Accessing information in plant metabolic pathway databases at the PMN, Gramene, and SGN

Part I: Contents, Search Strategies, and Data Sharing Opportunities
kate dreher, PMN / TAIR - www.plantcyc.org

Part II: Pathway Networks for Cereals
Palitha Dharmawardhana, Gramene - www.gramene.org/pathway

Part III: Using the desktop version of the Pathway Tools software
Lukas Mueller, SGN – solcyc.solgenomics.net

Part I overview

- Introduction
- Data
- Search tools
- Data submissions
- Acknowledgments
- More examples

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

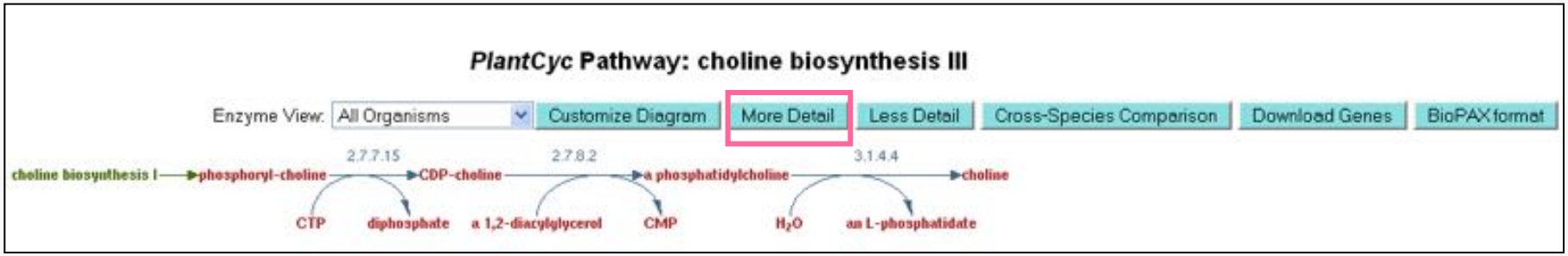
Plant metabolic databases

- Capture and organize published data
- Make metabolic predictions for “new” species
- Facilitate data analysis
- Focus on different aspects of metabolism
  - Pathways
    - KEGG
    - Reactome
    - **BioCyc / Pathway tools family**

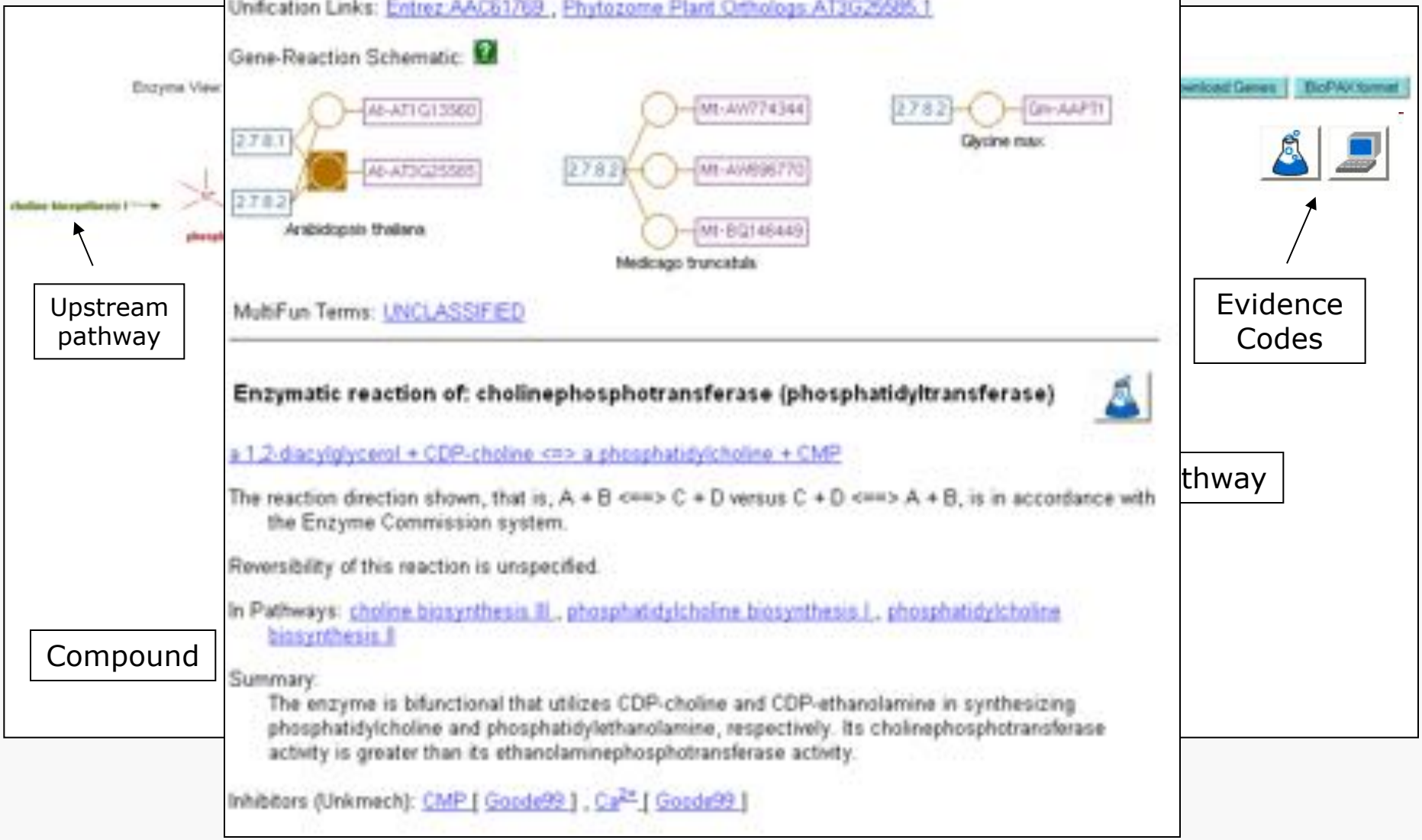
BioCyc / Pathway Tools databases

- Multiple data providers:
  - SRI International
  - Plant Metabolic Network (PMN)
  - Gramene
  - Sol Genomics Network (SGN)
- One software package:
  - Pathway Tools
  - SRI International
    - Common set of data types
    - Common tools
    - Common display modes

Pathway tools pathways



PlantCy


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:
At-AT1G13560
At-AT3G25585
2.7.8.1
2.7.8.2
Arabidopsis thaliana
Mt-AW774344
Mt-AW696770
Mt-BQ146449
Medicago truncatula
Gm-AAP71
Glycine max
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 [Gorde99] , Ca2+ [Gorde99]
Upstream pathway
Compound
Evidence Codes
thway

