FINAL PROJECT REPORT

Project Title: After RosBREED: Developing and deploying new apple DNA tests

PI:	Cameron Peace	**Co-PI(2)**:	Kate Evans
Organization:	WSU Pullman	**Organization**:	WSU TFREC
Telephone:	509 335 6899	**Telephone**:	509 663 8181 ext 245
Email:	cpeace@wsu.edu	**Email**:	kate_evans@wsu.edu
Address:	Dept of Horticulture, WSU	**Address**:	TFREC
Address 2:	PO Box 646414	**Address 2**:	1100 N Western Ave
City:	Pullman	**City**:	Wenatchee
State/Zip:	WA 99164-6414	**State/Zip**:	WA 98801
Co-PI(3):	Craig Hardner	**Co-PI(4)**:	Dorrie Main
Organization:	University of Queensland	**Organization**:	WSU Pullman
Telephone:	+61 7 3346 9465	**Telephone**:	509 335 2774
Email:	c.hardner@uq.edu.au	**Email**:	dorrie@wsu.edu
Address:	QAAFI	**Address**:	Dept of Horticulture, WSU
Address 2:	St Lucia	**Address 2**:	PO Box 646414
City:	Brisbane	**City**:	Pullman
State/Zip:	QLD, Australia 4072	**State/Zip**:	WA 99164-6414

Cooperators: Paul Sandefur, Sushan Ru, and Feixiong Luo (PhD students, WSU Pullman), Daniel Edge-Garza, Julia Piaskowski, and Sook Jung (WSU Pullman), Jay Norelli (USDA-ARS Kearneysville), Fred Bliss (Davis, California)

Other funding sources

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

Agency Name: WTFRC Apple Review
Amount awarded: $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: $53,254 (2014–2015)
Notes: "Adding apple map, marker and trait data to the Genome Database for Rosaceae" PI: Main. Co-PIs: Evans, Peace, 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, and Main.

Total Project Funding: $269,000

Budget History:

Item	**Year 1:** 2014	**Year 2:** 2015	**Year 3:** 2016
Salaries	51,008	52,249	53,540
Benefits	16,127	16,965	17,850
Wages			
Benefits			
Equipment	9,865	9,786	9,610
Supplies	2,000	2,000	2,000
Travel	5,000	8,000	8,000
Plot Fees			
Miscellaneous	5,000		
Total	89,000	89,000	91,000

RECAP ORIGINAL OBJECTIVES

Overall goal

Improve prospects for apple breeding efficiency, accuracy, creativity, and pace by developing and strategically deploying predictive DNA tests targeting valuable traits for the WSU Apple Breeding Program (WABP)

Specific objectives

1. DNA test development:
    a. Develop new DNA tests, first for current genomics discoveries (acidity, sweetness, firmness), and continue with future discoveries (maturity time, size, texture, storage disorders)
    b. Establish a streamlined statistical approach to predict performance from DNA test outcomes

2. DNA test deployment strategies:
    a. Deploy new DNA tests strategically by devising and trialing strategies for the WABP aligned with existing tests and breeding operations; host an international workshop on this topic
    b. Establish a streamlined statistical approach for DNA test deployment under complex scenarios

SIGNIFICANT FINDINGS

Objectives 1a, 1b, and 2a were accomplished in this three-year project (Figure 1). Objective 2b was partially accomplished. For 1a, at least two new trait-predictive DNA tests were developed each year and some previous DNA tests were refined. For 1b, software was developed for determining robust estimates of DNA test effects. For 2a, strategies for optimal deployment of multiple DNA tests were determined that account for various influencing factors. For 2b, working software was developed to support the critical cross-planning stage. Remaining elements of this software will be developed in 2017 with remaining funds.

Objective 1a

- New DNA tests developed or refined that account for <u>some</u> of the genetic influences variable in <u>most</u> WABP families and target <u>essential</u> thresholds for the following traits: **fruit firmness** (two genomic regions) and **crispness** (second region), **fruit acidity** (second region), **fruit texture** (combined test)
  Previously available DNA tests in this category: **fruit ethylene/storability** (two genomic regions), **fruit crispness** (first genomic region) and **juiciness**, **fruit acidity** (first genomic region), **fruit bitter pit incidence**

- New DNA tests developed or adapted that account for <u>most/all</u> of the genetic influences variable in <u>some specific</u> WABP families and target <u>essential</u> thresholds (just for those families) for the following traits: **powdery mildew resistance ('White Angel' source)**, **pink flesh color ('Pink Pearl' source)**

