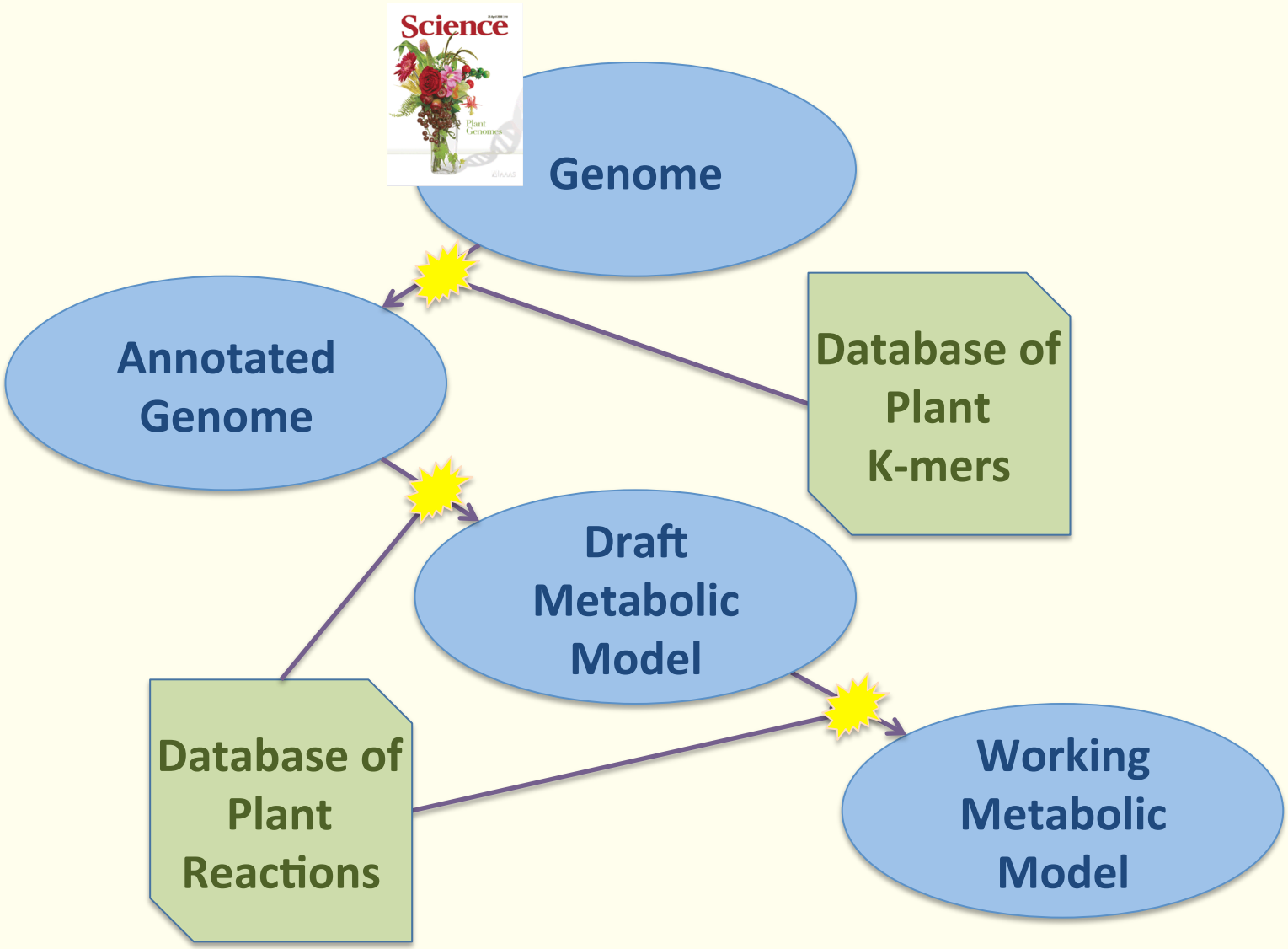
PlantSEED Workshop

Overview of PlantSEED curation, subsystems, and K-mers

PlantSEED Workflow

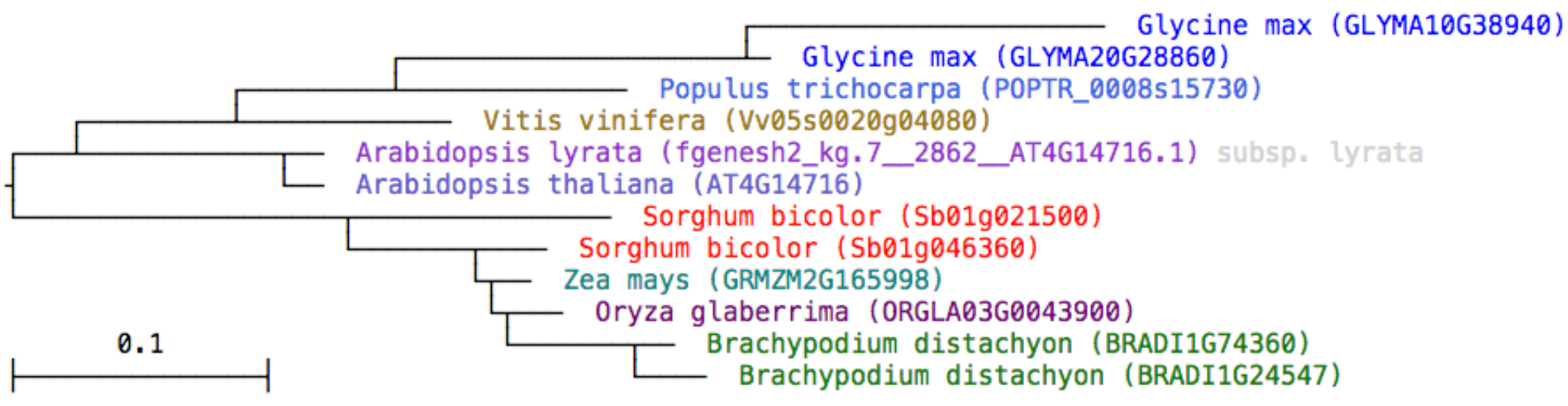


Plant K-mers

- A K-mer is a short (8 AAs) peptide sequence
  - It is a unique representation of a protein function:
    - MTIRNQRF
