USDA Cooperative State Research, Education and Extension Service
National Research Initiative Plant Genome Program
Plant and Animal Genome Conference XV
Town and Country Hotel, San Diego

Friday, January 12, 2007
6:00 pm – 10:00 pm
Windsor Rose Room
(Located on the 9th floor of the Regency Tower)

Rosaceae Project Directors Meeting

6:00pm: Meeting room open. Refreshments available. Load PowerPoint Presentations.

6:15pm: CSREES NRI Plant Genome Research – Ed Kaleikau

CSREES NRI Plant Biology – Gail McLean/Liang-Shiou Lin

6:30pm: Project Updates (10 minutes each)

- **Completion of the Peach Genome Database: A Reference Genome for *Rosaceae* – Bert Abbott**
- **Gender Determination, the Key to Germplasm Utilization in Strawberry: Genetic Mapping and Colinearity with Peach – Kim Lewers**
- **Genomic Resources to Improve Fruit Size and Quality in Sweet Cherry – Amy Iezzoni**
- **Development of Segregating Populations for Molecular and Genetic Analyses of X-Disease in Chokecherry (*Prunus virginiana* L.) –Wenhao Dai**
- **Candidate Genes for Fruit Softening in *Prunus* – Cameron Peace**
- **Genetic Diversity of Wild Apple Accessions in the National Plant Germplasm System – Gayle Volk**
- **Functional Genomic Response of Apple to Fire Blight – Jay Norelli**
- **High-Resolution Physical Mapping of the Apple Genome by BAC Fingerprinting – Schuyler Korban**

8:00pm: Break

8:15pm: Continue Project Updates (10 minutes each)

- **Identifying the Genes Associated with Day-Neutrality in Strawberry Using a QTL Approach – Jim Hancock**
- **Increasing the Diversity of EST Sequences for *Fragaria* – Janet Slovin**
- **Gene Pair Haplotypes and Sequence Samples from Strawberry (*Rosaceae*): Multi-Purpose, Transferable Resources for Genomics and Variety Improvement – Tom Davis**

- **Algorithms and Programs for Gene Expression QTL Analysis – Zhao-Bang Zeng**

9:15pm: General discussion
***Rosaceae*, Genomics, Genetics and Breeding Executive Committee – Amy Iezzoni and Kevin Folta**

- **Report: NRI funded project outcomes, accomplishments and deliverables.**
- **Questions:**
  - **What are the knowledge gaps and research, education, and extension opportunities in *rosaceae* genomics, genetics and breeding?**
  - **How can we move the field of *rosaceae* genomics forward to have the greatest impact? What research is needed to provide the best return on investment to the U.S. taxpayer?**
  - **What would a broad road map of this research look like?**
    - **Genome Tools, Resources and Informatics?**
    - **Functional Genomics?**
    - **Genome Structure & Organization?**

10:00pm: Adjourn

TABLE OF CONTENTS

USDA-CSREES NRI PLANT GENOMICS, GENETICS AND BREEDING

Ed Kaleikau
USDA Cooperative State Research, Education and Extension Service (CSREES)
Competitive Programs – National Research Initiative (NRI)
Washington DC
E-mail: ekaleikau@csrees.usda.gov
Telephone: 202-401-1931
Website: http://www.csrees.usda.gov/; http://www.csrees.usda.gov/funding/nri/nri.html

The CSREES NRI competitive grants plant genome program supports research, education and extension projects ranging from technology development to fundamental science and practical application for crop or forestry improvement in the U.S. Its priorities focus on technological advances and discoveries in areas such as a) analytical methods for mapping genes for complex traits for direct use by plant breeders, b) novel methods for analysis of the genome and its effect on biological function, c) cost-effective sequencing strategies to understand complex genome structure and organization, d) procedures to analyze the total expression patterns of genes under specific conditions, and e) appropriate data handling and analysis capabilities. The ultimate goal of the program is to contribute knowledge about the biology of agriculturally important plant processes and traits, which can be used to develop crops with enhanced economic value and expanded utilities.

To meet these identified needs of agriculture, the long-term (10 year) goals for this program are: increased fundamental knowledge of the structure, function and organization of plant genomes to improve agricultural efficiency and sustainability; effective integration of modern molecular breeding technologies and traditional breeding practice for U.S. crop and forestry improvement; and improved U.S. varieties for agricultural growers and producers.

The NRI plant genome program coordinates its activities with participating agencies (NSF, DOE, NIH, etc) of the Interagency Working Group on Plant Genomes. The NRI program focuses on genomics, genetics and breeding including: Tools, Genetic Resources, Bioinformatics, Functional Genomics, Genome Structure and Organization and Applied Plant Genomics.

The National Research Initiative Competitive Grants Program (NRI) is the office in the Cooperative State Research, Education and Extension Service (CSREES) of the USDA charged with funding research on key problems of national and regional importance in biological, environmental, physical, and social sciences relevant to agriculture, food, the environment, and communities on a peer-reviewed, competitive basis. To address these problems, NRI advances fundamental scientific knowledge in support of agriculture and coordinates opportunities to build on these scientific findings. The resulting new scientific and technological discoveries then necessitate efforts in education and extension to deliver science-based knowledge to people, allowing informed practical decisions. Competition is open to scientists at all academic institutions, Federal research agencies, private and industrial organizations, and as individuals. The NRI Program Description is distributed widely within the scientific community and among other interested groups. The FY 2005 Request for Applications contained 31 programs with 10 programs soliciting integrated research, education, and extension projects in addition to research projects.

USDA-CSREES NRI PLANT BIOLOGY

Liang-Shiou Lin and Gail McLean
USDA Cooperative State Research, Education and Extension Service (CSREES)
Competitive Programs – National Research Initiative (NRI)
Washington DC
E-mail: llin@csrees.usda.gov and gmclean@csrees.usda.gov
Telephone: 202-401-5045 and 202-401-6060
Website: http://www.csrees.usda.gov/; http://www.csrees.usda.gov/funding/nri/nri.html

The CSREES NRI competitive grants program "Plant Biology: Foundation for Agricultural and Forest Plant Production and Improvement" supports projects that will provide fundamental knowledge and training for improvement and sustainability of agricultural plant and forestry production. Knowledge of plant biology from the molecular to the systems level is essential for development of plants with increased productivity, fitness, and use. Such fundamental understanding of plant biology will allow scientists to make use of the increasing wealth of genomics data and tools and to develop new varieties of agricultural plants through techniques such as biotechnology and classical breeding.

