PMN Plant Metabolic Network

Combining Computational Prediction and Manual Curation to Create Plant Metabolic Pathway Databases

Peifen Zhang
Carnegie Institution For Science
Department of Plant Biology

Where We Are

Traffic
More...
Map
Satellite
Earth
Greenland
Iceland
Россия
Russia
Suomi Finland
Sverige Sweden
Norge Norway
United Kingdom
Polska Poland
Deutschland Germany
Україна Ukraine
France
Italia Italy
España Spain
Türkiye Turkey
Kazakhstan
Mongolia
中国
China
S. Korea
Japan
Iraq
Iran
Afghanistan
Pakistan
Saudi Arabia
Egypt
Libya
Algeria
Mauritania
Mali
Niger
Chad
Sudan
Nigeria
Ethiopia
DR Congo
Kenya
Tanzania
Angola
Namibia
Botswana
South Africa
Madagascar
India
Thailand
Indonesia
Papua New Guinea
Australia
New Zealand
North Atlantic Ocean
South Atlantic Ocean
Indian Ocean
North Pacific Ocean
South Pacific Ocean
Canada
United States
A
Mexico
Venezuela
Colombia
Peru
Bolivia
Brasil
Brazil
Chile
Argentina
2000 mi
2000 km
©2010 Google - Map data ©2010 AND, Geocentre Consulting, MapData Sciences Pty Ltd, PSMA, MapLink, Tele Atlas - Terms of Use

Who We Are

National Science Foundation
WHERE DISCOVERIES BEGIN

PMN:

- Sue Rhee (*PI*)
- Kate Dreher (*curator*)
- Lee Chae (Postdoc)
- Anjo Chi (*programmer*)
- Cynthia Lee (TAIR *tech team*)
- Larry Ploetz (TAIR *tech team*)
- Shanker Singh (TAIR *tech team*)
- Bob Muller (TAIR *tech team*)

Key Collaborators:

- Peter Karp (MetaCyc, SRI)
- Ron Caspi (MetaCYc, SRI)
- Lukas Mueller (SGN)
- Anuradha Pujar (SGN)

http://plantcyc.org

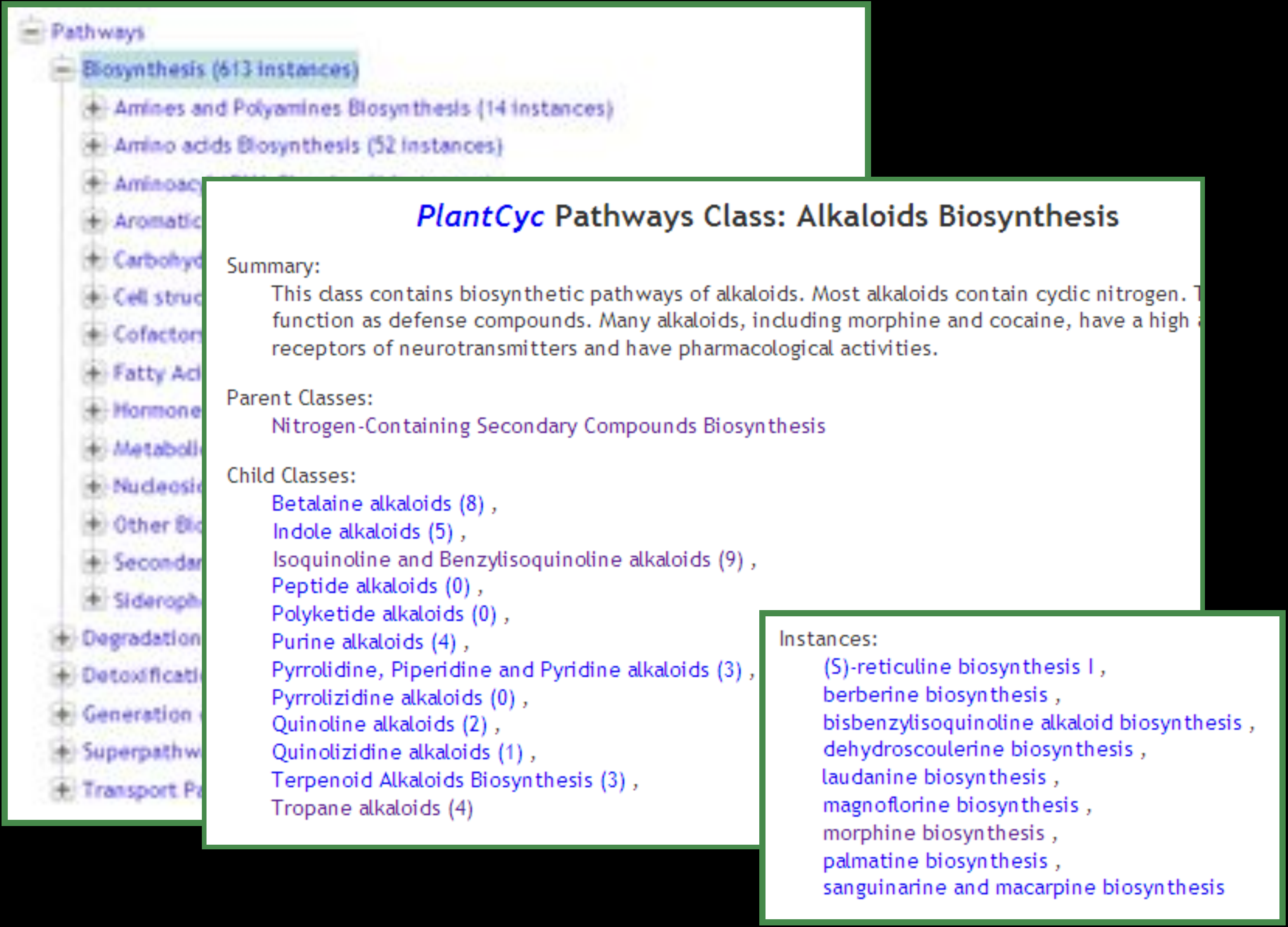
Outline

- Introduction and database snapshot
- Pathway database creation pipeline
- Manual curation
- Future work

Introduction

- Background and rationale
  - Plants (food, feed, forest, medicine, biofuel…)
  - An ocean of sequences
    - More than 60 species in genome sequencing projects, hundreds in EST projects
  - Putting individual genes onto a network of metabolic reactions and pathways
    - Annotating, visualizing and analyzing at system level
  - AraCyc (Arabidopsis thaliana, TAIR/PMN)
    - predicted by using the Pathway Tools software, followed by manual curation

Browsing Pathways



- Pathways
  - Biosynthesis (613 instances)
    - Amines and Polyamines Biosynthesis (14 instances)
    - Amino acids Biosynthesis (52 instances)
    - Aminoac
    - Aromatic
    - Carbohyd
    - Cell struc
    - Cofactor
    - Fatty Aci
    - Hormone
    - Metaboli
    - Nucleosi
    - Other Bi
    - Secondar
    - Sideroph
  - Degradation
  - Detoxificati
  - Generation
  - Superpathw
  - Transport P

PlantCyc Pathways Class: Alkaloids Biosynthesis

Summary:

This class contains biosynthetic pathways of alkaloids. Most alkaloids contain cyclic nitrogen. T function as defense compounds. Many alkaloids, including morphine and cocaine, have a high receptors of neurotransmitters and have pharmacological activities.

Parent Classes:

Nitrogen-Containing Secondary Compounds Biosynthesis

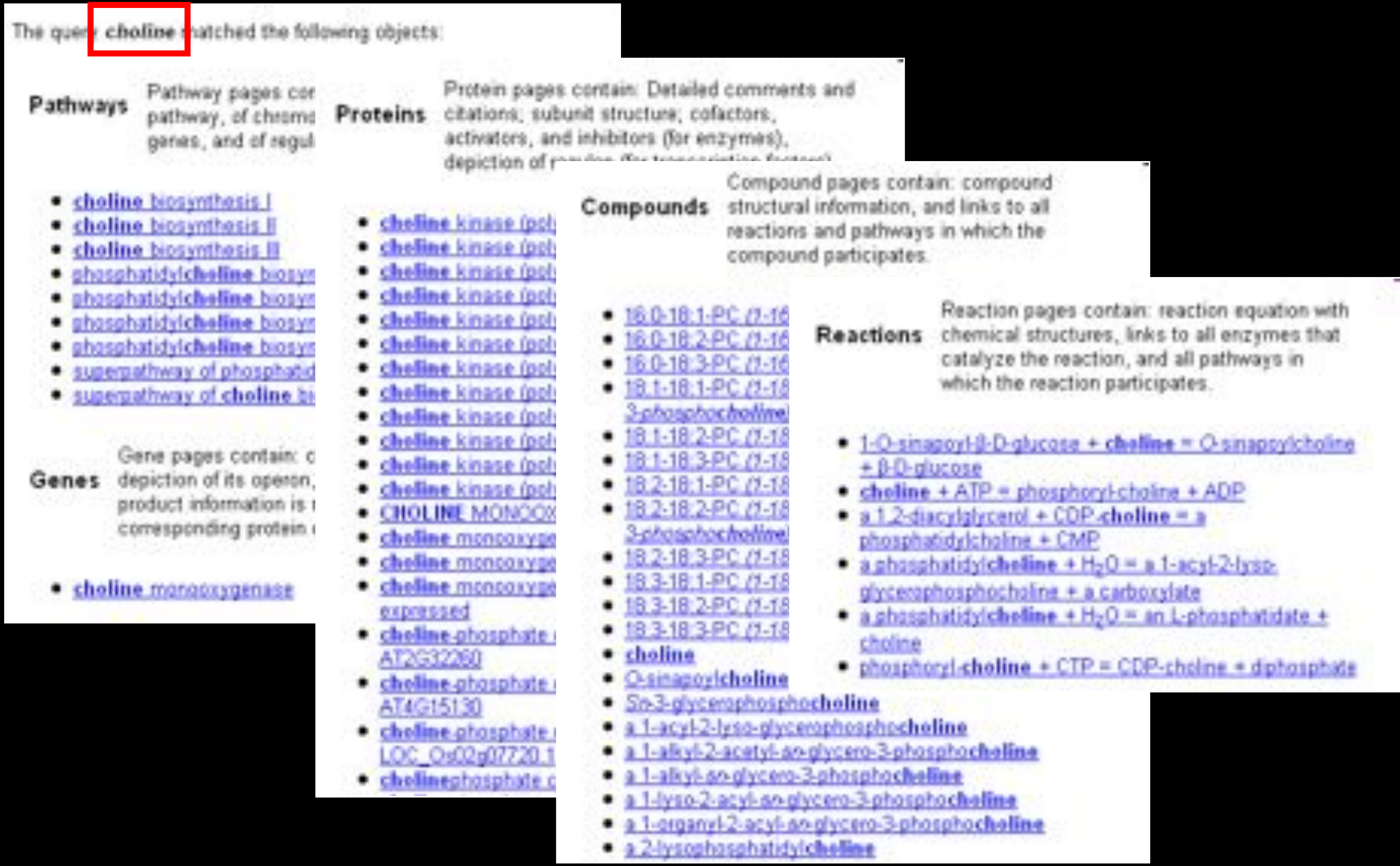
Child Classes:

Betalaine alkaloids (8) ,
Indole alkaloids (5) ,
Isoquinoline and Benzylisoquinoline alkaloids (9) ,
Peptide alkaloids (0) ,
Polyketide alkaloids (0) ,
Purine alkaloids (4) ,
Pyrrolidine, Piperidine and Pyridine alkaloids (3) ,
Pyrrolizidine alkaloids (0) ,
Quinoline alkaloids (2) ,
Quinolizidine alkaloids (1) ,
Terpenoid Alkaloids Biosynthesis (3) ,
Tropane alkaloids (4)

