

## **RosBREED: Enabling marker-assisted breeding in Rosaceae**

RosBREED will create a national, dynamic, sustained effort in research, infrastructure establishment, training, and extension for applying marker-assisted breeding (MAB) to deliver improved plant materials more efficiently and rapidly. The Rosaceae family (including apple, peach, sweet and tart cherries, and strawberry) provides vital contributions to human health and well-being, and collectively constitutes the economic backbone of many U.S. rural communities. Rosaceae genetics and genomics are developing rapidly but have not been translated to routine practical application. Specific objectives are to: (1) increase the likelihood of new cultivar adoption, enlarge market potential, and increase consumption of rosaceous fruits by using socio-economic knowledge of stakeholder values and consumer preferences to inform breeding; (2) establish sustainable technical infrastructure for an efficient MAB Pipeline in Rosaceae, including crop-specific SNP genome scan platforms for breeding-relevant germplasm exploiting the shared ancestry of Rosaceae crops; (3) integrate breeding and genomics resources by establishing a user-friendly U.S.-wide standardized statistical framework and breeding information management system; (4) implement MAB in core RosBREED breeding programs with a common focus on fruit quality traits; and (5) enhance sustainability of cultivar development by transferring MAB technologies to the public and private community of U.S. Rosaceae breeders through training current and future breeders as well as engaging the production, processing and marketing sectors, allied scientists, and consumers. This Coordinated Agricultural Project addresses SCRI Focus Area 1: Research in plant breeding, genetics and genomics to improve crop characteristics (80%), Focus Area 2 (10%), and Focus Area 3 (10%).

# EXECUTIVE SUMMARY

## PROJECT TITLE

RosBREED: Enabling marker-assisted breeding in Rosaceae

## PROJECT TYPE

Coordinated Agricultural Project

## FOCUS AREAS

Focus Area 1. Research in plant breeding, genetics, and genomics to improve crop characteristics (80%)

Focus Area 2. Efforts to identify and address threats from pests and diseases, including threats to specialty crop pollinators (10%)

Focus Area 3. Efforts to improve production efficiency, productivity, and profitability over the long term (10%)

## PROGRAM STAFF

---

### Project Director:

Amy Iezzoni - Professor, breeder  
iezzoni@msu.edu

Dept of Horticulture, Michigan State Univ.  
E. Lansing, MI 48824

---

### Co-PDs: Team Leaders:

James Luby - Professor, breeder  
lubyx001@umn.edu

Team Leader: Breeding  
Dept of Horticultural Science, Univ. of  
Minnesota, St. Paul, MN 55108

Chengyan Yue - Asst Professor  
yuechy@umn.edu

Team Leader: Socio-Economics  
Depts of Hort. Science and Applied Economics,  
Univ. of Minnesota, St. Paul, MN 55108

Eric van de Weg - Senior Scientist  
eric.vandeweg@wur.nl

Team Leader: Pedigree-Based Analysis  
Plant Research International, 6700AA,  
Wageningen, Netherlands

Gennaro Fazio - Geneticist, breeder  
gf35@cornell.edu

Team Leader: Breeding Information Management System  
USDA-ARS Plant Genetic Resources Unit,  
Cornell Univ., Geneva, NY 14456

Dorrie Main - Assoc. Professor  
dorrie@wsu.edu

Team Leader: Genomics  
Dept of Horticulture & Landscape Arch.,  
Washington State Univ., Pullman, WA 99164

Nahla Bassil: Geneticist  
nahla.bassil@ars.usda.gov

Team Leader: Genotyping  
USDA-ARS National Clonal Germplasm  
Repository, Corvallis, OR 97333

Cameron Peace - Asst Professor  
cpeace@wsu.edu

Team Leader: MAB Pipeline  
Dept of Horticulture & Landscape Arch.,  
Washington State Univ., Pullman, WA 99164

Cholani Weebadde - Asst Professor  
weebadde@msu.edu

Team Leader: Outreach  
Dept of Horticulture, Michigan State Univ.,  
East Lansing, MI 48824

---

**Co-PDs: Additional Team members:**

Marco Bink - Senior Scientist  
marco.bink@wur.nl

Plant Research International,  
6700AA, Wageningen Netherlands

Susan Brown - Professor  
skb3@cornell.edu

Dept of Horticulture, Cornell Univ., NYSAES,  
Geneva, NY 14456

David Byrne - Professor  
d-byrne@tamu.edu

Dept of Horticultural Sciences, Texas A&M,  
College Station, TX 77843

John Clark - Professor  
jclark@uark.edu

Dept of Horticulture, Univ. of Arkansas,  
Fayetteville, AR 72701

Carlos Crisosto - Professor  
carlos@uckac.edu

Dept of Plant Sci., Univ. of California - Davis,  
Kearney Agric. Center, Parlier, CA 93648

Thomas Davis - Professor  
tom.davis@unh.edu

Dept of Plant Biology, Univ. of New  
Hampshire, Durham, NH 03824

Kate Evans - Assoc. Professor  
kate\_evans@wsu.edu

Dept of Horticulture & Landscape Arch.,  
Washington State Univ., Wenatchee, WA 98801

Chad Finn - Research Geneticist  
finnc@science.oregonstate.edu

USDA-ARS Horticultural Crops Research  
Laboratory, Corvallis, OR 97330

Karina Gallardo - Asst Professor  
karina\_gallardo@wsu.edu

TFREC, Washington State Univ.,  
Wenatchee, WA 98801

Ksenija Gasic - Asst Professor

Dept of Horticulture, Clemson Univ.,

kgasic@clemson.edu	Clemson, SC 29634
Tom Gradziel - Professor tmgradziel@ucdavis.edu	Dept of Plant Sci., Univ. of California - Davis, 1 Shields Ave., Davis, CA 95616
James Hancock - Professor hancock@msu.edu	Dept of Horticulture, Michigan State Univ., E. Lansing, MI 48824
Raymond Jussaume - Professor rajussaume@wsu.edu	Dept of Community & Rural Sociology, Washington State Univ., Pullman, WA 99164
Vicki McCracken - Professor mccracke@wsu.edu	School of Economic Sciences, Washington State Univ., Pullman, WA 99164
Nnadozie Oraguzie - Assoc. Professor noraguzie@wsu.edu	Dept of Horticulture & Landscape Arch., Washington State Univ., Prosser, WA 99350
Gregory Reighard - Professor grghrd@clemson.edu	Department of Horticulture, Clemson University, Clemson, SC 29634
Alexandra Stone - Professor stonea@hort.oregonstate.edu	Dept of Horticulture, Oregon State Univ., Corvallis, OR 97331
Mykel Taylor - Asst Professor m_taylor@wsu.edu	School of Economic Sciences, Washington State Univ., Pullman WA, 99164
Dechun Wang - Assoc. Professor wangdech@msu.edu	Dept of Crop & Soil Sci., Michigan State Univ., E. Lansing, MI 48824
Kenong Xu - Asst Professor kx27@cornell.edu	Dept of Horticultural Science, Cornell Univ., NYSAES, Geneva, NY 14456

---

## **STAKEHOLDER NEED ADDRESSED**

The U.S. Rosaceae crop industries face numerous limitations to profitability and sustainability. Overcoming these barriers requires rapid development and deployment of new cultivars with improved characteristics to meet dynamic industry and market needs and consumer preferences. This project seeks to identify breeding trait targets based on knowledge of what industry sectors and consumers value (Act. 1), utilizing genomics information and tools (Act. 2) to develop a sustainable technical platform (Act. 3) to accelerate and increase efficiency of cultivar development and adoption (Act. 4).

## **OUTREACH PLAN**

Extension activities involve delivery of socio-economics and genomics information to accelerate the adoption of Marker-Assisted Breeding (MAB) in both core and non-core Rosaceae crop breeding programs (Act. 5). Continuously updated content will be delivered via individual and small group tutelage by project participants, two distinct short course modules, regional technical workshops, and national participatory workshops. RosBREED has already joined the existing Plant Breeding and Genomics eXtension Community of Practice (CoP) and will utilize this web-based platform to exchange and disseminate information to ensure widespread availability and access, and thereby enable technology transfer and promote understanding.

## **POTENTIAL ECONOMIC, SOCIAL, AND ENVIRONMENTAL BENEFITS**

A steady stream of new Rosaceae fruit cultivars with superior fruit quality will be developed more efficiently. These regionally-adapted products will increase profitability and sustainability throughout the entire supply chain, as well as enhance economic viability of supporting communities. Nutritious fruit and fruit products with quality traits preferred by consumers will increase per capita consumption and thus contribute to improved human health and well-being.

## **STAKEHOLDER ENGAGEMENT THROUGHOUT THE PROJECT**

We will hold nine Participatory Workshops and four Regional Workshops to target producers/processors, market intermediaries, and industry stakeholder trade organizations to showcase MAB activities in the core breeding programs and include short course educational programming (Act. 5). The Socio-Economics Team will directly engage industry sectors and consumers to obtain specific input on fruit trait values through a range of survey tools and face-to-face meetings at major industry events (Act. 1). In addition, all members of the Stakeholder Advisory Panel (Appendix II) – who represent a broad range of cross-commodity professional interests in the system components of Production, Distribution and Processing, and Consumers and Markets – have committed to active engagement on behalf of their clientele through attendance at the four annual RosBREED meetings and regional workshops (Act. 5).

## (i) INTRODUCTION

**Vision Statement:** Integration of modern genomics tools with traditional breeding approaches will transform crop improvement in Rosaceae, significantly improving profitability and sustainability of U.S. rosaceous crop industries and contributing to increased consumption and enjoyment of these fruit, nut and floral products.

**Mission Statement:** We will create a dynamic, sustained program in research, infrastructure establishment, training, and outreach for developing and applying MAB based on improved knowledge of industry value and consumer preferences to accelerate and increase the efficiency of rosaceous cultivar release and successful cultivar adoption.

### OBJECTIVES

1. Increase the likelihood of new cultivar adoption, enlarge market potential, and increase consumption of rosaceous fruits by using socio-economics approaches to enhance knowledge of industry values and consumer preferences to inform breeding.
2. Establish sustainable technical infrastructure for an efficient MAB Pipeline in Rosaceae, including crop-specific SNP genome scan platforms for breeding-relevant germplasm exploiting the shared ancestry of Rosaceae crops.
3. Integrate breeding and genomics resources by establishing a user-friendly U.S.-wide standardized statistical framework and breeding information management system (BIMS).
4. Implement MAB in core RosBREED programs with a common focus on fruit quality traits.
5. Enhance sustainability of cultivar development by transferring MAB technologies to the public and private community of U.S. Rosaceae breeders through training current and future breeders, while better engaging the production, processing, and marketing sectors, allied scientists, and consumers.

### BACKGROUND

**Fresh and processed products derived from the Rosaceae plant family (almonds, apples, apricots, blackberries, peaches, pears, plums, sweet cherries, tart cherries, strawberries, raspberries, roses and other ornamentals) provide vital contributions to human nutrition, health, and well-being, and collectively their production constitutes the economic backbone of many rural communities in the U.S.** Although current domestic production value of these crops is over \$7 billion, and global per-capita production and consumption is expanding in both domestic and export markets, the U.S. rosaceous crop industries face numerous limitations to profitability and sustainability. Overcoming these limitations requires rapid development and deployment of new cultivars with improved characteristics that meet dynamic industry, market, and consumer preferences. The improvement of rosaceous cultivars by targeted application of genomics research in breeding programs will exploit the extraordinary diversity of Rosaceae species. Hence, moving “genomics to the marketplace” by accelerating and increasing efficiency of cultivar development to match consumer demand within constraints of the production and marketing system is the primary goal of the project, which addresses the legislatively mandated focus area: “Research in plant breeding, genetics and genomics to improve crop characteristics”.

**The RosBREED mission and objectives have evolved through a process of active engagement among stakeholders and researchers, along with development of the U.S. and international Rosaceae community.** Key outcomes include:

- a) 2002: Genome Database for Rosaceae (GDR) created (GDR 2009a).
- b) 2003: American Nursery and Landscape Association partners with USDA to create the Floriculture and Nursery Research Initiative (USDA-ARS FNRI 2009).
- c) 2003: National Tree Fruit Technology Roadmap established. Genomics, genetics, and breeding identified as a top priority (WTFRC 2009).
- d) 2005: National Berry Crop Initiative established (NBCI 2009).
- e) 2005: U.S. Rosaceae Genomics, Genetics and Breeding Executive Committee (RosEXEC) formed to provide a communication and coordination focal point for the U.S. Rosaceae genomics, genetics, and breeding community (US RosEXEC 2009a).
- f) 2006: U.S. Rosaceae Genomics, Genetics, and Breeding Initiative White Paper released (US RosEXEC 2009b).
- g) 2007: Rosaceae Specialty Crops Planning Workshop sponsored by USDA leads to first community-wide research proposal (Iezzoni et al. 2007).
- h) Regular Rosaceae Genomics International Conferences spur community development (Spain 2002; USA, 2004; New Zealand, 2006; Chile, 2008).
- i) Submission of RosBREED proposal in 2008 SCRI RFA enlists significant industry stakeholder interest and support, both of which are expanded in resubmitted 2009 proposal.

These outcomes all included significant involvement from key stakeholders, many of whom are included in this project's Stakeholder Advisory Panel (Appendix II, III, and VI). While these activities have increased awareness among stakeholder communities of the importance of genomics and genetics to plant breeding, there remains some confusion about genomic approaches and skepticism about the value and return on investment of MAB. This project acknowledges and explicitly addresses this issue by directly engaging stakeholders and their organizations in educational and participatory activities (Obj. 5). Significantly, this project recognizes and addresses the importance of using knowledge of what industry stakeholders and consumers value to inform breeding (Obj. 1). The convergence of increased breeding efficiency informed by consumer and industry values, collectively targets our long term outcome of "improved profitability and sustainability of U.S. rosaceous fruit, nut, and floral crops and increased consumption and enjoyment" (Appendix I, Logic Models, pg. 22).