- New DNA test developed that accounts for some of the genetic influences variable in some specific WABP families and targets essential threshold for the following trait: **blue mold resistance (a *M. sieversii* source)**

- New DNA test developed that accounts for most/all of the genetic influences variable in most WABP families and targets enhancing threshold for the following trait: **fruit fructose content**
  Previously available DNA test in this category: **skin overcolor amount**

- New DNA test adapted for the WABP that reveals important allelic information for parents and elite selections/new cultivars: ***S*-genotyping for cross-compatibility**

- DNA test development for several other traits is still underway: harvest timing, fire blight resistance, and soft scald

- Note that there are no DNA tests available accounting for most/all of the genetic influences for essential attributes in most/all WABP germplasm. This situation reflects the biological nature of WABP goals and germplasm – numerous genetic factors contribute to the most important attributes considered to be essential for commercial success in Washington. Therefore, use of DNA tests in the WABP should enhance efficiency and accuracy of selection but is not expected to lock in particular attributes.

Objective 1b

- *DNA Test Effects*, software to calculate DNA test effects from datasets of multi-family replicated trials, developed. Provides a streamlined statistical approach to predict performance from DNA test outcomes. Currently being used to update all DNA test effect predictions

Objective 2a

- Workshop "DNA Test Deployment Strategies for Rosaceae Crop Breeding" hosted in 2014. Well attended by international researchers and affiliated scientists, keeping us on the cutting edge

- Framework developed for DNA test deployment strategies that consider essential vs. enhancing trait levels, cost and genetic gain efficiencies, operational logistics, and which particular germplasm is relevant; scientific papers published

Objective 2b

- *Multi-Trait Family Planning*, software to predict numbers of seedlings and their distributions of trait levels for hypothetical families, developed. Provides a streamlined statistical approach for preparing DNA test deployment during cross-planning and greenhouse-stage seedling selection

- An interim DNA test deployment strategy is available for the WABP, prior to more sophisticated software. The major deployment decisions are whether to use a DNA test only for parent selection or for both parent selection and seedling selection, and for most or for very specific families

