Tripal in the Legume Genomics Community

January 11th, 2015
Tripal Workshop
Ethyl Cannon
Iowa State University

1. A case study of Tripal/Chado
2. A description of two Tripal modules our groups are developing

A case study with PeanutBase and LegumeInfo

- **PeanutBase** is a new resource funded by the Peanut Foundation. Most personnel are at Iowa State University.
- **LegumeInfo** is the new implementation of the Legume Information System and is funded by the USDA-ARS. Most personnel are at the National Center for Genomic Resources.
- Both teams share some members.

A case study with PeanutBase and LegumeInfo

How can we share development and curation efforts across both websites and both locations?

Ruby on Rails?
Development objectives

• Enable sharing of tool development, curation, and data between our two similar data portals.
• Avoid redeveloping existing tools.
• Address the challenges of genomic/breeding data portals for small communities.
• Support efforts toward standard data collection, metadata standards, schema, structures with sharable loaders and viewers.

An overview of our Tripal/Chado experience

• Created (mostly empty) websites very quickly.
• Difficult for even experienced developers to learn how to customize Drupal/Tripal, more difficult to write new modules.
An overview of our Tripal/Chado experience

- Created (mostly empty) websites very quickly.
- Difficult for even experienced developers to learn how to customize Drupal/Tripal, more difficult to write new modules.
- Chado's flexibility makes it difficult to work with.
  - There are multiple ways to load the same data.
  - It is difficult to write custom loaders that are compatible with Tripal.
  - Controlled vocabularies for describing the data structures are essential but difficult to develop.

Wish list

- Chado: standards for loading common types of data (gene models, QTL, et cetera).
- Tripal/Chado: improved loaders with error checking to help debug data errors.
- Tripal: improved error reporting for both content management and module development.

Lessons learned

- Deciding to use Chado does not mean your data will look like other data in Chado.
- It is worthwhile to take the time to do things right; it makes your data and tools more sharable and puts you in a better position to use other people's data and tools.
- Don’t waste resources solving problems that have already been solved even if you don’t completely agree with the solution.
Lessons learned

- Deciding to use Chado does not mean your data will look like other data in Chado.
- It is worthwhile to take the time to do things right; it makes your data and tools more sharable and puts you in a better position to use other people’s data and tools.
- Don’t waste resources solving problems that have already been solved even if you don’t completely agree with the solution.
- Most important: Tripal/Chado permits productive cross-site and cross-database development, effectively increasing the size of both the LegumeInfo and PeanutBase teams.

1. A case study of Tripal/Chado

2. A description of two Tripal modules our groups are developing.

Tripal extension modules

- **PhyloTree** – in development at LegumeInfo (Iliana Toneva, Alex Rice)
- **QTL** – in development at PeanutBase, based on QTL module at CoolSeasonLegume.org (Ethy Cannon, Stephen Ficklin, QC by Scott Kalberer)

**PhyloTree**

For viewing phylogenetic trees of gene families.

**PhyloTree**

For viewing phylogenetic trees of gene families. Gene families are helpful for:
- doing cross-species comparative analysis,
PhyloTree
For viewing phylogenetic trees of gene families.
Gene families are helpful for:
• doing cross-species comparative analysis,
• make it possible for a poorly-characterized species like peanut to take advantage of resources for a well-characterized species like soybean.
PhyloTree

Status

- Hosted at LegumeInfo.
- Gene and gene family searches at both LegumeInfo and PeanutBase + homology through gene families link the two websites together.
- Will be made available to all Tripal installations; the process of meeting Tripal standards has started.

Tripal extension modules

- PhyloTree – in development at LegumeInfo (Iliana Toneva, Alex Rice)
- QTL – in development at PeanutBase, based on QTL module at CoolSeasonLegume.org (Ethy Cannon, Stephen Ficklin, QC by Scott Kalberer)

Collecting, loading and displaying QTL data
Collecting, loading and displaying QTL data

QTL data and metadata is very complex. The Chado schema is general-purpose and highly flexible. No standards and few recommended practices for mapping QTL data onto Chado.

• **The challenge**: the complexity of QTL data and metadata, and the lack of strong standards means the data is collected and displayed differently by each web resource.

• There is a recommendation, Minimum Information about a QTL or Association Study (MIQAS), animal-centric.

• **Required**: create a standard data collection template for plants, based on the MIQAS recommendation and what others are doing now.

• “There’s more than one way to do it.” —Perl of Wisdom.
Collecting, loading and displaying QTL data

- "There's more than one way to do it." – Perl of Wisdom.
- Different QTL information is provided and collected by different communities.
- QTL data has changed over time.
- We tried to find a consensus or "canonical" method, decided to mimic Genomic Database for Rosaceae's data structure, but still managed to create something different.

Tripal Extension module: QTL module (prototype)
QTL module

Status:

- We have a prototype which is active at both PeanutBase and LegumeInfo.
- Working with SoyBase as well as Tripal folks to define a standard data collection template.
- First kickoff meeting to plan the publicly-available Tripal QTL module at PAG.
  — Sook Jung, Stephen Ficklin, Lacey Sanderson, Ethy Cannon
- Input welcome from anyone.
## Who We Are

### PeanutBase
- Steven Cannon
- Sudhansu Dash
- Scott Kalberer

### LegumeInfo
- Andrew Farmer
- Alan Cleary
- Alex Rice
- Jugpreet Singh
- Iliana Toneva
- Pooja Umale
- Nathan Weeks

### Genomic Database for Rosaceae
- Dorrie Main
- Sook Jung

### CoolSeasonFoodLegume
- Dorrie Main
- Stephen Ficklin

**Funding:**
- Peanut Foundation
- USDA-ARS