Department of Plant Biology
Putting the PMN (and TAIR) to work for you:

Tips and Techniques for Accessing Data for Plant Biology Research

Kate Dreher

TAIR, AraCyc, PMN
Carnegie Institution for Science

kadreher@arabidopsis.org
Overview

- Introduction to the PMN
- Accessing data in the PMN
- Case study: Putting the PMN and TAIR to work for you
Welcome to the PMN

- The Plant Metabolic Network (PMN) maintains a set of metabolic pathway databases for Arabidopsis and other plants (AraCyc, PlantCyc, etc.)

- www.plantcyc.org

- Curators and programmers at the PMN:
  - Collect and store metabolic pathway information
  - Provide tools to analyze data
  - Work to generate new metabolic pathway databases for crops and other important plants
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:

<table>
<thead>
<tr>
<th>PGDB</th>
<th>Plant</th>
<th>Source</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiceCyc **</td>
<td>Rice</td>
<td>Gramene</td>
<td>some curation</td>
</tr>
<tr>
<td>SorghumCyc</td>
<td>Sorghum</td>
<td>Gramene</td>
<td>no curation</td>
</tr>
<tr>
<td>MedicCyc **</td>
<td>Medicago</td>
<td>Noble Foundation</td>
<td>some curation</td>
</tr>
<tr>
<td>LycoCyc **</td>
<td>Tomato</td>
<td>Sol Genomics Network</td>
<td>some curation</td>
</tr>
<tr>
<td>PotatoCyc</td>
<td>Potato</td>
<td>Sol Genomics Network</td>
<td>no curation</td>
</tr>
<tr>
<td>CapCyc</td>
<td>Pepper</td>
<td>Sol Genomics Network</td>
<td>no curation</td>
</tr>
<tr>
<td>NicotianaCyc</td>
<td>Tobacco</td>
<td>Sol Genomics Network</td>
<td>no curation</td>
</tr>
<tr>
<td>PetuniaCyc</td>
<td>Petunia</td>
<td>Sol Genomics Network</td>
<td>no curation</td>
</tr>
<tr>
<td>CoffeaCyc</td>
<td>Coffee</td>
<td>Sol Genomics Network</td>
<td>no curation</td>
</tr>
</tbody>
</table>

** Significant numbers of genes from these databases have been integrated into PlantCyc
### PMN database content statistics

<table>
<thead>
<tr>
<th></th>
<th>PlantCyc 4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathways</td>
<td>685</td>
</tr>
<tr>
<td>Enzymes</td>
<td>11058</td>
</tr>
<tr>
<td>Reactions</td>
<td>2929</td>
</tr>
<tr>
<td>Compounds</td>
<td>2966</td>
</tr>
<tr>
<td>Organisms</td>
<td>343</td>
</tr>
</tbody>
</table>

*Species (A - D) [back to top]*

- Abies grandis (yellow fir, white fir, silver fir, lowland fir, grand fir)
- Acer pseudoplatanus
- Adonis aestivalis
- Adonis annua (pheasants-eye, blood-drop)
- Ajuga reptans
- Albuca juniperina
- Allium
- Allium cepa
- Allium sativum
- Allium tuberosum
- Aloe arborescens
- Amaryllis muscaria
- Anemone majus
- Amorpha fruticosa
- Anabaena variabilis
- Anchusa officinalis
- Anisodus acutilobus
- Armeria maritima (snail-dragon)
- Apium graveolens
- Aquilegia vulgaris
- Arableopsis lyrata (lyrata rockress)
- Arabidopsis thaliana, Col (thale cress, mouse-ear cress)
- Arabidopsis thaliana, Ler (thale cress, mouse-ear cress)
- Arachis hypogaea
- Artemisia annua (sweet Annie)
- Atriplex hortensis
- Atropa belladonna (deadly nightshade)
- Berberis sanguinea (cutleaf doll)
- Berberis beani
- Berberis amoena
- Berberis wilsoniae
Getting information from PMN pathway pages

- Better . . . but what about compound structures?
  - Keep clicking on “More Detail” – sometimes several times
Getting information from PMN pathway pages

- **Pathway**
- **Enzyme**
- **Gene**
- **Reaction**
- **Compound**
- **Evidence Codes**
- **Regulation**
- **Upstream pathway**
- **Pathway**