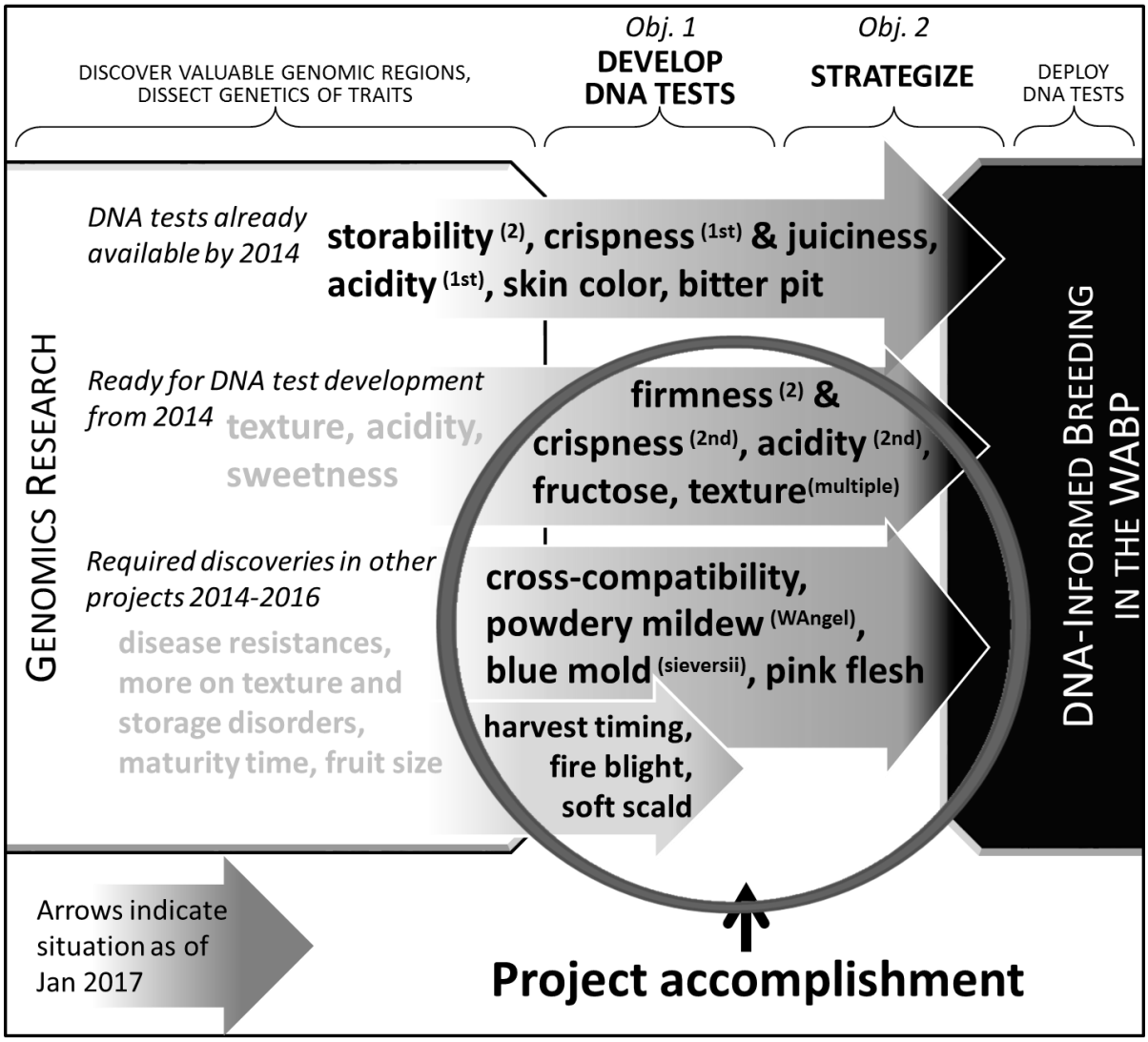

Figure 1: Progress made in 2014-2016 to develop, adapt, and refine DNA tests and establish strategies for their deployment in the WSU Apple Breeding Program (WABP).

RESULTS AND DISCUSSION

Activity 1a: DNA test development

Nine DNA tests were newly developed, adapted from publications, or refined for availability to the WABP. Previously available DNA tests were for fruit ethylene/storability (Md-ACS-indel and Md-ACO-indel), fruit crispness and juiciness (Ma-indel), fruit acidity (Ma-indel), and fruit bitter pit incidence (Bp16-SSR). Below are the 2014-2016 additions and refinements.

Fruit texture and storability (firmness, crispness, juiciness, especially after storage): A combined DNA test as well as individual tests are now available for four genomic regions that influence components of fruit texture and are variable within WABP germplasm. The first two genomic regions

are the genes encoding the ethylene biosynthesis enzymes ACS1 and ACO1 (targeted by DNA tests Md-ACS1-indel and Md-ACO1-indel). These DNA tests detect a small but significant effect on firmness especially at harvest. The third genomic region is the *Ma* locus (targeted by the DNA test Ma-indel). Among descendants of 'Honeycrisp', this DNA test detects a large difference in firmness at any storage duration and a large difference in crispness especially after storage. The fourth genomic region is the gene encoding the cell wall metabolizing enzyme of PG1 (targeted by the DNA test Md-PG1$_{SSR}$10kdb). This fourth DNA test is associated with large differences in firmness, crispness, and juiciness after long storage. These DNA tests can be run in a single assay all at once (we call it the texture "Wonder test") or individually. In addition, the Ma-indel test is predictive for some other traits (below).

Fruit acidity: The marker LG8A-SSR was developed for the second of two genomic regions associated with large genetic differences among WABP individuals for fruit acidity. LG8A-SSR when combined with the previous Ma-indel DNA test enhances prediction of apple acidity. Ma×A Acidity is the new combined DNA test that targets both genomic regions.

Fruit fructose content: A new DNA test, Md-LG1Fru-SSR, was developed for prediction of fruit fructose content. This DNA test differentiates almost all of the genetic differences observed in WABP seedlings for fruit fructose content (and some other sugars).

Disease resistances (foliar powdery mildew, blue mold) from particular sources: New DNA tests were developed for the 'White Angel' source of foliar powdery mildew resistance (presence of resistance allele is associated with strong resistance), Md-Plw-SSR, and a *Malus sieversii* source of resistance to the postharvest disease, blue mold (presence of resistance allele is associated with some resistance), Md-Pe3-SSR. The powdery mildew and blue mold DNA tests are only relevant for "pre-breeding" families that have used the wild sources in recent generations. These DNA test advances have been achieved by Pullman-based PhD student Feixiong Luo, supported by the China Scholarship Council, with guidance from pathologist and geneticist Dr. Jay Norelli (USDA-ARS Kearneysville, WTFRC- and RosBREED-funded) and PIs Peace and Evans. New progenies will be generated and inoculated with powdery mildew to validate the Md-Plw-SSR test.