The science-based knowledge and education contributed by this program can lead to increased economic opportunities for producers and consumers by reducing production costs, improving quality, and increasing value of agricultural plant products. This knowledge will allow U.S. agriculture to face critical needs in the areas of bioenergy, environmental change, loss of agricultural land, and increasing global competition.

The Plant Biology program consists of four program elements: Plant Biology (A): Gene Expression and Genetic Diversity; Plant Biology (B): Environmental Stress; Plant Biology (C): Biochemistry; and Plant Biology (D): Growth and Development. In FY 2007, the program elements Plant Biology (A): Gene Expression and Genetic Diversity and Plant Biology (B): Environmental Stress solicited both research and integrated projects. Integrated projects in these two program elements include a plant breeding education component. Program elements Plant Biology (C): Biochemistry and Plant Biology (D): Growth and Development are soliciting research projects only.

The National Research Initiative Competitive Grants Program (NRI) is the office in the Cooperative State Research, Education and Extension Service (CSREES) of the USDA charged with funding research on key problems of national and regional importance in biological, environmental, physical, and social sciences relevant to agriculture, food, the environment, and communities on a peer-reviewed, competitive basis. To address these problems, NRI advances fundamental scientific knowledge in support of agriculture and coordinates opportunities to build on these scientific findings. The resulting new scientific and technological discoveries then necessitate efforts in education and extension to deliver science-based knowledge to people, allowing informed practical decisions. Competition is open to scientists at all academic institutions, Federal research agencies, private and industrial organizations, and as individuals. The NRI Program Description is distributed widely within the scientific community and among other interested groups. The FY 2007 Request for Applications contained 26 programs with 15 programs soliciting integrated research, education, and extension projects in addition to research projects.

Completion of the Peach Genome Database: A Reference Genome for *Rosaceae*

Investigators: T. Zhebentenyayeva, G. Swire-Clark, B. Blackmon,W.V. Baird, G. Reighard, B. Sosinski, P. Arus, D. Main, J. Tomkins and A. Abbott

The goal of our research is to develop peach as a model genetic resource for the identification, characterization and cloning of important genes of Rosaceae species. The utility of having as complete a genomics database (integrated genetic/physical map and mapped EST database) for a Rosaceae model species genome is clearly without question. A physical map serves as an ideal tool to cross compare maps of different species and to identify cloned genomic regions containing important gene loci thus facilitating the process of gene marking and gene discovery in related species. In addition, physical maps provide marker resources that can serve to bridge the gap from the mapping of specific characters to the implementation of marker assisted breeding schemes. For this reason in 2001 -2004, we constructed an initial framework physical map/EST database for peach anchored on the general *Prunus* molecular genetic map. However, to complete the model genomics database for peach, this currently funded project is directed at the following specific aims:

1. Complete construction of the physical/genetic map of peach using HICF (High information content fingerprinting) of the existing physical map framework contig clones and those of larger insert libraries of the haploid cultivar 'Lovell'.
Progress: We have substantially completed this project aim. We have fingerprinted 20,000 BAC clones using HICF technologies. These fingerprints include the original physical map of peach and additional fingerprints from larger insert BAC libraries constructed from the haploid and dihaploid “Lovell”. Current estimates of genome coverage are at approximately 80%. Currently, we are re-assembling the physical map of peach with these new data to obtain the more complete map. Once finished, if necessary, we will utilize BAC end sequencing and hybridization to close the remaining gaps.

2. Incorporate the physical map data into the Genome Database for Rosaceae (GDR) to provide a reference genome for identification and cloning of genes important to rosaceous crop development and sustainability.
Progress: All our data is submitted regularly to the GDR database and the current peach genome database is being utilized by numerous laboratories in the search for candidate genes or markers.

3. Complete development of a high density genetic marker set anchored on the physical map to provide the tools necessary for marker assisted selection, comparative mapping and molecular map development in the less well characterized Rosaceae species.

Progress: We have commenced physical mapping of 400 genetically anchored SSR’s developed in the *Prunus* genomics community. These SSR sequences will strengthen the integration of the general *Prunus* genetics map and the physical map/EST database. We are already using the physical map in several comparative mapping projects to study genome evolution and to identify candidate genes for important characters in Rosaceae species.

GENDER DETERMINATION, THE KEY TO GERMPLASM UTILIZATION IN STRAWBERRY: GENETIC MAPPING AND COLINEARITY WITH PEACH, THE ROSACEAE MODEL SYSTEM

Investigators: Tia-Lynn Ashman, Co-PD: Kim S. Lewers

Performing Organization: University of Pittsburgh
Department of Biological Sciences
E-mail: tia1@pitt.edu
Telephone: 412-624-0984

The goal of this project is to develop gender-linked molecular markers so strawberry breeders can fully utilize extremely valuable wild germplasm. The supporting objectives are to 1) identify and characterize gender-related genes in octoploid wild *Fragaria* via QTL mapping, and 2) determine the level of similarity between peach (or *Prunus*) and *Fragaria* genomes with respect to markers linked to gender in strawberry. Plans to accomplish project goals include: 1) constructing mapping populations to capture the greatest diversity of putative gender genes/alleles; 2) quantitatively and qualitatively assessing gender with respect to both female and male traits in replicated greenhouse and field plantings; 3) using both a candidate gene approach and broad coverage of the genome with microsatellite markers developed from strawberry, peach and other Rosaceae.

Accomplishments as of 1 November 2006:

- To identify parents with the correct genotypes for developing two mapping populations, 28 testcrosses were made. The resulting progeny evaluated for the presence or absence of fruit and the amount of fruit produced.
- Concurrently, 15 mapping populations were made between all potential parents. When the testcross results were evaluated, seed for the two mapping populations derived from the selected parents was germinated. One population was then propagated to create six copies (three for greenhouse evaluation and three for field evaluation), the other will be assessed in the greenhouse only.
- The mapping population was planted in August in a field in Beltsville, Maryland, in a randomized complete block experimental design with each block planted in a separate hill from the other two blocks. Progeny and parents were randomized separately for each block.
- A total of 184 primer pairs were used in parental screens. All but 58 detected polymorphisms between the selected parents.
- DNA has been extracted from the population, but only eight primer pairs have been tested so far. However, this prepares us to map sex-determination as soon as we characterize the progeny for sex.

