Development of Genomic Tools for Almond Rootstock Improvement

Project No.: **11-HORT16-Aradhya/Ledbetter**

Project Leaders: Malli Aradhya
National Clonal Germplasm Repository (NCGR)
USDA-ARS
UC Davis
One Shields Avenue
Davis, CA 95616
530.752.9392
aradhya@ucdavis.edu

Craig A Ledbetter
Crop Diseases, Pests, and Genetics
USDA-ARS
9611 S. Riverbend Ave.
Parlier, CA, 93648
559.596.2817
craig.ledetter@ars.usda.gov

Project Cooperators and Personnel:
Gregory Browne and Daniel Kluepfel, Crops Pathology and Genetics, USDA-ARS
Michael McKenry, UC Riverside & Kearney Ag Center (*Emeritus*)
John Preece, NCGR Davis, USDA-ARS
Clay Weeks, NCGR Davis, USDA-ARS (*Retired*)

Objectives:

This project is part of a larger project focused on all aspects of rootstock breeding in almond and walnut funded by a USDA-CDFA Specialty Crop Block Grant. Specific of this portion of the work are:

1. Develop a set of molecular markers linked to disease-pest resistance for use in rootstock breeding programs
   a. Validate single nucleotide polymorphisms (SNPs) of peach using existing DNA sequence databases.
   b. SNP discovery in root-specific genes expressed in commercial rootstock materials and genomic sequences from a diverse set of peach, almond and wild species
2. Molecular characterization of a genetically diverse collection of commercially available rootstocks and mapping populations of newly derived interspecific hybrids.

Interpretive Summary:

The California Almond industry has identified rootstock development, testing, and commercialization, as a major priority to deal with serious soil borne pests and pathogens, including the complex replant syndrome. Currently soil fumigation is being phased out and reliance on rootstocks with field resistance to soil borne pests and diseases is increasing. Though widely used rootstocks (e.g., 'Nemaguard', other peach and peach x almond hybrids)

resist the attack of root knot nematodes, they are susceptible to other soil borne pests and diseases, such as lesion and ring nematodes, bacterial canker, crown gall, *Phytophthora* and *Armillaria*.

Countering soil borne disease-pest pressures on almond rootstocks requires identifying novel sources of resistance for incorporating into current and future rootstocks through breeding programs. Extensive disease evaluation combined with molecular characterization of germplasm and interspecific hybrids is critical to establish the genetic basis for resistance and to develop effective marker assisted selection (MAS) schemes for rapid development of rootstocks with disease resistance. Development of effective MAS strategies requires: (1) availability of marker systems for fine scale genotyping of large populations; (2) identification of molecular marker(s) that co-segregate with the disease resistance trait; and (3) reliable disease screening techniques to assay germplasm, current commercial rootstocks, and large breeding populations. Molecular markers permit rapid and accurate identification and selection of desired genes and gene combinations from among large pools of germplasm. These markers also make it possible to screen germplasm collections and large breeding populations without challenging them with pathogens; the importance of which cannot be overstated in tree breeding where one cycle of trait-based selection can take many years and is exceedingly costly.

A total of 64,851 molecular marker loci called Single Nucleotide Polymorphisms (SNPs) from publicly available sources have been assembled (**Table 2**). Of these, many are mapping uniquely and will be useful for genotyping peach-almond hybrids.

Materials and Methods:

We are using the following **Three Steps** to develop commercially viable Almond rootstocks with resistance to one or more of the key soil borne diseases/syndromes.

Step 1. Identify disease resistant rootstock genotypes

The importance of high quality reproducible disease resistance evaluation data for each of our plant genotypes cannot be understated. Using support from other sources we are screening commercially available rootstocks for disease resistance including replant disorder. To enhance our success in identifying genetic loci which mediate disease resistance in currently available rootstocks and future breeding efforts to incorporate disease resistance genes into new rootstocks we are generating new *Prunus* hybrids with broad genetic backgrounds, which we hypothesize contain novel genetic combinations imparting resistance to multiple diseases. These newly developed interspecific hybrids will be clonally propagated and subjected to disease evaluation by following standard replicated statistical designs.