Pink flesh color: Md-S3-indel, a DNA test for "Type 2" pink flesh, which identifies the allele associated with pink derived from 'Surprise' and 'Pink Pearl', was adapted from an existing marker for the S3 allele of the apple *S* locus. The *S* locus is closely linked to the *MYB110a* gene associated with flesh color, and the S3 allele is associated with the pink-flesh *MYB110a* from 'Pink Pearl'. This DNA test is only relevant for WABP families descended from 'Pink Pearl'.

Cross-compatibility: A recently reported "universal" DNA test for the *S* locus was adapted for the WABP. This DNA test, Md-S-universal, reveals most of the *S*-alleles present in the WABP, especially the common ones. Several allele-specific tests have also been adapted to detect further alleles. However, several WABP alleles have not yet been determined. In most cases, the test(s) can determine if two parents or selections carry the same two *S*-alleles and therefore would not be able to cross with each other.

Others in development: DNA tests are currently in development for two sources of fire blight tolerance and a genomic region for blue mold tolerance (tolerance alleles for all of these are present in elite WABP parents). DNA tests are also in development for major-effect genomic regions reported in the literature or detected by RosBREED collaborators to be associated with some of the genetic influences on harvest timing and soft scald.

Activity 1b: DNA test effects calculations

Software that we named *DNA Test Effects* was developed in R programming language implemented in ASReml. This software calculates the trait levels and variability associated with DNA test outcomes (i.e., genotypes, aka allelic combinations). Input data are from multi-family replicated trials – specifically for our current use is the dataset of RosBREED 1 in which performance data for many traits were recorded and genome scans were obtained on WABP germplasm (and two other U.S. apple breeding programs) for many hundreds of seedlings, selections, and cultivars. *DNA Test Effects* uses a statistical genetics model to estimate not only the relative effects on a trait of a DNA test of interest but also the "genetic background" (the cumulative effects of all the other genetic influences on the trait, large and small), fixed external effects such as year and location, statistical interactions between the DNA test and external effects (such as certain alleles whose influence is only manifested in some years, for example cold wet ones), and residual effects (effectively noise). This software provides a streamlined statistical approach to predict performance from DNA test outcomes, and is currently being used to update estimates of all DNA tests on their target trait(s). Next, the software will be used to estimate effects of our DNA tests on dozens of other traits of WABP interest that were measured in the RosBREED 1 project.

Activity 2a: Devising DNA test deployment strategies

Available DNA tests cannot simply be used all at once on all germplasm – it's much more complicated than that. The four major factors underlying deployment are value of the trait levels differentiated (trait levels), how well genetic differences among breeding program individuals are captured (predictiveness), which particular families or other germplasm are relevant (germplasm), and how much any given test is associated with another test or other traits (genetic complexity). The strategies associated with variations in these factors were discussed, modeled, calculated, and described in scientific papers. We developed a conceptual framework for capturing the above features of available DNA tests.

A one-day "DNA Test Deployment Strategies for Rosaceae Crop Breeding" workshop was hosted at WSU-TFREC in Wenatchee on 23 June 2014. The event was well attended, with more than 40 participants from at least 15 countries. Experiences, successes, constraints, and opportunities to deploying DNA information for parent selection, seedling selection, and elite candidate selection were discussed, with many valuable contributions from participants. The workshop outcomes, including subsequent presentations and scientific papers on the topic, are keeping our fruit breeding programs on the cutting edge.

Detailed considerations of cost-, time-, and genetic gain-efficiency, as well as logistical feasibility, were described in detail in this project's second-year continuing report. Scientific papers arising from this work are listed below. More than a dozen professional meeting presentations have also been made on this topic over the last there years.