Instances:

(S)-reticuline biosynthesis I ,
berberine biosynthesis ,
bisbenzylisoquinoline alkaloid biosynthesis ,
dehydroscoulerine biosynthesis ,
laudanine biosynthesis ,
magnoflorine biosynthesis ,
morphine biosynthesis ,
palmatine biosynthesis ,
sanguinarine and macarpine biosynthesis

Searching Databases



The quer choline atched the following objects:

Pathways Pathway pages co pathway, of chrom genes, and of regul

- choline biosynthesis I
- choline biosynthesis II
- choline biosynthesis III
- phosphatidylcholine biosyn
- phosphatidylcholine biosyn
- phosphatidylcholine biosyn
- phosphatidylcholine biosyn
- superpathway of phosphatid
- superpathway of choline bi

Genes Gene pages contain: c depiction of its operon, product information is corresponding protein

- choline monooxygenase

Proteins Protein pages contain: Detailed comments and citations; subunit structure; cofactors, activators, and inhibitors (for enzymes), depiction of r

- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- choline kinase (pol
- CHOLINE MONOOX
- choline monooxyge
- choline monooxyge
- choline monooxyge expressed
- choline-phosphate AT2G32260
- choline-phosphate AT4G15130
- choline-phosphate LOC_Os02g07720.1
- cholinephosphate c

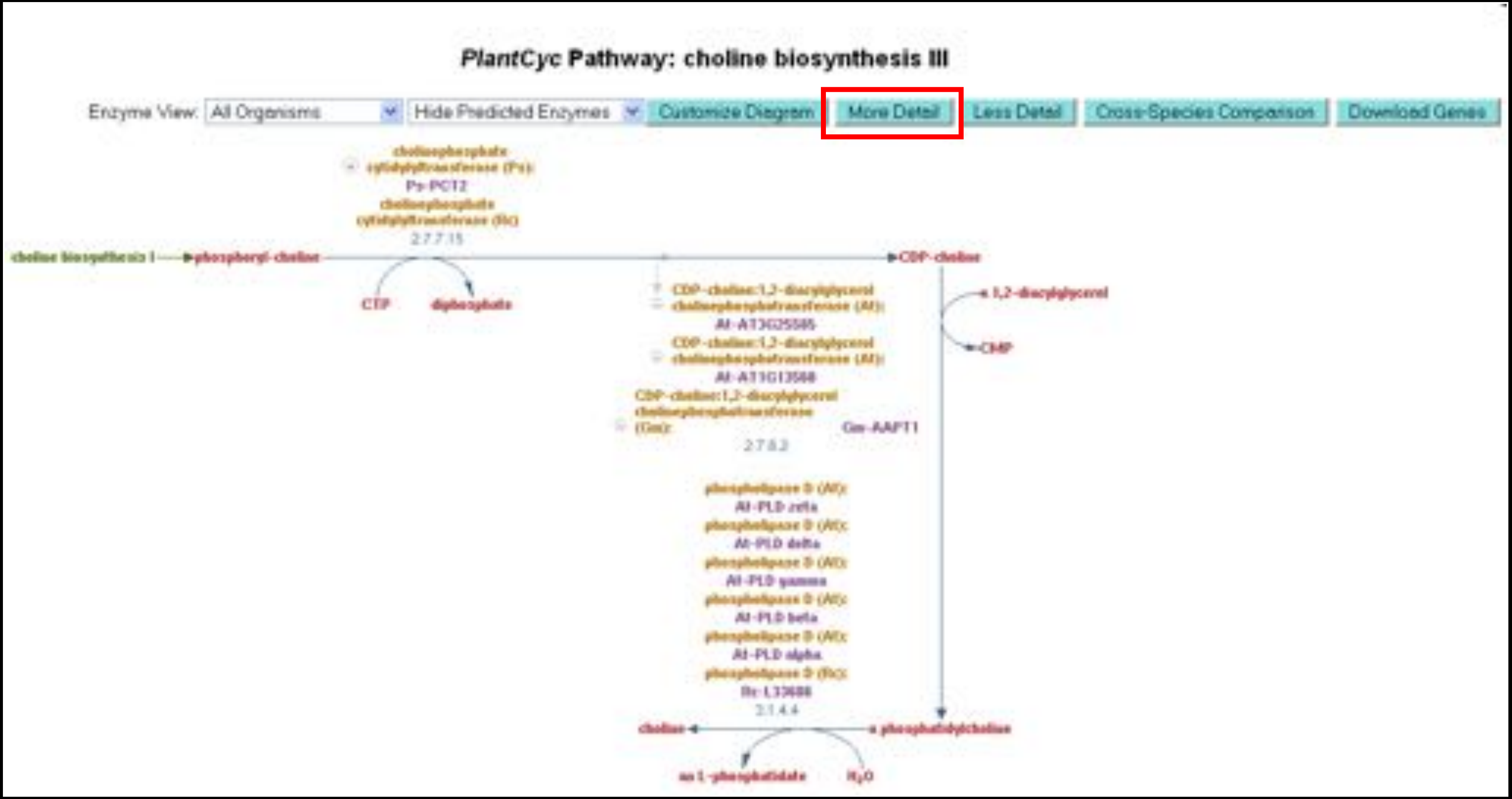
Compounds Compound pages contain: compound structural information, and links to all reactions and pathways in which the compound participates.

- 16:0-18:1-PC (1-16
- 16:0-18:2-PC (1-16
- 16:0-18:3-PC (1-16
- 18:1-18:1-PC (1-18 3-phosphocholine)
- 18:1-18:2-PC (1-18
- 18:1-18:3-PC (1-18
- 18:2-18:1-PC (1-18
- 18:2-18:2-PC (1-18 3-phosphocholine)
- 18:2-18:3-PC (1-18
- 18:3-18:1-PC (1-18
- 18:3-18:2-PC (1-18
- 18:3-18:3-PC (1-18
- choline
- O-sinapoylcholine
- Sn-3-glycerophosphocholine
- a 1-acyl-2-lyso-glycerophosphocholine
- a 1-alkyl-2-acetyl-sn-glycero-3-phosphocholine
- a 1-alkyl-sn-glycero-3-phosphocholine
- a 1-lyso-2-acyl-sn-glycero-3-phosphocholine
- a 1-organyl-2-acyl-sn-glycero-3-phosphocholine
- a 2-lysophosphatidylcholine

Reactions Reaction pages contain: reaction equation with chemical structures, links to all enzymes that catalyze the reaction, and all pathways in which the reaction participates.

- 1-O-sinapoyl-β-D-glucose + choline = O-sinapoylcholine + β-D-glucose
- choline + ATP = phosphoryl-choline + ADP
- a 1,2-diacylglycerol + CDP-choline = a phosphatidylcholine + CMP
- a phosphatidylcholine + H_2O = a 1-acyl-2-lyso-glycerophosphocholine + a carboxylate
- a phosphatidylcholine + H_2O = an L-phosphatidate + choline
- phosphoryl-choline + CTP = CDP-choline + diphosphate

