Metabolic Pathway Databases and Tools

Speaker and Schedule Update

PMN (Peifen Zhang)
KEGG (auto-slide show)
MetaCrop (cancelled)

and Its Databases

Peifen Zhang
Carnegie Institution For Science
Department of Plant Biology

Where We Are


Canada
Alberta
Manitoba
British Columbia
Saskatchewan
Newfoundland and Labrador
Ontario
Quebec
Gulf of St Lawrence
Washington
Montana
North Dakota
Minnesota
New Brunswick
Prince Edward Island
Nova Scotia
Maine
South Dakota
Wisconsin
Michigan
Oregon
Idaho
Wyoming
New York
Vermont
New Hampshire
Massachusetts
Rhode Island
Connecticut
New Jersey
Delaware
Maryland
District of Columbia
Nebraska
Iowa
Pennsylvania
Illinois
Indiana
Ohio
Nevada
Utah
Colorado
Kansas
Missouri
West Virginia
Kentucky
Virginia
California
Oklahoma
Arkansas
Tennessee
North Carolina
South Carolina
Arizona
New Mexico
Mississippi
Alabama
Georgia
Texas
Louisiana
Florida
Gulf of California
Gulf of Mexico
México
North Atlantic Ocean
Cuba

Who We Are

National Science Foundation
WHERE DISCOVERIES BEGIN

PMN:
- Sue Rhee (*PI*)
- Kate Dreher (*curator*)
- A. S. Karthikeyan (*curator, alumni*)
- Anjo Chi (*tech team*)
- Cynthia Lee (*tech team*)
- Larry Ploetz (*tech team*)
- Shanker Singh (*tech team*)
- Bob Muller (*tech team*)
- Vanessa Kirkup (*tech team, alumni*)
- Tom Meyer (*tech team, alumni*)

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

Outline

- General introduction
- Browse/Search/Analyze
- Pathway database creation and curation

Introducing the PMN

- Scope
  - PMN is a collection of plant metabolic pathway databases
  - PMN is a community for data curation
    - Curators, editorial board, ally databases, researchers

- Major goals
  - Create metabolic pathway databases for plants
    - For individual plant species
      - e.g. AraCyc (Arabidopsis thaliana)
      - e.g. PoplarCyc (Populus trichocarpa)
    - Combining data for all plant species - PlantCyc

  - Create a computational prediction "pipeline":
    - Start with protein sequences for a specific plant species
    - End with a comprehensive set of predicted enzyme functions and associated metabolic pathways

PMN Databases

- AraCyc, PoplarCyc, and more to come
  - Single-species
  - Comprehensive collection of pathways in a particular species
  - Complete collection of enzymes, known or predicted, in that species
- PlantCyc
  - Multiple-species
  - Comprehensive collection of pathways for all plants
  - Representative collection of known enzymes in plants

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).

Other Plant Databases Accessible From PMN

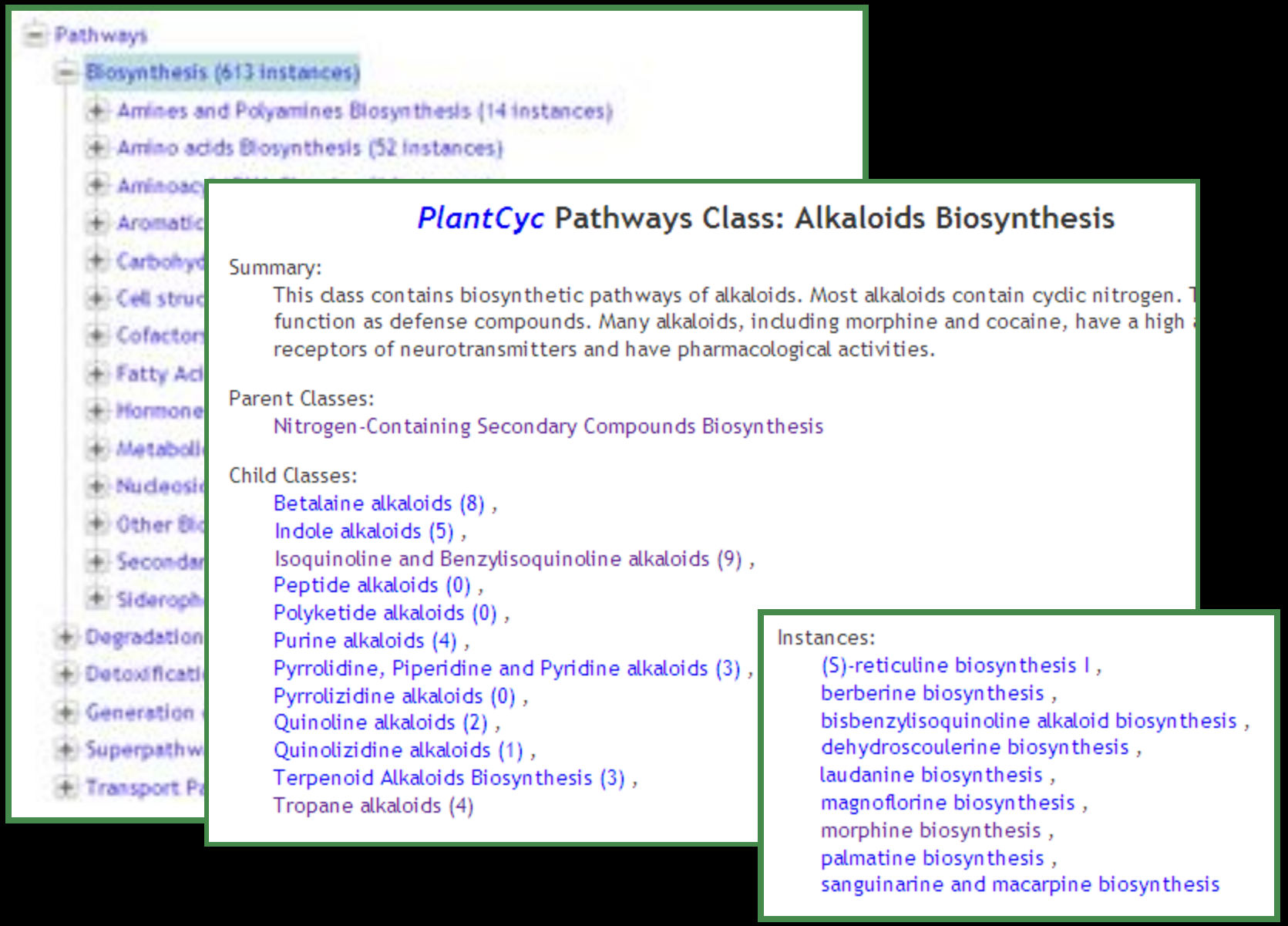
Database	Species	Source	Curation status
RiceCyc ***	Rice	Gramene	some curation
SorghumCyc	Sorghum	Gramene	no curation
MedicCyc ***	Medicago	Noble Foundation	some curation
LycoCyc ***	Tomato	Sol Genomics Network	some curation
PotatoCyc	Potato	Sol Genomics Network	no curation
CapCyc	Pepper	Sol Genomics Network	no curation
NicotianaCyc	Tobacco	Sol Genomics Network	no curation
PetuniaCyc	Petunia	Sol Genomics Network	no curation
CoffeaCyc	Coffee	Sol Genomics Network	no curation

