FINAL PROJECT REPORT

Project Title: Development of marker-based breeding technologies for pear improvement

PI: David Neale
Organization: UC Davis
Telephone: 530-754-8431
Email: dbneale@ucdavis.edu
Address: One Shields Avenue
Address 2: 262 Robbins Hall
City/State/Zip: Davis
State/Zip: CA 95616

Co-PI (2): Amit Dhingra
Organization: Washington State University
Telephone: 509-335-3625
Email: adhingra@wsu.edu
Address: Dept. of Horticulture
Address 2: 155 Johnson Hall
City/State/Zip: Pullman
State/Zip: WA 99164

Cooperators: Richard Bell (USDA/ARS Kearneysville, WV), Joseph Postman and Nahla Bassil (USDA/ARS Corvallis, OR), Kate Evans (Washington State University), Sara Montanari (UC Davis), Rachel Elkins (UC Cooperative Extension)

Other funding sources: none
Agency Name: UC Davis
Amount awarded: $100,000

Total Project Funding: $100,000

Budget History:

Item	2014:	2015	2016
Salaries			
Benefits			
Wages			
Benefits			
Equipment			
Supplies	$50,000	$50,000	$0, NCE
Travel			
Plot Fees			
Miscellaneous			
Total	50,000	50,000	0

OBJECTIVES

Pear production can be increased by developing new varieties with improved agronomic characteristics, such as disease/insect resistances and dwarfing stature, which can be combined with high fruit quality and many other traits. In traditional breeding the selection of such elite cultivars is based on the visual evaluation of phenotype, and in woody perennial crops, including pear, this process is time consuming and expensive, because of the trees' long juvenile phase, laborious trait assessment, and large land requirement. Marker-assisted selection (MAS) technologies are currently routinely and successfully applied for several plant crops, and they can potentially increase pear breeding efficacy. The objective of this project was to develop a high number of molecular markers to be used to screen ~2000 different pear cultivars collected from the National Clonal Germplasm Repository (USDA/ARS NCGR) in Corvallis, OR. These genotypic data will be useful to find strong marker-trait associations to be applied for MAS in pear, information which is currently lacking for most of the traits of interest in this crop.

All U.S. pear genetics researchers have teamed up under the new Pear Genomics Research Network, and collaborations with other foreign pear scientists have also been set up, with the objective of working together towards a common goal.

Activities:

1. Design a re-sequencing project and a SNP genotyping assay (accomplished).
2. Collect leaf samples from *Pyrus* spp. accessions from the National Clonal Germplasm Repository (NGCR) in Corvallis, OR (accomplished).
3. Conduct bioinformatics analysis of the re-sequencing data and design a SNP array (accomplished).
4. Genotype all the collected samples (in progress).
5. Submit the re-sequencing and genotypic data to the Genome Database of Rosaceae (https://www.rosaceae.org/).

SIGNIFICANT FINDINGS

1. We selected 55 accessions to represent the SNP discovery panel and we extracted high quality DNA from them.
2. We collected leaf samples from ~2000 *Pyrus* spp. accessions from the National Clonal Germplasm Repository (NGCR) in Corvallis, OR.
3. We processed the 55 selected accessions for whole-genome, low-coverage sequencing (re-sequencing).
4. We performed bioinformatics analysis of the re-sequencing data, SNP calling, filtering and we designed a high-density SNP array.
5. We extracted high-quantity DNA from a subset of the collected samples for genotyping.

METHODS

Design a re-sequencing project and a SNP genotyping assay for pear

Researchers working on pear breeding and genomics in the U.S., their extension collaborators, and the pear marketing boards created the Pear Genomics Research Network (PGRN), with the aim of bringing together their efforts for the enhancement of the pear-growing industry in the U.S. Within this collaboration, we started a re-sequencing project for the evaluation of Pyrus genetic diversity. We selected 55 pear accessions, representing founding cultivars and a total of 29 species and hybrids, within the NCGR in Corvallis, OR, and the Appalachian Fruit Research Station (AFRS) in Kearneysville, WV, to constitute the polymorphism discovery panel in this project (Table 1). These accessions were processed for whole-genome, low-coverage sequencing.