Plant metabolic databases

PGDB	Plant	Source
AraCyc	Arabidopsis	Plant Metabolic Network
PoplarCyc	Poplar	Plant Metabolic Network
PlantCyc	PLANT KINGDOM	Plant Metabolic Network
RiceCyc	Rice	Gramene
SorghumCyc	Sorghum	Gramene
MaizeCyc (beta)	Sorghum	Gramene
BrachyCyc (beta)	*Brachypodium*	Gramene
LycoCyc	Tomato	Sol Genomics Network
PotatoCyc	Potato	Sol Genomics Network
CapCyc	Pepper	Sol Genomics Network
NicotianaCyc	Tobacco	Sol Genomics Network
PetuniaCyc	Petunia	Sol Genomics Network
CoffeaCyc	Coffee	Sol Genomics Network
MedicCyc	Medicago	Noble Foundation

Data curation in databases

PGDB	Plant	Source	Status
AraCyc	**Arabidopsis**	**Plant Metabolic Network**	**extensive curation**
PoplarCyc	**Poplar**	**Plant Metabolic Network**	**some curation**
PlantCyc	**PLANT KINGDOM**	**Plant Metabolic Network**	**mixed curation**
RiceCyc **	**Rice**	**Gramene**	**some curation**
SorghumCyc	Sorghum	Gramene	all computational
MaizeCyc (beta)	Sorghum	Gramene	all computational
BrachyCyc (beta)	*Brachypodium*	Gramene	all computational
LycoCyc **	**Tomato**	**Sol Genomics Network**	**some curation**
PotatoCyc	Potato	Sol Genomics Network	all computational
CapCyc	Pepper	Sol Genomics Network	all computational
NicotianaCyc	Tobacco	Sol Genomics Network	all computational
PetuniaCyc	Petunia	Sol Genomics Network	all computational
CoffeaCyc	Coffee	Sol Genomics Network	all computational
MedicCyc **	**Medicago**	**Noble Foundation**	**some curation**

Database content statistics

- Plant Metabolic Network

	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*

- Gramene / SGN

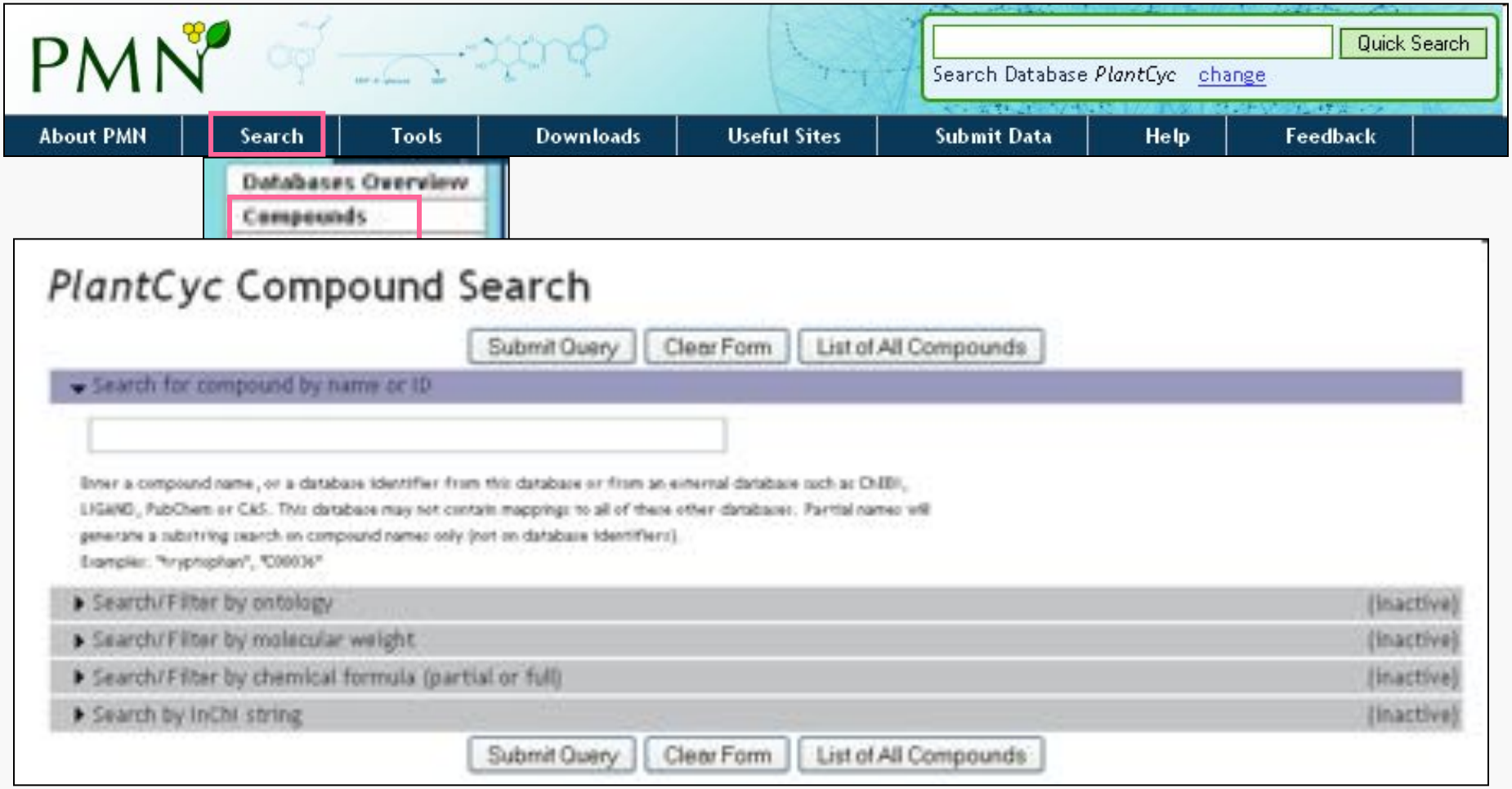
	RiceCyc 3.0	LycoCyc 2.0	PetuniaCyc 2.1
Pathways	339	343	136
Enzymes	9315	5216	294
Reactions	2109	1793	758
Compounds	1592	1379	637
Organisms	1	1	1