Key findings as of 1 November 2006:

- SCAR 2, a marker associated with seasonal flowering locus in diploid *Fragaria vesca* detected a polymorphism between the parents. Primers derived from CONSTANS and Apetally 3 gene sequences did not detect polymorphism.
- The maternal parent of the mapping population and a large portion of the progeny started flowering in both the greenhouse and the field. Therefore, it may be possible to map remontancy in this population, in addition to sex determination.

GENOMIC RESOURCES TO IMPROVE FRUIT SIZE AND QUALITY IN SWEET CHERRY

Investigators: Amy Iezzoni[1], Co-PDs: Esther van der Knaap[2], Wayne Loescher[1], Dechun Wang[1]

Performing Organizations: [1]Michigan State University
[2]Ohio State University
E-mail: iezzoni@msu.edu
Telephone: 517-355-5191 x 391
Website: www.cherrygenetics.org

The goal of this project is to develop the genomic resources necessary to implement marker-assisted selection for fruit size and quality traits in cherry breeding programs. We are using a linkage map based QTL strategy followed by QTL validation and allele mining utilizing software developed to analyze genotypic and phenotypic data from multigenerational pedigrees. An important resource is our sweet cherry pseudo-testcross mapping population which consists of 574 F_1 progeny from reciprocal crosses between 'Emperor Francis' (EF) and 'New York 54' (NY). The key findings to date are summarized below:

- All 378 of the currently publicly available SSR markers mapped on the *Prunus* consensus map (TxE) were screened to determine if they could be placed on the NY and EF linkage maps. Only ~ 25% of these SSRs could be mapped as the remaining 75% either did not amplify or were not heterozygous in the parents. To date, 94 of the polymorphic markers have been scored on the set of 190 progeny to construct initial framework maps.

- The current EF map is 503 cM and consists of 8 linkage groups while the NY map is 371 cM and consists of 10 linkage groups. The average distance between marker loci for the EF and NY linkage maps are 6 cM and 8 cM, respectively. Twelve of the linkage groups have been aligned with the *Prunus* map based on shared SSR markers.

- Due to the low levels of marker transferability and polymorphism, the van der Knaap lab is designing markers based on 75 ESTs with known TxE locations that target gaps in our current maps. To date, 49 primers pairs resulted in 38 amplicons of which 22 flanked the intended intron. Fifteen of the 38 primer pairs identified 2 alleles; however, only 9 of these markers were heterozygous within either NY or EF and could be placed on the linkage maps.

- The van der Knaap lab developed SSR primer pairs for one cherry vacuolar invertase gene and two sorbitol transporter genes. These three genes were placed on the linkage maps.

- The Iezzoni lab evaluated fruit from 300 NY x EF progeny (5 samples per progeny individual) for fruit weight, length, and width, skin and flesh color. Cell size and number counts from fruit mesocarp sections will be completed by April 2007.

- The Wang lab identified two highly significant QTL for fruit weight. A QTL on NY LG 6 explained 20% of the phenotypic variance (LOD score = 4.7, $P<0.003$). A second fruit weight QTL on NY LG Y explained 10% of the phenotypic variance (LOD score = 4.5, $P<0.003$). If these QTL are verified in subsequent analyses they will be targets for fine mapping.

- The Loescher lab has analyzed fruit from 113 NY and EF progeny (5 fruit per progeny individual), for sorbitol, glucose, fructose and malic acid.

- Marco Bink, Biometris, NL, collaborates on the use of their pedigree based software that was developed for the HIDRAS project. Our cherry pedigree set, consisting of 41 cultivars representing up to four familial generations, has been genotyped for 51 of the mapped SSR markers and scored for the fruit weight, width, length, sorbitol, glucose, fructose and malic acid.

- The project website (www.cherrygenetics.org) has been constructed and is linked to the Rosaceae data base, www.rosaceae.org. The linkage maps, comparative maps and QTL results, will be publicly available upon completion of the analyses.

DEVELOPMENT OF SEGREGATING POPULATIONS FOR MOLECULAR AND GENETIC ANALYSES OF X-DISEASE RESISTANCE IN CHOKECHERRY (*PRUNUS VIRGINIANA* L.)

Investigators: PD: Wenhao Dai[1], Co-PD: James A. Walla[2]

Performing Organization: [1] Department of Plant Sciences
North Dakota State University
Fargo, ND, 58105, USA
Email: wenhao.dai@ndsu.edu
Telephone: 701-231-8473
[2] Department of Plant Pathology
North Dakota State University
Fargo, ND, 58105
j.walla@ndsu.edu
Telephone: 701-231-7069

The goal of this project is to develop and evaluate segregating populations of X-disease resistance in chokecherry and develop strategies for identification of molecular markers assisting breeding/selection of stone fruit trees and other related species resistant to X-disease or other phytoplasma-associated diseases. The project progress in this period (04/15/2005 – present) is summarized below.

- Three segregating populations have been developed by hybridizing three resistant and one susceptible chokecherry lines selected from a large and well-evaluated chokecherry germplasm collection. Total 909 hybrid seeds were obtained.
- These seeds have been sowed out after two months stratification and seedlings are grown in the greenhouse.
- Selection and confirmation of X-disease phytoplasma sources by stem grafting in the green house, which will be used to inoculate chokecherry hybrid seedlings.
- Chromosome analysis confirmed that *Prunus virginiana* is a tetraploid and will be a unique *Prunus* species for genetic research.
- Molecular marker analysis for chokecherry species, such as RAPD and SSR, have been attempted to develop reliable strategies of identifying molecular markers linked to X-disease resistance.
- A linkage map of chokecherry is being constructed using TRAP marker and other marker systems.
- This project is a collaborative effort in the multidisciplinary research field including plant molecular biology, plant pathology, plant genetics, and biotechnology. Both undergraduate and graduate students are involved in this research.