Sample collections and DNA extraction

During the summer 2014 we collected leaves from 1870 different Pyrus spp. cultivars and hybrids maintained at NGCR and AFRS. For the 55 samples included in the discovery panel, we extracted DNA from freeze-dried leaves using the DNeasy Plant Mini Kit (Qiagen®). For each sample, paired-end libraries were constructed using the Nextera DNA Sample Preparation kit (Illumina®) at the UC Davis Dept. of Evolution and Ecology. Libraries were sent to the Institute for Genomic Medicine at UC San Diego for sequencing on an Illumina® HiSeq2500 in high output mode with v4 chemistry and 2x100 bps runs.

Bioinformatics analyses of re-sequencing data and SNPs calling and filtering

Sequences of the 55 different pear accessions were evaluated, and the low quality bases (usually at the boarders of the sequences) were trimmed off. Sequences from similar accessions were divided into 6 groups, as in Table 1: i) Group Communis, including all P. communis cultivars, P. communis subsp. caucasica and P. communis subsp. pyraster; ii) Group 1, including wild relatives of P. communis; iii) Group 2, including Middle East/Central Asia arid adapted species; iv) Group 3, including East Asian "pea" pears; v) Group 4, including East Asia large fruited cultivars and wild relatives; and vi) Group Hybrids, including all interspecific hybrids. The objective was to group together accessions with expected similar genomes and apply ad hoc parameters for both the sequence alignment and the SNP calling. The trimmed sequences were aligned to the 'Bartlett' v1.0 reference genome, applying more stringent parameters for the Group Communis. The aligned sequences within each group were pooled and searched for polymorphisms against the reference genome. The polymorphic sites (variants) were then subjected to a Quality filter (Fig. 1), with parameters calculated for each of the 6 groups. Afterwards, all the detected variants from each group were combined into a unique file and subjected to the Affymetrix filter (Fig. 1), aimed at discarding possible false SNPs.

SNP selection

The most informative set of SNPs was selected basing on their predicted effect on genes (according to the software SnpEff), their position on the genome (according to a Focal Point strategy), and the level of diversity across the 55 re-sequenced accessions (Fig. 1). Two different files were submitted to Affymetrix for the array design: a high priority file and a low priority file. SnpEff is a software that predict how a certain SNP, if it falls inside a coding region, might modify the protein, and it classifies the SNPs according to the impact of such a change. SNPs inside coding regions (those classified with HIGH, MODERATE and LOW effect by SnpEff), and SNPs close to coding regions (those with MODIFIER effect and not categorized as "intergenic") were given high priority for inclusion in the array. Also SNPs developed with other technologies and validated in mapping population were given high priority. These are Illumina Infinium II SNPs (Montanari et al, 2013) and SNPs developed by Genotyping-by-sequencing (GBS) at PFR.

The remaining SNPs (those with MODIFIER effect and intergenic) were given low priority for the array design, and a sorted list was submitted to Affymetrix. We divided the genome in windows of constant size, called Focal Points (FP). Of all the SNPs inside each FP, we removed those with the same genotypes (redundant information). Then we chose one SNP for each FP, the SNP with the higher number of heterozygous genotypes, and we put them at the top of the list; these were followed by the second SNPs with the higher number of heterozygous genotypes from each FP, and so on. This way, we selected SNPs that were evenly spread across the genome and more informative.

At Affymetrix, SNPs from the high priority file were tiled on the array first, then the SNP from the low priority file were selected starting from the top of the list and going down, until completion of the array.

Genotyping of the collected samples

DNA was first extracted from a subset of 284 highly diverse pear accessions (the "screening panel"). The SNPs and the DNA were sent to Affymetrix for the construction of a draft genotyping array, according to the Axiom myDesign™ protocol, and for genotyping. Basing on the results of this first round of genotyping, we will discard all non-functioning markers and the less informative SNPs. SNPs passing the "screening" step will be again sent to Affymetrix, along with the DNA of the remaining samples, for designing a final, highly-efficient SNP array and for genotyping.

