Accessing the Data You Need at the Plant Metabolic Network

kate dreher
biocurator
PMN

The Carnegie Institution for Science
Stanford, CA
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 agricultural and biofuel crops
  - ... many more applications

- These efforts require access to high quality plant metabolism data.
Plant Metabolic Network goals

- Transform published results into data-rich metabolic pathways
- Create and deploy improved methods for predicting enzyme function and metabolic capacity using plant genome sequences
- Facilitate data analysis
- Support research and education
- Provide public resources: www.plantcyc.org
  - PlantCyc
  - AraCyc
  - PoplarCyc
  - ... and 4 NEW databases
  - * CornCyc
  - * SoyCyc
  - * GrapeCyc
  - * CassavaCyc
Introduction

The Plant Metabolic Network (PMN) provides a broad network of plant metabolic pathway databases that contain curated information from the literature and computational analyses about the genes, enzymes, compounds, reactions, and pathways involved in primary and secondary metabolism in plants.

The PMN is funded by the National Science Foundation (Grant #: 1026003 and 0640769), 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

The PMN currently houses one multi-species reference database called PlantCyc and ten species/taxon-specific databases.

- PMN Content Statistics - stats about pathways, enzymes, reactions, compounds, and more

News

Come meet the PMN in Austin!

- Want to learn about the newest PMN data?
- Eager for a demo on data analysis using the Omics viewer?
- Interested in annotating your favorite enzyme?

Then meet with a PMN curator one-on-one at the Outreach Booth (215) during the ASPB Plant Biology 2012 conference.

Additional PMN learning opportunities:

- Plant Informatics Workshop: Saturday - 7:30 PM
- PMN poster (19014): Sunday - 6:30 PM
Plant Metabolic Network collaborators

- SRI International – BioCyc project
  - Provide Pathway Tools Software
  - Maintain and update MetaCyc (multi-kingdom) reference metabolic pathway database

- Other collaborators / contributors include:
  - Sol Genomics Network (SGN) / Boyce Thompson Institute
  - Gramene
  - Maize GDB
  - ChlamyCyc
  - MedicCyc / Nobel Foundation
  - SoyBase
  - PlantMetabolomics group
  - ... and more
PMN data content statistics

- Latest PMN release – March 2012

<table>
<thead>
<tr>
<th></th>
<th>Pathways</th>
<th>Enzymes *</th>
<th>Reactions</th>
<th>Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>PlantCyc 6.0</td>
<td>898</td>
<td>27652</td>
<td>3421</td>
<td>3334</td>
</tr>
<tr>
<td>AraCyc 9.0</td>
<td>502</td>
<td>7100</td>
<td>3320</td>
<td>3210</td>
</tr>
<tr>
<td>PoplarCyc 4.0</td>
<td>354</td>
<td>8691</td>
<td>2220</td>
<td>1596</td>
</tr>
<tr>
<td>CassavaCyc 1.0</td>
<td>323</td>
<td>7178</td>
<td>2103</td>
<td>1524</td>
</tr>
<tr>
<td>CornCyc 1.0</td>
<td>341</td>
<td>10519</td>
<td>2156</td>
<td>1731</td>
</tr>
<tr>
<td>GrapeCyc 1.0</td>
<td>330</td>
<td>5338</td>
<td>2101</td>
<td>1540</td>
</tr>
<tr>
<td>SoyCyc 2.0</td>
<td>412</td>
<td>13094</td>
<td>2528</td>
<td>1987</td>
</tr>
</tbody>
</table>

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

Over 350 species of plants appear in PlantCyc
PMN data content statistics

- Updated versions of AraCyc and PoplarCyc

<table>
<thead>
<tr>
<th></th>
<th>Pathways</th>
<th>Enzymes *</th>
<th>Reactions</th>
<th>Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>PlantCyc 6.0</td>
<td>898</td>
<td>27652</td>
<td>3421</td>
<td>3334</td>
</tr>
<tr>
<td>AraCyc 9.0</td>
<td>502</td>
<td>7100</td>
<td>3320</td>
<td>3210</td>
</tr>
<tr>
<td>PoplarCyc 4.0</td>
<td>354</td>
<td>8691</td>
<td>2220</td>
<td>1596</td>
</tr>
</tbody>
</table>
PMN data content statistics

- Plus FOUR new databases arrived!