*** Significant numbers of genes from these databases have been integrated into PlantCyc

Browse/Search/Analyze

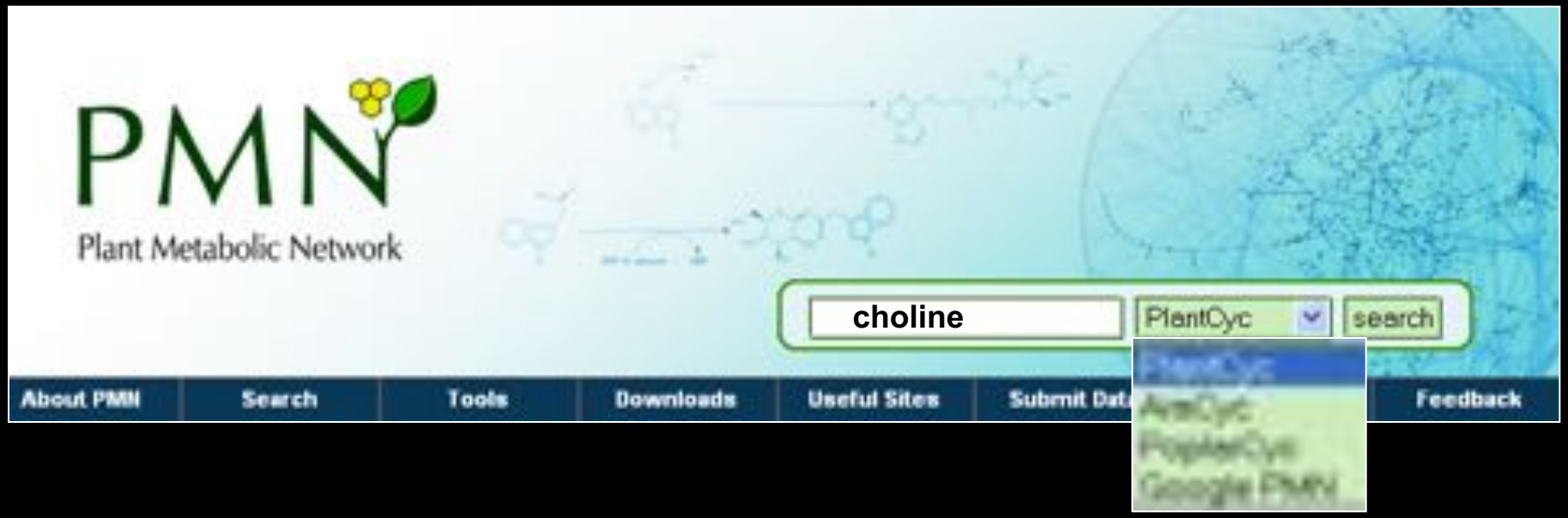
Browsing the PMN Data

Browsing Pathways



Quick Search

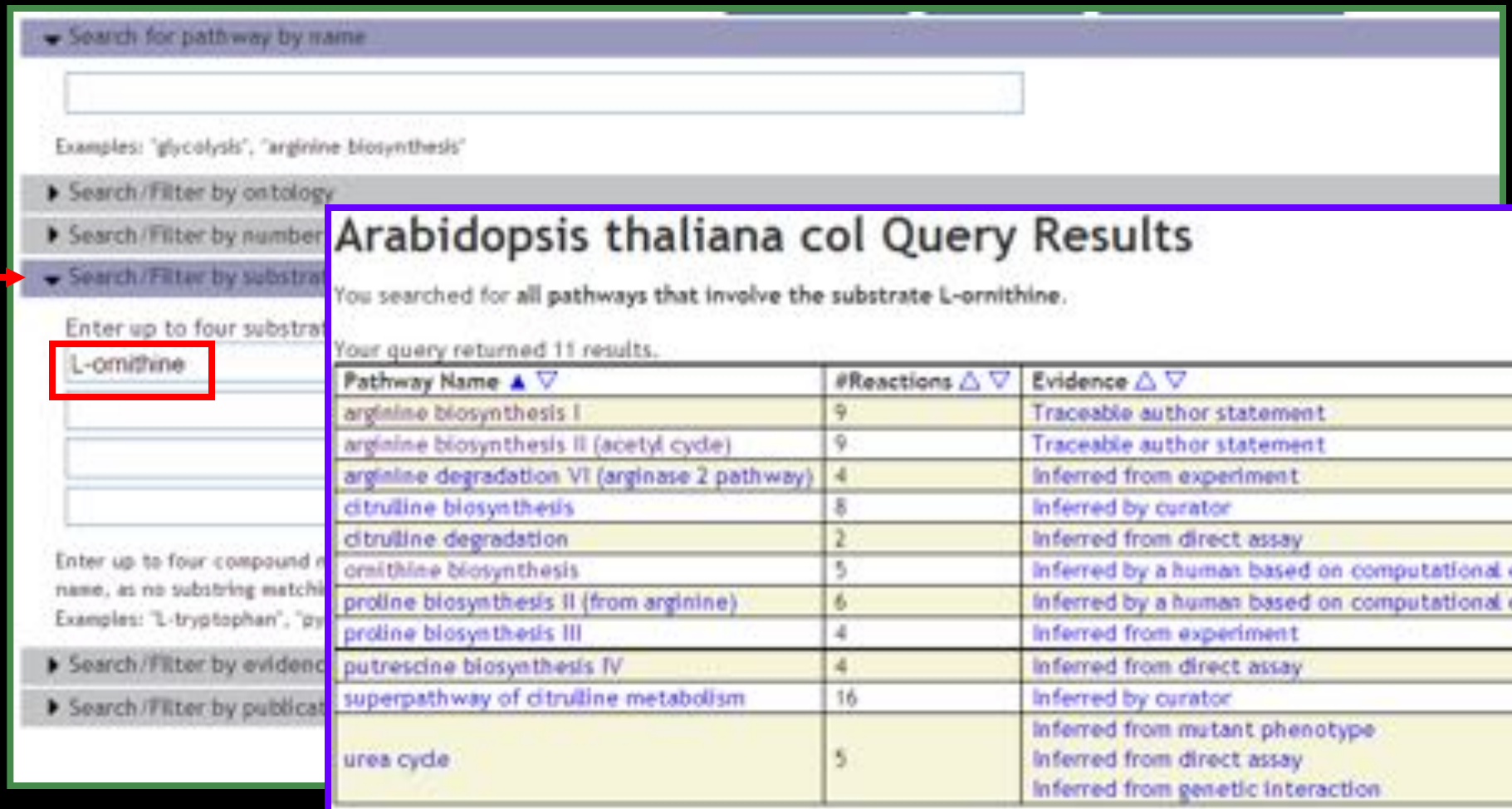
- Quick search bar



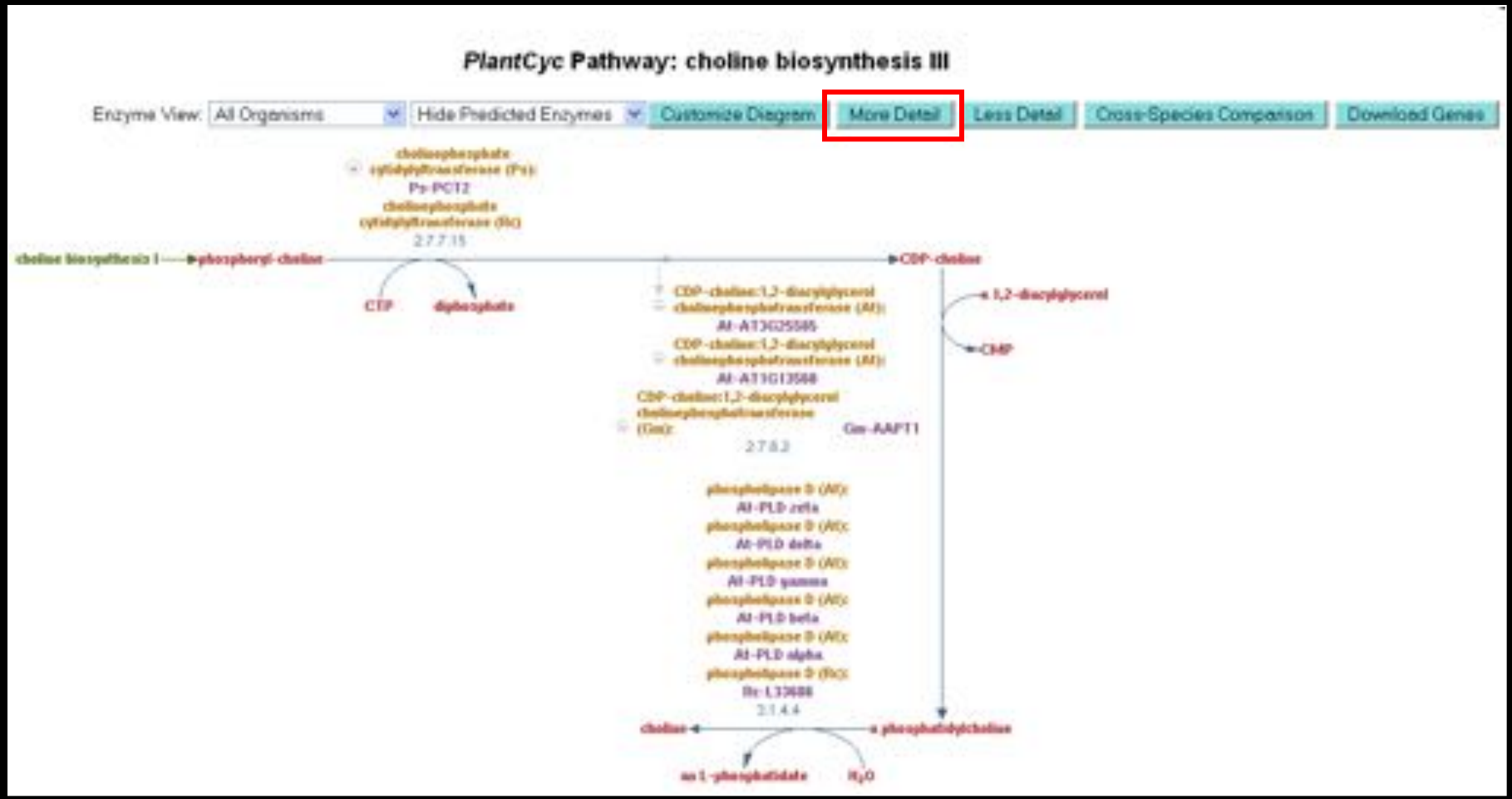
Searching in PMN databases

Specific Data Type Search, Pathway Search

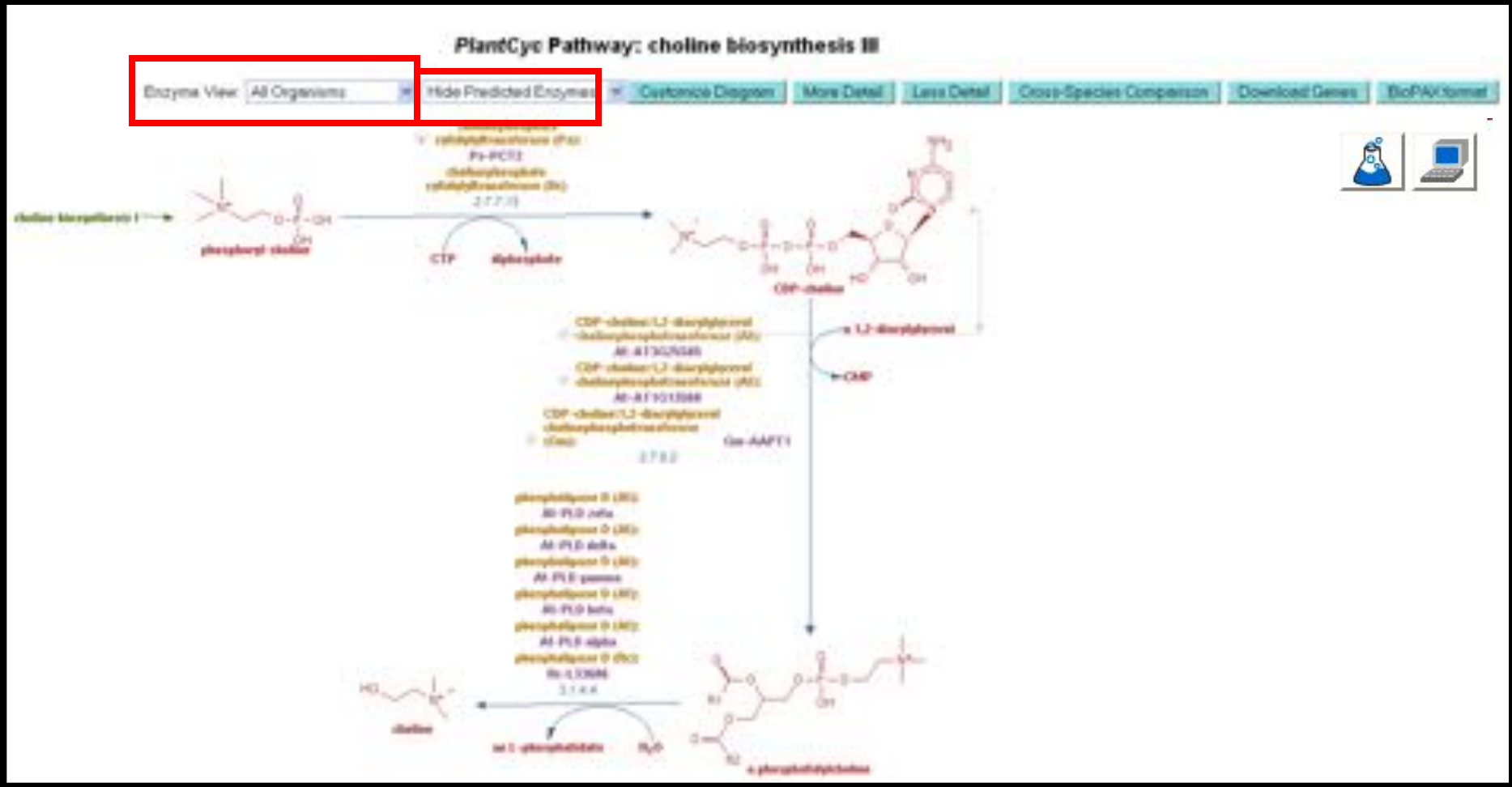
- For example, find pathways that includes a specific **intermediate ornithine**



A Typical Pathway Detail Page



A Typical Pathway Detail Page

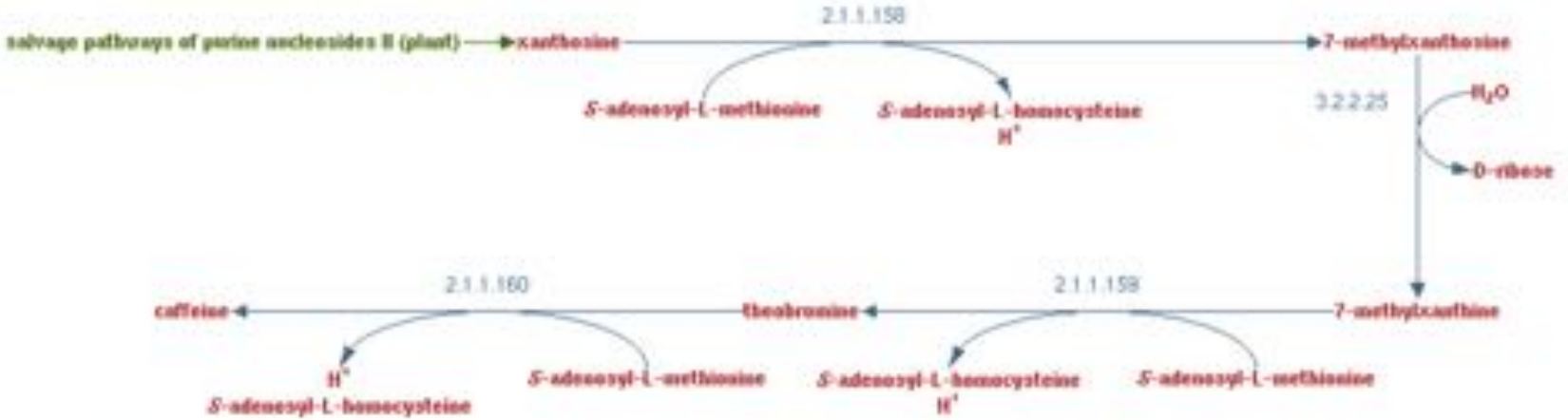


Conventions Used in Curation and Data Presentation

- 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 | More Detail | Less Detail | Species Comparison



