Building Better Biological Databases Together: Part 3:

Accessing the Data You Need at the Plant Metabolic Network

kate dreher

curator

PMN/TAIR

The Carnegie Institution for Science

Stanford, CA

Accessing the data you need at the PMN

- Introduction
- Data Overview
- Basic and Advanced Searching
- New and Improved Features
- Desktop Capabilities
- Acknowledgments

Plant metabolism

- Plants provide crucial benefits to the ecosystem and humanity
- A better understanding of plant metabolism may contribute to:
  - More nutritious foods
  - New medicines
  - More pest-resistant plants
  - Higher photosynthetic capacity and yield in crops
  - Better biofuel feedstocks
  - Improved industrial inputs (e.g. oils, fibers, etc.)
  - . . . many more applications
- These efforts require access to high quality plant metabolism data

Plant Metabolic Network goals

- Capture and organize published data to create metabolic pathways
- Facilitate data analysis
- Make proteome-based metabolic pathway predictions for **many** plant species
- Support research and education
- Provide public resources : **www.plantcyc.org**
  - **PlantCyc**
  - **AraCyc**
  - **PoplarCyc**
  - **more on the way**


PMN
Plant Metabolic Network
PlantCyc
search
About PMN
Search
Tools
Downloads
Useful Sites
Submit Data
Help
Feedback
Introduction
The Plant Metabolic Network (PMN) is a collaborative project among databases and biochemists with a common goal to build a broad network of plant metabolic pathway databases. A central feature of the PMN is PlantCyc, a comprehensive plant biochemical pathway database, containing curated information from the literature and computational analyses about the genes, enzymes, compounds, reactions, and pathways involved in primary and secondary metabolism.
PMN is funded by the National Science Foundation (Grant #
and
governed by an Editorial Board composed of internationally renowned scientists, and executed at the Carnegie Institution for Science, Department of Plant Biology.
Plant Metabolic Pathway Databases
PlantCyc
PlantCyc provides access to manually curated or reviewed information about shared and unique metabolic pathways present in over
AraCyc
News
We want your data at the
Please come to a database overview and
at the
meeting co-hosted by the PMN. We want your help to get the lastest information on plant metabolism into our resources.
PMN TAIR/SGN Workshop
August 7
7:00 PM - 8:30 PM
Auditorium 1
Please e-mail us at
to let us know if you can stay after the workshop to directly submit your latest information on plant compounds, enzymes, or pathways to the PMN.
You can also learn more about all the PMN resources at our poster
The Plant Metabolic Network: PlantCyc, AraCyc, and more metabolic pathway databases for plant research

Plant Metabolic Network collaborators

- SRI International – BioCyc project
  - Created Pathway Tools software
  - Provide on-going Pathway Tools software support and development
  - Maintain and update MetaCyc (multi-kingdom) reference metabolic pathway database
- Other collaborators / contributors include:
  - Sol Genomics Network (SGN) / Boyce Thompson Institute
  - Gramene
  - Maize GDB
  - MedicCyc / Nobel Foundation
  - SoyBase
  - ChlamyCyc
  - PlantMetabolomics group
  - . . . and more

SRI International

MetaCyc
A member of the BioCyc database collection

SolCyc

BTI

The Samuel Roberts Noble Foundation

SoyBase

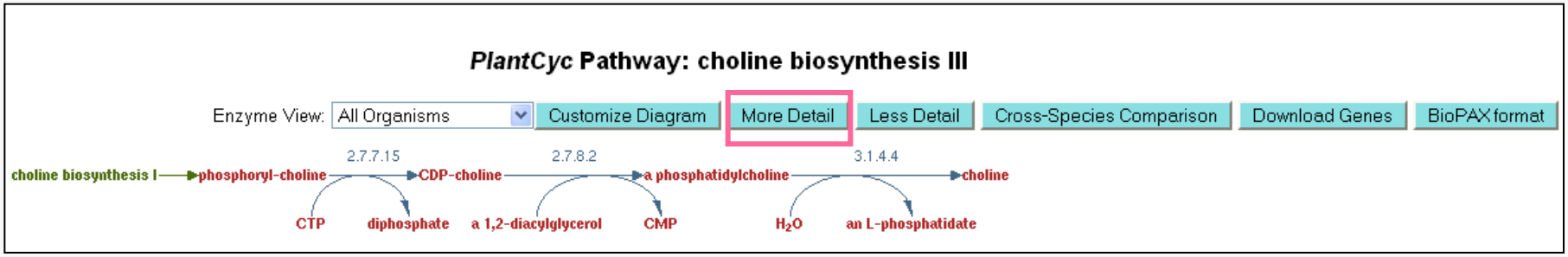
Gramene

MaizeGDB
Maize Genetics and Genomics Database

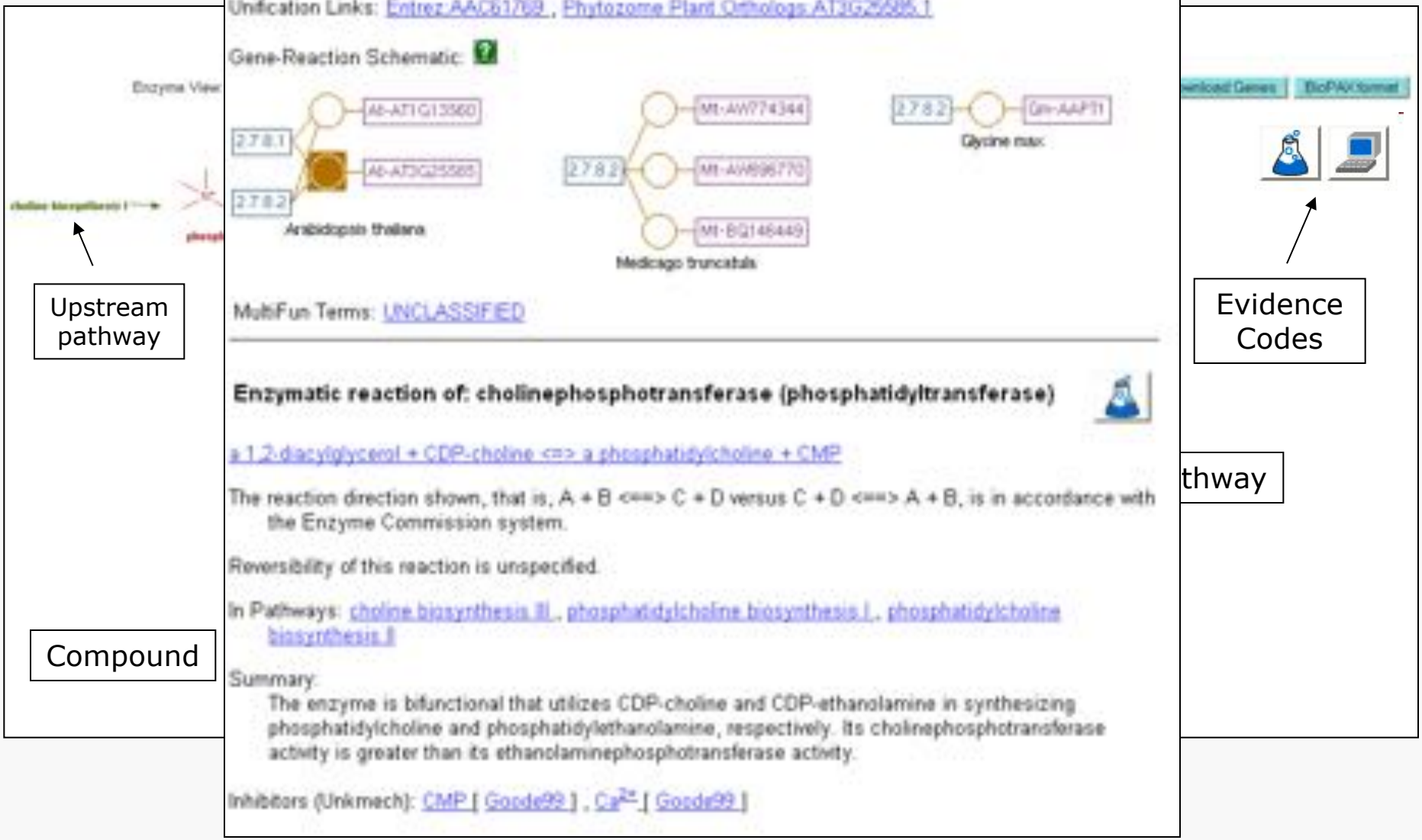
ChlamyCyc

Plant Metabolic Network data

- Pathway-based access to multiple data types



PMN pa



Upstream pathway

Compound

Evidence Codes

thway

PMN data content statistics

- Latest PMN release – April 2011

	PlantCyc 5.0	AraCyc 8.0	PoplarCyc 3.0
Pathways *	737	393	309
Enzymes **	11217	5520	3422
Reactions	3142	2525	1758
Compounds	3146	2825	1406
Organisms	366	1	1***

* These represent “base pathways” that are composed entirely of reactions. A base pathway may be joined with additional base pathways and/or reactions to create a superpathway. Superpathways are not included in this statistic.

