PlantCyc, AraCyc, PoplarCyc and more . . .

Building databases and connecting to researchers at the Plant Metabolic Network

kate dreher

curator

PMN/TAIR

Introducing the PMN

- The PM… earchers
  - Fun…
- Major g…
  - Cre…
    - …
    - …
  - Cre…
    - …
    - …
  - Cre…
    - …

PMN databases

- Current PMN databases: PlantCyc, AraCyc, PoplarCyc
  - Coming soon: databases for wine grape, maize, cassava, Selaginella, and more . . .

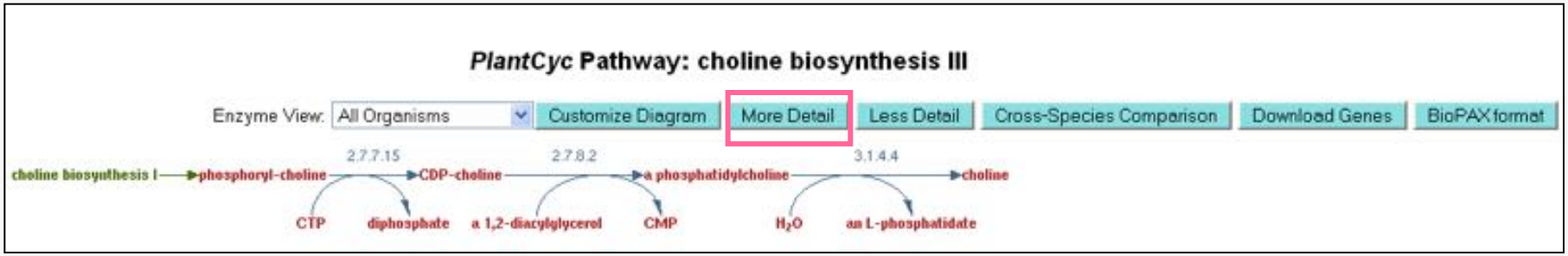
- Other plant databases accessible from the PMN:

PGDB	Plant	Source	Status
RiceCyc **	Ri		
SorghumCyc	So		
MedicCyc **	Me		
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

P60. Connecting the Solaneceae genome to the metabolic networks via SolCyc and MetaCyc.
Anuradha Pujar

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

PMN data types



PMN da

PlantCyc Enzyme: phosphatidyltransferase

Synonyms: aminoalcoholphosphotransferase

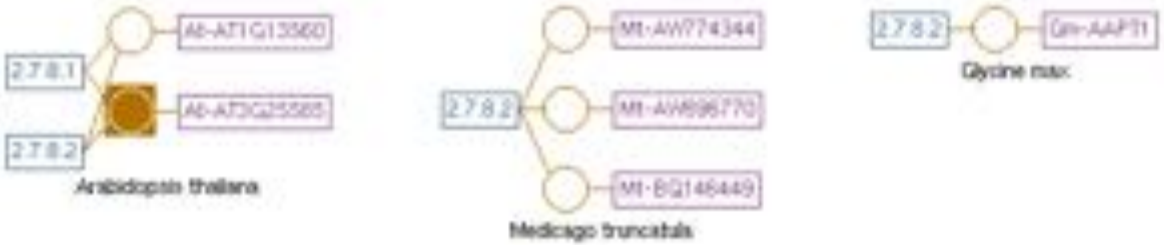
Species: Arabidopsis thaliana col

Gene: AT3G25585

Sequence Length: 1145 AAs

Unification Links: Entrez:AAC61769 , Phytozome Plant Orthologs:AT3G25585.1

Gene-Reaction Schematic:



MultiFun Terms: UNCLASSIFIED

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 , phosphatidylcholine biosynthesis I , phosphatidylcholine biosynthesis II

Summary:

The enzyme is bifunctional that utilizes CDP-choline and CDP-ethanolamine in synthesizing phosphatidylcholine and phosphatidylethanolamine, respectively. Its cholinephosphotransferase activity is greater than its ethanolaminephosphotransferase activity.

Inhibitors (Unkmech): CMP [Gorde89] , Ca^{2+} [Gorde89]

Upstream pathway

Compound

Evidence Codes

thway

PMN database content statistics

- New PMN release . . . almost here!
  - Increased data content
    - shhh . . . sneak preview at pmn.plantcyc.org!

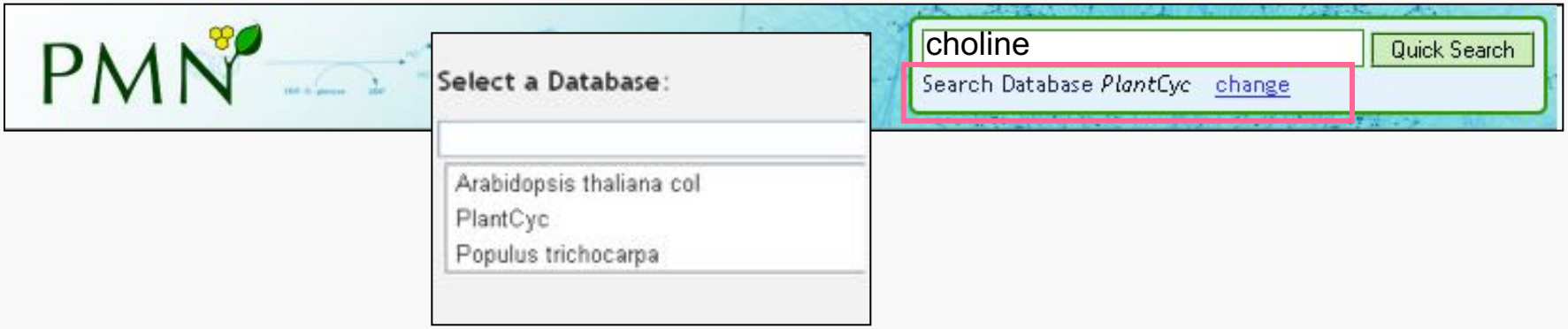
	PlantCyc 4.0	AraCyc 7.0	PoplarCyc 2.0
Pathways	**685**	**369**	**288**
Enzymes	**11058**	**5506**	**3420**
Reactions	**2929**	**2418**	**1707**
Compounds	**2966**	**2719**	**1397**
Organisms	**343**	**1**	**1***

  - Upgraded Pathway Tools software (version 12.5 -> version 14.0)