Searching in plant metabolic databases

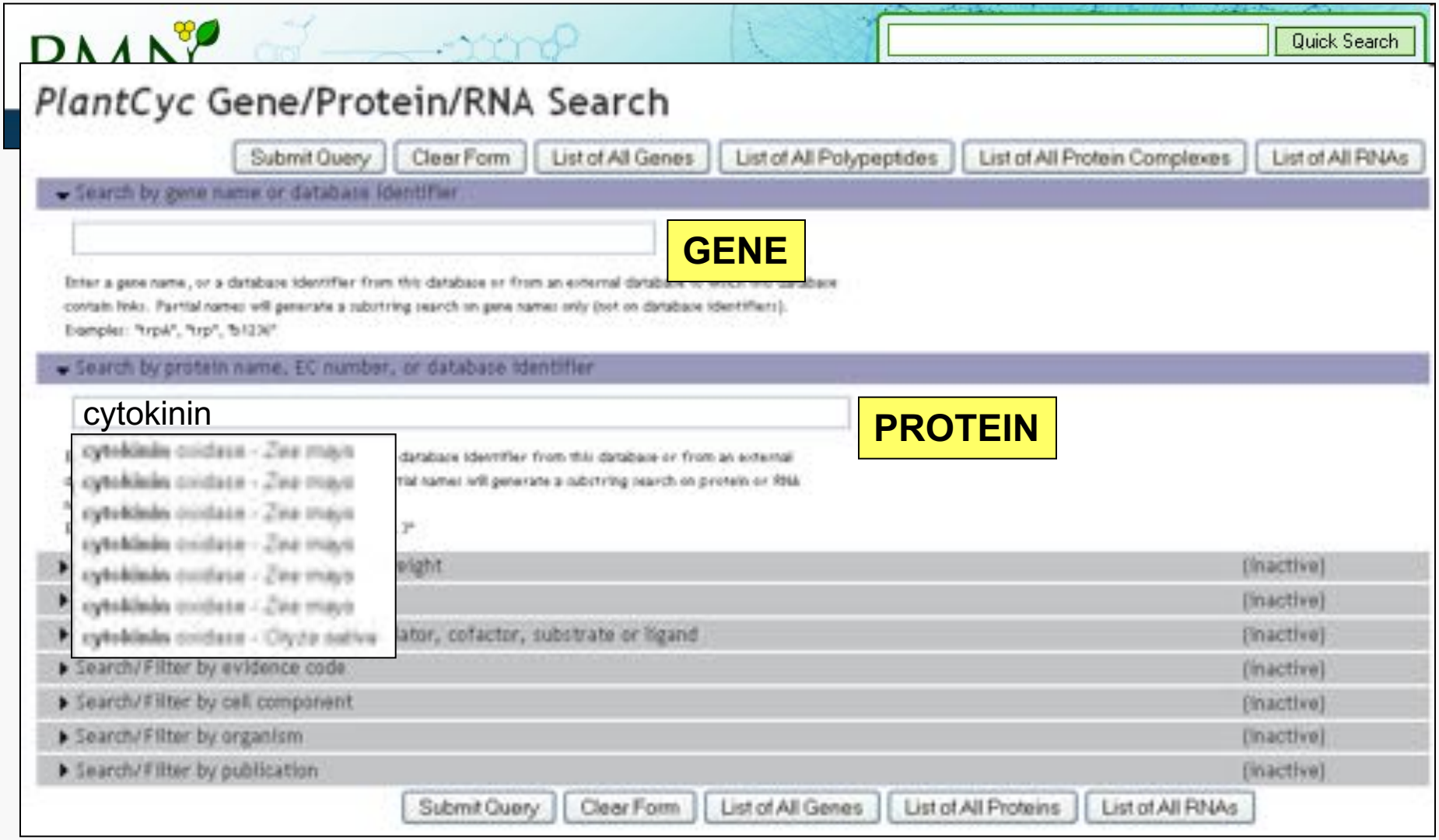
- Pathway Tools quick search bar

Quick search results