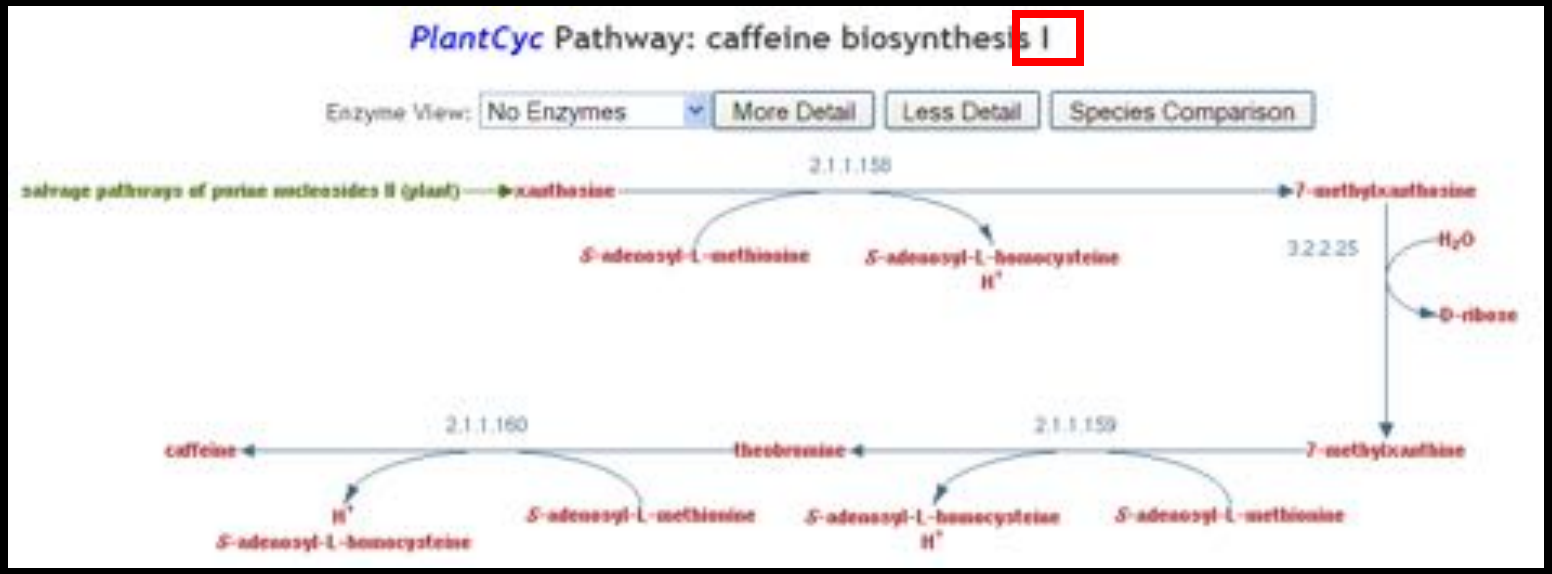
If an enzyme name is shown in bold, there is experimental evidence for this enzymatic activity.

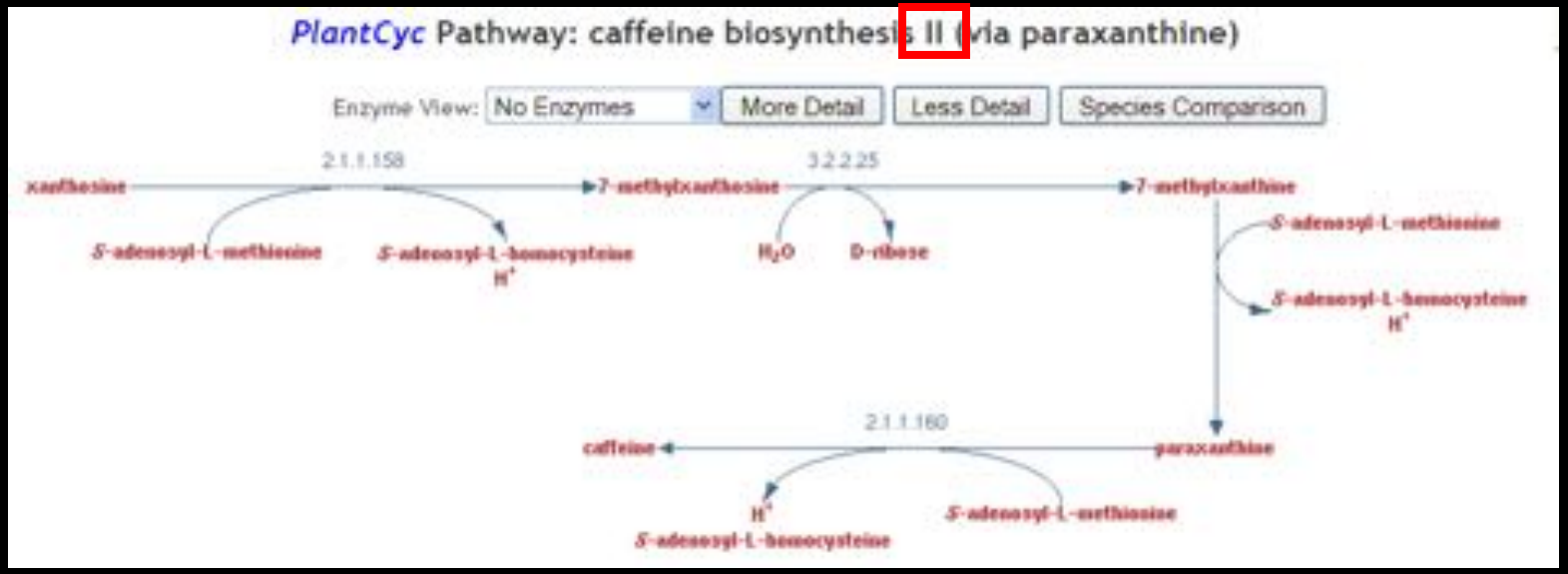
Superclasses: Biosynthesis -> Secondary Metabolites Biosynthesis -> Nitrogen-Containing Secondary Compounds Biosynthesis -> Alkaloids Biosynthesis -> Purine alkaloids -> Caffeine Biosynthesis

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 variants


PlantCyc Pathway: caffeine biosynthesis I
Enzyme View: No Enzymes
More Detail
Less Detail
Species Comparison
salvage pathways of purine nucleosides II (plant)
xanthosine
2.1.1.158
S-adenosyl-L-methionine
S-adenosyl-L-homocysteine
H+
7-methylxanthosine
3.2.2.25
H2O
D-ribose
caffeine
2.1.1.160
H+
S-adenosyl-L-homocysteine
S-adenosyl-L-methionine
theobromine
2.1.1.159
7-methylxanthine
S-adenosyl-L-homocysteine
H+
S-adenosyl-L-methionine

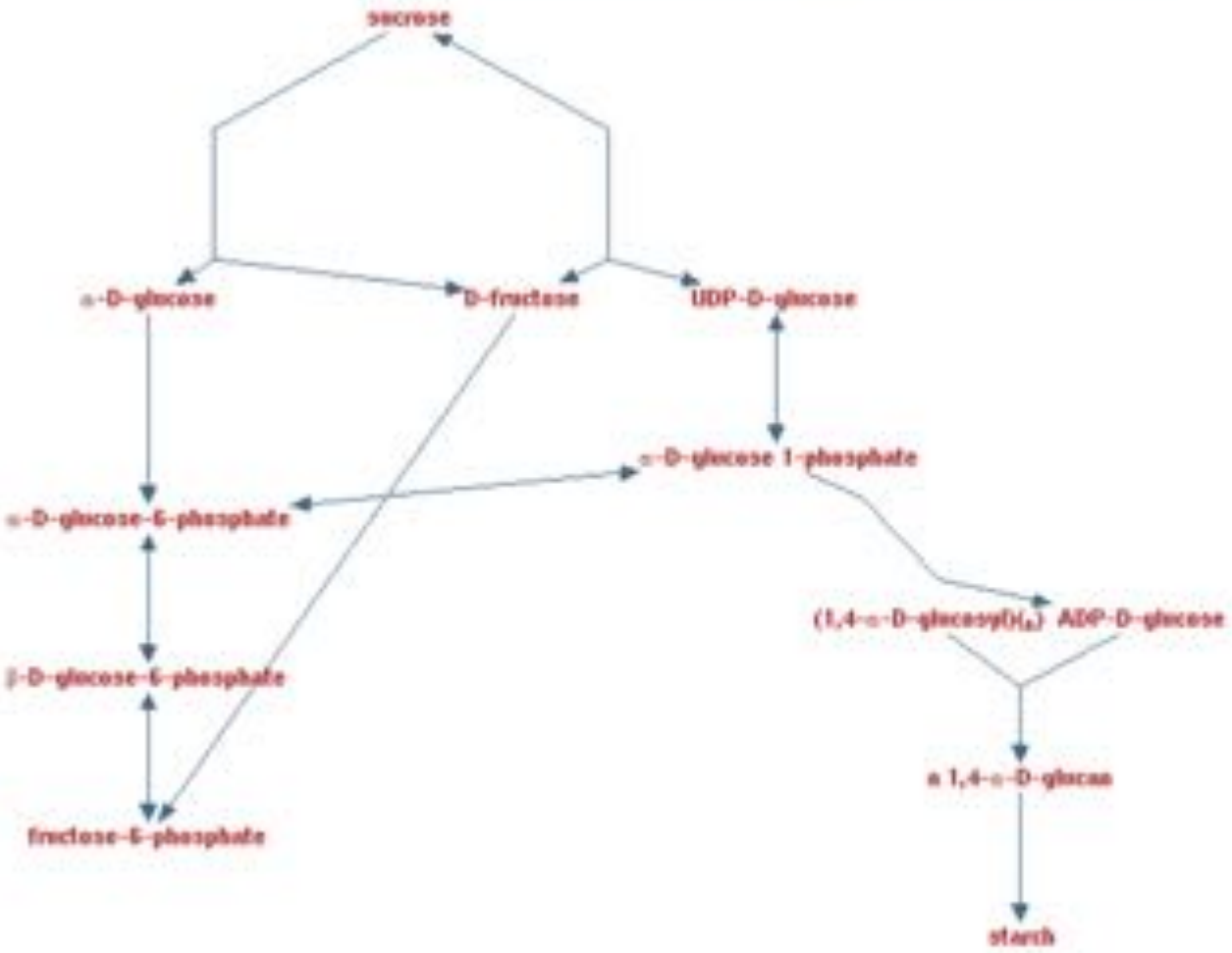

PlantCyc Pathway: caffeine biosynthesis II (via paraxanthine)
Enzyme View: No Enzymes
More Detail
Less Detail
Species Comparison
xanthosine
2.1.1.158
S-adenosyl-L-methionine
S-adenosyl-L-homocysteine
H+
7-methylxanthosine
3.2.2.25
H2O
D-ribose
7-methylxanthine
S-adenosyl-L-methionine
S-adenosyl-L-homocysteine
H+
caffeine
2.1.1.160
H+
S-adenosyl-L-homocysteine
S-adenosyl-L-methionine
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: superpathway of sucrose and starch metabolism I (non-photosynthetic tissue)

