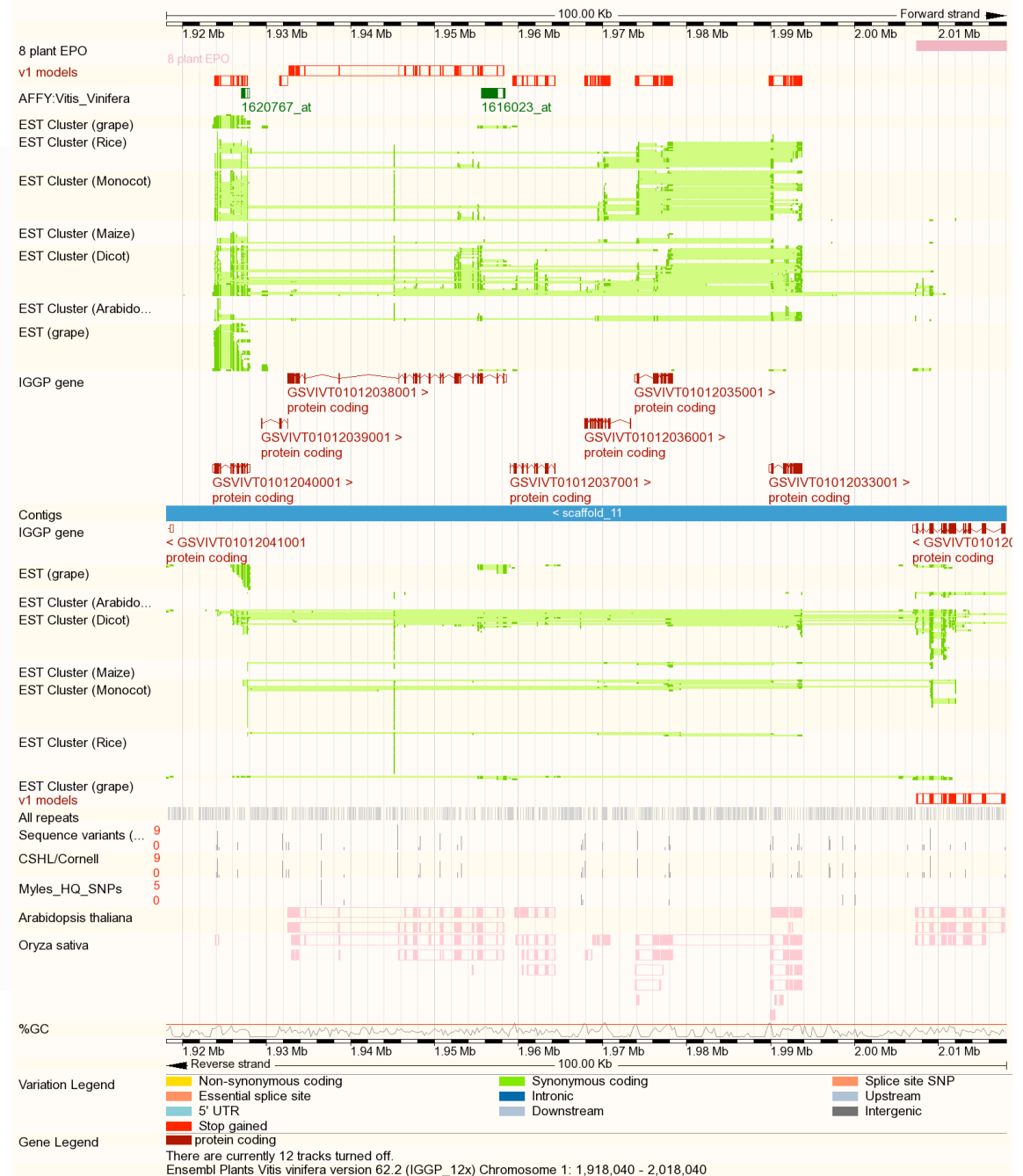
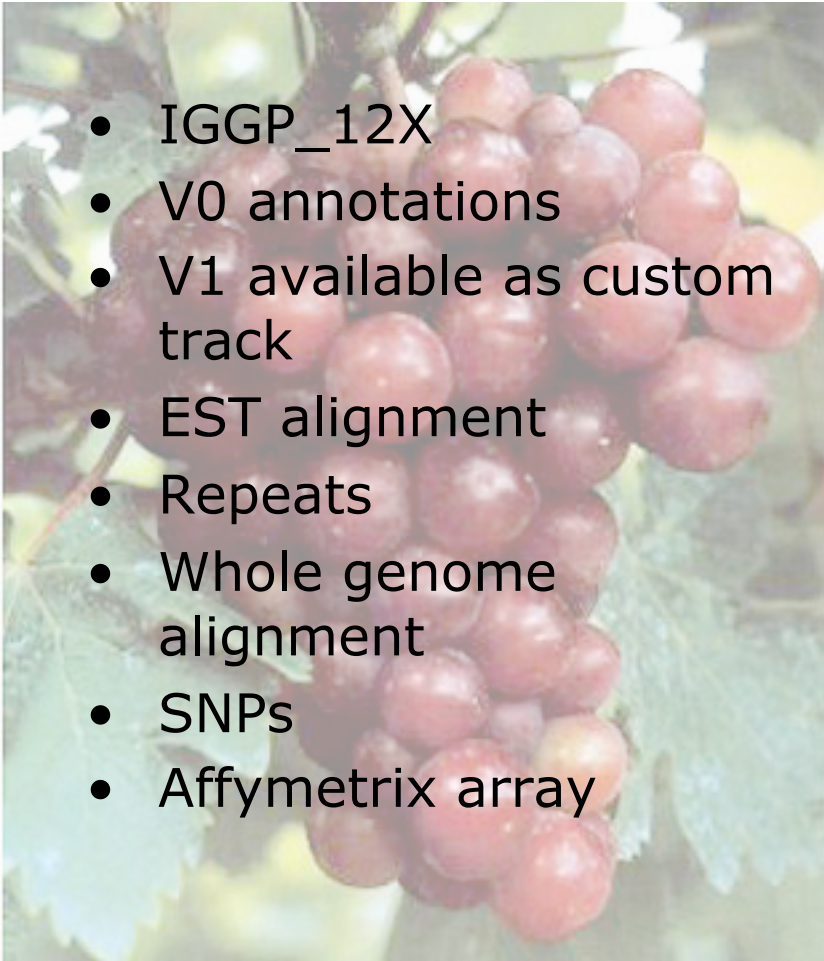
This Fall

- Tomato
- Soybean

Release 33 (May 2011)

# The Grape Browser

- IGGP\_12X
- V0 annotations
- V1 available as custom track
- EST alignment
- Repeats
- Whole genome alignment
- SNPs
- Affymetrix array



To view v1, add custom track:

[http://brie.cshl.edu/grape\\_rcn/all\\_cds.jigsawgaze\\_integrato\\_no\\_repeat.gff](http://brie.cshl.edu/grape_rcn/all_cds.jigsawgaze_integrato_no_repeat.gff)

**e!GRAMENE!** BLAST | BioMart | Documentation | Help | Feedback

Vitis vinifera (IGGP\_12x) Location: 7:381,035-383,836 Gene: GSVIVG01000752001 Trans: GSVIVT01000752001

### About this species

- Description
- Genome Statistics
  - Assembly and GeneBuild
  - Top 40 InterPro hits
  - Top 500 InterPro hits
- Sample entry points
- Karyotype
- Location (7:381035-383836)
- Gene (D1H9V1\_VITV)
- Transcript (D1H9V1\_VIT)
- Variation (ENSVVV0027)

**Search Vitis vinifera**

### Description

**About Vitis vinifera**

*Vitis vinifera* is the most widely cultivated and economically important grape species. It has a diploid genome with haploid chromosome number of 19, and an estimated genome size of 500 Mb.

**Genome Sequencing and Gene Prediction (release IGGP)**

This release is based on a 12x whole genome shotgun sequence assembly and annotation of the *Vitis vinifera* genome. These data were prepared by a French-Italian Public Consortium for Grapevine Genome Characterization under the auspices of the [International Grape Genome Program \(IGGP\)](#). Further details of the sequencing and assembly are available from [Genoscope](#).

**Gramene/Ensembl Genomes Annotation**

Additional annotations generated by the Gramene/Ensembl Genomes projects include:

- The standard set of Gramene analyses detailed [here](#).
- Mapping to the genome of multiple sequence-based feature sets.
- Phylogenetic gene trees and whole-genome alignments with several plant and metazoan species, using Ensembl's [Compara pipeline](#).
- Single nucleotide polymorphisms identified by re-sequencing a collection of grape cultivars and wild *Vitis* species from the [USDA germplasm collection](#).
- Probes from the [GeneChip Vitis vinifera Genome Array](#) was imported to the Ensembl functional genomics schema

This release of the assembly (IGGP 12X) for *Vitis vinifera* is composed of 2059 scaffolds with a total length of 485 Mb assembled into 33 chromosomes and chromosomal fragments with a total length of 486 Mb

- 33 chromosomes and chromosomal fragments with a total length of 486 Mb
- 2059 scaffolds with a total length of 485 Mb

**Links**

- [International Grape Genome Program](#)
- [Genoscope Grape Genome Browser](#)

Find:  ontol     Match case

**e!GRAMENE!** BLAST | BioMart | Documentation | Help | Feedback

Vitis vinifera (IGGP\_12x) Location: 7:381,035-383,836 Gene: GSVIVG01000752001 Trans: GSVIVT01000752001

### Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail
- Comparative Genomics
  - Alignments (image) (3)
  - Alignments (text) (3)
  - Multi-species view (2)
  - Synteny (3)
- Genetic Variation
  - Resequencing
  - Linkage Data
  - Markers
- Other genome browsers
  - Phylozome

**Chromosome 7: 381,035-383,836**

chromosome 7

**Region in detail**

Contigs IGGP.gene

Gene Legend

Location:

EST (grape) IGGP.gene Contigs EST (grape) %GC

Find:  ontol     Match case

**e!GRAMENE!** BLAST | BioMart | Documentation | Help | Feedback

Vitis vinifera (IGGP\_12x) Location: 7:381,035-383,836 Gene: GSVIVG01000752001 Trans: GSVIVT01000752001

### Gene-based displays

- Gene summary
- Splice variants (1)
- Supporting evidence
- Sequence
- External references
- Regulation
- Expression
- Literature
- Plant Compara
  - Genomic alignments
  - Gene Tree (image)
  - Gene Tree (text)
  - Gene Tree (alignment)
  - Orthologues (6)
  - Paralogues (12)
  - Protein families
- Pan-taxonomic Compara
  - Gene Tree (image)
  - Gene Tree (text)
  - Gene Tree (alignment)
  - Orthologues (3)
  - Paralogues (12)
  - Protein families
- Genetic Variation
  - Variation Table
  - Variation Image
- External Data
- ID History
- Gene history

**Gene: GSVIVG01000752001**

**Description** Whole genome shotgun sequence of line PN40024, scaffold\_141.assembly12x [Source: Uniprot/SPTREMBL\\_DSQMG\\_VITV](#)

**Location** Chromosome 7: 381,035-383,836 forward strand.

**Transcripts** There is 1 transcript in this gene

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype
Novel	GSVIVT01000752001	1583	GSVIVT01000752001	284	Protein coding

**Transcript and Gene level displays**

In Ensembl Plants we provide displays at two levels:

- Transcript views which provide information specific to an individual transcript such as the cDNA and CDS sequences and protein domain annotation.
- Gene views which provide displays for data associated at the gene level such as orthologues, paralogues, regulatory regions and splice variants.

This view is a gene level view. To access the transcript level displays select a Transcript ID in the table above and then navigate to the information you want using the menu at the left hand side of the page. To return to viewing gene level information click on the Gene tab in the menu bar at the top of the page.

**Gene summary**

Gene type Known protein coding

Prediction Method Gene annotation by IGGP through a process of automatic and manual curation

Find:  ontol     Match case

**e!GRAMENE!** BLAST | BioMart | Documentation | Help | Feedback

Vitis vinifera (IGGP\_12x) Location: 7:381,035-383,836 Gene: GSVIVG01000752001 Trans: GSVIVT01000752001

### Transcript-based displays

- Transcript summary
- Supporting evidence
- Sequence
  - Exons (5)
  - cDNA
  - Protein
- EBI Protein Database
- Protein Structure
- External References
  - General identifiers (5)
  - Oligo probes
- Ontology
  - Ontology chart (4)
  - Ontology table (4)
- Genetic Variation
  - Population comparison
  - Comparison image
- Protein Information
  - Protein summary
  - Domains & features (4)
  - Variations (10)
- External Data
- ID History
- Transcript history
- Protein history

**Transcript: GSVIVT01000752001**

**Description** Whole genome shotgun sequence of line PN40024, scaffold\_141.assembly12x [Source: Uniprot/SPTREMBL\\_DSQMG\\_VITV](#)

**Location** Chromosome 7: 381,035-383,836 forward strand.

**Gene** This transcript is a product of gene [GSVIVG01000752001](#) - There is 1 transcript in this gene

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype
Novel	GSVIVT01000752001	1583	GSVIVT01000752001	284	Protein coding

**Transcript and Gene level displays**

Views in Ensembl Plants are separated into gene based views and transcript based views according to which level the information is more appropriately associated with. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

**Transcript summary**

Statistics Exons: 5 Transcript length: 1,583 bps Translation length: 284 residues

Type Known protein coding

Prediction Method Gene annotation by IGGP through a process of automatic and manual curation

Gramene release 33 - April 2011 © EBI [About Ensembl/Genomes | Contact Us | Help](#)

Ensembl Plants is produced in collaboration with Gramene

Find:  ontol     Match case

# Analysis & Visualization



- InterPro/GO/dbXref
- Whole Genome Alignment (Blastz-chain-net)
- Phylogenetic Gene Trees (Compara)
- Synteny Analysis
- Consequences of SNP



# InterPro/dbXref/GO

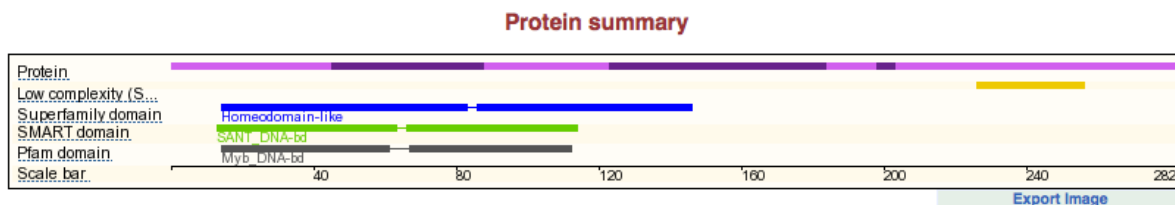
- Structural prediction: Pfam, PIRSF, PRINTS, PROSITE, SMART, SUPERFAMILY, TIGRFAM, TMHMM, SignalP
- Cross-reference genes to 3<sup>rd</sup> party identifiers: Entrez Gene, PlantGDB, PUTs, RefSeq, Gene Index, UniGene, UniProt
- Gene Ontology

**Domains & features**

**Domains**

Show/hide columns Filter:

Domain type	Start	End	Description	Accession	InterPro
Superfamily	14	83	Homeodomain-like	46689	<a href="#">IPR009057 [Display all genes with this domain]</a>
Superfamily	86	146	Homeodomain-like	46689	<a href="#">IPR009057 [Display all genes with this domain]</a>
Pfam	14	61	Myb_DNA-bd	PF00249	<a href="#">IPR014778 [Display all genes with this domain]</a>
Pfam	67	112	Myb_DNA-bd	PF00249	<a href="#">IPR014778 [Display all genes with this domain]</a>
Smart	13	63	SANT_DNA-bd	SM00717	<a href="#">IPR01005 [Display all genes with this domain]</a>
Smart	66	114	SANT_DNA-bd	SM00717	<a href="#">IPR01005 [Display all genes with this domain]</a>



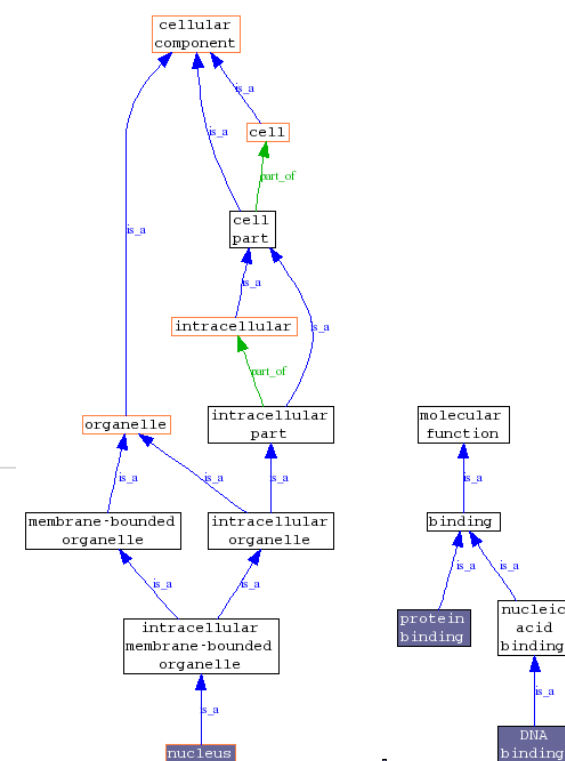
**General identifiers**

This transcript corresponds to the following database identifiers:

Show/hide columns

External database type	Database identifier
protein_id	<a href="#">CBI22245.3 [align] [view all locations]</a>
EMBL	<a href="#">FN595239 [align] [view all locations]</a>
UniProtKB/TrEMBL	<a href="#">D7SXEB_VITV [Target %id: 100; Query %id: 100] [align]</a> Whole genome shotgun sequence of line PN40024, scaffold_108.assembly12x <a href="#">[view all locations]</a>

The chart shows the ancestry of the ontology terms that have been mapped to this. The nodes are clickable links to the ontology websites.  
Terms: [Mapped terms](#) [GO Slim generic terms](#)