RESULTS & DISCUSSION

The Pear Genomics Research Network

The University of California (UC) Davis, UC Cooperative Extension, the NGCR in Corvallis, OR, the AFRS in Kearneysville, WV, Washington State University (WSU) and Oregon State University (OSU), have teamed up under the new Pear Genomics Research Network (PGRN), which also involves the industry organizations California Pear Advisory Board (CPAB), Pear Pest Management Research Fund (PPMRF), Pear Bureau Northwest (USA Pears), and Washington Tree Fruit Research Commission (WTFRC). A website for the PGRN (http://ucanr.edu/sites/peargenomics/) was developed in March 2015.

Re-sequencing, SNP calling and selection of SNPs for first draft genotyping array

Sequencing of the 55 accessions included in the discovery panel resulted in a total of 731.2 Million read pairs, with a per sample coverage of 3.3x to 5.4x. Variants were called from each of the 6 groups and i) 3,809,750 were discovered in Group Communis; ii) 5,484,730 in Group 1; iii) 5,957,246 in Group 2; iv) 7,004,301 in Group 3; v) 7,339,331 in Group 4; and vi) 5,732,197 in Group Hybrids. After the Quality filter and combination of the variants from all the groups into a single file, a total number of 9,662,991 unique variants were left and were submitted to Affymetrix for scoring. After the Affymetrix filter, 1,195,301 SNPs were left and were analyzed with SnpEff. 85,152 SNPs (643 tri-allelic and 84,509 bi-allelic) with HIGH or MODERATE effects were all kept; SNPs with LOW effect were subjected to further filtering (Fig. 1) and 93,302 were left (461 tri-allelic and 92,841 bi-allelic); SNPs with MODIFIER effect were subjected to further filtering (Fig. 1) and 552,485 were left (6138 tri-allelic and 546,347 bi-allelic, of which 98,557 intergenic). Also validated SNPs were scored by Affymetrix and filtered (Fig. 1): 1139 Illumina Infinium II SNPs (Montanari et al., 2013), filtered down to 558, and 9151 SNPs developed by GBS at PFR, reduced to 2452. In total, 733,949 were submitted to Affymetrix and 659,183 were successfully tiled on the first draft array.

Screening panel

The 284 samples constituting the screening panel were chosen to be representative of the entire diversity held at NCGR. A total of 35 different species and interspecific hybrid were included in the screening panel. Some cultivars with known pedigree information and their two parents ("trios") were also included, for a total of 21 trios, whose genotypic information will be useful to validate the SNP markers. Moreover, three samples were replicated, in order to double check, the reliability of the genotyping and identify possible causes of errors: P. communis 'Bartlett' was replicated three times, double haploid 'Bartlett' twice and P. pyrifolia 'Dan Bae' twice.

56,700 SNPs will be chosen for the final array.

Discussion

The number of SNPs we discovered is the highest ever found for pear. By performing the screening step, we will guarantee the design of a highly-efficient SNP array, with a success rate close to 100%, which is fundamental for the evaluation of a large genetic diversity. With this genotypic data we will be able to characterize the pear germplasm collection. From these studies, we will gain information about unknown genotypes identity and pedigrees, which is fundamental for their

employment in breeding. We will also be able to elucidate the degree of relatedness among different species, and the comparison of wild species with cultivars might also help us identifying regions linked to domestication patterns, which are assumed to be associated with important agronomic features.

Moreover, we will use this genotypic information to do associations with phenotypes and identify markers to be used in MAS. First of all, historic phenotypic data collected at NCGR will be used, although they are not expected to provide highly reliable information. Secondly, appropriate phenotypic experiments will be designed for the collection of new data and the identification of robust marker-trait associations.

REFERENCES

Chagné, D., Crowhurst, R. N., Pindo, M., Thrimawithana, A., Deng, C., Ireland, H., … Velasco, R. (2014). The draft genome sequence of European pear (Pyrus communis L. "Bartlett"). *PLOS ONE*, *9*(4), 1–12. http://doi.org/10.1371/journal.pone.0092644