Searching in PMN databases

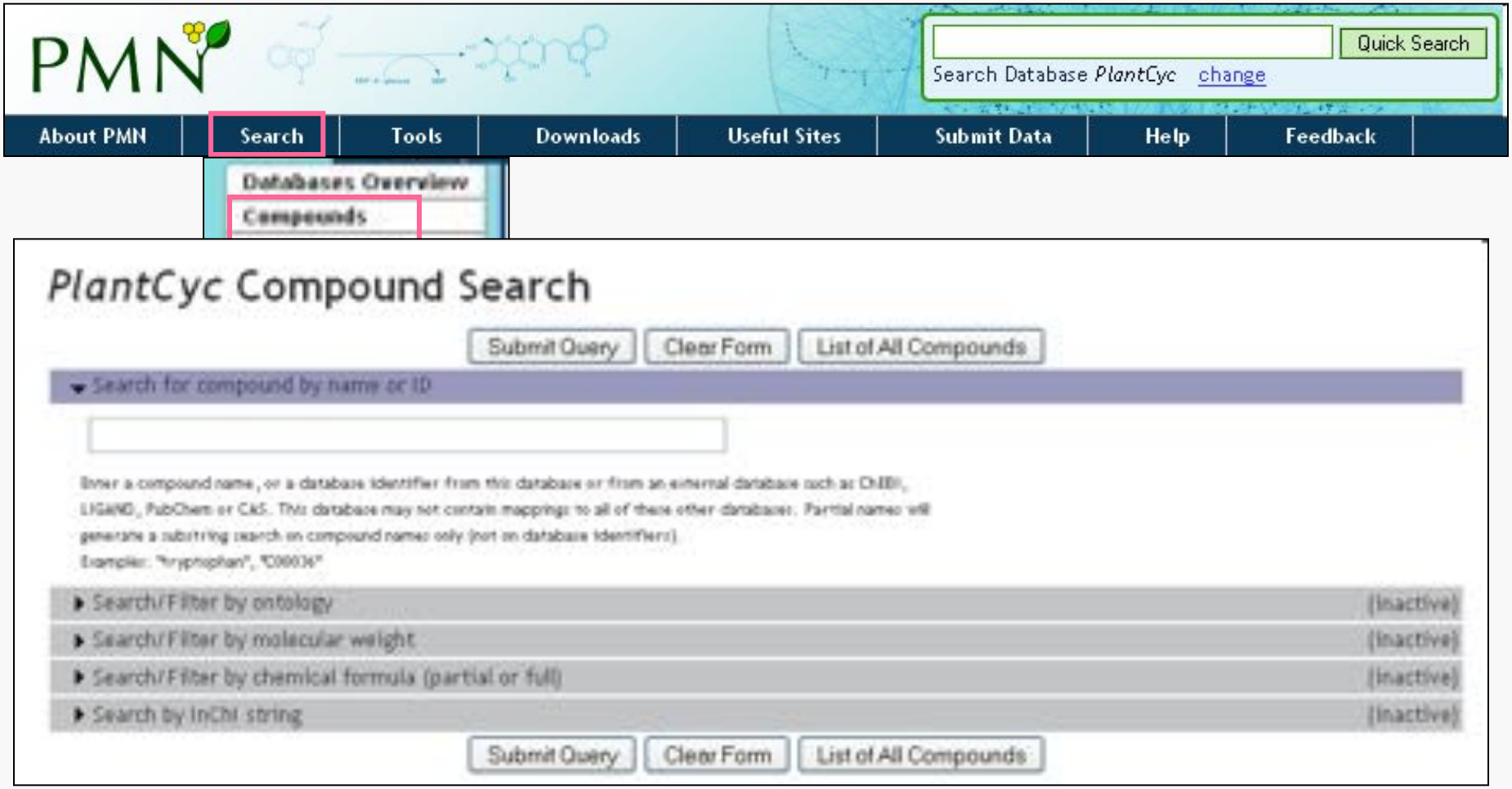
- PMN quick search bar

- Pathway Tools quick search bar



Searching in PMN databases

Specific search pages



Additional search options

Advanced Query Results

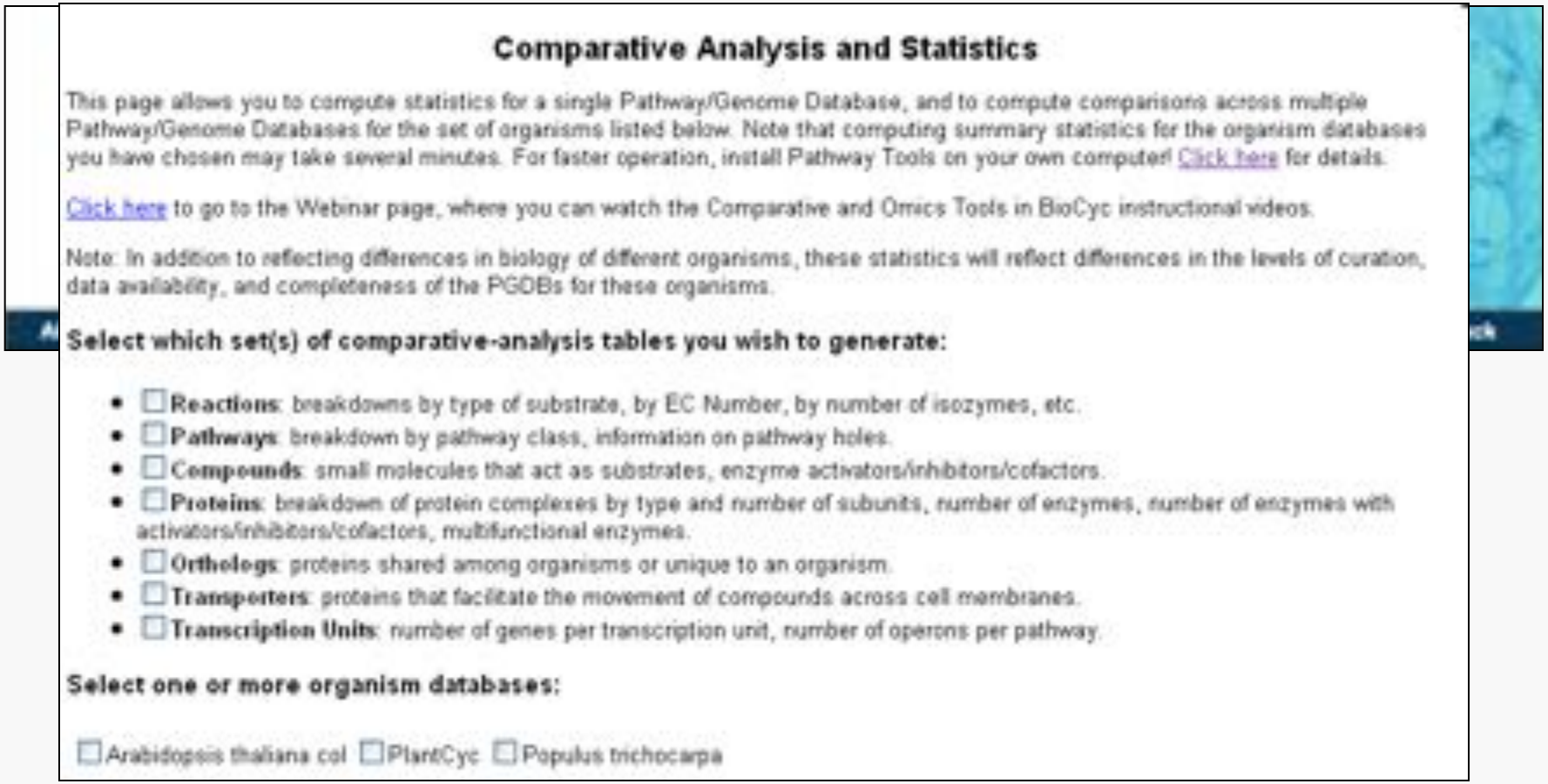
Your query in Ptools is html-sort-ascending([(i1^'NAME, i1^'REACTION-LIST, [e^'EC-NUMBER= e <- i2]) : i1<-APA^^Pathways, i2 := [i2 : i2 <- i1^'REACTION-LIST, ((0 < #[i3 : i3 <- i2^LEFT, ("C00" instringp:i i3^CHEMICAL-FORMULA)]) ; ("2.4.1." instringp:i i2^EC-NUMBER))], 0 < #i2],1) (click the query to edit it in the Free Form Advanced Query Page)

This query resulted in a single table of 64 rows.