** The term "enzyme" refers to both monomers and complexes found in the databases.

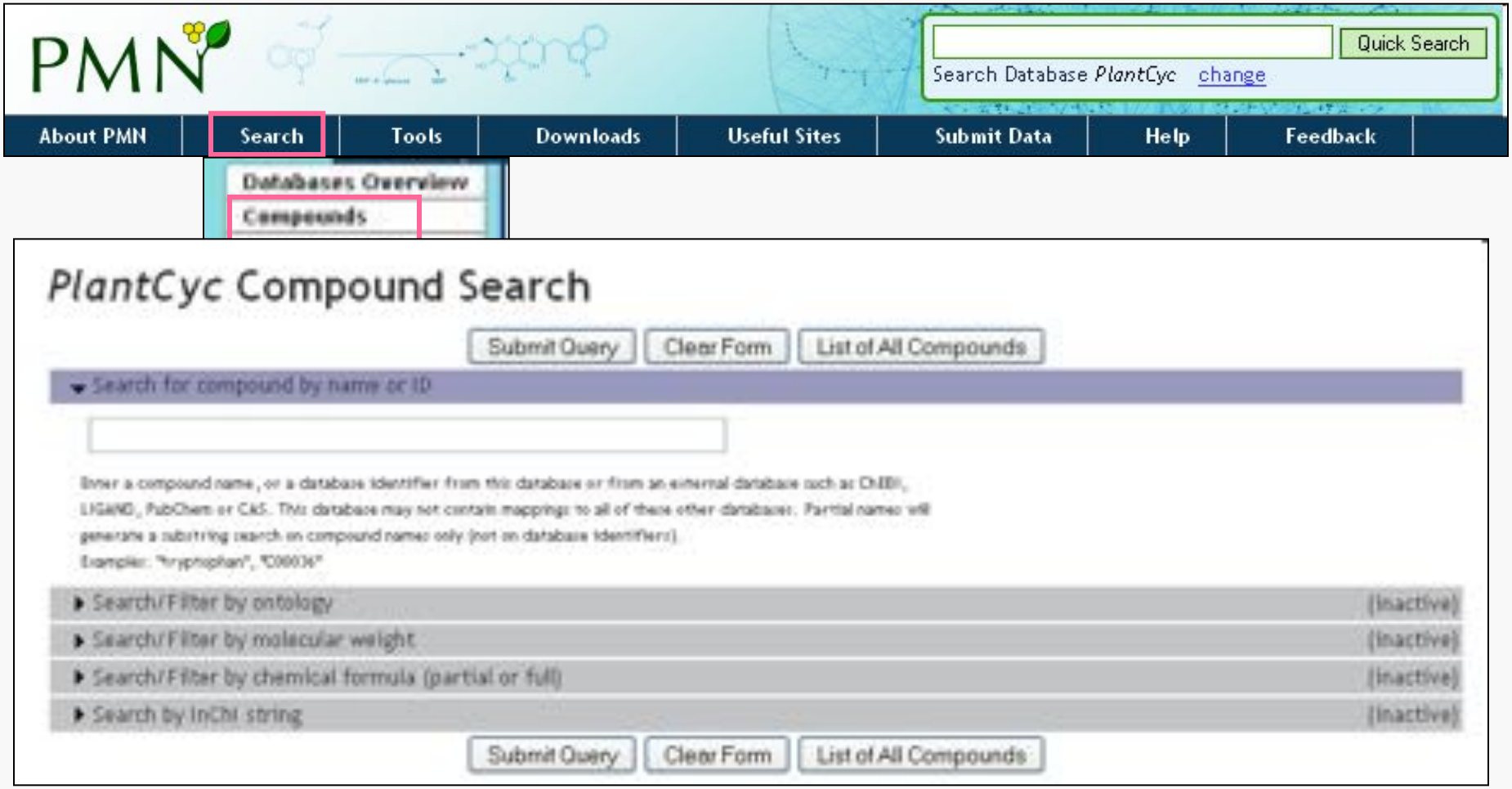
*** The vast majority of the enzymes present in the database are from *Populus trichocarpa* but experimentall supported enzymes and pathways from other species or hybrids in the *Populus* genus can be included in PoplarCyc.

