FINAL PROJECT REPORT **YEAR**: 2 of 2

WTFRC Project Number: AP14107-202

Project Title: Updating the apple map, trait and marker databases in GDR

PI:	Dorrie Main	**Co-PI (2):**	Cameron Peace
Organization:	WSU Pullman	**Organization**:	WSU Pullman
Telephone:	509 335 2774	**Telephone**:	509 335 6899
Email:	dorrie@wsu.edu	**Email**:	cpeace@wsu.edu
Address:	Dept Horticulture	**Address**:	Dept Horticulture
Address 2:	45 Johnson Hall	**Address 2**:	39 Johnson Hall
City/State/Zip:	Pullman/WA/99164	**City/State/Zip**:	Pullman/WA/99164

Co-PI(3):	Kate Evans	**Co-PI (4):**	Sook Jung
Organization:	WSU TFREC	**Organization**:	WSU Pullman
Telephone:	509 663 818	**Telephone:**	509 335 2774
Email:	kate_evans@wsu.edu	**Email:**	sook_jung@wsu.edu
Address:	Dept Horticulture	**Address:**	Dept Horticulture
Address 2:	1100 N. Western Av.	**Address 2:**	45 Johnson Hall
City/State/Zip:	Wenatchee/WA/98801	**City/State/Zip:**	Pullman/WA/99164

Cooperators: Jim Luby (University of Minnesota), Susan Brown (Cornell University), Gennaro Fazio (USDA-ARS), Yanmin Zhu (USDA-ARS), Francois Laurens (INRA, France, PI of FruitBreedomics project)

Total Project Request: **Year 1:** 26,183 **Year 2:** 27,071

Other funding sources

Agency Name: USDA-NIFA Specialty Crop Research Initiative
Amount awarded: $2.7M (Sep 2014 – Aug 2019)
Notes: “Genome Database for Rosaceae: Empowering specialty crop research through big-data driven discovery and application in breeding”. PI: Main. Co-PIs include Jung, Evans, Oraguzie, Wasko DeVetter, and Peace.

Agency Name: USDA-NIFA Specialty Crop Research Initiative
Amount awarded: $2.0 M (Sep 2009 – Aug 2014)
Notes: “Tree Fruit GDR: Translating genomics into advances in horticulture.” PI: Main. Co-PIs include Peace, Evans and Jung.

Agency Name: USDA-NIFA NRSP
Amt. requested: $1.99 M (Oct 2014- Sept 2019)
Notes: “Database resources for crop genomics, genetics and breeding research”. PI: Main. Writing team includes Jung, Peace and McFerson.

Agency Name: NSF DIBBS
Amt. *requested*: $1.48 M (Jan 2015–Dec 2017)
Notes: “Tripal Gateway, a platform for next-generation data analysis and sharing”. PI: Ficklin. Co-PIs include Main and Jung.

Agency Name: USDA-NIFA Specialty Crop Research Initiative
Amount awarded: $10.0 M (Sep 2014 – Aug 2019)
Notes: “RosBREED: Combining disease resistance with horticultural quality in new rosaceous cultivars.” PI: Iezzoni. Co-PIs include Peace, Oraguzie, Evans and Main.

Agency Name: USDA-NIFA Specialty Crop Research Initiative
Amount awarded: $7.2 M (Sep 2009 – Aug 2014)
Notes: “RosBREED: Enabling marker-assisted breeding in Rosaceae.” PI: Iezzoni. Co-PIs include Peace, Main, and Evans.

Agency Name: WTFRC
Amount *requested*: $771,688 (2015–2017)
Notes: “Apple scion breeding” PI: Evans. Co-PI: Peace.

Agency Name: WTFRC Apple Review
Amount *requested*: $107,000 (2015–2017)
Notes: “Combining fire blight resistance and horticultural quality in Washington apples” PI: Norelli. Co-PI: Evans.

Agency Name: WTFRC
Amount awarded: $862,261 (2012–2014)
Notes: “Apple scion breeding program” PI: Evans. Co-PI: Peace.

Agency Name: WTFRC/OSCC
Amount awarded: $125,000 (2014–2016)
Notes: “After RosBREED: Developing and deploying new sweet cherry DNA tests” PI: Peace.

Agency Name: WTFRC
Amount awarded: $862,261 (2012–2014)
Notes: “Apple scion breeding program” PI: Evans. Co-PI: Peace. The Washington Apple Breeding Program will be a primary beneficiary of this proposal as will the USDA-ARS rootstock breeding program.