A Typical Pathway Detail Page

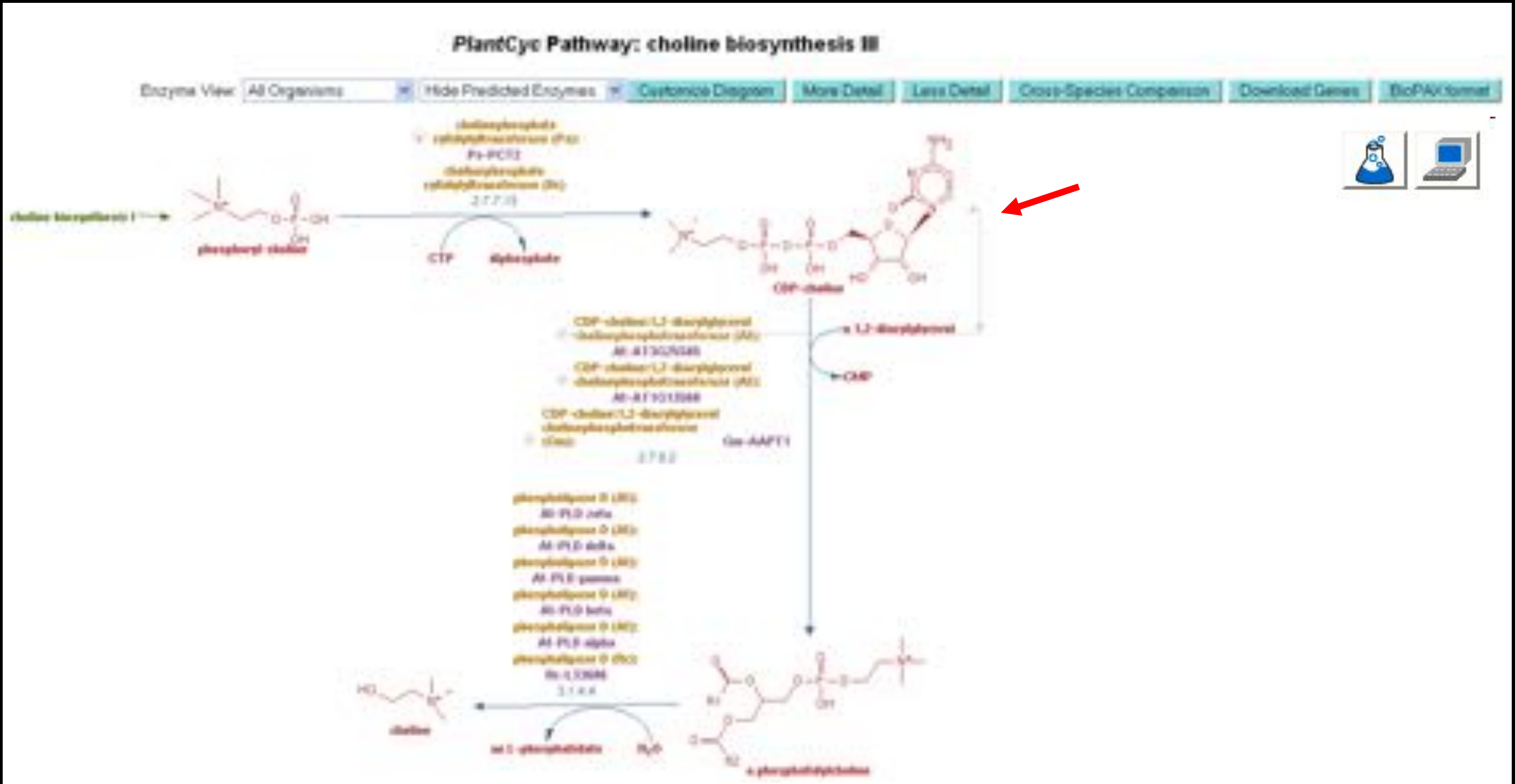


PlantCyc Pathway: choline biosynthesis III

Enzyme View: All Organisms

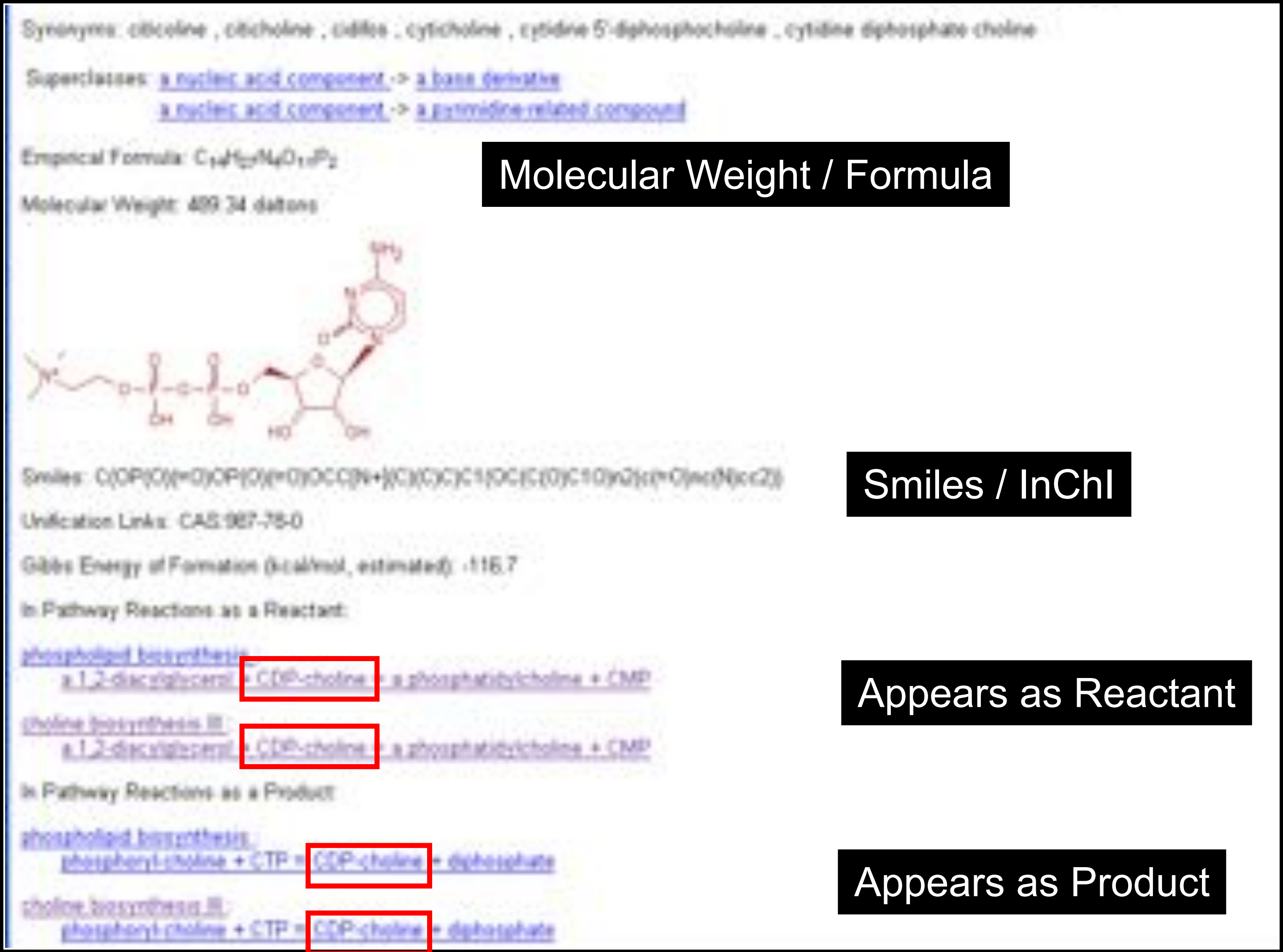
Hide Predicted Enzymes

Linking to Other Data Detail Pages



PlantCyc Pathway: choline biosynthesis III

Compound Detail Pages



Synonyms: citicoline , citicholine , cidifos , cyticholine , cytidine 5'-diphosphocholine , cytidine diphosphate choline

Superclasses: a nucleic acid component -> a base derivative
a nucleic acid component -> a pyrimidine-related compound

Empirical Formula: $C_{14}H_{26}N_4O_{11}P_2$

Molecular Weight: 489.34 daltons

Smiles: C(OP(O)(=O)OP(O)(=O)OCC[N+](C)(C)C)C1(OC(C(O)C1O)n2)c(=O)nc(N)cc2)

Unification Links: CAS:987-78-0

Gibbs Energy of Formation (kcal/mol, estimated): -116.7

In Pathway Reactions as a Reactant:

phospholipid biosynthesis :
a 1,2-diacylglycerol + CDP-choline = a phosphatidylcholine + CMP

choline biosynthesis III :
a 1,2-diacylglycerol + CDP-choline = a phosphatidylcholine + CMP

In Pathway Reactions as a Product:

phospholipid biosynthesis :
phosphoryl-choline + CTP = CDP-choline + diphosphate

choline biosynthesis III :
phosphoryl-choline + CTP = CDP-choline + diphosphate

Molecular Weight / Formula

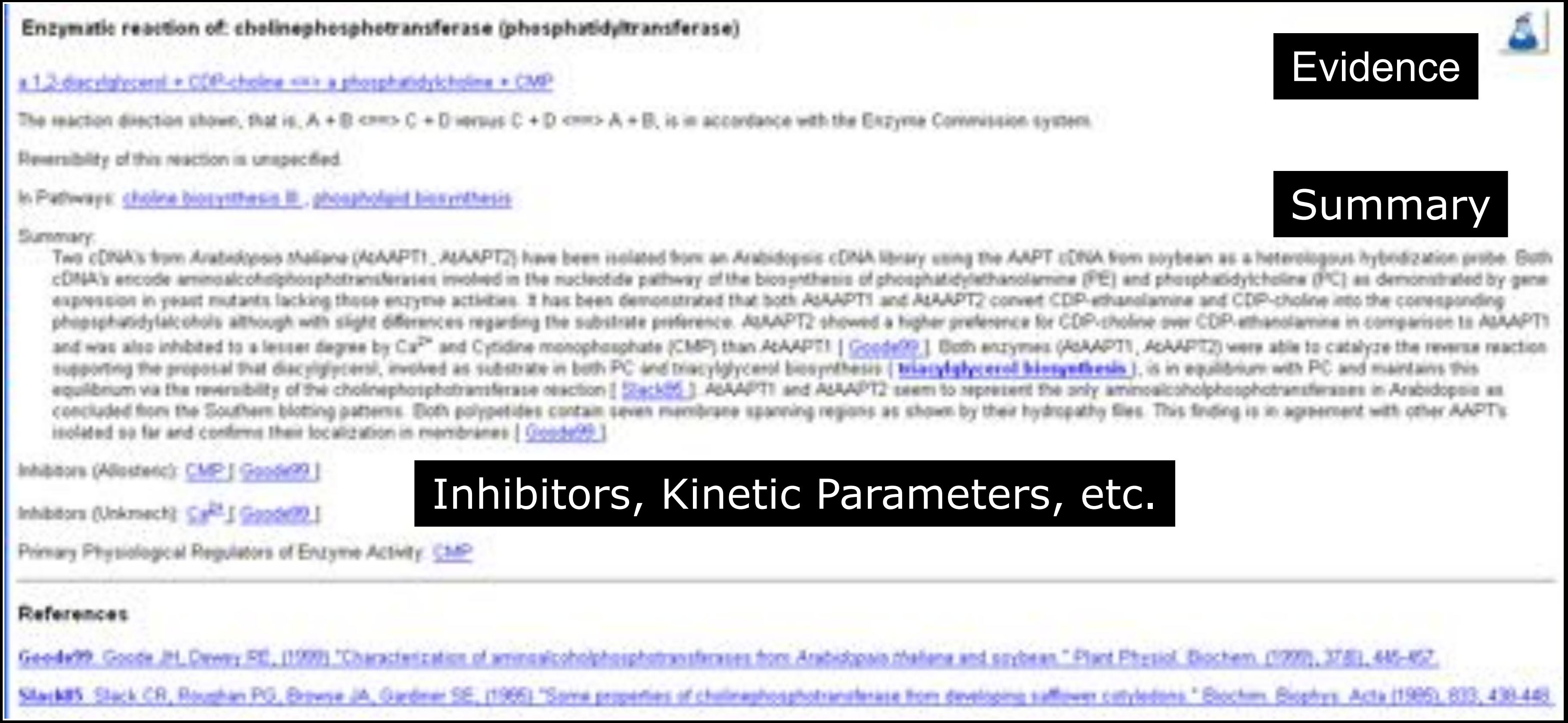
Smiles / InChI

Appears as Reactant

Appears as Product

Enzyme Detail Pages

Arabidopsis **Enzyme: phosphatidyltransferase**



Enzymatic reaction of: cholinephosphotransferase (phosphatidyltransferase)

a 1,2-diacylglycerol + CDP-choline <==> a phosphatidylcholine + CMP

The reaction direction shown, that is, A + B <==> C + D versus C + D <==> A + B, is in accordance with the Enzyme Commission system.

Reversibility of this reaction is unspecified.

In Pathways: choline biosynthesis III , phospholipid biosynthesis

Summary:

Two cDNA's from *Arabidopsis thaliana* (AtAAPT1 , AtAAPT2) have been isolated from an Arabidopsis cDNA library using the AAPT cDNA from soybean as a heterologous hybridization probe. Both cDNA's encode aminoalcoholphosphotransferases involved in the nucleotide pathway of the biosynthesis of phosphatidylethanolamine (PE) and phosphatidylcholine (PC) as demonstrated by gene expression in yeast mutants lacking those enzyme activities. It has been demonstrated that both AtAAPT1 and AtAAPT2 convert CDP-ethanolamine and CDP-choline into the corresponding phosphatidylalcohols although with slight differences regarding the substrate preference. AtAAPT2 showed a higher preference for CDP-choline over CDP-ethanolamine in comparison to AtAAPT1 and was also inhibited to a lesser degree by Ca^{2+} and Cytidine monophosphate (CMP) than AtAAPT1 [Goode99]. Both enzymes (AtAAPT1 , AtAAPT2) were able to catalyze the reverse reaction supporting the proposal that diacylglycerol, involved as substrate in both PC and triacylglycerol biosynthesis (**triacylglycerol biosynthesis**), is in equilibrium with PC and maintains this equilibrium via the reversibility of the cholinephosphotransferase reaction [Slack85]. AtAAPT1 and AtAAPT2 seem to represent the only aminoalcoholphosphotransferases in Arabidopsis as concluded from the Southern blotting patterns. Both polypeptides contain seven membrane spanning regions as shown by their hydropathy files. This finding is in agreement with other AAPT's isolated so far and confirms their localization in membranes [Goode99]

Inhibitors (Allosteric): CMP [Goode99]

Inhibitors (Unkmech): Ca^{2+} [Goode99]

Primary Physiological Regulators of Enzyme Activity: CMP

References

Goode99: Goode JH, Dewey RE, (1999) "Characterization of aminoalcoholphosphotransferases from Arabidopsis thaliana and soybean." Plant Physiol. Biochem. (1999), 37(6), 445-457.

Slack85: Slack CR, Roughan PG, Browse JA, Gardiner SE, (1985) "Some properties of cholinephosphotransferase from developing safflower cotyledons." Biochim. Biophys. Acta (1985), 833, 438-448.