Searching at the PMN

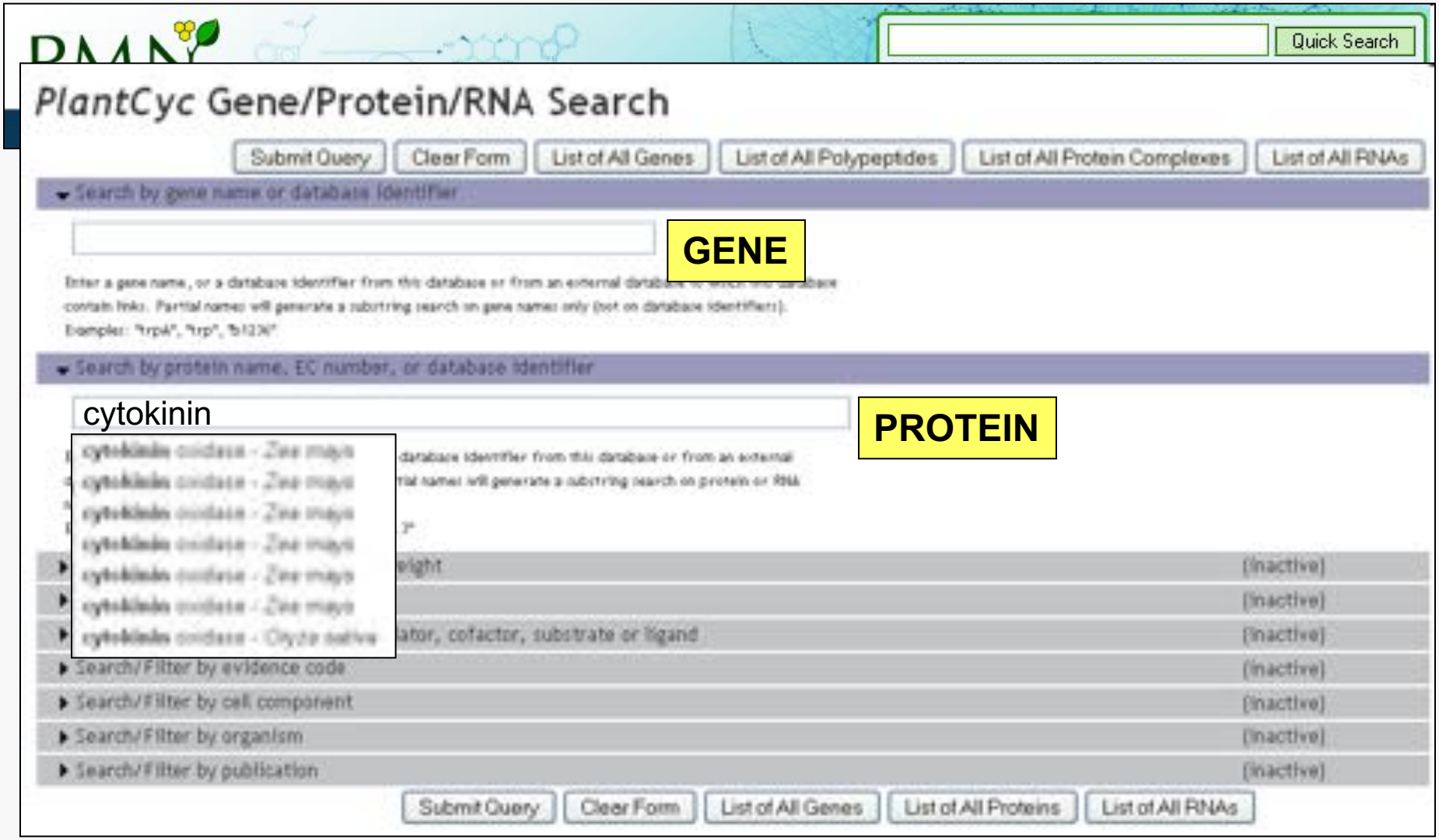
- Pathway Tools quick search bar

Quick search results

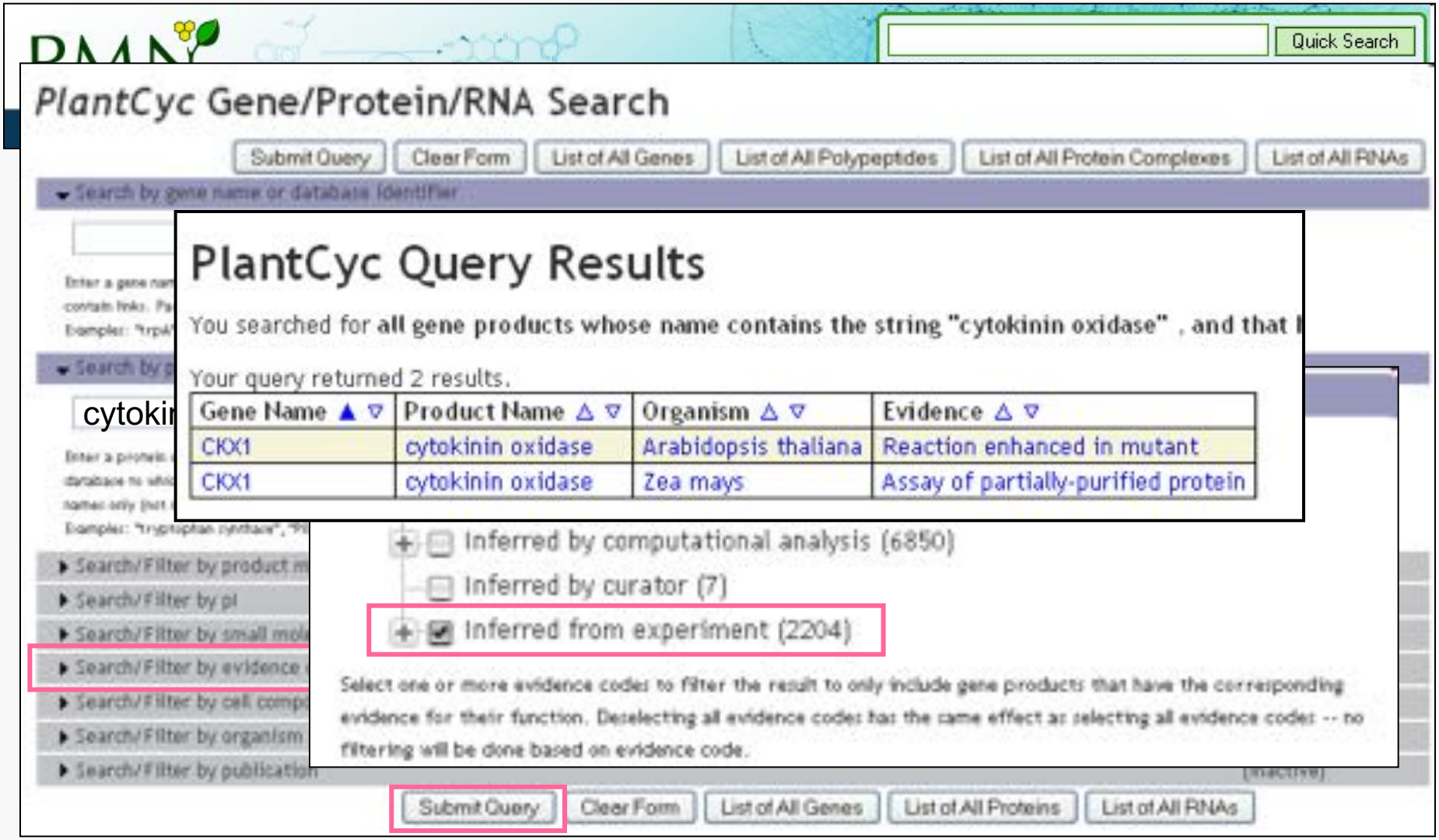
Specific search pages



Specific search pages

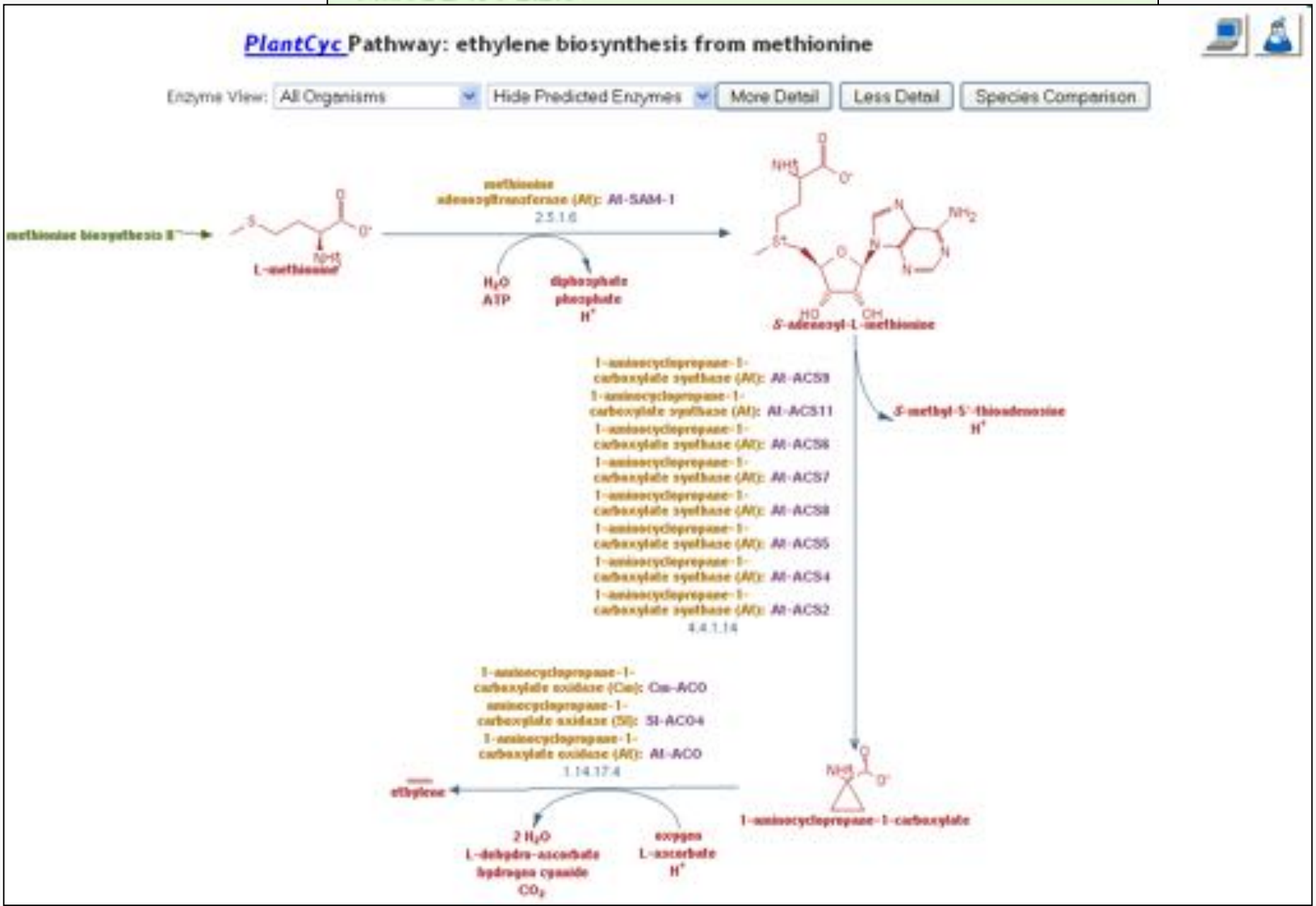


Specific search pages



Sequence-based search options

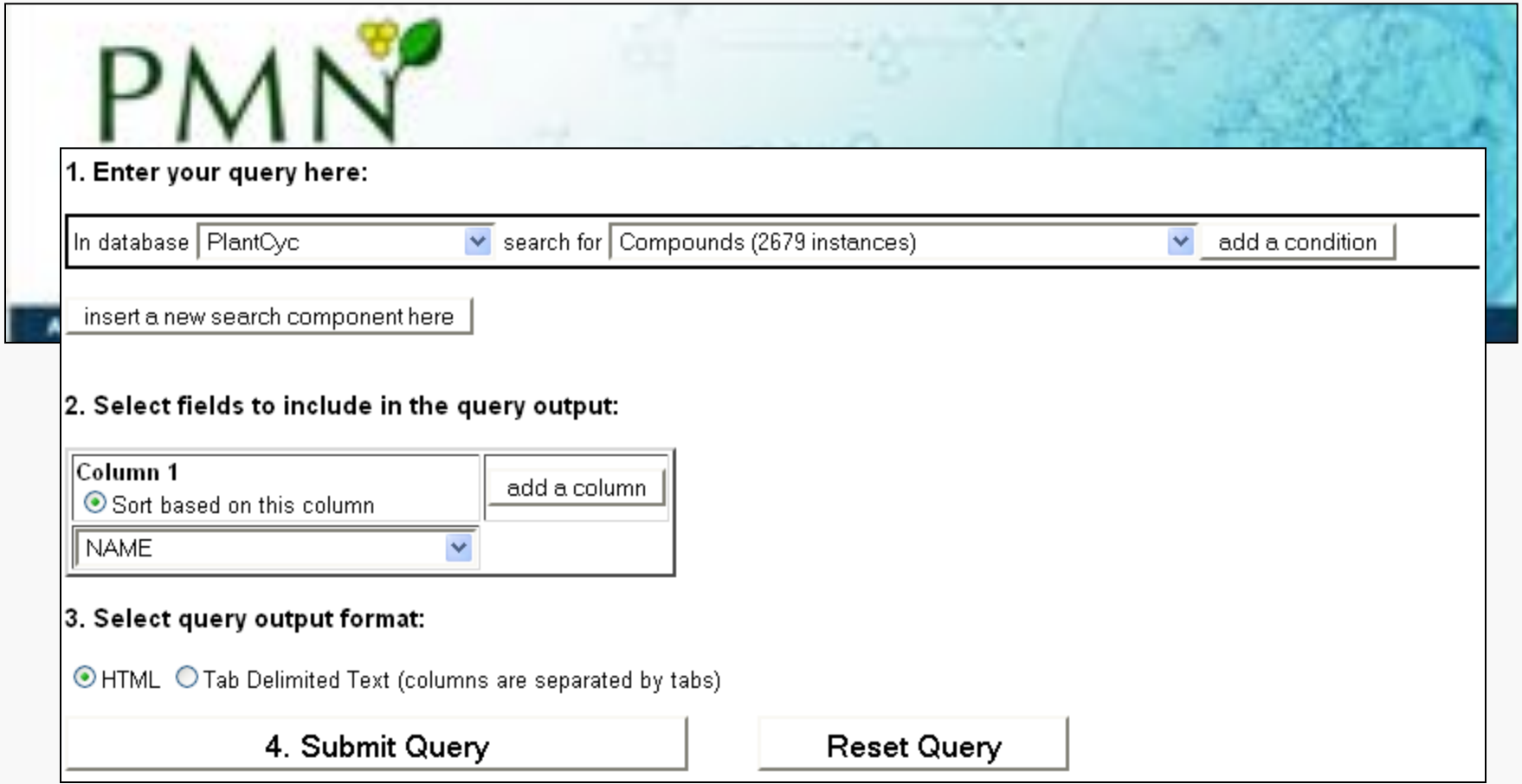
PMN BLAST 2.2.8



Q570P9	PlantCyc: AT5G51690-MONOMER	aromatic-amino-acid ...	392	e-110