More Detail | Less Detail | Species Comparison



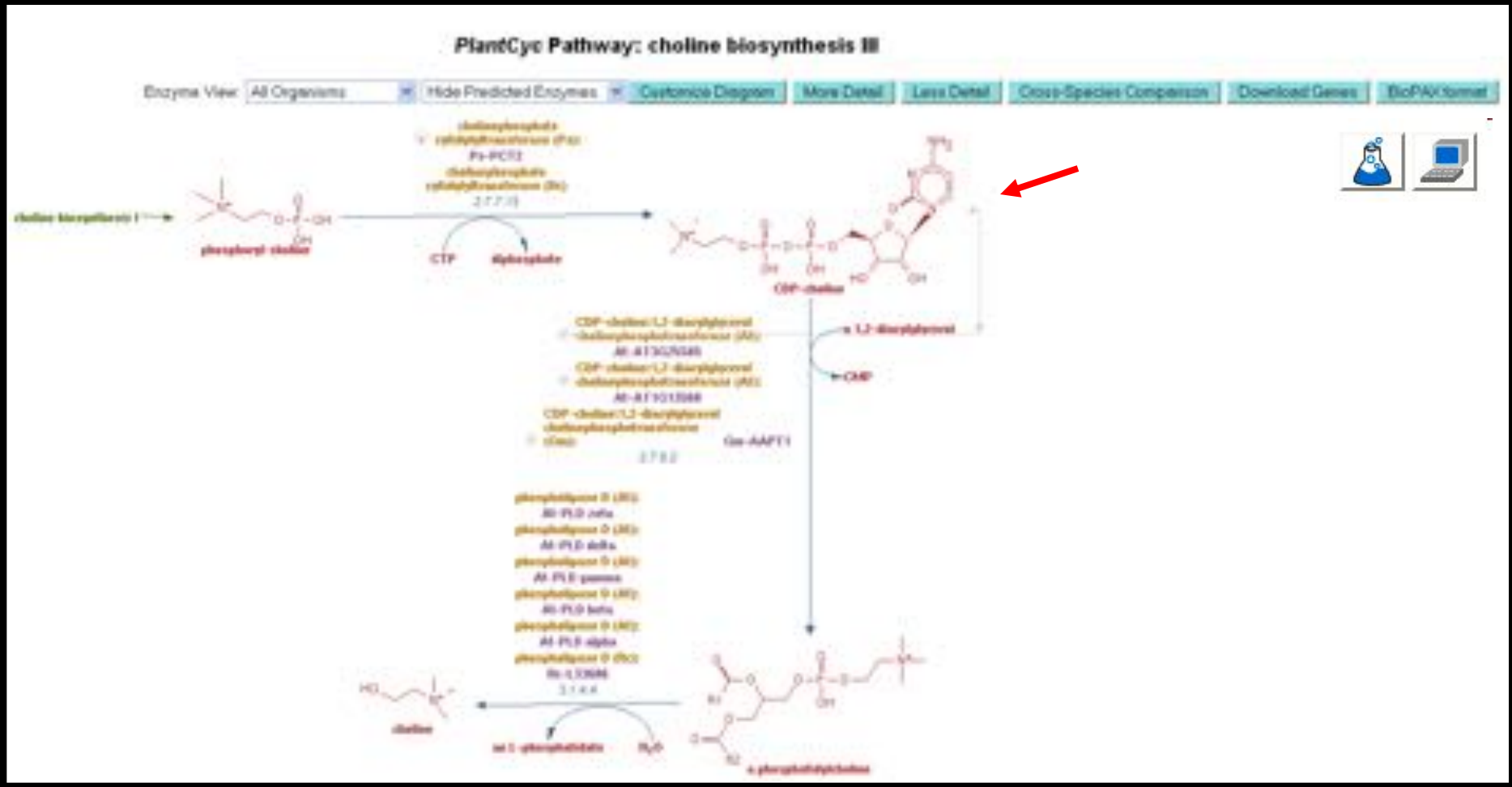
If an enzyme name is shown in bold, there is experimental evidence for this enzymatic activity.

Superclasses: Superpathways

Species Data Available for: Arabidopsis thaliana col

Subpathways: starch biosynthesis , sucrose degradation III

Linking to Other Data Detail Pages



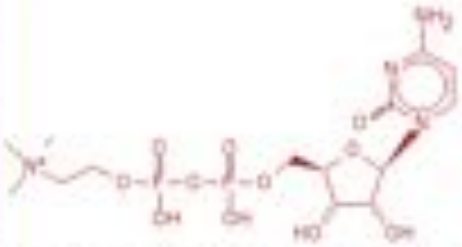
Compound Detail Pages

Synonyms: citicoline , citicholine , cidilin , 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: $C_{14}H_{25}N_4O_{11}P_2$

Molecular Weight: 489.34 daltons



Smiles: C(OP(O)(=O)OP(O)(=O)OCC[N+](C)(C)C)C1OC(C(O)C1O)n2ccc(=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