To generate these new *Prunus* hybrids a large number of interspecific crosses were made at the USDA-ARS National Clonal Germplasm Repository during the spring of 2012, primarily consisting of peach x wild almond species (**Appendix Table A1**). Crosses were made on 35 trees representing 14 different species combinations.

Step 2. Genetic mapping of genes which mediate disease resistance

While disease resistance evaluations are being conducted as described above we are laying the molecular genetics ground work for *Prunus*, which will allow us to identifying candidate disease resistance genes by mapping further allowing us to rapidly select disease resistant rootstock genotypes.

The advent of SNP technology has revolutionized our ability to characterize genetic loci involved in such characteristics as disease resistance. Association studies can now be performed where SNPs associated with potential genes controlling complex disease resistance traits can be identified and tagged. This approach to complex trait dissection has been widely applied in forest trees and in woody perennial tree crop species. The association genetic approach has been used to find SNPs associated with a broad array of quantitative traits of interest (wood properties, fruit/nut quality, growth, abiotic stresses and disease resistance). SNP-genotyping technology is now being used in the breeding of most major field crops such as corn, wheat, tomato, potato, and many more, and to a limited extent in the genetic improvement of woody perennial and tree crops.

Here we aim to identify and validate SNPs throughout the *Prunus* genome and use them to facilitate rapid selection of commercially viable rootstocks and evaluate current commercially available rootstocks. We will search for and validate a set functional SNPs (SNPs located in genes located in genes affecting disease resistance) using three genomics resources: (1) *in silico* SNPs assembled from various Rosaceae genomics databases (i.e., search through existing *Prunus* databases for SNPs); (2) search the genome sequences we assembled from a dozen selected *Prunus* species (wild almond, peach and plum species; **Table 1**) used in hybrid production that are potential donors of resistance to soil borne diseases; and (3) expressed sequence tags (ESTs; sequences derived from genes expressed by the plant) collected from the root tissues of resistant and susceptible rootstocks). DNA sequences from all these sources are being generated and will be assembled to identify sequence variations that are common to all the above sources. In the end, this will result in a group of SNPs that are found in root-specific genes associated with disease resistance. All current industry standard rootstocks and newly developed hybrids will be genotyped for these SNP loci.

Table 1. Diverse genotypes used for genome resequencing and SNP discovery.

Accession ID	Species	Type	Source
Unk.	*P. arabica*	Almond	USDA Parlier
DPRU 0194	*P. argentea*	Almond	USDA NCGRU
DPRU 0192.03	*P. bucharica*	Almond	USDA NCGRU
DPRU 0201	*P. dulcis 'Vesta'*	Almond	USDA NCGRU
DPRU 2578.02	*P. dulcis*	Almond	USDA NCGRU
Unk.	*P. fenzliana*	Almond	UCD Plant Sciences
DPRU 1467.09	*P. kuramica*	Almond	USDA NCGRU
DRPU 2327.16	*P. tangutica*	Almond	USDA NCGRU
DPRU 2493.07	*P. davidiana*	Peach	USDA NCGRU
DPRU 0582	*P. kansuensis*	Peach	USDA NCGRU
DPRU 2464	*P. salicina*	Plum	USDA NCGRU
DPRU 2316.12	*P. tomentosa*	Plum	USDA NCGRU

Genomic, EST, and SNP-containing sequences, derived primarily from peach with a small fraction of almond sequences, were downloaded from GDR (GDR; http://www.rosaceae.org/), ESTree (Lazzari et al. 2008; http://www.itb.cnr.it/estree/), and NCBI (NCBI; http://www.ncbi.nlm.nih.gov/) public databases. These public databases contain a variety of sequence types from genomic and EST contigs to sequence read archives from high throughput parallel sequencing systems such as Illumina, 454, and SoLiD. Putative peach SNPs have been bioinformatically identified prior to deposition in the databases and mapped to contigs available at the time.