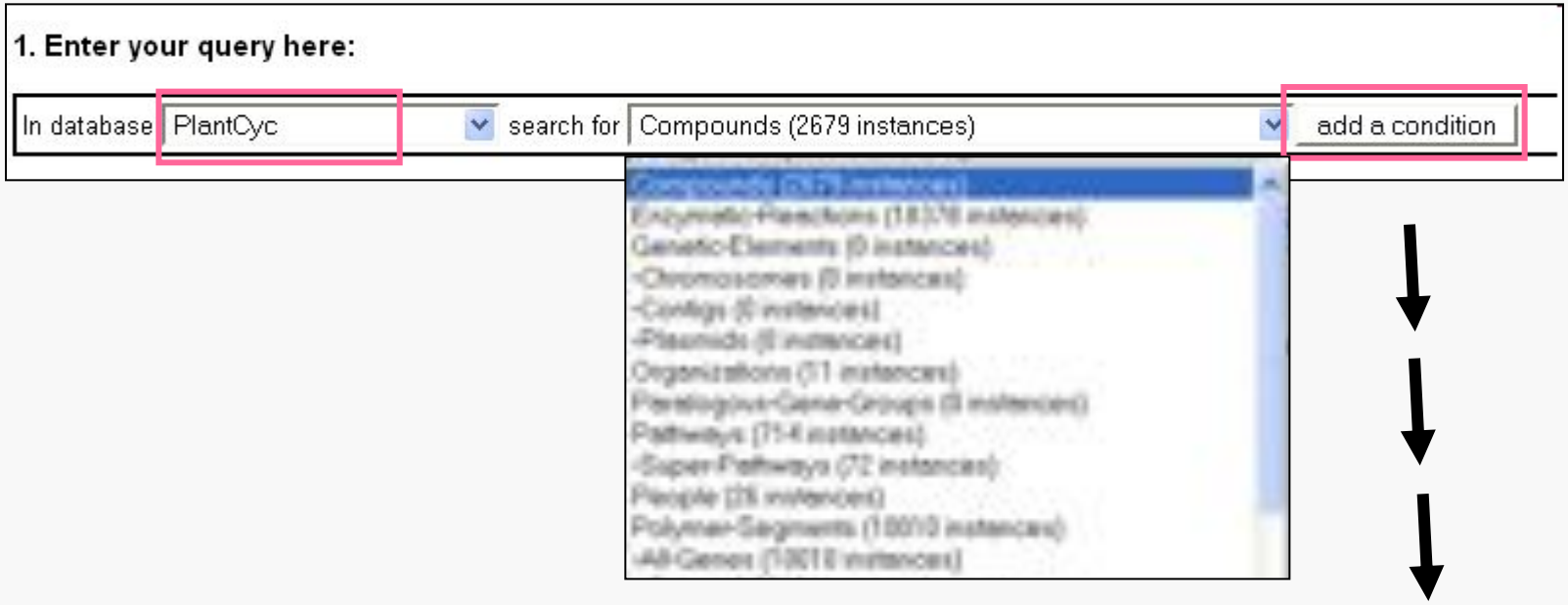
Advanced searching in PMN databases

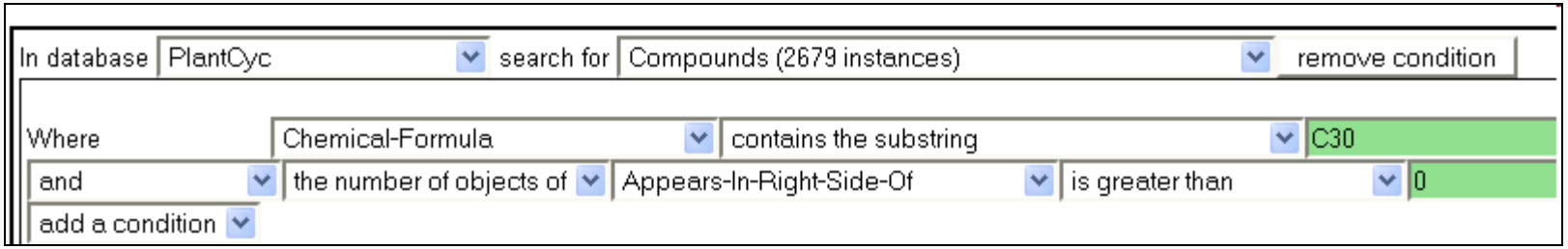
- Advanced search page
  - Allows the construction of very complex queries



Advanced searching in PMN databases

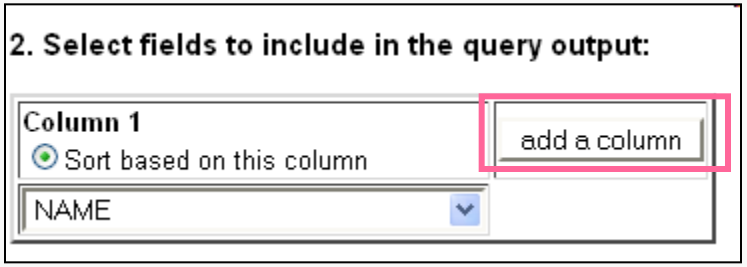
- Find all of the 30-carbon compounds that appear as products in reactions
  - Construct query