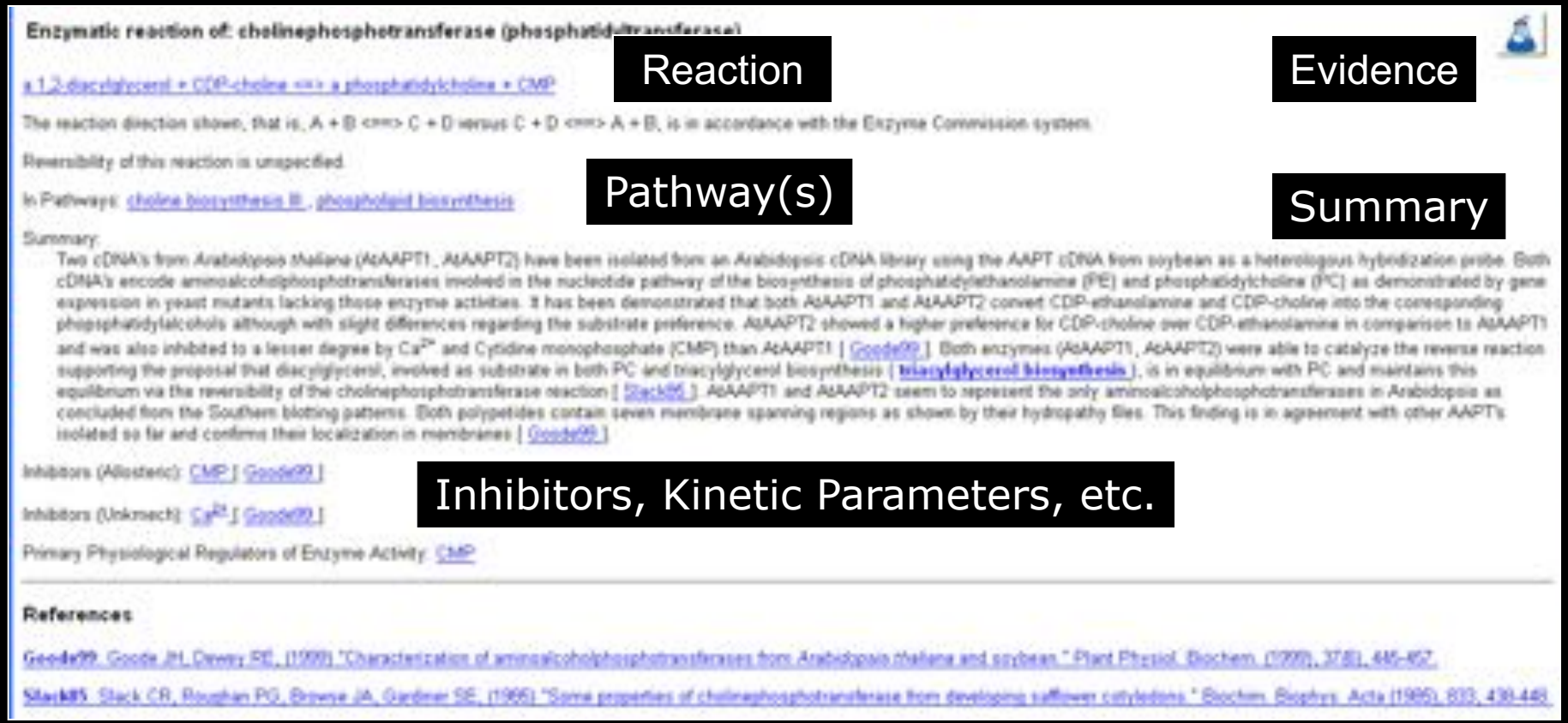
Smiles / InChI

Appears as Reactant

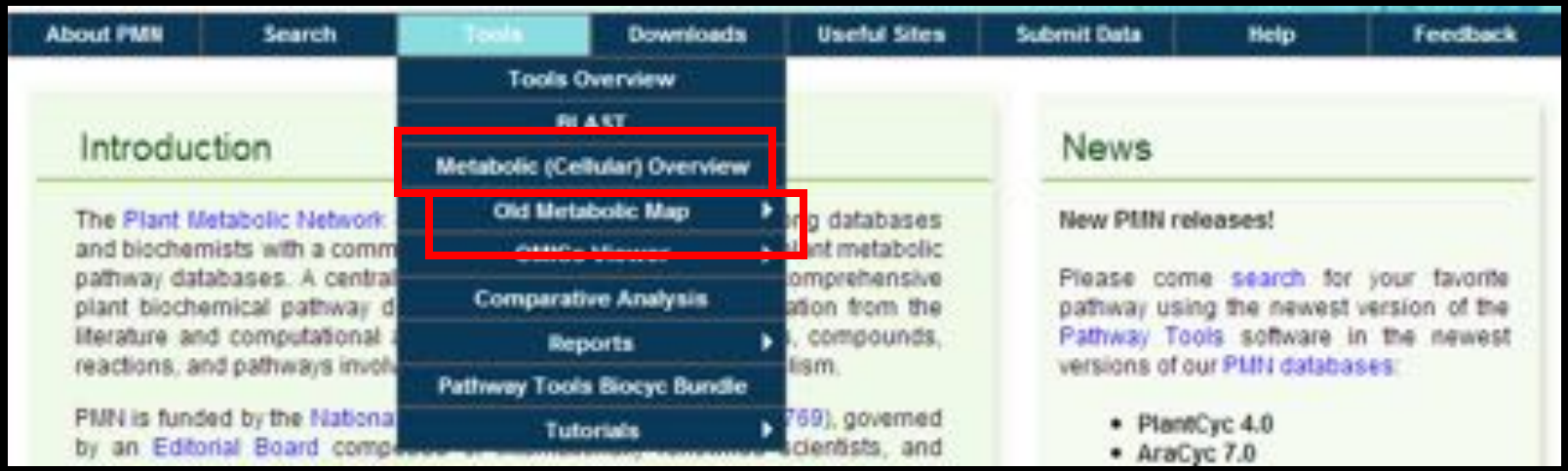
Appears as Product

Enzyme detail pages

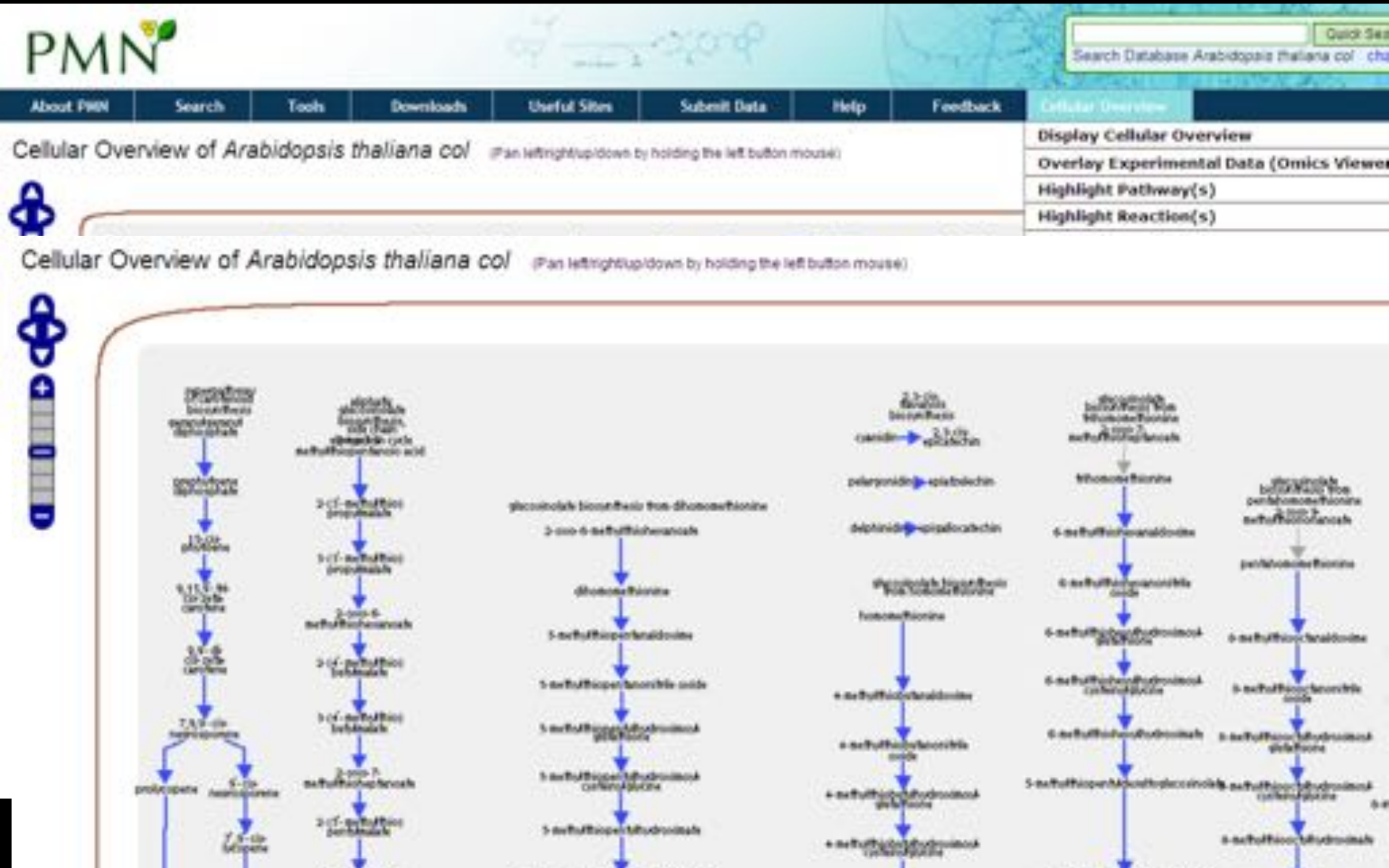
Arabidopsis **Enzyme: phosphatidyltransferase**



The Global Overview Map



The Global Overview Map



Visualizing Omics Data

Visualize and interpret large scale omics data in a metabolism context:

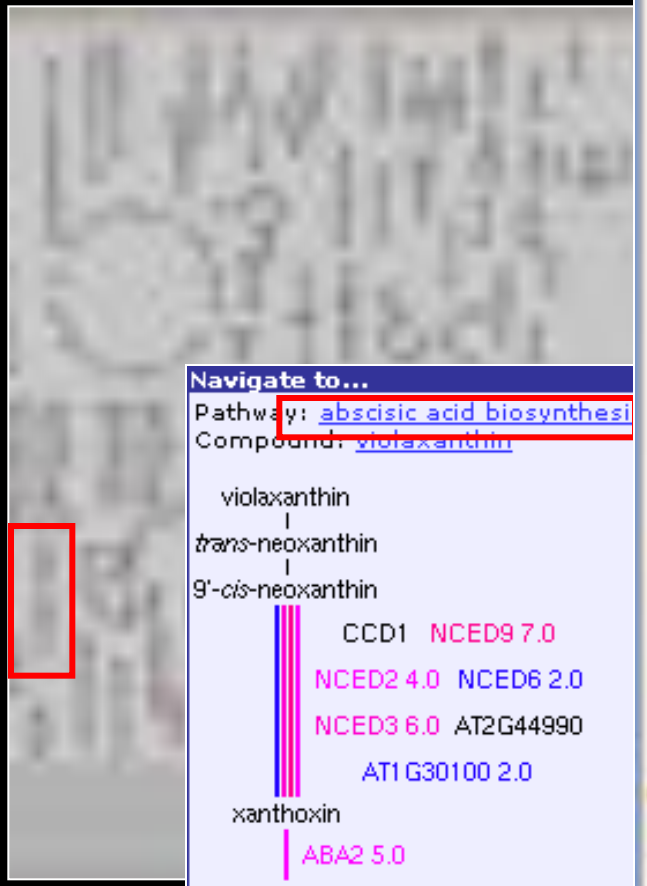
- Gene expression data
- Proteomic data
- Metabolic profiling data
- Reaction flux data

Input File Format For The Omics Viewer

- Tab-delimited text file

At1g77760	1.15	2.3	3.2	2.15	1.53	1.75
At2g13360	0.7	-0.53	0	-0.73	0.03	-0.72
At3g10230	-1.1	-0.05	1.05	1.15	1.25	0.05
At3g10230	-0.65	-0.58	1.13	1.23	0.67	-0.12
At3g01120	-1.08	-0.15	-1.2	-1.15	-1.15	-0.58
At3g01500	0.07	-0.72	-0.68	-1.4	-1.93	-9.23
At3g02470	0.03	-0.53	0.58	1.28	0.55	1.4
At3g02470	0.55	-0.12	0.62	0.65	-0.05	1.22
At3g02580	0.6	-0.55	0.08	0.55	-2.2	-1.65

Omics Viewer



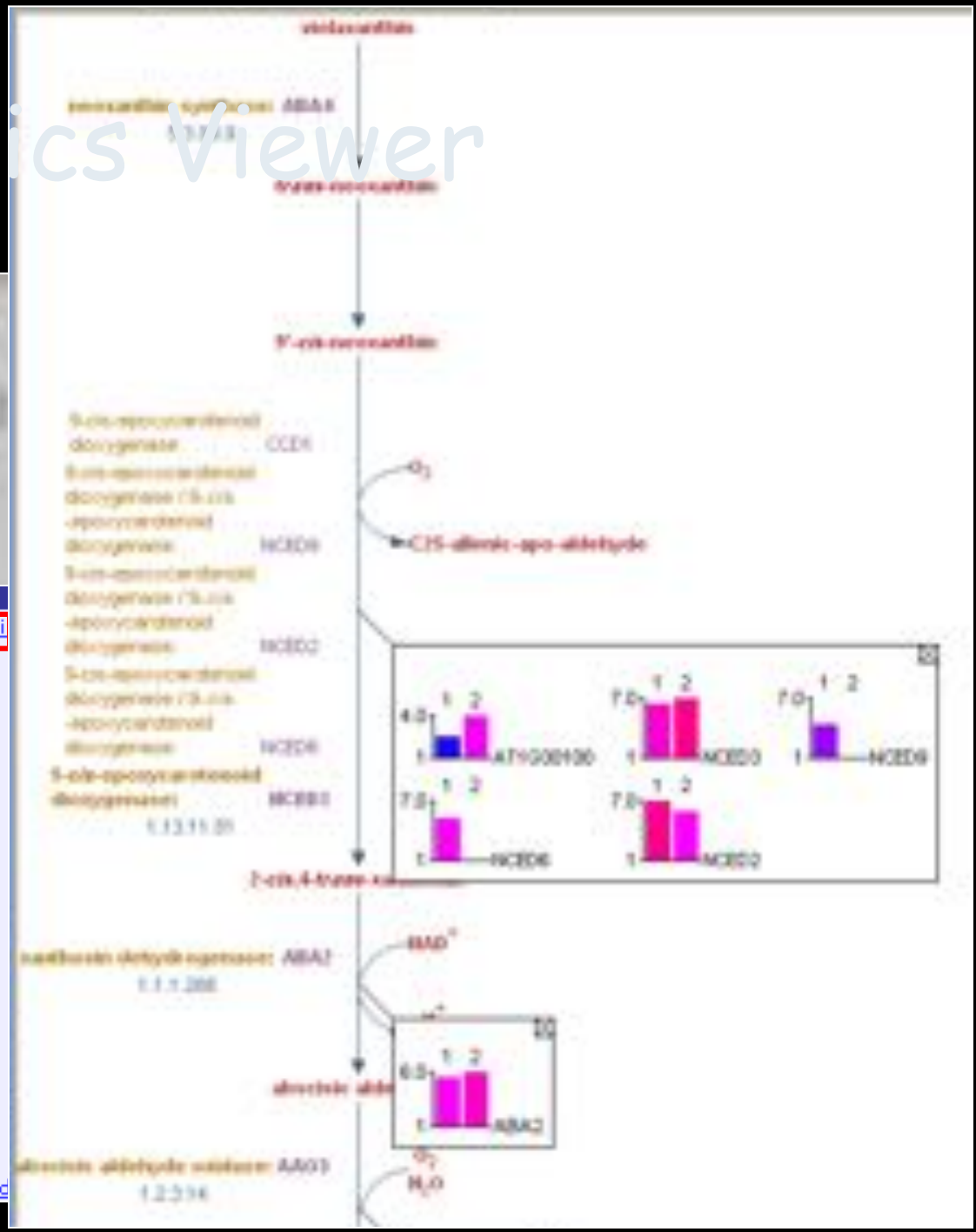
Navigate to...

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



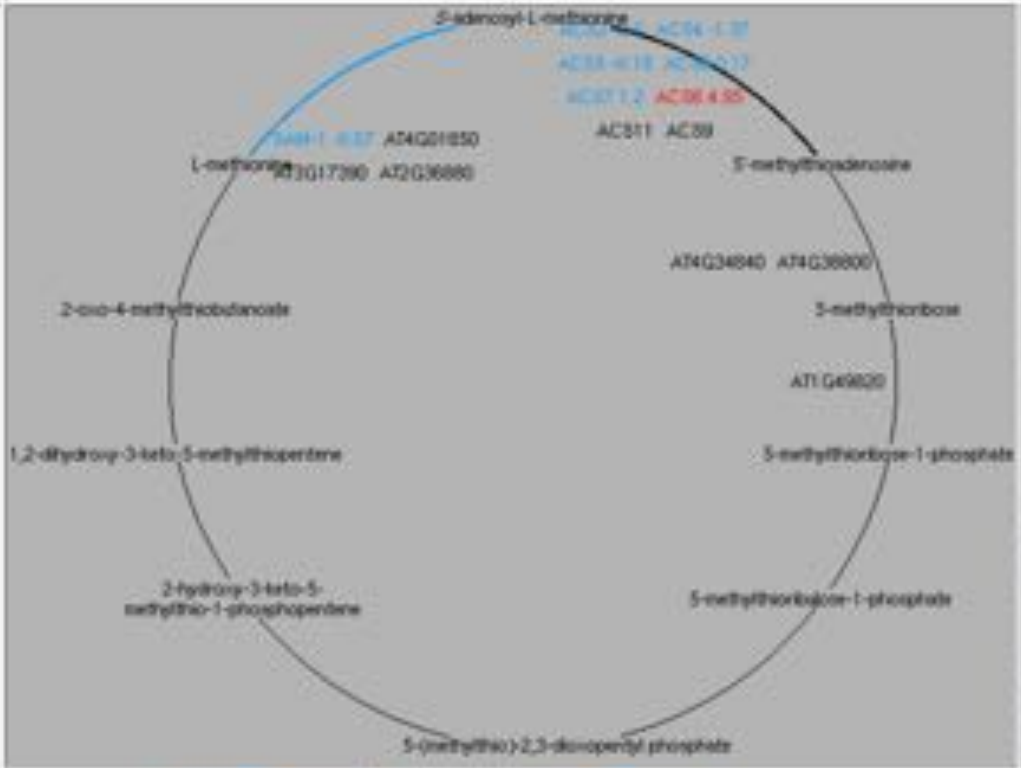
Omics Viewer: Color Coding Gene Expression Levels