CANDIDATE GENES FOR FRUIT SOFTENING IN *PRUNUS*

Investigators: PD: Cameron Peace, Co-PDs: Ann Callahan, Carlos Crisosto, and Thomas Gradziel

Performing Organizations: Department of Horticulture and Landscape Architecture, Washington State University
Department of Plant Sciences, University of California, Davis
USDA-ARS Appalachian Fruit Research Station, Kearneysville
E-mail: cpeace@wsu.edu
Website: http://www.mainlab.clemson.edu/gdr/community/funding/peace.shtml

The project is focused on identifying genes involved in fruit softening of *Prunus* crops, which include peach, nectarine, apricot, plum, and cherry, and the nut crop, almond. One particular gene, coding for the cell wall-modifying enzyme, endopolygalacturonase (endoPG), is known to have major effects on certain softening traits in peach and nectarine. EndoPG is therefore an excellent target for candidate gene analysis of fruit softening in *Prunus*, and comprehensive molecular and physiological analyses are underway to elucidate its role. In addition, other candidate genes are being examined to identify their roles in softening and their interaction with endoPG. We will provide molecular tools to facilitate development of new fruit quality phenotypes of important *Prunus* crops. The project involves collaboration between the fruit postharvest physiology lab of Carlos Crisosto, the peach and almond breeding programs of Tom Gradziel, the fruit molecular genomics lab of Ann Callahan, and numerous other domestic and international collaborators, tied together through the fruit molecular genetics lab of Cameron Peace. Each group involves training of students and/or postdocs. Progress to date includes the following activities and results:

- In the Peace lab, complete association of endoPG with *Freestone* and *Melting flesh* in peach was confirmed, after PCR screening of more than 300 progeny from a single cross and dozens of progeny from other crosses provided by peach breeders and researchers Marisa Badenes of IVIA, Spain, and Dick Okie of USDA-ARS Georgia. Collaborators with segregating populations in Clemson University, France, and Italy also confirmed complete co-segregation.

- In the Peace lab, more than 100 peach and nectarine varieties were assessed for their endoPG genotype, revealing their underlying fruit type. In all cases where fruit phenotype was already known, endoPG genotype matched correctly. EndoPG genotyping also resolved many ambiguous cases, particularly for early season and “semi-” stone adhesion types. The endoPG test is now being used in several peach breeding programs to determine the true fruit type of advanced breeding lines.

- Using a microsatellite associated with the gene, twelve alleles were identified in peach and nectarine, each falling into one of the major functional categories - freestone melting flesh, clingstone melting flesh, and clingstone non-melting flesh. To examine quantitative effects that these minor alleles may have on fruit softening characteristics, changes in firmness and stone adhesion were comprehensively monitored during ripening for more than 60 peach and nectarine varieties in the 2005 and 2006 summer seasons in the Crisosto lab. Data analysis is currently underway.

- A survey of endoPG allelic diversity in approximately 650 accessions of *Prunus*, involving the Peace lab and Gradziel breeding program, together with numerous collaborators and germplasm providers, revealed more than 200 different alleles and gene copies of endoPG. As *Prunus* species are able to hybridize, this allelic diversity signifies enormous potential for breeding new flesh softening and stone adhesion attributes in stone fruit. Sweet cherry was the only species that did not show abundant polymorphism for endoPG. Even almond had considerable diversity for this gene, despite its lack of a fleshy fruit. This DNA test involving a functional gene illustrates a means of screening for useful genetic variation in germplasm collections, and in a collaboration with Malli Aradhya of the USDA National Clonal Germplasm Collection, Davis, the endoPG locus was included in a marker survey of over 400 accessions covering 24 plum and apricot species to discover approximately 120 of the alleles and gene copies.

- Dan Potter of UC Davis has begun phylogenetic analysis of the endoPG locus in *Prunus* and Rosaceae. Sequencing of cloned fragments from representative accessions is underway. So far this has revealed that only the peach lineage (*P. persica* and just a few related species) seems to have the gene duplication that provides the separate *Freestone* and *Melting flesh* genes. Nevertheless, the

  presumed ancestral "wild type" allele, which confers the clingstone melting flesh fruit type, persists in cultivated and wild peach at high frequency.

- The Callahan lab has undertaken major efforts to isolate and sequence the entire endoPG locus. Re-screening of two Clemson peach BAC libraries specifically for this gene by collaborators Tatyana Zhebentyayeva and Bert Abbott at Clemson University identified many BACs containing sequence with significant homology to the endoPG gene. From fingerprinting, subcloning, and sequencing of these BACs, the basic organization of the locus is emerging. There are at least two copies of the gene at the locus for the dominant F allele of peach, as expected, together with several adjacent endoPG-like genes. Other alleles are missing one or both copies.

- During fruit firmness measurements in 2006, tissue samples were collected, frozen, and disseminated by the Crisosto lab for future investigations of gene expression (Callahan lab), enzyme activity (Crisosto lab with John Labavitch of UC Davis), and metabolite levels (with John Labavitch). Preparation of these samples and protocol optimization has begun in the respective labs.

- A collaborative effort, involving more than a dozen labs and coordinated by the Peace lab, has been initiated to comparatively map the endoPG locus across *Prunus*. Several labs working on *Malus* (apple) are also involved.

- To discover further genes that may affect fruit softening in *Prunus*, 40 additional candidate genes for softening were screened via PCR across *Prunus* accessions in the Peace lab. Candidate genes were also located on the *Prunus* reference map, mostly using the innovative bin-mapping approach (developed for *Prunus* by Werner Howad and Pere Arus of IRTA, Spain), to create the "softening gene map" of *Prunus*. The genetic map locations of more than 50 genes putatively involved in fruit softening are now known from these efforts.

- Data from the project is beginning to be uploaded to the Genome Database for Rosaceae, where GDR collaborators Sook Jung and Dorrie Main are developing tailored template files and websites for each Rosaceae NRI project.

- Coordination of the project is moving from UC Davis to Washington State University, where Cameron Peace has taken up an Assistant Professor position as of November 2006.

Genetic Diversity of Wild Apple Accessions in the National Plant Germplasm System

Investigators: Gayle M. Volk, Co-PD: Christopher M. Richards

Performing Organization: USDA-ARS National Center for Genetic Resources Preservation
E-mail: gvolk@lamar.colostate.edu
Telephone: 970-495-3205

The goal of this project is to determine genetic relationships among accessions of wild Rosaceae species in the National Plant Germplasm System (NPGS) and provide a maximally diverse panel of genotypes that represents the diversity of the wider collection.