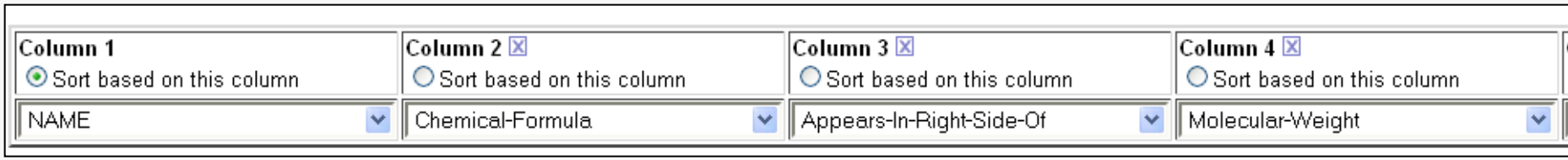




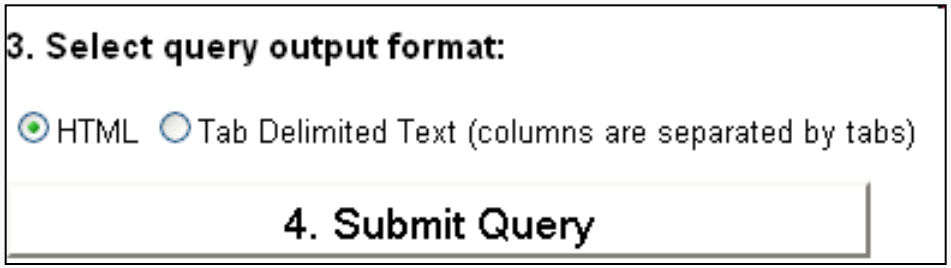
Advanced searching in PMN databases

- Select output fields





- Select file format and retrieve

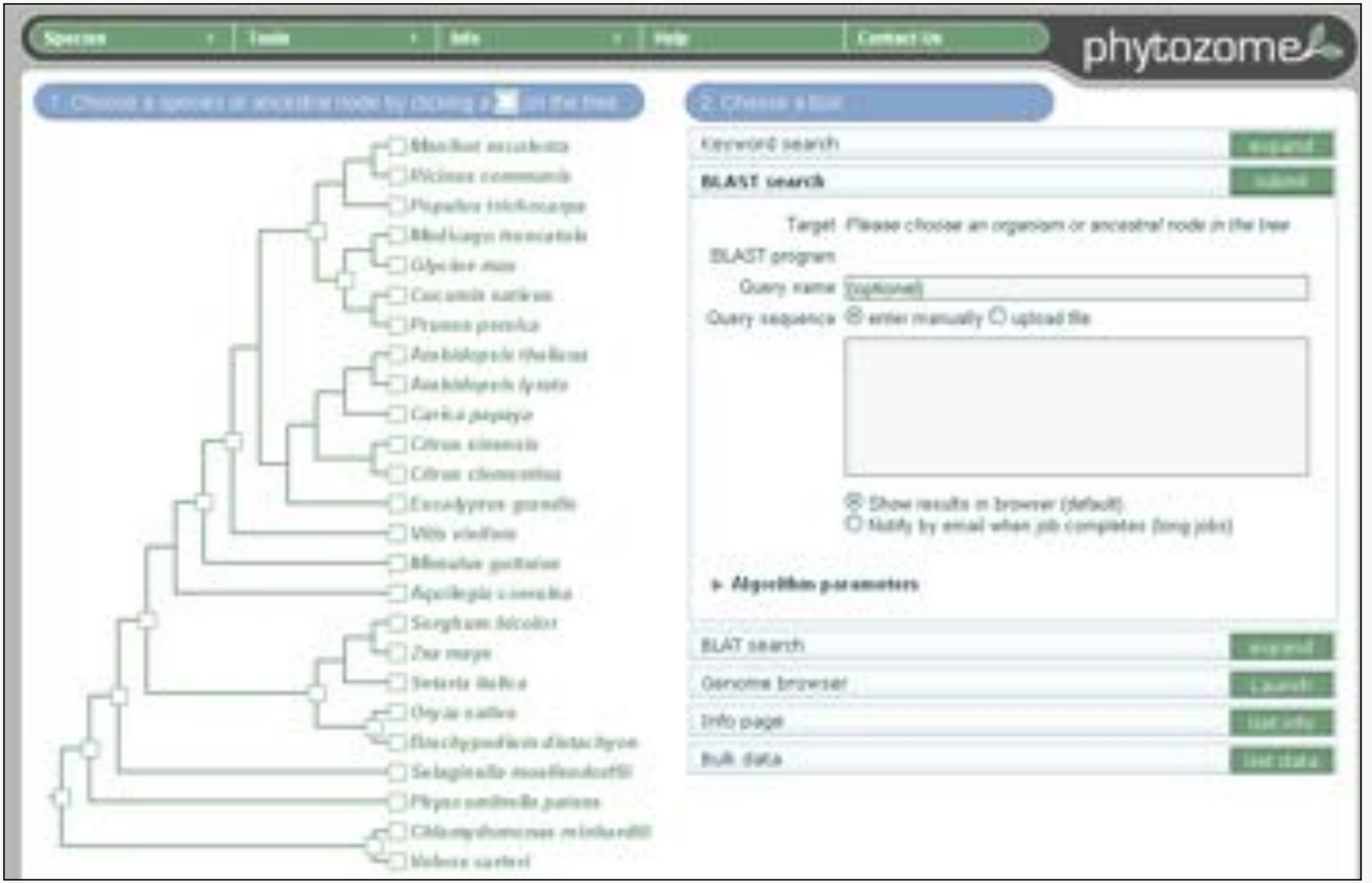


- A list of **30-carbon compounds that appear as products in reactions**

This query resulted in a single table of 53 rows

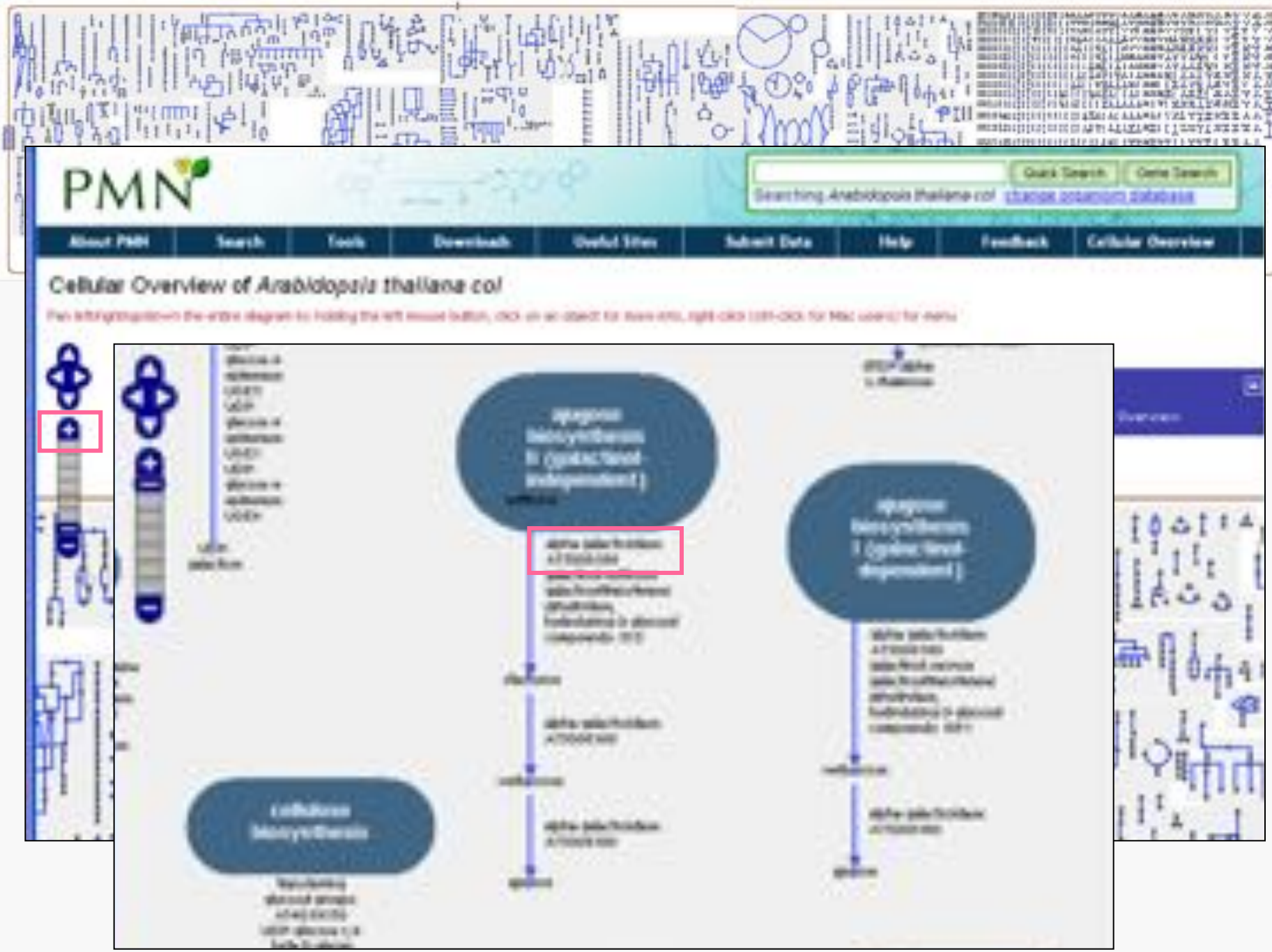
Column 1 for (x1 ^? NAME)	Column 2 for (x1 ^? CHEMICAL-FORMULA)	Column 3 for (x1 ^? APPEARS-IN-RIGHT-SIDE-OF)	Column 4 for (x1 ^? MOLECULAR-WEIGHT)
α-amyrin	C30H50O1	(S)-2,3-epoxysqualene = α-amyrin	426.724
β-amyrin	C30H50O1	(S)-2,3-epoxysqualene = β-amyrin	426.724
(E)-cinnamoyl-CoA	C30H42N7O17P3S1	trans-cinnamate + coenzyme A = (E)-cinnamoyl-CoA + H_2O	897.68
(S)-2,3-epoxysqualene	C30H50O1	squalene + NADPH + O_2 = (S)-2,3-epoxysqualene + $NADP^+$ + H_2O; O_2 + a reduced acceptor + squalene = (S)-2,3-epoxysqualene + H_2O + an acceptor	426.724
24-ethylidenelophenol	C30H50O1	24-methylenelophenol + S-adenosyl-L-methionine = 24-ethylidenelophenol + S-adenosyl-L-homocysteine	426.724
4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol	C30H48O2	4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADPH + O_2 = 4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol + $NADP^+$ + 2 H_2O	440.708
4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol	C30H50O2	lanosterol + NADPH + O_2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + $NADP^+$ + H_2O	442.724
4-(1-methyl-2-pyrrolidinyl)-3-oxobutanoyl-CoA	C30H49N8O18P3S1	1-methylpyrrolidine-2-acetyl-CoA + acetyl-CoA = 4-(1-methyl-2-pyrrolidinyl)-3-oxobutanoyl-CoA + coenzyme A	934.741
4-coumaroyl-CoA	C30H42N7O18P3S1	coenzyme A + 4-coumarate + ATP = 4-coumaroyl-CoA + diphosphate + AMP	913.679

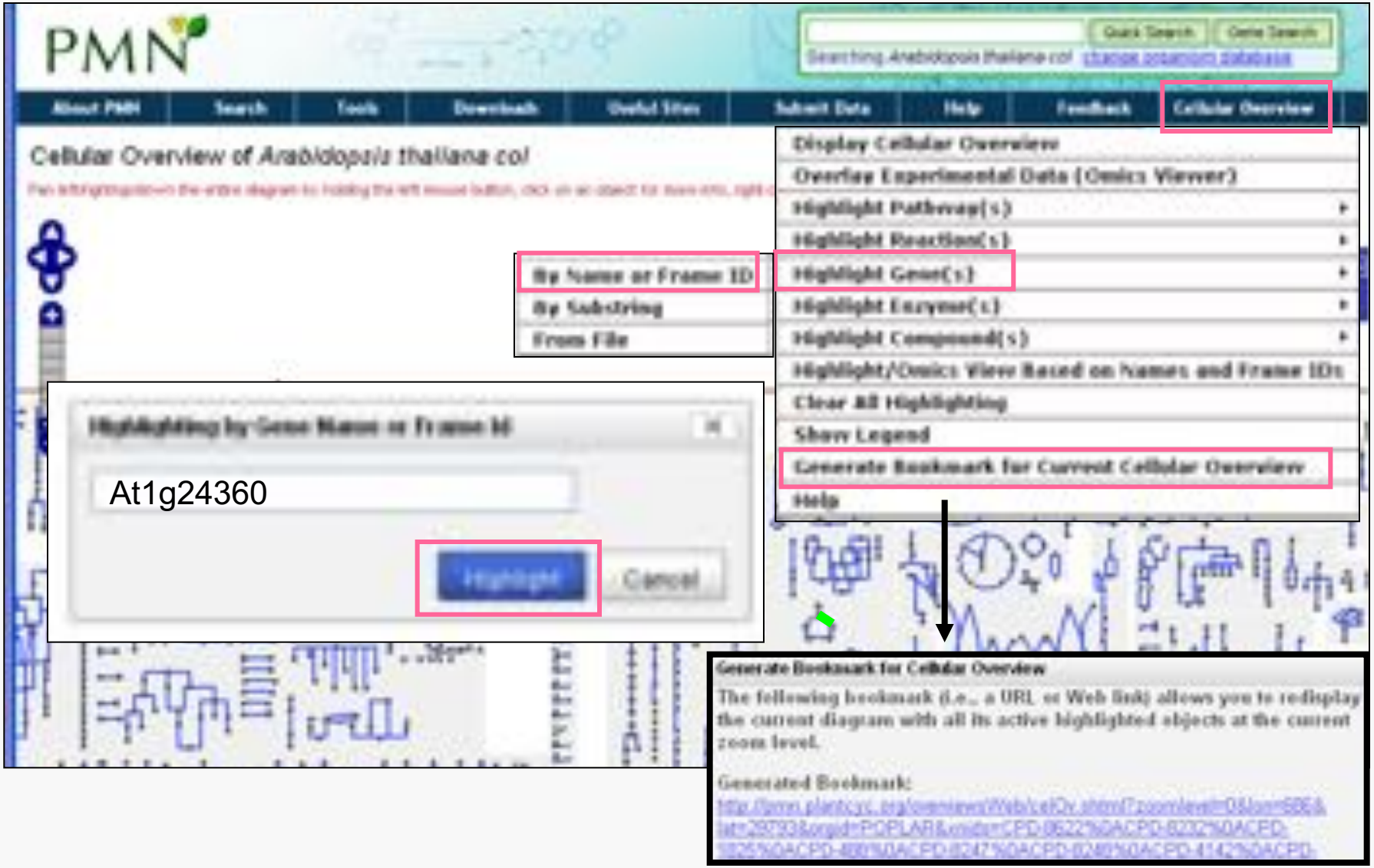
New search options at the PMN


phytozome
Keyword search
BLAST search
Target: Please choose an organism or ancestral node in the tree
BLAST program
Query name
Query sequence
enter manually
upload file
Show results in browser (default)
Notify by email when job completes (long jobs)
Algorithm parameters
BLAT search
Genome browser
Info page
Bulk data

New database analysis tools at the PMN

Improved Metabolic Map / Omics Viewer


PMN
About PMN
Search
Tools
Downloads
Useful Sites
Submit Data
Help
Feedback
Cellular Overview
Cellular Overview of Arabidopsis thaliana col
cellulose biosynthesis