**Rosaceae genetics and genomics information, and associated enabling technologies, are developing rapidly, leading to numerous discoveries with potential application.** In 2009, not only are genetic linkage maps available for all major crop species, but also the apple, peach and strawberry genomes are being sequenced, and hundreds of QTLs and major genes have been identified (Box 1 [full table at RosBREED 2009]). The vast majority of these QTL projects are being undertaken by RosBREED co-PDs and collaborators (Appendix I, Collaborators, pg. 15, and Linkage to Existing Programs and Projects, pg. 18). In addition, 2009 will be a watershed year for Rosaceae genomics with the upcoming publication of draft genome sequences of apple, peach and diploid strawberry, all involving RosBREED collaborators (Appendix I, Collaborators, pg. 15; Linkage to Existing Programs and Projects, pg. 18; Appendix V, Letters).

Despite these advances in Rosaceae genetics and genomics, MAB implementation has been minimal in U.S. Rosaceae breeding programs. Most rosaceous crops are clonally propagated for commercial production (often as a scion-rootstock composite plant) and have long generation times before the commercial product (fruit, nut, flower) can be phenotyped. For all, evaluation of individual plants requires intensive greenhouse and/or extensive field space; thus, phenotyping individual plants is expensive. Application of genetic markers is an obvious opportunity, but has not been implemented, due to several significant barriers: (1) lack of validated QTLs, (2)

### Box 1

Crop	Number of			Number used in		
	Traits	QTLs <sup>a</sup>	Major <sup>b</sup>	Genes <sup>c</sup>	MAPS <sup>d</sup>	MASS <sup>e</sup>
Apple	55	180	36	5	2	1
Pear	5	7	3	1	0	0
Peach	21	36	4	2	1	0
Almond	10	8	4	1	1	0
Cherry (sweet)	1	0	0	1	1	0
Cherry (tart)	3	14	0	1	1	0
Strawberry	11	31	2	0	0	2
Raspberry	0	0	0	0	0	0
Blackberry	0	0	0	0	0	0
Rose	11	41	9	0	0	0

<sup>a</sup> QTLs = quantitative trait loci with linked markers  
<sup>b</sup> Major = major trait loci with linked markers  
<sup>c</sup> Genes = known genes controlling a trait  
<sup>d</sup> MAPS = marker-assisted parent selection  
<sup>e</sup> MASS = marker-assisted seedling selection

QTL discovery performed in genetic backgrounds not relevant to breeder's germplasm, (3) lack of robust markers polymorphic in relevant germplasm, (4) undefined linkage drag with favorable QTL/gene alleles, (5) lack of knowledge on the mode of action of these QTLs (additivity, epistasis), (6) lack of knowledge on robustness of QTLs over environments, (7) lack of statistical tools and software for genetic studies in polyploid genomes, (8) lack of genotyping capacity, and (9) lack of breeding personnel trained in the application of MAB. **Bridging the gap between gene discovery and sustainable implementation of MAB in rosaceous crop breeding through establishment and implementation of a MAB pipeline is a focus of RosBREED (Obj. 4).**

RosBREED is rooted in our vision that the common ancestral origin of the diverse rosaceous genera can be harnessed to leverage knowledge and resources across commodity boundaries (Obj. 2). This project exploits homology among the genomes of three fruit-bearing genera of Rosaceae: *Malus* (apple), *Prunus* (peach and cherry), and *Fragaria* (strawberry) to develop practical outcomes. These three lineages collectively represent the majority of the rosaceous fruits produced and consumed in the U.S. RosBREED focuses on improvements in fruit quality because of its direct impact on producer profitability and consumer demand.

## (ii) RATIONALE AND SIGNIFICANCE

This project addresses SCRI Focus Area 1 (80%) as it will increase the efficiency of breeding to improve fruit quality in apple, peach, cherry and strawberry. Focus Area 2 is addressed (10%) as improved genetic resistance to disease threats, along with improved fruit quality, is represented in the plant populations selected for this project. Focus Area 3 is addressed (10%) as gains in long-term production efficiency, productivity, and profitability will result from more rapid and efficient development of new plant material improved for key crop characteristics. Despite successes achieved by plant breeders using classical approaches, tremendous opportunity exists to accelerate and expedite cultivar development through MAB, particularly for rosaceous crops, as breeding and testing cycles for

these perennial species require costly, labor-intensive management and are often long-term (e.g., tree fruit and nut species). This national and trans-disciplinary project seeks to provide a vertically-integrated set of practices (molecular marker enhancement, high-throughput genotyping, integrated data management, discovery and validation within breeding germplasm) across the breeding and supporting research community. This project will explicitly leverage the forthcoming genomic sequence information and translate it into application in breeding for increasing the effectiveness and efficiency of U.S. breeders to improve high-priority fruit quality traits (such as taste, texture, and appearance). Delivering premium quality fruit to the consumer is essential for the economic viability of the U.S. fruit industries and therefore is a key goal in the stakeholder-driven Technology Roadmap that defined research priorities to enhance tree fruit production through technological innovation (WTFRC 2009).

**Breeding:** The U.S. rosaceous breeding community consists of approximately 50 U.S. professionals each working on at least one Rosaceae crop (U.S. Rosaceae breeding programs, RosBREED 2009). All are breeding for improved fruit quality and other high impact traits (such as disease resistance and recurrent bearing) although with little formal understanding of the genetic control of their traits of interest. Breeding populations are mostly generated based on empirical experience of the combining abilities of a set of parents with inter-connected pedigrees. Breeders frequently use wild germplasm sources in crosses to obtain pest resistance and other characteristics (e.g., the introgression of resistance for apple scab; Crosby et al. 1992). However, in classical approaches with tree crops it typically takes 50 years from the initial cross to the release of the first resistant cultivars, and these only partly meet consumer preferences for eating quality. Such inefficiencies are due to incomplete purging of “wild” chromosomal segments imparting poor fruit quality. No disease resistant cultivar of peach, apple, or sweet or tart cherry has gained any substantial U.S. market share, as disease resistant cultivars to date lack sufficient fruit quality and handling characteristics. Breeders need a more efficient introgression system for rosaceous crops to allow them to capitalize on unique characters present in wild germplasm such as biotic resistances and unique fruit quality attributes.


RosBREED aims to enhance development of new breakthrough cultivars that have the potential to increase consumption and hence profitability (Box 2). The case of ‘Honeycrisp’ demonstrates that this is possible, an example of genetic improvement by selection for a key trait valued at a premium by consumers, prompting rapid

producer adoption and improved profitability. Recent molecular evidence points to a wild source of this cultivar’s exceptional crispness. Another potential example for achieving “breakthrough” fruit quality is the use of interspecific crosses between sweet and tart cherry to improve sweet cherry flavor (A. Iezzoni, pers. comm.). Finally, simply reducing the genetic predisposition of

**Box 2**

**Honeycrisp: a breakthrough cultivar**

- Honeycrisp apple - introduced in 1991 by Univ. of Minn.
- Required 30 years from crossing to commercialization.
- Dramatic attention and U.S. market share this decade.
- An *ultra-crisp juicy texture* and *pleasing flavor* have provided an eating experience that consumers crave, and for which they are willing to pay a premium. Growers, as well as consumers, reap the benefits, despite considerable challenges in growing the trees and storing the fruit.



In the past six years, more than 4 million Honeycrisp trees have been planted in major apple production regions across the U.S. An additional 1 million trees have been planted worldwide in Canada, Europe, South Africa, New Zealand, Chile, and Australia.

peaches to mealiness while increasing fruit sugar content holds promise to increase peach per capita consumption beyond its current levels.

RosBREED is built around 12 core breeding programs. The three core apple breeding programs (at Wash. State Univ., Cornell Univ., and Univ. of Minn.) collectively breed apples for all the major apple production areas in the U.S. The four core peach breeding programs (at Clemson Univ., Univ. of Calif. - Davis, Univ. of Arkansas, and Texas A&M Univ.) collectively breed peaches for the major peach growing regions in the U.S. The core sweet and tart cherry breeding programs (Wash. State Univ. and Mich. State Univ.) breed for the major cherry production regions in the U.S., the Pacific Northwest for sweet cherry and the Great Lakes region for tart cherry. The three core strawberry breeding programs (at Mich. State Univ., Univ. of New Hampshire, and USDA-ARS Corvallis) breed for strawberries adapted to all the northern, continental climates of North America. In addition, Driscoll's Strawberry Associates, the largest private strawberry breeding company in the U.S., will partner with RosBREED to provide a test at their California location (Letters from P. Stewart [Appendix VI] and J. Alcalá [Appendix III]).

The core breeders realize the potential of using genomics and genetic knowledge to reach their breeding goals more efficiently and hence are motivated to utilize the resources developed in this project. These 12 core breeders are already working together to develop crop-specific SNP Discovery Panels, and to identify pedigree linkages among germplasm and standardized phenotyping protocols (Crop reference sets in PediMap, RosBREED 2009). However, to implement MAB, breeders of Rosaceae fruit crops need: (1) understanding of the technology and implementation support from funding stakeholders; (2) robust, validated markers for high-impact traits; (3) the ability to obtain cost-efficient genotypic information, (4) efficient means to analyze and manage phenotypic and genotypic data to make selection decisions; and (5) empirically based knowledge of the relative importance of traits to guide and justify MAB implementation investments. Once trained, these breeders will become resources for successful MAB implementation, and the graduate students trained will become the future cadre of plant breeders, familiar with stakeholders and their priorities, as well as possessing skills to utilize genomics knowledge for marker development and application.

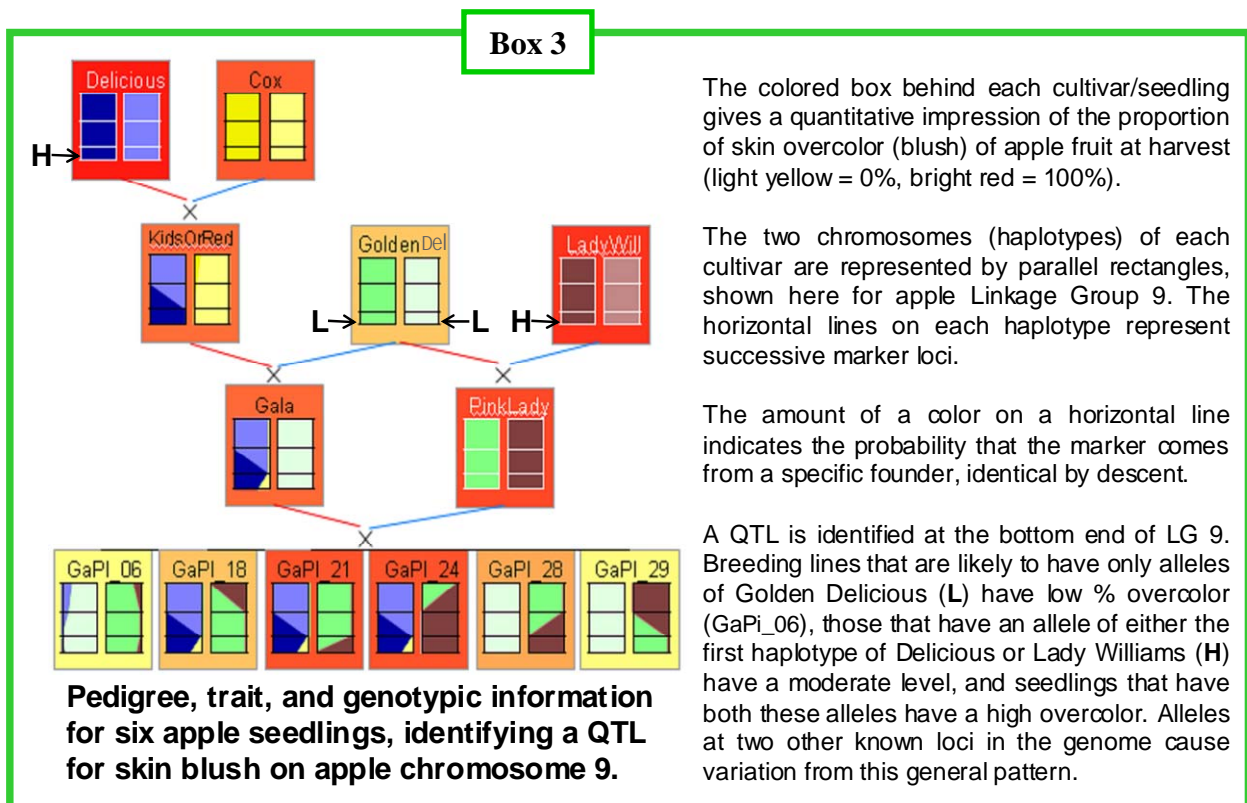
**Socio-economics:** Currently, Rosaceae breeders have little empirical basis on which to assign relative importance to various traits under selection. Breeding targets have been based largely upon a production-driven orientation whereby value has been considered primarily from the breeders' viewpoint – influenced by industry and market forces, but not transparently so. The decision-making process that guides breeding decisions will benefit from a science-based understanding of important stakeholder preferences, purchasing motives and attitudes, beliefs, concerns, constraints, and willingness to pay. Industry stakeholders would gain invaluable insights into values and preferences and become more meaningfully involved in “their” breeding programs. New cultivars would have targeted appeal (to both large-scale and small-scale niche market segments), be more quickly accepted, and have enhanced commercial impact.

Technology adoption in the presence of uncertainty is a dynamic process and a variety of factors are important at different times in the planning process to influence a decision to adopt (Besley and Case, 1993). It is important to understand and quantify these factors, which in fruit crops range from production level influences to the perceived final consumer demand for the product in the market, to accurately predict the probability of adoption. Breeders therefore need market-based information concerning impact of different traits on the entire supply chain to

assess the potential acceptance of specific traits and sets of traits by producers/processors, market intermediaries, and consumers. This information must acknowledge and involve different key stakeholders at early breeding decision-making stages.

**Pedigree-Based breeding information management systems:** Plant breeders collect and synthesize information on breeding selections and cultivars to maximize progress and expedite commercialization. Although this information-intensive process already requires robust information management systems, application of genomics technologies and MAB will increase the need for an efficient, accessible, and flexible software solution capable of integrating breeding and genomics data.

The success of QTL validation and functional allele characterization in rosaceous species will depend upon the flexibility and power of statistical analysis software. Unlike other crop species where recombinant inbred lines enable powerful experimental designs with statistical advantages for QTL studies, most plant populations used in rosaceous crops are segregating  $F_1$  populations from two highly heterozygous parents. Generation times of six or more years are not uncommon in species with protracted juvenility. Some Rosaceae species suffer inbreeding depression and many are obligate out-crossers due to self-incompatibility systems. The prevalent Rosaceae crop pedigree breeding approach utilizes a recurring set of ancestors within breeding programs, resulting in relatedness among individuals and extensive linkage disequilibrium. Therefore, a novel approach termed Pedigree-Based Analysis (PBA, Box 3) was developed based on the premise that for QTLs to be relevant for breeding, QTL validation and allele discovery must be performed in multiple breeding populations. PBA, as implemented in FlexQTL™ software, was developed by co-PDs Eric van de Weg, Marco Bink, and colleagues at Plant Research International, Netherlands (van de Weg et al. 2004; Bink et al. 2008).