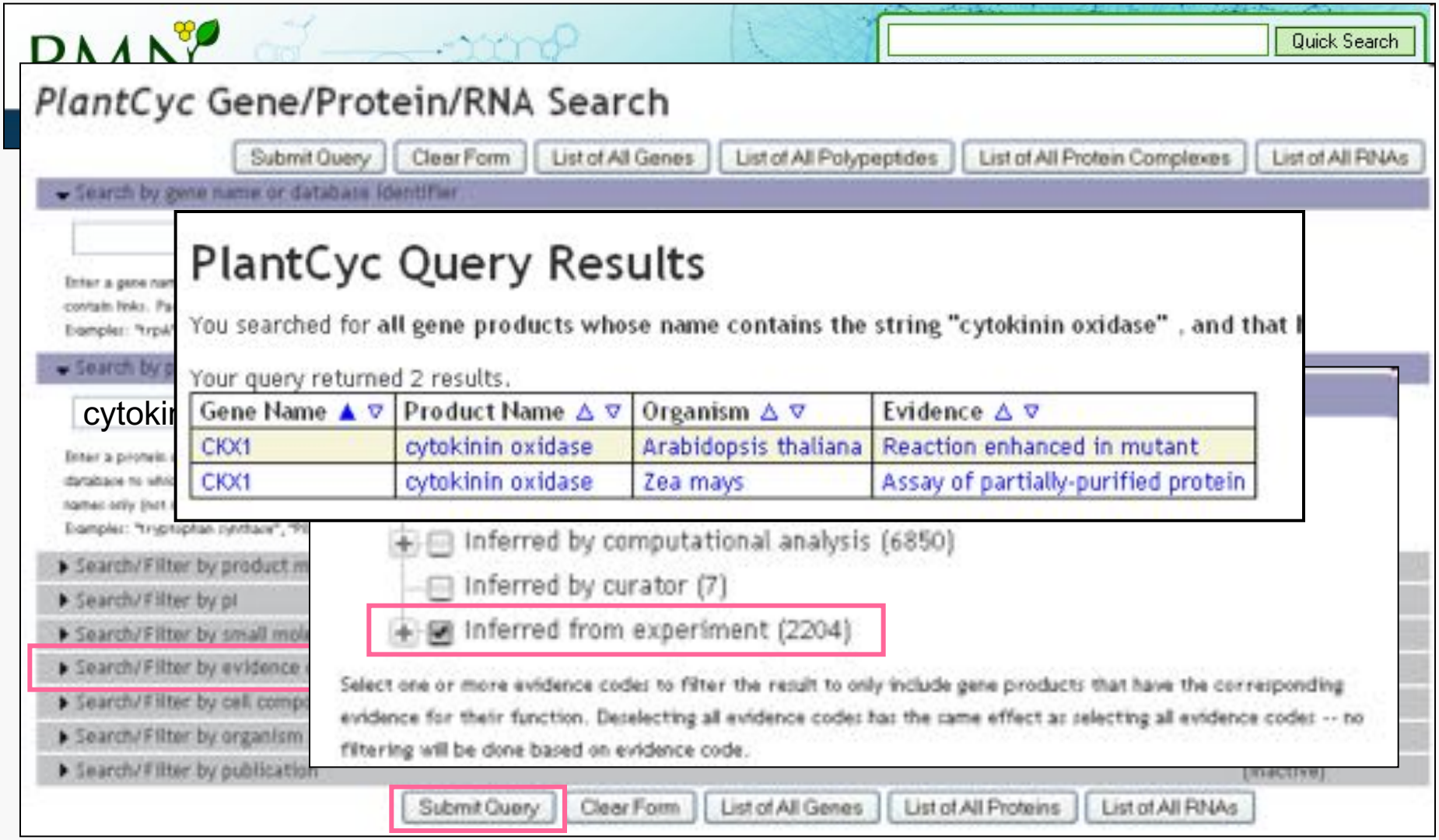
Specific search pages



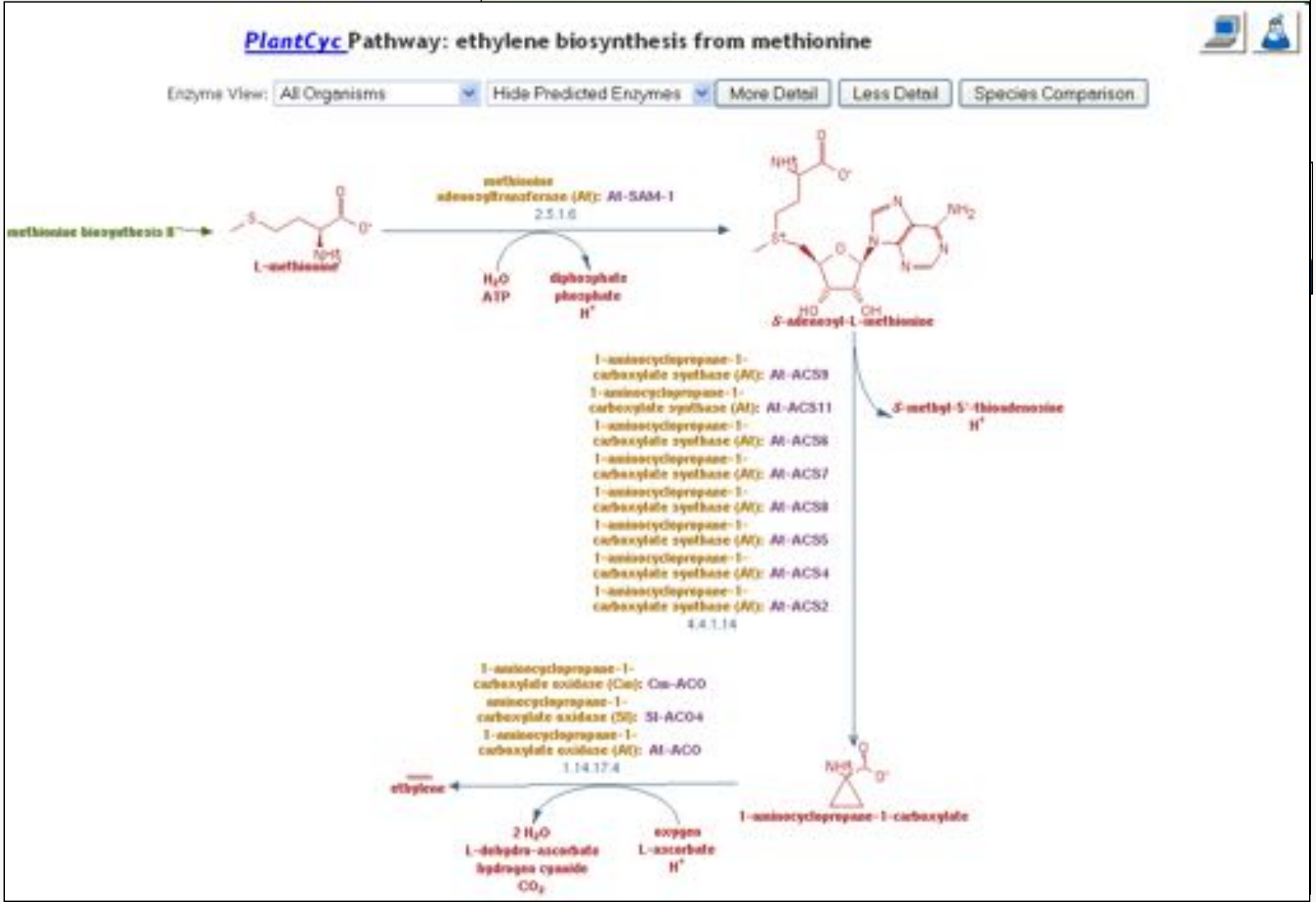
Specific search pages



Specific search pages