M. sieversii
949 samples of phenotypically characterized *M. sieversii* accessions collected from 8 locations in Kazakhstan were genotyped using 7 unlinked SSR markers. 120 alleles were scored. AMOVA results reveal a small but significant Fst (0.068), with most of the variation found within populations. Population graphs demonstrate that populations from northern locations cluster together and populations from southern locations cluster together. Bayesian cluster analyses using a Markov chain Monte Carlo optimization identified 12 stable cluster that were geographically correlated. Analyses to determine core collections for *M. sieversii* are underway.

M. orientalis

623 samples of phentypically characterized *M. orientalis* from the Caucasus region and Turkey were genotyped using 7 unlinked SSR markers. Data analyses to determine diversity and core collections are underway.

Chinese (primarily) *Malus*
M. pratii (38 accns), *M. toringoides* (142 accns), *M. hupehensis* (216 accns), *M. sieboldii* (100 accns), *M. transitoria* (36 accns), *M. kanuensis* (54 accns) were genotyped using 7 unlinked SSR markers. Many incidences of triploidy were identified. A collaboration has been set up with Gennaro Fazio and Angela Baldo (USDA-Geneva, NY) to characterize relationships among these accessions and to confirm ploidy levels.

***Malus x domestica* (non-dessert apples)**
500 NPGS + 50 Spanish hard cider *Malus x domestica*, 86 *M. sylvestris*, and 500 wild *Malus* species (core collections selected from above) are being genotyped using 20 SSR markers to identify genetic relationships among European cider types and wild species.

FUNCTIONAL GENOMIC RESPONSE OF APPLE TO FIRE BLIGHT

Investigators: Jay Norelli[1], Co-PDs: Herb Aldwinckle[2], Carole Bassett[1], and Bob Farrell[3]

Performing Organization: [1]USDA-ARS Appalachian Fruit Research Station, Kearneysville, WV; [2]Cornell University, Geneva, NY; [3]Pennsylvania State University, York, PA
E-mail: jay.norelli@ars.usda.gov
Telephone: 304-724-8340 x2142

The goal of this project is to use a functional genomic analysis to characterize the response of apple to fire blight disease and thereby identify new opportunities for improving fire blight resistance. Specifically, the project is 1) characterizing ESTs associated with apple's response to fire blight challenge, and 2) developing high-throughput methods for generating RNAi mutants in apple. The key accomplishments are summarized below:

- cDNA suppression subtractive hybridization (SSH) was applied to study the temporal progression of gene expression in susceptible ('Gala') apple in response to infection. In the early hours post-infection (1h, 2h and 24 h), host genes associated with photosynthesis and signaling, as well as a few host defense-related genes, were down-regulated. Some PR genes were up-regulated late (48h and 72 h) post-infection, including chitinase and Mal d1.
- In a second study, SSH and cDNA-AFLP analyses were used to identify genes expressed in resistant (Geneva 41) and susceptible (M.26) apple in response to infection. Two cDNA-AFLP kits (Licor and Infobiogen), which employ different restriction enzymes, were used in order to maximize discovery of EST's. Results obtained with SSH and cDNA-AFLP will be compared.
- In bioinformatics research, vector-screening tools were custom-modified to identify EST contamination. Computational methods were developed to identify apple genes expressed uniquely in disease-challenged tissue or that are similar to *Arabidopsis* ESTs associated with *Pseudomonas syringae* pv. *tomato* infection. The expression of the genes identified will also be ascertained through collaboration with existing microarray projects.
- A multi-plasmid transformation approach is being evaluated for the high-throughput generation of RNAi mutants in apple. M.26 apple tissue was transformed with three single pHellsgate8-derived plasmids and a mixture of all three. Transformation frequency was not reduced by use of multiple silencing plasmids. Universal PCR primers were developed for pHellsgate8-derived plasmids that can: 1) detect the presence of single or multiple EST silencing insertions in the RNAi transgenic, and 2) provide sequencing template to determine the EST contained in the silencing insertion.
- An assay was developed that can rapidly detect changes in fire blight resistance in apple RNAi transgenics.
- pHellsgate8-derived silencing plasmids are being constructed for 19 *Malus* ESTs identified as highly probable to be associated with response to infection.
- Two types of RNAi plant transformation vectors are being evaluated: 1) those that use GATEWAY technology to facilitate the rapid generation of hairpin RNA-encoding constructs and 2) those that use a 3' untranslated region inverted repeat to enhance sense-RNAi.
- Impact: Approximately 450 fire blight-associated *Malus* ESTs are being deposited into GenBank, 60% of the ESTs were not previously associated with fire blight resistance and 5 are new *Malus* ESTs.
- Impact, Collaboration and Community Building:

  o Collaborating with ongoing *Malus* microarray and EST projects (Tim McNellis, Pennsylvania State Univ. and Schuyler Korban, Univ. of Illinois).

  o Collaborating with Angela Baldo, USDA-ARS on bioinformatics (see bullet 3).

  o Working closely with Gennaro Fazio, USDA-ARS to facilitate vertical integration between NRI project and ARS/Cornell apple rootstock breeding program.

  o Established Non Funded Cooperative Agreement between USDA-ARS and HortResearch, New Zealand to identify molecular markers for fire blight ESTs and determine if the markers are associated with QTLs for fire blight resistance. Cornell Univ. is also involved in this research.

o Established collaboration with Rich Jorgensen, Univ. of Arizona (NSF Grant # 05018240) on development of sense-RNAi for functional genomics in *Malus*.