Evidence

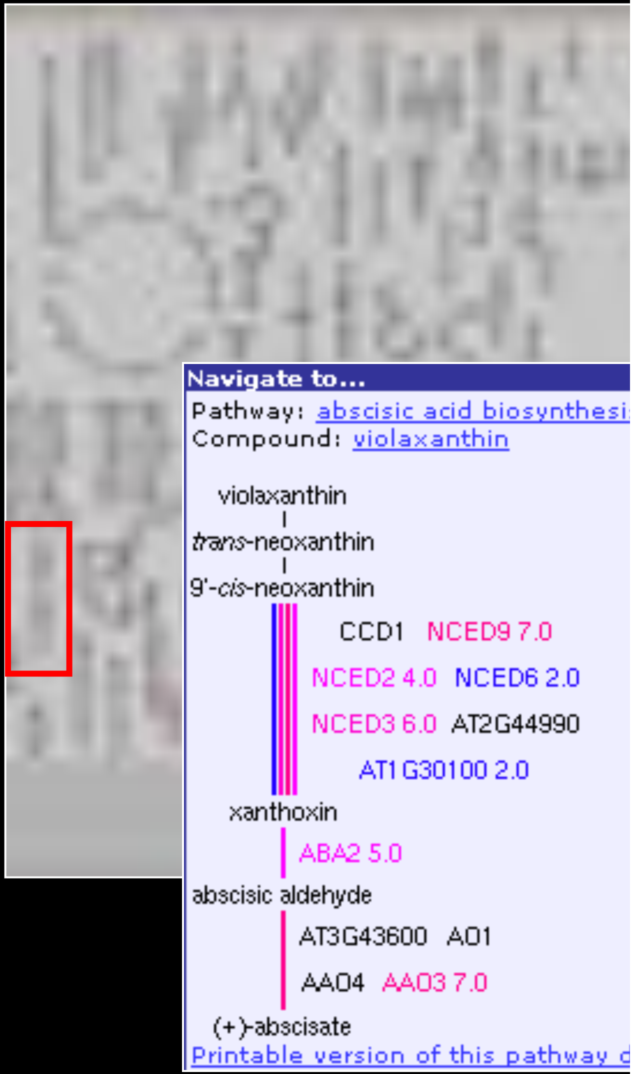
Summary

Inhibitors, Kinetic Parameters, etc.

Visualizing and Interpreting Omics Data in a Metabolic Context

- Gene expression data
- Proteomic data
- Metabolic profiling data
- Reaction flux data

Omics Viewer



Navigate to...

