GDR/CottonGen: Converting legacy sites to Tripal

Sook Jung, Jing Yu, Taein Lee, Chun-Huai Cheng, Stephen Ficklin, Dorrie Main

Talk Overview

- Introduction of GDR
- Moving to Chado/Tripal
  - GDR
  - CottonGen
- How to store data in Chado
- Tripal modules developed
- Future Directions

GDR (2002 ~)

- Genomics
  - EST Unigenes
  - WGS and annotation
- Genetics
  - Markers and maps
  - QTL
  - Molecular diversity
- Breeding
  - Genotypic data
  - Phenotypic data
  - Germplasm data

GDR in Tripal

- New Custom/Extension Tripal Modules for gene, marker, QTL, genotype, publication (2013)
- Breeders Toolbox developed (Not in Tripal) (2010-2011)
- Generic data moved to Chado (2010)
- Tripal instance created & Breeding data in Chado (2010)
- GDR in Drupal (2009)
- ND schema developed & Breeding data in Chado (2010)
- GDR in in-house schema (2002 ~)

CottonDB to CottonGen

- CottonDB to CottonGen Released (1st year)
- CottonDB on WSU servers (Oct 2011)
- CottonDB data in Chado
- Develop ICGC website
- Triad instance created
- CottonGen Released
- Gene, marker, search pages (Oct 2012)
- Phenotype, QTL, Publication search pages added (2013)

Chado: Modular, Generic and Ontology-driven schema

- Feature
  - Feature_id
  - Feature_name
  - Type_id
  - Organization_id
- Feature_relationship
  - Feature_relationship_id
  - Subject_id
  - Object_id
  - Type_id
- gene, mRNA, marker, QTL, etc

- Repeat_motif
  - Product_type
  - Value

- Comment
  - Comment_id
  - Name
  - definition
  - rv_id
  - Object_id
- Repeat_motif
  - Product_type
  - Next
Storing Stock (from samples to population; pedigree)

- Population, cultivar, breeding line, clone, sample, etc.
- stock
  - stock_id
  - Name
  - Uniquename
  - Type_id
  - Organism_id
  - residues
  - stock_relationship
    - Feature_relationship_id
    - Subject_id
    - Object_id
    - Type_id
- Gala
  - offspring
  - sample_of
  - Gala
- pedigree

Storing phenotype data (from measurements to projects)

- Population, cultivar, breeding line, clone, sample, etc.
- Gala
  - 001
- Description, population size
- Gala
  - Maternal_parent_of
  - Sonya
- pedigree

Genotypic data integrated with genomic/genetic data

- project
  - Nd_experiment
    - Nd_experiment_id
    - Nd_geolocation_id
    - Type_id
- genotype
  - genotype_id
  - name
  - Uniquename
  - description
  - NE_genotype
  - feature_genotype
    - feature_id
    - Name
    - Uniquename
    - Type_id
    - Organism_id
    - residues
  - Uniquename: CPSCT038
    - Type: microsatellite
  - feature_genotype

Tripal Modules Developed (Custom modules)

- Gene/Sequence Module
- Genetic Module
  - Marker Search
  - QTL Search
  - Genotype Search
Future Directions

• Make the current modules available
  o With a set of controlled vocabularies
  o Bulk loader templates

• Further refinement of the modules
  o QTL: add graphic interface to view the QTLs in the genome
  o Further develop diversity module (integrate with phenotypic diversity and germplasm module)
  o Germplasm (search page, integrate with image module, etc.)
  o Data transformation functionality

• Introduce flexibility to the modules
  o Allow adding users’ own CV
  o Options to display certain data according to the CV

Acknowledgements

• Main Lab team members
  Dorrie Main, Taein Lee, Stephen Ficklin, Jing Yu, Chunhua Cheng, Ping Zheng, Anna Blenda, Sushan Ru, Dorrie Main

• GDR Project coPIs: Dorrie Main (PI), Bert Abbott, Cameron Peace, Kate Evins, Des Layne, Nnadozie Iraguaze, Mercy Olmstead, Fred Gmitter Jr., The RosBREED Teams

• CottonGen – Don Jones, Richard Percy

• Rosaceae, Cotton and Bioinformatics Community

• USDA NIFA SCRI, NSF Plant Genome Program, MARS, USDA-ARS, Washington Tree Fruit Research Commission, Cotton Inc., WSU, Clemson University, University of Florida, Boyce Thompson Institute, Texas A&M