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Pathways</th>
<th>Enzymes *</th>
<th>Reactions</th>
<th>Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>PlantCyc 6.0</td>
<td>898</td>
<td>27652</td>
<td>3421</td>
<td>3334</td>
</tr>
<tr>
<td>CassavaCyc 1.0</td>
<td>323</td>
<td>7178</td>
<td>2103</td>
<td>1524</td>
</tr>
<tr>
<td>CornCyc 1.0</td>
<td>341</td>
<td>10519</td>
<td>2156</td>
<td>1731</td>
</tr>
<tr>
<td>GrapeCyc 1.0</td>
<td>330</td>
<td>5338</td>
<td>2101</td>
<td>1540</td>
</tr>
<tr>
<td>SoyCyc 2.0</td>
<td>412</td>
<td>13094</td>
<td>2528</td>
<td>1987</td>
</tr>
</tbody>
</table>
Plant Metabolic Network data

- Pathway-based access to data
Pathway

Enzyme

Gene

Reaction

Compound

Evidence

Codes

PMN pathways

Regulation

Upstream pathway

MultiFun Terms: UNCLASSIFIED

Enzymatic reaction of: cholinephosphotransferase (phosphatidyltransferase)

\[ a \, 1,2-diacylglycerol + CDP-choline \rightleftharpoons \text{a phosphatidylcholine} + \text{CMP} \]

The reaction direction shown, that is, \( A + B \rightleftharpoons \text{C + D} \) versus \( \text{C + D} \rightleftharpoons \text{A + B} \), 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, Ca^{2+}
Searching at the PMN

- Pathway Tools quick search bar

![Pathway Tools quick search bar](image)
Specific search pages
Specific search pages
Specific search pages

PlantCyc Gene/Protein/RNA Search

- Search by gene name or database identifier

- Search by protein name, EC number, or database identifier

- cytokinin

- PROTEIN

- GENE
Specific search pages

PlantCyc Gene/Protein/RNA Search

PlantCyc Query Results

You searched for all gene products whose name contains the string "cytokinin oxidase", and that includes:

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Product Name</th>
<th>Organism</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKX1</td>
<td>cytokinin oxidase</td>
<td>Arabidopsis thaliana</td>
<td>Reaction enhanced in mutant</td>
</tr>
<tr>
<td>CKX1</td>
<td>cytokinin oxidase</td>
<td>Zea mays</td>
<td>Assay of partially-purified protein</td>
</tr>
</tbody>
</table>

- Inferred by computational analysis (6850)
- Inferred by curator (7)
- Inferred from experiment (2204)

Select one or more evidence codes to filter the results to only include gene products that have the corresponding evidence for their function. Disabling all evidence codes will have the same effect as selecting all evidence codes, and the filtering will be done based on evidence code.

Submit Query
Sequence-based search options

- experimental or computational support
- plants only

- experimental support
- all kingdoms
PlantCyc Pathway: ethylene biosynthesis from methionine

- Methionine adenosyltransferase (AT): At-SAM-1
- 1-aminocyclopropane-1-carboxylate synthase (ACS): At-ACS8, At-ACS9, At-ACS11
- 1-aminocyclopropane-1-carboxylate oxidase (ACO): At-ACO2

Methionine biosynthesis:
- L-methionine

Ethylene synthesis:
- 2 H₂O
- L-dehydro-ascorbate
- CO₂
- L-ascorbate

S-adenosyl-L-methionine
- S-methyl-S'-thiodenosine

1-aminocyclopropane-1-carboxylate
- 1-aminocyclopropane-1-carboxylate dehydrase (ACD):
  - Cc-ACD
  - Sol-ACD
Advanced searching in PMN databases