# Whole Genome Alignments

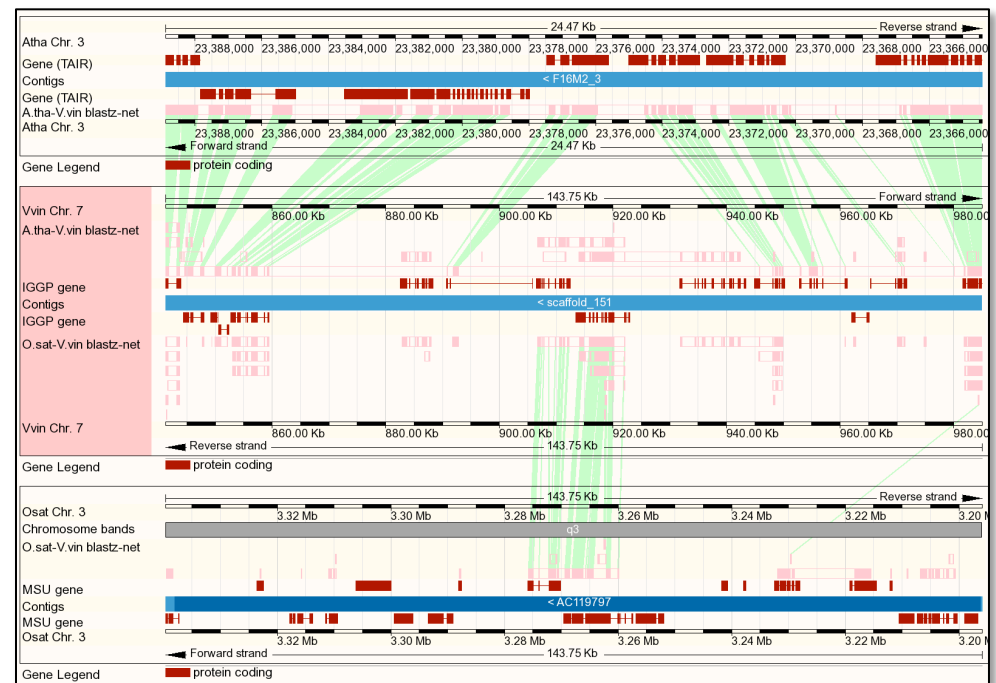
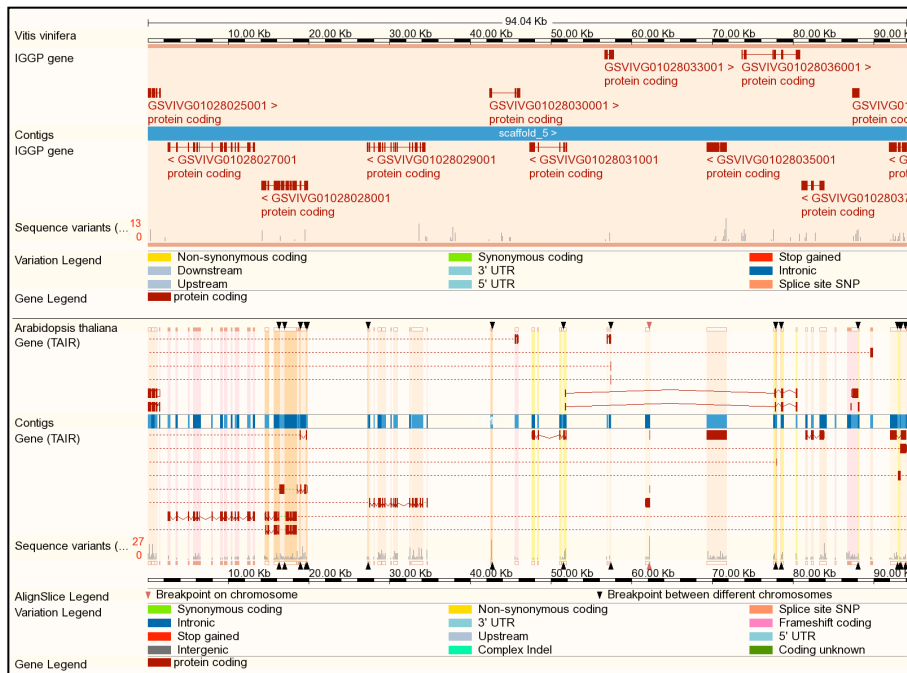
*BLASTZ-CHAIN-NET between 20 pairs of species*

Schwartz S et al., Genome Res.;13(1):103-7  
 Kent WJ et al., Proc Natl Acad Sci U S A., 2003;100(20):11484-9

## Alignment (Release)

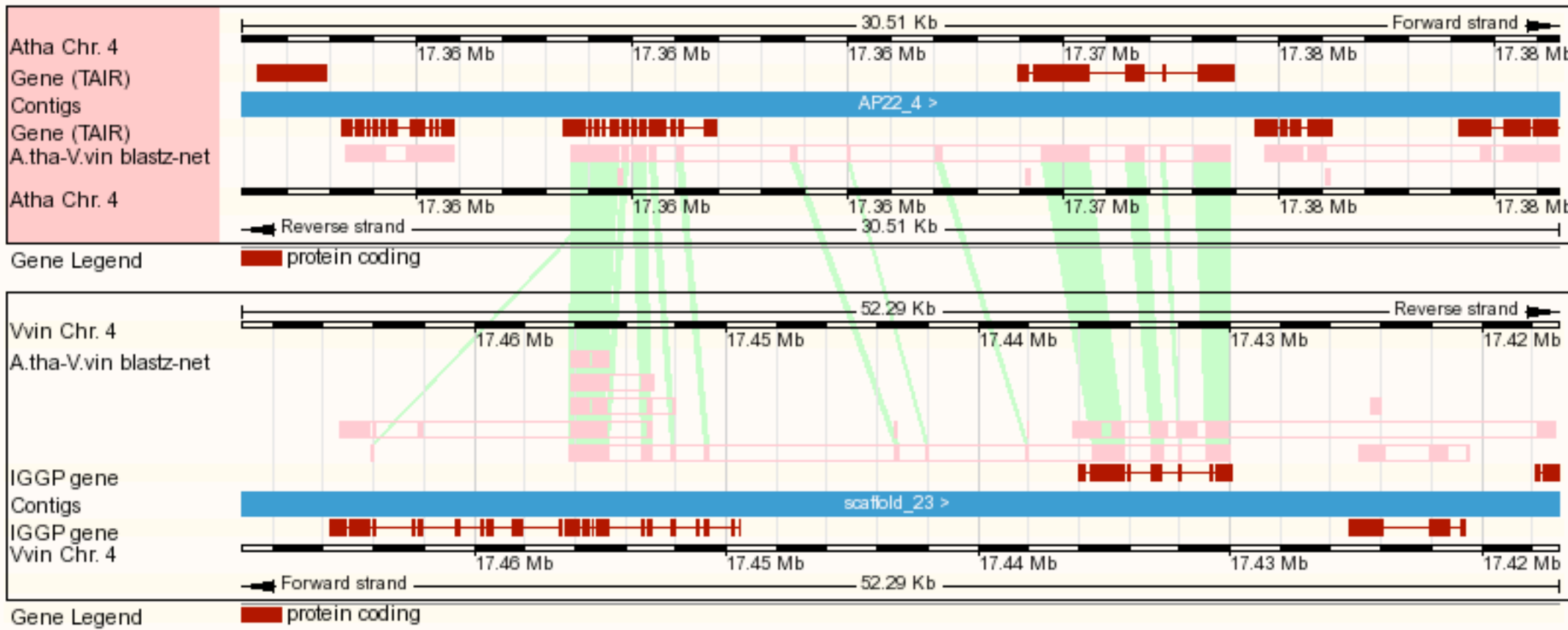
<b>Oryza sativa Japonica</b>	<b>O.jap</b>				
<b>Oryza sativa Indica</b>	<b>31</b>	<b>O.ind</b>			
<b>Sorghum bicolor</b>	<b>31</b>	<b>-</b>	<b>S.bic</b>		
<b>Brachypodium distachyon</b>	<b>31</b>	<b>31</b>	<b>-</b>	<b>B.dis</b>	
<b>Arabidopsis thaliana</b>	<b>31</b>	<b>31</b>	<b>31</b>	<b>31</b>	<b>A.tha</b>
<b>Arabidopsis lyrata</b>	<b>31</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>31</b>
<b>Vitis vinifera</b>	<b>31</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>31</b>
<b>Poplar trichocarpa</b>	<b>31</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>31</b>
<b>Oryza glaberrima 3s</b>	<b>31</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>
<b>Oryza minuta CC 3s</b>	<b>31</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>
<b>Oryza officinalis 3s</b>	<b>31</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>
<b>Oryza punctata 3s</b>	<b>31</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>-</b>
<b>Physcomitrella patens</b>	<b>32</b>	<b>-</b>	<b>-</b>	<b>-</b>	<b>32</b>

# Whole Genome Alignment Views



Vitis vinifera	GCAGAGCCCTTTTATGATGCATATGCTGGTGTCTTCTTCCCAATGTTGAGAAGGACATGGAGC-----AATCCACATTAATGCTTGAACATAGTTCATGATGAC
Arabidopsis_thaliana	GCAGAACCCTTTTCATCGACCGGTATGCCCTTGTCTTACCTCCTTACACAAGAAGGATTTGGACATCATGAACACAACAAGATTACTGTTCGAACAAGTTCATCGATGAT
Vitis vinifera	AAATTCGCAGTACCAATGACATATGATGAGTAAACAGGT-ATTTTCGCCATATCCATCATAGGCCTTACTGACTACTATTA-----CCATTTTGCCTTACTCCG
Arabidopsis_thaliana	AAATTTGTAGAAATAGCTAAGCCATGATGGCTAAACAGGTACTTTTGTGTTACTTACTGTCAGAGT--TTGAGATTAACAGTGAAGTAAATCTTTATGTCGCACTGATA
Vitis vinifera	TTGGTCTTACTCATAT-TGGCAGGTTGTTTATTAACAGATGGCATGATACAGCCATACAGACTCAGTTGGCCACTTCAACATAATAATTGGATAATCCCGGAAGTGTGTT
Arabidopsis_thaliana	T-----CATATCTTGCAGTGTGATTTTCCACAGATGGGATACTCGACTTATAGCTTAATGGCCAACTTCCACTATGATCTTGGATATACCGGAAAAGTTT
Vitis vinifera	TAAAGGACCCCTCAAAGGCTCAAAGGCTGTGGATAGTGGTAAAGATATCCAAAGGAGATTAATTTTTCAGTCTGTCCACATGAGTATTTCAAATACCTGTGGTGGACATAAGC
Arabidopsis_thaliana	CGAATAGCATCTCAAAGCTCAAAGTTTGGC-----
Vitis vinifera	ATGCATCAAGTCTCAACAGATTTCTTGTATAAACCTTAGGTGTGGGCTAAGATTCCTAAAAGTGGCTGTTCTTCAATGTCATTGGAGTCTTCTGATATACAGAAATTTGGC
Arabidopsis_thaliana	.....AGGCTTAGATATCTTCTTAACTGATCC--AGGCTGGAGCTAGGATCCCTAAAAGTGGCTTATCTTCAATCCCGTGAATTCGAAATCGAAGAACATAGAGCAAGCTTAGC
Vitis vinifera	CGAAAAGGTTTCAAGAAATGCACTACATATGGSTCTCCAGGTGGTTTAAAGTCTTAAATC--TCTCAANSTAAAGTGTAACTT-CACCAACAGACATAAATTTCCGGT
Arabidopsis_thaliana	GTCAAAGGTTTACGGAAATGCGCTAGTATAGGCAAGCAGGTA--TATGACTTATTCATCAAACTCTCTTCTCCAGCAATTAACATGATGTGACATGATTTTPTATP
Vitis vinifera	TCTTAAATGGTGATGAATTTGATTTTACCATTTGACAGGACTTCCCTGACGACTTTGGCGAATTTGACAGCTTGTCTTATGTAAGTAGTTAGCCATGAAAGGATGCTCTCT
Arabidopsis_thaliana	TATCAAGGGC-----TACCTTT-----GGATCACAAT--CAGTTTGAAGCTATCTATCGCCATAGTAGCTTGGCCATGAATGAATGCTATC
Vitis vinifera	TTCTGGGATATGCTCTGTGTGGCAGAAAGTGAAGTTGGAAGCAAGTCACTTCTGTTAATAATTAATTTCTTAAATTTTCTTAAATTTCTGGTAGCTACATTTTATCTATTA
Arabidopsis_thaliana	TCAAGAGAAATGCTTAC-----GAGTCAGTTTATCTTCTGACTATCTTGGACTTCTACTTAAAGGATCATCTTCTTGGTTGCTG
Vitis vinifera	AGAAGGACCTCTGTGTGATCTTCCACAGTCACTACAAAGCAATGGATGCAAGCTTTTCATGAGCAAGCTTTCAAGGTGAAATGATTCATTATGTTAAGTAAAGCAAGT
Arabidopsis_thaliana	AAATTTTATCTATATGCTGATGATGCTGATGTCAGTGGATTTGCAAAAGTGGATGGAACAAGCTTTCATGAGCAATGTTTTCGGTGAATAAGTATGATGAAAGATTTGCTGCAAGT
Vitis vinifera	TAGCAAGGACATA--CCATCAAGGCTGTCAGTACACTTTTCTTCCACACATTTCCGTTTCTGATGATCAGTGAAGACAGATTCAGTATCAGAGTCTGGAAATTTTGG
Arabidopsis_thaliana	TAGTGTGGTATACATTACAGCTAAACCAAGACAGGTTCTTTATCCCAACAGCTCAAGTTTTCAGATGATCAGTAA--
Vitis vinifera	TAAAGATTCCTAATAATTTTGTACTTTTCAGTAAACAATAAATCCGGTCTTGTATATCTTCCAGTGATGAACCCACAATAATGGAAATGATTCCTCAATCATTACAAAATG
Arabidopsis_thaliana	-----
Vitis vinifera	CCCTTATCATCTGAGCCCTTGGTTTCCCGATACATTTATGTTTTTGTATTAATTTTCTGGTAGCTACATTTTATCATCTCTCAGATGAAACTTGGAGGAGAAATTTCCAG
Arabidopsis_thaliana	-----ATCTTCCAGATGGAAGATGGAGACAGAAATTCAG
Vitis vinifera	AGAATAGAGAGACAGATGAAATGGCTTCAGGAGCTCTAAACACAAAATTCGTATTTTCCATCGAAACTTGCATAATTTCCGCTCTCTGTTACTTCTTTGTTGATTTTC
Arabidopsis_thaliana	AGACTAGAAAGATGAGATGAAACAGGATTTGAAGACTCTGA-----
Vitis vinifera	CTAACGAATTCACAAAGAGCTTGGTGGAGCTCCACCAACAGATCCAGGATAATACATCTTCCCTCGTTTGAATACGGAGGCTTCTTCCGCTTACACTTGGATAC

# Conserved non-coding regions

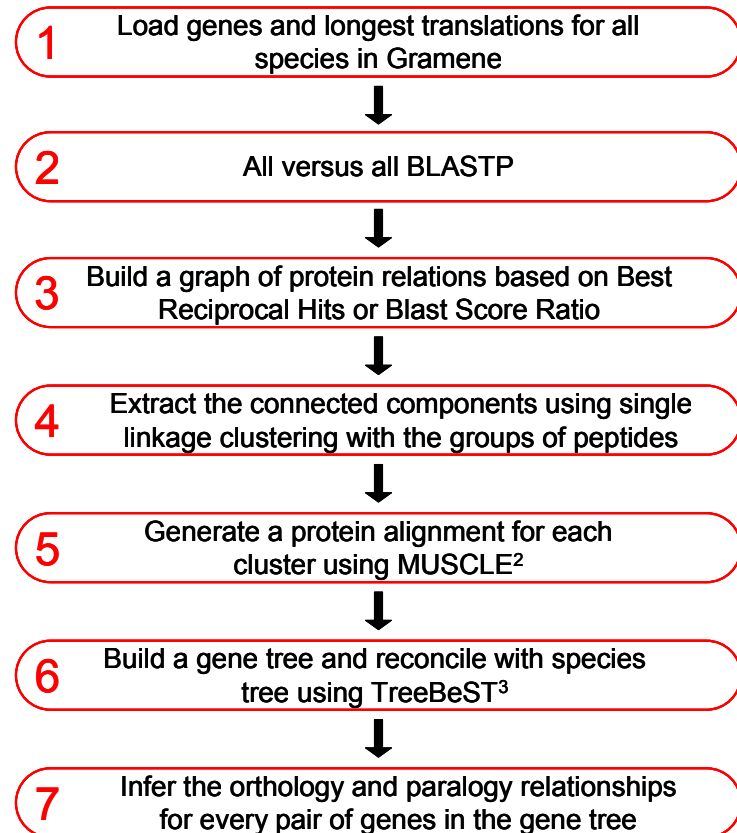
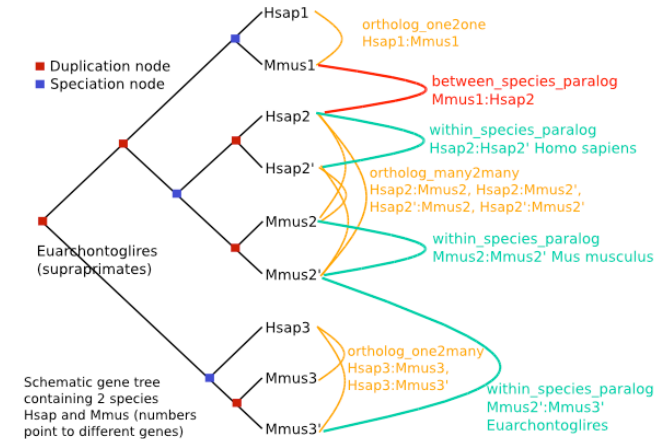


# Compara Gene Trees

## Reconstructing evolutionary histories

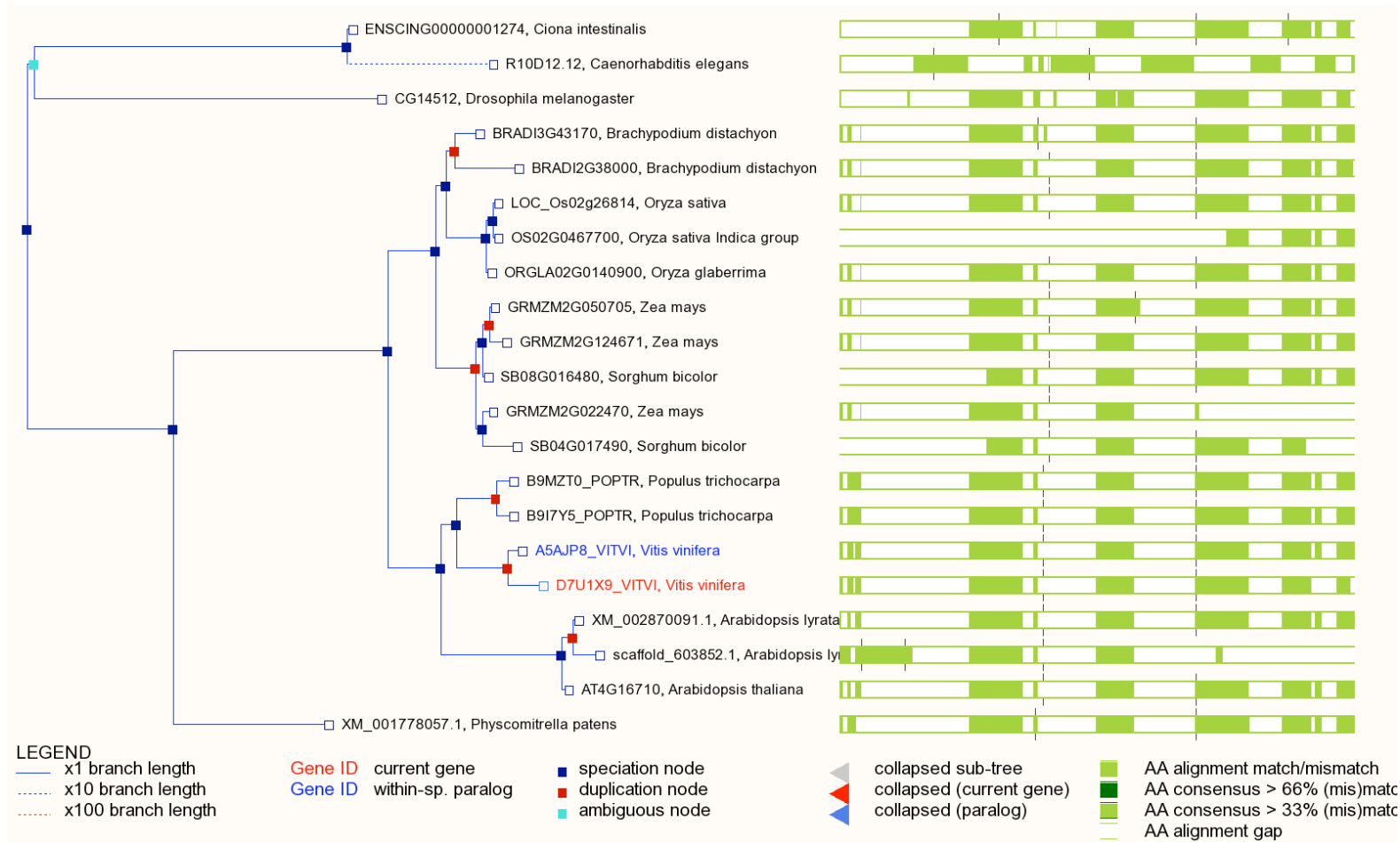
- Gene Trees for 11 plants plus human, ciona, fly, work, yeast
- Infers orthologues and paralogues by reconciling gene tree with species tree
- Vilella A.J., et al. (2008). *Genome Res.* Pre-print: doi:10.1101/gr.073585.107