    - Ubiquinol--cytochrome c reductase, cytochrome B subunit (EC 1.10.2.2)
- K-mers are generated via FigFams/Subsystems
- K-mers are utilized via RAST
  - **R**apid **A**nnotation using **Subsystem Technology**

Plant FigFams

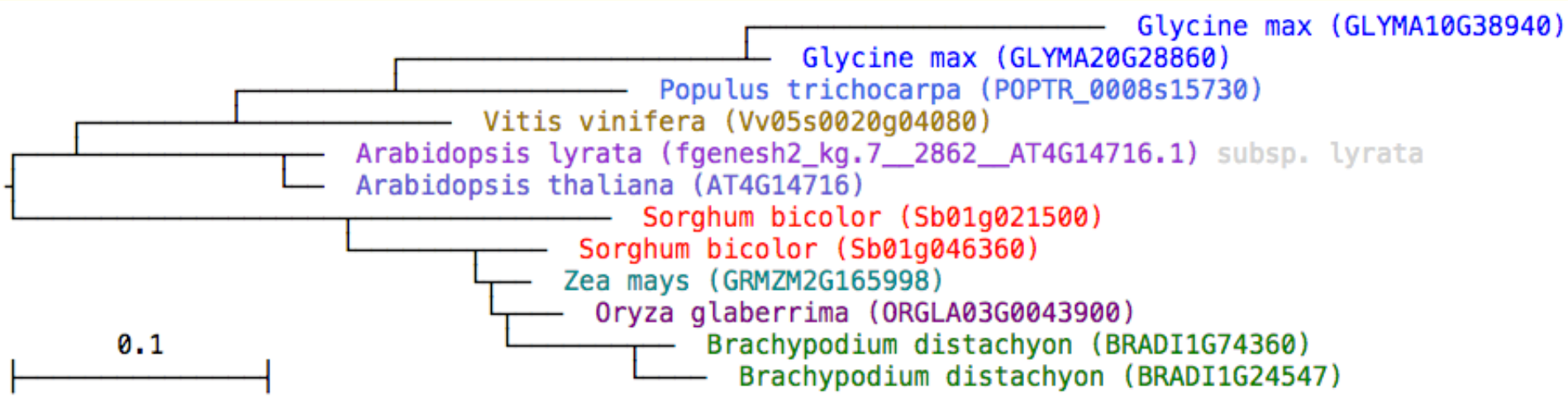
- A set of isofunctional protein sequences