The PBA approach exploits all data, including allelic diversity, in sets of multiple heterozygous parents and pedigreed family structure. Data from various-sized pedigreed populations comprising multiple generations, typical of rosaceous breeding populations, are tied together through common ancestry. PBA utilizes Identity by Descent to express the identity of an allele in a modern selection program in terms of alleles of founder cultivars. PBA informs breeding by permitting: (1) QTL discovery, validation, and allele mining in breeding populations, (2) estimations of the effects and values of QTL alleles in different genetic backgrounds, and (3) use of genomics information to design parental crossing schemes due to the ability to calculate probable QTL alleles of the parents and predict QTL performance of progeny from these parents. PBA is currently employed in corn breeding to improve populations prior to inbred selection (Dr. Bill Wilson, Pioneer Hybrid, pers. comm.). PBA was also successfully used in the EU apple project HiDRAS, in which the software was adapted to outcrossing species. PD Iezzoni is the first rosaceous crop breeder in the U.S. to adopt PBA and has demonstrated its practical value as a graphical allele mining and QTL software tool in sweet cherry (Iezzoni et al. 2008). Co-PD Peace has also introduced PBA for the Washington breeding programs of apple (co-PD Evans) and sweet cherry (co-PD Oraguzie).

**Crop-specific SNP genome scans:** Most of the genetic linkage maps available for rosaceous crops are based on SSR markers. Although these markers are useful, genotyping is expensive and time-consuming. With the forthcoming apple, peach, and strawberry sequences, the current large number of ESTs (425,780 at dbEST - NCBI 2009; DFCI 2009; GDR 2009b; Korban et al. 2009; NCBI 2009; Shulaev and Foltá 2009) and affordable DNA sequencing costs, it is now possible to develop genome-wide sets of SNP markers. However, most SNPs that can be mined from GenBank are from just a few representatives for each crop (e.g., ‘Goldrush’ and ‘Royal Gala’ in apple) or just one of the progenitor species (*F. vesca*) of the octoploid strawberry. Therefore SNP arrays based solely on these polymorphisms will have limited utility when applied to wider germplasm. Also, most SNPs that can be mined come from random genomic positions, and therefore it not possible to design an array that will have the full genome coverage required by PBA. To have practical value, SNP arrays must therefore be based on a wider range of breeding-relevant founders and on a genome-covering set of loci.

**Genomic information transfer among Rosaceae crops:** Knowledge of synteny among rice, corn and sorghum genomes accelerated the cross-genus discovery of corresponding seed size and grain shattering QTLs (Patterson et al. 1995). The forthcoming peach, apple, and diploid *Fragaria vesca* genome sequences will finally provide similar opportunities to leverage knowledge and resources across commodity boundaries and harness this knowledge for application in breeding. Current comparative mapping efforts have yielded promising results, with almost complete macro-synteny within *Prunus* crops (peach, almond, apricot, cherry), between apple and pear, and between strawberry and ornamental rose. There is also remarkable synteny between large chromosome segments of apple and *Prunus*, and between peach and strawberry (Dirlewanger et al. 2004; Arús et al. 2005; Arús and Gardiner 2007; Shulaev et al. 2008; Vilanova et al. 2008; Chen et al. 2008). Considerable micro-synteny also exists among rosaceous species (Davis et al. 2008; Vilanova et al. 2008; Sargent et al. 2009).

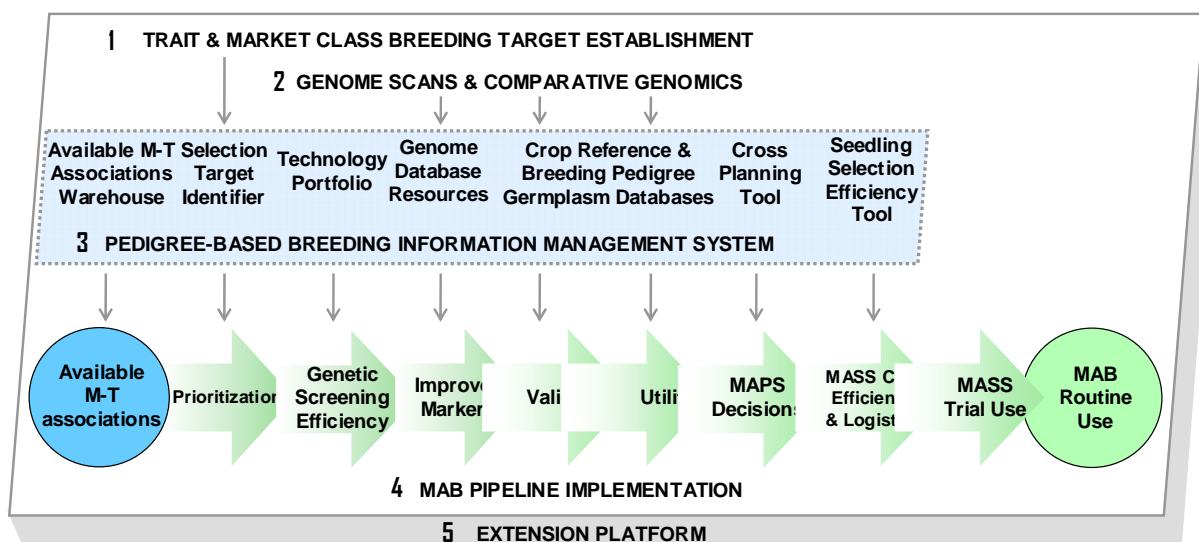
**MAB Pipeline establishment and implementation:** Implementing MAB in Rosaceae breeding programs requires establishment of a pipeline for connecting available knowledge of marker-trait

associations to application of MAB routinely. Besides science-driven stages, a successful pipeline must also include efficient, accessible, and affordable DNA extraction, genotyping, and data handling capabilities, and an economic and logistics assessment of using these technologies within routine breeding operations. Establishing and demonstrating each stage of the pipeline is critical to bridging the gap between the good intentions of genomics researchers in discovering new QTLs and gene-trait associations and breeders using this knowledge to inform breeding decisions.

**Extension:** The value of MAB for Rosaceae crop improvement is not currently obvious to the broad continuum of audiences targeted in this project: breeders, producers/processors, market intermediaries, and the consumer. It is critical that the project engage all four audiences, to engender support and encourage the use of outcomes of plant genomics research. Equally important, the project must establish a process and platform to sustain active communication flow among members of the entire supply chain.

### (iii) APPROACH

#### OVERVIEW OF ACTIVITIES



Establishing and implementing the marker-assisted breeding (MAB) Pipeline. Infrastructure that will be established for enabling routine MAB consists of a transparent process of prioritizing breeding selection targets that is based on values and preferences from industry sectors and consumers to determine economic weights for traits and market segments (1), development of state-of-the-art crop-specific SNP genome scanning capability for breeding-relevant germplasm that exploits genomic connections among Rosaceae crops (2), and creation of a sustainable data management system with statistical, database, and decision support functionalities (3). Using this infrastructure, an eight-stage MAB Pipeline will be implemented that results in routine MAB in Rosaceae breeding programs (4). An Extension platform engages stakeholders – Rosaceae breeders and allied scientists, the next generation of breeders, producers/processors, marketing groups, trade organizations, and consumers (5).

## ACTIVITY 1

### Trait and Market Segment Breeding Target Establishment

**Who:** Socio-Economics Team, Extension Team, Breeding Team

**Method:** Economic weights of traits will be determined from information on trait values from key market chain members. Information will be gained with a mix of survey research methods including one-on-one interviews and telephone surveys. Data will be collected for five crops: apple, peach, strawberry, tart cherry, and sweet cherry. Apple will be used as a model to test auction-based economic approaches and one-on-one interviews in which actual tasting of the fruit is required.

#### **1a. Breeders' current trait selection practices**

A survey will be conducted for U.S. breeders of project target crops to determine specific traits under selection, current relative weights placed on those traits, and traits for which they do not select due to limited perceived value, staffing, technology, or knowledge. Information on feasible selection targets from the perspective of breeders will be incorporated into subsequent surveys (1b-d). Surveys will be administered both in-person at RosBREED meetings and by phone to the U.S. population of breeders.

#### **1b. Producer/processors' preference and willingness to adopt new cultivars with specific fruit traits**

For producers/processors, a questionnaire will be administered to a randomly selected, stratified sample of producers/processors in the top-producing three states for each crop using a mixed mode (combination of mail and internet survey) method, backed up by in-person interviews. Sample size will depend on the degree of homogeneity and size of the producer/processor population. Survey development and administration will be collaborative with the Washington State Univ. Social and Economic Sciences Research Center (SESRC). Additionally, we will conduct in-person interviews with producers/processors at industry gatherings. The survey instrument will include questions on preferences for fruit traits (color, size, texture, etc.) and production traits (flowering, growth habit, annual bearing, postharvest drop, etc.). This information will provide baseline parameters to test the likelihood of choosing a new cultivar with a specific trait level or set of traits under alternative plausible scenarios. A model will be developed to measure the likelihood of adoption considering size of operation, diversification of products, credit availability, degree of risk aversion, synergies with other technologies, input and output price uncertainty, and neighborhood effects.

#### **1c. Market intermediaries' preference and values for fruit traits**

A survey similar to that of 1b will be developed and administered with the SESRC using a mixed mode methodology to market intermediaries, including suppliers, retailers, and wholesalers. Preferences and values for different fruit quality attributes, as well as marketing constraints associated with these attributes, will be collected along with size of operation, product sources, target markets, etc. A stratified random sampling procedure will be used, with a goal of 600 completed surveys. We will also conduct face-to-face-interviews with market intermediaries at industry events where fruit tasting and evaluation is conducted.

#### **1d. Consumer preference and willingness to pay for fruit traits and market segments**

Two major consumer-level research activities will be conducted: (1) a national survey of consumers, and (2) a focused non-hypothetical willingness-to-pay auction experiment. These two components will allow us to investigate how much consumers are willing to pay for fruit with different attributes (trait levels and trait combinations), assess consumer values of different

market segments, determine the degree of heterogeneity in fruit crop preferences, and analyze consumer attitudes and demographics that may be useful in explaining such heterogeneity. For (1), a survey will be developed to allow us to understand consumers' fruit selection and purchase decisions, relative importance of different attributes, and stated willingness to pay for attributes. This will be a combination of mail and internet survey methods based on a randomly selected national sample of consumers. Because consumers typically overestimate their willingness-to-pay in stated preference studies, we will complement the national survey with (2), a grocery-based auction experiment that will present randomly selected shoppers with real fruit, ask them to sample, and then bid on fruit with different attributes. Participants will pay real money for the product and be required to complete an abbreviated version of the national survey, allowing the assignment of value to attributes. These auctions will be performed in at least two participating states to capture potential regional differences in preferences. The combination of an auction (smaller population sample but without the hypothetical bias) and a survey (with a larger and more representative sample) provides complementary data and more robust results than either method used alone.

### **1e. Relative economic weights for fruit quality traits and production traits**

Information from 1b, 1c, and 1d will be used to develop and compare relative economic weights of fruit quality and production/processing traits posed by different market chain members for the five project crops. The relative economic weight for a trait will be calculated as a weighted average of marginal values that the three key groups (producer/processor, market intermediary, and consumer) place on the trait.

**Analysis of Results:** Preference and willingness to adopt or pay will be analyzed using various statistical techniques, including dynamic models, logistic regression models, mixed linear models, linear models, cluster analysis, and conjoint analysis. These models will enable identification of attributes preferred by different parties and the marginal value of each attribute.

**Use of Results:** Relative economic weights for each trait will provide objective information on overall importance of traits in their breeding programs and will replace or supplement weights currently used at the discretion of individual breeders. Results will be used as part of the infrastructure for 3c, to be practically applied in 4a and set up as a transparent objective process with parameters able to be readily adjusted with new information.

**Outreach Plan:** The target audiences for this activity are breeders, supporting researchers, and industry sectors, each of whom stand to benefit by more objective knowledge of the economic value of their selection, research, production, and marketing targets. Results from activities 1a-e will be presented at the RosBREED annual meetings and as a short course at RosBREED IV and the Regional Workshops of year 4 (Act. 5a). Reports and non-technical communication products will be disseminated to industry stakeholders through regional and national gatherings with content, style, and advocacy guided by Stakeholder Advisory Panel members. The eXtension platform will be utilized to provide dynamic web-based public access (5c). Evaluation instruments (short surveys and informal polling) will be administered to obtain feedback on socio-economics approaches, activities, results, and presentations.

**Expected Outcomes:** A transparent interactive process for determining economic values of production traits and fruit quality traits and setting breeding selection targets will be established. We expect Rosaceae breeders to routinely use (and update) economic weighting of production and fruit quality traits to optimize their set of selection targets. Rosaceae producer/processor, marketing group, trade organization, and public awareness of breeding, including efficiencies

gained by optimizing selection targets and strategies, will be increased by their participation in defining economic weights and through wide dissemination of findings.

**Potential Pitfalls:** Surveys of producers/processors, market intermediaries and consumers to be conducted are hypothetically based, that is, a person's statement has no immediate financial consequence, which can lead to bias in response. However, Lusk and Schroeder (2004) and Yue and Tong (2009) have shown that hypothetical surveys are incentive compatible and the bias is negligible when people evaluate real products rather than pictures of products. To lower possible bias, industry sector surveys will be supplemented with in-person interviews using real products; and consumer surveys will be supplemented with non-hypothetical experimental auctions.

**Limitations:** The array of attributes of Rosaceae fruit products is immense. We will only be able to assess five representative fruit crops (apple, peach, strawberry, sweet cherry, and tart cherry). Producer/processor, market intermediary and consumer preference and value for the same quality attribute of other crops may vary. However, the socio-economics tools and mechanisms to be developed can be readily applied to other fruit, nut, and similar specialty crops.