- Advanced search page
  - Allows the construction of very complex queries
Advanced searching in PMN databases

1. Construct query
2. Select output fields
3. Choose file format and retrieve

3. Select query output format:
   - HTML
   - Tab Delimited Text (columns are separated by tabs)

4. Submit Query
A list of 30-carbon compounds that appear as products in reactions:

<table>
<thead>
<tr>
<th>Column 1 for (x1 = ? NAME)</th>
<th>Column 2 for (x1 = ? CHEMICAL-FORMULA)</th>
<th>Column 3 for (x1 = ? APPEARS-IN-RIGHT-SIDE-OF)</th>
<th>Column 4 for (x1 = ? MOLECULAR-WEIGHT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>o-amyln</td>
<td>C30H60O1</td>
<td>(S)-2,3-epoxysqualene = o-amyln</td>
<td>426.724</td>
</tr>
<tr>
<td>l-amyln</td>
<td>C30H60O1</td>
<td>(S)-2,3-epoxysqualene = l-amyln</td>
<td>426.724</td>
</tr>
<tr>
<td>(E)-cinnamoyl-CoA</td>
<td>C30H42N7O17P3S1</td>
<td>trans-cinnamate + coenzyme A = (E)-cinnamoyl-CoA + H2O</td>
<td>997.68</td>
</tr>
<tr>
<td>(S)-2,3-epoxysqualene</td>
<td>C30H60O1</td>
<td>squalene + NADPH + O2 = (S)-2,3-epoxysqualene + NADP+ + H2O, O2 + a reduced acceptor + squalene = (S)-2,3-epoxysqualene + H2O + an acceptor</td>
<td>426.724</td>
</tr>
<tr>
<td>24-ethylidenosphenol</td>
<td>C30H50O1</td>
<td>24-methylenalosphenol + S-adenosyl-L-methionine = 24-ethylidenosphenol + S-adenosyl-L-homocysteine</td>
<td>426.724</td>
</tr>
<tr>
<td>4,4-dimethyl-14a-formyl:5a-cholesta:8,24-dien-3β-ol</td>
<td>C30H48O2</td>
<td>4,4-dimethyl-14a-hydroxymethyl-5a-cholesta:8,24-dien-3β-ol + NADPH + O2 = 4,4-dimethyl-14a-formyl-5a-cholesta:8,24-dien-3β-ol + NADP+ + 2 H2O</td>
<td>440.708</td>
</tr>
<tr>
<td>4,4-dimethyl-14a-hydroxymethyl-5a-cholesta:8,24-dien-3β-ol</td>
<td>C30H50O2</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14a-hydroxymethyl-5a-cholesta:8,24-dien-3β-ol + NADP++ + H2O</td>
<td>442.724</td>
</tr>
<tr>
<td>4-(1-methyl-2-pyrrolidinyl)-3-oxobutanylo-CoA</td>
<td>C30H49N8O18P3S1</td>
<td>1-methylpyrrolidine-2-acetyl-CoA + acetyl-CoA = 4-(1-methyl-2-pyrrolidinyl)-3-oxobutanylo-CoA + coenzyme A</td>
<td>934.741</td>
</tr>
<tr>
<td>4-coumaryl-CoA</td>
<td>C30H42N7O18P3S1</td>
<td>coenzyme A + 4-coumarate + ATP = 4-coumaryl-CoA + diphosphate + AMP</td>
<td>913.672</td>
</tr>
</tbody>
</table>
Enhanced comparative analysis opportunities
Enhanced comparative analysis opportunities

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

- Organisms: breakdown by principal data types.
- Reactions: breakdown 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.
- Genes/Proteins: breakdown of genes by product type, ontology, annotations; breakdown of proteins, 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
- Vitis vinifera
- Glycine max
- Papulus trichocarpa
- Zea mays mays
- Manihot esculenta esculenta

