FINAL PROJECT REPORT

Project Title: Further development of an online toolbox for tree fruit breeding

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

Co-PI (2): Sook Jung
Organization: WSU Pullman
Telephone: 509-335-2774
Email: sook_jung@wsu.edu
Address: 45 Johnson Hall
City/State/Zip: Pullman/WA/99164

Co-PI(3): Kate Evans
Organization: WSU TFREC
Telephone: 509-663-8181 x245
Email: kate_evans@wsu.edu
Address: 1100 N. Western Ave
City/State/Zip: Wenatchee/WA/98801

Co-PI (4): Nnadozie Oraguzie
Organization: WSU IAREC
Telephone: 509-786-9271
Email: noraguzie@wsu.edu
Address: 24106 N. Bunn Road
City/State/Zip: Prosser/WA/99350

PI(5): Cameron Peace
Organization: WSU Pullman
Telephone: 509-335-6899
Email: cpeace@wsu.edu
Address: 39 Johnson Hall
City/State/Zip: Pullman/WA/99164

Administrative Contact: Mary Lou Bricker (509-335-7667, mdeseros@wsu.edu)

Cooperators: Amy Iezzoni (MSU), Gennaro Fazio (USDA-ARS), Deven See (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:

Budget History:

Item	2011	Year 2:	Year 3:
Salaries[1]	31,180		
Benefits	11,069		
Wages	0		
Benefits	0		
Equipment	0		
Supplies[2]	1500		
Travel[3]	1500		
Plot Fees	0		
Miscellaneous	0		
Total	**45,249**		

*[1]20% salary for Sook Jung and 40% salary for Chun-Huai Cheng
[2]$1500 toward data backup costs and IT facilities rental costs
[3]$1500 for travel to Prosser, Wenatchee and Pullman for face-to-face group meetings

RECAP ORIGINAL OBJECTIVES

1. Develop a computational pipeline to process genotyping data from the WA apple and cherry breeding programs
2. Integrate marker/genotyping data with the breeders toolbox
3. Develop standard and customizable breeder reporting tools for each breeder within their toolbox.

SIGNIFICANT FINDINGS

Objective 1: Develop a computational pipeline to process genotyping data from the WA apple and cherry breeding programs

A computational pipeline has been developed to process and extract the genotyping data from the excel template. This now completes all the computational scripts needed for a programmer to upload all genetic and breeding data for the WA apple and cherry breeding programs. All available phenotypic and genotypic data for the apple breeding program have been uploaded and integrated in the database. The cherry breeding team are in the process of adding their data to the excel templates. The tools are in place to quickly create the online toolbox for the cherry team as soon as the data is made available.

Objective 2: Integrate marker/genotyping data with the breeders toolbox

All private marker and genotyping data have been uploaded to the database for the WA apple breeding program and collection of the same data for the cherry breeding program is in hand and will be complete by the end of March, ensuring the proposal goals are met. By the end of March, the private and public RosBREED marker data for the Apple and Cherry Crop Reference and Breeding Pedigree Sets will be also be available for our breeding programs**.** Apple and cherry QTL data have been collected from the literature and standardized nomenclature assigned to each trait. A total of 485 QTLs for 93 traits are now available for searching on GDR for apple and while yet integrated with markers and the Breeders Toolboxes, this will be included in the GDR makeover we are currently working on as part of the tfGDR SCRI project.

Objective 3: Develop standard and customizable breeder reporting tools for each breeder within their toolbox

Interfaces have been developed with links from the home page (see Figure 1) that allow the Apple breeding team to search and compare both their private and public (RosBREED) evaluation data by variety, trait, marker, allele, parentage, year, and site location. The results are customizable and downloadable as excel files or as a file for input into the Pedimap analysis program. The same customizable features will be made available to the Cherry breeding team after the data has been provided and uploaded to the database. We anticipate this will be available by the end of March and fulfill all the objectives of this proposal.