Agency Name: WTFRC
Amount requested: $269,000 (2014–2016)
Notes: “After RosBREED: Developing and deploying new apple DNA tests” PI: Peace Co-PIs: Evans, Hardner, Main.

Budget 1

Organization Name: Washington State University **Contract Administrator:** Carrie Johnston

Telephone: (509) 335 4564 **Email address:** carriej@wsu.edu

Item	**2014**	**2015**
Salaries[a]	16,900	17,576
Benefits	5,283	5,495
Wages		
Benefits		
Equipment		
Supplies[b]	3,000	3,000
Travel[c]	1,000	1,000
Miscellaneous		
Plot Fees		
Total	26,183	27,071

[a] 0.25 FTE Dr. Sook Jung (Senior GDR data curator), 4% salary increase in year 2.

[b] Computational hardware housing cost (10%), disk backup storage supplies, minor equipment maintenance

[c] Twice annual in-person meetings for feedback and training and quarterly teleconferences (gotomeeting).

OBJECTIVES

Overall goal: To ensure the public apple map, trait and marker databases remain current in GDR and are fully integrated with the private WA Apple Breeding program to help enable efficient apple breeding.

Specific objectives:

1. Collect and curate new publicly available map, marker, trait loci and genotypic and phenotypic data for apple
2. Upload and integrate apple data within the Genome Database for Rosaceae
3. Integrate the curated apple map, marker and trait data with the WABP
4. Ensure optimal utilization of the WA Apple Breeding ToolBox and GDR through hands-on training for the WABP team and allied researchers.

SIGNIFICANT FINDINGS

1. Data from 17 publications including 20 genetic maps, 22,896 molecular markers, 192 QTL, 42, MTL (Mendelian Trait Loci), 5174 genetic loci, 49 germplasm, 45 trait terms have been curated, uploaded and integrated with other GDR data. The data includes 20K apple SNP data. The curation effort includes the association of QTLs with Trait Ontology. These data will be used to help introgress genes/loci for traits of interest in the breeding programs.
2. New apple genome assembly and annotation (Malus x domestica V3.0.a1 has been made available through genome page and genome browser.
3. QTL data from WA Apple breeding program have been used in the analysis of finding sugar related QTL that exists in the conserved syntenic regions across the Rosaceae (peach and strawberry). The information on the conserved genomic regions in apple and strawberry have been transferred to the colleagues who will develop markers for pear and blackberry.
4. First version of data editing functionality of the breeder's toolbox has been developed. We are in the process of testing and will soon provide training for the personnel in WA Apple Breeding Program.
5. WA Apple Breeding Program one of 3 breeding programs being used to develop a comprehensive Breeding Information Management System

METHODS

1. **Collect and curate new map, marker, trait loci and genotypic and phenotypic data for apple**: Using our standardized Microsoft Excel data templates we will collect and curate publicly available apple map, marker, trait loci and genotypic and phenotypic data from various sources including: (a) the RosBREED (www.rosbreed.org) project (b) the EU version of RosBREED - FruitBreedomics (www.fruitbreedomics.com) project, (c) other collaborators and (d) extracted from peer-reviewed publications. For these data to be fully integrated with other existing data and be useful to breeders and other users, additional curation effort is necessary. For example, multiple names and aliases are often used for markers, primers, germplasm, mapping population and trait descriptors. Details such as marker source organism, sequences, marker types, mapped positions of marker and trait loci and germplasm details, such as pedigree and description, can be also missing in publications and user-submitted data. We will work to standardize the names and obtain the details required to integrate data so it is useful to researchers. We will also associate trait loci with Trait Ontology (TO) so that users can browse standardized TO to find all the associated trait loci. We are actively adding more tree fruit ontology terms to TO, originally

developed to describe traits in grass species, to expand to other plant species.

2. **Upload and integrate data within the Genome Database for Rosaceae**: We will upload and integrate the apple map, marker and trait data to the GDR using our bulk data loader so that the GDR apple data mining and browsing tools remain current with up-to date data. We currently use Perl scripts to upload data in the Excel file. We will use the newly available Tripal bulk data loader. Curators need to create bulk loader templates only once using the Tripal online page and use it repeatedly to upload data in the same format. This will make adding more data types, if needed, easier than modifying Perl scripts.