1, 2-dihydroxy-3-keto-5-methylthiopentene dioxygenase (EC 1.13.11.54)

Plant FigFams

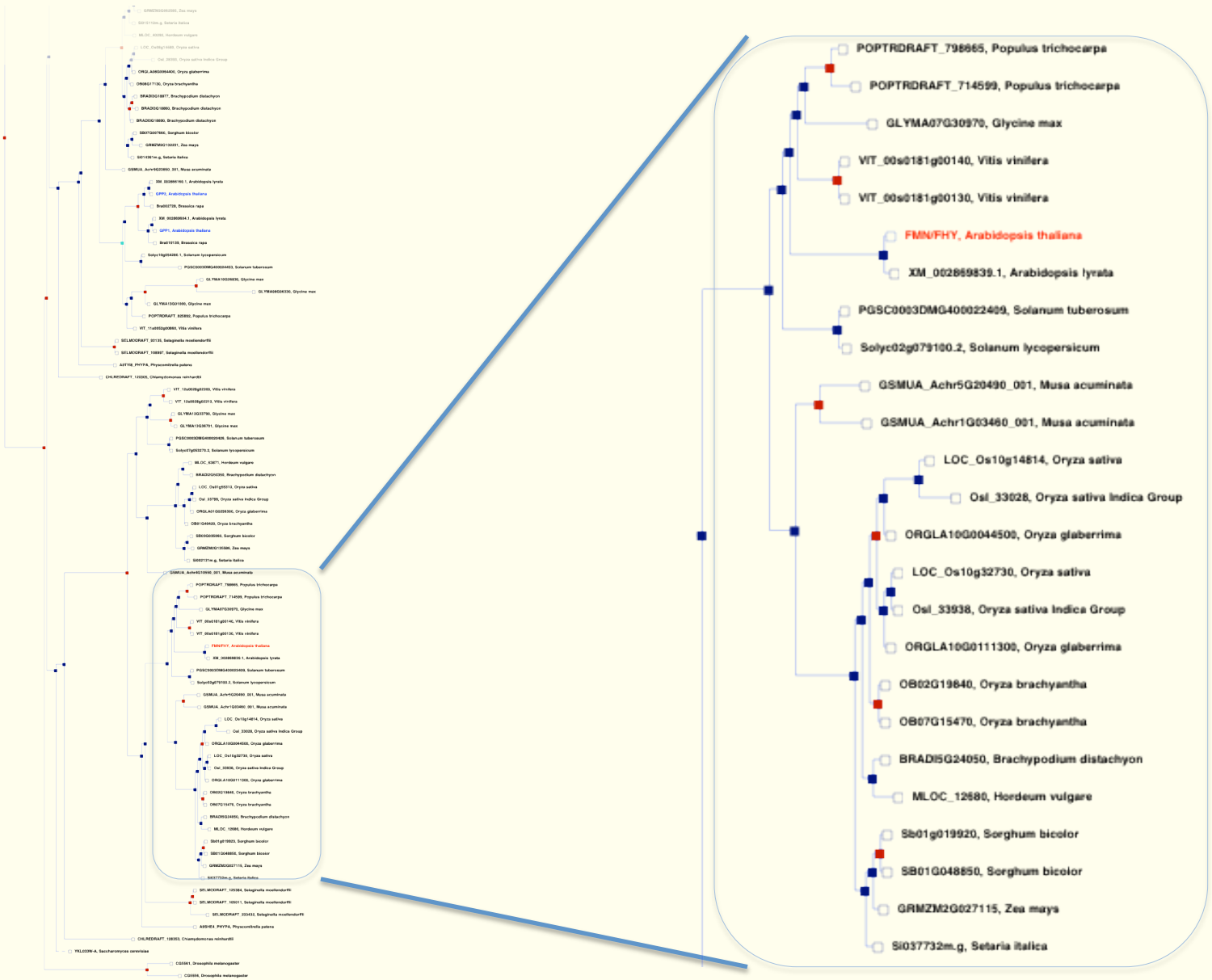
- A set of isofunctional protein sequences
  - A k-mer is a unique signature of a FigFam
  - It is *not found* in other FigFams