Publications and Presentations

Following development of a Natural Diversify module to expand the generic database schema, Chado, to accommodate large scale phenotyping and genotyping data, a paper entitled "The Chado Natural Diversity module: a new generic database schema for large-scale phenotyping and genotyping data" was published in the peer-reviewed Databases journal (Jung *et al*, 2011). This expansion to Chado has generated considerable interest within the database and bioinformatics community and is anticipated to significantly expand the use of Chado in the development of integrated genomics, genetics and breeding databases.

The apple toolbox was highlighted at several presentations during the course of the year. These included:

(1) The 2011 Plant and Animal Genome Conference ("GDR" computer demonstration by Dr. Jung, and the "Cacao Genome Database" presentation at the Cacao Workshop by Dr. Zheng),

(2) The 2011 National Citrus Genomics Conference ("Citrus Genome Database" presentation by Dr. Jung)

(3) The 2011 University Industry Consortium Meeting ("Knowledge Warehouses for Fruit Farms of the Future" presentation by Dr. Main),

(4) The 2011 American Society of Horticultural Science Annual Conference ("RosBREED: Enabling marker-assisted breeding in Rosaceae" presentation by Dr. Peace and a poster " RosBREED enables marker-assisted breeding for apple" by Dr Peace *et al*.)

(5) The 2011 American Society of Agronomy, Crop Science Society of America, and the Soil Science Society of America International Annual Meeting ("Building Database Resources For Translational Research in Crop Science" presentation by Dr. Jung)

(6) The 2011 Eucarpia Fruit Breeding and Genetics Meeting ("Apple breeding in the Pacific Northwest" by Dr. Evans)

(7) The 2011 Clemson Field Day Peach Growers Meeting ("GDR: A Community Resource for Rosaceae Genomics, Genetics and Breeding Research" presentation by Dr. Main)

(8) The NSF Future of Plant Genome Sequencing and Analysis Workshop (Plant Community Databases: The Stewards of Knowledge" presentaion by Dr. Main

During the course of the year, the whole team met in person to go over data and interface development and the apple team had several online meetings to further refine the functionality needs of the Toolbox.

RESULTS AND DISCUSSION
We are on track to meet all our objectives (end of March 2012). 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 and genotyping data are available. We have also developed data templates and a data uploading system that is flexible for various breeding programs. Another major work is that we identified the format of browse/query/download web interfaces that will be most useful for breeders using the WA apple and cherry breeding programs as models (Figure 2). The development of these web interfaces is completed, as are all the uploading scripts, although the breeding teams still need access to a programmer for the scripts to be run. 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 which has been integrated with private and public data.