Table 2. Putative SNP-containing sequences assembled from various *Prunus* databases.

Database	Crop	SNPs
ESTree[1]	Peach/almond	17,291
GDR[2]	Peach	40,794
NCBI[3]	Peach	6,657
NCBI[3]	Almond	109
Total		64,851

[1]http://www.itb.cnr.it/estree/
[2]http://www.rosaceae.org/species/prunus
[3]http://www.ncbi.nlm.nih.gov/

In silico discovered SNPs downloaded from public databases (**Table 2**) were aligned against the 'Lovell' draft peach genome available from GDR using Bowtie2 (Langmead and Salzberg, 2012), a fast reference-guided sequence aligner capable of handling both short and long reads in either end-to-end or local alignment modes. The reference genome was indexed using bowtie2-build tool with the default parameter settings. Alignments were then performed using bowtie2 in the local alignment mode (--local) with the alignment score settings of +5 for each matching nucleotide (--ma), -4 mismatch penalty (--mp), 0 penalty for nucleotides of N in either the read or reference (--np), and -5 for the read gap open penalty while the read gap extension penalty was varied depending upon the sequence source, -2 for ESTs or EST only derived SNPs, such as those from ESTree, and the default of -3 for all other sequences (--rdg). The alignment is output as a SAM file (.sam), a standard bioinformatics flat file format, filtered for the best alignment, containing the sequence identification, alignment start position, and quality for each query sequence.

Downloaded data was also aligned to the 'Lovell" draft peach genome using Mosaik (version 1.1.0014, Stromberg, Lee, and Marth, Boston College Biology Department). The algorithm used was for all positions at all possible location, with a maximum mismatch percent of 28%, hash size of 14, alignment candidate threshold of 20 bp, and hash position threshold of 100. Like with Bowtie2, gap open and extension penalties varied depending on the sequence source, SNP or EST; the gap open penalty used with EST sequences was 5 and the extension penalty was 2, while the default values were used for SNP sequences. The output files were then converted to SAM format for downstream use.

The SAM file was then converted to binary format (.bam), sorted, and analyzed for high quality variants using samtools (Li et al., 2009). The alignments and called variants were loaded into the integrated genome viewer (IGV; Robinson et al., 2011) to visually compare alignments of differently sources SNPs, and look for overlapping peach and almond SNPs. However, with 100 almond SNPs and nearly 41,000 peach SNPs there are few locations of overlap.

The SNPs will be *in silico* validated by aligning the SNP sequences with genomic resequencing data from a diverse panel of peach and almond genotypes (**Table 2**) pending receipt from Beijing Genome Institute (BGI), in partnership with UC Davis. They will also be further validated by using (1) primers designed to flanking regions of the SNPs and sequences aligned with those predicted and position of SNP checked and recorded; or (2) High-throughput SNaPshot multiplex system (Applied Biosystems) followed by capillary electrophoresis for SNP visualization.

Step 3. Identify SNPs associated with resistance.

Using our molecular data developed in Step 2 mentioned above, we will identify SNP markers associated with disease resistance/susceptibility. Once confirmed, these SNPS will be used to develop juvenile selection strategies for rootstock improvement. We anticipate performing Step 3 activities towards the end of this funding cycle.

Research Activities Planned for 2012/2013

Assemble genome sequences from diverse Prunus species and root-specific cDNA sequences using the peach draft genome as a reference. In addition, assemble the sequences against a consensus sequence of the peach draft genome and peach genome sequences available from the NCBI Sequence Read Archive (SRA). Once assembled, call SNPs based on the alignments, compare to existing and *in silico* SNPs, and output SNPs shared between species and SNPs specific to each species. Genotype all interspecific hybrid individuals and commercially available rootstocks, and perform association mapping of pathology data and genotypes to find candidate disease resistance genes.