NAME ▲ ▽	Reaction-List △ ▽	[(e ^? EC-NUMBER) : e <- i2] △ ▽
abscisic acid glucose ester biosynthesis	(+)-abscisate + UDP-D-glucose = abscisic acid glucose ester + UDP	2.4.1.-
ajugose biosynthesis I (galactinol-dependent)	verbascose + galactinol = ajugose + *myo*-inositol, stachyose + galactinol = verbascose + *myo*-inositol	2.4.1.- 2.4.1.-
ajugose biosynthesis II (galactinol-independent)	2 raffinose = stachyose + sucrose, 2 verbascose = ajugose + stachyose, 2 stachyose = verbascose + raffinose	2.4.1.- 2.4.1.- 2.4.1.-
anthocyanin biosynthesis (pelargonidin 3-O-glucoside, cyanidin 3-O-glucoside)	leucocyanidin + α-ketoglutarate + O_2 = a dihydroquercetin + succinate + CO_2 + H_2O, UDP-D-glucose + cyanidin = UDP + cyanidin-3-O-β-D-glucoside, leucocyanidin + α-ketoglutarate + O_2 = cyanidin + succinate + CO_2 + 2 H_2O + H^+, UDP-D-glucose + pelargonidin = UDP + pelargonidin-3-O-β-D-glucoside + H^+, leucopelargonidin + α-ketoglutarate + O_2 + H^+ = pelargonidin + succinate + CO_2 + 2 H_2O	2.4.1.115 2.4.1.115
arabidiol biosynthesis	(S)-2,3-Epoxysqualene + H_2O = arabidiol	4.2.1.-
baruol biosynthesis	(S)-2,3-Epoxysqualene = baruol	5.4.99.-
benzoyl-CoA biosynthesis	3-keto-3-phenylpropionyl-CoA + coenzyme A = benzoyl-CoA + acetyl-CoA, (*E*)-cinnamoyl-CoA + H_2O = 3-hydroxy-3-phenylpropionyl-CoA, 3-hydroxy-3-phenylpropionyl-CoA + $NADP^+$ = 3-keto-3-phenylpropionyl-CoA + NADPH + H^+	
cellulose biosynthesis	a long linear glucan + H_2O = a short glucan + a short glucan, UDP-D-glucose + $(1,4\text{-}\beta\text{-D-glucosyl})_{n1}$ = UDP + $(1,4\text{-}\beta\text{-D-glucosyl})_{(n+1)}$	2.4.1.12
cholesterol biosynthesis I	lanosterol biosynthesis, 4α-carboxy-4β-methyl-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ = 4α-methyl-5α-cholesta-8,24-dien-3-one + NAD(P)H + CO_2, lathosterol + NAD(P)H + O_2 = 7-dehydro-cholesterol + $NAD(P)^+$ + 2 H_2O, 4,4-dimethyl-5-α-cholesta-8,14,24-trien-3-β-ol + NADPH + H^+ = 4,4-dimethyl-5α-cholesta-8,24-dien-3-β-ol + $NADP^+$, epoxysqualene biosynthesis, cholesterol + $NADP^+$ = 7-dehydro-cholesterol + NADPH + H^+, 5α-cholesta-7,24-dien-3β-ol + NADPH + H^+ = lathosterol + $NADP^+$, zymosterol = 5α-cholesta-7,24-dien-3β-ol, 5α-cholesta-8,24-dien-3-one + NADPH + H^+ = zymosterol + $NADP^+$, 4α-carboxy-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ = 5α-cholesta-8,24-dien-3-one + NAD(P)H + CO_2, 4α-formyl-5α-cholesta-8,24-dien-3β-ol + NAD(P)H + O_2 = 4α-carboxy-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ + H_2O, 4α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NAD(P)H + O_2 = 4α-formyl-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ + 2 H_2O, 4-α-methyl zymosterol + NAD(P)H + O_2 = 4α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ + H_2O, 4α-methyl-5α-cholesta-8,24-dien-3-one + NADPH + H^+ = 4-α-methyl zymosterol + $NADP^+$, 4α-formyl-4β-methyl-5α-cholesta-8,24-dien-3β-ol + NAD(P)H + O_2 = 4α-carboxy-4β-methyl-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ + H_2O,	1.14.13.70 1.14.13.70 1.14.13.70

Comparing across species

- Use general Comparative Analyses tools



Comparative Analysis and Statistics

This page allows you to compute statistics for a single Pathway/Genome Database, and to compute comparisons across multiple Pathway/Genome Databases for the set of organisms listed below. Note that computing summary statistics for the organism databases you have chosen may take several minutes. For faster operation, install Pathway Tools on your own computer! Click here for details.

Click here to go to the Webinar page, where you can watch the Comparative and Omics Tools in BioCyc instructional videos.

Note: In addition to reflecting differences in biology of different organisms, these statistics will reflect differences in the levels of curation, data availability, and completeness of the PGDBs for these organisms.

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

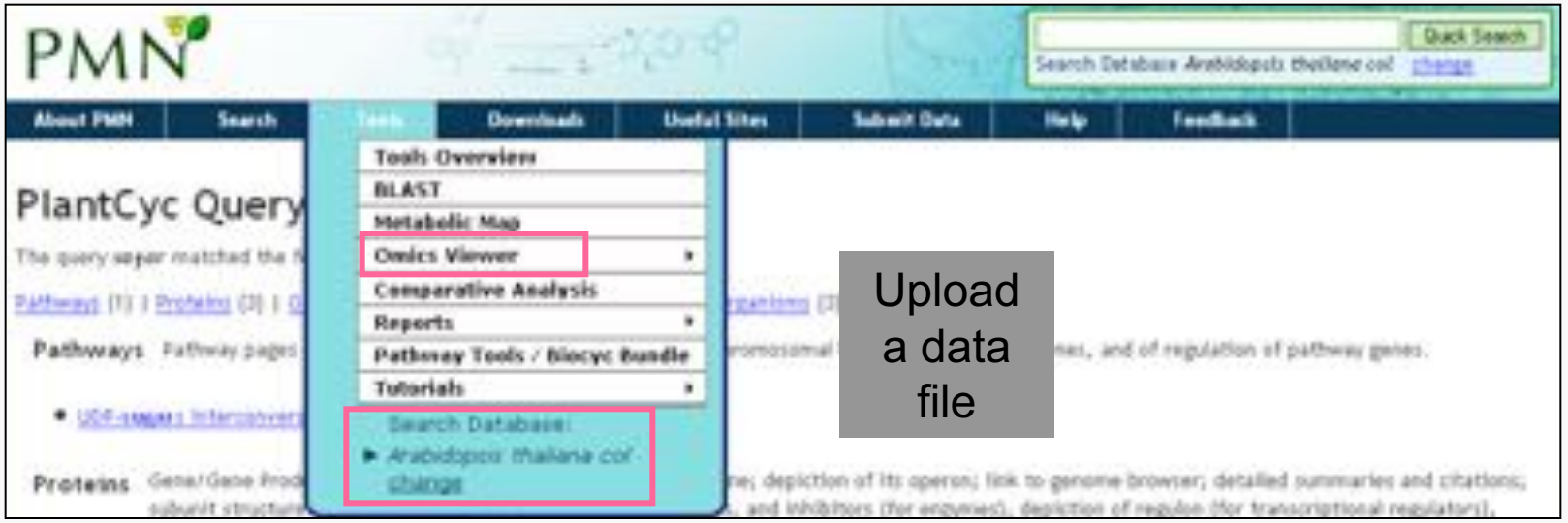
- ☐ **Reactions**: breakdowns by type of substrate, by EC Number, by number of isozymes, etc.
- ☐ **Pathways**: breakdown by pathway class, information on pathway holes.
- ☐ **Compounds**: small molecules that act as substrates, enzyme activators/inhibitors/cofactors.
- ☐ **Proteins**: breakdown of protein complexes by type and number of subunits, number of enzymes, number of enzymes with activators/inhibitors/cofactors, multifunctional 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