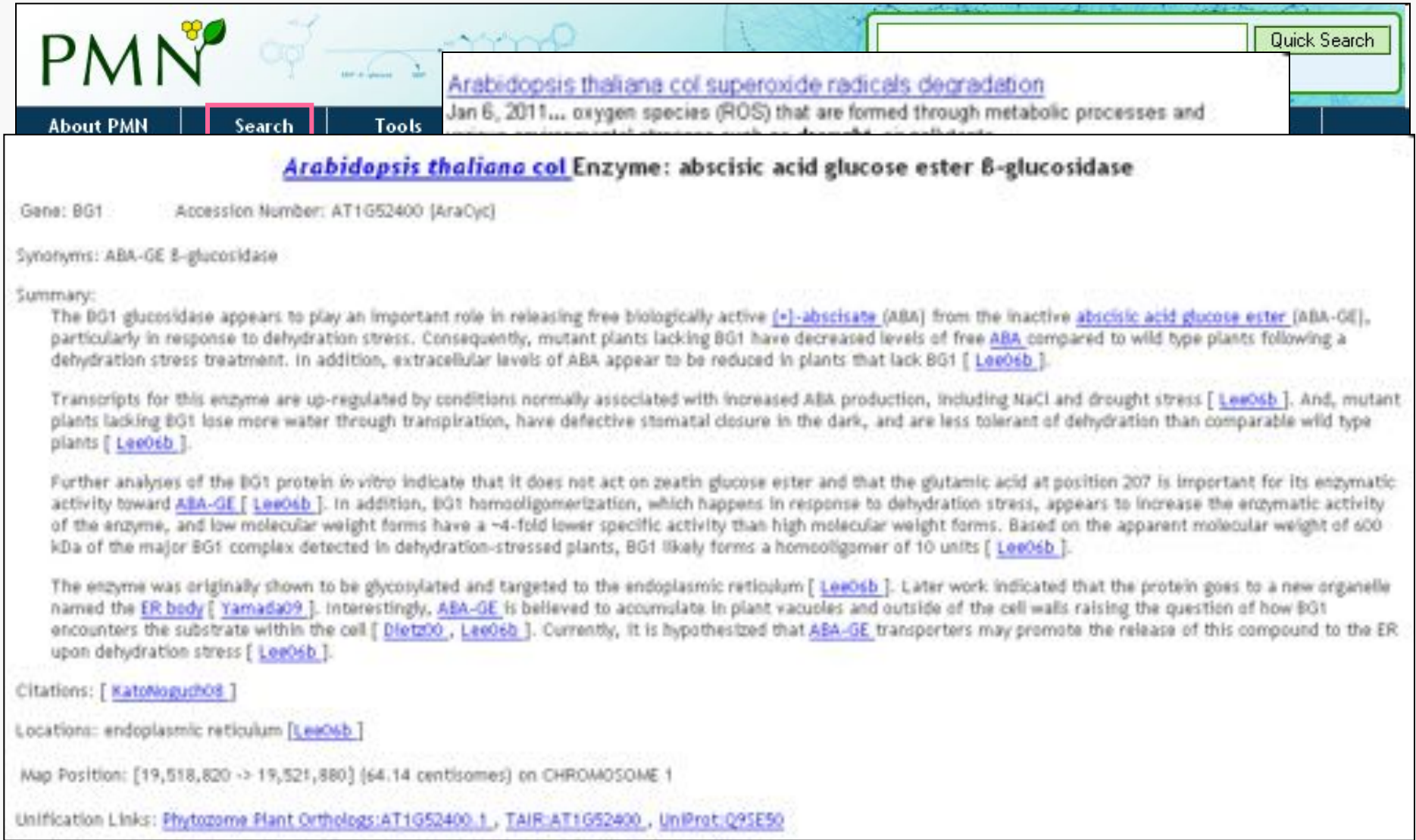


PMN BLAST 2.2.8



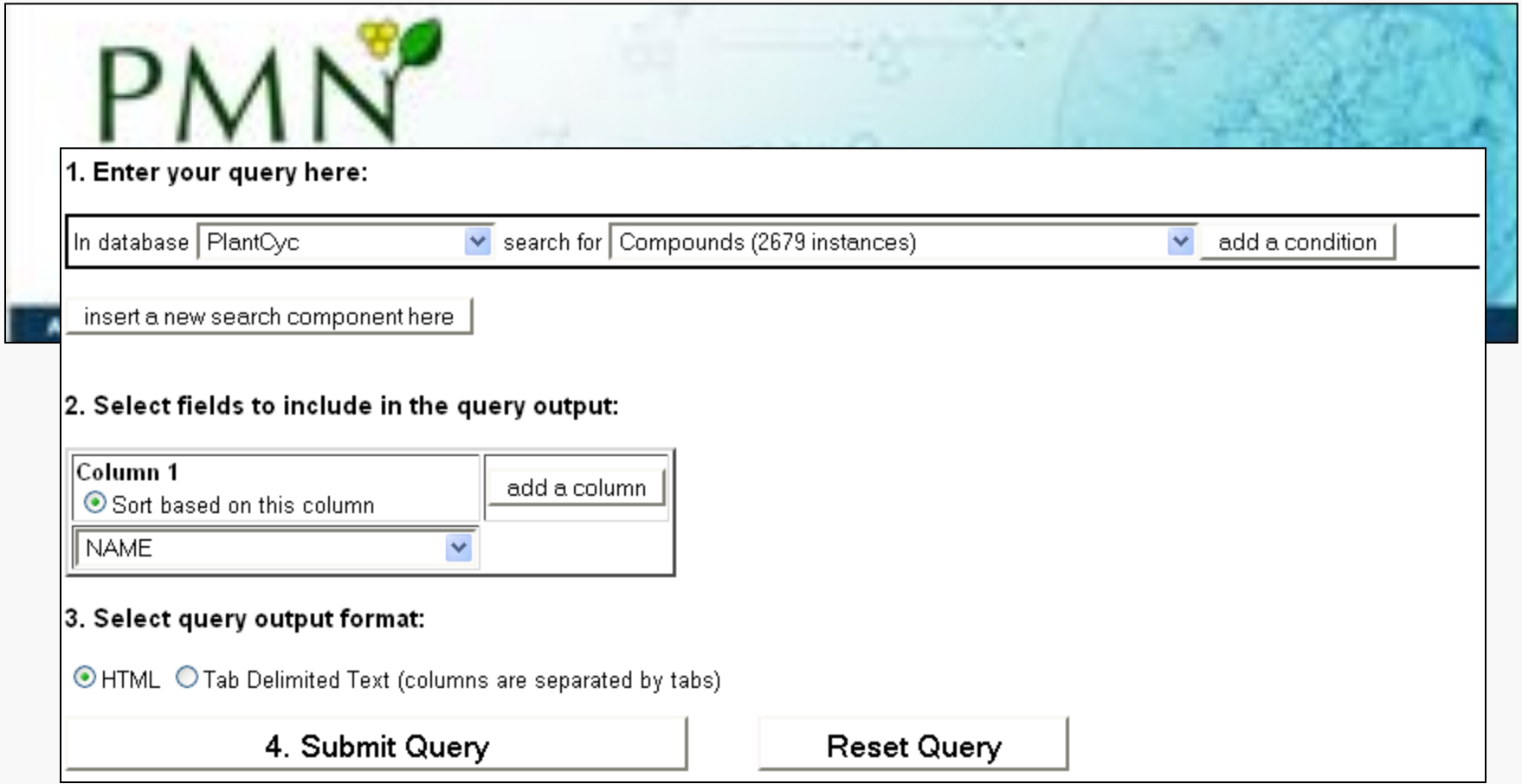
Q570P9	PlantCyc: AT5G51690-MONOMER	aromatic-amino-acid ...	392	e-110