Plant Subsystems
(Acetyl-CoA carboxylase complexes)

Organism	ACAC_hom	BCCP_hom	BC_hom	FabD	FabB_mt
	All	All	All	All	All
Oryza glaberrima (4538.3)	20106	20106	20106	13920	9880
Oryza sativa (39947.4)	25448	25448	25448	13923	7611
Brachypodium distachyon (15368.4)	362	362	362	25177	7672
Sorghum bicolor (4558.4)	10487	10487	10487	32585	16469
Zea mays (381124.6)	19643, 38069	19643, 38069	19643, 38069	2295, 4271	15600
Glycine max (3847.5)	13852, 9016	13852, 9016	13852, 9016	29908, 29909, 46607	25402, 25970, 34049, 49577, 6226
Populus trichocarpa (3694.3)	28459, 34988	28459, 34988	28459, 34988	19427, 39501	36220
Arabidopsis lyrata subsp. lyrata (81972.4)	31073, 31074	31073, 31074	31073, 31074	17904	23087
Arabidopsis thaliana (3702.11)	23857, 23858	23857, 23858	23857, 23858	18403	16440
Vitis vinifera (29760.3)	13415	13415	13415	5242	22661

Plant Protein Families



Gramene.org

Plant Protein Families

Role	Subsystem	Genes
1, 2-dihydroxy-3-keto-5-methylthiopentene dioxygenase (EC 1.13.11.54)	Methionine and cysteine metabolism in plants	AT2G26400, fgenesh2_kg.4__555__AT2G26400.1
1, 2-dihydroxy-3-keto-5-methylthiopentene dioxygenase (EC 1.13.11.54)	Methionine and cysteine metabolism in plants	AT4G14710, GLYMA07G13585, ORGLA10G0088200, scaffold_703119.1
1, 2-dihydroxy-3-keto-5-methylthiopentene dioxygenase (EC 1.13.11.54)	Methionine and cysteine metabolism in plants	AT4G14716, BRADI1G24547, BRADI1G74360, GLYMA10G38940, GLYMA20G28860, GRMZM2G165998, ORGLA03G0043900, OS03G0161800, OS10G0419400, POPTR_0008s15730, Sb01g021500, Sb01g046360, Vv05s0020g04080, fgenesh2_kg.7__2862__AT4G14716.1
1, 2-dihydroxy-3-keto-5-methylthiopentene dioxygenase (EC 1.13.11.54)	Methionine and cysteine metabolism in plants	AT5G43850, BRADI3G26930, GLYMA10G38960, GLYMA20G28850, GRMZM2G433446, GRMZM2G434572, ORGLA10G0088300, OS10G0419500, POPTR_0008s15720, Vv05s0020g04070, fgenesh2_kg.8__436__AT5G43850.1

- Conservative use will result in ‘splitting families’
- Reduces over-annotation

Curation of plant primary metabolism

- Gene – reaction associations available in many places:
  - KEGG
  - AraCyc
  - MetaCyc
  - BRENDA
  - UniProt
  - Gramene
  - Phytozome

Curation of plant primary metabolism

- Most gene-reaction associations are predictions
  - AraCyc:
    - 9,715/12,447 (78%)
  - PoplarCyc:
    - 14,154/14,162 (99.9%)
- Lack of Transparency

Enzymes and Genes:
L-tryptophan aminotransferase : TAR2
tryptophan aminotransferase : TAR1
tryptophan aminotransferase : TAA1

Enzymes and Genes:
Potri.008G187800.1 : Potri.008G187800.1
Potri.015G081900.1 : Potri.015G081900.1
Potri.010G044500.1 : Potri.010G044500.1
Potri.012G083300.1 : Potri.012G083300.1