Red: Enhanced expression over my threshold (i.e. 2-fold change)

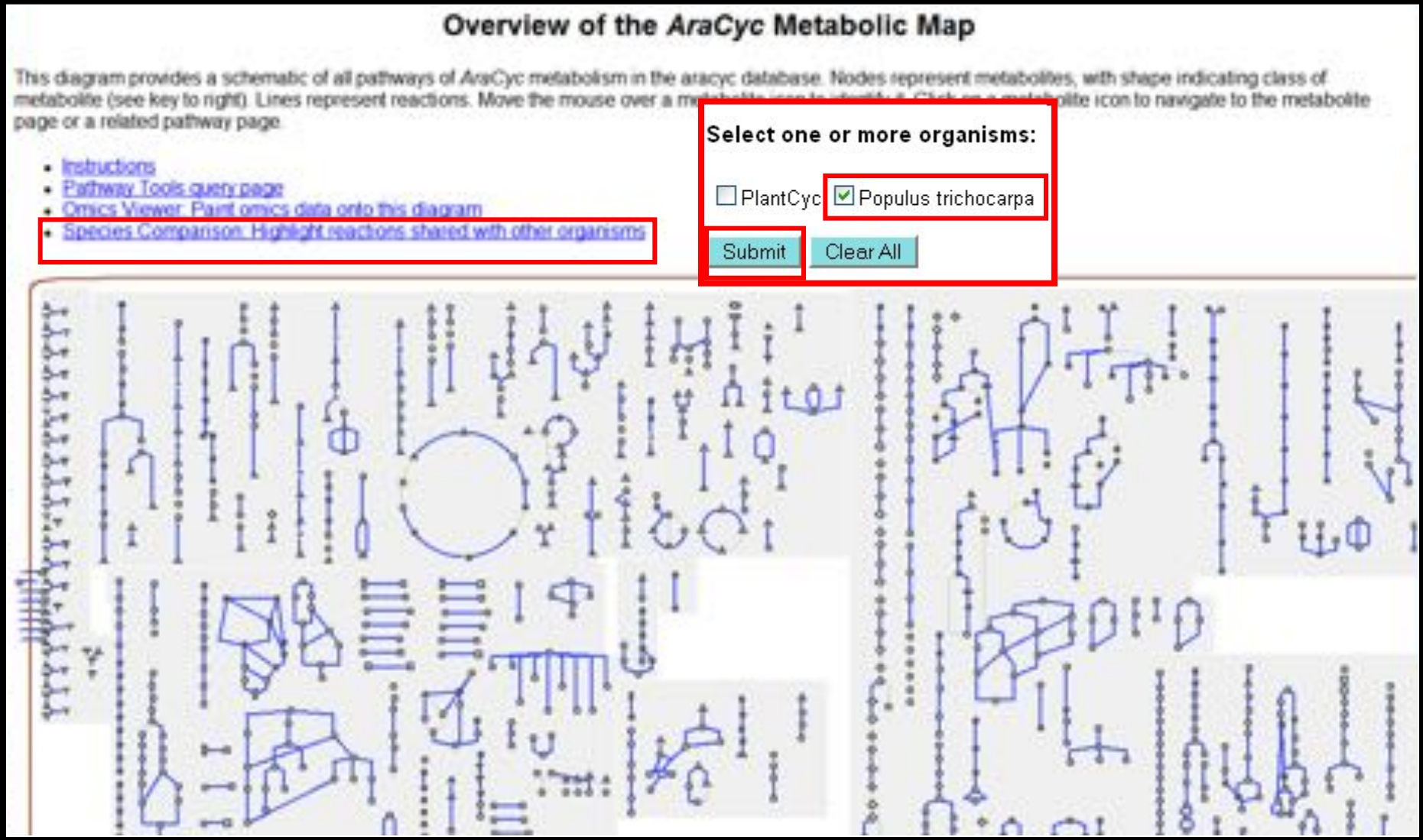
Yellow: Repressed expression over my threshold

Blue: Not significantly changed over my threshold

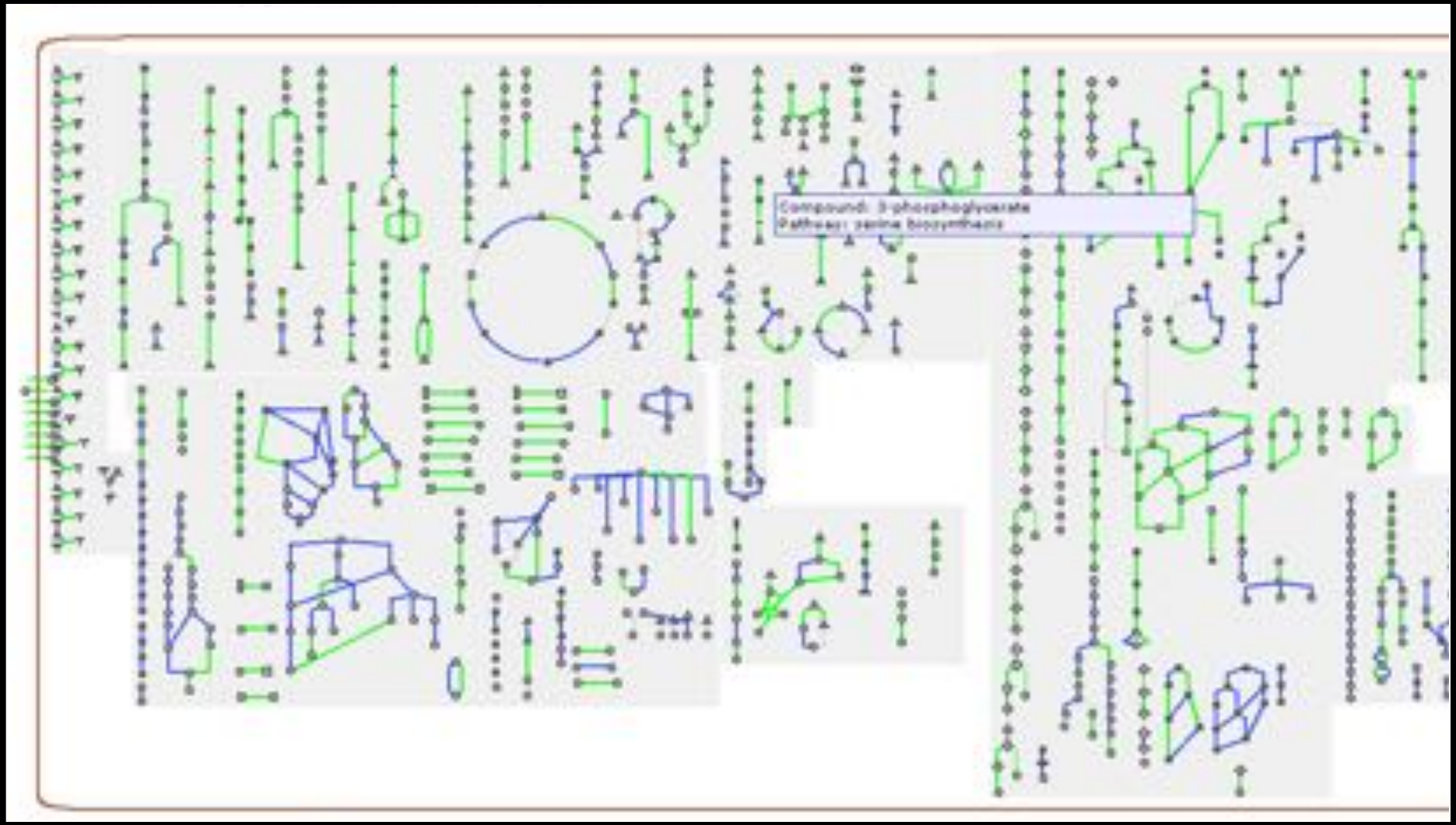
Omics Viewer: generating a table of individual pathways exceeding certain threshold

Pathway	Pathway Diagram	Enzymes, Genes, and Enzyme Cellular Locations	
methionine salvage pathway		1-aminocyclopropane-1-carboxylate synthase	ACS9
		1-aminocyclopropane-1-carboxylate synthase	ACS11
		1-aminocyclopropane-1-carboxylate synthase	ACS6
		1-aminocyclopropane-1-carboxylate synthase	ACS7
		1-aminocyclopropane-1-carboxylate synthase	ACS8
		1-aminocyclopropane-1-carboxylate synthase	ACS5
		1-aminocyclopropane-1-carboxylate synthase	ACS4
		1-aminocyclopropane-1-carboxylate synthase	ACS2
		catalytic / methylthioadenosine nucleosidase	AT4G388
		methylthioadenosine nucleosidase	AT4G348
		5-methyl-5-thioribose kinase	AT1G498
		methionine adenosyltransferase	AT2G368
		methionine adenosyltransferase	AT3G17
		methionine adenosyltransferase	AT4G018
		methionine adenosyltransferase	SAM-1
purine nucleotide metabolism (phosphotransfer and nucleotide modification)		IMP dehydrogenase	AT1G79470
		IMP dehydrogenase	AT1G16350
		GMP synthetase / GMP synthase (glutamine-hydrolyzing)	AT1G63660
		guanylate kinase	AT2G41880
		guanylate kinase	AT3G57550
		adenylosuccinate synthase	AT3G57610
		adenylosuccinate lyase	AT4G18440
		adenylosuccinate lyase	AT1G36280
		adenylate kinase / nucleotide kinase	AT2G37250

Comparing Across Species



Comparing Across Species



Comparing Across Species

Select which set(s) of comparative-analysis tables you wish to generate:

- ☐ **Reactions**: breakdowns by type of substrate, by EC Number, by number of isozymes, etc.
- ☑ **Pathways**: breakdown by pathway class, information on pathway holes.
- ☐ **Compounds**: small molecules that act as substrates, enzyme activators/inhibitors/cofactors.
- ☐ **Proteins**: breakdown of protein complexes by type and number of subunits, number of enzymes.
- ☐ **Orthologs**: proteins shared among organisms or unique to an organism.
- ☐ **Transporters**: proteins that facilitate the movement of compounds across cell membranes.
- ☐ **Transcription Units**: number of genes per transcription unit, number of operons per pathway.