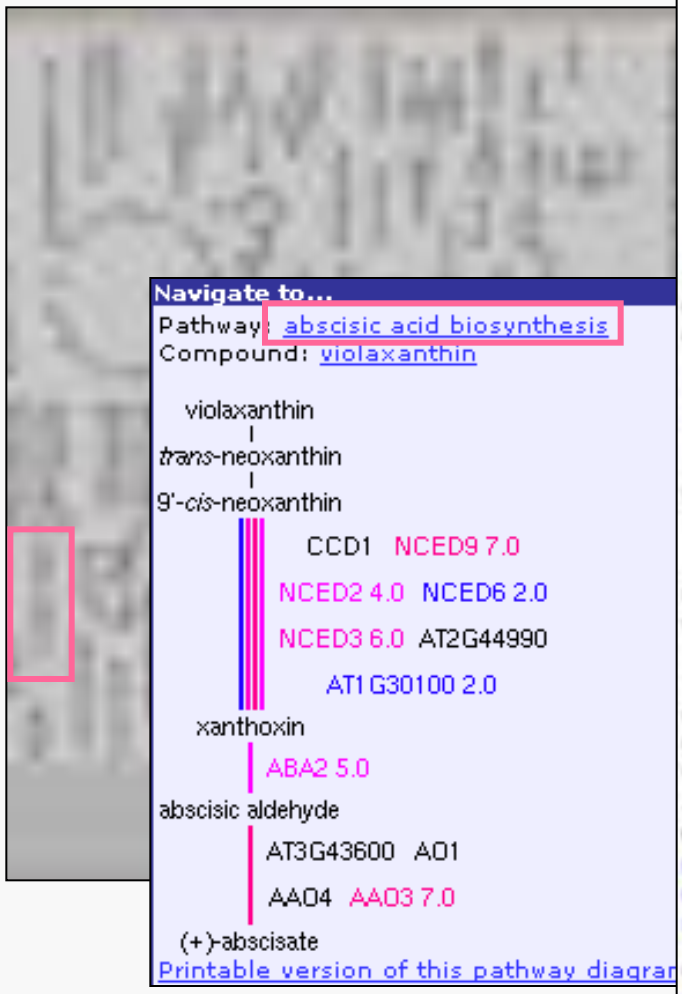
Visualizing OMICs data

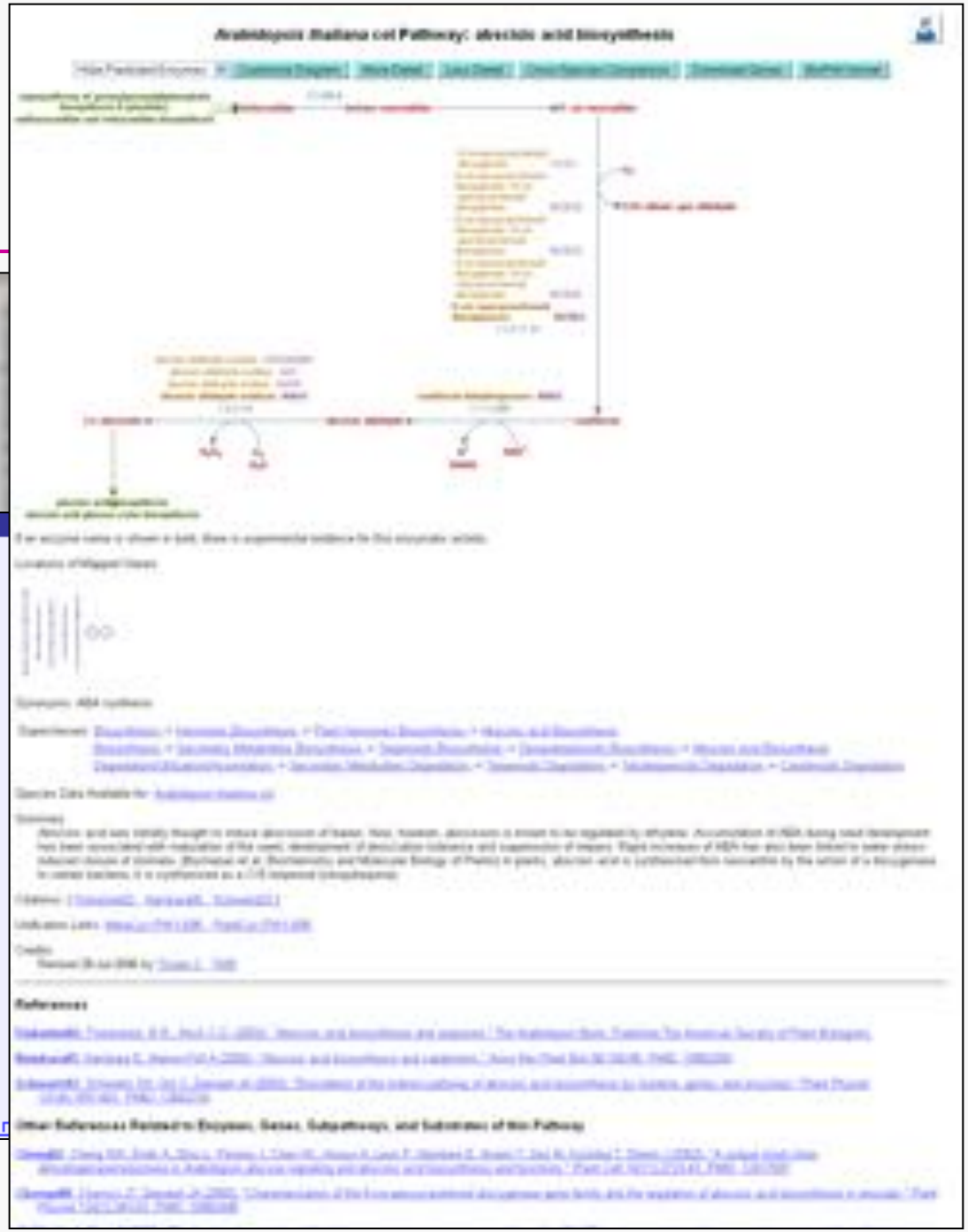
- Overlay “pre-cleaned” data sets on a metabolic map
  - Gene transcription data
  - Proteomic data
  - Metabolomic data
- Only available for single-species databases, not PlantCyc
- Demonstrations available at conference . . . please see me!