Curation of plant primary metabolism

- We intensively curated:
  - 361 pathways of primary metabolism pathways
    - ~100 Subsystems
    - ~2,400 Genes
    - ~1,300 Reactions

Curation of plant primary metabolism

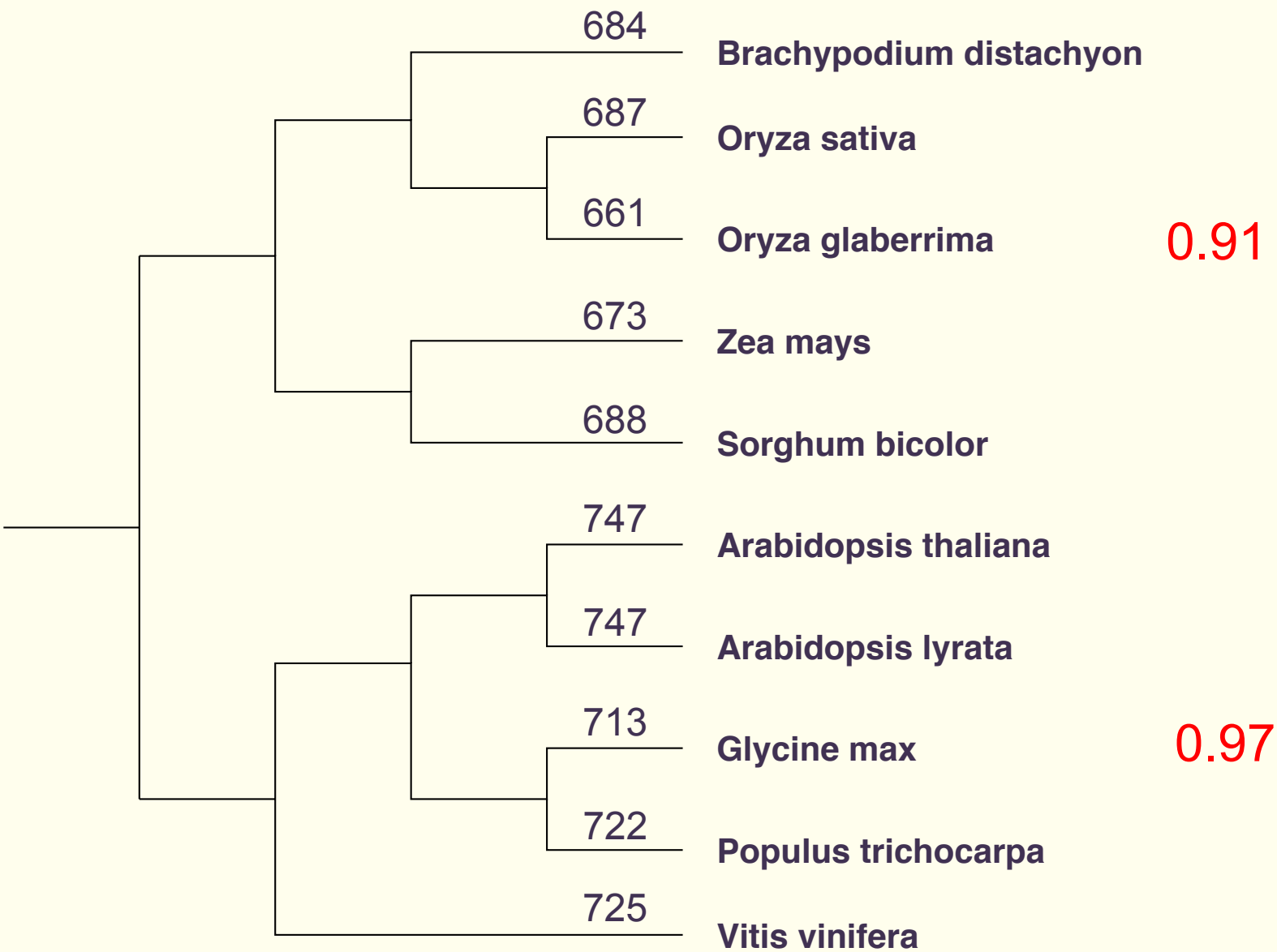
Riboflavin kinase (EC 2.7.1.26)