Submit  Clear All
Enhanced comparative analysis opportunities

Table 4: Distribution of Isozymes across SMM Reactions

Isozymes are multiple enzymes that catalyze the same reaction. This table shows the number of SMM Reactions that have the specified number of enzymes. For example, row 1 counts the number of SMM Reactions that are catalyzed by only a single enzyme, row 2 counts the number of SMM Reactions that are catalyzed by exactly two enzymes, the seventh row counts the number of SMM Reactions that are catalyzed by more than 10 enzymes, and so on. To compare the number of reaction across all organisms, expand the row with all multiple isozymes.

Unfortunately, if protein complexes have not been created for a database, then two subunits of a single enzyme may be mistakenly classified as isozymes, which will be misleading. Protein complex creation is a Pathologic task requiring manual assistance, and has not been done for the vast majority of the BioCyc websites. We recommend you give credence to this statistic only for databases that have been manually curated.

<table>
<thead>
<tr>
<th>Number of Enzymes per Reaction</th>
<th>Arabidopsis thaliana</th>
<th>PlantCyc</th>
<th>V. vinifera</th>
<th>G. max</th>
<th>P. trichocarpa</th>
<th>T. aestivum</th>
<th>L. esculentum</th>
<th>H. esculentum</th>
<th>L. esculentum esculentum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>475 (25%)</td>
<td>548 (23%)</td>
<td>779 (21%)</td>
<td>131 (7%)</td>
<td>215 (11%)</td>
<td>147 (11%)</td>
<td>271 (15%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>402 (13%)</td>
<td>202 (19%)</td>
<td>247 (16%)</td>
<td>230 (11%)</td>
<td>230 (12%)</td>
<td>196 (13%)</td>
<td>200 (11%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>245 (28%)</td>
<td>137 (23%)</td>
<td>154 (25%)</td>
<td>158 (7%)</td>
<td>147 (8%)</td>
<td>118 (6%)</td>
<td>162 (9%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>184 (45%)</td>
<td>88 (23%)</td>
<td>191 (35%)</td>
<td>145 (7%)</td>
<td>103 (5%)</td>
<td>90 (5%)</td>
<td>139 (8%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>110 (45%)</td>
<td>70 (25%)</td>
<td>59 (35%)</td>
<td>104 (5%)</td>
<td>113 (6%)</td>
<td>46 (4%)</td>
<td>87 (5%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Enhanced comparative analysis opportunities

<table>
<thead>
<tr>
<th>EC Category: 4 — Lysine</th>
<th>AnaCyc</th>
<th>PlantCyc</th>
<th>M. viridiflora</th>
<th>G. max</th>
<th>P. trichocarpa</th>
<th>Z. mays</th>
<th>M. esculenta</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.1.1: 3-ketoacetyl-2,3-dihydroxyacetone + H^+ -&gt; 2-trimethylamine + CO2</td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1.1.1: D-ribulose-1,5-bisphosphate + O_2 -&gt; 2-phosphoglycolate + 2-phosphoglycerate + 2 H^+</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>4.1.1.1: D-ribulose-1,5-bisphosphate + O_2 -&gt; 3-phosphoglycerate + 2-phosphoglycolate</td>
<td>×</td>
<td></td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>4.1.1.1: L-serine + H^+ -&gt; ethanolamine + CO2</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>4.1.1.1: UDP-D-glucose + H^+ -&gt; UDP-D-glucose + CO2</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>4.1.1.15: L-ascorbate + H^+ -&gt; CO2 + 4-aminobutyrate</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>4.1.1.17: L-ornithine + H^+ -&gt; CO2 + putrescine</td>
<td></td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>4.1.1.18: L-lysine + H^+ -&gt; CO2 + cadaverine</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Interpreting Omics results on the Metabolic Map
Cellular Overview of Arabidopsis thaliana col

Pan in/zoom in: the entire diagram by holding the left mouse button, click on an object for more info, right-click (Ctrl-click for Mac users) for menu.
Cellular Overview of Arabidopsis thaliana co!