Pathway: abscisic acid biosynthesi

Compound: violaxanthin

violaxanthin

trans-neoxanthin

9'-*cis*-neoxanthin

CCD1 NCED9 7.0

NCED2 4.0 NCED6 2.0

NCED3 6.0 AT2G44990

AT1G30100 2.0

xanthoxin

ABA2 5.0

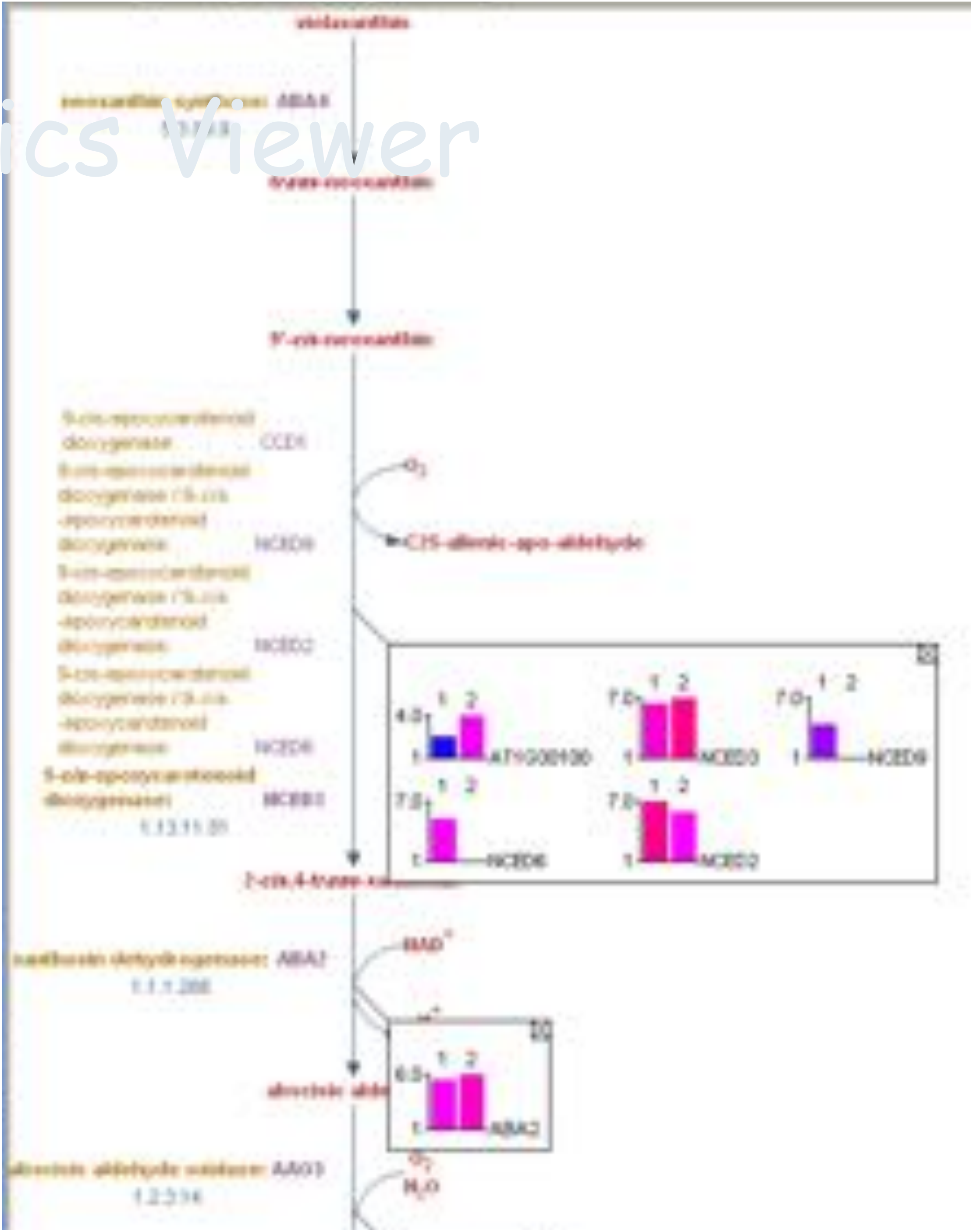
abscisic aldehyde

AT3G43600 AO1

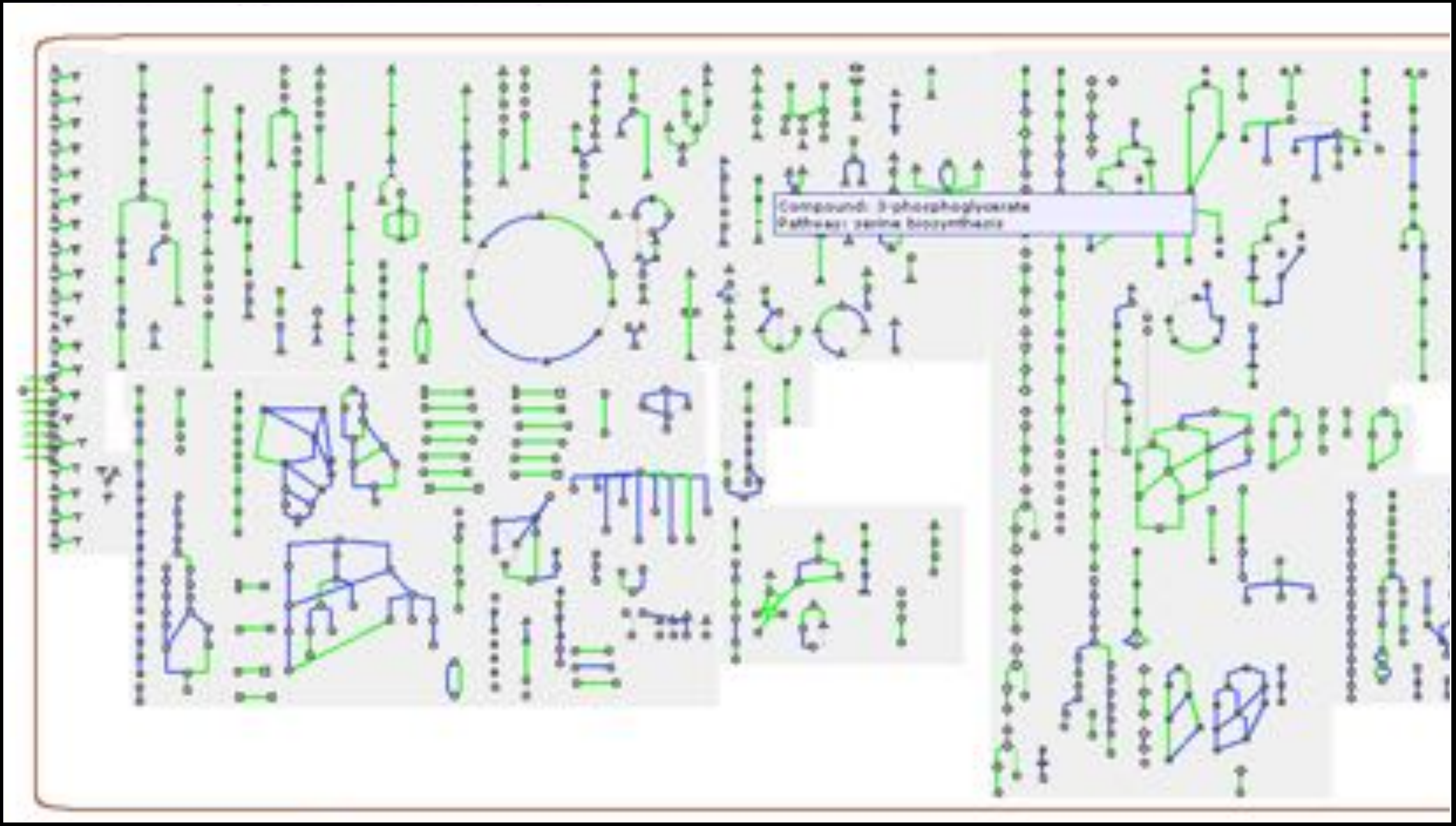
AAO4 AAO3 7.0

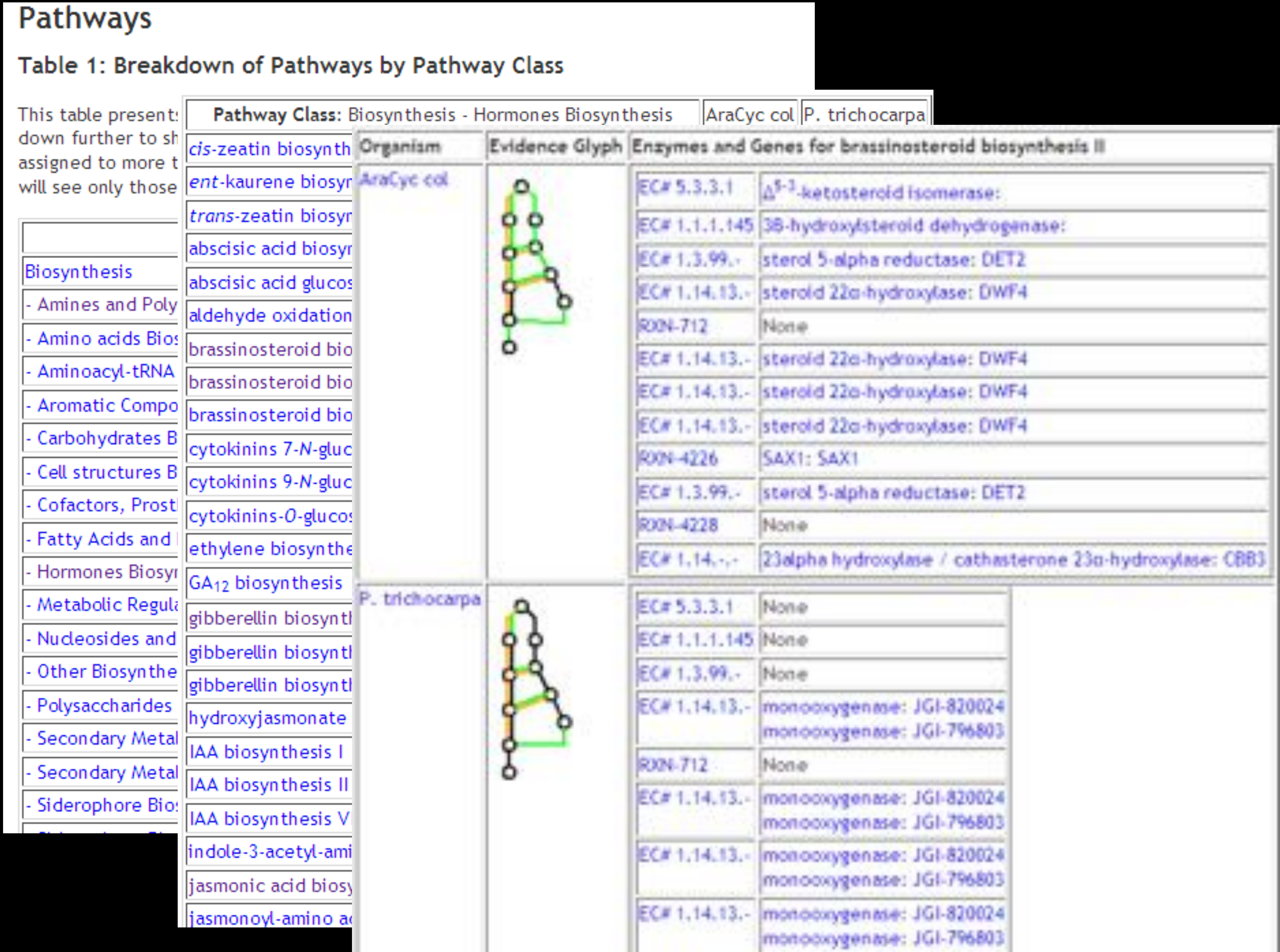
(+)-abscisate

Printable version of this pathway d



Comparing Across Species





Pathways

Table 1: Breakdown of Pathways by Pathway Class

This table present
down further to sh
assigned to more t
will see only those

Biosynthesis
- Amines and Poly
- Amino acids Bios
- Aminoacyl-tRNA
- Aromatic Compo
- Carbohydrates B
- Cell structures B
- Cofactors, Prost
- Fatty Acids and
- Hormones Biosy
- Metabolic Regul
- Nucleosides and
- Other Biosynthe
- Polysaccharides
- Secondary Metal
- Secondary Metal
- Siderophore Bio

Pathway Class: Biosynthesis - Hormones Biosynthesis	AraCyc col	P. trichocarpa
cis-zeatin biosynth		
ent-kaurene biosyn		
trans-zeatin biosyn		
abscisic acid biosyn		
abscisic acid glucos		
aldehyde oxidation		
brassinosteroid bio		
brassinosteroid bio		
brassinosteroid bio		
cytokinins 7-*N*-gluc		
cytokinins 9-*N*-gluc		
cytokinins-*O*-glucos		
ethylene biosynthe		
GA_{12} biosynthesis		
gibberellin biosynth		
gibberellin biosynth		
gibberellin biosynth		
hydroxyjasmonate		
IAA biosynthesis I		
IAA biosynthesis II		
IAA biosynthesis V		
indole-3-acetyl-ami		
jasmonic acid biosy		
jasmonoyl-amino a		

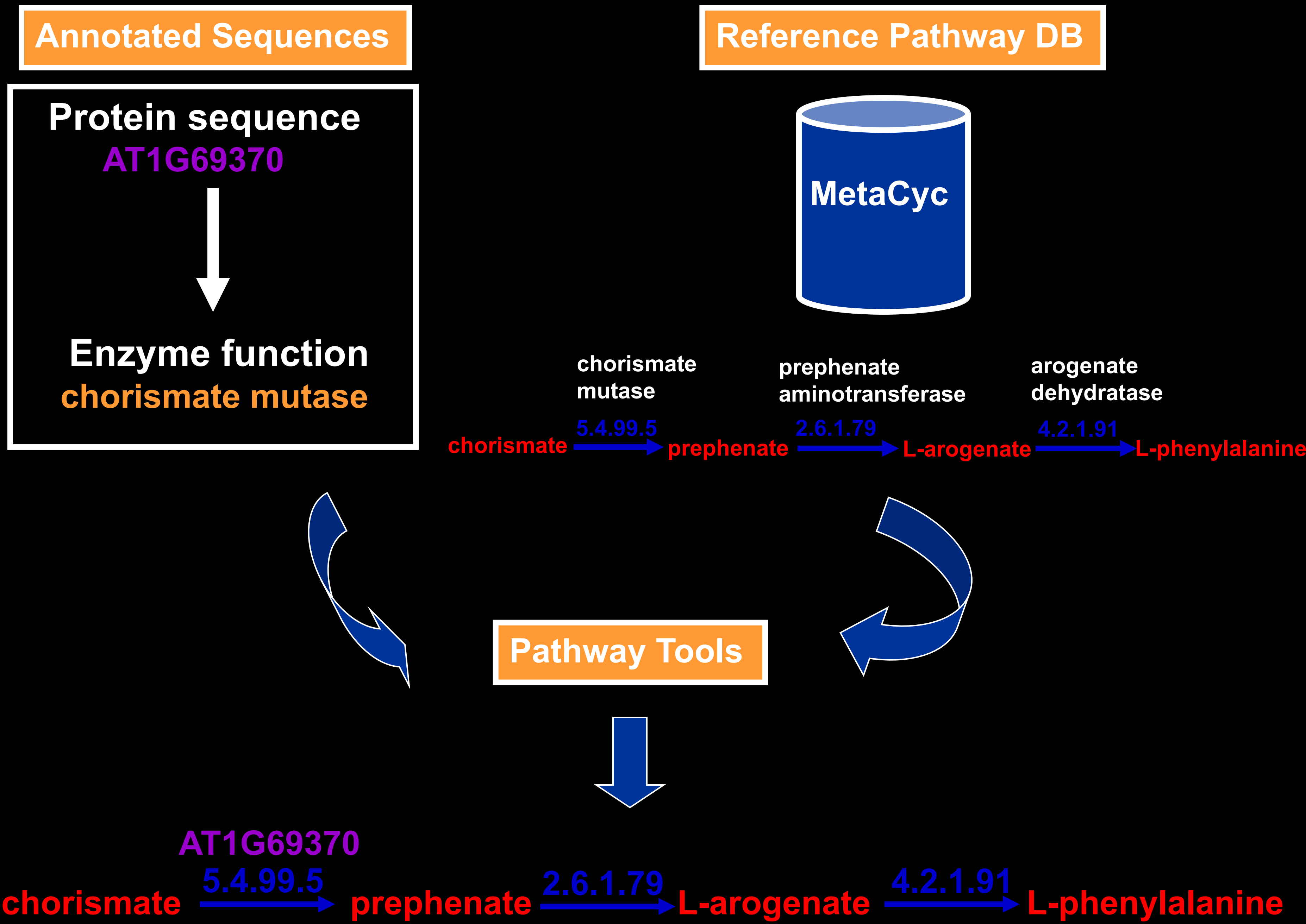
Organism	Evidence Glyph	Enzymes and Genes for brassinosteroid biosynthesis II	
AraCyc col		EC# 5.3.3.1	Δ^{5-3}-ketosteroid isomerase:
		EC# 1.1.1.145	3ß-hydroxylsteroid dehydrogenase:
		EC# 1.3.99.-	sterol 5-alpha reductase: DET2
		EC# 1.14.13.-	steroid 22α-hydroxylase: DWF4
		RXN-712	None
		EC# 1.14.13.-	steroid 22α-hydroxylase: DWF4
		EC# 1.14.13.-	steroid 22α-hydroxylase: DWF4
		EC# 1.14.13.-	steroid 22α-hydroxylase: DWF4
		RXN-4226	SAX1: SAX1
		EC# 1.3.99.-	sterol 5-alpha reductase: DET2
		RXN-4228	None
		EC# 1.14.-.-	23alpha hydroxylase / cathasterone 23α-hydroxylase: CBB3
P. trichocarpa		EC# 5.3.3.1	None
		EC# 1.1.1.145	None
		EC# 1.3.99.-	None
		EC# 1.14.13.-	monooxygenase: JGI-820024 monooxygenase: JGI-796803
		RXN-712	None
		EC# 1.14.13.-	monooxygenase: JGI-820024 monooxygenase: JGI-796803
		EC# 1.14.13.-	monooxygenase: JGI-820024 monooxygenase: JGI-796803
		EC# 1.14.13.-	monooxygenase: JGI-820024 monooxygenase: JGI-796803

Introduction (cont)

- Background and rationale
  - Plants (food, feed, forest, medicine, biofuel…)
  - An ocean of sequences
    - More than 60 species in genome sequencing projects, hundreds in EST projects
  - Putting individual genes onto a network of metabolic reactions and pathways
    - Annotating, visualizing and analyzing at system level
  - AraCyc (Arabidopsis thaliana, TAIR/PMN)
    - predicted by using the Pathway Tools software, followed by manual curation
  - Other plant pathway databases predicted by using the Pathway Tools
    - RiceCyc (Oryza sativa, Gramene)
    - MedicCyc (Medicago truncatula, Noble Foundation)
    - LycoCyc (Solanum lycopersicum, SGN), …

Pathway Prediction and Pathway Database Creation

- Infer the reactome of an organism from the enzymes present in its annotated genome
  - Mapping annotated enzyme sequences to reactions
- Infer the metabolic pathways of the organism from its reactome
  - Pathway-calling based on supporting evidence of reactions



Annotated Sequences

Protein sequence
AT1G69370

Enzyme function
chorismate mutase

Reference Pathway DB

MetaCyc

chorismate mutase
prephenate aminotransferase
arogenate dehydratase

chorismate $\xrightarrow{5.4.99.5}$ prephenate $\xrightarrow{2.6.1.79}$ L-arogenate $\xrightarrow{4.2.1.91}$ L-phenylalanine

Pathway Tools

AT1G69370

chorismate $\xrightarrow{5.4.99.5}$ prephenate $\xrightarrow{2.6.1.79}$ L-arogenate $\xrightarrow{4.2.1.91}$ L-phenylalanine

