FINAL PROJECT REPORT

Project Title: Developing an Online Toolbox for Tree Fruit Breeding

PI: Dorrie Main
Organization: WSU
Telephone: 509-335-2774
Email: dorrie@wsu.edu
Address: 45 Johnson Hall
City: Pullman
State/Zip: WA, 99164-6414

Co-PI(2): Sook Jung
Organization: WSU
Telephone: 509-335-2774
Email: sook@bioinfo.wsu.edu
Address: 50 Johnson Hall
City: Pullman
State/Zip: WA, 99164-6414

Co-PI(3): Kate Evans
Organization: WSU
Telephone: 509-663-8181
Email: kate_evans@wsu.edu
Address: 1100 N. Western Ave.
City: Wenatchee
State/Zip: WA, 98801

Co-PI(4): Cameron Peace
Organization: WSU
Telephone: 509-335-6899
Email: cpeace@wsu.edu
Address: 39 Johnson Hall
City: Pullman
State/Zip: WA 99614-6414

Co-PI(5): Nnadozie Oraguzie
Organization: WSU
Telephone: 509-785-9271
Email: noraguzie@wsu.edu
Address: 24106 N. Bunn Rd
City: Prosser
State/Zip: WA, 99350

Cooperators: Amy Iezzoni (MSU), Gennaro Fazio (USDA-ARS), Gayle Volk (USDA-ARS)

Other funding Sources

Agency Name: USDA-CSREES Specialty Crops Research Initiative
Amount awarded: $2,000,000 plus equal amount matching from universities and industry (Sep 2009 – Aug 2013)
Notes: “Tree Fruit GDR: Translating genomics into advances in horticulture”. PI: Dorrie Main. Co-PIs include Jung, Evans, Peace and Oraguzie. Synergistic project for practical application of bioinformatics to tree fruit crops.

Agency Name: USDA-CSREES, Specialty Crops Research Initiative
Amount awarded: $7,200,000 plus equal amount matching from universities and industry (Sep 2009 – Aug 2013)
Notes: “RosBREED: Enabling marker-assisted breeding in Rosaceae”. PI: Amy Iezzoni. Co-PIs include Peace, Main, Evans and Oraguzie. A synergistic project to establish sustainable marker-assisted breeding infrastructure for U.S. Rosaceae crops, using the Marker-Assisted Breeding Pipeline concept that involves Pedigree-Based Analysis.

Agency Name: WTFRC Apple Review
Amount requested: $635,201 (2009-2011)
Notes: “Apple Scion Breeding” PI: Kate Evans. Co-PIs: Peace, Ross, Zhu. The foundation program which this project supplements.

Agency Name: WTFRC Apple Review
Amount requested: $121,500 (2009-2010)
Notes: “Genetic marker assistance for the Washington apple breeding program” PI: Cameron Peace. Co-PIs: Evans, Olmstead , Mattheis. Fundamental development of seedling database for incorporation in the breeders toolbox.

Agency Name: MARS Inc. through USDA-ARS
Amount awarded: $550,000 (Jan 2008 – Jan 2012)
Notes: “The Cacao Genome Database. PI: Dorrie Main. Synergistic project for practical application of bioinformatics to tree fruit crops.

Total project funding:

Item	**2009**	**2010**
Salaries	26000	26000
Benefits	8808	8808
Wages	3000	3000
Benefits	0	0
Equipment	0	0
Supplies	0	0
Travel	1000	1000
Miscellaneous	0	0
Total	$38,808	$38,808

RECAP ORIGINAL GOALS

1. Integrate publicly available gene, trait, QTL and breeding data for apple and cherry with the current genomics and genetics data and tools in the Genome Database for Rosaceae.
2. Develop web interfaces for molecular geneticists and breeders to upload new data.
3. Develop web interfaces and online tools suitable for accessing and mining all breeder relevant data.

Year One Goals and Activities:

- Design data schema for new features
- Collect gene, trait, QTL and breeding data for apple and cherry and upload to database.
- Design queries for data retrieval
- Start web interface/tool design process

Year Two Goals and Activities:

- Continue to collect to gene, trait, QTL and breeding data for apple and cherry and upload to database.
- Complete web/interface tool design process

Anticipated Accomplishments:

After this two year project is complete the apple and cherry breeding program will have a robust, and secure data management and data querying database that will link directly to the integrated genomics and genetics data housed in the Genome Database for Rosaceae (GDR). GDR will have a publicly available breeding toolbox that is integrated with existing genomics and genetics data. This breeders database and toolbox will be part of a larger open-source genomics, genetics and breeding database for Rosaceae, that can also be utilized by other plant communities.