HIGH-RESOLUTION PHYSICAL MAPPING OF THE APPLE GENOME BY BAC FINGERPRINTING

Investigators: PD: Schuyler S. Korban, Co-PD: Jonathan E. Beever

Performing Organization: Department of Natural Resources & Environmental Sciences
University of Illinois
E-mail: korban@uiuc.edu
Telephone: 217-333-8298

The overall goal of this proposal is to develop a genome-wide physical map for the apple, *Malus* × *domestica* Borkh. The overall strategy involves fingerprinting of 10x coverage of the apple genome using BAC clones, followed by assembly of these BACs into contigs, and then anchoring of these BAC-fingerprinted clones onto a genetic map by STS content mapping. Previously, a BAC library for apple cv. GoldRush of 5x haploid genome equivalents was constructed. Since then, a second BAC library of GoldRush was constructed covering 8x haploid genome equivalents. These two libraries have been used for BAC fingerprinting. Developing a whole-genome map for the apple will play a critical role in our fundamental understanding of apple genome structure and function, allow for comparative genomics of plant chromosomes, provide a framework for understanding the biological basis of complex traits, and contribute to our fundamental knowledge on plant evolution. The key findings obtained in 2006 are summarized below:

- A total of 82,503 BAC clones (average insert size of 110 kb), derived from the two complementary BAC libraries (a *Bam*HI library and a *Hind*III library), were fingerprinted using the agarose gel-based restriction fingerprinting method. Of these clones, 8,222 (9.96%) were deleted during fingerprint editing due to either non-recombinant clones or cross-contamination between clones.
- A total of 74,281 clones, representing ~10.5× haploid genome equivalents, were successfully entered into the FPC V7.2 database for contig assembly. The average number of bands per clone was 24.9.
- To determine the appropriate cut-off value or “Sulston score”, the tolerance and cutoff values were varied, and their effects on known overlapping clones were evaluated. A tolerance of 7 and a cutoff value of 3e-9 were finally used for automatic contig assembly.
- Following the automatic assembly, 68,058 BAC clones (92%) were assembled into 3,943 contigs. The physical length of the automated contigs was estimated to be 943.8 Mb, based on 242,001 unique bands, and each band was equivalent to 3.9 kb.
- Subsequent to automated map assembly, a manual review of the assembly was conducted as it is an essential step for refining the relative order of clones within contigs, identifying joints between contigs, and disassembling larger chimeric contigs. Of 3,943 contigs, only 351 (8.9%) contained 1-4 questionable clones (more than 50% of bands were unmatched), and most of which had only one questionable clone.
- As a result, the total number of contigs of the physical map was reduced to 2,702. The assembled 2,702 contigs consisted of 237,763 unique bands and collectively spanning 927.3 Mb in physical length. Several different approaches were used to assess contig reliability.
- This draft physical map for the apple represents the first such complete map for any member of the Rosaceae family.
- In our first effort to develop molecular markers that will be used for constructing a genetic map for the apple, our apple EST database (developed from our NSF funded project) was used to develop EST-SSR markers. From a total of 122 unique apple sequences, 28 were found to produce polymorphic profiles between the two parents of a mapping population. Of these, 15 produced multiple polymorphic loci, thus culminating in a total number of 46 loci segregating in this mapping population (Naik et al., 2006). This outcome suggests that we will be able to identify 1,000 polymorphic markers that will be used for integrating genetic and physical maps.

IDENTIFYING THE GENES ASSOCIATED WITH DAY-NEUTRALITY IN STRAWBERRY USING A QTL APPROACH

Investigators: Jim Hancock, CoPIs: Kim Lewers, Steve van Nocker & Dechun Wang

Performing Organization: Department of Horticulture
Michigan State University, East Lansing, MI 48824
E-mail: hancock@msu.edu
Telephone: 517-355-5191

Our first objective was to develop a molecular map and identify QTL associated with day-neutrality in a population of 'Honeoye' × 'Tribute' after phenotyping them in multiple environments. Our intent was to begin mapping with the AFLP markers that we had already developed, but to incorporate SSRs as fast as we could identify them. We have now placed 387 single dose restriction fragments (SDRFs) on a consensus map of 1310.7 cM with 42 linkage groups. All of these are AFLPs, but we have about 60 new SSR markers that are ready to place on the map. Individuals from the mapping population were observed for their flowering habit throughout the growing season in Michigan (MI), Minnesota (MN), Maryland (MD), Oregon (OR) and California (CA). Eight QTL were found that were either location specific or shared among locations. None of these QTL explained more than 36 % of the phenotypic variation, indicating that the inheritance of day-neutrality is likely a polygenic trait. At all three eastern sites, one QTL was identified on Linkage group (LG) 17 that was a strong regulator of day-neutrality; additional QTL were identified in MI on LG 17 and in MN on LGs 7, 20 and 28. In the western states, only one significant QTL was identified on LG 7 which accounted for 22 % of the phenotypic variation in CA. This same peak was uncovered with a P value less than 0.2 in MI.

It is interesting that the QTL on LG 17 was so prominent in all three eastern states, but absent in CA and OR. Different loci may regulate day-neutrality in the various areas, as there is a strong temperature/photoperiod interaction that determines flowering in the strawberry. We speculate that the QTL on LG 17 may be required for floral initiation under the hot summer conditions found in eastern continental climates.

Our second objective was to develop a molecular map and identify QTL associated with day-neutrality in another large segregating population with the same SD parent ('Honeoye'), but a different DN *F. virginiana* parent (RH 30). We have now generated a family of 167 individuals of this cross and phenotyped them for their photoperiod sensitivity in the greenhouse. Over 80 % of the hybrids appeared to be day-neutral, compared to 50 % of the 'Tribute' x 'Honeoye' population. DNA has been extracted from this new hybrid population and we will begin mapping it this winter. We will also replicate the progeny individuals via runners and set them in the field next spring to further test their photoperiodic responses.

Our final objective was to anchor our two new octoploid linkage maps with existing maps of *F. ×ananassa* and diploid *F. vesca*. We have just begun this work with a set of 50 diploid SSR primers. So far, it appears that a high percentage of them will prove useful in mapping the octoploids.

INCREASING THE DIVERSITY OF EST SEQUENCES FOR *FRAGARIA*

Investigators: Janet Slovin[1]; Pablo Rabinowicz[2]

Performing Organizations: [1]USDA/ARS BARC-W, Beltsville, MD email: slovinj@ba.ars.usda.gov
[2]The Institute for Genomic Research, Rockville, MD email: pablo@tigr.org

This project aims to substantially increase the amount of strawberry expressed sequence tag (EST) data available to the community, and increase the diversity of EST sequences for the family Rosaceae. Currently, there are approximately 19,000 Fragaria ESTs in GenBank, 50% of which have been generated by this project. The project's goal is to produce 35-40,000 high quality EST sequences from 5 different cDNA libraries from *in vitro*- or greenhouse-grown plants treated with cold, heat, salt, drought, and a combination of salt and heat. The use of these libraries will increase the chances of capturing stress-induced genes, which are under-represented among the currently available Rosaceae EST sequences in the public databases.

