Putting the Plant Metabolic Network (PMN) to work for you

Kate Dreher

curator

PMN/TAIR
Presentation plan

- Introduction and content overview
- Search tools and tips
- Cross-species analyses
- OMICs data visualization
- Data and software downloads
- Creating new PMN content
What is the PMN?

- PMN = The Plant Metabolic Network
  - Created in 2008

- Funded by the National Science Foundation
What is the PMN?

“A Network of Plant Metabolic Pathway Databases and Communities”

Major goals:

- Create individual metabolic pathway databases for many plant species, e.g. AraCyc
- Create PlantCyc - a comprehensive multi-species metabolic pathway database
- Create a plant metabolism database “pipeline”: Newly annotated DNA sequences . . . to a set of predicted metabolic pathways
- Create a website to bring together researchers working on plant metabolism

PMN website: www.plantcyc.org
What is in the PMN?

- Databases contain detailed information about:
  - Pathways
  - Enzymes
  - Reactions
  - Compounds
  - Genes

- Data are entered and displayed using Pathway Tools software
  - Peter Karp, *et al*, SRI International

- Pathways are generated through:
  - Manual curation when a curator reads scientific literature
  - Computational predictions made by the Pathologic software
What is in the PMN?

- Databases created and maintained by the PMN
  - PlantCyc, AraCyc, PoplarCyc (new in October 2009!)

- Other plant databases accessible through 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
What is in the PMN?

- Most recent release of PMN databases: October 15, 2009

<table>
<thead>
<tr>
<th>Database contents</th>
<th>PlantCyc 3.0</th>
<th>AraCyc 6.0</th>
<th>PoplarCyc 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base pathways* (no superpathways)</td>
<td>643</td>
<td>360</td>
<td>285</td>
</tr>
<tr>
<td>Experimentally-supported (EV-EXP)</td>
<td>585</td>
<td>299</td>
<td>25</td>
</tr>
<tr>
<td>Enzymes** (monomers and complexes)</td>
<td>10964</td>
<td>5501</td>
<td>3434</td>
</tr>
<tr>
<td>in reactions (EV-EXP) in pathways</td>
<td>1974</td>
<td>858</td>
<td>1</td>
</tr>
<tr>
<td>Reactions</td>
<td>2709</td>
<td>2323</td>
<td>1668</td>
</tr>
<tr>
<td>in pathways with enzymes (EV-EXP)</td>
<td>1614</td>
<td>829</td>
<td>2</td>
</tr>
<tr>
<td>Compounds</td>
<td>2679</td>
<td>2630</td>
<td>1363</td>
</tr>
<tr>
<td>Organisms</td>
<td>376</td>
<td>1</td>
<td>1***</td>
</tr>
<tr>
<td>Citations</td>
<td>4803</td>
<td>2691</td>
<td>903</td>
</tr>
</tbody>
</table>

- How can I access all this information?
Searching in PMN databases

- Quick search bar

- Provides access to all PMN-created databases
- Allows Google-based text-mining of summaries, comments, etc.
  - For example, search “plant defense”
- PMN searches:
  - Match partial words
  - Search across all fields (compound, enzyme, etc.)
  - Return a list of items grouped by data type
Searching in PMN databases

- **Pathways**
  - choline biosynthesis I
  - choline biosynthesis II
  - phosphatidylcholine biosynthesis
  - superpathway of phosphatidylcholine
  - superpathway of choline biosynthesis

- **Proteins**
  - choline kinase (phosphatidyl)
  - choline kinase (phosphoryl)
  - choline dehydrogenase
  - choline dehydrogenase (phosphoryl)
  - choline acetyltransferase

- **Compounds**
  - choline
  - 1-O-sipapoyl-D-glucose + choline = O-sipapoylcholine + β-D-glucose
  - choline + ATP + phosphocholine + ADP
  - a 1,2-diacetylglycerol + CDP-choline = a phosphatidylcholine + CMP
  - a phosphatidylcholine + H2O = a 1-acyl-2-lyso-glycerylphosphocholine + a carboxylate
  - a phosphatidylcholine + H2O = an L-phosphatidate + choline
  - phosphoryl-choline + CTP = CDP-choline + diphosphate