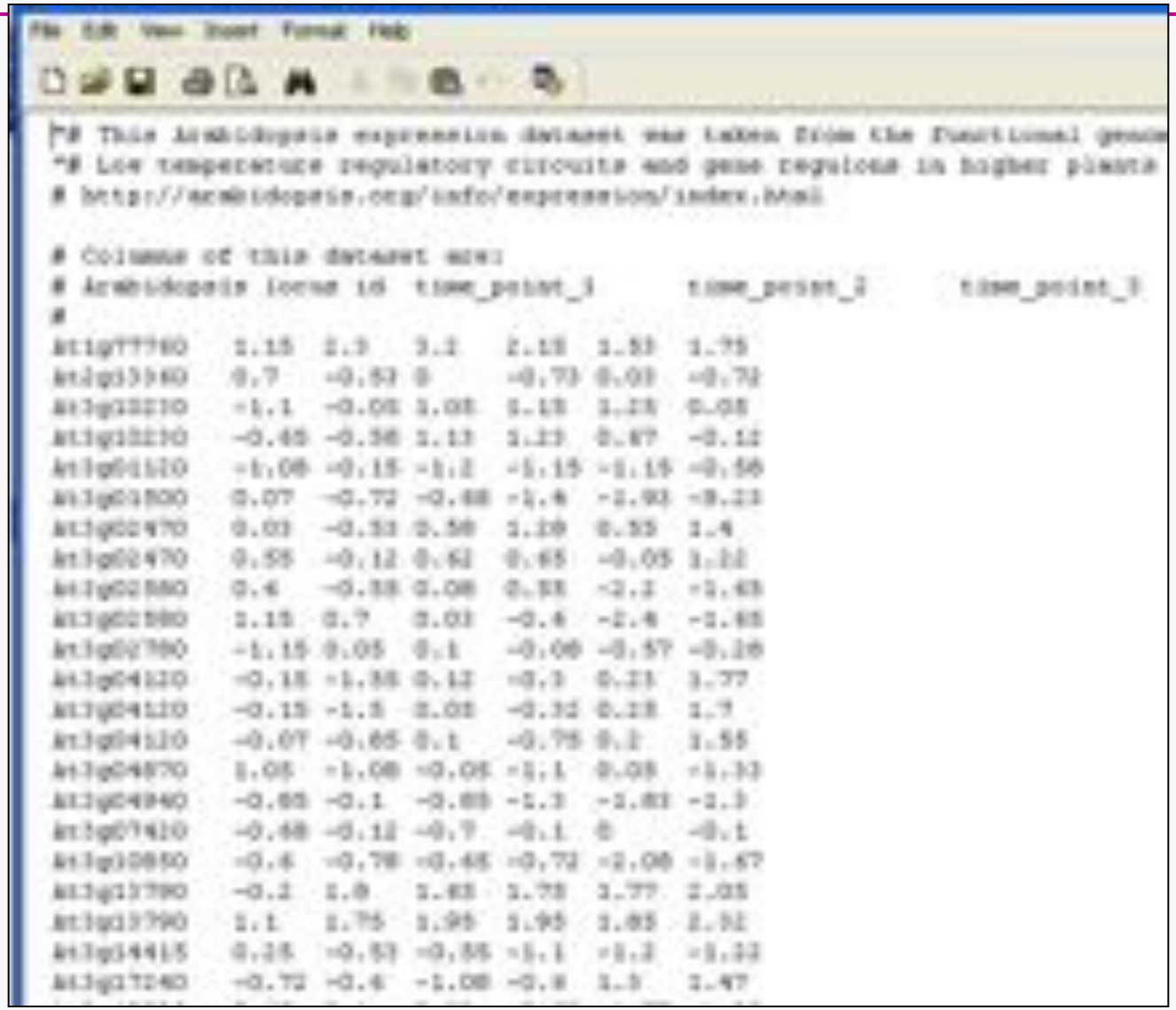

PMN
Cellular Overview
Display Cellular Overview
Overlay Experimental Data (Omics Viewer)
Highlight Pathway(s)
Highlight Reaction(s)
Highlight Gene(s)
Highlight Enzyme(s)
Highlight Compound(s)
Highlight/Omics View Based on Names and Frame IDs
Clear All Highlighting
Show Legend
Generate Bookmark for Current Cellular Overview
Help
By Name or Frame ID
By Substring
From File
At1g24360
Highlight
Cancel
Generate Bookmark for Cellular Overview
The following bookmark (i.e., a URL or Web link) allows you to redisplay the current diagram with all its active highlighted objects at the current zoom level.
Generated Bookmark:

Improved Metabolic Map / Omics Viewer

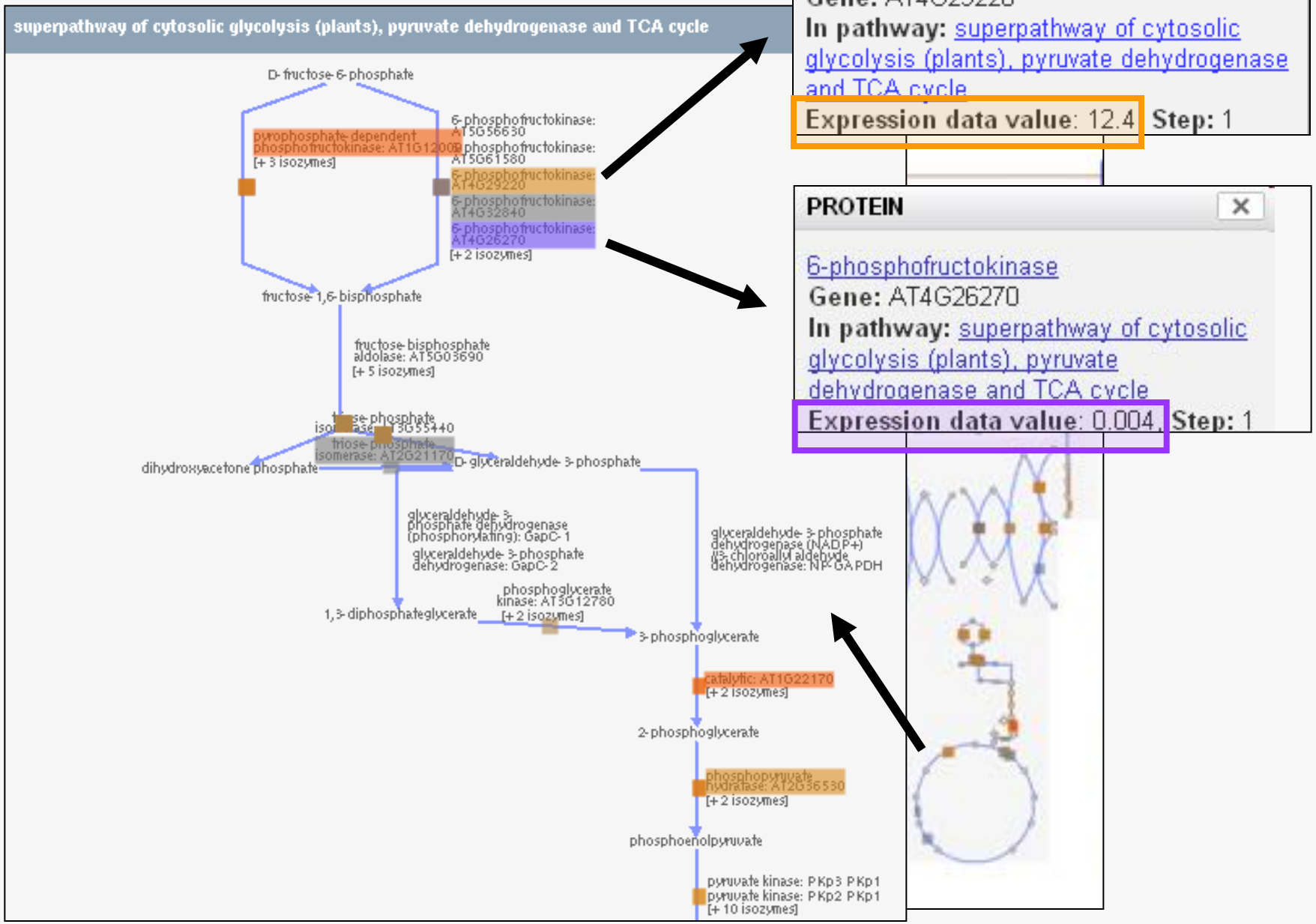
Data analysis with the Metabolic Map / Omics Viewer

- Display experimental data a on metabolic map
  - Data types:
    - Genes - transcriptomics
    - Enzymes – proteomics
    - Reactions - fluxomics
    - Compounds – metabolomics
  - Data inputs:
    - Single or multiple values for each object
    - Absolute or relative values
  - Data comparisons
    - Wild type vs. mutant
    - Stressed vs. unstressed
    - Time course after stimulus
    - Ecotype-specific responses
    - Development-driven changes
    - Organ / tissue – based analyses
    - . . . and many more

Preparing the input files



Displaying the data



PROTEIN

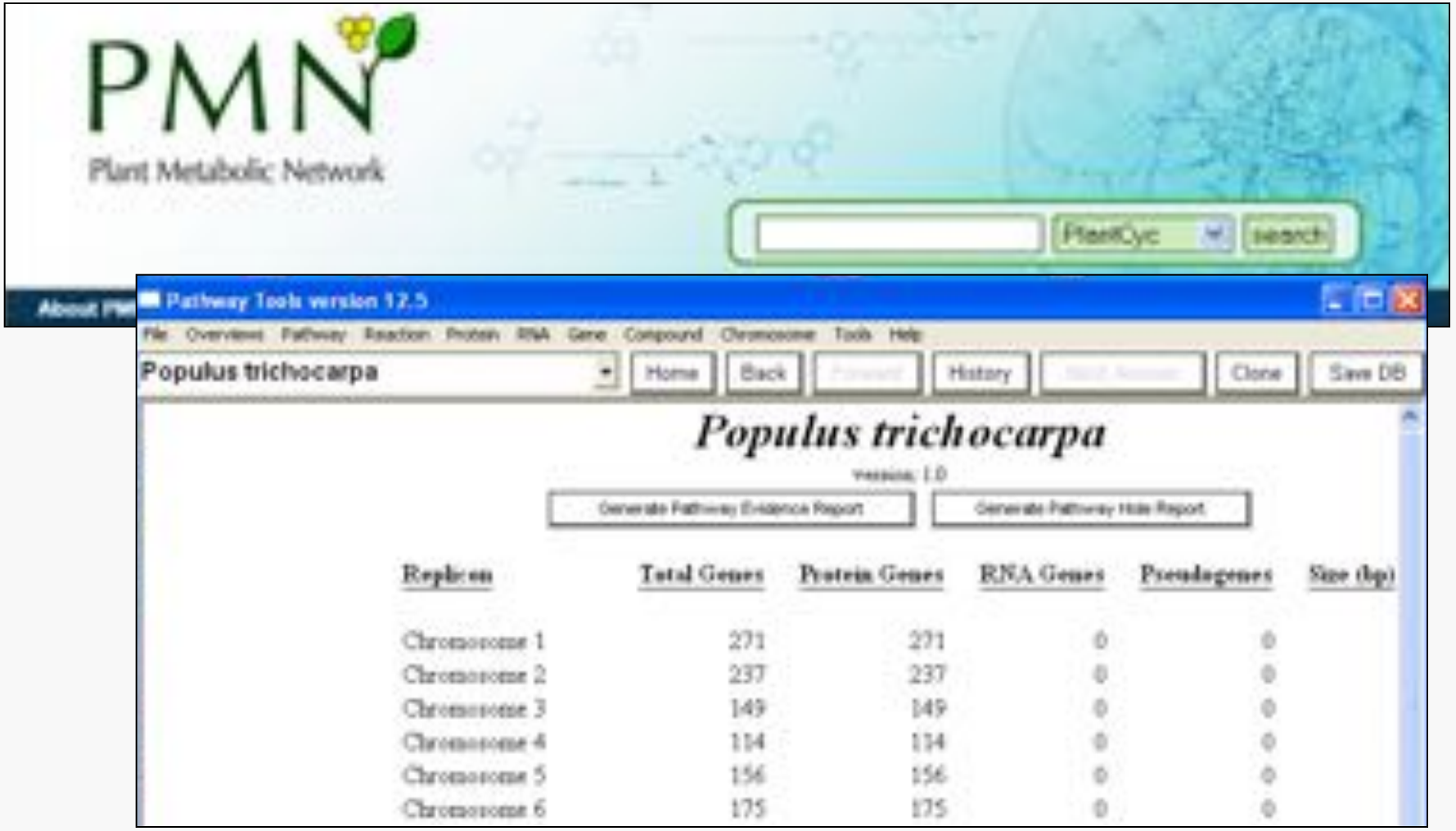
6-phosphofructokinase
Gene: AT4G29220
In pathway: superpathway of cytosolic glycolysis (plants), pyruvate dehydrogenase and TCA cycle
Expression data value: 12.4 **Step:** 1