[http://useast.ensembl.org/info/docs/compara/homology\\_method.html](http://useast.ensembl.org/info/docs/compara/homology_method.html)



# Compara Output

- Blue nodes are speciation events giving rise to orthologues
- Red nodes are duplication events giving rise to paralogues



# Newick tree & alignment

```
((((ENSCINP00000002474_Cint_:0.0000,
R10D12.12_Cele_:3.4477):0.7716,
FBpp0084782_Dmel_:0.8566):0.0000,
((((BRADI3G43170.1_Bdis_:0.0615,
BRADI2G38000.1_Bdis_:0.1536):0.0214,
((LOC_Os02g26814.1_Osat_:0.0000,
BGIOGA008178-PA_Oind_:0.0000):0.0000,
ORGLA02G0140900.1_Ogla_:0.0000):0.0938):0.0231,
(((GRMZM2G050705_P02_Zmay_:0.0099,
GRMZM2G124671_P01_Zmay_:0.0745):0.0043,
Sb08g016480.1_Sbic_:0.0348):0.0000,
(GRMZM2G022470_P01_Zmay_:0.0475,
Sb04g017490.1_Sbic_:0.1037):0.0000):0.0917):0.1118,
(((POPTR_0005s03870.1_Ptri_:0.0420,
POPTR_0013s02650.1_Ptri_:0.0427):0.0918,
(GSVIVT01006266001_Vvin_:0.0342,
GSVIVT01000019001_Vvin_:0.0817):0.1210):0.0363,
((scaffold_702792.1_Alyr_:0.0043,
scaffold_603852.1_Alyr_:0.0632):0.0277,
AT4G16710.1_Atha_:0.0204):0.2813):0.1261):0.5081,
E_GW1.232.43.1_Ppat_:0.3698):0.3605):0.0000;
```

ORGLA02G0140900.1_Ogla_	VFVTVGTTTCF	DALVKA	VDSP	QVKEAL	LEKG	YTDLII	QMGR	GT	Y-----
BRADI2G38000.1_Bdis_	VFVTVGTTTCF	DALVKA	VDSE	EVKQAL	LRKG	YTDLLI	QMGR	GT	Y-----
GRMZM2G050705_P02_Zmay_	VFVTVGTTTCF	DALVMA	VDSP	EVKKAL	LQKG	YSNLLI	QMGR	GT	Y-----
POPTR_0005s03870.1_Ptri_	VFVTVGTTTLF	DALVRT	VDTK	EVKQEL	LRNG	YTHLLI	QMGR	GS	Y-----
GRMZM2G022470_P01_Zmay_	VFVTVGTTTCF	DALVMA	VDSP	EVKKTLL	LQKG	YSNLLI	QMGR	GT	Y-----
BRADI3G43170.1_Bdis_	VFVTVGTTTCF	DALVKK	VDSP	QVKEAL	WQKG	YTDLFI	QMGR	GT	Y-----
GSVIVT01006266001_Vvin_	VFVTVGTTTCF	DALVKA	VDTQ	EFKKE	SARG	YTHLLI	QMGR	GS	Y-----
Sb08g016480.1_Sbic_	-----	----MA	VDSP	EVKMAL	LQKG	YSNLLI	QMGR	GT	Y-----
GRMZM2G124671_P01_Zmay_	VFVTVGTTTCF	DALVMA	VDSP	EVKKAL	LQKG	YSNLLI	QMGR	GT	Y-----
Sb04g017490.1_Sbic_	-----	----MA	VASP	EVKKAL	LQKG	YSNLVI	QMGR	GT	Y-----
BGIOGA008178-PA_Oind_	-----	-----	-----	-----	-----	-----	-----	-----	-----
E_GW1.232.43.1_Ppat_	VLVTVGTTTLF	DALVRE	ASSQ	PCRQVL	ADFG	YSSLVI	QRGK	GS	F-----
scaffold_702792.1_Alyr_	VFVTVGTTTSF	DALVKA	VVSE	DVKDEL	LQKR	FTHLLI	QMGR	GI	F-----
R10D12.12_Cele_	-----	-----	-----	-----	-----	-----	-----	---	NQDVIDR
ENSCINP00000002474_Cint_	IFVTVGTTTSF	DELTE	TITSK	PVQKVL	QSQG	YDKVTI	QYGR	GK	H-----
scaffold_603852.1_Alyr_	VFVTVGTTTSF	DALVKA	VVSE	DVKDEL	LQKR	FTHLLI	QMGR	GN	F-----
AT4G16710.1_Atha_	VFVTVGTTTSF	DALVKA	VVSQ	NVKDEL	LQKR	FTHLLI	QMGR	GI	F-----
LOC_Os02g26814.1_Osat_	VFVTVGTTTCF	DALVKA	VDSP	QVKEAL	LEKG	YTDLII	QMGR	GT	Y-----
GSVIVT01000019001_Vvin_	VFVTVGTTTCF	DALVKA	VDTH	EFKREL	FARG	YTHLLI	QMGR	GS	Y-----
FBpp0084782_Dmel_	VYITVGTTKF	DALIS	TASTE	PALKAL	QNRK	CTKLVI	QHGN	SQ	P-----
POPTR_0013s02650.1_Ptri_	VFVTVGTTTLF	DALVRT	VDTK	EVKQEL	LRKG	YTDLVI	QMGR	GS	Y-----

# Orthologs & Paralogs

## Orthologues

[View sequence alignments of these homologues.](#)

Show <input type="button" value="All"/> entries		Show/hide columns		Filter: <input type="text"/>				
Species	Type	dN/dS	Ensembl identifier	Gene name (Xref)	Compare	Location	Target %id	Query %id
Brachypodium distachyon	1-to-many	n/a	<a href="#">BRADI5G12417</a>	Novel Ensembl prediction DNA binding [Source:GO;Acc:GO:0003677]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	<a href="#">5:15889826-15891965:1</a>	46	55
Brachypodium distachyon	1-to-many	n/a	<a href="#">BRADI3G46910</a>	Novel Ensembl prediction DNA binding [Source:GO;Acc:GO:0003677]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	<a href="#">3:48642114-48644272:1</a>	50	56
Oryza glaberrima	1-to-1	n/a	<a href="#">ORGLA04G0119800</a>	NM_001059528.1 No description	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	<a href="#">Oqglab04_0230:187693-189773:1</a>	47	57
Oryza indica	1-to-many	n/a	<a href="#">BGIOSGA008493</a>	OS02G0579300 Putative uncharacterized protein [Source:Uniprot/SPTREMBL; acc: A2X6G0_ORYS]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	<a href="#">2:24013729-24015783:1</a>	45	55
Oryza indica	1-to-many	n/a	<a href="#">BGIOSGA016547</a>	OSJNBA0072F16.11 Putative uncharacterized protein [Source:Uniprot/SPTREMBL; acc: A2XU96_ORYS]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	<a href="#">4:21193125-21195215:1</a>	47	57
Oryza sativa	1-to-many	n/a	<a href="#">LOC_Os02g36890</a>	Novel Ensembl prediction Os02g0579300 [Source:UniGene;Acc:Os.83361]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	<a href="#">2:22262051-22264373:1</a>	45	55
Oryza sativa	1-to-many	n/a	<a href="#">LOC_Os04g38740</a>	Novel Ensembl prediction Os04g0461000 [Source:UniGene;Acc:Os.80847]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	<a href="#">4:22838914-22841536:1</a>	47	57
Physcomitrella patens	Many-to-many	n/a	<a href="#">E_GW1.106.133.1</a>	XM_001768758.1 subsp. patens predicted protein (PHYPADRAFT_133834) mRNA, partial cds [source:RefSeq_dna_predicted;Acc:XM_001768758.1]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	<a href="#">scaffold_106:1286042-1286798:1</a>	77	38

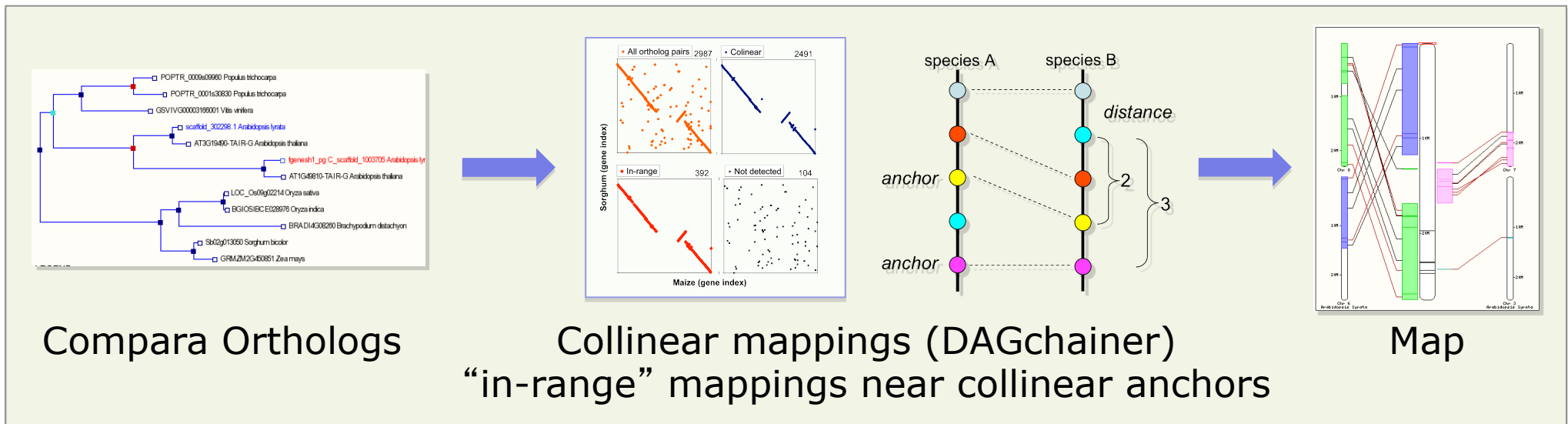
Show/hide columns		Filter: <input type="text"/>				
Ancestral taxonomy	Gene identifier	Gene name (Xref)	Compare	Location	Target %id	Query %id
rosids	<a href="#">GSVIVG01021831001</a>	D7TUL7_VITVI Whole genome shotgun sequence of line PN40024, scaffold_30_assembly12x [Source: Uniprot/SPTREMBL (D7TUL7_VITVI)]	Multi-location view Alignment (protein) Alignment (cDNA)	<a href="#">14:6773465-6781464:1</a>	70	73
Embryophyta	<a href="#">GSVIVG01025304001</a>	A5C6J2_VITVI Putative uncharacterized protein [Source: Uniprot/SPTREMBL (A5C6J2_VITVI)]	Multi-location view Alignment (protein) Alignment (cDNA)	<a href="#">6:2261465-2273045:1</a>	50	41
Embryophyta	<a href="#">GSVIVG01037558001</a>	D7T1P0_VITVI Whole genome shotgun sequence of line PN40024, scaffold_9_assembly12x [Source: Uniprot/SPTREMBL (D7T1P0_VITVI)]	Multi-location view Alignment (protein) Alignment (cDNA)	<a href="#">6:11443683-11452734:1</a>	27	39
Embryophyta	<a href="#">GSVIVG01033338001</a>	D7T1I0_VITVI Whole genome shotgun sequence of line PN40024, scaffold_7_assembly12x [Source: Uniprot/SPTREMBL (D7T1I0_VITVI)]	Multi-location view Alignment (protein) Alignment (cDNA)	<a href="#">8:21367807-21372343:1</a>	38	40



# Gene-Centered Synteny Build

2010: Implemented with automated pipeline runnables

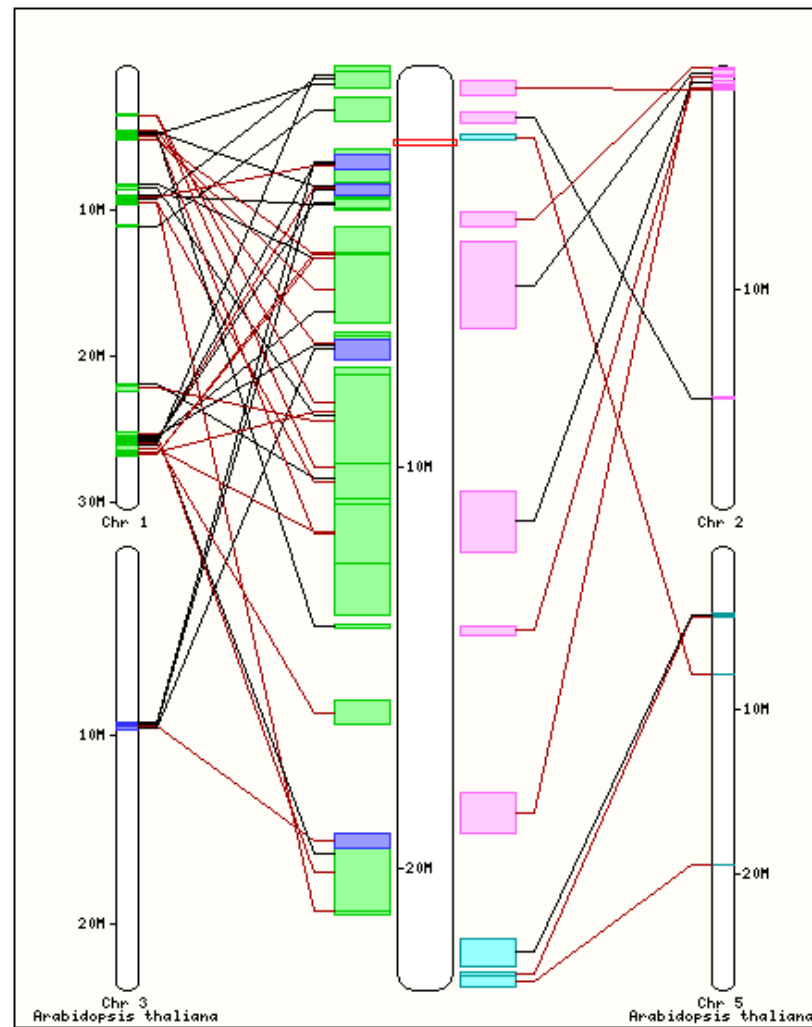
- Release 31: monocots
- Release 32: dicots



<b><i>Oryza sativa Japonica</i></b>	<b><i>O.jap</i></b>						
<b><i>Brachypodium distachyon</i></b>	YES	<b><i>B.dis</i></b>					
<b><i>Sorghum bicolor</i></b>	YES	YES	<b><i>S.bic</i></b>				
<b><i>Arabidopsis thaliana</i></b>	-	-	-	<b><i>A.tha</i></b>			
<b><i>Arabidopsis lyrata</i></b>	-	-	-	YES	<b><i>A.lyr</i></b>		
<b><i>Vitis vinifera</i></b>	-	-	-	YES	YES	<b><i>V.vin</i></b>	
<b><i>Poplar trichocarpa</i></b>	-	-	-	YES	YES	YES	<b><i>P.tri</i></b>

# Synteny Browser

Synteny



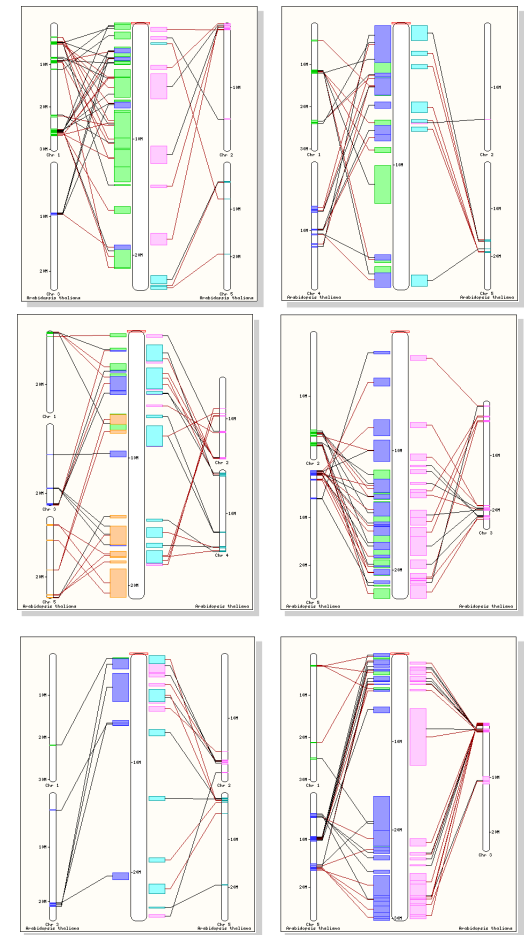
Change Species:

Arabidopsis thaliana

Go

# Grape Reference Highlights Duplicated Regions in Arabidopsis and Poplar

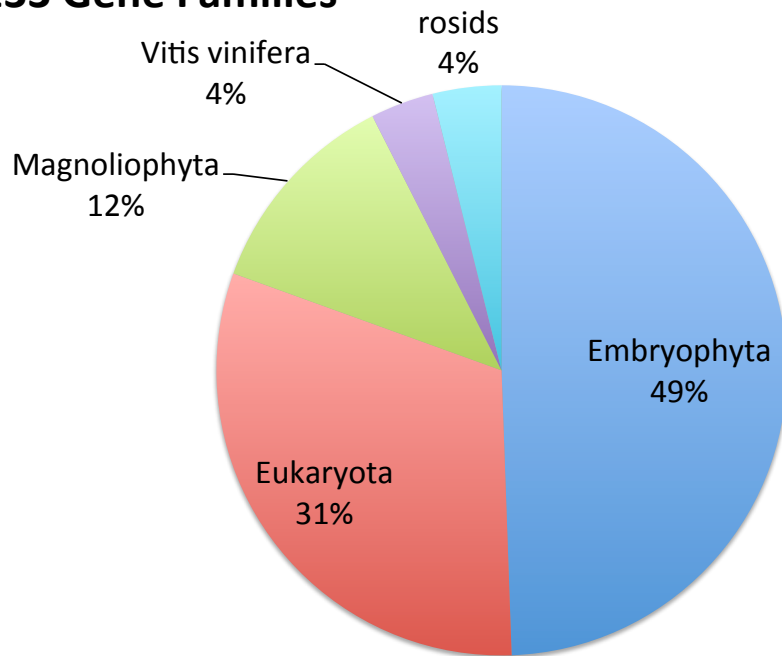
- Polyploid and segmental duplications manifest as co-syntenic regions
- SyntenyView links to browser: Thus users can easily navigate between duplicated regions