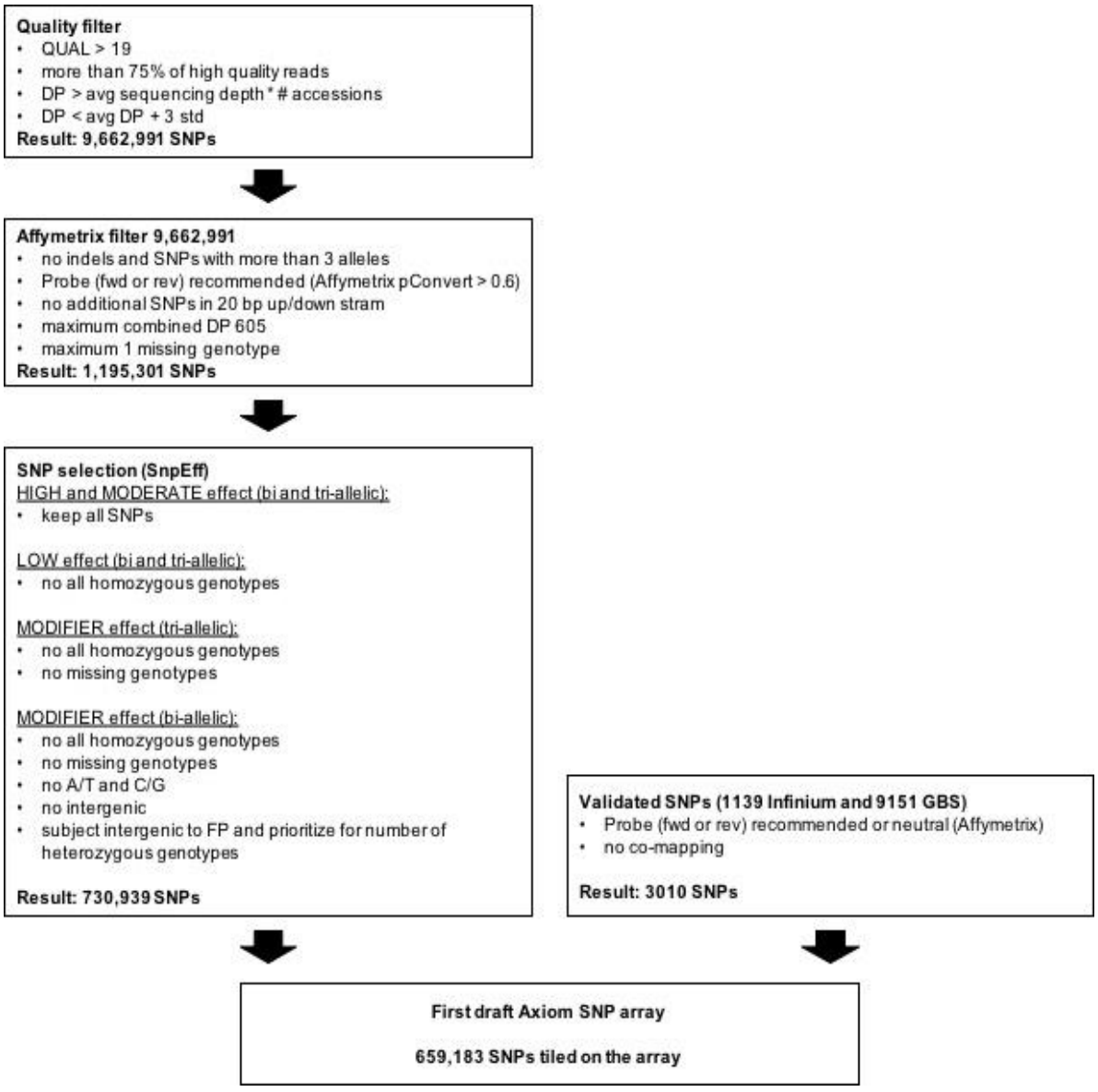
Montanari, S., Saeed, M., Knäbel, M., Kim, Y., Troggio, M., Malnoy, M., … Chagné, D. (2013). Identification of Pyrus Single Nucleotide Polymorphisms (SNPs) and evaluation for genetic mapping in European pear and interspecific Pyrus hybrids. *PLOS ONE*, *8*(10), 1–11. http://doi.org/10.1371/journal.pone.0077022

Table 1: List of 55 re-sequenced pear accessions, with subdivision into 6 groups of genetic similarity.