- **Reactions**
  - choline monooxygenase
  - sphingomyelinase
  - LOC_002602201
  - cholinephosphotransferase
  - AT2G32260
  - cholinephosphotransferase
  - AT4G15130
  - cholinephosphotransferase
  - LOC_002602201
  - cholinephosphotransferase

- **Genes**
  - depiction of its operon
  - product information is corresponding protein

- **Protein pages contain**
  - Detailed comments and citations; subunit structure; cofactors, activators, and inhibitors (for enzymes), depiction of regulation (for transcription factors).
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
Getting information from PMN pathway pages

**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 [McNe01].

The choline biosynthetic pathway enables plants to decouple choline synthesis from lipid metabolism (Kennedy pathway - triacylglycerol biosynthesis) and provides them with the metabolic flexibility to adapt to environmental conditions where large and variable amounts of choline are beneficial for survival [Rotten01].

**Pathway information:** The first step in choline biosynthesis is the direct decarboxylation of serine to ethanolamine [Rotten01], which is catalyzed by a serine decarboxylase unique to plants [Rotten02]. Ethanolamine is widely recognized as the entrance compound to choline biosynthesis.

**References**

Curator09. Curator (2009). “Following the initial computational build of PoplarCyc in 2008, 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

Compound
Compound: CDP-choline

Synonyms:
citicoline, citicholine, citidines, cyticholine, cytidine 5’-diphosphocholine, cytidine diphosphate choline

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

Molecular Weight / Formula:
Empirical Formula: C_{14}H_{27}N_{4}O_{7}P_{2}
Molecular Weight: 489.34 daltons

Smiles: C(OP)(O(h=0)OP(O)(h=0)OCN)(N+)(C)(C)(C)C1(O)C(O)C1OY)(C(=O)O)N(c=0)(c(c)c(c'))
Unification Links: CAS:987-75-0
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
  - a,1,2-diacylglycerol + CDP-choline + a phosphatidylcholine + CMP

In Pathway Reactions as a Product:
- phospholipid biosynthesis
  - phosphatidylcholine + CTP = CDP-choline + diphosphate
Getting information from PMN pathway pages

Enzyme
**PMN enzyme pages**

*Arabidopsis* Enzyme: phosphatidyltransferase

**Multifunctional protein**

![Gene-Reaction Schematic](image)

**Enzymatic reaction of cholinephosphotransferase (phosphatidyltransferase)**

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

**Enzymatic reaction of ethanolaminephosphotransferase (phosphatidyltransferase)**

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

Pathway(s)

Inhibitors, Kinetic Parameters, etc.

Summary

References
Getting information from PMN pathway pages
Getting information from PMN pathway pages

- Download a complete gene list
Getting information from PMN pathway pages

- Download the pathway in BioPAX format
Getting information from PMN pathway pages

- Set organism viewing preferences on PlantCyc pages
Getting information from PMN pathway pages

- View predicted or experimentally supported enzymes
Searching in PMN databases

- Search page
  - Provides more selective searches
  - Allows browsing
Searching in PMN databases

Pathway Tools Query Page

- Select a database: Populus trichocarpa
- Query: All (by name of PlantCyc) or Populus trichocarpa
- Browse: Pathways, EC Hierarchy, Compounds, Gene Ontology, MultiFun Gene Taxonomy