- Edge-Garza D, Luby J, and Peace C (2015). Decision support for cost-efficient and logistically feasible marker-assisted seedling selection in fruit breeding. Molecular Breeding 35:223
- Ru S, Hardner C, Carter PA, Evans K, Main D, and Peace C (2016). Modeling of genetic gain for single traits from marker-assisted seedling selection in clonally propagated crops. Horticulture Research 3:16015
- Evans K, and Peace C. Advances in marker-assisted breeding for apple. In (ed. K. Evans) Achieving Sustainable Cultivation of Apples, Burleigh Dodds (in press)

- Peace C. DNA-informed breeding of rosaceous crops: Promises, progress, and prospects. Horticulture Research (submitted)

Two examples of DNA test deployment strategies considering trait levels, predictiveness, germplasm, and genetic complexity are described below.

Example 1: Ma-indel for multiple traits

The Ma-indel DNA test is in a hotbed for trait influences of importance to the WABP. The test itself lies inside a gene strongly influencing acidity content – in fact, in some families it is possible to know the Ma-indel genotype by tasting fruit (of course it is more efficient to figure this out by DNA testing seedlings many years ahead of actual fruiting). Individuals with two copies of the allele associated with low acidity are often too bland, while those with two high-acidity alleles are often too acidic. A second genomic region influencing acidity determines whether the double-low or double-high Ma-indel seedlings will be pushed over the edge beyond acceptable WABP thresholds. Ma-indel, especially when incorporated into the more comprehensive Ma×A Acidity test, therefore differentiates essential trait levels, indicating that it warrants use in parent selection (to help choose crosses likely to result in seedlings with fruit acidity not bland and not too high) and seedling selection (to cull any seedlings generated with extreme acidity genotypes). The predictiveness of the Ma×A Acidity test is medium, whereby the DNA test accounts for about half of genetic influences on acidity in WABP germplasm. Combining that predictiveness with the fact that the influence on acidity of all genetic factors variable in WABP germplasm is high means that, according to the framework of Ru et al. (2016), the DNA test should be used to avoid and cull the extreme genotypes in parent and seedling selection, respectively. The extreme alleles are common in WABP germplasm, and so most families deserve attention when considering Ma×A Acidity.

Finally, the genetic complexity of Ma-indel is in two main ways. The first is its connection with the second genomic region influencing acidity, as described above. The second is that genes influencing several other traits – crispness, firmness, bitter pit incidence, phenolics content, fruit size, sweetness, and others – are located adjacent to the acidity gene that Ma-indel targets. Because Ma-indel reveals numerous alleles (several associated with high acidity, several with low, and one medium), each allele can be associated with certain levels of those other traits. For example, one of the Ma-indel alleles from ‘Honeycrisp’ is associated with medium acidity, lower crispness after storage, lower firmness at any point, lower bitter pit incidence, slightly increased size, and slightly decreased sweetness. The other ‘Honeycrisp’ allele (also inherited by ‘WA 38’) is associated with lower acidity, higher crispness after storage, higher firmness, higher bitter pit incidence, slightly decreased size, and slightly increased sweetness. We believe the second allele is an essential component to the ultra-crisp texture of ‘Honeycrisp’ and ‘WA 38’, and that case the Ma-indel DNA test targets an essential trait level for some families. The other traits influenced can be mitigated by alleles at other genomic regions (and the second Ma-indel allele carried by any individual), as attested by the high acidity, lack of bitter pit, and large size of ‘WA 38’. Therefore, the information provided by Ma-indel for those other traits can be considered as targeting enhanced trait levels, and weighed up as part of many contributors to those traits during parent and seedling selection.

Example 2: Pink flesh

Pink or red flesh color vs. white flesh color of apple fruit is conditioned by genetic variants at just two genomic regions. What is more commonly called “red flesh”, or “Type 1 red flesh”, is due to a rare allele (originally from a subtype of *M. sieversii* called *M. niedzwetzkyana*) at the same gene as conditions skin overcolor amount. Pink flesh color, or “Type 2 red flesh”, is conditioned by a different gene on another chromosome, and the pink-flesh allele is from ‘Surprise’, an old cultivar, and some of its offspring including ‘Pink Pearl’. The DNA test for pink (vs. white) flesh, Md-S3-indel, is only relevant for WABP families descended from ‘Pink Pearl’. Therefore, this DNA test

targets a trait level that is essential but only in particular germplasm being purposely advanced for combining pink flesh with other elite attributes, and has high predictiveness for the trait. The Type 2 flesh color gene's genetic complexity is its genetic linkage with the *S* locus such that individuals with pink flesh will also carry the S3 allele – which can be exploited by making crosses that allow fertilization only by the S3-carrying pollen, most of which should also carry the pink flesh allele. These features indicate that effective deployment of Md-S3-indel in the WABP would be to use the DNA test's information fully during parent and seedling selection but only for those families intended to introduce the pink flesh attribute.

