GWAS Identification of Loci Associated with Rooting in Populus

Anna C. Magnuson1, Cathleen Ma1, Christine Szaewski, Wellington Mucheno, Jialin Yuan, Yuan Jiang, Xinlin Li, Jonathan R. Cummings, Stephen P. Difazio2, and Steven H. Strauss1

1Department of Forest Ecosystems and Society, Oregon State University, Corvallis OR; 2Department of Biology, West Virginia University, Morgantown WV; 3Biosciences Division, Oak Ridge National Laboratory, Oak Ridge TN; 4Department of Statistics, Oregon State University, Corvallis OR; and 5Department 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 traits nulceotides (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 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 Phytome PO 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 of unobserved 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

Oregon State University
- 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

West Virginia University
- 700 wild black cottonwood genotypes were studied
- Cuttings were rooted in nutrient thin film (HTF) culture in 0.5 mm Ca(NO3)2, pH 5.6, in greenhouses with supplemental lighting
- Cuttings were rooted in groups of 192 in TNG systems ("gutters"), with 600 multiple experiments
- Once roots > 2.5 cm, they were photographed and images were evaluated using ImageJ
- Collected data manually included:
  - 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.

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 (mm²) and average stem diameter measurements were computed for 423 genotypes to supplement visible scores of rooting density.

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

Machine Vision Analysis

GWAS 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. Significant threshold values 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 Phytome PO trichocarpa v 3.0 reference genome. Below, some of the SNPs of interest are circled in red and their chromosome, position, significance value, nearby genes, annotations, and further details given.

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 data showed heritability estimates were of similar magnitude between OSU and WVU across different rooting platforms, with genetic causes explaining about 10% variance.
- Allele additive GWAS panels run against raw rooting data and principal component data yielded many genetically significant SNPs. Upon further investigation, some SNPs were found to be in or around genes with functional annotations in poplar 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 2014-67021-23657 [BURCH]). We also thank the members of the Oregon State Biosyslo and Genomics Research Cooperative at OSU for its long term support.