## ACTIVITY 2

### Genome Scans and Comparative Genomics

**Who:** Genomics Team

**Method:** Develop individual SNP-based genome-scanning capability for each of the crops of apple, peach, strawberry, and cherry, at a resolution of at least one polymorphic SNP marker every 5 cM in any random cross, for use in Pedigree-Based Analysis. The majority of genes targeted for SNP development will be those shared among species, and specific cases of synteny will be exploited to leverage ancestral relatedness at the genomic level.

#### **2a. SNP marker development**

SNP markers for genome scans will have two sources: validated SNP markers generated from existing projects and SNPs detected *de novo* in whole genome sequence using synteny and orthology information. These efforts will build on advances (Dirlewanger et al. 2004; Vilanova et al. 2008; Jung et al. 2009) involving researchers who are key personnel in this project. Existing validated SNPs will be identified by compiling those from published reports (e.g. Chagné et al. 2008; Han et al. 2009) and those generated within ongoing projects through collaborations with U.S. and international sequencing groups (Linkages [Appendix I, pg. 18] and Letters of Collaboration [Appendix V]). Additional gene-based SNPs and allelic variants of previous SNPs will be identified in breeding-relevant germplasm by genome sequencing of SNP Detection Panels. These Panels will consist of at least 28 individuals for each of the four crops, chosen by breeders with advice from the PBA Team to capture optimal diversity of founders (Act. 3g). SNP Detection Panel sequencing may use genomic DNA pools of Reduced Representation Libraries (Van Tassell et al. 2008). If validation of *de novo* SNPs is required, it will involve designing primers from genomic sequence, and PCR amplification in subsets of reference mapping populations (Letters of Collaboration from reference population owners: *Malus* (apple) - D. Chagné [Appendix V]; *Prunus* [peach and cherry] - P. Arús [Appendix III]; *Fragaria* (strawberry) - D. Sargent [Appendix III]). Population subsets will be chosen using the bin-mapping approach (Howad et al. 2005) that allow, when genotyped, genetic mapping of each SNP with a precision of  $\pm 2$  cM. SNP genotyping of the expected 20-30 bin-set offspring for each genus will then be performed with High Resolution Melting (HRM) analysis on a LightCycler 480, as described by Chagné et al. (2008). Such wet lab SNP validation will not be conducted if

a large number of individuals are included in the SNP Detection Panels (including the 20-30 bin-set individuals to allow genetic mapping). Following compilation of known SNPs and *de novo* validated SNPs, a final genome-spanning set of SNPs will be chosen for each crop. To capture existing haplotypes in breeding germplasm, we will usually represent each gene with several SNPs.

### **2b. SNP genotyping platform**

A SNP Platform Summit will be held in Washington in year 1, coordinated by the Genomics and Extension Teams. This Summit will allow us to select gene-marker sets and consider and compare state-of-the-art SNP genotyping platforms for choice of an efficient technology and service provider to conduct genome scans within RosBREED. Worldwide genomics experts in the Rosaceae community and beyond will be hosted together with the Genomics Team and the RosBREED Executive Committee for a three-day event (Letters of Collaboration from many Collaborators expected to attend [Appendix V]). However, for budgeting purposes we have considered available SNP genotyping platforms, and the 1536-SNP platform using the Illumina Golden Gate Assay appears cost-efficient. Once finalized at the Summit, the final list of gene-based SNPs covering the genome for each of the four crop groups will be submitted to the service provider (currently Illumina, for oligo design of 1536-SNP platforms using the GoldenGate Assay). If this platform is ultimately chosen, genotyping of breeding-relevant germplasm of the four crops (Act. 3g-h) will be conducted on the Illumina Bead Array platform (Letter from Cornell University's Sequencing and Genotyping Lab is in the MSU Budget Justification).

### **2c. Comparative genomics in Rosaceae**

Genes chosen for SNP marker development will include a majority that enhance our ability to conduct trans-Rosaceae genetic information transfer, as they will tend to be those that are orthologous, single copy, and encompass full genome coverage for each crop. The genomes of *Malus*, *Prunus*, and *Fragaria* will be aligned and displayed using Sybil (Crabtree et al. 2007). SNPs, synteny, and orthology data will be submitted to GDR (GDR 2009a) and made available for view on GBrowse. Specific opportunities for identifying new genes controlling fruit quality traits will be exploited, such as by studying co-locations of QTLs previously reported for any of the crops with candidate fruit quality genes in syntenic genomic regions of other crops.

**Analysis of Results:** Allelic diversity results from genome sequencing of SNP Detection Panels of breeding germplasm from each crop will be analyzed using the SNP detection pipeline available in the Rosaceae GDR bioinformatics laboratory of co-PD Main to identify existing SNP haplotypes in breeding germplasm. Criteria for SNP choice prior to validation will be: gene-based (required), orthologous gene present in multiple Rosaceae crops (desired; such as COS markers, up to 400 of which are currently being mapped in *Malus*, *Prunus*, and *Fragaria* genomes in a USDA-NRI project of PD Iezzoni), and gene putatively involved in fruit quality (desired). The wet lab SNP validation step will be used to identify true SNPs and the genetic locations of new markers on the internationally recognized reference map genomes for *Malus*, *Prunus*, and *Fragaria*. Physical map locations will also be determined by locating SNP-containing genes of interest in the assembled apple, peach, and strawberry genomes expected to be released in 2009 (Letters of Collaboration from worldwide whole genome sequencing efforts are in Appendix V; apple - R. Velasco, A. Dhingra; peach - B. Sosinski; strawberry - V. Shulaev). Criteria for validated SNPs to include in genome scans will be: spanning the genome with at least one SNP per 1-2 cM to achieve the target of at least one polymorphic SNP per 5 cM in any random cross within the crop (required), simple interpretation of genotype (required),

orthologous gene present in multiple Rosaceae crops, especially those known to map to syntenic regions (desired), and gene involved in fruit quality (desired). To ensure that at least one segregating gene-based marker is available every 5 cM, based on the known length of the diploid genomes of *Malus*, *Prunus*, and *Fragaria*, we will choose at least 1200, 500, and 500 genes, respectively for these genera, spaced across the genome, which provides one gene every cM. Final choice of markers for the SNP genotyping platforms will be determined in collaboration and coordination with U.S. and non-U.S. partners during the SNP Platform Summit scheduled in spring of 2010.

**Use of Results:** The full set of gene-based SNPs and their associated allelic variants will provide the markers for SNP genome scans of each crop (Act. 3g-h) and public genomics resources (3f). Significant gene-trait associations identified by exploiting synteny among Rosaceae species will enter the MAB Pipeline as “available marker-trait associations” (4a).

**Outreach Plans:** Besides the internal RosBREED BIMS, PBA, and MAB Pipeline Teams who need to efficiently convert the results of this activity into MAB resources that can be readily utilized by breeders (Act. 3 and 4), the target audience is the U.S. and international Rosaceae genomics research community, who outside of RosBREED are expected to use the results to expand identification of new marker-trait associations to fill the warehouse (3c). Results of SNP marker development, including their genome locations, will be presented at the RosBREED II meeting (5a) and data will be submitted to the GDR for display and searching by the international Rosaceae community. The SNP markers for apple, peach, strawberry, and sweet and tart cherry are major sustainable deliverables from this project; therefore, we will ensure open access to this information in the GDR.

**Expected Outcomes:** Based on prior experience, we expect at least half the genes to have multiple haplotypes, thereby providing more than enough SNPs required for genome-spanning resolution. The SNPs developed for targeted genes will enable development of genome-wide SNP genotyping platforms for apple, peach, strawberry, and cherry. The specific choice of genes will allow leveraging of genetic information across Rosaceae genera. Several cases of useful new gene-trait associations are expected to be uncovered, to provide examples of tangible benefits that can be reaped by considering the Rosaceae genomics, genetics, and breeding community as a whole rather than crop- or region-centric groups.

**Potential Pitfalls:** SNP Detection Panels of 28 individuals may not be large enough to cover an adequate proportion of allelic variants present in the breeding programs; further individuals will be sequenced as technological advances allow. It is not known if constituting genomes of tart cherry (tetraploid) or cultivated strawberry (octoploid) are co-linear or if detected polymorphisms in these crops will represent true SNPs or homoeologous sequence variants. Therefore, for these two crops, genotyping of the Crop Reference Set (Act. 3g) will include an existing mapping population to allow linkage mapping that differentiates the constituting ancestral genomes. By our calculations, a 1536-SNP genome scan is generous for our required genome coverage of peach and sweet cherry, suitable for apple and tart cherry, but too low for octoploid strawberry. Therefore, we seek to be flexible within the project to use the most cost-effective available technologies that allow increases in SNP numbers.

**Limitations:** Synteny among genera in some regions of the genome might be complicated and irresolvable at our level of saturation. Genetic control of similar fruit quality traits in different crops may be influenced by genetic variation in different genes in many cases, rendering comparative genomics approaches ineffective. However, the extent to which comparative genomics successfully identifies functional alleles across Rosaceae species and genera made

possible by this project will be instructive on the extent to which genomics knowledge gained in one crop is useful in another, and thus reveal the extent to which Rosaceae genomics researchers are truly connected by the shared ancestry of their target species.

### **ACTIVITY 3**

#### **Pedigree-Based Breeding Information Management System**

**Who:** BIMS Team, PBA Team, Breeding Team, Genomics team, MAB Pipeline Team, Extension Team

**Method:** A common BIMS aligned with the PBA approach will be established for streamlined collection, archiving, analysis, and interpretation of integrated breeding and genomics data. Statistical programs, component databases, and decision-support tools will be incorporated into this broader BIMS framework, which will be hosted on the GDR (GDR 2009a). Core participating breeding programs will be the alpha-testers for BIMS, ensuring practicality for the wider U.S. Rosaceae breeding community to which BIMS will be delivered. A step-wise process will introduce breeders to this data management system that will ultimately enable programs to leverage genomics advances for improved breeding efficiency.

#### **3a. BIMS framework**

The BIMS aligned with the PBA approach (3b) will enable integration of genomics data into breeding decisions. BIMS will have inbuilt analytical capabilities, online community-driven database interfacing, and decision-support tools for modern breeding needs. A BIMS Developer together with a BIMS Programmer (postdocs to be hired) will establish the computer-based framework for BIMS. In year 1, the BIMS Developer will visit all 12 core participating breeding programs to assess their current information management needs, desires, capabilities, and approaches. This process will be continued throughout the project with expansion in years 3 and 4 to include other U.S. Rosaceae breeding programs. The BIMS Developer will work with the BIMS Programmer to develop a common database template that efficiently addresses traditional breeding data management needs, and allows breeding programs to readily interface their operations with PBA and other infrastructure components of the MAB Pipeline. These individual components are described in 3b-j below, with 3b representing the statistical framework and 3c-j aligned with the first seven Pipeline stages. The broader BIMS framework will incorporate all of these components to provide seamless transitions between MAB Pipeline stages that interface with traditional breeding data and operations.

BIMS is analogous to Laboratory Information Management Systems now used throughout the scientific community. BIMS will include typical breeding data analytical capabilities and will be inter-operable with the GDR, to leverage the extensive publicly available genomics data and tools housed there. Database platforms already implemented in plant and animal breeding will be evaluated based on suitability criteria (such as cost, flexibility, user friendliness, and inter-operability with existing Rosaceae databases). A customized platform will be built for Rosaceae, and wherever necessary and possible, existing software (preferably open source) will be adapted which may include contracting or consulting with software engineering firms. Based on the general template schema, breeding program-individualized databases will be created for day-to-day operations to house breeding data such as pedigrees, statistics, genetic correlations, phenotypes, genotypes, QTLs, markers, images, and experiments, and be inter-operable with web interfaces and mobile devices (for field data).

BIMS is not intended solely for use of markers in breeding. The Selection Target Identifier (3d) will provide an objective means of prioritizing selection targets. The Cross Planner (3i) and Seedling Selection Efficiency Tool (3j) will be generalized devices for determining the best parents to cross and the best seedling selection strategies to efficiently achieve breeding targets. Exemplifying the RosBREED approach, these decision-support tools will be able to integrate genetic marker information to allow breeders to compare traditional selection targets and selection strategies with new MAB opportunities provided by genomics advances, and choose the most appropriate balance of approaches.

### **3b. PBA establishment**

Pedigree-Based Analysis is an existing statistical framework with associated software designed to identify, validate, and use QTL information from pedigree-linked individuals to inform breeding decision-making. Adoption of the PBA approach and establishment of U.S.-based PBA expertise will occur through training of and consultancy to Breeding Teams throughout the duration of the project. Breeders and supporting personnel will be professionally trained in the use of PBA at RosBREED annual meetings, within workshops spanning multiple days, and coordinated by the PBA Team. Documentation material (such as syllabi, PowerPoint presentations, and Frequently Asked Questions) will be developed to support the training and facilitate independence in use of software at breeding locations, and incorporated into BIMS and hosted in a central repository on the project website. The PBA Team will also be available for support to RosBREED participants in software use and actual PBA statistical analyses of RosBREED data throughout the project. A Community of Practice (Act. 5c) will be established to allow consultancy and/or email discussion groups.

In the first year of the project, current PBA software (FlexQTL™ and PediMap) will be provided for each core breeding program. At the RosBREED I meeting, co-PDs van de Weg and Bink will engage the breeders in discussions on development of optimum germplasm sets (for activities 3g-h), type and format of phenotypic and genotypic data required, efficient data storage, introduction to prospects and limits of association studies, and understanding of genetic map uses and linkage mapping procedures (how and why maps are built and pitfalls). At two-day PBA workshops in RosBREED II and III, co-PDs van de Weg and Bink will familiarize breeders with PBA and software options including actual analysis of data generated from the CR Sets in year 1 (4d) and year 2 (3g, 4d) and BP Sets in years 1 and 2 (4e) and early year 3 (3h). Annual meetings will also be used to survey and monitor needs and opportunities of U.S. Rosaceae breeding programs, with reports disseminated soon after each meeting and discussed through telephone conferences. Software training for breeders during the project will enable the breeders to incorporate additional pedigree, phenotypic, and genotypic data after the term of the project to refine QTL analyses and facilitate sustainability of MAB programs.