SIGNIFICANT FINDINGS

1. Developed a new diversity module for Chado, the open-source genomics database schema, in collaboration with bioinformaticists working on other databases including SGN (Solanaceae Genomics Network), VectorBase, and KnowPulse. Paper being prepared for submission to peer-reviewed Databases journal.

2. Excel templates for breeders to record and upload their data have been created.

3. Existing data from Washington Apple Breeding Program (WABP) have been converted into template, and the data from Pacific North West Sweet Cherry Breeding program (PNWSCBP) in the process of being converted into template.

4. Scripts created to upload data in template to database.

5. Data from WABP has been uploaded to database.

6. Held multiple online conferences between bioinformatics team and breeders to design the web interfaces.

7. Interfaces for browse, search and data download under development.

8. Private web management system created for breeding programs using the Drupal Groups function.

9. Following a presentation at the 5th International Rosaceae Genomics Conference by Dr. Evans on the "Development of an Online Tree fruit Breeders Toolbox", several breeders from International programs contacted Drs. Main/Evans about also wanting to use the toolbox system, indicating utility and expandability to other breeding programs

METHODS

1. A new module of Chado for breeding data has been developed in collaboration with bioinformaticists working on databases for multiple projects with multiple species to ensure the development of a module that can handle data from various projects.

2. The templates and the uploading scripts for breeding data have been designed and developed to have two different levels so that breeders in different programs can easily upload their data and yet the same main uploading scripts can be used in uploading data from multiple breeders. The main uploading scripts were developed in modules (ontology, contact, location, cross, germplasm, project, propagation, and phenotype) so that the common modules can be reused for uploading different types of breeding data. For example, all the modules above, except the phenotype module, can be used when uploading breeding data with genotype (Figure 1).

3. We are employing various tools and technologies that comes with Drupal, a pioneering Content Management System, to build a web site with various functionalities with minimum work for initial building and maintenance. For example, we used a feature called 'Organic Group' in Drupal to develop a private web management system for each breeder. We have also started a collaboration with a group in University of Saskatchewan to use a tool called 'View' in Drupal to develop versatile query and data detail pages that require minimum coding for maintenance and future modifications.

RESULTS AND DISCUSSION

We are on track to meet almost all our year 2 goals (end of March 2011), with the exception perhaps of not having all the trait data we'd like available by that date. One of the major findings to date is that we have successfully modeled a database system that will accommodate not just the data and analysis needs for the apple and cherry breeding programs in Washington State but also other crops and any other species for which phenotyping data are available. We have also developed data templates and data uploading system that are flexible for various breeding program. Another major work is that we identified the format of browse/query/download web interfaces that will be most useful for breeders (Figure 2). The development of these web interfaces are well underway and we anticipate it to be completed by the end of our year 2. In addition, we have developed a easier-than-wiki-type system where individual breeders, without any experience in creating a web site, can easily create their own web site, create users with different roles and directly link to their breeding database.

The provision of breeder focused web-databases with various browse, queries and download functionality will greatly accelerate breeders' work in evaluating breeding selections, comparing various lines and identifying elite lines for further testing. Our system allows breeders to directly upload the outputs of analytical tests whenever possible, reducing the time and labor managing the data. The breeding data reside within the same schema where the genomic and genetic data of GDR are stored, enabling the future connection between the genomics and genetics data with the actual improvement of cultivars. The two breeding programs in Washington are participants of RosBREED Project and they will utilize DNA markers to test the genotype of their breeding selections. When the genotyping data become available and uploaded to the breeding database, the data will be easily integrated with the genetics and genomics data of GDR.

The breeding data integrated within GDR will significantly accelerate identification and application of the genes and markers underlying important economic traits such as pre and post harvest fruit quality, and pest and disease resistance. Improvement of metric traits through the application of bio-informational methods will give a more predictable outcome to plant breeding than is currently the case with conventional one-gene-at-a-time genetics or phenotypic selection approaches. This database will allow the collection, storage and analysis of appropriate DNA, RNA, phenotype and germplasm datasets which can then be linked to traits that are of interest to breeders and industry stakeholders. This database resource will aid marker-assisted tree fruit breeding and facilitate the creation of new cultivars which meet consumers' needs and sustainable agricultural practices in the Pacific North West.

