Integrating Phenotypic Data With Genomic, Genetic and Genotypic Data Using Chado

Sook Jung, Taein Lee, Stephen Ficklin, Jing Yu, Dorrie Main

Outline
- Introduction of GDR and CottonGen
- Chado the generic schema
- Storing Stock Data
- Storing Phenotypic Data (trait, dataset, etc)
- Storing Genotypic Data
- Integration with genetic and genomic Data
- Conclusion

Database projects of Main lab
- Major databases with genomic, genetic, phenotypic and genotypic data
  1. GDR: Genome Database for Rosaceae
     Genomic, Genotect and Breeding data (Private data and data from RosBreed project)
     - Fruit and Nut, Sat, 12 PM
     - Computer Demo, Mon, 1:35 PM
     - 104HS, RosBreed GM System, Mon, 10-11:30 AM
  2. CottonGen: Replaced CottonDB and Cotton Marker Database
     - Cotton Genome Initiative, Sun, 3:50 PM
     - Computer Demo, Mon, 1:50 PM
- Other databases:
  - Citrus Genome Database, Cool season food legume database, Genome database for Vaccinium
- Built using Chado schema and Tripal (Drupal front end for Chado)
- Tripal: presentation, GMOD workshop, Wed 11:50 AM

Chado: Modular, Generic and Ontology-driven schema

Database tool
The Chado Natural Diversity module: a new generic database schema for large-scale phenotyping and genotyping data

Publication
Mapping the Arabidopsis thaliana genome to the Arabidopsis Genome Database using Chado

Feature
- feature_id
- name
- unique_name
- type_id
- organism_id
- residues
- feature_relationship_id
- subject_id
- object_id
- type_id
- featureprop_id
- featureprop_id
- type_id
- value
- rank
- cvterm_id
- name
- definition
- cv_id
- dbxref_id

Chado: Modular, Generic and Ontology-driven schema

Sequence Ontology, Gene Ontology, etc.
Storing Stock (from samples to population; pedigree)

Storing phenotype data (from measurements to projects)

Storing phenotype data (enabling comparison among datasets)

Genotypic data integrated with genomic/genetic data

Relationship between genotype and phenotype (haplotype and haplotype effect)

Conclusion

- Flexibility and generic characteristic of Chado enables us to store and integrate complex biological data from widely different projects and species.
- The ontology-driven characteristic makes adding new data types relatively easy.
- Performance issue mostly resolved by the use of materialized views.
Acknowledgement

- Natural diversity module working group
  Naama Menda, Seth Redmond, Robert M. Buels, Maren Friesen, Yuri Bendana, Lacey-Anne Sanderson, Hilmar Lapp, Tavin Lee, Bob MacCallum, Kirstin E. Bett, Scott Cain, Dave Clements, Lukas A. Mueller and Dorrie Main
- Main Lab team
- All Project CoPIs (tfGDR, RosBreed and CottonGen)
- Funding Sources
  USDA NIFA SCRI, NSF Plant Genome Program, USDA-ARS, Washington Tree Fruit Research Commission, Cotton Incorporated, Washington State University, Clemson University, University of Florida, Boyce Thompson Institute, North Carolina State University