We are using a diploid species of strawberry, *Fragaria vesca*, a useful model system for the family Rosaceae. It is a small plant with a small genome size (164 Mbp), a cycle of 3.5-4 months from seed to seed, and inbred lines are available. The genotype PI551574 (Hawaii 4) chosen is day-neutral, produces runners, produces abundant seed, and is easily transformed.

The first cDNA library (from cold-stressed tissue) has been constructed and 9,600 high quality sequences produced. An assembly (~2000 contigs and ~6,000 singletons) of all *F. vesca* ESTs and other cDNA sequences in Genbank is now available at http://plantta.tigr.org.

TIGR plant transcript assemblies (TAs) are built for individual species (NCBI taxon ID) and they exclude predicted transcript sequences from genome sequencing projects. They are freely available to the community through user friendly web interfaces. Each TA entry shows the component reads of the assembly and their orientation, as well as its sequence and annotation. Users can search the TAs by accession number or annotation keywords. A BLAST server that allows selecting one or more taxa, and an ftp server for data download are also available through the web pages. A TA has been constructed for *F.* x *ananassa*, with 350 contigs and 4,800 singletons, as well as for several other members of the Rosaceae.

A special TA, generated for our 9,600 *F. vesca* cold treated seedling ESTs contains a total of 5,800 assemblies and singletons. Sequences were aligned to *Arabidopsis* proteins (TAIR) using BLASTX with a cutoff E value $<10^{-5}$; alignments spanning at least 30% and at least 100 nt of the TA sequence. Over 90% of these TAs matched *Arabidopsis* proteins. A simplified version of the gene ontology annotation (GOslim) was transferred to the TAs in order to assign GO annotation to the strawberry sequences. Over 1,000 stress related sequences were identified.

F. vesca sequences produced in this project that are not represented in existing strawberry or other Rosaceae ESTs were identified by aligning our 5,800 *F. vesca* TAs to strawberry or Rosaceae unigenes in the Genome Database for Rosaceae (GDR; http://www.mainlab.clemson.edu/gdr), which does not yet include sequences from our project. Using a conservative, low stringency cutoff (BLASTN E value $<10^{-10}$), we found that ~73% of our TAs do not show similarity to other strawberry unigenes in GDR, and ~21% do not show similarity to any GDR Rosaceae unigene.

Among the 73% of our strawberry sequences not present in previous strawberry EST sets, 765 (or 13%) are associated with stress-related GO annotation. Also, 75 (or 1.3%) of those not represented in the whole Rosaceae unigene set are associated with stress-related GO annotation.

In conclusion, our project has already yielded a substantial number of new stress-related gene sequences that had not been previously identified in strawberry, and, expectedly a smaller but significant number of new stress-related genes not previously isolated in Rosaceae. We predict that additional ESTs to be produced under this project from tissues subjected to other stresses will continue to deliver new stress-related genes for strawberry and other members of the Rosaceae.

Genomic Tools and Resources for strawberry (*Fragaria*)

Presenter - Tom Davis

Progress on two strawberry genomics projects supported by the National Research Initiative (NRI) Plant Genome Program of the USDA Cooperative State Research, Education and Extension Service (CSREES) are described. The goals of these projects are to develop knowledge, tools and resources to advance the field of strawberry genomics, and to translate these resources into applications for cultivar improvement.

NRI Plant Genome Grant 2003-35300-3142: A diploid platform for strawberry genomics. PD: Thomas M. Davis (University of New Hampshire).

Under grant 2003-35300-3142, a genomic library was constructed in a fosmid vector for diploid strawberry *F. vesca* ssp. *americana* cv Pawtuckaway. This 6x coverage library consists of ~33,000 clones of ~35 kb average insert size, which have been picked and spotted in duplicate onto sets of hybridization filters. These filters have been hybridized with about 40 gene-specific probes, of which a fosmid containing the target gene was recovered in all but two cases. Three fosmids containing genes of potential horticultural significance (F3H, F3'H, ent-kaurene oxidase) were partially sequenced. Fosmid inserts comprising the chloroplast and mitochondrial genomes have been identified. Conditions for isolation of very high molecular weight DNA needed for BAC library construction have been assessed, and efforts to construct a satisfactory BAC library from 'Pawtuckaway' are continuing. A cDNA library was constructed from unopened flower buds of diploid *F. vesca* ssp. *vesca* cv 'Yellow Wonder', and about 3,300 cDNA sequences were generated, resulting in the identification of 1900 unigenes (GenBank numbers DV438013-DV440729).

NRI Plant Genome (Rosaceae) Grant (2005-35300-1546742. Gene pair haplotypes and sequence samples from strawberry (Rosaceae): multi-purpose, transferable resources for genomics and variety improvement. **PD: Thomas M. Davis (University of New Hampshire); Co-PDs: Kevin Folta (University of Florida)**, Phillip SanMiguel (Purdue University). Cooperator: Jeff Bennetzen (University of Georgia).

Under grant 2005-35300-1546742, shotgun sequencing of 31 randomly selected and 20 gene-targeted fosmid clones from the *F. vesca* ssp. *americana* cv Pawtuckaway genome library is complete. Probes for fosmid selection were based on the following genes: (flowering-related) Hy5, Constans-like, SOC, Leafy, PhyA, Pistillata, Apetala3; (fruit color and flavor) CHS, CHI, DFR, RAN (regulator of anthocyanins biosynthesis); (disease resistance) gRGA1, gRGA2, gRGA3, cRGA1, Ve-like; (other) ADH, GBSSI. Gene density in the gene-targeted fosmids was about one gene per 6 kb. Sequences of particular interest found among the random fosmids include the ribosomal RNA cluster and a transposable element family. Overall, SSR loci occurred at a frequency of about 1 per 6 kb. When combined with other results these data provide about 2 Mb of non-redundant genomic sequence from the diploid strawberry model species, *Fragaria vesca*, or about 1% of its 200 Mb genome. Gap filling and hand annotation are in progress in preparation for GenBank submission. Primer sets are being developed to amplify at least one gene pair interval from each fosmid insert, and initial results indicate that many *Fragaria* gene pair primer sets will amplify products in other Rosaceae species, including rose and cherry. At one genomic site, in-depth sequence sampling of a gene pair interval has been conducted across most of the *Fragaria* species, including three octoploid species, resulting in the identification of three discrete, genome-specific haplotypes. Other gene pairs have been amplified from diploids and octoploid cultivars to produce an accounting of alleles and a basis to test models of diploid contribution to the octoploid genome. These findings also produce a basis for the development of molecular markers that may soon be used to trace genes of interest in breeding populations.

