APPLICATION OF MARKER BREEDING IN THE WALNUT IMPROVEMENT PROGRAM

P.J. Martínez-García, Marc Crepeau, Daniela Puiu, Daniel Gonzalez-Ibeas, Jeanne Whalen, Kristian Stevens, Timothy Butterfield, Monica Britton, Russell Reagan, Charis Cardeno, Randi Famula, Sowmya Kuruganti, Mallikarjuna Aradhya, Chuck Leslie, Abhaya Dandekar, Steven Salzberg, Jill Wegrzyn, Chuck Langley, and David Neale

ABSTRACT

The primary goal of the walnut genomics research program at UC Davis is to bring *molecular breeding* to full application in the *Walnut Improvement Program* (MB/WIP) within five years. The MB/WIP project will proceed in five phases (roughly one-year each), each phase building upon the other. This project will be fully integrated into the ongoing WIP right from the outset and Neale and WIP Project Leader Leslie will co-design the research and collaborate on every aspect. This should facilitate transition of the WIP from a classical phenotype-based breeding program to an integrated phenotype and molecular breeding program. The Walnut Genome Implementation Group (WGIG) is a continuation of a collaborative group of researchers, all working on some aspect of walnut genomics, which was initially formed by Chuck Leslie. David Neale now chairs this group that meets bi-weekly. The purpose of the group is to facilitate information exchange and foster creativity and innovation toward walnut scion and rootstock genetic improvement.

OBJECTIVES

The five different phases (objectives) are:
I. Reference genome sequencing *(Year 2014).*
II. Resequencing and discovery of genetic variation in *J. regia* and related *Juglans* species *(Year 2015).*
III. Test for marker-trait associations that will be used in molecular breeding *(Year 2016).*
IV. Genomic selection model building *(Year 2017).*
V. Genomic selection model validation *(Year 2018).*

SIGNIFICANT FINDINGS

Phase I. Deliver an annotated genome sequence for *J. regia* that will be housed in the TreeGenes database. The reference genome sequence is a necessary resource for molecular breeding in walnut, providing a complete enumeration of all the coding genes that underlie agronomic traits. Molecular breeding is based on selection of superior allelic variants at these genetic loci versus selecting on the phenotype that is determined by both genetic and environmental variation (Phenotype = Genotype + Environment). Molecular breeding has the potential to be much more precise in the improvement of agronomic traits. Molecular breeding is applied in dozens of agricultural species where the genome sequence and other genomic resources have been developed.

PROCEDURES, RESULTS, AND DISCUSSION

'Chandler' reference genome sequence v1.0

The genome sequence of the cultivar 'Chandler' has been completed by a team of researchers at UCDavis, Johns Hopkins University and University of Connecticut. This is essentially the same group that recently completed the sequencing of the loblolly pine (*Pinus taeda*) genome. The experience gained in sequencing very large genomes has made sequencing of the much smaller walnut genome relatively quick and inexpensive. Additionally, to obtain a well annotated genome (identification of the different elements of the walnut genome), an assembly of the all expressed genes sequences (transcriptome) must be used. These sequences that represent expressed genes on specific tissues and/or expressed genes in response to various abiotic or biotic stress were generated by the Dandekar lab under earlier funding from the CWB and have a great value for the completeness of the walnut genome.

High-throughput genotyping platform for Lateral Bearing

Lateral bearing is strongly correlated with precocity, a major determinate of yield in young trees, and a key selection criterium in breeding. We developed a new genotyping platform to assist the Walnut Breeding Program in early screening for lateral versus terminal bearing trees. The LGC genotyping technology (KASP™ genotyping) that was used provides a fast, accurate, flexible and cost-effective genotyping service. Using this technology it was possible to genotype 3200 samples from the breeding program at very low cost (~1$/sample). DNA markers identified in previous work by Dvorak, Luo and Aradhya were refined and used for this platform. They worked well in the Chandler x Idaho mapping population and proved suitable for a sub-set of the current breeding populations but this work also showed that additional fine-mapping of the lateral bearing gene will be necessary in order to reliably apply genotyping technology across all breeding populations. To do the fine-mapping, Jan Dvorak is proposing to create a mapping population with 400 self-pollinated 'Chandler' nuts obtained from an isolated orchard with no other pollen source.

Breeding value estimation

Estimated breeding values (BVs) define the genetic value of each individual in a breeding population. We can use them as a source for objectively choosing which trees are the best candidates to select for breeding to produce the next generation of offspring. Estimation of BVs for the main parents in the breeding program has begun by focusing on the six most important traits of interest: yield, harvest date, lateral bearing, leafing date, color and blight. BVs will be estimated using 15 different families with a total of 3,048 individuals, and phenotypic data collected for almost 16 years. The most complete pedigree of the breeding program possible was assembled as a preliminary step for the estimation of breeding values. BVs will be used in Year 3 of the project to discover marker x trait associations that can be used in a marker breeding program.

Plant material collections for genetic diversity analysis

Several sets of samples have been collected that will be used in future phases of this project, including:

1. Species diversity:
   DJUG 29.11 (*J. microcarpa*)
   DJUG 11.03 (*J. cathayensis*)
   DPTE 1.09 (*Pterocarya* spp.)
   'Rawlins' (*J. hindsii*)

2. Parents of cultivars and selections in the Walnut Breeding Program:
   75 individuals (founders)
3. *J. regia* diversity:
   25 accessions have been selected, five from each of the five diversity clades for *J. regia*
4. *J. regia* germplasm collection:
   ~850 accessions from the USDA germplasm repository.

CONCLUSIONS

Phase I was completed in 2014. A well-annotated genome sequence and the complete transcriptome assembly were obtained. These are necessary resources for Phase II, which is to discover the genetic variation in *J.regia* for scion molecular breeding. In addition, four related *Juglans* species are also being sequenced, which will support the goals of the USDA/SCRI funded rootstock improvement project.