GWAS Identification of Loci Associated with Rooting in *Populus*

Anna C. Magnuson[1], Cathleen Ma[1], Christine Zawaski[2], Wellington Muchero[3], Jialin Yuan[4], Yuan Jiang[4], Fuxin Li[5], Jonathan R. Cumming[2], Stephen P. DiFazio[2], and Steven H. Strauss[1].

[1]Department of Forest Ecosystems and Society, Oregon State University, Corvallis OR; [2]Department of Biology, West Virginia University, Morgantown WV; [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge TN; [4]Department of Statistics, Oregon State University, Corvallis OR; and [5]Department of Electrical Engineering and Computer Science, Oregon State University,

Abstract

Two independent experiments were performed at Oregon State University and West Virginia University to study genetic variation and quantitative trait nucleotides (QTNs) associated with rate and intensity of root production in *P. trichocarpa*. Cuttings were taken from a field trial in Corvallis, Oregon with 1,035 wild *Populus trichocarpa* (black cottonwood) hybrids with resequenced genomes. This population displays low linkage disequilibrium making it ideal for genetic mapping, and has been effective in identifying genomic locations of genes affecting a variety of adaptive and productivity traits. Cuttings were rooted at both locations and rooting was determined manually and by and machine vision estimation.

GWAS analysis of OSU data yielded 45 strongly associated SNP loci below a false discovery threshold of 0.05. GWAS of principal component variables yielded another 28 SNP. SNP exploration using the Phytozyome *P. trichocarpa reference genome (v 3.0)* identified a number of SNP-proximal loci whose annotations based on Arabidopsis protein homologs appear to have functions related to rooting, as well as a number without obvious associations to known genes or physiological functions. GWAS analyses of the similar genetic material at WVU will help to identify those loci most likely to be biologically significant, whose locations and functions will then be analyzed in detail.

Project Overview

<u>Oregon State University</u>

- Dormant cuttings were collected from 537 wild genotypes in a plantation in Corvallis, OR
- Cuttings rooted in water and soil treatments, two blocks per treatment, 4 genotype replicates total
- Data collected manually for height, diameter, and rooting density score
- Water rooting treatments were photographed for root growth after 3 and 6 weeks
- Water-rooting images were analyzed for root area and average stem diameter using machine vision learning software
- Data underwent principal component analysis to generate new variables for GWAS

<u>West Virginia University</u>

- 700 wild black cottonwood genotypes were studied
- Cuttings were rooted in nutrient thin film (NTF) culture in 0.5 mM $Ca(NO_3)_2$, pH 5.6 in greenhouse with supplemental lighting
- Cuttings were rooted in groups of 192 in NTF systems ("gutters"), with ~600 multiple experiments
- Once roots > 2.5 cm, they were photographed and images were evaluated using ImageJ
- Data collected manually includes:
  - Days to Root Initiation (DRI)
  - Longest Root Length (LRL)
  - Root Density (RD=TRL/LRL)
  - Root Growth Rate (RGR=TRL/DRI)
- GWAS analysis is still underway

Phenotyping

Figure 1. Experimental setup at OSU. A) Water treatment cuttings after initial planting. B) Soil treatment cuttings. C) Rooting after three weeks.

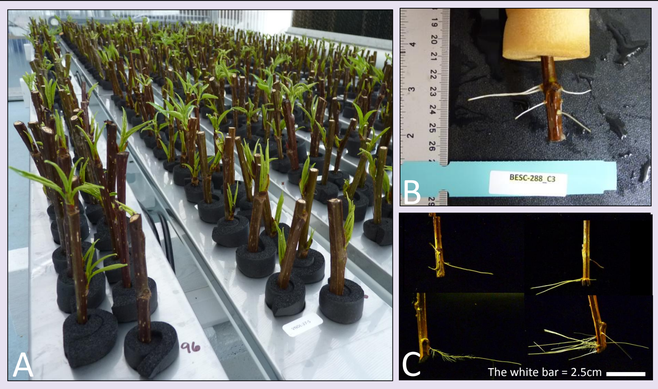


Figure 2. Rooting setup at West Virginia University
A) Cuttings lined up in gutters based on nutrient treatment. B) Root imaging. Rooting took 1-2 weeks and cuttings were imaged when roots measured 2.5 cm. C) ImageJ analysis.