Select one or more organism databases:

☑ Arabidopsis thaliana col ☐ PlantCyc ☑ Populus trichocarpa

Submit | Clear All

Pathways

Table 1: Breakdown of Pathways by Pathway Class

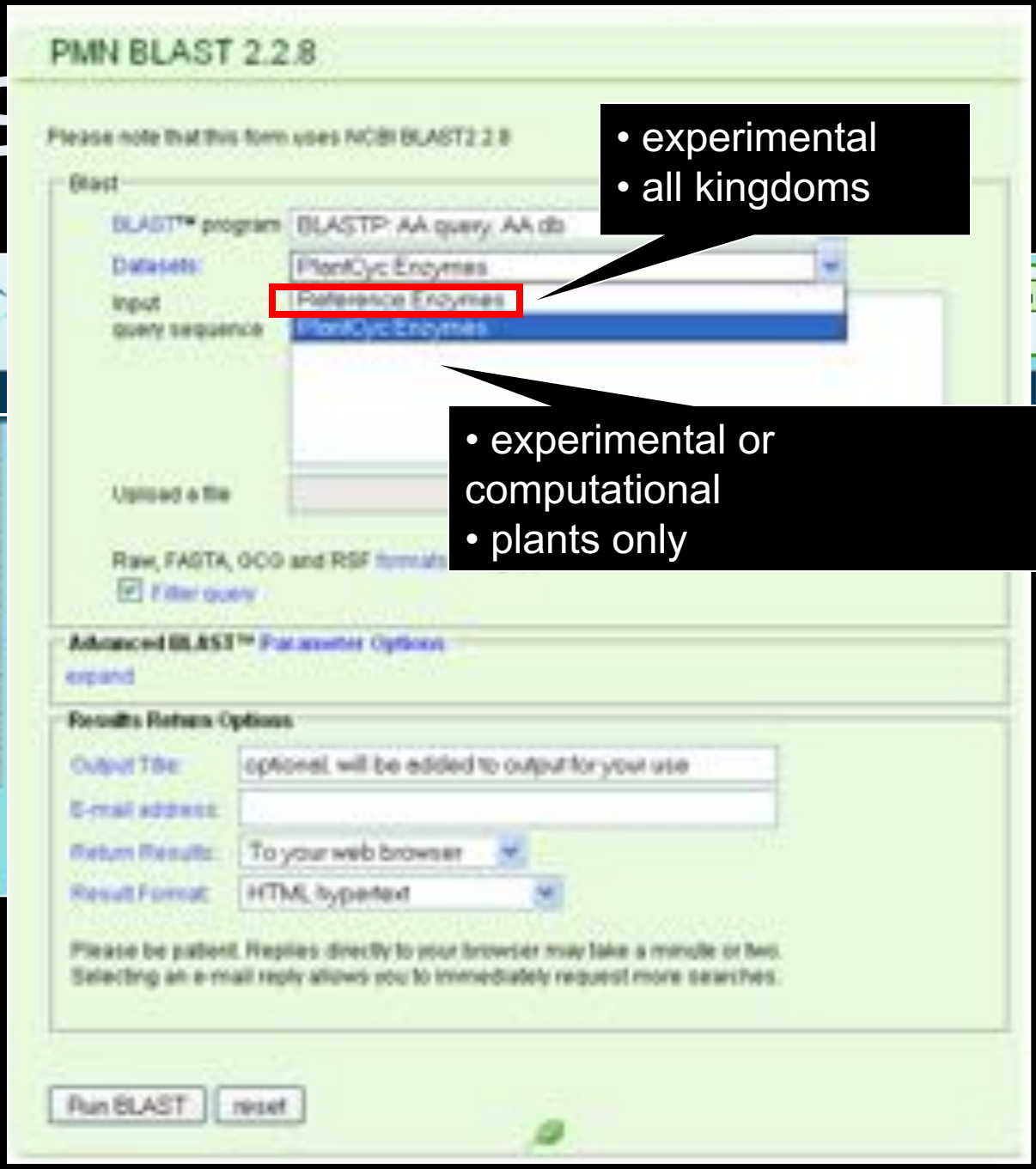
This table presents
down further to sh
assigned to more t
will see only those

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

Pathway Class: Biosynthesis - Hormones Biosynthesis	AraCyc col	P. trichocarpa
cis-zeatin biosynth		
ent-kaurene biosyr		
trans-zeatin biosyr		
abscisic acid biosyr		
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 — 3β-hydroxysteroid dehydrogenase:
		EC# 1.3.99.- — sterol 5-alpha reductase: DET2
		EC# 1.14.13.- — sterold 22α-hydroxylase: DWF4
		RXN-712 — None
		EC# 1.14.13.- — sterold 22α-hydroxylase: DWF4
		EC# 1.14.13.- — sterold 22α-hydroxylase: DWF4
		EC# 1.14.13.- — sterold 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

PMN B

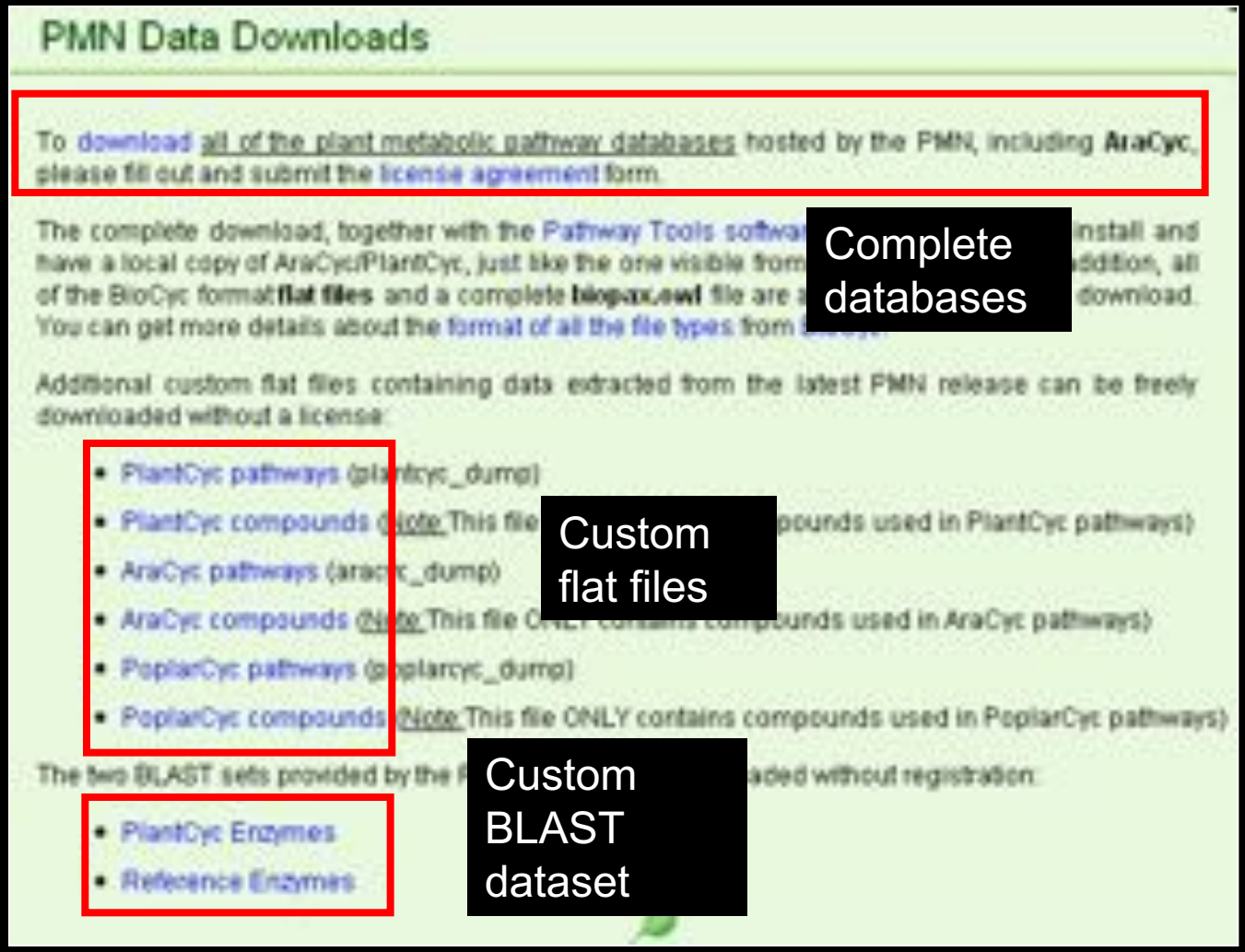


- experimental
- all kingdoms

- experimental or computational
- plants only

Online Tutorials

Data Downloads



Download and Install a Local Copy of the PMN Databases

- Run robust live database query by scripts, via Perl, Java, LISP interfaces
- Edit with private data
- Access to additional features not available on web mode
- Free, open database license
- Pathway Tools Software (SRI)

Developing The PMN

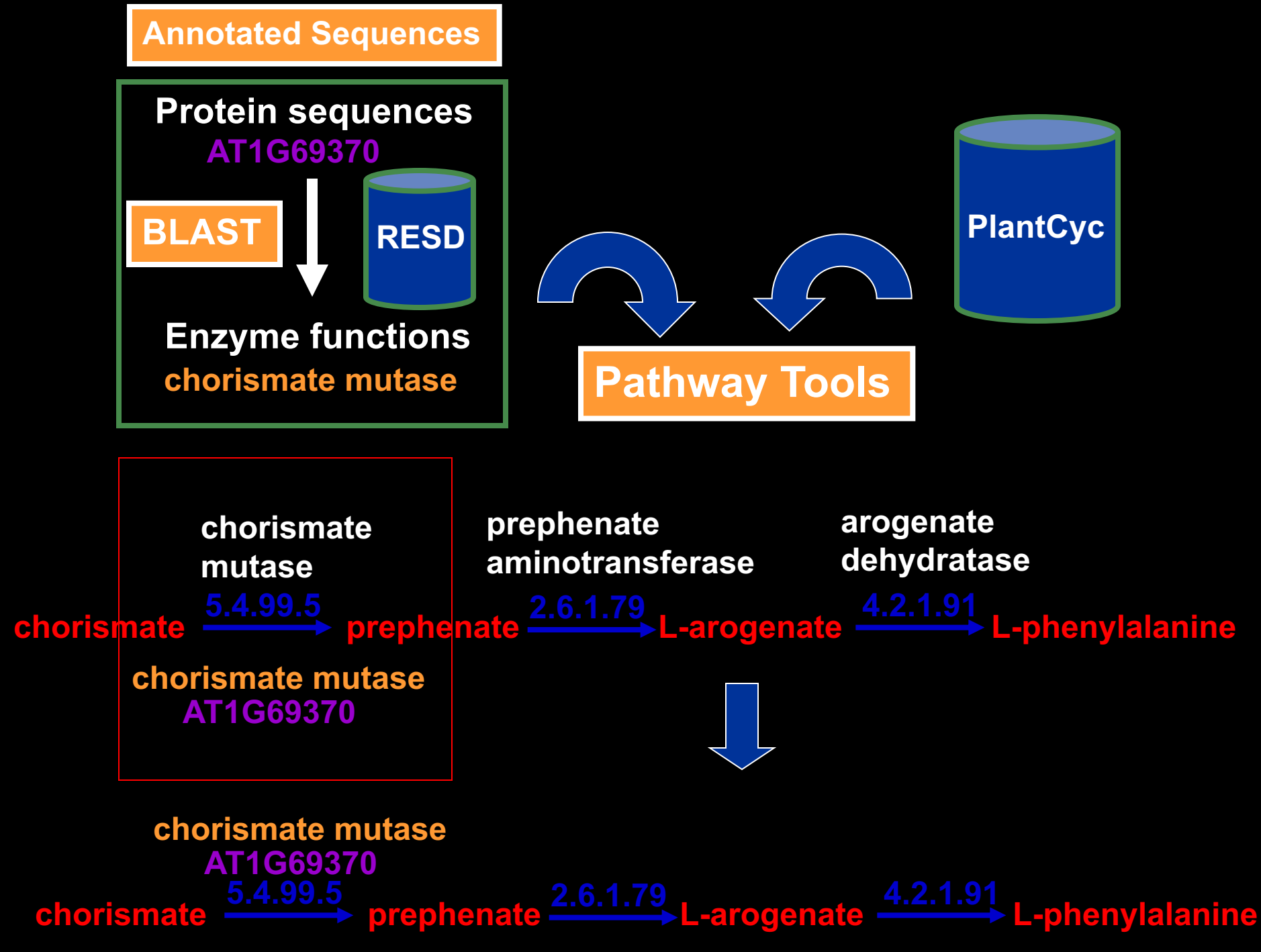
Creating Single-Species Databases

- New sets of DNA sequences become available
  - Genomes are sequenced
  - Large EST data sets are created
    - Unigene builds are generated