Example: PlantCyc Pathway: choline biosynthesis III

- **Enzyme**
- **Gene**
- **Reaction**
- **Compound**
Getting information from PMN pathway pages

Superclasses: Biosynthesis, Fatty Acids and Lipids Biosynthesis, Choline Biosynthesis

Species Data Available for: Arabidopsis thaliana, Glycine max, Medicago truncatula, Oryza sativa, Japonica Group, Pisum sativum, Populus trichocarpa, Ricinus communis

Summary:

General information: Choline is a fundamental metabolite in plants because of its contribution to the synthesis of the membrane phospholipid phosphatidylcholine, which accounts for 40 to 60% of lipids in non-plastid plant membranes [Mou02]. Choline is also a precursor for the formation of glycine betaine (glycine betaine biosynthesis) in certain plants such as spinach, where this osmoprotectant is accumulated and confers also tolerance to salinity, drought, and other environmental stresses. In addition choline has been recognized as an essential nutrient for humans [McNei01].

The choline biosynthetic pathway enables plants to decouple choline synthesis from lipid metabolism (Kennedy pathway - diacylglycerol biosynthesis) and provides them with the metabolic flexibility to adapt to environmental conditions where large and variable amounts of choline are beneficial for survival [Roeni01].

Pathway information: The first step in choline biosynthesis is the direct decarboxylation of serine to ethanolamine [Roeni01], which is catalyzed by a serine decarboxylase unique to plants [Roeni01]. Ethanolamine is widely recognized as the entrance compound to choline biosynthesis.

References:

Curator09. Curator (2009). "Following the initial computational build of PoplarCyc in 2009, pathways were validated by PMN curators based on a preliminary literature search. For pathways that lacked direct experimental support, curators considered a number of factors to judge the validity of the predicted pathways including: 1) critical compound(s) in the pathway are found in a Populus species, 2) a Populus trichocarpa gene is predicted to catalyze a critical or unique reaction of the pathway, or 3) the pathway is expected to exist in all plants."


Getting information from PMN pathway pages
PMN compound pages

Compound: CDP-choline

Synonyms: citicoline, citocholine, chidose, cyticholine, cytidine 5'-diphosphocholine, cytidine diphosphate choline

Classification(s):
- a nucleic acid component
- a base derivative
- a pyrimidine-related compound

Empirical Formula: C_{14}H_{24}N_{4}O_{7}P_{2}
Molecular Weight: 489.34 daltons

Smiles: C(OP)(O(=O)OP)(O(=O)OC(|N+([C(\{C\})C]|C)C)OC(C(=O)(C10)OC)C10)n(=O)(\{C\}cc2)nc3c(nc3c2c)oc
Unification Links: CAS:987-78-9
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
  - phosphatidylcholine + CTP + CDP-choline + diphosphate
- choline biosynthesis III:
  - phosphatidylcholine + CTP + CDP-choline + diphosphate

Appears as Reactant
Appears as Product
Getting information from PMN pathway pages
**PMN enzyme pages**

*Arabidopsis* Enzyme: phosphatidyltransferase

Enzymatic reaction of: cholinephosphotransferase (phosphatidyltransferase)

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

Enzymatic reaction of: ethanolaminephosphotransferase (phosphatidyltransferase)

\[
\text{a,1,2-diacylglycerol} + \text{CMP-ethanolamine} \rightarrow \text{an L-1-phosphatidyl-ethanolamine} + \text{CMP}
\]
**Arabidopsis Enzyme: phosphatidyltransferase**