PROTEIN

6-phosphofructokinase
Gene: AT4G26270
In pathway: superpathway of cytosolic glycolysis (plants), pyruvate dehydrogenase and TCA cycle
Expression data value: 0.004 **Step:** 1

Data and software downloads

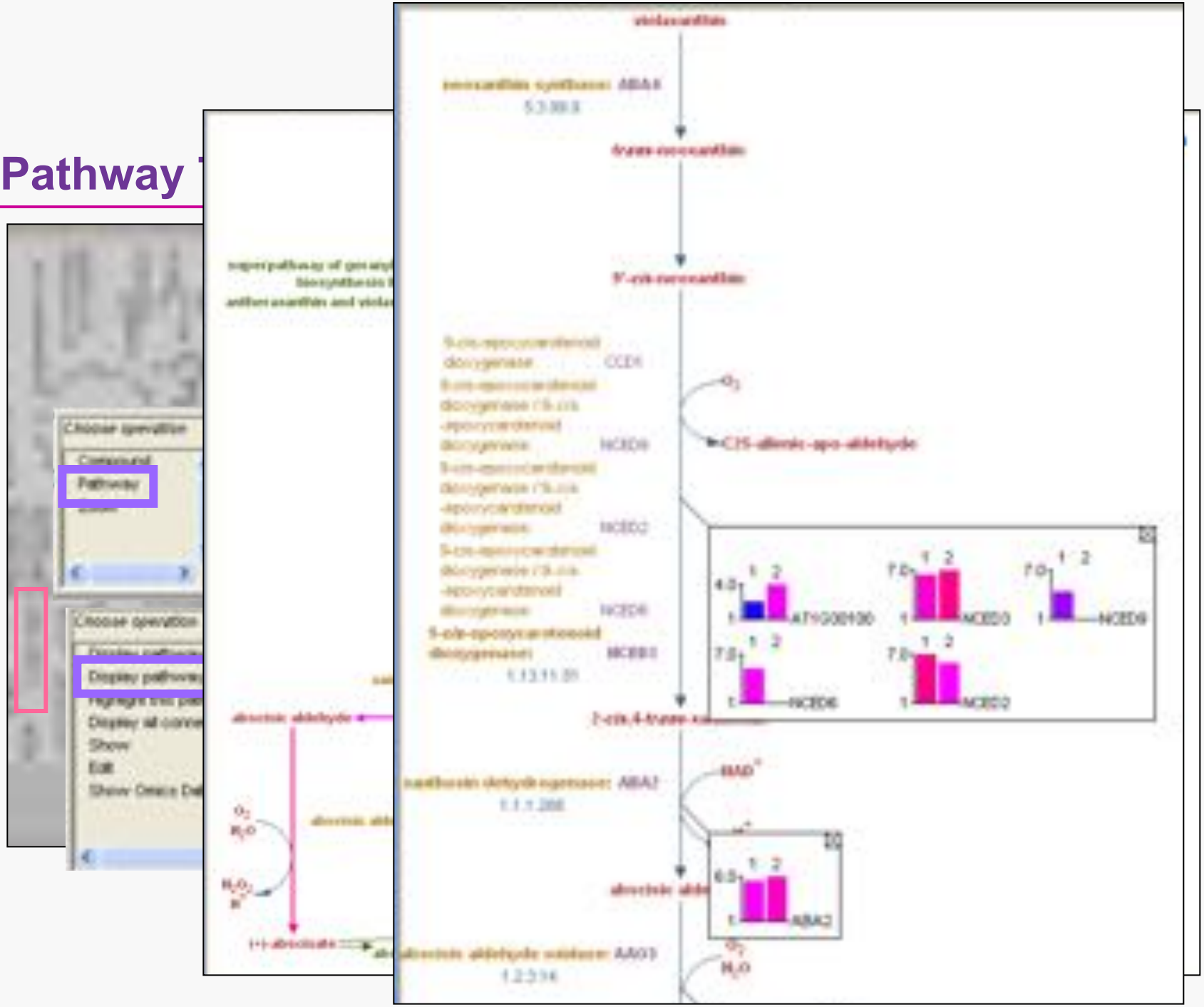
- Install a local copy of the Pathway Tools software



Pathway Tools – desktop version

- Desktop version offers additional features
  - Create new pathways
  - Modify existing pathways
  - Generate and work with groups of items
  - Perform metabolite tracing on metabolic map
  - See Omics Viewer results on pathway pages
- **Demonstrations available throughout the conference**

Pathway


Choose operation
Compound
Pathway
Choose operation
Display pathway
Display all
Show
Edit
violaxanthin
9'-cis-neoxanthin
CCD4
NCED3
NCED5
NCED9
NCED1
ABA2
AAO3
AtNCED3
NCED5
NCED9
NCED6
NCED2
ABA2

On the horizon . . .

- Develop a better enzyme annotation pipeline
- Predict plant metabolic pathways for many species with predicted proteomes
  - First round candidates:
    - moss
    - papaya
    - rice
    - tomato
    - Chlamydomonas
    - wine grape
    - *Selaginella moellendorfii*
    - soybean
    - maize
    - Medicago
    - poplar
    - cassava
- Solicit help from experts for pathway validation
  - Remove mis-predictions
  - Add missed pathways
- Release new databases at the PMN

Questions, contributions, volunteering to validate?

- To submit data, report an error, or ask a question . . .
  - . . . stay tuned for the next portion of my talk
  - Send an e-mail: curator@plantcyc.org
  - Use our feedback form:

  - Stop by the Plant Genome Resources Outreach Booth (415)
  - Visit the PMN poster – P02029
  - Schedule an individual meeting with me

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

curator@plantcyc.org

AraCyc

www.plantcyc.org

GRAMENE

SolCyc

PMN Acknowledgements

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

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

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

National Science Foundation
WHERE DISCOVERIES BEGIN

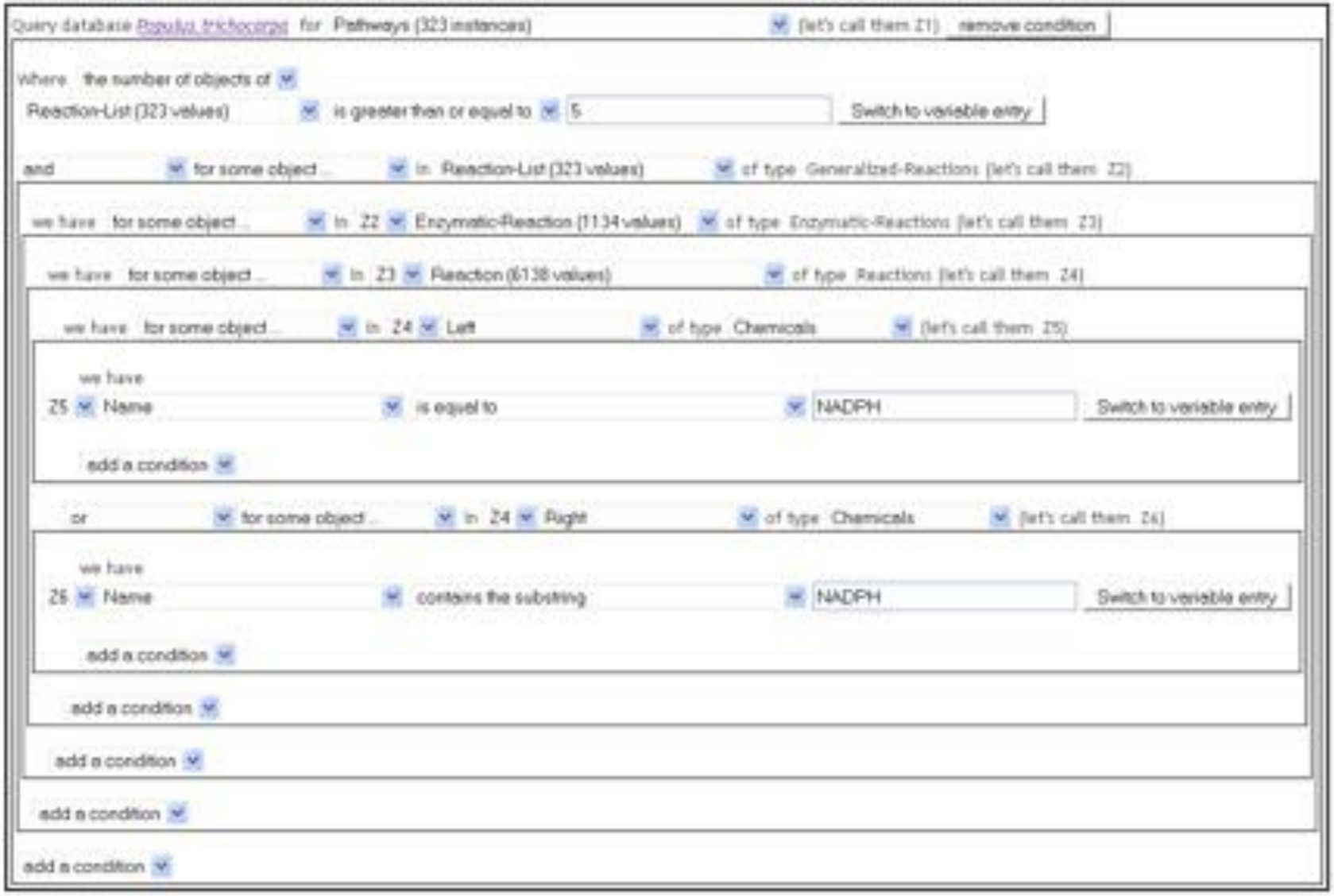
CARNEGIE
INSTITUTION FOR
SCIENCE

Advanced searching in PMN databases

- Other queries
  - Identify all of the “glycosyltransferase” enzymes associated with no more than two reactions in AraCyc
    - List their:
      - name
      - subcellular localization
      - reactions catalyzed
  - Find all of the biochemical pathways in PoplarCyc that have more than 5 reactions and where NADPH is used at least once in the pathway
    - List their:
      - name
      - reactions
      - citations and evidence codes