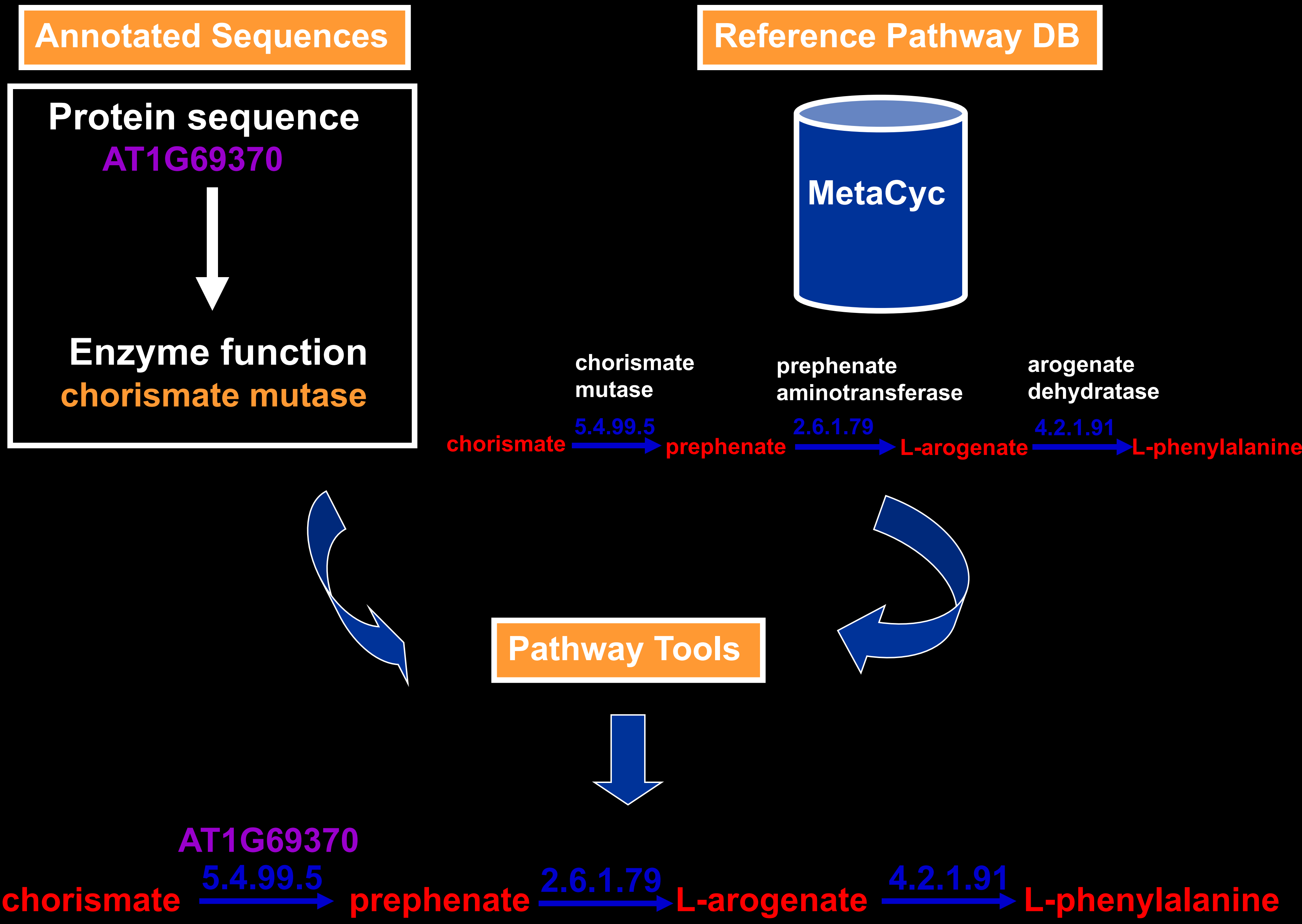
Limitations

- Creating pathway databases includes three major components, and is resource-intensive
  - Sequence annotation
  - Reference pathway database
  - Pathway prediction, validation, refinement
- Heterogeneous sequence annotation protocols and varying levels of pathway validation impact quality and hinder meaningful cross-species comparison
  - Using a non-plant reference database causes many false-positive and false-negative pathway predictions

Introducing the PMN

- Scope
  - A platform for plant metabolic pathway database creation
  - A community for data curation
    - Curators, editorial board, ally in other databases, researchers

- Major goals
  - Create a plant-specific reference pathway database (PlantCyc)
  - Create an enzyme sequence annotation pipeline
  - Enhance pathway prediction by using PlantCyc, and including an automated initial validation step
  - Create metabolic pathway databases for plant species
    - e.g. PoplarCyc (Populus trichocarpa), SoyCyc (soybean)



Annotated Sequences

Protein sequence

AT1G69370

Enzyme function

chorismate mutase

Reference Pathway DB

MetaCyc

chorismate mutase — prephenate aminotransferase — arogenate dehydratase

chorismate $\xrightarrow{5.4.99.5}$ prephenate $\xrightarrow{2.6.1.79}$ L-arogenate $\xrightarrow{4.2.1.91}$ L-phenylalanine

Pathway Tools

AT1G69370

chorismate $\xrightarrow{5.4.99.5}$ prephenate $\xrightarrow{2.6.1.79}$ L-arogenate $\xrightarrow{4.2.1.91}$ L-phenylalanine

PlantCyc Creation

- Nature
  - Multiple-species, plants-only, curator-reviewed pathways, primary and secondary metabolism

- Major Source
  - All AraCyc pathways and enzymes
  - Plant pathways and enzymes from MetaCyc
  - Additional pathways and enzymes manually curated and added
  - Enzymes from RiceCyc, LycoCyc and MedicCyc

PMN Database Content Statistics

	PlantCyc 4.0	AraCyc 7.0	PoplarCyc 2.0
Pathways	685	369	288
Enzymes	11058	5506	3420
Reactions	2929	2418	1707
Compounds	2966	2719	1397
Organisms	343	1	1*

Valuable plant natural products, many are specialized metabolites that are limited to a few species or genus.

- medicinal: e.g. artemisinin and quinine (treatment of malaria),
  codeine and morphine (pain-killer),
  ginsenosides (cardio-protectant),
  lupenol (antiinflammatory),
  taxol and vinblastine (anti-cancer)
- industrial materials: e.g. resin and rubber
- food flavor and scents: e.g. capsaicin and piperine (chili and pepper flavor), geranyl acetate (aroma of rose) and menthol (mint).

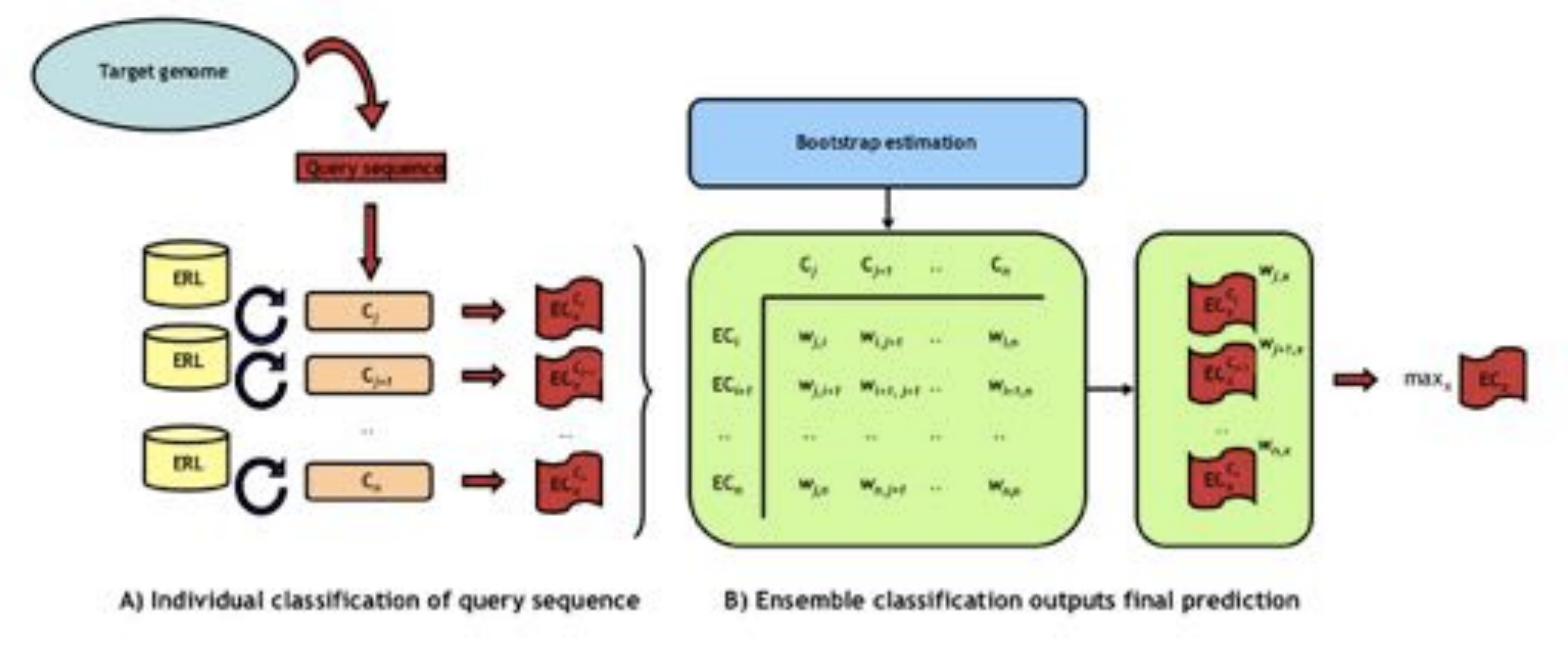
Enzyme Sequence Annotation (version 1.0)

- Reference sequences, enzymes with known functions
  - 14,187 enzyme sequences compiled from UniProt, Brenda, MetaCyc, and TAIR
  - 3805 functional identifiers (full EC number, MetaCyc reaction id, GO id)
- Annotation methods
  - BLASTP
- Cut-off
  - unique e-value threshold for each functional identifier

Enzyme Sequence Annotation (version 2.0)

- Reference sequences, proteins with known functions (ERL)
  - SwissProt
    - 117,000 proteins, 26,000 enzymes, 2,400 full EC numbers
  - Additional enzymes from MetaCyc, TAIR, Brenda and UniProt
  - Functional identifiers, full EC number, MetaCyc reaction id, GO id,
- Annotation methods
  - BLASTP
  - Priam (enzyme-specific, motif-based)
  - CatFam (enzyme-specific, motif-based)
- Function calling
  - Ensemble and voting