**Pathway(s):**
- Choline biosynthesis
- Phospholipid biosynthesis

**Summary:**
Two cDNA's from Arabidopsis thaliana (AtAAP1, AtAAP2) have been isolated from an Arabidopsis cDNA library using the AAPT cDNA from soybean as a heterologous hybridization probe. Both cDNA's encode aminoacoholphosphotransferases involved in the nucleotide pathway of the biosynthesis of phosphatidylethanolamine (PE) and phosphatidylcholine (PC) as demonstrated by gene expression in yeast mutants lacking those enzyme activities. It has been demonstrated that both AtAAP1 and AtAAP2 convert CDP-ethanolamine and CDP-choline into the corresponding phosphatidylalcohols although with slight differences regarding the substrate preference. AtAAP1 showed a higher preference for CDP-choline over CDP-ethanolamine in comparison to AtAAP2 and was also inhibited to a lesser degree by Ca²⁺ and Cytidine monophosphate ([CMP] than AtAAP1 [Goode99]). Both enzymes (AtAAP1, AtAAP2) were able to catalyze the reverse reaction supporting the proposal that diacylglycerol, involved as substrate in both PC and triacylglycerol biosynthesis (triacylglycerol biosynthesis), is in equilibrium with PC and maintains this equilibrium via the reversibility of the cholinephosphotransferase reaction (Stack95). AtAAP1 and AtAAP2 seem to represent the only aminoacoholphosphotransferases in Arabidopsis as concluded from the Southern blotting patterns. Both polypeptides contain seven membrane spanning regions as shown by their hydrophathy files. This finding is in agreement with other AAPT's isolated so far and confirms their localization in membranes (Goode99).

**Inhibitors, Kinetic Parameters, etc.:**
- **Inhibitors (Allosteric):** CMP [Goode99]
- **Inhibitors (Uncompetitive):** CDP [Goode99]

**Primary Physiological Regulators of Enzyme Activity:**
- CMP

**References:**
Getting information from PMN pathway pages

Gene → To TAIR, Gramene, SGN, etc. . .
Searching in the PMN databases

- PMN quick search bar

![Screenshot of PMN quick search bar with choline entered]
Searching in PMN databases

- Pathways: Description of the pathway, including enzymes, genes, and other regulatory elements.
  - Choline biosynthesis I
  - Choline biosynthesis II
  - Choline biosynthesis III
  - Phosphatidylcholine biosynthesis
  - Supergene of phosphatidylethanolamine
  - Supergene of choline

- Proteins: Detailed comments and citations, subunit structure, cofactors, activators, and inhibitors (for enzymes), depiction of regulation (for transcription factors).
  - Choline kinase
  - Choline monooxygenase

- Genes: Depiction of its operon product information, corresponding protein.
  - Choline monooxygenase

- Compounds: Structural information, links to all reactions and pathways in which the compound participates.
  - 1-O-sinapoyl-D-glucose + choline = Q-sinapoylcholine + D-glucose
  - choline + ATP = phosphoryl-choline + ADP
  - 1,2-diacylglycerol + CDP-choline = a phosphatidylcholine + CMP
  - a phosphatidylcholine + H2O = a 1-acyl-2-lyso-glycero-phosphocholine + a carboxylate
  - a phosphatidylcholine + H2O = an L-phosphatidate + choline
  - phosphoryl-choline + CTP = CDP-choline + diphosphate

- Reactions: Reaction equation with chemical structures, links to all enzymes that catalyze the reaction, and all pathways in which the reaction participates.
Specific search pages
Additional search options
- **PlantCyc, AraCyc, PoplarCyc**
  - Enzymes: include enzymes with available sequence information from each database

- **Reference Enzymes**:
  - Includes enzymes with experimental support from both plant and non-plant species
Finding enzymes through BLAST

PlantCyc Enzyme: 1-aminocyclopropane-1-carboxylate synthase

Species: Arabidopsis thaliana

Summary:
When recombinantly expressed in E. coli, this enzyme was shown to have ACS activity [Yamagami03].

Gene: ACS11

Sequence Length: 1825/3 AAs

Unification Links: Phytozome Plant Orthologs AT4G08040.1

Gene-Reaction Schematic:

Enzymatic reaction of: 1-aminocyclopropane-1-carboxylate synthase

$S$-adenosyl-$L$-methionine $\leftrightarrow S$-methyl-$S$-thioadenosine + 1-aminocyclopropane-1-carboxylate

The reaction direction shown, that is, $A + B \leftrightarrow C + D$ versus $C + D \leftrightarrow A + B$, is in accordance with the Enzyme Commission system.

Reversibility of this reaction is unspecified.