- PMN pipeline predicts enzyme functions
  - Based on sequence similarity to known enzymes, enzymes with experimental or literature support

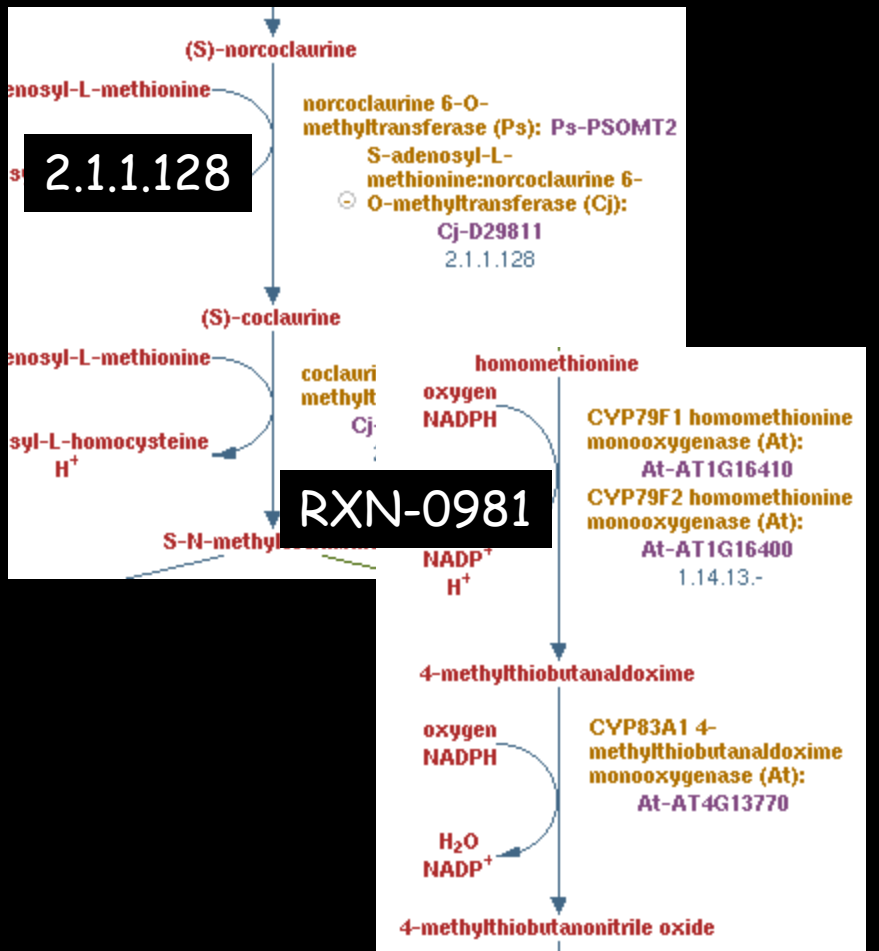
- Set of predicted enzymes is used to predict metabolic pathways
  - The pathway prediction software (Pathway Tools) uses:
    - Enzyme functional annotations
    - A reference set of pathways (e.g. PlantCyc)

- Curators validate predicted pathways in the new database
  - Remove incorrect information and add additional data


Annotated Sequences
Protein sequences
AT1G69370
BLAST
RESD
Enzyme functions
chorismate mutase
PlantCyc
Pathway Tools
chorismate mutase
5.4.99.5
prephenate aminotransferase
2.6.1.79
arogenate dehydratase
4.2.1.91
chorismate
prephenate
L-arogenate
L-phenylalanine
chorismate mutase
AT1G69370
chorismate mutase
AT1G69370
5.4.99.5
2.6.1.79
4.2.1.91
chorismate
prephenate
L-arogenate
L-phenylalanine

Identifiers Used in Automated Enzyme Annotation and Enzyme to Pathway Mapping

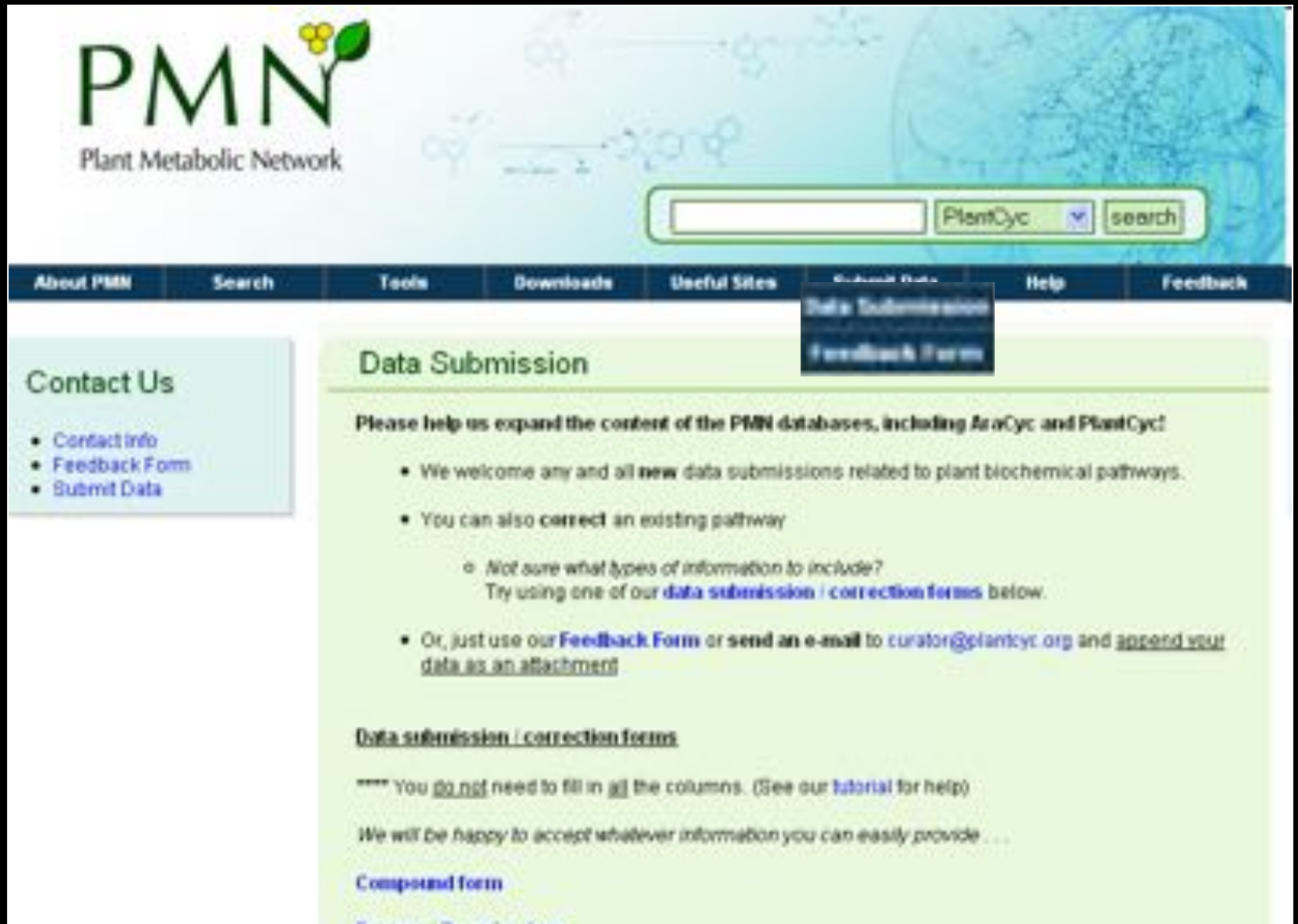
- Complete EC number
  - e.g. 2.1.1.128
- Unique PlantCyc reaction id, when complete EC is not assigned
  - e.g. RXN-0981
- GO term id
  - e.g. GO:560010



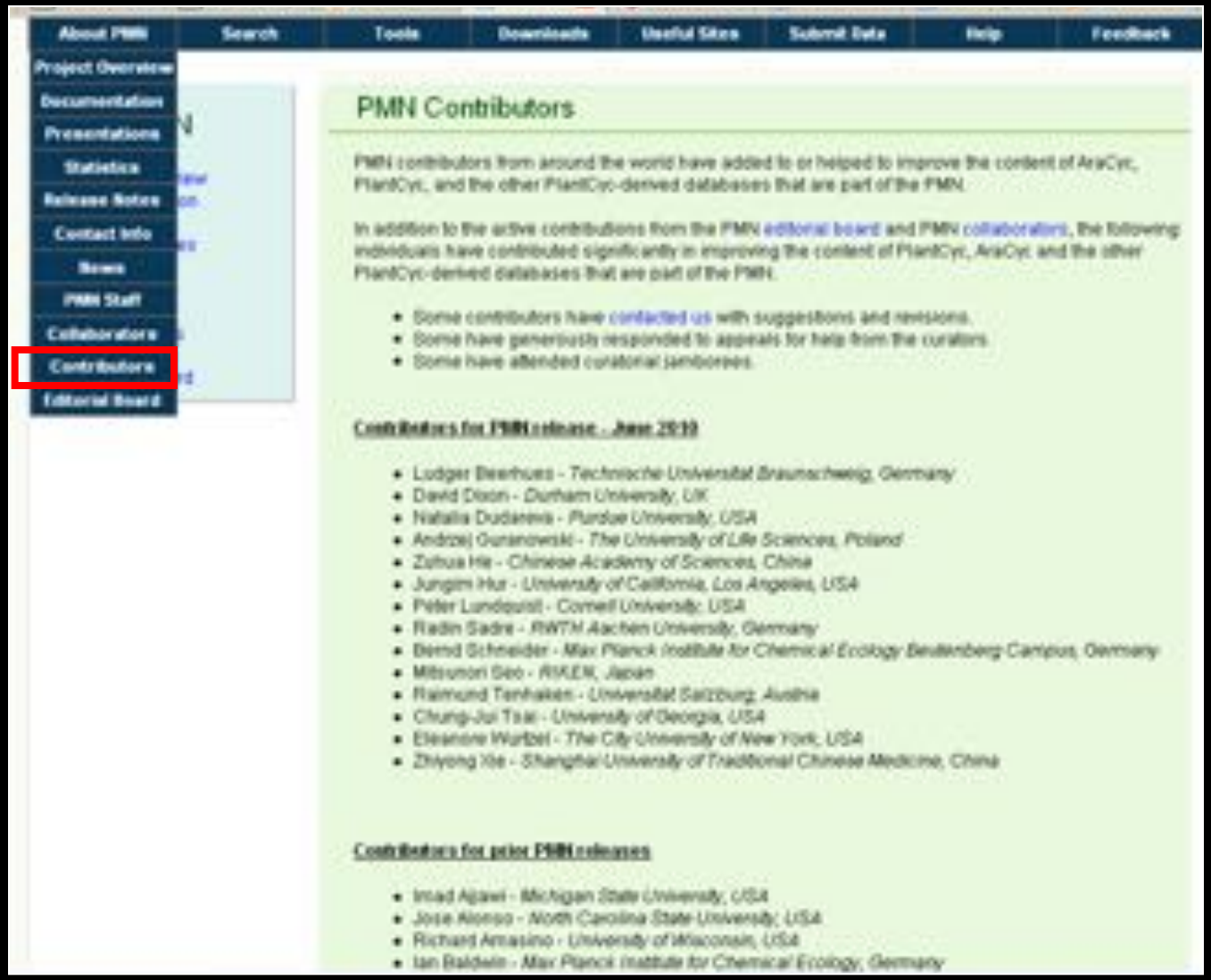
Manual Curation

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

Submitting Data To Us



Community Gratitude



www.plantcyc.org

curator@plantcyc.org

meet Kate, Booth# 219