Pathways:
- Biosynthesis
- Degradation, Utilization, Assimilation
  - Alcohols Degradation
  - Aldehydes Degradation
  - Amines and Polyamines Degradation
  - Amino Acids Degradation
  - Aromatic Compounds Degradation
  - C1 Compounds Utilization and Assimilation
  - Carbohydrates Degradation
  - Carboxylic Acids Degradation
  - Coenzymes, Prosthetic Groups, and Electron Carriers Degradation
  - Degradation, Utilization, Assimilation - Other
  - Fatty Acid and Lipids Degradation
  - Hormones Degradation
  - Inorganic Nutrients Metabolism
  - Nucleosides and Nucleotides Degradation and Recycling
  - Secondary Metabolites Degradation
- Detoxification
- Generation of precursor metabolites and energy
- Superpathways
Advanced searching in PMN databases

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

1. Enter your query here:
   - In database: PlantCyc
   - Search for: Compounds (2679 instances)
   - Add a condition

2. Select fields to include in the query output:
   - Column 1
     - Sort based on this column
   - NAME

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

4. Submit Query
   - Reset Query
Advanced searching in PMN databases

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

1. Enter your query here:

   | In database | PlantCyc | search for | Compounds (2679 instances) | add a condition |

   Where
   - add a condition

   Chemical-Formula
   - Contains the string
   - is a substring of
   - is not a substring of
   - is similar to (regular expression)
   - is not similar to (regular expression)
   - contains the substring (case-sensitive)
   - is a substring of (case-sensitive)
   - does not contain the substring (case-sensitive)
   - is not a substring of (case-sensitive)
   - is equal to
   - is not equal to
   - is equal to (case-sensitive)
   - is not equal to (case-sensitive)
Advanced searching in PMN databases

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

1. Enter your query here:

   - In database: PlantCyc
   - search for: Compounds (2679 instances)

   - Where: Chemical-Formula contains the substring: C30

   - “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

2. Select fields to include in the query output:

<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
<th>Column 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
<td>Chemical-Formula</td>
<td>Appears-In-Right-Side-Of</td>
<td>Molecular-Weight</td>
<td>NameC</td>
</tr>
</tbody>
</table>
Advanced searching in PMN databases

<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>
<th>Column 5 for (x1 ?NAMES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>o-amyrin</td>
<td>C30H5001</td>
<td>(S)-2,3-epoxysqualene = o-amyrin</td>
<td>426.724</td>
<td>o-amyrin, virmainol, alpha-Amyrenol</td>
</tr>
<tr>
<td>β-amyrin</td>
<td>C30H5001</td>
<td>(S)-2,3-epoxysqualene = β-amyrin</td>
<td>426.724</td>
<td>β-amyrin, β-amyrenol</td>
</tr>
<tr>
<td>(E)-cinnamoyl-CoA</td>
<td>C30H42N7O17P3S1</td>
<td>trans-cinnamate + coenzyme A = (E)-cinnamoyl-CoA + H2O</td>
<td>997.56</td>
<td>(E)-cinnamoyl-CoA</td>
</tr>
<tr>
<td>(S)-2,3-epoxysqualene</td>
<td>C30H5001</td>
<td>squalene + NADPH + O2 = (S)-2,3-epoxysqualene + NADP2 + H2O</td>
<td>426.724</td>
<td>(S)-2,3-epoxysqualene, squalene, 2,3-epoxide, squalene 2,3-oxide, (S)-squalene-2,3-epoxide, 2,3-EDSO, 2,3-epoxysqualene, 2,3-oxidosqualene</td>
</tr>
<tr>
<td>24-ethylidenolophenol</td>
<td>C30H5001</td>
<td>24-methylenelophenol + S-adenosyl-L-methionine = 24-ethylidenolophenol + S-adenosyl-L-homocysteine</td>
<td>426.724</td>
<td>24-ethylidenolophenol, (Z)-24-ethylidenolophenol, citrostadien</td>
</tr>
<tr>
<td>4,4-dimethyl-14a-formyl-5a-cholesta-8,24-dien-3β-ol</td>
<td>C30H48O2</td>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>440.708</td>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14a-formyl-5a-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
<tr>
<td>4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol</td>
<td>C30H5002</td>
<td>lanosterol + NADPH + O2 = 4,4-dimethyl-14α-hydroxymethyl-5α-cholesta-8,24-dien-3β-ol + NADP2 + 2 H2O</td>
<td>442.724</td>
<td>4,4-dimethyl-14α-formyl-5α-cholesta-8,24-dien-3β-ol</td>
</tr>
</tbody>
</table>
Advanced searching in PMN databases