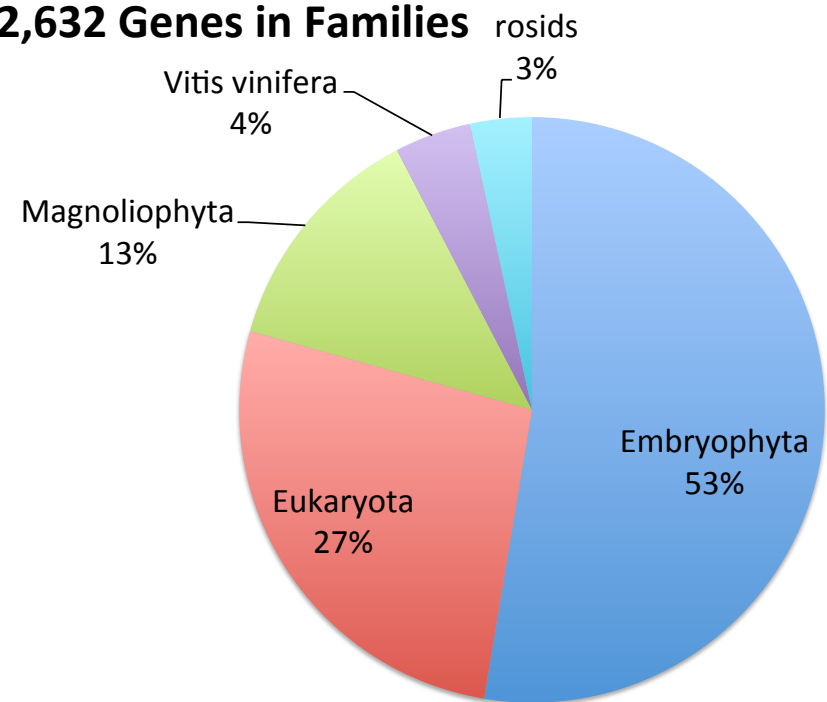
# Some Results

26346 genes  
86% in families  
3714 singletons

**9133 Gene Families**



**22,632 Genes in Families**

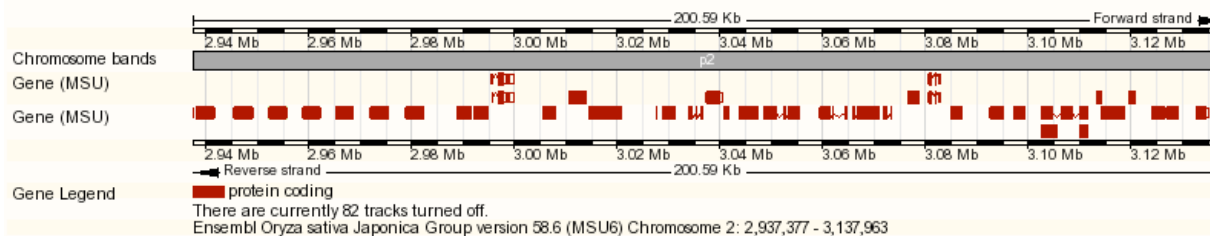


# Tandem Duplicate Detection

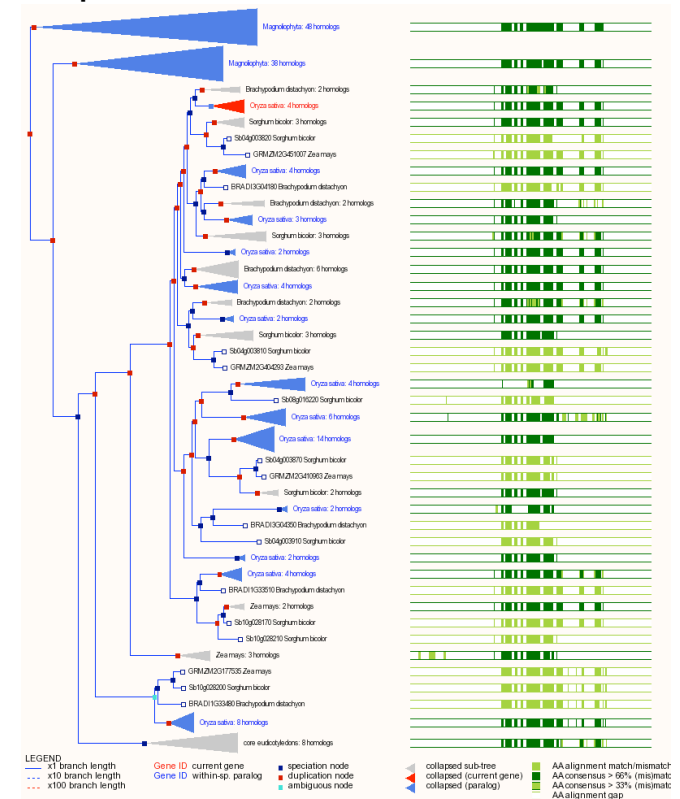
Species	Clusters	Genes	Largest	Function
Rice japonica	2519	7054	24	phytosulfokine receptor-like (LRR-kinase receptor)
Sorghum	2182	5927	19	Chalcone-stilbene synthase like
Maize	1871	4564	22	DUF1754 (domain of unknown function)
Arabidopsis	1738	4581	28	ECA1 gametogenesis related family
Grape	1999	5799	28	Terpenoid cyclase

- Adjacent paralogs with no more than 2 intervening unrelated gene
- Increase gene dosage
- Diversifying selection
- Often species-specific

## LRR-Kinase cluster in rice



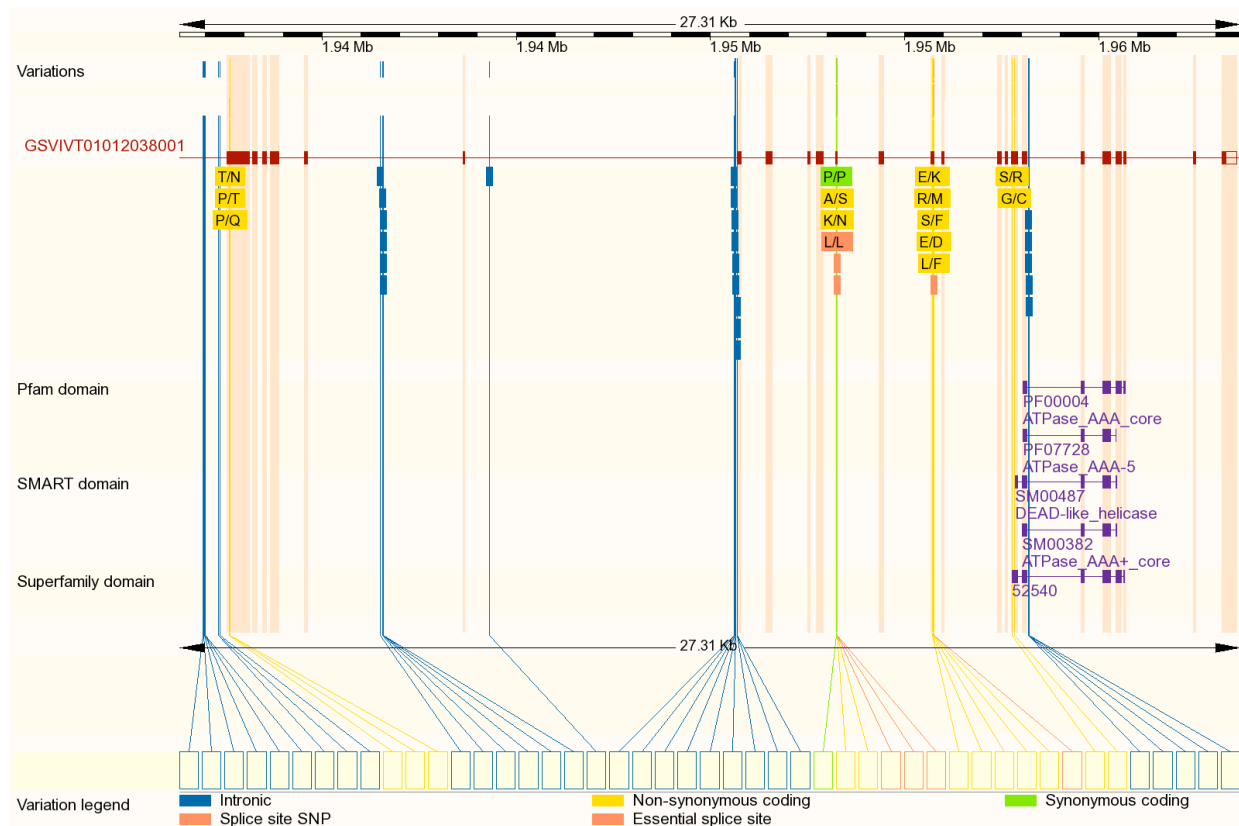
## LRR-Kinase species-specific expansions



# SNP View

71K SNPs from Myles et al  
PLoS One. 2010 Jan 13;5(1):e8219

## Functional consequences of variant



- Synonymous coding
- Non-synonymous coding
- Stop gain/loss
- Splice site
- UTR
- Intronic

**Individual genotypes**

Summary of genotypes by population

Number of genotypes	Population	Description
2	USDA_NPGS	USDA_NPGS
1	USDA_NPGS-Non_Vinifera	USDA_NPGS-Non_Vinifera
1	USDA_NPGS-Vinifera	USDA_NPGS-Vinifera_cultivars

Genotypes for USDA\_NPGS

Individual	Genotype (forward strand)	Description
DV10535:USDA_NPGS	A/A	THOMPSON SEEDLESS (USDA_NPGS-VINIFERA)
PI58233.01:USDA_NPGS	A/C	VITIS PALMATA (USDA_NPGS-NON_VINIFERA)

Genotypes for USDA\_NPGS-Non\_Vinifera

Individual	Genotype (forward strand)	Description
PI58233.01:USDA_NPGS	A/C	VITIS PALMATA (USDA_NPGS-NON_VINIFERA)

Genotypes for USDA\_NPGS-Vinifera

Individual	Genotype (forward strand)	Description
DV10535:USDA_NPGS	A/A	THOMPSON SEEDLESS (USDA_NPGS-VINIFERA)

### Population genetics

Population	Alleles A	Alleles C	Genotypes A/A	Genotypes A/C	Genotypes C/C	Count
USDA_NPGS-Non_Vinifera	0.100	0.900		0.200	0.800	5
USDA_NPGS	0.088	0.912	0.059	0.059	0.882	17
USDA_NPGS-Vinifera	0.083	0.917	0.083		0.917	12


# Data Access

Index of ftp://ftp.gramene.org/pub/gramene/CURRENT\_RELEASE  
/data/

 Up to higher level directory

Name	Size	Last Modified
 associations		4/18/11 9:57:00 PM
 database_dump		4/18/11 10:00:00 PM
 diversity		4/28/11 7:38:00 PM
 fasta		5/3/11 2:50:00 PM
 gff		4/24/11 4:13:00 AM
 maps		4/18/11 9:58:00 PM
 ontology		4/18/11 9:57:00 PM
 <i>pathways</i>		4/28/11 7:31:00 PM
 qtl		4/18/11 9:58:00 PM
 statistics		4/19/11 7:20:00 PM

# BioMart


[BLAST](#) | [BioMart](#) | [Documentation](#) | [Help](#) | [Feedback](#)

[New](#) | [Count](#) | [Results](#)

[★ URL](#) | [XML](#) | [Perl](#) | [Help](#)

Dataset: **Vitis vinifera genes (IGGP\_12x)**

Filters: [None selected]

Attributes:
 

- Ensembl Gene ID
- Chromosome/plasmid
- Gene Start (bp)
- Gene End (bp)
- Vitis vinifera Paralog Ensembl Gene ID
- Vitis vinifera Paralog Chromosome
- Vitis vinifera Paralog Chr Start (bp)
- Vitis vinifera Paralog Chr End (bp)
- % Coverage
- % Identity
- Vitis vinifera Paralog % Coverage
- Vitis vinifera Paralog % Identity
- Ancestor

Export all results to:  TSV  Unique results only

Email notification to:

View: 10 rows as HTML  Unique results only

Ensembl Gene ID	Chromosome/plasmid	Gene Start (bp)	Gene End (bp)	Vitis vinifera Paralog Ensembl Gene ID	Vitis vinifera Paralog Chromosome	Vitis vinifera Paralog Chr Start (bp)	Vitis vinifera Paralog Chr End (bp)	% Coverage	% Identity	Vitis vinifera Paralog % Coverage	Vitis vinifera Paralog % Identity	Ancestor
<a href="#">GSVIVG01030133001</a>	12	<a href="#">10124569</a>	<a href="#">10129564</a>	<a href="#">GSVIVG01030134001</a>	12	10130090	10132198	100	40	100	61	rosids
<a href="#">GSVIVG01030133001</a>	12	<a href="#">10124569</a>	<a href="#">10129564</a>	<a href="#">GSVIVG01033738001</a>	8	18222784	18224352	100	33	100	63	rosids
<a href="#">GSVIVG01030133001</a>	12	<a href="#">10124569</a>	<a href="#">10129564</a>	<a href="#">GSVIVG01035880001</a>	4	5206353	5210218	100	34	100	45	Magnoliophyta
<a href="#">GSVIVG01030133001</a>	12	<a href="#">10124569</a>	<a href="#">10129564</a>	<a href="#">GSVIVG01012820001</a>	11	5736757	5739075	100	32	100	43	Magnoliophyta
<a href="#">GSVIVG01030133001</a>	12	<a href="#">10124569</a>	<a href="#">10129564</a>	<a href="#">GSVIVG01012806001</a>	11	5557470	5575151	100	35	100	26	Magnoliophyta
<a href="#">GSVIVG01030133001</a>	12	<a href="#">10124569</a>	<a href="#">10129564</a>	<a href="#">GSVIVG01012808001</a>	11	5586712	5591626	100	31	100	50	Magnoliophyta
<a href="#">GSVIVG01030133001</a>	12	<a href="#">10124569</a>	<a href="#">10129564</a>	<a href="#">GSVIVG01012805001</a>	11	5550176	5552396	100	22	100	52	Magnoliophyta
<a href="#">GSVIVG01030133001</a>	12	<a href="#">10124569</a>	<a href="#">10129564</a>	<a href="#">GSVIVG01012807001</a>	11	5580210	5582473	100	32	100	52	Magnoliophyta
<a href="#">GSVIVG01030133001</a>	12	<a href="#">10124569</a>	<a href="#">10129564</a>	<a href="#">GSVIVG01035881001</a>	4	5211280	5215609	100	30	100	47	Magnoliophyta
<a href="#">GSVIVG01007870001</a>	17	<a href="#">8144795</a>	<a href="#">8155931</a>									

Dataset: [None Selected]

Homologs, structure/function prediction, SNP



# Tutorials

<http://www.openhelix.com/gramene>

The screenshot shows the OpenHelix website interface. At the top left is the OpenHelix logo. A search bar contains the text 'gramene' with 'Search' and 'Clear' buttons. Below the search bar, it says 'Advanced search options' and 'Help'. A suggestion 'Did you mean: [grameen](#)?' is shown. A message states 'Your query 'gramene' matched 517 documents in 33 resources.' Below this, a section titled 'Free OpenHelix tutorials that match your search' is visible. A sub-section 'OpenHelix Tutorials' is expanded to show 'Resource Information' for 'Gramene: <http://www.gramene.org/>'. The description reads: 'Gramene, grass species genomic database and tools for genetic analysis of the plants'. Below this, a 'Tutorial movie (audio-visual presentation)' is listed with a series of slides: Slide 3: Gramene is a public resource and genome browser that includes... Slide 4: ... you better acquainted with the Gramene tool. We will introduce... Slide 5: ... Gramene resource. The credit for Gramene goes to the Gramene... Slide 6: ... collaborators and contributors to the Gramene database... Slide 7: ... find out more about the Gramene resource and database... Slide 8: Gramene is publicly available and free to use. If you use Gramene... Slide 9: ... for the species found in the Gramene resource. Let's take a look... Slide 10: ... see a list of tips that the Gramene developers have offered... Slide 11: ... our introduction to the Gramene resource. Beginning of... Slide 12: The Gramene genome browsers can be accessed from the Gramene... Slide 13: ... top navigation bar to the entire Gramene resource is here. 22 additional slides match the query. Links for 'View the tutorial movie' and 'Go to tutorial info page' are provided.

Self-run tutorial as well as PowerPoint slides, handouts, and exercises

The screenshot shows the Gramene website navigation bar with 'GRAMENE Home' and a search bar. The main navigation menu includes 'Search', 'Genomes', 'Species', 'Download', 'Resources', 'About', 'Help', and 'Feedback'. Below the navigation bar, there are links for 'Gramene Home | FAQs | Help | Tutorials | Release Notes'. A section titled 'Gramene Tutorials and Exercises' contains a list of links: 'Gramene tutorials at Open Helix', 'Pathways', 'Phenote', 'Starting Tassel', 'Tassel GLM Analysis', 'Diversity', and 'QTL'. Below this, a section titled 'Ensembl Tutorials' contains a link: 'Tutorials at Ensembl'.

## Archived Tutorials and Exercises

The following resources are no longer maintained, and were created for **Gramene V26 (Sept 2007)**. Although the images and some of the navigation may be out of date, you may still find material here to be helpful in conducting your searches.

**Archived old tutorials**  
Gramene Navigation  
Overview of Modules  
Genomes  
GrameneMart  
BLAST  
Maps  
Markers  
Proteins  
Ontologies

**Select your preferred format:**  
PowerPoint  
PowerPoint  
PowerPoint  
PowerPoint  
PowerPoint  
PowerPoint  
PowerPoint  
PowerPoint  
PowerPoint

## Archived Exercises from Previous workshops.

RiceCAP 2006 Workshop (6/06) Conducting searches. QTL and markers. Introduction to Diversity and Pathway databases.  
Exercise in Power Point or MS Word. RTWG 2006 Workshop (2/06)  
Gramene workshop for new and novice users.