3. **Integrate the curated apple map, marker and trait data with the WABP**: We will link the WABP data with the available public data and tools. The data mining tools include breeding data search tools by dataset, germplasm names, trait values, alleles and parentage. Breeders can download genotypic and phenotypic data of germplasm that meet the various categories and thresholds that users specified. In addition to the efficient management and retrieval of genotypic and phenotypic data, the integration of the breeding data with GDR will allow breeders to directly use the up-to-date genome information in DNA assisted breeding. Currently available breeding decision-support modules in GDR include *Trait Locus Warehouse*, *Marker Converter*, *Technology Portfolio* and *Cross Assist*. The *Trait Locus Warehouse* allows breeders to search for available QTLs for their trait of interest by various categories. When they find QTLs they can view them in a graphic viewer, GBrowse (Stein et al., 2002). *Marker Converter* helps breeders to view markers and re-sequencing data around the QTL to find sequence alleles that can be utilized in developing better markers for a trait locus. When breeders develop primers using the downloaded sequences, they can go to the *Technology Portfolio* to find companies who can perform the genotyping. All of these tools will be extremely useful when all the data underlying the tools, such as markers, trait loci and sequences of various germplasm and reference species, are up-to-date. We will make every effort to update the underlying data. *Cross Assist* is a breeding decision-support tool designed to predict the efficient parent combinations that can produce a target number of seedlings with specific traits thresholds specified by users. We will integrate the breeding values and DNA-based functional genotype data from the available parent pool of the WABP into *Cross Assist* to enhance Dr. Evans' cross-planning efforts.

4. **Ensure optimal utilization of the WA Apple Breeding ToolBox and GDR through hands-on training to the WABP team and other allied scientists:** We will conduct hands-on in-person training on data template completion and use of the toolbox and hold quarterly conference calls to ensure toolbox is kept current with data and functionality.

RESULTS AND DISCUSSION

Publicly available trait and marker data that is relevant to the WABP has been added to GDR and more performance and genotypic data specific to the breeding program will continue to be added over the next four years using funding from the USDA SCRI award to PI Main (co-PI's Jung, Evans and Peace). These additions will continue to provide an up-to-date breeding information management system for the WABP to facilitate routine marker-assisted breeding and more efficient development of new cultivars for WA apple growers. The additions of the QTL data from other Rosaceae crops, as well as from apple, enabled us to perform analysis to find QTL in evolutionarily conserved genomic regions across the Rosaceae crops. Ties with the other breeding projects, such as RosBreed projects in US and European FruitBreedomics project, are strong, with data-sharing strategies in place. Significant funding of over $10 M has been obtained from federal sources by PI Main to continue to develop crop resources (GDR) and tools (markers for traits of economic importance) that will help make apple tree fruit breeding in WA even more efficient, accurate, creative, and rapid and keep WA apple growers competitive on a global scale. This funding allows us further development of breeder’s toolbox into a comprehensive Breeding Information Management System (BIMS) that will be incorporated within the Tripal generic database platform which has been led by Washington State University researchers. It will provide breeders full privilege for uploading, editing, management, analysis and archiving of various breeding data and will include access to statistical packages such as R.

EXECUTIVE SUMMARY

Tree fruit breeding programs generate copious amounts of data. Utilizing this data requires proper management plans and interrogation tools to enable breeders to efficiently mine their data and extract what they need to enable more efficient breeding. Concordant with this is the need to also access all relevant public information such as what’s known about traits, markers for these traits, germplasm containing useful traits in the same and related crops. Within the Rosaceae community database, GDR (www.rosaceae.org), a private breeding database for the WA Apple Breeding Program exists, connecting the programs private breeding data with all publicly available, quality checked, genomic, genetics and breeding data for Malus crops. Searchable interfaces allow the data to be searched by trait, trait levels, location, marker, pedigree, germplasm, year, etc and tools enable download of data for upload to analysis programs. The Cross-Assist tool takes this concept one step further. Using component data from the breeding program it outputs the optimal parents to cross and numbers of seedlings needed to generate the desired offspring and eventually new cultivar(s) which meet producer and consumer needs.

The Apple Toolbox in GDR will continue to be expanded to a more comprehensive Breeding Information Management System developed using the Tripal Database Platform and will be updated with the WAABP breeding data following the end of this WTFRC award. Funds from this WTFRC project were leveraged to support federal crop database proposals that have generated over $5M in new funds over the next 4 years. Tools to allow direct upload of the data are developed, which enable breeders to be responsible for managing their own data. The Field Book App currently being evaluated for this purpose is looking very promising for this purpose and we are working with its developers to optimize for tree fruit breeding. We will continue to develop the GDR breeding tools into a comprehensive breeding management system that provides secure, one stop access to all the data management and analysis tools that Rosaceae breeders and allied scientists need to more efficiently develop new cultivars.