In Pathways: ethylene biosynthesis from methionine

Citations: [Yamagami03]
<table>
<thead>
<tr>
<th>Organism</th>
<th>Evidence Glyph</th>
<th>Enzymes and Genes for choline biosynthesis III</th>
</tr>
</thead>
</table>
| **AraCyc_col**    | ![Evidence Glyph](image1) | EC#2.7.7.15: choline-phosphate cytidylyltransferase: AT4G15130  
|                   |                | choline-phosphate cytidylyltransferase: AT2G32260  
| P. trichocarpa     | ![Evidence Glyph](image2) | EC#2.7.8.2: phosphatidylintransferase: AT3G25585  
|                   |                | phosphatidylintransferase: AT1G13560  
|                   |                | EC#3.1.4.4: phospholipase D: AT1G52570  
|                   |                | phospholipase D: AT1G555180  
|                   |                | phospholipase D: AT3G05630  
|                   |                | phospholipase D: AT4G00240  
|                   |                | phospholipase D: AT4G11830  
|                   |                | phospholipase D: AT4G11840  
|                   |                | phospholipase D: AT5G25370  
|                   |                | phospholipase D: PLD alpha  
|                   |                | phospholipase D: PLD beta  
|                   |                | phospholipase D: PLD gamma  
|                   |                | phospholipase D: PLD delta  
|                   |                | phospholipase D: PLD zeta  
|                   |                | EC#2.7.7.15: None  
|                   |                | EC#2.7.8.2: diacylglycerol cholinephosphotransferase: JGI-225724  
|                   |                | diacylglycerol cholinephosphotransferase: JGI-720905  
|                   |                | EC#3.1.4.4: phospholipase D: JGI-811801  
|                   |                | phospholipase D: JGI-781949  
|                   |                | phospholipase D: JGI-810176  
|                   |                | phospholipase D: JGI-593768  
|                   |                | phospholipase D: JGI-827396  
|                   |                | phospholipase D: JGI-578949  
|                   |                | phospholipase D: JGI-730956  
|                   |                | phospholipase D: JGI-240457  
|                   |                | phospholipase D: JGI-833366  
|                   |                | phospholipase D: JGI-417354  
|                   |                | phospholipase D: JGI-763496  
|                   |                | phospholipase D: JGI-559591  
|                   |                | phospholipase D: JGI-415367  
|                   |                | phospholipase D: JGI-550827  
|                   |                | phospholipase D: JGI-755219  
|                   |                | phospholipase D: JGI-180605  
|                   |                | phospholipase D: JGI-829577  |
Visualizing OMICS data

- Overlay “pre-cleaned” quantitative data sets on a metabolic map
  - Gene transcription data
  - Proteomic data
  - Metabolomic data

- Only available for single-species databases, not PlantCyc

- Demonstrations available from 3:30 – 5:30 PM!
Visualizing OMICs data
Visualizing OMICs data

Animation feature is available!
Case study: Jasmonic acid biosynthesis

- You are studying jasmonic acid biosynthesis in your favorite plant

- You want to identify potential orthologs for all of the Arabidopsis enzymes associated with the pathway
Case study: Jasmonic acid biosynthesis
Case study: Jasmonic acid biosynthesis

Take this gene list to TAIR to get sequences
Case study: Jasmonic acid biosynthesis

- Obtain protein sequences for all of the enzymes
Case study: Jasmonic acid biosynthesis

- Obtain protein sequences for all of the enzymes
Case study: Jasmonic acid biosynthesis

- Blast enzymes against all Genbank Plant proteins in TAIR

- Or use UniProt, Genbank, your species-specific database, etc.
Putting the PMN and TAIR to work for you

- Use the PMN to learn more about metabolic pathways
- Use TAIR to find detailed information for specific genes / proteins
- Use TAIR and the PMN to enhance your plant biology research
- If you’re having trouble getting any information you want . . .
We are here to help!

www.arabidopsis.org
curator@arabidopsis.org

www.plantcyc.org
curator@plantcyc.org
We appreciate YOUR help!
PMN and TAIR Acknowledgements

Current Curators:
- Tanya Berardini (lead curator)
- Philippe Lamesch (lead curator)
- Donghui Li (curator)
- Dave Swarbreck (former lead curator)
- Debbie Alexander (curator)
- A. S. Karthikeyan (curator)
- Marga Garcia (curator)
- Leonore Reiser

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

PMN project post-doc
- Lee Chae

Sue Rhee (PI - PMN)
Eva Huala (PI-TAIR)
Peifen Zhang (Director-PMN)

PMN 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
We are here to help!

www.plantcyc.org
curator@plantcyc.org

www.arabidopsis.org
curator@arabidopsis.org