- Example of a database search on the Pi-ta gene in PDF
- Exercise - Milling Yield in PDF

# Web Services











- Distributed Annotation Server (DAS) serving Ensembl genes as well as Gramene markers, sequences, and QTL
- Gramene Mart integration with Galaxy
- Public MySQL server
- Diversity data via Tassel and GDPC
- Subversion for code access

<input type="checkbox"/> <input type="checkbox"/> <b>Web Services</b>
<input type="checkbox"/> <b>Gramene DAS</b>
<input type="checkbox"/> <b>Gramene/Ensembl DAS sources</b>
<input type="checkbox"/> <b>Gramene's QTL - Semantic Web Services</b>
<input type="checkbox"/> <b>Gramene's QTL - Semantic Web Services Via the Discovery Service at <a href="http://sswap.info">http://sswap.info</a></b>
<input type="checkbox"/> <b>The Diversity Advanced Search Tool</b>
<input type="checkbox"/> <b>Public Read-only MySQL Database Server</b>
<input type="checkbox"/> <b>Subversion Repository</b>


# Data Files

Index of [ftp://ftp.gramene.org/pub/gramene/outreach/grape\\_rcn/](ftp://ftp.gramene.org/pub/gramene/outreach/grape_rcn/)

 [Up to higher level directory](#)

Name	Size	Last Modified	
 <a href="#">VV-at-blastz-net.gff3</a>	14154 KB	6/2/11	3:07:00 AM
 <a href="#">v0_GO.txt</a>	33444 KB	6/1/11	11:00:00 PM
 <a href="#">v0_homeologs.txt</a>	254 KB	6/1/11	11:00:00 PM
 <a href="#">v0_ipr_long.txt</a>	4564 KB	6/1/11	11:00:00 PM
 <a href="#">v0_ipr_string.txt</a>	1633 KB	6/1/11	11:00:00 PM
 <a href="#">v0_paralog_clusters.txt</a>	252 KB	6/2/11	3:14:00 AM
 <a href="#">v0_putative_missing.txt</a>	114 KB	6/1/11	11:00:00 PM
 <a href="#">v0_putative_split.txt</a>	62 KB	6/1/11	11:00:00 PM
 <a href="#">v0_tree_ipr.txt</a>	2549 KB	6/1/11	11:00:00 PM
 <a href="#">vv_at_orthologs.txt</a>	3554 KB	6/1/11	11:00:00 PM

- Root\_node ID indicates family (each a different gene tree)
- Mine on InterPro domains
- Examples:
  - HLH: 33 trees; 140 genes; 127 with domain
  - MYB: 57 trees; 266 genes; 249 with domain

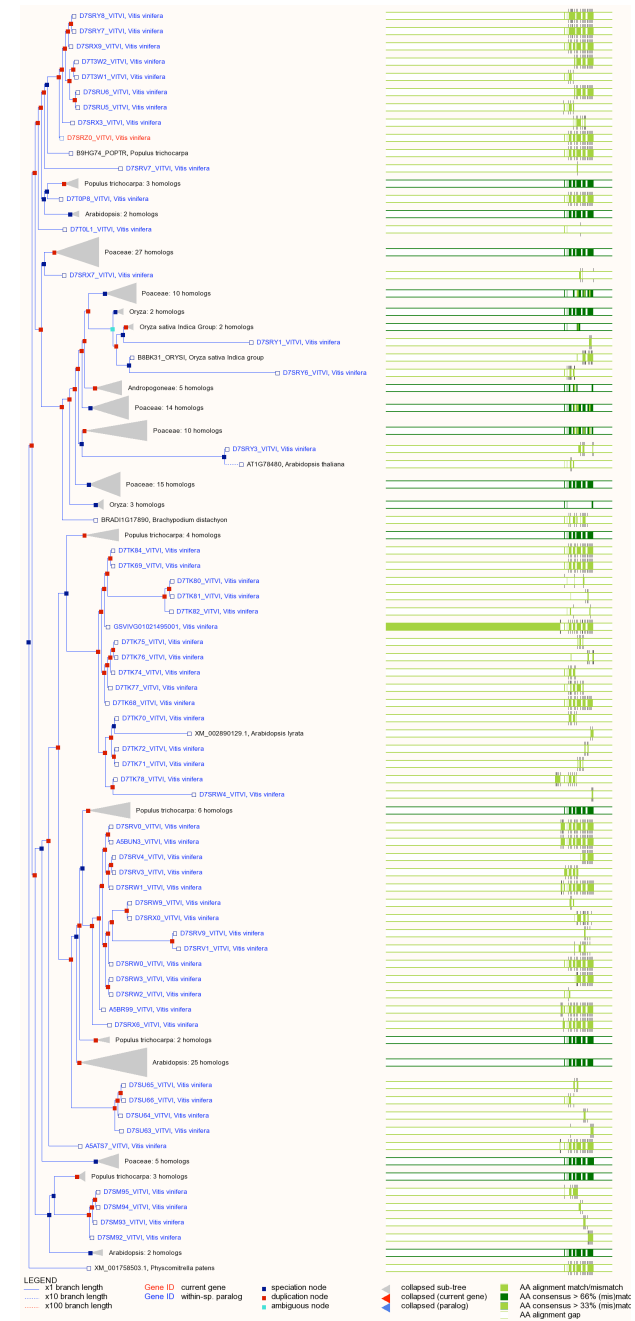
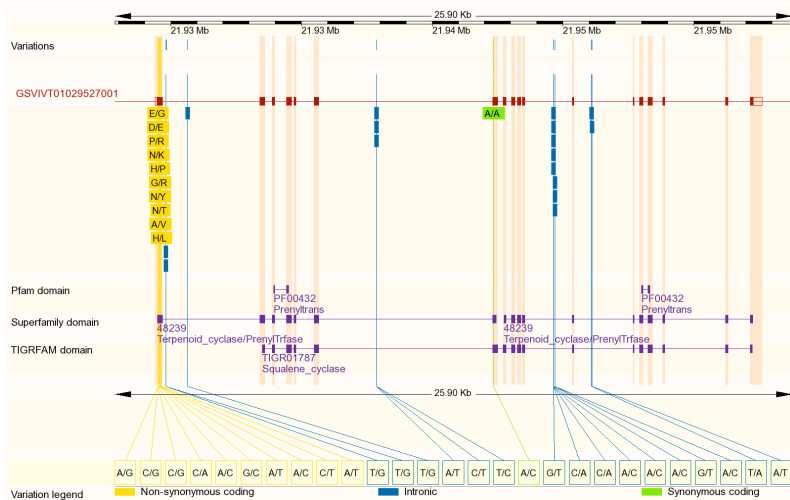
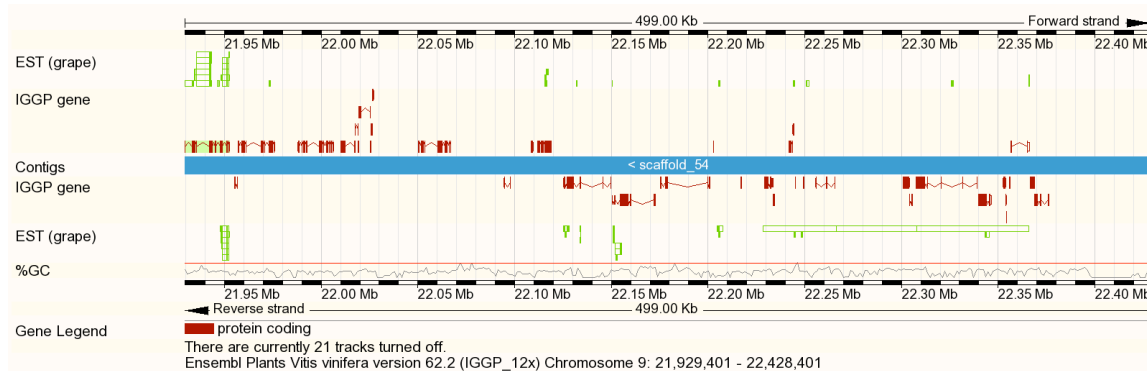


	A	B	C	D	E	F	G
1	gene_stable_id	reference	start	end	root_node	root_taxon	IPR_string
46	GSVIVG01008556001	17	904847	907405	868	Embryophyta	IPR004333=Transcpt_factor_SBP-box
47	GSVIVG01010496001	1	21064305	21071687	868	Embryophyta	IPR004333=Transcpt_factor_SBP-box
48	GSVIVG01010522001	1	21412776	21417820	868	Embryophyta	IPR004333=Transcpt_factor_SBP-box
49	GSVIVG01012247001	1	230334	235218	868	Embryophyta	IPR004333=Transcpt_factor_SBP-box
50	GSVIVG01032239001	11	13588595	13606856	868	Embryophyta	IPR004333=Transcpt_factor_SBP-box
51	GSVIVG01033064001	14	25486818	25489754	868	Embryophyta	IPR004333=Transcpt_factor_SBP-box
52	GSVIVG01033519001	8	20046489	20051103	868	Embryophyta	IPR004333=Transcpt_factor_SBP-box
53	GSVIVG01003836001	4	20536349	20537931	966	Embryophyta	IPR004333=Transcpt_factor_SBP-box
54	GSVIVG01014302001	19	2484297	2485658	966	Embryophyta	IPR004333=Transcpt_factor_SBP-box  IPR017238=Squamosa_promoter-bd_prot
55	GSVIVG01018205001	15	13245041	13249262	966	Embryophyta	IPR004333=Transcpt_factor_SBP-box
56	GSVIVG01020578001	12	4108297	4110237	966	Embryophyta	IPR004333=Transcpt_factor_SBP-box  IPR017238=Squamosa_promoter-bd_prot
57	GSVIVG01021087001	10	1346397	1348752	966	Embryophyta	IPR004333=Transcpt_factor_SBP-box  IPR017238=Squamosa_promoter-bd_prot
58	GSVIVG01013452001	18	347645	353985	997	Embryophyta	IPR002110=Ankyrin_rpt  IPR004333=Transcpt_factor_SBP-box  IPR020683=Ankyrin_rpt-contain_dom
59	GSVIVG01017678001	5	2565343	2574143	997	Embryophyta	IPR002110=Ankyrin_rpt  IPR004333=Transcpt_factor_SBP-box  IPR020683=Ankyrin_rpt-contain_dom
60	GSVIVG01020050001	1	11124690	11127549	997	Embryophyta	IPR004333=Transcpt_factor_SBP-box
61	GSVIVG01020051001	1	11080253	11083113	997	Embryophyta	IPR004333=Transcpt_factor_SBP-box
62	GSVIVG01028208001	7	4628142	4637596	997	Embryophyta	IPR004333=Transcpt_factor_SBP-box  IPR020683=Ankyrin_rpt-contain_dom
63	GSVIVG01018204001	15	13256865	13259160	1044	Embryophyta	IPR004333=Transcpt_factor_SBP-box
64	GSVIVG01017835001	5	3885149	3896071	1074	Embryophyta	IPR004333=Transcpt_factor_SBP-box
65	GSVIVG01002961001	Un	37441031	37441733	1086	Eukaryota	IPR011009=Kinase-like_dom  IPR020635=Tyr_Pkinase_cat_dom
66	GSVIVG01008413001	17	2321687	2342403	1086	Eukaryota	IPR001245=Ser/Thr/Tyr-Pkinase  IPR002290=Ser/Thr_prot_kinase_dom  IPR011009=Kinase-like_dom  IPR017442=Se/Thr_prot_kinase-like_dom  IPR020635=Ty
67	GSVIVG01009192001	18	5939861	5949524	1086	Eukaryota	IPR001245=Ser/Thr/Tyr-Pkinase  IPR002290=Ser/Thr_prot_kinase_dom  IPR011009=Kinase-like_dom  IPR017442=Se/Thr_prot_kinase-like_dom  IPR020635=Ty
68	GSVIVG01019821001	2	3631707	3639546	1086	Eukaryota	IPR000014=PAS  IPR001245=Ser/Thr/Tyr-Pkinase  IPR002290=Ser/Thr_prot_kinase_dom  IPR011009=Kinase-like_dom  IPR013655=PAS_fold_3  IPR017442=Se/T
69	GSVIVG01034988001	14	6026347	6048725	1086	Eukaryota	IPR001245=Ser/Thr/Tyr-Pkinase  IPR002290=Ser/Thr_prot_kinase_dom  IPR011009=Kinase-like_dom  IPR017442=Se/Thr_prot_kinase-like_dom  IPR020635=Ty
70	GSVIVG01033779001	8	17884683	17902051	1086	Eukaryota	IPR001245=Ser/Thr/Tyr-Pkinase  IPR002290=Ser/Thr_prot_kinase_dom  IPR02919=Prot_Inh_CR_TIL  IPR011009=Kinase-like_dom  IPR017442=Se/Thr_prot_kit
71	GSVIVG01034710001	13	8150739	8202754	1086	Eukaryota	IPR001245=Ser/Thr/Tyr-Pkinase  IPR002290=Ser/Thr_prot_kinase_dom  IPR011009=Kinase-like_dom  IPR017442=Se/Thr_prot_kinase-like_dom  IPR020635=Ty
72	GSVIVG01034988001	5	694778	701765	1086	Eukaryota	IPR000014=PAS  IPR001245=Ser/Thr/Tyr-Pkinase  IPR002290=Ser/Thr_prot_kinase_dom  IPR002373=cAMP/cGMP_kin  IPR011009=Kinase-like_dom  IPR013655
73	GSVIVG01035409001	4	1079723	1090460	1086	Eukaryota	IPR001245=Ser/Thr/Tyr-Pkinase  IPR002290=Ser/Thr_prot_kinase_dom  IPR011009=Kinase-like_dom  IPR017442=Se/Thr_prot_kinase-like_dom  IPR020635=Ty
74	GSVIVG01008750001	18	1698714	1706243	1243	Eukaryota	IPR000194=ATPase_F1/V1/A1_a/bsu_nucl-bd  IPR000793=ATPase_F1/V1/A1-cplx_a/bsu_C  IPR004100=ATPase_F1/V1/A1-cplx_a/bsu_N  IPR005723=ATPase_\
75	GSVIVG01023982001	3	1933536	1941032	1243	Eukaryota	IPR000194=ATPase_F1/V1/A1_a/bsu_nucl-bd  IPR000793=ATPase_F1/V1/A1-cplx_a/bsu_C  IPR004100=ATPase_F1/V1/A1-cplx_a/bsu_N  IPR005723=ATPase_\
76	GSVIVG01002448001	Un	34796058	34798389	1268	Eukaryota	IPR000194=ATPase_F1/V1/A1_a/bsu_nucl-bd  IPR000793=ATPase_F1/V1/A1-cplx_a/bsu_C
77	GSVIVG01029546001	9	21769589	21779189	1299	Eukaryota	IPR000194=ATPase_F1/V1/A1_a/bsu_nucl-bd  IPR000793=ATPase_F1/V1/A1-cplx_a/bsu_C  IPR003593=ATPase_AAA+core  IPR004100=ATPase_F1/V1/A1-cpb
78	GSVIVG0103538001	4	2125789	2133322	1299	Eukaryota	IPR000194=ATPase_F1/V1/A1_a/bsu_nucl-bd  IPR000793=ATPase_F1/V1/A1-cplx_a/bsu_C  IPR003593=ATPase_AAA+core  IPR004100=ATPase_F1/V1/A1-cpb

# Tandem Arrays

- Often rapidly evolving
- Species-specific expansion
- Frequent pseudogenization

Largest has 28 genes covering 580 kb!  
Terpenoid cyclase



# vv\_paralog\_clusters.txt

1999 detected clusters  
5799 genes

	A	B	C	D	E	F	G
1	clust_id	gid	chr	start	index	root_id	
2	1	GSVIVG01012250001	1	150382	7	153558	
3	1	GSVIVG01012249001	1	194107	8	153558	
4	2	GSVIVG01012243001	1	263082	14	101326	
5	2	GSVIVG01012240001	1	284312	17	101326	
6	3	GSVIVG01012210001	1	544467	47	252278	
7	3	GSVIVG01012209001	1	550211	48	252278	
8	4	GSVIVG01012192001	1	667767	62	138333	
9	4	GSVIVG01012191001	1	679483	63	138333	
10	5	GSVIVG01012169001	1	806983	83	4571	
11	5	GSVIVG01012168001	1	850561	84	4571	
12	5	GSVIVG01012167001	1	867583	85	4571	
13	5	GSVIVG01012165001	1	901077	87	4571	
14	6	GSVIVG01012150001	1	1017874	101	150053	
15	6	GSVIVG01012149001	1	1030308	102	150053	
16	7	GSVIVG01012139001	1	1107410	112	371829	

clust\_id = name of cluster  
root\_id = name of family

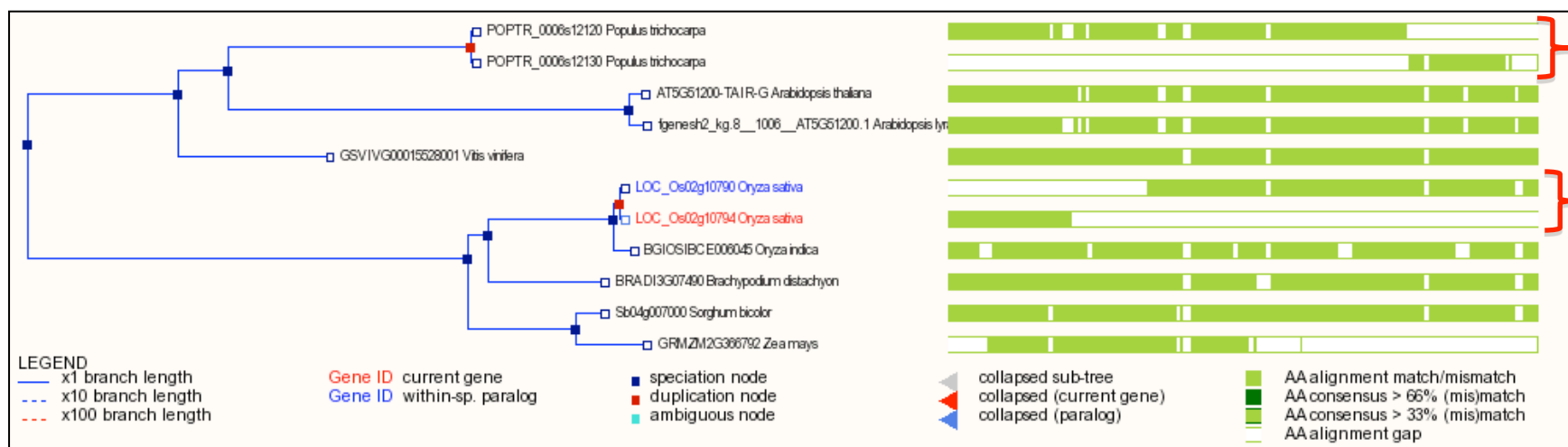
Count	representative	start range	InterPro	Family
28	GSVIVG01029527001	9:21929401-22509692	IPR001330=Prenyltrans!IPR008930=Terpenoid_cyclase/PrenylTrfase!IPR018333=Squalene_cyclase	240317
21	GSVIVG01021326001	10:4055584-4441034	IPR002283=Isopenicillin-N_synthase!IPR005123=Oxoglutarate/Fe-dep_oxygenase	182691
18	GSVIVG01032025001	13:23425164-23649470	IPR001360=Glyco_hydro_1!IPR017853=Glyco_hydro_catalytic_core IPR001099=Chalcone/stilbene_synthase_N!IPR011141=Polyketide_synthase_type-III!IPR012328=Chalcone/ stilbene_synth_C!IPR013747=ACP_syn_III_C!IPR016039=Thiolase-like	138137
16	GSVIVG01010590001	16:16268747-16695695	IPR001077=O_MeTrfase_2!IPR012967=Plant_MeTrfase_dimerisation!IPR016461=O-MeTrfase_COMT_euk	320615
16	GSVIVG01020647001	12:3482933-3698087	IPR002198=DH_sc/Rdtase_SDR!IPR002347=Glc/ribitol_DH!IPR002424=Insect_AlcDH_fam! IPR003560=DHB_DH!IPR013968=PKS_KR!IPR016040=NAD(P)-bd_dom	160055
16	GSVIVG01024899001	6:6168813-6302405	IPR002283=Isopenicillin-N_synthase!IPR005123=Oxoglutarate/Fe-dep_oxygenase	112713
16	GSVIVG01027734001	5:7321488-7468939	IPR004255=UPF0089	183365
15	GSVIVG01027145001	15:17436787-17635542	IPR000757=Glyco_hydro_16!IPR008264=Beta_glucanase!IPR008985=ConA-like_lec_gl!IPR010713=XET_C	358659
15	GSVIVG01029176001	11:18939068-19013959	IPR002110=Ankyrin_rpt!IPR020683=Ankyrin_rpt-contain_dom	248433
15	GSVIVG01013583001	5:21184286-21880043	IPR001330=Prenyltrans!IPR008930=Terpenoid_cyclase/PrenylTrfase!IPR018333=Squalene_cyclase	42684
15	GSVIVG01021473001	10:5939677-6150700	IPR003676=Auxin_inducible	240317
15	GSVIVG01024138001	3:858886-922119		293636