Pear accession	Group	Pear accession	Group
P. communis 'Anjou'	Communis	*P. elaeagrifolia* MSU6768	Group 2
P. communis 'Bartlett'	Communis	*P. glabra*	Group 2
P. communis 'Bosc'	Communis	*P. regelii*	Group 2
P. communis 'Coscia'	Communis	*P. sachokiana* GE-2006-115	Group 2
P. communis 'Gem'	Communis	*P. salicifolia* GE-2004-141	Group 2
P. communis 'Gin'	Communis	*P. spinosa* (*amygdaliformis*)	Group 2
P. communis 'Harrow Delight'	Communis	*P. syriaca*	Group 2
P. communis 'Harrow Sweet'	Communis	*P. betulifolia*	Group 3
P. communis 'Old Home'	Communis	*P. betulifolia*	Group 3
P. communis 'Para de Zahar de Bihor'	Communis	*P. fauriei*	Group 3
P. communis 'Roi Charles de Würtemburg'	Communis	*P. koehnei*	Group 3
P. communis 'Seckel'	Communis	*P.* × *bretschneider*i 'Ta Shian Sui Li'	Group 4
P. communis subsp. *caucasica*	Communis	*P.* × *bretschneideri* 'Xuehuali' (Snowflake)	Group 4
P. communis subsp. pyraster 'Erabasma'	Communis	*P.* × *bretschneideri* 'Ya Li'	Group 4
P. communis subsp. *pyraster* 'Mednik'	Communis	*P.* × *sinkiangensis* 'Ho Mon'	Group 4
P. communis subsp. *pyraster* ALB-2011-024	Communis	*P. hondoensis*	Group 4
P. communis US 309	Communis	*P. pashia* 'Naspati'	Group 4
P. communis US76128-009	Communis	*P. pseudopashia*	Group 4
P. communis US82720-002	Communis	*P. pyrifolia* 'Dan Bae' (Olympic)	Group 4
P. cordata (Turkey)	Group 1	*P. pyrifolia* 'Nijisseiki'	Group 4
P. cordata pure	Group 1	*P. pyrifolia* 'Zao Su'	Group 4
P. cossonii (Russia)	Group 1	*P. ussuriensis* 'Pai Li' (Beijing White Pear)	Group 4
P. gharbiana No. 1	Group 1	*P. ussuriensis* No. 2 (Korea)	Group 4
P. mamorensis	Group 1	*P. ussuriensis* x *P. pyrifolia* Illinois 76	Group 4
P. nivalis	Group 1		
(*P. ussuriensis* x *P. pyrifolia*) x *P. communis* NJ487601193	Hybrids	*P. communis* x *P. ussuriensis* NJB9R1T117	Hybrids
(*P. ussuriensis* x *P. pyrifolia*) x *P. communis* NJA2R59T69	Hybrids	*P. communis* x *P. ussuriensis* NY 10262	Hybrids
P. communis x *P. ussuriensis* 'Takisha'	Hybrids	*P. communis* x *P. ussuriensis* NY 10353	Hybrids

Communis = *Pyrus communis*; Group 1 = *P. communis* wild relatives; Group 2 = Middle East/Central Asia arid adapted species; Group 3 = East Asian "pea" pears; Group 4 = East Asia large fruited wild relatives; Hybrids = interspecific hybrids

Figure 1: SNP filtering pipeline.



EXECUTIVE SUMMARY

1. We designed a large number of SNP markers for pear and we included them in an array for high-throughput genotyping.
2. The genotypic data developed with this tool will be used to characterize the pear germplasm collection, evaluate *Pyrus* genetic diversity and build linkage maps for breeding populations.
3. Such studies will provide information that can be used for breeding in several ways: localize genomic regions associated with traits of interest; identify degrees of relationship among cultivars, in order to optimize their use for breeding; elucidate *Pyrus* domestication patterns, which are assumed to be associated with important agronomic features.
4. Available phenotypic data collected for the genotyped accessions will be used directly for association studies, and new phenotypic experiments will be designed for the confirmation of such associations and the study of new, important characters.
5. The final objective is to implement MAS in pear, for a faster development of new, high-performance cultivars.