Activity 2b: Implementing DNA test deployment strategies

Software is needed to capture the many variables in these considerations. A software tool, *Multi-Trait Family Planning* was developed to target the most critical deployment stage: cross planning and greenhouse-stage seedling selection. Given known DNA test genotypes of pairs of parents being considered to create a hypothetical family with a user-chosen initial number of seedlings, this Microsoft Excel-based tool predicts and graphically displays genotypic outcomes. Such outcomes are in terms of the numbers of seedlings and their distributions of trait levels. (Trait effect estimates are determined by results from *DNA Test Effects* (Activity 1b) and can be updated as desired.) The user can examine the predicted effect of selecting for/against certain seedling genotypes on the trait level distribution of remaining seedlings. Because some DNA tests are known to influence traits other than what they were developed for, the tool calculates and displays the predicted outcomes of selecting with one DNA test on up to four other traits. If there are DNA tests available underlying those other traits, the user can continue the exploration of predicted outcomes of selecting with the next DNA test, and the next, and the next. With information on certain DNA tests already pre-loaded, all of these applications of the Excel tool could also be used in real time during marker-assisted seedling selection operations each spring. In this case, rather than predicted proportions of seedlings in each genotypic class, actual data from the DNA testing lab is used (Figure 2).

With remaining funds, in 2017 the Excel tool will be improved in 2017 with additional functions and possibly conversion to stand-alone software in a programming language such as R or with R-based functions connected to the Excel tool. An additional functions will be the ability to consider seedlings with unintended parentage, which DNA testing of seedlings reveal in most families usually at low levels. Another function will be the ability to automatically populate genotypic information on parents when their cross number entered, rather than currently having to enter such information by hand.

Prior to more sophisticated software, an interim DNA test deployment strategy in place for WABP. The major consideration is whether to use the DNA test only for parent selection (P) or for both parent selection and seedling selection (P+S). Decision factors are whether the trait level is essential (P+S) or enhancing (P), the DNA test is highly predictive (P+S) or somewhat predictive (P), and which families are relevant (deployment only for families expected to carry both good and bad alleles).