# Automated Detection of Split Genes

## *Special class of “paralog” since Ensembl 58*

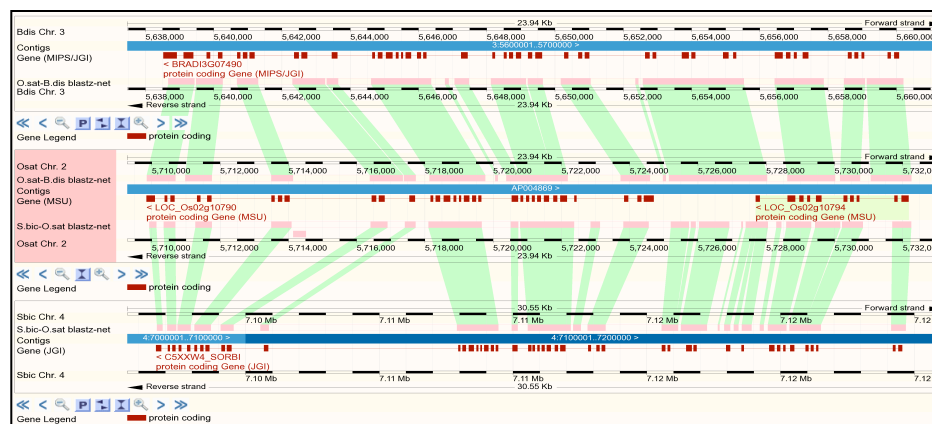
Contiguous split paralog: Non-overlapping, nearby (<1 Mb), same strand

Putative split paralog: Non-overlapping, different regions (e.g. scaffolds)



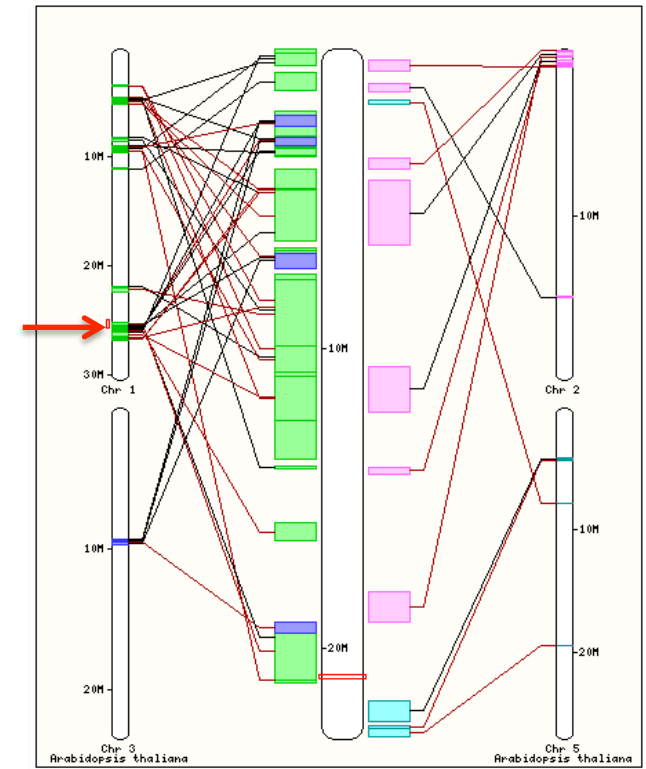
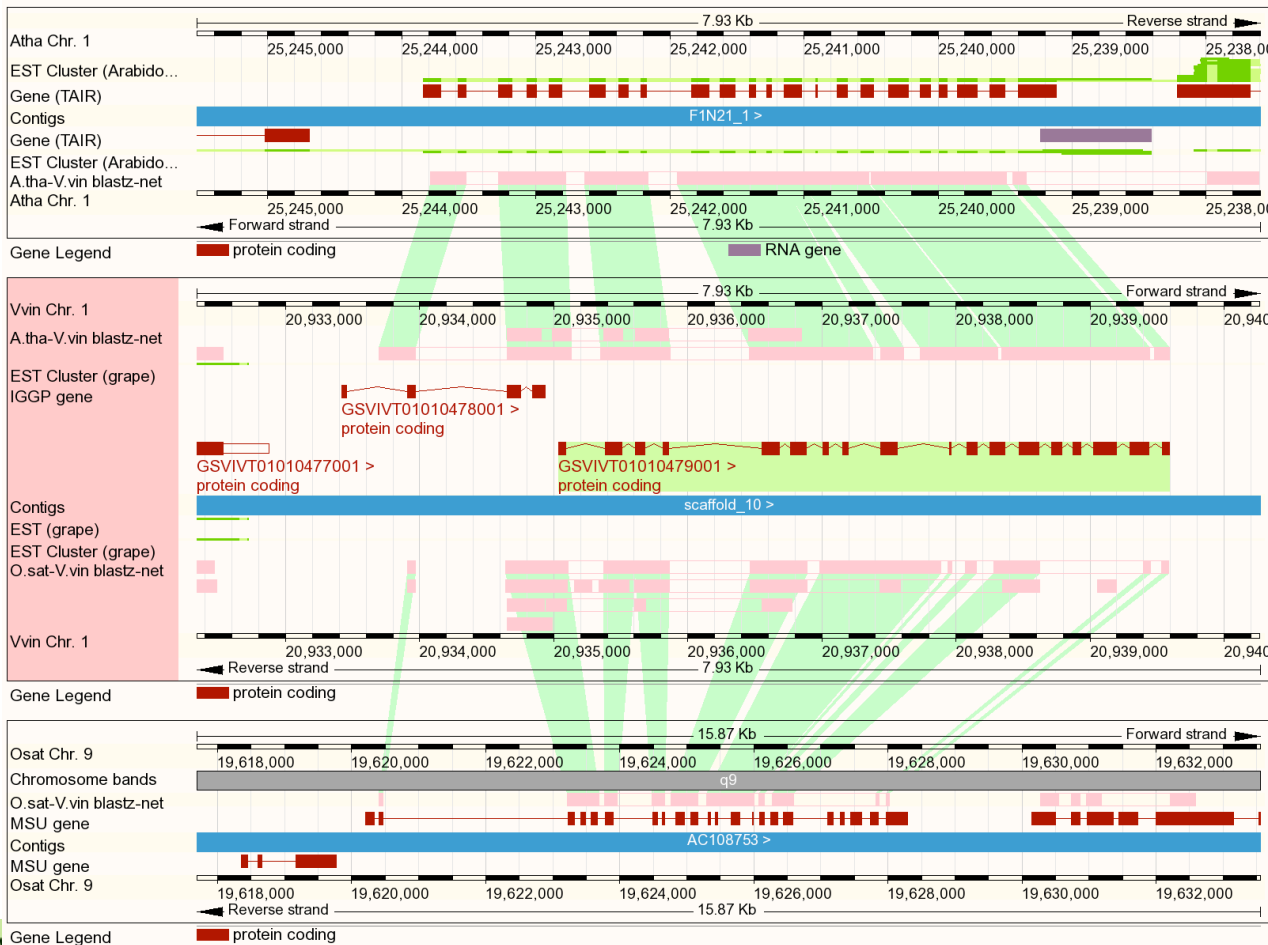
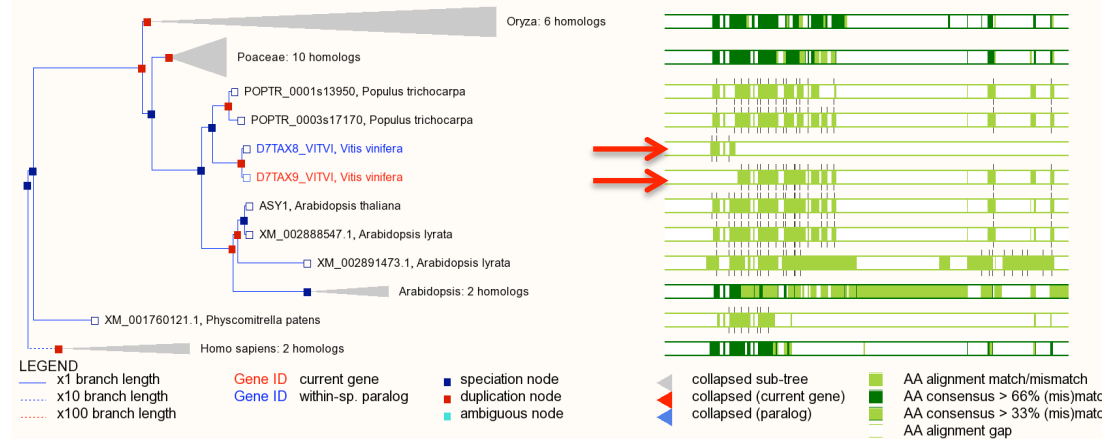
Species	Split Genes
Populus trichocarpa	1181
Sorghum bicolor	1087
Oryza sativa Japonica	916
Vitis vinifera	526
Oryza sativa Indica	365
Zea mays	280
Arabidopsis lyrata	202
Arabidopsis thaliana	137
Brachypodium distachyon	101

Genome alignment confirms inconsistent annotation



# Putative split ex 2:

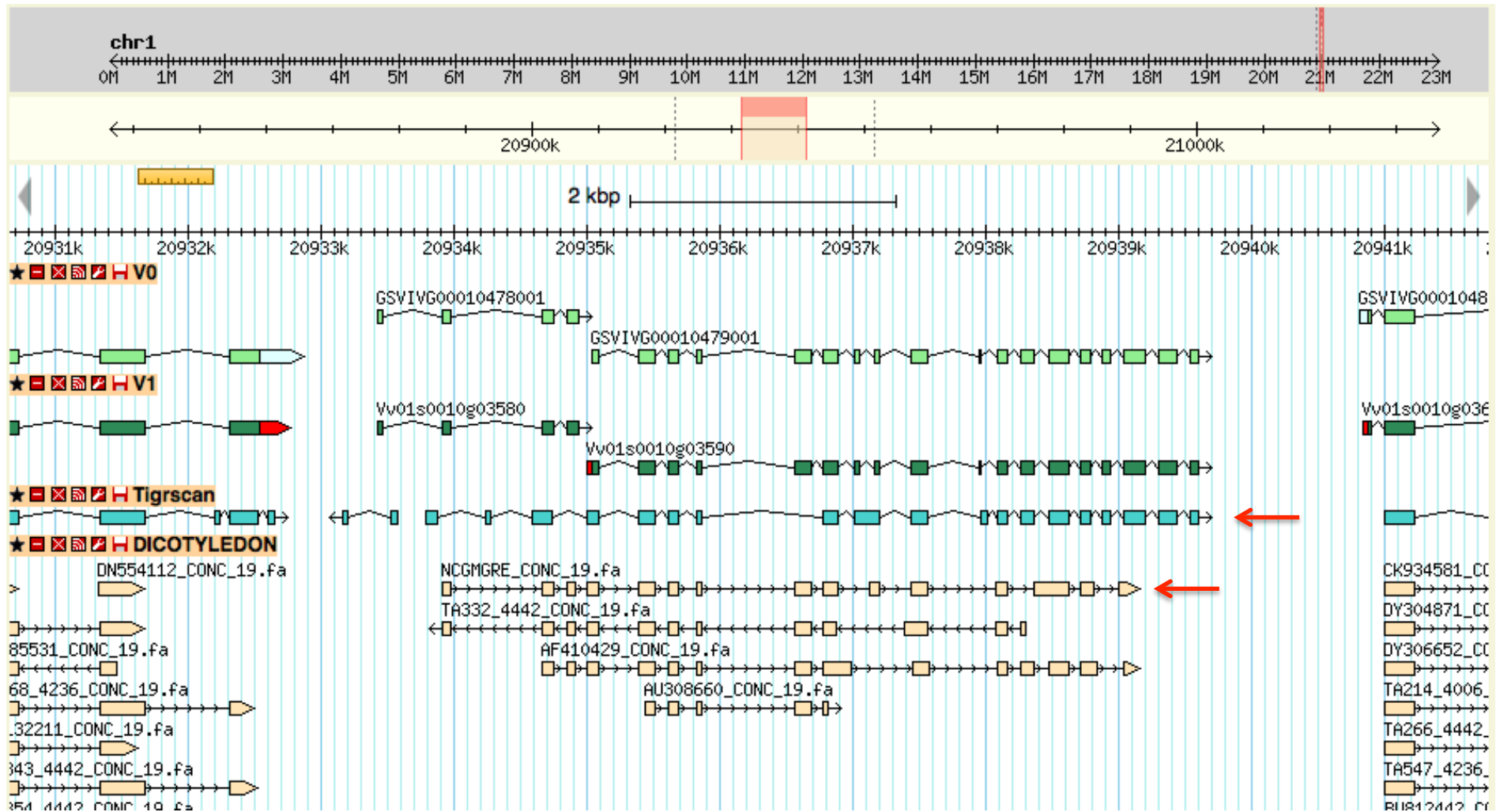
- *Arabidopsis* and *Rice* orthologs both show one gene
- *Arabidopsis* ortholog in correct syntenic context





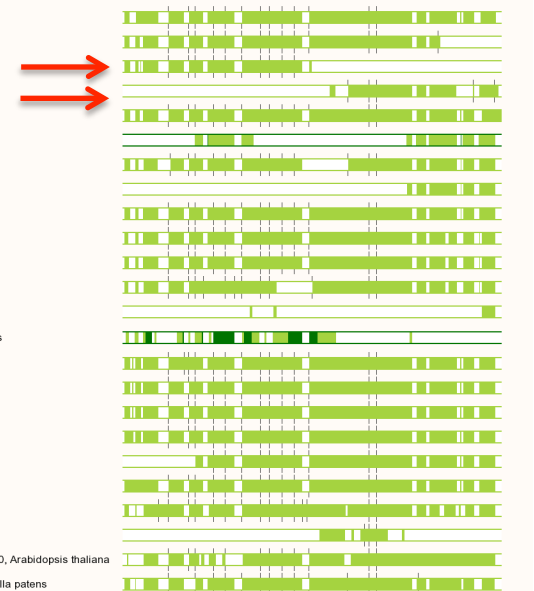
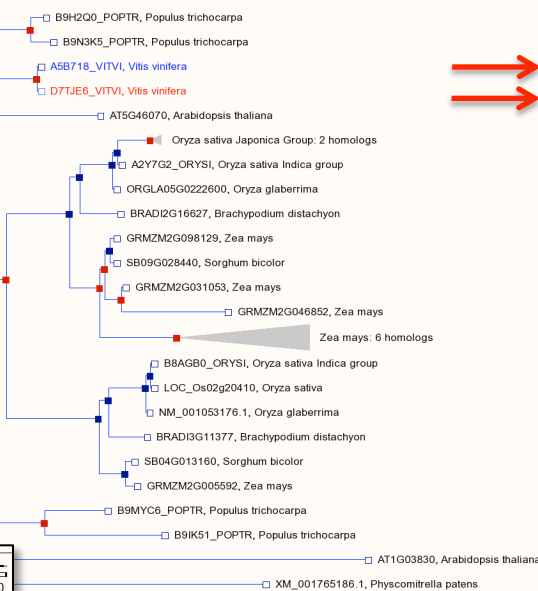
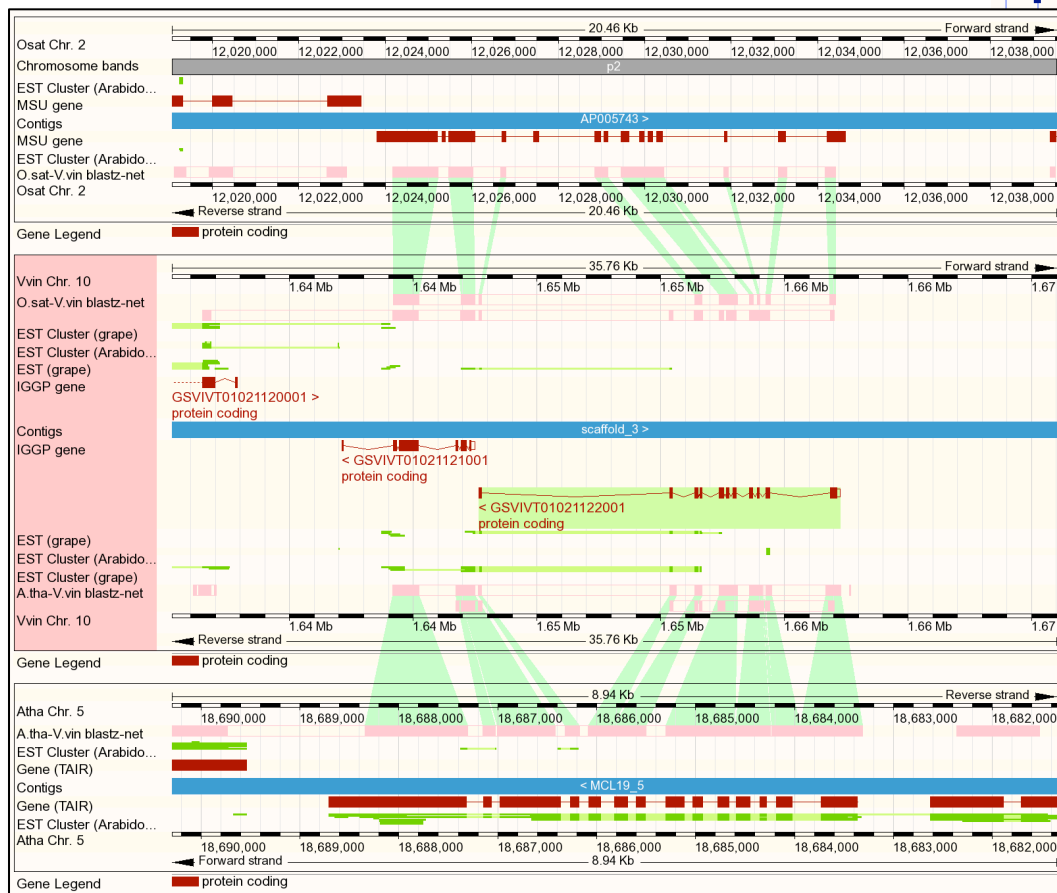
# Putative split model ex. 2