~~AT1G56500~~
~~AT2G38740~~
~~AT3G48420~~
~~AT4G25840~~
~~AT4G11570~~
~~AT4G39970~~
AT4G21470
~~AT5G57440~~

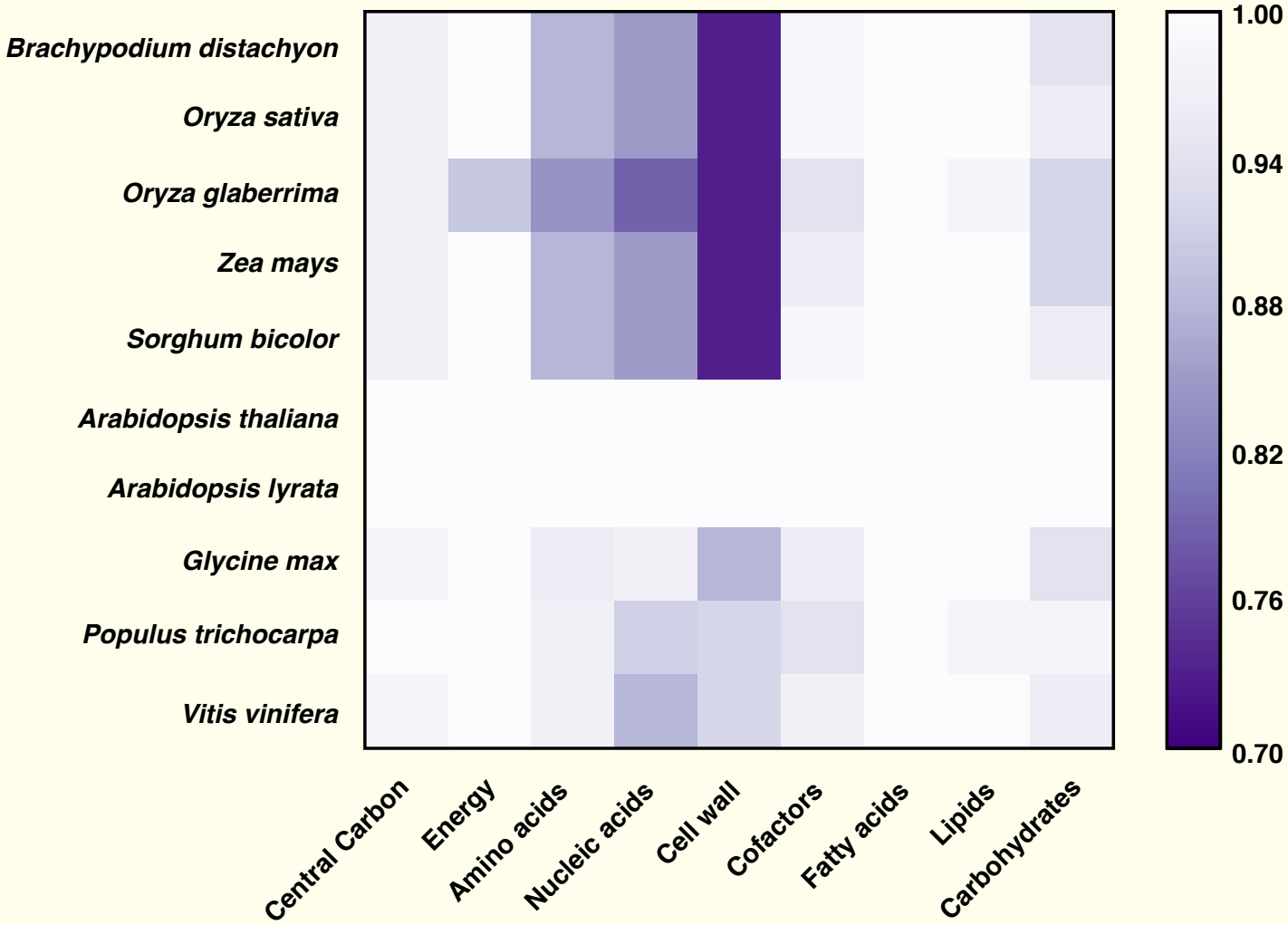
Propagation of plant primary metabolism

Organism	ACAC_hom	BCCP_hom	BC_hom	FabD	FabB_mt
	All	All	All	All	All
Oryza glaberrima (4538.3)	20106	20106	20106	13920	9880
Oryza sativa (39947.4)	25448	25448	25448	13923	7611
Brachypodium distachyon (15368.4)	362	362	362	25177	7672
Sorghum bicolor (4558.4)	10487	10487	10487	32585	16469
Zea mays (381124.6)	19643, 38069	19643, 38069	19643, 38069	2295, 4271	15600
Glycine max (3847.5)	13852, 9016	13852, 9016	13852, 9016	29908, 29909, 46607	25402, 25970, 34049, 49577, 6226
Populus trichocarpa (3694.3)	28459, 34988	28459, 34988	28459, 34988	19427, 39501	36220
Arabidopsis lyrata subsp. lyrata (81972.4)	31073, 31074	31073, 31074	31073, 31074	17904	23087
Arabidopsis thaliana (3702.11)	23857, 23858	23857, 23858	23857, 23858	18403	16440
Vitis vinifera (29760.3)	13415	13415	13415	5242	22661

Propagation of plant primary metabolism

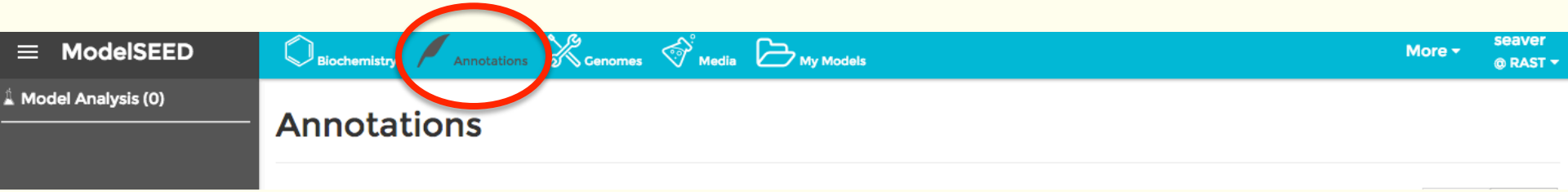


Propagation of plant primary metabolism



Viewing PlantSEED Annotation

- Master List of Subsystems:
  - http://plantseed.theseed.org
- PlantSEED Annotation
  - http://plants.theseed.org



Viewing PlantSEED Annotation

- Subsystems
  - linked to SEED
- Pathways