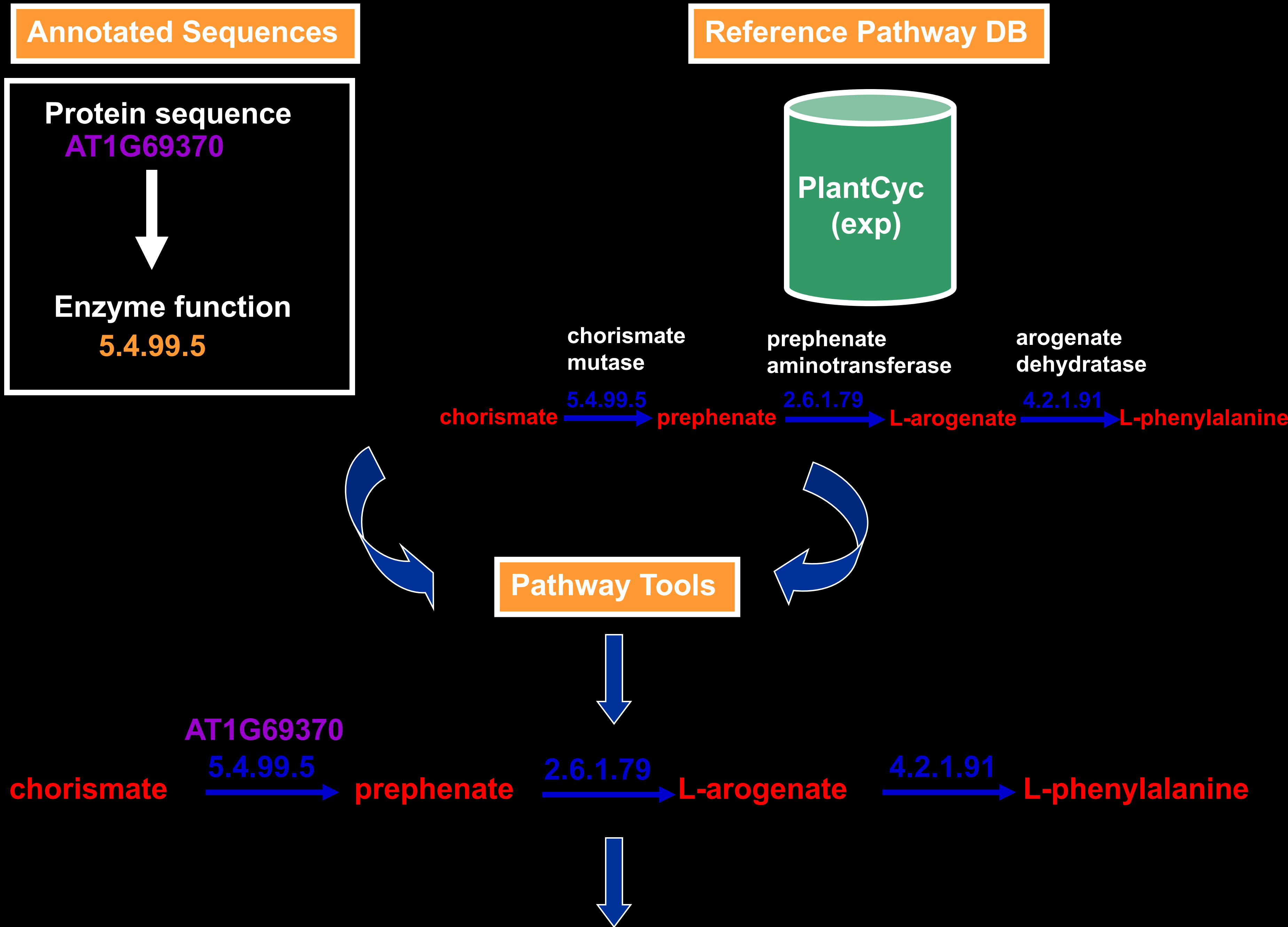
Enzyme Sequence Annotation (version 2.0)



A) Individual classification of query sequence

B) Ensemble classification outputs final prediction

Lee Chae (unpublished)



Annotated Sequences

Protein sequence
AT1G69370

Enzyme function
5.4.99.5

Reference Pathway DB

PlantCyc
(exp)

chorismate mutase

prephenate aminotransferase

arogenate dehydratase

chorismate $\xrightarrow{5.4.99.5}$ prephenate $\xrightarrow{2.6.1.79}$ L-arogenate $\xrightarrow{4.2.1.91}$ L-phenylalanine

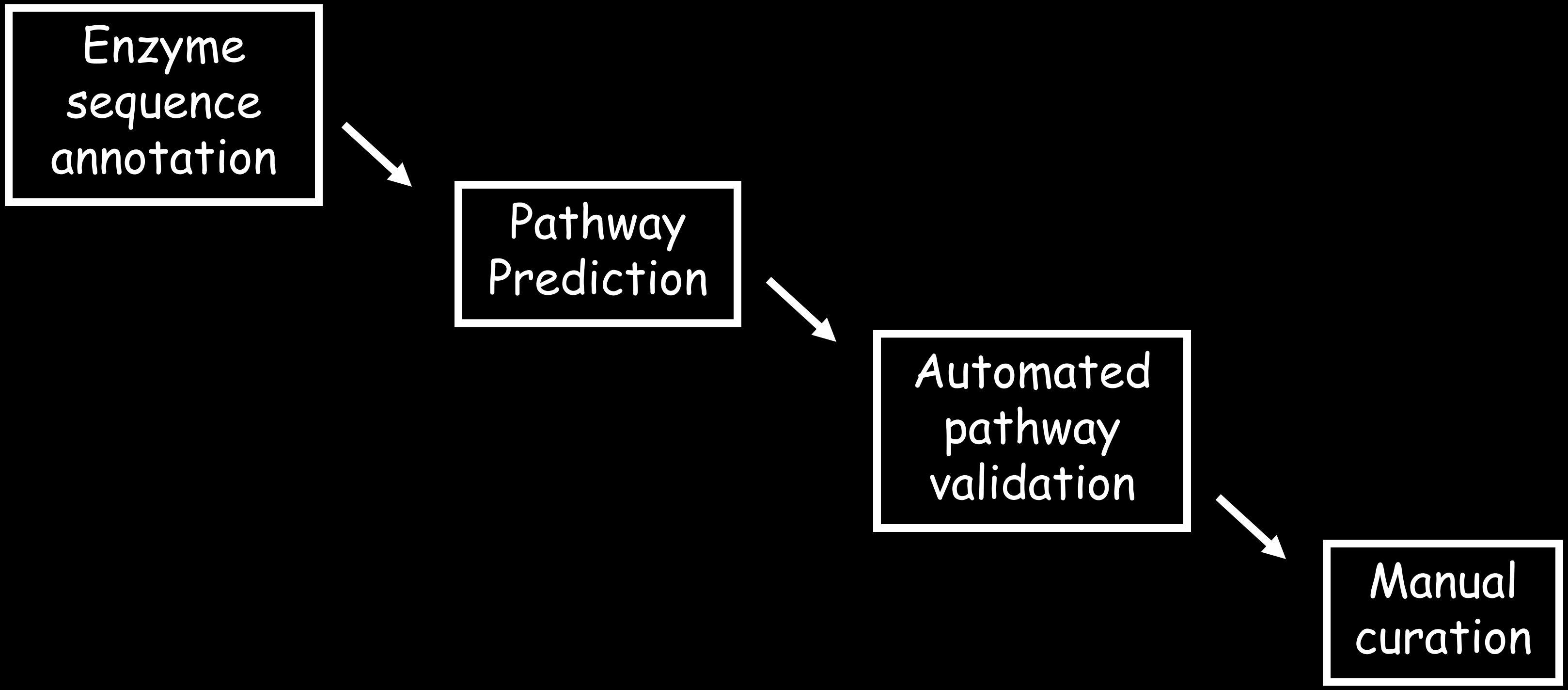
Pathway Tools

AT1G69370

chorismate $\xrightarrow{5.4.99.5}$ prephenate $\xrightarrow{2.6.1.79}$ L-arogenate $\xrightarrow{4.2.1.91}$ L-phenylalanine

Automated Initial Pathway Validation

- Remove non-plant pathways
  - A list of 132 MetaCyc pathways
- Add universal plant pathways
  - A list of 115 PlantCyc pathways



Enzyme sequence annotation

Pathway Prediction

Automated pathway validation

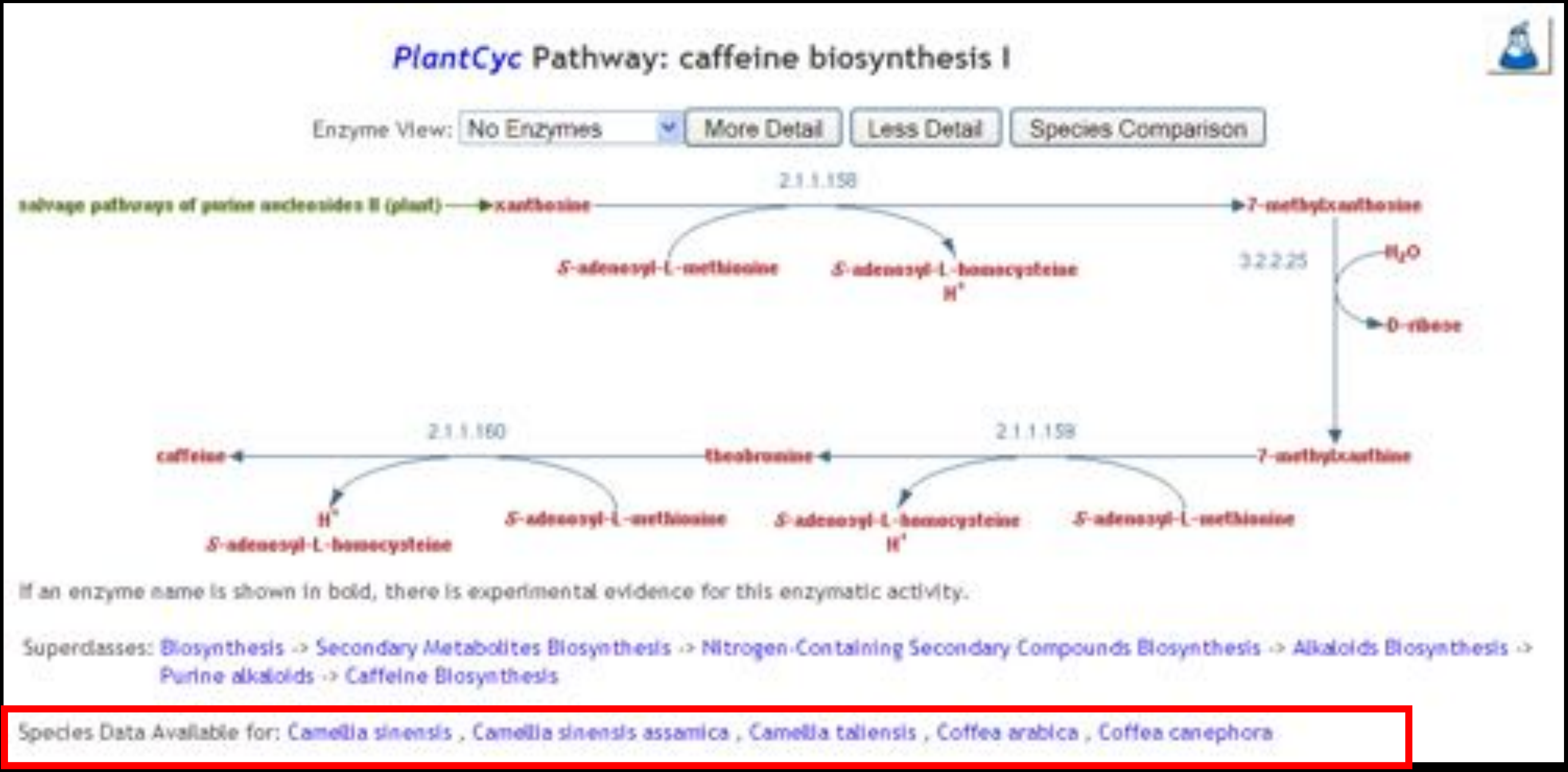
Manual curation

Manual Curation

- Who
  - Curators identify, read and enter information from published journal articles
- What
  - Remove false-positive pathway predictions
  - Remove false-positive enzyme annotations
  - Add missing pathways (pathway diagrams)
  - Add missing enzymes
  - Curate enzyme properties, kinetic data
  - Update existing pathways (pathway diagrams)
  - Add new reactions
  - Add new compounds and curate compound structures

Conventions Used in Curation and Data Presentation

- A pathway, as drawn in the text books, is a functional unit, regulated as a unit
- Pathway displayed is expected to operate as such in the individual species listed