Results and Discussion:

Prunus hybrids

From 35 mother trees, representing 14 different species combinations, 20 trees produced fruit (**Appendix Table A1**). Four hundred and forty-six (446) immature fruit were collected and delivered to California Seed and Plant Laboratory (CSPL) in Elverta, California, in late May 2012. Of the 446 immature fruit delivered, 371 potentially viable embryos were placed into culture at CSPL where they will be multiplied via micropropagation and then used for pathology screening. Additionally, interspecific hybrids produced in 2010 and 2011 are still undergoing multiplication, but 468 plants representing 23 genotypes from four confirmed crosses were received mid-summer of 2012 (**Appendix Figure A1**).

Molecular Data Development

A total of 64,851 SNP-containing DNA sequences from publicly available sources have been assembled (**Table 2**). Of the total, 17,291 are from peach and almond from the ESTree database, 40,794 are from peach from GDR, and 6,766 (109 almond and 6,657 peach) from NCBI. The data is being evaluated for duplicate SNPs, which will reduce the total number of

putative SNPs to between 40,794 and 64,851. This suggests that many SNP containing sequences in the database may be originating from duplicate regions of the genome. However, uniquely mapping SNP-containing sequences will likely be useful for genotyping peach-almond hybrids while overlapping peach and almond SNPs suggest immediate genotyping utility for these hybrid types. We are also in the process of mapping peach and almond EST sequences (**Table 3**) to identify new SNPs.

Table 3. SNP and EST sequences available from the NCBI database for almond, peach, and totals for all *Prunus* species. (SNP, UniGene, and SRA included for comparison.)

Resources (NCBI)	P. dulcis	P. persica	All Prunus
Single Nucleotide Polymorphisms (SNPs) (includes 9 SSRs and indels)	109	6 657	9 009
Expressed Sequence Tags (ESTs)	3 864	79 824	110 815
Unique Genes (UniGenes)	0	7 481	N/A
Sequence Read Archives (SRAs) (e.g., Illumina, 454, and SoLiD sequencing experiments)	0	15	33

Until genome sequences are received from BGI, analysis has consisted of aligning publicly available data from almond (*Prunus dulcis*), such as ESTs and SNP-containing sequences, and publicly available SNP-containing sequence data from ESTree, GDR, and NCBI to the draft genome of peach (*Prunus persica*), a summary of results from the aligners Bowtie2 and Mosaik are presented in **Tables 4 and 5**, respectively. Overall, this analysis supports the known sequence homology of the sister species *P. dulcis* and *P. persica*, suggesting that genome sequences of *P. dulcis* and other wild almond species will also have homology with *P. persica* supporting its use as the reference for genome sequence assembly, and SNP discovery.

Table 4. Results of Bowtie2 reference-guided alignment of publicly available almond and peach sequences.

Alignment	Almond SNPs*[1]	Peach SNPs[1]	Peach SNPs[2]	Peach & Almond SNPs[3]	Almond ESTs[1,2]	Almond ESTs[1]
No. of reads	100	6 657	40 794	17 291	6 658	3 864
Did not align	10 (10.00%)	0 (0.00%)	5 (0.01%)	5 (0.03%)	13 (0.20%)	9 (0.23%)
Aligned uniquely	74 (74.00%)	4 709 (70.74%)	18 083 (44.33%)	2 330 (13.48%)	1 137 (17.08%)	664 (17.18%)
Aligned non-uniquely	3463 (39.76%)	1948 (29.26%)	22 706 (55.66%)	14 956 (86.50%)	5 508 (82.73%)	3191 (82.58%)
Alignment rate	98.93%	100.00%	99.99%	99.97%	99.80%	99.77%

*Nine non-SNP sequences, such as insertion/deletions or simple sequence repeats, were removed prior to alignment.
[1] Source NCBI, [2] Source GDR, [3] Source ESTree

Table 5. Results of Mosaik reference-guided alignment of publicly available almond and peach sequences.