Machine Vision Analysis

Color-based clustering was used on each image to segment out the background and leaf regions. Then the ruler, stem, and root regions are segmented based on morphology information. Based on these segmentations, root area (mm^2) and average stem diameter measurements were computed for 423 genotypes to supplement visible scores of rooting density.

Figure 3. <u>Left panel</u>: Original image of water treatment cuttings. <u>Right panel</u>: Segment masks produced by color-based and morphology segmentation. Root area and mean stem diameters are computed from the segmentations.

Heritability

Calculations showed that ordinal heritability estimates were of similar magnitude between OSU and WVU across different rooting platforms, with genetic causes explaining about 10-25 % of the variance.

Variable	h^2 Root (binary)	h^2 Root (ordinal)
Combined	0.16	0.13
Water Treatment	0.94	0.18
Soil Treatment	0.09	0.15

Table 1. Generalized linear mixed model with genotype as random effects provides heritability estimates for OSU binary root data (roots either exist or not) and ordinal root data (root scoring spectrum from 0-3). Heritabilities were calculated for combined treatments, as well as individually for water and soil with 537 genotypes included.

Variable	h^2
Days to Root Initiation (DRI)	0.25
Longest Root Length (LRL)	0.17
Total Root Length (TRL)	0.15
Density Parameter (TRL/LRL)	0.12
Root Growth Rate (RGR)	0.17

Table 2. Heritability estimates for five measures of rooting from WVU experiments with 688 genotypes. Raw data evaluated using mixed models accounted for run (set in time) and gutter variances.

Principal Components Analysis

Principal components analysis was performed on 423 OSU water-treatment genotypes using the variables height, diameter, average stem diameter, and root area.

PC1 is a measure of the relationship between height and diameters. We are most interested in PC2 and PC3, which are both indicative of the relationship between height and root area.

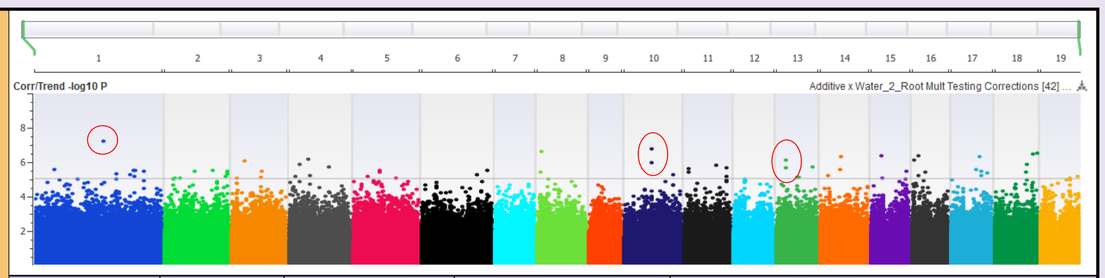
	PC1	PC2	PC3	PC4
Height	0.44	-0.43	0.77	0.15
Diameter	0.65	0.19	-0.13	-0.73
Ave. Stem Diameter	0.61	0.25	-0.34	0.66
Root Area	0.10	-0.85	-0.52	-0.04

Genome Wide Association Study

GWAS using efficient mixed model association (EMMA), accounting for kinship, was used to correlate a panel of 8.2 million SNPs to phenotypic variation data and associated principal component scores. This population has a panel of 29 million SNPs representing a marker every 17-base across the genome and rapidly decaying linkage disequilibrium that falls below 0.2 within 3Kb.

