PMN, A Unified Resource For Plant Metabolism

Peifen Zhang
Carnegie Institution for Science
The PMN Resources
plantcyc.org

• A pipeline for high quality enzyme function prediction and metabolic pathway inference from sequenced plant genomes
The PMN Pipeline

1. Genome protein sequences
2. Enzymes
3. Reactions & pathways
4. Validated pathways

- E2P2
- Pathway Tools* (MetaCyc)
- SAVI
Pathway Inference by Pathway Tools

A curated known pathway in the reference MetaCyc

E2P2 enzyme predictions of the Spirodela genome

Gene - reaction ID
Gene - reaction ID
......
......

An inferred pathway in SpirodelaCyc

An inferred pathway in BarleyCyc
The PMN Resources
plantcyc.org

• A pipeline for high quality enzyme function prediction and metabolic pathway inference from sequenced plant genomes

• Metabolic pathway databases inferred for 100 plant genomes
The PMN Resources
plantcyc.org

• A pipeline for high quality enzyme function prediction and metabolic pathway inference from sequenced plant genomes

• Metabolic pathway databases inferred for 100 plant genomes

• A pan-plant metabolic pathway database PlantCyc
  • Experimental information of enzymes and pathways for over 400 plant species
PlantCyc

Morphine 
(*Papaver somniferum*)

Taxol
(*Taxus brevifolia*)

Vicianin
(*Vicia sativa*)

Alliin
(*Allium sativum*)
PMN use case-1: my gene

- Search for the gene
  - find reaction
    - find other genes annotated to the same reaction
  - find pathway
    - find upstream and downstream genes
Use case-2: given a list of genes, find enrichment of pathways
### SmartTable: New SmartTable - 2018-07-13T11:50:37-07:00

31 rows of all-genes from AraCyc col

**ADD TRANFORM COLUMN**
- choose a transform...

**ADD PROPERTY COLUMN**
- choose a property...

**ENRICHMENTS**
- choose an enrichment...

### Table Content:

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Choose an enrichment...**

- Genes enriched for pathways
- Genes enriched for transcriptional/translational regulators, and GO (all)
- Genes enriched for transcriptional/translational regulators (direct and indirect)
- Genes enriched for transcriptional/translational regulators (direct only)
- GO terms - genes enriched for GO (biological process)
- GO terms - genes enriched for GO (cellular component)
- GO terms - genes enriched for GO (molecular-function)

### Enrichment Results:

- **Phenylpropanoid Derivatives Biosynthesis**
  - p-value: 3.96993e-4
  - Genes: AT2G2, PAL2, PAL2, FLS, 4CL3

- **Trehalose Biosynthesis**
  - p-value: 4.6177022e-4
  - Genes: TPPE, TPPO, TPPA

- **Organic Solutes Biosynthesis**
  - p-value: 4.6177022e-4
  - Genes: TPPE, TPPO, TPPA
Overlay omics data on pathways
Virtual data integration in PlantCyc
PlantCyc Pathway: ethylene biosynthesis I (plants)

Selected Organisms (an organism name in bold means that the pathway is predicted to occur in that organism): Arabidopsis thaliana, Chlamydomonas reinhardtii, Physcomitrella patens

This view shows enzymes only for those organisms listed below, in either the list of selected organisms or the list of taxa known to possess the pathway. If an enzyme name is shown in bold, there is experimental evidence for this enzymatic activity.
Comparative analysis
<table>
<thead>
<tr>
<th>Organism</th>
<th>Evidence Glyph</th>
<th>Enzymes and Genes for hydroxylated mugineic acid phytosiderophore biosynthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AraCyc col</td>
<td></td>
<td>This pathway is not marked as present in this organism. No enzymes or genes have been identified for this pathway</td>
</tr>
</tbody>
</table>
| H. vulgare               |                | **EC 1.14.11.24** mugineic acid synthase: id3s3  
|                          |                | MLOC_33486.1: MLOC_33486.1  
|                          |                | MLOC_80402.1: MLOC_80402.1  
|                          |                | MLOC_17126.1: MLOC_17126.1  
|                          |                | MLOC_76208.1: MLOC_76208.1  
|                          |                | MLOC_48351.2: MLOC_48351.2  
|                          |                | MLOC_1079.1: MLOC_1079.1  
|                          |                | MLOC_19745.1: MLOC_19745.1  
|                          |                | **EC 1.14.11.25** mugineic-acid 3-dioxygenase: id3s2  
|                          |                | **EC 1.14.11.25** mugineic-acid 3-dioxygenase: id3s2  
| O. sativa Japonica Group |                | **EC 1.14.11.24** LOC_Os07g07410.1: LOC_Os07g07410.1  
|                          |                | **EC 1.14.11.25** None  
|                          |                | **EC 1.14.11.25** None  
| P. patens                |                | This pathway is not marked as present in this organism. No enzymes or genes have been identified for this pathway |
| S. bicolor               |                | **EC 1.14.11.24** Sobic_008G109700.1: Sobic_008G109700.1  
|                          |                | Sobic_008G109400.1: Sobic_008G109400.1  
|                          |                | Sobic_002G046400.1: Sobic_002G046400.1  
|                          |                | Sobic_008G109300.1: Sobic_008G109300.1  
|                          |                | Sobic_004G348400.1: Sobic_004G348400.1  
|                          |                | Sobic_008G109600.1: Sobic_008G109600.1  
|                          |                | **EC 1.14.11.25** None  
|                          |                | **EC 1.14.11.25** None  
| Z. mays mays             |                | **EC 1.14.11.24** GRMZM2G110369_P01: GRMZM2G110369_P01  
|                          |                | GRMZM2G146234_P01: GRMZM2G146234_P01  
|                          |                | GRMZM2G086231_P01: GRMZM2G086231_P01  
|                          |                | **EC 1.14.11.25** None  
|                          |                | **EC 1.14.11.25** None  

Future Development

• Further enhance accuracy of enzyme function predictions

• Annotate transporters and subcellular location of reactions

• Develop genome-scale metabolic models

• Provide online database creation service

• Provide web app for integrated view of metabolism with other information such as gene expression
Team and Funding Sources

- Sue Rhee (PI)
- Pascal Schläpfer (Postdoc)
- Jiun Yen (Postdoc)
- Bo Xue (Research Assistant)
- Angela Xu (Research Assistant)
- Kevin Radja (Research Assistant)
- Galyna Vakulenko (Intern)
- Garret Huntress (IT support)

- Peter Karp (Collaborator)