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 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 are utilizing DNA markers to test the genotype of their breeding selections. When the RosBREED genotyping data become fully 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.
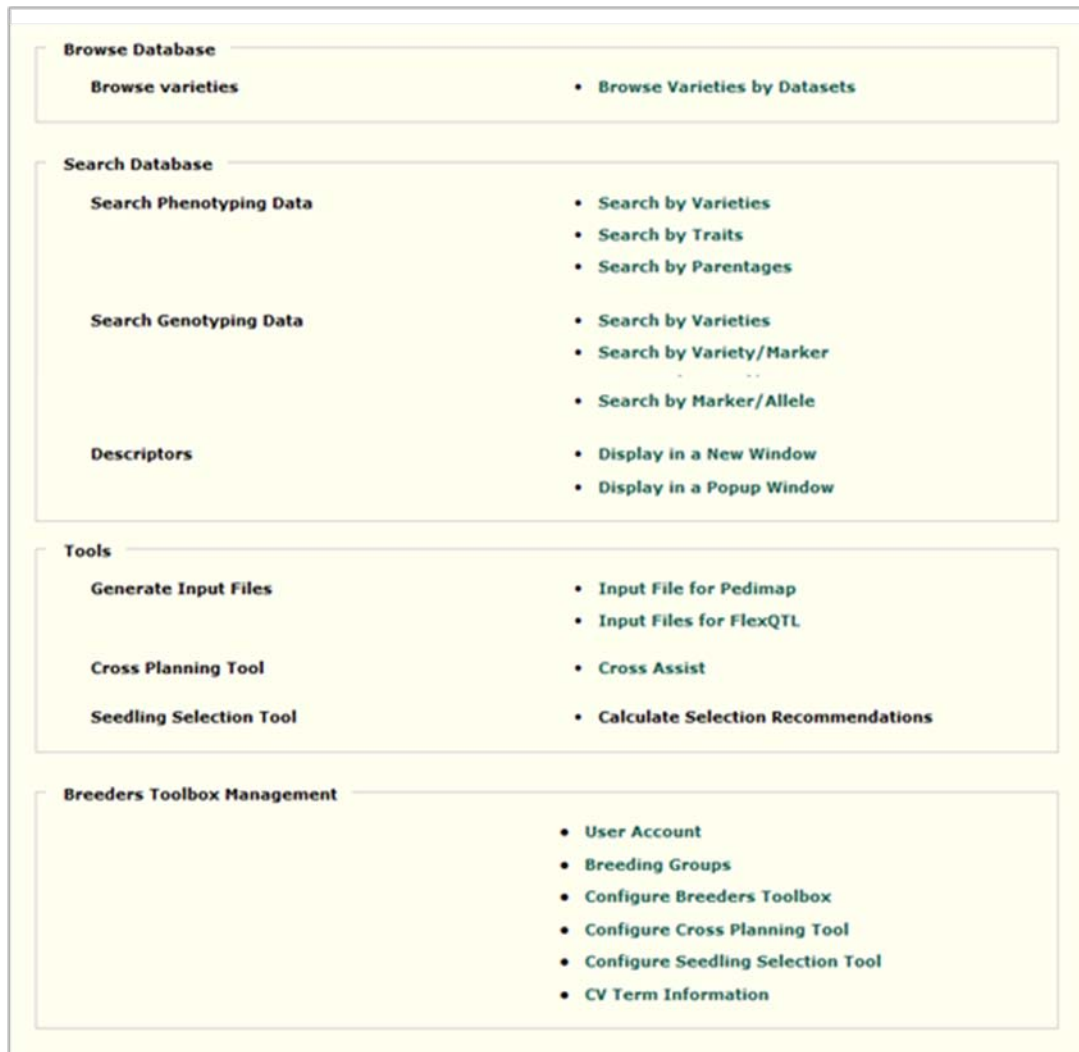

Browse Database

    Browse varieties
- Browse Varieties by Datasets

Search Database

    Search Phenotyping Data
- Search by Varieties
- Search by Traits
- Search by Parentages

    Search Genotyping Data
- Search by Varieties
- Search by Variety/Marker
- Search by Marker/Allele

    Descriptors
- Display in a New Window
- Display in a Popup Window

Tools

    Generate Input Files
- Input File for Pedimap
- Input Files for FlexQTL

    Cross Planning Tool
- Cross Assist

    Seedling Selection Tool
- Calculate Selection Recommendations

Breeders Toolbox Management
- User Account
- Breeding Groups
- Configure Breeders Toolbox
- Configure Cross Planning Tool
- Configure Seedling Selection Tool
- CV Term Information