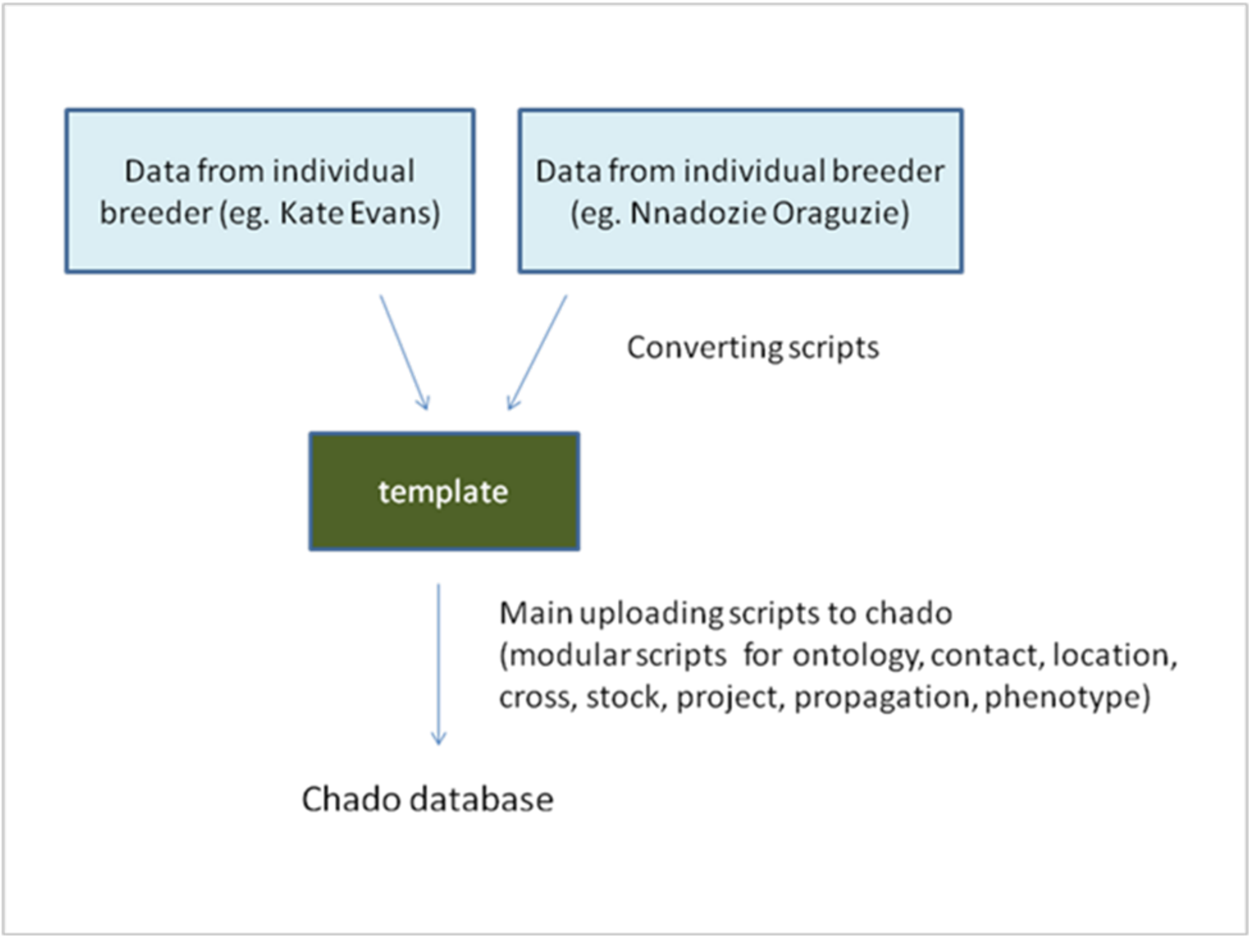
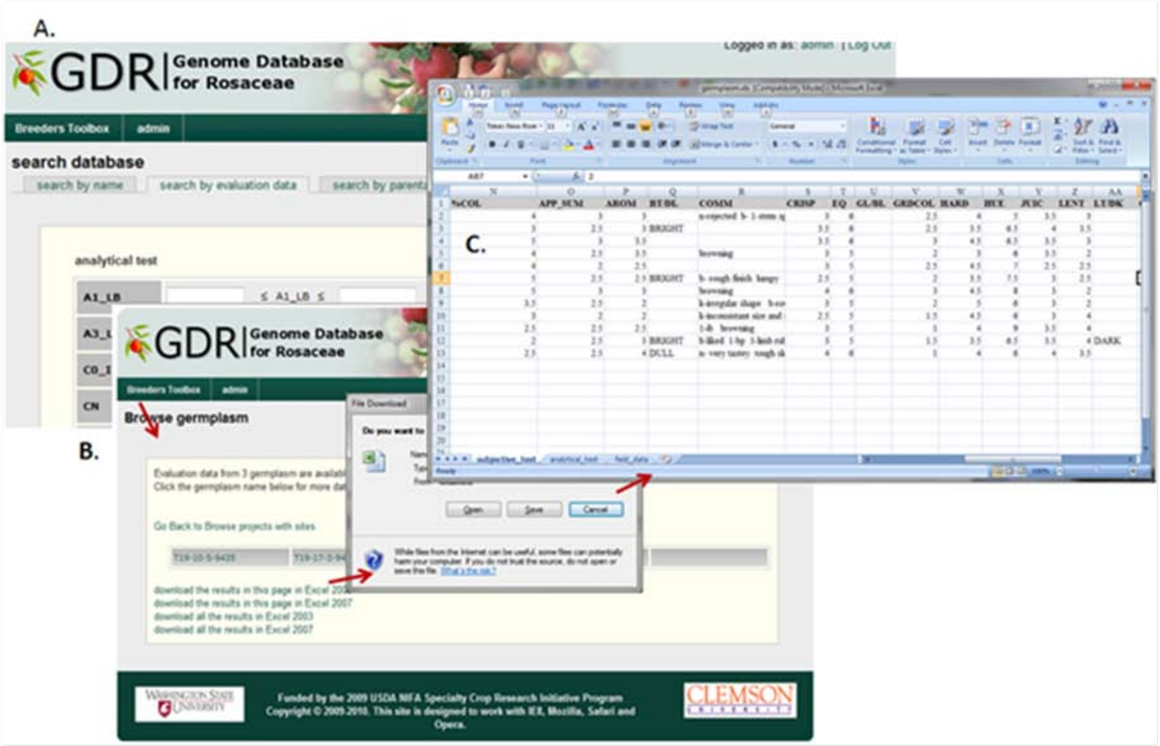