*Models and EST suggest that two genes could be joined*



# Putative split model ex. 1

- *Arabidopsis* and *Rice* orthologs both show one gene
- *Arabidopsis* ortholog in correct syntenic context



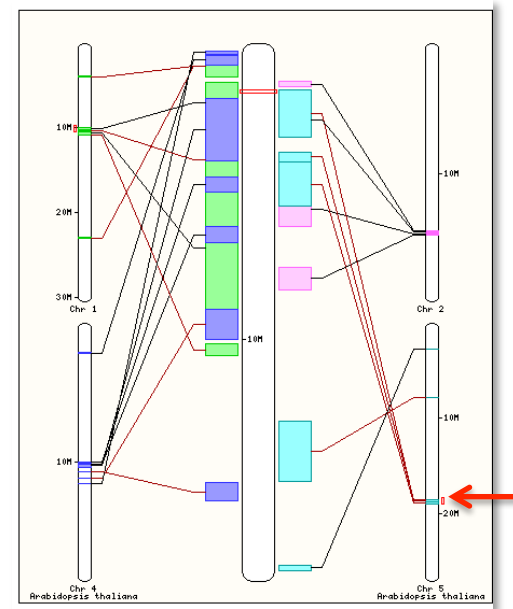
branch length  
branch length  
branch length

Gene ID current gene  
Gene ID within-sp. paralog

■ speciation node  
■ duplication node  
■ ambiguous node

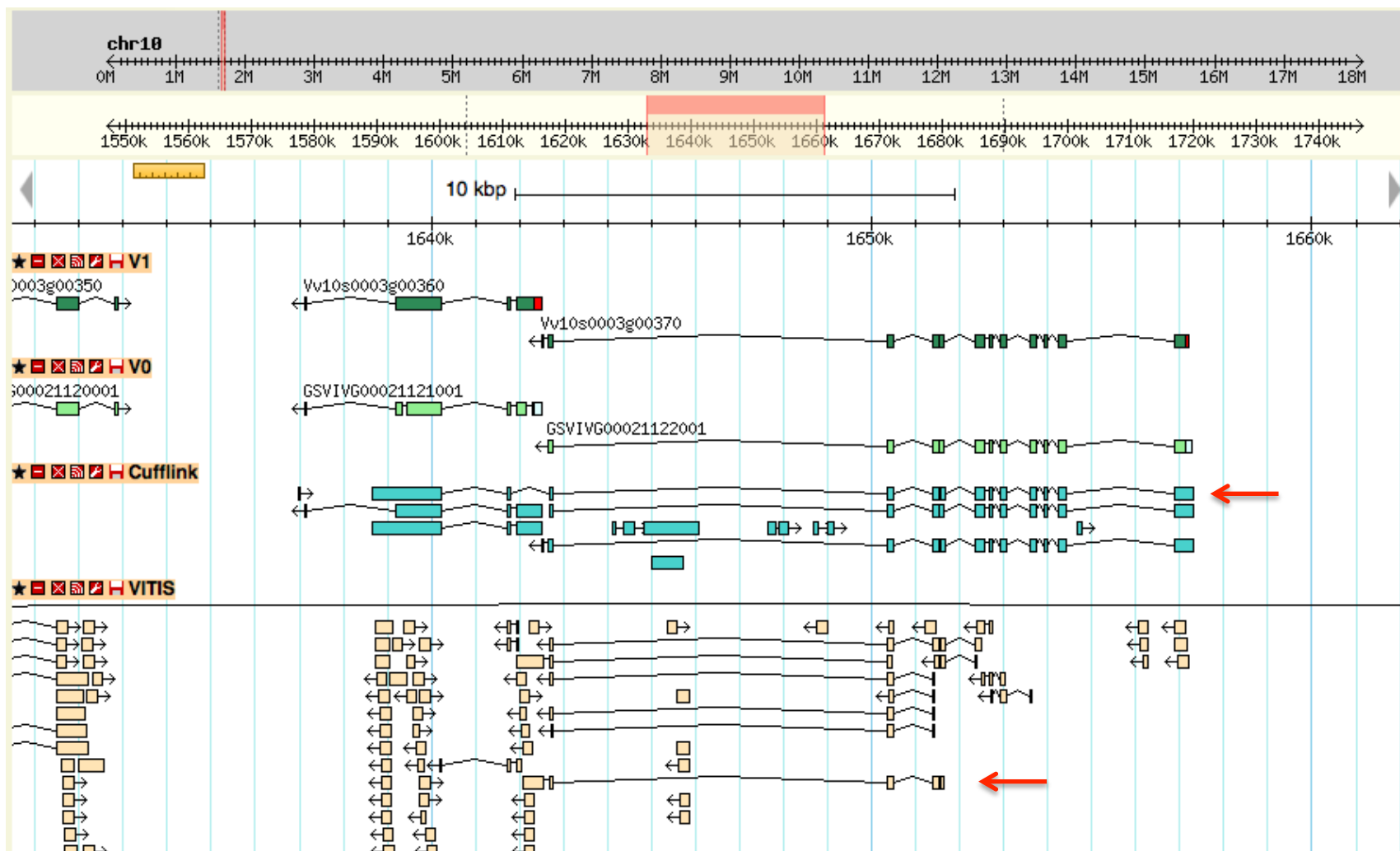
▲ collapsed sub-tree  
▲ collapsed (current gene)  
▲ collapsed (paralog)

■ AA alignment match/mismatch  
■ AA consensus > 66% (mis)/mat  
■ AA consensus > 33% (mis)/mat  
■ AA alignment gap



# Putative split model ex. 1

*Models and EST suggest that two genes could be joined*



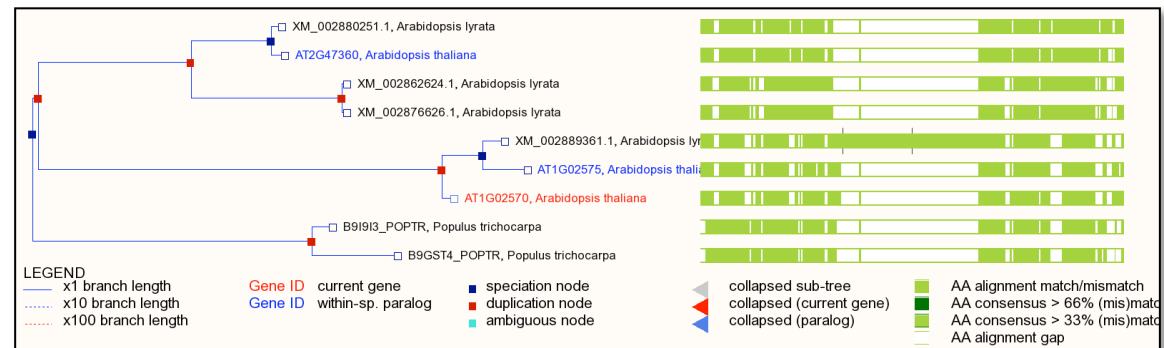
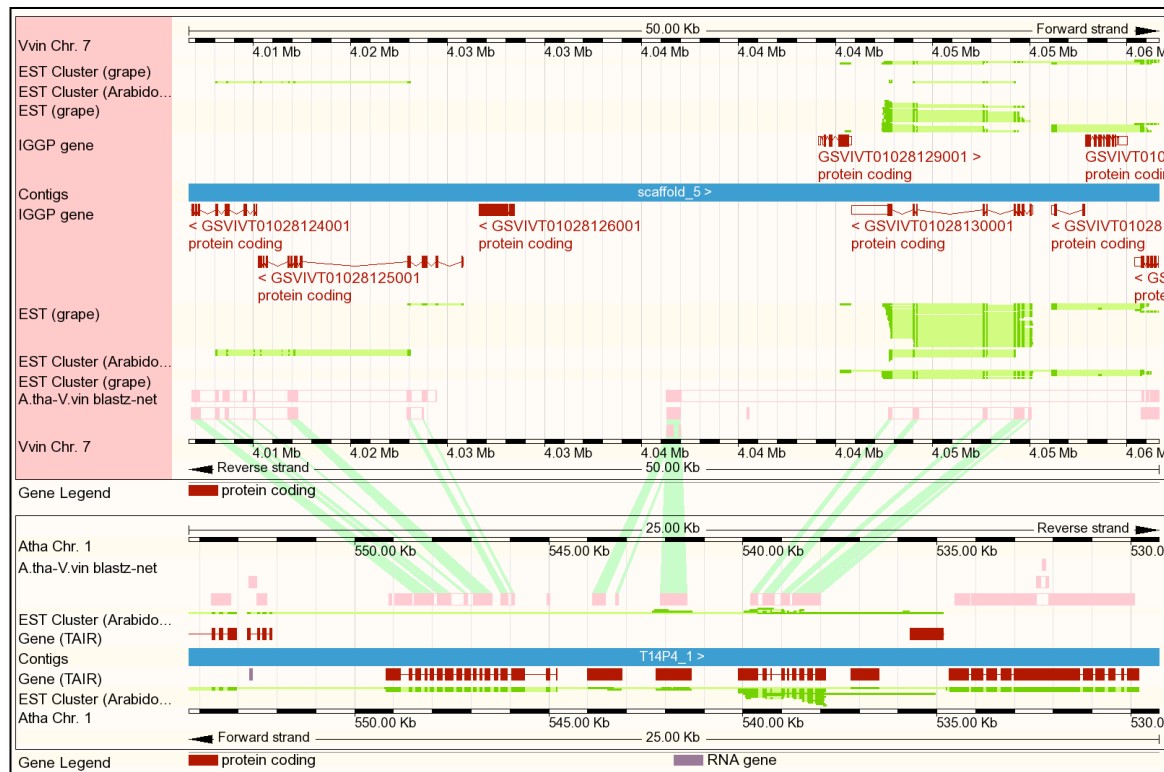
- 446 genes in 134 tandem arrays ("clust\_id")
- 80 genes not in tandem arrays ("NULL")
- URL to gene tree provided

	A	B	C	D	E	F	G	H
1	clust_id	gid	chr	start	index	root_id	URL	IPR string
225	622	GSVIVG0101	15	10879612	272	182691	<a href="http://www.">http://www.</a> IPR002283=Isopenicillin-N_synthase  IPR005123=Oxoglutarate/Fe-dep_oxygenase	
226	622	GSVIVG0101	15	10868314	271	182691	<a href="http://www.">http://www.</a> IPR005123=Oxoglutarate/Fe-dep_oxygenase	
227	618	GSVIVG0101	15	9314196	222	281198	<a href="http://www.">http://www.</a> IPR000423=Flag_FlgJ_1	
228	618	GSVIVG0101	15	9309870	221	281198	<a href="http://www.">http://www.</a> IPR006694=Fatty_acid_hydroxylase	
229	722	GSVIVG0101	16	14592615	528	101868	<a href="http://www.">http://www.</a> IPR001128=Cyt_P450  IPR002401=Cyt_P450_E_grp-III IPR002403=Cyt_P450_E_grp-IV  IPR017973=Cyt_P450_C	
230	722	GSVIVG0101	16	14576735	527	101868	<a href="http://www.">http://www.</a> IPR001128=Cyt_P450  IPR002401=Cyt_P450_E_grp-III IPR002402=Cyt_P450_E_grp-III IPR002403=Cyt_P450_E_grp-IV  IPR017973=Cyt_P450_C	
231	707	GSVIVG0101	16	10962970	389	183990	<a href="http://www.">http://www.</a> IPR005797=Cyt_b/b6_NII IPR016174=Di-haem_cyt_TM	
232	707	GSVIVG0101	16	10962793	388	183990	<a href="http://www.">http://www.</a> IPR016174=Di-haem_cyt_TM	
233	NULL	GSVIVG0101	NULL	NULL	NULL	NULL	<a href="http://www.">http://www.</a> .	
234	599	GSVIVG0101	15	277207	9	171061	<a href="http://www.">http://www.</a> IPR009543=VPSAP	
235	599	GSVIVG0101	15	252946	7	171061	<a href="http://www.">http://www.</a> .	
236	599	GSVIVG0101	15	196036	6	171061	<a href="http://www.">http://www.</a> .	
237	1114	GSVIVG0101	2	3511310	363	148463	<a href="http://www.">http://www.</a> IPR005202=TF_GRAS	
238	1114	GSVIVG0101	2	3512888	364	148463	<a href="http://www.">http://www.</a> IPR005202=TF_GRAS	
239	1082	GSVIVG0102	19	21600092	1068	362943	<a href="http://www.">http://www.</a> IPR005121=PheS_beta_Fdx_antiC_bd	
240	1082	GSVIVG0102	19	21573407	1066	362943	<a href="http://www.">http://www.</a> .	
241	NULL	GSVIVG0102	NULL	NULL	NULL	NULL	<a href="http://www.">http://www.</a> .	
242	1079	GSVIVG0102	19	21167540	1051	362943	<a href="http://www.">http://www.</a> .	
243	1079	GSVIVG0102	19	21087508	1049	362943	<a href="http://www.">http://www.</a> IPR002319=Phenylalanyl-tRNA_Synthase_acu IPR004530=Phe-tRNA-synth_IIc_mito IPR005121=PheS_beta_Fdx_antiC_bd	
244	1079	GSVIVG0102	19	20987714	1044	362943	<a href="http://www.">http://www.</a> IPR002319=Phenylalanyl-tRNA_Synthase_acu	
245	NULL	GSVIVG0102	NULL	NULL	NULL	NULL	<a href="http://www.">http://www.</a> .	
246	NULL	GSVIVG0102	NULL	NULL	NULL	NULL	<a href="http://www.">http://www.</a> IPR002885=Pentatricopeptide_repeat	
247	NULL	GSVIVG0102	NULL	NULL	NULL	NULL	<a href="http://www.">http://www.</a> IPR001611=Leu-rich_rpt IPR003591=Leu-rich_rpt_typical_subtyp IPR020474=Toll-like_rcpt_LRR-contain	
248	113	GSVIVG0102	10	1637127	147	60567	<a href="http://www.">http://www.</a> IPR000533=Tropomyosin	
249	113	GSVIVG0102	10	1642651	148	60567	<a href="http://www.">http://www.</a> IPR015894=Guanylate-bd_N	
250	124	GSVIVG0102	10	4358782	322	182691	<a href="http://www.">http://www.</a> .	
251	124	GSVIVG0102	10	4360092	323	182691	<a href="http://www.">http://www.</a> IPR005123=Oxoglutarate/Fe-dep_oxygenase	
252	124	GSVIVG0102	10	4361857	324	182691	<a href="http://www.">http://www.</a> IPR005123=Oxoglutarate/Fe-dep_oxygenase	
253	124	GSVIVG0102	10	4380211	326	182691	<a href="http://www.">http://www.</a> IPR002452=Alpha_tubulin IPR005123=Oxoglutarate/Fe-dep_oxygenase	
254	124	GSVIVG0102	10	4420780	327	182691	<a href="http://www.">http://www.</a> .	
255	124	GSVIVG0102	10	4421896	328	182691	<a href="http://www.">http://www.</a> IPR005123=Oxoglutarate/Fe-dep_oxygenase	
256	124	GSVIVG0102	10	4435330	330	182691	<a href="http://www.">http://www.</a> IPR005123=Oxoglutarate/Fe-dep_oxygenase	
257	124	GSVIVG0102	10	4441034	332	182691	<a href="http://www.">http://www.</a> .	
258	135	GSVIVG0102	10	5939677	419	240317	<a href="http://www.">http://www.</a> IPR001330=Prenyltrans IPR008930=Terpenoid_cyclase/PrenylTrfase IPR018333=Squalene_cyclase	

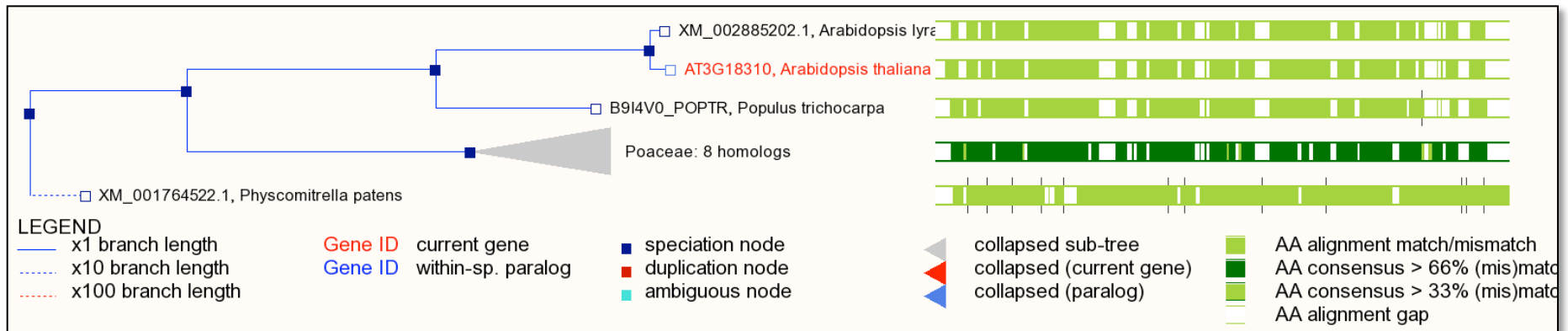
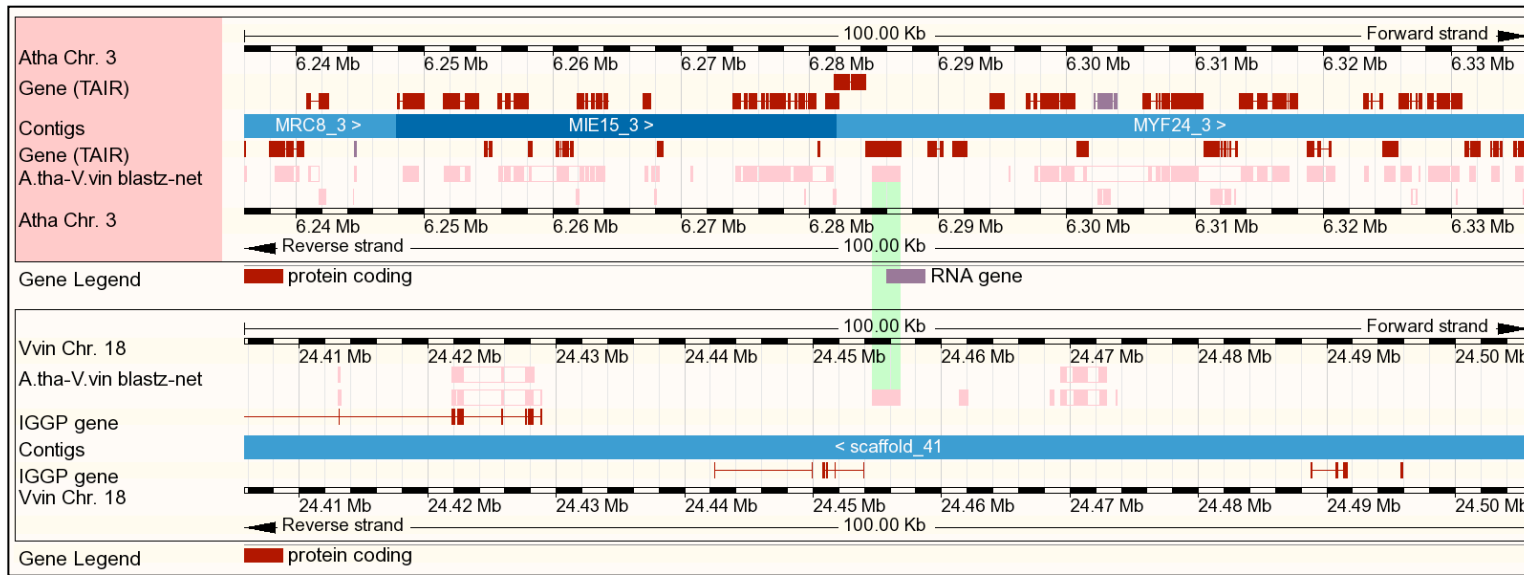
# Putative "missing" genes