- Other queries?
  - Identify all of the “glycosyltransferase” enzymes associated with more than two reactions in AraCyc
    - List their:
      - name
      - subcellular localization
      - molecular weight
      - inhibitors, activators, etc.
  
  - Find all of the biochemical pathways in PoplarCyc that have more than 5 reactions and where at least one of those reactions lacks enzymes
    - List their:
      - name
      - reactions
      - citations and evidence codes

- These searches can be used to span more than one PMN database

- What if I only have a gene or protein sequence?
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: Phytoreme Plant Orthologs AT4G008040.1

Gene-Reaction Schematic: ?

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

\[
\text{S-adenosyl-L-methionine} \leftrightarrow \text{S-methyl-5'-thiodenosine} + 1-\text{aminocyclopropane-1-carboxylate}
\]

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]
Comparing across species

- Use species selection tool on pathway pages
<table>
<thead>
<tr>
<th>Organism</th>
<th>Evidence Glyph</th>
<th>Enzymes and Genes for choline biosynthesis III</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcaCyc.col</td>
<td></td>
<td>EC# 2.7.7.15 choline-phosphate cytidylyltransferase AT4G15130 choline-phosphate cytidylyltransferase AT2G32280</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EC# 2.7.8.2 phosphatidylintransferase AT3G25595 phosphatidylintransferase AT1G13560</td>
</tr>
<tr>
<td>P. trichocarpa</td>
<td></td>
<td>EC# 2.7.7.15 None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EC# 2.7.8.2 diaclyglycerol cholinephosphotransferase JGI-225724 diaclyglycerol cholinephosphotransferase JGI-720065</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EC# 3.1.4.4 phospholipase D JGI-811801</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-781949</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-810176</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-593768</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-822706</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-578949</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-730956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-240457</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-633366</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-417304</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-558891</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-415367</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-556827</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-755219</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-180605</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phospholipase D JGI-829527</td>
</tr>
</tbody>
</table>
Comparing across species

- Use general Comparative Analyses tools
Comparing across species

<table>
<thead>
<tr>
<th>Pathway Class: Biosynthesis - Amines and Polyamines Biosynthesis</th>
<th>AraCyc col</th>
<th>P. trichocarpa</th>
</tr>
</thead>
<tbody>
<tr>
<td>glycine betaine biosynthesis III (plants)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>putrescine biosynthesis by agmatinase</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>putrescine biosynthesis II</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>putrescine biosynthesis IV</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>putrescine biosynthesis via N-carbamoylputrescine</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>spermidine biosynthesis</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>spermine biosynthesis</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>UDP-N-acetyl-D-glucosamine biosynthesis</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>urate biosynthesis</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pathway Classes: Metabolic Regulators Biosynthesis</th>
<th>AraCyc col</th>
<th>P. trichocarpa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Pathway Holes</td>
<td>435</td>
<td>646</td>
</tr>
<tr>
<td>Pathway Holes as a percentage of total reactions in pathways</td>
<td>29%</td>
<td>45%</td>
</tr>
<tr>
<td>Pathways with No Holes</td>
<td>180</td>
<td>93</td>
</tr>
<tr>
<td>Pathways with 1 Hole</td>
<td>80</td>
<td>76</td>
</tr>
<tr>
<td>Pathways with 2 Holes</td>
<td>46</td>
<td>43</td>
</tr>
<tr>
<td>Pathways with 3 Holes</td>
<td>13</td>
<td>22</td>
</tr>
<tr>
<td>Pathways with 4 Holes</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>Pathways with 5 Holes</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Pathways with &gt; 5 Holes</td>
<td>20</td>
<td>28</td>
</tr>
<tr>
<td>Total Pathways with Holes</td>
<td>178</td>
<td>192</td>
</tr>
</tbody>
</table>
Comparing across species