Visualizing OMICs data

Visualizing OMICs data



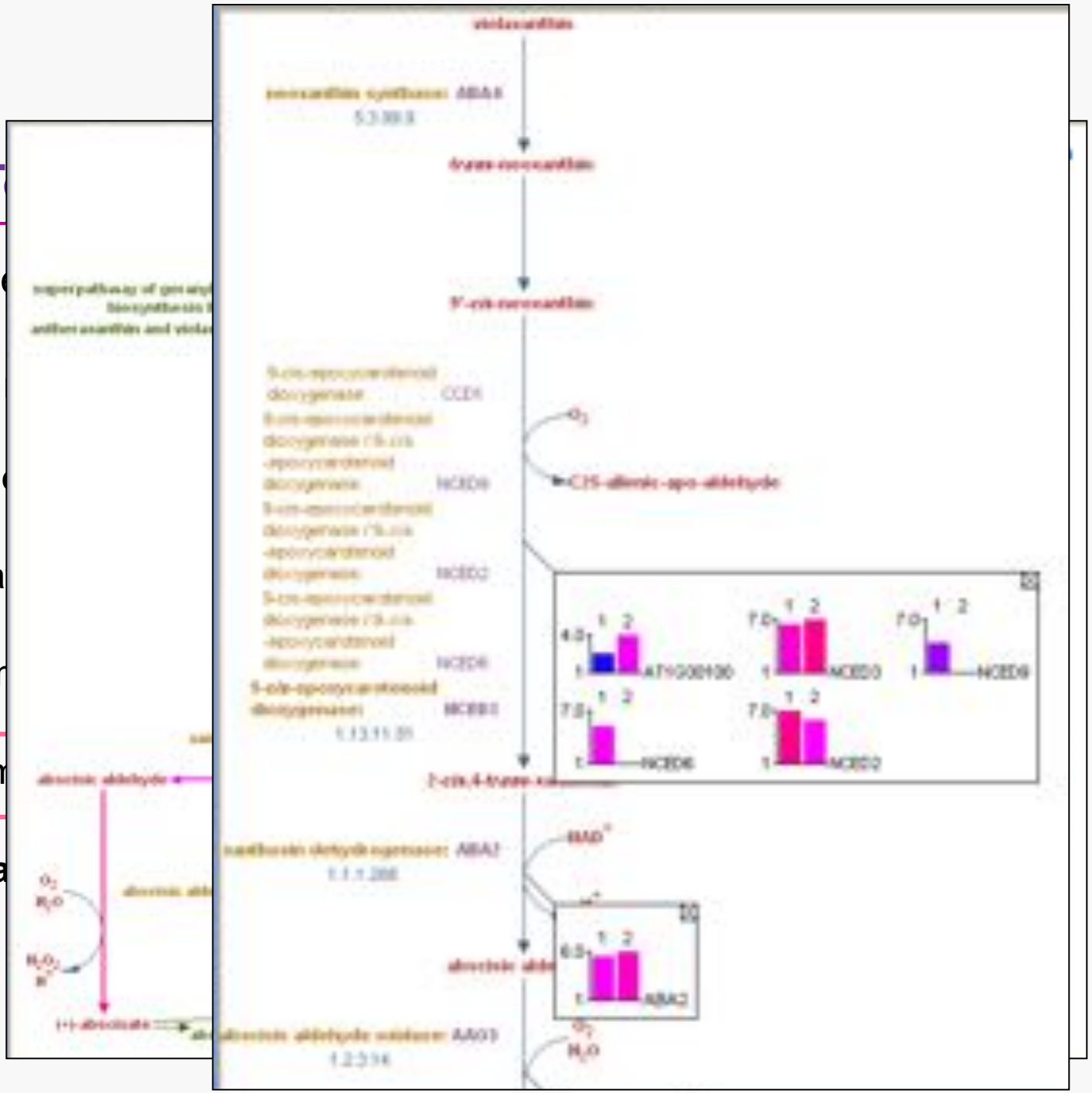


Data and software downloads

- Install a local copy of the Pathway Tools software

Pathway T

- Desktop v
  - Create
  - Modify
  - Genera
  - Perform
  - See On
- **Demonstra**



PMN data

- The PMN houses a large quantity of data, but . . .
  - Numerous previously identified enzymes and pathways are missing
  - Exciting new discoveries are continuously made
  - Plant genome sequence data is becoming available at a rapid rate
  - How do we bring this information in?

Database

- Manual curatio
  - Curators rea
  - Researchers
- Computational
  - Automated p
    - Predicted
    - Predicted
    - *Predict a*
  - Manual valid
    - Remove
    - Create or
    - Add refer

al articles

matic functions

nways in PlantCyc

d enzymes for species X

You!

Building better databases together

- To submit data, report an error, or volunteer to help validate . . .
  - Send an e-mail: curator@plantcyc.org
  - Use data submission “tools”

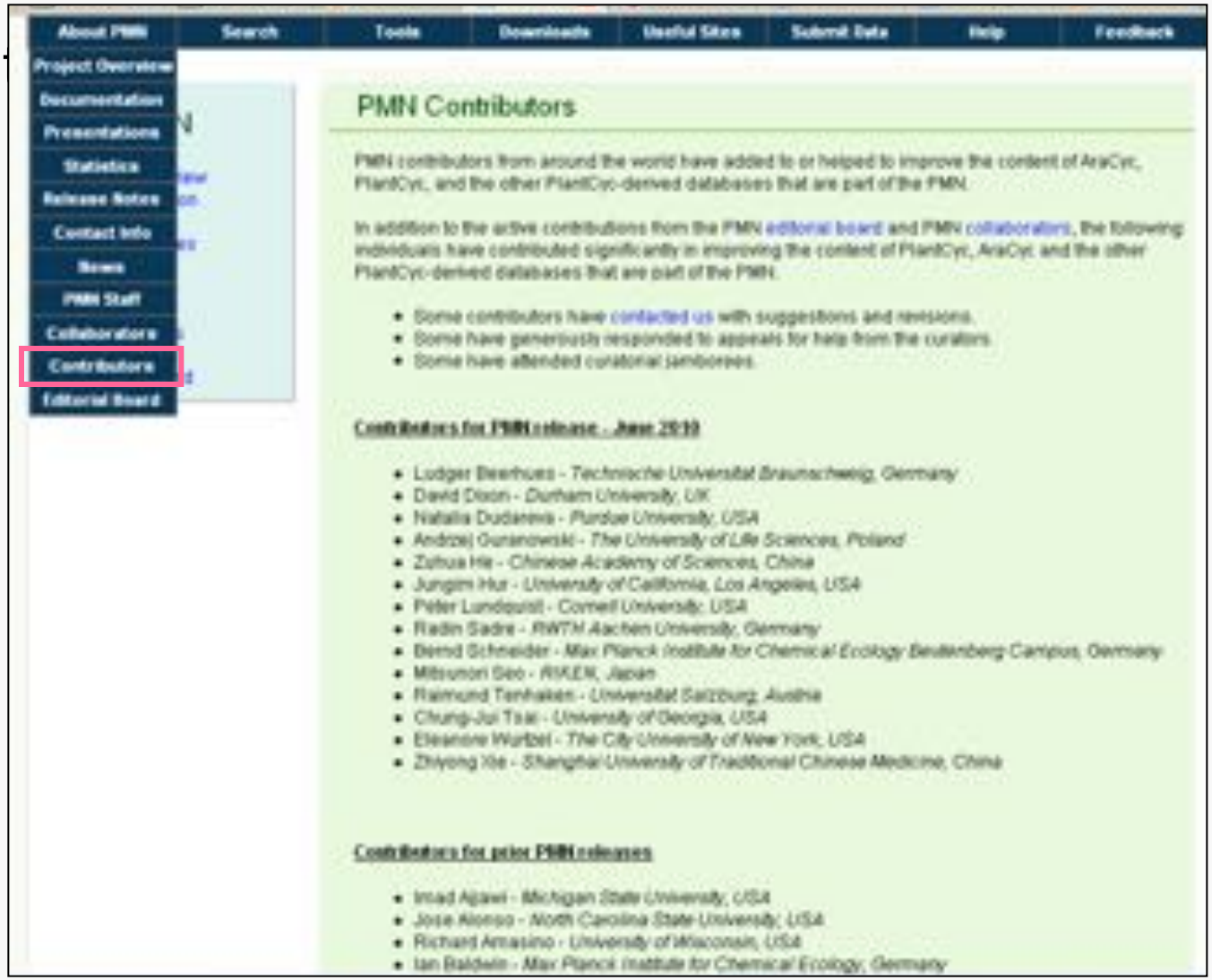
  - **Meet with me individually at this conference**
    - P2,4,12,14,18,20,22,24,26,30,34,36,38,40,44,48,50,54,56,62,74,84,92,9,21,25, 29,31,33,39,41,43,5157,59,63,65,75,81,87,89,91,93
    - CS30,15,32,27,28,13,14,7,4,12, 31, 35, 18, 36, 21, 1, 16, 11, 6,
    - PS IVb, Vb, VIa, VIIa, VIIIa, VIIIb

Building better databases together

- **Details are very, very welcome!!**
  - Reactions:
    - All co-factors, co-substrates, etc.
    - EC suggestions – partial or full
  - Compounds
    - Structure – visual representation / compound file (e.g. mol file)
    - Synonyms
    - Unique IDs (e.g. ChEBI, CAS, KEGG)
  - Enzymes
    - Unique IDs (e.g. At2g46480, UniProt, Genbank)
    - Specific reactions catalyzed

Community gratitude

- We



Plant metabolic NETWORKING

- Please use our data
- Please use our tools
- Please help us to improve our databases!
- Please contact us if we can be of any help!