Figure 1. The scheme of data uploading pipeline for breeding data



Evaluation data from 3 germplasm are availab
Click the germplasm name below for more da

While files from the Internet can be useful, some files can potentially harm your computer. If you do not trust the source, do not open or save this file. What's the risk?

Funded by the 2009 USDA NIFA Specialty Crop Research Initiative Program
Copyright © 2009-2010. This site is designed to work with IE8, Mozilla, Safari and Opera.

Figure 2. Screen Shots of the web interfaces of breeders' toolbox. A. Users can browse by projects (not shown) or search by name, evaluation data, or parentage. B. Users get a result page with list of germplasm names corresponding to various queries or browse selections made along with the options to download evaluation data in Excel. C. An example of downloaded Excel document with the evaluation data. In search pages, users can select the types of evaluation data that they want to include in the downloaded data.

EXECUTIVE SUMMARY

This project was initiated to provide the Washington Apple Breeding Program (WABP) and the Pacific North West Sweet Cherry Breeding program (PNWSCBP) with a database resource to house and manage their voluminous data within an easy-to-use, secure, online system that seamlessly integrates with all the available genomics and genetics data available in the Genome Database for Rosaceae (GDR). This required that gene and trait data continue to be updated in GDR while also building the private data management system and online toolbox for the individual breeding programs. This involved in-depth discussions so the development team could fully understand each program's breeding practices to ensure they were able to capture every relevant data point within the generic database structure (Chado) they decided to use. This was followed, in collaboration, which we initiated, with several other bioinformatics groups to remodel Chado and add a natural diversity module (currently being written up for publication) facilitating a more comprehensive genomics, genetics and breeding data structure.

Having modeled the database structure, an excel template was devised to house all the breeding data in sheets according to the breeders needs and scripts were written to (1) convert the data from the WABP Access database into this template and (2) upload the data from this template into the individual program databases. Interfaces are being developed specifically for the WABP to provide the required functionality, still an ongoing process. This was presented by Dr. Evans at the 5th International Rosacaeae Conference in South Africa, and was very well received by the community. Several breeders from outside the US have since expressed interest in also adopting this system if it is made available to them through GDR in the future.

By the end of year 2 we anticipate having the PNWSCBP data uploaded and available for browsing/querying, similar to the WABP functionality. Further developments will include providing an automated data upload and editing function for new data, the addition of genotyping data provided through the RosBREED and WTFRC projects, adding trait data and linking to the apple and peach gene and genome sequences available in GDR, and providing advanced querying and reporting functionality as required for the breeding programs.

We are confident that the interdisciplinary, team-based approach to creating this database and toolbox will prove to be a fundamental resource to our WSU managed tree fruit breeding programs and greatly facilitate the development of new and improved cultivars that will enhance the competitiveness and sustainability of our local tree fruit growers and producers.