  - linked to AraCyc
- Features
  - linked to curated Arabidopsis genes
- Reactions
  - can be found in Biochemistry
  - will be seen in metabolic reconstruction

Viewing User Annotation

- Nota Bene:
  - The annotation of the public PlantSEED genomes was done using FigFams
  - The annotation of the user genomes was done using K-mers
- Annotation and BLAST runs were done ahead of time
  - In the near future, we plan to have these available as part of the PlantSEED pipeline

Viewing User Annotation

- Features
  - Result of K-mer annotation process (2 minutes)
- Exemplar hits
  - Result of running BLAST against curated Arabidopsis genes (1/2 day – 2 days)

Genome Zmays_Thi2

Features | Annotations

Search annotations

1-10 of 653 results < prev next >

PlantSEED Role	Features	Exemplar Hits
Quinolinate phosphoribosyltransferase [decarboxylating] (EC 2.4.2.19)	GRMZM5G887647	GRMZM2G034079 GRMZM5G887647
5-Enolpyruvylshikimate-3-phosphate synthase (EC 2.5.1.19)	GRMZM5G877500	GRMZM5G877500

Viewing User Annotation

- Look through every functional role in Central Carbon Metabolism
  - Glycolysis, TCA Cycle, PPP
- Does your genome have feature(s) annotated with each function?
- Tip: Use two tabs/pages
  - One for master list of PlantSEED annotation
  - One for annotation of User's genome
- Do the same for a subsystem of choice