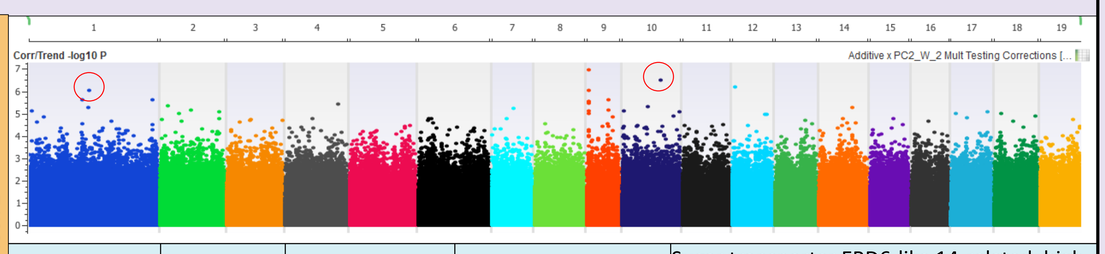
Over 70 SNP loci were strongly associated with one or more traits at a false discovery rate (FDR) threshold of 0.05. Significance thresholds were set at negative log p-values of 6 for PC score runs and 5.5 for raw data.

Significant SNP loci were investigated further using phytozome, *Populus trichocarpa* v3.0 reference genome. Below, some of the SNPs of interest are circled in red and their chromosome, position, significance value, nearby gene(s) annotations, and further details given.



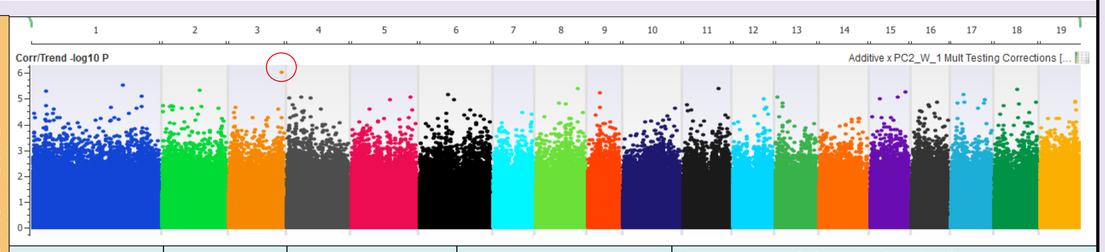
Raw Root data, Water Block 2

Chromosome	Position	-log10 P-value	Gene	Notes/Description
1	27299510	7.20	Potri.001G264500.1	SNP in gene, UBX domain, Arabidopsis homolog PUX1 gene, loss of which is known for accelerated growth of roots
10	10735457	5.94	Potri.010G081300.1	Protein Argonaute 10, known for small RNA-directed gene slicing, expressed in roots
10	10808758	6.71	Potri.010G081900.1	Expression in roots, Arabidopsis homolog dyggve-melchior-clausen syndrome protein
13	3730879	6.09	Potri.013G051100.2	SNPs very close, both hit in gene. Zinc finger FYVE domain containing protein, high expression in root tips
13	3730867	5.61		



PC2, Water Block 2

1	23802203	6.04	Potri.001G226600.1	Sugar transporter ERD6-like 14-related, high expression in roots
10	14966452	6.47	Potri.010G136100.1	Close to genes, first gene listed is a basic helix-loop-helix family protein, similar to bHLH transcription factor GBOF-1 with high root expression, Arabidopsis homolog regulates growth
			Potri.010G135900.1	WPP domain-interacting tail-anchored protein differentially downregulated in root in urea vs. standard, Arabidopsis homolog WIT2, association in root tip cells



PC2, Water Block 1

3	20611278	6.02	Potri.003G205000.1	No annotated domain but high root expression
			Potri.003G205100.1	Arabidopsis homolog ZTP29, involved in salt stress response and can be expressed in root tissue and mutations of gene affect widening of cells in the root elongation zone

Summary

- Machine vision learning yielded effective methods for acquiring root area measurements on high throughput experiments
- Independent heritability estimates for various rooting parameters using linear mixed models at OSU and WVU showed a similar range of values
- Allele-additive GWAS panels run against raw rooting data and principal component data yielded many statistically significant SNPs. Upon further investigation, some SNPs were found to be in or around genes with functional annotations in poplar and/or Arabidopsis for protein homologs that may have functions related to rooting

Acknowledgements

We thank the National Science Foundation Plant Genome Research Program for support (IOS # 1546900, Analysis of genes affecting plant regeneration and transformation in poplar), as well as the Department of Agriculture (National Institute for Food and Agriculture contract 2014-67013-21657 (JRC). We also thank the members of the Tree Biosafety and Genomics Research Cooperative at OSU for its long term support.