Searching by topic / keyword



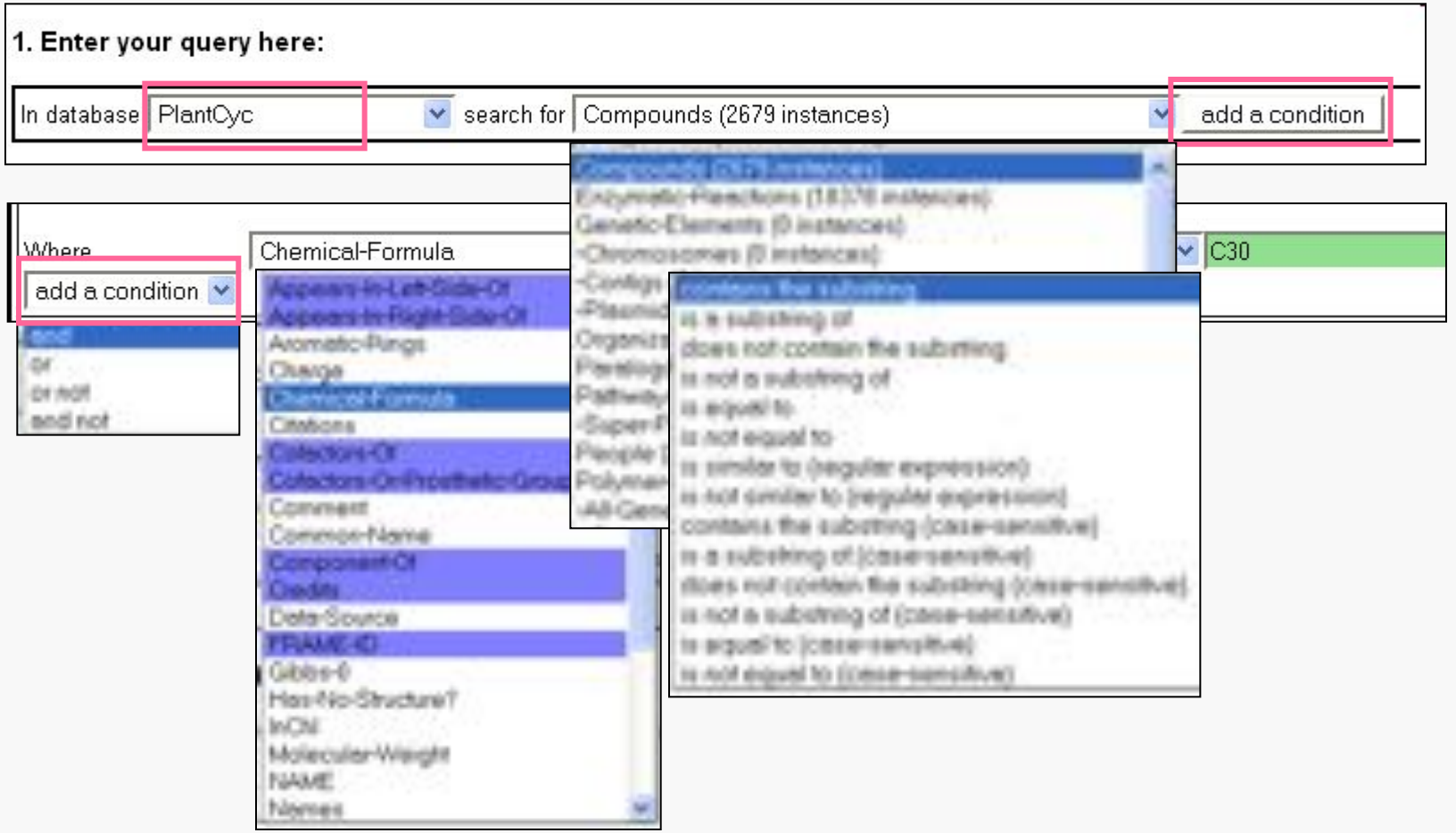
Advanced searching in PMN databases

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



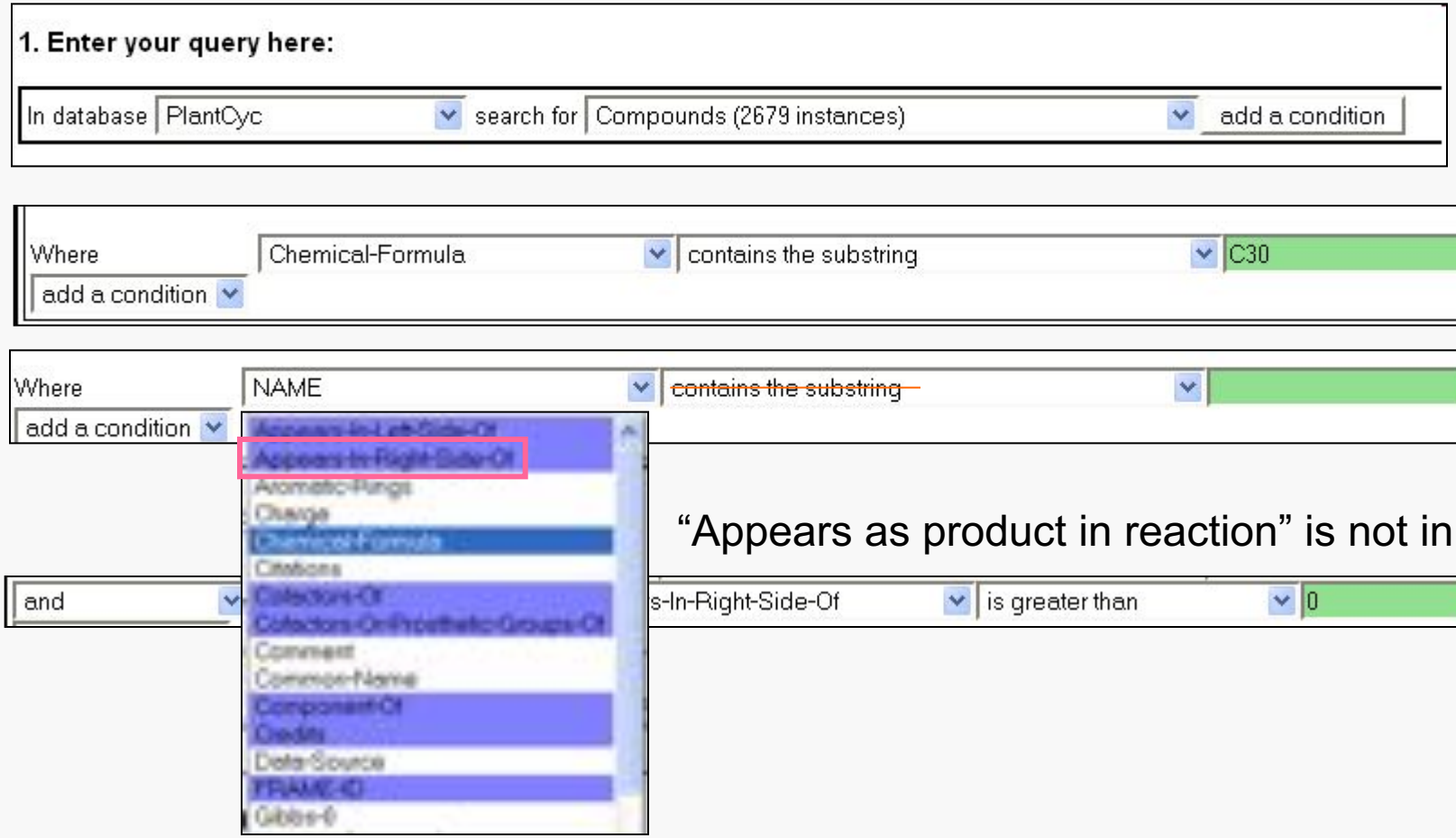
Advanced searching in PMN databases

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



Advanced searching in PMN databases

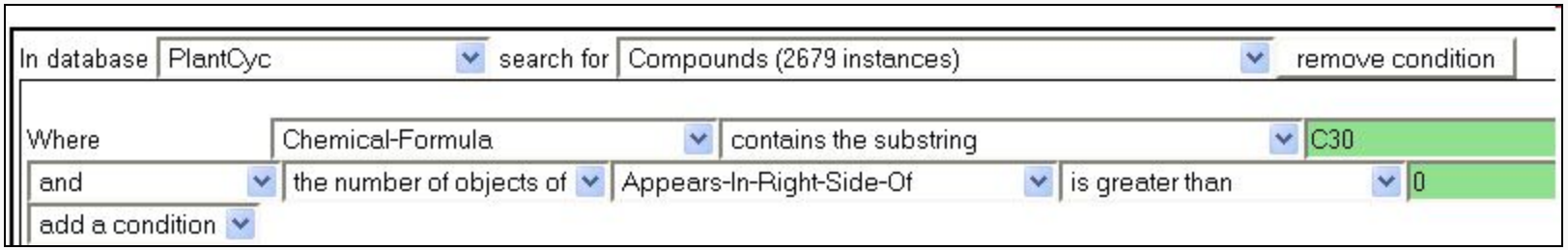
- Find all of the 30-carbon compounds that appear as products in reactions
  - Construct query