Improved Metabolic Map / Omics Viewer

1. Enter your query here:



Advanced Query Results

Your query in BioVelo is `html-sort-ascending([(Z1^NAME, Z1^REACTION-LIST, Z1^CITATIONS) : Z1<-POPLAR^^Pathways, Z2 := [Z2 : Z2 <- Z1^REACTION-LIST, [0 < #[Z3 : Z3 <- Z2^ENZYMATIC-REACTION, [0 < #[Z4 : Z4 <- Z3^REACTION, ([0 < #[Z5 : Z5 <- Z4^LEFT, ("NADPH" instringci Z5^NAME)]) | [0 < #[Z6 : Z6 <- Z4^RIGHT, ("NADPH" instringci Z6^NAME)]])]])], ((#Z1^REACTION-LIST) >= 5) & [0 < #Z2],1)` (click the query to edit it in the Free Form Advanced Query Page)

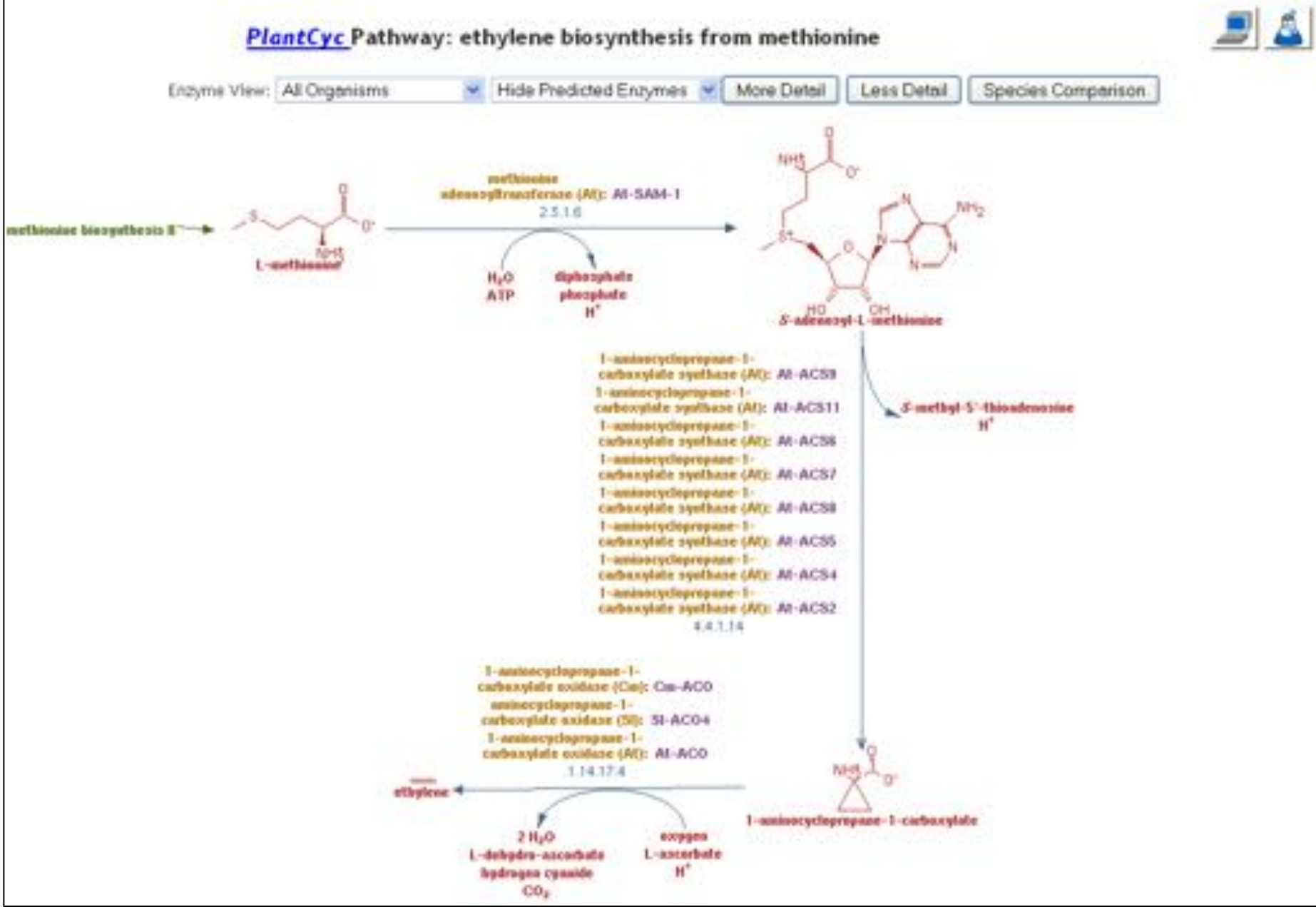
This query resulted in a single table of 29 rows.

NAME ▲ ▽	Reaction-List △ ▽	Citations △ ▽
arginine biosynthesis II (acetyl cycle)	2 ATP + L-glutamine + bicarbonate + H_2O = L-glutamate + 2 ADP + phosphate + carbamoyl-phosphate + 4 H^+, L-glutamate + acetyl-CoA = *N*-acetyl-L-glutamate + coenzyme A + H^+, L-glutamate + *N*-acetyl-L-ornithine = *N*-acetyl-L-glutamate + L-ornithine, *N*-acetyl-L-glutamate + ATP = *N*-acetylglutamyl-phosphate + ADP + H^+, *N*-acetyl-L-glutamate 5-semialdehyde + $NADP^+$ + phosphate = *N*-acetylglutamyl-phosphate + NADPH + H^+, *N*-acetyl-L-ornithine + 2-ketoglutarate = *N*-acetyl-L-glutamate 5-semialdehyde + L-glutamate, L-ornithine + carbamoyl-phosphate = citrulline + phosphate + H^+, L-arginino-succinate = L-arginine + fumarate, L-aspartate + citrulline + ATP = L-arginino-succinate + diphosphate + AMP + 2 H^+	curator09:EV-COMP-HINF:3452886714:pellenz
ascorbate glutathione cycle	L-ascorbate + O_2 + 5 H^+ = monodehydroascorbate + 2 H_2O, 2 monodehydroascorbate -> L-ascorbate + L-dehydro-ascorbate + 3 H^+, 2 monodehydroascorbate + NADPH = 2 L-ascorbate + $NADP^+$ + 3 H^+, 2 L-ascorbate + H_2O_2 + 4 H^+ = 2 monodehydroascorbate + 2 H_2O, 2 glutathione + $NADP^+$ = glutathione disulfide + NADPH + H^+, L-dehydro-ascorbate + 2 glutathione = glutathione disulfide + L-ascorbate + H^+	17899442:EV-EXP-IDA:3457189877:dreher, EDJOL001:EV-EXP-IDA:3457190309:dreher, SHANZH04:EV-EXP-IDA:3457190309:dreher, Nakagawara84:EV-EXP-IDA:3457190309:dreher
brassinosteroid biosynthesis I	6-deoxotyphasterol + O_2 = typhasterol + H_2O, 6α-hydroxy-castasterone = castasterone + 2 H^+, 6-deoxocastasterone + O_2 + 2 H^+ = 6α-hydroxy-castasterone + H_2O, 6-deoxotyphasterol + O_2 + 2 H^+ = 6-deoxocastasterone + H_2O, 3-dehydro-6-deoxoteasterone + 2 H^+ = 6-deoxotyphasterol, 6-deoxoteasterone = 3-dehydro-6-deoxoteasterone + 2 H^+, 6-deoxocathasterone + O_2 + 2 H^+ = 6-deoxoteasterone + H_2O, campestanol + NADPH + O_2 + H^+ = 6-deoxocathasterone + $NADP^+$ + H_2O, castasterone + NADPH + O_2 + H^+ = brassinolide + $NADP^+$ + H_2O, typhasterol + a reduced electron acceptor + O_2 = castasterone + an oxidized electron acceptor + H_2O, 3-dehydroteasterone + 2 H^+ = typhasterol, teasterone = 3-dehydroteasterone + 2 H^+, cathasterone + a reduced electron acceptor + O_2 = teasterone + an oxidized electron acceptor + H_2O, 6-oxocampestanol + NADPH + O_2 + H^+ = cathasterone + $NADP^+$ + H_2O, (6α)-hydroxycampestanol = 6-oxocampestanol + 2 H^+, campestanol + a reduced electron acceptor + O_2 = (6α)-hydroxycampestanol + an oxidized electron acceptor + H_2O, (5α)-campestan-3-one + 2 H^+ = campestanol, campest-4-en-3-one +	curator09:EV-COMP-HINF:3444506243:achill

Plant Metabolic Network goals

- Capture and organize published data
- Facilitate data analysis
- Make proteome-based metabolic pathway predictions for **many** plant species

PMN BLAST 2.2.8



Ontology-based search options

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

vledgements

(Director)

Current Tech Team Contributors:

ager)

s. Administrator)

ontributors

- Christophe Tissier
- Hartmut Foerster

- Damian Priamurskiy
- Caryn Johansen
- Purva Karia

Collaborators:

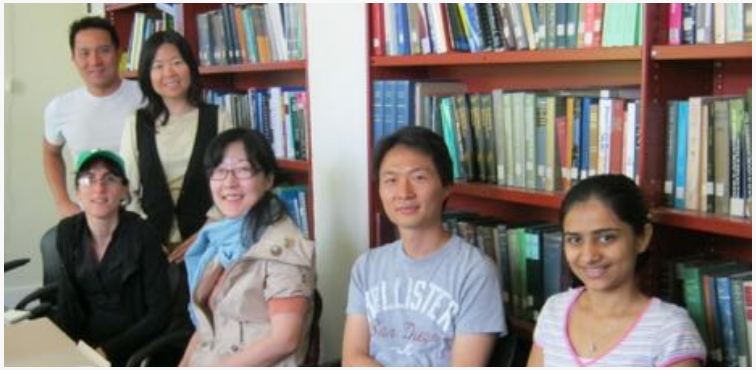
- Peter Karp (SRI)
- Ron Caspi (SRI)
- Suzanne Paley (SRI)
- SRI Tech Team
- Lukas Mueller (SGN)
- Anuradha Pujar (SGN / SRI)
- Gramene, MedicCyc, ChlamyCyc

National Science Foundation
WHERE DISCOVERIES BEGIN

CARNEGIE
INSTITUTION FOR
SCIENCE

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

Booth 415:
Plant Resources Outreach
Tuesday, 10:10 – 10:25 AM
Wednesday, 1:00 – 1:30 PM

Poster **P02029** :
Monday: 5:30 – 7:30 PM