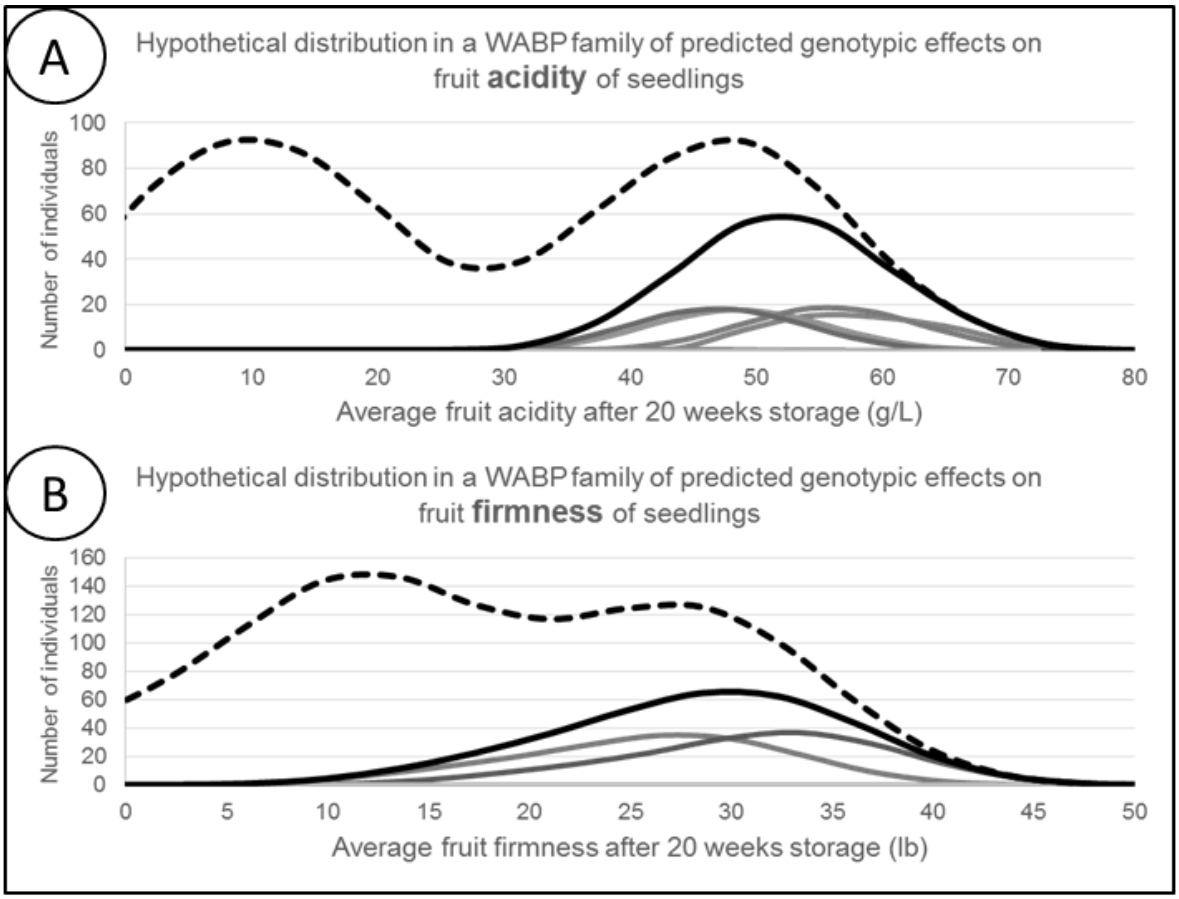

Figure 2: Example output of software Multi-Trait Family Planning*, Excel-based software to support cross-planning and seedling selection. Dotted lines represent estimated distributions of trait performance potential of seedlings prior to any culling. Thick continuous black lines represent the estimated distributions of seedlings after culling. Lighter gray lines represent estimated seedling distributions within specific genotypic classes for DNA tests underlying the trait. Note that input data in all cases here is hypothetical, created for the purposes of demonstration; our true estimated effects of DNA tests on these traits are not exactly the same.*

(A) Acidity distribution of a family after culling for low-acid genotypes associated with the DNA test Ma-indel. Note a substantial shift to the right (higher acidity) as well as a large reduction in total number of seedlings.

(B) Effects on seedling distribution for firmness in the same family as above after the culling for low-acid genotypes. Note that as well as fewer total seedlings there is a shift to the right – indicating that in this case the alleles associated with higher acidity were also associated somewhat with higher firmness. The software can model alternative situations too. The two gray lines represent estimated firmness distributions for two seedling genotypes of the next DNA test that could be considered by the breeder, Md-PG1SSR10kdb.

Continued next page.