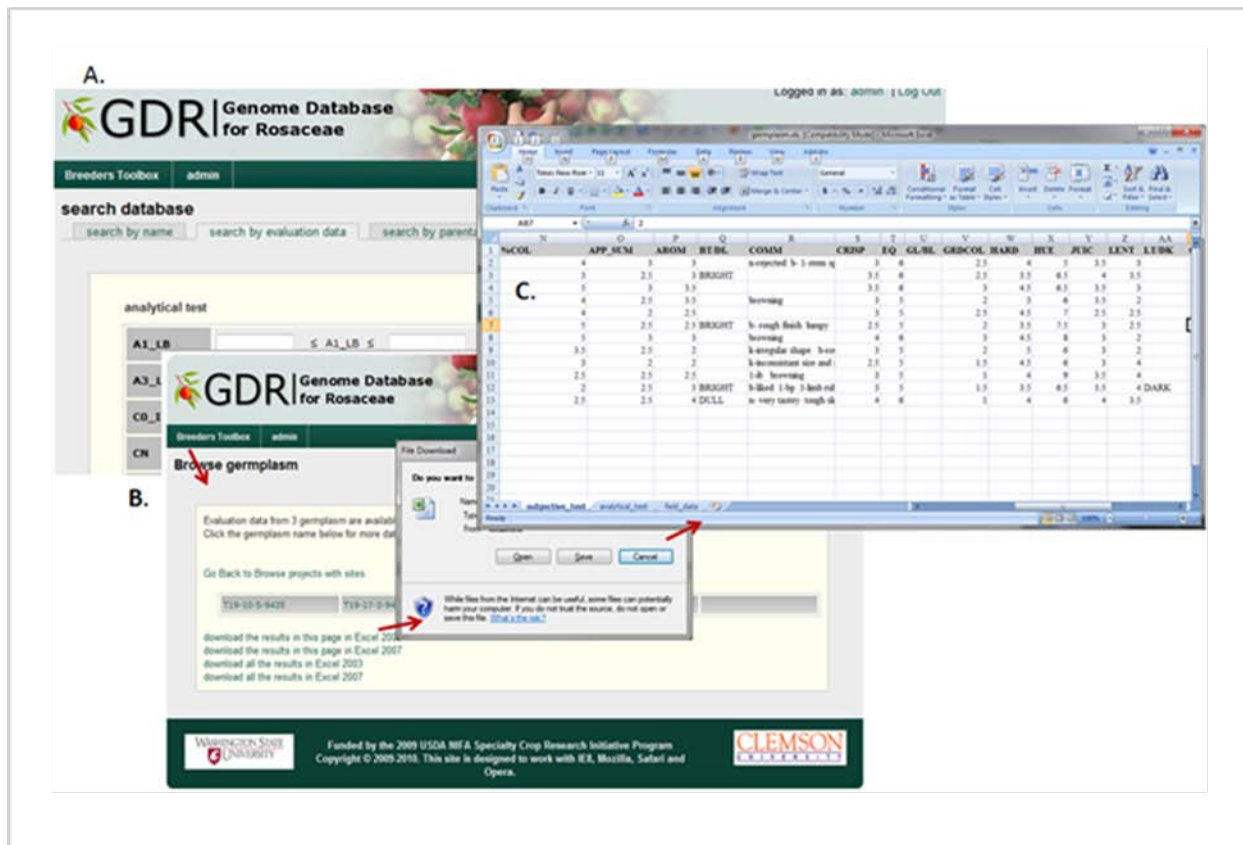**Figure 1**: The WA Apple Breeding Program Toolbox Home Page

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 Cherry Breeding program (PNWCBP) 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 (published in a peer reviewed journal in 2011) facilitating a more comprehensive genomics, genetics and breeding data structure.

Having modeled the database structure and created an excel template for breeders to store and update their breeding data, scripts were written to complete upload of all genetics and breeding data into the database by a programmer. Interfaces have been developed that enable the breeding teams to easily and quickly search and compare evaluation data by variety/germplasm, trait, marker, allele, parentage, year, and site location. The Toolbox seamlessly integrates private breeding data with the publicly available Crop Reference Set data provided by the RosBREED project, allowing the breeder to use all the available data in decision-making. This significantly enhances the breeder's ability to manage and track selection data and make more informed crossing and selection decisions, thereby facilitating the creation of new cultivars which meet consumers' needs and sustainable agricultural practices.

Further work is still needed to make the database more breeder autonomous and useful. To this end a number of features have been identified by the breeders and the development team. These include the ability for the breeding teams to directly edit germplasm and evaluation data and upload new data themselves, eliminating the need for a programmer, as is currently the case. We would also like to include more automated error checking of data to highlight where entered values are outside a specified threshold, thus allowing data to be easily user-edited. In addition, we would like to have results of searches returned to the screen (particularly useful when only a small output of data is expected), as well as be downloadable in excel and as input files for analysis in other programs that breeders routinely use, such as FlexQTL. Finally, the ability to save, update and compare searches would be a most useful and time-saving feature for the breeding teams.

We are confident that the interdisciplinary, team-based approach to creating this database and toolbox will continue to 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.