PMN
Plant Metabolic Network

AraCyc

curator@plantcyc.org

www.plantcyc.org

special pre-release sneak preview:
pmn.plantcyc.org

gements

Current Tech Team Members:
- Bob Muller (*Manager*)
- Larry Ploetz (*Sys. Administrator*)
- Cynthia Lee
- Shanker Singh
- Chris Wilks

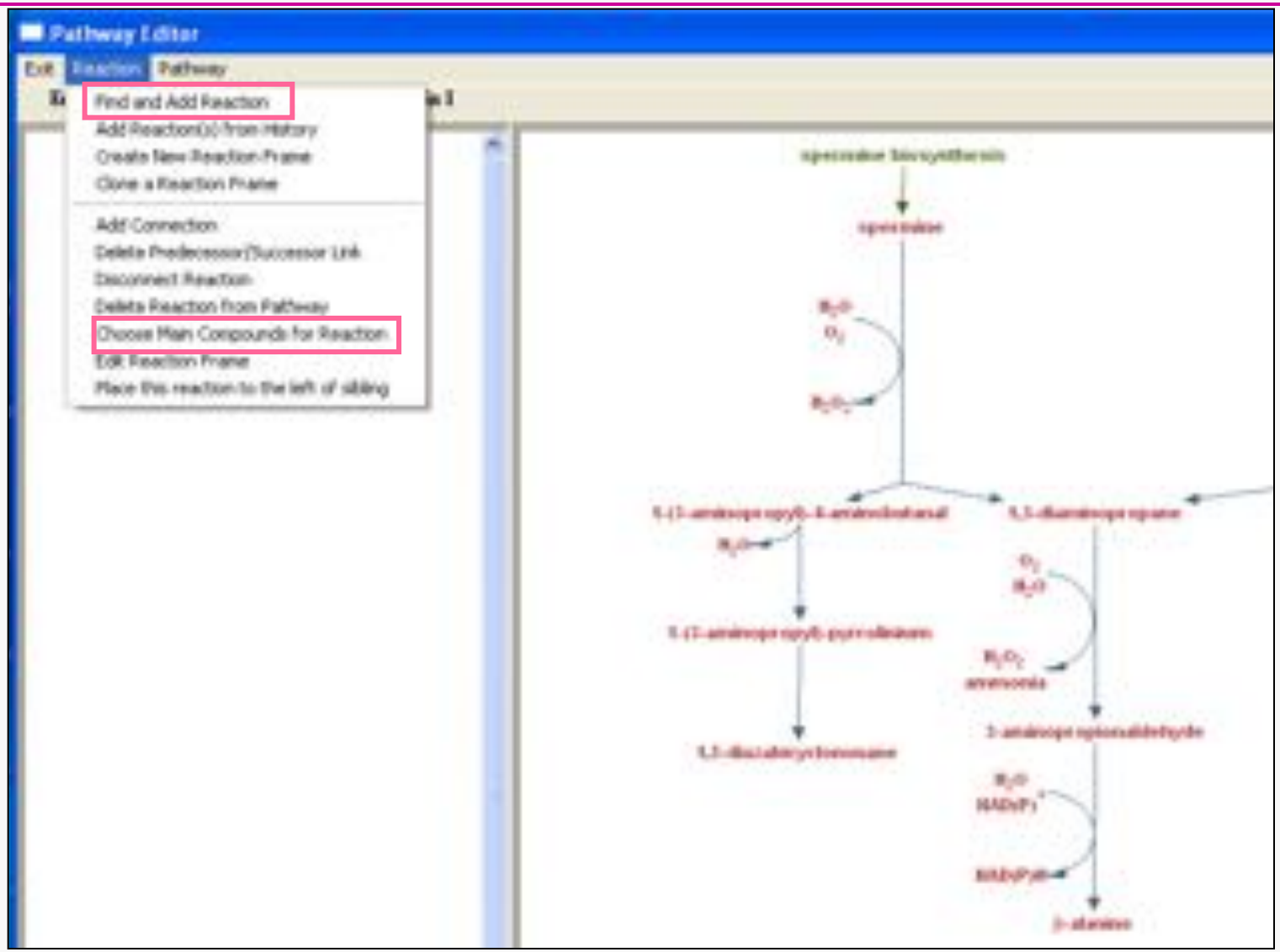
Tech Team: recent alumni
- Raymond Chetty
- Anjo Chi
- Vanessa Kirkup
- Tom Meyer

Collaborators:
- Peter Karp (SRI)
- Ron Caspi (SRI)
- Suzanne Paley (SRI)
- SRI Tech Team
- Lukas Mueller (SGN)
- Anuradha Pujar (SGN)
- Gramene and MedicCyc

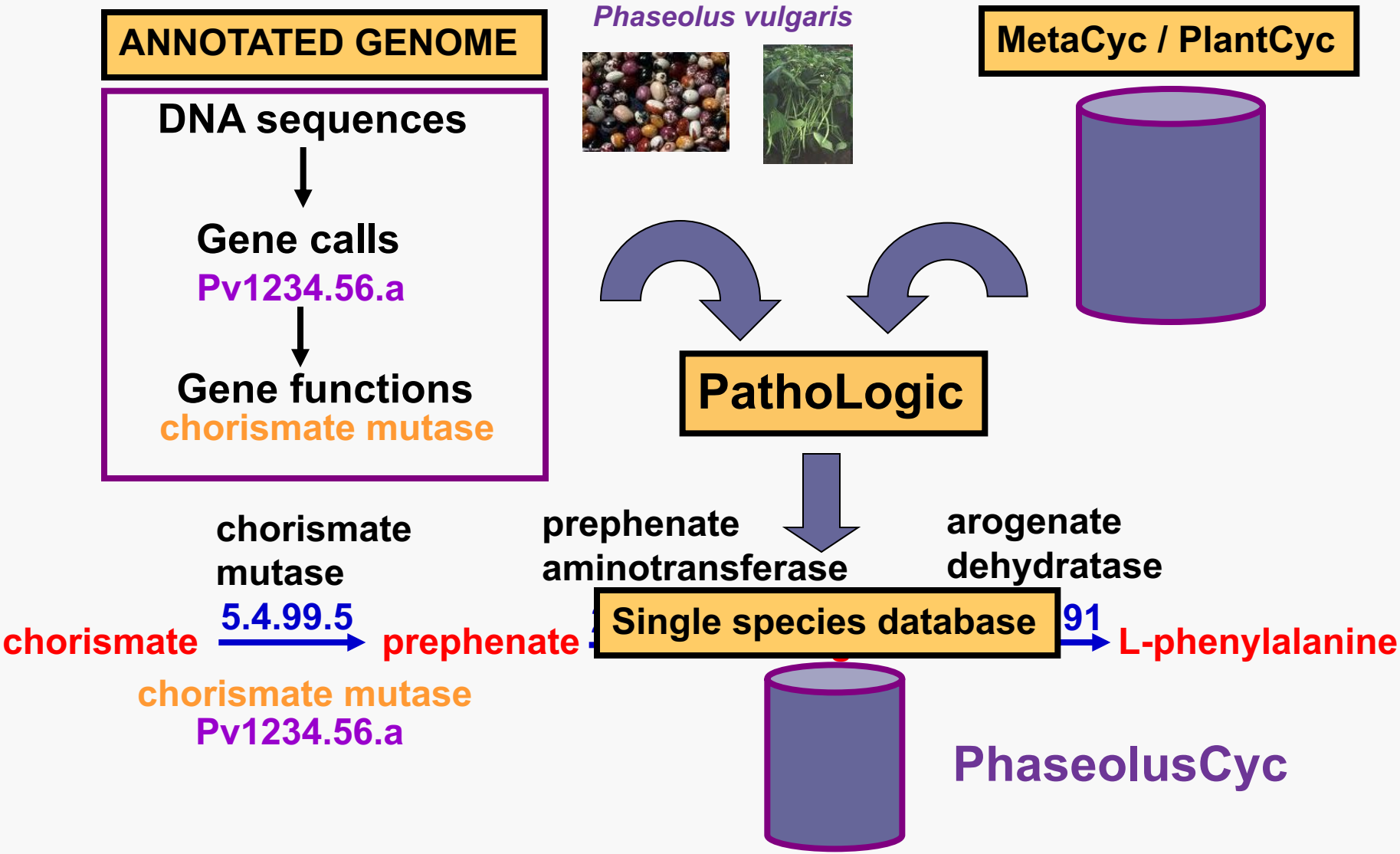
National Science Foundation
WHERE DISCOVERIES BEGIN

CARNEGIE
INSTITUTION FOR
SCIENCE

Data and software downloads



How does the PMN acquire new data?



Additional search o

How can you put the PMN to work for you?

- Learn background information about particular metabolic pathways
- Create customized metabolic data sets
- Compare metabolism across plant species
- Analyze experimental OMICs data in a metabolic context
- Manipulate and study data offline
- Create new metabolic pathway databases

Comm

Tools | Downloads | Useful Sites | Submit Data | Help | Feedback

Data Submission

Please help us expand the content of the PMN databases, including AraCyc and PlantCyc!

- We welcome any and all **new** data submissions related to plant biochemical pathways.
- You can also **correct** an existing pathway
  - *Not sure what types of information to include?*
    Try using one of our **data submission / correction forms** below.
- Or, just use our **Feedback Form** or **send an e-mail** to curator@plantcyc.org and append your data as an attachment

Data submission / correction forms

**** You do not need to fill in all the columns. (See our tutorial for help)

We will be happy to accept whatever information you can easily provide . . .

Compound form

Enzyme / Reaction form

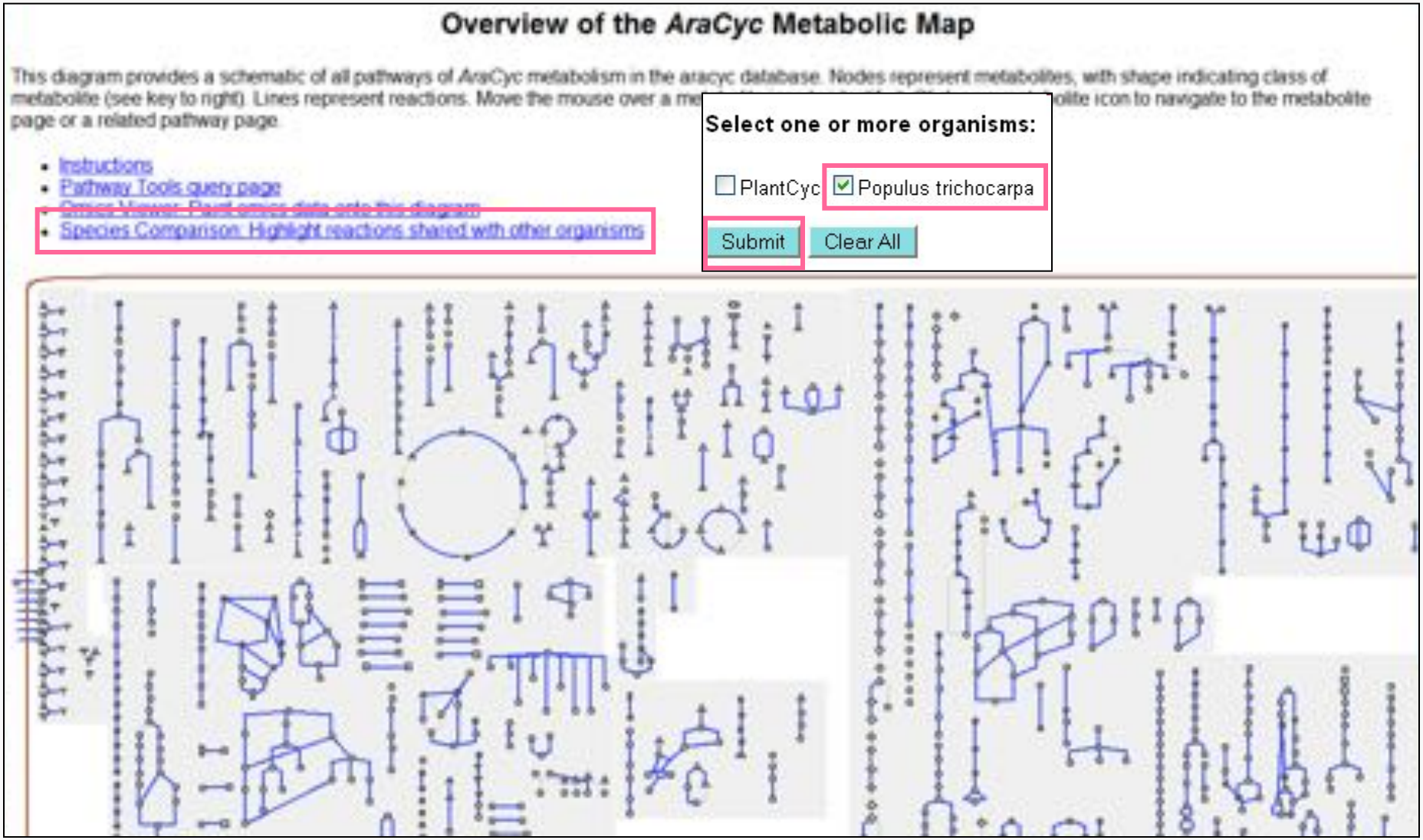
Pathway form

Please fill-in any information you would like to share on the appropriate forms above and send it as an attachment using our Feedback Form or send it as an e-mail attachment to: curator@plantcyc.org

We also welcome . . .