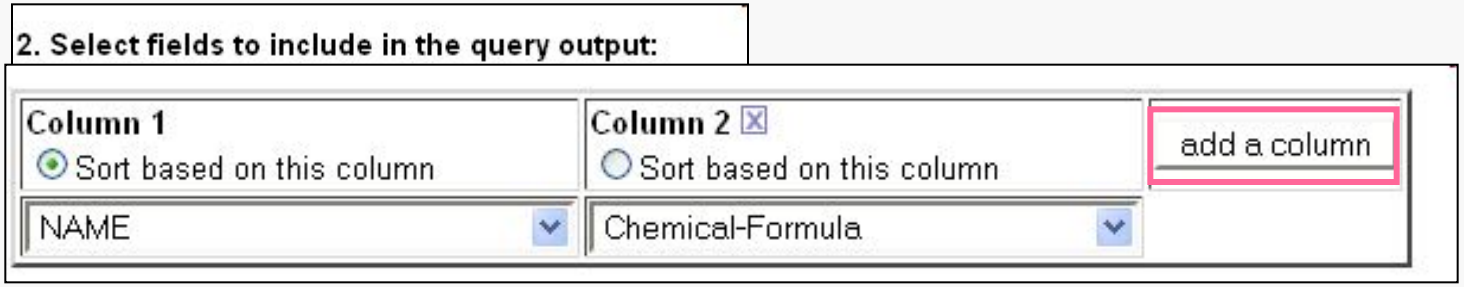
"Appears as product in reaction" is not in the list

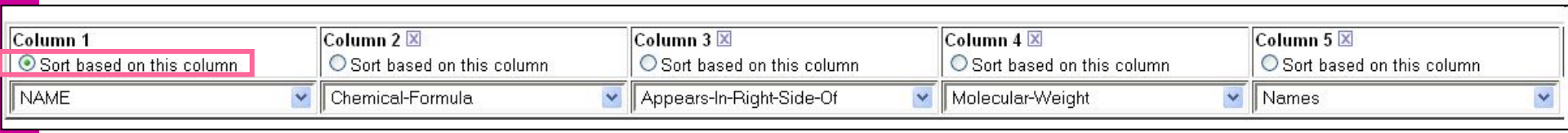
Advanced searching in PMN databases

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



  - Select desired data outputs





Advanced searching in PMN databases

Advanced Query Results

Your query in BioVelo is sort ([(x1^?NAME, x1^?CHEMICAL-FORMULA, x1^?APPEARS-IN-RIGHT-SIDE-OF, x1^?MOLECULAR-WEIGHT, x1^?NAMES) : x1<-PLANT^^Compounds, ("C30" instringci x1^CHEMICAL-FORMULA) & ((#x1^APPEARS-IN-RIGHT-SIDE-OF) > 0)],1)

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)	Column 5 for (x1 ^? NAMES)
α-amyrin	C30H50O1	(S)-2,3-epoxysqualene = α-amyrin	426.724	α-amyrin, viminalol, alpha-Amyrenol
β-amyrin	C30H50O1	(S)-2,3-epoxysqualene = β-amyrin	426.724	β-amyrin, β-amyrenol
(E)-cinnamoyl-CoA	C30H42N7O17P3S1	trans-cinnamate + coenzyme A = (E)-cinnamoyl-CoA + H_2O	897.68	(E)-cinnamoyl-CoA
(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	(S)-2,3-epoxysqualene, squalene 2,3-epoxide, squalene 2,3-oxide, (S)-squalene-2,3-epoxide, 2,3-EDSQ, 2,3-epoxisqualene, oxidosqualene, 2,3-oxidosqualene
24-ethylidenelophenol	C30H50O1	24-methylenelophenol + S-adenosyl-L-methionine = 24-ethylidenelophenol + S-adenosyl-L-homocysteine	426.724	24-ethylidenelophenol, (Z)-24-ethylidenelophenol, citrostadienol
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-formyl-5α-cholesta-8,24-dien-3β-ol
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	4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol
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-(1-methyl-2-pyrrolidinyl)-3-oxobutanoyl-CoA, N-methylpyrrolidineacetoacetyl-CoA
4-coumaroyl-CoA	C30H42N7O18P3S1	coenzyme A + 4-coumarate + ATP = 4-coumaroyl-CoA + diphosphate + AMP	913.679	4-coumaroyl-CoA, p-coumaryl-CoA, 4-coumaryl-CoA, coumaroyl-CoA, p-coumaroyl-CoA

Data and software downloads

- Install a local copy of the Pathway Tools software

Building better databases together

- To submit data, report an error, or ask a question . . .
  - Send an e-mail: curator@plantcyc.org
  - Use data submission “tools”

  - **Meet with me individually at this conference**
    - **Plant Genome Database Outreach Consortium booth: # 427**
    - **Poster: C891**