Beginning in year 1, the Dutch members of the PBA Team will survey specific needs and opportunities of U.S. Rosaceae breeding programs to further optimize PBA with regard to enhanced statistical functionality and user-friendliness of PBA software. During the first two years, the PBA team, comprising quantitative geneticists, statisticians, and software developers, will develop new software modules based on the GenStat program (VSN International 2009) that are inter-operable with FlexQTL™ and PediMap, pre-tested within individual crops and optimized as necessary. Such enhanced statistical functionality, for which protocols and prototype software will be developed depending on priorities of the U.S. Rosaceae breeding community, will address biological and genetic complexities (e.g. polyploidy, inter-locus interactions, and genotype x environment interactions). Modules will be used for professional

training during RosBREED II and III meetings (Act. 5c). Feedback from breeders will be used to further improve capabilities of PBA software.

### **3c. Marker-trait association warehouse**

A one-stop community-driven website for accessing reported marker-trait (M-T) associations for Rosaceae breeders will be hosted on the GDR. This database will contain up-to-date information on M-T associations in a format developed specifically for breeders that allows relevance assessment of each association. Data will include crop, trait (ontology-based), specific germplasm of discovery, method and location of phenotypic data collection, genomic location of trait locus, type of marker associated with the trait, effect on trait of each known allele/haplotype, genotypic variation observed, publication or information source, and contact details. Genome-viewing tools already exist on GDR (such as CMap for comparing multiple marker maps) and will be incorporated in this BIMS module to identify linked M-T associations and to identify similar trait loci or candidate genes in syntenic regions of other crops. The warehouse will first be established by collating information on reported or otherwise communicated (“available”) M-T associations, as well as the hundreds of associations expected to be identified in activities 3g-h. The warehouse will have Wiki-like public access for sustained community updates and refinement with newly discovered M-T associations, with moderator privileges for the Genomics Team.

### **3d. Selection target identifier**

An online community-driven database of economically weighted market classes, fruit quality traits, and other high-impact traits will be established for apple, peach, strawberry, sweet cherry, and tart cherry. Economically weighted breeding values (EWBVs) will be used to prioritize available M-T associations for any breeding program (Act. 4a) and thus provide breeding targets that are objectively defined and quantified (rather than simply ranked). For each trait, heritability values and available genetic variation (from published reports, breeders’ knowledge, and analyses of phenotypic data and QTLs in Act. 3g-h) will be compiled. Objective economic values of traits and market segments will be obtained from Act. 1e. The database will be initially populated with traits or market classes given high priority by breeders for each crop (from Act. 1a); such information has already been collected and compiled from the core breeders based on current breeding objectives and phenotyping interests (see Phenotyping tables, RosBREED 2009).

For each crop, EWBVs will be assigned to traits and market classes by multiplying marginal economic value by heritability and by available genetic variation. The first parameter may involve non-linear relationships between dollar value and unit increase in the trait. The second parameter will be dependent on conditions and germplasm in which estimates were obtained, and may have large margins of error. The third parameter will depend on germplasm available to breeders and will be cross-specific. Nevertheless, EWBVs will provide a more objective means of determining breeding selection targets and prioritizing M-T associations than existing methods that do not quantify these component parameters.

Once established, the database will be incorporated into the BIMS and will be modifiable by individual users as new traits and genetic variation are recognized or relative economic weights change due to shifting production or marketplace factors.

### **3e. Technology portfolio**

An online community-driven website of available technologies and service centers for genetic screening (sampling, DNA extraction, genotyping, and data provision) will be compiled to assist breeding programs in selecting appropriate technologies. Comparisons to be made among

technologies will include capital investment costs, equipment availability, operational logistics, cost-efficiency, quality, running costs (per data point), multiplexing, flexibility, reliability, technical difficulty, and applicability to the scale of Rosaceae breeding programs and to multiple breeding programs. Feasibility of applying robotics at a central DNA extraction center will be assessed.

Some breeding programs may not choose to make capital investments or develop in-house expertise. Those programs will be able to utilize service centers (public or those established in this project) identified through this grant to perform their routine genetic screening.

### **3f. Genome database resources**

To facilitate development of improved markers for high priority M-T associations, BIMS interfaces will be developed that leverage publicly available genome resources for the Rosaceae family on the GDR. The Locus Pages approach developed for SolCAP (SGN 2009) will be explored. The SNP and comparative genomics data from Act. 2 to be submitted to the GDR are expected to provide additional resources for developing improved markers for valuable loci. This will complement worldwide Rosaceae genomics knowledge already compiled on the GDR, and new data being added from other projects involving RosBREED participants and collaborators including whole genome sequences of apple, peach, and strawberry.

### **3g. Crop-wide breeding germplasm reference databases**

A reference database of phenotypic, genotypic, and pedigree data - both raw and integrated via PBA - will be established on the GDR for the large germplasm sets spanning U.S. breeding programs for apple, peach, strawberry, and cherry. This database will be a powerful public resource for marker-trait association discovery and validation in Rosaceae, with genome-wide SNP and high-impact trait data collected and analyzed for 480 individuals of each crop.

The first step for each crop will involve choosing pedigree-linked germplasm that represents the diversity in current and anticipated future breeding stock - the Crop Reference (CR) Sets. Establishment of CR Sets has already begun (Crop reference sets in PediMap, RosBREED 2009). This task was accomplished by the Breeding Teams for the different crops choosing germplasm that represents common cultivars used as parents in U.S. breeding programs, their ancestors and founders, new parental materials including representatives of wild species, and F<sub>1</sub> populations known to segregate for fruit quality traits and linked to other cultivars by common ancestry. Breeders provided pedigree information to PD Iezzoni for entry into the PediMap program provided by co-PD van de Weg. The PBA Team will assist breeders in choosing plant materials to maximize detection of true marker-trait associations with PBA, to arrive at a total of 480 individuals in each crop. PediMap, a free software component of PBA, is used to graphically present pedigree relationships. Detection of common founders with PediMap allows the Breeding Team to choose founders of each crop for the SNP Detection Panels of Act. 2a.

CR Sets will be phenotyped for fruit quality and other high-impact traits over two successive seasons (years 1-2), with a third year for those traits requiring it. An initial set of target traits was identified by the breeders for each crop (Phenotyping tables, RosBREED 2009) for phenotyping at each location. Standardized protocols (such as those developed within international Rosaceae initiatives; Peace and Norelli 2009; Volk and Bassil 2009) will be used for maximum congruence of data across multiple locations, and where possible, across crops. Existing phenotypic data from regular breeding activities will be compiled, and wherever necessary, additional data will be collected for important fruit quality attributes: flavor

components (sweetness, acidity, aroma, etc.), texture components (firmness, crispness, juiciness, etc.), color, size, shape, storability, and fruit storage disorders and possibly others (e.g. productivity, blooming and ripening dates, plant architecture, and disease susceptibility). Results from activity 1 may suggest additional target traits as the project proceeds. As most CR Sets are spread over multiple sites, breeders will obtain data on partial CR Set plants at their location. The entire strawberry CR Sets will be replicated in Michigan, Oregon, and California in collaboration with Driscoll's Strawberry Associates, Inc (Letter from P. Stewart, Appendix VI). Year and location effects on phenotypic data will be accounted for by ANOVA and other approaches including PBA GenStat modules once they are developed (3b).

In year 2, CR Sets will be genotyped with the genome-wide SNP markers developed in activity 2 and the platform chosen at the SNP Summit, with DNA extracted in activity 4d. This is expected to provide at least 1536 genotypic data points (as of August 2009, we are planning tens of thousands, coordinating with additional international partners) for each of the 480 individuals of the CR Sets for each of the four crops. If the GoldenGate platform is ultimately chosen for SNP genotyping, Illumina's GenomeStudio software will be used to score SNP genotypes for the 480 individuals of each CR Set. As the cherry and strawberry CR Sets will include mapping populations, SNP genotypic data for autotetraploid tart cherry and octoploid (amphidiploid) strawberry will be used for linkage map construction using JoinMap<sup>TM</sup> 4.0 (van Ooijen 2006). These new markers will first be mapped on existing U.S. reference maps (e.g. Weebadde et al. 2008; Spigler et al. 2008) to assess the homoeologous chromosome pairs on which these SNP markers are located.

All raw and analyzed pedigree, phenotypic, and genotypic data will be permanently archived in the GDR, the community database for all Rosaceae genomic and genetic data. Individual BIMS-enabled breeding programs will be able to download this data for their own use, and will be electronically notified on updates to the public archive. With pedigrees, phenotypes, and genome-wide genotypes in place, PBA software established in 3b will be used to integrate the data. Across this breeding germplasm of relevance to any breeding program including non-core programs, PBA statistical analyses will verify pedigrees, identify M-T associations, validate previously reported associations including estimating phenotypic effects of individual QTL alleles, characterize functional allelic diversity present, and calculate the probability of functional alleles present in each individual. The PBA team will support breeding programs in the performance of these analyses. New functionalities and modules (such as investigations of GxE and epistatic interactions) will be tested on RosBREED data as soon as they become available. PBA results for the CR Sets will be archived and made publicly available for reference on the GDR.

The expected value of this reference information is immense. Functional genotype-based breeding values for each CR Set individual will be described. Hundreds of M-T associations will be dissected to resolutions of less than 5 cM intervals, and in many cases the actual genes likely to be underlying the QTL will be apparent. These M-T associations will greatly expand the warehouse of 3c. Further phenotyping of CR Set germplasm beyond direct RosBREED funding support can be added to the data archive for immediate M-T association analyses with PBA. The resolution offered and saturation with fruit-quality gene SNPs is expected to accelerate functional marker identification, including refinement of the hundreds of previously reported fruit quality M-T associations in Rosaceae.

### **3h. Program-specific breeding germplasm reference databases**

PBA will be extended to enable utility testing of M-T associations in germplasm and conditions specific to individual programs, building on results from 3g. Breeders will be able to add their own germplasm and phenotyping interests to the CR Set data by collecting data and performing PBA on their own Breeding Pedigree (BP) Sets. This data will always remain the property of individual breeding programs, and will not be publicly available even to other RosBREED breeding programs. Germplasm for each breeding program of ~200-250 individuals including parent cultivars, founders and pedigree-linked populations that well represent the program will be assembled as advised by the PBA Team. BP Sets and CR Sets will substantially overlap for some programs, saving resources in data collection. Pedigree information will be entered into the generic data template developed in BIMS, and visualized with PediMap. Existing phenotypic data sets for each BP Set will be compiled for target traits and, where necessary, additional data will be collected primarily in years 2-3 (and year 1 if necessary) for target traits using standardized protocols that are immediately relevant to breeding selection criteria. Genome-wide SNP genotypic data for BP Sets will also be collected, using DNA extracted in activity 4e. PBA will be used to verify pedigrees of breeding program-specific germplasm and determine whether M-T associations identified in the CR Sets have utility within specific breeding programs, and in which specific germplasm and crosses they can be exploited.

Breeding Pedigree reference databases are not restricted to 200-250 individuals. Programs may, at their own cost, increase the number of individuals and the range and depth of traits measured, improving the power of the reference database for their own use. Once data is gathered it will be a permanent resource for the breeding program, and the flexible design of PBA and additional power obtained with more data encourages the inclusion of all breeding data for all germplasm - past, present, and future. Once the use of PBA software is established for a program (in 3b), the only requirement for expansion will be that data is entered into the simple PBA format (spreadsheet of germplasm in rows; parentage, traits, and markers in columns).

### **3i. Cross Planner**

A decision-support tool for planning crosses will be developed for breeders to design efficient single and multi-generation crosses for combining specific traits and alleles based on expected seedling outcomes under various selection protocols. The tool will estimate optimal parent combinations and family sizes to achieve breeding goals. The BIMS Team will initially consult Rosaceae breeders in person to identify their current strategies and decision needs in choosing parents. Relevant existing information that supports these decisions will be gathered from literature, experience, and results of CR Set and BP Set characterization (of 3g-h), including components of breeding value (heritabilities of important traits, known pedigrees, functional marker alleles for important traits, combining ability of parent in previous crosses) and intellectual property costs. The Cross Planner will process these data using a set of interactive computer program modules that include a heritability calculator, pedigree viewer and verifier/deducer, inbreeding coefficient calculator, genetic relationship calculator and viewer, and chromosome graphical viewer. Specific modules will use existing programs (especially Open Source) as much as possible and tailor them to the idiosyncrasies of Rosaceae crops.

Breeders and breeding trainees will test the Cross Planner using empirical breeding program data. Breeders will examine concordance of recommendations produce by the selection tool with their pre-existing decision-making process to identify new opportunities for improving parent selection efficiency. The computer program developed is expected to increase the efficiency of parent selection and cross-planning decisions by Rosaceae breeders by decreasing subjectivity in the crossing decision process.

### **3j. Seedling selection efficiency tool**

A decision-support tool for identifying optimal marker-assisted seedling selection (MASS) schemes within individual breeding programs, based on economic and logistical analyses, will be developed for generic use in specialty crop breeding programs. An initial approach to such a tool for fruit breeding programs was described by Luby and Shaw (2001). A more detailed approach was developed at Wash. State University (WSU), using the WSU apple breeding program as the test case (C. Peace, J. Luby, et al., manuscript in preparation).

Input parameters for each breeding program will include: a timeline and costs of each stage of the breeding program from crossing to selecting elite individuals, phenotypic cull levels, heritability of phenotyping (i.e., reliability of phenotyping), cash flow of the breeding program, numbers of individuals carried through each stage, number of genetic tests available, details of each genetic test (number of markers required for a test, its expected cull level – adjustable for multiple populations, and its reliability), costs for sampling, DNA extraction, and genotyping, and the minimum number/proportion of seedlings a breeder is comfortable with field-planting following genetic screening. Considerations for genotyping will include combined vs. sequential genotyping-selection, multiplexing, depreciation of equipment, and use of reliability factors to account for failed extractions and failed genotyping, in addition to the reliability of markers based on linkage distance. Genetic tests to be used are those for which the breeder is willing to make a cull decision and those where use of the genotypic information is less well-defined (i.e., more for descriptive purposes). Where results of multiple markers are combined to allow a cull decision, markers will form a single “genetic test”, e.g., when two markers flank a QTL, or where the breeder is interested in a selection index based on markers for multiple traits.