- Arabidopsis gene aligns to grape region that lacks a v0 gene
- Can be in syntenic region or not

## Example 1



# Putative "missing" example 2



# v0\_putative\_missing.txt

- 1837 Genes in Arabidopsis align to locations that lack an annotated v0 gene.
  - Frequently single exon genes
  - Frequently ncRNA genes
  - Sometimes transposon (rare in TAIR9 annotations)
- Prioritize 754 having InterPro domains in Arabidopsis, e.g.:
  - 9 have HLH DNA-binding domains; 5 bZIP TF; 3 MYB
  - 28 Disease R genes
  - 22 with protein kinase domains
  - 2 cyt-P450

	A	B	C	D
1	Vv_coord	At gene	Arabidopsis ipr_string	
26	6:7727097:7727163:1	AT5G07160	IPR004827=TF_bZIP  IPR008917=Euk_TF_DNA-bd  IPR011700=bZIP_2	
27	3:7530481:7530638:-1	AT5G18755		
28	8:20908785:20909542:1, 8:20909642:20909776:-1	AT5G18130		
29	15:4974459:4974631:1	AT4G06614		
30	2:7803522:7804476:-1	AT4G17920	IPR001841=Znf_RING  IPR011016=Znf_RING-CH  IPR018957=Znf_C3HC4_RING-type	
31	Un:7168439:7170706:-1	AT1G74350	IPR000442=Intron_maturse2  IPR000477=RVT	
32	6:6377061:6377358:-1	AT1G47595		
33	1:4191322:4191633:-1	AT2G01300		
34	5:24351620:24351684:1	AT4G15460		
35	12:14666061:14666197:1	AT5G11475		
36	5:2101385:2101497:1	AT3G10845	IPR000504=RRM_RNP1	
37	3:8267458:8267675:-1	AT3G19660		
38	9:6932975:6934203:1	AT1G80080	IPR001611=Leu-rich_rpt  IPR003591=Leu-rich_rpt_typical-subtyp	
39	9:5004363:5004689:1	AT1G49020		
40	10:2995809:2995910:-1	AT2G34204		
41	18:3219646:3220360:1	AT1G44830	IPR000843=HTH_LacI  IPR001471=TF_ERF_DNA-bd  IPR016177=DNA-bd_integrase-typ	
42	4:22820725:22820776:-1	AT5G66820		
43	4:21244402:21244689:1, 17:6320395:6320447:1	AT1G68170	IPR000620=DMT	
44	7:974789:975209:1	AT1G07130	IPR004365=NA-bd_OB_tRNA-helicase  IPR016027=NA-bd_OB-fold-like	
45	5:24807730:24807917:1	AT2G44080		
46	16:17796185:17796679:-1, 16:17795822:17795959:-1	AT5G23950	IPR000008=C2_Ca-dep  IPR008973=C2_Ca/lipid-bd_dom_CaLB	
47	16:21843281:21844104:-1	AT5G52990	IPR011012=Longin-like	
48	3:7530481:7530638:-1	AT5G18755		

# Project Function From Orthologs

16,810 grape genes orthologous to 19,080 Arabidopsis genes

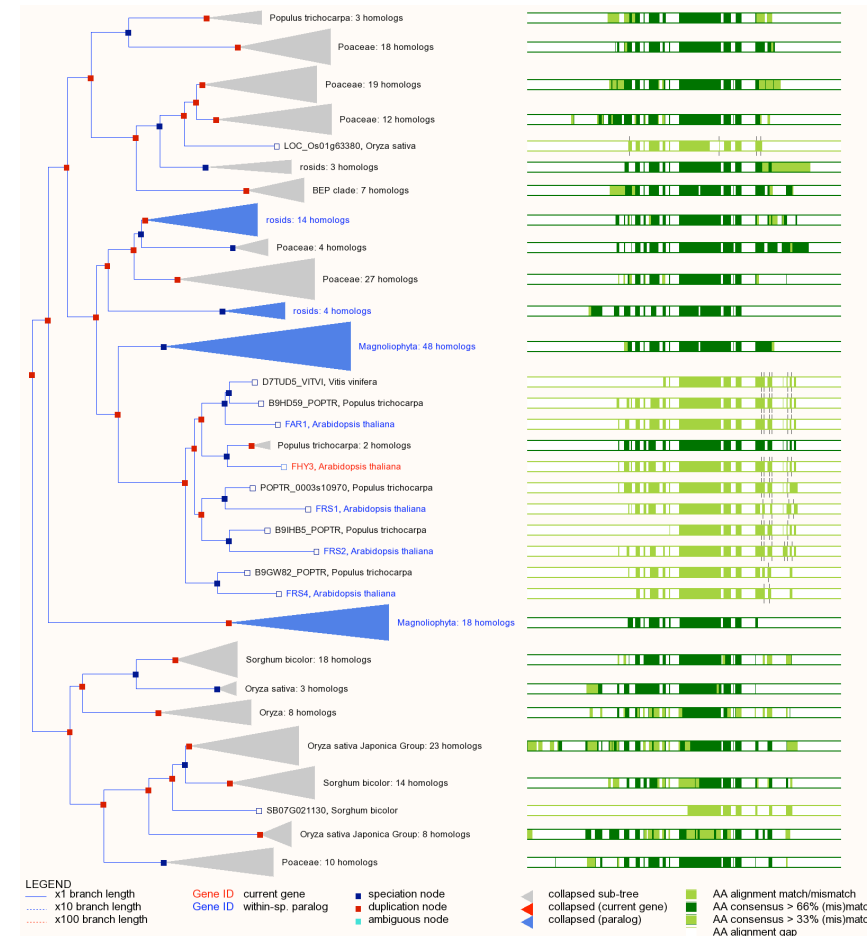
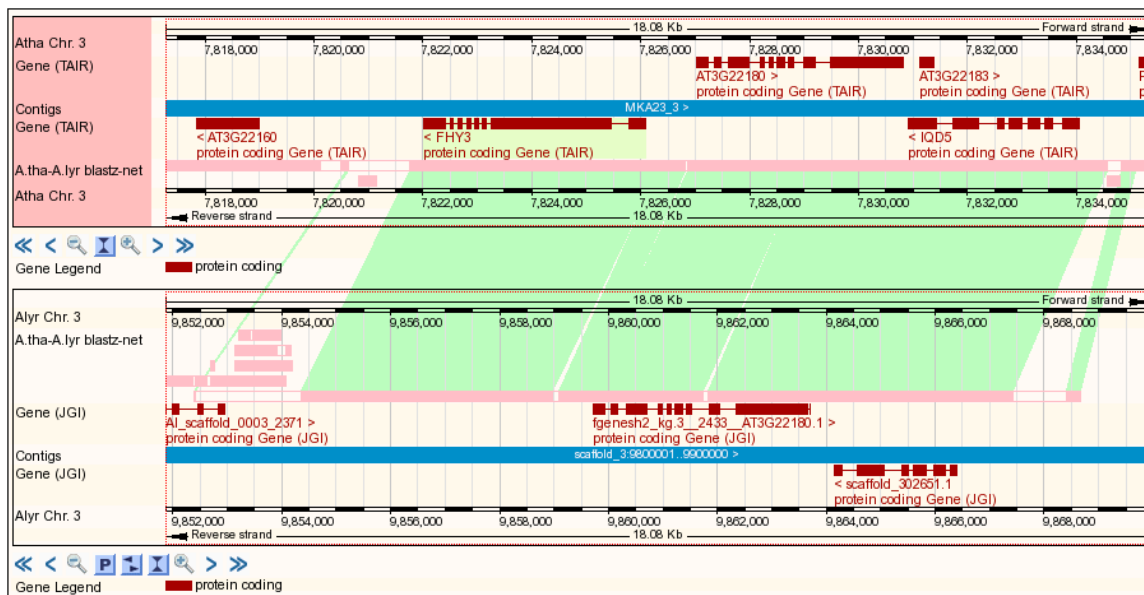
vv\_at\_orthologs.txt

	A	B	C	D	E	F	G	H	I
1	Grape Gene ID	Chromosome	Gene Start (bp)	Gene End (bp)	Arabidopsis Gene	Arabidopsis Chromosome	Arabidopsis Start (bp)	Arabidopsis End (bp)	Orthology Type
2	GSVIVG01000001001	14	2596280	2598951					
3	GSVIVG01000002001	14	2600404	2607034					
4	GSVIVG01000003001	14	2616036	2620460	AT5G27390	5	9674391	9676189	ortholog_one2one
5	GSVIVG01000004001	14	2638641	2642144	AT1G54780	1	20439434	20441159	ortholog_one2one
6	GSVIVG01000005001	14	2642583	2643125	AT1G54790	1	20440979	20444029	ortholog_many2many
7	GSVIVG01000005001	14	2642583	2643125	AT3G05180	3	1468398	1470551	ortholog_many2many
8	GSVIVG01000006001	14	2643130	2644348	AT1G54790	1	20440979	20444029	ortholog_many2many
9	GSVIVG01000006001	14	2643130	2644348	AT3G05180	3	1468398	1470551	ortholog_many2many
10	GSVIVG01000007001	14	2645322	2647029	AT1G54790	1	20440979	20444029	ortholog_many2many
11	GSVIVG01000007001	14	2645322	2647029	AT3G05180	3	1468398	1470551	ortholog_many2many
12	GSVIVG01000008001	14	2650299	2654390	AT3G63480	3	23440887	23444006	ortholog_one2one
13	GSVIVG01000009001	14	2656366	2692593	AT3G05190	3	1471397	1475784	ortholog_one2many
14	GSVIVG01000009001	14	2656366	2692593	AT5G27410	5	9676334	9682940	ortholog_one2many
15	GSVIVG01000010001	14	2692594	2693228					



# Domesticated Transposons

FAR1/FHY3 family, transcription factors that evolved from Mu-related transposons and in Arabidopsis function in light perception and signaling



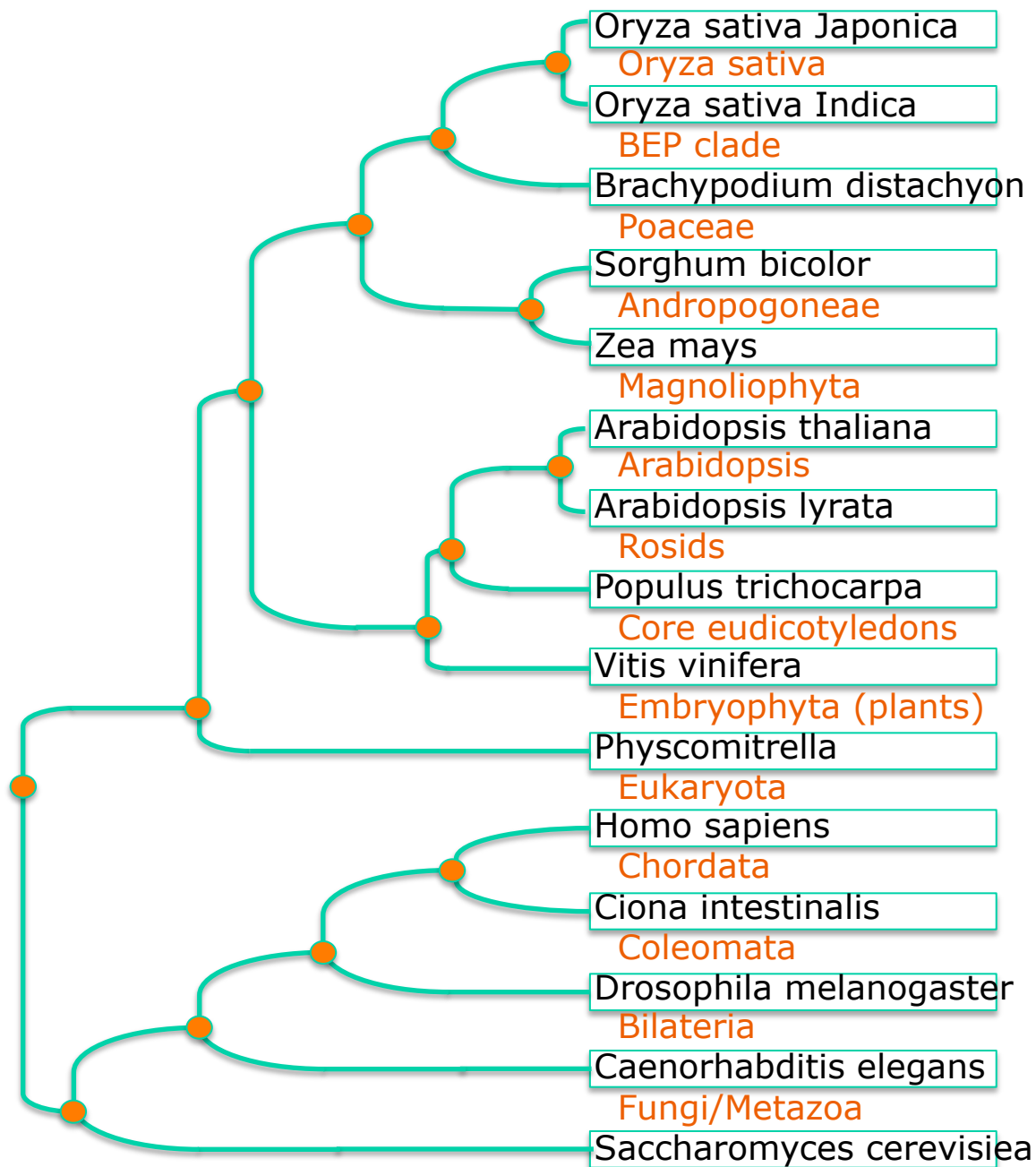
# Conclusion

- Gramene's goal is to use comparative genomics to improve structural and functional annotation.
- We need **FEEDBACK!**

# The Ware Lab


- Doreen Ware
- Sharon Wei
- Will Spooner
- Ken Youens-Clark
- Shiran Pasternak
- Jim Thomason
- Aaron Chuah
- Marcela Monaco
- Liya Ren (past)
- Jerry Lu





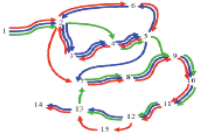
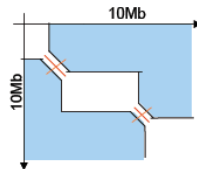

# EPO Multiple Alignment & Ancestor Reconstruction


- Gramene implementation in 2010
- Release 32: 8-way EPO alignment
  - Rice japonica, indica, Brachypodium, sorghum, Arabidopsis, A. lyrata, grape, poplar



## Pipeline Overview

- **Enredo**
  - Defines blocks of collinear sequences
  - Supports segmental duplications
- **Pecan**
  - Consistency based multiple aligner
  - Optimized to cope with long genomic sequences
- **Ortheus**
  - Ancestral sequences reconstructor (Tree Aligner)
  - Infers the history of insertion and deletions



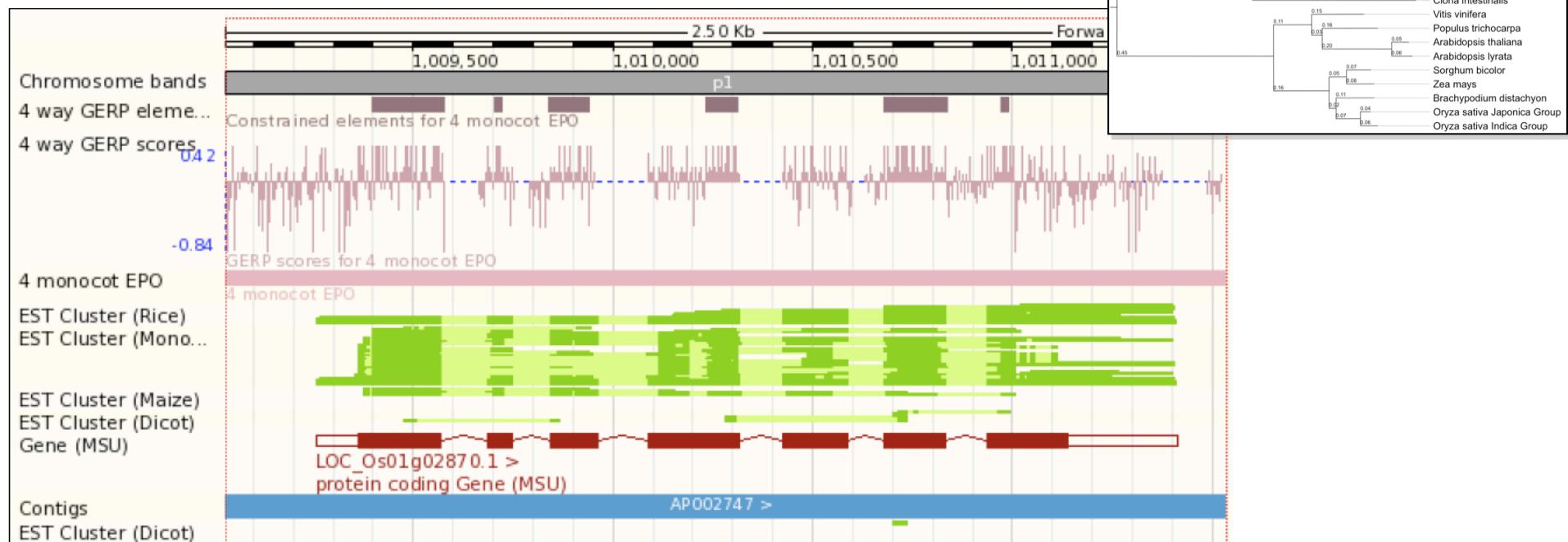
Nov 2009



Paten et al (2008) Genome Research 18:1814  
 Paten et al (2008) Genome Research 18:1829

# 2010 Genomes Development: Constrained Elements

- Genomic Evolutionary Rate Profiling (GERP): measures purifying selection
- Method testing using 4-way and 8-way EPO alignments as input with varying parameters
- Input tree generated from 1301 ortholog sets
- Planning release in 2011



Cooper et al (2005) Genome Research 15:901