- Use Phytozome links on enzyme pages
Comparing across species

Overview of the AraCyc Metabolic Map

This diagram provides a schematic of all pathways of AraCyc metabolism in the aracyc database. Nodes represent metabolites, with shape indicating class of metabolite (see key to right). Lines represent reactions. Move the mouse over a metabolite icon to navigate to the metabolite page or a related pathway page.

Select one or more organisms:

- PlantCyc
- Populus trichocarpa

Submit  Clear All

Instructions
- Pathway Tools query page
- Omics Viewer: Paste omics data onto this diagram
- Species Comparison: Highlight reactions shared with other organisms
Comparing across species
Visualizing OMICs data

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

- Case study: Analyzing an Arabidopsis mutant with “no phenotype”
  - Basic phenotypic analyses do not reveal any differences:
    - growth
    - development
    - response to hormones
    - etc.
  - Perform a microarray analysis
    - Measure transcript levels in wild-type and mutant plants
  - Clean and process data
    - Remove genes expressed below background levels
    - Measure fold-increase or decrease in mutant vs. wild-type
    - Discard statistically insignificant data

- Do these transcript differences highlight possible metabolic perturbations?
Visualizing OMICs data

- Prepare an input file
  - Enter identifiers in first column
    - Genes
    - AGI locus codes (e.g. At2g46990)
  - Enter data in next column
    - fold - increase / fold - decrease
- Save as a tab-delimited text file

```
> All genes with significantly altered expression levels in cm1-1 mutants relative to wild type
> gene = change in expression in cm1-1 mutant relative to wild type

1  AT2G188550
   AT2G188550
2  AT2G188550
   AT2G188550
3  AT2G188550
   AT2G188550
4  AT2G188550
   AT2G188550
5  AT2G188550
   AT2G188550
6  AT2G188550
   AT2G188550
7  AT2G188550
   AT2G188550
8  AT2G188550
   AT2G188550
9  AT2G188550
   AT2G188550
10 AT2G188550
    AT2G188550
11 AT2G188550
    AT2G188550
12 AT2G188550
    AT2G188550
13 AT2G188550
    AT2G188550
14 AT2G188550
    AT2G188550
15 AT2G188550
    AT2G188550
16 AT2G188550
    AT2G188550
17 AT2G188550
    AT2G188550
18 AT2G188550
    AT2G188550
19 AT2G188550
    AT2G188550
20 AT2G188550
    AT2G188550
```
Visualizing OMICS data

- Upload data in the OMICS viewer
Visualizing OMICs data

- Upload data in the OMICs viewer

Pathway Tools Omics Viewers

Select a dataset:

File containing experimental data (NOT a URL):

Do you want to display absolute or relative data values?

If displaying relative data values, use:

Data values use a:

The items in the first (zeroth) column of your datafile are:

Data column (numerator in ratios):

If using two columns, denominator data column:

Note: For column numbering purposes, the first column, which contains the gene name, is column number 0. The first potential data column is column number 1.
Visualizing OMICS data

- Set display parameters in the OMICS viewer

Choose a color scheme:
- Full color spectrum, computed from data provided (default)
- Full color spectrum with a maximum cutoff: [ ]
- Three color display with specified threshold: [ ]

Display Type

By default, data values are painted on the cellular overview chart. However, an alternative display is to either paint data values on the genomic map, or to generate a table containing all individual pathways which have one or more data values that exceed some threshold (or are less than the inverse of that threshold). To select one of these alternative displays, choose the corresponding option below and specify the threshold if appropriate. Note that if both the cellular and genome overviews are specified, the genome overview will appear in a new browser window (you must have popups enabled for this site or this will not work).

- Paint data on cellular overview chart (default)
- Paint data on genome overview chart
- Generate a table of individual pathways exceeding threshold: [4]

Submit

Note that this request will take several minutes to complete (possibly longer for large datasets). For faster operation, install Pathway Tools on your own computer! Click here for details.
Visualizing OMICs data

Arabidopsis thaliana col Pathway: trans-lycopene biosynthesis
The mutant with “no phenotype” has

- decreased levels of transcripts related to phospholipid biosynthesis
- elevated levels of transcripts related to carotenoid biosynthesis

Future targeted experiments can be planned!
Visualizing OMICs data

- Many applications

- View changes in transcript, protein, or metabolite levels related to:
  - Mutant phenotype
  - Biotic stresses
  - Abiotic stresses
  - Natural variation
  - Developmental stage
  - Tissue type

- Display static measurements or changes over time
  - Animation feature is available
Data and software downloads

- Get pathway data sets from pathway pages
Data and software downloads

- Obtain large data sets

**PMN Data Downloads**

To download all of the plant metabolic pathway databases hosted by the PMN, including AraCyc, please fill out and submit the license agreement form.

The complete download, together with the Pathway Tools software, will allow a user to install and have a local copy of AraCyc/PlantCyc, just like the one visible from the PMN web site. In addition, all of the BioCyc format flat files and a complete biopax.xml file are available as part of this download. You can get more details about the format of all the file types from BioCyc.

Additional custom flat files containing data extracted from the latest PMN release can be freely downloaded without a license:

- PlantCyc pathways (plantcyc_dump)
- PlantCyc compounds (Note: This file ONLY contains compounds used in PlantCyc pathways)
- AraCyc pathways (aracyc_dump)
- AraCyc compounds (Note: This file ONLY contains compounds used in AraCyc pathways)
- PoplarCyc pathways (poplarcyc_dump)
- PoplarCyc compounds (Note: This file ONLY contains compounds used in PoplarCyc pathways)

The two BLAST sets provided by the PMN can also be downloaded without registration:

- PlantCyc Enzymes
- Reference Enzymes
Data and software downloads

Data formats include:
- ocelot, Biopax (OWL), SBML, .dat
Data and software downloads

- Install a local copy of the Pathway Tools software
Data and software downloads

- Desktop version offers additional features
Data and software downloads

- Desktop version offers additional features

Create new pathways
Modify existing pathways
Data and software downloads

- Trace metabolites through metabolic pathways
Data and software downloads

- Trace metabolites through metabolic pathways
Data and software downloads

- Coming soon . . . create and work with “groups” of objects

  - For some AraCyc pathways and other plant metabolic pathways, you can create and work with groups of objects NOW at:

Creating new PMN content

- Manual curation
  - Entering experimental data

- Computational prediction
PMN curators read the experimental literature . . . but we can't keep up with all of your exciting research!

We are particularly eager to get new pathways and enzymes related to specialized metabolism. We have pathways for several of the compounds mentioned this morning: phenylethanol, xanthohumol, (-)\textalpha-pinene, podophyllotoxin, etc.

But, for these pathways, we may be missing enzymes, intermediates, etc. And, we are missing many more pathways, compounds, and enzymes. You are the experts and best data sources!

Please meet with me during the conference . . .

Please send us your information.

Adding experimental data
PMN Contributors

PMN contributors from around the world have added to or helped to improve the content of AraCyc, PlantCyc, and the other PlantCyc-derived databases that are part of the PMN.

In addition to the active contributions from the PMN editorial board and PMN collaborators, the following individuals have contributed significantly in improving the content of PlantCyc, AraCyc and the other PlantCyc-derived databases that are part of the PMN:

- Some contributors have contacted us with suggestions and revisions.
- Some have generously responded to appeals for help from the curators.
- Some have attended curatorial seminars.