Calculations are based on simple mathematical formulas within spreadsheet programs such as Microsoft Excel, or other appropriate software. The marginal gain in cost-efficiency will be determined for each genetic test, and cost savings of MASS shown for each possible stage in the breeding cycle. Graphs will display results of sensitivity analyses of input parameters to determine robustness of the decision process and comparisons of alternative scenarios within breeding programs will be displayed side-by-side. An option will allow for cost savings of MASS to be reinvested in increasing initial seed/seedling numbers (up to a specified physical limit), with an iterative process to determine the optimum MASS scheme. Another option will enable genetic screening to be spread over multiple stages where the most cost-efficient stage is too short to process all individuals.

Each participating breeder and associated breeding trainee will determine the most efficient implementation of MASS in specific breeding programs using available validated and utile markers. Recommendations will be made on whether to use each available genetic test (based on the marginal savings in their use), how (order, combination) and when (best stage in breeding cycle) to apply them. The tool will enable adjustment of MASS application based on real changes in input parameters and availability of validated markers for traits of interest. It will also be used in reverse, to determine which traits under selection would most benefit from the availability of genetic tests, and parameter thresholds of those tests to attain cost and logistical efficiency. While costs and savings will be based on net present value, where costs for MASS are projected to exceed available funds, the tool will estimate external investment needed and provide breeders with a supportable basis to request funding from institutional, industry, or other sources. Comparisons across breeding programs will be conducted empirically to determine general conditions for efficient implementation of MASS in Rosaceae breeding programs to

make general recommendations. The tool can be expanded beyond MASS, for decision-support of all types of selection.

**Analysis of Results:** Functionality of components and their transitions comprising the BIMS infrastructure of the MAB Pipeline will be assessed by alpha and beta testing (as described in the Outreach Plan below). At a minimum, the BIMS framework developed must allow breeders to use their own data in PBA, and therefore the primary criterion for BIMS evaluation is breeder utility of FlexQTL™, PediMap, and GenStat modules. However, the expectation by the release of BIMS v3.0 in early year 4 is for the series of components to seamlessly interface with each other, with existing and future breeding program data, and with other genomics application modules of the GDR developed outside of RosBREED.

**Use of Results:** The general framework of BIMS will be adopted and used by all participating breeding programs and further interested breeders during the project and beyond. Individual components of BIMS will be used for MAB Pipeline implementation in activity 4. BIMS will provide an interface with the online resources of 3c-h, and house the decision-support tools of 3i-j. All databases are expected to become sustainable, community-driven entities by the end of year 2, maintained on the GDR at the end of the grant (the GDR user community is seeking permanent funding opportunities for the GDR), and supported by the RosBREED CoP (Act. 5c) The decision-support tools will be refined by use with empirical data to identify additional breeder needs, and iterative improvements will be included in subsequent BIMS versions.

As described in 3g, the comprehensive PBA results from integrating pedigree, phenotypic, and genotypic data to characterize functional alleles in breeding germplasm (particularly for fruit quality traits) will enter the MAB Pipeline through the M-T association warehouse of 3c. SNP genotyping of the CR Sets will generate a reference database of >2.8M breeding-relevant genotypic data points (if only 1536 SNPs per crop), and BP Set genotyping will double the total. Standardized phenotyping for a dozen or more high-impact traits over 2-3 years will provide >100K breeding-relevant phenotypic data points. Use of PBA to integrate this data will be the ultimate key to industry-changing advances.

**Outreach Plan:** Our target audiences for this activity are 1) core breeders who will be alpha-testers for the whole system for which we expect to create success stories of MAB, and who must be trained in MAB Pipeline implementation, and 2) further U.S. Rosaceae breeders to whom technology transfer must occur to achieve the RosBREED Vision. A critical need of RosBREED will be to demonstrate to core breeders, and subsequently to further Rosaceae breeding programs, clear advantages to using the new system (of marker-assisted breeding, the BIMS database schema, and the Pedigree-Based Analysis approach) over what each potential user is already comfortable with. We will address that with a stepwise approach to gradually introduce breeders to BIMS functions and MAB capability. In summary, these steps are 1) use of BIMS modules that address traditional breeding needs, 2) addition of marker information to guide parent and cross selection, 3) use of marker information for improving seedling selection efficiency, 4) breeder implementation of earlier Pipeline stages involving increasing confidence with genomics interfacing, and 5) routine MAB using the whole Pipeline.

The first version of BIMS (v1.0) will include the Cross Planner and the Seedling Selection Efficiency Tool (at least in their initial functional forms), PediMap (pedigree visualization software for PBA), and the PBA database template. Data format conversion tools will be developed as soon as possible by the PBA Team to allow breeders to transfer existing breeding data into a PBA-compliant format efficiently and with less error than if performed ad hoc. BIMS v1.0 will be distributed at RosBREED II to core breeding programs for alpha-testing,

where breeders and trainees will receive instruction in its use by the BIMS Developer. The components and overall package of BIMS will be assessed empirically by periodic surveys of users for system utility, user-friendliness, and by soliciting suggestions for improvement. Alpha-testing will include experience-sharing among core breeding programs.

At RosBREED III, v2.0 will be publicly released to the international Rosaceae genomics, genetics, and breeding community for beta-testing. Version 2.0 will include all BIMS components (3c-j) in a PBA-compliant schema, including a user's manual. Core breeding programs will be heavily involved in beta-testing as they implement BIMS and the MAB Pipeline in increasing depth.

BIMS v3.0 will be released at the RosBREED regional workshops of year 4, and publicly housed on the GDR. For the PBA components of BIMS, users (i.e., all interested U.S. Rosaceae breeders) will be asked to register the software, and periodic surveys will be sent to users by the developers to generate feedback beyond the grant duration. We will also present this software at national and international professional scientific and industry meetings frequented by Rosaceae breeders. BIMS software will be publicly available and economically accessible to all other breeding programs.

Beginning in year 3, the BIMS Developer will visit non-core breeding programs to conduct "technology transfer". During these individual visits to home institutions, individual problems and opportunities can be fully explored. We recognize that constraints to adoption of MAB can fall outside of technical, logistical, and economic considerations, and into the realm of psychological and sociological barriers. The BIMS Developer's experience during the first two years visiting core breeding programs, and regular surveys conducted to assess attitudes to MAB among breeders, should provide much insight, and resources will be redirected to overcoming such barriers where possible. Nevertheless, we believe that powerful arguments for non-core programs to adopt the MAB Pipeline can be made from consideration of the bottom line (economics), face-to-face clear explanations of the steps required to proceed from available marker-trait associations through to routine MAB (illuminating the technical feasibility), and publicizing success stories from core programs (pioneers leading by example).

Other extension efforts will include a dynamic summary of known marker-trait associations in Rosaceae graphically represented on aligned Rosaceae genomes, and regular highlights of high-priority traits describing interesting facts, activities, and impacts. Extension platforms such as the Plant Breeding and Genomics CoP will be utilized for these efforts (Act. 5c).

**Expected Outcomes:** Software, a user-friendly database that leverages genomics information for maximum utility within the MAB Pipeline, and additional functions to aid breeding operation decisions will be developed for ease of adoption and optimized use of PBA in breeding programs. Personnel in Rosaceae breeding programs will be trained in operation of statistical software and database components necessary for the successful implementation of MAB. This common database format will strengthen ties among Rosaceae breeding programs by establishing a common platform for communication of information regarding MAB. Breeding programs adopting BIMS will be enabled to implement MAB, with technical, practical, logistical, and economic hurdles overcome.

**Potential Pitfalls:** The needs of individual breeding programs may be so varied that a single database format will not fulfill their individual requirements. If this occurs, we will ensure a minimum BIMS schema is achieved that interfaces breeding programs with PBA. A recently developed database fulfilling this requirement is AppleBreed (Antofie et al. 2007), although that

platform may not be freely available. Other private databases such as Hertha (AIGN 2009) have been developed but do not currently support PBA. The germplasm composition of some breeding programs may not be suitable for robust outcomes from PBA analysis. If this is the case, efforts will be made to improve the functionality of PBA (such as the GenStat modules) to better suit breeding needs. RosBREED germplasm of individual crops may be too small in size to investigate inter-locus and genotype x locus interactions. As breeding is a continuous process, germplasm for which data are available will increase in time. The PBA approach also allows increase in germplasm dataset sizes through inclusion of data from other programs.

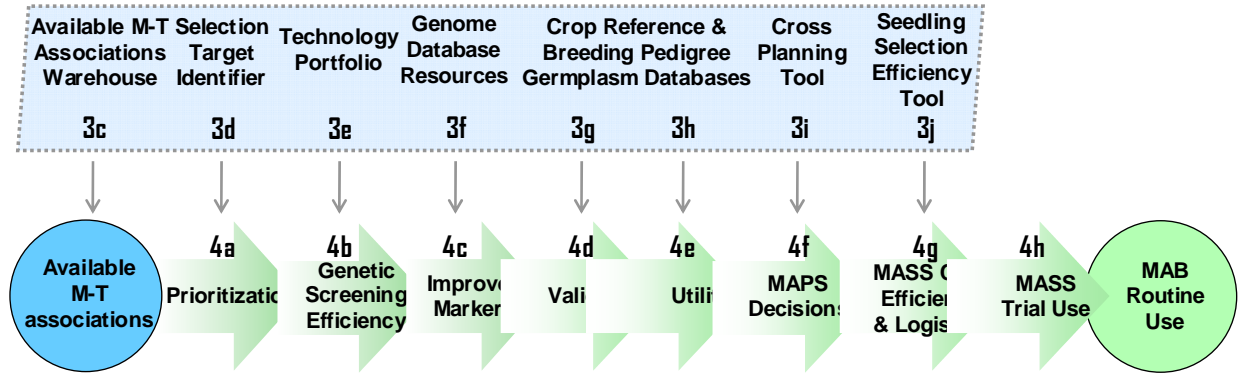
Traits with low heritability will require several seasons and multiple locations of data collection to adequately describe the genetic potential of individuals, which will hinder the timely validation of QTLs (in 3g-h) for use in MAB Pipeline implementation (Act. 4). The best CR Sets to adequately represent a crop may be spread over multiple sites, increasing the environmental variance contributing to phenotypic variation and reducing the power to detect true functional alleles – particularly those with small effects. However, QTLs that are not robust across sites are unlikely to be of interest to breeders. BP Sets phenotypic information will be more site-specific than for CR Sets, such that marker-trait associations and potential breeding values for lower heritability traits are more likely to be confounded by genotype x environment effects. If new target traits are identified in activity 1 during years 2 and 3 of the project, data for those traits will be more limited (fewer years of data).

The plan for individual database components (3c-h) to become community-driven and sustainable may not happen in one or more cases. Extension efforts will be directed as needed to encourage community involvement or identify alternative strategies.

For the cross-planning tool (3i), complexities of parent selection in Rosaceae crops may prove beyond the capability of a multi-module computer program to capture, or may not be sufficiently modeled within the timeframe of the project. To address this, we will develop (and implement in 4f) certain modules that are most immediately applicable to practical breeding needs. This will require prioritization of the planned modules by the BIMS Developer and the BIMS Programmer after initial breeder consultation and continuing breeder feedback.

For the seedling selection efficiency tool (3j), sufficient flexibility for the diversity of breeding programs, regions, and economic situations may be difficult to achieve. The tool will be alpha- and beta-tested and optimized to expand versatility to encompass various breeding situations. Initially, breeders may not have enough reliable financial data available for the tool to provide useful guidance. Breeding trainees will be responsible for obtaining the necessary financial information to fill knowledge gaps.

**Limitations:** Rosaceae breeders may prefer and exclusively use their own databases. The BIMS Team exists to identify and mitigate this limitation in the early stages of the project for core programs and in later years for non-core programs. Continual feedback from on-site visits and en masse during RosBREED meetings will be used to seek solutions. Breeders may rely on “gut instinct” rather than outcomes of the cross-planning tool. To address this, we will highlight success stories of practical use of the tool, and on the theoretical side will highlight comparisons between selection tool-based decisions and more subjectively-based decisions. For some market sectors, scale of production may too small, or competition for market share for new cultivars too severe, that the marginal value of genetic screening with available robust and validated markers will not warrant integration of marker selection into a breeding program.



Infrastructure components and implementation stages of the MAB Pipeline.

## ACTIVITY 4 MAB Pipeline Implementation

**Who:** MAB Pipeline Team, Breeding Team, Socio-Economics Team, Genotyping Team, BIMS Team, PBA Team

**Method:** An eight-stage MAB Pipeline will be implemented in core breeding programs by mobilizing Pipeline infrastructure established in activity 3. Implementation of all eight stages in core RosBREED breeding programs will fill the gap between available M-T associations and routine MAB. Core breeding programs will share a common focus of implementing MAB for fruit quality.

### **4a. Prioritization of marker-trait (M-T) associations**

Available M-T associations will be objectively prioritized for each core breeding program. EWBVs will be assigned to M-T associations in the Warehouse (3c) using the Selection Target Identifier (3d). Breeders and breeding trainees will regularly consult the Warehouse and Identifier (aided by automatic notifications) to choose new M-T associations to be progressed further in the MAB Pipeline for their breeding program. In year 1, the first M-T associations to be advanced are likely to include: Apple – *Md-ACS1* gene associated with firmness after storage (Zhu and Barritt 2008) which is currently in MASS use in the Washington apple breeding program, and a major acidity locus (*Ma*) well characterized in the recent European HiDRAS project; Peach – *endoPG* gene that defines flesh softening (melting vs. non-melting) and stone-flesh adhesion (freestone vs. clingstone) (Peace et al. 2005); and Sweet cherry – SSRs flanking two fruit size QTL (Zhang et al. 2009). All of these reflect critical fruit quality traits identified by core breeders (Letters of Collaboration in Appendix V). For sweet and tart cherry and apple, the *S-RNase* gene that defines self-fertility and cross-compatibility (Ikeda et al. 2004; Matsumoto et al. 2007) is also likely to be used.

### **4b. Genetic screening efficiency**

Efficient genetic screening technologies will be identified for each breeding program. Consulting the Technology Portfolio (3e) with guidance from the Genotyping Team, core breeders will choose locally accessible DNA extraction and genotyping services that are appropriate for the scope and logistics of their MAB needs (e.g., number of samples to be screened, number of markers, availability of breeding staff to collect and pre-process samples, and timeframes available for sampling and genotyping to obtain timely results to support selection decisions).

#### **4c. Improve markers**

At this stage, individual markers associated with each high-priority M-T will be improved if necessary. This improvement will be accomplished by consulting Genome Database Resources (3f) to mobilize available genomics data and with additional guidance from the Genotyping Team. Markers will also be made more efficient for high-throughput screening or more reliable (such as closer linkage or easier to score) if results of MASS cost efficiency and logistics analysis (4g) identify worthwhile opportunities.

#### **4d. Validation of M-T associations in crop-wide germplasm**

Promising M-T associations will be individually validated by obtaining and analyzing genotypic and phenotypic data for crop-wide germplasm and determining the extent to which the associations exist in the crop. The pedigree-linked CR Sets chosen and characterized (pedigreed, performance-assessed, and SNP genome-scanned) in activity 3g will form the crop-wide germplasm, which breeders will analyze with PBA to validate associations with new genotypic data, and sometimes new phenotypic data. CR Sets will be screened with individual improved markers from 4c, and where necessary, additional phenotypic data will be collected for the high-impact trait. In year 1 prior to genome scan results of 3g, CR Sets will be genotyped with improved markers for high-priority traits identified in 4a to provide initial genetic data for breeders, and validated associations determined by PBA will be advanced to the next stage in 4e. For these first genetic screens in year 1, Co-PD Bassil will extract DNA and genotype all CR Sets using existing facilities and equipment. This DNA will be provided for genome scans of 3g.

#### **4e. Utility assessment of M-T associations within breeding programs**

M-T associations validated crop-wide will be assessed for utility within the germplasm of specific breeding programs and growing conditions for target cultivars. The pedigree-linked BP Sets chosen and characterized in Act. 3h, in synergy with relevant CR Set individuals, will form the germplasm representing each program, which breeders will analyze with PBA to characterize the M-T associations. BP Sets will be genotyped with validated markers from 4d, and new phenotypic data collected where necessary. In later years, additional breeding germplasm may be characterized and added to analyses. PBA will be aided during the project duration by the PBA Team but with increasing responsibility placed on individual breeding programs towards independency of PBA use.

#### **4f. MAPS decisions**

Crossing decisions informed by marker data will be performed at this stage. Using data from 4e and the cross-planning tool of 3i, crossing decisions will be informed by markers (marker-assisted parent selection, MAPS). Prior to obtaining marker data in year 1, crossing decisions will be supported using functionalities of the cross-planning tool that do not rely on marker information. Crossing decisions for year 2 will include marker information obtained in 4d and 4e. As markers chosen will have immediate relevance to their programs, breeders may use this initial marker information to design efficient crosses for obtaining desired progeny genotypes or may increase targeted phenotyping efforts in certain existing segregating populations.

#### **4g. MASS cost efficiency and logistics**

Cost-efficient and logistically feasible marker-assisted seedling selection (MASS) schemes will be designed for each breeding program using available, high-priority, validated, utile M-T associations. Such schemes will be determined by the seedling selection efficiency tool (Act. 3j) based on input parameters from chosen genetic screening technologies (4b) for the improved markers (4c) with segregation and performance prediction that determines cull levels in each

cross resolved by utility assessment (4e). This stage will not be conducted in year 1 because several infrastructure components will not yet be established.

#### **4h. MASS trial use**

Breeding programs will trial efficient MASS schemes with seedling populations to test the effects of MAB on program efficiency, experience logistics of incorporating this new tool into routine operations, and optimize the Pipeline to suit idiosyncratic needs. Seedlings (several thousand to be targeted) in each program will be genetically screened with validated and utile markers at optimal times in the breeding cycle determined in Act. 4g, using appropriate technologies according to 4b. Breeders, their trainees, and supporting staff will gain firsthand experience of the impact of MAB. Each program will collect detailed economic and observational data to determine economic and logistical impacts of MASS, and compare to previous or concurrent non-MASS protocols. This stage will not be conducted in year 1 because several infrastructure components will not yet be established.

**Analysis of Results:** The MAB Pipeline Team will provide guidance to breeding programs as they implement each stage of the Pipeline. Other relevant Teams will provide further guidance for each stage. PBA software will be used to determine if genetic data supports breeder knowledge of ancestry, and to validate and assess utility of M-T associations. PBA will describe functional allelic diversity by estimating phenotypic effects of individual QTL alleles and calculating the probability of functional allele genotypes for each individual. Decision-support tools used in 4f and 4g will allow the breeder to translate new DNA information into practical advice for their programs.

**Use of Results:** Information gained from implementation of the eight MAB Pipeline stages will be used to guide breeding decisions to develop and apply optimal selection strategies aimed at objectively defined selection targets. Information collected on CR Sets will be included in the permanent CR Set data archives (3g), while BP Set data will remain the property of individual breeding programs. The initial and ongoing data collection will also be used for training breeders in PBA (3b). Data collected by programs on MASS costs and logistics will provide an unprecedented opportunity to compare implementation across breeding programs in an array of related perennial species and to compare impacts on several breeding programs within the same crop. Decision-support tools will be modified to reflect this collective experience.

**Outreach Plan:** The target audience for Activity 4 is the breeders, breeding trainees, and other supporting staff of the core breeding programs (whereas extension for non-core breeders is in Act. 5). Successful implementation of individual MAB Pipeline stages, and MAB use itself, will be highlighted at annual meetings (5a) and through the Plant Breeding and Genomics Community of Practice (PBG CoP; 5c), as will lessons learned from unsuccessful implementation. The MAB Pipeline Team will guide all implementation efforts and facilitate liaison with other supporting RosBREED Teams.

Best examples of MAB Pipeline implementation at the utility, cross-planning, and seedling selection stages of the pipeline will be showcased during the nine Participatory Workshops (5a) and developed into a public information section of the RosBREED PBG CoP (5c).

**Expected Outcomes:** We expect that high-priority M-T associations will be identified and utilized in MAB within individual Rosaceae breeding programs to provide success stories of routine MAB. For some programs, this will begin in year 1 to inform crossing decisions. MAPS based on functional genotypes of parents as well as marker-informed pedigree verification is the first expected application of MAB in this project. However, the cross-planning tool should also

be used to aid parent selection decisions that do not involve marker information. For all core breeding programs, results from PBA analyses of genome scans and standardized phenotypes for important traits is expected to lead to many useful M-T associations' being exploited via MAB during the term of the project. While this initial implementation of the MAB Pipeline focuses on genotyping for individual trait loci, opportunities for whole genome selection may be explored. The results of 3g-h (Crop Reference Set and Breeding Pedigree Set characterization) will provide whole genome functional genotype information for breeding parents and founders, offering the prospect of whole genome seedling selection.

Breeding programs implementing the MAB Pipeline are expected to be more efficient in their development of new cultivars. Impact will also include expanded information discovery in genetics and breeding research programs. Implementation of the MAB Pipeline is expected to result in the identification of selection targets better aligned with stakeholder expectations, and identification of efficient selection strategies for those targets. Breeders will optimize allocation of breeding program resources accordingly.

**Potential Pitfalls:** Insufficient or inaccurate phenotypic data in breeding programs may hinder progress, although such deficiencies also affect traditional breeding progress. RosBREED funds provide limited support for phenotyping. However, as targeted traits will be high-priority for each breeding program, it is probable that sufficient accurate phenotypic data will be available within programs. Implementation cannot occur if decision-support tools for parent and seedling selection do not identify opportunities for marker assistance.

**Limitations:** Only a few robust and cost-efficient M-T associations may be available for a given breeding program, whereas the impact of MAB may not be realized until several valuable markers are used. Even in such cases, the MAB Pipeline will be established in the breeding program, allowing future genomics discoveries in M-T associations to enter the pipeline and eventually provide high-impact genetic screening.

## **ACTIVITY 5**

### **Extension**

**Who:** Extension Team

**Method:** Meetings and workshops will be conducted to engage, educate, and train Rosaceae breeders and allied scientists, producers/processors, marketing groups, and trade organizations on RosBREED's socio-economics and genomics research and application. Extension programming will include short courses and hands-on computer training. Breeding trainees will be broadly trained in integrating breeding, genomics, and socio-economics for crop improvement. RosBREED personnel will actively participate in the eXtension Plant Breeding and Genomics CoP (PBG CoP; PBGworks 2009), and utilize this resource as a vehicle to create educational content for the larger community and public (via a Community of Interest, CoI). External evaluation of extension activities will provide feedback and prompt adjustments in activities.

#### **5a. Meetings and Workshops**

RosBREED will hold four Annual Project Meetings, four Regional Workshops in year 4, and nine Participatory Workshops in years 3 and 4, with specifically appropriate venues, hosts, attendees, focuses, and target audiences.

The annual meetings (RosBREED I-IV) will be targeted at project participants, collaborators, and Advisory Panel members, providing milestones of accomplishment and

software technical training. RosBREED I will be held in 7-8 January 2010 in conjunction with the Plant and Animal Genome Conference. This inaugural meeting will coordinate activities across breeding programs, define breeder software needs, and design content for future workshops. The subsequent three Annual Project Meetings will include workshops to facilitate integration, interaction, and information sharing among our diverse group of participants and Advisory Panel members. Training of core breeders and their breeding trainees in use of MAB Pipeline software components will be a major focus of RosBREED II and III, both of which will be held in March at Mich. State Univ. to take advantage of computer classrooms. Activities will include two-day breeder- and trainee-targeted workshops under the tutelage of the PBA and BIMS Teams. These statistical database activities will feature hands-on training in data entry and analysis for breeding program CR Sets and BP Sets (3g-h). Co-PD Stone and extension staff will provide training in the use of PBG CoP workspace tools (5c). RosBREED II and III will also include activities targeting the Stakeholder Advisory Panel and other non-technical participants, as educating these industry leaders will provide a direct connection to their clientele. For these audiences, co-PD Hancock will adapt a 15-lecture undergraduate course in “Applied Plant Improvement” and offer it as a two-day MAB short course incorporating RosBREED approaches and progress. RosBREED IV will be the project-culminating event, scheduled at a time and location fostering interaction among project participants and the wider interested community, including non-core U.S. rosaceous crop breeders. At RosBREED IV, co-PD Gallardo will present a one-day short course on integrating socio-economics research with breeding, production, and marketing, to discuss progress and prospects on socio-economic approaches to develop relative economic values of fruit quality and production traits to inform breeding and industry operations. Travel funds are budgeted to bring in additional individuals to enhance activities and interaction.

RosBREED will host an Introductory Workshop in August 2010 at the American Society for Horticultural Science annual conference, Palm Desert, CA, to introduce RosBREED to a broader research and extension audience. This Workshop will be the first official RosBREED meeting that includes breeding trainees. In addition to providing background and progress updates on RosBREED and networking opportunities for project participants, the workshop will address the opportunities to sustain plant breeding through professional and graduate training via systems-oriented, community-level activities like RosBREED.

Regional and Participatory Workshops in later years will showcase operations in core breeding programs, facilitate technology transfer to non-core programs, and engage industry stakeholders and allied scientists. A key target audience of these Workshops will be non-core Rosaceae breeders, for technology transfer of the socio-economically and genomically informed BIMS decision-support framework.

Nine Participatory Workshops will be hosted by the RosBREED core breeding programs (NY, MI, SC, AR, TX, CA, WA, MN, NH) to motivate and assist core and non-core breeders in MAB adoption, and educate industry stakeholders and allied scientists. Workshops will be scheduled and integrated within annual local industry meetings. Four workshops will be held in year 3 and five in year 4, each providing opportunities to target content to specific needs of individual breeding programs and their local industries. Activities of core breeders closest to the workshop locations will provide case studies to illustrate optimized selection strategies (with MAB) for economically weighted breeding targets. Co-PDs Weebadde and Hancock will adapt and present an existing half-day Short Course (“Molecular Plant Breeding”) currently offered

through MSU's PBROC (WorldTAP 2009). The BIMS Developer will participate actively at these workshops to assist local programs establish MAB capability.

Four Regional Workshops held in year 4 in NY, SC, WA, and CA will be hosted by co-CPDs with extension expertise (Xu [NY], Reighard [SC], Gallardo [WA], and Crisosto [CA]) and will focus on trait values and MAB training of regional extension specialists (i.e. "train-the-trainer") and technology transfer to regional non-core breeding programs. The Workshops will facilitate communication among breeders and allied scientists, producers/processors, marketing groups, and trade organizations. Activities will include: (1) the half-day PBROC short course "Molecular Plant Breeding"; (2) case studies in MAB from RosBREED core breeders in the region; (3) a site visit to a core breeding program; (4) a short course by the Socio-Economics Team on their economic valuation tools for fruit quality and production traits; and (5) personal and small group tutelage by the BIMS Developer to assist non-core breeders implement MAB.

### **5b. Trainee mentoring**

Graduate students embedded as breeding trainees will each receive mentoring from active breeders and other project participants, and will benefit from a unique integrated educational experience that includes the interface between breeding, socio-economics, and genomics. Trainees are expected to be heavily involved in adopting BIMS modules and utilizing other MAB Pipeline infrastructure components for their program, and implement the Pipeline towards routine MAB. Trainees will provide support to their breeder mentor with RosBREED activities. Trainees will share experiences (with other trainees, breeders, allied scientists and stakeholders) at meetings and workshops (5a), through project extension platforms (5c), and regular avenues. Trainees will participate in networked activities within the PBG CoP, with particular emphasis on learning and utilizing PBGworks (5c).

### **5c. Project extension platforms**

Sustainable platforms for project communication and public engagement will be developed for engagement among project participants. RosBREED personnel will actively participate in the eXtension PBG CoP, and utilize PBGworks as an online venue to create educational content for communication to less technical audiences via a CoI. In year 1, we will begin utilizing PBGworks to develop project management capabilities and expand RosBREED and community participation. We will build vigorous Rosaceae-specific forums within PBGworks for real-time breeder-to-breeder communications on topics such as MAB, PBA, BIMS, and project progress. In year 2 we will develop eXtension content under the leadership of Co-PD Weebadde featuring individual programs, cultivar development stories, successful application of socio-economics and genomics technologies to improve breeding program efficiency, and program impacts on stakeholders. Synergies in content development will be explored with relevant CAP projects within the PBG CoP (Letter of Collaboration from D. Douches, Appendix VI). In years 2 and 3, RosBREED annual meetings will provide opportunities to equip breeders and trainees with skills to fully utilize evolving capacities of PBGworks. In year 4, particular emphasis will be placed on providing non-technical content for the CoI.

### **5d. Evaluation of extension activity impact**

Dr. Michael Coe (Cedar Lake Research and Consulting Group, OR) will provide external evaluation of all extension activities (Letter of Support, Appendix VI). Evaluation surveys, interviews, and event feedback forms will be used on an ongoing basis to gather feedback from various participant groups and extension audiences about quality, relevance, and utility of extension activities, as well as recommendations for further research and development.

Evaluation feedback will be provided to the Extension Team and the RosBREED Executive Committee on a regular basis for continuous quality improvement, and will be summarized in annual reports to document the relevance, quality, and utility of extension activities. More detail on the evaluation framework is included in Appendix I (Evaluation and Metrics, pg. 26).

**Expected Outcomes:** Project breeders, breeding trainees, and allied scientists will acquire new knowledge and improve cross-commodity and cross-disciplinary collaboration that will increase project productivity and help achieve a key project outcome: the adoption of MAB in Rosaceae breeding programs. Project participants will improve communication, speed of data sharing, and project management utilizing PBGworks. The power and reach of internet-based extension will deliver science-based knowledge cost- and time-effectively to Rosaceae industries as well as a broader audience. These audiences will better understand genomics technologies and socio-economics approaches and how they enable breeding programs to more rapidly develop and commercialize new cultivars preferred by consumers and valued by industry stakeholders.

Core breeding programs will respond to engagement, education, and training by successfully implementing efficient selection strategies for objectively defined trait targets, involving MAB and socio-economics, respectively. Via the same outreach avenues but without RosBREED funding support for genotyping and phenotyping in their specific breeding program, non-core breeding programs will be inspired and enabled to adopt efficient MAB approaches. Breeding trainees will emerge as broadly trained scientists capable of leading state-of-the art programs. Engagement and education of industry stakeholders will create a pull-through demand for efficient and predictable genomics technologies and the project's novel socio-economics approaches. Breeders and industry stakeholders will identify new ways to jointly accelerate commercialization and impact of new cultivars. In total, these outcomes will measurably enhance sustainability of Rosaceae crop breeding.

**Potential Pitfalls:** The diversity of project participants and stakeholders by crop, scientific discipline, geographic location, and level of awareness and understanding of socio-economics and genomics in general and MAB in particular challenges uniform educational and training outcomes. Participants may be slow to appreciate and utilize web-based resources. Delivering effective outreach will require significant coordination, flexibility, and communication. Participants may adopt project approaches or technologies inconsistently, partially, not at all, or reject them after initial adoption. Rigorous evaluation, personalized attention to specific participant problems, and adjustment of project research and extension activities will help address such pitfalls.

**Limitations:** Breeding programs in some host institutions may lack academic or research capability in Rosaceae genomics and genetics, and require identification of scientists with expertise outside of Rosaceae to provide appropriate collaboration and mentoring.

## BIBLIOGRAPHY & REFERENCES CITED

1. AIGN (2009) "Hertha Horticultural Database". Retrieved April 6, 2009.  
<http://www.aign.org/Hertha.htm>
2. Antofie A, Lateur M, Oger R, Patocchi A, Durel CE, Van de Weg WE (2007) A new versatile database created by geneticists and breeders to link molecular and phenotypic data in perennial crops: the *AppleBreed DataBase*. *Bioinformatics* 23:882-889
3. Arús P, Gardiner S (2007) Genomics for improvement of Rosaceae temperate tree fruit. In: RK Varshney, R Tuberosa (eds.) *Genomics-Assisted Crop Improvement, Vol 2: Genomic Applications In Crops*. Springer Netherlands. pp 357-397
4. Arús P, Yamamoto T, Dirlewanger E, Abbott AG (2005) Synteny in the Rosaceae. *Plant Breeding Reviews* 27:175-211
5. Besley T, Case A (1993) Modeling technology adoption in developing countries. *American Economic Review* 83:396-402
6. Bink MCAM, Boer MP, ter Braak CJF, Jansen J, Voorrips RE, van de Weg WE (2008) Bayesian analysis of complex traits in pedigreed plant populations. *Euphytica* 161:85-96
7. Chagné D, Gasic K, Crowhurst RN, Han Yuepeng, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EHA, Gardiner SE, Korban SS (2008) Development of a set of SNP markers present in expressed genes of the apple. *Genomics* 92:353-358
8. Chen L, Zhang SM, Illa E, Song LJ, Wu SD, Howad W, Arús P, van de Weg E, Chen KS, Gao ZS (2008) Genomic characterization of putative allergen genes in peach and their synteny with apple. *BMC Genomics* 9:543
9. Crabtree J, Angiuoli SV, Wortman JR, White OR (2007) Sybil: methods and software for multiple genome comparison and visualization. *Methods Mol Biol* 408:93-108
10. Crosby JA, Janick J, Pecknold PC, Korban SS, O'Connon PA, Ries SM, Goffreda J, Voordeckers A (1992) Breeding apples for scab resistance: 1945-1990. *Acta Horticulturae* 317:43-70
11. Davis TM, Folta KM, Zhang Q, Tombolato DCM, Mishra V, Kahveci T (2008) Sequence samples and gene pair haplotypes in Strawberry. *International Plant & Animal Genome Conference XV*, January 10-15, San Diego, CA. Invited poster.
12. DFCI (2009) "DFCI Apple Gene Index". Retrieved April 6, 2009.  
<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=apple>
13. Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howard W, Arús P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proceedings of the National Academy of Sciences of the United States of America* 101:9891-9896
14. GDR (2009a) "GDR: Genome Database for Rosaceae". Retrieved April 6, 2009.  
<http://www.bioinfo.wsu.edu/gdr/>
15. GDR (2009b) "Rosaceae EST Unigene". Retrieved April 7, 2009.  
[http://www.bioinfo.wsu.edu/gdr/unigene\\_project.php](http://www.bioinfo.wsu.edu/gdr/unigene_project.php)
16. Han Y, Chagné D, Gasic K, Rikkerink EHA, Beaver JE, Gardiner SE, Korban SS (2008) BAC-end sequence-based SNPs and Bin mapping for rapid integration of physical and genetic maps in apple. *Genomics* 93:282-288
17. Howad W, Yamamoto T, Dirlewanger E, Testolin R, Cosson P, Cipriani G, Monforte AJ, Georgi L, Abbott AG, Arús P (2005) Mapping with a few plants: Using selective

- mapping for microsatellite saturation of the *Prunus* reference map. *Genetics* 171:1305-1309
18. Iezzoni A, Abbott A, Folta K (2007) "Rosaceae Specialty Crops Planning Workshop. June 22-23, 2007. Final Workshop Report." Retrieved April 6, 2009. [http://www.csrees.usda.gov/business/reporting/stakeholder/pdfs/rosaceae\\_workshop.pdf](http://www.csrees.usda.gov/business/reporting/stakeholder/pdfs/rosaceae_workshop.pdf)
  19. Iezzoni A, Zhang G, Wang D, Sooriyapathirana SS, Sebolt AM, Van de Weg WE, Bink MCAM (2008) Pedigree Based Analysis: Our experiences using multiple pedigreed populations in sweet cherry. 4<sup>th</sup> International Rosaceae Genomics Conference. Pucon, Chile. Abstract, p. 37. <http://www.genomavegetal.cl/rgc4/2007/RGC4-abstract-book.pdf>
  20. Ikeda K, Watari A, Ushijima K, Yamane H, Hauck NR, Iezzoni AF, Tao R (2004) Molecular markers for the self-compatible *S<sup>4</sup>*-haplotype, a pollen-part mutant in sweet cherry (*Prunus avium* L.). *Journal of the American Society for Horticultural Science* 129:724-728
  21. Jung S, Jiwan D, Cho I, Lee T, Abbott A, Sosinski B, Main D (2009) Synteny of *Prunus* and other model plant species. *BMC Genomics* 10:76
  22. Korban SS, Vodkin LO, Liu L, Aldwinckle H, Carroll N, Goldsbrough P, Orvis K, Clifton S, Main D (2009) "Apple EST Project". Retrieved April 6, 2009. <http://titan.biotec.uiuc.edu/apple/apple.shtml>
  23. Luby JJ, Shaw DV (2001) Does marker-assisted selection make dollars and sense in a fruit breeding program? *HortScience* 36:872-879
  24. Lusk JL, Schroeder TC (2004) Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks. *Artefactual Field Experiments* 0067. The Field Experiments Website. <http://ideas.repec.org/p/feb/artefa/0067.html>
  25. Matsumoto S, Eguchi T, Bessho H, Abe K (2007) Determination and confirmation of the S-RNase genotypes of apple pollinators and cultivars. *Journal of Horticultural Science and Biotechnology* 82:323-328
  26. NBCI (2009) "National Berry Crops Initiative". Retrieved April 6, 2009. <http://www.nationalberrycrops.org/>
  27. NCBI (2009) "Expressed Sequence Tags Database". Retrieved April 6, 2009. <http://www.ncbi.nlm.nih.gov/dbEST/>
  28. Paterson AH, Lin Y-R, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu S-C, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269:1714-1718
  29. PBGworks (2009) "RosBREED Forum". Retrieved April 6, 2009. <http://pbgworks.hort.oregonstate.edu/forum/41>
  30. PBROC (2008) "Plant Biotechnology Resource and Outreach Center". Retrieved April 6, 2009. <http://www.ptc.msu.edu>
  31. Peace CP, Crisosto CH and Gradziel TM (2005) Endopolygalacturonase: a candidate gene for *Freestone* and *Melting flesh* in peach. *Molecular Breeding* 16:21-31
  32. Peace CP, Norelli J (2009) Genomics approaches to crop improvement in Rosaceae. In: K Folta and S Gardiner (eds) *Genetics and Genomics of Rosaceae*. Springer. pp 19-53
  33. RosBREED (2009) "RosBREED: Enabling marker-assisted breeding in Rosaceae. A group of Rosaceae breeders, genomicists, quantitative geneticists, bioinformaticians, stakeholders, & extension educators". Retrieved April 6, 2009. [http://hrt2.hrt.msu.edu/faculty/Iezzoni/RosBREED\\_Consortium.html](http://hrt2.hrt.msu.edu/faculty/Iezzoni/RosBREED_Consortium.html)

34. Sargent DJ, Marchese A, Simpson DW, Howad W, Fernández-Fernández F, Monfort A, Arús P, Evans KM, Tobutt KR (2009) Development of “universal” gene-specific markers from *Malus* spp. cDNA sequences, their mapping and use in synteny studies within Rosaceae. *Tree Genetics & Genomes* 5:133–145
35. SGN (2009) “SOL Genomics Network”. Retrieved April 7, 2009.  
<http://www.sgn.cornell.edu/>
36. Shulaev V, Folta K (2009) Strawberry Genome Sequencing Consortium. Retrieved April 6, 2009. <http://strawberry.vbi.vt.edu/tiki-index.php?page=Strawberry+Genome+Sequencing+Consortium>
37. Shulaev V, Korban SS, Sosinski B, Abbott AG, Aldwinckle HS, Folta KM, Iezzoni A, Main D, Arus P, Dandekar AM, Lewers K, Brown SK, Davis TM, Gardiner SE, Potter D, Veilleux RE (2008) Multiple models for Rosaceae genomics. *Plant Physiology* 147:985-1003
38. Spigler RB, Lewers KS, Main DS, Ashman T-L (2008) Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity* 101:507-517
39. US RosEXEC (2009a) “U.S. Rosaceae Genomics, Genetics, and Breeding Executive Committee”. Retrieved April 6, 2009.  
<http://www.bioinfo.wsu.edu/gdr/community/rosexec/index.php>
40. US RosEXEC (2009b) “The U.S. Rosaceae Genomics, Genetics, and Breeding Initiative White Paper”. Retrieved April 6, 2009.  
[http://www.bioinfo.wsu.edu/gdr/community/rosexec/documents/RosWP\\_March\\_2006.doc](http://www.bioinfo.wsu.edu/gdr/community/rosexec/documents/RosWP_March_2006.doc)
41. USDA ARS FNRI (2009) “Floriculture and Nursery Research Initiative”. Retrieved April 6, 2009.  
[http://www.ars.usda.gov/research/programs/programs.htm?np\\_code=301&docid=13683&page=1](http://www.ars.usda.gov/research/programs/programs.htm?np_code=301&docid=13683&page=1)
42. Van de Weg WE, Voorrips RE, Finkers R, Kodde LP, Jansen J, Bink MCAM (2004) Pedigree genotyping: A new pedigree based-approach of QTL identification and allele mining. *Acta Horticulturae* 663:45-50
43. Van Ooijen JW (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV: Wageningen, Netherlands
44. Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* 5:247-252
45. Vilanova S, Arus P, Sargent DJ, Monfort A (2008) Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). *BMC Plant Biology* 8:67
46. Volk G, Bassil NV (2009) “Standardized Phenotyping”. Retrieved April 7, 2009.  
<http://www.bioinfo.wsu.edu/gdr/phenotype/>
47. VSN International (2009) “GenStat”. Retrieved April 6, 2009.  
<http://www.vsnl.co.uk/software/genstat/>
48. Weebadde CK (2009) “Molecular Plant Breeding – An International Short Course on Practical Applications of Molecular Tools for Plant Breeding”. Retrieved April 6, 2009.  
<http://www.worldtap.msu.edu/home/page/70>

49. Weebadde CK, Wang D, Finn CE, Lewers KS, Luby JJ, Bushakra J, Sjulín TM, Hancock JF (2008) Using a linkage mapping approach to identify QTL for day-neutrality in the octoploid strawberry (*Fragaria × ananassa* Duche x Rozier). *Plant Breeding* 127:94-101
50. WTFRC (2009) “Technology Roadmap. Research priorities to enhance tree fruit production through technological innovation”. Retrieved April 6, 2009.  
<http://www.treefruitresearch.com/technology-roadmap/>
51. Yue C, Tong C (2009) Organic or local? Investigating consumer preference for fresh produce using a choice experiment with real economic incentives. *HortScience* 44:366-371
52. Zhang G, Sebolt AM, Sooriyapathirana SS, Wang D, Bink MCAM, Olmstead JW, Iezzoni AF (2009) Fruit size QTL analysis of an F<sub>1</sub> population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. *Tree Genetics & Genomes* (in press).
53. Zhu Y, Barritt BH (2008) Md-ACS1 and Md-ACO1 genotyping of apple (*Malus × domestica* Borkh.) breeding parents and suitability for marker-assisted selection. *Tree Genetics & Genomes* 4:555-562