ALGORITHM AND PROGRAMS FOR GENE EXPRESSION QTL ANALYSIS

Investigator: Zhao-Bang Zeng

Performing Organization: Bioinformatics Research Center
North Carolina University
E-mail: zeng@stat,ncsu.edu
Telephone: 919-515-1942
Website: http://statgen.ncsu.edu/zeng/

The goal of this proposal is to develop statistical methods, bioinformatics tools and computer software to perform gene expression QTL (eQTL) mapping analysis and to interpret the mapping results. Significant progress has been made in several directions.

- **Multiple interval mapping for eQTL (MIM-eQTL)**: We have developed an efficient procedure specifically for gene expression QTL analysis (Zou and Zeng 2006). The method uses our previously developed MIM to search for eQTL and uses the false discovery rate (FDR) (the estimated proportion among the declared eQTL for all expression profiles that are falsely positive) to justify the model selection procedure. In this method, we proceed to scan the genome for one or multiple QTL on each expression trait stepwisely. In each step, we make the decision whether to continue the search for more QTL or stop the process based on a criterion that can be tuned up from FDR calculation. This process is similar to that proposed by Storey et al (2005, PLoS Biology 3:e267), but with a few critical differences or improvements. (1) The search is not restricted to markers, but covers the whole genome in the fashion of interval mapping. This improvement may not be very significant for dense markers, but is a nice generalization and can be important for less dense markers. (2) Our search is not restricted to two steps, which is the case of Storey et al and can proceed to multiple steps (eQTL) as justified. (3) The search in the second and subsequent steps is restricted to those expression traits that the previous search step is significant. This is drastically different from Storey et al that advocates the two-dimensional (or potentially three or higher dimensional search) for all expression traits. In Zou and Zeng (2006), we show that our conditional search is actually much more powerful statistically than Storey et al and we found more eQTL and eQTL epistasis on the same yeast data (Brems and Krugleyak (2005 PNAS 102:1572-1577) that Storey et al also analyzed with the same FDR level.
- **QTL Cartographer for eQTL**: We are currently developing a new package based on QTL Cartographer, that implements the procedures of MIM for eQTL. The package will contain the procedures for data input and editing of large amount of marker data and gene expression data; various data checking and quality control procedures; several methods for eQTL mapping, including our MIM-eQTL. MIM-eQTL includes procedures for threshold estimation based on permutation; stepwise eQTL search based on MIM; FDR calculation for the selected eQTL; and eQTL epistasis estimation. The new package will be released at our QTL Cartographer web site http://statgen.ncsu.edu/qtlcart/index.php as soon as possible.
- **eQTL Viewer**: The eQTL analysis will produce a list of eQTL (i.e. genomic regions) for all the typed and analyzed expression traits. The genomic region for each eQTL can be defined by a 1.5 LOD-support interval calculated from MIM-eQTL, and the genes in each region can be listed if the genome is sequenced and annotated. So, essentially the final results of eQTL analysis could be summarized in a gene list for each eQTL that are matched to its expression gene. Then it would be necessary to come up an efficient and informative way to display, annotate and interpret the results. Using the Scalable Vector Graphics (SVG) technology, we have developed a very informative and useful tool, called eQTL Viewer http://statgen.ncsu.edu/eQTLViewer/, for displaying eQTL mapping results (Zou, Aylor and Zeng 2006). The tool is a dynamic database of the gene lists with a graphic 2D display with x-axis for the genome location of eQTL genes and y-axis for the genome location of expression genes. Each gene in the database can be linked to the public genome databases. The scalable feature allows us to zoom-in to look at the detail of a particular region and zoom-out to look at the overall patterns.
- **Prioritize candidate genes in eQTL regions**: There are several ways to annotate the information of eQTL mapping results. One way is to numerically prioritize the genes in each eQTL gene list as potential causal genes for the eQTL. Recently, we have attempted to develop a Bayesian algorithm (Zou 2006; W. Zou and

Z.-B. Zeng, unpublished) for such a purpose. The algorithm first uses the information from prior studies and annotations on gene relationships (such as GO classifications, KEGG relationships, chip-on-chip study information) and weighs the information to come up a raw gene-pair relationship score matrix. Then the algorithm uses this relationship score matrix as a prior and combines it with the whole gene pair

- Information from the eQTL mapping study in a repeated recursive re-weight scheme to provide the final priority scores. Information can be reinforced particularly for those eQTL that have effects on multiple expression traits. Although this information is only suggestive, it can play a very important role for a variety of applications in advancing testable hypotheses.

- **MT-MIM for eQTL mapping**: We are currently working on adapting our multiple trait multiple interval mapping (MT-MIM) for eQTL analysis (J. Maia and Z.-B. Zeng, unpublished). This is a very complex procedure and can be used for several purposes. First, it can be used to estimate the contribution of individual eQTL to genetic correlations between expression traits. Second, MT-MIM has the potential to further improve the search for eQTL for a pair or multiple expression traits simultaneously. The main research is still in progress.

- **Using the QTL shielding test (QST) to infer genetic pathways**: For the potential pathway inference, we recently worked out a QTL shielding test (QST) that tries to infer whether the relationship between a QTL and multiple (expression and/or trait) phenotypes can be described by a pathway network or a star network (C. Woods and Z.-B. Zeng, unpublished.). The test focuses on a particular QTL at a time, say Q, and two or more expression and trait phenotypes that share the same QTL, say Y1 and Y2, to see whether Y1 can shield the effect of Q on Y2, i.e. whether Y1 is in the pathway from Q to Y2. We studied various statistical issues for the test, and showed that the test works very effectively for one Q and two Y's. Currently we are working on extending the test and analysis to multiple Q's and Y's. This is a very promising approach to robustly infer sub-networks that have relatively strong causal relationships and pathway structures for eQTL and their target traits.

We also made progress in developing new theory for modeling QTL with epitasis and linkage disequilibrium (Zeng et al 2005; Wang and Zeng 2006) and a multiple interval mapping method for categorical trait (CT-MIM) (Li and Zeng 2006). CT-MIM has been implemented and released in the current version of Windows QTL Cartographer (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm).