Structural Polymorphisms as Causes of Heterosis in *Populus*

Upendra Kumar Devisetty[1], Kori Ault[1], Analeslie Martinez[1], Alice Morel[1], Alejandro Riveros Walker[2], Luke M. Evans[2], Sandra Simon[2], Stephen DiFazio[2], Austin Himes[3], Brian Stanton[3], Laura Holloway[4], Todd Rosenstiel[4] and Steven H. Strauss[1]

[1] Oregon State University, Corvallis OR; [2] West Virginia University, Morgantown WV; [3]Greenwood Resources, Portland, OR and [4]Portland State University, Portland, OR

United States Department of Agriculture
National Institute of Food and Agriculture

Abstract

Heterosis refers to the superior performance and stress tolerance of highly heterozygous groups or individuals. It is widely known in maize (Fig 1), and many perennial biomass programs also employ heterosis either in the form of polyploids (e.g., *Miscanthus*) or in the form of interspecific hybrids (e.g., *Populus*, *Salix*, *Acacia* and *Eucalyptus*). This project focuses on the association of heterosis with structural polymorphisms (SPs), primarily gene insertions and deletions, in wild trees and interspecific hybrids in *Populus*. We are analyzing existing genomic and phenotypic data from three populations, and generating substantial new data on the extent of SPs. We established a new plantation of F_1 hybrid families that had a survival rate of 98%. We made progress in assembling reference genomes of *P. deltoides*, *P. maximowizii* and *P. koreana* required to align sequences of tree sequences in our F_1 population. DNA extractions were made from about one-third of the study population, and library preparation protocol and sequence-capture assay are nearly optimized for high-throughput sequencing. We made substantial progress in phenotypic assessments, including tree size, leaf size, leaf density, chlorophyll content, chlorophyll fluorescence, and net assimilation/respiration under ambient and high temperature stress. Preliminary phenotypic assessments indicate that there is significant variance among different genotypes, families, and/or cross-types for most of the phenotypes measured. We plan to do more extensive phenotypic measurements in future.

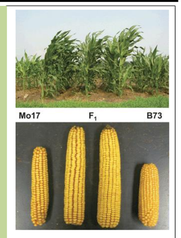


Figure 1. Classical heterosis in maize

Plant Materials

We are studying three types of populations:

1. Wild *P. trichocarpa* that was collected, resequenced, and established in several study plantations by the DOE Bioenergy Science Center project. We are focusing on the "Marchel" population in Oregon (Fig 2). This population has been extensively characterized genetically and phenotypically, and we are continuing to generate new phenotypic data from it.
2. A pseudo-backcross plantation, also supported by the DOE, that is growing in Boardman, Oregon (Fig 2). It is comprised of outcross progeny from a cross of (*P. trichocarpa x deltoides) x P. deltoides*.
3. An F1 plantation that was established in Jefferson, Oregon and Boardman, Oregon by GreenWood Resource (GWR) (Fig 2).

The three populations selected for the project will serve specific roles (Table 1).

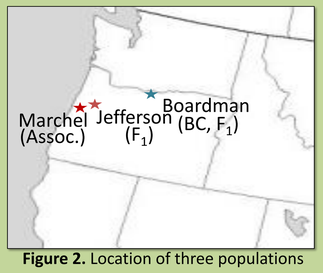


Figure 2. Location of three populations

Table 1. Rationale for study populations

Species	Nature of population	Purpose	No. of genotypes
P. trichocarpa	GWAS/Association	To examine independent SP effects because of low LD	1,100
(*P. trichocarpa* x *P. deltoides*) × *P. deltoides*	Backcross (BC)	To examine the effects of both intra- and inter- specific SP effects	337
P. deltoides × *P. nigra* (DN), *P. deltoides* × *P. maximowiczii* (DM) and *P. deltoides* × *P. koreana* (DK)	F_1 hybrids	To evaluate the SPs for Marker-Aided Selection in operational hybrid breeding	371

Experimental Methods

Phenotyping

The main phenotypes under study are (Figure 3):

1. Tree height & diameter
2. Leaf area & specific mass
3. Stomatal density
4. Chlorophyll content and dark/light fluorescence
5. Assimilation and respiration at ambient and high temperature stress

- **Experiments:** For each population we will obtain phenotypes over two growing seasons.
- **Productivity traits:** Productivity of the trees will be estimated as "volume index," the product of tree height x basal diameter squared.
- **Fluorescence traits:** We will determine leaf chlorophyll content and fluorescence parameters (Fv/Fm, PSII efficiency, and steady-state fluorescence) using portable chlorophyll meters and hand-held flourometers.
- **Stress related traits:** Stress will be imposed by exposing leaves to a series of increasing leaf temperatures and net assimilation, stomatal conductance, electron transport, and dark respiration will be determined at each temperature.

Figure 3. Phenotyping of plant material. A) leaf data collection at the Boardman F_1 population showing personnel taking dermal peels for stomatal counts, scanning leaves for area analysis, and taking SPAD readings; B) measuring fluorescence in the association population; C) sample dark-adapted fluorescence reading; D) clip used to dark-adapt leaves; E) measuring assimilation and respiration with a Licor at the Jefferson F_1 population; and F) a scanned *P. trichocarpa* leaf for area analysis.

Genotyping

1. **Sequencing**: A discovery population of 24 individuals per species will be sequenced to a depth of 15X for discovery of segregating SPs.
2. **Dual-barcoding system :** Illumina-compatible dual bar-coded libraries with 500 to 800 bp inserts will be prepared for each template using custom oligonucleotide (Fig. 4A).
3. **Sequence-capture assay:** For genotyping hybrid and BC populations, we will use a sequence-capture assay. The sequence capture probes will be designed to flank the insertion break points for the SPs such that a codominant genotype can be determined from sequence data (Fig. 4B).
4. **Multiplexing:** Libraries will be multiplexed in sets of 144 (each with a unique bar code) and sequenced with paired 150 bp reads on an Illumina HiSeq. This will provide approximately 10X depth per target.
5. **Non-coding regions:** We will capture a large number of non-coding sequences evenly spread across the genome to enable comparison of the performance of random SNP to selected SPs in explaining variance in heterosis in breeding populations.

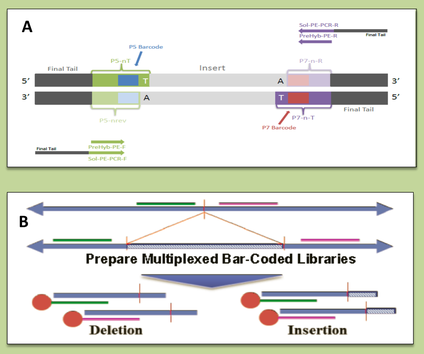


Figure 4. Genotyping for SP discovery using high - throughput sequencing. A) dual barcoding system; B) sequence-capture probe design assay.

Bioinformatics

1. **Reference genomes**: We are developing reference genomes for *P. deltoides*, *P. maximowizii*, *P. nigra* and *P. koreana* using paired end and mate pair sequencing on the Illumina HiSeq as well as PACBio long read sequencing. Assembly will be carried out using a combination of *denovo* assembly software, including ABySS, ALLPATHS-LG, and hybrid approaches such as those implemented in the Celera WGS pipeline.
1. **SP & SNP calling:** We will use GenomeSTRiP to identify large SPs and mpileup and bcftools to identify SNPs and small insertion/deletion polymorphisms, and calculate genotype likelihoods. We will then use BEAGLE to perform genotype imputation.
2. **Genomic selection:** We will use a genomic selection approach to estimate genotypic values for growth, leaf and physiological measures.


Jefferson F1 population


Boardman F1 population

Results

DNA extractions

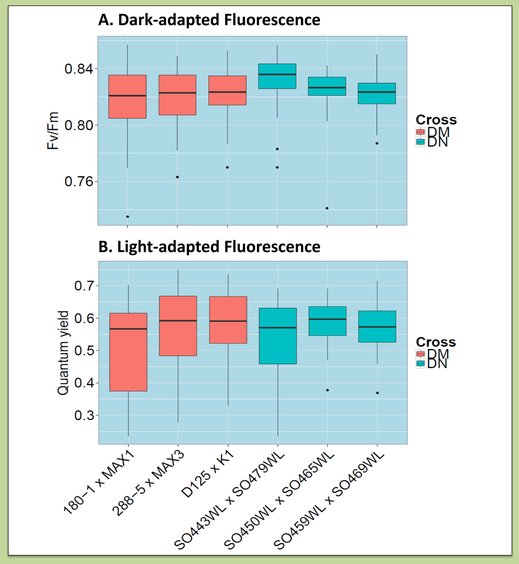
- Good quality DNA has been extracted from 174 F_1 genotypes so far using a CTAB DNA extraction protocol (Table 2).
- We will collect samples from 197 additional F_1 genotypes for DNA this spring (Table 2).
- Will collect samples from all (337) of the outbred hybrid backcross at Boardman this February and extract DNA from them.

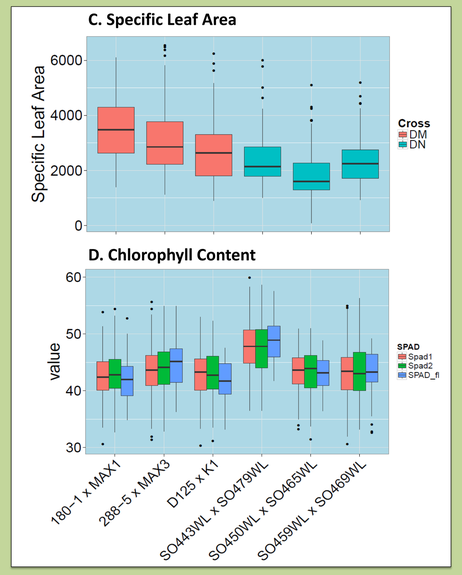
Cross	TaXa	Number of clones	Number of clones DNA isolated	Number of clone need for DNA preparation
180-1xMax1	DxM	88	54	34
288-5xMax3	DxM	80	49	31
D125xK1	DxM	61	37	24
SO450WLxSO465WL	DxN	55	14	41
SO459WLxSO469WL	DxN	49	13	36
SO433WLxSO479WL	DxN	38	7	31
Total		**371**	**174**	**197**

Table 2. Summary of DNA extractions from F_1 genotypes

Phenotypic assessment

Preliminary phenotypic analyses showed evidence of genetic variation in nearly all traits measured. Leaf fluorescence parameters (Fv/Fm and yield) varied among crosses results (Fig. 5A-B). Specific leaf area (SLA) varied among families and crosses (Fig. 5C). We determined that SPAD meter measurements of chlorophyll could be taken on any leaf samples, on or off the tree, with essentially identical results and the SPAD meter measurements varied among families and crosses (Fig. 5D). Leaf photosynthesis and respiration varied among temperatures and families (Fig. 5E).





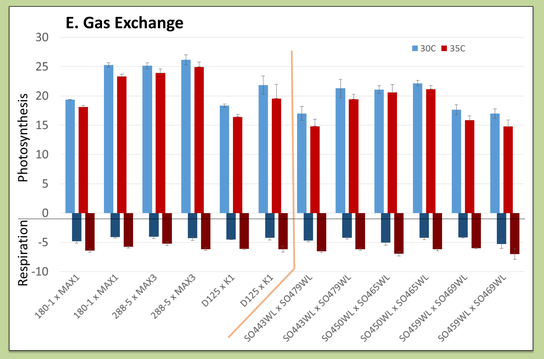


Figure 5. Phenotypic assessment of F_1 Jefferson plant material A) Fv/Fm fluorescence measurements for 6 families; B) quantum yield measurements for 6 families; C) SLA (Leaf area (cm2) / leaf punch weight (g) for 6 families; D) SPAD measurements for three sources of leaf material (SPAD1 and SPAD2 are measurements taken on leaves after they were harvested, and SPAD_fl measurement is taken on the trees); E) photosynthesis and respiration rates for 6 families consisting of 2 genotypes taken at 2 different temperatures.

Whole Genome Association Analysis (WGAS) for Dark-Adapted Chlorophyll Fluorescence

Objective: We gathered dark-adapted chlorophyll fluorescence data (Fv/Fm) in the *P. trichocarpa* association population in July of 2014 in order to evaluate the genetic bases of this high-throughput physiologically-informative genotype.

Method

- Dark adaptation clips were applied in the late afternoon for measurements following morning or late morning for afternoon measurements
- Using a model with F_0 and position (Thin Plate Spline adjustment, Figure 6A) as covariates, the broad sense heritability was 0.56.
- We performed a whole genome association analysis for corrected BLUPs from rep 3 only using 6 million SNPs and mixed model analysis with Principal Components for population structure and a kinship matrix as covariates (Fig. 6B). Three loci showed strong associations with the trait (Fig. 6C-E).

Genes Near Association Peaks

- Chr03: flavonol-7-O-rhamnosyltransferase involved in the formation of rhamnosylated flavonols, responsive to UV light (Fig. 6C)
- Chr04: Cysteine-type peptidases (Fig. 6D)
- Chr06: Protein of unknown function (DUF1278) nearest; a glycosyltransferase is also ~15Kb away (Fig. 6E)

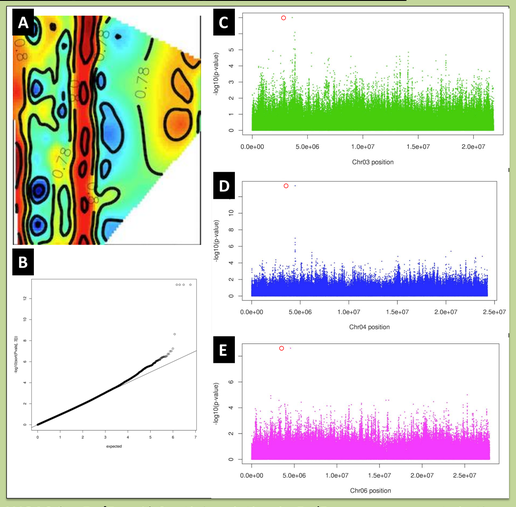


Figure 6. WGAS for Fv/Fm. A) Spatial variation in Fv/Fm measurements in the field, represented by a contour heat map generated using the Thin Plate Spline function in R. B) Q/Q plot for Fv/Fm associations. C-E) Manhattan plots depicting significant associations (circled) for Chr 3, 4, and 8 in *P. trichocarpa*.

Future research

1. Extensive phenotyping, including chlorophyll fluorescence, gas exchange as a function of temperature, leaf area, specific leaf mass, stomatal density, and height/DBH.
2. Indel discovery in the hybrid and backcross populations; assessment of different software methods for detection.
3. Optimizing bait-capture probe design for indel genotyping. These will then be used in conjunction with with high-throughput DNA library production to efficiently generate SP genotype data for all study populations.
4. Developing genomic selection models to assess whether SPs could be used to accelerate breeding beyond what could be achieved by randomly-selected SNP polymorphisms.

Acknowledgements

This project is supported by the USDA National Institute of Food Agriculture - Institute of Bioenergy, Climate and Environment, competitive grant no. 2013-67009-21008. We thank Amy Kloko, Alice Morel and Analeslie Martinez for their help in phenotypic analysis in the field, GreenWood Resources (GWR) for planting and management of plantations at Jefferson and Boardman, and personnel from the Bioenergy Science Center for providing materials and data.