Alignment	ESTree SNPs[3]	Almond ESTs[2]	Almond ESTs[1]	Almond SNPs*[1]
Number of reads	17 069	2 749	3 813	100
Filtered out	7 034 (41.2%)	844 (30.7%)	1 231 (32.3%)	5 (5.0%)
Unique	7 986 (46.8%)	1 821 (66.2%)	2 359 (61.9%)	90 (90.0%)
Non unique	2 049 (12.0%)	84 (3.1%)	223 (5.8%)	5 (5.0%)
Total aligned	10 035 (58.8%)	1 905 (69.3%)	2 582 (67.7%)	95 (95.0%)
Non unique alignments		1 195 (39.6%)	4 916 (67.6%)	29 (24.4%)
Total alignments		3 016	7 275	119

*Nine non-SNP sequences, such as insertion/deletions or simple sequence repeats, were removed prior to alignment.
[1] Source NCBI, [2] Source GDR, [3] Source ESTree

Currently we are sequencing the complete genomes of a diverse set of peach, almond, and wild *Prunus* spp genotypes (**Table 1**) used in the interspecific crosses to an average of 30X coverage (or 30 sequence reads representing each nucleotide) to use in the SNP discovery process. The sequencing depth will allow us to confidently identify heterozygous nucleotide positions in the genome from sequencing errors. Using bioinformatic tools, the genomic DNA sequences generated from these diverse genotypes will be aligned and searched for the presence of SNPs. Additionally, SNPs in **Table 2** will be compared to the genomic sequence alignments for validation and evaluation of usefulness for this project.

The final step is to further narrow the number of SNPs to those most usable for genotyping current and potential almond rootstock materials. We anticipate this step will reduce the number of SNPs by 97% down to approximately 2,000 SNPs. Rootstock genotypes that show contrasting disease reactions (i.e., susceptible vs. resistance) to target pathogens will be examined to identify those genes differentially expressed in the roots of susceptible versus resistant rootstocks. Once these genes are identified we will further narrow our list of SNPs to those that occur in these genes, which we hypothesize will play a large role in disease resistance. This will result in identification of useful SNPs linked to genes that confer disease resistance. These SNPs along with others identified using genomic sequences will allow us to carefully map resistance genes. Candidate rootstocks under consideration for use in identification of differentially expressed root specific are listed in **Table 6** along with their known disease resistance evaluations from various sources. Sequencing of both genomic and cDNA libraries will be completed by late 2012.

Table 6. Rootstocks used for root-specific expression profiling.

Trait	Lovell	Nemaguard	Hansen 536	Nickels
Soil Type Adaptability	Well drained soils	Well drained, sandy-loam soils	Well drained soils, drought tolerant, calcareous soils ok	Better adaptability than Hansen 536
Wet Soil	Moderately susceptible	Very susceptible	Very susceptible	Moderately susceptible
Anchorage	Good	Good	Very good	Very good
Vigor	Moderate	High	Very high	Very high
Root Suckers	None	None	None	None
Phytophthora Resistance	Low	Low	Very Low	Very Low
Bacterial Canker Resistance	High	Low	Unknown	Unknown
Crown Gall Resistance	Fair	Fair to Good	Fair	Fair
Root Knot Nematode Resistance	Low	High	High	High
Oak Root Fungus Resistance	Low	Low	Low	Low

In addition to developing SNP genotyping platform for genotyping currently available and newly produced interspecific rootstocks, we will be using genotyping-by-sequencing approach (Elshire et al., 2011) to genotype all the existing and new rootstocks. This approach is simple, extremely specific and highly reproducible with excellent genome coverage. This is an excellent altyernative to SNP genotyping using traditional SNP genotyping platform.

Research Effort Recent Publications:

None at this time.

References Cited:

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell (2011). A robust, simple genotyping-by-sequencing approach for high diversity species. PLoS ONE 6:1-10.