General Assistance:

- SJ Bloom - Keygene, The Netherlands
- Mark Poolman - Oxford Brookes University, UK
- Bryan Reynaud - Universidad de Chile, Chile

Contributors to Pathways of Primary and Secondary Metabolism

Primary Metabolism

- Analogs and Polyamines
  - John Jelkska - Virginia Tech, USA
- Amino Acids
  - Vipul Joshi - Cornell University, USA
- Carbohydrates
  - Christophe D'Hulst - Unité de Glycobiologie: Structure et Fonctionnelle, France

Cofactors, Prosthetic Groups, and Electron Carriers

- Imad Alawi - Michigan State University, USA
- J. Clark Lagarias - University of California, Davis, USA
- Rob Last - Michigan State University, USA
- Carlos Menck - Universidade de São Paulo, Brazil
- David Oliver - Iowa State University, USA
- Stéphane Rainsier - Institut de Recherches en Technologies et Sciences pour le Vivant, France
Creation of new PMN databases

- New sets of DNA sequences become available
  - Genomes are sequenced
  - Large EST data sets are created
    - Unigene builds are generated

- PMN pipeline predicts enzyme functions
  - Performed using computer algorithms based on sequence similarity

- Set of predicted enzymes is used to predict metabolic pathways
  - The pathway program (Pathologic) uses:
    - Enzyme functional annotations
    - A reference set of pathways (e.g. PlantCyc)

- Curators validate predicted pathways in the new database
  - Curators remove incorrect information and add additional data
Creation of new PMN databases

Annotated enzymes

DNA sequences

Gene identification

Enzyme activity

chorismate mutase

Pathway prediction program

Phaseolus vulgaris

PlantCyc

Reference pathways

Gene identification

Enzyme activity

chorismate mutase

DNA sequences

JGI, etc.

PMN pipeline

Pv1234.56.a

PlantCyc

New database

chorismate mutase

Pv1234.56.a

Pv1234.56.a

Prephenate aminotransferase

2.6.1.79

Pv1234.56.a

Prephenate dehydratase

4.2.1.91

Pv1234.56.a

L-phenylalanine

BeanCyc

+ validation

Phenylalanine biosynthesis
Creation of new PMN databases

- BeanCyc will be added to the PMN databases
- BeanCyc enzymes will be added to PlantCyc
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
We are here to help: www.plantcyc.org

- 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!
  - Stay around for the help session from 8-9 PM
  - Visit the PMN poster - #322
  - Make an appointment to meet with me during the conference
    - Sign-up on sheet in the back of the room
    - Send an e-mail

curator@plantcyc.org

www.plantcyc.org
We are here to help: www.plantcyc.org
PMN Acknowledgements

Peifen Zhang (Director)
Sue Rhee (PI)
Eva Huala (Co-PI)

Current Curators:
- A. S. Karthikeyan (curator)

Recent Past Contributors:
- Christophe Tissier (curator)
- Hartmut Foerster (curator)

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

Tech Team Members:
- Bob Muller (Manager)
- Larry Ploetz (Sys. Administrator)
- Raymond Chetty
- Anjo Chi
- Vanessa Kirkup
- Cynthia Lee
- Tom Meyer
- Shanker Singh
- Chris Wilks
We are here to help: www.plantcyc.org

- 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!
  - Stay around for the help session from 8-9 PM
  - Visit the PMN poster - #322
  - Make an appointment to meet with me during the conference
    - Sign-up on sheet in the back of the room
    - Send an e-mail

curator@plantcyc.org
www.plantcyc.org
What is in the PMN?

- Databases focus on “small” molecule metabolism
  - “Small”
    - No strict standards
    - Generally exclude macromolecules that are built using templates
  - “Large” molecules **Not In** the PMN:
    - chromosomes, proteins (as substrates), mRNA transcripts
      - but the building blocks of macromolecules are included
        - nucleotides and amino acids **are in the PMN**
  - “Large” molecules **In** the PMN:
    - cellulose
    - rubber
    - homogalacturonan / pectin