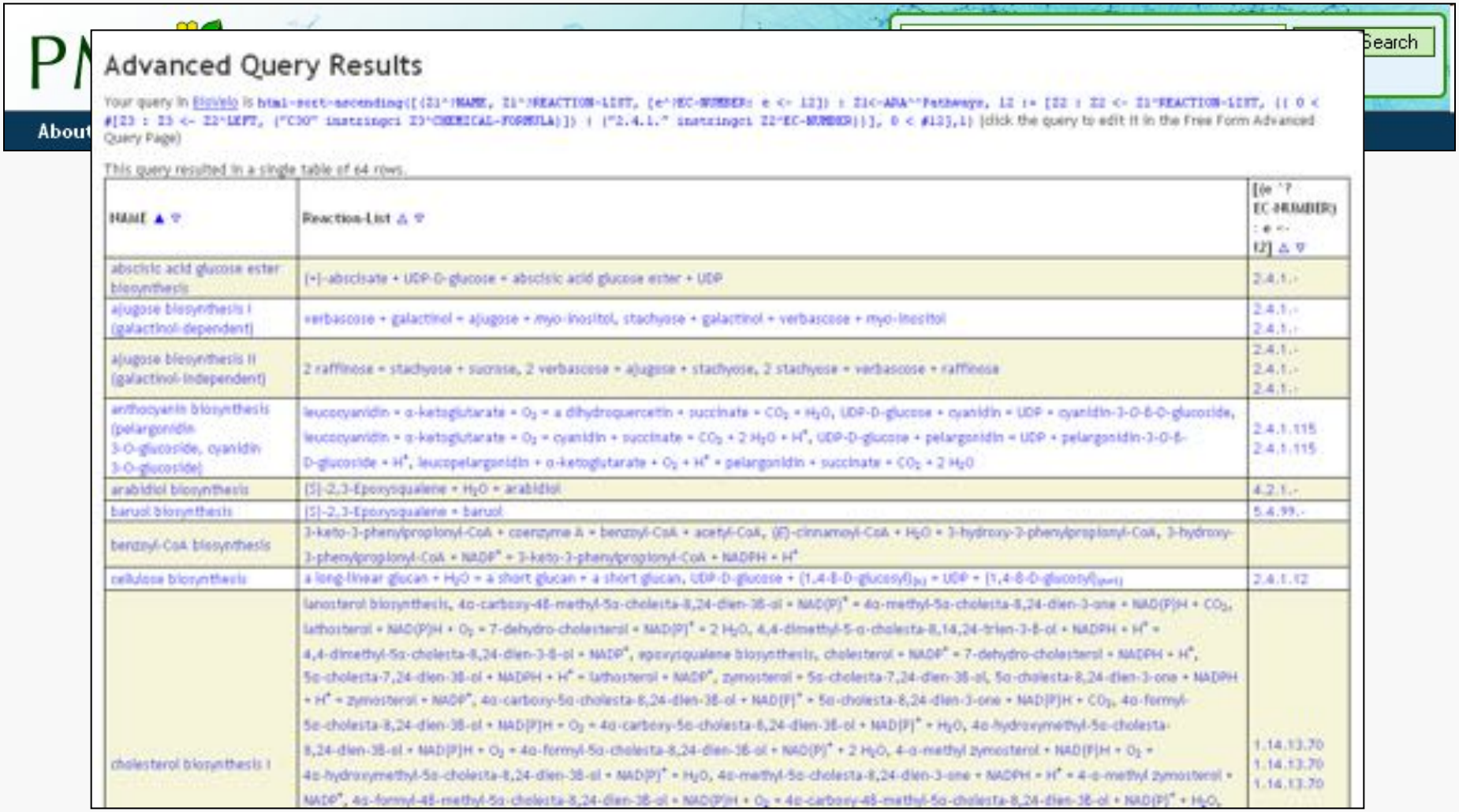
- diagrams and images (.gif, .jpg, .pdf, .ppt, .psd, etc.)

Comparing across species



Comparing across species

Additional search options



Search

About

Advanced Query Results

Your query in Ruleln is html-sect-ascending([(?1^?NAME, ?1^?REACTION-LIST, [e^?EC-NUMBER: e <- l2]) : ?1<-AHA^^Pathways, l2 := [?2 : ?2 <- ?1^?REACTION-LIST, ((0 < #[?3 : ?3 <- ?2^LEFT, ("C30" instringpci ?3^CHEMICAL-FORMULA)]) : ("2.4.1." instringpci ?2^EC-NUMBER))], 0 < #l2],1) (click the query to edit it in the Free Form Advanced Query Page)

This query resulted in a single table of 64 row(s).

NAME ▲ ▽	Reaction-List △ ▽	[(e ^? EC-NUMBER) : e <- l2] △ ▽
abscisic acid glucose ester biosynthesis	(+)-abscisate + UDP-D-glucose = abscisic acid glucose ester + UDP	2.4.1.-
ajugose biosynthesis I (galactinol-dependent)	verbascose + galactinol = ajugose + myo-inositol, stachyose + galactinol = verbascose + myo-inositol	2.4.1.- 2.4.1.-
ajugose biosynthesis II (galactinol-independent)	2 raffinose = stachyose + sucrose, 2 verbascose = ajugose + stachyose, 2 stachyose = verbascose + raffinose	2.4.1.- 2.4.1.- 2.4.1.-
anthocyanin biosynthesis (pelargonidin 3-O-glucoside, cyanidin 3-O-glucoside)	leucocyanidin + α-ketoglutarate + O_2 = a dihydroquercetin + succinate + CO_2 + H_2O, UDP-D-glucose + cyanidin = UDP + cyanidin-3-O-β-D-glucoside, leucocyanidin + α-ketoglutarate + O_2 = cyanidin + succinate + CO_2 + 2 H_2O + H^+, UDP-D-glucose + pelargonidin = UDP + pelargonidin-3-O-β-D-glucoside + H^+, leucopelargonidin + α-ketoglutarate + O_2 + H^+ = pelargonidin + succinate + CO_2 + 2 H_2O	2.4.1.115 2.4.1.115
arabidiol biosynthesis	(S)-2,3-Epoxysqualene + H_2O = arabidiol	4.2.1.-
baruol biosynthesis	(S)-2,3-Epoxysqualene = baruol	5.4.99.-
benzoyl-CoA biosynthesis	3-keto-3-phenylpropionyl-CoA + coenzyme A = benzoyl-CoA + acetyl-CoA, (E)-cinnamoyl-CoA + H_2O = 3-hydroxy-3-phenylpropionyl-CoA, 3-hydroxy-3-phenylpropionyl-CoA + $NADP^+$ = 3-keto-3-phenylpropionyl-CoA + NADPH + H^+	
cellulose biosynthesis	a long-linear glucan + H_2O = a short glucan + a short glucan, UDP-D-glucose + $(1,4\text{-}\beta\text{-D-glucosyl})_{n1}$ = UDP + $(1,4\text{-}\beta\text{-D-glucosyl})_{(n+1)}$	2.4.1.12
cholesterol biosynthesis I	lanosterol biosynthesis, 4α-carboxy-4β-methyl-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ = 4α-methyl-5α-cholesta-8,24-dien-3-one + NAD(P)H + CO_2, lathosterol + NAD(P)H + O_2 = 7-dehydro-cholesterol + $NAD(P)^+$ + 2 H_2O, 4,4-dimethyl-5-α-cholesta-8,14,24-trien-3-β-ol + NADPH + H^+ = 4,4-dimethyl-5α-cholesta-8,24-dien-3-β-ol + $NADP^+$, epoxysqualene biosynthesis, cholesterol + $NADP^+$ = 7-dehydro-cholesterol + NADPH + H^+, 5α-cholesta-7,24-dien-3β-ol + NADPH + H^+ = lathosterol + $NADP^+$, zymosterol + 5α-cholesta-7,24-dien-3β-ol, 5α-cholesta-8,24-dien-3-one + NADPH + H^+ = zymosterol + $NADP^+$, 4α-carboxy-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ = 5α-cholesta-8,24-dien-3-one + NAD(P)H + CO_2, 4α-formyl-5α-cholesta-8,24-dien-3β-ol + NAD(P)H + O_2 = 4α-carboxy-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ + H_2O, 4α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NAD(P)H + O_2 = 4α-formyl-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ + 2 H_2O, 4-α-methyl zymosterol + NAD(P)H + O_2 = 4α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ + H_2O, 4α-methyl-5α-cholesta-8,24-dien-3-one + NADPH + H^+ = 4-α-methyl zymosterol + $NADP^+$, 4α-formyl-4β-methyl-5α-cholesta-8,24-dien-3β-ol + NAD(P)H + O_2 = 4α-carboxy-4β-methyl-5α-cholesta-8,24-dien-3β-ol + $NAD(P)^+$ + H_2O,	1.14.13.70 1.14.13.70 1.14.13.70

Data and software downloads

- Get pathway data sets from pathway pages