Jung, S., Staton, M., Lee, T., Blenda, A., Svancara, R., Abbott, A. and Main, D. (2008). GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. Nucleic Acids Research. 36: D1034–D1040.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.
Lazzari, B., Caprera, A., Vecchietti, A., Merell, I., Barale, F., Milanesi, L., Stella, A. and Pozzi C. (2008). Version VI of the ESTree db: an improved tool for peach transcriptome analysis. BMC Bioinformatics. 26;9 Suppl 2:S9. PMID: 18387211
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9.
National Center for Biotechnology Information (NCBI) http://www.ncbi.nlm.nih.gov
Mosaik 1.1.0014 Strömberg, M., Lee, W.-P., and Marth, G http://code.google.com/p/mosaik-aligner/downloads/list
Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. Nature Biotechnology 29, 24–26.

APPENDIX

Table A1. Interspecific hybrid crosses made in 2012 and culture status.

Mother Tree (all *P. persica*)	Pollen Parent		Fruit collected	Embryos cultured
DPRU 1576. C Golden Queen Improved	*P. arabica*	USDA Kearney (C. Ledbetter)	1	1
DPRU 2546.3 A Sanguine de Chateauneuf (sdlg)	*P. argentea* F8 15-25	DPRU 0194	47	45
DPRU 2499.5 A Jing Yu (sdlg)	*P. bucharica* USSR 90-09-01 (sdlg)	DPRU 1871.1	5	5
DPRU 2499.1 A Jing Yu (sdlg)	*P. bucharica* (sdlg)	DPRU 0192.3	4	3
DPRU 2466.17 A Kawanakajima hakutou (sdlg)	*P. davidiana* Shanxi (sdlg)	DPRU 2493.X	12	10
DPRU 2151. B Indian Blood Cling	*P. dulcis* #53 (Ukraine) (sdlg)	DPRU 2578.2 A	2	1
DPRU 2544. A IC 209015	*P. dulcis* #53 (Ukraine) (sdlg)	DPRU 2578.2 A	37	23
DPRU 2151. A Indian Blood Cling	*P. fenzliana*	UCD Plant Sciences (T. Gradziel)	4	3
DPRU 1602. C (Chinese x cultivar) x OP	*P. fenzliana*	UCD Plant Sciences (T. Gradziel)	2	1
DPRU 2466.12 A Kawanakajima hakutou (sdlg)	P. hybrid [*P. dulcis* 'Mission' x *P. scoparia*] x OP (sdlg)	DPRU 2915.15	2	2
DPRU 2631. B Rio Oso Gem	*P. kansuensis*	DPRU 582	12	6
DPRU 2655.1 A Qing Zhou Hong (sdlg)	*P. kuramica* (sdlg)	DPRU 1467.9	3	3
DPRU 2499.4 A Jing Yu (sdlg)	*P. kuramica* (sdlg)	DPRU 1467.9	4	3
DPRU 2654.1 A Qiu Xiang Mi (sdlg)	*P. salicina* Korean wild plum – red	DPRU 2464 A	14	9
DPRU 2651.2 A Okayama #11 (sdlg)	*P. tangutica* (sdlg)	DPRU 2327.x	251	234
DPRU 2631. A Rio Oso Gem	*P. tomentosa* 96001 (sdlg)	DPRU 2316.12	29	12
DPRU 2656.2 A 12-28 (sdlg)	*P. tomentosa* 96001 (sdlg)	DPRU 2316.12	3	2
DPRU 2499.6 A Jing Yu (sdlg)	plum hybrid Marianna 26-24	UCD Foundation Plant	4	4
DPRU 2651.1 A Okayama #11 (sdlg)	plumcot NJ-PC 1	DPRU 1063	4	4
DPRU 2629. B Hale Early	plumcot NJ-PC 1	DPRU 1063	6	3
		Total	446	371

Figure A1. Embryo-rescued *Prunus* Interspecific hybrids multiplied by California Seed and Plant Laboratory on a mist bench at the USDA-ARS repository.