Pan left/right/up/down the entire diagram by holding the left mouse button; click on an object for more info, right-click to toggle Show Legend.
Data analysis with the Metabolic Map / Omics Viewer

- Display experimental data on a 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
Booth 215:
Sun: 12:30 - 1:30
Mon: 10:00 - 1:00
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
- See Omics Viewer results on pathway pages
On the horizon . . .

- Develop a better enzyme annotation pipeline
- Predict plant metabolic pathways for additional species with predicted proteomes
  - Almost ready . . . :
    - moss
    - papaya
    - Chlamydomonas
    - Selaginella
- Solicit help from experts for pathway validation
  - Remove mis-predictions
  - Add missed pathways
- Increase coverage of experimentally verified enzymes involved in secondary metabolism from diverse species
Questions, contributions, volunteering to validate?

- To submit data, report an error, or ask a question . . .
  - Send an e-mail: curator@plantcyc.org
  - Use our feedback form:
    - Stop by the Plant Genome Resources Outreach Booth (215) CHOCOLATE!
    - Visit the PMN poster – P19014
    - Schedule an individual meeting with me
Community gratitude

- We thank you publicly!
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!

curator@plantcyc.org

www.plantcyc.org
PMN Acknowledgements

Curators:
- kate dreher

Post-docs:
- Lee Chae
- Ricardo Nilo Poyanco

Interns
- Varun Dwaraka
- Tam Tran
- Caryn Johansen

Tech Team Members:
- Bob Muller (Manager)
- Larry Ploetz (Sys. Admin)
- Shanker Singh
- Bill Nelson

Rhee Lab Members:
- Flavia Bossi
- Hye-in Nam

Sue Rhee (PI)
Eva Huala (Co-PI)

Collaborators:
SRI
- Peter Karp
- Ron Caspi
- Suzanne Paley
- SRI Tech Team

MaizeGDB
- Mary Schaeffer
- Lisa Harper
- Jack Gardiner
- Taner Sen
- Lukas Mueller (SGN)
- Gramene and MedicCyc
PMN Acknowledgements

Curators:
- kate dreher

Post-docs:
- Lee Chae
- Ricardo Nilo Poyanco

Interns
- Varun Dwaraka
- Tam Tran
- Caryn Johansen

Tech Team Members:
- Bob Muller (Manager)
- Larry Ploetz (Sys. Admin)
- Shanker Singh
- Bill Nelson

Rhee Lab Members:
- Flavia Bossi
- Hye-in Nam

PMN Alumni:
- A. S. Karthikeyan (curator)
- Christophe Tissier (curator)
- Hartmut Foerster (curator)
- Damian Priamurskiy (intern)
- Ricardo Leitao (intern)
- Michael Ahn (intern)
- Purva Karia (intern)
- Anuradha Pujar (SGN curator)

Sue Rhee (PI)
Eva Huala (Co-PI)

Collaborators:

SRI
- Peter Karp
- Ron Caspi
- Suzanne Paley
- SRI Tech Team

MaizeGDB
- Mary Schaeffer
- Lisa Harper
- Jack Gardiner
- Taner Sen
- Lukas Mueller (SGN)
- Gramene and MedicCyc

Tech Team Alumni
- Anjo Chi
- Cynthia Lee
- Tom Meyer
- Vanessa Kirkup
- Chris Wilks
- Raymond Chetty
We're here to help . . .

curator@plantcyc.org

www.plantcyc.org

Booth 215:
Plant Resources Outreach
Sunday, 12:30 – 1:30
Monday, 10:00 – 1:00

Poster P19014:
Sunday: 6:30 – 7:30 PM