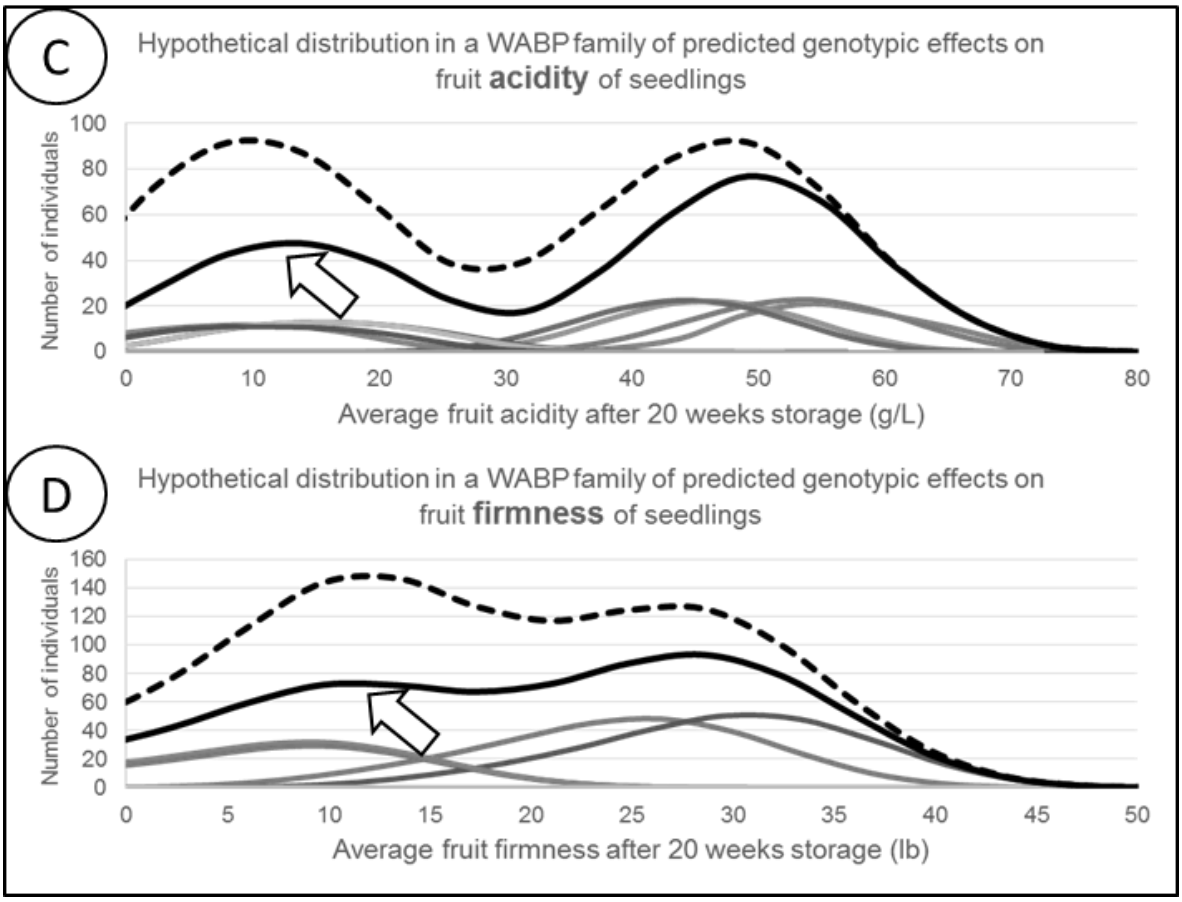

Figure 2 cont'd: (C) and (D) are the same as for previous (A and B), showing estimated distributions for acidity and firmness, except that there is retention of seedlings carrying a particularly desirable genotype at some other locus (in this case, high fructose content according to DNA test Md-LG1Fru-SSR). The main differences with (A) and (B) are that some seedlings with lower acidity and firmness are retained, pointed out by the two arrows, and more seedlings are retained overall.

EXECUTIVE SUMMARY

This project was about supporting the WABP with trait-predictive DNA information: New DNA tests were developed. Strategies to combine these new ones with already-available tests were devised. Software to improve both DNA test development and DNA test deployment were also developed.

New DNA tests were developed, adapted, or refined for the traits of: **fruit firmness** (two genomic regions) and **crispness** (a second region), **fruit acidity** (second region), **fruit texture** (combined test), **powdery mildew resistance ('White Angel' source)**, **pink flesh color ('Pink Pearl' source)**, **blue mold resistance (a *M. sieversii* source)**, **fruit fructose content**, and ***S*-genotyping for cross-compatibility**. DNA test development for several other traits is still underway: harvest timing, fire blight resistance, and soft scald.

The above DNA tests add to previous DNA-based tools for the traits of: fruit ethylene/storability (two genomic regions), fruit crispness (first genomic region) and juiciness, fruit acidity (first genomic region), fruit bitter pit incidence, and skin overcolor amount.

A new software tool, *DNA Test Effects*, was developed that calculates trait levels and variability associated with DNA test outcomes. This software provides a streamlined statistical approach to predict performance from DNA test outcomes, and is currently being used to update estimates of all DNA tests on their target traits. Next, the software will be used to estimate effects of our DNA tests on many other traits of WABP interest measured in the RosBREED 1 project.

Available DNA tests cannot simply be used all at once on all germplasm – it's much more complicated than that. The four major factors of each DNA test underlying their deployment in the WABP are value of the trait levels differentiated (trait levels), how well genetic differences among breeding program individuals are captured (predictiveness), which particular families or other germplasm are relevant (germplasm), and how much any given test is associated with another test or other traits (genetic complexity). The strategies associated with variations in these factors were discussed, modeled, calculated, and described in scientific papers. We developed a conceptual framework for capturing the above features of available DNA tests. The most critical components were distilled for objective DNA test deployment.

Software is needed to fully capture the many variables in the above considerations. A software tool, *Multi-Trait Family Planning* was developed to target the most critical deployment stage – cross planning and greenhouse-stage seedling selection. This software will be extended in 2017.