Coordination of Information Platforms to Support Use of Plant Genetic Resources

Prof. Pankaj Jaiswal
Dept. of Botany and Plant pathology
Oregon State University, USA
jaiswalp@oregonstate.edu
QUESTIONS:

List all individuals/accessions (e.g. of rice) that carry a given mutation and have a condition (Phenotype)?

What genes overlap my QTL?

Do these genes have any alleles for stress tolerance (e.g. drought)? Marker discovery.

What’s the function of candidate gene(s)?
Response to Xanthomonas
Response to Fusarium

PanGenome

Pathways
Expression

Inputs

Genome, Genotyping, Markers, Alleles
QTL & Phenotype
Ontologies

Outputs

Prior Knowledge
Scientific Journal
ISSN
WHERE DOES THE DATA COME FROM?

HOW DO WE TRACK THE DATA?

HOW IS THE DATA ASSIMILATED?

HOW IS THE DATA Displayed?

HOW IS THE DATA REPRESENTED?
Where does the information come from?

<table>
<thead>
<tr>
<th>Data Standards and Ontologies</th>
<th>Systems portal</th>
<th>Community databases</th>
<th>Topic centric databases</th>
<th>Data archives</th>
<th>Scientific community</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Database</td>
<td>Database</td>
<td>Database</td>
<td>Database</td>
<td>Individuals Labs Projects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Database</td>
<td>Database</td>
<td>Database</td>
<td>Publishers</td>
</tr>
</tbody>
</table>

- **Scientific community**: Includes individuals, labs, and projects.
- **Data archives**: Includes databases managed by publishers.
- **Topic centric databases**: Includes databases focused on specific topics.
- **Community databases**: Includes databases created by communities.
- **Systems portal**: Includes databases accessible through a portal.

**Individuals Labs Projects**

Data and Biocuration

- Gene
- Pathway
- Protein
- Map/Genome
- Ontology
- Protein
- Alternative forms and variants

Data Standards and Ontologies

- Standards
- Ontologies
WHERE DOES THE DATA COME FROM?

HOW DO WE TRACK THE DATA?

HOW IS THE DATA ASSIMILATED?

HOW IS THE DATA DISPLAYED?

HOW IS THE DATA REPRESENTED?
ABSENCE OF COMMON FRAMEWORK CONNECTING ALL DATA
120 reference genomes
Rice and Maize Pan-genomes
4M genes, 123K families
340 genome alignments
80 synteny maps
238M SNPs from 17 species
20K QTL, 1500 markers
Co-production with Ensembl Plants

Variant effect predictor
HMMER, an HMM-based query
Assembly converter to lift data from previous version

Sequence search

www.gramene.org
Gramene Plant Reactome

https://plantreactome.gramene.org

326 curated reference rice pathways
Pathway projection to 120 plant species
There is a challenge with agreement on what's important...

What should we record?

Should we have standardized data?
Let someone else handle it...

I'm going fishing!

We let other people do it.

Sorry--it's been done!

Wheee!

We didn't reinvent the wheel also, make data easily accessible.
Rescue effort by building Reference Ontologies for Plants

<table>
<thead>
<tr>
<th>GRANULARITY</th>
<th>Environment Ontology (ENVO)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plant Environmental Conditions (EO)</td>
</tr>
<tr>
<td></td>
<td>Plant Taxonomy</td>
</tr>
<tr>
<td></td>
<td>NCBI, uBio, USDA-GRIN, etc.</td>
</tr>
<tr>
<td>Anatomy</td>
<td>Plant Anatomical Entity</td>
</tr>
<tr>
<td></td>
<td>Plant Ontology (PO)</td>
</tr>
<tr>
<td>Cell</td>
<td>Plant Cell (PO)</td>
</tr>
<tr>
<td></td>
<td>Cell (CL)</td>
</tr>
<tr>
<td>Cellular Component</td>
<td>Cellular Component (GO)</td>
</tr>
<tr>
<td>Molecule</td>
<td>Molecular Entity</td>
</tr>
<tr>
<td></td>
<td>(ChEBI, PR)</td>
</tr>
<tr>
<td></td>
<td>Phenotypic Quality (PATO)</td>
</tr>
<tr>
<td></td>
<td>Biological Process (GO)</td>
</tr>
</tbody>
</table>

Key:
- Purple: New Ontologies to be developed
- Cyan: work with the existing collaborator ontologies for desired plant enrichment
- Yellow: ongoing in-house efforts
Can we use them for inter and intra-specific comparison?
Species/Genus-specific vocabulary
Common Reference Ontology

Trait Ontology

Image source via Google
Common Reference Ontology

Image source via Google
WHERE DOES THE DATA COME FROM?

HOW DO WE TRACK THE DATA?

HOW IS THE DATA ASSIMILATED?

HOW IS THE DATA DISPLAYED?

HOW IS THE DATA REPRESENTED?
Thank You

Questions?