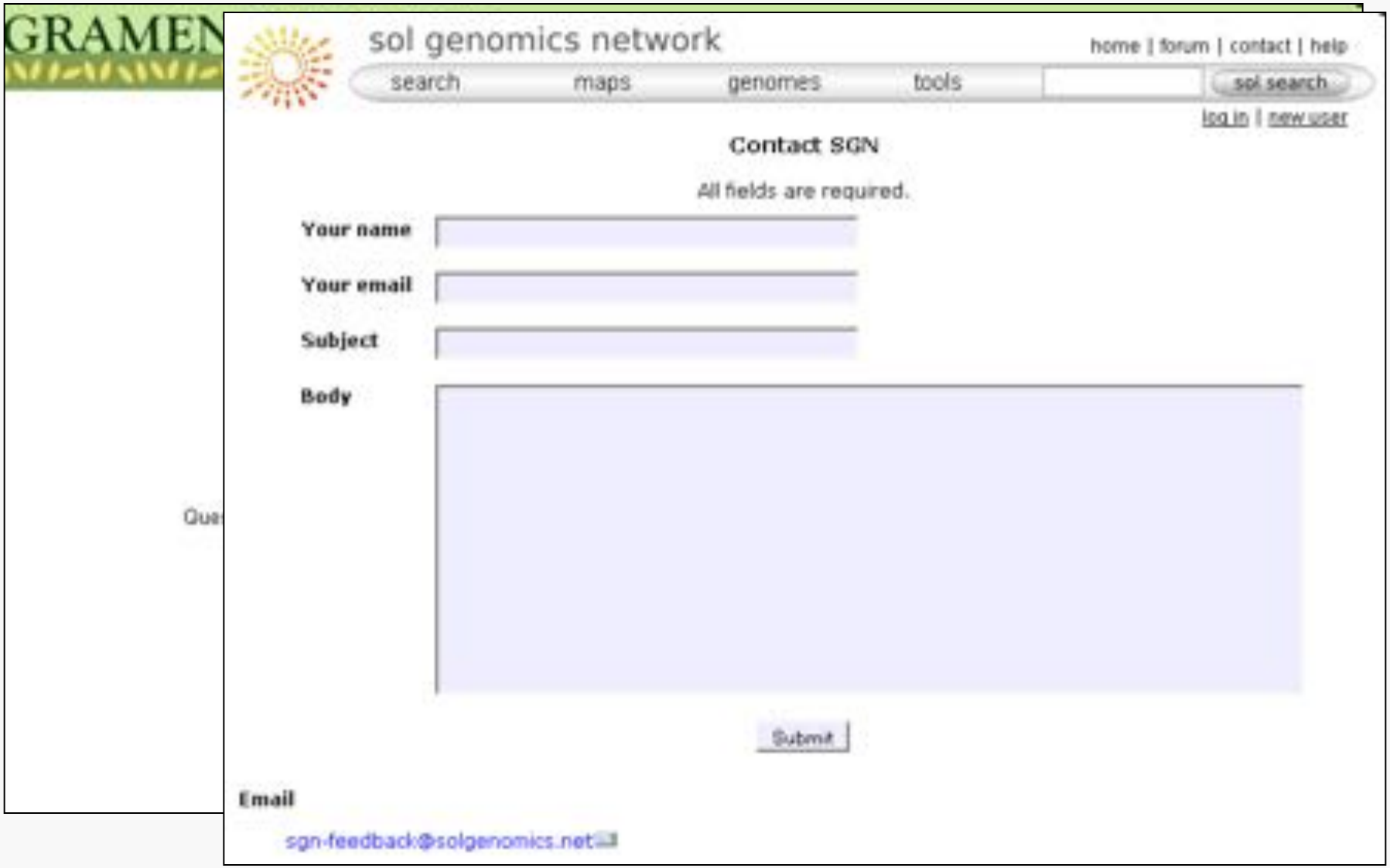
Building better databases together

- PMN future plans
  - Develop a better enzyme annotation pipeline
  - Predict plant metabolic pathways for many species with sequence data
  - Solicit help from experts who work on different species for pathway validation
    - Remove mis-predictions
    - Add missed pathways
  - Potential candidates:
    - wine grape
    - cassava
    - *Selaginella moellendorfii*
    - moss
    - eucalyptus
    - sunflower
    - apple
    - cowpea
    - ice plant
    - papaya
    - **YOUR FAVORITE SPECIES??**

Community gratitude

- We

Building better databases together



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

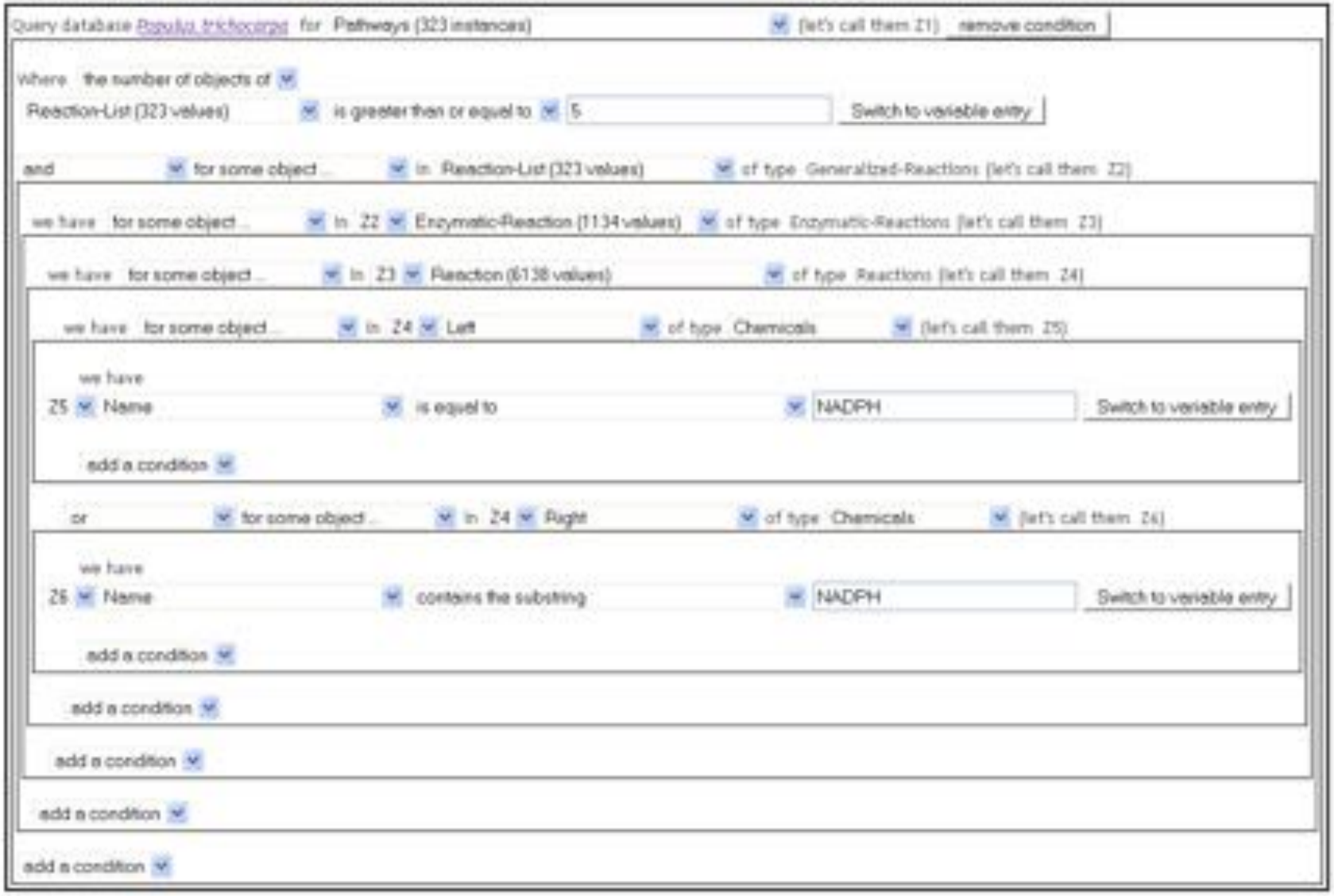
National Science Foundation
WHERE DISCOVERIES BEGIN

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

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