PlantCyc Pathway: caffeine biosynthesis I

Enzyme View: No Enzymes

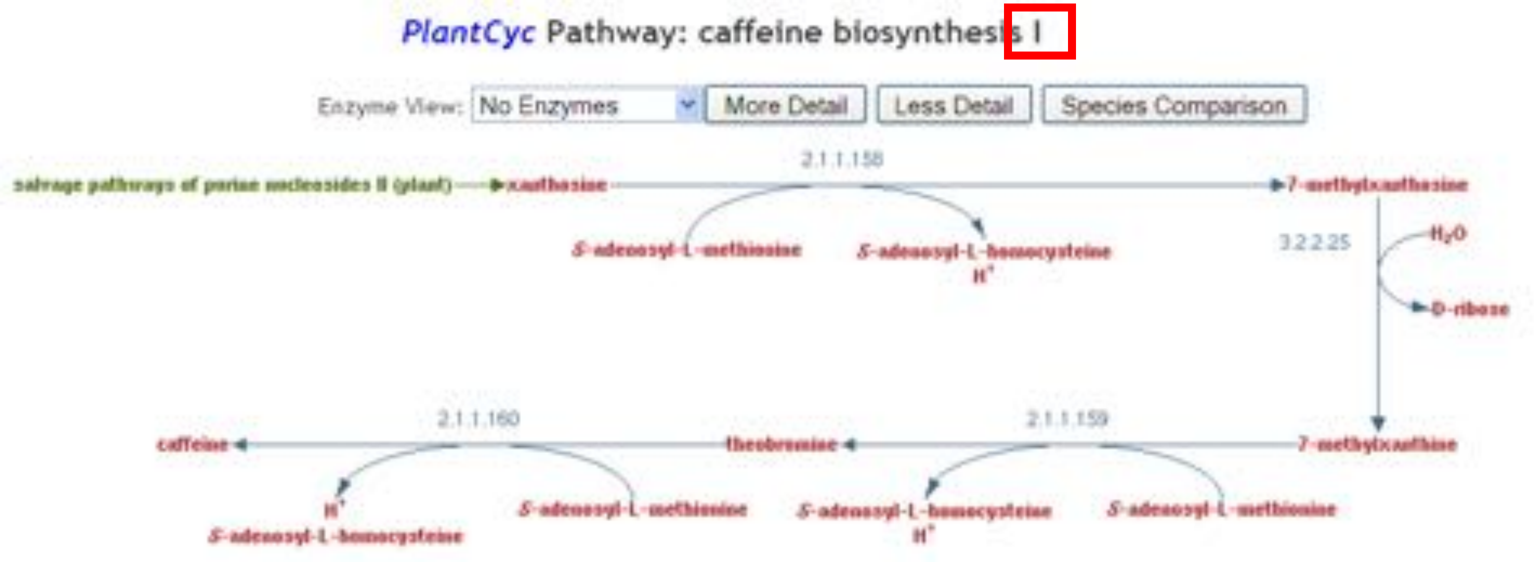
If an enzyme name is shown in bold, there is experimental evidence for this enzymatic activity.

Superclasses: Biosynthesis -> Secondary Metabolites Biosynthesis -> Nitrogen-Containing Secondary Compounds Biosynthesis -> Alkaloids Biosynthesis -> Purine alkaloids -> Caffeine Biosynthesis

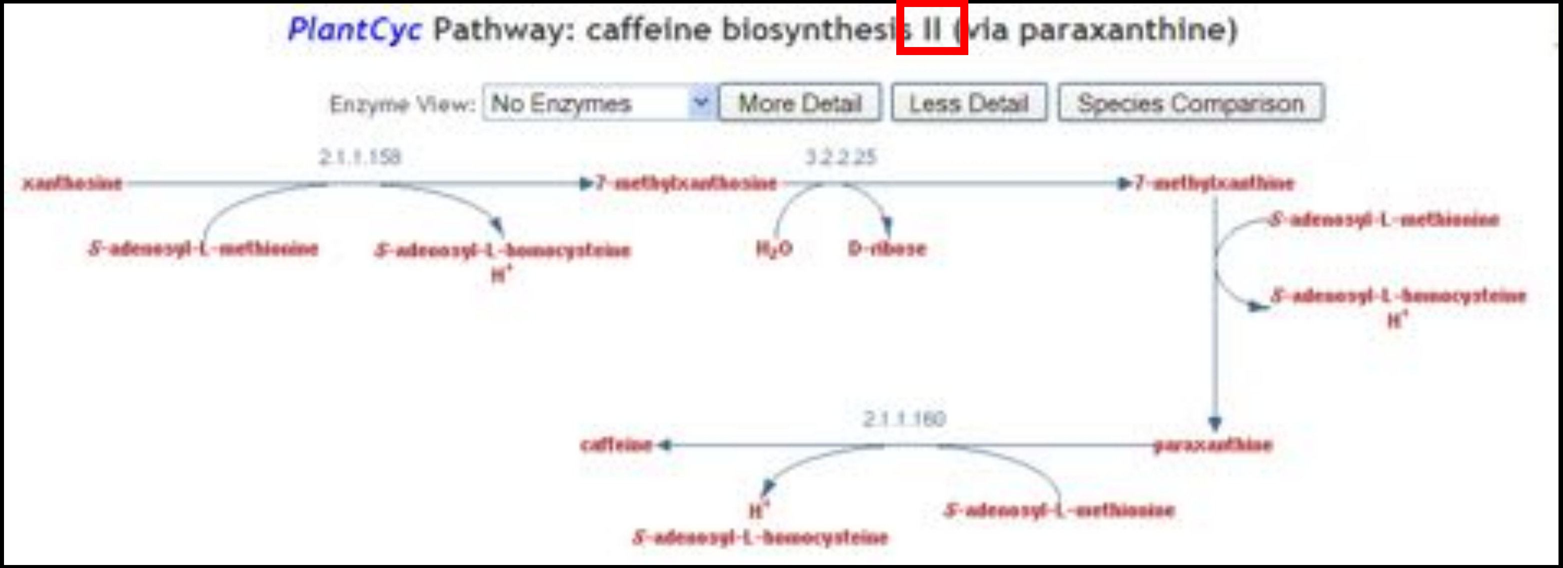
Species Data Available for: Camellia sinensis , Camellia sinensis assamica , Camellia taliensis , Coffea arabica , Coffea canephora

Conventions Used in Curation and Data Presentation

- Pathway, as drawn in the text books, is a functional unit, regulated as a unit
- Pathway displayed is expected to operate as such in the individual species shown
- Alternative routes that have been observed in different organisms are curated separately as pathway <u>variants</u>



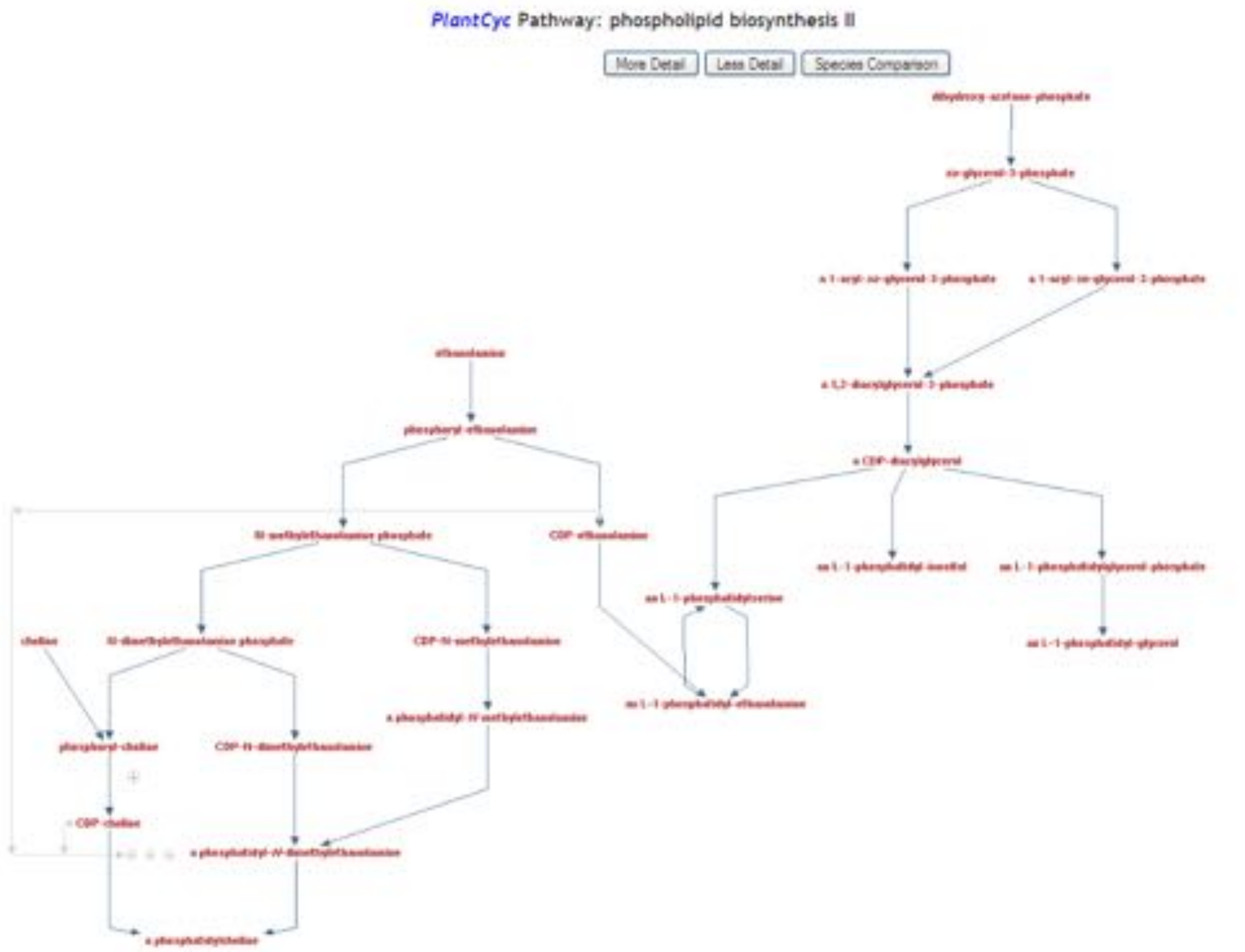
PlantCyc Pathway: caffeine biosynthesis I



PlantCyc Pathway: caffeine biosynthesis II (via paraxanthine)

Conventions Used in Curation and Data Presentation

- Pathway, as drawn in the text books, is a functional unit, regulated as a unit
- Pathway displayed is expected to operate as such in the individual species shown
- Alternative routes that have been observed in different organisms are curated separately as pathway variants
- Mosaics combined of alternative routes from several different species are curated as superpathways
- Connected pathways, extended networks, are curated as superpathways



PlantCyc Pathway: phospholipid biosynthesis II

More Detail
Less Detail
Species Comparison
dihydroxy-acetone-phosphate
sn-glycerol-3-phosphate
a 1-acyl-sn-glycerol-3-phosphate
a 1-acyl-sn-glycerol-2-phosphate
a 1,2-diacylglycerol-3-phosphate
a CDP-diacylglycerol
ethanolamine
phosphoryl-ethanolamine
N-methylethanolamine phosphate
CDP-ethanolamine
an L-1-phosphatidyl-inositol
an L-1-phosphatidylglycerol-phosphate
an L-1-phosphatidylserine
choline
N-dimethylethanolamine phosphate
CDP-N-methylethanolamine
an L-1-phosphatidyl-glycerol
an L-1-phosphatidyl-ethanolamine
a phosphatidyl-N-methylethanolamine
phosphoryl-choline
CDP-N-dimethylethanolamine
CDP-choline
a phosphatidyl-N-dimethylethanolamine
a phosphatidylcholine

Future Work

- Enhance pathway prediction and validation
  - Using additional evidence, such as presence of compounds, weighted confidence of enzyme annotations
- Refine pathways, hole-filling
  - Including non-sequence homology based information in enzyme function prediction, such as phylogenetic profiles, co-expression
- Add new data types, critical for strategic planning of metabolic engineering
  - Rate-limiting step
  - Transcriptional regulator
- Create new pathway databases
  - moss (P. patens), Selaginella, maize, cassava, wine grape …

Thank you!